IV

Primate Lentivirus Complete Genomes

Alignment of PLV Complete Genomes

Complete genomes of HIV-1, HIV-2, and SIV sequences are represented in the alignment. For the selection, several of the major subtypes and a few CRFs of the HIV-1 M group were chosen, plus one N group and two O group sequences. Most of the SIV-CPZs were included, including new sequences from wild caught chimpanzees. Three HIV-2 A group and two HIV-2 B group were included along with the single genomes sequenced to date for HIV-2 G and HIV-2 U. Three SIV(Mac-SMM-MNE) genomes were included in this alignment and more are available in the HIV-2/SIV-SMM alignment. For the rest of the SIVs, all complete genomes were included. The principle in deciding which sequences to include in this alignment was to provide a representative sampling of all primate lentiviruses that have had complete or nearly complete genomes sequenced. Many HIV-1, HIV-2, SIVsmm and SIVmac genomes were not included, but they can be found in the HIV-1/SIVcpz and HIV-2/SIVsmm alignments. Also, many shorter sequences from PLVs are included in the PLV protein sequence alignments later in this compendium.

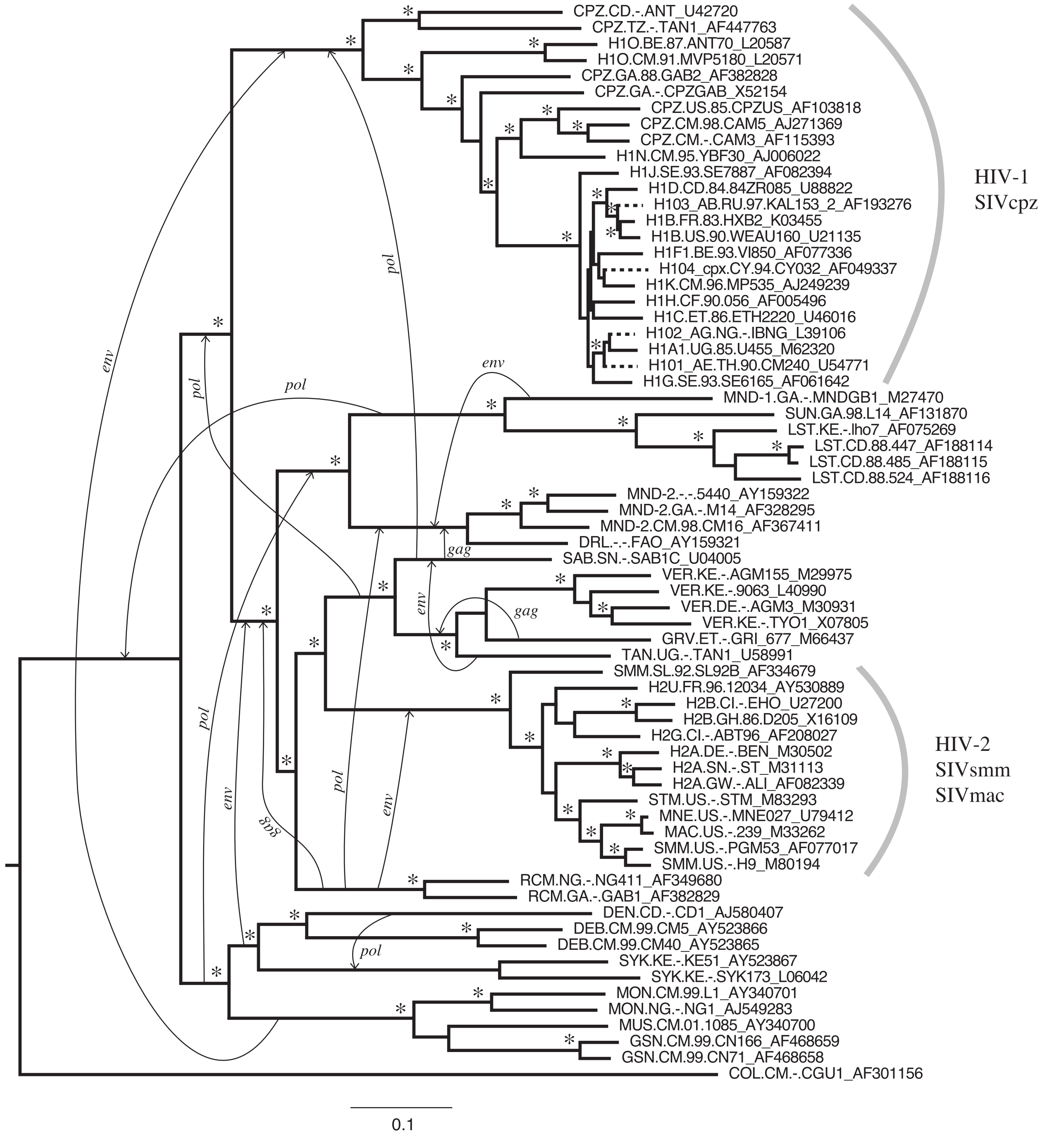
The alignment was based on the previous version published last year, by a hidden Markov model using HMMER and manual editing using the programs BioEdit and Se-Al. It is codon-aligned, meaning that the correct translation reading frame has been maintained as much as possible; in the case of overlapping frames, obviously one will not translate correctly. We have tried to "reset" the alignment so that each new gene starts in the first frame; this means in some places empty columns have been inserted. The HIV-2/SIVsmm vpx gene is postulated to be a duplication of the vpr gene (Tristem et al., 1990) and thus there may be two alternative alignments of this region of the genome, as there are for the duplicated stem-loops of the TAR element.

Like last year, the annotation is based on known protein coding regions in HIV-1 and on annotations found in SIV sequence database entries. The protein cleavage sites that create Gag p17, Gag p24 and other mature peptides from the Gag and Gag-Pol precursor polyproteins have been experimentally determined for HIV-1 and at least one strain of HIV-2; the study of analogous cleavages in SIV polyproteins have not been published. Two representative genomes have been translated: HIV-1 subtype B strain HXB2, and SMM-239. The translations are provided as a visual aid for finding landmarks in the genomes.

The tree in Figure 1 is a demonstration of the apparent phylogenetic relationships between the PLVs. As has been pointed out by several authors, several of the PLV sequences appear to be recombinant. For instance, SIVrcm has been shown to display discordant clustering in different genomic regions (Beer *et al*., 2001), SIVcpz has been described to have hybrid origins (Bailes *et al*., 2003), and SIVsab was described as mosaic thirteen years ago (Jin *et al*., 1994). As can be seen in Figure 1, trees based on *gag*, *pol* and *env* gene sequences differ in their cladistic classification of the PLVs. The topology shown is derived from the complete genome alignment in this section, and the thin line arrows indicate alternative clusterings as inferred in *gag*, *pol* and *env* gene trees. Note that this is not intended to be an exhaustive analysis of potential recombination, but merely a display of the problematic relationships among PLVs. The direction of the arrows is in most cases arbitrary and does not indicate direction of potential crossover events or cross species transmissions. Although important, similar problems within the HIVs, such as the recombinant structure of HIV-1 group N genomes (Roques *et al*., 2004) and the multiphyletic histories of the CRFs, are not shown in Figure 1. The resulting network structure is likely the result of multiple reasons, including true recombination between SIVs, lack of appropriate parental representatives, effects of mutational rate differences, possible convergent evolution, and inadequate phylogenetic reconstruction methods. For instance, early attempts to describe mosaic structures of some SIVs have been revised when virus from other simian species have become available. Also note that even though bootstrap support is strong for many clusters, sequences within those clusters are involved in conflicting topologies, and is thus not a method to reveal uncertainties resulting from any of these problems. Therefore, any tree containing PLV sequences should be interpreted with caution. Recently, several reviews and analyses on the PLVs have been published (*e.g*., Bailes *et al*., 2002; Clewley, 2004; Foley, 2000; Peeters and Courgnaud, 2002; Salemi *et al*., 2003).

The tree in Figure 1, as well as the separate *gag*, *pol* and *env* gene trees and the gapstripped alignments from which the trees were calculated are available at our website, http://hiv.lanl.gov.

Figure 1. Genetic relationships among Primate Lentiviruses (PLVs). The tree shows the inferred relationships in the following PLV complete genome alignment. The thin line arrows indicate the main differences inferred from the gag, pol and env gene fragments from the same alignment. Dashed tree branches indicate inferred positions for known recombinant HIV-1 circulating forms. The resulting network should therefore not be interpreted as a singular phylogenetic history of the PLVs. Asterisks indicate bootstrap support ≥70% in the complete alignment. The scale bar at the bottom shows genetic distance according to a F84 model with relative rates. The tree shown was constructed from a gapstripped version of the DNA alignment in this section of the compendium, leaving 5327 unambiguously aligned positions. A neighbor-joining tree, calculated using DNADIST (F84 model) and NEIGHBOR in the PHYLIP package, was used as a guide to calculate relative site rates. The relative site rates were estimated using a generalized version of DNArates (Korber et al., 2000) with eight categories (0.133; 0.287; 2.373; 6.826; 19.634; 56.472; and 115.653). Finally, these rates were given to fastDNAml (version 1.2.2) and a maximum likelihood tree was inferred from the gapstripped alignment (transition/ transversion ratio = 2). The tree space was searched by both Jumble and Global Rearrangements (G 3 3). Bootstrap support was calculated using SEQBOOT with the above 8 rate categories on 1000 resampled sets, and DNADIST and NEIGHBOR in PHYLIP (version 3.6.2). The separate gag, pol and env alignments were created using a new version of GeneCutter (see www.hiv.lanl.gov) and trees were calculated the same way as in the complete alignment with optimized site rates (not shown). The Colobus SIVcol sequence, which is the most divergent PLV sequence, was used as outgroup to visualize the tree using TreeEdit (version1.0) and Adobe Illustrator.


CPZ.CD.-.ANT_U42720
CPZ.TZ.-.TAN1_AF447763
H1O.BE.87.ANT70_L20587
H1O.CM.91.MVP5180_L20571
CPZ.GA.88.GAB2_AF382828
CPZ.GA.-.CPZGAB_X52154
CPZ.US.85.CPZUS_AF103818
CPZ.CM.98.CAM5_AJ271369
CPZ.CM.-.CAM3_AF115393
H1N.CM.95.YBF30_AJ006022
H1J.SE.93.SE7887_AF082394
H1D.CD.84.84ZR085_U88822
H103_AB.RU.97.KAL153_2_AF193276
H1B.FR.83.HXB2_K03455
H1B.US.90.WEAU160_U21135
H1F1.BE.93.VI850_AF077336
H104_cpx.CY.94.CY032_AF049337
H1K.CM.96.MP535_AJ249239
H1H.CF.90.056_AF005496
H1C.ET.86.ETH2220_U46016
H102_AG.NG.-.IBNG_L39106
H1A1.UG.85.U455_M62320
H101_AE.TH.90.CM240_U54771
H1G.SE.93.SE6165_AF061642
HIV-1
SIVcpz
MND-1.GA.-.MNDGB1_M27470
SUN.GA.98.L14_AF131870
LST.KE.-.lho7_AF075269
LST.CD.88.447_AF188114
LST.CD.88.485_AF188115
LST.CD.88.524_AF188116
MND-2.-.-.5440_AY159322
MND-2.GA.-.M14_AF328295
MND-2.CM.98.CM16_AF367411
DRL.-.-.FAO_AY159321
SAB.SN.-.SAB1C_U04005
VER.KE.-.AGM155_M29975
VER.KE.-.9063_L40990
VER.DE.-.AGM3_M30931
VER.KE.-.TYO1_X07805
GRV.ET.-.GRI_677_M66437
TAN.UG.-.TAN1_U58991
SMM.SL.92.SL92B_AF334679
H2U.FR.96.12034_AY530889
H2B.CI.-.EHO_U27200
H2B.GH.86.D205_X16109
H2G.CI.-.ABT96_AF208027
H2A.DE.-.BEN_M30502
H2A.SN.-.ST_M31113
H2A.GW.-.ALI_AF082339
STM.US.-.STM_M83293
MNE.US.-.MNE027_U79412
MAC.US.-.239_M33262
SMM.US.-.PGM53_AF077017
SMM.US.-.H9_M80194
HIV-2
SIVsmm
SIVmac
RCM.NG.-.NG411_AF349680
RCM.GA.-.GAB1_AF382829
DEN.CD.-.CD1_AJ580407
DEB.CM.99.CM5_AY523866
DEB.CM.99.CM40_AY523865
SYK.KE.-.KE51_AY523867
SYK.KE.-.SYK173_L06042
MON.CM.99.L1_AY340701
MON.NG.-.NG1_AJ549283
MUS.CM.01.1085_AY340700
GSN.CM.99.CN166_AF468659
GSN.CM.99.CN71_AF468658
COL.CM.-.CGU1_AF301156
env
pol
pol
pol
env
gag
env
gag
pol
env
gag
pol
env
pol
0.1

References

Bailes, E., Chaudhuri, R. R., Santiago, M. L., Bibollet-Ruche, F., Hahn, B. H., and Sharp, P. M. (2002). The evolution of primate lentiviruses and the origins of AIDS. In *The molecular epidemiology of human viruses* (T. Leitner, Ed.). Kluwer Academic Publishers, Boston.

Bailes, E., Gao, F., Bibollet-Ruche, F., Courgnaud, V., Peeters, M., Marx, P., Hahn, B., and Sharp, P. M. (2003). Hybrid origin of SIV in Chimpanzees. *Science* **300**, 1713.

Beer, B., Foley, B., Kuiken, C., Tooze, Z., Goeken, R. M., Brown, C. R., Hu, J., St. Claire, M., Korber, B., and Hirsch, V. M. (2001). Characterization of novel simian immunodeficiency viruses from Red-Capped mangabeys from Nigeria (SIVrcmNG409 and -NG411). *Journal of Virology* **75**, 12014–12027.

Clewley, J. P. (2004). Enigmas and paradoxes: the genetic diversity and prevalence of the primate lentiviruses. *Current HIV Research* **2**, 113–125.

Foley, B. (2000). An overview of the molecular phylogeny of lentiviruses. In *HIV Sequence Compendium* (C. Kuiken, B. Foley, B. Hahn, P. Marx, F. McCutchan, J. Mellors, J. I. Mullins, J. Sodroski, S. Wolinsky, and B. Korber, Eds.), pp. 35–43. Theoretical Biology & Biophysics, LANL, Los Alamos, NM.

Jin, M. J., Hui, H., Robertson, D. L., Muller, M. C., Barre-Sinoussi, F., Hirsch, V. M., Allan, J. S., Shaw, G. M., Sharp, P. M., and Hahn, B. (1994). Mosaic genome structure of simian immunodeficiency virus from West African green monkeys. *EMBO Journal* **13**, 2935–2947.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S., and Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789–1796.

Peeters, M., and Courgnaud, V. (2002). Overview of primate lentiviruses and their evolution in non-human primates in Africa. In *HIV Sequence Compendium* (C. Kuiken, B. Foley, E. Freed, B. Hahn, P. Marx, F. McCutchan, J. Mellors, S. Wolinsky, and B. Korber, Eds.), pp. 2–23. Theoretical Biology & Biophysics, LANL, Los Alamos, NM.

Roques, P., Robertson, D. L., Souquiere, S., Apetrei, C., Nerrienet, E., Barre-Sinoussi, F., Muller-Trutwin, M., and Simon, F. (2004). Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N. *AIDS* **18**, 1371–1381.

Salemi, M., De Oliveira, T., Courgnaud, V., Moulton, V., Holland, B., Cassol, S., Switzer, W. M., and Vandamme, A.-M. (2003). Mosaic genomes of the six major primate lentivirus lineages revealed by phylogenetic analyses. *Journal of Virology* **77**, 7202–7213.

Tristem, M., Marshall, C., Karpas, A., Petrik, J., and Hill, F. (1990). Origin of vpx in lentiviruses. *Nature* **347**, 341–342.

Table 1: Table of Primate Lentivirus (PLV) Sequences in the Complete Genome Alignments

Name	Accession	Country	Author	Reference
B.FR.83.HXB2	K03455	France	Wong-Staal, F	*Nature* **313**(6000):277–84 (1985)
D.CD.84.84ZR085	U88822	D.R.C.	Gao, F	*J Virol* **72**(7):5680–98 (1998)
F1.BE.93.VI850	AF077336	Belgium	Laukkanen, T	*Virology* **269**(1):95–104 (2000)
G.SE.93.SE6165	AF061642	Sweden	Laukkanen, T	*Virology* **247**(1):22–31 (1998)
H.CF.90.056	AF005496	C.A.R.	Murphy, E	*ARHR* **9**(10):997–1006 (1993)
J.SE.93.SE7887	AF082394	Sweden	Laukkanen, T	*ARHR* **15**(3):293–7 (1999)
K.CM.96.MP535	AJ249239	Cameroon	Peeters, M	*ARHR* **16**(2):139–51 (2000)
01_AE.TH.90.CM240	U54771	Thailand	Laukkanen, T	*J Virol* **70**(9):5935–43 (1996)
02_AG.NG.x.IBNG	L39106	Nigeria	Howard, TM	*ARHR* **10**(12):1755–7 (1994)
03_AB.RU.97.KAL153_2	AF193276	Russia	Liitsola, K	*AIDS* **12**(14):1907–19 (1998)
04_cpx.CY.94.CY032	AF049337	Cyprus	Gao, F	*J Virol* **72**(12):10234–41 (1998)
N.CM.95.YBF30	AJ006022	Cameroon	Simon, F	*Nat Med* **4**(9):1032–7 (1998)
O.BE.87.ANT70	L20587	Belgium	Vanden Haesevelde, M	*J Virol* **68**(3):1586–96 (1994)
O.CM.91.MVP5180	L20571	Cameroon	Gurtler, LG	*J Virol* **68**(3):1581–5 (1994)
CPZ.CM.05.SIVcpzEK505	DQ373065	Cameroon	Keele, BF	*Science* **313**(5786):523–6 (2006)
CPZ.CM.05.SIVcpzLB7	DQ373064	Cameroon	Keele, BF	*Science* **313**(5786):523–6 (2006)
CPZ.CM.05.SIVcpzMB66	DQ373063	Cameroon	Keele, BF	*Science* **313**(5786):523–6 (2006)
CPZ.CM.05.SIVcpzMT145	DQ373066	Cameroon	Keele, BF	*Science* **313**(5786):523–6 (2006)
CPZ.CD.90.ANT	U42720	D.R.C.	Vanden Haesevelde, M	*Virology* **221**(2):346–50 (1996)
CPZ.CM.98.CAM3	AF115393	Cameroon	Corbet, S	*J Virol* **74**(1):529–34 (2000)
CPZ.CM.98.CAM5	AJ271369	Cameroon	Muller-Trutwin, MC	*J Med Primatol* **29**(3-4):166–72 (2000)
CPZ.GA.88.GAB1	X52154	Gabon	Huet, T	*Nature* **345**(6273):356–9 (1990)
CPZ.GA.88.GAB2	AF382828	Gabon	Gao, F	*ARHR* **20**(12):1377–81 (2004)
CPZ.TZ.01.TAN1	AF447763	Tanzania	Santiago, ML	*J Virol* **77**(3):2233–2242 (2003)
CPZ.US.85.CPZUS	AF103818	U.S.A.	Gao, F	*Nature* **397**(6718):436–41 (1999)
GSN.CM.99.CN166	AF468659	Cameroon	Courgnaud, V	*J Virol* **76**(16):8298–309 (2002)
GSN.CM.99.CN71	AF468658	Cameroon	Courgnaud, V	*J Virol* **76**(16):8298–309 (2002)
A.GW.x.H2ALI	AF082339	Guinea-bissau	Azevedo-Pereira, JM	Unpublished
A.DE.x.H2BEN	M30502	Germany	Kirchhoff, F	*Virology* **177**(1):305–11 (1990)
A.SN.x.H2ST	M31113	Senegal	Kumar, P	*J Virol* **64**(2):890–901 (1990)
B.GH.86.H2D205	X61240	Ghana	Kreutz, R	*ARHR* **8**(9):1619–29 (1992)
B.CI.x.H2EHO	U27200	Cote d'ivoire	Rey-Cuille, MA	*Virology* **202**(1):471–6 (1994)
G.CI.x.H2ABT96	AF208027	Cote d'ivoire	Brennan, CA	*ARHR* **13**(5):401–4 (1997)
U.FR.96.H212034	AY530889	France	Damond, F	*ARHR* **20**(6):666–72 (2004)
MAC.US.x.239	M33262	U.S.A.	Kestler, H	*Science* **248**(4959):1109–12 (1990)
SMM.SL.92.SL92B	AF334679	Sierra leone	Chen, Z	*J Virol* **70**(6):3617–27 (1996)
SMM.US.x.PGM53	AF077017	U.S.A.	Novembre, FJ	*J Virol* **72**(11):8841–51 (1998)
MNE.US.x.MNE027	U79412	U.S.A.	Kimata, JT	*J Virol* **72**(1):245–56 (1998)
RCM.NG.x.NG411	AF349680	Nigeria	Beer, BE	*J Virol* **75**(24):12014–27 (2001)
RCM.GA.x.GAB1	AF382829	Gabon	Gao, F	Science 300(5626); 1713 (2003)
DRL.x.x.FAO	AY159321		Hu, J	*J Virol* **77**(8):4867–4880 (2003)
MND-1.GA.x.MNDGB1	M27470	Gabon	Tsujimoto, H	*Nature* **341**(6242):539–41 (1989)
MND-2.x.x.5440	AY159322		Hu, J	*J Virol* **77**(8):4867–4880 (2003)
MND-2.GA.x.M14	AF328295	Gabon	Souquiere, S	*J Virol* **75**(15):7086–96 (2001)
MND-2.CM.98.CM16	AF367411	Cameroon	Takehisa, J	*ARHR* **17**(12):1143–54 (2001)
DEB.CM.99.CM40	AY523865	Cameroon	Bibollet-Ruche, F	*J Virol* **78**(14):7748–62 (2004)
DEB.CM.99.CM5	AY523866	Cameroon	Bibollet-Ruche, F	*J Virol* **78**(14):7748–62 (2004)
MON.CM.99.L1	AY340701	Cameroon	Courgnaud, V	*J Virol* **77**(23):12523–12534 (2003)
MON.NG.x.NG1	AJ549283	Nigeria	Barlow, KL	*J Virol* **77**(12):6879–88 (2003)

TAL.CM.01.8023	AM182197	Cameroon	Liegeois, F	*Virology* **349**(1):55–65 (2006)
TAL.CM.00.266	AY655744	Cameroon	Liegeois, F	*Virology* **349**(1):55–65 (2006)
MUS-1.CM.01.1085	AY340700	Cameroon	Courgnaud, V	*J Virol* **77**(23):12523–12534 (2003)
SAB.SN.x.SAB1C	U04005	Senegal	Jin, MJ	*EMBO J* **13**(12):2935–47 (1994)
GRV.ET.x.GRI_677	M66437	Ethiopia	Fomsgaard, A	*Virology* **182**(1):397–402 (1991)
TAN.UG.x.TAN1	U58991	Uganda	Stivahtis, GL	*Virology* **228**(2):394–9 (1997)
VER.DE.x.AGM3	M30931	Germany	Baier, M	*Virology* **176**(1):216–21 (1990)
VER.KE.x.9063	L40990	Kenya	Hirsch, VM	*J Virol* **69**(2):955–67 (1995)
VER.KE.x.AGM155	M29975	Kenya	Johnson, PR	*J Virol* **64**(3):1086–92 (1990)
VER.KE.x.TYO1	X07805	Kenya	Fukasawa, M	*Nature* **333**(6172):457–61 (1988)
SUN.GA.98.L14	AF131870	Gabon	Beer, BE	*J Virol* **73**(9):7734–44 (1999)
LST.CD.88.447	AF188114	D.R.C.	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.CD.88.485	AF188115	D.R.C.	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.CD.88.524	AF188116	D.R.C.	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.KE.x.lho7	AF075269	Kenya	Hirsch, VM	*J Virol* **73**(2):1036–45 (1999)
SYK.KE.x.SYK173	L06042	Kenya	Hirsch, VM	*J Virol* **67**(3):1517–28 (1993)
SYK.KE.x.KE51	AY523867	Kenya	Bibollet-Ruche, F	*J Virol* **78**(14):7748–62 (2004)
COL.CM.x.CGU1	AF301156	Cameroon	Courgnaud, V	*J Virol* **75**(2):857–66 (2001)

```
                           / 5' LTR U3 start
H1B.FR.83.HXB2             .TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTG   129

H1D.CD.84.84ZR085          ..................................................................................................................................     0
H1F1.BE.93.VI850           ..................................................................................................................................     0
H1G.SE.93.SE6165           ..................................................................................................................................     0
H1H.CF.90.056              ..................................................................................................................................     0
H1J.SE.93.SE7887           ..................................................................................................................................     0
H1K.CM.96.MP535            ..................................................................................................................................     0
H101_AE.TH.90.CM240        ..................................................................................................................................     0
H102_AG.NG.-.IBNG          ..................................................................................................................................     0
H103_AB.RU.97.KAL153_2     ..................................................................................................................................     0
H104_cpx.CY.94.CY032       ..................................................................................................................................     0
H1N.CM.95.YBF30            ..................................................................................................................................     0
H1O.BE.87.ANT70            C--------T-----T-------TAA----GC---A-----G-----T---G-G--TA----T--G--A-T-----------G------------------A-----A-C---G-TC--------A----   130
H1O.CM.91.MVP5180          C----T---T-----T-------TAAG---GC---A-----G-----C-----A--T-----T--G--A-T-----------G----TGT--------G--A-----ACCT----TC--------A----   130
CPZ.CM.04.EK505            ..................................................................................................................................     0
CPZ.CM.04.LB7              ..................................................................................................................................     0
CPZ.CM.04.MB66             ..................................................................................................................................     0
CPZ.MT145                  ..................................................................................................................................     0
CPZ.CD.-.ANT               ..................................................................................................................................     0
CPZ.CM.-.CAM3              ..................................................................................................................................     0
CPZ.CM.98.CAM5             ..................................................................................................................................     0
CPZ.GAB_X52154             .--------T--G--T-----AGGA----------G--------C--C---G----T-------------T---------C-G-------------A----A-----A-CA----TC-----CTGT----   129
CPZ.GAB2                   ..................................................................................................................................     0
CPZ.TZ.-.TAN1              ..................................................................................................................................     0
CPZ.US.85.CPZUS            .--------T--G--T-T---AG-A----------G-----A-----C---G-T-----T--------AAT---------C-G------T-----C--------T--AG-A--------C--A--A-A--   129
GSN.99CM166_AF468659       ..................................................................................................................................     0
GSN.99CM71_AF468658        ..................................................................................................................................     0
H2A.ALI_AF082339           .--------A-GT--T--AGTG-GA-------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G----GATAA-T--A----G---------T--T-AT--------A--A--G--C--GA--TT-----   129
H2A.BEN_M30502             .--C-----A-GT--T--AGTAGGA-G-----TAGA-----A--CA-A-ACC-AG-AA-AGAGG----GATAA-A--A----G------T--T--T-AT--------AG-A--G--C---A--TA---C-   129
H2A.ST_M31113              ..................................................................................................................................     0
H2B.D205_X61240            .--------A-GTA-T-TAGTG-GA-------CAGA--AT-A--CACA-ATT-TG-GA-TGA-G-----ATTG-GT---GC-G---A-----T--T-AT-----------A--GC----CAAATA-----   129
H2B.EHO_U27200             .--------A-TT-CT-TAGTG-GA----G--TA-A--A--A--CACA-ACT-AG-AA-TGA-G-----ATTG-GT---GA-G--------------AT--------AG-A--------CAA-TT-----   129
H2G.ABT96_AF208027         ..................................................................................................................................     0
H2U.FR.96.12034            ..................................................................................................................................     0

MAC.239_M33262             .--------A-TTA-T--AGTGC-A-------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT---A-----A--T-----C---AA---A----   129

SMM.SL.92.SL92B            ..................................................................................................................................     0
SMM.US.x.PGM53             ...........................................................................--A----G---A--T------T----------A--------C--TA--TT-----    55
MNE.US.-.MNE027            ..................................................................................................................................     0
RCM.NG.x.NG411             ..................................................................................................................................     0
RCM.GAB1_AF382829          ..................................................................................................................................     0
DRL.FAO_AY159321           ..................................................................................................................................     0
MND-1.F17                  ..................................................................................................................................     0
MND-2.5440_AY159322        ..................................................................................................................................     0
MND-2.14CG_AF328295        ..................................................................................................................................     0
MND-2.MND2_AF367411        ..................................................................................................................................     0
DEB.CM.CM40                ..................................................................................................................................     0
DEB.CM.CM5                 ..................................................................................................................................     0
MON.CM.L1                  ..................................................................................................................................     0
MON.NG1                    ..................................................................................................................................     0
TAL.CM.8023                ..................................................................................................................................     0
TAL.CM.266                 ..................................................................................................................................     0
MUS.CM.1085                ..................................................................................................................................     0
SAB.SN.-.SAB1C             .--------A-TTA-T--AGTG--A-G---A-GA-G---T-G-----C-ATGCTCTTA-TGAGTGG---ATAG-AGA--GA-G------T--T---GAT--------T-C---G--C---AA-TG-----   129
GRV.ET.x.GRI_677           .----T---A--TA-T----TG--A-----G--A-G-----GA--T---ATGC--TGA--GAGTGG--AATAA-AGA-----G---AGCT---T-------C--G-----A--G--C--GAGAGT-----   129
TAN.UG.x.TAN1              .----G--CA--TA-TGG--A-C-AAG---G-GC-A-----GA----C-ATGC-CTAA-TGA-TGG---ATTA-AGA-----G---AGCA--TT-C--T--------C-C---GA-G--C-GATG---C-   129
VER.DE.-.AGM3              ..................................................................................................................................     0
VER.KE.-.9063              .----T---A-TTA-T-----G--A-G--GA-TAGA-----CA-C----ATGCTCTTA-TGA-TGG--AAT-A-TGA-----G-A-CGC--GGT--A----A-----CG-A--------C-GCTG-----   129
VER.KE.-.AGM155            .----T---A-TTA-T-----G-TA-G---A-TA-G------A------ATGC-CT-A-TGA-TGG--AAT-A-TGA-----G-A-CGCA-GGT--AA---A--T-----A-----C--GAG-TG-----   129
VER.KE.-.TYO1              ..................................................................................................................................     0
SUN.GA.98.L14              .----------CTGGT-----AG-ACC---G----A---T-G---T---ATGCAG-AA-TGAGTGG----T-A-AA---GC-G---AG-T-----CAA------T--AG-G---------AA-G------   129
LST.CD.88.447              ..................................................................................................................................     0
LST.CD.88.485              ..................................................................................................................................     0
LST.CD.88.524              ..................................................................................................................................     0
LST.KE.-.lho7              .-----------TA-TGG-----GA-------G--C---T-AAT---C-ACTGTG-GA-TGAGTGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A-CA-----C--T-----A----   129
SYK.KE.-.SYK173            ........................................................................................................---A-C-C-G--C--GT-A-T-----    26
SYK.KE.KE51_AY523867       ..................................................................................................................................     0
COL.CM.-.CGU1              ..................................................................................................................................     0
```

```
H1B.FR.83.HXB2             GATGGTGCTACAAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAA...GCCAACAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGC......CTGCATGGAATGGATGACCCGGAGAGAGAAGT   250
H1D.CD.84.84ZR085          ..................................................................................................................................     0
H1F1.BE.93.VI850           ..................................................................................................................................     0
H1G.SE.93.SE6165           ..................................................................................................................................     0
H1H.CF.90.056              ..................................................................................................................................     0
H1J.SE.93.SE7887           ..................................................................................................................................     0
H1K.CM.96.MP535            ..................................................................................................................................     0
H101_AE.TH.90.CM240        ..................................................................................................................................     0
H102_AG.NG.-.IBNG          ..................................................................................................................................     0
H103_AB.RU.97.KAL153_2     ..................................................................................................................................     0
H104_cpx.CY.94.CY032       ..................................................................................................................................     0
H1N.CM.95.YBF30            ..................................................................................................................................     0
H1O.BE.87.ANT70            ------TG-TT--A-----------GTCAGA---AG--GC----AG-CTA-GAG-T-C-T-T----GGG-T-ATC-CC-G--T--A-CAT-T......GCC------T-T--A--TA-AC-T-A----A-   254
H1O.CM.91.MVP5180          ------TG-TT--A--G--------GTCAG----AG--GC---GAG-CTG-GT--T-C-AAT--AG-TG-T--TC-TC----T--A-CTT-T......AAT------GCT--G--TG-AC-CG-G--GA-   254
CPZ.CM.04.EK505            ..................................................................................................................................     0
CPZ.CM.04.LB7              ..................................................................................................................................     0
CPZ.CM.04.MB66             ..................................................................................................................................     0
CPZ.MT145                  ..................................................................................................................................     0
CPZ.CD.-.ANT               ..................................................................................................................................     0
CPZ.CM.-.CAM3              ..................................................................................................................................     0
CPZ.CM.98.CAM5             ..................................................................................................................................     0
CPZ.GAB_X52154             ---------T---A-----G--CC-GACAGAG---C--G-----C--...-----TG-------T----A-T--C----G--T--CA-AT-T......-A------G-----A--TGA---C-A---G--   250
CPZ.GAB2                   ..................................................................................................................................     0
CPZ.TZ.-.TAN1              ..................................................................................................................................     0
CPZ.US.85.CPZUS            -GC------TT---T----C--TC-CACAGA----G-AG----GCGG...-----T----------C--A--TAC--C-G-----CA--T--......-A------------A--TGAAC-TG-C-----   250
GSN.99CM166_AF468659       ..................................................................................................................................     0
GSN.99CM71_AF468658        ..................................................................................................................................     0
H2A.ALI_AF082339           -G---CTG-GG--------------A--T......GTCCC-C-----...-GGG-GG-CACT----CTCA-T--C--C-------A--ACAA......ACAAGCA-GCAT------A--C-TG-G--GAC   244
H2A.BEN_M30502             -G---CTG-GG---------T----A--A......CTC-C-C----G...--AG-GG---AT---GC--A-T----AG-------A-CACAA......ACAAGCA--CAT-----TGA-C-TG-G--GAC   244
H2A.ST_M31113              ..................................................................................................................................     0
H2B.D205_X61240            -T---CTG-GG-----G--------A---...-T-CCAGC--CGACC...CGAG-GG-G-AG--A-C-CATT--C-AA-G-----G-CACAG......A-CTCCTC-TG-------ATCC-TG-G--GAC   247
H2B.EHO_U27200             -C---CTA-GG--------G---A-AA-T......-T-A---C----...C-AG-GG-C-AG--A-C-CATT-TC--G-G--T--A-CACAG......ACCTCC-C-TG---------CC-CGA---GAC   244
H2G.ABT96_AF208027         ..................................................................................................................................     0
H2U.FR.96.12034            ..................................................................................................................................     0

MAC.239_M33262             -C---CTA-GG--AT----C--T--AA-T......GTA-C---T--G...--AC-GG-G-AT---G-GCATTAT--AA-G--T--A-CTCAA......ACTTCCCAGTG---------TTG-G----G--   244

SMM.SL.92.SL92B            ..................................................................................................................................     0
SMM.US.x.PGM53             -C---CTA-GG--AT-G--C-----AAGT......GTC-C---T---...--TC-AG---AC----CACATTA---AG-G--T--A-CTCAG......A-ATCCCAGTG------T--CTG-G----G--   170
MNE.US.-.MNE027            ..................................................................................................................................     0
RCM.NG.x.NG411             ..................................................................................................................................     0
RCM.GAB1_AF382829          ..................................................................................................................................     0
DRL.FAO_AY159321           ..................................................................................................................................     0
MND-1.F17                  ..................................................................................................................................     0
MND-2.5440_AY159322        ..................................................................................................................................     0
MND-2.14CG_AF328295        ..................................................................................................................................     0
MND-2.MND2_AF367411        ..................................................................................................................................     0
DEB.CM.CM40                ..................................................................................................................................     0
DEB.CM.CM5                 ..................................................................................................................................     0
MON.CM.L1                  ..................................................................................................................................     0
MON.NG1                    ..................................................................................................................................     0
TAL.CM.8023                ..................................................................................................................................     0
TAL.CM.266                 ..................................................................................................................................     0
MUS.CM.1085                ..................................................................................................................................     0
SAB.SN.-.SAB1C             ---------TT--A--G-----T--G--C......CT--C---G--G...--T--G--TA-T------CA-T--C--C-G--T----CTCAA......G--GCCTATGA-------G-CTG--A---GAC   244
GRV.ET.x.GRI_677           -C-TC----TT--------C-----G--C......CT-CAT--G--G...--ACG---CT-T----GACA-T-TC--A-G--T--A-CACAGATGGGGAAG--CCTGAT-GAATAGATC-TG--------   250
TAN.UG.x.TAN1              -C-TT----TTG--T----G--G--A--T......GTC-CCC-G---...---C-GG-T-AG...-GACA-T-T---C-C-------CCCAG......A-TG-AT-GGAAAGC-----TTG--A---GAC   241
VER.DE.-.AGM3              ..................................................................................................................................     0
VER.KE.-.9063              -C-TT----T------------G--A---......-T-CAT-----G...--AG-G-CCT-T----GACATTA----GCG--T----CACAAATAAAAGA-G-CCCCGAT-GCAT-AGTC-TG----GAC   250
VER.KE.-.AGM155            -C-TC----T------------G----CC......CT-CAT--G---...--AG-A-C-T-T--A-GGCATT-----G-------A-CACAACTGCATGAAG-CCCTGAT-G-ATAAATC-TG-----A-   250
VER.KE.-.TYO1              ..................................................................................................................................     0
SUN.GA.98.L14              -C---CTA-GG-------C------CACC...ATTG-TGA---CAGG...-ATCCT--TCATCCTTGTCAAGC-C-AC-G--TT-CAGCCAG......-AAGGG-TT-AT--A-----TTG-G----GAG   247
LST.CD.88.447              ..................................................................................................................................     0
LST.CD.88.485              ..................................................................................................................................     0
LST.CD.88.524              ..................................................................................................................................     0
LST.KE.-.lho7              ------TG-GGC--T---A------G-CT...TGTG-TGAGT-TA--...-ATCC-TC--AT----CACAAT--C-T--G--TT-AAGCCAG......---GGG-TCC----A-----CTG-G----GAG   247
SYK.KE.-.SYK173            ------TA-GGG-A-----T---AA----......-TAGAG-G-TGC...TTGG-GT-T-AG-----T--ACT-C-CC-C--T--A-CCTCT.........GGGAC-GG---GTT-GAAC-TG-G--GAA   138
SYK.KE.KE51_AY523867       ..................................................................................................................................     0
COL.CM.-.CGU1              ..................................................................................................................................     0
```

```
                                                                                          |- TCF-1 alpha  -|
H1B.FR.83.HXB2            GTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTAC...TTCAAGAACTGC........................TGACA...................   334

H1D.CD.84.84ZR085         ..................................................................................................................................     0
H1F1.BE.93.VI850          ..................................................................................................................................     0
H1G.SE.93.SE6165          ..................................................................................................................................     0
H1H.CF.90.056             ..................................................................................................................................     0
H1J.SE.93.SE7887          ..................................................................................................................................     0
H1K.CM.96.MP535           ..................................................................................................................................     0
H101_AE.TH.90.CM240       ..................................................................................................................................     0
H102_AG.NG.-.IBNG         ..................................................................................................................................     0
H103_AB.RU.97.KAL153_2    ..................................................................................................................................     0
H104_cpx.CY.94.CY032      ..................................................................................................................................     0
H1N.CM.95.YBF30           ..................................................................................................................................     0
H1O.BE.87.ANT70           AC-GAT-----A------T--ATCT----GCAACACC--TG-T--TATGATAACT--C--A---CT-...---C----G..............................G-CTAAAAACTGC........   343
H1O.CM.91.MVP5180         AC--A-A---CA------T--ATCAT---GC--AACA--T--A----TGC-AAA---C--A---CT-...---CCC--G..............................T-ACTGACACTGC........   343
CPZ.CM.04.EK505           ..................................................................................................................................     0
CPZ.CM.04.LB7             ..................................................................................................................................     0
CPZ.CM.04.MB66            ..................................................................................................................................     0
CPZ.MT145                 ..................................................................................................................................     0
CPZ.CD.-.ANT              ..................................................................................................................................     0
CPZ.CM.-.CAM3             ..................................................................................................................................     0
CPZ.CM.98.CAM5            ..................................................................................................................................     0
CPZ.GAB_X52154            C--G-TC---C-C---------A-G--G--C--AAGA--T--T---A----A-AA------------...-A---AG--...........................----TTTAG...............   335
CPZ.GAB2                  ..................................................................................................................................     0
CPZ.TZ.-.TAN1             ..................................................................................................................................     0
CPZ.US.85.CPZUS           C---ATC---CA--------CTGAA--G--TCGGAGA----GA--TAA-----------A-------...---CG----...........................---.....................   329
GSN.99CM166_AF468659      ..................................................................................................................................     0
GSN.99CM71_AF468658       ..................................................................................................................................     0
H2A.ALI_AF082339          A----TT-----A------CCTAAG--G--TCA-G--T---AA---TTTATT--A--C--A---G-A...--TGG-T--AAG...TCAGGCCTGCCAGAAGATGAG--GA-GGCAAGACTGAAAGCAAGA   368
H2A.BEN_M30502            A----T----CA-------TC-ATG--G--C-A-A-CT---A----TTCACT---T-C--A--AG-G...--TGG-C--AAG...TCAGGATTGCCAGAGAAAGAA--GA-GGCAAAACTGAAAGCAAGA   368
H2A.ST_M31113             ..................................................................................................................................     0
H2B.D205_X61240           TC-TATC---CA------TTC--T---G----A-G--T-TG----TTTCA-TAG-TT---A--AG-G...---GG-T--CAG...TCAGGGCTACCAGAGGAGGAG--GA-GGCTAGACTAAAAGCAAGA   371
H2B.EHO_U27200            CC-T-TC---CA-------TC--T--------A-G-CT-TG-----TTCAGCAG-TTC--A---G-G...--TGG-T-TCAG...TCAGGAATGCCAGAGAAAGAG--GA-GGCTAAACTGAGAGCAAGA   368
H2G.ABT96_AF208027        ..................................................................................................................................     0
H2U.FR.96.12034           ..................................................................................................................................     0

MAC.239_M33262            TC---CA----A------TCCAACT--G--C-ACAC-T-TGA---ATAT-TTAGAT-C--A--AG-G...--TGGA-G-AAG...TCAGGCCTGTCAGAGGAAGAGGTTAGAAGAAGGCTAACCGCA...   365

SMM.SL.92.SL92B           ..................................................................................................................................     0
SMM.US.x.PGM53            TC-G-CA----A------TTCA-AAT----C-ACAGAT-TGA---TTTTATTAGA--C--A--AG-G...--TGGT-GTAAG...TCAGGCTTGTCAGGGGAAGAGGT-A-GAGAAGGCTAACCGCAAGA   294
MNE.US.-.MNE027           ..................................................................................................................................     0
RCM.NG.x.NG411            ..................................................................................................................................     0
RCM.GAB1_AF382829         ..................................................................................................................................     0
DRL.FAO_AY159321          ..................................................................................................................................     0
MND-1.F17                 ..................................................................................................................................     0
MND-2.5440_AY159322       ..................................................................................................................................     0
MND-2.14CG_AF328295       ..................................................................................................................................     0
MND-2.MND2_AF367411       ..................................................................................................................................     0
DEB.CM.CM40               ..................................................................................................................................     0
DEB.CM.CM5                ..................................................................................................................................     0
MON.CM.L1                 ..................................................................................................................................     0
MON.NG1                   ..................................................................................................................................     0
TAL.CM.8023               ..................................................................................................................................     0
TAL.CM.266                ..................................................................................................................................     0
MUS.CM.1085               ..................................................................................................................................     0
SAB.SN.-.SAB1C            CC-G-T-----A-------CCA-TA------G-GG-CT-TG-T--TT-GAG------C--A---C-G...G-GCC-TCAGCA...CAAGGCTAACCACAAAACCACATC-TATG................   352
GRV.ET.x.GRI_677          C--G-TC----A-------...............-CGA-GT----GGTG---TAC-GC-----CATG...--T...--GGA-...ATGCACGAACATGCAAAGCGC...T-GTGTCAG............   341
TAN.UG.x.TAN1             A--G-T-----AA---A-T...............-CA-T-C-A--AGTTC--TACA-C-----CAG-...--T---G--ATG...CATGGTCTTGTTAAGAGGAAG-AG-CTAACCGCAGGCTTGCGGTT   350
VER.DE.-.AGM3             ..................................................................................................................................     0
VER.KE.-.9063             C--G-TC----A-------...............-C-ATGT-A--AGTGC--TATG-C--AA-CAGA...CAGT-TTTTGAAGACATGCATGCAATAGTGAAGAGGAAGT-GCTAACCGCAGGCTTGTG.   361
VER.KE.-.AGM155           A--G-CA----A------T...............-CAATGT----TGTTC--TACG-C--CTCAAGG...GAGT-CTTTACAGACTTATATTCAACAGTTGGTACAG--A-CTAGCCGACCACAGGCTTG   362
VER.KE.-.TYO1             ..................................................................................................................................     0
SUN.GA.98.L14             AC-TATC----CA-----TCC-ACT--G--C-A-G--TTT-G----AT-C--AA---C--T--AG-G...------C-TGTT...ACTAGTCTGCAGTGGGAGGTG-AG-TTAACCGCAAACCG......   365
LST.CD.88.447             ..................................................................................................................................     0
LST.CD.88.485             ..................................................................................................................................     0
LST.CD.88.524             ..................................................................................................................................     0
LST.KE.-.lho7             -C-TATC---CAC--CA-TCCAATGT-G---G-GG-CTTT--A----TCA-AAA---G--T-C-A-G...A-AC-A--TACT......GCTTTCGCTTTTGACTGCAAGAGGAAGTAGCCAAACCGC...   365
SYK.KE.-.SYK173           CC-CAT-----AC--CA--CC--A---G--C-A-ACA-CAGGCTGGGAGAT-GCC-GCTT-C---TA...GAG-GAC-AACA...GGAAAACCACAAGAGCTGCAG-C-GCACTTAGTAAGAACATCTCC   262
SYK.KE.KE51_AY523867      ..................................................................................................................................     0
COL.CM.-.CGU1             ..................................................................................................................................     0
```

```
                                See Pohlmann, J Virol 72(7):5589-5598(1998) and similar publications for information on this enhancer region
                                                                                                                      NF-kappa-B-II                     NF-kappa-B-I
                                                                                                                       |-      -|                        |-
H1B.FR.83.HXB2            ........................................................................TCGAGCTTGCT...ACA...AGGGACTTTCCG....................CTGGGG   366
H1D.CD.84.84ZR085         ..................................................................................................................................     0
H1F1.BE.93.VI850          ..................................................................................................................................     0
H1G.SE.93.SE6165          ..................................................................................................................................     0
H1H.CF.90.056             ..................................................................................................................................     0
H1J.SE.93.SE7887          ..................................................................................................................................     0
H1K.CM.96.MP535           ..................................................................................................................................     0
H101_AE.TH.90.CM240       ..................................................................................................................................     0
H102_AG.NG.-.IBNG         ..................................................................................................................................     0
H103_AB.RU.97.KAL153_2    ..................................................................................................................................     0
H104_cpx.CY.94.CY032      ..................................................................................................................................     0
H1N.CM.95.YBF30           ..................................................................................................................................     0
H1O.BE.87.ANT70           .................................................TGACCTGAAGATTGCTGACACTG-....................--A-------A..GCAAAGACTGCTGACACTGC----   402
H1O.CM.91.MVP5180         .............................................................................................----------A.......GACTGCTGACACTGC----   373
CPZ.CM.04.EK505           ..................................................................................................................................     0
CPZ.CM.04.LB7             ..................................................................................................................................     0
CPZ.CM.04.MB66            ..................................................................................................................................     0
CPZ.MT145                 ..................................................................................................................................     0
CPZ.CD.-.ANT              ..................................................................................................................................     0
CPZ.CM.-.CAM3             ..................................................................................................................................     0
CPZ.CM.98.CAM5            ............................................................GCGCCTGCGCAG-A--AAC----G..---A...---------TAA......................---    43
CPZ.GAB_X52154            ........................................................ACTGGCGCATGCGC.ACAAGAAC----G..--TCTGC----------AA......................---   384
CPZ.GAB2                  ..................................................................................................................................     0
CPZ.TZ.-.TAN1             ..................................................................................................................................     0
CPZ.US.85.CPZUS           .................................................TGGCTGTAACCGCGCAGGCGC.AATA-AAC----G..--TG...A--------TAA...GGGACTTTCCAA.......---   394
GSN.99CM166_AF468659      ..................................................................................................................................     0
GSN.99CM71_AF468658       ..................................................................................................................................     0
H2A.ALI_AF082339          GGGATACCATTTAGTAAGAACAGGAACAGCTGATTTGG.......TCAGGGCAGGAAGTAACTACTG.....AAA-C.AGCTGA..GACTGC-----------A.....................GAA--   462
H2A.BEN_M30502            GGGATACCATATAGTGA...ATAACAGGAACAACCATACT.....TGGTCAAGGCAGGAAGTAGCTA.CTAAGAA-C.AGCTGA..GGCTGC-----------A.....................GAA--   465
H2A.ST_M31113             ..................................................................................................................................     0
H2B.D205_X61240           GGGATACCTACAGATTAGGCAAGAGACAGCAGCATAAACA.....GGAACTAGCTGACACTGCACAA.GAAGGAA-CTAGCAGA..CACTGC-----------A.....................AAA-A   472
H2B.EHO_U27200            GGAATACCTACAGAGTAGACAGGAAATAACAGCACAA.....AGAGGAACTAGCTAACACTGCATAGAGAAGGAA-CTAGCTGA..TACTGC---.-------A.....................GAA--   469
H2G.ABT96_AF208027        ..................................................................................................................................     0
H2U.FR.96.12034           ..................................................................................................................................     0

MAC.239_M33262            ..........................................AGAGGCCTTCTTAACATGGCTGACAAGAAGGAA-CTCGCTGA..-ACAGC-----------A.....................CAA--   430

SMM.SL.92.SL92B           ..................................................................................................................................     0
SMM.US.x.PGM53            GG....CCTTTTAAAAATGGCTGACAAGAAGGAAACGAGC.........................................TGA..GACAGC-----------A.....................CAAA-   356
MNE.US.-.MNE027           ..................................................................................................................................     0
RCM.NG.x.NG411            ..................................................................................................................................     0
RCM.GAB1_AF382829         ..................................................................................................................................     0
DRL.FAO_AY159321          ..................................................................................................................................     0
MND-1.F17                 ..................................................................................................................................     0
MND-2.5440_AY159322       ..................................................................................................................................     0
MND-2.14CG_AF328295       ..................................................................................................................................     0
MND-2.MND2_AF367411       ..................................................................................................................................     0
DEB.CM.CM40               ..................................................................................................................................     0
DEB.CM.CM5                ..................................................................................................................................     0
MON.CM.L1                 ..................................................................................................................................     0
MON.NG1                   ..................................................................................................................................     0
TAL.CM.8023               ..................................................................................................................................     0
TAL.CM.266                ..................................................................................................................................     0
MUS.CM.1085               ..................................................................................................................................     0
SAB.SN.-.SAB1C            ...................................................GAGTTGTCATGGTGATGACAT-AAGAAC----G..--TGAG-----------A..........................   403
GRV.ET.x.GRI_677          .................................................................................-ACTTTGCGGTT-----------.CCAGGGACTTTCCACA.....-T--   384
TAN.UG.x.TAN1             AAGCTAACTAGGCCGTTGCCTAGGAGATGACATAAAGAAGTGC......................................TGA...TCAGC-----------A..........................   413
VER.DE.-.AGM3             ..................................................................................................................................     0
VER.KE.-.9063             .............................................GTTAAGCCGTTGCCAGGGAGATGACAT-T--AAC----G..---G...----------A......................A---   419
VER.KE.-.AGM155           ......................................CG.....GTTTCCTGGTTGCCTAGGAGATGACAT-AAGAAC----G..--.....----------A...................G-AC---   423
VER.KE.-.TYO1             ..................................................................................................................................     0
SUN.GA.98.L14             ............................CATCCTCTTGCA.....TCGCCTAGGCAACGGGGCTAGC.GCAT......GC--GC..TAGAGT------------..........................   427
LST.CD.88.447             ..................................................................................................................................     0
LST.CD.88.485             ..................................................................................................................................     0
LST.CD.88.524             ..................................................................................................................................     0
LST.KE.-.lho7             ........AAACCGCAGCACATCCTCTTGCAGCCCGGTTG.....CTAAGGCAACCGGGCTAGCGCATGCGCAATG-------G..--GAGC-----------A.............GGACGGG-G----   467
SYK.KE.-.SYK173           TAGGAGACTCCATGGTGACAAGCTCGGCCCAGGGGA.........................................................-----C-GGGC..................GGGG-AA-   317
SYK.KE.KE51_AY523867      ..................................................................................................................................     0
COL.CM.-.CGU1             ..................................................................................................................................     0
```

HIV-2 and SIVs have extra TAR stem-loop(s) see Berkhout, *Nucleic Acids Res* **20**(1):27-31(1992)

```
                                                                                                                                                    /+1 mRNA start site
                                                                                                                                     5' LTR U3 end \/ 3' LTR start
                                                                                                                                              TAR element stem
                    NF-kappa-B-I  -|                  Sp1-III      Sp1-II      Sp1-I                             TATA Box                     |   ||||^|||||
H1B.FR.83.HXB2            ACT...TTCCAG.............GG.AGGCGTGG.CCTGGGCGGGACTGGGGAGTGGCGA.........GCCCTCAGATCCTGCATATAAGCAGCTGCT...TTTTGCCTGTACT...GGGTCTCTCT  463
H1D.CD.84.84ZR085         ..................................................................................................................................    0
H1F1.BE.93.VI850          ..................................................................................................................................    0
H1G.SE.93.SE6165          ..................................................................................................................................    0
H1H.CF.90.056             ..................................................................................................................................    0
H1J.SE.93.SE7887          ..................................................................................................................................    0
H1K.CM.96.MP535           ..................................................................................................................................    0
H101_AE.TH.90.CM240       ........................................................................................................................----------   10
H102_AG.NG.-.IBNG         ..................................................................................................................................    0
H103_AB.RU.97.KAL153_2    ..................................................................................................................................    0
H104_cpx.CY.94.CY032      ..................................................................................................................................    0
H1N.CM.95.YBF30           .......................................................................................................T-C-C--T------...----------   24
H1O.BE.87.ANT70           ---...------...............TG--A-G-ACAGG------TT-G---AGTG-C..........TAA--------AG---------------C---...--C---T-----C...-------GG-  498
H1O.CM.91.MVP5180         ---...------.............C-TG--A-G-ATAAG------TT-G---AGTG-C..........TAA---------G-------------------...--CC--T-----C...------TAG-  471
CPZ.CM.04.EK505           ..................................................................................................................................    0
CPZ.CM.04.LB7             ..................................................................................................................................    0
CPZ.CM.04.MB66            ..................................................................................................................................    0
CPZ.MT145                 ..................................................................................................................................    0
CPZ.CD.-.ANT              ..................................................................................................................................    0
CPZ.CM.-.CAM3             ..................................................................................................................................    0
CPZ.CM.98.CAM5            ---...-----A..............-GG--A-GCTCA..-------........ACA-G-CGTGGTTT.CA--------AG---------------C---...--C---T-----G...T---------  139
CPZ.GAB_X52154            ---...----G-.............--AGA-GTG-TCGGGA------G.......T---G-TGTGGCTT.T--------.AG-------------------...--C----------...--------T-  483
CPZ.GAB2                  ..................................................................................................................................    0
CPZ.TZ.-.TAN1             .........................................................................................................................-C---TG-C    9
CPZ.US.85.CPZUS           --G...----...............AAGG--GTG--T-A.------A........ACA-G-CGTGGTTT.TA------A-GA-------A-------C---...C-C---T------...--------T.  489
GSN.99CM166_AF468659      ..................................................................................................................................    0
GSN.99CM71_AF468658       ..................................................................................................................................    0
H2A.ALI_AF082339          GGC...-GTA-C.............CAGG--AAG-AC.A----A--AG----T-G-GAA-..........GC--TCATAC-----T---A-T-T-C-C---.GC--GCATTGTATTC...A-TCGCTCTG  561
H2A.BEN_M30502            GGC...-GTA-C.............CAAG--A-G-AC.A----A--AG----T-G-GAA-.........GCC-T-AT-CT-A---T---A-T-T-C-C---.TC--GCATTGTATTC...A-TCGCTCTG  565
H2A.ST_M31113             ........................................................................................................................A-TCGCTCTG   10
H2B.D205_X61240           GGC...-GTA-C.............A.TG--A-GA-C.AA---A---GT---TTG-GAA-.........GCC-AT-A-ATCT---T---A-T-T-C-C---.TC-CGCATTGTATTC...A-TCGCTCTG  571
H2B.EHO_U27200            GGC...-GTA-C.............A-TG--A-G-AC.A----A--AG-A---AG-.AA-.........GCC-ATCA---CT---T---A-T-T-C-C---.TTG-GCATTGTATTC...A-TCGCTCTG  568
H2G.ABT96_AF208027        ..................................................................................................................................    0
H2U.FR.96.12034           ........................................................GAA-.........GCC-T-AA-TGAT---T---A-T-T-C-C---CTT-.GCATTGTATTC...A-TCGCTCTG   61

MAC.239_M33262            GGA...-GTT-CGGG..GAGGTACT--GGA-GAGCCG.G-C--GAACG......................CC-A--TTCT-GA--T---A-TATCA----A.TT-CGCT-TGTATTC...A-TCGCTCTG  528

SMM.SL.92.SL92B           .........................................................................................................................-TCGCTCTA    9
SMM.US.x.PGM53            GGGA..-GTT-T.............--GGA-GTACTG..G--AG-AACTG-TT-GACC............GCC-A--T-TT-T---T---A-TA--A---A.TT-CGCT-TGTATTC...A-TCGCTCTG  454
MNE.US.-.MNE027           ........................................................................................................................A-TCGCTCTG   10
RCM.NG.x.NG411            ..................................................................................................................................    0
RCM.GAB1_AF382829         ..................................................................................................................................    0
DRL.FAO_AY159321          ..................................................................................................................................    0
MND-1.F17                 ..................................................................................................................................    0
MND-2.5440_AY159322       ..................................................................................................................................    0
MND-2.14CG_AF328295       .....................................................................................G-G-GGCTACCAGAT-.GAGCC--GG-TGTTCTCT--TGAGTC--   44
MND-2.MND2_AF367411       ........-GTCGGGTGGGGGAGTG-TC-AC-CAACCTGGCT--ATATAA-CA-CTC-CTTTGCGCTTGAA-TGAGTCTC---CTG-G-GGCTACCAGAT-.GAGCC--GG.TGTTCTCT--TGAGTCT-  120
DEB.CM.CM40               ..................................................................................................................................    0
DEB.CM.CM5                ..................................................................................................................................    0
MON.CM.L1                 ..................................................................................................................................    0
MON.NG1                   ..................................................................................................................................    0
TAL.CM.8023               ..................................................................................................................................    0
TAL.CM.266                .....................................................................................TG-C-GCAG--TG-T..............................   15
MUS.CM.1085               ..................................................................................................................................    0
SAB.SN.-.SAB1C            GGG...-GGAGA.............CTGG-CG--ACTGGGA-TG-CTT....................GCC--G-GCTGCT-AAT---AGCA..-------.TTGCGCCTG-AA-AC...A-TCTCTG--  490
GRV.ET.x.GRI_677          GTGG.....................ATCG-AG-C--TA-A---GC--TACT--------TT........TCC--TCAGAGCTGCAT-A-AGCAG-TGCT-GCTGGC---TAAC-C.....A-TCTCT-AC  481
TAN.UG.x.TAN1             ...........-.............--TG-TGCGA-.GGC--TACTTGGGA-T-GC-T..............---CTCAGAG-------A------A..TGCTCGC-G--T----ACTC...AGTCTCT-  499
VER.DE.-.AGM3             ....................................................................................................................C...A-TCTCT-AC   11
VER.KE.-.9063             ---...------.............--CG---CAT-.GGC--T.AC-GGGA-T-GT-T.............TA--CTCAGAG-------A------A----CGC-GGCTTG-AACTC...A-TCTCT-AC  515
VER.KE.-.AGM155           ---...-----A.............--CG--ACAT-.GGC--T.AC-GGGA-T-GC-T.............TA--CTCAGAG-------A------A----CGC-GGCTTG-AACTC...A-TCTCT-AC  519
VER.KE.-.TYO1             ....................................................................................................................C...A-TCTCT-AC   11
SUN.GA.98.L14             .......................................GACAG--AGGG--A-...............GCTTA-C--CT-A-------------------.CT-CGCCTG-T.CGG..GA-TCTCTG--  499
LST.CD.88.447             ..................................................................................................................................    0
LST.CD.88.485             ..................................................................................................................................    0
LST.CD.88.524             ..................................................................................................................................    0
LST.KE.-.lho7             -GG...--T................................................................G-C--CT-G-------------------..C-CCGCTTGT-CGG...-A-.......  518
SYK.KE.-.SYK173           G--...CA--CT.............-CTGC........................................................---A----C-A----.CTG-AGCTA-CCG-GG..A-T-GCAA--  372
SYK.KE.KE51_AY523867      ...................................................................................................................-GA..A-T-GCAG--   13
COL.CM.-.CGU1             .........................................................................................................................T-GAGT---    9
```

```
                         TAR element stem                              TAR element bulge  loop       TAR element stem
                            ||||||^                                          |||||^^^||||      |||||||||||| ||||||| |||||||
H1B.FR.83.HXB2              GGTTAGA..........................................CCAGATCTGAGCCTGGGAGCTCTCTGGCTA.ACTAGGG.AACCCACTGCTTA...........AGCC..T...........  525
H1D.CD.84.84ZR085           .................................................................------------CG.G------.-------------...........----..-...........   39
H1F1.BE.93.VI850            ..................................................................................................................................    0
H1G.SE.93.SE6165            ..................................................................................................................................    0
H1H.CF.90.056               ..................................................................................................................................    0
H1J.SE.93.SE7887            ..................................................................................................................................    0
H1K.CM.96.MP535             ..................................................................................................................................    0
H101_AE.TH.90.CM240         T------..........................................----G--.-----C----------------.G-A----.-------------A..........----..-...........   72
H102_AG.NG.-.IBNG           ..................................................-----------------------------.G-GGA--.-------------...........----..-...........   54
H103_AB.RU.97.KAL153_2      ..................................................................................................................................    0
H104_cpx.CY.94.CY032        ..................................................................................................................................    0
H1N.CM.95.YBF30             T-C-G--..........................................------TA----------------------.G-AG--A.-C--ACTG.----...........----..-...........   85
H1O.BE.87.ANT70             TAGAG--..........................................----G--------C--------C-----CT.CTAGCT-.-----G-------...........-CG-..-...........  560
H1O.CM.91.MVP5180           TAGAG--..........................................----G--------C--------C-----CT.CTAGCT-.-----G-------...........-CG-..-...........  533
CPZ.CM.04.EK505             ..................................GTCTCTCTTGTTAGA------TA----------------------GCA..A--.-------------...........----..-...........   69
CPZ.CM.04.LB7               ..................................GTCTCTCTTGTTAGA------TCTGAG-CT---------------GCT..AA-G-------------...........A---..-...........   71
CPZ.CM.04.MB66              ..................................GTCTCTCTTGTTAGA------TCTGAG-CT---------------GCT..AA-G-------------...........----..-...........   70
CPZ.MT145                   ..................................GTCTCTCTTGGTAGA------------------------------ACA..A--G-------------...........----..-...........   70
CPZ.CD.-.ANT                ..................................................................................................................................    0
CPZ.CM.-.CAM3               ..................................................................................................................................    0
CPZ.CM.98.CAM5              T---G--..........................................------TA------------------..GCC-GC----.-------------...........----..-...........  200
CPZ.GAB_X52154              CAC-G--..........................................------T------------------A.-C-GTGA--.-.-------------...........----..-...........  544
CPZ.GAB2                    ..................................................................................................................................    0
CPZ.TZ.-.TAN1               TAA-CTG..........................................---------------------------TAGTGGCT--..CTAGAGACCGC-G.............................   66
CPZ.US.85.CPZUS             C-C----..........................................------AA-----C----------------GTG--..-.-------------...........----..-...........  550
GSN.99CM166_AF468659        ..................................................................T---TCG-CTT-......CAC.-G-GGC---GGAGCCCTTGGAGCT--GT..AAAGCGGCCTGG   55
GSN.99CM71_AF468658         ..................................................................T--CTC-GTCTC......CAC.-G-GGC---GGAGCCCTTGGAGCT--GT..AGAGCGGCCTGG   55
H2A.ALI_AF082339            C-GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAGTG--TGGCCGGCACTGGGCAGA-...GGCTCCACGCTT  682
H2A.BEN_M30502              C-GAGAGGCTGGCAGATCGAGCCCTGAGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTGG.---CTCA.CCAGT---TGGCCGGTACTGGGCAGA-...GGCTCCACGCTT  686
H2A.ST_M31113               C-GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAGTG--TGGCCGGCACTGGGCAGA-...GGCTCCACGCTT  131
H2B.D205_X61240             C-GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAG----AGGCCAGTGTTGGGCAGA-...GGCTCCACGCTT  692
H2B.EHO_U27200              C-GAGAGGCTGGCAGATCGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAG----TGGCCGGTGCTGGGCAGA-...GGCTCCACGCTT  689
H2G.ABT96_AF208027          ........................................................................---CTAG.---CTCA.CCAG-G--TGGCCGGTGCTGGGCAGAGY..GGCTCCACGCTT   54
H2U.FR.96.12034             C-GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTG..G---G-.A---------T-T--C---CTAG.---CTCA.CCAGT---TGGCCGGTACTGGGCAGAGT..GACTCCACGCTT  183

MAC.239_M33262_             C-GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAG----TGGCCGGTGCTGGGCAGAGT..GACTCCACGCTT  650

SMM.SL.92.SL92B             C-GAGAGGCTGCCAGATAGAGCCCCGAGAGGTTCTCT.GCAGCACTTA.GT--G-..---------T-T-----ACTA-.---CTCAACCAG--ACAGGCCATTGCTGGGTAGA-...GGCTCCACGCTT  131
SMM.US.x.PGM53              --CAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---GC.A---------T-T--C---CTAG.---CTCA.CCAG----TGGCCGGTGCTGGGCAGAG-..GGCTCCACGCTT  576
MNE.US.-.MNE027             C-GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAG----TGGCCGGTGCTGTGCAGAGT..GGCTCCACGCTT  132
RCM.NG.x.NG411              ..................................................................................................................................    0
RCM.GAB1_AF382829           ..................................................................................................................................    0
DRL.FAO_AY159321            ..................................................................................................................................    0
MND-1.F17                   .....................................................................GGAG-CT---CTAC--A-GCTAAGGG-TG-ATCTCTGAGCAGATC--CT-AGAGCAAGGAC   61
MND-2.5440_AY159322         ..................................................................................................................................    0
MND-2.14CG_AF328295         T-AAG--GTGCCTGCTTGTAGCCCTGGGCGGTTCGCAGGCCCCTGGCTTGT--C----G-TAGCTCGT-AGGTGTT--GG-AAG.-TCTCG-T-AGTGAGCACTCTTGCTTGG..TTT-GGTAGACCTCT  171
MND-2.MND2_AF367411         --AAG--GTGCCTGCTTGTAGCCCTGGGTGGTTCGCAGGCCCCTGGCTTGT--C----G--AGCTCGT-AGGTGTT-CTG-AG---CTC-T-GGGG---CCCTTGCTTGGTCTTGGTAGACCTCTAGCAG  250
DEB.CM.CM40                 ..................................................................................................................................    0
DEB.CM.CM5                  ..................................................................................................................................    0
MON.CM.L1                   .......................................AGCGGCTCGT-TCC-CAGCG-----...........................GG-GCC---GGAGCTAGGTAA---GGCCTGGGAGCCCTT   64
MON.NG1                     ..................................................................................................................................    0
TAL.CM.8023                 ..................................................................................................................................    0
TAL.CM.266                  ...................................................................................................C-CAGCGTCCTGGGAGACC-TGGCCACCGGT   46
MUS.CM.1085                 .........................................AGTACTAG-TCCGCAGCG-----...........................GG-GCC--CGGAACTAGGTAT---GGCCTGGGAGCCCTT   62
SAB.SN.-.SAB1C              T-GAG..GCTGCCAGATTGAGCC.TGGG.TGTTCTCTGGTAAGTCTGAA----C-T.---------T-T--G----GG-AGGC-A-C.CTAAG-GCA--CTGGTGAGGTCTTG-TA..GACCCTCGCCTG  612
GRV.ET.x.GRI_677            TAGG--...ACCAGCTAGAGCCTGGGT...GTTCGCTGGTTAGCCTAAC--G-T-GGCCA--G..........................................GGGGTAAG-A-TCCTTGGCTTCATA  563
TAN.UG.x.TAN1               AC-AG--GACCAGCTTGAGCCTGGGAGTT.................................................................................................CGCT  532
VER.DE.-.AGM3               TAGG---..........................................----C-T.---------T-T--G----T--.G-CTAAC.CTGGT.TG--CA.CCAGGGGTAAGGA-T..C...........   81
VER.KE.-.9063               TAGG---..........................................----C-T.---------T-T--G----T--.G-CTAAC.CTGGT.TG--CA.CCAGGGGTAAGGA-T..C...........  585
VER.KE.-.AGM155             TAGG---..........................................----C-T.---------T-T--G----T--.G-CTAAC.CTGGT.TG--CA.CCAGGGGTAAGGA-T..C...........  589
VER.KE.-.TYO1               TAGG---..........................................----C-T.---------T-T--G----T--.G-CTAAC.CTGGT.TG--CA.CCAGGGGTAAGGA-T..C...........   81
SUN.GA.98.L14               CAGA..GGCTAGGCGACG.A.CCCT........................GAGAGGTCTCA---.A-----GG--T-TCG.CTCCT-AGTGAG----TGC-GAGCACCTCACCC-AG..CCTGATCAGCTT  597
LST.CD.88.447               ..................................................................................................................................    0
LST.CD.88.485               ..................................................................................................................................    0
LST.CD.88.524               ..................................................................................................................................    0
LST.KE.-.lho7               ................................TCCTT.GCTCTGGGGAC-AGC-GTC-CT------T.GAGCACACTGG.T-AGCTA.-G-AGT-A--CCTGGGAGGCTACACTTA..GCACCCACCTGG  610
SYK.KE.-.SYK173             A-C-CTCTCCAGCGGCCTGGGAGCCC.TGAGAGAC..............................................TGGTA-.-G-GGC-C-GGAGCCCTTGCCAGT......AAAGGCTAGTGA  448
SYK.KE.KE51_AY523867        A-C-CTCCTCAGCGGCCTGGGAGCCC.TAGGAGAC..............................................TGGTA-.-G-GGC---GGAGCCCTTGCCAGT.....AAAAGGCTAGTAA   90
COL.CM.-.CGU1               -AC-CTT..........................................G-CTGG---C-----A-CC--GAG-----TGC-AGCTA.TCAG-CT.TGCCTGAGTCTGAGTAGCTTG.....CTTGGTTG   90
```

```
                                                                                                          secondary structure in this region, see Rizvi, J. Virol 67:2681
                                                                                                                      Poly A signal |     |
H1B.FR.83.HXB2              ................................................................................................C.AATAAA.G.CTTGCCTTG..AGT....GCTTC  550
H1D.CD.84.84ZR085           ................................................................................................-.------.-.---------..---....----T   64
H1F1.BE.93.VI850            ..................................................................................................................................    0
H1G.SE.93.SE6165            ..................................................................................................................................    0
H1H.CF.90.056               ..................................................................................................................................    0
H1J.SE.93.SE7887            ..................................................................................................................................    0
H1K.CM.96.MP535             ..................................................................................................................................    0
H101_AE.TH.90.CM240         ................................................................................................-.------.-.---------..---....----A   97
H102_AG.NG.-.IBNG           ................................................................................................-.------.-.---------..---....-----   79
H103_AB.RU.97.KAL153_2      ..................................................................................................................................    0
H104_cpx.CY.94.CY032        ..................................................................................................................................    0
H1N.CM.95.YBF30             ................................................................................................-.------.-.---------..---....---AA  110
H1O.BE.87.ANT70             ................................................................................................-.------.-.---------..---GAGAAGCA.  588
H1O.CM.91.MVP5180           ................................................................................................-.------.-.---------..---GAGAAGCA.  561
CPZ.CM.04.EK505             ................................................................................................-.------.-.---------..---....----.   93
CPZ.CM.04.LB7               ................................................................................................-.------.-.---------..---....---G-   96
CPZ.CM.04.MB66              ................................................................................................-.------.-.---------..---....-AC-T   95
CPZ.MT145                   ................................................................................................-.------.-.---------..---....---G-   95
CPZ.CD.-.ANT                ..................................................................................................................................    0
CPZ.CM.-.CAM3               ..................................................................................................................................    0
CPZ.CM.98.CAM5              ................................................................................................-.------.-.---------..---....-AC-T  225
CPZ.GAB_X52154              ................................................................................................-.------..G---------..---...GTA--T  570
CPZ.GAB2                    ..................................................................................................................................    0
CPZ.TZ.-.TAN1               ......................................................CTTA............ACGC......................TC------.-.-C-----G...--A....-TG-T   99
CPZ.US.85.CPZUS             ................................................................................................-.------.-.---------..---....----T  575
GSN.99CM166_AF468659        G.................AGCCCTTTGCTAGAAGACGGTAGAATTTC..G....CTTG..CT.T...GCCTAAGGCT...TATGACT...........------..CT-CCAAAG-..-TAA.CGAG-CT  137
GSN.99CM71_AF468658         G.................AGCCCTTGCCTAGAGGACGGTAGATTCTC..GG...CTTG..CT.T...GCCCTAGGGC...TTTAGTGGCT......-.------.-CT-C-AAAG-..TACAGCAAG-CT  144
H2A.ALI_AF082339            G.....................................................CTTG..CT.T.....AAAG........ACCTCTT..........------.-.--.---AAT...TAGAAGCAGGT  727
H2A.BEN_M30502              G.....................................................CTTG..CT.T.....AAAG........ACCTCTT........-.------.-.--.---A.-..TTAGAAGCAAGT  732
H2A.ST_M31113               G.....................................................CTTG..CT.T....AAAAG........ACCTCTT..........------.-.--.---AGT...TAGAAGCAAGT  177
H2B.D205_X61240             G.....................................................CTTG..CT.T....AAAA.........CCCTCTT..........------...--.---A.A..TTAGAAGCAAGT  736
H2B.EHO_U27200              G.....................................................CTTG..CT.T....AAAA.........CCCTCTT..........------.A.--.----.A..TTAGAAGCAAGA  734
H2G.ABT96_AF208027          G.....................................................CTTG..CT.T...ACAAAG........ACCTCTT..........------.-.--GC-AAAT..TAGAAGTAAGCA  103
H2U.FR.96.12034             G.....................................................CTTG..CT.T.....AAAA........CCCTCT.........-.------.-.--.---AGT...TAGAAGCAAG-  228

MAC.239_M33262              G.....................................................CTTG..CT.T.....AAAG........CCCTCTT........-.------.-.--.---A-T..TTAGAAGTAAG-  697

SMM.SL.92.SL92B             G.....................................................CTTG..CT.T.................GACTCT.........-.------G-.--.--AACT...TAGAAGCAAGT  173
SMM.US.x.PGM53              G.....................................................CTTG..CT.T.....AAAG........ACCTCTT........-.------.-.--.---GCT...TAGAAGTAAG.  621
MNE.US.-.MNE027             G.....................................................CTTG..CT.T.....AAAG........ACCTCTT........-.------.-.--.---A-T...TAGAAGTAAG-  178
RCM.NG.x.NG411              ..................................................................................................................................    0
RCM.GAB1_AF382829           ..................................................................................................................................    0
DRL.FAO_AY159321            ..................................................................................................................................    0
MND-1.F17                   CAGAGTCCTGAGTGACTGGGTCTGAGCACCTCACTCGGGGCTGATCACCTCGAGGTAGTGGAACTCCTTGCTTGCTTGCTATTGTCTT........-.------.-.TAACTTAGAATTAGAGCAAG-GA  180
MND-2.5440_AY159322         ..................................................................................................................................    0
MND-2.14CG_AF328295         AGCAATCTC....TAGGCCAGGAGATTGTGGGTTAGACTACCG...........CTTG..CT.T.....GCT.......ATTGATGCT........-.------.-G---A---GA..TTAGAA.TAGCA  258
MND-2.MND2_AF367411         TCTCGG.......T.GGCCAAGAGGCTGTGGGATTGACTACCG...........CTTG..CT.T....GCTTT.......TAAGTGCT........-.------A-.---A--CAAGTTAAGAA.AAGCA  336
DEB.CM.CM40                 ..................................................................................................................................    0
DEB.CM.CM5                  ..................................................................................................................................    0
MON.CM.L1                   G.....................CCTAGAGAGTTGGTGAAGCTGACCCCGGCTCGCTTG..CT.T........TAGTA...AGGCCTTACAGCC...-.------.-........................  130
MON.NG1                     ..................................................................................................................................    0
TAL.CM.8023                 ..................................................................................................................................    0
TAL.CM.266                  AAAGCGTCCTGGGAGACCTTGCCGGTAAACGCTCAGTAGACATCCCGCTG....CTTG..CT.TCGCGAAGCGCTTTAAA......CATAGCC...-.------..............GTGTGTTCAAAG  145
MUS.CM.1085                 G.....................CCTAG...........GAGTACTGATTCTCGGCTTG..CT.T..TGCTTCGAAGC...TTATTGAAGCCCA...-.------.-........................  123
SAB.SN.-.SAB1C              GCGAC.TGGCCATTGCCAGTAGCAGAGACTCCG.....................CTTG..CT.T....GCTTGA.......TTGCCT.........-.------GT.--------A..GT-TACAAG-AA  693
GRV.ET.x.GRI_677            TAGCT...........................................................................................-.------CCTGC..T-GCT..TAGTCGCTA-AT  597
TAN.UG.x.TAN1               GGTGAGCCTAACCTGGTCTGGCCATCCAGGGGTAAGAGCCTC............CTTG..CT.T....AGAA.........GGCTATTG.......-T------.-..------GC..TTAAGAGT--AT  622
VER.DE.-.AGM3               ......................................................CTTGG.CT.T...AGAAAG.......................-T------.T.---CG--GC..-T-AGAGCT-CT  125
VER.KE.-.9063               ......................................................CTTGG.CT.T...GGAAAG.......................-T------AC.T------GC..-T-AGAGCT-AT  630
VER.KE.-.AGM155             ......................................................CTTGG.CT.T...AGAAAG.......................-T------..--------GC..-T-AGAGCT-AT  632
VER.KE.-.TYO1               ......................................................CTTGG.CT.T...AGAAAG.......................-T------..--------GC..-T-AGAGCT-AT  124
SUN.GA.98.L14               G...................................GGGAGCTAGAGGCTC...CTTG..CT.T...TGCAGTACA...AG...CCTT........-.------G..-C------A..GT-TTACTGCAA  664
LST.CD.88.447               ..................................................................................................................................    0
LST.CD.88.485               ..................................................................................................................................    0
LST.CD.88.524               ..................................................................................................................................    0
LST.KE.-.lho7               GCCTGATCA..............GCCTAGGGAGC.TAAGGGCTCCTAG......CTTG..CT......ACTTTAAA...G....CCTT........-.------AA.G----TCAT..T--TAGAAAG-A  692
SYK.KE.-.SYK173             G.GTCCGGGTTTG.....GCCGCCCGGTCTTCGGCAGCTCTCCGTTG.......CTTG..CT.T..TGCTTACGCT...GATCAGCCCA.........------GT.G-G-TAAAC..GCAAGTAA-G--  545
SYK.KE.KE51_AY523867        G..ACAAGGACTG.....ATCCACCTAGCCTCTACAATTCTCCGTTA.......CTTA..CT.T...TACTACGCG..CTGATCAGCC........-.------GT.G-GTTAACT..-CAAGTAA-GAG  186
COL.CM.-.CGU1               AAGGGCTTGAGGGACTC..........CAGTCCCCTCGCGCTAAGGGTAGATA.........GGTG.C.T...TTGCATTAT......AAGAG...-.------GT.GA-TTTGAA..GC-TCAA--AG-  183
```

```
      5'LTR R Repeat end \ / 5' LTR U5 start
H1B.FR.83.HXB2            A.AGTA...GTGTGTGCCCGT.CTGTT.GT.......GTGACTCT.GGTAACTAGAGATCCCTCAGACCCTTTTAGTCA.GTGTGG.................................AAAA.TCTC   629
H1D.CD.84.84ZR085         -.----..------------.-----TT.......--------.---------------------AGC-------C.-G----.................................----.----   143
H1F1.BE.93.VI850          .................................................................................................................................     0
H1G.SE.93.SE6165          ............................................................................................................GGCTT.---.----    12
H1H.CF.90.056             .................................................................................................................................     0
H1J.SE.93.SE7887          .................................................................................................................................     0
H1K.CM.96.MP535           .................................................................................................................................     0
H101_AE.TH.90.CM240       -.---G..------------.----G.T-.......AG------.---------------------AC-C---A-T.-A--A.................................----.----   175
H102_AG.NG.-.IBNG         -.----..------------.---A-T--.......--------.---------------------AC-C---A-T.----A.................................----.----   158
H103_AB.RU.97.KAL153_2    .................................................................................................................................     0
H104_cpx.CY.94.CY032      .................................................................................................................................     0
H1N.CM.95.YBF30           -..--G..----------A-.-CA--C-.........--A----.----C-----------------A-C-AGAC.T.-A---A..............................----.----   187
H1O.BE.87.ANT70           ........--------T-A-CTG.--CAACC..........--.---GT---------------T-AC--AGAC.T.-AAGCAG............................----.----   659
H1O.CM.91.MVP5180         ........--------T-A-CTG.--CAACC..........--.---GT---------------T-AC--AGAC.T.-AAGCAG............................----.----   632
CPZ.CM.04.EK505           -.----..----------A-C-AT-...CGGTAA......----.----T----------------T.AA--CTA-TGAGTA...............................----.----   170
CPZ.CM.04.LB7             -.----..----------A-T-AC-...-TGTGA......----.----T---------------A-...--GT-GTCTGCGC-G............................----.----   174
CPZ.CM.04.MB66            -.----..--------T-A-C-AT-...GTGTGA......----.----T---------------TA..AC-GTAGTCTGTGC-GCC..........................----.----   176
CPZ.MT145                 -.----..------------CTAT-...CAGTGA......----.----T---------------A-.A-C--TATTGA-AGA..............................----.----   173
CPZ.CD.-.ANT              .................................................................................................................................     0
CPZ.CM.-.CAM3             .................................................................................................................................     0
CPZ.CM.98.CAM5            G.--C-..---------TA-C-AT-CA.........--------.--------------------A....CA-T-AT.AA--A-.........................GTG.----.----   303
CPZ.GAB_X52154            G.--C-..---------TA-CTAT-CA.........-AC----.--------------------T....-A-A-T-.TA--CA........................AGG..----.----   648
CPZ.GAB2                  .................................................................................................................................     0
CPZ.TZ.-.TAN1             ..-AC-..----------A-TTCA-ACCGCGT....C--C-....CTGGGG---------------T...............................TTGTA..GTGGCTAAGTA.----.----   178
CPZ.US.85.CPZUS           -.----..---------TA-C-.T-CA.........--------.--------------------A....C-CT-A-.AA--A........................GGTG.----.----   652
GSN.99CM166_AF468659      GCTC-C....AT-CATTGG-.T-GAGGTCGGGCTTG.................................................................AGTCCG..GGTGGACACC-G-.--CA   191
GSN.99CM71_AF468658       GCTC-C....AT-CATTGG--CGG-A-C-GGCTTGG.................................................................AGTCGC..GGTGGATACC-G-.--CA   199
H2A.ALI_AF082339          T.-AAG..------T---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGTGTTCACCTG.AGTA--AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGAAAC..GGAGGCAGG----.--C-   845
H2A.BEN_M30502            T.-AGT..------T---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGTGTTCATCTGAGT-AC.AAGACCCTGGT.C---TAGGACCCTTCTCGCTTTGGGAATC..CAAGGCAGG----.--C-   850
H2A.ST_M31113             T.-AGT..-----CT---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGTGTTCATCTAAAGTA--AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGAAAC..CAAGGCAGG----.--C-   296
H2B.D205_X61240           ..-AGT..------T---A-CT.C-CCTAGTCGC.GCC--GTCAC..TCGGTGCTCC-CTGA......TAGACCCTGGT.C---TAGGACCCTTCTTGCTTTGGGAAAC..CGAAGCGGG----.--C-   847
H2B.EHO_U27200            C.-AGT..------T---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGTGCTCC-CTAG.......AAACCCTGGT.C---TAGGACCCCTTCTGCTTTGGGAAAC..CAAGGCAGG----.--C-   846
H2G.ABT96_AF208027        -.GTGT..------T---A-CT.C-CCTAGTCGCCGCC--GTCAC.CTCGGTGTTC-CGTAGC.....AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGRGAAC..CAGAGCGGG----.--C-   218
H2U.FR.96.12034           TA-.GT..------T---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGAG-TTC-C-TTAT-.....AACCCTGGTAC---TAGGACCCTTCTTGCTTTGGGA..AACCAGAGCAGG----.--C-   343

MAC.239_M33262            T.-.GT..------T---A-CT.C-CCTAGCCGCCGCC--GTCAA.CTCGGTACTCA--AATAAG...AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGAAAC..CGAAGCAGG----.--C-   813

SMM.SL.92.SL92B           C.-AG-..------T-T-A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGGGTCCCGGATAA-....GAG.CCCTGGTAC---CAGAACCCTCTCACTCTTGGGGGCAACTCCTGAGTG----A----   291
SMM.US.x.PGM53            C.-AGT..------T---A-CT.C-CCTAGTCGCCGCC--GTCA-.CTCGGTACTC--A-ATAA--..AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGC..AACCGAGGCAGG----.--C-   739
MNE.US.-.MNE027           C.-.GC..------T---A-CT.C-CCTAGTCGCCGCC--GTCAA.CTCGGTACTC..AATAAT-AG.AAGACCCTGGT.C---TAGGACCCTTTCTGCTTTGGGA..AACCGAAGCAGG----.--C-   294
RCM.NG.x.NG411            .................................................................................................................................     0
RCM.GAB1_AF382829         .................................................................................................................................     0
DRL.FAO_AY159321          .................................................................................................................................     0
MND-1.F17                 GT-AGTGTTA-CCA-TGTGCGC--C-CTTCT.................................-A---TG--GT--TCTCAT-TAGAGAACAGAAGGACTTCTAGTTA........ACCCT-GAAGC-   268
MND-2.5440_AY159322       .................................................................................................................................     0
MND-2.14CG_AF328295       -.GTA-..----ACT-GTCCAT.T--GTCCTTGAT............AG-AACTCT.GGTTA-T-G.AGA-CCCTCAGATC-GT-GCAGAGATTTGATATCTAGTGAG..AGTAGAGAA----.----   365
MND-2.MND2_AF367411       -.GTG-..----ACT-G-C-AT.T--GTCTTTGGT............AG-AACTCT.GGTTA-T-G.AGA-CCCTCAGATC-GT-GCAGAGATTTGATATCTAGTGGC..AGTAGAGAA----.--C-   443
DEB.CM.CM40               .........CG-GAGT-G--CCT--CGA--CCTGGTGA--AGAG-GGCC-GGGAGCCCT-G-TCACT.GGGA-AGCC-AG.C-CCA-CGGCCTGGGAGCCCTGAGCTGGG..GT.AGGCACTCCT.CGG-   115
DEB.CM.CM5                .........CG-GAGT-G--CCTA-CGA--CCTGGTAG--AGAG-GGCC-GGGAGCCCT-G---GCT-GTGAC-ACC-AG.C-CCA-TGGCCTGGGAGCCCTGAGCTGGG..GTTAGGTACTCCT.CGG-   117
MON.CM.L1                 ........................................-GCTT-A--GAGACTTG-GAGTCTG----CA-CTGTG.CG-ATCAGGAGCGGGCTTGAGCCGGGGAG..GGGACAGCAG-C-.G-C-   213
MON.NG1                   .................................................................................................................................     0
TAL.CM.8023               .................................................................................................................................     0
TAL.CM.266                -GCAGCA..--------AGCCCA-C-GTTGGAGCCGCG-C-GG-GCT---CC-GC-CCC--GG..................................GTCGAGGGAACTTGGTGGGTG--G-AAT-T   236
MUS.CM.1085               ...........................................TTG----AC-AAGG--AGT-TGC-CCTATT.CAT-CACACTGAGAGCCTATCTCAGCGTG..GTGTGAGTCTGG-.AACA   199
SAB.SN.-.SAB1C            -.GCA-..----------A-TT.A--CC-CAGGACAACCCT...................................................GGTTACTAAGGA..TCCCTGATAGAA----.--CT   762
GRV.ET.x.GRI_677          TGGAGTCAA---CTCATT..............GCTGC-CCGAG-CTCTAG-GGT--ACCT-TCTTACTG.........................GGTT..CTCCTG..TACCCAGGTGGG-G--ACTC-   683
TAN.UG.x.TAN1             C.T-AGCAA---CC.T-.AT-G-............GC-CCC.-TCCTC..-GGGA-CCCT-.GTTACTGG........................GGTT...TCTTA..T..CCAGGCTGAG---.----   700
VER.DE.-.AGM3             C.T-AG...TCAA------TCAT.TGAC-CCTCACTCTTGA--GGG.TAA---.TTCCT-A-TG..............................GGTTCTCTCTCA...ACCCAGGCGAG-G--.A---   213
VER.KE.-.9063             C.T-AG...TCAA------TCAT.TGAC-CCTCACTC..AAG-AGG.--A-C-GTTCCT-A-TG..............................GGTTCTCTCTCT..GACCCAGGCGGG-G--.A---   718
VER.KE.-.AGM155           C.T-AG...TCAA------TCAT.TAAT-CCTCACTC..-TGAA-G.--AG-AGTTCCT-A-TG..............................GGTTCTCTCTCA..AACCCAGGCGAG-G--.A---   720
VER.KE.-.TYO1             C.T-AG...TCAA------TCAT.TGAC-CCTCACTCTC-TGAA-G.--A-T-.TTCCT-A-TG..............................GGTTCTCTCTCT..GACCCAGGCGAG-G--.A---   213
SUN.GA.98.L14             G.CAA......----T-----GT-C-C.TCTTTCTCCTTCT....A.AACCCTGTTC-............AC-GCTC-T-.ACTA--GAGCAATTACTGGCCGCCTA..GTGATCCGGGGCTG-C.GGC.   764
LST.CD.88.447             .................................................................................................................................     0
LST.CD.88.485             .................................................................................................................................     0
LST.CD.88.524             .................................................................................................................................     0
LST.KE.-.lho7             C.-AGC..TA----CT-ATAGT.AG-CTCTTTCTCCTTC-C-CTG.-T-C-G-GATCTCT-ACT---GAGA--GGAG-C.T---.........................GATCCGGG-C-.GGG-   785
SYK.KE.-.SYK173           T.GC-C..TCA-C--TT-TGCTG-G-CACCTCCCGAA-GTG---A.--G-G-GC-..................................................................-   597
SYK.KE.KE51_AY523867      T...GC...TCTCA-CTGTTCTGCCGGC--CTCCCAGAAG.--G-AC--G--AC...........................................................G-G-   237
COL.CM.-.CGU1             G.TT-G...AGTAC-AGTTTCA-GTCCGCCTTCTC...............................-GCCCT....C....AGAGCGGGGCCTTTTTGCTG......TCCGGGAAGG---.GGC-   258
```

```
                     see Zhang, J. Virol 72(7):5664-5471(1998)
                      5' LTR U5 end \/ Lys tRNA binding site
H1B.FR.83.HXB2              TAGCAG...TGGCGCCCGAACAGG.G.ACCTGA................................................................................................  657
H1D.CD.84.84ZR085           ------...---------------.-.------................................................................................................  171
H1F1.BE.93.VI850            .......................-.-.--T---................................................................................................    8
H1G.SE.93.SE6165            ------...---------------.-.--T---................................................................................................   40
H1H.CF.90.056               .............................T---................................................................................................    4
H1J.SE.93.SE7887            .................................................................................................................................    0
H1K.CM.96.MP535             .................................................................................................................................    0
H101_AE.TH.90.CM240         --C---...---------------.-C--TC--................................................................................................  204
H102_AG.NG.-.IBNG           ------...---------------.A..-T---................................................................................................  185
H103_AB.RU.97.KAL153_2      .................................................................................................................................    0
H104_cpx.CY.94.CY032        .............................T---................................................................................................    4
H1N.CM.95.YBF30             ------...---------------.-.--T---................................................................................................  215
H1O.BE.87.ANT70             ------...---------------.-.--T---................................................................................................  687
H1O.CM.91.MVP5180           ------...---------------.-.-.-GCG................................................................................................  659
CPZ.CM.04.EK505             ------...---------------.-.--T---................................................................................................  198
CPZ.CM.04.LB7               ------...---------------.-.--T---................................................................................................  202
CPZ.CM.04.MB66              ------...---------------.-.--T---................................................................................................  204
CPZ.MT145                   ------...---------------.-.--T---................................................................................................  201
CPZ.CD.-.ANT                ......................--.-.--T---................................................................................................    9
CPZ.CM.-.CAM3               ..----...---------------.-.--T---................................................................................................   26
CPZ.CM.98.CAM5              ------...---------------.-.--T---................................................................................................  331
CPZ.GAB_X52154              ------...---------------.-.------................................................................................................  676
CPZ.GAB2                    .............................T---................................................................................................    4
CPZ.TZ.-.TAN1               --C---...---------------.-.--T---................................................................................................  206
CPZ.US.85.CPZUS             ------...---------------.-.--T---................................................................................................  680
GSN.99CM166_AF468659        -C----...---------------.-.--T---................................................................................................  219
GSN.99CM71_AF468658         -C----...---------------.-.G-T---................................................................................................  227
H2A.ALI_AF082339            ------G..T--------------.-.--T---................................................................................................  875
H2A.BEN_M30502              ------G..T--------------.-.--T---................................................................................................  880
H2A.ST_M31113               ------G..T--------------.-.--T---................................................................................................  326
H2B.D205_X61240             ------A..T--------------.-.-----G................................................................................................  877
H2B.EHO_U27200              ------A..T--------------.-.------................................................................................................  876
H2G.ABT96_AF208027          ------A..T--------------.-.--T---................................................................................................  248
H2U.FR.96.12034             ------G..T--------------.-.--T---................................................................................................  373

MAC.239_M33262              ------A..T-------T------.-.--T---................................................................................................  843

SMM.SL.92.SL92B             -G----T..T--------------.-.--T---................................................................................................  321
SMM.US.x.PGM53              ------A..T--------------.-.--T---................................................................................................  769
MNE.US.-.MNE027             ------A..T--------------.-.--T---................................................................................................  324
RCM.NG.x.NG411              ..........--------------.-.--T---................................................................................................   22
RCM.GAB1_AF382829           ..........--------------.-.--T---................................................................................................   22
DRL.FAO_AY159321            ..........--------------.-.--T---................................................................................................   22
MND-1.F17                   -TT---...---------------.-.--T---................................................................................................  295
MND-2.5440_AY159322         ..........--------------.-.--T---................................................................................................   22
MND-2.14CG_AF328295         C-----...---------------.-.--T---................................................................................................  393
MND-2.MND2_AF367411         C-----...---------------.-.--T---................................................................................................  471
DEB.CM.CM40                 AC--TTG..C-T--TA-GCTT-TATA-CC-AGA-...............................................................................................  147
DEB.CM.CM5                  AC--TTG..C-T--T.G-ATTTTATT-CC-AGA-...............................................................................................  148
MON.CM.L1                   -TCTCCA..G--------------.-.--T---................................................................................................  243
MON.NG1                     .................................................................................................................................    0
TAL.CM.8023                 .................................................................................................................................    0
TAL.CM.266                  ACC--CCAGC--------------.-.--T---GAACACCCAGGTTCTCTGTCCTGAGGGCCAAGCGGCTTCGGGGACCGCGGGTTTTGGGTCCCAAAGAAGAGAGAGAAACTCCGAGGGGTCGTGAGA  364
MUS.CM.1085                 CCT-GCA..G--------------.-.--T---................................................................................................  229
SAB.SN.-.SAB1C              ------...---------------.-.--T---................................................................................................  790
GRV.ET.x.GRI_677            .----....G--------------.-.--T---................................................................................................  710
TAN.UG.x.TAN1               C-----...---------------.-.--T---................................................................................................  728
VER.DE.-.AGM3               C-----...---------------.-.--T---................................................................................................  241
VER.KE.-.9063               C-----...---------------.-.--T---................................................................................................  746
VER.KE.-.AGM155             C----....---------------.-.--T-..................................................................................................  745
VER.KE.-.TYO1               C-----...---------------.-.--T---................................................................................................  241
SUN.GA.98.L14               ------...---------------.-.--A---................................................................................................  792
LST.CD.88.447               .................................................................................................................................    0
LST.CD.88.485               .................................................................................................................................    0
LST.CD.88.524               .................................................................................................................................    0
LST.KE.-.lho7               -TC---...---------------.-.--T---................................................................................................  813
SYK.KE.-.SYK173             CTT---...-------G.------.-.--T---................................................................................................  624
SYK.KE.KE51_AY523867        CTT---...---------------.-.--T---................................................................................................  265
COL.CM.-.CGU1               -T----...---------G-----.-.--T---................................................................................................  286
```

```
H1B.FR.83.HXB2             ............................................................................................................A.........AGCGAAAGG..G  668
H1D.CD.84.84ZR085          ............................................................................................................-.........--------T..A  182
H1F1.BE.93.VI850           ............................................................................................................-..........-------T..A   18
H1G.SE.93.SE6165           ............................................................................................................-.........--------T..T   51
H1H.CF.90.056              ............................................................................................................T.........--T-----T..A   15
H1J.SE.93.SE7887           ..................................................................................................................................    0
H1K.CM.96.MP535            ..................................................................................................................................    0
H101_AE.TH.90.CM240        ............................................................................................................-.........--------T..T  215
H102_AG.NG.-.IBNG          ............................................................................................................C.........G-TA-T---..-  196
H103_AB.RU.97.KAL153_2     ..................................................................................................................................    0
H104_cpx.CY.94.CY032       ............................................................................................................-.........--T-----T..T   15
H1N.CM.95.YBF30            ............................................................................................................-.........-A------T..A  226
H1O.BE.87.ANT70            ............................................................................................................-...........GTG--A-..T  696
H1O.CM.91.MVP5180          ............................................................................................................-.........-AGTG--A-..T  670
CPZ.CM.04.EK505            ...........................................................................................................A-G..TG..AA--TA-.......  210
CPZ.CM.04.LB7              ...........................................................................................................A-A..CG..AA--TA--......  215
CPZ.CM.04.MB66             ...........................................................................................................A-A..CG..AA--TA-G-.....  218
CPZ.MT145                  ...........................................................................................................A-G..TG..AA--TA---.....  215
CPZ.CD.-.ANT               ..........................................................................................................................GG--....   13
CPZ.CM.-.CAM3              ............................................................................................................-.........--TTG--A-.T-   38
CPZ.CM.98.CAM5             ............................................................................................................-.........--T-----..AA  342
CPZ.GAB_X52154             ............................................................................................................-..........-T-----..TA  686
CPZ.GAB2                   ............................................................................................................-.........--T-----..CA   15
CPZ.TZ.-.TAN1              ............................................................................................................G........A---AGGG...AA  217
CPZ.US.85.CPZUS            ............................................................................................................-........A-CGA--G...TA  691
GSN.99CM166_AF468659       ............................................................................................................-........ACAGCCC---..A  231
GSN.99CM71_AF468658        ............................................................................................................-........GCATCCC--T.GA  240
H2A.ALI_AF082339           ............................................................................................................-.........GAA--CT-A..-  886
H2A.BEN_M30502             ............................................................................................................-.........GAG--CT-A..-  891
H2A.ST_M31113              ............................................................................................................-.........GAA--CT-A..-  337
H2B.D205_X61240            ......................................................................................................................GAA--GT-A..-  887
H2B.EHO_U27200             ............................................................................................................G.........GAA--GT-A..-  887
H2G.ABT96_AF208027         ............................................................................................................G.........-AG--GT-A..-  259
H2U.FR.96.12034            ............................................................................................................-............--GT-A..-  381

MAC.239_M33262             ............................................................................................................-.........G-A--GT-A..-  854

SMM.SL.92.SL92B            ............................................................................................................-.........GA--G-CT....  330
SMM.US.x.PGM53             ............................................................................................................-.........G-A--GT-A..-  780
MNE.US.-.MNE027            ............................................................................................................-.........G-A--GT-A..-  335
RCM.NG.x.NG411             ............................................................................................................G.........--T--CT-A..-   33
RCM.GAB1_AF382829          ............................................................................................................G.........--T-GCT-A..-   33
DRL.FAO_AY159321           ............................................................................................................-............--C-.....   27
MND-1.F17                  ............................................................................................................-.........GAGAGGCACTGA  308
MND-2.5440_AY159322        ............................................................................................................C..........AA-GGTTC..A   32
MND-2.14CG_AF328295        ............................................................................................................C..........AA-GGTTC..-  403
MND-2.MND2_AF367411        ............................................................................................................C..........AA-GGTTC..-  481
DEB.CM.CM40                .........................................................................................................TAA-GTGTGGTAACTGC--GT-TGT  172
DEB.CM.CM5                 .........................................................................................................TAA-GTGTGGTAACTGC--GT-TGT  173
MON.CM.L1                  ..........................................................................................................GT-ACGACCCAGC-ACTCGA-CCC  267
MON.NG1                    ..................................................................................................................................    0
TAL.CM.8023                ..................................................................................................................................    0
TAL.CM.266                 GGACGACAGTTCCGGAGTCGAGCCCCGGATCGGCGAAGACGACGCGCGTCGCAAAGTGAAAGTAAGGTAAGGAGCCGATCGGTCACCCCGGGCGAGTAGCCGCCGTGAGCGAGTGCGGT--T-GGAAGA-  494
MUS.CM.1085                ..........................................................................................................TT-AGCCCGCGAG-A-GT---CG-  253
SAB.SN.-.SAB1C             ............................................................................................................C......TAA--A.CGTCCC.-  804
GRV.ET.x.GRI_677           ...........................................GAAGAGGCATCGGCACCGACCGCTGAGTTGCTGAGCGTCGGAGAGGGACGACTCAGGTAGGGTGAGAGCCTACGAGTTTTTT-CTAC  797
TAN.UG.x.TAN1              ...........................................GTCAGAAAGACTTCACTGAAGGGGAGGAAGGGCATCACCAGTCGAGTTCCCCGGACGCGGTAAGTCGGGACGCGGG-A-GCGAAGA-  815
VER.DE.-.AGM3              ............................................................................................................C.........T.T--GT-A..A  251
VER.KE.-.9063              ............................................................................................................GGTCCGAGGACTT--GT-A..A  766
VER.KE.-.AGM155            .........................................................................................................................--GT-A..A  752
VER.KE.-.TYO1              ............................................................................................................G.........T-A--GT-T..A  252
SUN.GA.98.L14              ............................................................................................................G.........G-A--G--A..-  803
LST.CD.88.447              ..................................................................................................................................    0
LST.CD.88.485              ..................................................................................................................................    0
LST.CD.88.524              ..................................................................................................................................    0
LST.KE.-.lho7              ............................................................................................................-.........G-A---GA-..-  824
SYK.KE.-.SYK173            ............................................................................................................C.........TAG-CC---..-  635
SYK.KE.KE51_AY523867       ............................................................................................................C.........T-A-CT---..-  276
COL.CM.-.CGU1              ............................................................................................................-.........--A-G.TCC..C  296
```

```
                                                                                                   Gag p7 nucleocapsid binds these loops, see DeGuzman, Science 279:384-388(1988)
                                                                                                    packaging signal secondary structure, see Harrison J Virol 72:5886-5896(1998)
H1B.FR.83.HXB2              A..AA.......................CC.....AGA...GGAG.........CTCTCT.CGACGCA...GGACTCGGCTT.GCTGAA....GCGCGCACG.GCAAGAGGC..GAGG.....GG...   734
H1D.CD.84.84ZR085           G..--.......................--.....---...----.........A-----.-------..-----------GCTGA--....------T--.---------..----.....--...   249
H1F1.BE.93.VI850            G..--.......................--.....---...-A--.........A-----.------GG..-----------GCTGA.-....-T--AACAC.-GC-AGA-G..CGA-.....A-...    85
H1G.SE.93.SE6165            -..-CAGGGAC.TCGAAAGC...GAAAGTT--.....---...-A--.........T-----.-------..-----------GCTGA.G....-T--A---A.---------..---A.....-...   137
H1H.CF.90.056               -..--.......................--.....---...-A--.........A-----.-------..-----------GCTGA.-....-T--A---A.---------..---A.....-...    80
H1J.SE.93.SE7887            ......................................................T-----.-------..-----------GCTGA.-....-T--A----.---------..----.....--...    54
H1K.CM.96.MP535             .................................................................................................................................     0
H101_AE.TH.90.CM240         -..-TAGGGAC.TCGAAAGC...GAAAGTT--.....---...-A--.........T-----.-------..-----------GCTGA.G....-T--A---A.---------..---A.....-...   301
H102_AG.NG.-.IBNG           .........AC.TCGAAAGC...GAAAGTT--.....---...-A--.........A-----.------A..-----------GCTGA.G....-T-----A.---------..---A.....-...   276
H103_AB.RU.97.KAL153_2      .................................................................................................................................     0
H104_cpx.CY.94.CY032        -..-TA.GGAC.TCGAAAGC...GAAAGTT--.....---...-A--.........T-----.-------..-----------GCTGA.G....-T--A---A.---------..---A.....-...   100
H1N.CM.95.YBF30             G..--.......................--.....G--...--CTGA.......A-----.-------..----------CGTTG......-T--A---A.--G------..----..........   290
H1O.BE.87.ANT70             G..--A......................--.....--G...-A--AA.......AA-CTC.-------AC.--G--------A--G-.-....-T--A-C--.CT-------..---A.....--A..   768
H1O.CM.91.MVP5180           G..G-A......................--.....--G...-A--AA.......AA-CTC.-------AC.--G--------A--G-.-....-T--A-CT-.CT-------..---A.....--A..   742
CPZ.CM.04.EK505             ............................--GG...--G.....CTGAAT.....-C---GG--CA-.....-.---------C.-T---.....-T--A---A.--G------..----.....C-G..   275
CPZ.CM.04.LB7               ............................-TGG...--C.....--TT.........-----GA--CG-.....-----------.------.....-T--A----.---------..----.....--A..   280
CPZ.CM.04.MB66              ............................--AG...---.....--TT.........-----GA--CA-.....-.---------.------.....-T--A----.---------..----.....--A..   282
CPZ.MT145                   ............................--GG...--G.....CTGAAT.......-----GA--CAC.....-----------.------.....-T--A-G-A.---------..----.....C-G..   281
CPZ.CD.-.ANT                ............................-TAC...---.....A-C..........--G-.GA--...-CA..---------TC-.-----.C...A-----GTA.---GA----..----.....C-G..    76
CPZ.CM.-.CAM3               G..--.......................--.....---.....--CTGA.......----TC.-------CG.-----------.-----.-....-T--A---A.--G------..----.....C-A..   108
CPZ.CM.98.CAM5              G..--.......................--.....---.....--CTGA.......----TC.-------C..-----------.A---.-....-T--A-G-A.-T-------..----.....C-A..   411
CPZ.GAB_X52154              -...........................--GG...--GCT.-A-T...........---C-GA--...--C..-----------.-----.-....-T--A---A.---------..----.....C-G..   753
CPZ.GAB2                    -..--.......................--G....GAGGAA--CTGA.........-----C.------GC..--------A---.----T-....-T--A--ACA--G------..----.....C-...    89
CPZ.TZ.-.TAN1               CG....................................................-GGC-C.-TGGA-GCA.-----------..G---CAGCGCAATCACAAGAGGC-----..G-A.........   276
CPZ.US.85.CPZUS             C..--.......................--AG...--GCTGAA-C..........T--.-GA--...-G..----------C.-----.-....-T--A---A.--G------..----.....C-A..   759
GSN.99CM166_AF468659        C..TTCAAGCC.CGAAGGCG...GGACCCAAG.....GAT...............TG-C-C.T-G-CG-GT.--T-CG----CCA-GA--....TTTGCTG--CAGCCCCTTGAA---AAGGTAA-GAG   328
GSN.99CM71_AF468658         C..TTCAAGCT.CGCAGGCG...GGACCCAAG.....GAT...............TG-C-C.T-G-CG-GT.--T-CG-.--CCA-AA......CTTGCTG--CAGC-CCTAGAA---AAGGTAA-GAG   334
H2A.ALI_AF082339            -..-GTCTAG..GAACACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGT---CCT....AGAAAGG--.CAGGCCAAG..-TAC.....CAA..   978
H2A.BEN_M30502              -..-GCCCTG..GAACTCGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGT---CCT....AGAAAGG--.CGGGCC-AG..-TAC.....CAA..   983
H2A.ST_M31113               -..-G.CCTTG.GAACACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGT---CCT....AGAAAAG--.CAGGCC-AG..-TAC.....CAA..   429
H2B.D205_X61240             -..GTCTTCAA.GAGCACAG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAATCAA.CC--GAC-GAGA---CCT....-TAAAAG--.CAGGCC..G..-TAC.....CAG..   978
H2B.EHO_U27200              -..GTCTTCG..GGACACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAATCAA.CC--GAC-GAGT---CCT....-TAAAA---.CAGGCC..G..-TAC.....CAG..   977
H2G.ABT96_AF208027          -..GTCCTTT..GAACACGG...CAGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAT-CAA.CC--GAC-GAGT-T-CCT....AGAAAGG--.CGGGTC..G..-TAC.....CAG..   349
H2U.FR.96.12034             -..GTCTTCTG.AAACTCGG...CTGAGTGAG.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGT---CCT....-TAAAGG--.CAGGCT..G..-TAC.....CAA..   472

MAC.239_M33262              -..G.ACTCCT.GAGTACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA-CAA.CC--GAC-GAGT---CCT....ATAAAGG--.CGGGTC..G..-TAC.....CAG..   944

SMM.SL.92.SL92B             .................CGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAATCAA.CCG-GAC-GAGTTTGACCCGGT--CA-TGG-CTG-GCG--A..-TAC.....A-GTA   416
SMM.US.x.PGM53              -..G..CTCCT.GAGTACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGT---CCT....ATAAAGG--.CGGGCC..G..-TAC.....CAG..   869
MNE.US.-.MNE027             -..G.ACTCCT.GAGTACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA-CAA.CC--GGC-GAGT---CCT....AGAAAGG--.CGGGTC..G..-TAC.....CAG..   425
RCM.NG.x.NG411              -..G-CCTCCG.ACGCGGGT...CCGGA..--.....---...A---TG.......ACTGG-.A--ACGCGG.ACC-AGTC-AGGTAA--G....CA--CTGGT.-GCTCGTAA..ATCT.....A-C..   124
RCM.GAB1_AF382829           -..G-CCTCCG.AGGCTAAG...GTTCGGCG-.....G-C....AG-TC.......AC-G-G.....-G-GT.---AC-TTGAC..........CA-GTA-GA.--TGCCT-GTG-CTTCAAAGT-C..   121
DRL.FAO_AY159321            ...GTCGACCC.CACCTGTG...AGGGACGAA.GGCGGC-GCCATCCGGACCGACC-A-C-GGAA-A-AGAGT-CTAAG..AAGCC-C--CGCGA-AC-CGC-TC-GT-AGTAA..A-A-GAAAGTAAAA   145
MND-1.F17                   CACTTGAGGCAGAGCACTCCGCCTGGAAGAAGCAGGTT--AGGA---TGGACTGGT--GAAGA--C-AGGAGGT--G--A-TGGGA----CTTTACAAGAATTAGTTGTACCCTAGTGTAAGGGGCAGCA   438
MND-2.5440_AY159322         G..GTCGCACC.CACCTGTA...AGGGACGAGCGGCGGC-GCA-.TCGGACAGACC-A-C-GG---AA.GAGTTA-T-GAAGAGCC-C--CGCGCAG-ACTCAAA-GT-AGCAG..TGCACTGTAATTGA   152
MND-2.14CG_AF328295         -..GTCGCTCC.GACCTGTA...AGGGACAAGCGGCGGC-GCC--TCGGACAGACC-A-C-GA---A-AGAGTTA-TCAA-GAGCC-C--CGCGCAG-ACTCAA.-GT-AGC-G..TGCACCGTATTGTG   524
MND-2.MND2_AF367411         -..GTCATTCC.AACCTGTA...AGGGACG.AAGGCGGC-GCCA-CCGGACCGACC-A-C-GG---A-TGAGTTA--CAA-GAGCC-C--CGCGCAG-ACTCAA.-GT-AGC-G..TGCACCGAACAGTT   601
DEB.CM.CM40                 CTGCCCTCATCATTGACTCCAGGGAATTCGGGTCGAGC-GGCTA-GCTC.AAGCGAGAGGTCCTAG-AGTTGGC-C-CGAA-AGG-ACTTGA.GCTAC-CAGA-CT-G.-CCCGCA-GC-GTGTGC-GAG   299
DEB.CM.CM5                  CTGCCCTCATCATTGACTCCAGGGGTCGGGAGCTCAAGCGGCTA-GCTCTGGGCAGGAGAACCTAG-AGTTGGC-C-CGAA-AGG-ACTTGAAGCATC-CAGT-TTTG--CCCGAA-GC-GAGTGT-GAG   303
MON.CM.L1                   GCAGGCGAACAGATCCGCGAGAAGCCTTGC-GAGTGCGCCCGACA-AGCTGCGCAG-AAGAGGTA-G-AGAGATCGGG---TCCCCTA-GGTGGCAGA-TAG--T-GG-CG--AACCTATCCAAGA-GGA   397
MON.NG1                     .................................................................................................................................     0
TAL.CM.8023                 ..........................CAGGGG.....G--...A-GT.........-G--TG.A-GGA-GCG..C-GG-A-GGGTCTGCGG...G-AAGT.CA-C-AC-CTTC-...--A..........    67
TAL.CM.266                  TGGC-GAGAGGCCTAGCTTAA...CCCAGGGGGGAGG-CGGAC---.........GA--TG...........C-GG-A-GGGTCTGC-....G-AAGTTCGTGCGT-CCTC-...--A..........   583
MUS.CM.1085                 TGTG-CGGAGAGAAGGCCCCAGCCGAGTGCGTGCCAAGCCACC----AACAAGTGCTG-G-AG-C-TAAGGTA--TGA-TTGGAC-GGTCTGGGAAG-GTGG-AGA-TTCCTCAGATC-AGCAGGCATAG   383
SAB.SN.-.SAB1C              G..CCGCAGAG.TAGACCAG...CCTGAGGA-.....TCC...TTT-AC.......TGAGTA.A---CGGAC.---AGAAAGGCG-ACCGT....--CG-AGT-.CAGGT-A-A..-GC-.....CCT..   897
GRV.ET.x.GRI_677            CTAGTCAGCGAGAAAGGCTAGGCCGCGACAGGGGCGCG-GTCCCATTAGTGGCAAC-AAC-CAGTTG-ACGAA--GT-G-TAGGG-AC-GGTCGGAGCA.............................   896
TAN.UG.x.TAN1               GATCCGCCAGGAAAAAAGTACTGCCAGGGAGGCAGGGGTGGTGCAG-TAGATCTAGG-GGGCGAC-GATCAGCACCTGG-TGAGCA...........................................   901
VER.DE.-.AGM3               G..GCACGTAC.AGCTGAGA...AGACGTCGG.....-CG...C--AGG.......AAGG-G.--GG-TGCG.ACGTGACCAAGAAGG-......CTTG-TGA-.TAGGCTTCT..CGA-.....T-C..   342
VER.KE.-.9063               G..GCACGTAC.AGCTGAGA...AGACGTCGG.....-CG...C-TAGG.......AA.C-G.--GG-TGCG.ACGTGACCAAGAAGG-......CTTG-TGA-.TGGGCTTCT..CGA-.....T-C..   856
VER.KE.-.AGM155             G..GCACGTAC.AGCTGAGA...AGACGTCGG.....-CG...C--AGG.......AA.C-G.--GG-TGCG.ACGTGACCGAGAAGG-......CTCG-TGA-.TAGGCTTCT..CGA-.....T-C..   842
VER.KE.-.TYO1               G..GCACGTAC.AGCTGAGA...AGGCGTCGG.....-CG...C--AGG.......AA.G-G.--GG-TGCG.ACG-GACCAAGAAGGAG-....CTTG-TGA-.TAGGCTTCT..CGA-.....T-C..   344
SUN.GA.98.L14               T..G-CGAAGA.GACGTGGA...GTGAAAC-T.....TC-...-A--AA.......G-GAGC.A-CT--TGA.-A-GAA--AAGGTT-CG-....C-TG-TGGA.AGGC-CC-G..--A-.....TTA..   896
LST.CD.88.447               .................................................................................................................................     0
LST.CD.88.485               .................................................................................................................................     0
LST.CD.88.524               .................................................................................................................................     0
LST.KE.-.lho7               C..GCTGGCAA.GAACCTCT...GGAGGGAGG.....-A-...A-C-CG.......-AGAAG.A-G-CAGGA.A-CTAG-A-AGA-AAAGT....CT--T-GG-.---G--A-A..--A-.....ACC..   917
SYK.KE.-.SYK173             -..CCCGAGAG.GAGGTGAC...GCCAGTCAA.....--G...-AGTCC.......GGA--C.AAGA-GGAT.C-G--GTC-G-CCTA-GC....CGA-C.GGT.-.CT-GT-G..CTC-TG...-TG..   728
SYK.KE.KE51_AY523867        -..CCCGAGAG..AGGAGAC...GCCAGTCAA.....--G...-ACCGC.......AGAGGA.ACGT--GCA.C--AGAA-AGAA--G-TGT....C---C-TATGC-G-TC---GACC--ATTGA-TAGG   379
COL.CM.-.CGU1               T..G-GGGTCG.CGGAAGAA...GAAGAGGAG.....-CT...-ATCGC.......---CTC.A--A-AGAG.-T-AGTC-AAAAAGG-G-....CTTACGGGT.-G-GTCAAG..----.....ACT..   389
```

```
H1B.FR.83.HXB2              .........CG...GCGAC.TG......GTGAGT.ACGCCA....AAA...............ATTTT...GACTAGC.GGAGG.....CTA.....GAAGGAG..AGAG....................   789
H1D.CD.84.84ZR085           .........-A...-----.--......------.----T..GAA--T...............T----...-------.-----.....---.....-------..----....................   306
H1F1.BE.93.VI850            .........--...-----.--......------.-----GAATTTTT...............T----...-------.-----.....---.....-------..----....................   144
H1G.SE.93.SE6165            .........--...-----.--......------.-----....A---...............T----...-------.-----.....---.....-----G-..----....................   192
H1H.CF.90.056               .........--...-----.--......------.-----..ATTTT................G----...-------.-----.....---.....-------..----....................   136
H1J.SE.93.SE7887            .........--...-----.--......------.-----.AAAA-T-...............T----...-------.-----.....---.....-------..----....................   112
H1K.CM.96.MP535             ..................................................................................................................................     0
H101_AE.TH.90.CM240         .........--...-----.--......------.-----.....--................-----...-------.-----.....---.....-------..----....................   354
H102_AG.NG.-.IBNG           .........--...-----.--......------.-----.....--................-----...-------.-----.....---.....-------..----....................   329
H103_AB.RU.97.KAL153_2      .......................................................................................................................GCG..GAGGCTAGAA    13
H104_cpx.CY.94.CY032        .........--...-----.--......------.-----....A--T...............T----...-------.-----.....---.....-------..----....................   155
H1N.CM.95.YBF30             .........--...---GA.A.......---T-A.GTA-GC.AATTTT...............GAC-G...-CGGT-G.CC--A.....AAG.....T-G-AGA..GAG-....................   347
H1O.BE.87.ANT70             ...A...CT-A...C.AGA.G-......------.-A.............................----...-.--G--.--T--.....-C-.....--CCT--.G-GAAGGG.CGA..AGTCTCTAGG   833
H1O.CM.91.MVP5180           ...A...CT-A...CAAGA.G-......------..............-...............-A---...-.--G--.--T--.....-C-.....--CCT--.G-GAAGGG.CGA..AGTCCCTAGG   808
CPZ.CM.04.EK505             ...C...GGA.....AA....-......------.---...C.AA.....................----...----G--.--T--.....-C-.....-G--A--............................   320
CPZ.CM.04.LB7               ...G...GCG....A-T....-......------.---...CCTA-C...................----...-------.--TA-G.....---.....----...............................   327
CPZ.CM.04.MB66              ...G...GCG....A-T....-......------.---...CCTA-T...................----...------GC-GT-G.....---.....---................................   329
CPZ.MT145                   ...C...GA-....T-A....A......---GTG.-GTA-GCAA......................----...-.-----.--T.-G.....---.....--CCT--.--GAAGGG.CGA..AGTCTCTA..   348
CPZ.CD.-.ANT                .........A....A-.....-......------.--CTT.CATTTTG..................----...-.T-GCT.A-T-A.....GC-.....ATCCT--.G-GAAGGT.CGA..AGTCTCTAGG   144
CPZ.CM.-.CAM3               ...A...CT-....CA--...-......------.G............................T----...-.-C---.--T--.....--G.....--CCT--.G-GAAGGG.CGA..AGTCTCTAGG   171
CPZ.CM.98.CAM5              ...C....T-....CT--...-......------........AT.TT-................T----...-.-C---.--T--.....--G.....--CCT--.--GAAGGG.CGA..AGTCTCTAGG   477
CPZ.GAB_X52154              ...C...GA-....T......-......------.----A-T........................----...-.-----.--T--.....---.....--CCT--.--GAAGGG.CGA..AGTCTCTAGG   818
CPZ.GAB2                    ..............-TAC...-......------.--A..........................T---CTTT-.-----.A-T--.....---.....--CCT--.--GAAGGG.CGA..AGTCTCTAGG   153
CPZ.TZ.-.TAN1               ..............CTCCGG--......------.--AAA..........................----...-.T-GT-.--T--.....GC-.....ACCCT--.--GAAGGG.CGA..AGTCTCTAGG   341
CPZ.US.85.CPZUS             ...A...CT-....CA--...-......------.--A............................----...-.-C---.A-T--.....--G.....--CCT--.G-GAAGGG.CGA..AGTCTCTAGG   823
GSN.99CM166_AF468659        ACCGGACGGGTCTA-AAGGG--GCAGACT-CTAG.G-CAGGTGGGTTGA.AGAAGGGTACCCACACAGAGG-GAG-AACT-T-AAAGAGGAGTAACA--CTCCCACCAGAGGATAGGGCAGA........   448
GSN.99CM71_AF468658         ACCGGACGGGTCTG-AAGGG--GCAGACT-CCAG.G-C.GGTGGGTTGAGAGAAGGGTACCCACACAGAGG-GAG-AACT-T-AAAAAGAAGTAACA--CTCCCACCAGAGGATAGGGCAGA........   454
H2A.ALI_AF082339            ...A...GG.C...CG-CG.--......TG---C.GG-AGTGAAG-GG...............CC-CC...-GG-G....A---.....TA-.....-T-CCTA..CACCAAAA.T.T..GTAGCC.GAA  1051
H2A.BEN_M30502              ...A...GG.C...-GCGT.GT......-GAGCG.GGAGTGAAAG-GG...............CC-CC...-GG-G....A---.....TA-.....-T-CCTA..CACCAAAA.ACT..GTAGCCAGAA  1058
H2A.ST_M31113               ...G...GG.C...-GCGT.GT......-GAGCG.GGAGTGAAAG-GG...............CC-CC...-GG-G....A---.....TA-.....-TGCCTA..CACCAAAT.ACA..GTAGCCAGAA   504
H2B.D205_X61240             ...G...CA......GCGT.GA......-GAGCG.GGAGG-GAAG-GG................C-CC...-GGAGTA.A..--.....TA-.....-T-CTCC..-A-....................T  1034
H2B.EHO_U27200              ...G...CA......GCGT.GA......-GAGCG.GGAGG-GAAG-GG................C-CC...-GGAGTA.AA---.....TA-.....-T-CTCC..-A-TTT..................  1037
H2G.ABT96_AF208027          ...A...CA......GCGT.GA......-GA-CG.GGAGT.GAAG-GG...............CC-CC...-GG-GCG.CAG-......TA-.....-TGCACC..-A-AATTT.GTA..GCCAAAAAGG   423
H2U.FR.96.12034             ...G...CA......GCGT.GA......-GAGCG.GGAGT.GA.G-GG...............TC-CC...-G...TT.-C---.....TA-.....-TGCA-C..-A-AAATA.GA...ACCAA..AGG   540

MAC.239_M33262              ...A...CG......GCGT.GA......-GAGCG.GGAGAGGAAG-GG...............CC-CC...-G...TT.-C---.....TA-.....-TGCA-C..-C-AAAAA.GAAA..........T  1009

SMM.SL.92.SL92B             AGAGA.........-GCCT.G.......TG-GAC.-GT.............................................A---.....T....................CCAAAA...........GGTG   451
SMM.US.x.PGM53              ...G...CG......GCGT.GA......-GAGCG.GGAGTCGGAG-GG...............CC-CC...-G...TT.-CG--.....TA-.....-TGCA-C..-AGAA..........GAGTCATA   934
MNE.US.-.MNE027             ...A...CG......GCGT.GA......-GAGCG.GGAGAG.AAG-GG...............CC-CC...-G...TT.-C---.....TA-.....-TGCA-C..-C-AAAAA.GAA..........AT   489
RCM.NG.x.NG411              ...G...AGAA...AA-TG.A-......-GCTAG.TA-AGCAAGGCG-...............CCCGG...TCAGGTA.CTG--CCTCTAGG.....-----GT..C--AGTCC.TAG..AGAAGGGAGG   208
RCM.GAB1_AF382829           ...T...AGA-A..AA-TG.A-CT....AGCGAA.GGAAGCAGGGC--...............CCCGG...TC-GGTA.CTG--CCTCTAGG.....-------..C--AGTCC.TAG..AGAAGGGAGG   208
DRL.FAO_AY159321            T..CAGTGCACTGT-A-G-GAAGGGTCTAGTG-G.C--TGTGAGC-GCGGCTAGGCAAAGGCA-C-CC...-TAGGCAAAA--AAAGAAAA-GGGAGA-GTA--GCCTGTTGCT.GAT..AAGTGAGGCA   266
MND-1.F17                   TAGTCAGAG-A.......................................................................................................................   449
MND-2.5440_AY159322         T..TGAGAGA-AGTC-AGGGATCCGCTT-A-CAG.G--AG-TCGCG--GGCAACCCCAGTAGGCAAAA...--AA--AA--G-AAGTAAGGCCGACGC--A-T-AA--TAAGGG.CTT..CCTCCGAAGG   273
MND-2.14CG_AF328295         G..TAGTGAGT...C-AGG.ATCCGCTT-A-CAG.G--AG-TCGCT--GGCAACCCCAGTAGGTAAAA...--AA--A.--G-AAGTAAGGCCGAAGC--A-T-AA--TAAAAG.GGC..CTTCCGAAGG   640
MND-2.MND2_AF367411         GTGTGTGTGA-T..C-AGG.GTCCGCTA-A-CAA.G--AG-CCGCCG-GGCAACCCCAGTAGGCAAAA...-GAA-AGA--G-AAGTAAGGCCGAGGC--A-T-AA-AGAGGTC.CTT..GTGAAAGAGG   721
DEB.CM.CM40                 TCTGGAAAC-CTGA--CGATCTCTTCCC-CCG-A.T.AG-TGCGC-GCAAAGGTAAGCGGACGCCGGAGAAT--AG-GCC-C-AATAAGGGCGGGTGTGT--TCTCT-CAGGCAGAGTACCACTGTGGAG   427
DEB.CM.CM5                  TCTGGAAAC-CTGA--CGATCTCTTCCC--CGAG.TGAG-TGCGC-GCATAGGTAAGCGGACGCCGGAGAAT--AG-GCC-C-AGTGGGGACGGGTGTGT-ACCTCT-CTGGCAGAGTGTCACTGTGGAG   432
MON.CM.L1                   AGGTTCCAGA-GTGCTCCAAGAGGGAGCA-A-AA.TAAGGGCACG.....................................................................................   441
MON.NG1                     ..................................................................................................................................     0
TAL.CM.8023                 ...C...AAGC..AC-CG........GTA--AA.GGTGACAGCTC-..................GAGGA...-GAG..........................CTC.AG-GACCTGCGGGTGGTGCTAAGTG   133
TAL.CM.266                  ...C...AAGC..AC-CG........GTA--AA.GGAGACTGCTC-....................GGA...-GAG..........................C-C.AG-GACCTGCGGGTGGTGTTAAGTG   647
MUS.CM.1085                 AGAGATCCTGCGGGAG-GGAGAAACTCG-A---G.-A-AG-CTCCC-CAGAGGACAGGGCCAG-..................................................................   446
SAB.SN.-.SAB1C              ...G...CGGA...C-CGA.GA......CACCC-.CG-G-TGTGTGTG...............TG-GA...--AG--A.---AAAAGAGAAG.....TC--ACA..G-GCTGAC.AAG..GCTGGCTTAG   981
GRV.ET.x.GRI_677            ..................................................................................................................................   896
TAN.UG.x.TAN1               ..................................................................................................................................   901
VER.DE.-.AGM3               ...C...GG.-...AAA-A.GC......TC---C.CTAGTTAGAGG-C...............TAGGA...AGGGCCG.TAGCC.....A--.....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGA   420
VER.KE.-.9063               ...C...GA.A...AG.-A.GC......TC---C.CTAGTTAGAGG-C...............TAGGA...AGGGCC..TAGCC.....A--.....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGA   932
VER.KE.-.AGM155             ...C...GG.-...AAA-A.GC......TC---C.CTAGTTAGAGG-C...............TAGGA...AGGGCCG.TAGCC.....G--.....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGA   920
VER.KE.-.TYO1               ...C...GG.-...AAA-A.GC......TC---C.CTAGTTAGAGG-C...............TAGGA...--GGCCG.TAGCC.....G--.....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGT   422
SUN.GA.98.L14               ...A...GGTA...AGTGA.G-......-CACT-.--CAGGCACTGTC...............-GGGG...ACAGGC-.A--A.....AG-C.....CT---CT..G-G-GGAG.GTC..CCGCTAGGTT   975
LST.CD.88.447               ..................................................................................................................................     0
LST.CD.88.485               ..................................................................................................................................     0
LST.CD.88.524               ..................................................................................................................................     0
LST.KE.-.lho7               ...T...GC.T...TGAGA.C-......AG--TA.-GAAGGACTTG-C...............TGC-A...--AGGCG.-C--A.....AGG.....T------..--TAAGAC.CTC..ACCGGGAGAG   995
SYK.KE.-.SYK173             ...T...ACAC...A---G.AC......-A-GTA.-GTGATCCAGCGC...............CAGAC...TG-CCA-.A-CA-TCGGGGA-.....AT---G-..-CC....................   794
SYK.KE.KE51_AY523867        TGAGTTCTC--GCGT-AGAC--CACGCG-C-GTC.GG-GATAAAT-GGGGGTC.............................................................................   431
COL.CM.-.CGU1               ...C...TA--...-TCTG.C-......CAA-AG.TGAAAGTAAA-GT...............GCGCG...CT-GTCG.C-CA-TCAGGAAG.....TGGCAGA..C---AGAG.CAC..AGAGAGGGAG   473
```

```
H1B.FR.83.HXB2            ..................................................................................................................................  789
H1D.CD.84.84ZR085         ..................................................................................................................................  306
H1F1.BE.93.VI850          ..................................................................................................................................  144
H1G.SE.93.SE6165          ..................................................................................................................................  192
H1H.CF.90.056             ..................................................................................................................................  136
H1J.SE.93.SE7887          ..................................................................................................................................  112
H1K.CM.96.MP535           ..................................................................................................................................    0
H101_AE.TH.90.CM240       ..................................................................................................................................  354
H102_AG.NG.-.IBNG         ..................................................................................................................................  329
H103_AB.RU.97.KAL153_2    .G..GA...GAGAG....................................................................................................................   21
H104_cpx.CY.94.CY032      ..................................................................................................................................  155
H1N.CM.95.YBF30           ..................................................................................................................................  347
H1O.BE.87.ANT70           .G..GA...GGAAG....................................................................................................................  841
H1O.CM.91.MVP5180         .G..GA...GGAAG....................................................................................................................  816
CPZ.CM.04.EK505           .......GAGGGAGG...................................................................................................................  328
CPZ.CM.04.LB7             .......GAGAGAGA...................................................................................................................  335
CPZ.CM.04.MB66            GG.....AGAGAGG....................................................................................................................  338
CPZ.MT145                 ....GGGGGAGAAG....................................................................................................................  358
CPZ.CD.-.ANT              AA..CA...GGAAG....................................................................................................................  153
CPZ.CM.-.CAM3             .G..GA...AGAGG....................................................................................................................  179
CPZ.CM.98.CAM5            .G..GA...AGAGG....................................................................................................................  485
CPZ.GAB_X52154            .G..GG...AGAAG....................................................................................................................  826
CPZ.GAB2                  TG..GA....GAGG....................................................................................................................  161
CPZ.TZ.-.TAN1             TAACAG...GGGAA....................................................................................................................  352
CPZ.US.85.CPZUS           .G..GA...AGAGG....................................................................................................................  831
GSN.99CM166_AF468659      ..................................................................................................................................  448
GSN.99CM71_AF468658       ..................................................................................................................................  454
H2A.ALI_AF082339          AG..GG...CTTGTTATCCTACCTTTAG...ACAGGTAGA...AGATT..GTGGGAG......................................................................... 1095
H2A.BEN_M30502            AA..GG...CTTGTTATCCTACCTTTAG...ACAGGTAGA...AGATT..GTGGGAG......................................................................... 1102
H2A.ST_M31113             .G..GG...CTTGTTATCCTACCTTTAG...ACGGGTAGA...AGATT..GTGGGAG.........................................................................  547
H2B.D205_X61240           TT..TG...CATTGTCTACTCTAAAGAG...GGAGTAGGG...CATAA..GTGGGAG......................................................................... 1078
H2B.EHO_U27200            ....GC...AATTCGTCTACTCTAGGAC...AGGGTACGG...AATAA..GTGGGAG......................................................................... 1079
H2G.ABT96_AF208027        GC..TG...TGAATCCTACCTTGAATAG...TCAGGTAGA...AGT....GTGGGAG.........................................................................  465
H2U.FR.96.12034           GT..TG...TCTGCCTCTGAGAAAGGAG.....AGGTAGT...ATATT..GTGGGAG.........................................................................  582

MAC.239_M33262            AG..CT...GTCTT.TTATCCAGGAAGG...GGTAATAAG...ATAGA..GTGGGAG......................................................................... 1052

SMM.SL.92.SL92B           TG..AA...TATCCTACCTACGGGATAG...GTAGAAGTA......AA..GCGGGAG.........................................................................  492
SMM.US.x.PGM53            GG..AC...TGAGTTCCCTACTTTTGAG...GAAAGAGTA...GGAGA..GTGGGAG.........................................................................  978
MNE.US.-.MNE027           AG......CTGT........CTTTTATCCAGGAAGGGATAATAAGATAGAGTGGGAG.........................................................................  532
RCM.NG.x.NG411            AA..A.............................................................................................................................  211
RCM.GAB1_AF382829         AA..A.............................................................................................................................  211
DRL.FAO_AY159321          AAAGAA...GCCTA.GGGGTAAGAAAAG......................................................................................................  290
MND-1.F17                 ..................................................................................................................................  449
MND-2.5440_AY159322       GAGAAG...CAGGGGGCAAGAAA...........................................................................................................  293
MND-2.14CG_AF328295       GAAGGG...CAGGGGGCAAGAAA...........................................................................................................  660
MND-2.MND2_AF367411       A........CAGGGGGCAAGAAA...........................................................................................................  736
DEB.CM.CM40               CAAGAGAGGCTCCTGGTCCACTAGGGACGTTTGTGCAAGTAGGCACGTCAGGGTAAG.........................................................................  484
DEB.CM.CM5                CAAGA..GGCTCCTGGTCCACTAGGGACGTTTGTGCAAGTAGGCACGTCAGGGTAAG.........................................................................  487
MON.CM.L1                 ..................................................................................................................................  441
MON.NG1                   ..................................................................................................................................    0
TAL.CM.8023               CGCACACGAGGCAAG...................................................................................................................  148
TAL.CM.266                CACACACGGGGCAAG...................................................................................................................  662
MUS.CM.1085               ..................................................................................................................................  446
SAB.SN.-.SAB1C            CA..GT...GGCAGAAGGCTAGGTCGAG...CGGGAAAGG...AAA.................................................................................... 1016
GRV.ET.x.GRI_677          ..................................................................................................................................  896
TAN.UG.x.TAN1             ..................................................................................................................................  901
VER.DE.-.AGM3             CG..GG...TACGCA...................................................................................................................  430
VER.KE.-.9063             C...GG...TACGCA...................................................................................................................  941
VER.KE.-.AGM155           CG..GG...TACGTA...................................................................................................................  930
VER.KE.-.TYO1             .G..GG...TACGCA...................................................................................................................  431
SUN.GA.98.L14             AG..TA...CTGTCAAGTAGGCAGACTA...GTAGTACGT...GGAAT..AAAATAAGAAGTGTGTGTATAACTCTGTGTCTCGGAAGAGTGTCTCAAAC.............................. 1062
LST.CD.88.447             ..................................................................................................................................    0
LST.CD.88.485             ..................................................................................................................................    0
LST.CD.88.524             ..................................................................................................................................    0
LST.KE.-.lho7             TG..GC...AGCAGGCCGGGGGCAGCCT...TGTGAGAAG...TGTGT..AATTCTGATCTAGTGTAACTGGGACACTAGTGGTAGCTTCACT..................................... 1075
SYK.KE.-.SYK173           ..................................................................................................................................  794
SYK.KE.KE51_AY523867      ..................................................................................................................................  431
COL.CM.-.CGU1             TG..CT...CGCCAGAGAAGGAGAAGAG...CTCATTAGG..AGGTCTTCAGGCAGGGAGTAGCGCCCCTGGGCAGGTCGAGGGGGGCCTGCAAGTGGAGTAAGCTGCGGCGCTCCAGGACCGGTAGAGG  593
```

