III

HIV-1/HIV-2/SIV Complete Genomes

Alignment of HIV-1/HIV-2/SIV Complete Genomes

Complete genomes of HIV-1, HIV-2, and SIV sequences are represented in the alignment as follows:

HIV-1: One or two sequences per subtype from M, N, and O groups
HIV-2: Three A, two B, and one G sequence
SIV: Sequences from viruses infecting chimpanzees, the subspecies of African Green Monkeys—vervet, grivet, sabaeus, tantalus—sooty mangabeys and macaques infected with sooty mangabey virus, L'Hoest monkey, and mandril.

Table 1 lists the sequence name, accession number, first author, and publication for each of the sequences in this alignment. See page 34 of this compendium for a review of the phylogeny of SIV, HIV and other lentiviruses. Together, these genomes represent the full breadth of diversity discovered to date in the primate lentivirus lineage. Non-primate lentiviruses such as equine infectious anemia virus (EIAV) and caprine arthritis/encephalitis virus (CAEV) are more distantly related to each other and to the primate lentiviruses, and are not included in this alignment.

The new alignment was based on the previous version, but entirely redone using manual editing with the program BioEdit,

http://www.mbio.ncsu.edu/RNaseP/info/programs/BIOEDIT/bio-edit.html

created by Tom Hall. It is now entirely 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 annotation is unchanged from last year's compendium.

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. Three representative genomes have been translated in all three reading frames; HIV-1 subtype B strain HXB2, HIV-2 ALI, and SIV from a Sykes monkey. The translations are provided as a visual aid for finding landmarks in the genomes.

The HIV-2/SIV-SMM *vpx* gene is postulated to be a duplication of the *vpr* gene (Tristem *et al*. *Nature* **347**:341–342 (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.

HIV Database compendia from previous years, as well as electronic copies of these alignments in a variety of formats are available on our WWW site at

http://hiv-web.lanl.gov/HTML/compendium.html

and

http://hiv-web.lanl.gov/ALIGN_CURRENT/ALIGN-INDEX.html

respectively.

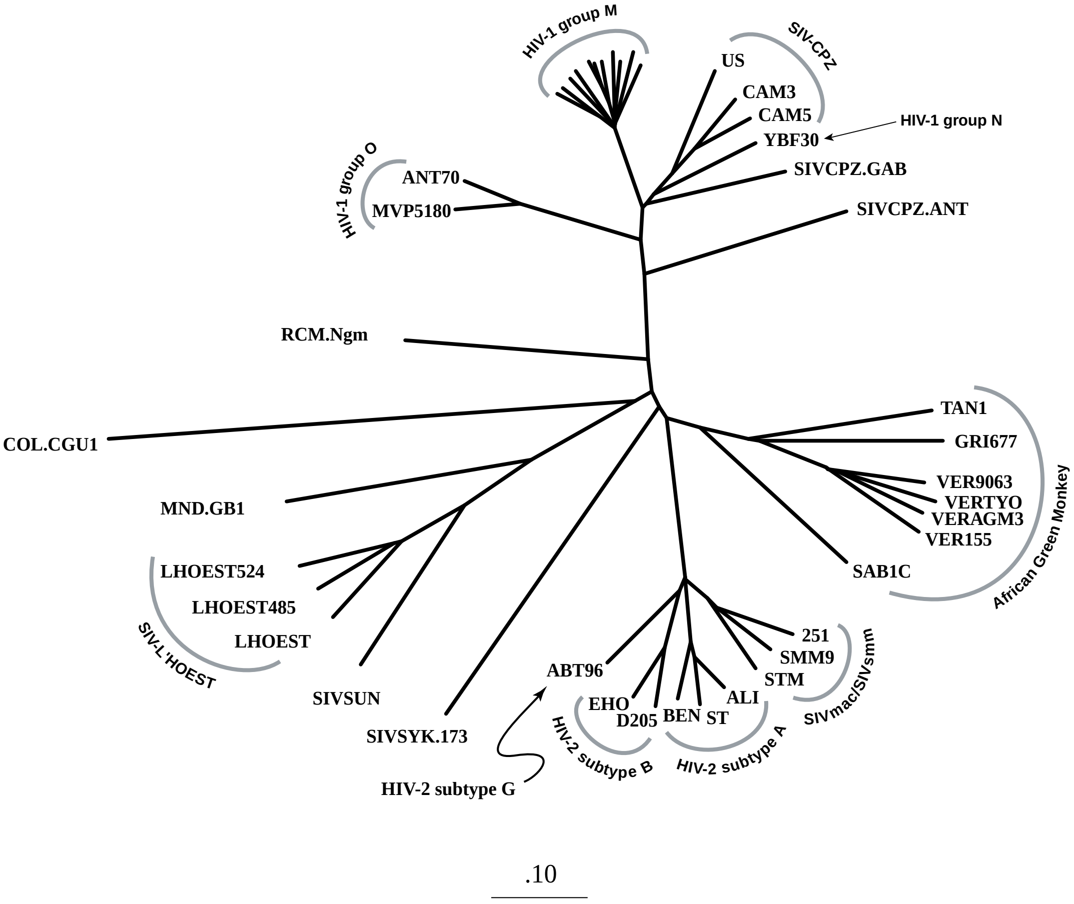


Figure 1. *This phylogenetic tree was constructed from the complete lentivirus genome alignment. Columns in the alignment for which any sequence is represented by a gap were first removed. The tree was constructed using PHYLIP DNADIST with maximum likelihood distance estimation and a transition/transversion ratio of 1.6. The DNADIST output was then used to create a neighbor-joining tree using the WEIGHBOR program (The PHYLIP neighbor program gave similar results). WEIGHBOR is available for UNIX, Mac, and PC at*

http://t10.lanl.gov/billb/weighbor/index.html.

A cut and paste interface to weighbor is available at

http://bioweb.pasteur.fr/seqanal/interfaces/weighbor.html

Some of the genomes are known to represent recombinants or mosaics between two or more different lineages, so this tree should not be assumed to represent the true phylogenetic history of these viruses. For example, the SAB1C genome from the sabaeus subspecies of African green monkeys is recombinant between an African green monkey lentiviral lineage and a sooty mangabey viral lineage. When the AGM-like regions of its genome are used to build a phylogeneitc tree, the SAB1C lineage shares the same major branch with the other AGM lineages. Given these limitations, the tree is still useful for graphically representing the diversity and relationships between these viruses. It is readily apparent that the HIV-2 viruses are related to sooty mangabey viruses.