PLV
complete genomes

```
                                     / Gag and Gag-Pol start (p17)
H1B.FR.83.HXB2            ...........ATGGGTGCGAGAGCGTCAGTA...TTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCAG  905
Gag                                  _M__G__A__R__A__S__V__.__L__S__G__G__E__L__D__R__W__E__K__I__R__L__R__P__G__G__K__K__K__Y__K__L__K__H__I__V__W__A__S__R  p17

H1D.CD.84.84ZR085         ...........---------------------...------------A----G---GCT------------------C-------------------G-----GC-------------------------  422
H1F1.BE.93.VI850          ...........------------------A--...------------A--------GA--------------A---------G-------------G-------A-G------C--A-------------  260
H1G.SE.93.SE6165          ...........---------------------...----C-------A--------GCT-----------------G-------------G----TC-------A--------C----------------  308
H1H.CF.90.056             ...........---------------------...--------C---A--------GCT-----G---------C---------------------------GGC--------C----------------  252
H1J.SE.93.SE7887          ...........------------------A--...-----T------A--------GAT-----------------G-----------G-------------GGA----G---C----------------  228
H1K.CM.96.MP535           ...........---------------------...------------A--------GC---------G---------C--------------------------C-G------C----------------  116
H101_AE.TH.90.CM240       ...........---------------------...-----T------A--------GC------------------GC------------GA----------GGC-G------T----------------  470
H102_AG.NG.-.IBNG         ...........---------------------...-----T------A--------GC------G-------------------------------------G-C--------C----------------  445
H103_AB.RU.97.KAL153_2    ...........---------------------...------------A--------GC----------------------------------G---------G-A-------CC----------------  137
H104_cpx.CY.94.CY032      ...........---------------------...-----T------A--------GC------G-GG----------------------------------G-C-G------C----------------  271
H1N.CM.95.YBF30           ...........--------------------G...----CA------A---------A-------TC----TAT--G--A--------------------C-G-A-G------T----------------  463
H1O.BE.87.ANT70           ...........-----------T-----T--G...--G-CA--AA-CA----G---GC-------C-----A------A------ATCT--A--G-------G-C-------CT----------------  957
H1O.CM.91.MVP5180         ...........-----------------T--G...--G-CA---A-TA----G---GC-------CG----A-------------ATCT--A--GGC-----GGC--------T----------------  932
CPZ.CM.04.EK505           ...........---------------------...----CA--A---A---------A-------------TAT--G--A--------------------C-TGA-G-----CT----------------  444
CPZ.CM.04.LB7             ...........------------------A-C...----CA--C---A-G------GAT---------G-----C-------C-------GA----G-----TGA-G------T----------------  451
CPZ.CM.04.MB66            ...........--------------------T...----CA------A-------CA-------G------TATC-G-------------------------TGA-G------T----------------  454
CPZ.MT145                 ...........------------------A--...----C-------AG-------GCT------------A--C----A--------------------C-TGA-G------T----------------  474
CPZ.CD.-.ANT              ...........-----A---G-G-----T--T...--G--G--A-AGA-GC-----AC--------GT--CA--C-TC----C--T--C--------G--C-TGA--------C-G--T--------ATC  269
CPZ.CM.-.CAM3             ...........---------------------...----CA------CG-------GCT------------A--C---------------------------TGA-G------T----------------  295
CPZ.CM.98.CAM5            ...........---------------------...----CG--A---CG---G---GCT-----G------A--C----M-----A----------------TGA-G------T----------------  601
CPZ.GAB_X52154            ...........--------------------T...C---CA------A----------C---------G-----C-T-----C-------GA----G-----TGA-G------T----------------  942
CPZ.GAB2                  ...........-----------------TA--...----C---C--GCG-------GCC-----G------A--C-C--A----------G-----G---C-TGA-G-----------------------  277
CPZ.TZ.-.TAN1             ...........--------------------G...--G--G--A-ATA-GC-G---AC-------TCC--AA--C-T-AAT-CA-A--C-G----------TT-A--------C----------CG-A--  468
CPZ.US.85.CPZUS           ...........--------------------T...C-G-CA------CG-------GCT--------G------C-T--A--T--------------G----TGA-G------T----T-----------  947
GSN.99CM166_AF468659      ...........-----A-----GCAC--G-CGATGC-G-CA--TACTA-G--G---A-G-AT--G---G-G--TC-C--A---AA------------G--C-TGA-C--G--CT----------CTC--A  567
GSN.99CM71_AF468658       ...........-----A-----GCAC--G-CGATGC-G-CG--TACTA-G--G---A-G-AT--G---G-GA-TC-C--A--GA-A--------G--G--CGCGA-C--G--CT----------CTC--A  573
H2A.ALI_AF082339          ...........-----C------AAC--C--C...--G--A---A--A--GC---CGA--TA----G----A-----C----C--C---------------C-GC----------T--G------GCG-A  1211
H2A.BEN_M30502            ...........-----C------AAC--C--C...--G--A---AA-A--GC---CGA--TA------G--A-----C----C--C-----------G--C-GG-----------T--G------GCG-A  1218
H2A.ST_M31113             ...........-----C------AAC--C--C...--G--A---AA-A--GC---CGA--TA-----G---A-----C----C--C----------------GGC----------T--G------GCG-A  663
H2B.D205_X61240           ...........-----C-------GC--C--C...C-GTCA---AAGA--AC----GA--TA------G--A-----C----C--C-----------G----TGC--------G-----------GTG-A  1194
H2B.EHO_U27200            ...........-----C-------GC--C--C...--GTCA---AAGA-GAC----GA--TA-----GG--A-----C----C--C----------G---C-TGC--------------------GTG-A  1195
H2G.ABT96_AF208027        ...........-----C------AGC--C--C...C-GTCA---AA-A--GCT---GA--TA------G--A-----C----C--C--------------C-TG---------G-T--G------GCG-A  581
H2U.FR.96.12034           ...........-----C------AAC--.--C...--GTCA---AAGA--G-----GA--TA------G--A-----C----CAAC--------------C-TG---------G-T--G------GCG-A  697

MAC.239_M33262            ...........-----C-T----AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A--C--C-A--CAAC-----------G--C-TG--G--G---G-----------GCA-A  1168
Gag                                  _M__G__V__R__N__S__V__.__L__S__G__K__K__A__D__E__L__E__K__I__R__L__R__P__N__G__K__K__K__Y__M__L__K__H__V__V__W__A__A__N  p15

SMM.SL.92.SL92B           ...........-----C-------GC--C--C...--GTCA---AA-A--GC----GA--TA------G--A-----C----C--C----G------G--C-TGC-T--G-----TA--------GC---  608
SMM.US.x.PGM53            ...........-----C------AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A-----C----CAAC-----------G--C-TG--G--G---G-----------GCA-A  1094
MNE.US.-.MNE027           ...........-----C------AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A--C--C-A--C--C--G--------G--C-TG--G--G---G-----------GCA-A  648
RCM.NG.x.NG411            ...........-----------G------A-T...--GTCA---AAGA-G------GCT------C-GG--A-----C----C--C-----------G----TG-----G---C-----------T----  327
RCM.GAB1_AF382829         ...........--------------C---T-G...--GTCA---AAGA-GC----CGC-------TC-G--A-----C----C--C--G--A--G-----C-TGC-G--G---T-G---------T---A  327
DRL.FAO_AY159321          ...........-----A-----C-------G-...C-T--G--A-A-A-GC-T---GCT-TT-----GG--A-----C----CTCC-----A--G--------GC-C--G---C-T--C----T-GCG-A  406
MND-1.F17                 ...........------AATG-GAAC--T-CC...--GTTA---ACT--T--G---AA--TT--G-----AA-A----A-AG---T--T--A---TGT----G---G-----CC-CTGT---TGT-AAG-  565
MND-2.5440_AY159322       ...........-----C-----C-------G-...C-T--G--A-A-A----G---GA-CT------G---A-----C----CTCC-----------G--CC-G---------G-TA------T-----A  409
MND-2.14CG_AF328295       ...........-----C-----C-------GT...C-T--G--A-AGA----G---GA-CT-----C----A-----C----CTCC--------G-----CC-G------------A------T-----A  776
MND-2.MND2_AF367411       ...........-----C-----C-------GT...C-T--G--A-A-A----G--CGAGCT------G---A-----C----CTCC-----A--G-----CC-GC-G-----C---A------T-----A  852
DEB.CM.CM40               ...........-----G---G-GCG---G-C-CCA----C-----CG-CT-----CA---TT------G-AA-A--G--A---AAA--C--A--G--G----TGA----G--CC-C--C------G-T-A  603
DEB.CM.CM5                ...........-----G---G-GCGA--G-CGCCA----CG----C--CT-----CA---TT--G-----AA-A--G--A---AAA--C--------G---TTGA----G--CC-C--C--------T--  606
MON.CM.L1                 ...........-----G-----GCAC--G-CGATGC-T--T--TACTA-------CAAG-AT--G--GG-G---C-T--A--GC-C--T-----G--G--CCTGA-T--G--------------TGC--A  560
MON.NG1                   ..............................................................................................................................  0
TAL.CM.8023               ...........-----G--AC-GCAC--G-CGATGC-C-C----A-G-CC------AAG-AT--G---G-GA-AC-------GAAA-----A--G--G---G-GG-G-G----G-G--------TTC--A  267
TAL.CM.266                ...........-----G--AC-GCAC--G-CGATGC-C-C---TA-G-CC------AA--AT--G---G-GA-AC-G-----GAAA--G--A--G--G----TGG-G--G-----T--------TTC--A  781
MUS.CM.1085               ...........-----A-----GCAC--G-CGATGC-G-CA--TAACA-G--G---A-G-AT--G--GG-GA--C-C--A---AAA--------G-----CCTGA-C--G--C--TA-------TTC---  565
SAB.SN.-.SAB1C            ...........-----------TAAC-----C...-----T--TA-GA--C-----GC--TC---TC-G-CA-----C----CAAC--------G--G-----G----G----C---------------A  1132
GRV.ET.x.GRI_677          ...........-----C-G-G-TCAC----C-...C-GTCA---A--AGCC-C--CACG-TC--G--G---A--C--C-T--GAAC--G--A--G--G--CC--A-T------T--A--------G-A-A  1012
TAN.UG.x.TAN1             ...........-----A---G--CAC--G-C-...C-GTCA---A-GA-T--G--CAC--TT--G-----A--C---C-C--GAAC-----------G---C----------CT--A--------G---A  1017
VER.DE.-.AGM3             ...........-----G---GCTA-C----C-...C---ATA--A--C-------CAA--TT--GC-T--A--AC-TC-C--GACC-----------G--CC--A-T------T--A--------G---A  546
VER.KE.-.9063             ...........-----G---GCTA-C----C-...C---AA---A--C-------C-A--TC--G-----A--A---C-C--GAAC---------------C--A-T--G---T--A--------G-T-A  1057
VER.KE.-.AGM155           ...........-----G---GCTA-C----C-...C-G-ATA--A--C--------GA--TT--GC-T--A--AC-TC-C--GAAC-----------G---C--A-T------T--A--------G---A  1046
VER.KE.-.TYO1             ...........-----G---GCTA-C----C-...C---ATA--A--C-------C-A--TT--G-----A--AC-TC-C--GAAC-----------G--CC--A-T------T--A--------G---A  547
SUN.GA.98.L14             ...........-----A---G--AAC--CAC-...G-GGATA---A--TTG-GAGGA-C-TT----G-G-AGCTC-G-AA--T--A--G--A--G-CC--CC-GGC--G----G-----------G-T-A  1178
LST.CD.88.447             ...........-----AT--G-GAAC--T--C...C----TA--CA-ATTGAG-GAGA--TCTGC-GT--CA--C----A--T-ACA-T--A--G-CT---C-GAAG-G----T----G-------CA--  116
LST.CD.88.485             ...........-----AT--G-GAAC--T--C...C----TA--CA-ATTGAG-GAGA--TCTGC-GT--CA--C----A--T--CA-T--A--G-CG---C-GAA--G------T--G-------CA-A  116
LST.CD.88.524             ...........-----AT--G-GAAC--C--C...C-C--TA--CA-ATTGAGA-AGAT-TTTGT--TG-CA--C----A--T---A-T--A--G-CT---C-GAA--G----G---AC-------CG-A  116
LST.KE.-.lho7             ...........-----AT--G-TAAC--T--C...C-C--TA--CA-ATTGA-A-AGAT-TTTGT-GTG-CA--C----A--T---A-T--A--G-CT---C--AA--G----G---A--------CG-A  1191
SYK.KE.-.SYK173           ...........-----A--AGCG-G---G-C-ATA----CA--TA----GC----CA---AT--G---------C-CC-A--GAAA--G-----GCGC--CCT-G-TCG----T----T-------AG-A  913
SYK.KE.KE51_AY523867      ...........-----A--AG-G-GA--G-CGATCC---CA--TA--A-GC-T---AA--AT--G---------C-TC----GAA---G-----GCGC--CCT-G-TCG----T----T------G-A-A  550
COL.CM.-.CGU1             GAGGTCATTA.-----CAACGAGCA-GG-C-G...---G-GAA-AA-AC------AGACCTAC-G---G-C---C-G-A-AA----AAG--AGG-TGT-----GA--------G-TAG----ATGT---C  719
```

```
H1B.FR.83.HXB2           GGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTA  1035
Gag                      __E__L__E__R__F__A__V__N__P__G__L__L__E__T__S__E__G__C__R__Q__I__L__G__Q__L__Q__P__S__L__Q__T__G__S__E__E__L__R__S__L__Y__N__T__V_   p17

H1D.CD.84.84ZR085        A-----G--------T---C-----------T------------------------A-------A---A-------------G-T---A--------------------------A--------------   552
H1F1.BE.93.VI850         ---------------T---C--G----------TC--------------------CA-A-----A-AA-------------------------------------G----AG-------T----------   390
H1G.SE.93.SE6165         ------G--GA----T---C----C----A---T---------G-------T---CA-------A--A-----T-G--------T--C---------A----G---A---A-------------G-----   438
H1H.CF.90.056            ------G---A----T---C----C--C-----T---------C-----------CT---G---A-A-A----A----G---G-TA--A--------A------------A--------T---CT-----   382
H1J.SE.93.SE7887         ------G--CA----T---C----C--------TC----GT--G--A--------CA---------A-T------C------G-T--C---------A--C-----A---A------G------------   358
H1K.CM.96.MP535          ---------------T---C-C--C--------T-----G---A-----------C-G------A-AAC---AA----G------A----A--------------GA---A----C-----------A--   246
H101_AE.TH.90.CM240      A---T-----A--------C----C---A--T-T---------G-------A---CA-------A-A-A----T----GT--A-T--CA---------T-----------A--------T----------   600
H102_AG.NG.-.IBNG        ------G---A--------C----C--------T---------G----G--A---CA----C--A---A----T-----T--G-T--CAG--------------------A--------T-------A--   575
H103_AB.RU.97.KAL153_2   ------G---A-------GC----C---A----T-----------------A---CA-----------A----T--------A-T--CA---------------------A-------------------   267
H104_cpx.CY.94.CY032     A---T-G---A--------C----C--------T---------G-------A---CA----T--A---A----T-----T--A-T--CA-A-----------------------------------TA--   401
H1N.CM.95.YBF30          ------G---A-------TTG---C--A--T--CA-G--C---G-G--C------GCCA-GT----AAAT--AT--G-----G-T--CA-------G-----------GC-C--T--------CG-TC--   593
H1O.BE.87.ANT70          A-----G---A--------TG--------AG--AC-------TG----G--TAA-GAGA-GC-G--ACAG---T--G-G---G-T--CA-------G-----CAGC--GCAG---C-C-GG--CG--A--  1087
H1O.CM.91.MVP5180        ------G---A---A----TG---------T--A--------TG-------TAC-GAG---C-G--ACAG---T--G-G---G-T--CA-------G-----G--C--G-A---TC-C-GG--CG--A--  1062
CPZ.CM.04.EK505          ------G---A-------GTG---C--A--T--CA-G--C---G-------T---GCC---T----CA----AT--G-----G-T--CA-A-----G-------G---GC-C--TC---T---C--CC-G   574
CPZ.CM.04.LB7            ------G--GA----T--TTG---C--C--G--TA-G------A-----------GAG---T-GA-TAC----T--G-----G-T---A-------G------A-C----A---T----T---C--CA--   581
CPZ.CM.04.MB66           ------G---A-------TTG---C--CTC---CA-G------A------------AG--GC-TT--CAG------G-----G-T--CA-A-----T-------G---------T--G-T---C--CA--   584
CPZ.MT145                A-----G---A-------TTG---C-----T--AA-G----ATG-------T---GAGA-GT--T--AA--------G-----G-T---A-------G-------G---GC-C--T----T---C--TT-G   604
CPZ.CD.-.ANT             ------GC-G--T--T--GC-C-GCT-CTC---TC----------------T---GA-A-GGCTA-CCAT--AT-GAGC--T---A-AG-A-T-A----CCCT---A-A-T---T--G-T---C--CA-T   399
CPZ.CM.-.CAM3            A-----G--CA-------TTG---C-----T--AA-G-----GG-------A---GAG--GT--T--AA-G--AT--G-----G-T--CA---------------GC--GC----GC-T-T------TC--   425
CPZ.CM.98.CAM5           ------G---A----T--TTG---C-----T--AA-G--G---G-T-----A---GA-A-TC----AAA----T--G-----G-T--CAG------------G-GCT-GC----GC---T---C--TC--   731
CPZ.GAB_X52154           ------G---A--------TG-G-C--C--G--AA-G----GTAAG-----A----CTA--T-GT-ACA---AT--G-G---G-T--CA-A-----C-------G---GC-G--C--G-T---C--TC-G  1072
CPZ.GAB2                 ------G---A-------TTG-G-C-----T--AA-G------AA----------TAT--GT--T--AA-------G-----G-T---A-A-------------G-T-AC-G--C----T------TT-G   407
CPZ.TZ.-.TAN1            C--A---C-G--T-----GA-G-----C--T--CA-G--G-ACGT---------CT-GA----CA-CCTC-----G--G--T--GG-AG-C-TT--T--TCC----A-C-TT--T--G-T------CA-C   598
CPZ.US.85.CPZUS          ------G---A----T--TTG---C--A--T--AA-G------G----C------CTC---T-T--AAA----T--G-----G-T--AA-A--------G----GC--GC----GC-C-T---C--CT--  1077
GSN.99CM166_AF468659     A-----T--GA---AT--GC-GTCCGAC-CT--TC-T----ATCA---G--G---C---GG---T-A-A-ATTG-CTTC--CCTAGAA-CA-----G-------CT--C-A---TC-C-T-GGA-TC--G   697
GSN.99CM71_AF468658      ------T--GA---A---GC-ATCAGGCA-T--TC-T---T-TAA------G---CA---GG--T---A-GTCACCTCC----TAGAA-CA-AT--G--C---TCT--C-A-AGCC-C---GGA-TC--G   703
H2A.ALI_AF082339         T--AT-G--CA------G-T-GGCAGAAA-------G--GT--AA------T--CCA-AG---T--TAA-GTTT--G------TAG-G-CA------------A-TT-A-A-AGCC-T-T------T--C  1341
H2A.BEN_M30502           T--AT-G--TAA-----G-T-GGCAGAGA-------G--GT--AA------T--CCA-A-G--T--CA--GTTT--G-T----TAG-A-CA-----G------A-TT-A-A-AGCC-T-T------C--C  1348
H2A.ST_M31113            T--AT-G--CA------G-T-GGCAGAGA-------G--GT--AA---G--T--CCA-A----T--TAC-GTTT--G-T----TAG-A-C------G------A-TT-A-A-AGCC-T-T------T--C   793
H2B.D205_X61240          C--AT----CA----T-G-T-AGCAGAGA-------G---T--AA------A---CA-A----T--TAA-GTCT--GCT----TAG-A-CA-----G------A-TT-A-A-AGCC-T-T-----TC--C  1324
H2B.EHO_U27200           T--AT----GA----T-G-T-GGCAGAGA---G---G-G-T--AA------A--CC-GA----T-G-AA-GTCT--GG-----TAG-A-CA-----T------A-TT-A-A-AGTC-T--------C--C  1325
H2G.ABT96_AF208027       C--AT-G--CA------G-T-AGCAGAGA-------G---T--AA------T--CCA-A----T--TTC-GTAT--GCT----TAG-A-CA-----T------A-TT-A-AGAGCC-T--------T--C   711
H2U.FR.96.12034          C--AT-G--CA----T-G-T-AGCAGAAA-------G--GG-CAA------A---CA-A----TT-ATCGGT-T--GCT----TAG---CA-----T------A-TT-A-AGAGCC-T--------T--C   827

MAC.239_M33262           T--AT----TA----T-G-T-AGCAGAAA-------G--G-ACAA------A---CA-A-------TTCGGTCT--GCT----TAG-G-CA-----C------A-TT-A-A-AGCC-T--------T--C  1298
Gag                      __E__L__D__R__F__G__L__A__E__S__L__L__E__N__K__E__G__C__Q__K__I__L__S__V__L__A__P__L__V__P__T__G__S__E__N__L__K__S__L__Y__N__T__V    p15

SMM.SL.92.SL92B          A--AT----CA----T-G-TCGGCAGAAA-------G----GCAA------A---CA-AG----T-A-C-GTA---GCT----TAA-G-CA-----C--G---A-TT-A-A-AGC--G-T--G---T--C   738
SMM.US.x.PGM53           C--AT-G--CA----T-G-T-AGCAGAAA-------G--G--CAA------T---CA-A----TA-TTC-GTTT--GCT----TAG---C-G----T------A-TT-A-AGAGCC-T--------T--C  1224
MNE.US.-.MNE027          T--AT----TA----T-G-T-AGCAGACA-------G--G-ACAA------A---CA-A-------TTCGGCCT--GCT----TAG-G-CA-----C------A-TT-A-AGAGCC-T--------T--C   778
RCM.NG.x.NG411           ---A-----GA----T-GCT-G-G-GA-ACTT---------GTG-------AATCCAGA----TA-A--GGTTA-T-TG--TCTAG---C------G--T--G-GCT-A-A---CC-C-T----TTGTGC   457
RCM.GAB1_AF382829        AA-A---A-TAA---T-GCT-G-G-GA-CATT-----------G--AC---A---GA-A-----T-A---GTC--G-TG--TCTAG---C------G-----G-GG--A-A-AGCC-C-T----TTGTGC   457
DRL.FAO_AY159321         ---A-----T-------GCC-GC--GAGA-AT-A-------GTCAG-----G--CCA-A----TT-A---GTA--CTTT--TCTAG---CT-----G-----GA-TT-A-TT--TC-C-T------CTGC   536
MND-1.F17                T--AT----TA----T-GCT-ATCGGA-AAA--CC-T------CAGC----A---GA-A----C--CTC-GTATGTTGG----TATA-G-CCA---------TA-T--A-A-G-T--GGTAGGG-----C   695
MND-2.5440_AY159322      ---A-----CA----T-GCC-AC--GAGAAGT---------GTCAG-----A--CGA-A----T--TA-CGTA--CTTT---CTAG---CT-----G------A-TT-A-TT--GC-G--C--C--CTGT   539
MND-2.14CG_AF328295      ---A-----T-----T-GCC--C--GAAAAA--A--------TCAG-----A--CGAGA----T--TA-CGTA--CTTT---CTAG---CT-----G-----GA-T---CTTG-GC-G--C--C--CTGT   906
MND-2.MND2_AF367411      ---A-----CA----T-GCC--C--GAAAAG----------GTAAG-----A--CGAGA----T--TA-CGTA--CTTT---CTAG---CT-----G------A-TT-A-TC--GC-G--C--C--CTGC   982
DEB.CM.CM40              ---A-----CA----T-GCC--GCAG-A-CAT-A-------ATAAG---------CAGG----TT--ACTAG-T----G--T-TAGAAGCA-----CAGT--GA-TT-G-A---TC-----GGC-T---C   733
DEB.CM.CM5               ---A-----CA----T-GCC--GCAGGA-CAT-A--G-----CAA------A---CA-G-G--CT-AACTAGAT-------T-TAGAGGCAT----CAGT----GC----AG--CC----CGGC-T----   736
MON.CM.L1                ------T--TA------GTC-CTCGGACTCG--CC-G--G---CAG--C-------AGA-G--TT---A-GTTA-T-TT--TCTA-AG-C---------C--GTCTA-C-A---TC-C-T-GG--TC-CT   690
MON.NG1                  ..................................................................................................................................     0
TAL.CM.8023              A---A-G---A-G--T-GCC-G-G-GAC-C---T--G--G--CAA------G--CGTTA-G--CA-A-A-GCCA-TTCT--G-TAGAA-CA-AT--C-------G-A---AG--TC-C-T-GGA-TC--G   397
TAL.CM.266               A---A-G--GA-G--T-GCC-G-G-GAC-C---------G--CAA------G--CGTTA-G--CA-A-A-GCCA-TT-CATG-TAGAA-C--AT--C-------GGA-A-AG--TC-C-T-GGA-TC--G   911
MUS.CM.1085              ------T--TA----T-G-C-CTC-GA-TC---T--G---T--AGG--T--G---CA-A-G--CA-A-AGGTCA-C-TC--C-TAGAA-CAG----G--C--GTCC--C-A---C--C-T-GGA-TC--G   695
SAB.SN.-.SAB1C           ---A-----T-----TAGCC--TCAG-AAAT--C-----G---AAG-----GGTCGTCA----T--TA-TGTA--C-TT--CCTAG---CT-----G-----GA-TT-A-TTG-TC-C-T----TT-TGC  1262
GRV.ET.x.GRI_677         A--AA-G--------T-GGT-AC--GAGAAA--T---------AA----------CA-A----CA-A-A-GTTT--ACC--G-TGGAA-C------C--C--G-GG--A-A-G-TC-G-T----TTGTGC  1142
TAN.UG.x.TAN1            AA-AA-G--G--C--T-GGC-CC-CGAGAAA--C---------AA---G------CA-A-G--CA-A-A-GTAT-GAGC--GCTAGAA-CA-------------G---A-A-AGTC-G-T----TTGTGC  1147
VER.DE.-.AGM3            ---AA-G--G--C----GCC-CC--GAGA-AT-AC-----T--GA------A----AGA-G--CA-A-A-GTA--CT-C--GCTAGAA-CA-----G--G--G-GCT-A-A-AGTC-G-T---CCTT--G   676
VER.KE.-.9063            A---A-G--G--C----GCC-CC--GA-A-GT-----------GA---T--A----A-A-G--CA-A-A-GTC--TTTT--T-TAGAA-CA-----C--G--G-GCT-A-A-AGTC-G-TC---CTT--G  1187
VER.KE.-.AGM155          -A--A-G--C--C----GCC-CC--GAGAAGT-A--G--G---GAG-----T----A-A-G--CA-A-A-GTT--CTCT--CCTAGAA-CA-----G--G----G-A-G-A-AGTC-G------CTG--G  1176
VER.KE.-.TYO1            ----A-G--G--C----GCC-CC--GAGA-GT----G--G---GAG--G--G----A-AG---CA-A-A-GTC--CT-C--CCTAGAA-CA--------G--G-GCT-A-A-AGTC-G-TC---CTT--G   677
SUN.GA.98.L14            A--AT----TA-G--T-G-C----CAAG-AA-------AG---GTG---------CAGA-G--T---TC-GTCTGTTGG--T-TGTA-GCAT-T--T--TA-GA-C--G-AGG-C---GTAGG-------  1308
LST.CD.88.447            A--AT----TA-G--T-G-T-AGGAG-ACA-T--C----G---G----T-----C-AGA----T--T---GTCTGTTGG---CTATACTCC-----C--GA-GA-TT-A-A-G--C-TGTAGGA--T---   246
LST.CD.88.485            A--AT----TA-G--T-GTT-AGGAG-ACA-T--C-G--G---G----T-----C-AGAG---T--T---GTCTGTTGG---CTATACTCC-----C---A-GA-TT-A-A-G--C-TGTAGGA--T---   246
LST.CD.88.524            A--A-----TA-G--T-GGC-AGGAG-AAA---TC--------G----G-------AGA-G--TT-----GTC-GTTGG--T-TATA--CT-----C--TA--A-T--G-A-G-TC--GTAGGG-----G   246
LST.KE.-.lho7            A--AT----CA----T-G-T-GGGAT-ACAA--C--G------G-------G----AGA----C--ATCTGTCTGCTGG--T-TATA-GCA--T--T--TA-GA-C--A-AGG-C---GTAGGG------  1321
SYK.KE.-.SYK173          ----T-G--CA------GCC--TCGGACCAA--CA-G--GT-TAA----------GA-A----T--CAC-GTC----TT----TAGAAGCT-AT--G------A-TT-A-A-AGCC-T-T-GGC-T-A-T  1043
SYK.KE.KE51_AY523867     ------G--CA------GCC-CTCAGACCAG--TC-T----GTAA------T---GA------CA-AAA-ACCA-CTT-----TAGAAA-ACAT--C------A-TT-A-A--------T-GGC-T-AC-   680
COL.CM.-.CGU1            A---G--AGC--C-GT-TTT-A-TATT--AG------A-GT--G--AC---AGTGGCT------T---A-A--G-GACG----TAG-GG-C-------------TGT-G--G--CC-----GGA...TGT   846
```

```
H1B.FR.83.HXB2             GCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAA...........AACAAAAGTAAGAAAAAAGCACAGCAAGCA..........  1143
Gag                        _A__T__L__Y__C__V__H__Q__R__I__E__I__K__D__T__K__E__A__L__D__K__I__E__E__E__Q__.__.__.__.__N__K__S__K__K__K__A__Q__Q__A__.__.__.__   p17

H1D.CD.84.84ZR085          -----------------A---A----------------------------------------------------A---.............-----C----------GG-------------..........   660
H1F1.BE.93.VI850           ---GT--------A---A---------GC--G-G---C------------------------C-----------A---.............---------C--C----GA----------G..........   498
H1G.SE.93.SE6165           -----------------A---------------G-------------A-----C----AG-AG-G--AA-GATA---.............--G------C--G-----AT------G---..........   546
H1H.CF.90.056              ---GT---------C--A----G--AA-----TG-G--------------G--------T------------ATA---.............---------C--C-----A-------------..........   490
H1J.SE.93.SE7887           ---------------C--A-----G--------A--------------T-----------G------------ATT---.............-----G-AC--AC-GC-G------A-----..........   466
H1K.CM.96.MP535            ---GT--------T---A-------A-------G-------------------C--------AC----------A---.............---------C--CG---GA----A---..........   351
H101_AE.TH.90.CM240        ----------GG--C--A--C-------------G--------------------T--A-----------TA---.............--T--G--CC--CG---GA-------G---..........   708
H102_AG.NG.-.IBNG          --------T-GG--C-----------------C----------------------C-----T--A--G------ATA---.............--T--G--C---C----G-T------GA--..........   683
H103_AB.RU.97.KAL153_2     -----------------A------C-----------------------------------T--A---------ATA---.............--T--G--C---C----GA-C--A--G---..........   375
H104_cpx.CY.94.CY032       A---------GG--C--A--------A-----TG--C--------------------T--A---------ATA---.............-GT--G--C---C----GA-------G---..........   509
H1N.CM.95.YBF30            ---GTT--T--------C---AGT------C-----C-CA----AC----------G---------A-A--GA-A--G...GAACAGCAC--GCCCGAGCCA-----CC--G-AGC--GG..........   710
H1O.BE.87.ANT70            -T-GTG----GG-----T--CA-C--ATATA-A--TGG---T--GC--C-G--AA--C-A---T--A------TAATG...........GGG-GC--G---TCTGCG-ACGCCGCTAAG..........  1195
H1O.CM.91.MVP5180          ---GTA----GG--C--T--CA-C--AT-T--C--CCG---T--AC--C-G--AA--C-A---T--A------TAATG...........GCA-GC--G---TCTGC--AGGCCGCTAAG..........  1170
CPZ.CM.04.EK505            ---GTT--T--C-----T---A-T-AC---A--G--C-GA-T--AC-------------A--A...TTAAG-......GAGAAAATGAAA...GC-GAA.............--AA--..........   667
CPZ.CM.04.LB7              -TGGT-----GG--C--------G--A--C-C-------------------G--------T-CA...CTAAC-......GCAGTTAGACAG..---CAG.............--AA-G..........   674
CPZ.CM.04.MB66             -T-GT---T-GG--C---------G----CCCAG-------T--A--A-----------TC--...CTAC--......GAAGCGCAGCAG..---G-A...............A--..........   674
CPZ.MT145                  ---GT---T-GG-----A---A-G--A-----AG-GG-------AC-AC-G------C-G--A...CT-A--......GCTGCCGTGAAC...CC-GAA.............---GGT..........   697
CPZ.CD.-.ANT               TGTGTT--G-GG--C--A---A--G--GA-A------------AG-AC----CG-TA-A-CAG-GA-AATGA-AGT-.............-TGC-G-CAC-AGC-G--A--GGAAGTAGCCAAACCGCAA   517
CPZ.CM.-.CAM3              ---GTG--T-GG-----T---A-G--A--TTCAG--G-G-----AC--C-G------ACA--AT-GA-A----CAGTT.............-CGGCCTCCCGAG--C-G-A-GT-GCCCA-..........   533
CPZ.CM.98.CAM5             --GGTA--T-GG-----T---A-G--A--CC-AG--G-G-----AC--C-----C--TCA--AC-GA-A---ACAGTT...........GC-GCCCAAG-A...C-G-A-GT-GCCCA-..........   836
CPZ.GAB_X52154             ---GTA--G-GG--CA-A---AGTGAC--CACTG--G-------AC--A-----C---AC--C--A--.........................CGGCA-C-TGG-G--CA---AGCAA-ACTGAAAGTA  1178
CPZ.GAB2                   --TGTGT-A-GG-----T---GCT--A-----AG--G----T--AC--C----C---C-A---T-GA-A---TTTGTG...........C-A--GGAA-CAGC--C-------G-----..........   515
CPZ.TZ.-.TAN1              TGTGTA-----C--C--A--CGC-G-AGA-AGAG-CC----T--GG-A-----AG-CA-A-TTG-GA-AATGA-A-T-.............-CTGT-CAG--A--T--CT-CACAGCGA--......TCTA   710
CPZ.US.85.CPZUS            --GGT---T-GG-----C--CAGT---G-GACAG-GG-G--T--GC--C-G--C----TG--AC-GA-A--G-TA................GTGC-GTTG......C-GAA---A--G-A-..........  1176
GSN.99CM166_AF468659       T--GTG---------A-C---GC-GAC---C--G-GG-G--T--AG--C-G---AAGC-GC-AG-CAG-AT-AGATGTCATCTAGAGGAG-C---GG----AG----GCAGA-TA---AC..........   817
GSN.99CM71_AF468658        T--GTG---------A-C---GC-GAC---TG-G--G----T--AG--C-G---AAGA-GC-AG-TC--AT-AG-TGTCATCTAGTAGGA-CA---GAG--AG---C-CAGA-CA---AC..........   823
H2A.ALI_AF082339           TGCGTAG-T-GG--C-----CGC-GAAGAGA-AG-G-----T--TG-A-G---AAA-C-A-TA---...C-GAGA--TCTAGCGGCAGAA....................................  1428
H2A.BEN_M30502             TGCGT-A-T-GG--CT----CGC-GAAGAGA-AG-G-----T--TG-------AAAGA-ACTAGC-...C-GAGA--TCTAGTGGCAGAA....................................  1435
H2A.ST_M31113              TGCGT-A-T-GG---A-A--CGC-GAAGAGA-AGCG-----T--TG-A-----AAA-C-A---G--...C-GAGA--TCTAGTGGCAGAA....................................   880
H2B.D205_X61240            TGCGT-A-T-T---CC----CGC-GAAGAGA-AG-G-----T--AG-------AAA-A-G-TAGC-...C-GAGA--TCTAGCGGCGGAC....................................  1411
H2B.EHO_U27200             TGCGT-A---T---CC----CGC-GAAGAGA-AG-G-----T--TG-------AAA-A-G-TAGC-...C-GCGA--TCTAGCAGCGGAC....................................  1412
H2G.ABT96_AF208027         TGCGT-A-T-GG---T----CGCCGAACAGA-AG-----C-T--TG-A-----AAA-C-AGTAG-G...C-GAGA--TCTAGTGGTAGAA....................................   798
H2U.FR.96.12034            TGCGT-A-T-GG---T----CGC-GAAGAGA-AG-G---C-T--TG-------AAAGC-G-TAG--...C-GAGA--TCTAGCGGTGGAA....................................   914

MAC.239_M33262             TGCGT-A---GG--CA-T--CGC-GAAGAGA-AG-G---C----TG-------AAA-C-G-TAG-G...C-GAGA--CCTAGTGGTGGAA....................................  1385
Gag                        _C__V__I__W__C__I__H__A__E__E__K__V__K__H__T__E__E__A__K__Q__I__V__.__Q__R__H__L__V__V__E__.__.__.__.__.__.__.__.__.__.__.__.__.__._   p15

SMM.SL.92.SL92B            TGCGTAG-T-GG---C----CGC-GAA--GA-AG-G--------AG--------AAGA-A-CAG--...C-GAGC--TCTAGTGGTAGAA....................................   825
SMM.US.x.PGM53             TGCGT-A-T-GG--CA-T--CGC-GAAGAGA-AG-G---C-T--TG-------AAA-C-A-TAG-G...C-GAGA--TCTAGTGGTGGAA....................................  1311
MNE.US.-.MNE027            TGCGT-A---GG--CA-T--CGC-GAAGAGA-AG-G---C----TG-------AAA-C-G-TAG-G...C-GAGA--CCTAGTGGTGGAA....................................   865
RCM.NG.x.NG411             TGCGTA--T-GG--C-----CA-GGAAG-GA-AG-G-----T--AG-------GG-G-CGC-TG--A-AA-GC--TGCCATCTAGTGGAC--AG-TGAG--TGC-GG--A---AG--AA-..........   577
RCM.GAB1_AF382829          TGCGTA----GG--C--A--CA-GGAAG-GA-AG-G--------AG--------G---CA--AG-GA-A---TGCTGCCATCTAGTGGAA--AGC-GAA--T-C--C--A-A-AG--AAG..........   577
DRL.FAO_AY159321           TGCGTAG-T-GG--C--T--CTCG-AAG-GA-AG---C------AG--------G-TC-A--AG--A-AC--C-CTGCCATCTAGTGGAT--A--TGAG--TGC-GCTT--A-AA-T-A-AATGGGGAAA   666
MND-1.F17                  TGTGTTG-AGCC--CA-A--CGC-G-T-----A--T--GAG---AC-A--T------A-A--AT--A-A-TCATAAC-.................................................   773
MND-2.5440_AY159322        TGTTG-A-T-GG--C--A---GCG-AAG-GA-AG---C---T--TG-------AAA---G--AG--A-AC--CG-T-CCACCTAGTGGTTG-A-G-GAG--TGC-GCCT--G-AG---A-AAAGGAGCA.   668
MND-2.14CG_AF328295        TGTTG-A---GG--C--A---GCG-AAG-GA-AG---CT--T--AG-A--G--AAA-A-G--AG---AC--TGCTTCCATCTAGCGGCC--AGG-GAA-GTGC-GCCT--G-AA---A-GAGAAAGCA.  1035
MND-2.MND2_AF367411        TGCTG-G-A-GG--C--A---GCG-AAGAGA-AG---C---T--AG-------GAA---G--AG--A-AC--A---TCCATCTAGTGGCCG-A--GGAA--TGC-GC-T--G-AA---A-CAAAGAGCA.  1111
DEB.CM.CM40                TGTGT-A---GGGCATGC---GCC-ACC-CTC-G-GG-T------G---G--CAA-A-GGCAG-GT-CAGGAG-AT-AAGCAGGTGGAA......................................C   824
DEB.CM.CM5                 TGTGT-G-T-GGGCATGC--CGCC-ACC-CTC-G--G----T---G---G--AAA-A-GGCAG-G-CTAG-AG-AT-AAGCAGGCAGAG.......................................   826
MON.CM.L1                  T-GGTAT-G-----CA-A---GCCG-T-----A--TG-G-----AG-AC-G---AAGC-GC-AG--A-AATCAG-TGTCACTTAGCAGGG.......................................   780
MON.NG1                    .......................................................................................................................................     0
TAL.CM.8023                TGTGTT--A--C--CA-T---GC-G-T-----A---G-G--T--AG-AC-G--CAAG-CAG-AG--A--A--CGCTGCCACCCAGTGGAG--A--GGAGG---GC--G-G-ACAAC--AG..........   517
TAL.CM.266                 TGTGTT--A--C--CA-T---GCTG-T-----C--TG-G--T--AG--C-G--CAAG-CAG-AG--AG-A--CGCTGCCACCTAGGGGAG--A--GGAA-CA-C-G-G-AGG-AA--.............  1028
MUS.CM.1085                T--GTG----GG---A-C--CGC-GA-G-GC-TG-GG-------AG----G---AAGA-GC-A--CAGA-C-AG-TGCCATCTAGAGGAG.......................................   785
SAB.SN.-.SAB1C             TGCGTGT-AGC----A-T--CGC-GAA---A-AG-G-----T--AG-------AAA--CG--AG--A-A--G--AGTGCCAGCAGAAATG-CAG-----GCC-C-GCGA--TCTAGT-GC..........  1382
GRV.ET.x.GRI_677           TGCGT-A-T-GG--CA-T--CGC-GAACAGA-AG-G--------AG--------G---TA-CAG-TA--C--C-CT-CCATCTAGTGGAC--A--T.................................  1238
TAN.UG.x.TAN1              TGCGT-A-T-GG---A-T--CGC-GAACAGA-AG-G-----T--AG-------AG-G-TATTAG--A-AC--CGCTGCCATCTAGTGGAA--AG---AA-CTGC-GC---G-CATCT-GT..........  1267
VER.DE.-.AGM3              TGCGTAT-G-T---C--A--CA--GAT-AG--AG-G--------AG-A-----AG---CA-TAG--AGAC--TGCTGCCATCTAGTGGAG--AG----A--TGC-G--AG-A-TAC-A--..........   796
VER.KE.-.9063              TGCGTG--A-----CA-C--C-GCGAACAGA-AG-G--------AG-------GG---CA-CAG--AGAC--TGCTGCCATCTAGTGGAA--GG---AA-GAGC--C--AG-CATCTAGT..........  1307
VER.KE.-.AGM155            TGCGTAT-GCT---C--C--C---GAA-AGA-AG-G--------AG-----------CA-TAG--AGAC--TGCTGCCACCTAGTGGAC--AG---AA-CTGC-GTTA-G-CA-CT-GT..........  1296
VER.KE.-.TYO1              TGCGTA--A-----CT----CA-GGAACAGA-AG-G--------AG-------AG---CA-CAG--AGAC--C-CTGCCATCTAGTGGAA--AG---AA-GTGC--C--AGACATCTAGT..........   797
SUN.GA.98.L14              TGTGTG--TGCC---TGT------G-C--TCCAG----------AC----G---C--A-A---G-CAGACT-----C-GCAAAAAATGAAG-A.....................................  1401
LST.CD.88.447              TGTGTG--G-G---CTGC---TT-G-----A-AG--GC-------C-A------A-GC-A------A-AAT---A-C-ACTCAGGAAGAA-GGC-G.................................   342
LST.CD.88.485              TGTGTG--G-G---CTGC---TT-G-----A-A---GC-A-----GCA------A-GC-A------A-AAT---A-C-ACTCAGGAAGAC-GGC-G.................................   342
LST.CD.88.524              TGTGT-A---G---CTGT--C-T-G-A---AGAG-C--T--T--AC-------AA-TA-T--AG-GA-AATT-----GAACAAACCTGTG.......................................   336
LST.KE.-.lho7              TGTGT-A---GC--CTGT--C-T-G-AG--AGA---TC------TC-------CA-CAGT--AG--A-AAT---A-C-GCACCAGAAGCAGCAGGG.................................  1417
SYK.KE.-.SYK173            T-GGT-G-A-GGGCC--T---GCC-AA-AG--AG--G-------TG--C----AAA-C-G--AG--A-A----CATGTAATTGGAAAGATG--CC-CCAGC--C-TCT-GTGGA---AGT..........  1163
SYK.KE.KE51_AY523867       ---GTAG---GGGCC--C--CGC-GAA-A---CG-GG----T--GG-AC----AAA-C-G--AG-GA-A--G-CTTGCAATTGGGAAGATG-GG---CAGT--C-TCTAGTGGC---AA-..........   800
COL.CM.-.CGU1              T-TGTGTGT--C--CC----CAG--AATGGA-T--TG-G-----TC-A--G--AGA-A-G--AG----A----CTT-T.................................................   924
```

```
H1B.FR.83.HXB2               ..GCAGCTGAC.....................ACAGGA.........................................................CACAGCAAT...CAGGTCAGC.........CAAAA 1181
Gag                          .__A__A__D__.__.__.__.__.__.__.__T__G__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__H__S__N__.__Q__V__S__.__.__.__Q__N  p17

H1D.CD.84.84ZR085            ..-------G-.....................------.........................................................A------GC...---------.........-----  698
H1F1.BE.93.VI850             ..---------.....................-A---G........................................................................-----T.........-----  524
H1G.SE.93.SE6165             ..---ATG--T.....................-A----.........................................................A-------C...--A------.........-----  584
H1H.CF.90.056                ..--------T.....................-AG-A-.........................................................A-AGA---C...A-------T.........-----  528
H1J.SE.93.SE7887             ..-A-A-----.....................-A-AA-.........................................................G---A--G-...--------T.........-----  504
H1K.CM.96.MP535              ...................................-A-GCAGCT......................................................GA---A...GG------T.........-----  380
H101_AE.TH.90.CM240          ..-------G-.....................------.........................................................AG-----GC...A-A------.........-----  746
H102_AG.NG.-.IBNG            ..-------C-.....................------.....................................................................AGCAG----.........-----  712
H103_AB.RU.97.KAL153_2       ..---A---G-.....................------.........................................................AG-----GC...A-------T.........-----  413
H104_cpx.CY.94.CY032         ..-------C-.....................G-----.........................................................GGT----GC...A-T------.........-----  547
H1N.CM.95.YBF30              ..-----G-CA.....................G--ACTGAT............................................................-GC...A-TA----T.........AGG--  745
H1O.BE.87.ANT70              ..-A--ACACA.....................-GC...............................................................GCA-GG...----CGG-T.........----- 1227
H1O.CM.91.MVP5180            ..-A--AAACA.....................-GC...............................................................CCT-GG...---ACA--T.........----- 1202
CPZ.CM.04.EK505              ........--ACCTGAACCAGAACAAGCAGCAGGG-C-GCGGCAGCA...................................................CCTG-AAGCAGTA----T.........AGG--  729
CPZ.CM.04.LB7                ...........GAA......AAGGAAGCAGCTG---CCGCGGCAAGG...................................................GA----GACAGT-GT---.........AG---  727
CPZ.CM.04.MB66               ..CAG-AA...............GTGGCTGCAG-TACCGCCGACGGT......................................................-C-AGCACT-----T.........AGG--  724
CPZ.MT145                    ...........GAA.......................................................................................-CACCATCA-CT---.........GCG--  720
CPZ.CD.-.ANT                 GCAG--GCATGCTTCTGCGGCTGCTCCTGTTA-ACAA-CAG......TGGTGTCAGCGA.......................................CATCT-AGTGGC-AAG-G.........AG---  593
CPZ.CM.-.CAM3                ..C-GCAGC-G.....................CAGCAGCAA.........................................................GA--GC...GCT-----T.........AG---  571
CPZ.CM.98.CAM5               ..C-GCAGC-A.....................CAGCAGCAG...............GAA.......................................----GC......A-A--T.........AG---  874
CPZ.GAB_X52154               ACT---GAAG-CGTGAAGGGGGAGCCAGTCAAGGC-CTAGTGCC............TCT.......................................GCTGGC......A-T--T.........GG--- 1242
CPZ.GAB2                     ..-----AA-G.....................CA-A-TGAG......................................................TCA-A--CCACTG--A-A---.........AG---  559
CPZ.TZ.-.TAN1                GT-G-CAAAGA............CAGAATGCAGGT-A-AAAGAGGAAACAGTGCCACCTAGTGGCAATACAGGAAACACAGGGAGAGCAACAGAGACACCT-G-...GG-...--T.........AG-CT  813
CPZ.US.85.CPZUS              ..-A-AAA--ACAACAGCAACAG.........GAG-C-AGT...............GGC.......................................--T--C...ATT-GT---.........AGC-- 1229
GSN.99CM166_AF468659         ................................................................................CCCCCGGGGGCCGCAGGGG-AC-GGCGGTA--T--T.........-----  858
GSN.99CM71_AF468658          ................................................................................CCCCCAGGGGCCGCAGGGG-AC-GGCGGTA--T--T.........--G--  864
H2A.ALI_AF082339             ..AT--AAACA.....................G---AGAAAATGCCAAGCACAAGTAGACCAACAGCA...........................-CACCT-G-...G-ACAGG-G.........GG--- 1496
H2A.BEN_M30502               ..A-T-GAACT.....................G---AGAAAATGCCAAATACAAGTAGACCAACAGCA...........................-CACCT-G-...GG-AAA--A.........GG--- 1503
H2A.ST_M31113                ..A--AAAACT.....................----A-AAAATGCCAAGTACAAGTAGACCAACAGCA...........................-CACCT-GC...GG-AA-G-A.........GG---  948
H2B.D205_X61240              ..A---AAA-A.....................-TGCC-GCTACAAATAAACCAACAGCA....................................-CACCT-GC.........G--.........GG--- 1464
H2B.EHO_U27200               ..A---AAA-A.....................-TGCC-GCTATGAGTAAACCAAGTAAA.......................................CCA-C-...AGCCG-............-T-GC 1465
H2G.ABT96_AF208027           ..A-T-GAACT.....................----A-AAAGTGCCTGCCACAAGCAGACCAATAGCA..............................CCACC-AGTGGCAGAG-A.........GG---  866
H2U.FR.96.12034              ................................--T-A-ACTGCAGAAAAAATGCCAGCCACAAGTAGACCAACAGYT..................-CACCT-G-GGCGGCAGA............GG---  982

MAC.239_M33262               ..A---GAACA.....................----A-ACTATGCCAAAAACAAGTAGACCAACAGCA...........................-CATCT-GC...GGCAGAG-A.........GG--- 1453
Gag                          .__T__G__T__.__.__.__.__.__.__.__T__E__T__M__P__K__T__S__R__P__T__A__.__.__.__.__.__.__.__.__.__P__S__S__.__G__R__G__.__.__.__G__N  p15

SMM.SL.92.SL92B              ..AGT-GAACT.....................G---A-AAATTGCCAGCTCAAAGCAGACCAACAGCT...........................-CACCT-G-.........G-A.........GG---  887
SMM.US.x.PGM53               ..A---GAACA.....................G---ACAAAATGCCAATAACAAGCAGACCAACAGCA...........................-CACCT-G-...GGCAGAG-A.........GG--- 1379
MNE.US.-.MNE027              ..A---GAACA.....................G---A-ACTATGCCAAAAACAAGTAGACCAACAGCA...........................-CATCT-G-...GGCAAA--A.........GG---  933
RCM.NG.x.NG411               ..-G---CACA.....................GTGAC-.........................................................TCT--TGGA...--AAGA............GG---  612
RCM.GAB1_AF382829            ..-G---AACA.............................................................................................GCG-CACCT--TGGACAAAGAGG---  612
DRL.FAO_AY159321             CA--GA-ATCTAGCGGCAGG.............................................................................................TCA.........AG---  694
MND-1.F17                    ...................................................................................AGAAAGGAAGAA...-AGC-GGAGG-T-AA--T.........A-G--  808
MND-2.5440_AY159322          ..A----GACA.................................CCTGCTGTTCGC.........................................................TCA.........A----  697
MND-2.14CG_AF328295          ..A----GACA.................................CCTAGTGGCCGC.........................................................TCA.........A---- 1064
MND-2.MND2_AF367411          ..AT--TGACA.................................CCTAGTGGCCGC.........................................................TCA.........A---- 1140
DEB.CM.CM40                  AG-AG-AA--ATTAGAAATGGCCCTAATCCAA-G-CAGAAAGAGAAGGAAGTAGAACAGAAACAACAAAAGCAACAGGAACAAAGACAGCAGGAG-CTCAGGCAGCA-CCCAG-CAACACCATCAGTT--  954
DEB.CM.CM5                   ..CAG-AG--A.........GAGTTAGAGATGG--TT-ATTCAGAAGAAAGCAGAAAAGGAGTTAGAGAAAAAGCAGAAAACTCAGGCAGCAGCAGCAGA--CACAG--AAAA-CAAACCAACCAGCT--  945
MON.CM.L1                    ................................GA-CAGGGAGAACAGAAAGCTGCAGCCGCGGCCGCT...........................-CACCA-CA...GGC...G--GTACCCAGTGG---  845
MON.NG1                      ..................................................................................................................................    0
TAL.CM.8023                  ..CAG-GA-CA.................................................................................GTCAGA--TGCAATAGT-ACATCTAGTGGCCATAGTGG  564
TAL.CM.266                   ............................................................................................GTAAGA-AGGCAACAGT-ACATCTAGTGGCCAAAC-GG 1066
MUS.CM.1085                  ...............................................AAAAAGGGAGAA....................................A-AGAATCA...--A-CGCAGGCAGCTAGTGCC--  829
SAB.SN.-.SAB1C               ..CA-A--A-G.....................GA-CTGCAGGCAAAAAAGAAAAATGAGCCAACAGTG...........................ACACCT-G-...GGT-G-TCG.........-G--- 1450
GRV.ET.x.GRI_677             ...................................-AGAAAGCAGCTAAAAAGAAAAATGAGACAACAGCG........................-CACCT......GGT-G-GAATCA......AG--- 1297
TAN.UG.x.TAN1                ..-GCCAGC-G.....................CA-AATTACAACACAGCTGCG..........................................ACACCT-G-...GGCAGACAT.........GGT-- 1320
VER.DE.-.AGM3                ..-AGA-ATCT.....................-GT---CAAAAGAAAAATGACAAGGGAGTAACAGTG..............................CCACC-...GGT-G---T.........-----  861
VER.KE.-.9063                ..-G--AAA-G.....................-A-A-TAACAGGGAGACAACAGCG.......................................-CACCTGG-...GGC...--T.........----- 1360
VER.KE.-.AGM155              ..-G-CAGC-G.....................-A-AATAACACAGGAGGAACAGCG.......................................ACACCTGG-...GGC...---.........----- 1349
VER.KE.-.TYO1                ..-G-CAAA-G.....................-A-AATGACAAGGGAATAGCAGCG.......................................-CACCTGGG......TG---T.........--G--  850
SUN.GA.98.L14                .....................................................................................................GCA...--A-CTTCT.........----- 1418
LST.CD.88.447                ...........................................................................................................-CTAAA--T.........GGC--  356
LST.CD.88.485                ...........................................................................................................---AAA--T.........GGC--  356
LST.CD.88.524                ...........................................................................................................-CA-AA--T.........A----  350
LST.KE.-.lho7                ...............................................................................................A-G-AAC-A...---ACAG-G.........GG--- 1440
SYK.KE.-.SYK173              ..-A-AA-AG-.....................-GTCA-AACATGGCTAGTGAGACATCTAGTGGCCAAAAGGTAGTACAGCAGGAAAAACAAAAGGCAGCA-CA...-C-CCACCTAGA......GG--- 1261
SYK.KE.KE51_AY523867         ..-A-AA-AG-.......................................AACGACACAGCGACATCTAGTGGCAGAGAAGGAAAAATGCAGCTG-CAGCAGCAATG-CATCT--TGGCGGCTCAGG---  889
COL.CM.-.CGU1                ...................................AA-AAGCAGGCCATGATAGAAATGGCC.....................AGCAAGGAGGAGG-G-AAGCAAAAA---AAGCAGAG......A--TT  992
```

```
                            Gag p17 \   / Gag p24
H1B.FR.83.HXB2              TTAC...CCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCA  1308
Gag                         __Y__.__P__I__V__Q__N__I__Q__G__Q__M__V__H__Q__A__I__S__P__R__T__L__N__A__W__V__K__V__V__E__E__K__A__F__S__P__E__V__I__P__M__F__S_   p24

H1D.CD.84.84ZR085           ---T...--------A------C-A--A--------------C------C----------------G--C---------------A-------A----------A---------A---------------   825
H1F1.BE.93.VI850            ----...--------A-----TC-T-----A-----------C---T-TC-------------------------------G--GA----------------T-----------A---------------   651
H1G.SE.93.SE6165            ---T...--------------TGCA--A--------------C---------A-------------G--------------------------A-----C-----T--------A---------------   711
H1H.CF.90.056               ---T...--------A-----TGCT--A-----G--------C-----------------G--C-----------------------------A--------T-----------A---------------   655
H1J.SE.93.SE7887            ---T...--------------TC-G--A------CC------C------C----------------------------------GA-------A--A---------------------------------   631
H1K.CM.96.MP535             ----...--------A-----TC-G-----------------C------C----------------------------G--G---A----G-----------------------A------------A--   507
H101_AE.TH.90.CM240         ----...-----------A--TGCA--A----------C-------C-TT----------------G-----------G--------------A----G---T-A---------A-----------C---   873
H102_AG.NG.-.IBNG           ----...-----------A--TGC-A-A---------AC-------T----G--------------------------G------A-------A----G---------------A---------------   839
H103_AB.RU.97.KAL153_2      ----...--C--------A--TGCA--A---------AC---C---T----G--------G-----G-----------G------A-------A--------------------A---------------   540
H104_cpx.CY.94.CY032        ----...-----------A--TGCA--A------------------AG---T--------------G------------------A-------A--------------------A-----------C---   674
H1N.CM.95.YBF30             ---T...---C----C----CTGCT--A--A---------------C-GC-GA----C-----C--------T-----G-----GA----G--------C--T--T--------A-----A------ATG   872
H1O.BE.87.ANT70             ----...--------ATCA--TGCG-----A--------------------C--C--C--G--------------------G-C---------A--A--C--T-A---T---A-C--T--T-----CATG  1354
H1O.CM.91.MVP5180           ----...--------AACA--TGCA-----A--------------A-----C--C--C--G--------------------G-C---------------C--T-A---T---A-T--T--T------ATG  1329
CPZ.CM.04.EK505             C--T...---C-G--A--A--TGCT--A------------------C-GC-G-----C-----C--------T-----------G--------A-----C--T-AT--------A-----A------ATG   856
CPZ.CM.04.LB7               C---...---------ACA--TGCA--A-----GCC------C-----C-C---------------------C-----C--G-----------A-----A--T-----------A-----------CG--   854
CPZ.CM.04.MB66              C-TT...--------AGCA--TGCT--------G-----------AC----C--G--------------------------------------A-----A--T-----------A--T--------CATG   851
CPZ.MT145                   C-T-...---G----A-GA---GCT-----A--GT--ACG---A-TC-A--------C-----CC-------C----------C---------A--A-------AT-----G--C-----A-----CATG   847
CPZ.CD.-.ANT                ----...--C--CA-AGT-G-TGCAGGA--AAT-GCAAGG------C-AC-GA----A-----C--------C--------GTGT-----------AAA-----AT--------C--C--T--------T   720
CPZ.CM.-.CAM3               C---...--AG----A-----TGCA--A--A--GC-A--G------C----G-----C--G--C--------T-----T------A-------A---AA---T-A------G--------A-----CATG   698
CPZ.CM.98.CAM5              C---...--A-----A-----TGCA--A--A------A-------------G--------G--CC-------T-----C--G---A-------A---AAC--T-A---------A-----A------ATG  1001
CPZ.GAB_X52154              ----...--CC-C--A--A--TGCT--A-----G-----G-----------C-----C-----C--G-----------------G-----G--A-----A--T-----T--G--A-----A---------  1369
CPZ.GAB2                    ----...--AG-G--------TGCA-----A--G--A--------------G-----C----------G---C-----T----C---------A--A--C--T-----------C--T--A-----CATG   686
CPZ.TZ.-.TAN1               A---...--AG-GA-AACTG-TGCA-----AGTTGCAAGG------C-T--T------------C-------C-------GG---A-------A--A-GG--T-AT--------A-----A-----C---   940
CPZ.US.85.CPZUS             ----...--AG----A-A----GCT-----A--G--------------A--G--T--C--------------C----------CG--------A-----C--T-A---G-----C--C--G-----CATG  1356
GSN.99CM166_AF468659        C---...---G-GA-TAG---TGC---A--A--GTATCA----------C-CAATAG-----T-C-T--GA-T-------GCA----G--G-----AAAG--TGCT-----GACAG-TG--T-----CAG   985
GSN.99CM71_AF468658         C---...--C-----TAGA--TGCT-----A--GT-TCA---C------C-CAACAGC---GTCC-T--GA-T-------GCA----G-----A---AAG--TGCT-----GACAG-CG-TC-----CAG   991
H2A.ALI_AF082339            C-T-...--CG--CAA--AGTAGC-...--CA-CTACACC---GT-C-GC-GAGC--CC----C---G----T---------T----------A---AAG---G-GG-------AG-G--AGGA---CAG  1620
H2A.BEN_M30502              C---...--CG-GCAA--AGCGGGT...--CA-CTAT--C---GT-C-AC-GAGC--CC-----C-----------------T----G--G--A---AAG---G-GG-------AG-G--AGGA---CAG  1627
H2A.ST_M31113               C-T-...--CG--CAA--AGTGGC-...--CA-CTATACC---GT-C-AC--AGT--CC----CC-------T---------C-------G--A---AAG---G-GG-------AG-G--AGGA---CAG  1072
H2B.D205_X61240             ---T...--AG-GCA---ACTGGCT...--CA-CTAC--C--C-T-C-GC--AGC--CC----C--------T--------GT----------A---AAG---G-GG-------AG----AGGA---CAG  1588
H2B.EHO_U27200              ---T...--AG-GCA---A-TAGCT...--CA-TTATTCC--C-T-C-GC--AGC--CC----C--------C---------T--------------AAG---G-GG-------AG----AGGA---CAG  1589
H2G.ABT96_AF208027          C---...--AG-GCA---AGTAGG-...--CA-CTAT--T----T-C-AT--AGT---C-----------------------T-G--------A---AAG---G-GG--------G-G--AGGA---CAG   990
H2U.FR.96.12034             ----...--GG-GCA-----TAGGT...--CA-TTAT-CC--C-TAC-TC--AGT--------AC-C-----C---------C-G-----G-----AAAA--TG-AG-------AG-G--AGGA---CAG  1106

                            Gag p15 \   / Gag p27
MAC.239_M33262              ----...--AG--CAA--A-TAGGT...--TA-CTAT--C--C-T-C-AT--AGC--G-----A--------C---------T-GA----G--A---AAA--TG-AG-------AG-G--AGGA---CAG  1577
Gag                         __Y__.__P__V__Q__Q__I__G__.__G__N__Y__V__H__L__P__L__S__P__R__T__L__N__A__W__V__K__L__I__E__E__K__K__F__G__A__E__V__V__P__G__F__Q   p27

SMM.SL.92.SL92B             C--T...--AG-GCA---AGTAGGA...AATA-TTAT--G--CACTC-AC-G--C--AC-G--AC-C-----G-----C---C-G--------A---AAA--TG-AG-------AG-G--AGGA---CAG  1011
SMM.US.x.PGM53              ----...--AG-GCA---A-TAGGT...--CA-TTACACC--C-TAC-TT--AGT--A-----A--------T---------T-G-----------AAAA--TG-GG----G--AG-G--AGGA---CA-  1503
MNE.US.-.MNE027             ----...--AG--CAA--A-TAGGT...--TA-CTATACC--C-TAC-AT--AGC--G-----A--------C---------T-GA----G------AAA--TG-AG-------AG-G--AGGA---CAG  1057
RCM.NG.x.NG411              ----...--A-----AACA-TT-AT...CA---GCCT-A-------C-A-----T--------A--------T-----------------G--A---AAA--TG-AG-------AG-G--T--------G   736
RCM.GAB1_AF382829           ---T...------A-TACT-TA-AT...CA---GCCT-AG---A-TC-T--------A--G---C-------C-----C--G--G-----G-----AAAA---TCAG-------AGCG--------C--G   736
DRL.FAO_AY159321            C---...--C--TCA-GTTGTT-AT...CAA--GGCA-----C-----T--T-----A-----C--------C-----G--G---A-T--G--A---AAG-----TG--------G-G--T-----C-TG   818
MND-1.F17                   --T-...---G--CAAAG-G-TGCAGCA--A--GTATCAGT--ACTC-A---AGT-----G-T-A--C-GA-----------AC---G-----A---AAG-GG-AA--G--G--C--C--TC-A--C---   935
MND-2.5440_AY159322         ----...--C--TCA-GTA-TA-AT...CAAAC-CCA----------GA--T-----A-----C--G---------------TGTA----G-----AAAG------------A-AG-G--T------ATT   821
MND-2.14CG_AF328295         ---T...-----TCA-ATA-TA-AT...CAAAC-CCA--G-----A-GA--T-----A--------------C-----C---TGTA-T--G--A---AAA-----T------A-AG----T-----CAT-  1188
MND-2.MND2_AF367411         ----...--A---CA-ATA-TA-AT...CA-ACCCCA--C-------GA--T-----GC-C--GC-G-----G---------TGCA-----------AAG--T---------A-AG-G--------CAT-  1264
DEB.CM.CM40                 C---...-----CT-AAGAC-GGGG...CAA--GT-T--G---AT-C-AT--AGC--C--G-TAG----AA-C---A----T-CT-----G--A---AAA--TTCT-----GA-AG-T--TC-T---CA-  1078
DEB.CM.CM5                  --TT...--A--CC--AGAC-AGGA...CA----TAT------AT-C-TT--AGT-----G-T-G-G--AA-C---A----T-C---G--G--A---AAA--TTC------GA-AG-G---C-A---CA-  1069
MON.CM.L1                   C---...---G-C---AGA-CACAGGG---AGGCT-CCA---------AG-GGAG---C-GCTCC-C--GA--------C-G--CA-T--G------AAA--TGC----------G-GG--C-C--CCAG   972
MON.NG1                     ..................................................................................................................................     0
TAL.CM.8023                 GA-TTAC--C--TA-AAGA--TGCA--ACAA--GTATCAG---------T-G-----C---CTGC-G--AA-----A--GCCACTA-------A---AAA--TGC------GACAG-GG-GC-T--C--T   694
TAL.CM.266                  AA-TTAC--C---A-AAGA--TGCA---CA---GTATCA---------AC-G--------GCTGC----GA-----A-TGCCACTA----G-----AAAA--TGC---T---ACAG-GG-GC-C--C--G  1196
MUS.CM.1085                 C---...--AG-CA-AAG---GCA--A------GT-CCA----------C-CAATG------TA-----AA-------GTC-A----G--G--A---AAA--TGCTG-------TG-GG-A------CAG   956
SAB.SN.-.SAB1C              ----...--C--T--CAGTGTT-AT...AAT---TG---C--C---C-AT-G--C--A--G--AC-T-----T-----------GA-T--G--A---AAG-----TG----G--AG-----------AGC  1574
GRV.ET.x.GRI_677            ----...--AG----AA-TC-G-AT...AATGCCTG------C---C-TT-G--T--GC-C--G--------G-----C---TGC--G--G--A--AAGG-GGG-AG-------AG-C--------CCA-  1421
TAN.UG.x.TAN1               ----...---G----A--AC-G-AT...AAT---TG-------ACTC-TT-G-----AC-T--CC-------------T---AC---T--G--A--AAGG---G-GG-----A-AG----------CCA-  1444
VER.DE.-.AGM3               --T-...--AGC-CAA--AC-GGGA...AATGC-TG-A-----GT-C--T-G-----AC-C--C--------G----------C------G-----AAAA---G-AG-----A-AG-G--------CCAG   985
VER.KE.-.9063               --T-...--AGCGCAA--AC-AGGA...AATGCCTG-------GTTC-AC----C---C-C--A--------G----------C------G--A---AAA--TG-AG----GA-AG-G--A-----CCA-  1484
VER.KE.-.AGM155             --TT...--CGC-CAA---C-AGGG...AATGC-TG---G---GTAC-AC-T------C-C--CC------------------C------------AAAA--TG-GG-----A-AG----------CCA-  1473
VER.KE.-.TYO1               --TT...--AGCGCAA--AC-AGGA...AATGCCTG-------GTAC--T-G-----GC-C--C--------G----------C------G-----AAAA--TG-AG-----A-AG-----------CA-   974
SUN.GA.98.L14               --TT...--AG-GCA-AGAG-GGGA...CAAA-TTACA-T------C-TC-C--G--------AG-GC-AA-------G--GA----G----------GA-GG-AA------ACAG-GG-A------G--  1542
LST.CD.88.447               ---T...--AC-C--CAGAG-A-AT...CA-AG-TG------CACAC-AT-G-----------AA--C-AA-T---------A----------T-GA-GA-GG-AG------ACAG-GG-A--------T   480
LST.CD.88.485               ---T...--AC-CA-CAGAG-A-AT...CA-AG-TG------CAC-C-AT-G-----------AA--C-AA-T---------A----------T-GA-GA-GG-AA------ACAG-GG-A--------T   480
LST.CD.88.524               --TT...---T-GA-TAGAG-A-AT...CAAAG-TG---C---ACCC-TC-------C-----AA-TC-AA-------G---A----------T-G--GA-GG-AG------ACAG-GG-T---------   474
LST.KE.-.lho7               ---T...--CC-GA-AAGAG-A-AT...CA-AG-TG-------ACAC-TT----G--C-----AA--C-AA-------G---A----G-----C-GA-G--GG-AA------ACAG--G-A---------  1564
SYK.KE.-.SYK173             C--T...--AC-GC-CAGA--TCCT--AAAT--GTG-A-T---ACA-GAG-CC--GT---G---C----AA-------GG-G-CT--GA-CTCC--AAAA--TGATG--TCCA-AG-G--AT-A---CAG  1388
SYK.KE.KE51_AY523867        ----...---C--A-TAGA---CCA--AAAT---TG-A-T---GTA-GGG-CAACA-C--G--------AA-C------G-G-C---GA-CAGC--AAAA--TGATG-CTC-C--G-C--AC-A--CCA-  1016
COL.CM.-.CGU1               GG--ATG--A-----AACAGGGCCT...CAAGGGCCA--G-----AC--C-GAGC--C--G--G---GGA--C-----T--GTGT-----G...GGA-GAA-TGCT--TTCCT--GCT--------C-T-  1116
```

```
H1B.FR.83.HXB2          GCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTGCATC  1438
Gag                     _A__L__S__E__G__A__T__P__Q__D__L__N__T__M__L__N__T__V__G__G__H__Q__A__A__M__Q__M__L__K__E__T__I__N__E__E__A__A__E__W__D__R__V__H__   p24

H1D.CD.84.84ZR085       -----------------------------------------------------------------------------------------------------------------------C--GC-A----   955
H1F1.BE.93.VI850        --------------G-----T--CAC---------------------T-----------------------------------------C-----------------------------C---T-A----   781
H1G.SE.93.SE6165        -----------------------------------TCT----T----------------G-----------T---------C----G--T--T--------------------G-----C--GA------   841
H1H.CF.90.056           -----------------------------C-----TG-T--------T------------------------------G-----------T--A--------------------------C--G--A----   785
H1J.SE.93.SE7887        -----------------------G-----------T---------------A-A--------C--------T-----------------T--T--------------------------C--G--A----   761
H1K.CM.96.MP535         --------------------T---------C-----------------------------------------------------------T--T--------T-----------------C--GT-A--C-   637
H101_AE.TH.90.CM240     -----------G-----------------------T-TG--------T-T------------C--G-----A-----------------A---------------C----------------G--A--C-  1003
H102_AG.NG.-.IBNG       -----------G-----------G-----------T-TG----------T------------C--G-----A-----G-----------T-------------------------------G--A----   969
H103_AB.RU.97.KAL153_2  -----G-----G-------------------------TG-----G----T------------C--G-----T-----------G-----T--------------------T-----------GT-A----   670
H104_cpx.CY.94.CY032    -----------G--------------------G----TG--------T-T------------C--G-----A-----------------T-----------------------C-----C--GACA----   804
H1N.CM.95.YBF30         --C--G--------G--A--G--CTC----C----T--T---T----T------A-----------G-----A-----G---C-G--G--AGT---------------A-----C--------GACA----  1002
H1O.BE.87.ANT70         -----G-----G-----A-TTT-CT-T---A-T--T--T--------TG-CA-A--A----------G---TT-A---G--C----G--AGTA--------------A-T---G---------ACT--C-  1484
H1O.CM.91.MVP5180       --------------G--TGT---CT-T---A-C--T--------G--TG-CA-A--------C----GG--TT-A---G----G--G--AGTA--------------A---------------ACT----  1459
CPZ.CM.04.EK505         --C--G--------G--A--G--C-----C------------T----T------A-----------G-----A-----G---C-G--G--AGTT--------------A-----G-----C--G-GA----   986
CPZ.CM.04.LB7           --C---G----------------T-----------T-----------T------A-------C--G-----------G--C---------GT---T-----------ACA-------------GC------   984
CPZ.CM.04.MB66          --TC-G-----------------T------C-------------------------------C--G-----A---------C-------AGTA--------------------G-----C--GC-C----   981
CPZ.MT145               --------------G--A--G--T-----CC-C--T--T--------T---A-A--A---------G--------------TC-------AGTA--------T--G-----T--G-----C--GACA--C-   977
CPZ.CD.-.ANT            --T-----------G--A--T--T--T-----------------T--TG----T----AC-------G---------GG--C-------AGTA--------------A--T--G--------GT-A--C-   850
CPZ.CM.-.CAM3           --CC-G-----------A--T---------G-T--T--T-----T--TG-TA-A--A--G--C--G-G---T-----GG-C--G--G--AGT---------------A---------------ACA--C-   828
CPZ.CM.98.CAM5          --CC-G--G--G-----T--T--C--G---G-C--T--T-----T--TG-C--A--A--G--C--G-G---T-----GG-T-----G--AGT---------A-----A-----G--------GC-A--C-  1131
CPZ.GAB_X52154          --TC-------G------TTA--T--G---G-T---------T----TG----A-------------G----------G--C----G--AGT---T-----------C--C--G------C-CT-A----  1499
CPZ.GAB2                --T-----------G--A--A--C------G-G-----A-----G--TG-TA-A--A-----C----G---T-----GG-AC--------GTA---------A-----A--G--G----------T-A--C-   816
CPZ.TZ.-.TAN1           -----G--T--G-----A-----TT-T---C----T-GT-----C--TG-T--T----A------------A-----------G--G--AGT---------------A-----G-----C----CA----  1070
CPZ.US.85.CPZUS         --C-----------G--G--T--C------G-C-----T-----C--TG-TA-A-----------G-G---A------G-T-----G--AG----------------A-----T--A-------ACA----  1486
GSN.99CM166_AF468659    --T-----------CTGT-TA--CT-T---A-G---CAA-----C--TG-TA-A--A-ATT------GG--AG-A--G--TA-T-----TGT---------AC-G--A-----G-----CCTTT-A----  1115
GSN.99CM71_AF468658     --TC--A-------CTGT-TA--TT-T---A-G---CAA-----C---G-T--T--A-ACT------G---AG-A--G--TA-T-----TGT---------AC-G--A-----G------CTCT-A----  1121
H2A.ALI_AF082339        ---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AC--------------------AA-C-GG----TT---------A-----A-----C------GTT-CA----  1750
H2A.BEN_M30502          ---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AT--C---------T--------AA-C-G-----TT--T-----A-----A-----C------TCGCA---C-  1757
H2A.ST_M31113           ---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AC-------------T--------AA-C-GG--A-TT--T-----A-----A-----T-----CGC-CAA--C-  1202
H2B.D205_X61240         ---C-----------TG------TT-T---A----TCAG--------TTGT--A--A-A------G---------------TA-T-G---A-TA---------------A-----C-----CCAGCAA--C-  1718
H2B.EHO_U27200          ---C----------GTG------TT-T---A-T--TCAG---T----TTGT--A--A-A------G---------------TA-T-GG----TA--T------------A-----C-----CCA-CA-----  1719
H2G.ABT96_AF208027      --GC-G--------CTG-CTA---T-T---A-C--TCAA--------TTGT--A--A-A----------------------TA-T-G---A-TA--T--------G--A-----K------CTTCAA----  1120
H2U.FR.96.12034         ---C-C-----G--CTG---G--CT-T---A-T--TCAA---T----TTGC-----A-A---------T--T---------TA-C-GG--A-TA--T-----T-----A-----T-----C-C-CA---C-  1236

MAC.239_M33262          ---C-G--------TTG------CT-T--CA-T--TCAG---T----TTGT-----A-AC--------G--T-----G--TA-C-G---T-TT--A--C-----G--------T-----CTTGCA---C-  1707
Gag                     _A__L__S__E__G__C__T__P__Y__D__I__N__Q__M__L__N__C__V__G__D__H__Q__A__A__M__Q__I__I__R__D__I__I__N__E__E__A__A__D__W__D__L__Q__H_   p27

SMM.SL.92.SL92B         ---C----------CTGT--A--CT-T--CA-T--TCAG----T--TTGT--A--T-A------G----------------TA-C-G---A-TA--T-----------A-----C-----CCT-CAA--C-  1141
SMM.US.x.PGM53          --GC----------CTG---T--CT-T--CA-C--TCAA--------TTGT--A--A-A------G---------------TA-T-G---A-TT--A-----A--------T--T------TTGCAA--C-  1633
MNE.US.-.MNE027         ---C-G--------CTG--T---CT-T--CA-T--TCAG---T----TTGT-----A-A-----------T-----G--TA-C-G---A-TT--A--C-----G--------T-----CTGCA---C-  1187
RCM.NG.x.NG411          ---C-G--------CTG--TA--CT-T---G----TCAG---T----TG-CA-A----A--------GT--T-A------AG----G--AGTA--------T--G--A-----T-----CTTGAGA----   866
RCM.GAB1_AF382829       ---C----------CTG--TA--CT-T---A----TCAA--------TG-CA-A----A---C--G-GT--GC----G--AG----G--AGTG--------------A-----C------GCTAGA----   866
DRL.FAO_AY159321        --CC-G--------CTG--T-AGCT-T--CA-C---CAG---T----TG-TA-T----AT--C----GT---C-C-----AG-------C-T-----------------A--T--T---------GAGA----   948
MND-1.F17               -----GA-----------A-T-AGT--T------G--T-T------G--TG----A--A-AT-----G-G---A------G-C--------TGTA--T------C----A-------------CT-ACA----  1065
MND-2.5440_AY159322     --C--G----------TG--T---GT-T--CC-C--TGG------C--TG-CA-T--A-AT-----G-G---TC-C-----AG-G--G--TGT---------------------C------CTTAGA--C-   951
MND-2.14CG_AF328295     --CC------------TGCT-----T-T--CC-C--TGGT---T----TG-TA-T----A------G-G---TC-C-----AG-G--G--TGT---------------------C------CTTAGA--C-  1318
MND-2.MND2_AF367411     --C--G--T-----GTGCT----CT-T---C-C--TGG------C--TG-TA-T--A-AC-----G-G---GC-T-----AG-G--G--TGT---------------------C---A-CTTGAGA--C-  1394
DEB.CM.CM40             -TC--GG----G--CTGT-----CT-T--CA-T--TGGTC-A--T--TG--A-T--A-AT-T---G-GG--T-----G--CA-T-----TGTA--T-----------A-----G------TT-CAA--C-  1208
DEB.CM.CM5              -T---GG-------GTGT--T--CT-T---A-T--TGGGC-------TG-CA-A--A-AT-TA----GG--A---------TA-------TGTG--T-----------A-----G-----CTT-CAA--C-  1199
MON.CM.L1               --TC----------CATG-TT--CT-T--CA-C--TCAGC-A------G-CA-A--T-AT-----G-GG--AC-C-----TA-T--G--TGT----------------G--CA-T-----CTT------C-  1102
MON.NG1                 ..................................................................................................................................     0
TAL.CM.8023             ---C--G---------TG--TT--TT-T---A-C--TCAG----T--TG----A--A-A------G-GG-------G--CA-T-----T-T---A------C-G-----C---------CAGA----C-   824
TAL.CM.266              ---C--G-G-----GTG--T---CT-T--CA-C--TCAA----T--TG----A--A-AG-----G-GG--A-----G--CA----G--C-T---T------C-G--AT----G------CAGA------  1326
MUS.CM.1085             -----GG----G--G--A-TA---T-T--CA-C--TCAAC----G--TG-CA-A--A-A------G-G---T--A-----TA-------TGTA--T-----AC----A-----------CCTGT--A-C-  1086
SAB.SN.-.SAB1C          --CC--G----------T-T---TT-T---A-C--TCAA-----C--TG----A--A-A---C--G-GG--AC-A-----AG-G-----TGTT---------A-----G-----T-----CCTTAGA--C-  1704
GRV.ET.x.GRI_677        ---C-C-----G---TGTCT-T-CT-T---G----TCAG-----C--TGT-A-A--A-AC-----G-GG--AT-A-----TC-T--G--AGT---T-----A-----A-----G-----C--GACA--CA  1551
TAN.UG.x.TAN1           --CC-C--------GTG-TTAAGTT-T--CA-T--TCAG---T----TGTCA-A--A-AC-------GG--A-----G--TA-C--G--AGTA--------T-----A--CC-G--------T-ACA--C-  1574
VER.DE.-.AGM3           --T-----------GTG---A--CT-T--CA-C--TCAA-----T--TGTCC----A-AC-------GG--GC-A-----AG-------A-T-----------------A--CC-G--------T--CT--C-  1115
VER.KE.-.9063           ---C-C--------GTGT-----CT-T--CA-C--TCAA-----T--TGTGT-A--A-AC-------GG--AC-A-----AG-------A-T-----------------A--CC-G------GT-ACA--C-  1614
VER.KE.-.AGM155         --CC-C--------CTG-------T-T--CA-C--TCAG-----T--TGTCT-A--A-AT-----G-GG---T-------AG-G-----A-TA--T-------------A--CC-G------GT-ACC--C-  1603
VER.KE.-.TYO1           --CC----------CTG---A--CT-T--CA-T--TCAG-----T--TGTGC-A--A-AT-------GG--AT-A-----AG-G-------T---T-----A-----A--CC-G------GT-ACA--C-  1104
SUN.GA.98.L14           --T---ATGAC------A-TT---G----CA-C--TGTG-----G--TG-CA-A----AG-------G---T-----G---A-------TCAT--AGTG------AG-------------C---A----C-  1672
LST.CD.88.447           --CC-CA-----AA---TTTG---G-T--C-----TGT----T----TG-CA-A--A-AC-------GG--A-----G--AA-T-----TCAT--TGTG--A----GG-----------C--GAAA----   610
LST.CD.88.485           --CC-CA-----AA---TTTG---G-T--C-----TGT----T----TG-CA-A--A-AC-------TG--A-----G--AA-T-----TCAT--TGTG--A----GG-----------C--GACA----   610
LST.CD.88.524           --TC-GA-----AA----TTG---G----C-----TGT----T----TG-TA-A--A-AC-------GG--A-----G--AA--------TCAT--AGCG--A----GA--T-------------CA-----   604
LST.KE.-.lho7           --C---A-----AA---TTTG--TG-T--------TGTT---T----TG--A-T--A-AT-----G-GC-----------AA-------TCA---AGTG-------G------G---------CA---C-  1694
SYK.KE.-.SYK173         --TC--A---------TTT-TT--TT-T---C-C--TGG------T--TG-T--A--A-AT-----G-G---------GG-AA-T-----C-T--------------GG--G-C--------CTT-AA----  1518
SYK.KE.KE51_AY523867    ATT---A-------TTTT-TA--TT-T--CC-C--TGA------T--TG-CA-T--A-AT-------G---A------G--A-C-----TGTG--T----------GGT-T--------CCT-CAA--C-  1146
COL.CM.-.CGU1           --T-AT--TACG--G--A-TAG-TT-T---A-G---TTG-----C--T-TTT---ATACT-------G-TT-CCT--GG-C--G--G--TGAG--T---A-AA-G--A-A---G-AT--CCT-C-A----  1246
```

```
H1B.FR.83.HXB2           CA...GTG...CATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAAT...............AA  1547
Gag                      P__.__V__.__H__A__G__P__I__A__P__G__Q__M__R__E__P__R__G__S__D__I__A__G__T__T__S__T__L__Q__E__Q__I__G__W__M__T__N__.__.__.__.__.__N   p24

H1D.CD.84.84ZR085        --...---...--G---------G-------------------------------------T--T----------------------------------------------GC...............--  1064
H1F1.BE.93.VI850         --...---...------------GCCC----------------G-----T-----G-----T-----T------------------------------CA---------GGGC...............--   890
H1G.SE.93.SE6165         --...CAA...--G---------T--C-------------A--------------------T------------------T----G------------ACT--------CGGC...............--   950
H1H.CF.90.056            --...---...---------------C----------A--------------------C--T-----------------------G-------------C----------GGC...............--   894
H1J.SE.93.SE7887         --...--A...---------------------------G----------------------T-----------------------C------------------------GGC...............--   870
H1K.CM.96.MP535          --...---...--------------CC----------A-----------G-----G--------------------C--C-------------------C-----------GC...............--   746
H101_AE.TH.90.CM240      --...--A...---------------C----------------G--------------------------------------------A-----------------------C...............--  1112
H102_AG.NG.-.IBNG        --...--A...---------------C----------------G--------------------------------------------A----------------------GC...............--  1078
H103_AB.RU.97.KAL153_2   --...-CA...--G---------T--C-------G-------------------------------------G--G------------A----------------------GC...............--   779
H104_cpx.CY.94.CY032     --...--A...---------------C----------------------------------T--------------------------A----------------------GC...............--   913
H1N.CM.95.YBF30          --...--C...-C--TG--A--AC-AC-C-----G--AC-------C--T--A--------T-----------A-----C-----GGCA-----GG-G-CT--------TGC-...............--  1111
H1O.BE.87.ANT70          --...CCA...-CG-T------GT-GC-------G-----A--G-------CA-----------T--T--G--A-----C----AG--A--G-----TCAC----CT--C-GGCCC............--  1596
H1O.CM.91.MVP5180        --...CCA...GCAATG-----GT-AC-------G-----A--G-------CA-----------T--T-----A-----C--A-AG--A--G-----TAT-----CT--T-GAGGG............GC  1571
CPZ.CM.04.EK505          --...--A...---ATG-----A--AC-------A--AG-A--G--G-----A--------T-----------A-----------GGC---G---G---C---------TGCC...............--  1095
CPZ.CM.04.LB7            -T...A-C...---ATG-----A--A-----------A-----------------------------------C--------TG-A--A--------------------T-GC...............--  1093
CPZ.CM.04.MB66           --...--T...--------A---G--C-------A--A-----------------------------------C--C--C---A-C--A--G--GG----G--------T-GC...............--  1090
CPZ.MT145                -T...ACC...--------------A--T-----A--A-----G--------A--C--------------C--------CT----AG----G---G----G---------GCA...............--  1086
CPZ.CD.-.ANT             -C...ACT...--------A--AG-ACAGG----A--AT-A--G-------CA--------T--------G--A--A--C--AG-G-----G--G--GCA-------T---CACCT......CAACAG--   968
CPZ.CM.-.CAM3            -T...---...--------T--AG-A--C-----A---C-A-----GG--C-A--------------T--G--------C--T--AGTA-----G--T-CT---------GCA...............--   937
CPZ.CM.98.CAM5           -T...---...-----------A--A--C-----A--AC-------GG----A--G-----------T-----------C--T--A--A------G-T-CT---------GCA...............--  1240
CPZ.GAB_X52154           -C...ACT...-----T--T--G--A--C-----A---C-A--G--G--------T-----T--------G--C--C--C--A--G-----------T--G----CA---GCA...............--  1608
CPZ.GAB2                 --...T-A...--------A--AG-G--------T--A--------G-----A--C--------T-----T------------T-G--A--G---G--------------TCC...............--   925
CPZ.TZ.-.TAN1            -C...-CT...--------A--CCAGCA-G----GAT-C-A-----G--C-CA--GGCA--T--T-----G--C--------G--A--A------G--CTG----------CCCCACAGGCA......C-  1188
CPZ.US.85.CPZUS          -T...C--...--------C-----A--------G--------------C--------------T--------C--C--C---T-G--A------G-G--C---------GCA...............--  1595
GSN.99CM166_AF468659     -G...--A...-CA----CA---CAAC-GGTG-CAGG-C-T-----C--T--C---GCA------------GT----TCA--T-CCA-T--G--G----AG-----C----GGCAG............--  1227
GSN.99CM71_AF468658      -G...CA-...-CC--G-CA--CCAAC--GT--CAGG-C-T-----C--T--C---GCA------------GTA--ATCA--T-C-A-T--G--G----AG-----T----GGCAG............--  1233
H2A.ALI_AF082339         -C...A-A...-CA...--C--CT-AC--G----G---C-T-----------A--GTC---------------A--A--C--AG-AG-A-----G--CCAG------TTC-GGCCA............CG  1859
H2A.BEN_M30502           --...A-A...-CA...--C--CT-AC--G---A---C-C------C-----A--GTC---------------A--A--C--AG-AG-T-----G--CCAG------TAT-GGCCA............C-  1866
H2A.ST_M31113            --...A-A...-CA...--C--CT-AC--G-G--G---C-C--G--G---------TC------------G--A--A--C--AG-AG-A--G--G--CCAG------TTT-GGCCA............C-  1311
H2B.D205_X61240          -G...TCA...-CA...--C--A--GC-GG----A--AC-T--G--C-----A--GTCA--T-----------C--C--C--AG-AG-------G---CAG------TAC-GGGCC............C-  1827
H2B.EHO_U27200           --...TC-...-CA...--C--A--GC--G----G--AC-C--G--------A--GTCA--------G-----C--C--C--AG-AG-A-----G---CAG------TAC-GACCC............C-  1828
H2G.ABT96_AF208027       --...CAA...--A...-----GCCGC--G----A--AM-C--G-----G-CA---TCA--T--------G--C--------AA-AG-A--------TCAG----CACAC-GGCAG............C-  1229
H2U.FR.96.12034          --...AAC...--A...--C--ACCAC--G----G--AC-T-----G-----A--TTC---T--T--------A-----C--AG-GG-A--G--G---CAG------TAC-GGCCG............C-  1345

MAC.239_M33262           --...CAA...-CA--T...--A...CA--A---A--AC-T--G--G--GTCA---TCA--T--T--------A------T-AG-AG-T--------CCAG------TAC-GACAA............C-  1813
Gag                      P__.__Q__.__P__A__.__P__.__Q__Q__G__Q__L__R__E__P__S__G__S__D__I__A__G__T__T__S__S__V__D__E__Q__I__Q__W__M__Y__R__Q__.__.__.__.__Q   p27

SMM.SL.92.SL92B          --AGA-GA...--ACAGCC-G-GCAAC-GG-T--AGG-C-A--G------TCA---TCA-----------G--A-----C----CAAGT-----G----A-------TAT-GGGCA............C-  1256
SMM.US.x.PGM53           -G...CA-...-CA...--T--A--AC--G----A--AC-C-----T-----A---TCA--T--T-----------------AG-GG-A--G-----TCAG------TAT-GGCAG............C-  1742
MNE.US.-.MNE027          --...CAA...--A--T...--A...CA--A---A---C-T--G--G--GTCA---TCA-----T--------A--------AG-AG-T--------CCAG------TAC-GACAA............C-  1293
RCM.NG.x.NG411           -C...AAT...-CA...--A--AC-AC--G----T--AT-A--G-------CA----C---T--T--------A--A--C-ATA--GCT----------CT----CC--T-GAGCC............--   975
RCM.GAB1_AF382829        --...--A...-CA...--C--G--AC--G----G--AC-T----------CA-----------------G--A-----CT-AA-AGCA-----G----CT----CC--C-GAGCA............--   975
DRL.FAO_AY159321         -T...CAA...GTA-G-CCT...C-CC---A---GGT-C-T---A-C--CTCA--GTCA-----T--------C--C--CT-TA-AG-A--G-------AG----CC--C-GAGCC............C-  1057
MND-1.F17                -T...CAA...--ACA-CCAG-ACAAC--GG---AGGAT-A--GACC--TTCA--CTC---T---------------TC---AG-GG-A-----T-G-C---------AT-TGCAA............C-  1177
MND-2.5440_AY159322      -C...CCT...GTG...--A-----GC-C-A---AGT-C-A---A-C----CA----A------T--------A--A--CT--A-AG-A--G--G----A-----CA--C-GGCAG............C-  1060
MND-2.14CG_AF328295      -C...CAA...GTA...--A---T-GC-T-A---GGT-C-----A-C----CA-----------T--------A--A--C-G-A-AG-A-----G----A-----CA--T-GACAG............C-  1427
MND-2.MND2_AF367411      --...CAA...GTG...--T-----GC-C-A---TGTCT-----A-C----C---C-----T--T--T-----A--A--CT--A-AG-A--G--G----A-----CT----GGGAG............C-  1503
DEB.CM.CM40              --...CAA...--ACAGCC-......C---AG--G---C-T--G------TCA---GCA--------------A-AC--C--AG-GG-A-----G----C---------C-GACCAGCAGGTCAAGGAC-  1326
DEB.CM.CM5               --...CA-...--ACAGCCA......C---A---T---C-T-----G---TCA---GC---------------A-AC--C--AG-GG-A--G--G----CT--------C-GGCCAGCAGGACAAGGAGC  1317
MON.CM.L1                --...CA-CCG-CACA-...--GCAAC-CAAT-CAGG-C--G----C----C---GGCC--T---------GTCT-GTCC--A-CC--AC-G--G----AG-----A--C-GGGCT............--  1214
MON.NG1                  ..................................................................................................................................     0
TAL.CM.8023              -C...CA-.........AT---T--AT-AC--G----A---C----G--G---GCA--C-----------------A--C--C----CAGCA------G-G-AG----CC--C-GGCAG............C-   933
TAL.CM.266               -C...CA-.........AT---T--AC-AC--G----A---T-A--G------GCA--G--------------G--A--A--C--T-CCGCT-----GG-G-A-----CC--C-GAGCA............C-  1435
MUS.CM.1085              --...CA-CCC--ACAG...--GCAGC-CAAT-CAGGCC-C--GA-C----CA---TCA------------GTAT-ATC-----CA--A-----G----A------C--T-GGGCA............--  1198
SAB.SN.-.SAB1C           -G...CC-...-CGCA-CA---CCCA--T-AG--AGT-C-A--G--C---CAA--GTCA--------G--C--------C--AA-A--A--G-------A-----CC--C-GAGCA............C-  1816
GRV.ET.x.GRI_677         G-...CCA...-CA--T--C--GT-AC--G----G---C-A-----C--G-CA--GTCA--T--------------C--CT-AA---------------AG----CCTTC---GCC............--  1663
TAN.UG.x.TAN1            -C...CCA...-CA-----A--GC-AC--G----G---C-A--G--C-----A--TTCA--------G-----A--C--CT-TG--GCT--G-------AG----CTTTT---GCT............--  1686
VER.DE.-.AGM3            --...CCA...-CA-----A--AT-AC--G----A--AC-C-----C--T--A--CTC---------------C--C--C--AG-G--A-----GC-G-A------ATAC-CAGCC............--  1227
VER.KE.-.9063            --...CCA...-CA-----A---T-AC-------A---C-T-----C--T-----TTCA--------------C--------AG-G--A------T---A------TTAC-CAGCC............--  1726
VER.KE.-.AGM155          --...CC-...-CG-----C--CT-GC--G-G--A---C-C--G--T--GG-----TCA--T--------G--C--------AG-G--A--G--GC---AG-----CTAT-C-GCT............--  1715
VER.KE.-.TYO1            --...C-A...-CC-----A--CC-AC--G----A---C-C--G--C--TC-C--CTCA--T--------G--C--C--CT-AG-A--A-----GT---A------CTAT-C-GCT............--  1216
SUN.GA.98.L14            -T...CA-...--GCAGCCTG-ACAGC--GG---AGGCC-A--GACT--T-AT--T-CA-----------GGTA--ATCA--AG-AG-A--G--G---CAG-----CTATGGCCTT......AATGGAGC  1790
LST.CD.88.447            -G...CAA...--ACA-CCTG-ACAAC--GG---AGG-C-T---ACC--T-AT-C--CA--TG-C------GTC--CTCA--AG-AG----G--GC---CT----CC--TGCAGAC..............   720
LST.CD.88.485            -G...CAA...--ACA-CCTG-ACAAC--GG---AGG-C-T---ACC--T-AT-C--CA--TG-T------GTC--CTCA--AG-AG----G--GC---CC----CC--TGCAGAC..............   720
LST.CD.88.524            -C...CA-...--GCA-CCAG-ACAAC--GG---AGG-T-A---ACT----AT-C--CA--TG-G-----G-TC--ATCA--AG-AG-A-----GCA--CC----CA---GCAGAC..............   714
LST.KE.-.lho7            -G...CAA...--ACAGCCAG--CAAC--GG---GGGTT-A---ACT--T-AT-C--CA--TG-G-----GGTC--CTCA--AG-AG-A------C---CC----CT---GCAGAC..............  1804
SYK.KE.-.SYK173          -C...CAA...-CGCA-CAA---GC-C---A--CAGGAT----G--T--CTCC-CCTCA--------------A--A--CT-AA--GCT--G-------A------T--T-GGCAG............--  1630
SYK.KE.KE51_AY523867     --...CAA...-C-CA-CA--AGCAGC-TGT--CAGG-C-A-----T--T--T-CCTCA--------------A--C------A-C--A----------AG-----A--T-GACAG............--  1258
COL.CM.-.CGU1            -G...--C...--GCAGCCA...CAACAG-A---AGCAT-A--GC-G--T-CA-C-TCA--T--CA----C-AC--A--CT-AG-AGCA------G-G-CT---GGAGAG....................  1347
```

```
H1B.FR.83.HXB2              TCCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGACTAT  1677
Gag                         __P__P__I__P__V__G__E__I__Y__K__R__W__I__I__L__G__L__N__K__I__V__R__M__Y__S__P__T__S__I__L__D__I__R__Q__G__P__K__E__P__F__R__D__Y_   p24

H1D.CD.84.84ZR085           ------------------------C------------------------------------------------------GT-------T-------C---------------------------------  1194
H1F1.BE.93.VI850            C------G-------G-----C--C--------------C-----A-----------------------------T---GT-------T---------A------G-----A------------------  1020
H1G.SE.93.SE6165            C--------------G--------------------------------G------------------------------GT-------T-A--------------G-----A------------------  1080
H1H.CF.90.056               ----G----------G-----C--C--------------------------------G-----------------T---GT-----------------A------G-----A------------------  1024
H1J.SE.93.SE7887            ---------------------G--------------------T------C-----------------------------GT---T---T----T-----------------A-----T------------  1000
H1K.CM.96.MP535             C------G-------G--G-----C-----------------------T-----C------------------------GT-------T----------------G-----A------------------   876
H101_AE.TH.90.CM240         ------C--------G-----C--C--------G---------------------------------------------GTT------T----------------G-----A--------C---------  1242
H102_AG.NG.-.IBNG           ---------------G--------C---------------G-------------------------G------------GT-------T----------------G-----A--------------T---  1208
H103_AB.RU.97.KAL153_2      C--------------G-----C--C--------------------A-----------G---------------------GTT------T-A--------------G-----A--------C---------   909
H104_cpx.CY.94.CY032        C------G-------G--------C------------------T----G-------------------C---------C-TT-----CT----------------------A--------C-----T---  1043
H1N.CM.95.YBF30             ---T---G-T-----------T-------G----------G-------G-----C-G---T--G-----------T---GT---------A--G--C-A------------A--------C---------  1241
H1O.BE.87.ANT70             C-A------------------C--C----G--A-------G-GT-A---C----C-----G----A------C-----AGTG-----CT-A--T--T-AG--G--------------A------------  1726
H1O.CM.91.MVP5180           -AACT----------------C--C----G--A-------G-G--A---C----C-----G----A------C--T--AGTG-----CT-A--T--T--G--G--------A-----A--C-----T---  1701
CPZ.CM.04.EK505             ---T---G-T-----------T-------G----------G-T--------------G--T--------------T---G----------A-----C-A---G--------A--GA-T--------T--C  1225
CPZ.CM.04.LB7               C--T-----A-----------------C--------------TA----GC-G-----GG----C--------C------GTT-----C----------AG-----G-----A-----A--C---------  1223
CPZ.CM.04.MB66              C--T-----T-----------T--C-----------------TT----GC-------G-------A-------T-----GT------A--A-------A------G-----A--GT-A--C-----T---  1220
CPZ.MT145                   C--T-----T-----------T-------G----------G-------G--------------T-A------C--T---GTT------T-------T-A------------A--G--T--C--G------  1216
CPZ.CD.-.ANT                -GG-GGAG----------G--C--C-----G--------C---A-------------GG-G--C--G-N------T--AGT---------A--G----A------------A--G-----C-----T---  1098
CPZ.CM.-.CAM3               ------CG-T--------------A--C-GG------G-GG-T--A--GC-T--C---G-------G------T----AGTT--T--CT-A-----C-AG--G-----T--A--G--T--C--G--T---  1067
CPZ.CM.98.CAM5              ------C-----------G-----C----G-------G-GG-T--A--GC-C--C---G----G-A------CT----CGT---T-----A--T--C-A---G-----T--A--G--A--C--G--T---  1370
CPZ.GAB_X52154              ---T--C-----------G--TG------G-------G----TT-A--G-----C---G-G--T---------T----AGTA--T--C--T-----C--------G--------G-----------T---  1738
CPZ.GAB2                    ------AG----T--------G-----C-G-------G-GG-AT-A--GC-T-----GG-G------------T-----GTT-----CT---------A------G--G-----G--T------------  1055
CPZ.TZ.-.TAN1               AGG-GGAG-G-----------C--C--------G--------TT-A------------T----C--------C------GTT------T---------A---G--------A-----A--C-----T---  1318
CPZ.US.85.CPZUS             C-----C--T-----------C--A----G-------G-GG----A--GC-T------G------A------CT-T--CGTTG-T--CT-A--T--C-AG--G-----T--A--G---------------  1725
GSN.99CM166_AF468659        C-A----G--AAT--GTC---T--A----G--AG-----C--TT----CC-GC-GCG-TGT----A--------A----GT--AT---T-------C-A------C--T--A--G-------AG--A---  1357
GSN.99CM71_AF468658         CAAT--AG-TAAT--G-C---T--A----G--A------C---T----C--GC-GCG-TGT----A--------A----GT--AT---T---G---C-A---G--C--T--------A----AG--A---  1363
H2A.ALI_AF082339            GAAT---G-G--------GA-C--C----G---------CCAGA-A--GC--C-G--GTGT--C--G-----C-A---A----A---C--A-------A---------------G--A--CCA-AG----  1989
H2A.BEN_M30502              AAAT--CG-A--G-----GA-C--C--C-G---------CCAAA-A--GC-GC-A--GTGT--C----A---C-A---A--T-A---CT-A-------A---G--------A-----G--CCA-AG----  1996
H2A.ST_M31113               AAAT---G-A---------AGC--C----G---------CCAGA-A--GC--C-G--GTGC--C--G-----C-A---A----A---C--A-------A---G-----------G--A--CCAGAGT---  1441
H2B.D205_X61240             AAAT---G-------G---A-C-------G---------TCAAT-A-----GC-G---TGT--CC-------C-AT-------A---AT-A-------AG--G-----------G-----CCA-AG----  1957
H2B.EHO_U27200              AAAT---G-------G---A-C--C----G---G-----TCAGT-A--GC--C-G---TGT--CC-G-----C-AT-----T-AT--A----------AG-----G-----A--------CCA-AG----  1958
H2G.ABT96_AF208027          AAAC-----A---------A-T--Y----G---------CCART-A---C-TC-A---TGT--C--G-----Y-AT--A--A-A---C--A--TG---A------G-----A---------CAGTCT---  1359
H2U.FR.96.12034             AAAT--A--A--G--G--TA-C--C---CGG--------CCAAT----CC--C-A---TGT--T--------C-AT--A--T-AT--CT-A--T----AG-----G--------G--A---CA-AGT---  1475

MAC.239_M33262              GAAC--C--A--------CA-C-----C-GG--------CCAA-----G--GC-A---TGT--C----------A---A--A-A------A--TG---A------G-----A--G--A---CAGAG----  1943
Gag                         __N__P__I__P__V__G__N__I__Y__R__R__W__I__Q__L__G__L__Q__K__C__V__R__M__Y__N__P__T__N__I__L__D__V__K__Q__G__P__K__E__P__F__Q__S__Y    p27

SMM.SL.92.SL92B             AAAC--AG-------G-----C--C----G---G-----CCAG------C-CC-A---TGC--C--G-------A---A--A-A---C-----TG-T-A------T--------G------CA-AGT---  1386
SMM.US.x.PGM53              GAAC--C--A-----G---A-C-----C-GG--------CCAA--T--G---C-A--GTGT--C--------C-AT--A--A-A----------G---A------G--G--A--G--G--CCAGAG---C  1872
MNE.US.-.MNE027             GAAC--C--A--------CA-C-------GG--------CCAA-----G--GC-A---TGT--C-A--------A---A--A-A------A--T-----------G-----A--G--A---CAGAG----  1423
RCM.NG.x.NG411              CAAC--A--TG----G---A-C-------GG-AT-----TG-G--T--G---C-A---TGT--G-A--------A---AGTA-A---CT-A-----C-AG--G--------A--G--T----AG------  1105
RCM.GAB1_AF382829           CAAC--C--TAAT--G--CA-TC-G--C-G--AT--------AG-A--G---C-A---TGG----A------C-AT--AGTG-A---C--A--T----AG-----------A--GT-A--C-AG--T---  1105
DRL.FAO_AY159321            GGATT-C---AAT--G----C--------G-CA----G-GG--T----G---C-A-G-TGT----C------C-AT--AGT--A---CT----TG-G-AG-----T--T--A-----A--C-AG------  1187
MND-1.F17                   AAATG-A---AAT------AC---C-----G--T-----T--A-----CA-G----G-T-G----A--GTC--T-T--A-TA--T--AACA--TG-------G-----------G--T----A-------  1307
MND-2.5440_AY159322         AGAC-AGG-AAAC-------GT--A--C---CA-------G-A--A--GC-GC-A---TGT-----C-------A----GTA-AT---T-A-----C-A---G--------A--G--T----AG------  1190
MND-2.14CG_AF328295         GGA--AGG-AAAT-----G-C---C-----GCAG------G----A--GC-GC-A---TGT--G--C-----C-A---AGTG-A--------------A------G-----A--G--T--C-AG------  1557
MND-2.MND2_AF367411         AGATG-AG-AAAT--G----G---A------CA-------G-TT-A--GC-TC-G--GTGT-----C-------AT--AGTA-A---C----------AG--------------G--G--C-AG------  1633
DEB.CM.CM40                 AGG---C--AGAT------C----A----G-------G-C--A--A--TC-GC-A-G-TGT--G-A------C-AT--A----A---CT----TG-T-A------G--C--------A--C-A---T---  1456
DEB.CM.CM5                  AGGC--A--AGAT------C-G--A--C-GG------G-C-----A--C---C-G-G-TGT----A------C-AT--A--T-A----T----TG-T-AG--G--T--T-----------C-A-------  1447
MON.CM.L1                   CAAC--CG-T-AG--CTC---C--C--C-G--A----G-C---T----CC-GC-GCG-TGC--C-A------C-A----GT--A----T----T--C-A---G-----T--A-----A--C-A---T---  1344
MON.NG1                     ..................................................................................................................................     0
TAL.CM.8023                 GAAT--AG-AGAT--G--GTCC--A--C--G--G---G-C---A-A--C---C-GCGCTGT----A------C-AT--AGTG-A---C-----TG-G-A------C--C-GA--G-------AG--G--C  1063
TAL.CM.266                  GAAC--CG-AGAT--G----C---A--C---------G-C---A----C---C-GCGCTGT----A------C-AT--AGTT-A---A-----TG-G-A------C--C-G------A--C-A-------  1565
MUS.CM.1085                 CAAC-----AAAT----C------C--C--G-AG---G-G---A----GC-GC-G-G-TGT--C-A------C-AT--AGT--A---C--A--T--C-AG--G--------A-----A----AG--T---  1328
SAB.SN.-.SAB1C              GAATG--G--AAT------A-T---------G-------C--------TC-CC-G---TGT--C-A------C-A---AGTA-A----T-A-----T-A------G--T--A-----T--C-AG--T---  1946
GRV.ET.x.GRI_677            ----AGA--AGAC-----G-C-CAA--C-G--A----G-T--TT----C---C-A--GG----GCAG-----C-AT--CCAA-AGG-C--A-----TC----G-----T--A--------CCAG------  1793
TAN.UG.x.TAN1               ----AGAG-AGAT------CGT--C--C-G-G-----G-T---------C-CC-A---TGT----A------C-AT----T-TCAG-C--------C--------GG-C-----G--T----AG-----C  1816
VER.DE.-.AGM3               ---CAGAG-AGAT--G--T-CC--C----G---G-----T--------G--GC-A---TGT----A------C-A---AGTGTCTG-CT-A-----------G--G--C--A---G-A--C-A------C  1357
VER.KE.-.9063               C--CAGAG-AGAT--T--G-CT--C----G---------T-----A--GC-GC-G--GTGT----A------C-A---AGT-TCAG-AT-A-----C-----G-----T--A-----T--C-AG--T---  1856
VER.KE.-.AGM155             C---AGGG-AGAT-----G-CC--C---CG---------C-----A--G---C-A---TGT----A------C-AT--AGTGTCTG--T-A--T--C--------G--C--A-----A--C-A---T---  1845
VER.KE.-.TYO1               C--C-GGG-AGAT-----T-CC--C--CCGG--------T--T--A---C-TC-A--GTGT--C-A------C-A---AGTATCAG-C--A-----T--G--G-----T--A--G-----C-AG--T---  1346
SUN.GA.98.L14               AA--AGAG-TG-----CA---C--C------G-C---G-C--AGA---CA-GG-A-G-G----T---C-CC--CAGA-----TCAG-AG-A--A--T-----G--------A--G--T--C---------  1920
LST.CD.88.447               .A-C---G-AGAT------A-G---------GAG---G----T-AA-CTA-GG-A---G----T-----CC--CAA--AGTT--TG-AA-A-----C-AG--G--G-----A-----A--C-A---A---   849
LST.CD.88.485               .A-C--AG-GGAT------A-----------GA----G----T-AA-CTA-GG-A---G----T-----TC--CAA--AGTT--TG-AA-------C-A---G--G-----A-----A--C-A---A---   849
LST.CD.88.524               .A----C--AGAT-----GA----C------GA----G------A--CTA-GG-G---G-G--------AC--CAG--AGT---TG-AA---------A---G--T-----A-----A--C-AG------   843
LST.KE.-.lho7               .A-T---G-AGAT------A-----------GA----G-C--T-A--CTA-GG-A---G----C-----CC--CAA--AGTA--TG-CA-------C-A---------C--A--G--T----A---A---  1933
SYK.KE.-.SYK173             CAAC--AG-A-A------------A----G---------C--T--A--GC-GC-G--GTGT--GCAGG-C--C-A----GTTTCT--CT-A-------A---G--C-----A---ATT--C-A-------  1760
SYK.KE.KE51_AY523867        CAAC--A--T-AG--T---C-G--C--C-GGCAG--------T--A------C-A--GTGT--GCA-G-C--C-AT---GT--AT---T-------C--G--------T--A---A-A----AG--T---  1388
COL.CM.-.CGU1               ....--C--A......-CCA-C-------GGG-C------G-A-A-A-TC-GG-A--GG--A-CCAG--TGCA--A--AT--TCTG-A--A--T--T-----G---AGT--A--GGAT----A-AG----  1467
```

```
H1B.FR.83.HXB2              GTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTGGGACCAG  1807
Gag                         _V__D__R__F__Y__K__T__L__R__A__E__Q__A__S__Q__E__V__K__N__W__M__T__E__T__L__L__V__Q__N__A__N__P__D__C__K__T__I__L__K__A__L__G__P__   p24

H1D.CD.84.84ZR085           --------------C---------------------------A--A---------------------------------------A-------------------C-------------------  1324
H1F1.BE.93.VI850            ------A-A----T----GTC--------T-----------------C-----GGG-----------C--A----------------------------------C------------------A---  1150
H1G.SE.93.SE6165            -----TA------T----TG-T-G-----T-----------------T-----GGC-----------C-------------------------------------C--C----G------A----A--  1210
H1H.CF.90.056               ------A----T-T-------T-------T--------CA-------T--G--G-------------------------------------A--T--------C--------A----G------A----A--  1154
H1J.SE.93.SE7887            ------A------T----G----------T---------A-------T-----------------------T------C---------------A--T--------C-----C---------------A---T---  1130
H1K.CM.96.MP535             -----TA------T------C--------T--A-----CA-------A-----G-------------C----C------------C--A------------------C------------G----------  1006
H101_AE.TH.90.CM240         -----TA-----------------C-----G--A------A----------------C------------------C-A--------C-----T-----C-----T-C-----------------A---A---  1372
H102_AG.NG.-.IBNG           -----TA------T-------T-G-----T--A------A-G---------------C------------------C--------------------T-----C-----T-C-------G------A---A---  1338
H103_AB.RU.97.KAL153_2      -----TA------T---------T-----T---------A-------T----G--C---------------C--C---------------A--T------------A--C-------G------A---T---  1039
H104_cpx.CY.94.CY032        -----TA------T----TG---C-----A--A------A-C--------G-----------------------C--C--------------------T-----C------T-C--C-------------A---A---  1173
H1N.CM.95.YBF30             ------A-------C--------------A-----G--AA-------A-----G-----------------AC-C--A--A---------A-----------------ACAGC-CC----------A--G----  1371
H1O.BE.87.ANT70             -----T--------C-----AT-------T---------A-T--A--A-----------------------TC--C-T--T---------C--T---------C--ACAG-----G---T----A--G----  1856
H1O.CM.91.MVP5180           -----T-----T--C-----AT-------T---------A-T--A--A-----G-------------------C-T--T--G---T-A-----------------ACAA---C-G--------A--------  1831
CPZ.CM.04.EK505             -----TA----T--C--------------A-----G--CA-T-----A-----G---------------G--AC-C--A--A---------A--T--T---------A-A---CC-G-G------A-------  1355
CPZ.CM.04.LB7               --G---A----T--C-----AT-------A--A-----CA-------A---------------------C------C-T--A--G-----T-------------C--A-G---CC-G-G---------T----  1353
CPZ.CM.04.MB66              --G---A-A--T---------T----G--A--A------A----A--A---------------------C--T--A-----G--------T--T-------------T-C--A----G------A--G----  1350
CPZ.MT145                   ------A-------C--G--CT----G--T--A-----A--------T--T------------------C---C-----A-A--G-----T---------C-----A-AC--C----G------A-------  1346
CPZ.CD.-.ANT                --G--TA-A-----------AA-T-----A--A--G---------CCT--G---GCC--------------------AA----------C--T---------C--ACAC--CC-G--G--T------A---  1228
CPZ.CM.-.CAM3               -----TA-A-----C-----C-----G--A--A-----AA---A--A----------------T--C---C-C--A--A--G-----T--T-----C-----A-AC--C----G---CC---------  1197
CPZ.CM.98.CAM5              -----TA----T---------T-------A--A--G--AA-------A----------------T--C---C-T--A--T---------T--T-----C--C--A-GC--C--G-G---C--A-------  1500
CPZ.GAB_X52154              --G---A-A-----C------T----G--T---------------------A--G---------------T--C--TC-C--A--G--G-----A-----------------CAA---C------------------T-  1868
CPZ.GAB2                    -----TA-A-----C---GT-T-------A--A--G--------A--T--G-----C-----------G---C-CC---A--G-----C-----------------CAA--AC-T-GG---C-A-------  1185
CPZ.TZ.-.TAN1               ------A-A-----C-----AA-C-----A--A-----A--T--ACCA-------C-----------------T--AC----A---------A-----------------CA---C---------C-----G-A--  1448
CPZ.US.85.CPZUS             ------A-------C-----AT-------A--A---------T-CA--A-------CC-----------G---C-T-----A-----C--T--T-------------A-----------G---C---------T-  1855
GSN.99CM166_AF468659        ------A-A----T---GTGCT----G-----A-----AGAT----CC--G------------------C--T-GC--C----T---------T---------A--C---CTG--C--G-----CA--CC-GG-.  1486
GSN.99CM71_AF468658         ------A-A----T----TGCT----G--T--A--G--AGAT----CA-----G---------------C--TTGC--C----T---------T--T-----A--C---CTG--C--G--G--CA--CC-GG-.  1492
H2A.ALI_AF082339            -----TA-A-----C----GCT----G--A--A---A-AGAT-CA-CA-----G---------------TC---AC--C----A--G-----C---------C--C--ATTAG-GC-G----G---A--GATGA  2119
H2A.BEN_M30502              ------A-------C----GCT-G--G--A--A---A-AGAC-CA-CA---------------------CC---GC--C-AA-A--G-----C---------C--C---TTAG-AC------G-C----GATGA  2126
H2A.ST_M31113               -----TA-A-----C--G-GCT-G--G--A--A---A-AGAT-CA-CA---------------------CC---AC--C-A--G--G-----C---------C------TTAG-AC------G-C-A--GAT-A  1571
H2B.D205_X61240             -----TA-A-----C----GCT--C-G--A--A---A-AGAC-CA-CA--G------------------C---AC--C--A-T--G-----T-----------C---TTAG-GC-T--G-GC------ATGA  2087
H2B.EHO_U27200              -----TA-A-----C----GCT--C-G--A------A-AGAC-CA-CA---------------------C---AC--C--A-T--G-----T-----------C--ATTAG-GC-T--G-GC------ATGA  2088
H2G.ABT96_AF208027          -----TA-A-----C--G-G---R-----A--A---A--GAT-CA-CA---------------------C-G--AC-AC--A-T---------C-----------------ATTAG-AC----G-GG------ATGA  1489
H2U.FR.96.12034             -----TA-A-----C----G-T-G-----A--A---A-AGAT-CA-CA--G------------------TC-G--GC--C--A-T--G-----T---------C--C--ACTCG-G-------GT-----TATGA  1605

MAC.239_M33262              ------A-------C----G-T-------A--A--GA-AGATGCA-CA-----G---------------TC---AC--C--A-T---------T------------C---CTAG-GC-G--G-GGC----TGTGA  2073
Gag                         _V__D__R__F__Y__K__S__L__R__A__E__Q__T__D__A__A__V__K__N__W__M__T__Q__T__L__L__I__Q__N__A__N__P__D__C__K__L__V__L__K__G__L__G__V_    p2

SMM.SL.92.SL92B             -----TA-------C----G-T-G-----A-----GA-AGAT-C--CA--G--G---------------C----AC--C--A-T--G--C--C-----T-----C--ACTAG-A--------GG-----CATGA  1516
SMM.US.x.PGM53              -----TA-------C----GCT-------A--A---A-AGAT-CA-CA--G--G---------------TC---AC--C--A-T---------T----------------ATTGG-GC-C--G-GTC----CATGA  2002
MNE.US.-.MNE027             ------A-------C----GCT-------A--A---A-AGAT-CA-CA-----G---------------TC---AC--C--A-T---------T------------C---CTGG-GC-G--G-GGC----TATGA  1553
RCM.NG.x.NG411              -----TA-------C--GG-CT-G--G--A--A--G--AGAC-CA-CA--G--G---------------C--T-AC--C--A-T---------T--T-----C---------AG-AC-T--G-GTC-----ATGA  1235
RCM.GAB1_AF382829           --G--TA-A--T------G-CT-G-----A--A--G--AGAC-C--CA---------------------C--T-AC--C--A-A---------T---------C-----A-TGG-AC-C--G-GTC-----ATGA  1235
DRL.FAO_AY159321            -----TA-A--T------G-C--C--G--A--A--CA-AGATGCA-CA---------------------C---AC--C--A-A---------A--T-----C-----AGTG---C-C----GCC-A--GATGA  1317
MND-1.F17                   -----TA-A-----C--TGTAA-G-----A--A---------GGA--A-----G-TG------CAGC-GCATC--C-TA-AG-------A---------A--C---CAG-----G-G-AGC--A--GAA--  1437
MND-2.5440_AY159322         --G--TA-A-----C---G----GC-G--G----GCA-AGAT-CAC----C-----C----------GC-G--A--AC-CA----G-----C---------------AG---CCC-C----GG--A--CATGA  1320
MND-2.14CG_AF328295         --G---A-A--T------G----GC-G--G----G-A-AGAT-CAC-A--C-----CC-----------C----A---C-CA-----------C-------------------GC--C--G----G------CATGA  1687
MND-2.MND2_AF367411         --G---A----T--C---G----GC-G--G----GGA--GAT-CAC-A--C-----CC-----------C----T---C-CA----G-----T---------------AT-C--CC-T----GGC-A--CATGA  1763
DEB.CM.CM40                 --T---A-----------G--CT-------T--A-----AGAT----CA-----------------CACC---C-CA--A-T---------A-----T-----C-G--TC--CC-G----G---A--T-AGA  1586
DEB.CM.CM5                  -----TA-A-----C-----CT-G--G--T--A-----AGAT----CA--C--------------CACC---C-AA--A-T--G-----T----------------GA-TA---C------GG--A----AGA  1577
MON.CM.L1                   ------A-A-----C---TGC--C-----A--A---A-AGAT----CA--C--------------GC-G--GC--C---------------C-----------C--ACTC--CC----G--CA--...----  1471
MON.NG1                     ...................................................................................................-----G--C--C---CTC--CC-C------A--...----    37
TAL.CM.8023                 ------A-------C-G-TGC--T-----A--A---A-AGAT----CA--C----------------C-G--GC---A--G--G-----C---------A--C---G-C---C-G--G--CA----T--CC  1193
TAL.CM.266                  ------A-------C-G-TG---T--G-----A---A-AGAT----CA-------------------C-G--GC----------G-----T---------G--C---G-C--C--------CA----C----  1695
MUS.CM.1085                 --G---A-A----T----TG------G--A-----GT-AGAC--A-C---C----------------CTC----C---A--A--G-----T---------G-----ACTC---C-G--GAGCA--...----  1455
SAB.SN.-.SAB1C              -----TA-A-----C--GG---G------A-----GA-AGAC-CT-CT-----G--C----------C-GT-TC-----A---G-----C-------------------AG-GC-C----GG--A--CATGA  2076
GRV.ET.x.GRI_677            ------A-A---------G-C--G-----A--A-----AC-------T--T----------------C----T---C-TA---G-----C--T--G--------ATTG---C-G----G-------ATGA  1923
TAN.UG.x.TAN1               ------A-A--T---C--G-C--T-----A--A---A-AC-C--A--T---------------------------AC--C-CA---G-----C------------C---TTGG---G----G-C-T---AT-C  1946
VER.DE.-.AGM3               -----TA-------C---G-AA-------T-----G--C---GGA--A------C-G-------------T-A--AC-CA-T--G-----T--T-----C-----AGTC--CC----G-GCC-----ATGC  1487
VER.KE.-.9063               -----TA-------C---G-AA-C-----A--A--G--C---GGA--A------C-A-------------T-A--AC-CA-A---------T-------------C--AGTC--CC-G----G---A---ATGC  1986
VER.KE.-.AGM155             ------A-A---------G-AA-------A--A---------GGA--A--C---C-A-------------T-T---C-CA-T--G-----C-------------C--AGTA-----G--G-GCC-A--GATGC  1975
VER.KE.-.TYO1               --G---A-A--T--C--GG-AA-T-----A--A-----C---GG---A--G---C-A-------------T-A--AC-CA-T---------T--T---------------GTC--CC-G--G-GCC-A---ATGC  1476
SUN.GA.98.L14               AC---TA------TC--GG--T-------A--ATC--GC---G---A-----G-G-------A---G-AAA--C-A--G---------A----------------ATTGG-AA----G---C-----GA-A  2050
LST.CD.88.447               -C----A------T----G-AT-------A--AGG--GC-----T--A-----G-A-------AG--G-AGA--C-A--A---------A-------------G--TGG-GA-------C--A---GA--   979
LST.CD.88.485               -C---TA------T----G-AT-------A--AGG--GC-----T--------G-A-------A---G-AGA--C----A---------A-------------G--TGG-AA-------C--A---GA--   979
LST.CD.88.524               -CT--TA------T---GG-CT-------A--AGG--G------T--A--G--G-G-------A---G-AAA--C-A--G--G-----A-----T--C--C--A-TGG-AA----G--C--A---GA--   973
LST.KE.-.lho7               -C---TA------T----G-A--G-----A--AGG--GC--T--T--A-----GG-G-------AG--G-AAA--C-T--G--G-----A-----T--C--C-G-TTAG--A---------C-A---GA--  2063
SYK.KE.-.SYK173             ------A-A-----CC-TTGCT-------A-----G--GGAC-CATCA--C---GGG---C-A--TC----AC-----A-T---------C---------A--C-GACAG--CC-C--G---A---TCAA-.  1889
SYK.KE.KE51_AY523867        --T--TA-A------C-TTG---T-----A--A-----GGAC-C--CT-------------C-C---C-GT-A--AC-AA-T--G-----A-----T--G--C-GA-GC---C-G------A---TCAA-.  1517
COL.CM.-.CGU1               ------A-A--T---TC-G-C-----G--A--A-CT--AG--GGA--AA-C--GGCC------G-TA-T-AT---C-AA-A--GC----C-----T--C--C--A-GA---C-G----G-C--CAGAAG.  1596
```

```
                                                                   see Kaye, J Virol 72(7):5877-5885(1998) Gag p2 role in RNA encapsidation
                                                                                       Gag p24 \/ Gag p2
H1B.FR.83.HXB2             CGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCA......GCTACC...............AT  1916
Gag                        A__A__T__L__E__E__M__M__T__A__C__Q__G__V__G__G__P__G__H__K__A__R__V__L__A__E__A__M__S__Q__V__T__N__S__.__.__A__T__.__.__.__.__.__I    p2

H1D.CD.84.84ZR085          G-------T-------------------------------G--G--G---A-------A-----------------G-------------C-----G-G-C......T--G-TGCC............--  1436
H1F1.BE.93.VI850           G---------------------------------------G--------TA----------------------C--G--------T----C--ATTCAG-C...........................--  1253
H1G.SE.93.SE6165           GA---T----------------------------------G---------A-------A-----------A-----G-----------G-CCT--GG-G--......--AG-AGCC............--  1322
H1H.CF.90.056              G----T--A-------------------------------G--------TA-T-----A-----------------G---------------------A--......AA---AGCC............--  1266
H1J.SE.93.SE7887           GA--------------------------------------G--------T--T--------G-----------------------------G--C---A-C......AAC..................--  1236
H1K.CM.96.MP535            G----T--T-------G-----------------------G-----G--TA-------A------A----------G-----------G---------C--......-T-..................G-  1112
H101_AE.TH.90.CM240        GA------T----------------------C--------G--------TA-------A-----G-----------------------C-C-CA-C--G--......A--..................--  1478
H102_AG.NG.-.IBNG          G-------T-------------------------------G--------T--------A-----G-----------G--------T------CA-C-GG-C......AAC..................G-  1444
H103_AB.RU.97.KAL153_2     G-------T------------------T---C--------G-----------------A-----G--------C--G--------T------CA----G--......AAC..................--  1145
H104_cpx.CY.94.CY032       G-------T-------------------------------G---------A-------A-----------A-----G-----------G-C-T-----G--......--AG-AGCC............--  1285
H1N.CM.95.YBF30            GA-----CT-------G--------G--C--C--------G--G-----A-CA--------------GC-A--A--G--T---TCA--G--GCAGC-GC--......A-A--TAGT............G-  1483
H1O.BE.87.ANT70            GA-----CT-------G------GT---C--------------------AACT--------C-----GC-A--A---------GCTAC--CCCAGC-AGAT......TTG-AAGGAGGATACACAGCAG-  1980
H1O.CM.91.MVP5180          AA-----TT--------------GT---C-----A-----------G--AACT--C-------A-A-AC-A--A---------GCTTCT-CCCAGC-AGAT......TTA-AAGGAGGATACACAGCAG-  1955
CPZ.CM.04.EK505            GA--CT-C--G--G-----------------C--A--G--G--------A-CT--------------AC-A--A-----C----CA--G-C-CA--CAG--.....................ACTAGTG-  1464
CPZ.CM.04.LB7              GA--CT-CT----G-----------------C--------C--G--T--A-C------A--C--G--C--A--A-----T-----T----C--ATGCAG--.....................ACAGTC--  1462
CPZ.CM.04.MB66             GA-----CT------GGGATGA-C----C-TC--------G---------A-------A--T-----CC----A-----------T----C-CA-C-CAGT.............................  1451
CPZ.MT145                  GA--C--T--------------------T-----A--G------------TCA--------T-----GC----A-----C--------G----AC--AA--........................GAA--  1452
CPZ.CD.-.ANT               GA--CT-CT-----------T-A-----T-----A-----------C--A-C---------------G-----A-----T---GCTTCT-CT-AT---G--CAGGGAA-C...............GCAG-  1343
CPZ.CM.-.CAM3              GA-----CT----------------------C-----------G-----A-C---------T-----AC----------T---------C--CA---CC-T......A-A...............GGTG-  1306
CPZ.CM.98.CAM5             GA-----CT-------------------C--C--------G--------A-C---------T------C----------C------------CA------C......T-A...............AATG-  1609
CPZ.GAB_X52154             G---C--C-----G--------------T-----A--------------ATCT-----------G---C-A--------C---TCAATG--TCAG---CA-GGGAGA--A...............GATG-  1983
CPZ.GAB2                   GA-----C-----G--------------C--C--A-----------C--G-CA--------T------C-A--A-----T---------A--CA-C-GA--.....................AGCTCA--  1294
CPZ.TZ.-.TAN1              GA--A---T-----------C-C-----C-----A-----G---------TCT----------AGA--C--------------GC-TC--C----GCAGGGGGA...-TA..................-A  1557
CPZ.US.85.CPZUS            GA--C--C------------G-------------A-----G----------CT-----------G--C--A--------C---T-----A-G-A---CC-C......T-A...............AGCG-  1964
GSN.99CM166_AF468659       ..--C-ACT-G---------C-T--------C--A--G-----G-----AACT--C--AT-T--GC-C--A--A-----C---GCAAC--CCCTC--AGGC......A---GTTCA.........TAC-A  1599
GSN.99CM71_AF468658        ..--C-GCT-G---------C-T--------C--A--G-----G-----AACT-----AT-T---C-C-----A--G--C---GCAAC--C-CTC--AGGC.........-G-TCA.........TAC-A  1602
H2A.ALI_AF082339           ATC----CT-------G---C-A--C--C--------------G-----A-----G--A--C---T-AA----A-----CT-A-AGG-G-CC-TG-CAC--......---C-TATC......CCATTTGC  2237
H2A.BEN_M30502             ATC-C--C--------G---C-A--C--C--C-----G-----C-----A-----G--A--C--GC-AA----------CC-A-AAG-G-CT-TGGGAC--......AGCC-TATC......CCATTTGC  2244
H2A.ST_M31113              ATC----CT-----------C-A--C--C--------G-----T-----A-----G--A--C---T-AA----A-----CT-A-AGG-G-CC-TGGCAC--......--CC--ATC......CCATTTGC  1689
H2B.D205_X61240            ATC-C--CT----G------C-A--G--C--C--A--GA-------C--A--G--G--------GC-AA----C-----CT-A-AAG-G-CCCT--CAC-T......--AC--ATA......CCGTTTGC  2205
H2B.EHO_U27200             ACC-C---T-----------C-A-----C--C-----GA-------C--A--G--G--------GC-AA-------G--TT-A-AAG---CCTTG-CAC-T......T-C---AAT......CCGTTTGC  2206
H2G.ABT96_AF208027         ATC----TT----G------T-------C-----A--GR----------A--G--A--A--C--GC--A-----------T-A-AAG---CCTT---CC-T......A-AG--CTA......CCATTTGC  1607
H2U.FR.96.12034            ATC----CT-------G---C-A-----C-----A---A----------A--A--A--A--T--GC-GA----A-----CT-A-AAG-G-CCTT-C-AC-T......ATGC-TATA......CCCTTTGC  1723

                                                                                       Gag p27 \/ Gag p2
MAC.239_M33262             ATC-C--C------------C---G--T------A--------G--G--G--A--G-----T---T-AA----A-----CC---AAG-G-CCCTCGCAC--......-TGC-AATC......CCTTTTGC  2191
Gag                        N__P__T__L__E__E__M__L__T__A__C__Q__G__V__G__G__P__G__Q__K__A__R__L__M__A__E__A__L__K__E__A__L__A__P__.__.__V__P__I__.__.__P__F__A    p2

SMM.SL.92.SL92B            ATC-C--C--------G---C-A-----C-----A--G--------C--A--A--G-----CC-CC-CA----A--G--C----AAG-C-CCCTC-CGGGGTCTTTA-TGG-AGCACAGTTTAGGGGAGC  1646
SMM.US.x.PGM53             ATC-C--TT-----------C-------C-----------------G--A--A--A--A--T--GT-AA----A--G---T---AAG-G-C-CTG-CGC--......-GGCAACTC......CCATTTGC  2120
MNE.US.-.MNE027            ATC-C--C------------C---G--T---C--A--------------A--A--A-------A-T-AA----A-----CC---AAG-G-CCCTTGCAC--......-GGC-ACTC......CCTTTTGC  1671
RCM.NG.x.NG411             ACC----C--G--G--G---C-A--------C--A--G-C------T--ACAG--C-----T--G--GC-A------------CAGATG--GCA--G-AATATCATG--CCAACAGGGAGGACCT...-G  1362
RCM.GAB1_AF382829          ACC--T-TT-----------C-A-----C--------G------------CAG-----A--TC-G--AC-A--A--G--C---CAAATGA-GCA--G-AAT...........................--  1338
DRL.FAO_AY159321           ACC-C-GT--G---------C--TT---T--C-----T-----------A--------------G--CA----A-----T-----AG--CAGCAGGCAG-T......ATA-TGATGCAGCAGAAT.....  1436
MND-1.F17                  GA--A--TT----G------T--GA------------------T--G--ACAA-----A--C---T-AA----A---ATG-----AAC---GGT-GGACA-...TCACAA-ATTTTGTGCAGCAGAGAGG  1564
MND-2.5440_AY159322        ATC-C---T-G---------T--CTG--C--C--A--G-----------TAAAT-C------CA-A-GA----A--G--C---CAG----C-CAGGCAG-T......-TA-TGATGCAGAACTCGGGGGG  1444
MND-2.14CG_AF328295        ACC-CT-TT-----------T-ACT---------A-----G--G-----AAAAT--------CA-A-GA----A--G--C---CAGA-GA-GCAGTCAGA-......-TC-TGATGCAGAACTCGGGGGG  1811
MND-2.MND2_AF367411        ATC-CT-TT-------G---CC-CT---C--C--A--G--------C--TAAAT----A---CA-A-GA----A--G--T----AGG---CCCAGTCAG-T......-TA-TG...............--  1872
DEB.CM.CM40                ATC-C--CT-G-----G---C-ACAT-----C--A--G--G-----C--ACAG--A-----C---C--A----G--G--T---GC-TC--CCTT-...AA-GAGGCA-GGT--CTAGGCATGGTACAACA  1713
DEB.CM.CM5                 ACC-C--TT-----------T--CAT-----C--------------T---CAA--G--A--CC-TC-GA----G--G--T---GC-TCG-CCTT-...AA-GAAGCA-GCT--CTGGGGATGGTGCAGCA  1704
MON.CM.L1                  GT----GCT-G--G--C---C-ACAG--------------G-----T--AATG--C--AT-T--GA--C----A-----T---GCAGGG-C--T-GCCAACATGCCCATG-ATATGGTACAGGCCAGAGG  1601
MON.NG1                    GA--C-GCT-----------C-CCA---------A--G--------C----CA------AGC--GT-A--A--A--G--T---GC-AC--C--TT...AATAGCAACATGC-AATG............-A   152
TAL.CM.8023                A---G--C--G---------T--CAG--------A--G--C--G-----TCAG------AGC---C-CA----A-----C---GCGG-G-CCTTG--ACAG.............................  1294
TAL.CM.266                 GA--C--CT-----------C-CCAG--C-----A--G------------CAG------AGC--GC--A----A-----T---GCAG---CCTT---ACA-.............................  1796
MUS.CM.1085                GA--C--T--G--G------T--CAG-----C--------G---------ACT-----AT-C--GC--A----A--G--TC--GC-TCG-CCTTC...AA-CAGGCA-GC--ACTG............-A  1570
SAB.SN.-.SAB1C             ATC-A--C--------G---C-T--T-----C--A--CA-T------G-ACAA---------C-TC-AA----G-----T----CAGCG-CCTTTC-GCA-......CAA--AGTT......GGGAAC--  2194
GRV.ET.x.GRI_677           ATC-A--CT-G--G------C-A-T---T--C--------------G--ACAA--------T-AGC-AA---TA---ATG-----TA-T-G-CAG---ATG......-TCCAAGTG..............  2033
TAN.UG.x.TAN1              ATC-A--T--------G---C-------T--------------------A--A-----A----AGC-CA---TA---------CAA---A-GCA-GGAGTT...AACATGGTGCAA......GGAGCACC  2067
VER.DE.-.AGM3              ATC-C--T------------T-A--T--C-----A-----G--------AA-TT-C--A----A---GA----A---ATG---CAAA-TA-GCA--GCCAG...........................-A  1590
VER.KE.-.9063              ACC----CT-----------C----T--C-----A--------G--C--AA--T-C--A---------A----A---ATG---CAAA-CA-GCAG-C-CA-...........................-A  2089
VER.KE.-.AGM155            ACC-C--T--T---------C-------C-----A--G--G-----C--AA-TT-C--A--C-A---CA----A---ATG---CAGA-CC-GCAG-G-CAG...........................-A  2078
VER.KE.-.TYO1              ACC-C--C--T---------T-A--G--T--------G--------C--AA--T-C--A----A---AA----A---ATG---CAGACCA-GCA----CA-...........................-A  1579
SUN.GA.98.L14              -TC-AT--T----G--G-------G---C--C--A--G--T--G--G--GA-------A-G--AGA-CC----A---------GC-TCT-CTTTT-GGCAG......CAAGTAGGA......AGACAAGC  2168
LST.CD.88.447              G------TT-------G-------GG--C-----A--------------A-C------A-G---GA-------A---------GCAAC--CC-TGC-AAGT......CAA-TGAGA.........CAG-A  1094
LST.CD.88.485              G-----GTT-------C-------AG--C--------------------A-C------A-G--AGA-------A--G------GCAAC--CC-TGC-AAGT......CAA-TGAGA.........CAG-A  1094
LST.CD.88.524              GT--C--TT---------------A------------------------T-C------AAG--A-A-C--A--A--G------GCAAC--CC-TGC-GGGG......CAG-TGAAA.........CATC-  1088
LST.KE.-.lho7              G---AT-T--------G-------A---T--------------------T-CT-----A-G--A-A-A-----A---------GCTTC--CC-T-C-GCA-......CAAGTAAGA.........CAA-A  2178
SYK.KE.-.SYK173            ..C----CT-G--G------C--CA---T-----A-----C---------TTA--C-------A-C-CA----A--G--C---GTAATG-CCCAGC-G---......-TA..................-A  1993
SYK.KE.KE51_AY523867       ..C-A--CT-G---------C-TCA---T--C--A--G-----------ACAAT-C--A----A-C-A--A--A---------GTAATGAC-CA-C-CAGC......TTG..................GG  1621
COL.CM.-.CGU1              ..C-AT-TT-G-----C---C-AG-------C--------G--------A-A---C-------AG--GC-A--A--G--T---CAG--GT-CCAGC-AGAG.....................AGAACC-A  1703
```

```
                              / Gag-Pol fusion TF protein start
                      Gag p2 \/ Gag p7 NC
H1B.FR.83.HXB2            AATGATGCAGAGAGGCAATTTT...AGGAACCAA.............................................AGA............AAG...ATTGTTAAGTGTTTCAATTGTGGCAAAGAA  1983
Gag                       __M__M__Q__R__G__N__F__.__R__N__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__.__.__.__.__K__.__I__V__K__C__F__N__C__G__K__E_    p7

H1D.CD.84.84ZR085         ---------A-A-A--------...-A-GG-AC-.............................................---............--A...-----------------C------------  1503
H1F1.BE.93.VI850          -----------A-A-T------...-A-GG----.............................................---............-GA...G-------A-----T---------------  1320
H1G.SE.93.SE6165          -------------A--------...-A-GG--C-.............................................---............-GA...-C-A-C--A-----------------G---  1389
H1H.CF.90.056             -----------A------C---...-A-GG----.............................................---............--A...T-------A--C-----C-----------G  1333
H1J.SE.93.SE7887          ---------A-----T--C---...---G----T.............................................-A-............-GA...---------------------------C--  1303
H1K.CM.96.MP535           -----------A------C---...-A-GG---T.............................................---............--A...-----------C-----C------------  1179
H101_AE.TH.90.CM240       ---------------------C...-A-GG---G...........................................................--A...-GAA-------C-----C-----T-G----  1542
H102_AG.NG.-.IBNG         ----------------------...---GG---G...........................................................-GA...-CAA-A-----------C------------  1508
H103_AB.RU.97.KAL153_2    -----------A-A-T------...---GG--C-...........................................................--A...-GAA-------------C------------  1209
H104_cpx.CY.94.CY032      -----------A-A----A---...-A-GG----.............................................---............-GA...-C-A-------------C--------G---  1352
H1N.CM.95.YBF30           CT-TGCA--A--G--A--C---...-AAGG-AT-.............................................--G............--A...CCCA----A--------------------G  1550
H1O.BE.87.ANT70           -T-C-----A-----GC-AAA-...CCA-TTAGG.............................................-A-............GGA...-C-A-A--A--------C-----A-----G  2047
H1O.CM.91.MVP5180         -T-C-----A-----GC-GAA-...CCA--TAG-.............................................-A-............GG-...CCCA-A--A--C-----------A-----G  2022
CPZ.CM.04.EK505           TT-T-----A-----A---......................................................TTTAAAG-C......ATAAGA---...-CCA-C--A--C--T--------T-----G  1531
CPZ.CM.04.LB7             T--------A--G--A---......................................................TTTAAAG--......CCAAAA-GA...G--A----A---------------------  1529
CPZ.CM.04.MB66            .-ATGATGCA-AGA-GC-A......................................................TTTAAAG-C......CCAAAA-GA...-----C--A--C--T-----------G---  1517
CPZ.MT145                 CT-T-----A-----GGGACAA...................................................AATAGACCT......CCTAAT--A...-AGA----A--C--T--------A------  1522
CPZ.CD.-.ANT              CT-TC--------------GGA...-AT-GAGG-GGA..........................................-A-............-GA...CC-C-C--A-----T--C--C--T-----G  1413
CPZ.CM.-.CAM3             GT-CC----A--G--A---AA-...G----A-C-.............................................-CT............-G-...-AAA----A--C--T--C--------G---  1373
CPZ.CM.98.CAM5            GT-T-----------A---AA-...G---GAGT-.............................................-AT............---...-AAA----A--C-AT---------------  1676
CPZ.GAB_X52154            TT-CT-C----AG--AC-AGG-....................................GCGGGCCCA........................AAA-GA...-AAA-A--A--C--T--------G------  2050
CPZ.GAB2                  -T-C-----A----CAGGGGGA....................................AGAACTCCT........................CCT-G-...-GA--C-----C-----------A--G---  1361
CPZ.TZ.-.TAN1             T---C-----G----A--AAGA...CCACC-TT-.............................................-A-......AAG...GGT...CAGC-GC-------T--C-----G----T-  1627
CPZ.US.85.CPZUS           -T-TC-A----A---A--CGC-...G----G-C-.............................................G-G............-GA...-AAA-C--A-----T--------------G  2031
GSN.99CM166_AF468659      C---G-----...................................................AAAGGCCCCCCTGGAAGG.........CAAGGA--A...-CCCCA--A----AT--------AC--TTC  1666
GSN.99CM71_AF468658       T---G----A...................................................AGAGGCCCCCCAGGAAGG.........CAAGGA--A...-CCCCA--A----A---------A--GTTT  1669
H2A.ALI_AF082339          GGCAGCC--ACA-..................................................................---............-G-...-CAA-------C-GG--------A--G---  2286
H2A.BEN_M30502            -GCAGCC--ACA-..................................................................---............---...GCAA---G--A--GG--C-----A--G--G  2293
H2A.ST_M31113             -GCAGCC--ACAG..................................................................---............-G-...-CAA-------C-GG-----C--A--G---  1738
H2B.D205_X61240           TGCCG-T--ACA-AAAGCAGGG....................................AAG..................---............GG-...-CA--G-CA--C-GG--C---------C-G  2266
H2B.EHO_U27200            CGCCGCT--ACC-A-AGCAGGG.........................................................-AG............-G-...-CA--G-CA--C-GG--C--C-----G-CG  2264
H2G.ABT96_AF208027        -GCTGCT--ACA-AAA-CAGGA....................................GGA..................-A-............-G-AGC-CAA-A--A----GG-----------G--G  1671
H2U.FR.96.12034           -GCAG-C--ACAGA-AGGCGGCCCA......................................................--G............-G-...CCA--C-GA----GG--C--------G---  1784

                      Gag p2 \/ Gag p8 NC
MAC.239_M33262            -GCAGCC--ACAGA-GGGACCA.........................................................---............---...CCAA---------GG--------G-----G  2249
Gag                       __A__A__Q__Q__R__G__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__.__.__.__.__K__.__P__I__K__C__W__N__C__G__K__E_    p8

SMM.SL.92.SL92B           -GC--AAGGCCA---G...............................................................-AT......AAG...CCT...--CA-C-G------T--C-----A--GAC-  1701
SMM.US.x.PGM53            -GCAG-C--ACAGAAAGGACCA.........................................................---............---...-CAA-------C-GG--------A--G--G  2178
MNE.US.-.MNE027           -GCAGCC--ACAGAAGGGACCA.........................................................---............---...CCAA---------GG--------G-----G  1729
RCM.NG.x.NG411            --AAGGA-CCCCTAAACTAGGG...G-AGGA................................................CCT............-G-...T-CC-C--A--C-AT------------AC-  1426
RCM.GAB1_AF382829         C---GCT---CA-TCAGCAAAC..............................AGGGGGCCTCCA............AGA---TCAGGAGGAAATCCAAAT...T-A-GA----A---------T--GCC-  1423
DRL.FAO_AY159321          .CCCCCCAGAG-TCCACCGCGG.................................GGACCCAGGGGACCCCCTCCA..................--A...-ACCC--GA---CCA--C--C--GC-GTTT  1511
MND-1.F17                 GCCTCAAAGAG--CCAGT-AGA..............................CAACCTACTGGA...............--G............--A......CC--TC--C-----C---AAT------  1631
MND-2.5440_AY159322       -CC-CC--G-G-TCC-CCGAGA.................................CAACCCCCT..............................-GA...-ACCC--GA--CCC------C--A--GTTT  1508
MND-2.14CG_AF328295       -CCACC--G-G-TCCTCCGAGA.................................CAACCCCCT..............................-GA...-ACCCC-GA---CC---C-----A-GGT-T  1875
MND-2.MND2_AF367411       GCA--ATTC-G----GCCACCG..............................CGGGGCCCCCCG...............---CAACCTCCT...-GA...-ACA-C-GA--CCCT--C-----A---TTT  1951
DEB.CM.CM40               --G--GAGGACCCCCTGG--C-.....................AGAAGACAG...................................................A---GA--C-----------AC--ATT  1771
DEB.CM.CM5                --GAGGAGGACA---AGGAGG-..................CCTAGAAGACAG...................................................C---GA--C--T--------AC-GATT  1765
MON.CM.L1                 -CCACCC--A--GAAGGGACAA.................................................................................CCC--A--C-----------A---TTT  1650
MON.NG1                   C---G-A--AG-CA-AGGAGG-......................................................RGTCAGCCACGCCGACAAGGCRWGCAAA-C-G-----AT--------G---TTT   228
TAL.CM.8023               .-AC-CC---CA--T--TA......GCC-TGGT-...CAGGGACAGGGAGGTCCGAAAGGGAGAGGGGGA......CCA---AGAACTCCCCCAGGA...CAAA-C-GA----AT-----------GTTT  1405
TAL.CM.266                .-AC-CC--ACA--T--TA......GCC-TGGT-CAACAGGGGCAAGGAGGCCCAAGGGGGAGAGGGGGA......CCT---AGGAACCCCCCAGG-...CAGA---G-----AT--C-----A---TTT  1910
MUS.CM.1085               C---G----AG---C---GGGA...GCC-GGGG-AGGGGCAACGGCAGCCAGGGA...AACCGAGGAAAT.................................CCC-GA--C-A---C-----GC--T-T  1661
SAB.SN.-.SAB1C            CT-TG-----CA---GGCAAGA...CCC-GGGGTCCCTTAGGGGGAAGAGGTAGACCTCTG..................-AC............CCA...-A-A----A----A---------A--GCCT  2288
GRV.ET.x.GRI_677          .GGACCT----A-AAGGGCCCC.........................................................C--............GG-...CCGC-A--A--C--T------------TTT  2090
TAN.UG.x.TAN1             GCGAGGAGGT--G----GAGGC.........................................................---............GGC...CCACCCCGC--C-----A-----AC--ATT  2125
VER.DE.-.AGM3             C--------ACAG---GG-CAG...--AGGAAG-CCA..........................................---............CCC...CCA--A-------A---------A---TTT  1660
VER.KE.-.9063             C---G-----CA---GGCAGCA...G-AGGAGT-AGGAGACAG....................................---............CCC...CCAC-A--A--------C-----A---TTC  2165
VER.KE.-.AGM155           C---G-A---CAG--AGG-GGA...---GGAAG-CCA..........................................---............CCC...CCGCCA-------A---C-----A---TTT  2148
VER.KE.-.TYO1             C---G-----CAG--AGG-CCA...-AA-GA---.............................................---............CCC...CCAC-A-GA----AT--------A---TTT  1646
SUN.GA.98.L14             T---G-----CA-AA-TTACCA...CCC-GAA-CAGCCAGGGGAGGTTTGTAAGAATAGGAGGAGGAGGA......CCT--G............---...CCCA-G-CT-----T--------T---CC-  2274
LST.CD.88.447             T---G-A--A...-TG-CACCA...CCA-GGA-TGCACAGGGAAGGTTTGTAAGAACAGGAGGAGGAGGC......CCT---............---...CC-T-A-CT--C--T--------A--GCC-  1197
LST.CD.88.485             T---G-A--A...-TG-CACCA...CCA-GGA-TGAGCAGGGAAGGTTTGTAAGAACAGGAGGAGGAGGC......CCT---............---...CC-T-A-CT--C--T--------A--GCC-  1197
LST.CD.88.524             C---G-----...CAA-CACCA...CCT-GAA-TGCACAAGGGAGATTTGTAAGAACGGGGGGAGGTGGC......CCA---............-GA...CC-C-C-CT--------------G--GCC-  1191
LST.KE.-.lho7             C---G-----...-T--C-CCG...CTT-GAA-TGCGCAAGGCAGGTTTGTGAGAACAGGAGGAGGAGGT......CCC--G............--A...CCCC-C-CT--C-----------A--GCC-  2281
SYK.KE.-.SYK173           T---G-A--AG-GCCATCAAAA...G--...................................................---............TCA...--GA----A--C-A---------GC--ATT  2054
SYK.KE.KE51_AY523867      C-----A--AG-GCCA...AGACAAG-A..........................................AGCAATCCC--G............-GAGGGCCAACA-GA--C--T--------GC-GTT-  1694
COL.CM.-.CGU1             C-----TG-AGT-AAG-CAGCA....................................................................................--A--C---------CAGGG-AT-  1749
```

```
H1B.FR.83.HXB2             GGGCACACAGCCAGAAATTGC...AGGGCCCCTAGGAAAAAG............................................................GGCTGTTGGAAATGTGGAAAGGAAGGAC  2050
Gag                        _G__H__T__A__R__N__C__.__R__A__P__R__K__K__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G__C__W__K__C__G__K__E__G__    p7

H1D.CD.84.84ZR085          -------T---A--------T...------------------............................................................----------------------------  1570
H1F1.BE.93.VI850           --A----T-------------...-----------A------............................................................-------------------GA-------  1387
H1G.SE.93.SE6165           --A--TCT-------------...------------------............................................................----------------------------  1456
H1H.CF.90.056              --A----T-------------...------------------............................................................-------------------GA-------  1400
H1J.SE.93.SE7887           --A----T---A-A---C---...-----------A------............................................................----------------------------  1370
H1K.CM.96.MP535            -------T-------------...-----------A------............................................................-----------------G----------  1246
H101_AE.TH.90.CM240        --A---CT-------------...-----------A---C--............................................................--T-----------C--G----------  1609
H102_AG.NG.-.IBNG          --A---CT-------------...-A--------------G-............................................................-----------------G--A-------  1575
H103_AB.RU.97.KAL153_2     --A---CT-------------...------------------............................................................--------------------------C-  1276
H104_cpx.CY.94.CY032       --A--TCT-------------...------------------............................................................-----C-----G-----------G----  1419
H1N.CM.95.YBF30            --C--TTTG--A-----C--T...-A---------A-G-GGA............................................................-----------G-----GC-A-------  1617
H1O.BE.87.ANT70            --A--T-T---A--------T...C-A--A------------............................................................--T--C------------C-------T-  2114
H1O.CM.91.MVP5180          --A--T-T---A-A---C--T...C-A--A-----A----G-............................................................--T--C------------C-------T-  2089
CPZ.CM.04.EK505            --C--TCT---A-----C--T...-A---A-----A------............................................................-----C-----------GC-A--G----  1598
CPZ.CM.04.LB7              -------T---A--------T...-AA-----C----G---A............................................................-----------------GC---------  1596
CPZ.CM.04.MB66             --T----T---T--G------...-A------C--A-G---A............................................................-----------G-----GC-A-----G-  1584
CPZ.MT145                  --A---CT---A--G--C---...-----A--CC-A-G----............................................................-----C-----G------C----G----  1589
CPZ.CD.-.ANT               --C--T--T--A---------...-A------A--A-GG--A............................................................-----C----G-------C---------  1480
CPZ.CM.-.CAM3              --A--TTT---------C---...--A-----C--A-GG---............................................................--A--C-----G-----GC-A--G--G-  1440
CPZ.CM.98.CAM5             -----TGT---T-----C---...--A-----A--A-G----............................................................--A--C--------C---C-------G-  1743
CPZ.GAB_X52154             --T---CT---T--------T...-AA--A--A--A-G---A............................................................---------CGG-----GC-A--G----  2117
CPZ.GAB2                   --T--T-T---A--------T...-AA--------A-G----............................................................--A--------G--------A-------  1428
CPZ.TZ.-.TAN1              --C--T-----A--------T...-----T--A--A--G--A............................................................--T--C----GG------C-A--G----  1694
CPZ.US.85.CPZUS            --A--TCT-------------...-----T--A--A-G---A............................................................--G--C-----G-----GC---------  2098
GSN.99CM166_AF468659       --A---CT---A---G----T...CCTAAA--C-AAG---GA............................................................AAG--C-TT--------C-GA-C---G-  1733
GSN.99CM71_AF468658        --A--T-TC--A---G-C--T...CCCAAA--C-A-G---GA............................................................AAG----TT--------C----C-----  1736
H2A.ALI_AF082339           ------T-G--A---C-A---...C-A--A--C--A---C--............................................................-----C-----------C---CC-----  2353
H2A.BEN_M30502             --A---T-G--A---C-G---...C-A--A-----A-G-C--............................................................-----C-----G-----C---CC-----  2360
H2A.ST_M31113              ------T-G--A---C-A---...C-A--A-----A-G-C-A............................................................-----C-----------C----C-----  1805
H2B.D205_X61240            --A-----------GC-A---...-----------A-G-C--............................................................--A--C--------------AAC-----  2333
H2B.EHO_U27200             --A--T--------GC-G---...-A---------A-GGC--............................................................--A--C------------C--C------  2331
H2G.ABT96_AF208027         --A--T----TT---C-G---...--A-----C--A-G-C--............................................................--G--C-----G-----G---CC---C-  1738
H2U.FR.96.12034            --A-----T--A--GC-G--T...-AA--------A-G-C--............................................................--A--C----------A----CC-----  1851

MAC.239_M33262             --A---T-T--A--GC-A---...--A-----A--A-G-C--............................................................--A--C--------------AATG-AC-  2316
Gag                        _G__H__S__A__R__Q__C__.__R__A__P__R__R__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G__C__W__K__C__G__K__M__D_     p8

SMM.SL.92.SL92B            --A---T-T--G---C-G---...--A--T--C--A-GG---............................................................--A--C-----------GG-A-------  1768
SMM.US.x.PGM53             --C---T-T--A-A-C-A---...--A--------A-G-C--............................................................-----C--------------AAC---C-  2245
MNE.US.-.MNE027            --A---T-T--A--GC-A---...--A-----A--A-G-C--............................................................-----C------------C-AATG--C-  1796
RCM.NG.x.NG411             --T--T--T--T---T-----...-AAA-AAG---A--G--A............................................................--G--C----G-------G-------T-  1493
RCM.GAB1_AF382829          --A----TTT-T---T----T...-AA--------A-GG---............................................................--A--C------------TCCCC--AC-  1490
DRL.FAO_AY159321           --A--T--CCTG---C-A--T...-CAAAG-----A------............................................................--A--C-TT-GG------GCTTT--AT-  1578
MND-1.F17                  -----TGT---A--GTTC-T-...-A---------A-GG--A............................................................--G--C-----T------GCAATG-AT-  1698
MND-2.5440_AY159322        --A--TGT-CTG--GG-----...--A-----A--A--G-GA............................................................-------TT--G------G-TCC-----  1575
MND-2.14CG_AF328295        --A--TGTTCTG---G-C--T...--ACT---C--A--G---............................................................--A--C-TT--G-----GG-CAC-----  1942
MND-2.MND2_AF367411        --T--TGGGCTG--GG-C--T...-TAAG----C-A------............................................................--A--C-TC--G------G-TCT-----  2018
DEB.CM.CM40                --A--TCTCCAA-AGG-C---...-AAAGA---...-G----.........................................................ACCAAA----TT--------GC-------C-  1838
DEB.CM.CM5                 -----TGTGCAA--GG----T...-AAAAG---...-G---A.........................................................GTTAAA--C-TC--------T-----G--C-  1832
MON.CM.L1                  -------TG--G-A-------...-AA-----CCAA-GG-G-............................................................AAA----AC--T-----CC-ACCG----  1717
MON.NG1                    --A--TGTG--A-AG--C--T...-CT--T--C--A--G-CA............................................................--T--C-TCCGC--C--C--------G-   295
TAL.CM.8023                -----T-TT--A-A---C--T...CC---T---CA---GCCA.............................................CGGGGGCCTCCGGGGA------TC--G--C--GC--ATG----  1487
TAL.CM.266                 --C--T-T---A-AG--C---...CC------GCC--G-C--.............................................AAGGTGCCGCCAGGCACT--C-AT--G-----C--ACC---C-  1992
MUS.CM.1085                --C---GT------GG-----...CCCAAA--C...-GG--C.........................................................AAAAAG----TC--------G-GA-------  1728
SAB.SN.-.SAB1C             --T---CTG--T---TT---T...-A-----------G-C-A............................................................-----C-----------C-GCCC--AC-  2355
GRV.ET.x.GRI_677           --A--T-TGCAA--GG-A---...-A---A--A--AC-G-TC............................................................AAA--C-TT--G--C--C--AATT--C-  2157
TAN.UG.x.TAN1              -----T-T-CAA-A-G----T...CCAAGAG-AG-ACC---C............................................................AAA--CCTA--------G---CC-----  2192
VER.DE.-.AGM3              --C--T-TGCAA---C-A---...CCT-AA--A--A--G-T-............................................................A-A--C-T---------G--ACC---G-  1727
VER.KE.-.9063              --A----TGCAA--GC-A---...-CA-AA--A--A--G-TA............................................................AAA--CCTT--------G---CT-----  2232
VER.KE.-.AGM155            --C----TGCAG--GC-G--T...CCT-AG--A--A----TA............................................................AAA---CTT------------CC---G-  2215
VER.KE.-.TYO1              --C--T-TGCAA---C-A--T...CC--AA--A-------CA............................................................AAA---CTA--G--------ATTG----  1713
SUN.GA.98.L14              --A--TCT---A-ATC-G---...--A-AG--A-A----GGA...................................................CCACCCGGCTC---C---------------ATG----  2350
LST.CD.88.447              --T--T-----A----TG--T...--ACAG-----AC-GG--............................................................--A--C-----C-----T-GCA-G-A--  1264
LST.CD.88.485              --T--T-----A----TG--T...--ACAG-----AC-GG--............................................................--A--C-----C-----T-GCA---A--  1264
LST.CD.88.524              --A--TT----A-A--TG---...---CAG-----AC-GG-A............................................................-----------C-----TGCTA---AC-  1258
LST.KE.-.lho7              --T--------A----TG---...--ACAG--C--AC-GG-A............................................................--T--------C-----C-GTA---AG-  2348
SYK.KE.-.SYK173            -------TGCAA-A-G-C---...-AAAAG--ACTT---GCT............................................................AAA--C-TT--T--------AAC-----  2121
SYK.KE.KE51_AY523867       --A--TCTTCAA-A-G-C---...CCTAGG--A-AA--GCTT............................................................AAG--C-TT--C------GGAACG----  1761
COL.CM.-.CGU1              ------TT--------TG--TCCC-AAAGA--A-TAGG-GGAGCCGGTCGAGGAAGAGGCCGAGGACGGGGAGGCTTTAGAGGAGCTCCCAGAAGGCCAGTGA-G----TT-CC--CAATC---------  1879
```

```
                                                                           stem-loop after TTTTTT slip potentiates slippage
                                                                                      Gag-Pol -1 ribosomal slip site
                                                                           Gag p7 end \/ Gag p1 start                                                 Gag p1 end \/ Gag p6 start
H1B.FR.83.HXB2            ACCAAATGAAAGATTGT......ACT....................GAGAGACAGGCTAATTT.TTTAGGGAAGAT...CTGGCCTTC...........CTACAAGGGAAG...GCCAGGGAATTTTC  2134
Gag                       H__Q__M__K__D__C__.__.__T__.__.__.__.__.__.__.__E__R__Q__A__N___F__L__G__K__I__.__W__P__S__.__.__.__.__Y__K__G__R__.__P__G__N__F__    p1
Pol                                                                       _F__R__E__D__.__L__A__F__.__.__.__.__L__Q__G__K__.__A__R__E__F__S Pol TF

H1D.CD.84.84ZR085         ----------------C......---....................--A--------------.---------A--...T--------...........-C----------...---G-----C----  1654
H1F1.BE.93.VI850          -------------C--C......---....................--A--------------.---------A--...T--------...........-A------G---...---C--A-----C-  1471
H1G.SE.93.SE6165          -T-----------C--C......--G....................-----------------.---------A--...T--------...........-A------G---...--------------  1540
H1H.CF.90.056             -T--G--------C--C......--A....................-----------------.---------A--...T--------...........-AG--A--G---...------A-------  1484
H1J.SE.93.SE7887          -------------C--C......---....................-----------------.------------...T--------...........-AG--A--G---...---------C----  1454
H1K.CM.96.MP535           -T--G--------C--C......---....................-----------------.---------A--...---------...........-C------G---...--------------  1330
H101_AE.TH.90.CM240       -T-----------C--C......---....................-----------------.---------A--...T--------...........-A----------...---G----------  1693
H102_AG.NG.-.IBNG         -------------C---......---....................-----------------.---------A--...T--------...........-AG-----G---...------A-------  1659
H103_AB.RU.97.KAL153_2    -T-----------C--C......---....................-----------------.--------GA--...T--------...........-AG--A--G---...------A-------  1360
H104_cpx.CY.94.CY032      -T-----------C--C......---....................-----------------.--------GA--...G--------...........-AG--A--G---...------A-------  1503
H1N.CM.95.YBF30           -T---------------......-AAAATGAA..............-GA--------------.-----------G...----T--C-...........--T--A--G---...A-----A--C--C-  1707
H1O.BE.87.ANT70           ----------------C......-GAAAT.................-GA-A-----A------.------C--ATA...-------C-...........GGGGGGCACG--...------C----A-G  2201
H1O.CM.91.MVP5180         ----------------C......-AAAAT.................-GA-------A------.----------TA...-------C-...........GGGGGGCACG--...------C----A-G  2176
CPZ.CM.04.EK505           -T-------G--C--C......-GAAGTGGA..............--A--G----G------.----------G-CTG..---...TCTCTCCAA........A--...GAG-----A-----C-  1688
CPZ.CM.04.LB7             -T-GG---C--------......-ACATG.................-G--AG...-T------.---------AC-CTG..---...TTCCTCCCA........---...AAG-----A--C----  1680
CPZ.CM.04.MB66            -T--G----G-A----C......--AAATGAA................---------------.---------AC-CTG..---...TTCCAGCAA........---...AAG----------C-  1671
CPZ.MT145                 -T--G------------......-A-CAAAATAAT..........-GA--G...AT------.------A--A-ATTG..----C-CTGGGG.........---...GAG--------C-A-G  1679
CPZ.CD.-.ANT              ----GC-T---A-C---......C-AGCAACA..........AATACAG--A-A-TA-----.---------ACC...GACC--CA-GTGGTG...GGGG-G--G........A--------C---G  1576
CPZ.CM.-.CAM3             -T---------A-C--C......C-AGCTGGA..............--A--------------.---------AGG...T---T--C-GTGGAG...CGGGGGATCAAAG--...A--------C--C-  1539
CPZ.CM.98.CAM5            -TA-C--------A---......--AACTGGA..............-GC--------------.----------GA...T---T--C-GTGGAG...CGGGAGATCAAA---...A--------C----  1842
CPZ.GAB_X52154            -------------C--C......--A....................-GA-------TG-----.---------AGG...T--------CCGCAG...CGGGAG..........-------C--C---G  2201
CPZ.GAB2                  -T--G-----G--C--C......CAAAAT..............GAA-GA------TG-----.-------GG-GG...---------............-CGG-GC--G--...---------C----  1518
CPZ.TZ.-.TAN1             -T--------G--C--C......--CACCAGAAAC.........AACAGC-CTGG--TA-----.---------ACG.....CA-CC-CTTGTG...GGGG-G--G........A--------C---G  1790
CPZ.US.85.CPZUS           -T-----------C--C......--AGCAGGA..............A-C--------------.---------ACA...T---T-CC-ATCGTGGAGCGGGGGATCAAAG--...A--------C--C-  2200
GSN.99CM166_AF468659      --TTT-GC---CCA---AGA...--A...........GGGACACCAATC-AGG--CT-T----.-----ACCTTGGCCGGG---G...............-CG-CGA-CCCCCAG-AACTTTCCAG-AA  1832
GSN.99CM71_AF468658       --TT-GCT-GGC-A--CAAA...--A...........GGAACACCA-CC-AGG--CT------.-----ACCTTGGCCGGG---G.................ACGA-CCCCCAG-AACTTTCCAG-AA  1832
H2A.ALI_AF082339          -T-TT---GC-A-C--C......C-A....................--A--------GG---.-------CTTGGCCCA---GG...............AA-G---CCTC-...CAACTTCCCCG-AA  2437
H2A.BEN_M30502            --ATC---GC-A-C--C......C-G....................--A-------AGG---.-------TT-GGCCCAC--GG...............AA-G---CCTC-...CAACTTCCCCG-GA  2444
H2A.ST_M31113             --ATC---GC-A-A--C......C-A....................--A-------GGG---.-------TT-GGCCCA---GG...............AA-G---CCCC-...CAATTTCCC-G-GG  1889
H2B.D205_X61240           --ATC---TC-A-A--C......C-A....................--A-------GGG---.-------TTAGGACC----GG...............AA-G---CCTC-...CAACTTCCCCA-GA  2417
H2B.EHO_U27200            --ATC---TC-A-A--C......C-A....................--A-------GGG---.-------TTCGGCCC----GG...............AA-G---CCTC-...CAATTTCCCCG-C-  2415
H2G.ABT96_AF208027        --ATT---GCCA-A---......C-A....................----------AGG---.-------TTTGGACCT---GG...............AA-G---CCTC-...CAACTTCCCCA-GA  1822
H2U.FR.96.12034           --A-C---GCTAGC--C......C-A....................--A-CCA---TGGG---.-------CTTGGCCCT---GA...............AA-A---CCTC-...CAACTTCCCCA-GG  1935

                                                                            Gag p8 end \/ Gag p1 start                                          Gag p1 end \/ Gag p6 start
MAC.239_M33262            -TGTT---GCCA-A--C......C-A....................--C-------GGG---.------CCTTGGTCCA---GG...............AA-G---CCCC-...CAATTTCCCCA-GG  2400
Gag                       H__V__M__A__K__C__.__.__P__.__.__.__.__.__.__.__D__R__Q__A__G___F__L__G__L__G__P__W__G__.__.__.__.__.__K__K__P__R__.__N__F__P__M__    p1
Pol                                                                       _F__R__P__W__S__M__G__.__.__.__.__.__K__E__A__P__.__Q__F__P__H__G Pol TF

SMM.SL.92.SL92B           GTATTCA-GCCA-C--C......C-A....................A-CCAGA----GGG---.------CCTTGGCCCA---GG...............AA-G---CCTC-...CAACTTCCCCA-G-  1852
SMM.US.x.PGM53            -TGTC---GCCA-A--C......C-C....................--A-------GGG---.-------CTTGGCCCA---GG...............AA-G---CCCC-...CAATTTCCCCA-GG  2329
MNE.US.-.MNE027           -TGTT---GCCA-A--C......C-A....................--C-------AGG---.------CTTTGGCCCA---GG...............AA-G---CCCC-...CAATTTCCCCA-GG  1880
RCM.NG.x.NG411            -T-T------G-----C......C-A.......................-A-G---T------.---------TGCTCC----AA...............GC-GGG-AA-CC...-AGGAACCT-CCCT  1574
RCM.GAB1_AF382829         -T-TCT-------C--C......--AAAG....................--AATA-----.---------GAC-CCC----GG-CA...........GGGG---CCG--...-AACTTTCC---GA  1574
DRL.FAO_AY159321          -TATGT-A-G-A----C......C-AAAA....................----TG-----.---------T-C......................TCC-TG---C--CGG-AA-CCC-GGAA-T  1653
MND-1.F17                 -T--G-AAGCTC-A--C......C--AAG.........CCAGCTCAGC--CAGAG--T------.-------T-TGG....................-CCTTG---TCCCTCCAA-CC-GGGAA-T  1788
MND-2.5440_AY159322       -T-T-----GGA-C--C......C-AAAG.....................AT--TG-----.---------T-C....................-CC-TG---C--TGGCAAGCCC-GGAACT  1650
MND-2.14CG_AF328295       -TATG----G-A-C--C......C-AAAA.....................AT--TG-----.---------TGC....................TCC-TG---C--TGGCAAGCCC-GGAACT  2017
MND-2.MND2_AF367411       --AT-----G-A----C......C-AAAG.....................AT--TG-----.---------T-C....................-CC-TG---C--TGGCAA-CCC-GGAACT  2093
DEB.CM.CM40               --ATTGCA---A-----...GGACAG..................ACTCCGAG---A-----.---------T-C....................TCC-TG--CCTCCAG-AG-CCC-GGAACT  1919
DEB.CM.CM5                --AT-GCA--GA-C---...GGACAA..................-TACC-AG---A-----.---------TGC....................TCC-TG--CCTCCCG-AG-CCC-GGAA-T  1913
MON.CM.L1                 ---TCGCC-----C--C......C-G.....CAGCCCCCCAAACAAA-C-A-GG--TG-----.---------TCCTTT-G-T--...CAAGAAGGGGCCAAGG--CTTCCC...T-TCACCTC-G-G-  1828
MON.NG1                   -TRT-TCC---A-C--C...CCA-ACGGAGGACAAAAC...CAACCCAGA-ACAG--TG-----.---------TCCTTTTG-..........................................   371
TAL.CM.8023               ---GTGCCGC-C-G--C......-GACAG.....................CC------A-----.----------GCTTTT-TAGG...............GA-G-TCCCCCC...-GTCAA-GGACCGA  1571
TAL.CM.266                --AT-GCA---C-A--C......T-CTCA.....................GC------C-----.----------TCTTTT-TAGG...............GA-A-TCCC-CC...-GTCA-AGGACCGA  2076
MUS.CM.1085               --TTGGCA-G-C-A---......-GG...........TCTGATAATACC-A-GC--TG-----.-------G-TCCCTT-G-.........................................  1792
SAB.SN.-.SAB1C            -T--G------------......CAA.......................-A-----T------.-------TTTGGCCC----GG...............-AGAGG-AA-CC...CAGGAATTTCCC--  2436
GRV.ET.x.GRI_677          -TATGGCA-----C--C......-AG.....................A-TG------A-----.-------T-TGGCCAT---GG...............AGGAGC-AA-CC...AAG-AATTT-G-G-  2241
TAN.UG.x.TAN1             --TT-GCA--------C......CGC.....................AGTG-------C-----.--------GA--GCCAACCTGGGG............AACA---CCGC-...CAATTTTTTAGAG-  2279
VER.DE.-.AGM3             -TTT-GCA--------C......-GA.......................G-----TA-----.-------T-TGGCCGG---AT...............GGGAGC-AA-CC...CAG-AATTT-CCCG  1808
VER.KE.-.9063             ---TGGCA-----C--C......-GG.......................G-----TG-----.-------T-TGGACGG---A-...............GGGG-CAAA-CC...-AG-AACTT-C-CG  2313
VER.KE.-.AGM155           --TT-GCA-----C--C......-GG.......................G-----TG-----.-------T-TGGCCGG---AT...............GGGG-CAAA-CC...AAG-AATTT-CCCG  2296
VER.KE.-.TYO1             ---T-GCA-----C--C......-GG.......................G-----TG-----.-------T-TGGACGG---AT...............GGGGGCAAA-CC...-AG-AATTT-CCCG  1794
SUN.GA.98.L14             -TA--CAAGCTC-A--C......C-ACAA...................-GA-A-----TG-----.-------T-TGGCCC----AA...............-AGGGG-CCGCC...AGGGAACTT-CCAG  2437
LST.CD.88.447             -TAGGT-TGCTC-A---......C-A.....AAG.........CCTA--G--A---TA-----.------CT-TGGACC----AG...............GAGTGGTCCTCC...AGG-AACTT-CC-T  1354
LST.CD.88.485             -TAGGT-TGCTC-A---......C-A.....AAG.........CCTA--G--A---TC-----.------CT-TGGACC----AG...............GA-TGGTCCTCC...AGG-AACTT-CC-T  1354
LST.CD.88.524             --AG----GCCC-G--C......C-A.....AAA.........CCAA--G-GA---TG-----.-------T-TGGTCC----AG...............GA-TGGACCTTC...AGGCAACTT-CCAA  1348
LST.KE.-.lho7             -TAG-T-TGC-C-G--C......C-A.....AAA.........CCAA-AG--A---TG-----.------AT-TGGTCC----AG...............-AG---ACCTCC...AGGCAATT--CC--  2438
SYK.KE.-.SYK173           ---T-GCA-GG-CC---AGGCAAC-CAAGAGAAATCAAGGCCCTCCA-TTGC------------.---------AGGATGGG--GT............CTC-AGG-G-CCCCCAGCCAACTTCCCAG-GA  2238
SYK.KE.KE51_AY523867      --AT-GCA-GGC-G---AGACAAC-AAGAAAAGGACAAGGCAACCCTCCTCC------C-----.---------AGGGTGGG--AA...............-AGG--ACCCCCAGCAAATTTTCC-G-GA  1875
COL.CM.-.CGU1             --ATGCAA-G-------......C-C....................A-C-AG-----A-----.------..............................................................  1922
```

```
H1B.FR.83.HXB2              TTCA............GAGCAGACC.........AGAGCCAAC.....................AGCCCCACC............AGAAGAGAGCTTCAGGTCTGGGGTAGAGACAAC............  2198
Gag                         L__Q__.__.__.__.__S__R__P__.__.__.__E__P__T__.__.__.__.__.__.__.__A__P__P__.__.__.__.__E__E__S__F__R__S__G__V__E__T__R__.__.__.__.    p6
Pol                         __S__.__.__.__.__E__Q__T__.__.__.__R__A__N__.__.__.__.__.__.__.__S__P__T__.__.__.__.__R__R__E__L__Q__V__W__G__R__D__N__.__.__.__._  Pol TF

H1D.CD.84.84ZR085           ----............---------.........---------.....................---------............--C---AGCT-CGG-T-TG---AGGAGATA-C-............  1718
H1F1.BE.93.VI850            ----............---------.........---------.....................---------............--C----------G---TCA-A-AG----T---............  1535
H1G.SE.93.SE6165            ----............--A---GA-.........---A-----.....................---------............T-C---A---C--G---TC--A-AG----T-G-............  1604
H1H.CF.90.056               -C--............------G--.........---A-----.....................---------............--C----------G---TC--A-AG----TG--............  1548
H1J.SE.93.SE7887            -C--............---------.........---------.....................---------............--C-------C--G--CTC--A-AG----T...............  1515
H1K.CM.96.MP535             ----............---------.........---------.....................---------............--C----------G---TC----AG----T---............  1394
H101_AE.TH.90.CM240         C---............---------.........---------.....................---------............--C---A-A--GGG--ATG----A-----T---............  1757
H102_AG.NG.-.IBNG           C---............---------.........G--A-----.....................---T-----............--C---------TG--ATG----A-----T...............  1720
H103_AB.RU.97.KAL153_2      C---............---------.........-------T-.....................---------............--C---A-A---TG--ATG----A-----T...............  1421
H104_cpx.CY.94.CY032        ----............--A---G--.........---------.....................------G--............C-CG--AT----AGA-AGGAAA-AG--------............  1567
H1N.CM.95.YBF30             CC--............--CA-C-A-.........-AG-AA-GAGCCCAC...............------G--............-CT------T-ATG---T-CA--AG--..................  1771
H1O.BE.87.ANT70             -G--............---ACC-G-.........-C-C---T-.....................---------.........................................................  2232
H1O.CM.91.MVP5180           -G--............--AACA-GT.........GTCC---T-.....................---------.........................................................  2207
CPZ.CM.04.EK505             CC--GAC......AACA-CA--GAA.........---------.....................------T--............-AT------T-ATG---A-CA..................GGAGGA  1752
CPZ.CM.04.LB7               C---GAG............------.........------C--.....................------G--............-AT---------TG---T-CA-CC-A-..................  1738
CPZ.CM.04.MB66              ----GAA...............GAG......ACC---A--..................AACAGCTC-A--...............-AT----------G---TCA-........................  1723
CPZ.MT145                   -G--AGG............---............GA--GA-C-...............AACAGCTC-A--TATGGA...............ATC----G-A-T-CA........................  1731
CPZ.CD.-.ANT                -G--............--AGGAGGAAGTAGT...G--------.....................---T-----............CAT-----T--A......-CA--AG---CA...............  1637
CPZ.CM.-.CAM3               -G--............-G----GAA.........G--A--C--.....................---T--G--............CCT------T-ATG---A-CA--A--CA---GG............  1603
CPZ.CM.98.CAM5              CG--............-G--G-GAA.........G--A--C--.....................---T-----............CATG-----T-ATG---T-CA--A-ACA----T............  1906
CPZ.GAB_X52154              -G--............--A----A-.........---A-----.....................G--------............GAT------T-ATG---ACCA--AG--AGA...............  2262
CPZ.GAB2                    CC--............A--G-AGGT.........G--A-----.....................---T-----............TCC-----C----G---A--CA-AGC-----CA............  1582
CPZ.TZ.-.TAN1               -G--............--A--CC--AGA......GA-AGGG-A............................................................G-CTCAG--..................  1824
CPZ.US.85.CPZUS             -...............-GAA-ACAG......GAAG-----C--......AGCTCCACCAATAGAG-ATTTCGGGTATCAAGA......-----CAG-.................................  2264
GSN.99CM166_AF468659        C-TTCCTGCTGCCAACAGC-CCGG-.........G--...................................................G---GAAGAACCCCAGCA-C-CA-TCTCT-............  1890
GSN.99CM71_AF468658         C-TCCCTGCTGCCATCAGC-CCGG-.........G--...................................................G---GTGGAACCCCAGCA-C-CA-CCTGT-............  1890
H2A.ALI_AF082339            CC-G............AGTTCCG-A.........G-GATT---ACCAAC....................................--C-CCCCCAGCAGA-C-A-CA-CG--CCTGCTGGAGCAGTATAT  2510
H2A.BEN_M30502              CC--............AGC-CCT-A.........G-G--TG-TACCAAC....................................--C-CCTCCGGCAGATC-A-CA-CG--ACTGTTGGAGAGATATAT  2517
H2A.ST_M31113               CC--............A-T-CCG-A.........G-G--TG--ACCAAC....................................--C-CCCCCGA-AGACC-A-TA-AG--CCT-CTAGAGAAGTACAT  1962
H2B.D205_X61240             CC--............AGTGCCT-A.........G-GAGTG--ACCATC...............T--A--C--..................--TGAA-CCAG-A-A--GCAT----C-TCGGGGGGCGAC  2493
H2B.EHO_U27200              AGGC............ACC-CAGGG.........GATAGTGC-ATC..................T--G--C--..................--TGAA-CCAG-ATTC-GCAT----C-TCAGGGTGCGAT  2488
H2G.ABT96_AF208027          CC--............AGT-CCT-A.........G-GT-TG--TCCATC....................................--C-CCACCGA-GGACC-A-CA-----CCT-CTGAAGAATTACAT  1895
H2U.FR.96.12034             CC--............AGT-CCG-A.........G-GA-TG--CCCAAC....................................--CCCCTCCAGCAGAAC-A-CA--G--TCTGCTGACCCCGACAGC  2008

MAC.239_M33262              C---............AGTGCAT-A.........G-G--TG-TGCCAAC...............T--T--C--.........................................................  2437
Gag                         A__R__.__.__.__.__V__H__Q__.__.__.__G__L__M__P__T__.__.__.__.__.__A__P__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.    p6
Pol                         __S__.__.__.__.__S__A__S__.__.__.__G__A__D__A__N__.__.__.__.__.__C__S__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._  Pol TF

SMM.SL.92.SL92B             AGAC............A-CATC-TT.........GACA--GT-AGCACCCCCAGATCC......---AG-C-G.........................................................  1898
SMM.US.x.PGM53              CC--............--TGCCT-A.........G-G--TG--GCCAAC...............T--T--C--AGA......................................................  2369
MNE.US.-.MNE027             CC--............A-TGCAT-A.........G-G--TG--ACCAAC...............T--T--C--.........................................................  1917
RCM.NG.x.NG411              -GAC............TTC-TTGAG...............-C-TAC..................T--T---GG.................................-AG----GA-GGATGGAATCCCCA  1626
RCM.GAB1_AF382829           C-TC............CTTG-CT--.........CTCTG-TC-AGGGATGGAGAGCAATTA........................C--CCCTGCAGAAGA-ATGCTAAAGA-..................  1641
DRL.FAO_AY159321            ---C............TCAGGCCT-................................................GCTGACCCCATC--CTCCTCCA--GCCAGGATACC-CC--GAGGATCCGGCCGAGAG  1723
MND-1.F17                   A--C............................................................G--A-A-GAGGTGACTCCAAC--CTCCACCA--...AGA.........-GAG-AACCTCTGCAGAA  1846
MND-2.5440_AY159322         -C-C............TGC--TG--................................................ATTGACCCCATC--C-CCCCCGA-GCCAGGGTT--A---CC-GG-AGAAAAGATGCT  1720
MND-2.14CG_AF328295         ---C............TGC--TG--................................................ACTGACCCCATC--C-CCCCCGA-GCCAGGGAT--AG--TC--G-AGAGAAGATGTT  2087
MND-2.MND2_AF367411         -C-C............................................................G--GATG--ATTGACCCCAACG-CCCCTCCAA-GCCAGGGAT--A---CC--G-AGAGAGAATGCT  2163
DEB.CM.CM40                 ---T............AGAAG--AT.............................................................................................TGGGAAGTACAT  1944
DEB.CM.CM5                  --TT............-GAAG-GAT.............................................................................................TGGGAAATACAT  1938
MON.CM.L1                   AG-CCAGCGCACCGACTCTGCC-G-.........................................................................................................  1853
MON.NG1                     .......................................................................................................CCCC-GCA---GGGGGCCAAGGAACTT   398
TAL.CM.8023                 GGA-............................................................................................................CTTCC-AGTGGTGCAACC  1593
TAL.CM.266                  GGA-............................................................................................................TTTTC-AGTGGTGCAACC  2098
MUS.CM.1085                 .......................................................................................................CCTCCA...............GAAAAG  1804
SAB.SN.-.SAB1C              -GAC............CTC--TCAG.........-CCAA--G-TCCCCCAATGGA...ACGGGATTA-AGCAG............GCCG---GAAAATT---A--CA-ACAGAC-TC-GACCAGAGGACC  2530
GRV.ET.x.GRI_677            AAT-............C--AG--GA.................................................................................CAC--TTGGTCTGGAACCAACAGC  2278
TAN.UG.x.TAN1               AAGG............CG-AGCCGT...TCCGACG-CT--CC-GATGCC...............T--G-ACGGATTCCC........................CACA-GTTC-C--G-TGCGGGAGC...  2352
VER.DE.-.AGM3               CCGC............T-CTCTTGG.........G-T-GAGC-AAC..................T--G--C--............TCC-CC---TCC.................................  1854
VER.KE.-.9063               C-GC............C-CTCATGG.........--T-GA-C-GAG..................C--G--C--............TCC-CC--CACC.................................  2359
VER.KE.-.AGM155             CAGC............C-CTCTTGG.........G-C-GA-C-AAG..................T--G--C--............TCC-CC--A-AA-TCTA-ACC........................  2351
VER.KE.-.TYO1               CCGC............T-CTCTTGG.........--C-GA-C-GAG..................T--G--T--............TCC-CC----GG--CCA-CCC........................  1849
SUN.GA.98.L14               -CAT............-CCTTC..........................................T--T--G--.....................TC-AGA-GA-CT-AC-TT-GGG-A............  2480
LST.CD.88.447               -GAT............-G-GG-GAA.......................................T--AGG-GTAGTACC.......................................ATCAGCTCCCCC  1394
LST.CD.88.485               -GAT............-G-GG-GAA.......................................T--AGG-GTAGTACC.......................................ATCAGCTCCCCC  1394
LST.CD.88.524               -CAT............-G-AG-GG-.......................................T--AGG-GTGATTCC....................................TT-AGCGCCACCCAT  1391
LST.KE.-.lho7               -A-T............-G-AG--G-.......................................T--AGGGAGAATTCC.......................................ATCAGCTCCCCC  2478
SYK.KE.-.SYK173             GGAGCGAGCCATCTGCTCC-CCT-T.........G--...................................................G--C-T-GAAGATGGGCCATGGCTC---TGGTCAGCACAGAT  2308
SYK.KE.KE51_AY523867        -GAGGGAGACACCAACAGC-CC---...AGCGGAG--.........................................................T-GGCCA-GGCA-CA-C-----TGGGGAAATTATGC  1945
COL.CM.-.CGU1               ................-GCA-CCAG.........................................................................................................  1931
```

```
H1B.FR.83.HXB2          ..................................................................AACTCCCCC.....................TCAGAAGCAGGAGCCGATAGACAAGGAACTGTAT  2241
Gag p6                  __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__P__P__.__.__.__.__.__.__.__Q__K__Q__E__P__I__D__K__E__L__Y_    p6
Pol TF                  _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__N__S__P__.__.__.__.__.__.__.__S__E__A__G__A__D__R__Q__G__T__V__ Pol TF

H1D.CD.84.84ZR085       ................................................................CCCT-C-AGAA.....................A---G-A---A--GAC-A----------------  1763
H1F1.BE.93.VI850        ..................................................................CC-CT-T--.....................GA--C--G--C--AAAGAC-GGG-ACTGTACCC-  1578
H1G.SE.93.SE6165        ................................................................................................C-CCTCC-C-A---A-GAGATG------AA-G-G  1638
H1H.CF.90.056           .....................................................................C---T-.....................--C-----------A-C-GA-GG-CA-GGAACC-  1588
H1J.SE.93.SE7887        .....................................................................C---T-.....................C-C---A----------AG---------------  1555
H1K.CM.96.MP535         ..................................................................CC-CT-T--.....................GAG-C--G--ACCAAAGACA-GG-AC-GAGCCC-  1437
H101_AE.TH.90.CM240     ...................................................GGGGGAAGAGATAACCT-CTTA--.....................GA--C--G--C--AAAGACA-GG-AC-T-CTCC-  1815
H102_AG.NG.-.IBNG       ...............................................................ACCTC-CT-T--.....................A---C--G-ACC-AG-GACA-GGG-CT-TATCC-  1766
H103_AB.RU.97.KAL153_2  ...............................................................AACCC-CT---T.....................GA-AC--G-AC--AA-GACAGGG-AC-G-ATCC-  1467
H104_cpx.CY.94.CY032    .....................................................................CT--T-.....................--T---------A----GG-----------A---  1607
H1N.CM.95.YBF30         .............................................GAAGAGCACACAGGGGAAGGAG-TG-AGGA.....................GA-CC--G--AG-A-AGAGA--TCTCTGTACCCA  1835
H1O.BE.87.ANT70         .............................................GATGGAGGAGGAAGTGAAGGG-CAGGAGAA.....................----G-A---A-AGG-GGCCCG--C--G--C---  2296
H1O.CM.91.MVP5180       .............................................AATGGAGGAGGCAGTGAAGGA-CAAGAGAA.....................-----GT---A--GG-GATC-GG-A--G-----C  2271
CPZ.CM.04.EK505         GAAGACCACACAGGGGACAGAGAGGGAAGAAAAGGAGAAGACAGAGAGCTC.....................................................................TCTGTACCCA  1813
CPZ.CM.04.LB7           ...AACCACCCAGAAACAGGAGGACAGGGAAGAGGACAAGACAGAGAACAA.....................................................................-CTGTACCC-  1796
CPZ.CM.04.MB66          ...GGAGGAGGAACCTGCCCAGAAGGAGGAAGTGAAGAGCGAGGAGACAGAGAACAA...............................................................-CTGTATCC-  1787
CPZ.MT145               .........GGAGGAGAAGAAAGTCAAGGAGGAGGTGAGGGAGGAGAGCAAGCC..................................................................-CCGTC...-  1783
CPZ.CD.-.ANT            .........CAAGAGGACTCAGAAGGGTCTCAAGGGGGAGGA..............................................................................------ACC-  1680
CPZ.CM.-.CAM3           ...AAAGAGACAGGAGGAGGAGAGGGAGAAGGAAAAGGAACC..............................................................................CCTGTATCC-  1652
CPZ.CM.98.CAM5          ...............GGAGAAAAAGAAGGAGGAGAAGGAACC..............................................................................CCTGTATCC-  1943
CPZ.GAB_X52154          ......GAAGAGCCAGGAGAAGAAAGAGGGAGAGAGCAG.................................................................................TCT-TA...-  2302
CPZ.GAB2                GACAGGAAAGAGTCAGAAGGGGGAGGAGAGTACAACAGAAAA..............................................................................---GAGCCC-  1634
CPZ.TZ.-.TAN1           ...............GCAGGAGACAGCACAGACACCAGT......................................................GGTG-CA-CTGCCCCA--AC-G--G-T-ACGA--A-A  1885
CPZ.US.85.CPZUS         .....................GACACAGGAGAAACAGGGAAA......................................................GG-A---G-...............-CCCT-...-  2301
GSN.99CM166_AF468659    .........GACACCAGTAGGGGA.....................ACCCCA.............................................GAGC---GGAC--AA-GAG--G-GCAGGAACAC-  1945
GSN.99CM71_AF468658     .........TCCGCCGGTGGGGGA.....................GCCCCA.............................................GA-C---G--C--AA-GC---A---AGGAACAC-  1945
H2A.ALI_AF082339        GCAGCAGGGGAGAAAGCAAAGAGAGCAGAGAGAGAGGCCATACAAAGAGGTGACGGAGGACTTACTGCAC-T-GA.....................G---GGAG--ACA--ACACA-AG---TGACAG-G  2619
H2A.BEN_M30502          GCAGCAAGGGAGAAAGCAGAGGGAGCAGAGGGAGAGACCATACAAAGAGGTGACGGAGGACTTGCTGCAC-T-GA.....................G----GAG--ACA--TCACAGAG----GACAG-G  2626
H2A.ST_M31113           GCAGCAAGGGAAAAGGCAGAGAGAGCAGAGAGAGAGGCCATACAAAGAAGTGACAGAGGACTTCCTGCAG-T-GA.....................GA-AC-AG--ACA--ATGCAGAG--ACGACAG-G  2071
H2B.D205_X61240         ACCATCTGCGCCCCCTGCAGATCCAGCAGTGGAGATGCTGAAAAGTTACATGCAGATGGGGAGACA-CAGAGAGA.....................GAGCCGAG--AGA--CTACA-GG---TGACAG-G  2602
H2B.EHO_U27200          TCCATCTGCACCCCCTGCAGATCCAGCAGAGGAGATGCTAAAGAACTACATGCAACTAGGGAAGAAGCAGAAGGA.....................GA-C-GAG--AGA--CTACA-GG---TGACAG-G  2597
H2G.ABT96_AF208027      GCARCTGGGGAGGAAGCAAAAGGAGCA........................................................................--GAA-CA-A--CTACA-GG---TGACAG-R  1953
H2U.FR.96.12034         CCCTCCAGCGGACCCAGCAGTGGA.........TCTGCTAAAGAGCTACATGCAACAGGGCAA...G-AA-AAAA..................GGAGA-C-GAG--AGA--ATACA-GG---TGAC-G-G  2108

MAC.239_M33262          ...............AGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAGA.........AAAGCAGAGAGAAAGC-GAG--A----TTACA-GG---TGACAG-G  2543
Gag p6                  __.__.__.__.__.__E__D__P__A__V__D__L__L__K__N__Y__M__Q__L__G__K__Q__Q__R__E__.__.__.__K__Q__R__E__S__R__E__K__P__Y__K__E__V__T__E_    p6
Pol TF                  _.__.__.__.__.__R__G__P__S__C__G__S__A__K__E__L__H__A__V__G__Q__A__A__E__R__.__.__.__K__A__E__R__K__Q__R__E__A__L__Q__G__G__D__R__ Pol TF

SMM.SL.92.SL92B         ....................................GATAGTCAAGGAGTATCTAGAGAAAGCACAG-GGGAAAAGAC..................AAG--G-AGCAG---TTACA-GG---TGAC-G-G  1974
SMM.US.x.PGM53          ..................GGATCCAGCTGTGGATCTACTGAGGAATTACATGAAGATGGGCAGAAAGCAGAGGGA.....................GA-C-GAG--AGA--TTACA-GG---TGAC-G-G  2460
MNE.US.-.MNE027         ...............AGAGGACCCAGCTGTGGATCTGCTAAAAAACTACATGCAGTTGGGCAAACAGCAGAGAGA.....................AAGC--AAG-A----TTACA-GG---TGGCAG-G  2011
RCM.NG.x.NG411          GTACGATCCAGCAGAAGAGATGCTCAGGAAATACCTAGCATTGGGGAGACAGCACAAACAGGAGCAG-GGAAAGA.....................GA-C---G--A-AGT-GGGAGAGCCT-CGA-G--  1735
RCM.GAB1_AF382829       ..............................TTATCTGAGGAGGGCAGGGGAACAAAAGAGACAACAG-GG-AGGA.....................AG---GCA--A--AGAGAG-GAGCAT-T-A-G-A  1720
DRL.FAO_AY159321        GAT.........GCTCCAAAAGTACATGGAGCAGGGAGCTCAGCAGAAGAGACAACAGAGACAACA-CA....................................-C--AA--GG-G-CCCT-CGA-G-G  1808
MND-1.F17               AACTCTGAGCACTTATCAGAAATTAGGGAGAGGGCT............................................................CAG-C--A--AT-AA-GAG--G---AG-GA-G-G  1916
MND-2.5440_AY159322     GTTAGACTATATGAAGAAAGGGCAACAACAGAGGGCAGCAGCAGGAGC................................................C-----AG----AAA--AG-GTCCCT-CGA-GCA  1802
MND-2.14CG_AF328295     ATTAGACTATATGAGAAAGGGGCAACAACAGAGGGCAGCAAAGGAGAC................................................CA--C--G--A--GAC-AG-GTCCCT-CGA-GCA  2169
MND-2.MND2_AF367411     GTTAGATTACATGAAGAAAGGGCAACAGCAGAAAGCGGAAAGCAA...................................................A---G-AA--A-AGA--GG-GTCCAT-CGA-GCA  2242
DEB.CM.CM40             CCCGGAGAGCACAGCAGATCAGGTATACTCGACTACCAGGGTACCAGTACAGGGGCAGGAGAACAACC-C-AGAG........................-G--G-AACAA---GGAGAGG-A-GTACCCG  2050
DEB.CM.CM5              CCCAGAGAGCACAGCAGATCAGGTTTATTCAACAACCAGGGTACCTGTGCAAGGGGAAAGGAATGCTC-C-AGGA........................-G-CACAAC-A-C-GGAGAGG-A-GTACCCA  2044
MON.CM.L1               ...............AGG......GACTGCGGAGCCAGT...CAATCTCAA...CCAAGAGATTGGTT-C--........................GA---C-GG-ACC--A-AG--G-C-AGG-G-G--  1932
MON.NG1                 TCCAGTAACCTCCTTGCAGCCAACAGCCCCAGCGGAACCAACCCARGGGGCYTACAAGGTGAACTTCT-A--T--AGTGGGACCACCAGAGACCCCAGC-CC-ACA-CC----AG--G-C-AGGAA-G--   528
TAL.CM.8023             CACCGCACCGCCGATAGAGCCTCTCACAGCATCATGGCAGGAGGGGACTACCACATGGACCCCTCCG--CAAGGA.....................-GCCGCAGCACCCAA-CA-C-G-C-AGGGA-AG-  1702
TAL.CM.266              ATCGGCACCGCCAATAGAGCCTCTCACAGCATCATGGCAGGAGGGGACTACCACGTGGACCCCACC-G-CAAGAA........................-CCC--A--AA-A-CCACA-CAAGGGA-AG-  2204
MUS.CM.1085             GAGGCCGAGGAACTTCCCCCAGACCTCTCTCTTGCCAACTGCTCCGCCAGAACAGGAGGGGGAATTGGTCAAT-TCTCTCCAGCAGTGGGTCCCGGGG--GG.........--A-C-GG-CAGGAA-AG-  1925
SAB.SN.-.SAB1C          AGGGCCAGACGATCCAGCAACAGCCCTGTTAAAGCAGTATGCTGTTCAGGGGAAACGGCAGAAACAGCAGTGG-A.....................AA-CC-CTC-CCA-AACAGAG-CCCT-CGA-G-G  2639
GRV.ET.x.GRI_677        CCCCCCAATGGAAACAGCTTACGATCCAGCAAAGAAGCTCCTCCAGCAGTATGCAGAGAAGGGACAGCGC-TGAGAGA............GGAGAGAG-AC--ACAAG-AAACAGA-GG--A--GAAGTG  2396
TAN.UG.x.TAN1           ......ATACGACCCAGCCAGGAAACTGCTGGAGCAGTATGCCAA..................AAAGGG...........................AG-TC-ATT-AGAAA-CAGA-GG--A-GGAACTG  2431
VER.DE.-.AGM3           ......ATACGACCCTGCAAAGAAGCTCCTGCAGCAATATGCAGACAAGGGGAAGCAGTTGAGGGA-CAAAGGAA.....................AA-ACCA-CA-CAGT--ATCC-G-TTGGACAG-G  1957
VER.KE.-.9063           ......ATACGACCCAGCAAAGAAGCTCCTGCAGCAGTATGCAGAGAAGGGAAAACAGTTGAGGGA-CAAGGGAA.....................GAGA-CA-CACCAA-A-ATCCAG-CTGGGC-G-A  2462
VER.KE.-.AGM155         ......TTACGACCCAGCAAAGAAGCTCCTGCAGCAGTATGCAGAGAAAGGGAAACAAATGAGAAATCAGAA-AG.....................AA-CCCC-CA-C-AAC-ATCCAG-TTGGAACG-G  2454
VER.KE.-.TYO1           ......ATACGACCCAGCAAAGAAGCTCCTGCAGCAATATGCAGAGAAAGGGAAACAACTGAGGGAGCAAAAGAG.....................GA-TCCA-C--CAAT--ATCCGG-TTGGACCG-G  1952
SUN.GA.98.L14           ...CAGAATGACACCACCCCAGTCAAAGGCAGAGAGAGCTCTGGAGACCTACAGACTGTTAGGACAGGGC-T-AG.....................AGCAC-----A--AG--AG--A-G--GGGAA-G-  2586
LST.CD.88.447           AATGGAGAGGAGCCCAACACAAGCAGAGAGAGCCTTAGA...AACATACAGGAACCTCGGACAACAGCTGAGGAG.....................G--AC----...-ATCCC-C-G--ATGTG--G-C  1497
LST.CD.88.485           AATGGAAAGGAACTCAACAAAAGCAGAGAGAGCCTTAGA...AACATACAGGAACCTCGGACAACA-CTGAGGAA.....................G--AC----...-GTTCC-C-G--ATGTG--G-C  1497
LST.CD.88.524           GGAGAAAGTGCCTCCAACAAGAGCAGAGAGGGCAATAGA...GACATACAGGAACCTAGGACAACA-CTGAGGAA.....................G--AC-A--...AGTTCC-C-G---TGTG--G-C  1494
LST.KE.-.lho7           AATGGAGAGTGCTCCAACAAAAGCAGAGAGAGCACTGGA...GACTTACAGGACTTTAGGACAACAGCTGAAAAG.....................G---C-----C-AGTCCC-C-G--ATGTG--G-C  2584
SYK.KE.-.SYK173         GAGCCAACAAGCACAAGCGAAGGCACAGAACTCTCCCTCGAAGAAACCCCCCACCAACAGGGAAGTGCTCT-G--.....................GA--G--AGCAGTGG--AG--GG--ACCAA-AG-  2417
SYK.KE.KE51_AY523867    TTCACCCCAACAGAAGCACATAGCAATGGCCACCCCGCTACAGCCACAGTCCTCTCCAAAGACAGTC-AGGA--TCTTGGTCCCTGGGAAACCTTCAG-----A-A--AGA---GA-GG--A-GGGACC-  2075
COL.CM.-.CGU1           .....................GGGGGTGGAGCTACAGACAGCCATCTT...CCCCCCGAAGATGAGC-AAGA--T.....................G-CACC......CAG--GG--G----GGGA-AG-  2010
```

```
                            Pol TF \/ Pol Protease p10          Gag p6 end \
H1B.FR.83.HXB2            CCTTTAACTTCCCTCAGGTCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTTTG.....  2366
Gag p6                    _P__L__T__S__L__R__S__L__F__G__N__D__P__S__S__Q__*_
Pol TF                    S__F__N__F__P__Q__V__T__L__W__Q__R__P__L__V__T__I__K__I__G__G__Q__L__K__E__A__L__L__D__T__G__A__D__D__T__V__L__E__E__M__S__L__.__. Protease

H1D.CD.84.84ZR085         --C---G---------AA--------------------G-T--------------------A--G-----A--------------------------------------------C--A-A----.....  1888
H1F1.BE.93.VI850          --C---G---------AA--------------------T-A-----------A--------A--GA-------G-----------------------------------------C--A-A----.....  1703
H1G.SE.93.SE6165          -TA-ATC-C-------AA-------------G--------A------G----A--------A--G----TA-----C--------C--------------C-----------------A-A---A.....  1763
H1H.CF.90.056             --C---G----------A-------------G--------T--T---G----A----A---A--GT---G---------------------------------------------G--A-A----.....  1713
H1J.SE.93.SE7887          ---C------------AA-------------G--------T----------GA-----------G--G-G---------------------------------------------C--AGAC---.....  1680
H1K.CM.96.MP535           ----------------AA--------------------A-T-----------AG-------A--GT---GA-----------------------------------------------A-A----.....  1562
H101_AE.TH.90.CM240       ------GT--------AA----------------------T------G----A-----A--A--G--G--A--------------------------------------------T--A-A----.....  1940
H102_AG.NG.-.IBNG         ----------------AA--------------------T-A--T---G---GA----A---A--G----TA-----C--------C-----------------------------C--A-A---A.....  1891
H103_AB.RU.97.KAL153_2    ----C--T--------AA------------G---------T------G---GA-----A--A--G-----A--------------------------------------------C--A-A----.....  1592
H104_cpx.CY.94.CY032      ----------------AA-------------G--------T-----------AC-------A--GA---G---G-----T--------------------------------------A-A----.....  1732
H1N.CM.95.YBF30           -----------------A--------------------G--A-----G----A------AAAG--G---GA--------T--------------T------------A-------GC-ACAA--A.....  1960
H1O.BE.87.ANT70           --G--TG-C-------AA--C--------G-CA----AA-A--T---GC--GAG-T-----C--C---TGT----T-T-GC-G-----------------C------C--A-CA-C--ACAA---.....  2421
H1O.CM.91.MVP5180         --A--TG-C-------AA--C--------G-CA----AA-A------GC----G-T-----T--T---TGT--G---T--C-G--------G------------------A-TA-C--ACAA--A.....  2396
CPZ.CM.04.EK505           --C-------------AA-------------G------A--C-T---G----A-----A--AG--A----A--------TC-------------T------------A----------ACAG---.....  1938
CPZ.CM.04.LB7             --C-------------AA--T-----------------A-T-----------AG-C-----CT-CT-G--A-G--A---G-----------G--------------GC-------TT-AGAA--A.....  1921
CPZ.CM.04.MB66            -AGC------------AA--T-----------------G-T------G---GA----A---A--GT----A-----A--G--------------------C-----GC----------AGAA---.....  1912
CPZ.MT145                 ----------------AA-----------------------A--C--G----A--T-AA--A---T---TT-----CT-GC----C-----G--T-----C------A-T-----CT-A-A----.....  1908
CPZ.CD.-.ANT              --C-CGTA------G-AA--C-----------A----AA-GA-GGA-G-TCTC---CA---A---AA-TGTC-G--------------------T-----C------G----G-G---TCA----.....  1805
CPZ.CM.-.CAM3             --G-----C-------AA-------------G---------C-GG--G----A----AA-----G----TA--------GC-------------T------------A----TA-TG-ACAAC-A.....  1777
CPZ.CM.98.CAM5            -AA-----C-------AA-------------G---------C-TGG-G----AG---AA-----GT---TA--------TC-G--------G--T-----C------A----TA-TG-ACAA---.....  2068
CPZ.GAB_X52154            --ACC---C-------AA-------------G---------A-C---G----AG---AA---------TGT------T-GC-------------T------------A----GAG---ACAA--A.....  2427
CPZ.GAB2                  -TC-ATC---------AA-------------G---------C--C--G---GAG---CA--A------TGT-----C---C----------G--T-----------------TA-T--ACAAC-A.....  1759
CPZ.TZ.-.TAN1             GGCGGGTTC--------A---A---------G-----AA-GA-G---G----AG-CCA---A---G-CTGTC-------T--------T-----------C-GT--T--TTGTA-C--C-AA--A.....  2010
CPZ.US.85.CPZUS           -AGC------------AA-------------G---------C---G------AG---CA-----GA--GTA--------T-----C--------T------------C----CA-C--ACAAA-T.....  2426
GSN.99CM166_AF468659      -TC-ATC-C-----GTCA--T--------GG-A----AA-GAAGGA-G-CTCC--------A---G-GGT-TCCATGT-------C-----G--G--C--C--CA-CG-CC----CTCAGCAA-T.....  2070
GSN.99CM71_AF468658       -T--ATC-C-----GTCA--T--------GG-A----AA-GAGAGA-G-CTCC--G-----A---G-GGT-TCCATGT-------C-----G--G--C-----CA-TG-CC-----TCAGCAA-T.....  2070
H2A.ALI_AF082339          GAC--GCTGCA-----AT--T--------A--A----AG-A------GCCT-C--T-A---C--G-C-GT---T-T----C----C-----G--T--C--CT--A--G---C--G---AGAG--A.....  2744
H2A.BEN_M30502            GAC--GCTGCA-----AT--T--------A--A----AG-A------GCGT-C--C-A--AT--G-CGGTA----TCT--C----C-----G--T-----CT--A--G---C--G---AGAA--A.....  2751
H2A.ST_M31113             GAC--GCTGCA-----AT--T--------A--A----AG-A------GC-C-TG-T-A---C--G-C-GTA----T-T-GC----C-----G--T--C--CT--A--G---C--GCG-AGAG--A.....  2196
H2B.D205_X61240           GA---GCTGCA-----AT--T--------AG-A----AG-A----A-GC-TGT--C-A---T--GTC-GTA----TAT--C----C-------TT--C--CT--A--G---C--GG--AGAA--A.....  2727
H2B.EHO_U27200            GA---GCTGCA-----AT--T--------AG-A----AG-A----A-GC--CT--T-A---T---TC-GTA----TAT--C----C--------T-----CT--A--G---C--GG--AGAA--A.....  2722
H2G.ABT96_AF208027        GAH--GCTGCA------C--T--------AG------AG-A-----TGC-T-C----AAA-T--G-C-GTA----TA--------C-----G--T-----CT-TA--G---C--G---AGAA--A.....  2078
H2U.FR.96.12034           GA---GCTGCA-----AT--T--------A-------AA-A-----TGC-T-C---A--------C-GTAA----TA---C-GA----C--G--T------T--A--G---C--GG--T-AA--A.....  2233

                                                                  Gag p6 end \
MAC.239_M33262            GA---GCTGCA-----AT--T--------AGGA----AG-A-----TGCTC-T--T-AA--A--G-CTGTA----TAT--C-G--------G--T------T-TA-TG--AC--G---AGAG--A.....  2668
Gag                       _D__L__L__H__L__N__S__L__F__G__G__D__Q__*
Pol TF                    G__F__A__A__P__Q__F__S__L__W__R__R__P__V__V__T__A__H__I__E__G__Q__P__V__E__V__L__L__D__T__G__A__D__D__S__I__V__T__G__I__E__L__.__. Protease

SMM.SL.92.SL92B           GA---GCTGCA-----AT--T-----C--AG-A----AG-A----G-GC-CTA----A---T--G-CCCTA--G-T----C-G--C-----G-----C--CT-CA--G---C--GG--AGAG--A.....  2099
SMM.US.x.PGM53            GA---GCTGCA-----AT--T--------AGGA----AG-A---G-TGCCT-C--T-AA-AA--G-CCGTA----TAT-------------G--T------T--A-TG-G-C--GG--AGAAC-A.....  2585
MNE.US.-.MNE027           GA---GCTGCA-----AT--T--------AG-A----AG-A-----TGCTC-T--T-A---A--G-CTGCA----TA-----G--------G--T------T-TA-TG---C--G---AGAG--A.....  2136
RCM.NG.x.NG411            G-G--GT-C-------AT--T--------A-G-----TG-AACATGGG--G-A----A-------AAGGT---G--CT---------------------G------TA--C--A-TT-AGA---A.....  1860
RCM.GAB1_AF382829         G-C-----C-------AT--G--------A-G-----AACTACA-T-GCTC-A----AA-----GAA-GT---G-TC--------C--------------C------A-T----G---AGAA--A.....  1845
DRL.FAO_AY159321          G---AC-AC--------T--T-----C---TCA----AACTACA-T-G--G--G---AA------AAGGTA--G---T--C----C---------------------A-TC---G-T--GAA--A.....  1933
MND-1.F17                 GA---TCA------G--TA-T------CAAG-A----AA-A-AAGAGG-CTCAG-G-AT--TGTCACT-TAAG------C--------------T--------CA-C--TA-T----GA-A-A-A.....  2041
MND-2.5440_AY159322       G---AC-AC--------T--T--------A-CA----AACTACAGT-G--G-A----A-------AAGGTA--G--A--TC-------T--------C---------A-TA----TT-AGAG--A.....  1927
MND-2.14CG_AF328295       G---AC-AC--------T--T--------A-CA----AACTACAGT-G--G-A----A---A---AAGGTA-----C---C-G--C--T------------------A-CA----TG-AGA---A.....  2294
MND-2.MND2_AF367411       G---AC-AC--------T--T--------A-CA----AACTACAGT-G--G------AA------AA-GT---G--CT--C-G--C--------T------------A-CA----TC-AGA---A.....  2367
DEB.CM.CM40               GA-----TGGGG--G-AA--T--------AG-A----AA-GAC----G--C-C----AA--AAG-AA-GTCC-G-TG---A----------G-----------CA-TA-TTC-----AAGA-A-A.....  2175
DEB.CM.CM5                GA-----TGGGG--G-AA--T--------AG-A----AA-GAC-G--G--T-TG-G-AA--AA--AGGGTCC-G-TG---A----C-----T------------A-TA-TTC---G-AAGA---A.....  2169
MON.CM.L1                 -T--ATC-C---T-AGCA--T--------GG-A----AA-AAGAT-TG-GCT---------A---AAGGT-TCCATG-----G--------T--C--C--C---A-TA--C----TCA-CACG-A.....  2057
MON.NG1                   -T--ATC-C---T-AGCA--C--------GG------AG--AGA---G--G-A--C-CT-----GTCCGT-TCCATGT----G--C--G--R--G--C-----CA-CG-CC----RCAAGACA-A.....   653
TAL.CM.8023               -T--ATC-C--GT-AGCCAGT--------AG------AA-A--A-AGG-GCTC----AA--AAGGGAGGT---G-TA--CC----------G--------C---A-CA--C-----GG-GACA-A.....  1827
TAL.CM.266                -TC-ATC-C--TT-GGCCAGT--------AG------AA-A----A-G-GCT-----AA--A---GAGGT---G-TA--CC----C--------------C---A-TA--C---T-GGAGAAA-A.....  2329
MUS.CM.1085               -TC-ATC-------GTCA--C------CAGG------AG--AGAC--G--G------AA-----GG-GGT-TC-ATG--GC-C--C--C-----------C--CA--A-CC-G--TGCCCA-A-A.....  2050
SAB.SN.-.SAB1C            G---AC-GC-----------T--------AG-A----AA-GAAA--TG-GT-C----A---A--GAA-GTCACT-----------C--------------CT-----A--C-G-GT--AGAA--A.....  2764
GRV.ET.x.GRI_677          GAGGATGT----T-G--C--C--------AGGA----AA-GAAACG-G-C-TC----AA--AACG-C-GT-C----CT-G--------------------C--TA--A-TC------A-GAC---CACTT  2526
TAN.UG.x.TAN1             GAGGACTA----T-G--T--C--------AGGA----AA-GAGT-AG--CT-C---CA---C--G-C-GTAC---T---G--------------T--------CA--A--A-G---GAAGA-A-CCATTT  2561
VER.DE.-.AGM3             GGA-ATT---TGAA-TC...C--------AG-A----AA-AAAA---G-TT-C----AA---GTC-CC-TCAG---AT----------G--G-----------CA-TA--A-----GCAGA---A.....  2079
VER.KE.-.9063             GGA-ATT---TGAA-TC...C--------AG-A----AA-AAAA---G-C-TT--T-AA--AGTC-CTGTAA---TC--G-----------G--------C--CA-CA--A------CAGACA-A.....  2584
VER.KE.-.AGM155           GGA-ATT---TGAA-TC...C--------AG-A----AA-AAGG--CTGT-TA-----A--AACTGCCGTA--G--AT-------------G--------C--TA--A--A-G--T-CAGA---A.....  2576
VER.KE.-.TYO1             GGA-ATT---TGAA-TC...C--------AG-A----AA-AAAG---G-GT-T----AA---GTC-CC-TTA-G--A--GC----C-----G--------C--CA--A-TA------ATGA---A.....  2074
SUN.GA.98.L14             -AGGAGC-A-G-T-G-ATATG------CCAG-A-----T-A-AG-AGG-CTTC--TAAT-------CTGT-ACT---T----G--C--G-----------C---A--G-GTCT---GATTC-G--.....  2711
LST.CD.88.447             GAGCC-TGCCTGAGTTTC-TCT-----CCAG------AA-A-AA-AC--TCT-G-CAAT-----G-C-GT-TCT-----TC-T--C--T--G-----------TA-C--TTC-----ATTCAG-T.....  1622
LST.CD.88.485             GAGCC-TGCCTGAGTTTC-TCT-----CCAG------AA-A-AA-AC--TCT-G-CAATA----G-C-GT-TCT-----TC-T--C--TA-G-----------CA-C--TTC-----ATTCAG-T.....  1622
LST.CD.88.524             GAGCCTTGCCTGAATTTC-TCT-----CCAG------AA-A-AG-A-G--CT-G-CAAT------GA-GTAACT--C-----G--C--T---------------A-C--TTCT----ATTCAG-T.....  1619
LST.KE.-.lho7             GAGCCCTGCCTGAATTCACTCT-...-CCAG------AA-A-AA-GT---TT-G-CAAT-------C-GT-TC---A--G--G--C--T--G--T-----C---A-T--TTC-----ATTCAG-A.....  2706
SYK.KE.-.SYK173           -TC-ATC-C--T-----C--T--------AG-A----AA-GA-AGA-G--G-TG---AA--TG-C--GGT-C--ATGT--G----------G------------TA--A-TAG---GAAGA-A-T.....  2542
SYK.KE.KE51_AY523867      -T--ATC-C-----TCAA--T--------AG------AA-GA-GGA------TG---AA---ACCAAGGTAAC-ATGT-GC-G--C-----G--------C--CA--A--AG----CAAGACA-C.....  2200
COL.CM.-.CGU1             -TC-ATC-C---T-G-A---T-----C--AG------AA-A--T-T-G--G-AG---AA--A---AAGTGCC----G--G--------------------------T--TA-----GAAGA-G-T.....  2135
```

```
H1B.FR.83.HXB2              .......CCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATA...............GGTACA  2474
Pol                         __.__.__P__G__R__W__K__P__K__M__I__G__G__I__G__G__F__I__K__V__R__Q__Y__D__Q__I__L__I__E__I__C__G__H__K__A__I__.__.__.__.__.__G__T_ Protease

H1D.CD.84.84ZR085           .......-------AG-----------------------------------------------------------C--------------------------G------...............------  1996
H1F1.BE.93.VI850            .......-------A-----------------------------------------------A----------A-C--------------T--------C--G------...............------  1811
H1G.SE.93.SE6165            .......-----------------------------------------------------------A--------AG---CT--------TA-----A-A--G------...............--G--G  1871
H1H.CF.90.056               .......--G----A---------------------------------------------------------G--AG--GC----------------A-A--G------...............------  1821
H1J.SE.93.SE7887            .......---C---A-------------------------------------------------------A-CG--G---CG--------TGAG---A-A--G------...............------  1788
H1K.CM.96.MP535             .......-------A---------------------------------------------------A--------AG----T--------T--------A--G------...............------  1670
H101_AE.TH.90.CM240         .......-------A------------------------------------------G----AG--A--------------T---------------A-A--G------...............------  2048
H102_AG.NG.-.IBNG           .......-------A---------------------------------------------------A--------A-----T--------T------A-A--G--C---...............------  1999
H103_AB.RU.97.KAL153_2      .......-------A------------------------G-----------------G-----------------------T--------T------A-A--G------...............-----G  1700
H104_cpx.CY.94.CY032        .......-------A------G--------------------C-----------------------A-------------CT--------T------A-A--G--C---...............--C---  1840
H1N.CM.95.YBF30             .......GAG----A---------------------A-----------A-----------G-----A------A-T---ACAG----C--ACAG---AGA-----AG-T...............------  2068
H1O.BE.87.ANT70             .......GA-----A------------------------T--A-----------A-------A-G-A------A-TG-GACAG-------AGAA---AGGG-G-TACAG...............--A---  2529
H1O.CM.91.MVP5180           .......GA-----------C------------------T--A-----C-----A-------A-G-----A-CA-TG-GACAG-----G-ACAA---A-GG---TACAG...............--A---  2504
CPZ.CM.04.EK505             .......GA-----A---------------------------------G-------------A----------A-TG-GA-A-----G--ACAA--GA-A-----AG-T...............--C---  2046
CPZ.CM.04.LB7               .......GGG-----G-----------G--------A-------------------------AG--A------A-T----CA--------AG-A---A-A--G--AC--...............------  2029
CPZ.CM.04.MB66              .......GG---T--------------------------------------C--A-----------A------A-TG-TACA-----C--A------A-A-G---AG--...............------  2020
CPZ.MT145                   .......-------A-----GT--------------------------------T-------A---A-T---AAGAG--AA---------AGAA---A-A--GATA-CC...............A----T  2016
CPZ.CD.-.ANT                .......-A-----AC-----G-------CA--T--T--------G--------TTCC---CA-------A-CA-AG---C----C----TG-A-ACAGG-C--TAC--...............-CA---  1913
CPZ.CM.-.CAM3               .......A------A-----G---------------------------------------T-A---A-----CA-C---GCA--------TGAG---AGA---A-A-C-...............--C--G  1885
CPZ.CM.98.CAM5              .......A------AG----G-----G---------------------------T-----T-A---A------A-C---G-A--------AGAG--CAGA---A-A-C-...............--C---  2176
CPZ.GAB_X52154              .......-A----CTT--------------------A-----------------------C-A---A-T----A-TG---A------G--AGAA--GAGA----TAG--...............--G---  2535
CPZ.GAB2                    .......GA---G-A-----G-------T-A-----A--C--------A-----T-------A---A------A-TG--GAA--------AGAG---A-A-G--TAGA-...............A-C---  1867
CPZ.TZ.-.TAN1               .......AAG---CAG----C--------CC-----A-----A-----A---G-ACC---T--TG----CT--A-T--T-CAG--C----TG-CAATA-AG---TC-G-...............-CC--T  2118
CPZ.US.85.CPZUS             .......GA---G-C-----G------------G--------------------A-------A---A--------CG-CAAT--------AGAG---AGA-----ACAG...............---T--  2534
GSN.99CM166_AF468659        .GAATTAGATAC-CC-----C----------G----------A--G-----A--AC-------GG--C--A-A--TG--GAAG--ATCT-TAA--AGA-A-G-ATC-A-...............-C---T  2184
GSN.99CM71_AF468658         .GAATTAGACAC-CC-----C--------C-G-------G--A--G-----A--AC------AGG-AC--AGA--TG--GAAG-GATCT-TAA--A-A-A-G-ATC-A-...............-CC--T  2184
H2A.ALI_AF082339            .......GGGA-C-AC-AT-CC--------AG----------A--G--A--C--A--TACC-A-G-A-----AG-TG--GAA---A--G-ACTAAATA-A-G--TA-A-...............-CC--C  2852
H2A.BEN_M30502              .......GGG-AC-AT-AC-CT--------AG----------A--G--A-----A--CACC-A-G-A--CA-AA-TG--GAA---A--G-ACTAAATA-A-G--TA-G-...............-CC--C  2859
H2A.ST_M31113               .......GGGA-C-AT-AT-GT-----G--AG----------A--G--A--C--A--TACC-A-G-A---A-AA-TG--GAA---AG-G-A-TAAATA-A-G--TA-G-...............-CC--C  2304
H2B.D205_X61240             .......GGTA-C-AT-AC-CC--------AG----A--G--A-----G--C--A--TACC-A-G-A--CA-AG-TG--GAA------G-AGTG---A-A-G--TA-GG...............-CA--T  2835
H2B.EHO_U27200              .......GGCA-C-AT-AC-CC--------AG----T--G--A-----A-----A--TACC-ATG-A--CA-AA-TG--GAA------G-AGTA---A-A-G--TA-G-...............-CA---  2830
H2G.ABT96_AF208027          .......GG-A-T-AT-AT-GT--------AG----A-----A-----A-----T--TACC-A-G-A--CA-AA-TG--GAA--TA--G-GGTA---A-A-G-ATACAG...............-C---T  2186
H2U.FR.96.12034             .......GA-CC-TAT-AT-CT-------CTG------A---A-A---G--C--T--TAC--A-A-A-T-C-CA-TG--AA----A--G-A-TG--CA-A-C-AT--G-...............-----C  2341

MAC.239_M33262              .......GGTCC-CAT-AT-CC--------AG----A-----A-----------T--TACT-A-G-A--CA-AA-TG--GAA------G-T-TA--CA-A-GGAT--A-...............--G---  2776
Pol                         __.__.__G__P__H__Y__T__P__K__I__V__G__G__I__G__G__F__I__N__T__K__E__Y__K__N__V__E__I__E__V__L__G__K__R__I__K__.__.__.__.__.__G__T Protease

SMM.SL.92.SL92B             .......GGTCC-G-G-AC-C-----G---TG-------G--------A--C--A--TAC--A-G-A---AGAACAG--AAA------G-T-TA---A-GGT-ATA-G-...............-AA---  2207
SMM.US.x.PGM53              .......GGTCC--AT-AT-CC--T-----AG----A-----A-----C--C--T--CACC-A-G-----A-AG-TG--AAA---A--G---TA--CA-GGT-AT--AG...............--A---  2693
MNE.US.-.MNE027             .......GGTCC-CAT-AT-CC--------AG----A-----A-----------T--TACT-A-G-A--CA-AA-TG--GAA------G-T-TA--CA-A-GGAT--A-...............--G---  2244
RCM.NG.x.NG411              .......GG--A--AT-----G--------T-----A-----A--G--A-A---A--T--T-A---A-T-A-C--TTGTAAA---ACT--AGCA--GA-G-CTA-ACAT...............-C-CAT  1968
RCM.GAB1_AF382829           .......GG-AATGAT----CC--------A-----A-----A--G--A-A---T--T----A---A---A-AA-CTGTGAA--T-----AGC----A-A-GGA--CAT...............-C-CAT  1953
DRL.FAO_AY159321            .......A------A------G---C-G--T-----A-----------AAC---T--T--G-----A---TT-A-TTGTGAA--TA-G--AGCA--GA-A-TTA--CAT...............-C-T-G  2041
MND-1.F17                   .AAATTAAA-----AT---C-G--------T-----------A--T--AAACT-A-G-----A----------A-TG--TATG-------AA-A--GA-GGG-A-AT-T...............--G---  2155
MND-2.5440_AY159322         .......A------AT-----G---CG---T-----A-----------AGCA----G-------------TT-A-TTGTAAAG-------AGCA---A-A-TTA--CAT...............-C-G--  2035
MND-2.14CG_AF328295         .......A------AT--------CC----T-----A-----------AGCA--T-GG--G-A-------TTCA-TTGTAAA---AC-G-AGCA---A-A-GTA--CAT...............-C-T--  2402
MND-2.MND2_AF367411         .......AA---T-AT---------C-G--T--T--A-----------A-CA-----T----A-----T-TTCA-TTGTAAAG--AC---AGCA--CA-A-CTA-ACAC...............-C-T--  2475
DEB.CM.CM40                 .GATTTAGAG-C-CCT----CC-------CAG--------T-A--------C--A--T--C-A-TGT--CCCAGG----GAA--T-C---GGC-AATA-GGT---AG--...............TCAGAT  2289
DEB.CM.CM5                  .GATTTAGAT-C-CCT----CC--G----CAG-G------T-A--G--A--------T----A-TGC--CCCAGG----GAG--TA----GGCA-ACA-GGT----G-T...............TCAGAT  2283
MON.CM.L1                   .AATTTAGGG-AG-AT----C---T---G-AG----------A--G--AA-G--AG----T-A---A--CA-AGCA----AAG-GAT-T-TGAA-ACAGAG-G-T-TGG...............-CA---  2171
MON.NG1                     .AATCTAGG--A--AT----C---C------G----------A-----C--A--AC-G--T-AG--A-TCA-AGG-G--ACGG--TCCT--CAG--TA-A-C--TG-G-...............-CW---   767
TAL.CM.8023                 .GAATTAGA----GTGCCCC-C-A-----CAG--------T-G-C---A-AC---C-G---CA-AGC---AC-GGAG--GAA--C---TGGGAACACA-A-G-----GG...............-AC---  1941
TAL.CM.266                  .GAATTGGA----GT-CCCC-C-A-----CAG--------T-G-C---A-AC--AC----GCA-AGT---AG-GCC---GAA---C-GTGGGAAAACA-A-G---A-GG...............-AC---  2443
MUS.CM.1085                 .CACCTAGA----CC------T--C---TCAG----------A--G--AAAC--ATC---C-A-G-A---AGGGGAG---AGG-TTGTT--AA----A-A-CTAT--GG...............-CA---  2164
SAB.SN.-.SAB1C              .......GG--AC-AT----------G---A--T--A-----A-----A-G---T--T----A-GCA--CC-CA-TCAGGAGG--A----AGAA-ATA-A-CTTG--A-...............-C---C  2872
GRV.ET.x.GRI_677            TCCCCCA-ATAA-CC----CGTT-C--GG-AG----A--T--A-----AGGG--TC-T--C-A-G-A---C-GGG-G---AAG--C--T-GGAG-ATA-A-TCATC-CC...............--CT--  2641
TAN.UG.x.TAN1               A......GA----C-C---CGG--T---G-AG----A--T--A-----GGGA---TC-----A---A-----CA-T--TT-AG-GA-T--AGA----A-AG--A---G-...............----AT  2670
VER.DE.-.AGM3               .CAATTAT------C---------------A--------C--------GGGAC----T----A-G-----AG-G-T-GGGAAG--AG-T-GGAA-ACA-A-TTTTG-G-...............--G--C  2193
VER.KE.-.9063               .CAACTAA-----CCT----G-------G-A-----A-----------AGGAC-T--T----A-G-----AG-A-TG--GAAG--A----AGAA-ATA-A-TTTTG-G-...............--C---  2698
VER.KE.-.AGM155             .CAATTAAGG---TC-----G---------AG----A--------G--AGGGT-A--C----A-G-A------A-TG--GAAG--C--T-GGAA-ACA-G-T-TTA-G-...............--A---  2690
VER.KE.-.TYO1               .CAATTAT----TCC-----G---C-----T--------C--A-----AGGCC-T--T----A-G-A---A-CG-C-GGGAAG--A----AGAA-ATA-A-TTTTG-G-...............--A---  2188
SUN.GA.98.L14               .TCCATAGA---TCAC-----G--C-GGG-AG-G--T--T--A--G--ACAA--T-G-----A-G-A----CAG-TG-CT-TG-------TA-G-ATA-A-C-----C-...............--A---  2825
LST.CD.88.447               .AGAATAGAG---CCT-AT-C---T-G-G-AG----------A--G--ACAA--T-G-----A-G-A--CA-GG-TG-CT-T--------AGCA---A-A-C-A-AG--...............--G---  1736
LST.CD.88.485               .AAAATAGAG---CCT-AT-C---T-G-G-AG----------A--G--ACAA--T-G-----A-G-A--CA-GG-TG-CT-T--------AGCA--GA-G-C-A-AG--...............--G---  1736
LST.CD.88.524               .AGAATAGA----CCT-AT-CT----G-G-AG-------T--A--G--ACAA--T-------A-G-A-TCAGGG-TG-TT-T---C----AGCA--GA-G-CCA-AG--...............--G---  1733
LST.KE.-.lho7               .AGAATAGA----CCC-AT-CT----G-G-AG-------C--A--G--GCAA--T-GG----A-G-A---AGAG-TG-CT-T--------AGCA---A-A-CTA-AG--...............--G---  2820
SYK.KE.-.SYK173             .CAATTA-ACCA-CC----TC---------AG----A---T-A-----AAAC--T-C---T--G--A---AGAA-T--TAGAT-CACTG-TGT-AA--CCTC--GA-A-AGAAAGCAAGTAGAG--G--T  2671
SYK.KE.KE51_AY523867        .CAGTTA-ATCA-CCT----CC--------AG--------T-A-----AAAC--A-GG--------A---A-C-----CAGAT-TC----AGAAA---CAG-T-GG-G-ATAAAGGAAGTAGAG----GT  2329
COL.CM.-.CGU1               .CAATTAA----CGCT----C-G--GT-CAA--TCA------A--G--AGCA----G-----A-------AGAG-T-A-A-TC--A-G--GG-A---A-AG--TAC-G-...............--AGAT  2249
```

```
                                                                                  Pol p10 Protease end \/ Pol p51 and p66 RT and RT-RNaseH start
H1B.FR.83.HXB2             GTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTTCCCATTAGCCCTATTGAGACTGTA.........CCAGTAAAATTAAAGCCAGGAA  2595
Pol                        _V__L__V__G__P__T__P__V__N__I__I__G__R__N__L__L__T__Q__I__G__C__T__L__N__F__P__I__S__P__I__E__T__V__.__.__.__P__V__K__L__K__P__G__    RT

H1D.CD.84.84ZR085          --------------C---------------------------T-------C--------C-----------C-----A-----T--------A------.........----------------------  2117
H1F1.BE.93.VI850           --G--------------G------------------------A-------------------T--------------AG----T--------A------.........-----------G----------  1932
H1G.SE.93.SE6165           A--------------------A-------------G------A----------------------------------A-----T--------A------.........----------------------  1992
H1H.CF.90.056              --------------------------T-----------G---A-A--------A-----------C-----------A-----T--------A------.........----------------------  1942
H1J.SE.93.SE7887           --G---A-------------------------------G--CA-----------C-------T--------------A-----T--------A------.........----------------------  1909
H1K.CM.96.MP535            ------------------------------------C----C--------------------T--------------A-----T--------A------.........----------------------  1791
H101_AE.TH.90.CM240        ------------------------------------C-----A-------------------T-----------C--A-----T--------C------.........-------C--------------  2169
H102_AG.NG.-.IBNG          ------------------------------------C-----A-------------------T--------------A-----T--------A-----G.........----------------------  2120
H103_AB.RU.97.KAL153_2     -----------------C------------------------A-----------C-------T--------------A--A--T--------A------.........-------C--------------  1821
H104_cpx.CY.94.CY032       --G---------------------------------C----CA-----------C-------T--------------A-----T--------A------.........----------------------  1961
H1N.CM.95.YBF30            --------------A--------T--T--T--A-----------T--A--C--------C--T--------------A--A--T--------A------.........--------------A-------  2189
H1O.BE.87.ANT70            -----G--G--------T-----T--T--T------------A-A-----AGGAT-A-----T--AC----C--C--T--A-----C--A-CCC-A--G.........---------C----A-------  2650
H1O.CM.91.MVP5180          -----G--G--------T-----T--T--TC----G-----CA-A-----AGGAT-A--A--T--AC-------C--T--A--T--C--A-CCC-A--G.........--------GC----A-------  2625
CPZ.CM.04.EK505            --C--------G--A--------T--T--T------------T-TC----A--A--------T---C----------A--A--T--------A--CA--.........-----G-----G--------T-  2167
CPZ.CM.04.LB7              --T-----------C-----A--A--T--C--A-----G--CA-T--A-----A-----A--T---C----------T--------------A------.........------C---------------  2150
CPZ.CM.04.MB66             --T-----------C--------A--T--C--A-----G---A-T--------A-----A--------------C-----C--T--------A------.........-------GT-------------  2141
CPZ.MT145                  --------------A--C--------T-----A---------G-CC-A--AA-AT-A--G--T---C-------C--A--A--T--A--A--C--A---.........AA---GTTTC----------G-  2137
CPZ.CD.-.ANT               ---C-GT-G-----A-AC--A--A-----T--A--T------G-T--ATG--TTT-A--A--T--------------A-----TAAAG----A--A--G.........--C--T--------AGA-----  2034
CPZ.CM.-.CAM3              -----G--G-----A-----A--G--T--------C------A-A-----A--A-----C--T--------------T--A--T--C-----A--A--G.........----------------------  2006
CPZ.CM.98.CAM5             --------------A-----A--A--T--------C------A-A--A--A--A--------T---C----------T--A--T--CG-------A--G.........----------------------  2297
CPZ.GAB_X52154             --------------A--------A--T---------------A-T-----A--AT-G-----T---C--GTG--C--A-----TT-A-----A-----C.........-----C-----G--A-------  2656
CPZ.GAB2                   ---C----G-----A--T--A--T--T--T--A---------G----A--C--AT-A--A--T---C----------A--A--T--A---ACTC----G.........----------------------  1988
CPZ.TZ.-.TAN1              --CC----G---GAA--C--CA-T--T-----A--T------A-T--A-AG--AT-A--A--T--C-----------T--------A--A---GTA---.........AA----C-------AGA-----  2239
CPZ.US.85.CPZUS            --T-----G-----A-----A--A--T--T-----C--G---A-T--A--A--A-----A--T--------------T--C---T-----A-A------.........-------G------A-------  2655
GSN.99CM166_AF468659       ---C-G--T-----C-----A--A--T---T-A---------TGC--A-G-A--T-A--GGTA--------CA-GGTGCAGCAGAAG--A--AC-CA--.........GA---GC-CC-----G------  2305
GSN.99CM71_AF468658        ---C-T-----G--C-----A--G--T---T-G---------TGT--A-GCA--T-A--AGTA--C-----CA-GGTGCAGCAGAAAC-A---C-C---.........GA---GC-CC----AT------  2305
H2A.ALI_AF082339           A--A-GAC---TGAC--C--AA----T--TT----C-----CA-T-----AGCCT-A--CATGT-A-----CC-A--AG--GC-AAG--A---C-AA--.........GAG----G--------------  2973
H2A.BEN_M30502             A--A-GAC----GA---C--AA-------CT----C------A-TC----AGCCT-A--CATGT-A--------A--AG--GC-AAG--A---C-AA--.........AA-----C---G--------G-  2980
H2A.ST_M31113              A--A-GAC---TGA---C--AA-------TT----C-----CA-TC----AGCCT-A--CATGT-A------C-A--AG-CGC-AAG--A--AC-AA--.........AA-A---TGC-G----------  2425
H2B.D205_X61240            A--A-GAC----GA---C--AA-A-----TT----C------A-T--A-A-ACCT-G--CATG-----------C--AG-GGCAAAGG-A--AC-A---.........AA---TG-G-----A--T----  2956
H2B.EHO_U27200             ---A-GAC---GGAC--C--AA-A-----TT----C------A-T--A-A-AGCT-A--CATG---C-------C--AG-AGCAAGG--A--AC-A---.........AA---CC-G--------T-A--  2951
H2G.ABT96_AF208027         --TA-GAC---GGA---C---A-A--T--CT-------G--CA-T--AGT-A--C-A--CATGT--C----------A---GCAAAGG-A---C-A---.........AA----GGC-----A--T--C-  2307
H2U.FR.96.12034            A--A-GAC----GAC--C--AA----T--TT----T--G-----T-----AG-AC-G--CATGT--C----C-----A--AG-AAAAGAA-CACTA--C.........ATT-GGGG-C----A-----C-  2462

MAC.239_M33262             A-CA-GAC---GGAC--C--GA-T-----TT----T------T--C-A--AGCTC-G--GATGT--C-------------AGCTAAAG-A---C-----.........AA---CGCC-------------  2897
Pol                        _I__M__T__G__D__T__P__I__N__I__F__G__R__N__L__L__T__A__L__G__M__S__L__N__F__P__I__A__K__V__E__P__V__.__.__.__K__V__A__L__K__P__G_     RT

SMM.SL.92.SL92B            C--A-GAC----GA---C--GA-A--T--TT----C------A-TC-AG-AGCAT-A--GGT-T---------AC--AG-GGCAAAGG-A--ATA-AC-.........AA------G--G---GAG----  2328
SMM.US.x.PGM53             A-TA-GAC----GA---C--AA-T-----TT----C-----------A--AGCT--G--CATGT-C------C-C---G-AGCTAAGG-G---C--A--.........AA----C-C-----A-------  2814
MNE.US.-.MNE027            A-CA-GAC---GGAC--C--GA-T-----TT----T------T--C-A--AGCTC-G--AATGT--C-------C-----AGCTAAGG-A---C-----.........AA---C-CC-------------  2365
RCM.NG.x.NG411             -----G--------A-----A--A--T--------C--G---G------GCA-AT-A--AGC-T--C----------A--A--TAAAGCA--A--AA--.........AAG---G--------------C  2089
RCM.GAB1_AF382829          --GC----G-----A-----A--G--T--T--A--G------G-T--A-AGA-AT-A--AGC---AC----------A--A----AA--A--A---A--.........AAG---G--------T-T---C  2074
DRL.FAO_AY159321           --G--G-----C-----C-----A--T--T--A--T------G-AC-CTG-A-AT-A--AGCT---C-T--------TG----T-AAG-A--AC-A---.........AA----C-C-----A-------  2162
MND-1.F17                  -----GA----------T--AA-AG-T-----A--G-----CA-AA--GAAA-AT-A--AG-A-AA----TA--GG-ACAATTGT--GA-A-A-TACC-...ATAACAAA---G--------A------G  2282
MND-2.5440_AY159322        --TC-G--G--C-----C-----A--T--T--A--T------G-AC-T-AGA--T-A--A--T--------------T-----TAAG--A-----A---.........AA----C-C-----A-------  2156
MND-2.14CG_AF328295        --GC----G--C-----C-----A--T--T--A--T------G-TC-T-AGA--T-A--A--T---C-G--------T-----TAAG--A-----A---.........AA----C-C-----A-------  2523
MND-2.MND2_AF367411        --CC----G--C--C--------A--T--TG-A--T------G--C---AGA-AT-A--A--T--AC----------AG----TAAAG-A--A--A---.........AAG---C-C-----A-------  2596
DEB.CM.CM40                C-----------GA---C--AA-T--T--TC-A---------TAT--AG-AA-A--G--GGTAT--A----------AG-ATTAGA-TA-ACCC-A---.........AA-T--......--AGAG--GG  2404
DEB.CM.CM5                 --G--------TGA---C--AA--------T-G-----G---TAT---G-AA----G--GGTAT-CA----------AG-ATTGGACT--ACAC-A---.........AAGT-G......---GAG---G  2398
MON.CM.L1                  --------T-----A--T---A-A--T--TT-A--G------G-C--AG--A----G--GGTA--CC-C--CA-GG-TGC-G-GGGAGACCTCCAACCCCCATCTTTAGAG--C-CT-----AG------  2301
MON.NG1                    --T--GC-G-----R--C--AA-T--T--TT-G--G-----CG-CC-CT-AAG---G--GATA--CC-G--CA-GY--CAGG-AATCGCCC-AT-CCC-...TCCTTGA---G-CCC--T--A-------   894
TAL.CM.8023                ------A----G--------AA--------T-A---------T-CC-AG-CA-AT----GGTA--AC-T---A-GGT--GC-CAGAA--ACCAGTAA--.........AA------G--G---GA-----  2062
TAL.CM.266                 --T--GA-T-----A--G--AA-T-----TC-A--G------T-T--AG-AA-AT--A--GTT--AC-T--CA-GGT--GCTCAGAC--ACC--TAACC.........AAG--T--------AGAG--T-  2564
MUS.CM.1085                -----------G--A------A-A--T-----G--G-----CTGT--A-GCA-AT----CATT---C----CA-GATACAGG-AAAAT-A--AC-AA--.........AA---G-CCC----AGA----G  2285
SAB.SN.-.SAB1C             A-T--G------GAA-----A--T--T--T--A---------G-C--AG-A--AT-A--AGTT---C-------GA-ACAA--AGAG--A--AC--A--.........AA---CC-C------------C  2993
GRV.ET.x.GRI_677           A-TC--A-----AG------AA----T--T--A---------A-T--AG-----GCA--CATG-AA---GT-A-GGGAG--CTAT---G-C---T--AG...GAAACAAA----C--C----AGA---G-  2768
TAN.UG.x.TAN1              ---C--A-T-----A-----AA-A--T--TG-A---------A-TC-AG-A-CTT-A--AGCA-GA----TAA-GGGACAGCTAAA-GAGA-A-TACCC...ATAACCAAG---C----G--A--T--G-  2797
VER.DE.-.AGM3              A----GA-----AGC--T--CA-A------------------A-A--AG-A-CAGCA--AGC--AA---GTAA-GGGTCAACTGT-AGAAC-A-T-CCC...ATTACC--T--G--------AGA---GG  2320
VER.KE.-.9063              --TC-TT-----G-A--C---A----T---------------T-AC-AG-A-CAGCA--AGCT-GAC--GTGA-GGGACAATTGT-AGAAC-A-T-CCC...ATCACC-----C---C-G--AGAT---G  2825
VER.KE.-.AGM155            --CC-CA-----G-A--T--CA----T--C--A--------CT-T--AG-C---GCA--AGC--AA---GTGA-GGGGCAATTGT-GCAGACA-TACC-...ATCACC--G---CGC------GA---GG  2817
VER.KE.-.TYO1              A----GT-----G-A--T--CA-T--T-----A--T------T--C--G-C-C-GCA-TGCC-CGG---GTAA-GGGACAATTAT-AGAAA-A-T-CCT...GTCACA--T--C----G---GA----GG  2315
SUN.GA.98.L14              --G--G--G--G--A-----A--AG-T--C--A--G------A-TC-TT-AGCCT-A--GG-A-GA---GT---AG---C-CTTAG-GAAA---TCCC-...ATTACAAA---T--G-----AGA---GG  2952
LST.CD.88.447              ------T-G-----------A--AG----TG-A---------A-T--A---A-TG-G--AGCA-GA---GTAC-AT-ACAATTAAG-GACA-A-TACCC...ATCACAAAG--C-----G--------CT  1863
LST.CD.88.485              ---C--T-------------A--AG----TG-A---------A-T--A---AGTG-G--AGCA-AA---GTAC-AG-ACAATTAAG-GACA-A-TACCT...ATCACAAAG--C-G---G--------CT  1863
LST.CD.88.524              --T--GC-------A--C--A--AG-T---G-A--------CA-T--A-AC-CTC-A--GGCT-GA---GTG--AG-ACAATTAAG-GAGA-A-TACCT...ATAACAAAG--GC-T--G--------GT  1860
LST.KE.-.lho7              --C---T----------C--A--AG-T--TG-A--G--G---A-T--A---GCA--G--GGC--AA--G-T---AG-TCAGTTAAG-GA-A-A-T-CCT...ATTACAAAG---TCT-----A--T--GT  2947
SYK.KE.-.SYK173            C-T-----G--T-----G--A--A--T---T-A-----G---A-AC-T--CA-AT-G--AGT--AA---GTAA-GGTTCAA-CAG-AT-A---C-A--C.........AA-----GTC-G--A----AC-  2792
SYK.KE.KE51_AY523867       C-TC----------------A--A--T--TT-A--------CA-A----GCA-AT-G--AGCA-AA---GTAA-GATAGCACAAAAAT-A--AC-AA-G.........AA---CC--C----------G-  2450
COL.CM.-.CGU1              A-C---A------AC--C--CA-A--T---G-A--T------A----A-AAG-AGGGAAA-TAGTAG--GC-CAATTATCAGATAGG--CCCAGTCACC.........AAG--G-CG-----AGAG--C-  2370
```

```
H1B.FR.83.HXB2           TGGATGGCCCAAAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACAGAGATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCC  2725
Pol                      M__D__G__P__K__V__K__Q__W__P__L__T__E__E__K__I__K__A__L__V__E__I__C__T__E__M__E__K__E__G__K__I__S__K__I__G__P__E__N__P__Y__N__T__P    RT

H1D.CD.84.84ZR085        ----------------------------G---------------------------AC---------------T--------------A----------G------------------------------  2247
H1F1.BE.93.VI850         -------------G-----------------------------G------------AC------A---CT---A--------A-----A-----------------------------------------  2062
H1G.SE.93.SE6165         ------------GG------------------------------------------AC-----------A---A------G-------A-----------------------------------C-----  2122
H1H.CF.90.056            -------------G------------------------------------------ACG-----------------------A-----A-----C----G---A--------G----------GC-----  2072
H1J.SE.93.SE7887         --------------A-----------------------------------------AC-C--------G----A------G-------A----------G-G-------------------T--C-----  2039
H1K.CM.96.MP535          -------------G-----------------A------------------------AC---------------A--------A-----A-----------------------G--------T--C-----  1921
H101_AE.TH.90.CM240      -------A-----G--------G---------------------------------AC-----------A----------G-------A-----C-----------------------------------  2299
H102_AG.NG.-.IBNG        -------------G------------------------------------------AC---C--------------------------A-----------------A-----------------------  2250
H103_AB.RU.97.KAL153_2   -------------G-----------------A--------G---------------AC---C-------AG-----------------A-----------------------------------------  1951
H104_cpx.CY.94.CY032     -------------G--------------------------------------C---AC---G--A--------C--------------C--G--------------------------------------  2091
H1N.CM.95.YBF30          -------------G--A-----------T------AC----------G-G------AG---------------A--------------A--------T-G---A--------G--------T--C-----  2319
H1O.BE.87.ANT70          -------A--------A-----------CC-AT-TA-----------G----C--GAC--C---A---CAG--A------C-A-----A----------G---A--A-----------T--T-----A--  2780
H1O.CM.91.MVP5180        -------A--------A-----------CC-AT-TAG---G------G-----C--ACT-C---A---CA---A------C-------A-----C----G---A--A-----------T--T-----A--  2755
CPZ.CM.04.EK505          ----------G-G---A-----G-----T--------------------G---C--AC---------------A--------A-----A--------C-G---A-----------------T--------  2297
CPZ.CM.04.LB7            -------G--------A-----------C--A-----G--------T-----T---AC------C--------A------C-------A----------G---A-----------C-----T--------  2280
CPZ.CM.04.MB66           -------G----GG--C-----G--------A----------------G-------AC---------------A--------------A----------G-G-A-----------------------C--  2271
CPZ.MT145                ----C--G--C--G--A-----G-----TC--T--A-----------G-G------ACT-----A----A---A--------------A--------------A-----A--------T--------A--  2267
CPZ.CD.-.ANT             -------A----G---A-----------GC-CT--A----G------G-G--CC--AA----------GATA--T-A---GCA---AAT--G-----T-G---A-----A--T--C-----------A--  2164
CPZ.CM.-.CAM3            ----------T-G---A-----------T--A----C---------T---------AC----------CAG--A--------A-----A--G-------G---A-----C-----------T-----A--  2136
CPZ.CM.98.CAM5           ------------G---A-----------T--A----C---------C--G--C---AC---------CCAG--A-----------G--A-----------G--A--C--A--G--C--T--T--C--A--  2427
CPZ.GAB_X52154           -------T-----G--A--G--------C--AT---C---------T-----C---AC----------CA---A--------------------A--G-----A--A-----------------------  2786
CPZ.GAB2                 CA-----A----GG--A--G------------T--A-------G--G--G------AC-----------GG--A--------A-----A----------G---A--T--------------T--------  2118
CPZ.TZ.-.TAN1            -------G--------A--G--G-----CC-CT-CA-G--G-----TG-G------AC------A----AGACAT-------------A--------TGC-G----A--A-----C-----T--C--A--  2369
CPZ.US.85.CPZUS          ----------T-G---A--G----------------C---------T--G------AC---------CCA---A--------A-----A--G---A---G-G-A-----A--------T--------A--  2785
GSN.99CM166_AF468659     AA--G--A-----GA-C-----G-----CC--T-TAGG-----G--G--G--C---ACCC----AACACA-------CTT-GATT---TC--T-GGA------A--A--A--------------CT-A--  2435
GSN.99CM71_AF468658      AA--A--A-----GA-------------C--T-TA-G-----G--G--G--C---ACCC----AACACA---A---CTC---TT---CC--T-GGA---G--A-----A--------------CT-G--  2435
H2A.ALI_AF082339         AA--C--G------T-A-G---------C--A---A-----------G-G---C--AA------C---GA-A-A-C----GA------CC--T-AGAGG-GGCACCT--AACT--C--T--T-----C--  3103
H2A.BEN_M30502           AA-----A----GGC-G-----------CC-A---A----G------G-----C--AA---G--C---GA-A-A--------A--G--CC-GC-AGA-G-GGCACCT--AACT-----T--T-----C--  3110
H2A.ST_M31113            A------A------C-G-G---------C--A---A-----------G-G---C--AA---G--C---GAGA-A------GA---G--CC-GC-AGAGG-GGCACCT--AACT-----T--T-----C--  2555
H2B.D205_X61240          AA-----G-----GA-C-G---------TC-AT-CAGG-----G---CT---CC-CAA------C---GA-A-A-----------G--AC--T-AGA-G--GCACCC---ACT-----------C--A--  3086
H2B.EHO_U27200           AA-----G------A-C-G---------CC-AT-CA----G------CT---CC-CAA------C---GA-A-A--------A--G--AC-GT-AGA-G-GGCGCCT---ACT------------T-G--  3081
H2G.ABT96_AF208027       -------G------A-C-GG--------CC-AT-TA-G---------C-----C-CAGG---------GACA-A------C--------C--T-GGA-G--GCACCC--CACC-----G--T--C--A--  2437
H2U.FR.96.12034          A------A------T-A-G---G-----T--AT-TA-G------T--G--------AA---G--A---GAGA-A-----------G--AC--T-GGA-G--GC-CCC--CACT-----------C-----  2592

MAC.239_M33262           A------A------T-G--G--G--------AT--A-------G---GTT------AG------C---GA-A-------------T--TC-GT-GGAGG--GC-CCC--GACC-----------C--C--  3027
Pol                      K__D__G__P__K__L__K__Q__W__P__L__S__K__E__K__I__V__A__L__R__E__I__C__E__K__M__E__K__D__G__Q__L__E__E__A__P__P__T__N__P__Y__N__T__P    RT

SMM.SL.92.SL92B          -------G----G-T-A-----------C--AT-TA-G-----G---C--------AC---G--C--CGAGA-A-----------G--AC--T-GGA-CG-GCCCCT---AC------T--T-----A--  2458
SMM.US.x.PGM53           AA--C--A------T-G-G---G-----GC-AT--A-----------GTT------AG------C---GA-A-A--------A--T--CC-GT-AGAGG--GCCCCT--AACC--------T--C--C--  2944
MNE.US.-.MNE027          AA-----A------T-G-GG--G--------AT--A-------G---GTT------AG------C---GA-A-------------T--TC-GT-GGAGG--GC-CCC--GACC-----------C--C--  2495
RCM.NG.x.NG411           AA-----G--T-G---A-----------T----GTA----G-----TG----T---AC---G-------AT-CC-----------G--A--------T-G---A--A--A--------T-----------  2219
RCM.GAB1_AF382829        AA-----A---CG---A---------------T--A-----------G----T---AC---------C-AT-CA-----G--------A--------------A-----A-----C--C-----C--A--  2204
DRL.FAO_AY159321         -------A-----GA-A-----------CC-CT-CA-G--G-----TG----C--GAA------A---GAGC-A-----G--------TC----AAGT-G---A--T--A--------------C--A--  2292
MND-1.F17                -A-----A--C-G-A-A-----------T--A-GTA----G------GTT-GTC-TCAGA----A---GATAGAT-A--GG-------A------AGT-GGG-A-AT--A-G----AAT--------A--  2412
MND-2.5440_AY159322      CA-----A--G-G-A-A-----G------C--T-TA-------G--CCT---C---CA------A--C-ATC-A--------A-----C--G--C--T-G-G-A--T--A--------------C--C--  2286
MND-2.14CG_AF328295      -------A----G-A-------G------C--T-TA-------G--TTT---C---CAGAC---A---GAT--A--------------C-----C--T-G---A--T--A-----------T-----A--  2653
MND-2.MND2_AF367411      CT--C--A--T---A-A---------------T-TA-G-----G--TTT---C---CA------A----GTC-A-----G--------CC-G--C--T--G--A--T--A--------T-----C--A--  2726
DEB.CM.CM40              CA-----A-----GA--------------C-TT--AGG---------G-G--C---AA---T--C-TA-AGCCC---AT-G---C---AC-G--AGTCCCGGCA-C---CACC--C-----T--CT----  2534
DEB.CM.CM5               CT----------GGA-------G-----CC-TT--AG----------G----C---AA---G--CATA-GGCCT----T-G-A-C---AC-G--AGTCCCC-CA-A---CACT--C-----T-GCT----  2528
MON.CM.L1                AA--G--A-----GT-G-----G--------AT-CAG------G---G-G--TC-TAC-C-----AC-CA-------CTT--ATT---AC--T-GGA-CC--CC-A---CA-T--C-----T--CT-A--  2431
MON.NG1                  AA-----G------T-G-----------CC-CT-TA-------G---G-G--TC-CACCACC---RCCCGG------AT-G-ATT---CC--T-GGA-G--GCA-A---AACC--------T---T----  1024
TAL.CM.8023              A---ACC------GA-C--G--------C-----CA-G---------G---GGC-GCAG-CC--CATAGAC--C----TG----C---AC-GT-GGAGG-G--------G-----C--------CT----  2192
TAL.CM.266               A---ACC---C---A-C--G--------C--A--CAG----------G---GGC--CA-ATC--CATAGAC--C----TG----C---AC-GT-AGAGG-GG----A--G-----------T---T----  2694
MUS.CM.1085              CAA-A--G----TG--C--G--------T-----CC-G-----G--C-GG--C--G-AG--C--AACCCAGC-----CT---A-T---CC----AGAGG----A--A--A--G-----T--T---T----  2415
SAB.SN.-.SAB1C           AA-----G----GGA-A-GG--------T---T--A-----------G-G--C---AAG-CC------GA---CT-A-----AC----AC-TT-GGA--G---A-----A-----------T--C--C--  3123
GRV.ET.x.GRI_677         AA-----A--T---T-G-----------C--AT--AG---------TG----T---AC------A--C-A-C-A------G----G--A---T-A--T--GG-A--AGGA--------T--T-----A--  2898
TAN.UG.x.TAN1            --TCA--A--G---T-A-----------CC-AT-------------------CC-GAC------AAC-CAG-CTC-A-----------A--------T--GG----AGGA--G-----T-----------  2927
VER.DE.-.AGM3            CTAGA--A--TTTCT-A-----------CC-CT-CA----------------C---CAG-----A---GACC-AT-A--G--A-----A------AGC--G--A--AGGA--G---G-------C-----  2450
VER.KE.-.9063            CAAGG--A---C-TA-A-----------TC-AT--A-G--G--------G--C---AC---G--A---GA-C--T-A-----A-----------CAGC--G--A--AGGA------G-T--T-----C--  2955
VER.KE.-.AGM155          CCAGA--A---CG-T-G--G---------C-CT-TA------------T---CC-GCA---------C-A-ACAT-A--GG-A-----A---T-AAGC-GGG-A---GGA--C---G----------A--  2947
VER.KE.-.TYO1            CTCGG--A--CTGT--A-G---------TC-CT-TA----G--G--TG----T---CAG-----A---T-CC-AT-A--GC-------A-----CAGT-G-G-A--AGGA------G----------C--  2445
SUN.GA.98.L14            CC-C---A--------G-----------CC--T--AGG--G------G---GCC--CAGA--------GATAGAT-----GCA-----A--G---AGC-GGGCA-AC--A-GG--C--T--T--C--A--  3082
LST.CD.88.447            GT-----T--C-G---A-----G-----T---T-TA-------G---G---GCC--CA--C-------GATAGAT-A-----A-----------A------GCA-ATG-A-GG--C-----T--C--A--  1993
LST.CD.88.485            GT-----T--C-G---A-----G-----T---T--A-------G---G-G-GCC--CA--C-------GATAGAT-A-----A-----------A------GCA-ATG-A-G---C-----T--C--A--  1993
LST.CD.88.524            GT-----G----G---A-----G------C-AT-CA-----------G---G----CA--C-------GATAGAT-A-----------------A--CCC-G---AT--A-G------T-----C--C--  1990
LST.KE.-.lho7            GT-----A----G---A--G--G-----TC-AT--A----G------G---GCC--CA--CT------GATAG-T-A-----A-----A-----A--TCC-G-G-AT--A-GG--C--------------  3077
SYK.KE.-.SYK173          AA--ATTG--CCGGC-A--------------AT-G-T---G---T--G-G--CC-CAAG-C---AGTAGA---C---CTC----C---AC--T-GGAG---GC-A-T---ACC-----T--T-----C--  2922
SYK.KE.KE51_AY523867     A---GCTA--T---T-A-----G-----TC-----AG-------C-TG-G--C--GCG--C---AGT-T----T---CTTG-AA-G---C--T-GGA---GGCCTCC---AC------T--T-----C--  2580
COL.CM.-.CGU1            -------G--------A--G--G-----C--A-GCA-----------G---GT---CA-A-G--A---GA-AGAT-----GCA--G--A-----AGA----GCA-AATTA-GG--------T-----A--  2500
```

```
H1B.FR.83.HXB2               AGTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAACTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGGTTAAAAAAG  2855
Pol                          __V__F__A__I__K__K__K__D__S__T__K__W__R__K__L__V__D__F__R__E__L__N__K__R__T__Q__D__F__W__E__V__Q__L__G__I__P__H__P__A__G__L__K__K_    RT

H1D.CD.84.84ZR085            -A-------------------------------G---------------------------------------------------------------------G-----------T-----A-----G---  2377
H1F1.BE.93.VI850             -----------------------------G---------G----------------A------------A-----------T--------G-----------------------T---------------  2192
H1G.SE.93.SE6165             -A-------------------G-----------------------G--------------G--C-----A--------------------G--C--------------T-----T---------------  2252
H1H.CF.90.056                -A-------------A--G--G--T-----------------------G--------------C-----A--------------------------G--------------C--A---------------  2202
H1J.SE.93.SE7887             ------------------------------------------------------T--------C-----A--------------------------G-----------------A---------------  2169
H1K.CM.96.MP535              ---G--------------------------------------------------T--------C-----A---------------------------------------------A-----A---------  2051
H101_AE.TH.90.CM240          ---------T-----------G-----C--C--------G--------------------G--C-----A--------G-----T-----------------------G-----A-----T---------  2429
H102_AG.NG.-.IBNG            ---------------------G--T--------------------------------------C-----------------T--------G--C--G-----------T--------G--A---------  2380
H103_AB.RU.97.KAL153_2       --------------------------------------------------G------------------------------------------------------------C--T---------------  2081
H104_cpx.CY.94.CY032         -A-------T----------------C-----------------------------------C-----A--------G-----------------G-----------G--C--A-----------G---  2221
H1N.CM.95.YBF30              -A-T-----T-----A--G-----T--C-----------------------------G---T-A-----A--G--C-----T--T-----------G--GC-------T---------A-----A-----GC--  2449
H1O.BE.87.ANT70              TA-C-----T-----A--G-----TG-------------------G--------T--G---T-A-----------A------G--------G--A--GC----T--C--------G-GG--T--G--GC-A  2910
H1O.CM.91.MVP5180            TA-T-----T-----A--G-----T--C-----G-----------G------C---------T-A-----A-----A-----T--------G--G-----------T--T---------A-GG--T-----GC-A  2885
CPZ.CM.04.EK505              -A-T-----T-----A--G-----T--C--------------------------------GT-A--C--A-----------T--T--------A-----------C---------A------C----G---  2427
CPZ.CM.04.LB7                T--G-----T-----A--G-----T--C--A-----------------------T-------T-A-----A--------------T------------------------T--C--T--G--A-----G---  2410
CPZ.CM.04.MB66               -A-T-----T-----A--G-----T--C--------------------------------GT-A-----A--------------T--------C--GC---------C-----T------------C--  2401
CPZ.MT145                    TA-T-----A--T--------G-----C--C------------C-T--T---------------------G--------A--T-----------A---C-G--C--------C--A--T--------G---  2397
CPZ.CD.-.ANT                 -A-T-----A-----A--G-------C-T-A-----------GC----T-------------T-A-----A-----A-----T--T-----GA-A-----------T--T-----A--C--A-----GC-A  2294
CPZ.CM.-.CAM3                -A-T-----T-----A--G-----T-----A------------C-G-----------G---T-A-----A-----A--G-----T--------C--G-----------C--------C--G-----T---------  2266
CPZ.CM.98.CAM5               -A-T-----T-----A--G-----------A------------C-G--------T-------T-A--C--A-----A--G--A--------G--C-----------C--------C--A-----AC----G--A  2557
CPZ.GAB_X52154               TA-T-----A--C--A--G-------------------------------C-----------------A-----------A--------T--------G--G-----C-----T--C--A---------------  2916
CPZ.GAB2                     CA-------A----G---G-----------A--G--------------------T-------T-G-----A-----A-----T--T--------A--G-----C--T-----C--T-----T-----G--A  2248
CPZ.TZ.-.TAN1                -A-T--------T--------G--T-CCT----------------------------------G-----A-----------T--T-----GT-A--GC----------C-----G--------------G---A  2499
CPZ.US.85.CPZUS              -A-T-----T--C--A--G-----T--C-----------------------C--T-------T-A-----A--G--A---------------------G-----C--------C--A-----A-----G---  2915
GSN.99CM166_AF468659         ---C-----------A--G--G----AAT-AC---------TGC--A-------T---A--T-A---G--GCT--G--------GGCT--G--C--G--G-----T--G--C--A-----AA-TCCGC-C  2565
GSN.99CM71_AF468658          ---C-----------A--G-------AAT-AC---------TG---A-----------GA--T-G---G--GCT--G--------GGCT--G--C--G--G--G--T--G--C--A-----A.........  2556
H2A.ALI_AF082339             CAC------A--------G--G----AA-AC----------TGC--A-------T---------T-A--C---GT---------T---ACA--GA----G-----G--T--------A-----A---GCC---  3233
H2A.BEN_M30502               CAC------A--T--------G----AG-AC--------G-TGC-GA-------T-----------A------GTG--------T---ACA---A----GC-------T-----C--G-----AC--GCC--A  3240
H2A.ST_M31113                CAC------A--C--------G----AA-AC----------TGC--A-------T-----------A--C---GT--------------ACA---A-C--G--------T-----C--A-----AC--GCC---  2685
H2B.D205_X61240              CACC--------------G-----T-AG-AC----------TGC--A----C--T-----GT-A--C---GT---C--G-------ACA-----CA-C-.G--T--T--C--C--G-----TG..GC-G-A  3213
H2B.EHO_U27200               CACC--C--------A--G-------AA-AC--------G-TGC--A-----------------A--C---GT---C------A--TACA--G--C--GC-G--T--T--T--C--A-----AC-GGC-TCA  3211
H2G.ABT96_AF208027           CACC---A-T--C-----G--G----AA-AC----------TG---A----C---------------C-G-GT---C--G-----TACT-----C--G--G--C-----T--C--T-----T--GGC-G-A  2567
H2U.FR.96.12034              CACC--------------G--G----AA-AC----------TG---A----C--T-----GT-A--C-GAGT---------T--TA-T-----C--G-----T-----C--C--A-----AC--GCTG--  2722

MAC.239_M33262               CAC------T-----------G--T-AG-AC----------TGC-GA-------T--G--------A----G-GTC-----G-----TAC------C-----------------C--T-----AC--GC---A  3157
Pol                          __T__F__A__I__K__K__K__D__K__N__K__W__R__M__L__I__D__F__R__E__L__N__R__V__T__Q__D__F__T__E__V__Q__L__G__I__P__H__P__A__G__L__A__K    RT

SMM.SL.92.SL92B              GACC-----T--T--------GA-T-AGGAC--G-------TGC-CA---------------T-A--C--A-TG--A-----A---ACA-----A--GC-G--T--C--------G-----CC----GG-A  2588
SMM.US.x.PGM53               CAC------T-----A--G-------AA-A---------G-TGC--A-------T-------T-G----G-GTT--------T--TACA-----A--G-----G--------C--T-----AC--GC----  3074
MNE.US.-.MNE027              CAC------T-----------G----AG-AC----------TGC-GA-------T--G-----A------GTC-----G-----TACA-----C-----------------C--T-----AC--GC---A  2625
RCM.NG.x.NG411               -A-----TGT--C--A--G-----TTCC--A-----------GC-------------------T-G-----A-----A-----T----TT--G--A--GC-G-----C--G-----A-G---A------C-A  2349
RCM.GAB1_AF382829            -A----CTGT--T--A--G------TCC--A--------------G--------T--------A-----------A--G-----T-TT--G--G--GC-G--------------A-G---A-----GC-A  2334
DRL.FAO_AY159321             G--G---TG---T-----G-----TG----C---------------------------GC--T-A------GC---A--G--------TT--G--G--G--G--G--------C--G-G---C------GCA  2422
MND-1.F17                    TA-C-------------G--G--T-AA-A-G--------------A----C--T-------T-A--C---TT---A--G--T--TCAT---T-A--G-----T-----T--C--A-----AA--------  2542
MND-2.5440_AY159322          ------CTGT-----A--G-----TG-A--C-GC--------------------T---C----G--C--AGTC-----G--T----TT--G--G--G--------CA-------A-G---CC----GC-A  2416
MND-2.14CG_AF328295          -------TGT-----A--G-----TG-A--C-GC--------------------T--GC--T-A----GAGTG-----G--T----TT--G--G--GC-------C--------G-G---C------C-A  2783
MND-2.MND2_AF367411          G--G---TG---C--A--G-----TG-A--C-GC-----------G--------T---C-GT-A--C--AGT------G--------TT--G--G--G--G-----C-----C--T-G---TC-C---C-A  2856
DEB.CM.CM40                  ---G----TA--------G--G----AA-G---G---C---TGC--A----C--A----CCT-A--C---GCC--A--G--GG-G-----G--G--GAC---T--------C--AT-T-CAC--CC-C-A  2664
DEB.CM.CM5                   T--G----TT--------G--G----AG-A-----------TGC-TA----CC-G----CC---------GCT--C--G--TG----------G---AC---G--C--G--C---T-G-----GCC-C-A  2658
MON.CM.L1                    CTGT-----------------G----AG-GC----------TGC-CA-------------------C---GC------G-------TT--G--G--G--G-----C-----C--A--G--AC-TG-G---  2561
MON.NG1                      TTGC--C--G----G------G--T-AA--C--------G-TGC-CA----------G---T-A--C--AGC---A-----T----TT--G--A-----------T-----C--G--T--AC-CC-G--C  1154
TAL.CM.8023                  ---G-----T--C-GA--G--G----AG--AC---------TG--GA--------C---CC--G--C---GC------G-----TGCA--G--C---CAG--G--------C---T-T--AC--CC-C--  2322
TAL.CM.266                   ---G--C--T--C-GA--G--G----AA---C---------TGC-CA----C---C-C-CC--A---C-AGC---C---------GC------A---CAG--G-----T--C---T----CC-GCCTC--  2824
MUS.CM.1085                  -TGC-----A--T-G------G----AAT----G-------TGC--A----------GC--T-A---G--GCT--C-----AC--GCA---------C-------C--------A-----A----T-G--  2545
SAB.SN.-.SAB1C               ---C-----A----G---G-----T-AA---C-G-------T-C-CA-G--C-----GC-GT-A-----A--T-----------TCA------C--G-----G-----C--C--A-----C---C-GC-A  3253
GRV.ET.x.GRI_677             ---G-----------------G--T-AA--AC---------TGC-T-----------G-----A--C--AGCT-----------T-TT--G-----GC-G-----T--T--C--A--G--CC-TC-G--A  3028
TAN.UG.x.TAN1                C--G---TGT--CCG-----------AA--AC-G-----G-TG---A-------T----CT--C-----AGC---A--G--------TT--G--A--------T--T--C-----A-----A--G----GA  3057
VER.DE.-.AGM3                ---G---TG-----------------AGT-AC---------TG-----------T--G-----A--C--AGC---A-----T--T-TC-----A--G-----C-----T-----AT-------CG-----  2580
VER.KE.-.9063                -A-C---TGT--C-GA--G--G----AA---C-------G-TG--------C---------T-A-----AGT---A----------TT-----------G--------T--C--G--------GC-G--A  3085
VER.KE.-.AGM155              ------CTGT----G-----------AAT-AC-G-------TGC-G-----------G-----C--C--AGCT--A----------TT-----C--------T-----C-----A-----------G--A  3077
VER.KE.-.TYO1                -A-----TG---------G--G----AATCC-G-----G-TGC-------C--T-----GT-A------GC---C-----T----TT-----G--------G-----C--C--A-----A----G----  2575
SUN.GA.98.L14                -A-T-----T--------G-------AA-A-G----------G---A-------T----T---A---GGT-TG-----------TCAT--GT-A--G-----G--T--T-----A-----A-----G---  3212
LST.CD.88.447                -A----------------G-------AA-A-G-G------------A-T-----T---C----A---GCTTT---A--------TCAT-----A--------C--T-----C--T-----AA----G--A  2123
LST.CD.88.485                -A-C--------------G--G----AA-A-G-G---------C-GA-T-----T---C----A---GCTTT---A------C-TCAT-----A--------C--T--------T-----AA----G--A  2123
LST.CD.88.524                -A-C-----T-----------G----AA-A-G----------G--GA------------A--T-A---G--TT---A--------TCAT--GC-A--G--------------C--A-----AA----G--A  2120
LST.KE.-.lho7                -A-------T--C-----G-------AA-A-G----------G---A-T---------A-G--C---G-ATT--------------TCAT--GC-A--------T--T-----C--A-----TA-C--G--A  3207
SYK.KE.-.SYK173              --------TA--T-G------G----AA-AG--------G-TGC-CA-T-----T---A--T-A---G--GCC--A-----T--T-TC--G--A-----G--------C--C--A-G---AC-----CG-  3052
SYK.KE.KE51_AY523867         ---T----TA---C-A--G--G----AATTA--------G-TGCCTA-G-----T--GA----C---G-AGC---A--G-----T--------A--G-----------T-----A-G----C-C-GT-CA  2710
COL.CM.-.CGU1                -A-----TG---T-G---------T-AA-A-G-------G------A----------G--G--A--C--------A---------AT------C--------G--------C--A-G---TC-T-T-G-A  2630
```

```
                                                                   | Asp110 catalytic site
H1B.FR.83.HXB2              AAAAAATCAGTAACA..................GTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACAC  2967
Pol                         _K__K__S__V__T__.__.__.__.__.__.__V__L__D__V__G__D__A__Y__F__S__V__P__L__D__E__D__F__R__K__Y__T__A__F__T__I__P__S__I__N__N__E__T__    RT

H1D.CD.84.84ZR085           -----G---A-----..................--------------C---------------A--------TG---------T--------C-------------------------------------  2489
H1F1.BE.93.VI850            -----G---------..................--------------G---------------------------A-G--T----A------C--------C------------G-C-------------  2304
H1G.SE.93.SE6165            ---------------..................-----A--------G-----------------------------------T--A--------------C--T--------------T----------  2364
H1H.CF.90.056               ------------T--..................--------------G-----------------C--T------A----A-----A--------------C----------------------------  2314
H1J.SE.93.SE7887            -----------C---..................--------------G-----------------C--T---T-------T--------A-----------C--T-------------------------  2281
H1K.CM.96.MP535             ---------------..................--------------G-----------------C--T------A-------------------------C--T-----------T--------A----  2163
H101_AE.TH.90.CM240         ---------------..................-----A--------A--------------------T---------AG---T--A--------------C----------------------------  2541
H102_AG.NG.-.IBNG           ---------------..................--G--A--------G---------------------------A-------T--A-----------------T---------G----T----------  2492
H103_AB.RU.97.KAL153_2      --------T------..................-------------------------------------------C----------A----------------------------C--------------  2193
H104_cpx.CY.94.CY032        ---------------..................---T----------G---------------------------CC---G-----------C--------C-------------CC-------------  2333
H1N.CM.95.YBF30             -----------G---..................--TT-------A--A-----T---------TG------G--CA----T--T--A--------A--T-------------------------------  2561
H1O.BE.87.ANT70             --GC----T--T---..................--CT-A-----A--A-----T-----C---TG---------CCC---T--T--A--A--------T--C--T--T------G-G--T--------C-  3022
H1O.CM.91.MVP5180           -GGC----T--T---..................--CT-A-----A--A-----T-----C---TGC--T------CC------T--A--A--C-----C--C--T--T------G-G-----------C-  2997
CPZ.CM.04.EK505             ---------------..................--G--A-----A--A-----T--C------TG-------------A-T--T--A--------A--T---------------G----T-----A----  2539
CPZ.CM.04.LB7               ----GG-----C---..................---T-------A--A-----C--C------TGC---C--------A----T--A--A-----------C-----------------------A----  2522
CPZ.CM.04.MB66              ---------------..................--TT-A-----A--G--C--T--C------TG----C-G------A----T--------C-----T-----------C---G---------------  2513
CPZ.MT145                   ---------------..................--CT----------A-----C---------TG---TC-T--CCC------T--A--A-T---------C-----T-----C-----T-----A--T-  2509
CPZ.CD.-.ANT                --G--------G---..................---T----------A-----C-----C--CA-A---------C-G-----T--A--A-----A--T--C--A--T--A--CG-G-----------C-  2406
CPZ.CM.-.CAM3               -----G---------..................--G--A--------A-----T--------CTGC--A---------A-T--T-----A-----A--------T-----C---G----T-----A----  2378
CPZ.CM.98.CAM5              --G--G---------..................-----A--------A-----T-----C--CTGC--A---------A-T--T--A--A--C--A--T-----T-----C---G----T----------  2669
CPZ.GAB_X52154              -----------G---..................---T-A-----A--A-----C--C--C--TTG----C-G---A----T--------A--------------A--T--C--------T--------C-  3028
CPZ.GAB2                    --G------------..................--G-C---------A-----T--------CTGC------------A----T--A--------A--C--C-----C------G----T-----A--T-  2360
CPZ.TZ.-.TAN1               -G---TATG--G---..................-----------A--G-----C--C-----CA-----C-G---CC---------A--------A--T-----------C---C-C--T---A-C----  2611
CPZ.US.85.CPZUS             ----G----------..................--C--A-----A--G-----C--------TTG---AC-G---A-G--A--T--A--A--C--A-----------C------G-------------C-  3027
GSN.99CM166_AF468659        CCTGC-GG-C--GAGCAAAAGGAGCATGTAACTA--A-A---A--AAA-----T----A-AGTA-A--A--GT----G--T-----------C--C--T----T-AG----TCAG----T---C--G-C-  2695
GSN.99CM71_AF468658         .........C--GA-CAAAAAGAGCACGTAACTA--A-A---A--AAA--C--T----ACAGT--A--A--GT-------A-----A-----C----------T-AG---ATCAG--------C----C-  2677
H2A.ALI_AF082339            ----G-AG-A-C--T..................--G--------A--G-----T--C-----CA-A--AC-GC----GAG---T--AC----------------TC----ATCAG---------CAGA--  3345
H2A.BEN_M30502              --G---AGGA-CT-T..................A--T-A-----A--G-----C--------CA-A--AC--C-------T--T---C-----------------C----AGCAG--------AT-GA--  3352
H2A.ST_M31113               --G---CG-A-T--T..................--C--A-----A--G-----T--C-----CA-A--AC--C----G--T--T--AC----------------TC----ATCA----------CTGA--  2797
H2B.D205_X61240             ----GGAG-A-----..................---A-A--------A-----C--C--CAGTA-C--AC-----CC-A-------AC-------A-----C---T-G--ATCAG---------CAGAG-  3325
H2B.EHO_U27200              --G---AG-A-----..................-----A-----A--A-----C--C--CAGT--C--AC-----CC---------AC-A-----A--------TT-G--AGCAG----T----CAGA--  3323
H2G.ABT96_AF208027          ----GGAG-A-----..................---------A-A--G-----C--C--CAGCA----TC------T---T------C----C--A--T--C--TT----ATCAG---------CAGA--  2679
H2U.FR.96.12034             -----GAGGA-----..................--T--A-----A--G-----C--C--CAGCA-C--GC-GTG-------------C-A--------T-----TT----ATCC--T--T----CAGAG-  2834

MAC.239_M33262              -GG---AG-A-T---..................---------A-A--------------C--CA-A--TC----------A--T---C----C-----C-----TT----ATCAG----T----CAGAG-  3269
Pol                         _R__K__R__I__T__.__.__.__.__.__.__V__L__D__I__G__D__A__Y__F__S__I__P__L__D__E__E__F__R__Q__Y__T__A__F__T__L__P__S__V__N__N__A__E_    RT

SMM.SL.92.SL92B             -TGG--AGGA-----..................--G--A---A-A--G-----C-----------A---C----CCCG--A--T--AC-A--C--A-----C--T-----ATCAG-G-----CC-AGA--  2700
SMM.US.x.PGM53              -GG-G-AG-A-C---..................---T-------A--------------C--CA-A--TC----------A------C-A--C-----T-----TT----ATCAG---------CAGA--  3186
MNE.US.-.MNE027             -GG--GAGGA-C---..................-----------A-----C--------C--CA-A--TC----------A--T---C----C-----T-----TT----ATCAG----T----CAGA--  2737
RCM.NG.x.NG411              TGC-GGCA-A-T---..................--CT-A---A-T--G-----G---------TGC---C--T----GCCA--T--A-----C-----T--C--G----TAGC---T--T---C-AGG--  2461
RCM.GAB1_AF382829           TGTG-GAG-A----G..................---T-A--CA-C--G---------------TG--TTC-GT----GCCT--T-----A--------------A-----AGCAG----T---C-AGG--  2446
DRL.FAO_AY159321            TGCG--CAGA-C--T..................-----------A--------C-----C---TGC--TC-G---CCG-----T--A-----C--------C-----T--ATCGG-G--T---C--GGC-  2534
MND-1.F17                   TGT---AG-A-----..................--C--A---A-A--G-----C------AGTA-A--TC-G---CC---T-AT--ACCC--------C-----GG----ATCAG-T--T---C-AG---  2654
MND-2.5440_AY159322         TGTG-GCA-A-T---..................---T----CA-A--A-----C---------TG------G----TG-----T--A-----------G--C-----T--ATCGG-G-----CC--GGC-  2528
MND-2.14CG_AF328295         TGTG--CA-A-C---..................A--T-----A-A--A-----C---------TG---T--G-----GAG---T--A-----C--------------T--ATCGG----T---C-AGGC-  2895
MND-2.MND2_AF367411         TGTG-GCA-A-T---..................--------CA-A--G-----C-----C---TG---T--G-----G-----T--A-----C--------C-----T--ATCGG-G--T---C--GGC-  2968
DEB.CM.CM40                 -TGG--CAGA-C--T..................--CT-A---C-T-CA-----C----A----A-AG-AC----CCC-ACA--TGCA--A--C--A-----C--AG----ATC-G-C--T---ATACA--  2776
DEB.CM.CM5                  -TGG--CAGA-----..................--C--A--CT---CA-----C----A----A-A-----G---CC-C-G--TGCACGT-----A-----C--A------TCGG-G------ACACA--  2770
MON.CM.L1                   -TGG-CCAT--G---..................A--G-A--CA--AAG--C--C-TC-ACAGCA-------GTGG---CCG-----A--A--C--G--G--CT-A-----CTC-T---------CTGA--  2673
MON.NG1                     -TGGC-AATA----T..................A--G----CA--AAA-----C-T--ACAGCA-C--TC-G------A-T--T-----A------------T-A------TCAG----T----CAGA--  1266
TAL.CM.8023                 -----GCAGA-----..................--CT-A---A--AAG-----C----A-AGCA-C--A------CC-------GCCCGT--C-----T--C-----C--GTC-C-G--T-----C-G--  2434
TAL.CM.266                  -TG--GCA-A----G..................--C--A--CA--AAG-----C----ACAGCA-C--------CCC-------TCAC----C--A--C--C-----C--GTC-C-G--------T-GG-  2936
MUS.CM.1085                 C-TG-GCAC--T---..................A--G-----A-CAAG--C--C-T--ACAGT--C---C-----CC------T--------C--A--C---T-GC-G---CAGG------CTT-GG---  2657
SAB.SN.-.SAB1C              -GGG-GCAGA-T---..................--GT----CA-A--A-----C------AGCTGC--T------CC------TCA---A-----A-----C-----C--ATCAG-C--T---AG-GAG-  3365
GRV.ET.x.GRI_677            --G--GCA-A-C---..................---A-A--CA-A--G-----C----A----A-A--A---TGCA-G--A-----A--A-----A-----------C--CTCAG----T---ACAGGG-  3140
TAN.UG.x.TAN1               --G---CAGA-C---..................-----A---A-A--A-----C----A-AGTA-A--A--G---CC---A------CCC-----A--T--------C--ACAAG----T----TAGGG-  3169
VER.DE.-.AGM3               -TG-CGGA-A-----..................---T-A--CA-A--G-----C----A----A-A--A-----CCC---G--T--A--------C--T--------T--ATCAG----T---C-AGGG-  2692
VER.KE.-.9063               -TG---CAGA-----..................---T-A---A-A--G-----T----A---CA-A--A------CC---G--T--A--A--C--A--------A--C---TCAG----T---C-AGG--  3197
VER.KE.-.AGM155             -TG--GCA-A----C..................A-TA-A--------G----------A-AGCA-A--AC-G---CCT--G--T--A--A--C--A--T--C-----C----CGG-----------GG--  3189
VER.KE.-.TYO1               -TG-G-CAGA-----..................--TT-A-----A--A--C--C----A---CA-A--A--G---CC-A-T--T-----A--------T-----T--T----CAG-G--T---C--GG--  2687
SUN.GA.98.L14               TGT---CAGA-T---..................---G-A--------G--------C--C---A-A--A--G---CC---T-AT--A--------A-----C--A-----GTCAG----T---C--G---  3324
LST.CD.88.447               TGC--GCAG------..................--GG-A---A-A--A-----C------AGCA-----C-G---CC-----A---AC----C--A-----C--AG----ATC---T--T--CC--GA--  2235
LST.CD.88.485               TGC--GCA-A-----..................--GG-A---A-A--A-----C------AGTA-----C-G---CC-----AT--AC----C--A-----C--AG----ATC---T--T--CC--GA--  2235
LST.CD.88.524               TGC---CA-A-T---..................---G-A---A-A--A-----C-----CAGCA-A--T------CCT--T-A---AC-A--------G--C--T-----GTCAC--------C--GA--  2232
LST.KE.-.lho7               TGC---CAGA-T---..................---G-A--CA-A--A-----C-----CAGTA-C--T------CC-A-T-A---A--------A--------T------TC-C-C--T---C-AGAG-  3319
SYK.KE.-.SYK173             C-----...C-----..................A-TA-A---T-AAAA----------AC-----C---C-G--CA----G--T---CCA--C--G--------AG-G--A--C-----T----CCT-C-  3161
SYK.KE.KE51_AY523867        ----G-...A-G---..................---T-A--CT--AAA--C--C----ACA----A--TC----------G-----AC-------A--------AG-G--C------------TCTT-T-  2819
COL.CM.-.CGU1               --G-CCCATA-----..................---T-A---A-A--A-----C-----C---A-------G---CCG--T-ATCAA-----C--A--T-----A--C--ATCAG-G------TT-G---  2742
```

```
H1B.FR.83.HXB2             CAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTTATCTATCAATA  3097
Pol                        P__G__I__R__Y__Q__Y__N__V__L__P__Q__G__W__K__G__S__P__A__I__F__Q__S__S__M__T__K__I__L__E__P__F__R__K__Q__N__P__D__I__V__I__Y__Q__Y    RT

H1D.CD.84.84ZR085          -------------------------A--------------------------------------------T-----------------C-------------------AG----------------  2619
H1F1.BE.93.VI850           -T--A-----G--C-------------------A-----------------------------T-----------------------C-------TGA----C-----------C-----C-----  2434
H1G.SE.93.SE6165           -----G-----------------------------------------------------T--G-------------G-----------C------GC-A-T---------A--G--G----C-----  2494
H1H.CF.90.056              ------------------T----------------------------------------------G------------------C---C------G-----------T--A--G-----T--C-----  2444
H1J.SE.93.SE7887           -------C--------------C-----A-------------------C-----------T--GT-------------------A-A---------G--AG---C-----A-----------C--G--  2411
H1K.CM.96.MP535            -----G-----------------------A-----A-----------------------------CA----------------------C-------T-A-----------A--G-----A--C-----  2293
H101_AE.TH.90.CM240        ----A--C-------------------G-----------------------G-----------G------------------------C-------T-A-----------A--G----------------  2671
H102_AG.NG.-.IBNG          -------------C--------------------------------G-----G---------T--GGCA-------------------C-------C-A-----------G-----G-----C-----  2622
H103_AB.RU.97.KAL153_2     ---------------------------------------------------------T----------------------------------------------------G----------------  2323
H104_cpx.CY.94.CY032       ----A--------------------------C------------------------------------------T------------C------TTCA----C-----A-----C--A--C-----  2463
H1N.CM.95.YBF30            -T--T--------C-----T--------G-----A--C---------G-----------T--T--G----CA-----------TC----A--A--C---G-GA--C-------G---A-C--T--C--G--  2691
H1O.BE.87.ANT70            ----A--A-----C------------C--C--G--A------------T--------T------------TCA-----------TC----T--A--C--G-G-G-C--C-----AT---AA--T-G---G--  3152
H1O.CM.91.MVP5180          ----AG-A-----C------------C--C--G--A--G---------T---------C-----T--G---TCA--------G--TC----T--A----------AGC--C-----AG---AA--T-----G--  3127
CPZ.CM.04.EK505            -T--T--------C--------------G-----A--C-------------------C--T--G--C-CT-----------A-----A--A-----G--GA-C--------AC----C--A-----G--  2669
CPZ.CM.04.LB7              ----A--C---------A--------T--A--G-------------------------T-----T------------------G-T--------A--C----G---G--C-----T-----A--T-----G--  2652
CPZ.CM.04.MB66             ----A--------C-----T--------A--G--A--G---------G--------------T-----T------------------T-----------C--------G---------A---A---------------  2643
CPZ.MT145                  ----A------------A--T-----A-----T--A---------G-----T--TT-C--C--T--G----------------G---------A--C-------C-A-GC-------A---A----T--C--G--  2639
CPZ.CD.-.ANT               ----A--A------TGT--T-----T--A-----A--C------------------T--T--T---GCA---------C--G-------CA--A-----GG-TA-GT------CAG---AA--T---------  2536
CPZ.CM.-.CAM3              ----A------------A--T------T-A-----A------------G-----------T--T-----C-------T------TC--ACC--A--C---C-G--GC---------T--A-A--T--C--G--  2508
CPZ.CM.98.CAM5             ----AG--------------T------T-A--R--A------------G-----G-----C--T-----C----------------A-----T--A--C--------GC-------TC--A-A-----C-----  2799
CPZ.GAB_X52154             ----AG-----------A--T-----TT-G-----A------------G--T---T-T--C--------C--T-----------TC----A--C--C---G--A-G-----T-----TAC---T--C--G--  3158
CPZ.GAB2                   ----T--A--G--C-----T-----T--A-----A------------G-------------C---------C-----------GG-T-----C-----C---C-GA----C-----AT-G--A--T--C-----  2490
CPZ.TZ.-.TAN1              ------AA----T------T--C---T-A--T--A--T-----G-----T--------T--T--G--C--T------------C-----T-----C------G--C-C-----TG-G-AC--T--C-----  2741
CPZ.US.85.CPZUS            ----A--------C-----T------T-A-----A------------G-----------T-----G--C--T-----------TC----T-----------------C-------TG-TA-A-----------  3157
GSN.99CM166_AF468659       ---CTGAA--G-------T---A--CT-A--G-----T-------CG--G---A-C--C--T----AC-TTG-AG-TTCCT-GC--CGA-AGA-----G-T--GG-G-----TG-G--AT-AATA--G--  2825
GSN.99CM71_AF468658        ---CTGAG--G-------T---A--TT-A--T-----T-----G-CG--T---A-C--T--T--GGCC-CTG-AG-TTCCT-G---T-T-AGA-----G-C--GG-G-----TG-G--A--AATA--G--  2807
H2A.ALI_AF082339           ----A-AA------ATA--T--A--CT-A--G--------------G--------------T--T---CAC-CA----G-C-G-------------A--C-----GGC---C-AG--TG-CA--C--AT------  3475
H2A.BEN_M30502             ----A-AA------ATA--T--A--CT-G-----A---------G--------------T--T---TAC-CA----GGC--G-------A-----C------GC---C-----TG-CA--C--ATC--G--  3482
H2A.ST_M31113              ----A-AA-----CATA--T--A--CTCA---------------G--------------T--T--GTAC-CA----GGC-GG-------A--A--C------GC---C--G--T--CA--C--AT---G--  2927
H2B.D205_X61240            ----A-AG-----CATT--T--A--T--A-----A--G-----G--G--C-AGT----T-GT---TACTC-----G---GG-A-----T-----------GGCC--CAGC--TG-CA----AAT---G--  3455
H2B.EHO_U27200             ----A-AG-------TT-----A--C--A----------------G-----C---------T-----GTAC-C----G----GG-AC----C-----C------GCC--CAAT--TG-CAC---AATC--G--  3453
H2G.ABT96_AF208027         ----A-AA-----CATC-----G--C--A--G--A--------G--C--------T--C------CA--CA----GG--GG-AC----C--C--C------GCC---GAG--TR--A----YATC-----  2809
H2U.FR.96.12034            ----A-AA-----CATC-----G---T-G--T--A--C--------------A----T--T--GTA--CA----GG--CC-AC-G-----A--------GGC----GA---TGCT--C--AAT---G--  2964

MAC.239_M33262             ----A-AAC----CATT--T--G--T--G--T-----------G--G--------C--C------TAC-CT----G-C-TG-GC----A--C--C--G--GGC---------TG-GACCT-AGTC--G--  3399
Pol                        P__G__K__R__Y__I__Y__K__V__L__P__Q__G__W__K__G__S__P__A__I__F__Q__Y__T__M__R__H__V__L__E__P__F__R__K__A__N__P__D__V__T__L__V__Q__Y    RT

SMM.SL.92.SL92B            ----A-AG--G---ATT-----G--C--G-----A--------G-----T--T-----T-----GGCA-C-----GGC-GG-A-----A--C--C------GCT--C-----TG-GC--C--GTC--G--  2830
SMM.US.x.PGM53             ----A-AA-----CATC--T--G--AT-A--T--A--G--------G--------T--T--T--GCA--CT----G---TG------GA-----C------GC---------TG-GACCC-GATC-----  3316
MNE.US.-.MNE027            ----A-AAC----CATT--T--G--C--G--T-----G-----G--G--------C--C------CAC-CT----G---TG-GC-G--A--C--C--G--GGC---------TG-GACCC-AGTC--G--  2867
RCM.NG.x.NG411             -T--C------------A--T------T-G--C-----------G-----T-----T-----T----CA-CAGCT-AC-----T-----A--A-----G---G-G---------TG----C-------G--  2591
RCM.GAB1_AF382829          ----AG-G--G-----A--T--------G--T-----------G-----T--C--C--C--T--GGCATCAGCC-AT--G------C----A-----GG--G-G---------TG-CA-C--T--C--G--  2576
DRL.FAO_AY159321           ----A--C--G--------T-----C--C------------------C-----T-----C------GC--CAGCTGAC--G------C----A--C---G-GA---------TG-GA----A-----G--  2664
MND-1.F17                  ----A-AA-----CAT---T-----T-----T--A--G-----G---AGT---TGT--C--T---G-G-CAG-AG--TC-C-GC-G---GTA---------GA-CC-----ACAG--CAGT-A---------  2784
MND-2.5440_AY159322        ----A--C-----C-----T-----C--A------------------C--T--------T-----GGCA-CAGCAGAC-----TC--A--G------AGG--A-GC-------AG-CT-A---------G--  2658
MND-2.14CG_AF328295        ----A--C--G--C-----T-----T---------------------C--T--G--------T---GCA-CGGCAGAC-----TC--A-AA-A--C-A-G-GG--T-C-----AG--T-A--T---------  3025
MND-2.MND2_AF367411        ----A--C--G--C-----T-----C--A--------------G--C--C-----C--T-----GGCA-CAGCAGAT---------C----C--C---G-GAG-C-------TG----G-----------  3098
DEB.CM.CM40                ----AGAA------G---T--GA--A--G--T-----------GC-CT--------C--T--T---GCCTCAG-AGGG-G-CAG---C--ATC-----GG--A-G---------GC--A-A--AGTC--G--  2906
DEB.CM.CM5                 ----AGAG------G---T--GA--A--A--T--A--C----GT-C----------C--------GGCA-CAG--GGC-GGCAGC-GA-A-T-----AGG--A--T-C------T--A----AGTC-----  2900
MON.CM.L1                  -T-CT-AA--G-------T---G--C--C--G-----C---------G--C---A-TG-G--T---CA--CAGC-G--G-GC-ACC-C--GAGA-C--G---A--T-C-----GG--AC-T-AATA--G--  2803
MON.NG1                    ---CCG-G----------T--GA---T-G--C-----R------------C---A-------T--GTACTCAGCAT--G-GC-GC-RA-AGRAA-C--G-G-A--T-------GG-CACCC--ATC--G--  1396
TAL.CM.8023                ------GGC-G--C----T---G--A--C--C--A--G---------------CA-C-----T---TAC-CTTCAGATC-GC-G---AG-AAG-----GG--A--T-C---C-GG---CGC---TC--G--  2564
TAL.CM.266                 ----A-GA---------A-T---G-----A--C--A---------G--T--G--CA-C--C------CA--C-TC-GATC--T-A--GAG-AAA--C--GG-CA-GT-C--TC-GG---CAC---T------  3066
MUS.CM.1085                ---CA-GA---------A-T---A-----A---------C--------G--T--TA-C--C-----GTTC-CAGCAG-C---C-------AGAGA---AG---GC---C-----TG-CAC-C--ATC-----  2787
SAB.SN.-.SAB1C             ----C--A------------T--G-----C---------G---------C-----CA----T--T--G-CA-CAGCC-AC-----TC-GC--GAA-----GC--A-G--C-----TG---A---A--------  3495
GRV.ET.x.GRI_677           -------A--G-------T----TGT--G--T------------------T--TA----T-----G-A--CGGCAG----C--T------GAGA-C-A--GG--C-C---T-GGT---AA--TGTC-----  3270
TAN.UG.x.TAN1              -T--A-----G-----A-T----TGT--A--T--A------------------CA-C-----T--GCAC-CAGCACAGCG-G--C--A-AGAAA-AG-----A-GGT--A---TT---AA--AAT---G--  3299
VER.DE.-.AGM3              ----T-C-----------T---CTGT---------A---------G-----C---A-T--T--T-G-AC-CAGCAG-TTCC--TC----AGAAA-A-A---GG-GTTAAA-CC-C--ACC--TGTG-----  2822
VER.KE.-.9063              ----TG----G-----A-T----TGC--------A--C--------G--T---A-T--T--T-G-A--CAGCAG-------T------AAAA-A-A-G-GG--TTAAA-CCAT-GACA--AGTG--G--  3327
VER.KE.-.AGM155            ----C--A---------A-TT---TGC--A--G-----C-----G-----C--GA----T--------AC-CAGCAT-------TC----AGAAA-A-AG---G--TTAAA-C-GC-GACG--TGTC--G--  3319
VER.KE.-.TYO1              -C--------G-----A-T---CTGT--C--G--A--G-----------T--TA----C--------A--CAGCAG--TCC--T--G---GAGA-A-A--G-A-CTTG----CAC--ACC--TGTA-----  2817
SUN.GA.98.L14              ----A-AA------GT---T-----C--A----------C----------AGT--TTGC--C--T--GG-A-CGG-AG-G-GCC--------AAA-------G---GT-------AG-TCAAT-G--C-----  3454
LST.CD.88.447              ------GA------------T-----A--A--C--A---------G--GAGT---TGC--T--T---GC--CAG-AG--GG-C--C--TCAGAA----------TT---C--T------A-CT-G--------  2365
LST.CD.88.485              ------GA------------T-----A--A--C--A---------G--GAGT---TGC--T--T---GC--CAG-AG--GG-C--C--TCAGAA----------TT---C--T------A-CC-G--------  2365
LST.CD.88.524              ----A-AA------------T------T-G-----A---------G--GAGT--TTGT--T--T---GAA-CAG-T-GTGG-C--C-T-CAGAG--C--G---TT-------------C-CC----C--G--  2362
LST.KE.-.lho7              ----A-AA--------------------G-----A--G-----G--GAGC--TTGT-----T--GG-A-CAG-AG--GG-C-TC-CTCAGAG-----G---TT-------------GA-C--T--C-----  3449
SYK.KE.-.SYK173            -G--AGAA---------A-TT-CA---T-A----------------------T--CA-T--C---------C-CA--A-ATC-G--AC--C----A---------GA-GT--T----TC-CAC-C-AATA--G--  3291
SYK.KE.KE51_AY523867       ----AGAG-----C----T---G---T-G-----A---------G-----G--CA-T--C--T--GGCA-CAG-A-ATC-G--AC-CC-A---A-A--GG--A-G---T----T---A-A--AGTA--G--  2949
COL.CM.-.CGU1              ------C-------GT------A-----A--G-----T---------G--T---A----C------G-A-CAG-AG---G----C----A--G--C--G--GAG-GGA-A-TTG...CAA--AGC---G--  2869
```

```
                        Asp185  Asp186 catalytic site
                             |  |
H1B.FR.83.HXB2          CATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCA  3227
Pol                     __M__D__D__L__Y__V__G__S__D__L__E__I__G__Q__H__R__T__K__I__E__E__L__R__Q__H__L__L__R__W__G__L__T__T__P__D__K__K__H__Q__K__E__P__P_    RT

H1D.CD.84.84ZR085       ---------------------------T-----------A---------G-----------A-AT-A---G-------------------G---------------------------------------  2749
H1F1.BE.93.VI850        ---------------------G------------------------------------------T-A---G-------AC----A------T----T---------------------------------  2564
H1G.SE.93.SE6165        ------------A------------------------------------G--------------T-A---G-------A----AA-----GT-------------T------------------------  2624
H1H.CF.90.056           ------------------------------------------A------G--------------T-A---GCT---T------AA------T----------------------------------C---  2574
H1J.SE.93.SE7887        ---------C--------G-----------G-------AA----------G-------A----------GG-------A----A-------T-------------T------------------------  2541
H1K.CM.96.MP535         ------------A--------G--------------------A-C------------------A--A---G-------A----AA------T-------------------G--------------C---  2423
H101_AE.TH.90.CM240     --A------C-----------------T-----------------C--------------------A---GCT-----A-----C------T----T--------------G--------G---------  2801
H102_AG.NG.-.IBNG       ------------A--------------T---------------------G--------------T----GG-------AC---AA------T-------------------G------------------  2752
H103_AB.RU.97.KAL153_2  ---------------------------------G------------------G----------A------G--------C-----------T--------------------------------------  2453
H104_cpx.CY.94.CY032    T--------------------G--------------------A------G----------A-----A---G-G-----A-----A------T-C--------------------------------C---  2593
H1N.CM.95.YBF30         ---------CC-C-----G---------------C--CA--A------GAGGC-G----A--C--T---G-T-----T----A------CT----G--C--T-----------------G-----C--G  2821
H1O.BE.87.ANT70         ---------CC-A-----------A--T---CCCC-GACAG-A-------A--GG--T--ATT---T---G----C--A-ATCA-------T-C--T--C--T--------G-----A--G--------C  3282
H1O.CM.91.MVP5180       ---A-----C--A-----------A--T---CC-T-G-CAG-A-------AG-GGG-C--ATT---T--GG-----T-A-ATCA-------T----T--C--T--T-----G--------G--------C  3257
CPZ.CM.04.EK505         ----------C-C--------C-----TC--------ACA--A------GAGGC-G-G--AAGA--T--GAGT-----T----CT------T----A--C--T--------G--------------A---  2799
CPZ.CM.04.LB7           ------------A--------G--A--T--GA------A---A--C--G-AC--GG----A---------GCT--CT--C-A-A------GT-------C--T-----------------G-----A---  2782
CPZ.CM.04.MB66          T-----------A--------C---------A------A--TA------GA----G-G--A--AT----GGCT--C--A----AA-----TT-C-----T--T--------G-----A--G--G--A---  2773
CPZ.MT145               ---------C--------G--C--A--T---CCT----A-G-A-------AG-G-G----A-----C---GC---------A-AA------T----A-----------G--G--------------C---  2769
CPZ.CD.-.ANT            ----------C-C--------------TA-G-----TACTGCA------GA--TG-----AA----T---------T-ACA-GTC--------AGAG--T--T--------G------------------  2666
CPZ.CM.-.CAM3           T---------C-T-----G--------T--G---C--A-AG-A------GA---GG-----CTA--C-----G-----AC-T-CT-----GT----T--C--T--T-----G--------G--G--A--T  2638
CPZ.CM.98.CAM5          T--------CC-T---------------C-G--G-----AA-A------GA---GG----ATTA--C------------C-T-CT-----TT----T--C--T--------G-----A--G-----A--C  2929
CPZ.GAB_X52154          ---------CC-A-----G--G-----TC-T-----T-AT--A-------A---GG-G--A--A--A---------T--C-T-AA-----GT-C--A--C-----------------A--G--G--A--C  3288
CPZ.GAB2                ---------C--A-----------A--T---TCTT-G-AAG-A--C--G-A--G-G----AC-A--C---G-----T-AC-C--A-----TT-------C--T-----G--G--C--A--G--------T  2620
CPZ.TZ.-.TAN1           T---------C-T--CA----T--A--TC-TA-TGA--A-G-A-----G-A-CTG---A--A----------G--------A-CA------T-AGAG--C--T--------GT-----G--A--------  2871
CPZ.US.85.CPZUS         T---------C-C--------G--A--TC--A-CT---AAA-------GGA---GG----ACT---C------T--T--C-T-CT------T-C--T--C--T--T--G--G------G-G-----A---  3287
GSN.99CM166_AF468659    ------------ACTGA-------A--TAGGA--C-G--AG-A--C--GCA-GT-G-GC-CA-AA-A---A-TTTA--AACATCC-ATAACA-CCAA--C--C--GGCC-----C--ACC---TTA---G  2955
GSN.99CM71_AF468658     ---------C----TGA-------A--TAGGA--T-G-CAG-------GCA-GT-G-GC-CA-AA-A---A-TTTGT--AC-TCC-ATAACA-CCAG--C--C--GGCC-----C--AC----TTA----  2937
H2A.ALI_AF082339        ----------A-C-TAA---CTAG----AGGAC-GATTTAG-A---GAC-AGGTGG-CCT-C--T-A-AGG---TC--AAATG-CCTA---T--T----C-----TG-G--GTTC--A-----C------  3605
H2A.BEN_M30502          ----------A-C-TAA---CTAG----AGGAC-GGTTTAG-----GAC-A-GTGG-CCT-C----A-A-G---T---AAATG-CCTA--GT--T-T--T-----TG-G--GTTC--A-----C------  3612
H2A.ST_M31113           ----------A-C-TGA---CCAGC---AGGAC-GATTTAG-A---GAC-G-GTGG-TCT-C----A-AGG---T---AAATG-CCT----T--T----C-----TG-G--GTTC--A-----C------  3057
H2B.D205_X61240         ---------CA-CCT-A---C-AG----AG-AGTGATCT-G----CGAC-GGGT-G-GTCCC-A--A-A-G-GTTAT-AAATGACAT----T-CT-T--C-----AG----GTTC--A-----C-----G  3585
H2B.EHO_U27200          ---------CA-TCTC--G-C-AG----AGGAGCGATCT-G-----GAC-GGGT-G-GTCTC-A--A-A-G-G-TAT-AAAT-ACAT----T-CT-T--T-----AG----GTTC--A-----C------  3583
H2G.ABT96_AF208027      T--------CA-TCT---G-CCAG----AG-ACTGATTTAG-A---GATCA-GT-G-TTT-C----A-AGG---TCT-AAAT-AAACR--GT-CT-T--C--T--TG----GTTC--A-----C------  2939
H2U.FR.96.12034         T--------CA-TCT-A---CCAG---TAG-A-TGATTTAG-----GATCAGTT-G-TTCAC----C-AGG-G-TA---AAT-A-ACA---T---G---C------G----TT----------C--A---  3094

MAC.239_M33262          T--------CA-C-TAA---CTAG----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC--TCA-AGG---TCT--AAT--CATA--GT--T-T--C-----AG-G---TTC--A-----T--C---  3529
Pol                     __M__D__D__I__L__I__A__S__D__R__T__D__L__E__H__D__R__V__V__L__Q__S__K__E__L__L__N__S__I__G__F__S__T__P__E__E__K__F__Q__K__D__P__P     RT

SMM.SL.92.SL92B         T-----------ACTAA-T--CAG-A--AG--GCC--ACAG-----GAC-A--TGG--ACAC--T-A---G-CATG--CAAC-ATCT---GT-C-GT--C-----AG-C--GTT---A--GA-T--A---  2960
SMM.US.x.PGM53          ---------CA-CCTAA---CTAG----AG-AC-GATTTAG-----GAC-GGGT-G-TTTAC--T-A-AGG---T----AAC--CATA---T--T----C-----AG-G--GTTC--------T--C---  3446
MNE.US.-.MNE027         T--------CA-C-TA----CTAG----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC--T-A-AGG---TCT-AAAT--CATA--GT--T-T--C-----AG-G--GTTC--A-----T--C---  2997
RCM.NG.x.NG411          T-------------T------G--A--TAGGAC-GC-CATG-A-----TCAG-TG--CA-AAG---C---G-G------C---CC-----TT--GAG--T-----T--G---TT----G-CA-G-----C  2721
RCM.GAB1_AF382829       ----------C-C-T------C--A--TAG-AC--AGTT-G-A---GAC-A--TG--CA-AC-AT-A---G-T-----AC--TTC-----CT--GAG--C------------TT----G-TA--------  2706
DRL.FAO_AY159321        ----------C---T-------AG---TAG-CC--A-CATTTA---GAGCAG-TG--TA-A--AT-A---AC----T-AAA-TTC-----G--CGAG--C--G---------TT----G-CA-T-----T  2794
MND-1.F17               --------------T------G--A----ATAC-GC--AAG-----GAG-A-GCT----TA--AT-A--GGCTTTAT-AA---CA---AACT-AGAA-----T--A--G---T----------------C  2914
MND-2.5440_AY159322     T--------C----T---G--GAG---TC--AGTGCCTCTG-A--C--T-A--TG----TA--AT-A--GG-----T--CA-TTC-----G--CGAG--C-----------GTTC--A--G--------C  2788
MND-2.14CG_AF328295     ----------C---T---G---AG---TC-GAGTGCGTCAG----C-AC-A--TGG--A-TAG-T----GG-----T-AAG-TTC-----G--CGAG--C-----------GTTC--A-----------T  3155
MND-2.MND2_AF367411     T---------C-C-T---T--GAG---TAG--TTGCCCCAG-A--C--C-G--TG--TC-A---T-A---G-C--C--C---TTT-----G--CGAG--C--------G--GTT---A--G-----A--C  3228
DEB.CM.CM40             T-----------ACTA-----G--A--T---AGG-A--AT--A--CCT--AG--GG---CAC-AT-A-----GTTC--CC-AGA-A-----T-G-GG------CCTG-G--GT-----G----C--A--C  3036
DEB.CM.CM5              T--------CC-A-TG--------AA-T-----T-AG-TCTCA---TT-GA-C-GG--A-AC----T--GA--TT---CC-AGAAAA---GT-A-GG-----CCCGG-G---T-C---G-G--T--C---  3030
MON.CM.L1               ---------CC-ACTGA-------AA-TCACTCTC-GCCAG-A--C--G-A--T-G---G--TCA-A---GCGACC--AC-A---AAA--GA--CAA-----CCCAG-G---TT----G-T--CTAC--G  2933
MON.NG1                 ----------C-ACTCA----T--A---CATTCCC--CCTG-A--C--G-AG-T-G--ACA-CCA-T---ACG-TR--A-CAGAAAA---GA--CAA------ATAG-G---T-----G-CA-GTAC---  1526
TAL.CM.8023             ---------CC-ACTAA-----AG---TGA-AGCT---AA--------G-G--T-G-G-GA-----T---GCC-TC--TAAC-A-GTA--GA-CCAG-----G--GGCC---TTC---G-CA-G-----T  2694
TAL.CM.266              ---------CC-A-TG------AGCA-TGAGAGCT-G-A-G-A-------GG-T-G---GA--A--A---GCT-TCT-AAAC-A-GTA--GA-CCAG--C-----GGCT---T-C--AG-CA----G--C  3196
MUS.CM.1085             ---------CC-CCTGA-------CA--AG----T--TCAAGC--C--G-A-GT-G---CAC--A-T---A-CATG--C--A-A-CTT--GA--CAA--C--TC-AG-----TT----G----CTAC---  2917
SAB.SN.-.SAB1C          ----------A---TAA---CCAG---TAGGCC--AG-CAG-A---CT-GT--TGG--C--C--T-A---G-CT----AGAA-CT-----GT---AA--C--T--A--G--GTT---A--G--T--A---  3625
GRV.ET.x.GRI_677        ------C-------GGT-G-CG--A---CAT--TGAGACTAGA----ATCA-C-GG----CATAG-A---A-GATG---C-AGAAAAA--T--AGAA--C--------G---GTC--A-G------G---  3400
TAN.UG.x.TAN1           ---------CA---GGA-------A--T--GCC-GAGAAAG-A---GATGA---GG----A---G-A---A-G-T----C-ATTAA-A--CT-CGAG--G-----T--G--GGTG------AG-------  3429
VER.DE.-.AGM3           ---------CC-A-GG-----G---C-GGA---TGA-TACAC----GATCGGTTGG----AC-A--A---ATGA-AT-AAGTGCC------T-AGAA-----------G---GTA------A----A--T  2952
VER.KE.-.9063           --------------GG-----G---C-AGA---TGA-TACACC---GAC-AGTT-G----AC----T---GTCA-G--AAAA-CA------T-AGAA--C-----------GGTG--A--GA-G--A---  3457
VER.KE.-.AGM155         ---------CC-C-GG--------AC-AGA---GGGTCCAA-----GATCAGCT-G--C-AACA--T--GA-TAGAT--CAAGAA------T-AGAA--------G-----GGTG--A-G---------C  3449
VER.KE.-.TYO1           ------------A-GG-----T---C-AGA-A-TGA-CACACC---GAC-A-TT-G----AC--T-A---AC-A-AT-ACAAGCC-----CT-AGAA--C-----A--G--GATG--A--------A--T  2947
SUN.GA.98.L14           ------------ACTCA----------T-AT--G-A-AA-A-----GA-GA--T-G-GA-AC-AT----G--G-TA--CA--GAA---AATT-AGAA--------G------T-----GG---G--A--C  3584
LST.CD.88.447           ------------A-TCA-------A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-----G-T-TCT-AG---CA---AATT-A-AA--------G--G---TT---AGC---------C  2495
LST.CD.88.485           ------------ACTCA----G--A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-A---G-G-TCT-AA---CA---AATT-AGAA--------A--G---TT---AGC---G-----C  2495
LST.CD.88.524           --------------TCA----T--A---AGG--G-A--AAGGA-----TCAGGC-G--A-A--AT-A---G---T-T-AA-A-CT---AACT-AGAA--T-----G--G---TT---AG-CA----A--C  2492
LST.KE.-.lho7           T-----------A-TCA-------A--TAG---G-G-AAAGGA---GATCAGGC-G--A-A--A--C---G---T---TA---CA---AACT-AGAG--------A--G--GTTC--AGC---G--A--C  3579
SYK.KE.-.SYK173         ---------C--A-TGA-T--CA----TAG-TC-GAGAA-GCA---CAGGAG-T-G--C-AC-AA-AGTGAC-GCA--A--A-AAGT----T---AGGTC---A-AG-----TGG---G-CC--TA----  3421
SYK.KE.KE51_AY523867    ---------CC-T-TG-----G---A--AG-TC-GA-AA-G-A---G-TCA--TTG--AGT--AA-TGT-A--AG---A--AGCAGTA---T-CT-T-TT---CCAG-----TGG---G-CA--TTC--C  3079
COL.CM.-.CGU1           ------------A---C-------A--TC-GCCTT---A-G---------GGGC-G--C-------C---G-TGCA--C---TAC-AT--GT-AGAG--------A--G---T-C---GC---T------  2999
```

```
H1B.FR.83.HXB2              TTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGC......TGGACTGTCAATGACATACAGAAGTTAGTGGGGAAATTGAATTGGGCAA  3351
Pol                         _F__L__W__M__G__Y__E__L__H__P__D__K__W__T__V__Q__P__I__V__L__P__E__K__D__S__.__.__W__T__V__N__D__I__Q__K__L__V__G__K__L__N__W__A__    RT

H1D.CD.84.84ZR085           --T--------------------------------G------------T-----ACA--------G-----A---......--------------T-----------------A-----A----------  2873
H1F1.BE.93.VI850            --------------GC----------------------------G---------CAAT-----A-C--G------......-----------------------------A--A---C-A----------  2688
H1G.SE.93.SE6165            --------------A-----G-----------C--------G-----A------CA---------C--G--A---......--------------T-----A-----------A---C-A----------  2748
H1H.CF.90.056               --T-----------A-----------------C---------------A--G--AAA------------------......--------------T-----------------A---C-A----------  2698
H1J.SE.93.SE7887            --T-----------A-----G--T--------C-----------C--A------CAA--------------AGA-......--------------T-----A-----------A---C-A----------  2665
H1K.CM.96.MP535             --T-----------A-----------------C-----------G---------CAA--------C--G--T---......--------------T-----------------A-----A----------  2547
H101_AE.TH.90.CM240         --------------A-----------------C-G---------C----------AA------------------......--------------T--------A--------A---C-A----------  2925
H102_AG.NG.-.IBNG           --T-----------------------------------------C------G---AA--A---------------......--------------T--------A--------A---C-A----------  2876
H103_AB.RU.97.KAL153_2      -----------------------------------------T---------------T-----------------......------------------------C-------A----------------  2577
H104_cpx.CY.94.CY032        --T-----------G-----------------C-----------G---------CAA-C-G-------G--T---......-----------C--T--C--------------A---C-A----------  2717
H1N.CM.95.YBF30             -----C--------A--------------A--C-----------C-----A---AA-T-A--------G--TGTA......--------------T--------A-----A--A--G--A----------  2945
H1O.BE.87.ANT70             --T--A--------G-----G--------A--C---------------T-C--CCAAT----TA-C--G--TGTG......-----A--A-----T-----A--AC--A-A--A--GC-A----------  3406
H1O.CM.91.MVP5180           --TT-A--------A-----G-----C--A--C--G--------------C--CCAAT----T--C-----AGTG......-----A--A-----T-----A--A-----A--A-----A----------  3381
CPZ.CM.04.EK505             --TT-G--------A--------------A--------------C---A-A---CAAT-A-----G-------CA......--------------T-----AC-------A--A-----A--C-----T-  2923
CPZ.CM.04.LB7               --T--C--------A--------------A--C-----------------C---CA-T-A--------G--A--T......-----------------------A--G-----A-----A----------  2906
CPZ.CM.04.MB66              --T--C--------A-----------C--A--C-----------T---------CA-T-A-----------A---......--------------T--C-----A---A-A--A-----A---------T  2897
CPZ.MT145                   --T--C--------A-----GT-A-----A-----G--------C----AA---AAAT--------C-G--AGTG......--------T-----T--T-----------A--C-----A----------  2893
CPZ.CD.-.ANT                --T-AA--------A-----GT-A-----A--C--------T------AAA---AA--A-----GCC---TGAT......-----A--T--------C-----AC----A--A-----A----------  2790
CPZ.CM.-.CAM3               --TT-G--------C-----G--------------G--------G-----A---CA-T-A---C-G-----G-C-......--------------T--T-----AC-------A---C-A----------  2762
CPZ.CM.98.CAM5              --TT-G--------A-----G--------A--C-----------G-----A---CAAT-A---C-G-----G-T-......--------------T--T-----A-----A--A-----A----------  3053
CPZ.GAB_X52154              --TT-A--------A-----G-----C--A--C-----------C--------TCAAT-A-----------GGTA......--------------T-----A--AC-GA-A--A--G--A----------  3412
CPZ.GAB2                    --T--C--------G-----G--T-----A--C-----------G-----A---CA-----TC-G-----G-AT......--------A--------T-----AC-G--A--T-----A----------  2744
CPZ.TZ.-.TAN1               ---A-G--------C-----G--A-----AA-------------T--AAA---CACAT-A------CC---GCAG......-----A--G---C-T--C--------G--A--C---C-T--------C-  2995
CPZ.US.85.CPZUS             ---T-A--------C--------------A-----G--------G-----C---CA-T-A---C-------A-TT......-----A--------T--T-----A-----A------C----C-------  3411
GSN.99CM166_AF468659        G--AAG---C-C--G-----GT-G-----A--GGG-----GGC-C-----GG-G-AA--C-----CC-G--T-C-......-----A--G--------C-----A--G--A--------A-------T-T  3079
GSN.99CM71_AF468658         G--AGA---C-C--G-----GT-G--C--A--GGG-----GGC-C--A--AG-G-A---C-----TC------T-......-----A--A--------C--A--A--G--A-----GC-C-------T-T  3061
H2A.ALI_AF082339            -A-AAA--------C-----G--GTGG--AAC--------AGC-G--AAAA---CAAT----CC-G-----AGTA......-----A-----------C--A--ACAT-----TGTCC-A---------G  3729
H2A.BEN_M30502              --T-AA--------C-G------ATGG--AAC--------AGC-G---AAAC-CAA------CC-G-------TA......-----A-----------C--A---C-------AGTC--A--------GG  3736
H2A.ST_M31113               -A--AA--------C--------GTGG--AAC--------AGC-G--AAGA---CAAT----CC----G--AGTA......-----A-----------C--A--AC-G-----TGTCC-A---------G  3181
H2B.D205_X61240             ---AAA--------------G---TGG--AA-A--G----A-C-G--AAAA---CAA--------------AGTT......-----A--G----CA--T--A--AC-G--A--AGT---A--C------G  3709
H2B.EHO_U27200              ---AAA--------G-----G---TGG--AA-G-------A-C-G--AAAA---CA---A-----------GGTT......-----A--A--------T--------------AGT---A---------G  3707
H2G.ABT96_AF208027          -AT-AA--------A------T-RTGG--AA-A--G----A-C----AAAG--TAAC-----M--------G-CT......-----A--A-----T--T--A--AC----A---GT-C----Y------G  3063
H2U.FR.96.12034             CA-AAA--------C--------ATGG---A-A--G----A-T-G---AAA--T-A---C-------G---TGTT......-----A--A--------T-----------A--AGT-C-A---------G  3218

MAC.239_M33262              --T-AA--------G--C---T-GTGG--AACA-------AGT-G--AAAG----A-T-----C---G---G-C-......-----A--G-----T--------------A--AGT---A---------G  3653
Pol                         _F__Q__W__M__G__Y__E__L__W__P__T__K__W__K__L__Q__K__I__E__L__P__Q__R__E__T__.__.__W__T__V__N__D__I__Q__K__L__V__G__V__L__N__W__A_     RT

SMM.SL.92.SL92B             --A-AG--------A---TTGT-AT-C--CA-G--G----A-T-----AAA----AA--C--T-----G--GC-A......-----A--A-----------A--AC-------AGTTC-A---------G  3084
SMM.US.x.PGM53              ----AG--------A------T-GTGG--AACC-------A-C-G---AAA----A-T-----C---G---A-C-......-----A--A-----------A--A-----A--AGT-C-A---------G  3570
MNE.US.-.MNE027             --T-AA--------G------T-ATGG--AACA-------AGT-G--AAAG----A-T-----C-------G-C-......-----A--G-----T--------------A--AGT---A---------G  3121
RCM.NG.x.NG411              --TGAA--------C----TG--------A--C--------------AGAG--CAA-T-A-T---------AGAA......-----A--A-----T--------A--G--A--------A----------  2845
RCM.GAB1_AF382829           -ATT-G--------G-----G--G--C--AA-A-GT------------GAG--CAA-T-A-----G-----AGAA......-----A--T--C--T--T-----A-----A--A--G--A--C-------  2830
DRL.FAO_AY159321            -ATGAA--------G--CC-G--G-----AA-G--G--------G--AGAA--TCGCT----T--T-----A-TT......--------T-----T--T-----A-----A--AGTGC-T--------T-  2918
MND-1.F17                   --T-A---------G-----GT-A--C--A-----G----AGA--G-AAAGG-TCAA--A------TT--CAGAACAG...CCA--A--A-----A--------A--G--A--T-----A--------TG  3041
MND-2.5440_AY159322         --TGAA--------A----TG--G--C--AA-G-----------T---AAA---CA---A-----------A---......-----A--A-----T--T-----A-----A------C-T----------  2912
MND-2.14CG_AF328295         --TGAG--------G----TG--G--C--AC-G-----------C---AAAG--CA---------------G-CA......-----A--A-----T--T-----AC-G--A--A---C-C----------  3279
MND-2.MND2_AF367411         --TGAG--------A--CAT---G--C---A-G-----------G---AAAG--CA---T-----G-----AGAA......-----A--A--------C--A-----G--A------C-T----------  3352
DEB.CM.CM40                 --T-AC--------G-----GT-A--C---A-A------CGGT-GAT---AG-TACC--A-----TG----AGAA......-----A--GC--A-A--T-----------A--AC-G--A----------  3160
DEB.CM.CM5                  -A--A---------G-----GT-A-----AA-G------CACT-GTT---A----A---A------G-G--GGAG......-----A---C-CA-A--C-----AC-G--A--AC-G--A----------  3154
MON.CM.L1                   G-G-AA---T-A--G-----G--A--C--ATCCC-T----A-A-TGCCAAA----A---A-----CC-G--ACAA......--------------G--C--A-----GA-A--A--GC-T---------G  3057
MON.NG1                     --G-AG---T----G-----------C--CACCC-T--------C--AAAG---AA----------C-----CAG......-----A-------GG--C--A--A-----A--------A--C------G  1650
TAL.CM.8023                 A-TAAG---C-T--G--C--G--GT-C--A--G-------A-------AAA----A---A---------C-G-CA......-----A--A-----TT--------C-G--A--A---C-A----------  2818
TAL.CM.266                  A-TAAA---C-A--G------T-GT----A-----G----A------------T-AAT-------T---CT--CA......-----A--------GT-----------A-A--A---C-T---------T  3320
MUS.CM.1085                 G---AA---T-A--G-----G--A--C--CCGGGGC----A-A-------CG---CAT-----C--C----AGTT......-----A--T-----TG-G--A---C-G--A--------A---------G  3041
SAB.SN.-.SAB1C              -A---C--------G------T-GT----AA-G------CA-C-G---GAG--CACC--A-------GG--AGAA......-----A--G-----T--T--A--A-----A--A--G--A----------  3749
GRV.ET.x.GRI_677            -GGGAA--------G---A--T-G-----GA----------CA-TA-CAAA----AAT-A--CCCCTT---AG-A...GAA-----A--A--CA-A---------G----A--AGTTC-A----------  3527
TAN.UG.x.TAN1               -ATGA---------A--CA-GT-GT-C---TC-C-C----GTC-CA-CAAA----A-A-A-----GC----ACA-......-----A--G-----T--T--A--A-----A--AGTG--A--------T-  3553
VER.DE.-.AGM3               -ATGAG--------A--CA--T-GTGG--AC-C--G---CAGA--AGCAGC----AAT-AGA---C-----AGAA......--------A-----T-----A-GAC-----------C-A---------G  3076
VER.KE.-.9063               -ATGAA--------A--CA--T-ATGG--AC-A------CA-C--A--TAG---AACT-AGA------G--ACAA......--------T-----T--C--A--AC----A-----------C------G  3581
VER.KE.-.AGM155             --TGAG--------A---A--T-ATGG---C---------AGT-----AAG----AAT-AGAGA-G-----ACAA......-----A--G-----TC-T-----A--G--A--------A---------G  3573
VER.KE.-.TYO1               -ATGAG--------A--CA----TTGG---C-C------GA-C--AGCAGA---CAA---GAG--------TGAA......-----------------C-----------T------C-A---------G  3071
SUN.GA.98.L14               -ATAAG--------G--CAT-T-A--C--A--C--G---GAGA--G--AAG---AAAT-A--CCCTCT---AGAAGAG...CCA--A--G--------------AA-T--A--AGTTA-C--------TT  3711
LST.CD.88.447               -AT-A---------G----T---T-----A----GG---GA-A--G--AAA-CTAAA--------C-T----CTTCAGAAGACA--A--T--CA-----------C----A--AGTC------------G  2625
LST.CD.88.485               -AT-A---------G----T---T-----A----GG---GA-A--G--AAAG-TAAA--------C-T----CTTCAGAAGACA--A--A--CA-----------C----A--AGTC------------G  2625
LST.CD.88.524               -A--A---------G----TC--G--C--A-----G---GA-A--G--AGAG--AAA--------G-T-----CAAGGAAAACT--A--A-----A--T-----AC-G--A--AGTT--A---------G  2622
LST.KE.-.lho7               -AT-A--------------T---G----------GG---GA-A--G-AAAG--TAAAT-A-----G-TG--TCTAACAAAAACT--A--A---C-A--------AC-G-----AGT-C-T---------G  3709
SYK.KE.-.SYK173             A-G-AG---T-A-----CACT--T-----A--C--G---CAGT-G---AAA----AA--C--TA-T-T---TGA-GAA...ATC--A--G---C-GT-------A--GA-A--AGT-C-T----------  3548
SYK.KE.KE51_AY523867        --G-AA---T-A--G--CACG--A--C--------G----GTC----AAAGG-TCAA--A-------TTACAGAATCA...CCA--A--T-----GT-------AA--A-A--AGTCC-A----------  3206
COL.CM.-.CGU1               -A-AAG--------C------T-A------A-GC------A-T----AAAGG-CAAAA-T------C-G--T-AG......-----A--A-----T-----A--A--G--A--AGCC--A------TT--  3123
```

```
H1B.FR.83.HXB2            GTCAGATTTACCCAGGGATTAAAGTAAGGCAATTATGTAAACTCCTTAGAGGA...ACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGAGAGATTCT  3478
Pol                       S__Q__I__Y__P__G__I__K__V__R__Q__L__C__K__L__L__R__G__.__T__K__A__L__T__E__V__I__P__L__T__E__E__A__E__L__E__L__A__E__N__R__E__I__L    RT

H1D.CD.84.84ZR085         -C--------T-----A----------A---------------------G---...-----G-----------G--------------------------AT-------------------G--------  3000
H1F1.BE.93.VI850          ----------T-------------------C------------------G---...G----G-----------CA--G-G-----G--T-C---G------T-----T-----A----T--G--------  2815
H1G.SE.93.SE6165          ----------T----------------C---C-----------------G--G...G----------------CA--G--T----G----C---------AA-G--GT-------G-----G--A-----  2875
H1H.CF.90.056             ----------T---AAT----------A----C----------------G--G...G--------T-------CA----------G---A----G-----AT-G---T-------------G--------  2825
H1J.SE.93.SE7887          ----------T-----A----------A----C---------------A---G...G-T--------------CA--G-----T-G--TAG---------AT-G----------------A---------  2792
H1K.CM.96.MP535           ----A--A--T----------------AA-----G--------------G---...GT---------------CA--G-----T----T-C----------T-----T-A-----G--------------  2674
H101_AE.TH.90.CM240       ----A-----TG----------G----A----C-G-----------C--G---...G-T--------------CA--G-------G--T-----------AT----GT-------G-----G--------  3052
H102_AG.NG.-.IBNG         ----------TG----A------A---A----C-G----G------C--G---...G----------------TA--G--G----G--T--G--------AT-----T-------G-----G--A-----  3003
H103_AB.RU.97.KAL153_2    ----------TG-------------------------------------G---...G---------------------------------C------------------------------G--------  2704
H104_cpx.CY.94.CY032      ----------T----------------A---------------T-----G---...G-T-----C--------CA--G----------TAC---G------T-----T-A-----G-----G--------  2844
H1N.CM.95.YBF30           -------C--T-----A--C-G-----AA--GC-C------T-AA-C------...--------TT-G---------G-CAACT-T--------------AT-------A-----------G-----AT-  3072
H1O.BE.87.ANT70           ----A--C--T-A---A----G---G--AG----G-----GT-AA-------C...-----GT--T-----------G----TT---GTAG---G--------G---T-A-AG-----------A-GGT-  3533
H1O.CM.91.MVP5180         ----A--C--T-A---A----G-----AAG----G--C--GT-AA-C------...------T--T-G-----G---G----TT---GTA----G-----A------T-A-A------------A-AG--  3508
CPZ.CM.04.EK505           -C-----C-------------------AA---C-T--C---T-AA--------...G-A-----T------------G-CA----T---AG---------AT----GT-A--T--G--T--G-----AT-  3050
CPZ.CM.04.LB7             ----A--A--------C--C--GA---AA--GA----C-GG--T-----G---...GTA--GAATTC----------G-CA-CT-T--TAG---------AT-----T-A--------T--G--A--AT-  3033
CPZ.CM.04.MB66            -------A--T-----A--C-G-----AA---C-G--C--G---A--------...--A-----TT----------TG--A-CT-T---AC---------AT-----T----------T-----A--AT-  3024
CPZ.MT145                 ----A--A--T-----A------A-T-A---GC-T------T-AA----G---...G-G--G-GT--G---------G--G-T--C--CCC------T--A-----------------TCA------A--  3020
CPZ.CD.-.ANT              -------C--------A--C---ACT-A---G--G----G----A-C------...GT---GAGT--------TAG-G---A-A-G--TAGG--------AT-----T-A-A------T-A-C-A-----  2917
CPZ.CM.-.CAM3             ----A--A--T-----A--A-------AA--G-----C--G--GA-A-A---G...--A-----T--------CA--G-GA-CA-G---C-G--------AA-G---T-A-A---G-----R------T-  2889
CPZ.CM.98.CAM5            ----A-----TT-------A-------AA--G-----C------A-A-A---G...--A-----CT-G--T------G--G-CA-G---C-G--------AA-G---T-A-A------T-----A---T-  3180
CPZ.GAB_X52154            ----------T-----A--A---A---A------------G--GA-A------...--A--GAA---G-----T---G-T--T--C---CC---------AT-----T-A--------T--G-----AG-  3539
CPZ.GAB2                  -------C--T-----A-----GA---AA-----G------T-AA-C------...G-T--G--TT--------A--G-CGAGA-G--TAG---------AT----GT-A-A------T--G--A---T-  2871
CPZ.TZ.-.TAN1             ----A-----T-AT--A--A---ACT-AAG--C----C---T-GA--------...GTA----G-T----T--GCC-G--GA-A-G--CAGG--------AT-G--GT-A-A------T-AGC-------  3122
CPZ.US.85.CPZUS           -------A--T-----A--A---A---A---G-----C---T-AA-A-A----...G-T-----T-----T-----TG--AATT-T---C-T--G-----AA-G--GT-A-A------------A-----  3538
GSN.99CM166_AF468659      CA--A--AC----T--T--A-GGAC--AA--G--G--C---TGTA----G---...G-T----GG--------CCAGG--GAG--C--T--G--G---C--A-----T-A--------T--G-----CT-  3206
GSN.99CM71_AF468658       CG--A--AC-T--T--C--A-GGAC--AA--G-----C-G-TGTA--------...G------GG---------CAGG-GGAG--C--T--G------C-AA----GT-A--------------------  3188
H2A.ALI_AF082339          CA--A--C--------A--A--GACC-AA--C-------G---AA--------...-AA-TGA----C-------G-G-G-AGTGG------CT------AGC------A-AG--G------ATT--CT-  3856
H2A.BEN_M30502            CA--A--C--TT----A--A---ACC-AA--C-------G---AA--------...-AA-TGA----C-------A-G-G-AGTGG------CT------AGC---G--A-A--------A-ATT--CT-  3863
H2A.ST_M31113             CA--A--C-----------A--GACC--AA-C-------GGT-AA-C------...-AA-TGA----C-------AGG---AGTGG------TT------AGCG-----A-A--------A-ATC--CT-  3308
H2B.D205_X61240           C---AC-C-TT--T--A-----GAC------CA----C-----AA----G---...-AG-TGA-C----------A-G---AGTGG------CT------AGC---G--ACAG--G--T-A-ATC--CT-  3836
H2B.EHO_U27200            C---AC---T---G--------GACC-----TA----------AA-A--G---...-AG-TGA-C----------AGG---A-TGG--T---TTG------GC----T-CCAG-------A-ATC--C--  3834
H2G.ABT96_AF208027        C----C-A--T-----A--C---AC--AAA-CC-G--Y---T-AA-------T...-AR-TGA-CT-G--G----A-G---AGTGG------TTG-----AGC----T--CA--------AGATT--C--  3190
H2U.FR.96.12034           CA--AG----T-----A--A--GAC--AA--TC-T------T-AA---AG--G...-AA-TGA----C-------AGG---AGTGG------ATG-----AGC----T-T-----G--T-AGATT---T-  3345

MAC.239_M33262            C---A-----T-----T--A---ACC-AA--TC-C----GGT-AA--------...-AA-TGA-T--------G-A-G-T-AGTGG--T--GATG-----AGC----TAT-AG-----T-A-ATA-----  3780
Pol                       A__Q__I__Y__P__G__I__K__T__K__H__L__C__R__L__I__R__G__.__K__M__T__L__T__E__E__V__Q__W__T__E__M__A__E__A__E__Y__E__E__N__K__I__I__L    RT

SMM.SL.92.SL92B           CA-----A--T------------AC--AAA-CC-G--C---A-GA-C-----G...-AA-TGA-T--G-------A-G-T-AGTGG------TT------AGC----T-------G--T---ATT-----  3211
SMM.US.x.PGM53            CA--A-----T-----A--A--GACT-AA--TC-T--C---T-AA-C------...-AA-TGA-TT---------AGG-T-AGTGG--T--GATG-----AGC----TAT-A--------AGATA-----  3697
MNE.US.-.MNE027           C---A-----T-----T--A---ACC-AA--TC-C----GGT-AA--------...-AA-TGA-T--------G-A-G-T-AGTGG--T--GATG------GC----TAT-AG-----T-A-ATA-----  3248
RCM.NG.x.NG411            -------A--TT----A--A--GAC--A-G-G-----C-----AA-A--G---...GTT--G---T-GGAT--CAAGG--GA-T-T---AG---------AT-----TAT-A--------A--CTG-C--  2972
RCM.GAB1_AF382829         ----A------AGT---C-A-GGACT-A-G----G--C--GT-GA-A--G---...G-A-----T--GAT----AA-G--GA-A-G---A----------AA-----TAT-A--------AGAT------  2957
DRL.FAO_AY159321          ----A-----TT-C--A------AC--A-G-GC-C------T-AA----G---...G-----C-C---GAT----A-G--GA-TGG--TAGG--------A------TAT-A------T-A-CTA--AA-  3045
MND-1.F17                 CA---T-A--T--T-----C---AC--AA---C-G--C--G--AA-A------...GGACT-AACA-------GAA-G-CA--A-G------------AGA--G---TAT-A-C----T-A------CT-  3168
MND-2.5440_AY159322       -C--A--C--TT-C--A------AC--AAG-GC-C------T-GA-A--G--G...G-A-G-C-C---GATA--A--G-GGA-TGG--TAG---------A------TAT-A--------AGCTA--AG-  3039
MND-2.14CG_AF328295       ----A--C--TT-C--A------AC--AAG-GC-C--C---T-GA-C-----G...G-A---C-T---GAT------G--GA-TGG--TAGG--------AT-----TAT-A------T-AGTTA--AG-  3406
MND-2.MND2_AF367411       ----------TT-C--A------AC--AAG-GC-C----G---AA-------G...G-A---C-C---GAT---AA-G--GA-TGG--TAG---------AT-----TAT-A--------A-CTC--AG-  3479
DEB.CM.CM40               -C--A-----TT-------A---AC--AA--C-------GGGCAA-C-----G...GT-CCG-G-T-------CCCGG-TA-C--CT-T--------TC-AGC------TAG---G----AG--A--C--  3287
DEB.CM.CM5                -------A---T----A--A--GACC-AA--C--G--C--GGCAA-C-----G...GTTCC--GG--------TAA-G--AGC--TT----G------C-AGC-------AGG--G--T-A---A-C--C  3281
MON.CM.L1                 CA--A--C--TT----A--A---ACC-AAA-CC-G--C---AGTA-C-----G...G-ACGGAGC--G-----TAC-G-GGTC--CT----TTT---TC-AGCT--G--A--G-----T--G--------  3184
MON.NG1                   CG-----C--------A--C--GAC--A---TC----C-R-TGTA-CC-R--G...RTA-----C-----G--CA--G-GGAG---T----CTT---CC-AGC----T-A--------T--G--A-----  1777
TAL.CM.8023               ----A--A--TT-------C---AC--AAGCCC----C---TG------G--T...GTA----GCT-GTT-----AGG-TA--T-G--------------AGC----T---------------CA-----  2945
TAL.CM.266                -C--A--A--TT-T--T------AC--A-GCCC-C--C-GGTA---C--G---...GTA----G---TCT-----A-G---AGT--T----G---------GC----T-A-----------G--A--CT-  3447
MUS.CM.1085               CA--AGCA--------T--C--GAC--A---------C---TG-A-C--G---...GTA---T-T--G--T----AGG-C-AGT-G-----G------C--GC----T----------TCAG-----C--  3168
SAB.SN.-.SAB1C            ----A------A-------A--GACT-AA--T--G----G---AA--------...G-T-G-C-TT-G-----GA--G-C-A-TGG--------G-----A------T-A-A------T---C----CT-  3876
GRV.ET.x.GRI_677          ----A-----T-----A------ACC-AA--TACC---GCCA-GT-G-----G...-AA--GAAC--CCT-----A----GT-TGG--------G------GC----TATAAGA-C--TCA--G---AG-  3654
TAN.UG.x.TAN1             ----AT-A--T-----A-----GAC--AAGCCC----C---T-A--AC-----...G-TC---ATT------CT---G-G-A-TGG--G--GC-G------G----GT-CAG-C----TCA------CT-  3680
VER.DE.-.AGM3             CA---C----------AC-C-G-ACT-AAA-C--G------T-AA-C------...-AA--GAACT--CT----AC-G--A-CTGG-----G--------AGC----TAT----------A------CT-  3203
VER.KE.-.9063             C----C-------C---T-A-G-AC--AA--GA----C-GGT-AA-C--G---...-AG---AAT---TT---G-A-G-G-A-TGG--T-----G--G--AGC---GTAT-----G----A---A---T-  3708
VER.KE.-.AGM155           CA--AT-A--T-----AT-G-G-AC--AAA-TA-C-----G--A---------...-AG---AATT--TT---C--GG--GA-TGG--CCC---G-----AGC---GTAC-A--------AG-----C--  3700
VER.KE.-.TYO1             CA--AT-G--T-----TC----GAC---A...A----C--GT-AA---C----GGA-AG---AAT--GTT---GC--G--G-TTGG---CCT--G-----AGCT---TAT--------TGC---------  3198
SUN.GA.98.L14             CA--AT-A--TGA----C-C-G-AC--AAG-------------AA-------G...-AT---C--T--G----GAA-G--AAGA-G-----------TAG-GA----TACCAGAGT--TCAG---G-AT-  3838
LST.CD.88.447             CA--AC----TGAG--A----G-ACT-AAG--C-T------T-AA----G---...GTA-----T---GAT-----TG--A-CTGG-----T-C----TTAGA----TAT-A-C-----CA-C----A--  2752
LST.CD.88.485             CA--AC----TGAG--A----G-ACT-AAG--C-T------T-AA----G--G...GTA-----T---GA------TG--A-CTGG-----T-C----TTAGA----TAT-A-C-----CA-C----A--  2752
LST.CD.88.524             CA--AC-G--TGAT--A--C-G-AC--AAG-GC-C-----GT-GA--------...GTA--GC-CT-GGA-C--C--G-GA-TTGG-----G------CTAGA----TAT-AGACC--T-AGACA-----  2749
LST.KE.-.lho7             C---AT-G--TGAT--T----GGAC--AAG--C-C--C---T-AA-A--G---...GTA---C-CT-GGA-----A-C--AACTGG-----G-----CTT-GA----TAT-GGC-G----A----G-A--  3836
SYK.KE.-.SYK173           -C--------TT----A-----GAC--AAG-GC-C-----GTGTA-C------...--T---CA-T-----------C-CA----T-------C---T---GC----T-A-A---G--T---C-A--CT-  3675
SYK.KE.KE51_AY523867      ----A--A--T--T--T--C---TGT-AAGC---GA----GTG-A-A-----G...G-A--------T--C--CA-TG--GA---G--------G-----AGC----A----T-----T--------A--  3333
COL.CM.-.CGU1             -C--AT-A--------A--C-GGAC--A-G-GA--A-----TG-AC--A---T...-AG--GCATT-GTTG--G-A-G-G-AG--T-GT-C---------AGAG--G--ACA---T----A-C----AA-  3250
```

```
H1B.FR.83.HXB2           AAAAGAACCAGTACATGGAGTGTATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCA...TTTAAAAATCTGAAAACAGGA  3605
Pol                      __K__E__P__V__H__G__V__Y__Y__D__P__S__K__D__L__I__A__E__I__Q__K__Q__G__Q__G__Q__W__T__Y__Q__I__Y__Q__E__P__.__F__K__N__L__K__T__G_    RT

H1D.CD.84.84ZR085        ---G------A-G------------------------------------------T-------A--A--------T--------------------------A---...---------------------  3127
H1F1.BE.93.VI850         --G---------------G--A-----------------------------------------A--A--AG-C--G-----------------------GA-C---...-----------A---------  2942
H1G.SE.93.SE6165         --G------T-----------C--------------------A------------G----A--A--A----T--A-------------------------------...-AC------------------  3002
H1H.CF.90.056            G-G-------A----------A--------T----------------------------G------A--------------------------------G------...--------------G------  2952
H1J.SE.93.SE7887         ------------------G--A---------T--G-------A------------G-G-----A--A----TG-A------------------------G------...--------C-----------G  2919
H1K.CM.96.MP535          ------G-----G-----G--A-----------------------------------------A--A---A-T-A------------------A------------...CA------------------G  2801
H101_AE.TH.90.CM240      ----ACC--T--G--------A------------------------G--------G-------A--A-----G-A-------------------------------...-----------A---------  3179
H102_AG.NG.-.IBNG        ---------T--------G--A------------A-----------G--------T-------A--A-------A-------------------------------...---------T-A---------  3130
H103_AB.RU.97.KAL153_2   ----------------------------------------------G----------------------A------------------------------------...---------------------  2831
H104_cpx.CY.94.CY032     ------------------G-CA--------------------------------------------A--------T-----------------A------------...CA------------------G  2971
H1N.CM.95.YBF30          ---------CC-G--------C------------GG------A---G----------T--A-----A--A-----T--G-----------G--------G---TT-...CA-------T-A---------  3199
H1O.BE.87.ANT70          ----C-------G-----G--A--C---C-A--TGAT--G--TC--TGG-TTA-T--T--------A--AGGG-AG--------T--C--G--A-----G--AGA-...CA---G--C--C--------G  3660
H1O.CM.91.MVP5180        ------G--------------A-----CC-G--TGAC--------GTGG-TTAGT--T--------T--AG----G--------T--C--GG-A-----G--TGA-...CA---G--C--T---------  3635
CPZ.CM.04.EK505          ---------T--------G-CC------A-----GAC-----A-----------G--C-----------A-----T--------------G-----------CTT-...CA-------T-A--G------  3177
CPZ.CM.04.LB7            ----A-C-----------G--C-----------GG-T------C-G---------G----A--A--A--A--GA-T--------C--------C-----G--A-A-...---------------------  3160
CPZ.CM.04.MB66           ------------------G-CC-----------C--------A-----------------------A--T-----A--G-----T----------TC--G--A-AG...-A---G--CT----------G  3151
CPZ.MT145                --------------------CC--C---------TGA--G--A-----------------------A--A--G--T--G--------------A-T---G------...CA------C--A--G------  3147
CPZ.CD.-.ANT             -C-GC-GAA-A--G-G---TAC--C---C-G--TGGTCT-CCAC---A----ACC--------A--A---TC---A--------C--------A------A-TGA-...GGA---CTGT-A---G-----  3044
CPZ.CM.-.CAM3            --RG--T-----G-----G-----C---------GA---G--A------------N-------A-----AA-TA-T--G-----T----------TC--G--A-AG...CA---G--C--A--------C  3016
CPZ.CM.98.CAM5           ---G--T--T--G--------T--C-----T---GA---G--G-----------GG----A--A-----AA-TA-T--G-----T--C--G----TC--G--A-AG...CA---G--CT-A--------G  3307
CPZ.GAB_X52154           --GCAC------------G--A--C-----T---GAC-----GC-T-----------------A-----CA-CT----G-----T-----G--A-T---G--A---...CA---G--T-----G-----G  3666
CPZ.GAB2                 ------------G-----CACC--C---------GA------AC-----------G-T--A------AATA-T--G--------T-----G----T---G--AA--...GGA-GG-----A---------  2998
CPZ.TZ.-.TAN1            -------AAG--T--A----CA--C-----T--TAA-TT-CCTC-GCA-----C---------------------A--G-----------G--A-----G--AGA-...GGG------T-A---------  3249
CPZ.US.85.CPZUS          ---G--------------G--C------------GA------A---G--------G----A--A--A--AAGGA-T--G--------C-------T------A-GG...CA---G-----A--G------  3665
GSN.99CM166_AF468659     ----C--T-TTC-G-G--T-GC---------G--GA---GCCGC-CG-G-TG-----TACCTCC-TA--AG--CAG------GGG--CATGT-CAGC-----AGGC......--G-TGT-A---T---CC  3330
GSN.99CM71_AF468658      ----C--T-CTC-G-G--CAGC--C--C---G--GA---GCCAC-TG-G-T------TACCTCC-TA--AG--CAG------GG---CATGT-CAG---G--CAGC......--G-TG--A-G-T----C  3312
H2A.ALI_AF082339         --GTC-GGA-CA-G-G--GCAC-----CC-AGA-GA---G--G---GA----AC-G-C--A--AG-TCAAG-CAAT-------------A---AC-C--G-GAGA-......----T---A---GTG-A-  3980
H2A.BEN_M30502           G-GCC-GGA-CA-G-A---TAT-----CC-AGA-GA------A---GAG---AC---C--A--AAGCCAAGG-CAT------------CA---AC-C--G--AGAG......----TC--A---GT----  3987
H2A.ST_M31113            --GCC-GGA-CA-G-A---TGC-----CC-AGAGGA---G---GC-GA----AC-G-C--A--AG-TCAAG-CAAT--G---------A-G--AC-C--G-GAGG-......----T---A---GT----  3432
H2B.D205_X61240          -G--C-GGA-CA-G-A---TCC--C--CA-GGA-AGGGT-CCGC--GA----AC-G-------AA-CCTAGC-AAT--G--------CA-----C----G-GAAAT......---GTC--A---GT----  3960
H2B.EHO_U27200           -G--C--GAGCAGG-A---TCC-----CA-AGA-GGGGT-CCT---GA----AC-G-G-----AA-TCTAGC-AAT--G--------CA-G---C----G-GAGAT......----TC--A---GT----  3958
H2G.ABT96_AF208027       -R--C-GGARCAGG-A--C-CC-----CA-AGA-GR-G-GCCT--RGA---CACTG-------AA-TTTAG-CAAT--G--------CA--G--C------GAAAC......-G--TC--A---GT---T  3314
H2U.FR.96.12034          G-GTC-GGA-CA-G-G--G-CT-----CAGGGA-GA-G--A-TC--GA----AC-G-G-TT--AA-TCAAG-TAAT---------C--CA-G--CC---G-GTGAC......-GG-CAT-A---GT----  3469

MAC.239_M33262           C-GTC-GGA-CA-G-A---TGT-----CC-AGA-GGC--GCCA---GA---CACGG--ATA---AGTCA-G-CAAT--G---T-T---A-----C-C-----AGAC......----TA------GT----  3904
Pol                      __S__Q__E__Q__E__G__C__Y__Y__Q__E__G__K__P__L__E__A__T__V__I__K__S__Q__D__N__Q__W__S__Y__K__I__H__Q__E__D__.__.__K__I__L__K__V__G     RT

SMM.SL.92.SL92B          ---TC--GAGCAGG-A---AGA-----CAGAGA-GATG--CCGC--GA----ACGG-T-TA---A-TCAAG-CAAT--G----GT---A-G---C------GTGAC......-GG-T---C--GGT---G  3335
SMM.US.x.PGM53           C-GTC--GA-CAGG-G---TGT--C--CC-AGAGGG---GCCA---GAG---AC-G--ATA---AGTCA-G-TAAT------T-----A----CC-C-----AGAC......--G-TA------GT---C  3821
MNE.US.-.MNE027          C-GTC-GGA-CA-G-A---TGT-----CC-AGA-GGC--GCCA---GAG--CACGG--ATA---A-TCA-G-CAAT--G---T-T---A-G---C-C-----AGAC......----TA------GT----  3372
RCM.NG.x.NG411           G--G--GAA-A-------G--T------A-----GA----CCA--------TA--G-T-----ATTA-AAGG---G-------GC--C--G--AG-G--G-GAGAC...-CA---CC---A--G-----G  3099
RCM.GAB1_AF382829        -------AAGT-G-----G-----------TGA-AAG---CC----G-G---A-C--T------TTA-AAGGT--A--G---T-C--------AG-A--G---T--...GGA---CCA-----------C  3084
DRL.FAO_AY159321         ------G-A-A-G-----G--A------C-G---GA----CC-C---A----CA-G----A--A-T---TA-T--A------T-------G--AG-A-----TGAT...AAC--GCCA--A---------  3172
MND-1.F17                GGCT---GA-CA-G-A--GTCT--------T--TAAT--G--A---TAT-T-AG-T-T-----AACAACAGG---AG-TATAT---T----TGGA-G----GAAAT......--GGT-T-A-G-G----G  3292
MND-2.5440_AY159322      GC-G--GGA---------------C--CC-G---GA----CCT-----G---A-GG-G--A---TTAACA-----A--G----GT--C-----AG-A--G--AGAC...AAC---CC---C--GGT----  3166
MND-2.14CG_AF328295      GC-G--GGA------------T------C-G---GG---CCTC-G--G--CA-GG--------A-T-ACA--G--A-------GC-----G--AG-A--G--AGAG...AA---GCC---T--GGT---G  3533
MND-2.MND2_AF367411      GC-G--GGA---G--------A--C---C-A---GAG--GCCT-----G---A-GG-T--A--ATT-ACA-----G-------GT--C-----AG-G--G---GAT...AA---GCCG-----GGT---G  3606
DEB.CM.CM40              G--GC--GA---GTCA-----T--C---A-AGAGCAGG--CCA------------T--TCC--ATTA---A----A------GG----GTG--CAGA---CCTAAG......GG--TAT-A--G------  3411
DEB.CM.CM5               C---C-GGA---CTCA--G--A--C--CA-AGA-GATG-GCC-C------T---G--TTCC--AATA--AAG---A--G---GG----GT----AGA---ACAAAG......GG-TTG--A--------G  3405
MON.CM.L1                C--G--GGA-TCCACC--G-GA-----C-----CAACG-GCCA----AG-TG--GT--ACCTCA-TA-CAG--AA-------GGC--CAG-T-C-TC-----AAA-......TC-GTAT----G------  3308
MON.NG1                  --GTC---AGTCGAC---G-GA-----C-----CCAGCTC-----G-AG-T---G--TACCTCC-T--YAG----A--G---GGG--CRGGTA--TC--G--A......RA--GGGT---A--G-----C  1901
TAL.CM.8023              -GC---GGAGACC-----TTCA-----CA-GGA-GG-G--CCA---GA------GC--ACC--ATT--CAG-----------GGC--CATGG-AA-A--G-GAAA-......CGGCTG--A--G-----G  3069
TAL.CM.266               -GC----GA-ACT--C--TACA-----CC-GGA-GG-G-GCC-C-GGAG------C-GACC--ATT--CAG----T------GGC---ATG--AA-A----GCAA-......CGGCTCT-A--------C  3571
MUS.CM.1085              TC--C-GA----C--G--G-CA--C---------GA-C-GCCA--G-G---------TGTCTCA-TA--AG-CTCA--G---GG----A-CT-CACC--GAGC......AAGGGGGTAT-A-GG------  3292
SAB.SN.-.SAB1C           --G-C--AAGCAG--G---CAA-----------TG-TCTTCCT----GG--TA-GG-T-TT--A-T---AG-T---------GG------G--A--C--GCCAGA-...AA-----TA-----GGTT--C  4003
GRV.ET.x.GRI_677         GC-G---A--CA-G-A---ACA--C--------TCTC-----A---------AC-G-T--A-----A--AG----G-----------C--T-CACC------AGGG......GC-GTAT-A--GGTG---  3778
TAN.UG.x.TAN1            -------GAGCA---G----CT--C---------AGT---TCT--G-G---CACT--TACC----T--ATG-T--G------GG-------T-CA-A--G---GG-......---GTA--A---------  3804
VER.DE.-.AGM3            ----ACGGA-CAGG-A--GACC--C--CA-A---GG--G-CC-A-C-G-----C-G-G--A--A-TA-AAGG---T-------GT--C---T-CA-G------......GGAC--GTAT-A---GT---T  3327
VER.KE.-.9063            G---AC-GA-CA-G-A---ACT------CA----GG---GCCTC-T-G-----C-G-C-----ATTA-AAG-T--G-------GT-----GT--A-A--G--A......GGA----TCT-----GT----  3832
VER.KE.-.AGM155          ----AC-GAGCA-G-A--TACT------CA----GA----CC-C-T-GG----C-G-------ATTA--AG-T--G------T----C---T-CA-G--G--A......GGA----TCT-A--GGT---G  3824
VER.KE.-.TYO1            T---AC-GA-CAGG-A---ACC-----CA-A---GG--T-CCTA-T-GG----C-G-------ATT--AAGG---A--G----GT--C---T-CA-A-----A......GGAC--GTCT-----GT----  3322
SUN.GA.98.L14            -C-G--GT-----TCA--CAGT--C-----A---GAT-----A-----TTGCAG-G-G--A---GTAAA------AAT--TA---C-T---TGG-TG--G-GA......AAAC--GTAT-A-GGGT----  3962
LST.CD.88.447            -------AA-C-G--G----CT--------T--TA-T-----A--GG---T-AG-G-----C--AGTAA-A-G--AATTATT--C-TC--GTGGAGA----GAAAT......TCT-T-----G-G-----  2876
LST.CD.88.485            -------AA-C-G--G----CT-----------TA-T-----A--GG---T-AG-G-----C--A-TAAAA-G--GGTCATT--C-TC--GTGGAGG----GAAAT......TCT-T-----G-G-----  2876
LST.CD.88.524            -------AA------G--G-CT--C-----T---AGT-----AC-TG---T-AGGG----AC--A-CAA-A----GGTTATA--T-TC---TGGAGG----GAAAT......GCC-T---A-G-G-----  2873
LST.KE.-.lho7            T------AAGA-G--G----CC------------GA---G--AC-T--T-TCAGGG-----C-AA-CAAAA-G--GAT-ATT--T-TC--GTGGAGA----GAAAT......--C-TCT-A-G-G-T--G  3960
SYK.KE.-.SYK173          ---G---GA-CAGGCA--GTCT-----------TAA----CCT---GA----C-C--CACT--A-TA---TCCCAG------GG---CATG---A-G-------A-AAAGGCCC-CCC--A-T-------  3805
SYK.KE.KE51_AY523867     C----T-GA-CA-ACA--GTCT------C-G---GA----CCA---GA----C-C--TAGC---TTA--A---CAA--G---GGG--CAT----A-A---G-----CAAGAA---CC-----TC------  3463
COL.CM.-.CGU1            C------GA---T--G--------C---ATA---CATG-G---A-TTGG-T---C--TTTC-G-TTA---A----A------GGC---GC-G-C-GC---A-A......CA-GG--CG--A-G--G----  3374
```

```
H1B.FR.83.HXB2             AAATATGCAAGAATGAGGGGTGCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGCATAGTAATATGGGGAAAGACT...CCTAAATTTAAACTGCCCATACAAA  3732
Pol                        _K__Y__A__R__M__R__G__A__H__T__N__D__V__K__Q__L__T__E__A__V__Q__K__I__T__T__E__S__I__V__I__W__G__K__T__.__P__K__F__K__L__P__I__Q__    RT

H1D.CD.84.84ZR085          --G--------------------------------------G------------------------G---T-----------G--------------...----------G---A----------  3254
H1F1.BE.93.VI850           --G-------A-G-----TCG-------------------------------A-----A-----G---G-TTT--------------------G----GATCT-----G---------A-------TG-  3072
H1G.SE.93.SE6165           --G-------A--G-G--TC--------------------G--------A-T--------------G-------G--------------------...--C------------A-------G--  3129
H1H.CF.90.056              ----------A------AAC----------------A---------------A----------G---T-T-------------------------A-T-...----------G---A--T-------  3079
H1J.SE.93.SE7887           ----------A--G---A-----------------------------G---A-TG-------------G--TTG---GC----------------A---...----------G---A----------  3046
H1K.CM.96.MP535            --G--------------ATC--------------------G----------A--------------G--------G--------------------A---...----------G---A----------  2928
H101_AE.TH.90.CM240        --------C----GAG--TC---T------------------G-----------A-TG---------G--G-----------------------------C...----------G---A----------  3306
H102_AG.NG.-.IBNG          ----------A--A----TC--------------------G----------------T----------G--G-T-TG----------------------C...----------G---A----------  3257
H103_AB.RU.97.KAL153_2     ------------C---------------------C--------G----------------------------G----T--------------------...--------------A----------  2958
H104_cpx.CY.94.CY032       --G-----------CC--ATC-------------------T-G------------A-------------G---G---TG---T----------------...-----G-----G-T-A----------  3098
H1N.CM.95.YBF30            --G-------A------ATC------T----------A-------G---GTT--A-TG--AAGG---G-GG-A---------T--------T------------...----------G-T-A--AG------  3326
H1O.BE.87.ANT70            ------A-T--GCAA-A--CCT-------A------A---G-------G----A-T-A-C--G--GG-GT-TCA----TCT---A-T--C--------ATTG...------------G-----AG-CACT-  3787
H1O.CM.91.MVP5180          --------T--GCAA-A--CCT-------A------A---G------GG----A-T---C--G--GG-GT-TCA----GCT-----T--------G--ATTA...---------C-GG-----AG-TACT-  3762
CPZ.CM.04.EK505            ----------A-------TC-A-T--T---------CA---G---------T--A-TG--A-----GG-GG-ACT------T-----T--C---------A---...---------C-G-T-A--AG-G----  3304
CPZ.CM.04.LB7              --------C---CAAG--TC---T--T--C---------------G---G------T-A-------G-------G---T-------------C-----------A-TA...--A--G----G-T-------T----  3287
CPZ.CM.04.MB66             --------T--G-----ATC---A--T------------------G-----T----T---A------G--G---G--CT-------------C--C--------AGTA...----G---C-GG--A--T-------  3278
CPZ.MT145                  -------GC-A-CAA--A-C---A--T----------A-T-GG--GC-TG-T--AA----A----------G-T-----------T---------------A--A...--A-------G---C--TG-T----  3274
CPZ.CD.-.ANT               --------T--GCCT-CA--AA-T------------G--T-GG------G-T-GA-T-------------GGATT------T-----T--C------G--GTA...--A------C--T-A------ACT-  3171
CPZ.CM.-.CAM3              --------T--GCAA--ATCA--A--T---------CA---G-------G-----T---A---------A---------G-T--T-----T---------C---------A--A...--A-----C-G-T-A--TG-T----  3143
CPZ.CM.98.CAM5             --------T--GCA---ATCA--A--T---------CA---G-------G-----TG---------------------G----T-----T---------C---------A--A...--A-----C-G-T-A--TG-T----  3434
CPZ.GAB_X52154             --------C---CAA---TCA--A-----A------CA-C-GG---C-GG-G--A-----A--------TG-T--T---------------T---------A--A...-----------GGT-A--AG----G-  3793
CPZ.GAB2                   --------C---CAA--A-CA--T-CT---------CA---G----C--G-T--A-TG----C-----G-TT--C-T-----T---A-CC-T-------CTGT-...------------GT-A--AG-G----  3125
CPZ.TZ.-.TAN1              ----------A-TCACCA---A-------C------GA---G-------G---GACTGA-A--G------GG--AT--G------A----T-----T-TTGTG...----------TT-T-A--TG--TCC-  3376
CPZ.US.85.CPZUS            --------C---CAA--ATCA--A--T----------A-C-G---GC--GTT--A-TG-------------G-T--T--G-----T--C--T----------GTA...--A----------T---TAG-G--G-  3792
GSN.99CM166_AF468659       ----TC----AG-CC--AAA---A----TA--CTCCTATC-G-----TG----T--TT--ACC---G--GGT-GG--G--TCAG--T-CC----------GTACCAGAC-------GGA-C--AG--GTC-  3460
GSN.99CM71_AF468658        ----T-----A--CC--AAC---A--T-TA--CTC-TATC-G-----GG----T--TA--ACC---G--GGT-GG-----TCAG--T-CT---------AGTACCTGAC-----C-GGA-C--AG--GTC-  3442
H2A.ALI_AF082339           --G-------A-----AAAA-A----T--C--C-GG--C-G-TTG---G--C---T---T----------GGA-A---GCAC-G--C--T------CGA-TA...--A-G----C-C--A--AG--G---  4107
H2A.BEN_M30502             --G-------AG--A-AAAA-A----T--C----GG--C-G-TT-C--G--C---T---T-G-------GGA-A---GGCAC----C--T------CG--TA...--A------C-C-----AG-GG-G-  4114
H2A.ST_M31113              ----------AGG-A-AAAA-A--------C--C-GA--C-G--TCC--G--C-A-T---T----------GGA-A---GCAC----C--T------CGA-TA...--A------C-C--A--AG--G---  3559
H2B.D205_X61240            ----------AGG-T-AAAACA-G-----C--C-GG---G--T-C-GG--C-T-T---T--G-------GG--A---GC-C----C--C------G---TA...--AGTG--CC-T-----AG--G---  4087
H2B.EHO_U27200             ----------AGG-T-AAAACA-T-----C----GA----G--T---GG-TC-T-T---C---------GGA-AG---GCAT-G--C--C------G---TA...--A-TG--CC-T--A--AG--G---  4085
H2G.ABT96_AF208027         ----------AG--A-AAAA-A----T--A----GA----G-TT---GG-TC-T-TG--C-----R---GGA-AG---GCAC------------------ATTA...--CTTC---C-T-----AG--G---  3441
H2U.FR.96.12034            ----T---C-AGG-A-A-AA-A-T--T--A----GA--TCGC-TGC--G--C-T-TG--A--G--G---GGA-A----GC--------------------A-T-...--A-TG--CC-T-----AG--G---  3596

MAC.239_M33262             ----T-----AG--A-A-AA-A-A--T--C----GA--G-G--T----G--C-T-T-A-A--G------GGA-AG---GCA-----G--C------C--GTC...--A-----CC-CT-A--AG-TG-G-  4031
Pol                        _K__F__A__K__I__K__N__T__H__T__N__G__V__R__L__L__A__H__V__I__Q__K__I__G__K__E__A__I__V__I__W__G__Q__V__.__P__K__F__H__L__P__V__E_    RT

SMM.SL.92.SL92B            --G-T-----A---A-AAAA-A-A--T--A----GAA---GGTT-C-GG-TA-T-T---T-----G---GGA-A------TC----G---------------A...--ATTT---C-C--T--AG--G---  3462
SMM.US.x.PGM53             ----T-----AG--T-AAAA-A-A--T--A----GA--C-G-TT----G--C-T-T---CG-G------GGG-A----GCA--------T------C--GTG...--A------C-TT----AG--G-G-  3948
MNE.US.-.MNE027            ----T-----AG--A-AAAA-A-A--T--C----GA--T-G--T----G--C-T-T-A-A--G------GGA-AG---GCA-----G--C------C--GTC...--A-----CC-CT-A--AG-CG-G-  3499
RCM.NG.x.NG411             --G-------A-CAA-A-ACA--A--T--A-----AA-T-G-ATGC-GG---GCTT---A--G------TG--A----GCT--------T---------CT-...--A------G-G-----A---G---  3226
RCM.GAB1_AF382829          --------T-AGCA--AAACA--A-----C-----AA-T-G-ATGC-GG---G-TT---A--------TG-A-A----GC------T--T-------G-CTG...--A-C---C-G------A---G-G-  3211
DRL.FAO_AY159321           --------T-AGG-A-A-AA-AGT-----C------A-G-G-ATG---G---GACT---A-----G---G-A-A---GGCAC-G-----C-------G-TTG...--C-TT--CT-CT-A--AG--G---  3299
MND-1.F17                  -------GG-A-CA--AAACA--A--T-G------CC-C-TGA----GG---GT--TAC---G--GG--GGA-G------T--------C-----TTTTGTA...--A---A-GC-GA-A----CTAC--  3419
MND-2.5440_AY159322        --G-----C--G-CA-A-AA---------A-----AT-G-GGGT-C-TG---G-TTG--A-----GG--G-T-AG--GT-TC-------C---------CTA...--G-----CT-T--C--GT--G-G-  3293
MND-2.14CG_AF328295        --G-----T--G-CA-AACA---------A-----GT---GGGTGC-TG---GATT---A----------G---A---GT-------G--T------G-ATTG...--A-----CT-T--C---C-GG-G-  3660
MND-2.MND2_AF367411        --------C----CA-A-AA---T--T--A-----GC-T-G-GTGC-TG---GACT---A--G--G--TG-A-A----GCTT----G--C--------ACTC...-----G---T-T-----A---G---  3733
DEB.CM.CM40                ----TCT-T---GAT-AA--A--T---TA-------T-CC-T------G-TA----TA--T-T---------GGA----------------TT-T-------GA-TA...--AC-G--C-GG--C--AG-GGT--  3538
DEB.CM.CM5                 ----TCT---AGGAC-A---AA-T--TTA------CT-TC-T---C--G-TA-----A--T-T-------GG------G--T-----CTAT--------AGT-...--C-------GGT----AG--GT--  3532
MON.CM.L1                  ----T---C-A---T--AA--A-----T-C---ACCTATC-G--G---G----T--GC-AGCC---T-GGGA-A---GGC------C-C-------CGACT-...--C-TT---CGC--A--AG-GGTG-  3435
MON.NG1                    ----TC----AG--T-AAAA-------T-C--CACATACC-G-----GG----T--T--AACC---C--GGG-AG--GGC-T-G--C--T---------CTA...--A-T-----G-T-A--AG--A-C-  2028
TAL.CM.8023                --G-T-----AGCA---AACA--A----G---CCCCTATC-----C--GTG-GA--CA-----------GGA-A---GGC------CT-T------C--GTG...--C-T-----GGG-A--AG-GGTC-  3196
TAL.CM.266                 --G-T---C-A-CA---AACA--A----G---CCCCTATC----G----T--GA--TA----G------GG--AG--GTC------CT-T-----GC--GTA...--AGT---C-GGA-C--AG-GGTC-  3698
MUS.CM.1085                ----T---C-A-G-A---TCA-TA---------AGCTATC-G-----GG----C---A--GCC------GGA-G-------C----C--T-----G---C-CCCTGAC--GG-C-G-A-C--AG-GATC-  3422
SAB.SN.-.SAB1C             ----------A---C-AAACA--A-----C-----AT---G-ATG---G---GCCT---A--G--G---GGA-AG-----------C--C------C---TA...----TTA-GG-------AG--G---  4130
GRV.ET.x.GRI_677           -G------C-AGCAA--A-AAA-T--T----------C---GGACTC--G--C-CCTT--C--------CTGT-AG---GCAC-TACC--T-------GACT-...--ACG-G-AC----C--AG--G-C-  3905
TAN.UG.x.TAN1              ------A---AG-GC-AAAA-A-T--T--A-----GT-T-G-GT---GG---GAATG--C----GGG-GTG--AG---GCG--CACCT-T--------ATTA...--C-TCA-GG--G-A--AG--G---  3931
VER.DE.-.AGM3              -----CA---AGCA--AAAACA-T--T--C-----GT-CCGTGT---GG---GATT---A------C-TTGT-A----TCTT----T---------G--TTG...--AGTCC--G----C--A---G-G-  3454
VER.KE.-.9063              --G-------A-CAA-A-AACA----T--A-----AC-TCGCAC---GG-C-GCCT---A--------CTGT-A---GGCAC-G--C--T------G-ACTA...--A-TT---G----C--AG--G---  3959
VER.KE.-.AGM155            --G-TC--C-A-CA--AA-C-A-T-----C-----GT-GCGTGT-C--G---GA-T---A--G------GGG-A---GGC-C--------T------C-ATTA...--C-CT---G----C--AG-GG-G-  3951
VER.KE.-.TYO1              -----CA-C-AGCAA-A-AACA----T--A-----AC-TCGCAC----G-T-GTTT------G--G--TTG--A----GCTC----T--T-----G-TATTA...--AGTTC-AG----C--G---G---  3449
SUN.GA.98.L14              -G----CA--AG-GAG---CA--A--TGAG---CC-TGTC----G---G---CT--CC-A-----G---GGA-G-----------T---------TTTGTG...--A--GA-AC-GG-A----------  4089
LST.CD.88.447              -GG---CA----CAA-AATCA--T-----A---CCCC--C--A---GG----------A---------GGA-A---G--T-----C--T-----TC-TGTG...--C--GA--C-GG-A--GG--AC--  3003
LST.CD.88.485              -GG---CA---GCAA-AA-CA--T-----A---CCCC--C--A-G--GG----------A---------GGA-A---G--T-----C--T-----TC-TGTG...-----GA-CC-GG-A--AG--AC--  3003
LST.CD.88.524              -G----CA----CAA-A-A----T----TG---CCCC-TC-G------G---A--T--A--G------GGA-A-------------C--T-----TTTTGTA...--A--GA-AC--G-A--TG-CACT-  3000
LST.KE.-.lho7              -GG---CA----CA--A--CA--A-----A---CCCC--C-GA-----GT---A--TA-T--G--G---GGA-A-------------C--C-----CTTTGTG...--A---A--C--G-C--AG-GACT-  4087
SYK.KE.-.SYK173            --GACA----A--CCTTT-CA--A--T-G------CTATC--TCTC-TG-TC-ATTGT-AA-T------GGG-TCC--TCAC--TGGTAT---------GTA...--A-C---CC-T--C--AG--A---  3932
SYK.KE.KE51_AY523867       ---ACA-G--A-GCATAT-CGA----TCAC------TACC-GGC---GG-TC--CTGA--A-T------GGA-T-C--GC-C--TGG--T--------A-TC...---G----CC-C--C--AG--A---  3590
COL.CM.-.CGU1              ---C--AAC-C-GGC-AAAA-CAG---TT----AG-A-GC-GG-G---G-TTCA-TTA-A-----G---GGT-GG---GC------C-CC---------GTG...--C---A-G---G-C--AGC-A-G-  3501
```

```
H1B.FR.83.HXB2          AGGAAACATGGGAAACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACCCCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGA  3862
Pol                     K__E__T__W__E__T__W__W__T__E__Y__W__Q__A__T__W__I__P__E__W__E__F__V__N__T__P__P__L__V__K__L__W__Y__Q__L__E__K__E__P__I__V__G__A__E     RT

H1D.CD.84.84ZR085       ------------------------T---C--------------------------------A-----C-----------T-----A-----------------------G---------A----------  3384
H1F1.BE.93.VI850        -A--G--------T--------------T--------------------------A-----------C--------C--TC--------C----------------A-C---G-------C---------  3202
H1G.SE.93.SE6165        -A-------------T------------C--------G-----------------------------C-----------TC----A--G--------T-G------A-C----------CC---------  3259
H1H.CF.90.056           -A-----------G--C--------------------------A-----------A--------------C-------ATC----A-----------T--------A-C---G-------C---------  3209
H1J.SE.93.SE7887        GA-----------G--------------C--C-----G-----------------------------C--------C--TC----A--G-----------A-----A--G---------A-G--------  3176
H1K.CM.96.MP535         -A-----------G--------------A--------------------C-----A-----------C-----T------C----A---C----------------A-C---------------G-----  3058
H101_AE.TH.90.CM240     GA----------------------TG-----------G--T--------------A-----------------------TC----A--------------A-----A-----C-----------------  3436
H102_AG.NG.-.IBNG       GA--G-------------------TG-----------G--T--------------A-----------C-----------TC----A--------------------A-----C-----------------  3387
H103_AB.RU.97.KAL153_2  -A-----------------------------------------------------------A-----C-----------------A--------------------------------------------  3088
H104_cpx.CY.94.CY032    -------------C--------------A--------G-----------C-----A-----A-----C-----------TC----A--------------------A-C---C-------C---------  3228
H1N.CM.95.YBF30         -----GTG-----GG----------C--TC----------A--T-----------------A-----C--C--T------C-T--A-----------T--------A-C---G--A--CAGT--G-----  3456
H1O.BE.87.ANT70         GA-----T---------------G-G--C-----------------------A--A-----A-----C-GC--A--C--A--GA-C------------AG-C----A-GT-----T--TA-G--G-----  3917
H1O.CM.91.MVP5180       GA-----T--------T------G----A--------G-----------------A-----A-----C-GC--A--C--A--GA-C--------------------A-C------T--T-----G-----  3892
CPZ.CM.04.EK505         -A---GT------G--C--------------------G--A--T-----------T-----------C--C--------AC----A-----------T----------C------A--TAGC--G-----  3434
CPZ.CM.04.LB7           -A---G--------G-------------T--------------T-----C-----A-----A-A-------------T--CAT--C--G--T--------A------C-------T--CC----G-----  3417
CPZ.CM.04.MB66          -------------GG-G-----------C-----------A--A-----C-----A-------A------C--------------A---C-T--------A------C----C-----CC----------  3408
CPZ.MT145               -A----TG------G-C------G----T--C-----G--T--------------------A--CA-A--C-----A---C----A--G-------------------C---G--T----A---------  3404
CPZ.CD.-.ANT            G------T-----TG-C------T----C--------G-----------A--A--A-----------------A--C--A---A-T-G-C-C-----TA-TC--TT-GC---C--T--TCC--AG--T--  3301
CPZ.CM.-.CAM3           ----GGTC--------C------T-------------------------C--A--C-----N--C--G------------C-T--A--------------A--G--A-C----------CC---------  3273
CPZ.CM.98.CAM5          -----GTC--------C------T----A--------------------A--A--C-----A-----A------------C-T--A--------------A--G--ACC----------CC---G--G--  3564
CPZ.GAB_X52154          -A----GT------G--------G----A--C-----G--A--------C-----A-----A---A-C--------A--------C------------AGT-------C------T--CC-ACCA-----  3923
CPZ.GAB2                G-----GT------G-C-----------T--------------------C-----A-----A--C--C-GA-----A---C-T--A--G--------T--A--G--A--------T--CA----------  3255
CPZ.TZ.-.TAN1           -A--G------AGCCAG--------T--T--C-------TT------G-A-----------A---A----C-----A--AC--A-C-GGC--------A-TC-GTT-TCT--C-----CCC--A------  3506
CPZ.US.85.CPZUS         -----GTC---------------T----A--------G--------------A--T-----A-----C------------C-C--A---C-T-----TA---------C---GG------AG--G-----  3922
GSN.99CM166_AF468659    -----CA------TG---------TGA-T--------G--A--A-----A--CACCATA--AGCAA-AC-C--A--A-A-C--C---G-CA-------ACCC---TCTCG------C---A--AT----T  3590
GSN.99CM71_AF468658     ----GCA------T--G-------TGA-C--------G--A--A--------CACCATA--AGCA--AC-C--A--A-AT--GT---G-CAG------ACCC---TCTCG------C---A-AAT----T  3572
H2A.ALI_AF082339        GA-----C------CAG------GAT--C--C-------TG--A-----C--A--C-----C-----ATC------A--AC-G--C-GGC--GCA-TTA-CC---TA-----T--T---C----C-----  4237
H2A.BEN_M30502          GA--G--C-----GCAG------GATA-C--C-------TG--A-----C--A--------C-----ATC------A--AC-G--C-GG---ACA-TTA-CC---TAGG---T--T---CC---C-----  4244
H2A.ST_M31113           GA--T--C------CAG------GATA-C--C-------TG--A-----C--A--C-----C--CA-ATC------G--AC-G--C-G----GTA-TTA-CC-G-T------T-------C---C-----  3689
H2B.D205_X61240         GA--G--------CCAG-----------T--C-------TA--------C--A--------C-----CTCG-----A--A---A-A-G-C--GCC---A-CC---TC-----C---C---A---GAG---  4217
H2B.EHO_U27200          GA--G--------TCAG-----------T--C-------TA--------C--A--A-----T-----CTCA-----A--A---A-A-GG---GCC--TA-CC-G-TC-----C---C---A-----T---  4215
H2G.ABT96_AF208027      G---C--C------CA------------C----------TA--T-----C--A--T-----C-----GTC---A--G-ATC----A-GG---GTC---A-CC---TC--------TC---AGCAGA----  3571
H2U.FR.96.12034         GA--GGTC------CA---------T--C--------G-T-------G-C-TA--C-----T-----ATCC--A--A---C-G----G-C-GGCA--TA-TC---TA--------AT--C--...-A---  3723

MAC.239_M33262          ----TGT-------CAG-----------C--------G-TA--------A--G--A-----T---A-CTCA--A--A--GC--A-G------GTC-T-A-TC---T---G--C--T----AG----A---  4161
Pol                     K__D__V__W__E__Q__W__W__T__D__Y__W__Q__V__T__W__I__P__E__W__D__F__I__S__T__P__P__L__V__R__L__V__F__N__L__V__K__D__P__I__E__G__E__E     RT

SMM.SL.92.SL92B         G---GGTT-----TCA------------T-----------T--T--------A--------C--CA-CTCC--A--A--G---A-C-GG---GT--TTA-T----TA-----T--A----A-AAG-AG--  3592
SMM.US.x.PGM53          G-----TT------CA------------T----------TA--------A--A--A-----C-----GTCA--A-----------C-G----GTC-T-A-CC---TA--------T--CAG---------  4078
MNE.US.-.MNE027         ----TGT-------CAG-----------C--------G-TA--------A--A--A-----T---A-CTCA--G--G--AC--A-G------GTC-T-A-TC---T---G--C--T----AG----A---  3629
RCM.NG.x.NG411          GA---GT-------CAG------T-G--T--C-----G-TTT-------------A-----A-------G---T--A--AC--A-A-GGC--------A-TC---TA--G--T--T---CC-----A---  3356
RCM.GAB1_AF382829       GA--GGT------CTA----C--T-CC-A--C-----G-TT--------A--A--T-----A-------GC--A--A--TC-TA-T-G-C-CG-----A-CC---TT--G--T------CC-----AG--  3341
DRL.FAO_AY159321        G----GTT------CA-------CAG-----C-----G-TA--A-----C-----T-----A---A-ATCA--A--C-ATC-CA-C--G--------TA-C--GCT-GT---C--A---CC---G-A---  3429
MND-1.F17               G---G-T-------GAT------CAT-----------GTGT--A-----A--A--AGTA--A---A-C-GC--A---ATG----AA-GGGA------TAGC--GTCCCC------TC---AG--G-T---  3549
MND-2.5440_AY159322     G---GGT------CCA-------CAT--C----------TA--A-----C--A--------C--CA--TCA--A--A--A---A-A-GG---------A-CC-GCTA-----C--A---CC---G-A---  3423
MND-2.14CG_AF328295     G----GT------TCAG------CAT--T--------G--A--A-----C--A--A-----A--CA-CTCA--A--A--AC--A-A-GG--G-----TA-CC-GTTA--------A---CC---G-AG--  3790
MND-2.MND2_AF367411     GA---GTC-----CCAG------C----A--------------A--------C--A-----A-----ATCC--T--A-A-C-TA-CGGG--G-----TA-CC-GTT--G------TG-GCC-----A---  3863
DEB.CM.CM40             -A---GA------C-AT------CATA-TC-C---------G-------C--A--C------GCAA-AC-------A-ATC-G--C--GC-C-----TG-C--G-T-TC----------CC--AT-----  3668
DEB.CM.CM5              -----GA------T-AC------CATA-CC-C-----G-----T-----C-----C------GCG--GC-------A-AAC----C-GGC-------TG-C--T-TATC---G--A---CC--AT--T--  3662
MON.CM.L1               -A--GCA------CG--------G----TA--------AT--AT---G-A--A--CATA---GCG--GT--T----A-A-C-GC-C-G-CAG-----TACTC-G-TAC----C-------A---G--TCC  3565
MON.NG1                 -A---CAG-----C---------G-C--CA---------T---T-----A--A---ATA---CCA--AC-CT-T--A-A-C-GC-T-G-CA-------ACC----T-C--R-T-----T-A-CAC--TCC  2158
TAL.CM.8023             ----GGAG-----GCA-------T--------------TG-T-A---C-G--CACCATT-TCCCCA-AC----------------C-GGC-------TA-CC---TATC---G-------A---------  3326
TAL.CM.266              ----GGA-------CAG-----------AC-C------TG-T-A-----A--CACCATC-T-CCCA--C-C--A--A--A-----T-GGC-------TA-CC---T-C-G--G-------A---------  3828
MUS.CM.1085             -A---CAG-----CCAG-------T--CAC-C-----G-T-T-G-----A---ACCATA--AGCA--GC-----A-C-A-C-GT-A---CAG----T-ACA--G-TACC------AT--AA--A---TCC  3552
SAB.SN.-.SAB1C          GA--GCTC------CA-------T-T--C--------G-T---------C--G--A-----AA-G--C-G---T----AA--GA-T-G---G------A-A----TA-----T-----CCC-----A--C  4260
GRV.ET.x.GRI_677        --A----------T-TG------CAG--C----------TAT-------A--A--A-------------GC--A--A-T-C----A---C-G-----TTCC----TA--------A--CAA-----A---  4035
TAN.UG.x.TAN1           GA---GT-------CA--------GT--C----------T---A---C-A--A---ATA--A--CA-CTCC--A--G------A-C-----------TA-C--G-TA-----T------CC-AA----A-  4061
VER.DE.-.AGM3           G----GT-------CA-------G-T--T--C-----G-TA-GT--------A--C-----A-----C-G------A---C----A-----------TACCC-GACA------------CC-AAG-A---  3584
VER.KE.-.9063           G----GT-------CA-------G----C--C-------TG-G------------A-----A-----G-G-GTA--A--TC----------G------ACTC-GACC--G-----T---CC-----AG--  4089
VER.KE.-.AGM155         G---C---------CA-------G----C----------T--GT-----A--C--A-----C-----C-G-GTT--G--------A-CT--G-----TACAC-GACT--G--------CCCG----AG--  4081
VER.KE.-.TYO1           GA--GGT------GCA-------G-G--T--C-----G-TA-G---------C--A-----T-----C-GC-----A--T--GC-C---C--------ACA---ACA------------CCCAAG-AG--  3579
SUN.GA.98.L14           GA--G-T----AGTCAG------G----C--C-----GTGT--A-----C-------TA--A-----C-GC--T---AAG--G-A-C--GAG-----TACTC---CC-C------AG-GCC----...--  4216
LST.CD.88.447           GA---GTG-----TCA--------GT--TC-------G-T---------A-----A-T---A---A-CTCA-----A-AG----AAC-GGA------TAGC-----AGC------T----A---G-T---  3133
LST.CD.88.485           GA---GTG-----TCA--------GT--TC-------G-T---------A-----A-T-------A-CTCA-----A-AG----AAC-GGA------TAGC-----AGC-----------A---G-T---  3133
LST.CD.88.524           GA-------------AT-------T-A-C-T--------TA--T-----A--A--C-TA--A--CA-CTCC--A----AA----AAC-GGA------TAGC--G--AGC-----------A---TA----  3130
LST.KE.-.lho7           GA---GTC-----GCAC-------GC--CC-C-----G-TT--A--------A--C-TA--A--CA--TC------G-AA----AAC--GAG------ATT-GG---GC----------A----G-T---  4217
SYK.KE.-.SYK173         GA--GGA--------A------------C--------------A---G-A--A--AGTCA-A--CA-A-G------G--A--G--T-GG-GGAC---TA-C--G-TACC----------CC--A-----T  4062
SYK.KE.KE51_AY523867    G----GA--------A------------TC-C--------A--A---T-G--A--AGTAA-A-G---GC-C-----CATA-----T-G--GGAT----A-T--G-TTTC---G--AG--...-A------  3717
COL.CM.-.CGU1           GA--GGAT------CAG-------GT--C---------AGTG-T-----C--A---ATA--------CTCAT--T--TATG----A------GTA-GGA-C--G-TTTT---G--TC-GAAG-AG---CC  3631
```

```
                     Pol p51 RT \/ Pol p15 RNAse
H1B.FR.83.HXB2            AACCTTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACTAATAGAGGAAGACAAAAAGTTGTCACCCTAACTGACACAACAAATCAGAAGACTGAGTTACAAGCA  3992
Pol                       __T__F__Y__V__D__G__A__A__N__R__E__T__K__L__G__K__A__G__Y__V__T__N__R__G__R__Q__K__V__V__T__L__T__D__T__T__N__Q__K__T__E__L__Q__A_  RNAse

H1D.CD.84.84ZR085         ---T-----------------------T-AA---------------------------------G-C-----------G---------C--T-T------------------------------------  3514
H1F1.BE.93.VI850          C---------------------T----T--A-----C---AA----------------------G---A-----A-------------T-------G--G--C------------G--------------  3332
H1G.SE.93.SE6165          ---T-A---------------------T--------A-----------G---------------G-C-A-----AG------A--A-T-----------A--------C--A---G----A-----G---  3389