Table 1: Table of HIV-1/HIV-2/SIV Complete Genome Alignments

Name	Accession	Author	Reference
01_AE.TH.CM240	U54771	Carr, JK	*J Virol* **70**(9):5935–5943 (1996)
02_AG.NG.IBNG	L39106	Howard, TM	*ARHR* **10**(12):1755–1757 (1994)
03_AB.RU.KAL153-2	AF193276	Liitjola, K	*AIDS* **12**(14):1907–1919 (1998)
04_cpx.CY.94CY032-3	AF049337	Gao, F	*J Virol* **72**(12):10234–10241 (1998)
A.UG.U455	M62320	Oram, JD	*ARHR* **6**(9):1073–1078 (1990)
B.FR.HXB2	K03455	Wong-Staal, F	*Nature* **313**(6000):277–284 (1985)
C.ET.ETH2220	U46016	Salminen, MO	*ARHR* **12**(14):1329–1339 (1996)
D.CD.84ZR085	U88822	Gao, F	*J Virol* **72**(7):5680–98 (1998)
F1.BE.VI850	AF077336	Carr, JK	*Virology* **269**(1):95–104 (2000)
G.SE.SE6165	AF061642	Carr, JK	*Virology* **247**(1):22–31 (1998)
H.CF.90CF056	AF005496	Murphy, E	*ARHR* **9**(10):997–1006 (1993)
J.SE.SE7887	AF082394	Laukkanen, T	*ARHR* **15**(3):293–297 (1999)
K.CM.MP535	AJ249239	Peeters, M	Unpublished
N.CM.YBF30	AJ006022	Simon, F	*Nature Med* **4**(9):1032–1037 (1998)
O.CM.ANT70	L20587	Vanden Haesevelde, M	*J Virol* **68**(3):1586–1596 (1994)
O.CM.MVP5180	L20571	Gurtler, LG	*J Virol* **68**:1581–1585 (1994)
H2A.DE.BEN	M30502	Kirchhoff, F	*Virology* **177**(1):305–311 (1990)
H2A.GW.ALI	AF082339	Azevedo-Pereira, JM	Unpublished (1998)
H2A.SN.HIV2ST	M31113	Kumar, P	*J Virol* **64**(2):890–901 (1990)
H2B.CI.EHO	U27200	Rey-Cuille, MA	*Virology* **202**(1):471–476 (1994)
H2B.GH.D205	X61240	Kreutz, R	*ARHR* **8**(9):1619–1629 (1992)
H2G.CI.ABT96	AF208027	Brennan, CA	*ARHR* **13**(5):401–404 (1997)
COL.-.SIV-COL-CGU1	-	Peeters, M	Unpublished (2000)
CPZ.CD.CPZANT	U42720	Vanden Haesevelde, MM	*Virology* **221**(2): 346–350 (1996)
CPZ.CM.CAM3	AF115393	Corbet, S	*J Virol* **74**: 529–534 (2000)
CPZ.CM.CAM5	AJ271369	Souquiere, S	Unpublished
CPZ.GA.CPZGAB	X52154	Huet, T	*Nature* **345**(6273):356–359 (1990)
CPZ.US.CPZUS	AF103818	Gao, F	*Nature* **397**(6718):436–441 (1999)
GRIVET.ET.AGM_GRI-677	M66437	Fomsgaard, A	*Virology* **182**(1):397–402 (1991)
LHOEST.CD.SIVlhoest447	AF188114	Beer, BE	*J Virol* **74**(8):3892–3898 (2000)
LHOEST.CD.SIVlhoest485	AF188115	Beer, BE	*J Virol* **74**(8):3892–3898 (2000)
LHOEST.CD.SIVlhoest524	AF188116	Beer, BE	*J Virol* **74**(8):3892–3898 (2000)
LHOEST.KE.SIVlhoest	AF075269	Hirsch, VM	*J Virol* **73**(2):1036–1045 (1999)
MAC.US.MM251	M19499	Franchini, G	*Nature* **328**(6130):539–543 (1987)
MANDRIL.GA.MNDGB1	M27470	Tsujimoto, H	*Nature* **341**:539–541 (1989)
RCM.NG.SIV-RCM-NGM	-	Beer, B	Unpublished (2000)
SABAEUS.SN.SAB1C	U04005	Jin, MJ	*EMBO J* **13**(12):2935–2947 (1994)
SMM.US.SIVSMMH9	M80194	Courgnaud, V	*J Virol* **66**:414–419 (1992)
STM.US.STM	M83293	Novembre, FJ	*Virology* **186**(2):783–787 (1992)
SUN.GA.SIVSUN	AF131870	Beer, BE	*J Virol* **73**(9):7734–7744 (1999)
SYKES.KE.SYK173	L06042	Hirsch, VM	*J Virol* **67**(3):1517–1528 (1993)
TANTALUS.UG.TAN1	U58991	Stivahtis, GL	*Virology* **228**:394–399 (1997)
VERVET.DE.AGM3	M30931	Baier, M	*Virology* **176**(1):216–221 (1990)
VERVET.KE.AGM155	M29975	Johnson, PR	*J Virol* **64**(3):1086–1092 (1990)
VERVET.KE.AGMTYO	X07805	Fukasawa, M	*Nature* **333**:457–461 (1988)
VERVET.KE.AGM_VER-9063	L40990	Hirsch, VM	*J Virol* **69**(2):955–967 (1995)

```
                       |-> 5’LTR U3 start
B.FR.HXB2               TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTG  129

A.UG.U455                                                                                                                                                  0
C.ET.ETH2220                                                                                                                                               0
D.CD.84ZR085                                                                                                                                               0
F1.BE.VI850                                                                                                                                                0
G.SE.SE6165                                                                                                                                                0
H.CF.90CF056                                                                                                                                               0
J.SE.SE7887                                                                                                                                                0
K.CM.MP535C                                                                                                                                                0
01_AE.TH.CM240                                                                                                                                             0
02_AG.NG.IBNG                                                                                                                                              0
03_AB.RU.KAL153-2       ----T---T-----T-----A-GAA---------A--------------G----T-----------------------G----..............--------T----TC-----A--A----  114
04_cpx.CY.94CY032-3                                                                                                                                        0
N.CM.YBF30                                                                                                                                                 0
O.CM.MVP5180           C----T---T-----T-------TAAG---GC---A-----G-----C-----A--T-----T--G--A-T-----------G----TGT--------G--A-----ACCT----TC---------A----  130
O.BE.ANT70              --------T-----T-------TAA----GC---A-----G-----T---G-G--TA----T--G--A-T-----------G-----------------A-----A-C---G-TC---------A----  129
CPZ.US.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
CPZ.GA.CPZGAB           --------T--G--T-----AGGA----------G--------C--C---G----T-------------T---------C-G-------------A----A-----A-CA----TC-----CTGT----  129
CPZ.CM.CAM3                                                                                                                                                0
CPZ.CM.CAM5                                                                                                                                                0
CPZ.CD.CPZANT                                                                                                                                              0

H2A.GW.ALI              --------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.SN.HIV2ST                                                                                                                                              0
H2A.DE.BEN              --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
H2B.GH.D205             --------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.CI.EHO              --------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.CI.ABT96                                                                                                                                               0
MAC.US.MM251            --------A-TTA-T--AGTGC-A-------TAGA---T-A--CA---ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT----A-----A--T-----C---AA---A----  129
SMM.US.SIVSMMH9                                                                                                                                            0
STM.US.STM                                                                                                                                                 0
TANTALUS.UG.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
GRIVET.ET.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
VERVET.KE.VER-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
VERVET.KE.AGMTYO                                                                                                                                           0
VERVET.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
VERVET.DE.AGM3                                                                                                                                             0
SABAEUS.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
SUN.GA.SIVSUN           ----------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
MANDRIL.GA.MNDGB1                                                                                                                                          0
LHOEST.KE.hoest         -----------TA-TGG-----GA-------G--C---T-AAT---C-ACTGTG-GA-TGAGTGG--TCT-A-TGGA--C-TCAT-----------GAT--A-----A-CA-----C--T-----A----  129
LHOEST.CD.hoest447                                                                                                                                         0
LHOEST.CD.hoest485                                                                                                                                         0
LHOEST.CD.hoest524                                                                                                                                         0
RCM.NG.SIV-NGM          --------A-TTA-TT---AG--A-G---AGGCGC--TT-A--C--C-AC--GG-GA-TGA---T--AATAA-A--A----G------T-----T---------------A----------ACA-TG----  129
COL.__.SIV-CGU1                                                                                                                                            0
SYKES.KE.SYK173                                                                                                        ---A-C-C-G--C--GT-A-T-----   26
```

```
B.FR.HXB2          GATGGTGCTACAAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAA...GCCAACAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGC......CTGCATGGAATGGATGACCCGGAGAGAGAAGT    250

A.UG.U455                                                                                                                                             0
C.ET.ETH2220                                                                                                                                          0
D.CD.84ZR085                                                                                                                                          0
F1.BE.VI850                                                                                                                                           0
G.SE.SE6165                                                                                                                                           0
H.CF.90CF056                                                                                                                                          0
J.SE.SE7887                                                                                                                                           0
K.CM.MP535C                                                                                                                                           0
01_AE.TH.CM240                                                                                                                                        0
02_AG.NG.IBNG                                                                                                                                         0
03_AB.RU.KAL153-2  -G----------------------C--C----CTG-AG-------G...----CAG-G----------A----C-A---------A-AT--......-AA--C-----------TGA---A-A------    235
04_cpx.CY.94CY032-3                                                                                                                                   0
N.CM.YBF30                                                                                                                                            0
O.CM.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
O.BE.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-    253
CPZ.US.CPZUS       -GC------TT---T----C--TC-CACAGA----G-AG----GCGG...-----T----------C--A--TAC--C-G-----CA--T--......-A------------A--TGAAC-TG-C-----    250
CPZ.GA.CPZGAB      ---------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.CM.CAM3                                                                                                                                           0
CPZ.CM.CAM5                                                                                                                                           0
CPZ.CD.CPZANT                                                                                                                                         0

H2A.GW.ALI Nef     G__W__L__W__K__L__V__P__V__D__.__.__V__P__Q__E__.__G__E__D__T__E__T__H__C__L__L__H__P__V__Q__.__.__T__S__R__H__D__D__T__H__G__E__T    Nef
H2A.GW.ALI         -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.SN.HIV2ST                                                                                                                                         0
H2A.DE.BEN         -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
H2B.GH.D205        -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.CI.EHO         -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.CI.ABT96                                                                                                                                          0
MAC.US.MM251       -C---CTA-GG--AT----C--T--AA-T......GTA-C---T--G...--AC-GG-G-AT----GGCATTAT--AA-G--G--A-CTCAA......ACTTCCAAGTG---------TTG-G----G--    244
SMM.US.SIVSMMH9                                                                                                                                       0
STM.US.STM                                                                                                                                            0
TANTALUS.UG.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
GRIVET.ET.GRI-677  -C-TC----TT--------C-----G--C......CT-CAT--G--G...--ACG---CT-T----GACA-T-TC--A-G--T--A-CACAGATGGGGGAAG--CCTGAT-GAATAGATC-TG-------    250
VERVET.KE.VER-9063 -C-TT----T------------G--A---......-T-CAT-----G...--AG-G-CCT-T----GACATTA----GCG--T----CACAAATAAAAGA-G-CCCCGAT-GCAT-AGTC-TG----GAC    250
VERVET.KE.AGMTYO                                                                                                                                      0
VERVET.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
VERVET.DE.AGM3                                                                                                                                        0
SABAEUS.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
SUN.GA.SIVSUN      -C---CTA-GG-------C------CACC...ATTG-TGA---CAGG...-ATCCT--TCATCCTTGTCAAGC-C-AC-G--TT-CAGCCAG......-AAGGG-TT-AT--A-----TTG-G----GAG    247
MANDRIL.GA.MNDGB1                                                                                                                                     0
LHOEST.KE.hoest    ------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
LHOEST.CD.hoest447                                                                                                                                    0
LHOEST.CD.hoest485                                                                                                                                    0
LHOEST.CD.hoest524                                                                                                                                    0
RCM.NG.SIV-NGM     -C---CTA-GGC-AT----------A--T......GTCAGT--T---...--A-CAG-T-AT----GGCATT-TC-CCCG--T--A-CAGAA......ACTAG---G-----C-----CTG---G--G--    244
COL.__.SIV-CGU1                                                                                                                                       0

SYKES.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
```

```
                                                                                      TCF-1 alpha
B.FR.HXB2            GTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACA..............................................  334

A.UG.U455                                                                                                                                                0
C.ET.ETH2220                                                                                                                                             0
D.CD.84ZR085                                                                                                                                             0
F1.BE.VI850                                                                                                                                              0
G.SE.SE6165                                                                                                                                              0
H.CF.90CF056                                                                                                                                             0
J.SE.SE7887                                                                                                                                              0
K.CM.MP535C                                                                                                                                              0
01_AE.TH.CM240                                                                                                                                           0
02_AG.NG.IBNG                                                                                                                                            0
03_AB.RU.KAL153-2    ----AT-----A---------T-----G--TC-AACA----GA---------------------T--A---AG---------CAGACT........................................  325
04_cpx.CY.94CY032-3                                                                                                                                      0
N.CM.YBF30                                                                                                                                               0
O.CM.MVP5180         AC--A-A---CA------T--ATCAT---GC--AACA--T--A----TGC-AAA---C--A---CT----CCC--G...-A--TGACACTGC......................................  343
O.BE.ANT70           AC-GAT-----A------T--ATCT----GCAACACC--TG-T--TATGATAACT--C--A---CT----C----GGA--A-A-ACTGCTG.......................................  344
CPZ.US.CPZUS         C---ATC---CA--------CTGAA--G--TCGGAGA----GA--TAA-----------A---------CG------A...................................................  329
CPZ.GA.CPZGAB        C--G-TC---C-C---------A-G--G--C--AAGA--T--T---A----A-AA------------A---AG----ACTTT-G.............................................  335
CPZ.CM.CAM3                                                                                                                                              0
CPZ.CM.CAM5                                                                                                                                              0
CPZ.CD.CPZANT                                                                                                                                            0

H2A.GW.ALI Nef       __L__V__W__R__F__D__P__K__L__A__H__D__Y__K__A__F__I__L__H__P__E__E__F__G__Y__K__S__G__L__P__E__D__E__W__K__A__R__L__K__A__R__G__I_  Nef
H2A.GW.ALI           A----TT-----A------CCTAAG--G--TCA-G--T---AA---TTTATT--A--C--A---G-A--TGG-T--AAG-C-GGCCTGCCAGAAGATGAGTGGAAGGCAAGACTGAAAGCAAGAGGGAT.  373

H2A.SN.HIV2ST                                                                                                                                            0
H2A.DE.BEN           A----T----CA-------TC-ATG--G--C-A-A-CT---A----TTCACT---T-C--A--AG-G--TGG-C--AAG-C-GGATTGCCAGAGAAAGAATGGAAGGCAAAACTGAAAGCAAGAGGGAT.  373
H2B.GH.D205          TC-TATC---CA------TTC--T---G----A-G--T-TG----TTTCA-TAG-TT---A--AG-G---GG-T--CAG-C-GGGCTACCAGAGGAGGAGTGGAAGGCTAGACTAAAAGCAAGAGGGAT.  376
H2B.CI.EHO           CC-T-TC---CA-------TC--T--------A-G-CT-TG-----TTCAGCAG-TTC--A---G-G--TGG-T-TCAG-C-GGAATGCCAGAGAAAGAGTGGAAGGCTAAACTGAGAGCAAGAGGAAT.  373
H2G.CI.ABT96                                                                                                                                             0
MAC.US.MM251         TC---C-----A------TCCAACT-----C-ACAC-T-TGA---ATAT-CTAGAT-C--A--AG-G--GG-AGCAA-TCAGGCCTGTCAGAGGAAGAGGTTAGAAGAAGGCTAACCGC...........  363
SMM.US.SIVSMMH9                                                                                                                                          0
STM.US.STM                                                                                    TGGAAGGGATTTATTACAGTGAGAGAAGGCTAACCGC............   37
TANTALUS.UG.TAN1     A--G-T-----AA---A-TCCA-T-------G----GT---ACC-GGAC......AGCTTTA--G--A-GC-TGGTCTTGTTA-GAGGAAGTAGCCTAACCGCAGGCTTGCGGT................  349
GRIVET.ET.GRI-677    C--G-TC----A-------CCGAAGT-G--GG-GG-GT--CGCC-GGAC......ATGTTTA--G--A-GC-CG-ACATGC-A-GCGC...TAGTTTCGCTGGTTAGCCTAACCCGGTTG..........  361
VERVET.KE.VER-9063   C--G-TC----A-------CCTATGT-----G-G--GT-TGACC-AAACAGA-A-T--TTT--AG--A-GC-TGCAATAGTGA-GAGGAAGTAGCTAACCGCAGGCTTGTG...................  361
VERVET.KE.AGMTYO                                                                                                                                         0
VERVET.KE.AGM155     A--G-CA----A------TCCAATGT-G--TG----GT--GACC--TC-AG-GA-T-CTTTACAG----AT-TTCAACAGTTGGTACAGGAAACTAGCCGACCACAGGCTTG.................  362
VERVET.DE.AGM3                                                                                                                                           0
SABAEUS.SN.SAB1C     CC-G-T-----A-------CCA-TA------G-GG-CT-TG-T--TT-GAG------C--A---C-GG-GCC-TCAGCACA-GGCTAACCACAAAACCACATCCTATG......................  352
SUN.GA.SIVSUN        AC-TATC----CA-----TCC-ACT--G--C-A-G--TTT-G----AT-C--AA---C--T--AG-G------C-TGTTACTAGTCTGCAGTGGGAGGTGTAGCTTAACCGCAAACCG............  365
MANDRIL.GA.MNDGB1                                                                                                                                        0
LHOEST.KE.hoest      -C-TATC---CAC--CA-TCCAATGT-G---G-GG-CTTT--A----TCA-AAA---G--T-C-A-GA-AC-A--TACTGCTTTCGCTTTTGACTGCAAGAGGAAGTAGCCAAACCGC............  365
LHOEST.CD.hoest447                                                                                                                                       0
LHOEST.CD.hoest485                                                                                                                                       0
LHOEST.CD.hoest524                                                                                                                                       0
RCM.NG.SIV-NGM       A----CA----A---CA-TCC--T-T-----G-AG--T--GCA-GGTATAGA--------T----T---TGGAG-GA-AAAGA-CAAGACCAAGTAACCACTTCCTCTGGAGTTGCCTTG..........  364
COL.__.SIV-CGU1                                                                                                                                          0

SYKES.KE.SYK173      CC-CAT-----AC--CA--CC--A---G--C-A-ACA-CAGGCTGGGAGAT-GCC-GCTT-C---TAGAG-GAC-AACAG--A-ACCACAAGAGCTGCAGTCAGCACTTAGTAAGAACATCTCCTAGGA.  267
```

```
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  Sp1-III
B.FR.HXB2           ..............................................................TCGAGCTTGCT.ACA...AGGGACTTTCCG....................CTGGGGACTTTCCAGGG.     377

A.UG.U455                                                                                                                                                            0
C.ET.ETH2220                                                                                                                                                         0
D.CD.84ZR085                                                                                                                                                         0
F1.BE.VI850                                                                                                                                                          0
G.SE.SE6165                                                                                                                                                          0
H.CF.90CF056                                                                                                                                                         0
J.SE.SE7887                                                                                                                                                          0
K.CM.MP535C                                                                                                                                                          0
01_AE.TH.CM240                                                                                                                                                       0
02_AG.NG.IBNG                                                                                                                                                        0
03_AB.RU.KAL153-2   .......................................................GCTGACA-GA-.G-----G---....----------.....................GC---------------G     374
04_cpx.CY.94CY032-3                                                                                                                                                  0
N.CM.YBF30                                                                                                                                                           0
O.CM.MVP5180        .................................................................................----------A.......GACTGCTGACACTGC-------------C-T     385
O.BE.ANT70          ...........................................................ACC-GA--A-----G---CTGT--A-------A..GCAAAGACTGCTGACACTGC-------------T-G     413
CPZ.US.CPZUS        .......................................TGGCTGTAACCGCGCAGGCGC.AATA-AAC----G--TG...A--------TAA...GGGACTTTCCAA.......-----G----..AAG     404
CPZ.GA.CPZGAB       ..............................................ACTGGCGCATGCGC.ACAAGAAC----G--TCTGC----------AA......................----------G---A     396
CPZ.CM.CAM3                                                                                                                                                          0
CPZ.CM.CAM5                                                           GCGCCTGCGCAG-A--AAC----G---A...---------TAA......................-----------A.-G      54
CPZ.CD.CPZANT                                                                                                                                                        0

H2A.GW.ALI Nef      __P__F__S__K__N__R__N__S__$                                                                                                                Nef
H2A.GW.ALI          ACCATTTAGTAAGAACAGGAACAGCTGA..TTTGGTCAGGGCAGGAAGTAACTACTG.....AAA-C.AGCTGAGACTGC-----------A.....................GAA--GGC-GTA-CCAG     474

H2A.SN.HIV2ST                                                                                                                                                        0
H2A.DE.BEN          ACCATATAGTGA...ATAACAGGAACAACCATACTTGGTCAAGGCAGGAAGTAGCTA.CTAAGAA-C.AGCTGAGGCTGC-----------A.....................GAA--GGC-GTA-CCAA     477
H2B.GH.D205         ACCTACAGATTAGGCAAGAGACAGCAGCATAAACAGGAACTAGCTGACACTGCACAA.GAAGGAA-CTAGCAGACACTGC-----------A.....................AAA-AGGC-GTA-CA.T     483
H2B.CI.EHO          ACCTACAGAGTAGACAGGAAATAACAGCACAAAGAGGAACTAGCTAACACTGCATAGAGAAGGAA-CTAGCTGATACTGC---.-------A.....................GAA--GGC-GTA-CA-T     481
H2G.CI.ABT96                                                                                                                                                         0
MAC.US.MM251        ...............................AAGAGGCCTTCTTAACATGGCTGACAAGAGGGAA-CTCGCTGAGATAGC-----------A.....................CAA--GGA-GTT-T--G     441
SMM.US.SIVSMMH9                                                                                                                                                      0
STM.US.STM          ...............................AAGAGGCCTTCTAAAGATGGCTGACAAGAAGGAA-CAAGCTGAGACAGC-----------A.....................CAA--GGC-GTA-.CAG     114
TANTALUS.UG.TAN1    ...........................TAAGCTAACTAGGCCGTTGCCTAGGAGATG.ACATAAAGAAG----G-TCAGC-----------A..................................---T     417
GRIVET.ET.GRI-677   ..................................................................................................................................     361
VERVET.KE.VER-9063  ...................................GTTAAGCCGTTGCCAGGGAGATGACAT-T--AAC----G---G...----------A......................A--------------C     431
VERVET.KE.AGMTYO                                                                                                                                                     0
VERVET.KE.AGM155    .................................CGGTTTCCTGGTTGCCTAGGAGATGACAT-AAGAAC----G--.....----------A...................G-AC-----------A--C     435
VERVET.DE.AGM3                                                                                                                                                       0
SABAEUS.SN.SAB1C    .........................................GAGTTGTCATGGTGATGACAT-AAGAAC----G--TGAG-----------A..........................GGG-GGAGACTG     415
SUN.GA.SIVSUN       .......................CATCCTCTTGCATCGCCTAGGCAACGGGGCTAGC.GCAT......GC--GCTAGAGT------------......................................     427
MANDRIL.GA.MNDGB1                                                                                                                                                    0
LHOEST.KE.hoest     ...AAACCGCAGCACATCCTCTTGCAGCCCGGTTGCTAAGGCAACCGGGCTAGCGCATGCGCAATG-------G--GAGC-----------A.............GGACGGG-G-----GG--T......     473
LHOEST.CD.hoest447                                                                                                                                                   0
LHOEST.CD.hoest485                                                                                                                                                   0
LHOEST.CD.hoest524                                                                                                                                                   0
RCM.NG.SIV-NGM      ..........................................GAAACAGCGCATGCCTACTAAGA--A-GCTGACACAGA-----------ACTGGCGCCTGCGCGCTGGTGTAA------------ACT     452
COL.__.SIV-CGU1                                                                                                                                                      0
SYKES.KE.SYK173     GACTCCATGGTGACAAGCTCGGCCCAGGGGA..................................................-----C-GGGC..................GGGG-AA-G--CA--CT-CT     329
```

```
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
                    Sp1-III      Sp1-II      Sp1-I                             TATA Box                     |||||^|||||||||||^
B.FR.HXB2           AGGCGTGG.CCTGGGCGGGACTGGGGAGTGGCGA.........GCCCTCAGATCCTGCATATAAGCAGCTGCT...TTTTGCCTGTACTGGGTCTCTCTGGTTAGA.......................    470

A.UG.U455                                                                                                                                                  0
C.ET.ETH2220                                                                                                                                               0
D.CD.84ZR085                                                                                                                                               0
F1.BE.VI850                                                                                                                                                0
G.SE.SE6165                                                                                                                                                0
H.CF.90CF056                                                                                                                                               0
J.SE.SE7887                                                                                                                                                0
K.CM.MP535C                                                                                                                                                0
01_AE.TH.CM240                                                                                               ----------T------.......................     17
02_AG.NG.IBNG                                                                                                                                              0
03_AB.RU.KAL153-2   ---T----.TT-------AGT-----------T-.........A---------G-------------------...---C--------------------------.......................    467
04_cpx.CY.94CY032-3                                                                                                                                        0
N.CM.YBF30                                                                                 T-C-C--T----------------T-C-G--.......................     31
O.CM.MVP5180        G--A-G-ATAAG------TT-G---AGTG-C..........TAA---------G-------------------...--CC--T-----C------TAG-TAGAG--.......................    478
O.BE.ANT70          GA-G-AC..AGG------TT-G---AGTG-C..........TAA--------AG---------------C---...--C---T-----C-------GG-TAGAG--.......................    504
CPZ.US.CPZUS        G--GTG--T-A.------A........ACA-G-CGTGGTTT.TA------A-GA-------A-------C---...C-C---T--------------T.C-C----.......................    496
CPZ.GA.CPZGAB       GA-GTG-TCGGGA------G.......T---G-TGTGGCTT.T--------.AG-------------------...--C------------------T-CAC-G--.......................    490
CPZ.CM.CAM3                                                                                                                                                0
CPZ.CM.CAM5         G--A-GCTCA..-------........ACA-G-CGTGGTTT.CA--------AG---------------C---...--C---T-----GT---------T---G--.......................    146
CPZ.CD.CPZANT                                                                                                                                              0

H2A.GW.ALI          G--AAG-AC.A----A--AG----T-G-GAA-..........GC--TCATAC-----T---A-T-T-C-C---.GC--GCATTGTATTCA-TCGCTCTGC-GAGAGGCTGGCAGATTGAGCCCTGGGAG    591
H2A.SN.HIV2ST                                                                                A-TCGCTCTGC-GAGAGGCTGGCAGATTGAGCCCTGGGAG     40
H2A.DE.BEN          G--A-G-AC.A----A--AG----T-G-GAA-.........GCC-T-AT-CT-A---T---A-T-T-C-C---.TC--GCATTGTATTCA-TCGCTCTGC-GAGAGGCTGGCAGATCGAGCCCTGAGAG    595
H2B.GH.D205         G--A-GA-C.AA---A---GT---TTG-GAA-.........GCC-AT-A-ATCT---T---A-T-T-C-C---.TC-CGCATTGTATTCA-TCGCTCTGC-GAGAGGCTGGCAGATTGAGCCCTGGGAG    601
H2B.CI.EHO          G--A-G-AC.A----A--AG-A---AG-.AA-.........GCC-ATCA---CT---T---A-T-T-C-C---.TTG-GCATTGTATTCA-TCGCTCTGC-GAGAGGCTGGCAGATCGAGCCCTGGGAG    598
H2G.CI.ABT96                                                                                                                                               0
MAC.US.MM251        GA-GAGCCG.G-C--GAAC-......................CC-A--TTCT-GA--T---A-TATCA----A.TT-CGCT-TGTATTCA-TCGCTCTGC-GAGAGGCTGGCAGATTGAGCCCTGGGAG    546
SMM.US.SIVSMMH9                                                                             CA-TCACTCTGC-GAGAGGCTGGCAGATTGAGCCCTGGGAG     41
STM.US.STM          G--A-GTAC.TG--AG-A-CTG-T--G-.AA-.........GCC-TT-T-CT-T---T---A-TA--A-C---.TA-AGCT-TGTATTCA-TCACTCTGC-GAGAGGCTGGCAGATTGAGCCCTGGGAG    231
TANTALUS.UG.TAN1    G-TGCGA-.GGC--TACTTGGGA-T-GC-T..............---CTCAGAG-------A------A----CGC-GGCTTG-AACTCA-TCTCT-ACTAGG---.......................    508
GRIVET.ET.GRI-677   ............-CCACC.GGG--TA--GAC.............T---TG-C-T..------GCTCAA-AAA.........................................................    405
VERVET.KE.VER-9063  G---CAT-.GGC--T.AC-GGGA-T-GT-T.............TA--CTCAGAG-------A------A----CGC-GGCTTG-AACTCA-TCTCT-ACTAGG---.......................    522
VERVET.KE.AGMTYO                                                                            CA-TCTCT-ACTAGG---.......................     18
VERVET.KE.AGM155    G--ACAT-.GGC--T.AC-GGGA-T-GC-T.............TA--CTCAGAG-------A------A----CGC-GGCTTG-AACTCA-TCTCT-ACTAGG---.......................    526
VERVET.DE.AGM3                                                                              CA-TCTCT-ACTAGG---.......................     18
SABAEUS.SN.SAB1C    G-CG--ACTGGGA-TG-CTT.....................GCC--G-GCTGCT-AAT---AGCA...GCTGCTTTGCGCCTG-AA-ACA-TCTCTG--T-GAG..GCTGCCAGATTGAGCC.TGGG.T    516
SUN.GA.SIVSUN       ............GACAG--AGGG--A-..............GCTTA-C--CT-A-------------------.CT-CGCCTG-T.CGG-A-.....................................    491
MANDRIL.GA.MNDGB1                                                                     -GAGTC-C-ACTACAG-GGCTAAGGG-TGTA-CTCTGAG.CAGATCCCCTTAGAGCAAG     58
LHOEST.KE.hoest     .............................................G-C--CT-G-------------------..C-CCGCTTGT-CGG-A-.....................................    518
LHOEST.CD.hoest447                                                                                                                                         0
LHOEST.CD.hoest485                                                                                                                                         0
LHOEST.CD.hoest524                                                                                                                                         0
RCM.NG.SIV-NGM      GACGTG--AGGG---T-T-G--T-CCCTCTC-..................TGCTGCAT---AGCAGCTGCAT-TCGC-C--TAACC-G-CTC--C--GG-AGGCT-CCAGATTGAGCCTGGGTTGTTCT    563
COL.___.SIV-CGU1                                                                                                                                   TGGAGT      6
SYKES.KE.SYK173     GC........................................................---A----C-A----...................CTGTAGCTA-CC-CGGAGTTGCAACTAGCTCTCTCCA    383
```

```
                                 TAR element bulge  loop     TAR element stem
                                       |||||^^^||||      |||||||||||| ||||||||||||||
B.FR.HXB2           ...................CCAGATCTGAGCCTGGGAGCTCTCTGGCTA.ACTAGGGAACCCACTGCTTA...........AGCCT............................................    525

A.UG.U455                                                                                                                                                 0
C.ET.ETH2220                                                                                                                                              0
D.CD.84ZR085                                           ------------CG.G-------------------...........-----............................................     39
F1.BE.VI850                                                                                                                                               0
G.SE.SE6165                                                                                                                                               0
H.CF.90CF056                                                                                                                                              0
J.SE.SE7887                                                                                                                                               0
K.CM.MP535C                                                                                                                                               0
01_AE.TH.CM240      ...................----G--.-----C----------------.G-A-----------------A..........-----............................................     72
02_AG.NG.IBNG                           -----------------------------.G-GGA---------------...........-----............................................     54
03_AB.RU.KAL153-2   ...................-------.----------------------.G---A---------------...........-----............................................    521
04_cpx.CY.94CY032-3                                                                                                                                       0
N.CM.YBF30          ...................------TA----------------------.G-AG--A-C--ACTG.----...........-----............................................     85
O.CM.MVP5180        ...................----G--------C--------C-----CT.CTAGCT------G-------...........-CG--............................................    533
O.BE.ANT70          ...................----G--------C--------C-----CT.CTAGCT------G-------...........-CG--............................................    559
CPZ.US.CPZUS        ...................------AA-----C----------------GTG--.---------------...........-----............................................    550
CPZ.GA.CPZGAB       ...................------T------------------A.-C-GTGA--.--------------...........-----............................................    544
CPZ.CM.CAM3                                                                                                                                               0
CPZ.CM.CAM5         ...................------TA------------------..GCC-GC-----------------...........-----............................................    200
CPZ.CD.CPZANT                                                                                                                                             0

H2A.GW.ALI          GTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAGTG--TGGCCGGCACTGGGCAGA-.GGCTCCACGCTTG................................    683
H2A.SN.HIV2ST       GTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAGTG--TGGCCGGCACTGGGCAGA-.GGCTCCACGCTTG................................    132
H2A.DE.BEN          GTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTGG.---CTCACCAGT---TGGCCGGTACTGGGCAGA-.GGCTCCACGCTTG................................    687
H2B.GH.D205         GTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAG----AGGCCAGTGTTGGGCAGA-.GGCTCCACGCTTG................................    693
H2B.CI.EHO          GTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAG----TGGCCGGTGCTGGGCAGA-.GGCTCCACGCTTG................................    690
H2G.CI.ABT96                                                  ---CTAG.---CTCACCAG-G--TGGCCGGTGCTGGGCAGAGYGGCTCCACGCTTG................................     55
MAC.US.MM251        GTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAG----TGGCCAGTGCTGGGCAGAGTGGCTCCACGCTTG................................    639
SMM.US.SIVSMMH9     GTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTGG.---CTCACCAG----TGGCCGGTGCTGGGCAGAGTRGCTCCACGCTTG................................    134
STM.US.STM          GTTCTCT.CCAGCACTA..G---G-.A-------A-T-T--C---CTAG.---CTCACCAG----TGGCCAGTGCTGGGTAGAGTGGCTCCACGCTTG................................    324
TANTALUS.UG.TAN1    ...................----.CT------------T--G----TG-.G-CTAACCTGGTCTG--CATCCAGGGGTAAG-----C...........................................    574
GRIVET.ET.GRI-677   ..................................................................................................................................    405
VERVET.KE.VER-9063  ...................----C-T.---------T-T--G----T--.G-CTAACCTGGT.TG--CA.CCAGGGGTAAGGA-TC............................................    585
VERVET.KE.AGMTYO    ...................----C-T.---------T-T--G----T--.G-CTAACCTGGT.TG--CA.CCAGGGGTAAGGA-TC............................................     81
VERVET.KE.AGM155    ...................----C-T.---------T-T--G----T--.G-CTAACCTGGT.TG--CA.CCAGGGGTAAGGA-TC............................................    589
VERVET.DE.AGM3      ...................----C-T.---------T-T--G----T--.G-CTAACCTGGT.TG--CA.CCAGGGGTAAGGA-TC............................................     81
SABAEUS.SN.SAB1C    GTTCTCTGGTAAGTCTGAA----C-T.---------T-T--G----GG-AGGC-A-CCTAAG-GCA--CTGGTGAGGTCTTG-TAGACCCTCGCCTGGCGAC.TGGCCATTGCCAGTAGCAGAGACTCCG    644
SUN.GA.SIVSUN       ..TCTCT.GCT.CAGAGGCTAG-CGAC--C----A---G----ACCTAG.-GCT--CTTGT-G--C--......GAGTGAGCA-T-GCTGAGCACCTCACCCGAGCCTGATC...............AGC    595
MANDRIL.GA.MNDGB1   GACCAGAGTCCT.............GA-T-A-----TCTGAG-ACCTC-.CTCG---CTGATCACCTCG-GGTAGTGGAACTC-T-GCTTGCTTGC...TATTGTCTTC.....................    150
LHOEST.KE.hoest     ..TCCTT.GCTCTGGGGAC-AGC-GTC-CT------T.GAGCACACTGG.T-AGCTA-G-AGT-A--CCTGGGAGGCTACACTTAGCACCCACCTGGGCCTGATCA..............GCCTAGGGAG    629
LHOEST.CD.hoest447                                                                                                                                        0
LHOEST.CD.hoest485                                                                                                                                        0
LHOEST.CD.hoest524                                                                                                                                        0
RCM.NG.SIV-NGM      CTGGTAA.GTCTCTAGGAGA-TAGCT----------T-T--G--A-T-C.T--...---GGGCT----GGGAGCGCCTTGGCTGGGTCTTGGTAGACCTTCAGCTGGGTA...CTTGGCCAGTCCCAGTG    685
COL.__.SIV-CGU1     TCTGACT.CTTGCCTGGCTG-GCC-GA-CC-TGA-TG---TG-CA----.T-AGCCTTG--TGAGT--G-GTAGCTTG....................................CTTGGTTGAAGGGCTT     98
SYKES.KE.SYK173     GCGG..........................---------C--GA-AGAC...-...-GTAGAG-G--CCGGGAGCCCTTGCCAGTAAAGGCTAGTGAG.GTCCGGGTTTG.....GCCGCCCGGTCTTCG    475
```

```
                                                               secondary structure in this region, see Rizvi, J Virol 67:2681-2688(1993)
                                                                         Poly-A signal  5’LTR R Repeat end \/ 5’LTR U5 start
B.FR.HXB2             ......................................................C.AATAAA.GCTTGCCTTGAGT....GCTTCAAGTAGTGTGTGCCCGT.CTGTT.GT........GTGACTCTGGT    585

A.UG.U455                                                                   --.-------------....----G-----------------.-----AT-.........----------     55
C.ET.ETH2220                                                                                                                                                0
D.CD.84ZR085          ......................................................-.------.-------------....----T-----------------.-----TT.........-----------     99
F1.BE.VI850                                                                                                                                                 0
G.SE.SE6165                                                                                                                                                 0
H.CF.90CF056                                                                                                                                                0
J.SE.SE7887                                                                                                                                                 0
K.CM.MP535C                                                                                                                                                 0
01_AE.TH.CM240        ......................................................-.------.-------------....----A----G------------.----G.T-........AG---------    132
02_AG.NG.IBNG         ......................................................-.------.-------------....----------------------.---A-T--........-----------    115
03_AB.RU.KAL153-2     ......................................................-.------.-------------....--NNNNNNNNNNNNNNNNNNNN.NNNNNNNN........NNNNNNNNNNN    582
04_cpx.CY.94CY032-3                                                                                                                                         0
N.CM.YBF30            ......................................................-.------.-------------....---AA-.--G----------A-.-CA--C-.........--A--------    144
O.CM.MVP5180          ......................................................-.------.-------------GAGAAGCA......--------T-A-CTG.--CAACC............-----    588
O.BE.ANT70            ......................................................-.------.-------------GAGAAGCA......--------T-A-CTG.--CAACC............-----    614
CPZ.US.CPZUS          ......................................................-.------.-------------....----T--------------TA-C-.T-CA..........-----------    609
CPZ.GA.CPZGAB         ......................................................-.------.-------------...GTA--TG--C----------TA-CTAT-CA..........-AC--------    605
CPZ.CM.CAM3                                                                                                                                                 0
CPZ.CM.CAM5           ......................................................-.------.-------------....-AC-TG--C----------TA-C-AT-CA..........-----------    260
CPZ.CD.CPZANT                                                                                                                                               0

H2A.GW.ALI            ...................CTTG.CT.T..AAAG.....ACCTCTT..........------G.--.---AAT.TAGAAGCAGGTT-AAG------T---A-CT.C-CCTAGTCGCCGCC--GTCA-.TC    770
H2A.SN.HIV2ST         ...................CTTG.CT.T.AAAAG.....ACCTCTT..........------G.--.---AGTTAGAAGCAAG-T--.GT-----CT---A-CT.C-CCTAGTCGCCGCC--GTCA-.TC    220
H2A.DE.BEN            ...................CTTG.CT.T..AAAG.....ACCTCTT........-.------G.--.---A.-TTAGAAGCAAGTT-AGT------T---A-CT.C-CCTAGTCGCCGCC--GTCA-.TC    775
H2B.GH.D205           ...................CTTG.CT.T.AAAA......CCCTCTT..........------..--.---A.ATTAGAAGCAAGT.-AGT------T---A-CT.C-CCTAGTCGC.GCC--GTCA.CTC    777
H2B.CI.EHO            ...................CTTG.CT.T.AAAA......CCCTCTT..........------A.--.----.ATTAGAAGCAAGAC-AGT------T---A-CT.C-CCTAGTCGCCGCC--GTCA-.TC    777
H2G.CI.ABT96          ...................CTTG.CT.TACAAAG.....ACCTCTT..........------G.--GC-AAATTAGAAGTAAGCA-GTGT------T---A-CT.C-CCTAGTCGCCGCC--GTCACCTC    147
MAC.US.MM251          ...................CTTG.CT.T.......AAAGACCTCTT........-.------G.--.---A-TTTAGAAGTAAG-C-.GT------T---A-CT.C-CCTAGTCGCCGCC--GTCAACTC    728
SMM.US.SIVSMMH9       ...................CTTG.CT.T.AAAAG.....ACCTCTT........-.------G.--.---AGTT.AGAAGTAAG-C-.GT------T---A-CT.C-CCTAGMCGCCGCC--GTCA-CTC    223
STM.US.STM            ...................CTTG.CT.T.......ATATATCTCTT........-.------G.--.---AATTTAGAAGTAAG.C-AGT------T---A-CT.C-CCTAGTCGCCGCC--GTCAACTC    413
TANTALUS.UG.TAN1      ...................CTTG.CT.TAGAAGGCTATTG..............-T------.-.------GCTTAAGAGT--ATCT-AG.CAA-----.TC.A-TGC.............-CGC-CTTC    651
GRIVET.ET.GRI-677     .........................................................................CC-GCTC----AGTCGCTATAT--GAGTCAA.--G..CTCATTGCT-C-C-GAGCC-    459
VERVET.KE.VER-9063    ...................CTTGGCT.TGGAAAG....................-T------ACT------GC-T-AGAGCT-ATCT-AGTCAA------TCAT.TGAC-CCTCACTC..AAG-AGG--A    672
VERVET.KE.AGMTYO      ...................CTTGGCT.TAGAAAG....................-T------..-------GC-T-AGAGCT-ATCT-AGTCAA------TCAT.TGAC-CCTCACTCTC-TGAA-G--A    168
VERVET.KE.AGM155      ...................CTTGGCT.TAGAAAG....................-T------..-------GC-T-AGAGCT-ATCT-AGTCAA------TCAT.TAAT-CCTCACTC..-TGAA-G--A    674
VERVET.DE.AGM3        ...................CTTGGCT.TAGAAAG....................-T------.T---CG--GC-T-AGAGCT-CTCT-AGTCAA------TCAT.TGAC-CCTCACTCTTGA--GGGTAA    169
SABAEUS.SN.SAB1C      ...................CTTG.CT.T.GCTTG..ATTGCCT...........-.------GT--------AGT-TACAAG-AA-GCA-----------A-TT.A--CC-CAGGACAACCCT.......    730
SUN.GA.SIVSUN         TTGGGGAGCTAGAGGCTC.CTTG.CT.TTGCAGTACAAG...CCTT........-.------G.-C------AGT-TTACTGCAAGCAA...----T-----GT-C-C.TCTTTCTCCTTCT....AAAC    701
MANDRIL.GA.MNDGB1     ........................................................------GTAACTTAGAATTAGAGCAAG-G-GTA-----TAT--AT-GTGCGCCTCTCTTCT.............    211
LHOEST.KE.hoest       C.TAAGGGCTCCTAG....CTTG.CT....ACTTTAAAG...CCTT........-.------AAG----TCATT--TAGAAAG-AC-AGCTA----CT-ATAGT.AG-CTCTTTCTCCTTC-C-CTG-T-    736
LHOEST.CD.hoest447                                                                                                                                          0
LHOEST.CD.hoest485                                                                                                                                          0
LHOEST.CD.hoest524                                                                                                                                          0
RCM.NG.SIV-NGM        GGAGTAGGCTCCG......CTTGC....TTTCTGGTTA..ATTGCT........-.------CTGCATG--CTTAGTTGAA-CAG--TC-A.-----TG-CCAT.T--ATCTCTCAGTGAAACGA--CTG    792
COL.__.SIV-CGU1       GAGGGACTCCAGTCCCCTCGCGCTAAGGGTAGATAGGTGCTTTGCATTATAAGAGC------GTGA-TTTGAAGC-TCAA--AG-GTT-GAGTAC-AGTTTCA-GTCCGCCTTCTC..............    214
SYKES.KE.SYK173       GCAGCTCTCCGTTG.....CTTGC.TTTGCTTACGCTGATCAGCCCA.........------GTG-G-TAAACGCAAGTAA-G--TGC-CTCA-C--TT-TGCTG-G-CACCTCCCGAA-GTG---A--G    590
```

```
                                                                                       see Zhang, J. Virol 72(7):5664-5471(1998)
                                                                                     5'LTR U5 end <-| Lys tRNA primer binding site
B.FR.HXB2            AACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGG................................AAAATCTCTAGCAG..TGGCGCCCGAACAGGG.ACCTGAA.........AGCGAAAGG.  667

A.UG.U455            --------------------AC-A---A-T----A.................................--------------..----------------.--TC---.........--------T.  136
C.ET.ETH2220                                                                      -------------..----------------G-------.........--------T.   46
D.CD.84ZR085         -------------------AGC-------C-G----................................--------------..----------------.-------.........--------T.  181
F1.BE.VI850                                                                                                       --.--T----.........--------T.   17
G.SE.SE6165                                                                    GGCTT.-------------..----------------.--T----.........--------T.   50
H.CF.90CF056                                                                                                           T---T.........--T-----T.   14
J.SE.SE7887                                                                                                                                          0
K.CM.MP535C                                                                                                                                          0
01_AE.TH.CM240       --------------------AC-C---A-T-A--A.................................----------C---..----------------C--TC---.........--------T.  214
02_AG.NG.IBNG        --------------------AC-C---A-T----A.................................--------------..---------------A..-T---C.........G-TA-T---.  195
03_AB.RU.KAL153-2    NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN.................NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN..NNNNNNNNNNNNNNNN.................NNNNNNNNN.  672
04_cpx.CY.94CY032-3                                                                                                    T----.........--T-----T.   14
N.CM.YBF30           -C------------------A-C-AGACTGAGTGA.................................--------------..----------------.--T----.........-A------T.  225
O.CM.MVP5180         GT----------------T-AC--AGAC.T-AAGCAG...............................--------------..----------------.-.-GCG-.........-AGTG--A-.  669
O.BE.ANT70           GT----------------T-AC--AGAC.T-AAGCAG...............................--------------..----------------.--T----.........GTGA--GT-.  696
CPZ.US.CPZUS         ------------------A....C-CT-A-AA--A............................GGTG.--------------..----------------.--T----.........-A------..  689
CPZ.GA.CPZGAB        ------------------T....-A-A-T-TA--CA...........................AGG..--------------..----------------.-------..........-T-----..  684
CPZ.CM.CAM3          ..............................................................................----..----------------.--T----.........--TTG--A-.   36
CPZ.CM.CAM5          ------------------A....CA-T-ATAA--A-............................GTG.--------------..----------------.--T----.........--T-----..  340
CPZ.CD.CPZANT                                                                                                   ---.--T---..............GG--..   13

H2A.GW.ALI           GGTGTTCACCTG.AGTA--AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGGAAACGGAGGCAGG------C-------GT----------------.--T----.........GAA--CT-A.  885
H2A.SN.HIV2ST        GGTGTTCATCTAAAGTA--AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGGAAACCAAGGCAGG------C-------GT----------------.--T----.........GAA--CT-A.  336
H2A.DE.BEN           GGTGTTCATCTGAGT-AC.AAGACCCTGGTC---TAGGACCCTTCTCGCTTTGGGAATCCAAGGCAGG------C-------GT----------------.--T----.........GAG--CT-A.  890
H2B.GH.D205          GGTGCTCC-CTGA......TAGACCCTGGTC---TAGGACCCTTCTTGCTTTGGGAAACCGAAGCGGG------C-------AT----------------.-----G..........GAA--GT-A.  886
H2B.CI.EHO           GGTGCTCC-CTAG.......AAACCCTGGTC---TAGGACCCCTTCTGCTTTGGGAAACCAAGGCAGG------C-------AT----------------.------G.........GAA--GT-A.  886
H2G.CI.ABT96         GGTGTTC-CGTAGC.....AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGRGAACCAGAGCGGG------C-------AT----------------.--T---G.........-AG--GT-A.  258
MAC.US.MM251         GGTACTC-G-AATAAG...AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGAGAAACCGAAGCAGG------C-------AT---------------A..-T----.........G-A--GT-A.  840
SMM.US.SIVSMMH9      GGTACTC--A-ATAAG...AAGSCCCTGGTC---TAGGACYCTTTCTGCTTTGGGAAACCGAGGCAGG------C-------GT----------------.--T----.........G-A--GT-A.  336
STM.US.STM           GGTACTC............AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGGCTACTGAGGCAGA------C-------AT----------------.--T----.........G-A--GT-A.  517
TANTALUS.UG.TAN1     CT-.--G--A-----GTTA-TGG.....................GGTTTCTTAT.....CCAGGCTGAG-------C-----..----------------.--T---G.........TCA-----A.  738
GRIVET.ET.GRI-677    CTAG--GTGAA----TC.TTAC-G....................GGTT..CTCCTGTACCCAGGTGGG-G--A---C-----..----------------.--T---G.........-AGAGGCAT.  550
VERVET.KE.VER-9063   -C-GTTCCT-A-TG..............................GGTTCTCTCTCTGACCCAGGCGGG-G--A---C-----..----------------.--T---GGTCCGAGGACTT--GT-A.  765
VERVET.KE.AGMTYO     -T-.TTCCT-A-TG..............................GGTTCTCTCTCTGACCCAGGCGAG-G--A---C-----..----------------.--T---G.........T-A--GT-T.  251
VERVET.KE.AGM155     G-AGTTCCT-A-TG..............................GGTTCTCTCTCAAACCCAGGCGAG-G--A---C----...----------------.--T-...............--GT-A.  751
VERVET.DE.AGM3       ---.TTCCT-A-TG..............................GGTTCTCTCTCA.ACCCAGGCGAG-G--A---C-----..----------------.--T---C.........T.T--GT-A.  250
SABAEUS.SN.SAB1C     ............................................GGTTACTAAGGATCCCTGATAGAA------CT------..----------------.--T---C......TAA--A.CGTCCC  803
SUN.GA.SIVSUN        CCTGTTC-............AC-GCTC-T-ACTA--GAGCAATTACTGGCCGCCTAGTGATCCGGGGCTG-CGGC.------..----------------.--A---G.........G-A--G--A.  802
MANDRIL.GA.MNDGB1    ...................AAACC-GT-GT--TCTCATTTAGAGAACAGAAGGACTTCTAGTTAACCCT-G-AGC--TT---..-------------.--.--T----.........GAGAGGCAC.  305
LHOEST.KE.hoest      C-G-GATCTCT-ACT---GAGA--GGAG-CT---...........................GATCCGGG-C-GGG--TC---..----------------.--T----.........G-A---GA-.  823
LHOEST.CD.hoest447                                                                                                                                   0
LHOEST.CD.hoest485                                                                                                                                   0
LHOEST.CD.hoest524                                                                                                                                   0
RCM.NG.SIV-NGM       GGG-T..-GGAT-C-TC-GATA---GT.................GGCAGAAGTACTGAAGGCTAAGAA------C-A-C---..----------------.--T---G.........--T--CT-A.  887
COL.__.SIV-CGU1      .....................-GCCCT...C....AGAGCGGGGCCTTTTTGCTG....TCCGGGAAGG---GGC--T----..---------G------.--T----.........--A-G.TCC.  295
SYKES.KE.SYK173      -G-GC-.....................................................................-CTT---..-------G.-------.--T---C.........TAG-CC---.  634
```

```
                                                                     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)
B.FR.HXB2           .GAAA.....................CC..AGA...GGAG...CTCTCT.CGACGCA...GGACTCGGCTTGCTGAAG.CGCGCAC.GGCAAGAGGCGAGG..GG....CGGCGACTGGTGAGTACG    752

A.UG.U455           .T.......................--.---...-A--...T-----.-------..----------------.-.GTGCACACA------------A..-.....------------------    217
C.ET.ETH2220        .--G-....................--.---...-----...A-----.-------..----------------.A.GTGCACTC-------------A..-.....------------------    130
D.CD.84ZR085        .AG--....................--.---...-----...A-----.-------..----------------.A.GCGCGCTC---------------..--....-A----------------    267
F1.BE.VI850         .AG--....................--.---...-A--...A-----.------GG..----------------.A.GTGCA--AC-GC-AGA-GCGA-..A-....------------------    103
G.SE.SE6165         .T--CAGGGACTCGAAAGCGAAAGTT--.---...-A--...T-----.-------..----------------.-.GTGCACACA------------A..-.....------------------    155
H.CF.90CF056        .A---....................--.---...-A--...A-----.-------..----------------.A.GTGCACACA------------A..-.....------------------     98
J.SE.SE7887         ..........................................T-----.-------..----------------.A.GTGCACAC---------------..--....------------------     72
J.SE.SE7887                                                  .T-----.-------..----------------.A.GTGCACAC---------------..--....------------------     72
K.CM.MP535C                                                                                                                                               0
01_AE.TH.CM240      .T--TAGGGACTCGAAAGCGAAAGTT--.---...-A--...T-----.-------..----------------.-.GTGCACACA------------A..-.....------------------    319
02_AG.NG.IBNG       .-.......ACTCGAAAGCGAAAGTT--.---...-A--...A-----.-------A..----------------.-.GTGCGCACA------------A..-.....------------------    294
03_AB.RU.KAL153-2   .NNNNNNNNNNNNNNNNNNNNNNNNNN..NNN...NNNNNN.NNNNNN.NNNNNNN..............NNNNNNN.NNNNNNNNNNNNNNNNNNNNNN..NNNNNNNNNNNNNNNNNNNNNNNN    774
04_cpx.CY.94CY032-3 .T--TA.GGACTCGAAAGCGAAAGTT--.---...-A--...T-----.-------..----------------.-.GTGCACACA------------A..-.....------------------    118
N.CM.YBF30          .AG--....................---..G--...--CTGA.A-----.-------..----------C-T--....GTGCACACA--G--------------.......-----GAA.---T-AGTA    307
O.CM.MVP5180        .TGG-A...................--..--G...-A--AA.AA-CTC.-------AC.--G--------AGC-G.A.GTGCAC-T-CT----------A..--AACT-ACAAGAG-------...    760
O.BE.ANT70          .A--.....................--..--G...-A--AA.AA-CTC.-------AC.--G--------AGC-G.A.GTGCAC-C-CT----------A..--AACT-ACA.GAG--------..    785
CPZ.US.CPZUS        TAC--....................--AG--GCTGAA-C...T--.-GA--...-G..----------C.GCTG.A.GTGCACACA--G-----------..C-AACT-.CA--..---------A    777
CPZ.GA.CPZGAB       TA-......................--GG--GCT.-A-T...---C-GA--...--C..-----------.GCTG.A.GTGCACACA--------------..C-GCGA-.T.....----------    768
CPZ.CM.CAM3         T-G--....................--.---....--CTGA.----TC.-------CG.-----------.GCTG.A.GTGCACACA--G-----------..C-AACT-.CA--..-------G..    124
CPZ.CM.CAM5         AAG--....................--.---....--CTGA.----TC.-------C..-----------.ACTG.A.GTGCACGCA-T------------..C-AC.T-.CT--..--------...    425
CPZ.CD.CPZANT       .........................-TAC---....A-C...--G-.GA--...-CA..--------TC-.GCTG.CAGCGCGCGTA---GA---------..C-GAA.....-...---------C     89

H2A.GW.ALI          .---GTCTAG.GAACACGGCTGAGTGAA..G-C...A-TAAG.GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCCT.A-AAAGGC-CAGGCCAAG-TAC..CAAAGG.CCG-CG--TG---CGG-    998
H2A.SN.HIV2ST       .---G.CCTTGGAACACGGCTGAGTGAA..G-C...A-TAAG.GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCCT.A-AAA-GC-CAGGCC-AG-TAC..CAAGGG.C-GCGTGT-GAGCGGGA    449
H2A.DE.BEN          .---GCCCTG.GAACTCGGCTGAGTGAA..G-C...A-TAAG.GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCCT.A-AAAGGC-CGGGCC-AG-TAC..CAAAGG.C-GCGTGT-GAGCGGGA   1003
H2B.GH.D205         .--GTCTTCAAGAGCACAGCTGAGTGAA..G-C...A-TAAG.GG-GGC.A-GAATCAA.CC--GAC-GAGAGCTCCT.GTAAA-GC-CAGGCC..G-TAC..CAGGCA...GCGTGA-GAGCGGGA    996
H2B.CI.EHO          .--GTCTTCG.GGACACGGCTGAGTGAA..G-C...A-TAAG.GG-GGC.A-GAATCAA.CC--GAC-GAGTGCTCCT.GTAAA-AC-CAGGCC..G-TAC..CAGGCA...GCGTGA-GAGCGGGA    995
H2G.CI.ABT96        .--GTCCTTT.GAACACGGCAGAGTGAA..G-C...A-TAAG.GG-GGC.A-GAT-CAA.CC--GAC-GAGTG-TCCT.A-AAAGGC-CGGGTC..G-TAC..CAGACA...GCGTGA-GA-CGGGA    367
MAC.US.MM251        .--G.ACTCCTGAGTACGGCTGAGTGAA..G-C...A-TAAG.GG-GGC.A-GAA-CAA.CC--GAC-GAGTGCTCCT.ATAAAGGC-CGGGTC..G-TAC..CAGACG...GCGTGA-GAGCGGGA    949
SMM.US.SIVSMMH9     .--G..CTCCTGAGTACGGCTGAGTGAA..G-C...A-CAAG.GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCCT.ATAAAGGC-CGGGCC..G-TAC..CAGGCG...GCGTGA-GAGCGGGA    444
STM.US.STM          .--G.ACTCCTGAGTACGGCTGAGTGAA..G-C...A-TAAG.GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCCT.GTAAAGGC-CGGGC..........-AAGCG...GCGTGA-GAGCGGGA    620
TANTALUS.UG.TAN1    .CTTCACTGAAGGGGAGGAAGGGCAT-A..CC-...-TC-AG.T--C-C.G-----GGT.AAGTCG--ACGCGG-G--.GCGAAGAG-ATCC-CCAG--AA..AAAGTA-T--C-GG-AG-CAGGG.    852
GRIVET.ET.GRI-677   .CGGCACCGACCGCTGAGTTGCTGAG-G..TCG...-AGAGG.GA-GAC.TC-G-T-GG.-TGAGA-C--ACGA-TTT.TTT--TACCTAGTC--CGAGAA..A-GCTAG-C--CGACAG-G-CG--    665
VERVET.KE.VER-9063  .AGGCACGTACAGCTGAGAAGACGTCGG..-CG...C-TAGG.AA.C-G.--GG-TGCG.ACGTGACCAAGAAG-G...-TT-GTGA-TGGGCTTCTCGA-..T-CCGA.AAG.-AGCTC---CCTA    875
VERVET.KE.AGMTYO    .AGGCACGTACAGCTGAGAAGGCGTCGG..-CG...C--AGG.AA.G-G.--GG-TGCG.ACG-GACCAAGAAG--GA.-TT-GTGA-TAGGCTTCTCGA-..T-CCGG.-AAA-AGCTC---CCTA    364
VERVET.KE.AGM155    .AGGCACGTACAGCTGAGAAGACGTCGG..-CG...C--AGG.AA.C-G.--GG-TGCG.ACGTGACCGAGAAG-G...-T--GTGA-TAGGCTTCTCGA-..T-CCGG.-AAA-AGCTC---CCTA    862
VERVET.DE.AGM3      .AGGCACGTACAGCTGAGAAGACGTCGG..-CG...C--AGG.AAGG-G.--GG-TGCG.ACGTGACCAAGAAG-G...-TT-GTGA-TAGGCTTCTCGA-..T-CCGG.-AAA-AGCTC---CCTA    362
SABAEUS.SN.SAB1C    .-GCCGCAGAGTAGACCAGCCTGAGGA-..TCC...TTT-AC.TGAGTA.A---CGGAC.---AGAAAGGC-GACCGT.GC--G-GT-CAGGT-A-A-GC-..CCTGCGGAC-CGAGACACCC-CG-    918
SUN.GA.SIVSUN       .-TG-CGAAGAGACGTGGAGTGAAAC-T..TC-...-A--AA.G-GAGC.A-CT--TGA.-A-GAA--AAG-T-TCGA.-CT-GTGGAAGGC-CC-G--A-..TTAAGGTAAGTGAG--CACT---C    917
MANDRIL.GA.MNDGB1   .TG-CACTTGAGGCAGAGCACTCCGC-T..G--...A--AGC.AGG-TG.AAGGAG-GT.-----G-T--GAAGACGC.-AG-AGGT-AGTCAGT-G--CT..-ACT.TTACAAGAATTA-TTGTAC    419
LHOEST.KE.hoest     .-CGCTGGCAAGAACCTCTGGAGGGAGG..-A-...A-C-CG.-AGAAG.A-G-CAGGA.A-CTAG-A-AGAGAA-GT.-TGCTCGG----G--A-A--A-..ACCTGC.TTGAGAC-AG--TA-GA    937
LHOEST.CD.hoest447                                                                                                                                        0
LHOEST.CD.hoest485                                                                                                                                        0
LHOEST.CD.hoest524                                                                                                                                        0
RCM.NG.SIV-NGM      .--G-CCTCCGACGCGGGTCCGGA..--.---...A---TG.ACTGG-.A--ACGCGG.ACC-AGTC-AG-TAAG--.-AGC-TGGT-GCTCGTAAATCT..A-CGAGAAAA-TGA--GCTAGTA-   1000
COL.__.SIV-CGU1     .CTG-GGGTCGCGGAAGAAGAAGAGGAG..-CT...-ATCGC.---CTC.A--A-AGAG.-T-AGTC-AAAAAG-GGA.-TTA-GGGT-G-GTCAAG----..ACTCTA---TCTGC-CAA-AGTGA    410
SYKES.KE.SYK173     .--CCCGAGAGGAGGTGACGCCAGTCAA..--G...-AGTCC.GGA--C.AAGA-GGAT.C-G--GTC-G-C--AGGC.--AC-.GGT-.CT-GT-GCTC-TG-TGTACACA---GAC-A-GTA-GT    749
```

```
                    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)
B.FR.HXB2           CCA....AAAATTTTGACTAGCGGAGG.....CTAGAAGGAGAGAG..............................................................................   789

A.UG.U455           --..TAA--TT----------------.....-------------A..............................................................................   256
C.ET.ETH2220        --..AATTTTTA---------------.....--------------..............................................................................   169
D.CD.84ZR085        -T..GAA--TT----------------.....--------------..............................................................................   306
F1.BE.VI850         --GAATTTTTT----------------.....--------------..............................................................................   144
G.SE.SE6165         --....A---T----------------.....--------G-----..............................................................................   192
H.CF.90CF056        --...ATTTTG----------------.....--------------..............................................................................   136
J.SE.SE7887         --.AAAA-T-T----------------.....--------------..............................................................................   112
K.CM.MP535C         ............................................................................................................................     0
01_AE.TH.CM240      --......-------------------.....--------------..............................................................................   354
02_AG.NG.IBNG       --......-------------------.....--------------..............................................................................   329
03_AB.RU.KAL153-2   NNNNNNNNNNNNNNNNNNNNNNNNNNN.....NNNNNNNNNNNNNNNNNNGCGGAGGCTAGAA.GGAGAGAG....................................................   840
04_cpx.CY.94CY032-3 --....A--TT----------------.....--------------..............................................................................   155
N.CM.YBF30          -GC.AATTTTGAC-G-CGGT-GCC--A.....AAGT-G-AGAGAG-..............................................................................   347
O.CM.MVP5180        .........--A----.--G----T--.....-C---CCT--G-GAAGGGCGAAGTCCCTAGG.GGAGGAAG....................................................   816
O.BE.ANT70          ..........------.--G----T--.....-C---CCT--G-GAAGGGCGAAGTCTCTAGG.GGAGGAAG....................................................   840
CPZ.US.CPZUS        ...........-----.-C---A-T--.....--G--CCT--G-GAAGGGCGAAGTCTCTAGG.GGAAGAGG....................................................   831
CPZ.GA.CPZGAB       -A-T.......-----.-------T--.....-----CCT----GAAGGGCGAAGTCTCTAGG.GGGAGAAG....................................................   826
CPZ.CM.CAM3         ..........T-----.-C-----T--.....--G--CCT--G-GAAGGGCGAAGTCTCTAGG.GGAAGAGG....................................................   179
CPZ.CM.CAM5         ....AT.TT-T-----.-C-----T--.....--G--CCT----GAAGGGCGAAGTCTCTAGG.GGAAGAGG....................................................   485
CPZ.CD.CPZANT       TT.CATTTTG.-----.T-GCTA-T-A.....GC-ATCCT--G-GAAGGTCGAAGTCTCTAGGAACAGGAAG....................................................   153

H2A.GW.ALI          AGTGAAG-GGCC-CC-GG-G...A---.....TA--T-CCTACACCAAAAT.TGTAGCC.GAAAGGGCTTGTTATCCTACCTTTAGACAGGTAGAAGATTGTGGGAG.................  1095
H2A.SN.HIV2ST       GTGAAAG-GGCC-CC-GG-G...A---.....TA--TGCCTACACCAAATACAGTAGCCAGAA.GGGCTTGTTATCCTACCTTTAGACGGGTAGAAGATTGTGGGAG.................   547
H2A.DE.BEN          GTGAAAG-GGCC-CC-GG-G...A---.....TA--T-CCTACACCAAAAACTGTAGCCAGAAAAGGCTTGTTATCCTACCTTTAGACAGGTAGAAGATTGTGGGAG.................  1102
H2B.GH.D205         GG-GAAG-GG.C-CC-GGAGTAA..--.....TA--T-CTCC-A-.................TTTTGCATTGTCTACTCTAAAGAGGGAGTAGGGCATAAGTGGGAG.................  1078
H2B.CI.EHO          GG-GAAG-GG.C-CC-GGAGTAAA---.....TA--T-CTCC-A-TTT.................GCAATTCGTCTACTCTAGGACAGGGTACGGAATAAGTGGGAG.................  1079
H2G.CI.ABT96        GT.GAAG-GGCC-CC-GG-GCGCAG-......TA--TGCACC-A-AATTTGTAGCCAAAAAGGGCTGTGAATCCTACCTTGAATAGTCAGGTAGAAGT..GTGGGAG.................   465
MAC.US.MM251        GAGGAGG-GGCC-CC-G...TT-C---.....TA--TGCA-C-C-AAAAAGAAA........TAGCTGTCTTGTTATCCAGGAAGGGATAATAAGATAGAGTGGGAG.................  1040