H1H.CF.90.056             ---T-A----A----------------T-----A--------------------------C---G---------AG------------T-------G--A--------------------A---------  3339
H1J.SE.93.SE7887          ---T------------------T-----------------AC------G-----G---------G-C-A---------G-----A--T--------------------------------AC----C--C  3306
H1K.CM.96.MP535           ---T-----C---------------C-T-----A------AAG----G----------------G-C--------------G------T--A-------A--------------AG----A---------  3188
H101_AE.TH.90.CM240       G--T----------------------GT-----------GC-------------G-----C---G-C--------------G--A--TT----------G-----------A--------A-----T--G  3566
H102_AG.NG.-.IBNG         ---T--------------A--------T--A--------GA-------------G-----C---G-C--------------G-----TT----------G-----------A--------A-----T---  3517
H103_AB.RU.97.KAL153_2    ---T--------------A--------T-------------C----------------------G-C--------------G------T-----------------------------------------  3218
H104_cpx.CY.94.CY032      ---T-----------------------T--A--A--A---CAG---------------------G-------C---------------T-----T----A--------------------A---------  3358
H1N.CM.95.YBF30           ---T--------------A--------T-----A--A-----G-----------T-T---G--AG----G--------G-----G---T-TA-TG-A-----C--C-----A---G-------------T  3586
H1O.BE.87.ANT70           -----AT-----------A--------T--A-----A---C-------G--------------AG-ACA---G-A---G--GA-AA-A-AAT--GA---G--C--C-----A--AG----A---ATG--G  4047
H1O.CM.91.MVP5180         ------T-----------A--------T---A-T--A---C-------G--G-----------AG-ACA-----A---G--CA-AA-A-AGT--GAA--G-----C-----A---G----A---ATG--T  4022
CPZ.CM.04.EK505           -----AT-----------A--------T-----A-----------G----------T---G--AG----G--T--G-----G--GAC-T-TA-CT-A--A--T--C--C--AC-AG----A-----G--T  3564
CPZ.CM.04.LB7             ---A--T-----------------------A--A-----G--------------------A--AG----G--G-AG--G---A-AA--C--T----A--A-----C---------G-------------C  3547
CPZ.CM.04.MB66            ---A--T-----------------------A--A--C--------G--------------A--AG-C-A------------GA-CA--T-TT----A--A--C--T--------AG-A--A--------T  3538
CPZ.MT145                 ---A-A------------------C-----A--A--A---AG------------C-----G--A--C--G--G--------GA--CAT-GT--TGAA--T--C--T-----A--AG-A--A---------  3534
CPZ.CD.-.ANT              ------T-----T--------------T--AA-CT--C----G--G--G-----C-----G--AG-C---------AGT-GG--AAAACA----CAAA-G--C--C-----AC-AG-A--A-----G--G  3431
CPZ.CM.-.CAM3             ---G--------G-----------------A-----A--GAA------------C--C--C---G-------C-----G---A-A--A-A-T--GAAA-T--C--T--------AG-A--AC-GAC---T  3403
CPZ.CM.98.CAM5            ---A--------------------------A--A--A--GAA------------T--CA----AG----G--C-----G---A-AA-A-ATT--GAAA-T--T--T--------AG-A--AC--AC---T  3694
CPZ.GAB_X52154            T--T-AT-----------A-----A--T-----A--A---ACT--G--------------A--AG-C-A---G-A-------A-CA-T-G-T--GAAA-T--C--T------C-AG-A--A---A-G--T  4053
CPZ.GAB2                  -----AT-----T-----------C--T--------A-G-AG--------------------T-----AT----A---G--G--CA-TT--T-GGAA--G--C--T--C--A---G-A--A---G-G---  3385
CPZ.TZ.-.TAN1             ------T-----------------A-----A--C-G---AAG---G------------A--A--C-----C---T-C-GGTCAAAGGA-T--GAGA----C--T-----AC-AG-A--A---TGG---  3636
CPZ.US.85.CPZUS           ---A--T-----G-----A-----CC-AC-A--A--C--GAA------G---------------G----G--T---------A--A-A--TT--GAAA-T--T--T-----A---G-A---C-CAC---G  4052
GSN.99CM166_AF468659      T--T-AT-----------A-----A--T---ACCT-C---C----------T--T-----G--C----C-----A---T-GG-CCA-AGAAT--GAA--A--T---------C-AG-A--AC-C--T---  3720
GSN.99CM71_AF468658       C----AT--------C--------A--T---ACCT----------------T--C-----G--C----C-----A-T-T-G--CCA-AGAA---GAA-GA--T--------AC-AG-A--AC-C--G---  3702
H2A.ALI_AF082339          G--------CAC---C---C-CTG---T---C-AT-A---GA----------------A-A--AG-------G---G-C--G--GAAGGTA---GAAC-A--T--C------C-AG-A--A---G----C  4367
H2A.BEN_M30502            G--------CAC------AT--TGC--T--AC--T-A---GAG-----------------A--AG---------A-G-C-----AAAAGTAT--GAAC-A--T--C------C--G-A--A---G---TC  4374
H2A.ST_M31113             ---------CAC------AT-CTGC--T-A-C-AT-A-G-GA---------------CA-A--AG-----------G-C--G--GAGGCTAT--GAGC-A--C--C------C-AG-A--A---G----C  3819
H2B.D205_X61240           -----A---CAC-------T-CTGC--T--AACCT-A--GGA------------------C---G-C--G----A-G-T--G---AAAGTGT--GAAC-G--------C--AC-AG-A--AC-TG-----  4347
H2B.EHO_U27200            ---T-A---CAC------AT-CTG-----AA-CCT-A---GA---G--------------C--AG-C--G----AGG-T------AAAC-AT--GAAC-A------------C-AG-A---C-TG-----  4345
H2G.ABT96_AF208027        G----AT---AC---C---T-CTGC--T-AAACCT-C--GGAR-----------C--C--C--AG-C-------A-G-C--G--AAARGT-T--GAGC-G--------C---C--G-R--AC-TG-G---  3701
H2U.FR.96.12034           G--A-AT---AC------CT-TTG--------CTT-A---CAG--G--------C-----G--AG---------A-G-C-G---AAAAC-TT--GAAC-G--C--C--C--AC-AG-A--A---G----C  3853

MAC.239_M33262            -----AT---AC------AT--TG---T-AAC--T-A---GA---G------------A-C--AG----G--C-A-G-C-----AAAAGTGT--GAAC-G--T--T-----AC-AG-A--A--GG-----  4291
Pol                       __T__Y__Y__T__D__G__S__C__N__K__Q__S__K__E__G__K__A__G__Y__I__T__D__R__G__K__D__K__V__K__V__L__E__Q__T__T__N__Q__Q__A__E__L__E__A_  RNAse

SMM.SL.92.SL92B           -GTA-AT--CA-T-----CT-CTGC--T--AA-C-G----GA---G--------------C--AG-C--G----A-G----G--CC-GC-AT--GAAC-GG-T--C------C-AG-A---C-G--G--C  3722
SMM.US.x.PGM53            G--A--T-----------AT-CTG---T---C-AT-A-G-GA------G-----C-----A--AG-------C---G-C--G-CAAAAC-TT-GGAAC-G--T--C-----AC-AG-A-----GG----C  4208
MNE.US.-.MNE027           -----AT------------T--TG---T-AAC--T-A---GA---G------------A-C--AG----G--C-A-G-C-----GAGAGT-T--GAAC-G--T--T-----AC-AG-A--A--GG-----  3759
RCM.NG.x.NG411            -GT--AT-----------------A--T---A-T-G---G--G-----G-----------A--AG----G----A-G----G--AAAAGAAT--GAAA-T--T--C-----A--AG-A------GG----  3486
RCM.GAB1_AF382829         -GT--A------G-----------------AA-T-G----A-------G--G-----------A----------A-G----G--AAAAGAAT--GAA--A--T--T-----A--AG-A--A---G-----  3471
DRL.FAO_AY159321          -GTT-AT-----G-----A-----------A-TT-G----GA---------T--C--------CTCC--G----A-G-------AA-AG--A--GAA--A--C--C---------G----A-----G--C  3559
MND-1.F17                 ---A-AT-----T-----A--------------C-G----A-G---------------A----AG-------TTTT----GG--A-AAGAATATCTAA-T--C--C------C----A--A-----T---  3679
MND-2.5440_AY159322       TGTA-AT-----------------A-----AA-CT-C--GGA---C--G-----C---TA---CTC------T-A-AGC--G--A--AG--T-GGAA--A--G--C-----A---G-A--A--GAC---T  3553
MND-2.14CG_AF328295       TGT--AT-----G-----------A--T--AA-CT----GGA---G-----------CTA---AGC---G----A-AGT-----AA-AG--T--GAAA-------T--------AG-A--A---G-G---  3920
MND-2.MND2_AF367411       TGTG-AT-----G-----A-----------AA-CT-C---GA---G--------------A--AGC----AAT-A-TCC-G---AA-AG-----GAGA-T--C-----C--A--AG-A--A--GG-G---  3993
DEB.CM.CM40               C----A----------------------------T-C--------G-----T--C-----C--AG-GT-G----A-----GT--AAAATG-T--GAGA-T--C--T--------AG-A--A---G----C  3798
DEB.CM.CM5                C----AT-----------A--------T-----AT-A---A-G-----------C--CA-AT-AG-AT-G----A---GGG---AAAA-GT--GGAAA-T--T--C-----A---G-A--A---G----C  3792
MON.CM.L1                 ---A-A---C--G--------C--C--T---ACAT-A--GGA---G--------T-----A--C--CTTT--GGC-AG---G-CC-----AT--GAAC-G--C--C--C--A---G-A--C--G--G--C  3695
MON.NG1                   C----A---------------G--------A-CAT-A---GA---G--G-----------C--C-G--C---CCA-AG---G--AA-TC-AT--GAGC-G--T--C--C------G----AC--G-----  2288
TAL.CM.8023               ---A-A------G-----T--T--CC----A--AT-A--GGA---C--------C-----G--C-CGTC-----A-G--C-T--A--AGGA--T-GCA-T--T--T---------G-A--A---G-----  3456
TAL.CM.266                C--G-AT--C-----C--T-----CC--C-A---T-A---GA---C--G--------C--G---GCA-C-----A-G--C-T--CA---AGT--GAAA-A--T--C--C--A---G-A--A--GG-----  3958
MUS.CM.1085               C--A-AT-----G-----A-----A--T----TAT-C-----G--------------C--C--A--C-C---G-A-G----G--AA-T-G----GAAAGT--T--C--C-----AG-A--A--GG-G---  3682
SAB.SN.-.SAB1C            -GT--AT--------------------T--AA-TT-C---GA---------------CT-A--AG-------GGAT-----G--G--AG-AT--GAGA-T--T--C--C-----AG-A------G-G--C  4390
GRV.ET.x.GRI_677          TGTT-AT-----G-----------ATC--AA-T---C--------T--G---------C-GT-AG-G-------A-AGT-G-A--AGGGAAT--GAAA----C--T--C--AC-AG-A--A---AC----  4165
TAN.UG.x.TAN1             GGTG-AT---TGC---------TGC--T--AA-TT-C-----------G-----C-TCC-A-----C-C----GA-TC-C-G--CA-AGAAT--GAG--T--C--G--C--AC-AG-A--A---------  4191
VER.DE.-.AGM3             TGT--A------G-----A--TTG---T--AA-TT-A-GGGA----------------A-C--AC-ATAT--G-A-----GG--G-AA-AAT--GAAA-T--------C---C-AG-A--A---ATG--C  3714
VER.KE.-.9063             TGTA-AT--------------CTG---T--AA-CT-A---GA---------------CA----AC-ATAT----A---G-G---A-AG-AAT--GAGC-G--T-----C---C-AG-A--A--GGT---C  4219
VER.KE.-.AGM155           TGT--A------------A--CTG---T--AC--T-G---GAG-----------C--CA-A--CC-ACA---C-A-----G---ACAACAG---GAAA-------------AC-AG----AC-GAC---C  4211
VER.KE.-.TYO1             CGTT-A-------...--A---TGC-----AA-TT-A---GA---------------CA-CT-AC-ATAC----A---G-G---A-AA--AT--GAAA----T--C------C-AG-AA-A---AC---T  3706
SUN.GA.98.L14             C----A------G-----A------G-A-A-CTAGAA--GAG---G------------A-A--CC-GG--------TCC-G----AAG-AAT-GGAAA----C--C------C-AG-A--AC--G-G--T  4346
LST.CD.88.447             T----AT-----------A--T--AG-A-AA-TAGGG---AC------G-----T-----C--AC-ATC-----AGG-G------AAAGAAT----A-----C--C-----AC-AG-A--A---G----T  3263
LST.CD.88.485             T----AT-----------A--T--AG-A-AA-TAGGG---AC------------T-----C--AC-ATC-----AGG-G------AAAGAAT----A--T--C--C-----AC-AG-A--A---G-G--T  3263
LST.CD.88.524             T----A------G-----A--T--AG-A--A-TAGG----AC------------G---A-A---C-ATC-----A-G-G---A-AAAAGAAT--CAA--T--T--T------C--G-A--A---G-G---  3260
LST.KE.-.lho7             C----A------------A-----AG-A-A--TAGGA---AC----------------A-A--AC-ATC---G-A-G-G-----AAAGGAGT----------C---------C-AG-A--AC--G-G---  4347
SYK.KE.-.SYK173           ---------C--------A--------T--A--C-GC--GAC------T-----------AG-A-G-GAT--G-C---G-G---ACAGTAT--GGAGC-A--------C--AC--G-A--A---G-G-GC  4192
SYK.KE.KE51_AY523867      ---A--T-----C--------C-----T---A-T-GC---GA----C------T---CA-A--AG----G--T-A---G-C-T-AAAG-GGT-GGAAA-------C-----A---G--------G-----  3847
COL.CM.-.CGU1             ---T-AT-GGAC------A-G-TGCTC----A-ATTAGG-GCG-----G-------GGA-A-A-TCA-A----GA-G--G----A----AT--TCA---AGGGT--------C-AG-A---C-GAC--G-  3761
```

```
H1B.FR.83.HXB2           ATTTATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCACAACCAGATCAAAGTGAATCAGAGTTAGTCAATCAAATAATAGAGCAGT  4122
Pol                      _I__Y__L__A__L__Q__D__S__G__L__E__V__N__I__V__T__D__S__Q__Y__A__L__G__I__I__Q__A__Q__P__D__Q__S__E__S__E__L__V__N__Q__I__I__E__Q__ RNAse

H1D.CD.84.84ZR085        ---A---------------------------------------------T-----------C-------------------------A-G----------A-------G-----------A----  3644
H1F1.BE.93.VI850         ------T---------------A----C------G-------------------------------------G--G--------A-G----------AA--------------------A-  3462
H1G.SE.93.SE6165         ---C-G-----------A--C--AA---C----------------G--------G------C-------------------------AGG-----G----A------------------A---C  3519
H1H.CF.90.056            -----------------A----A--G-------G--------G----T-----G------C---------------------C---A-G----------------T--------------G-A-  3469
H1J.SE.93.SE7887         ------T-------A-G------A--GC---------------------T--------------G--G--T--A--------------A-G----------A------------------G---  3436
H1K.CM.96.MP535          --C-G-T-------A---------A----C--------------------------------G---------------------G------A-----------T-----T---------------A-  3318
H101_AE.TH.90.CM240      --CC--T----C-----------A----C--------T------------------------------------G-----------CAGG-----------AG------C------------G--C  3696
H102_AG.NG.-.IBNG        ---C--T----C--A--------A----C--------T-----------------G------------------G-----------CAGG--------------------------------A--C  3647
H103_AB.RU.97.KAL153_2   ---C--------------------------------------------------------------------------------A-G------------------G----------------  3348
H104_cpx.CY.94.CY032     -----CT---------------A----C-------------------------G------A-----------------------AG-----------T----T------------------  3488
H1N.CM.95.YBF30          --CCT-A-G--C--A--A--G--A---CGG--T----------C--T-----T--G-----TA-G-----A-----TT----G-----A------------A--G--G-GC----------AG--C  3716
H1O.BE.87.ANT70          --A-TA-----C--A--------CAA-GA-AC------T----------T---------------G--CG----CTCCT-C-----TACA--G---------CCCTA----TC-G-----------G-AC  4177
H1O.CM.91.MVP5180        G-A-TAA----C-----------CAAGGAGC-------------------------------T---G--C-----ATCCT-C------ACA--G-----C--CCCTA----TC-G--G---------G-AC  4152
CPZ.CM.04.EK505          G-CCTCA-G--C--A--A---G-A---CA---------T-----C--T-----T--G----TCC-------A-----CT-----------A-----------------A-----A-----G--------AG-AC  3694
CPZ.CM.04.LB7            ---C-AT----C--A--A--C--A---CCT--G-----T-----T--T--T------------------------C-------C-----G-----G-----T-----A--G--G--------C-----AA-TC  3677
CPZ.CM.04.MB66           --AC-AT-G-----A--A-----A-A-G---------T-----C--T--T-------------C-G-----A-------GT-----T---AC-----------AA----A-----G---------G-AC  3668
CPZ.MT145                G---TAA-G--C--A-----C--A---CAG-----C--TG-G--C--T-----C-------TGC-G------T-A---T----C-----C--C-----------TA----A-----G---------AG---  3664
CPZ.CD.-.ANT             ---CT-A-G---C-AG-------AAC-GGCCC---C--T-----C-----T-------------------G--G--T-G----GTACC-------------------CCCC----GG-AG-------CC--A-A-  3561
CPZ.CM.-.CAM3            G---------A--A--A---G-A-A-CATAT---T--T-----C--T--T--C-------TG-----G--A-----CT-C--G----------------G-----A--G--T----T---------AG---  3533
CPZ.CM.98.CAM5           G-------G--A---------G-A-A-CATAT---C--T-----T--T--T--T-------TGC-------A-----CT-T--G-----------------------A-----AGGGTCT---------AG---  3824
CPZ.GAB_X52154           T-GCT---G--C-----------A-ATCA-C-G--T---------G--T-------------TG-----G--T-----GT----G---------C-----------A-----------G---------AG---  4183
CPZ.GAB2                 C-CAGAA----CC-A-----C--A--GCA---------TG-G-----------C--------TG--G---------AA-TT-C---------C---------G-----AA----G-----------------AG--C  3515
CPZ.TZ.-.TAN1            G-AG-------C--AA-A--C--A---GC-C-G-----T-----C-----T--C-------TTA-G---G-TT-A--G-G-TT-----------------C--CCCCA----AG-G-----T--TC-AA---  3766
CPZ.US.85.CPZUS          G-A--CT----A--AA-A-----A-AGAATAC---T--TG----C--T-G---C-------TC-------------C--CT-T--G-----C---------------A--G--------G---------AG-A-  4182
GSN.99CM166_AF468659     G--CT-T-G--A--AA-A--AGG-CC-CC-A--A-G--TC-G--C-----T--C--G----T-A-G--G---T-G---AGT--G-----GGT-TCCACC--T-C-C----GG-A-----C---C-AG-AC  3850
GSN.99CM71_AF468658      G--CT-T-G--A--AA-A--AGGACC-CC-AG-A-G---T-G--C-----T--C-----C-T-A-G--G---T-A---AGC--------AGT-TCCACC--TCCC-----GG-A-----C---C-AG-AC  3832
H2A.ALI_AF082339         T-CGCG--G--AG-AACA--C--A--TCC-A---CC--T--TA--GT---T-----G-----T-A-G-----AG-AGC--GC--G---ACAG--TCA---AATAGAA----A---------C-----AG-AA  4497
H2A.BEN_M30502           T--CGGA-G--AC--GCA--C--A--CCC-A-G--T--T--CA--GT---T-----G-----T-A-G--G--AG-AGC--GC--G---ACAG-GTCA---AATAGAA----G--C--G--C-----AG-AA  4504
H2A.ST_M31113            T--GCGA-G--AG-AACA--C--A--TCC-A-G-CC-----TA--GT---------------T-A-G-----AG-AGC--GC------ACAG-GTCA--GAGTA-AA----A---------C-----AG-AA  3949
H2B.D205_X61240          T--GCAT----A--AACA--C--A-A-CC-C----T-----CA--GT---T----------TCA-G-----A--AGCT-----G---ACAG---CA------CCAA----AGCAA-------T--AG-AA  4477
H2B.EHO_U27200           T--GCA-----AC-A-----C--A---CC-C-G--C--T--CA--GT---T----------TCA-G-----AG-AGCT---------ACAG---CA------CC-A----A-GAG-------T--AG-AA  4475
H2G.ABT96_AF208027       T--GCAA-G--C--------C--A--TCC-A----C--T--CA--GTR--T--C--G----TCA-G--R-----AGC--GR------ACGG---CA-----CCC------A---A-------T--AG-AA  3831
H2U.FR.96.12034          T--GC-A-G--C--AG-A-----A---CC-A---------T--G-----T---------AT-A-G-----A--AGC--G---G---ACAG---CA--G--CCCAC----A-G-A----C-----AG-AC  3983

MAC.239_M33262           T--CTCA-G--A---ACA--C--A--GCC-A-G-C---T--TA--GT---T----------TTA-G-----A--AAC--G-TGC--TACAG--TCA--GAGCAG-C----T---------------AG-AA  4421
Pol                      _F__L__M__A__L__T__D__S__G__P__K__A__N__I__I__V__D__S__Q__Y__V__M__G__I__I__T__G__C__P__T__E__S__E__S__R__L__V__N__Q__I__I__E__E_  RNAse

SMM.SL.92.SL92B          C-G-TAT-G--C---A-A-----TCC-AGTA-G--C--TG-------G--------G----TC--GAAC--A--AAC--G---G---TCAG--TCG--T-----TA----GGCA-----T--T--A--AC  3852
SMM.US.x.PGM53           T-C--------C--AGCA-----A---CC-A-G-CC--T--CA--GT---T--C-------TTA-G--C--A--AGC--GT-----CAC-G--TCA---AGTAGA-----A--C--G---------G-AA  4338
MNE.US.-.MNE027          T--CTCA-G--A---ACA--C--A--GCC-A-G-C---T--TA--GT---T----------TTA-G-----A--AAC--G-TGC--TACAG--TCA--GAGCAG-C----T--C-----------AG-AA  3889
RCM.NG.x.NG411           G-A-TA--G--C--AA-A--C--A---AA-AGG-----------------T--T-G-C---TG--T--C--AT-AGC--G-AGT-----CAC------GAGCCCCC----AC-A--G--T-----A--AA  3616
RCM.GAB1_AF382829        G-A-TA--G--A--AA-A-----A--GCCTA-------------C-----T-----G----TT-AT-----TC-AG--------------ACT-----C----G------G-CAG----T---A-T---A  3601
DRL.FAO_AY159321         ---CTGT----C---A----C--T--TCC-A----C--T-----G--G--T-----G--C--CA-G--T--A--ATCCT-CTCC-----AATC-----CAATCCTA----G---------T-----A--AA  3689
MND-1.F17                G-AA-A------C-AG-A---AGT---AGTT-T--T--------------T----------T-G----T--AC-AGC-AGCAG---TAC-G---CA--TCACCCCA----A--GG----------ATTAA  3809
MND-2.5440_AY159322      --AA-GA-G--------A-----A---CCCAGG--G--T-----C--T--T--C--G------C----C--AC--TC----GCC---------GTCA--CAACCCCA----T-GGG-----------A-TCA  3683
MND-2.14CG_AF328295      --ACTGT----C---A-A-----A--TCCTAGG-C---T-----C------------------C----C--AC--TC---CACT-------G-TCA--CAACCCTA----T-GAG-------CA-C--AA  4050
MND-2.MND2_AF367411      --AA-AA-G--A--A--A--C--A---CC-A-------T-----G-----T-----G-----TA-G--T---C-ATCG---GC---C--C---TCA--CAACCCCA----A-GAG----T-----A-TCA  4123
DEB.CM.CM40              --A-TGT-G--C---G-A--AGGCCC--C-A--A-G--T---------------C--G------C-G--G--A---TT--A---C---TC-G-G-CA---CATA--A----AG-AA-GG-----C--GCC-  3928
DEB.CM.CM5               --CCTA--G--C---G----GGGACCC-CCA-GA-G------------------C--G-----T--------AT-GTT--A---C--CACGG---CA---CAT---A----AG-AA-G--C---C-AGCC-  3922
MON.CM.L1                G-G-TA-----CC-CAGA---GGACC-CCTC-GA-G------A-T-----T--T--G----T-C-G--C--TC-AGCCT-CTGC--G--GATCTCA---AGTCCCA----GG-GGCC--C---C--G-AC  3825
MON.NG1                  G-GCT-T----A-----AA--GG-CCCGCTAGTA-G--T-----------------G----TCA-G------C-AGC-T--A-G-----GGTCTCG--G--CCCAA----AG-GA----C--CC-----C  2418
TAL.CM.8023              G-GCTAT-G--A--AA-AC-C--T--TCC-A-G-----------C-------T-------TC-AT--C---T-AGT--G-A-T--CACAG-G-CA--CAACA--A----GG-AG-C--C---C-----C  3586
TAL.CM.266               G-ACTAT-G--AC-CA-A-----A--GCCCA-G-----T-----C-------T--G----TC-AT--C--AT-GGC--G-AGC--CACAG---CA--TAACA--A--A-AG-AG-C--T--CC---TAC  4088
MUS.CM.1085              G-C-TAT-G--C--AA-A--AGG-CCTCCTAGGA-G--TT----C-----T--T--G----TC-----G--AG-AGCTT-T--G---C-AG--TCCACT---CCA-----AG-G---------C--G--C  3812
SAB.SN.-.SAB1C           ---CTGT----C--AAGA--C--T---AGCA-------T---A-------------------TA-G--T--T--AGC--GGG----TACAG--TCA--TAATA-CA----AC-A-----T------G-AC  4520
GRV.ET.x.GRI_677         G--A-GA-G--A---G----CAGT---GA-A-T-----T-----C-----T--T-------T-A-GAAC---T-GAC----TGT---C-GG-----A-C---CCC-----GG-A--G------C-AGCCC  4295
TAN.UG.x.TAN1            G-CCT------A--AA-AC--AGTAAGAGCA----T--T--------G--T-----G----TCA-G------C-GAC---------TACA---------C--C-T-C----A-----------T--AG-AA  4321
VER.DE.-.AGM3            --AA-AA-G--AC-AG-A---AGT--GCCTA-T-----------------T------------A-G-----AT-AACT--C-----CACA--G-----C---CCC---A-AG-A-----T----CA-TAA  3844
VER.KE.-.9063            --AA-AA-G--C---G-----AGT---GA-A-------------------T-----------CA-G-----AT-GAC------G--CACA--------C--CCCA----AG-A--G--C----CATTAA  4349
VER.KE.-.AGM155          --AA-AA-G--C---G-----AGC--CCCTA----C--T-----------T----------GA-G--C--AT-GAC------G--CACA--G-----C--CCCAC---AG-A----------CA---A  4341
VER.KE.-.TYO1            --AA-AA-G------G-A--CAGT--GCCTA-T--G-----------------T---------A-G-----TT-GAC---------CACA--------T---CCA-----AG-G-----T----CCTTAA  3836
SUN.GA.98.L14            --AA-AA-G--C---G-----AGTA-GAGCTC---T--T----------------G------C---A--C-GC-CAGTAAGAG---CACAG---CA--T------C----A--GG-G--CG------TCA  4476
LST.CD.88.447            G-C-TAT-G--A--AAGA--CAGTAAGAGCAGG--C----------------------TGA--AA-G-TT-ATCTCA--GG--TACAG---CA---CATCCTG----A--A--G--T------G-A-  3393
LST.CD.88.485            G-C-TAT-G--A--AA-A--CAGTAA-AGCAGG--C----------------------TGA-GAA-G--T-ATCTCA--GG--TACAG---CA---CACCCTG----G--G--G--T------G-A-  3393
LST.CD.88.524            G-A-TGT-------A--A---AGCAA--C-AC---T-----------T-----C-------T-A-GAAGG--T-AAGTCAGAG---TACAG---C---GCATCCTA----G--GA-C--C---A--G--C  3390
LST.KE.-.lho7            G--CTAA-G--A--A--A---AGCAATAGTA-------T-----------T----------T-A-GAA---AT-GTC-CA-AG----ACAG---CA---CACCCCA----A--AG-C--C-----A--A-  4477
SYK.KE.-.SYK173          T-GCTCA-G--A--------C--AAA-GACA-------T------GT---T--------CT-T-AT--C--TC--ATGA--TGC--CAC-A-C-CA---CACCCAA----GG-G--G------C-AG--G  4322
SYK.KE.KE51_AY523867     G-A-TAA-G--A--A--A--C--A---CC------T--------------T--T--G--------------AT-AATGAACTGT--TACATCCTCG---CACCCAG-----G-A--G--T--GC-AA--G  3977
COL.CM.-.CGU1            G---TG------C-AA-AC--GG-CC-AA-AG-A-G--TT-G---------AGT-TT-----------G--A---ACT-GG--G---T--G-TCAGCTTC--TCTAGTAGAGGGG-G--T--TC--GCAG  3891
```

```
                                                                              Pol p66 RT & Pol p15 RNAse H end \/ Pol p31 Integrase start
H1B.FR.83.HXB2          TAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGA...ATCAGG......AAAGTACTATTTTTAGATGGAAT  4243
Pol                     L__I__K__K__E__K__V__Y__L__A__W__V__P__A__H__K__G__I__G__G__N__E__Q__V__D__K__L__V__S__A__G__.__I__R__.__.__K__V__L__F__L__D__G__I Integrase

H1D.CD.84.84ZR085       ----------------------C-----------------T---------------------------------------------AG----...------......--------------G--------  3765
H1F1.BE.93.VI850        -----C---------G------C---T----------------------G------------------------------------------...G-----......---A----G-----------G--  3583
H1G.SE.93.SE6165        ----------------------C---T----------------------G-----------------------C------------AG----...------......------------C-------C--  3640
H1H.CF.90.056           -------G--------------C---T-------------------------------------------------------T---T-----...G----A......-----G------C-------G--  3590
H1J.SE.93.SE7887        -------G--------------C---T-G--------------------------------------G------------------T-----...------......-----G--G---C-------G--  3557
H1K.CM.96.MP535         ---------------G-A----C--AT-----------------------------------------------------------------...------......-----G------C----------  3439
H101_AE.TH.90.CM240     -------------G--A-----C---T-------------------G--G------------------------------------T-A---...------......--G--G--------------G--  3817
H102_AG.NG.-.IBNG       -------------C--A-----C---T----------------------G------------------------------------AA----...------......--G-----------------C--  3768
H103_AB.RU.97.KAL153_2  -G--------------------C-----------------------------------------------T---------------------...------......G----------------------  3469
H104_cpx.CY.94.CY032    -----CGG-----C--------C---T----------------------G-----------------------------------CAA----...-----A......--G--G-----------------  3609
H1N.CM.95.YBF30         -C-------------GA--T-----CT-T--------T-----T-----T--------------G--G-----C--------T--CT-A---...--T--A......---A--T----CC-------T--  3837
H1O.BE.87.ANT70         ---C----------C----G-----TA-------T--T--C--T-----C--A------------A--A-------------A--CAAG-AT...--T--A......-G---C--G--CC----A-----  4298
H1O.CM.91.MVP5180       ---C----------CGA--G-----TA-------T--T--T--------C--A------------A--A-------------A--CAAA-AC...--T--A......-G---C--G--CC-G--A-----  4273
CPZ.CM.04.EK505         -T-------------GAA-A-----TT-C--------T-----T-----T-----------------GA----C-----------CA-A---...------......-----CT----CC-------T--  3815
CPZ.CM.04.LB7           ----------------A--T--CT-AA----------T-----T--G-----A--------------------C------------T-A---...--T---......-----C--T---C-G--------  3798
CPZ.CM.04.MB66          -G--C-----A--G--A--G-----CT----------T------------------------------A----------------CT-A---...--T---......-----C--C--CC-G-----G--  3789
CPZ.MT145               ----------------AA-T-----AT------------------G---T--A--------------G---------C----T---T-A--G...-----A......-----T--T--C-----------  3785
CPZ.CD.-.ANT            ---------GA---C--A-T--C--CT-C--------------T-----C--A--------------G-----C------------CAA--G...-----A......C----C--C---C-G--A--C--  3682
CPZ.CM.-.CAM3           -------G--------AAGT--C--TT-------G--------T------G-------------------------------T---T----G..A-----A......-----T--C-----------T--  3655
CPZ.CM.98.CAM5          ----------------AAGT--C--CT-------G------------------------------------------C----T---T-A--C...--T--A......-----C--C--C--------T--  3945
CPZ.GAB_X52154          ----T-----A-----AA----C--CT-C--------------------T--A-----------G-----G-----------------A---...------......-----G-----CC----------  4304
CPZ.GAB2                -------G--A--G---------T--A-------------------------A--------------G--------G---------AG---C...-----A......--G--G--------------T--  3636
CPZ.TZ.-.TAN1           ---C-C-----AC-GCAA-T-----A--------------C--T-----T--A--G--T------G-------C-----G--T---AAAAAT...--T--A......---A--T----CC-G--------  3887
CPZ.US.85.CPZUS         -------G--A-----AAGT---A-CT----------------T------------------------A--------C--------T-A--G...-----A......-----T--T--CC-------C--  4303
GSN.99CM166_AF468659    -GC-----------GCCA----CT-AAGT-----T--G--------G-----A--------CAC-G-G--G--C-----------CAAG---...---C-C......C----G--T--CC-G--AAAC--  3971
GSN.99CM71_AF468658     -GT--GGG--A--GGCC-----CT--AGT-----T--------------G--A--------CAC-G-G-----C-----------CAGA--G...---C-C......C-------C--CA-G--GAAC--  3953
H2A.ALI_AF082339        -G------------GCCA-----G-T--G-----C-----C--------C--A--G------C-GG-------CC-T-----A---CAG--C...-----A......C-----T-G---C----AAA---  4618
H2A.BEN_M30502          -G-----G------GCA------G-T--------C-----C--T-----C--A--------CC-GG-------CC-T-----A---CAA--C...-----A......C-----T----CC-G--AAAG--  4625
H2A.ST_M31113           -G------------GCAA-----G-T--------C-----C--T-----C--A---------C-GG-G------C-C-----A---CAG--C...-----A......C-----T----CC----GAA---  4070
H2B.D205_X61240         -G--C-----A--GGCA--A---G-A-G------------T-----G---C-G--T--T---C-GG-------CC-CC----A---CAA---...-----A......C-G--CT-G--CC----AAA---  4598
H2B.EHO_U27200          -G--C-----------AA-A---G-A-G------------T-----G---C-G--T--T---C-GG-------CC-CC----G--CCAA---...--T--A......C--A-C------C----AAA---  4596
H2G.ABT96_AF208027      -R--T-----A---RCTC-A---G-A-G------------T--T--------A-----T---C--G-G-----CCRTC----A--CCAG---...--T--A......C----------C--GR-AAA---  3952
H2U.FR.96.12034         -T----GG--A---GCA--G---G-T-GC--------T--T--T--G--T-----T--C---C--G--------C-T--G--G---CAG---...--T---......C-------G---C----AAA---  4104

MAC.239_M33262          -G--T------TC-G-AA-T---G-A-----------------------T--A--------CC--G--A----CC-CC----T---CAA--G...--T--A......C----T--C--C--G--AAAG--  4542
Pol                     M__I__K__K__S__E__I__Y__V__A__W__V__P__A__H__K__G__I__G__G__N__Q__E__I__D__H__L__V__S__Q__G__.__I__R__.__.__Q__V__L__F__L__E__K__I Integrase

SMM.SL.92.SL92B         --G-TC-G--A---GCA--A--CA-A-G------T-----T--T-----C--A-----G--CA-TG-G-----CCGT-----T---CAA---...--A--A......C-G--C--G--CC----AA-T--  3973
SMM.US.x.PGM53          -G--T---------GCAA-T---G-------------T--T--T--------A---------C--G--------C-CC-G--T---CAA---...--T--A......C----C-----C-----AAA---  4459
MNE.US.-.MNE027         -G--T------AC-G-AA-T---G-A--------G--------------T--A--------CC--G--A----CC-CC----T---CAA--G...--T--A......C----C--C--C--G--AAAG--  4010
RCM.NG.x.NG411          -G---GGG------G-A--G--C--CT-C--------G-----T-----C-----G--------G--G--------------G---CAG---...--T--A......C-G--------CC----C--G--  3737
RCM.GAB1_AF382829       -G---GGG--A---GCA--G--C--TT-C-----G--C-----T--G--C--C-----------GG----------------T---AAA---...-----A......C-G-----G--CC-------G--  3722
DRL.FAO_AY159321        -G--C-GT--A---GCA--T--CT-AAAT--------------T--------A-----T-----GG-G-AG--C---C----T---AGA---...--A--A......C----GT----CC-----AAC--  3810
MND-1.F17               -G-A-GG---A-----AA-T---T-AAGT---C----------------G--A-----G-----G---A-------GC----A---T-A---...-----A......-----CT----CC--C-AAAT--  3930
MND-2.5440_AY159322     -G---GGC------G--A-T-----CT-------C-----C-----G--T--A-----T--------GA-------------A---AAG---...--A--A......C-G--------------G-----  3804
MND-2.14CG_AF328295     -G---GCC--A---GCT--A--CT-A--------------T--T-----C--A-----C---------A----C--------A--CAAA---...-----A......C----------CC-G--A-----  4171
MND-2.MND2_AF367411     -G--CC-C------GCA--A--C--A--T--------------------C--A-----G-----G------------C----G--CAGA---...G-A---......C-------G--------A--C--  4244
DEB.CM.CM40             -GCA-GGG--A---C-AA-T-----A--------------C-----G-----A-----------C--G------------------AAA---...--A---......-GGA-C--C--C--G--AA-GG-  4049
DEB.CM.CM5              --CA-G--------C----A-----TAAT-----------C--------T--A-----G-----T--G-----C-----G--A---AGA---...--A---......-G-A-C-----C--G--AA--G-  4043
MON.CM.L1               -CT----------CC----A-TCT--T-------------T-----------A------------G--------CGCC----GG-CCAG---...--T---......C----C--G---A-G--AAATT-  3946
MON.NG1                 -CC--GG---AA-GC----A---T--T-------C--G--C--------C--A--G---------G----R--C--------A--YCAA--G...-----A......C-G--C--C--CA-G--GAAC--  2539
TAL.CM.8023             --T--GGC------GCT--G--C--TA-----C-------C--T--------A--------------------C--GC-G--G---CAA---...------......-G-A-T--C---G----ACA---  3707
TAL.CM.266              --C---GC-----GGCA-----CT----T---C-G--G--C--------C--A--------------G--G-----G-----A--CCAA---...------......--GA-C--T---G----ACAG--  4209
MUS.CM.1085             --C-GTC---A---GCCA-A--CT-AT-------G-----C--T--G--T--A-----------GG----G--C--------A-A-AG----...------......C----C-----CA-G--ACAG--  3933
SAB.SN.-.SAB1C          -------G-----GGCA--G---A-A-----------T--C--T-----G--A--T--C-----GG--A-T------C----T---CAA---...--A--A......C----------CC-G--CA----  4641
GRV.ET.x.GRI_677        ----G------AGGC-------CT-ACA------------T--T--G--G--A-----C---AC-G--A-------------A--CAAA---...--A--A......C-GA-C--C--C------A----  4416
TAN.UG.x.TAN1           -G--T--T--A--TGCTA-----T-AA-C--------------T-----G--A--G------C--G--A-------GC----A--CCAG--G...--T---......C----G-----C-----CA----  4442
VER.DE.-.AGM3           -GG--C----AC-TC--A-A--CT--CA-------------G-------G--A-----C------G-GA--------C----A---CAA--G...--G---......---A-TT----------AAA---  3965
VER.KE.-.9063           -GG----G--AAGGGCAA-A--C--ACA------G--T-----------T--A--G---------G--A----C--------A---CAA---...--A--A......-G---T------A---GAA----  4470
VER.KE.-.AGM155         -GG--C-G--A---GCCA--------CA---------T--T--T-----T--A--G--C------G--A----C--------A--CAAG---...G-T--A......-G-A--T-G--CA-T-GCA-G--  4462
VER.KE.-.TYO1           -G---C-----C--C-AA-A---T--CAG--------------T--------A-----------GG-GA-------------G---AAA--C...--T--A......-G---TT----C-----AAA---  3957
SUN.GA.98.L14           ---GGC-------CC----A--C--A-GT--------G--------------A---------C--G--A-----C--C----T--CCAG---...-----AAAGAGAC-G---A-G---A----AAAG--  4603
LST.CD.88.447           GT-A---G--A--CC-A--G-----A-G------------T--------T--A--G------C--G-------CC-CC--------AAA---...--T--ACAAAAAC-------G---C---GAAAG--  3520
LST.CD.88.485           GT-AG--G--A--CC-A--A-----A-G------------T--------T--A--G------C--G-------CC-CC--------AAA---...--T--ACAAATAC-------G---C----AAA---  3520
LST.CD.88.524           -GCAG--G--A--TGCA--T--C--A-G------------T--T-----T--A--G-----CC-GG-GA-----C-------T-T-AAA---...--T---ACAAAAC-G--CA-G---C-G--GAAG--  3517
LST.KE.-.lho7           GC-AGC-G--A--TC-A--T-----A-G------G--T--T--T--------A-----T---C--G-G-----CC-CC----A---AAA--T...--A--ACAGAAAC-G--CA-G---C----AAAG--  4604
SYK.KE.-.SYK173         CC------------GCCA-A---G--A-------C-----C--T--------A------------GCT--T------C--------AAA---...--A--A......---A-C-----CC----AA----  4443
SYK.KE.KE51_AY523867    CT--GG-G--A------A----CA-AA-C-----T-----C--T--------A---------C--G--------C-------A---AAG---...--T---......---A----G---C-G---AAC--  4098
COL.CM.-.CGU1           GT--GGCT--A---GCCA-TC--G-A--C---TGT--------T-----G--A-------------TGA-----C--AAG--A......---...G-A--A......C-G---A-G-GGA-----AA---  4006
```

```
H1B.FR.83.HXB2              AGAT...AAGGCCCAAGATGAACATGAGAAA...TATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAA  4367
Pol                         __D__.__K__A__Q__D__E__H__E__K__.__Y__H__S__N__W__R__A__M__A__S__D__F__N__L__P__P__V__V__A__K__E__I__V__A__S__C__D__K__C__Q__L__K_ Integrase

H1D.CD.84.84ZR085           ----...-----T-----G------------...--------C-----------------------------------------G--------------------T------------------------  3889
H1F1.BE.93.VI850            ----...-----A-----A--------A---...-------AC-----------G-----------------T---------A----------------------T-----------G------------  3707
H1G.SE.93.SE6165            ----...--A--------A--G-------G-...-------AC--C--------------C-----------TT--------A-------------------G---------------------------  3764
H1H.CF.90.056               ----...--A--T-----A--------A-GG...-----T-AC------------G----------------T--A------A----------------------T------------------------  3714
H1J.SE.93.SE7887            ----...--A--T-----A--T-----A---...-----T--C-----------------------------T-----------------------G--------T-----------------------G  3681
H1K.CM.96.MP535             ----...-----T-----A--------A---...-----T-AC--C--------------------------T---------A-------C--------------T--------C---------------  3563
H101_AE.TH.90.CM240         ----...-----T-----A--------A-G-...--------C---------A-------------------TT--------A----------G---------A---A---------------A------  3941
H102_AG.NG.-.IBNG           ----...--A--------A--G-----A-G-...--------C-------AG--------------------T---------A-------------------G---------------------A-G---  3892
H103_AB.RU.97.KAL153_2      ----...-----A-----A------------...------G-------------------------------------------G--------------------------------------AT-----  3593
H104_cpx.CY.94.CY032        ----...-----T-----A------------...-------A---C--------------------------T------T-A--G-----------G--------T------A-----------------  3733
H1N.CM.95.YBF30             ---A...--A--------A--T-----C-G-...--------C-------A---------C------------T-A--C--CA----G--------------------------C-----C---------  3961
H1O.BE.87.ANT70             ---C...C----A-----A--T-----A---...-----T----------A----C-A--------A---GGA--A-----A--G--G--C--G-----CA-T--T------CC---------TA-----  4422
H1O.CM.91.MVP5180           ----...C----A-----A--T-----A---...-----T---------------T-A--------C---GGAT-A-----AA-------C--G-----CA-T--T--T---CC------C--TA-----  4397
CPZ.CM.04.EK505             ----...-----T-----A--------A-G-...-----T----------AG--------C-----------T--A--C--CA-------T-----------------T--------G-----AT-G---  3939
CPZ.CM.04.LB7               ----...CGA-----G--A--G-----A---...--C--T-A------------T-----C--------C--T--------CA-T-----------G--------T--T--------------------G  3922
CPZ.CM.04.MB66              ----...--------G--A--G-----A---...--C--T-A---------G--C-----C-----------T--------CA-------------G--------A-AT--------G-----AT-----  3913
CPZ.MT145                   ---A...--A--T-----A--------A-G-...--C---TC--------AG---------CAG--C-----T--A-----AA----G--T---------------CAA---------------------  3909
CPZ.CD.-.ANT                ----...--A--T-----G--C-----T---...-----T--C---------T--T-A--AGA---A-AC--T--T--C---A-T--G--T--------TA----ACAG--------------CG----G  3806
CPZ.CM.-.CAM3               ----..T---CA-A-G--A---------G--A..-----T-AC-----------T-----A--------C--TA-A----A-A----G--C------C--------C-T-----------C--A------  3781
CPZ.CM.98.CAM5              C---...-----A--G--A--------A--G...-----T-AC--C----A---T-----C--------C--T--A-----------G--C-----------------T-----C--G--C---T-----  4069
CPZ.GAB_X52154              ---C...-G---------A--------A-GG...-----T----------A---T-----------------TT-A-----CA-------------------G---CAT--------G--C---G-----  4428
CPZ.GAB2                    ----...-----A-----A--C-------G-...-----T--C--C--------C------GA---A--CC-A--C-----G--------C----------------CA-----------------T---  3760
CPZ.TZ.-.TAN1               TA--...G-A--A--G--A--C-----T---...----------------A---TT-A---GA---A-A---T-----C--A--T--G--T--------TA-T--TCAG---CCA-----C--TA-----  4011
CPZ.US.85.CPZUS             ----...--A--A--G--A--C-----C---...-----T----------C---T-----C--------------------A--G--C--T--G--G-----G----------------------C----  4427
GSN.99CM166_AF468659        ---G...CCT--AGT---A--C-----A--G...-----T----------AGTACC-AAGAGA-C-A-AC--AA-C--CG-AC--T-G--T-----------GAA--AA---TCC--G------G--C-T  4095
GSN.99CM71_AF468658         ---G...CCC--AGT---G--C-----A---...-----T----------AGTATT-AAGGGA-C-A-AC--AA-C--GG-AC--T-G--T------------AA--AG--CTCC---------A--C-T  4077
H2A.ALI_AF082339            ---G...CCA--T--G--A--------A--G...-----T--C---GT--A--A-C-AT-CCA-A-A---GGAT-A--CAA-C-G--G----G-C-G------AA--CA---CCC----------AG--G  4742
H2A.BEN_M30502              ---G...CCC--T-----G--------A---...-----T--C-T-ATA-A--A-C-AA-CCA-A-A---GGAA-A--C-T-C---------G-C-G------AA-TCA---CCC-----C--A-AG---  4749
H2A.ST_M31113               ---A...CCC--T--G--G--------A---...-----T--C---GTA-A--A-C-AT-CCA-A-A---GGA-----CAAAT----G----G-C--------AA--CA-CACCC----------AG---  4194
H2B.D205_X61240             ---A...CCA-----G--A--G-----A---...-----TG-C---GTA-A--A-C---TCCA-A-A--CGGAA-T----AAT----G------C-G------AATTC------------C--A-A----  4722
H2B.EHO_U27200              ---A...CCA--T-----A--------A---...-----T-A----GTA-A--A-C-A-TCCA-A-A---GGGA-T----AAT----G----G-C--------AATTC------------C--A-A----  4720
H2G.ABT96_AF208027          ---A...CCT--T-----A--------A--R...-T---T------GTA-A--AGT-AA--CA-A-G---GG-A-T--T-AAT-----------C-G------AATTCA-ACC-C-GT--C--A-AG---  4076
H2U.FR.96.12034             ---A...CCA--T--G--A------------...--C--T--C---GTT-A--AGC-ACAGCA-A-A---GGTA-A----AAT-------C---C--------AAT-CT---C--GTT--C----AG---  4228

MAC.239_M33262              ---G...CCA--A-----A--------T---...--C--T------GTA-A--A-T---TATTCA-A---GGAT-A--CAGAA----G--C-G-C-G-------A--C---------------T-AG---  4666
Pol                         __E__.__P__A__Q__E__E__H__D__K__.__Y__H__S__N__V__K__E__L__V__F__K__F__G__L__P__R__I__V__A__R__Q__I__V__D__T__C__D__K__C__H__Q__K Integrase

SMM.SL.92.SL92B             ---G...CCA--A-----A--C-----C---...-----T--C---GTA-AG-AGT-A---CAGA-A-A----A-C----AAT-------T---C-G------AATGCT---A-C-----C--A-AG--G  4097
SMM.US.x.PGM53              ---A...CCA--A-----A--G-----A--G...--C--T--C---GTA-A--A-T---TATTCA-A---GGTA-A--TAGGC-----------C-G-------A--CA---C-------C--C-AG---  4583
MNE.US.-.MNE027             ---G...CCA--A-----A--------T---...--C--T------GTA-A--A-T---TATTCA-A---GGAT-A--CAGAC----G--C---C-G-------A--CA--------------T-AG---  4134
RCM.NG.x.NG411              --GA...--A--A--G--A--------A---...-----T--C-----------------AGAA--A---C-AA-A--T-AAA-----------------------CAA--CCCC--G------G-----  3861
RCM.GAB1_AF382829           ---A...--A--A-----A--------A--G...-----T-A------------T-----AGAA------C-AA-T----AAA----G-----------------ACAG---CCA--------AG-----  3846
DRL.FAO_AY159321            G--A...--A--A--G--A--------TTT-...--C--T-A----------T-CT-A--ACAA--A---GGTT-A---GGGA----G--------------G--ACAG---CC---------AA--C-T  3934
MND-1.F17                   ---A...CCA--A--G--A------------...-----T--C---GAAGC-CA-T-AAGAGAGA-A--CC--T-A---G--C-------C---C-G--T--GCAA--T--CAG---G--CTGT-ATC-T  4054
MND-2.5440_AY159322         ----...--A-----G--A--------C---...-------A------------GT--T--CAG--C--------A--C--CA----G--------G-----T--ACAA---CCA---------AC---G  3928
MND-2.14CG_AF328295         ----...--A--T--G--G--------C---...--C--T-A----------T--T-AAG-CAG--A----G-A-A--C---A----G--------------T--ACAG--CCCA--G------G-----  4295
MND-2.MND2_AF367411         ----...--A--A-----A--G-----T---...-------A---------G---T-A--ACAA--C--CTGTA-A---AACA----G------------------CAG---CCA--G-----AAC----  4368
DEB.CM.CM40                 ---G...G-A--------G--C-----A---...--C--TG-G-------A-CA-T-AAGGGA---A-ACC-A--A--TA-AC-CA----T--GC----TA--A-ACAA--CCCC---------AC--G-  4173
DEB.CM.CM5                  ---A...G----T-----G--T---------...-----TG---------A-CA-T--AGAGA---A---C-GT-A--TA-CT--A----C--GC-G--TA-T--GCAG---CCA--G--C---A----G  4167
MON.CM.L1                   G--A...CCC--AAGG--A--T-----A--G...-----T----------A-TACT-AAGAGA-ACC-ACC--A-T--TA--C-GT-G--C-----------CAA-CAT---C-C--------AAACC-T  4070
MON.NG1                     T--A...CCR----G---G--T-----A---...--------C-------A-TATC-CAGGGA-AC--A-GG-A-T---A-AC-CC----C-----G-----CAA-CAT--CCCC--G-----CAC-C--  2663
TAL.CM.8023                 CCCA...G-A--A--G--A--G-----A-GG...--C--T-AC-------AG-ACC-AAAGGCCAGG-----A--A--CA-CA-------C----CC--CA---AGGCA--CCCC---------G-CC--  3831
TAL.CM.266                  CCCA...G-A--A-----A----------G-...--C--T-A---------G-ATC-TAAGGCAAGA-----GT-A--CA-CA-------C--G-CT--TA---AAGCA--CCCC--------AG--C--  4333
MUS.CM.1085                 ---A...CCT--AA-G--A--T-----A---...----------------AGTACC-CAGGGA-A-A-ACC-AA-C---G--T-GT----C------------AA-CAT---CCA-----C---AC-C-T  4057
SAB.SN.-.SAB1C              T--A...G-A--A-----A--G-----T---...--C--TGCC--------GAGT---CAACAA--A---GG-T-A--TG--A----------------------GGCA---CC------C---A-----  4765
GRV.ET.x.GRI_677            ---A...G-A--A--------C----CA--G...--C--T-AC----------AGT---TACAG--A---GGAT-A--TAA-A-------------G--------GGCA---CCC-----C--AA---G-  4540
TAN.UG.x.TAN1               G--A...G-A-----G--ATC------T---...-----T-C-------CA-TTT--AAGAGA--CA--CGGTA-A---G-AC-G--------------------TGCT--CCCA--------AA-TCG-  4566
VER.DE.-.AGM3               ---A...G-A-----G--G--------A-GG...--C--T-A---------GAACT-A--AGACAC--A-GGG--A----AAA-T--G-------------------TG---CCA---------A-----  4089
VER.KE.-.9063               ---A...G-A--T-----A--------C-G-...-----T----------A-AATT-A--GGACACA---GGAT-A----AGA----T--C-----G--------A-TG---CCA--TGCCA-AA-----  4594
VER.KE.-.AGM155             ---A...G-A--A-----A--------T-GG...-----------C------AATC-A--AGACACA---GGAT------AAA-------T--------T-----A-TG--CCCA--------AG-----  4586
VER.KE.-.TYO1               ---A...G-A--T-----AA-G-----A-G-...-----T-A--------A-AACC-A--AGA-ACA-A-GGG--T----AAA-------------G-----G----TG---CCA--------AA----G  4081
SUN.GA.98.L14               ---A...CCA--TGT---A-------G---G...-TC--T-AC---GCAGCCT-TT--CAGGAGATG---G--A-A--C-TA--G--------GC-G-----CAATGAA---CAC--------A-AG---  4727
LST.CD.88.447               ---G...CCC--A--G--A--GT--AGC---...-T---T-A----GCT-A--ATT-A-AAGAAA-A---C--A-A--C---A-G-----T--GC-------GAAT--T---A-----------AC---G  3644
LST.CD.88.485               ---G...CCA--A--G--A--G---AGC---...-T---T-A----GCT-A--ATC-A-AAGAAA-A---C----A--C---A-G-----T--GC-------GAA---T---A-----------AC---G  3644
LST.CD.88.524               ---A...CCA--TGT---A--G----GA--G...-T-----A----GCA-A--ACT-A-AAGAAA-G-A-GGGA-A--C--CA-G-----T--GC--------AATGAG----C---G-------AG--G  3641
LST.KE.-.lho7               ---A...CCT--AGT---A------AGC---...-TC--T-AC---GCA-AG-ATC-A-AAGAAA-A--------A--C--CA-G-----C---C----T--CAATGA-----CA--C-----AAAG---  4728
SYK.KE.-.SYK173             TCCC...C-A--A--G--A--T-----A-GG...--C---TCA---AT-GAGTATC-TAGGCAG--G---C-T--A--C-GACAG----------C---TA--CAACAA---CCA--------AAAT-G-  4567
SYK.KE.KE51_AY523867        CCCA...--A--A--G--A------------...-------C----GTAGA-TAC--AAGGCAA--A---C-T--A--CAGACAG-----C----CT------AA--TG---CCC-----C--AA-C-G-  4222
COL.CM.-.CGU1               ---A...GCA--AG----A--C---C-A--G...-T---T------GT-CAGTATC-AAAAGAAC-A---GGGT-A--CA-------------------TTGG-AA--A---TCAG-G-----AAC----  4130
```

```
                                                                            | Asp64 catalytic site
H1B.FR.83.HXB2            GGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATTC  4497
Pol                       _G__E__A__M__H__G__Q__V__D__C__S__P__G__I__W__Q__L__D__C__T__H__L__E__G__K__V__I__L__V__A__V__H__V__A__S__G__Y__I__E__A__E__V__I__  Integrase

H1D.CD.84.84ZR085         --------A-----------------------------------------T----C---------------------------------------------------C----------------------  4019
H1F1.BE.93.VI850          --G-----------C-----------------------G---------T-------C---------------C--G-----------------C--------T----AG--C-------------------  3837
H1G.SE.93.SE6165          --G---------------------------------------------T--------------------------A-----A-A-----------------------C----------------------  3894
H1H.CF.90.056             --G-----------------------------C---------------T-------C--------G------C-------T-----------C--------------C-----------------C--C-  3844
H1J.SE.93.SE7887          --G-----------------------------------G---------T--------------C-------------------------------------------C--------------------C-  3811
H1K.CM.96.MP535           --G--------A------------------------------------T--------------------------A----------------C--------T--C--C--------------------C-  3693
H101_AE.TH.90.CM240       --G-----T-----------------------------G---------T-------C------C-------------C--------------C--C--G-------------------------------C-  4071
H102_AG.NG.-.IBNG         --G---------------------------G-----------------T----------------------G---A-----T-A--------C--------------C--------------------C-  4022
H103_AB.RU.97.KAL153_2    --------------C------------------------------------------------------------A-------A----------------------------------------------  3723
H104_cpx.CY.94.CY032      --G--------------------G--------------G--------GT----------------------T---------A----------------G--T--------C-----------------C-  3863
H1N.CM.95.YBF30           --G-----------------G--CA-T-------------G-G------T-------------C-----G------A-C-----T--G--G--C-----G--------C--CT------------------  4091
H1O.BE.87.ANT70           --G-----A--T-----T--G--------C-------A-G--------A-------C------A-G-----C---A-C--AA-A--T--T--C-----G--A-----G-TC--------------G--A-  4552
H1O.CM.91.MVP5180         --G-----A-C------T----------AC--C----AG---------A-G--------------------C---A-C--AA-A--T--T--C--------A----AC-T------------G--G--A-  4527
CPZ.CM.04.EK505           -----------A-----G---A-CA-------------GG--------T-------C--------G--G------A----------T--------C-----G--------C--CT-------------C----  4069
CPZ.CM.04.LB7             -----G--A-----------G-----------C--------------GT-------------CC--------------C---T-A--------G--------T-----C-----G-----------C--A-  4052
CPZ.CM.04.MB66            -----G-----A--------G---------------------------T----C--C-----C--G--------GA-C-----A--G-----A--------------G-----G-----------A--A-  4043
CPZ.MT145                 --------A--------------T--------C--------------GA-T-----C--C---------------A-A--TA-A--G-----C-----G-----------C--------------C--C-  4039
CPZ.CD.-.ANT              ----------G------------------C------------------G-------C--C-----------T-----A----A-A--------G------T-T-----C-TC--------T-----A--GG  3936
CPZ.CM.-.CAM3             -----G-----------G-----------C--C-----C---------T----C--C-----CC----------------------G-----A--------------G--CC----G-----G--C----  3911
CPZ.CM.98.CAM5            --G-----------------------------------C--T------T----C---------C-----------A-C-----A--G-----A-----G-----------CC----G-----G--C----  4199
CPZ.GAB_X52154            -----------------G--G-----------C-----G--T------G----------C--CC-------C-----G---A-A--G--------C-----------C--------------------C-  4558
CPZ.GAB2                  --------A--A-----------G--T--C--------G--C-----G--G--C--C--C--C--------G---A-A--------G-----C-----G--A----------CT-----------A----  3890
CPZ.TZ.-.TAN1             -----G--T--A--------G--G----AC-------A---C------A----C-----C--CC----------G--C---A-A--------G--------T-----T-TC--------------C--A-  4141
CPZ.US.85.CPZUS           -----G-----A-----G--G-----T--C--------T--C-----G-----------------------G---A-C--TA-A--G-----A-----G-----------CC----------G--C----  4557
GSN.99CM166_AF468659      --G---C-A-AA-----G--G---A-TGCAGAATT---C--G-----GA-G--C--C--G-----G--------G--C---T-A--G--G--C--C------A-----G---G-CTGG--TAGGA-A--A-  4225
GSN.99CM71_AF468658       --G---C-A-AA-----G------A-TGCAGAATT---C---------A-G--C--C------C-G-----C---------T-A--G-----C--C-----------G---G-TTGG--TAGGA-A--A-  4207
H2A.ALI_AF082339          -----G--T--A-----G-----GA-TGCAGAA-T---C-CT------A-G--C--C-----C----------------C---A-AA-------G-----T-----------TC-----------G--C--C-  4872
H2A.BEN_M30502            -----------A-----G------A-TGCAGAAAT---CG-T------A-G--C-AC-----C--------------A-C--TA-A--------A-----T--A-------TC--------------C--C-  4879
H2A.ST_M31113             --G--G--T--A-----G------A-TGCAGAATT---C-CT------A-G--C--C-----C--------------A-C--TA-A--------A-----T--A-------T---------------C--C-  4324
H2B.D205_X61240           --G-----T--T--------G---A-TGCAGAC-T---G-C------GA-G--C-------------------------A----AA-A--G-----C--------------G-T------------G--A--A-  4852
H2B.EHO_U27200            --G-----T--T--------G---A-T-CAGAA-T---G-C-------A-G--C---------------G----G-----AA-A--G----------------------------TC--------------A--A-  4850
H2G.ABT96_AF208027        -----------T--------G---A-TGCAGAATTG--T-CT------A-G--C--C--C-----------------A----A-A----------------G--------C-T-G-------------A--C-  4206
H2U.FR.96.12034           --------------------G---A-TACAGAAGT---G-C-------A-G--C-----C--C-----G---------G--A-C--G-----A-----G--------C-T---------------G--C-  4358

MAC.239_M33262            -----G--T--A-----G--G-C-A-T-CAGA--T---G-CT------A-G--------C---C----G------A-A---A-A--T-----A--------T-------TC------------G--A----  4796
Pol                       _G__E__A__I__H__G__Q__A__N__S__D__L__G__T__W__Q__M__D__C__T__H__L__E__G__K__I__I__I__V__A__V__H__V__A__S__G__F__I__E__A__E__V__I_  Integrase

SMM.SL.92.SL92B           --------G--A-----T--GAC-A-TGCAGAGGT---C-C------GA-G--C--C--C--C-----------G------A-A--------A-----G--T-----C-T------------G--A--A-  4227
SMM.US.x.PGM53            -----------A-----G------A-TGCAGAA-T---G-CT------A-G--C--------CC----G--C---A-A---A-A--------A-----G--T-------TC-----G--------A----  4713
MNE.US.-.MNE027           --------T--A-----G--G---A-T-CAGA--T---G-CT------A-G--------C---C----G------A-A---A-A--T-----A--------T-------TC--------------A----  4264
RCM.NG.x.NG411            --------AG-A--------G-----TGC---------T-CT------A-G-----C------C-----------A-C---A-A--------C-----------------C---------------A--A-  3991
RCM.GAB1_AF382829         --G-----A-C------G-----G--TGCA--------G-CT------A-G--C--C--------G---------A-A---A-A--------C-----G--A----------------G--------A--A-  3976
DRL.FAO_AY159321          -----GC-A--A-----C--------TGCA--C-------CT------A-G--C--C-----------------GA-A---A-A--G-----C--C-----A------C-C--------------A--A-  4064
MND-1.F17                 -----GC----AA-G-----GAC---TGC-TCA-TT--G-C------GA-------C------C-G---AATC--A----TA-A--------G--------TTCA--C-TC--GA-G-----------AA  4184
MND-2.5440_AY159322       ------C-TG-T-----G--G---A-TGCAGA------G-C------GA-G--C--C--C---------------A-C---A-A--------C------------------T----------G--A--A-  4058
MND-2.14CG_AF328295       ------C-T--C--------G--G--TGCGGCC-------C-------A-G-----C--C---C----------G------A-A--------C-----G------------C-------------G--A-  4425
MND-2.MND2_AF367411       -----GC-A--A-----C--G-----TACATCC------CC-------A-G--C--C------A-G---------------A--A-A-C--------C--------A---A-G---C-------------A--A-  4498
DEB.CM.CM40               --G--GC---CA-----T--G--G--TGCCTCAATG--C--C-----GA-G--C--C--C---C-G---AAC---A-A--AA-A--G-----C-----G--T--------------GA-CA-GA-CC-A-  4303
DEB.CM.CM5                -----GC-T--------T--------TGCCTCCATG--GG--------A-------C--C--CC-----AAC-----C----A-A--------G--------T-------T---G--GA-TA-G--CT-AT  4297
MON.CM.L1                 ------C---AATCA-----G---A--GCAGAAGT---G--------GA-G--C--C--C---C----G----GCA-A--TT----------A--------------T----G--TGG--CA-GA-CC-A-  4200
MON.NG1                   ------C---AA--------G-----Y-CAGA-AT---G---------A-G-----C--T----AT-------GTA-C-----A-----C-----------T-----C--CG-TTGG--CA--A--C---  2793
TAL.CM.8023               ------C-A-A-ACA-----GACCA--GCAGCAGT---G-C------GA-G-----C--C--CC-G------C-G--A--ATGT----------A-----G--A-----T--C--T--GA--A-GA-CT-G-  3961
TAL.CM.266                ------C-T-A-ACA-----GAACA-TGCAGC-GT---G-C-------A-G-----C--C-----G--G---C--A-C---TGT----------A--C--G--------G--C-----GA-TA-GA--C-G-  4463
MUS.CM.1085               ------C-A-AAACA--G-----CA-TGCAGAGAT---GG--------A-G-----C--C--CC-G--G------A-C--T--A--------C-----G--T-----G--C--GTGG---A-GA-AT-G-  4187
SAB.SN.-.SAB1C            --G--GT-TG-A-----G--------TGC----------G-G------A-G-----------CC-----------A-C---A-A--G--T--------T--T-------T---------------A--C-  4895
GRV.ET.x.GRI_677          ------C-T-A---------G------GCCTCCATT-A--CT-----GA-G--C--C--C--------------------AA-A--------A------------------TC-----------G--G--C-  4670
TAN.UG.x.TAN1             ------C-G--A-----G--------TGCC---GT---GG-C------A-G--C-----------G---------A-C--TA-A--------C-----G--------C-TCT-------------C--AG  4696
VER.DE.-.AGM3             --G---C-AG-C-----G--------TGCCTCG-----GG--------A-G--C-----------------C--G--A---A-A--------C------------------TC-----------------A-  4219
VER.KE.-.9063             --G---C-A--A-----G---------GCCTCC-----GG--------A-G--C--C--G---G-----------A-AG--A-A----------------G-----C----TC--------------C--A-  4724
VER.KE.-.AGM155           --G---C-A--A--------------TGC-TCA------G-G-----GA-G--C--C-----A------------A-AG-GA-A-----G--C--------------G-T------------G-----C-  4716
VER.KE.-.TYO1             ------C-AG-------------G--TGCCTCA--T----C------GA-G--------T--C------AA------AG--A-A--T--G--C------------------TC--------------C--A-  4211
SUN.GA.98.L14             --G--------AACA-----------TGCCTCAGTG--C--C-----GA-------C--C--CA-G----A------C--AA-A-----G--G-----G--TTCA--G-----G-T----------AC-A-  4857
LST.CD.88.447             -----------AACA-----------TAC-TCA-----T--T------A----C--C--T--CA-G--G---C-G--A---A-TAAT-----------G---TC-----T---G-T---------A----  3774
LST.CD.88.485             -----------AACA-----------TAC-TCA-----T--T------A----C--C--T--CA-G--G---C-G--A---A-TAAT---A-C-----G---T------T---G-T---------A----  3774
LST.CD.88.524             -----------AACA--T--------TGTATCAGTG--T--T--------G--------T--CC-------GC--A-C---A--AAT-----C---A-G--ATCA--C-TC--G-TT---------G----  3771
LST.KE.-.lho7             -----------AACA--------G--TGTCTCAGTG--T--T-----G-----C-----T--C---------C-G--C--TA-AAAT-----C-----G---TCA--C-TC--G-TT-----------A-  4858
SYK.KE.-.SYK173           ------C-A-AA-----G--------TGTAGA-AT-TAT-AT-----GA-G--C-----T---GAG-----C-----A---TGT--------AA-CAC---T-----------------A-CA-GA-CT-AA  4697
SYK.KE.KE51_AY523867      ------C-A-AA-----G---------ACAGA--T---T-C-------A-G--------C--C----------------A--TTGT--------AA-CAC---------G----C----A-TA--A-AT-GA  4352
COL.CM.-.CGU1             ---C----TG-T-----C--GC----T-A---CTAT--CT--------T----------T---GAG--------G-------T-A--------A--C--CTGT-C-CTG-TCTGCTGG---ACCA-AT-AA  4260
```

```
                                                                                                                                  | Asp116 catalytic site
H1B.FR.83.HXB2               CAGCAGAAACAGGGCAGGAAACAGCATATTTTCTTTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATA...CATACTGACAATGGCAGCAATTTCACCGGTGCTACGGTTAGGGCCGCCTGTTGGTG  4624
Pol                          P__A__E__T__G__Q__E__T__A__Y__F__L__L__K__L__A__G__R__W__P__V__K__T__I__.__H__T__D__N__G__S__N__F__T__G__A__T__V__R__A__A__C__W__W Integrase

H1D.CD.84.84ZR085            -------------------G-------C-----C----------------------------GT-G-G...-----A---------------------A-------A----A---------C-----  4146
H1F1.BE.93.VI850             -------------A----------C--C--CA-A-----G-----------------------T----...-----A---------------------A----CG------A---AT----------  3964
H1G.SE.93.SE6165             -------------A-------------C---A-A---------------G----------C-GT----...-----A--T------------------A-----G-A--A-A---A--A--------  4021
H1H.CF.90.056                -------------AA------------C--C--G--G---C-----A-C--------------GT----...-----A--------------------GA-----G------A---A-----------  3971
H1J.SE.93.SE7887             -------------A------G------T----A-A---------------C-------------GT----...-----A--------------C------A---G-G-T--G-A---A-----------  3938
H1K.CM.96.MP535              ----------G--A--A-------C--C---A-AC---------------------------GT----...-----A----------C----------A-CA--GT-----A---A-----------  3820
H101_AE.TH.90.CM240          -------------A-----G-------C-----GC-----C-----------------------GT----...--C--A-----C--T--------------A-C---G-A----AA--A-----------  4198
H102_AG.NG.-.IBNG            -------------A-----G-------C---A-A---------------------------G---GT----...--C--A---------------------A-----G-A--A-A---G--A--------  4149
H103_AB.RU.97.KAL153_2       -------------A--------------C---G-C--------------------------------T----...-----A---------------------A--A--G------A---T-----------  3850
H104_cpx.CY.94.CY032         -------------A----------C--C--CA-AC--------------------------G----TG---...---G-A-----C---CC----------A-----G------A---A-----------  3990
H1N.CM.95.YBF30              -T-----G-----A-----------------A-------G-----T-----------------GTT---...--C-----T-----ATC---------TA----C--T--A-AA--A-----------  4218
H1O.BE.87.ANT70              -------------A--A-----T--C--C--C--G------C-G--T-C---------T--T---GT----...-----A-----C--GCCT-----T--AA--A-A--TA-G-A---T--A--------  4679
H1O.CM.91.MVP5180            -------------A--------T--C-----C--G---------------C---------T--C---GT----...-----A---------ACCT-----T--AA----AG-CA-G-AA--T--A--------  4654
CPZ.CM.04.EK505              -T-----------A--------T--------A-CC---G-----T------------------GTC---...--------T-----G-----C------A----C--T--A-AA--A-----------  4196
CPZ.CM.04.LB7                -T-----------A----------T------A-C-----------TCG----------------TT---...--------T-----T---------T--TA-CA--G-A--C-A---T--A--------  4179
CPZ.CM.04.MB66               -T--T--------A--A-----------C---A-C---------G-----C-------------GTC---...--------T-----A--T--C--T---A--AG---T--C-A---A--T--------  4170
CPZ.MT145                    -T--T--------A----------------------G-CC-T--G-----------------------G--T---...--C--A---------GCCT---------TA--AA-G-A--A-A---A--A--------  4166
CPZ.CD.-.ANT                 ---AT--G------A-AAGT--------C--C--G-----------CA-C------------------------...--C------------AGCT---------AA----AG-A--A-AA--G--A--------  4063
CPZ.CM.-.CAM3                -T-----G--T--A-----------T------A--C------------------------T------GT---C...--C-----T-----GCCT--C--T---A--AA---A----A---A-----C-----  4038
CPZ.CM.98.CAM5               -T-----------A--A--------T------A--------------------------T------GT----...--------T-----GCCT--C--T--TA--AAC--A--C-AA--A-----------  4326
CPZ.GAB_X52154               ----T--G-----A--A--------T-----C--G--------------T----------------T--T...--C--A--T-----GCCA-----T--AA-----G-A--C-A---T-----------  4685
CPZ.GAB2                     ----T--G--T--A--------T---------A-C--------------G---------------G-AG---...--C--A--T-----A-CA----------A----AG-A--A-AA--A-----------  4017
CPZ.TZ.-.TAN1                ---A---------AAGA-----C--T--C--CA-CC-------G--------------T-----G-A----...-----A--T-----ACCA-----T--TA--A-AG-A--G-A---A-----C-----  4268
CPZ.US.85.CPZUS              -T-----------A--A--G-----C------A-C------------------------T------GTG--T...--------T-----ATCT--C--T--AA--AG---A-----A--A--T--C-----  4684
GSN.99CM166_AF468659         -CCAG--------CAGAC----T----TGAAG--C---G---------C--C-------T----CCCACC--...-----A--T------CC-------GT-TCCAAGGAAT-GGAA--A-----C-----  4352
GSN.99CM71_AF468658          -CCA---------CAGAC----T----TAAAG--A---G---------C--C-------------CTCACC--...-----A--T------CC--------T-TC-AAGGAAC-GGA---A--T--------  4334
H2A.ALI_AF082339             -GCAG---T----AAGAC--------CTC--C--A------C-G--TA-T---------A---C-CACT-G...--C--A--T-G---TGT---C------TCACAGGAA--A-A-ATG-TAGCA-----  4999
H2A.BEN_M30502               --CAG---T----AAG-C-G------CTC--C--A------C-G--CA-T--G------A---CGCACT-G...--C--A----------CC---C-----TTCACAGGAA--G-A-ATG-TGGCA-----  5006
H2A.ST_M31113                --CAG---T----AAG-C----G---CTC--C--AC-----C-G--CA-T--G------A---C-CATT-G...--C--A---------TGC---C-----TTCACAGGAA--A-A-ATG-TGGCA-----  4451
H2B.D205_X61240              -CCA---G-----AAGAC-G-----TCTC--C--AC----G--G--CA-C---------TA-C-C-CACC--...--C--A-----C--TGC---C------TCAC-A-GT--A-A-ATG-TAGCC-----  4979
H2B.EHO_U27200               -CCA---------AAGAC-G-----TCTC--C--G-----GC-G--CA-C---------TA-C-C-CACC-G...--C--A-----C--TGC---C-----TTCACAAGAT--G-AAATG--AGCC-----  4977
H2G.ABT96_AF208027           --CA---------AAG-C----------TA--C--G-----------------G------TA-C-C-CACC--...-----A---------TGC---C------TCGCAAGA---G-A-ATG-TAGCC-----  4333
H2U.FR.96.12034              -GCAG-----T--TAG-C-------CCTG--C--AC----GC----------G------TA-C-CTCATT-G...--C--A--T-----TGC---C-----TTCACAGGAA--A-A-ATG-TAGCA-----  4485

MAC.239_M33262               --CA---G-----AAGAC-G------CTA-----G--------G-----C---------TA-T-C-CATC--...--C--A--T-----TGCT--C--TG-TTCGCAAGAA--A-A-ATG-TTGCA-----  4923
Pol                          P__Q__E__T__G__R__Q__T__A__L__F__L__L__K__L__A__G__R__W__P__I__T__H__L__.__H__T__D__N__G__A__N__F__A__S__Q__E__V__K__M__V__A__W__W Integrase

SMM.SL.92.SL92B              -TAGG--------AAGAC-G--------TG--C--C------A-----TC---------TA----GCATC-G...-----A-----C--TGC----------TCCCAGGA---G-A-ATG-TAGCA-----  4354
SMM.US.x.PGM53               -GCAG--------AAG-C---------CTG-----GC----------TA-C---------CA-C-C-CATC-G...--------T-----TGC---------ATCACAAGAA--G-AAATG-TTGCC-----  4840
MNE.US.-.MNE027              --CA---G-----AAGAC-G------CTA-----G--------G---A-C---------TA-T-CGCATC--...--C--A--T-----TGC---C--T--TTCGCAAGAA--A-A-ATG-TTGCA-----  4391
RCM.NG.x.NG411               -------G-----AA-A--G-----TC----C--G--------------CC------------G---CATT--...--C--A---------A-CA--C--T--AA----AG-A--GCAA--A-A---------  4118
RCM.GAB1_AF382829            -------------AA-A--G------C----C--G--G---C------C--------------GG-AGC--...--C--A--T-----AGCA---------AA----AG-A--ACA---G-T---C-----  4103
DRL.FAO_AY159321             -------------AG--A-------T---------GC----------T--C-----------TC-CACT-G...--C--A--T-----ACCA----------A---AA-A---AGCAA-A-TA--C-----  4191
MND-1.F17                    ----------T--AA-AA-G------G-G-----G-----G-------C-CA------TA-T-GT-A-C--...--C--A--T-----GCCT--C--T--TA--CAGGAA--AGAAA--ATG--------  4311
MND-2.5440_AY159322          -------G-----AA-A--------GC-C--C--G-----GC-------T--G-------------CATC--...--C-------------TCCT--C--TGT-A---AA-A---GGCCA-A-T---C-----  4185
MND-2.14CG_AF328295          -------------AA-A-----------C----C--GC-----C-------------------G---CATT--...----------------TCCA--C--TGT-A---AG-A---AGCCA-A-TG--------  4552
MND-2.MND2_AF367411          -CA----------AA-A---------C-C--C--G---------------T---------------CATC--...--C--A--T------CCT----------A---AA-AA--AGCCA-A-T---------  4625
DEB.CM.CM40                  -------------AA-A--G-----TCTG--C--G---C-G---G--C--------------CTCAG---...--C--A---------ACC---C--T-T-A--CAAC-A--C-A---A--T--C-----  4430
DEB.CM.CM5                   ----------T--AA-----------C-TA--CT-A---C-G--G-GG-C---------CA-T---CAG---...--C--A---------ACCA-------TAA--CAGC-A--C-A---A--A--------  4424
MON.CM.L1                    --AGG--------AA-ATGT----GGATAGCC--G--GG-GC-G--C-CT-TG-----G--C-CTCAG---...-----A--T-----GCCA----------A-CCAAGAAT-CGAA--A--TGCA-----  4327
MON.NG1                      -CAG----------A-A-------GCAT-GCC--G--GG--C-T----CG-C-------A--C-TCATG--...--C-------------ACCA----------A--CAGGAAT--GCA--Y--TGCA-----  2920
TAL.CM.8023                  -CAG------C--A-G------C--CCTG-----C--GC-GG-G--CA-C---------A--G--CATC--...--C--A---------GCCG--C--TGT-TCG--AGA-A-GCA----A-AGCG-----  4088
TAL.CM.266                   --AGG--G--G----G---G--C--CCTC-----C---C--G-G--CA-C--------TA---GCCACC-G...-----A---------GCCA--C--TGT-TCA--AGAAA-GCA----ATGGTA-----  4590
MUS.CM.1085                  --AG----------AG-C-G-----T-TAGGC--AC-GG-G------TC-CT------T-CG-C-CA---C...-----A--T-----GGCT--C----T-A----AGAAT-CGAA--A--A--C-----  4314
SAB.SN.-.SAB1C               --CAG--------AA-A-CT------C-C--C--GC-G--G-----CA-C---------CA-C-C-CAGC-G...-----A--T-----T-C---C--T--TA-CCAGCAA--AGCA--TATT--C-----  5022
GRV.ET.x.GRI_677             --AG------T---A----G------C-C-----GC-G---C-GTT--C-------------G---CATC--...--C-----T------CCA--C--T---TC-CAG-AT--GGCA--G-TG--C-----  4797
TAN.UG.x.TAN1                -TAGG--------AA-A---------A-A--C--G------A-CATTA-T--G-----TA-T-C--A-T--...----------------GCCA-----TGTTA--CAGGA---ACAAA-TATT--------  4823
VER.DE.-.AGM3                -TAG----------A-A---------A-A---T-A-----GA--CT-A-T---------CA---CCCA-C-G...-----A---------ACC------T--GTC-CAAGAA--AGCA--AATG--------  4346
VER.KE.-.9063                -TAGG--------AAGA--G-----CA-A--C---C-----A--TTGA-T---------A--GTCCAGT--...--C--A--T-----ACCA-----T--TTCCCAAGA---AGC---AAT---------  4851
VER.KE.-.AGM155              -TAGG--------AA-A--G------A-G--CT-G------A--AT-------------CA-C-CTCACC-C...-----A--T-----ACCA---------TTC-CAGGAA--AGCT--TATG--C-----  4843
VER.KE.-.TYO1                -TAGG--------AA-A-----G---A-G-----A------A--CTGA-T--------TA---C-CAGT--...--C--A---------GCCT--C--T---TCCCAAGAA--GGCA--AATA--------  4338
SUN.GA.98.L14                -CAAT--GCA----A--AC------CAC--GG--G-----GC-CTGT-CC-TG--------------CAG---...-----A--T-----GCCA-------T-A--AAAGAT--GGA---A-TA--------  4984
LST.CD.88.447                ---AT---------A-AAC-------C-C--C--G------C-GTGCA-C-------------C-GCA-G-C...-----A---------ACCA--C--TGT-A-CAAGGA---ACAA----TAAT------  3901
LST.CD.88.485                ---AT---------A-AAC-------A-C--C--G------C-GTGCA-C-------------C-GCA-G-C...-----A---------ACCA--C--TGT-A-CAAGGA---ACAA--A-TAACC-----  3901
LST.CD.88.524                ---AT---------A-AACT------A-----T-G-----G---TGTA-C---------T------CA---T...--C--A--T-----TCCA-----TGTGA--AAAGAT--ACA---A-TGACC-----  3898
LST.KE.-.lho7                ---AT--G-----AA-AAC---GT--C----C--G---------TGCA-T---------T------CA----...-----A---------TCCA-----TGTAA--AAGGA---ACA---A-TAACC-----  4985
SYK.KE.-.SYK173              A-AGG-----G---G-T----------TG-----CA-GC--A----CA-T---------A-----CAG---...--C-----T-----ACCA--C--TGT-A---AC-A-T-C-AA--A--A--------  4824
SYK.KE.KE51_AY523867         A-AG-----------A-------G-CTA--C-----GC--A------C-----------A-T---CAT---...-----A---------TCCA--C----T-A---A-G-CT--GC---A-----------  4479
COL.CM.-.CGU1                AGAG----------G-A-------GTAGGGCC--CA-------G--TA-TCAG---GAG--T-G-CA-G-T...--C--A--T-----GCCA-----TGTAA--CAGCACT---A---A--AGTG-----  4387
```

```
                                                                                     | Glu152 catalytic site
H1B.FR.83.HXB2            GGCGGGAATCAAGCAGGAATTTGGAATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCA  4754
Pol                       __A__G__I__K__Q__E__F__G__I__P__Y__N__P__Q__S__Q__G__V__V__E__S__M__N__K__E__L__K__K__I__I__G__Q__V__R__D__Q__A__E__H__L__K__T__A_  Integrase

H1D.CD.84.84ZR085         ---A--T-------------------------------------------------------------------------------------------------------A---------------------  4276
H1F1.BE.93.VI850          ---A--T---C-A--------------C--------------------------------------A--------G-----A--G--C--------------------A---------------------  4094
H1G.SE.93.SE6165          ---AAAT----CA--------------------------------C--------------------------G--------------C--C--G-----C--G-----A---------------------  4151
H1H.CF.90.056             ---A-AT---C-A-----------G-----------------------G-----------------------------------G--C-----G-----------C--A--A-----C------------  4101
H1J.SE.93.SE7887          ---A-AT--------------------------------------------------------------------------------C-----------------A--A--------C------------  4068
H1K.CM.96.MP535           ---A--TG----------------G--------------------C-----------------------------------------C--------------G-----------------C--A------  3950
H101_AE.TH.90.CM240       ---CAATG--C-A-----------G--C--------------------------------------------G--------------C-----G-----------G--A--------C-----A------  4328
H102_AG.NG.-.IBNG         ---AAATG---CA--A-----------------------------C-----------G------------------------------------------C--------A--------C------------  4279
H103_AB.RU.97.KAL153_2    ---A--G-----------------C------------------------------------------------C--------C---C---------------------A---------------------  3980
H104_cpx.CY.94.CY032      ---A-AT-----C--------------------------------C-----------G-----------------------------C-----G-----C--G-----A--------C------------  4120
H1N.CM.95.YBF30           ---AAAT-----A-----------G--A-----------T--------G----C------G--C--------------------------------AA-C--------A--A--------A---------  4348
H1O.BE.87.ANT70           ---CAAC--AC-A--T--G--------A--A--T-----A---------------------G-C--------G--------ATC-------CAG-----G--G--C--A--A-----CT-A-GA------  4809
H1O.CM.91.MVP5180         -A-A--C--AC-A--T--G-----G--A--A--T-----A---------------------G-C-----------------ATCT------CAG-----G--G--C--A--A--G---T-A--A------  4784
CPZ.CM.04.EK505           ---AAAT---C-A-----G--------A-----------T--------G--T-C---------C--------------------------------AA-T--------A--A------T-A--A------  4326
CPZ.CM.04.LB7             ----CAG---C-------------------------------G-----------G--------A-----C------C-T--------A-----G-----------------A--G--C--A--A------  4309
CPZ.CM.04.MB66            ------T---C-A--A-----------------------T-----------G-----------------------------------A-----T---A----------A--A-----------A------  4300
CPZ.MT145                 ---A--T--T-GA-----------------A-----C--A-----------------G-----------------------A--G--C----AG--A--C------------------T-A--A------  4296
CPZ.CD.-.ANT              ---TAAT---C-A--------------A--A--------A-----------------G-----C---------C----G---C-------------AA-T-----C--A--A-----AT-A---------  4193
CPZ.CM.-.CAM3             ---A--C---C-A--------------------T-----A--------------G--------A--------G--------A--G--C--------AA-T-----A--A--A-----CT-G-G-------  4168
CPZ.CM.98.CAM5            ---T--C---C-A-----G--------C-----------A-----------------G-----C-----------------------C---------A-T-----C--A--A--G--G--A--------T  4456
CPZ.GAB_X52154            ---A-AC--------------------A-----T-----A--G-----------G--------CT-A--------GC----------A-----------T--G-----A--A------T-A--A------  4815
CPZ.GAB2                  ---A--C--A--A--A-----------A--T--T-----A-----------G-----------C-----C-----------A--G--------G--A--T-----C--A--A-----CT-A---------  4147
CPZ.TZ.-.TAN1             ---ACA---TC-A--T--------G-----A--T-----T-----------------------------------C--------C---------AG--A--C--G--C--A--A--G--A--G-G-------  4398
CPZ.US.85.CPZUS           ---A--C--AC-A--A--G-----------A--------A--------------G--------C-----------------------C--------AA-C--G-----A--A--G---T-A--------T  4814
GSN.99CM166_AF468659      ---A-AC--AC-A--CACCACA--GG-G--A-----C--G-----------G---------AAC-----C---C--C-C---G---CA--CACGA-AA-C-----AG---TCAC-T-CT-AG-A------  4482
GSN.99CM71_AF468658       ---AAAC--AC-A--CAGTACA--GG-G--A-----C--G-----------G---------AAC-----C---C--C-C---G---CA--CCA-A--A-T------G---TCAC-T-CT-AG-A------  4464
H2A.ALI_AF082339          --TT--T--AG----ATCC------G-A--T--------A-----C---------------G-A------C-CC-CC----A--TCAG---A-CAGAA-T-----G-----AA-TACAG-GG-A---AT-  5129
H2A.BEN_M30502            --TA--T--AG-A--ATCC------G-A--T-----C--A-----C--G------------G-A------C-CC-CC-------TCAG---A-TAGAA-T-----A-----AA-TACAA-AG-A---AT-  5136
H2A.ST_M31113             -ATA--T--AG-A--ATCC--C---G-A--T--------A-----C-----------G---G-A------C-CC-CC----A--TCAG---A-CAGAA-T-----G-----AA-CACAG-AG-A---AT-  4581
H2B.D205_X61240           --TA-----AG-A--AACT------G-A-----T--C--A-----------------G---G-A-----CC-TC-CC-G--A--TCAA----ACAGAC-C-----C--A--A-T-TCAA-AG------TT  5109
H2B.EHO_U27200            -ATA--G--AG-A--AAC---C---G-G-----T-----AG-------G------------G-A-----CC-TC-TC-G--A--TCAG----ACAGAA-T-----------A-T-TCAA-AG------TT  5107
H2G.ABT96_AF208027        --TA--C---G----A-C-------G-A--A-----C-----G--------G-----------A------C-TC-TC--------CAA----ATA-AA-T--------A--AA-TTCAA-AG-A--CAT-  4463
H2U.FR.96.12034           --TA--T--AG----TACC------G-A-----T--C--G-----C--------G------G-C------C-CC-TC-------CCAA----AGAGAA-C-----A-----AA-TACAA-GG----CAT-  4615

MAC.239_M33262            ---A--G--AG----CACC-----GG-A--A--------A--G-----G--------G---G-A------C-CC-CC-G--A--TCAA----ATAGAA-C--G--A--A--AA-TTCAG-AG-A--CAT-  5053
Pol                       __A__G__I__E__H__T__F__G__V__P__Y__N__P__Q__S__Q__G__V__V__E__A__M__N__H__H__L__K__N__Q__I__D__R__I__R__E__Q__A__N__S__V__E__T__I  Integrase

SMM.SL.92.SL92B           -TTA---G-GG-A---TCC------G-A--G-----C--A--------G------------G-C---G-CTT-C-TC-T-------AC----ATA-AA----G--A--A--A--GTCAG-AG----CCT-  4484
SMM.US.x.PGM53            ---A-----TG-----ACC--C--GG-A--T--T-----A--G--C-----------G---G-A-----CC-TC-TC----A-CCCAG----ATAGAA-T-----A-----AA-TTCAA-AG----TAT-  4970
MNE.US.-.MNE027           ---A--G--AG----CACC-----GG-A--A--------A--G-----G--------G---G-A------C-CC-TC----A--TCAA----ATAGAA-C--G--A--A--AA-TTCAA-GG-A--CAT-  4521
RCM.NG.x.NG411            ---TCA---AG----CACC------G-A-----------T-----C--------G--------A------C-TC----G----T---A-----G--AA-T-----------A---A-GT-AG--------  4248
RCM.GAB1_AF382829         ---TCAG--AG----C-CC------G-A--T--------T-----------------G---AGC---------C-------A-T---C----A---A--------A--A--A---A-AT-AG-A------  4233
DRL.FAO_AY159321          ---TAAG--AG----TACCACA---G-------T--C-----G-----------G--------A-----C--TC-GC-------------------AA-T-----------A---A-A--AG--------  4321
MND-1.F17                 -TTA--G--AG-A--CAC---------A-----T--C--A--------G--G-----G---AA--AA-----GT-TC----AG--T-G--T-AGA-AA-------AG-TTGCA--G-AT-A--A------  4441
MND-2.5440_AY159322       ---TCA---AG----CACCACA---G-A--T-----C--A--G------------------G-A-AA---C-TC-TC-T-----G--C----A---A--T--G-----A------AGG--AG-A------  4315
MND-2.14CG_AF328295       ---TCAG--AG----CACCACA-----A-----T--C-----------------T--G---G-A-AA---C-TC-TC-T---GT--------AG--A--T--G-----A--A--GAGA--GG--------  4682
MND-2.MND2_AF367411       ---TCA---AG----CACTACA-----------------T--G------------A----GG-A-AA--CC-TC-TC-C--AC----C-----G--A--T-----C--A--A--GA-A--AG-A------  4755
DEB.CM.CM40               -CTA--T--TG----CACCACA--GG-A--A--T--C--A-----C--G-----G------AG--A---C-GGATTC----AG-T-CC--CAA---AA-C------G-T--A--GAGAT-GG-A------  4560
DEB.CM.CM5                -TT------AG-T--TACTACA--GG-A--A--T--C--A-----------G---------AGC-AA----GG-TCC----AG----C--CAA---AA-T-----AG-T--A---AGA--AG-A------  4554
MON.CM.L1                 ---AAAT--TC-T--CACCACA--GG-A--------C--A-----C-----G--T-----GAA--------G-C--C----AG-G-CC--TAAGA-AA-C--G---G---T----T-C--CCCA------  4457
MON.NG1                   ---AAAC--AG-A--CAGTAC----G-G--------C--A-----------T----------AAC-------GC--A-C---A-CC-CAT----GA-AA-C--G---G-A-T-C-GTTC--CTCT------  3050
TAL.CM.8023               -CT-AA---AG----CACCACA--GG-A--A--T-----A--TCC------GTC------GAA--A---C---C-G--G-----G-CA--TCA---AA-T--G---G-A-TGC-GT-C--GTCC-----G  4218
TAL.CM.266                -CTCAA---AG----CTCCGCA---G----A--T-----T--TCC---G----C-------AAC-AA---C-GC-C--------G-CA--CAC---AA-T--G---G-A-TAC--T-C--GTCC------  4720
MUS.CM.1085               ---CA-C--AG-A--TACCACA--GG-A--A--T--C--T--G-----------G------AAC-AA--C---C-G------AG--GCC---CACA-AA-C--G--AG---TCACTT--T-GG-A------  4444
SAB.SN.-.SAB1C            --GAAA---AG----CAC------GG-A--------C-----------G---------------------GC-G------AG----C---------AA-T------G-T--A---AGAT-GG-A------  5152
GRV.ET.x.GRI_677          --GTAAT--AG----CACCAC------A--T--T--C--A--GTCA--G--TAGT------AGC-----C-G-C-GC-C--G---C---CTCT--AA--------G-TTG---GAGAT-GG--------  4927
TAN.UG.x.TAN1             --GACA-G-AG----CAC-ACA--G--C-----T--C--A--TCA------AGT-------AG--------G-C-------AG---C-----A-A-AA-T-----AG-CTG--C-TT-T-GG-A--G---  4953
VER.DE.-.AGM3             --GAAA---AG-A--CACCACA--TG-A-----T--C--T--TCA------CTCTA----GAG----------C-G-----AG----A--T--GA-AA--------G-CTG-C--T-CACAG-A------  4476
VER.KE.-.9063             --GAAAG--AG----CACCACA---G-G--A--------A--GTCA--G---TC-------AGC-----C---C--C-C--AG----A--T---A-AA--------G-TTG-C-GT--ACAG-A------  4981
VER.KE.-.AGM155           --GAAAGG-AG-A--CAC-ACG--GG-A--A--T-----A--GTCC--G---TCTA-----AG------C---C----G--AG-G--A--T---A-AA--------G-CTG-C--T--ACAG-A------  4973
VER.KE.-.TYO1             --GAAA---TG-A--TAC-ACA--T--A--A--T--C-----TC--------TC-A-----AGC-----C--GC-------AG-G--A--T--GA-AA--------G-TTGCC--T--ACAG--G-----  4468
SUN.GA.98.L14             -TTA--C--AC----TAC-ACA--G--A--T-----C--T--G--------------G---G-A-A---C--G-T------AC-------CA-CAGAA-T-----AG-T--AC-GG-AT-G--A--T---  5114
LST.CD.88.447             -CTA--T--TG-A--TAC-ACA--------T--T--C--T--G-----G--------T---G-C-AA------C-C-----AC---------A-AG---C-----AG-T---C----AT-A--------T  4031
LST.CD.88.485             -CTA--T--TG-A--CTC-ACA--G-----T-----------G-----G--------T---G-C-AA-------TGC-----C----C----A-AG---T-----AG-T---C----AT-A--------T  4031
LST.CD.88.524             -TTAA-T--AG----CAC-ACA--G--A--A--T--C--T--------------------GG-A-AA-------TT------C----C----A-AGA--------AG-T--AC-G--G--A---------  4028
LST.KE.-.lho7             -ATA--G--AG-A--TAC-AC---G--A--T--T--C--T--G--------G--T------G-A-AA--C----TT-----A-GT-------A-AGA--------AG-T--AC-G--A--A--A------  5115
SYK.KE.-.SYK173           -TGT--C--AG-A--CACTACA--T-----A-----C--A-----C--G---A-------G--C-AA----GGT-T-----AG--GC---CTC---AA-T-----CG-C-TCAC----T-GC-A------  4954
SYK.KE.KE51_AY523867      ---A---G--G-A--CACCACA-----A--A--------A-----C--G---A--------AAC-AA----G-C-------AG---CC--CAAT--AA-T--G--AG-A-TA---AGAT-AG-------G  4609
COL.CM.-.CGU1             -CTA-----AGCA--CACCACA---CAC-----T-----T--TCA---------T-----GCAG-GA--C--G--TG-T--A-G--AG--TAAGA-AA-G-A------A--A---ACAT-AG-A-GCAA-  4517
```

```
H1B.FR.83.HXB2              GTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGA...ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATT.  4880
Pol                         _V__Q__M__A__V__F__I__H__N__F__K__R__K__G__G__I__G__G__Y__S__A__G__E__R__.__I__V__D__I__I__A__T__D__I__Q__T__K__E__L__Q__K__Q__I__ Integrase

H1D.CD.84.84ZR085           --------------------T-----------G---------------------------------------...---A--------------T----T----------G------------------C.  4402
H1F1.BE.93.VI850            ------------------------------------------------------------------------...---A-----------T------------------G-------------------.  4220
H1G.SE.93.SE6165            -----G--------------T--------------------------------------------A------...---A--------------T----T---------------C--------------.  4277
H1H.CF.90.056               --------------------T---------------------------------------------------...---A--------------------------------------------------.  4227
H1J.SE.93.SE7887            --------------------A--------------------------------A------------------...---A------------------------------G-------------------.  4194
H1K.CM.96.MP535             ------------------------------------------------------------------------...--------T---------------------------------------------.  4076
H101_AE.TH.90.CM240         --------------------T---------------------------------------------------...---A--------------------------------------------------.  4454
H102_AG.NG.-.IBNG           -----G--------------T---------------------------------------------------...---A----T---------T----T---------------C----------G---.  4405
H103_AB.RU.97.KAL153_2      ------------------------------------------------------------------------...---A--------------------------------------------------.  4106
H104_cpx.CY.94.CY032        -----G--------------T---------------------------------------------------...---A--------------T----T---------------C--------------.  4246
H1N.CM.95.YBF30             --G--------G--T-----T-------------------------------------C-------------...---A-------------------------G--A-C-A-T-------C-------.  4474
H1O.BE.87.ANT70             -----------------TG-T-------------------------------------C------A-----G...---A--------T-----T--C-A--------A-C-------------------.  4935
H1O.CM.91.MVP5180           --------------C--TG-T-------------------------------------C---------G---...C--A--------C-----T--C-A--------A-C----C--------------.  4910
CPZ.CM.04.EK505             --------------C-----T-----------G--------------------A----C-------------...---A-------------------T--------A-C-A-GC------C-------.  4452
CPZ.CM.04.LB7               --------------------T--------------------------------A-----------A------...---A-------------------T----------C-------------------.  4435
CPZ.CM.04.MB66              --C-----------------T--------------------------------------------A------...---A--------C----------T----------C-A--------C--------.  4426
CPZ.MT145                   --------------C-----T--------------------------------------------A--G---...---A-------C--------------------A------C--------------.  4422
CPZ.CD.-.ANT                ---GT---------GCA---T-------------------------------------CAC-T--AC-G---...---T--------C----------------G--A-CTC----------T------.  4319
CPZ.CM.-.CAM3               --C--G--------G-----T-------------------------------------C------A------...--TA-------C--------------------A-CTA-TC--------------.  4294
CPZ.CM.98.CAM5              --C-----------C-----T-----------------------------------T-C------A-----G...--TA-------C--------------------A-CTA-T---------------.  4582
CPZ.GAB_X52154              --------------G-----T-------------------------------------C-------------...---A----------------------------A-GC------------------.  4941
CPZ.GAB2                    -----------T--------T-------------------------------------C------A------...-----------GC-----------T-------AC-----C--------------.  4273
CPZ.TZ.-.TAN1               ---ATC--------G-AT----------------------------------A-----C-------------...C--T--------C--A-T---A-T-----G--A---C-----------------.  4524
CPZ.US.85.CPZUS             -----G--------------T-----------------------------------T-C------A------...--CA-------C------T----AC-------AG-CTT----------------.  4940
GSN.99CM166_AF468659        ---GC-CA----TGC-A---T--T--------------G--A--A--A----ATATGT-CC-TACA------...------A-T--G--TCAC-----AT--G----AC--C-C---A-C-C-----A-.  4608
GSN.99CM71_AF468658         ---GC-CAA---TGC-A---T--T--------------G--A--A--A----ATATGT-CC-TACA--G--G...T----GA----G--CCAC-----AT--G----AC--C-C---A-C-C-----G-.  4590
H2A.ALI_AF082339            ----T---------TCATTG-ATG-------------GG--A--A--A----ATATG-CCC---CA------...C-CA-CA-T--G--CT-C-----ACA-G---TAC--TTCC-C----C-A---A-.  5255
H2A.BEN_M30502              ----TG--------TCATTG-ATG-------------GG--A--A--A----ATATG-CCC---CA------...C--A-CA----G--TA-C-----ACA-G---TAC--TTCC-C----G-A---A-.  5262
H2A.ST_M31113               ----T---------TCATTG-ATG-----------G-GG--A--A--A----ATATG-CCC---CA------...C--A-CA-T--GG-CA-TG----ACAGG---TAC--TTCC-C---GC-A---A-.  4707
H2B.D205_X61240             ----T-------ACTCA-TG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA------...C----TA----G---A-C-----GCA-G---TAC-GTTC--C---GC-A---A-.  5235
H2B.EHO_U27200              --GTT-------ACTCA-TG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA------...-----CA----G---A-T-----ACA-G---TAC--TTCC-C----C-A---A-.  5233
H2G.ABT96_AF208027          --GTT----------CATTG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA------...---A-TA----G--TA-T-----ACA-G---TAC--TTTCA-----C-A---A-.  4589
H2U.FR.96.12034             --C-T--------CTCA-TG-ATG-------------GG--A--A--A----ATATG-C-C-T-CA------...--CA-TA----G---T-C-----ACA-G---TAC--TTCC-C---TC-A---A-.  4741

MAC.239_M33262              ---TT---------TCATTG-ATG-------------GG--A--A--A----ATATG-C-C---CA------...T--A-TA----G--CA-T-----ACA-G-G-TAC--TTTCA----TC-A---AC.  5179
Pol                         _V__L__M__A__V__H__C__M__N__F__K__R__R__G__G__I__G__D__M__T__P__A__E__R__.__L__I__N__M__I__T__T__E__Q__E__I__Q__F__Q__Q__S__K__N_  Integrase

SMM.SL.92.SL92B             ----T--------C-CATTG-ATG-------------GG--A--A--A----ATATG-C-C-T-CA------...------A-T--G--CA-C-----AC--G----CC--T-C---A-TTC-----A-.  4610
SMM.US.x.PGM53              ----T---------TCATTG-ATG-------------GG--A--A--A----ATATG-C-C---CA------...T----CA-T--G--CA-C-----ACA-G---TAC--TTCCA----TC-A---A-.  5096
MNE.US.-.MNE027             ---TT---------TCATTG-ATG-------------GG--A--A--A----ATATG-C-C---CA-----G...T--C-TA----G--CA-T-----ACA-G---TAC--TTCCA----TC-A---AC.  4647
RCM.NG.x.NG411              --T-----------GC----T--------------------------------------------A-----G...---A--------------T-----T-------A-CCA----------C------.  4374
RCM.GAB1_AF382829           --C-----------T--GG-T--------------------------------------------A------...---A----T---------------T--GC---C--TA----------T------.  4359
DRL.FAO_AY159321            --T------------C-G--T--------------------------A----A-----------CA------...-----------T-----CT-T---C-CTT---A-GCA--C-----C--A-T---.  4447
MND-1.F17                   ---GCC-----CAC------T--T---------CA--GG--A--AC-A------ATG-CA-----A--G---...------A-T--G--CAAT-----AT--G--TA-C--T-TCA------T------.  4567
MND-2.5440_AY159322         --G-----------GC-A--T--------------------------A----A-------C-T--A------...---------------A-C--------C-T---A-CCA------G-C--A-T---.  4441
MND-2.14CG_AF328295         ---------------C----T--------------------------A----A-------C----A------...-----G--T------A-------T--T-T-----CCA--------C--A-T---.  4808
MND-2.MND2_AF367411         --T--------------A--T--------------------------A----AA------C----A------ATAG---ACATT---GC-A--GACCT-C--AC-------TT-CA---C--...T---.  4881
DEB.CM.CM40                 ---GC------TACC-A---AATT-------------GG--A--A--A----ATATGTC-C-GATA------...C-----A----G---A-------A---G----A---ACT------C--A-----.  4686
DEB.CM.CM5                  ---GC------CACC-----TATT-------------GG--A--A--A----ATATGGCA---ATA--G---...C-----A----G---A-T------T-GG----AC--ACC------C--A-----.  4680
MON.CM.L1                   T--GC-CAA---C-G------TTA--------------G--A--A--A----ATATGGCCC-T-TA--C---...T-TA--A----G---CAC-----AT--G--CTAC--ACC-C-A-C--C------.  4583
MON.NG1                     C-GGC-CAAA-GC-C-A---T-TA-------------RG--A--A--A----ATAT----C--ATA--G--G...T----CA----G--TCAT-----TT--G--CTAC--C-C-C-A-C--C------.  3176
TAL.CM.8023                 --GGC-CA---TACC-----T-TT--------------G-----AT-A----ATATGT--C---CA---GC-...C-CA--A----G--CTAC-----AC-------A-C-ACCC-------C------.  4344
TAL.CM.266                  --GGC-CA---GACT-----TTTG----A---------G-----AT-A----ATATGT-CC---CA---GC-...---A-TA----G--CTAC-----AC---------C-C--C-------C------.  4846
MUS.CM.1085                 ---GC-CAA-----C-----TTTA----------AG--G--A--AG-G----ACCTATCAAAT-CT--G--G...---A-CA----GC--CAC-----AC--G--CTAC--C-C------C------AA.  4570
SAB.SN.-.SAB1C              --CAT---------GCA---T------------------------------------TC-----CA------...C--A-CA-T------CAT-----AC--G----C---AC-C-T---C--A-----.  5278
GRV.ET.x.GRI_677            --G--------TACGCAT--------------------G--A--A--A-----TAT-TC-AGT-CA------...T-G--TA-T--GC--A-----C-AC--G--CTA--TACTC-------C-----C.  5053
TAN.UG.x.TAN1               --C-TC-----CTGCCA---T-----------------G--A--A--A-----TATG-CCC---CA--G--G...C--A--A----G---A-C---C-AC--G---TAC--C-CA------C----CAA.  5079
VER.DE.-.AGM3               ----TT------TGCCA---------------------G--A--A--A-------TA-CAC-G-CA--G---...T--A-CA-T--G--TA-T---C-AT--G--TTAC--C-CC------CCA-----.  4602
VER.KE.-.9063               ----TT------TGCCA---------------------G--A--A--A-------TA-CAC---CA------...C--A-TA-T--G---A-----C-AT--G--CTAC--AC--------CCA-----.  5107
VER.KE.-.AGM155             ----TT-----CTGCCA---T-----------------G--A--A--A------CTA-CA--T-CA--G---...C--A--A-T--G---A-----C-AT--G---TC--CACTC------CCA----C.  5099
VER.KE.-.TYO1               ----TG-----TTGCA-AC-T-----------------G--A--A--A-----AC-G-C-T---CA--G---...C--A-TA-T------A-----C-AT--G---TAC--C-T-------CCA-----.  4594
SUN.GA.98.L14               ----TC-----TC-GCA---T--T---------CA--GG--A--AC-A-----AATG-CA----CA--G--G...T-TA-CA----G--TAATG----TC--G----AC--T-C-----G---AT--A-.  5240
LST.CD.88.447               --C-T-------AC-CA---T--T---------CA--GG--A--AC-A------CT-TCAC---CA-----G...T-TA-TA-T--G--TAATG-----T--G----AC--T-T----------T--A-.  4157
LST.CD.88.485               --C-T----------CA---T--T---------CA--GG--A--AC-A------CTTTCAC---CA-----G...T-TA-TA-T--G--TAATG-----T--G----AC--T-T----------T--A-.  4157
LST.CD.88.524               --GTT---------CCA---T--T---------CA--GG--A--AC-G-------TATCAC---CA--G---...TACA-CA-T--G--CAATG----TT--G-----C--T-CA--------TT--A-.  4154
LST.KE.-.lho7               ----T---------CCA---T--T---------CA--GG--A--AC-A------CTT-CAC---CA--G--G...T-TA-TA-T--G--TAATG----AC--G----AC--T-TC--------TT--A-.  5241
SYK.KE.-.SYK173             --GGC-------ACC------TTA--------------G--A--A--A------ATATCAC----A-----G...TACA-TA-T--GC--TAT-----AC--.........C--------C--A-T-CAA  5072
SYK.KE.KE51_AY523867        --GGC------CACG-----TTTA--------------G--A--A--A------ACT-CAC---CA--T--G...TATA-TA----GC--TAC-----AC--.........C--C-------TAC-CAC.  4726
COL.CM.-.CGU1               ---GC---------C-ATGCG-T---------------G--A--AT-A------A-G---C--T--------CAAG---A-AGAGC----AT-GA-...T--G-T--AC--A-CC--...-C-A--T-AC  4641
```

```
H1B.FR.83.HXB2          .....ACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGT...GACATAAA  5002
Pol                     .__.__T__K__I__Q__N__F__R__V__Y__Y__R__D__S__R__N__P__L__W__K__G__P__A__K__L__L__W__K__G__E__G__A__V__V__I__Q__D__N__S__.__D__I__K Integrase

H1D.CD.84.84ZR085       .....------------------------------------------G-----A----------------------T-----------------------------------C------...--------  4524
H1F1.BE.93.VI850        .....--G--------------C--------------------------C---G----------------------A--------------------------C--------C------...--A-----  4342
H1G.SE.93.SE6165        .....------------------------------------------G-C---G----------------C--A--A-----------------------------------C----AC...--A-----  4399
H1H.CF.90.056           .....T----C--------A---------------------------G-C---A-------------------A--------------------------------------C------...--A-----  4349
H1J.SE.93.SE7887        .....------------------------------------------G-C---A-------------------A--A-C------------------------------G--C------...--A-----  4316
H1K.CM.96.MP535         .....TT---C--------A---------------------------G-A---A----------------------A-----------------------------------C--C---...--A-----  4198
H101_AE.TH.90.CM240     .....------------------------------------------G-C---A-------------------A--A-----------------------------------C------...--T-----  4576
H102_AG.NG.-.IBNG       .....-T----------------------------------------G-C--CA-------------------A--A--------------------------------G--C------...--T-----  4527
H103_AB.RU.97.KAL153_2  .....-T------------------------------A---------G-----A-------------------A--A-----------------------G--------G--C----AC...--T-----  4228
H104_cpx.CY.94.CY032    .....------------------------------------------G-A---A-------G-----------A--A--------------------------------G--C--C---...--T--C--  4368
H1N.CM.95.YBF30         .....TT----G-------------------------A---------G----CA----------------C--A--T--G--------A-----------G-----T--------CG-G...--T-----  4596
H1O.BE.87.ANT70         .....TT----N----C--A--------C--------A---------G-C--TA-C-----------G---C-------G-----------G--A--------C-----------GG-A...-----T--  5057
H1O.CM.91.MVP5180       .....TT------CA-C-----------C--------A--T------G-C--TA-------------G---C-A-----G-----------G-----------C-----------AG-A...-----T--  5032
CPZ.CM.04.EK505         .....TT----G-------------------------A---------G-G--CAC---------------C--A--A--G--------A-----------G-----T--------CG-G...--T-----  4574
CPZ.CM.04.LB7           .....T-----G-------A---------------------C-----G----TA----------------C--A--A--G-----G--------A--------GC-G------C-GGAA...--A-----  4557
CPZ.CM.04.MB66          .....T----T--------A---------------------------G-C--TA----------------T--A--A--------------G--A--------GC-C-----CC-AGAA...--A-----  4548
CPZ.MT145               .....TT----------C-A--------C--------A-------A-G-----A----------------G--A-----G--------------------G-----T------C-GGAC...--A-----  4544
CPZ.CD.-.ANT            .....TT----------C-A---------C----------T------G-C--TG-G---------------C-A--T--G-----------------------C--CA----CC-AGAG...--A--T--  4441
CPZ.CM.-.CAM3           .....TT----G----------------C------------------G----CA-------------T--C-GA--T--G--------------------G-----TA----G-GAGAG...--AG-T--  4416
CPZ.CM.98.CAM5          .....TT----G-----------------------------------G-----A-------------T--C-GA--T--G-----------------------G--TA----G--CGAG...--AG-T--  4704
CPZ.GAB_X52154          .....TT----G-------A-----------------A---------G-C---A-------------G----CCT-A--G-----------------------G--C--G---C-AG-G...--AC----  5063
CPZ.GAB2                .....CT------------A-----------------A---------G-C--CA----------------CTCT-----G-----------G--T--------T--T--G--C--CG-A...--T-----  4395
CPZ.TZ.-.TAN1           .....TT----G----------------------T------GC----G-----A-------G--------GCGA--A--G---------------------------A----AGGAGAA...-----T--  4646
CPZ.US.85.CPZUS         .....TT----G----------------C------------------G-----A----------------C--A--T--G--------------------------CA-G--A--CGAG...--GG-T--  5062
GSN.99CM166_AF468659    .....T-C---T-------A------------------C-AG-AGCT--C--T------C-----------GTA--------------------A---------G-C---AC-C-AGCAGGT--G--T-T  4733
GSN.99CM71_AF468658     .....T-C---T-------A------------------C-AG-AGCT-----T-A----C-----------GTA--T--------G--------A--------GG-C---ACCC-AGCAGGT--G--C-T  4715
H2A.ALI_AF082339        .....TTG---T--A-------C-C---C-----------AG-----G---AG--G-----------T-GGG----A--G--------G--C--A-----CA--G-TA-G-TAGGG-CA...--------  5377
H2A.BEN_M30502          .....T----TT--A-------C-A---C--------A--AG-----G---AG--G-----------T-GTG-A--A--G-----G--A-----A-----CA--G-CA-G-TAGGG-CA...--------  5384
H2A.ST_M31113           .....T-----T-A--------------C----T---A--AG-----G---AG--G-----------T-GGG-A--A--G-----G--G--C--A-----CA--G-CA-G-TAGGGGC-...--------  4829
H2B.D205_X61240         .....TT----T----------C-A---C--------A--AG-----G---A---C-----G-----T-GTG-A--AT-G--------G-----A-----CA-C---A-G-TAGGG-CA...--A--C--  5357
H2B.EHO_U27200          .....TT----T-C--------C-----C--------A--AG-----G---A---C-----G-----T-GTG-T--AT-G--------G-----A-----CA-C---A-G-TAGGG-CA...--A--C--  5355
H2G.ABT96_AF208027      .....T-----T--A-------------C--------A--AG-----G-C-AG--C-----------T-GTG----TT-G--------G-----A--T--CA-C---A-G-TAGGG-CA...--A--C--  4711
H2U.FR.96.12034         .....T-----T--A-------------C--------A--AG-----G-C-A-T-G--------C------G-CT-A--G-----G--A-----A--G--CC-T--TA-G-TAGGG-CA...-----T--  4863

MAC.239_M33262          .....T-----T--A-------------C--------A--AG-----G---A---G-----G-----C-GTG----AT-G--------G-----A-----CA-CT--A-G-TAGGG-CA...-----T--  5301
Pol                     .__.__S__K__F__K__N__F__R__V__Y__Y__R__E__G__R__D__Q__L__W__K__G__P__G__E__L__L__W__K__G__E__G__A__V__I__L__K__V__G__T__.__D__I__K Integrase

SMM.SL.92.SL92B         .....T-----T-------------------------A--AG-A---G-C-A---G-----G---------G----------------A--------T--T-----TA-G-TAGGG-CA...--A-----  4732
SMM.US.x.PGM53          .....T-----T--A-------------C--------A--AG-----G-C-AG--G-----------C-GTG----AT-G--------G-----A-----CA-CT--A-G-TAGGG-CA...--G--C--  5218
MNE.US.-.MNE027         .....T-----T--A-------------C--------A--AG-----G---A---G-----G-----C-GTG----AT-G--------G-----A-----CA-CT--A-G-TAGGG-CA...-----T--  4769
RCM.NG.x.NG411          .....T---------------------------T---A--AG-A--GG---AG-AG-----G--------T--A---A-----G----A--------T--C--G--C--G---C-AGAG...--TC----  4496
RCM.GAB1_AF382829       .....T-------------------------------A--AG-A--GG---A---G----G---T-----T-----GA----------A-----A--------C--C--G--GG-G-C-GGA---T-G--  4484
DRL.FAO_AY159321        .....CTC-----------------------------A--GG-A---G---AGT-G---------------G-A--TG----------G-----A--G--G--C--CA-G--AGG--C-...---T-G--  4569
MND-1.F17               .....T------A-TT-------AA------C-T---A--AG-A---G---AG--G-----------T-GT-TC--TT-G--------A--------------TT--A--T--C-AGAA...--G-----  4689
MND-2.5440_AY159322     .....T------------------A------------A--AG-A--GG-C-A--AG---------------G-----A-T--------A-----C--T--G-----CA----AGGAGT-...--TT----  4563
MND-2.14CG_AF328295     .....T-------------------------------A--AG-A--GG-G-AG--C---------------G-A---A-T--------A-----A-----G-----CA----AGGA-CA...---T----  4930
MND-2.MND2_AF367411     .....CA------------------------------A--AG-A--GG-C-A--AG---------------G-A---A-T--------A-----A--G--G-----CA----GGGG-CA...---T-G--  5003
DEB.CM.CM40             .....T-----C--TT-GGC--CAA---C--------A--AG-TGCTG-----AC---------G-------CC--G--G--------A-----T-----G--GTGCA-GACAG-GGTAGGA--T--C--  4811
DEB.CM.CM5              .....T-----C--TT-GG--ACAA---A--------A--AG-AGCTG-C---AC------G-----C--T-CCT-A--G--------A-----A--G--G--GTGTA--ACAG-GGTAGGA--------  4805
MON.CM.L1               .....------T--TC---A-------------T----ACAG-TGCCG-C--T-AG--------------GCGC--A-----------------T--C-----C---A--AC-G-GGAGGGC--T--CCT  4708
MON.NG1                 .....------T--TCG--A-------------T---AACAG--GCTG----TTCG-----G---------C-T--T--------G--G-----A--TT-G--G--TA-GACAG--CAAGGT--GG---T  3301
TAL.CM.8023             .....CAC--TT--TC-G-----AA---G--C--T--AA-GG-TGCT-----T------C-G--------TC-C---G-G--------A-----A-----C--CC--AGGACAG-CGAAGGA--GG---T  4469
TAL.CM.266              .....CA---TT--TCGG-----AA---C--------AA-GG-TGCT-----T------C-G--------TC-C---A-T--------A-----A--G--T--CT-G-G-ACAG-CGAAGGA--GG---T  4971
MUS.CM.1085             .....TAC---T----------------------T---ACAG-TTCTG----GTC------G-----C--CGTCT-A--G----G---------A--------C--CA--ACAG-CTCAGGA--GG---T  4695
SAB.SN.-.SAB1C          .....T----------------------C-----------AG-----G-C--TG-G--------T--G--T------A----------A-----A--------T--------AC-GG--...--AT-G--  5400
GRV.ET.x.GRI_677        .....CA-------TTG------AA---C--C-----A--AG-T---G-----G-G--------------GCGA---A----------A-----C--G--G-----TA----GGGGGAA...-----C--  5175
TAN.UG.x.TAN1           .....CA-------TC-------AA---C--C-----A--AG-A---G-G---G-G--------------C-CT-----------G--A-----A--G-----C---A----AGG-GAG...--GT-G--  5201
VER.DE.-.AGM3           .....CA-------TT-------A-A--C--C-----A--AG-G---G----TG-C-------------G-C--T-AA-T--------G-----T-----G--C--CA---GAGG-GTG...--AT----  4724
VER.KE.-.9063           .....CA-------TT-------A-A--C--C-----A--AG-G---G-----G-G-----G---------C-AT-AA----------G-----T-----G--C--CA----AGGAGAA...--GC-G--  5229
VER.KE.-.AGM155         .....CA-------TTG------A-A--C--C-----A--AG-----G-----G-G-----G-----T--TCGC--GA----------A-----C--G-----TC-CA-G--AGG-GAA...--AC-G--  5221
VER.KE.-.TYO1           .....CA-------TT-------A-A--C--C-----A--AG-G---G-C--TG-G-----------G---C-AT-AA----------G-----A-----G--CC-CA-G--CGGA---...---C----  4716
SUN.GA.98.L14           .....T--------TT---A---AA---C--C-----A--AG-A---G----T-AG-----------T--T--A--T--G--------G-----A--------C--CA--C-GGG-GAG...A----CCT  5362
LST.CD.88.447           .....T--------TT---A---AA------C-----AC-AG-AC--G----C-AG-----G-----T---C----T--G--------A-----T--T-----GG--A----AGGAGAC...AC---CTT  4279
LST.CD.88.485           .....T--------TT---A---AA------C-----AC-AG-AC--G----T-AG-----G-----T---C-AA-T--G--------A-----T--T-----GG--A----AGGAGA-...AC---CTT  4279
LST.CD.88.524           .....T--------TT---A---AA------C-----AC-AG-A---G------AG----G------T--TC-A--T--G--------A-----T--T-----GG--A-G--GGGAGAG...A----TTT  4276
LST.KE.-.lho7           .....T--------TT---A---AA------------AC-AG-A---G----T-AG--------------GC-A--T--G--------A-----T--T-----GG--A----AGGGGAG...A----CTT  5363
SYK.KE.-.SYK173         CATCAC-----T--TCG------A-------C---C-CC-GG-G-AG...AACGAG-----G--------TCGT--A--T--------------T--------GG-G---ACAG-GGAGGGA-----TTT  5199
SYK.KE.KE51_AY523867    ..ACACA----T--TCG------AAA--A-----T--AC-TG-T-CT...T-TGA----C-G--------TTCT--A--G-----G--A-----A--------G--C--GACACCAGA-CAAC-GG---T  4851
COL.CM.-.CGU1           AAAATCA----T--A--------AA--CC--C-GG-AA--GCA-...-CAGG-GAG---C---------G-G-A--GG-G--------------A--G-----C--CAG-A--TC-CAAGGTAC-T-GTT  4768
```

```
                                                             / Vif start                         Pol p31 Integrase end \
H1B.FR.83.HXB2           AGTAGTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAACAG...ATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAG.................................  5096

Vif                                                          _M__E__N__R__.__W__Q__V__M__I__V__W__Q__V__D__R__M__R__I__R__.__.__.__.__.__.__.__.__.__.__.   Vif
Pol                      __V__V__P__R__R__K__A__K__I__I__R__D__Y__G__K__Q__.__M__A__G__D__D__C__V__A__S__R__Q__D__E__D__*                            Integrase

H1D.CD.84.84ZR085        G-----A------------------------------------------...--------------------------------------------A.................................  4618
H1F1.BE.93.VI850         GA----A-----G--G---------------------------------...-------T----------------G-------------------A.................................  4436
H1G.SE.93.SE6165         G-----A------------------------------------------...------------------------G--------------------.................................  4493
H1H.CF.90.056            ------A---------G-G-----A------------------------...--------------------------------------------A.................................  4443
H1J.SE.93.SE7887         G-----A--------------------------A---------------...------------------------G-------------------A.................................  4410
H1K.CM.96.MP535          G-----A--------------------T---------------------...------------------------G-------------------A.................................  4292
H101_AE.TH.90.CM240      ------A------------------------------------------...------------------------G--------------------.................................  4670
H102_AG.NG.-.IBNG        G-----A-----------------A------------------------...------------------------G--------------------.................................  4621
H103_AB.RU.97.KAL153_2   ------A------------------------------------------...---------------------------------------------.................................  4322
H104_cpx.CY.94.CY032     ------A--------------------T--------C-----C------...---------A--------------G-------------A------.................................  4462
H1N.CM.95.YBF30          ------C---C-T--G--------A--A---------------------...-------------G-------------G----------AA--C--GAAATGGAATAG.....................  4702
H1O.BE.87.ANT70          G-----A--------G--G-----A--A--C--A--G------------...---------AC----A--A--------G-----ACA--AAG-G-AAGCGTGGAACAGCCTAGTGAAATACCATAA...  5181
H1O.CM.91.MVP5180        ---G--A-----------G-----A--A--C--A---------------...---------AC----A--A------A-------ACA--AAG-G-AAGCATGGAACAGCCTGGT...GAAATACCATAA  5156
CPZ.CM.04.EK505          ------C---C-T-----G--C--A--A-----A---------------...-------------G-------------G-----------AG-C--GACATGGAA.........TAG............  4680
CPZ.CM.04.LB7            G-----C---C-T--------T--A--T---------------------...----A-----C-A-----C-----GA-------------A----AAGCATGGAA.........CAG............  4663
CPZ.CM.04.MB66           G-----C--TC----G--------A------------------------...---------------------------------------A--C-AAACATGGAA.........TAG............  4654
CPZ.MT145                ---G-----T-----G-----T-----A---------------------...----------CA---A-----------------------AG---AAACCTGGAA.........TAG............  4650
CPZ.CD.-.ANT             G-----C--TC----G-----------T--A-A---G---A-----A--...--A-A--A-AGA---GA-T-----G--------A----------ATATCTGGAAATC.....................  4547
CPZ.CM.-.CAM3            ----A-A--C-----G--------A--A---------------------...-----T---------A--A-----G--G-----------AG-C-AGGCCTGGAATAG.....................  4522
CPZ.CM.98.CAM5           ------A--C--------G--T-----A--C--A---------------...-------------G-A--A-----G--------A--------C-AAACCTGGAATAG.....................  4810
CPZ.GAB_X52154           G-----A--------G-----------------A---------------...---------------------------------A----------AGACTTGGAATAG.....................  5169
CPZ.GAB2                 ------A--C-----------T--A--T----AAC----------G---...G-----------------C----TG--G-----------A--C--AAC..............................  4492
CPZ.TZ.-.TAN1            ------A--C--G-----------A-----A-AA--G------------...----------CA-G-G--A---ATGA-------A-----AC----AACATGGACAAG.....................  4752
CPZ.US.85.CPZUS          ------A--C--------------A--T--A-AA--C------------...----------C----A--A-----G--------------AG---AGGCTTGGAATAG.....................  5168
GSN.99CM166_AF468659     TAC---C--T-----G--------------C-A-CCA------G-GGCA...-A-A---AA----TGG---GT----CCCATACAA--A-T--C-GAGAGGAAGGTAGACTGGTTGCTTAG.........  4851
GSN.99CM71_AF468658      CAC---C--T-----G--------------C-A-CCA------G-GGCA...-A-----AA----TGG--AGT---GCCCATACAAG-A-T--CAGAGAGGAAGGTAGAATGGCTGATTAG.........  4833
H2A.ALI_AF082339         ------A--------G--G--C--------C--A--C------GG-AG-...CAA-AGCTG---AG-G--CCCCACTTGGAGGGT-CCAG---GG-T...GGAGAAGTGGCATAG.....CCTTGTCAAG  5496
H2A.BEN_M30502           ------A--------G--G--C-----T--C-----C------GG-AGA...CA--A-CTG---AG-A--CCCCACCTGGAGGGT-CCAG---GG-T...GGAGAAATGGCATGC.....CCTTGTCAAG  5503
H2A.ST_M31113            -A--A-A--------G-----T--------C-AA--C------GG-AG-...CAA-AGATG---AGCG--TCCAACTTGGAGGGT-CCAG---GG-T...GGAGAGGTGGCATAG.....CCTTATCAAG  4948
H2B.D205_X61240          ------A--C--G-----------A--T--A---C-C------GG-GGA...-AA-G-TTG---TG-A---CC-AC-TGGAGG-TACCAG-C-GGCT...AGAGAGATGGCACAG.....TCTGATTAAG  5476
H2B.EHO_U27200           ----A-A--C--------------------A---A-C------GG-GGA...-AA-A-TTG---TGCA---CC-ACGTGGAGG-TACCAT-C-GGCT...AGAGAGGTGGCACAG.....TCTAATTAAA  5474
H2G.ABT96_AF208027       ---T--A--C--------------A-----A-AA---------GG-GGA...-AA-AG-TG---AGCA--ACCAAC-TGGAGG-TACCAG-C-GACT...GGAGAAGTGGCATAG.....TCTGATAAAA  4830
H2U.FR.96.12034          G---A-A--C--G-----------------C--A---------GG-GGA...-AA-A-TTG---AG-A--TCCAACTTGGAGGGT-CC---A-GGTT...AGAGAGATGGCACTC.....CCTGATCAAA  4982

                                                                                                              Pol p31 Integrase end \
MAC.239_M33262           G-----A--C--------G--T--A--T--C-AA---------GG-GGA...-AA-AG-TG---AGCA--TCCCAC-TGGAGG-TACC-GA--GGCT...AGAGAGGTGGCATAG.....CCTCATAAAA  5420

Vif                                                          _M__E__E__E__.__K__R__W__I__A__V__P__T__W__R__I__P__E__R__L__.__E__R__W__H__S__._.__L__I__K_   Vif
Pol                      __V__V__P__R__R__K__A__K__I__I__K__D__Y__G__G__G__.__K__E__V__D__S__S__S__H__M__E__D__T__G__E__A__.__R__E__V__A__*           Integrase

SMM.SL.92.SL92B          ---G--A-----------------A-----C-AA---------GG-AG-...CAA-AGATG-G-AGCA---CCAGT-TGGAGG-TC-GT--A-GA-TAGTGGAGCGCTGGCATAG.....CTGCATTAAA  4854
SMM.US.x.PGM53           G-----A--------G-----T--A--T--C-AA---------GG-GGA...-AA-A-TTG---AG-G--CCCCACTTGGA-G-TACC-G---GGCTAGAGAAGTGGCATAG........CCTCATCAAA  5337
MNE.US.-.MNE027          G-----A--C--------G--T-----T--C-AA---------GG-GGA...-AA-AG-TG---AGCA--TCCCAC-TGGAGG-TACC-GA--GGCT...AGAGAGGTGGCATAG...CCTCATAAAATA  4890
RCM.NG.x.NG411           G-----C--T--------GTGT--------A-AA--------C-G-A-A...GA--TG-A-AGA--GACCAATATGGAAG-TAGAC-G-----AAGTTGA........GCAGTGGCATTCCTTAGTAAAG  4615
RCM.GAB1_AF382829        G-----C--C--G-----------A-----A-AA--A-----C-G-A-A...GAT-TG-A-AG---GGCCAATTTGGCGG-TAGAC-G-----AA-T...TGAACA........................  4584
DRL.FAO_AY159321         ----A-A--T----------------------AA------------A-T...G---ATA--C--ACCAACA---AG---T--G---TAC-T-G-AGAA................................  4664
MND-1.F17                GA----A--T--------GTGT--A--A--A-AA---------G-GAGT...GGA...AAGA--AG-CAG--TAACTTGGA-AGT-TCT--.......................................  4774
MND-2.5440_AY159322      G--G--T--T--------------A--A----A---C---------ACA...G---ATA-CA--CCCCACA---G---ATAG................................................  4642
MND-2.14CG_AF328295      ---T-----C-----------T--A--A--C---------------ACA...G---ATA--A--CCCCACA---AGGA---------GACA-C--GA.................................  5024
MND-2.MND2_AF367411      G-----A-----G-----------A-----C--A------------ACAGTGGATAGT-A-CCCA-CGTG-AA--TT-ATAG................................................  5085
DEB.CM.CM40              ---G--C--T-----G--------AG-A--C-AA--A---------G-C...----ATA--A-G-T-AACACACA--A-TCCA-T--AT-AAT-AGATAGGTGGTGCAGTGCTATAAAAAGAGAAAAATA  4938
DEB.CM.CM5               G-----A--------------C---G-G--A-AA------------GGA...G---ATA--A-GAC-GACACACA--A-CCCAGT--AT-AAT-AGATAAGTGGTGTAGTGCCATAAAAAGAGGAAAGTA  4932
MON.CM.L1                GAC---C--T--G-----G--------A--C-AACCC-----C-CCAC-...GAAAAT-TG-G--G-GACACCAACCAATATA-TCTCAGAA-GC-AGATGGATTGGCTAATTAGATGCACTAAATGGCA  4835
MON.NG1                  -AC---A---C-C--G--G------------C-CCC------CTCCA--...GAA-AT-TG-G-AG-AAGCCCAGTGCACAT---ATCAGA--GGT-GATGGAATGGCTGATTAGAGCAACCAAATATAA  3428
TAL.CM.8023              TAC---T--C-----G--G--T--A--A--C-AACCC------C--GCT...----GTAACA-GAC-GACT---A-G--TCCA----AC-----GCAGAGATGGGTAGGGACAATTAG............  4584
TAL.CM.266               CAC---C--T-----G--G-----A--A----A-CCC------C--GCT...T---GTAACA-GAT-GACC---A----TCCA----AC-----GCAGAGATGGGTAGGGACAATTAG............  5086
MUS.CM.1085              CAC---T--T--------G-----A--A--C-AACCC-----CGCCA--...GAA-AT-TG-G-AG-AAGTCC-AT-CAG--G-CCTCAGAA-GG-AGATAGATTGGATAATTAGGGCAACAAAATACTA  4822
SAB.SN.-.SAB1C           -AC-A-C--------------C--A--T--A-AA------------.........---TTG---AG-CAG-CCC-TCTGGA-GGTA-C-G-CGGACAGCAGGAGAGGTGGACTAG...............  5506
GRV.ET.x.GRI_677         G-----C--C--G-----G--T-----T--C-AA---------G-GAGA...-AAA--ATG---AG-GAG-GTAGT-TGGAGGGT-TC...AGAG--GCAAATAAGCAGATGGAGGGGGATAGTGACTTA  5299
TAN.UG.x.TAN1            ------A-----G-----G-----A--T--C-AA--A------G-GAGA...-AAA-T-TG-G---CAAGACTAAC-TGGAGGGTT-G..........................................  5286
VER.DE.-.AGM3            --A-TAC--------G--------A--T--A-A--------A-CC-AGA...-AAAG-ATG-G----GAGAGTAACTTGGA-GGT-CC-GA-GAGCTGATAACTAAATGGCAAGG...GATAGTGAGGTA  4848
VER.KE.-.9063            -A-T--C-----------------A-----C-AA-------A-CC-AGA...-AAAG-ATG-G----GAG-GTAAC-TGGA-GGTCTCAG--GACCTGATAACCAAATGGCAAGG...AATAGTCAAATA  5353
VER.KE.-.AGM155          G-----T--G-----G--------A-----A-AA--C----AGCC-AGA...-AAA--TTG-G----GAGACTCACCTGGA-GGT-CA-GA-GAAGTGATCACCAAATGGCAGGG...GATAGTTAGATA  5345
VER.KE.-.TYO1            G--T--A--------G-----T--A--T----A--------A-CCCA-A...CAAAG--TG-G-A--GAG-GT-ACGTGGA-GGTACCAG--GA-CTGATAACTAAATGGCAGGG...AATAGTCAGATA  4840
SUN.GA.98.L14            G-----A--T--G-----------AC-AG-A-A---C------GG-..................--A-CCTCCT--GTGGAGAT--TG-GATGA....................................  5438
LST.CD.88.447            CTC---C--C-----------T---C-AG-A-AA---------G-G......GGAC--AAG---AG-GAG-GT-GCCTACATA-CA-CT--.......................................  4364
LST.CD.88.485            CTC---C--C-----------T---C-AG-A-AA---------G-G......GGAC--AAG---AG-GAG-GTAGCCTGGATA-CA-CT--.......................................  4364
LST.CD.88.524            CTCTA-C--C--G--G--------AT-AG-G-AA---------G-G......GGAC--AAG---AG-CAGAGTAGC-TGGATA-CC-AT--.......................................  4361
LST.KE.-.lho7            CTC---C--C--------------AC-AG-A-AA---------G-G......GGAC--AAG---AG-GAGAGTAGCCTGGATA-CA-CT--.......................................  5448
SYK.KE.-.SYK173          T-C---A-----------G-----A--A----CA--CC-----G--.........AG-ATG---AG-G--TCCCACGTGGAGA-T--CCCCA-GACAGATTGATAG........................  5296
SYK.KE.KE51_AY523867     --C---A--T-----G--------A-----C--CAG-G-----G--.........AG--TG---AG-G--ACTCACCTGGA-ATTACATCCA-A-CAAATTGATAA........................  4948
COL.CM.-.CGU1            T--GAAA--T----------T--------CC--ACTCC-ATATGG-G-AGATG---GCA----GA---TGT-AT-C-A-G-G--AA-G--A-CCG-AACCGTTAAGGGAATGGATTAA............  4886
```

```
H1B.FR.83.HXB2            .............................................................................AACATGGAAAAGTTTAGTAAAACACCATATGTATGTTTCAGGGAAA...GCTA  5146
Vif                       _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _._ _T__W__K__S__L__V__K__H__H__M__Y__V__S__G__K__.__A__   Vif

H1D.CD.84.84ZR085         ................................................................................CACATGG-AAAGTT--GT-A-AT-CCATATGCA-ATTTCA--GAAA----  4668
H1F1.BE.93.VI850          ................................................................................CACATGG-AAAGTT--GT-A-AT--CATATGTA-GTTTCA--GAAA--C-  4486
H1G.SE.93.SE6165          ................................................................................-ACATGGCACAGTT-GGT-A-A---CATATGTA-GTCTCA--GAAA----  4543
H1H.CF.90.056             ................................................................................CACATGG-AAAGCT--GT-A-GT-CCATATGCA-ATTTCA-GGAAA----  4493
H1J.SE.93.SE7887          ................................................................................CACATGG-AAAGTT--GT-A-GT-CCATATGAA-GTTTCA--GAAA----  4460
H1K.CM.96.MP535           ................................................................................TACTTGG-AAAGTT--GT-A-AT-CCATATGCA-ATTTCA--GAAA----  4342
H101_AE.TH.90.CM240       ................................................................................-ACATGG-ACAGTT--GT-A-A---CATATGTA-ATCTCA--GAAA----  4720
H102_AG.NG.-.IBNG         ................................................................................-ACATGG-A-AGTT--GT-A-A---CATATGTA-GTCTCA--GAAA----  4671
H103_AB.RU.97.KAL153_2    .............................................................................---------C--------------------A---A-----AA----...----  4372
H104_cpx.CY.94.CY032      .............................................................................---------C---C----G-----T---------------AA----...----  4512
H1N.CM.95.YBF30           .........................................................................................C-----------T-----------G---AAA--G...--A-  4740
H1O.BE.87.ANT70           ...........................................................................................................---CAGG--TA-A--G...A-CG  5201
H1O.CM.91.MVP5180         ...........................................................................................................A--CA-G--TAA---G...--CG  5176
CPZ.CM.04.EK505           ............................................................................................TT-GGT-A-G---CACATGTA-GTGTCA--GAAG--AC  4718
CPZ.CM.04.LB7             ............................................................................................TT-GGT-A-AT--CATATCTA-GT-TCC-GGAAA--A-  4701
CPZ.CM.04.MB66            ............................................................................................CC--GT-A-A---CACATGTA-GTGTCT-GGAAA--C-  4692
CPZ.MT145                 ............................................................................................TT--GTGA-AT--CATATGTACAG-TCA--GAAA--A-  4688
CPZ.CD.-.ANT              ............................................................................................CT--GTGA-A--CCACATATGGGAGACT---GTACT--  4585
CPZ.CM.-.CAM3             ............................................................................................CC--GT-A-ATG-CATATGTA-AA-ACA--GAAA--A-  4560
CPZ.CM.98.CAM5            ............................................................................................CC--GT-A-AACCCATATGTA-AG-ACA--GAAG--A-  4848
CPZ.GAB_X52154            ............................................................................................TT--GTTA-AT--CATAT-TA-AGGTCA---AAG--A-  5207
CPZ.GAB2                  ...................................................................................ATGG-A-AGTT--GT-A-GTT-CATATGTACAAGTCA---AAG--C-  4539
CPZ.TZ.-.TAN1             ............................................................................................CC--GTTA-A---CATATCT--A--ACC---TGCTG--  4790
CPZ.US.85.CPZUS           ............................................................................................TC-GGT-A-AT--CACATGTA-AG-TCA--GAAA----  5206
GSN.99CM166_AF468659      .......................AGCTACTAAACACCACATATGGTCAGGCAAAACACCATTTGTATATGTCCATCACTACCAATTGCAACACCA-...AGATT--CAC-AAACAAGATTCG-TTA--A-  4955
GSN.99CM71_AF468658       .......................AGCAACTAAACATCATATATGGTCAGGAAAAACTCCATTTGTGTATATCCACCACTATCAACT-CAACATCA-...AGGTT--CAC-AAACAA-ATTCG-TTA--AT  4937
H2A.ALI_AF082339          TACCTAAAGTAC......................................................................................................AG-ACA---GATCTAG  5524
H2A.BEN_M30502            TACCTGAAATAC......................................................................................................AG-ACA---GACCTAG  5531
H2A.ST_M31113             TATCTAAAATAC......................................................................................................AG-ACAGG-GATCTAG  4976
H2B.D205_X61240           TATCTTAAGTAT......................................................................................................AG-ACAGG-GAGTTGC  5504
H2B.EHO_U27200            TACCTGAAATAT......................................................................................................AGGACA---GACTTGC  5502
H2G.ABT96_AF208027        TTCTTAAAGTTT......................................................................................................AA-ACA---GACCTAC  4858
H2U.FR.96.12034           CACCTGA.GTAT......................................................................................................AACACA---GAGCTAC  5009

MAC.239_M33262            TATCTGAAATAT......................................................................................................AA-ACT---GATCTAC  5448
Vif                       _Y__L__K__Y__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__K__T__K__D__L_    Vif

SMM.SL.92.SL92B           TTCCATAAGTAT......................................................................................................AA-ACA-G-GAGCTAG  4882
SMM.US.x.PGM53            TACTTGAAATTT......................................................................................................AG-ACT--GGACCTAC  5365
MNE.US.-.MNE027           TCTGAAATATAA........................................................................................................-ACT---GATCTAC  4916
RCM.NG.x.NG411            TATCATCAGTAT...................................................................................................AAGGG-AA---GGAA----  4646
RCM.GAB1_AF382829         ................................................................................----C-CTC---G------T-T---CA---CAAAGG-AA----GCA--A-  4634
DRL.FAO_AY159321          ..............................................................................-GG-----CTCCC-GTGC---T-T----A-C--AAGGG--AA--GTATTTAG  4716
MND-1.F17                 .............................................................................TCAGA-A-T-GAAAAGTGGC-CTGGTTAG-AAGAAGACAGAT-GC-TGG--C-  4827
MND-2.5440_AY159322       ...........................................................ACAGGACGTCAAGTTGAG-GAT-----TTCAA--------T-T----A----AAAGG--A----CATTTAG  4713
MND-2.14CG_AF328295       .............................................................................--A------TTCA-----C---T----C-AA---AAGGA--AA---CACTTAG  5077
MND-2.MND2_AF367411       .................................................................ACATGCAGTAGA--G------TTCAC-G------T----C-AA---AAAGG--AA---CATTTGG  5150
DEB.CM.CM40               CAAGACCAAAGA.................................................................................................................ATTAG  4955
DEB.CM.CM5                TAAGACCAAAGA.................................................................................................................ATTAC  4949
MON.CM.L1                 CATACTCACAGG.................................................................................................................AAG-G  4852
MON.NG1                   CATATATACAGG.................................................................................................................GAAAG  3445
TAL.CM.8023               .....................................................................................................GT--CACATGTACAA-ACC--GTACTTAC  4613
TAL.CM.266                .....................................................................................................GTT-CACATACA-AA-ACC--GTACTTAC  5115
MUS.CM.1085               CATCCTTTCAGG.................................................................................................................GAGAG  4839
SAB.SN.-.SAB1C            ............................................................................................CC--GTGA-AT-CCACATGCA-GTGTCC---CAGTG-G  5544
GRV.ET.x.GRI_677          CAAGATCAGGAATAA.................................................................................................................AC  5316
TAN.UG.x.TAN1             .......................................................................TGGAGAGCATATAG-C-AA-GGAAGGGCATAGTG-AA--CCACATGA----CCGGTTAC  5345
VER.DE.-.AGM3             CTGGATGAGGAC..............................................................................................................TAGAAAAT  4868
VER.KE.-.9063             TTGGATGACTAA..............................................................................................................-AGGAACC  5373
VER.KE.-.AGM155           TTGGATGAATAA..............................................................................................................-AGGAA-C  5365
VER.KE.-.TYO1             TTGGATGAGACA..............................................................................................................-AGAAA-T  4860
SUN.GA.98.L14             .................AAGAGAATGGGAAGATAGAAGACAGTATAAGATAGTTAGAATAGTATGGCTTATAGATAG--TTGCTGT-GAGAA-T-TTT-G-TTTA-GAAGAA-GCATA--G--ACAAAAG  5551
LST.CD.88.447             .............................................................................G-GCCTAGT-GAGAAGT-CCT-G-TATC-G-AGA-CACATAAAG-GACTCAA-  4417
LST.CD.88.485             .............................................................................G-ACCT--T-GAGAA-T-CCT-G-TAT--G-AGA-CACATAAAG--ACTCAGG  4417
LST.CD.88.524             .............................................................................G-GC-TAGT-GAGAA-T-TCT-G-TATC-AAAGGAA-CATTCAG--AGTAAA-  4414
LST.KE.-.lho7             .............................................................................-T-C-TA-T-GAAAAGT-TTT-G--ATAG-CAGGT-AAACACAG-GGCAAAAG  5501
SYK.KE.-.SYK173           ...................................................................................ATTGCAGCATA---T-A-AACACATA-GTA-AA-A-C---GAATTGG  5343
SYK.KE.KE51_AY523867      ...................................................................................A-TGCAGCATA--TT-A-GGTACACA-GTA-AA-A-C---CAATTAG  4995
COL.CM.-.CGU1             ..............................................................................................................GAAGCAGTTTCC-GAATGCC  4906
```

```
H1B.FR.83.HXB2            GGGGATGG...TTTTATAGACATCACTATGAAAGCCCTCATCCA...AGAATAAGTTCAGAAGTACACATCCCACTAGGG......GATGCTAGA...TTGGTAATAACAACATATTGGGGTCTGCATAC  5261
Vif                       R__G__W__.__F__Y__R__H__H__Y__E__S__P__H__P__.__R__I__S__S__E__V__H__I__P__L__G__.__.__D__A__R__.__L__V__I__T__T__Y__W__G__L__H__T   Vif

H1D.CD.84.84ZR085         AA------TTT-A-AGACAC--CT-TG-CAGCCCT-AC-CAAA-......------------------------T----A......---------CTG...---G-------------------------  4783
H1F1.BE.93.VI850          AA------TCT-A-AGACAT--CTTTC-AAGC--G-A--CAAG-......G---------------------------A-......--A-T--A-TTA...--------------------G-----CC-  4601
H1G.SE.93.SE6165          -A--C---TTT-A-AGACCT--CT-TGCAAGC--G-A--CAAG-......G----------------------------A......-------C-CTA...---G----------------G--------  4658
H1H.CF.90.056             -A------TTT-A-AGACAT---TTTG-AAGC-CT-A--CAAGG......------------------------T----A......--A-----GTTA...--C-----C-----C---------A----  4608
H1J.SE.93.SE7887          -AAA----CTG-A-AGGCAT--CT-TG-CAGC-AT-A--CAAA-......-----------------------------A......--G----T-TTA...--------------------------G--  4575
H1K.CM.96.MP535           ACC-----TAT-A-AGACAT--CT-TG-AAGC--G-A--CAAA-......------------------------------......-----AGAGTTA...---G---------------------T---  4457
H101_AE.TH.90.CM240       AAAAG---TTT-A-AGACAT---T-TG-AAGCCAG-A--CAAAG......G----C-----------T-----------A......--G------TTA...-------G------------------A--  4835
H102_AG.NG.-.IBNG         A---T---TTT-A-AGACAT--CTTTG-AAGT--G-A--CAAA-......G--T--------------------------......--------GTTA...---G---G---------------------  4786
H103_AB.RU.97.KAL153_2    --------...G------A----------------AGGA-----...---------------------------T-----......-------A-...----------A---------------------  4487
H104_cpx.CY.94.CY032      AA-----A...--C--------------------TAGG--C---...-A-G----------------T------------......--G------...--A---G---G------------------GC-  4627
H1N.CM.95.YBF30           AA------...-A------------T-------CA-A---C---...-A------------------T------G----T......C-G--A---...--A--G-C-GTC--T--------G--AACA--  4855
H1O.BE.87.ANT70           A-AAC---...-GG--C--------T------TC-AGAA-----...---G-T------AGT---T-T--T---G-----......-TG--CCAT...G-A---G----C-----------AT--ATGC-  5316
H1O.CM.91.MVP5180         C-AAC---...CG------G-----T------TC-AGGA-----...-A-G-C-----G-CG--GT-T--T---G---CA......--A---GAT...A-A--GG-C--C-----------AT-AATGC-  5291
CPZ.CM.04.EK505           -A--T---TTC-A-AGACAT---T-TG-AAC-GATAA--CAAA-......------------A-----------T----A......--A--A-A-TTA...------GTT--C--------AT-AATGC-  4833
CPZ.CM.04.LB7             -A--C---TTT-A-AGACAT---T-TG-GAGTGAAAA--CAAA-......G-----------A----T------------......--G--C-A-TTA...------GT---T--------A--TACAC-  4816
CPZ.CM.04.MB66            -AAAT---TTC-A-AGACAC--CT-TG-AAGC-ATAA--CAAA-......--------------------------G--A......--A----A-TTA...--------------------A--CACA--  4807
CPZ.MT145                 AAAAT---TTC-A-AGACAT--CT-TG---GC-C-AAC-C-AGG......--GGC----C-GA--T-T--T---A--AAT......C-G----AGCTA...A-TG-C--------------A--AATGC-  4803
CPZ.CD.-.ANT              AACC----AAG-A-AGACAT---T-TG-AA-TGAT-AC-CAAAG......-A-G-AGA---------T--A------CCAACATTG---AAA-A-TTG...---G-G---GTG-TC-----AT-A--ATG  4706
CPZ.CM.-.CAM3             A--AC---TTT-A-AGGCAC---T-TG-AAGTTCT-A--C-AA-......G-----------A----------TA---AA......A-G--A-A-TTA...--GG-------C--------AT-AATGC-  4675
CPZ.CM.98.CAM5            AAAAG---TTT-A-AGGCAT---T-TG-GAGTGATAA--C-AA-......G-G--------------T-----TT---AT......ACA--A-A-TTA...--GG-C--------------A---ACA--  4963
CPZ.GAB_X52154            -A--T---TTT-ACAGACAT---T-TG--C-CCCTAA--CAAA-......G-TGC---T---A-------A--TT-CA-AGAT...T--T---A-TTA...A--G-G-----C-------CCT-AAGCC-  5325
CPZ.GAB2                  -A------TTT-A-AGGCAC---T-TG--AGTCC-AA--C-AA-......G-------------C-----A--CA--A-A......-----A-A-ATA...--GG-CTGT-----C---AA-T-AATGC-  4654
CPZ.TZ.-.TAN1             AA-AT---AAG-A-AGACAT---T-TG-AACTGATA-A-CAAA-......-G-GCAGGG---A-------A--T---ACA...GAAAGAT-A-A-TTA...--GG-TTT-CAT-----------AGCCTG  4908
CPZ.US.85.CPZUS           AA------TTC-ACAGGCAT---T-TG-GAGC-CTAA--CCAG-......G---A------------T-----TA---AT......A-A--A-A-ATA...--G--T-T---G-------GT--A-A---  5321
GSN.99CM166_AF468659      T--ATACTAGGAAGGTAG-GG-AG-AGTA---GCAA-AT--AT-...GAG----C-ATCCT-TGGG---CAA-CAGT.........C---GACC-GCTAGCT--TCCCGGT-CACC-ATTAC-AA--AG-  5073
GSN.99CM71_AF468658       TA-ATACTAGGCAGGTAG--G-GG-AGTA---GC-A-AT--ATT...GAG----CCATCCT-TGGG---CAA-TAGT.........C---GACC-GCCAGTC--TCCCG-T-CACC-ACTAC-AA--GG-  5055
H2A.ALI_AF082339          AAAAGGT-TGC-A-GT-CCC--C--TA-G-TGG-ATGGGCATGG...T-G-CTT-CAGCAGG---ATAT-----T--CAA......-GAAGA--TCAT...C--GAG-T-CAGGCA-ATT-GAAC-TA--  5642
H2A.BEN_M30502            A--AGGT-CGC-A-GT-CCC--C---A-G-TGG-ATGGGCATGG...T-G-CTT-CAGCAGG---ATAT--------CAA......-GAAAA--TCAT...C--GA--T-CAGGCA-ATT-GAAC-TA--  5649
H2A.ST_M31113             A-AAGGT-TGC-A-GT-CCC--C--TA-G-TGG-ATGGGCGTGG...T-G-CTT-CAGCAGG---ATAT-----T--AAA......-GA-AA--TCAT...C-GGAG-T-CAGGCA-ACT-GAAC-TA--  5094
H2B.D205_X61240           AACAGGTCTCT-A-GTCCCT--C---A-G-T-G-ATGGGC-TGG...T-G-CTT-CAGTAG-A--ATAT-T--C---AAC......A-A-GAGC-TGG...C--GA-GTCCA-GGA-ATT-GAAC-TA--  5622
H2B.EHO_U27200            AACAGGTCTCT-A-GT-CCT--C--TA-G-T-G-ATGGGC-TGG...T-G-CTT-CAGTAG----ATAT-T--C--GAAA......--A-GAGC-CAT...C--GA-GTCCA-GGA-ACT-GAAC-TG--  5620
H2G.ABT96_AF208027        A-AAGGC-GTT-A-GT-CCT------A-A-TTG-ATGGGC-TGG...T-G-CTT-CAGCAGG---ATAT-----T--ACA......A-A-AAGC-CAT...T-GGA--T-CA-GGG-ATT-GAACTTG--  4976
H2U.FR.96.12034           A-AA-GT-TGT-A-GTGCCC------A-G-T-G-ATGGGC-TGG...T-G-CTT-CAGCAGG---ATAT-T--T---AAA......A-A-GATC-CAT...T-GGAGGTGCA-GCC-ATT-GAATTTG--  5127

MAC.239_M33262            AAAAGGTTTGC-A-GTGCCC---TTTA-G-TCG-ATGGGCATGG...T-G-CCT-CAGCAG----AT-T--------CA-......--A-GA--CCAT...T--GA-GT-CA-GGG-ATT-G-ATTTG--  5566
Vif                       Q__K__V__C__Y__V__P__H__F__K__V__G__W__A__W__.__W__T__C__S__R__V__I__F__P__L__Q__.__.__E__G__S__H__.__L__E__V__Q__G__Y__W__H__L__T   Vif

SMM.SL.92.SL92B           A-AA-GCCTGC-A-GTGCC-------A-A-T-G-ATGGGCATGG...TAC-C-GCC--TAGG--GATTT----CT-G-A-......--G-GA--TCAC...T-GGAGGT-CA-GTG-ACT-GAAT-TG--  5000
SMM.US.x.PGM53            A-AAGGCTTGT-A-GTACCC-----TA-A-TTG-ATGGGCATGG...T-G-CTT-CAGCAG----AT-T--------A-A......----AATCTCAT...T-GGA-GT-CA-GGA-ATT-GAATTTGG-  5483
MNE.US.-.MNE027           AAAAGGTTTGC-A-GTGCCC-----TA-G-TCG-ATGGGCATGG...T-G-CCT-CAGCAG----AT-T--------CAA......--A-AA--CCAG...T--GA-GT-CA-GGG-ATT-GAATTTA--  5034
RCM.NG.x.NG411            AA-AT---GAG-A-GTACCC---TTTA-A-T-CCTTGGGGGTGG...T-GTCTC----------C--------CT-G--A......A--AAC-CTAAG...A---AGGTT---ACA-ATT-GAGT-TC--  4764
RCM.GAB1_AF382829         AA-AG---GAG-A-GTGCCT--CTTTA-A-T-CCATGGGGATGG...T-GTCGCA---------T-----A--TT---A-......--A-GATC-AAG...T---A--T---CACC-ATT-GAATTTG--  4752
DRL.FAO_AY159321          AAA-----CAT-A-GCACCT--CTT-C-ATGCTCA...GG-TGG...T-G-CTCA-AGTC-GTGGAC------CT-CAAA...GACA-AAG--A-......A--G--GT----GTCCT-T-GAAC-TC--  4831
MND-1.F17                 CT-C-AATAATGAGG-AG--TG-TGG-GGCTGTAT------TTT...-TGGCTTA-AAT---TGGT-T-CTTGCAGTAAA.....................---G-G-TT-T-ATAAATA-GGACATA-G  4933
MND-2.5440_AY159322       AAAAT---CAA--A--CCCC---TTTC-ATGCTCA...GGATGG...T-G-C-CAC--TC--AAGA-G-----CT-TAA-......----G-TC-AAG...A-T--C-TC--CGTC-T-T-GAACTTG--  4828
MND-2.14CG_AF328295       ATAAG---GAAC-A-TCCAC--CTTTC-ATGCTCA...GGATGG...T-G-CGCAC--TC--AAGATT-----CT-CAAA......----GCTC-AAG...A-T--T-TT--TGCCCTAT-GAACTTG--  5192
MND-2.MND2_AF367411       ACAA----GAT--AA-GGC----TTTC-ATGCTCA...GGATGG...T-G-C-CAC---C--AA-T-T--T--C----AA......--A-A-GA-CAG...A-C----TC--CATCCT-T-GAACTTA--  5265
DEB.CM.CM40               ATACTGTACAA-GGATACAC---TTTG-ATT---GAACTCATAT...TAT-C-CAAA--A--A--GTAT----CA-CACA......A-A-AAG-GGGA...---G--GT-GA--TACTAT-GTGCTTG--  5073
DEB.CM.CM5                AA-C-ATAAAG-GGATACAC---T-TG-ATT---AAGGGCATTC...TAC-CTCAGA--A--A--GT-T----CG-CACA......--A-AAG--AAA...---G--GTGGA-ATA-TCT-GAGC-TGG-  5067
MON.CM.L1                 CACCC-TTGTC-A-GTACAT--CT--C-GCTGCA-AA---GAG-...TTCTC-CAGAACA--A-TA-AC----CT---ACCTAGGAATCA--CA-GAAGGA-AGTC-TGGG--ACA-ATTTGGA-...-T  4976
MON.NG1                   T-CC--TTGTG-A-GTACAT--CT-TC-GCTT-CT-AC--G-GC...TTTTCGCAA-GCA-GA-TA-GC-T--TT---A-ATGGGTATCA-ACA-RAGGR--A-CA-TGGG--ACA-ACATRGA-ATC--  3572
TAL.CM.8023               AAAAG-TAATA-GGGTCCAT---TT-C-ATT---AATGGCCT-TTTC-CGCAGGCCAA-TGGA-C-T-CC--T-GA--AAAAAGAGA--AGAGT-GAG-ACA-T-CC-TC--TC-G-ATT-GTGCTTAG-  4743
TAL.CM.266                AAAAG-T-ATT-GGGT-CAT--CTTTC-ATTG--A-ACGCCT-CTAT-C-CA-GCAAAGTGGA---T-CCGTT-GA--AAAAGGACACCAGAGT-GAA-ATT---CC-T----C-G-ATT-GTGTTTAG-  5245
MUS.CM.1085               CA-CC-TTACA-A-GTGCAT--CT-TC-ATTGTA--A---GAGG...TTCTCCCAGAGCA--A-CAGA--A---T-GACACAAAATACCCGGGA-TCAGGT-C-G--GA--GTAC--ACATAGAA-TG--  4966
SAB.SN.-.SAB1C            TTCAT---CGG-A-ACCCCT---ACTA-AAT---ATGGA--TGG...TACTCCTA-CA----TGGGT---A--C--GAA-......----GAGCTCTAA-AAA-G-T--T-AT------CA-...TTA--  5662
GRV.ET.x.GRI_677          AATTGCCTTGGGAA--C--------T-GGC-GGTG-AATGG-AGTTTT-G-CCTACAGCC-GT-CATT-----CT--TCA......A-A-A-GATTAC...A--GA-GTG-ATAT--ATCACAAC-TC--  5437
TAN.UG.x.TAN1             A--AT---ACA-A-CT--TG--CT--C-ATGTG-ATGGGC-TGG...TAC-C-T--AGTAG-T-C-T------CT----A......-GA-AAG--AAAA-A---G--GATTGT--C---CAC...-TA--  5463
VER.DE.-.AGM3             TA-AC---AAA-A-CGA-TG--CT--C-AATT-CATGGGCATGG...TAC-C--TGAGTAG-TATG-G--A--C------......C-ACA-G--AGTA-ACATG--GATCT-------CA-...-TG--  4986
VER.KE.-.9063             TTAA----CAA-ACATG-TG--CT--C-AATT-ATTGGGC-TGG...TAT-CT-TG-GCAG-TACAT-------A----A......TTA-AAG--GAAA-ACAG---GACTGC------CA-...TTA--  5491
VER.KE.-.AGM155           T-AA----GAA-ACA-A-TG---T-TC-AATC-CTTGGGCATGG...TAC-CT-TGAGCAG-TATGTA--A--C--CCCA......-GAAG-G--GAAA-CCATG-GGAT-TC------CA-...TTAG-  5483
VER.KE.-.TYO1             T-AA----AAC-A-----TG---T--C-AATT-CATGGGC-TGG...TAC-CC-TGAGTAG-TATGTA--A---A----A......A-ACA-G-GGAAA-ATGTG--GACCT-------CA-...TTA--  4978
SUN.GA.98.L14             AT-AT---GTA-C-ATGTATGGAACAGG-ACTG-ATGGG-GTGG...TAC-C-TACAATA--A-TATT--T---G--ACT......T---GG-C-GTA...---G---GG...ATC-AT--A-AC-TC--  5666
LST.CD.88.447             AT-AT---GTAAGC-TACATGGAACAGGG-T-G--TGGG-ATTT...TAT-CTTA-AATA-TA--GTG--A--TT---A-......ACA-GA-CCTTA...---G-G-G-...-GC-AT-CA-ACTTGG-  4532
LST.CD.88.485             TT-AT---GTAAGC-TACATGGAACAGGG-T-G--TGGG-ATTT...TAT-CTTA-AATA--A--GTA--A--TT---A-......ATA-GA-CCTTA...---G---G-...-GC-AT-CA-ACTTGG-  4532
LST.CD.88.524             AT-AT---GTTAG--TACATGG-ACAGGA-T-G-ATGGG-GTTT...TAT-CTTACAATA--T---TA--T---T---AT......ACA-GA-CTTTA...---G---G-...AT--AT-C--AT-TGG-  4529
LST.KE.-.lho7             AAAAT---GTAGGC-TACATGGAACAGG--T-G-TTGGG-ATTC...TAT-CCTA-CATA--T--GTG--A--T----A-......TCA-GA-CTTTA...---G-C-G-...ATA-AT-CA-ATTTGG-  5616
SYK.KE.-.SYK173           A-AA-GCCACA-A-A-ACAT--CT--C-AAT-GAATGG--ATGG...TAC-CTTA--GCC--TGGACA--A---G----A......----GA-CTATA......TGG-TC--G-TC-ATCACAATTTGG-  5458
SYK.KE.KE51_AY523867      A-AA-GCCACA-A-A-ACAT---T-TC-GATT-CATGGG-GTGG...T-G-CTC-AG--C--TGGG-A--A---G----A......C-A-G-GT-TTA......--T-T--A--TC-ATCA-AATTTG--  5110
COL.CM.-.CGU1             AATAT-ATTAT--AGGA-A-ACAG-G-CACCCGAATGG...TGGACAGA-GGT-AAAT---GT-TT--T-AGGG...TCAGCTATTA-GTTA......AGCA-----TACTTGGGCACT......TTA--  5018
```

```
H1B.FR.83.HXB2            AGGAGAAAGAGACTGGCAT............................................................................................................... 5280
Vif                       __G__E__R__D__W__H__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._ Vif

H1D.CD.84.84ZR085         ------------A------CTGGGTCAGGGA................................................................................................... 4814
H1F1.BE.93.VI850          ------------A------CTGGGTCAGGGA................................................................................................... 4632
H1G.SE.93.SE6165          --------A---------ATTGGGTCATGGG................................................................................................... 4689
H1H.CF.90.056             ------------A------TTAGGCCAGGGA................................................................................................... 4639
H1J.SE.93.SE7887          ------------T-----CTTGGGTCAGGGA................................................................................................... 4606
H1K.CM.96.MP535           ---------G---------TTGGGTCAGGGA................................................................................................... 4488
H101_AE.TH.90.CM240       --------AG--------CTTGGGTCATGGA................................................................................................... 4866
H102_AG.NG.-.IBNG         -------C----------CTTGGGTCATGGG................................................................................................... 4817
H103_AB.RU.97.KAL153_2    -------------------............................................................................................................... 4506
H104_cpx.CY.94.CY032      ---G---CA---------C............................................................................................................... 4646
H1N.CM.95.YBF30           -------CAGTCT------............................................................................................................... 4874
H1O.BE.87.ANT70           ---G--------TGAA---............................................................................................................... 5335
H1O.CM.91.MVP5180         ---G--------GGAA--C............................................................................................................... 5310
CPZ.CM.04.EK505           ---------GCCA------TTAGGACATGGA................................................................................................... 4864
CPZ.CM.04.LB7             -------C-G--------CTTAGGTCATGGG................................................................................................... 4847
CPZ.CM.04.MB66            -----------C-------CTAGGCCATGGG................................................................................................... 4838
CPZ.MT145                 ------------T------CTGGGACATGGA................................................................................................... 4834
CPZ.CD.-.ANT              T------C-CCCA-----CTTAGGACATGGT................................................................................................... 4737
CPZ.CM.-.CAM3             ----C-GCAT-C-------CTGGGGCATGGA................................................................................................... 4706
CPZ.CM.98.CAM5            ---GC--CAT-CT-----CCTGGGACATGGA................................................................................................... 4994
CPZ.GAB_X52154            ---------G-C-------TTAGGCCATGGA................................................................................................... 5356
CPZ.GAB2                  -------CA-CCA------CTAGGCCATGGA................................................................................................... 4685
CPZ.TZ.-.TAN1             T---------CCA------CTAGGTCATGGC................................................................................................... 4939
CPZ.US.85.CPZUS           -------CA--C------CTTAGGGCATGGA................................................................................................... 5352
GSN.99CM166_AF468659      --TCAT--T---A---GTATACAACAGAAGCAAA................................................................................................ 5107
GSN.99CM71_AF468658       --TCAC--T---G---GT-TACAACAGAAGTAAG................................................................................................ 5089
H2A.ALI_AF082339          -CC-----A--GA----TCTCCTCTTATGCA................................................................................................... 5673
H2A.BEN_M30502            -CC-----A--GA----TCTCCTCTCATGCA................................................................................................... 5680
H2A.ST_M31113             -CC-----A--GA----TCTCCTCCTATTCA................................................................................................... 5125
H2B.D205_X61240           CCC------G-GA-TCTTGAGCTCCTATGCT................................................................................................... 5653
H2B.EHO_U27200            CCC------G-GA-TCTTGAGTTCCTATGCT................................................................................................... 5651
H2G.ABT96_AF208027        CCCT----AG-GA---TTAAGCCAGTATGCA................................................................................................... 5007
H2U.FR.96.12034           -CC---G-AG-GA----TATCCTCTTATGCA................................................................................................... 5158

MAC.239_M33262            -CC-----A--GG----TCAGTACTTATGCA................................................................................................... 5597
Vif                       __P__E__K__G__W__L__S__T__Y__A__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._ Vif

SMM.SL.92.SL92B           CCC-----A--GA---TTATCAAGCTATGCA................................................................................................... 5031
SMM.US.x.PGM53            -CC-----A--GA----TCAGTACTTATGCA................................................................................................... 5514
MNE.US.-.MNE027           -CC--------GG----TCAGTACTTATGCA................................................................................................... 5065
RCM.NG.x.NG411            -AT-----AG-GA---TTAGGAACCTATGGA................................................................................................... 4795
RCM.GAB1_AF382829         --T-----AG-GA----TAGGGACATATGGG................................................................................................... 4783
DRL.FAO_AY159321          CCC---C-A--GG---TTAAGTAAATATGCCATA................................................................................................ 4865
MND-1.F17                 -TT-AT-GTTAGAA-CT--TGGCATTTGCAA................................................................................................... 4964
MND-2.5440_AY159322       CCC-----A--GG----TGTCTCAATACGCC................................................................................................... 4859
MND-2.14CG_AF328295       CCC-----A--GA---TTATCTCAATATGCA................................................................................................... 5223
MND-2.MND2_AF367411       CCC-----A--GA---TTATCAACTTATGCC................................................................................................... 5296
DEB.CM.CM40               CCC---G-A--GA----TA............................................................................................................... 5092
DEB.CM.CM5                TCC-----A--GG---ATG............................................................................................................... 5086
MON.CM.L1                 -AC-ATTTATTGGGACTTAACAAACGTGGGACCTAGTGCACTCTCACCAACAGTGTACAACAAACAGGCCTACACAATAGAATGGGTTTATTGGATGCAA...AAGGTGGAACGG......GTCCCTTTA 5097
MON.NG1                   GATCCTTT-G---AYTACCAATGTAGGGCCCGCGTCACTATCTCCATCAGTCAAATGGAAGCAAGCATACATCATAGAATGGGTATATTGG...AGGCAGCCCAGGGTGCCACAGCAAGAGGGAGTAGTA 3699
TAL.CM.8023               -CC---C-AG-GA-ACACATCAGACTGGGAT................................................................................................... 4774
TAL.CM.266                -CC---CCAT-GA-ATACGTCCGATTGGGAC................................................................................................... 5276
MUS.CM.1085               --TCCTGT-G--TGT-ACCAATGTGGGACCAGCAAGTTGCTCTAAAAGCACATGGTGGAAGCAATCCTTTATCCTAGAGTGGGTATATCTA....................................... 5057
SAB.SN.-.SAB1C            CCC---G-A--GA---TTAGAAACCTATGCA................................................................................................... 5693
GRV.ET.x.GRI_677          CCC--------GA----TCTCAAGTCATGGA................................................................................................... 5468
TAN.UG.x.TAN1             GCC----CAG-GA----TCTCTACCTATGCA................................................................................................... 5494
VER.DE.-.AGM3             -CC-----AG-GA----TATCAACATATGCT................................................................................................... 5017
VER.KE.-.9063             -CC--------G-----TCTCTACTTACGCC................................................................................................... 5522
VER.KE.-.AGM155           TCC-A--CA--GA----TCTCAACTTATGCA................................................................................................... 5514
VER.KE.-.TYO1             -CC---GCA--GA----TATCCACATATGCA................................................................................................... 5009
SUN.GA.98.L14             -CC--C--A--GG-T-AT-AACCAGTGGGGA................................................................................................... 5697
LST.CD.88.447             --C--G-----GA---ATCTCTCAGTGGGCA................................................................................................... 4563
LST.CD.88.485             --C--G-----GA---ATCTCTCAGTGGGCA................................................................................................... 4563
LST.CD.88.524             --C--GG----GA---AT-TCACAGTGGGCA................................................................................................... 4560
LST.KE.-.lho7             --C--G---G-GA---ATCTCACAATGGGCA................................................................................................... 5647
SYK.KE.-.SYK173           CCC--------GA---TTACATATGCAGGGA................................................................................................... 5489
SYK.KE.KE51_AY523867      -CC---G-A--G-CT-TTAAGAACAGAGGGA................................................................................................... 5141
COL.CM.-.CGU1             CCC---T---T-T...---............................................................................................................... 5034
```

```
                                                                                                     | Thr96 phosphorylated
H1B.FR.83.HXB2            ..........................TTGGGTCAGGGAGTCTCCATAGAATGGAGGAAAAAGAGATATAGC...ACACAAGTAGACCCTGAACTAGCAGACCAACTAATTCATCTGTATTACTTTGACT  5380
Vif                       _.__.__.__.__.__.__.__.__.__.__L__G__Q__G__V__S__I__E__W__R__K__K__R__Y__S__.__T__Q__V__D__P__E__L__A__D__Q__L__I__H__L__Y__Y__F__D__   Vif

H1D.CD.84.84ZR085         .............................-TCTCC...A-AGAATGGAGGAAA----G-T-T--C...-CACAAGT-G-CCCT-G-......--G------------------A-------T-----T-  4902
H1F1.BE.93.VI850          .............................-TCTCC...A-AGAATGGAGGCA-G----GT-T--G...-CACAA-T-G-CCCT-G-......--G--G--------G--C---A-A-----T-----T-  4720
H1G.SE.93.SE6165          .............................-TCTCC...A-AGAGTGGAGGCA---A-G-T-T---...-CACAAGT-G-GCCT---......-----------T--G---------C----T-----T-  4777
H1H.CF.90.056             .............................-TCTCC...A-AGAATGGAG-CT--AA-GGT-T--C...-CACAAGT-G-GCCT-G-......--G------------------A--C----T-----T-  4727
H1J.SE.93.SE7887          .............................-TCTCC...A-AGAATGGAGGCA---A-GGT-C---...-CACAAGT-G-CCCT-G-......--G------------------A---G---------T-  4694
H1K.CM.96.MP535           .............................-TCTCC...A-AGAATGGAGGCT--AA-GGT-C---...-CACAAGT-G--CCT---......---------------------A-A-----T-----T-  4576
H101_AE.TH.90.CM240       .............................-TCTCC...A-AGAATGGAGGCA---A---T-T--C...-CACAA-T-G-TCCT---......---------A----G--------AC-A--T----G--  4954
H102_AG.NG.-.IBNG         .............................-TCTCC...A-AGAATGGA-GCA--AA-G-T-T--C...-CACAA-T-G-CCCT---......T-------------G-----C--------T---A-T-  4905
H103_AB.RU.97.KAL153_2    ..........................-----C-------C---------------A---G-----------...----------------A-C----------------------------T-----T-  4606
H104_cpx.CY.94.CY032      ..........................--------T--------------------CTC-GA----------...---------G--T-----C--G------------------A--C----------T-  4746
H1N.CM.95.YBF30           ..........................C-A--A--T-----A-------------ACT--GA-A---C-AG...--------T--T------A-G------A-G-----A-----TC----T-----T-  4974
H1O.BE.87.ANT70           ..........................-----A--T--G--TAGT---------C-TT-T----AG----AA...-----GA-T---------AC-------AGGA-G--A------C----T---AC--  5435
H1O.CM.91.MVP5180         ..........................-----A--T--G--TAGT---------CAAT-C---GAG----AA...-----GA-T--T------AC-------AGGA-G--A------C----T--CACA-  5410
CPZ.CM.04.EK505           .............................-TATCC...A-AGAATGGAG-CAAG-A-T-T-C--G...-CACAA-T-G-TCCT--A......T-G------A-G-----A--C--T-----T-----T-  4952
CPZ.CM.04.LB7             .............................-TATCT...A--GAGTGGA--CA--A----T-T--T...-CACAA-T-G-CCCT--G......T-G-----T--TA-G------T-A-----T---A-T-  4935
CPZ.CM.04.MB66            .............................-TAAGT...A-AGAATGGAG-CA---A-GTT-T--G...-CACAA-T-G-TCCT--T......--G---------T----C---T----C--T-----T-  4926
CPZ.MT145                 .............................-TAGCA...A-AGAATGGAG-GAA--AGGTT-C--T...-C-CAGGT-G-CCCT--A......AC------TAGG-----A--C--TC----T-----T-  4922
CPZ.CD.-.ANT              .............................-TGAGC...A-AGAATGGAG---TG-A--GT-T-TC...-CACAAGT-G-CCC---A......AC---T--T---A-G--A--C-AA-----T--CCC--  4825
CPZ.CM.-.CAM3             .............................-TGTCC...A-AGAATGGAG---TG--TC-T-T--T...-CACAAGT-G-CCCT--G......A-G-----TA----------C--TC-A--T-----T-  4794
CPZ.CM.98.CAM5            .............................-TGTCC...A-AGAATGGAG---TG--TC-T-C-A-...-CACAAGT-G-TCCT--A......ACG-----TAG---------C--TC----T-----T-  5082
CPZ.GAB_X52154            .............................-TCTCC...A-TCAATGGAGGCTAG-ATCTT-TGTC...-CACAGGTGG-TCCCTT-......AC------TAG-T-G--C--CTCTC-A--T-----T-  5444
CPZ.GAB2                  .............................-TGTCC...A-AGAGTGGAGGCA---A-C-T-T---...-CTCAG-T-G-TCCTC-A......AC------T-GG--G--A--CTGT-----T-----T-  4773
CPZ.TZ.-.TAN1             .............................ATAGGA...T-AGAATGGAG-CAAG-A---T-C--T...-CACAA-T-G-CCCT--A......AC----------T-G-----CACTAGG--T---AC--  5027
CPZ.US.85.CPZUS           .............................-TATCA...A-AGAATGGAG-CAAG-A-CTT-C-AG...-CACAA-T-G-TCCT--G......A-G------A-G-----------------T-----T-  5440
GSN.99CM166_AF468659      .............................CTAGGA-C-......-G-...GA-G-AG-C-TA-TT-GGTA-AGC--TA-TC-GAC---A-G-G-----ATG---A-T--A---GGAA-A------CCA-  5198
GSN.99CM71_AF468658       .............................CTAGGA-C-......-GG...GAA-CAG-C-TA-C--GGTA-AGT--TA-TT-GAC---A--CG-----ATG---A-C--A--CACAAGG------CCA-  5180
H2A.ALI_AF082339          .........................................GTA-G-AT-ACTT--T-T-CAGA-A-GTT-TGG---G-T--TAC---A--CTGT------TCC-----A---GGCAC---T--CTCT-  5761
H2A.BEN_M30502            .........................................GTA-G-TT-ACCT--T-T-CAGA-A-GTT-TGG---G-T--TAC---A--CTGT------ATC-----A---AGCAC---T--CTCT-  5768
H2A.ST_M31113             .........................................GTA-G-CT-ACTT--T-T-CAGA-A-ATT-TGG---G-T--TAC---A--CTGT--G---TCC-----A---AGCAC---T--CTCT-  5213
H2B.D205_X61240           .........................................GTA-G-CT-ACAT--T-TG----GA-CTTTTAT---G-T---ACA-----TG-G--------G---C-G---GG--C---T--CTCT-  5741
H2B.EHO_U27200            .........................................GTA-G-CT-ACAT--T-TG----GAGCTTTTAT---G-T---ACT-----TG---------G-T--C-G---GG--C---T--CTCT-  5739
H2G.ABT96_AF208027        .........................................GTA-G-TT-ACCT--T-C-CA---A-ATTTTAT---G-T--GACA--A--GAC------T---T--T-G---GGA-C---T-----T-  5095
H2U.FR.96.12034           .........................................GTA-GGTT-ACTT--T-C-CA---GG-TTTTGG---G-T---ACA-----TTGC---------T--C-G---GG-AG---T--CTCT-  5246

MAC.239_M33262            .........................................GTG-GGAT-ACCT--T-CTCA-AGA-CTTTTGG---G-T---ACA--AA-CTAT------ATTT--C-G---AGCAC---T--CCCT-  5685
Vif                       _.__.__.__.__.__.__.__.__.__.__.__.__.__.__V__R__I__T__W__Y__S__K__N__F__W__T__D__V__T__P__N__Y__A__D__I__L__L__H__S__T__Y__F__P_   Vif

SMM.SL.92.SL92B           .........................................GTT-GGATTACTT--T-CTCAGA-A-ATTTTGG---G-T---AC---A--TG-------T--GT--GCA---AGTAC---T--CCCT-  5119
SMM.US.x.PGM53            .........................................GTA-G-AT-ACCT--T-CTCC---A-ATT-TGG---G-T---ACA--A--TTAT------ACTT--C-G---GGCAC---T--CCCT-  5602
MNE.US.-.MNE027           .........................................GTG-G-AT-ACCT--T-CTCA--GA-CTTTTGG---G-T---ACA--A--CTAT------ATTT--C-G---AGCAC---T--CCCT-  5153
RCM.NG.x.NG411            .............................-TAGGAAT--CA-A------TAAC--ATGTG-TCCC-G-TA-TTC---G-TA-T-----AAC-G-------TA--T--G-A---AATC-G--T---CCT-  4895
RCM.GAB1_AF382829         .........................................GTAGG-ATC-TAT---T---AG--G--TATGTA---G-T---TTT---TGGACT-----TAGTT----A---AAAAT---T---CCA-  4871
DRL.FAO_AY159321          ............................................-C-AT-GAACAT-TTC--GAGA-GTTTTGT--CTT-A-------AAC-AC-------AG-A----C--CT--G-A--T--CCCT-  4950
MND-1.F17                 ...............................................AT-GA-GTAGG-TGCTT-AG--CTTATG--...---AG-ATA---GC--T--TTAG--CGCCA-CCT-TG-GA-AGAGTGG-  5043
MND-2.5440_AY159322       .........................................-TT-C--T-GAATATC-----GA-A--TATTTC--TT-CA-------A-TGACG-----TAG-A----C--CGG-G-A--T--CCCA-  4947
MND-2.14CG_AF328295       .........................................AT--C--C-GAATAC--C--AG-GG-CTATTAT--T--CA-------A-T-AC-------AG-A-G------TG-G-A--T--CCCA-  5311
MND-2.MND2_AF367411       .........................................ATG-C-AT-GAATAT--GTTC-C-A-CTA-TAC--T--CA-------CC--AC-------AGGA-G--A---TG-A-A--TC-CCC--  5384
DEB.CM.CM40               .............................CCCACTATG-CAGTAGC-AT-GCAT--G---CA-A-ACAT-GAAG---G--C--ACA--A--TT-------T--C--T--A--C--AAGA------CCA-  5192
DEB.CM.CM5                .............................CCCTCTATG-CGGTAGG-ATC-CTT--G-GC---C-GGGT-GCAG---G--T--ACA-----C-----T--T--TT-G--A--CT-AAGA-----CCCT-  5186
MON.CM.L1                 GATCCCGTGTGGAATCCAAACCCAAGA.....G-TA---AGGT-...AG-ACAT--T-TGCT-CCC--CTG...--G.........--G--C--G--GA-T--GA--G-A---ACAC----T----CA-  5206
MON.NG1                   GATCGAGTGTGGTTTCCTAGTCCCAGAAGACCGG-CC-G-AAATTGATAG-GTCT--T-TGCT-CC--CCTA...---.........--A--TG----T------A----C--CACCCGC------TCA-  3817
TAL.CM.8023               .................ACAATATCCA-T-AGTG-TC------GGAT-TGACCTT-G-TGCA-AGGTCTATTCC---G-GT--ACA--A--GG---------C--G-----CACCA-G--T--CCCA-  4886
TAL.CM.266                .................ACAATATCCA-T-AGTG-TC------GGATAT-GTCC-AG-CG-A-AGGTGTA-GCC--TG-----AC---AA-TG-G-----T--CT-G------ACCA-A--T--CCCA-  5388
MUS.CM.1085               .............................A-GAGACATCCACAGGACAGGGAACA-G-TT-C-TC-GGTATTCT---T-CT-GAG---CAC---G--CATGAG-A-C--A---ACCC-G--T---TCT-  5157
SAB.SN.-.SAB1C            .............................ACA............GG-AT-G-ATAC-GT--AG-GG-AT-GTTT---G-GC-T--T---TGGAC---G-----TA-T------TG--CC--T--CCC--  5781
GRV.ET.x.GRI_677          .............................-TAGG-TT-...---TATT-CCATCAA--GGGA......TATAAG---G-------T--A-G-AC-------AG-A-G--A--C--A-----T---A---  5559
TAN.UG.x.TAN1             .............................-TAGCAAT-...AGTT-T--GAATT--C----C...ACGTA-AAA---G-----ACT-----TG-G--G-----TA-G------TGCC-C-----CCCT-  5588
VER.DE.-.AGM3             .............................-AGGG-AT-...CAGTATCT-A-C-AT-GGG-TCCT-GGTATAGG---G--T-G--T----C-AC------TAGC--G--A---ACCC----T---ACT-  5114
VER.KE.-.9063             .............................ATAGGAAT-...CAATACTTGG-AGCAGT-GGA...A--TT-AGA---G--T-------A-C-AC------TGG-A-G--A--C--AC-C--T---TCA-  5616
VER.KE.-.AGM155           .............................-TAGGAAT-...CAATAT-TTA-CCTAGT---TGATA-ATATAGA---G--T----T--CA-TAC-------TCCA-G--A---TGTC----T---AC--  5611
VER.KE.-.TYO1             .............................-TAGGTAT-...CAATAT-T-A-C-ATTT-G-ATCTA-ATATAGA---G--T---------CTAC------TAGTA----A---GGTC-C--T---A-T-  5106
SUN.GA.98.L14             .............................T--............-GTATGGAAT---TTT-T-ACAG-TATCAA---G--A----T---CTTG-G-----T--GA-G--C--CT-TA-G--T-----T-  5785
LST.CD.88.447             .........................................GTATC-AT-GAAT---TTT-TG-GA-CTATCAA---G--A----T--CATT------A-T---A-G--A--CTGTA-A--T---A---  4651
LST.CD.88.485             .........................................GTATC-AT-GAAT---TTT-TG-GA-CTATCAA---G--A----T--CATT--------T---A-G--A--CTGTA-A--T---A---  4651
LST.CD.88.524             .........................................GTATC-AT-GAAT---TCT-TG--A-CTATCAG---G----G--T--CATT-----T--T---A-GC-----G--A-G--T-------  4648
LST.KE.-.lho7             .........................................GTTTCCAT-GAAT--GTCT-T---G-CTATCAA---G--A-------CAT-------------A-G--C--CTATA-A--T---A---  5735
SYK.KE.-.SYK173           .............................AT-...A-GA-TCAATACC------ACC-GTG--AC...-CTGACTT-AC-CC--CA......G---------GC--T--A---AAT-T---T--CCCT-  5577
SYK.KE.KE51_AY523867      .............................-TAGGATT-A-T-GGGCCC-TCC-TCTGG-TG---T...-CAGAAGT-AC-CC-C-A......AC------TGC-T--G-A---AGTC-A-----CCCT-  5232
COL.CM.-.CGU1             ..........................AGA--A-CACC----...TCTT-TGAATTAC-G-TAG-TA-GT-GGAA---G-TT--AG-TGGC---A-----TAGGGTAC--G--CA-CATAAGGGGA-G-  5134
```

```
       Cys114 needed for APOBEC 3G degradation                                                                              SLQYLAL needed for APOBEC 3G degradation
                                |                                                    | Cys133 needed for APOBEC 3G degradation             |-              -|
H1B.FR.83.HXB2          GTTTTTCAGACTCTGCT...ATAAGAAAGGCCTTATTAGGACACATAGTTAGCCCTAGGTGTGAATATCAAGCAGGACATAAC.................AAGGTAGGATCTCTACAATACTTGGCACT  5489
Vif                     C__F__S__D__S__A__.__I__R__K__A__L__L__G__H__I__V__S__P__R__C__E__Y__Q__A__G__H__N__.__.__.__.__.__.__K__V__G__S__L__Q__Y__L__A__L   Vif

H1D.CD.84.84ZR085       -----G-----------...--C-----A---A---------T--------T----------------------------------..................-----------CT----G--T--A-----  5011
H1F1.BE.93.VI850        ----------A-----C...-----G--A---A---------T-G-A-------------A-C-----------------T..................-----------C--------T--------  4829
H1G.SE.93.SE6165        ----------------C...-----G--A---A---------A------------------------------------T..................C------------------T------T-  4886
H1H.CF.90.056           ----------A-----C...-----G--A---A--------GTG-------A---------A-C----C-----------A..................C--------A-----------------T-  4836
H1J.SE.93.SE7887        ----------------C...-----G------A---------A--------------------C--C---------------..................-------------------TC-------  4803
H1K.CM.96.MP535         ----------A-----A...G----G------A---------T-G------T-------------G----------------..................-----------CT-------T--------  4685
H101_AE.TH.90.CM240     ----------------C...-----G--A---A---------AG-------A-G--------------C-T-----------..................-----------C--------T--------  5063
H102_AG.NG.-.IBNG       ----------------C...--------A---A-------G-A-------A--------------------------------T..................-----------G--------T--------  5014