SMM.US.SIVSMMH9     GTCGGAG-GGCC-CC-G...TT-C---.....TA--TGCA-C-A-A.........GAGTCATAGGACTGAGTTCCCTACTTTTGAGAAAAGAGTAGGAGAGTGGGAG.................   534
STM.US.STM          GGCGAAG-G-.C-CC-G...TT-C-A-.....TG--TGCA-C-AGTAGCCAGA..........GGGCTGATTGTTCTACTCTTGAGGGAAGAGGTAGAAGGTGGGAG.................   708
TANTALUS.UG.TAN1    .................G-G-..TGCA.....GGTAG-TCTAG-T-GGCGACAGATCAGCACCTGGGTGAGCA...................................................   901
GRIVET.ET.GRI-677   GGT.......CCCA-T-G-G--AACCA.....ACCC-GTTG--CGAAGGGTTGGTAGGGGACGGGTCGGAGCA...................................................   726
VERVET.KE.VER-9063  GTTAGAGG-CTAGGAAGGGCC.TAGCC.....A--ACTACTCT-G-CAAGTAGGGCAGGCGGAC.GGTACGCA...................................................   941
VERVET.KE.AGMTYO    GTTAGAGG-CTAGGA--GGCCGTAGCC.....G--ACTACTCT-G-CAAGTAGGGCAGGCGGT.GGGTACGCA...................................................   431
VERVET.KE.AGM155    GTTAGAGG-CTAGGAAGGGCCGTAGCC.....G--ACTACTCT-G-CAAGTAGGGCAGGCGGACGGGTACGTA...................................................   930
VERVET.DE.AGM3      GTTAGAGG-CTAGGAAGGGCCGTAGCC.....A--ACTACTCT-G-CAAGTAGGGCAGGCGGACGGGTACGCA...................................................   430
SABAEUS.SN.SAB1C    G-TGTGTGTGTG-GA--AG--A---AAAAGAGAAGTC--ACAG-GCTGACAAGGCTGGCTTAGCAGTGGCAGAAGGCTAGGTCGAGCGGGAAAGGAAA..........................  1016
SUN.GA.SIVSUN       AGGCACTGTC-GGGGACAGGC-A--A.....AG-CCT---CTG-G-GGAGGTCCCGCTAGGTTAGTACTGTCAAGTAGGCAGACTAGTAGTACGTGGAATAAAATAAGAAGTGTGTGTATAACT  1036
MANDRIL.GA.MNDGB1   --TAGTGT--GGGGCAG-ATAGTC--AGCA..............................................................................................   449
LHOEST.KE.hoest     AGGACTTG-CTGC-A--AGGCG-C--A.....AGGT--------TAAGACCTCACCGGGAGAGTGGCAGCAGGCCGGGGGCAGCCTTGTGAGAAGTGTGTAATTCTGATCTAGTGTAACTGGGA  1056
LHOEST.CD.hoest447                                                                                                                                   0
LHOEST.CD.hoest485                                                                                                                                   0
LHOEST.CD.hoest524                                                                                                                                   0
RCM.NG.SIV-NGM      AGCAAGGCG-CCCGGTCAGGTACTG--CCTCTAGG-----GTC--AGTCCTAGAGAAGGGAGGAAA..........................................................  1066
COL.__.SIV-CGU1     AAGTAAA-GTGCGCGCT-GTCGC-CA-TCAGGAAGTGGCAGAC---AGAGCACAGAGAGGGAGTGCTCGCCAGAGAAGGAGAAGAGCTCATTAGGAGGTCTTCAGGCAGGGAGTAGCGCCCCTG   534
SYKES.KE.SYK173     GATCCAGCGCCAGACTG-CCA-A-CA-TCGGGGA-AT---G--CC...............................................................................   794
```

```
                                                                                          |-> Gag and Gag-Pol start (p17)
B.FR.HXB2 Gag                                                                              M__G__A__R__A__S__V__.__L__S__G__G__E__L__D__R__W__E__K__I   p17
B.FR.HXB2           .....................................................................ATGGGTGCGAGAGCGTCAGTA...TTAAGCGGGGGAGAATTAGATCGATGGGAAAAAAT   845

A.UG.U455           .....................................................................---------------------...---------AA-A--------TC------G-----   312
C.ET.ETH2220        .....................................................................------------------A--...-----A--C-A-A--------GCC-----------   225
D.CD.84ZR085        .....................................................................---------------------...------------A----G---GCT-----------   362
F1.BE.VI850         .....................................................................------------------A--...------------A--------GA------------   200
G.SE.SE6165         .....................................................................---------------------...----C-------A--------GCT-----------   248
H.CF.90CF056        .....................................................................---------------------...--------C---A--------GCT-----G-----   192
J.SE.SE7887         .....................................................................------------------A--...-----T------A--------GAT-----------   168
K.CM.MP535C         .....................................................................---------------------...------------A--------GC---------G--    56
01_AE.TH.CM240      .....................................................................---------------------...-----T------A--------GC------------   410
02_AG.NG.IBNG       .....................................................................---------------------...-----T------A--------GC------G-----   385
03_AB.RU.KAL153-2   .....................................................................---------------------...------------A--------GC------------   896
04_cpx.CY.94CY032-3 .....................................................................---------------------...-----T------A--------GC------G-GG--   211
N.CM.YBF30          .....................................................................--------------------G...----CA------A---------A-------TC---   403
O.CM.MVP5180        .....................................................................-----------------T--G...--G-CA---A-TA----G---GC-------CG---   872
O.BE.ANT70          .....................................................................-----------T-----T--G...--G-CA--AA-CA----G---GC-------C----   896
CPZ.US.CPZUS        .....................................................................--------------------T...C-G-CA------CG-------GCT--------G--   887
CPZ.GA.CPZGAB       .....................................................................--------------------T...C---CA------A----------C---------G-   882
CPZ.CM.CAM3         .....................................................................---------------------...----CA------CG-------GCT-----------   235
CPZ.CM.CAM5         .....................................................................---------------------...----CG--A---CG---G---GCT-----G-----   541
CPZ.CD.CPZANT       .....................................................................-----A---G-G-----T--T...--G--G--A-AGA-GC-----AC--------GT--   209

H2A.GW.ALI Gag                                                                             M__G__A__R__N__S__V__.__L__R__G__R__K__A__D__E__L__E__R__I   p17
H2A.GW.ALI          .....................................................................-----C------AAC--C--C...--G--A---A--A--GC---CGA--TA----G---  1151

H2A.SN.HIV2ST       .....................................................................-----C------AAC--C--C...--G--A---AA-A--GC---CGA--TA-----G--   603
H2A.DE.BEN          .....................................................................-----C------AAC--C--C...--G--A---AA-A--GC---CGA--TA------G-  1158
H2B.GH.D205         .....................................................................-----C-------GC--C--C...C-GTCA---AAGA--AC----GA--TA------G-  1134
H2B.CI.EHO          .....................................................................-----C-------GC--C--C...--GTCA---AAGA-GAC----GA--TA-----GG-  1135
H2G.CI.ABT96        .....................................................................-----C------AGC--C--C...C-GTCA---AA-A--GCT---GA--TA------G-   521
MAC.US.MM251        .....................................................................-----C------AAC--C--C...--GTCA---AAGA--GC----GA--TA--------  1096
SMM.US.SIVSMMH9     .....................................................................-------T----AAC--C--C...--GTCA---AAGA--GC----GA--TA--------   590
STM.US.STM          .....................................................................-----C------AGC--C--C...--GTCA---AA-A--GC----GA--TA------G-   764
TANTALUS.UG.TAN1    .....................................................................-----A---G--CAC--G-C-...C-GTCA---A-GA-T--G--CAC--TT--G-----   957
GRIVET.ET.GRI-677   .....................................................................-----C-G-G-TCAC----C-...C-GTCA---A--AGCC-C--CACG-TC--G--G--   782
VERVET.KE.VER-9063  .....................................................................-----G---GCTA-C----C-...C---AA---A--C-------C-A--TC--G-----   997
VERVET.KE.AGMTYO    .....................................................................-----G---GCTA-C----C-...C---ATA--A--C-------C-A--TT--G-----   487
VERVET.KE.AGM155    .....................................................................-----G---GCTA-C----C-...C-G-ATA--A--C--------GA--TT--GC-T--   986
VERVET.DE.AGM3      .....................................................................-----G---GCTA-C----C-...C---ATA--A--C-------CAA--TT--GC-T--   486
SABAEUS.SN.SAB1C    .....................................................................-----------TAAC-----C...-----T--TA-GA--C-----GC--TC---TC-G-  1072
SUN.GA.SIVSUN       CTGTGTCTCGGAAGAGTGTCTCAAAC...........................................-----A---G--AAC--CAC-...G-GGATA---A--TTG-GAGGA-C-TT----G-G-  1118
MANDRIL.GA.MNDGB1   .....................................................................------AATG-GAAC--T-CC...--GTTA---ACT--T--G---AA--TT--G-----   505
LHOEST.KE.hoest     CACTAGTGGTAGCTTCACT..................................................-----AT--G-TAAC--T--C...C-C--TA--CA-ATTGA-A-AGAT-TTTGT-GTG-  1131
LHOEST.CD.hoest447  .....................................................................-----AT--G-GAAC--T--C...C----TA--CA-ATTGAG-GAGA--TCTGC-GT--    56
LHOEST.CD.hoest485  .....................................................................-----AT--G-GAAC--T--C...C----TA--CA-ATTGAG-GAGA--TCTGC-GT--    56
LHOEST.CD.hoest524  .....................................................................-----AT--G-GAAC--C--C...C-C--TA--CA-ATTGAGA-AGAT-TTTGT--TG-    56
RCM.NG.SIV-NGM      .....................................................................-----------G------A-T...--GTCA---AAGA-G------GCT------C-GG-  1122
COL.__.SIV-CGU1     GGCAGGTCGAGGGGGGCCTGCAAGTGGAGTAAGCTGCGGCGCTCCAGGACCGGTAGAGGGAGGTCATTA-----CAACGAGCA-GG-C-G...---G-GAA-AA-AC------AGACCTAC-G---G-   659

SYKES.KE.SYK173     .....................................................................-----A--AGCG-G---G-C-ATA----CA--TA----GC----CA---AT--G-----   853
SYKES.KE.SYK173 Gag                                                                        M__G__A__A__G__S__A__I__L__T__G__R__E__L__D__R__Y__E__K__I   p17
```

```
B.FR.HXB2 Gag       __R__L__R__P__G__G__K__K__K__Y__K__L__K__H__I__V__W__A__S__R__E__L__E__R__F__A__V__N__P__G__L__L__E__T__S__E__G__C__R__Q__I__L__G_   p17
B.FR.HXB2           TCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGA   975

A.UG.U455           ---------------------C----------G-C-G------T----------------------G---AA----A--C----C--------T---------G-------A---CAG------------   442
C.ET.ETH2220        -AA----------------------C-C----TGC-G-----CC----C-------A---------G---AA---T---C----C----A---T-----T-------C--------A-------A-TAA-   355
D.CD.84ZR085        -------C-------------------G-----GC-------------------------A-----G--------T---C-----------T------------------------A-------A---A-   492
F1.BE.VI850         --A---------G-------------G-------A-G------C--A----------------------------T---C--G----------TC--------------------CA-A-----A-AA--   330
G.SE.SE6165         ------G-------------G----TC-------A--------C----------------------G--GA----T---C----C----A---T---------G-------T---CA-------A--A--   378
H.CF.90CF056        ----C---------------------------GGC--------C----------------------G---A----T---C----C--C-----T---------C-----------CT---G---A-A-A-   322
J.SE.SE7887         ------G-----------G-------------GGA----G---C----------------------G--CA----T---C----C--------TC----GT--G--A--------CA---------A-T-   298
K.CM.MP535C         -------C--------------------------C-G------C-------------------------------T---C-C--C--------T-----G---A-----------C-G------A-AAC-   186
01_AE.TH.CM240      ------GC------------GA----------GGC-G------T----------------A---T-----A--------C----C---A--T-T---------G-------A---CA-------A-A-A-   540
02_AG.NG.IBNG       --------------------------------G-C--------C----------------------G---A--------C----C--------T---------G----G--A---CA----C--A---A-   515
03_AB.RU.KAL153-2   ----------------------G---------G-A-------CC----------------------G---A-------GC----C---A----T-----------------A---CA-----------A-  1026
04_cpx.CY.94CY032-3 --------------------------------G-C-G------C----------------A---T-G---A--------C----C--------T---------G-------A---CA----T--A---A-   341
N.CM.YBF30          -TAT--G--A--------------------C-G-A-G------T----------------------G---A-------TTG---C--A--T--CA-G--C---G-G--C------GCCA-GT----AAAT   533
O.CM.MVP5180        -A-------------ATCT--A--GGC-----GGC--------T----------------------G---A---A----TG---------T--A--------TG-------TAC-GAG---C-G--ACAG  1002
O.BE.ANT70          -A------A------ATCT--A--G-------G-C-------CT----------------A-----G---A--------TG--------AG--AC-------TG----G--TAA-GAGA-GC-G--ACAG  1026
CPZ.US.CPZUS        ----C-T--A--T--------------G----TGA-G------T----T-----------------G---A----T--TTG---C--A--T--AA-G------G----C------CTC---T--T-AAA-  1017
CPZ.GA.CPZGAB       ----C-T-----C-------GA----G-----TGA-G------T----------------------G---A--------TG-G-C--C--G--AA-G----GTAAG-----A----CTA--T-GT-ACA-  1012
CPZ.CM.CAM3         -A--C---------------------------TGA-G------T----------------A-----G--CA-------TTG---C-----T--AA-G-----GG-------A---GAG--GT--T-AA-G   365
CPZ.CM.CAM5         -A--C----M-----A----------------TGA-G------T----------------------G---A----T--TTG---C-----T--AA-G--G---G-T-----A---GA-A-TC----AAA-   671
CPZ.CD.CPZANT       CA--C-TC----C--T--C--------G--C-TGA--------C-G--T--------ATC------GC-G--T--T--GC-C-GCT-CTC---TC----------------T---GA-A-GGCTA-CCAT   339

H2A.GW.ALI  Gag     __R__L__R__P__G__G__K__K__K__Y__Q__L__K__H__I__V__W__A__A__N__E__L__D__R__F__G__L__A__E__S__L__L__E__S__K__E__G__C__Q__R__I__L__K_   p17
H2A.GW.ALI          -A-----C----C--C---------------C-GC----------T--G------GCG-AT--AT-G--CA------G-T-GGCAGAAA-------G--GT--AA------T--CCA-AG---T--TAA-  1281

H2A.SN.HIV2ST       -A-----C----C--C----------------GGC----------T--G------GCG-AT--AT-G--CA------G-T-GGCAGAGA-------G--GT--AA---G--T--CCA-A----T--TAC-   733
H2A.DE.BEN          -A-----C----C--C-----------G--C-GG-----------T--G------GCG-AT--AT-G--TAA-----G-T-GGCAGAGA-------G--GT--AA------T--CCA-A-G--T--CA--  1288
H2B.GH.D205         -A-----C----C--C-----------G----TGC--------G-----------GTG-AC--AT----CA----T-G-T-AGCAGAGA-------G---T--AA------A---CA-A----T--TAA-  1264
H2B.CI.EHO          -A-----C----C--C----------G---C-TGC--------------------GTG-AT--AT----GA----T-G-T-GGCAGAGA---G---G-G-T--AA------A--CC-GA----T-G-AA-  1265
H2G.CI.ABT96        -A-----C----C--C--------------C-TG---------G-T--G------GCG-AC--AT-G--CA------G-T-AGCAGAGA-------G---T--AA------T--CCA-A----T--TTC-   651
MAC.US.MM251        -A--C--C-A--C--C-----------G--C-TG--G--G---G-----------GCA-AT--AT----TA----T-G-T-AGCAGAAA-------G--G-ACAA------A---CA-A-------TTCG  1226
SMM.US.SIVSMMH9     -A-----C----C--C-----------G--C-TGC-G--G---------------GCA-AT--AT-G--CA----T-G-T-AGCAGAAA-------G--G-ACAA------T---CA-A----T--TTC-   720
STM.US.STM          -A-----C----C--C--G--------G--C-TG--G------G----G------GCA-AC--AT-G--CA----T-G-T-GGCAGAAA-------G--GT--AA------G---CA-A----TA-AACG   894
TANTALUS.UG.TAN1    A--C---C-C--GAAC-----------G---C----------CT--A--------G---AAA-AA-G--G--C--T-GGC-CC-CGAGAAA--C---------AA---G------CA-A-G--CA-A-A-  1087
GRIVET.ET.GRI-677   -A--C--C-T--GAAC--G--A--G--G--CC--A-T------T--A--------G-A-AA--AA-G--------T-GGT-AC--GAGAAA--T---------AA----------CA-A----CA-A-A-   912
VERVET.KE.VER-9063  A--A---C-C--GAAC---------------C--A-T--G---T--A--------G-T-AA---A-G--G--C----GCC-CC--GA-A-GT-----------GA---T--A----A-A-G--CA-A-A-  1127
VERVET.KE.AGMTYO    A--AC-TC-C--GAAC-----------G--CC--A-T------T--A--------G---A----A-G--G--C----GCC-CC--GAGA-GT----G--G---GAG--G--G----A-AG---CA-A-A-   617
VERVET.KE.AGM155    A--AC-TC-C--GAAC-----------G---C--A-T------T--A--------G---A-A--A-G--C--C----GCC-CC--GAGAAGT-A--G--G---GAG-----T----A-A-G--CA-A-A-  1116
VERVET.DE.AGM3      A--AC-TC-C--GACC-----------G--CC--A-T------T--A--------G---A---AA-G--G--C----GCC-CC--GAGA-AT-AC-----T--GA------A----AGA-G--CA-A-A-   616
SABAEUS.SN.SAB1C    CA-----C----CAAC--------G--G-----G----G----C---------------A---A-----T-----TAGCC--TCAG-AAAT--C-----G---AAG-----GGTCGTCA----T--TA-T  1202
SUN.GA.SIVSUN       AGCTC-G-AA--T--A--G--A--G-CC--CC-GGC--G----G-----------G-T-AA--AT----TA-G--T-G-C----CAAG-AA------AG----GTG---------CAGA-G--T---TC-  1248
MANDRIL.GA.MNDGB1   AA-A----A-AG---T--T--A---TGT----G---G-----CC-CTGT---TGT-AAG-T--AT----TA----T-GCT-ATCGGA-AAA--CC-T------CAGC----A---GA-A----C--CTC-   635
LHOEST.KE.hoest     CA--C----A--T---A-T--A--G-CT---C--AA--G----G---A--------CG-AA--AT----CA----T-G-T-GGGAT-ACAA--C--G------G-------G----AGA----C--ATCT  1261
LHOEST.CD.hoest447  CA--C----A--T-ACA-T--A--G-CT---C-GAAG-G----T----G-------CA--A--AT----TA-G--T-G-T-AGGAG-ACA-T--C----G---G----T-----C-AGA----T--T---   186
LHOEST.CD.hoest485  CA--C----A--T--CA-T--A--G-CG---C-GAA--G------T--G-------CA-AA--AT----TA-G--T-GTT-AGGAG-ACA-T--C-G--G---G----T-----C-AGAG---T--T---   186
LHOEST.CD.hoest524  CA--C----A--T---A-T--A--G-CT---C-GAA--G----G---AC-------CG-AA--A-----TA-G--T-GGC-AGGAG-AAA---TC--------G----G-------AGA-G--TT-----   186
RCM.NG.SIV-NGM      -A-----C----C--C-----------G----TG-----G---C-----------T--------A----GA----T-GCT-G-G-GA-ACTT---------GTG-------AATCCAGA----TA-A--G  1252
COL.__.SIV-CGU1     C---C-G-A-AA----AAG--AGG-TGT-----GA--------G-TAG----ATGT---CA---G--AGC--C-GT-TTT-A-TATT--AG------A-GT--G--AC---AGTGGCT------T---A-   789

SYKES.KE.SYK173     ----C-CC-A--GAAA--G-----GCGC--CCT-G-TCG----T----T-------AG-A----T-G--CA------GCC--TCGGACCAA--CA-G--GT-TAA----------GA-A----T--CAC-   983
SYKES.KE.SYK173 Gag __R__L__R__P__K__G__K__K__R__Y__L__V__R__H__L__V__W__A__K__K__E__L__D__R__F__G__L__S__D__Q__L__M__E__S__K__E__G__C__E__K__I__L__T_   p17
```

```
B.FR.HXB2 Gag       _Q__L__Q__P__S__L__Q__T__G__S__E__E__L__R__S__L__Y__N__T__V__A__T__L__Y__C__V__H__Q__R__I__E__I__K__D__T__K__E__A__L__D__K__I__E__    p17
B.FR.HXB2           CAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGG   1105

A.UG.U455           --AT--------G-T--C---------A-----------------------------------GT-----------A-------------TG----------------------A-T--A-------    572
C.ET.ETH2220        ------------G-T------------A----G------A--------T---------G-----T-----------A-------A-------------------------C----------------    485
D.CD.84ZR085        ------------G-T---A-------------------------A-------------------------------A---A-------------------------------------------------    622
F1.BE.VI850         -----------------------------------G----AG-------T-------------GT---------A---A---------GC--G-G---C-----------------------C------    460
G.SE.SE6165         ---T-G--------T--C---------A----G---A---A-----------G----------------------A-----------------G-------------A------C----AG-AG-G--AA    508
H.CF.90CF056        ---A----G---G-TA--A--------A------------A--------T----CT-------GT---------C--A----G--AA-----TG-G----------------G--------T----------    452
J.SE.SE7887         -----C------G-T--C---------A--C-----A---A-----G--------------------------C--A-----G--------A-------------T-----------G----------    428
K.CM.MP535C         --AA----G------A----A--------------GA---A----C-----------A-----GT--------T---A-------A-------G------------------C---------AC------    316
01_AE.TH.CM240      ---T----GT--A-T--CA---------T-----------A--------T-------------------GG--C--A--C------------G-------------------------T--A-------    670
02_AG.NG.IBNG       ---T-----T--G-T--CAG--------------------A--------T-------A----------T-GG--C-----------------C--------------------C-----T--A--G----   645
03_AB.RU.KAL153-2   ---T--------A-T--CA---------------------A-------------------------------A------C---------------------------------------T--A-------   1156
04_cpx.CY.94CY032-3 ---T-----T--A-T--CA-A-------------------------------TA--A---------GG--C--A--------A-----TG--C------------------------T--A-------   471
N.CM.YBF30          --AT--G-----G-T--CA-------G-----------GC-C--T--------CG-TC-----GTT--T--------C---AGT------C-----C-CA---AC-----------G---------A-A-   663
O.CM.MVP5180        ---T--G-G---G-T--CA-------G-----G--C--G-A---TC-C-GG--CG--A-----GTA----GG--C--T--CA-C--AT-T--C--CCG---T--AC--C-G--AA--C-A---T--A---   1132
O.BE.ANT70          ---T--G-G---G-T--CA-------G-----CAGC--GCAG---C-C-GG--CG--A---T-GTG----GG-----T--CA-C--ATATA-A--TGG---T--GC--C-G--AA--C-A---T--A---   1156
CPZ.US.CPZUS        ---T--G-----G-T--AA-A--------G----GC--GC----GC-C-T---C--CT----GGT---T-GG-----C--CAGT---G-GACAG-GG-G--T--GC--C-G--C----TG--AC-GA-A-   1147
CPZ.GA.CPZGAB       --AT--G-G---G-T--CA-A-----C--------G---GC-G--C--G-T---C--TC-G---GTA--G-GG--CA-A---AGTGAC--CACTG--G-------AC--A-----C----AC--C--A--.   1141
CPZ.CM.CAM3         --AT--G-----G-T--CA------------------GC--GC----GC-T-T------TC-----GTG--T-GG-----T---A-G--A--TTCAG--G-G-----AC--C-G------ACA--AT-GA-A-   495
CPZ.CM.CAM5         ---T--G-----G-T--CAG-------------G-GCT-GC----GC---T---C--TC----GGTA--T-GG-----T---A-G--A--CC-AG--G-G-----AC--C------C--TCA--AC-GA-A-   801
CPZ.CD.CPZANT       --AT-GAGC--T---A-AG-A-T-A----CCCT---A-A-T---T--G-T---C--CA-TTGTGTT--G-GG--C--A---A--G--GA-A-------------AG-AC----CG-TA-A-CAG-GA-AA   469

H2A.GW.ALI Gag      _V__L__E__P__L__V__P__T__G__S__E__N__L__K__S__L__F__N__T__V__C__V__V__W__C__V__H__A__E__E__K__V__K__D__T__E__G__A__K__.__.__.__.__    p17
H2A.GW.ALI          GTTT--G------TAG-G-CA------------A-TT-A-A-AGCC-T-T------T--CTGCGTAG-T-GG--C-----CGC-GAAGAGA-AG-G-----T--TG-A-G---AAA-.............   1398

H2A.SN.HIV2ST       GTTT--G-T----TAG-A-C------G------A-TT-A-A-AGCC-T-T------T--CTGCGT-A-T-GG---A-A--CGC-GAAGAGA-AGCG-----T--TG-A-----AAA-.............    850
H2A.DE.BEN          GTTT--G-T----TAG-A-CA-----G------A-TT-A-A-AGCC-T-T------C--CTGCGT-A-T-GG--CT----CGC-GAAGAGA-AG-G-----T--TG-------AAAG.............   1405
H2B.GH.D205         GTCT--GCT----TAG-A-CA-----G------A-TT-A-A-AGCC-T-T-----TC--CTGCGT-A-T-T---CC----CGC-GAAGAGA-AG-G-----T--AG-------AAA-.............   1381
H2B.CI.EHO          GTCT--GG-----TAG-A-CA-----T------A-TT-A-A-AGTC-T--------C--CTGCGT-A---T---CC----CGC-GAAGAGA-AG-G-----T--TG-------AAA-.............   1382
H2G.CI.ABT96        GTAT--GCT----TAG-A-CA-----T------A-TT-A-AGAGCC-T--------T--CTGCGT-A-T-GG---T----CGCCGAACAGA-AG-----C-T--TG-A-----AAA-.............    768
MAC.US.MM251        GTCT--GCT----TAG-G-CA-----C------A-TT-A-A-AGCC-T--------T--CTGCGT-A---GG--CA-T--CGC-GAAGAGA-AG-G---C----TG-------AAA-.............   1343
SMM.US.SIVSMMH9     GTTT--GCT----TAG---C------T------A-TT-A-AGAGCC-T-----C--T--CCGCGT---T-GG--CA-T--CGC-GAAGAGA-AG-G---C-T--TG-A-----AAA-.............    837
STM.US.STM          GTCT--G------TAG-G-CA-----T------A-TT-A-AGAGTC-T-T------T--CTGCGT-A-T-GG---A-C--CGC-GAAGAGA-AG-G---C-T--TG-------AAAG.............   1011
TANTALUS.UG.TAN1    GTAT-GAGC--GCTAGAA-CA-------------G---A-A-AGTC-G-T----TTGTGCTGCGT-A-T-GG---A-T--CGC-GAACAGA-AG-G-----T--AG-------AG-G-TATTAG--A-AC   1217
GRIVET.ET.GRI-677   GTTT--ACC--G-TGGAA-C------C--C--G-GG--A-A-G-TC-G-T----TTGTGCTGCGT-A-T-GG--CA-T--CGC-GAACAGA-AG-G--------AG--------G---TA-CAG-TA--C   1042
VERVET.KE.VER-9063  GTC--TTTT--T-TAGAA-CA-----C--G--G-GCT-A-A-AGTC-G-TC---CTT--GTGCGTG--A-----CA-C--C-GCGAACAGA-AG-G--------AG-------GG---CA-CAG--AGAC   1257
VERVET.KE.AGMTYO    GTC--CT-C--CCTAGAA-CA--------G--G-GCT-A-A-AGTC-G-TC---CTT--GTGCGTA--A-----CT----CA-GGAACAGA-AG-G--------AG-------AG---CA-CAG--AGAC    747
VERVET.KE.AGM155    GTT--CTCT--CCTAGAA-CA-----G--G----G-A-G-A-AGTC-G------CTG--GTGCGTAT-GCT---C--C--C---GAA-AGA-AG-G--------AG------------CA-TAG--AGAC   1246
VERVET.DE.AGM3      GTA--CT-C--GCTAGAA-CA-----G--G--G-GCT-A-A-AGTC-G-T---CCTT--GTGCGTAT-G-T---C--A--CA--GAT-AG--AG-G--------AG-A-----AG---CA-TAG--AGAC    746
SABAEUS.SN.SAB1C    GTA--C-TT--CCTAG---CT-----G-----GA-TT-A-TTG-TC-C-T----TT-TGCTGCGTGT-AGC----A-T--CGC-GAA---A-AG-G-----T--AG-------AAA--CG--AG--A-A-   1332
SUN.GA.SIVSUN       GTCTGTTGG--T-TGTA-GCAT-T--T--TA-GA-C--G-AGG-C---GTAGG-------TGTGTG--TGCC---TGT------G-C--TCCAG----------AC----G---C--A-A---G-CAGAC   1378
MANDRIL.GA.MNDGB1   GTATGTTGG---TATA-G-CCA---------TA-T--A-A-G-T--GGTAGGG-----CTGTGTTG-AGCC--CA-A--CGC-G-T-----A--T--GAG---AC-A--T------A-A--AT--A-A-    765
LHOEST.KE.hoest     GTCTGCTGG--T-TATA-GCA--T--T--TA-GA-C--A-AGG-C---GTAGGG------TGTGT-A---GC--CTGT--C-T-G-AG--AGA---TC------TC-------CA-CAGT--AG--A-AA   1391
LHOEST.CD.hoest447  GTCTGTTGG---CTATACTCC-----C--GA-GA-TT-A-A-G--C-TGTAGGA--T---TGTGTG--G-G---CTGC---TT-G-----A-AG--GC-------C-A------A-GC-A------A-AA    316
LHOEST.CD.hoest485  GTCTGTTGG---CTATACTCC-----C---A-GA-TT-A-A-G--C-TGTAGGA--T---TGTGTG--G-G---CTGC---TT-G-----A-A---GC-A-----GCA------A-GC-A------A-AA    316
LHOEST.CD.hoest524  GTC-GTTGG--T-TATA--CT-----C--TA--A-T--G-A-G-TC--GTAGGG-----GTGTGT-A---G---CTGT--C-T-G-A---AGAG-C--T--T--AC-------AA-TA-T--AG-GA-AA    316
RCM.NG.SIV-NGM      GTTA-T-TG--TCTAG---C------G--T--G-GCT-A-A---CC-C-T----TTGTGCTGCGTA--T-GG--C-----CA-GGAAG-GA-AG-G-----T--AG-------GG-G-CGC-TG--A-AA   1382
COL.__.SIV-CGU1     A--G-GACG----TAG-GG-C-------------TGT-G--G--CC-----GGA...TGTT-TGTGTGT--C--CC----CAG--AATGGA-T--TG-G-----TC-A--G--AGA-A-G--AG----A-    916

SYKES.KE.SYK173     GTC----TT----TAGAAGCT-AT--G------A-TT-A-A-AGCC-T-T-GGC-T-A-TT-GGT-G-A-GGGCC--T---GCC-AA-AG--AG--G-------TG--C----AAA-C-G--AG--A-A-   1113
SYKES.KE.SYK173 Gag _V__L__L__P__L__E__A__N__G__S__E__N__L__K__S__L__F__G__I__I__S__V__V__W__A__V__H__A__K__K__E__V__E__D__T__E__Q__A__K__Q__K__V__K__    p17
```

```
B.FR.HXB2 Gag       E__E__Q__.__.__.__.__N__K__S__K__K__K__A__Q__Q__A__.__.__.__.__A__A__D__.__.__.__.__.__.__.__T__G__.__.__.__.__.__.__.__.__.__.__.__.    p17
B.FR.HXB2           AAGAGCAA............AACAAAAGTAAGAAAAAAGCACAGCAAGCA............GCAGCTGAC.....................ACAGGA................................   1158

A.UG.U455           --AT----............--T--G-AC---C---GGA-------G---............------A--.....................------................................    625
C.ET.ETH2220        ----A---............---G-----C--C-----A-------G---............-G---A-CT.....................GACA--................................    538
D.CD.84ZR085        ----A---............-----C----------GG------------............-------G-.....................------................................    675
F1.BE.VI850         ----A---............---------C--C----GA----------G............---------.....................-A---G................................    513
G.SE.SE6165         -GATA---............--G------C--G-----AT------G---............---ATG--T.....................-A----................................    561
H.CF.90CF056        --ATA---............---------C--C-----A-----------............--------T.....................-AG-A-................................    505
J.SE.SE7887         --ATT---............-----G-AC--AC-GC-G------A-----............-A-A-----.....................-A-AA-................................    481
K.CM.MP535C         ----A---............---------C--CG---GA----A---................................................-A-GCAGCT..........................    360
01_AE.TH.CM240      ---TA---............--T--G--CC--CG---GA-------G---............-------G-.....................------................................    723
02_AG.NG.IBNG       --ATA---............--T--G--C---C----G-T------GA--............-------C-.....................------................................    698
03_AB.RU.KAL153-2   --ATA---............--T--G--C---C----GA-C--A--G---............---A---G-.....................------................................   1209
04_cpx.CY.94CY032-3 --ATA---............-GT--G--C---C----GA-------G---............-------C-.....................G-----................................    524
N.CM.YBF30          -GA-A--G...GAACAGCAC--GCCCGAGCCA-----CC--G-AGC--GG............-----G-CA.....................G--ACTGAT.............................    728
O.CM.MVP5180        ---TAATG............GCA-GC--G---TCTGC--AGGCCGCTAAG............-A--AAACA.....................-GC...................................   1182
O.BE.ANT70          ---TAATG............GGG-GC--G---TCTGCG-ACGCCGCTAAG............-A--ACACA.....................-GC-C-................................   1209
CPZ.US.CPZUS        -G-TA...............GTGC-GTTG......C-GAA---A--G-A-............-A-AAA--ACAACAGCAACAG.........GAG-C-AGT...............GGC...........   1209
CPZ.GA.CPZGAB       .......................CGGCA-C-TGG-G--CA----AGCAA-ACTGAAAGTAACT---GAAG-CGTGAAGGGGGAGCCAGTCAAGGC-CTAGTGCC............TCT...........   1225
CPZ.CM.CAM3         ---CAGTT............-CGGCCTCCCGAG--C-G-A-GT-GCCCA-............C-GCAGC-G.....................CAGCAGCAA.............................    551
CPZ.CM.CAM5         --ACAGTT............GC-GCCCAAG-A...C-G-A-GT-GCCCA-............C-GCAGC-A.....................CAGCAGCAG...............GAA...........    857
CPZ.CD.CPZANT       TGA-AGT-............-TGC-G-CAC-AGC-G--A--GGAAGTAGCCAAACCGCAAGCAG--GCATGCTTCTGCGGCTGCTCCTGTTA-ACAA-CAG......TGGTGTCAGCGA...........    570

H2A.GW.ALI Gag      .__.__.__.__.__.__.__Q__I__I__Q__R__H__L__A__A__E__.__.__.__.__I__E__T__.__.__.__.__.__.__.__A__E__K__M__P__S__T__S__R__P__T__A__.    p17
H2A.GW.ALI          ....................C-A-T--TAC---G-C-TCT-GC-GC--A-............AT--AAACA.....................G---AGAAAATGCCAAGCACAAGTAGACCAACAGCA..   1473

H2A.SN.HIV2ST       ....................C-A--GGTAC---G-C-TCT-GT-GC--A-............A--AAAACT.....................----A-AAAATGCCAAGTACAAGTAGACCAACAGCA..    925
H2A.DE.BEN          ....................--ACT-GCAC---G-C-TCT-GT-GC--A-............A-T-GAACT.....................G---AGAAAATGCCAAATACAAGTAGACCAACAGCA..   1480
H2B.GH.D205         ....................--G-T-GCAC---G-C-TCT-GC-GCG-AC............A---AAA-A.....................-TGCC-GCTACAAATAAACCAACAGCA...........   1447
H2B.CI.EHO          ....................--G-T-GCAC--CG-C-TCT-GCAGCG-AC............A---AAA-A.....................-TGCC-GCTATGAGTAAACCAAGTAAA...........   1448
H2G.CI.ABT96        ....................C-AGT-GTGC---G-C-TCT-GT-GT--A-............A-T-GAACT.....................----A-AAAGTGCCTGCCACAAGCAGACCAATAGCA..    843
MAC.US.MM251        ....................C-G-T-GTGC---G-C-CCT-GT-GTG-A-............A---GAACA.....................G---A-ACTATGCCAAAAACAAGTAGACCAACAGCA..   1418
SMM.US.SIVSMMH9     ....................C-A-T-GTGC---G-C-TCT-GT-GTG-A-............A---GAACA.....................G---ACAAAATGCCAGCAACAAGTAGACCAACAGCA..    912
STM.US.STM          ....................C-AGTGGTA----G-C-TCTTGT-GT--AG............A-T-GAACT.....................G--AACAAAATGCCTGCCACAAGTAGACCAACAGCA..   1086
TANTALUS.UG.TAN1    --CGCTGCCATCTAGTGGAA--AG---AA-CTGC-GC---G-CATCT-GT............-GCCAGC-G.....................CA-AATTACAACACAGCTGCG.................   1297
GRIVET.ET.GRI-677   --C-CT-CCATCTAGTGGAC--A--TGAG--AGC-GCTAA-A--A--AAT............-AGA-AACA.....................G-G...................................   1104
VERVET.KE.VER-9063  --TGCTGCCATCTAGTGGAA--GG---AA-GAGC--C--AG-CATCTAGT............-G--AAA-G.....................-A-A-TAACAGGGAGACAACAGCG..............   1340
VERVET.KE.AGMTYO    --C-CTGCCATCTAGTGGAA--AG---AA-GTGC--C--AGACATCTAGT............-G-CAAA-G.....................-A-AATGACAAGGGAATAGCAGCG..............    830
VERVET.KE.AGM155    --TGCTGCCACCTAGTGGAC--AG---AA-CTGC-GTTA-G-CA-CT-GT............-G-CAGC-G.....................-A-AATAACACAGGAGGAACAGCG..............   1329
VERVET.DE.AGM3      --TGCTGCCATCTAGTGGAG--AG----A--TGC-G--AG-A-TAC-A--............-AGA-ATCT.....................-GT---CAAAAGAAAAATGACAAGGGAGTAACAGTG..    841
SABAEUS.SN.SAB1C    -G--AGTGCCAGCAGAAATG-CAG-----GCC-C-GCGA--TCTAGT-GC............CA-A--A-G.....................GA-CTGCAGGCAAAAAAGAAAAATGAGCCAACAGTG..   1427
SUN.GA.SIVSUN       T-----C-GCAAAAAATGAAG-A...........................................................................................................   1401
MANDRIL.GA.MNDGB1   TCATAAC-AGAAAGGAAGAA--GC-GGAGG-TG--...............................................................................................    800
LHOEST.KE.hoest     T---A-C-GCACCAGAAGCAGCAGGG........................................................................................................   1417
LHOEST.CD.hoest447  T---A-C-ACTCAGGAAGAA-GGC-G........................................................................................................    342
LHOEST.CD.hoest485  T---A-C-ACTCAGGAAGAC-GGC-G........................................................................................................    342
LHOEST.CD.hoest524  TT-----GAACAAACCTGTG..............................................................................................................    336
RCM.NG.SIV-NGM      -GC--TGCCATCTAGTGGAC--AG-TGAG--TGC-GG--A---AG--AA-............-G---CACA.....................GTGAC-TCTAGTGGACAAAGAGGA..............   1465
COL.__.SIV-CGU1     ---CTT-TAAAAAGCAGGCC-TG-T-GAA-T-GCC-GCAAGG--G-G-AG............AA---AA-A.....................-AG-A-GCAGAGAAATTGGACATGCCAATAGTAACAGG   1013

SYKES.KE.SYK173     ---CATGTAATTGGAAAGATG--CC-CCAGC--C-TCT-GTGGA---AGT............-A-AA-AG-.....................-GTCA-AACATGGCTAGTGAGACATCTAGTGGCCAAAA   1210
SYKES.KE.SYK173 Gag E__A__C__N__W__K__D__D__P__P__A__T__S__G__G__Q__S__.__.__.__.__E__N__S__.__.__.__.__.__.__.__S__Q__N__M__A__S__E__T__S__S__G__Q__K    p17
```

```
                                                                                           Gag p17 \/ Gag p24
B.FR.HXB2 Gag       __.__.__.__.__.__.__.__.__.__H__S__N__.__Q__V__S__.__.__Q__N__Y__P__I__V__Q__N__I__Q__G__Q__M__V__H__Q__A__I__S__P__R__T__L__N__A_    384
B.FR.HXB2           ............................CACAGCAAT...CAGGTCAGC......CAAAATTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA    p24

A.UG.U455           ...........................................AG---T......-----------C--------A--TGCA--A------CCA-----C------T----------G--C--G------    706
C.ET.ETH2220        ............................GGA-AGG-C.........--T......--------T--------------T--G---------------------C--------G----------G------    625
D.CD.84ZR085        ............................A------GC...---------......--------T--------A------C-A--A--------------C------C----------------G--C---    768
F1.BE.VI850         ...........................................-----T......-----------------A-----TC-T-----A-----------C---T-TC-----------------------    594
G.SE.SE6165         ............................A-------C...--A------......--------T--------------TGCA--A--------------C---------A-------------G------    654
H.CF.90CF056        ............................A-AGA---C...A-------T......--------T--------A-----TGCT--A-----G--------C-----------------G--C---------    598
J.SE.SE7887         ............................G---A--G-...--------T......--------T--------------TC-G--A------CC------C------C-----------------------    574
K.CM.MP535C         ...............................GA---A...GG------T......-----------------A-----TC-G-----------------C------C-----------------------    450
01_AE.TH.CM240      ............................AG-----GC...A-A------......--------------------A--TGCA--A----------C-------C-TT----------------G------    816
02_AG.NG.IBNG       ........................................AGCAG----......--------------------A--TGC-A-A---------AC-------T----G---------------------    782
03_AB.RU.KAL153-2   ............................AG-----GC...A-------T......-----------C--------A--TGCA--A---------AC---C---T----G--------G-----G------   1302
04_cpx.CY.94CY032-3 ............................GGT----GC...A-T------......--------------------A--TGCA--A------------------AG---T--------------G------    617
N.CM.YBF30          ..................................-GC...A-TA----T......AGG-----T---C----C----CTGCT--A--A---------------C-GC-GA----C-----C--------T    815
O.CM.MVP5180        ...............................CCT-GG...---ACA--T......-----------------AACA--TGCA-----A--------------A-----C--C--C--G------------   1272
O.BE.ANT70          ..................................-GG...----CGG-T......-----------------ATCA--TGCG-----A--------------------C--C--C--G------------   1296
CPZ.US.CPZUS        ...............................--T--C...ATT-GT---......AGC--------AG--A-A------GCT-----A--G--------------A--G--T--C--------------C   1299
CPZ.GA.CPZGAB       ...............................GCTGGC......A-T--T......GG---------CC-C--A--A--TGCT--A-----G-----G-----------C-----C-----C--G------   1312
CPZ.CM.CAM3         ...............................GA--GC...GCT-----T......AG---C-----AG----A-----TGCA--A--A--GC-A--G------C----G-----C--G--C--------T    641
CPZ.CM.CAM5         ...............................----GC......A-A--T......AG---C-----A-----A-----TGCA--A--A------A-------------G--------G--CC-------T    944
CPZ.CD.CPZANT       ...............................CATCT-AGTGGC-AAG-G......AG---------C--CA-AGT-G-TGCAGGA--AAT-GCAAGG------C-AC-GA----A-----C--------C    663

H2A.GW.ALI Gag      __.__.__.__.__.__.__.__.__.__P__P__S__.__E__Q__G__.__.__G__N__F__P__V__Q__Q__V__A__.__G__N__Y__T__H__V__P__L__S__P__R__T__L__D__A_    p24
H2A.GW.ALI          ............................-CACCT-G-...G-ACAGG-G......GG---C-T---CG--CAA--AGTAGC-...--CA-CTACACC---GT-C-GC-GAGC--CC----C---G----T   1563

H2A.SN.HIV2ST       ............................-CACCT-GC...GG-AA-G-A......GG---C-T---CG--CAA--AGTGGC-...--CA-CTATACC---GT-C-AC--AGT--CC----CC-------T   1015
H2A.DE.BEN          ............................-CACCT-G-...GG-AAA--A......GG---C-----CG-GCAA--AGCGGGT...--CA-CTAT--C---GT-C-AC-GAGC--CC-----C--------   1570
H2B.GH.D205         ............................-CACCT-GC.........G--......GG------T--AG-GCA---ACTGGCT...--CA-CTAC--C--C-T-C-GC--AGC--CC----C--------T   1531
H2B.CI.EHO          ...............................CCA-C-...AGC...C--......-T-GC---T--AG-GCA---A-TAGCT...--CA-TTATTCC--C-T-C-GC--AGC--CC----C--------C   1532
H2G.CI.ABT96        ...............................CCACC-...AGT...G--AGAGGAGG---C-----AG-GCA---AGTAGG-...--CA-CTAT--T----T-C-AT--AGT---C--------------    933
MAC.US.MM251        ............................-CATCT-GC...GGCAGAG-A......GG---------AG--CAA--A-TAGGT...--TA-CTAT--C--C-T-C-AT--AGC--G-----A--------C   1508
SMM.US.SIVSMMH9     ............................-CACCT-G-...GGCAGAG-A......GG---------AG-GCA---AGTAGGT...--CA-TTATACC--C-TAC-TT--AGT--A-----A--------T   1002
STM.US.STM          ............................-CACCT-G-...GGCAGAG-A......GG---------AG--CA---AGTAGG-...--CA-TTAT--C--C-TAC-AT--AGC--A-----C--G------   1176
TANTALUS.UG.TAN1    ............................ACACCT-G-...GGCAGACAT......GGT---------G----A--AC-G-AT...AAT---TG-------ACTC-TT-G-----AC-T--CC--------   1387
GRIVET.ET.GRI-677   ............................-CACCTGG-...GGC-AATCA......AG---------AG----AA-TC-G-AT...AATGCCTG------C---C-TT-G--T--GC-C--G--------G   1194
VERVET.KE.VER-9063  ............................-CACCTGG-...GGC...--T......-------T---AGCGCAA--AC-AGGA...AATGCCTG-------GTTC-AC----C---C-C--A--------G   1427
VERVET.KE.AGMTYO    ............................-CACCTGGG......TG---T......--G----TT--AGCGCAA--AC-AGGA...AATGCCTG-------GTAC--T-G-----GC-C--C--------G    917
VERVET.KE.AGM155    ............................ACACCTGG-...GGC...---......-------TT--CGC-CAA---C-AGGG...AATGC-TG---G---GTAC-AC-T------C-C--CC--------   1416
VERVET.DE.AGM3      ...............................CCACC-...GGT-G---T......-------T---AGC-CAA--AC-GGGA...AATGC-TG-A-----GT-C--T-G-----AC-C--C--------G    928
SABAEUS.SN.SAB1C    ............................ACACCT-G-...GGT-G-TCG......-G---------C--T--CAGTGTT-AT...AAT---TG---C--C---C-AT-G--C--A--G--AC-T-----T   1517
SUN.GA.SIVSUN       ..................................GCA...--A-CTTCT......-------TT--AG-GCA-AGAG-GGGA...CAAA-TTACA-T------C-TC-C--G--------AG-GC-AA--   1485
MANDRIL.GA.MNDGB1   ..............................................--T......A-G----T----G--CAAAG-G-TGCAGCA--A--GTATCAGT--ACTC-A---AGT-----G-T-A--C-GA--    878
LHOEST.KE.hoest     ............................A-G-AAC-A...---ACAG-G......GG------T--CC-GA-AAGAG-A-AT...CA-AG-TG-------ACAC-TT----G--C-----AA--C-AA--   1507
LHOEST.CD.hoest447  ........................................-CTAAA--T......GGC-----T--AC-C--CAGAG-A-AT...CA-AG-TG------CACAC-AT-G-----------AA--C-AA-T    423
LHOEST.CD.hoest485  ........................................---AAA--T......GGC-----T--AC-CA-CAGAG-A-AT...CA-AG-TG------CAC-C-AT-G-----------AA--C-AA-T    423
LHOEST.CD.hoest524  ........................................-CA-AA--T......A------TT---T-GA-TAGAG-A-AT...CAAAG-TG---C---ACCC-TC-------C-----AA-TC-AA--    417
RCM.NG.SIV-NGM      .............AATTACCCAATAGTAACA-TT---...---CAGCCT......G--C--C-G--A---TCT-CT-GA-CATTAAATGCTTG----...A-A-TAG--GAGGAA-AG-AA--TGGA---   1570
COL.__.SIV-CGU1     G......................................................-CTC-AGGG--AG-GCAT--ACC-C-GAGCCCCAGG-C-T--GGAGCCTGGG-TAAGTG-GT-GAGGG-GGAATT   1089

SYKES.KE.SYK173     GGTAGTACAGCAG...GAAAAACAAAAGGCAGCA-CA...-C-CCACCTAGA...GG---C--T--AC-GC-CAGA--TCCT--AAAT--GTG-A-T---ACA-GAG-CC--GT---G---C----AA--   1331
SYKES.KE.SYK173 Gag __V__V__Q__Q__.__E__K__Q__K__A__A__T__.__P__P__P__R__.__G__N__Y__P__L__L__R__N__P__Q__N__Q__W__I__H__T__G__V__P__V__R__T__L__K__T_    p24
```

```
B.FR.HXB2 Gag       _W__V__K__V__V__E__E__K__A__F__S__P__E__V__I__P__M__F__S__A__L__S__E__G__A__T__P__Q__D__L__N__T__M__L__N__T__V__G__G__H__Q__A__A__    p24
B.FR.HXB2           TGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCA   1381

A.UG.U455           -----G--------------C--------------------A--------------------------G-----------------------T-TG-----G--TGT------------C--G-----T-    836
C.ET.ETH2220        -----------------G--A-----------------G--A------------A----T--------------------------------------------T-------------------------    755
D.CD.84ZR085        ------------A-------A----------A---------A----------------------------------------------------------------------------------------    898
F1.BE.VI850         --------G--GA----------------T-----------A-----------------------------G-----T--CAC---------------------T-------------------------    724
G.SE.SE6165         --------------------A-----C-----T--------A--------------------------------------------------TCT----T----------------G-----------T-    784
H.CF.90CF056        --------------------A--------T-----------A--------------------------------------------C-----TG-T--------T-------------------------    728
J.SE.SE7887         -----------GA-------A--A--------------------------------------------------------G-----------T---------------A-A--------C--------T-    704
K.CM.MP535C         -----G--G---A----G-----------------------A------------A----------------------T---------C------------------------------------------    580
01_AE.TH.CM240      -----G--------------A----G---T-A---------A-----------C--------------G-----------------------T-TG--------T-T------------C--G-----A-    946
02_AG.NG.IBNG       -----G------A-------A----G---------------A--------------------------G-----------G-----------T-TG----------T------------C--G-----A-    912
03_AB.RU.KAL153-2   -----G------A-------A--------------------A--------------------G-----G-------------------------TG-----G----T------------C--G-----T-   1432
04_cpx.CY.94CY032-3 ------------A-------A--------------------A-----------C--------------G--------------------G----TG--------T-T------------C--G-----A-    747
N.CM.YBF30          -----G-----GA----G--------C--T--T--------A-----A------ATG--C--G--------G--A--G--CTC----C----T--T---T----T-----A-----------G-----A-    945
O.CM.MVP5180        --------G-C---------------C--T-A---T---A-T--T--T------ATG--------------G--TGT---CT-T---A-C--T--------G--TG-CA-A--------C----GG--TT   1402
O.BE.ANT70          --------G-C---------A--A--C--T-A---T---A-C--T--T-----CATG-----G-----G-----A-TTT-CT-T---A-T--T--T--------TG-CA-A--A----------G---TT   1426
CPZ.US.CPZUS        ----------CG--------A-----C--T-A---G-----C--C--G-----CATG--C-----------G--G--T--C------G-C-----T-----C--TG-TA-A-----------G-G---A-   1429
CPZ.GA.CPZGAB       -----------G-----G--A-----A--T-----T--G--A-----A-----------TC-------G------TTA--T--G---G-T---------T----TG----A-------------G-----   1442
CPZ.CM.CAM3         -----T------A-------A---AA---T-A------G--------A-----CATG--CC-G-----------A--T---------G-T--T--T-----T--TG-TA-A--A--G--C--G-G---T-    771
CPZ.CM.CAM5         -----C--G---A-------A---AAC--T-A---------A-----A------ATG--CC-G--G--G-----T--T--C--G---G-C--T--T-----T--TG-C--A--A--G--C--G-G---T-   1074
CPZ.CD.CPZANT       --------GTGT-----------AAA-----AT--------C--C--T--------T--T-----------G--A--T--T--T-----------------T--TG----T----AC-------G-----    793

H2A.GW.ALI Gag      _W__V__K__L__V__E__E__K__K__F__G__A__E__V__V__P__G__F__Q__A__L__S__E__G__C__T__P__Y__D__I__N__Q__M__L__N__C__V__G__D__H__Q__A__A__    p24
H2A.GW.ALI          ---------T----------A---AAG---G-GG-------AG-G--AGGA---CAG---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AC-------------   1693

H2A.SN.HIV2ST       ---------C-------G--A---AAG---G-GG-------AG-G--AGGA---CAG---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AC-----------T-   1145
H2A.DE.BEN          ---------T----G--G--A---AAG---G-GG-------AG-G--AGGA---CAG---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AT--C--------T-   1700
H2B.GH.D205         --------GT----------A---AAG---G-GG-------AG----AGGA---CAG---C-----------TG------TT-T---A----TCAG--------TTGT--A--A-A------G-------   1661
H2B.CI.EHO          ---------T--------------AAG---G-GG-------AG----AGGA---CAG---C----------GTG------TT-T---A-T--TCAG---T----TTGT--A--A-A------G-------   1662
H2G.CI.ABT96        ---------T-G--------A---AAG---G-GG--------G-G--AGGA---CAG--GC-G--------CTG-CTA---T-T---A-C--TCAA--------TTGT--A--A-A--------------   1063
MAC.US.MM251        ---------T-GA----G--A---AAA--TG-AG-------AG-G--AGGA---CAG---C-G--------CTG------CT-T--CA-T--TCAG---T----TTGT-----A-AC--------G--T-   1638
SMM.US.SIVSMMH9     ---------T-G-----------AAAA--TG-GG----G--AG-G--AGGA---CA---GC----------CTG---T--CT-Y---A-C--TCAA--------TTGT--A--A-A------G-------   1132
STM.US.STM          --------GT----G--G--A---AAA--TG-TG--------G-G--AGGA---CAG---C----------CTGT-----CT-T--CA-C--TCAAC--T----TTGT-------AG-------------   1306
TANTALUS.UG.TAN1    -----T---AC---T--G--A--AAGG---G-GG-----A-AG----------CCA---CC-C--------GTG-TTAAGTT-T--CA-T--TCAG---T----TGTCA-A--A-AC-------GG--A-   1517
GRIVET.ET.GRI-677   -----C---TGC--G--G--A--AAGG-GGG-AG-------AG-C--------CCA----C-C-----G---TGTCT-T-CT-T---G----TCAG-----C--TGT-A-A--A-AC-----G-GG--AT   1324
VERVET.KE.VER-9063  ----------C------G--A---AAA--TG-AG----GA-AG-G--A-----CCA----C-C--------GTGT-----CT-T--CA-C--TCAA-----T--TGTGT-A--A-AC-------GG--AC   1557
VERVET.KE.AGMTYO    ----------C------G-----AAAA--TG-AG-----A-AG-----------CA---CC----------CTG---A--CT-T--CA-T--TCAG-----T--TGTGC-A--A-AT-------GG--AT   1047
VERVET.KE.AGM155    ----------C------------AAAA--TG-GG-----A-AG----------CCA---CC-C--------CTG-------T-T--CA-C--TCAG-----T--TGTCT-A--A-AT-----G-GG---T   1546
VERVET.DE.AGM3      ----------C------G-----AAAA---G-AG-----A-AG-G--------CCAG--T-----------GTG---A--CT-T--CA-C--TCAA-----T--TGTCC----A-AC-------GG--GC   1058
SABAEUS.SN.SAB1C    -----------GA-T--G--A---AAG-----TG----G--AG-----------AGC--CC--G----------T-T---TT-T---A-C--TCAA-----C--TG----A--A-A---C--G-GG--AC   1647
SUN.GA.SIVSUN       -----G--GA----G----------GA-GG-AA------ACAG-GG-A------G----T---ATGAC------A-TT---G----CA-C--TGTG-----G--TG-CA-A----AG-------G---T-   1615
MANDRIL.GA.MNDGB1   ---------AC---G-----A---AAG-GG-AA--G--G--C--C--TC-A--C--------GA----------A-T-AGT--T-----G--T-T------G--TG----A--A-AT-----G-G---A-   1008
LHOEST.KE.hoest     -----G---A----G-----C-GA-G--GG-AA------ACAG--G-A-----------C---A-----AA---TTTG--TG-T--------TGTT---T----TG--A-T--A-AT-----G-GC----   1637
LHOEST.CD.hoest447  ---------A----------T-GA-GA-GG-AG------ACAG-GG-A--------T--CC-CA-----AA---TTTG---G-T--C-----TGT----T----TG-CA-A--A-AC-------GG--A-    553
LHOEST.CD.hoest485  ---------A----------T-GA-GA-GG-AA------ACAG-GG-A--------T--CC-CA-----AA---TTTG---G-T--C-----TGT----T----TG-CA-A--A-AC-------TG--A-    553
LHOEST.CD.hoest524  -----G---A----------T-G--GA-GG-AG------ACAG-GG-T-----------TC-GA-----AA----TTG---G----C-----TGT----T----TG-TA-A--A-AC-------GG--A-    547
RCM.NG.SIV-NGM      ....................................-----AG-G--T--------G---C-G--------CTG--TA--CT-T---G----TCAG---T----TG-CA-A----A--------GT---T   1664
COL.__.SIV-CGU1     GCT..............................--TTCCT--GCT--------C-T---T-AT--TACG--G--A-TAG-TT-T---A-G---TTG-----C--T-TTT---ATACT-------G-TT-C   1189

SYKES.KE.SYK173     -----GG-G-CT--GA-CTCC--AAAA--TGATG--TCCA-AG-G--AT-A---CAG--TC--A--------TTT-TT--TT-T---C-C--TGG------T--TG-T--A--A-AT-----G-G-----   1461
SYKES.KE.SYK173 Gag _W__V__E__A__V__N__S__K__K__F__D__A__S__I__V__P__L__F__Q__A__L__T__E__G__F__I__P__Y__D__L__N__G__M__L__N__A__V__G__D__H__Q__G__A__    p24
```

```
B.FR.HXB2 Gag       M__Q__M__L__K__E__T__I__N__E__E__A__A__E__W__D__R__V__H__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    p24
B.FR.HXB2           TGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAG   1511

A.UG.U455           ----------------T----------------------G-----C--GT-A----------------------C-----------------------------------------------------    966
C.ET.ETH2220        ----------------C-------------G-------------C--GT-A-------------------G------------A--------C--------------------------A-----    885
D.CD.84ZR085        ---------------------------------------------C--GC-A----------G--------G--------------------------------------T--T-------------   1028
F1.BE.VI850         ----------------C----------------------------C---T-A-------------------GCCC---------------G-----T-----G-----T-----T-----------    854
G.SE.SE6165         --------C----G--T--T-------------------G-----C--GA--------CAA--G---------T--C--------------A-------------------T-----------------    914
H.CF.90CF056        ----G-----------T--A-------------------------C--G--A----------------------C------------A-------------------C--T-----------------    858
J.SE.SE7887         ----------------T--T-------------------------C--G--A-------A-------------------------------G-------------------T-----------------    834
K.CM.MP535C         ----------------T--T-------T-----------------C--GT-A--C-------------------CC-----------A-----------G-----G--------------------C--    710
01_AE.TH.CM240      ----------------A---------------C----------------G--A--C-----A----------------C---------------G----------------------------------   1076
02_AG.NG.IBNG       ----G-----------T--------------------------------G--A--------A----------------C---------------G----------------------------------   1042
03_AB.RU.KAL153-2   ----------G-----T---------------------T----------GT-A-------CA--G---------T--C-------G-------------------------------------G--G---   1562
04_cpx.CY.94CY032-3 ----------------T----------------------C-----C--GACA--------A----------------C----------------------------------T-----------------    877
N.CM.YBF30          ----G---C-G--G--AGT---------------A-----C--------GACA--------C-C--TG--A--AC-AC-C-----G--AC-------C--T--A---------T-----------A-----   1075
O.CM.MVP5180        -A---G----G--G--AGTA--------------A----------------ACT------CCAGCAATG-----GT-AC--------G-----A--G-------CA-------------T--T-----A-----   1532
O.BE.ANT70          -A---G--C----G--AGTA--------------A-T---G----------ACT--C---CCA-CG-T------GT-GC--------G-----A--G-------CA-------------T--T--G--A-----   1556
CPZ.US.CPZUS        -----G-T-----G--AG----------------A-----T--A-------ACA-----TC-----------C-----A--------G---------------C----------------T--------C--C--   1559
CPZ.GA.CPZGAB       -----G--C----G--AGT---T-----------C--C--G------C-CT-A-----CACT-----T--T--G--A--C-----A---C-A--G--G--------T-----T--------G--C--C--   1572
CPZ.CM.CAM3         ----GG-C--G--G--AGT---------------A----------------ACA--C--T-----------T--AG-A--C-----A---C-A-----GG--C-A-----------------T--G---------    901
CPZ.CM.CAM5         ----GG-T-----G--AGT---------A-----A-----G--------GC-A--C--T--------------A--A--C-----A--AC-------GG---A--G------------T-----------   1204
CPZ.CD.CPZANT       ----GG--C-------AGTA--------------A--T--G--------GT-A--C--CACT---------A--AG-ACAGG---A--AT-A--G-------CA---------T--------G--A--A--    923

H2A.GW.ALI Gag      M__Q__I__I__R__E__I__I__N__E__E__A__A__D__W__D__V__A__H__P__I__P__.__G__P__L__P__A__G__Q__L__R__E__P__R__G__S__D__I__A__G__T__T__S    p24
H2A.GW.ALI          -------AA-C-GG----TT--------A-----A-----C------GTT-CA-----CA-A-CA...--C--CT-AC--G----G---C-T-----------A--GTC----------------A--A--   1820

H2A.SN.HIV2ST       -------AA-C-GG--A-TT--T-----A-----A-----T-----CGC-CAA--C---A-A-CA...--C--CT-AC--G-G--G---C-C--G--G---------TC------------G--A--A--   1272
H2A.DE.BEN          -------AA-C-G-----TT--T-----A-----A-----C------TCGCA---C---A-A-CA...--C--CT-AC--G----A---C-C-----C-----A--GTC----------------A--A--   1827
H2B.GH.D205         -------TA-T-G---A-TA--------------A-----C-----CCAGCAA--C--GTCA-CA...--C--A--GC-GG----A--AC-T--G--C-----A--GTCA--T-----------C--C--   1788
H2B.CI.EHO          -------TA-T-GG----TA--T-----------A-----C-----CCA-CA-------TC--CA...--C--A--GC--G----G--AC-C--G---------A--GTCA---------G-----C--C--   1789
H2G.CI.ABT96        -------TA-T-G---A-TA--T--------G--A-----K------CTTCAA------CAA--A...-----GCCGC--G----A--AM-C--G-----G-CA---TCA--T---------G--C-----   1190
MAC.US.MM251        ----G--TA-C-G---T-TT--A--------G--------T-----CTTGCA---C---CAA-CA--T...--A...CA--A---A---C-T--G--G--GTCA---TCA--T--T--------A-----   1762
SMM.US.SIVSMMH9     -------TA-T-G---A-TT--A-----A--G-----T--T------TT-CAA--C--RCA--CA...--T--A--AC--G----A--AC-T-----G-----R---TCC-----T---------------   1259
STM.US.STM          -------TA---------T---A---------------T--T-----C-TGCA-------CAA-CAC-G--T--GC-GC--G----G--AC-T------------C--GTCA--T---------G--------   1436
TANTALUS.UG.TAN1    ----G--TA-C--G--AGTA--------T-----A--CC-G-------T-ACA--C--CCCA-CA-----A--GC-AC--G----G---C-A--G--C-----A--TTCA---------G-----A--C--   1647
GRIVET.ET.GRI-677   -A-----TC-T--G--AGT---T-----A-----A-----G-----C--GACA--CAG-CCA-CA--T--C--GT-AC--G----G---C-A-----C--G-CA--GTCA--T---------------C--   1454
VERVET.KE.VER-9063  -A-----AG-------A-T---------------A--CC-G------GT-ACA--C---CCA-CA-----A---T-AC--------A---C-T-----C--T-----TTCA---------------C-----   1687
VERVET.KE.AGMTYO    -A-----AG-G-------T---T-----A-----A--CC-G------GT-ACA--C---C-A-CC-----A--CC-AC--G----A---C-C--G--C--TC-C--CTCA--T---------G--C--C--   1177
VERVET.KE.AGM155    -------AG-G-----A-TA--T-----------A--CC-G------GT-ACC--C---CC--CG-----C--CT-GC--G-G--A---C-C--G--T--GG-----TCA--T---------G--C-----   1676
VERVET.DE.AGM3      -A-----AG-------A-T---------------A--CC-G-------T--CT--C---CCA-CA-----A--AT-AC--G----A--AC-C-----C--T--A--CTC----------------C--C--   1188
SABAEUS.SN.SAB1C    -A-----AG-G-----TGTT--------A-----G-----T-----CCTTAGA--C--GCC--CGCA-CA---CCCA--T-AG--AGT-C-A--G--C---CAA--GTCA---------G--C---------   1777
SUN.GA.SIVSUN       ----G---A-------TCAT--AGTG------AG------------C---A----C--TCA---GCAGCCTG-ACAGC--GG---AGGCC-A--GACT--T-AT--T-CA-----------GGTA--ATC   1745
MANDRIL.GA.MNDGB1   -----G-C--------TGTA--T------C----A------------CT-ACA-----TCAA--ACA-CCAG-ACAAC--GG---AGGAT-A--GACC--TTCA--CTC---T-----------------TC   1138
LHOEST.KE.hoest     -------AA-------TCA---AGTG-------G------G---------CA---C--GCAA--ACAGCCAG--CAAC--GG---GGGTT-A---ACT--T-AT-C--CA--TG-G-----GGTC--CTC   1767
LHOEST.CD.hoest447  ----G--AA-T-----TCAT--TGTG--A----GG-----------C--GAAA-----GCAA--ACA-CCTG-ACAAC--GG---AGG-C-T---ACC--T-AT-C--CA--TG-C------GTC--CTC    683
LHOEST.CD.hoest485  ----G--AA-T-----TCAT--TGTG--A----GG-----------C--GACA-----GCAA--ACA-CCTG-ACAAC--GG---AGG-C-T---ACC--T-AT-C--CA--TG-T------GTC--CTC    683
LHOEST.CD.hoest524  ----G--AA-------TCAT--AGCG--A----GA--T-------------CA-----CCA---GCA-CCAG-ACAAC--GG---AGG-T-A---ACT----AT-C--CA--TG-G-----G-TC--ATC    677
RCM.NG.SIV-NGM      -A-----AG----G--AGTA--------T--G--A-----T-----CTTGAGA-----CAAT...C----A--AC-AC--G----T--AT-A--G-------CA-----C--T--T--------A--A--   1791
COL.__.SIV-CGU1     CT--GG-C--G--G--TGAG--T---A-AA-G--A-A---G-AT--CCT-C-A-----G--C...CAGCA---ACAACAG-A---AGCAT-A--GC-G--T-CA-C-TCA--T--CA----C-AC--A--   1316

SYKES.KE.SYK173     ----GG-AA-T-----C-T--------------GG--G-C-------CTT-AA-----CCAA-CGCA-CAA---GC-C---A--CAGGAT----G--T--CTCC-CCTCA--------------A--A--   1591
SYKES.KE.SYK173 Gag M__Q__V__I__K__D__I__I__N__E__E__G__A__A__W__D__L__E__H__P__Q__P__Q__Q__P__A__P__Q__A__G__L__R__D__P__S__A__S__D__I__A__G__T__T__S    p24
```

```
B.FR.HXB2 Gag       __T__L__Q__E__Q__I__G__W__M__T__N__.__.__.__N__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_   p24
B.FR.HXB2           TACCCTTCAGGAACAAATAGGATGGATGACAAAT.........AATCCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCAGC  1632

A.UG.U455           C---G----A---------------------GGC.........-----------------G-----C--C----G-------------------------------------------------GTT---  1087
C.ET.ETH2220        --------------------C----------GGG.........--C------G-T-----G-----C--C-----------------------GC-----------------------------GT----  1006
D.CD.84ZR085        --------------------------------GC.........--------------------------C------------------------------------------------------GT----  1149
F1.BE.VI850         -------------------CA---------GGGC.........--C------G-------G-----C--C--------------C-----A-----------------------------T---GT----   975
G.SE.SE6165         -T----G------------ACT--------CGGC.........--C--------------G--------------------------------G------------------------------GT----  1035
H.CF.90CF056        ------G-------------C----------GGC.........------G----------G-----C--C--------------------------------G-----------------T---GT----   979
J.SE.SE7887         ------C------------------------GGC.........-----------------------G--------------------T------C-----------------------------GT---T   955
K.CM.MP535C         C-------------------C----------.........AGC--C------G-------G--G-----C-----------------------T-----C------------------------GT----   831
01_AE.TH.CM240      ---------A-----------------------C.........--------C--------G-----C--C--------G---------------------------------------------GTT---  1197
02_AG.NG.IBNG       ---------A----------------------GC.........-----------------G--------C---------------G-------------------------G------------GT----  1163
03_AB.RU.KAL153-2   ---------A----------------------GC.........--C--------------G-----C--C--------------------A-----------G---------------------GTT---  1683
04_cpx.CY.94CY032-3 ---------A----------------------GC.........--C------G-------G--------C------------------T----G-------------------C---------C-TT---   998
N.CM.YBF30          C-----GGCA-----GG-G-CT--------TGC-.........-----T---G-T-----------T-------G----------G-------G-----C-G---T--G-----------T---GT----  1196
O.CM.MVP5180        C--A-AG--A--G-----TAT-----CT--T-GAGGG......GC-AACT----------------C--C----G--A-------G-G--A---C----C-----G----A------C--T--AGTG---  1656
O.BE.ANT70          C----AG--A--G-----TCAC----CT--C-GGCCC......--C-A------------------C--C----G--A-------G-GT-A---C----C-----G----A------C-----AGTG---  1680
CPZ.US.CPZUS        C---T-G--A------G-G--C---------.........GCA--C-----C--T-----------C--A----G-------G-GG----A--GC-T------G------A------CT-T--CGTTG-T  1680
CPZ.GA.CPZGAB       C--A--G-----------T--G----CA---.........GCA-----T--C-----------G--TG------G-------G----TT-A--G-----C---G-G--T---------T----AGTA--T  1693
CPZ.CM.CAM3         C--T--AGTA-----G--T-CT---------.........GCA--------CG-T--------------A--C-GG------G-GG-T--A--GC-T--C---G-------G------T----AGTT--T  1022
CPZ.CM.CAM5         C--T--A--A------G-T-CT---------.........GCA--------C-----------G-----C----G-------G-GG-T--A--GC-C--C---G----G-A------CT----CGT---T  1325
CPZ.CD.CPZANT       C--AG-G-----G--G--GCA-------T---CACCTCAACAG---GG-GGAG----------G--C--C-----G--------C---A-------------GG-G--C--G-N------T--AGT----  1053

H2A.GW.ALI Gag      __T__V__E__E__Q__I__Q__W__M__F__R__P__.__.__R__N__P__V__P__V__G__N__I__Y__R__R__W__I__Q__I__G__L__Q__K__C__V__R__M__Y__N__P__T__N_   p24
H2A.GW.ALI          C--AG-AG-A-----G--CCAG------TTC-GGCCA......CGGAAT---G-G--------GA-C--C----G---------CCAGA-A--GC--C-G--GTGT--C--G-----C-A---A----A-  1944

H2A.SN.HIV2ST       C--AG-AG-A--G--G--CCAG------TTT-GGCCA......C-AAAT---G-A---------AGC--C----G---------CCAGA-A--GC--C-G--GTGC--C--G-----C-A---A----A-  1396
H2A.DE.BEN          C--AG-AG-T-----G--CCAG------TAT-GGCCA......C-AAAT--CG-A--G-----GA-C--C--C-G---------CCAAA-A--GC-GC-A--GTGT--C----A---C-A---A--T-A-  1951
H2B.GH.D205         C--AG-AG-------G---CAG------TAC-GGGCC......C-AAAT---G-------G---A-C-------G---------TCAAT-A-----GC-G---TGT--CC-------C-AT-------A-  1912
H2B.CI.EHO          C--AG-AG-A-----G---CAG------TAC-GACCC......C-AAAT---G-------G---A-C--C----G---G-----TCAGT-A--GC--C-G---TGT--CC-G-----C-AT-----T-AT  1913
H2G.CI.ABT96        ---AA-AG-A--------TCAG----CACAC-GGCAG......C-AAAC-----A---------A-T--Y----G---------CCART-A---C-TC-A---TGT--C--G-----Y-AT--A--A-A-  1314
MAC.US.MM251        -T-AG-AG-T--------CCAG------TAC-GACAA......C-GAAC--C--A--------CA-C-----C-GG--------CCAA-----G--GC-A---TGT--C----------A---A--A-A-  1886
SMM.US.SIVSMMH9     C--AS-AG-C--------TCA-------TAY-GGCAG......C-AAAC-----A--------CA-C--C--C-GR--G-----CCAAT-A--GC-GC-G---TGC-------------AT--A--A-A-  1383
STM.US.STM          CT-A-CAG----------TCA-------TACCGGCAG......C-AAAT-----A--G------A-C--C--C-GG--------TCAA--A--GC--C-A---TGT--C--------C-AT--AGTT-A-  1560
TANTALUS.UG.TAN1    CT-TG--GCT--G-------AG----CTTTT---GCT......------AGAG-AGAT------CGT--C--C-G-G-----G-T---------C-CC-A---TGT----A------C-AT----T-TCA  1771
GRIVET.ET.GRI-677   CT-AA---------------AG----CCTTC---GCC......------AGA--AGAC-----G-C-CAA--C-G--A----G-T--TT----C---C-A--GG----GCAG-----C-AT--CCAA-AG  1578
VERVET.KE.VER-9063  ---AG-G--A------T---A------TTAC-CAGCC......--C--CAGAG-AGAT--T--G-CT--C----G---------T-----A--GC-GC-G--GTGT----A------C-A---AGT-TCA  1811
VERVET.KE.AGMTYO    CT-AG-A--A-----GT---A------CTAT-C-GCT......--C--C-GGG-AGAT-----T-CC--C--CCGG--------T--T--A---C-TC-A--GTGT--C-A------C-A---AGTATCA  1301
VERVET.KE.AGM155    ---AG-G--A--G--GC---AG-----CTAT-C-GCT......--C---AGGG-AGAT-----G-CC--C---CG----------C-----A--G---C-A---TGT----A------C-AT--AGTGTCT  1800
VERVET.DE.AGM3      C--AG-G--A-----GC-G-A------ATAC-CAGCC......-----CAGAG-AGAT--G--T-CC--C----G---G-----T--------G--GC-A---TGT----A------C-A---AGTGTCT  1312
SABAEUS.SN.SAB1C    C--AA-A--A--G-------A-----CC--C-GAGCA......C-GAATG--G--AAT------A-T---------G-------C--------TC-CC-G---TGT--C-A------C-A---AGTA-A-  1901
SUN.GA.SIVSUN       A--AG-AG-A--G--G---CAG-----CTATGGCCTTAATGGAGCAA--AGAG-TG-----CA---C--C------G-C---G-C--AGA---CA-GG-A-G-G----T---C-CC--CAGA-----TCA  1875
MANDRIL.GA.MNDGB1   ---AG-GG-A------T-G-C--------AT-TGCAA......C-AAATG-A---AAT------AC---C-----G--T-----T--A-----CA-G----G-T-G----A--GTC--T-T--A-TA--T  1262
LHOEST.KE.hoest     A--AG-AG-A------C---CC----CT---GCAGAC.........A-T---G-AGAT------A-----------GA----G-C--T-A--CTA-GG-A---G----C-----CC--CAA--AGTA--T  1888
LHOEST.CD.hoest447  A--AG-AG----G--GC---CT----CC--TGCAGAC.........A-C---G-AGAT------A-G---------GAG---G----T-AA-CTA-GG-A---G----T-----CC--CAA--AGTT--T   804
LHOEST.CD.hoest485  A--AG-AG----G--GC---CC----CC--TGCAGAC.........A-C--AG-GGAT------A-----------GA----G----T-AA-CTA-GG-A---G----T-----TC--CAA--AGTT--T   804
LHOEST.CD.hoest524  A--AG-AG-A-----GCA--CC----CA---GCAGAC.........A----C--AGAT-----GA----C------GA----G------A--CTA-GG-G---G-G--------AC--CAG--AGT---T   798
RCM.NG.SIV-NGM      C-ATA--GCT----------CT----CC--T-GAGCC......--CAAC--A--TG----G---A-C-------GG-AT-----TG-G--T--G---C-A---TGT--G-A--------A---AGTA-A-  1915
COL.__.SIV-CGU1     CT-AG-AGCA------G-G-CT---GGAGAGCCCATA...GCC--CATTTA--GGGGCTGGAT--T-CAGAG-CTGGA-AA-G----......................CAG--TGCA--A--AT--TCT  1422

SYKES.KE.SYK173     CT-AA--GCT--G-------A------T--T-GGCAG......--CAAC--AG-A-A------------A----G---------C--T--A--GC-GC-G--GTGT--GCAGG-C--C-A----GTTTCT  1715
SYKES.KE.SYK173 Gag __S__I__A__E__Q__I__E__W__I__T__R__Q__.__.__N__N__P__V__Q__V__G__E__I__Y__R__R__W__I__I__L__G__L__Q__K__C__V__Q__V__Y__N__P__V__S_   p24
```

```
B.FR.HXB2 Gag       _I__L__D__I__R__Q__-__P__K__E__P__F__R__D__Y__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__   p24
B.FR.HXB2           ATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAA  1762

A.UG.U455           ---T----------------G-----A--------C--G--T--------TA-A----T---------C-----T---------A-------T--------C------------------C---------  1217
C.ET.ETH2220        ---T---------A------------------T-------------------------T------CT------T--A------A----A--T--------------------C----------------  1136
D.CD.84ZR085        ---T-------C-----------------------------------------------C-----------------------------A--A-------------------------------------  1279
F1.BE.VI850         ---T---------A------G-----A------------------------A-A----T----GTC--------T-----------------C-----GGG------------C--A-------------  1105
G.SE.SE6165         ---T-A--------------G-----A-----------------------TA------T----TG-T-G-----T-----------------T------GGC-----------C----------------  1165
H.CF.90CF056        -------------A------G-----A------------------------A----T-T-------T-------T--------CA-------T--G--G-------------------------------  1109
J.SE.SE7887         ---T----T-----------------A-----T------------------A------T----G----------T---------A-------T--------------------T------C---------  1085
K.CM.MP535C         ---T----------------G-----A-----------------------TA------T------C--------T--A-----CA-------A-----G--------------C---C------------   961
01_AE.TH.CM240      ---T----------------G-----A--------C--------------TA-----------------C-----G--A------A----------------C------------------C-A-------  1327
02_AG.NG.IBNG       ---T----------------G-----A--------------T--------TA------T-------T-G-----T--A------A-G--------------C------------------C---------  1293
03_AB.RU.KAL153-2   ---T-A--------------G-----A--------C--------------TA------T---------T-----T---------A-------T-----G--C---------------C--C---------  1813
04_cpx.CY.94CY032-3 --CT----------------------A--------C-----T--------TA------T----TG---C-----A--A------A-C--------G---------------------C--C---------  1128
N.CM.YBF30          -----A--G--C-A------------A--------C---------------A-------C--------------A-----G--AA-------A-----G-----------------AC-C--A--A----  1326
O.CM.MVP5180        --CT-A--T--T--G--G--------A-----A--C-----T--------T-----T--C-----AT-------T---------A-T--A--A-----G---------------------C-T--T--G-  1786
O.BE.ANT70          --CT-A--T--T-AG--G--------------A-----------------T--------C-----AT-------T---------A-T--A--A-----------------------TC-C--T--T----  1810
CPZ.US.CPZUS        --CT-A--T--C-AG--G-----T--A--G---------------------A-------C-----AT-------A--A--------T-CA--A-------CC-----------G---C-T-----A----  1810
CPZ.GA.CPZGAB       --C--T-----C--------G--------G-----------T-----G---A-A-----C------T-----G--T-----------------A--G--------------T--C--TC-C--A--G--G-  1823
CPZ.CM.CAM3         --CT-A-----C-AG--G-----T--A--G--T--C--G--T--------TA-A-----C-----C------G--A--A-----AA----A--A-----------------T--C---C-C--A--A--G-  1152
CPZ.CM.CAM5         -----A--T--C-A---G-----T--A--G--A--C--G--T--------TA----T---------T-------A--A--G--AA-------A------------------T--C---C-T--A--T----  1455
CPZ.CD.CPZANT       -----A--G----A------------A--G-----C-----T-----G--TA-A-----------AA-T-----A--A--G---------CCT--G---GCC--------------------AA------  1183

H2A.GW.ALI Gag      _I__L__D__I__K__Q__G__P__K__E__P__F__Q__S__Y__V__D__R__F__Y__K__S__L__R__A__E__Q__T__D__P__A__V__K__N__W__M__T__Q__T__L__L__V__Q__   p24
H2A.GW.ALI          --C--A-------A---------------G--A--CCA-AG---------TA-A-----C----GCT----G--A--A---A-AGAT-CA-CA-----G-----------TC----AC--C----A--G-  2074

H2A.SN.HIV2ST       --C--A-------A---G-----------G--A--CCAGAGT--------TA-A-----C--G-GCT-G--G--A--A---A-AGAT-CA-CA-------------------CC----AC--C-A--G--G-  1526
H2A.DE.BEN          --CT-A-------A---G--------A-----G--CCA-AG----------A-------C----GCT-G--G--A--A---A-AGAC-CA-CA-------------------CC----GC--C-AA-A--G-  2081
H2B.GH.D205         --AT-A-------AG--G-----------G-----CCA-AG---------TA-A-----C----GCT--C-G--A--A---A-AGAC-CA-CA--G----------------C----AC--C--A-T--G-  2042
H2B.CI.EHO          --A----------AG-----G-----A--------CCA-AG---------TA-A-----C----GCT--C-G--A------A-AGAC-CA-CA-------------------C----AC--C--A-T--G-  2043
H2G.CI.ABT96        --C--A--TG---A------G-----A---------CAGTCT--------TA-A-----C--G-G---R-----A--A---A--GAT-CA-CA-------------------C-G--AC-AC--A-T----  1444
MAC.US.MM251        -----A--TG---A------G-----A--G--A---CAGAG----------A-------C----GCT-------A--A---A-AGATGCA-CA-----G-----------TC----AC--C--A-T----  2016
SMM.US.SIVSMMH9     ---T----TG-G-A------------A--GT-A---CA-AG---------TA-A-----C--G-G---------A--A---A-AGAC-C--CAS-G--------------TC----AC--C--A-T----  1513
STM.US.STM          ---T-A---G---AG-----G-----A---A-G--CCA-AGT---------A-------C--G-G-T-G-----A--A----AGAT-CATCA--G--------------TCGG--AC--C-AA-T----  1690
TANTALUS.UG.TAN1    G-C--------C--------GG-C-----G--T----AG-----C------A-A--T---C--G-C--T-----A--A---A-AC-C--A--T-----------------------AC--C-CA----G-  1901
GRIVET.ET.GRI-677   G-C--A-----TC----G-----T--A--------CCAG------------A-A---------G-C--G-----A--A-----AC-------T--T---------------C----T---C-TA----G-  1708
VERVET.KE.VER-9063  G-AT-A-----C-----G-----T--A-----T--C-AG--T--------TA-------C---G-AA-C-----A--A--G--C---GGA--A------C-A------------T-A--AC-CA-A----  1941
VERVET.KE.AGMTYO    G-C--A-----T--G--G-----T--A--G-----C-AG--T-----G---A-A--T--C--GG-AA-T-----A--A-----C---GG---A--G---C-A------------T-A--AC-CA-T----  1431
VERVET.KE.AGM155    G--T-A--T--C--------G--C--A-----A--C-A---T---------A-A---------G-AA-------A--A---------GGA--A--C---C-A------------T-T---C-CA-T--G-  1930
VERVET.DE.AGM3      G-CT-A-----------G--G--C--A---G-A--C-A------C-----TA-------C---G-AA-------T-----G--C---GGA--A------C-G------------T-A--AC-CA-T--G-  1442
SABAEUS.SN.SAB1C    ---T-A-----T-A------G--T--A-----T--C-AG--T--------TA-A-----C--GG----G-----A-----GA-AGAC-CT-CT-----G--C---------C-GT-TC-----A----G-  2031
SUN.GA.SIVSUN       G-AG-A--A--T-----G--------A--G--T--C---------AC---TA------TC--GG--T-------A--ATC--GC---G----A------G-G-------A---G-AAA--C-A--G----  2005
MANDRIL.GA.MNDGB1   --AACA--TG-------G------------G-T----A------------TA-A-----C--TGTAA-G-----A--A---------GGA--A-----G-TG------CAGC-GCATC--C-TA-AG---  1392
LHOEST.KE.hoest     G-CA-------C-A---------C--A--G--T----A---A----C---TA------T----G-A--G-----A--AGG--GC--T--T--A-----GG-G-------AG--G-AAA--C-T--G--G-  2018
LHOEST.CD.hoest447  G-AA-A-----C-AG--G--G-----A-----A--C-A---A----C----A------T----G-AT-------A--AGG--GC-----T--A------G-A-------AG--G-AGA--C-A--A----   934
LHOEST.CD.hoest485  G-AA-------C-A---G--G-----A-----A--C-A---A----C---TA------T----G-AT-------A--AGG--GC-----T---------G-A-------A---G-AGA--C----A----   934
LHOEST.CD.hoest524  G-AA---------A---G--T-----A-----A--C-AG-------CT--TA------T---GG-CT-------A--AGG--G------T--A--G---G-G-------A---G-AAA--C-A--G--G-   928
RCM.NG.SIV-NGM      --CT-A-----C-AG--G--------A--G--T----AG-----------TA-------C--GG-CT-G--G--A--A--G--AGAC-CA-CA--G--G------------C--T-AC--C--A-T----  2045
COL.__.SIV-CGU1     G-A--A--T--T-----G---AGT--A--GGAT----A-AG----------A-A--T---TC-G-C-----G--A--A-CT--AG--GGA--AA-C--GGCC------G-TA-T-AT---C-AA-A--GC  1552

SYKES.KE.SYK173     --CT-A-------A---G--C-----A---ATT--C-A-------------A-A-----CC-TTGCT-------A-----G--GGAC-CATCA--C---GGG---C-A--TC----AC-----A-T----  1845
SYKES.KE.SYK173 Gag _I__L__D__I__K__Q__G__P__K__E__I__F__K__D__Y__V__D__R__F__Y__H__C__L__R__A__E__Q__A__D__P__S__V__K__G__W__L__T__Q__T__L__L__I__Q__   p24
```

```
                                                                                       see Kaye, J Virol 72(7):5877-5885(1998) Gag p2 role in RNA encapsidation
                                                                                                                              Gag p24 \/ Gag p2
B.FR.HXB2 Gag       N__A__N__P__D__C__K__T__I__L__K__A__L__G__P__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     p2
B.FR.HXB2           ATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAG   1892

A.UG.U455           -------T-----C------T-C-------G------A--G----G-------T----------------------C--------G-----------------A-----G-----------G--------   1347
C.ET.ETH2220        -------------------A--C-------G------A--G----G----T--T-------------------------------G--------T-C---C--A--------G--------G--------   1266
D.CD.84ZR085        ----A--------------------C-------------------G-------T-------------------------------G--G--G---A-------A-----------------G--------   1409
F1.BE.VI850         ----------------------C------------------A---G---------------------------------------G--------TA----------------------C--G--------   1235
G.SE.SE6165         ----------------------C--C----G------A----A--GA---T----------------------------------G---------A-------A-----------A-----G--------   1295
H.CF.90CF056        ----A--T--------C--------A----G------A----A--G----T--A-------------------------------G--------TA-T-----A-----------------G--------   1239
J.SE.SE7887         ----A--T--------C-----C--------------A---T---GA--------------------------------------G--------T--T--------G-----------------------   1215
K.CM.MP535C         -C--A-----------------C-----------G----------G----T--T-------G-----------------------G-----G--TA-------A------A----------G--------   1091
01_AE.TH.CM240      -C-----T-----C------T-C--------------A---A---GA------T----------------------C--------G--------TA-------A-----G--------------------   1457
02_AG.NG.IBNG       -------T-----C------T-C-------G------A---A---G-------T-------------------------------G--------T--------A-----G-----------G--------   1423
03_AB.RU.KAL153-2   ----A--T-----------A--C-------G------A---T---G-------T------------------T---C--------G-----------------A-----G--------C--G--------   1943
04_cpx.CY.94CY032-3 -------T-----C------T-C--C-----------A---A---G-------T-------------------------------G---------A-------A-----------A-----G--------   1258
N.CM.YBF30          ----A--------------ACAGC-CC----------A--G----GA-----CT-------G--------G--C--C--------G--G-----A-CA--------------GC-A--A--G--T---TC   1456
O.CM.MVP5180        --T-A--------------ACAA---C-G--------A-------AA-----TT--------------GT---C-----A-----------G--AACT--C-------A-A-AC-A--A---------GC   1916
O.BE.ANT70          ----C--T--------C--ACAG-----G---T----A--G----GA-----CT-------G------GT---C--------------------AACT--------C-----GC-A--A---------GC   1940
CPZ.US.CPZUS        -C--T--T-----------A----------G---C--------T-GA--C--C------------G-------------A-----G----------CT-----------G--C--A--------C---T-   1940
CPZ.GA.CPZGAB       ----A---------------CAA---C----------------T-G---C--C-----G--------------T-----A--------------ATCT-----------G---C-A--------C---TC   1953
CPZ.CM.CAM3         ----T--T-----C-----A-AC--C----G---CC---------GA-----CT----------------------C-----------G-----A-C---------T-----AC----------T-----   1282
CPZ.CM.CAM5         ----T--T-----C--C--A-GC--C--G-G---C--A-------GA-----CT-------------------C--C--------G--------A-C---------T------C----------C-----   1585
CPZ.CD.CPZANT       ----C--T--------C--ACAC--CC-G--G--T------A---GA--CT-CT-----------T-A-----T-----A-----------C--A-C---------------G-----A-----T---GC   1313

H2A.GW.ALI Gag      N__A__N__P__D__C__K__L__V__L__K__G__L__G__M__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     p2
H2A.GW.ALI          ----C--------C--C--ATTAG-GC-G----G---A--GATGAATC----CT-------G---C-A--C--C--------------G-----A-----G--A--C---T-AA----A-----CT-A-A   2204

H2A.SN.HIV2ST       ----C--------C------TTAG-AC------G-C-A--GAT-AATC----CT-----------C-A--C--C--------G-----T-----A-----G--A--C---T-AA----A-----CT-A-A   1656
H2A.DE.BEN          ----C--------C--C---TTAG-AC------G-C----GATGAATC-C--C--------G---C-A--C--C--C-----G-----C-----A-----G--A--C--GC-AA----------CC-A-A   2211
H2B.GH.D205         ----T-----------C---TTAG-GC-T--G-GC------ATGAATC-C--CT----G------C-A--G--C--C--A--GA-------C--A--G--G--------GC-AA----C-----CT-A-A   2172
H2B.CI.EHO          ----T-----------C--ATTAG-GC-T--G-GC------ATGAACC-C---T-----------C-A-----C--C-----GA-------C--A--G--G--------GC-AA-------G--TT-A-A   2173
H2G.CI.ABT96        ----C--------------ATTAG-AC----G-GG------ATGAATC----TT----G------T-------C-----A--GR----------A--G--A--A--C--GC--A-----------T-A-A   1574
MAC.US.MM251        ----T-----------C---CTAG-GC-G--G-GGC----TGTGAATC-C--C------------C----G--T-----A--------G-----A--A--G-----T---T-AA----A-----CC---A   2146
SMM.US.SIVSMMH9     ----C-----G--C-----ATTGG-GC-C--G-GTC----YAT-AATC-C--TT-----------T-------C--------------R--G--A--A--A--AS-T---C-AA----A--G---T---A   1643
STM.US.STM          ----T--T-----C--C--ATTGG-A--G----GTC----CATGAATC-C--C--------G---T-A-----------A--G--T--G-----A--A--G--A--T---C-GA----A-----CT---A   1820
TANTALUS.UG.TAN1    ----C-----------C---TTGG----G----G-C-T---AT-CATC-A--T--------G---C-------T--------------------A--A-----A----AGC-CA---TA---------CA   2031
GRIVET.ET.GRI-677   ----C--T--G--------ATTG---C-G----G-------ATGAATC-A--CT-G--G------C-A-T---T--C--------------G--ACAA--------T-AGC-AA---TA---ATG-----   1838
VERVET.KE.VER-9063  ----T-----------C--AGTC--CC-G----G---A---ATGCACC----CT-----------C----T--C-----A--------G--C--AA--T-C--A---------A----A---ATG---CA   2071
VERVET.KE.AGMTYO    ----T--T------------GTC--CC-G--G-GCC-A---ATGCACC-C--C--T---------T-A--G--T--------G--------C--AA--T-C--A----A---AA----A---ATG---CA   1561
VERVET.KE.AGM155    ----C-----------C--AGTA-----G--G-GCC-A--GATGCACC-C--T--T---------C-------C-----A--G--G-----C--AA-TT-C--A--C-A---CA----A---ATG---CA   2060
VERVET.DE.AGM3      ----T--T-----C-----AGTC--CC----G-GCC-----ATGCATC-C--T------------T-A--T--C-----A-----G--------AA-TT-C--A----A---GA----A---ATG---CA   1572
SABAEUS.SN.SAB1C    ----C-----------------AG-GC-C----GG--A--CATGAATC-A--C--------G---C-T--T-----C--A--CA-T------G-ACAA---------C-TC-AA----G-----T----C   2161
SUN.GA.SIVSUN       ----A--------------ATTGG-AA----G---C-----GA-A-TC-AT--T----G--G-------G---C--C--A--G--T--G--G--GA-------A-G--AGA-CC----A---------GC   2135
MANDRIL.GA.MNDGB1   ----A--------A--C---CAG-----G-G-AGC--A--GAA--GA--A--TT----G------T--GA------------------T--G--ACAA-----A--C---T-AA----A---ATG-----   1522
LHOEST.KE.hoest     ----A-----T--C--C-G-TTAG--A--------C-A---GA--G---AT-T--------G-------A---T--------------------T-CT-----A-G--A-A-A-----A---------GC   2148
LHOEST.CD.hoest447  ----A-------------G--TGG-GA-------C--A---GA--G------TT-------G-------GG--C-----A--------------A-C------A-G---GA-------A---------GC   1064
LHOEST.CD.hoest485  ----A-------------G--TGG-AA-------C--A---GA--G-----GTT-------C-------AG--C--------------------A-C------A-G--AGA-------A--G------GC   1064
LHOEST.CD.hoest524  ----A-----T--C--C--A-TGG-AA----G--C--A---GA--GT--C--TT---------------A------------------------T-C------AAG--A-A-C--A--A--G------GC   1058
RCM.NG.SIV-NGM      ----T--T-----C--------AG-AC-T--G-GTC-----ATGAACC----C--G--G--G---C-A--------C--A--G-C------T--ACAG--C-----T--G--GC-A------------CA   2175
COL.__.SIV-CGU1     ----C-----T--C--C--A-GA---C-G----G-C--CAG...AA-C-AT-TT-G-----C---C-AG-------C--------G--------A-A---C-------AG--GC-A--A--G--T---CA   1679

SYKES.KE.SYK173     ----C--------A--C-GACAG--CC-C--G---A---TCAA-...C----CT-G--G------C--CA---T-----A-----C---------TTA--C-------A-C-CA----A--G--C---GT   1972
SYKES.KE.SYK173 Gag N__A__N__P__E__C__R__Q__I__L__K__A__M__V__K__.__P__T__L__E__E__M__L__Q__A__C__Q__G__V__G__G__P__L__H__K__A__K__L__M__A__E__A__M__V     p2
```

```
               see Kaye, J Virol 72(7):5877-5885(1998) Gag p2 role in RNA encapsidation
                                                                      / Gag-Pol fusion TF protein start
                                                             Gag p2 \/ Gag p7 NC
B.FR.HXB2 Gag       __Q__V__T__N__S__.__.__A__T__.__.__.__.__.__I__M__M__Q__R__G__N__F__R__N__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._     p7
B.FR.HXB2           CCAAGTAACAAATTCA......GCTACC...............ATAATGATGCAGAGAGGCAATTTTAGGAACCAA......................................................   1947

A.UG.U455           T------CA-C-GA--......AGC..................---------------------------GG--CG......................................................   1399
C.ET.ETH2220        --------AC---A--......A-C..................-------------A-A---------A-GG--CT......................................................   1318
D.CD.84ZR085        -----C-----G-G-C......T--G-TGCC............-----------A-A-A---------A-GG-AC-......................................................   1467
F1.BE.VI850         T----C--ATTCAG-C...........................-------------A-A-T-------A-GG----......................................................   1284
G.SE.SE6165         ---G-CCT--GG-G--......--AG-AGCC............---------------A---------A-GG--C-......................................................   1353
H.CF.90CF056        -------------A--......AA---AGCC............-------------A------C----A-GG----......................................................   1297
J.SE.SE7887         ------G--C---A-C......AAC..................-----------A-----T--C------G----T......................................................   1267
K.CM.MP535C         ---G---------C--......-T-..................G------------A------C----A-GG---T......................................................   1143
01_AE.TH.CM240      ---C-C-CA-C--G--......A--..................-----------------------C-A-GG---G......................................................   1509
02_AG.NG.IBNG       T------CA-C-GG-C......AAC..................G--------------------------GG---G......................................................   1475
03_AB.RU.KAL153-2   T------CA----G--......AAC..................-------------A-A-T---------GG--C-......................................................   1995
04_cpx.CY.94CY032-3 ---G-C-T-----G--......--AG-AGCC............-------------A-A----A----A-GG----......................................................   1316
N.CM.YBF30          A--G--GCAGC-GC--......A-A--TAGT............G-CT-TGCA--A--G--A--C----AAGG-AT-......................................................   1514
O.CM.MVP5180        TTCT-CCCAGC-AGAT......TTA-AAGGAGGATACACAGCAG--T-C-----A-----GC-GAA-CCA--TAG-......................................................   1986
O.BE.ANT70          TAC--CCCAGC-AGAT......TTG-AAGGAGGATACACAGCAG--T-C-----A-----GC-AAA-CCA-TTAGG......................................................   2010
CPZ.US.CPZUS        ----A-G-A---CC-C......T-A...............AGCG--T-TC-A----A---A--CGC-G----G-C-......................................................   1995
CPZ.GA.CPZGAB       AATG--TCAG---CA-GGGAGA--A...............GATG-TT-CT-C----AG--AC-AGG-GC-GG--C-......................................................   2014
CPZ.CM.CAM3         ----C--CA---CC-T......A-A...............GGTG-GT-CC----A--G--A---AA-G----A-C-......................................................   1337
CPZ.CM.CAM5         -------CA------C......T-A...............AATG-GT-T-----------A---AA-G---GAGT-......................................................   1640
CPZ.CD.CPZANT       TTCT-CT-AT---G--CAGGGAA-C...............GCAG-CT-TC--------------GGA-AT-GAGG-GGA...................................................   1377

H2A.GW.ALI Gag      __E__A__M__T__P__.__.__A__P__I__.__.__P__F__A__A__A__Q__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._     p7
H2A.GW.ALI          GG-G-CC-TG-CAC--......---C-TATC......CCATTTGCGGCAGCC--ACA-........................................................................   2250

H2A.SN.HIV2ST       GG-G-CC-TGGCAC--......--CC--ATC......CCATTTGC-GCAGCC--ACAG........................................................................   1702
H2A.DE.BEN          AG-G-CT-TGGGAC--......AGCC-TATC......CCATTTGC-GCAGCC--ACA-........................................................................   2257
H2B.GH.D205         AG-G-CCCT--CAC-T......--AC--ATA......CCGTTTGCTGCCG-T--ACA-AAAGCAGGG-A-............................................................   2230
H2B.CI.EHO          AG---CCTTG-CAC-T......T-C---AAT......CCGTTTGCCGCCGCT--ACC-A-AGCAGGG...............................................................   2228
H2G.CI.ABT96        AG---CCTT---CC-T......A-AG--CTA......CCATTTGC-GCTGCT--ACA-AAA-CAGGAG-A............................................................   1632
MAC.US.MM251        AG-G-CCCTCGCAC--......-TGC-AATC......CCTTTTGC-GCAGCC----AGA-GGGACCA...............................................................   2201
SMM.US.SIVSMMH9     AG-T-C-CTG-CGCA-......-GGC-ACTC......CCATTTGCRGCAG-C--ACAGAAAGGACAA...............................................................   1698
STM.US.STM          AG-G--CTTCC-AC--......-ACC-ACTG......CCCTTCGC-GCAGCC--ACAGCAGGGA..................................................................   1872
TANTALUS.UG.TAN1    A---A-GCA-GGAGTT...........................-AC---G----AG---CACCGCGAG-AGGTAGGGGCAGAGGC.............................................   2089
GRIVET.ET.GRI-677   TA-T-G-CAG.................................-AT---G-C--AGTG--ACC-CAG-AA--GGGCCCC...................................................   1884
VERVET.KE.VER-9063  AA-CA-GCAG-C-CA-...........................-AC---G-----CA---GGCAGCAG-AGGAGT-AGGAGACAG.............................................   2129
VERVET.KE.AGMTYO    GACCA-GCA----CA-...........................-AC---G-----CAG--AGG-CCA-AA-GA---......................................................   1610
VERVET.KE.AGM155    GA-CC-GCAG-G-CAG...........................-AC---G-A---CAG--AGG-GGA---GGAAG-CCA...................................................   2112
VERVET.DE.AGM3      AA-TA-GCA--GCCAG...........................-AC--------ACAG---GG-CAG--AGGAAG-CCA...................................................   1624
SABAEUS.SN.SAB1C    AGCG-CCTTTC-GCA-......CAA--AGTT......GGGAAC--CT-TG-----CA---GGCAAGACCC-GGGGTCCCTTAGGGGGAAGAGGTAGACCTCTG...........................   2252
SUN.GA.SIVSUN       -TCT-CTTTT-GGCAG......CAAGTAGGA......AGACAAGCT---G-----CA-AA-TTACCACCC-GAA-CAGCCAGGGGAGGTTTGTAAGAATAGGAGGAGGAGGACCT...............   2238
MANDRIL.GA.MNDGB1   AAC---GGT-GGACA-......T-A...............CAA-ATT-TG-----CAGA-AGGGCC-CAA-GAGG-CCAGTTAGACAACCTACTGGA.................................   1598
LHOEST.KE.hoest     TTC--CC-T-C-GCA-......CAAGTAAGA.........CAA-AC---G-----...-T--C-CCGCTT-GAA-TGCGCAAGGCAGGTTTGTGAGAACAGGAGGAGGAGGTCCC...............   2245
LHOEST.CD.hoest447  AAC--CC-TGC-AAGT......CAA-TGAGA.........CAG-AT---G-A--A...-TG-CACCACCA-GGA-TGCACAGGGAAGGTTTGTAAGAACAGGAGGAGGAGGCCCT...............   1161
LHOEST.CD.hoest485  AAC--CC-TGC-AAGT......CAA-TGAGA.........CAG-AT---G-A--A...-TG-CACCACCA-GGA-TGAGCAGGGAAGGTTTGTAAGAACAGGAGGAGGAGGCCCT...............   1161
LHOEST.CD.hoest524  AAC--CC-TGC-GGGG......CAG-TGAAA.........CATC-C---G-----...CAA-CACCACCT-GAA-TGCACAAGGGAGATTTGTAAGAACGGGGGGAGGTGGCCCA...............   1155
RCM.NG.SIV-NGM      GATG--GCA--G-AAT...........................--C---GCC--ACAG--AGGACC---A--AGG-CCCCCTAAACTAGGGGGAGGA.................................   2245
COL.__.SIV-CGU1     G--GT-CCAGC-AGAG.....................AGAACC-AC-----TG-AGT-AAG-CAGCA-AATG-TTCAATTGTCAGGGAATAGGGCACTTAGCCAGAATGTGTCCC...............   1773

SYKES.KE.SYK173     AATG-CCCAGC-G---......-TA..................-AT---G-A--AG-GCCATCAAAAG---GATC-ATGATTAAATGCTACAATTGTGGGCAAATTGGGCACATGCAAAAAGACTGCAAA   2078
SYKES.KE.SYK173 Gag __M__A__Q__Q__S__.__.__V__.__.__.__.__.__.__N__M__V__Q__G__P__S__K__G__R__S__M__I__K__C__Y__N__C__G__Q__I__G__H__M__Q__K__D__C__K_     p7
```

```
B.FR.HXB2 Gag       _.__R__K__.__I__V__K__C__F__N__C__G__K__E__G__H__T__A__R__N__C__R__A__P__R__K__K__.__.__.__G__C__W__K__C__G__K__E__G__H__Q__M__K__    p7
B.FR.HXB2           ...AGAAAG...ATTGTTAAGTGTTTCAATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTAGGAAAAAG.........GGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAG   2062

A.UG.U455           ......-GA...-GAA------------C--------------A---CT-----A------T-----------------.........--------------C--G--A------------------   1511
C.ET.ETH2220        ...-A--GA...GCAA----A--------C--------G---------CT--------------------------A.........-----------------------------------------   1433
D.CD.84ZR085        ...-----A...-----------------C------------------T---A--------T-----------------.........-----------------------------------------   1582
F1.BE.VI850         ...----GA...G-------A-----T----------------A----T----------------------A------.........------------------GA------------------   1399
G.SE.SE6165         ...----GA...-C-A-C--A-----------------G-----A--TCT------------------------------.........-----------------------------T----------   1468
H.CF.90CF056        ...-----A...T-------A--C-----C-----------G--A----T------------------------------.........------------------GA--------T--G-------   1412
J.SE.SE7887         ...-A--GA...---------------------------C----A----T---A-A---C-------------A------.........-----------------------------------------   1382
K.CM.MP535C         ...-----A...-----------C-----C-------------------T----------------------A------.........-----------------G-----------T--G-------   1258
01_AE.TH.CM240      ......--A...-GAA-------C-----C-----T-G------A---CT----------------------A---C--.........--T-----------C--G-----------T----------   1621
02_AG.NG.IBNG       ......-GA...-CAA-A-----------C--------------A---CT--------------A-------------G-.........-----------------G--A------------------   1587
03_AB.RU.KAL153-2   ......--A...-GAA-------------C--------------A---CT------------------------------.........-----------------------------C--T----------   2107
04_cpx.CY.94CY032-3 ...----GA...-C-A-------------C--------G-----A--TCT------------------------------.........-----C-----G------------G-----T----------   1431
N.CM.YBF30          ...--G--A...CCCA----A--------------------G--C--TTTG--A-----C--T-A---------A-G-GGA.........------------G-----GC-A---------T----------   1629
O.CM.MVP5180        ...-A-GG-...CCCA-A--A--C------------A-----G--A--T-T---A-A---C--TC-A--A-----A----G-.........--T--C--------------C--------T-------------   2101
O.BE.ANT70          ...-A-GGA...-C-A-A--A--------C-----A-----G--A--T-T---A--------TC-A--A-------------.........--T--C--------------C--------T-------------   2125
CPZ.US.CPZUS        ...G-G-GA...-AAA-C--A-----T-----------------G--A--TCT-----------------T--A--A-G---A.........--G--C-----G-----GC-----------T----------   2110
CPZ.GA.CPZGAB       ...-A--GA...-AAA-A--A--C--T--------G--------T---CT---T--------T-AA--A--A--A-G---A.........---------CGG-----GC-A--G----------------   2129
CPZ.CM.CAM3         ...-CT-G-...-AAA----A--C--T--C--------G-----A--TTT----------C-----A-----C--A-GG---.........--A--C-----G-----GC-A--G--G--T---------A   1452
CPZ.CM.CAM5         ...-AT----...-AAA----A--C-AT---------------------TGT---T-----C-----A-----A--A-G---.........--A--C---------C---C-------G--TA-C-------   1755
CPZ.CD.CPZANT       ...-A--GA...CC-C-C--A-----T--C--C--T-----G--C--T--T--A---------A------A--A-GG--A.........-----C----G-------C--------------GC-T---A   1492

H2A.GW.ALI Gag      _.__R__R__.__T__I__K__C__W__N__C__G__K__E__G__H__S__A__R__Q__C__R__A__P__R__K__Q__.__.__.__G__C__W__K__C__G__K__P__G__H__L__M__A__    p7
H2A.GW.ALI          ...----G-...-CAA-------C-GG--------A--G---------T-G--A---C-A---C-A---C-A--A--C--A---C--.........-----C-----------C---CC-------T-TT---GC-A   2365

H2A.SN.HIV2ST       ...----G-...-CAA-------C-GG-----C--A--G---------T-G--A---C-A---C-A---C-A--A-----A-G-C-A.........-----C-----------C----C--------ATC---GC-A   1817
H2A.DE.BEN          ...--------...GCAA---G--A--GG--C-----A--G--G--A---T-G--A---C-G---C-G---C-A--A-----A-G-C--.........-----C-----G-----C---CC--------ATC---GC-A   2372
H2B.GH.D205         ...---GG-...-CA--G-CA--C-GG--C----------C-G--A-----------GC-A---------------A-G-C--.........--A--C---------------AAC--------ATC---TC-A   2345
H2B.CI.EHO          ...-AG-G-...-CA--G-CA--C-GG--C--C-----G-CG--A--T---------GC-G----A---------A-GGC--.........--A--C--------------C--C---------ATC---TC-A   2343
H2G.CI.ABT96        ...-A--G-AGC-CAA-A--A----GG------------G--G--A--T----TT---C-G-----A-----C--A-G-C--.........--G--C-----G-----G---CC---C---ATT---GCCA   1750
MAC.US.MM251        ...--------...CCAA---------GG--------G--G--G--A---T-T--A--GC-A-----A-----A--A-G-C--.........--A--C----------------AATG-AC--TGTT---GCCA   2316
SMM.US.SIVSMMH9     ...R-------...--AA--W------RG--C---R-A--G-----A---T-T--A---C-A-T---A--------A-G-C-R.........-----C-----G--------A-C---C--TGTT---GCTA   1813
STM.US.STM          ...C---G-...-CA--A--A--C-GG--------G----C----------T--A-A-C-G--T-AA-G---G--A-G-C-A.........--T----------------C-AC-G--C--T--G---GCCA   1987
TANTALUS.UG.TAN1    ...---GGC...CCACCCCGC--C------A-----AC--ATT-----T-T-CAA-A-G---TCCAAGAG-AG-ACC---C.........AAA--CCTA---------G---CC-------TT-GCA----   2204
GRIVET.ET.GRI-677   ...C--GG-...CCGC-A--A--C--T-------------TTT--A--T-TGCAA--GG-A----A---A--A--AC-G-TC.........AAA--C-TT--G--C--C--AATT--C--TATGGCA----   1999
VERVET.KE.VER-9063  ...---CCC...CCAC-A--A----A------C--A---TTC--A----TGCAA--GC-A----CA-AA--A--A--G-TA.........AAA--CCTT---------G---CT---------TGGCA----   2244
VERVET.KE.AGMTYO    ...---CCC...CCAC-A-GA----AT---------A---TTT--C--T-TGCAA---C-A--TCC--AA--A-------CA.........AAA---CTA--G---------ATTG-------T-GCA----   1725
VERVET.KE.AGM155    ...---CCC...CCGCCA-------A---C-----A---TTT--C----TGCAG--GC-G--TCCT-AG--A--A----TA.........AAA---CTT---------------CC---G---TT-GCA----   2227
VERVET.DE.AGM3      ...---CCC...CCA--A-------A----------A---TTT--C--T-TGCAA---C-A---CCT-AA--A--A--G-T-.........A-A--C-T----------G--ACC---G--TTT-GCA----   1739
SABAEUS.SN.SAB1C    ...-ACCCA...-A-A----A----A----------A--GCCT--T---CTG--T---TT---T-A-----------G-C-A.........-----C------------C-GCCC--AC--T--G-------   2367
SUN.GA.SIVSUN       ...--G---...CCCA-G-CT-----T---------T---CC---A--TCT---A-ATC-G-----A-AG--A-A----GGACCACCCGGCTC---C-----------------ATG-----TA--CAAGCTC   2362
MANDRIL.GA.MNDGB1   ......-G-...-AACC--TC--C------C---AAT-----------TGT---A--GTTC-T--A---------A-GG--A.........--G--C-----T------GCAATG-AT--T--G-AAGCTC   1710
LHOEST.KE.hoest     ...--G--A...CCCC-C-CT--C------------A--GCC---T--------A----TG-----ACAG--C--AC-GG-A.........--T--------C-----C-GTA---AG--TAG-T-TGC-C   2360
LHOEST.CD.hoest447  ...-------...CC-T-A-CT--C--T--------A--GCC---T--T-----A----TG--T--ACAG-----AC-GG--.........--A--C-----C-----T-GCA-G-A---TAGGT-TGCTC   1276
LHOEST.CD.hoest485  ...-------...CC-T-A-CT--C--T--------A--GCC---T--T-----A----TG--T--ACAG-----AC-GG--.........--A--C-----C-----T-GCA---A---TAGGT-TGCTC   1276
LHOEST.CD.hoest524  ...----GA...CC-C-C-CT--------------G--GCC---A--TT----A-A--TG------CAG-----AC-GG-A.........------------C-----TGCTA---AC---AG----GCCC   1270
RCM.NG.SIV-NGM      ...CCT-G-...T-CC-C--A--C-AT-------------AC---T--T--T--T---T------AAA-AAG---A--G--A.........--G--C----G--------G-------T--T-T------G-   2360
COL.__.SIV-CGU1     ...-A--GA...CCAA-AGGAG-AGC-GG-C-A--A-G--GCC-AGGACGG-GAG-CTT-A-AG-A--T--C--A-GGCCAGTG......A-G----TT-CC--CAATC------------ATGCAA-G--   1891

SYKES.KE.SYK173     AAGCC-CTT...-AA-C---A--C--T--------A---AC---A---CT---A--GGCC--T---CAA--C-A--G---T................................................C   2157
SYKES.KE.SYK173 Gag _K__P__L__.__K__A__K__C__F__N__C__G__K__T__G__H__L__A__R__A__C__R__Q__P__K__R__N__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__    p7
```

```
                               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
B.FR.HXB2 Gag       D__C__T__.__.__.__.__E__R__Q__A__N__F.__L__G__K__I__.__W__P__S__.__.__.__.__Y__K__G__R__P__G__N__F__L__Q__S__R__P__.__.__.__E__P__     p6
B.FR.HXB2 Pol                                         _F_._F__R__E__D__.__L__A__F__.__.__.__.__L__Q__G__K__A__R__E__F__S__S__E__Q__T__.__.__.__R__A__N    Pol
B.FR.HXB2           ATTGTACT............GAGAGACAGGCTAATTT.TTTAGGGAAGAT...CTGGCCTTC............CTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACC.........AGAGCCAA   2155

A.UG.U455           -C--C---............-----------------.---------A--...T--------............-A-------G----------------C------------.........--------   1604
C.ET.ETH2220        -C-----C............-----------------.--------GAC-...T--------............-A----------------------C--------------.........--------   1526
D.CD.84ZR085        ----C---............--A--------------.---------A--...T--------............-C-------------G-----C-----------------.........--------   1675
F1.BE.VI850         -C--C---............--A--------------.---------A--...T--------............-A-------G-----C--A-----C--------------.........--------   1492
G.SE.SE6165         -C--C--G............-----------------.---------A--...T--------............-A-------G----------------------A---GA-.........---A----   1561
H.CF.90CF056        -C--C--A............-----------------.---------A--...T--------............-AG---A--G--------A--------C--------G--.........---A----   1505
J.SE.SE7887         -C--C---............-----------------.------------...T--------............-AG---A--G-----------C-----C-----------.........--------   1475
K.CM.MP535C         -C--C---............-----------------.---------A--...---------............-C-------G-----------------------------.........--------   1351
01_AE.TH.CM240      -C--C---............-----------------.---------A--...T--------............-A-------------G----------C------------.........--------   1714
02_AG.NG.IBNG       -C------............-----------------.---------A--...T--------............-AG------G--------A-------C------------.........G--A----   1680
03_AB.RU.KAL153-2   -C--C---............-----------------.--------GA--...T--------............-AG---A--G--------A-------C------------.........-------T   2200
04_cpx.CY.94CY032-3 -C--C---............-----------------.--------GA--...G--------............-AG---A--G--------A-------------A---G--.........--------   1524
N.CM.YBF30          ------AAAATGAA......-GA--------------.-----------G..-----T--C-............--T---A--G--A-----A--C--C-CC----CA-C-A-.........-AG-AA-G   1728
O.CM.MVP5180        ----C-AAAAT.........-GA--------A-----.----------TA..--------C-............GGGGGGCACG--------C----A-G-G----AACA-GT.........GTCC---T   2197
O.BE.ANT70          ----C-GAAAT.........-GA-A------A-----.------C--ATA..--------C-............GGGGGGCACG--------C----A-G-G-----ACC-G-.........-C-C---T   2221
CPZ.US.CPZUS        -C--C--AGCAGGA......A-C--------------.---------ACA...T---T-CC-ATCGTGGAGCGGGGGATCAAAG--A--------C--C--......GGA-AACAG...GAAG-----C-   2221
CPZ.GA.CPZGAB       -C--C--A............-GA-------TG-----.---------AGG...T--------CCGCAG...CGGGAG.........------C--C---G-G----A----A-.........---A----   2222
CPZ.CM.CAM3         -C--CC-AGCTGGA......--A--------------.---------AGG...T---T--C-GTGGAG...CGGGGGATCAAAG--A--------C--C--G---G----GAA.........G--A--C-   1560
CPZ.CM.CAM5         -A-----AACTGGA......-GC--------------.----------GA...T---T--C-GTGGAG...CGGGAGATCAAA---A--------C----CG---G--G-GAA.........G--A--C-   1863
CPZ.CD.CPZANT       -C---C-AGCAACA...AATACAG--A-A-TA-----.---------ACC...GACC--CA-GTGGTG...GGGG-G--G......A--------C---G-G----AGGAGGAAGTAGT...G-------   1603

H2A.GW.ALI Gag      N__C__P__.__.__.__.__E__R__Q__A__G__F.__L__G__L__G__P__W__G__.__.__.__.__.__K__K__P__R__N__F__P__V__T__R__V__P__Q__.__.__.__G__L__     p6
H2A.GW.ALI Pol                                        _F_._F__R__A__W__P__M__G__.__.__.__.__.__K__E__A__S__Q__L__P__R__N__P__S__S__A__.__.__.__G__I__N    Pol
H2A.GW.ALI          -C--CC-A............--A---------GG---.-------CTTGGCCCA---GG...............AA-G---CCTC-CAACTTCCCCG-AACC-GAGTTCCG-A.........G-GATT--   2458

H2A.SN.HIV2ST       -A--CC-A............--A--------GGG---.-------TT-GGCCCA---GG...............AA-G---CCCC-CAATTTCCC-G-GGCC--A-T-CCG-A.........G-G--TG-   1910
H2A.DE.BEN          -C--CC-G............--A--------AGG---.-------TT-GGCCCAC--GG...............AA-G---CCTC-CAACTTCCCCG-GACC--AGC-CCT-A.........G-G--TG-   2465
H2B.GH.D205         -A--CC-A............--A--------GGG---.-------TTAGGACC----GG...............AA-G---CCTC-CAACTTCCCCA-GACC--AGTGCCT-A.........G-GAGTG-   2438
H2B.CI.EHO          -A--CC-A............--A--------GGG---.-------TTCGGCCC----GG...............AA-G---CCTC-CAATTTCCCCG-C-AGGCACC-CAGGG.........GATAGTGC   2436
H2G.CI.ABT96        -A---C-A............-----------AGG---.-------TTTGGACCT---GG...............AA-G---CCTC-CAACTTCCCCA-GACC--AGT-CCT-A.........G-GT-TG-   1843
MAC.US.MM251        -A--CC-A............--C--------GGG---.------CCTTGGTCCA---GG...............AA-G---CCCC-CAATTTCCCCA-GGC---AGTGCAT-A.........G-G--TG-   2409
SMM.US.SIVSMMH9     -A--CC-C............--A--------GGG---.-------CTTGGCCCA---GG...............AA-G---CCCC-CAATTTCCCCA-GGCC----TGCCT-A.........G-G--TG-   1906
STM.US.STM          -A--CC-A............--A-------TGGG---.------CTTTGGCCCA---GG...............AA-G---CCCCACAACTTCCCCA-GGCC--A-TACCT-A.........G-G--TG-   2080
TANTALUS.UG.TAN1    ----CCGCAGT............G-------C-----.--------GA--GCCAACCTG...............GGGA-CAAAGCC--GCAATTT---AGAG--AG--G--G-.........C-TT--G-   2297
GRIVET.ET.GRI-677   -C--C-AGAAT............G-------A-----.-------T-TGGCCAT---GG...............AGGAGC-AA-CCAAG-AATTT-G-G-AAT-C--AG--GA.........CACAGTTG   2092
VERVET.KE.VER-9063  -C--C-GG...............G------TG-----.-------T-TGGACGG---A-...............GGGG-CAAA-CC-AG-AACTT-C-CGC-GCC-CTCATGG.........--T-GA-C   2334
VERVET.KE.AGMTYO    -C--C-GG...............G------TG-----.-------T-TGGACGG---AT...............GGGGGCAAA-CC-AG-AATTT-CCCGCCGCT-CTCTTGG.........--C-GA-C   1815
VERVET.KE.AGM155    -C--C-GG...............G------TG-----.-------T-TGGCCGG---AT...............GGGG-CAAA-CCAAG-AATTT-CCCGCAGCC-CTCTTGG.........G-C-GA-C   2317
VERVET.DE.AGM3      ----C-GA...............G------TA-----.-------T-TGGCCGG---AT...............GGGAGC-AA-CCCAG-AATTT-CCCGCCGCT-CTCTTGG.........G-T-GAGC   1829
SABAEUS.SN.SAB1C    -----CAA...............-A-----T------.-------TTTGGCCC----GG...............-AGAGG-AA-CCCAGGAATTTCCC---GACCTC--TCAG.........-CCAA--G   2457
SUN.GA.SIVSUN       -A--CC-ACAA.........-GA-A-----TG-----.-------T-TGGCCC----AA...............-AGGGG-CCGCCAGGGAACTT-CCAG-CAT-CCTTC....................   2447
MANDRIL.GA.MNDGB1   -A--CC--AAGCCAGCTCAGC--CAGAG--T------.-------T-TGGCCCT---GG...............TCC-TCCAA-CC-GGGAATT--CC................................   1792
LHOEST.KE.hoest     -G--CC-A......AAACCAA-AG--A---TG-----.------AT-TGGTCC----AG...............-AG---ACCTCCAGGCAATT--CC---A-T-G-AG--G-.........T-CAGGG-   2459
LHOEST.CD.hoest447  -A---C-A......AAGCCTA--G--A---TA-----.------CT-TGGACC----AG...............GAGTGGTCCTCCAGG-AACTT-CC-T-GAT-G-GG-GAA.........T-CAGG-G   1375
LHOEST.CD.hoest485  -A---C-A......AAGCCTA--G--A---TC-----.------CT-TGGACC----AG...............GA-TGGTCCTCCAGG-AACTT-CC-T-GAT-G-GG-GAA.........T-CAGG-G   1375
LHOEST.CD.hoest524  -G--CC-A......AAACCAA--G-GA---TG-----.-------T-TGGTCC----AG...............GA-TGGACCTTCAGGCAACTT-CCAA-CAT-G-AG-GG-.........T-CAGG-G   1369
RCM.NG.SIV-NGM      ----CC-AAAA...............G---T------.---------TGCTCC----AAGCA............GGGG--ACCG---AACCTTCCC--GAC-TCCTTG-----.........TACTG-TC   2453
COL.__.SIV-CGU1     -----C-CAAC............-AG-----A-----.-------GCA-CCAGGG--GT...............GG-GCTACAG-C...--CC-TC--C-CC-C--AG-TGAG.........CA-AGACC   1981

SYKES.KE.SYK173     -AG-CC--.........CCA-TTGC------------.---------AGGATGGG--GT...............--C--G-A-GCCC----CC--C--C-CAGT---G--CGA.........GCC.....   2248
SYKES.KE.SYK173 Gag Q__G__P__.__.__.__P__V__A__Q__A__N__F.__F__G__K__G__W__G__V__.__.__.__.__.__S__R__R__P__P__A__N__F__P__V__R__S__E__.__.__.__P__.__     p6
SYKES.KE.SYK173 Pol                                    F_._F__R__E__R__M__G__G__.__.__.__.__.__L__Q__E__A__P__S__Q__L__P__S__E__E__R__.__.__.__A__.__.    Pol
```

```
B.FR.HXB2 Gag       T__.__.__.__.__.__.__.__A__P__P__E__E__S__F__R__S__G__V__E__T__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.    p6
B.FR.HXB2 Pol       __.__.__.__.__.__.__.__S__P__T__R__R__E__L__Q__V__W__G__R__D__N__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   Pol
B.FR.HXB2           C.....................AGCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAAC...................................................................  2198

A.UG.U455           -.....................-----------C---A-T---TG--ATG----A-A---T.....................................................................  1644
C.ET.ETH2220        -.....................------------GAGTCT-AGACCAGAGCCAAC--CCC--C-ACCAGA.........................................................GAG  1578
D.CD.84ZR085        -.....................-----------C---AGCT-CGG-T-TG---AGGAGATA-C-..................................................................  1718
F1.BE.VI850         -.....................-----------C----------G---TCA-A-AG----T---..................................................................  1535
G.SE.SE6165         -.....................---------T-C---A---C--G---TC--A-AG----T-G-..................................................................  1604
H.CF.90CF056        -.....................-----------C----------G---TC--A-AG----TG--..................................................................  1548
J.SE.SE7887         -.....................-----------C-------C--G--CTC--A-AG----T.....................................................................  1515
K.CM.MP535C         -.....................-----------C----------G---TC----AG----T---..................................................................  1394
01_AE.TH.CM240      -.....................-----------C---A-A--GGG--ATG----A-----T---...............................................................GGG  1760
02_AG.NG.IBNG       -.....................---T-------C---------TG--ATG----A-----T.....................................................................  1720
03_AB.RU.KAL153-2   -.....................-----------C---A-A---TG--ATG----A-----T---..................................................................  2243
04_cpx.CY.94CY032-3 -.....................------G--C-CG--AT----AGA-AGGAAA-AG--------..................................................................  1567
N.CM.YBF30          AGCCCAC...............------G---CT------T-ATG---T-CA--AG--...............................................................GAAGAGCAC  1780
O.CM.MVP5180        -.....................---------..........................................................................................AATGGAGGA  2216
O.BE.ANT70          -.....................---------..........................................................................................GATGGAGGA  2240
CPZ.US.CPZUS        -.....................---T------AT----GAT---G---A-CAA-A-------GT.........................................................GACACA...  2270
CPZ.GA.CPZGAB       -.....................G--------GAT------T-ATG---ACCA--AG--AGA.....................................................................  2262
CPZ.CM.CAM3         -.....................---T--G--CCT------T-ATG---A-CA--A--CA---GG.........................................................AAAGAGACA  1612
CPZ.CM.CAM5         -.....................---T-----CATG-----T-ATG---T-CA--A-ACA----T..................................................................  1906
CPZ.CD.CPZANT       -.....................---T-----CAT-----T--A......-CA--AG---CA..................................................................CAA  1640

H2A.GW.ALI Gag      T__P__T__.__.__.__.__.__.__.__.__A__P__P__A__E__P__A__A__D__L__L__E__Q__Y__M__Q__Q__G__R__K__Q__R__E__Q__R__E__R__P__Y__K__E__V__T    p6
H2A.GW.ALI Pol      __T__N__.__.__.__.__.__.__.__.__S__T__P__S__R__A__S__S__G__P__A__G__A__V__Y__A__A__G__E__K__A__K__R__A__E__R__E__A__I__Q__R__G__D_   Pol
H2A.GW.ALI          -ACCAAC........................--C-CCCCCAGCAGA-C-A-CA-CG--CCTGCTGGAGCAGTATATGCAGCAGGGGAGAAAGCAAAGAGAGCAGAGAGAGAGGCCATACAAAGAGGTGAC  2564

H2A.SN.HIV2ST       -ACCAAC........................--C-CCCCCGA-AGACC-A-TA-AG--CCT-CTAGAGAAGTACATGCAGCAAGGGAAAAGGCAGAGAGAGCAGAGAGAGAGGCCATACAAAGAAGTGAC  2016
H2A.DE.BEN          TACCAAC........................--C-CCTCCGGCAGATC-A-CA-CG--ACTGTTGGAGAGATATATGCAGCAAGGGAGAAAGCAGAGGGAGCAGAGGGAGAGACCATACAAAGAGGTGAC  2571
H2B.GH.D205         -ACCATC.....................TG--CCCCC--TGAA-CCAG-A-A--GCAT----C-TCGGGGGGCGACACCATCTGCGCCCCCTGCAGATCCAGCAGTGGAGATGCTGAAAAGTTACATGCA  2547
H2B.CI.EHO          -ATC........................TG-GCCCCC--TGAA-CCAG-ATTC-GCAT----C-TCAGGGTGCGATTCCATCTGCACCCCCTGCAGATCCAGCAGAGGAGATGCTAAAGAACTACATGCA  2542
H2G.CI.ABT96        -TCCATC.....................-G--CC-CC--TGGA-CCAG-A-TA-ACCTACTG-AGAATTACATGCARCTGGGGAGGAAGCAAAAGGAGCA..............................  1922
MAC.US.MM251        -GCCAAC...........................................................................TGCTCCCCCAGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCA  2464
SMM.US.SIVSMMH9     -ACCAAC...........................................................................TGCTCCYCCAGAGGATCCAGCTGTGGATCTACTGAAGAATTACATGAA  1961
STM.US.STM          -GCCAAC...........................................................................TGCCCCTCCAGAAGACCCAGCTGCGGATCTGCTGAGAAGTTACATGCA  2135
TANTALUS.UG.TAN1    -GGC..................TC----GATGCCT-C-CA-GGATTCC-CACA-GTTC-C--G-TGCGGGAGC.................................ATACGACCCAGCCAGGAAACTGCT  2376
GRIVET.ET.GRI-677   GTCTGGAACCAAC.........------C---ATG--A-CAGC.......................................TTACGATCCAGCAAAGAAGCTCCTCCAGCAGTATGCAGAGAAGGGACA  2174
VERVET.KE.VER-9063  -GAG..................C--G--C--TCC-CC--CACC.......................................ATACGACCCAGCAAAGAAGCTCCTGCAGCAGTATGCAGAGAAGGGAAA  2407
VERVET.KE.AGMTYO    -GAG..................T--G--T--TCC-CC----GG--CCA-CCC..............................ATACGACCCAGCAAAGAAGCTCCTGCAGCAATATGCAGAGAAAGGGAA  1897
VERVET.KE.AGM155    -AAG..................T--G--C--TCC-CC--A-AA-TCTA-ACC..............................TTACGACCCAGCAAAGAAGCTCCTGCAGCAGTATGCAGAGAAAGGGAA  2399
VERVET.DE.AGM3      -AAC..................T--G--C--TCC-CC---TCC.......................................ATACGACCCTGCAAAGAAGCTCCTGCAGCAATATGCAGACAAGGGGAA  1902
SABAEUS.SN.SAB1C    -TCCCCCAATGGA...ACGGGATTA-AGCAGGCCG---GAAAATT---A--CA-ACAGAC-TC-GACCAGAGGACCAGGGCCAGACGATCCAGCAACAGCCCTGTTAAAGCAGTATGCTGTTCAGGGGAA  2584
SUN.GA.SIVSUN       ..............................................TG--CC-CCTCTAGAGGATCTGACATTGGGGAACAGAATGACACCACCCCAGTCAAAGGCAGAGAGAGCTCTGGAGACCTACAG  2531
MANDRIL.GA.MNDGB1   .............................................................GG-ACAAGAGGTGACTCCAACAGCTCCACCATTAGAGGAGAAACCTCTGCAGAAAACTCTGAGCACTTA  1861
LHOEST.KE.hoest     GAATTCC.........................................................ATCAGCTCCCCCAATGGAGAGTGCTCCAACAAAAGCAGAGAGAGCACTGGA...GACTTACAGGAC  2529
LHOEST.CD.hoest447  TAGTACC.........................................................ATCAGCTCCCCCAATGGAGAGGAGCCCAACACAAGCAGAGAGAGCCTTAGA...AACATACAGGAA  1445
LHOEST.CD.hoest485  TAGTACC.........................................................ATCAGCTCCCCCAATGGAAAGGAACTCAACAAAAGCAGAGAGAGCCTTAGA...AACATACAGGAA  1445
LHOEST.CD.hoest524  TGATTCC.........................................................TTCAGCGCCACCCATGGAGAAAGTGCCTCCAACAAGAGCAGAGAGGGCAATAGAGACATACAGGAA  1442
RCM.NG.SIV-NGM      -AGGGAGAGAGGAAGGATGGAATC---AGTAC--TCCAGCAGAAGA-ATGCTCAGGA-.................................................ATACCTAGCATTGGGGAGACAGCA  2535
COL.__.SIV-CGU1     TGCCACC...........................................................................................................................  1988

SYKES.KE.SYK173     ................ATCTGCTC----T-TG--G--C-T-GAAGATGGGCCATGGCTC---TGGTCAGCACAGATGAGCCAACAAGCACAAGCGAAGGCACAGAACTCTCCCTCGAAGAAACCCCCCAC  2362
SYKES.KE.SYK173 Gag .__.__.__.__.__x__S__A__P__P__L__E__D__I__E__D__G__P__W__L__T__W__S__A__Q__M__S__Q__Q__A__Q__A__K__A__Q__N__S__P__S__K__K__P__P__T    p6
SYKES.KE.SYK173 Pol __.__.__.__.__.__I__C__S__P__S__G__G__H__R__R__W__A__M__A__H__M__V__S__T__D__E__P__T__S__T__S__E__G__T__E__L__S__L__E__E__T__P__H_   Pol
```

```
                                                                                                                          Gag p6 end <-|
B.FR.HXB2 Gag       __.__.__.__.__T__P__P__.__.__.__Q__K__Q__E__P__I__D__K__E__L__Y__P__L__T__S__L__R__S__L__F__G__N__D__P__S__S__Q__$              p6
                                                                                              Pol TF \/ Pol p10 protease
B.FR.HXB2 Pol       _.__.__.__.__N__S__P__.__.__.__S__E__A__G__A__D__R__Q__G__T__V__S__F__N__F__P__Q__V__T__L__W__Q__R__P__L__V__T__I__K__I__G__G__Q__   p10
B.FR.HXB2           ............AACTCCCCC.........TCAGAAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGGTCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGCAAC   2307

A.UG.U455           ............G--CT---.........-GC---A-------T--A-----G-----A-AC-------GT--------AA---------------------T------G----A-----A--A--G-   1753
C.ET.ETH2220        CTTCAGGTTCGAGGAAG-AA-........A-CTTCT-C-A---A-GAGCTG--A--CAG-G-AG-C-------------AA------------------A--T--T--------A--------A--G-   1699
D.CD.84ZR085        ..........CCCT-C-AGAA........A---G-A---A--GAC-A-----------------C---G---------AA-------------------G-T---------------------A--G-   1829
F1.BE.VI850         ............CC-CT-T--........GA--C--G--C--AAAGAC-GGG-ACTGTACCC---C---G---------AA-------------------T-A-----------A--------A--GA   1644
G.SE.SE6165         .............................C-CCTCC-C-A---A-GAGATG------AA-G-G-TA-ATC-C-------AA------------G--------A------G----A--------A--G-   1704
H.CF.90CF056        ...............C---T-........--C-----------A-C-GA-GG-CA-GGAACC---C---G----------A------------G--------T--T---G----A----A---A--GT   1654
J.SE.SE7887         ...............C---T-........C-C---A----------AG-----------------C-----------AA------------G--------T----------GA-----------G-   1621
K.CM.MP535C         ............CC-CT-T--........GAG-C--G--ACCAAAGACA-GG-AC-GAGCCC------------------AA-------------------A-T-----------AG-------A--GT   1503
01_AE.TH.CM240      GGAAGAGATAACCT-CTTA--........GA--C--G--C--AAAGACA-GG-AC-T-CTCC-------GT--------AA---------------------T------G----A-----A--A--G-   1881
02_AG.NG.IBNG       .........ACCTC-CT-T--........A---C--G-ACC-AG-GACA-GGG-CT-TATCC-----------------AA-------------------T-A--T---G---GA----A---A--G-   1832
03_AB.RU.KAL153-2   ............CC-CT---T........GA-AC--G-AC--AA-GACAGGG-AC-G-ATCC-----C--T--------AA------------G---------T------G---GA-----A--A--G-   2352
04_cpx.CY.94CY032-3 ...............CT--T-........--T---------A----GG----------A-------------------AA------------G--------T------------AC-------A--GA   1673
N.CM.YBF30          ACAGGGGAAGGAG-TG-AGGA........GA-CC--G--AG-A-AGAGA--TCTCTGTACCCA-----------------A-------------------G--A-----G----A------AAAG--G   1901
O.CM.MVP5180        GGCAGTGAAGGA-CAAGAGAA........-----GT---A--GG-GATC-GG-A--G-----C--A--TG-C-------AA--C---------G-CA----AA-A------GC----G-T-----T--T-   2337
O.BE.ANT70          GGAAGTGAAGGG-CAGGAGAA........----G-A---A-AGG-GGCCCG--C--G--C-----G--TG-C-------AA--C---------G-CA----AA-A--T---GC--GAG-T-----C--C-   2361
CPZ.US.CPZUS        ...GGAGAAACAGGGAAA...........GG-A---G-...............-CCCT-...--AGC-------------AA------------G---------C---G------AG---CA-----GA   2367
CPZ.GA.CPZGAB       ...GAAGAGCCAGGAGAAGAAAGAGGGAGAGAGC-G.................TCT-TA...---ACC---C-------AA------------G---------A--C--G----AG---AA-------   2368
CPZ.CM.CAM3         GGAGGAGGAGAGGGAGAAGGA........AA--G-A-C...............CCTGTATCC---G-----C-------AA------------G---------C--GG-G----A----AA-----G-   1718
CPZ.CM.CAM5         ...GGAGAAAAAG-AGGAGGA........GA--G-A-C...............CCTGTATCC--AA-----C-------AA------------G---------C-TGG-G----AG---AA-----GT   2009
CPZ.CD.CPZANT       GAGGACTCAGAAGGG--T-AA........GGG-G--G-...............------ACC---C-CGTA------G-AA--C------------A----AA-GA-GGA-G-TCTC---CA---A---A   1746

H2A.GW.ALI Gag      __E__D__L__L__H__L__E__.__.__.__Q__G__E__T__P__H__K__E__V__T__E__D__L__L__H__L__N__S__L__F__G__K__D__Q__$                         p6
H2A.GW.ALI Pol      _G__G__L__T__A__P__R__.__.__.__A__G__R__D__T__T__Q__R__G__D__R__G__L__A__A__P__Q__F__S__L__W__K__R__P__V__V__T__A__Y__I__E__G__Q__   p10
H2A.GW.ALI          GGAGGACTTACTGCAC-T-GA........G---GGAG--ACA--ACACA-AG---TGACAG-GGAC--GCTGCA-----AT--T---------A--A----AG-A------GCCT-C--T-A---C--G-   2685

H2A.SN.HIV2ST       AGAGGACTTCCTGCAG-T-GA........GA-AC-AG--ACA--ATGCAGAG--ACGACAG-GGAC--GCTGCA-----AT--T---------A--A----AG-A------GC-C-TG-T-A---C--G-   2137
H2A.DE.BEN          GGAGGACTTGCTGCAC-T-GA........G----GAG--ACA--TCACAGAG----GACAG-GGAC--GCTGCA-----AT--T---------A--A----AG-A------GCGT-C--C-A--AT--G-   2692
H2B.GH.D205         GATGGGGAGACA-CAGAGAGA........GAGCCGAG--AGA--CTACA-GG---TGACAG-GGA---GCTGCA-----AT--T---------AG-A----AG-A----A-GC-TGT--C-A---T--GT   2668
H2B.CI.EHO          ACTAGGGAAGAAGCAGAAGGA........GA-C-GAG--AGA--CTACA-GG---TGACAG-GGA---GCTGCA-----AT--T---------AG-A----AG-A----A-GC--CT--T-A---T---T   2663
H2G.CI.ABT96        .................................--GAA-CA-A--CTACA-GG---TGACAG-RGAH--GCTGCA------C--T---------AG------AG-A-----TGC-T-C----AAA-T--G-   2019
MAC.US.MM251        GTTGGGCAAGCAGCAGAGAGA........AAGC-GAG--A----TTACA-GG---TGACAG-GGA---GCTGCA-----AT--T---------AGGA----AG-A-----TGCTC-T--T-AA--A--G-   2585
SMM.US.SIVSMMH9     GGTGGGCAGAAGGCAGAGGGA........GA-C-GAG--AGA--TTACA-GG---T-AC-G-GGA---GCTGCA-----AT--T---------AG-A----AG-A-----TGCCT-C--T-AA-AA--G-   2082
STM.US.STM          GCTGGGCAAGAAGCAGAGAGA........GAGC-G-A--ACA--CTACA-GG---TGACAG-GGA---GCTGCA-----AT--T---------AG-A----AG-A------GCCC-T--T-A---T--G-   2256
TANTALUS.UG.TAN1    GGAGCAGTATGCC-AAAAGGG........AG-TC-ATT-AGAAA-CAGA-GG--A-GGAACTGGAGGACTA----T-G--T--C---------AGGA----AA-GAGT-AG--CT-C---CA---C--G-   2497
GRIVET.ET.GRI-677   GCGCCTGAGAGAGGAGAGAGA........A----CAAG-A-A-A--AG--G--A---G--G-GGA-G-TT-C-TGAG-TC...C---------AGGA----AA-GAAACG-G-C-TC----AA--AACG-   2292
VERVET.KE.VER-9063  ACAGTTGAGGGA-CAAGGGAA........GAGA-CA-CACCAA-A-ATCCAG-CTGGGC-G-AGGA-ATT---TGAA-TC...C---------AG-A----AA-AAAA---G-C-TT--T-AA--AGTC-   2525
VERVET.KE.AGMTYO    ACAACTGAGGGAGCAAAAGAG........GA-TCCA-C--CAAT--ATCCGG-TTGGACCG-GGGA-ATT---TGAA-TC...C---------AG-A----AA-AAAG---G-GT-T----AA---GTC-   2015
VERVET.KE.AGM155    ACAAATGAGAAATCAGAA-AG........AA-CCCC-CA-C-AAC-ATCCAG-TTGGAACG-GGGA-ATT---TGAA-TC...C---------AG-A----AA-AAGG--CTGT-TA-----A--AACTG   2517
VERVET.DE.AGM3      GCAGTTGAGGGA-CAAAGGAA........AA-ACCA-CA-CAGT--ATCC-G-TTGGACAG-GGGA-ATT---TGAA-TC...C---------AG-A----AA-AAAA---G-TT-C----AA---GTC-   2020
SABAEUS.SN.SAB1C    ACGGCAGAAACAGCAGTGG-A........AA-CC-CTC-CCA-AACAGAG-CCCT-CGA-G-GG---AC-GC-----------T---------AG-A----AA-GAAA--TG-GT-C----A---A--GA   2705
SUN.GA.SIVSUN       ACTGTTAGGACAGGGC-T-AG........AGCAC-----A--AG--AG--A-G--GGGAA-G--AGGAGC-A-G-T-G-ATATG------CCAG-A-----T-A-AG-AGG-CTTC--TAAT-------   2652
MANDRIL.GA.MNDGB1   TCAGAAATTAGGG-GAGGG-T........CAG-C--A--AT-AA-GAG--G---AG-GA-G-GGA---TCA------G--TA-T------CAAG-A----AA-A-AAGAGG-CTCAG-G-AT--TGTCA   1982
LHOEST.KE.hoest     TTTAGGACAACAGCTGAAAAG........G---C-----C-AGTCCC-C-G--ATGTG--G-CGAGCCCTGCCTGAATTCACTCT-...-CCAG------AA-A-AA-GT---TT-G-CAAT-------   2647
LHOEST.CD.hoest447  CCTCGGACAACAGCTGAGGAG........G--AC----...-ATCCC-C-G--ATGTG--G-CGAGCC-TGCCTGAGTTTC-TCT-----CCAG------AA-A-AA-AC--TCT-G-CAAT-----G-   1563
LHOEST.CD.hoest485  CCTCGGACAACA-CTGAGGAA........G--AC----...-GTTCC-C-G--ATGTG--G-CGAGCC-TGCCTGAGTTTC-TCT-----CCAG------AA-A-AA-AC--TCT-G-CAATA----G-   1563
LHOEST.CD.hoest524  CCTAGGACAACA-CTGAGGAA........G--AC-A--...AGTTCC-C-G---TGTG--G-CGAGCCTTGCCTGAATTTC-TCT-----CCAG------AA-A-AG-A-G--CT-G-CAAT------G   1560
RCM.NG.SIV-NGM      CAAACAGGAGCAG-GGAAAGA........GA-C---G--A-AGT-GGGAGAGCCT-CGA-G--G-G--GT-C-------AT--T---------A-G-----TG-AACATGGG--G-A----A-------A   2656
COL.__.SIV-CGU1     .........................................CAG--GG--G----GGGA-AG--TC-ATC-C---T-G-A---T-----C--AG------AA-A--T-T-G--G-AG---AA--A---A   2076

SYKES.KE.SYK173     CAACAGGGAAGTGCTCT-G--........GA--G--AGCAGTGG--AG--GG--ACCAA-AG--TC-ATC-C--T-----C--T---------AG-A----AA-GA-AGA-G--G-TG---AA--TG-C-   2483
SYKES.KE.SYK173 Gag __N__R__E__V__L__S__P__.__.__.__K__E__S__S__G__K__E__E__T__K__S__L__Y__P__S__L__S__S__L__F__G__E__D__Q__$                         p6
SYKES.KE.SYK173 Pol _Q__Q__G__S__A__L__A__.__.__.__E__G__E__Q__W__E__G__G__D__Q__E__S__L__S__L__S__Q__L__S__L__W__R__R__P__M__I__E__V__D__V__E__G__D__   p10
```

```
B.FR.HXB2 Pol       L__K__E__A__L__L__D__T__G__A__D__D__T__V__L__E__E__M__S__L__.__.__.__.__P__G__R__W__K__P__K__M__I__G__G__I__G__G__F__I__K__V__R__Q    p10
B.FR.HXB2           TAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTTTG............CCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACA   2425

A.UG.U455           -G-TA-----------------------------------C--------C--A-A----............-------A---------------A-----------------------------------   1871
C.ET.ETH2220        -------G-----C-----C--------------------------------A-A----............-------A---------------------A-----------------T-----------   1817
D.CD.84ZR085        ----A--------------------------------------------C--A-A----............-------AG--------------------------------------------------   1947
F1.BE.VI850         -------G-----------------------------------------C--A-A----............-------A-----------------------------------------------A---   1762
G.SE.SE6165         ---TA-----C--------C--------------C-----------------A-A---A............-----------------------------------------------------------   1822
H.CF.90CF056        ---G---------------------------------------------G--A-A----............--G----A---------------------------------------------------   1772
J.SE.SE7887         -G-G---------------------------------------------C--AGAC---............---C---A---------------------------------------------------   1739
K.CM.MP535C         ---GA-----------------------------------------------A-A----............-------A---------------------------------------------------   1621
01_AE.TH.CM240      -G--A--------------------------------------------T--A-A----............-------A------------------------------------------G----AG--   1999
02_AG.NG.IBNG       ---TA-----C--------C-----------------------------C--A-A---A............-------A---------------------------------------------------   1950
03_AB.RU.KAL153-2   ----A--------------------------------------------C--A-A----............-------A------------------------G-----------------G--------   2470
04_cpx.CY.94CY032-3 ---G---G-----T--------------------------------------A-A----............-------A------G--------------------C-----------------------   1791
N.CM.YBF30          ---GA--------T--------------T------------A-------GC-ACAA--A............GAG----A---------------------A-----------A-----------G-----   2019
O.CM.MVP5180        --TGT--G---T--C-G--------G------------------A-TA-C--ACAA--A............GA-----------C------------------T--A-----C-----A-------A-G-   2455
O.BE.ANT70          --TGT----T-T-GC-G-----------------C------C--A-CA-C--ACAA---............GA-----A------------------------T--A-----------A-------A-G-   2479
CPZ.US.CPZUS        --GTA--------T-----C--------T------------C----CA-C--ACAAA-T............GA---G-C-----G------------G--------------------A-------A---   2485
CPZ.GA.CPZGAB       --TGT------T-GC-------------T------------A----GAG---ACAA--A............-A----CTT--------------------A-----------------------C-A---   2486
CPZ.CM.CAM3         ---TA--------GC-------------T------------A----TA-TG-ACAAC-A............A------A-----G---------------------------------------T-A---   1836
CPZ.CM.CAM5         ---TA--------TC-G--------G--T-----C------A----TA-TG-ACAA---............A------AG----G-----G---------------------------T-----T-A---   2127
CPZ.CD.CPZANT       A-TGTC-G--------------------T-----C------G----G-G---TCA----............-A-----AC-----G-------CA--T--T--------G--------TTCC---CA---   1864

H2A.GW.ALI Pol      P__V__E__V__L__L__D__T__G__A__D__D__S__I__V__E__G__I__E__L__.__.__.__.__G__S__N__Y__T__P__K__I__V__G__G__I__G__G__F__I__N__V__K__E    p10
H2A.GW.ALI          C-GT-----T-T--C----C-----G--T--C--CT--A--G---C--G---AGAG--A............GGGA-C-AC-AT-CC--------AG----------A--G--A--C--A--TACC-A-G-   2803

H2A.SN.HIV2ST       C-GTA----T-T-GC----C-----G--T--C--CT--A--G---C--GCG-AGAG--A............GGGA-C-AT-AT-GT-----G--AG----------A--G--A--C--A--TACC-A-G-   2255
H2A.DE.BEN          CGGTA----TCT--C----C-----G--T-----CT--A--G---C--G---AGAA--A............GGG-AC-AT-AC-CT--------AG----------A--G--A-----A--CACC-A-G-   2810
H2B.GH.D205         C-GTA----TAT--C----C-------TT--C--CT--A--G---C--GG--AGAA--A............GGTA-C-AT-AC-CC--------AG----A--G--A-----G--C--A--TACC-A-G-   2786
H2B.CI.EHO          C-GTA----TAT--C----C---------T-----CT--A--G---C--GG--AGAA--A............GGCA-C-AT-AC-CC--------AG----T--G--A-----A-----A--TACC-ATG-   2781
H2G.CI.ABT96        C-GTA----TA--------C-----G--T-----CT-TA--G---C--G---AGAA--A............GG-A-T-AT-AT-GT--------AG----A-----A-----A-----T--TACC-A-G-   2137
MAC.US.MM251        CTGTA----TAT----G--------G--T------T-TA-TG--AC--G---AGAG--A............GGTCC-CAT-AT-CC--------AG----A-----A-----------T--TACT-A-G-   2703
SMM.US.SIVSMMH9     CC-TA----TAT-------------G--T------T--A-TG---C--GG--AGAA---............GGTCC--AT-AT-CC--T-----AG----A-----A-----C--C--T--CACC-A-G-   2200
STM.US.STM          CTGTA----TA-----G--C-----G--T--C--CT--A--G-G-CG-GG--AGAGC-A............GGGCT-CA--AC-CC-----GG-AG----T-----A-----A-----A--TACT-AGG-   2374
TANTALUS.UG.TAN1    C-GTAC---T---G--------------T--------CA--A--A-G---GAAGA-A-C......CATTTAGA----C-C---CGG--T---G-AG----A--T--A-----GGGA---TC-----A---   2621
GRIVET.ET.GRI-677   C-GT-C----CT-G--------------------C--TA--A-TC------A-GAC---CACTTTCCCCCA-ATAA-CC----CGTT-C--GG-AG----A--T--A-----AGGG--TC-T--C-A-G-   2422
VERVET.KE.VER-9063  CTGTAA---TC--G-----------G--------C--CA-CA--A------CAGACA-A......CAACTAA-----CCT----G-------G-A-----A-----------AGGAC-T--T----A-G-   2649
VERVET.KE.AGMTYO    CC-TTA-G--A--GC----C-----G--------C--CA--A-TA------ATGA---A......CAATTAT----TCC-----G---C-----T--------C--A-----AGGCC-T--T----A-G-   2139
VERVET.KE.AGM155    CCGTTA-G--AT-------------G--------C--TA--A--A-G--T-CAGA---A......CAATTAAGG---TC-----G---------AG----A--------G--AGGGT-A--C----A-G-   2641
VERVET.DE.AGM3      CC-TCAG---AT----------G--G-----------CA-TA--A-----GCAGA---A......CAATTAT------C---------------A--------C--------GGGAC----T----A-G-   2144
SABAEUS.SN.SAB1C    A-GTCACT-----------C--------------CT-----A--C-G-GT--AGAA--A............GG--AC-AT----------G---A--T--A-----A-----A-G---T--T----A-GC   2823
SUN.GA.SIVSUN       CTGT-ACT---T----G--C--G-----------C---A--G-GTCT---GATTC-G--......TCCATAGA---TCAC-----G--C-GGG-AG-G--T--T--A--G--ACAA--T-G-----A-G-   2776
MANDRIL.GA.MNDGB1   CT-TAAG-------C-------------T--------CA-C--TA-T----GA-A-A-A......AAATTAAA-----AT---C-G--------T-----------A--T--AAACT-A-G-----A---   2106
LHOEST.KE.hoest     C-GT-TC---A--G--G--C--T--G--T-----C---A-T--TTC-----ATTCAG-A......AGAATAGA----CCC-AT-CT----G-G-AG-------C--A--G--GCAA--T-GG----A-G-   2771
LHOEST.CD.hoest447  C-GT-TCT-----TC-T--C--T--G-----------TA-C--TTC-----ATTCAG-T......AGAATAGAG---CCT-AT-C---T-G-G-AG----------A--G--ACAA--T-G-----A-G-   1687
LHOEST.CD.hoest485  C-GT-TCT-----TC-T--C--TA-G-----------CA-C--TTC-----ATTCAG-T......AAAATAGAG---CCT-AT-C---T-G-G-AG----------A--G--ACAA--T-G-----A-G-   1687
LHOEST.CD.hoest524  A-GTAACT--C-----G--C--T---------------A-C--TTCT----ATTCAG-T......AGAATAGA----CCT-AT-CT----G-G-AG-------T--A--G--ACAA--T-------A-G-   1684
RCM.NG.SIV-NGM      AGGT---G--CT---------------------G------TA--C--A-TT-A............GATTTAGG--A--AT-----G--------T-----A-----A--G--A-A---A--T--T-A---   2774
COL.__.SIV-CGU1     AGTGCC----G--G--------------------------T--TA-----GAAGA-G-T......CAATTAA----CGCT----C-G--GT-CAA--TCA------A--G--AGCA----G-----A---   2200

SYKES.KE.SYK173     -GGT-C--ATGT--G----------G-----------TA--A-TAG----GAAGA-A-T......CAATTA-ACCA-CC----TC---------AG----A---T-A-----AAAC--T-C---T--G--   2607
SYKES.KE.SYK173 Pol L__V__Q__M__L__V__D__T__G__A__D__D__T__I__I__R__E__E__D__I__.__.__Q__L__H__Q__P__W__S__P__K__I__V__G__G__L__G__G__N__I__T__V__R__Q    p10
```

```
B.FR.HXB2 Pol       __Y__D__Q__I__L__I__E__I__C__G__H__K__A__I__.__.__.__.__.__G__T__V__L__V__G__P__T__P__V__N__I__I__G__R__N__L__L__T__Q__I__G__C__T_   p10
B.FR.HXB2           GTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATA...............GGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACT  2540

A.UG.U455           ---------------T--------T------A-A--GA-----...............----------G-----------------------------G---A-------------------T---  1986
C.ET.ETH2220        ---------A---A---------T------A-A--G------...............---------C-----------------------------C-----CA-----------C----AC----A  1932
D.CD.84ZR085        ---------C--------------------------G------...............--------------------C-------------------------T-------C---------C------  2062
F1.BE.VI850         -------A-C-------------T--------C--G------...............--------G-------------G-------------------------A-------------------T---  1877
G.SE.SE6165         A--------AG---CT--------TA-----A-A--G------...............--G--GA-------------------A-------------G-----A-----------------------  1937
H.CF.90CF056        ------G--AG--GC-----------------A-A--G------...............--------------------------------T-----------G---A-A---------A-----------C  1887
J.SE.SE7887         ----A-CG--G---CG--------TGAG---A-A--G------...............--------G---A-------------------------------G--CA-----------C-------T---  1854
K.CM.MP535C         A--------AG----T--------T--------A--G------...............-----------------------------------------C----C-------------------T---  1736
01_AE.TH.CM240      A--------------T-----------------A-A--G------...............-----------------------------------------C-----A-------------------T---  2114
02_AG.NG.IBNG       A--------A-----T--------T------A-A--G--C---...............-----------------------------------------C-----A-------------------T---  2065
03_AB.RU.KAL153-2   ---------------T--------T------A-A--G------...............-----G-----------------C------------------------A-----------C-------T---  2585
04_cpx.CY.94CY032-3 A-------------CT--------T------A-A--G--C---...............--C-----G---------------------------------C----CA-----------C-------T---  1906
N.CM.YBF30          A------A-T---ACAG----C--ACAG---AGA-----AG-T...............-------------------A--------T--T--T--A-----------T--A--C---------C--T---  2134
O.CM.MVP5180        ----A-CA-TG-GACAG-----G-ACAA---A-GG---TACAG...............--A--------G--G--------T-----T--T--TC---G-----CA-A-----AGGAT-A--A--T--A  2570
O.BE.ANT70          A------A-TG-GACAG-------AGAA---AGGG-G-TACAG...............--A--------G--G--------T-----T--T--T-------------A-A-----AGGAT-A-----T--A  2594
CPZ.US.CPZUS        A---------CG-CAAT--------AGAG---AGA-----ACAG...............---T----T-----G-----A-----A--A--T--T-----C--G---A-T--A--A--A-----A--T---  2600
CPZ.GA.CPZGAB       A-T----A-TG---A------G--AGAA--GAGA----TAG--...............--G-------------------A--------A--T-----------------A-T-----A--AT-G-----T---  2601
CPZ.CM.CAM3         A-----CA-C---GCA--------TGAG---AGA---A-A-C-...............--C--G-----G--G-----A-----A--G--T--------C------A-A-----A--A-----C--T---  1951
CPZ.CM.CAM5         A------A-C---G-A--------AGAG--CAGA---A-A-C-...............--C-------------------A-----A--A--T--------C------A-A--A--A--A---------T---  2242
CPZ.CD.CPZANT       ----A-CA-AG---C----C----TG-A-ACAGG-C--TAC--...............-CA------C-GT-G-----A-AC--A--A-----T--A--T------G-T--ATG--TTT-A--A--T---  1979

H2A.GW.ALI Pol      __Y__E__D__V__E__I__K__V__L__N__K__R__V__K__.__.__.__.__.__A__T__I__M__V__G__D__T__P__I__N__I__F__G__R__N__I__L__T__A__L__G__M__S_   p10
H2A.GW.ALI          A-----AG-TG--GAA---A--G-ACTAAATA-A-G--TA-A-...............-CC--CA--A-GAC---TGAC--C--AA----T--TT----C-----CA-T-----AGCCT-A--CATGT-A  2918

H2A.SN.HIV2ST       A---A-AA-TG--GAA---AG-G-A-TAAATA-A-G--TA-G-...............-CC--CA--A-GAC---TGA---C--AA-------TT----C-----CA-TC----AGCCT-A--CATGT-A  2370
H2A.DE.BEN          A--CA-AA-TG--GAA---A--G-ACTAAATA-A-G--TA-G-...............-CC--CA--A-GAC----GA---C--AA-------CT----C------A-TC----AGCCT-A--CATGT-A  2925
H2B.GH.D205         A--CA-AG-TG--GAA------G-AGTG---A-A-G--TA-GG...............-CA--TA--A-GAC----GA---C--AA-A-----TT----C------A-T--A-A-ACCT-G--CATG---  2901
H2B.CI.EHO          A--CA-AA-TG--GAA------G-AGTA---A-A-G--TA-G-...............-CA------A-GAC---GGAC--C--AA-A-----TT----C------A-T--A-A-AGCT-A--CATG---  2896
H2G.CI.ABT96        A--CA-AA-TG--GAA--TA--G-GGTA---A-A-G-ATACAG...............-C---T--TA-GAC---GGA---C---A-A--T--CT-------G--CA-T--AGT-A--C-A--CATGT--  2252
MAC.US.MM251        A--CA-AA-TG--AAA------G-T-TA--CA-A-GGAT--A-...............--G---A-CA-GAC---GGAC--T--GA-T-----TT----T--G---T--C-A--AGCTC-G--GATGT--  2818
SMM.US.SIVSMMH9     A---A-AG-TG--AAA---A--G-A-TA--CA-AGT-AT--AG...............--A---A-TA-GAC----GA---C--AA-T-----TT----C---------C-A--AGCT--G--CATGT--  2315
STM.US.STM          A-T-A-GA-TG--AA------GG-A-TA--CA-G---ATA-A-...............--A--TA-TA-GACT---GAC--C--CA-T-----TT----C--G--C--A-----AGCTT-G--GATGT-A  2489
TANTALUS.UG.TAN1    A-----CA-T--TT-AG-GA-T--AGA----A-AG--A---G-...............----AT---C--A-T-----A-----AA-A--T--TG-A---------A-TC-AG-A-CTT-A--AGCA-GA  2736
GRIVET.ET.GRI-677   A---C-GGG-G---AAG--C--T-GGAG-ATA-A-TCATC-CC...............--CT--A-TC--A-----AG------AA----T--T--A---------A-T--AG-----GCA--CATG-AA  2537
VERVET.KE.VER-9063  ----AG-A-TG--GAAG--A----AGAA-ATA-A-TTTTG-G-...............--C-----TC-TT-----G-A--C---A----T---------------T-AC-AG-A-CAGCA--AGCT-GA  2764
VERVET.KE.AGMTYO    A---A-CG-C-GGGAAG--A----AGAA-ATA-A-TTTTG-G-...............--A---A----GT-----G-A--T--CA-T--T-----A--T-------T--C--G-C-C-GCA-TGCC-CGG  2254
VERVET.KE.AGM155    A------A-TG--GAAG--C--T-GGAA-ACA-G-T-TTA-G-...............--A-----CC-CA-----G-A--T--CA----T--C--A--------CT-T--AG-C---GCA--AGC--AA  2756
VERVET.DE.AGM3      ----AG-G-T-GGGAAG--AG-T-GGAA-ACA-A-TTTTG-G-...............--G--CA----GA-----AGC--T--CA-A------------------A-A--AG-A-CAGCA--AGC--AA  2259
SABAEUS.SN.SAB1C    A--CC-CA-TCAGGAGG--A----AGAA-ATA-A-CTTG--A-...............-C---CA-T--G------GAA-----A--T--T--T--A---------G-C--AG-A--AT-A--AGTT---  2938
SUN.GA.SIVSUN       A----CAG-TG-CT-TG-------TA-G-ATA-A-C-----C-...............--A-----G--G--G--G--A-----A--AG-T--C--A--G------A-TC-TT-AGCCT-A--GG-A-GA  2891
MANDRIL.GA.MNDGB1   -------A-TG--TATG-------AA-A--GA-GGG-A-AT-T...............--G--------GA----------T--AA-AG-T-----A--G-----CA-AA--GAAA-AT-A--AG-A-AA  2221
LHOEST.KE.hoest     A---AGAG-TG-CT-T--------AGCA---A-A-CTA-AG--...............--G-----C---T----------C--A--AG-T--TG-A--G--G---A-T--A---GCA--G--GGC--AA  2886
LHOEST.CD.hoest447  A--CA-GG-TG-CT-T--------AGCA---A-A-C-A-AG--...............--G---------T-G-----------A--AG----TG-A---------A-T--A---A-TG-G--AGCA-GA  1802
LHOEST.CD.hoest485  A--CA-GG-TG-CT-T--------AGCA--GA-G-C-A-AG--...............--G------C--T-------------A--AG----TG-A---------A-T--A---AGTG-G--AGCA-AA  1802
LHOEST.CD.hoest524  A-TCAGGG-TG-TT-T---C----AGCA--GA-G-CCA-AG--...............--G-----T--GC-------A--C--A--AG-T---G-A--------CA-T--A-AC-CTC-A--GGCT-GA  1799
RCM.NG.SIV-NGM      A-T-A-C--TTGTAAA---ACT--AGCA--GA-G-CTA-ACAT...............-C-CAT-----G---------A-----A--A--T--------C--G---G------GCA-AT-A--AGC-T--  2889
COL.__.SIV-CGU1     ----AGAG-T-A-A-TC--A-G--GG-A---A-AG--TAC-G-...............--AGATA-C---A------AC--C--CA-A--T---G-A--T------A----A-AAG-AGGGAAA-TAGTA  2315

SYKES.KE.SYK173     A---AGAA-T--TAGAT-CACTG-TGT-AA--CCTC--GA-A-AGAAAGCAAGTAGAG--G--TC-T-----G--T-----G--A--A--T---T-A-----G---A-AC-T--CA-AT-G--AGT--AA  2737
SYKES.KE.SYK173 Pol __Y__R__N__I__R__F__T__V__V__K__P__S__G__K__R__K__Q__V__E__G__T__L__L__V__G__P__T__P__V__N__I__L__G__R__N__I__L__T__K__L__G__V__K_   p10
```

```
                     Pol p10 protease end \/ Pol p51 and p66 reverse transcriptase start
B.FR.HXB2 Pol       _L__N__F__P__I__S__P__I__E__T__V__.__.__P__V__K__L__K__P__G__M__D__G__P__K__V__K__Q__W__P__L__T__E__E__K__I__K__A__L__V__E__I__C__   p51
B.FR.HXB2           TTAAATTTTCCCATTAGCCCTATTGAGACTGTA......CCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTA   2664

A.UG.U455           -----------A-----T--------A------......--------------A----A---------------G--------------------------G---------------AC-----------   2110
C.ET.ETH2220        -----C-----A-----T--C-----A------......--------------------------------------C--------------------------G------------AC--C-------G   2056
D.CD.84ZR085        -----C-----A-----T--------A------......--------------------------------------------------G---------------------------AC-----------   2186
F1.BE.VI850         -----------AG----T--------A------......-----------G-----------------------G-----------------------------G------------AC------A---C   2001
G.SE.SE6165         -----------A-----T--------A------......----------------------------------GG------------------------------------------AC-----------   2061
H.CF.90CF056        -----------A-----T--------A------......-----------------------------------G------------------------------------------ACG----------   2011
J.SE.SE7887         -----------A-----T--------A------......------------------------------------A-----------------------------------------AC-C--------G   1978
K.CM.MP535C         -----------A-----T--------A------......-----------------------------------G-----------------A------------------------AC-----------   1860
01_AE.TH.CM240      --------C--A-----T--------C------......-------C---------------------A-----G--------G---------------------------------AC-----------   2238
02_AG.NG.IBNG       -----------A-----T--------A-----G......-----------------------------------G------------------------------------------AC---C-------   2189
03_AB.RU.KAL153-2   -----------A--A--T--------A------......-------C---------------------------G-----------------A--------G---------------AC---C-------   2709
04_cpx.CY.94CY032-3 -----------A-----T--------A------......-----------------------------------G--------------------------------------C---AC---G--A----   2030
N.CM.YBF30          -----------A--A--T--------A------......--------------A--------------------G--A-----------T------AC----------G-G------AG-----------   2258
O.CM.MVP5180        C-------C--T--A--T--C--A-CCC-A--G......--------GC----A--------------A--------A-----------CC-AT-TAG---G------G-----C--ACT-C---A---C   2694
O.BE.ANT70          C----C--C--T--A-----C--A-CCC-A--G......---------C----A--------------A--------A-----------CC-AT-TA-----------G----C--GAC--C---A---C   2718
CPZ.US.CPZUS        -----------T--C---T-----A-A------......-------G------A-----------------T-G---A--G----------------C---------T--G------AC---------CC   2724
CPZ.GA.CPZGAB       C--GTG--C--A-----TT-A-----A-----C......-----C-----G--A--------------T-----G--A--G--------C--AT---C---------T-----C---AC----------C   2725
CPZ.CM.CAM3         -----------T--A--T--C-----A--A--G......--------------------------------T-G---A-----------T--A----C---------T---------AC----------C   2075
CPZ.CM.CAM5         C----------T--A--T--CG-------A--G......----------------------------------G---A-----------T--A----C---------C--G--C---AC---------CC   2366
CPZ.CD.CPZANT       -----------A-----TAAAG----A--A--G......--C--T--------AGA------------A----G---A-----------GC-CT--A----G------G-G--CC--AA----------G   2103

H2A.GW.ALI Pol      _L__N__L__P__V__A__K__I__E__P__I__.__.__E__V__R__L__K__P__G__K__D__G__P__K__L__R__Q__W__P__L__T__K__E__K__I__E__A__L__K__E__I__C__   p51
H2A.GW.ALI          -----CC-A--AG--GC-AAG--A---C-AA--......GAG----G--------------AA--C--G------T-A-G---------C--A---A-----------G-G---C--AA------C---G   3042

H2A.SN.HIV2ST       ------C-A--AG-CGC-AAG--A--AC-AA--......AA-A---TGC-G----------A------A------C-G-G---------C--A---A-----------G-G---C--AA---G--C---G   2494
H2A.DE.BEN          --------A--AG--GC-AAG--A---C-AA--......AA-----C---G--------G-AA-----A----GGC-G-----------CC-A---A----G------G-----C--AA---G--C---G   3049
H2B.GH.D205         --------C--AG-GGCAAAGG-A--AC-A---......AA---TG-G-----A--T----AA-----G-----GA-C-G---------TC-AT-CAGG-----G---CT---CC-CAA------C---G   3025
H2B.CI.EHO          C-------C--AG-AGCAAGG--A--AC-A---......AA---CC-G--------T-A--AA-----G------A-C-G---------CC-AT-CA----G------CT---CC-CAA------C---G   3020
H2G.CI.ABT96        C----------A---GCAAAGG-A---C-A---......AA-----GGC----A--T--C--------G------A-C-GG--------CC-AT-TA-G---------C-----C-CAGG---------G   2376
MAC.US.MM251        C-----C-------AGCTAAGG-A---C-----......AA---C-CC-------------A--T---A------T-G--G--G--------AT--A-------G---GTT------AG------C---G   2942
SMM.US.SIVSMMH9     C-----C-C-----AGCTAAGG-G--RC--A--......AA-----C-C----A-------AA-----A------T-A-G---G-----GC-AT-GA-------G----TT------AG------C---G   2439
STM.US.STM          -----------AG-AGCTAAAG-A--AC-A---......AAG--G-C------A-------AA-----A------A-A--------------AT--A-------G---G-G---C--AA---G--C---G   2613
TANTALUS.UG.TAN1    ----TAA-GGGACAGCTAAA-GAGA-A-TACCCATAACCAAG---C----G--A--T--G---TCA--A--G---T-A-----------CC-AT-------------------CC-GAC------AAC-C   2866
GRIVET.ET.GRI-677   ---GT-A-GGGAG--CTAT---G-C---T--AGGAAACAAA----C--C----AGA---G-AA-----A--T---T-G-----------C--AT--AG---------TG----T---AC------A--C-   2667
VERVET.KE.VER-9063  C--GTGA-GGGACAATTGT-AGAAC-A-T-CCCATCACC-----C---C-G--AGAT---GCAAGG--A---C-TA-A-----------TC-AT--A-G--G--------G--C---AC---G--A---G   2894
VERVET.KE.AGMTYO    ---GTAA-GGGACAATTAT-AGAAA-A-T-CCTGTCACA--T--C-----G---GA---GGCTCGG--A--CTGT--A-G---------TC-CT-TA----G--G--TG----T---CAG-----A---T   2384
VERVET.KE.AGM155    ---GTGA-GGGGCAATTGT-GCAGACA-TACC-ATCACC--G---CGC------GA---GGCCAGA--A---CG-T-G--G---------C-CT-TA------------T---CC-GCA---------C-   2886
VERVET.DE.AGM3      ---GTAA-GGGTCAACTGT-AGAAC-A-T-CCCATTACC--T--G--------AGA---GGCTAGA--A--TTTCT-A-----------CC-CT-CA----------------C---CAG-----A---G   2389
SABAEUS.SN.SAB1C    C-------GA-ACAA--AGAG--A--AC--A--......AA---CC-C------------CAA-----G----GGA-A-GG--------T---T--A-----------G-G--C---AAG-CC------G   3062
SUN.GA.SIVSUN       ---GT---AG---C-CTTAG-GAAA---TCCC-ATTACAAA---T--G-----AGA---GGCC-C---A--------G-----------CC--T--AGG--G------G---GCC--CAGA--------G   3021
MANDRIL.GA.MNDGB1   ----TA--GG-ACAATTGT--GA-A-A-TACC-ATAACAAA---G--------A------G-A-----A--C-G-A-A-----------T--A-GTA----G------GTT-GTC-TCAGA----A---G   2351
LHOEST.KE.hoest     --G-T---AG-TCAGTTAAG-GA-A-A-T-CCTATTACAAAG---TCT-----A--T--GTGT-----A----G---A--G--G-----TC-AT--A----G------G---GCC--CA--CT------G   3016
LHOEST.CD.hoest447  ---GTAC-AT-ACAATTAAG-GACA-A-TACCCATCACAAAG--C-----G--------CTGT-----T--C-G---A-----G-----T---T-TA-------G---G---GCC--CA--C-------G   1932
LHOEST.CD.hoest485  ---GTAC-AG-ACAATTAAG-GACA-A-TACCTATCACAAAG--C-G---G--------CTGT-----T--C-G---A-----G-----T---T--A-------G---G-G-GCC--CA--C-------G   1932
LHOEST.CD.hoest524  ---GTG--AG-ACAATTAAG-GAGA-A-TACCTATAACAAAG--GC-T--G--------GTGT-----G----G---A-----G------C-AT-CA-----------G---G----CA--C-------G   1929
RCM.NG.SIV-NGM      C----------A--A--TAAAGCA--A--AA--......AAG---G--------------CAA-----G--T-G---A-----------T----GTA----G-----TG----T---AC---G-------   3013
COL.__.SIV-CGU1     G--GC-CAATTATCAGATAGG--CCCAGTCACC......AAG--G-CG-----AGAG--C--------G--------A--G--G-----C--A-GCA-----------G---GT---CA-A-G--A---G   2439

SYKES.KE.SYK173     ---GTAA-GGTTCAA-CAG-AT-A---C-A--C......AA-----GTC-G--A----AC-AA--ATTG--CCGGC-A--------------AT-G-T---G---T--G-G--CC-CAAG-C---AGTAG   2861
SYKES.KE.SYK173 Pol _L__V__M__V__Q__T__A__L__E__P__V__.__.__K__V__S__L__K__P__D__K__E__L__P__R__L__K__Q__W__P__L__S__V__E__K__L__E__A__L__K__A__I__V__   p51
```

```
B.FR.HXB2 Pol       T__E__M__E__K__E__G__K__I__S__K__I__G__P__E__N__P__Y__N__T__P__V__F__A__I__K__K__K__D__S__T__K__W__R__K__L__V__D__F__R__E__L__N__K    p51
B.FR.HXB2           CAGAGATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAACTTAATAA   2794

A.UG.U455           AT-----------------A--------------------------------------------------T-----------G-----C-----------G-----------------------C-----   2240
C.ET.ETH2220        A---A-----GC-------A----------G---------------C-----T-----------------------A--G--G-----------G-----------------------G-----C-----   2186
D.CD.84ZR085        ----T--------------A----------G-------------------------------A-------------------------------G-----------------------------------   2316
F1.BE.VI850         T---A--------A-----A----------------------------------------------------------------------G---------G----------------A------------   2131