H103_AB.RU.97.KAL153_2  ----------A------...--------T---A---------T-G-----------T---------G---------------..................-----------------G-----------  4715
H104_cpx.CY.94.CY032    ----------A-----C...-----G--A---A---------T-G-----T-------------------------------T..................-----------CT--------C-------  4855
H1N.CM.95.YBF30         -----A---C------C...-----GC-A--GG-C-----GAGACC---ATTA----------------C------G--C--A..................C-------CA-C--------TC-A-----  5083
H1O.BE.87.ANT70         -----A---CA--A--A...G-C--G--A---A-CC-------G-G---ACTGA-C-A---------C-CTA---------GT..................C-------GA-A------CTGC-A--T--  5544
H1O.CM.91.MVP5180       -----A----A--A--A...--C--G------A-TC----G--G-G---GCTGA-C-A---------C-TG----------GT..................C-------GA-A--------T---A--CT-  5519
CPZ.CM.04.EK505         -----A---C------C...--C--GC-A--GG-C-----GAGACC---A-TA----A---------C-CT-----G-----ACAG...................------C--C---------TC-----T-  5061
CPZ.CM.04.LB7           -----G----A-----C...-----G------A---------GAG-------------A--------T--C---------C--GAAG...................-------A-C---------T------A-  5044
CPZ.CM.04.MB66          -----G----A-----C...-----G------A---------A----------------------------C-A-----------AAG...................------CAG-T-------T------T-  5035
CPZ.MT145               -----A---CAA----C...--C--G--C--TA-C----A-A-GGC---CCT-----A---------TC-C--A---G--C-G-AAG...................------G---T---------G---T--  5031
CPZ.CD.-.ANT            ----------TCAA--A...G-C---C-A--AA-CC-----G-A-G-A-ATTGA-ATAC--CC-C--CA-GAAG-------GT..................C-------GA-A--G--G--T--A--TT-  4934
CPZ.CM.-.CAM3           -----G---CT-----C...--C--G--A--TG-T-----GA-ACA---ATAT----A---------C-C-------G---C-A..................C-------G--AT-------T--A--C--  4903
CPZ.CM.98.CAM5          -----G---CT-----C...--C--G--A--TG-T-----GA-ACA---ATAT----A---------C-C-A----G--CC-A..................C---K---G--A--------T--A--C--  5191
CPZ.GAB_X52154          -C---G----AA----C...--C----GA--TA-------G--GC-T--AGCG---C-T--------CA---A------C-GG..................C----------CT----G-T------C--  5553
CPZ.GAB2                -----G---C-------...G-C----GA---A-CC-----GG-G--T--G-CAAAGGT--------TC-CTAGG--G---CCT..................C-A--TCAG--A--------TC----CA-  4882
CPZ.TZ.-.TAN1           -----G-T-CAGGA--A...G-TC-GC-A--AA--------G-A-G-A-ATTGA-ATTC--CC-C-T----T-------C-GA..................C-------GA----G-----T---A--TT-  5136
CPZ.US.85.CPZUS         -----A---C------C...--C-----A--TG-T-----GAGACC---A-TT--C-AA--------C-CTA----G--C--T..................---------C---T-G----T-------TT-  5549
GSN.99CM166_AF468659    -C---CAGA--GAA-AC...--C-G-G---AA--AGG--TG-GCAGA-CCTAGGA-AT------C-C-CG-A--CT--CCTG..................C--------A-A-CT--GACGC-T-AG-A  5307
GSN.99CM71_AF468658     -----CAGA--GAA-AC...G--C-G-G---AA--AGG--TG-ACAGC-CTTAGGATAT-----CC-C-CG-A--CT--CCCA..................C--------A-A-CT--CACGC-T-A-AG  5289
H2A.ALI_AF082339        -C---A-G-CAGG--AA...G------GA---A-CAG---GG-A-AGT-ATTGT-CT-C--CA-T----CCCAG-CC-----GTCA..................C-----CCG--A--C----TTC----CT-  5873
H2A.BEN_M30502          -C---A-G-CAGG--AA...G------GA---A-CAG---GG-A-AGT-ATTGT-CT-C--CA-C----CCCA--CT-----AGCA..................C-----CC---A--T------C-A--C--  5880
H2A.ST_M31113           -C---A-G-CAGGC-AA...G------GA---A-CAG---GG-A-AGC-ATTAT-CT-C--CA-C--C-CCCA--CC-----GTAC..................C-----CCG--A--C--G-TTC----CT-  5325
H2B.D205_X61240         -C-------C-AA--AA...G----G-GA---A-CAGG---G-A-AGA-ATTGT-CTAC--CA-C----C-T---CT--CG-AGGG..................C-----CC-AGCT----G-TTC-A--C--  5853
H2B.EHO_U27200          CG---A---CTAA--AA...G----G-GA---A-CAGG---G-A-AGA-ATTGT-CCAC--CA-C--C-C-T---CT----CAGGC..................C-----CC-AG-T----G-TTC-A--C--  5851
H2G.ABT96_AF208027      -C---A---CTGG--AA...G------GA---A-CAGG---G-GCAGA-ATTGT-CT-C--CA-C----C-A---CT-----AAGG..................C----GCCTAGCT-G----TTC-A--CT-  5207
H2U.FR.96.12034         -C---A---CGGG--AA...G-G--G-GA---A-CAG----G-ACAGT-ATTGT-CT-C--CA-C-T--C-T-G-CT----GACAG..................C-----CC----T------TTC-A-----  5358

MAC.239_M33262          -C---A---CGGGA-AA...G-G----G----A-CAGG---G-ACA-C-GCTGT--T-C--CAGG-TC-CGAG--CT-----GTAC..................C-----CC-AGC-----G-----A-----  5797
Vif                     C__F__T__A__G__E__.__V__R__R__A__I__R__G__E__Q__L__L__S__C__C__R__F__P__R__A__H__K__Y__.__.__.__.__.__Q__V__P__S__L__Q__Y__L__A__L   Vif

SMM.SL.92.SL92B         -C---G---CGCAC--G...G-G---C-A---A-CAG----G-GCAG--GTTAT-CTAC----GC--CGCG-T--CC---C--TCT..................-GT--CCA-AGC-----GCTG-----C--  5231
SMM.US.x.PGM53          -C--------GGGA-AA...G--C---G----A-CAGG---G-G-A-T-GCTGT--T-C--CA-G-TC-CGAA--CT-----GAAC..................C-----CC-AGC-----G--TC-A-----  5714
MNE.US.-.MNE027         -C---A---CGGGA-AA...G-G----G----A-CAGG---G-ACA-C-GCTGT--T-C--CAGG-TC-CGAG--CT-----GAAC..................C-----CC-AG------G-----A-----  5265
RCM.NG.x.NG411          -----A----TAAA--A...--T---C-A---A-C------G-GCA---GCTG-T-...----GG-T-G-GAA--------GAGAT..................C-------GAGC--------TC-----A-  5004
RCM.GAB1_AF382829       -----A----TAGA--A...--C---C-AAGTC--C----GG-A-A----CTAGTG...----CC-TC----GG-------GAGAT..................C-------GA----G----TT------A-  4980
DRL.FAO_AY159321        -----A----TCAA--A...------GGT-T----C-G---T-T-G-A--CAAG--...--CT-T-TC-CTAG---G--C--GGGT..................C-------GAG---T--G--TC-T--CT-  5059
MND-1.F17               --ACAGAGATAA--C-A...GAGGT-GCA-ATCATC--AT---TT--CA-TTTTA-GAC--CTTCATGG-CAGT-C-GT--TGAAAGCCATCAGGGGAGAAG-A--GTT-AAAG-TTGTAGA--TC--GC  5170
MND-2.5440_AY159322     -----A----TCAA--A...---------A-----G--T---G-A-G-C-GGTAG--...---T-C-TC-CCTGG-------GAGGT..................C-------GA-------G-TT-----CT-  5056
MND-2.14CG_AF328295     -----A---CTG-C---...G-G-------TGC-G-AC---G-A-G-A-AGTAG-G...---T-CAGC-CCTGG--------AGGT..................C-------GA--------CT------TT-  5420
MND-2.MND2_AF367411     -----A----ACAA--A...-------C-A---C-GC-----A-G-G-C-A-CAGTG...--CT-C-TC--CTGG--------AGGG..................--------GAGC--G--G--------TT-  5493
DEB.CM.CM40             ----C--CC-ACAG--A...G-G---C-A---A--AGG---G-G--CC-ACTG-AACA----C-GC-GT-CCATA-GGGACCA...GAAGGTAAAGGACCACCTTC-CT-CAGTACAT-GCAC-TAG-G-  5316
DEB.CM.CM5              ----C--CC-ACAG--A...G-C-C-C-A---C--AAG---G-G-A-C-ACTA-AACA----C-GC-AT-CCATA-GGGACCA...GAAGGTAAAGGACCACCCTC-TT-CAGTACTT-GCAC-AAGGAC  5310
MON.CM.L1               -C--CCA-C-G-TG-AC...G-GC-C-GA--AA-CAG----G-G-AGC-CTTGGGACA----C--C---T----AC---CT--CCGAAGGTAGGGACCCCGTT-ACCTT-GAGAA-TT-GC---CCTTGC  5333
MON.NG1                 ----CCAG-C-----AC...G--C-T-GA--AA-TAG---GG-AGCGM-CCTAGGG-AC---C-CC-CAT-CAGAC---CT--AATAAGGTAGGAACTCCTCCAACCCT-GAGAA-TT-GC---CCTTGC  3944
TAL.CM.8023             -----A---C---A--A...G-C---C-A---A-CAG---GG-GG--C--CTTG-GGT-------GTCAG-CATAA-GG--GGGAC..................TC-CC-CC---CT----G-T-C-A--CT-  4998
TAL.CM.266              -----A---CG--A--A...G-C---C-A---A-CAG---GG-GG-GC-ACTAG-AGTT----G-GTCAG-CACAA-GG--GGGAC..................-C-CC-CCG--------G-T-C-A--CT-  5500
MUS.CM.1085             -C--C-ATAG-CAG-AC...--CCC--G--T-A-CAG------GCACAC-CTAGGGGAT------C---C---C--AC-C---C-CAAAGGTAGGACCTCCTCCTTCTT--CA-GTC---GCT-TA--A---T-  5284
SAB.SN.-.SAB1C          -----A----TAGA---...G-CCA-C-A---A-CAGG---G-A-AGTA-CTTTGG...---A--C-C--G-TG------C-GCCAACA.........GGGC-A---CC---AT----G--T--A-----  5896
GRV.ET.x.GRI_677        -----A----TAGA--C...--CCA-C----TA-CAG---GG-G-AGTA--CGTGG...--CAC--TCA-G-A---------AGGT..................C-----CA---A--G---CTT--------  5668
TAN.UG.x.TAN1           -C---A----TAGA--C...--CCA-C-----A-CAG---GG-A-GCT--TTGTGG...--CAC---CA-G-A------CGTAGCAGAAAACCATTGGGGTC-----C-G--C--G--G-TT--A--C--  5712
VER.DE.-.AGM3           -----A----AAGG--C...--C--G--A---C----G-----G-GGT-C-C-TTC...---C-G-TC-CC-AG-----C--GAAAACA.........GGAC-----CCC---T-G--------A--T--  5229
VER.KE.-.9063           -----A----AAGG--C...--TCA---A--TA--AGG---G-G-G-T-CGT-TTC...---C---TC-CC-AG--------AACCACA.........GGA------CCC--A--G--G-----A--C--  5731
VER.KE.-.AGM155         -----A----TAGA--C...--CCA-C----AC--AGG---A---GGT-C-T-TTC...---C---T--C--G---------ACTAACA.........GGTC-----CCC--CT-G-----T--A---T-  5726
VER.KE.-.TYO1           -----AA---AAGA--C...--CCA-C-A--TC-GAGG--C----G-T--GT-TTC...---C-G-T--C--A---G-----AAGCACA.........GGAC-----CC----T-G--G---C-A--T--  5221
SUN.GA.98.L14           ---GGA-CTCTAGGTG-...--C----GA--TA-G-----GG-A-A-A-ACTA-A-GA---CAG-A------T--CT--C--AGGC..................CT---TCTC---T-G--G-TTC-CTGCT-  5897
LST.CD.88.447           -C-----TACAAGG-AC...--T--G-----T----GG---G-A-AGA-A-TGG-ATTC----CTATG-CCA-T--------AGGT..................TGT--TCTC--AT-G--G-T-C-CTGCT-  4763
LST.CD.88.485           -C-----TACAAGG-A-...--T--G-----T----GG---G-A-AGA-A-TGG-A-TC--C-CTATG-CCA-T--------AGGT..................TGT--TCTC--AT-G--G-TTC-CTGCT-  4763
LST.CD.88.524           ---GGGTGC-TAGG-AC...---GTT-G----A--AGG---G-A-AGA--CTG-AC--C---A--ATG-CTA-T--------AGGT..................TGT--TCTC--AT-G--G---C-CTGCT-  4760
LST.KE.-.lho7           -C--CAT-C-AAGA-AC...--T----G---AC-TC-G---G-A-G-A-A-CTATGT-T--CC-G...-CTTT---------AGGT..................TGT--TCTC---T-G--G-TTA-CTGCT-  5844
SYK.KE.-.SYK173         -C---A---CAAGA--G...G-C-ACC----AG-CAG---GG-AC--T-G-CAT-ACAC--CTGGACC-C-......----CTGAC..................C-----CCC--A--T--G--T--A---T-  5683
SYK.KE.KE51_AY523867    ----------TAGG--A...G-CCA-C-A--AC-TAG---GG-G-AGT-G-CTT-GCAC--CTGGA--TTC......-----AGAA..................C-----CTT--AT-G-----T-----TT-  5338
COL.CM.-.CGU1           --G-ACGCATAGAG-AAGAAGC------A--TA--AGG--CTTACCCTGG-AT---...-----T-TC--G-T------CCTA..................GGA--GTTGAAGT-G---G-AC-ATGT--  5243
```

```
H1B.FR.83.HXB2          AGCAGCATTAATA...............ACACCAAAAAAGATA...AAGCCACCTTTGCCTAGTGTTACGAAACTGACAGAGGATAG..........................................  5558
Vif                     __A__A__L__I__.__.__.__.__.__T__P__K__K__I__.__K__P__P__L__P__S__V__T__K__L__T__E__D__R__.__.__.__.__.__.__.__.__.__.__.__.__._   Vif

H1D.CD.84.84ZR085       -A-----------...............G------------G-...--------------C-------A---G--A-----A-----..........................................  5080
H1F1.BE.93.VI850        -A-----------...............G-T---G-G----C-...--------GC----------CCA------AGT------C--..........................................  4898
H1G.SE.93.SE6165        GAA--T----G--...............---T------G--GT...-G------------C---------G--T-AG----A-----..........................................  4955
H1H.CF.90.056           -A--------G-G...............G--------------...----------------------GA--G--AGT---------..........................................  4905
H1J.SE.93.SE7887        -A-----------...............-G--------G--G-...---------------------CA------AGT---------..........................................  4872
H1K.CM.96.MP535         CA--------G--...............G-T----G--G-CC-...---------G------------A------AGT---------..........................................  4754
H101_AE.TH.90.CM240     GAA--------C-...............----------G----...-G-------C----------A...G--A-A-----A-----..........................................  5129
H102_AG.NG.-.IBNG       GAAT--T---G--...............G------C-----C-...----------------------G---GT-AG----A--C--..........................................  5083
H103_AB.RU.97.KAL153_2  -----------G-...............---------------...-----------------------------------------..........................................  4784
H104_cpx.CY.94.CY032    -------------...............T-C----------C-...----------------------A------AGTG--------..........................................  4924
H1N.CM.95.YBF30         -A----C-GGG-G...............GG-G----G----G-...--------C--A--------G--T--G--A-----A-----..........................................  5152
H1O.BE.87.ANT70         -AG----G--G--...............-A-G---G--GC-GG...-----T--CC-A--C-----CCA----T-A-----A-----..........................................  5613
H1O.CM.91.MVP5180       GAA----G--G-G...............-A-GT-----GA-AT...-----T--CC-A--C-----CCA--G-T-A-----A-----..........................................  5588
CPZ.CM.04.EK505         -AT---C-GGG-GGGAGTACAGAAGAGA-G----CCCTTACCTAGTGT-A-CAAG--AA-AGAG-A--G............................................................  5130
CPZ.CM.04.LB7           CT-T--C---T--CAAAGACAGCCAACA-GG--GCCTTT-CCTAGTGTCAGCAAA--AG-AGAA-A--G............................................................  5113
CPZ.CM.04.MB66          GT--A--------AAAAGCAAACCAAGA-G----CCCTT-CCCAGTGTTAG-AAA--AG-AGAG-A--G............................................................  5104
CPZ.MT145               -AAGCAG--GG-TAAAACCAAGAAAAAT-G----CCTCT-CCCTCAGTAG--AAG--AG-AGAA-A--G............................................................  5100
CPZ.CD.-.ANT            TTGTAAGA-CC--GAATTCAGA......GG-TACCCC---GGTCCA-GAAG--AG--C------C-GT-T-TCT-A-----A-----..........................................  5015
CPZ.CM.-.CAM3           -AG---C-GGG--...............-G-GT-GG-----A-...--A-----C-----------C--T--GT-A-----A-----..........................................  4972
CPZ.CM.98.CAM5          -AG---T-GGG-G...............GG-GT-C------A-...-CA-----C-----C---A--GSC--GT-AG----A-----..........................................  5260
CPZ.GAB_X52154          -AA---T------...............T--GAG-G--G-CAT...-GA--------A--------CG-C---T-A-----------..........................................  5622
CPZ.GAB2                --T-CA-G--G--...............-AGA----G--A-AG...--------CC-A--------C--T-TCT-AG----A-----..........................................  4951
CPZ.TZ.-.TAN1           CAG-AAGG--G-TGAGAGCCAA...GAT-A--AGCC----GG-CCA-G-AGG--C-----ATC------A-----A--------C--..........................................  5220
CPZ.US.85.CPZUS         -AAG--C---G--...............GG--AG-GC--A-G-...-GA-----C--A--------G--T---T-------A-----..........................................  5618
GSN.99CM166_AF468659    GTT---C--CT-TGCATATGTCAAACATT....................................................................................................  5336
GSN.99CM71_AF468658     GCT---C--CT-TGCGTACGTTAGATACT....................................................................................................  5318
H2A.ALI_AF082339        --TG-T-G-GCA-CAAA......................................................................ATGGCA...AACCCCAGAGAAACAGTACCACCAGGAAACAGT  5929
H2A.BEN_M30502          --T--T-G--CA-CAAA......................................................................ATGACA...GACCCCAGAGAAAGGGTACCGCCAGGAAACAGT  5936
H2A.ST_M31113           --TG-T-G-GCA-CAAA......................................................................ATGGCA...GGCCCCAGAGAGACAATACCACCAGGAAACAGT  5381
H2B.D205_X61240         -AGG-TCG--CAGGAAGGAAAAA................................................................ATGGAT......CCCAGGGAGAGAGTGCCACCAGGAAACAGC  5912
H2B.EHO_U27200          -AG--TTG--CA-GAAGGAAAAG................................................................ATGGAT......CCCAGGGAGAGAGTACCACCAGGAAACAGC  5910
H2G.ABT96_AF208027      -CA--T-G-GCA-AAAGGAC...................................................................ATGGAT......CCAAGGGAGAGAGTMCCACCAGAAAACAGC  5263
H2U.FR.96.12034         CAG------GCA-G.........................................................................ATGGGA...GATCCAAGAGAGAGAATCCCACCAGGAAATTCA  5411

                                                                                                               / Vpx start
MAC.239_M33262          GAA--T-G---GCG.........................................................................ATGTCA...GATCCCAGGGAGAGAATCCCACCTGGAAACAGT  5850
Vif                     __K__V__V__S__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._._D__V__.__R__S__Q__G__E__N__P__T__W__K__Q_   Vif
Vpx                                                                                                            _M__S__.__D__P__R__E__R__I__P__P__G__N__S   Vpx

SMM.SL.92.SL92B         -AAG-T-G--T--CAAA......................................................................ATGACA...GACCCAAGGGAAAGAATCCCACCAGGAAACAGT  5287
SMM.US.x.PGM53          -A---T-G-G-GTC.........................................................................ATGTCA...GATCCCAGAGAGAGGATCCCACCTGGAAACAGT  5767
MNE.US.-.MNE027         GAG--T-G---GTT.........................................................................ATGTCA...GATCCCAGAGAGAGAATCCCACCTGGAAACAGT  5318
RCM.NG.x.NG411          -AA---T-GGGCTAGGCAGCAGCTAAAGG..........................................................ATGGCA...GAGGGCAGAGAAAGAGTGCCAGAGGCCCCCACT  5072
RCM.GAB1_AF382829       -CA---G-GGGCTAAA...AGTCAGTTAGATAGGT....................................................ATGGCAGAA............AGAGCCCCCGAG...GTCCCC  5039
DRL.FAO_AY159321        GCA---CCAC---AAGTACCTTAAGC.............................................................ATGGCAGAAAGACAGTCAGTGGAGAGAGCTCCAGCGGAGCCA  5127
MND-1.F17               T-GCCATAA-GC-CAAGGTGTTCTCTCTTTG-AGTTTCTCTGCTTG-GAGTCATC-.........................................................................  5226
MND-2.5440_AY159322     -CA---C-ATC-CAGAGGTAGAAAGC.............................................................ATGGCAGAG...AGG.........GCACCAGAG...GCGCCA  5109
MND-2.14CG_AF328295     -AG---C-AC--CAAGTTTTGCAGAA.............................................................ATGGCAGAA...AGA.........GCACCAGAG...GCACCA  5473
MND-2.MND2_AF367411     GCT-AGT-AT-C-GCCTACTGTAACA.............................................................ATGGCAGAAGGG............GCCCCAGAG...ATCCCA  5546
DEB.CM.CM40             -----T-AA-C......................................................................................................................  5327
DEB.CM.CM5              -----T-AA-T......................................................................................................................  5321
MON.CM.L1               CCT--TGCG-TATGGAGCAACCCCCACA-AGT--TCCGCTCC-TTGG-C-TC--GG.........................................................................  5389
MON.NG1                 TTGCTTCAAGTATGGAAAATGCACCACCGAG---TCCTCC-C-TTGG-C-TC--G..........................................................................  3999
TAL.CM.8023             -AAG-T-GC-G-TAGGC................................................................................................................  5015
TAL.CM.266              GA-----GC-G-TAGAT................................................................................................................  5517
MUS.CM.1085             -AAGAAGC.........................................................................................................................  5292
SAB.SN.-.SAB1C          TAGG-T--AT-C-A...................................................................................................................  5910
GRV.ET.x.GRI_677        --TT----AT-C-A...................................................................................................................  5682
TAN.UG.x.TAN1           GA---T--AT-C-GACTTC...CTAAGA-....................................................................................................  5738
VER.DE.-.AGM3           CCTT---........................-ACC---...........................................................................................  5243
VER.KE.-.9063           TCTG--C........................-ATC-G-...........................................................................................  5745
VER.KE.-.AGM155         -CT---C........................-ATC---...........................................................................................  5740
VER.KE.-.TYO1           GCTC---........................-ATC---...........................................................................................  5235
SUN.GA.98.L14           GAG--TTC-CC......................................................................................................................  5908
LST.CD.88.447           GAG-CA-C--CA-C...................................................................................................................  4777
LST.CD.88.485           GAG-CA-C--CA-C...................................................................................................................  4777
LST.CD.88.524           GAG-CA-C--CA-C...................................................................................................................  4774
LST.KE.-.lho7           GAG-CA-C--CA-C...................................................................................................................  5858
SYK.KE.-.SYK173         -CA--TG-ACC--AAAG................................................................................................................  5700
SYK.KE.KE51_AY523867    -CA-AAG-ATT--AGCAAG..............................................................................................................  5357
COL.CM.-.CGU1           TCTG--CC---G-GAGAAA...GAGAGG--C-ACC-G---.........................................................................................  5280
```

```
H1B.FR.83.HXB2            ..................................................................................................................................  5558
H1D.CD.84.84ZR085         ..................................................................................................................................  5080
H1F1.BE.93.VI850          ..................................................................................................................................  4898
H1G.SE.93.SE6165          ..................................................................................................................................  4955
H1H.CF.90.056             ..................................................................................................................................  4905
H1J.SE.93.SE7887          ..................................................................................................................................  4872
H1K.CM.96.MP535           ..................................................................................................................................  4754
H101_AE.TH.90.CM240       ..................................................................................................................................  5129
H102_AG.NG.-.IBNG         ..................................................................................................................................  5083
H103_AB.RU.97.KAL153_2    ..................................................................................................................................  4784
H104_cpx.CY.94.CY032      ..................................................................................................................................  4924
H1N.CM.95.YBF30           ..................................................................................................................................  5152
H1O.BE.87.ANT70           ..................................................................................................................................  5613
H1O.CM.91.MVP5180         ..................................................................................................................................  5588
CPZ.CM.04.EK505           ..................................................................................................................................  5130
CPZ.CM.04.LB7             ..................................................................................................................................  5113
CPZ.CM.04.MB66            ..................................................................................................................................  5104
CPZ.MT145                 ..................................................................................................................................  5100
CPZ.CD.-.ANT              ..................................................................................................................................  5015
CPZ.CM.-.CAM3             ..................................................................................................................................  4972
CPZ.CM.98.CAM5            ..................................................................................................................................  5260
CPZ.GAB_X52154            ..................................................................................................................................  5622
CPZ.GAB2                  ..................................................................................................................................  4951
CPZ.TZ.-.TAN1             ..................................................................................................................................  5220
CPZ.US.85.CPZUS           ..................................................................................................................................  5618
GSN.99CM166_AF468659      ..................................................................................................................................  5336
GSN.99CM71_AF468658       ..................................................................................................................................  5318
H2A.ALI_AF082339          GGCGAAGAGACTATCGGAGAGGCCTTCGAGTGGCTAGACAGGACAGTAGAGGCCTTAAACAGAGAGGCGGTGAATCACCTG...CCCCGGGAGCTCATTTTCCAGGTGTGGCAAAGGTCCTGGAGATACT  6056
H2A.BEN_M30502            GGAGAAGAGACCATTGGAGAGGCCTTCGAGTGGCTAGAGAGGACCATAGAAGCCTTAAACAGGGAGGCAGTGAACCATCTG...CCCCGAGAGCTCATTTTCCAGGTGTGGCAAAGGTCCTGGAGATATT  6063
H2A.ST_M31113             GGCGAAGAAACTATCGGAGAGGCCTTCGAGTGGCTAGACAGGACGGTAGAAGCCATAAACAGAGAGGCAGTGAACCACCTG...CCCCGAGAGCTTATTTTCCAGGTGTGGCAAAGGTCCTGGAGATACT  5508
H2B.D205_X61240           GACGAAGAAACAGTAGGAGAAGCATTCGCTTGGCTAGAAAGAACAATAACAGAGCTCAACAGGGTAGCGGTCAACCATTTG...CCCCGAGAACTTATTTTCCAGGTCTGGCAGAGGTCTTGGGCATACT  6039
H2B.EHO_U27200            GGCGAAGAAACAGTAGGAGAGGCATTCGAATGGCTAGAGACAACATTAGAACATCTCAACAGAGTAGCAGTCAATCACTTG...CCCAGGGAACTTATTTTCCAGGTCTGGCAGAAGTCTTGGGCATATT  6037
H2G.ABT96_AF208027        GACGAGGAGACAATCGGAGAGGCATTCGAATGGCTAGAAAGAACAGTMACAGAGATCAACAGGATAGCAGTCAATCATTTA...CCCAGGGAGCTTATTTTCCAGGTTTGGCGAAGGTCCTGGGAATATT  5390
H2U.FR.96.12034           GAAGAAGAAACTGTAGGAGAGGCTTTCAGTTGGCTAGAAAGAACTGTGAGAGATATCAACAGGGCAGCAGTGCAACATCTT...CCCAGGGAACTTATTTTCCAGGTCTGGCAAAGGTCCTGGCAATACT  5538

MAC.239_M33262            GGAGAAGAGACAATAGGAGAGGCCTTCGAATGGCTAAACAGAACAGTAGAGGAGATAAACAGAGAGGCGGTAAACCACCTA...CCAAGGGAGCTAATTTTCCAGGTTTGGCAAAGGTCTTGGGAATACT  5977
Vif                       W__R__R__D__N__R__R__G__L__R__M__A__K__Q__N__S__R__G__D__K__Q__R__G__G__K__P__P__.__T__K__G__A__N__F__P__G__L__A__K__V__L__G__I__L   Vif
Vpx                       _G__E__E__T__I__G__E__A__F__E__W__L__N__R__T__V__E__E__I__N__R__E__A__V__N__H__L__.__P__R__E__L__I__F__Q__V__W__Q__R__S__W__E__Y_   Vpx

SMM.SL.92.SL92B           GGAGAAGAAACAATAGGAGAGGCCTTCGAATGGTTACACAATACAGTGGAGGCCCTGAATCAAACAGCAGTACAACACCTA...CCAAGAGAGTTAATTTTCCAGGTCTGGAGAAGGTGTTGGGAATACT  5414
SMM.US.x.PGM53            GGAGAAGAAACAATAGGAGAGGCATTCGAATGGCTAAACAGAACAGTAGAAGAAATAAACAGGGCAGCAGTGAATCACTTG...CCGAGGGAGCTAATTTTCCAGGTTTGGCGAAGGTCTTGGGAATACT  5894
MNE.US.-.MNE027           GGAGAAGAGACAATAGGAGAAGCCTTCGAATGGCTAAACAGAACAGTAGAGGAGATAAACAGAGAGGCAGTAAACCACCTA...CCAAGGGAGCTGATTTTCCAGGTTTGGCAAAGGTCTTGGGAATACT  5445
RCM.NG.x.NG411            GGGGCTGGAGATGTAGAGTTTGCCCCCTGGCTTCACAGAATGCTAACAGAAGTCAACTTAGAAGCC...AGGTTGCACTTTCATCCA...GAGTTCATTTTCCGTCTGTGGCGTACTTGTGTGGAACACT  5196
RCM.GAB1_AF382829         ACTGGGGCTGGAGAAGCAGAGTTCCAGCCTTGGCTACGGGACATGCTAGAAAAGGTCAACTTGGAAGCCAGGTTACACTTTCATCCA...GAGTTCATTTTCCGTCTGTGGCGCACTTGTGTGGAACACT  5166
DRL.FAO_AY159321          ATGGGAGCAGGAGAGGTAGAGTTAGAAGAATGGCTACAGAGGAGTCTCTTAAGAATCAACCAGGAGGCTCGATTACACTTCCACCCA...GAGTTCCTCTTCCGTCTTTGGAACACCTGCATGGAGCACT  5254
MND-1.F17                 ..................................................................................................................................  5226
MND-2.5440_AY159322       GAAGGAGCAGGAGAGGTAGGACTGGAGCAATGGCTGGAAACGTCACTGGAGAGAATCAACCGGGAGGCCCGGTTACACTTCCACCCA...GAGTTCCTTTTCCGTCTCTGGAACACATGTGTAGAACACT  5236
MND-2.14CG_AF328295       CAGGGGGCAGGAGAAGTGGGACTAGAACAGTGGCTGGAAAGATCATTGGAACAAGTGAACAGAGAGGCTCAATTACACTTCCACCCA...GAGTTCCTTTTCCGTCTTTGGAACACCTGTGTAGAACACT  5600
MND-2.MND2_AF367411       GAAGGAGCAGGAGAGGTGGATCTAAATACCTGGTTGGAAAGATCTCTGGAGAAAATCAATCAGGAGGCAAGATTACACTTCCACCCA...GAGTTCCTTTTCCGTCTCTGGAACGCATGCATAGAACACT  5673
DEB.CM.CM40               ..................................................................................................................................  5327
DEB.CM.CM5                ..................................................................................................................................  5321
MON.CM.L1                 ..................................................................................................................................  5389
MON.NG1                   ..................................................................................................................................  3999
TAL.CM.8023               ..................................................................................................................................  5015
TAL.CM.266                ..................................................................................................................................  5517
MUS.CM.1085               ..................................................................................................................................  5292
SAB.SN.-.SAB1C            ..................................................................................................................................  5910
GRV.ET.x.GRI_677          ..................................................................................................................................  5682
TAN.UG.x.TAN1             ..................................................................................................................................  5738
VER.DE.-.AGM3             ..................................................................................................................................  5243
VER.KE.-.9063             ..................................................................................................................................  5745
VER.KE.-.AGM155           ..................................................................................................................................  5740
VER.KE.-.TYO1             ..................................................................................................................................  5235
SUN.GA.98.L14             ..................................................................................................................................  5908
LST.CD.88.447             ..................................................................................................................................  4777
LST.CD.88.485             ..................................................................................................................................  4777
LST.CD.88.524             ..................................................................................................................................  4774
LST.KE.-.lho7             ..................................................................................................................................  5858
SYK.KE.-.SYK173           ..................................................................................................................................  5700
SYK.KE.KE51_AY523867      ..................................................................................................................................  5357
COL.CM.-.CGU1             ..................................................................................................................................  5280
```

```
H1B.FR.83.HXB2             .................................................................................................................................. 5558
H1D.CD.84.84ZR085          .................................................................................................................................. 5080
H1F1.BE.93.VI850           .................................................................................................................................. 4898
H1G.SE.93.SE6165           .................................................................................................................................. 4955
H1H.CF.90.056              .................................................................................................................................. 4905
H1J.SE.93.SE7887           .................................................................................................................................. 4872
H1K.CM.96.MP535            .................................................................................................................................. 4754
H101_AE.TH.90.CM240        .................................................................................................................................. 5129
H102_AG.NG.-.IBNG          .................................................................................................................................. 5083
H103_AB.RU.97.KAL153_2     .................................................................................................................................. 4784
H104_cpx.CY.94.CY032       .................................................................................................................................. 4924
H1N.CM.95.YBF30            .................................................................................................................................. 5152
H1O.BE.87.ANT70            .................................................................................................................................. 5613
H1O.CM.91.MVP5180          .................................................................................................................................. 5588
CPZ.CM.04.EK505            .................................................................................................................................. 5130
CPZ.CM.04.LB7              .................................................................................................................................. 5113
CPZ.CM.04.MB66             .................................................................................................................................. 5104
CPZ.MT145                  .................................................................................................................................. 5100
CPZ.CD.-.ANT               .................................................................................................................................. 5015
CPZ.CM.-.CAM3              .................................................................................................................................. 4972
CPZ.CM.98.CAM5             .................................................................................................................................. 5260
CPZ.GAB_X52154             .................................................................................................................................. 5622
CPZ.GAB2                   .................................................................................................................................. 4951
CPZ.TZ.-.TAN1              .................................................................................................................................. 5220
CPZ.US.85.CPZUS            .................................................................................................................................. 5618
GSN.99CM166_AF468659       .................................................................................................................................. 5336
GSN.99CM71_AF468658        .................................................................................................................................. 5318
H2A.ALI_AF082339           GGCATGATGAACAAGGAATGTCACAAAGTTACACAAAATATAGATACTTGTGCTTAATGCAGAAAGCTATGTACACACATTTCATGAAAGGGTGCACCTGCCTGGGG.........GGAGGG...CATGG 6174
H2A.BEN_M30502             GGCATGATGAACAAGGGATGTCAGCAAGCTACACAAAGTATAGATATTTGTGCCTAATGCAAAAAGCTATATTTACACATTTCAAGAGAGGGTGCACTTGCTGGGGG.........GAGGAC...ATGGG 6181
H2A.ST_M31113              GGCATGATGAACAAGGAATGTCAATAAGTTACACAAAGTATAGATATTTGTGCCTAATGCAGAAAGCTATGTTCATACATTCTAAGAGAGGGTGCACTTGCCTGGGG.........GGAGGA...CATGG 5626
H2B.D205_X61240            GGCGTGAGGAACAGGGCATGTCAATTAGCTATACCAAATATAGATACTTGTTGCTAATGCAGAAAGCAATGTTTGTGCACTATACAAAGGGCTGTAGGTGCCTGCAG.........GAGGGC...CATGG 6157
H2B.EHO_U27200             GGCGTGAGGAACAGGGCATGTCAATTAGTTATACCAAATATAGGTATTTGTGTCTAATGCAAAAAGCAATGTTTATACACTTTGCAAAGGGTTGTGGATGTCTGCGG.........GAAGGC...CATGG 6155
H2G.ABT96_AF208027         GGCATGATGAACAGGGCATGTCAGCTAGTTACACTAAGTATAGATATTTGTTGCTGATGCAAAAGGCTATGTTTGTACATTGTAAGAAAGGATGCACATGCTTGCAG.........AAGGGC...CATGG 5508
H2U.FR.96.12034            GGCGTGATGAACAGGGCATGTCAGAAAGTTATACTAAGTATAGATATCTCTGTCTGATACAGAAAGCTTTATTTATGCATTGCAAAAAGGGATGTCAATGCTTGGGG.........GAAGGC...CATGG 5656

                                 \ Vif end in HIV-2 and SMM sequences
MAC.239_M33262             GGCATGATGAACAAGGGATGTCACCAAGCTATGTAAAATACAGATACTTGTGTTTAATACAAAAGGCTTTATTTATGCATTGCAAGAAAGGCTGTAGATGTCTAGGG.........GAAGGA...CATGG 6095
Vif                        __A__*                                                                                                                              Vif
Vpx                        W__H__D__E__Q__G__M__S__P__S__Y__V__K__Y__R__Y__L__C__L__I__Q__K__A__L__F__M__H__C__K__K__G__C__R__C__L__G__.__.__.__E__G__.__H__G  Vpx

SMM.SL.92.SL92B            GGGTAGATGAGCAAGGATACTCACCAAGCTATGCAAAATATAGATATGTACAGTTAATGCAAAAAGCAATGTTTCAGCACTTCAGAAAGGGATGCACATGCAGAGGA.........GAAGGG...CATTC 5532
SMM.US.x.PGM53             GGCGTGATGAAATGGGGATGTCAGAGAGCTACACAAAATACAGATACTTGTGCTTAATACAGAAAGCTCTGTTTGTGCATTGCAAGAGAGGGTGTAGGTGCTTAGGA............GAAGAGCATGG 6012
MNE.US.-.MNE027            GGCATGATGAACAAGGGATGTCACCAAGCTATGTAAAGTACAGATACTTGTGTTTAATACAAAAGGCTTTATTTATGCATTGCAAAAAAGGCTGTAGATGTCTAGGG.........GAAGGA...CATGG 5563
RCM.NG.x.NG411             GGCATGATAGGCTTGGAAGAAGCCTTGAGTATGCAGGCTATAGATATCTGCTTCTGATGCAAAAAGCTCTGTTTATTCATTGCCAATCAGGGTGTTCTCAGAGA.....................CATGG 5305
RCM.GAB1_AF382829          GGCATGACGTACATCAGAGAAGTTTAGAATATGCTGCTTATAGGTACTTACTGCTGATGCAAAAGGCTTTGTTTATCCACTGTCAGACAGGGTGTAGTCAGAGA.....................CATGG 5275
DRL.FAO_AY159321           ACCATGATGCTCTTCAGTTATCTTTTACTTACAGCAAGTATAGATACCTACTTTTGTTACAGAAGGCCATGTTCATGCACTTTCAGCAAGGATGCTCATGTCTGCAGGGAAGGCATCCACCTCCCCTCAG 5384
MND-1.F17                  .................................................................................................................................. 5226
MND-2.5440_AY159322        GGCATGATAGACATCAGAGATCTCTTGATTATGCCAAGTATAGATACCTGCTGTTGATGCATAAGGCCATGTATACTCACATGCAACAGGGATGCCCATGTAGAAAT............GGG...CGCCC 5351
MND-2.14CG_AF328295        GGCATGATAGACTTAGAAGAACCTTAAATTATGCCAAATATAGGTATCTTCTGCTGATGCAGAAAGCCATGTTTGTTCATATGCAACAAGGATGCCCATGTAGAAGT............GGA...CATTC 5715
MND-2.MND2_AF367411        GGCATGACAGACATCAGAGAAGCTTAAGCTATGCAAAGTATAGATACCTGCTGTTGATGAATAAAGCAATGTTTACTCACATGCAACAAGAATGTCCTTGTAGAAGT............GGG...CATCC 5788
DEB.CM.CM40                .................................................................................................................................. 5327
DEB.CM.CM5                 .................................................................................................................................. 5321
MON.CM.L1                  .................................................................................................................................. 5389
MON.NG1                    .................................................................................................................................. 3999
TAL.CM.8023                .................................................................................................................................. 5015
TAL.CM.266                 .................................................................................................................................. 5517
MUS.CM.1085                .................................................................................................................................. 5292
SAB.SN.-.SAB1C             .................................................................................................................................. 5910
GRV.ET.x.GRI_677           .................................................................................................................................. 5682
TAN.UG.x.TAN1              .................................................................................................................................. 5738
VER.DE.-.AGM3              .................................................................................................................................. 5243
VER.KE.-.9063              .................................................................................................................................. 5745
VER.KE.-.AGM155            .................................................................................................................................. 5740
VER.KE.-.TYO1              .................................................................................................................................. 5235
SUN.GA.98.L14              .................................................................................................................................. 5908
LST.CD.88.447              .................................................................................................................................. 4777
LST.CD.88.485              .................................................................................................................................. 4777
LST.CD.88.524              .................................................................................................................................. 4774
LST.KE.-.lho7              .................................................................................................................................. 5858
SYK.KE.-.SYK173            .................................................................................................................................. 5700
SYK.KE.KE51_AY523867       .................................................................................................................................. 5357
COL.CM.-.CGU1              .................................................................................................................................. 5280
```

```
                                                                                        / Vpr start
H1B.FR.83.HXB2              ............................................................ATGGAACAAGCC..........................................................  5570
Vif                         .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__W__N__K__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.   Vif
Vpr                                                                                     _M__E__Q__A__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   Vpr
H1D.CD.84.84ZR085           ............................................................------------..........................................................  5092
H1F1.BE.93.VI850            ............................................................------------..........................................................  4910
H1G.SE.93.SE6165            ............................................................------------..........................................................  4967
H1H.CF.90.056               ............................................................------------..........................................................  4917
H1J.SE.93.SE7887            ............................................................------------..........................................................  4884
H1K.CM.96.MP535             ............................................................------------..........................................................  4766
H101_AE.TH.90.CM240         ............................................................------------..........................................................  5141
H102_AG.NG.-.IBNG           ............................................................------AG----..........................................................  5095
H103_AB.RU.97.KAL153_2      ............................................................------------..........................................................  4796
H104_cpx.CY.94.CY032        ............................................................------------..........................................................  4936
H1N.CM.95.YBF30             ............................................................-------G---A..........................................................  5164
H1O.BE.87.ANT70             ............................................................-----------A..........................................................  5625
H1O.CM.91.MVP5180           ............................................................------------..........................................................  5600
CPZ.CM.04.EK505             ............................................................--------G--A..........................................................  5142
CPZ.CM.04.LB7               ............................................................----------T-..........................................................  5125
CPZ.CM.04.MB66              ............................................................---------CT-..........................................................  5116
CPZ.MT145                   ............................................................------------..........................................................  5112
CPZ.CD.-.ANT                ............................................................------------..........................................................  5027
CPZ.CM.-.CAM3               ............................................................--------G--A..........................................................  4984
CPZ.CM.98.CAM5              ............................................................---------ATA..........................................................  5272
CPZ.GAB_X52154              ............................................................-----------A..........................................................  5634
CPZ.GAB2                    ............................................................------------..........................................................  4963
CPZ.TZ.-.TAN1               ............................................................-----------A..........................................................  5232
CPZ.US.85.CPZUS             ............................................................----------TG..........................................................  5630
GSN.99CM166_AF468659        ............................................................------AG-.............................................................  5345
GSN.99CM71_AF468658         ............................................................------AG-.............................................................  5327
H2A.ALI_AF082339            GCCAGGAGGATGG...AGATCAGGACCTCCCCCTCCTCCCCCTCCAGGTCTAGTCTA...---...................................................................  6231
H2A.BEN_M30502              CCGGGAAGGATTGGAAGACCAAGGACCTCCCCCTCCTCCCCCTCCAGGTCTAGTCTA...---...................................................................  6241
H2A.ST_M31113               GCCGGGAGGATGG...AGATCAGGACCTCCCCCTCCTCCCCCTCCAGGTCTAGTCTA...---...................................................................  5683
H2B.D205_X61240             GCCAGGGGGATGG...AGATCAGGACCTCCTCCTCCTCCTCCCCCAGGCCTGGCCTA...---...................................................................  6214
H2B.EHO_U27200              GCCAGGGGGATGG...AGATCAGGACCTCCTCCTCCTCCTCCCCCAGGTCTAGCCTA...---...................................................................  6212
H2G.ABT96_AF208027          GCCTGGAGGATGG...AGACAAGGACCTCCTCCTCCTCCTCCTCCAGGTCTAGCATAG..---...................................................................  5566
H2U.FR.96.12034             ACCAGGAGGATGG...AGATCAGGGCCTCCTCCTCCTCCCCCACCAGGATTGATCTA...---...................................................................  5713

                                                                            Vpx end \   / Vpr start
MAC.239_M33262              GGCAGGGGGATGG...AGACCAGGACCTCCTCCTCCTCCCCCTCCAGGACTAGCATAA..---...................................................................  6153
Vpx                         __A__G__G__W__.__R__P__G__P__P__P__P__P__P__P__G__L__A__*                                                                            Vpx
Vpr                                                                                     _M__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.   Vpr
SMM.SL.92.SL92B             ACAGGGGGGCTGG...AGAACTGGACCACCACCTCCTCCCCCACCAGGTTTGGCCTA...---...................................................................  5589
SMM.US.x.PGM53              GGCAGGGGGATGG...AGATCAGGGCCTCCTCCTCCTCCCCCTCCAGGACTAGCATAAA.A---..................................................................  6072
MNE.US.-.MNE027             GGCAGGGGGATGG...AGACCAGGACCTCCTCCTCCTCCCCCTCCAGGACTAGCATAA..---...................................................................  5621
RCM.NG.x.NG411              ACAGGGACAAGCAAGGGAAGCAGGAGAAAGAATCCAGATTCTTCCGGGAATG...TAA............................................................TCAAGCACTCCT  5372
RCM.GAB1_AF382829           A...CCCAATCCTAGGGCAGTAGGAGAAAGGATAACCATCCTACCGGGGATG...TAA..............................................................TGATGGCCCT  5337
DRL.FAO_AY159321            ACCAGCAGGAGAT...AGACTT...CCTCCTCCTCCTCCTCCA............TGA...........................................................TTGGTCTAATAAT  5440
MND-1.F17                   ............................................................----GC--GAAGAGA..............................................GATGAGCAA  5250
MND-2.5440_AY159322         A...............AGG...GGA.........CCTCCTCCTCCAGGGATGGCTTAA....................................................................TTTC  5386
MND-2.14CG_AF328295         A...............AGA...GGA.........CCTCCTCCTCCAGGATTGGCCTAA......................................................................CT  5748
MND-2.MND2_AF367411         G...............AGA...GGA.........CCTCCTCCTCCAGGGATGGTCTAA......................................................................CC  5821
DEB.CM.CM40                 ............................................................------.............................................................AGG  5336
DEB.CM.CM5                  ............................................................------.............................................................AGG  5330
MON.CM.L1                   ............................................................----TT................................................................  5395
MON.NG1                     ..................................................................................................................................  3999
TAL.CM.8023                 ............................................................------AGGTTG..........................................................  5027
TAL.CM.266                  ............................................................------AGGTTG..........................................................  5529
MUS.CM.1085                 ............................................................---...G--............AGCTTGCCCCCCAGCCACCCA............................  5319
SAB.SN.-.SAB1C              ............................................................----CCTC--GAGGGTGGCTCCCACCAGTAGGAGGGGATCCTCCCAAGGATCCCCCCAAGAATCCCAGA  5979
GRV.ET.x.GRI_677            ............................................................----C-TC--GAAGAGATCCAAGAGAACCTTTACCAGGATGG............................  5724
TAN.UG.x.TAN1               ............................................................----C-G---GAAGA..............................................GATTCCAGG  5762
VER.DE.-.AGM3               ............................................................----CCTC--G-AGA..............................................GATCCCAGA  5267
VER.KE.-.9063               ............................................................----C-TC--GGAGA..............................................GGTCCCAGA  5769
VER.KE.-.AGM155             ............................................................----CCTC--GAAGA..............................................GATCCCAGA  5764
VER.KE.-.TYO1               ............................................................----CCTC--GGAGA..............................................GATCCAAGA  5259
SUN.GA.98.L14               ............................................................----CCAGC....................................................AGGAGAGAG  5926
LST.CD.88.447               ............................................................---TCCAG-A-AAGG..............................................CAGCAAAGG  4801
LST.CD.88.485               ............................................................---TCCAG---AAGG..............................................CAGCAAAGG  4801
LST.CD.88.524               ............................................................---TCCAG-CAAAGA..............................................CAGCAGCGC  4798
LST.KE.-.lho7               ............................................................---TCCAGGCAGAGG..............................................CAGCAAAGA  5882
SYK.KE.-.SYK173             ............................................................----CGG-G--TTTC............................TTCAATCCCTCCCAGCATGTGCAAGGA  5742
SYK.KE.KE51_AY523867        ...........................................................G----C---G--TTTC............................TTCAGAGTTTACCAGCGGCAGCCAGGG  5400
COL.CM.-.CGU1               ............................................................---AG-.......................................................GTCCTGAAA  5295
```

```
                                                                      Vif end in HIV-1 \
H1B.FR.83.HXB2              ........................CCAGAAGACCAAGGGCCACAGAGGGAGCCACACAATGAATGGACACTAGAGCTTTTAGAGGAGCTTAAGAATGAAGCTGTTAGACATTTT...CCTAGGATTTGGC  5673

Vif                         __.__.__.__.__.__.__.__.__Q__K__T__K__G__H__R__G__S__H__T__M__N__G__H__*                                                      Vif
Vpr                         _.__.__.__.__.__.__.__.__P__E__D__Q__G__P__Q__R__E__P__H__N__E__W__T__L__E__L__L__E__E__L__K__N__E__A__V__R__H__F__.__P__R__I__W__  Vpr

H1D.CD.84.84ZR085           ........................--G-C------------G------------T--------------T------------------------G---------C---------...--------A----  5195
H1F1.BE.93.VI850            ........................----G------G------------------T-----------G-------AA-------------------------------G------...------CCA----  5013
H1G.SE.93.SE6165            ........................-----------G-----------A--A--CT-----------G--T----A--G-----A--------A---------------------...------C------  5070
H1H.CF.90.056               ........................-----------G---------------------------------------------------A----------G--------G------...-----AG-A----  5020
H1J.SE.93.SE7887            ........................-----------C-----------A------T--C----------------------G-----------A-----------------C---...------CCA----  4987
H1K.CM.96.MP535             ........................-----------G------------------A--------------T-----A----G-------------GA------------------...------CCA----  4869
H101_AE.TH.90.CM240         ........................-------...-G-----------A--A--CT--------------T----A--G--------------A---------------------...------CCC----  5241
H102_AG.NG.-.IBNG           ........................-----------G----------------GTC------------------A--G-----A--------AC----------------C---...------CCG----  5198
H103_AB.RU.97.KAL153_2      ........................-----G--------A---------------T-------G-------------------------------G---------------C---...------G-G----  4899
H104_cpx.CY.94.CY032        ........................-----------G-----G-----A--A---A-T-----------------------G-----------A--------------G------...-----ACCC----  5039
H1N.CM.95.YBF30             ........................--------TGC---------------A--CT-T---------G-------AT-A-----A--AT-A--A-----------GC-C------...--A----------  5267
H1O.BE.87.ANT70             ........................--T--GA-T--G--A---GCT-AA-----CTT----------G----------C-----A-----A--AGCA-----A--A--------C...------CC-----  5728
H1O.CM.91.MVP5180           ........................-TG---A-T--G--A---GCT---------TT---------------------CC-G--A-----G--AG-A-----A--A--------C...------CC----T  5703
CPZ.CM.04.EK505             ........................--------T-----A---------------T----C------G-TT----AT-G--G--A--C--A--------G---C-GC-C--C---...---C--CC-----  5245
CPZ.CM.04.LB7               ........................------------------------------T-T-----G---G--T-----G-A-----A--A--C--AC-G-----A--A-----C---...-----ACCA----  5228
CPZ.CM.04.MB66              ........................-----------G--A-----A-----A---T-T------------------A-A-----A--A-----A-GG-----A--A-----C--C...---C-AGA-----  5219
CPZ.MT145                   ........................-----G-----G--A---TCA--A-----CT-TG-GC-G---CTGT----CACA--G--A--AA----AC-A-----A--------C---...--C--ACCC---T  5215
CPZ.CD.-.ANT                ........................--GC-G--TG-G-----------A--A---ATG-----G---TTGT----AAC------A--AA-A--A-----G--A--A--------C...--ACA-CC-ACAT  5130
CPZ.CM.-.CAM3               ........................--------T--G--T----CA--A------T-----------G--T----A--GC-G-----A--A--------------AC-C------...---C-AGCA----  5087
CPZ.CM.98.CAM5              ........................-----------G-------CA--A--A---T-----------R-------AA-A-----A-----A--A-C---G---C-A---------...---C-CCC-----  5375
CPZ.GAB_X52154              ........................-----G--------T----CA--A--A--TT--C-G------G--T----ACC---G--A--AT-A--A--------A--A--------C...-----ACCA---T  5737
CPZ.GAB2                    ........................-----------G--A---GCAG-A--A--TT-------G---G-------AA-A--G-----CT-A--A--C--G-----C---------...-----ACCA----  5066
CPZ.TZ.-.TAN1               ........................--GA-C---A-C-----G-----A--A---T---C---G---CTGT----CA-CC----A--AA-A--AC-A-----A--G-A---C---...--A--ACCAATAT  5335
CPZ.US.85.CPZUS             ........................------------------GCC--A--A---T-------G---G--T----TG-C-----A--A--A--AG-A--G---T-GC-C------...-----ACCC----  5733
GSN.99CM166_AF468659        ATCCCACCCAGTCATCCTATGCCT......TGG-TCA-C-G-AG-GT-CCCA--AC--TG-CTATCG-T-A-A-CGC-C-GTG---AA-A--CG-G--G--A-AG-A---C--C...T-A---GACGA--  5466
GSN.99CM71_AF468658         ACCCCACCAAGTCACCCTCTACCT......TGGATCA-C-G-AG-GT-CCCA--AC--TG-C-GTTG-T-A-AGCGCCA-GTG---AA-A--CG-A--G--A-AA-A---C--C...T-A---GAAGA--  5448
H2A.ALI_AF082339            ACTGAAGCACCAACAGAGTTTCCC--G-CC-GAATG--A--C--CCA--GAG-CAGGG----G---GT-A----AG-CC-GAGA--AA-A--AG-A------T-A--G------GAC---C-C--G...-  6358
H2A.BEN_M30502              ACTGAAGCACCAACAGAGTTTCCC--------TGGGACC----G------CTT-GGG-G---C---GT-A----AAC-C-GAG---AA-A---G-A-----CT-A---------GAT--CC-CT-G...-  6368
H2A.ST_M31113               ACTGAAGCACCAACAGAGTCTCCC--G--G--TAGGACC----C----------GGGG----G---GT-A----AACCC-GAGA---A-A--AT-A------T-A-AG--C---GAC---C-CT-G...-  5810
H2B.D205_X61240             GCAGAAGCAGCCCCAGAGATCCCT-----GA--G-GAAC-----A--A--A--GTGGG-A--G----T-GGG---A-CC-G-----AA-A---C-A-----CT-A-AG------GAT---C-CT-G...-  6341
H2B.EHO_U27200              GCAGAAGCAGTCCCAGAGATTCCT-----G--TA--AAC-----A--A--A--GTGGG-AC-G---GT-G-G--CG-CC-G-----AA-A--AC-A-----CT-A-AG------GAC--CC-CT-A...T  6339
H2G.ABT96_AF208027          GCAGAAGAAATCCCCCCA.........-----TG-G-CC--C-----A------TGGG--------GT-G----AG--C-G--A--AA-C--AG-A-----CT-A-A-------GAC---C-CT-G...-  5684
H2U.FR.96.12034             GCA............GAAATACCT-----G---GGG-CC-----A--A-----TGGGG-C--G---GTGAG---AA-AC-G-----A--A--AG-A--------A--G------GAT---C-CT-G...-  5828

MAC.239_M33262              ............GAAGAAAGACCT------A-TG----A-----A-----A---TGGG--------GT-G-G---G--C-G--A--A--G--AG-A------T-A-A-------GAT---C-CT-G...-  6268
Vpr                         _.__.__.__.__E__E__R__P__P__E__N__E__G__P__Q__R__E__P__W__D__E__W__V__V__E__V__L__E__E__L__K__E__E__A__L__K__H__F__D__P__R__L__._  Vpr

SMM.SL.92.SL92B             ............GAA...CATGCT-----G--TG-GACCAAC-CA--A--A---TGGG----G----T-AG---TG-AC-G--A--AT-G--AG-A--G--AC-G-A-------GAC---C-CT-G...-  5701
SMM.US.x.PGM53              ............ACAGAAAGACCT--------TG---CC-----------A--GTGGG--------GT-G--A-AG--C-G-----AG-A--AG-A------C-G-A---C---GAT---C-CT-G...-  6187
MNE.US.-.MNE027             ............GAAGAAAGACCT--------TG----C-----A-----A---TGGG--------GT-G-G---G--C-G-----A--G--AG-A------T-A-A-------GAT---C-CT-G...-  5736
RCM.NG.x.NG411              CTCATTGCAGAGATGGAGATGCTC--G-----TG---AA--------A--A--CT-TG--------CT-A-G--TACCC-TATA---A-AC-AG-A------AAG-A-------ACA---GAAT-G...-  5499
RCM.GAB1_AF382829           CTCTTTACAGAGATGGAGCTGCCC--T--G--TG---T--G--AC-A--A--TT-TG--------TT-A-G--TACCC--ATA---T-GC-AG-A------AAG-A-------ACATA-GCTT-G...-  5464
DRL.FAO_AY159321            TGTGTCTTACAGATGGAAAGGCCA--G--G--TG---A--G-CA-----A--CT-TG----G---CT-G----TAC-----TA--AA-AC-AG-G------T-G-AG------GATAGGC-CT-A...T  5567
MND-1.F17                   GTATCA.....................-----T-----A--T-CC--A------T-----C-G---CT-GC---TAC-A-G-----AA-A---G-A-----AAGA-AG--C--C...---...C-CATTA  5353
MND-2.5440_AY159322         TCTGTTTTACAGATGGAACAGCCA--T--G--TG-G-CT--C-----A-GA---T-T---A-----CTGG---GTAC---G-CA--AA-AC-AG-G------T-A-AG-----CGATAGGC-CT-G...-  5513
MND-2.14CG_AF328295         TCTGTCTTGCAGATGGAACAGCCG-----G--TG-G-CT-----A--A--A---T-T-----G---CTGA----TACA----CA--AA--C-AG-A------T-G-AG------GATAGGC-CT-A...T  5875
MND-2.MND2_AF367411         TCTGTGTTACAGATGGAGCAACCA--C--G---G-G-CC--T--A--A--A---T-----------CT-A-T--CACC----CA--AG--C-AG-A------T-A-AG------GATAGAC-CT-A...T  5948
DEB.CM.CM40                 TACCCACCCTCACACCCA......---C-TTT-ACCTCAAG-ACA-TACCCATGACA-GGCT-G--TTG-AGC--GCCA-GC-A--T--A--CG-G-----CC-A-A------C...A-C--AGAAGAA-  5457
DEB.CM.CM5                  TATCCACCCTCACACCCA......---C-TTT-AC-TCAAG-ACAGTACCAATGACA-GGCT-GCACT--A-C-AGCCA-GC----T--A--CG-G------T-A-A---C--C...A-A--AGAAGA-T  5451
MON.CM.L1                   ................................................CCCAT-G-A-GAC-GGCCCT--AG-CCGCCA-TTG------A--CG-A-----CC-A-A---C--C...T-A---GAAGA--  5474
MON.NG1                     .........................................CGT-GTTCCCAT-G-GCGC-CCGCTCTG-AG-CCGCCA--TG---C--A--CG-G--G--CC--TYG--C--C...T-A-A-GAAGA--  4085
TAL.CM.8023                 .........CCACCCAGTCATCCT-TGCC-ATAAC-TCAAGG-TC-TACCAA---CA-GA-CTCAATT-GA-AGA--CA-GAGA--AG-A-G-G-C--GATGC-C-T---C--C...A-A--AGAAGA-G  5145
TAL.CM.266                  .........CCACCCAGTCATCCT-T-CCGCTGACCTCCAG--TC-TTCCAA---CA-GGCCTCA-CT-GA-AG---CA-GAGA--AG--CGAG-C---ATGC---T---C--C...A-A--AGACGA-G  5647
MUS.CM.1085                 ........................G-GCCCTGG--CTCCAG-GT-GTCCCAA-CAC--TGCGCCA-G---A---AGCCA-GTG---AG-A--CG-A--G--A-AA-A---C---...T-C--AGAAGA--  5422
SAB.SN.-.SAB1C              GAAGAGATACCAGGATGG......-T----ACATGG-ATTTG-CC--A-----GTT-G-C------CTC-G---CA-G---C----C--C--TTCA--G--CCAGT-C--C--C...--A----A-CTC-  6100
GRV.ET.x.GRI_677            ........................-TG---AT-TGG-AT-T-G-C---------TGGG-C------CT--A---CA-GC-CAG---T--A--CG-A-----CAGA--G--C---...GGA-T--ACAT--  5827
TAN.UG.x.TAN1               GAAAGAAGACCCGGATGGTTAGAAATTTGG--T-TCA-C......--------GTGGG-C-----ATT-AG---CA-GG---CA--A--C--TC-G-----ACAAC-C------...GGG--AGAACTC-  5883
VER.DE.-.AGM3               GAAGCAAGACCGGGGGAA......-T----ATATGG-ATTTGAGC--------GTGGG-C------CT-AG---CA-GC----A--TA-C--TC-G-----CAAG-TG------...GGGC-CGAGCTC-  5388
VER.KE.-.9063               GAAAACAGGCCAGGAGAA......GT---GAT-TGG-AT-T-A-C--A--A---TGGG-C------CT-AGG--CA-G-----A--T--C--TC-G-----CAAATT-------...GGC--AGAGCTC-  5890
VER.KE.-.AGM155             GAGGAGAGACCAGGAGGA......-T----ATTTGG-AT-TCAGC--------GTGGG-C------CTCAG---TA-GG----A--AA-C--C--C-----CAAACTG------...GGCC-AGAACTC-  5885
VER.KE.-.TYO1               GAGGCAAGACCAGGAGAA......GT----ATTTGG-AT-T-AGC--------GTGGG-CA-----CT-AG---TA-G---C-A--T--C--CC-G--G--GAGGCTG------...GGG--AGAACTC-  5380
SUN.GA.98.L14               CCAGTAGAACAGCCACCA.........------G-G-------C---A------TTTGCAC-----CTGGC---TACCA-G--A--AA-C-G-G-C-----G-AA--G-----C...--AGCAGCAATAG  6044
LST.CD.88.447               AGTCCAGAAAAACCTCCA.........--G---TTT-------CA---------TGGGG-C-----CT-GC---CACCA-G-----AA-A--AG-G--G--CAGAC-C-----C...--A...GAAACAA  4916
LST.CD.88.485               AGTCCAGAAAGACCTCCA.........--G---TTT--A----CA--A------TGGGG-C-----TT-GC---CACCA-G-----AA-A---G-G--G--CAGAC-C-----C...--A...GAAGCCA  4916
LST.CD.88.524               AGCCCAGAAAGGCCCCCA.........--G---TTT--T--T-CA--A------TGGGC-C-----CT-GC---CACAA-G-----AA-C---G-A--G--AAGAC-C-----C...--A...GAAGCAA  4913
LST.KE.-.lho7               AGTCCAGAGAGACCACCA.........--G---TTT--A----CA---------TGGGG-C-G---TT-GC---CACCA-G-----AA-A--AG-G--G--AAGAC-C-----C...--A...GAAGTAA  5997
SYK.KE.-.SYK173             ACACCATGGTTCTTCATTCCA.........AGAA-T-TAGAGTT--CCCCA...A--GTGATCAATGTCACT-TAAAGGC---ACT-G-AGT--CA-----ATCC-AG------...A--CCTCAAGAAA  5857
SYK.KE.KE51_AY523867        GAACCACTATTTTACATTCCA.........AGAA-T-TCAATTG-GAC---G--GGAGTAAT-A-TG-CACT-CCAAAGC---AC--G-AGCCC-A-----ATGC-AG------...A-AGC-GAAGAAA  5518
COL.CM.-.CGU1               GGGACTTCACA....................................................---TTG-AG--C-AGG-----A-CA-A-CCCGG--G--CAGA--G--C--CCCTTTG-CTCACATCT  5373
```

```
H1B.FR.83.HXB2             TCCATGGCTTAGGGCAACATATCTATGAAACTTATGGGGAT........................ACTTGGGCAGGAGTG...GAAGCCATAATAAGAATTCTGCAACAACTGCTGTTTATCCAT..TTT  5774
Normal Vpr                 L__H__G__L__G__Q__H__I__Y__E__T__Y__G__D__.__.__.__.__.__.__.__.__T__W__A__G__V__.__E__A__I__I__R__I__L__Q__Q__L__L__F__I__H_..__F   Vpr
HXB2 Vpr                   L__H__G__L__G__Q__H__I__Y__E__T__Y__G__D__.__.__.__.__.__.__.__.__T__W__A__G__V__.__E__A__I__I__R__I__L__Q__Q__L__L__F__I__H_.._F_   Vpr
                                                                                                                  frameshift (extra T) ^
                                                                                                                  causes premature end in HXB2

H1D.CD.84.84ZR085          -T---A-------A---T-----------------------........................--C-----------T...-----T-----------------------A--------T---...--  5295
H1F1.BE.93.VI850           -T--C--------C------------A-C--C--------C........................-------AG-----T...-----T--------G--AT----------A------------...--  5113
H1G.SE.93.SE6165           -T-----------A--G---------A-C------------........................-------A------T...-----------------A--A--------A------------...--  5170
H1H.CF.90.056              -----CAA-----A--G---------A-C--C-----A---........................-------T------T...-----TT---------CG-----------A--------T---...--  5120
H1J.SE.93.SE7887           -T---A-T-----A--GT--------AGC--------A---........................--A----A------C...-----------------A-----------AT-------T---...--  5087
H1K.CM.96.MP535            -----AATC-G--A--G---------AC-------------........................-------A----C-T...-----T-----------AT----------A--------T---...--  4969
H101_AE.TH.90.CM240        ----A--------A--GT-C------A-C-A----------........................-------A---G--T...-----T-----------GT----------A------G-T---...--  5341
H102_AG.NG.-.IBNG          -------A-----A-----------------A---------........................-------A---G--T...A----T------------T----------AT-----G-T---...--  5298
H103_AB.RU.97.KAL153_2     -T---A-------A---T-C-----------------A---........................---------------...--------------------------------------T---...--  4999
H104_cpx.CY.94.CY032       --------C----A--G---------A-C--------A---........................--C----A---G--T...-----T------------T----------A--------T---...--  5139
H1N.CM.95.YBF30            -A-----G-----A-----C------A-C--A-----A--C........................--C----AG--G--A...--G--A--T--C--G--A--A------T-A------------...-A  5367
H1O.BE.87.ANT70            -A--C-C---G--A---T-C--T-----G-----------C........................-------T------T...ATG--A--T--------CT-A------T---------C----...-A  5828
H1O.CM.91.MVP5180          -A--A-C--GT-----GT-C--T-----G--------A--C........................-------A------T...ATG--A--T--------CT-A--------A-------C----...-A  5803
CPZ.CM.04.EK505            -A-----AC--------T-CT-----A-T--A-----A---........................--C----AG-----A...--G-----C--T--G-CA--A---------T-------A---...-A  5345
CPZ.CM.04.LB7              ----CCAA--G--A---T----A-----C--C-----A--C........................--C----AG-----T...C----T--C--------C-----------C--------T---...-A  5328
CPZ.CM.04.MB66             -----CAGC----A--G---------ACC--------A---........................-----------G---...--G--T-----------C------------T-------T---...-A  5319
CPZ.MT145                  -A---CAC-----AG-------T--------A-----A--C........................T-C-----G-----A...C----T------------T--------T-A------G-----...--  5315
CPZ.CD.-.ANT               -A--ACA-C-T--AA-TTGGG-A----C--A---C--A---........................T-A---AG-------...C-G-TTT-G----TTC-C--T---A--GCT-------CA---...-A  5230
CPZ.CM.-.CAM3              -A-----A-----A--GTT---T-----C--------A---........................--A--------T--A...-----A--C--T-----A--------G--T------------...--  5187
CPZ.CM.98.CAM5             -A---A-AC----A--GT-C-----------------A---........................--A---AT---G--A...-----A--T--T-----AT-------T--T------G-----...--  5475
CPZ.GAB_X52154             -G---CAA-----A---TTC--T-----C--------A---........................--A----T------A...--G-----C--------CT-A-----CT--------------...--  5837
CPZ.GAB2                   -A---CAG-----A---TTA-----CA-T-A------A---........................--A----A------A...--G--------C-----A--T-----CT--------G-G---...--  5166
CPZ.TZ.-.TAN1              -A--G--GG----AA-TTGGG---TCACC-T------A--C........................T-C----AG-----A...C-G-AAT----C-AG--CT----GAG-GCTT------C---C...-A  5435
CPZ.US.85.CPZUS            -A--A--------A--GT-C-----CAGT--------A--C........................-------AG-----A...-----A--T-----G--CT-A--------TT-----------...-A  5833
GSN.99CM166_AF468659       ---G---AA-TT----TG--G-GACG--GCTCCC--CC---...CCAGATTGGACAGTGGATCAGG-AGCAAT--C-TGTGCAAT--ATTAC--C--G...G-----AC---T--C----GA--C...--  5587
GSN.99CM71_AF468658        ---G---AA-TT----TG--G-AACG--GCTCCC--CC---...CCGGATTGGACAGTGGATCAGG-AGCAAT--C-TGTGCAAT--ATTATG-C--G-GAG----GAC-T-A--C----GA--C...--  5572
H2A.ALI_AF082339           -AAT--CTC-T--CGGCT--------ACT-GAC----A--C...........................ACCCTT-A-AGAGCCAGG-AGC-C--T-ATGCC--A----G-GCC--C--C--G--C...--  6458
H2A.BEN_M30502             -AAT--CTC-T--CT-CT-----C--A-T-GAC----A--C...........................ACCCTT-A--GCGCCAG--AGC-C--T-A--CC--A----G-GCC--C--CG-G--C...--  6468
H2A.ST_M31113              -AAT-ACTC-T--CA-CT---------CT-GAC----A--C...........................ACCCTT-A--GCGCCAG--GGC-C--T--G--C--A----G-GCC--CC-CT-G--C...--  5910
H2B.D205_X61240            -AAC--CGC-T--TA-CTT------CAGT-GGC----A---...........................ACCCTT-C--GAGCA-G--AGC-C--T-A---C--C----G-GCC--C--CC----C...--  6441
H2B.EHO_U27200             -AAC--CAC-T--CA-TTT------CA-T-GGC----AA--...........................ACCCTT-A--GAGCA-G--AGC-C--T-A-C-C--C--G-GGGCA--C--CC----C...--  6439
H2G.ABT96_AF208027         -AAC--CGC-T--TA-CT----------T-GGC----A--C...........................ACCCTC-AG-GAGCA-G--AAC-C--T-A---C-----GAG-GCT--C--CC-G--C...--  5784
H2U.FR.96.12034            -AAC--CGC-T--TA-CT--------A-T-GGC----A---...........................ACCCTT-A--GAGCA-G--AGC-C--T-A---C------AG-GCAA-C--CC----C...--  5928

                                                         / Tat exon1 start
MAC.239_M33262             -AAC--CAC-T--TA-T---------A-T-GAC----A--C...........................ACCCTT-AG-GAGCA-G--AAC-C--T-----C--C----G-GC---C--C--G---...--  6368

Tat                                                     _M__E__T__.__.__.__.__.__.__.__.__.__P__L__R__E__Q__E__N__S__L__E__S__S__N__E__R__S__S__C__I__.__   Tat ex1
Vpr                        L__T__A__L__G__N__H__I__Y__N__R__H__G__D__.__.__.__.__.__.__.__.__T__L__E__G__A__G__E__L__I__R__I__L__Q__R__A__L__F__M__H__.__F   Vpr

SMM.SL.92.SL92B            -TAC--CT-----AA-TT-CG-A-----T-----C------...........................ACAAT--A--G-GCA-GG-AA-----T-AG--AT-A---A-GGCAT----CT-A--C...--  5801
SMM.US.x.PGM53             -AAC--CGC-T--TA-CT----------T-GAC----A--C...........................ACCCTT-A--GAGCA-G--AAC-C--T--G--C--C----G-GC---C--C--A--C...--  6287
MNE.US.-.MNE027            -AAC--CGC-T--TA-TT----------T-GAC----A--C...........................ACCCTT-AG-GAGCA-G--AAC-C--T-A---C--C----GGGC---C--C--G--C...--  5836
RCM.NG.x.NG411             -TACACAAGA---AA-TT----A-----GCAGC----A---...........................A-TTT--AG-GAGTAA----T--G----CCT-G--TA-TA--GCA--T--CT-G--C...--  5599
RCM.GAB1_AF382829          -AACGCAAA----AG-TT--G-A-----GCAAC--------...........................A-CAT--AG-GAGTGC-G--A--G--TC-GC-G--T---AG-GCCT-----C-T---...--  5564
DRL.FAO_AY159321           -G---AATG-----TC-TGGG---------GA-----A---...........................-CATT--A--GAGTGC-CCAGC-T----T-C-CT-----AG-GCA--T--CT-G---...--  5667
MND-1.F17                  ---TAAATGC--TATC-G-ATAT-G--TG.....................CAAAACACAGGGAGTGAGGAA-AG-CCTGT...--GAAAT-T--T-CCT-AA--A-TAG-GCCA-T-GGG----C...C-  5456
MND-2.5440_AY159322        -A----CAG----CTC-TGGG-G-----GCAAC-A--A--C...........................ACTTT--A--GTGTCC--CAGC-------C---T-----AG-GCA-----CT-A--C...--  5613
MND-2.14CG_AF328295        -G----CAG----AAGCTGG--T--------A--------C...........................ACCTTG-AG-G-GTGC-GAAGC-T--C-CCC-CT-A--GAG-GCT--A---C-G---...--  5975
MND-2.MND2_AF367411        -G----CAG----AAGCTGG-----CA----C---------...........................ACATT--A--G-GTGCG-CAGC-T--CTC------T---A--GCT----C---G--C...-A  6048
DEB.CM.CM40                -ATT---AG-TT--...A--CAT-G--T-GAC-TACCC-CCCAACCAGATTGGACGGGAGAACAAG-A-----T-C-TC-GTCATT-A-TAC----A---AG-T---A--A--T-A-GGT-G--C...C-  5581
DEB.CM.CM5                 -ATGG--AG-CT--...A--CA--G--T-GACCT-CCT-CCCAACCAGATTGGACAGGAGAACAAG-A-----G-CGTCAGTAATT-A-TAT----A---AG-A---AG-A----T-GGC-T---...C-  5575
MON.CM.L1                  ---G----A-TT--G----GG-AACG--GCTGCC--CT--CCCGGCC...TGGAATGCAGACCAGG-A-------C-TGTGCCATT-ATTAT-CT---TGGG-C---ACCA-C--T-A--GG--C...-A  5598
MON.NG1                    ---G----A-CT--G-T--GG--AC----TTGCC--CC--CCCAGGG...TGGAACGCAGACCAGG-A-------C-TGYGCTATT-ATTAC-CC--GCGGG-----AC-A-RA-A----GG--C...A-  4209
TAL.CM.8023                --AT---AG-TT--A-T--CTGTGTA--GTTACCA-CA--GCCTGAT...TGGACAGGCGAACAAG-A-----T-CTTCAGTCAT--ATT-TGCTC-G-AGGGTA-CG-GA--T-A--CC-A---...--  5269
TAL.CM.266                 --AT---AG-CT--A-T---TGTGTA---TTGCCA-CA--CCCTGAT...TGGACAGGAGAACAGG-A-----T-CCTCAGTCAT--ATT-TGCT--G-AGGGTA-T---A--T-AC-CC-A---...--  5771
MUS.CM.1085                ---G---AA-TT--A-TG-GG-AACG--GCTGCCA-CC---CCTGAT...TGGACAGTGGATCAGG-GGCAAT--C-TGTGCCAT--A-TATG-T--G-GAG-----AC---TG-C----GG---...--  5546
SAB.SN.-.SAB1C             -TTTCC-TC-TT--TGGA----TGTG---GAACCA-CCAT-.........GATCATGGGCAAACC-GGCT--A---CTG-...T-TAAATATTGT------------A-GGCTT-A---G-G---...A-  6215
GRV.ET.x.GRI_677           -AATCC-AG--T--A-T...TA--G--T-GAGG-G--AAGGAGA............CATAATACCC-A---AAT-AGA-AGGCT-CAAGTACTAT------G-T---A-GTCTA----G-A----...--  5939
TAN.UG.x.TAN1              --TTCCAAG--T--A-C...T---G-C-GGAAG-A-----AAGG.........AATGGGGCACCA-TGATA-A-A-G-CT...T-TAGATATTAT--GT-GG-A--GA-GGCTT-A---G-G---...--  5995
VER.DE.-.AGM3              -GTTCCAAG--T--A-C...TAT-G-C-GGAGG-G--A--AAGG.........AATCGCACTCCC-TGCTA-A-A-G-CT...T-TAAATATTAT-A-T-GG-----A--GCT--C---G-G---...--  5500
VER.KE.-.9063              --TTCCAAG-TT--A-T...TAT-G-C-GGAGG-A--A--GAGA.........CGGGGAGCCCCA-TGAC--AGA-G-CC...T-TAGATACTAT--GT-GG-T--GA--GCA--C---G-T---...--  6002
VER.KE.-.AGM155            -AT-CCAAG--T--A-T...TAT-G-C-GGAGG-A-----GAGA.........CAGGGAAGACCC-TAGC--A-A-G-CA...T-TAAGTATTATC-CT-AG-T--GA--GCT--C---G-G---...--  5997
VER.KE.-.TYO1              -GTTCCAAG--T--A-C...TA--G-C-GGAGG-A--A--AAGA.........CATGGTACTCCC-TGAT--A-A-G-CC...T-CAAATATTAT--GC-AG-A---A-GGCT--C---G-G---...--  5492
SUN.GA.98.L14              -TGCC...CA--TCTTG--ATA--G--CG.....................GACTCTACAGGGTCAGAAAA--A---GTGT...ATGAGAGC---C-CTT-G--CA-TAG-GCCT----C--T---...C-  6144
LST.CD.88.447              -TTTG-C-CA--TCTC-G-ATA--G--TC.....................AGCAGTACCGGGTCAGAGGA-AA--CCTGC...ATGAAAT-C--T-CC--AG-AA-TAG-GCA--C-----T---...--  5019
LST.CD.88.485              -TTTG-C-CA--TCTCGG-ATA--G--TC.....................AGCAGTACCGGGTCAGAGGA--A--CCTGC...ATGAGAT-C--T-CC--AG-AA-TAG-GCA--C-----T---...--  5019
LST.CD.88.524              -TTTG-C-CA--TCTCGG-ATA--G--TC.....................AGCAGTACCGGGTCAGAAGA--AG-CTTGC...ATGAGAT-C--C-CTT-AG-TA-TAG-GCT--C-A-T-A---...C-  5016
LST.KE.-.lho7              -GTTG-C-CA--TCTCG--ATAT-G--T-.....................GACAGTACAGGGTCAGAAGA--AG-CCTGC...ATGAAAT-C--C-CCC-AG-CA-TAG-GCTT-----C-T--C...C-  6100
SYK.KE.-.SYK173            -AT--...GG--TTTGGA--CAA-CCTT--ACG-A-A--CA...GGGACTGATTCCCCGACAATGG------A-A--ACCATGTTG-A---GG-----GCA--CA-T-TTA----C---GAA---...--  5978
SYK.KE.KE51_AY523867       -ATT-...GG--TTTGGA--CAA-GCTT-GAGGTG-AA-CAGGTCCAGAAGAAACACCCACCATGG-----ATTA--TGCATGCT--AT--GCAT---GCA--CA-CTTTA----T---...G-ACA---  5642
COL.CM.-.CGU1              CAGTGATAA-GT--......GGT-G-ACT-GAAGG............CACTGG...GGGAGACCGTGG...A-G-CT-GT...CTG--TTATG-G-----CA--G-GTGTGCC--C-ACT-G--C...G-  5473
```

```
                                                                                                                    HIV-1 Vpr normal end \
H1B.FR.83.HXB2            CAGAATTGGGTGTCGACATAGCAGA...ATAGGCGTT...............ACTCGA..................CAGAGGAGAGCAAGAAATGGAGCCAGTAGATCCTAG..................  5850
                                                                                                                      / Tat exon 1 start
Tat ex1                                                                                                              _M__E__P__V__D__P__R__.__.__.__.__.__.  Tat ex1
Normal Vpr                __R__I__G__C__R__H__S__R__.__I__G__V__.__.__.__.__.__T__R__.__.__.__.__.__.__Q__R__R__A__R__N__G__A__S__R__S__*                    Vpr
HXB2 Vpr                  _Q__N__W__V__S__T__*

H1D.CD.84.84ZR085         --------------A----------...------A--...............----C-..................-----A---------------T-------------A..................  5371
H1F1.BE.93.VI850          ------------C--T---------...------A--...............GT--C-..................-----A----T-------------------------..................  5189
H1G.SE.93.SE6165          ------------C-A----------...------A--...............--A-C-..................-G---A--G-T---GG-----C--G----------A..................  5246
H1H.CF.90.056             ------------C-A----------...-----AA--...............------..................-----A----T----------C-------------A..................  5196
H1J.SE.93.SE7887          ------C-----C-AT---------...------A--...............-TA-C-..................--A--A----GG--G--------------------A..................  5163
H1K.CM.96.MP535           ------------C-AC---------...------A--...............-TA-C-..................-----A----GG---------T-------------A..................  5045
H101_AE.TH.90.CM240       --------------A----------...------A--...............-TG-C-..................GG---A----GC--G------A-TG----------A..................  5417
H102_AG.NG.-.IBNG         --------------A----------...------A--...............-T--A-..................GG---A----GC--G--------TG-----------..................  5374
H103_AB.RU.97.KAL153_2    ---G----------A----------...------A--...............-T--A-..................-G----------------------------------..................  5075
H104_cpx.CY.94.CY032      ------------C-A------T---...------A--...............----CT..................--A------AG-G-C-GG---TGG--CCAG-AG.....................  5212
H1N.CM.95.YBF30           T--G-----C--C-AG--C------...-----GA-C...............----CT..................--A------AGG-AT..----A--------------..................  5440
H1O.BE.87.ANT70           T--------A--C-A------T---...-----AA--...............-AC-C-...............AGAGGA--AG--AG----------T------------GA..................  5907
H1O.CM.91.MVP5180         T--------A--C-A------T---...-----AA--...............CTC-C-...TCTAACACAAGAGGAAGAG-A---AG--AT..----T------------GA..................  5891
CPZ.CM.04.EK505           T--G-----C----A---------G...-----AA-C...............----CT..................--A..---GAG---G-------------------GA..................  5418
CPZ.CM.04.LB7             ---------C----A----------...------A--...............--A-C-..................--A--A---AGG---G---------A----------..................  5404
CPZ.CM.04.MB66            --------------AT---------...------A-C...............-A-ATT............AATCAG-----A---AG--------------A---------A..................  5401
CPZ.MT145                 ---------C--C-AT---------...------A-A...............-AC-CT..................-----A---AGG---G-----A---A----G---GA..................  5391
CPZ.CD.-.ANT              T---CAC--A---GCT--------G...-----ACA-...............GGA-CC.......................................CGT--AC-C...-GA..................  5282
CPZ.CM.-.CAM3             T--GC----C----A----------...-----GA-C...............-TC-CT..................GG------GAG--AT..-----------------GA..................  5260
CPZ.CM.98.CAM5            T--GC----C----A----------...-----GA--...............CT--YT..................G----A--GAG--ATGGAACCAGT--ATCY-AA.....................  5548
CPZ.GAB_X52154            T---C-------C-AG---------...------A-C...............CTC-C-...............CAAAGA--A---T-C-ATGGATCCAAT--ATCC-GA.....................  5913
CPZ.GAB2                  T--G-----C----A---------G...-----GA--...............GAA---..................---------AG---G-GA-...................................  5225
CPZ.TZ.-.TAN1             TC-CCA---T---ATC--C---...AGA-----ATCA.............................................................................................  5469
CPZ.US.85.CPZUS           T--G-----C---AAT--------G...-----GA--TCCTTGGCAAGGAGA----CT..................--AG-----AGG-AT..----T-------------A..................  5921
GSN.99CM166_AF468659      T-A-GA---A--CTTT--C--GGTCAAT...TTATCA............GGG-GATACCCAGCAATTAGGCCCTCAAG-G-C-CC--CCC-CCA-ACAGT-ACTCTGTT-CTCATGCCGATCCTGAGCAG  5702
GSN.99CM71_AF468658       T---GA-------TTC--C--ATACAAT-G-ATT--C............AGA-GGTACCCAGTAATCAGGCCCTTAAG-G-T-CC--CCC-CCA-ATAGT-ACTCTGTTCCTCATGCTGATCCTGAGCAG  5690
H2A.ALI_AF082339          ----GCA--A---G-C-G-TCA---...G-T---CAAACAAGGGGAAGA......AATCCTCTCTCAGCTATACCGACCCCT---AAC-TGC-ATA-.................................  6546
H2A.BEN_M30502            ----GCG--A---AAC-GCTCA---...--T---TAAACAAGGAGAAGA......ACTCCTTGCCCAGCTGCACCGACCCCT-----C-TGC-CTA-.................................  6556
H2A.ST_M31113             ----GCA--A--CG-C-GCTCA--G...--T--TCAGCCCAGGGGACGA......AATCCTTTATCAGCTATACCAACCCCT----GC-TGCGATA-.................................  5998
H2B.D205_X61240           ----GCC--T----A---CTCA--G...--T--ACAATCAGGGGGAGGA......AATCCTCTCTCAACTATACCG-CCCCTTA--GC-TGCGATA-.................................  6529
H2B.EHO_U27200            ----GGG--T--C-A---CTCA--G...--T--ACAACCGGGGGGAGGA......AATCCTCTCTCAGCTATACCG-CCTCTTA--GC-TGC-GTA-.................................  6527
H2G.ABT96_AF208027        TC--GGA-AC---A--TGCTC---G...--T--A-GGAACGCAGGCGGA......AATCCTCTATCAACTATACCG-CCTCT----GCGTGTTCTA-.................................  5872
H2U.FR.96.12034           ----GGG--A-----G--CTCA--G...--C---AGCAGCAGGGGAGGA......AATCCTCTCACAACTGTACCG-CCTCTT---GCGTGCTCTA-.................................  6016

                                                                                                                Vpr end \
MAC.239_M33262            ----GGC--A--CATC--CTC----...--C---CAACCTGGGGGAGGA......AATCCTCTCTCAGCTATACCG-CCTCT---AGC-TGCTATA-.................................  6456

Tat ex1                   S__E__A__D__A__S__T__P__E__.__S__A__N__L__G__E__E__.__.__I__L__S__Q__L__Y__R__P__L__E__A__C__Y__N__.__.__.__.__.__.__.__.__.__.__.  Tat ex1
Vpr                       __R__G__G__C__I__H__S__R__.__I__G__Q__P__G__G__G__.__.__N__P__L__S__A__I__P__P__S__R__S__M__L__*__.__.__.__.__.__.__.__.__.__.__._  Vpr

SMM.SL.92.SL92B           T--GCA---A--CACC---TCA--G...--T--ACAGCCAAGGGGACCA......AATCCCCTGGGATCTATACCGTCCGCT----ACGTGTTATA-.................................  5889
SMM.US.x.PGM53            ----GGA--A---GCT--CTC----...--C---CAAAGTAGAGGAGGA......AATCCTCTGTCAACTATACCG-CCTCT----GCGTGCTCTA-.................................  6375
MNE.US.-.MNE027           ----GGC--C--CAAC--CTCT---...--C---CAATCTGGAGGAGGA......AATCCTCTCTCAACTATACCG-CCTCT---AGC-TGCTATA-.................................  5924
RCM.NG.x.NG411            ---GCA---A---GA-GGG-----G...--T--A-GAGCGAGAGGAGGA......AATCCTCTTAGATCCATTCTT--CTC----AAC-TCTTGTA-.................................  5687
RCM.GAB1_AF382829         T----A---A---GC-GGG--T--G...--T--AACATCCAGAGGA......-G-AATCCTCTACGATCCATTCCG--A-C----AAC-TC-TGTA-.................................  5652
DRL.FAO_AY159321          T---TA---A--C-AGG-A--T---...C-T--A...CCAGCAGGACGC...TGGAACCCATTTCGAACCTTTCCAAG-GC--ACCA-CA-TTGTA-.................................  5755
MND-1.F17                 AGCTCAA------GATGGA-C-TTCAGGGA-......AGAAGA...CCACAACTG-CCCCCTCAGGATTCAGGCCAAGAG-AGATAG-TT-......TAA..............................  5541
MND-2.5440_AY159322       T---TA---A---A-GGGA--TC-C...--C--ACAAGAAGGAGGAAGA...TA-AACCCACTTCGATCCTTTCCGAGCCC--ACAACCCCTTGTA-.................................  5704
MND-2.14CG_AF328295       T---CA---A--CA-GG-A---C--...--T--ACAAGAAGGAGGCAAA...TA-AACCCCCTTAGATCCTTTCCAAG-CC--ACAACCCCTTGTA-.................................  6066
MND-2.MND2_AF367411       T---TA---A--CA-GG-A---C-T...--C--ACAAGAAGGAGGGGGA...TG...TTACCCCTTACGATCCTTT-C---ATCG-AC-ACCCCCTGTGA..............................  6139
DEB.CM.CM40               G---GAA-CT--CTTT--C--GGAGAGAGAG-CTACCAGAGGGTATCCTAAC-T-A-GCCATTGACAGGAAGAAATAGAGAAGTTAGGGACGGA-A-TAACC-TTCA---TTGCTTCAG...........  5700
DEB.CM.CM5                G---GAA-CA--CTTC-----AGAGAGGGA--CTAC-AGGAGATACCCCAAT-T-A--CCCCTAACAGGAAGAAACAG-GA-GTCAGGGATGGA-A-TAACC-CTC-TT--CATCATAG...........  5694
MON.CM.L1                 ---GGAA--C---TAC-----ATATGCAGAGCAAA-AAGAAGATACCCTGTTCT-A-GCCAATGAGAGGGACTGCT-CCG-CCC-A-C--TTC--TTC-ACAGGCCGATCCAGACAATCCCAGAAGACCT  5728
MON.NG1                   G--GGA-------TAT--C--GCATGCCCG--TGR--CGCAGGTACCCTYGRCTG---CCACTAAGAGGCACTGCC-CAG-ACCCA-C--TTC--TCC-TCA-GC-GA-CCAGAGAGGCCTTTGCACCCG  4339
TAL.CM.8023               -CAGCAA-----CTTT---GAGCAGCAA-G-A-AAC-GCCCACTACCCCAAC-TA---CCCCTTTCCGGCAGGAGG...-AC-CCATGCA-T-A....................................  5360
TAL.CM.266                -CAGCAA-----CTAT---GAGCAGCAA-G-A-AAC-CCGCACTACCCCAAT-TC---CCCCTTTCAGGAAGAGGGG------CCATGCA-T-A....................................  5865
MUS.CM.1085               ----GAC--A--CTAT-----GTATAAC-G--TAA-GAGAAGATATCCAACC-TAA-GCCTTTGAGAGGCACACAA-CCCCTCCTT-T-ATTCAACCC-A-A-GCTGA-CCTGTACCTCCTTTGAGTCCT  5676
SAB.SN.-.SAB1C            G-A-GG-A-A---T-TA-G.........CC-AAGAC-.........CACCCTG-GTATGGACCAGGAGCAGGAGGC-C-CCCCC--GTCTGGGA-----TTCAG--GGAGCTGCATCGGCCGCTCCAGGC  6327
GRV.ET.x.GRI_677          ----TG---T---A--AGG--AG--...CCTTTTTCCCCTTACGAAGAGAGG-GAAATGGACAAGGAGGAGGAGCC-CACCCCCTC-TCC-GGACTT--ATAG...........................  6039
TAN.UG.x.TAN1             TC--TG---C--CA-GAGA--A-CT...CCTTTT-AGCCCTACGAAGAAAGG-GAAATGGAGTCGGAGGGGGACGGG-TG-C----AGCCTCC-CC--GACT-GC--AG.....................  6101
VER.DE.-.AGM3             -C-GTG---A--C--CAGA--ACA-...CCCTTT-AA...............C-ATAC..................G-AGAA--GAGGGATGGACA--GGG-AG--CGAGCAGGGCGCGTACCACCAGGA  5594
VER.KE.-.9063             -C-TTG---T-----TAGA--ACAG...CCCTT--AA...............C--TAC..................G--GAA---AG-GATGGACA--GGG-AG--AGAGC-AACCGTGCTCCACCAGGA  6096
VER.KE.-.AGM155           -C-GTG---A-----CAGG--ACA-...CCCTTT-AG...............C-ATAC..................G--GA---GAG--ATGGACA--GGG-AG--AGACCAGGACGTGTCCCACCAGGA  6091
VER.KE.-.TYO1             TC--TG------CA-GAGA--GCAG...CCCTTT-AA...............C-ATAC..................G--GA---GAG-GATGGACA--GGG-AG-CAGAGCAAATCGTGTCCCACCAGGA  5586
SUN.GA.98.L14             GCC-...AT-----CTTC-TA--TCAGA--CA-AAG-GGATCAGGAAATGTC-AC-C-AGGCCACCAACAAGACCAAGACCAG-GCA-G-GG-CATTAGA--AG-CCTAC-AGACTAA............  6259
LST.CD.88.447             GCCT...ACC----CCAGCTATGC-...GC--CAAGA......GCAGGAACAG---ACCCGACAGAAGCAGCACCAAGACCC-CTAA----G-CAT-CAA--AG-CCGT--CTAA...............  5122
LST.CD.88.485             GCCT...ACC----CCAGCTATGC-...GC-ACAAGA......GCAGGAACAG---ACCCGACAGAAGCAGCACCAA-ACCC-CT-A----G--AT-CAA--AG-CCGT--CTAA...............  5122
LST.CD.88.524             GCCT...CAA----CCAG-TATGC-...GC--C-AGA......GCAGGAACAG-A-ACCCAACAGAAGCAGCACCTAGACCA-CT-G----G--AT-CAA--AG-CCAT--CTGA...............  5119
LST.KE.-.lho7             -CC-...ACT----CCAGCTATGC-...GC--C-AGA......GCAGGAACAG-C-ATCCAACAGAAGCAGCACCAAGACCT-AT-A-G--G-CAT-CAAC-AG-CAGTG-CTAA...............  6203
SYK.KE.-.SYK173           TGCTGCA--A----CT--AC-G-CC...-G-TAT-CC...............-GA-AC..................AGAG--TACC-CCATCC-TCTTAG..............................  6042
SYK.KE.KE51_AY523867      TGCTGCA--A----C---G--G-C-AAG...TA--CGAGACACAGGGGATACC-A-ATCCTTCTTAG...............................................................  5706
COL.CM.-.CGU1             TCATGCG---CAAT-T-GCTTTCCC...TCTACGCAAAACCCTTAG....................................................................................  5516
```

```
H1B.FR.83.HXB2           .................................................................ACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAAACTGCTTGT...ACCAATTGCTATTGTA  5912
Tat ex1                  _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__L__E__P__W__K__H__P__G__S__Q__P__K__T__A__C__.__T__N__C__Y__C__  Tat ex1

H1D.CD.84.84ZR085        .................................................................CA----T--------C-------------------GG---------...-A------T-------  5433
H1F1.BE.93.VI850         .................................................................CT----T--------T-----------C-------C----C-----...-----A--T-------  5251
H1G.SE.93.SE6165         .................................................................C--------------T-----G--G---------------C-C---...-A---A--T-T-----  5308
H1H.CF.90.056            .................................................................---------------C------------------C-----------...-A------T-------  5258
H1J.SE.93.SE7887         .................................................................CAG------------T------------------------------...--T-----T-------  5225
H1K.CM.96.MP535          .................................................................CA-------------C--A---------------------------...-A-C-G----------  5107
H101_AE.TH.90.CM240      .................................................................C--------------T-----G-------------C----------...-G---G--T--C----  5479
H102_AG.NG.-.IBNG        .................................................................C--------------C--C--G-------------C----------...-G---G--T--C----  5436
H103_AB.RU.97.KAL153_2   .................................................................---------------------------C------------------...----------------  5137
H104_cpx.CY.94.CY032     ..........................................................ATCCTGAC--------------C-----G-------------C-----A----...-A---G--T-TC----  5281
H1N.CM.95.YBF30          .................................................................-T-------------T-----------C--A--------A-----C...-AT-------------  5502
H1O.BE.87.ANT70          .................................................................GG-GCCC--T---C-C--C--T-----------CC-G-TCC-----...-AT-----------C-  5969
H1O.CM.91.MVP5180        .................................................................GA-GCCC--T---C-T--C--T--G--CA----CC----CC-----...-AT-----------C-  5953
CPZ.CM.04.EK505          .................................................................-T-------------T--C-----------A-----G--A-----C...-AT------C------  5480
CPZ.CM.04.LB7            .................................................................C-----T--------T------------------------C-C---...-A--------T-----  5466
CPZ.CM.04.MB66           .................................................................TT-------------T------------------------------...-AT-------------  5463
CPZ.MT145                .................................................................CT----A--A-----C--------G--CA-A----G---A-AA--C...-AT--G----T---C-  5453
CPZ.CD.-.ANT             .................................................................-AC-CCT--A---TTA--C--TCCTGC-ACA--AGC---GC-A--C...-AT--C--T--C----  5344
CPZ.CM.-.CAM3            .................................................................T------------CTA-----------C-------G---------C...-AT--C----------  5322
CPZ.CM.98.CAM5           .................................................................T------------CTA-----------------------------C...-AT--C--T-------  5610
CPZ.GAB_X52154           .................................................................C--G-----------A--C-----C--------A-G---A-T----...-AT-----------C-  5975
CPZ.GAB2                 .............................................ATGGATCCTATAGATCCTAGT--T-----------T-----------------A-----A-----C...-ATTCA-----C--C-  5307
CPZ.TZ.-.TAN1            ..............................................TGAATCCCATAGATCCTCAGG---CA--A---G-A---------GC-GCA---G----AC-----...--A--C--T--C----  5550
CPZ.US.85.CPZUS          .................................................................CA----------------G--------C--------G--------C...-AT--------C----  5983
GSN.99CM166_AF468659     CCTAGGAGACCCTCAAGATATCGCATGGAT.........GAATAG.....AGACCCGGAAGAGTGT--CTT----AACTG---G-AGCC-G-AGC-G-GCCCG-CA-CCCGTGCT-TGC------C--C-  5818
GSN.99CM71_AF468658      CCCAGAAGACCTTCAAGATACCGCATGGAT.........GAATAG.....AGACCCGGCAGAGTGT--TTT---TAA-TG---A-AGCC-G-AGCCG-GCC--ACA-CCCGTGCT-TGC---------C-  5806
H2A.ALI_AF082339         ....................................................................................................................--CA----------  6560
H2A.BEN_M30502           ....................................................................................................................---A----------  6570
H2A.ST_M31113            ....................................................................................................................---A--T--C----  6012
H2B.D205_X61240          ....................................................................................................................T-CA-----C----  6543
H2B.EHO_U27200           ....................................................................................................................T-CA-----C----  6541
H2G.ABT96_AF208027       ....................................................................................................................T--A--T-----C-  5886
H2U.FR.96.12034          ....................................................................................................................T--A----------  6030

MAC.239_M33262           ....................................................................................................................--CA----------  6470
Tat ex1                  _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__C__Y__C_   Tat ex1

SMM.SL.92.SL92B          ....................................................................................................................--GC--T-T-----  5903
SMM.US.x.PGM53           ....................................................................................................................T-CA--T--C----  6389
MNE.US.-.MNE027          ....................................................................................................................--CG--------C-  5938
RCM.NG.x.NG411           ....................................................................................................................T-GC--T-----C-  5701
RCM.GAB1_AF382829        ....................................................................................................................--C---T-------  5666
DRL.FAO_AY159321         ....................................................................................................................T-CA----GG--C-  5769
MND-1.F17                ...................................GCAACTGCTA..................................G-AG-GTATTATCAG-CTTTGC--G--TG-...GAG-AT--A---GG--C-  5600
MND-2.5440_AY159322      ....................................................................................................................T--G----TC----  5718
MND-2.14CG_AF328295      ....................................................................................................................---A--T-----C-  6080
MND-2.MND2_AF367411      ....................................................................................................................--CA--T--C----  6153
DEB.CM.CM40              .................................ATGGAGAGCATA.................................G-T-CATTT-A-GAAAGAG----T--AC-CCAC...C--GC---TATG--C-  5761
DEB.CM.CM5               .................................ATGGAAGAAATA.................................G-T-CATTCAAGGAAACAG----T--AC-CCA-...C-T-----TATG--C-  5755
MON.CM.L1                TCCAGATATAGAATGGATGAATGA.........................TGGAACCTGTTGACCC-GACTTA---AAAG-A--A-ACCCCCCAGCAA-CCC-CGCA-ACC-TGT--T--C----T-----  5833
MON.NG1                  TCCAGATACCACCACGATGAATAG.......................TGCAASGATAGTGAACCC-GC-TTA---CCCG----G-ATCCCTCAGC-A-ACCGCTCA-CCCGTGT-A---C----T---C-  4446
TAL.CM.8023              .......................................................................................................................G--T--C--C-  5371
TAL.CM.266               .......................................................................................................................A--------C-  5876
MUS.CM.1085              TCTAGGTACAGGATGGATGAATAA.........................TGGATCCATCAGTAGAGGAGTT---AAA-G-A--A-G-CCGG--GCCG-ACCGG-CA-GCC-TGT--ATC-----TC----  5781
SAB.SN.-.SAB1C           CTGTGA...........................................................................................................T-AT-C-----TC--C-  6350
GRV.ET.x.GRI_677         ...............................................................................................---ATT-CAAC-A---ACT-AT--A--------C-  6074
TAN.UG.x.TAN1            ...............................................................................................----TT---AC-A---ACT-A---A----T---C-  6136
VER.DE.-.AGM3            CTTGATTGA..........................................................................A-A-CTC-AAGCA--CCTG-AGCGG---ACA-A---G----------  5650
VER.KE.-.9063            CTTGATTAG..........................................................................A-A-TAT-AGA----GCT---AA-----AGA-AT--G----T---C-  6152
VER.KE.-.AGM155          CTTGATTAA..........................................................................A-A-TAC--GA-A--CCTTGAGA-A---ACA-AT--A----T---C-  6147
VER.KE.-.TYO1            CTTGAGTGA..........................................................................AGACTATCAGA-----CTGCAGA-----AAA-AT--A--T-T---C-  5642
SUN.GA.98.L14            ..................................................................................................CTTGGAG--C---GAT-A---A----GG----  6291
LST.CD.88.447            ............................................................................................G-CA-TG-GTG-CTG-...AAC-A---G----GC----  5157
LST.CD.88.485            ............................................................................................G-CA-TG-GTGTCTG-...GAC-A---G----GC--C-  5157
LST.CD.88.524            ............................................................................................--CA-TCC--T-CTGC...CAG-AT--G---C-C--C-  5154
LST.KE.-.lho7            ............................................................................................AGCC-TGC--G-GTG-...CAG-AT-CC--T-GC----  6238
SYK.KE.-.SYK173          .........................................................................AAG-A-CATTC-T--AG-AAGGA--ACC---AC-A---...-AT--A----T---C-  6096
SYK.KE.KE51_AY523867     .........................................................................AAG-A-CATT--TCTC--A-GG----C-G---C-----...-G--GA--------C-  5760
COL.CM.-.CGU1            .......................................................................................................AG--G---AATGA---A----GG----  5543
```

```
                                                                                  /Rev exon 1 start
H1B.FR.83.HXB2              AAAAGTGTTGCTTTCATTGCCAAGTTTGTTTCATAACAAAAGCCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGA.....CAGCGACGAAGAGCTCATCAGAAC.........AGTCAGACTCATCA  6027
Tat ex1                     K__K__C__C__F__H__C__Q__V__C__F__I__T__K__A__L__G__I__S__Y__G__R__K__K__R__R__.__.__Q__R__R__R__A__H__Q__N__.__.__.__S__Q__T__H__Q  Tat ex1

Rev ex1                                                                             _M__A__G__R__S__G__D__.__.__S__D__E__E__L__I__R__T__.__.__.__V__R__L__I__  Rev ex1
H1D.CD.84.84ZR085           ------------A---------------C--------G----G----------------------------------.....------------C---C---C.........AGC--------------  5548
H1F1.BE.93.VI850            --CGA------------------TGG--C--T-C---G--G-G----------------------------------.....-------AC---A---C---A.........AGC-----AGTG-----  5366
H1G.SE.93.SE6165            --GT--------GG---------------C--TC-G-AC-----G--------------------------------AG.....--C-----G--A---C----.........AGC---A-AGGC-----  5423
H1H.CF.90.056               -------C----A-----------A-G--C--TT---AG----G------A--T-----------------------C.....-------AC---A---C-GCA.........AGTTTG--AGA------  5373
H1J.SE.93.SE7887            ------------A------T-----G--C--TC--CAG--G-G---------T------------------------.....---------TC-----C--CT.........GGC---A----------  5340
H1K.CM.96.MP535             --CG--------A-----------A-A--C--TT---AG--G-G---------T---A--------------A-----.....-CC-------C-A---C-T-C.........AAC---G---AC-----  5222
H101_AE.TH.90.CM240         ----A-------GG----------C-A--C--TC-G-A-----G---------------------------------AG.....--C-----G--A---C----.........AGC---A--GA------  5594
H102_AG.NG.-.IBNG           ---T---C----G.----------T-G--C--TC-G-AC--G-G---------------------------------.....-GC-----G--A---C----.........AGCC-----GA------  5550
H103_AB.RU.97.KAL153_2      ----------TC---------------------G-A-----G-----------------------------------.....---------------C---A.........GAC-A-------G----  5252
H104_cpx.CY.94.CY032        ------------GG---------------C--TC-G-A-----G---------------------------------A-.....--T-----G--T---T---A.........GGC--CA--GGC-----  5396
H1N.CM.95.YBF30             ---GA-------A---C---TT-TA---C----C--AG----G------------A----------------------T.....--A---------A---C----.........AGC---A---G------  5617
H1O.BE.87.ANT70             ---GA--C----A-------T-T---------G---G---G-GT--G--A---------------------------AG--.....-GA-C-GCTGCT--A...AGC............CA--CAGA----A-  6078
H1O.CM.91.MVP5180           ---GA--C----A-------T-T----------C--AG--G-GT--G--A-----C---------------------A---.....AGA-C-GC-GCT---GCAAGC............TA--CAGA-A--A-  6065
CPZ.CM.04.EK505             --GTT-------A---C--TGTGTA---C----C--A-----G------------A----------------------T.....--A---------A---C-----G-.........-A-A---G------  5595
CPZ.CM.04.LB7               ------------A---C--T-C-T----C----C--AG----G------A--T-----------------T-----.....TC-----------A--TC-A---G-.........----C-CA-T----  5581
CPZ.CM.04.MB66              --C-A--C----AC--C--T---C----C----CG-AG----G------T--T-------------G--------A-.....--------------A--TC-G---G-.........-------A------  5578
CPZ.MT145                   ----------T-A------TAT-C----C--T-C--AG----G----------------------------------AG-G.....AGA---AA-CCT---TCGAGC--T.........CAGG-T...A--A-  5565
CPZ.CD.-.ANT                --TGC-----T-AC-------CTC-C--C----C--A-----G---G--G--T-------------------GAA-AGCTCGA...AG-AACA--C--A-AAC-GCTG-A......AGC---G-A-A-A----  5465
CPZ.CM.-.CAM3               ---T---C------------AT---------TGAG-A-----G------A---------------------------A--G.....GGA------C-CTTGTC-GCTGCA......AGC---AC-CA-A----  5440
CPZ.CM.98.CAM5              --TGT--C------------ATG-----C--TGAG-A-----G------A-----A---------------------A--G.....GGA---A--CACT-G-T-GCTTCA......AGC---AC-CA-A----  5728
CPZ.GAB_X52154              --GCT-----T-A---C--TAT-TA---C----C--A-----G------A---------------------------A-CCACA...AGA-----C-CG----C-GCAGG-.........---A-A-A-A----  6093
CPZ.GAB2                    ---G---C----A------TATGC----C----CC-GC----G------T--T------------------------G--.....-GAA------CG----T-A--GG-.........---A-A-GCA----  5422
CPZ.TZ.-.TAN1               ----A--C--------------C------C--T-CG-A------A-----A-------------------GAA-AG--.....-GCAA-TCTGCT-TA...--C-GT...........AC--A-A----  5659
CPZ.US.85.CPZUS             ----------------C--TGTG--------T-CG-AG--G-G------A--T------------------------G-GCGA...-GCA------C-A--GC-A---G-.........---GC-GA-A----  5945
GSN.99CM166_AF468659        -G------GC-C----C--T--GT-G--C--TC---GG--G-GAC----TC-T---CTCCATG-C-GATCA--C-AGAGGAAA-GA-C-AA-GCTA-AG-A-CTG-T...CCTAGCGTGTC-T--A----  5945
GSN.99CM71_AF468658         ---G----GCT---------T--GT-G-----TC---GG--G-GAC----TC-T---CT-CATG-C-GATCC--C-A-CGGAAGAGATC-GA-GCTT-AG-A-CTGCT...CCTAGCGTG-C----A----  5933
H2A.ALI_AF082339            -GG-A--C----AC--------GC-G-----TT---AT--G-GAC-C--G--A-GG----A-C-A---GGCA--C--.....AGAA--TCTCC-AAGA-A.................-T-A--G-----TC  6669
H2A.BEN_M30502              -GCGA--CA-T--C--------GC-G------TCG-A---G-GGC-C--A--A--A----AGC-A---GGCA-AC--.....AGAA-GACTCC-AGGA-A.................-C-A------C-TC  6679
H2A.ST_M31113               -------C----AC--------GA-G-----TT---AC--G-GGC-C--G--A-GG----AAC-A---GGCA-A---.....AGAA--ACTCCGAAGA-A.................-C-A--G-----TC  6121
H2B.D205_X61240             -G--A--C----AC--------GC-------TC-T-A---G-GTC-T--G--A-GT----A-C-CTC--GAA-----.....TCTGC-AA-------A-G.......................---AC-GC  6646
H2B.EHO_U27200              ----A----C--AC--------GC----C--TC-T-A---G-G-C-G--G--A-GT----AAC-CTC--GAAA----.....TCTTC-AA---G---A-G.......................---AC-AC  6644
H2G.ABT96_AF208027          ----A--C----A------T--GT-A--C--TC-T-AC--G-GAC-C--AG-A-GT----AGCAACCT-GAA-A---...AGAACT-CGAA--AGA--A-G.......................G--A--TC  5992
H2U.FR.96.12034             ----A-------A---------GC-A-----TC-T-A---G-G-C-T--G--A-GT----AGCAACCC-GAA-A--G.....AGT-C-AAG-AG-T-A-G.......................G--T--TC  6133

MAC.239_M33262              ------------AC--------GT-------TC-T-A----G---G--G--A-GT----AGCAATCACGAAA----.....AGAA--ACTCCGAAAA-G...............GC-A--G--A--AC  6579
Tat                         K__K__C__C__Y__H__C__Q__F__C__F__L__K__K__G__L__G__I__C__Y__E__Q__S__R__K__R__.__.__R__R__T__P__K__K__.__.__.__.__.__A__K__A__N__T  Tat ex1

Rev                                                                                 _M__S__N__H__E__R__D__.__.__E__E__L__R__K__R__.__.__.__.__.__L__R__L__I__  Rev ex1
SMM.SL.92.SL92B             -------------C--------GC-------TC-T-GG--G-G-C-T-----TA-T---CA-AACCC-GAA-A---.....GTTAAGAAG-AG---A-G.......................G--A---C  6006
SMM.US.x.PGM53              ----A--C--T-AC--------GC-------TC-T-A---G-G---G--G--A-GT----AGCA-C--CGCA-----.....AGAA--ACTCCGAAGA-G...............GC-A--G--A--AC  6498
MNE.US.-.MNE027             ------------AC--------GT-------TC-T-A----G---G--G--A-GT----AGCA-TCACGCA-A---.....AGAA--ACTCCGAAGA-G...............GT-A--G--A--AC  6047
RCM.NG.x.NG411              ---GA---GTG-A---C--T--GC----C--TC-TCAG--G--T--------AAA-----CTTCT-G--GAA-A---.....AGAG-CA--CC--GAA--A----G...AAGAATCAGATT-A--CATT  5822
RCM.GAB1_AF382829           ----A-----T-A---C------C-C--C---C-GCAG---GAC-------AAAT----CTTCC-G-GCAA-AC--AGAAGATCTAAGGA-GA-AA-A-G.....................GCTGA-A-ATT  5778
DRL.FAO_AY159321            --T-C--CGCT-AC--C--T--GC----C--TC-TCA-----G------G--ACGT---AATGTC-GC-G--CC---.....A--A--ACCGCCA-CGTGGTT......TCAAGAGTATCTGAGAC-GG  5887
MND-1.F17                   -G--A--C-----------TATGC----C--TCA--AG--G-GT-----A--AAGG--CCATGTCT-C-G-AAAC-T.............GT-CC--GAAC-......AATAAGAAG-TA-CTGG-AG-GG  5712
MND-2.5440_AY159322         --GTT-------A------T--GC-G--C---T-GCAG-----T-----GG-ACAT---CATGTCT-C-GA-A---G.....-GA-CT---C---GG-T--T-......GAGAAAGTA-CA-GCA--AG  5836
MND-2.14CG_AF328295         -G--A-------A---C-----GTG---C---T-GCAG-----T-----G--ACAT---CATGTCT-C-GAACC--G.....AGAACT--G-AGAGA-T--T-......GGAAAA-TATCACAGG--AG  6198
MND-2.MND2_AF367411         ---GA-------A---C-----GC-A--C---T-GCAG--G--------AG-ACAT---CATGCCT-C-GAAAC--G.....AGATCTA--CAGAGA-T--T-......GAAAAA-TATCAGAAG--AG  6271
DEB.CM.CM40                 --GCC--CATTC-A--------GC-G-----T---CA---G-GTC----G--T---------CCACGC-G-A---AGAGGAAGTGCA--A--GA-TACACG-CAGCTGCTGAAAGT-A----C--G----  5891
DEB.CM.CM5                  --GCC---ACTA-G--------GC-G--C--T--GCA-----GTC----G---A-------TCACGC-G-A---AGAGG.....A----CG--T-AAGCTTTGCT...GCGAATGA-A--GA------  5876
MON.CM.L1                   G-GTT-----------C--T--GA----C---C-C-------G-C-C--G--T--A...TATG-C-G---A----AGAGACGAAGAGCGACGTCTC--GTC...CCT...GGCCTT--CAGC-G-A-GA-  5954
MON.NG1                     -G-TT---GC-C-G--CY-T--G-----C---C-CCA-----GAC-T--A--A------------GGCCGAAA---GCGCCGCGGT-C-A--CATC--ACG......ATTGGCGATC-GAGCTA---GG-  4570
TAL.CM.8023                 ----A-----T-A---C-----GC----C---T-GCAG--G-GT-----T---CGG...TATGTCCCTG-CA-----ATCCGGA-AA-GA-----T-CAC-TCT-G-.........--CA-TCAGA-GG-  5489
TAL.CM.266                  -G--A-----T-AC--C-----GC----C---T-GCAG--G-GT-----T---CGG...TATGTCCCTG-CA----GTTCCGGA--A-GA-----T-CAC--CT-G-.........--CA-TCAGG-GG-  5994
MUS.CM.1085                 --------GCT---------GT-C-C-----TCA--AG----A-----T--T------------CGTTTCA-A-A-CGATCGAGAGCTGATTCGAGC................CTGCCGAGCCA-TCAG-  5896
SAB.SN.-.SAB1C              --GTT--C-----------TATTC-C--C---CAT-A---G---C-------TCG---TATGTCCCT-G--CA---.....AGAGCTTCT-AGAAG.......................ATTT-G---A-  6453
GRV.ET.x.GRI_677            ----A--C--T-A---C--TG-GC----C---C-GCAG--G-G------AG-GCGT---CATGTCTCT-G-AAA---.....AGAAA-AC---CA----A.........................GA-A----  6177
TAN.UG.x.TAN1               ---GA--C----A---C--T--------C--TC-GCAG--G-G----------A-----CATGTCTCT-G-ATC---.....AGA-CTAAG-AGAA----......................TCA-A------  6242
VER.DE.-.AGM3               --TGT--C--T-A---C--T--GC----C--TT--CA---G-GA-----TG-TA-----CATGCCCCT-G-ATC---.....AGAAAGAAG-TT----CG.........................CT-G---G  5753
VER.KE.-.9063               ------------A---C--T--GC-------TC-GCA-----G-C----TG-TA-----CATGCCCCT-G-ACC---.....AGAAAGAA---T-T---G.........................C--A---G  6255
VER.KE.-.AGM155             ----A--C--T-A----------T-C--C---T--CGG---GAC----T--TA-----CATGCCTTT-G-ACC---.....AGAAAGAAG-TT---TCG.........................G--G---G  6250
VER.KE.-.TYO1               ----A-----T-AC--C--T--GC----C--TC-GCA-----G-C----TG-TA-----CATGCCCCT-G-ACC---.....AGAAAGAAG-TTCG-TCG.........................CT-A--TT  5745
SUN.GA.98.L14               GG-GA--C--------C--T---C----C--TC-GCAG---GG-----G---CA----TATGTCT-C-G-G-AC--.....----AT-A-TC-ATATC-......................A--GA-A-G--  6397
LST.CD.88.447               -------------------TATTC----C--TCAGCAG--G-GA-----G--ACGA---TATGTCC-C-G-A-----.....A-AA--GTCGCCCAGGTATTT.....................GAAG--A-  5263
LST.CD.88.485               ---G---------------TATTC----C--TCAGCAG--G-GA-----G--ACGA---TATGTCC-C-G-A-----.....AGAA--GTT-CCCAGGTATTT.....................GAAG--A-  5263
LST.CD.88.524               --GTT-----------C--TAT-C----C---CAGCAG--G--G-----G--ACGG---TATGTCC-C-G-ACC---.....AG-A--GTT-CCTACATA-TT.....................GAAAC-G-  5260
LST.KE.-.lho7               ----A--------------TATTC----C--TCAGCAG---GG-----GC-GCA---TATGTCC-C-G-AAAC-G.....AGA---GTT-CCT-GATA-TT.....................GAGAC---  6344
SYK.KE.-.SYK173             ----C------A---------GC--------C-TCA--G-GA--G--A---A-T----C---ACCA-G-AA-C--.....GCAGCT--C--TAT-......................TCTGAGG--G-  6199
SYK.KE.KE51_AY523867        ------------A------TT-TAGA--C--TC-TCA-----G------G--AA-A----T-C--G---GAAA----AGACACGCAA--AC--CT---..............GAGGACTTTG--GCG--  5875
COL.CM.-.CGU1               -GGCC-----------C--T--GC-A--C--TC-TT------GGC-G--A---CG---......------AAT-AG.........A-TGCT-C-CAA................................  5626
```

```
                          Tat, Rev exon 1 \/ intron                                       / Vpu start (ACG start codon in HXB2)
H1B.FR.83.HXB2            AGCTTCTCTATCAAAGCAGTAAGTA.......................GTACATGTAACGCAACCT...ATACCAATAGTAGCAATAGTAGCATTAGTAGTAGCAATAATAATAGCA...ATAGTTGTG  6127
Tat ex1                   __A__S__L__S__K__Q
Rev ex1                   K__L__L__Y__Q__S__
Vpu                                                                      _T__Q__P__.__I__P__I__V__A__I__V__A__L__V__V__A__I__I__I__A__.__I__V__V_   Vpu

H1D.CD.84.84ZR085         --A-C--A--C--------------.................----------T----T--...T---A----T--------------------T----GC-----------...---------  5648
H1F1.BE.93.VI850          -AA-------C--------------GTGTTA.............AATA---T------T-TC-TA-...T-GTT-GC-A---GG----C-------A--------C----------C..---------  5476
H1G.SE.93.SE6165          --A-C--G--C---------G----GAAAAA.............ATTA---T------T---GT-A...T--GT----T-----------G----A--------T-C---GC---C..---------  5533
H1H.CF.90.056             -AA----A-----------------.................T--TCATA--T-T-TATA...T--GG-T---G-AT-GG--CGCT-G-----AC-TTT--C---GCC.....G--A-----  5470
H1J.SE.93.SE7887          --A-CT-A--C--------------AATCAA..............GTA-C-T------T-AT----...T-G-A-----C---------------TA----------TT-CC-T---..----G-A--  5449
H1K.CM.96.MP535           --A-C-----CG--------G----GTGCTA.............AGTA----------T-GTGT-C...T--G----TTC-AT-G---C-TT-G------C-CT----T--GC-.........---A----A  5329
H101_AE.TH.90.CM240       -AA-C--A--C--------------.................ATAA---T------T-AC----...T-GGA---TAGT----------G-C-GA-------GC----CT-----..-----A---  5698
H102_AG.NG.-.IBNG         -AA-C--G--C---------G----GTAATA.............ATTA-C-T----G-T-------...T--A-----ACT----------G-C-----------T-C---GC---C..---------  5660
H103_AB.RU.97.KAL153_2    --T-------C--------------.........................---CATGT-ATG...CA-T-CT---C-AT-GC--C-AT-G---C-T---T-G--G--GG-AT-..----CAA-A  5346
H104_cpx.CY.94.CY032      -AA-CT-A--C--------------.................TTAA---T------T-TT-TTC...TGGGA---CTGG----------G-C-G--------GC----T----T-..-----A---  5500
H1N.CM.95.YBF30           --A-CT-A--C--G-----------...............AAACCT---T--..............--G-TGTC-T-G-G-T-CA----G----G--C----G-T-GC------..G--A-A--C  5711
H1O.BE.87.ANT70           --A-C--G--C--------------........................ACGC---T---T-A-...-GGGACC-GC--------A--ATTA-TAGT-CTTTGC-GT-T---AAT..G--A--T-A  6175
H1O.CM.91.MVP5180         --A-C--G--C--G-----------........................ACGC-G-T---T-AA...GAGAACC-GC----CT--A----T---AGT-CTTTGTGTC-T---AAT..G--C--A-A  6162
CPZ.CM.04.EK505           --A-C-----C----------C--.................................A--G--ATATATATGT-GT-GCTT---AAGTT-GG-T-CA---G-T--GC-----AG...-CTC-AA-A  5685
CPZ.CM.04.LB7             --A-C--A-----------------........................AC-CCT-TAAGTTG-AGTCA-TAT--G-ACCT-----A-TT-GG-T--A---GCT--G-T----A-...C--A-CA-A  5680
CPZ.CM.04.MB66            --A-C--G--C---------G---..AATTC........ATGCAGATATATGC-TAGG-TTT-ATGGAT---GTGCAGCA--T-GGCT--TT-G----TC--ATT---GA-T---TT..-----CA-A  5694
CPZ.MT145                 G-A-C-----CG-------------.................................--AAGGCTCACATGCAGC---A---T--GCT-A--A--T--TTC---GCTT--ATG..C-T--A-CT  5654
CPZ.CD.-.ANT              --A-C--G--------T--------.........................-TCT---T-ACTAA-...---TTT.................GAGTAT-CTTTTC-TGCCT-TAGT..-----AC-A  5544
CPZ.CM.-.CAM3             --G-------CAGC-----------..................ACTTC-----CT-T--T-A--...TGGGA-CA-A---GCT--A-T-GTA---G-A---A---T--T-----T..--T--A-CC  5543
CPZ.CM.98.CAM5            --A----A--CG-C-----------..................ATTT-G----CT-T--T-AT-...TGGGAGCA-A---GTT--A----T----G-A---A-C-T--T--T-TT..--T--A--T  5831
CPZ.GAB_X52154            --A----A--C--------------.............................T-T-ACT-TG..TTAGT-.GGCT---TTC-CA--TT-G---GCT--...--TGC-TGGAAC..--TTG-A-A  6182
CPZ.GAB2                  --A-C-----CG-C------G----......................TCCC-TCT-T-T--TTT-CATGTGGGT-GCTA---G-C--A-T-G-A---GTACTCTTT--G-T---AAT..-----G---  5526
CPZ.TZ.-.TAN1             --A-C--G--CG-C-----------CCCATG...........ATAAAAA--GTA--GGGAAGTGTGTCA-CTAATG-CA---GC--TC-TTGTA--T--C-GATTT-----GGG-G-...GGCT-GC-A  5774
CPZ.US.85.CPZUS           --A-C-----CGC-----------GCCCTTTTG....................--CT-T--TTAA-...TGGTTTGA-A---GTT--A---------G-A---A-GGT--TT---T-...G-GA-AA-C  6204
GSN.99CM166_AF468659      --AC---GC--AGGTC------..................................GTAAGGC--G-ATCCT-C----G-TTGGTGG-GG-GT-CT--T--T----CCTTCATATATC--TGT  6034
GSN.99CM71_AF468658       --ACC--A--CA--TC------..................................GTAAGGC--GAGTGCT-C---TC-CTGGTGG-GG-GG-CT--TG------CCTTTATATATT-CTGC  6022
H2A.ALI_AF082339          GT-----GC----G-CA---G---.........................................................................................................  6693
H2A.BEN_M30502            GC-----GC-C--G-CA---G---.........................................................................................................  6703
H2A.ST_M31113             GT-----GC----G-CA---G---.........................................................................................................  6145
H2B.D205_X61240           -C-----GC-C--G-CA---G---.........................................................................................................  6670
H2B.EHO_U27200            -T-----GC-C----CG---G---.........................................................................................................  6668
H2G.ABT96_AF208027        -TT----GC------CG-------.........................................................................................................  6016
H2U.FR.96.12034           -T--G---C------CA-------.........................................................................................................  6157

MAC.239_M33262            -T-----GC------CA-------.........................................................................................................  6603
Tat                       __S__S__A__S__N__K
Rev                       H__L__L__H__Q__T__

SMM.SL.92.SL92B           -TT-GA-AC------C--------.........................................................................................................  6030
SMM.US.x.PGM53            -TT----GC-C----CG-------.........................................................................................................  6522
MNE.US.-.MNE027           -T-----GC------CA-------.........................................................................................................  6071
RCM.NG.x.NG411            T-A-A-ATCTGAGC--T-A..............................................................................................................  5841
RCM.GAB1_AF382829         TC--GTA-CTAATC--T-A..............................................................................................................  5797
DRL.FAO_AY159321          T-ACA-G-CTGTGGCAAG----...........................................................................................................  5909
MND-1.F17                 T-AAGAAGCTAT-CGAAG----...........................................................................................................  5734
MND-2.5440_AY159322       TCAAG-AGCT-TGGGAAG----...........................................................................................................  5858
MND-2.14CG_AF328295       T-AAG-AGCTGTGGGAAG----...........................................................................................................  6220
MND-2.MND2_AF367411       T--AG-G-CTGTGGGAAG----...........................................................................................................  6293
DEB.CM.CM40               --A--T-G--C--G---------C-AATCTGCATGTAAATGCTAAAGGGAACAT-A-G-G--TTATACAGC-TTTGC--C-TATTT---GGTT-A----GA---GT-C-CAGGGGAA-CAG-A-AAGCAA  6021
DEB.CM.CM5                --AG-T-A--C--G---------...GTAAAGCTGCATATGTATATAGTATC---ATGCTTAGCTTA-TAATA-TA-TAGCA--TCTG-CCCTGG-AG-GGCAGCACCA-CA--GCAACAATGGGTGACA-  6003
MON.CM.L1                 TC--G-AAG-AAGC--TGA......................................................................................................GTAA--AA  5981
MON.NG1                   TAA-C-GGG-GTC---T-A...................................................................................................GTGAT-AATAC  4600
TAL.CM.8023               TAAGGA--CTGT-TCAA--C-GTA-........................................................................................................  5514
TAL.CM.266                TAAGGA--CTGT-TCAA--C-GTA-........................................................................................................  6019
MUS.CM.1085               TC--GTA-A-ATC---T-A....................................................................................................GTAGCA--AA  5925
SAB.SN.-.SAB1C            TCAAGT-TCTCT-C-CA-C--GTA.........................................................................................................  6477
GRV.ET.x.GRI_677          --ACC--A--CGGC----------.........................................................................................................  6201
TAN.UG.x.TAN1             GAA-CT-G-----C----------.........................................................................................................  6266
VER.DE.-.AGM3             CTT-C--GA-CA---A----G---.........................................................................................................  5777
VER.KE.-.9063             CCTA----A-CAGG-C--------.........................................................................................................  6279
VER.KE.-.AGM155           CAT-C--G--C-GC----------.........................................................................................................  6274
VER.KE.-.TYO1             G---C--T--CAGC-C--------.........................................................................................................  5769
SUN.GA.98.L14             --AGG--G--CG--G-T-A..............................................................................................................  6416
LST.CD.88.447             GCAAGA--CTGTGG-CGG-----A.........................................................................................................  5287
LST.CD.88.485             GCAAGA--CTGTGG-CGG-----G.........................................................................................................  5287
LST.CD.88.524             GCAGGA--CTGTGG-CAG----AA.........................................................................................................  5284
LST.KE.-.lho7             GCAGGA--CTGTGG-CAG-----G.........................................................................................................  6368
SYK.KE.-.SYK173           TAGCG-A-CTA--GG-T-A..............................................................................................................  6218
SYK.KE.KE51_AY523867      TAGC--AAC-A-TCG-T-A..............................................................................................................  5894
COL.CM.-.CGU1             ..............-CA--------........................................................................................................  5637
```

```
H1B.FR.83.HXB2          TGGTCCATAGTAATCATAGAA......TAT......AGGAAAATATTA...AGACAAAGAAAAATAGACAGGTTAATTGATAGACTAATAGAA.....................AGAGCAGAAGACAGTG  6221
Vpu                     _W__S__I__V__I__I__E__.__.__Y__.__.__R__K__I__L__.__R__Q__R__K__I__D__R__L__I__D__R__L__I__E__.__.__.__.__.__.__.__R__A__E__D__S__   Vpu

H1D.CD.84.84ZR085       ---A--------T--------......---......C-T-G----AA-...--G---------------T--------------A---G----.....................----A-----------  5742
H1F1.BE.93.VI850        ---A-T------TAT------......---......-A----C-GG--...--G------------A-T-AA---TA-A-A---A---G----.....................----------------  5570
H1G.SE.93.SE6165        ---A--------T-T------......---......---G-----AG-...-A---G-A--G-----GA-A----C--------A---G----.....................----------------  5627
H1H.CF.90.056           ---A--------TAT------......---......-AA---T-GG--...--G---AG-------------C-------A---A--GG----.....................----------------  5564
H1J.SE.93.SE7887        ---A-T-----GTAT------......---......-A----C-G---...--------G--------T-A-------------A---G----.....................----------------  5543
H1K.CM.96.MP535         ---A--------TAT------......---......------T-GG--...-A-----A--G----A--T-------A------A---G----.....................----------------  5423
H101_AE.TH.90.CM240     --AA-T------GCT------......GT-......-A-------C--...--G---------------------G--A-G---A---G----.....................----------------  5792
H102_AG.NG.-.IBNG       ---AG-------TAT------......---......---------AGG...-A---G-AG---------------C--------A---G----.....................-----------T----  5754
H103_AB.RU.97.KAL153_2  GTTGTGGGGTCC--AG--TTC......ATAGAATAT-------------..---------------------------------A---G----.....................----------------  5446
H104_cpx.CY.94.CY032    ---A-TT-----T-T------......---......-A----T-GAG-...--G-----G-G---------C--GTACA-----A---G----.....................----------------  5594
H1N.CM.95.YBF30         ---G-AT--C--TAT-G----......---......-A-------AA-...TTG--GGA-------A-ACACA---GAC-G---A---G----.....................----A------T----  5805
H1O.BE.87.ANT70         ---GGGT-TA-TC-T-G-A--......---......TTAG--CA-AAG...GA----GAC-G--AG--G--AGA---ACT-GAGAGGT--AG-AGAATAAGAGAAATTAGGGATGATAGT--CT-TGAAA  6290
H1O.CM.91.MVP5180       ----TGT-TAACC-T-G-ATT......---.........TT-G-GCA-AGA-A----GAT-G--GG--GCA-GA---ACT-GA-AG-T--AGGAGAATAAAGGAAATCAGGGATGACAGT--CT-TGAAA  6277
CPZ.CM.04.EK505         GTAATAG----TTGGGC-AT-GTATATAGAATATAC--AG--G--AA-GTAGA-G---A--T-TCTC-GCTAAG-CAGAGA-T-AG-GACAG-GCAGAAGATAGTGGCA.....................  5794
CPZ.CM.04.LB7           GTAATAG-----TGGT--A--GCATATC-GCTTTAT-AAG-G-ACA--AGACA-A-GGC--T--ATA-ATTAA-TGAGAGG-T-AG-GA-AG-GGGGAAGACAGTGGCA.....................  5789
CPZ.CM.04.MB66          GTA......A--TGGG--A--GTTTATA-ACTCTGT-AAG-GGATAGGAGGCA-A-G-A--T-GATAGATTAA-TGCAAGA-T-AG-GA-AG-GCAGAAGACAGTGGCAATGAA--C-............  5806
CPZ.MT145               ATA-TT...-CTTGG----C-GCTTATA-AGAGTAT-AA--GT--CA-CAAGT-AGGC---T-GA-AGACTTCAGGACAGA-TTAG--GTAG-GCAGAGGACAGTGGCAATGAA--T-............  5769
CPZ.CD.-.ANT            ---ATT--CTGT--ACCTAT-CTCTATA-GCTTTAT-AA-T-TATAAGCAGCAG----T-G-T-ATA-G--AAATCAAAGA-TCA--GA--TGTTAAGTAGA............---CTTAGTATAGA-A  5662
CPZ.CM.-.CAM3           ---GGG----CTT-T-A---G......-GG......--A---GG-AA-GAAGA-G---AC-G-GC-C-A-AAA-T-GGAC-TT-A--GA-AG-ATAATATCA............---CAG-----T----  5649
CPZ.CM.98.CAM5          ---GGA----CTTAT-A-A-G......-GG......--AG-----AA-GAAGA-AGG-AC-G--CCC-A--AA-TTACA--TT-GC-GA-AG-ATAAGAATA............---CAG-----T----  5937
CPZ.GAB_X52154          ---GGATACA-T--T-A-TGGGGATATAGAAGGTAT-AA-G-CATAGGCTTGAGAC-GA-............A-TGAGAGACTTAACC--AT-TTGAGA...............................  6269
CPZ.GAB2                ---GGT--T---GGA--CTCTGTCTATA-AAGGTGG-AA-G-CACAA-GAAGAG------.........-TCA-TGA-TTG-TTA---A-ATCACAGAA...............................  5616
CPZ.TZ.-.TAN1           ATAGGT----GT--A-G-AG-......G-G.........TT-GA-AGGGAA--G---CATC---G--TATTAGA--GGCTAGCTAG--G-TT-AGCATAGAC............--T-G--T---AGAA-  5877
CPZ.US.85.CPZUS         ---GGAC----TGCT-G-TTG......-GGAGACAA-TA------AA-GAA-ATACTCA-C--...............--G-TT-A--ATCTTTTAGAAAGA............-TCAG-AT-AGAGAA-  6301
GSN.99CM166_AF468659    GTAG-GC-TC-TGCTC-CT-CTTAGCT-GGGATAAGT--GT--A-GG-AAACC-A--CC--C-CA--TAGCAG-----AGACT-..............................................  6118
GSN.99CM71_AF468658     -TAG-A--CT-TGCTC-CT-CTTGGCG-GGGACAAGT---T--AGGG-AAGCC-A-G-T-CC-...-TAGCAG-T---AGGCT-..............................................  6103
H2A.ALI_AF082339        ..................................................................................................................................  6693
H2A.BEN_M30502          ..................................................................................................................................  6703
H2A.ST_M31113           ..................................................................................................................................  6145
H2B.D205_X61240         ..................................................................................................................................  6670
H2B.EHO_U27200          ..................................................................................................................................  6668
H2G.ABT96_AF208027      ..................................................................................................................................  6016
H2U.FR.96.12034         ..................................................................................................................................  6157

MAC.239_M33262          ..................................................................................................................................  6603

SMM.SL.92.SL92B         ..................................................................................................................................  6030
SMM.US.x.PGM53          ..................................................................................................................................  6522
MNE.US.-.MNE027         ..................................................................................................................................  6071
RCM.NG.x.NG411          ................................................................................GTATGG---GTCCTGGTCTGTACCTGCTTTTAGAT-CTT--CTT-T-AG-  5891
RCM.GAB1_AF382829       .......................................................................................G--TGGAGTGCCCTGGACTAGTACTGCTTTTAGAGCCT---CT  5840
DRL.FAO_AY159321        ..................................................................................................................................  5909
MND-1.F17               ..................................................................................................................................  5734
MND-2.5440_AY159322     ..................................................................................................................................  5858
MND-2.14CG_AF328295     ..................................................................................................................................  6220
MND-2.MND2_AF367411     ..................................................................................................................................  6293
DEB.CM.CM40             CAA-A-T-GAC-G-ATATT-TGGAGTACCAGTCTGGGTAG-TGCCAA-GTTGATTTGTTCTGT-C--C--AC-C-TCA--GTC-GG-TGG-C-GTGA....CAGCATGCCTGCC-C..AT-CCTTAGTGA  6145
DEB.CM.CM5              -ATATT--G-AGTG-CCGT-TGGCACAATGCCACTGTT-GCTTAT-CTGTACAG-T--T-GTTCAGA-TCA-GATGGGCTGT---AGCAT-CTTGCCACATGCAATGGT....T----A-C---T-CC..  6127
MON.CM.L1               -TA-TGG-G--C-CTGGTAGCAATAAC-TACAGCCTGATCCTCATAGCCCTCCC-GT--C-GCT-G-G-TT-G-GGCGATATT-TA--ATC-CTAAGAGGTTTAAAAGAATAGACC-AG-A-TACAGAG-  6111
MON.NG1                 CTA-TGGATAGC-G-CATATCTGTTTGGTGTGTTATT-TGCTT-TAGCGCCTTG--CCCT-T-CCTTCTCTA-C-G-CTTATA-GG-GC-C--GCGGGTGTCCCGCTTCGAACAG--CTTTC-A-GGC-T  4730
TAL.CM.8023             ..................................................................................................................................  5514
TAL.CM.266              ..................................................................................................................................  6019
MUS.CM.1085             CTA-TGG--T-T-G--GCTGT.......................................C-TTGTA-----CATC-ACT-T-TGATTGCT-TATTTGCCTTTGTGTTAGCTTACC-AAG-TG-TGTAA-  6016
SAB.SN.-.SAB1C          ..................................................................................................................................  6477
GRV.ET.x.GRI_677        ..................................................................................................................................  6201
TAN.UG.x.TAN1           ..................................................................................................................................  6266
VER.DE.-.AGM3           ..................................................................................................................................  5777
VER.KE.-.9063           ..................................................................................................................................  6279
VER.KE.-.AGM155         ..................................................................................................................................  6274
VER.KE.-.TYO1           ..................................................................................................................................  5769
SUN.GA.98.L14           ..................................................................................................................................  6416
LST.CD.88.447           ..................................................................................................................................  5287
LST.CD.88.485           ..................................................................................................................................  5287
LST.CD.88.524           ..................................................................................................................................  5284
LST.KE.-.lho7           ..................................................................................................................................  6368
SYK.KE.-.SYK173         ..................................................................................................................................  6218
SYK.KE.KE51_AY523867    ..................................................................................................................................  5894
COL.CM.-.CGU1           ..................................................................................................................................  5637
```

```
                                   / Env gp160 start signal peptide
H1B.FR.83.HXB2             GCA.....ATGAGA.................................................................GTGAAGGAGAAATATCAGCACTTGTGG..................AGATGGGG  6265
Vpu                        G__.__._N__E__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__S__E__G__E__I__S__A__L__V__.__.__.__.__.__.__E__M__G_
Env                                _M__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__V__K__E__K__Y__Q__H__L__W__.__.__.__.__.__.__R__W__G sig pept

H1D.CD.84.84ZR085          ---.....------.................................................................-------G--T-A-GAG-A---ATCAA..................CCC...TT  5783
H1F1.BE.93.VI850           ---.....------.................................................................----G--G--TGC-GAG-A-T-G-CA-..................CAC...TT  5611
H1G.SE.93.SE6165           -A-.....------.................................................................----C--G--T-C-GAG-A-T-G--AA..................CAC...TT  5668
H1H.CF.90.056              ---ACGAG-GTG......................................................................-T-----C-C-GAG-A-T-ATCCA..................--C...TT  5605
H1J.SE.93.SE7887           -T-ACGAG-GTG......................................................................-T-----C-C-GA--A-T-G-CA-..................-CC...TT  5584
H1K.CM.96.MP535            --.........-......................................................................-T-AGAGTGAGGGG-ATGCA-A--............AATTGGCAGACATT  5464
H101_AE.TH.90.CM240        -A-.....------.................................................................----------C-C-GAT-A-T-G-CCA..................-AC...TT  5833
H102_AG.NG.-.IBNG          ---.....------.................................................................----T--G--T-C-GA--A-T-ATCCA..................CTC...TT  5795
H103_AB.RU.97.KAL153_2     ---.....------.................................................................----------TCAGGA--...........................CAC...TT  5478
H104_cpx.CY.94.CY032       ---.....------.................................................................----T--G--TGC-GAG-A-T-ATCCA..................CAC...TT  5635
H1N.CM.95.YBF30            ---.....----A-.................................................................----T--G--TGC-GAGTGGT-G-AT-..................G-G...AT  5846
H1O.BE.87.ANT70            ---.....---GAG.................................................................AA--G-A-C--GA-GTTATGGACC-T-..................TAC...TT  6331
H1O.CM.91.MVP5180          -T-.....----AG.................................................................AA---CA-C-GGA-GTCATGGAGC-TA..................TAC...AT  6318
CPZ.CM.04.EK505            ...............................................................................A----A-T--CGG-GAT---GAA-AATTGGCTAATCTGC.. TGCCTCCT-AT  5844
CPZ.CM.04.LB7              ...............................................................................A---GA-T---GGGGACATGGAC-A-C.........TGCATGCCATAC-AC-T  5833
CPZ.CM.04.MB66             ...............................................................................A--G--ACAC-GAGGA-TTG-AG-ACCTTATCACTGAAGGAGATA-TC-TAT-  5859
CPZ.MT145                  ...............................................................................A--GG-AT-------T-TTGG-G-A--AACTCATGCAGGTCCATC--CATCAA  5822
CPZ.CD.-.ANT               -TGCAATAGAAGA-.................................................................-..-T-AG---GCCGATA--TA-TATT.. TGGGGT.........CTGGCTTT  5715
CPZ.CM.-.CAM3              -A-.....----A-.................................................................----T------GATCA-AGG-AGC-T-..................-T-GA-AA  5693
CPZ.CM.98.CAM5             -A-.....----A-.................................................................-CA-T-----GGA-GA---T-AAT---..................-T-ATTT-  5981
CPZ.GAB_X52154             ..............................................GAAAGAGCAGAAGATAGTG..GAA.........A----A-TA-TGG-GA--A-GAA-A-AGACTGGAACAGCTTATCC-T-ATTAC  6344
CPZ.GAB2                   ..............................................AGAGCAGAAGATAGTGGAA..ATGAAAGTG...A--GG-A---C-AGGAGA-G--G-CAA..................CCC-ATT-  5679
CPZ.TZ.-.TAN1              ...............................................................................A-----A-TTT-AT-..................GGA.........-T-ACT..  5901
CPZ.US.85.CPZUS            AAGATAGTGGA-...................................................................A----A-T--TGG-GA--A-GAA-A-ACTCTGGCTAAGT............TA  6354
GSN.99CM166_AF468659       ......................................ATAGAAG..................................A---G-A---C-GTGG-ATCTA--.....................-TGAT-CT  6157
GSN.99CM71_AF468658        ......................................GTTGAAG..................................A---T-AG----GTGGGATTT---.....................-AGAT-CT  6142
H2A.ALI_AF082339           ...............................................................................A---T-TCT-GTAGAA-T--GC--.....................CTTGTTAC  6725
H2A.BEN_M30502             ...............................................................................A--G--CCTGGTAGGA-T--GC--.....................TTTGTT-T  6735
H2A.ST_M31113              ...............................................AAG.............................A--...TGTGGTAGGA-T--AC-A.....................TTTGTT-C  6177
H2B.D205_X61240            ...............................................................................A--GCAT-TTTTAGCAGC-G-C--..................CCT-TTGC-CT  6705
H2B.EHO_U27200             ...............................................................................A--GCAC-TGTTA--A-TT--C-A.....................CTTGTTAC  6700
H2G.ABT96_AF208027         ...............................................................................A--GCAT-TCTTGGCA-T--GC--.....................CTTATC-C  6048
H2U.FR.96.12034            ...............................................................................A--GCATGTCTTGGCA-T--GC--.....................CTTATC..  6187

                                                                                                          / Env gp160 start signal peptide
MAC.239_M33262             ...............................................................................A--GGATGTCTTGGGA-T--GC--.....................CTTATC-C  6635
Env                        .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.____M__G__C__L__G__N__Q__L__.__.__.__.__.__.__.__L__I__A sig pept

SMM.SL.92.SL92B            ...............................................................................A--GC-TGTCCTGGA-TT---C--.....................CTTATA..  6060
SMM.US.x.PGM53             ...............................................................................A--GGATGTCTCGGGA-T--GC--.....................CTTATC..  6552
MNE.US.-.MNE027            ...............................................................................A--GGATGTCTTGGGA-T--GC--........................CTTAT  6100
RCM.NG.x.NG411             TG-TAGGAGAT-T-ATTAGAGAAGTTTTATTTGTAGAGTTTATAGAAAATAGAAGGTAAATCTTTAGAT..........A--GCT......A-GTTAATAAGT...........................AT  5981
RCM.GAB1_AF382829          AA-AAGGTT-TTATAGATCTTTTAGTTACAATAATTAAAGGAAAATAGAATAAGATAAG....................A--G-TA--....................................-A--TA-T  5916
DRL.FAO_AY159321           ..................................................................................GT-A-A--GATGATTAGACATATA......ATTATAGGCTTAT-CATTTT  5953
MND-1.F17                  ....................................................................................-T-TG--GTGTCCAGG-CT..................TTTTCT-ATAC  5764
MND-2.5440_AY159322        ............................................................................GT.TAA-CA-TA-GTATGTTAAGA-AT.........CTTAGGTATATAGTC-TA--  5904
MND-2.14CG_AF328295        ............................................................................GT.A-A-TA-TA-GTATG-TTAGACAT.........TTAATCACAGCA-T-GTAAT  6266
MND-2.MND2_AF367411        ............................................................................GTAT-A-CA-TA-GTATG-TTAGG-AC............ATAATATTAG--ATAAT  6337
DEB.CM.CM40                -AGAAGAAG-CCC-.................................................................A--CCAA-TGTGACC--AA----T.............................  6183
DEB.CM.CM5                 .T.............................................................................A--CCTA-TGTG-CC---T-T--T.............................  6152
MON.CM.L1                  CTGATTCAGAT-C-TGAGAGAAGG...AGACATGACTCAGGG..GTTGACACAGAATCTGAATCAGAGCAAC.......A---G-A-A.......................................CTCAT  6192
MON.NG1                    -T-CAGGY--ATC-AGAGYATGATTCAGGGT................................................A---G-AT--GGATGA--AC-A-A...........................AT  4787
TAL.CM.8023                ..........................................................GTTACATGTTTAGCCACA...A----ATGCTT-C-AGT--TA................................  5553
TAL.CM.266                 ...........................................................GTATATGTTTAGCAGCA...A----AT-CTT-GCAGTTATT................................  6057
MUS.CM.1085                C-CAAGAAGGT---GGTCAGTGTCATCCGCCTCTTGGAAGAGGGAGACGGTGACAGCGGCATCTTCG.AAG........A--CC-AG--TGACATG--TGAGAGC-..................--CACCAC  6121
SAB.SN.-.SAB1C             ...............................................................................AGT-T-A--CTTCT-ACAGTAC--.....................CTT---CT  6509
GRV.ET.x.GRI_677           ...............................................................................A--GG-AGATTGCT-ATAA-AA-A.....................CT-ATAAT  6233
TAN.UG.x.TAN1              ...............................................................................A--GGACCATT-AGGGGAA-AGG-.....................GT--TATT  6298
VER.DE.-.AGM3              ...............................................................................A-----CT--C--TA-T-ATAGG-.....................-T-CTATT  5809
VER.KE.-.9063              ...............................................................................A----ACCTCTT-TAGGTATAG-T.....................CTTATA-C  6311
VER.KE.-.AGM155            ...............................................................................A---CAA--TTC-TAGGAATT--T.....................-T-GTATT  6306
VER.KE.-.TYO1              ...............................................................................A---G-T-T-C-ATAATAAC---A.....................G--ATAAT  5801
SUN.GA.98.L14              ............................................................GT..AAG............A--CG-TGC..........................................CC  6432
LST.CD.88.447              ...............................................................................A----CTGCCCTGGA-TAATAC-T...............CTCTTG-TCATTTT  5325
LST.CD.88.485              ...............................................................................A--G-CTGTCTTA-G-TC-TAC-T...............TTCTTGCTC-TTTT  5325
LST.CD.88.524              ...............................................................................A--GCATGTCC-GG-A-CATA---.....................CT--T-CT  5316
LST.KE.-.lho7              ............................................................................AGGA--GCATGTCC-GGATTAGGA........................-TTCT-CT  6400
SYK.KE.-.SYK173            ............................................................GTA.AAA............A--GCA-CTTTTAGAACTT--A-AGTTTGT.......................  6254
SYK.KE.KE51_AY523867       ..............................................................GCACA............A----ACTTTTTA--ATCATAAGTGTAGTG.....................TT  5931
COL.CM.-.CGU1              ...............................................................................A--CTTAGAT-TCTCTTTATAAAC...........................TT  5663
```

```
                         signal peptide                                Vpu end \   \/ Env gp120 start
H1B.FR.83.HXB2           GTGGAGATGGGGCACCATGCTCCTTGGGATGTTGATG......ATCTGTAGTGCTACAGAAAAATTGTGG...GTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACC.........ACCACT  6377
Vpu                      _V__E__M__G__H__H__A__P__W__D__V__D__.__.__D__L__*                                                                                  Vpu
Env                      __W__R__W__G__T__M__L__L__G__M__L__M__.__.__I__C__S__A__T__E__K__L__W__.__V__T__V__Y__Y__G__V__P__V__W__K__E__A__T__.__.__.__T__T  gp120

H1D.CD.84.84ZR085        -----A--------T-----------------------...ATG-C--A-------G----T--C------...--------T-----------------A-----A---------.........-----G  5898
H1F1.BE.93.VI850         -G---A-------CTTT-AT----G-----A--A--A......-------A----G----C--C------...--------------------------A---------.........--T---  5723
H1G.SE.93.SE6165         -----A------G--TT--A--------T--G----A.....--T--------CT--A-T--T------...--T-----------------------G-A--T---GAT.........------  5780
H1H.CF.90.056            A---------------T-AA--T-G--T------C--......------------G--C----C------...--T-----A------------------A--G----AA.........------  5717
H1J.SE.93.SE7887         -------G-----CT----A--T----A-----A---.....--T----AG--A-A----G-C------...--------------------------A--T----AG.........------  5696
H1K.CM.96.MP535          -G---AC-------T-T-AT--T-G--A--A-----A.....---------CAA-G----C---------...--------------------------A--------A.........C-----  5576
H101_AE.TH.90.CM240      -----A------G--TT--A--------T--G----A.....--T--------CT----C--C------...--T-----T-----------G----------GA--T---GAT.........-----C  5945
H102_AG.NG.-.IBNG        A-----------T--G-ATA-AT--T----AA----A.....--T----A----...---C---------...-----G-----C---------------------ACG---GAG.........-----C  5904
H103_AB.RU.97.KAL153_2   A---------------T--T-----------C-----.....------------------T--A---...--------T-----------------------A----------.........------  5590
H104_cpx.CY.94.CY032     ----GA------A--TT--A--------T--G----A.....-----------CT--A-C--C------...--------T-----------------------GA--C---GAG.........-----C  5747
H1N.CM.95.YBF30          -AA---TG-TT-GTTAC-CT--TA-CTTC-AG-A-GCTTGATC-AGGTA-T--GGT-T---C--CAT---...--A-----G--C-----------A--A----GA------GAG.........--A---  5964
H1O.BE.87.ANT70          AGCC-TG......G-TT--A-AACCCCATGT...............-TG--CCT-...AG-C-GC-A-AT...-CA----------GC------G-----A---G-A--T------A.........C-AGTA  6425
H1O.CM.91.MVP5180        AGCC-TG......G-TT----AA-CCCATGT...............-TG---TA--GTA--C------AT...-C------T----C-------------A---G-A--G---G-A.........C-AGTA  6415
CPZ.CM.04.EK505          AG-ATTGATCAAG-TA--TGGG...............................TCTG---------...--A-----G--C---------------C----GA--T---GAG.........--AGT-  5926
CPZ.CM.04.LB7            AGT-G-GATCCTG-G-T--TG-TCATT---AA-T-GTAGCACT-AAAA-..........-----T--A---...--G-----G--C-----A--G-----C-----A------GAA.........G-A---  5942
CPZ.CM.04.MB66           CACATTGGTATTAGGG--AA-AGGAATA--TAA-TGTGAA...................--T--CA------...--G-----T--C-----A---------C----GA--G---GA-.........--A---  5959
CPZ.MT145                --TA-TCCT-ATTGGATG-A--GCATCCTGT--TGGTGAA...................-----T-G----...-----T--T-----------C--A--A----G--G----AA.........-----C  5922
CPZ.CD.-.ANT             -CTA-TC...CAGTTT--AGAGAAG----C-.........................A-TG--GAC-AT...--A-----A-TC-----A--C-----C----GAA-T--G--A.........C-T---  5803
CPZ.CM.-.CAM3            A-AC-TG-TTATGGGTT--A-AA-CCCATGT----CTGGTAGT..................--TC-G-------...-CA-----G--C-----------------------A--G---GA-.........--A--A  5796
CPZ.CM.98.CAM5           C-AC-TGCTAATGGGTT--A-AACCCCATGT----CTGGC.................C-TG-G-G----...-CA-----G--C-----A---------A-----A--G---GAA.........--A--A  6081
CPZ.GAB_X52154           AATC-T-ACAAT--TTT----AACCCCATGT----CC....................TCTG-G--A---...--A-----A---------A---------T---C-T--T--TGA-.........C-GGTA  6441
CPZ.GAB2                 CATA-T-...ATGG-TT--A-AA-CCCATGT----A-ACAGAA...........T-T--TC---G----...--A-----A--C---------T--A--C---------G--T--T.........--AC-A  5782
CPZ.TZ.-.TAN1            .-T--TCCTC...-TA--TACAA-CCTAGG-A-T...................GG-TTT-GCACT-ATTACAC------G-T------A---------T-----A--G--CCAA.........C-A--C  5997
CPZ.US.85.CPZUS          T--TCTTCTCTCG-G-T--A-AA-CCCAGGA---.......................TCT-GC--A---...-CT-----A--------------------A----GA--T-T-GAG.........--A--C  6448
GSN.99CM166_AF468659     AGCAGTGA-CTTACTGGCTT-AA-G-CTT--C-A--C......CCGG--TTGAAGTGT...G--AAT---...ACG-----G--C--------------G--A----G---C------T.........C-AC-A  6266
GSN.99CM71_AF468658      AGTAGTGA-CCTA--G...--TA-G-CTT--C-A--C......CCGG--TCGAGGT-TA--GGGAAC---...ACG-----G-TC--------------G--A----GA--T----AA.........C-AC-A  6251
H2A.ALI_AF082339         TATCTT-CTA-CT-GTGCTTG-T-A-TATAT.............----AA............CAA-AT...--G--T--T-T------C--G--A-CA-----AA-T---T--.........-TTC-C  6816
H2A.BEN_M30502           CATTTT-CTAACA-GTGCTTG-T-A-TATAT.............-----C............CA--AT...--G--T--T-TC-----CA----C-C------AA-T---T-T.........-TTC-C  6826
H2A.ST_M31113            CA-CTTGCTA-CT-GTGCTTG-T-AATATAT.............--CGTC............CAA-AT...--G--T--T-TC-----C--G--C-------GAA-T---T--.........-TTC-C  6268
H2B.D205_X61240          CCT-CTTATA--T...--CAGTGGGTTTG-A.............----AA............CAA-AT...--T--T----TC-----CA----C-CA----G-A-C------A.........GTTC-C  6793
H2B.EHO_U27200           ACTCCTGCTTATA-GT--CTATGGGTAT---.............G-C-AG............AAC-TT...-----T----TC-----TA----C-CA-----AA-T---T-A.........-TTC-C  6791
H2G.ABT96_AF208027       AACCTTRCTCTTA-GT-CTTATGGGTTCTAT.............----CC............AAA-AT...-----T--T-T------CA----C-CA----G-A-C--G-G-.........GTAC--  6139
H2U.FR.96.12034          .GCTCT--T-T-TTTA-ATGCATGC-T---A-AC.............----C-............CA--AT...-----T----T------CA------CA----G-A-C--G--A.........-TTC-C  6280

MAC.239_M33262           CATCTTGCTTTTA-GTG-CTATGGGATCTAT.............----C-............C-A-AT...----------T------T-----A-CT----G-A-T--G--A.........-TTC-C  6726
Env                      __I__L__L__L__S__V__Y__G__I__Y__.__.__.__.__C__T__.__.__.__.__L__Y__.__V__T__V__F__Y__G__V__P__A__W__R__N__A__T__.__.__.__I__P  gp120

SMM.SL.92.SL92B          .GAC-TC-T-TTTTTA-GTG-GT-A----CC-G-.............---GCA............CA--AT...--A---A---T------TA-T----CA----G-A-C--T--G.........-T-C-C  6153
SMM.US.x.PGM53           .GCC-TC-T-CTTTTA-GTGCTTA------C-AT.............--C-C-............CAA-AT...--------T-T------T-----A-CA------A-T--G--A.........-TTC-C  6645
MNE.US.-.MNE027          CGCC-TC-T-TTTCTA-GTGC-TA------C-AT.............--C-T-............CAA-AT...----------T------T-----A-CT----G-A-T--G--A.........-TTC-C  6194
RCM.NG.x.NG411           AG-A-TTATAATAGGAT--T-AGG-ATA--T................-TA-TAGG---G---CAA-AT...--G-----A-T------T--C--A--A------A-CAGT--A.........GTTC--  6081
RCM.GAB1_AF382829        AATAGT--TAATAGTAG-AA-AGGGATA--AC-A......GTACAAG-ATCACAA-A-...CCGCAA-AT...A-A-------TC-----T--C--A--A----GAA-CAGC--A.........GTGC--  6025
DRL.FAO_AY159321         AG-ATTT-TA--A-TA-GTA-AGGAAAA..................................CAT---...--A--G--G-TC-----AACC--CAAA----GACC-----T--.........--ACA-  6035
MND-1.F17                TAT-TTTA-CCTACTAG-AGGT.........................-T-ATAGG-A--...CAA-AT...--G--------TC-----A-----A--A-----------T-AA.........--ACA-  5852
MND-2.5440_AY159322      AATA-T-GTAA-TGTA--AG-AGGA-AA..................................CAA---...--G-----G--C-----CAC---CAAA---C-TA-G----GG.........--ACAC  5986
MND-2.14CG_AF328295      AATAG-TATA--T-TA...GGA..........................-AA............CAA---...--A-----A---------CAC----AAA---C-CCC-----GA.........--GCA-  6342
MND-2.MND2_AF367411      TGTAG--CTA--GTTA...GGA..........................-AC............CAA---...--G-----A---------CAC----AAA---C-TA-G---GAG.........--ACAC  6413
DEB.CM.CM40              ...................................................................................................................................  6183
DEB.CM.CM5               ...................................................................................................................................  6152
MON.CM.L1                -GCTTTG-CAAC-CTGTGTT-AA-GAT--CT.................TTG-TG-AT-GGTCTAAC---...AC---C-----C--C--T--T--A--A-----ACCG-----T.........C-TC--  6292
MON.NG1                  AGTTTTGACAAT-CT-TGTT-GA-GAT-GG-ACCC--ACCAGTGGG-C-RA-T-AGG-A-C----CA---...---------A-TC-----A--G--G--C-AT--ACCT--G--G.........C-TC-A  4905
TAL.CM.8023              ....T-C-TA-CT-TTTGTG-AA-AT--G-AGCA-ATATAGGA-AAA--CT-A-CT-C-C-G--CA-CA-TTCC-------G--C--C--A--G--A--A-----AA-T-----G.........C-TC-A  5670
TAL.CM.266               .-TTT-C-TA-CTTA-T-A....................GGA-GG-T--TA...-T--GG--CCA--AT...------CA-----------A-----A---------AA-T-----A.........C--C-A  6150
MUS.CM.1085              -CTTTTGCTAAC-CTGC-TT-GAGCA-T-GT.................-TAAT-TGG-G--G-CAA---...-----------T-------------G---------A--T--T---.........C-AC-A  6221
SAB.SN.-.SAB1C           AA-TG-G--TT-G-G-T-AG-GTGGTTAG-C...............................CA--AT...--T-----G-T------TA-C--A--A-----AA-TAGTT-A.........GT-CAA  6594
GRV.ET.x.GRI_677         AGCA-T-G--ATA-GT--AGGAA-A--T-AC....................CTGTAT..................--G-----G-T------AA-C--A--A-----AA-TT----A.........GTTCAG  6318
TAN.UG.x.TAN1            AGTA-TT-T---ATTA-GCT-AA-A--AC--....................TTATAT-GG-C-CA--AT...A-T-----G-T------TA-C--A--A-----AA-CAGCT-A.........GTGCAA  6395
VER.DE.-.AGM3            AATAG-GATA--AGTAG----TAA-ACA-G-....................CAA.........CAA---...---------A-T------A-----A--A-----AA-CAGCT-A.........GTACAG  5897
VER.KE.-.9063            TATC-T-...--A-TA-GTTC-...TTAT-A....................AG-GA-AG-...CAA---...---------G-TC------A--------A------A-CAGCT-A.........GTACAA  6399
VER.KE.-.AGM155          AG-A-T-...--G-TAGGAA-A..-----A....................AG----A--C-GCA----...A-A------G-TC-----A-----A--A-----AA-CAGCT-A.........GT-CAA  6397
VER.KE.-.TYO1            AGT--T-...--A-TAGG-A-A...-T-T-A....................AG--AG......CAA---...A-A-------T------AA----A--A-----AA-CAGCT-G.........GTGCAG  5886
SUN.GA.98.L14            TGAA-T--T-ATAGGTT-CTCTT-GCT-T-A.................G--AG--T--C--T-CA--AT...--A-----A-TC------ACT---AA----G-ACC----GTA.........GTGC-A  6532
LST.CD.88.447            AA-AGT-CAA--G..............................GAG-AAA-A-AGA-G...CA--AT...--A-----G-TC---------G---AAA---G-A------GTA.........GTGC-A  5407
LST.CD.88.485            AG-AGC-CAA--G..............................GAGGCAAAG-AG......CA--AT...--A-----A-T----------G---AA----G-A--T--TGTA.........GTGC-A  5404
LST.CD.88.524            -CTCT-C-T---ACTA-CAGCAGGA..................GAA-AG-AG-A-......CAA-AT...--A-----A-TC-----A-----AAAA---G-A--T--TGTA.........GTAC--  5407
LST.KE.-.lho7            T-T-CTTCTT----TA--A......T--...............G-A-AA............CAA-AT...--A-----A-T----------G--AAAT---G-T--TAATGTTTCA......GTGC-C  6485
SYK.KE.-.SYK173          .......CTTTTT-G-T-AA-AAG-TTAGGA.................TT-ATGGA-A--C--CAA-AT...--A-----A-T-------A----ACAT---G-A--T---TAT.........G-AC--  6347
SYK.KE.KE51_AY523867     AG-T-T-GTATTA-GTT-AGG-A-A-TAGCT.........................-GGTT-G--CAA-AT...---------A-T------A--T--CAAT---G-A--T--TCAA.........GTAC--  6025
COL.CM.-.CGU1            -CTAGT--TA--T-TAG--T-A.............................AG--AT......AAA---...--A-----A---C-A--A-------CA---G-A------GATGTAAATGATCAGCAG  5754
```

PLV
complete genomes

```
H1B.FR.83.HXB2             CTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCC...ACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTA  6504
Env                        _L__F__C__A__S__D__A__K__A__Y__D__T__E__V__H__N__V__W__A__T__H__A__C__V__P__.__T__D__P__N__P__Q__E__V__V__L__V__N__V__T__E__N__F__ gp120

H1D.CD.84.84ZR085          ------------------------T-----A-A-------C-------A-C-----T------------------...--------------------A---A-C--AA------C-------------  6025
H1F1.BE.93.VI850           -----C--------------------------A-G-----C------C---------------------------...------------------------T-TC--AA------A------------G  5850
H1G.SE.93.SE6165           --------------C---------T-----AG-G-T--AAGC--------C-----T------------------...--------------------A--A-TA---A------A------T------  5907
H1H.CF.90.056              -----C------------------G--------G-----AAAG--------C-----T--------A---------...-------------------GA-G--CA---AG-----A-----G-GC----  5844
H1J.SE.93.SE7887           T-------C-----T---------------AG---T--AAA---------C-----T--------T---------...-----T--A-G---------GA-GAATC--CC------A--------C----  5823
H1K.CM.96.MP535            -----C------------------------G-AG--------------C-----T--C-----------------...--G------------------------AGA---A------A--------C----  5703
H101_AE.TH.90.CM240        --------------------C--G---C----G-----A--G--C-----C------------------------...--------------------A--CACC---A------A-------------  6072
H102_AG.NG.-.IBNG          ------------------------------------A-----------C-----T--------------------...---------T-----------A--CAT----A------A--------G----  6031
H103_AB.RU.97.KAL153_2     ------------------------------AG--AG---------------------T-----------------...--G-------G-----------A--CC----AA------------------  5717
H104_cpx.CY.94.CY032       ------------------A---------------G-A---A---------A-C-----T----------------...------------------------CTC--A-------A-----G--C----  5874
H1N.CM.95.YBF30            --T--C-----T-------------CC--AG--------CT--C--CA-C------------A--A-----T--T...--T--T-----T------------GC----ACCC-----A--T-----A----  6091
H1O.BE.87.ANT70            -----C-----T-----------CCT-ACAAGC--T--AAAG------A-------AT----A--------T--T...------------CT---T-T---TATCC----CAC-----------TG-C----  6552
H1O.CM.91.MVP5180          -----C-----T-----------CCT-ACAAGC--T--ACAG------A-------AT----A-----C--T--T...-----T-----T-----T---T-TCC-C-A-GC-----------T--C---G  6542
CPZ.CM.04.EK505            -----C-----T-----------G--CC--AG--------CT--C---A----------G--A--A--C--C---...--T----------------------C--A-ACC---------------CG---G  6053
CPZ.CM.04.LB7              --C--------G---------A------C-AA-GG------C---C---A-A------T-T--A--T---------...-----T--TG-T--TA---C-A---A-C-AAGC-----A-------------  6069
CPZ.CM.04.MB66             --G------------------A------C-GA-CC----A--G--C-----A-----TT-T--A--T------T--...---A----T------G-G--GA---A-C-AAC------A--T---T-C--C-  6086
CPZ.MT145                  --T--C-----C--C--------GAGT---AG------A-CT-----CA-C---------C--A--T--------T...--T--T----CT--T-----G---C-C--ACCT-----C--T---G-G----  6049
CPZ.CD.-.ANT               -----------CA--A----CTCCATGACAAG----------G--C-----A-----A--TACCAGT-----G--A...-T---T--AG-T--TATT...-----TAG-C-C---...--CTC-.......  5917
CPZ.CM.-.CAM3              --C--------C-----G--GTC---TCT-A-C-A-----CC------A--------T----A--------T--T...---------T-----------------GCAGA-CCC------A--T-----C----  5923
CPZ.CM.98.CAM5             --T--C-----C-----G--CTC----CT-A-C-A---A-CC--C---A--------T----A--------C--T...GT---T--T-----C-----G--GTATA--CC------AT-T----GC----  6208
CPZ.GAB_X52154             --C--------C-----C-----G---C--AG--------CT------A-------------G--A--------T...-----T----GT--T--G------T-TC-TCC------A-T----TCA----  6568
CPZ.GAB2                   T-G--C------G--A----A-C---T-TAA-AGAG--ACCT------A-------------A--------G--T...----------TCT---G----G--CA-CC-AA-T------------G-A----  5909
CPZ.TZ.-.TAN1              T-G--------C--T------G-TATTAC-AG--G---TAA---C--CA-A-----A------AA------G--T...TT---T-----T--TT-T------ACCC-A-CC------T--AT--GG----  6124
CPZ.US.85.CPZUS            T----C-----C--T-----A--G-----CA-GCAG----CC--C--CA-------------G--A-----C--T...--T--------T-------------CAT---CC---------T-----G---G  6575
GSN.99CM166_AF468659       T-G-----C--C------C-AG-TAT-GC-AG--AT--ACC-GGA--CA-----ATTT-TACA--A---T-G---...T-G------TCA--GGC---G--GCCG--A...--CA-A-----G--A--C-  6390
GSN.99CM71_AF468658        T-G--C-----C--------AG-CAT-AC-AG--AT--ACC-GGA--CA-----AT-T-TACG--A--CT-G---...T--------TCA--GTCT--G--GCC-C-A...--CA-A-----GG-A----  6375
H2A.ALI_AF082339           --C---------....................--CA-AAATAGAG--AC-----GA--CATACAG--CT----A...GACA-TGATG-TTAT--G---...A--GCTT-G-----------GGC---CG  6919
H2A.BEN_M30502             T-----------....................--TA-AAATAGAG-CAC-----GG--CATACAG--CT-G--A...GACA-TGATG-TTAT--G---...A--A-TT--------------GGC----G  6929
H2A.ST_M31113              --C---------....................--TA-AAATAGAG--AC-----GA--CATACAG--CT-G--A...GACA-TGATG--TAT--G---...A--GCTT--------------GGCC--CG  6371
H2B.D205_X61240            --CA--------....................--CACAAACAGAG-CACC----GA--TGTACAG---C-C--A...GACA-TGGTG--TACACT--G...A-CAG-C----CA-A-----GGC----G  6896
H2B.EHO_U27200             --C--------T....................--CAGAAACAGAG--ACC----GA--TGTACAA--CC-C--A...GATA-TGATG--TATACC---...A-CCAAT-----A-A-----GGC----G  6894
H2G.ABT96_AF208027         --C--C------....................--CA-AAATAGGG-CACA----G---TACACAG--CC-T--T...GATA-TGATG-TTACAGT---...C--GCAT-G-----A------GC----G  6242
H2U.FR.96.12034            --C---------....................--CA-CAATAGAG--ACA----GA--CACACAG--CC-T--T...GATA-TGGTG-TTATTC---G...T--GCAT-------C-----GGCC---G  6383

MAC.239_M33262             --C---------....................--CA--AATAGGG--AC-----GA---AC-CAG--CC----A...GATA-TGGTG-TTATTC----...--GGCCC-T-----T-------GC---G  6829
Env                        _L__F__C__A__.__.__.__.__.__.__.__T__K__N__R__D__T__W__G__T__T__Q__C__L__P__.__D__N__G__D__Y__S__E__.__V__A__L__N__V__T__E__S__F_  gp120

SMM.SL.92.SL92B            --C--C-----G....................--CC--AATAGAG-CACA----GG--CGT-CAG--CT-G--A...GATA-TGGAG--TATTC----...T-GGCCC-T-----T-----GGCC--CG  6256
SMM.US.x.PGM53             --C--C------....................--CA--AATAGGG-CAC-----GA---AC-CAA--CT-G--A...GATA-TGATG-TTATTC----...T-GGCCA-------T-------GA---G  6748
MNE.US.-.MNE027            --C--C----T-....................--CAG-AATAGGG--AC-----GA---AC-CAG--CC----A...GATA-TGATG-TTATTC----...T-GGCCC-T---A-T-------GC---G  6297
RCM.NG.x.NG411             A-G--------T....................--T--TAACGCAC-ATCA----GA--GTTAAAT---A----A...GA--GAGG--TAT---C----...---CCA---------T----G--A---G  6184
RCM.GAB1_AF382829          A-G-------TG....................--T--TAATAC-C-ATCA----GA--T-TAAAT---A----A...GAG-GAGGA-TAT-T-C---G...--T-CAA--------T-----CGA---G  6128
DRL.FAO_AY159321           --GA-C-----T....................--G--TAAT---TCAT-C----TG--TACAAGT--CA----A...-GTTTATTGC-TTATG-----AC--A-A-TAAC---A-AGAG----------  6141
MND-1.F17                  T-GA-------T....................-----TAATTCA-G-C-C----TA--CAC-AAT--CA----T...T--TTG--AG-TTATG-T--G----A-A-TCCTG--A-A-AG----------  5958
MND-2.5440_AY159322        --G---------....................-----CAATA--TCAT-C----TG--CACAAG------C---...-GTCTGTTGC--TATG-----CAGCATA-CCCC---A-A-----G-------  6092
MND-2.14CG_AF328295        --T---------....................--T--CAACA-CTCCC------TA--CACAAGT--------A...-GCTTATTGC-TTATG-G--GCA-C-CA-CCCC---A-A-------------  6448
MND-2.MND2_AF367411        --G--C------....................-----AAATG--TCCT-A----TA---AC-AGT-----T---AGCTTGTTG-A-T-TGA-G-GC--CC-A-TCC-AAC-TAACATGGA-TTT...C-  6519
DEB.CM.CM40                ..................................................................................................................................  6183
DEB.CM.CM5                 ..................................................................................................................................  6152
MON.CM.L1                  --C--------C--T---C-A--TTACGGGTC--A-----C-GGG--CAA----CTGG-GTC-T----CC-C--G...--------T-C---G---TCTC--TAT---AAC-TCACTGAG--GTTCAA-G  6419
MON.NG1                    T-G--C-----T-----CC--G-YTATGGCTCA-A-----C-GGG---AA----CTAG--TC-T----CC-C--T...T----T--TGTT----GC-CTC-CCCTC-CAAT-TAACAGA-AGCTT-AA-G  5032
TAL.CM.8023                T-G--C--C--C....................--TAT--C-AGGG-CCAG---T-T---TCCCAA--------A...T-------TC-GAAT-T---G...--GCCA--C---T----T--GT-C---G  5773
TAL.CM.266                 T-G--------C....................--CAT--C-AGGG-CCAA---T-A---TCCCAA--C--C--A...T----T--TC-GAAT-TC--G...---CCA--G---C-C--T--GT-C--CG  6253
MUS.CM.1085                --C--C---------------C--CAT-GCAA---G---ACC-GGG-----A---ATA--CACA--A--CC-C---...--G--T--TGCA--C--G-----GCAG---AAT-TCTCTGGG-----G---G  6348
SAB.SN.-.SAB1C             GCC------AAG....................--CCCCAATACG--CC-A------T-TAC-AAT--CA-C---...GAT--TGAACCAGA-GG-AC-...A--GCA-A-GTACCC-T-CCT---A-C-  6697
GRV.ET.x.GRI_677           GC------CATG....................--GCCCAATACC---A-G-----A--CACCAA---CA----A...GAT--T-ATG-TAATAC---G...--GCCTC----CA-T------GC---CG  6421
TAN.UG.x.TAN1              GCT--C---ATG....................---CCTAATACC--CC-------A---AC-AAT---A----A...GAT--T-ATG-TTATACT---...---CA-T-------CT-T--G--A---G  6498
VER.DE.-.AGM3              GCT--C--CATG....................---CCCACCACC-GAC-A-----A--TAC-AA--CGA----A...GAT--T-ATG--TACAC---G...---CCAT----CA-C--T---CCA---G  6000
VER.KE.-.9063              GCC--C---ATG....................--TCCTACTACC-GGC-A-----T---AC-AAT--CA----A...GAT--T-ATG-TTATAC---G...--GCCAT----CA-T--T--GCCC---G  6502
VER.KE.-.AGM155            GCT-----CATG....................---CCTACTAC--GGT-G-----A--TAC-AAT--CA----A...GAT--T-ATG--TATAC----...---CCAC-G---A-A-----GCCA---G  6500
VER.KE.-.TYO1              GCT--C--CATG....................--TCCCACCACA-GCT-A-----T--TAC-AAT--CA----A...GAT--T-ATG--TATAC----...---CCTC-----A-C--T---CCA---G  5989
SUN.GA.98.L14              T--A-A-----T....................T---CTAATA---GCT-G----TG--TAC-T-A--CT-G--A...GATTTA-AA-CTTATGC------GCCCA-AAC-GGGT-AGA---G-------  6638
LST.CD.88.447              T--A-------T....................T---CTAATA---G-T-A----T----ACAT-A--CA----A...GACTTG-AGTCTTATGC-C-G...A-TCCTA-TC-CAAC-T-AGTTTAAA-T  5510
LST.CD.88.485              T--A-------T....................T---CTAATA---GCT-A----T----ACAT-A--CC----A...GATTTA-AGTCTTATGC-C-G...A-TCCTA-TC-CAAC-T-AGTTTAAA-T  5507
LST.CD.88.524              T--A-C-----T....................T---CTAATA---GCT-G----T---TACAT-T---C-C--A...GATCTG-AATCTTATGCTC-G...A-CCCTA-TC--AAT-TTACTCTAAA-T  5510
LST.KE.-.lho7              T-GA-------T....................T---CCAATAC--GCC-C----T---GACAT-T--CT-G--A...GATTTG-AATCTTATGC---G...---CCTA-CT--AAT-TTTC-G-AAACT  6588
SYK.KE.-.SYK173            --C------A-C....................--CTCCCATA-AGGA-GA-----T--TA-GAAT-----G--GTCAG----T-AA-TAGA-GT-AGG--CAACA-TAC-GGA......---T-C---C  6450
SYK.KE.KE51_AY523867       A-G---------....................--CCCCCATACGGGA-G------A---A-GAA---------A......TCTT-TG-A-AGAT------CAG-G-TAAT-T-AGTGGG---T----C-  6128
COL.CM.-.CGU1              T-C--C--CTTT....................T-TTCCTCC-CAG-GA--CAACAGGTGTTG-GG---T----ACCTC-GCCTGGA--G---GT----CA-AACA--CC------A------GC----G  5863
```

```
H1B.FR.83.HXB2                ACATGTGG..................AAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGT............TTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCACTCTGTGT  6604
Env                           N__M__W__.__.__.__.__.__.__K__N__D__M__V__E__Q__M__H__E__D__I__I__S__.__.__.__.__L__W__D__Q__S__L__K__P__C__V__K__L__T__P__L__C__V gp120

H1D.CD.84.84ZR085             --------..................-----CA-------G--C--------------------T---............--------------T-----A----------------------G------  6125
H1F1.BE.93.VI850              -T------..................------A-------------A------ACA------------............--------------------------------G-----------------  5950
H1G.SE.93.SE6165              --------..................------A-------G-----A---------------------............---------G----------------------GC-------T--------  6007
H1H.CF.90.056                 -T------..................G-----A-------G--G---------ACA--C---------............---------------T-G--A-----------------------------  5944
H1J.SE.93.SE7887              --------..................-----------------C--------G--A------------............G--------G-----T----------------GA-------T--A-----  5923
H1K.CM.96.MP535               --------..................------A----------G---------ACA------------............---------G----------A--T------G-G--------T--------  5803
H101_AE.TH.90.CM240           --------..................------A----------G--------G------G--------............--------------T-----------------G-----T--T-----C--  6172
H102_AG.NG.-.IBNG             --------..................------A----------------------A--------T---............C-------C-----------------------G--------T-----C--  6131
H103_AB.RU.97.KAL153_2        -----G--..................------A-----------------------------------............--------------------------------G-----------------  5817
H104_cpx.CY.94.CY032          --------..................------------------------------------------............------A--G--G-------A-------C---GC-----T-T--------  5974
H1N.CM.95.YBF30               -T------..................G-----A-A----C---C--A-----A--------T------............C-G-----A--G---T-------C-----T------------T-A-----  6191
H1O.BE.87.ANT70               -T--A---..................------T-------------A-----G--A--C--T--T---............--------C--G--T--T--A--T-----TC--A-G--TTTC--G-----  6652
H1O.CM.91.MVP5180             -T--A---..................------T-------G--C--A--------A--C--C--T---............--G-----A--G--TT-------T----AG---A-G--TTTCT-A-----  6642
CPZ.CM.04.EK505               -T------AAA..................--CA-T--------T--A-----A--A--C--T--T--C............--G-----A--G--TT----A--C-----T--G---------T-A-----  6153
CPZ.CM.04.LB7                 -T------ACA..................----CA------A-T--------AC-----G----T--C............--------------------A------A----GC-------T--T-----  6169
CPZ.CM.04.MB66                -TGCA---GAA..................--CA-T--------G--A---------------G---A-............C-T-----C--G--TG----A----------------TT--TT-G-----  6186
CPZ.MT145                     --------GAA..................---T----------T--A-----G--A--C-----A---............C-C-----A-----T--T--A----------------A--CT--A-----  6149
CPZ.CD.-.ANT                  ..G-C---TTT..................----CTTATAA-A-TT-T---GTA--AAG---G-CAGAAGATATGNTACAA----TCC-A-------AT----------------C---A--TA-G---A-  6027
CPZ.CM.-.CAM3                 -T------..................--C---AC-----C------A---A--ACT--C--T--T---............C-G-----------TT-G--A--------T--GC-G--G---A-A-----  6023
CPZ.CM.98.CAM5                --------..................------A-T----C------A-----G--A-----T-GT---............C-G-----------T-----A--------T---C-G--A---A-A-----  6308
CPZ.GAB_X52154                --------..................------A-T----G--C--A--------A--C--T------............C-C--------G--T--C-----T--------G--G-----TT-A-----  6668
CPZ.GAB2                      -TG-A---..................G-C----CT--------G--A-----G--------T-CT--C............C-------C--G-----T-GA--T-----------G-A---TT-A-----  6009
CPZ.TZ.-.TAN1                 -T---GAA..................G-------T-----GC--G----A-A-A--------T-ATCA............C-T-TTC-A--G--TT-T-----T--------------A---T-T--CA-  6224
CPZ.US.85.CPZUS               --------..................G-----A-T----C------A-----G--A-----T--T--C............C-------C--G---T----A--T---A----G---------T-G-----  6675
GSN.99CM166_AF468659          ----C-AT..................------T----------TG--G---GA--T--C--GG-ATCA............C---TTA----GGC---G--A--------C-----G--T---A-G-----  6490
GSN.99CM71_AF468658           -T--C-AT..................------T-T--------TG--G--AGA--T--C--GG-ATC-............C---TTA----GGC-----GA--------C---C-G--T--GA-G-----  6475
H2A.ALI_AF082339              -TGCA---GAT..................---ACAG-AAC------AGCAGTA--A---G-CTGG--A............C---TT--GAC-TCAA----A--------C--G-----A--TT-A---A-  7019
H2A.BEN_M30502                -TGCA---AAT..................---ACAG--AC------AGCAGTA--A---G-CTGGCA-............C---TT--GAC-TCAA----A--------C--GC----A--T--A-----  7029
H2A.ST_M31113                 --GCA---AAT..................---ACAG-AAC------AGCAGTA--A---G-CTGG---............C---TT--GAC-TCAA----A-----C--C---C----A--CT-A-----  6471
H2B.D205_X61240               -TGCA---GAT..................---ACAG--AC-C-----GCAGTA--T---G-GTGG--A............C-C-TT--AACCTC-A----A--------C---C----------G-----  6996
H2B.EHO_U27200                -TGCA---GAT..................---ACAG--AC---T---GCAACAA-----G-GTGG---............C-C-TT--GACCTCAA----A--------T--------------G-----  6994
H2G.ABT96_AF208027            -TGCC---GAT..................--CACAG-AAC-------GCAGTA--A---G-TTGG-A-............C-G-TT--GAC-TCTAC---A--T-----T--------A--T--G--CA-  6342
H2U.FR.96.12034               -TGCC---AAT..................--CACAG-AAC---G--AGCAGTG--A---G-CTGG---............C-T-TT--GAC-TCTA----A-----------G--G--A--C--T---A-  6483

MAC.239_M33262                -TGCC---AAT..................---ACAG-CAC-------GCAATA------G--TGGCAA............C-C-TT--GACCTCAA-------T------------T-----T-A--CA-  6929
Env                           D__A__W__N__.__.__.__.__.__.__N__T__V__T__E__Q__A__I__E__D__V__W__Q__.__.__.__.__L__F__E__T__S__I__K__P__C__V__K__L__S__P__L__C__I gp120

SMM.SL.92.SL92B               -TGCT---GAT..................---ACAG-AACT-----AGCAATA--A---G-TTGG-A-............C-C-TT--AAC-TCTA-T--A--T--------G-----T---T-A---A-  6356
SMM.US.x.PGM53                -TGCC---AAC..................---ACAG-CAC-------GCAATA------G--TGG-A-............C-C-TT--GACCTCTA----A--C-----------------CT-A--CA-  6848
MNE.US.-.MNE027               -TGCT---GAG..................---ACAG-CAC-------GCAATA------G--TGGCA-............C-C-TT--GACCTCAA-------T------------------T-A--CA-  6397
RCM.NG.x.NG411                -TGCT---GAA..................---AG-T--TAT-----AGCAA-GTCTA--G-GTGGCA-............C-T-AT---TCC-CTT-G--A----------GG-----A---A-G---A-  6284
RCM.GAB1_AF382829             -TGCT---AAT...............--T-GCTTATAT-A-...--AGCAA-A--TA--G-GTGG-A-CTT.............-AT---TCC-CT-----A--------C-G-----G---T--A---A-  6228
DRL.FAO_AY159321              CTG-CCCTATG........................-CA-A-A-TG--G--ATAA-AC-GGC-TGGG--GCGCTGTCATCCA-GATA---GC-GTG--T--A--------T---A-C-AT--TTAT-----  6247
MND-1.F17                     CAGGACTT........................ATA-G--A-A-T-----AGT-T-TC-AGC-TGGCA-GCTATGGGAAGTA-G-TA---ACC-TA--T--------------GA-T-A----TAT-----  6064
MND-2.5440_AY159322           CAGGCCCTATA........................GAA-AGA-TG-A--AGTGACAC-AGCTTGGG-AGCAATTTCTTCCA-GATA---GC-GTAT----A----------GG-----A--TTAT-----  6198
MND-2.14CG_AF328295           CAG-CCCTATA........................CAA-AGA-TG-AG--GTAAG-C-AGC-TGGG-AGCTATTTCCTCTA-GATA---GC-GTT-----A--G-----------G--A---TA------  6554
MND-2.MND2_AF367411           CAGGTCCCATG........................GAA-A-A-TG--G-AGTAAT-C-AGC-TGGG-AGCTATCTCCTCCA-GATA--CGC-GTAT----A--T-----G---C-G------TA------  6625
DEB.CM.CM40                   .............................----CTT-T-ATA-C-CC--AG-G---C-GC--TGGCAGGATATGACTAGT--G-ACA-A--GTCAT-C--A-----------GC----A---TAT-----  6284
DEB.CM.CM5                    .............................G-CTCTT-TTC-A-TGTT---G-AC-AC-AT--TGGGACGACATGACTAGC----ACAGA---TCTT-T--A--C--------G---------TAT-----  6253
MON.CM.L1                     CTTATCAA.....................--CT-------G---G-A----GTA-------G-AGTC-............C-G-TCTCG--GGCTT-------T-----G--------A--CA-G-----  6516
MON.NG1                       CATATAAC.....................--CT-------G---G-----GTG--A--C----AATCA............C-C-TCTCA--GGC-T----------------G--------TA-G-----  5129
TAL.CM.8023                   -TG---AT..................-----CT-------G--C--C--AGCA--T-----GGA-GCA............--G-TCTTAAC-GCAACC-----T--CACC---C----T--TA-G-----  5873
TAL.CM.266                    -TG---AT..................------T-------G--C--C--AGCA--A--C--GGAT--C............C-C-TTTTGAC-GC-AC--GA------ACT--G-----A---A-A-----  6353
MUS.CM.1085                   ----C-ATAAG..................---T-T--------C------AC---T---G-G-GGTC-............A-C-TTA----GGC-T----A--------C-----------TA-G-----  6448
SAB.SN.-.SAB1C                CAGAAAAATTTGATGCTTGGAAAAATCGG--CCCTC-A----GC--AGCAG-GAGTA-C---CA-CT-............--G-TT--AAGT-CTT-------------T--GC-CT-A---A-G--CA-  6815
GRV.ET.x.GRI_677              -GGCT---GAT..................---CCGC-----A----AGCAG-GAGTA-----CATCTA............C-C-TT--A---CGA-G-G----T-----T--GC-CT----CA-A---A-  6521
TAN.UG.x.TAN1                 -AGCA---AAGGACAGG............---CCAT-A--G-C----GC-G-GAGTA-C---CATTTG............C-C-TT--AAGC-CT--G--------C------C-G--A--TA-G--CA-  6604
VER.DE.-.AGM3                 -AGCA---GCTGACAGA............--CCC-T-A----C---AGCAGGAAGTA-----CA-CTG............C---TT--G--G-CT--G------------------T-A--TT-G--CA-  6106
VER.KE.-.9063                 -AGCC---GCAGATAGG............---CCAT-A--G-C----GCAGGAAGTA----TCATTTG............--G-TT--A----CAT-G--A-----C---------T-T--T--A---A-  6608
VER.KE.-.AGM155               -AGCA---GCAGACAGA............---CC-T-A----C---AGCAGGAAGTA-C--TCA-CTG............C-G-TT--A--G-CAT-------C--------GC--T-A--T--A---A-  6606
VER.KE.-.TYO1                 -GGCA---GGTGATAGA............--CCCAT-AA---C---AGCAGCCAGTA-C--CCATTTA............C-C-TT--G----CTA-G--A--T-----G--G---T-A-----A--CA-  6095
SUN.GA.98.L14                 CAGA-G-AATAAGCAAT............---C-A--A--GC-G--AGCATGGC-A-C---G-C-TC-............A-GGTA---GCC-TTA-G-----T--------GA-C-AT--CTAT-----  6744
LST.CD.88.447                 T--CTAAAGAG.....................ATA-AA-ATA-T--A--TATAC--C-AGC-TGG---GCTATGACCAGTA-GGT----ACT-T-A-G-----T---------G---AT---TA------  5619
LST.CD.88.485                 T--CTCAAGAG.....................ATA-AA-ATA-T--A--CATAC--C-AGC-TGG---GCTATGACCAGTA-GGT----ACT-T-A-G-----T--------GA---AT---TA------  5616
LST.CD.88.524                 TT-CTCAAGAA.....................ATT-A--ACA-T--A--AATAC--C-GGC-TGGG-AGCTATGACGAGCA-GGT----GCT-TTA-G---------AC----A---AT--TTAT-----  5619
LST.KE.-.lho7                 TT-CAATACCA.....................-TA-AA-ACA-T---G-TATAC--C-AGCCTGGTC-GCTATGAATGCTA-GGTA--CAGT-TTA-G--A------------A-C-AT---TAT-----  6697
SYK.KE.-.SYK173               CAGCA---AATAGCAGT............C-CATG--AAGGC-G--A--ATTA-----C--GTCAGC-............----TCTTG---GC-AAT-GA--C-----G--GC--G-A--CA-G---A-  6556
SYK.KE.KE51_AY523867          CTGC----AATTCTTCC............C---G---AAG-C-----C-TTTGC----C--GTCTGA-............C-T-TCCTG--GGCTGAT-G---T------------T-T--TA-G---A-  6234
COL.CM.-.CGU1                 -TC-T-TT..................--G---AG-T-CAGT-GGG--G-A............TGG-TAATAACACAGACTACC-TA--A-----G----GA-------CT--G--G--AG--TA-----C  5963
```

```
                                       |-                       V1 loop
H1B.FR.83.HXB2              TAGTTTAAAGTGCACT...............GATTTGAAGAATGATACTAATACCAATAGTAGTAGCGGGAGA.........................................................  6662
Env                         __S__L__K__C__T__.__.__.__.__.__D__L__K__N__D__T__N__T__N__S__S__S__G__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._ gp120

H1D.CD.84.84ZR085           C-C------T------GATGCATCGAGAAACAGCACTG-T---A-C-GC-CA..............................................................................  6177
H1F1.BE.93.VI850            G-C------C--T---AATGCCACCAATAACAG-CA-G-A--GCCAGGGGCA..............................................................................  6002
H1G.SE.93.SE6165            --CC-----C--T---GATGTAACCAACAAA-GCAAC--A-GAA---A--GC--TG---A---C-CT-AA-CCAACAATAGCACTGTGGATAACCCA.................................  6104
H1H.CF.90.056               --C-C----C--T---AATGTCAGAAACAATACC-CT--C-GCAC--GC-G-..............................................................................  5996
H1J.SE.93.SE7887            C-C------T--T-G-AACATCACTAGTAACAG-AAT-CA-C-AG--AC-G--GT...........................................................................  5978
H1K.CM.96.MP535             C-C------T--T---GATTACAAAGGAACCA--AGC-CC---A--G-A-C--G--C-........................................................................  5861
H101_AE.TH.90.CM240         --C------T--T--CAATGCTAATTTGACCA--GGC-GT-GCA-A--C---GT-TC--AC.....................................................................  6233
H102_AG.NG.-.IBNG           --C----G-T--TCA-AATTTCAACAACAGCT-CAGC--C-G-AGC-ACCTA--T-G-........................................................................  6189
H103_AB.RU.97.KAL153_2      --C------T------GATTTGAAGAAGGAG-T-ACT-GT-CC.......................................................................................  5860
H104_cpx.CY.94.CY032        --C---T-CA--T-T-AATGCAACTACTACTA--AGT-CC----GC---.................................................................................  6023
H1N.CM.95.YBF30             A-C-A-GCTT--T-ACGATAGCTATGGGGAG--AAG---C---ACA-A-.................................................................................  6240
H1O.BE.87.ANT70             ACAAA-GG----T--AAACATAGCTGGAACAACAAATG-A..........................................................................................  6692
H1O.CM.91.MVP5180           ACAAA-G--C--TGTAGATCTGCAAACAAATA-AACAGGC..........................................................................................  6682
CPZ.CM.04.EK505             A-CCC--TCA--T-GCTCATGGAGGAGTGTGA-CAAT-GT..........................................................................................  6193
CPZ.CM.04.LB7               --C----C-T----G-ATCCCTAAATTTGATA-C-CT-GC..........................................................................................  6209
CPZ.CM.04.MB66              --C------C--TT-ACTGTTTAAATGCATTA-AGAA--T........................................................................GGCAACACCACCAACTGC  6244
CPZ.MT145                   G-CCC-T-CT--T-A-AATCCCACTAATACTTCC-GC-CT..........................................................................................  6189
CPZ.CD.-.ANT                A-AAA-G--T--T--AGGATACAATGGAACACC-ACA-CACCA.........-GT-CA-CA-CA--TACA............................................................  6088
CPZ.CM.-.CAM3               C-CCA-GG-A--T-GAAAGGTTACTTTCAATAGCACTTCT---AGG-A---A--TTC--C--TG-CTACT............................................................  6093
CPZ.CM.98.CAM5              C-CCA-GC-A--T--GACCCTTAAGAACATTAGCACT-CT---AC--AC...-GT-CAGAGTTA-AAAAT............................................................  6375
CPZ.GAB_X52154              A-CC--GC------G-AAGGCTAACTTTAGC.....................CAGGCA-AA-ACCTAACA............................................................  6717
CPZ.GAB2                    GCAA----C---T---TCTATAAATGCTACT-GAAACG-A-C-A--G---C-GGA...........................................................................  6064
CPZ.TZ.-.TAN1               A-AGA-G-CA--T...ACAATGACTAATACCACAAAT--A-CC....................................................................................CTG  6267
CPZ.US.85.CPZUS             --C-A-G-CT--TCT-AACCCC.......................................GA---TAAT............................................................  6706
GSN.99CM166_AF468659        GG-GA-GT-T--T-ACCTCACAAACACATCCACAAGCG-ACCAACA--A-C-C-A-GGCCACCA-AT-TCTCCACAACCACG................................................  6572
GSN.99CM71_AF468658         AC-AA-G--A--T-AGCTCCCAAATACGAGCACAACA-CA-CACCGT-C-CA--A-CAGCC-CA-CATCACCCTCATCCCCAACGCCAACACCA....................................  6569
H2A.ALI_AF082339            AGCAA-G-----T-GCAAC...ATAAGCACA--GAGC-CA-CCACAT-CCCG-G-CCAG-G--C-CACTC............................................................  7086
H2A.BEN_M30502              GGCAA-G--T--T-GCAGG...GTTCAAGGGA--ACC-C--CCCCG-A-CCC-GG-CCTCG---TC-ACA-CCTCGAGACCACCCACATCC.......................................  7117
H2A.ST_M31113               AGCAA-GCGT--T-ACAGC...ACAACTGCAA-AAAC-CA-CCTCC--ACCA...-CA-CC-CC-CAACAGC-.........................................................  6538
H2B.D205_X61240             GGCAA-G--C--T-G-AAAACCGAAACAAACCCAGG---TGCCAG--G--C-----CC-C--AGCCTACT-CCACCTCT...................................................  7075
H2B.EHO_U27200              G-CAA-G-----T-A-AAGACGTGGAGCTCA-C-AGC--AG-GACC----CGT--TCCGCGTCCCT-A-ATCTTCTACTCAG................................................  7076
H2G.ABT96_AF208027          -GCCA-G-GA----A-AAAACTGAGACAAACA-A-G-GGTCTAACAGGG-CC----C-GTA-CCCCT-CA-C-GTTCCCACAACTAAAATGGTGACAGCT..............................  6442
H2U.FR.96.12034             AGCAA-G-GA--T-G-AAAACTGAAACAAATAGA-G-GGATTAACAGGA-CAG-AGCA-CA-CA-CTA-CCC-GCAATGCAAGAATCAAGT.......................................  6574

MAC.239_M33262              --C-A-G-GA----A-AAAAGTGAGACAGATAGA-G-GGATTGACA-AATCA......-TA-CA-CAACAGC-TCAACAACATCAACGACAGCA....................................  7017
Env                         __T__M__R__C__N__K__S__E__T__D__R__W__G__L__T__K__S__.__.__I__T__T__T__A__S__T__T__S__T__T__A__.__.__.__.__.__.__.__.__.__.__.__.  gp120

SMM.SL.92.SL92B             -GC-A-G--A--T-A-AAAAATGAGACAGACAGA-G-GG-TTGACA-GAGCAG-T-C--C--C----TCACC-ACTACTACTAGCCCCTTAACTGCTGCTAGCCCATCAGGAGAG...............  6471
SMM.US.x.PGM53              --CCA-G-GA----A-AAGACAGAAACAGACAGA-G-GG-TTAACA-AG---GAA-TACCA-CA-CAACA-C-ACAACAACAACAAAAAGCCCAAAAGCA..............................  6948
MNE.US.-.MNE027             --C-A-G--A----ACAAAAGTGAAACAGATA-A-G-GGATTGACA-AATCAT-A-CA-CA-CAGCA-CACC-ACAACAAAAACAACAACAACA....................................  6491
RCM.NG.x.NG411              A-CAA-GT-T--T---GCATTGAATGGTAGTTGGGATGGT-C--CC--CC-AT-ACC--CA-CA-CAACA---GCAACAACAAGAACAACA.......................................  6375
RCM.GAB1_AF382829           --CCA-G--T--TT-AGCAATAAATGGTAGCTGGGATGGA.....................-TCCCTACCTC-GCACCACCAACAACAACAAAAACA.................................  6304
DRL.FAO_AY159321            G-AAA-GG-A--T-G-GATCTC.........A-AA-AGTA----CAG-C-GC--A-CACCA-CACCAACT-C-......................................................GTA  6314
MND-1.F17                   --AGA-GC-A--TCAGGAAACAGAAAATGTATCAGCA-CA-CA-C--AGCC--TA-C--CACC--CTACT-C-TCTACAGTTGCA.............................................  6149
MND-2.5440_AY159322         G-AGA-G--A--T--AGAAGGTCAAAATGAAACAGA-C-AGCAACAG----G----CA-CACC-GTACCC-C-ACAACAACCCCCTCCACTACCACT.................................  6295
MND-2.14CG_AF328295         C-AGA-GC-A--T--AAAGGGAGAAAACAGTAG-ACC-GC-C-ACC---CCC-TGCCC-C--CC-C-ACTCCCACCACTCCA................................................  6636
MND-2.MND2_AF367411         C-AGA-GC----T--AAAAGGAAAAGACCTT-C-ACCCCT-TACCC----CAT---CC-CA-C--C-ACA-CTACCACC...................................................  6704
DEB.CM.CM40                 ---CA-GC------TAAAAACATCAACAAATCCAACACCA-CAA-C--ATCA--A-CA-CA-CA-CAATAGC-ACAACAACA................................................  6366
DEB.CM.CM5                  GC-GA-G--C--T...........................-G-AG---A-CA--A-CA-CA-CAGCA-CA-C-ACAACAATA................................................  6308
MON.CM.L1                   ---GA-GCTC--TGT-GAAGTTAACACCGTTTCGAATGCC-G-ACC---CC-G--CC-----CCCC-ACACCC.........................................................  6589
MON.NG1                     A--GA-GCG---TCAAARGATCAATGAAAGCACAACAGT--CRCCA----CC----CYWCC.........................................................GCCACCACACCA  5202
TAL.CM.8023                 G--GA-GGC---T--AGACATCACAAAAAATCTCACTGGC-CA-TAG-ACCCC-A-CA-CA-A-TCAACA-C-AACTCAGTAACATCC..........................................  5961
TAL.CM.266                  A--GA-G-----TGAGGACATAACAAAGAACCT-ACA-TC-G--CC-A--CC--A-CACCAGCCCCGAC-............................................................  6423
MUS.CM.1085                 A-AAA-G--A--TG-GAAAGTAACGAATTACACCAGATCT..................................................................ACGGGAGGGAATAGCACCGGAACC  6512
SAB.SN.-.SAB1C              A-AAA-G--C---TA-CGCTTAGAAGGAGGA-CAGCA-CA-CAACAT-ACCAT-A-CATCA-CAGCAC-CCCCGAGGTAGTGAGT.............................................  6900
GRV.ET.x.GRI_677            --AAA-GTCC--TGTAGAGCTGAATGGTACA-CCAC--CA--G-CC--C-C---TGCA-C--CA-CAAT--CT..............................ACCCCCTGT..................  6603
TAN.UG.x.TAN1               A-AGA-G--T--T---AAATTAACAAGTACC-CCCCT-CATCAAG---CCC---ATCA--C--C-CTAC-.................................GATCCCTGTCCAAATACCGACGAA...  6698
VER.DE.-.AGM3               --AAA-GTCC--TGTAGAATTGAACTCCTCT--GCCT-CC-CCAC-C----A-GT-CC-CGGCCTCAACA-CCAATATCACAGCCTCAACAACCACT......TTGCCGTGTGTCCAGAACAAGACAAGT  6230
VER.KE.-.9063               A-AAA-G-GC--TGTAGAGTTGAACTCAAGTAGAAGA--CCCAAC-C-A-C-T-AGC-TCA-CA-CAAAAGCT...CCAAAAACAGGT...............GATCCCTGCATTAAATCTACCAATAAT  6720
VER.KE.-.AGM155             C-AAA-G--T--TGTAGAGTTAAAAGGCTCC-CAACCTCT-CCCCAG-A-CCT-T-C--CGGCAG-AACC-A-..............................CTACCCTGTGTTAGAAATAAAACAGAC  6706
VER.KE.-.TYO1               C-AGA-G--T--TGTAGAGTTAAATTCCACAAGAGAA-G-GCGACA--ACC---A-CG-CGCCG-AATCT-CCGGC...........................CTACCCTGTGTAGGCCCGACGTCAGGT  6198
SUN.GA.98.L14               C-AAA-G-----TC-ACCGAAACCGACAACTCCC-CCTCC....................................AACTCAACAGTCAAAAGTAGCTGTGATTATTGGACGACCACCACGGCTAAGACT  6838
LST.CD.88.447               G--GA-G--A--TGAAGGGGAAGGAATAGGAACAACA--AGCAACA-AA-CA-AA-CG...TCA-CAACACCC..............................ATGCCTTGCTTTATAACAGAAAAAACC  5716
LST.CD.88.485               G--AA-G--A--T--AGGGGGAGGAATAGGACCAAAA-CA-CACCAT-ACCA-TAGGA-AA-CA-CAACACCC..............................ATGCCTTGCTTTATAGCAGAAACAACC  5716
LST.CD.88.524               -C-CA-G--A--T--AGGGGAAGGACTAGGTA-GAA--GA-CAACA--A-CA--A-CACCA-CA-CAACACCC..............................ATGCCTTGCTTTATAACAGAATCTACT  5719
LST.KE.-.lho7               G--AA-GC----TGGGGAAGTTACAAAAACACCAACA-CA-CACCA-AA............-C--CTACACAG..............................ATGCCTTGTTTTATCAACGAACAGGTA  6785
SYK.KE.-.SYK173             ---GA-GTT---T--ATTGGACAACAGCCCT-C-ACA-GT.................................................................................ACCCCTACT  6605
SYK.KE.KE51_AY523867        C--AACGTTA--TGTAGAGGACAAGGACAGTAG-AGTGGT...............................................................AATCTTACCACAACAATCAGTCCAACA  6301
COL.CM.-.CGU1               -CCAA-G-TT-----CAAAGTAAAT.......................................--GACTGA-AATGGCACTAGTACAGTAGCCCCAACAACA..................ACAAATAAT  6036
```

```
                                                                  V1 loop                                 -|-                V2 loop
H1B.FR.83.HXB2             .........ATGATAATGGAGAAAGGAGAG..............................ATAAAAAAC...TGCTCT.........TTCAATATCAGCACAAGC......ATAAGAGGTAAGGTGCAGA  6735
Env                        _.__.__.__M__I__M__E__K__G__E__.__.__.__.__.__.__.__.__.__.__I__K__N__.__C__S__.__.__.__F__N__I__S__T__S__.__.__I__R__G__K__V__Q__  gp120

H1D.CD.84.84ZR085          ....................................CTGCCAACGGTGAAGCCAGGAGAA--G------...------.........--------A-C----GTA......G---CG-A----AG-A--C  6253
H1F1.BE.93.VI850           ............................................................--GC-----...--T---.........--------G-C----GAA......G------A----AA--T--  6054
H1G.SE.93.SE6165           ......................................................GGAGAA---------...------.........------G-A-C----GAA......-------A----AAAA---  6162
H1H.CF.90.056              ..........................................ATGGAGGCAGGAGGGGAAC---C---T...------.........------G-A-CT---GTA......C------A----CA-----  6066
H1J.SE.93.SE7887           ..........................................GTAAGTAGTCCTGATATC--G-C----...------.........--------A-CT---GAA......------AA---AAG-A--C  6048
H1K.CM.96.MP535            ..........................................GTGGTAAGCCCAGCAGAA--------T...------.........--------A-C----GAA......----A--A----AA-A-A-  5931
H101_AE.TH.90.CM240        ....................................ATAATAGGAAATATAACAGATGAAG---G----...--TA--.........--T-----G-C----GAA......C---C--A----AA-----  6309
H102_AG.NG.-.IBNG          .............................................GACATGAATGGGGAA---------...------.........--------A-C----GAA......G------A----AAAA---  6256
H103_AB.RU.97.KAL153_2     .................................AATACTAGTAGCATAAAAATGATGGAA--G------...------.........-----C----C----GA-......C-G----A---A---A-A-  5939
H104_cpx.CY.94.CY032       .............................................GTGATTAAAGAAGGA---------...------.........---G----A-C----GAA......-----G-A----AA-A---  6090
H1N.CM.95.YBF30            ........................ATGACAACAAGAGAACCAGACATAGGATACAAACAA--G-----T...-----A.........------GCA-C---TGAG......C---C--A---AAA-A--C  6328
H1O.BE.87.ANT70            ......................................................AACCTT--G--G--G...--TGAG.........--T---G-A-C---TGTT......--C-A--AC--AAA-G---  6750
H1O.CM.91.MVP5180          ....................................CTATTAAATGAGACAATAAATGAG--G-G---T...--TAG-.........--T---G-A-CT---GT-......C-C-C--AC--AAA-G--C  6758
CPZ.CM.04.EK505            .......................................GTTAACCAAACAAACCACGTACA--TGC-AAAT-----A.........--T---G-G-C---TGAG......T------A---AAA-A--C  6269
CPZ.CM.04.LB7              ..........................................ATTAATAGCTCTAATGAG-CT-TG--AAAT--TAGC.........-------CA-CA---GAA......C------A----AA-A-AC  6282
CPZ.CM.04.MB66             ACTGTTCAA--T-GC-CT-GT--T-ATAGCACTGCTAACAACATTACGGTTGGCACAATAGAC-TGT--AAC--T---.........------GCA-C----GAA......T------AC-GAAA-A-AC  6359
CPZ.MT145                  ....................................AACTCTACTGACGATCGCTTGGGAGAC-TG-GAAAT--TAGC.........--T---G---CA--TGAA......C------A----AAAAGAC  6268
CPZ.CD.-.ANT               ............G---CACCA--GAC-ACAACACCAATAGTAGATGGCATGAAG......C--C--G-A...--TAAC.........--T---CAG------G-A......T-T-A--A----AAA--A-  6182
CPZ.CM.-.CAM3              .....................--CA-TCCTAAT.........GAAAAAAYAGACAGCACAG----G--T...--TA-C.........--T-----G-CT--TGAG......G------A---A-A-A-A-  6175
CPZ.CM.98.CAM5             ................................................ATAAAAGAAATA--G-----T...--TA-A.........--T-----G-CT--TGAA......G------A---AAA-A--G  6439
CPZ.GAB_X52154             ............-ACCA-ACATCTTCTCC-CCTCTCGAA.....................--G------...--T--A.........--T---G-G-CA---GAA......T------A----AA-A-AC  6796
CPZ.GAB2                   .....................--C---ATTGAAAAG......GATGGGTTGGCACAAGAC--G-G---T...--TA--.........-------CA-C----GAA......T------A---AAAA---C  6149
CPZ.TZ.-.TAN1              AATTCGGCA-CA-C--CCTTA-C-CC-ACAGTAAATTTGAGTTCTATACCTAACTATGAGG-GT-T--T...--T--A.........--T---CAG-CA--TGAG......T-T----A----AAAA-AC  6379
CPZ.US.85.CPZUS            .........-GT-GTGCT-TA--TACTACTGAT...ATA.....................--G-G----...--T---.........--T-----A-CT--TGAA......T------AC--AAA-A-AC  6785
GSN.99CM166_AF468659       ............CAGTG--GA-GTT-G-G-GGGGAAAATGGAACAGGACAACCC......T--T-T---...--T--C.........--T---CAG-C----GAA......T-T----ACC-AAA-A-AC  6666
GSN.99CM71_AF468658        ...............TG--GA--TT-G-GAGGT...AATGGAACAGGACAGCCA......---T-T---...--T--C.........--T--CCAG-CT---GAA......T-T--G-AC--AAA-AGAC  6657
H2A.ALI_AF082339           .....................---CCC...CTGATAAATGAGAGCGATCCATGCATAAAGGC-G-C---...---C-C......AGGGGACTAGGGGATGA-GAG......--GGTCAA-TGTCG-TTC-  7177
H2A.BEN_M30502             .....................GC--CCTCCATAATAAATGAAACTTCTAACTGCATAGAA-AC--C-CA...---G-A.........GGATTAGGGTATGAGGAG......--G-TGCAATGT-A-TTC-  7208
H2A.ST_M31113              .....................--CAC-ACAATAGGA...GAGAATTCTTCATGCATACGC-C-G-C---...---A-A.........GGGTTGGGAGAAGA-GAG......--GGTC-ACTGTCA-TTC-  6626
H2B.D205_X61240            ..................CGTGGGCTGA-AACGATTAACGAAACAGACCCATGCATAAAA-ATG-C-G-...---A-A.........GGACTAGGAGAAGAGGAA......----TGCAATGTAATTTT-  7169
H2B.EHO_U27200             ...........................ACCCTACTCAACGAAGATAGCAAATGTATTCAA-ATG-C-G-...--TG-A.........GGG-TAGGACTAGA-GAA......--G-T--ACTGTCAATTC-  7161
H2G.ABT96_AF208027         ...........................--ACTAGTGAATAGCACTAGCCAGTGCTTAATGTACG-T--T...--TA-A.........GGT-TACAGTCAGA-TC-......--GGTT--CTGCAAATTT-  6527
H2U.FR.96.12034            ...........................A-AGTAATAAATGAAACAGAGCCTTGTATACGA-AT--T-G-...-----A.........GGATTAGAACAAGA-CCA......T-GGTGA--TGTAGATTT-  6659

MAC.239_M33262             ...............TCA-CA----T---CATGGTCAATGAGACTAGTTCTTGTATAGCCCAGG-T--T...---A-A.........GG-TTGGAACAAGAGCAA......--G-T-A-CTGTAAATTC-  7114
Env                        _.__.__.__.__.__S__A__K__V__D__M__V__N__E__T__S__S__C__I__A__Q__D__N__.__C__T__.__.__.__G__L__E__Q__E__Q__.__.__M__I__S__C__K__F_  gp120

SMM.SL.92.SL92B            ...........................--AATCGTTAACGACACTATGTCTTGTACAAAG-AC--C--T...------.........GG--TAGAGCAGGA-CCA......--G-T----TGTCAATTT-  6556
SMM.US.x.PGM53             ............GA-GCAATA-C--C-A-AGTTATAAATGAAAGTGATCCTTGTATAAGC-AC--T--T...--TA-A.........GG-TTAGAGCAGGA-CCA......--GGT-A-CTGTAGATTT-  7048
MNE.US.-.MNE027            ..................A--G--AT---AGTGGTCAATGAAAATAGTACTTGTGTAAATCGTG-T--T...---A-A.........GG-TTGGAACAAGAGCCA......--G-T-A-CTGTAAATTC-  6585
RCM.NG.x.NG411             .................................AAAGAAATAGGGAAAGATTGTTATGCG-ATC---CAGAG---AG-.........AGTGTCCCTGA-G--GAT......G-C-TGAA-TGT-AATTT-  6457
RCM.GAB1_AF382829          ..................ACA-C-CA-AGAACTATAGGTGTAGAAAAGGAATGTACTGCTGGC--CG-AACA--TGAG.........GAAGT-CAGGATG--GAT......G-G-TGTC-TGT-AATTTG  6401
DRL.FAO_AY159321           GCTACTACTG-T-AT-CTAGTGTT-CCTCAACAACAACAGCAAATGACACTTTGGAAATAGCTGT---T...AATA-A.........GAACCAG-A-TGGA--CA......-AC----TATGTAAAT-C-  6426
MND-1.F17                  ...............-GTAGT-C--AGATTTACTTAGATGTAGATAAAAATAATACAGAAGA---GGTA..............................GAG--G......-ATCAT-TATGTAG-T-T-  6228
MND-2.5440_AY159322        ..................AGCTCCA-CACAAATAAGACAACAACTCCTGTGTTGGTTGTTGAG---C-AAATAATGAG............-CA-CA-CACAGCAA......-AT----TATGTAAATTC-  6389
MND-2.14CG_AF328295        ...............GCA-T--CG--GCT-GAGGTGACGGTACAGAAT........................AATGAG............-G--C--CACA-GAA......-AT--G-TATGTAAATTT-  6709
MND-2.MND2_AF367411        ...........................A--ACGGTGGCAAATACTACTGAGTTGGACATAGAC-CC--TAACACTGAA............-C--CA-CACA-CAG......-AC--G-TCTGTAAATTC-  6789
DEB.CM.CM40                ...............-AAACA-CTAC---TTGGTCAGGGGAAAACATAACCATGACACAATATTGG--T...--T--A.........--T---G--TCAGGTCCT......TAT----A----AAAG-A-  6463
DEB.CM.CM5                 ...............-AAACA-C-ACT--CTGGTCAGGAGAGAATAAGACAGTGACACAGTATTGG--T...--TA-C.........------G-T-CAGGGCCA......TAT--G-AC---AA-G-A-  6405
MON.CM.L1                  .....................TGG---A-TTGGGGA...GGGAAT...GGAACAGGACAGCC-GT-T-TAAT-----ATTTAACCAGACT-CAGAATT--GGGAT......-A--AGAAGC-AA--T-T-  6686
MON.NG1                    .....................TGG---AGTTGGGGA...GGCAAT...GGGACGGGACAGCCTCTTT--AAT--T--ATTTAACCAAAC--CAGAATTT-G-GA-......-A--RGAAAC-AA--T-T-  5299
TAL.CM.8023                ............-AG-GAACC--T-ACACCAATGTGTGGTGGGGAAACAATGCCACAGAACC-GTGT--AAC--TA-ATGGAATATGACAGGAGGATTTT-GGGAT......-AG-AGTAC-CCTATTTTT  6073
TAL.CM.266                 ............GCC---AGC--TA-CA-TATGCCGTGGTGGGGAGACAATACAACAGAACC-GT-T-TAAT--TA-ATGGAACATGACAGGAGCATTT-GGGAT......-AG--GTA--CCTACTTTT  6535
MUS.CM.1085                ACAGCACCCCCA-C--CACCATGG---TCATGGGGCAATGGCAATGATAGTGCTGGACAACCT-TGTT-AAT-----CTTCAATCAGAC--CAGAGTT--GGGAT......-A--AGA-AC--A--T-TT  6636
SAB.SN.-.SAB1C             ............G-GGGCTTT--T-ACTCAGTGATAGAACAAGAGATGGAAAAAGAACAGGCC-TG---...--T--C.........--TGC---GGCAGG-TAT......-G---G-A-GTAAA-A-A-  7000
GRV.ET.x.GRI_677           .............................................CAGAATTGCAGTACAGAGC-G-TAGAAG-AGAA............-TGGCAGAGGA-CCA......GC-TCCAACTGCACTTTTG  6670
TAN.UG.x.TAN1              .............................................AGTAGCTGTAACGCC-CCTT-GTTACAAATAGC............-TGGATTATGAG-AT......-GTTCTATATGCTCCTTTG  6765
VER.DE.-.AGM3              ACTGTG...T-AGA-TCATGT--T-A-ACAATCATAGAAAAGGAATTAAATGAAGAGCCTGCTTCT--T...--TA-A.........--TGCA--GGCAGGGTAT......G------A-C--AAAA---  6339
VER.KE.-.9063              AATGTTAATT-ACAGCCCTGT--T-CCTCCCTTATAGAAGAGGAACTGGAGGAAGAAGCAGCTTC---T...--TA-A.........--TGC---GGCAGGGTA-......--T----A-C--AA-A-A-  6832
VER.KE.-.AGM155            TCC...AACC-ACAGTCATGC--C-ACACCATCATAGAAAAGGAGATGAATGACGAGGCAGCGTC----...---A-C.........--TGC---GGCTGGGTA-......--T--G-ACC-AAA-A---  6815
VER.KE.-.TYO1              GAA...AATC-ACAGTCCTGT--T-C-AGCATTATAGAAAGGGAGATGGAGGATGAGCCCGCCTCT--T...--TA-A.........---GCA--GGCTGGCTAT......G------A-C--AA-A-A-  6307
SUN.GA.98.L14              ACAACTCAG-CT-C--GTAGT-CCA-CTCCACTGCTAGTACAACTACTCCAATGCCTCTAGATTGG--T...--TA--...GACACAGAG--C--AGCGGAG---......-AT-A--TATGTAAAT-T-  6956
LST.CD.88.447              ACAACCACC-CC-A--GATCT-C-ACT-G-GGACCAGAA............TTAATAGAGGAC-T---T...--TAT-AAATGGGTAAATG-G-CT-CTGAG-TT......-AT----AGTGTAGAT-T-  5825
LST.CD.88.485              ACAACCACC-CC-A-GG-A-T-C-ACTAT-GAAAGT...............TTAATAGAGGAC-C---T...--TA--AGATGGTTAAATG-G-CT-CTGAG-TT......-AT----AATGTAAAT-T-  5822
LST.CD.88.524              ACTACTACT-CA-AG-GTTCAGGGACT--ACAACCATCT............TTAATAGAAGAT-C---T...--TA-ACGATGGTTAAATG-G-CT-CAGA--TT......-AT--G-AATGTTCAT-T-  5828
LST.KE.-.lho7              ..................ACAGTTAAGA-CCCGGGGAATGAAACA...AGATTAGAGGAAGATCTT--T...---A-AAGAGGGCTTAATG-G-C--CAGAG--G......-ATGC--AATGCCA-T-T-  6885
SYK.KE.-.SYK173            ACAAGTCCCCCT-CT-CACCTCC-AAT--AACGTGGTGGGGAGACAACAGTACAGAGCCA-G-TTC--T...------................................................TTT-  6684
SYK.KE.KE51_AY523867       ACAACAACA-AC-A-GCA-CA-C-CC-CCTGCATGGTGGGGG...GACAACTCTACAGAA-CT-GGT--AAT--T--ATTTAACATGACAGGGGGTTTT-AGGAT......-AG-A-CA-C--...T-T-  6419
COL.CM.-.CGU1              AGTGCTTCAGATTGGGAT---TC-..................AATTGGAAAGAATATCCGTGGT-T--T...---AGA......ATGAATTCA-CAGCTTTCTTA......C-T-A--A--GAAA-G--T  6133
```

```
                                                V2 loop
H1B.FR.83.HXB2             AAGAATATGCATTTTTTTATAAACTT...GATATAATACCAATAGATAATGATACT................................................ACC.......................  6791
Env                        K__E__Y__A__F__F__Y__K__L__.__D__I__I__P__I__D__N__D__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__.__.__.__.__.__.__.__. gp120

H1D.CD.84.84ZR085          ---T-C-----C---------G----...---G--G---A------------GGG-AAAAATGAAATCAAT........................GATACTTATGGT.......................  6333
H1F1.BE.93.VI850           --TT--C---GC---------GG---...------G---------GC---A-C-A-AGC.............................................-GT.......................  6113
H1G.SE.93.SE6165           -------C--G-----C----G----...---G--G-------TA------CA...................................................--T.......................  6215
H1H.CF.90.056              ---T-C-----C-C-------G----...---G--G-------T-----CA---G-................................................--T.......................  6122
H1J.SE.93.SE7887           ----G--C--GC----C----G--AA...---G--G-------T----G-A-C-A-................................................-AA.......................  6104
H1K.CM.96.MP535            -G----C----C----C----G----...---G--C-G---C-TA--GGC--GGG-AATAACAGTAGT....................................--T.......................  5999
H101_AE.TH.90.CM240        -G-TCC-----C-C--------G---...------G---A---T--AG--A-G-AGACTAGT..........................................-GT.......................  6371
H102_AG.NG.-.IBNG          -GATGC-----C---------G----...---G--G---A---TA--G-AA---A-GGT.............................................-GT.......................  6315
H103_AB.RU.97.KAL153_2     -----------C-C------------...---G--G---A--------------GCTAT.......................................................................  5995
H104_cpx.CY.94.CY032       ----------GC----C----G-A--...------G-G-----TA--GC-AGAGTGCCAATTAATGGTAGTAATAGGAATAAT...............AGTACAGAA.......................  6179
H1N.CM.95.YBF30            ---TT--CT-TC-G------GT-GAA...---G--G-------CA--GCC................................................TATAAT-AA.......................  6384
H1O.BE.87.ANT70            --A--C-G--TC-A--C---GT-TCA...---T-G--GGA-C-GA--G-GACA-GCAGCACAAAT...........................AAGACAAACAGC-AA.......................  6827
H1O.CM.91.MVP5180          --A--C-G--TC-A--C---GT-TCA...---C-G-GTAAGG-TA--G-CTCA-A-GCAGTAAAT....................................GGA--A.......................  6826
CPZ.CM.04.EK505            ---TC--CT-TC-G-------TGGGA...-----------CT----C-C-...............................................................AATAATAGCAGTGGTAA  6333
CPZ.CM.04.LB7              -GACT--CT----G------GT-GA-...---T---CT-AG---A-C--G..............................................................................AC  6331
CPZ.CM.04.MB66             -G-TG---T----G-------GG-AA...---C-TGAG---C----AGGA.........................................................AATAAGCCACCAGAAGGAGATAA  6429
CPZ.MT145                  ---TG---T-C--A--C---GTGGAG...-------C-G------GA---................................................................................  6315
CPZ.CD.-.ANT               --ATGA-A--CA-A---------GGA...---C-T--GAAGTGTC-GG-CA-C-A-GAGACTAAC.................................................................  6244
CPZ.CM.-.CAM3              ---T----T-CC--------GT-GA-...---C--G--TTG--T-----C-----CGAT.......................................................................  6231
CPZ.CM.98.CAM5             ---T----T-TC-A------GT-GA-...------G----T--T-GAGG-A---AGAGTAATGAT.................................................................  6501
CPZ.GAB_X52154             -G-TC---T--C-A------GTGGAG...---G-GG--AACC---GG-----G-ACAAC.......................................................................  6852
CPZ.GAB2                   --ATT---T-C--A--C-GG---AA-...---C-TG--GGT-CTA-C---................................................................................  6196
CPZ.TZ.-.TAN1              --AT----T-C--G-------G-GAA...-----TG--AA-GAG---GG-A-C...................................................-ATAAT....................  6435
CPZ.US.85.CPZUS            ---TG---T-C--A------GT-GA-...---C--GCT-AT--CA-----......AAT.......................................................................  6835
GSN.99CM166_AF468659       -GATG---AGCC-C--C-GG-G-GAA...--C--T--GGA-GAGACCC---G-.............................................AACCAG-GT.......................  6725
GSN.99CM71_AF468658        -GATG---AGCC-C----GG---GA-...-----T--GAGGGC-C-GG---G-.............................................AATGAC--T.......................  6716
H2A.ALI_AF082339           -CATGACA-G---ACAGAGAG-TAAGCCAA-ACAGTATAATGA-ACA...TGGTACTCAAAAGATGTGGTTTGTGAACCATTTAACACCACCACAAAC...CAG---.......................  7278
H2A.BEN_M30502             -TATGA-G-GG--AGAAC-AG-TAAGAAAAGG-GGTATAAGGACACA...TGGTA-TTAGAAGATGTGGTTTGTGACAACACAACA...............GCTGG-.......................  7297
H2A.ST_M31113              -TATGACA-G---AGAGAGGG-TAAGAAAA-AC--TATAATGA-ACA...TGGTACTCAAAAGATGTAGTCTGTGAATCAAATGACACCAAGAAAGAG......-AA.......................  6724
H2B.D205_X61240            GTATGACG-G-C-AAGAAGAG-TGAGCTAA-ACA-TATAA-GACACC...TGGTACTCAGAAGATTTAGAGTGTAATAATACCAGGAAGTATACC.........-G-.......................  7264
H2B.EHO_U27200             --ATGACA-G---AAAAAGAG-TGAGTCAA-ACA-TATAAGGACACC...TGGTA-AAACAAGATTTAGTGTGTGAAAAGGGGACAAGGAGTAAT......GAA-G-.......................  7259
H2G.ABT96_AF208027         -CATGACA-GGC-AAAAAGGG-T-AGAGAA-GGA-TACAATGAGAC-...TGGTA-TCACAAGATTTGCAATGYGAACAGTCCAATAAGAGTGAGAAT...GAG-GT.......................  6628
H2U.FR.96.12034            -TATGACA-GC--AAAAAGGG-TAAGAAAC-AGAGTATAA-GA-ACA...TGGTA-TCTCATGACATAGTTTGTGAACAGAATACTAGTGAAAGT......GAA-GT.......................  6757

MAC.239_M33262             -CATGACA-GG--AAAAAGAG-CAAGAAAA-AGAGTACAATGA-AC-...TGGTACTCTGCAGATTTGGTATGTGAACAAGGGAAT...AACACTGGTAATGAA-GT.......................  7215
Env                        N__M__T__G__L__K__R__D__K__K__K__E__Y__N__E__T__.__W__Y__S__A__D__L__V__C__E__Q__G__N__.__N__T__G__N__E__S__.__.__.__.__.__.__.__. gp120

SMM.SL.92.SL92B            -CATGACA-G-C-AAAAAGAG-C-AGAAAAGGCAGTACAATGA-ACC...TGGTACTCCAGAGATCTAGTCTGTGAGCAAGGAGGA............AATGAA-G-AGT....................  6651
SMM.US.x.PGM53             -CATGACA-GG--AAAAA-GG-CAAAAAGA-AGA-TATAATGA-ACA...TGGTA-TCAAGAGATATAGTATGTGAACAAAGTACTACTAGTAGAAAT...GAA-GT.......................  7149
MNE.US.-.MNE027            -CATGACA-GG--AAAAAGAG-CAAGAAAAGAGAGTACAATGA-AC-...TGGTACTCTGCAGATTTGGTTTGTGAACAAGGTAAT...AGCACTGAAGATGAA-GT.......................  6686
RCM.NG.x.NG411             CG-TGGCA-GG--AAAAAGAG-TGAAAAAATG-A-G--AATGACACA...TGGTACTCAAGAGATCTCTGGTGTAAGGCAGGTAATAATAATACT.........-GG.......................  6552
RCM.GAB1_AF382829          CT-T-GCA-G-C-AAAGAGAG-TGAGAAAC-C-AGTATAATGATACC...TGGTA-TCTAGAGACCTTTGGTGTGAAAAGGAAACA...AATTCTACAAATTCT--AAAAAAG.................  6508
DRL.FAO_AY159321           -T-TGACA-G---A-GCAGGG-CTG-AAG--AGAG--CA--GA-A-C...TT--GGTATGATGATGTAGTTTGT...AAATGTGGCAATGAAACTGAATCACCAGAAAGCAATAACTGTACTAATAAAAA  6550
MND-1.F17                  -CAT-ACA-G-C-A-GCAGGG-TTCGAAG--AGA----GT--C-A--...TT--GAGGGGATGATGTGAAATGTGAAAATAAT...............................................  6308
MND-2.5440_AY159322        -CACGACA-G---A-G-AGGG-CTGCAAATTGGA----GAGGA-A-C...TT--GATATGAAGATGTAACTTGC.........ACAAAGCTTAATAAAACTGGT-GTGCTACTAATAGTACTGAACCTGA  6507
MND-2.14CG_AF328295        -CAC-ACA-G---A-G-AGGG-TTG-AAAATAGAG--TAA-GAGAGC...TT--GATATGAGGATGTGACCTGT.........ACTAAAAAGGAAAATGACACTGAA.....................AC  6806
MND-2.MND2_AF367411        -TACTACA-G---A-G-AGGG-TTGCAAATTGGAG---GA-GA-A--...TT--GATATGAAGATGTAACTTGT.........ACAAATGGCACAAATAACACATAT.......................  6884
DEB.CM.CM40                -GAGC...T--GCAG-A-GGCT-GAGGAT-----CCAGTGGGCG..................................................................GATAATAAAGATGGTAGTGG  6524
DEB.CM.CM5                 TGC--...AGTGCAG-G-GGTT-GAGGAT--C---CAGTGGGAG.........................................................ACAGAAAACCAAACAAATACTAACAAAAG  6475
MON.CM.L1                  GCTT--T-TGGAAGGAAG-C-T-A-GAAG--GGA-GG-AGC-AT-GC...................................................................................  6733
MON.NG1                    GCTTG-T-KGGAAGGAGG-C-TTA-GCAA--C-AGGATAAC-C-AC-...................................................................................  5346
TAL.CM.8023                CTC-T-GGTATC-AGCAG-CTT-A-GAAG--ACC-GGCAATGGCA--...................................................................................  6120
TAL.CM.266                 CTC-C-GGTACC-AGCAG-CCT-A-GAAG--ACC-GGCAATGG-TCC...................................................................................  6582
MUS.CM.1085                C-......TTG--C---GT-G-TGA-CTGATG-G-GACAGCTC---G......-G-GAGAATGAT.................................................................  6689
SAB.SN.-.SAB1C             -TT-C.................................T---C--TA...TGGGA-GATCAAGAGGTGGTTTGTGAAGAG...GGAAGGGAAAAATCAAAC............GCAACA.........CA  7070
GRV.ET.x.GRI_677           C-ATTGCA-G--A-CAAAGAG-TG-AAAAA-G-ATTATAGC--GACC...TGGTA-GATCAGGAGTTAGTCTGCAATAATAAAACAGGAAGTGAAAAGGGAAGT-AG.......................  6774
TAN.UG.x.TAN1              CTATGGCA-G--A-AGGAGAG-TG-AAAAA-G-A-TATAAT-GTAC-...TGGTA-GATCAGGAGTTGGTATGTGAGAAGGAAAACAACACCACA...GGCACG-GA.......................  6866
VER.DE.-.AGM3              -GT-T.................................T--G-G-TG...TGG-A-GATGCAGAAATCATGTGTAAGAAG...GGT......AACAAT..............................TC  6394
VER.KE.-.9063              -TT-T.................................T--G---TA...TGG-A-GATGCAGAGATCTATTGTAAGAAC...AGCTCTAGCTCTAACTCT...........................AG  6896
VER.KE.-.AGM155            -TT-C.................................T--G---TA...TGG-A-GATGCAGAAATCTTTTGTAAGCGT...AGTACA...TCGCATAAT...........................GG  6876
VER.KE.-.TYO1              -TT-T---..............................T-TG-G-TG...TGG-A-GATGCAGAAATCTATTGCAAAAAT...AAGACT......AATAGC...........................AC  6368
SUN.GA.98.L14              -T-TGACA-GGC---G-AGAG-TTG-AAAACGGAGG-GGA-CA-A-C...TT--GGGATACAGAAGTTACTTGC............AATGGTAATGAC................................  7039
LST.CD.88.447              -T-TTACA-G-C-A-G-AGGG-CTG-AGGACAGA----AAGCAGAG-...TT--GGTATGATGAGACAACATGTGATAAGAATGGA............................................  5908
LST.CD.88.485              -T-T-ACA-G-C-A-G-AGGG-CTG-AAA--AGA----AAGCAGAG-...TT--GGTATGATGAGACAACCTGTGATAAGAATGGA............................................  5905
LST.CD.88.524              -T-TGACA-GGC---G-AGAG-TTGCAAG--AGAG---AAGCA-A--...TT--GGTATGATGAGACAACATGTAATGAGAATGGT............................................  5911
LST.KE.-.lho7              -T-T-ACA-GG--A-GCAGAG-CTGCAGGACGGA----AA-CA-AGC...TT--GATATGATGATGTAACGTGCTCAGGAGAGAGGGAGAACAGA...................................  6977
SYK.KE.-.SYK173            -CTTGACA-GGGGG--CA-GG-TAAAAAAC-GCA-TATAG-GCTTTC...TTCTA-AAGGATGACTTGATGAAAGAGGAAGGCAAT............................................  6767
SYK.KE.KE51_AY523867       -T.........GCA--C-T-T-CAAGTCA--------GAG-GAG--A...-GA-A-GAAAGT....................................................................  6469
COL.CM.-.CGU1              T-----TA-G-----CAGTAG--GA-TTAAC-G-GC-TGG--AT......A-A-A-GACAGTAATAGT....................................-T-.......................  6198
```

```
                               V2 loop              -|
H1B.FR.83.HXB2              .......AGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTAATAAT  6914
Env                         __.__.__S__Y__K__L__T__S__C__N__T__S__V__I__T__Q__A__C__P__K__V__S__F__E__P__I__P__I__H__Y__C__A__P__A__G__F__A__I__L__K__C__N__N_ gp120

H1D.CD.84.84ZR085           .......-CT----G---A-T--A------T-------C---------A--T--------A--------------------------------------A-----------A--------G-----CG--  6456
H1F1.BE.93.VI850            .......GAA----G-C-A-T--A------T------AC------------T--------------T-GG--T-----C--T--------------T--A-------A---A---------------G--  6236
H1G.SE.93.SE6165            .......-AT----G-C-A---CA------TGT----AC---C-A------T-----------A-T-----C-----------------------T--AA--------------T----G----GGG--  6338
H1H.CF.90.056               .......CAG----G-C-A-T--A------T--------------------T--C--------G--------A--T--------------------T--A-----C--------------G-----C---  6245
H1J.SE.93.SE7887            .......-AT----TTC-A-T--AC-----T-------------A---A--T-----------G--T---C-A--------------------------A---------------------------G--  6227
H1K.CM.96.MP535             .......GAA--C-G-C-A-T--A------T------AC----------A-T--------A---A----------------------------------A-----C--------------G-----GG--  6122
H101_AE.TH.90.CM240         .......GAG----G---A-T--A------T--T----------AG-----T---------A----------T--------T------------A-T--A-------A-------T----G------G--  6494
H102_AG.NG.-.IBNG           .......CAG----G---A-T--A------T-------C------------T-----------------------------------------------A--------------------G-----GG--  6438
H103_AB.RU.97.KAL153_2      .............-G-----T---------T---------G-------A------------A-------------------------------------A--------------------G-----CG--  6112
H104_cpx.CY.94.CY032        .......GAG----T---A-T--A-------G-----AC-----A------T--C--------G--T--------------------------------A-----------A---T----G------G-G  6302
H1N.CM.95.YBF30             .......-CA----G-C-A-T--A------T---A---CTG-G-----A--T-----T---ACT-----------------A--------C-----A--AC-A--C-----C---A-G---------G-A  6507
H1O.BE.87.ANT70             .......-TG----CA--A--T-A-------T--A--AC---C--G--A-----------------T-----A--------A-----C--------T--A--A--A-A---T--CT-T--G-----C-GC  6950
H1O.CM.91.MVP5180           .......-CA----T---A--T-A-------T--A--A-T--C-AG-----------G------AGT--------C-----------C--------T--AA-A--A-A---C--CT-T--G------G-C  6949
CPZ.CM.04.EK505             TAACAGTCAA----G---A-T--A----------A---CTG-G--------------T--AA-T-----------------A---T----------A--AC-A--A-----C---A-----------G--  6463
CPZ.CM.04.LB7               AACCGAG---------C-T-T--A------T--TA---C---A--C-----T--C------A-T--T-----------A--T--------------T--A-----------TC------GG-----C---  6461
CPZ.CM.04.MB66              GAACGCCTTA----GA--ATAT-A----------A---C---G--T--A--T--CT----A--C--TC----A--------T--------------T--A-----A-----CC-CT-G--G------G--  6559
CPZ.MT145                   .AATAGT-CT--C-G-----TT-A------T---A---C---C--T--A--T---------ACT--------A--------A--------C-----A--A--A--A-----AC-C------------G--  6444
CPZ.CD.-.ANT                .......T-----T-C--ATGGCAC--C------A--ACT--C-----AT-----GA----TCTA-T-----A--------T--------------T--A--A--A-A---T--AT-G-G----G-AG--  6367
CPZ.CM.-.CAM3               .......-CA----GAC-A-T--A---C--T---A---CT-----------G-----T---ACC-----------------A--------C-----AA-CC-A----A---A--CA-G-----C----TG  6354
CPZ.CM.98.CAM5              .......-C-----GA--A-T--A------T--TA--ACTG-A--------A-----T--AACT-----------------A--------C-----A--CC-A--------A--AA-G---------G--  6624
CPZ.GAB_X52154              .......-CA----G-A-A-TT-A---C--T--TA---C---A-----A--A--------AACG-----------T-----A-----C--------A--A--A--G-----A-----T-----C---G--  6975
CPZ.GAB2                    .......-CG-T--G---A-TT-A---C--T--TA---CA--C-----A--T--------AACC--------A--------A--------C-----A--A-----C-----CC-GT-G---------G--  6319
CPZ.TZ.-.TAN1               .......--T---T-T--ACAT-A---C--T--------------T--A-AA---GAT--ATCTA-T-----A-----------CAGA--C-----T--A--A--C-----CC-GT---------GAG--  6558
CPZ.US.85.CPZUS             .......-C-----GAC---T--A---C------A-C-CT--C--------T-----T---ACC-----------------A-----C--------A--AC-A--C-----C--C------------G-A  6958
GSN.99CM166_AF468659        .......G-A---T-TA-T-G--A------T------TAT--C--T--AAAA---GTC---TCC-----CC-A--GG----------C--C-----T--AC-C--C-A---T--G--------C---G--  6848
GSN.99CM71_AF468658         .......GA----T-TA-T-T--A---C-----A---TATG-C--G--AAAG---GTC---ACC-----CC-A--GG----------C--C-----T---C-C--C-----T--G--------C---G--  6839
H2A.ALI_AF082339            .......--G-G-T-CA---ACCA---C-----A--------C---G--T-A---GAT---CAC-AT-GG--TG-T--AAGGT-TAGA--C-----A--AC------AC--CC-A----G---CG--G--  7401
H2A.BEN_M30502              .......-CA-G-T-CA---G-CA---C-----A---A----C-A-G--T-A---GAT---CAC-AT-GG--TG-T--GAGGT-TAGA--C-----A--AC-G--C-----CC-AT---G-------G--  7420
H2A.ST_M31113               .......-CA-G-T-CA---ACCAC--C-----A--------C---G--T-A---GAC---CAC-AT-GG--TA-T--GAGGT-TAGA--------A--AC-G--------CC-G----G---C---G--  6847
H2B.D205_X61240             .......--A-GCT-TA-A-G--CC--C-----AA--A-T--CCA-G--T-A---GAC--ACAT-AT-GG--CAGCT-AAGGT-TAGG--------T--CC-G--G---TTTC-A----G-------G--  7387
H2B.EHO_U27200              .......-AA-G-T-TA-A-A--CC-----T--A---A-T--CCA-G-AT-A---GAC--ACAT-AT-GG--TAGTT-AAGAT-TAGG--------T--CC-A--A-----TT-G----G---C---G--  7382
H2G.ABT96_AF208027          .......--A-GCT-TA---GGCAC-----T--A---------CAGG-AT-A---GAT--ACAT-AT-GG--TG-T--AAGAT-TAGA--------T---C-A--G-----TT-G----G-------G--  6751
H2U.FR.96.12034             .......-AA-G-T-CA---ACCAC--C--------T--T--CCA-G--T-T---GAC--ACAT-AT-GG--TG-C--AAGAT-TAGG--------A---C-A--C-A---TC-GT---G-------G--  6880

MAC.239_M33262              .......--A-G-T-CA---ACCAC--------T--T--T--CCA-G--T-T---GAC--ACAT-AT-GG--TG-T---AGAT-TAGG--------A--TC-A----A---TT-G--T-G-------G-C  7338
Env                         __.__.__R__C__Y__M__N__H__C__N__T__S__V__I__Q__E__S__C__D__K__H__Y__W__D__A__I__R__F__R__Y__C__A__P__P__G__Y__A__L__L__R__C__N__D_ gp120

SMM.SL.92.SL92B             .......--A-G-T-CA---ATCA---C--T--AAGT--T--ACA-G--T-A---GAC---CAC-AT-GG--TG-C--CAGATATAGG--C-----A--AC-A--C-A---TT-G--T-G-------G--  6774
SMM.US.x.PGM53              .......-AA-G-T-CA---ATCA----------AGT--T---CA-G-AT-----GAC---CAT-AT-GG--TG-T---AGAT-TAGA--C-----A---C-A----A---TT-G--T-GG------G--  7272
MNE.US.-.MNE027             .......--A-G-T-CA---ATCAC--------T--T--T---CA-G-AT-T---GAC--ACAT-AT-GG--TG-T---AGAT-TAGG--------A--TC-A----A---TT-G--T-G-------G-C  6809
RCM.NG.x.NG411              .......CAA-G-TTTA-A-GGCA---------AA--AGT--ACA----TT----GA-CCAAA--A--GG--A---T--AGGT--AGG--------A---C-A----A---C--A--G------CT-G--  6675
RCM.GAB1_AF382829           .......-AA-G-TTTG-A-GGCAC--C--T--AA-TTC---ACA---ATTT---GA-CCAAAG-A--GG--A---T--AGGT--AGA--------T--AC-A--G-----CT-A--GGTC--C--AG--  6631
DRL.FAO_AY159321            AAGA...-CA-G-T-TG-----C----C---T-AA----G-----C---AT----AAT----CC---ACA--TATG--GAAAT-CAGA-TG-----T--AC-A--A-A-ATAT-G----GG------G-A  6677
MND-1.F17                   .......-CT-GCT-TA---ATCA------TGAG-----T-A----G-A-A-----AG----G-CTT--GATA...-GATGT--TTTAGG-----TG--TC-A--A-A--TC--GT---G--A----G-G  6428
MND-2.5440_AY159322         ATAT...GAG-G-T-CA----TTCA------G-AA----A--A-------A-C--AAC--A-CT---ACA--TAG---GA-AT-TAGG-TG-----A--CC-A--G----TCT-A------------G-A  6634
MND-2.14CG_AF328295         TGAA...GAA-G-T-TA-----CAC-----T--A-----A--A-------AT---AAT----CC--AACA--CAA---GA-AT-TAGA-TA--------TC-A--G-ACATTC-G----G----CG-G-G  6933
MND-2.MND2_AF367411         .......TCT-G-T-TA-----CAA--C--T--A-----A--A--C----A----AAT--A-C----ACA--TGA---AAAAT-TAGACTG-----A--TC-A--A-A--TTC-CT-G-GG--CCTAG-A  7007
DEB.CM.CM40                 GAACAGA-CAGGGT-CA---A-CAC-----TGA---T-----C--C---T-T---GA--CTAGCCG----A-A--TT--AAG--CAGG--C-----A--A-----A-A--GTC-A----G----G--G--  6654
DEB.CM.CM5                  TATAAGACATGGGT-TA---ACCAC--C----AT-----T--A------T-----GA--CCAGTCG----A----TT--AAG---AGA--C-----A--A-----A-A--GTC-GT---G----G--G-C  6605
MON.CM.L1                   ....AGTCA----T-CA-CCTG-A---C-----A---TA---A--C--A--T---GAG---TCCAAT-A-------G----TT-G-----------A--CC-A--C-A---TT-AT-G-G----G--G-C  6859
MON.NG1                     .GATGACTAT---T-TC--CT--A------T----Y-TA---Y--C--A------GA----TCCAAT-A------T-----T--T-----C-----T---C-C--C-A---T--AT-G-G----G--G--  5475
TAL.CM.8023                 .GACTCATA----T-TGCA---G-A--C--TTTG-----A--AG-T-----T---GA----AC-CAT-A-C-AG-GT----A--T--G--C-----T--A--A--C-----CC-GA---G-------G-C  6249
TAL.CM.266                  .......TT----T-TGCC---G-A--C--TGTG-----T--AG--A--T-----GA---AACTCA--ACC--G--T----T-----G--C-----T--A--A--C-----CC-GA---G-------G--  6705
MUS.CM.1085                 .......--T---T-TA-C---AC---C-----A---TA---C--C-----A---GAG---AGCAG---CC-A--GG----G-----A--C-----A--C--G--G-ACAGCT-G--------C---G--  6812
SAB.SN.-.SAB1C              TACTGTGG-A-G-T-CA---TCCA------T--------A--C-A-G----T---GAT--AACT-AT-GG--TA-CT--AGGC-TAGG--------T--A--A----A---TT-G----G----GCAG--  7200
GRV.ET.x.GRI_677            .......GAT-G-T-CA---T-CA------TGAT----G-A-A-G-A--T-----GAT--AAC--AT-GG--TA-TT-AAGAG--AGA--C-----A--A--A--G-A---TT-G------------G--  6897
TAN.UG.x.TAN1               .......G-T-G-T-CA---TTCAC--C---GA---T--A--A-A-G-A--T---GAG--AACT-AT-GG--TA-CT-AAGAT--AGA--C-----A--A--A--C-----T--CT----------GG--  6989
VER.DE.-.AGM3               TAACAGAGAA-G-T-TA---TTCA------TGAT-----T--A-A-G-A------GAT--AAC--AT-GG--TGAGT-AAGAT--AGG--C-----------A--G-----TT-AT-------C--CG--  6524
VER.KE.-.9063               CACAAAAGAA-GCT-TA---TTCAC-----TGAT-----T--A-A-G-A------GA---AAC--AT-GG--TGAGC-AAGGC--AGA--C--------A--A--A-----TC-GT-------CC--G--  7026
VER.KE.-.AGM155             GACAAAAGAG-GCT-TA---TCCAC-----TGAT-----T--A-AGG-A--T---GAT---AC--AT-GG--TGA-T-AAGAC--AGA--------T--A--A--A-AC--TT-G--T---------G--  7006
VER.KE.-.TYO1               TAGCAAAGAG-G-T-CA---TTCA------TGA------T--A-A-G-A--A---GAC--AAC--AT-GG--T-AGT-GAGGT--AGG--------T--A--A----A---TT-G------------G--  6498
SUN.GA.98.L14               .......-CA-G-T-TA-----CA---C--TGAT---A-T--------A-A---C-AT----GTATTA-GC-AAATGC-TA-T-CAGGCT------A--T--G--G-A-AT-T-AT---G----G--G-A  7162
LST.CD.88.447               .......-CT-G-T-TA----CCAC-----TGATAGTA-T---CAG--A-AT--CAAT----G-GTAA-ATCTAATGC-TATT-TAGGCT------A--A--A--C-A-AT-C--T----G--C--CG-A  6031
LST.CD.88.485               .......-C--G-T-TA----CCAC-----TGATAGTA-T---CAG--A-AT---AAT----G-GTAA-ATCTAATGC-TATT-TAGGCT------A--A--A--C-A-AT-C--T-------C---G-A  6028
LST.CD.88.524               .......-CT-GCT-CA----CCA---C--TGATAGTA-T---CTG--A-A---CAAT----G-GTTA-GTCAAATGC-TATT-TAGGCT------A--A-----C-ACAT-C--T-------C---G-A  6034
LST.KE.-.lho7               .......-C--G-T-CA----TCA------TGAT---A-A--A-----A-AT---AAT----GGGTAA-GC-AAATGC-TATT-TAGGCT------T--A--A--C-A-AT-C-AT-G-G---C---G-A  7100
SYK.KE.-.SYK173             .AGTAGCTAT---T-CC--TT-CA------T--AAGT--T--AT-CGCA--A---GAG--ACAGA-----C-A--CT----G--T--A--C-----T--AC-G----ACT-AT-GT-G-----C---G--  6896
SYK.KE.KE51_AY523867        .......TAT--CT-T--ACAGCA------T--TAGT--A--GCGCA-T--T---GA---ACAGA-T---CGC---T-------T--A--C-----T--AC-G--A-ACT-AC-GT-G---------G--  6592
COL.CM.-.CGU1               .......--GGCA-CAA---A-GA----GCA-AT-ATACAG----C--A-TA---GAT-T-ACTATAG-G--T---G--AGG-C-GGG-TC-----AG-TC-A--A-ACAT-C--T---G----G--G--  6321
```

```
H1B.FR.83.HXB2          AAGACGTTCAATGGAACAGGACCA......TGTACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAA...GAAGAGGTAG  7035
Env                     _K__T__F__N__G__T__G__P__.__.__C__T__N__V__S__T__V__Q__C__T__H__G__I__R__P__V__V__S__T__Q__L__L__L__N__G__S__L__A__E__.__E__E__V__ gp120

H1D.CD.84.84ZR085       --A-G---------------C---......--C-A----------T-------------------------------G--G------------------------------------...------A-C-  6577
H1F1.BE.93.VI850        --A-GA--------G-----G---......--C-AG----------------------------------AA-----G------------T--T-------------C---------...----GTA---  6357
H1G.SE.93.SE6165        ---GA-------------------......----A---------T-------------------------A------G--------------A--AC-G---------T--------...-G-A-AA--A  6459
H1H.CF.90.056           --A--A---------------TT-......--------------T--------------------------A-----G-----------------A--------A--C---------...---C--A-CA  6366
H1J.SE.93.SE7887        ----AT---------------T--......--C-A---------T-----------------C-------A------G--------------A---C-----------A--------...-G---CA--A  6348
H1K.CM.96.MP535         --A-G---------G---------......----A-----------------------------------AA-----G-----------------------------C---------...------A--A  6243
H101_AE.TH.90.CM240     ----AT--------G-----G---......----A----------T----------C-------------A------G------------T--------------------------...------A--A  6615
H102_AG.NG.-.IBNG       --AGG---------------G---......--C-A---------------------C-----------C-A----------------------------------------------...-G---A----  6559
H103_AB.RU.97.KAL153_2  --A-A------C--------G---......--------------T-------------------------A---------------------------------T------------...----------  6233
H104_cpx.CY.94.CY032    --A-AT----C----TT---G---......--C------------T-G----G---C--T----------A------G------------T-------------A--CT-----ACG...----------  6423
H1N.CM.95.YBF30         GGA-AC--T-G-----AT---AGC......-----------G--T--T--------C-----------A-A------GA----C-----GT-AA-CC-------A--CT--AATAC-...--T-GAA-T-  6628
H1O.BE.87.ANT70         -CAGAA--T-----------CA--......--C-G---CA-A-CGGT---TACT-----------C---------AC----AGT-----G--AA-A--------G-CA--CT-TA--...-G-A-AA--A  7071
H1O.CM.91.MVP5180       -CAGAC--T-----------C-T-......--CCAC---A-TTCAGTG--TACT-----------C--C-A----AC----AGT--------AA-AC-G-----G-CA--CT-TAG-...---A--A--A  7070
CPZ.CM.04.EK505         C-AGAT--T--------T--GGA-......----AC-----A--T--T-----------C--------A-A------GA----C--A--GT--A-CC----------CT-----AC-...TC-A-TA-C-  6584
CPZ.CM.04.LB7           --A--A-AT-----T-GT--T---......----AC-----T--T--------C-----T----------AA-----TA-T--T------T--A-CC-G-----T------------...---A-AA-CA  6582
CPZ.CM.04.MB66          --A-AT--T-C---G-T-----AG......----A------T--T--T--G--T-----------C--C--A--T-----C--------GT-T--T--------A-CCT---A----AAG-TGACA--CC  6683
CPZ.MT145               -TAGAT-A---A--G-AT-A-A-T......----A------A--T--------T-----------C----AA-----G-C-A-C--C------A-T--G-----G---AC------C...A-TC--ACT-  6565
CPZ.CD.-.ANT            G--GAT--T-CA---GT---GATG......----A-------TCAGT------T--C--T--------A--C---A-G--GG----ATGGT-A--A--------A-C-TACCA-AC-AACACTTCA----  6491
CPZ.CM.-.CAM3           CCA-AC--T------------A--GGAAGA----AC---A-A--------C--------T-----C--C--A-----G---A-T--A---T--A-CC-T-----A---G----C---...A-CA--ACTA  6481
CPZ.CM.98.CAM5          C-A-A---T---------------......---------A-A--------------C--T-----C--A--A---------A-T--A---G--A-CC-T-----A--C-----T---GAG-GGCCC-CT-  6748
CPZ.GAB_X52154          --AGAC--TTCA----A---TAA-......-----T-----A--T-----G--T----------------AA-----G--TA-C-----GT-A--TA-T-----A---T----T---...-G-A-CA--A  7096
CPZ.GAB2                --AGAT-ATCCA----A----AAG......----A------A--T-----G--T--C-------------A----ACT---A-T-----GT--A-AC-C--------CT--------...-----A-AG-  6440
CPZ.TZ.-.TAN1           C---AT----CA--G-A-----A-......--CT-C------TCAGT---T--C-----------G---TAT--TA-GA--G-C--AGC-T-A-AC--------GTCC--G-A----...-----AAC-A  6679
CPZ.US.85.CPZUS         --AGAT-----A----AG---GAG......----A------T--T-----G--G-----T-----C--A-AA-----G--GA-T--A--G--CA-AA-----------T-----ACT...A--A-T--TA  7079
GSN.99CM166_AF468659    GTT-AT----CA---GT---CA--......--CTAC-----GTCAG-----ACC--C--------C---CAA--CT-----G----ATGG--C-AT--------G-CCTATCAGCC-...-G-A-T...A  6966
GSN.99CM71_AF468658     GCC-AT----CA---GTG--CAAG......---T-C-----ATCAG-----AC---C--G-----C---CAA--TT-----G----ATGG--T-ATC-------G-C-TATCAGCC-...-GGA-C...A  6957
H2A.ALI_AF082339        -TC-AT-ATTCA--CTTT-C---C...AAT--CT-T--A--AGTAG-T-CTAC---C---AGGATG--GGA-A-GCA-ACT--T---TGGT-TGGC--T-------C-AGG------...A-TAGAAC-T  7525
H2A.BEN_M30502          -CC-AC-ATTCA--CTTT-A---T...AAG--C--T--A--AGTAG-T-CTTC---C---AGGATG--GGAAA-GCA-ACT--T---TGGT-TGGC--T-------C-AG-------...A-TAGAAC-T  7544
H2A.ST_M31113           -CC-AT-ATTCA--CTTT-AG--C...AAT---T-T--G--AGTAG-T-CTAC-------AGGATG--GGAAA-GCA-ACC--C---TGGT-TGGC--T-------C-AGG------...A-TAGAAC-T  6971
H2B.D205_X61240         -CC-AC-ATTCA--CTTCATG--C...AAC--C-GT--G--AGTAG-GTCCTCC--C---AGAATG--GGAAA--CAGTCC--T--ATGGT-TGGC--C-----T-CAAGG-----G...A-CAG-AC-T  7511
H2B.EHO_U27200          -CT-AA-ATTCA--CTTCATG--C...AAC----GT--G--AGTAGT-TCT-TG...TACAGAATG--GGAAA--CAGACC--T--ATGGT-TGGC--C-----T-CAAGG------...A-CAG-AC-T  7503
H2G.ABT96_AF208027      -CT-AC-ATTCA---TTT-C---T...AAT--CT-T--A--AGTGGTCTCTAC-------AGAATG--GGAAA--CA-ACC--T--ATGGT-CGG---T-----A-CAAGG------...A-TAGAAC-T  6875
H2U.FR.96.12034         -CT-AT-A-TCA---TTTATG--T...AAT--CT-T--A--AGTAGTCTCTAC---C---AGAATG-GGG-AA--CA-ACC--T---TGGT-TGGA--T-------C-AG-T-----...A-TAGAACTT  7004

MAC.239_M33262          -CA-AT-ATTCA--CTTTATG--T...AAA---T-T--G--GGTGGTCTCTTC---C---AGGATG--GGA-A--CAGACT--T---TGGT-TGGC--T-----A-C-AG-------...A-TAGAACTT  7462
Env                     _T__N__Y__S__G__F__M__P__.__K__C__S__K__V__V__V__S__S__C__T__R__M__M__E__T__Q__T__S__T__W__F__G__F__N__G__T__R__A__E__.__N__R__T_  gp120

SMM.SL.92.SL92B         -CA-AT-ATTCA--CTTT-C---T...AAT----GT--G--AGTAGT-TC-TC-------AGGATG--GGAAA--CAGACT--T--ATGGT-TGGC--T-----T-C-AG------G...A-TAGAAC-T  6898
SMM.US.x.PGM53          -CA-AT-ATTCA--CTTT-CT--T...AAC---T-T--G--AGTGGTTTCTTC---C---AGAATG--GGA-A-GCA-ACC--T---TGGT-TGGT--C-----T-C-AGG------...A-TAGAAC-T  7396
MNE.US.-.MNE027         -CA-A--ATTCA--CTTTATG--T...AAC---T-T--G--GGTGGTCTCTTC---C---AGAATG--GGA-A--CAGACT--T---TGGT-TGGC--T-----A-C-AG-------...A-TAGAACTT  6933
RCM.NG.x.NG411          GCA-AT-AT-CA--CTTT-ATA-C......---CA------T-CAG--ACGTC---C-----CATG--A-ATA--AC------C--ATC-T-TGGA--------ATC-A-TAGT--G...A-TAGAAC-T  6796
RCM.GAB1_AF382829       --A-AT-AT-CA--CTTT-ATA-C......---GTT-----T-CTG-CACTTC---C------ATG----ATA-TACT--GG-CT-AGGGT-TGGA--------ATCAA-TAAT-T-...A-T---ACTT  6752
DRL.FAO_AY159321        --ACTAAAT-TAA-T-GTAAG.........----A----A-A-CAG----G--G-----TAGACAT--GCCTG-CAC-A-C--T-GC--TT-TGGC--C-------CAAAGCAT-GTCAG--T---T--A  6798
MND-1.F17               ---TTAAAT---AAT-A-TTG.........----GC---A-ATCAG----G--G-----T--GCACT-AGTAG-CAC----AGT-GCTTTT-TGGC--T-----A-C-A-GCATA-G...----GA-A-T  6546
MND-2.5440_AY159322     --ATT-AA---GAC--A-TT-.........---GGG-----GTCTG----G--G--C---GCACCGC-GCCAG-CAC-A-C--T--AATGT-TGGC--C-----A-C-AAGCAT--TTAT--T---C-TA  6755
MND-2.14CG_AF328295     --ATT-AAT--AACT-A-TT-.........---G-C------TCAG-------G-----TG-CCC---GCCTG-TACTA--------ATGT-TGGC--T-----A-CAAA-CATA-T...T-T--T-AGT  7051
MND-2.MND2_AF367411     -GATT-AATGTCTCT-A-AA-.........-----C---A---CAG-------G--------GCC-T-GCCAG-CACCACC------ATGTGTGGA-CT-G---G-CAAA-CAT--TTATA-T--AT--A  7128
DEB.CM.CM40             --A-AT--T-----T-----GTT-......--C-AC--C--G-CTG----GGC-------A--CTC--ACACA-CA-G-C---C--ATGGG-A-AA--T-----GTC-GAT-AG--GAGG-C---A-AGC  6778
DEB.CM.CM5              -GA-AC--T-----T--C--CTTG......-----C-----G-CTG----GGC-------A--CTC--ACATA-CA-G-C------ATGGA---AA--C-----TTCAGAT-A----AGG-C---A-ACA  6729
MON.CM.L1               CCCG-C--T-CG---CA---GAGT......---T-T-----TTCTG-----AC------T----CT--CCAA--TA-----G-T---TGGT-C-A---G--CA-T-CGGGTAAT-CT...CCCA-TAC-A  6980
MON.NG1                 CCAK-C----CA--GCAGAACAAT......--CT-R------TCAG-----AC---C--G-----T--A-T----A-------T---TGGT-T-A-C-T----CT-CCTCT--CAC-...A-TR--ACCA  5596
TAL.CM.8023             --CC-A-GG-CA--T-G---CAAG......---CAT--CA--TCAG-------C--C---G--CCC--C-CCA-CA---C-G----ATGGT--G-TC-------T-CAAAG-A-A-T...C-GACACA-A  6370
TAL.CM.266              -GCC-A-GG-CA--T-G----AAG......---CAC--CA--TCAG-------C------G-CCCT--C-CCA--A---C-G-T--GTGGT-AA-TC-C-----G-CAAAG-A-A-T...C-CACACAG-  6826
MUS.CM.1085             GCC-AT--T-C---GCA----GA-......----AT-----G-CTG-TAGG--T--C--G-----C--ATTA--CT-G--TG-T---TGG--A-A-C-T-----A-C-TATTT---GGGAA-TA-TAC--  6936
SAB.SN.-.SAB1C          -CAGAT-ATTCA---CATAAGG--......----GG-----A-CAGT-TC-GCT--C--TAGGTT---C-ATA-TACT---AGC--AGG-A-AGGAA-T-----T---TAT-T--C-...A-TAGAAC--  7321
GRV.ET.x.GRI_677        ---GAT-AT-GA--CTTT-CT---...AAG--C-AG-----TTCAGT---G--T-----TAGATT---C-ATA-TACTA--A-T--AGGGA-AGGA--------T---AG-T-----...A-TAGAAC--  7021
TAN.UG.x.TAN1           -CT-AT-AT-CA---TTT--TGTT......----G------TTCAGT---TAGT-----TGGATTG--G-ATA-TAC---GAGCT-AGC-T-TGGCA------------AG------...A-CAGAAC--  7110
VER.DE.-.AGM3           T-TGAT-ATGCA--GTTTAAGA--...AAC---T-T-----TTCAGTG--G--T-----TA-CTTG--A-ATA--AC---GA-T---GG----T----G-----G--CTACT----G...A-TCGAACCC  6648
VER.KE.-.9063           C-TGAT-AT-----TTATAAG-A-...AAT----GT-----TTCAGT------C-----TGGACTT--G-ATA--AC---T--T--AGG---A-----G-----A--CTATT-T--G...A-TCGAAC-C  7150
VER.KE.-.AGM155         TG-GAT-ATGCA---TTTAAG---...GAA---T-T-----TTCAGT---G--T--C---AC-TT---G-ATA--AC----A-C---GGT---T-A--G-----A--CTATT-----...A-TCGAACCC  7130
VER.KE.-.TYO1           G-AGAT-AT-----GTATAA--A-...AAT--CT-------ATCAGT---G--T------GGCTT---G-ATA--AC---GA----AGGGT--T--C-G-----A--CTATCAT--G...A-TCGAACCC  6622
SUN.GA.98.L14           --ATTAAATGC-ACT-AGAA-.........----A----A-A-CTG-TACC-CT-----TA--TAT--G-CAAGTAC---TAGT-GCTTTT-TGGC--T-----A-C-AG-CATA-G...-----T-AGT  7280
LST.CD.88.447           --ATT-AATTTCTC-G-TAAT.........-----C---A-A-CAG--ACC-CT--C---G-CTAT--G-TTAGTTC---GAGT-G-TTCT-TGGC--T-----A-CCAATCATAC-...--C--T-A-T  6149
LST.CD.88.485           --ATT-AATTTCTC-G-TAAT.........-----C---A-A-CAG--ACT-CT--C---G-CTAT--G-TTAGTAC---GAGT-G-TTCT-TGGC--T-----A-CCAATCATAC-...--G--TAAGT  6146
LST.CD.88.524           ---CTAAA-TT-TC-G-TAAT.........-----T---A-A-CAG-TAC--C---C---G--TAC--G-TAAGTTC---TAGT-GCTTCT-TGGC--T-----A-CCAATCATAC-...-----C-AGT  6152
LST.KE.-.lho7           C-ATTAAA-TT-A-T-A-AA-.........---GA----A-A-CAG-TAC--C------TGGATAT--GCTTAGTAGT---AGT-G-TTCT-TGGC--T-----A-C-AATCATACC...AGG--T-AGC  7218
SYK.KE.-.SYK173         -CA-AT--TG-A---GAC-ATGTC......-----------A-CTG-----TC---C-----G-A-T-C-ATA--T-G-CTAGC--ATGGT-C-AAC-G-----G-CCTACAA--C-...A---TAA--  7017
SYK.KE.KE51_AY523867    -CAGAA--TG-A---GAT-CTGAG......---CA------A-CAG----TTC---C------CCTT-C-ACA-TT-G-CTAGT--ATGGT-T-A-----------CCTATAA--C-AAA--CA-A--TA  6716
COL.CM.-.CGU1           --A-A--GGG----G------G--......--C-AC-----G-CAG----GTCC-----C----AGT---ATAT-AC---CATGT----TG-AT-AG-C----C-TCAAA--AGTTG...TCT--TTGG-  6442
```

```
                                                                                                                                    |-  V3 loop
H1B.FR.83.HXB2            TAATTAGA.....................TCTGTCAATTTCACGGACAATGCTAAAACCATAATAGTACAGCTGAACACATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGAAT  7144
Env                       V__I__R__.__.__.__.__.__.__.__S__V__N__F__T__D__N__A__K__T__I__I__V__Q__L__N__T__S__V__E__I__N__C__T__R__P__N__N__N__T__R__K__R__I gp120

H1D.CD.84.84ZR085         ----A---.....................----AA---C----AA-----------T------------T--T--TCA------------------------G---T----A--GGA----C----G-C  6686
H1F1.BE.93.VI850          ----C---.....................---CAA---A--T-AA-T-----A----------------C--T--TGA-------C-G--------------------------------G---G-T--  6466
H1G.SE.93.SE6165          A-G-----.....................----AA--------A------A-C---GT--------------T--T-A-A-----------------C-------------------TG-----G--  6568
H1H.CF.90.056             --------.....................A--AAA---A--T-A------A-C----A--------------T--G---C-A---A-C----CA-----C--G--T-----T-----G----C---T--  6475
H1J.SE.93.SE7887          --------.....................----AA--CA--T-A--------C----A------------A--T--T-A-A-----------GTG---TAC-----T--T-----------G---G-T--  6457
H1K.CM.96.MP535           -------G.....................----AA---A-T--A--T---A-A----A--------------T--TGA-A-----C-----------------G--A--------------------T--  6352
H101_AE.TH.90.CM240       ----C---.....................----AAG--C----AA-------C--------------G--C--T--T-A------------C--------C------TC--------------C---T--  6724
H102_AG.NG.-.IBNG         -G--C---.....................----AA---A----AA-------C-----------------T--GCT-ATC-----A-------------C-----T------------------G-TG-  6668
H103_AB.RU.97.KAL153_2    --------.....................----------------------A-----------------------AGA-C----G----------------------------------------G-T--  6342
H104_cpx.CY.94.CY032      --------.....................---AAA---A----A------A-C----AT--------------TGCA-AGG-----A-------------C-----TGG------------------TG-  6532
H1N.CM.95.YBF30           -T------.................................AT--T-G-CAC-GT-ATC-GT-G--G--ATG---TGAGA-A--GCC---A------------G--AGGA--T------G--GG-CAGG-  6725
H1O.BE.87.ANT70           G---G-TG.....................G-AAAAG--A-TTT---AGG--GA----AT--C-----GACC--A---T-TA-CC--A-C--G-CC---GA------AC-A-TA...GAC-T-C--GAG--  7177
H1O.CM.91.MVP5180         G-----TG.....................GGAAAA---A-T--A--ATCA--A--G-AT--C------ACC--A-----TC--A--A-C--G-CC--C-T----GAAGGA-TTGCAGAGGT-C--GAT--  7179
CPZ.CM.04.EK505           ----A---.....................AA-AATTCAAAAGAC.............--AC-CT-G--G--AT-A--TGA-AG-A-CCC---A-----C--------AGGA--T--G...-C--G-G--CA  6678
CPZ.CM.04.LB7             -G-----G.....................-ACAAGGC-...GAT.............-AGG-TC----------C--T---AG-A--C-C--C--------T--GGTAGGA--T--G---T----G--..  6674
CPZ.CM.04.MB66            --GA---G.....................AA----TCAAATGATATGG-C......--T--------GA-AT-A---GA-A-----CGTC-C-----C--T---A-AGGA--T--C----T----G-T..  6784
CPZ.MT145                 -GGCC---.....................ATA-A-CC--CAGA-..........--TTTAGC----A-------C--AGATC-A---A-G----CA----GG-----AGGA--T--C...-C--G-G-TCA  6662
CPZ.CD.-.ANT              ----G-AT..................GGTCGCAAA---.......--ATC--TGCTTGTA-G-T-T-G-A-AGAATT-GA-AACT--AC-----CA----T------AGGA--T-GG---GT--G-...-A  6594
CPZ.CM.-.CAM3             -TGC----...................................AGG--C-GGT-C-R-T-CC-GA-N--RT-TC-G-ANN-G--CAGC--A----------------AGGA--T-NC...--C-G-G--CA  6572
CPZ.CM.98.CAM5            -TGC----.....................AG--AAG--C-A...AGAC-CAAA-TC-TT------CA-...T-T--AGA-GGGA-TA-C--A-T-----TG-----AGGA--T--C...--C-G-G--CA  6848
CPZ.GAB_X52154            CTG-A---.....................GTG-AG---AAGT-AA-A---A--G-TGT-TGG-----------AGTAGA-G-A--TAGTC-A------CAT-----AGGA------...-C--GGG--GA  7202
CPZ.GAB2                  -GG-GTTC............AGAACTAAGAA-A-G-CAGCTC-A-GTCTCAG-G-C--AG-T---------T-A--A-GGG-AA-TCCC--------C-GT-----AGGA------...-C-GG----GC  6555
CPZ.TZ.-.TAN1             A-GC-TAC........................T-TGT-AAT--CTCAGT-AA--C-C--T--T-----A-AT-T--TGT---AA--A-TT-A-CG---GA----A-AGGA------------GGTCA-G-  6785
CPZ.US.85.CPZUS           CTG-A---........................AGT...AAA...A--TT---AG-C-TT--TC-------AT-CTCAGAGGGA--CA-T--G-C-----TT-----AGGA---------GTGGG--ATG-  7179
GSN.99CM166_AF468659      AT-CC---..............ATAATGATGAATGG-A-G-A-A-TG-AT---T-GTT---GGCT-TGGAGAAG-TTATCA-C-TA-TC---CA----TC-----TGGA--T--A--C-T--G--ATT-  7081
GSN.99CM71_AF468658       AT-CC---..............GTAATGATGAATGGCAAA-A-A-TG-AT---T-GT----GGCT-TGG-GAAG-TTATCAGC-TACTC---CA----TC-----TGGA--T--A--CCT-----ATT-  7072
H2A.ALI_AF082339          AT--CTAT.......................TGGC--GGT-GA--T---AGA-CT-T---C-GCT--A-CAAAC-TTATAA-C-TACT--GC------AG--G--AGGA--T--G---GTTGT-CC---  7631
H2A.BEN_M30502            AT--CTAT.......................TGGC--GG--GA--T---AGG-CT-T---T-GCT--A-CAA-T-TTATAA-C-CAC---GCG------AG-----AGGA--T--G---GTTTT-CC---  7650
H2A.ST_M31113             AT--CTAT.......................TGGC--GGT-G---T---AGA-CT-T---T-GCT--A-CAA-TTTTATAA-C-CACCG-AC------AG--G--AGGA--T--G---GTTGT-CC---  7077
H2B.D205_X61240           AT--ATAT.......................TGGC--GAA-AA------AGG-CC-T-----GCT--A-TACAT--TATAA-T-GTC---AC-C----AG--G--AGGA-----G--GGTTGT-CC---  7617
H2B.EHO_U27200            AT--CTAT.......................TGGC--GGT-AA------AGG-CT-T-----GCT--A-TTCAT-TTATAA-C-GAC---GC-C----A---G--AGGA-----A-TGGTTGTGCC---  7609
H2G.ABT96_AF208027        AT--CTAT.......................TGGC--GGG-GA------AGG-CT-TA----GTT--A-TAA-T-TTATAA-T--AC---A-RA----GG-----AGGG--T--A---GT-TT-CC---  6981
H2U.FR.96.12034           AT--ATAC.......................TGGC--GGT-GA------AGA-C--TA--T-GCT--A-TAA-T-TTATAA-T--AC---GT-C----G---G--AGGG--T--G---GTCTT-CC-G-  7110

MAC.239_M33262            AT---TAC.......................TGGC--GGT-G---T---AGG-CT-TA--T-GTT--A-TAA-T-TTATAA-C--AC---G--A----G------AGGA--T--G---GTTTT-CC-G-  7568
Env                       Y__I__Y__.__.__.__.__.__.__.__.__W__H__G__R__D__N__R__T__I__I__S__L__N__K__Y__Y__N__L__T__M__K__C__R__R__P__G__N__K__T__V__L__P__V gp120

SMM.SL.92.SL92B           AT--ATAT.......................TGGC--GG-GGAAGT---AGA-C--TA--T-GCT--A-TAA-T-TTATAA-T-GAC---A--G--C-G------AGGA--T--G---GTCCT-CC-G-  7004
SMM.US.x.PGM53            AC---TAT.......................TGGC--AGT-GAAGT---AGA-CC-TA--T-GCT--A-TAA-T-TTATAA-C--AC---A-GA----G--G--AGGA--T--G---GTCTT-CC-G-  7502
MNE.US.-.MNE027           AT---TAC.......................TGGC--AG--AA--T---AGG-CT-TA--T-GTT-GA-TAA-T-T-ATAA-C--AC---G--A----G------AGGA--T--G---GTTTT-CC-G-  7039
RCM.NG.x.NG411            GG--CTAT.......................CAA-GAAAACAAAGT---AGG-C-GTT---GGGT--A-TAGAC-TTTTAA-T--AC-G-A-C-----GG-----A-GT----GA---GTG---G----  6902
RCM.GAB1_AF382829         GG--ATAT.......................CAG-GAAGGCAAAGT---AGG-C-GTT---GGTC-CA-TAGTTTTTATAA-T-GTC-G-A-CA--C-GG-----TTCA--T-GA---GTG---G-G--  6858
DRL.FAO_AY159321          --GAA-C-...............AGGGGAAAACCAG-C--TCT-A-TC--AAAT-TGTGT-T-GG--TA-TAA---ATGGAACC-CA------TC----G----AAAGGA--T-GG-GC-T--T-TCT-C  6913
MND-1.F17                 -G--ACCCATAGATGATAAATATAGGGGCC-A-AGG-A--TCATC-A-GGAAGTTTGT-TAT-AG--G-CAGGA--ATATGGCT--A-G--AG-A---CAC---AAAGGA--T-GGT--GT-GTG--T-C  6676
MND-2.5440_AY159322       -TCAG-C-AACCCA......CGAAAAGGAAAA-ATG-G--TCAT---C--AAAT-TGTATAT-G---GG-TAAA--ATGGGGAT--C-GG-A-GG----G---GAAGGGA--T-GAT-T-TC-TTTCC-C  6879
MND-2.14CG_AF328295       -----TTG...........GTGAACCCTCAAAAGG-G--TCAT---C--AAAT-TGTTTAT-GG--GA-CAAA--ATGGGGGT-GA-GG-GGTA---GAT--GAAGGGA----GGT-T-T-GTTTC-G-  7169
MND-2.MND2_AF367411       --CAG-C-AATACA......AAGAAAGGCAAA-AGG-G--TCAT---C--AAGT-TGTGTAT-G----G-TGA---GTACAACT--C--G-AGTA----G----AAAGGA----GAT-TGTG-TCTC--C  7252
DEB.CM.CM40               -TCA--T-......ATAAGAAAAGAGGTAAAA-GTG-GG-GCA-A-TGGAAGC-TT--T----G---G-CCGCT--ATACAACT-GAC-T-A-CA---GT---G--AGGA--T--G--CTACCG-GCC--  6902
DEB.CM.CM5                --CA--TC......ATAAGGAAG.........-G-...C-GGAAA-T--GA-C-TT--TG-C-G-A-C-CA-CA--GTATAA-C--TCTT-GG-----GTT--G--AGGG-----G--GTAT-G-GCTG-  6841
MON.CM.L1                 C-G-A-TG....................ATGAAT--GCAG-AAA-TG-GT---T-GTAG-T-G-T-GGCTAA-C--CT-CA---CA-C----C---C-TT--G--AGGA-----A--T-TT-GG-ATT-  7089
MON.NG1                   C-G---TG....................ATGAATGGGAAA-GAA-TG-AT-C-T-GTA----GGC-TGM-AA--C-CTC--A--GA-C--A-C---C-TT-----AGGA--T--G--CTT--G--ATT-  5705
TAL.CM.8023               ----ACAT.......................CAGGCACC--AAA-TG-AT-AGCC-TA--CC-CT--GG-AA-G-ACACAACA-CACC--AC-G--C-TT-----AGGA--T-GA----TC---GATC-  6476
TAL.CM.266                -C--ACAT.......................CAGGC-CCA-A-A-TG-AAG-GCT-TA--CC-CC--GGAAA-G-ACACAACC-CAC---C-CA--C-T------AGGA--T-GG-----T---GGACT-  6932
MUS.CM.1085               C-G-A-TG....................ATGAATGGGGAT-AAA-TG-AT-A-T-GGA----A-T-TGG-GAA--TTTTAGA---A-TC---C------T------AGG-------G-----T--GG-ACC-  7045
SAB.SN.-.SAB1C            A---CT-G.................CAGAAAAATGGCAATT-CA-TG--T--GTT-TT--T-GGC-CA-TAGAT-TTTTAA-T--ACT--C-GA----G------AGGA--T--G---GT-TTGCCTG-  7433
GRV.ET.x.GRI_677          AG--AT-G.................CAGAAA-GAGGAAATGATA-TG--A-AGTT-TA----AGT-GA-TAA-TTTTACAACT-GAC-G-G-GA--CCG------TGGT--T--A---GTGTTGCC-G-  7133
TAN.UG.x.TAN1             A---AT-G.................CAAAAGCATGGAG-G-GCA-T---T--GTG-TA----A-C--A-TAAAC-TTATAAGT-GA-G--AGTG----G------AGGA--T--G---GTCTT-CC-G-  7222
VER.DE.-.AGM3             AG--AT-G.................CAGAAACAT-GAG-A-GCA-TG-CT-AGTGTTAG-GT--T-TA-TAAAC-TTACAA-C--AC-G---C---C-A------AGGA-----A---GTCTT-CC-G-  6760
VER.KE.-.9063             A---AT-G.................CAGAAGCAT-GAG-G-ATA-TG--T--GT--TAG-TC-TT-TA-CAA-T-TTACAA-T-GAC-G-C-C---C-A----G--AGGA--T--G---GTCTT-CC-G-  7262
VER.KE.-.AGM155           AG--CT-G.................CAAAAACATGGAG-G-GCA-TG-CT-AGTGTTA--CT-GC-CA-TAA-C-TTATAACC-GAC-G---CA--C-A----G--AGGG--T--G---GTCTTGCC-G-  7242
VER.KE.-.TYO1             AG--AT-G.................CAGAAACAT-GGG-A-ATA-T--CA-AGT-TTG--CT-GT-CA-CAA-C--TATAA-C--TC-G-C-CC----GG-----AGGA-----G---GTCCT-CCGG-  6734
SUN.GA.98.L14             ----ACC-ATCAAT.........AACAAGGTA-GAC-AGGGG-T--GGGA-AAT-TGTATGG-A----GCAGCT--GTGGGGGT---T-----C-G--C-T---GAAAGGG--T-GGT--CA-GT---T-C  7401
LST.CD.88.447             -----CC-CTAACTCCA...AAAAAGATG...-GA---C--CTA--AGCAAAGTTTGTATAT-A----GCAGGA--ATGGGGAC-CATT----GA----TT--GAAAGGA--T-GGT--GA-GT-TCT-C  6273
LST.CD.88.485             -----CC-CTAACT......CCAAAAAAGATG-GA---C--CT----GCGAAGTTTGTATAT-A---TCAGGA--ATGGGGAC-CATT--C-GA----TT--GAAAGGA----GGT--GA-GTCTCT-C  6270
LST.CD.88.524             -----CC-CTAACT......CCAAAAAAGATG-GAG--C-T-AT-GAGC-AAATTTGTTTAC-AG---GCAGGA--ATGGGGAC--ATC----GA-----T----AAAGGA--T-GGT--GA-GTCTCC-C  6276
LST.KE.-.lho7             -T---CC-CTAACT......CCAAATAAGATG-AAG--C-A-AT-GTGCAAAGTTTGTGTAT-A----GCAGGA--ATGGGG-T--ATT----GA--C-T----AAAGGA--T-GAT--GA-GT-TCC-C  7342
SYK.KE.-.SYK173           -T-GATTTATCAAA...........CAAAAG...G--AAA-AT--GTCA...GT--TT--TT----T-CAGAAGCACTTAGAT-GC-G----TA----GA------AGGA--TG-GT-T-TC-----AT--  7129
SYK.KE.KE51_AY523867      GGT---TT....................AGACAA--GGAT-AAA-TG-AT-AGTG-TT---C-----G-CAGAA--ATT-GGAT--C--C-GGTG---GA------AGGA--TG-GT-T-T-----AC--  6825
COL.CM.-.CGU1             CG-AAGAT..............AGAGAGGGA---TGGAAG-AT---TCA-GG-C--TTGAGTATTATTG-T-TCCA-A-GA-A---CTT-AGGG----T----AGAGGA--T-G-T-TCAC-G--ATC-  6557
```

```
                                             loop tip                                                     V3 loop end -|
H1B.FR.83.HXB2              CCGTATCCAGAGA...GGACCAGGGAGAGCATTTGTTACAATA...........GGAAAAATA...GGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAG  7256
Env                         __R__I__Q__R__.__G__P__G__R__A__F__V__T__I__.__.__.__.__G__K__I__.__G__N__M__R__Q__A__H__C__N__I__S__R__A__K__W__N__N__T__L__K__Q_ gp120

H1D.CD.84.84ZR085           A-C---A.........----A----CA---GC-CTA-----C-............A-GT-T-C-ACAA-G-TA--AG--------T-------T------G---T---------------C---CG----  6795
H1F1.BE.93.VI850            A-A-T-A.........--------ACA-A----CTA-G---C-............---GC----ATA---G-C--A---A-G-------------------G--A--C--------------C-GG-GT-T  6575
H1G.SE.93.SE6165            AA-A--G...G--ATT--------ACA-A-G--CTA-G---C-............--TGCC---ATA---G-C--A------------------TG---C--A-AG---------AG-GG-C---C-GA-T  6683
H1H.CF.90.056               A-A-T-A.........--G-----AC-------CTA-G---C-............--TG-C--CATA---G----A------------------T---------A--G-C--------G------C-C---  6584
H1J.SE.93.SE7887            A-AC--G.........--------ACA--TGC-CTACG---C-............---G-----ATA--------A---G--A-----------------GA-AG-G-T----G----------CGTAGA  6566
H1K.CM.96.MP535             A-A---G.........--------A-A------CTA-----C-............--TG-T---ATA---G----A-----G--------C----------G---A---------C-TG-------GCAGA  6461
H101_AE.TH.90.CM240         AAC---A.........--------AC---T---CTA--G--C-............---G-T---ATA--------A---A-----T-----G-G----A-G--A--------------AGT----------  6833
H102_AG.NG.-.IBNG           A-A---A.........---------CA------CTA-G---C-............--TG-C---ATA--GG----A------------------TG-C----A-A--G----------A------C-T---  6777
H103_AB.RU.97.KAL153_2      T-A---A.........-------------------TA-G---C-...........---G-C---ACA---G----A-------------------------T-A-----------------A---------  6451
H104_cpx.CY.94.CY032        A-A---A.........--G-----ACT-A---GGTA-G---C-............--TG-----ATA---G----A----------------------G--AATG-T------G----C------GT-  6641
H1N.CM.95.YBF30             G-AG--A.........-----T-CT-TGA-----TA--AC--.............-A-------GTA---G-C--T---------T-C-----TG-CTC--A--A-CT----G-ACCA-TG-GG--TAGA  6834
H1O.BE.87.ANT70             GA-A--A.........--T---AT-GCCTGG-ACAGC-TGGG-......ATAGGG----C-GC-GGAAAC-GCTCA--GGC---TT----C--GTA--A-GCCA-TG-T---GGA--A-TA--------A  7292
H1O.CM.91.MVP5180           ATA--CA.........--T---AT----TGGCGCAG--TG-C-CTTAAAAGAAGTAAC--T-C-TCACC--GATCA--GGT---TT------CATA--A--AGA-TGT----G-A--TG-CC--C----A  7300
CPZ.CM.04.EK505             AGTACAAATA......-----T--T-TGA----CTA--AC--.............-A---T---ATA---G-C-CT--G------T-----G-GG-A-AC--GA--TGGGAAC-A-TATGGAAT-C-ACA  6790
CPZ.CM.04.LB7               .ATCCC--TA......--------ACAGCTG--CTA-GG--C-............-A--CTG--GTA---G-C-CA---G-G--T------G-A----ACCAGA--C-T---T----A-TAC----C---  6785
CPZ.CM.04.MB66              .ATACCTATA......------A-TCAGAT----TA-GG---.............-A--CTG--ATA---G-C-CA--G--G--CTT---CC-GT-G-AC-A-A--GT-----C---T--A--T---A-A  6895
CPZ.MT145                   AATACAAATA......-------CA-TGA-T---TA--AC--.............-A---TG--GTA--GG-C-CT---A-----T-----G-G--A-A-G-GA--C-G---GCA--AG-CC-C--TG-A  6774
CPZ.CD.-.ANT                T-TACAAATA......--------A-TGA-T--CTA--ACG--............-A--T-GC-ACA---G-C-CT--GA----GTTC----CAG-C-A--AGA-GCT----G-AC-AG-ACGT--CA-A  6706
CPZ.CM.-.CAM3               AATACAGATA......--------C-TGA-C---TA--AC--.............-A---T--TGTA---G---CC----G----T-----C-G--C-A---GA--GTC---G-CG-A-GAC----TG-A  6684
CPZ.CM.98.CAM5              A-TACAGATA......---------GT-T-C--CTA--AT--.............-A---C---GTA--GG--CCC------------------TG-C--C-AG--CC-G----TG--A-AG-----CG--  6960
CPZ.GAB_X52154              GGTGCAGATA......--G-----A-TGA-C---TA--AT--.............-A---TG--GTA---G---CC---TCT--CT-C-----G--A-A-G-GA-C-C-------CGA--GG-GG--G-A  7314
CPZ.GAB2                    AATA-A--TT.....A-C-------C-A----CT-C-AC-CG.............-A-GCCC--ATA-----CCCT--GA----GTCC--CC--C---A-G-TA--CT-----------T-C----CAGA  6667
CPZ.TZ.-.TAN1               A-AG--A.........--T-----T-TGA-C---TA--AT--.............-A---TG--GTA--GG-C-CC--GA----TT-----TCAG-C-A-GC-A---C----T-C-GG-AC---G-TTG-  6894
CPZ.US.85.CPZUS             A-AAC-A.........--------A-TGA-T---TA--AC--.............CC---G---GTA---G--G-A---G------------------CTCA-A-CTG-C----G-G--ACAGAG----T-C  7288
GSN.99CM166_AF468659        G-AA--A..........---G-G--A-TGA-----TA-T-CCA-...........-TG-T-G--GGA-------CT--GA----CT----CCGAC--GA-CCT-AT-GG-------GGG-AA---G-G-A  7190
GSN.99CM71_AF468658         A-AG--A..........---G----A-TGA---CTACT-CCA-............ATT-T-G--GGG---G---CT--GA-G--CT----C--AC-C-ACCATA----G---G---TAG-AA---GG--A  7181
H2A.ALI_AF082339            AACAC-TAT-TC-...--GTT-ATATTT...CACTCC..................CAGCC---CAATAA--GACCC---------TGG--CTGGT-C...-A--GCG------GG--AG-CA-GC-GG--  7734
H2A.BEN_M30502              AACAC-TAT-TC-...---TT--T-TTT...CACTC-..................CAGCC---CAACAC--GGCCT--G--G---TGG--CCGGT--...G---GC-G-----GGG-AG-CA-GC-GG--  7753
H2A.ST_M31113               AACAC--AT-TC-...--GTT--T-TTT...CACTCC..................CAGCC---CAATA---GACCC--G------TGG--CTGGT-C...-A--GCG-G-----GG-AG-CA-G--GG--  7180
H2B.D205_X61240             AA-A-C-GT-TC-...----T-CTTTTC...CA-TCA..................CAGCCT--CAATAAG-GACCC--------TTGG--CTGGT--...-AG-G---C----CAG-AG-CA----GG--  7720
H2B.EHO_U27200              AA-A-C-GT-TC-...--TATTCTCTTC...CA-TCA..................CAGCCT--CAATAA--GACCT-A------TTGG--CTGGT--...-A--G---C----CAG-AG-CA--C-GG--  7712
H2G.ABT96_AF208027          TAC---TAT-TC-...--G-T--TCTTC...CACTC-..................CA-CC---CAAT-A--GACCC-----G--TTGG---TGGT--...G---GC--------GG-AG-CA-GC-GG--  7084
H2U.FR.96.12034             AACC--TAT-TC-...----T--T-TTC...CA-TC-..................CA-GCT--CAATAAC---CCA-A------TTGG---TGGT--...G-T-G---T-----AGGGG-AA-CC-GG--  7213

MAC.239_M33262              -ACC--TAT-TCT...---TTG-TTTTC...CACTCA..................CA-CC---CAAT-AT-GGCCA-AG--G---TGG---TGGT--...G---G---------GG-TG-AA-----G--  7671
Env                         __T__I__M__S__.__G__L__V__F__.__H__S__.__.__.__.__.__.__Q__P__I__N__D__R__P__K__Q__A__W__C__W__F__.__G__G__K__W__K__D__A__I__K__E  gp120

SMM.SL.92.SL92B             -ACC--TAT-TC-...--TTTG-TCTTT...CA-TCG..................CA-CC----AAT-AG-GGCCA---------TGG--CTGGT--...G---G--------GAG-AG--A-GC-GG--  7107
SMM.US.x.PGM53              -ACC--TAT-TC-...--GTTGATCTTC...CA-TCG..................CA-CCC---AAT-AG-GACCA-A------CTGG--CTGGT--...G---G--GT----GAG-GG-CA-CC-GG--  7605
MNE.US.-.MNE027             -ACC--TAT-TCT...---TTG-TTTTC...CACTCA..................CA-CC---CAAT-A--GGCCA-A---G---TGG----GGT--...GA--G---T-----GG-GG-AA-----G--  7142
RCM.NG.x.NG411              -TCAT-AGCA-C-...---GT-...TTC...A-CTC-..................CT--GGG-GCAAAA--GGCCA-A-GG---TTGG----GAT--...GAT-G-G-T-----AGGTG-G-GG---G--  7002
RCM.GAB1_AF382829           ATCGC-AGCA-C-...---GTCTTT-TC.....TCA...................CT--G-G--GAGAAG-GACCA-A-GG---TTGG----GAT--...GA--GG--T----CGG-TG-A-GG---G-A  6958
DRL.FAO_AY159321            A-CCTCAGCCTC-...--GTTGTT-TTTTATCA--GG..................TT-G--CC-GGAAA----T-A-A-AG--G-ATG-----AT--...GA--G-C-----GGA--GG-----GG-AGT  7019
MND-1.F17                   T-CATCAGCT-C-...---TT-TTATTTTATCA--GG..................TT-G--CCTGGAAAG---T-A-AGA---GCATG--C-C-T-C...-A--G-CGT---GGGTTAG-AC-TTGGAGT  6782
MND-2.5440_AY159322         A-CA-GTTCT-C-...---TTGCT-TTCTATCA--GG..................TTGG-GCC-GGAAAG---C-T---A-G-GCA-G---C-GT-A...GA--GG------GGAC-GG-----C-TAGT  6985
MND-2.14CG_AF328295         --CA-GTGCATC-...---TT-ATTTTCTATCA--GG..................TT-G--CC-GGCA-G---C---A-A---G-ATG--C-GGT-A...-T--G-C-----GGA-GAG----G--CGCA  7275
MND-2.MND2_AF367411         A-C--GTGCA-CG...---TTGATCTTTTATAG--GG..................TT-G--CC-GGGAA----C-T-AGA---GTATG---C-GT-G...-AG-G-C-----GGA-TAG-CA-GC-TG-T  7358
DEB.CM.CM40                 A-AC--GGCT-C-...--T-TGTCCTTT................TATACTACCTTT-T-CA-AGGCTG-GG--A-AGAG-----------GAC-A...-AC-G--GC---GCA---G-AAC---GG-A  7008
DEB.CM.CM5                  T-A---GGCC-C-...--G-T-TCCTTT................TATACAACCTTC-T-CC-AGGCTG-GG--A-AGAG-----------GG--G...----GT--T----CAGGAG-AG-T--GG--  6947
MON.CM.L1                   A-AG--AGGAGC-...--CATGACATTCTAT...TCG.............CAACTT-T-G-GGGA--T----CTC-TA----CT-C--C--AG---A--AGA-CC-----G-G-CTG-CC--C-GGCA  7198
MON.NG1                     A-AG--TGGTGC-...---ATGACATTCTAT...TCG.............CAAATC-T-G-GGGA-----C-CY---A-G---T-C---C-GG-A-A--AGA--C-G---G-CGCTG-C---C-TGCA  5814
TAL.CM.8023                 T-AG--AGCAGC-...--T-T-AT-TTT...CACAG-.............CAGATC-T-GC-GGTAA-G--T-A-AGAG---CT-------G--A...GA--G---T----CG-GGG-GA-GG--G-T  6582
TAL.CM.266                  A-AG--AGCAGC-...----T-AT-TTT...CACAGC.............CAGAT--T-GC-GGGAAG--CC-C-ACAGG--CT-------GC-A...GAG-GC--T----C---AG-CA-GG--TC-  7038
MUS.CM.1085                 T-AA--AGGAGC-...--CATGACTTTCTAT...TCG.............CAGCTT-T-G--GAT--CG---CAC-CAG---CT----C--A--A-A--AGA--C-G---G--CCAG-A--G-G---A  7154
SAB.SN.-.SAB1C              AACC---AT-GCT...--CTTG-TCTTC...CA-AGC..................CAG---TATAACACC-GGT---A------TTGG--CTGGT--...G---GT--T----GAGGAG-C-GG--GG-A  7536
GRV.ET.x.GRI_677            AACA---AT-GC-...--GTT--TATTT...CACTC-..................CAG---TATAATACC-GGT-A-A------GTGG--CC--T-C...CA--G-G-T-----AGGGG-A-GG---G-A  7236
TAN.UG.x.TAN1               AACC---AT-GC-...--T-TC-T-TTC...CACTCA..................CA-C--TATAATAC---AT-A---------TGG--CC-TT-C...CAG-GTG-T-----AGGAG-C-GG-GGG-A  7325
VER.DE.-.AGM3               AACA---AT-GC-...--G-T--T-TTT...CA-TC-..................CAG-GGTACAATAC--GGC----------TTGG---C--T-C...CAG-GC--C----GAGGAG-C-GG---G-A  6863
VER.KE.-.9063               AACC--AAT-GC-...--TTT--TTTTT...CA-TCG..................CAG---TACAATACG--GC-T-----G--CTGG---C-TT-C...CA--GC-GG----GAG-AG-C-GG--GG--  7365
VER.KE.-.AGM155             AACG--AAT-GC-...---TT--TCTTC...CACTCA..................CAG--GTATAATAC--GAC-A--G--G--CTGG--CC--T-C...CAG-GC--T-----AGGAG---GG--GG-A  7345
VER.KE.-.TYO1               AACG--AAT-GCG...----TG-TTTTC...CACTC-..................CA----TACAACATG--GC-T-----G--TTGG---C--T-C...GA--GC--T----GAGGTG-C-GGCGGG-A  6837
SUN.GA.98.L14               TATA-GTAGT-C-...----TGCT-TTTTAT-A--GG..................CT-G-GCC-GGTTCT--GT-A----TG-----G-----AT--...GT--G-C-----GGA-GAG-C--TGCCACC  7507
LST.CD.88.447               -AT--GTTCA-C-...----TTCT-TTTTAC-A--GC..................TTGG--CATGGTA-C-GGC-----TTG--T--G-----AT--...GA--G-C-G---GGA-GA-TG--TC-TA-T  6379
LST.CD.88.485               -ATC-GTTCA-C-...----TTCT-TTTTAC-A--GC..................TTGG--CATGGTA-C-GGC-----TTG--T--G-----AT--...GA--G-C-----GGA-GG-TG--T--TA-T  6376
LST.CD.88.524               AATC-GTTCA-C-...----TTCT-TTTTAC-A--GG..................TT-G--CATGGGTCT-GAC-A---TTG-----A--C--AT--...GAG-GTC-----GGA-GA-TG--T--CA-T  6382
LST.KE.-.lho7               AAT--G-TCT-C-...---TATTTATTTTAT-A--GC..................CT-G--CATGGCA-C-GAT-A---TT------G-----AT--...GA--GCC-----GGA-GA-TG--T--TA-T  7448
SYK.KE.-.SYK173             T-AAT-GGCTGC-...--CTATTTCCTCC--G-AA-ACA-GGG........AAATTG----CCGGGA-GG-CGCA-A--GT--CTTC---CGAG-A...-C--G---T----CAG-ATT---T--G---  7244
SYK.KE.KE51_AY523867        --AAT-GGCAGC-...---TTTTTCTTGC--G-GA--CA-GGG........AGATTG----C-GGGAA-GC-GC--A-AG----TT---C--AG-C...-A--G-G-----GGA--GTT---TG-G---  6940
COL.CM.-.CGU1               AAAC-CTGCC-AT...---G--AAATTCTAC-A--AG..................CTT-TTCCTTATTC---AGGC-TTT-T-G-AGG---C-GT--......-T-CC-AT--C-GGGCAGAAT--GA-A  6660
```

```
H1B.FR.83.HXB2              ATAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATC..................................................................................  7304
Env                         _I__A__S__K__L__R__E__Q__F__G__N__N__K__T__I__I__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._  gp120

H1D.CD.84.84ZR085           G-------A------G--A-C-TT--GAACC-A-CA---...........................................................................................  6834
H1F1.BE.93.VI850            G--AAGGCAG------AGTCG--T--CCCT------..---GC-......................................................................................  6617
H1G.SE.93.SE6165            G-C---GCAG--C--G-GA--ATC---AAT--G-GC-GCGAG-AC.....................................................................................  6728
H1H.CF.90.056               G---T--CAC-----G--AT---C--GAAC----GA-C-...........................................................................................  6623
H1J.SE.93.SE7887            G------CA---C----------C---AAT--A-CA..............................................................................................  6602
H1K.CM.96.MP535             G--AAGGAA--GC---A------T---AAG---GGA-C--T-........................................................................................  6503
H101_AE.TH.90.CM240         G--A--GAA-------A---G--C---AAT--G-CA..............................................................................................  6869
H102_AG.NG.-.IBNG           G---T--CAC------AGACGT-C---AAG--C-CC-C-...........................................................................................  6816
H103_AB.RU.97.KAL153_2      ----T--T----------A----------G-----A-C-...........................................................................................  6490
H104_cpx.CY.94.CY032        ---AG-GAAG----G-A-AG--TC--CCCT-----A-C--T--A-.....................................................................................  6686
H1N.CM.95.YBF30             -C-AGAGAGG--A---AGA--ATCC-G--G--A--C--C..........................................................................................A  6874
H1O.BE.87.ANT70             -C----GAA-GG-ATTT-----T-G-AAAC----CAGGT.......................................................................................AGTA  7335
H1O.CM.91.MVP5180           -C-----TA-GG-ATTT-A-T-TTG-AAACC-A-CAG-G-AT.......................................................................................G  7343
CPZ.CM.04.EK505             -A-CAA-TA-T-A---ACA-C........................-GAAAG...............................................................................  6817
CPZ.CM.04.LB7               G-GAAA--AG-GC---CCACGGTC--CAATG-ATC---C-A-........................................................................................  6827
CPZ.CM.04.MB66              G--AGGCA-GCC----AT---AC--A-AAGGGC...T-TCTCGG--ATGAA...............................................................................  6943
CPZ.MT145                   -C-AAAGAAGTTC----GA-CAT-C-GA-G--G.................................................................................................  6807
CPZ.CD.-.ANT                -C--AGCA-GTTC-TGCG--G--T-GGAA---AGTAG-C-AC-A--C-AAT...............................................................................  6757
CPZ.CM.-.CAM3               -CT-G-CAAGCC--GC----G-TT---AC---CCT--C-CA-G-C-A-TTTACAGTAAGT......................................................................  6744
CPZ.CM.98.CAM5              -CC---TTAGCCA--GC-A-GT-TGA-AA--GC-TAGC--A-G-C-C-TTTACACCAAAC......................................................................  7020
CPZ.GAB_X52154              G--AAA-AGGCT---GC-ACCTC--CAAATCGA-CAGC-G-C-AC--AACACTCAACAGAGCC...................................................................  7377
CPZ.GAB2                    ---AAACAA--GA-T-A-A-CAGCACAACGTGGC-C-GGGGGGAC---ACA...............................................................................  6718
CPZ.TZ.-.TAN1               GCTATGGCTGCCA---ACAC-ACCA-GA-GGCC-GA--TGA--CGG-A..................................................................................  6942
CPZ.US.85.CPZUS             -CTTTAGAG-T-A---A-A-GG--GCAAACCTG-CA--GGT-GAGT-AATTCCAAAT.........................................................................  7345
GSN.99CM166_AF468659        GCTATG-AAGC-A-G-AT-----T-GG-A------A-CTGGG..............................................................................AGAAATGATA  7242
GSN.99CM71_AF468658         GCCATG-AGGCCA-G-AGA-C--T-GG-A---G-TC--C-ATGAT-C-ACC.....................................................................CCTAATAAGA  7242
H2A.ALI_AF082339            G-GAAGGAA-CCC-TGT-A----TCCCA-GT----AGG---C-ATGA-ACA.............................................................................AACC  7789
H2A.BEN_M30502              G-GAAGCAA-CCC-TGT-C----TCCCA--T-C--AGG--TC-ATGATACA.............................................................................GGGA  7808
H2A.ST_M31113               G-GAAGCTA-CCC-TGC-A----TCCCA-GT----AGG---C-ACGA-ACA.............................................................................GAAA  7235
H2B.D205_X61240             G-GAAA--G-CCA-C-T-A----TCCCA-GT----AGG-GGTGC--AAAAT...ATCACA..................................................................AGCG  7781
H2B.EHO_U27200              G-GAAAGAG-CCA-T-A-A-T--TCCCA--T--TCAGG-----C--ATATC.............................................................................TCAC  7767
H2G.ABT96_AF208027          G--AAGGAA-CCG-GGTCA-G--CCCCA-CT----AGG-----ATGA-ACA.............................................................................AAGA  7139
H2U.FR.96.12034             G-GAAGGAG-CCA-TGCCA-C--TCCCA-GT--C--GGT--C-AG-ATATT.............................................................................AGTC  7268

MAC.239_M33262              G-GAAGCAG-CCA-TGTCA----TCCCA-GT---C-GG---T-AC-ATACT.............................................................................GATA  7726
Env                         _V__K__Q__T__I__V__K__H__P__R__Y__T__G__T__N__N__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__D__  gp120

SMM.SL.92.SL92B             G-TAAG-AA-CCA--GTCA----CCCCA-GT---C-GG---C-ATGATACT.............................................................................AGGA  7162
SMM.US.x.PGM53              G-GAAGGAA-CCA-TGTCA----TCCCA--T---CGGG---T-ATGAGACT.............................................................................AAGA  7660
MNE.US.-.MNE027             G-GAAGCAG-CCA-TGTCA----TCCCA-GT---C-GG---T-AC-ATACT.............................................................................GATA  7197
RCM.NG.x.NG411              G--AAAGAGGC-A-T-ACA-GAC-AAA---T----AGGT---CG-CAGCCT.............................................................................GAGA  7057
RCM.GAB1_AF382829           G--AAAGAG-G-G-G-A-AC-AC-AAA--GT--CGAGGT--T-GT-A-ACA..........................................................................GACAAAA  7016
DRL.FAO_AY159321            C----C-AAG------A-TC-GT-AG----TCC-T-TGG-G--AT-CTACT.........AATAGCACATGCATAAGTCTCAAAGGCAAAAAGATCCAGGACAGAAACGTGACAGGGTGCATGTTAAAGC  7140
MND-1.F17                   C------AAG--C---ATA--TT-AA--ACTCC-TC---GTG-ACCAGACCTGTAAAAATTTTACTAGCACTGGAGAGGAGAACAAACAAAACACGGACAAGCAAAAGGAGTTTGCCAAATGCATAAAGA  6912
MND-2.5440_AY159322         C--T-ATTAG-GC-----A-GAT-AA--AT-GC-T-T-C-A-GAC-A-CAC...AATATGACATGTAAAAGCAGTAACAACAAA......AAGAACACAACAGGGTGTCATCTT............AAAA  7094
MND-2.14CG_AF328295         T--T-AC-TG------AGA--GT-AA--C--G--T-T-T-G-GGTT-TAATGCAAGTGGCCCTTGCAGGGGGATAAACAATAGAGGAGTGGTAAACAAAACAGGGTGTGCCTTG............AAAA  7393
MND-2.MND2_AF367411         C----A-TAG--------A--AT-GACA-CTCC-TATGG-G--ATG-AACA...AAGGGCTATTGCAAGGGGTTAAAAAGAAAG......GAAAATAGAACGGGTTGTGCCCTT............AAAA  7467
DEB.CM.CM40                 --GCGACAA--GA--TTG---AT---------GGCA--T-GG-C--ATAAT...............................................................................  7059
DEB.CM.CM5                  ---AGG----G-----AG--GAT--A-.........--TGTG-C--AA..................................................................................  6986
MON.CM.L1                   G-TCACGAGGC-G---A-ACTG-G-GG-A---G--A--C-ATGG--C-AATGTGACCACAATATCT................................................................  7264
MON.NG1                     GCCCA--TRGCTA-C-RGR-GG---GG-A--GC--A--T-AT-ATG--ACAGTGGCCAACATAACT................................................................  5880
TAL.CM.8023                 G-GAA-GAG--GA-TGC-A--GCC-ACAAGG--C-C-T--AC-GC-GTAGT..........................................................................GGAACAG  6640
TAL.CM.266                  GCCAA-GAAG--A--GC-A--AGG-ACAAGG-GC---TC-AC-AC-CAGAT.............................................................................AAAG  7093
MUS.CM.1085                 GCTATGCTAGCCA-G---A----T-GGCATGG--GATTGCA--A-CAGAACCACAGCTCGATAAACGAGACTAAAATAAAGATAAGA...........................................  7241
SAB.SN.-.SAB1C              G-CAAAGAA-C-A--GTGAG--T-CCACC---A--GT-T-GTGGC-CAAAT.............................................................................GACA  7591
GRV.ET.x.GRI_677            G-CAGAGAAG--G---AGA--GTGAAAAATCT--CAG--GT--GC--AGAA..............................................................................A  7288
TAN.UG.x.TAN1               G-GAGA-AA-C-A--GTG--G-TTCCAAA-G-G--AT-TCG-GGG-CAAAT.............................................................................AACA  7380
VER.DE.-.AGM3               G--AAA-ATG--A--GT-A--TT-CCAAA-G---GAT-CCA-GG--C-AAT.............................................................................GATA  6918
VER.KE.-.9063               G-TAAAGAAG--A---TCA--TT-CCAAA-G---GGT-TCGGGG--C-AAT.............................................................................GATA  7420
VER.KE.-.AGM155             G--CAAGAGG--A--GT-A--TT-CCAAA-G-ACGGT-CCA-GGC-C-AAT.............................................................................GATA  7400
VER.KE.-.TYO1               G--AAACAA---A--GT---GTT-CCAAA-G-C-GGT-T-A-GG--C-AAT.............................................................................AATA  6892
SUN.GA.98.L14               T---GG-AACT---G--GC-GGTGGAACCTTTAGC---C-TG...............AGTAATGGCTGTACATTTGATAATATAACAAAGACATGCAGATTTACCAATGGTACAGATTTTAAAAAGATGA  7622
LST.CD.88.447               T---GA-A--TGC-T-A-AGG-TTAA--C--CAGCG-TG-ATTAT-C-CATGGAAAC......TGTATAGATAACACA.....................AAGCCTTGTGGTAGACAACTAAAG...GGAT  6479
LST.CD.88.485               T---GG-AA-TGC-TCA-A-G-TTAA--C--CAGCA-TG-ATTAT-CAAAT...GGACAA...TGTCCTGAGGGAAAA.....................CAGCCTTGTGGTAGACATCTAAAG...GGAT  6476
LST.CD.88.524               T---GA--G-TG----AGA-G-TTAA--AG-C-GC--TG-ATTAT-AAGAA...GGAAAT...TGTTCCACACCAGGA.....................AAACCATGTGGTAGACAATTAAAG...GGTC  6482
LST.KE.-.lho7               C---GC-AA-TGC---AG--GTTGAA--C-G-AGCC-TG-ATTAT-CAGAAGGGACAGGGACATGTGATTCTAAGAAA.....................ACAACTTGTGGACGGAAATTAAAA...GGCT  7554
SYK.KE.-.SYK173             G--CA-GAAC-GGCC-CGA--ACC...T-G--A---GT---C-AC.....................................................................................  7286
SYK.KE.KE51_AY523867        G-CCA--ATG-GAGT-T-A-GGTG-GGAA----GTG-CCT---C-TCA..................................................................................  6988
COL.CM.-.CGU1               -ACAAA-CAC----CGC-AT-G--A-CAA---G---TT----GCTTGGCTAGAGCGCATTTCACGCAAAAATATA......................................................A  6736
```

```
                                                                                                                                   |-          V4 loop
H1B.FR.83.HXB2           ..........TTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTG  7423
Env                      .__.__.__.__F__K__Q__S__S__G__G__D__P__E__I__V__T__H__S__F__N__C__G__G__E__F__F__Y__C__N__S__T__Q__L__F__N__S__T__W__F__N__S__T__W gp120

H1D.CD.84.84ZR085        .....ATAATT-----A-C---------------T--------AC---A----------------------------------C---A--T--GG------------G-A---AA--TATCAGGACA  6959
H1F1.BE.93.VI850         .....ATAAAA-----C------------------T-------AC--T---T-----------A----A-------------CG--A--T--GG----------GAC--AG-C-CC----A-GGCAC  6742
H1G.SE.93.SE6165         .....ATAACC-----CTC---TG-----------T-------AC---A--T-----C-T---A----A--G-------T-----A--T--GGC---------TCA-GCCT---G-GATCA-A-A-  6853
H1H.CF.90.056            .....ATAAGC-------C-AA------------ATG---G--AG---A--T-----------A----A----------T--C---A--T--GGG-------------G----GAA-TGCA----AA  6748
H1J.SE.93.SE7887         .....ATAAAC----CATC-C-------------TAT---------G--A--T----------------A-------A--T-----A--T--A-------C-------G----GA---G-A--GCAT  6727
H1K.CM.96.MP535          .....ACATTTAAACCA-C-AA-C-------A-----------C---------TG---------C---A-------------C---A-----A-----------GAG--AG--GAG---G-----..  6626
H101_AE.TH.90.CM240      .....ATAATC---C-A-C-C----------A--T-T-------AC--T---TCA---------A-----------------T--C---A-----A------------A------CC-AGGA-A-GAAAC  6994
H102_AG.NG.-.IBNG        .....ATAATC---GCTA-CC---T---------TGT-------AC---A--T----------------A----------T--C---A--T--A-------------C------GA------C-A-A-  6941
H103_AB.RU.97.KAL153_2   .....ATAGTC-----T-----------------------------T------------------------------------A-----A--------------C------AA-GG--C-GAAGA  6615
H104_cpx.CY.94.CY032     ..........---GCT-C-C--GT-----------T-------AC---A--T--C---------AA---A----------T--C---A------C---------------ACACA-GC-A-A-GG-AC  6805
H1N.CM.95.YBF30          TAACCTTCAGGGC-CGAG-GAGGAAT-A---A----T----G-GAC-CACTTA-TG--C------A----A--G--------T-----CA-TT-CA--T-A------GAGGAA-TAC----CGAG--AG-  7004
H1O.BE.87.ANT70          TTAACATGACA--C--T--CAG-AGC--T--A--T-T---GG-AACCCATTTACAC-----C---CAT--A-----C--T--T-----CA--G-TA-GA--------TA---C-TT...TCAT---ACG-  7462
H1O.CM.91.MVP5180        TTACCATAATA--C-GCAG-A-TAGT--T--A--TG-----G-AAGCCATTTACA------C---CAT--A-----C--T--T-----CA--T-TGGGA-------CTA-----TTA-C--CT----AAA  7473
CPZ.CM.04.EK505          ..AATATTACA----TA-CCAATC----C--A----T----G-AAC--ACTTA-TGA-C------------A-----C--T---------A-C-GT---T------CA-A-CAAAAT.............  6931
CPZ.CM.04.LB7            ..ACTGTACAC---GCAA-CAGT-----T--A--T------G-G-C--AC-TTCA----------------A-------T--T-----CA-CT-GGCCT----C---GAC--C-T-......TTACGCAA  6949
CPZ.CM.04.MB66           ..ACTATCACA---GGA-C-AG-A-G-----T---TT---GG-AAC--AC-TTCACC-C-TC-------A--G--C--T--T--C---A-CT-CAT---C--------CCT-CATC.............  7057
CPZ.MT145                ..AACATAAGC----T-GTCC-T--T--G--A--T------G-GACT-AC--TCAC-----C--C--G--A-----C--------------A---GTG-GA-CA-C----TC---AA-.............  6921
CPZ.CD.-.ANT             ..GCGAAAACAA-ATG-AC--T-CA--AT--A--T--T---G-AAA-GT---TT-G--------CCA---A-----C--T--T---G--AT-----CTTG---C---GCC--A-ACACG......GGAAA  6879
CPZ.CM.-.CAM3            ......................C-------A---------G-AAC--ATATG-TG--------------A-----C--T--T--C---A-T--TAC-T---------TA---G---AAA--C-AC-A-AT  6851
CPZ.CM.98.CAM5           ......................C-C-----A--T---------AC--ATATG-TG--------------A-----C--T--T-----A-C--T-C---C--C-----AGTT-AAT-A--C-T--GA-AA  7127
CPZ.GAB_X52154           .......................--T-----A--T------GG-CAC-CAT--T-TG--------------A-------------T-----A-T-GC---A-A---......---GACAAC..........  7467
CPZ.GAB2                 ...........--C-CAA-GCATC-------A---------G-G----ATTT--TG-----C-----------G--C--T--T-----A--T--AG-T-AA--.......--C--TAAC..........  6820
CPZ.TZ.-.TAN1            ..CAACAAACG--CC-ATGGCAGAGG-AT--A-----T---GG-CACT-GCTT-T-G--C------CA---A-----C--T---------CTC---A-TTG-AC-----CCTGGACAGC-...-A--GAAC  7067
CPZ.US.85.CPZUS          ......................G-------A---------G-G----ATATG-TGC--------------A-------T--T------A---TT-CC--A-----C-TG--C-ACAAC..........  7441
GSN.99CM166_AF468659     CTCAAATAAGA-GG-CCTC-GAGC-TAA---A--T-T---GG--CA---T--TT-G--CC-A---CA---A----------T--C----TGT--AT---C---C-ACT--A-AATAAC...-CC-T-AA  7369
GSN.99CM71_AF468658      CTCAAATAAGG-GG-CCTCCGAAC-TAA---A--T-T---GG-CCA---T--TT-G--CC-A---CAG--A----------T-------TGT--GT---C---C-ATTC-ACAACAA--T--CGG-AGT  7372
H2A.ALI_AF082339         AAATTAACTTTACA--A-C-GGAAG---CTCA--TG-----G-G---TATATGT-GAC--C--CA---A------C--C-T--C--CATG--TTGGT-CC-C---T-GGTGGAAAAC--A-CGGG-CA  7919
H2A.BEN_M30502           AAATTAACTTTACG--A-CGGGAG----CTCA-----G---G-G-C-TTTATGT-GAC--C--CA---A-----C---------CATG--TTGGT-CC-C---T-GGTAGAAGAC--G-ACCAAAC  7938
H2A.ST_M31113            AAATTCGTTTTA-AGC--C-GGAGA-C-CTCA--------G-G-C-TACATGT-GAC---C--CA---A-----C-------C---ATG--TTGGT-CC-C---T-GGTAGAAAAC-GA-CG-A-CA  7365
H2B.D205_X61240          TAAAGTTAGTA-CAG-A--TGGAAA---TTCA--T-------CAAC-TATATGT-GACC------A---A-----C--A---------ATG--CTGGT-C--A--CT-GGTAGA-AAC--A-CA-A-AC  7911
H2B.EHO_U27200           AGATAAGGTTAGCAG----CG-GAG-A-CTCA--T------G-AAG-TATATGT-GAC---C--A-G--A--G--C-----------ATG--TTTTT-C--A--CT-GGTAGAAAA--GA-C-GGGCT  7897
H2G.ABT96_AF208027       AMATAACCTTTACA-CA-C-GGAGA---TTCA--T------G-AAAGTTCATGT-GAC------CA---A-----C--G--T-----AATG--TTGGT-TC-A---T-G-TAGA-GAC-GA-A--TGAC  7269
H2U.FR.96.12034          AGATAAATCTGGCAG-A-C-G-AGG-...--T--TT----GG-AAA-TTTATGT-GAC---C--CA---A-----CC---------GATG---TGGT-CC-C---T-GGTAGAAAAC-GG--C-TGGA  7395

MAC.239_M33262           AAATCAATTTGACGGCT-CTGGAGG-...--A--T--G---G--ACCTTCATGT-GACA-----CA---A--G--CC---------AATG-ATTGGT-TC-A---T-GGTAGAAGA--GG-A---AGC  7853
Env                      K__I__N__L__T__A__P__G__G__.__G__D__P__E__V__T__F__M__W__T__N__C__R__G__E__F__L__Y__C__K__M__N__W__F__L__N__W__V__E__D__R__N__T__A gp120

SMM.SL.92.SL92B          AAATTAATCTAACGGCC-CGGGAGG-...-----T--G---G-CAC-TTCATGT-GACA-----CA---A-----CC-T--------AATG-ATTGGT-CC-C---T-GGTAGA-GA--GG-A---GA-  7289
SMM.US.x.PGM53           AAATTAATCTGACAGCT-C-GGAGG-...-A--T------G-CACTTTTATGT-GACA------A---A--G---C----T--C--AATG-ATTGGT-CC-----T-GGTAGAAGA--GA-AC--A-C  7787
MNE.US.-.MNE027          AAATCAATTTGACGGCT-CTCGAGG-...--A--T--G---G--ACCTTCATGT-GACA-----CA---A--G---C---------AATG-ATTGGT-TC-A---T-GGTAGAAGA---G-A-CTGAC  7324
RCM.NG.x.NG411           ATATAACTATAAGATCAGTC-ATGG-..--A--TGAT--GGCAAA-TGTTTTT-G--A-------AT--A------C-T--T--C--AATG-ATTGGT-TC-A--CTA-CTAAATAA---A-CAGAAG-  7184
RCM.GAB1_AF382829        TAAAGATAAGAACAGTAT-TGGT......--A--TGAT--GGCAAG-TATTT-T-GC-A------AAT--A--------A--T--C--G-T--ATTGGT-T--A---TTGTTAAATAA-G-G-CAGTAG-  7140
DRL.FAO_AY159321         AGCTTAAGGTGGA---TT-CA--A-CAAG-----T-TG-C-TCA-AG-AC-T--TGA-GTTG--------A---AC----T-------GTTT-TA--A-A-GG-GA-CATGGAACAA--GA-CA-G-AA  7270
MND-1.F17                CTCTTAAGATAGA---TT-TA-TA--TC---A--TAG--C-GCA-A--T-ATG-TGA-G-CA---CA---T---A-G-----T-------GT----AG-A-CA-G-GGGCATGGAATGA-CC--A-GAGAA  7042
MND-2.5440_AY159322      CTATAAGTATAAG-G--TCCA--GTTAA---C--A----G-GC--A---TATT-TGC-CCTC--------A----A-----TT-------GG--TA--A-A-GG--GGCATGGAATAGC--ACAA-G--C  7224
MND-2.14CG_AF328295      CAATAGAAGTAAG---CT-CA--A-CATT--C--A----G-GCA-A---AATT-TGA-CCTC--------A----A-----TT-------GG---AG-A-T-GG--A-CCTGGAATAACC-A-ACT-AAA  7523
MND-2.MND2_AF367411      CAATAAAGGTC-C-G-CT-CA--A-GAA---C--A----G-GCA-A---TATT-TGC-CCTC--------A----AC----TT-------GG---A--A-T-GG-GA-CATGGAATGA-C-A-A-T-A-C  7597
DEB.CM.CM40              ........CTGACA-TA--C-A-C--AA---A--TAGG---G-ACA--GTGT-T-G---C-G---CAT--A-----C-----T-----CATTT-CA--GCCC-AG--CTCCTA................  7164
DEB.CM.CM5               ........ATAACAGTA--C-ATC-------T--TAGG--GG-ACA--ACACGT-G---C-A--CCAT--A--G--C--T--T------GTTT-CA--GCA---......-A-................  7085
MON.CM.L1                ...........-GG-GATTTCAAC-CCA-------AAG--GG-CCAG--T---T-G-----C-CCAG--A-----C--T--T------GTCT--GCTT-A--C-T--A--GAA-AACC--C-AA--GGA  7383
MON.NG1                  ...........-GG-GAT-CCAGC--AA---A---AAT---G-CCAG--C--TT-G--C------CAG--A--G--C--T--T--C---CTGT-TAG---A--C-TYTCG-G-GACCCG-R--AG-A-AA  5999
TAL.CM.8023              CACAAATAAGG-GGC-AGG-GT-AG-AAG--A----T---G-ATC-CTCTTT-GCA-GCG--CCA---A--G--C-----T------ATTT-TAGCA----C-TG-AA--CAAC-A--CATAC-A--A  6770
TAL.CM.266               TGAACATCACA-GG-C-GC-GT-AGGA---A----T---GG-ACA-CTTTT-T-GCA-GCA---CAG--A----------T--C-CAT--GTTCTA-----CCTG-A-----CC-A-G-GTA--AC-A  7223
MUS.CM.1085              ...........-GG-GATC-GAGC--AAG--A--T-A---GG-CCAG--A--TT-G------------G--A-----C-----T------CTCT-TGTG-----CC-ATT-GT-AACCAC-CAGAG-TAAA  7360
SAB.SN.-.SAB1C           CAAACAAGATT---CTA---AGACA-T-G--A--T------TC--A-TTCTT-TT----------CA---T-----C-----T-----AATGGATTGGT-CC-C--CTA-TTGAATAA---GTCAGTAGA  7721
GRV.ET.x.GRI_677         ATATACATCTGAGA-G-AT--GG......--A--T------TCA-CG-ATTTTT-G--C------CA---T-----------T-----GATGGACTGGT-TA-C---TA-CTAAACAA-CGA-CAGAAGA  7412
TAN.UG.x.TAN1            CAAGGCAGATT-GGCTAAGTAGACA-T-G--A--T------GCA-CT-ACATTT-GC-C-----CCA---A--G--------T----C-C-TGATTGGT-TG-----T-GCTGAATAA-G-GTC--A-A-  7510
VER.DE.-.AGM3            CTGAAGAGATT-A-CT---GAGACT-TTT--A--T------GCA-C--ATTTAT-G---------CAG--------C-----T-----AATGGATTGGT-TC-A---TACCTGAATAA-CG--CAGTAGA  7048
VER.KE.-.9063            CAGAGAAAATA-A-CT---GAGACAGTTT--A--T------GCA-C--ATTTGT-G--------CCAG--A--G--C--------C--GATGGACTGGT-TC-A---TA-TTGAATAA-C-ATCAGTAGA  7550
VER.KE.-.AGM155          CAAACAAAATC---TT----AGACA-TTT--A---------GCA-C--AT-TAT-G--C--C---CA----------C---------AATGGACTGGT-T--A---TA-CTGAATAA-TTA-CAGTGGA  7530
VER.KE.-.TYO1            CAGAACACATA-ACCT----AGACA-T-G--A---------GCATCC-ACTTGT-G---------CA---A-----C--T--T-----GATGGATTGGT-T--A---TACTTAAATAA---A-CATGGGA  7022
SUN.GA.98.L14            TTAAGTTTAAACCA-T-G-GAAGCAT----CA--TG---C--CA-AG-T-TTG-TGA-G-C--------AA--GA-G----TT-------T----AG-A-T--C--AGTATGGAATGAC-CA-CCT-CAA  7752
LST.CD.88.447            TACCTATAGCAAA--T-AC-AGAAG---G--AA-T-TT-C--CA-AG-T-TTG-TGCA--CA-------AA---A-G----T---C---GT----AG-A-A--CC-GGAATGGAATAA---A-ACT-AGA  6609
LST.CD.88.485            TACCTATAGCAAA--T-AC-AGACGT--G-AA-GT-TT-C--CA-AG-T-TTG-TGCA--CA-------A----A-G--T-T---C---AT----AG-A-AC-CC-AGAATGGAATAA---A-ACT-AGA  6606
LST.CD.88.524            TACCCATAGCAAA--T-ACGAGAAG-----TA--T-TC-C--CA-AG-T--T--TGCAC-C------T-A----A-G----TT------GTC---AG-A-C--CC-GGAATGGAACAA---A-ACT-AGA  6612
LST.KE.-.lho7            TACCAATAGCTAA--T-ACTAGGCAT----CA---TT--C--CA-AG-T-TTA-TGCA--C--------A----A-G----TT------GT---TAGGA-A--CC-AGAATGGAACAA---A-A-T-AGA  7684
SYK.KE.-.SYK173          .....ACCACA-GG-G-TCCCAGC----G--A--T-T---GG-ACG---A--TT-G---CC-A------G--A--G--C-----------GT-T-TA-G--C---GC--AC-T-AC-AA-GGA-ACG-AA-  7411
SYK.KE.KE51_AY523867     ...........-GG-G-TCGCAAC--CA---A----TG---G-GAG---A--TT-G---C-A-----C--A---------T-----C---GTTT--A-GA-T---GCA-A-GTG-CAAA-GGCTA-G--AA  7107
COL.CM.-.CGU1            CTATCACTCCAAGG--TGG-AATAG-AC-A-T--T-----GGC-ACTTTCAC-TT-G-G-T----CATA--TT---C--T---------G--T-TAGTT-A-GG--ACA-GA--CC.............  6852
```

```
                                                  V4 loop                                                   -|
H1B.FR.83.HXB2          GAGTACT......GAAGGGTCAAATAACACTGAAGGAAGTGACACA...........................ATCACCCTCCCATGCAGAATAAAACAAATTATAAACATGTGGCAGAAAGTAGGAAAA  7520
Env                     __S__T__.__.__E__G__S__N__N__T__E__G__S__D__T__.__.__.__.__.__.__.__.__.__I__T__L__P__C__R__I__K__Q__I__I__N__M__W__Q__K__V__G__K_ gp120

H1D.CD.84.84ZR085       T-----A......-GGTTAAATG---CA-TC..........................................--A---A--------------------------------------G-----------  7041
H1F1.BE.93.VI850        T........................................................................-----T--------TC-G------------G--------------GG-------CG-  6800
H1G.SE.93.SE6165        T---GAG......A-T---A-T...................................................-----A--------T-AG------------G-G-GA----------GG--G---C--  6926
H1H.CF.90.056           TTAC--A......TC-AATGAC-CA--GGGAA-C-A--AC.................................--T--A--G---------------------G---------------G-------CG-  6839
H1J.SE.93.SE7887        TGAAG--......AC-AATGAC-CA-GTNAC-C-ACT....................................-----AA-A-----T-A-------------G-G-GA--------A-GGAC----C--  6815
H1K.CM.96.MP535         .........................................................................-----A--------T--------G--G-----------------------G------  6683
H101_AE.TH.90.CM240     C-TGG-G......-GGT-TAATG-C-CT.............................................-----A--T-------AG-----G---------------------GG--C----C--  7073
H102_AG.NG.-.IBNG       C-C-G-C......A-CCACA--GGGTCA-A---CACT....................................--A--T----A------------G------G----T------------------C--  7029
H103_AB.RU.97.KAL153_2  -TTA-A-......A-CACTGA-GGAG-T..........................................ATAG----A---------------------------------------G-----------  6697
H104_cpx.CY.94.CY032    A-AC-T-......AC-A-TA--G--TCT--AA-TTC--CC.................................-----A----A-------C--------T--G---GG---------G----G--GC--  6896
H1N.CM.95.YBF30         TGAGC--..................................................................--T--T--G--T--T-------G---G---G----TT-----ACA-GG---------  7068
H1O.BE.87.ANT70         A-CC--C......TGTA-TGTT-G---TGT-......---C-AGGTAACAATGGC.....................--T--A--T----A-C-G-G---GG-GG---GGTCA---ATA-GG-G-CAGTCG  7559
H1O.CM.91.MVP5180       -TCCGGA......TGCCA-GAG-TC--AGGGAGCAATGAGAC--ATAAAAATGGT.......................--TA-A--T----AGT---G---GC-AG---GATCA---AT---G-G--AGTCG  7576
CPZ.CM.04.EK505         ................--CAAC-CA-CTGGC.......................................AAT------T-A-A-----------G-------G----TT-A---ACA-G----------  7006
CPZ.CM.04.LB7           T-AA--CTTGCTCA-T-AAA-CTTGC---A-.......................................AACT-G--A--A--T--T--G--T-G-------G-----C-A----TA-G---------G  7040
CPZ.CM.04.MB66          .......ATCTTTA-T-A-A-T--GG-TGA-.......................................AAT-----TA-T--T----------GG------G---GGC-A-----A-G--------G-  7141
CPZ.MT145               .............AT-AACAA--C-G-G-A-.......................................ATG-CA-TAA----T--T-------G------AG-G--TTCC---AT-CGT-----G---  6999
CPZ.CD.-.ANT            CCTC-TC......AC-AACGG-GCCCT-................................................-TAGCA-AT--------T--G--G--AG-T--TCAT---GGC-T---TTC----  6955
CPZ.CM.-.CAM3           A-C--AG......-G--ATAAC-CC-C-................................................TT-T-------T-------GG------G-----TCC---AT--G---G--G--G  6927
CPZ.CM.98.CAM5          CCAA-AC......A-CT-CA-C-GA--T-ACAC-AATGACACA-ACTCAGACAATAGTACACTA............--AA-AT----T-------GG------G-----TCC---AT--G---G--G---  7239
CPZ.GAB_X52154          ....-T-......AC-AATGGC-T--TA................................................-TA--G-----------T-G---G--AG-T-GTTCA---AT--GG-------G-  7539
CPZ.GAB2                ....-G-......AGT-ACA---G-G-G.............................................TA--TT----------A---T-GG---G-AG----TTCT---AT--G---G--G---  6895
CPZ.TZ.-.TAN1           C-A--A-......ACTCATGGT-C-CTT................................................GTTGCA---------C-G-GG--G--AG----TCAT---GGT-T---GTC----  7143
CPZ.US.85.CPZUS         .-A---C......--CAACA-C-C--T-................................................--A--TAAG--T-------G-------G----TCA----AT--G----------  7516
GSN.99CM166_AF468659    C--C-GC.........AATATTGGG---........................ATCACAAGCAAATACAAAGGACAATGG--GG-T----A---T-GG---T--G-C-C-CAA---GGAT-T--GTCT---  7466
GSN.99CM71_AF468658     -GA-GAC......AC-AACATT--C--T........................GTCACAAGCAAATACAAAGGACAATGGA-GG-T--------T-GG---T--G-C-CACAA---GGAT-T--GTCT-GG  7472
H2A.ALI_AF082339        -GAACAG...............................................................CAC-ATTATGCA--G---CAT----G------A--T--T--C-----C----C---G---  7986
H2A.BEN_M30502          AC-GCGC..................................................................-A-TAT......---CAT----G--G--A--T--T-CC-----T--------G---  7996
H2A.ST_M31113           --CACAG...............................................................CAC-ATTATG-G------CAT----G------A--T--T-CC-----C--G-----G---  7432
H2B.D205_X61240         A-C-CGG...............................................................AGA-ATTATGCG----TCAT--C-G---G--A--C----C-------C--G--T--G---  7978
H2B.EHO_U27200          C-AG-GA..................................................................-ATTATGCGT-----CAC--C-G---G--AG-C----C------C--GA-T----G-  7961
H2G.ABT96_AF208027      TCTCCTA..........................................AAGCCCCAGGAACGACAAAGACGC-ATTATG-A--C---CAT--T-GG--G--A--T----CA-----C------------  7357
H2U.FR.96.12034         -G-ACAA..........................................ACGTCAAGGGAGAGACATAAAAGG-A-TATG-G--T---CAT--T-GG--GG-A--T----CA-----T--------G-G-  7483

MAC.239_M33262          T-AC..........................................CAGAAGCCAAAGGAACAGCATAAAAGG-ATTA-G-G-----TCAT--T-G------A--C----CT-----T--------C---  7941
Env                     __N__.__.__.__.__.__.__.__.__.__.__.__.__.__.__Q__K__P__K__E__Q__H__K__R__N__Y__V__P__C__H__I__R__Q__I__I__N__T__W__H__K__V__G__K  gp120

SMM.SL.92.SL92B         C--CC-A...........................A-GT-GACA--TCAAACCAAAAAGGAACAGCACAAAAGA-ATTATG-A--T---CAC--C-GG-----A--T--T-C------C-GG---------  7392
SMM.US.x.PGM53          AG--T--...........................CT-T-GACACA-CAAACAAAGGCAGACGAAAAGAAGAGA-A-TATG-G-----TCAT--T-G---G--A--C----C------C-G------T---  7890
MNE.US.-.MNE027         TG-A....................................ACT--CCAGAAGCCACAGGAACAGCATAAAAGG-ATTA-G-G-----TCAT--T-G------A--C----CT-----T--------C---  7418
RCM.NG.x.NG411          C-A--AA................................................AATGAAGCTAGGCAAGCC--GTTTG-T--T--T-TC-CT---ATGG--G-T--TGAT---T-T-C---GTCT-G-  7266
RCM.GAB1_AF382829       -.................................................ACAACAAATGAGAAGAGAAAAGCACCTTTTG-A------TC-C----ATG--AG-C--TGAT---T-T-C----TCG-GG  7222
DRL.FAO_AY159321        TGTGTGG...............................................................TATCCATATGCAT-C---CAC-----------A---G-TGAT---GCA-G------C-GG  7337
MND-1.F17               --AGTGG...............................................................TATCCTTATGC-T----TCA---T-GG-----AG--G-TGAC---AT-C---------G-  7109
MND-2.5440_AY159322     AGT-TGG...............................................................TACCCTTA-A-GT-TC---AT--T-G------AG--GGTGAT-----T--------C--G  7291
MND-2.14CG_AF328295     TGT-TGG...............................................................TATCCATATA-GT----T-A---C-G---G--AG-GG-TGAT-----T-----G----GG  7590
MND-2.MND2_AF367411     AGTATGG...............................................................TATCCATGGA-GT----T-AT--T-G------A---G-TGAT-----T-----------G  7664
DEB.CM.CM40             ................................CTCCT-CAAA-T-ACACTAGAAATAGCACTTGGTCTGACAAGTGGTTGA-G--T--T--G-----T--GT-AG------CT---T-T-CG-----GC--  7263
DEB.CM.CM5              ................................CTGCTC-TGA-T-A-ACTAATAACATAACCTGGAGTAATAAC-CTTTGA-G--T--T-AG--T--T---T-GG------CA---TTC--G-----C--  7184
MON.CM.L1               CG-G.....................-TAT--CC-TTCGAC-TG-ATAACAAGCCCAACACCACGTATCATGGTGGGTGG--AG----T-CC----G------GG-G-CACAA---GGAT-T--GTC---G  7492
MON.NG1                 --A-............-AAA--TT-G-AT-CACCTTC-A--TTT-TACCAATGCCTCCACTGCCTATAACGGAGAGTGGT-AA-G---C--C-C-GG---C-AG---CTCA----GGAT-CA-CTCC--G  6117
TAL.CM.8023             C-CAG-AGAAAAA-G--TACTGC-CC-GGAGA-GAT-G-CCGACA-GAGCACAGGTATAGTGACAGAACCAAAT-TTT-A-GT-C---CAGT---GG-----G--C--T-G----AGC-------AG--G  6900
TAL.CM.266              C-CAT-AAACCAA-GCAAA-T-C-CC-AGGG-TCAT-G-AA-G...AAGCCGAGGTCGAGTGATAGGACCAAGT--TA-A-GT-----CA-C-G-GG-----G--C----G---AGC-GG--G-AG--G  7350
MUS.CM.1085             TGAA..............................A-C-ACAT-CATACTGTAACAACTAAGTACCACAAGGATCAATGGA-GGTC--T-----T-G----T--G-C---CA----GGAT-T--GTC---G  7460
SAB.SN.-.SAB1C          CCCA.....................G--CA-A-CAACT---CA-A-AATAATACAAAGCCATGTTGGCAGAGA-C-TATG-T--C---CAC---GG---G-AG---TGAT---T-C-CCT-GTCT---  7830
GRV.ET.x.GRI_677        TGCAGAA.........--TA-T----GG--CTGT-AC-AA-GG-AGCCAGGACCAGGACCATGTGTTCAGAGA-CTTATG-TG-C---CAT---CG----G-AG---TGAT---T-C-CT--CTCT---  7533
TAN.UG.x.TAN1           TG-A-GAAATGTA--T-TAGA-GG---T-A-TGCACC-C--GA......AAGGATAAACGCTGCTACAAAAGG-CATATG----C---CAT--T-GGTC----G-C--TGAC---T-C-CGC-CA-T--G  7634
VER.DE.-.AGM3           TCCGGAC......C-TAATC-GTG---TGG-ACGAAGG-AA-AGGTAAGGCACCAGGACCCTGTGCACAAAGA-CATATG-TG-T---CAT---CG-TCTG-C--T--TGAT---T-C-C-C--TC--GG  7172
VER.KE.-.9063           TGCA-A-......C-TAATCA-TG--GT-A----AC--AAA-AGGGAATGCCCCAGGACCATGTGTACAAAGG-CTTATG-TG-T---CAC--C-G-TCTG-C-----TGAT---T-T-CCA--TC---G  7674
VER.KE.-.AGM155         TGC-GA-......C-TAATCATTG---A-ACA-C-C-G-GA-AGGTCGAAGTCCAGGTCCCTGTGTACAGAGA-CTTATG-TG-C---CAT--CCG-TCTG-C-----TGAT---T-T-CTA--TC---G  7654
VER.KE.-.TYO1           TGCATAC......C-TAAT-TTTG--G--GCA--AAG...A-AGG-CACGCACCAGGACCATGTGTACAAAGG-CGTATG-TG-T-A-CAT--C-GGTCTG-A-----TGAT-CCT-T-CCC--TC---G  7143
SUN.GA.98.L14           T-AATGG...............................................................TATCCTTGGGC-AAT---CAC------AGTG-G---G-TGAT---GC-TC---G--C--G  7819