G.SE.SE6165         A---A------G-------A-----------------------------------C------A-------------------G-----------------------G--------------G--C-----   2191
H.CF.90CF056        -------------A-----A-----C----G---A--------G----------GC------A-------------A--G--G--T-----------------------G--------------C-----   2141
J.SE.SE7887         ----A------G-------A----------G-G-------------------T--C-----------------------------------------------------------T--------C-----   2108
K.CM.MP535C         ----A--------A-----A-----------------------G--------T--C--------G--------------------------------------------------T--------C-----   1990
01_AE.TH.CM240      A----------G-------A-----C--------------------------------------------T-----------G-----C--C--------G--------------------G--C-----   2368
02_AG.NG.IBNG       -------------------A-----------------A--------------------------------------------G--T--------------------------------------C-----   2319
03_AB.RU.KAL153-2   AG-----------------A-------------------------------------------------------------------------------------------G------------------   2839
04_cpx.CY.94CY032-3 ----C--------------C--G---------------------------------------A-------T-----------------C-----------------------------------C-----   2160
N.CM.YBF30          ----A--------------A--------T-G---A--------G--------T--C------A-T-----T-----A--G-----T--C-----------------------------G---T-A-----   2388
O.CM.MVP5180        A---A------C-------A-----C----G---A--A-----------T--T-----A--TA-T-----T-----A--G-----T--C-----G-----------G-----C---------T-A-----   2824
O.BE.ANT70          AG--A------C-A-----A----------G---A--A-----------T--T-----A--TA-C-----T-----A--G-----TG-------------------G--------T--G---T-A-----   2848
CPZ.US.CPZUS        A---A--------A-----A--G---A---G-G-A-----A--------T--------A---A-T-----T--C--A--G-----T--C-----------------------C--T------T-A-----   2854
CPZ.GA.CPZGAB       A---A--------------------A--G-----A--A-----------------------TA-T-----A--C--A--G--------------------------------C-----------A-----   2855
CPZ.CM.CAM3         AG--A--------A-----A--G-------G---A-----C-----------T-----A---A-T-----T-----A--G-----T-----A------------C-G-----------G---T-A-----   2205
CPZ.CM.CAM5         AG--A-----------G--A-----------G--A--C--A--G--C--T--T--C--A---A-T-----T-----A--G-----------A------------C-G--------T------T-A--C--   2496
CPZ.CD.CPZANT       ATA--T-A---GCA---AAT--G-----T-G---A-----A--T--C-----------A---A-T-----A-----A--G-------C-T-A-----------GC----T------------T-A-----   2233

H2A.GW.ALI Pol      E__K__T__E__R__E__G__Q__L__E__E__A__P__P__T__N__P__Y__N__T__P__T__F__A__I__K__K__K__D__K__N__K__W__R__M__L__I__D__F__R__E__L__N__K    p51
H2A.GW.ALI          A-A-A-C-----GA-----CC--T-AGAGG-GGCACCT--AACT--C--T--T-----C--CAC------A--------G--G----AA-AC----------TGC--A-------T------T-A--C--   3172

H2A.SN.HIV2ST       AGA-A-------GA--G--CC-GC-AGAGG-GGCACCT--AACT-----T--T-----C--CAC------A--C--------G----AA-AC----------TGC--A-------T--------A--C--   2624
H2A.DE.BEN          A-A-A--------A--G--CC-GC-AGA-G-GGCACCT--AACT-----T--T-----C--CAC------A--T--------G----AG-AC--------G-TGC-GA-------T--------A-----   3179
H2B.GH.D205         A-A-A-----------G--AC--T-AGA-G--GCACCC---ACT-----------C--A--CACC--------------G-----T-AG-AC----------TGC--A----C--T-----GT-A--C--   3155
H2B.CI.EHO          A-A-A--------A--G--AC-GT-AGA-G-GGCGCCT---ACT------------T-G--CACC--C--------A--G-------AA-AC--------G-TGC--A----------------A--C--   3150
H2G.CI.ABT96        ACA-A------C--------C--T-GGA-G--GCACCC--CACC-----G--T--C--A--CACC---A-T--C-----G--G----AA-AC----------TG---A----C--------------C-G   2506
MAC.US.MM251        A-A-------------T--TC-GT-GGAGG--GC-CCC--GACC-----------C--C--CAC---------------------T-AG-AC----------TGC-GA-------T--G-----A----G   3072
SMM.US.SIVSMMH9     A-A-A-----G--A--T--CC-GT-AGAGGR-GC-CCT--AACC--------T--C--C--CAC------T------R-GR-----G-A-A---------G-TGC--A-------T------T-A--C--   2569
STM.US.STM          A-A-A-----G-----T---C-GC-AGA-G--GC-CCT--AAC------T--------C--CACT-----T-----------G----AA-AC--G-----G-TGC-GA----------------A-----   2743
TANTALUS.UG.TAN1    AG-CTC-A-----------A--------T--GG----AGGA--G-----T-----------C--G---TGT--CCG-----------AA--AC-G-----G-TG---A-------T----CT--C-----   2996
GRIVET.ET.GRI-677   A-C-A------G----G--A---T-A--T-GG--A--AGGA--------T--T-----A-----G-----------------G--T-AA--AC---------TGC-T-----------G-----A--C--   2797
VERVET.KE.VER-9063  A-C--T-A-----A-----------CAGC--G--A--AGGA------G-T--T-----C---A-C---TGT--C-GA--G--G----AA---C-------G-TG--------C---------T-A-----   3024
VERVET.KE.AGMTYO    -CC-AT-A--GC-------A-----CAGT-G-G-A--AGGA------G----------C---A-----TG---------G--G----AAT-CC-G-----G-TGC-------C--T-----GT-A-----   2514
VERVET.KE.AGM155    A-ACAT-A--GG-A-----A---T-AAGC-GGG-A---GGA--C---G----------A--------CTGT----G-----------AAT-AC-G-------TGC-G-----------G-----C--C--   3016
VERVET.DE.AGM3      ACC-AT-A--G--A-----A------AGC--G--A--AGGA--G---G-------C--------G---TG-----------------AGT-AC---------TG-----------T--G-----A--C--   2519
SABAEUS.SN.SAB1C    A---CT-A-----AC----AC-TT-GGA--G---A-----A-----------T--C--C-----C-----A---G----G-----T-AA---C-G-------T-C-CA-G--C-----GC-GT-A-----   3192
SUN.GA.SIVSUN       ATAGAT-----GCA-----A--G---AGC-GGGCA-AC--A-GG--C--T--T--C--A---A-T-----T--------G-------AA-A-G----------G---A-------T----T---A---GG   3151
MANDRIL.GA.MNDGB1   ATAGAT-A--GG-------A------AGT-GGG-A-AT--A-G----AAT--------A--TA-C--------------G--G--T-AA-A-G--------------A----C--T------T-A--C--   2481
LHOEST.KE.hoest     ATAG-T-A-----A-----A-----A--TCC-G-G-AT--A-GG--C---------------A-------T--C-----G-------AA-A-G----------G---A-T---------A-G--C---G-   3146
LHOEST.CD.hoest447  ATAGAT-A-----A-----------A------GCA-ATG-A-GG--C-----T--C--A---A----------------G-------AA-A-G-G------------A-T-----T---C----A---GC   2062
LHOEST.CD.hoest485  ATAGAT-A-----A-----------A------GCA-ATG-A-G---C-----T--C--A---A-C--------------G--G----AA-A-G-G---------C-GA-T-----T---C----A---GC   2062
LHOEST.CD.hoest524  ATAGAT-A-----------------A--CCC-G---AT--A-G------T-----C--C---A-C-----T-----------G----AA-A-G----------G--GA-----------A--T-A---G-   2059
RCM.NG.SIV-NGM      AT-CC-----------G--A--------T-G---A--A--A--------T------------A-----TGT--C--A--G-----TTCC--A-----------GC-----------------T-G-----   3143
COL.__.SIV-CGU1     A-AGAT-----GCA--G--A-----AGA----GCA-AATTA-GG--------T-----A---A-----TG---T-G---------T-AA-A-G-------G------A----------G--G--A--C--   2569

SYKES.KE.SYK173     A---C---CTC----C---AC--T-GGAG---GC-A-T---ACC-----T--T-----C----------TA--T-G------G----AA-AG--------G-TGC-CA-T-----T---A--T-A---G-   2991
SYKES.KE.SYK173 Pol E__D__M__L__K__A__G__Q__L__E__K__A__S__P__T__N__P__Y__N__T__P__V__F__V__I__R__K__K__D__K__K__K__W__R__M__L__I__D__F__R__K__L__N__E    p51
```

```
B.FR.HXB2 Pol       __R__T__Q__D__F__W__E__V__Q__L__G__I__P__H__P__A__G__L__K__K__K__K__S__V__T__V__L__D__V__G__D__A__Y__F__S__V__P__L__D__E__D__F__R_   p51
B.FR.HXB2           GAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATGAAGACTTCAGG  2924

A.UG.U455           A--------G--------------------------C--G---A-A--G--TC----------------------------A--------G--C-----------------T---------AG---T--A  2370
C.ET.ETH2220        A--------------T-----------------G--------C--A--------------------------G--------A--------G-----------C--------T----------GT-----A  2316
D.CD.84ZR085        ---------------------------------G-----------T-----A-----G--------G---A-------------------C---------------A--------TG---------T---  2446
F1.BE.VI850         A-----------T--------G-----------------------T--------------------G-----------------------G---------------------------A-G--T----A-  2261
G.SE.SE6165         A--------------------G--C--------------T-----T-----------------------------------A--------G-----------------------------------T--A  2321
H.CF.90CF056        A--------------------------G--------------C--A---------------------------T----------------G-----------------C--T------A----A-----A  2271
J.SE.SE7887         A--------------------------G-----------------A--------------------------C-----------------G-----------------C--T---T-------T------  2238
K.CM.MP535C         A--------------------------------------------A-----A--------------------------------------G-----------------C--T------A-----------  2120
01_AE.TH.CM240      A--------G-----T-----------------------G-----A-----T-----------------------------A--------A--------------------T---------AG---T--A  2498
02_AG.NG.IBNG       ------------T--------G--C--G-----------T--------G--A--------------------------G--A--------G---------------------------A-------T--A  2449
03_AB.RU.KAL153-2   ------------------------------------------C--T-----------------------T------------------------------------------------C----------A  2969
04_cpx.CY.94CY032-3 A--------G-----------------G-----------G--C--A-----------G---------------------T----------G---------------------------CC---G------  2290
N.CM.YBF30          A--G--C-----T--T--------G--GC-------T--------A-----A-----GC-------------G-----TT-------A--A-----T---------TG------G--CA----T--T--A  2518
O.CM.MVP5180        A-----A-----T--------G--G--------T--T--------A-GG--T-----GC-A-GGC----T--T-----CT-A-----A--A-----T-----C---TGC--T------CC------T--A  2954
O.BE.ANT70          ------A-----G--------G--A--GC----T--C--------G-GG--T--G--GC-A--GC----T--T-----CT-A-----A--A-----T-----C---TG---------CCC---T--T--A  2978
CPZ.US.CPZUS        A--G--A--------------------G-----C--------C--A-----A-----G-------G------------C--A-----A--G-----C--------TTG---AC-G---A-G--A--T--A  2984
CPZ.GA.CPZGAB       ------A--------T--------G--G-----C-----T--C--A--------------------------G------T-A-----A--A-----C--C--C--TTG----C-G---A----T------  2985
CPZ.CM.CAM3         A-----A--G-----T--------C--G-----C--------C--G-----T--------------G-----------G--A--------A-----T--------CTGC--A---------A-T--T---  2335
CPZ.CM.CAM5         A-----A--G--A--------G--C--------C--------C--A-----AC----G--A--G--G--------------A--------A-----T-----C--CTGC--A---------A-T--T--A  2626
CPZ.CD.CPZANT       A-----A-----T--T-----GA-A-----------T--T-----A--C--A-----GC-A--G--------G------T----------A-----C-----CA-A---------C-G-----T--A  2363

H2A.GW.ALI Pol      __V__T__Q__D__F__T__E__I__Q__L__G__I__P__H__P__A__G__L__A__K__K__R__R__I__T__V__L__D__V__G__D__A__Y__F__S__I__P__L__H__E__S__F__R_   p51
H2A.GW.ALI          -GT---------T---ACA--GA----G-----G--T--------A-----A---GCC-------G-AG-A-C--T--G--------A--G-----T--C-----CA-A--AC-GC----GAG---T--A  3302

H2A.SN.HIV2ST       -GT-------------ACA---A-C--G--------T-----C--A-----AC--GCC-----G---CG-A-T--T--C--A-----A--G-----T--C-----CA-A--AC--C----G--T--T--A  2754
H2A.DE.BEN          -GTG--------T---ACA---A----GC-------T-----C--G-----AC--GCC--A--G---AGGA-CT-TA--T-A-----A--G-----C--------CA-A--AC--C-------T--T---  3309
H2B.GH.D205         -GT---C--G------ACA-----CA-C-.G--T--T--C--C--G-----TG..GC-G-A----GGAG-A--------A-A--------A-----C--C--CAGTA-C--AC-----CC-A-------A  3282
H2B.CI.EHO          -GT---C-----A--TACA--G--C--GC-G--T--T--T--C--A-----AC-GGC-TCA--G---AG-A----------A-----A--A-----C--C--CAGT--C--AC-----CC---------A  3280
H2G.CI.ABT96        -GT---C--G-----TACT-----C--G--G--C-----T--C--T-----T--GGC-G-A----GGAG-A--------------A-A--G-----C--C--CAGCA----TC------T---T------  2636
MAC.US.MM251        -GTC-----G-----TACA-----C-----------------C--T-----AC--GC---A-GG---AGGA-T------------A-A--------------C--CA-A--TC----------A--T---  3202
SMM.US.SIVSMMH9     -GTC--------T--TACA-----A--G-----G--------C--TR----AC--GC-----G--G-AGGA-C-----GT-------A--------------C--CA-A--T-----------A------  2699
STM.US.STM          AGTG--A-----T--TACA---A----GC----T--------C--T-----AC--GC-----GG-G-AG-A-C------T-------A--G--------------CA-A---C-------G-G---T---  2873
TANTALUS.UG.TAN1    AGC---A--G-------TT--G--A--------T--T--C-----A-----A--G----GA--G---CAGA-C--------A---A-A--A-----C----A-AGTA-A--A--G---CC---A------  3126
GRIVET.ET.GRI-677   AGCT-----------T-TT--G-----GC-G-----T--T--C--A--G--CC-TC-G--A--G--GCA-A-C------A-A--CA-A--G-----C----A----A-A--A---TGCA-G--A-----A  2927
VERVET.KE.VER-9063  AGT---A----------TT-----------G--------T--C--G--------GC-G--A-TG---CAGA--------T-A---A-A--G-----T----A---CA-A--A------CC---G--T--A  3154
VERVET.KE.AGMTYO    -GC---C-----T----TT-----G--------G-----C--C--A-----A----G-----TG-G-CAGA-------TT-A-----A--A--C--C----A---CA-A--A--G---CC-A-T--T---  2644
VERVET.KE.AGM155    AGCT--A----------TT-----C--------T-----C-----A-----------G--A-TG--GCA-A----CA-TA-A--------G----------A-AGCA-A--AC-G---CCT--G--T--A  3146
VERVET.DE.AGM3      AGC---A-----T--T-TC-----A--G-----C-----T-----AT-------CG------TG-CGGA-A--------T-A--CA-A--G-----C----A----A-A--A-----CCC---G--T--A  2649
SABAEUS.SN.SAB1C    A-T------------TCA------C--G-----G-----C--C--A-----C---C-GC-A-GGG-GCAGA-T-----GT----CA-A--A-----C------AGCTGC--T------CC------TCA-  3322
SUN.GA.SIVSUN       T-TG-----------TCAT--GT-A--G-----G--T--T-----A-----A-----G---TGT---CAGA-T------G-A--------G--------C--C---A-A--A--G---CC---T-AT--A  3281
MANDRIL.GA.MNDGB1   -TT---A--G--T--TCAT---T-A--G-----T-----T--C--A-----AA--------TGT---AG-A-------C--A---A-A--G-----C------AGTA-A--TC-G---CC---T-AT--A  2611
LHOEST.KE.hoest     ATT------------TCAT--GC-A--------T--T-----C--A-----TA-C--G--ATGC---CAGA-T------G-A--CA-A--A-----C-----CAGTA-C--T------CC-A-T-A---A  3276
LHOEST.CD.hoest447  TTT---A--------TCAT-----A--------C--T-----C--T-----AA----G--ATGC--GCAG--------GG-A---A-A--A-----C------AGCA-----C-G---CC-----A---A  2192
LHOEST.CD.hoest485  TTT---A-------CTCAT-----A--------C--T--------T-----AA----G--ATGC--GCA-A-------GG-A---A-A--A-----C------AGTA-----C-G---CC-----AT--A  2192
LHOEST.CD.hoest524  -TT---A--------TCAT--GC-A--G--------------C--A-----AA----G--ATGC---CA-A-T------G-A---A-A--A-----C-----CAGCA-A--T------CCT--T-A---A  2189
RCM.NG.SIV-NGM      A-----A-----T----TT--G--A--GC-G-----C--G-----A-G---A------C-ATGC-GGCA-A-T-----CT-A---A-T--G-----G---------TGC---C--T----GCCA--T--A  3273
COL.__.SIV-CGU1     ------A---------AT------C--------G--------C--A-G---TC-T-T-G-A--G-CCCATA--------T-A---A-A--A-----C-----C---A-------G---CCG--T-ATCAA  2699

SYKES.KE.SYK173     -GCC--A-----T--T-TC--G--A-----G--------C--C--A-G---AC-----CG-C-----...C-----A-TA-A---T-AAAA----------AC-----C---C-G--CA----G--T---  3118
SYKES.KE.SYK173 Pol __A__T__Q__D__F__F__E__V__Q__L__G__I__P__H__P__G__G__L__K__R__Q__K__.__L__T__I__I__D__L__K__D__A__Y__Y__S__V__P__L__D__K__E__F__R_   p51
```

```
B.FR.HXB2 Pol       _K__Y__T__A__F__T__I__P__S__I__N__N__E__T__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__   p51
B.FR.HXB2           AAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAG  3054

A.UG.U455           -----------G--C--------------------------------AG-C--G--------------------G---------------------T----------G----------------------  2500
C.ET.ETH2220        --A-----------C-------------C---------A--------------------A--T-----C--C------------------------C----------G---------C-CC---------  2446
D.CD.84ZR085        -----C--------------------------------------------------------------A-------------------------------------------------T-----------  2576
F1.BE.VI850         -----C--------C------------G-C--------------T--A-----G--C--------------------A------------------------------T---------------------  2391
G.SE.SE6165         --------------C--T--------------T---------------G-------------------------------------------------------T--G-------------G--------  2451
H.CF.90CF056        --------------C-----------------------------------------------T--------------------------------------------G----------------------  2401
J.SE.SE7887         --A-----------C--T--------------------------------C--------------C-----A--------------------C-----------T--GT--------------------A  2368
K.CM.MP535C         --------------C--T-----------T--------A---------G----------------------A-----A------------------------------CA--------------------  2250
01_AE.TH.CM240      --------------C--------------------------------A--C--------------------G-----------------------G-----------G----------------------  2628
02_AG.NG.IBNG       -----------------T---------G----T-----------------------C--------------------------------G-----G--------T--GGCA-------------------  2579
03_AB.RU.KAL153-2   ----------------------------C------------------------------------------------------------------------T----------------------------  3099
04_cpx.CY.94CY032-3 -----C--------C-------------CC-----------------A--------------------------------C---------------------------T---------------------  2420
N.CM.YBF30          --------A--T--------------------------------T--T--------C-----T--------G-----A--C--------G-----------T--T--G----CA-----------TC---  2648
O.CM.MVP5180        --A--C-----C--C--T--T------G-G-----------C-----AG-A-----C-----------C--C--G--A--G--------T--------C-----T--G---TCA--------G--TC---  3084
O.BE.ANT70          --A--------T--C--T--T------G-G--T--------C-----A--A-----C-----------C--C--G--A-----------T--------T------------TCA-----------TC---  3108
CPZ.US.CPZUS        --A--C--A-----------C------G-------------C-----A--------C-----T------T-A-----A-----------G-----------T-----G--C--T-----------TC---  3114
CPZ.GA.CPZGAB       --A--------------A--T--C--------T--------C-----AG----------A--T-----TT-G-----A-----------G--T---T-T--C--------C--T-----------TC---  3115
CPZ.CM.CAM3         --A-----A--------T-----C---G----T-----A--------A-----------A--T------T-A-----A-----------G-----------T--T-----C-------T------TC--A  2465
CPZ.CM.CAM5         --A--C--A--T-----T-----C---G----T--------------AG-------------T------T-A--R--A-----------G-----G-----C--T-----C--------------A----  2756
CPZ.CD.CPZANT       --A-----A--T--C--A--T--A--CG-G-----------C-----A--A-----TGT---T-----T--A-----A--C-----------------T--T--T---GCA--------C--G-------  2493

H2A.GW.ALI Pol      _Q__Y__T__A__F__T__L__P__S__V__N__N__A__E__P__G__K__R__Y__I__Y__K__V__L__P__Q__G__W__K__G__S__P__A__I__F__Q__H__T__M__R__Q__I__L__   p51
H2A.GW.ALI          C----------------TC----ATCAG---------CAGA------A-AA------ATA--T--A--CT-A--G-----------G--------------T--T---CAC-CA----G-C-G-------  3432

H2A.SN.HIV2ST       C----------------TC----ATCA----------CTGA------A-AA-----CATA--T--A--CTCA--------------G--------------T--T--GTAC-CA----GGC-GG------  2884
H2A.DE.BEN          C-----------------C----AGCAG--------AT-GA------A-AA------ATA--T--A--CT-G-----A--------G--------------T--T---TAC-CA----GGC--G------  3439
H2B.GH.D205         C-------A-----C---T-G--ATCAG---------CAGAG-----A-AG-----CATT--T--A--T--A-----A--G-----G--G--C-AGT----T-GT---TACTC-----G---GG-A----  3412
H2B.CI.EHO          C-A-----A--------TT-G--AGCAG----T----CAGA------A-AG-------TT-----A--C--A--------------G-----C--------T-----GTAC-C-----G---GG-AC---  3410
H2G.CI.ABT96        C----C--A--T--C--TT----ATCAG---------CAGA------A-AA-----CATC-----G--C--A--G--A--------G--C--------T--C------CA--CA----GG--GG-AC---  2766
MAC.US.MM251        C----C-----C-----TT----ATCAG----T----CAGAG-----A-AAC----CATT--T--G--T--G--T-----------G--G--------C--C------TAC-CT----G-C-TG-GC---  3332
SMM.US.SIVSMMH9     C-A--C-----C------T----ATCGG---------CAGA------A-AA-----CATC--T--G--AT-A--T--A--G-----------------T--T--T--RCA--CTR---G---TG-A----  2829
STM.US.STM          C----------T--C---T----ATCAG---------CAGA------A-AA-----CATC-----G--T--G--------------G--------------T--T---TA--CA----G---T--A----  3003
TANTALUS.UG.TAN1    CCC-----A--T--------C--ACAAG----T----TAGGG--T--A-----G-----A-T----TGT--A--T--A-----------------CA-C-----T--GCAC-CAGCACAGCG-G--C--A  3256
GRIVET.ET.GRI-677   --A-----A-----------C--CTCAG----T---ACAGGG--------A--G-------T----TGT--G--T-----------------T--TA----T-----G-A--CGGCAG----C--T----  3057
VERVET.KE.VER-9063  --A--C--A--------A--C---TCAG----T---C-AGG------TG----G-----A-T----TGC--------A--C--------G--T---A-T--T--T--G-A--CAGCAG-------T----  3284
VERVET.KE.AGMTYO    --A--------T-----T--T----CAG-G--T---C--GG---C--------G-----A-T---CTGT--C--G--A--G-----------T--TA----C-------A--CAGCAG--TCC--T--G-  2774
VERVET.KE.AGM155    --A--C--A--T--C-----C----CGG-----------GG------C--A--------A-TT---TGC--A--G-----C-----G-----C--GA----T-------AC-CAGCAT-------TC---  3276
VERVET.DE.AGM3      --------C--T--------T--ATCAG----T---C-AGGG-----T-C-----------T---CTGT--------A--------G-----C---A-T--T--T--G-AC-CAGCAG-TTCC--TC---  2779
SABAEUS.SN.SAB1C    --A-----A-----C-----C--ATCAG-C--T---AG-GAG-----C--A-----------T--G-----C--------G--------C-----CA----T--T--G-CA-CAGCC-AC-----TC-GC  3452
SUN.GA.SIVSUN       --------A-----C--A-----GTCAG----T---C--G-------A-AA------GT---T-----C--A--------C---------AGT--TTGC--C--T--GG-A-CGG-AG-G-GCC------  3411
MANDRIL.GA.MNDGB1   CCC--------C-----GG----ATCAG-T--T---C-AG-------A-AA-----CAT---T-----T-----T--A--G-----G---AGT---TGT--C--T---G-G-CAG-AG--TC-C-GC-G-  2741
LHOEST.KE.hoest     --------A--------T------TC-C-C--T---C-AGAG-----A-AA--------------------G-----A--G-----G--GAGC--TTGT-----T--GG-A-CAG-AG--GG-C-TC-CT  3406
LHOEST.CD.hoest447  C----C--A-----C--AG----ATC---T--T--CC--GA--------GA-----------T-----A--A--C--A--------G--GAGT---TGC--T--T---GC--CAG-AG--GG-C--C--T  2322
LHOEST.CD.hoest485  C----C--A-----C--AG----ATC---T--T--CC--GA--------GA-----------T-----A--A--C--A--------G--GAGT---TGC--T--T---GC--CAG-AG--GG-C--C--T  2322
LHOEST.CD.hoest524  C-A--------G--C--T-----GTCAC--------C--GA------A-AA-----------T------T-G-----A--------G--GAGT--TTGT--T--T---GAA-CAG-T-GTGG-C--C-T-  2319
RCM.NG.SIV-NGM      -----C-----T--C--G----TAGC---T--T---C-AGG---T--C-----------A--T------T-G--C-----------G-----T-----T-----T----CA-CAGCT-AC-----T----  3403
COL.__.SIV-CGU1     -----C--A--T-----A--C--ATCAG-G------TT-G---------C-------GT------A-----A--G-----T--------G--T---A----C------G-A-CAG-AG---G----C---  2829

SYKES.KE.SYK173     CCA--C--G--------AG-G--A--C-----T----CCT-C--G--AGAA--------A-TT-CA---T-A--------------------T--CA-T--C--------C-CA--A-ATC-G--AC--C  3248
SYKES.KE.SYK173 Pol _P__Y__T__A__F__T__V__P__S__I__N__N__A__S__P__G__E__R__Y__Q__F__T__V__L__P__Q__G__W__K__G__S__P__T__I__F__Q__S__T__I__N__Q__I__L__   p51
```

```
B.FR.HXB2 Pol       E__P__F__R__K__Q__N__P__D__I__V__I__Y__Q__Y__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    p51
B.FR.HXB2           AGCCTTTTAGAAAACAAAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTG   3184

A.UG.U455           ----C------TC----C----------------------------------C--------G--------T--------------A------G-------------AT-A---GCT-----A-----C--   2630
C.ET.ETH2220        ----C-----GGCC-CC--C-----A-----------------T--------C--------------------------------A------G-CCC------A---T-A---G-------A--A-A---   2576
D.CD.84ZR085        ----C--------------------AG-------------------------------------------T-----------A---------G-----------A-AT-A---G----------------   2706
F1.BE.VI850         ----C-------TGA----C-----------C-----C--------------------------G------------------------------------------T-A---G-------AC----A--   2521
G.SE.SE6165         ----C------GC-A-T--------A--G--G-----C-----------------A------------------------------------G--------------T-A---G-------A----AA--   2581
H.CF.90CF056        C---C------G----------T--A--G-----T--C-----------------------------------------------A------G--------------T-A---GCT---T------AA--   2531
J.SE.SE7887         -A---------G--AG---C-----A-----------C--G-----------C--------G-----------G-------AA----------G-------A----------GG-------A----A---   2498
K.CM.MP535C         ----C-------T-A----------A--G-----A--C-----------------A--------G--------------------A-C------------------A--A---G-------A----AA--   2380
01_AE.TH.CM240      ----C-------T-A----------A--G----------------A------C-----------------T-----------------C--------------------A---GCT-----A-----C--   2758
02_AG.NG.IBNG       ----C-------C-A----------G-----G-----C-----------------A--------------T---------------------G--------------T-----GG------AC---AA--   2709
03_AB.RU.KAL153-2   -------------------------G--------------------------------------------------G------------------G----------A------G--------C-------   3229
04_cpx.CY.94CY032-3 ----C------TTCA----C-----A-----C--A--C-----T--------------------G--------------------A------G----------A-----A---G-G-----A-----A--   2550
N.CM.YBF30          -A--A--C---G-GA--C-------G---A-C--T--C--G-----------CC-C-----G---------------C---CA--A------GAGGC-G----A--C--T---G-T-----T----A---   2778
O.CM.MVP5180        -T--A---------AGC--C-----AG---AA--T-----G-----A-----C--A-----------A--T---CC-T-G-CAG-A-------AG-GGG-C--ATT---T--GG-----T-A-ATCA---   3214
O.BE.ANT70          -T--A--C--G-G-G-C--C-----AT---AA--T-G---G-----------CC-A-----------A--T---CCCC-GACAG-A-------A--GG--T--ATT---T---G----C--A-ATCA---   3238
CPZ.US.CPZUS        -T---------------C-------TG-TA-A-----------T---------C-C--------G--A--TC--A-CT---AAA-------GGA---GG----ACT---C------T--T--C-T-CT--   3244
CPZ.GA.CPZGAB       -A--C--C---G--A-G-----T-----TAC---T--C--G-----------CC-A-----G--G-----TC-T-----T-AT--A-------A---GG-G--A--A--A---------T--C-T-AA--   3245
CPZ.CM.CAM3         CC--A--C---C-G--GC--------T--A-A--T--C--G--T---------C-T-----G--------T--G---C--A-AG-A------GA---GG-----CTA--C-----G-----AC-T-CT--   2595
CPZ.CM.CAM5         -T--A--C--------GC-------TC--A-A-----C-----T--------CC-T---------------C-G--G-----AA-A------GA---GG----ATTA--C------------C-T-CT--   2886
CPZ.CD.CPZANT       CA--A-----GG-TA-GT------CAG---AA--T------------------C-C--------------TA-G-----TACTGCA------GA--TG-----AA----T---------T-ACA-GTC--   2623

H2A.GW.ALI Pol      E__P__F__R__K__A__N__Q__D__V__I__L__I__Q__Y__M__D__D__I__L__I__A__S__D__R__T__D__L__E__H__D__K__V__V__L__Q__L__K__E__L__L__N__G__L    p51
H2A.GW.ALI          ----A--C-----GGC---C-AG--TG-CA--C--AT----------------A-C-TAA---CTAG----AGGAC-GATTTAG-A---GAC-AGGTGG-CCT-C--T-A-AGG---TC--AAATG-CCT   3562

H2A.SN.HIV2ST       -A--A--C------GC---C--G--T--CA--C--AT---G------------A-C-TGA---CCAGC---AGGAC-GATTTAG-A---GAC-G-GTGG-TCT-C----A-AGG---T---AAATG-CCT   3014
H2A.DE.BEN          -A-----C------GC---C-----TG-CA--C--ATC--G------------A-C-TAA---CTAG----AGGAC-GGTTTAG-----GAC-A-GTGG-CCT-C----A-A-G---T---AAATG-CCT   3569
H2B.GH.D205         -T-----------GGCC--CAGC--TG-CA----AAT---G-----------CA-CCT-A---C-AG----AG-AGTGATCT-G----CGAC-GGGT-G-GTCCC-A--A-A-G-GTTAT-AAATGACAT   3542
H2B.CI.EHO          -C-----C------GCC--CAAT--TG-CAC---AATC--G-----------CA-TCTC--G-C-AG----AGGAGCGATCT-G-----GAC-GGGT-G-GTCTC-A--A-A-G-G-TAT-AAAT-ACAT   3540
H2G.CI.ABT96        -C--C--C------GCC---GAG--TR--A----YATC-----T--------CA-TCT---G-CCAG----AG-ACTGATTTAG-A---GATCA-GT-G-TTT-C----A-AGG---TCT-AAAT-AAAC   2896
MAC.US.MM251        -A--C--C--G--GGC---------TG-GACCT-AGTC--G--T--------CA-C-TAA---CTAG----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC----A-AGG---TCT-AAAT--CAT   3462
SMM.US.SIVSMMH9     -A------------GC---------TG-GACCC-GATC--------------CA-C-TAA---CTAG---TAG-AC-GATTTAG-A---GAC-GGGT-G-TTTAC-AT-A-AGG---T----AAC--TAT   2959
STM.US.STM          ----A--C--G-G-GCC--C-----TG--AC-C-GATC---------------A-C-TGA---CCAG----AG-AC-GATCTAG-----GAC-GGGT-G-TTTAC--T---AGG---TC--AAAT-AT-T   3133
TANTALUS.UG.TAN1    -AGAAA-AG-----A-GGT--A---TT---AA--AAT---G-----------CA---GGA-------A--T--GCC-GAGAAAG-A---GATGA---GG----A---G-A---A-G-T----C-ATTAA-   3386
GRIVET.ET.GRI-677   --GAGA-C-A--GG--C-C---T-GGT---AA--TGTC-----------C-------GGT-G-CG--A---CAT--TGAGACTAGA----ATCA-C-GG----CATAG-A---A-GATG---C-AGAAAA   3187
VERVET.KE.VER-9063  --AAAA-A-A-G-GG--TTAAA-CCAT-GACA--AGTG--G----------------GG-----G---C-AGA---TGA-TACACC---GAC-AGTT-G----AC----T---GTCA-G--AAAA-CA--   3414
VERVET.KE.AGMTYO    --GAGA-A-A--G-A-CTTG----CAC--ACC--TGTA-----------------A-GG-----T---C-AGA-A-TGA-CACACC---GAC-A-TT-G----AC--T-A---AC-A-AT-ACAAGCC--   2904
VERVET.KE.AGM155    -AGAAA-A-AG---G--TTAAA-C-GC-GACG--TGTC--G-----------CC-C-GG--------AC-AGA---GGGTCCAA-----GATCAGCT-G--C-AACA--T--GA-TAGAT--CAAGAA--   3406
VERVET.DE.AGM3      -AGAAA-A-A---GG-GTTAAA-CC-C--ACC--TGTG--------------CC-A-GG-----G---C-GGA---TGA-TACAC----GATCGGTTGG----AC-A--A---ATGA-AT-AAGTGCC--   2909
SABAEUS.SN.SAB1C    --GAA-----GC--A-G--C-----TG---A---A------------------A---TAA---CCAG---TAGGCC--AG-CAG-A---CT-GT--TGG--C--C--T-A---G-CT----AGAA-CT--   3582
SUN.GA.SIVSUN       --AAA-------G---GT-------AG-TCAAT-G--C-----------------ACTCA----------T-AT--G-A-AA-A-----GA-GA--T-G-GA-AC-AT----G--G-TA--CA--GAA--   3541
MANDRIL.GA.MNDGB1   --GTA--------GA-CC-----ACAG--CAGT-A----------------------T------G--A----ATAC-GC--AAG-----GAG-A-GCT----TA--AT-A--GGCTTTAT-AA---CA--   2871
LHOEST.KE.hoest     CAGAG-----G---TT------------GA-C--T--C-----T-----------A-TCA-------A--TAG---G-G-AAAGGA---GATCAGGC-G--A-A--A--C---G---T---TA---CA--   3536
LHOEST.CD.hoest447  CAGAA---------TT---C--T------A-CT-G--------------------A-TCA-------A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-----G-T-TCT-AG---CA--   2452
LHOEST.CD.hoest485  CAGAA---------TT---C--T------A-CC-G--------------------ACTCA----G--A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-A---G-G-TCT-AA---CA--   2452
LHOEST.CD.hoest524  CAGAG--C--G---TT-------------C-CC----C--G----------------TCA----T--A---AGG--G-A--AAGGA-----TCAGGC-G--A-A--AT-A---G---T-T-AA-A-CT--   2449
RCM.NG.SIV-NGM      -A--A-----G---G-G--------TG----C--------G--T-------------T------G--A--TAGGAC-GC-CATG-A-----TCAG-TG--CA-AAG---C---G-G------C---CC--   3533
COL.__.SIV-CGU1     -A--G--C--G--GAG-GGA-A-TTG...CAA--AGC---G--------------A---C-------A--TC-GCCTT---A-G---------GGGC-G--C-------C---G-TGCA--C---TAC-A   2956

SYKES.KE.SYK173     ----A--------GA-GT--T----TC-CAC-C-AATA--G-----------C--A-TGA-T--CA----TAG-TC-GAGAA-GCA---CAGGAG-T-G--C-AC-AA-AGTGAC-GCA--A--A-AAGT   3378
SYKES.KE.SYK173 Pol Q__P__F__R__K__K__Y__S__D__L__T__L__I__Q__Y__M__D__D__L__L__I__G__T__D__R__S__E__K__A__H__Q__E__I__V__Q__Q__I__V__T__A__L__L__K__V    p51
```

```
B.FR.HXB2 Pol       __G__L__T__T__P__D__K__K__H__Q__K__E__P__P__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_   p51
B.FR.HXB2           GGGACTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGC......TGGACT  3308

A.UG.U455           ----T-C-TT--C-----------G--------------------T-----------G--------T--------C-----------T---------CA-------------------......------  2754
C.ET.ETH2220        ----T----------------G-----------------------T-----------G--------T--------C---------------------CA------------G--T---......------  2700
D.CD.84ZR085        ---G-----------------------------------------T--------------------------------G------------T-----ACA--------G-----A---......------  2830
F1.BE.VI850         ----T----T-----------------------------------------------GC----------------------------G---------CAAT-----A-C--G------......------  2645
G.SE.SE6165         ---GT-------------T--------------------------------------A-----G-----------C--------G-----A------CA---------C--G--A---......------  2705
H.CF.90CF056        ----T----------------------------------C-----T-----------A-----------------C---------------A--G--AAA------------------......------  2655
J.SE.SE7887         ----T-------------T--------------------------T-----------A-----G--T--------C-----------C--A------CAA--------------AGA-......------  2622
K.CM.MP535C         ----T-------------------G--------------C-----T-----------A-----------------C-----------G---------CAA--------C--G--T---......------  2504
01_AE.TH.CM240      ----T----T--------------G--------G-----------------------A-----------------C-G---------C----------AA------------------......------  2882
02_AG.NG.IBNG       ----T-------------------G--------------------T-----------------------------------------C------G---AA--A---------------......------  2833
03_AB.RU.KAL153-2   ----T-------------------------------------------------------------------------------T---------------T-----------------......------  3353
04_cpx.CY.94CY032-3 ----T-C--------------------------------C-----T-----------G-----------------C-----------G---------CAA-C-G-------G--T---......------  2674
N.CM.YBF30          ---CT----G--C--T-----------------G-----C--G-----C--------A--------------A--C-----------C-----A---AA-T-A--------G--TGTA......------  2902
O.CM.MVP5180        ----T----T--C--T--T-----G--------G--------C--TT-A--------A-----G-----C--A--C--G--------------C--CCAAT----T--C-----AGTG......-----A  3338
O.BE.ANT70          ----T-C--T--C--T--------G-----A--G--------C--T--A--------G-----G--------A--C---------------T-C--CCAAT----TA-C--G--TGTG......-----A  3362
CPZ.US.CPZUS        ----T-C--T--C--T--T--G--G------G-G-----A------T-A--------C--------------A-----G--------G-----C---CA-T-A---C-------A-TT......-----A  3368
CPZ.GA.CPZGAB       ---GT-C--A--C-----------------A--G--G--A--C--TT-A--------A-----G-----C--A--C-----------C--------TCAAT-A-----------GGTA......------  3369
CPZ.CM.CAM3         ---GT----T--C--T--T-----G--------G--G--A--T--TT-G--------C-----G--------------G--------G-----A---CA-T-A---C-G-----G-C-......------  2719
CPZ.CM.CAM5         ---TT----T--C--T--------G-----A--G-----A--C--TT-G--------A-----G--------A--C-----------G-----A---CAAT-A---C-G-----G-T-......------  3010
CPZ.CD.CPZANT       ------AGAG--T--T--------G--------------------T-AA--------A-----GT-A-----A--C--------T------AAA----AA--A-----GCC---TGAT......-----A  2747

H2A.GW.ALI Pol      __G__F__S__T__P__D__E__K__F__Q__K__D__P__P__Y__K__W__M__G__Y__G__L__W__P__T__K__W__K__L__Q__K__I__Q__L__P__Q__K__E__V__.__.__W__T_   p51
H2A.GW.ALI          A---T--T----C-----TG-G--GTTC--A-----C-------A-AAA--------C----G---GTGG--AAC--------AGC-G--AAAA---CAAT----CC-G-----AGTA......-----A  3686

H2A.SN.HIV2ST       ----T--T----C-----TG-G--GTTC--A-----C-------A--AA--------C--------GTGG--AAC--------AGC-G--AAGA---CAAT----CC----G--AGTA......-----A  3138
H2A.DE.BEN          A--GT--T-T--T-----TG-G--GTTC--A-----C--------T-AA--------C-G------ATGG--AAC--------AGC-G---AAAC--CAA-----CC-G-------TA......-----A  3693
H2B.GH.D205         ----T-CT-T--C-----AG----GTTC--A-----C-----G---AAA--------------G---TGG--AA-A--G----A-C-G--AAAA---CAA--------------AGTT......-----A  3666
H2B.CI.EHO          ----T-CT-T--T-----AG----GTTC--A-----C---------AAA--------G-----G---TGG--AA-G-------A-C-G--AAAA---CA---A-----------GGTT......-----A  3664
H2G.CI.ABT96        R--GT-CT-T--C--T--TG----GTTC--A-----C-------AT-AA--------A------T-RTGG--AA-A--G----A-C----AAAG--TAAC-----M--------G-CT......-----A  3020
MAC.US.MM251        A--GT-CT-T--C-----AG-G---TTC--A-----T--C-----T-AA--------G--C---T-GTGG--GACA-------AGT-G--AAAG----A-T-----C---G---G-C-......-----A  3586
SMM.US.SIVSMMH9     A---T--T----C-----AG-G--GTTC--A-----T--C------AG---------A------T-GTGG--CACC-------A-C-G--AAAA----A-T----GC---G---A-C-......-----A  3083
STM.US.STM          A---T--T----C--T--AG-G---TT------G--T--------T-AA--------A-----GT-ATGG--AACA--G----A-C-G--AAAA----AAT-A--TC-G-G----GTT......-----A  3257
TANTALUS.UG.TAN1    A--CT-CGAG--G-----T--G--GGTG------AG--------ATGA---------A--CA-GT-GT-C---TC-C-C----GTC-CA-CAAA----A-A-A-----GC----ACA-......-----A  3510
GRIVET.ET.GRI-677   A--T--AGAA--C--------G---GTC--A-G------G----GGGAA--------G---A--T-G-----GA----------CA-TA-CAAA----AAT-A--CCCCTT---AG-AGAA...-----A  3314
VERVET.KE.VER-9063  ----T-AGAA--C-----------GGTG--A--GA-G--A----ATGAA--------A--CA--T-ATGG--AC-A------CA-C--A-TAG----AACT-AGA------G--ACAA......------  3538
VERVET.KE.AGMTYO    ---CT-AGAA--C-----A--G--GATG--A--------A--T-ATGAG--------A--CA----TTGG---C-C------GA-C--AGCAGA---CAA---GAG--------TGAA......------  3028
VERVET.KE.AGM155    ----T-AGAA--------G-----GGTG--A-G---------C--TGAG--------A---A--T-ATGG---C---------AGT----AAG-----AAT-AGAGA-G-----ACAA......-----A  3530
VERVET.DE.AGM3      ----T-AGAA-----------G---GTA------A----A--T-ATGAG--------A--CA--T-GTGG--AC-C--G---CAGA--AGCAGC----AAT-AGA---C-----AGAA......------  3033
SABAEUS.SN.SAB1C    ---GT---AA--C--T--A--G--GTT---A--G--T--A---A----C--------G------T-GT----AA-G------CA-C-G---GAG--CACC--A-------GG--AGAA......-----A  3706
SUN.GA.SIVSUN       -AATT-AGAA--------G------T-----GG---G--A--C-ATAAG--------G--CAT-T-A--C--A--C--G---GAGA--G--AAG---AAAT-A--CCCTCT---AGAAGAG...CCA--A  3668
MANDRIL.GA.MNDGB1   -AACT-AGAA-----T--A--G---T----------------C--T-A---------G-----GT-A--C--A-----G----AGA--G-AAAGG-TCAA--A------TT--CAGAACAG...CCA--A  2998
LHOEST.KE.hoest     -AACT-AGAG--------A--G--GTTC--AGC---G--A--C-AT-A--------------T---G----------GG---GA-A--G-AAAG--TAAAT-A-----G-TG--TCTAACAAAAACT--A  3666
LHOEST.CD.hoest447  -AATT-A-AA--------G--G---TT---AGC---------C-AT-A---------G----T---T-----A----GG---GA-A--G--AAA-CTAAA--------C-T----CTTCAGAAGACA--A  2582
LHOEST.CD.hoest485  -AATT-AGAA--------A--G---TT---AGC---G-----C-AT-A---------G----T---T-----A----GG---GA-A--G--AAAG-TAAA--------C-T----CTTCAGAAGACA--A  2582
LHOEST.CD.hoest524  -AACT-AGAA--T-----G--G---TT---AG-CA----A--C-A--A---------G----TC--G--C--A----G----GA-A--G--AGAG--AAA--------G-T-----CAAGGAAAACT--A  2579
RCM.NG.SIV-NGM      ---TT--GAG--T-----T--G---TT----G-CA-G-----C--TGAA--------C----TG--------A--C--------------AGAG--CAA-T-A-T---------AGAA......-----A  3657
COL.__.SIV-CGU1     T--GT-AGAG--------A--G---T-C---GC---T-------A-AAG--------C------T-A------A-GC------A-T----AAAGG-CAAAA-T------C-G--T-AG......-----A  3080

SYKES.KE.SYK173     ----T---AGGTC---A-AG-----TGG---G-CC--TA----A-G-AG---T-A-----CACT--T-----A--C--G---CAGT-G---AAA----AA--C--TA-T-T---TGA-GAA...ATC--A  3505
SYKES.KE.SYK173 Pol __G__F__K__V__P__K__E__K__W__Q__D__Q__Y__P__M__Q__W__L__G__Y__T__L__H__P__D__K__W__Q__L__Q__K__I__E__L__P__N__I__D__D__E__.__I__T_   p51
```

```
B.FR.HXB2 Pol       _V__N__D__I__Q__K__L__V__G__K__L__N__W__A__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__   p51
B.FR.HXB2           GTCAATGACATACAGAAGTTAGTGGGGAAATTGAATTGGGCAAGTCAGATTTACCCAGGGATTAAAGTAAGGCAATTATGTAAACTCCTTAGAGGA...ACCAAAGCACTAACAGAAGTAATACCACTAA   3435

A.UG.U455           --------T--------A-----A--A---C-A-------------------TG---------------A----C-G--------T--C--G---...G----------------TA--G--A-C--G-   2881
C.ET.ETH2220        --------T-----A-----------A-----A--C----------------------------------C-G--------------G---...G----------------CA--G--A------   2827
D.CD.84ZR085        --------T----------------A-----A-----------C--------T-----A---------A--------------------G---...-----G-----------G-------------   2957
F1.BE.VI850         ----------------------A--A---C-A-------------------T------------------C------------------G---...G----G-----------CA--G-G-----G-   2772
G.SE.SE6165         --------T-----A-----------A---C-A-------------------T---------------C---C----------------G--G...G----------------CA--G--T----G-   2832
H.CF.90CF056        --------T----------------A---C-A-------------------T---AAT----------A----C---------------G--G...G--------T-------CA----------G-   2782
J.SE.SE7887         --------T-----A-----------A---C-A-------------------T-----A---------A----C--------------A---G...G-T--------------CA--G-----T-G-   2749
K.CM.MP535C         --------T----------------A-----A--------------A--A--T---------------AA-----G--------------G---...GT---------------CA--G-----T---   2631
01_AE.TH.CM240      --------T--------A--------A---C-A--------------A-----TG----------G---A----C-G-----------C--G---...G-T--------------CA--G-------G-   3009
02_AG.NG.IBNG       --------T--------A--------A---C-A-------------------TG---A------A---A----C-G----G------C--G---...G----------------TA--G--G----G-   2960
03_AB.RU.KAL153-2   -----------------C-------A--------------------------TG-------------------------------------G---...G---------------------------------   3480
04_cpx.CY.94CY032-3 -----C--T--C-------------A---C-A-------------------T---------------A---------------T-----G---...G-T-----C--------CA--G---------   2801
N.CM.YBF30          --------T--------A-----A--A--G--A-----------------C--T-----A--C-G-----AA--GC-C------T-AA-C------..--------TT-G---------G-CAACT-T-   3029
O.CM.MVP5180        --A-----T-----A--A-----A--A-----A--------------A--C--T-A---A----G-----AAG----G--C--GT-AA-C------..------T--T-G-----G---G----TT---   3465
O.BE.ANT70          --A-----T-----A--AC--A-A--A--GC-A--------------A--C--T-A---A----G---G--AG----G-----GT-AA-------C..-----GT--T-----------G----TT---   3489
CPZ.US.CPZUS        --------T--T-----A-----A------C----C---------------A--T-----A--A---A---A---G-----C---T-AA-A-A----...G-T-----T-----T-----TG--AATT-T-   3495
CPZ.GA.CPZGAB       --------T-----A--AC-GA-A--A--G--A-------------------T-----A--A---A---A------------G--GA-A-------..--A--GAA---G-----T---G-T--T--C-   3496
CPZ.CM.CAM3         --------T--T-----AC-------A---C-A--------------A--A--T-----A--A-------AA--G-----C--G--GA-A-A---G..--A-----T---------CA--G-GA-CA-G-   2846
CPZ.CM.CAM5         --------T--T-----A-----A--A-----A--------------A-----TT-------A-------AA--G-----C------A-A-A---G..--A-----CT-G--T------G--G-CA-G-   3137
CPZ.CD.CPZANT       --T--------C-----AC----A--A-----A-----------------C---------A--C---ACT-A---G--G----G----A-C-------...GT---GAGT--------TAG-G---A-A-G-   2874

H2A.GW.ALI Pol      _V__N__D__I__Q__K__H__V__G__V__L__N__W__A__A__Q__I__Y__P__G__I__K__T__K__H__L__C__R__L__I__R__G__.__K__M__T__L__T__E__G__V__Q__W__   p51
H2A.GW.ALI          -----------C--A--ACAT-----TGTCC-A----------GCA--A--C---------A--A--GACC-AA--C--------G---AA---------...-AA-TGA----C--------G-G-G-AGTGG-   3813

H2A.SN.HIV2ST       -----------C--A--AC-G-----TGTCC-A----------GCA--A--C------------A--GACC--AA-C--------GGT-AA-C-------...-AA-TGA----C--------AGG---AGTGG-   3265
H2A.DE.BEN          -----------C--A---C-------AGTC--A---------GGCA--A--C--TT----A--A---ACC-AA--C--------G---AA---------...-AA-TGA----C--------A-G-G-AGTGG-   3820
H2B.GH.D205         --G----CA--T--A--AC-G--A--AGT---A--C-------GC---AC-C-TT--T--A-----GAC------CA----C-----AA----G---...-AG-TGA-C-----------A-G---AGTGG-   3793
H2B.CI.EHO          --A--------T---------------AGT---A----------GC---AC---T---G---------GACC-----TA-----------AA-A--G---...-AG-TGA-C-----------AGG---A-TGG-   3791
H2G.CI.ABT96        --A-----T--T--A--AC----A---GT-C----Y-------GC----C-A--T-----A--C---AC--AAA-CC-G--Y---T-AA-------T...-AR-TGA-CT-G--G----A-G---AGTGG-   3147
MAC.US.MM251        --G-----T---------------A--AGT---A----------GC---A-----T-----T--A---ACC-AA--TC-C----GGT-AA---------...-AA-TGA-T--------G-A-G-T-AGTGG-   3713
SMM.US.SIVSMMH9     --A---A-T-----A--A-----A--AGTGC-A----------GCA--A-----T-----A--A--GACT-AA--TC-T--C-G-T-AA-C-------...-AA-TGA-TT----------AGG-T-AGTGG-   3210
STM.US.STM          --G-----T--C--A--AC-G--A---GTC--A--C-------GCA--A-----T-----A------ACT-AA--TC-G--C---T-AA-C-----T...-AG-TG-----G--T--G-AGG---AGTGG-   3384
TANTALUS.UG.TAN1    --G-----T--T--A--A-----A--AGTG--A---------T-----AT-A--T-----A-----GAC--AAGCCC----C---T-A--AC------...G-TC---ATT------CT---G-G-A-TGG-   3637
GRIVET.ET.GRI-677   --A--CA-A----------G----A--AGTTC-A---------------A-----T-----A------ACC-AA--TACC---GCCA-GT-G-----G...-AA--GAAC--CCT-----A----GT-TGG-   3441
VERVET.KE.VER-9063  --T-----T--C--A--AC----A-----------C-------GC----C-------C---T-A-G-AC--AA--GA----C-GGT-AA-C--G---...-AG---AAT---TT---G-A-G-G-A-TGG-   3665
VERVET.KE.AGMTYO    -----------C-----------T------C-A----------GCA--AT-G--T-----TC----GAC---A...A----C--GT-AA---C----GGA-AG---AAT--GTT---GC--G--G-TTGG-   3155
VERVET.KE.AGM155    --G-----TC-T-----A--G--A---------A----------GCA--AT-A--T-----AT-G-G-AC--AAA-TA-C-----G--A------------...-AG---AATT--TT---C--GG--GA-TGG-   3657
VERVET.DE.AGM3      --A-----T-----A-GAC-----------C-A----------GCA---C----------AC-C-G-ACT-AAA-C--G------T-AA-C-------...-AA--GAACT--CT----AC-G--A-CTGG-   3160
SABAEUS.SN.SAB1C    --G-----T--T--A--A-----A--A--G--A----------------A------A-------A--GACT-AA--T--G----G---AA---------...G-T-G-C-TT-G-----GA--G-C-A-TGG-   3833
SUN.GA.SIVSUN       --G--------------AA-T--A--AGTTA-C---------TTCA--AT-A--TGA----C-C-G-AC--AAG---------------AA-------GAAT-AACC-TT-GC-GAGA------AGATGAC-G   3798
MANDRIL.GA.MNDGB1   --A-----A--------A--G--A--T-----A---------TGCA---T-A--T--T-----C---AC--AA---C-G--C--G--AA-A-------...GGACT-AACA-------GAA-G-CA--A-G-   3125
LHOEST.KE.hoest     --A---C-A--------AC-G-----AGT-C-T----------GC---AT-G--TGAT--T----GGAC--AAG--C-C--C---T-AA-A--G---...GTA---C-CT-GGA----A-C---AACTGG-   3793
LHOEST.CD.hoest447  --T--CA-----------C----A--AGTC-------------GCA--AC----TGAG--A----G-ACT-AAG--C-T------T-AA----G---...GTA-----T---GAT-----TG--A-CTGG-   2709
LHOEST.CD.hoest485  --A--CA-----------C----A--AGTC-------------GCA--AC----TGAG--A----G-ACT-AAG--C-T------T-AA----G--G...GTA-----T---GA------TG--A-CTGG-   2709
LHOEST.CD.hoest524  --A-----A--T-----AC-G--A--AGTT--A----------GCA--AC-G--TGAT--A--C-G-AC--AAG-GC-C-----GT-GA---------...GTA--GC-CT-GGA-C--C--G-GA-TTGG-   2706
RCM.NG.SIV-NGM      --A-----T--------A--G--A---------A------------------A--TT----A--A--GAC--A-G-G-----C-----AA-A--G---...GTT--G---T-GGAT--CAAGG--GA-T-T-   3784
COL.__.SIV-CGU1     --A-----T-----A--A--G--A--AGCC--A------TT---C--AT-A---------A--C-GGAC--A-G-GA--A-----TG-AC--A---T...-AG--GCATT-GTTG--G-A-G-G-AG--T-   3207

SYKES.KE.SYK173     --G---C-GT-------A--GA-A--AGT-C-T-----------C---------TT----A-----GAC--AAG-GC-C-----GTGTA-C-------...--T---CA-T-----------C-CA----T-   3632
SYKES.KE.SYK173 Pol _V__N__Q__L__Q__K__L__I__G__V__L__N__W__A__S__Q__I__Y__S__G__I__K__T__K__E__L__C__K__C__I__R__G__.__T__K__Q__L__T__E__V__L__T__L__   p51


B.FR.HXB2 Pol       T__E__E__A__E__L__E__L__A__E__N__R__E__I__L__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   p51
```

```
B.FR.HXB2           CAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGAGAGATTCTAAAAGAACCAGTACATGGAGTGTATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATA   3565

A.UG.U455           -T-----------AT-----T-------G-----G-------------C--T--G---------A--------------------C--G----------------A--A-------A------------   3011
C.ET.ETH2220        -T-----------AT-----T-------G-----G--A--------G-----------------A-T--------------------------T-----------A------A-T-A-----------T   2957
D.CD.84ZR085        -------------AT-------------------G-----------G------A-G-----------------------------------------T-------A--A---------T-----------   3087
F1.BE.VI850         -T-C---G------T-----T-----A----T--G----------G---------------G--A----------------------------------------A--A--AG-C--G-----------   2902
G.SE.SE6165         ---C---------AA-G--GT-------G-----G--A-------G------T-----------C-------------------A-----------G----A--A--A----T--A------------   2962
H.CF.90CF056        --A----G-----AT-G---T---------------G--------G-G-------A-----------A--------T--------------------------G------A-------------------   2912
J.SE.SE7887         -TAG---------AT-G-----------------A-------------------------G--A---------T--G-----------A----------G-G-----A--A----TG-A------------   2879
K.CM.MP535C         -T-C----------T-----T-A-----G-----------------G-----G-----G--A----------------------------------------A--A---A-T-A------------   2761
01_AE.TH.CM240      -T-----------AT----GT-------G-----G-----------ACC--T--G---------A-----------------------G-------G-------A--A-----G-A------------   3139
02_AG.NG.IBNG       -T--G--------AT-----T-------G-----G--A-----------------T--------G--A------------A-----------G-------T-------A--A-------A------------   3090
03_AB.RU.KAL153-2   ---C------------------------------G-----------------------------------------------------------G-----------------------A-----------------   3610
04_cpx.CY.94CY032-3 -TAC---G------T-----T-A-----G-----G-------------------------G-CA--------------------------------------------A---------T-----------   2931
N.CM.YBF30          -------------AT-------A-----------G-----AT-----------CC-G---------C------------GG------A---G----------T--A-----A--A-----T--G--------   3159
O.CM.MVP5180        GTA----G-----A------T-A-A-----------A-AG--------G---------------A-----CC-G--TGAC---------GTGG-TTAGT--T---------T--AG----G---------T--   3595
O.BE.ANT70          GTAG---G---------G---T-A-AG-----------A-GGT-----C--------G-----G--A--C---C-A--TGAT--G--TC--TGG-TTA-T--T---------A--AGGG-AG--------T--   3619
CPZ.US.CPZUS        --C-T--G-----AA-G--GT-A-A-------------A---------G-------------G--C-------------GA------A---G---------G----A--A--A--AAGGA-T--G--------   3625
CPZ.GA.CPZGAB       --CC---------AT-----T-A---------T--G-----AG---GCAC------------G--A--C-----T---GAC-----GC-T-----------------A-----CA-CT----G-----T--   3626
CPZ.CM.CAM3         --C-G--------AA-G---T-A-A---G-----R------T---RG--T-----G-----G-----C--------GA---G--A------------N-------A-----AA-TA-T--G-----T--   2976
CPZ.CM.CAM5         --C-G--------AA-G---T-A-A------T-----A---T----G--T--T--G---------T--C-----T---GA---G--G-----------GG----A--A-----AA-TA-T--G-----T--   3267
CPZ.CD.CPZANT       -TAGG--------AT-----T-A-A------T-A-C-A------C-GC-GAA-A--G-G---TAC--C---C-G--TGGTCT-CCAC---A----ACC---------A--A---TC---A---------C--   3004

H2A.GW.ALI Pol      T__E__L__A__E__A__E__L__E__E__N__R__I__I__L__S__Q__E__Q__E__G__H__Y__Y__Q__E__E__K__E__L__E__A__T__V__Q__K__D__Q__D__N__Q__W__T__Y    p51
H2A.GW.ALI          -----CT-------AGC------A-AG--G------ATT--CT---GTC-GGA-CA-G-G--GCAC-----CC-AGA-GA---G--G---GA----AC-G-C--A--AG-TCAAG-CAAT-----------   3943

H2A.SN.HIV2ST       -----TT-------AGCG-----A-A--------A-ATC--CT---GCC-GGA-CA-G-A---TGC-----CC-AGAGGA---G--GC--GA----AC-G-C--A--AG-TCAAG-CAAT--G--------   3395
H2A.DE.BEN          -----CT-------AGC---G--A-A--------A-ATT--CT-G-GCC-GGA-CA-G-A---TAT-----CC-AGA-GA------A---GAG---AC---C--A--AAGCCAAGG-CAT-----------   3950
H2B.GH.D205         -----CT-------AGC---G--ACAG--G--T-A-ATC--CT--G--C-GGA-CA-G-A---TCC--C--CA-GGA-AGGGT-CCGC--GA----AC-G-------AA-CCTAGC-AAT--G--------   3923
H2B.CI.EHO          -T---TTG------GC----T-CCAG-------A-ATC--C---G--C--GAGCAGG-A---TCC-----CA-AGA-GGGGT-CCT---GA----AC-G-G-----AA-TCTAGC-AAT--G--------   3921
H2G.CI.ABT96        -----TTG-----AGC----T--CA--------AGATT--C---R--C-GGARCAGG-A--C-CC-----CA-AGA-GR-G-GCCT--RGA---CACTG-------AA-TTTAG-CAAT--G--------   3277
MAC.US.MM251        -T--GATG-----AGC----TAT-AG-----T-AGATA-----C-GTC-GGA-CA-G-A---TGT-----CC-AGA-GGC--GCCA---GA---CACGG--ATA---AGTCA-G-CAAT--G---T-T--   3843
SMM.US.SIVSMMH9     -T--GATG------GC----TAT-A--------AGATA----CC-GTC--GA-CC-G-G---TGT--C---C-GGAGGG---GCCAC--GAG---CC-G--ATA---AGTCA-G-TAAT------T----   3340
STM.US.STM          -----ATG------GC----TAT--------T-A-ATT-----C-GTC-GGAGCAGG-A--GCAA--C--CCGAGA-GAC--GCCTC--GAG---AC-G--GT---AG-TCA-G-CAAT------T-C--   3514
TANTALUS.UG.TAN1    -G--GC-G------G----GT-CAG-C----TCA------CT--------GAGCA---G----CT--C---------AGT---TCT--G-G---CACT--TACC----T--ATG-T--G------GG---   3767
GRIVET.ET.GRI-677   -------G------GC----TATAAGA-C--TCA--G---AG-GC-G---A--CA-G-A---ACA--C--------TCTC-----A---------AC-G-T--A-----A--AG----G-----------   3571
VERVET.KE.VER-9063  -T-----G--G--AGC---GTAT-----G----A---A---T-G---AC-GA-CA-G-A---ACT--------CA---GG---GCCTC-T-G-----C-G-C-----ATTA-AAG-T--G-------GT--   3795
VERVET.KE.AGMTYO    --CCT--G-----AGCT---TAT---------TGC---------T---AC-GA-CAGG-A---ACC-----CA-A---GG--T-CCTA-T-GG----C-G-------ATT--AAGG---A--G----GT--   3285
VERVET.KE.AGM155    -CCC---G-----AGC---GTAC-A--------AG-----C------AC-GAGCA-G-A--TACT-------CA---GA----CC-C-T-GG----C-G-------ATTA--AG-T--G------T----   3787
VERVET.DE.AGM3      ----G--------AGC----TAT----------A------CT-----ACGGA-CAGG-A--GACC--C--CA-A---GG--G-CC-A-C-G-----C-G-G--A--A-TA-AAGG---T-------GT--   3290
SABAEUS.SN.SAB1C    -------G-----A------T-A-A------T---C----CT---G-C--AAGCAG--G---CAA------------TG-TCTTCCT----GG--TA-GG-T-TT--A-T---AG-T---------GG---   3963
SUN.GA.SIVSUN       A-----CTAGG--AGA-T-C-A-AGTA-TC-GGAG-TAT-A...C-G--GT-----TCA--CAGT--C-----A---GAT-----A-----TTGCAG-G-G--A---GTAAA------AAT--TA--C-T   3925
MANDRIL.GA.MNDGB1   -----------AGA--G---TAT-A-C----T-A------CT-GGCT---GA-CA-G-A--GTCT---------T--TAAT--G--A---TAT-T-AG-T-T-----AACAACAGG---AG-TATAT---T   3255
LHOEST.KE.hoest     ----G-----CTT-GA----TAT-GGC-G----A----G-A--T------AAGA-G--G----CC-------------GA---G--AC-T--T-TCAGGG-----C-AA-CAAAA-G--GAT-ATT--T-T   3923
LHOEST.CD.hoest447  ----T-C----TTAGA----TAT-A-C-----CA-C----A----------AA-C-G--G----CT---------T--TA-T-----A--GG---T-AG-G-----C--AGTAA-A-G--AATTATT--C-T   2839
LHOEST.CD.hoest485  ----T-C----TTAGA----TAT-A-C-----CA-C----A----------AA-C-G--G----CT-------------TA-T-----A--GG---T-AG-G-----C--A-TAAAA-G--GGTCATT--C-T   2839
LHOEST.CD.hoest524  ----G------CTAGA----TAT-AGACC--T-AGACA-------------AA------G--G-CT--C-----T---AGT-----AC-TG---T-AGGG---AC--A-CAA-A----GGTTATA--T-T   2836
RCM.NG.SIV-NGM      --AG---------AT-----TAT-A--------A--CTG-C--G--G--GAA-A-------G--T------A-----GA----CCA--------TA--G-T-----ATTA-AAGG---G-------GC--   3914
COL.__.SIV-CGU1     GT-C---------AGAG--G--ACA---T----A-C----AA-C------GA---T--G--------C---ATA---CATG-G---A-TTGG-T---C--TTTC-G-TTA---A----A------GGC--   3337

SYKES.KE.SYK173     ------C---T---GC----T-A-A---G--T---C-A--CT----G---GA-CAGGCA--GTCT-----------TAA----CCT---GA----C-C--CACT--A-TA---TCCCAG------GG---   3762
SYKES.KE.SYK173 Pol T__E__A__A__E__A__E__L__E__E__N__R__Q__I__L__K__E__E__Q__A__G__S__Y__Y__D__P__K__K__P__L__E__A__H__I__T__K__L__G__S__Q__Q__W__G__Y    p51
```

```
B.FR.HXB2 Pol       __Q__I__Y__Q__E__P__.__F__K__N__L__K__T__G__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_   p51
B.FR.HXB2           TCAAATTTATCAAGAGCCA...TTTAAAAATCTGAAAACAGGAAAATATGCAAGAATGAGGGGTGCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGCATAGTA  3692

A.UG.U455           -------------------...----------------------------------AA---TC---T-----------------------------A-TG---------G-GT-----------------  3138
C.ET.ETH2220        ----T----C-----A---...--C-----------------G--G-T-----A--GAG--AC---------------------G--G-------C--T---------G---G--CTG------------  3084
D.CD.84ZR085        ---------------A---...-----------------------G-----------------------------------------G------------------------G---T------------G  3214
F1.BE.VI850         ------------GA-C---...-----------A-----------G-------A-G-----TCG--------------------------------A-----A-----G---G-TTT-------------  3029
G.SE.SE6165         -------------------...-AC--------------------G-------A--G-G--TC------------------------G--------A-T-------------G-------G---------  3089
H.CF.90CF056        ------------G------...--------------G----------------A------AAC----------------A----------------A-----------G---T-T---------------  3039
J.SE.SE7887         ------------G------...--------C-----------G----------A--G----A-----------------------------G----A-TG------------G--TTG---GC-------  3006
K.CM.MP535C         ------A------------...CA------------------G--G--------------ATC---------------------G-----------A---------------G--------G--------  2888
01_AE.TH.CM240      -------------------...-----------A-----------------C----GAG--TC---T----------------G------------A-TG---------G--G-----------------  3266
02_AG.NG.IBNG       -------------------...---------T-A-------------------A--A----TC---------------------G-------------T----------G--G-T-TG------------  3217
03_AB.RU.KAL153-2   -------------------...---------------------------------C---------------------C---------G------------------------G----T------------  3737
04_cpx.CY.94CY032-3 ------A------------...CA------------------G--G----------CC--ATC------------------T-G------------A-----------G---G---TG---T--------  3058
N.CM.YBF30          ---G--------G---TT-...CA-------T-A-----------G-------A------ATC------T---------A-------G---GTT--A-TG--AAGG---G-GG-A--------T------  3286
O.CM.MVP5180        C--GG-A-----G--TGA-...CA---G--C--T-----------------T--GCAA-A--CCT-------A------A---G------GG----A-T---C--G--GG-GT-TCA----GCT-----T  3722
O.BE.ANT70          C--G--A-----G--AGA-...CA---G--C--C--------G------A-T--GCAA-A--CCT-------A------A---G-------G----A-T-A-C--G--GG-GT-TCA----TCT---A-T  3746
CPZ.US.CPZUS        C-------T------A-GG...CA---G-----A--G--------------C---CAA--ATCA--A--T---------A-C-G---GC--GTT--A-TG------------G-T--T--G-----T--C  3752
CPZ.GA.CPZGAB       ---G--A-T---G--A---...CA---G---T----G-----G--------C---CAA---TCA--A-----A-----CA-C-GG---C-GG-G--A-----A--------TG-T--T------------  3753
CPZ.CM.CAM3         --------TC--G--A-AG...CA---G--C--A--------C--------T--GCAA--ATCA--A--T--------CA---G-------G------T---A---------G-T--T-----T------  3103
CPZ.CM.CAM5         C--G----TC--G--A-AG...CA---G--CT-A--------G--------T--GCA---ATCA--A--T--------CA---G-------G------TG------------G----T-----T------  3394
CPZ.CD.CPZANT       ------A------A-TGA-...GGA---CTGT-A---G-------------T--GCCT-CA--AA-T-----------G--T-GG------G-T-GA-T-------------GGATT------T-----T  3131

H2A.GW.ALI Pol      __K__I__H__Q__G__.__.__E__K__I__L__K__V__E__K__Y__A__K__M__K__N__T__H__T__N__G__V__R__L__L__A__Q__V__V__Q__K__I__G__K__E__A__L__V_   p51
H2A.GW.ALI          -A----AC-C--G-GA......GAA----T---A---GTG-A---G-------A-----AAAA-A----T--C--C-GG--C-G-TTG---G--C---T---T---------GGA-A----GCAC-G--C  4067

H2A.SN.HIV2ST       -A-G--AC-C--G-GA......GGA----T---A---GT--------------AGG-A-AAAA-A-------C--C-GA--C-G--TCC--G--C-A-T---T---------GGA-A----GCAC----C  3519
H2A.DE.BEN          CA----AC-C--G--A......GAG----TC--A---GT------G-------AG--A-AAAA-A----T--C----GG--C-G-TT-C--G--C---T---T--G------GGA-A---GGCAC----C  4074
H2B.GH.D205         CA-----C----G-GA......AA----GTC--A---GT--------------AGG-T-AAAACA-G-----C--C-GG----G--T-C-GG--C-T-T---T--G------GG--A----GC-C----C  4047
H2B.CI.EHO          CA-G---C----G-GA......GA-----TC--A---GT--------------AGG-T-AAAACA-T-----C----GA----G--T---GG-TC-T-T---C---------GGA-AG---GCAT-G--C  4045
H2G.CI.ABT96        CA--G--C------GA......AAC-G--TC--A---GT---T----------AG--A-AAAA-A----T--A----GA----G-TT---GG-TC-T-TG--C-----R---GGA-AG---GCAC-----  3401
MAC.US.MM251        -A-----C-C-----A......GAC----TA------GT--------T-----AG--A-A-AA-A-A--T--C----GA--T-G--T----G--C-T-T-A-A--G------GGA-AG---GCA-----G  3967
SMM.US.SIVSMMH9     -A-----C-C------......GAC----TA--A---GT---C----T-----AG--T-AAAA-A-A--T--A----GA--C-GCTT----G--C-T-T------G------GGA-A----GCA---A--  3464
STM.US.STM          -A-G--AC-C-----A......GAC----TAT-----GT---C--G-T-----A---A-AAAACA-A--T--A----GA--T-G-TT---GG--C-T-TGA-T--G--G---GGG-A----GCA-----G  3638
TANTALUS.UG.TAN1    ----T-CA-A--G---......GGA---GTA--A---------------A---AG-GC-AAAA-A-T--T--A-----GT-T-G-GT---GG---GAATG--C----GGG-GTG--AG---GCG--CACC  3891
GRIVET.ET.GRI-677   C---T-CACC-----A......GGGGC-GTAT-A--GGTG----G------C-AGCAA--A-AAA-T--T---------C---GGACTC--G--C-CCTT--C--------CTGT-AG---GCAC-TACC  3695
VERVET.KE.VER-9063  ---GT--A-A--G--A......GGA----TCT-----GT------G-------A-CAA-A-AACA----T--A-----AC-TCGCAC---GG-C-GCCT---A--------CTGT-A---GGCAC-G--C  3919
VERVET.KE.AGMTYO    C---T-CA-A-----A......GGAC--GTCT-----GT---------CA-C-AGCAA-A-AACA----T--A-----AC-TCGCAC----G-T-GTTT------G--G--TTG--A----GCTC----T  3409
VERVET.KE.AGM155    C---T-CA-G--G--A......GGA----TCT-A--GGT---G--G-TC--C-A-CA--AA-C-A-T-----C-----GT-GCGTGT-C--G---GA-T---A--G------GGG-A---GGC-C-----  3911
VERVET.DE.AGM3      C---T-CA-G------......GGAC--GTAT-A---GT---T-----CA---AGCA--AAAACA-T--T--C-----GT-CCGTGT---GG---GATT---A------C-TTGT-A----TCTT----T  3414
SABAEUS.SN.SAB1C    ---G--A--C--GCCAGA-...AA-----TA-----GGTT--C----------A---C-AAACA--A-----C-----AT---G-ATG---G---GCCT---A--G--G---GGA-AG-----------C  4090
SUN.GA.SIVSUN       ----TGG-TG--G-GA......AAAC--GTAT-A-GGGT-----G----CA--AG-GAG---CA--A--TGAG---CC-TGTC----G---G---CT--CC-A-----G---GGA-G------------T  4049
MANDRIL.GA.MNDGB1   ----TGGA-G----GA......AA---GGT-T-A-G-G----G-------GG-A-CA--AAACA--A--T-G------CC-C-TGA----GG---GT--TAC---G--GG--GGA-G------T------  3379
LHOEST.KE.hoest     C--GTGGAGA----GA......AA---C-TCT-A-G-G-T--G-GG---CA----CA--A--CA--A-----A---CCCC--C-GA-----GT---A--TA-T--G--G---GGA-A------------C  4047
LHOEST.CD.hoest447  C--GTGGAGA----GA......AA-TCT-T-----G-G------GG---CA----CAA-AATCA--T-----A---CCCC--C--A----GG----------A---------GGA-A---G--T-----C  2963
LHOEST.CD.hoest485  C--GTGGAGG----GA......AA-TCT-T-----G-G------GG---CA---GCAA-AA-CA--T-----A---CCCC--C--A-G--GG----------A---------GGA-A---G--T-----C  2963
LHOEST.CD.hoest524  C---TGGAGG----GA......AA-GCC-T---A-G-G------G----CA----CAA-A-A----T----TG---CCCC-TC-G------G----A--T--A--G------GGA-A------------C  2960
RCM.NG.SIV-NGM      C--G--AG-G--G-GAGAC...-CA---CC---A--G-----G--G-------A-CAA-A-ACA--A--T--A-----AA-T-G-ATGC-GG---GCTT---A--G------TG--A----GCT------  4041
COL.__.SIV-CGU1     -GC-G-C-GC---A-A......CA-GG--CG--A-G--G-------C--AAC-C-GGC-AAAA-CAG---TT----AG-A-GC-GG-G---G-TTCA-TTA-A-----G---GGT-GG---GC------C  3461

SYKES.KE.SYK173     CATG---A-G-------A-AAAGGCCC-CCC--A-T---------GACA----A--CCTTT-CA--A--T-G------CTATC--TCTC-TG-TC-ATTGT-AA-T------GGG-TCC--TCAC--TGG  3892
SYKES.KE.SYK173 Pol __M__I__K__Q__E__Q__K__G__P__P__L__I__T__G__K__T__A__K__T__F__A__A__H__S__N__D__Y__Q__S__L__A__Q__L__L__N__K__I__G__I__Q__S__L__W_   p51
```

```
B.FR.HXB2 Pol       _I__W__G__K__T__.__P__K__F__K__L__P__I__Q__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__   p51
B.FR.HXB2           ATATGGGGAAAGACT...CCTAAATTTAAACTGCCCATACAAAAGGAAACATGGGAAACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACCCCTCCCTTAGTGA   3819

A.UG.U455           -------------TC...----------G---A------------------------G---------TG-----------G--T--------------A-----------C------------C----A-   3265
C.ET.ETH2220        ---------------...----------G-T-A-----C--G--A------------G-------------C--------------------------A------------------------C----A-   3211
D.CD.84ZR085        ---------------...----------G---A----------------------------------T---C--------------------------------A-----C-----------T-----A-   3341
F1.BE.VI850         --------G----GATCT-----G--------A-------TG--A--G--------T--------------T--------------------------A-----------C--------C--TC------   3159
G.SE.SE6165         ---------------...--C-----------A-------G---A-------------T------------C--------G-----------------------------C-----------TC----A-   3216
H.CF.90CF056        -----------A-T-...----------G---A--T--------A-----------G--C--------------------------A-----------A--------------C-------ATC----A-   3166
J.SE.SE7887         -----------A---...----------G---A----------GA-----------G--------------C--C-----G-----------------------------C--------C--TC----A-   3133
K.CM.MP535C         -----------A---...----------G---A-----------A-----------G--------------A--------------------C-----A-----------C-----T------C----A-   3015
01_AE.TH.CM240      --------------C...----------G---A----------GA----------------------TG-----------G--T--------------A-----------------------TC----A-   3393
02_AG.NG.IBNG       --------------C...----------G---A----------GA--G-------------------TG-----------G--T--------------A-----------C-----------TC----A-   3344
03_AB.RU.KAL153-2   ---------------...--------------A-----------A-----------------------------------------------------------A-----C-----------------A-   3864
04_cpx.CY.94CY032-3 ---------------...-----G----G-T-A-----------------------C--------------A--------G-----------C-----A-----A-----C-----------TC----A-   3185
N.CM.YBF30          --T------------...----------G-T-A--AG-----------GTG-----GG----------C--TC----------A--T-----------------A-----C--C--T------C-T--A-   3413
O.CM.MVP5180        --------G--ATTA...--------C-GG-----AG-TACT-GA-----T--------T------G----A--------G-----------------A-----A-----C-GC--A--C--A--GA-C-   3849
O.BE.ANT70          --C--------ATTG...-----------G-----AG-CACT-GA-----T---------------G-G--C-----------------------A--A-----A-----C-GC--A--C--A--GA-C-   3873
CPZ.US.CPZUS        --T---------GTA...--A---------T---TAG-G--G------GTC---------------T----A--------G--------------A--T-----A-----C------------C-C--A-   3879
CPZ.GA.CPZGAB       --T--------A--A...----------GGT-A--AG----G--A----GT------G--------G----A--C-----G--A--------C-----A-----A---A-C--------A--------C-   3880
CPZ.CM.CAM3         --C--------A--A...--A-----C-G-T-A--TG-T--------GGTC--------C------T-------------------------C--A--C-----N--C--G------------C-T--A-   3230
CPZ.CM.CAM5         --C--------A--A...--A-----C-G-T-A--TG-T---------GTC--------C------T----A--------------------A--A--C-----A-----A------------C-T--A-   3521
CPZ.CD.CPZANT       --C------G--GTA...--A------C--T-A------ACT-G------T-----TG-C------T----C--------G-----------A--A--A-----------------A--C--A---A-T-   3258

H2A.GW.ALI Pol      _I__W__G__R__I__.__P__R__F__H__L__P__V__E__R__E__T__W__E__Q__W__W__D__D__Y__W__Q__V__T__W__I__P__D__W__D__F__V__S__T__P__P__L__V__   p51
H2A.GW.ALI          --T------CGA-TA...--A-G----C-C--A--AG--G---GA-----C------CAG------GAT--C--C-------TG--A-----C--A--C-----C-----ATC------A--AC-G--C-   4194

H2A.SN.HIV2ST       --T------CGA-TA...--A------C-C--A--AG--G---GA--T--C------CAG------GATA-C--C-------TG--A-----C--A--C-----C--CA-ATC------G--AC-G--C-   3646
H2A.DE.BEN          --T------CG--TA...--A------C-C-----AG-GG-G-GA--G--C-----GCAG------GATA-C--C-------TG--A-----C--A--------C-----ATC------A--AC-G--C-   4201
H2B.GH.D205         --C------G---TA...--AGTG--CC-T-----AG--G---GA--G--------CCAG-----------T--C-------TA--------C--A--------C-----CTCG-----A--A---A-A-   4174
H2B.CI.EHO          --C------G---TA...--A-TG--CC-T--A--AG--G---GA--G--------TCAG-----------T--C-------TA--------C--A--A-----T-----CTCA-----A--A---A-A-   4172
H2G.CI.ABT96        -----------ATTA...--CTTC---C-T-----AG--G---G---C--C------CA------------C----------TA--T-----C--A--T-----C-----GTC---A--G-ATC----A-   3528
MAC.US.MM251        --C------C--GTC...--A-----CC-CT-A--AG-TG-G-G---TGT-------CAG-----------C--------G-TA--------A--G--------T---A-CTCA--G--A--AC----A-   4094
SMM.US.SIVSMMH9     --T------C--GTG...--A-G----C-TT----AG--G-G-RA----TT------CA------------T----------TA--------A--A--A-----C-----GTCA--A-----------C-   3591
STM.US.STM          --C------C---TC...--A-----CC-CT----AG--G-G-GA--GGTC------CA-------G----T--------G-TA--A-----A--A--A-----T-----ATCC--A-----T-----A-   3765
TANTALUS.UG.TAN1    T-T--------ATTA...--C-TCA-GG--G-A--AG--G---GA---GT-------CA--------GT--C----------T---A---C-A--A---ATA--A--CA-CTCC--A--G------A-C-   4018
GRIVET.ET.GRI-677   --T-------GACT-...--ACG-G-AC----C--AG--G-C---A----------T-TG------CAG--C----------TAT-------A--A--A-------------GC--A--A-T-C----A-   3822
VERVET.KE.VER-9063  --T------G-ACTA...--A-TT---G----C--AG--G---G----GT-------CA-------G----C--C-------TG-G------------A-----A-----G-G-GTA--A--TC------   4046
VERVET.KE.AGMTYO    --T-----G-TATTA...--AGTTC-AG----C--G---G---GA--GGT------GCA-------G-G--T--C-----G-TA-G---------C--A-----T-----C-GC-----A--T--GC-C-   3536
VERVET.KE.AGM155    --T------C-ATTA...--C-CT---G----C--AG-GG-G-G---C---------CA-------G----C----------T--GT-----A--C--A-----C-----C-G-GTT--G--------A-   4038
VERVET.DE.AGM3      ---------G--TTG...--AGTCC--G----C--A---G-G-G----GT-------CA-------G-T--T--C-----G-TA-GT--------A--C-----A-----C-G------A---C----A-   3541
SABAEUS.SN.SAB1C    --C------C---TA...----TTA-GG-------AG--G---GA--GCTC------CA-------T-T--C--------G-T---------C--G--A-----AA-G--C-G---T----AA--GA-T-   4217
SUN.GA.SIVSUN       ---------TTTGTG...--A--GA-AC-GG-A----------GA--G-T----AGTCAG------G----C--C-----GTGT--A-----C-------TA--A-----C-GC--T---AAG--G-A-C   4176
MANDRIL.GA.MNDGB1   --C-----TTTTGTA...--A---A-GC-GA-A----CTAC--G---G-T-------GAT------CAT-----------GTGT--A-----A--A--AGTA--A---A-C-GC--A---ATG----AA-   3506
LHOEST.KE.hoest     --C-----CTTTGTG...--A---A--C--G-C--AG-GACT-GA---GTC-----GCAC-------GC--CC-C-----G-TT--A--------A--C-TA--A--CA--TC------G-AA----AAC   4174
LHOEST.CD.hoest447  --T-----TC-TGTG...--C--GA--C-GG-A--GG--AC--GA---GTG-----TCA--------GT--TC-------G-T---------A-----A-T---A---A-CTCA-----A-AG----AAC   3090
LHOEST.CD.hoest485  --T-----TC-TGTG...-----GA-CC-GG-A--AG--AC--GA---GTG-----TCA--------GT--TC-------G-T---------A-----A-T-------A-CTCA-----A-AG----AAC   3090
LHOEST.CD.hoest524  --T-----TTTTGTA...--A--GA-AC--G-A--TG-CACT-GA-------------AT-------T-A-C-T--------TA--T-----A--A--C-TA--A--CA-CTCC--A----AA----AAC   3087
RCM.NG.SIV-NGM      --T---------CT-...--A------G-G-----A---G---GA---GT-------CAG------T-G--T--C-----G-TTT-------------A-----A-------G---T--A--AC--A-A-   4168
COL.__.SIV-CGU1     -CC---------GTG...--C---A-G---G-C--AGC-A-G-GA--GGAT------CAG-------GT--C---------AGTG-T-----C--A---ATA--------CTCAT--T--TATG----A-   3588

SYKES.KE.SYK173     TAT---------GTA...--A-C---CC-T--C--AG--A---GA--GGA--------A------------C--------------A---G-A--A--AGTCA-A--CA-A-G------G--A--G--T-   4019
SYKES.KE.SYK173 Pol _Y__W__G__K__V__.__P__T__F__H__L__P__V__K__R__E__E__W__E__K__W__W__T__D__Y__W__Q__A__T__W__V__P__E__V__K__F__I__S__T__P__P__L__V__   p51
```

```
                     Pol p66 reverse transcriptase continued (RT is dimer of p51 and p66)
                            Pol p51 reverse transcriptase end \/ Pol p15 Rnase H
B.FR.HXB2 Pol        K__L__W__Y__Q__L__E__K__E__P__I__V__G__A__E__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   p15
B.FR.HXB2            AATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACCTTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACTAATAGAGGAAGACAAAAAGTTGT  3949

A.UG.U455            -------------------------C-------C---------G--A-----------------------T-----A-----GC-------------G-----C---G-C--------------G-----  3395
C.ET.ETH2220         --------------C------------------C-----T------T--------------A--------T-----A------A-------------G---------G----------G--G---A----  3341
D.CD.84ZR085         ----------------------G---------A-------------T-----------------------T-AA---------------------------------G-C-----------G--------  3471
F1.BE.VI850          --C----------------A-C---G-------C---------C---------------------T----T--A-----C---AA----------------------G---A-----A------------  3289
G.SE.SE6165          -G--------T-G------A-C----------CC------------T-A---------------------T--------A-----------G---------------G-C-A-----AG------A--A-  3346
H.CF.90CF056         ----------T--------A-C---G-------C------------T-A----A----------------T-----A--------------------------C---G---------AG-----------  3296
J.SE.SE7887          -G-----------A-----A--G---------A-G-----------T------------------T-----------------AC------G-----G---------G-C-A---------G-----A--  3263
K.CM.MP535C          --C----------------A-C---------------G--------T-----C---------------C-T-----A------AAG----G----------------G-C--------------G-----  3145
01_AE.TH.CM240       -------------A-----A-----C-----------------G--T----------------------GT-----------GC-------------G-----C---G-C--------------G--A--  3523
02_AG.NG.IBNG        -------------------A-----C--------------------T--------------A--------T--A--------GA-------------G-----C---G-C--------------G-----  3474
03_AB.RU.KAL153-2    ----------------------------------------------T--------------A--------T-------------C----------------------G-C--------------G-----  3994
04_cpx.CY.94CY032-3  -------------------A-C---C-------C------------T-----------------------T--A--A--A---CAG---------------------G-------C--------------  3315
N.CM.YBF30           ----------T--------A-C---G--A--CAGT--G--------T--------------A--------T-----A--A-----G-----------T-T---G--AG----G--------G-----G--  3543
O.CM.MVP5180         -------------------A-C------T--T-----G-----------T-----------A--------T---A-T--A---C-------G--G-----------AG-ACA-----A---G--CA-AA-  3979
O.BE.ANT70           -----------AG-C----A-GT-----T--TA-G--G----------AT-----------A--------T--A-----A---C-------G--------------AG-ACA---G-A---G--GA-AA-  4003
CPZ.US.CPZUS         --C-T-----TA---------C---GG------AG--G--------A--T-----G-----A-----CC-AC-A--A--C--GAA------G---------------G----G--T---------A--A-  4009
CPZ.GA.CPZGAB        -----------AGT-------C------T---CC-ACCA----T--T-AT-----------A-----A--T-----A--A---ACT--G--------------A--AG-C-A---G-A-------A-CA-  4010
CPZ.CM.CAM3          -------------A--G--A-C----------CC------------G--------G-----------------A-----A--GAA------------C--C--C---G-------C-----G---A-A--  3360
CPZ.CM.CAM5          -------------A--G--ACC----------CC---G--G-----A--------------------------A--A--A--GAA------------T--CA----AG----G--C-----G---A-AA-  3651
CPZ.CD.CPZANT        G-C-C-----TA-TC--TT-GC---C--T--TCC--AG--T--------T-----T--------------T--AA-CT--C----G--G--G-----C-----G--AG-C---------AGT-GG--AAA  3388

H2A.GW.ALI Pol       R__L__A__F__N__L__V__K__D__P__I__L__G__A__E__T__F__Y__T__D__G__P__C__N__R__Q__S__K__E__G__K__A__G__Y__I__T__D__R__G__R__D__K__V__K   p15
H2A.GW.ALI           GGC--GCA-TTA-CC---TA-----T--T---C----C-----G--------CAC---C---C-CTG---T---C-AT-A---GA----------------A-A--AG-------G---G-C--G--GAA  4324

H2A.SN.HIV2ST        G----GTA-TTA-CC-G-T------T------C----C--------------CAC------AT-CTGC--T-A-C-AT-A-G-GA---------------CA-A--AG-----------G-C--G--GAG  3776
H2A.DE.BEN           GG---ACA-TTA-CC---TAGG---T--T---CC---C-----G--------CAC------AT--TGC--T--AC--T-A---GAG-----------------A--AG---------A-G-C-----AAA  4331
H2B.GH.D205          G-C--GCC---A-CC---TC-----C---C---A---GAG--------A---CAC-------T-CTGC--T--AACCT-A--GGA------------------C---G-C--G----A-G-T--G---AA  4304
H2B.CI.EHO           GG---GCC--TA-CC-G-TC-----C---C---A-----T------T-A---CAC------AT-CTG-----AA-CCT-A---GA---G--------------C--AG-C--G----AGG-T------AA  4302
H2G.CI.ABT96         GG---GTC---A-CC---TC--------TC---AGCAGA----G----AT---AC---C---T-CTGC--T-AAACCT-C--GGAR-----------C--C--C--AG-C-------A-G-C--G--AAA  3658
MAC.US.MM251         G----GTC-T-A-TC---T---G--C--T----AG----A--------AT---AC------AT--TG---T-AAC--T-A---GA---G------------A-C--AG----G--C-A-G-C-----AAA  4224
SMM.US.SIVSMMH9      G-C--GTT-T-A-CC---TA-----G--T---CAG-----------A--T---------A-AT-CTG---T---C--T-A-G-GA------------C-----G--AG-------C---GGC--GACAAA  3721
STM.US.STM           G----GTC-TTA-T----TA---------T-G-A----A-------A--T-----G-----CT--TG-------C-AT-----GA---G--------C--CA-A---G---------A-A-C-----AAA  3895
TANTALUS.UG.TAN1     ----------TA-C--G-TA-----T------CC-AA----A-GGTG-AT---TGC---------TGC--T--AA-TT-C-----------G-----C-TCC-A-----C-C----GA-TC-C-G--CA-  4148
GRIVET.ET.GRI-677    --C-G-----TTCC----TA--------A--CAA-----A---TGTT-AT-----G-----------ATC--AA-T---C--------T--G---------C-GT-AG-G-------A-AGT-G-A--AG  3952
VERVET.KE.VER-9063   ----G------ACTC-GACC--G-----T---CC-----AG--TGTA-AT--------------CTG---T--AA-CT-A---GA---------------CA----AC-ATAT----A---G-G---A-A  4176
VERVET.KE.AGMTYO     --C--------ACA---ACA------------CCCAAG-AG--CGTT-A-------...--A---TGC-----AA-TT-A---GA---------------CA-CT-AC-ATAC----A---G-G---A-A  3663
VERVET.KE.AGM155     CT--G-----TACAC-GACT--G--------CCCG----AG--TGT--A------------A--CTG---T--AC--T-G---GAG-----------C--CA-A--CC-ACA---C-A-----G---ACA  4168
VERVET.DE.AGM3       ----------TACCC-GACA------------CC-AAG-A---TGT--A------G-----A--TTG---T--AA-TT-A-GGGA----------------A-C--AC-ATAT--G-A-----GG--G-A  3671
SABAEUS.SN.SAB1C     G---G------A-A----TA-----T-----CCC-----A--C-GT--AT--------------------T--AA-TT-C---GA---------------CT-A--AG-------GGAT-----G--G--  4347
SUN.GA.SIVSUN        --GAG-----TACTC---CC-C------AG-GCC----...--C----A------G-----A------G-A-A-CTAGAA--GAG---G------------A-A--CC-GG--------TCC-G----AA  4303
MANDRIL.GA.MNDGB1    GGGA------TAGC--GTCCCC------TC---AG--G-T------A-AT-----T-----A--------------C-G----A-G---------------A----AG-------TTTT----GG--A-A  3636
LHOEST.KE.hoest      --GAG------ATT-GG---GC----------A----G-T---C----A------------A-----AG-A-A--TAGGA---AC----------------A-A--AC-ATC---G-A-G-G-----AAA  4304
LHOEST.CD.hoest447   -GGA------TAGC-----AGC------T----A---G-T---T----AT-----------A--T--AG-A-AA-TAGGG---AC------G-----T-----C--AC-ATC-----AGG-G------AA  3220
LHOEST.CD.hoest485   -GGA------TAGC-----AGC-----------A---G-T---T----AT-----------A--T--AG-A-AA-TAGGG---AC------------T-----C--AC-ATC-----AGG-G------AA  3220
LHOEST.CD.hoest524   -GGA------TAGC--G--AGC-----------A---TA----T----A------G-----A--T--AG-A--A-TAGG----AC------------G---A-A---C-ATC-----A-G-G---A-AAA  3217
RCM.NG.SIV-NGM       GGC--------A-TC---TA--G--T--T---CC-----A----GT--AT-----------------A--T---A-T-G---G--G-----G-----------A--AG----G----A-G----G--AAA  4298
COL.__.SIV-CGU1      -----GTA-GGA-C--G-TTTT---G--TC-GAAG-AG---CC---T-AT-GGAC------A-G-TGCTC----A-ATTAGG-GCG-----G-------GGA-A-A-TCA-A----GA-G--G----A--  3718

SYKES.KE.SYK173      GG-GG-AC--TA-C--G-TACC----------CC--A-----T---------C--------A--------T--A--C-GC--GAC------T-----------AG-A-G-GAT--G-C---G-G---ACA  4149
SYKES.KE.SYK173 Pol  R__W__Y__Y__N__L__V__P__E__P__I__P__E__A__V__T__F__Y__V__D__G__A__A__N__R__D__S__K__T__G__N__A__G__Y__V__A__S__D__G__T__Q__R__V__Q   p15
```

```
B.FR.HXB2 Pol       __T__L__T__D__T__T__N__Q__K__T__E__L__Q__A__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_    p15
B.FR.HXB2           CACCCTAACTGACACAACAAATCAGAAGACTGAGTTACAAGCAATTTATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCACAACCA   4079

A.UG.U455           TT----------G-----------A--------A-----T-----CC-------C-----------A----C---------T-----------------G-----------G--------G---------   3525
C.ET.ETH2220        TT-T--------A--------------------A--------G--CC-G-----G--A--------A----C---------------------------G-----------G-----C-TG---------   3471
D.CD.84ZR085        -C--T-T---------------------------------------A----------------------------------------------T------------C-----------------------   3601
F1.BE.VI850         -T-------G--G--C------------G--------------------T----------------A----C------G-----------------------------------------G--G------   3419
G.SE.SE6165         T-----------A--------C--A---G----A-----G------C-G-----------A--C--AA---C------------------G--------G------C-----------------------   3476
H.CF.90CF056        -T-------G--A--------------------A--------------------------A-----A--G--------G--------G-----T-----G------C----------------------C   3426
J.SE.SE7887         T--------------------------------AC----C--C------T-------A-G------A--GC----------------------T--------------G--G--T--A------------   3393
K.CM.MP535C         -T--A-------A--------------AG----A-----------C-G-T-------A--------A----C---------------------------G--------------------------G---   3275
01_AE.TH.CM240      TT----------G-----------A--------A-----T--G--CC--T----C-----------A----C---------T--------------------------------------G---------   3653
02_AG.NG.IBNG       TT----------G-----------A--------A-----T------C--T----C--A--------A----C---------T-----------------G--------------------G---------   3604
03_AB.RU.KAL153-2   -T--------------------------------------------C-----------------------------------------------------------------------------------   4124
04_cpx.CY.94CY032-3 -T-----T----A--------------------A--------------CT----------------A----C---------------------------G------A-----------------------   3445
N.CM.YBF30          -T-TA-TG-A-----C--C-----A---G-------------T--CCT-A-G--C--A--A--G--A---CGG--T-----------C--T-----T--G-----TA-G-----A-----TT----G---   3673
O.CM.MVP5180        A-AGT--GAA--G-----C-----A---G----A---ATG--TG-A-TAA----C-----------CAAGGAGC------------------------------T---G--C-----ATCCT-C------   4109
O.BE.ANT70          A-AAT--GA---G--C--C-----A--AG----A---ATG--G--A-TA-----C--A--------CAA-GA-AC------T-----------T--------------G--CG----CTCCT-C-----T   4133
CPZ.US.CPZUS        A--TT--GAAA-T--T--T-----A---G-A---C-CAC---GG-A--CT----A--AA-A-----A-AGAATAC---T--TG----C--T-G---C-------TC-----------C--CT-T--G---   4139
CPZ.GA.CPZGAB       T-G-T--GAAA-T--C--T------C-AG-A--A---A-G--TT-GCT---G--C-----------A-ATCA-C-G--T--------G--T-------------TG-----G--T-----GT----G---   4140
CPZ.CM.CAM3         A-A-T--GAAA-T--C--T--------AG-A--AC-GAC---TG----------A--A--A---G-A-A-CATAT---T--T-----C--T--T--C-------TG-----G--A-----CT-C--G---   3490
CPZ.CM.CAM5         A-ATT--GAAA-T--T--T--------AG-A--AC--AC---TG-------G--A---------G-A-A-CATAT---C--T-----T--T--T--T-------TGC-------A-----CT-T--G---   3781
CPZ.CD.CPZANT       ACA----CAAA-G--C--C-----AC-AG-A--A-----G--G---CT-A-G---C-AG-------AAC-GGCCC---C--T-----C-----T--------------G---G--T-G----GTACC---   3518

H2A.GW.ALI Pol      __V__L__E__Q__T__T__N__Q__Q__A__E__L__E__A__F__A__L__A__V__T__D__S__G__P__K__A__N__I__I__V__D__S__Q__Y__V__M__G__I__V__A__G__Q__P_    p15
H2A.GW.ALI          GGTA---GAAC-A--T--C------C-AG-A--A---G----CT-CGCG--G--AG-AACA--C--A--TCC-A---CC--T--TA--GT---T-----G----T-A-G-----AG-AGC--GC--G---   4454

H2A.SN.HIV2ST       GCTAT--GAGC-A--C--C------C-AG-A--A---G----CT--GCGA-G--AG-AACA--C--A--TCC-A-G-CC-----TA--GT--------------T-A-G-----AG-AGC--GC------   3906
H2A.DE.BEN          AGTAT--GAAC-A--T--C------C--G-A--A---G---TCT--CGGA-G--AC--GCA--C--A--CCC-A-G--T--T--CA--GT---T-----G----T-A-G--G--AG-AGC--GC--G---   4461
H2B.GH.D205         AGTGT--GAAC-G--------C--AC-AG-A--AC-TG-----T--GCAT----A--AACA--C--A-A-CC-C----T-----CA--GT---T----------TCA-G-----A--AGCT-----G---   4434
H2B.CI.EHO          AC-AT--GAAC-A------------C-AG-A---C-TG-----T--GCA-----AC-A-----C--A---CC-C-G--C--T--CA--GT---T----------TCA-G-----AG-AGCT---------   4432
H2G.CI.ABT96        RGT-T--GAGC-G--------C---C--G-R--AC-TG-G---T--GCAA-G--C--------C--A--TCC-A----C--T--CA--GTR--T--C--G----TCA-G--R-----AGC--GR------   3788
MAC.US.MM251        AGTGT--GAAC-G--T--T-----AC-AG-A--A---G-----T--CTCA-G--A---ACA--C--A--GCC-A-GAC---T--TA--GT---T----------TTA-G-----A--AAC--G-TGC--T   4354
SMM.US.SIVSMMH9     ACTG--GGAAC-G--T--C-----AC-AG-A-----GG----CT----------C--AGCA-----A---CC-A---C---T--TA--GT---T--C-------TTA-G--C--AG-AGC--GT-----C   3851
STM.US.STM          AG-TT--GAAC-G--T--T-----AC-AG-A--A---G-----T--GCCA-G--C---GC---C--A--TCC-A---C---T--T--GGT--------------TGA-G--------AAC--G---G--C   4025
TANTALUS.UG.TAN1    AGAAT--GAG--T--C--G--C--AC-AG-A--A---------G-CCT------A--AA-AC--AGTAAGAGCA----T--T--------G--T-----G----TCA-G------C-GAC---------T   4278
GRIVET.ET.GRI-677   GGAAT--GAAA----C--T--C--AC-AG-A--A---AC----G--A-GA-G--A---G----CAGT---GA-A-T-----T-----C-----T--T-------T-A-GAAC---T-GAC----TGT---   4082
VERVET.KE.VER-9063  G-AAT--GAGC-G--T-----C---C-AG-A--A--GGT---C--AA-AA-G--C---G-----AGT---GA-A-------------------T-----------CA-G-----AT-GAC------G--C   4306
VERVET.KE.AGMTYO    A--AT--GAAA----T--C------C-AG-AA-A---AC---T--AA-AA-G------G-A--CAGT--GCCTA-T--G-----------------T---------A-G-----TT-GAC---------C   3793
VERVET.KE.AGM155    ACAG---GAAA-------------AC-AG----AC-GAC---C--AA-AA-G--C---G-----AGC--CCCTA---C---T-----------T-----------GA-G--C--AT-GAC------G--C   4298
VERVET.DE.AGM3      A-AAT--GAAA-T--------C---C-AG-A--A---ATG--C--AA-AA-G--AC-AG-A---AGT--GCCTA-T-----------------T------------A-G-----AT-AACT--C-----C   3801
SABAEUS.SN.SAB1C    AG-AT--GAGA-T--T--C--C-----AG-A------G-G--C---CTGT----C--AAGA--C--T---AGCA-------T---A-------------------TA-G--T--T--AGC--GGG----T   4477
SUN.GA.SIVSUN       G-AAT-GGAAA----C--C------C-AG-A--AC--G-G--T--AA-AA-G--C---G-----AGTA-GAGCTC---T--T-----------------G-----C---A--C-GC-CAGTAAGAG---C   4433
MANDRIL.GA.MNDGB1   AGAATATCTAA-T--C--C------C----A--A-----T---G-AA-A------C-AG-A---AGT---AGTT-T--T--------------T----------T-G----T--AC-AGC-AGCAG---T   3766
LHOEST.KE.hoest     GGAGT----------C---------C-AG-A--AC--G-G---G--CTAA-G--A--A--A---AGCAATAGTA-------T-----------T----------T-A-GAA---AT-GTC-CA-AG----   4434
LHOEST.CD.hoest447  AGAAT----A-----C--C-----AC-AG-A--A---G----TG-C-TAT-G--A--AAGA--CAGTAAGAGCAGG--C-------------------------TGA--AA-G-TT-ATCTCA--GG--T   3350
LHOEST.CD.hoest485  AGAAT----A--T--C--C-----AC-AG-A--A---G-G--TG-C-TAT-G--A--AA-A--CAGTAA-AGCAGG--C-------------------------TGA-GAA-G--T-ATCTCA--GG--T   3350
LHOEST.CD.hoest524  AGAAT--CAA--T--T--T------C--G-A--A---G-G---G-A-TGT-------A--A---AGCAA--C-AC---T-----------T-----C-------T-A-GAAGG--T-AAGTCAGAG---T   3347
RCM.NG.SIV-NGM      AGAAT--GAAA-T--T--C-----A--AG-A-----GG-----G-A-TA--G--C--AA-A--C--A---AA-AGG-----------------T--T-G-C---TG--T--C--AT-AGC--G-AGT---   4428
COL.__.SIV-CGU1     --AT--TCA---AGGGT--------C-AG-A---C-GAC--G-G---TG------C-AA-AC--GG-CC-AA-AG-A-G--TT-G---------AGT-TT-----------G--A---ACT-GG--G---   3848

SYKES.KE.SYK173     GTAT--GGAGC-A--------C--AC--G-A--A---G-G-GCT-GCTCA-G--A--------C--AAA-GACA-------T------GT---T--------CT-T-AT--C--TC--ATGA--TGC--C   4279
SYKES.KE.SYK173 Pol __Y__L__E__Q__T__T__N__Q__Q__A__E__L__E__G__L__L__M__A__L__Q__D__S__K__D__K__V__N__I__V__V__D__S__Q__Y__S__Y__G__I__L__M__T__C__P_    p15
```

```
B.FR.HXB2 Pol       _D__Q__S__E__S__E__L__V__N__Q__I__I__E__Q__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__   p15
B.FR.HXB2           GATCAAAGTGAATCAGAGTTAGTCAATCAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTG   4209

A.UG.U455           --CAGG-----------AA--------------------A--C-----G----------A-----C---T-------------G--------G--------------G---------------------T   3655
C.ET.ETH2220        ---A-G------------A-------------------A----------GC-------G------C---T----------------T--------------------------------------A---A   3601
D.CD.84ZR085        ---A-G-----------A-------G------------A--------------------------C-----------------T---------------------------------------------A   3731
F1.BE.VI850         ---A-G-----------AA----------------------A------C---------G------C---T----------------------G-------------------------------------   3549
G.SE.SE6165         ---AGG------G----A--------------------A---C----------------------C---T----------------------G-----------------------C------------A   3606
H.CF.90CF056        ---A-G-----------------T---------------G-A--------G--------------C---T-------------------------------------------------------T---T   3556
J.SE.SE7887         ---A-G-----------A---------------------G----------G--------------C---T-G--------------------------------------G------------------T   3523
K.CM.MP535C         ---A-------------T-----T-----------------A----------------G-A----C--AT------------------------------------------------------------   3405
01_AE.TH.CM240      --CAGG-----------AG-------C------------G--C-------------G--A-----C---T-------------------G--G------------------------------------T   3783
02_AG.NG.IBNG       --CAGG---------------------------------A--C-------------C--A-----C---T----------------------G------------------------------------A   3734
03_AB.RU.KAL153-2   ---A-G-------------------G------------------G--------------------C-----------------------------------------------T----------------   4254
04_cpx.CY.94CY032-3 ---AG------------T-----T------------------------CGG-----C--------C---T----------------------G-----------------------------------CA   3575
N.CM.YBF30          ---A-------------A--G--G-GC-----------AG--C-C-------------GA--T-----CT-T--------T-----T-----T--------------G--G-----C--------T--CT   3803
O.CM.MVP5180        ACA--G-----C--CCCTA----TC-G--G---------G-AC---C----------CGA--G-----TA-------T--T--T--------C--A------------A--A-------------A--CA   4239
O.BE.ANT70          ACA--G--------CCCTA----TC-G------------G-AC---C----------C----G-----TA-------T--T--C--T-----C--A------------A--A-------------A--CA   4263
CPZ.US.CPZUS        --C--------------A--G--------G--------AG-A--------G--A-----AAGT---A-CT----------------T------------------------A--------C--------T   4269
CPZ.GA.CPZGAB       -----C-----------A-----------G--------AG-------T-----A-----AA----C--CT-C--------------------T--A-----------G-----G----------------   4270
CPZ.CM.CAM3         -----------G-----A--G--T----T---------AG----------G--------AAGT--C--TT-------G--------T------G-------------------------------T---T   3620
CPZ.CM.CAM5         -----------------A-----AGGGTCT--------AG-------------------AAGT--C--CT-------G------------------------------------------C----T---T   3911
CPZ.CD.CPZANT       ---------------CCCC----GG-AG-------CC--A-A----------GA---C--A-T--C--CT-C--------------T-----C--A--------------G-----C------------C   3648

H2A.GW.ALI Pol      _T__E__S__E__N__R__I__V__N__Q__I__I__E__E__M__I__K__K__E__A__I__Y__V__A__W__V__P__A__H__K__G__I__G__G__N__Q__E__V__D__H__L__V__S__   p15
H2A.GW.ALI          ACAG--TCA---AATAGAA----A--------C-----AG-AA-G------------GCCA-----G-T--G-----C-----C--------C--A--G------C-GG-------CC-T-----A---C   4584

H2A.SN.HIV2ST       ACAG-GTCA--GAGTA-AA----A--------C-----AG-AA-G------------GCAA-----G-T--------C-----C--T-----C--A---------C-GG-G------C-C-----A---C   4036
H2A.DE.BEN          ACAG-GTCA---AATAGAA----G--C--G--C-----AG-AA-G-----G------GCA------G-T--------C-----C--T-----C--A--------CC-GG-------CC-T-----A---C   4591
H2B.GH.D205         ACAG---CA------CCAA----AGCAA-------T--AG-AA-G--C-----A--GGCA--A---G-A-G------------T-----G---C-G--T--T---C-GG-------CC-CC----A---C   4564
H2B.CI.EHO          ACAG---CA------CC-A----A-GAG-------T--AG-AA-G--C-----------AA-A---G-A-G------------T-----G---C-G--T--T---C-GG-------CC-CC----G--CC   4562
H2G.CI.ABT96        ACGG---CA-----CCC------A---A-------T--AG-AA-R--T-----A---RCTC-A---G-A-G------------T--T--------A-----T---C--G-G-----CCRTC----A--CC   3918
MAC.US.MM251        ACAG--TCA--GAGCAG-C----T--C-----------AG-AA-G--T------TC-G-AA-T---G-A-----------------------T--A--------CC--G--A----CC-CC----T---C   4484
SMM.US.SIVSMMH9     AC-G--TCA---AGTAGA-----A--C--G---------A-AA-G--T---------GCAA-T---G-A-----------T-----T---RR---A----A----C--G--------C-TC----T---C   3981
STM.US.STM          ACAG--TCG--GAGCA-------A-----G--------AG-AA-G--T---------GCAA-T---G-A-----------------T--------A---------C-GG-G------C-T--G--A---C   4155
TANTALUS.UG.TAN1    ACA--------C--C-T-C----A-----------T--AG-AA-G--T--T--A--TGCTA-----T-AA-C--------------T-----G--A--G------C--G--A-------GC----A--CC   4408
GRIVET.ET.GRI-677   C-GG-----A-C---CCC-----GG-A--G------C-AGCCC----G------AGGC-------CT-ACA------------T--T--G--G--A-----C---AC-G--A-------------A--CA   4212
VERVET.KE.VER-9063  ACA--------C--CCCA-----AG-A--G--C----CATTAA-GG----G--AAGGGCAA-A--C--ACA------G--T-----------T--A--G---------G--A----C--------A---C   4436
VERVET.KE.AGMTYO    ACA--------T---CCA-----AG-G-----T----CCTTAA-G---C-----C--C-AA-A---T--CAG--------------T--------A-----------GG-GA-------------G---A   3923
VERVET.KE.AGM155    ACA--G-----C--CCCAC----AG-A----------CA---A-GG--C-G--A---GCCA--------CA---------T--T--T-----T--A--G--C------G--A----C--------A--CA   4428
VERVET.DE.AGM3      ACA--G-----C---CCC---A-AG-A-----T----CA-TAA-GG--C----AC-TC--A-A--CT--CA-------------G-------G--A-----C------G-GA---------C---A---C   3931
SABAEUS.SN.SAB1C    ACAG--TCA--TAATA-CA----AC-A-----T------G-AC-------G-----GGCA--G---A-A-----------T--C--T------G-A--T--C-----GG--A-T------C----T---C   4607
SUN.GA.SIVSUN       ACAG---CA--T------C----A--GG-G--CG-----TCA---GGC--------CC----A--C--A-GT--------G--------------A---------C--G--A-----C--C----T--CC   4563
MANDRIL.GA.MNDGB1   AC-G---CA--TCACCCCA----A--GG----------ATTAA-G-A-GG---A-----AA-T---T-AAGT---C----------------G--A-----G-----G---A-------GC----A---T   3896
LHOEST.KE.hoest     ACAG---CA---CACCCCA----A--AG-C--C-----A--A-GC-AGC-G--A--TC-A--T-----A-G------G--T--T--T--------A-----T---C--G-G-----CC-CC----A---A   4564
LHOEST.CD.hoest447  ACAG---CA---CATCCTG----A--A--G--T------G-A-GT-A---G--A--CC-A--G-----A-G------------T--------T--A--G------C--G-------CC-CC--------A   3480
LHOEST.CD.hoest485  ACAG---CA---CACCCTG----G--G--G--T------G-A-GT-AG--G--A--CC-A--A-----A-G------------T--------T--A--G------C--G-------CC-CC--------A   3480
LHOEST.CD.hoest524  ACAG---C---GCATCCTA----G--GA-C--C---A--G--C-GCAG--G--A--TGCA--T--C--A-G------------T--T-----T--A--G-----CC-GG-GA-----C-------T-T-A   3477
RCM.NG.SIV-NGM      --CAC------GAGCCCCC----AC-A--G--T-----A--AA-G---GGG------G-A--G--C--CT-C--------G-----T-----C-----G--------G--G--------------G---C   4558
COL.__.SIV-CGU1     T--G-TCAGCTTC--TCTAGTAGAGGGG-G--T--TC--GCAGGT--GGCT--A---GCCA-TC--G-A--C---TGT--------T-----G--A-------------TGA-----C--AAG--AG-A-   3978

SYKES.KE.SYK173     AC-A-C-CA---CACCCAA----GG-G--G------C-AG--GCC------------GCCA-A---G--A-------C-----C--T--------A-------------GCT-T------C--------A   4409
SYKES.KE.SYK173 Pol _T__N__T__E__H__P__I__V__E__Q__I__I__Q__E__A__I__K__K__E__A__I__Y__V__T__W__V__P__A__H__K__G__I__G__G__N__E__A__V__D__K__L__V__S__   p15
```

```
             Pol p66 RT and Pol p15 Rnase H end \/ Pol p31 integrase start
B.FR.HXB2 Pol       A__G__.__I__R__.__.__K__V__L__F__L__D__G__I__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   p31
B.FR.HXB2           CTGGA...ATCAGG......AAAGTACTATTTTTAGATGGAATAGAT...AAGGCCCAAGATGAACATGAGAAA...TATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGT  4324

A.UG.U455           -----...------......--G--G--------------G------...-----T-----G--C-----A---...------T-C--C--------------------------T-----------G--  3770
C.ET.ETH2220        G----...------......-----G--G---C--------------...-----T-----A--G-----A---...--------C-------------------A---A-----TA-C-----C-----  3716
D.CD.84ZR085        G----...------......--------------G------------...-----T-----G------------...--------C-----------------------------------------G--  3846
F1.BE.VI850         -----...G-----......---A----G-----------G------...-----A-----A--------A---...-------AC-----------G-----------------T---------A----  3664
G.SE.SE6165         G----...------......------------C-------C------...--A--------A--G-------G-...-------AC--C--------------C-----------TT--------A----  3721
H.CF.90CF056        -----...G----A......-----G------C-------G------...--A--T-----A--------A-GG...-----T-AC------------G----------------T--A------A----  3671
J.SE.SE7887         -----...------......-----G--G---C-------G------...--A--T-----A--T-----A---...-----T--C-----------------------------T--------------  3638
K.CM.MP535C         -----...------......-----G------C--------------...-----T-----A--------A---...-----T-AC--C--------------------------T---------A----  3520
01_AE.TH.CM240      -A---...------......--G--G--------------G------...-----T-----A--------A-G-...--------C---------A-------------------TT--------A----  3898
02_AG.NG.IBNG       A----...------......--G-----------------C------...--A--------A--G-----A-G-...--------C-------AG--------------------T---------A----  3849
03_AB.RU.KAL153-2   -----...------......G--------------------------...-----A-----A------------...------G-------------------------------------------G--  4369
04_cpx.CY.94CY032-3 A----...-----A......--G--G---------------------...-----T-----A------------...-------A---C--------------------------T------T-A--G--  3690
N.CM.YBF30          -A---...--T--A......---A--T----CC-------T-----A...--A--------A--T-----C-G-...--------C-------A---------C------------T-A--C--CA----  3918
O.CM.MVP5180        AA-AC...--T--A......-G---C--G--CC-G--A---------...C----A-----A--T-----A---...-----T---------------T-A--------C---GGAT-A-----AA----  4354
O.BE.ANT70          AG-AT...--T--A......-G---C--G--CC----A--------C...C----A-----A--T-----A---...-----T----------A----C-A--------A---GGA--A-----A--G--  4378
CPZ.US.CPZUS        -A--G...-----A......-----T--T--CC-------C------...--A--A--G--A--C-----C---...-----T----------C---T-----C--------------------A--G--  4384
CPZ.GA.CPZGAB       -A---...------......-----G-----CC-------------C...-G---------A--------A-GG...-----T----------A---T-----------------TT-A-----CA----  4385
CPZ.CM.CAM3         ----G..A-----A......-----T--C-----------T------..T---CA-A-G--A--------..GGAAA-----T-AC-----------T-----A--------C--TA-A----A-A----  3738
CPZ.CM.CAM5         -A--C...--T--A......-----C--C--C--------T--C---...-----A--G--A--------A--G...-----T-AC--C----A---T-----C--------C--T--A-----------  4026
CPZ.CD.CPZANT       AA--G...-----A......C----C--C---C-G--A--C------...--A--T-----G--C-----T---...-----T--C---------T--T-A--AGA---A-AC--T--T--C---A-T--  3763

H2A.GW.ALI Pol      Q__G__.__I__R__.__.__Q__V__L__F__L__E__K__I__E__.__P__A__Q__E__E__H__E__K__.__Y__H__S__N__V__K__E__L__S__H__K__F__G__L__P__N__L__V   p31
H2A.GW.ALI          AG--C...-----A......C-----T-G---C----AAA------G...CCA--T--G--A--------A--G...-----T--C---GT--A--A-C-AT-CCA-A-A---GGAT-A--CAA-C-G--  4699

H2A.SN.HIV2ST       AG--C...-----A......C-----T----CC----GAA------A...CCC--T--G--G--------A---...-----T--C---GTA-A--A-C-AT-CCA-A-A---GGA-----CAAAT----  4151
H2A.DE.BEN          AA--C...-----A......C-----T----CC-G--AAAG-----G...CCC--T-----G--------A---...-----T--C-T-ATA-A--A-C-AA-CCA-A-A---GGAA-A--C-T-C----  4706
H2B.GH.D205         AA---...-----A......C-G--CT-G--CC----AAA------A...CCA-----G--A--G-----A---...-----TG-C---GTA-A--A-C---TCCA-A-A--CGGAA-T----AAT----  4679
H2B.CI.EHO          AA---...--T--A......C--A-C-----C-----AAA------A...CCA--T-----A--------A---...-----T-A----GTA-A--A-C-A-TCCA-A-A---GGGA-T----AAT----  4677
H2G.CI.ABT96        AG---...--T--A......C----------C--GR-AAA------A...CCT--T-----A--------A--R...-T---T------GTA-A--AGT-AA--CA-A-G---GG-A-T--T-AAT----  4033
MAC.US.MM251        AG--G...--T--A......C----T--C--C--G--AAAG-----G...CCA--A-----A--------T---...--C--T------GTA-A--A-T---TATTCA-A---GGAT-A--CAGAA----  4599
SMM.US.SIVSMMH9     AA-A-...--T--A......C----C-----C-----AAA------A...CCA--G--G--A--G-----A---...--C--T--C---GTA-A-AA-T---TATTCA-A---RGTT-A--TAGGC----  4096
STM.US.STM          AG---...--T--A......C-G--C-----CC----AAA------A...CCA--T-----A--G-----A---...-----T--C---GT--A--AGC-A-TCTTCA-G---GG-A-A---AGGC----  4270
TANTALUS.UG.TAN1    AG--G...--T---......C----G-----C-----CA----G--A...G-A-----G--ATC------T---...-----T-C-------CA-TTT--AAGAGA--CA--CGGTA-A---G-AC-G--  4523
GRIVET.ET.GRI-677   AA---...--A--A......C-GA-C--C--C------A-------A...G-A--A--------C----CA--G...--C--T-AC---------AGT----TACAG--A---GGAT-A--TAA-A----  4327
VERVET.KE.VER-9063  AA---...--A--A......-G---T------A---GAA-------A...G-A--T-----A--------C-G-...-----T----------A-AATT-A--GGACACA---GGAT-A----AGA----  4551
VERVET.KE.AGMTYO    AA--C...--T--A......-G---TT----C-----AAA------A...G-A--T-----AA-G-----A-G-...-----T-A--------A-AACC-A--AGA-ACA-A-GGG--T----AAA----  4038
VERVET.KE.AGM155    AG---...G-T--A......-G-A--T-G--CA-T-GCA-G-----A...G-A--A-----A--------T-GG...-----------C------AATC-A--AGACACA---GGAT------AAA----  4543
VERVET.DE.AGM3      AA--G...--G---......---A-TT----------AAA------A...G-A-----G--G--------A-GG...--C--T-A---------GAACT-A--AGACAC--A-GGG--A----AAA-T--  4046
SABAEUS.SN.SAB1C    AA---...--A--A......C----------CC-G--CA----T--A...G-A--A-----A--G-----T---...--C--TGCC--------GAGT---CAACAA--A---GG-T-A--TG--A----  4722
SUN.GA.SIVSUN       AG---...-----AAAGAGAC-G---A-G---A----AAAG-----A...CCA--TGT---A-------G---G...-TC--T-AC---GCAGCCT-TT--CAGGAGATG---G--A-A--C-TA--G--  4684
MANDRIL.GA.MNDGB1   -A---...-----A......-----CT----CC--C-AAAT-----A...CCA--A--G--A------------...-----T--C---GAAGC-CA-T-AAGAGAGA-A--CC--T-A---G--C----  4011
LHOEST.KE.hoest     AA--T...--A--ACAGAAAC-G--CA-G---C----AAAG-----A...CCT--AGT---A------AGC---...-TC--T-AC---GCA-AG-ATC-A-AAGAAA-A--------A--C--CA-G--  4685
LHOEST.CD.hoest447  AA---...--T--ACAAAAAC-------G---C---GAAAG-----G...CCC--A--G--A--GT--AGC---...-T---T-A----GCT-A--ATT-A-AAGAAA-A---C--A-A--C---A-G--  3601
LHOEST.CD.hoest485  AA---...--T--ACAAATAC-------G---C----AAA------G...CCA--A--G--A--G---AGC---...-T---T-A----GCT-A--ATC-A-AAGAAA-A---C----A--C---A-G--  3601
LHOEST.CD.hoest524  AA---...--T---ACAAAAC-G--CA-G---C-G--GAAG-----A...CCA--TGT---A--G----GA--G...-T-----A----GCA-A--ACT-A-AAGAAA-G-A-GGGA-A--C--CA-G--  3598
RCM.NG.SIV-NGM      AG---...--T--A......C-G--------CC----C--G----GA...--A--A--G--A--------A---...-----T--C-----------------AGAA--A---C-AA-A--T-AAA----  4673
COL.__.SIV-CGU1     TAA--...CAGGTA......-TGTGGA--GA-AA-ATA-A-GC--CA...G-A......--A--C---C-A--G...-T---T------GT-CAGTATC-AAAAGAAC-A---GGGT-A--CA-------  4087

SYKES.KE.SYK173     AA---...--A--A......---A-C-----CC----AA----TCCC...C-A--A--G--A--T-----A-GG...--C---TCA---AT-GAGTATC-TAGGCAG--G---C-T--A--C-GACAG--  4524
SYKES.KE.SYK173 Pol K__G__.__I__R__.__.__K__I__L__F__L__E__R__I__P__.__Q__A__Q__E__D__H__E__R__.__Y__H__S__N__M__E__Y__L__R__Q__E__F__H__L__P__R__Q__V   p31
```

```
B.FR.HXB2 Pol        __A__K__E__I__V__A__S__C__D__K__C__Q__L__K__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_   p31
B.FR.HXB2            AGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCTGGTA  4454

A.UG.U455            ---G--G------------------A----------A--------G-----------------------------------G---------T-------C------C-------------A-----A---  3900
C.ET.ETH2220         -C-C--------------TT-------------------------G--------A------------A-T-----------G---------T-------------C--------G---A-C---------  3846
D.CD.84ZR085         ------------------T--------------------------------A---------------------------------------T----C---------------------------------  3976
F1.BE.VI850          ------------------T-----------G--------------G-----------C-----------------------G---------T-------C--------------C--G------------  3794
G.SE.SE6165          ---------------G-----------------------------G---------------------------------------------T--------------------------A-----A-A---  3851
H.CF.90CF056         ------------------T--------------------------G-----------------------------C---------------T-------C--------G------C-------T------  3801
J.SE.SE7887          ---------G--------T-----------------------G--G-----------------------------------G---------T--------------C-----------------------  3768
K.CM.MP535C          ---C--------------T--------C-----------------G--------A------------------------------------T--------------------------A-----------  3650
01_AE.TH.CM240       ------G---------A---A---------------A--------G-----T-----------------------------G---------T-------C------C-------------C---------  4028
02_AG.NG.IBNG        ---------------G---------------------A-G-----G---------------------------G-----------------T----------------------G---A-----T-A---  3979
03_AB.RU.KAL153-2    ------------------------------------AT-------------------C------------------------------------------------------------A-------A---  4499
04_cpx.CY.94CY032-3  ---------G--------T------A-------------------G--------------------G--------------G--------GT----------------------T---------A-----  3820
N.CM.YBF30           G--------------------------C-----C-----------G-----------------G--CA-T------------G-G------T-------------C-----G------A-C-----T--G  4048
O.CM.MVP5180         ---C--G-----CA-T--T--T---CC------C--TA-------G-----A-C------T----------AC--C----AG---------A-G--------------------C---A-C--AA-A--T  4484
O.BE.ANT70           G--C--G-----CA-T--T------CC---------TA-------G-----A--T-----T--G--------C-------A-G--------A-------C------A-G-----C---A-C--AA-A--T  4508
CPZ.US.CPZUS         C--T--G--G-----G----------------------C---------G-----A-----G--G-----T--C--------T--C-----G-----------------------G---A-C--TA-A--G  4514
CPZ.GA.CPZGAB        ---------------G---CAT--------G--C---G----------------------G--G-----------C-----G--T------G----------C--CC-------C-----G---A-A--G  4515
CPZ.CM.CAM3          G--C------C--------C-T-----------C--A-----------G-----------G-----------C--C-----C---------T----C--C-----CC----------------------G  3868
CPZ.CM.CAM5          G--C-----------------T-----C--G--C---T-------G-----------------------------------C--T------T----C---------C-----------A-C-----A--G  4156
CPZ.CD.CPZANT        G--T--------TA----ACAG--------------CG----G----------G------------------C------------------G-------C--C-----------T-----A---A-A---  3893

H2A.GW.ALI Pol       __A__R__Q__I__V__N__T__C__A__Q__C__Q__Q__K__G__E__A__I__H__G__Q__V__N__A__E__L__G__T__W__Q__M__D__C__T__H__L__E__G__K__V__I__I__I_   p31
H2A.GW.ALI           G----G-C-G------AA--CA----CCC---------AG--G-----G--T--A-----G-----GA-TGCAGAA-T---C-CT------A-G--C--C-----C--------------C---A-AA--  4829

H2A.SN.HIV2ST        G----G-C--------AA--CA--CACCC---------AG-----G--G--T--A-----G------A-TGCAGAATT---C-CT------A-G--C--C-----C------------A-C--TA-A---  4281
H2A.DE.BEN           -----G-C-G------AA-TCA----CCC----C--A-AG--------------A-----G------A-TGCAGAAAT---CG-T------A-G--C-AC-----C------------A-C--TA-A---  4836
H2B.GH.D205          G------C-G------AATTC------------C--A-A------G-----T--T--------G---A-TGCAGAC-T---G-C------GA-G--C---------------------A----AA-A--G  4809
H2B.CI.EHO           G----G-C--------AATTC------------C--A-A------G-----T--T--------G---A-T-CAGAA-T---G-C-------A-G--C--------------G-----G-----AA-A--G  4807
H2G.CI.ABT96         -------C-G------AATTCA-ACC-C-GT--C--A-AG--------------T--------G---A-TGCAGAATTG--T-CT------A-G--C--C--C-----------------A---A-A---  4163
MAC.US.MM251         G--C-G-C-G-------A--C---------------T-AG-----------T--A-----G--G---A-T-CAGA--T---G-CT------A-G--C-----C---C-----------A-AG--A-A--T  4729
SMM.US.SIVSMMH9      -------C---------A--CA-----------C--C--G-----------Y--A-----G------A-TGCRGRA-T---G-CT------A-G--C--------CC----G--C---A-A--TA-A--T  4226
STM.US.STM           ------GC-------G-AT-CA-----C--G--C--C-AG-----------T--A-----G------A-TGCAGAGTT---G-CT-----GA-G--C---------C-------C--GA-A--TA-A---  4400
TANTALUS.UG.TAN1     ------------------TGCT--CCCA--------AA-TCG-------C-G--A-----G--------TGCC---GT---GG-C------A-G--C-----------G---------A-C--TA-A---  4653
GRIVET.ET.GRI-677    ---------G--------GGCA---CCC-----C--AA---G-------C-T-A---------G------GCCTCCATT-A--CT-----GA-G--C--C--C--------------------AA-A---  4457
VERVET.KE.VER-9063   T--C-----G--------A-TG---CCA--TGCCA-AA-------G---C-A--A-----G---------GCCTCC-----GG--------A-G--C--C--G---G-----------A-AG--A-A---  4681
VERVET.KE.AGMTYO     ---------G-----G----TG---CCA--------AA----G------C-AG-------------G--TGCCTCA--T----C------GA-G--------T---C-----AA------AG--A-A--T  4168
VERVET.KE.AGM155     ---T--------T-----A-TG--CCCA--------AG-------G---C-A--A--------------TGC-TCA------G-G-----GA-G--C--C------A-----------A-AG-GA-A---  4673
VERVET.DE.AGM3       G-------------------TG---CCA---------A-------G---C-AG-C-----G--------TGCCTCG-----GG--------A-G--C-----------------C--G--A---A-A---  4176
SABAEUS.SN.SAB1C     ------------------GGCA---CC------C---A-------G--GT-TG-A-----G--------TGC----------G-G------A-G-----------CC-----------A-C---A-A--G  4852
SUN.GA.SIVSUN        ------GC-G-----CAATGAA----CAC-------A-AG-----G--------AACA-----------TGCCTCAGTG--C--C-----GA-------C--C--CA-G----A------C--AA-A---  4814
MANDRIL.GA.MNDGB1    ---C---C-G--T--GCAA--T--CAG---G--CTGT-ATC-T-----GC----AA-G-----GAC---TGC-TCA-TT---G-C-----GA-------C------C-G---AATC--A----TA-A---  4141
LHOEST.KE.hoest      ---C---C----T--CAATGA-----CA--C-----AAAG--------------AACA--------G--TGTCTCAGTG--T--T-----G-----C-----T--C---------C-G--C--TA-AAAT  4815
LHOEST.CD.hoest447   ---T--GC-------GAAT--T---A-----------AC---G-----------AACA-----------TAC-TCA-----T--T------A----C--C--T--CA-G--G---C-G--A---A-TAAT  3731
LHOEST.CD.hoest485   ---T--GC-------GAA---T---A-----------AC---G-----------AACA-----------TAC-TCA-----T--T------A----C--C--T--CA-G--G---C-G--A---A-TAAT  3731
LHOEST.CD.hoest524   ---T--GC--------AATGAG----C---G-------AG--G-----------AACA--T--------TGTATCAGTG--T--T--------G--------T--CC-------GC--A-C---A--AAT  3728
RCM.NG.SIV-NGM       -------------------CAA--CCCC--G------G-------------AG-A--------G-----TGC---------T-CT------A-G-----C------C-----------A-C---A-A---  4803
COL.__.SIV-CGU1      ------------TTGG-AA--A---TCAG-G------AAC------C----TG-T-----C--GC----T-A---CTAT--CT--------T----------T---GAG--------G------T-A---  4217

SYKES.KE.SYK173      --------C---TA--CAACAA---CCA--------AAAT-G-------C-A-AA-----G--------TGTAGA-AT-TAT-AT-----GA-G--C-----T---GAG-----C-----A---TGT---  4654
SYKES.KE.SYK173 Pol  __A__K__A__I__I__Q__Q__C__P__K__C__Q__N__R__G__E__P__K__H__G__Q__V__D__V__D__I__Y__N__W__Q__M__D__C__T__H__E__E__G__K__V__I__C__V_   p31
```

```
B.FR.HXB2 Pol       _A__V__H__V__A__S__G__Y__I__E__A__E__V__I__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__    p31
B.FR.HXB2           GCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAAACAGGGCAGGAAACAGCATATTTTCTTTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACTGACAATGGCA   4584

A.UG.U455           -----C--------------C--C-----------------C--------------A-----G--------C---A-AC-----------------------------GT------C--A----------   4030
C.ET.ETH2220        -----C--------------C--C-----G-----G--------------------A--A-----------C-----AC-------------G-----------C-GGGT---------A--T-------   3976
D.CD.84ZR085        --------------------C------------------------------------------G-------C-----C------------------------------GT-G-G-----A----------   4106
F1.BE.VI850         -----C--------T----AG--C--------------------------------A-----------C--C--CA-A-----G-------------------------T---------A----------   3924
G.SE.SE6165         --------------------C-----------------------------------A--------------C---A-A-----------------G----------C-GT---------A--T-------   3981
H.CF.90CF056        -----C--------------C-----------------C--C--------------AA-------------C--C--G--G---C-----A-C---------------GT---------A----------   3931
J.SE.SE7887         --------------------C--------------------C--------------A------G------T----A-A--------------C---------------GT---------A----------   3898
K.CM.MP535C         -----C--------T--C--C--------------------C-----------G--A--A--------C--C---A-AC-----------------------------GT---------A----------   3780
01_AE.TH.CM240      -----C--C--G-----------------------------C--------------A-----G--------C-----GC-----C-----------------------GT------C--A-----C--T-   4158
02_AG.NG.IBNG       -----C--------------C--------------------C--------------A-----G--------C---A-A--------------------------G---GT------C--A----------   4109
03_AB.RU.KAL153-2   --------------------------------------------------------A--------------C---G-C-------------------------------T---------A----------   4629
04_cpx.CY.94CY032-3 -----------G--T--------C-----------------C--------------A-----------C--C--CA-AC-------------------------G----TG------G-A-----C---C   3950
N.CM.YBF30          --G--C-----G--------C--CT-------------------T-----G-----A------------------A-------G-----T------------------GTT-----C-----T-----AT   4178
O.CM.MVP5180        --T--C--------A----AC-T------------G--G--A--------------A--------T--C-----C--G-------------C---------T--C---GT---------A--------AC   4614
O.BE.ANT70          --T--C-----G--A-----G-TC--------------G--A--------------A--A-----T--C--C--C--G------C-G--T-C---------T--T---GT---------A-----C--GC   4638
CPZ.US.CPZUS        -----A-----G-----------CC----------G--C-----T-----------A--A--G-----C------A-C-----------------------T------GTG--T--------T-----AT   4644
CPZ.GA.CPZGAB       --------C-----------C--------------------C-----T--G-----A--A--------T-----C--G--------------T-----------------T--T--C--A--T-----GC   4645
CPZ.CM.CAM3         -----A--------------G--CC----G-----G--C-----T-----G--T--A-----------T------A--C----------------------T------GT---C--C-----T-----GC   3998
CPZ.CM.CAM5         -----A-----G-----------CC----G-----G--C-----T-----------A--A--------T------A-------------------------T------GT------------T-----GC   4286
CPZ.CD.CPZANT       -----G------T-T-----C-TC--------T-----A--GG---AT--G------A-AAGT--------C--C--G-----------CA-C-----------------------C-----------AG   4023

H2A.GW.ALI Pol      _A__V__H__V__A__S__G__F__I__E__A__E__V__I__P__Q__E__S__G__R__Q__T__A__L__F__L__L__K__L__A__S__R__W__P__I__T__H__L__H__T__D__S__G__    p31
H2A.GW.ALI          -----G-----T----------TC-----------G--C--C--GCAG---T----AAGAC--------CTC--C--A------C-G--TA-T---------A---C-CACT-G--C--A--T-G---TG   4959

H2A.SN.HIV2ST       -----A-----T--A-------T---------------C--C---CAG---T----AAG-C----G---CTC--C--AC-----C-G--CA-T--G------A---C-CATT-G--C--A--------TG   4411
H2A.DE.BEN          -----A-----T--A-------TC--------------C--C---CAG---T----AAG-C-G------CTC--C--A------C-G--CA-T--G------A---CGCACT-G--C--A---------C   4966
H2B.GH.D205         -----C--------------G-T------------G--A--A--CCA---G-----AAGAC-G-----TCTC--C--AC----G--G--CA-C--------TA-C-C-CACC----C--A-----C--TG   4939
H2B.CI.EHO          ----------------------TC--------------A--A--CCA---------AAGAC-G-----TCTC--C--G-----GC-G--CA-C--------TA-C-C-CACC-G--C--A-----C--TG   4937
H2G.CI.ABT96        -----------G--------C-T-G-------------A--C---CA---------AAG-C---------TA--C--G--------------G--------TA-C-C-CACC-------A--------TG   4293
MAC.US.MM251        -----A--------T-------TC--------------A------CA---G-----AAGAC-G------CTA-----G--------G-----C--------TA-T-C-CATC----C--A--T-----TG   4859
SMM.US.SIVSMMH9     -----A-----G-GT-------TC-----G--------A--C--GCRG--------AAGAC--------CTG-----GC----------CA-C--------CA-C-C-CATC-G--------T-----TG   4356
STM.US.STM          -----A--------T-----G-T------G--------A--C---CA----------AGAC-G-------TG-----G--G--G------A-C--------T--C-C-CACC-G--C--G--T------G   4530
TANTALUS.UG.TAN1    -----C-----G--------C-TCT-------------C--AG-TAGG--------AA-A---------A-A--C--G------A-CATTA-T--G-----TA-T-C--A-T----------------GC   4783
GRIVET.ET.GRI-677   -----A----------------TC-----------G--G--C---AG------T---A----G------C-C-----GC-G---C-GTT--C------------G---CATC----C-----T------C   4587
VERVET.KE.VER-9063  -----------G-----C----TC--------------C--A--TAGG--------AAGA--G-----CA-A--C---C-----A--TTGA-T---------A--GTCCAGT----C--A--T-----AC   4811
VERVET.KE.AGMTYO    --G--C----------------TC--------------C--A--TAGG--------AA-A-----G---A-G-----A------A--CTGA-T--------TA---C-CAGT----C--A--------GC   4298
VERVET.KE.AGM155    --G--C--------------G-T------------G-----C--TAGG--------AA-A--G------A-G--CT-G------A--AT------------CA-C-CTCACC-C-----A--T-----AC   4803
VERVET.DE.AGM3      -----C----------------TC-----------------A--TAG----------A-A---------A-A---T-A-----GA--CT-A-T--------CA---CCCA-C-G-----A--------AC   4306
SABAEUS.SN.SAB1C    --T--------T--T-------T---------------A--C---CAG--------AA-A-CT------C-C--C--GC-G--G-----CA-C--------CA-C-C-CAGC-G-----A--T-----T-   4982
SUN.GA.SIVSUN       --G--G-----G--TTCA--G-----G-T---------AC-A--CAAT--GCA----A--AC------CAC--GG--G-----GC-CTGT-CC-TG------------CAG--------A--T-----GC   4944
MANDRIL.GA.MNDGB1   -----G--------TTCA--C-TC--GA-G-----------AA----------T--AA-AA-G------G-G-----G-----G-------C-CA------TA-T-GT-A-C----C--A--T-----GC   4271
LHOEST.KE.hoest     -----C-----G---TCA--C-TC--G-TT-----------A----AT--G-----AA-AAC---GT--C----C--G---------TGCA-T--------T------CA---------A--------TC   4945
LHOEST.CD.hoest447  -----------G---TC-----T---G-T---------A-------AT---------A-AAC-------C-C--C--G------C-GTGCA-C------------C-GCA-G-C-----A--------AC   3861
LHOEST.CD.hoest485  ---A-C-----G---T------T---G-T---------A-------AT---------A-AAC-------A-C--C--G------C-GTGCA-C------------C-GCA-G-C-----A--------AC   3861
LHOEST.CD.hoest524  -----C---A-G--ATCA--C-TC--G-TT--------G-------AT---------A-AACT------A-----T-G-----G---TGTA-C--------T------CA---T--C--A--T-----TC   3858
RCM.NG.SIV-NGM      -----C-----------------C--------------A--A--------G-----AA-A--G-----TC----C--G-------------CC-----------G---CATT----C--A--------A-   4933
COL.__.SIV-CGU1     -----A--C--CTGT-C-CTG-TCTGCTGG---ACCA-AT-AAAGAG----------G-A-------GTAGGGCC--CA-------G--TA-TCAG---GAG--T-G-CA-G-T--C--A--T-----GC   4347

SYKES.KE.SYK173     -----AA-CAC---T---------------A-CA-GA-CT-AAA-AGG-----G---G-T----------TG-----CA-GC--A----CA-T---------A-----CAG-----C-----T-----AC   4784
SYKES.KE.SYK173 Pol _A__V__N__T__A__S__G__Y__I__E__T__K__I__L__K__R__E__T__G__D__E__T__A__L__F__L__M__Q__I__A__S__R__W__P__I__K__Q__I__H__T__D__N__G__    p31
```

```
B.FR.HXB2 Pol       S__N__F__T__G__A__T__V__R__A__A__C__W__W__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    p31
B.FR.HXB2           GCAATTTCACCGGTGCTACGGTTAGGGCCGCCTGTTGGTGGGCGGGAATCAAGCAGGAATTTGGAATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATTAAAGAAAATTAT   4714

A.UG.U455           -----------A-C---G-A----AA--A-T------------AAAT---C-A-----------G--C-----------------------------G-----C-----C--G--------------C--   4160
C.ET.ETH2220        -T--C------A--AA-G-A----AA--A--------------A--T--TC-A--------------------------------------------------------------------------C--   4106
D.CD.84ZR085        -----------A-------A----A---------C--------A--T-----------------------------------------------------------------------------------   4236
F1.BE.VI850         -----------A----CG------A---AT-------------A--T---C-A--------------C--------------------------------------A--------G-----A--G--C--   4054
G.SE.SE6165         -----------A-----G-A--A-A---A--A-----------AAAT----CA--------------------------------C--------------------------G--------------C--   4111
H.CF.90CF056        ----------GA-----G------A---A--------------A-AT---C-A-----------G-----------------------G-----------------------------------G--C--   4061
J.SE.SE7887         ----C------A---G-G-T--G-A---A--------------A-AT--------------------------------------------------------------------------------C--   4028
K.CM.MP535C         C----------A-CA--GT-----A---A--------------A--TG----------------G--------------------C-----------------------------------------C--   3910
01_AE.TH.CM240      -----------A-C---G-A----AA--A--------------CAATG--C-A-----------G--C--------------------------------------------G--------------C--   4288
02_AG.NG.IBNG       -----------A-----G-A--A-A---G--A-----------AAATG---CA--A-----------------------------C-----------G--------------------------------   4239
03_AB.RU.KAL153-2   -----------A--A--G------A---T--------------A--G-----------------C------------------------------------------------C--------C---C---   4759
04_cpx.CY.94CY032-3 C----------A-----G------A---A--------------A-AT-----C--------------------------------C-----------G-----------------------------C--   4080
N.CM.YBF30          C---------TA----C--T--A-AA--A--------------AAAT-----A-----------G--A-----------T--------G----C------G--C--------------------------   4308
O.CM.MVP5180        CT-----T--AA----AG-CA-G-AA--T--A---------A-A--C--AC-A--T--G-----G--A--A--T-----A---------------------G-C-----------------ATCT-----   4744
O.BE.ANT70          CT-----T--AA--A-A--TA-G-A---T--A-----------CAAC--AC-A--T--G--------A--A--T-----A---------------------G-C--------G--------ATC------   4768
CPZ.US.CPZUS        CT--C--T--AA--AG---A-----A--A--T--C--------A--C--AC-A--A--G-----------A--------A--------------G--------C-----------------------C--   4774
CPZ.GA.CPZGAB       CA-----T--AA-----G-A--C-A---T--------------A-AC--------------------A-----T-----A--G-----------G--------CT-A--------GC----------A--   4775
CPZ.CM.CAM3         CT--C--T---A--AA---A----A---A-----C--------A--C---C-A--------------------T-----A--------------G--------A--------G--------A--G--C--   4128
CPZ.CM.CAM5         CT--C--T--TA--AAC--A--C-AA--A--------------T--C---C-A-----G--------C-----------A-----------------G-----C-----------------------C--   4416
CPZ.CD.CPZANT       CT--------AA----AG-A--A-AA--G--A-----------TAAT---C-A--------------A--A--------A-----------------G-----C---------C----G---C-------   4153

H2A.GW.ALI Pol      V__N__F__T__S__Q__E__V__K__M__V__A__W__W__V__G__I__E__Q__S__F__G__V__P__Y__N__P__Q__S__Q__G__V__V__E__A__M__N__H__H__L__K__N__Q__I    p31
H2A.GW.ALI          T---C------TCACAGGAA--A-A-ATG-TAGCA-------TT--T--AG----ATCC------G-A--T--------A-----C---------------G-A------C-CC-CC----A--TCAG--   5089

H2A.SN.HIV2ST       C---C-----TTCACAGGAA--A-A-ATG-TGGCA------ATA--T--AG-A--ATCC--C---G-A--T--------A-----C-----------G---G-A------C-CC-CC----A--TCAG--   4541
H2A.DE.BEN          C---C-----TTCACAGGAA--G-A-ATG-TGGCA-------TA--T--AG-A--ATCC------G-A--T-----C--A-----C--G------------G-A------C-CC-CC-------TCAG--   5096
H2B.GH.D205         C---C------TCAC-A-GT--A-A-ATG-TAGCC-------TA-----AG-A--AACT------G-A-----T--C--A-----------------G---G-A-----CC-TC-CC-G--A--TCAA--   5069
H2B.CI.EHO          C---C-----TTCACAAGAT--G-AAATG--AGCC------ATA--G--AG-A--AAC---C---G-G-----T-----AG-------G------------G-A-----CC-TC-TC-G--A--TCAG--   5067
H2G.CI.ABT96        C---C------TCGCAAGA---G-A-ATG-TAGCC-------TA--C---G----A-C-------G-A--A-----C-----G--------G-----------A------C-TC-TC--------CAA--   4423
MAC.US.MM251        CT--C--TG--TCGCAAGAA--A-A-ATG-TTGCA--------A--G--AG----CACC-----GG-A--A--------A--G-----G--------G---G-A------C-CC-CC-G--A--TCAA--   4989
SMM.US.SIVSMMH9     C---------ATCACAAGAA--G-AAATG-TTGC---------A--R--TG-A---ACC---R-GG-A--T--T-----A--G--C---R-G-----G---G-G-----CC-TC-TC----A-CCCAA--   4486
STM.US.STM          CT--C------TCGCAGGA---A-A-ATG-TAGCC--------------AG-A--AACC------G-A--T--------A--G--C-----------G---G-C------C-TC-T-------CCCAG--   4660
TANTALUS.UG.TAN1    CA-----TGTTA--CAGGA---ACAAA-TATT----------GACA-G-AG----CAC-ACA--G--C-----T--C--A---TCA------AGT------AG--------G-C-------AG---C---   4913
GRIVET.ET.GRI-677   CA--C--T---TC-CAG-AT--GGCA--G-TG--C-------GTAAT--AG----CACCAC------A--T--T--C--A--GTCA--G--TAGT------AGC-----C-G-C-GC-C---G----C--   4717
VERVET.KE.VER-9063  CA-----T--TTCCCAAGA---AGC---AAT-----------GAAAG--AG----CACCACA---G-G--A--------A--GTCA--G---TC-------AGC-----C---C--C-C--AG----A--   4941
VERVET.KE.AGMTYO    CT--C--T---TCCCAAGAA--GGCA--AATA----------GAAA---TG-A--TAC-ACA--T--A--A--T--C------TC-------TC-A-----AGC-----C--GC-------AG-G--A--   4428
VERVET.KE.AGM155    CA--------TTC-CAGGAA--AGCT--TATG--C-------GAAAGG-AG-A--CAC-ACG--GG-A--A--T-----A--GTCC--G---TCTA-----AG------C---C----G--AG-G--A--   4933
VERVET.DE.AGM3      C------T--GTC-CAAGAA--AGCA--AATG----------GAAA---AG-A--CACCACA--TG-A-----T--C--T---TCA-----CTCTA----GAG----------C-G-----AG----A--   4436
SABAEUS.SN.SAB1C    C---C--T--TA-CCAGCAA--AGCA--TATT--C-------GAAA---AG----CAC------GG-A--------C-----------G-----------------------GC-G-----AG----C--   5112
SUN.GA.SIVSUN       CA-------T-A--AAAGAT--GGA---A-TA---------TTA--C--AC----TAC-ACA--G--A--T-----C--T--G--------------G---G-A-A---C--G-T------AC-------   5074
MANDRIL.GA.MNDGB1   CT--C--T--TA--CAGGAA--AGAAA--ATG---------TTA--G--AG-A--CAC---------A-----T--C--A--------G--G-----G---AA--AA-----GT-TC----AG--T-G--   4401
LHOEST.KE.hoest     CA-----TGTAA--AAGGA---ACA---A-TAACC------ATA--G--AG-A--TAC-AC---G--A--T--T--C--T--G--------G--T------G-A-AA--C----TT-----A-GT-----   5075
LHOEST.CD.hoest447  CA--C--TGT-A-CAAGGA---ACAA----TAAT-------CTA--T--TG-A--TAC-ACA--------T--T--C--T--G-----G--------T---G-C-AA-------C-C----AC-------   3991
LHOEST.CD.hoest485  CA--C--TGT-A-CAAGGA---ACAA--A-TAACC------CTA--T--TG-A--CTC-ACA--G-----T-----------G-----G--------T---G-C-AA-------TGC-----C----C--   3991
LHOEST.CD.hoest524  CA-----TGTGA--AAAGAT--ACA---A-TGACC------TTAA-T--AG----CAC-ACA--G--A--A--T--C--T--------------------GG-A-AA-------TT------C----C--   3988
RCM.NG.SIV-NGM      CA--C--T--AA----AG-A--GCAA--A-A------------TCA---AG----CACC------G-A-----------T-----C--------G--------A------C-TC----G----T---A--   5063
COL.__.SIV-CGU1     CA-----TGTAA--CAGCACT---A---A--AGTG------CTA-----AGCA--CACCACA---CAC-----T-----T---TCA--------T-----GCAG-GA--C--G--TG-T--A-G--AG--   4477

SYKES.KE.SYK173     CA--C--TGT-A---AC-A-T-C-AA--A--A---------TGT--C--AG-A--CACTACA--T-----A-----C--A-----C--G---A-------G--C-AA---GGT-T------AG--GC---   4914
SYKES.KE.SYK173 Pol P__N__F__V__S__D__K__F__K__A__A__C__W__W__C__G__I__E__H__T__T__G__I__P__Y__N__P__Q__S__Q__G__I__V__E__S__K__N__R__Y__L__K__E__A__I    p31
```

```
B.FR.HXB2 Pol       __G__Q__V__R__D__Q__A__E__H__L__K__T__A__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_   p31
B.FR.HXB2           AGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCA  4844

A.UG.U455           ---G-----------G--A--------C--------------------------------T--------------------------------A---------------------A--------------  4290
C.ET.ETH2220        ---G-----------A--A-----G--C--------------------------------T----------------G-------------------------------------A----T---------  4236
D.CD.84ZR085        ------------------A-----------------------------------------T-----------G------------------------------------------A--------------  4366
F1.BE.VI850         ------------------A------------------------------------------------------------------------------------------------A-----------T--  4184
G.SE.SE6165         C--G-----C--G-----A--------------------------G--------------T--------------------------------------------A---------A--------------  4241
H.CF.90CF056        ---G-----------C--A--A-----C--------------------------------T------------------------------------------------------A--------------  4191
J.SE.SE7887         ---------------A--A--------C--------------------------------A--------------------------------A---------------------A--------------  4158
K.CM.MP535C         ------------G-----------------C--A--------------------------------------------------------------------------------------T---------  4040
01_AE.TH.CM240      ---G-----------G--A--------C-----A--------------------------T------------------------------------------------------A--------------  4418
02_AG.NG.IBNG       ---------C--------A--------C-----------------G--------------T------------------------------------------------------A----T---------  4369
03_AB.RU.KAL153-2   ------------------A------------------------------------------------------------------------------------------------A--------------  4889
04_cpx.CY.94CY032-3 ---G-----C--G-----A--------C-----------------G--------------T------------------------------------------------------A--------------  4210
N.CM.YBF30          ------AA-C--------A--A--------A-----------G--------G--T-----T-------------------------------------C----------------A--------------  4438
O.CM.MVP5180        -CAG-----G--G--C--A--A--G---T-A--A--------------------C--TG-T-------------------------------------C---------G---C--A--------C-----  4874
O.BE.ANT70          -CAG-----G--G--C--A--A-----CT-A-GA-----------------------TG-T-------------------------------------C------A-----G---A--------T-----  4898
CPZ.US.CPZUS        ------AA-C--G-----A--A--G---T-A--------T-----G--------------T-----------------------------------T-C------A--------CA-------C------  4904
CPZ.GA.CPZGAB       ---------T--G-----A--A------T-A--A--------------------G-----T-------------------------------------C----------------A--------------  4905
CPZ.CM.CAM3         ------AA-T-----A--A--A-----CT-G-G---------C--G--------G-----T-------------------------------------C------A--------TA-------C------  4258
CPZ.CM.CAM5         -------A-T-----C--A--A--G--G--A--------T--C-----------C-----T-----------------------------------T-C------A-----G--TA-------C------  4546
CPZ.CD.CPZANT       ------AA-T-----C--A--A-----AT-A------------GT---------GCA---T-------------------------------------CAC-T--AC-G------T--------C-----  4283

H2A.GW.ALI Pol      __S__R__I__R__E__Q__A__N__T__V__E__T__I__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__S_   p31
H2A.GW.ALI          -A-CAGAA-T-----G-----AA-TACAG-GG-A---AT-----T---------TCATTG-ATG-------------GG--A--A--A----ATATG-CCC---CA------C-CA-CA-T--G--CT-C  5219

H2A.SN.HIV2ST       -A-CAGAA-T-----G-----AA-CACAG-AG-A---AT-----T---------TCATTG-ATG-----------G-GG--A--A--A----ATATG-CCC---CA------C--A-CA-T--GG-CA-T  4671
H2A.DE.BEN          -A-TAGAA-T-----A-----AA-TACAA-AG-A---AT-----TG--------TCATTG-ATG-------------GG--A--A--A----ATATG-CCC---CA------C--A-CA----G--TA-C  5226
H2B.GH.D205         --ACAGAC-C-----C--A--A-T-TCAA-AG------TT----T-------ACTCA-TG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA------C----TA----G---A-C  5199
H2B.CI.EHO          --ACAGAA-T-----------A-T-TCAA-AG------TT--GTT-------ACTCA-TG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA-----------CA----G---A-T  5197
H2G.CI.ABT96        --ATA-AA-T--------A--AA-TTCAA-AG-A--CAT---GTT----------CATTG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA---------A-TA----G--TA-T  4553
MAC.US.MM251        --ATAGAA-C--G--A--A--AA-TTCAG-AG-A--CAT----TT---------TCATTG-ATG-------------GG--A--A--A----ATATG-C-C---CA------T--A-TA----G--CA-T  5119
SMM.US.SIVSMMH9     --ATAGAA-T-----A-----AA-TTCAA-AG-A--TAT-----T-------A-TCATTG-AT--------------GG--A--A--A----ATATG-C-C---CA------T----CA-T--G--TA-C  4616
STM.US.STM          --ACAGAA-T--------A--AA-CACAG-AG-A--T-T-----TC--------TCA-TG-ATG--------G-A--GG--A--AC-A----ATATG-CCC---CA------T----TA-T--G--CA-C  4790
TANTALUS.UG.TAN1    --A-A-AA-T-----AG-CTG--C-TT-T-GG-A--G-----C-TC-----CTGCCA---T-----------------G--A--A--A-----TATG-CCC---CA--G--GC--A--A----G---A-C  5043
GRIVET.ET.GRI-677   CTCT--AA--------G-TTG---GAGAT-GG----------G--------TACGCAT--------------------G--A--A--A-----TAT-TC-AGT-CA------T-G--TA-T--GC--A--  4847
VERVET.KE.VER-9063  T---A-AA--------G-TTG-C-GT--ACAG-A----------TT------TGCCA---------------------G--A--A--A-------TA-CAC---CA------C--A-TA-T--G---A--  5071
VERVET.KE.AGMTYO    T--GA-AA--------G-TTGCC--T--ACAG--G---------TG-----TTGCA-AC-T-----------------G--A--A--A-----AC-G-C-T---CA--G---C--A-TA-T------A--  4558
VERVET.KE.AGM155    T---A-AA--------G-CTG-C--T--ACAG-A----------TT-----CTGCCA---T-----------------G--A--A--A------CTA-CA--T-CA--G---C--A--A-T--G---A--  5063
VERVET.DE.AGM3      T--GA-AA--------G-CTG-C--T-CACAG-A----------TT------TGCCA---------------------G--A--A--A-------TA-CAC-G-CA--G---T--A-CA-T--G--TA-T  4566
SABAEUS.SN.SAB1C    ------AA-T------G-T--A---AGAT-GG-A--------CAT---------GCA---T------------------------------------TC-----CA------C--A-CA-T------CAT  5242
SUN.GA.SIVSUN       CA-CAGAA-T-----AG-T--AC-GG-AT-G--A--T-------TC-----TC-GCA---T--T---------CA--GG--A--AC-A-----AATG-CA----CA--G--GT-TA-CA----G--TAAT  5204
MANDRIL.GA.MNDGB1   T-AGA-AA-------AG-TTGCA--G-AT-A--A---------GCC-----CAC------T--T---------CA--GG--A--AC-A------ATG-CA-----A--G---------A-T--G--CAAT  4531
LHOEST.KE.hoest     --A-AGA--------AG-T--AC-G--A--A--A----------T---------CCA---T--T---------CA--GG--A--AC-A------CTT-CAC---CA--G--GT-TA-TA-T--G--TAAT  5205
LHOEST.CD.hoest447  --A-AG---C-----AG-T---C----AT-A--------T--C-T-------AC-CA---T--T---------CA--GG--A--AC-A------CT-TCAC---CA-----GT-TA-TA-T--G--TAAT  4121
LHOEST.CD.hoest485  --A-AG---T-----AG-T---C----AT-A--------T--C-T----------CA---T--T---------CA--GG--A--AC-A------CTTTCAC---CA-----GT-TA-TA-T--G--TAAT  4121
LHOEST.CD.hoest524  --A-AGA--------AG-T--AC-G--G--A-----------GTT---------CCA---T--T---------CA--GG--A--AC-G-------TATCAC---CA--G---TACA-CA-T--G--CAAT  4118
RCM.NG.SIV-NGM      ---G--AA-T-----------A---A-GT-AG----------T-----------GC----T--------------------------------------------A-----G---A--------------  5193
COL.__.SIV-CGU1     TAAGA-AA-G-A------A--A---ACAT-AG-A-GCAA----GC---------C-ATGCG-T---------------G--A--AT-A------A-G---C--T--------CA------A-G-GC-AT-  4607

SYKES.KE.SYK173     CTC---AA-T-----CG-C-TCAC----T-GC-A--------GGC-------ACC------TTA--------------G--A--A--A------ATATCAC----A-----GTACA-TA-T--GC--TAT  5044
SYKES.KE.SYK173 Pol __S__Q__I__R__D__D__V__T__H__L__Q__T__A__V__A__M__A__T__F__I__L__N__F__K__R__K__G__G__I__G__G__I__S__P__G__E__R__Y__I__N__M__L__Y_   p31
```

```
B.FR.HXB2 Pol       _T__D__I__Q__T__K__E__L__Q__K__Q__I__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__    p31
B.FR.HXB2           ACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAAGGTGAAGGGGCAG   4974

A.UG.U455           ------------------------------------T-----------------------------------------G-C--CA-------------------A--A----------------------   4420
C.ET.ETH2220        T----------G---------C-C-----C------TT------------------------------A---------G-C--TA----------------C--A--A----------------------   4366
D.CD.84ZR085        T----T----------G------------------C------------------------------------------G-----A----------------------T----------------------   4496
F1.BE.VI850         ----------------G---------------------G--------------C--------------------------C---G----------------------A----------------------   4314
G.SE.SE6165         T----T---------------C--------------------------------------------------------G-C---G----------------C--A--A----------------------   4371
H.CF.90CF056        ------------------------------------T----C--------A---------------------------G-C---A-------------------A-------------------------   4321
J.SE.SE7887         ----------------G-------------------------------------------------------------G-C---A-------------------A--A-C--------------------   4288
K.CM.MP535C         ------------------------------------TT---C--------A---------------------------G-A---A----------------------A----------------------   4170
01_AE.TH.CM240      ------------------------------------------------------------------------------G-C---A-------------------A--A----------------------   4548
02_AG.NG.IBNG       T----T---------------C----------G----T----------------------------------------G-C--CA-------------------A--A----------------------   4499
03_AB.RU.KAL153-2   -------------------------------------T------------------------------A---------G-----A-------------------A--A----------------------   5019
04_cpx.CY.94CY032-3 T----T---------------C--------------------------------------------------------G-A---A-------G-----------A--A----------------------   4340
N.CM.YBF30          -----------G--A-C-A-T-------C-------TT----G-------------------------A---------G----CA----------------C--A--T--G--------A----------   4568
O.CM.MVP5180        T--C-A--------A-C----C--------------TT------CA-C-----------C--------A--T------G-C--TA-------------G---C-A-----G-----------G-------   5004
O.BE.ANT70          T--C-A--------A-C-------------------TT----N----C--A--------C--------A---------G-C--TA-C-----------G---C-------G-----------G--A----   5028
CPZ.US.CPZUS        T----AC-------AG-CTT----------------TT----G----------------C------------------G-----A----------------C--A--T--G-------------------   5034
CPZ.GA.CPZGAB       --------------A-GC------------------TT----G-------A-----------------A---------G-C---A-------------G----CCT-A--G-------------------   5035
CPZ.CM.CAM3         --------------A-CTA-TC--------------TT----G----------------C------------------G----CA-------------T--C-GA--T--G-------------------   4388
CPZ.CM.CAM5         --------------A-CTA-T---------------TT----G-----------------------------------G-----A-------------T--C-GA--T--G-------------------   4676
CPZ.CD.CPZANT       -----------G--A-CTC----------T------TT----------C-A---------C----------T------G-C--TG-G---------------C-A--T--G-------------------   4413

H2A.GW.ALI Pol      _T__E__Q__E__I__Q__F__L__Q__T__K__N__L__K__F__K__N__F__P__V__Y__Y__R__E__G__R__D__Q__L__W__K__G__P__G__E__L__L__W__K__G__D__G__A__    p31
H2A.GW.ALI          -----ACA-G---TAC--TTCC-C----C-A---A-TTG---T--A-------C-C---C-----------AG-----G---AG--G-----------T-GGG----A--G--------G--C--A----   5349

H2A.SN.HIV2ST       G----ACAGG---TAC--TTCC-C---GC-A---A-T-----T-A--------------C----T---A--AG-----G---AG--G-----------T-GGG-A--A--G-----G--G--C--A----   4801
H2A.DE.BEN          -----ACA-G---TAC--TTCC-C----G-A---A-T----TT--A-------C-A---C--------A--AG-----G---AG--G-----------T-GTG-A--A--G-----G--A-----A----   5356
H2B.GH.D205         -----GCA-G---TAC-GTTC--C---GC-A---A-TT----T----------C-A---C--------A--AG-----G---A---C-----G-----T-GTG-A--AT-G--------G-----A----   5329
H2B.CI.EHO          -----ACA-G---TAC--TTCC-C----C-A---A-TT----T-C--------C-----C--------A--AG-----G---A---C-----G-----T-GTG-T--AT-G--------G-----A----   5327
H2G.CI.ABT96        -----ACA-G---TAC--TTTCA-----C-A---A-T-----T--A-------------C--------A--AG-----G-C-AG--C-----------T-GTG----TT-G--------G-----A--T-   4683
MAC.US.MM251        -----ACA-G---TAC--TTTCA----TC-A---ACT-----T--A-------------C--------A--AG-----G---A---G-----G-----C-GTG----AT-G--------G-----A----   5249
SMM.US.SIVSMMH9     -----ACA-G---TAC--TTCCA----TC-A---A-T-----T--A-------------C--------A--AG-----G-C-AG--G-----------C-GTG----AT-G--------G-----A----   4746
STM.US.STM          -----ACA-G---TAC--TTTCA----TC-A---A-T-----T-CA-------------C-----------AG-----G---A---A-----G-----C-GTG----AT-G--------G-----A----   4920
TANTALUS.UG.TAN1    ---C-AC--G---TAC--C-CA------C----CAACA-------TC-------AA---C--C-----A--AG-A---G-G---G-G--------------C-CT-----------G--A-----A--G-   5173
GRIVET.ET.GRI-677   ---C-AC--G--CTA--TACTC-------C-----CCA-------TTG------AA---C--C-----A--AG-T---G-----G-G--------------GCGA---A----------A-----C--G-   4977
VERVET.KE.VER-9063  ---C-AT--G--CTAC--AC--------CCA-----CA-------TT-------A-A--C--C-----A--AG-G---G-----G-G-----G---------C-AT-AA----------G-----T----   5201
VERVET.KE.AGMTYO    ---C-AT--G---TAC--C-T-------CCA-----CA-------TT-------A-A--C--C-----A--AG-G---G-C--TG-G-----------G---C-AT-AA----------G-----A----   4688
VERVET.KE.AGM155    ---C-AT--G---TC--CACTC------CCA----CCA-------TTG------A-A--C--C-----A--AG-----G-----G-G-----G-----T-TCGC---GA----------A-----C--G-   5193
VERVET.DE.AGM3      ---C-AT--G--TTAC--C-CC------CCA-----CA-------TT-------A-A--C--C-----A--AG-G---G----TG-C-------------G-C--T-AA-T--------G-----T----   4696
SABAEUS.SN.SAB1C    -----AC--G----C---AC-C-T---C--A-----T----------------------C-----------AG-----G-C--TG-G--------T--G--T------A----------A-----A----   5372
SUN.GA.SIVSUN       G----TC--G----AC--T-C-----G---AT--A-T--------TT---A---AA---C--C-----A--AG-A---G----T-AG-----------T--T--A--T--G--------G-----A----   5334
MANDRIL.GA.MNDGB1   -----AT--G--TA-C--T-TCA------T------T------A-TT-------AA------C-T---A--AG-A---G---AG--G-----------T-GT-TC--TT-G--------A----------   4661
LHOEST.KE.hoest     G----AC--G----AC--T-TC--------TT--A-T--------TT---A---AA------------AC-AG-A---G----T-AG--------------GC-A--T--G--------A-----T--T-   5335
LHOEST.CD.hoest447  G-----T--G----AC--T-T----------T--A-T--------TT---A---AA------C-----AC-AG-AC--G----C-AG-----G-----T---C----T--G--------A-----T--T-   4251
LHOEST.CD.hoest485  G-----T--G----AC--T-T----------T--A-T--------TT---A---AA------C-----AC-AG-AC--G----T-AG-----G-----T---C-AA-T--G--------A-----T--T-   4251
LHOEST.CD.hoest524  G----TT--G-----C--T-CA--------TT--A-T--------TT---A---AA------C-----AC-AG-A---G------AG----G------T--TC-A--T--G--------A-----T--T-   4248
RCM.NG.SIV-NGM      T-----T-------A-CCA----------C------T---------------------------T---A--AG-A--GG---AG-AG-----G--------T--A---A-----G----A--------T-   5323
COL.__.SIV-CGU1     -T---AT--G-T--AC--A-CC--AC----TT-CAA-ATC---AATTT--AAA-TTTAAGGCC---T--A-AGAGCAC-CAGG-GAG---C---------G-G-A--GG-G--------------A--G-   4737

SYKES.KE.SYK173     -----AC-----TTAC--C--AATAC--C-TC-...C-----T--TCG------A-------C---C-CC-G...G-G--GAACGAG-----G--------TCGT--A--T--------------T----   5168
SYKES.KE.SYK173 Pol _T__E__L__Q__L__Q__Q__N__T__T__S__.__P__K__F__S__N__F__R__V__Y__Y__R__Q__.__G__K__N__E__W__K__G__P__A__R__L__L__W__K__G__E__G__A__    p31
```

```
                                                                                                                          Pol p31 integrase end <-|
B.FR.HXB2 Pol       V__V__I__Q__D__N__S__.__D__I__K__V__V__P__R__R__K__A__K__I__I__R__D__Y__-__K__Q__M__A__-__D__D__C__V__A__S__R__Q__D__E__D__$          p31
                                                                                          |-> Vif start
B.FR.HXB2 Vif                                                                             M__E__N__R__W__Q__V__M__I__V__W__Q__V__D__R__M__R__I__R__.__    Vif
B.FR.HXB2           TAGTAATACAAGATAATAGT...GACATAAAAGTAGTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAG.....   5096

A.UG.U455           -------------C------...--T-----G-----A------------------------------------------------------------A-----G-------------A------.....   4542
C.ET.ETH2220        --------------------...--------G-----A-----G--G--------A----------------------------------C-------------G-------------A------.....   4488
D.CD.84ZR085        -------------C------...--------G-----A--------------------------------------------------------------------------------------A.....   4618
F1.BE.VI850         ----C--------C------...--A-----GA----A-----G--G----------------------------------------T----------------G-------------------A.....   4436
G.SE.SE6165         -------------C----AC...--A-----G-----A------------------------------------------------------------------G--------------------.....   4493
H.CF.90CF056        -------------C------...--A-----------A---------G-G-----A--------------------------------------------------------------------A.....   4443
J.SE.SE7887         ----------G--C------...--A-----G-----A--------------------------A---------------------------------------G-------------------A.....   4410
K.CM.MP535C         -------------C--C---...--A-----G-----A--------------------T---------------------------------------------G-------------------A.....   4292
01_AE.TH.CM240      -------------C------...--T-----------A------------------------------------------------------------------G--------------------.....   4670
02_AG.NG.IBNG       ----------G--C------...--T-----G-----A-----------------A------------------------------------------------G--------------------.....   4621
03_AB.RU.KAL153-2   -G--------G--C----AC...--T-----------A---------------------------------------------------------------------------------------.....   5141
04_cpx.CY.94CY032-3 ----------G--C--C---...--T--C--------A--------------------T--------C-----C---------------A--------------G-------------A------.....   4462
N.CM.YBF30          -G-----T--------CG-G...--T-----------C---C-T--G--------A--A----------------------------------G-------------G----------AA--C--GAAAT   4695
O.CM.MVP5180        ----C-----------AG-A...-----T-----G--A-----------G-----A--A--C--A------------------------AC----A--A------A-------ACA--AAG-G-AAGCAT   5131
O.BE.ANT70          ----C-----------GG-A...-----T--G-----A--------G--G-----A--A--C--A--G---------------------AC----A--A--------G-----ACA--AAG-G-AAGCGT   5155
CPZ.US.CPZUS        -------CA-G--A--CGAG...--GG-T--------A--C--------------A--T--A-AA--C----------------------C----A--A-----G--------------AG---AGGCTT   5161
CPZ.GA.CPZGAB       ----G--C--G---C-AG-G...--AC----G-----A--------G-----------------A------------------------------------------------A----------AGACTT   5162
CPZ.CM.CAM3         -G-----TA----G-GAGAG...--AG-T------A-A--C-----G--------A--A--------------------------T---------A--A-----G--G-----------AG-C-AGGCCT   4515
CPZ.CM.CAM5         ----G--TA----G--CGAG...--AG-T--------A--C--------G--T--A--A--C--A----------------------------G-A--A-----G--------A--------C-AAACCT   4803
CPZ.CD.CPZANT       ----C--CA----CC-AGAG...--A--T--G-----C--TC----G-----------T--A-A---G---A-----A----A-A--A-AGA---GA-T-----G--------A----------ATATCT   4540

H2A.GW.ALI Pol      V__I__V__K__V__G__T__.__D__I__K__V__V__P__R__R__K__A__K__I__I__R__D__Y__G__G__R__Q__E__L__D__S__G__P__H__L__E__G__A__R__E__D__.__G    p31
H2A.GW.ALI Vif                                                                            M__E__E__G__K__S__W__I__V__V__P__T__W__R__V__P__G__R__M__.__    Vif
H2A.GW.ALI          -CA--G-TA-G-TAGGG-CA...--------------A--------G--G--C--------C--A--C------GG-AG-CAA-AGCTG---AG-G--CCCCACTTGGAGGGT-CCAG---GG-T...GG   5473

H2A.SN.HIV2ST       -CA--G-CA-G-TAGGGGC-...---------A--A-A--------G-----T--------C-AA--C------GG-AG-CAA-AGATG---AGCG--TCCAACTTGGAGGGT-CCAG---GG-T...GG   4925
H2A.DE.BEN          -CA--G-CA-G-TAGGG-CA...--------------A--------G--G--C-----T--C-----C------GG-AGACA--A-CTG---AG-A--CCCCACCTGGAGGGT-CCAG---GG-T...GG   5480
H2B.GH.D205         -CA-C---A-G-TAGGG-CA...--A--C--------A--C--G-----------A--T--A---C-C------GG-GGA-AA-G-TTG---TG-A---CC-AC-TGGAGG-TACCAG-C-GGCT...AG   5453
H2B.CI.EHO          -CA-C---A-G-TAGGG-CA...--A--C------A-A--C--------------------A---A-C------GG-GGA-AA-A-TTG---TGCA---CC-ACGTGGAGG-TACCAT-C-GGCT...AG   5451
H2G.CI.ABT96        -CA-C---A-G-TAGGG-CA...--A--C-----T--A--C--------------A-----A-AA---------GG-GGA-AA-AG-TG---AGCA--ACCAAC-TGGAGG-TACCAG-C-GACT...GG   4807
MAC.US.MM251        -CA-CT--A-G-TAGGG-CA...-----T--G-----A--C--------G--T--A--T--C-AA---------GG-GGA-AA-AG-TG---AGCA--TCCCAC-TGGAGG-TACC-GA--GGCT...AG   5373
SMM.US.SIVSMMH9     -CA-CC--A-G-TAGGG-CA...--G--C--G-----A-----G--G-----T--A--T--C-AA--C------GG-GGA-AA-A-TTG---AG-G--TCCCACTTGGAGG-TACC-G---GGCT...AG   4870
STM.US.STM          -CA-CG--A-G-TAGGG-CA...-----T--------A-----G-----G--T--A-----C-AA---------GG-GGA-AA-AG-TG---AG-G--TCCCACCTGGAGG-TACCAG---GGCT...GG   5044
TANTALUS.UG.TAN1    ----C---A----AGG-GAG...--GT-G--------A-----G-----G-----A--T--C-AA--A------G-GAGA-AAA-T-TG-G---CAAGACTAAC-TGGAGGGTT-G..............   5286
GRIVET.ET.GRI-677   -G-----TA----GGGGGAA...-----C--G-----C--C--G-----G--T-----T--C-AA---------G-GAGA-AAA--ATG---AG-GAG-GTAGT-TGGAGGGT-TCAGA--GGCAAATAA   5104
VERVET.KE.VER-9063  -G--C--CA----AGGAGAA...--GC-G---A-T--C-----------------A-----C-AA-------A-CC-AGA-AAAG-ATG-G----GAG-GTAAC-TGGA-GGTCTCAG--GACCTGATAA   5328
VERVET.KE.AGMTYO    -G--CC-CA-G--CGGA---...---C----G--T--A--------G-----T--A--T----A--------A-CCCA-ACAAAG--TG-G-A--GAG-GT-ACGTGGA-GGTACCAG--GA-CTGATAA   4815
VERVET.KE.AGM155    ----TC-CA-G--AGG-GAA...--AC-G--G-----T--G-----G--------A-----A-AA--C----AGCC-AGA-AAA--TTG-G----GAGACTCACCTGGA-GGT-CA-GA-GAAGTGATCA   5320
VERVET.DE.AGM3      -G--C--CA---GAGG-GTG...--AT------A-TAC--------G--------A--T--A-A--------A-CC-AGA-AAAG-ATG-G----GAGAGTAACTTGGA-GGT-CC-GA-GAGCTGATAA   4823
SABAEUS.SN.SAB1C    ----T--------AC-GG--...--AT-G---AC-A-C--------------C--A--T--A-AA-----------.......---TTG---AG-CAG-CCC-TCTGGA-GGTA-C-G-CGGACAGCAGG   5493
SUN.GA.SIVSUN       ----C--CA--C-GGG-GAG...A----CCTG-----A--T--G-----------AC-AG-A-A---C------GG-...............--A-CCTCCT--GTGGAGAT--TG-GATGA........   5438
MANDRIL.GA.MNDGB1   ----TT--A--T--C-AGAA...--G-----GA----A--T--------GTGT--A--A--A-AA---------G-G...-GT-G-AAGA--AG-CAG--TAACTTGGA-AGT-TCT--...........   4774
LHOEST.KE.hoest     ----GG--A----AGGGGAG...A----CTTCTC---C--C--------------AC-AG-A-AA---------G-G...GGAC--AAG---AG-GAGAGTAGCCTGGATA-CA-CT--...........   5448