LST.CD.88.447           T-AATGG...............................................................TATCC-TGGGC-AAT---CAC------TC-G-----G-TGAT---GCATCT-----G--G  6676
LST.CD.88.485           T-AATGG...............................................................TATCC-TGGGC-AAT---CAC------TC-G-C---G-TGAT---GCA-CT-----G--G  6673
LST.CD.88.524           T-AATGG...............................................................TATCC-TGGGC-AAT---CAC------TC-------G-TGAT---GCA-CT--T--G--G  6679
LST.KE.-.lho7           T-AATGG...............................................................TATCCTTGGGC-AAC---CAC------AGC--C--TG-TGAT---GCT-C-A-------G  7751
SYK.KE.-.SYK173         T-AG-A-...............---T-TG-C....................................AGTAACC-TCGT--GT----TGCC--TCGC-----A--T--TGAT---AGAT-T---A-----  7490
SYK.KE.KE51_AY523867    TCCA...............AGC---T-TG-A....................................AAAAATT-ATAT---T-G--TGCC---CG---G---------TAT---GG-T-T---ACC--G  7186
COL.CM.-.CGU1           .........................................................................CCAGT-A-GAAC--T-CC--T-G-A-GC-GG-G--T-GT---GTC-CTCAT-CC-GG  6909
```

```
                                                                                                                       |-    V5 region
H1B.FR.83.HXB2             GCAATGTATGCCCCTCCCATCAGTGGA.........CAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAAT.................AGCAACAATGAGTCCGAGA  7623
Env                        _A__M__Y__A__P__P__I__S__G__.__.__.__Q__I__R__C__S__S__N__I__T__G__L__L__L__T__R__D__G__G__N__.__.__.__.__.__.__S__N__N__E__S__E__ gp120

H1D.CD.84.84ZR085          --------------------TGAA--G.........-------AT----------------------A-----G---------------GC-..................-AT--T-C-C--AAT--C-  7144
H1F1.BE.93.VI850           ---------A--T-------TGCA---.........A-C----CC---AAC-------------TT-A-----G------------------....................G-G-G-A-CATT----  6900
H1G.SE.93.SE6165           ---------------------GCA---.........A-C---GA----AAC----GC--------A--AA--------------------G---AAT.....AACAATACA-AT-CA-G----AGT----  7041
H1H.CF.90.056              ---------------------CAA---.........A-C----TG---GT--------------A--AA----G----TT--C-AG-----C.................GCGTCTGCA--AAATT-T-  6942
H1J.SE.93.SE7887           -----C--------------TGCA---.........A-C----CG--CA---------------AT-A-----G---C-T-----A--A--CAGG.....GGGAATGGC--TG-G----GCA-T----  6930
H1K.CM.96.MP535            -----A--------------TGCA---.........AGT----AC---AGC-------------AA--A----G---------------..........................-AT--T-C-C-TAAT----  6783
H101_AE.TH.90.CM240        -----------T-------------------.........AG-----AT---GT--------------AA-A-----G---------------GT-AAT.....................--T-CG--TAA-----  7176
H102_AG.NG.-.IBNG          ---------------------CAA---.........AT---A------GAT-----C--------A--A---------------A--G---......................--T-G-ACAAAT--A-  7129
H103_AB.RU.97.KAL153_2     ----G----------------GCA---.........---------------------------------A-------------------------C.................CAG-G-----TTA------  6800
H104_cpx.CY.94.CY032       --------C---T-C-----TGCA--G.........AGC----AC--CAGC---G----------AA-AA----G----------------C-.......................---A-TA-T----  6993
H1N.CM.95.YBF30            -G---T-----A--A--A--TC-G---.........GTTC---AC---A-C--C--------T--A---G-TC--GA-TATAG------GCC-.....................G---CCA--GAAACA-  7168
H1O.BE.87.ANT70            -G-C-C-----A----------AA--T.........A-TC-A-C----ATG-----C--A--T--AA--A-CC--CA--TG---AACACATGGAACAGC...AGCAACAAC-ATGTA-CATTTAGACCA-  7677
H1O.CM.91.MVP5180          AG---C-----A---------CCC--C.........A-CT-A-C----CAT--C--C--A--T--AA--A-TC--CAGTT----CAACCATGG.................-ATTC--CA-GTGAAA-T-  7679
CPZ.CM.04.EK505            -G---T-----T--G--A--T-AA---.........-CT----AC---CTC--T--C--C--T---A-AA--C-TGAC.........................TATACAAAGTCTGGA-C----AAGT-C-  7103
CPZ.CM.04.LB7              -G---A-T---T--A--T--T--A--T.........A-C--CC-C---AAT--C-C------T---T--A-CC--GA-.........................AAACAC-CA----CA--CAATATC-  7134
CPZ.CM.04.MB66             -G---A-T-CTT--C-----T--A---.........AC---A-AC---ATC--G-----C------A-AT-G--TG-G..............CAGCAA...AAAACAGAC--G-TG--CA-A--A-CC-  7244
CPZ.MT145                  -G---T-T---A--A--A--T--A--T.........A-C----CC--CA-T--T-----C-----TA-----C--GAG-T-C--AAGAACCGAGAAGATCAGGGAGAAGATCAGG-TC-GA-TAA-ACAT  7120
CPZ.CD.-.ANT               -GC--T--CTTAG-C--T-GG--A--G.........A-TG--TCC---A-T--C-GC--A--T--AA-TA-G--G...GA--G-CAAATAT--.........................-----AA-T-TT-  7049
CPZ.CM.-.CAM3              -G---T--------A--A--T--A---.........GTT--A--T--CA-------CG----T---A-TA--C-TGAG-C--GGCA---C-TC..............AAC-A--G--TCACCAA-ATC-  7033
CPZ.CM.98.CAM5             -G---C--------A--A--A--A--G.........A-C--A-C---CAT---------A--T--A--C--CC-CGA-GC--GA--AAAATGGGAA...........AAC.........A-TA-TATA-  7339
CPZ.GAB_X52154             -G---A-----A--C--A-----A---.........A-C--C-CT--CAAT--T--C-----T--A--T--C-----TTC---CAC-CC-GTGACT...........AAC...----G--GTAA-CT--  7645
CPZ.GAB2                   -GT--C-T------G--T-GA--A---.........ACT--A-CC---AAT--C-CC--------A--C--C---GA-GT-C-GAA-.......................G---CGGGGA-CAA-ACAG  6992
CPZ.TZ.-.TAN1              -GGG-T--CCTT--C--A-GG--G---.........AC-G-A-A----CAC-----C--C-----A--TA-CA-G---GC---AAAA-AC--CAAT.............................A-T-  7234
CPZ.US.85.CPZUS            -G---C-T------A--A----AA--T.........GTGC-A--T---AAC--------A--G--AA--A-TC-TGAC-T-AGCATAA-CGCAGTC.........AATAAC...G-T-G-AG-AATATA-  7625
GSN.99CM166_AF468659       T----C--CTTA-----ACGGCAG--T.........--T--A-AT--CA----C--C--------T---T-GA--GATG-C-C-ATGTA-G-AAGC...........................--TATC-  7560
GSN.99CM71_AF468658        T-G--C--CTTA-----ACGGCAA--T.........--T--A-AT---A----T--C--C--C--T--C--C---A--GACG---CGATGTA-GGACAT..........................---ATC-  7566
H2A.ALI_AF082339           AATG-A---TTG-----T-GGGAA---.........G-GT-G-CC--CAAC----CAG-A--CA-CT--A-TGCT-AC-TT--CACG-A-GGC............................AA-C---  8077
H2A.BEN_M30502             AATG-A---TTG-----T-GGGAA--G.........G-GT-GGCC---GA-----CAG-A--CA-CA-AA-TGCT-AC-TT--CATA-A---AAAT......................CG-A-TC-T-  8093
H2A.ST_M31113              AATG-A---TTG-----T-GGGAA---.........--GT-A-CC--CAAC--T-CAG-G--CA-CA-AA-TGCT-AC-TT--C--A--AG-G.............................AA-C---  7523
H2B.D205_X61240            AAT--A---TTG-----A-GAGAA--T.........G-GC-CTC----AAT--C-C-G-G--TA-C--CA--GCC-AC-TCA-CTCA-AC.............................A-CAGTACC-  8069
H2B.EHO_U27200             AATG-----TTG-----A-GAGAG--T.........G--C-CTCC---AAT--C-C-G----CA-C--CA--GCC-AC-TT--CT-GATAG--AAG..........................AA-CTT-  8055
H2G.ABT96_AF208027         AATG----CTTG--C----GAGAA--T.........G-TC-G-CG---AAT----CAG-C--CA-C--AA--GCG-AT-T---A-AGAA--GGTCTAAC.........................AA-C-T-  7454
H2U.FR.96.12034            AATG-----TTG-----A-GAGAA---.........G-TC-C-CC--CAAT--T-CAG-A--TA-CA--A-TGC--AT-TT---C-GAT---CAAA..........................AATC-A-  7577

MAC.239_M33262             AATG-T---TTG-----A-GAGAG---.........G-CC-C-CG---AAC--C-CAG-G--CA-T--CA--GC--AC-T----T-GAT-G--GGA..........................AA-C-A-  8035
Env                        _N__V__Y__L__P__P__R__E__G__.__.__.__D__L__T__C__N__S__T__V__T__S__L__I__A__N__I__D__W__I__D__G__.__.__.__.__.__.__.__.__.__N__Q__ gp120

SMM.SL.92.SL92B            AATG-C---TTG-----A-GGGAA---.........G-CT-A-CT---AAC--C-CAG-G--CA-C--AA--GC--AT-TT---T-GAT-G-CAAC..........................AAT----  7486
SMM.US.x.PGM53             AATG-----TTG-----T-GGGAA---.........G-CC-G-C----AAT--C-C-G-A--TA-C--CA--GC-GAT-T---CT-GAT-G--GGC..........................AA-C---  7984
MNE.US.-.MNE027            AATG-T---TTG-----A-GAGAG---.........G-CC-C-CG---AAC--C-CAG-A--CA-T--CA--GC--AC-T----T-GAT-G--GGA..........................AA-C-A-  7512
RCM.NG.x.NG411             AA-G-A--CA-T--A--A-GGCCA-AT.........GCCT-G----CAACG-C-C-G-G---TACT--T--GC-GAT-TT--CTA-ACAG--AAT.........................ATGACA-  7360
RCM.GAB1_AF382829          AAGG-A--CA-G--A--G-GGCCA-AT.........GCGT-A-AG--CAGTG--C-GG-AT-CTAT---T-GGC-GAC-T---CTA-...................-TT-----CAGT----  7316
DRL.FAO_AY159321           AAG--T---ATG-----TG-ATCA---TTTAATAATG----A------A--CA-G--G-C----AAA--T-C--TGAG-TCC-AAAATTGG-GACAGAG...GAAACAGGAGATTC---C...-ATC---  7461
MND-1.F17                  AAG--A---TTA--A--T-CATCA---TTTAATAAT--C--A--G---A--CAT-GGG-A----AAA--TAC--TGA--TGC-AAAGATAG--AGTAAT...............G-A-CAA-AATGC-A-  7224
MND-2.5440_AY159322        AA---T---ATG-----TG-GTC---GTTTAATAATG----A------A-TAATG--G-G----AAA--T-C--TGAGGTCC-AAAAAC-G--GATGAC..................----GA-ATATC-  7403
MND-2.14CG_AF328295        AAG--T---ATG-----TG-GTC----TTTAATAAT--T--A-----CA-TAATG--G-C----A-A--T-T--TGAGGTCC-ACTA-T-G-AGGA.....................--AAGA-A-CTA-  7699
MND-2.MND2_AF367411        AAG--T---ATG-----TG-ATC----TTTAATAATG-G--A--------TAATG--G-A----AAA--T-C--TGAGGTTC-AAAGACAG-AGAGGGA........................-A-ATC-  7770
DEB.CM.CM40                CAT--C---CTA-----G-AAGAA---.........G-GT-G-AG--CAGC---C-C---T-T-CAT-TG-T--TGATGT----CACTA----GGG.............................AGT-  7354
DEB.CM.CM5                 CATG-A---TTG-----A-AGGAA---.........G--T-G-AG---CAT--TTC-G--T-T-CA--TG-T--TGAGGT--C-CA-TA-GGGCAA.............................A-C-  7275
MON.CM.L1                  AGT--C--CCTA-----ACGG-AA--C.........--TG-GCAG---A----T-----C--T-CC--C--GA-T---G-G--GTTGTA-C-AAAC.............................A-CG  7583
MON.NG1                    AGTR-C---CTG--A--A-GACAA--G.........--TG-GCAG---A-C--T-----C----CC--CT-GG--GATG-G--GTTGTA-GGCAGC............................ACAG  6208
TAL.CM.8023                CT-------CTG--G--A-GAGAG---.........--CC-G--T--CAA-AGC--CG-C-----TA-C--GG-GGAC-TCA-CCACTA---GGGGAAT.........................GA-ATC-  6997
TAL.CM.266                 TT------CCT---A----GGGAA---.........--CC-GCAG--CAA-AGT--C--C-----AT--T-GG-GGAT-TCA-CCACCACCGCGGCTCT....................AGTGA-ATA.  7449
MUS.CM.1085                T-T--C--CTTG-----ACG--AG--C.........--TTG--C-----GAG--T-----A--------T--CA-TGACG---CCATGTA--GAGAT.............................TCA-  7551
SAB.SN.-.SAB1C             AAG-CA--C----------GAGAA--C.........--CT-GGAG---AAC--G-CAGCC----C----TATG--GA-TTGA-CTA-AAC-G-AAA..........................AA-AG--  7924
GRV.ET.x.GRI_677           AAGG-A-----T--A--A-GGGAA--T.........--TT-GGAG---AAC---TCAG-C--G-CA--ATACG-GG---T----TA-AAC--CAAGTCT........................GG-CCA-  7630
TAN.UG.x.TAN1              AA--CC-----A--A--A-GGGAA---.........--CT-AGA---CA------CAG-G---TCTA--A-GG--T--CTG---TA-AAC-GCAAA..........................GAAAG--  7728
VER.DE.-.AGM3              AA--CC-----A--G--A-GAGAA--G.........--CT-GCA---CA----C-CGG-A--G--TA--TC-G-GGAGCT-A--TACAA--G-AAG..........................AA-AG--  7266
VER.KE.-.9063              AA-G-------T--A--A-GGGAA---.........--TC-GGA---CA----C-CGG-C--T--TA--AC-G--GA-CT-A--TA-AAC--AAC..........................-ATAGA-  7768
VER.KE.-.AGM155            AA--CA-----T--A--A-GAGAA---.........--TT-GCAG--CA-G--C-CAG----T---A--AC-G--GAGCT-A-CTA-AA---CCAG..........................AA-AG--  7748
VER.KE.-.TYO1              AA--CT-----T--G--A-GAGAA---.........--TT-GCA---CAGG--C-CAG-C--T---A--AC-G-TGAGCTTA--TA-AA--G-AAA..........................AA-AGA-  7237
SUN.GA.98.L14              AAG--T--CTTG--C--A-CATCA---GTTTAATAACAG---A------G--AAC-GAG-G----AAGCATGG--C---CTG--AA-G-TGG-AGACTGG...AAGGTAAATG-ATCA--CATC--A-TTG  7946
LST.CD.88.447              AA---T---CT------A-CATCA---CTTTAACAACAG---A------A-TCAT-GAG-A----AAA--TGG--TCA--TG--AAAGTT-G-ACCAGAA...GAAAACCTGG-AGGA---TT---T-TC-  6803
LST.CD.88.485              AA---C---AT------A-CATCA---CTTTAATAACAG---A------A-TCAT-GAG-A----AAA--TGG--TGA-CTG--AAAGT-GG-GCCAGAG...GAAAACCTGG-GGGA---TT---T-TC-  6800
LST.CD.88.524              AA---T---CTT-----A-CATCA---TTTAACAATAG---A------A-TCAC-GAG-G----AAA--TGG--TGA--TG--AAAAT-GG-ACCACAT...GAAGACTTAG-AGGA---CT---TATT-  6806
LST.KE.-.lho7              AAG--T---CTA--A--T-CTTCA---TTTCAACAATAG---C------A-TCAT-GAG-A----AAA--T-C--TGAG-TG--AAAAT-GG-ACCACAT...GAGGACTTAG-AGG---CCTAAGTATA-  7878
SYK.KE.-.SYK173            TTG--A--CCTG--C--T-CGGCA--G.........--T----A-----A-T--C---G-C--T-CAG-AT-GACGGAC-T---ATACTACCCAGGA...........................--TAC-T  7584
SYK.KE.KE51_AY523867       CT-------TTG--A--T-CAGAA--G.........--C----AG--CA-T--T--------T-CTG-AT--AC-GAT-T---GTACTA-CCAAAT...........................--AC--T  7280
COL.CM.-.CGU1              ATCC------GG--G--TCCAG-A--T.........--TC-ACA----AAC-GGG-G-AACAGCCAG-TA-TGCGTTC-TG-GCACCATAG-AGGA.........GACAATGATGGT----G--GT...G  7018
```

```
                                                      Rev responsive element secondary structure in RNA, see Charpentier, J Mol Evol 266:950-956(1997)
                                                                   see HIV-1 complete genome annotation in compendium for stem-loop positions
                        V5 region     -|                                                                                        |- Rev responsive element start
H1B.FR.83.HXB2              TCTTCAGACCTGGAGGAGGAGATATGAGGGACAAT...TGGAGAAGTGAATTATATAAATAT...AAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAG..  7745
Env                         I__F__R__P__G__G__G__D__M__R__D__N__.__W__R__S__E__L__Y__K__Y__.__K__V__V__K__I__E__P__L__G__V__A__P__T__K__A__K__R__R__V__V__Q__. gp120

H1D.CD.84.84ZR085           C---T----------------------A-----C...------------------------...--------------------C--------------------------------------G-A..  7266
H1F1.BE.93.VI850            C-----------AG------A------AA------...------------C-------G---...---------G---------------------------------A---CA-----------..  7022
H1G.SE.93.SE6165            -----------------------------------...--------------------G---...---AC---------CA--T--C-----------------G-----G-------------G--..  7163
H1H.CF.90.056               C----------------------------------...--------------G---------...--------------------C-G---A---------------A---G-------------G--..  7064
H1J.SE.93.SE7887            C----------AC-------A------A-------...------------------------...---------G-G-----G---C------------------------A-------------G--..  7052
H1K.CM.96.MP535             C----------------------C-----------...------------C-G---------...---------C-G---------C-----A-------------G---C-G---------------..  6905
H101_AE.TH.90.CM240         C-------------------A-C--A-A-------...------------------------...---------C-----------C-----A-------------G-------------------G--..  7298
H102_AG.NG.-.IBNG           C------G---------------------------...--------------------G---...---------------------C----T------------CGT-----A-------------G--..  7251
H103_AB.RU.97.KAL153_2      -T--T---A--------------------------...------------------------...--------T----------------------------------G---------------------..  6922
H104_cpx.CY.94.CY032        -----------------------C-----------...------------C-----------...-----------G---------A----------------AT-------G----C-------------..  7115
H1N.CM.95.YBF30             -AG-ATAT--CTC-------A-C---GTTA-TCTC...------CAA--G--G-----G--C...--------T-GC--A-----CA-------------AGGT--A--T--A---C-CACA---AGT..  7290
H1O.BE.87.ANT70             -AGGGG--GACATGAA--ATATATG---AACTG-ATTG-TC-ACTACA--G--GTA-GGGTA...---CCTT-T-GTG-G-C---CAC-C-TA-T---AGGC-AGTCAT--GC-CT---ACTCAT.....  7799
H1O.CM.91.MVP5180           CAC-T-----A-T---G----------AA--T-TA...-------C-A---G--C--C--C...----------C-G--AA----T--TA-T--------T--A--AATGTCA---CC-A-AA-AA-CAT  7803
CPZ.CM.04.EK505             C-A-ATAT--CAC---G-----C----CTA--CTC...-----GCAG--G--G-----G---...--G--G--C-GC--A-----CA-------------AGGA--A--T--A---CACACA---ACT..  7225
CPZ.CM.04.LB7               C----C-T--AAT------G-------CA----TC...--------------------C---...---A---------------------------G--T----G---G-GA--G--GACT---G-T..  7256
CPZ.CM.04.MB66              -G----C-----TT--G-----A-----AA-T---...--------------C-------...----------G----------C---------------------------A------AC------T..  7366
CPZ.MT145                   A-G-TT-T-TAAC---G--TA-C----AA--T-T-...------------C----C-----C...---A----TG-G--AC-----C-------------T--T--AAGT-GA---TAT-CA--AG-AAA  7244
CPZ.CD.-.ANT                AAG-GTC-----CT-C-A---TAGCA...---C-A...------GCG--G--G-CC-GG--C...C-G--G---G-G----RT--C--GTC------C--A--A-CA-GN--A--GCC--AAA-AA-ACA  7170
CPZ.CM.-.CAM3               CGC-ATAC---AC-------A-C---GTA---TTG...-----GCT------GC-------C...--G-----T-GC--A-----CA-------------A-GT--A--T--A--GCACACA--ATCA..  7155
CPZ.CM.98.CAM5              CAG-GTAC---TC---G-----A-----A---TTG...-----GCT-------A-C-GG---...--------T-GC--A-----CA-------------A--T-----T--A--GCACAC----GCA..  7461
CPZ.GAB_X52154              CA--T------AC-------A-C----AA--T-T-...-----------GC-C-----G---...--------TCG---A--G--T---TC---G-----A--A--A----GA--GCATACA---GCAAG  7769
CPZ.GAB2                    AAG-TTAC-TGTC---G-----------A--T-T-...-----G------------------...---A------------G---C-G--G--G-----A--A--A--------GTACACA---GCAAA  7116
CPZ.TZ.-.TAN1               GT-AT-CC--CCA-TTTTCT-C-G-AGTA--AG-CTAT----A-GTA------GCA-G----...-----G--GG-----C-G--C--GTC---G--T--A-GGCCA-G---A--GCCT-AAA-TA--GC  7361
CPZ.US.85.CPZUS             CAG-G-TG---AC--------------C--CTTTA...----A--A------C----G---...--G--G--C-GC--A-----TA-------G-----AGGT-----C--A--GCATACA---A-A..  7747
GSN.99CM166_AF468659        A-A-G-CC--GTCG-CC...---G--GCT--TGCG...----ACTA------G-CC-GG--C...-----G--GG-G--A--T--GC-CTCGA-G--C--A--TCC---T--A--G-A--AACAC-CC..  7679
GSN.99CM71_AF468658         ATA-G-CC--ATCG-CC...---G-AGCC--TGCA...----AGTA---G----CC-GG--C...--G--G--GG-G--A--T---C-TTCGC-G--C--A---TC---TC-A--G---CCACAT-CAGG  7687
H2A.ALI_AF082339            CAAAT-TTA-CTTTA-T-C---GG--GCA--ACTA...-ACC--TTA-----GGGGG-T---...---T-----G-G--AAC----A-T--CT-C-----T--ATCA-A--G----TACTCCTCT...AC  8198
H2A.BEN_M30502              C-AA--TTA-CTTTA-T-C---AG--GCA--ACTG...-ACC--TTA---C-GGGAG-C--C...---T--A--G----AAC----A-T--CT-C-----T--AG-TCAG-GA--GTACTCCTCA...AC  8214
H2A.ST_M31113               CAAAT-TTA-CTTTA-T-C---GG--GCA--ACTA...-ACC--TTA-----GGGGG-T---...---T-GA--G--G-AAC----A-T--CT-T-----T--ACCA-T---A---TACTCCTCT...GC  7644
H2B.D205_X61240             C-AAT-TTT---TGA-T-C---AG--TCA--ACTG...-AT---TTG---C-GGGGG-C---...---T-----G----AAC----A-T--CT-C-----T--AG-T-T--GA---TACTCCTCA...GT  8190
H2B.EHO_U27200              CTAAT-TTA---TGA-T-C---AG--TCA--ACTG...-AT-A-TTG---C-GGGGG-C--C...---T-----G-G--AAC----A-T--CT-T-----T--A-GTAT---A---TATTCCTCA...GT  8176
H2G.ABT96_AF208027          CAAAT-TCAT-TTCA-T-C---GG--GCA--ACTG...-ATC--TTA-----GGGGG-T---...---T-----G----AACT---A-T--CT-T--------A--T-T-------TATTCCTCT...GT  7575
H2U.FR.96.12034             CAAAT-TCT-CATGA-C-C---AG--GCA--ACTG...-AT---TTA--G--GGGGG-T---...---T-G---G----AAG----A-T--CT-T--T-----ACCA-TG--A---TATTCCTCA...GT  7698

MAC.239_M33262              CTAAT-TCA-CATGA-T-C---GG--GCA--ACTG...-ATC--TTG-----GGGAG-T---...---T-----G-G--CACT---A-T--CT-G--C-----AG-T-TG-----GTACACTACTGGTGG  8159
Env                         T__N__I__T__M__S__A__E__V__A__E__L__.__Y__R__L__E__L__G__D__Y__.__K__L__V__E__I__T__P__I__G__L__A__P__T__D__V__K__R__Y__T__T__G__G gp120

SMM.SL.92.SL92B             C-AAT-TTA-CATGA-T-C---AG--GCA--ATTG...-ATC--TTG-----GGGGG-T---...---T-G---G-G--CACT---A-T--CA-G--T-----AC-T-TG--A--GTACACCACA...AG  7607
SMM.US.x.PGM53              CTAA--TCA-CATGA-T-C---GG--GCA--ACTG...-AC---TTG--G--GGGAG-T---...---T-----G-G---ACT---A-T--CT-G--C-----A-GT-T--G---GTACACAACTGGTGC  8108
MNE.US.-.MNE027             CTAAT-TCA-CATGA-T-C---GG--GCA--ACTG...-ATC--TTG-----GGGAG-T---...---T-----G-G--CACT---A-T--CT-G--C-----A--T-TG-----GTACACTACTGGTGG  7636
RCM.NG.x.NG411              CAAATGTCA-CTTGTCT-CT--AG-AG-A--T-TA...---GC-GCA-----GGGC-G---C...--------GG-G---ATG---A-T---TAT-----A--A--T-TC-GA--GTAC-A-ACA.....  7479
RCM.GAB1_AF382829           CAAA--TCA-CCTCTC--CG--AG--G-T--TT--...---GC-GCA-----GGGG-G----...--G-C-A--G----CAG----A-T--CT-T-----A--AG-AAT---A--GTACCA-AC-.....  7435
DRL.FAO_AY159321            -TAAATTTATACCGCA--AT--AG--CAAA-TC-G...-TC-C-GCA-T-GG-GCAC-T--C...---T-G--C--GG-G--G---A-T--CT-T--T--A--TG-T-T-C-T---TACAACC---C-GG  7585
MND-1.F17                   -TAAATTCTTGCCTCCCA-T--A-CCTCCA-TC-A...-TTGTTGC-T-TGG-GC-C-T---...---T----C-----AATG---A-T--CA-------T--AG-T-TG--A---CACACTT-A...CC  7345
MND-2.5440_AY159322         -AAAATTTATACC-CA--ATTGG--TCAAA-TC-A...-AC-C-GCA-TGGGGGC-C-T--C...---T-G--G--GG-G--T--GA----TT-T--C-----AG-CAT-C-C---CATCATT---CTAA  7527
MND-2.14CG_AF328295         -AAAGTTT-TGCC-CA--AT--GG-ACAAA-TC-A...-TC-C-GCA-TGGG-GCAC-T---...---T-G------G-G--T--GA-----T-C--C--T--TG-T-T-C-C---TATCATC-A-CAGA  7823
MND-2.MND2_AF367411         --AAATTTGTGCC-CA--AT--GG-ACAAA--CGA...-TT-C-GCG-T-GG-GCAC-C---...C--C--------G-G--C---A-T--CT-T--T--T--AG-A-TGGCA---TATCATC---CAGA  7894
DEB.CM.CM40                 -AAC-CTTA-ACC-TC--C---C--C----CAGTA...------GCA--CC---T-------...---A-CA--G--G-AA----TA-----T-C--T---T--GCA-TG-GA--GTAC-A--GA...CC  7475
DEB.CM.CM5                  -GAA--T-A-ACCCTCCTC------T-A--CAGT-...-----GGCA--TC----C--G---...--GA-CA--G--G--A----CA-C--GA-G--C--T--A-GTAT--GA--GTAC-A--GA...CC  7396
MON.CM.L1                   -GAC-CTTGT-CCCTCG-C-C-GG----T---TCA...------TCA--GC-CAG-CG----...--G--G--GG-------C--C---TCTA-G-----T--A-CA--GC-------GACT-G-GTA..  7705
MON.NG1                     --ACGTTGAT-CC-TC--C-CGAG--TCT---GCA...---C-CTCA--G--GAGC--G--C...--G--G--CG-G--C--C--C--GTCT--G-----T--AGCA--CC-A--G---CAT-G--TA..  6330
TAL.CM.8023                 -AAGTGTGGAACCCTCTTC-C-A--C--A--TGC-...---C--GCA-CGG--GCACGG--C...--G--GA--G-G--CAC----A-T--CT-T-----A--G--TATC-GA--GTAT-AT-G--C-..  7119
TAL.CM.266                  ..ACAGT-GAGCCCTCGTC-C-G--AC-T---TCG...---C--GCA-C-G-T-CACGG--C...--G--GA--G-G--C-TG--T--T--GTAT--G--G--A--TATC-GA--GTAC-AT-G--CC..  7569
MUS.CM.1085                 --AAT-TGA--CC-TC--C-A--G--TT----GCC...---C---TG--GCCTA-G--GGTACAA--G--G--GG-G--C--T---C-TTCCA-G--------GG-A-TGC-A-----GCCT--C-CA..  7676
SAB.SN.-.SAB1C              C-AATGTGA-GTT-TC-CCTC-G--TGAAAG--T-...---GCT-A------GGGAG-T--C...---T----GG----CA-G---A-T---T-T--T--A-----A-T--GA---TACACT-GC.....  8043
GRV.ET.x.GRI_677            -AAATGTGA-CCT-A-TCCTC-GG-AC-CAG--TA...---GCGTAC---C-GGGAG-C---...---T-----G-G--AAC----A-T--CT-T--T--T--AG-T-T--GA---TATACT-GC-CCAC  7754
TAN.UG.x.TAN1               CTAATGTGA-ATTGAC--CTA--T--GA-A---TA...---GCTTA------GGGA-G----...--GC-CA--G-----------A-C--TT-C--G--A--AG--AT--GA---TAT--T-G--CAAC  7852
VER.DE.-.AGM3               CTAATGT-A-ATT-A-TCCCC-G--AGAAAC--TC...---GC-GCA-----GGGC-GG--C...---T-----G-----AC----A-T--CT-C--------AG-A-T--GA--GTATAC--GA...GG  7387
VER.KE.-.9063               C-AATGTGA-ATT-A-TCCCC-G--AGAA-GA-TC...---GC-GCA-----GGGACG----...---T-G---G----CAC----A-C---T-T--------G--T-T--GAC-TTATAC--GA...GG  7889
VER.KE.-.AGM155             CAAATGT-A-ATTGA-TCCCC-G--AGAAAC--TC...---GCGGCA-----GGGC-G---C...---T-G---G-G---AC----A-T---T-T--------AG-A-TC-G-C--TACAC--GA...GG  7869
VER.KE.-.TYO1               C-AA-GTGA-ACT-A-TCCCC-G--AGAATCT-TC...---GCGGC------GGGC-G---C...---T----GG----CAC----A-T--CT-T--------AG-A-T--G-C-TTATAC--GA...GG  7358
SUN.GA.98.L14               -AGC-TTC-AGCC-CCTACCA-C-CCTTAA-TC-A...-TTGTG--CAC-GG-GCAC-T--C...---C------G----AGG---A-T--GT-T--T--A--AG-T-AGC-CC-GTAC-CACCC.....  8065
LST.CD.88.447               GA---TTG---CCCTCCT-G--A-CA...A-TC-G...-TTGTGGCAAC-GG-GC-C-C---...---T--A---GGC--AG----A-T--CT-T--T--A--AG-T-TGC-TC-GTAC-CACCC.....  6919
LST.CD.88.485               GA---TTG---CCCTC-T-G--G-CA...A-TC-A...-TTGTGGCAACGGG-GC-C-T---...---T--A---GG---AG----A-T--CT-T--T--A--AGCT-AGC-CCAGTAC-CACCC.....  6916
LST.CD.88.524               AG--TTTG---CCCTC-T-G--G-CA...A-TC-A...-TTGTGGCAAC-GG-GC-C-T--C...---T-GA-C-G-C-GAG----A-T--CT-T--T--A--ATCA-AGC-TC-GTAC-CACCC.....  6922
LST.KE.-.lho7               AG---TTG---CC-TC-T-G--A-CT...A-TC-A...-TTGTGGCA---GGG-C-------...---T-GA-----T-GA-T---A-T--GT-T--T--T--AG-T-AGC-C--GTAC-CACCC.....  7994
SYK.KE.-.SYK173             -GAA-TTTA--CC-AC--C-A--G--GA---TGT-...------GCA--TC-G-TC--T---...---T-GA-TC-----A-----A-T--CT-T--G-----TG-TCA-CGTC-GTAT-A-T---CT..  7706
SYK.KE.KE51_AY523867        -GAATTTTG-ACCTACT-CTA--G--GAA--TGTG...------GC---T----TC--T---...---T-GA-TCGG---A----TA-T--CT-T--G--T--A-GTCA--G-C-CTAT-A-C---CT..  7402
COL.CM.-.CGU1               CT-ATCC-G-A-C-CC-AATTT--A-CAT-CATTA...A-T-CTTTG-----GGGA-G---C...---T-G--------GAGGA--ACTACTTAT-TC--A--AG-TAT---A---TC----AATGTAAA  7142
```

```
                 Rev responsive element secondary structure in RNA, see Charpentier, J Mol Evol 266:950-956(1997)
                 see HIV-1 complete genome annotation in compendium for stem-loop positions
                                                     gp120 \/ gp41
H1B.FR.83.HXB2             ...................AGAGAAAAAAGAGCA...GTGGGA...ATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGCCAGA 7850
Env                        __.__.__.__.__.__.__R__E__K__R__A__.__V__G__.__I__G__A__L__F__L__G__F__L__G__A__A__G__S__T__M__G__A__A__S__M__T__L__T__V__Q__A__R_ gp41

H1D.CD.84.84ZR085          ...................---------------...A-A---...C--------A----------------------------------G-----------G--------------------------- 7371
H1F1.BE.93.VI850           ...................---------------...-C----...C----------------------TC-T--G-ACAGCA-GGAGCA---------G--G--G--A--------------------- 7127
H1G.SE.93.SE6165           ...................---------------...--T---...C-G------G-C--------------A--------------------------G--G-----A----------------T---- 7268
H1H.CF.90.056              ...................---------------...------...--G-------CT-----------------------------------------G--G-----A--------------------G 7169
H1J.SE.93.SE7887           ...................---------------...------...---------G--------------------A----------T--------------G-----A----------------T---- 7157
H1K.CM.96.MP535            ...................---------------...--A---...T--------G-----T----A--------------------------------G--G-----A--------------------- 7010
H101_AE.TH.90.CM240        ...................-----------G---...------...---------A--A--T----------A--------------------------G--G-----A--------------------- 7403
H102_AG.NG.-.IBNG          ...................--------------G...--T---...C-G------G-C--------------A--------------------------GCGG-----A--------------------- 7356
H103_AB.RU.97.KAL153_2     ...................-----G---------...------...---------G-------------------------------------------G--G-----A--------------------- 7027
H104_cpx.CY.94.CY032       ...................---------T-----...------...-----G--CA----------------------------------------------G--------------------------- 7220
H1N.CM.95.YBF30            ...................---------------GCCT-T---...C----T--GC----T--------TC-T-----------G-----------------G-----A------------------C-G 7398
H1O.BE.87.ANT70            ...................---------------...--A---...T-G---ATGC-A---T-G---G-TC-AA-T--------T-----------------GG---CA------G--------A--CAC 7904
H1O.CM.91.MVP5180          TCACACC......CCTCAC--G------------...--A---...T-G---ATGC-A---T-G---G-GC-AA-T--------T-----------------GG---CAG------------G-A--CAC 7921
CPZ.CM.04.EK505            ...................---C-----------...-CCTTTGGAC----T--GC----T--------TC-T-----------G-----------------G-----A------------------C-G 7333
CPZ.CM.04.LB7              ...................---------------......-C-GGGC-G----TG--T--------------------------G-----------------A---T-A--------------------- 7361
CPZ.CM.04.MB66             ...................--G------------...-CC-T-GGAT-G------------------------------------------------------G---C----------------------- 7474
CPZ.MT145                  ACAACACCAT.........--------G-----C......CTTGGAC-------AC----T---------C-------------------------------A---G-CGT------------A---C-G 7359
CPZ.CD.-.ANT               ACACTCC............---C---------GCATT..---....-----GCTG--C---T-G--TC-TC-CA-T-----T--C--T--A-----------G-----AG---------C----A----G 7282
CPZ.CM.-.CAM3              ...................---------------GCCT-T---...C-G--T--GC----T-----------A-----------------------------A---G--GT----------------C-G 7263
CPZ.CM.98.CAM5             ...................-----G---------GCCT-T---...C-------AC-------C-----TC-A-----------------------------A---G--GT----------------C-G 7569
CPZ.GAB_X52154             ACAG.........AAAGAC---C-G---------GCCT-C--T...C-G-----AC--------------C----T--------------------------GG--G-A------------------C-- 7887
CPZ.GAB2                   AGCA.........CTTGAT---TCC---------GCCT-T---...C-G--G--GC----T-----A---C-T-----------------------------A--TG---T------------A--TC-- 7234
CPZ.TZ.-.TAN1              CAATCAT......ACT...--GTC--G-----ATGTG..---C...------CTG-----T-----A--TC-TA-T-----------T--A-----------G-----AG---------C---------- 7476
CPZ.US.85.CPZUS            ...................---------------GCCT-C---...C----T--AC-------G-----TC-T-----------------------------A---G-AGT----------------C-- 7855
GSN.99CM166_AF468659       ................GCTGTG--G--------C...C-TTC-...T-G---ATA-CC---T-GAC---TC-CA-T--G-----T-C---G-----------GG---CAG---------A----T--C-G 7787
GSN.99CM71_AF468658        GGTG............CATGTGA----G--G--C...A-TTC-...T-G--CATA-CC---T-GAC---TC-CA-T--G-----TGC---G-----------GG---CAG-------------AT--C-G 7799
H2A.ALI_AF082339           T............CCAAGG--GA-T-------GTGTGT-C...GTGC----G......---T-A--T--TC-C-CGA-------TTCTG-A------A-G--AG-TT-A------T-T-CT---T-TC-G 8307
H2A.BEN_M30502             T............CCAGTG--GA-C-------GTGTGT-C...GTGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------GCGG--CC--------T-A-CC---T--C-G 8323
H2A.ST_M31113              T............CCAGTG--GA-T-------GTGTAT-C...GTGC----G......---T-A--T--TC-CACGA--------GCTG-A--------G--G--CT--------T---CT---T-TC-G 7753
H2B.D205_X61240            A............AAACCG--GA-T-------GTGTAA--...GTGC-----......---T-G--A---C-T-C-ATG-----TTCTG-A---------A-G--TT--------T-A-CT---T--C-G 8299
H2B.EHO_U27200             G............ACACCG--GA-T-------GTGTAC-T...GTGC----G......--TT-G--A---C-T-CGA-G-----TTCTG-A--------G--G--CT-----T--T---CT---T--C-G 8285
H2G.ABT96_AF208027         G............ACACCG-A-A-T-------GTGTAT-T...GTGC----G......--TT-G--A--TC-T-C-ATG-----TTCTG-A------A----G--GT--------T---CT---T--C-G 7684
H2U.FR.96.12034            G............ACACCG--GA-T-----G-G-GTAT-T...GTGC-T--G......---T-G--T--TC-C-CCA-G-----TTCTG-A-----T--G--G--GT--------T-A-CT---T--C-G 7807

MAC.239_M33262             C............ACCTCA---A-T-------GGGTCT-T...GTGC----G......---T-G--T--TC-C-C-A-G-----TTCTG-A--------G--G--GT----------C-CT---T--C-- 8268
Env                        __.__.__.__.__T__S__R__N__K__R__G__V__F__.__V__L__G__.__.__F__L__G__F__L__A__T__A__G__S__A__M__G__A__A__S__L__T__L__T__A__Q__S__R  gp41

SMM.SL.92.SL92B            T............ACCTCA-AGA-T-------G-GTCT-T...GTGC----G......---T-G--T--TC-C-CGA-G-----TTCTG-A-----------G--GC-A--------T-CT---T--C-G 7716
SMM.US.x.PGM53             C...............TCA---A-T--G----GGGTCT-T...GTGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------G--G--CG----T-G-T-C-CT---T--C-G 8214
MNE.US.-.MNE027            C............ACCCCA---A-T-------GGGTCT-T...GTGC----G......---T-A--T--TC-C-C-A-G-----TTCTG-A--------G--G--GT----------C-CT---T--C-G 7745
RCM.NG.x.NG411             ...................-A-C----G----T-CCTT--...GTGC-G--T......--T--A--T---C-CTCG--------T-CTG-A-----------GG-G-CAG-C-----A--C---T--C-G 7581
RCM.GAB1_AF382829          ...................-A-C-G-----G...GTACCTTTGGTGC-G--T......--T--A--T---C-CTC---------T-CTG-A-----------GG-G-CAG-C-----T--C---T--C-G 7537
DRL.FAO_AY159321           A...............CAC--GC-G--G----GTGCA...-TCGTGC-T--G......A-------CC-----A-CTTG-----TTC-G-A---T--T---TGG-G-C-G-CA----T--C---T--CAG 7691
MND-1.F17                  T.........GAACATCAT-A---G--G----G-GCA--A...ATAC-T--T......A-------TC-GC-CTCGCTG------TC-G-G------T---TG--GG--G-A-----T--C--AT-TCAG 7457
MND-2.5440_AY159322        T...............ACA--GC---------G-GCG--C...TTAC-T---......A-G--C--CC--C-A--TTTG-----TTC-G-G------T---TGG-GG--G-C-----T--C---T--CAG 7633
MND-2.14CG_AF328295        T...............GCA-AGC-G--G----G-GCA--C...TTGC-T---......A-G--C--CC-------CTTG-----TTC-G-G------T---TGG-GG--G-C-----T--C---T--CAG 7929
MND-2.MND2_AF367411        A...............GCT--GC-G--G--G-G-GCG--C...TTGC-T--G......A-GT-C--CC--C-A--TTTG-----TTC---G------T---TGG-GG--G-C-----T--C---T-TCAG 8000
DEB.CM.CM40                G.........GAGTCGGTA---C-C--G--G---GCA-GAATTGCTT-T--T......C-GG-A-CA--TC-CA-TA-------TGCTG-------------G----CAG--------------T-TC-C 7590
DEB.CM.CM5                 G.........GAGTCTGCA---C-------G--GGCG-GAATTGCTT-T--T......C-AG-G-CT--TC--A-TA-------TGCTG-------------G----CAG--------------T--C-C 7511
MON.CM.L1                  ................CAC--G-----G---......-CAATTACTT-G--TATGGCC--T--G--C---C-CA-CA-G------G----A-----------GG-G-CAG-C------------T-TC-G 7813
MON.NG1                    ................ATG-------G-C--......-CTATCACAT-G---ATGGCC--T--G--C--TC-CA-CA-G------G----G-----------GG-G-CAG-C------------T--C-G 6438
TAL.CM.8023                .............GCAACA-A---G--------CGCGCCTCTGGCTT-G---......---A-A--A--TC-CA-CA-------C-CTG-C----------TGG-G-CAG-C--------G---T-TC-T 7230
TAL.CM.266                 .............GCTAAC-AGC-----C----GGCCCC-CT-GCCT----C......---A-A--A--TC-CA-CA-------C-CTG-C----------TGG-G-CAG-C--------G---T-TC-C 7680
MUS.CM.1085                ................CATGTGC-TGC----......-CCATCTCAT-G--CATAACC--TT-A--A--TC-CA-C---------G----A-----------GG-G-CAG-C------------T-TC-G 7784
SAB.SN.-.SAB1C             .............CCCGAA---C-------G...GTGCCATTCGTGC----G......-----A--T--------------T---GCTG-A-----T-----GG-G-CAG-C--------C---T-TCAG 8151
GRV.ET.x.GRI_677           A..................-----------G-TGCCAT-C...GTGC----G......--T--A--C-----------T--T----CTG-A-----------GG---CA--------A--C---T-TC-G 7857
TAN.UG.x.TAN1              T..................C-------G--G-TGCCCT-C...GTGT-G--G......-----A--C--T-----G-----T---GCTG-A-----T---A-AG-G-CAG--T-------C---T--CAG 7955
VER.DE.-.AGM3              T............CATGAC---AC---GC--...GTCCC-TTCGTGC----G......-----A--C-----A-----T--T--G-CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG 7496
VER.KE.-.9063              A............CAAGAC---TC----C--...GTCCC-TTTGTGC----G......-----A--C-----A-----T--T----CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG 7998
VER.KE.-.AGM155            C............CAAGAG--GC-----C--...GTCCC-TTCGTGC----G......-----A--C-----------T--T--G-CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG 7978
VER.KE.-.TYO1              A............CATGAG--GC----G---...GTCCC-TTTGTGC----G......---T-A--C--------G--T--T----CTG-A-----A-----GG-G-GC-GC--------C---T--C-G 7467
SUN.GA.98.L14              ................AGG-AG-----G--G---GCACCT-TTGCCC-----......GCTT-G-CAC-TC-CA-T--T--G--T-CTG-A------TT--TA--G-C--TA--A--T--C------CA- 8173
LST.CD.88.447              .............AGTAGT-AGC-G--G------GCACC-CTGGCTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C--A---CA- 7030
LST.CD.88.485              .............AGTAGT-AGC-G--G------GCACC-CTGGCTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C------CA- 7027
LST.CD.88.524              .............AGGGGA--GC---C-------GCTCCACT-GCTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C------CA- 7033
LST.KE.-.lho7              .............AGGGGA--GC---CC------GCGCCTCTGGCTT-----......GC-T-G--AC-TC-CA-T--T-----C-CTG-A------TT--TA--G-C--TA--A--T--C------CA- 8105
SYK.KE.-.SYK173            .............AATACC--G--------G---GCACCATTGGCAC-T--G......-----G--A--AC-CTCG--T--G--C-CTG-A-----G-GC--TG-G-CAG----A--AC-C---T-TCAG 7817
SYK.KE.KE51_AY523867       ................ACG-A-C-G---------GCACCATTGGCAC-T--G......-----G--A--AC-CTCG-----G--T-CTG-A-----GAGC--TG-G-CAG----T--AC-C---T--CAG 7510
COL.CM.-.CGU1              TTGGCAT......CATGGG---C----G---......-GCATC...T-T-C-TTC-CCA--T-A-CCC-----A-T-GG--C--TGCTG-A------T----G--TG--G--------A-C-----TCA- 7257
```

```
H1B.FR.83.HXB2              CAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTG...AGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAA  7977
Env                         _Q__L__L__S__G__I__V__Q__Q__Q__N__N__L__L__.__R__A__I__E__A__Q__Q__H__L__L__Q__L__T__V__W__G__I__K__Q__L__Q__A__R__I__L__A__V__E__  gp41

H1D.CD.84.84ZR085           ---G-------------------A-----A------------...--------A-----------------------------G-----------T--A--------------------------A----  7498
H1F1.BE.93.VI850            --------------A--------A-----A------------...-----------A--------------------G-----------------T--A---------------G---------------  7254
H1G.SE.93.SE6165            --------------C--------A-----AGG---C------...--------A--------G--------------------------------T--A---------------C----------A----  7395
H1H.CF.90.056               -----------------------------A-G----------...--A-----AC------G-------A-------G-----G-----------T--A---------------G----C----------  7296
H1J.SE.93.SE7887            -----G--------C--------A-----A-G----------...-A------A---------------------A-------T-----------T------------------G---------------  7284
H1K.CM.96.MP535             -----------------------A-----A-G----------...--A-----A--A--------A--------A--G-----G-----------T--A-------G------------------A----  7137
H101_AE.TH.90.CM240         -----------------------A-----A-G----------...--------A--------G--------------------------------T--A---------------G---------------  7530
H102_AG.NG.-.IBNG           -----------C--C--------A-------G----------...-AA-----A-----T-----A---------A-------G-----------T--A---------------G--------C------  7483
H103_AB.RU.97.KAL153_2      -----------------------A--------T---C-----...--------------------A--------------------------------------------G---G-------------G-  7154
H104_cpx.CY.94.CY032        -----------C--C----------------G----------...--------A-----T-----A---------AG------G-----------T--A---------------G--------C------  7347
H1N.CM.95.YBF30             AC------A-----G--------A--------T-T-C--T--...--A--A--A-----------A---T-----------T--A----------T--A-------------A-G----T---A-A----  7525
H1O.BE.87.ANT70             ACT--GC--AAG-----------A------G----CC----A...--A--A--AC----C--G-----AT--C--AGG--AT-TR-A-----T----GA--A----GA--TC-CC-G--A--CT-A----  8031
H1O.CM.91.MVP5180           AGTG--C--AAG-----------A------G----CC-----...--A--G--AC----C--G--A--CT--C--AGGT-AT-T--A-----T--T-GA--A----GA--TC-CC-G-AA--CT-A----  8048
CPZ.CM.04.EK505             A-------A-----G-----------------T---C-----...--A--A--A-----------A--CT--C--------T----T--------T--A--------A------G----T---A-A----  7460
CPZ.CM.04.LB7               ---------A----C--------A-------GT---------...--A-----A--------G--A--CT--C--------T----------------A--------A------G----T---A-C----  7488
CPZ.CM.04.MB66              ---C-GC------------------------G----C-----...--A--C--A---------------------------T-------------T--A-----T------C-GG-T--T--------G-  7601
CPZ.MT145                   -----G--AA-G--C-----------------T---C----A...--A--A--A--A--T-----A-----------G--TT-T--------A-----A-----------G---G-T--T----------  7486
CPZ.CD.-.ANT                A-T--GY-CCA------T--A--A------GC----C-----...CAA--C--A---A-A--G--A------C-A--G---T-G--------AG-A--A--A--------------G--T--A--C--G-  7409
CPZ.CM.-.CAM3               -----------A--G--------A-----------CC-----...C-A--A--A---------------T-----------T----T--------T--A--------A------G-T--T-----A----  7390
CPZ.CM.98.CAM5              --C--G-----A--G--------A-----------CC-----...C-A--A--A-----------A---T-----------T----T-----------A--------A------G-T--T-----A----  7696
CPZ.GAB_X52154              --G--GC----G--------------A-----T---C--T--...-AA--A--A-----T--------C--A--------AT--A-------AG-A--A--A--A--A--C---T-G--T-----A----  8014
CPZ.GAB2                    A-CC-GC-T--G--A--------------A--T--C------...C-A--C--A--A--T--G--A--C---C----G--AT-------------A--A-----------G--GG----T----------  7361
CPZ.TZ.-.TAN1               GG----C-C--------T--A--------AC-A--CC----T...CA---C--A--A--------A--CT-------G---T-T--A--------T--------------C-----G--T--A--A--G-  7603
CPZ.US.85.CPZUS             -----------A--G--------A--------T---C----A...--A--A--A-----C---------T-A--------AT----T--------T--A-----T--A--C--G--A--T----------  7982
GSN.99CM166_AF468659        TCT--G---G----G---A-----------G-G--CC-----...------G-A-----C------AGCT-A--A--G-CAT-T--A--------A--A-----A--AA-CC-CC-GTCAT-AC-A--G-  7914
GSN.99CM71_AF468658         TC---G---G----G---------------G-G--CC-----...------G-A-----C------AGCT-A--A--G--AT-T--A--------A--A-----A--A--CC-CC-GTCTT-AC-A--G-  7926
H2A.ALI_AF082339            ACT------G-C--G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA------G-AA------G---G--C--------A-CG--AA-T------------G--ACT---A-C--G-  8434
H2A.BEN_M30502              ACT---C--G-C--G-----------A---C-AC-GC--T--...GAC-TAG-CA--AGA-----AG-AA------G---G--C--------A-CG--AA-C------------G--ACT---A-C--G-  8450
H2A.ST_M31113               ACT------G-C--G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AA------G---G--C--------A-CA--AA-T------------G--ACT---A-C--G-  7880
H2B.D205_X61240             ACT---C--G-C--G---------------C-AC-GCCTG--...GAC-TGG-CA-AAGA------G-A-------GG--G--C--------G-CG--AA-C------------G--ACT--CA-C--G-  8426
H2B.EHO_U27200              ACT---C--G----G---------------C-AC-GC-CG--...GAC-TGG-CA-AAGA-----AG-A-------GG--G--C--------G-CG--AA-C------------G--ACT--CA-C--G-  8412
H2G.ABT96_AF208027          ACT--G---G----G-----------A---C-AC-GC--T--...GAT-TGG-CA--AGA------G-AA------G---G--C--------G-CG--AA-C------A-T---G--ACT--CA-C--G-  7811
H2U.FR.96.12034             ACT---C--G----G-----------A---C-AC-GC--T--...GAC-TGG-CA-AAGA-----AG-AA------GG--G--C--------A-CT--AA-C------A-T---G--ACT--CA-C--G-  7934

MAC.239_M33262              ACT------G----G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AT------G---G--C--------A-CA---A-C------A-T--GG--ACT--CA-C--G-  8395
Env                         _T__L__L__A__G__I__V__Q__Q__Q__Q__Q__L__L__.__D__V__V__K__R__Q__Q__E__L__L__R__L__T__V__W__G__T__K__N__L__Q__T__R__V__T__A__I__E_   gp41

SMM.SL.92.SL92B             ACT------G----G--------A--A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AT------G---G--C--------A-CT--AA-T--T---A-T---G--ACC--CA-C--G-  7843
SMM.US.x.PGM53              ACT--G---G----G-----------A---C-AC-GC--T--...GAT-TGG-CA-AAGA-----AG-AT------G-T-G--C--------A-CT---A-C------A-T---G--ACT---A-C--G-  8341
MNE.US.-.MNE027             ACT------G----G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AT------G---G--C--------A-CA---A-C------A-T---G--ACT--CA-C--G-  7872
RCM.NG.x.NG411              --T--GC-TG----G---T-------------G--G------...GACATAG----ACA---G---G-A---C-AA-G-----C--------G-CT--AA-C--------GC-TG--ACT---A-T--G-  7708
RCM.GAB1_AF382829           --T--GC-TG-A--G---T----------A--G--CC-----...GACATAG--A--CG---G---A-----C-AA-G-----C--------A-CT--AA-T--------GC-TG--ACT---A-T--G-  7664
DRL.FAO_AY159321            GCT--GC--A----G--G---G-------A--GC-AC-----...C-CCTGG-G--ACA-------G-GT-AC-CA-GT-A--TA--------G-A---A-C--------CC-CC--ACA---C-T--GG  7818
MND-1.F17                   TCT--GG--A----G------G-A--A--A--AC-G---T--...-A-CTC--A---CAA--GTCTG-A--C--AA-------CA-A-----AG-A---A-TT-A---A-TC-CC-GACCAG-T----G-  7584
MND-2.5440_AY159322         ACT---C--AA---G--T---G----------GGT-C-----...--CCTG--A--TCA---CTCCG-GT-A--AA----A--TA-------TG-A---A-T--T----TCC-CC--ACA--CT----GG  7760
MND-2.14CG_AF328295         GCT--G---AA---G--T---G----------G-T-C-----...--CCTG--A--TCA---CTCCG-GT-A--AA----A--TA-------TG-A--AA-T--T-----CC-CC--ACA--CT----GG  8056
MND-2.MND2_AF367411         GCT--G---AA---G--T---G----------GGT-C-----...--CCTG--A--TCA---CTCCG-GT-A--AA----A--TA-------TG-A---A-T--T-----CC-CC-TACA--CT----GG  8127
DEB.CM.CM40                 TCT--GC--AG---G-----------A---C-AG-GC--T--...-A----G-G-----T--CGG---G-----AAC-T-A----C------AG-G-GAA-C---A-CA-CC-CC-GACT---A-A--G-  7717
DEB.CM.CM5                  TCT--GC--AG---G-----------A---C-AG-GC--T--...-A----G-G-----T--CGG----------TC---A----CA-----AG-G-GAA-C---A-CA-CC-CC-GACT---A-A----  7638
MON.CM.L1                   TC----C--G----G---------------G-G--CC-----...-----CG-AACA--C------AGCT-AC-A--G--T--T--T-----AG----------A-----CC-CC-GAC------A--G-  7940
MON.NG1                     TCT--G---G----G---------------G-G--CC-----...-----CG-AACT--C--G---AGCT-A--A--G---T-T--------------A--AT-A-----CC-CC-GAC------A----  6565
TAL.CM.8023                 TCT--GC--AGC--G---------------G-GC-CC-----...-----A------CAT--G--A--C-----A---T-A-----G------------A-T---A-T--CC-CC-GACT--GC-T--G-  7357
TAL.CM.266                  TCT--GC--AG---G---------------G-AC-CC-----...-----A------CAT--G--A--C-----A---T-A-----G------------A-T---A-T--CC-CC-GACC--GC-T--G-  7807
MUS.CM.1085                 TCT--G---G-C--G---------------GCA---C-----...--A--CG-A--A--C-----A-----A--A--GT-GT-T--T--------T--A-----A-----TC-CC-GACA--GC-T--G-  7911
SAB.SN.-.SAB1C              --G---C-CG----G---C-------------A---C----T...GC----G-A---CA---G-----AA--C-AA-G-----TA-T-----TG-A--AA-C---A-T--CC-CG--ACT---C-T--G-  8278
GRV.ET.x.GRI_677            --T--GC-TG----G---T-------------G--C------...GC----G-G--ACA------A--GT-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-C--G-  7984
TAN.UG.x.TAN1               ------C-TG-A--G---T-------------G---C-----...GC---AG-----CA---G-----GA--C-AA-G-----CA-T-----TG-G--AA-C---A-T--CC-CG--ACT--CC-T----  8082
VER.DE.-.AGM3               --T---C-TG----G---C-------------G---C-----...GC----G-G-----T--------GA-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-  7623
VER.KE.-.9063               --T---C-TG----G---A-------------A---C-----...GC----G-G-----T--------AA-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-  8125
VER.KE.-.AGM155             --T---C-TG----G---T-------------G---C-----...GC----G-G-GA--T--------GA-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-  8105
VER.KE.-.TYO1               --T--GC-TG----G---C-------------G---C-----...GC----G-G-----T--------GA-----A-G--G--CA-T-----TG-T--AA-C---A-T--CC-CG--ACA--CC-T--G-  7594
SUN.GA.98.L14               GC-G-C---CAA--G---T-------------GC-G------...GT-CT-G-----AA---G--AG-G--AC-AAG------CA-A-----AG-G---A-T-----A---C-CC-GACA--CC-A--GG  8300
LST.CD.88.447               GT-G--A-ACAA--G---T-------------GC-G------...GT-CT-G-----AAA--G--AG-G--C--AAG------CA-------AG-G---A-T-----A---C-CC-GACA---A-T--GG  7157
LST.CD.88.485               GT-G--A-ACAA--G---T-------------GC-G------...GT-CT-G-----AAA--G--AG-G--C--AAG------CA-A-----AG-G---A-T-----A---C-CC-GACA---A-T--GG  7154
LST.CD.88.524               GT-G--A-ACAA--G---T-------------GC-G------...GT-CT-G-----AAA--G--AG-G---C-AAG------CA-------AG-G---A-T-----A---C-CC-GACA---A-T--GG  7160
LST.KE.-.lho7               GC-G-C---CAA--G---T-------------GC-G------...GT-CT-G-----AAA--G--AG-A-----AAG------CA-A-----AG-G---A-T-----A---C-CC-GACA--CC-T--GG  8232
SYK.KE.-.SYK173             ACT--G---G----G---------------C-A--G------...GA---AG-------A--G--A--C--C--AGGCT-A--T--------TG----AA-C--TA-C--CC-GC--ACA---T-A--G-  7944
SYK.KE.KE51_AY523867        ACT--G---G----G---------------C-A--G------...GA----G-A-----A--------C--C--AGGCT-A--G--------TG-G--AA-C--TA-C--CC--C--ACT---C-A--G-  7637
COL.CM.-.CGU1               TCCC-TAATGGAA-GGCTAGTGCCAGCAGC---CGAA--T--CTG-AACTGG-C---A----GTCTGCA---C-A--G-----G--------GG-T--AA-C-------T-C--G-TGCCA--A-A---G  7387
```

```
                                                         Rev responsive element end -|
H1B.FR.83.HXB2             GATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCT.............................  8078
Env                        R__Y__L__K__D__Q__Q__L__L__G__I__W__G__C__S__G__K__L__I__C__T__T__A__V__P__W__N__A__S__W__S__N__K__S__.__.__.__.__.__.__.__.__.__.__.  gp41

H1D.CD.84.84ZR085          -------G-----------------A--A---------------------A------------A---A--C-----CT---------------G----.............................  7599
H1F1.BE.93.VI850           -------------------------A-----C-----C-----------------C---------AA------C-----CT-------------------.............................  7355
H1G.SE.93.SE6165           -------------------------A-----------C-----------------C---------AA------C-----C-T------------------.............................  7496
H1H.CF.90.056              ---------G-------------------------C-----------------C---------AA-----------CT------------------A.............................  7397
H1J.SE.93.SE7887           -------------------A-----A-----------C-----------------C---------AA------C-----------------------------.............................  7385
H1K.CM.96.MP535            -------------------------A-----------C-----------------C---------AA------C-----CT---C-----------...........................TC  7237
H101_AE.TH.90.CM240        -----------------A--T----A--AC-------C-----------A-----C-----------------C-----CT-C-C----------G----.............................  7631
H102_AG.NG.-.IBNG          ---------GA--------------A--A--------C-----------------C---------A---A--C-----CT---C-----------A--.............................  7584
H103_AB.RU.97.KAL153_2     ------------------------------------------------------------------------------------A-----------------C--.............................  7255
H104_cpx.CY.94.CY032       -C-----------------------A--A--------C-----------------C---------AA---A--------CT------------------.............................  7448
H1N.CM.95.YBF30            -------T-G------G--AA----AA-TC-A-----C-----A------ACA--A---TAT--CA---------------AG-C------C--C--TA-CTCT.........................  7629
H1O.BE.87.ANT70            CC-TA---C--A----G--A-----AA-CC-A-----C--TAAA-----G--AG-C---TA---AT-A--AAAA------AGA-CA----TAGGA.............................  8126
H1O.CM.91.MVP5180          CCCTTA--C--A----G--A-G---AAACC-A-----C--TAAA--------A--C--TTA---AT-A--AAAA-----CA-ATCA---TCAGGA........................AGATATAA  8151
CPZ.CM.04.EK505            -------G-G---C--G--AA----A--CC-A-----C-----A------TCTG-G--TTA---CAA-------------A-C-C------C--C........................AATAACTC  7563
CPZ.CM.04.LB7              -------------------A-----A--A--A-----------A-----------C---------T-------A------AGA-C---------C...........................AAAAC  7588
CPZ.CM.04.MB66             -----------------------T-A--AC-A-----C------------------------------T-------------CA---CC----C----...........................AAATC  7701
CPZ.MT145                  -G-------G------G--A-----A--TC-A-----C--TA-A------ACT-----TC----------A-GC-----CAAG-CA---G-A----TT..........................AGTGA  7589
CPZ.CD.-.ANT               AG-------GA--------A--AT--A-CC-C--------TG---AC--GG-G-CC--TCA----A-G------------AA-TCC---GTA--CTTCA-GCAAACATGT...GCAAAGAACAGCAGTGA  7536
CPZ.CM.-.CAM3              -------C-GA-----G--AA----A--CC-A-----C-----A------GC------TTAT---A----------------AA--C--------GC---T..........................ACATC  7493
CPZ.CM.98.CAM5             ----------------G---A----AA-TC-A-----C-----A------GC---C--TTAT--CA-G-----------------C--------GC---C..........................ACATC  7799
CPZ.GAB_X52154             -G-----GC-------G---A-T-----CC-G-----C------------GGCTG-C--TTAT--CA-G------------CAACTC----CC-GGG...AGC.....................AATTC  8117
CPZ.GAB2                   ----T-----------G--AG----A-CAC-A-----C------------ACTG-G---TA-T--A----A--C------A--TCC----AC-G---CAAA...........................TC  7464
CPZ.TZ.-.TAN1              A----A---GA--C-----------AA-CC-C-----A--TG--AAC---T-GG-G--TCA--G-AG------A-----CCTC-CC---GC-G-AG-T---ACAAAGTGCAATCACAGTGATGCAAAGTA  7733
CPZ.US.85.CPZUS            -------T--------G---A----A--TC-G-----C-----A------ACA-----TTAT--CA----------------A--CC---------C--C...........................CTCTC  8085
GSN.99CM166_AF468659       A---TT-G-G-------ACCA-T--CCAAGCA-----C--TG-GAAC-GG-CA--C--TCA---CATA--A---------CA-CTCA---GC----GGC--G.............................  8015
GSN.99CM71_AF468658        A------G-G-------ACAA-T--CCAAGC------C--TG-GAATC-G-CA--C--TCA---CATA--A-----------ACTCA---GC---G--C--G.............................  8027
H2A.ALI_AF082339           A----T-------C--GGC--GG--AAATTCA-----A--TG-ATTT-G--AAG-C---CA----A----A--A---GTAAA--AC-CC.............................  8523
H2A.BEN_M30502             AG---------C----GGCA-AG--AAATTCA-----A--TG-GTTT-G--AGG-C---CA----A----A--G---GTAAA-GAC-CT.............................  8539
H2A.ST_M31113              A----T-------C--GGC--AA--AAATTCA-----A--TG-GTTT-G--AAG-C---CA----A----A--A---GTAAA-GACACC.............................  7969
H2B.D205_X61240            A---------------GGCA-AA--AAATTCA-----A--TG--TTC-G--AGG-C--TCA---GA----A--A---CCAAA-GAAACC.............................  8515
H2B.EHO_U27200             A---T--C--A--C--GGCA-AA--AAATTCA-----A--TG--TTC-G--AGG-C--TCA---GA----A--A---GTAAA-GAA-CC.............................  8501
H2G.ABT96_AF208027         A---T--------C--GGCA-GG--AAATTCA-----A--TG-GTTT-G--AGG-C---CA----A----A------G----CCT-G-TGC-------A-A.............................  7912
H2U.FR.96.12034            A---------------GGCAAGT--AAATGCA-----A--TG-CTTT-G--AGG-C---CAT---A-------A----TAAACGACACA.............................  8023

MAC.239_M33262             AG---T-------C--GGC--AG---AATGC------A--TG-GTTT-G--AAG-C---CA----A----A--A---CCAAA-GCAA-T.............................  8484
Env                        K__Y__L__K__D__Q__A__Q__L__N__A__W__G__C__A__F__R__Q__V__C__H__T__T__V__P__W__P__N__A__S__.__.__.__.__.__.__.__.__.__.__.__.__.__.  gp41

SMM.SL.92.SL92B            A---------------GGCA-AG--AAATTCA-----A--TG-ATTT-GG-AAG-C---CA----A----A-TA---CCAAA-GACA-T.............................  7932
SMM.US.x.PGM53             AG----G---------GGC--AG--AAATTCA-----A--TG--TTT-GG-AGG-C---CA----A----A--A---CCAAA-GCC-CA.............................  8430
MNE.US.-.MNE027            AG---T-------C--GGC--AG--AAATGC------A--TG-ATTT-G--AAG-C---CAT---A----A--A---CCAAA-GCAA-T.............................  7961
RCM.NG.x.NG411             A-------GC---C---TCT--AT-AAAT-CA-TT--C--TG-AT-G-G--AAG-G--TCA---AA-G---GAG----TATA----CAA.............................  7797
RCM.GAB1_AF382829          A-------GCA--C---TCT--AT--AAT-CA-TT--G--TG-CT-G-G--AAG-C---CAT--A-TG-----G----CATTC...........................AA  7749
DRL.FAO_AY159321           A----A--GGA--C---GCAA-GT--A-CC-C-----A-----CTTTGC--AAG-A---CAT--AAA----GTA---CCCAA-GAG-CA.............................  7907
MND-1.F17                  AT--TA-C-----C---GCTT-G---TCTCAA-----G--T--AT-GGC--AGG-G--TCAT---T----AGAG----C-AA--CAA-C.............................  7673
MND-2.5440_AY159322        A----G--GC---C---TCAAGG--CTC-G-A-----A--T--ATTCTCC-AAG----TCAT---AG---AAAA---CCCAA--A-A-T.............................  7849
MND-2.14CG_AF328295        AC--TG--GC---C---TCAAGA--A-CAG-A-----A-----ATTCTCC-AAG-G--TCAT---AA---T--G---CCAAA-GAAA-T.............................  8145
MND-2.MND2_AF367411        AC--TG--GCT------GCAAGA--ATCC--G-----C-----ATTCGCC-AGG-A---CA---ACA---A--A---CCAAA-GAC-CT.............................  8216
DEB.CM.CM40                AG---T----A--C---GCAAAAT-AAATGAA-----C---G-CTTC----AG--C--TCAT--CA-G--C---------AACTC-CT-.............................  7806
DEB.CM.CM5                 AG--------A-----GAGCAAGT-AAATGAA-----A--TG-CTTT----AA--C--TCA----A-G--C--------CCA--CA---.............................  7727
MON.CM.L1                  AG-T-A-------C---ACAT-G--AAATGCA-----A--TG-CAAT---GC-G-G--TCA---AA-A-----G-----CAA-TC-.....................TGGGCAAAAGG  8037
MON.NG1                    AG-T-A-------C---ACAT-GT-AAATTCA-----A--TG-CAATCGCGCAG-R---CA---ACAA----TA-----CAA--CG.....................TGGGCTAAAGG  6662
TAL.CM.8023                A----T--G----C---GCCAGGT-AAATTCA-----A--TG--T-G----AG--C---TA---CT----A--G------AAG-CA.....................TGGACAAACTC  7454
TAL.CM.266                 A----T--G----C--GGCTAAG--AAATTCA-----G---G-AT-G----AG--C---TA---CT-G--A--G-----CAAA-CA.....................TGGTCGAATTA  7904
MUS.CM.1085                AG-T-A-------C---GCC--T---AATC-C-----A--TG-AAAC-GG-AG-----TCA---AGA-----A-------A-TCA.....................TGGGCAAATCA  8008
SAB.SN.-.SAB1C             A-------G-------GGCA-GGT--AAC--------A--TG--TTC-GG-AGG-G--TCA---GA-G--C-T------GTA--A-AAC.............................  8367
GRV.ET.x.GRI_677           AG------G-------GGCA-GG--AAATTCA-----A--TG-GT-G----AAG-A--TCA---CA-A-----A-----GTA--A-AAC.............................  8073
TAN.UG.x.TAN1              A-------G-------GACA-GG--AAATT-G-----A--TG-ATTC----AGG-G--TCA----A-G-----G----C-TTC-ACAAT.............................  8171
VER.DE.-.AGM3              AG------G----C--GGC--GGT--AATGC------G---G-AT-G--G-AAG-C--TCAT--AA-C--A--G---C-GTGG-A-AAT.............................  7712
VER.KE.-.9063              AG------G-------GGCA-GAT--AACG-C-----G---G-AT-G----AGG-C--TCAT--CA-A--A--A---C-ATGGCAAAAC.............................  8214
VER.KE.-.AGM155            AG-----GGC------GGCA-GGT-AAACGC------G---G-GT-G----AAG-A--TCAT--AA-A--A--C----CGTGG-A-AAT.............................  8194
VER.KE.-.TYO1              AG------G-------GGCA-GA--AAACTCC-----G---G-AT-G----AAG-A--TCAT--CA-A---GAG---CCCTGG-CAAAT.............................  7683
SUN.GA.98.L14              AG--TG-GC-A--C---AGC--TT-A-CTTC------G--TCAAT-G----AAG-A---CAT---AA------C------TA--A-.............................  8386
LST.CD.88.447              AG-----------C---ACT--G--T-CTTCC-----G--TCAAT-G----AAG-A--TCAT---AA---AGAA-----CTA--A-.............................  7243
LST.CD.88.485              AG-----------C---GCT--G--T-CTTCC-----G--TCAAT-G----AA--A--TCAT---AA----GAA------TA--A-.............................  7240
LST.CD.88.524              AG---T-G--A--C---GCA--A--T-CCTC------G--TCAGT-G----AAG-G--TCAT---AA---A--A------TA--A-.............................  7246
LST.KE.-.lho7              AG--TG-G--AC-----GC------T-CCTC------G--TCAAT-G----AGG----TCAT---AA----GAG----CATA--A-.............................  8318
SYK.KE.-.SYK173            CG--T--C-GA--C--GGCCA--T--TCA-A------A---G-ATTC--G-AG--C---CAT--A--A---A-G---G-GAAAGCA--TG-A--C--C--CAAC.....................TTCTG  8053
SYK.KE.KE51_AY523867       CG--T--C-GA--C---GCCA-TA--TCA-A------A---G-ATTT--G-AA-----TCA---A--A--AA-G---C-GCAAGCA.....................TGTGGCAATAACTCTCGTTG  7743
COL.CM.-.CGU1              -T--TT-GG----A---GCAAAG--T-CC-GCATT--G--TG-AAAC-TG-AA-----T-GA---AT------C-----CAAG-CC---.............................GG  7478
```

```
H1B.FR.83.HXB2             .CTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTG  8207
Env                        __L__E__Q__I__W__N__H__T__T__W__M__E__W__D__R__E__I__N__N__Y__T__S__L__I__H__S__L__I__E__E__S__Q__N__Q__Q__E__K__N__E__Q__E__L__L_  gp41

H1D.CD.84.84ZR085          .G-A--GT-T------GGGA---T----------C-------A---------G---------GG--------T--AA---G----------------T-------------------A----------A  7728
H1F1.BE.93.VI850           .-A----G----------A---T-----------------A-AG--G---G-------T---A-A----T--AAA--G--------G--A-----------G----------------------------  7484
H1G.SE.93.SE6165           .TATA-TG-A------G--A---T---T-----A--A-----A--G-----C-----------TA-CA----T--AG-C--C-----------A--G--------G-----------------C------  7625
H1H.CF.90.056              .-A-AGTG-A--C---G-CA---T---T--------A-----T-A-C------G----------GAGGA----T--AGG--GC------TC-------C------G---------------G--C-----A  7526
H1J.SE.93.SE7887           .TAT--GG-C------G-GA---T---------AC-A-----A--G------------------G-AA-----T--AGTC------------G-A-----------G-----T------A----C------  7514
H1K.CM.96.MP535            TTG----G----------CA---T-----------------A-A---G---GG-------T--GA-AC----T-TAAG----------------A----C------G-----------------T------  7367
H101_AE.TH.90.CM240        .T-T---G----------CA---T---A-----A--A-----G---------G-----------A-CA----T-TGAGA--C--AC----------G-----A--G--C-G-------A-G--T--G--A  7760
H102_AG.NG.-.IBNG          .T-TA-TG-C--A---G--A---T---------AC-A-----G-A--------G----------GA-A-----T-TAATC-----------------G----G---G-----------------C--G---  7713
H103_AB.RU.97.KAL153_2     .--A--TG-------A---A---T-----------------A-----------T---------G-T------T--AAT-----A-----------G-----A----------------------A-----  7384
H104_cpx.CY.94.CY032       .TATA-TG-T--A---G-CA-T-T-------T--C-A-----T-A-------------------CAAA-----T-TGGG---C-----------A--G--------G------------G-----C------  7577
H1N.CM.95.YBF30            .TAT--TACA--C-------A-TTTA------CAAC-A-----TGAGA--G-A-GA--C--TT--G-TG-C--TTTTGGAC-T--A---C-GG-A---G-A--A--GA-C-CA-----GA--TC-C-C---  7758
H1O.BE.87.ANT70            .AAC---AGC------G-CAC-TTA--A---CA---A-----TC-GC-G--A-G---CATA-GCTC-ACC---T-TGAGGA---AC--A-GG-A---GTA-----G---C-A-----GA--A-G--GC--  8255
H1O.CM.91.MVP5180          TGAT--CAGT------G-CA--CTT--A---CA-C-A------CA-C-C--A------GTA-GCTC-A-T---T-TGATGA---AC---C-G-A---G----A--G----------T-A---C---G---  8281
CPZ.CM.04.EK505            CTAT--CACA--A---GGAA-T-T---A---CAAA-T-----TGAGC--G-A-GA-----TT--G-TG-C--TTTTGGGC-GC-A--GC--G-A---G-G--A---AGC-TA------A-GTC-C-C---  7693
CPZ.CM.04.LB7              ATATA-TG-A------G-CA---T---A--------A--------G---G-A-GA---------GAAA-T--TT-TGGG----A--GC--G-A---G--------G-----T-----GA--A--C-T--A  7718
CPZ.CM.04.MB66             TTAT--TG-T------T--A---T---A------C-A-------A----G---G----------GATG-T--CT-TAATC--T-A---A-GG-A----C------G-----T------A-----C-CC-A  7831
CPZ.MT145                  CTATC-GGTT---------A-TTAC--A---CAAC-A-----T------G-A--T---------G-AC-C--CT--A-A--GC-A--G--GG-TA-T-CA-----G--G--A------A-G--G------  7719
CPZ.CD.-.ANT               TA-AC--TGT------G-AA-T-T---A---CAA--A---------TT-G-AC-G----CA---G-ACAG---T-TAATA-CT-AC--AT-G-A--TG-G--A-----G-GA---A--A-G-------AT  7666
CPZ.CM.-.CAM3              CT-T--TG----C-------A-TTTA--A---CAA--C-------A-AGGG-A--A-----TT--G-TG-C--TTTTAG-C-T--A---C--G-A---G-G--A---A-C-CA-----GA--AGTC-----  7623
CPZ.CM.98.CAM5             CTAT--TG----C-------A-TTTA--A---CAA--T-----T-A-A-GG-A--A------T--G-AG-C--TTTTAG-C-T--A--GC--G-A---G-G--A--GA-C-CA------A----TC-T---  7929
CPZ.GAB_X52154             CACA--TG-C------GGGA-TCTA--A---CA-C-A-----T-A-TT-G-GTCT--C------G-GAA---TTTTGGTC-GT-A--G---G-A---TCA--A-----G--A------AGG--CC-----  8247
CPZ.GAB2                   TTAT--GG-C-------G-A-TCT-------CA-C-A-----T-AGTT-G-AG-A---C-T---G-AACC--TTTTCAA---T-AC-GAG-G-A--TG-G--A--GA-T-GC------A----GC-----  7594
CPZ.TZ.-.TAN1              CTAT--CTGT--A-----CA-TTT---T---CA---A-----TC--TT-G-AG-A--C-CT---G-AACC---T-----C-GT-A--GA--G-A----CA--A--G--G--A--CA-------G--G--A  7863
CPZ.US.85.CPZUS            CTAT--TGCT------GGCA-TCTA--T---CAA--A---------A--G-A-GA--C--TT--G-TACT--TTTTAGTC-T--A---C--G-A---G-A-----GA-T-CA------A--TC-C-C---  8215
GSN.99CM166_AF468659       ....CTT-CAGAC---G-GA---T-------CAAA-----AGT-TGTT-G--G-G---G-----TATAC------GCAG--GT-A--GC--G-AA-CC-G--A----C-TCA--CCTGA-T--GC-GA--  8141
GSN.99CM71_AF468658        ....ACG-CAGA----G-G----T-------CA-------AGT-AGTT----G-A---G-----TATACC-----GCAA---T-A---A-TG-AA-CC-T--A---TC------CATGA-T--C-----A  8153
H2A.ALI_AF082339           .T-AA---CTGA----G-CA---T---G---CAA--------ACA-C--G-CCGTT-CCTAGAGGCAAAT--CAGTGAACAGT-A---CGGG-A----TT----------------ACGT-T---C--CAA  8652
H2A.BEN_M30502             .T-ATCG-CTGAC-----AA-T-T---A---CA---------G-A-C--G-CCG-T-CCTAGAGGCAAAT--CAGTCAAAGTT-A------G-C----TT--A-------------ATGT-T------CAA  8668
H2A.ST_M31113              .T-AACG-CTGA------CA---T---A---CA---A-----GCA-CG---CCG---CCTAGAGGCAAAT--CAGTGAAAGTT-A---C-GG-A----T----------------CATGT-T---C--CAA  8098
H2B.D205_X61240            .--CAC--CA-A--------A---T---A---CAAC-------A-AGC--G-CC-TTTC-TGGATGCAAAT---ACGG-TC-GT-A-----GG-T--G-TA--A--G--------CATGT-T------CA-  8644
H2B.EHO_U27200             .--TA-G-CAGAC-------A---T---A---CAAC-A-----G--GC--G-CCG-TTC-TGGATGCAAAT---ACAAAA---C-A-----GG-A--G-TA--A-----------CATGT-T--G---CA-  8630
H2G.ABT96_AF208027         .T----G-CACAA-------A-T-T---A---CAA--A-----A-AGC-G--C---TT--TGGAGGA-AAC--CACACGGC--T-A------G-A----TT-------------A---ATGT-T--G---CA-  8041
H2U.FR.96.12034            .T-AAC--CT-A----G--A-T-T---G---CAA--A-----GGA-A-GG-G---T-C-TGGA-GAAAAT--CACACAGC--T-A----C-G-T----TA--A-----G--A--CATGT-T------CA-  8152

MAC.239_M33262             .--AAC--CA-AG-----CA-TGA---T---CAA--------GC--A-GG--G--TTC-TGGA-GAAAAT---ACAG--C-CC-A--G--GG-A----TT--A-----G-----CATGT-T------CAA  8613
Env                        __L__T__P__K__W__N__N__E__T__W__Q__E__W__E__R__K__V__D__F__L__E__E__N__I__T__A__L__L__E__E__A__Q__I__Q__Q__E__K__N__M__Y__E__L__Q  gp41

SMM.SL.92.SL92B            .--T-TC-CAGAC-----CA-T-T---A---CAA--A-----G-A-A-GG--G-ATTCCTAGAGGCAAAT---ACTCAAA-GT-G--G---G-A-G-CTA--------G--A--CATGT-T------CA-  8061
SMM.US.x.PGM53             .T---TG-CT-A------CA-T-T---T---CAA--A-----A---C-GG--G--G-CCTAGAGGCAAAT---ACTCAGGC-T-G------G-A----TT----------------ATGT-T-----GCAA  8559
MNE.US.-.MNE027            .--AAC--CA-A------CA-TGA---T---CAA--------GC--A-GG--G--TTC-TGGAGGAAAAT---ACGG--C-TC-A-----GG-A----TT--A-----------CATGT-T------CAA  8090
RCM.NG.x.NG411             ....AC--CAGAG-----CA-ACA---A---C-T--------A---A-T---TCA-GA-TAGA-G-AAAC--CTCAGTAGCTT-AC-G--TG-A---G-A-----T---CGT----TG--T--C---GAA  7923
RCM.GAB1_AF382829          CAAAACG-CTGAG---C-GA-AGAAT-A---T--C-------A---A-T--CTCTT---TAGAGGCTAAC--TACAATAGC-T-AC-G--GG-C--GG-T--A--T--G--A----TG--T-----GGA-  7879
DRL.FAO_AY159321           .G--ACT-CA-A-----CCTCAGA---A-------------C-A-A-AG-G-AG-T-G-ATAT---A-AAC--TACTCTTGATT-AC-GA-GG-AT-TG-A--AG--C--------ATCTTT--G---CAA  8036
MND-1.F17                  .A-CACT-CA-A-----CATCAGAA--T----A---A-----G-C-AG--C-G-TT--CTGCA-CAAAAC--TACAGAAA-GT-AA--C-GG-AT-TG-T-GAG-GC---GA--CAC-T-T------CA-  7802
MND-2.5440_AY159322        .A-A-TG-CC-A-----CCTCAGA---A---C----A---------AG-G-A----GCATAGTG-CAAAT--GAC-ATAGACT-AC-GAGGG-AT-TG-ATTAG--C-G-G---CATCTTT--G--GCAA  7978
MND-2.14CG_AF328295        .A-AACT-CT-A-----CCTCAGAA--A---C----A-----T---AG-G-G-CAGC-AT---T-A-AAT--GACAAT-GATT-AC-GAGGG-AT-TG-ATT-G--C---------ATGT-T------CAA  8274
MND-2.MND2_AF367411        .A-AAC--CA-A-----CCTCAGA---A---C----A-----T-A-AG-G-A-CAGC-CTA--TGA-AAT--GACAGTAAACT-AC-GA-GG-AT-TG-ATT-G-GC---------AT-T-T--GC--GAA  8345
DEB.CM.CM40                .GA---T-CAGAC---G-CA-T-T-------CA---------A-TGA-GG-AGCA-----T---GATGAGTGGG-AGGAGCTT-AC-GAG-G-A---G-G-----G--G-G-----T---T-CCC--CAA  7935
DEB.CM.CM5                 .GGA----CTGA-------A-T-T-------CAA--------A--GA--G-AGCA--C--T---GATGA-TGGG-AGGAGCCT-AC-GAG-G-A--GG-A--A-----G-GA----TG--T-CC---CA-  7856
MON.CM.L1                  T-ACTTC-CTGAG---G-CA-T-T---A---CAAC-----AGTGAGTT-G-AG-T---G------TGACC--T--GCAGC-CT-G----CTG-----G-G-------GC--A--CC----T--G---A--  8167
MON.NG1                    C-ACTTC-CCGAG---G-CA---T---A---CAAC-A---AGT-TGCT-G--G-----G-T--GGCAC-T--C--GCADC-CT-G--G--GG-T--GG-G--A----G---A----C---T--GC-CA--  6792
TAL.CM.8023                CACAA-C-CTGA----C-GA---T---A---CAA--------G-AGTTGG-AG-----GC-T--GA-ACT--CACAGTA--GT-AC-----G-A---G-A--A--G---GA----TC--T--G--GCAA  7584
TAL.CM.266                 CACA--T-CTCAA---C-GA-T-T---A---CAA--A-----G-TGA--G-AG-----C-----G-GC-C---TCACAG--GT-AC-----G-A---G-G--A--G---GA----TC--T--G--GCAA  8034
MUS.CM.1085                CACCC-G-CAGG----G-AA-T-T---A---CAAC-----AG---GTTGG-AG-----G-T--T-CAACT-----GGAG---T-G---TTGG-A----GA-----G---G-----C-G--TA-----CAA  8138
SAB.SN.-.SAB1C             ....ACT-CCGAC---G-AA---T---G---CAA--A-----G--GC----AG-A--G--TGA-GCTAAC--CAGTAGGA-TT-A--GC--G-A--TG-A---G--C--------TTGG-TAGT-ATCAA  8493
GRV.ET.x.GRI_677           ....ACT-CT-AG---G-CA-T-T---T---T----------G---C-------TGCC-TGGA-G--AAC---ACTCAAC--T-G------G-A-----T---G--TC-------CTGG-TCTG-ACCA-  8199
TAN.UG.x.TAN1              ....AC--CAGAC---G-CA-T-T-------CA---A-----G--CC----A-CTGCC-TGGA-G-AAAT--TAGTA-TACTC---TCA--G-AT-TG-G--AG-GC-G--A---ATGG-TACT-ATCAA  8297
VER.DE.-.AGM3              .AG-ACC-CTGA-------A-T-T---T---C----A-----A---C-G--ATCGT---TGGA-G-TAAC---ACAA-ACA-T-A--G---G-CAG-GCA---G-G--G------TTGG-T-C--ACCAA  7841
VER.KE.-.9063              .A--ACG-CC-A----C-AA---T-------T----------A---C-G---GGAG-G-TGGA-G--AAC---ACAGAACAGT-A-T-A--G-AAG-G-A--AG---G--A---TT-G-T-CC-ATCAA  8343
VER.KE.-.AGM155            ....AC--CAGAG------A-T-T-------T----------A-A-C-G--AG-AGGA-TGGAGG--AAC---ACAAAACA-T-G---C-GG-AAGGG-A--AG-G---------TTGG-T-CT-ATCAA  8320
VER.KE.-.TYO1              .-G-ACT-C-GA----C-AA-T-T---T---T----------A---C----AGCTG---TGGA----AAC--TACGGGACA-T-A-TGA-GG-TAG-G-A---G-G---------CT-G-T-CC-ATCA-  7812
SUN.GA.98.L14              .A-CACG-CC-A-----C-A-GGAC-----------A-----T--GC--G-A--A-TG--TGATGATAAT---ACAG-TC-TC-CC-G--GG-AT-TGTAACAG-GTT-G-----C--A-CA-G--CAAA  8515
LST.CD.88.447              .A-AACT-CT-A-----CAAGGGAT--------T------------C--G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC-----G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA-  7372
LST.CD.88.485              .A-AACT-CT-A-----CAAGGGAT--------T-----------CG-G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC-----G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA-  7369
LST.CD.88.524              .G-TACT-CT-A-----CAAGGGAT--------A--A-----A---C--G--GGATC--TAGA-GCAAAC--CAC-A-AC-GT-GC----GG-TT-C-CAACAG-GTT-G-----AGGA-TA-C--TAAA  7375
LST.KE.-.lho7              .A-CAC--CC-A-----CCA-GGAT-------GA--A-----ATC-A-GG-AGCA-T---TGAC-AGAAC--TACT--TC-GT-AC-----G-AT-C-C-ACAG-GCT-G-----C-GA-TA----TAA-  8447
SYK.KE.-.SYK173            T-CCA---CACAG-----AA-T-T---A---CACAGA-----ACA---GG-AG------TG--TGA-CAT--TG-TGGA--GC-CAG----G-A---G-A--A--G--G-GA----TG--T--TC--ACA  8183
SYK.KE.KE51_AY523867       C-CAAC--CCCAA---G-AA---T---A---CATACA-----AC-GC-GG-AG-T----TG---GA-CAC---G-TAAT---C-GAG----G-T--GG-G--------G-------TG--T--TC-TACT  7873
COL.CM.-.CGU1              AGA---GG-TCCA---C-GA---T--------AAC-----C-TGAGCG-G-A-GA---------GATA-----G-GG-AGACT-A-T---GG-AT-TG--TTA-----GG-A------A-GA----GGCA  7608
```

```
H1B.FR.83.HXB2             GAATTAGATAAATGGGCAAGT...TTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTA  8334
Env                        _E__L__D__K__W__A__S__.__L__W__N__W__F__N__I__T__N__W__L__W__Y__I__K__L__F__I__M__I__V__G__G__L__V__G__L__R__I__V__F__A__V__L__S__  gp41

H1D.CD.84.84ZR085          -----G--C--G-----G---...----------------G-------C-A------------------A--------------------------A----------------------A----------  7855
H1F1.BE.93.VI850           -C------C------------...C---------------G-----T---C------------------A-----------------G---------A---C---------------------------  7611
H1G.SE.93.SE6165           -CC--G--CC-G---------...----------------GG-------GA-----A------------A----T----------------T--AA-------------------------G-------  7752
H1H.CF.90.056              -C---G--C------------...C------C--------G-----T--C-------------------A-------------------T--AA----------------A---------G-------  7653
H1J.SE.93.SE7887           -C---G--C--G---A---A-...---------------------T----C----------C--C----A-------------A-------T--AA--------------A----------G---G---  7641
H1K.CM.96.MP535            -C---G--C------------...C---------------G---------A-----A------------A-------------A----------A-----------------C---------G------G  7494
H101_AE.TH.90.CM240        -----G---------------...C---------------G--------------------------- A----T----------------T--AA----------------A---------G-------  7887
H102_AG.NG.-.IBNG          -C---G--C--G---------...C-A-------------G----------------A----------G-A----T----------------T--AA-------------------------G---A---  7840
H103_AB.RU.97.KAL153_2     -C-------------------...----------------G-----T---A------------------A------------------------------------------A-----------------  7511
H104_cpx.CY.94.CY032       -CC--G--C--G--------C...C----------------G---------A-----A-----------A----T---------------------A----C----------A----------G-------  7704
H1N.CM.95.YBF30            -----G---C------AC---...C------GC------GGT--T-----A------------------A--GCT--------------C----A-T-----CA-------CA-AAG-ATA---A-AA---  7885
H1O.BE.87.ANT70            --G------G-------CTC-...A-T---------C--G-------T--A---T--------------A--GCA-----C-----------CAC-A-----GG-G---G-TA-CA-GATA------AA--  8382
H1O.CM.91.MVP5180          --GC-----G-------CTC-...C-T-------------G-------T--A---T-------------A--GCT-----C-----G----CAC-AA-----A-----G-TA--A-GATAA-----AA-C  8408
CPZ.CM.04.EK505            -----G---C-----T-----...--------C-------G----T-----A----------C------A-T---------G-------C-----A-T-----CA---------A-AAGCATAA--A-G----  7820
CPZ.CM.04.LB7              ---C-G---C-T---A----C...C---------------G------GTC-C----------C------A-------T------A-----------AA---TG-GT--G---A----------G--CG---  7845
CPZ.CM.04.MB66             --------C--G-----G--C...C-C-------------G--------TCA-----------------A----T--C--A------------A--AA--------------------------T-G-----A-  7958
CPZ.MT145                  -----G---TC------T-A-...--A----G--------G---------------T-------C-----GA-G---T----AG-----------A-TA----C----------TGC-----AA--GGA--AC  7846
CPZ.CD.-.ANT               ---C----C------AGCTCA...--A-------------G---------C-A-----A----------A----T--T---------------CTA-T-----AC-------TT-GC---TAT--G--AG-T  7793
CPZ.CM.-.CAM3              -----G---C-----T-----...--A-------------G-T--T--CCGC----------C--------G--T-------------C-----A-----CA-----------AGGG---A-CA-A--AT  7750
CPZ.CM.98.CAM5             -----G---C-----T-----...--A----G------C--T-----TC-G---T--------------A----TT-G--AG------C---T--AA-------T---C--A-AGGAAT---TA-G--AG  8056
CPZ.GAB_X52154             -----G---C-------C--C...----------------G---------A------------------A-T---C-T---GC----------AA-CA----G---------TA--A-GA-A--TT-C--AG  8374
CPZ.GAB2                   -----G---C-----T-CTCA...--A-------------G---------C------------------A-G--T-----CT----G-----GG-TA---------------T--AA-A-G----G-AAA--  7721
CPZ.TZ.-.TAN1              --------C------AGC---...C-T---G---------G-T------C-A-----------------A--GCT-----C-------C---A--A-----AC-T-----TC-CA-GTT-A--G--AA-G  7990
CPZ.US.85.CPZUS            -----G---C-----T-----...C-A-----C-------G-T--T-C--C------------------A----TT----AG------C-A----A-----AA-C-----T--AGG--TGA--T----AC  8342
GSN.99CM166_AF468659       A-G--GTCC-------ACTCC...C-T----G--------G-T--TT-CG------A-A----C--T--GA-C--TG---AG----T-C--C---AA---C-------------AA-GTT-A-C---AA--  8268
GSN.99CM71_AF468658        A-G---TCC-------ACTCC...C-T----G--------G-T--TT-C--------C-----C--T--GA-T----------G----T-C--CT--A----C---G------A-AA-GTT---T---AA--  8280
H2A.ALI_AF082339           A-----A---GC----ATGT-...--TACC--C---C--G--T----CGCC---G-CAA------TC---ATGGAG-TTAT---A---T---AA-A----C-C-T--------AA-ATA----G-GCAA-  8779
H2A.BEN_M30502             A-----A---GC----AT-T-...C-TG-C--C-------G--T----CTCC---G-CAA------TC---ATGGAG-GCAT-------TG--AA-AA---C------------CAA-CTA----G-GCAAT  8795
H2A.ST_M31113              A-----A---GC----ATGT-...--TG-C--C-------G-TT----CTCC---A-CAAA-----TC-G-ATGGAG-TTAT-------T---AA-AA---T-------------AA-ATA----G-ACAA-  8225
H2B.D205_X61240            A--A--A---GC----ATGT-...--TG-T----------G--C-C--CTCC---A-AAAA-----CC-T---GGAC--TAT-------C---G--A----TG-----------AG-ATACA-TG-GCAG-  8771
H2B.EHO_U27200             A-----A--C------AT-T-...--CA-T----------G--T-C--CTCC---A--GCA--C--C-GG---GGAT--TAT------AT---AA-A----TA-----------CAA-ATACA-TA-ACAG-  8757
H2G.ABT96_AF208027         A-----A---GC----ATGTA...--TG-C----------G--C----CTCC---G-TAAA--CG-TT-C---GG-C-TTATG----G-C---AG-AA---TG-----------CA-ATA----G-ACAA-  8168
H2U.FR.96.12034            A-----A----C----AT-TC...--TG-T--C-------G-TC-C--CTCA---G-TAAA---G--T-C---GGAC-GTATG------C---AA-AA--ATAC-T---------GA-ATA----G-ACAGC  8279

MAC.239_M33262             A-G--GA---GC----ATGTG...--TG-C----------G--C-TG-TTC----A-AAA-------C---ATGGAG-TTAT-----T-T---AG-AA--CTG-----------GA-CTA-A--G-ACAA-  8740
Env                        _K__L__N__S__W__D__V__.__F__G__N__W__F__D__L__A__S__W__I__K__Y__I__Q__Y__G__V__Y__I__V__V__G__V__I__L__L__R__I__V__I__Y__I__V__Q_  gp41

SMM.SL.92.SL92B            A----GA---GC----ATGTG...--TG-A----------G--C-T--CTCC---G-AA-A--C---C-G-ATGGAG-CT-TC----CAT---AA-A---TTG------------A-CTA----G-ACAA-  8188
SMM.US.x.PGM53             A-----A---GC----AT-T-...--TG-C----------G--C-T--TTC----A-AAAA------C-G-ATGGTG--C-T-----TCT---AG-A-----G-----------GA-ATA----G-GCAA-  8686
MNE.US.-.MNE027            A-G--GA---GC----ATGTG...--TG-C----------G--C-TG-TTC----A-AA--------C---ACGGAG-TTAT-----T-T---AG-AA--CTG-----------CA-CTA-A--G-ACAA-  8217
RCM.NG.x.NG411             A----GA--TC-----G-GAC...A--CTC-G----C-A-----GGATTGG---T-AAA--------GGA--GGA--TT-C--TA-TTT---AA-AA----C----------A-C---TTAC-CTGGAGCT  8050
RCM.GAB1_AF382829          A-----AG---T----G-GA-...GCA-TC-G----C-G--TC-TGACTGG---A--CAA----------A--GG-T-CT-T------AT---TA-CA----A------G---C--GGCTGT--TGGAA-T  8006
DRL.FAO_AY159321           A------GGG-T-TAA--TCA...-G-GCT--C-------G-TT-C--TTGG---TCTAAA-----T--GA--GGAT-TT-C------ATG-CAA-AA----A---------TT-AGCA---T--TGGAGC-  8163
MND-1.F17                  A-G----GAG-CCTTA--TC-...-G-GCA-G--------G--T-T--TTGG---G-TCAA--CT------GGGGAG-TT-CT----GTT---AA-TA----A---------TT-G--A--CT--TGGAA--  7929
MND-2.5440_AY159322        A------GGG-C-TAAACTTCCATGG-CTA-C-G----CG--C-T---TGG-----TAA----G-T--GA--GGAC-GT-AG------T--TTA-CA----A---------GT-AGC-TGCT--TGG--AG  8108
MND-2.14CG_AF328295        A----G-GAG-T-T-A-TTCA...-G-GCT-GC-----CG--C-C--GTGG----CAAA----G-T--GA--GGA---T-----A--ATG-TAG-AA----AC-T-----TT-AGC-TG-T--TGG----  8401
MND-2.MND2_AF367411        A------GAG-T---A-TTCC...-G-GCT-GC-----CG--T-T--GTGG----TAAA---G-T--GA--GGTT--C-A------AT--TGA-TA---TG-----G--TT-AGC-TGCT--TGG--AG  8472
DEB.CM.CM40                TC-C--C-AG-T----ACTC-...--A------------G--C-GT-C-GA---T-C----GG--T-G----G-AG--TAC---A---CT-CT--AA--TTGC---------C-A-GTT--G-G-GAA--  8062
DEB.CM.CM5                 TC-C-GAC-G-T----ATTCA...--A------------G--C--T-C-GA---T-C----GG--T-G----G-AG-GTAT---A---C--CT---A--TTGC-T--G----CAA-GTT--GGG--AA--  7983
MON.CM.L1                  A-GCCG-GAC------ACTTC...C--------------G---CT-C--A---T-------C--C-----A-------T--AG------C--CT---A----C----------C--A-GTTCA----AGGAG  8294
MON.NG1                    A----G-GAC------ACT-G...--A------------G-T--CT-C--A----T--------T--GA-C--------AG-G----C--CT--------CC----GG-GC-CA-GTTCA-C--AGGG-  6919
TAL.CM.8023                A-G---A--G-C----AC--C...C-C----GC-----C--TT-GT-TGC----T-TA-A-GGC-T-G-A--GCAG----AG------CTA----AA-TTTGC-T-----T--AA-GTACA-TA-CAA-G  7711
TAL.CM.266                 A-G---A--G-C----AC---...C-C----GC-----C--TT-GT-TGCC---T-TA-A-GGC-T-GGA--GCAG----AG------CTA-TC-TA-CTTAC-T--G--T--AA-GTACA--G--AA-G  8161
MUS.CM.1085                A----GTTAG-G----ACTCC...C-C---G-A------G-T--CT-C--A-----C--------C---A-C---TGC---G----T-C---TC-A--GCT---T--GC----CA-GTT---GT-AGGA-  8265
SAB.SN.-.SAB1C             A----G-TGTCT---T--GAC...--T---TCG------G--C-C-----A---T-TG-A-GG--G--GA--GC---C---G-TA---C----A-CA---TAGCT---G-TC-GC-A-TAA--A-AGGA-  8620
GRV.ET.x.GRI_677           A----G---G-T---T--G-G...--C---TCA-----CTCAC-GT---C----T-AG-C---G-T---A--GGAT-TT-AG-GA---TGATTA-TC----A------T-T-CA-GG-TAT--TGGGGAT  8326
TAN.UG.x.TAN1              A-G----G-G-T---A-TTC-...-G------CATC---G--G-CT--TCA---T-C----GG--T-----GGGGAT-TTAT------AT---AC-TA--TTG--T--G--G-C--GGCT-A-TTGGGGAT  8424
VER.DE.-.AGM3              A-----AG--GT---T--GA-...--C---TC------CG-TT-CT----G------AACAT-C-----A--GGAT-TT--GAT---CT---TA-TA----A------T-GC---A-A-A---TA----T  7968
VER.KE.-.9063              AG---GACATCT---T---AC...--C---TC------CG-TT-CT----A-----AAACAT-C-T---A--GGAT-CT-AG------T---TA-CA----AC-T--GT--C-G-A-A-AA--TA-AG-T  8470
VER.KE.-.AGM155            A-G--GTCAG-C---T-G---...--T---TC------CG-TT-TT----A------AACAT-T----GA--GG-T-TT--GC----AT----G-TA----G------T-GC---ACA-AT--TA-A--T  8447
VER.KE.-.TYO1              A-G---AC--GT---T--GA-...--C---TC------CG-TT-CT----A-----TAACAT-T-----A-GGGAT-TT-AG--A---T---AA-AA----G------T--C---ACA-A---TA-GGAT  7939
SUN.GA.98.L14              C-GC--C-AG---TTAAC......--C----G------GG--C-GT-TC-A---T-TCTA--C-------ATGCAG-CC----TA----CATTA-AA---CAGCT-----TC--AG-TT-A-CA-ACAAC  8639
LST.CD.88.447              A-GC--C-GG-G-TCAAT......--T-----------AG----CTT-TCA---T-CCAA-----T--G-ATGCAG--C-A---A--AT---TA-TA---T-------G---AAGCTT-A-TG-GCAA-  7496
LST.CD.88.485              A-GC--C-GG-G-TTAAC......--T----G------GG-T--TTT-TCA---T-CCAA-----T--G-ATGCAG--C-A---A--AT---TA-TA---TG-----------AAGCTT-A-TG-GCAG-  7493
LST.CD.88.524              A-G---C-AG-T-TTAAT......--T----G----A-GG--C-C--T-C----T-CCAA-----T--G-ATGCAG-GC-T---A-CAT---TA-AA--ATAC-------TC-AAGCTT-A-CA-ACAA-  7499
LST.KE.-.lho7              A-G---C-AG---TTAAC......--T----G------GG-T--CT--C-C---T-TACT---G-T--G-ATGCAG-GT-A--TA--CTT-TAA-CA----G------G--T-GAGCTT-A--A-ACAA-  8571
SYK.KE.-.SYK173            A-G---C-GG------ACTCG...-------GC------G--T-GT-G--A---T-C-TT---C----GA--GGAT-TTATG--A------CT--A--GCT------GC-----AG-TT-AGTG-AGGA-  8310
SYK.KE.KE51_AY523867       A-----C-AG------ATTCC...C------GC------G-TC-GT-G--A---T-CCA---CC-T--GA--GGAT-CT-TGCCA---C--CTA-A--GATA------G--C--AG-TT--CTTGGGGC-  8000
COL.CM.-.CGU1              -------GGG-T---A-T-A-...-G--TTTCAG-C---GG-C--TTC--CATCT-TAA----G--CTC-ATGCTGC-TATG-T--------T--AA----AC-T-----TA--A-G-TA---A-AG--T  7735
```

```
                                                                    / Tat, Rev exon 2 start                                    Tat exon 2 end \
H1B.FR.83.HXB2           TAGTGAATAGAGTTAGGCAGGGATATTCACCA...TTATCGTTTCAGACCCACCTCCCAACC......CCGAGGGGACCCGACAGG...............CCCGAAGGAATAGAAGAAGAAGGTGGAGA  8440
Env                      I__V__N__R__V__R__Q__G__Y__S__P__.__L__S__F__Q__T__H__L__P__T__.__.__P__R__G__P__D__R__.__.__.__.__.__P__E__G__I__E__E__E__G__G__E  gp41
Rev ex2                                                                 N__P__P__P__N__.__.__P__E__G__T__R__Q__.__.__.__.__.__A__R__R__N__R__R__R__R__W__R   Rev
Tat ex2                                                                   P__T__S__Q__P__.__.__R__G__D__P__T__G__.__.__.__.__.__P__K__E__*                   Tat

H1D.CD.84.84ZR085        ----A-----------------G--C-----T...----------------T-------G--......-------------------...............--A--------------------------  7961
H1F1.BE.93.VI850         ----A------------A-------C-----T...--G--A--A-------TTA------G-......-------------------...............-----------C-------G---------  7717
H1G.SE.93.SE6165         ----A--------------------C-----T...--G--A----------TTAC--ACCA-......-A-----A------------...............-----------C-------G------C--  7858
H1H.CF.90.056            ----A--------------------C-----T...--G--T----------TTG----G-A-......--AC---------------...............----------C--------G------C--  7759
H1J.SE.93.SE7887         ----A--------------------C-----T...--G--A----------TTA-------A-......--A-C--A-G--------...............----G------C-------G---------  7747
H1K.CM.96.MP535          ----A--------C-----------C-----T...--G--A----------TTA--------------......T--------G-A-----A...............-----------C-----------------  7600
H101_AE.TH.90.CM240      ----A--------------------C-----T...--G--T--C-------CTTC--ATCAT......-A--A--A---------A...............-----------C-------G------C--  7993
H102_AG.NG.-.IBNG        --A-A--------------------C-----T...--G-----C-------TTAC--ACCA-......-A-----A---------A...............-----A----C-------G---------  7946
H103_AB.RU.97.KAL153_2   ------------G-------------------...-----A----------G-------G--......-A------------------...............---------------------A------  7617
H104_cpx.CY.94.CY032     ----A--------------------C-----T...--G--T--G-------TTA-------A...ACC-AAC------T-------...............----G-----C---------------C--  7813
H1N.CM.95.YBF30          --A-AGCA-------------------T--C...C-T-----G-------TTA-------A......G-A--------A------...............--A---A--C-----G--GC-T------  7991
H1O.BE.87.ANT70          -------A-ACA--------------CA---C...C-C-----A----T--C-AA--ATCA-......-AAGA--A-G-A-GA-C-...............--A-G-A---C--G--G--G---------  8488
H1O.CM.91.MVP5180        -------G-ACA--------------CA---C...C-C-----G----T--CTG-----CA-......-G-CA--A-G-A--A-C-...............--A-G-A---C--G---------------  8514
CPZ.CM.04.EK505          -G--AGCA-------------------T--C...C-C-----G-------TTA-------G......A-TC-------A------...............--A--A---C---------C-C------  7926
CPZ.CM.04.LB7            ----A--------C-----------C-----T...C-G--A----------TTTC---C-T-......----------T--CTCA-...............----G-----C------------------  7951
CPZ.CM.04.MB66           ----A--------------------------T...C-G--A----------TTA-----G--......-G-----AT-GA------...............--A---A---C-------G---------  8064
CPZ.MT145                --A-C-----G------A-------C-----T...--G--A--G-------TTA-----G--......AATCA------A---G--...............-T--G--A--C----A---G---------  7952
CPZ.CD.-.ANT             GCT-A-GA-AG---------------CAT--T...C-G--A------......A----T---CAAAAC-A-CA--AT--A--GCA-...............--A---A----AG------------AG  7899
CPZ.CM.-.CAM3            -T--AGCA-AG---------------C--T--C...C-C-----G-------TTA----G--A......A-A----A---A------...............--A-G-----C----------C-T------  7856
CPZ.CM.98.CAM5           -TA-AGCA-AG---------------C--T--C...C-C-----G-------TTA----G--A......A-A----A---A------...............-GA-G-----C----------C--------  8162
CPZ.GAB_X52154           ----C-GG--------------C--C-----T...C-C-----G-------TTA-----GT-......-A-----A--AA-GA---...............-T--G--A---C--C---G------C--  8480
CPZ.GAB2                 --A-A-GA---AG------------C-----T...-----------------TTA-----G-A......A-AG-------A--AGT-...............--A---------C-G--G--G---------  7827
CPZ.TZ.-.TAN1            ----T--GCA------------T---A----C...C---TT-CA----T--CTAC--A-G-G......GA-CA--AT--A--ACA-...............--A-G------C-C--G--G--------G  8096
CPZ.US.85.CPZUS          ----AGCA-A------------T-----T--C...C-C-----A-------TTT-------A......A-TC---A---A------...............--A-------C---------GC-C----A-  8448
GSN.99CM166_AF468659     -GC-T-GA--GA----------C--C-----T...C-C--CCCG----T--TTA-----CAAGCAGTGGAAC-C--T--G...................--A--C-C--CC-------GC-T---C-C  8374
GSN.99CM71_AF468658      -GC-T-GA--G-----------C--C-----T...C-C--CCCG----T--TTAC----CAGCAAGTGGAAC-C-AT--G...................--A--C----C-AG-A----C-----C-C  8386
H2A.ALI_AF082339         -GT-A-G----C-C---A----C---AGG--TGTT--C--C-CC-CTC--GGTTA-ATCCAACAGATC-AT-TCCACAAG...GACCAGGAA...CAG...--AACCA--GG----AC-----AA-AC-T  8900
H2A.BEN_M30502           -GT-A-G----T----AA----C---AGG--TGTT--C--T-CC-CCC--GGTTAT-TCCAACAGATC-AT-TCCACAAG...GACCGGGGA...CAG...--A-CCAACGA--G-AC-----AA-AC-T  8916
H2A.ST_M31113            -GT-A-G----C----AA----C---AGG--TGTT--C--T-CC-CCC--GCTTA-TTCCAACAGATC-AT-TCCACAAG...GACCGGGAA...CAG...--A-CCA--GA----AC-----AA-AC-T  8346
H2B.D205_X61240          -GC-AGCA---C-----A----C---AGG---GTA--C--C-CC-CTC--TCTTATA-TCAACAGATC--T-TCC-CAAG...GACCGGGGA...CAG...--A-CCAACGA----AC-----AA----G  8892
H2B.EHO_U27200           -GC-AGCA--GC-----A----C---AGG---GTA--C--C-CC-CTC--TCTTATA-TCAACAGATC--T-TCC-CAAG...GACCGGGGA...CAG...--A-CCAACGA----AC-----AA----G  8878
H2G.ABT96_AF208027       --C-AGGA---C-------A--G---AGG--TGTT-YT--T-CC-CTC--TCTTATGTTCAGCAGATC-AT-TCC-CAAG...GACCAGGAA...CWG...--AACCAA-GA----AT------A-A-AG  8289
H2U.FR.96.12034          -GT-AGG---TC-----A-A--T---AGG---GTT--C--T-CC--CC--TCTTATGTTGAACAGATC--T-TCCA--GG...GACCAGGAA...CAG...--AACCAA-GG----AT-----AA--C-C  8400

MAC.239_M33262           -GC-AGC--AGT-A--------G---AGG---GTG--C--T-CC-CAC--TCTTATTTCCAGCAGACC-AT-TCCA--AG...GACCCGGCA...CTG...--AACCA--GA--GCA-----A-A-AC-G  8861
Env                      M__L__A__K__L__R__Q__G__Y__R__P__V__F__S__S__P__P__S__Y__F__Q__Q__T__H__I__Q__Q__.__D__P__A__.__L__.__P__T__R__E__G__K__E__R__D__G  gp41
Tat ex2                                                                      P__I__S__N__R__.__T__R__H__.__C__.__Q__P__E__K__A__K__K__E__T   Tat
Rev ex2                                                                       P__Y__P__T__.__G__P__G__.__T__.__A__N__Q__R__R__Q__R__K__R__R   Rev

SMM.SL.92.SL92B          -GT-A-G---GT-A--------T---AGG--TGTT--C--C-CC-CCC-ATCTTAT-ATCAGCAGATC-AT-TCCAG-GG...GACCAGGAA...CTG...--A-CCAA-GA--G-AC------A-A--G  8309
SMM.US.x.PGM53           -GA-AGC---GT-A-----A--T---AGG---GTG--C--T-CC-CTC--GTTTATGTTCAGCAGATC--T-TCCAC...A-GG-CCAGGAA...CAG...--AACCAA-GA--G---------A-AC-G  8807
MNE.US.-.MNE027          -GC-AGC--AGT-A-----A--G---AGG---GTG--C--T-CC-CAC-TTCTTATTTCCAGCAGACC-AT-TCC---AG...GACCAAGCA...CTG...--AACCAA-GA--G-AC------A-AC-G  8338
RCM.NG.x.NG411           GT--T-G--AGC-----GCA--T--CA-G--TTTGC-C--TCCC-CAC-TTGTTAT-ATCAGCAGATC--T-TCC-C...A-TC--GAGGAA...CAG...--AATCAA-GA--G----A----A--CAG  8171
RCM.GAB1_AF382829        GTC-T-G--ATC-------A--G---AGG--T...C-C--C......C-A-C-TCTTATGTTCAGCAGATCCATATC-A-A---C-GGGGAACCGCAAACT--A-G--A--A-AG------C--------  8127
DRL.FAO_AY159321         C---AGG---GT----------T---CGC--TCTTCCT-AT...ATCTTTA-GGA-G--TA-CGC...--AGA-ACGA--A--TAC...CAC...AGACAG--A--CA--GA-A---G-----AGCA---  8281
MND-1.F17                C-A-A-G---GT----------C---CG---T...G-C-TT-CA---..............GACTGC-A-CA-AAC-TAT--C-CAAACGG.........--A--CAACGG----------A-CAACAG  8032
MND-2.5440_AY159322      --T-AGGG-AGT--------------CGC--TCTTCCT-AT..............GTCTT-AAGGGAGACTATCTC-GGCC-CACAACCTC...AAACGG--A--CA--GA--G--G-----AGCC---  8220
MND-2.14CG_AF328295      C-A-AGG---GT----------T---CGC--TCTTCCT-AT..............GTCTTAAAGGGAGACTATCTC-GACCTC-CAACCTC...AGACAG--A--C-A-GA-AG-----G--AGCTCA-  8513
MND-2.MND2_AF367411      -CT-AGGA-AGT----------T---CGC--TCTTCCT-AT..............GTCTT-AAGGGAGACTATCTC-GTCC-CACAACCTC...AAACAG--A-GCAA-GA-AG--------AGCCC--  8584
DEB.CM.CM40              -T-GA-G--AGT-GT-T-GA--------C--T...--GCA-ACC-TT---..........----AG......GA-GA--A-GTGCTTC-A.................-T--C--A-GA--G--GCAG-A------G  8162
DEB.CM.CM5               -C-G--C--AGT-GT-T-GA--------C--T...--GCA-ACC-TT---..........----AG......GA-GT--A-GTGCTTCAA.................-T--G----GA--G---CA--A-------  8083
MON.CM.L1                -TA-C-G---GT-AG----------C--T-TTCTT-CT--T.............-AG-T-CCTATC--ATCCCACG-G-GACA-CCAACG...CCCGACGGAAC---CGC--GC-GT-GC-AC--CAG  8406
MON.NG1                  -CA-TCG---G--AG-------T--C--T-TCCTC-CTC-A.............-AG-T-CCTACC-AAGTCCCT-GG--ACCAGGAGCG...CCAGAAGAAAC----GA--GC-GT-G--AGCA--G  7031
TAL.CM.8023              --T-A--GTTGC-C--------C--C-T---TCTACCC-TT.............AAG---CCTATC-----CAACAGG--ACCAGGAGTG...CCAGAAGAA-----T-C-----GT-G-AAGCCC--  7823
TAL.CM.266               -GC-C--GTTGC-C--------C--C-T---TCTTCCG-TT.............AAG---CCTATC-----CAACAGG--TCCAGAAGCG...CCAGGAGGA--C-CT-C-----GT-G-AAGCA--G  8273
MUS.CM.1085              --C-A-GG--------------C--C--T--CCTT-CTC-T.............-AG-T-CTTATC-GACC-CAGA-G-GACCAGAAGCG...CCAGAAGGAAC--A-GG---C-GT-G-A-C---AG  8377
SAB.SN.-.SAB1C           -TC-T-GG-AGT----AA-A--G---G----TCTC-CC--CC-C-CCT-TTCT-ATTA-..........-A--TCCATAT-C--CT-GACAAG...GGACAG--A--CAA-GA------GC----CA-C-G  8738
GRV.ET.x.GRI_677         GTA-C-GA-ATA--------------AAT--TCTC...C-C.....................CAGATC-AT-TCCAC...AGTTCAGCGGAA...CGG...--A...-AC-AC-G--G--GGCAA-ACAG  8420
TAN.UG.x.TAN1            GCA-AGC---G-----------T--C-TT--TTTG-CTC-T.....................CAGATCAAT-TCC-C...CTGG--AGGGAA...CAG...--A...-AC-AC-C--G--G--AA-ACA-  8521
VER.DE.-.AGM3            GCA-AGC---G-----------T--C--T--T...C-T--TCCA----T---TA---ACC-GTGG...AA-G-ACAG--A----ACGCAGAA...GGG...--A-G--A-GGT-G---CA-GC-CAAGA-  8086
VER.KE.-.9063            GCA-CGG---G-----------C-----T--C...-----TCCA----T---TA---ACC-GTGG...AA-G-AC-G--A-G--ACGCAGGA...GAG...--A----A-GGT-G---GAG--ACAAC--  8588
VER.KE.-.AGM155          GCA-AGC---G-----------T--C--T--T...-----TCC-----T---TA---ATC-GTGG...AA-G-ACAG--A----ACGCAGGA...GAG...--A----A-GGT-G-AG-AC---CAA-AG  8565
VER.KE.-.TYO1            GTA-AGTG--G---------------GTT--T...C----TCCA----T---TA---ACCAAGTGGGGAA-G-AC-G--A----ACGCCGAC...GAG...--A-G--A-GGT-G---CA-CA-CA-GAT  8060
SUN.GA.98.L14            AGA-CT----GA-GT-T---------AGGGTG...--G--CCC-TCTG-TT-TG-TGA-CAGGATTGG-TCCA--A-.........ACCTGC...CCCAAA---AC--AC-A-----------AA-AGAC  8754
LST.CD.88.447            ATA-AGTA-A-A-GT-TAG-------AGGGTT...C-TG-CCCCTCTG-TT-TG-TGA-CAGGACTGCAA-T---AGAAAAGAGA-AACCAA...GAACAG--A--CAA-GA--G---GAT--AGAA---  7620
LST.CD.88.485            ATA-AGTA-A-A-GT-TAG-------AGGGTT...C-TG-CCCCTCTG-TT-TG-TGA-CAGGACTACAA-T---AGAAA-GAGA-AACCCA...GAACAG--A--CAA-GA--G---GAT--AGAA---  7617
LST.CD.88.524            GT--AGTG-AGA-GT-TAG---C---AGGGTT...C-CG-CCCCTCTG-TT-TG-TGA-CAGGACTTCAA-T---AGAA--GAGA-AGCCAA...GAACAG--A--CAA-GA--G---GAC--AGAA---  7623
LST.KE.-.lho7            AT--AGTA-A-A-GT-TAG---G---AGGGTG...C-C--CCCCTCTGTTT-TA-TGA-CAGGACTACAA-T---AGAAA--AGA-AACCAA...GAACAG--A--CA--GA------GA----A-C---  8695
SYK.KE.-.SYK173          -CA-A--G-ATT-GTT-GGA--C---GTT--TATT--GCA-AAC-CT-----G.........................GG-CGG-A-GACCCA...GGCAAA----C--ACGA--------GCA-------  8413
SYK.KE.KE51_AY523867     ----A-GA-ATA-GTT-GGA--C-----T--TCTT--GCA-AAG-C-CAG-C--CTA-CCT-TACCACAA-GCCCC-G-A-GGCCAGAAGAA...GGAATAGAAAC--AT-C---CA-TT-CAAGCCCAG  8127
COL.CM.-.CGU1            GCA-A-GAG---C-TTCAGA-TTA-GGG-TTT...CAGCA-A-AGGC-GGACTAATGTGT-TTCACAGATCC-T--C-AA---C--GAGTGG...GAACAG--A--CAACGC-------AT-CACA----  7859
```

```
H1B.FR.83.HXB2          GAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCCTTGGCACTTATC.....................TGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTC  8549
Env                     _E__R__Q__R__Q__I__H__S__I__S__E__R__I__L__G__T__Y__.__.__.__.__.__.__.__L__G__R__S__A__E__P__V__P__L__Q__L__P__P__L__E__R__L__T__  gp41
Rev                     __R__D__R__D__R__S__I__R__L__V__N__G__S__L__A__L__I__.__.__.__.__.__.__.__W__D__D__L__R__S__L__C__L__F__S__Y__H__R__L__R__D__L__L   Rev

H1D.CD.84.84ZR085       -CA--------GT-C-----------G---------T--CA---------.....................------------A---A-------------------------------A---A--  8070
H1F1.BE.93.VI850        -CA--G--A----------G-G--C-------C-----T---A--T---GC-.....................------------A---A--------------------G--A------------CA-A  7826
H1G.SE.93.SE6165        -CA--G-----G------AG-G---------G-----T---AC-G------.....................--------C------------------------------------------CCA--  7967
H1H.CF.90.056           -CA----------------G-GA----------------T---AC--G--G--.....................--------C--C---------CA-----------G--T----------------  7868
H1J.SE.93.SE7887        -CA--G----AC------G------------------T-C-A--T---GC-.....................--------C-------A---------------T--------A-----------CG--  7856
H1K.CM.96.MP535         -CA-----A-A-------AG-------------G---C-T---A--G---GC-.....................---------------A--------------------G--AA------A-----A--  7709
H101_AE.TH.90.CM240     -CA--G-----------AG-G---------G-----T---A------GC-.....................-----------A------------------------------G-------------AC-  8102
H102_AG.NG.-.IBNG       -CA-----A----------G-G---------G-----T---A------GC-.....................----------------------------------------A-------------A--  8055
H103_AB.RU.97.KAL153_2  --------------C--------------------A---T---A---------.....................--------C---A-----------T------T--------A----------------  7726
H104_cpx.CY.94.CY032    -CA-------AG------------C-------------T----C---------.....................--------C-------A--------------------G--A-------A--------  7922
H1N.CM.95.YBF30         -CA--------G-------G-G----------G-----T--CA--T---G--.....................-----G--C--C----A-----TGA-----CT---------------C-----C----  8100
H1O.BE.87.ANT70         AGA--G---GCC---G-GG--A-CC-CGCC-C-A--G-T----C----GT-GTACACGGACCTCAGAACAATAATATTGTGGACTTACCA---C-TGAG-AA-TTAGCATCAG-GA-CCAGA-GG-GA--  8618
H1O.CM.91.MVP5180       AG-------GCC--AG-GG-CAGCC--GCCACCA----T----CA--AGT-G.....................-ACACG-----CA---CAA-AAT-T-G-GG-CT-------T------CA-----A-A  8623
CPZ.CM.04.EK505         -CT-----AC-G-------G-G--C-------G-----T---A--T---GCG.....................-----G--CT-CA---A---TCTA------CT----------------C-----GT---  8035
CPZ.CM.04.LB7           -CA--G--C---------AG-GA----GC--G------T---A------GCT.....................---A----C---A--GA-------------CTT--------A----C-------G--  8060
CPZ.CM.04.MB66          -CC-----AC-T------A-----C-------G-----T---A------GC-.....................---A----C--A---GA------T------CTT--------A----------------  8173
CPZ.MT145               --AC-T-----G------A--CA----G----G---G-T----C-T---G--.....................-----A-----T---A---CCT-AG----CT--------AA----------GTGCA  8061
CPZ.CD.-.ANT            A-A------GAT---G-GG-GGGCC--GCA-C----G-T---C-----CT-G.....................----TG--C---AC----A-AAT--AG-GG-T------GAT--GC---AC--GT---  8008
CPZ.CM.-.CAM3           -CC--G--A--G------------C-------G-----T---A--T------.....................-----A-----C---A---CCTG------T-T--------A---------------A  7965
CPZ.CM.98.CAM5          -CC--G-----G--------CGA----G--C-G---G-T---A--T---G--.....................-----A-----T---A---CCTGA-----TT---------A----T----------  8271
CPZ.GAB_X52154          -CA-------AG------AG-GA--------G-A--G-G----C---------.....................--------C--C--A-A----G-GA-T-GG--------GA------C-AG----GC-  8589
CPZ.GAB2                -CA--------G------TG-GA-------C-G-----TT--A--T--CTG-.....................--------CA-C----A----ACGA-----CT-----G--T---------T-GT---  7936
CPZ.TZ.-.TAN1           C-------ACAT---G-GG-CG-CC-CGCCAGCA----T---CAGTA-CG--.....................-----G--C--CA---A---CCT-A---GG-TA-----GAC---TCA-A----CA--  8205
CPZ.US.85.CPZUS         A-C-----AC-T------A-CGA-------C-G-----T---A------G--.....................-----A-----T---A---CCTGA-----CT----------A---GAG--------A  8557
GSN.99CM166_AF468659    -CA--G--AG-T---G---G-GA---ATC-A-CA----T--C-AGT--CT-G.....................--------C---A---A----GTGA-T-GG-T------GAT----GCCAG----GC-  8483
GSN.99CM71_AF468658     -CA--G--AC-T---G---G-CA---ATC-C-CA----T--C-AGT--CT-G.....................--------C---A----T---GTAA-T-GG-T------GAT----GCCAC----G--  8495
H2A.ALI_AF082339        TG------AC-TTG-GGA-............................-GATTGTGGCCCTGGCCGATC...GCAT-TTTA-ATTTCCTGA-CCA---GC-AGCTCG--T-TTGA-CG-GCTG-ACAG-  8997
H2A.BEN_M30502          CG---G-GACAGTG-T-A-............................GA-TTGTGGCCTTGGCCAATA...AACT-T-TG-A-TTCCTGA-CCA---AC-G-CTCG--T-TTGA-CG-GCTA-ACAA-  9013
H2A.ST_M31113           TG--A----C-TTG--GA-............................-ATTGGTGGCCCTGGCCGATA...A-AT-TATA-ATTTCCTGA-CC---AGC-G-TTCG--T-TTGAAC---CTA-ACAA-  8443
H2B.D205_X61240         TG--A--GAC-GGGAC-A-............................-GATCCTGGCCCTGGCAGATA...GAAT--AT--ATTTTCTA---C---AGC-G--GA---T-TTAA-TT-GCTA-ACAA-  8989
H2B.EHO_U27200          TG--A---AC--GG-C-A-............................-GATCTTGGCCCTGGCAGATC...GAAT--ATC-ACTTCCCGA-TC---AAC-G--GG---T-TTGA-TT-GCTA-ACAG-  8975
H2G.ABT96_AF208027      TG--A---A--GTG-C-A-............................-GATCTTGGCCTTGGCAGATA...GAAT-TAT--ATTTCCTGM-CC---AGC-AG-GA---T-TTGACTT-GCTA-ACAG-  8386
H2U.FR.96.12034         TG-------C-GTG-C-A-............................-GATCTTGGCCTTGGCAGATA...A-AT-TATA-ATTTCCTGA-CCA--AAC-G-TTCG--TGTTGACTT-GCTA-ACAG-  8497

                                                                            Tat end \
MAC.239_M33262          TG----AG-C-GTG-CAA-............................-G-TCCTGGCCTTGGCAGATA...GAAT-TAT--ATTTCCTGA-CC---AAC-G-TACG--T-TTGACTT-GCTA--CAG-  8958
Env                     __G__E__G__G__G__N__.__.__.__.__.__.__.__.__.__S__S__W__P__W__Q__I__.__E__Y__I__H__F__L__I__R__Q__L__I__R__L__L__T__W__L__F__S  gp41
Tat                     V__E__K__A__V__A__T__.__.__.__.__.__.__.__.__.__A__P__G__L__G__R__*                                                                  Tat
Rev                     _W__R__R__R__W__Q__.__.__.__.__.__.__.__.__.__Q__L__L__A__L__A__D__.__R__I__Y__S__F__P__D__P__P__T__D__T__P__L__D__L__A__I__Q_   Rev

SMM.SL.92.SL92B         TG--A--G---GTG---A-............................-GATTGTGGCCTTGGCAGATA...GAAT-TAT--ATTTCCTGA-CC---AGC-G-TACGGATATTGACTT-GCTA-ACAA-  8406
SMM.US.x.PGM53          TG---------GTG-CAA-..............................AGATCTTGGCCTTGGCAGATAGAAT-TAT--ATTTCCTGA-CC---AGC-G-TTCG--T-TTGACTT-GCTG--CAA-  8904
MNE.US.-.MNE027         TG-------C-GTG-CAA-............................-GTTCCTGGCCTTGGCAGATA...GA-T-TAT--ATTTCCTGA-CC---AAC-G-TACG--T-TTGACTT-GCTA--CAG-  8435
RCM.NG.x.NG411          TG----AG-C-GTG--CA-..............................AAATATTACAACTGGCAGAGAG-AT-TTT-TATATCCTGA-ACT--G-CCG-TAG-GAT--TT-GC--GAGGC-CTAT  8268
RCM.GAB1_AF382829       AGA--GTG-CA---AG-A--ACAAT-GGC---GA-A--AT-GCTGGA--CAACTGATCCACCCGTTGAGCAGGATTTGGACG-A--TATCG-A-ATTTG-AGA----G-AG-TCAA-A-TCTT-CAGAG-  8257
DRL.FAO_AY159321        C----G-GAGTGG...AA---CA---C--AACC-TCC.........AAGCCAGGATCCTCCAAAGCTTGGAGCGAA--AACAG-AG---C-TG--TGAAGGAATCGAGAGGATA-A-CT-GCTGAAGAA-  8399
MND-1.F17               CTT---ACT--GA...GAGCACAAC-CC-A----TTGAAGGAA-A-TCCT-AAACAGATCCTTGATAGAGGACCT-AC-AGCT-TGCC--GGA---A-CAA-TT-GCTTTGG-TGAAA-ATCTGAG-GCA  8159
MND-2.5440_AY159322     CTT---A-A-C-G...AA---CAAG-C--A--G-TCC.........AGACAAGAATCTAGGAAGCCTTGGAAGCCA--GC-AG-CA-----TG-CTGAAAAGGTCGACG-TTTA-A-TT-GCTGAAGAA-  8338
MND-2.14CG_AF328295     CTT---AGA-C-G...AA---CAAG-C--A--G-TCC.........AAG-AAGAATTTGGCAGACCTTGGAGGCCA--GC-AA-AA---ATTG--TGAAGGAGTCAACAATTTA-A-TT-GCTGAAGAAT  8631
MND-2.MND2_AF367411     CTC---A-A-C-G...AGT-CCAA--C--AC-G-TCC.........AAAG-AGAATTTGGGAAGCCTTGGAGCAAA--GC-AA-CA-AGATTG-CTGAAGACTTCAAGGGG-TA-G--T-GCTGAAGAAT  8702
DEB.CM.CM40             AGAC-G---G-C---G-GG-CAGC---GCC--GA----T---CAGT--C---.....................-----G-----C-AACAA---CTGACG-GGCT------GAC---TC-CA---GCA-T  8271
DEB.CM.CM5              AG-C-T--AGAG---G-GG-CGAC--CGCCC-GA----T----AGG--C---.....................--------CA-CTCACAG--TCTGACG-GGCT------AAC------GA---GCGC-  8192
MON.CM.L1               C-ACTC----TCGCC-G-ATA--TGAA--GATTTTTTA--A-TATCTGGGAAGATCTCAGGAACCTTGTCCTC---ACAT-C-A-ATCTTGAAAGA-TCAG-ATTAGTGAT-TA-CG--TCCT-CAGAGA  8536
MON.NG1                 C--GA----GTCR-C-GT-TA-GCGAG--GATTTTT-A--A-TATTTGG-CAGACTGCCGAGACCTAATAGTC---AC-TTC-A--TATTGAG-GA-TC-GG-TTGACGAT-TA-CG---CCT-CAGAGA  7161
TAL.CM.8023             C--GAG---GTGG---A-ATCG-CTCA--GATT-TTTG--A-TATCTGGGAGGACCTCAACAGCCTCCTCAGC-ACAGTT-C-A-ACCTTGAGAAA-T-AG-TT-GCTGAGTT-GGACCTCCTGAAG-G-  7953
TAL.CM.266              C--CT---AGTGG-ACAGATCG-A-GC--GATT-TTGAG-A-TATTTGGGAAGATCTGACAGGTCTGATCAGA-ACAT-T-C-A-ACCTTAGCCAA-T-AGCTT-GCTGATTT-GGACCTT-CGAA-GCT  8403
MUS.CM.1085             C---------TC-GTGAGATA--TGAGC-GATT-TTC--TC-CTTCTGGGA-GACCTCCGCAACCTAGTGATT---AG-T-C-A-ACCTTAA-AA--TCAGC-TC-AC--TTT-G-G-CT-CTTCAGAGA  8507
SAB.SN.-.SAB1C          TG-GCTA-ACAG-GACAG--GGTC--ACACTTGGCAGAGAGAATTCT-G-GACACCTGTGTCACAGATTGATCACTTGGCTCAG-AATTTGACCA-TTGG---TCGACAAT-TT-AGC-ACCT-CAT-G-  8868
GRV.ET.x.GRI_677        AG-T-GAGA-AG---CAG-............................-G-AAATTGATAAGATTGCAGGAAGA-TC-TCAACA-CTTCGAG-AT-AA-AA-T-G-GG-T-AA---C-AGAG--GCAG-  8520
TAN.UG.x.TAN1           AGATTC---CAG---CAGA............................GA-AAGTCGCCGCCCTCAGTGAAAGAATCTTTAT---CC-A-AGAG-AGGGA--CAAGCGG-GGAGAGAGCTTGGCGG-AG  8621
VER.DE.-.AGM3           C--CTC-GAGCCTT-GCAG-AAGA--CT-GC-CA-C-GAG-G-AAGAGC-A-TGGTGCAAGCGATTGACCAAT---TG-TCGA-CA-C---A-C--G--A-A--A-AGTTG--TGACCCTCCTAG-T-AT  8216
VER.KE.-.9063           C--CTCAGATTCTT-GCAG-AAG---CT-GA-CGA--CAAAA-AG-GGC-GTTGGAGCAAACGCTTGACCAAT---CT-TCGAC-TTC---ACC--G--A-A--A-AC-TT--TGACCCTCCTAA-T-AT  8718
VER.KE.-.AGM155         C-A-TCT-CGC-TTAGCAG-AAGA---T-G-GGAC--GA-AA-AGGACC-GTTGGTGCAGGCAATTGACCAAT---TT-TTGACA-TC---A-C--GT-ACA--A-AG-TG--TGACCCTCCTTC-CAAG  8695
VER.KE.-.TYO1           C-A-TTAGAGTCTT-G-AG-AAGAC-CCAA--G-A-G-GTA--CAGT-G-C-GCCTGGCTGACGAGGCTCAAC.................ACT--G--A-A--A-AGTTG--TGACCCTCCT-A-T-AG  8172
SUN.GA.98.L14           AGAGA-AGA-AGA-TC-A---CAAC----A-C-GTCCAAGAAA-A-TC-T-GCCACCACCTTGGACAGTAGAC--------G-CTTT---AGAC-CT--G--AGTGACT-TT-T-AA-T-GCTGAA-GCT  8884
LST.CD.88.447           C-A---A-ACAT-...-A---CAAC----AAC-GTACAAGAAA-A-TC-TC-ACTCCACCGTGGAACGTGGAC---AG---G-CTTTACAAGAC-CA--G--TGTGACT-T--TGAA-T-GAT-AAGGC-  7747
LST.CD.88.485           C-A---AGACAT-...-A---CAAC----AAC-GT-CAAGAAA-A-TC-TC-ATTCCACCGTGGAACGTGGAC---AG---G-CTTTACAGGAC-CA--G--TGTGACT-T--TGAA-T-GAT-AAGGCA  7744
LST.CD.88.524           C-----A-ACAT-...-A---CAAC----A-C-GT-CAAGAAA-A-TC-T--AGGCCACCGTGGAACGTGGAC---A----G-CTTTACAGGAC-CA--G--TGTGACT-T--TGAA-T-GCTGAAGGCG  7750
LST.KE.-.lho7           T-C---G-CTAT-...-A---CAAC----A-C-GT-CAAGAAA-A-TC-TC-AGGCCACTGTGGAACGTGGAC---A----G-CTTTACAAGAC-CG--G--AGTGACT-T--TGAA-T-GCT-AAGGAA  8822
SYK.KE.-.SYK173         C-----AG--CT--ACGT-TC-ACC--CAGC-GA-A---T---AGG-AATCTTTGGAAGCAGGACAGCAGCTT---AGGACAG-CT-C---AGC-T-AGA-CTCTGATTAGA-AGC-T-CTAT-ACCTGG  8543
SYK.KE.KE51_AY523867    CGTCTCT-CGTT---CAGAGAATTT--GAGAC-ATCGCT-GCA-A-GCGCAGAAACTTTGGAACAGAGTTTCGGA-AG-TGCAGATCACTGA-TA-AGGAC-TGTAATTGC-T-GGGATTCAT-AGTTAT  8257
COL.CM.-.CGU1           -CA---A-ACAG---CAG--GATTT-GGCGATTGATGGAAGA-TGGTGG-G-..........................................................................  7911
```

```
H1B.FR.83.HXB2              TTGATTGTAACGAGGATTGTGGAACTTCTGGGA...................................................CGCAGG...GGGTGG...............................  8594
Env                         _L__I__V__T__R__I__V__E__L__L__G__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__R__.__G__W__.__.__.__.__.__.__.__.__.__.__  gp41
Rev                         L__D__C__N__E__D__C__G__T__S__G__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__Q__.__G__V__.__.__.__.__.__.__.__.__.__.__.   Rev

H1D.CD.84.84ZR085           --A----C-G-----------------------...................................................------...------...............................  8115
H1F1.BE.93.VI850            --A----C-G-------------CAGGGGCCTG...................................................A-G---...------...............................  7871
H1G.SE.93.SE6165            ---------G------CA---------------..............................CGCAGCAGCCTCAAGGGACTGA-ACT-...------...............................  8033
H1H.CF.90.056               --A------GT-----C---------------G...................................................A-A---...---A--...............................  7913
H1J.SE.93.SE7887            -------C-G------C-----G-AC-------...................................................-T----...------...............................  7901
H1K.CM.96.MP535             --A-----G--------CC-----AGGGGACTG...................................................A-GG--...--T---...............................  7754
H101_AE.TH.90.CM240         -------C-G------CG---------------..............................CACAGCAGTCTCAAGGGACTGA-AC--...------...............................  8168
H102_AG.NG.-.IBNG           -------C-G------C----------------..............................CACAACTGTCTCAAGGGACTGA-ACT-...------...............................  8121
H103_AB.RU.97.KAL153_2      -------C-G-----------------------...................................................------...------...............................  7771
H104_cpx.CY.94.CY032        --A------G------C---------------G...................................................ATA---...------...............................  7967
H1N.CM.95.YBF30             -----AC-G-G-----C-C-------C------..............................CAGAGTCTCAGCAGGGGACTG-AACTA...CT-AAT...............................  8166
H1O.BE.87.ANT70             AGCTA-C-G-G-CTTGGAC--TGGA-C--A--G..............................CAGAAGATAATTAATGTT.......................................TGCAGAATT.  8678
H1O.CM.91.MVP5180           -CAGGGA-CCG----C-GA-C--CTAC------..............................CTGGGACTGTGGATCCTGGGA-AA-A-...ACAATT...............GAAGCTTGTAGACTT.  8704
CPZ.CM.04.EK505             -CA--CC-G-G-----CCT-----------A--...........................CAG........................-ACATCCACAA-..................GGACTGCAACTAC  8096
CPZ.CM.04.LB7               --A--A-CG-G-----CAT----GA--G-----...........................CAA........................TACACTCT-AA-..................GGACTGAGACTGG  8121
CPZ.CM.04.MB66              -----A--GCT-----CAT-------GG-----...........................CAG........................-CCCTCCTCAA-..................GGACTGAGACGGG  8234
CPZ.MT145                   -CTC-GA-TTG---AC-AC----GA-A------...........................CAA........................TACAGCCTCA--..................GGAGTACAACAAA  8122
CPZ.CD.-.ANT                ----ACC-TTG-GCAG-CC-CC---AC--C.......................................................................................TGCAGAATTACTT  8051
CPZ.CM.-.CAM3               -----AC-GGG-----CAT-----AAC------..............................CAGAGCCTCAACAAGGGACTG-AACA-...CT-A--...............................  8031
CPZ.CM.98.CAM5              -CA--CC-GTG-----CA---C--A--------..............................CAGAGCATCAACAAGGGACTG-AACA-...TT-AA-...............................  8337
CPZ.GAB_X52154              -GC-A---GTG---ACAAC-CA-GACGT-----..........................................CATCTAATA-TTCAC...A-CCTC...............................  8643
CPZ.GAB2                    -C-G-CC-TTG---CC-GC------AAT-----..............................CAACAGGTCCTGAGAGGACTGA-ACTA...CTCA--...............................  8002
CPZ.TZ.-.TAN1               -G---CC-CTG--TC-GCC--C--GCA---AA-..............................CAGGGGATAATCAGCTTGGCA...CACAGCCTAGTA..................ATAGTGCATAGAA  8284
CPZ.US.85.CPZUS             --A---C-G-G-----CA--AC--A-C------..............................CAGAACATCAACAAGGGACTG...CAA...CT--T-...............................  8620
GSN.99CM166_AF468659        -G--C-C-CCGTG----C............--G...................................................AAGTTCCTGT--GAACAGCTGCAGAAAGCAATCCAGAAAGCAATCC  8550
GSN.99CM71_AF468658         -G--C-C-CCA------C............--C...................................................AAGTTC...CT----.........GAGCAGCTGCAGAAGGCGTTCC  8550
H2A.ALI_AF082339            A-CTGCAGGGACTTAC-ATCCAGGA-CTCCCCG............ATCCTCCAACCGATCTTCCGGAGTCTTCAGAGAGCGCTGACA-CA...ATCA--...............................  9081
H2A.BEN_M30502              A-CTGCAG-GACTTAC-ATCCA-GAACTCCCCG........................ACCCGCCGACTGATCTCCCAGAGTCTAACAGCA...ATCA--...............................  9085
H2A.ST_M31113               A-CTGCAGGGACTTAC-ATCCAGGAGCT-CCAG............ACCCTCCAACTAATCTCCCAGAGTCTTCGGAGAGCATTGACAGCA...-TCA-A...............................  8527
H2B.D205_X61240             GGCTGCAG---CTTAC-GT--A-GACCT-CCA-....................................ATCCTCCACCAAATCTCA-CC...AATCTC...............................  9049
H2B.EHO_U27200              GGCTGCAG---CTTAT-ATCAA-GACCT-CCA-.............................................ACCCTC-AACCA...-TACTC...............................  9026
H2G.ABT96_AF208027          AACTGCAG-G-CTTAC-ATCCAGGA-CY-CCAG.............................................ACTCTC-AACC-...CT--TCCAAAGA...ATCTCCAGGACCCTCCAAGCAA  8465
H2U.FR.96.12034             AGCTGCAGGGACTT-C-ATCCAGGA--TGCCAG............AGCCTCCAGCCGTTGTTCCAGAGC........................ATCA-A...............................  8560

MAC.239_M33262              AACTGCAG---CTT-C-ATC-AG-G-ATACCAG............ATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTA-AG---...ATTC-A...............................  9042
Env                         _N__C__R__T__L__L__S__R__V__Y__Q__.__.__.__._I__L__Q__P__I__L__Q__R__L__S__A__T__L__Q__R__.__I__R__.__.__.__.__.__.__.__.__.__.__  gp41
Rev                         Q__L__Q__N__L__A__I__E__S__I__P__.__.__.__.__D__P__P__T__N__T__P__E__A__L__C__D__P__T__E__.__D__S__.__.__.__.__.__.__.__.__.__.__.   Rev

SMM.SL.92.SL92B             AATT-GACTCGCTTAGCATCCAG-GCCTACCAG............AACCTCCAACAACTGTGCCAGAGACTCTCAGAGATCAGT-AGCC-...ATTA-A...............................  8490
SMM.US.x.PGM53              AACTGCAGG--CTATC-CT--AG-A-ATACCAGATCCTCCAACCAGTGTTCCAGAGGCTCTTACAGAGAATCCGAGAAGTCGTCA-G-CTGAGATA-CC...............................  9003
MNE.US.-.MNE027             AACTGCAG---CTT-C-ATC-AG-GCATACCAG............ATCCTCCAACCAATATTCCAGAGATTCTCCACGACCCTA-AG--A...-TCC-A...............................  8519
RCM.NG.x.NG411              CAA--CTGC-GC-ACTCAC-CTC-G-GA-CTACCAGAGCCTCCAGTCAATCCCTTCTCGAATCCTTCTTCATTTGCGGTGGATAG-AGCTAAGCTACAA...............................  8367
RCM.GAB1_AF382829           C-CCGGTGG-TTTTAGC-AA-AT--AATAT--G...................................................T-GCAA........................................  8296
DRL.FAO_AY159321            G---AA-C-GTT.....................................................................CTTGAATATGGGT---CA........................GAACTCC  8436
MND-1.F17                   GCA----A-TATG--T-CC-C-C-G-C-AA-A-GCGATCAGAAGCTTGGGTTCGCTACTCGTCAGCTTTGCGGCAATTAGTTGGA-GGCCGGTTACACC...............................  8258
MND-2.5440_AY159322         C-CCAA-C-GTA.....................................................................ATTGAATATGGGT--CAA...............................  8368
MND-2.14CG_AF328295         C--CAG-C-GTA.....................................................................ATTGAATATGGGT--CAA...............................  8661
MND-2.MND2_AF367411         C-ACAA-C-GTA.....................................................................ATTGAATATGGGT--CAA...............................  8732
DEB.CM.CM40                 -G-C-GA-T-GA--AGGCT-T---T-A---CA-......GAAGGAATATATAGACTGTCATTTGCTCTTTGGCAAAGACTGAGA-T-TT-GGG-AAC-C...............................  8364
DEB.CM.CM5                  -G-C-GA-T-GA--AGGCT-T--G--C---CA-......GAAACAATATATAGACTTTCATTTGCATTTTGGCAAAGACTGAGA-T-TT-GGG-AACAT...............................  8285
MON.CM.L1                   G--TCCCAG-G-CT-CCACCCCTC--G-ACAT-CGCCTGCTCCAACTGTGGGAAAGCCTTAGACGCTTGCTAGCCTACTGCCAGTATG--ATCCAAGAA...............................  8635
MON.NG1                     G--CGCAGCCACCT---CCCTCTT--C-GC-ACCTGTGCCGACAGCTTCGGGAAGCCAGTAGCCGCCTTCTTGCCTACCTGCAGTATG--CTCCAAGA-...............................  7260
TAL.CM.8023                 C---AA-CG--C-CCCAACA--C-GCG-AA-CCACATACGCCTACCTGAGAATTAAATTCACAGCAGCAGTAGCAACAATCCAGTATG--TTCA-TGAA...............................  8052
TAL.CM.266                  GC--AA-CC-A---CCAACA--T-GCG-AA-CCACATACGCCTTCCTACGACTCCACCTGACAGCAGCAGCAGCGACTCTCCAGTATG--TTCA-TGAA...............................  8502
MUS.CM.1085                 GCAG-CCACCT--TACAGAA--GGTCAA-AA--TTGTGGTTCACGCTCAGGGCACTGATTGAGCCTCTCAGGGAAGCCTTGCAAA-AGCCCTC-CT--C...............................  8606
SAB.SN.-.SAB1C              ..............................................................................TGCCTC-AGGAC...ATCCAA...............................  8886
GRV.ET.x.GRI_677            ----GAA---G--CTTGGTGTT-CAACA-CT-CCTGACCCTCCTCATATTCATCAGGACAGCAGTGGGATACCTGCAGTATGGG-T-CA-CAACTCCAA...............................  8619
TAN.UG.x.TAN1               CATT-GACC-ATT--TGCT--AC-A-CAGCA-TTGGTTATTGAGACTTTACCAGATCCTCCGCAGGAGCCTCACGACTCTTCTT-AACT-CTTA-ACA-...............................  8720
VER.DE.-.AGM3               ..................................................................................................................................  8216
VER.KE.-.9063               ..................................................................................................................................  8718
VER.KE.-.AGM155             ..................................................................................................................................  8695
VER.KE.-.TYO1               ..................................................................................................................................  8172
SUN.GA.98.L14               -G-GGAA--GTCCTAGCACAAA-CA-CTACCACCTGCTAAGTTTCTTGTGGCACCTACTTACGACTTCCTTCCATCATGGGCAA---CTT...T--CA-...............................  8980
LST.CD.88.447               --TGGAA-TCTTCTCC-CAGCCTTG-GTG-CAGTTCCTTTCGTGGCTAGGCCATTTAGTGATCCTTTTCTTCCAACATGGGCAA-ATCTT...C--AAA...............................  7843
LST.CD.88.485               --TGGAA-TCTTCTCC-CAGCCTTG-GTG-CAGTTCCTTTCGTGGCTAGGCCACTTAGTGATCCTTTTCTTCCAACATGGGCAA-ATCTT...T--AAA...............................  7840
LST.CD.88.524               -G-GGAA--CT-CTCC-CACCCTCG-GTG-CA-TTCCTCACTTGGCTGGGACATATAGTGATCCTTTCTTACCACCATGGGCAA-ATCTT...T--CAA...............................  7846
LST.KE.-.lho7               GG-GGAA--CTCCTCC-ATCACT-G-GTG-CAGAGCCTCAGTTGGCTGTGGCATTTGTTGATCCTTTTCTTCCAAAATGGGCAA---CTT...T--CA-...............................  8918
SYK.KE.-.SYK173             ..................................................................................................................................  8543
SYK.KE.KE51_AY523867        GG-G-CAC-GA-CTC-AA-AA-C-GCCA-CTCGCTGGGACGAGAAGTGGCGCACCAGATGGTGGCCATTTGGCAAGCCCTACTCGC-TATGCCA-AC-A...............................  8356
COL.CM.-.CGU1               .........................................................AGCTTGTTGCGTCATTGCAGCAATGGGAT-CTT...ACTGTC...............................  7950
```

```
                                                                                    Rev exon 2 end \
H1B.FR.83.HXB2          .....GAAGCCCTCAAATATTGG............TGGAATCTCCTACAGTATTGGAGTCAGGAACTAAAGAATAGTGCTGTTAGCTTGCTCAATGCCACAGCCATAGCAGTAGCTGAGGGGACAGATAG 8707
Env                     .__.__E__A__L__K__Y__W__.__.__.__.__W__N__L__L__Q__Y__W__S__Q__E__L__K__N__S__A__V__S__L__L__N__A__T__A__I__A__V__A__E__G__T__D__R gp41
Rev                     __.__G__S__P__Q__I__L__.__.__.__.__V__E__S__P__T__V__L__E__S__G__T__K__E__*                                                      Rev

H1D.CD.84.84ZR085       .....---------------CT-............-----------G--A----------GA-----G------------A--------G-TG----A-----A-------------------------- 8228
H1F1.BE.93.VI850        .....-----T---------CT-............G--------ACG-G-------------------------------A-C---C--T-T---A-------A----T---------A----------- 7984
H1G.SE.93.SE6165        .....----GT--T-------T-............--------T--GTT-------G-CAG-------------------A----T-----TG--A-AGT---A------------A-TT---------- 8146
H1H.CF.90.056           .....---------------CTC............--------T-----A--C---G-A---------------------A--GAT-----T--CA-------A-----------------A------G- 8026
H1J.SE.93.SE7887        .....--GAT----------CT-............GT---C--TG--TG-------G-G---------------------A----T-----T---A-------A--------------A--A-------- 8014
H1K.CM.96.MP535         .....---------------CT-............-----C--TG----------------------------------CA----------T---A-------A-------------GA--A-------- 7867
H101_AE.TH.90.CM240     .....----G----------CT-............G-------T--GTTA------G-C-----------A-T-------A--TCT-----TG----T-----A-------C---G-G-T-------C-- 8281
H102_AG.NG.-.IBNG       .....-G-------------CT-............--------TA--TCA--C---GT----------------------A---AT-----T---A-A-T---A----T-------A-CT---------- 8234
H103_AB.RU.97.KAL153_2  .....------------------............----------------------T--------------G-------A---ATC--A-AGG-A---T---A-------------G-T---------- 7884
H104_cpx.CY.94.CY032    .....-----------G---CT-............-----CT----G-T-------G-A-----G---------------A---AT---T-T---A-------A-----------------A-------- 8080
H1N.CM.95.YBF30         .....---CT-AGA-C-C-C-T-............---GGAA-A--TGCA------G-AA-A--GT---G-G--------A-C--------T---A-A-----T--T-T------A--A--A-------- 8279
H1O.BE.87.ANT70         ....................--T............GCAGC-G-AAC---A--C---CTA--A---T-GC-----------ACA--------AG-CA-ACTT--AG-G--------CA-TT----T--CG- 8776
H1O.CM.91.MVP5180       ....................--T............G-AGC-G-AA-G--A------CTA--A---T-G--A---------ACA-A-C----TG--A-T-TT--AG-GT----T--CA-TT----T--CG- 8802
CPZ.CM.04.EK505         TTAAT---CT-AGA-T-----T-............---GGAA-AA--GCT--C---G-AAG---G-----A-T-------A-A-A------TG--A-T-----TG-T--------A--A----------- 8214
CPZ.CM.04.LB7           TGTGG--------ACT----CT-............CA-GGAA-T------------G-AA----G-----A-C------CAC---------TG--A-A-----A-----------------A-------- 8239
CPZ.CM.04.MB66          GACGG--------A-T-C-CCT-............A--GGCA-T-----A------G-A-----G-----A-C------CA-A--------TG-CA-A-----T-----------A--A--A-------- 8352
CPZ.MT145               TAGGTACTCT---AC-CC-ACT-............A-AGGGAC---G----------CAACA---A-C-----------CA-A-----AT-AG--A-T--C--T--T-------GA--A--A--T----- 8240
CPZ.CD.-.ANT            TCAGACTGTG-AA-C-TCTGGA-............AAC------AGCACCCTC---...ACAAT-A-C-GA-C-GAGATCA---AGAACA-TG-CAGACTT--T--TTGG----GG--AAAA-------- 8166
CPZ.CM.-.CAM3           .....A-CTT-AG--G---CCT-............---GGAG-GA--ACC--C---G-AAG---G---C-A-C------CA-A-----AT-AG----A--------T-----------A--A-----C-- 8144
CPZ.CM.98.CAM5          .....--CTTTAGT-C---CCTA............AA-GGAG-A---ACA--C---G-AAGA--GA--C-A-T-------ACA-----AT-AG--A-A-----A--T---------A-T--A-------- 8450
CPZ.GAB_X52154          .....AG-CTG--ACGGG-GA-ACTGTGCCTGCTGG--GGAA-TA-T--A------G-AA----G-----A-T-------A-A---C----TG----A-----T--T--------A--A--T-------- 8768
CPZ.GAB2                .....--GCTT--G-GCC-GCT-............AAAGGGA-AGG---A------CTG--A-------GA-C-------A------C---TG--A-T-----A--TAT------C--AA-A--T----C 8115
CPZ.TZ.-.TAN1           CTATCAT--TAGGAGTTAGACA-ATC............-T-GAGTGGAGCAG-AAT-C-T-T-CTAGCTTA-GAGT-TTGC-A-TACAAGC--TA-A--G-CTTGCTAACT-TA-A-G-T--TGGACAGA 8402
CPZ.US.85.CPZUS         ..AAT---CT-AGAGC-CGC--C............---GGGG-AA-CGCT--C---GCAAG---G----AGT--------ACA-----AT-AG--A-A-----A--------------A--A--T----- 8736
GSN.99CM166_AF468659    AGCAA-CGAGA-AGCTCAGGGAAGTAGCAGCCAGAGT--TAGC-TACATTAGC-ATG-GATCC--GA-CTCC-AGCA--A-CA-CAGGAA--CTC-A--GCCTTGCTATCT-CA-GTG-AACTGGACAGA 8680
GSN.99CM71_AF468658     AACAA-CGAGA-AGCTGGGGGAAGTGGCAGCCAGAGCAGCAGC-TATATTAGC-ATG-GATCC--GA-CTCC-AGCA--A--CTCAGGAA--CTC-A--GCCTTGCT-TCT-CA-ATG-AACTGGACAGA 8680
H2A.ALI_AF082339        .....--CTGG--G-G-CT-AAA............GCAGCCTA---G------G--T-CG--TGGA-CC-AG-AGCGTTCCGGGC---TGCA-GGATTG-GAGAGAGA-TC-TA-AA-CACCTGGAGAGA 9194
H2A.BEN_M30502          .....--CTGG--G-G-CT-AA-............GC-GCC-AA--G--A---G--T-CG--TGGA-CC-AG-AGC-TTCCAAGCA--CGCG-GGA-T--GAGAGAGA-TC-T--G-GC-C-TGG-GAT- 9198
H2A.ST_M31113           .....--CTGG--G-G--T-AAC............ACAGCCTA---G--A---G--G-CG--TGGA-CC-AG-AGCGTTCCGAGC---CGCG-GG--T--G-GAGAGA-TC-TA-AA-C-CCTGGAGAG- 8640
H2B.D205_X61240         .....C--C-A----GGCTCCC-............GTTGCGTA---C------G---TCAGCTGGT-CC-AG-AGCGCTCCGAGCGGCAGCA-GG--T--G-GAGA-A-TC-T--GAGT-CCGG---G-C 9162
H2B.EHO_U27200          .....C--C-G----GGCT-CCA............CCTGCATA---C-GA---G---TCAGCTGGT-CC-AG-AGCAATCCAAGCAGCAGC--GG--TG-G-GAGAGA-TC-T--GAGC-C-G-GAGG-C 9139
H2G.ABT96_AF208027      TTAGA---CAT---CGGCTAGAA............GCAGC-TATT-CAGC---G--TTC-GCTGG--CC-AG-AGCGTGCACCGCAGCAACT-GG--TG-GCAAGAGA-TC-TA-AAGCAC-TGGAGAGC 8583
H2U.FR.96.12034         .....---CGT--GC-CCT-GAA............ATAGC-TA------A---G--T-G---T-TT-C--AG-AGCATTCCAAGCG--CGGG--A--TG-GCGAGAGA-TC-TAGCCGCAC-GGGAGAGA 8673

                              Rev exon 2 end \                             / Nef start
MAC.239_M33262          .....----T-----GGAC-GAA............CT--CCTA------A---G--T-GAGCT-TT-CC-TG-GGCG-TCCAGGC-G-CTGG-GAT-TG-GA-AGAGA-TC-T--G-GC-C-TGG-GAGA 9155
Env                     .__.__E__V__L__R__T__E__.__.__.__.__L__T__Y__L__Q__Y__G__W__S__Y__F__H__E__A__V__Q__A__V__W__R__S__A__T__E__T__L__A__G__A__W__G__D gp41
Rev                     __.__R__S__P__Q__D__*                              _M__G__G__A__I__S__M__R__R__S__R__P__S__G__D__L__R__Q__R__L__L__R__A__R__G__E_  Nef

SMM.SL.92.SL92B         .....--GCTTG---G-AGAGAA............GCAGGCTATA-C-GT---G--T-GA-TT-CT-C-TCG-AGCCTGCCAGGAAGCATGG-GAT-TG-GCAAGA---TA-T-TC-GC-C-TGG-GACT 8603
SMM.US.x.PGM53          .....T-CCTG-AAT-T....................................G--T-C-GCTGGT-CC-AG-AGCA-TCCAGGCAGC-TGG-GG--TCGG-GAGA-A-TC-T--AAGC-C-TGGAGAGA 9092
MNE.US.-.MNE027         .....----T-----GGAC-GAA............CTA-CCTA------A---G--T-GAGCT-CT-CC-AG-AGCG-TCCAAGT-GCCTGG-GAT-TG-GA-AGAGA-TC-T--G-GC-C-TGG-GAGA 8632
RCM.NG.x.NG411          .....--T-GGTGGC--G-G.....................T--AA-GGC-TCA-C--CTG-CTTGCGG---TGGCGAGGAC--ATGCTTA-T-CA--TGGAGAGG-CT-TGC--G-TTT-C-G---CCC 8471
RCM.GAB1_AF382829       ..............................................................--GT-C--AG-ATTCAGCAGCT-G--TGCTG-GATGG-GCTTCA-AAT-CTTACT-CACCTGGAGAG- 8364
DRL.FAO_AY159321        AAGAA-CG-GAAGAGGCATA--C................................................-CGCT-CTACAAGAGG-TGCGC-GAGG-T-TGGCGC-A--GGAGGC-ACTTGG-CTATC 8518
MND-1.F17               ..............................................................................................................-G-TGGCTACAAG--A---- 8278
MND-2.5440_AY159322     .....--G...------GCAGC-..............................G-AGCAGCATT-TATC-AGTATTACAA-GCTTTGC--AG-GGCTATGGAG--G--GGTACCAACTT--ACT-TCAG- 8460
MND-2.14CG_AF328295     .....--G...------GCAGC-..............................G-AGCAAG-AT-TGCC-AGTATTACAAAGCTTTGC--AG-GGCTATGGAG-GG--GGCACCAACTT--ACT-TCAGT 8753
MND-2.MND2_AF367411     .....--G...------ACAGCA..............................G-AG-AA-AAT-T-T--AGTACTACAA-GCTATGC--AGGGGCTATGGAG--G--GGCACCAATG---ACT-TCA-C 8824
DEB.CM.CM40             ...........T-T-G-GTGCTT............G--GGGTA-T--------G--T-GA-A---T-C--AG---CCATCACGT-GGCAGGTGGG-AGCTTTGGCAGTGGC-CCAG-G-ATTG-T--G-T 8471
DEB.CM.CM5              ...........T-A-G-GTGGTC............---GG-TATTGT--A---G--T-GA-A---T-C--AG-C-CCATCA--T-GGCAGGGC-A-AA-TCTGGGAGTGGC-CCAA-G-ATTG-TC-GGT 8392
MON.CM.L1               .................................................................--CC-AGCAGCA-TCACCTCTC-C---G-C---CTT--T-GGTTCAC-ATC-TCT----G---GC 8700
MON.NG1                 .................................................................Y-CC-ARCAGCCTGCACAG-AGCCA-TG-----CTC--ACGGTTCAC--TCATCT-------CGC 7325
TAL.CM.8023             .................................................................A-CC-AGC-GC-TGG-GAGA--GCG---CA--GCTT--A-GGTTCACCTGGCGCT----GTC-G- 8117
TAL.CM.266              .................................................................A-CC-AGC-GC-TGGCGAGA-GGCA---CA--GCTT--A-GGTTCAC-TGGTG-T----GA--G- 8567
MUS.CM.1085             ............................................T-----C--G---TC--A--G--CC-A-CAGCCAGCACAGCAG-C--TG--CAGCTC----CCTTCACCTGGA-CT----G--A-C 8692
SAB.SN.-.SAB1C          .....C-GAGAAC--G-C-GCTA.........ACAGCTC----TGA-T-TGGG---..----A--G--C-AGCAGCA--A-CGTCAC-CACTGTG--TGTG-T-CA---T-CGA-AAGC-CC-GT---TC 8999
GRV.ET.x.GRI_677        .....--G--AGCA-C-GGG.....................--TGCT--AGC-CT-GCGAG--CTGCG-G-G-AGCCTGG-GC--AC--.......................................G- 8684
TAN.UG.x.TAN1           .....--GTG--AAT-CAT-.....................-AGTATGG--GGCA-CAATTCA--GAGGGAGCAGCAAGGTC-TTTGA-GCTTTG--G-GC--TGCGCA-AGC--CAGTC-T--GCTAT- 8824
VER.DE.-.AGM3           ...........--T-GGAGCGCT.........TTCCA-T-CA-A-A-T-TGGGCTT..GG------C--AGCCGCA--ACAAGAAGCAG-TGTC--TTTG--ACGCCTT-C-CAAA-C-C-GGCT-CCA 8323
VER.KE.-.9063           ...........--T-G-GGCGCT.........TGGCA-T-CA---A-T-TGGGCTT..GG------C--A-CCGCA--ACAAGAAGCA-GTCTC--TGTG--ACGCTTT-C-CAAA-C-CAGGTC-CCA 8825
VER.KE.-.AGM155         ...........--T-G-AGCGCC.........TGGCA-T-CT-A-A-T-TGGGCTT..GG---G--C--AGCCGCA--ACAAGAAGCAG-T--CCATCTG--G-GCTTT-C-CGCA-C-C-G--C-CCA 8802
VER.KE.-.TYO1           ...........--T-GGA-AGCT.........TTTCAAT-C--G-A-T-TGGGCTC..GCA-----C--A-CCG-C--ACAAGAAA-A---C-AA-TCTG--AGGC-TT-C-CAAA-C-CATGTC-CCA 8279
SUN.GA.98.L14           .....ACCCT-AGAGGT-GGCT-............G-TTCG-A---TATTC-GGCT-CGAGCAGGA-C-G---CGCGTGCAGA-C-AGCAGGG-GAGAGT-T-ATCGT-TCAGAAG-CAA-A-GTAGA-C 9093
LST.CD.88.447           .....-CCTG-AG--G-CGGGT-............GT-G-AAA-GCT---A-GATTGCGAGCTGGAGCTG---C-AGATCAGA--GAACAGA-GA-A--T-T-AGC---T.........CATG--AGA-A 7947
LST.CD.88.485           .....-CCAG-AG-GG--GGAT-............GT-G-AAA-GCT---A-GATTGCG-GCTGGAGCTG---C-AGATCAGA--GAACAGA-GA-A--T-T-AGC---T.........CATG-GAGAGA 7944
LST.CD.88.524           .....ACCCT-AG--G--TCCT-............GT-G-GAGTGCT---A-AACTGCGAGCTGGA-C-G---C-AGAGCAGA--GAACAGAGGA-AGCT-T-A-A-A-T.........AATGG-ACA-A 7950
LST.KE.-.lho7           .....ACCAG-AG--G--GGAT-............GT-G-AAA-GCT---A-GATTCAGAGCTGG--C-G-G-G-AATGCAGA--GAACAGAGGACAGTT-T-G-GCA-T.........-AT-G-A-A-A 9022
SYK.KE.-.SYK173         ......................................GGAT--A-CAGC---G--TTCA-C--G--C--A-TCGCA--A-C-TCGC-CGGA-GG-AA-T-CG-GATTGG--G--G-CTATTTGGC-AGC 8635
SYK.KE.KE51_AY523867    .................................................................................................................................. 8356
COL.CM.-.CGU1           .....CTGC-T-AAC-TCCGCCA...ACAACCTCAGAC--GAGGAGGAGTGC----..---ACT--TCG-T--CTA.........---G-ACCCAGGGGTC-A-GCCTTCAGAT-CTT-A----CT-CA 8060
```

```
                                                                                                                   Env end \
H1B.FR.83.HXB2              GGTTATAGAAGTAGTACAA...........GGAGCTTGTAGAGCTATTCGCCACATACCTAGAAGAATAAGACAGGGCTTGGAAAGGATTTTGCTATAA.G............................. 8796
Env                         __V__I__E__V__V__Q__.__.__.__.__G__A__C__R__A__I__R__H__I__P__R__R__I__R__Q__G__L__E__R__I__L__L__*                                gp41

H1D.CD.84.84ZR085           -A-------CA-----AGG...........A---------A----G-G-TT--------C-C---------------T--A------GC----------.A............................ 8317
H1F1.BE.93.VI850            AA----------TT-G---...........A-----G--------G---T-A--------------------------GCA------GC------G---.A............................ 8073
H1G.SE.93.SE6165            --------------C----...........A----------------C-T-A----------C---G--------A-----------AGC----------.A............................ 8235
H1H.CF.90.056               -A------T-A----G---...........A-------G----------T-----------------------------T-----A-GC---------TA............................ 8116
H1J.SE.93.SE7887            AA-C------A---C----...........A------T-----------TT-------------------------T--A-----AGC----------.A............................ 8103
H1K.CM.96.MP535             AA--------A---G----...........A------T-------T-A-TT-------------------------TC-C-----AGC---A------.A............................ 7956
H101_AE.TH.90.CM240         --------------C----...........--------G----C----T-----------------C-----------A-------C----------.C............................ 8370
H102_AG.NG.-.IBNG           A-C-------A---G----...........A---TGG-------------A---------------C-----------C------GC----------.C............................ 8323
H103_AB.RU.97.KAL153_2      ----------A---G----...........A--TT-----------G--TA----------G-----C---------GCA---A-GC-----A----.C............................ 7973
H104_cpx.CY.94.CY032        -A---------C------G...........A-----------------T--A----------------C---------C-T-----AGC------T---.A............................ 8169
H1N.CM.95.YBF30             -A--------T---C----...........A--ATAG-A--G-GA--ATTA-----------------C-----A---C-A-----AGCAC--A-----.-............................ 8368
H1O.BE.87.ANT70             CA-A--C-C--GGA-----...........A--ATAG-A-C--GA-----TA----C--A--G-----T------------A-----A-G---AT-G---.C............................ 8865
H1O.CM.91.MVP5180           CA-C--CTT--GTC-----...........A--ATAG-ACA--GAT-C-TT-----C--A--------T-----A--TGCA-----A--C--AG-G---.C............................ 8891
CPZ.CM.04.EK505             AA-C-----GC----G---...........A--ATAG-AC-G-GA--A-TA------------G-----C-----A--TC-A-----AGCAC-TT-----A............................. 8303
CPZ.CM.04.LB7               AA-C------A--------...........A--AT-G------GA--ATTA--T-----------G--T-----A--T--C------GC--AT-----A............................. 8328
CPZ.CM.04.MB66              AA--------A---C---G...........A--TT-G------GC--A-TGA----C------------C-----A---C-T-----AGCA--AT-----A............................. 8441
CPZ.MT145                   AA-CT------C-A--AC-...........C-CTTGG-C--G-GC---TTA--T--------------T-----A-----A-----AGC------T---G............................. 8329
CPZ.CD.-.ANT                CA-AC-TCT--CTC-C---...........ACTATAGTC---ATC--AA-GG-AG-------GC-C--C-----A--G-------TTGCA--AAAT---.AC............................ 8256
CPZ.CM.-.CAM3               AA--C-------T-C----...........AT-ATAG------GC----TG--T-----A--------T---------T--A-----A-GC---------.C............................ 8233
CPZ.CM.98.CAM5              AA-------GT-T------...........AT--TAG-G----GC----TA--T--------G-----T-----A--T--A-----A-G----T------.C............................ 8539
CPZ.GAB_X52154              AA-A-------CTT-T---...........-TTA-ACT----AT----A-AA----C-----GC-C---------A--GC-------AGC---A--T---.-CACT......................GCA 8864
CPZ.GAB2                    AA-A----------C-ACT...........A--ATAG-G----GC--ATTG--T-----A--------C--G--A--T--A-----AGC---A---G--TAA............................ 8206
CPZ.TZ.-.TAN1               TT-A--CAT--A--G-GTGGTT........TACATAGCC--G-GA--CA-AA-T--T-----------T---------TC-----CTAGCC--AAAT---..A............................ 8494
CPZ.US.85.CPZUS             AA--------T--AC-AG-...........A-GCTC-T-CT--G----ATA--------A--------T--G--A----C-A--G--A-GC--AT----............................. 8823
GSN.99CM166_AF468659        --CCG--CTTCAC-CTTGTAGA........C---TA--GC---AAT---TGGCT--C--CC--C-C--C-----A--GGC---G-TAC-A--TAAC-G-.A............................ 8772
GSN.99CM71_AF468658         -CCCC-C-TTCAGACTGTTGGA........C---TA--GC---AAT-CTTGGCT--C--CC--C-C--C---------GGC---G-TAC-AC-TAAC-G-.A............................ 8772
H2A.ALI_AF082339            CT-GTGG-GG-C---G--G...........T-G-TCG-G--GAGG--A-T-GCAG-C--A--G--G--C--G-----GGCA---TTGCCC-C--G-G-............................. 9282
H2A.BEN_M30502              -T-ATGG----C--CG-G-...........C-CATCG-G--G-GA--A-T-GCAG-T--A--------C--G-----AGCA---CTCGCCC-C--G-G-............................. 9286
H2A.ST_M31113               CT-CTGG-GGAC-C-GGG-...........CA-AT-G-G--G-GA--A-TTGCAG-C--A-----G--C--G-----GGCA---TCGCCC-C--G-G-............................. 8728
H2B.D205_X61240             CT-ATGG--G-CTC-CAGG...........A----GGCG-----A--CAT-GCG--C--C--G-----C-----A--AC-T---CTC-CCC-CT-G-G-............................. 9250
H2B.EHO_U27200              CTCGTGG-G---CC-CAGG...........A----GGCAG---AG--CATTGCA-----C--G--G--C-----A--GGCT---CTCGCCC-CT-G-G-............................. 9227
H2G.ABT96_AF208027          CT-ATGGA--ACTC-GGG-...........A-G-T-G-A----GG--A-T-GCA-----G--------C--G--A--GC-T--GCTC-C-C-CT-G-G-............................. 8671
H2U.FR.96.12034             AC-ATGG--GACTC-GGGG...........C-G-T-G-A---TGGC----AGCT-----A--G--G--C--G--A--A--T--GCTCGCGC-TT-G-G-.A............................ 8762

                                                                                                                   Env end \
MAC.239_M33262              CT-ATGG--GACTC-TAGG...........A---G-G-A---TGG--A-T-GCA--C--C--G--G--T-----A--GC-T--GCTC-C-C-CT-G-G-............................. 9243
Env                         __L__W__E__T__L__R__.__.__.__.__R__G__G__R__W__I__L__A__I__P__R__R__I__R__Q__G__L__E__L__T__L__L__*                                gp41
Nef                         T__Y__G__R__L__L__G__.__.__.__.__E__V__E__D__G__Y__S__Q__S__P__G__G__L__D__K__G__L__S__S__L__S__C__E__.__.__.__.__.__.__.__.__.__.  Nef

SMM.SL.92.SL92B             CA-ATGG--GACTC-GGG-...........C-G-T-G-A---GG--CGCAGCA--C--A--G--G--C---------C-T---CTC--GC-TAAC--G.............................. 8691
SMM.US.x.PGM53              CTCATGG---ACTC-GGG-...........A-G-T-G-A---TGGG-A-T-GCA--C--G--G-----T--G-----AC-----CTTGCAC-C--T-G-............................. 9180
MNE.US.-.MNE027             CT-ATGG--GACTC-GGG-...........A-G-T-G-A---TGG--A-T-GCA--C-----G--G--C-----A-AGC-T--GCTT-C-C-CT-G-G-............................. 8720
RCM.NG.x.NG411              T.................................GCAG--TGGCCAGC-ACGT-GTGC-----G--C--G--------C--G--AT--C-TAAC--G.............................. 8535
RCM.GAB1_AF382829           -T-ATGT-CG--G-C--G-...........-ACTT-GC-G--TGGCCAGC-ATGG-GTGC-----G--C--G------C-C--G--AC---GTAAT--G.............................. 8452
DRL.FAO_AY159321            ATCAGCTAG--CCC-C-G-...........-CCCT-GC-CAG-AAG-GGCAGC---T--C--G--G--T-----A---GCA--GGTCC-C--CAAC--G.............................. 8606
MND-1.F17                   AA-CT-CAC--GGTGCAG-GAAGCAATCATT-CT--GG-G-CGTGC--AT-G-------TC-----G--T-----AA-TGCA---C-ACCCC-TAAC---............................. 8378