LHOEST.CD.hoest447  ----GG--A----AGGAGAC...AC---CTTCTC---C--C-----------T---C-AG-A-AA---------G-G...GGAC--AAG---AG-GAG-GT-GCCTACATA-CA-CT--...........   4364
LHOEST.CD.hoest485  ----GG--A----AGGAGA-...AC---CTTCTC---C--C-----------T---C-AG-A-AA---------G-G...GGAC--AAG---AG-GAG-GTAGCCTGGATA-CA-CT--...........   4364
LHOEST.CD.hoest524  ----GG--A-G--GGGAGAG...A----TTTCTCTA-C--C--G--G--------AT-AG-G-AA---------G-G...GGAC--AAG---AG-CAGAGTAGC-TGGATA-CC-AT--...........   4361
RCM.NG.SIV-NGM      -C--G--C--G---C-AGAG...--TC----G-----C--T--------GTGT--------A-AA--------C-G-A-AGA--TG-A-AGA--GACCAATATGGAAG-TAGAC-G-----AAGTTGA..   5448
COL.__.SIV-CGU1     ----C--CAG-A--TC-CAAGGTAC-T-GTTT--GAAA--T----------T--------CC--ACTCC-ATATGG-G-AGAT-TG--CAG---GAA-T--TT-TCC-ATGG-C-AA----AGCCGAAAC   4867

SYKES.KE.SYK173     ----GG-G---ACAG-GGAGGGA-----TTTT-C---A-----------G-----A--A----CA--CC-----G--......AG-ATG---AG-G--TCCCACGTGGAGA-T--CCCCA-GACAGATTG   5292
SYKES.KE.SYK173 Pol V__V__V__Q__T__E__E__G__D__I__F__A__V__P__R__R__K__A__K__I__I__T__D__H__G__E__.__.__R__M__D__S__G__S__H__V__E__N__D__P__K__T__D__$    p31
SYKES.KE.SYK173 Vif                                                                       M__E__K__.__.__E__W__I__V__V__P__T__W__R__M__T__P__R__Q__I__    Vif
```

```
B.FR.HXB2 Vif       __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__W__K__S__L__V__K__H__H_    Vif
B.FR.HXB2           ......................................................................................................AACATGGAAAAGTTTAGTAAAACACCAT   5124

A.UG.U455           ......................................................................................................---------C--------------T--C   4570
C.ET.ETH2220        ......................................................................................................---------T-----------G------   4516
D.CD.84ZR085        ......................................................................................................C---------------------T-----   4646
F1.BE.VI850         ......................................................................................................C---------------------T-T---   4464
G.SE.SE6165         ......................................................................................................-------C-C-----G--------T---   4521
H.CF.90CF056        ......................................................................................................C-----------C--------GT-----   4471
J.SE.SE7887         ......................................................................................................C--------------------GT-----   4438
K.CM.MP535C         ......................................................................................................T--T------------------T-----   4320
01_AE.TH.CM240      ......................................................................................................---------C--------------T---   4698
02_AG.NG.IBNG       ......................................................................................................---------T--------------T---   4649
03_AB.RU.KAL153-2   ......................................................................................................---------C------------------   5169
04_cpx.CY.94CY032-3 ......................................................................................................---------C---C----G-----T---   4490
N.CM.YBF30          GGAATAG...........................................................................................................C-----------T---   4718
O.CM.MVP5180        GGAACAGCCTGGTGAAATACCATAA.........................................................................................................   5156
O.BE.ANT70          GGAACAGCCTAGTGAAATACCATAA.........................................................................................................   5180
CPZ.US.CPZUS        GGAATAG...........................................................................................................-C-G------T-T--C   5184
CPZ.GA.CPZGAB       GGAATAG...........................................................................................................------T---T-T---   5185
CPZ.CM.CAM3         GGAATAG...........................................................................................................CC--------TGT---   4538
CPZ.CM.CAM5         GGAATAG...........................................................................................................CC--------AC----   4826
CPZ.CD.CPZANT       GGAAATC...........................................................................................................C-----G--------C   4563

H2A.GW.ALI Pol      __E__V__A__$                                                                                                                          p31
H2A.GW.ALI Vif      E__K__W__H__S__L__V__K__Y__L__K__Y__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.    Vif
H2A.GW.ALI          AGAAGTGGCATAGCCTTGTCAAGTACCTAAAGTACAG.............................................................................................   5510

H2A.SN.HIV2ST       AGAGGTGGCATAGCCTTATCAAGTATCTAAAATACAG.............................................................................................   4962
H2A.DE.BEN          AGAAATGGCATGCCCTTGTCAAGTACCTGAAATACAG.............................................................................................   5517
H2B.GH.D205         AGAGATGGCACAGTCTGATTAAGTATCTTAAGTATAG.............................................................................................   5490
H2B.CI.EHO          AGAGGTGGCACAGTCTAATTAAATACCTGAAATATAG.............................................................................................   5488
H2G.CI.ABT96        AGAAGTGGCATAGTCTGATAAAATTCTTAAAGTTTAA.............................................................................................   4844
MAC.US.MM251        AGAGGTGGCATAGCCTCATAAAATATCTGAAATATAA.............................................................................................   5410
SMM.US.SIVSMMH9     AGAAGTGGCATAGCCTCATCAAACACCTGAAATATAA.............................................................................................   4907
STM.US.STM          AGAGGTGGCATAGCCTCATAAAACATTTAAAGTATAA.............................................................................................   5081
TANTALUS.UG.TAN1    ................................................................................................TGGAGAGCATATAG-C-AA-GGAAGGGCATAGTG   5320
GRIVET.ET.GRI-677   GCAGATGGAGGGGGATAGTGACTTACAAGATCAGGAATAA..........................................................................................   5144
VERVET.KE.VER-9063  CCAAATGGCAAGGAATAGTCAAATATTGGATGACTAA.............................................................................................   5365
VERVET.KE.AGMTYO    CTAAATGGCAGGGAATAGTCAGATATTGGATGAGACA.............................................................................................   4852
VERVET.KE.AGM155    CCAAATGGCAGGGGATAGTTAGATATTGGATGAATAA.............................................................................................   5357
VERVET.DE.AGM3      CTAAATGGCAAGGGATAGTGAGGTACTGGATGAGGAC.............................................................................................   4860
SABAEUS.SN.SAB1C    AGAGGTGGACTAG........................................................................................................CC--GTGA-AT-C   5519
SUN.GA.SIVSUN       ..........................................AAGAGAATGGGAAGATAGAAGACAGTATAAGATAGTTAGAATAGTATGGCTTATAGATAG--TTGCTGT-GAGAA-T-TTT-G-TTTA   5526
MANDRIL.GA.MNDGB1   ....................................................................................TCAGAGAATAGAAAAGTGGCAC---TT-GTAAG-AG-C-GATGGCA   4820
LHOEST.KE.hoest     ......................................................................................................-T-C-TA-T-GAAAAGT-TTT-G--ATA   5476
LHOEST.CD.hoest447  ......................................................................................................G-GCCTAGT-GAGAAGT-CCT-G-TATC   4392
LHOEST.CD.hoest485  ......................................................................................................G-ACCT--T-GAGAA-T-CCT-G-TAT-   4392
LHOEST.CD.hoest524  ......................................................................................................G-GC-TAGT-GAGAA-T-TCT-G-TATC   4389
RCM.NG.SIV-NGM      ......................................................................................................GCAG---C-TTCC--------GT-T---   5476
COL.__.SIV-CGU1     CGTTAAGGGAATGGATTAA.................................................................GAAGCAGTTTCCAGAATGCCA--ATT-TTA-----G----ACAG-G   4932

SYKES.KE.SYK173     ATAGA........................................................................................................TTGCAGCATA---T-A-AACA   5318
SYKES.KE.SYK173 Vif D__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._L__Q__H__I__I__K__T_    Vif
```

```
B.FR.HXB2 Vif       _M__Y__V__S__G__K__.__A__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__   Vif
B.FR.HXB2           ATGTATGTTTCAGGGAAA...GCTAGGGGATGG...TTTTATAGACATCACTATGAAAGCCCTCATCCAAGAATAAGTTCAGAAGTACACATCCCACTAGGG......GATGCTAGA...TTGGTAATAA  5239

A.UG.U455           --------C--CAA----...---CAA--T---...-----------------------TAGA---T-----G-----------------------------......--A------...--A---G---  4685
C.ET.ETH2220        ---C--A-----A---G-...----AT------...G-------------T-----C---AGA-------AGG-----------------------T-----......--G------...--AA------  4631
D.CD.84ZR085        ---C--A-----AA----...----AA------...-----------C--------C--------C----A-------------------------T----A......---------...C-----G---  4761
F1.BE.VI850         ------------AA----...--C-AA------...-C--------------T-C-----AGG---------G---------------------------A-......--A-T--A-...--A-------  4579
G.SE.SE6165         --------C---AA----...-----A--C---...----------C--------C----AGG---------G----------------------------A......-------C-...C-A---G---  4636
H.CF.90CF056        ---C--A-----A-----...-----A------...--------------T-T-------A----------G------------------------T----A......--A-----G...--A--C----  4586
J.SE.SE7887         ---A--------AA----...-----AAA----...C-G-----G-----------C---AA--------A------------------------------A......--G----T-...--A-------  4553
K.CM.MP535C         ---C--A-----AA----...----ACC-----...-A----------------------AGG-------A-------------------------------......-----AGAG...--A---G---  4435
01_AE.TH.CM240      ------A-C---AA----...----AAAAG---...--------------T----------AG-------AGG----C-----------T-----------A......--G------...--A-------  4813
02_AG.NG.IBNG       --------C---AA----...----A---T---...----------------T------TAGG-------A-G--T--------------------------......--------G...--A---G---  4764
03_AB.RU.KAL153-2   --A---A-----AA----...------------...G------A----------------AGGA--------------------------------T-----......-------A-...----------  5284
04_cpx.CY.94CY032-3 ------------AA----...----AA-----A...--C--------------------TAGG--C----A-G----------------T------------......--G------...--A---G---  4605
N.CM.YBF30          --------G---AAA--G...--A-AA------...-A------------T-------CA-A---C----A------------------T------G----T......C-G--A---...--A--G-C-G  4833
O.CM.MVP5180        ..A--CA-G--TAA---G...--CGC-AAC---...CG------G-----T------TC-AGGA------A-G-C-----G-CG--GT-T--T---G---CA......--A---GAT...A-A--GG-C-  5269
O.BE.ANT70          ..---CAGG--TA-A--G...A-CGA-AAC---...-GG--C--------T------TC-AGAA--------G-T------AGT---T-T--T---G-----......-TG--CCAT...G-A---G---  5293
CPZ.US.CPZUS        ------AGA---AA----...----AA------...--C--C--G-----T-----G---A--A----C---G---A------------T-----TA---AT......A-A--A-A-...A-A--G--T-  5299
CPZ.GA.CPZGAB       --T---AGG---AAA--G...--A--A--T---...-----C--------T-----TCA----A------A-G-TGC---T---A-------A--TT-CA-AGAT...T--T---A-...--AA--G-G-  5303
CPZ.CM.CAM3         ------AAAA--AA----...--A-A--AC---...--------G--C--T--------TT-------T-A-G-----------A----------TA---AA......A-G--A-A-...--A--GG---  4653
CPZ.CM.CAM5         ------AGAA--AA---G...--A-AAAAG---...--------G-----T-----G--TGA-A----T-A-G-G--------------T-----TT---AT......ACA--A-A-...--A--GG-C-  4941
CPZ.CD.CPZANT       --A-GG-AGA-TAAAGT-...CT--AACC----...AAG-----------T-------ATGA---C----AG-A-G-AGA---------T--A------CCAACATTG---AAA-A-...------G-G-  4684

H2A.GW.ALI Vif      _.__.__.__.__T__K__D__L__E__K__V__C__Y__V__P__H__H__K__V__G__W__A__W__W__T__C__S__R__V__I__F__P__L__Q__.__.__G__R__S__H__L__E__I__   Vif
H2A.GW.ALI          ...........-ACA---GATCTAGAAAAGGT-TGC-A-GT-CCC--C--TA-G-TGG-ATGGGCATGGT-G-CTT-CAGCAGG---ATAT-----T--CAA......-GAAGA--TCATC-A-AG---C  5623
H2A.SN.HIV2ST       ...........-ACAGG-GATCTAGA-AAGGT-TGC-A-GT-CCC--C--TA-G-TGG-ATGGGCGTGGT-G-CTT-CAGCAGG---ATAT-----T--AAA......-GA-AA--TCATC---AG---C  5075

H2A.DE.BEN          ...........-ACA---GACCTAGA--AGGT-CGC-A-GT-CCC--C---A-G-TGG-ATGGGCATGGT-G-CTT-CAGCAGG---ATAT--------CAA......-GAAAA--TCATC-A-A----C  5630
H2B.GH.D205         ...........-ACAGG-GAGTTGCAACAGGTCTCT-A-GTCCCT--C---A-G-T-G-ATGGGC-TGGT-G-CTT-CAGTAG-A--ATAT-T--C---AAC......A-A-GAGC-TGGC-A-A-G-CC  5603
H2B.CI.EHO          ...........GACA---GACTTGCAACAGGTCTCT-A-GT-CCT--C--TA-G-T-G-ATGGGC-TGGT-G-CTT-CAGTAG----ATAT-T--C--GAAA......--A-GAGC-CATC-A-A-G-CC  5601
H2G.CI.ABT96        ...........-ACA---GACCTACA-AAGGC-GTT-A-GT-CCT------A-A-TTG-ATGGGC-TGGT-G-CTT-CAGCAGG---ATAT-----T--ACA......A-A-AAGC-CAT----A----C  4957
MAC.US.MM251        ...........-ACT---GATCTACAAAAGGTTTGC-A-GTGCCC---TTTA-G-TCG-ATGGGCATGGT-G-CCT-CAGCAG----AT-T----C---CA-......--A-GA--CCAT--A-A-G--C  5523
SMM.US.SIVSMMH9     ...........CACT---GACCTACA-AAGGCTTGT-A-GTACCC-----TA-A-TTG-ATGGGCATGGT-G-CTT-CAGCAG----AT-T----GT--A--......----AGGCTCAT----A-G--C  5020
STM.US.STM          ...........CACT---GAGCTA--CAAGGCATGC-A-GTGCCC--C--TA-G-TTG-GTGGGCATGGT-G-CTT-CAGTAG---GATTT-T--CT-GCAA......-GA-AAGC-CACC-A-A-G-TC  5194
TANTALUS.UG.TAN1    -AA--CCACATGA----CCGGTTACA--AT---ACA-A-CT--TG--CT--C-ATGTG-ATGGGC-TGGTAC-C-T--AGTAG-T-C-T------CT----A......-GA-AAG--AAAA-A---G--G  5444
GRIVET.ET.GRI-677   ....................ACAATT-CCT---GAA-ACAGACAT---TGGC-G-TGCAATGG--GTTTT-G-CCTACAGCC-GT-CATT-----CT--TCA......A-A-A-GATTACA-A-A-G-G-  5248
VERVET.KE.VER-9063  .................-AGGAACCTTAA----CAA-ACATG-TG--CT--C-AATT-ATTGGGC-TGGTAT-CT-TG-GCAG-TACAT-------A----A......TTA-AAG--GAAA-ACAG---G  5472
VERVET.KE.AGMTYO    .................-AGAAA-TT-AA----AAC-A-----TG---T--C-AATT-CATGGGC-TGGTAC-CC-TGAGTAG-TATGTA--A---A----A......A-ACA-G-GGAAA-ATGTG--G  4959
VERVET.KE.AGM155    .................-AGGAA-CT-AA----GAA-ACA-A-TG---T-TC-AATC-CTTGGGCATGGTAC-CT-TGAGCAG-TATGTA--A--C--CCCA......-GAAG-G--GAAA-CCATG-GG  5464
VERVET.DE.AGM3      .................TAGAAAATTA-AC---AAA-A-CGA-TG--CT--C-AATT-CATGGGCATGGTAC-C--TGAGTAG-TATG-G--A--C------......C-ACA-G--AGTA-ACATG--G  4967
SABAEUS.SN.SAB1C    CACATGCA-GTGTCC---CAGTG-GTTCAT---CGG-A-ACCCCT---ACTA-AAT---ATGGA--TGGTACTCCTA-CA----TGGGT---A--C--GAA-......----GAGCTCTAA-AAA-G-T-  5643
SUN.GA.SIVSUN       -GAAGAA-GCATA--G--ACAAAAGAT-AT---GTA-C-ATGTATGGAACAGG-ACTG-ATGGG-GTGGTAC-C-TACAATA--A-TATT--T---G--ACT......T---GG-C-GTAG-A----GG.  5649
MANDRIL.GA.MNDGB1   TG-GCCAC-G--AAT--TGAG-AAG-AT-T---TGGC-G---CCT---TTTATG-CTTATAA-G-ATGGTAT-CTT-CAGTA-----GTG--TAT-A--AAT......AGG-AC-T-AGA--AA--G-T-  4944
LHOEST.KE.hoest     G-CAGGT-AAACACAG-GGCAAAAGAAAAT---GTAGGC-TACATGGAACAGG--T-G-TTGGG-ATTCTAT-CCTA-CATA--T--GTG--A--T----A-......TCA-GA-CTTTAG-A--C-G-.  5599
LHOEST.CD.hoest447  -G-AGA-CACATAAAG-GACTCAA-AT-AT---GTAAGC-TACATGGAACAGGG-T-G--TGGG-ATTTTAT-CTTA-AATA-TA--GTG--A--TT---A-......ACA-GA-CCTTAG-A--G-G-.  4515
LHOEST.CD.hoest485  -G-AGA-CACATAAAG--ACTCAGGTT-AT---GTAAGC-TACATGGAACAGGG-T-G--TGGG-ATTTTAT-CTTA-AATA--A--GTA--A--TT---A-......ATA-GA-CCTTAG-A----G-.  4515
LHOEST.CD.hoest524  -AAAGGAA-CATTCAG--AGTAAA-AT-AT---GTTAG--TACATGG-ACAGGA-T-G-ATGGG-GTTTTAT-CTTACAATA--T---TA--T---T---AT......ACA-GA-CTTTAG-A----G-.  4512
RCM.NG.SIV-NGM      CA----AAGGG-AA---GGAA----AA-AT---GAG-A-GTACCC---TTTA-A-T-CCTTGGGGGTGGT-GTCTC----------C--------CT-G--A......A--AAC-CTAAGA-AAAGG-T-  5600
COL.__.SIV-CGU1     TCACCC-AA-GGT---C-GAA-G--AAAT-GA-TTT-AC-TAG-GTCAGCTAT-A-GTTAAGCATAAT-TACT-GG-CA-T......TTA-C----GATA-A......T--CA----GGACCACC-G-CT  5050

SYKES.KE.SYK173     CATA-GTA-AA-A-C---GAATTGGA-AA-GCCACA-A-A-ACAT--CT--C-AAT-GAATGG--ATGGTAC-CTTA--GCC--TGGACA--A---G----A......----GA-CTATA-G-A-C-CG.  5441
SYKES.KE.SYK173 Vif _H__K__Y__K__S__K__E__L__E__K__A__T__Y__K__H__H__Y__Q__I__E__W__Q__W__Y__T__Y__C__Q__W__T__I__P__V__G__.__.__D__G__T__I__W__I__T__   Vif
```

```
B.FR.HXB2 Vif       T__T__Y__W__G__L__H__T__G__E__R__D__W__H__.__.__.__.__.__.__L__G__Q__G__V__S__I__E__W__R__K__K__R__Y__S__T__Q__V__D__P__E__L__A__D   Vif
B.FR.HXB2           CAACATATTGGGGTCTGCATACAGGAGAAAGAGACTGGCAT..................TTGGGTCAGGGAGTCTCCATAGAATGGAGGAAAAAGAGATATAGCACACAAGTAGACCCTGAACTAGCAGA  5351

A.UG.U455           G-----------------------------A---------C..................--------T--G------------------CTG--A--------------------T-----C--------  4797
C.ET.ETH2220        A-------------T----A--------------T------..................--------T--------------------ATTG-GA--C----A-----------------GC--G-----  4743
D.CD.84ZR085        ----------------------------------A------..................C---------------------------------G--------------------------GC--G-----  4873
F1.BE.VI850         -------------G-----CC-------------A------..................C-----------------------------C-GGG--AG----G------A---------GC--G--G--  4691
G.SE.SE6165         -------------G----------------A---------A..................--------T--G-----------G------C-G-GA--------A-----------G-----C--------  4748
H.CF.90CF056        -C-----C---------A----------------A------..................--A--C-----------------------ACTG--A--G-----------------G----GC--G-----  4698
J.SE.SE7887         -------------------G--------------T-----C..................------------------------------C-G-GA--G--C--A----------------GC--G-----  4665
K.CM.MP535C         ------------------T------------G---------..................------------------------------CTG--A--G--C--A-----------A-----C--------  4547
01_AE.TH.CM240      G------------------A----------AG--------C..................--------T---------------------C-G-GA-A-------------A----T-----C--------  4925
02_AG.NG.IBNG       G----------------------------C----------C..................--------T--G----------------A-C-G--A---------------A----------CT-------  4876
03_AB.RU.KAL153-2   A----------------------------------------..................-----C-------C---------------A---G--------------------------A-C--------  5396
04_cpx.CY.94CY032-3 G------------------GC----G---CA---------C..................--------T---------------------CTC-GA-----------------G--T-----C--G-----  4717
N.CM.YBF30          TC--T--------G--AACA---------CAGTCT------..................C-A--A--T-----A--------------ACT--GA-A---C-AG--------T--T------A-G-----  4945
O.CM.MVP5180        -C-----------AT-AATGC----G--------GGAA--C..................-----A--T--G--TAGT---------CAAT-C---GAG----AA-----GA-T--T------AC------  5381
O.BE.ANT70          -C-----------AT--ATGC----G--------TGAA---..................-----A--T--G--TAGT---------C-TT-T----AG----AA-----GA-T---------AC------  5405
CPZ.US.CPZUS        T---G--------GT--A-A---------CA--C------C..................--A--G--T-----A--A------------AC--GGA-CT--C-AG------A----T-----GA-G-----  5411
CPZ.GA.CPZGAB       ----C-------CCT-AAGCC----------G-C-------..................--A--C--T-----------TC--------CT-GGATCT---GT------G--G--T--CTTCAC------  5415
CPZ.CM.CAM3         ----C--------AT-AATGC-----C-GCAT-C-------..................C----G--T-----G--------------ATGTGG-TC------T-----------------GA-G-----  4765
CPZ.CM.CAM5         -------------A---ACA-----GC--CAT-CT-----C..................C----A--T-----G--------------ATGTGG-TC---C-AA-----------T------ACG-----  5053
CPZ.CD.CPZANT       --GTG-TC-----AT-A--ATGT------C-CCCA-----C..................--A--A--T--T--GAG------------ATGTGGA-AG----T---------------A---AC---T--  4796

H2A.GW.ALI Vif      Q__A__Y__W__N__L__.__T__P__E__K__G__W__L__S__S__Y__A__V__.__.__.__.__.__.__R__I__T__W__Y__T__E__K__F__W__T__D__V__T__P__D__C__A__D   Vif
H2A.GW.ALI          AGG--------AAC--A...---CC-----A--GA----TCTCCTCTTATGCAGTA..................AGA---ACT---TAT-C-G-A-AG-TCT-G---G-T--TAC---A--CTGT-----  5732

H2A.SN.HIV2ST       AGG----C---AAC--A...---CC-----A--GA----TCTCCTCCTATTCAGTA..................AGAC--ACT---TAT-C-G-A-A--TCT-G---G-T--TAC---A--CTGT--G--  5184
H2A.DE.BEN          AGG--------AAC--A...---CC-----A--GA----TCTCCTCTCATGCAGTA..................AGAT--ACC---TAT-C-G-A-AG-TCT-G---G-T--TAC---A--CTGT-----  5739
H2B.GH.D205         A-GG-------AAC--A...--CCC------G-GA-TCTTGAGCTCCTATGCTGTA..................AGAC--AC----TATG-G-G--AC-T-TAT---G-T---ACA-----TG-G-----  5712
H2B.CI.EHO          A-GG---C---AAC---...--CCC------G-GA-TCTTGAGTTCCTATGCTGTA..................AGAC--AC----TATG-G-G---C-T-TAT---G-T---ACT-----TG-------  5710
H2G.CI.ABT96        A-GGG------AACT--...--CCCT----AG-GA---TTAAGCCAGTATGCAGTA..................AGAT--ACC---TAC-C--GA-A--T-TAT---G-T--GACA--A--GAC------  5066
MAC.US.MM251        A-GGG------CA-T--...---CC--------GG----TCAGTACTTATGCAGTG..................AGG---ACC---TACTC--G--AC-T-T-G---G-T---ACA--A--CTAT-----  5632
SMM.US.SIVSMMH9     A-GG-------AA-T--...---CC---GRA--GG----TCAGTACTTATGCAGTG..................AGA---ACC---TACTCC-GA-AT-TCT-G---G-T-C-ACA--A--TTAT-----  5129
STM.US.STM          A-GGG------AA-T-A...---CC-----A--GA----TCAGTGAATATGCTGTG..................AGG---ACT---TAT-CC-GA-AT-TCT-GT--G-T---ACA--A--CTGT-----  5303
TANTALUS.UG.TAN1    ATTGT--C---CAC--A...--GCC----CAG-GA----TCTCTACCTATGCAGTAGCAA-A.........AGT-TTGAGA-T---CAA--C...-CG--C-AA---G-----ACT-----TG-G--G--  5559
GRIVET.ET.GRI-677   AT-TT---CACAAC--C...--CCC--------GA----TCTCAAGTCATGGAGTAGGG--A.........TC--ATTACC-TCAA-A-GG-......----AG---G-------T--A-G-AC------  5360
VERVET.KE.VER-9063  ACTGC------CA-T-A...---CC--------G-----TCTCTACTTACGCCATAGGAA-A.........CAA-A-T-G-G-GCAGTAGG-...-AT-TC--A---G--T-------A-C-AC------  5587
VERVET.KE.AGMTYO    ACCT-------CA-T-A...---CC---GCA--GA----TATCCACATATGCAGTAGGTA-A.........CAA-ATG--AGCAATTTAG--TCT-A------A---G--T---------CTAC------  5077
VERVET.KE.AGM155    AT-TC------CA-T-A...G-TCC-A--CA--GA----TCTCAACTTATGCAGTAGGAA-A.........CAA-ATG-TAGCCTAGTA--TG-T-A------A---G--T----T--CA-TAC------  5582
VERVET.DE.AGM3      ATCT-------CA----...---CC-----AG-GA----TATCAACATATGCTGAGGGGA-A.........CAG-ATC--AGCAAT---G-TCCTT-G-----G---G--T-G--T----C-AC------  5085
SABAEUS.SN.SAB1C    -T-AT------CA-T-A...--CCC---G-A--GA---TTAGAAACCTATGCAACA..................GGA----G--AC--T---GG-GA--GGTTT---G-GC-T--T---TGGAC---G--  5752
SUN.GA.SIVSUN       ..-TC---G-ACAC--C...---CC--C--A--GG-T-AT-AACCAGTGGGGATGT..................AGT--G-------TTT-T--C--T---CAA---G--A----T---CTTG-G-----  5756
MANDRIL.GA.MNDGB1   G--GC------CA-T--...CA-AT---GGT--GA--CTTAAGTACTTATGCAGTAAGCA-A-AAGCA-T---TAGACCGCC-CCCTTTG-G--AGAG-GGT-T---G-GA--ACT--A--GG-------  5071
LHOEST.KE.hoest     ..-T----GCACA-T--...G---C--G---G-GA---ATCTCACAATGGGCAGTT..................------------GTCT-T-GAGAC---CAA---G--A-------CAT---------  5706
LHOEST.CD.hoest447  ..TGC---GCACACT--...G---C--G-----GA---ATCTCTCAGTGGGCAGTA..................--A----------TTT-TGG--AC---CAA---G--A----T--CATT------A-  4622
LHOEST.CD.hoest485  ..TGC---GCACACT--...G---C--G-----GA---ATCTCTCAGTGGGCAGTA..................--A----------TTT-TGG--AC---CAA---G--A----T--CATT--------  4622
LHOEST.CD.hoest524  ..-TT---GCTCA----...G---C--GG----GA---AT-TCACAGTGGGCAGTA..................--A----------TCT-TGGA-AC---CAG---G----G--T--CATT-----T--  4619
RCM.NG.SIV-NGM      -----------A----C...---AT-----AG-GA---TTAGGAACCTATGGAGTAGGAA-A......-C-TA-ATAGATA-CA-AT-TG-TCCCT-T--CTT----G-TA-T-----AAC-G-------  5721
COL.__.SIV-CGU1     -TTATG-A-TACAGA-AGG--AGT-G----C---T-TAAGCTGGCAACAAGCAGTAGGG-AC....................................................................  5112

SYKES.KE.SYK173     ..TTC---CACAA-T--...G-CCC--------GA---TTACATATGCAGGGAATT..................AGG--TC---ACCAATGG--CCAG-GG-A---TG-CT--ACA--A-C-G-------  5548
SYKES.KE.SYK173 Vif .__F__Y__H__N__L__.__A__P__E__R__G__W__L__H__M__Q__G__I__.__.__.__.__.__.__R__I__Q__Y__Q__W__N__Q__W__N__T__D__L__T__P__A__V__A__D   Vif
```

```
B.FR.HXB2 Vif       __Q__L__I__H__L__Y__Y__F__D__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_   Vif
B.FR.HXB2           CCAACTAATTCATCTGTATTACTTTGACTGTTTTTCAGACTCTGCTATAAGAAAGGCCTTATTAGGACACATAGTTAGCCCTAGGTGTGAATATCAAGCAGGACATAAC..................AAG   5463

A.UG.U455           ---T--------C---C----T-----------------A-----C-----G-GA---A----------A--------A-----------------------------..................---   4909
C.ET.ETH2220        ---C---------A--C----T-----T------G----A-----C-----G--A---A--------T-T-G-----------------C-----------------T..................---   4855
D.CD.84ZR085        -------------A-------T-----T------G-------------C-----A---A----------T--------T-----------------------------..................---   4985
F1.BE.VI850         ------G--C---A-A-----T-----T-----------A-----C-----G--A---A----------T-G-A-------------A-C-----------------T..................---   4803
G.SE.SE6165         ---T--G---------C----T-----T-----------------C-----G--A---A----------A-------------------------------------T..................C--   4860
H.CF.90CF056        -------------A--C----T-----T-----------A-----C-----G--A---A---------GTG-------A---------A-C----C-----------A..................C--   4810
J.SE.SE7887         -------------A---G---------T-----------------C-----G------A----------A------------------C--C----------------..................---   4777
K.CM.MP535C         -------------A-A-----T-----T-----------A-----AG---G------A----------T-G-----T-------------G-----------------..................---   4659
01_AE.TH.CM240      -A----G--------AC-A--T----G------------------C-----G--A---A----------AG-------A-G-------------C-T-----------..................---   5037
02_AG.NG.IBNG       ------G-----C---------T---A-T----------------C--------A---A--------G-A-------A-----------------------------T..................---   4988
03_AB.RU.KAL153-2   ----------------------T-----T----------A-------------T---A----------T-G-----------T----------G-------------..................---   5508
04_cpx.CY.94CY032-3 -------------A--C-----------T----------A-----C-----G--A---A----------T-G-----T-----------------------------T..................---   4829
N.CM.YBF30          -A-G-----A-----TC----T-----T------A---C------C-----GC-A--GG-C-----GAGACC---ATTA-----------------C------G--C--A..................C--   5057
O.CM.MVP5180        -AGGA-G--A------C----T--CACA------A----A--A--A--C--G------A-TC----G--G-G---GCTGA-C-A---------C-TG----------GT..................C--   5493
O.BE.ANT70          -AGGA-G--A------C----T---AC-------A---CA--A--AG-C--G--A---A-CC-------G-G---ACTGA-C-A---------C-CTA---------GT..................C--   5517
CPZ.US.CPZUS        -A-G-----------------T-----T------A---C------C--C-----A--TG-T-----GAGACC---A-TT--C-AA--------C-CTA---G--C--T..................---   5523
CPZ.GA.CPZGAB       TAG-T-G--C--CTCTC-A--T-----T--C---G---AA----C--C----GA--TA-------G--GC-T--AGCG---C-T--------CA---A------C-GG..................C--   5527
CPZ.CM.CAM3         TA----------C--TC-A--T-----T------G---CT-----C--C--G--A--TG-T-----GA-ACA---ATAT----A---------C-C------G---C-A..................C--   4877
CPZ.CM.CAM5         TAG---------C--TC----T-----T------G---CT-----C--C--G--A--TG-T-----GA-ACA---ATAT----A---------C-C-A----G--CC-A..................C--   5165
CPZ.CD.CPZANT       T---A-G--A--C-AA-----T--CCC-------------TCAA--AG-C---C-A--AA-CC-----G-A-G-A-ATTGA-ATAC--CC-C--CA-GAAG-------GT..................C--   4908

H2A.GW.ALI Vif      __S__L__I__H__G__T__Y__F__S__C__F__T__A__G__E__V__R__R__A__I__R__G__E__K__L__L__S__C__C__N__Y__P__Q__A__H__K__S__.__.__.__.__.__Q_   Vif
H2A.GW.ALI          -TCC-----A---GGCAC---T--CTCT--C---A-G-CAGG--AAG------GA---A-CAG---GG-A-AGT-ATTGT-CT-C--CA-T----CCCAG-CC-----GTCA...............C--   5847

H2A.SN.HIV2ST       -TCC-----A---AGCAC---T--CTCT--C---A-G-CAGGC-AAG------GA---A-CAG---GG-A-AGC-ATTAT-CT-C--CA-C--C-CCCA--CC-----GTAC...............C--   5299
H2A.DE.BEN          -ATC-----A---AGCAC---T--CTCT--C---A-G-CAGG--AAG------GA---A-CAG---GG-A-AGT-ATTGT-CT-C--CA-C----CCCA--CT-----AGCA...............C--   5854
H2B.GH.D205         ---G---C-G---GG--C---T--CTCT--C-------C-AA--AAG---G-GA---A-CAGG---G-A-AGA-ATTGT-CTAC--CA-C----C-T---CT--CG-AGGG...............C--   5827
H2B.CI.EHO          --G-T--C-G---GG--C---T--CTCT-CG---A---CTAA--AAG---G-GA---A-CAGG---G-A-AGA-ATTGT-CCAC--CA-C--C-C-T---CT----CAGGC...............C--   5825
H2G.CI.ABT96        T---T--T-G---GGA-C---T-----T--C---A---CTGG--AAG------GA---A-CAGG---G-GCAGA-ATTGT-CT-C--CA-C----C-A---CT-----AAGG...............C--   5181
MAC.US.MM251        -ATTT--C-G---AGCAC---T--CCCT--C---A---CGGGA-AAG-G----G----A-CAGG---G-ACA-C-GCTGT--T-C--CA-G-TC-CGAG--CT----GGTAC...............C--   5747
SMM.US.SIVSMMH9     -ACTT--C-G---GGCAC---T--CCCT--C--------GGGA-AAG--C---G----A-CAGG---G-A-A-T-GCTGT--T-C--CA-G-TC-CGAA--CT-----GAAT...............C--   5244
STM.US.STM          T--G---C-G---GGCAC---T--CCCT--C---A---CGGGA-AAG-G----GA---A-CAGG---G-A-AGT-ACTGT--T-C--CAGG-TCACGAA--CT-----GAAC...............C--   5418
TANTALUS.UG.TAN1    ---TA-G------TGCC-C-----CCCT--C---A----TAGA--C--CCA-C-----A-CAG---GG-A-GCT--TTGTGG...--CAC---CA-G-A------CGTAGCAGAAAACCATTGGGGTC--   5686
GRIVET.ET.GRI-677   -AG-A-G--A--C--A-----T---A--------A----TAGA--C--CCA-C----TA-CAG---GG-G-AGTA--CGTGG...--CAC--TCA-G-A---------A...............GGTC--   5472
VERVET.KE.VER-9063  TGG-A-G--A--C--AC-C--T---TCA------A----AAGG--C--TCA---A--TA--AGG---G-G-G-T-CGT-TTC...---C---TC-CC-AG--------AACCACA.........GGA---   5705
VERVET.KE.AGMTYO    TAGTA----A---GGTC-C--T---A-T------AA---AAGA--C--CCA-C-A--TC-GAGG--C----G-T--GT-TTC...---C-G-T--C--A---G-----AAGCACA.........GGAC--   5195
VERVET.KE.AGM155    -TCCA-G--A---TGTC----T---AC-------A----TAGA--C--CCA-C----AC--AGG---A---GGT-C-T-TTC...---C---T--C--G---------ACTAACA.........GGTC--   5700
VERVET.DE.AGM3      TAGC--G--A---ACCC----T---ACT------A----AAGG--C--C--G--A---C----G-----G-GGT-C-C-TTC...---C-G-TC-CC-AG-----C--GAAAACA.........GGAC--   5203
SABAEUS.SN.SAB1C    ---TA-T------TG--CC--T--CCC-------A----TAGA---G-CCA-C-A---A-CAGG---G-A-AGTA-CTTTGG...---A--C-C--G-TG------C-GCCAACA.........GGGC-A   5870
SUN.GA.SIVSUN       T--GA-G--C--CT-TA-G--T-----T----GGA-CTCTAGGTG---C----GA--TA-G-----GG-A-A-A-ACTA-A-GA---CAG-A------T--CT--C--AGGC...............CT-   5871
MANDRIL.GA.MNDGB1   T--T-----A---T-AC---TT-A------C--CATG---AG---AG-T-TG--A---A-CAGG---G-AGA---GTTAAAAGTT---AG--T--C---T--C-----AGCA...........CAAGGT   5189
LHOEST.KE.hoest     ----A-G--C--CTATA-A--T---A----C--CAT-C-AAGA-AC--T----G---AC-TC-G---G-A-G-A-A-CTATGT-T--CC-G...-CTTT---------AGGT...............TGT   5818
LHOEST.CD.hoest447  T---A-G--A--CTGTA-A--T---A----C-----TACAAGG-AC--T--G-----T----GG---G-A-AGA-A-TGG-ATTC----CTATG-CCA-T--------AGGT...............TGT   4737
LHOEST.CD.hoest485  T---A-G--A--CTGTA-A--T---A----C-----TACAAGG-A---T--G-----T----GG---G-A-AGA-A-TGG-A-TC--C-CTATG-CCA-T--------AGGT...............TGT   4737
LHOEST.CD.hoest524  T---A-GC-----G--A-G--T----------GGGTGC-TAGG-AC---GTT-G----A--AGG---G-A-AGA--CTG-AC--C---A--ATG-CTA-T--------AGGT...............TGT   4734
RCM.NG.SIV-NGM      TA--T--G-A---AATC-G--T---CCT------A----TAAA--A--T---C-A---A-C------G-GCA---GCTG-T-.....----GG-T-G-GAA--------GAGAT...............C--   5833
COL.__.SIV-CGU1     ....A-GCACATCA-AAGGGGGAAGTGTGTACGCAT---GGAA-AAGC------A--TA--AGG--CTTACCCTGG-AT---..-----T-TC--G-T------CCTA..................GGA   5217

SYKES.KE.SYK173     --GC--T--A---AAT-T---T--CCCT--C---A---CAAGA--GG-C-ACC----AG-CAG---GG-AC--T-G-CAT-ACAC--CTGGACC-C-.........----CTGAC...............C--   5657
SYKES.KE.SYK173 Vif __R__L__I__H__N__F__Y__F__P__C__F__T__A__R__A__V__N__Q__A__V__R__G__E__L__L__T__S__H__C__W__T__P__.__.__H__T__D__.__.__.__.__.__Q_   Vif
```

```
                                                                                                                                                                   |<- start Vpr
B.FR.HXB2 Vpr                                                                                                                                                      M__E__Q__A__.__    Vpr
B.FR.HXB2 Vif      _V__G__S__L__Q__Y__L__A__L__A__A__L__I__.__.__.__.__T__P__K__K__I__.__K__P__P__L__P__S__V__T__K__L__T__E__D__R__W__N__K__P__._    Vif
B.FR.HXB2          GTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATA............ACACCAAAAAAGATA...AAGCCACCTTTGCCTAGTGTTACGAAACTGACAGAGGATAGATGGAACAAGCC....   5570

A.UG.U455          -----------------T-----T--GAA-------G--.............--C----C--G-GC-...---------------------A----T-A-----------------------....   5016
C.ET.ETH2220       -------------------------GA-------G---.............-AG---------GC-...---------C-----------GT---T-AGT--------A-------------....   4962
D.CD.84ZR085       --------CT----G--T--A------A------------............G------------G-..---------------C-------A---G--A-----A-----------------....   5092
F1.BE.VI850        --------C--------T---------A------------............G-T---G-G----C-..---------GC----------CCA------AGT------C--------------....   4910
G.SE.SE6165        -----------------T------T-GAA--T----G--.............---T------G--GT...-G------------C---------G--T-AG---A-----------------....   4967
H.CF.90CF056       ------A-----------------T--A--------G-G.............G---------------..---------------------GA--G--AGT---------------------....   4917
J.SE.SE7887        -----------------TC--------A------------.............-G--------G--G-..---------------------CA------AGT---------------------....   4884
K.CM.MP535C        --------CT-------T--------CA--------G--............G-T----G--G-CC-..---------G------------A------AGT---------------------....   4766
01_AE.TH.CM240     --------C--------T--------GAA--------C-.............----------G----..-G-------C----------A...G--A-A-----A-----------------....   5141
02_AG.NG.IBNG      --------G--------T--------GAAT--T---G--............G------C-----C-..---------------------G---GT-AG----A--C--------AG----....   5095
03_AB.RU.KAL153-2  --------------G----------------------G-.............---------------..------------------------------------------------------....   5615
04_cpx.CY.94CY032-3 --------CT--------C-----------------------............T-C----------C-..---------------------A------AGTG--------------------....   4936
N.CM.YBF30         -----CA-C--------TC-A------A----C-GGG-G............GG-G----G----G-..---------C--A--------G--T--G--A-----A-----------G---A....   5164
O.CM.MVP5180       -----GA-A-------T---A--CT-GAA----G--G-G.............-A-GT-----GA-AT..------T--CC-A--C-----CCA--G-T-A-----A-----------------....   5600
O.BE.ANT70         -----GA-A------CTGC-A--T---AG----G--G--.............-A-G---G--GC-GG..------T--CC-A--C-----CCA---T-A-----A----------------A....   5624
CPZ.US.CPZUS       -----C---T-G----T------TT--AAG--C---G--............GG--AG-GC--A-G-...-GA-----C--A--------G--T---T-------A---------------TG....   5630
CPZ.GA.CPZGAB      --------CT----G-T------C---AA---T------............T--GAG-G--G-CAT...-GA--------A--------CG-C---T-A---------------------A....   5634
CPZ.CM.CAM3        -----G--AT-------T--A--C---AG---C-GGG--.............-G-GT-GG-----A-..--A-----C-----------C--T--GT-A-----A-------------G--A....   4984
CPZ.CM.CAM5        -K---G--A--------T--A--C---AG---T-GGG-G............GG-GT-C------A-...-CA-----C-----C---A--GSC--GT-AG----A---------------ATA....   5272
CPZ.CD.CPZANT      -----GA-A--G--G--T--A--TT-TTGTAAGA-CC--GAATTCAGA...GG-TACCCC---GGTCCA-GAAG--AG--C------C-GT-T-TCT-A-----A-----------------....   5027

                                                                                                                                                                   |<- start Vpx
H2A.GW.ALI Vpx                                                                                                                                                     M__A__N__.__._     Vpx
H2A.GW.ALI Vif     _V__P__S__L__Q__F__L__A__L__V__V__V__.__.__.__.__.__._#_N__K_.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.#_W__Q__.__.__._     Vif
H2A.GW.ALI         ---CCG--A--C----TTC----CT---TG-T-G-G..................-A-C---..............................................----C-A.........   5897

H2A.SN.HIV2ST      ---CCG--A--C--G-TTC----CT---TG-T-G-G..................-A-C---..............................................----C-..........   5348
H2A.DE.BEN         ---CC---A--T------C-A--C----T--T-G--..................-A-C---..............................................---AC-..........   5903
H2B.GH.D205        ---CC-AGCT----G-TTC-A--C---AGG-TCG--CAGGAAGGAAAAA.....................................................-----T..........   5882
H2B.CI.EHO         ---CC-AG-T----G-TTC-A--C---AG--TTG--CA-GAAGGAAAAG.....................................................-----T..........   5880
H2G.CI.ABT96       --GCCTAGCT-G----TTC-A--CT--CA--T-G-GCA-...AAAGGAC.....................................................-----T..........   5233
MAC.US.MM251       ---CC-AGC-----G-----A------AA--T-G---GCG..............................................................---TC-..........   5793
SMM.US.SIVSMMH9    ---CC-AGC-----G--TC-A------A---T-G---GTC..............................................................---TCR..........   5290
STM.US.STM         ---CC-AGCT----G-----A--TT--AA--T-G-GGA-C..............................................................---TC-..........   5464
TANTALUS.UG.TAN1   ---C-G--C--G--G-TT--A--C--GA---T--AT-C-GACTTCCTAAGA-...................................................----C-G---GAAGA.   5753
GRIVET.ET.GRI-677  ---CA---A--G---CTT----------TT----AT-C-A..............................................................----C-TC--GAAGA.   5527
VERVET.KE.VER-9063 ---CCC--A--G--G-----A--C--TCTG--C.....................-ATC-G-.............................................----C-TC--GGAGA.   5760
VERVET.KE.AGMTYO   ---CC----T-G--G---C-A--T--GCTC---.....................-ATC---.............................................----CCTC--GGAGA.   5250
VERVET.KE.AGM155   ---CCC--CT-G-----T--A---T--CT---C.....................-ATC---.............................................----CCTC--GAAGA.   5755
VERVET.DE.AGM3     ---CCC---T-G--------A--T--CCTT---.....................-ACC---.............................................----CCTC--G-AGA.   5258
SABAEUS.SN.SAB1C   ---CC---AT----G--T--A-----TAGG-T--AT-C-A..............................................................----CCTC--GAGGGT   5926
SUN.GA.SIVSUN      --TCTC---T-G--G-TTC-CTGCT-GAG--TTC-CC.................................................................----CCAGC.......   5917
MANDRIL.GA.MNDGB1  --TCTC---T-G--G-TTC-CTGCT-GAG--TCA-CT.................................................................---.............   5229
LHOEST.KE.hoest    --TCTC---T-G--G-TTA-CTGCT-GAG-CA-C--CA-C..............................................................---TCCAGGCAGAGG.   5873
LHOEST.CD.hoest447 --TCTC--AT-G--G-T-C-CTGCT-GAG-CA-C--CA-C..............................................................---TCCAG-A-AAGG.   4792
LHOEST.CD.hoest485 --TCTC--AT-G--G-TTC-CTGCT-GAG-CA-C--CA-C..............................................................---TCCAG---AAGG.   4792
LHOEST.CD.hoest524 --TCTC--AT-G--G---C-CTGCT-GAG-CA-C--CA-C..............................................................---TCCAG-CAAAGA.   4789
RCM.NG.SIV-NGM     -----GAGC--------TC-----A--AA---T-GGGCTAGG............-AGC-GCTA-AG...G.....................................----C-G-G-G-AGA.   5903
COL.__.SIV-CGU1    --GTTGAAGT-G---G-AC-ATGT--TCTG--CC---G-GAGAAAGAGAGG--C-ACC-G---...........................................---AG-..........   5286

SYKES.KE.SYK173    ---CCC--A--T--G--T--A---T--CA--TG-ACC--AAAG...........................................................----CGG-G--TTTC.   5715
SYKES.KE.SYK173 Vif _V__P__S__L__Q__Y__L__A__L__Q__V__Y__L__K__D__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__#.__G__G__G__F__.    Vif
SYKES.KE.SYK173 Vpr                                                                                                                                                 _M__A__E__A__F__    Vpr
```

```
                                                                                                                             Vif end in HIV-1 <-|
B.FR.HXB2 Vif       _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__Q__K__T__K__G__H__R__G__S__H__T__M__N__G__H__$       Vif
B.FR.HXB2 Vpr       .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__P__E__D__Q__G__P__Q__R__E__P__H__N__E__W__T__L__E__L    Vpr
B.FR.HXB2           .............................................................................CCAGAAGACCAAGGGCCACAGAGGGAGCCACACAATGAATGGACACTAGAGCT   5623

A.UG.U455           .............................................................................-----------G-----------------GT--GC-----A-G--T-----T-   5069
C.ET.ETH2220        .............................................................................-----------G-----G--------A---T-----------G----------   5015
D.CD.84ZR085        .............................................................................--G-C------------G------------T--------------T-------   5145
F1.BE.VI850         .............................................................................----G------G------------------T-----------G-------AA-   4963
G.SE.SE6165         .............................................................................-----------G-----------A--A--CT-----------G--T----A--   5020
H.CF.90CF056        .............................................................................-----------G-----------------------------------------   4970
J.SE.SE7887         .............................................................................-----------C-----------A------T--C-------------------   4937
K.CM.MP535C         .............................................................................-----------G------------------A--------------T-----A-   4819
01_AE.TH.CM240      .............................................................................-------...-G-----------A--A--CT--------------T----A--   5191
02_AG.NG.IBNG       .............................................................................-----------G-----------------GTC------------------A--   5148
03_AB.RU.KAL153-2   .............................................................................-----G--------A---------------T-------G--------------   5668
04_cpx.CY.94CY032-3 .............................................................................-----------G-----G-----A--A---A-T--------------------   4989
N.CM.YBF30          .............................................................................--------TGC---------------A--CT-T---------G-------AT-   5217
O.CM.MVP5180        .............................................................................-TG---A-T--G--A---GCT---------TT---------------------   5653
O.BE.ANT70          .............................................................................--T--GA-T--G--A---GCT-AA-----CTT----------G----------   5677
CPZ.US.CPZUS        .............................................................................------------------GCC--A--A---T-------G---G--T----TG-   5683
CPZ.GA.CPZGAB       .............................................................................-----G--------T----CA--A--A--TT--C-G------G--T-----AC   5687
CPZ.CM.CAM3         .............................................................................--------T--G--T----CA--A------T-----------G--T----A--   5037
CPZ.CM.CAM5         .............................................................................-----------G-------CA--A--A---T-----------R-------AA-   5325
CPZ.CD.CPZANT       .............................................................................--GC-G--TG-G-----------A--A---ATG-----G---TTGT----AAC   5080

H2A.GW.ALI Vif      __.__.__.__.__.__.__.__.__.__.__.__.__.__.__#__P__Q__R__N__S__T__T__R__#__.__.__#__Q__W__R__R__D__Y__R__R__G__L__R__V__A__R__Q__D_    Vif
H2A.GW.ALI Vpx      .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__P__R__E__T__V__P__P__G__.__.__.__N__S__G__E__E__T__I__G__E__A__F__E__W__L__D__R__T    Vpx
H2A.GW.ALI          .............................................ACCCCAGAGAAACAGTACCACCAGGA.........A-CAGTGGC-AAGAGACT-TC-GAGAGGC-TTC--G---CT-GACAG-AC   5973

H2A.SN.HIV2ST       ............................................GGCCCCAGAGAGACAATACCACCAGGA.........A-CAGTGGC-AAGA-ACT-TC-GAGAGGC-TTC--G---CT-GACAG-AC   5425
H2A.DE.BEN          ............................................GACCCCAGAGAAAGGGTACCGCCAGGA.........A-CAGTGG--AAGAGACC-TT-GAGAGGC-TTC--G---CT-GAGAG-AC   5980
H2B.GH.D205         ...............................................CCCAGGGAGAGAGTGCCACCAGGA.........A-CAG-G-C-AAGA-ACAGTA-GAGA-GCATTC-CT---CT-GA-AGAAC   5956
H2B.CI.EHO          ...............................................CCCAGGGAGAGAGTACCACCAGGA.........A-CAG-GGC-AAGA-ACAGTA-GAGAGGCATTC------CT-GAGACAAC   5954
H2G.CI.ABT96        ...............................................CCAAGGGAGAGAGTMCCACCAGAA.........A-CAG-G-C-A-GAGACA-TC-GAGAGGCATTC------CT-GA-AGAAC   5307
MAC.US.MM251        ............................................GATCCCAGGGAGAGAATCCCACCTGGA.........A-CAGTGG--AAGAGACA-TA-GAGAGGC-TTC------CT-AACAGAAC   5870
SMM.US.SIVSMMH9     ............................................SRTCCCAGGGAGAGRATCCCACCTGGA.........A-CAGTGG--AAGA-ACRRTA-GCGAGGCYTTC--C---CT--ACAGAAC   5367
STM.US.STM          ............................................GATCCCAGAGAGAGAATACCGCCCGGA.........A-CAGTGG--AAGAGGCA-TA---GA-GCATTC--G---CT--ACAGAAC   5541
TANTALUS.UG.TAN1    ............................................GATTCCAGGGAAAGAAGACCCGGATGGTTAGAAATTTGG--T-TCA-C......--------GTGGG-C-----ATT-AG---CA-   5833
GRIVET.ET.GRI-677   ............................................GATCCAAGAGAACCTTTACCAGGATGG......-TG---AT-TGG-AT-T-G-C---------TGGG-C------CT--A---CA-   5607
VERVET.KE.VER-9063  ............................................GGTCCCAGAGAAAACAGGCCAGGAGAA......GT---GAT-TGG-AT-T-A-C--A--A---TGGG-C------CT-AGG--CA-   5840
VERVET.KE.AGMTYO    ............................................GATCCAAGAGAGGCAAGACCAGGAGAA......GT----ATTTGG-AT-T-AGC--------GTGGG-CA-----CT-AG---TA-   5330
VERVET.KE.AGM155    ............................................GATCCCAGAGAGGAGAGACCAGGAGGA......-T----ATTTGG-AT-TCAGC--------GTGGG-C------CTCAG---TA-   5835
VERVET.DE.AGM3      ............................................GATCCCAGAGAAGCAAGACCGGGGGAA......-T----ATATGG-ATTTGAGC--------GTGGG-C------CT-AG---CA-   5338
SABAEUS.SN.SAB1C    GGCTCCCACCAGTAGGAGGGGATCCTCCCAAGGATCCCCCCAAGAATCCCAGAGAAGAGATACCAGGATGG......-T----ACATGG-ATTTG-CC--A-----GTT-G-C------CTC-G---CA-   6050
SUN.GA.SIVSUN       ............................................AGGAGAGAGCCAGTAGAACAGCCACCA.........------G-G-------C---A------TTTGCAC-----CTGGC---TAC   5994
MANDRIL.GA.MNDGB1   ............................................GGCCAGAAGAGAGATGAGCAAGTATCA.........-----T-----A--T-CC--A------T-----C-G---CT-GC---TAC   5306
LHOEST.KE.hoest     ............................................CAGCAAAGAAGTCCAGAGAGACCACCA.........--G---TTT--A----CA---------TGGGG-C-G---TT-GC---CAC   5950
LHOEST.CD.hoest447  ............................................CAGCAAAGGAGTCCAGAAAAACCTCCA.........--G---TTT-------CA---------TGGGG-C-----CT-GC---CAC   4869
LHOEST.CD.hoest485  ............................................CAGCAAAGGAGTCCAGAAAGACCTCCA.........--G---TTT--A----CA--A------TGGGG-C-----TT-GC---CAC   4869
LHOEST.CD.hoest524  ............................................CAGCAGCGCAGCCCAGAAAGGCCCCCA.........--G---TTT--T--T-CA--A------TGGGC-C-----CT-GC---CAC   4866
RCM.NG.SIV-NGM      ...............................................GAAAGAGTGCCAGAGGCCCCCACTGGG............GCT--AGATGTA......GAGTTTGCCCCC---CTT-ACAGAA-   5968
COL.__.SIV-CGU1     ................................................GTCCTGAAAGGGACTTCACA................................................---TTG-AG--C-A   5320

SYKES.KE.SYK173     ..........................TTCAATCCCTCCCAGCATGTGCAAGGAACACCATGGTTCTTCATTCCAAG............-AATGT-G--TT-ACC---A--GTGATCAATGTCACT-TAAA   5807
SYKES.KE.SYK173 Vif                           L__Q__S__L__P__A__C__A__R__N__T__M__V__L__H__S__K__.__.__.__.__K__C__R__V__D__P__K__R__D__Q__C__H__C__K_    Vif
SYKES.KE.SYK173 Vpr .__.__.__.__.__.__.__.__.__F__N__P__S__Q__H__V__Q__G__T__P__W__F__F__I__P__.__.__.__.__R__N__V__E__L__T__P__N__V__I__N__V__T__V__K    Vpr
```

```
B.FR.HXB2 Vpr       __L__E__E__L__K__N__E__A__V__R__H__F__.__P__R__I__W__L__H__G__L__G__Q__H__I__Y__E__T__Y__G__D__.__.__.__.__.__.__.__.__T__W__A__G_   Vpr
B.FR.HXB2           TTTAGAGGAGCTTAAGAATGAAGCTGTTAGACATTTT...CCTAGGATTTGGCTCCATGGCTTAGGGCAACATATCTATGAAACTTATGGGGAT........................ACTTGGGCAGGA   5726

A.UG.U455           G-----A---------C------A-------------...------CAG----------A-----A------------A-C--C---------........................-------AG---   5172
C.ET.ETH2220        ------------C---C-G--------C-----C---...-----ACCA---------AA------A---T------------C---------........................------T-G---   5118
D.CD.84ZR085        -----------------G---------C---------...--------A-----T---A-------A---T----------------------........................--C---------   5248
F1.BE.VI850         ----------------------------G------...------CCA-----T--C--------C-----------A-C--C--------C........................-------AG---   5066
G.SE.SE6165         G-----A--------A---------------------...------C-------T-----------A--G---------A-C-------------........................-------A----   5123
H.CF.90CF056        ----------A----------G--------G------...-----AG-A---------CAA-----A--G---------A-C--C-----A---........................-------T----   5073
J.SE.SE7887         ---G-----------A-----------------C---...------CCA-----T---A-T-----A--GT--------AGC--------A---........................--A---A----   5040
K.CM.MP535C         ---G-------------GA------------------...------CCA---------AATC-G--A--G---------AC-------------........................-------A----   4922
01_AE.TH.CM240      G--------------A---------------------...------CCC--------A--------A--GT-C------A-C-A----------........................-------A---G   5294
02_AG.NG.IBNG       G-----A--------AC----------------C---...------CCG----------A-----A----------------A---------........................-------A---G   5251
03_AB.RU.KAL153-2   -----------------G---------------C---...------G-G-----T---A-------A---T-C-----------------A---........................------------   5771
04_cpx.CY.94CY032-3 ---G-----------A--------------G------...-----ACCC-----------C----A--G---------A-C--------A---........................--C----A---G   5092
N.CM.YBF30          A-----A--AT-A--A-----------GC-C------...--A-----------A------G----A-----C------A-C--A-----A--C........................--C----AG--G   5320
O.CM.MVP5180        CC-G--A-----G--AG-A-----A--A---------C...------CC----T-A--A-C--GT-----GT-C--T-----G--------A--C........................-------A----   5756
O.BE.ANT70          C-----A-----A--AGCA-----A--A---------C...------CC------A--C-C---G--A---T-C--T-----G-----------C........................-------T----   5780
CPZ.US.CPZUS        C-----A--A--A--AG-A--G---T-GC-C------...-----ACCC-----A--A---------A--GT-C-----CAGT--------A--C........................-------AG---   5786
CPZ.GA.CPZGAB       C--G--A--AT-A--A---------A--A--------C...-----ACCA---T-G---CAA-----A---TTC--T-----C--------A---........................--A----T----   5790
CPZ.CM.CAM3         GC-G-----A--A--------------AC-C------...---C-AGCA-----A------A-----A--GTT---T-----C--------A---........................--A--------T   5140
CPZ.CM.CAM5         A-----A-----A--A-C---G---C-A---------...---C-CCC------A---A-AC-----A--GT-C-----------------A---........................--A---AT---G   5428
CPZ.CD.CPZANT       ------A--AA-A--A-----G--A--A---------C...--ACA-CC-ACAT-A--ACA-C-T--AA-TTGGG-A----C--A---C--A---.......................T-A---AG----   5183

                                                                             Vif end in HIV-2 and SMM sequences <-|
H2A.GW.ALI  Vif     _S__R__G__L__K__Q__R__G__G__E__S__P__A__.__P__G__A__H__F__P__G__V__A__K__V__L__E__I__L__.__.__.__A__$                              Vif
H2A.GW.ALI  Vpx     __V__E__A__L__N__R__E__A__V__N__H__L__.__P__R__E__L__I__F__Q__V__W__Q__R__S__W__R__Y__.__.__.__W__H__D__E__Q__G__M__S__Q__S__Y__T_   Vpx
H2A.GW.ALI          AG------CCT-A--C-GA--G--G--G-AT--CC-G...--CC--GAGCTCA-TTTCCAGG-GT-----AGGTC--GGAG-TAC.........TGGCATGATGAACAAGGAATGTCACAAA-TTACAC-   6091

H2A.SN.HIV2ST       GG----A-CCA-A--C-GA--G--A--G-AC--CC-G...--CC-AGAGCTTA-TTTCCAGG-GT-----AGGTC--GGAG-TAC.........TGGCATGATGAACAAGGAATGTCA-TAA-TTACAC-   5543
H2A.DE.BEN          CA----A-CCT-A--C-GG--G--A--G-AC---C-G...--CC-AGAGCTCA-TTTCCAGG-GT-----AGGTC--GGAG-TA-.........TGGCATGATGAACAAGGGATGTCAG-AA-CTACAC-   6098
H2B.GH.D205         AA--ACA-----C--C-GG-T---G--C-AC-----G...--CC-AGAACTTA-TTTCCAGG-CT----GAGGTCT-GG-C-TAC.........TGGCGTGAGGAACAGGGCATGTCA-T-A-CTATACC   6074
H2B.CI.EHO          A-----AC-T--C--C-GA-T---A--C-AT--C--G...--C---GAACTTA-TTTCCAGG-CT----GA-GTCT-GG-C-TA-.........TGGCGTGAGGAACAGGGCATGTCA-T-A-TTATACC   6072
H2G.CI.ABT96        AG-MACA---A-C--C-GGAT---A--C-AT-----A...--C---GAGCTTA-TTTCCAGG-TT---G-AGGTC--GG---TA-.........TGGCATGATGAACAGGGCATGTCAG--A-TTACACT   5425
MAC.US.MM251        AG--------A-A--C-GA--G--A--A-AC--CC-A...--A---GAGCT-A-TTTCCAGG-TT-----AGGTCT-GG---TAC.........TGGCATGATGAACAAGGGATGTCACAAA-CTAT-T-   5988
SMM.US.SIVSMMH9     AS----A--AA-A--C-GG-C---A--G-AT-----G...--R---GAGCTAA-TTTCCAGG-CT---G-AGGKC--GG---TAC.........TGGCATGATGAAATGGGGATGTCAGTCA-CTACACT   5485
STM.US.STM          GG-------CA-A--CCGA--G--A--A-AC---C-A...--G---GAACT-A-TTTCCAGG-CT---G-AGGTCT-GG---TAC.........TGGCATGATGAGCAAGGGATGCCCGGGA-CTAT-TT   5659
TANTALUS.UG.TAN1    GG---CA--A--C--TC-G-----ACAAC-C-------...GGG--AGAACTC---TTCCAAG--T--A-CTTCTGTC-G---GAAGGG-AAAGG.............AATGGGGCACCA-TGATA-A-A-G   5948
GRIVET.ET.GRI-677   GC-CAG---T--A--CG-A-----CAGA--G--C----...GGA-T--ACAT---AATCC-AG--T--A-TT-CTGTGTA--GGAG............GGAAGGAGACATAATACCCCATGGAAT-AGAT-   5722
VERVET.KE.VER-9063  G-----A--T--C--TC-G-----CAAATT--------...GGC--AGAGCTC---TTCCAAG-TT--A-TT--TGTC-G--GGAA............GGAGAGAGACGGGGAGCCCCA-TGAC--AGA-G   5955
VERVET.KE.AGMTYO    G---C-A--T--C--CC-G--G--GAGGCTG-------...GGG--AGAACTC--GTTCCAAG--T--A-CT-CTGTC-G--GGAA............GGAGAAAGACATGGTACTCCC-TGAT--A-A-G   5445
VERVET.KE.AGM155    GG----A--AA-C--C--C-----CAAACTG-------...GGCC-AGAACTC--AT-CCAAG--T--A-TT--TGTC-G--GGAA............GGGGAGAGACAGGGAAGACCC-TAGC--A-A-G   5950
VERVET.DE.AGM3      GC----A--TA-C--TC-G-----CAAG-TG-------...GGGC-CGAGCTC--GTTCCAAG--T--A-CT--TGTC-G--GGAG............GGAGAAAGGAATCGCACTCCC-TGCTA-A-A-G   5453
SABAEUS.SN.SAB1C    G---C----C--C--TTCA--G--CCAGT-C--C--C...--A----A-CTC--TTTCC-TC-TT--TGGA----TGTG---GAACCA-CCAT-GATCATGGGCAAACCAGGCTGGAAGGC---TATAA-   6177
SUN.GA.SIVSUN       CA-G--A--AA-C-G-G-C-----G-AA--G-----C...--AGCAGCAATAG-TGCCCAAG-CTT----.........................TACTGTGCGGACTCTACAGGGTCAGAAAA--A---G   6097
MANDRIL.GA.MNDGB1   -A-G-----AA-A---G-A-----AAGA-AG--C--C...---CTC---ATC--AA---CAG--TCAG--.........................TATTGTGTGCAAAACACAGGGAGTGAGGAA-AG-CC   5409
LHOEST.KE.hoest     CA-G-----AA-A--AG-G--G--AAGAC-C-----C...--A...GAAGTAA-GTTG-C-CA--TCTCG--A.....................TATTGTGTAGACAGTACAGGGTCAGAAGA--AG-CC   6053
LHOEST.CD.hoest447  CA-G-----AA-A--AG-G--G--CAGAC-C-----C...--A...GAAACAA-TTTG-C-CA--TCTC-G-A.....................TACTGTGTCAGCAGTACCGGGTCAGAGGA-AA--CC   4972
LHOEST.CD.hoest485  CA-G-----AA-A---G-G--G--CAGAC-C-----C...--A...GAAGCCA-TTTG-C-CA--TCTCGG-A.....................TACTGTGTCAGCAGTACCGGGTCAGAGGA--A--CC   4972
LHOEST.CD.hoest524  AA-G-----AA-C---G-A--G--AAGAC-C-----C...--A...GAAGCAA-TTTG-C-CA--TCTCGG-A.....................TACTGTGTCAGCAGTACCGGGTCAGAAGA--AG-CT   4969
RCM.NG.SIV-NGM      GC--ACA--AG-C--CTTA-----CAGGTTG--C---CAT--A...GAG-TCA-TTTCC-TC-GT---GTAC-TGTGTG---CAC.........TGGCATGATAGGCTTGGAAGAAGCCT-GA-TAT-C-   6086
COL.__.SIV-CGU1     GG-----A-CA-A-CCCGG--G--CAGA--G--C--C...---TT--C-CACA--TCA-TGA--AT-TGGGG-TGTAC-AG--GG......C-CTGGGGGAGACCGTGGACGGCTGGT............   5429

SYKES.KE.SYK173     GGC---ACT-G-AGT--CA-----ATCC-AG------ACT---CAAGAAATATAT...--AG-TT--A-T--ATC--TAA-CGAA...-A--CAGGGACTGATTCCCCGACAATGGCTTGGGAAAG-ACC   5931
SYKES.KE.SYK173 Vif _G__R__T__G__S__D__R__S__I__Q__A__F__Y__S__S__R__N__I__.__W__S__L__E__S__I__L__K__R__.__R__G__R__D__$                              Vif
SYKES.KE.SYK173 Vpr __A__E__L__V__V__T__E__A__S__K__H__F__T__P__Q__E__I__Y__.__G__V__W__N__Q__S__L__N__E__.__E__A__G__T__D__S__P__T__M__A__W__E__R__T_   Vpr
```

```
                                            HXB2 Vpr premature end (without frameshift) <-|
B.FR.HXB2 Vpr       _V__E__A__I__I__R__I__L__Q__Q__L__L__F__I__H_#._F__Q__N__W__V__S__T__$                                                               Vpr
B.FR.HXB2 Vpr       _V__E__A__I__I__R__I__L__Q__Q__L__L__F__I__H__.__F__R__I__G__C__R__H__S__R__.__I__G__V__.__.__.__.__.__T__R__.__.__.__.__.__.__Q__   Vpr
                                      frameshift insert in HXB2  |
B.FR.HXB2           GTGGAAGCCATAATAAGAATTCTGCAACAACTGCTGTTTATCCATT..TTCAGAATTGGGTGTCGACATAGCAGA...ATAGGCGTT...............ACTCGA..................CAGA   5818
A.UG.U455           --T-----T------------T--------T-A--------T----..--------------C-A----------...------A--...............-T--C-..................GG--   5263
C.ET.ETH2220        --A-----TT---------C------------AA------------..-----------A--C-AG---------...------A--...............TTA---..................----   5209
D.CD.84ZR085        --T-----T-----------------------A--------T----..----------------A----------...------A--...............----C-..................----   5339
F1.BE.VI850         --T-----T--------G--AT----------A-------------..--------------C--T---------...------A--...............GT--C-..................----   5157
G.SE.SE6165         --T-----------------A--A--------A-------------..--------------C-A----------...------A--...............--A-C-..................-G--   5214
H.CF.90CF056        --T-----TT---------CG-----------A--------T----..--------------C-A----------...-----AA--...............------..................----   5164
J.SE.SE7887         --C-----------------A-----------AT-------T----..--------C-----C-AT---------...------A--...............-TA-C-..................--A-   5131
K.CM.MP535C         C-T-----T-----------AT----------A--------T----..--------------C-AC---------...------A--...............-TA-C-..................----   5013
01_AE.TH.CM240      --T-----T-----------GT----------A------G-T----..----------------A----------...------A--...............-TG-C-..................GG--   5385
02_AG.NG.IBNG       --TA----T------------T----------AT-----G-T----..----------------A----------...------A--...............-T--A-..................GG--   5342
03_AB.RU.KAL153-2   -----------------------------------------T----..-----G----------A----------...------A--...............-T--A-..................-G--   5862
04_cpx.CY.94CY032-3 --T-----T------------T----------A--------T----..--------------C-A------T---...------A--...............----CT..................--A-   5183
N.CM.YBF30          --A--G--A--T--C--G--A--A------T-A------------...-AT--G-----C--C-AG--C------...-----GA-C...............----CT..................--A-   5411
O.CM.MVP5180        --TATG--A--T--------CT-A--------A-------C----...-AT--------A--C-A------T---...-----AA--...............CTC-C-..................TCT-   5847
O.BE.ANT70          --TATG--A--T--------CT-A------T---------C----...-AT--------A--C-A------T---...-----AA--...............-AC-C-..................AGAG   5871
CPZ.US.CPZUS        --A-----A--T-----G--CT-A--------TT-----------...-AT--G-----C---AAT--------G...-----GA--TCCTTGGCAAGGAGA----CT............CAA......G   5892
CPZ.GA.CPZGAB       --A--G-----C--------CT-A-----CT---------------..--T---C-------C-AG---------...------A-C...............CTC-C-............CAA...AGA-   5884
CPZ.CM.CAM3         --A-----A--C--T-----A--------G--T-------------..--T--GC----C----A----------...-----GA-C...............-TC-CT..................GG--   5231
CPZ.CM.CAM5         --A-----A--T--T-----AT-------T--T-------G-----..--T--GC----C----A----------...-----GA--...............CT--YT..................G---   5519
CPZ.CD.CPZANT       ---C-G-TTT-G----TTC-C--T---A--GCT-------CA---...-AT---CAC--A---GCT--------G..---..................................................   5258

H2A.GW.ALI Vpx      _K__Y__R__Y__L__C__L__M__Q__K__A__M__Y__T__H__.__F__M__K__G__C__T__.__.__.__.__C__L__G__.__.__.__.__.__G__G__.__.__.__.__.__.__H__   Vpx
H2A.GW.ALI          AAAT-TAGATACT-GT-CT-AA---GA--GCTA---AC-CA---..----TG-AA----CACC............TGCCTG-GG...............GGAG-G..................--TG   6173

H2A.SN.HIV2ST       AA-T-TAGATATT-GT-CC-AA---GA--GCTA----C--A---...-CT-AG-GA-----CACT............TGCCTG-GG...............GGAG--..................--TG   5625
H2A.DE.BEN          AA-T-TAGATATT-GT-CC-AA----A--GCTA-A----CA---..----AG-GA-----CACT............TGCT-G-GG...............GAGGAC..................AT-G   6180
H2B.GH.D205         AAAT-TAGATACT-GTTGC-AA---GA--GCAA-----G-G--C...-AT-C--AG--C---A-G............TGCCTGCAG...............GAGG-C..................--TG   6156
H2B.CI.EHO          AAAT-TAGGTATT-GT-TC-AA----A--GCAA------A--C..--TGC--AG--T---G--............TGTCTGCGG...............GAAG-C..................--TG   6154
H2G.CI.ABT96        AA-T-TAGATATT-GTTGC-GA-----A-GGCTA-----G-A---...-GT-AG-AA--A--CAC-............TGCTTGCAG...............-AGG-C..................--TG   5507
MAC.US.MM251        AAAT-CAGATACT-GT-TT-AA-----A-GGCTT-A-----G---...-G--AG-AA--C---A--............TGTCTA-GG...............GAAG--..................--CG   6070
SMM.US.SIVSMMH9     AAATRCAGATACT-GT-CT-GA-A--GA--GCT--------G---...-G--AG-AA------A-G............TGCTTA-GA...............GGAGAG..................--TG   5567
STM.US.STM          AAAT-TAGATACT-GT-CT-GA-A--GA--GCTT-------G---...AG--AG-GG--A--CA--............TGCTTG-GG...............GAAG--..................--TG   5741
TANTALUS.UG.TAN1    -CTT-TAGATATTAT--GT-GG-A--GA-GGCTT-A---G-G---..--TC--TG---C--CA-GAGA--A-CT...CCTTTT-AG...............C-CTAC..................G-AG   6039
GRIVET.ET.GRI-677   -GCT-CAAGTACTAT------G-T---A-GTCTA-----G-A---..------TG---T---A--AGG--AG--...CCTTTTTCC...............C--TAC..................G-AG   5813
VERVET.KE.VER-9063  -CCT-TAGATACTAT--GT-GG-T--GA--GCA--C---G-T---..---C-TTG---T-----TAGA--ACAG...CCCTT--AA...............C--TAC..................G--G   6046
VERVET.KE.AGMTYO    -CCT-CAAATATTAT--GC-AG-A---A-GGCT--C---G-G---..--TC--TG------CA-GAGA--GCAG...CCCTTT-AA...............C-ATAC..................G--G   5536
VERVET.KE.AGM155    -CAT-TAAGTATTATC-CT-AG-T--GA--GCT--C---G-G---..---C-GTG---A-----CAGG--ACA-...CCCTTT-AG...............C-ATAC..................G--G   6041
VERVET.DE.AGM3      -CTT-TAAATATTAT-A-T-GG-----A--GCT--C---G-G---..---C-GTG---A--C--CAGA--ACA-...CCCTTT-AA...............C-ATAC..................G-AG   5544
SABAEUS.SN.SAB1C    TATTGTAGA--TC-G.........---A-GGCTT-A---G-G---..A-G-A-GG-A-A---T-TA-G.........CC-AAGAC-.........CACCCTG-GTATGGACCAGGAGCAGGAGGC-C-C   6277
SUN.GA.SIVSUN       TGTATGAGAGC---C-CTT-G--CA-TAG-GCCT----C--T---..C-GCC-...AT-----CTTC-TA--TCAGA--CA-AAG-...............GGATC-..................GGA-   6188
MANDRIL.GA.MNDGB1   TGT--GAAAT-T--T-CCT-AA--A-TAG-GCCA-T-GGG----C..C-AGCTCAA------GATGGA-C-TTC...-GG-AAAGA...............-GA-C-..................--AC   5500
LHOEST.KE.hoest     TGCATGAAAT-C--C-CCC-AG-CA-TAG-GCTT-----C-T--C..C--CC-...ACT----CCAGCTATGC-...GC--C-AGA...............G-AG--..................ACAG   6141
LHOEST.CD.hoest447  TGCATGAAAT-C--T-CC--AG-AA-TAG-GCA--C-----T---..--GCCT...ACC----CCAGCTATGC-...GC--CAAGA...............G-AG--..................ACAG   5060
LHOEST.CD.hoest485  TGCATGAGAT-C--T-CC--AG-AA-TAG-GCA--C-----T---..--GCCT...ACC----CCAGCTATGC-...GC-ACAAGA...............G-AG--..................ACAG   5060
LHOEST.CD.hoest524  TGCATGAGAT-C--C-CTT-AG-TA-TAG-GCT--C-A-T-A---..C-GCCT...CAA----CCAG-TATGC-...GC--C-AGA...............G-AG--..................ACAG   5057
RCM.NG.SIV-NGM      -GCT-TAGATATC-GCTTC-GA-----A--GCT--------T---..-G-CA-TCA------TCT--G..............AGA.......................................--TG   6159
COL.__.SIV-CGU1     ...CTG--TTATG-G-----CA--G-GTGTGCC--C-ACT-G--C............................................................GTT............CATGCGGG-C   5484

SYKES.KE.SYK173     A--TTG-A---GG-----GCA--CA-T-TTA----C---GAA---..--TGCTGCA--A----CT--AC-G-CC...-G-TAT-CC...............-GA-AC..................AGAG   6022
SYKES.KE.SYK173 Vpr _M__L__D__M__V__R__A__L__N__L__M__L__F__E__H__.__F__A__A__G__C__P__Q__R__T__.__R__Y__A__.__.__.__.__.__R__H__.__.__.__.__.__.__R__   Vpr
                                                                                |<- Tat exon 1 start
SYKES.KE.SYK173 Tat                                                             M__S__S__T__D__Q__.__I__C__Q__.__.__.__.__.__T__Q__.__.__.__.__.__.__R  exon1
```

```
                                 |-> Tat exon 1 start
B.FR.HXB2 Tat                    M__E__P__V__D__P__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__  exon1
                           HIV-1 Vpr normal end <-|
B.FR.HXB2 Vpr       R__R__A__R__N__G__A__S__R__S__$__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.    Vpr
B.FR.HXB2           GGAGAGCAAGAAATGGAGCCAGTAGATCCTAG..................................................................................................   5850

A.UG.U455           -A----GC--G--------------------A..................................................................................................   5295
C.ET.ETH2220        -A-----------------------------A..................................................................................................   5241
D.CD.84ZR085        -A---------------T-------------A..................................................................................................   5371
F1.BE.VI850         -A----T-------------------------..................................................................................................   5189
G.SE.SE6165         -A--G-T---GG-----C--G----------A..................................................................................................   5246
H.CF.90CF056        -A----T----------C-------------A..................................................................................................   5196
J.SE.SE7887         -A----GG--G--------------------A..................................................................................................   5163
K.CM.MP535C         -A----GG---------T-------------A..................................................................................................   5045
01_AE.TH.CM240      -A----GC--G------A-TG----------A..................................................................................................   5417
02_AG.NG.IBNG       -A----GC--G--------TG-----------..................................................................................................   5374
03_AB.RU.KAL153-2   --------------------------------..................................................................................................   5894
04_cpx.CY.94CY032-3 -----AG-G-C-GG---TGG--CCAG-AG.....................................................................................................   5212
N.CM.YBF30          -----AGG-ATGGAACCAGT--ATCC-AG.....................................................................................................   5440
O.CM.MVP5180        AC-C-AG-G---GA---AGA--A-ATGGA-CCAGTAGATCCTGA......................................................................................   5891
O.BE.ANT70          -A----G-----GAAAT-GATCC--TAGA-CCTGA...............................................................................................   5906
CPZ.US.CPZUS        -----AGG-ATGGATCCAGT--ATCC-AA.....................................................................................................   5921
CPZ.GA.CPZGAB       -A---T-C-ATGGATCCAAT--ATCC-GA.....................................................................................................   5913
CPZ.CM.CAM3         ----GAG--ATGGA-CCAGT--ATCC-GA.....................................................................................................   5260
CPZ.CM.CAM5         -A--GAG--ATGGAACCAGT--ATCY-AA.....................................................................................................   5548
CPZ.CD.CPZANT       .....-G-CATGGACCCCGT--AC-C-GA.....................................................................................................   5282

                                                                  Vpx ends ->|                          |<- Vpr starts
H2A.GW.ALI Vpx,Vpr  G__P__G__G__W__.__R__S__G__P__P__P__P__P__P__P__G__L__V__$                          M__T__E__A__P__T__E__F__P__P__A__G__M__G__P__H  Vpx,Vpr
H2A.GW.ALI          --CC--G-G--TGG...AGATCAG--C-TCCCCCTCCTCCCCCTCCAGGTCTAGTCTA...A......................TGACTGAAGCACCAACAGAGTTTCCCCCGGCCGGAATGGGACCCCA   6275

H2A.SN.HIV2ST       --CCG-G-G--TGG...AGATCAG--C-TCCCCCTCCTCCCCCTCCAGGTCTAGTCTA...A......................TGACTGAAGCACCAACAGAGTCTCCCCCGGAGGATAGGACCCCACC   5727
H2A.DE.BEN          -CC-G-A-G--TTG-A--A-CAAG--C-TCCCCCTCCTCCCCCTCCAGGTCTAGTCTA...A......................TGACTGAAGCACCAACAGAGTTTCCCCCAGAAGATGGGACCCCACG   6285
H2B.GH.D205         --CC--GGG--TGG...AGATCAG--C-TCCTCCTCCTCCTCCCCCAGGCCTGGCCTA...A......................TGGCAGAAGCAGCCCCAGAGATCCCTCCAGAGAACGAGAACCCACA   6258
H2B.CI.EHO          --CC--GGG--TGG...AGATCAG--C-TCCTCCTCCTCCTCCCCCAGGTCTAGCCTA...A......................TGGCAGAAGCAGTCCCAGAGATTCCTCCAGAGGATAAAAACCCACA   6256
H2G.CI.ABT96        --CCT-G-G--TGG...AGACAAG--C-TCCTCCTCCTCCTCCTCCAGGTCTAGCATAG.AT......................GGCAGAAGAAATCCCCCCA..........GAAGATGAGGCCCCCCA   5601
MAC.US.MM251        --GC--G-G--TGG...AGACCAG--C-TCCTCCTCCTCCCCCTCCAGGACTAGCATA...A.................................ATGGAAGAAAGACCTCCAGAAAATGAAGGCCCACA   6161
SMM.US.SIVSMMH9     --GC--GGG--TGG...AGACCAG-GC-TCCTCCTCCTCCCCCTCCAGGACTAGCATA...A...............................A.ATGACAGAAAGACCTCCAGAAGATGAAGCCCCACA   5659
STM.US.STM          --GC--G-G--TGG...AGACCAG-CC-TCCTCCTCCTCCCCCTCCAGGCTTAGCATA...G...............................A.ATGACACACAGACCTCCAGAGGATGAAGGCCCACA   5833
TANTALUS.UG.TAN1    AA--GAG--ATGGA-TC-GAG-GG--CGGG-TGGC...............................................................................................   6074
GRIVET.ET.GRI-677   A---GAG--ATGGACA--GAG-AG--G--...CCACCCCCTCCTCCAGGACTTGCATA...G....................................................................   5869
VERVET.KE.VER-9063  AA---AG-GATGGACA--GGG-AG--AGAGC-AACCGTGCTCCACCAGGACTTGATTA...G....................................................................   6105
VERVET.KE.AGMTYO    A---GAG-GATGGACA--GGG-AG-CAGAGCAAATCGTGTCCCACCAGGACTTGAGTG...A....................................................................   5595
VERVET.KE.AGM155    A---GAG--ATGGACA--GGG-AG--AGACCAGGACGTGTCCCACCAGGACTTGATTA...A....................................................................   6100
VERVET.DE.AGM3      AA--GAGGGATGGACA--GGG-AG--CGAGCAGGGCGCGTACCACCAGGACTTGATTG...A....................................................................   5603
SABAEUS.SN.SAB1C    CCCC--GTCTGGGA-----TTCAG--GGAGCTGCATCGGCCGCTCCAGGCCTGTGA..........................................................................   6333
SUN.GA.SIVSUN       ATGTCAACCC--GGCC-C-A-CA---C-AAGACCAGGGCAAGGGGACATTAGAAGAGG...CCTACAAGACTAA........................................................   6259
MANDRIL.GA.MNDGB1   T-CCCC-CTC-GGATTCAGGCCA---GGAG-TAGATTATAA.........................................................................................   5541
LHOEST.KE.hoest     CCCATC---C-G-A-C---ACCA---C-TA-TGAAGGAGACATACAACGAGGCAGTGA...CTAA.................................................................   6203
LHOEST.CD.hoest447  CTCACC-G-C-G-A-C---ACCA---C--ACTAAAAGAGACATACAAAGAGGCCGTTA...CTAA.................................................................   5122
LHOEST.CD.hoest485  CTCACC-G-C-G-A-C---ACCA-A-C--ACTGAAAGAGATATACAAAGAGGCCGTTA...CTAA.................................................................   5122