MND-2.5440_AY159322         T----G--G--C--C-..............-CCTT-G-C--G-GC--CT-GA----C---C-CC-C--C-----A--TGCT--GGCCC-AC-CAAC-G-............................. 8545
MND-2.14CG_AF328295         TA---G--G--C--CG..............-CTAT-G-A--G--C---G-GA----C---C-GC-C--C-----A--AGCT---GTCC-GC-CAAC---............................. 8838
MND-2.MND2_AF367411         A-C-GC-TGCT-CAG-..............-CCAT-GC---G-GC---AT-A----C--A--------TC-------AGCA---GTCC-CC-CAAC-G-............................. 8909
DEB.CM.CM40                 --CGT--C-G-G-C-TGTC...........A----AG-AG---A----T-AGAG-G---GCT-----C--------AGCA--GCTC--CC-CAAC-G-............................. 8559
DEB.CM.CM5                  --CG---C---G-C-TGTC...........T-G-GAG-C---AACC---T-GC---T---GC------C-----A--AGCA--GTT-T-AC-CAAC-G-............................. 8480
MON.CM.L1                   -C-AC-TC-T-CT-G-GG-...........C--CTA--G-----A--CGTGGCT--C--C---C-C--CC-------AGC-----TAT-CC-TAAC-G-............................. 8788
MON.NG1                     ---CC-TCGCC-G-G-GG-...........C--TTA--GC---GAC-CGTGGCTG-C--C---C-C--CC-------GGC---G-TTC-CC-TAAC-G-............................. 7413
TAL.CM.8023                 T--AT-G-CG-CT-CG-G-...........T-G--AGTGCAG--A---GCAGCG--C--GC-G----CG--G--A--GC-T---TATTGCC-TGCT-G-T-CAGG.....ATGGCAAGCATCCGACGCA 8230
TAL.CM.266                  T--AT-G-CG-CT-CG-G-...........T-G--AG--GA---G--CGCAGCT--C--C--GC-C-CG--G--A--AC-T---TATTGC--TACT-G-T-AAGA.....ATGGCAAGCAACAGGAGCG 8680
MUS.CM.1085                 A--AC-TC---CT-GT-G-...........A---TA--GC---AAT-CTTGGCT--C--C---C-CG-CC-G-----GGC---G-TAC-CC-TAAC-G-............................. 8780
SAB.SN.-.SAB1C              A--ACG-C-T-CTTGCAG-...........TCCAT-GT---G--AG-CAT-GCACAT--A--------GC-C--A-AGC-T--G--ATGG--TAAT--G.............................. 9087
GRV.ET.x.GRI_677            T-C---T-TCCG-.................TCC----A-C-G--AG-CAT-A---GT--A------G-GC-G--A---C-T----AAG-CC--GGG---............................. 8766
TAN.UG.x.TAN1               -AA-GCTTGCAG-.................TCC----A-C-G--A--C-T-G-ACAT--A--------GC----A-AAC-----C--TGG--CAAC--G.............................. 8906
VER.DE.-.AGM3               -A-ATGGCTT-CTTGCAG-...........TCC----A---G--A--CAT-A--TCT--A------G-GC----A---C-T---G-A--CC-TAAT--G.............................. 8411
VER.KE.-.9063               AA-ATGGCTT-CTTGCAG-...........TCCA---A---GCAC--CAT-AG--GT--A------G--C----A--GC-T--GGAA--CC-TAAT--G.............................. 8913
VER.KE.-.AGM155             -A-ATGGCTT-CTTGCAG-...........TCC----A-C-G--A--CAT-A--TCT--A------G-GC----A--GC-T--GGAAG-CC-TAAT--G.............................. 8890
VER.KE.-.TYO1               -A-ATGGCTT-CTTGCAG-...........TCC----A---GAAC--CGT-A---GT--A------G-GC----A---C-T--GGAA--CC-TAAT--G.............................. 8367
SUN.GA.98.L14               CT---GTCTC-GCAG-A--...........T-GAGGCCC-A-TGG-A-A-AACA-G-GGA--T-----TCCCAGT-AGACCAC-GA--CCACTT-G---............................. 9181
LST.CD.88.447               CC-C-G-CTG-G-CA-A-G...........AAGAGA--G---TTC-GATT-AGAGG-AGA--TG-GT-TCCCAGT-AGACCAC-GA--CCGCT--G--G.............................. 8035
LST.CD.88.485               CA-C-G-CTG-G-CA-A-G...........AA-AGA--G---TTC-GATT-AGAGGGAGA--TG-GT-TCCCAGT-AGACCAC-GA--CCGCT--G--G.............................. 8032
LST.CD.88.524               CA---GGTTG-G-CA-A-CAGA........A--AGA--G---TTC-GATT-AGA......--TG-GC-TCCCAGT-AGACCAC-GA--CCACT--G--G.............................. 8035
LST.KE.-.lho7               CA-CCA-TTG-GCAA-A-G...........AAGAGA--G---CTC-GATT-GGAGG-AGA--TG-G--TTCCAGT-AAGCCAC-GA--CCGCTT-G--G.............................. 9110
SYK.KE.-.SYK173             AA-ATAC-CC-CTACCAG-...........C---T-GTCGAG--CG--GCAGC-T-G--GC-G--GC-TC----A--------G-TCTACC-TAAC-G-............................. 8723
SYK.KE.KE51_AY523867        .................................-T-GC-GAGAA-G-CGCAGC-T----AC-----C-TC-------GC----G-TCTACC-TAAC-G-............................. 8422
COL.CM.-.CGU1               -AGGTGGCT-CG-TCTGCGGCG........C-T-GC--GC-TCGAGCG-CTG-ATAT-TGC--G--TGG-T-T-T-A-CGTCCCCAAGG-CCTGCC---............................. 8151
```

```
                                / Nef start
H1B.FR.83.HXB2            ......ATGGGTGGCAAGTGGTCAAAAAGTAGTGTGATTGGATGGCCTACTGTAAGGGAAAGAATGAGACGA....................................GCT.........GAGCCAGCAG  8875
Nef                             _M__G__G__K__W__S__K__S__S__V__I__G__W__P__T__V__R__E__R__M__R__R__.__.__.__.__.__.__.__.__.__.__.__.__A__.__.__.__E__P__A__   Nef

H1D.CD.84.84ZR085         ......-----------A---------------A-AG-----------G--A-------------A---AA-............ACTGATCCAAGGGAAAGAAGAAGAC--.........----------  8420
H1F1.BE.93.VI850          ......-----G-----------------C---A-AG-----------G-----G---------------A-....................................A-C.........CCTA------  8152
H1G.SE.93.SE6165          ......-----A--------------------CA-AG----------CGAG--------------A---AAC....................................A-C.........CCTA------  8314
H1H.CF.90.056             ......-----A-----A--------------GA--GG---G---T-----A----------------G---....................................---.........--A----T--  8195
H1J.SE.93.SE7887          ......-----GAA---A------------............------CAG-----------------G---.................................GCTC-C.........-CT-------  8173
H1K.CM.96.MP535           ......---------------------------A-AG-----------G--A--------------------.........GCACGACCAGCAGCAGACAGGGTGGGAA-A.........C-A-------  8062
H101_AE.TH.90.CM240       ......-----AA-T-----------------CA-AG-G---------CAG--C-------A---A-AG....................................CAAA--.........CCT------A  8449
H102_AG.NG.-.IBNG         ......-----------------------C--CA-AG-G----------AG--T-T-A-----------....................................CAAA-C.........CCTA------  8402
H103_AB.RU.97.KAL153_2    ......-----G-----A--------------CA-AG-G---------CAG--T-----------A------....................................---......CCT-CT-----G-  8055
H104_cpx.CY.94.CY032_     ......-----A-----A--------------CA-AG-----------GAGA----------------G---GCTCGAGCTGAGCCAGAAAGAATGAGGCGAGCTCAA---.........----------  8284
H1N.CM.95.YBF30           ......-----AAAG-TT--------G--C--CC-AG-A--------AGAAA-C--A------------A--.....................CAAACGCAAGAACCA--A.........-TAGAGC---  8462
H1O.BE.87.ANT70           ......-----AAA-GCA-T-AG----G---AAT-TGAG------G-AG-A-----A------------A--.....................ACTAGAACTTTCCCT-AG......TCT--A---TGC-  8962
H1O.CM.91.MVP5180         ......-----GAATGCA---AGC-----C-AAT-TGCA------T-AGAA-----A--T------------.....................TCCTCCTCTGATCCTCAA.........C-A---TGT-  8985
CPZ.CM.04.EK505           ......-----AAAAGTT--------G-----CC-TG-G--G-----ACAA--C-TA--T---------A--.............................................CAGC-AGATC---  8382
CPZ.CM.04.LB7             ......-----AA----A--------------CC--G-A---------GAA--T--AAG------T----A-.............................................GAAACTGT-----  8407
CPZ.CM.04.MB66            ......-----A-----------------C---C--G-G--C-----AGAG-----A--C---C-T--GA-G.............................................ACAC-AA-T----  8520
CPZ.MT145                 ......-----A-C------------------CC-AG-A--------AGAAA----A--G-----A-A-A--.............................................GCTCCAA------  8408
CPZ.CD.-.ANT              ......------TCTGCA-----T----TC-AGTG-G-----GCA.........--AC--GC---T--GAA-....................................ATA......CAT--AA--AATC  8329
CPZ.CM.-.CAM3             ......-----CAA--G---------------CC--G-A--G-----AGA---C--A------C-A--CGA-....................................CTC.........A-C-TGCTT-  8312
CPZ.CM.98.CAM5            ......------AA-------------------C--G-G------G-AG----T--AC----G--A----A-....................................A--.........CCA-------  8618
CPZ.GAB_X52154            TTAAG.-----AAC---A-----T---------C--G-A---------GAG--C--AAG------A--GGA-....................................---.........CCAA------  8948
CPZ.GAB2                  .....A-----CAA---A-----T-----C--CA-TG-A--------GCAG--C--A--G---C-T--G--G..............................ACT...............C--GA-----  8286
CPZ.TZ.-.TAN1             ......-----AAA--TA-TTGGT-G-......TG-CC---GGCC-GG-AA-CC-TC---GATC-T..........................................CA-.........A-CA-CT--A  8561
CPZ.US.85.CPZUS           ......------AA---------T--G-----CA-AG-A--------AGAA--C--AA-C---T--....................................AGA...CAAACTCAGACAACAG------  8908
GSN.99CM166_AF468659      ..................................................................................................................................  8772
GSN.99CM71_AF468658       ..................................................................................................................................  8772
H2A.ALI_AF082339          .............................................................................................................................--G-C  9287
H2A.BEN_M30502            .............................................................................................................................G-G-C  9291
H2A.ST_M31113             .............................................................................................................................G-G-C  8733
H2B.D205_X61240           ..................................................................................................................................  9250
H2B.EHO_U27200            ..................................................................................................................................  9227
H2G.ABT96_AF208027        .............................................................................................................................--GGC  8676
H2U.FR.96.12034           ........................................................................................................................-CT-AG-GCC  8772

MAC.239_M33262            .............................................................................................................................G-G-C  9248
Nef                       _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G_    Nef

SMM.SL.92.SL92B           ..........................................................................................................................-GAT-G-T  8699
SMM.US.x.PGM53            .............................................................................................................................G-G-C  9185
MNE.US.-.MNE027           .............................................................................................................................G-G-C  8725
RCM.NG.x.NG411            ..................................................................................................................................  8535
RCM.GAB1_AF382829_        ..................................................................................................................................  8452
DRL.FAO_AY159321          ..................................................................................................................................  8606
MND-1.F17                 ..................................................................................................................................  8378
MND-2.5440_AY159322       ..................................................................................................................................  8545
MND-2.14CG_AF328295       ..................................................................................................................................  8838
MND-2.MND2_AF367411       ..................................................................................................................................  8909
DEB.CM.CM40               ..................................................................................................................................  8559
DEB.CM.CM5                ..................................................................................................................................  8480
MON.CM.L1                 ..................................................................................................................................  8788
MON.NG1                   ..................................................................................................................................  7413
TAL.CM.8023               AATTGG...ATG--G---GCACTCTCG.......................................................................................................  8254
TAL.CM.266                GACTAC...ATG------GCA-T--C-.......................................................................................................  8704
MUS.CM.1085               ..................................................................................................................................  8780
SAB.SN.-.SAB1C            ..................................................................................................................................  9087
GRV.ET.x.GRI_677          ..................................................................................................................................  8766
TAN.UG.x.TAN1             ..................................................................................................................................  8906
VER.DE.-.AGM3             ..................................................................................................................................  8411
VER.KE.-.9063             ..................................................................................................................................  8913
VER.KE.-.AGM155           ..................................................................................................................................  8890
VER.KE.-.TYO1             ..................................................................................................................................  8367
SUN.GA.98.L14             ..................................................................................................................................  9181
LST.CD.88.447             ..................................................................................................................................  8035
LST.CD.88.485             ..................................................................................................................................  8032
LST.CD.88.524             ..................................................................................................................................  8035
LST.KE.-.lho7             ..................................................................................................................................  9110
SYK.KE.-.SYK173           ..................................................................................................................................  8723
SYK.KE.KE51_AY523867      ..................................................................................................................................  8422
COL.CM.-.CGU1             ..................................................................................................................................  8151
```

```
H1B.FR.83.HXB2              CAGATAGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATGGAGCAATCACA...AGTAGCAATACAGCAGCTACCAATGCTGCT........................TGTGCCTGGCTAGAAGCACA  8978
Nef                         A__D__R__V__G__A__A__S__R__D__L__E__K__H__G__A__I__T__.__S__S__N__T__A__A__T__N__A__A__.__.__.__.__.__.__.__.__C__A__W__L__E__A__Q   Nef

H1D.CD.84.84ZR085           -----G-----------T-----------------G------G---------..---------------A----T--------C.........................----------G----------  8523
H1F1.BE.93.VI850            ----AG-----------TG---------T-A--C-G--GG-------T---..--C--------TAG-A----T---C---AC.........................-TG--------G--------  8255
H1G.SE.93.SE6165            ----AG-A--A-------T-----A---TT-A--T-GG---------------..--C-------------C-AT---C---A-.........................-----------G--------  8417
H1H.CF.90.056               ----AG----A-------TG--------TT----T-G--GC--G---G-----..-T--AT---------T----T---CG--A-.........................GCC--------G--------  8298
H1J.SE.93.SE7887            -----G-A----------TG----A---TT---CT--G-----G---------..--C-----------------A-------AC.........................-----------G--------  8276
H1K.CM.96.MP535             ----CG------------T-----A--------CT-G------G---G-----..--C--------TT--CAC-AT---C---A-.........................-----------G--------  8165
H101_AE.TH.90.CM240         ----AG-A--A-------T-----A---T--A--T-------------A---..------T---.....ATAGAT-------A-.........................----T------GAG------  8546
H102_AG.NG.-.IBNG           --ACAG-A--A-------------A---TT-A--T-G----------------..--C--T----------CAA--T--CC---A-.........................-----------G--------  8505
H103_AB.RU.97.KAL153_2      --AGAG-A--A---C---T-----A---TT----T--GT-------G------..--C--T-------------AT-------A-.........................--C--------G--G--G--  8158
H104_cpx.CY.94.CY032        ---CAG-A--A-------TG----A----T----C--------G---------..-T--A---------------T---C---AC.........................AAAA-------G--------  8387
H1N.CM.95.YBF30             ---TAG-A-CA--------T----A---T--A-CT--T-GA--G--C-----C..-TA--A-----TAG--AC-AT----AAAG-.........................ATA--T-------------  8565
H1O.BE.87.ANT70             --CC-G-A--A---CAGATC--CA-G--AT-A-C-GCTAGA---GG--AC--..---TC-C----TC-TCAA-A------A--C.........................CT---A-TC------AGT--  9065
H1O.CM.91.MVP5180           --CC-G-A--A-----T-TC--CA-G--GT-A-C--CTAGA--G-G---AT--..---TC-C-C--TC-TCAA-A------A--C.........................CT---A-TC-----CAGC--  9088
CPZ.CM.04.EK505             ----AG---A-------TT----A---T--A-CT--T-GA--G--T--T--C..-TA--A-----TAAG-AG-AT---CAA--C.........................CTA--T----------AG--  8485
CPZ.CM.04.LB7               ----AG-A--A---C----C--A-A------A-CT-GG-----G-----T---..------T-----CC--CAA--T----AGA-C.........................AT-------T----G-AG--  8510
CPZ.CM.04.MB66              ----AG-A--A---C---T---A--G-----A-C--G------------A--C..------A--C---T-CCAG--A----AGA-C.........................CT-------T------A-GT  8623
CPZ.MT145                   -----G-A--A--G-A--T---A-A----A-A-C--GC-G-------G-----..-T---A-----CC--CAA-------AGA--.........................-TA--T----------A-AT  8511
CPZ.CD.-.ANT                -----GATA-A--GC-CTGTGGAAAC--AT---CC-GTAGA--G---C-A--C..TCCTCT-C--T-...-G---AG-AAAA-A-.........................GTAAT-ACCTATTC--A-G-  8429
CPZ.CM.-.CAM3               --CCAG-A--A---C---TC--A-A----T-A--GGC---------G----C..-C---GC-C---C--CAG-AT---CAGA--.........................CTG------T------A-AT  8415
CPZ.CM.98.CAM5              --TCAG-A--A---C---TC--A-A---T--A--GGC--G-----------C..-C---A---CAC...A-A-AAT--CAGA-C.........................CTG------T------A-AT  8718
CPZ.GAB_X52154              ----GG-A--A----A--TT--GAAG-----A----G---C-----T-----T..------G--C--CC---AG--T---CAAA--.........................CTA--T--------G-A-AT  9051
CPZ.GAB2                    ----AG-A--A----A--TG----AG--TT-A-C--GG-G---------A--T..TCC--AC----CC-TCAG--T---CAAA--.........................CT---A----------A-AT  8389
CPZ.TZ.-.TAN1               GT--GCCT--A---CAG--C--A-A------CC-G--TA-A----GTC----T..-C--A--CCCT-...-G----TCA--A-A-.........................GTGTTAGAATACTCT---G-  8661
CPZ.US.85.CPZUS             ----AG-A--A---C-T-TC----A----T-A-C-G-------G--T-----T..-C---A------C--CAG-A---CCAAA--.........................CTG--A---T----T-AGAT  9011
GSN.99CM166_AF468659        .........................................................................................................GAGCCAAAACCCAAGAG--G-AG  8795
GSN.99CM71_AF468658         .........................................................................................................GAA-CAC---CCAGGC--AATTG  8795
H2A.ALI_AF082339            AGCGGTATCA-CAG-G--AT-T-ATGA-TACCCC-TGGAGAACCC--GCA---..GAA--AG-G-A--A-TTGTA---GCAACAG.........................AA-ATGGATGAT-T--ATTT  9390
H2A.BEN_M30502              AGC-GTATCA-CAG-G--AG-ACATGA--AGCCC-TGGAGAAACC--GCA---..GAA--AC-G-A--ATTTGTAT-GGCAGCAA.........................AA-ATGGATGAT-T--ATTC  9394
H2A.ST_M31113               GGCGGTATCAACAG-G--AT-T-ATGA-TACCCC-TGGAGA-CCC--GCAGA-..G-GGAG...-A--GCT-GTA---GCAACAA.........................AA-ATGGATGAT-T--ATTC  8833
H2B.D205_X61240             ...........................................G-G-CAA-AT..GCCC-GGGGG---A--GGGGAGGGCAACAA...............................GATGC---T-A-AG  9304
H2B.EHO_U27200              ...........................................T-G-CAA-AG..-C-TTGGGAG---A--GGGGAGGAAAACAA...............................GATTC---T-AGG-  9281
H2G.ABT96_AF208027          A-C-ATACAC-CAG-GTCA--T-ATGA-TACCCC-TGGAGAAACC--GCA---..GAA--AGCA-AGTT---ATA--GGYAGCAA..........................AATAATGAT--T-T-G-  8776
H2U.FR.96.12034             A--GCTATAGT-A--G--AG-TCATGA-TACACCTTGGAGAACCC--GCA---..G-G-AAG---A-TT-CAATAT--GCAACAA.........................AACATGGATGAT-T--ATG-  8875

MAC.239_M33262              AGA-ATACAATCAG-G-CAG-A-ATGA-TACTCC-TGGAGAAACC--GCTGA-..GAG--AG-A-A-TT---ATA--GAAAACAA.........................AA-ATGGATGATAT--ATG-  9351
Nef                         Q__K__Y__N__Q__G__Q__Y__M__N__T__P__W__R__N__P__A__E__.__E__R__E__K__L__A__Y__R__K__Q__.__.__.__.__.__.__.__.__N__M__D__D__I__D__E   Nef

SMM.SL.92.SL92B             ATA-ATACAAT-T--GG-A--T-ATGA--ACCCC-TGGAGAAACC-TGCT---..GAA--AC-A-AGCA-CAGTAT-GGCAGCAG.........................---ATGGATGATAT--ATG-  8802
SMM.US.x.PGM53              AG-GATACAGT-AG-G-CA--TCATGA--ACCCC-TGGAGAAATC--GCT-G-..GAG--AG-A-AGTT-AAATAT-GACAACAA.........................AACATGGATGAT-T--ATG-  9288
MNE.US.-.MNE027             AGA-ATACAATCAG-G-CAG-T-ATGA-TACTCC-TGGA-AAACC--GCTGG-..GAG--GG-A-A-TT---ATA--GAAAACAA.........................AA-ATGGATGATAT--ATG-  8828
RCM.NG.x.NG411              .....................................-GGA--GTTT.........-CA-TAG-A--TCAGCAGT-AGTG-A---A.........................ATA-AT---TAT----ATAC  8595
RCM.GAB1_AF382829           ..................................-GG..-..................................GGCAT-A-AGCAGAAACGCAAGCTTCAATAGATGACATTGACTGGT-T-A-G-  8509
DRL.FAO_AY159321            ................................................--GCTTAGA--CT-C-CATATGCAAG-AGCA--A--AGAGAAAGATGTCACCCTTCCCACAGAA-AAGAA-A-CCCT--G-  8687
MND-1.F17                   ...........................................................................................AGATGAGAAACTTGACTTAACA-AGC-G-AT-C  8411
MND-2.5440_AY159322         .........................................................................GCGGGATCTAAGGCACCAGCAGGATGTAACTCTCCCCTCAGAAAAAGA--A-CC-AGTG-  8605
MND-2.14CG_AF328295         ......................................................................................................................AG----T-ATG-  8850
MND-2.MND2_AF367411         ..............................................................................................................................CG-  8912
DEB.CM.CM40                 ...........................................GT--TGG..G-A--TCC--G-AG--AA-TA--GTGCA-C.........................AA-AAGCAA-A-C---ACTT  8616
DEB.CM.CM5                  .......................................G-A-C--C--TAT..GCAGA-CCCCACAG--AG-TA--ATGCA--.........................CAGAGAGCC-AG--GTGTT-  8543
MON.CM.L1                   ...............................................................G-GCAGAAATGGCAAACAACGCCGGCAACCT......CT-GAC--C-ACG-  8833
MON.NG1                     ...............................................................G-GGCGGAACCACCGCTCAAACCGCCGCCCAACCA-TGTCG-T--T-ACGC  7464
TAL.CM.8023                 ...........................................................................................GAAGGAGAAGTC--GTATGAAAAG--GCA-G-  8286
TAL.CM.266                  ...........................................................................................CAAGGAGAGGTT--GTATGAAAA----CA-G-  8736
MUS.CM.1085                 ..........................................................................................................AG-A---A-TAA-AG-C-TGT  8801
SAB.SN.-.SAB1C              ...........................G-AA-G-G-CTCCTCTAG--CAGA-..GAAG-TGTA-A-ATGAAACATC-A-GAAGA.........................CAGC---C-TGGT-T-ATG-  9162
GRV.ET.x.GRI_677            ........................................A--TAA---G--T..CCAGA-GGCCGCCGCCTGCAAG-A-GA-AC.........................ACCTTTGATGAGTGG-ATG-  8829
TAN.UG.x.TAN1               .....................................GAGG---AAG-G---C..CC-GAAGGA-G-AA-TTGG-AGC--A--AC.........................ACATGGGATGACTGG-A--C  8972
VER.DE.-.AGM3               ..............................-A-G-G-A----TAA---G--T..CCAGAAGGA-G-AA-TTACAAG-A-GA-A-.........................AAATGGGATGA-TGGT-TG-  8483
VER.KE.-.9063               ..............................-A-G-GGA----CAAC--G--T..CCAGATGGA-G-CGCCTGCAAG-A-GA-AC.........................GAATGGGATGA-TGGT--G-  8985
VER.KE.-.AGM155             ..............................-A-G-G-A----CAAC--G--T..CCAGAAGGA-G-CGTCTACAGG-C-GG-AC.........................CAATGGGATGA-TGGT--G-  8962
VER.KE.-.TYO1               ..............................-A--C--A----CAAC--G--G..GAAGAGG-A-GGAGGCT-CAAG-A-GA-AC.........................ACCTGGGAAGAGTGGT-GG-  8439
SUN.GA.98.L14               .................................................................................................................................  9181
LST.CD.88.447               .................................................................................................................................  8035
LST.CD.88.485               .................................................................................................................................  8032
LST.CD.88.524               .................................................................................................................................  8035
LST.KE.-.lho7               .................................................................................................................................  9110
SYK.KE.-.SYK173             ...........................................GCC--TTGACCCACA-GGGC-G-ACAGAGA-TGG-GACA-TCGGGAGGACAAAAATTCTCACCTG-A-ATATAG-GC-G-ATG-  8807
SYK.KE.KE51_AY523867        ...................................................ACCC-T-TTGGATC-CAAA-GAG-CTT--ACTCTGGGGGGAAAAAATGGAATGCGG-A-A-ATCTGTC-T-ATG-  8497
COL.CM.-.CGU1               .........................................................-TGCC-TTGCC-GCT-TGCTA-AGACATCCAGATGTATATTACCACAACC-AGAA-C---G-ACGT  8217
```

```
                                                                                                                             polypurine tract
                                                                                                                             |-            -|      |- 3'LTR U3
H1B.FR.83.HXB2            AGAGGAGGAGGAG......GTGGGTTTT...CCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATT  9099
Nef                       __E__E__E__E__.__.__V__G__F__.__P__V__T__P__Q__V__P__L__R__P__M__T__Y__K__A__A__V__D__L__S__H__F__L__K__E__K__G__G__L__E__G__L__I_   Nef

H1D.CD.84.84ZR085         G--A-----C---GAG...-----C---...-------G--------------------------------------------------------------G-A-----------------------G--  8647
H1F1.BE.93.VI850          ------------A......--A--C---...-------G-------------G-------------C--T-----------------C--T---------------------------------T-----  8376
H1G.SE.93.SE6165          ---------CTCAGAA...--A--C---...-------G---A--A--------G-------------TT----GT---T-T-----C---TT---------------------------T---------  8541
H1H.CF.90.056             ------C-G----GAA...--A--C---...-------GG--------------------------C--T----G----T-T-----C-----T--------------------------T---T-----  8422
H1J.SE.93.SE7887          -ACA---------......--A--A---...-------A------A-----A--------------A-------G------T---------TT---------------------------T---T-----  8397
H1K.CM.96.MP535           G-----A-----A......--A------...-------G------------A-----------------T---------T-T-----CG--TT---------------------------T---------  8286
H101_AE.TH.90.CM240       ------C------......--A--C---...-------TG--G---------C----------------T----G----T-T---------TT---------------------------T---------  8667
H102_AG.NG.-.IBNG         ------T---A-T......--A--C---...-------GG--A-----G-----G--------------T----G------C-----C--------------------------------T---T-----  8626
H103_AB.RU.97.KAL153_2    ------A------......--A--C---...-------GG--G-----------G-----G-----A--T----G----T-C-----C--------------------------------T---T-----  8279
H104_cpx.CY.94.CY032      ---A-----A--AGAG...--A------...-------GG--A--------------G--------C-TT--A-G----T-------C--------------------------------T---------  8511
H1N.CM.95.YBF30           ---A--A--A---GAA...--A--C---...-----ACGC-----------A-----G-----A--C--T--ACAG---T-T------TC-TT------------T---------------------G--  8689
H1O.BE.87.ANT70           CC-A-----A--A......--A------...-----AG-------A--G---C----G--------C--T--A-G---AT-T--C--C---TT------------------A------------T-----  9186
H1O.CM.91.MVP5180         CA-A--T-----T......--A--C--C...-----A-G------A--G---C----G--------C-TT--A-----CT-T--C--C---TT------------------A--------T---T-----  9209
CPZ.CM.04.EK505           -A-A-----A-CA......--A--C---...-----ATGT-----------A--------------A-----A-A----T-T------TC-TT--------G-------------------------G--  8606
CPZ.CM.04.LB7             -A-------A--A......A-AT-G---...-----A-A------------A-----G----------TT--A-AG---T-T------G--T--------------------------------T--G--  8631
CPZ.CM.04.MB66            -C-AA-T--A--A......--A--C---...-----A-GG-----A------C-------------C-TT----G----T-T-----C---------------------------------------G--  8744
CPZ.MT145                 GC----T--A--A......-----A---...-------A---A--A--G--C--------T-----C-TT----GT--CT-T-----CTCT---------------------------------T----C  8632
CPZ.CD.-.ANT              TC-TACA-----A...GGGACA--C---...--T--G-G------A--C--CA-G-----C-----AG-A--ATT---------C--GTCATGN--------------------------TN--------  8553
CPZ.CM.-.CAM3             GC-A--AC-CA-TGAAGAA--A--C---...-----T-G------A------C-T-----T-----C-TT--A------T-T---T-A---------------A--------------------------  8542
CPZ.CM.98.CAM5            G--AA-TC-T--AGAAGAA--A--C---...-----T-G---C--------CC-------------A------CA----T-T---T-A------------------------------------T--G--  8845
CPZ.GAB_X52154            G--CA-T--A......GAA--A--A--C...-----A-G---A-----T--AACG--------------T--A------T-T------TCA---------------------------------T--G--  9172
CPZ.GAB2                  G--AC-------T......--A--G---...-----A-G---------------------------A--T--A-----CT-T-----C------------------------------------T--G--  8510
CPZ.TZ.-.TAN1             CC-TACT--A--AGAA...--A------...-------G---AGCA-----CA-G-----C-----AG-G---CT---AA-------GTCATGG--C-----------------------T------T--  8785
CPZ.US.85.CPZUS           GACTA-TC-T--AAGTGAA--A--C--C...--T--TCGC-----A-----A-----G--C-----------ACA---AT-C---T-AG--TT-------------------------------T--G--  9138
GSN.99CM166_AF468659      CCTC------T-TGAACCATCAT-C............A-G---A-G------AC-T--GACCCC---C-----A-TCATG--G-----CTCT--------G-----------A--------G--CA-GT--  8919
GSN.99CM71_AF468658       CATC--T...T-TGAACCATCAT-C............G-G---A-----G--GC-T--GACCCC---C-----ACTTATG--G-----CTCG--------G-----------A--------G-CCA-GT--  8916
H2A.ALI_AF082339          ---T--T--T--TAGCCTA--A--GG-C...T-T--------AAGA---AA------A--------A-----ATTG--A------A-GTCA--T--AA---------G----------------A-GT--  9517
H2A.BEN_M30502            T--T--T--T--C...CTAA-A--AG--...--T--T-----AAGA-----ACGG--GA-------C--T--ATTG--AA-----A-GTCA--T---A----------A---------C-----A-GT--  9518
H2A.ST_M31113             ---T--T--T--C...CTA--A--GG-C...--T--------AAGA-----A-----GA-------A--T-G-TTG--AAG----A-GTCA--T--GA----------------------------GTA-  8957
H2B.D205_X61240           T--T-----C--TGAA...-----AGCCATATGTAAG--C--AATA-----AC-G-----------A-----ACT---A-----CA-GTCT--T---A--------C-----------------A-GTA-  9431
H2B.EHO_U27200            T--T-----CA-TGAA...--A--GG-C...-GT--A-G---CGG---C--AC-G--G--------A-T---ACT---A-----CA-GTCT--T-------------------A----------A-TT-C  9405
H2G.ABT96_AF208027        TAGT--T--TA-TGACTTA--A--AG-A...------TAC--CA-A--C--GC-----GT-----G------ATTG--AA----CA-GTCA--T---A--------------------------A-TTA-  8903
H2U.FR.96.12034           ---A--TA-T--T...TTA--A--AG-A...G----A-G---AAGA---G-AC-G---A-------------ATTG--A------A-GTCA--T---A----------A-----------T---A--T--  8999

   premature stop in Mac239 |
MAC.239_M33262            GT-A--T--T--C...TTG--A--GG-A...T----G-GG--AA-A--T--CC-----A------G------ATTG--AA----CA-GTCT--T---A--------------------------A-TTA-  9475
Nef                       __*__D__D__D__.__L__V__G__V__.__S__V__R__P__K__V__P__L__R__T__M__S__Y__K__L__A__I__D__M__S__H__F__I__K__E__K__G__G__L__E__G__I__Y_   Nef

SMM.SL.92.SL92B           G--T--T--T---...CTG-CT--AG-A...GTT---TGG--AA-A-----CC-----G-----T-A-----ATTG--AA----CA-GTCA--T---A----------------------T---A-TTA-  8926
SMM.US.x.PGM53            T--T--T--T---...CTGA-A--AG-A...T----ACAC--AA----G--A------G-C---T-A-----ATTG--GA----CA-GTCT--T---A--------------------------A-TTA-  9412
MNE.US.-.MNE027           G--A--T--T-GC...TTG--A--GG-A...-----G-GG--A-GA--T--C------G-------------ATTG--A------A-GTCT--T---A--------------------------A-TTA-  8952
RCM.NG.x.NG411            ---T--TACCTT-......--A--G---...-----G-G---A--A--C--CA----G------T-------ATT---AA-------ATCT---------------------------------A-TTA-  8716
RCM.GAB1_AF382829         TACT--T--CACC...TTG--A--A---...-----G-A------A-----AC-T----------G-------CT---AA----CA-GTCT---------------------------------A-TTA-  8633
DRL.FAO_AY159321          ---A--A--A--AGAA...--A------...-----ATGT-----ACGC---------GAGCCA--C-----A-ATCTA-----CT--TC------------G---------------A-----A-TTGG  8811
MND-1.F17                 ---A--------AGAA...--T--A---...--T--GTGT-GC--A--TT-C------GTGCCAT-A-----A-ATCTGA----CT-CTCT--T---A---------------------G----A--TA-  8535
MND-2.5440_AY159322       T--A-----A---......--A--C---...------TAC--AAC-CAG--AG-GCCTGA-GCA--C-------ACTTGA----CA-GTC----------------------------------A-TTGG  8726
MND-2.14CG_AF328295       GC-AA-A--A--A......AG---C---...-----TTAT--CACACAG---G-GC-TGA-GCA--C-----AAACCTGA-----A-GTC----------------------------------A-TTGG  8971
MND-2.MND2_AF367411       ------A--AAC-.........--C---...-----ATAT--A---TGC--AG-----GA-CCC--C-------ACTTG--T---A-GCC----------------------------------A-TTGG  9030
DEB.CM.CM40               ---TC--ACA--AGCA...--A--C--C...-----ACGC--C------T-CA-CC--GACCCC--A--T---TT-ATGA----CTACTCT--T-----G-----------T--------G-ATG-TT--  8740
DEB.CM.CM5                T--T-----AC-TGGG...A-A--C---...G-T----G---AAG---C--AA-T--GGA-CCC--------ATTGATGA----CTACTCT--T-A---------------T--------G-ATA-TT--  8667
MON.CM.L1                 C--C--C--C-GT......--C----G-...-----A-GG--C-GA--T--GC-TC-CGACCCT----GG---CTCATGA-G--C--CTCT--T-A---G-----------A-------G--A-A-GT--  8954
MON.NG1                   C--C-----CC-AGATTTGCCT--C-G-...--G--TCGC--AT----G--AC-T---GACCCT--A-GG---ATCATGA-G--C---TCT--T-----------------A--------TAA---GT--  7591
TAL.CM.8023               ---AA-T--CAT-GAGCTT--A--C---...-----A-GG------A-T--A--G--G--CTGT-A------ACTG-G-A----CT-CTCGTT------G--------A---------A-G---A--T--  8413
TAL.CM.266                ---AA-T--CAT-GAACTA-----C--C...--T--G-GG---A-A--T--G--G---A-CTGT-----T--ACT--G-A----CT-CTCGTT------G-----------A------A-G---A--T--  8863
MUS.CM.1085               GATA--A---C-T...GAGCGTACC-GC...--G--GCGT--GAG---G--AA-CC-TGACCCC--C------CTGATG--G---A-GTCG--T--C--------------A------------A-GT-C  8925
SAB.SN.-.SAB1C            G--T--A--A--A......--A--C---...-----A-GG----TGCT---AC--C--G-------A--T---TTG--AA----CT--G-----------------------------------A-TTA-  9283
GRV.ET.x.GRI_677          T--T--A--A--A......--A--C--C...--T--GCA-----GA--C--C-------AG-----C--T--ATT---A--G--CT--TC-------------TC---------------T---A--TA-  8950
TAN.UG.x.TAN1             ------A--A--A......-----C---...-----A-G---A-GA--G---------A-------A--T---TTG--A-------T-CTC--------------------A--------G--CA--TA-  9093
VER.DE.-.AGM3             T--A--A--T--A......--A--A---...-----A-G---AAGA--G--GC-----A-------C--T--ATT---G--G--CT--TCG-----------------------------T---A-TTA-  8604
VER.KE.-.9063             T--T--A--T--A......-----A---...--T--A-A---AAGA-----A--G--G-A------------ATT---A--G--CT--TCG--------------------A--------T---A-TTA-  9106
VER.KE.-.AGM155           T--A--A--T--A......-----A---...-----A-G---AAGA--G--AC------A---A--A-----ACTT--A------T--TCG--------------------A--------T---A-TTA-  9083
VER.KE.-.TYO1             T--T-----A--A......-----A---...-----G-G---AAGA-----C-----G-A---------T--ACTT--A--G---T--TCG--------------------A--------T---A-TTA-  8560
SUN.GA.98.L14             ............................................................GCCA--A--T---CAGCTAA-------CTCA--T---A-------------T--------------CTGG  9251
LST.CD.88.447             ............................................................CCCT--C--T---A-TCTGA----C--CTCA--T---A-------------T---------------TA-  8105
LST.CD.88.485             ............................................................CCCT-----T---A-TCTGA----C--CTCA--T---A-------------T---------------TA-  8102
LST.CD.88.524             ............................................................CTCT--C--T--AA-CTTGA----C--CTCA--T---A-------------T---------------TA-  8105
LST.KE.-.lho7             ............................................................CCCT--C--T---A-TCT-A----C--CTCA--T---A-------------T---------------TA-  9180
SYK.KE.-.SYK173           --GA--CACA-GT...CTC--A-----C...--T--GTGT-----AAC---AC-C---A--T-A--C-----ATTG--AA----C--CTCT--T---A-----A-C-----A------C-G--CA-G-A-  8931
SYK.KE.KE51_AY523867      G-GA--T-----T...CTA--A--C---...--T--TTT---A--A-----C--G-----CT-A--------ACTT--CA-------CTC---T---A-----AG------A------C-G--CA-GT--  8621
COL.CM.-.CGU1             ---A--T-TA--A...GCA-GAT---A-......--G-TGT---GAC-G--AG--C---AGGC---CATT---CT-CTG--G--CA-CTCTTG------G-----------T--------G---A-T---  8338
```

```
                                                                         Nef end premature in HXB2 \
H1B.FR.83.HXB2             CACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGG.....................GTCAGATATCCACTGA  9208
Nef                        _H__S__Q__R__R__Q__D__I__L__D__L__W__I__Y__H__T__Q__G__Y__F__P__D__*__Q__N__Y__T__P__G__P__G__.__.__.__.__.__.__.__V__R__Y__P__L__   Nef

H1D.CD.84.84ZR085          T-------G-A---------------------T---G------------------T-----------G---A---------------------.....................A-----------T-A-  8756
H1F1.BE.93.VI850           T-----A-G-A----GG-----C-------------G-T----------------------------G-------------------------.....................A---------------  8485
H1G.SE.93.SE6165           T-----A-G-A---------G--------C------G-----A-T--------A--------A----G------------G------------.....................ACT----TC--C----  8650
H1H.CF.90.056              T-----A-GCA------G--C---------T-A---G----TA----------------------C-G-------------------------.....................-AG----T---C----  8531
H1J.SE.93.SE7887           T-----A-G-A---------G---------------G-TC-TA----------T-----------C-G---A---------------------.....................A-------C---T---  8506
H1K.CM.96.MP535            T-----A-G-A---------A---------------G-T--T-----------T-T-----------G---A---------------------.....................A-------C-------  8395
H101_AE.TH.90.CM240        T-----A-G-A---------G--------CT-A---G----TA-T----------T-----------G---A---------------------.....................A------TC------T  8776
H102_AG.NG.-.IBNG          T-----A-G-A---------A---------------G----TA----------A-T------A----G-------------------------.....................ACT----TC-------  8735
H103_AB.RU.97.KAL153_2     T-----A-G-A---------A---------------G----T-------------------------G----.........------ATTA-A...........................-TC-----A-  8373
H104_cpx.CY.94.CY032       T-----A-G-A---------G---------------G----T-----------T-T-----------G-G-T--------------A------.....................-AG----TC------T  8620
H1N.CM.95.YBF30            TGG---AG--A---G--------T--A--C--C-----G--T------------AT-C-------C-G---T--------------------A.....................A-T-----C--CG-A-  8798
H1O.BE.87.ANT70            T-------T-A----GC---A-----G-----T---G-G--TA----T--G--A-T-----------G------------------A-----A.....................AC---G-TC-------  9295
H1O.CM.91.MVP5180          T-------T-AG---GC---A-----G-----C-----A--T-----T--G--A-T-----------G----TGT--------G--A-----A.....................CCT----TC-------  9318
CPZ.CM.04.EK505            TGG--AACC-AG--------A--T--A--C--C---G-T--T--T--T-----AAT---------C-G------------T-----A-----C.....................--------C--C----  8715
CPZ.CM.04.LB7              T----TA---AG--------A--------C--CG--G-T--T-----------AAT---------C-G------T-----C------------.....................A-T---------A-A-  8740
CPZ.CM.04.MB66             T-----A---AG--------G-----------T---G----T--T--------A-T-----------G---A---------------------.....................A---------------  8853
CPZ.MT145                  TGG--AAGG-A------G--G---------T-A---G-T--T-----------T-----------C-G---A--------C--G--------A.....................--GC-C--C-----A-  8741
CPZ.CD.-.ANT               T-NN-A-TT-A----GC--C------A---A-------GTTA-----C--G--AAT-----------G------T------GA--A------A.....................--T-----------CT  8662
CPZ.CM.-.CAM3              T-----AT--NNNNNNNNNNNNNNN-G----ACG--TA---T--------G----T-----------G------------------------A.....................--------C-----A-  8651
CPZ.CM.98.CAM5             T-----AC--A---G-----G-----A-----C---G-T--T--------G--AATT--------C-G------T-----------------A.....................--------C-----A-  8954
CPZ.GAB_X52154             T-----AGG-----------G--------C--C---G----T-------------T---------C-G-------------A----A-----A.....................ACA----TC-----CT  9281
CPZ.GAB2                   T-T---A-G-----------G-----G--C--C---G-T--T--T--C-----A-------------G-A-T-----T--T--T--A-----A.....................--GC-T-----GT---  8619
CPZ.TZ.-.TAN1              TT---T-C--A----GC--CC-----A--CACC-----G--TA-T-----G--TGT---T--A--C-G------------C--T--A-----A.....................A-------C------T  8894
CPZ.US.85.CPZUS            T-T---AG------------G-----A-----C---G-T-----T--------AAT---------C-G------T-----C--------T--A.....................--A--------C--A-  9247
GSN.99CM166_AF468659       T---GTG-GGAC--G--CC-G-AGT-G--A-AA-ATGC---T-TGGAGTGG---CTGG-G----GA-G--TCTCA-T---T-----C-----A.....................AC-C-C-----TAC--  9028
GSN.99CM71_AF468658        T-T-GTG--GAT-----CC-A-AG--A--GAGC-ATTG---TT-TGAGTGG---ATTG-G----GA-G--TCC-G-GG--T-----A-----A.....................A--C-C--C---ACT-  9025
H2A.ALI_AF082339           T--AGTG-G--------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G----GATAA-T--A----G---------T--T-AT--------A.....................A-A--G--C--GA--T  9626
H2A.BEN_M30502             T--AGTAGG--G-----TAGA-----A--CA-A-ACC-AG-AA-AGAGG----GATAA-A--A----G------T--T--T-AT--------A.....................--A--G--C---A--T  9627
H2A.ST_M31113              T--AGTG-T--G----GTAGAG----A--CA-A-ACT-AG-AA-GGA-G-G--AATAA-TGGA--C-G---------T--T-AT--A-----A.....................--A--G------AA-T  9066
H2B.D205_X61240            T-TAGTG-G--------CAGA--AT-A--CACA-ATT-TG-GA-TGA-G-----ATTG-GT---GC-G---A-----T--T-AT---------.....................A-A--G-----CAAAT  9540
H2B.EHO_U27200             T-TAGTG-G-----G--TA-A--A--A--CACA-ACT-AG-AA-TGA-G-----ATTG-GT---GA-G--------------AT--------A.....................--A--------CAA-T  9514
H2G.ABT96_AF208027         T--AGTG----------TAGA-----A--CA-A-AC--GG-AA-GGAG-----TATAA-T--A----G------T------G----A-----T.....................A-A---------R---  9012
H2U.FR.96.12034            T-T--T--G--------TAGA---T-A---A-C-ACT-AG-AA-TGAGG-G--AATAA----A----G------T-----CT----A-----A.....................ACT----GG---A--T  9108

MAC.239_M33262             T--AGTGC---------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT----A-----A.....................A-T-----C---AA--  9584
Nef                        _Y__S__A__R__R__H__R__I__L__D__I__Y__L__E__K__E__E__G__I__I__P__D__W__Q__D__Y__T__S__G__P__G__.__.__.__.__.__.__.__I__R__Y__P__K_    Nef

SMM.SL.92.SL92B            T--AGTG-G--G-----TAGA-----A---A-A-ACC-AG-AA-AGAGG----TAT-A----A----G------T--T---AGT--------A.....................A-A--G--C---AA-T  9035
SMM.US.x.PGM53             T--AGTG-G--------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-AC---A----G------T-----T-----A-----A.....................A-------C---A--T  9521
MNE.US.-.MNE027            T--AGTG----------TA-A---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-G-G--A----G----G-T-----CT-G--A-----A.....................CCT-----C---AA--  9061
RCM.NG.x.NG411             TT---AG----G---AGGCGC--TT-A--C--C-AC--GG-GA-TGA---T--AATAA-A--A----G------T-----T------------.....................A-A---------ACA-  8825
RCM.GAB1_AF382829          TGGAGTATC---------AGA--AT-G---A---ACC-GG-AA-TGAG--T---ATAA-A-------G---A--------T-----------A.....................A-A---------ACAC  8742
DRL.FAO_AY159321           TGG--T-TC------G--ACA---T-G-----C-ATGCTC-AA-TGA-TGG----T-A-AAAA-GC-G---A-GT-----CAAG--A--T--C.....................A----G--C---AA--  8920
MND-1.F17                  T-TAG-AGG------G----A-----A-----C-ATGCAG-GA-TGAGTGG--A-TTGAA----GA-G---AC-G--T---GA---T-----A.....................AC---------TAA--  8644
MND-2.5440_AY159322        TGG--TA--------G----G--T-G---C--T-ATGCAC-AA--GAGTGG--ACTAA-A---A---G-------------AAG--A--G---.....................A-A--G-----GAA--  8835
MND-2.14CG_AF328295        TT---TAG-------G----G--T-AA--C----ATGCTC-GA-TGA-TGG--A-T-A-A--A----G---AG-A-----TT-----------.....................A-A-----C--GAA--  9080
MND-2.MND2_AF367411        --T--TA--------G----G--T-G---CT-A-ATGCTC-AA-TGA-TGG--A-T-A-T---ACC-G---ATCT--T--TGAT--C------.....................A-A-----C--GAA--  9139
DEB.CM.CM40                T----TGCC-----G--T-C---TT-A--G--TCATGC-C-GA-TGA-TGG--AATTA-A--A-GC-G--TAC-G------AAG--A--T--A.....................--T--G------A-TT  8849
DEB.CM.CM5                 T----TGCT--------T-C---AT-A--AA-CCATGC-C----TGA-TGG--AATAA-A--A-G--G-TT---G------GA---C-----A.....................CCA--G------ACAT  8776
MON.CM.L1                  T---GTG--GAC-----TAGA-AGA-A--G-AA-ATGC---T-TGGA-TGG----TAA----A-GA-G--TAC-A--T--GGA----------.....................-----G--C--GACT-  9063
MON.NG1                    T---GTGTWGACC-Y--TC-A-AGT-A--G-AA-ATGC---TGT--TGTGG--ACTTG-G----GA-G--TAC-G-TT--CY-C--A-----C.....................--GC-G--C---ACC-  7700
TAL.CM.8023                T--AA-A-C-----G--T-CA---T-AA-C----ATGC-C-TA-TGAGTGG---ATAC-G--A----G----TCA------GA---------A.....................ACA--G--C---T-AT  8522
TAL.CM.266                 T--AA-A--C-T--G--T-CA---T-GA------ATGC-C--A-TGA-TGG---ATAC-G--A--C-G-------------GAG--A-----A.....................ACA-----C---T--T  8972
MUS.CM.1085                GT-AGTAC-GA---G-TCC-G-AGT-G--GACC---GC---TGTGGAGTGG--AATAA-A--A-GA-G--T--TG--T---GAT--C------.....................ACTC-G-----CAA--  9034
SAB.SN.-.SAB1C             T--AGTG----G---A-GA-G---T-G-----C-ATGCTCTTA-TGAGTGG---ATAG-AGA--GA-G------T--T---GAT--------T.....................AC---G--C---AA-T  9392
GRV.ET.x.GRI_677           T----TG--------G--A-G-----GA--T---ATGC--TGA--GAGTGG--AATAA-AGA-----G---AGCT---T-------C--G---.....................A-A--G--C--GAGAG  9059
TAN.UG.x.TAN1              TGG--A-C--AG---G-GC-A-----GA----C-ATGC-CTAA-TGA-TGG---ATTA-AGA-----G---AGCA--TT-C--T--------C.....................AC---GA-G--C-GAT  9202
VER.DE.-.AGM3              T-----G-C--G--GA-TC-G-----AA-C----ACGC-CT-A-TGAGTGG--AAT-A-TGA-----G-A-TGCT-GGT--GA---A-----A.....................A------C---AGAT  8713
VER.KE.-.9063              T-----G----G--GA-TAGA-----CA-C----ATGCTCTTA-TGA-TGG--AAT-A-TGA-----G-A-CGC--GGT--A----A-----C.....................--A--------C-GCT  9215
VER.KE.-.AGM155            T-----G-T--G---A-TA-G------A------ATGC-CT-A-TGA-TGG--AAT-A-TGA-----G-A-CGCA-GGT--AA---A--T---.....................A-A-----C--GAG-T  9192
VER.KE.-.TYO1              T-----G-T--G---A-TA-G-----GA------ATGCTCTTA-TGA-TGG--GATAA-TGA-----G-A-TGC--GGT-GAAG--A-----A.....................A-A----TC--TAAAT  8669
SUN.GA.98.L14              T-----AG--CC---G----A---T-G---T---ATGCAG-AA-TGAGTGG----T-A-AA---GC-G---AG-T-----CAA------T--A.....................--G---------AA-G  9360
LST.CD.88.447              TGG-----G--------G--C---T-ATG---T-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT--CT------GAT---------.....................ACGC-T-TC---T-AG  8214
LST.CD.88.485              TGG-----G--------G--C---T-GTG---T-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT--C-------GAT--A------.....................ACAC-T--C---T-A-  8211
LST.CD.88.524              TGG-----G-----GA-T--C---T-AACA--C-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A.....................ACTC-T--C-----CG  8214
LST.KE.-.lho7              TGG-----G--------G--C---T-AAT---C-ACTGTG-GA-TGAGTGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A.....................ACA-----C--T----  9289
SYK.KE.-.SYK173            T-T-GTG---AG---G-T--A--T---C----C-ACC-AC-GA-TGAG--T--AAT-A-A...--CAG-ATC--T-----CT----------A.....................AC-C-G--C--GT-A-  9037
SYK.KE.KE51_AY523867       TT--GT----AG-----T--G--T--AC-G----ATT-AA-GA-TGAG--T--TGTGA-AGA---C...ATC-C---T--GAGT--------A.....................ACA-----C--G----  8727
COL.CM.-.CGU1              AGAA---CTGAG---G-T--CC-GA-A--G-A--ATGC----ATTGAGTGG----G---AAAA-G--G-TT-G-G--TGA-GAT-A--TG--AGAGGATGGAGCATTAAAAGAG-A----A-G--T---G  8468
```

```
H1B.FR.83.HXB2              CCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGATAGAAGAGGCCAATAAAGGAGAGAACACCAGC.....TTGTTACACCCTGTGAGCCTGCATGGGATGGATGAC..........  9321
Nef                         T__F__G__W__C__Y__K__L__V__P__V__E__P__D__K__I__E__E__A__N__K__G__E__N__T__S__.__.__L__L__H__P__V__S__L__H__G__M__D__D__.__.__.__.   Nef

H1D.CD.84.84ZR085           -------G-------T-G---------------T-----GGT-G-----A------C-G-------AG---A-T--.......---C--------A-CT---A------A-----G---..........  8869
H1F1.BE.93.VI850            ----G----------T-----------------T-----AG--G-G---A--------G------------A-T--.......-----------CA------AA-----A-----G--T..........  8598
H1G.SE.93.SE6165            ---------------T---A---------A-G--T----CAG--G----G--A-----------------A---T......C-A--------CA-CT---A------A-----G---..........  8763
H1H.CF.90.056               -------G-------T--------------AA-T---C-GG--G-----C--------G-----------A---.......---C-------CA-------------A-----G--T..........  8644
H1J.SE.93.SE7887            -A-------------------------------T---AGCG-AG----G--A--T---G-------A----A-T--......C----------CA-AT---A------A--A--G--T..........  8619
H1K.CM.96.MP535             -------G-------------------------T----CAG-AG--------A---CAG--------G---A-T--......C----------CA-A-A--A------A-----G--T..........  8508
H101_AE.TH.90.CM240         GT-------------T-----------------C-A-AGAG-AG----G----A---C--------A----A-T--......C----------CA------A------A--A--G---..........  8889
H102_AG.NG.-.IBNG           -------G-------T---A---------A-G--T----CAG--G----G--A-----G-----------A---T.......--A--------CA-CT-T-AA-----A-----G---..........  8848
H103_AB.RU.97.KAL153_2      -A-----G---------------------C--C----C-G-AG------------CAG-G----------A---.......C-A----------A-AT---AA--C--A--------T..........  8486
H104_cpx.CY.94.CY032        G--------------T---------------A--T---C-GG--G-G----------C-G--------------T-T.......---C-G------A-A----A------A-----G--T..........  8733
H1N.CM.95.YBF30             ---------------T---A---------T-GTCAG-T--AG-AG----------T---G-------C----ATGC-......C-C---------CA-AT-T-AA-----AGCA-----T..........  8911
H1O.BE.87.ANT70             -A----------TG-TT--A------------GTCAGA---AG--GC----AGACTAGGA--TAC-TGTG-G-GGGCT...AATC-CC-G--T--A-CAT-TGCC-----AT-T--A--T..........  9411
H1O.CM.91.MVP5180           -A----------TG-TT--A--G---------GTCAG----AG--GC---GAGACTGGG---TAC-A-TG-AGATGCT...AGTC-TC----T--A-CTT-TAAT-----AGCT--G--T..........  9434
CPZ.CM.04.EK505             ---------------T-------------C-GTCA--T--GG--G-------A--T---C-------C----ATGT-......C-------T--CA--T---AA-----A-----A--T..........  8828
CPZ.CM.04.LB7               -------CG------T-------------C-AACC-----AG--G-G--G-----T-----------------ATCTA......C-AC-G-----AA-AT---A------A-----A--T..........  8853
CPZ.CM.04.MB66              -------C-------T---A---------T-A--A-----AG-AG-G---AGA--A---G-------C----A--TA......---------T--CA-CT---A---C--ACAA--G--T..........  8966
CPZ.MT145                   ----C--G-------TT--A--G--C--TT-GCCA--T--GG--G-G--GA-A------G-------A----A-T--......C--C----T--CA--T---A------A-----G--T..........  8854
CPZ.CD.-.ANT                GTAGA---------T-TT---T-----------A--AT--CCACCTGAT...............-AT----GG-A--TA......C-CC-C-------CTT-TACA-----AGAT-GA---..........  8760
CPZ.CM.-.CAM3               -A-A---GG------TT--A--G--T---C-GACA-----AG-AG-G---C--------G-------T--T-A--T-......C-AC-G-----CA-CT---A------A-----A--T..........  8764
CPZ.CM.98.CAM5              -A-------------T------G-----TC-GACA-----GG-AG-G--GA-A-----CG-------C-C--AT-TA......C--C-T-----CA-AT---AA-----A-----G--T..........  9067
CPZ.GAB_X52154              GT-------------T---A-----G--CC-GACAGAG--GC--G-----C-A------G-------T----A-T--......C----G--T--CA-AT-T-A------------A--T..........  9394
CPZ.GAB2                    ---A-----------T---------G-----ATCA-----GGCAG-------A-----C--G------G-T...-ATCTC...C-A-----T--CC--T--ACT-----AT-T--G--T..........  8732
CPZ.TZ.-.TAN1               GTAGG-------TA-TT---T-G-----G--A--C---CCAG-AGAT...........G-T-AGA-----.........ATC---C----T--A-CCT-TAGC-----A-CTACC--T..........  8992
CPZ.US.85.CPZUS             -A-A---G-------TT---T----C--TC-CACAGA---GG-AG---GC-----------------C--A--TA......C--C-G-----CA--T---A------A-----A--T..........  9360
GSN.99CM166_AF468659        TACCA----TC---AT-TGT--GCGG-----A-CCA--AC-G--GACTCT---C-TGGGG-T-AT--AT-T...CTG......C--ACC-------CCTAT-A-GGAC---GT--A---..........  9138
GSN.99CM71_AF468658         TGCCA--C-TC----GGTGCT--CGG-----G-CCATGAC-G--GACTCT---C--GGGG-T-AT...C-ATATCTG......---AAC--T--C-CCTAT-A-GGACA-CAA--G---..........  9135
H2A.ALI_AF082339            T------G---CTG-GG--------------A--TGTCCCAC-AGA--GG----A--C-G-GACT......CA-T--......C--C-------A--ACAAACAAGCA--CAT------..........  9733
H2A.BEN_M30502              A---C--G---CTG-GG---------T----A--A-TCTCAC-AGAG-C-----AAG--G-G-CC.......-A-T--.......--AG-------A-CACAAACAAGCA-ACAT-----T..........  9734
H2A.ST_M31113               T------G----TA-GG--------------A--TGTCCCAC-AGAG-G---T-A--G-G-GACT......CA-T--.......--AG-G--T--A-CACAAACAAGCA--T-T------..........  9173
H2B.D205_X61240             A------T---CTG-GG-----G--------A---GTGCCAGCAGCGACCCGA-AGG-GG-G-A-ACC...CATT--......C-AA-G-----G-CACAGA-CTCCTCATG-------..........  9650
H2B.EHO_U27200              T------C---CTA-GG--------G---A-AA-TATGATAGCAGA-CC-----A-G-GG--ACC......CATT-T......C--G-G--T--A-CACAGACCTCC-CATG-------..........  9621
H2G.ABT96_AF208027          -G-----C---CTA-GG--A-----------CAC-GTCTC-G-TGAG-C-C---AGG-CG-GACT......CATT--.......---G----T--A-CACAAACTAGCCCATG-------..........  9119
H2U.FR.96.12034             G----------CTT-GG-------A------A--TG--A--G-T-A--C-C---AGG--G-GA-G......CTGTAT......C--G-GGGTT-A-CACAGACCAGCT-CGA---G---..........  9215

MAC.239_M33262              -A-----C---CTA-GG--AT----C--T--AA-TGT-TCAG-TGAG-C-C---AGG--G-G-AG......CATTAT......--AA-G--T--A-CTCAAACTTCCCA-TG-------..........  9691
Nef                         T__F__G__W__L__W__K__L__V__P__V__N__V__S__D__E__A__Q__E__D__E__E__.__.__H__Y__.__.__L__M__H__P__A__Q__T__S__Q__W__D__D__.__.__.__.   Nef

SMM.SL.92.SL92B             T------G---CT--GGC----G--G-----A--CGTCTCGG-TGAG-C-A-CAATG--G-GACC......CATT--.......---G----T--A-CCCAGACAT--CA-CATTCA---..........  9142
SMM.US.x.PGM53              T------C---CTA-GG--AT-G--C-----AAGTG-CTCAG-TGA--CTC-A-AAG-CG-GAC-......CATTA-......--AG-G--T--A-CTCAGA-ATCCCA-TG------T..........  9628
MNE.US.-.MNE027             -A-----C---CTA-GG--AT----C--T--AA-TGT-TCAG-TGAG-C-C---AGGG-G-G-AG......-ATTAT......--AC-G--T--A-CTCAAACTTCCCA-TG-------..........  9168
RCM.NG.x.NG411              TG-----C---CTA-GGC-AT----------A--TGTCAG-G-TGA--C-ACA-AG---G-GA-G......CATT-T......C-CC-G--T--A-CAGAAACTAG---------C---..........  8932
RCM.GAB1_AF382829           TG---------CT--GGC-AT-G--G-----A--TGT-TC-G-TGA--C-AGA-AAG--G---AGC-T-GT............---C----T--A-CAGAAACAAG---------A---..........  8849
DRL.FAO_AY159321            TG-----G-TCCTG-T------G--C--T--G---GT----G-AGATCTG-TAAATC-GC-GT-TA-C-GG...........C----GA--T-CAGCCAGA--GG-CCT--T------..........  9027
MND-1.F17                   -A-------TCCTG-TT-----G-A------GAGCAG--C--TAGG---T---TATGCAGCT.........-A--ATCAT...C--------T-CTCCCAGT-ATG-CCTCA---A--T..........  8751
MND-2.5440_AY159322         -T--C--G-TC-TG-T------G--C--G--G-CAGTGTCGCCCCCCTG----AGG--G--T-CA-C-GG...........C--C--A--T--TCTCAAA--GGAATCCA---A---..........  8942
MND-2.14CG_AF328295         GA-----G-TC-TA-T----T--A-C-----AC-AGT-CCACCAGACC-G---AA-C-GG-GT-TA-C-GG...........C--C-TA--T-CTCTCAG--AGGAATCCA---A--A..........  9187
MND-2.MND2_AF367411         -A---A-G-TC-TA-T------TTGC--G--G-CAGTGCC-CCAGACC-G---AA----G-GT-TA-C--A............---C-C--GT-CTCTCAG---GGAATCCA---A--A..........  9246
DEB.CM.CM40                 A---C--G-TCCT--TT--A--G--G-----A--AAT--CAG-CCC---TT-T-AG---G-T-AG......-GG-A-..ATC--------TGA--CCCA--A-GGGAT------A---CCCCAT.....  8965
DEB.CM.CM5                  A------G-TCCTA-TT---------G-----A---AT--CAG-TCCT--CT-T-AG---G-T-AG......-GA-AT..ATCC-C-----TGA--CCCA--AAGGACAAGCA--A---CCCCAT.....  8892
MON.CM.L1                   TGCCC--G-TCCTG-GGTGCT-GCGC--G--A-CCA-CACGG--GATTCC----AGGGAG-T-AG--CTT-.........CTCC-AACC--T--G-CCTAT-AAGGCC-------G---..........  9173
MON.NG1                     T-CC---G-TCCTG-GGTGC--GCGT--G--A-CCATCC-GG-AGACTCG----A-GGAG-T-AY---TT-.........CTCC--ACC-----G-CCTA--A-GG-C--GAT--A---..........  7810
TAL.CM.8023                 G---C--TATC-TG-GG-----GTGC--C--G---ATCC-CG--GAC--CAC--AGG-CGG-CAC..............CTCC-CA-G-----A-CATATGACGG-CA-CA---A---..........  8626
TAL.CM.266                  G------TATCCTG-GG--A--CTGC-----A---ATCC--G--GAT--CAC--AGG-CGG-CAT..............CTCC-CCCG--T--A-CATATGATGGGCA-GCA--A---..........  9076
MUS.CM.1085                 TACCA--C-TCCTA-GGTGT--GAG-GA---G-CCATGACGG--GGCTCT----AGGGAG-T-ACC-AT--.........CTCC-AGAC-------CCTTT-A-GGGA-AGAA--A---..........  9144
SAB.SN.-.SAB1C              G--------------TT--A--G-----T--G--C-TGTCAG--GAG-CTA--AAT-G-G-GAAC......CA-T--......C--C-G--T----CTCAAG--GCCTATGA-------..........  9499
GRV.ET.x.GRI_677            T------C-TC----TT---------C-----G--C-TGC--G--GAG-C-CGCAA-TG-G-GA--......CA-T-T......C--A-G--T--A-CACAGA--GGG-AAGATCC---T......GGAAT  9171
TAN.UG.x.TAN1               G---C--C-TT----TTG--T----G--G--A--TGTCTCCC--GA--CCC---AT...G-GA--......CA-T-T......---C-C-------CCCAGA-TG-AT--GAAAGC---..........  9306
VER.DE.-.AGM3               G---C--C-TC----TT--AT-G--------A--C-TGC--G--GA--C----A-TTG-G-GA--......CATT--......C--G-G--T--A-CACAAG--AGG-AAGACCC---TGGAATC.....  8826
VER.KE.-.9063               G------C-TT----T-----------G--A---ATGC--G-AGAG-C----A--TG-G-GA--......CATTA-......---GCG--T----CACAAA-AA-A-A-GACCCC--TGGCATC.....  9328
VER.KE.-.AGM155             G------C-TC----T-----------G----CC-TGC--G--GA--C---AA-ATG-G--A-G......CATT--.......---G-------A-CACAA-------AAGACCC---TGGTATA.....  9305
VER.KE.-.TYO1               G------G-TC----TT---------G-----G--CTT-C--G--GA--C-C-AA-ATG-G--A--......CATT--......C-AG-C--T--A-C-CAGA--GGA-AAGATCCA--TGGTATC.....  8782
SUN.GA.98.L14               -------C---CTA-GG-------C------CACCATT---G-AGACAGG--TC-T---......C-TCCTTGTCAAGCCCTAC--CATTC-AGCCA-CAAGG-GT-AATGAA--C............  9470
LST.CD.88.447               -A----------TG-GGC----G-A------G-CCTGT---G--TACAC---TC--TCAG-TTACAGA...CAGT--......C--C-G---T-AAGTCAG---GGG-TCCA---G---..........  8324
LST.CD.88.485               -A----------TG-GGC----G-A------G-CCTGT---G--TACAC---TC--TCAG-T-CCAGC...CAAT--......C--C-G---T-AAGCCAG---GGG-TCCA---G---..........  8321
LST.CD.88.524               -----------CTG-GGC--T-G-AG-----A-CATGT---G--T-TA----TC--TCAG-T-AGAGA...CAAT-T.......---C----TT-AAGCCAGT--GGG-TCC-A--A---..........  8324
LST.KE.-.lho7               -A----------TG-GGC--T---A------G-CTTGT---G--TATA----TC--TCAG-T-AGACA...CAAT--......C-T--G--TT-AAGCCAG---GGG-TCC----A---..........  9399
SYK.KE.-.SYK173             T-----------TA-GGG-A-----T---AA----AT---GGGATGCTTG---TATG-GG-GCATACA.............CTCC-CC-C--T--A-CCTCTGGA--G----GTTCGAG-......ATGGG  9149
SYK.KE.KE51_AY523867        T---C--G----TA-GGG-AT-G-CC--TAA---AAT---AGG-TACTTG-TA-ATG-GG---ACACA.............CTGA--A-G-----A-CAGCAGGAGTG---GCTTCA--A.........GA  8836
COL.CM.-.CGU1               TTGC-------CTT-GG--AT----CTACA-A----A-CT-GGAGAGT-T-CATA--GCT-T.......................................................................  8530
```

```
                                                                                                                                   Nef end \
H1B.FR.83.HXB2         .CCGGAG...AGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCAC...GTGGCCCGAGAGCTGCATCCGGAGTAC...TTCAAGAACTGC...TGA.....................  9417
Nef                    __P__E__.__R__E__V__L__E__W__R__F__D__S__R__L__A__F__H__H__.__V__A__R__E__L__H__P__E__Y__.__F__K__N__C__.__*                       Nef

H1D.CD.84.84ZR085      .GA----...-AG---------T------A---A-----A-A--------G-G---...AA----AA---AAA-T-----------...------------...---CATCGAGCTC...........  8975
H1F1.BE.93.VI850       .GAA--C...---------AC-GAG-----A----------A----G---C-GAGA---...A-A---A-----AGA---------T-...-A-C-AG----A.......................  8691
H1G.SE.93.SE6165       .GAA--C...---------C-G-T------A--------TA-------CGGAGA---...A-A-----------------------...-A---AG-----...---..................  8859
H1H.CF.90.056          .GAC-G-...---------C-GAT-----AA--------T--A--------GACA---...T---------TAAA--------------......--AG-----...---..................  8737
H1J.SE.93.SE7887       .GAA--A...------------C------A---------TC-------CGGAGA---...A-A--------------------TT...-A---AG-----...---..................  8715
H1K.CM.96.MP535        .GAAC-C...------A-A---AT-----A---------TTC-------CGAAGA---...--A------------------C---...-A---AG--.......................  8598
H101_AE.TH.90.CM240    .GAA--A...---------C-GAT-----A---------TGC-------CGAA-A---...--A--------A-A------A-------...-AT--AG-----...---..................  8985
H102_AG.NG.-.IBNG      .GAC--T...---------C-GATC-----A--------TA-A--G-----CAGA---...ACA--TA--------------------...-A---AG-----...---..................  8944
H103_AB.RU.97.KAL153_2 .GA---A...-A---------AT-----A---------T-----G--TC-AACA---...AGA---------------------T-...-A---AG-----...---..................  8582
H104_cpx.CY.94.CY032   .GAA---...--------A---A------A---------T-----G---ACA-G---...--A---------------------TT...-A---AG-----...---..................  8829
H1N.CM.95.YBF30        .GATC-T...-A---------G-T----C-A---------TC-------AGAAGA--T...--A--AA-------------------TT...-A-----------...---..................  9007
H1O.BE.87.ANT70        .A-AC-T...-A----A-AC-GAT-----A------T--ATCT----GCAACACC--T...--T--TATGATAACT--C--A---CT-...---C----GGA-...-A-..................  9507
H1O.CM.91.MVP5180      .G-AC-C...G-G--GA-AC--A-A---CA------T--ATCAT---GC--AACA--T...A-A----TGC-AAA---C--A---CT-...---CCC--G-AA......................  9527
CPZ.CM.04.EK505        .--A--C...-A-------C-G-T----C-C---------TC-------AGAGTA---AGA-CAAGA...----------------T-TAC...................CAGAAC  8918
CPZ.CM.04.LB7          .GAA--C...-A-------C-GATA----AA-A--------AG--C--CC-AAGA---ATA-CTCGA...----------T--------TAC...................AAGAAC  8943
CPZ.CM.04.MB66         .GAA-CA...---------C---TC----CA-----------G--T---C-GA-G---AGA-CTCGA...-----A------------TAC...................AAAGAT  9056
CPZ.MT145              .GA---A...--------AC-G-TA----CT-----T---AAG--G--TC-CAGA--TCTA-CAAGA...---AAA---------A--TTAC......................AGA  8941
CPZ.CD.-.ANT           .--AC-T...-A---GA-C---AGA---GAA-----TGCAA----GATGAGAAGA---...A-T--AA----AAGA--C--T--A--T...----GAG---AAGAC--C....................  8859
CPZ.CM.-.CAM3          .G-AC-T...G-------C--G-TC-----A---------A-G--G--C--AAGA---...A-A---A----A-AA-----------T...-A-C-------A.......................  8857
CPZ.CM.98.CAM5         .GAAC-T...G-------C----TC-----A---------A-G--G--TC-GA-A---...C-A---A----A-AA-----------T...-ATGGAG--CA-AACCC-GCGCCTGCGCAGTAGAACTGC  9187
CPZ.GAB_X52154         .GA---C...-A---G--C--G-TC---C-C---------A-G--G--C--AAGA--T...A-T---A----A-AA-------------...-A---AG----ACTT-AGACTGGCGCATGCGCACAAGAA  9514
CPZ.GAB2               .GAA---...-AG--G--TC-GATA----A---------T-AA-----CC-G-GG--T...C-T--AA----AAAA-----T-----T...-A--GAG----A...AAC...................  8828
CPZ.TZ.-.TAN1          .--A--T...G----GACTC-GATC---C-C---------A---------AGAAGG---...A-A---A----AAGAT--------------...------AT-AG-ACTTCCGGGTGCCATGA..........  9101
CPZ.US.85.CPZUS        .GAAC-T...G-C-----C---ATC---CA--------CTGAA--G--TCGGAGA---...AGA--TAA-----------A---------...---CG------ATGGC...................  9457
GSN.99CM166_AF468659   .-A-C-T...-AG--GT-C--G-TC-TCTCC--CTG-TC-AAG--G--CA-AA-GTCAGGCA-ACAG-T---C-A--TG-A-C--G-G...GAGCG---GATGCGCCT-ACCGCAAACCGCTTCCTATAG  9261
GSN.99CM71_AF468658    .-ACC-T...--G--GA-C--G-TC-TCTCC--CTG-TCTA-G-----C--GA-ATCAGGCTG-CAGATGA-C-A--TG-A-C--G-G...GAGCG---GAAGCGCCT-ACCGCAAACCGCTTCCTATAG  9258
H2A.ALI_AF082339       .A--C-T...G-G--GACA----TT-----A------CCTAAG--G--TCA-G--T--...AAA---TTTATT--A--C--A---G-A...--TGG-T--AAG...-C-GGCCTGCCAGAAGATGAGTGG  9850
H2A.BEN_M30502         .GA-C-T...G-G--GACA----T----CA-------TC-ATG--G--C-A-A-CT--...AA----TTCACT---T-C--A--AG-G...--TGG-C--AAG...-C-GGATTGCCAGAGAAAGAATGG  9851
H2A.ST_M31113          .---C-T...G-----ACA----TT-----------CC-ACG-----T---AGCT--...-A----TTTATT-GAT-C--A---G-G...--TGG-T--AAG...-C-GGCCTGCCAGAGGATGAATGG  9290
H2B.D205_X61240        .ATCC-T...G-G--GACTC-TATC---CA------TTC--T---G----A-G--T-T...------TTTCA-TAG-TT---A--AG-G...---GG-T--CAG...-C-GGGCTACCAGAGGAGGAGTGG  9767
H2B.EHO_U27200         .--CC-C...GA---GACCC-T-TC---CA-------TC--T--------A-G-CT-T...------TTCAGCAG-TTC--A---G-G...--TGG-T-TCAG...-C-GGAATGCCAGAGAAAGAGTGG  9738
H2G.ABT96_AF208027     .-------...-CC--------G-C-----A-------CCTACAT----C-A-G-CT--...AG---ATTTATAT-------T---G-G...----G-TGGAAG...-C-GGGCTACCAGAAGCAGTATGG  9236
H2U.FR.96.12034        .-ACTG-...G----G-CC----TA----A------TTC-TCT--G----A-AGCT-T...CA---ATTTATTAA-TGC--A--AG-G...--TGG--GTAAG...-C-GGTCTGTCAGAGGAAGAGGTT  9332

MAC.239_M33262         .--TTG-...G----G--TC---CA----A------TCCAACT--G--C-ACAC-T-T...-A---ATAT-TTAGAT-C--A--AG-G...--TGGA-G-AAG...-C-GGCCTGTCAGAGGAAGAGGTT  9808
Nef                    __P__W__.__G__E__V__L__A__W__K__F__D__P__T__L__A__Y__T__Y__.__E__A__Y__V__R__Y__P__E__E__.__F__G__S__K__.__S__G__L__S__E__E__E__V   Nef

SMM.SL.92.SL92B        .--TTG-...G--A-G--CC---CC----A----A-TCCT-A------C-ACACAT-T...-AA--ATTT-TAAG---C--A--AG-G...--TGG-TGGAAG...-C-GGCTTGACAGAGGAAGAGGTT  9259
SMM.US.x.PGM53         .--CTG-...G----G--TC-G-CA----A------TTCA-AA-----C-ACAGAT-T...-A---TTTCATTAGA--C--A--AG-G...--TGGC-GTAAG...-C-GGCTTGTCAGAGGAAGAGGTA  9745
MNE.US.-.MNE027        .--TTG-...G----G--TC---TA----A------TCCAACT-----C-ACAC-T-T...-A---ATATATTAGAT-C--A--AG-G...--TGGA-G-AAG...-C-GGCCTGTCAGAGGAAGAGGTT  9285
RCM.NG.x.NG411         .--CTG-...G-G--G--A----CA----A----A-ACC--T-T-----G-AG--T--...-CA-GGTATAGA--------T----T-...--TGGAG-GA-A...AAGAACAAGACCAAGTAA......  9043
RCM.GAB1_AF382829      .--ATG-...G-G--G--C--G-CC----A----A-TCCTATG--G---G-AG--T--...A-A-G-TATAGA--------A----T-...--TGG-G-AA-GAAGAACAAGACCCAGTAACCACATCCT  8969
DRL.FAO_AY159321       .--A--A...G-----AGAC-CAT----CA------TCCAG---G----A-ACATTTCAG-CTTGT...ATT-AC-----A---G-G...--TGGAC-TGTTAAAA-TATAAAGTCAGTAGAAGAGCCA  9147
MND-1.F17              .--A--A...G----GACCC-CAT----TCTGGGAC-CT-AT-T-GC-TA-GAC-TTG...................................C--C-T-AACATATCACCCTGAGTGTTTCAATAAGG  8841
MND-2.5440_AY159322    .--A---...G----GAG---GAT-----A------TTCAG-G--T--T-A-ACGTT-...TAT--T-CTATAA-TAGG--A---G-A...-A----TGTGTG...AC-AGTCTGAGCTATGAAGCCTAC  9059
MND-2.14CG_AF328295    .T-CAT-G..G-G--GAG-C-GAT-----A-------CCTGAAT-G--T-A-ACATTT...TAT----CCATAAAATGC--AA--G-A...-A--G-C-TGTT...AC-AGCCTCTCCTATACTGCCTAC  9305
MND-2.MND2_AF367411    .T-CTT-..GG----GAG-C---T-----A-------TCTG---T--T-ACACCTTT...TAT----CCATAAA-AGA--A-GTG-A...------C-TGTG...CA-AGCATCAGCCATATTGCTTAT  9364
DEB.CM.CM40            ..........-----GAG---G-TT----A----------ACT--G--C-A-TG-T-T...AA---AG-CC-TGCTG-A-AAAGAG-G...CAT-CT-GAA-ATGTATGTTCCCTAAGAGGAAGTAG...  9076
DEB.CM.CM5             ..........-AG--GAG-C---TA----A------TTC-TC---T--C-A-TGCT-T...AAA--AG--C-TGAAGCC-ATA-ACCAGAGCAT--C-GAA-ATGCATGTTCCCTAAGAGGAAGTAG...  9006
MON.CM.L1              .--CC-T..---C-GT-C--G-TC-TCTCA--CTG-TC-AAG--G---G-CA-GAG-GGCAGACAG-T--CA-A--T--A-C-AG-G...GAGCG---GAAGCGC-T-GCCGCAAACCGCATCCTCTAG  9296
MON.NG1                .--CC-T...-AGC-GT-C--G-TC-TYTCC--CTGTTCTAAG--G-G-G-CA-GAG-GGCCCTCAG-T---TGA-TTA-A-C-AG-A...GAACG---GA-GCGC-T-ACCGCAAACCGCATCCTATAG  7933
TAL.CM.8023            .--CTG-...G-C----CCC-G-TC---GTC-----TGAGAAG--G--C-ACACC-CAGGA-C-AAAATG-CTGA-T--GACAG--TGGAGAGG-----GAAGGAGCTGCTAGCTCCACCGCAAACCGCA  8752
TAL.CM.266             .--ATG-...G-----TCCC-G-TC---GTC-----TGAGAAG--G--C-ACACC-C-GGA-C-AAGATG-CCGA-TGGGACAG--TGGAGAGAG----GC-GATGCTGCTAGCTCCACCGCAAACCGCA  9202
MUS.CM.1085            .-ACC-T...-AGC-GT-CC-G-TC-TCTCC--CTG-TC----T-G--CC-GA-GTCAGGCT-ACAA-T---C-AAATG-A-C--G-G...GAGCG---GAAGCGC-T-GCCACAAACCACATCCTATAG  9267
SAB.SN.-.SAB1C         .G-CTG-...-A---GACCC-G-T-----A-------CCA-TA------G-GG-CT-T...--T--TT-GAG------C--A---C-G...G-GCC-TCAGCA...CA-GGCTAA..............  9601
GRV.ET.x.GRI_677       AGATC-T...G-------C--G-TC----A-------CCGAAGT-G--GG-GG-GT--...CGCC-GGACAT-T-TA-GGACAT-C--...GAAC-TGCAAAG...C-CTAG................  9271
TAN.UG.x.TAN1          .--TTG-...-A---GACA--G-T-----AA---A-TCCA-T-------G----GT--...AACC-GGACAGCT-TA-GGACAT-C-T.....GGTCTTGTT...AAGAGGAAGTAG...........  9408
VER.DE.-.AGM3          .AACC-T...G-------C--G-TC----A------TCC-ATGT----G-C--AT-T...-ACC-TAACAGAAAAT---TCACTG--...A-GC-TG-TCTT...G-CAAGAGGAAGTAG.........  8934
VER.KE.-.9063          .AGTC-T...G----GACC--G-TC----A-------CCTATGT----G-G--GT-T...-ACC-AAACAGA-A-T--TTT--AG--...A-GC-TGCAATA...GTGAAGAGGAAGTAG.........  9436
VER.KE.-.AGM155        .AATC-T...G-----A-A--G-CA----A------TCCAATGT-G--TG----GT--...-ACC--TC-AG-GA-T-CTTTACAG--...--AT-TTCAACA...GTTGGTACAGGAAACTAG......  9416
VER.KE.-.TYO1          .AGCC-T...G----GA-C--G-T-----A------TCCTATGT-G---A-A--GT--...-ACC--AATCG-GA-T-CTTTACTG--...A-GC-TGGGCTG...GTGAAGAGGAAGTAG.........  8890
SUN.GA.98.L14          .--TTG-...G----GAGAC-TATC----CA-----TCC-ACT--G--C-A-G--TTT...AG----AT-C-AAA---C--T--AG-G...------C-TGTTACTA-TCTGCAGTGGGAGGTGTAGCTT  9590
LST.CD.88.447          .--CTG-...G-----AG-C-TATC---CAC---A-TCCTATG------G-GG-CT-T...--A----TCAGAAAA--A--TA-AAG-...A-AC-AGCA-CAGCTATTGGCTTG..............  8429
LST.CD.88.485          .--CTG-...G----GAG-C-TATC---CAC--CAGTCCTATG------G-GG-CTTT...--A----TCAGAAAA--A--TA--AG-...A-AC-AGCA-CATCTCTTAGCTTG..............  8426
LST.CD.88.524          .--CTG-...G-----AG-C-CATC---CAC---A-TCC-ATGT-----G-GG-CT-T...--A----TGAGAAA---A--A-CAAGT...G-TC-AGCAACTGCTATGAGATTT..............  8429
LST.KE.-.lho7          .--CTG-...G----GAG-C-TATC---CAC--CA-TCCAATGT-G---G-GG-CTTT...A-A----TCA-AAA---G--T-C-A-G...A-AC-A--TACTGCT-TCGCTTTT..............  9504
SYK.KE.-.SYK173        AGAACCTCATGTG---C-TCA-CCCCCACCTGGCT-T-CA-CAGGTTGG...G-GATGGCCCGCTTG-AGTTAGA-AGA-AAACAGGA...AAACCAC-AGAGCTGCAGTCAGCACTTAGTAAGAACATC  9273
SYK.KE.KE51_AY523867   C--AC-T...--G--GAATC-CAT-----AC--CA-TCCA-A-T----C-A-ACA-CAGGCTG--AAATG-CT-GT--G-AATTAG-GAGACAA-CAGGAAAGCGC-AGCTCACCGCAGAAACCACAGGG  8963
COL.CM.-.CGU1          ....--C...TTGAGTC----GAGCGTT-CAAGCAGG-AGAAGAAGAA-CCA...............-A-C--G-TGCCATA---ATG...G-GG-CTGA.............................  8605
```

```
                                                                                                                                                                      NF-kappa-B-II
H1B.FR.83.HXB2             ........................................................................................................CATCGAGCTTGCTACAA..............GG  9436
H1D.CD.84.84ZR085          .........................................................................................................................................  8975
H1F1.BE.93.VI850           ..........................................................................GACTGCTGACACAGAGATT.GCTG.ACACAGAA--ATC-..A-A-................--  8734
H1G.SE.93.SE6165           .....................................................................................................CACAGAA-TTGC--AC-AG...............--  8880
H1H.CF.90.056              .....................................................................................................CACAGAA--T.C-..G-A-...............--  8755
H1J.SE.93.SE7887           .....................................................................................................CAAAGAA-TTTC-..AG-G...............--  8734
H1K.CM.96.MP535            .........................................................................................................................................  8598
H101_AE.TH.90.CM240        .....................................................................................................CAAGGAA-TTTC-.ACT.................A-  9003
H102_AG.NG.-.IBNG          .....................................................................................................CACAGAA-TTGC--AC-.................--  8963
H103_AB.RU.97.KAL153_2     .....................................................................................................CACAGA-TGCTGACA-GA-GTTGCTGACA.....--  8613
H104_cpx.CY.94.CY032       ......................................................................................................CA--GAAGTTGCTGACA--..............--  8850
H1N.CM.95.YBF30            ..................................................................................................CAAGGGAC-TT-CTGCTGACA-...............--  9031
H1O.BE.87.ANT70            ............................................................................................AAACTGCTGACCTGAA--TTGCTGACACTGT............--  9540
H1O.CM.91.MVP5180          .................................................................................................................CTGACACTGC............--  9539
CPZ.CM.04.EK505            TGCTGAC..................................................................................................-A............................--  8929
CPZ.CM.04.LB7              TGCTGAC..................................................................................................-A............................-A  8954
CPZ.CM.04.MB66             TGCTGAC................................................................................AAAGGACTTG.TG....-TG............................--  9080
CPZ.MT145                  GACTGAC..................................................................CACAGCGGCGCATGCGCACTAAGAACTGCTGATCAACAT-----GAC-A.............--  8999
CPZ.CD.-.ANT               .....................................................................................TAACAGGCGGATATAAAAC-GCT--CAGA--AGACTTGTGGTCT......--  8905
CPZ.CM.-.CAM3              .....................................................................................TGCCCAGCG.CCTG.CGCAGTAGA-A--GCTG----..............--  8893
CPZ.CM.98.CAM5             .....................................................................................TGACAA............................................--  9195
CPZ.GAB_X52154             CTGC.................................................................................TGACTCTG..................-.......................--  9529
CPZ.GAB2                   ...............................................................................CGCAACTGGCGCATGCGCTCTAGAA-TG-TGAGCCTGA--TTTTACAA........--  8878
CPZ.TZ.-.TAN1              ..........................................................CTCAGAACTGCTGACAG.AGGAC.TTTT.....GGACTC......................................--  9135
CPZ.US.85.CPZUS            .....................................................................................TGTAACCGC.GCAGGCGCAATA.A-A--GCTG--TG..............A-  9493
GSN.99CM166_AF468659       AACTCTGCCC.............................................................................................................................--  9273
GSN.99CM71_AF468658        AACTCTGCTC.............................................................................................................................--  9270
H2A.ALI_AF082339           AAGGCAAGACTGAAAGCAAGAGGG....ATACCATTTAGT.AAGAACAGGAACAGCTGATTTGGTCAGGGCAGGAAGTAACTACTGAA....AACAGC-GAG.AC---A...............--  9951
H2A.BEN_M30502             AAGGCAAAACTGAAAGCAAGAGGG....ATACCATATAGTGAATAACAGGAACAACCATAC.TTGGTCAAGGCAGGAAGTAGCTA......CTAAG-AAC-.GC--AGG-TGCA............--  9955
H2A.ST_M31113              AAGGCAAGACTGAAAGCAAGAGGG....ATACCGTTTAGCTAAAAACAGGAACAGCTATAC.TTGGTCAGGGCAGGAAGTAACTA......ACAGA-AAC-.GC--AG--TGCA............--  9394
H2B.D205_X61240            AAGGCTAGACTAAAAGCAAGAGGG....ATACCTACAGATTAGGCAAGAGACAGCAGCATA.AACAGGAACTAGCTGACACTGCACAA....GAAGG-AACTAGCA-AC--TGCA............--  9875
H2B.EHO_U27200             AAGGCTAAACTGAGAGCAAGAGGAATACCTACAGAGTAGACAGGAAATAACAGCACAAAGAGGAACTAGCTAACACTGCATAGA........GAAGG-AACTAGC--A---TGCA.............-  9846
H2G.ABT96_AF208027         AAAGAGAAACTAAAACAGAGAGGACTGCCTATAGAATAAGCAGAAAGAGG........................................................AC-AGC--AC---GCA............--  9304
H2U.FR.96.12034            AAGAGGAGGCTAACCGCAAGAGGCCTCCCAGTAAAGAATTGCTGA....................................................CAGAAAGG-AAC-AGC--AG---GCA............--  9405

                                                                                                                                  Nef end \
MAC.239_M33262             AGAAGAAGGCTAACCGCAAGAGGCCTTCTTAACATGGCT...................................................GACAAGAAGGAAA-TCGCT-AA..AC-GC-................    9875
Nef                        _R__R__R__L__T__A__R__G__L__L__N__M__A__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__D__K__K__E__T__R__*                               Nef

SMM.SL.92.SL92B            GAGAGAAGGTTAGCTAACAAGCCAAAACCGCAGAAGAAGATGGCG.............................................GATAAGAAGGAAA--AGCT-AG..AC-GC-...............--  9334
SMM.US.x.PGM53             AAGAGAAGGCTAACCGCAAGAGGCCTTTTAAAAATGGCTGACAAGAAGGAAACAAGCTGA...................................................GA....CAGCA.............--  9814
MNE.US.-.MNE027            AGAAGAAGGCTAACCGCAAGAGGCCTCTTAAAAATGGCT...................................................GACAAGA.AGGAAAC-A-CTGA.-.ACAGCA.............--  9354
RCM.NG.x.NG411             .........................................................CCACTTCCTCTGGAGTTGCCTTGGAAACAGCGCATGCCTACTGAGAAGAT-CTGA...CAC-GAA............--  9107
RCM.GAB1_AF382829          CTGGGGTTGCCTTGGTAACCAGGCAGAAGAATCTGCTGATGCAAAA.........................................................................................--  9017
DRL.FAO_AY159321           GGT..................TGTTGCAAGAGGAAGTGGTGGCAGTTCAAGCCAACCACAGAAGGTTGTCATGGAGACAACCTTCAGAAGTGCTGACTGAGA-..........................--  9234
MND-1.F17                  CTAAGAGTATTGAACATCTGCCATTTTGGAAGAGGAAGTAGCCTAACCGCAAAACCACATCCTACTGCAGAACTGTAGTTGCTTGGCAACCTGCTTAGCAACCTGGACTGG-GCTTGCGCGCTAGGAA--  8971
MND-2.5440_AY159322        AAGAAGGAGGAAAAGCCTGACTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCTAACCACAAAGGGTTGCTATAAGCAACTGCTAACTCTGGA......................................--  9153
MND-2.14CG_AF328295        CAGGAG......AAATCAGACTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCTAACCACAGAAGGTTGCTATGGAGATGCTCTTT.........AATTCA-CTGA...CTCTGGA............--  9404
MND-2.MND2_AF367411        GCTAAGGAGCATAAGCCAGAGTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCCGACCACAGAAGGTTGCCATGGAGATGACCTGCAGAAA...TGCTGA-TCTGG.........A............--  9469
DEB.CM.CM40                ..........................................................CTAGCTTAACCGCAAACCACATCCTACAGCAGCGCCCG...................................--  9116
DEB.CM.CM5                 ..........................................................CTAGCTTAACCGCAAACCGCATCCTCTTGCAGAGAGCCC..................................--  9047
MON.CM.L1                  AACTCTGCCC.............................................................................................................................--  9308
MON.NG1                    AACTCTGCCC.............................................................................................................................--  7945
TAL.CM.8023                TCCTCTTGAAAGGACTC......................................................................................................................--  8771
TAL.CM.266                 TCCTCTTGA.AGGACTC......................................................................................................................--  9220
MUS.CM.1085                AACATTGCCC.............................................................................................................................--  9279
SAB.SN.-.SAB1C             ......................................................................CCACAAAACCACATCCTATGGAGTTGTC.ATGGTGAT--CA--AAG-ACTGCTGACTGAG...A--  9663
GRV.ET.x.GRI_677           ................................................................................................................-GT-AG--C.TTTGCGGTT....--  9291
TAN.UG.x.TAN1              ......................................CCTAACCGCAGGCTTGCGGTTAAGCTAACTAGGCCGTTGCC.TAGGAGATGACATAAAGA-GT-CTGA...-CAGCA.............--  9483
VER.DE.-.AGM3              ...................................................................CTAACCGCAGGCTTGTGGTTAAGCACATCACCATGGTGAT--CA--AAG-ACTGCTGAC.........--  8995
VER.KE.-.9063              ...................................................................CTAACCGCAGGCTTGTGGTTAAGCCGTTGCC.AGGGAGAT--CA--TGA-ACTGCTGACAG.......--  9498
VER.KE.-.AGM155            ...................................................................CCGACCACAGGCTTGCGGTTTCCTGGTTGCC.TAGGAGAT--CA--AAG-ACTGCTGAC.........--  9476
VER.KE.-.TYO1              .................................................................CCAGACCGCAAGCCTGCGGTTAGAACAT.CACC.ATGGAGAT--CA--AAA-ACTGCTGAC.........--  8951
SUN.GA.98.L14              AACCGCAAACCGCATCCTCTTGCATCGCCTAGGCAACG................................................................GGGC-A-C--A---GCGCTAGAGTA.........--  9655
LST.CD.88.447              ......CAGAGGAAGTAGCCAAACCGCAGACCGCAACACATCCTCTTGCAGCCTGGTTGCCAGGGCAACCAGGCTAGCGCATGCGC......AGGA-T-T-CTGA...-GAGCA.............--  8530
LST.CD.88.485              AACTGTAAGAGGAAGTAGCCAAACCGCAGACCGCAGCACATCCTCTTGCAGCCTGGTTGCCAGGGCAACCAGGCTAGCGCATGCGC......AGGA-T-T-CTGA..--AGCA.............--  8533
LST.CD.88.524              AACTGCGAGAGGAAGTAGCCAAACCGCAAACCGCAGAACATCCTCTTGCAGCCCGGTTGCCCTAGCAACCGGGCTAGCGCATGCGC......ACAAGC-T-CTGA...ACAGCA.............--  8536
LST.KE.-.lho7              GACTGCAAGAGGAAGTAGCCAAACCGCAAACCGCAGCACATCCTCTTGCAGCCCGGTTGCTAAGGCAAC...CGGGCTAGCGCATGCGC....AATGGC-T-CTGA...CGAGCA.............--  9611
SYK.KE.-.SYK173            TCCTAGGAG...ACTCCATGGTGACAAGCTCGGCCCAGGGGA...............................................................................................  9312
SYK.KE.KE51_AY523867       TACATCTCCTAGGAGACACCTAGGAGACAGAGACCGCCCAGGGGAGGAGTCGGGGCGGGGGAGGCTCACCCAACCTGCATAAAAGCCGATGCTCTGTAGCTATC-...............................  9068
COL.CM.-.CGU1              .........................................................................................................G-T-CTGA...GAC-GCA............--  8622
```

```
                        NF-kappa-B-II                     NF-kappa-B-I |-         Sp1-III      -|  |-  Sp1-II  -|   |- Sp1-I   -|
H1B.FR.83.HXB2            GACTTTCCG.CTG...................GGGACTTTCCAGGGAGGCG.TG................GCCT..GGG...CGGGACT.GGGG.AGTGGC.GA......G......CCCTCAG.ATCCT   9508
H1D.CD.84.84ZR085         ..................................................................................................................................   8975
H1F1.BE.93.VI850          --------A.---...................------------A-G-CG-TCC................AGAG..--....-------.----.------.TC......A......-------.--G--   8806
H1G.SE.93.SE6165          --------AC--....................----------G--------CG-.................---..---...A--AGT-.----.------.T-......A......-------.-AG--   8952
H1H.CF.90.056             ---------.--....................----------G--------.--A................T--..---...----GA-.----.------.C-......A......-------.--G--   8826
H1J.SE.93.SE7887          ---------.---...................--------------GA-GAG--................----..---...----GT-.----.------.T-......A......-------.--G--   8807
H1K.CM.96.MP535           ..................................................................................................................................   8598
H101_AE.TH.90.CM240       A----.---.---...................--------------GA-GTG--................---G..---...---AGT-.----.---A--.T-......A......-------.--G--   9075
H102_AG.NG.-.IBNG         --------A.---...................----------.---GA-GTG-.................-GT-..T--G..A--AGT-.----.------.T-......A......-------.-C-G-   9035
H103_AB.RU.97.KAL153_2    ---------.-.-...................--------------GA-GTG-.................-GT-..T--G.----AGT-.----.------.T-......A......-------.--G--   8685
H104_cpx.CY.94.CY032      ---------C-C....................--------------GA-GCGC-................----..---...A---GT-.----.------.T-......A......-------.--G--   8923
H1N.CM.95.YBF30           ------ATACT--....GGGACTTTCCGCCA.--------------.A-GTG-.................-GT-..---GGAGT--CT-.-..........................-------.-.G--   9107
H1O.BE.87.ANT70           A-------AG-AAAGACTGCTGACACTGCG..------------T-G-AG-GA........................CAGGGG-C-GT-.C---G------.T-......A......-------A-.G--   9629
H1O.CM.91.MVP5180         --------AGAC.TGCTGACACTGCG......------------C-T--GAGG-AT..............AAGG..--CGG......T-.C---G------.T-......A......-------.--G--   9625
CPZ.CM.04.EK505           -------TA.--TG..................-----------............................G-AGG-ACTTT-C......----AG---TGGTTTGGGGA-TGGCTTG--CTC.AGAG--   9005
CPZ.CM.04.LB7             ------T-TGACCA..................-----------.........................CCAAGGGCTTCCGGG-A-G-GT-TCAAG-GA-GGTTTGGGGA-TGGTTTA--CTCAG-AG--   9041
CPZ.CM.04.MB66            --------A...CA..................---GA-----G.........................GG-AGGCGT-ACCAT---CGGG--A-AG-A-TGGCTT............G--CTCAG-AG--   9152
CPZ.MT145                 --------A......................A-------------------.............AGGGGGGTTGGGTTTGGGGC-TAACT---G-GGAACTGG-A-TGGCTT............G--CA-CA-AG--   9080
CPZ.CD.-.ANT              --------ACTAT...................----..................................-TGA..CTATGGGC-.GGCA-T--G------.TTTC...........----G-CT..G--   8965
CPZ.CM.-.CAM3             -------TAA......................-----------A--G--G.--.................-TTC..A--G..-------.---C.-----T.TTTA...........-------.-AG--   8963
CPZ.CM.98.CAM5            -------TAA......................-----------A--G--A..G-.................-TC..A--G..------A.---C.-----T.TTCAG..........TTT----.-....   9261
CPZ.GAB_X52154            --------AA......................----------G-----AG-.-.................-GTC..---AGG----GT-G---T.------.TTTG...........-------.-.G--   9601
CPZ.GAB2                  --------AA.....................G---GG-GG..............................-TTC.AA--GA.G-A-T-G.---T.------.TTTG...........-------A..G--   8938
CPZ.TZ.-.TAN1             --------AA................................................TGTGGGT.....-GT-..ACTGG.GC--GACA.---G-----T.TTT............G--.-GCTGAG--   9194
CPZ.US.85.CPZUS           -------TAA......................-----------A--GACGTTC...CAAGGGGGTG....-GTC...A-G..GC---ACA---C.-----T.TTTA...........------A.-GA--   9574
GSN.99CM166_AF468659      --------A.............GTCACCATGACTACGGGG--C--TT-CT-AG-CAACCGGGGCGGACTC-TGG............GAGG-ACTGG-A--T.CGGGGAGT-GCTCTA---CGCTAT.--A   9376
GSN.99CM71_AF468658       --------A.............GTTACCATGACTACGGGG--C--TT-CT-AG-CAACCGGGGCGGACTC-TGG............GTGG-ACTGG-A---.C.GGGAGT-GCTCTA---CGCTAT.--A   9372
H2A.ALI_AF082339          --------AGAA-....................--G--G-.A-CCAG--GAAG-AC..............ATGG..-A-GAG-T--TGG.--AA.C-....................---.TCAT-.-TC  10020
H2A.BEN_M30502            --------AGAA....................---G--G-.A-CCA---GAGG.................-A-A..T--GA..---GC.T--T.G-......GGAACG.........------TAC.TTA  10025
H2A.ST_M31113             --------AGAA....................---G--G-T.-CCAG--GAGG.................-A-A..T--GA..---GC.C--T.G-......GGAACG.........------TAC.TT-   9464
H2B.D205_X61240           --------AAAA....................-A-G--G-.A-CA.T--GAGG.................AG-A..A--GA..--GG-.T--T.T-......GGAACG.........---ATTAA-.T--   9944
H2B.EHO_U27200            --------AGAA....................---G--G-.A-CA-T--GAG..................-GAC..AT-GG..A-GAG.CA--.GAG-......AACG.........---ATCAAG.A--   9915
H2G.ABT96_AF208027        --------AGTA....................T--G--G-.A-CA-.--GAG...................GTA..CTATG..---GG.-ACT.G--T-G.-..AACG.........---C.-AA-.T--   9373
H2U.FR.96.12034           --------AC-A...................GA--GAGG-.A-CCA.--GAG..................-AGC..AA-GG..A-GAG.CCAT.G-G-......AACG.........---.TCAA-.TGA   9473

MAC.239_M33262            ...............................G--ACT--C-ACAA-G--ATG-TAC...GGGGAGGTACT-GG..GA--AGC---TCG.G-AACG......................---A-TTTC.TTG   9945

SMM.SL.92.SL92B           --------ACAA..................................G--GAG--TACAA..GGGAGGAGTCT.GG..-CGGGG--A--GG.-AAC.-....................---CT-AAC--TG   9403
SMM.US.x.PGM53            --------ACAAAGGGGATGTTA......................CG--GAGGTACTGGGGAGGAGCTGGCTGGAAC-.......................................---A-TTAT.TT-   9882
MNE.US.-.MNE027           --------ATAA..................................G--GATGTCAT...GGGGGGGTACTGGG..-A-GAG-T--T-G.--AAC-.....................---A-TTTC.TTG   9422
RCM.NG.x.NG411            --------ACTG-CGCCTGCGCGCTGGTGTAA------------ACT-A--TG-GAGGGGGGTGTGGCTT-..............................................-----TCCT.G--   9190
RCM.GAB1_AF382829         --------ACTG-TGCATGCGCACTGGGGAA.----------G--AT-A--TG-GAGGGGGAGTGGTCAG....................................................CCCTCT-C   9094
DRL.FAO_AY159321          --------ACAG-AGCATGCGCTCTGTGGCAG--.-----------T-A--TG-GT..............TGGG..--AGTGGTCTG..............................-------...-TG   9314
MND-1.F17                 --------AA.ACAGGGAG...........GG--AGGC-CG............................................................................---.--TGC-G--   9012
MND-2.5440_AY159322       --------ACTG-TGCATGCGCACTGGACT.G----------.A-ACT-ACG-CGGG.............TGGG..--AGTGGCCAG..............................---AA...C.-TG   9232
MND-2.14CG_AF328295       --------ACTA-TGCATGCGCACTGGACT.G--------T---ACT-A--TC-GG..............TGGG..--AGTGGTCAG..............................---CA..T..-TG   9483
MND-2.MND2_AF367411       --------ACTA-TGCATGCGCACTAGACT.G--------T-CA-ACT-A................................................................................   9518
DEB.CM.CM40               --------A.............GTTG........................CC-AGGAAATATCATGGCAAC-AGGG-A-GGC-T..GGGT--TAC-G--AGTGGCCA..........G--CGCTG-A---   9197
DEB.CM.CM5                T-------A.............GTAG........................TCAAGGAGA.CTCCATGGTTA-AGGG-A-GGA-T--G-GGTAC-GG-A-TGGTTTG...........----GCTGT-A--   9128
MON.CM.L1                 --------A...........GTCACCATGACTAC-GGGCC-...............GGTTGCTGCGGCAAC-GGGC--TCCAAA-----T--T-GGTG-A-C.CGGGGAGTGGCTTG-----TCGT-T..   9409
MON.NG1                   --------A........GTTGCCCTGACTACG---C-CGGTTGCC-----AACCGGGGCGGACTCGTGGGAGGGACT--GAGGTACG.......................-GGAGTGG-T-ACCCTCAGA   8044
TAL.CM.8023               ---------...............................................CGCTGCTAAGGCAACGGGGGAC-GACT---G-GG--A-CG-GA-GA-CTGGGAGTGGCTAC--TC......
TAL.CM.266                --------A...............................................CGTTGCTAAGGCAACGGGGGAC-GACT---G-GG--A-CG-GA-GA-TTGGGAGTGGCTTG-------A-C-..   9301
MUS.CM.1085               --------A........GTCACCATGACTACG---C-CGGTTGCT-----AA-CGGGGCGGACTCGTGGGTGGGACT--GCGGTACT.......................-GGAGTGG-T-ACCCTCAGA   9378
SAB.SN.-.SAB1C            --------AG..................................-.........................-TGG..A-ACT.G--CGG-.ACT-GGAGT-G.CT......TGC....---.GT-CT.G--   9717
GRV.ET.x.GRI_677          ---------C-A....................-----------CA-T--GTGGATCGGAGGCGGTACAGG-G-GGTACTGGGA-T-G--TT..........................---CTCAG-G-TG   9375
TAN.UG.x.TAN1             --------AG......................................-GTGGTGCGAGGCGGTACT.T-GGAGT--CTT.....................................---CTCAG-G-TG   9538
VER.DE.-.AGM3             --------AG-AA...................--------------C--G.TCATG..............-G-G..-TACGGG-A-TGG.CTTTA......................-------.-.G--   9064
VER.KE.-.9063             --------AA......................-----------...........................-GGC..---CCAT.--G-G.-TAC.G-G-AG.TG......-TT.TTA-------.-.G--   9564
VER.KE.-.AGM155           --------AG-AC...................-----------A--C--................................GACATGGG.C--TAC-G--A.-T......-GCTTTA-------.-.G--   9545
VER.KE.-.TYO1             --------AG-GAA..................-----------A-.........................--GG..-ACAT...--G-G.-TCC.G-G-AG.TG......-CT.TTA-------.-.G--   9021
SUN.GA.98.L14             ---------GACAGGG...........................A--G--A-GCT............TA.................................................---A-..TT.A--   9694
LST.CD.88.447             --------AGGACGGGCGG............G--AGG---G............................................................................---A-..TT.G--   8569
LST.CD.88.485             --------AGGACGGGCGG............G--AGG--CA............................................................................---A-..TT.G--   8572
LST.CD.88.524             --------AGGACGGGCGG............G--AGG---G............................................................................---A-..TT.G--   8575
LST.KE.-.lho7             --------AGGACGGGCGG............G--AGG---G............................................................................---A-..TT.G--   9650
SYK.KE.-.SYK173           ........................................................................................GGA-CCTG-GC-GG-GAAGGCTCA.....----.........   9340
SYK.KE.KE51_AY523867      ..................................................................................................................................   9068
COL.CM.-.CGU1             ---------TTC....................-----------A-TT--GAGG..................................GACCT--GC-GA-G.--AGGGCGTG....----GCTA-A.--   8691
```

```
                                                               3'LTR U3 end \/ 3'LTR R repeat start
                                                                      TAR element   stem        bulge     loop      stem
                              TATA Box                                |||| |^||||  |||||||^   |||||^^^||||        | |||||||||    |||
H1B.FR.83.HXB2              ....GCATATAAGCAG..CTGC....TTTTTG.CCTGT....AC..TGGG.TCTCTC..TGGTTAGA..CCAGATCTGAGC.CTGGGAG.CTCTCTGGC....TAA.......................  9586
H1D.CD.84.84ZR085           .................................................................................................................................  8975
H1F1.BE.93.VI850            ...-------------.-C--....----C-.-----....--.----.------..-A------.------T----.-----G-.---------....--G.......................  8884
H1G.SE.93.SE6165            ...-------------.-C--....--C-C-.-----....--.----.------..-T------.------T----.-------.---------....--G.......................  9030
H1H.CF.90.056               ...-------------.----....----C-.-T---....--.----.-----T..--------.-----------.-------.---------....--G.......................  8904
H1J.SE.93.SE7887            ...-------------.----....----C-.-T---....--.----.------..-T------.-----------.-------.---------....---.......................  8885
H1K.CM.96.MP535             .................................................................................................................................  8598
H101_AE.TH.90.CM240         ...-----A-------.-C--....----C-.-T---....--.----.------..-T------.----G--.----.-C-----.---------....--G.......................  9152
H102_AG.NG.-.IBNG           ...-------------.----....--C-C-.-----....--.----.------..-T-C----.-----------.-------.---------....--G.......................  9113
H103_AB.RU.97.KAL153_2      ...-------------.----....----C-.-----....--.----.------..--------.-------.C---.-------.---------....--G.......................  8762
H104_cpx.CY.94.CY032        ...-----A-------.-C--....--C-C-.-----....--.----.------..--------.-----------.-------.---------....--G.......................  9001
H1N.CM.95.YBF30             ...-----A-------.-C--....--C-C-.-T---....--.----.------..-T-C-G--.------TA---T.-------.-ATAT---G.............................  9182
H1O.BE.87.ANT70             ...-------------.-C--....---C--.-T---....--.C---.----GG..-TAGAG--.----G-------.-C-----.---C-----....CT........................  9706
H1O.CM.91.MVP5180           ...-------------.----....---CC-.-T---....--.C---.---TAG..-TAGAG--.----G-------.-C-----.---C-----....CT........................  9702
CPZ.CM.04.EK505             ...-------------.----....-GC-C.G-T---....--.TG--TC-----T..GTTAG.A..-------.-AGAGC-------.---------TAGCA-GGAAC.CCACT..GCTT........  9100
CPZ.CM.04.LB7               ...-------------.-C--....-C-GC.G-T---....--.TG--TC-----T..GTTAG.A..-------.-CT-AG-CT----.---------TAGC---GGAACCCACT..GCTTA.......  9138
CPZ.CM.04.MB66              ...-------------.-C--....---GC-.-T---....--.TG--TC-----T..GTTAG.A..-------.-CT-AG-CT----.---------TAGC---GGAACCCACT..GCTT........  9248
CPZ.MT145                   ...-------------.-GCG....C-C-C-.-T---.....-AAG--TC-----T..G-TAG.A..-------------.-------.---------TAACA-GGGAACCCACT..GCTT........  9176
CPZ.CD.-.ANT                ...-------------.----....---GC-.-T---....TA..A---C-----G..CCTAGGAG..-------A----.-------.--------TA.AT-TTT.......................  9048
CPZ.CM.-.CAM3               ...-------------.-C--....-C-C--.-T---....--.----.------..-T-C----.-----------.-------.--------T...C-G.........................  9041
CPZ.CM.98.CAM5              .................................................................................................................................  9261
CPZ.GAB_X52154              ...-------------.----....---C--.-----....--.----.-----T..-CAC-G--.------T----.-------.------AC-.....-G.......................  9678
CPZ.GAB2                    ...-------------.----....-C-C--.-T---....T-.----.-----T...A-GC---.-----------.-------.---------T...G-C.......................  9016
CPZ.TZ.-.TAN1               ...-------------.----....---GC-CT----....-A..A--..----TG.CCTAATCTG.-----------.-------.--------.....--GTG.....................  9274
CPZ.US.85.CPZUS             ...-----A-------.-C--....-C-C--.-T---....--.----.------...-C-C----.------AA----.-C-----.---------....--G.......................  9651
GSN.99CM166_AF468659        ...-----A-----A..AC--T..C..........---...AG-G.AC...---TCGGAGTC..........................................................  9414
GSN.99CM71_AF468658         ...-----A-------.AC--T..C..........---...AG-G.AC...---TCGGAGTC..........................................................  9410
H2A.ALI_AF082339            ....CTG-----ATGTAC-C--T..GC--.GCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGT 10129
H2A.BEN_M30502              ....CTG-----ATGTAC-C--...T-C--GCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.C----C---A---GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGT 10134
H2A.ST_M31113               ....CTG-----ATGTAC-C--...TAC-CGCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGT  9573
H2B.D205_X61240             ....CTG-----ATGTAC-C--...T-C-CGCAT.---...ATTC...........................................................................  9978
H2B.EHO_U27200              ....CTG-----ATGTAC-C--...T--G-GCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.C----C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGT 10024
H2G.ABT96_AF208027          ....CTG-----AT-TAG----...A---CGCT-.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-C-----GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGT  9482
H2U.FR.96.12034             T...CTG-----ATGTAC-C--...TC---GCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGT  9583

MAC.239_M33262              ....ATG-----AT-TCA----...A---CGCT-.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGT 10054

SMM.SL.92.SL92B             ....TAG-----AT-TAC-C--...T--GCC-CT----...ATTC.A-TC.G----AC.G-AGAG-CTG------.A----C-C-A---GT------C...AGC-CTT...AGTAGGTGAGCCTGGGTGT  9513
SMM.US.x.PGM53              ..CT-T---A-TA--A..----A.TT-CGC-C..---...ATTC.A-T.CG----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA.G..CAGGTAGAGCCTGGGTGT  9991
MNE.US.-.MNE027             ....ATG-----AT-TCA----...A---.C-CT----...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGT  9531
RCM.NG.x.NG411              ....-------------.----A.TT-CGC-C..---...A--C.A-T..------CC---GAG-CTA------.-----.-----TTGT-------T..AAGTCTC..TAGGAGACTAGCTTG.AGCC  9296
RCM.GAB1_AF382829           TGCT-------------.----T..C.-GCGC.T.---...A-AACG--..------CC---GAG-CTA--G---.-----.------...TGT---CT.GGTAAGTCTCTAG...GAACTCCAGCTTG.A  9202
DRL.FAO_AY159321            CA..-----A-------.CT---T..T.AC-GC.A.---......AAAC--.------CC---GAG-CTA------TGAGC-TGG-...TGT-------TGAGT..CTGCTGAAGGAGGCTACCTGCTCGA  9424
MND-1.F17                   ......-----------.----A.....---CG-T---......TCG--AG-----ACTACAGAG-CTAAGG-T-G-ATCTCTGA-C--ATC-C--TAGAGCAAGGACCAGAGTCCTGAGTGACTGGGTC  9123
MND-2.5440_AY159322         GC.T-------------.CT---T..T.GCCGC.-.---......AAGT-AG------CC--AGAG-CTG------.-----.------..TTGT---CT.GGTGAGTTCTTGAAGGAG.TGCCTGCTTGTA  9343
MND-2.14CG_AF328295         AC.T-------------.C-C--T..T.GCCGC.T.---...ATTG.A-TC.---.C-CT...........................................................  9529
MND-2.MND2_AF367411         .................................................................................................................................  9518
DEB.CM.CM40                 ....A----A----C-..A---....-C-G-AG-TATCCA.................................................................................  9227
DEB.CM.CM5                  ....-----------T-..A---....-C-G-AG-TATCCA................................................................................  9158
MON.CM.L1                   ...CA----A-------.AC--C..C..........---...AG-T.ATCCT--GGAGTC...........................................................  9448
MON.NG1                     TCCT-----A-------.AC--T..C..........---...AGMT.ATCCGC-GGAGTCA-C-CTCGTCT-CACAGC-GC-TGG-A-CCCT-GGAGCTAGGTAA-GCGGCCTGGGAGCCCTTGCCTAGAG  8157
TAL.CM.8023
TAL.CM.266                  ...A-------------.A---T..C..........---...AG-T.ATCCT-TG---
MUS.CM.1085                 TTCC-----A-------.AC--T..C..........---...AG-T.ATCCC--GGAGTC...........................................................  9420
SAB.SN.-.SAB1C              ....CA-----------.----.....----.GCG-----...A-A..ACA-.------GC.-T-GAG-CTG------.-----.-----T-.T-------TAAGTCTGAACCAGCTTGAGCCTGGGTGTTCG  9826
GRV.ET.x.GRI_677            .....----A-------.A---....-CGC--G-T---..AACTC.A-T..----TAC-A-GAGACCAG-T..A..----.-----T-.T--G----T..TAGCCTA..ACCCGGT.TGGCCA.CCGGG  9476
TAN.UG.x.TAN1               .....----A-------.A---....-CGC--G-T---..AACTC.A-T..----TAC-A-GAGACCAG-T..-..----.-------.T--G----T..GAGCCTA..ACCTGGTCTGGCCATCCAGG  9641
VER.DE.-.AGM3               ...-----A-------.A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---.----C-.-----.-----T-.T--G----T....--GCCTA..ACCTGGT.TGGCCA.CCA  9164
VER.KE.-.9063               ...-----A-------.A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---.----C-.-----.-----T-.T--G----T....--GCCTA..ACCTGGT.TGGCCA.CCA  9664
VER.KE.-.AGM155             ...-----A-------.A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---.----C-.-----.-----T-.T--G----T....--GCCTA..ACCTGGT.TGGCCA.CCA  9645
VER.KE.-.TYO1               ...-----A-------.A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---.----C-.-----.-----T-.T--G----T....--GCCTA..ACCTGGT.TGGCCA.CCA  9121
SUN.GA.98.L14               ...-------------.----TCTTCGCC.....---...TCGGGA-...-----GCTCA-AGGATAGG-GACGACCCTGAGA--TCTCAC--AGA--TGG.CTTGTCGCTCCTGAGTGAGCACTTGCT  9806
LST.CD.88.447               ...-------------.----T..C.-..............................................................................................  8588
LST.CD.88.485               ...-----A.-----.----T..C.-...............................................................................................  8589
LST.CD.88.524               ...-------------.----T..C.-..............................................................................................  8594
LST.KE.-.lho7               ...-------------.----TCTCCGC-......---TCGGGAG......--CT-GCTCT-GGGACCAG---TCGCTCCTGGGT-AGCA-A--GGTCAGCTAAGCAGTCAGCCCTGGGAGGCTACA...  9760
SYK.KE.-.SYK173             .ACT-----A----C-..A---T..C..........---...AG-TA-CC-CGGAG-TGCAAC-AGCTCT-TCC-G-G-CCTGGGA...-C-CTGAGA.GACTGGT-GAGCGGCCCG..GGAGCCCTTGCC  9447
SYK.KE.KE51_AY523867        .................................................................................................................................  9068
COL.CM.-.CGU1               GCTT-T---A---TCAG.AGC-T..CGAGAGCTT.C-AGG.C...............................................................................  8728
```

```
                                                    TAR element
                             ||||                  |||||||     |||
H1B.FR.83.HXB2            ...CTAG..................GGAACCC.....ACTGCTTAA....................................................................................  9606
H1D.CD.84.84ZR085         ..................................................................................................................................  8975
H1F1.BE.93.VI850          ...--G-..................-------.....--------.....................................................................................  8903
H1G.SE.93.SE6165          ...-AG-..................-------.....---------A...................................................................................  9051
H1H.CF.90.056             ...---A..................-------.....---------....................................................................................  8924
H1J.SE.93.SE7887          ...----..................-------.....---------....................................................................................  8905
H1K.CM.96.MP535           ..................................................................................................................................  8598
H101_AE.TH.90.CM240       ...-A--..................-------.....---------A...................................................................................  9173
H102_AG.NG.-.IBNG         ...-GGA..................-------.....---------....................................................................................  9133
H103_AB.RU.97.KAL153_2    ....................................................................................................CTAAGGAACCCACTGCTTAA..........  8782
H104_cpx.CY.94.CY032      ...----..................-------.....---------....................................................................................  9021
H1N.CM.95.YBF30           ..................................................................................................................................  9182
H1O.BE.87.ANT70           ..CTAGC..................T------.....G--CG----....................................................................................  9727
H1O.CM.91.MVP5180         ..CTAGC..................T------.....G--------....................................................................................  9723
CPZ.CM.04.EK505           ............................................--....................................................................................  9102
CPZ.CM.04.LB7             ............................................--....................................................................................  9140
CPZ.CM.04.MB66            ............................................--....................................................................................  9250
CPZ.MT145                 ............................................--....................................................................................  9178
CPZ.CD.-.ANT              ...GGCAG.................-AG----.......-----......................................................................................  9065
CPZ.CM.-.CAM3             ...G---..................A------.....---------....................................................................................  9061
CPZ.CM.98.CAM5            ..................................................................................................................................  9261
CPZ.GAB_X52154            ...TG-A..................-------.....---------....................................................................................  9698
CPZ.GAB2                  ....---...................------.....---------....................................................................................  9034
CPZ.TZ.-.TAN1             .GCTGGC................TA-AG----.....G--------....................................................................................  9297
CPZ.US.85.CPZUS           ...TGTA...................------.....---------....................................................................................  9670
GSN.99CM166_AF468659      ..................................................................................................................................  9414
GSN.99CM71_AF468658       ..................................................................................................................................  9410
H2A.ALI_AF082339          TCC--.................GCTA--CT-TC.ACC-G-----GGCCGGCACTGGGCA..GA....................................CGGCTCCACGCT.TGCTTGC...TTAA.AGA 10198
H2A.BEN_M30502            TCC--.................GCT---CT-TC.ACC-G-A---GGCCGGTACTGGGCA..GA....................................CGGCTCCACGCT.TGCTTGC...TTAA.AGA 10203
H2A.ST_M31113             TCC--.................GCTA--CT-TC.ACC-G-----GGCCGGCACTGGGCA..GA....................................CGGCTCCACGCT.TGCTTGC...TTAAAAGA  9643
H2B.D205_X61240           ..................................................................................................................................  9978
H2B.EHO_U27200            TCC--.................GCTA--CT-TC.ACC-GCA---GGCCGGTGCTGGGCA..GAC....................................GGCTCCACGCT.TGCTTGC...TTAA.AA. 10092
H2G.ABT96_AF208027        TCC--.................GCTA--CT-TC.ACC-GCA---GGCCGGTGCTGGGCA..GAGC...................................GGCTCCACGCT.TGCTTGC...TTACAACG  9553
H2U.FR.96.12034           TCC--.................GCTA--CT-TC.ACC-AAAG........................................................................................  9607

MAC.239_M33262            TCC--.................GCTA--CT-TC.ACC-GCA---GGCCGGTGCTGGGCA..GAG...................................TGACTCCACGCT.TGCTTGC...TTAA.AGC 10124

SMM.SL.92.SL92B           TCT--.................ACTAA-CT-TCAACC-GCAACAGGCCATTGCTGGGTA..GAC....................................GGCTCCACGCT.TGCTTGC...TTGA....  9580
SMM.US.x.PGM53            TCC--................GCTA-ACT-T-ACCAGCACTTGGCCGG.............................................TGCTGGGCAGAGTGGCTCCACGCTTGCTTGCTTAAAG 10060
MNE.US.-.MNE027           TCC--.................GCTA--CT-TC.ACC-GCA---GGCCGGTGCTGTGCA..GAG...................................TGGCTCCACGCT.TGCTTGC...TTAA.AGA  9601
RCM.NG.x.NG411            TGGG-............GTTCGCTA-TTCT-TGAAGGG--TGC-GGGAGCGCCTTGGCTGGGTCTTGGTAGACCTTCAGCTGGGTACTTGGCCAGTCCCAGTGGGAGTAGGCTCCGCTTGCTTTCTGGTT  9414
RCM.GAB1_AF382829         GCC.-G-GT.GTTCGCTGGTGTCTCT----A.GGCTTG---GGGTGCCTCTCGCTCTTCGGGTAGACCGCCAGTTGAGGCTCGGCCGGCCTCAACGGGAGAGATCACCGCTTGCTTATAGCCTTGAAGCT  9329
DRL.FAO_AY159321          GCC-CG-GTGGTTCGCAGGGCCTTG-CCTTG.GGCT..---GG.T-GCTCGGTCAGGAGGCTTCTGGAGG.TAGATCACTAGGGGCTGCTCTGGTCTTGGTCTGGTAGACCTTCGCTACTTCTAGTGGCC  9549
MND-1.F17                 TGAGC-CCTCACTCGGGGCTGATCACCTCGAGGTAGTGGAA--CCTTGCTTGCTTGCTATTGTCTTC...............................................................  9190
MND-2.5440_AY159322       GCC--G-GCGGTTCGCAGGCCCCTG-C.TTGTGGCT..---GG.T-GCTCGTCAGGTGTTCTGGAAAGGCTTACGGGAGGACACTCTTGCTTGGTCTTGATAGACCTCTAGCAATCCCTTAGGCCTGGAG  9469
MND-2.14CG_AF328295       ..................................................................................................................................  9529
MND-2.MND2_AF367411       ..................................................................................................................................  9518
DEB.CM.CM40               ..................................................................................................................................  9227
DEB.CM.CM5                ..................................................................................................................................  9158
MON.CM.L1                 ..................................................................................................................................  9448
MON.NG1                   AGGTGCCGGCTCTCGGCTCGCTTGCT---.............................................................................................TAGGCTTT  8194
TAL.CM.8023
TAL.CM.266
MUS.CM.1085               ..................................................................................................................................  9420
SAB.SN.-.SAB1C            CTGGGGAA.....GGCAAGCCTAAGA-C--T-TGGTG-GGT---GGTAGACCCTCGCC..............TGGCGACTGGCCATTGCCAGTAGCAG.AGACTCC.......GCTTGCTTGCTTG..AT  9927
GRV.ET.x.GRI_677          GGTAAG..................G.........................................................................................................  9483
TAN.UG.x.TAN1             GGTAAGA.................G.........................................................................................................  9649
VER.DE.-.AGM3             GGGG--A..................---CT--.....TTG-----GAA..................................................................................  9189
VER.KE.-.9063             GGGG--A..................---CT--.....TTG----GGAA..................................................................................  9689
VER.KE.-.AGM155           GGGG--A..................---CT--.....TTG-----GAA..................................................................................  9670
VER.KE.-.TYO1             GGGG--A..................---CT--.....TTG-----GAA..................................................................................  9146
SUN.GA.98.L14             AAG-ACCTCACCCGAGCCTGATCAGCTTGGGGAGCTAGA.......................................GGCTCC...TTGCTTTGCAGTACAAGCCTTC.....................  9873
LST.CD.88.447             ..................................................................................................................................  8588
LST.CD.88.485             ..................................................................................................................................  8589
LST.CD.88.524             ..................................................................................................................................  8594
LST.KE.-.lho7             ....CTTAGCACCCACCTGGGCCTGATC-G--TAGGG-GCTAAG.................................GGCTCCTAGCTTGCTACTTTAAA...GCCTTC.....................  9829
SYK.KE.-.SYK173           AGTAA--GCTAGTGAGGTCCGGGTTT-GC-.GCCCGGT--T-GGC-GCTCTCCGTT..GCTT...........................................GCTTTGCTTACG..CTGA..TCAGC  9527
SYK.KE.KE51_AY523867      ..................................................................................................................................  9068
COL.CM.-.CGU1             ..................................................................................................................................  8728
```

```
                                          TAR elements of HIV-2 and SIV have extra stem-loop, see Berkhout, Nucleic Acids Res 20(1):27-31(1992)
                                                                  Poly-A signal     3'LTR R repeat end \/ 3'LTR U5 start
H1B.FR.83.HXB2            GCC.TC..................................AATAAAG...CTTGCC.......TTGAG..TG.CTTCAA.GTAGTG...TGTGCCCGTC.TG.TT..GTG.TGACTCT...GGT..AA.C  9673
H1D.CD.84.84ZR085         ..................................................................................................................................  8975
H1F1.BE.93.VI850          ..................................................................................................................................  8903
H1G.SE.93.SE6165          ---.--..................................-------...------.......AA-CC..............................................................  9074
H1H.CF.90.056             ---.--..................................-------...------.......-----..--.---T.....................................................  8953
H1J.SE.93.SE7887          ---.--..................................-------...------.......-----..G-CGCATGC.AAGCC-............................................  8943
H1K.CM.96.MP535           ..................................................................................................................................  8598
H101_AE.TH.90.CM240       ---.--..................................-------...------.......-----..--.---A-....................................................  9203
H102_AG.NG.-.IBNG         ---.--..................................-------...------.......-----..--.------.------...----------.--A--..---.-------...---..--.-  9201
H103_AB.RU.97.KAL153_2    ---.--..................................-------...------.......-----..--.-........................................................  8808
H104_cpx.CY.94.CY032      ---.--..................................-------...------.......-----..--.----.....................................................  9050
H1N.CM.95.YBF30           ..................................................................................................................................  9182
H1O.BE.87.ANT70           CG-.--..................................-------...------.......-----..--.AG.......................................................  9754
H1O.CM.91.MVP5180         CG-.--..................................-------...------.......-----..--.AGAAGC.AGT---....--CT-A.--.--.--CAACC.CTGG-G-...CTA..G-..  9789
CPZ.CM.04.EK505           ---.--..................................-------...------.......-----..--CT-AA.....................................................  9132
CPZ.CM.04.LB7             ---.--..................................-------...------.......-----..--CTGCA.....................................................  9170
CPZ.CM.04.MB66            ---.--..................................-------...------.......-----..--A---A-....................................................  9281
CPZ.MT145                 ---.--..................................-------...------.......-----..--CTGCA.....................................................  9208
CPZ.CD.-.ANT              .-T.-.............................................................................................................................  9068
CPZ.CM.-.CAM3             ---.--..................................-------...------.......-----..--A---G.-.-C----...------TA--.CA.--C..A-.-A-----...---..--.-  9128
CPZ.CM.98.CAM5            ..................................................................................................................................  9261
CPZ.GAB_X52154            ---.--..................................-------...------.......-----.TGTAT--G.-.-C----...------TA--TA..--C..A-.AC-----...---..--.-  9766
CPZ.GAB2                  ---.--..................................-------...------.......-----..--CTC-A.-.------...-----T-A--TG..--C..A-.-A-----...---..GT.-  9101
CPZ.TZ.-.TAN1             CG-.--..................................--------..-C----........----..A-.TG-T.-...................................................  9326
CPZ.US.85.CPZUS           ---.--..................................-------...------.......-----..--CT--A.-.------...------TA--.C..--C..A-.-------...---..--.-  9736
GSN.99CM166_AF468659      ..................................................................................................................................  9414
GSN.99CM71_AF468658       ..................................................................................................................................  9410
H2A.ALI_AF082339          C-TC-T..................................-------...--.---AA.....--AGA..A-CAGGTT-.AAG---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CA-..TCGG 10275
H2A.BEN_M30502            C-TC-TC.................................-------...--.---AG.....--AGA..A-CAAGTT-.AGT---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CA-..TCGG 10281
H2A.ST_M31113             C-TC-T..................................-------...--.---AG.....--AGA..A-CA........................................................  9672
H2B.D205_X61240           ..................................................................................................................................  9978
H2B.EHO_U27200            C--TCTT.................................------A...--.---TA.....--AGA..A-CAAGAC-.AGT---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...CA-..TCGG 10170
H2G.ABT96_AF208027        A--TCTT.................................-------...--.---AAA....--AGA..A-TAAG.C-.AGT---TGTGT.......................................  9599
H2U.FR.96_12034           ..................................................................................................................................  9607

MAC.239_M33262            C-TC-TC.................................-------...--.---AT..T..--AGA..A-TAAG-T-..GT---TGT-CCCAT-TCT.CCTAGCC-CC.GCCTGG-...CAAC.TCGG 10203

SMM.SL.92.SL92B           .-TC-.C.........................................................AAT-AAGG..--.GC-AC.....TTAGA..AG-AAGTC.............................  9613
SMM.US.x.PGM53            A--TCTTC................................-------....C----GC.....--AGAAG...TAAGC-A--GTGTGTTCCCAT-TC--C-AG-CGCCGCC--GTCA-CTC---ACTCGA 10146
MNE.US.-.MNE027           C-TC-TC.................................-------...--.---AT.....--AGA..A-TAAG-C-..GT---TGC-CCCAT-TCT.CCTAGTC-CC....................  9664
RCM.NG.x.NG411            AATTG-TCAATAAACTGCATGCTCTTAGTTGAACCAGAATC-AGTGT...G----CAT.....--T-......TC--TCA--GAAACGACTCTGGG--TGG-A-CCCTCAGAT-T-TGTGGCAGAAGTA-  9530
RCM.GAB1_AF382829         C.......................................-------C......ATGCCAGTTAGTTT.ACTGTAAGC-.AGT---TGCCTGTTTTAC-TCTCAGCA--TAACGACTCTGG---..-GGG  9410
DRL.FAO_AY159321          ATAGAAGTAGGGGAGTAGGCACCGCTTGCTTATCCTAATTGC-C--TAAAGC-A--CATTTAGA--GATGA-TAAGTGT.ACTC-TGTC-T-TGTTC-TAGTGAGA.....CTCTGG-TA.CTG..G-GA  9670
MND-1.F17                 ........................................------.....G-AA-........-T--AA-TAGAG--
MND-2.5440_AY159322       ATTG-GGGATTGACTACCGCTTGCTTGCTATTATTGCCC.-------C..T-AC-TGA..TTAGAATAGCAAGTAAG-G.TACTC-TCCAT-TTGTC-TCAATAGAA...ACTCTGG-TA.CCG..G-GA  9587
MND-2.14CG_AF328295       ..................................................................................................................................  9529
MND-2.MND2_AF367411       ..................................................................................................................................  9518
DEB.CM.CM40               ..................................................................................................................................  9227
DEB.CM.CM5                ..................................................................................................................................  9158
MON.CM.L1                 ..................................................................................................................................  9448
MON.NG1                   A--AG-CC................................------.....G-GCTTGAAAGCAAAA-AAAAAAAAAA--AAA-AAAAAAAAAAAAAAAAAAA...........................  8260
TAL.CM.8023
TAL.CM.266
MUS.CM.1085               ..................................................................................................................................  9420
SAB.SN.-.SAB1C            TG-C--..................................-------T..------.......--.--TT-A.-AAGT-AAGCA..AGTGTGTG--CATT-AT-CCTCA-GAC-AC-CTGGT........ 10002
GRV.ET.x.GRI_677          .................ACTCCTTGGCTTCATATAGCTC.------C...C---TCGC.....--AG......TCG-T-TA-T-GAGTCAAGTG-TCATTGCTGCGCCGAGCCT--AGAGGT-AACCTCT  9581
TAN.UG.x.TAN1             .................CCTCCTTGCTTAGAAGGCTATTGCT-AT--A....G-TG-CTGCTTAAGAG.....TC-ATCT-AGCAAGTGCC-CATT-CGCGCCC-TCC-CAG-GAA-CCTC-T-ACTGGG  9753
VER.DE.-.AGM3             AG-.-...................................------.T..---CG-TG..C.A--AG-..CAGAGA-TT.--GAGAGACAT----T-GT.GCA-CCA-CA.CA-G-GA...--G..--GA  9267
VER.KE.-.9063             AG-.-...................................------AC..T-----TG..CA.--AGA..GCTTA--TG.AGTCAAGTGCCCT-ATTGA.C-CC-CAC-C.AAG-AGG...--...--C-  9767
VER.KE.-.AGM155           AG-.-...................................------....------TG..CA.--AGA..GCTTA--TG.AGTCAAGTGCCCT-ATTAA.--CC-CAC-C.-TGAA-G...--...-GAA  9746
VER.KE.-.TYO1             AG-.-...................................------....------TG..CA.--AG...............................................................  9170
SUN.GA.98.L14             ........................................-------CCTGCCTTAGTTTTAC-........................GCAAGCAAGTGTTCCCG-..............GTTCTCTCT.  9924
LST.CD.88.447             ..................................................................................................................................  8588
LST.CD.88.485             ..................................................................................................................................  8589
LST.CD.88.524             ..................................................................................................................................  8594
LST.KE.-.lho7             ........................................------......AA-TTGCTCATTG-T--AAA-TACAAGCTAGT--CTCA-AGTAGTC--T-TC-CCTTC-CCC................  9897
SYK.KE.-.SYK173           C-.A....................................-------T..G-G-TAAA..CGCAA-TA..ACGTC-GCT.C-CA-CTGT-C---TG-GT.CACC-CCCGA.A-GTG..............  9597
SYK.KE.KE51_AY523867      ..................................................................................................................................  9068
COL.CM.-.CGU1             ..................................................................................................................................  8728
```

```
                                                                                                     3'LTR end \
H1B.FR.83.HXB2            TAGAGA...............TCCCT..CAGA....................CCCTTTTAGTCAGTGTGGAAAATCTCTAGCA      9719
H1D.CD.84.84ZR085         ...................................................................................      8975
H1F1.BE.93.VI850          ...................................................................................      8903
H1G.SE.93.SE6165          ...................................................................................      9074
H1H.CF.90.056             ...................................................................................      8953
H1J.SE.93.SE7887          ...................................................................................      8943
H1K.CM.96.MP535           ...................................................................................      8598
H101_AE.TH.90.CM240       ...................................................................................      9203
H102_AG.NG.-.IBNG         ...................................................................................      9201
H103_AB.RU.97.KAL153_2    ...................................................................................      8808
H104_cpx.CY.94.CY032      ...................................................................................      9050
H1N.CM.95.YBF30           ...................................................................................      9182
H1O.BE.87.ANT70           ...................................................................................      9754
H1O.CM.91.MVP5180         ....--TC...........................................................................      9793
CPZ.CM.04.EK505           ...................................................................................      9132
CPZ.CM.04.LB7             ...................................................................................      9170
CPZ.CM.04.MB66            ...................................................................................      9281
CPZ.MT145                 ...................................................................................      9208
CPZ.CD.-.ANT              ...................................................................................      9068
CPZ.CM.-.CAM3             ------TCC...................................CTCAGAA.-TA--A--A..GTA-GT-----------...      9170
CPZ.CM.98.CAM5            ...................................................................................      9261
CPZ.GAB_X52154            ------TCC...................................CTCAGAT.TAAA--ATA..GTCAA---------------      9811
CPZ.GAB2                  ------TCC...................................CTCAGA..-TAAACATAA.G-A-----------------      9146
CPZ.TZ.-.TAN1             ...................................................................................      9326
CPZ.US.85.CPZUS           ------TCC...................................CTCAGAA.-T---AA-A..GTA-GT--------------      9781
GSN.99CM166_AF468659      ...................................................................................      9414
GSN.99CM71_AF468658       ...................................................................................      9410
H2A.ALI_AF082339          -GTTC-CCTGA....GTAACAAGA-CCTGGTCTGTTAGGACCCTTTCTGCTT.TGGGAA-CGG--GCA--------C------     10353
H2A.BEN_M30502            -GTTC-TCTGA....GTAACAAGA-CCTGGTCTGTTAGGACCCTTCTCGCTT.TGGGAATCCA--GCA--------C------     10359
H2A.ST_M31113             ...................................................................................      9672
H2B.D205_X61240           ...................................................................................      9978
H2B.EHO_U27200            -..GCTC..CA..C.TA..GAA.A-CCTGGTCTGTTAGGACCCCTTCTGCTT.TGGGAA-CCA--GCA--------C------     10242
H2G.ABT96_AF208027        ...................................................................................      9599
H2U.FR.96.12034           ...................................................................................      9607

MAC.239_M33262            -..-CTCAATA...ATA.AGAAGA-CCTGGTCTGTTAGGACCCTTTCTGCTT.TGGGAA-CCG-AGCA--------C------     10279

SMM.SL.92.SL92B           ...................................................................................      9613
SMM.US.x.PGM53            ATC-T-AGAAGACCCTGGTCTGTTAGGA-CCTTTCTGC
MNE.US.-.MNE027           ...................................................................................      9664
RCM.NG.x.NG411            -GA--GCTAAGAAAAAATCCCAA--AG
RCM.GAB1_AF382829         A....................--....C-TC-GATTCTTGTGGCAGAAGAGC-TTGGGCTAAG-AAA-.TCCCTA-CAGT...      9465
DRL.FAO_AY159321_         .....................-----CAG-CTAGGTGAGAGCTATCCTGAGCTTTAG-G--AGTAAAG.A--------C----      9731
MND-1.F17
MND-2.5440_AY159322       .....................----..T----TCTGTGGCAGAGATTTGATAT-TAG-G--AGTAGAGAA--------C----      9647
MND-2.14CG_AF328295       ...................................................................................      9529
MND-2.MND2_AF367411       ...................................................................................      9518
DEB.CM.CM40               ...................................................................................      9227
DEB.CM.CM5                ...................................................................................      9158
MON.CM.L1                 ...................................................................................      9448
MON.NG1                   ...................................................................................      8260
TAL.CM.8023
TAL.CM.266
MUS.CM.1085               ...................................................................................      9420
SAB.SN.-.SAB1C            .TACT-AG...........................................................................     10009
GRV.ET.x.GRI_677          CTT-CTGGGTTCTCCTGTACCCAGG-GGG---AACTCCAGCA
TAN.UG.x.TAN1             GTTTCTTATCCAGGCTGAGAAAT-TCCAGCA
VER.DE.-.AGM3             CCCT--TGGAA..T................................CAACCAT..GGAG-AGT....................      9294
VER.KE.-.9063             GTTCCTTACTG..G.....................................................................      9779
VER.KE.-.AGM155           GTTCCTTACTG..G.....................................................................      9758
VER.KE.-.TYO1             ...................................................................................      9170
SUN.GA.98.L14             .TTCTCCTTCTAAACCCTGTTCGA--GCTCTCAGCTAGGGAGCAATTACTGG--GCC-AGTGATCC-G--CTG-CGG------     10006
LST.CD.88.447             ...................................................................................      8588
LST.CD.88.485             ...................................................................................      8589
LST.CD.88.524             ...................................................................................      8594
LST.KE.-.lho7             ....TGGTTCAGTGATCTCTCA-TAGAGAGATTGGAGCCTTGTGATCCGGGA-AGGGC-TCCAG...................      9957
SYK.KE.-.SYK173           ...................................................................................      9597
SYK.KE.KE51_AY523867      ...................................................................................      9068
COL.CM.-.CGU1             ...................................................................................      8728
```