LHOEST.CD.hoest524  CACACC---C-G-A-C---ACC----C-AACTGGAAGAGATATACAAAGAGGCCATTA...CTGA.................................................................   5119
RCM.NG.SIV-NGM      -ACAG-G-CA-GCAA-G-AAGCAG--GAAAGAATCCAGATTCTTCCGGGAATGTAA..................TCAAGCACTCCTCTCATTGCAGAGATGGAGATGCTCCCGGAAGATGAAGAACCACA   6271
COL.__.SIV-CGU1     AAT-TCGCTTTCCCTCTA-GCAA-ACC-T---..................................................................................................   5516

SYKES.KE.SYK173     --TACC-CCATCC-TCTTAG..............................................................................................................   6042
SYKES.KE.SYK173 Tat __V__P__P__S__F__L__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._  exon1
SYKES.KE.SYK173 Vpr G__Y__P__H__P__S__$                                                                                                                   Vpr
                Vpr end in SIVSYKES <-|
```

```
B.FR.HXB2 Tat        .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  exon1
B.FR.HXB2            ..................................................................................................................................  5850

A.UG.U455            ..................................................................................................................................  5295
C.ET.ETH2220         ..................................................................................................................................  5241
D.CD.84ZR085         ..................................................................................................................................  5371
F1.BE.VI850          ..................................................................................................................................  5189
G.SE.SE6165          ..................................................................................................................................  5246
H.CF.90CF056         ..................................................................................................................................  5196
J.SE.SE7887          ..................................................................................................................................  5163
K.CM.MP535C          ..................................................................................................................................  5045
01_AE.TH.CM240       ..................................................................................................................................  5417
02_AG.NG.IBNG        ..................................................................................................................................  5374
03_AB.RU.KAL153-2    ..................................................................................................................................  5894
04_cpx.CY.94CY032-3  ..................................................................................................................................  5212
N.CM.YBF30           ..................................................................................................................................  5440
O.CM.MVP5180         ..................................................................................................................................  5891
O.BE.ANT70           ..................................................................................................................................  5906
CPZ.US.CPZUS         ..................................................................................................................................  5921
CPZ.GA.CPZGAB        ..................................................................................................................................  5913
CPZ.CM.CAM3          ..................................................................................................................................  5260
CPZ.CM.CAM5          ..................................................................................................................................  5548
CPZ.CD.CPZANT        ..................................................................................................................................  5282

                                                                                                                                 |<- Start Tat  exon1
H2A.GW.ALI Tat                                                                                                                     M__E__T__P__L  exon1
H2A.GW.ALI Vpr       __Q__G__A__R__D__E__W__V__I__E__V__L__R__E__I__K__E__E__A__L__R__H__F__D__P__R__M__L__I__A__L__G__G__Y__I__Y__T__R__H__G__D__T__L_    Vpr
H2A.GW.ALI           CCAGGGAGCCAGGGATGAGTGGGTAATAGAAGTCCTGAGAGAAATAAAAGAAGAAGCTTTAAGGCATTTTGACCCTCGCATGCTAATTGCTCTTGGCGGCTATATCTATACTAGACATGGAGACACCCTT  6405

H2A.SN.HIV2ST        GAGGGAGCCAGGGGATGAGTGGGTAATAGAAACCCTGAGAGAGATAAAATAAGAAGCTTTAAAGCACTTTGACCCTCGCTTGCTAATTACTCTTGGCAACTATATCTATGCTAGACATGGAGACACCCTT  5857
H2A.DE.BEN           GAGGGACTTAGGGAGTGACTGGGTAATAGAAACTCTGAGGGAAATAAAGGAAGAAGCCTTAAGACATTTTGATCCCCGCTTGCTAATTGCTCTTGGCTACTATATCCATAATAGACATGGAGACACCCTT  6415
H2B.GH.D205          AAGAGAACCGTGGGAAGAGTGGATAGGGGAGATCCTGGAGGAAATAAAGCAAGAAGCCTTAAAGCATTTTGATCCTCGCTTGCTAACTGCGCTTGGTAACTTTATCTACAGTAGGCATGGAGATACCCTT  6388
H2B.CI.EHO           AAGAGAACCGTGGGAACAGTGGGTAGTGGACGTCCTGGAGGAAATAAAACAAGAAGCCTTAAAGCATTTTGACCCCCGCTTATTAACTGCACTTGGCAATTTTATCTACAATAGGCATGGAAATACCCTT  6386
H2G.CI.ABT96         GAGAGAGCCATGGGATGAATGGGTAGTAGAAGTTCTGGAAGAAATCAAAGAAGAAGCCTTAAAACATTTTGACCCTCGCTTGCTAACTGCGCTTGGTAACTATATCTATGATAGGCATGGAGACACCCTC  5731
MAC.US.MM251         AAGGGAACCATGGGATGAATGGGTAGTGGAGGTTCTGGAAGAATTGAAAGAAGAAGCTTTAAAACATTTTGATCCTCGCTTGCTAACTGCACTTGGTAATCATATCTATAATAGACATGGAGACACCCTT  6291
SMM.US.SIVSMMH9      GAGGGAACCGTGGGATGAATGGGTRGTAGAAGTTCTGGAGGAACKAAAAGAAGAAGCCCTGAACCAKTTTGATCCTCGCTTGCTAATYGCGCTTGGTAACTATATCTATGATAGACATGGAGACACCCTT  5789
STM.US.STM           GAGAGAACCATGGGATGAGTGGGTAGTGGAAGTTCTGGAGGAAATAAAACAAGAAGCCCTGCGGCATTTTGATCCTCGCTTGCTAAGCGCGCTTGGTAACTATATCTATGATAGACATGGAGACACCCTT  5963
TANTALUS.UG.TAN1     ..................................................................................................................................  6074
GRIVET.ET.GRI-677    ..................................................................................................................................  5869
VERVET.KE.VER-9063   ..................................................................................................................................  6105
VERVET.KE.AGMTYO     ..................................................................................................................................  5595
VERVET.KE.AGM155     ..................................................................................................................................  6100
VERVET.DE.AGM3       ..................................................................................................................................  5603
SABAEUS.SN.SAB1C     ..................................................................................................................................  6333
SUN.GA.SIVSUN        ..................................................................................................................................  6259
MANDRIL.GA.MNDGB1    ..................................................................................................................................  5541
LHOEST.KE.hoest      ..................................................................................................................................  6203
LHOEST.CD.hoest447   ..................................................................................................................................  5122
LHOEST.CD.hoest485   ..................................................................................................................................  5122
LHOEST.CD.hoest524   ..................................................................................................................................  5119
RCM.NG.SIV-NGM       GAGAGAACCCTATGATGAATGGCTAATGGATACCCTTATAGAGATACAAGAAGAAGCTAAGAAACATTTTACACCTGAATTGCTTACACAAGAAGGAAATTATATATATGAGCAGCATGGAGATAGTTTA  6401
COL.__.SIV-CGU1      ..................................................................................................................................  5516

SYKES.KE.SYK173      ..................................................................................................................................  6042
SYKES.KE.SYK173 Tat  _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__  exon1
```

```
B.FR.HXB2 Tat       .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  exon1
B.FR.HXB2           ..................................................................................................................................   5850

A.UG.U455           ..................................................................................................................................   5295
C.ET.ETH2220        ..................................................................................................................................   5241
D.CD.84ZR085        ..................................................................................................................................   5371
F1.BE.VI850         ..................................................................................................................................   5189
G.SE.SE6165         ..................................................................................................................................   5246
H.CF.90CF056        ..................................................................................................................................   5196
J.SE.SE7887         ..................................................................................................................................   5163
K.CM.MP535C         ..................................................................................................................................   5045
01_AE.TH.CM240      ..................................................................................................................................   5417
02_AG.NG.IBNG       ..................................................................................................................................   5374
03_AB.RU.KAL153-2   ..................................................................................................................................   5894
04_cpx.CY.94CY032-3 ..................................................................................................................................   5212
N.CM.YBF30          ..................................................................................................................................   5440
O.CM.MVP5180        ..................................................................................................................................   5891
O.BE.ANT70          ..................................................................................................................................   5906
CPZ.US.CPZUS        ..................................................................................................................................   5921
CPZ.GA.CPZGAB       ..................................................................................................................................   5913
CPZ.CM.CAM3         ..................................................................................................................................   5260
CPZ.CM.CAM5         ..................................................................................................................................   5548
CPZ.CD.CPZANT       ..................................................................................................................................   5282

H2A.GW.ALI Vpr      _E__R__A__R__E__L__I__N__A__L__Q__R__A__L__F__M__H__F__R__A__G__C__G__R__S__R__V__G__Q__T__R__G__R__N__P__L__S__A__I__P__T__P__R__    Vpr
H2A.GW.ALI Tat      __K__E__P__G__S__S__L__M__P__Y__N__E__P__S__S__C__T__S__E__Q__D__V__A__V__Q__E__L__A__K__Q__G__E__E__I__L__S__Q__L__Y__R__P__L__E_  exon1
H2A.GW.ALI          GAAAGAGCCAGGGAGCTCATTAATGCCCTACAACGAGCCCTCTTCATGCACTTCAGAGCAGGATGTGGCCGTTCAAGAGTTGGCCAAACAAGGGGAAGAAATCCTCTCTCAGCTATACCGACCCCTAGAA   6535

H2A.SN.HIV2ST       GAAGGCGCCAGAGGGCTCATTAGGATCCTACAACGAGCCCTCCTCTTGCACTTCAGAGCAGGATGCGGCCGCTCAAGGATTGGTCAGCCCAGGGGACGAAATCCTTTATCAGCTATACCAACCCCTAGAG   5987
H2A.DE.BEN          GAAGGCGCCAGAGAGCTCATTAAAACCCTACAACGAGCCCTCTTCGTGCACTTCAGAGCGGGATGTAACCGCTCAAGAATTGGCTAAACAAGGAGAAGAACTCCTTGCCCAGCTGCACCGACCCCTAGAG   6545
H2B.GH.D205         GCAGGAGCAGGAGAGCTCATTAAAATCCTCCAACGAGCCCTCTTCCTCCACTTCAGAGCCGGTTGTCAACACTCAAGGATTGGACAATCAGGGGGAGGAAATCCTCTCTCAACTATACCGCCCCCTTAAG   6518
H2B.CI.EHO          GAAGGAGCAGGAGAGCTCATTAAACTCCTCCAGCGGGCACTCTTCCTCCACTTCAGAGGGGGTTGCCAACACTCAAGGATTGGACAACCGGGGGGAGGAAATCCTCTCTCAGCTATACCGCCCTCTTAAG   6516
H2G.CI.ABT96        GAGGGAGCAGGAGAACTCATTAAAATCCTGCAGAGAGCTCTCTTCCTGCACTTTCGAGGAGACTGTAGATGCTCCAGGATTGGAGGGAACGCAGGCGGAAATCCTCTATCAACTATACCGCCCTCTAGAG   5861
MAC.US.MM251        GAGGGAGCAGGAGAACTCATTAGAATCCTCCAACGAGCGCTCTTCATGCATTTTAGAGGCGGATGCAACCACTCCAGAATCGGCCAACCTGGGGGAGGAAATCCTCTCTCAACTATACCGCCCTCTTGAG   6421
SMM.US.SIVSMMH9     GAAGGAGCAGGAGAGCTCATTAGAATCCTCCAGGGAGCACTCTTCATCCATTTCAGAGGTGGATGCAGACACTCCAGAATCGGCCAGTCTGGAGGARGAAATCCTCTCTCAACTATACCGCCCKGTAGAG   5919
STM.US.STM          GAAGGAGCAGGAGAGCTCATTAAGATCCTCCAGCGAGCCCTCTTCATGCACTTCAGAGGCGGTTGCCGCCACTCCAGGATTGGCCAACCAGGAGGAGGAAATCCTTTGGCAACTATACCGCCCACTAGAG   6093
TANTALUS.UG.TAN1    ..................................................................................................................................   6074
GRIVET.ET.GRI-677   ..................................................................................................................................   5869
VERVET.KE.VER-9063  ..................................................................................................................................   6105
VERVET.KE.AGMTYO    ..................................................................................................................................   5595
VERVET.KE.AGM155    ..................................................................................................................................   6100
VERVET.DE.AGM3      ..................................................................................................................................   5603
SABAEUS.SN.SAB1C    ..................................................................................................................................   6333
SUN.GA.SIVSUN       ..................................................................................................................................   6259
MANDRIL.GA.MNDGB1   ..................................................................................................................................   5541
LHOEST.KE.hoest     ..................................................................................................................................   6203
LHOEST.CD.hoest447  ..................................................................................................................................   5122
LHOEST.CD.hoest485  ..................................................................................................................................   5122
LHOEST.CD.hoest524  ..................................................................................................................................   5119
RCM.NG.SIV-NGM      GAGGGAGTAAAAGCTATGATAACCTTGCTTAATAAAGCACTTTTCTTGCACTTCAGGCATGGATGTGAAGGGAGCAGGATTGGAGGAGCGAGAGGAGGAAATCCTCTTAGATCCATTCTTCACTCGAGAA   6531
COL.__.SIV-CGU1     ..................................................................................................................................   5516

SYKES.KE.SYK173     ..................................................................................................................................   6042
SYKES.KE.SYK173 Tat .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  exon1
```

```
B.FR.HXB2 Tat       .__.__.__.__.__.__.__L__E__P__W__K__H__P__G__S__Q__P__K__T__A__C__.__T__N__C__Y__C__K__K__C__C__F__H__C__Q__V__C__F__I__T__K__A__L  exon1
B.FR.HXB2           ...................ACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAAACTGCTTGT...ACCAATTGCTATTGTAAAAAGTGTTGCTTTCATTGCCAAGTTTGTTTCATAACAAAAGCCTT  5958

A.UG.U455           ...................C--------------A--C--G-------------C----------...-G---C--T---------GT--------GG---------C----C--TC-G-A-----G---  5403
C.ET.ETH2220        ...................C-----A--------C--------------------G---------...-ATC-A--T-----------A---A---A------T-T------C--TC-G--.----G---  5348
D.CD.84ZR085        ...................CA----T--------C-------------------GG---------...-A------T-------------------A---------------C--------G----G---  5479
F1.BE.VI850         ...................CT----T--------T-----------C-------C----C-----...-----A--T---------CGA------------------TGG--C--T-C---G--G-G---  5297
G.SE.SE6165         ...................C--------------T-----G--G---------------C-C---...-A---A--T-T-------GT--------GG--------------C--TC-G-AC----G---  5354
H.CF.90CF056        ......................................C------------------C-----------...-A------T--------------C----A----------A-G--C--TT---AG----G---  5304
J.SE.SE7887         ...................CAG------------T------------------------------...--T-----T-------------------A------T-----G--C--TC--CAG--G-G---  5271
K.CM.MP535C         ...................CA-------------C--A---------------------------...-A-C-G------------CG--------A----------A-A--C--TT---AG--G-G---  5153
01_AE.TH.CM240      ...................C--------------T-----G-------------C----------...-G---G--T--C--------A-------GG---------C-A--C--TC-G-A-----G---  5525
02_AG.NG.IBNG       ...................C--------------C--C--G-------------C----------...-G---G--T--C-------T---C----G.---------T-G--C--TC-G-AC--G-G---  5481
03_AB.RU.KAL153-2   ...................---------------------------C------------------...--------------------------TC----------------------G-A-----G---  6002
04_cpx.CY.94CY032-3 ............ATCCTGAC--------------C-----G-------------C-----A----...-A---G--T-TC----------------GG--------------C--TC-G-A-----G---  5327
N.CM.YBF30          ...................-T-------------T-----------C--A--------A-----C...-AT----------------GA-------A---C---TT-TA---C----C--AG----G---  5548
O.CM.MVP5180        ...................GA-GCCC--T---C-T--C--T--G--CA----CC----CC-----...-AT-----------C----GA--C----A-------T-T----------C--AG--G-GT--  5999
O.BE.ANT70          ...................GG-GCCC--T---C-C--C--T-----------CC-G-TCC-----...-AT-----------C----GA--C----A-------T-T---------G---G---G-GT--  6014
CPZ.US.CPZUS        ...................CA----------------G--------C--------G--------C...-AT--------C--------------------C--TGTG--------T-CG-AG--G-G---  6029
CPZ.GA.CPZGAB       ...................C--G-----------A--C-----C--------A-G---A-T----...-AT-----------C---GCT-----T-A---C--TAT-TA---C----C--A-----G---  6021
CPZ.CM.CAM3         ...................T------------CTA-----------C-------G---------C...-AT--C-------------T---C------------AT---------TGAG-A-----G---  5368
CPZ.CM.CAM5         ...................T------------CTA-----------------------------C...-AT--C--T---------TGT--C------------ATG-----C--TGAG-A-----G---  5656
CPZ.CD.CPZANT       ...................-AC-CCT--A---TTA--C--TCCTGC-ACA--AGC---GC-A--C...-AT--C--T--C------TGC-----T-AC-------CTC-C--C----C--A-----G---  5390

H2A.GW.ALI Vpr      N__M__Q__$                                                                                                                          Vpr
H2A.GW.ALI Tat      _T__C__N__N__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._T__C__Y__C__K__E__C__C__Y__H__C__Q__L__C__F__L__N__K__G__L  exon1
H2A.GW.ALI          ACATGCAATAA...........................................................--CA-----------GG-A--C----AC--------GC-G-----TT---AT--G-GAC-  6606

H2A.SN.HIV2ST       GCATGCGATAA...........................................................---A--T--C-----------C----AC--------GA-G-----TT---AC--G-GGC-  6058
H2A.DE.BEN          CCATGCACTAA...........................................................---A-----------GCGA--CA-T--C--------GC-G------TCG-A---G-GGC-  6616
H2B.GH.D205         GCATGCGATAA...........................................................T-CA-----C-----G--A--C----AC--------GC-------TC-T-A---G-GTC-  6589
H2B.CI.EHO          GCATGCAGTAA...........................................................T-CA-----C--------A----C--AC--------GC----C--TC-T-A---G-G-C-  6587
H2G.CI.ABT96        GCGTGTTCTAA...........................................................T--A--T-----C-----A--C----A------T--GT-A--C--TC-T-AC--G-GAC-  5932
MAC.US.MM251        GCGTGCTATAA...........................................................--CA----------------------AC--------GT-------TC-T-A---G-GA--  6492
SMM.US.SIVSMMH9     GCGTGTTATAA...........................................................---G--Y--C-----G--A-------A------T--GCA------TC-Y-A---G-G---  5990
STM.US.STM          GAGTGTTGTAA...........................................................---G----T---C-----A-------A---C-----GC-G-----TG-T-------GTC-  6164
TANTALUS.UG.TAN1    ......................----AG-CTCCTC--GGACTTGCA-AG----TT---AC-A---ACT-A---A----T---C----GA--C----A---C--T--------C--TC-GCAG--G-G---  6182
GRIVET.ET.GRI-677   .................................................---ATT-CAAC-A---ACT-AT--A--------C-----A--C--T-A---C--TG-GC----C---C-GCAG--G-G---  5950
VERVET.KE.VER-9063  .....................................A-A-TAT-AGA----GCT---AA-----AGA-AT--G----T---C-------------A---C--T--GC-------TC-GCA-----G-C-  6198
VERVET.KE.AGMTYO    .....................................AGACTATCAGA-----CTGCAGA-----AAA-AT--A--T-T---C-----A-----T-AC--C--T--GC----C--TC-GCA-----G-C-  5688
VERVET.KE.AGM155    .....................................A-A-TAC--GA-A--CCTTGAGA-A---ACA-AT--A----T---C-----A--C--T-A----------T-C--C---T--CGG----GAC-  6193
VERVET.DE.AGM3      .....................................A-A-CTC-AAGCA--CCTG-AGCGG---ACA-A---G------------TGT--C--T-A---C--T--GC----C--TT--CA---G-GA--  5696
SABAEUS.SN.SAB1C    ...................................................................T-AT-C-----TC--C---GTT--C-----------TATTC-C--C---CAT-A---G---C-  6396
SUN.GA.SIVSUN       ....................................................CTTGGAG--C---GAT-A---A----GG----GG-GA--C--------C--T---C----C--TC-GCAG----GG--  6337
MANDRIL.GA.MNDGB1   .........................--AACT-CTAG-AGAGTATTA-------TTGCAA------GAG-AT--A----GG--C--G--A--C-----------TATGC----C--TCA--AG--G-GT--  5646
LHOEST.KE.hoest     .................................................AG-CCTGCAA--G---CAG-AT-CC--T-GC--------A--------------TATTC----C--TCAGCAG----GG--  6284
LHOEST.CD.hoest447  .................................................---ACTG-G---C---AAC-A---G----GC-----------------------TATTC----C--TCAGCAG--G-GA--  5203
LHOEST.CD.hoest485  .................................................---ACTG-G--TC---GAC-A---G----GC--C----G---------------TATTC----C--TCAGCAG--G-GA--  5203
LHOEST.CD.hoest524  .................................................T--ACTCCAAT-C--CCAG-AT--G---C-C--C---GTT-----------C--TAT-C----C---CAGCAG--G--G--  5200
RCM.NG.SIV-NGM      ACATCTTGTAA...........................................................T-GC--T-----C----GA---GTG-A---C--T--GC----C--TC-TCAG--G--T--  6602
COL.__.SIV-CGU1     .........................................................AG--G---AATGA---A----GG-----GGCC-----------C--T--GC-A--C--TC-TT------GGC-  5589

SYKES.KE.SYK173     ..............................AAGGAACATTCCT-GAGA-AGGACC-C-AA-ACCATGT-AT--A----T---C-----C-------A---------GC--------C-TCA---G-GA--  6142
SYKES.KE.SYK173 Tat .__.__.__.__.__.__.__.__.__.__E__G__T__F__L__E__K__G__P__P__T__P__C__N__K__C__F__C__K__N__C__C__Y__H__C__Q__L__C__F__L__Q__K__G__L  exon1
```

```
                                |-> Rev start except in macaques and mangebeys                   Tat and Rev exon end \/ intron
B.FR.HXB2 Rev                   M__A__G__R__S__G__D__.__.__S__D__E__E__L__I__R__.__.__T__V__R__L__I__K__L__L__Y__Q__S__S                               exon1
B.FR.HXB2 Tat       __G__I__S__Y__G__R__K__K__R__R__.__.__Q__R__R__R__A__H__Q__.__.__N__S__Q__T__H__Q__A__S__L__S__K__Q                                exon1
B.FR.HXB2           AGGCATCTCCTATGGCAGGAAGAAGCGGAGA......CAGCGACGAAGAGCTCATCAG......AACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAG.TAA..........................  6049

A.UG.U455           ------T----------A-----------A-......-CC------G--C---C---A......GG----A--GA-------A--CT-A--C--------.---..........................  5494
C.ET.ETH2220        ------T---------------------C--......----------------C---A......-G----A--GA-------AA-CT-A-----------.---..........................  5439
D.CD.84ZR085        -------------------------------......------------C---C---C......-G-----------------A-C--A--C--------.---..........................  5570
F1.BE.VI850         -------------------------------......-------AC---A---C---A......-G------AGTG------AA-------C--------.---..........................  5388
G.SE.SE6165         -----------------------------AG......--C------G--A---C----......-G----A-AGGC-------A-C--G--C--------.-G-..........................  5445
H.CF.90CF056        ---A--T-----------------------C......-------AC---A---C-GCA......-GTTTG--AGA-------AA----A-----------.---..........................  5395
J.SE.SE7887         ------T------------------------......---------TC-----C--CT......GG----A------------A-CT-A--C--------.---..........................  5362
K.CM.MP535C         ------T---A--------------A-----......-CC-------C-A---C-T-C......------G---AC-------A-C-----CG-------.-G-..........................  5244
01_AE.TH.CM240      -----------------------------AG......--C------G--A---C----......-G----A--GA-------AA-C--A--C--------.---..........................  5616
02_AG.NG.IBNG       -------------------------------......-GC------G--A---C----......-G-C-----GA-------AA-C--G--C--------.-G-..........................  5572
03_AB.RU.KAL153-2   -------------------------------......----------------C---A......G---A-------G------T-------C--------.---..........................  6093
04_cpx.CY.94CY032-3 -----------------------------A-......--T------G--T---T---A......GG---CA--GGC------AA-CT-A--C--------.---..........................  5418
N.CM.YBF30          ---------A--------------------T......--A---------A---C----......-G----A---G--------A-CT-A--C--G-----.---..........................  5639
O.CM.MVP5180        G--A------C----------------A---......AGA-C-GC-GCT---GCAAGC.........TA--CAGA-A--A---A-C--G--C--G-----.---..........................  6087
O.BE.ANT70          G--A-----------------------AG--......-GA-C-GCTGCT--AAGC--T............-CAGA----A---A-C--G--C--------.---..........................  6099
CPZ.US.CPZUS        ---A--T---------------------G-GCGA...-GCA------C-A--GC-A--AGC......---GC-GA-A------A-C-----CGC------TA-.GTG.......................  6126
CPZ.GA.CPZGAB       ---A-----------------------A-CCACA...AGA-----C-CG----C-GCAGGC......---A-A-A-A------A----A--C--------TA-.GTA.......................  6118
CPZ.CM.CAM3         ---A-----------------------A--G......GGA------C-CTTGTC-GCTGCA...-G----AC-CA-A------G-------CAGC-----.---..........................  5462
CPZ.CM.CAM5         ---A-----A-----------------A--G......GGA---A--CACT-G-T-GCTTCA...-G----AC-CA-A------A----A--CG-C-----.---..........................  5750
CPZ.CD.CPZANT       G--G--T----------------GAA-AGCTCGA...AG-AACA--C--A-AAC-GCTGAA...-G----G-A-A-A------A-C--G--------T--TA-.GTA.......................  5490

                                |<- Rev exon 1 start                                             Tat and Rev exon end \/ intron
H2A.GW.ALI Rev                  M__T__E__R__A__G__E__.__.__E__D__L__Q__R__K__.__.__.__.__L__R__L__I__R__L__L__H__Q__T__N                               exon1
H2A.GW.ALI Tat      __G__I__W__Y__D__R__K__G__R__R__.__.__R__R__S__P__K__K__.__.__.__.__I__K__A__H__S__S__S__A__S__D__K                                exon1
H2A.GW.ALI          C--G--A-GG----A-C-A---GGCA--C--......AGAA--TCTCC-AAGA-A............-T-A--G-----TCGT-----GC----G-CA--.-G-..........................  6691

H2A.SN.HIV2ST       C--G--A-GG----AAC-A---GGCA-A---......AGAA--ACTCCGAAGA-A............-C-A--G-----TCGT-----GC----G-CA--.-G-..........................  6143
H2A.DE.BEN          C--A--A--A----AGC-A---GGCA-AC--......AGAA-GACTCC-AGGA-A............-C-A------C-TCGC-----GC-C--G-CA--.-G-..........................  6701
H2B.GH.D205         T--G--A-GT----A-C-CTC--GAA-----......TCTGC-AA-------A-G..................---AC-GC-C-----GC-C--G-CA--.-G-..........................  6668
H2B.CI.EHO          G--G--A-GT----AAC-CTC--GAAA----......TCTTC-AA---G---A-G..................---AC-AC-T-----GC-C----CG--.-G-..........................  6666
H2G.CI.ABT96        C--AG-A-GT----AGCAACCT-GAA-A---...AGAACT-CGAA--AGA--A-G..................G--A--TC-TT----GC------CG--.---..........................  6014
MAC.US.MM251        G--G--A-GT----AGCA-TCACGAAA----......AGAA--ACTCCGAAAA-G............GC-A--G--A--AC-T-----GC------CA--.---..........................  6577
SMM.US.SIVSMMH9     G--A--A-G-----AGCA-C-ACGAA-A---......ACT-CGAAG-AGA--A-G..................G--A--AC-TT----GC----G-CA--.---..........................  6069
STM.US.STM          T--A--AA------AGC-ATCA-GAA-A---......GTTAAGAA--AG---A-G..................---T---C-AT----GC------CA--.---..........................  6243
TANTALUS.UG.TAN1    -------A-----CATGTCTCT-G-ATC---......AGA-CTAAG-AGAA----...............TCA-A------GAA-CT-G-----C-----.---..........................  6264
GRIVET.ET.GRI-677   ---AG-GCGT---CATGTCTCT-G-AAA---......AGAAA-AC---CA----A..................GA-A------ACC--A--CGGC-----.---..........................  6029
VERVET.KE.VER-9063  ---TG-TA-----CATGCCCCT-G-ACC---......AGAAAGAA---T-T---G..................C--A---GCCTA----A-CAGG-C---.---..........................  6277
VERVET.KE.AGMTYO    ---TG-TA-----CATGCCCCT-G-ACC---......AGAAAGAAG-TTCG-TCG..................CT-A--TTG---C--T--CAGC-C---.---..........................  5767
VERVET.KE.AGM155    ---T--TA-----CATGCCTTT-G-ACC---......AGAAAGAAG-TT---TCG..................G--G---GCAT-C--G--C-GC-----.---..........................  6272
VERVET.DE.AGM3      ---TG-TA-----CATGCCCCT-G-ATC---......AGAAAGAAG-TT----CG..................CT-G---GCTT-C--GA-CA---A---.-G-..........................  5775
SABAEUS.SN.SAB1C    ------TCG----TATGTCCCT-G--CA---......AGAGCTTCT-AGAAG..................ATTT-G---A-TCAAGT-TCTCT-C-CA-C.--G..........................  6475
SUN.GA.SIVSUN       ---G---CA----TATGTCT-C-G-G-AC--......----AT-A-TC-ATATC-..................-AGG--A-GCAAGAGGCTGT-CGAAG-.---..........................  6416
MANDRIL.GA.MNDGB1   ---A--AAGG--CCATGTCT-C-G-AAAC-T......GTA-C-G---CTAA-A-GA--.........-TA-CTGG-AG-GGT-AAGAAGCTAT-CGAAG-.---..........................  5734
LHOEST.KE.hoest     ---GC-GCA----TATGTCC-C-G-AAAC-G......AGA---GTT-CCT-GATA-TT...............GAGAC---GCAGGA--CTGTGG-CAG-.---..........................  6366
LHOEST.CD.hoest447  ---G--ACGA---TATGTCC-C-G-A-----......A-AA--GTCGCCCAGGTATTT...............GAAG--A-GCAAGA--CTGTGG-CGG-.---..........................  5285
LHOEST.CD.hoest485  ---G--ACGA---TATGTCC-C-G-A-----......AGAA--GTT-CCCAGGTATTT...............GAAG--A-GCAAGA--CTGTGG-CGG-.---..........................  5285
LHOEST.CD.hoest524  ---G--ACGG---TATGTCC-C-G-ACC---......AG-A--GTT-CCTACATA-TT...............GAAAC-G-GCAGGA--CTGTGG-CAG-.---..........................  5282
RCM.NG.SIV-NGM      ------AAA-----CTTCT-G--GAA-A---......AGAG-CA--CC--GAA--...AAGAAG--G-A-----T-A---C-TT-GA-AC---TG-----.---.GTATGGAATGTCCTGGTCTGTACCT  6721
COL.__.SIV-CGU1     G--A---CG----AAG-A---T---...................................................AG-GCTA-ACAAAC-CAGT-A.................................  5635

SYKES.KE.SYK173     G--A---A-T----C---ACCA-G-AA-C--......GCAGCT--C--TAT-.....................T--G-GG-T-A-AGCGC-C-T-CAGG-.---..........................  6218
                                                                                             Tat and Rev exon 1 end \/ intron
SYKES.KE.SYK173 Tat __G__I__T__Y__A__R__P__R__K__R__.__.__A__A__R__S__I__.__.__.__.__.__.__.__S__E__D__D__S__A__P__T__G                                exon1
SYKES.KE.SYK173 Rev             M__P__D__Q__G__S__E__.__.__Q__L__A__V__F__.__.__.__.__.__.__.__L__R__M__I__A__H__L__Q                                  exon1
                                |<-Rev start except in macaques and mangebeys
```

```
                                                                                                                (ACG start codon in HXB2, ATG in others)
                                                                                                                              |-> Vpu start
B.FR.HXB2 Vpu                                                                                                                 T__Q__P__I__P__I__V__A__    Vpu
B.FR.HXB2           .....................................................................................GTAGTACATGTAACGCAACCTATACCAATAGTAGCAA   6086

A.UG.U455           .................................................................................GTAAT---CTT------T-AC----T-GGA---CTGG----   5535
C.ET.ETH2220        .............................................................GTAC..CAAATAATAGATGTAATG--T-ATTTAC--G-AA--GTAGATTAT-G-A---T--   5498
D.CD.84ZR085        .....................................................................................-------------T----T--T---A----T------   5607
F1.BE.VI850         ...........................................................................GTAGTGTTAAA-----T------T-TC-TA-T-GTT-GC-A---GG-   5435
G.SE.SE6165         ...........................................................................GTAGAAAAAAT-----T------T---GT-AT--GT----T------   5492
H.CF.90CF056        .....................................................................................---T--TCATA--T-T-TATAT--GG-T---G-AT-G   5432
J.SE.SE7887         ...........................................................................GTAAATCAAGTA.-C-T------T-AT----T-G-A-----C-----   5408
K.CM.MP535C         ...........................................................................GTAGTGCTAA-------------T-GTGT-CT--G----TTC-AT-G   5291
01_AE.TH.CM240      .................................................................................GTAATA----T------T-AC----T-GGA---TAGT----   5657
02_AG.NG.IBNG       ...........................................................................GTAGTAATAAT---C-T----G-T-------T--A-----ACT----   5619
03_AB.RU.KAL153-2   .....................................................................................-------------T----T-CT--G------C-----   6130
04_cpx.CY.94CY032-3 .................................................................................GTATTA----T------T-TT-TTCTGGGA---CTGG----   5459
N.CM.YBF30          ...........................................................................................GTAAA-C-TGT-TA---G-TGTC-T-G-G-T   5670
O.CM.MVP5180        ........................................................................................---ACGC-G-T---T-AAGAGAACC-GC----CT   6121
O.BE.ANT70          ........................................................................................---ACGC---T---T-A--GGGACC-GC------   6133
CPZ.US.CPZUS        ..................................................................................CCC...T-TTG--CT-T--TTAA-TGGTTTGA-A---GTT   6163
CPZ.GA.CPZGAB       ................................................................................................T-T-ACT-TGT--GTCGGCT---TTC   6144
CPZ.CM.CAM3         ..................................................................................GTAACTTC-----CT-T--T-A--TGGGA-CA-A---GCT   5502
CPZ.CM.CAM5         ..................................................................................GTAA-TT-G----CT-T--T-AT-TGGGAGCA-A---GTT   5790
CPZ.CD.CPZANT       ..................................................................................CTC.........T---T-ACTAA----TTT..........   5511

                                                                                                                      |-> Env start
H2A.GW.ALI Env                                                                                                        M__M__S__S__R__N__Q__L__  gp120
H2A.GW.ALI          ...............................................................................................GT-T-ATGT---GTAG--ATCAGCTGC   6718

H2A.SN.HIV2ST       ...............................................................................................GT-A-ATGTG-GGTAGG-ATCA-CT-T   6170
H2A.DE.BEN          ...............................................................................................GT-T-G-G---GGTAGG-ATCAGCTGT   6728
H2B.GH.D205         ...............................................................................................GT-T-GC-TA-T-TAGC-GCCGCCTGC   6695
H2B.CI.EHO          ...............................................................................................GT-T-GC--A-G-TAAT-ATTACCT-C   6693
H2G.CI.ABT96        ...............................................................................................GT-T-GC-TA-C-TGGC-ATCAGCTGC   6041
MAC.US.MM251        ...............................................................................................GT-T-GG-TG-C-TGGG-ATCAGCTGC   6604
SMM.US.SIVSMMH9     ...............................................................................................GT-T-GG-TG-C-TGGG-ATCAGCTGC   6096
STM.US.STM          ...............................................................................................GT-T-GCCTGCCCTGG--ATCA-CTGC   6270
TANTALUS.UG.TAN1    ...............................................................................................GT-T-GG---AT--AGGGG-AA--GGG   6291
GRIVET.ET.GRI-677   ...............................................................................................GT-T-GGGAGAT-G-TT---AA-AT-C   6056
VERVET.KE.VER-9063  ...............................................................................................GT-T-A-----C-TTT-GGTA---TTC   6304
VERVET.KE.AGMTYO    ...............................................................................................GT-T-AGGTA--C-AT----ACCTT-G   5794
VERVET.KE.AGM155    ...............................................................................................GT-T-AC-AAGT-CTT-GG-A-TTTT-   6299
VERVET.DE.AGM3      ...............................................................................................GT-T-A-G-TG-C-TT-C-G...AT-G   5799
SABAEUS.SN.SAB1C    ............................................................................................TAAGT-T-A-G-T-C-TA--G--C-GCTTT   6505
SUN.GA.SIVSUN       ...............................................................................................GT-A-ATG-GGTGC--TGA-A--TTG-   6443
MANDRIL.GA.MNDGB1   ...............................................................................................GT...ATGAAGTGT---GGTC-TTTTC   5758
LHOEST.KE.hoest     ...............................................................................................GG-G-ATGG-ATGT---GG-T---G--   6393
LHOEST.CD.hoest447  ..................................................................................................GAATGAACTGC--TGG-C--AT-C   5309
LHOEST.CD.hoest485  ..................................................................................................G-ATGGACTGT-TT-AGC-CCT-C   5309
LHOEST.CD.hoest524  ..................................................................................................AAATGG-ATGT---GGTAACAT-T   5306
RCM.NG.SIV-NGM      GCTTTTA..GATGCTTAGCTTATAAGGTGATAGGAGATATAATTAGAGAAGTTTTTATTTGTAGAGTTTATAGAAAATAGAAGGTAA-TCTTTA-AT-T-GCTAAGT--AT--GTA---G--   6841
COL.__.SIV-CGU1     .........................................................................................GTATGC-T-GAT-T-TCT-TAT--ACT-GCT-G   5668

SYKES.KE.SYK173     ..............................................................................................----AAATGG-AGCTTTT-G-ACTTAC-   6246
SYKES.KE.SYK173 Env                                                                                                      M__A__A__F__R__T__Y__  gp120
                                                                                                                         |-> Env start
```

```
B.FR.HXB2 Vpu      I__V__A__L__V__V__A__I__I__I__A__.__.__.__.__I__V__V__W__.__S__I__V__I__I__E__.__.__Y__.__.__.__R__K__I__L__R__Q__R__K__I__D__R__L    Vpu
B.FR.HXB2          TAGTAGCATTAGTAGTAGCAATAATAATAGCA............ATAGTTGTGTGG...TCCATAGTAATCATAGAA......TAT.........AGGAAAATATTAAGACAAAGAAAAATAGACAGGTT   6186

A.UG.U455          --AC--GGC-GA--------C----CT-----............-----A------...A-T------GGT------......--AAA.......-AA---T-GC---AG----A---------------   5637
C.ET.ETH2220       ----------CA--------C----C------............------------...A------C-TAT------......---.........------T-G-------------G------T-----   5598
D.CD.84ZR085       --------------T----GC-----------............------------...A--------T--------......---.........C-T-G----AA---G-----------------T----   5707
F1.BE.VI850        ---C-------A--------C----------C............------------...A-T------TAT------......---.........-A----C-GG----G------------A-T-AA--   5535
G.SE.SE6165        ------G----A--------T-C---GC---C............------------...A--------T-T------......---.........---G-----AG--A---G-A--G-----GA-A---   5592
H.CF.90CF056       G--CGCT-G-----AC-TTT--C---GCC...............G--A--------...A--------TAT------......---.........-AA---T-GG----G----AG------------C-   5529
J.SE.SE7887        ----------TA----------TT-CC-T---............----G-A-----...A-T-----GTAT------......---.........-A----C-G-----------G--------T-A---   5508
K.CM.MP535C        ---C-TT-G------C-CT----T--GC-...............---A----A---...A--------TAT------......---.........------T-GG---A-----A--G----A--T----   5388
01_AE.TH.CM240     ------G-C-GA-------GC----CT-----............-----A-----A...A-T------GCT------.....GT-.........-A-------C----G-------------------   5757
02_AG.NG.IBNG      ------G-C-----------T-C---GC---C............------------...AG-------TAT------......---.........---------AGG-A---G-AG--------------   5719
03_AB.RU.KAL153-2  ------------------T-GG----------............---------G--...---------T--------......---.........-----------------------------------   6230
04_cpx.CY.94CY032-3 ------G-C-G--------GC----T----T-............-----A-------...A-TT-----T-T------......---.........-A----T-GAG---G-----G-G---------C--   5559
N.CM.YBF30         -CA----G----G--C----G-T-GC------............G--A-A--C----...G-AT--C--TAT-G----......---.........-A-------AA-TTG--GGA-------A-ACACA-   5770
O.CM.MVP5180       --A----T---AGT-CTTTGTGTC-T---AAT............G--C--A-A----...-TGT-TAACC-T-G-ATT......---.........TTAGTGCA-AG--A----GAT-G--GG--GCA-GA   6221
O.BE.ANT70         --A--ATTA-TAGT-CTTTGC-GT-T---AAT............G--A--T-A----...GGGT-TA-TC-T-G-A--......---.........TTAG--CA-AAGGA----GAC-G--AG--G--AGA   6233
CPZ.US.CPZUS       --A---------G-A---A-GGT--TT---T-............G-GA-AA-C----...GGAC----TGCT-G-TTG......-GGAGACAAATA-AA-T--A-GA--ATACTCA-C--...........   6261
CPZ.GA.CPZGAB      -CA--TT-G---GCT--...--TGC-TGGAAC............--TTG-A-A----...GGATACA-T--T-A-TGG...GGA---AGAAGGTAT-AA-G-CATAGGCTTG-G-C-G--...........   6242
CPZ.CM.CAM3        --A-T-GTA---G-A---A---T--T-----T............--T--A-CC----...GGG----CTT-T-A---G......-GG......AGA-AAGG--A-GA-GA-G---AC-G-GC-C-A-AAA-   5605
CPZ.CM.CAM5        --A----T----G-A---A-C-T--T--T-TT............--T--A--T----...GGA----CTTAT-A-A-G......-GG......AGAGAA-T--A-GA-GA-AGG-AC-G--CCC-A--AA-   5893
CPZ.CD.CPZANT      ........GAGTAT-CTTTTC-TGCCT-TAGT............-----AC-A----...ATT--CTGT--ACCTAT-CTCTATA-GCTTTATAAA-TAT-T-AGCAGCAG----T-G-T-ATA-G--AAA   5618

H2A.GW.ALI Env     L__V__T__I__L__L__A__S__A__C__L__V__.__.__Y__C__K__Q__Y__V__T__V__F__Y__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  gp120
H2A.GW.ALI         -T--TA-TA-CT--C----T-GTGCTTGCTT-GTA......TATTGTAAACAA-ATGTGA-TG-TT-TTAT............................................................   6783

H2A.SN.HIV2ST      -T--T--CAGCT-GC----T-GTGCTTGCTT-ATA......TATTGC--CCAA-ATGTGA-TG-TT-CTAT............................................................   6235
H2A.DE.BEN         -T--T-TCA-TT--C--A---GTGCTTGCTT-GTA......TATTGTAGCCA--ATGTGA-TG-TT-CTAT............................................................   6793
H2B.GH.D205        CTA-T--GC-CC-GC-TAT-GGT--C-GT-GGTTT......GTATGTAAACAA-ATGTTA-TG-CT-CTAT............................................................   6760
H2B.CI.EHO         -T--TA--C-CC-GC-TAT--GT--CTAT-GGTAT......ATGGGCAAGAAC-TTGTCA-TG-CT-CTAT............................................................   6758
H2G.CI.ABT96       -TA-C---ACCT-RC-CTT--GT-CTTAT-GGTTC......TATTGTACCAAA-ATGTCA-TG-TT-TTAT............................................................   6106
MAC.US.MM251       -TA-C--CA-CT-GC-TTT--GTG-CTAT-GGATC......TATTGTAC-CAA-ATGTCA-AG-CT-TTAT............................................................   6669
SMM.US.SIVSMMH9    -TA-C--GC-CTCGC--TY--GTGCTTC--GGATT......TATTGT---CAA-ATGTAA-AG--T-CTAT............................................................   6161
STM.US.STM         -TA-C--TA-CT-GC--TT--GTGCTTGCTTGACC......TATTGCAC-CA--ATGTCA-GG--T-TTAT............................................................   6335
TANTALUS.UG.TAN1   --T--TT-G--A-TT-G-G-T---GCT--AT-GGACTGTTATATGGGACACA--ATATTA-AG-GT-TTAT............................................................   6362
GRIVET.ET.GRI-677  --A--AT-GC-A---GGAT--GT---GG-AT-GGTAAC............C---ATGTGA-AG-GT-TTAT............................................................   6115
VERVET.KE.VER-9063 -TA----TA-CA---G-AT--GTTCCT--TT-AGTGAAAGA.........CAA---GTCA-AG-GT-CTAT............................................................   6366
VERVET.KE.AGMTYO   G-A--AT-G-GA---G-AT-GGG---G-GTT-AGTAAG............CAA---ATAA-AG-CT-TTAT............................................................   5853
VERVET.KE.AGM155   -----TT-GG-A---GGAT-GG----GGGAT-AGTACAAAA......CAGCA----ATAA-AG-GT-CTAT............................................................   6364
VERVET.DE.AGM3     GGA--CT----A---GGAT-GG-G--G-GCTTAATACAAGG......CAACAA---GTCA-AG--T-TTAT............................................................   5864
SABAEUS.SN.SAB1C   GGC--AGTGGGTGTTGGAGCT--G-GTGGTT-GTC...............CA--ATGTTA-AG-GT-TTAT............................................................   6561
SUN.GA.SIVSUN      ---GTTTC-CTT-GC-GTT-................GGTAGTATAGC-A-ACA--ATGTAA-AG--T-CTAT............................................................   6499
MANDRIL.GA.MNDGB1  -TA--CT--GTT---CC.T-C--G--..........GGTATTATAGG-AAACAA-ATGTGA-AG-CT-CTAT............................................................   5819
LHOEST.KE.hoest    -TC-GCTT--GC-TC-T...............GGCATAATATGGGG-AAACAA-ATGTAA-AG--T-TTAT............................................................   6449
LHOEST.CD.hoest447 -TC-CTTGA-CA-TT--AG-G--CA-......GGGGAGAAAACA-AGAAGCA--ATGTAA-AG-GT-CTAT............................................................   5374
LHOEST.CD.hoest485 -TT-CTTGC-CT-TT---G-GC-CA-......GGGGAG...GCA-AGAAGCA--ATGTAA-AG--T-TTAT............................................................   5371
LHOEST.CD.hoest524 -GC--TTGC-GC-CTGCTTGGG-C---C----GGAGAA...AAGGAGAAACAA-ATGTAA-AG--T-CTAT............................................................   5374
RCM.NG.SIV-NGM     -TA--AT-GG-T-GT---GT-----T----T-GGAGAGAAA.........CAA-ATGTGA-AG--T-TTAT............................................................   6903
COL.__.SIV-CGU1    --T---GTA----GT--AGT-AT...........................AAA---GTAA-AG--TATCAA............................................................   5712

SYKES.KE.SYK173    ----T...-GTC-TT-TAGCT------GTTT-GGATTTATGGAA-A-CAACAA-ATGTAA-AG--T-TTAT............................................................   6314
SYKES.KE.SYK173 Env I__V__.__C__L__F__S__L__I__S__L__G__F__M__E__K__Q__Q__Y__V__T__V__F__Y__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  gp120
```

```
                                                                                |-> Env gp160 start in HIV-1
B.FR.HXB2 Env                                                                   M__R__V__._#__K__E__K__Y__Q__H__L__W__.__.__.__.__.__.__.__R__W__G__W_  gp120
B.FR.HXB2 Vpu       __I__D__R__L__I__E__.__.__.__.__.__.__.__R__A__E__D__S__G__N__E__S__E__#_.__G__E__I__S__A__L__V__E__.__.__.__.__.__.__.__M__G__V__    Vpu
B.FR.HXB2           AATTGATAGACTAATAGAA.....................AGAGCAGAAGACAGTGGCAATGAGAGTG.....AAGGAGAAATATCAGCACTTGTGG.....................AGATGGGGGTGG   6269

A.UG.U455           -T-AA-C---A---G----.....................----------------------------.....-T--G--T-C-GAG-A-T-ATCCT.....................T-C...TT----   5717
C.ET.ETH2220        ----A-A---ACT-GG---.....................-----------------------A----.....-T--G--T-C-GAG-A-T-GTCAA.....................CA-...T-----   5678
D.CD.84ZR085        ----------A---G----.....................----A-----------------------.....----G--T-A-GAG-A---ATCAA.....................CCC...TT----   5787
F1.BE.VI850         -TA-A-A---A---G----.....................----------------------------.....-G--G--TGC-GAG-A-T-G-CA-.....................CAC...TT-G--   5615
G.SE.SE6165         -C--------A---G----.....................-----------------A----------.....-C--G--T-C-GAG-A-T-G--AA.....................CAC...TT----   5672
H.CF.90CF056        ------A---A--GG----.....................--------------------C-------.....-T-----C-C-GAG-A-T-ATCCA.....................--C...TTA---   5609
J.SE.SE7887         ----------A---G----.....................-----------------T--C-------.....-T-----C-C-GA--A-T-G-CA-.....................-CC...TT----   5588
K.CM.MP535C         ---A------A---G----........................AG--C---AGACA-TGGC......A.....-T-AGAGTGAGGGG-ATGCA-A--.....................-AT---CA-ACA   5462
01_AE.TH.CM240      -G--A-G---A---G----.....................-----------------A----------.....-------C-C-GAT-A-T-G-CCA.....................-AC...TT----   5837
02_AG.NG.IBNG       -C--------A---G----.....................-----------T----------------.....-T--G--T-C-GA--A-T-ATCCA.....................CTC...TTA---   5799
03_AB.RU.KAL153-2   ----------A---G----.....................----------------------------.....-------TCAGGA-------A---.....................---......---   6307
04_cpx.CY.94CY032-3 GTACA-----A---G----.....................----------------------------.....-T--G--TGC-GAG-A-T-ATCCA.....................CAC...TT----   5639
N.CM.YBF30          --GAC-G---A---G----.....................----A------T-----------A----.....-T--G--TGC-GAGTGGT-G-AT-.....................G-G...AT-AA-   5850
O.CM.MVP5180        ---ACT-GA-AG-T--AGGAGAATAAAGGAAATCAGGGATGACAGT--CT-TGAAA-T-----AGAA-.....--CA-C-GGA-GTCATGGAGC-TA.....................TAC...ATAGCC   6322
O.BE.ANT70          ---ACT-GAGAGGT--AG-AGAATAAGAGAAATTAGGGATGATAGT--CT-TGAAA------GAGAA-.....-G-A-C--GA-GTTATGGACC-T-.....................TAC...TTAGCC   6334
CPZ.US.CPZUS        ....--G-TT-A--ATCTTTTAGAAAGA............-TCAG-AT-AGAGAA-AAG--AGTG-AA.....-T-A-AGTGATGG--A-GAA-AA-AGACTCTGGCTAAGT............TAT--T   6358
CPZ.GA.CPZGAB       .-----G-----T-ACCT-ATATTGAGA............GA-AG--C---AGA-A-TGGAA...........-T-A-AGT-ATGG--A-GAA-AA-AGAGACTGGAACAGCTTATCC-T-ATTACAATC   6348
CPZ.CM.CAM3         T-GGAC-TT-A--GA-AG-ATAATATCA............---CAG-----T-----A-----A----.....-T------GATCA-AGG-AGC-T-.....................-T-GA-AAA-AC   5697
CPZ.CM.CAM5         TTACA--TT-GC-GA-AG-ATAAGAATA............---CAG-----T-----A-----A--CA.....-T-----GGA-GA---T-AAT---.....................-T-ATTT-C-AC   5985
CPZ.CD.CPZANT       TCAAAGA-TCA--GA--TGTTAAGTAGA............---CTTAGTATAGA-A-TGCAATAGAA-AAG..-T-AG---GCCGATA--TA-TATT......TGGGGT.........CTGGCTTT-CTA   5719

H2A.GW.ALI Env      __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._  gp120
H2A.GW.ALI          ..................................................................................................................................   6783

H2A.SN.HIV2ST       ..................................................................................................................................   6235
H2A.DE.BEN          ..................................................................................................................................   6793
H2B.GH.D205         ..................................................................................................................................   6760
H2B.CI.EHO          ..................................................................................................................................   6758
H2G.CI.ABT96        ..................................................................................................................................   6106
MAC.US.MM251        ..................................................................................................................................   6669
SMM.US.SIVSMMH9     ..................................................................................................................................   6161
STM.US.STM          ..................................................................................................................................   6335
TANTALUS.UG.TAN1    ..................................................................................................................................   6362
GRIVET.ET.GRI-677   ..................................................................................................................................   6115
VERVET.KE.VER-9063  ..................................................................................................................................   6366
VERVET.KE.AGMTYO    ..................................................................................................................................   5853
VERVET.KE.AGM155    ..................................................................................................................................   6364
VERVET.DE.AGM3      ..................................................................................................................................   5864
SABAEUS.SN.SAB1C    ..................................................................................................................................   6561
SUN.GA.SIVSUN       ..................................................................................................................................   6499
MANDRIL.GA.MNDGB1   ..................................................................................................................................   5819
LHOEST.KE.hoest     ..................................................................................................................................   6449
LHOEST.CD.hoest447  ..................................................................................................................................   5374
LHOEST.CD.hoest485  ..................................................................................................................................   5371
LHOEST.CD.hoest524  ..................................................................................................................................   5374
RCM.NG.SIV-NGM      ..................................................................................................................................   6903
COL.__.SIV-CGU1     ..................................................................................................................................   5712

SYKES.KE.SYK173     ..................................................................................................................................   6314
SYKES.KE.SYK173 Env __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._  gp120
```

```
                                                       Vpu end <-|
B.FR.HXB2 Vpu       E__M__G__H__H__A__P__W__D__V__D__D__.__.__L__$                                                                                    Vpu
B.FR.HXB2 Env       _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__L__F__  gp120
B.FR.HXB2           AGATGGGGCACCATGCTCCTTGGGATGTTGATG......ATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACC.........ACCACTCTATTTT   6384

A.UG.U455           --------A--T---A--T-G---T--A-A--A.........AT-T--AA-G--C--C-------------G-----C--------------------A--T---GTT.........-----C-------   5829
C.ET.ETH2220        -T------A-T-T-AGG-T--T-------A---......--T----A--GA-TG-GG--C--------------------------------A-----A--T--T-G-.........C-T----------   5793
D.CD.84ZR085        -A--------T----------------------...ATG-C--A-------G----T--C--------------T-----------------A-----A---------.........-----G-------   5905
F1.BE.VI850         -A-------CTTT-AT----G-----A--A--A......-------A----G----C--C--------------------------------------A---------.........--T--------C-   5730
G.SE.SE6165         -A------G--TT--A--------T--G----A......--T--------CT--A-T--T--------T---------------------------G-A--T---GAT.........-------------   5787
H.CF.90CF056        ------------T-AA--T-G--T------C--......------------G--C----C--------T-----A-----------------------A--G----AA.........-----------C-   5724
J.SE.SE7887         ---G-----CT----A--T----A-----A---......--T----AG--A-A----G-C--------------------------------------A--T----AG.........------T------   5703
K.CM.MP535C         TTGG--AA-TGGGGCA--T-ATTCT--GGA--ATTGATA--------CAA-G----C-----------------------------------------A--------A.........C----------C-   5583
01_AE.TH.CM240      -A------G--TT--A--------T--G----A......--T--------CT----C--C--------T-----T-----------G----------GA--T---GAT.........-----C-------   5952
02_AG.NG.IBNG       --------T--G-ATA-AT--T----AA----A.........AT-T--AA-G-T---C-------------G-----C---------------------ACG---GAG.........-----C-------   5911
03_AB.RU.KAL153-2   ...G-CAC-TTGT-C--TGGGAT-C--A----C.........-----------------T--A-----------T-----------------------A---------.........-------------   6416
04_cpx.CY.94CY032-3 GA------A--TT--A--------T--G----A......-----------CT--A-C--C--------------T----------------------GA--C---GAG.........-----C-------   5754
N.CM.YBF30          --TG-TT-GTTAC-CT--TA-CTTC-AG-A-GCTTGATC-AGGTA-T--GGT-T---C--CAT-----A-----G--C-----------A--A----GA------GAG.........--A-----T--C-   5971
O.CM.MVP5180        -TG......G-TT----AA-C......CCATGT.........-TG---TA--GTA--C------AT-C------T----C------------A---G-A--G---G-A.........C-AGTA-----C-   6422
O.BE.ANT70          -TG......G-TT--A-AACC......CCATGT............TTGAGCCTTAG-C-GC-A-AT-CA---------GC------G-----A---G-A--T-----A.........C-AGTA-----C-   6431
CPZ.US.CPZUS        CTTCTCTCG-G-T--A-AA-CCCAGGA---........................TCT-GC--A----CT-----A-----------------A----GA--T-T-GAG.........--A--CT----C-   6455
CPZ.GA.CPZGAB       -T-ACAAT--TTT----AACCCCATGT----CC.....................TCTG-G--A-----A-----A--------A--------T---C-T--T--TGA-.........C-GGTA--C----   6448
CPZ.CM.CAM3         -TG-TTATGGGTT--A-AA-CCCATGT----CTGGTAGT...............--TC-G-------CA-----G--C--------------------A--G---GA-.........--A--A--C----   5803
CPZ.CM.CAM5         -TGCTAATGGGTT--A-AACCCCATGT----CTGGC..................C-TG-G-G-----CA-----G--C-----A--------A-----A--G---GAA.........--A--A--T--C-   6088
CPZ.CD.CPZANT       -TCCA-TTT-TAGA-AAGGGGAC-..............................A-TG--GAC-AT--A-----A-TC-----A--C-----C----GAA-T--G--A.........C-T----------   5810

H2A.GW.ALI Env      _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G__V__P__A__W__K__N__A__S__.__.__.__I__P__L__F__  gp120
H2A.GW.ALI          .................................................................................--C--G--A-CA-----AA-T---T--.........-TTC-C--C----   6823

H2A.SN.HIV2ST       .................................................................................--C--G--C-------GAA-T---T--.........-TTC-C--C----   6275
H2A.DE.BEN          .................................................................................--CA----C-C------AA-T---T-T.........-TTC-CT------   6833
H2B.GH.D205         .................................................................................--CA----C-CA----G-A-C-----A.........GTTC-C--CA---   6800
H2B.CI.EHO          .................................................................................--TA----C-CA-----AA-T---T-A.........-TTC-C--C----   6798
H2G.CI.ABT96        .................................................................................--CA----C-CA----G-A-C--G-G-.........GTAC----C--C-   6146
MAC.US.MM251        .................................................................................--T-----A-CT----G-A-T--G--A.........-TTC-C--C--C-   6709
SMM.US.SIVSMMH9     .................................................................................--T-----A-CA----G-A-T--G--A.........GTTC-C--C--C-   6201
STM.US.STM          .................................................................................--T-----A-CA----G-A-T--G--G.........-TTC-C--C--C-   6375
TANTALUS.UG.TAN1    .................................................................................--TA-C--A--A-----AA-CAGCT-A.........GTGCAAGCT--C-   6402
GRIVET.ET.GRI-677   .................................................................................--AA-C--A--A-----AA-TT----A.........GTTCAGGC-----   6155
VERVET.KE.VER-9063  .................................................................................---A-------A------A-CAGCT-A.........GTACAAGCC--C-   6406
VERVET.KE.AGMTYO    .................................................................................--AA----A--A-----AA-CAGCT-G.........GTGCAGGCT--C-   5893
VERVET.KE.AGM155    .................................................................................--A-----A--A-----AA-CAGCT-A.........GT-CAAGCT----   6404
VERVET.DE.AGM3      .................................................................................--A-----A--A-----AA-CAGCT-A.........GTACAGGCT--C-   5904
SABAEUS.SN.SAB1C    .................................................................................--TA-C--A--A-----AA-TAGTT-A.........GT-CAAGCC----   6601
SUN.GA.SIVSUN       .................................................................................---ACT---AA----G-ACC----GTA.........GTGC-AT--A-A-   6539
MANDRIL.GA.MNDGB1   .................................................................................--A-----A--A-----------T-AA.........--ACA-T-GA---   5859
LHOEST.KE.hoest     .................................................................................-----G--AAAT---G-T--TAATGTTTCA......GTGC-CT-GA---   6492
LHOEST.CD.hoest447  .................................................................................-----G---AAA---G-A------GTA.........GTGC-AT--A---   5414
LHOEST.CD.hoest485  .................................................................................-----G---AA----G-A--T--TGTA.........GTGC-AT--A---   5411
LHOEST.CD.hoest524  .................................................................................--A-----AAAA---G-A--T--TGTA.........GTAC--T--A-C-   5414
RCM.NG.SIV-NGM      .................................................................................--T--C--A--A------A-CAGT--A.........GTTC--A-G----   6943
COL.__.SIV-CGU1     .................................................................................--A-------CA---G-A------GATGTAAATGATCAGCAGT-C--C-   5761

SYKES.KE.SYK173     .................................................................................---A----ACAT---G-A--T---TAT.........G-AC----C----   6354
SYKES.KE.SYK173 Env _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G__I__P__H__W__E__D__A__Y__.__.__.__A__P__L__F__  gp120
```

```
B.FR.HXB2 Env       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__N__M__W__K  gp120
B.FR.HXB2           GTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGGAA   6514

A.UG.U455           --------------------------G----A--G--------C-----T---------------------------------------A---AT-----C------------G-A--------------   5959
C.ET.ETH2220        -------------------------------A-----------C----GT---TT---------C-------------G----------T-G-GT----A------------------------------   5923
D.CD.84ZR085        -----------------T-----A-A-------C-------A-C-----T---------------------------------------A---A-C--AA------C-----------------------   6035
F1.BE.VI850         -------------------------A-G-----C------C---------------------------------------------------T-TC--AA------A------------G-T--------   5860
G.SE.SE6165         -------C---------T-----AG-G-T--AAGC--------C-----T---------------------------------------A--A-TA---A------A------T----------------   5917
H.CF.90CF056        ----------------G--------G-----AAAG--------C-----T--------A-----------------------------GA-G--CA---AG-----A-----G-GC-----T------G-   5854
J.SE.SE7887         -C-----T---------------AG---T--AAA---------C-----T--------T--------------T--A-G---------GA-GAATC--CC------A--------C--------------   5833
K.CM.MP535C         -------------------------G-AG--------------C-----T--C-----------------G----------------------AGA---A------A--------C--------------   5713
01_AE.TH.CM240      -------------C--G---C----G-----A--G--C-----C---------------------------------------------A--CACC---A------A-----------------------   6082
02_AG.NG.IBNG       -------------------------------A-----------C-----T--------------------------T------------A--CAT----A------A--------G--------------   6041
03_AB.RU.KAL153-2   -----------------------AG--AG-----------------------T-----------------G-------G----------A--CC----AA-------------------------G----   6546
04_cpx.CY.94CY032-3 ----------A--------------G-A---A---------A-C-----T-------------------------------------------CTC--A-------A-----G--C--------------   5884
N.CM.YBF30          ----T--------------CC--AG--------CT--C--CA-C-----------A--A-----T--T--T--T-----T-----------GC----ACCC-----A--T-----A-----T------G-   6101
O.CM.MVP5180        ----T-----------CCT-ACAAGC--T--ACAG------A-------AT----A-----C--T--T-----T-----T-----T---T-TCC-C-A-GC-----------T--C---G-T--A-----   6552
O.BE.ANT70          ----T-----------CCT-ACAAGC--T--AAAG------A-------AT----A--------T--T----------CT---T-T---TATCC----CAC-----------TG-C-----T--A-----   6561
CPZ.US.CPZUS        ----C--T-----A--G-----CA-GCAG----CC--C--CA-------------G--A-----C--T--T--------T------------CAT---CC---------T-----G---G--------G-   6585
CPZ.GA.CPZGAB       ----C-----C-----G---C--AG--------CT------A-------------G--A--------T-----T---GT--T---G------T-TC-TCC------A-T----TCA--------------   6578
CPZ.CM.CAM3         ----C-----G--GTC---TCT-A-C-A-----CC------A--------T----A--------T--T--------T--------------GCAGA-CCC------A--T-----C-----T--------   5933
CPZ.CM.CAM5         ----C-----G--CTC----CT-A-C-A---A-CC--C---A--------T----A--------C--TGT---T--T-----C-----G--GTATA--CC------AT-T----GC--------------   6218
CPZ.CD.CPZANT       ----CA--A----CTCCATGACAAG---------G--C-----A-----A--TACCAGT-----G--A-T---T--AG-T--TATT...-----TAG-C-C---ACCT---TCTGG-----TGCT-AT--   5937

H2A.GW.ALI Env      C__A__.__.__.__.__.__.__.__T__K__N__R__D__T__W__G__T__I__Q__C__L__P__D__N__D__D__Y__Q__E__.__I__A__L__N__V__T__E__A__F__D__A__W__D  gp120
H2A.GW.ALI          -----.....................--CA-AAATAGAG--AC-----GA--CATACAG--CT----AGACA-TGATG-TTAT--G---...A--GCTT-G-----------GGC---CG-TGCA---G-   6929

H2A.SN.HIV2ST       -----.....................--TA-AAATAGAG--AC-----GA--CATACAG--CT-G--AGACA-TGATG--TAT--G---...A--GCTT-------------GGCC--CG--GCA-----   6381
H2A.DE.BEN          -----.....................--TA-AAATAGAG-CAC-----GG--CATACAG--CT-G--AGACA-TGATG-TTAT--G---...A--A-TT-------------GGC----G-TGCA-----   6939
H2B.GH.D205         -----.....................--CACAAACAGAG-CACC----GA--TGTACAG---C-C--AGACA-TGGTG--TACACT--G...A-CAG-C----CA-A-----GGC----G-TGCA---G-   6906
H2B.CI.EHO          ----T.....................--CAGAAACAGAG--ACC----GA--TGTACAA--CC-C--AGATA-TGATG--TATACC---...A-CCAAT-----A-A-----GGC----G-TGCA---G-   6904
H2G.CI.ABT96        -----.....................--CA-AAATAGGG-CACA----G---TACACAG--CC-T--TGATA-TGATG-TTACAGT---...C--GCAT-G-----A------GC----G-TGCC---G-   6252
MAC.US.MM251        -----.....................--CA--AATAGGG--AC-----GA---AC-CAG--CC----AGATA-TGGTG-TTATTC----...T-GGCCC-T-----T------GC----G-TGCT---G-   6815
SMM.US.SIVSMMH9     -----.....................--CAGAAATAGGG--AC-----GA--CACACAA--CT-G--AGATA-TGGTG-TTACTC----...T-GGCAA-T-----C-----GGC----G-TGCT---G-   6307
STM.US.STM          -----.....................--CA--AATAGGG--AC-----GA---AC-CAA---C----AGACA-TGGTG-TTACTC----...T-GGCAA-------C-----GGC----G-TGCC---G-   6481
TANTALUS.UG.TAN1    --ATG.....................---CCTAATACC--CC-------A---AC-AAT---A----AGAT--T-ATG-TTATACT---...---CA-T-------CT-T--G--A---G-AGCA-----   6508
GRIVET.ET.GRI-677   -CATG.....................--GCCCAATACC---A-G-----A--CACCAA---CA----AGAT--T-ATG-TAATAC---G...--GCCTC----CA-T------GC---CG-GGCT---..   6259
VERVET.KE.VER-9063  --ATG.....................--TCCTACTACC-GGC-A-----T---AC-AAT--CA----AGAT--T-ATG-TTATAC---G...--GCCAT----CA-T--T--GCCC---G-AGCC---GC   6512
VERVET.KE.AGMTYO    -CATG.....................--TCCCACCACA-GCT-A-----T--TAC-AAT--CA----AGAT--T-ATG--TATAC----...---CCTC-----A-C--T---CCA---G-GGCA---GG   5999
VERVET.KE.AGM155    -CATG.....................---CCTACTAC--GGT-G-----A--TAC-AAT--CA----AGAT--T-ATG--TATAC----...---CCAC-G---A-A-----GCCA---G-AGCA---GC   6510
VERVET.DE.AGM3      -CATG.....................---CCCACCACC-GAC-A-----A--TAC-AA--CGA----AGAT--T-ATG--TACAC---G...---CCAT----CA-C--T---CCA---G-AGCA---GC   6010
SABAEUS.SN.SAB1C    --AAG.....................--CCCCAATACG--CC-A------T-TAC-AAT--CA-C---GAT--TGAACCAGA-GG-AC-...A--GCA-A-GTACCC-T-CCT---A-C-CAGAAAAATT   6707
SUN.GA.SIVSUN       ----T.....................T---CTAATA---GCT-G----TG--TAC-T-A--CT-G--AGATTTA-AA-CTTATGC----...--GCCCA---CA-G-TT----G-GAA-TTT-CAGA-GG   6645
MANDRIL.GA.MNDGB1   ----T.....................-----TAATTCA-G-C-C----TA--CAC-AAT--CA----TT--TTG--AG-TTATG-T--G...---GAAA-TCC--AT-T-A-GG-AAA-TTT-CAG-ACT   5965
LHOEST.KE.hoest     ----T.....................T---CCAATAC--GCC-C----T---GACAT-T--CT-G--AGATTTG-AATCTTATGC---G...---CCTA-CT--AAT-TTTC-G-AAACTTT-CAATACC   6598
LHOEST.CD.hoest447  ----T.....................T---CTAATA---G-T-A----T----ACAT-A--CA----AGACTTG-AGTCTTATGC-C-G...A-TCCTA-TC-CAAC-T-AGTTTAAA-TT--CTAAAG-   5520
LHOEST.CD.hoest485  ----T.....................T---CTAATA---GCT-A----T----ACAT-A--CC----AGATTTA-AGTCTTATGC-C-G...A-TCCTA-TC-CAAC-T-AGTTTAAA-TT--CTCAAG-   5517
LHOEST.CD.hoest524  ----T.....................T---CTAATA---GCT-G----T---TACAT-T---C-C--AGATCTG-AATCTTATGCTC-G...A-CCCTA-TC--AAT-TTACTCTAAA-TTT-CTCAAG-   5520
RCM.NG.SIV-NGM      ----T.....................--T--TAACGCAC-ATCA----GA--GTTAAAT---A----AGA--GAGG--TAT---C----...---CCA---------T----G--A---G-TGCT---G-   7049
COL.__.SIV-CGU1     -CTTT.....................T-TTCCTCC-CAG-GA--CAACAGGTGTTG-GG---T----AC-TCCG--TGGAAAG-C--T--A-CA-AACA-GCCAAATGT-AC-G-AGC-TTTGATCTTTT   5870

SYKES.KE.SYK173     --A-C.....................--CTCCCATA-AGGA-GA-----T--TA-GAAT-----G--GT---CA......GAT---AT--A----AG---C--CA-T----G-G-A-ACTTTCCAGCATG   6457
SYKES.KE.SYK173 Env C__T__.__.__.__.__.__.__.__T__S__H__K__G__G__W__A__T__K__N__C__V__P__S__A__.__.__D__Q__I__E__V__R__V__N__I__T__G__E__Y__F__P__A__W  gp120
```

```
B.FR.HXB2 Env       __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__S__L__K_  gp120
B.FR.HXB2           AAATGAC..................ATGGTAGAACAGATGCATGAGGATATAATCAGT............TTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCACTCTGTGTTAGTTTAAAG   6614

A.UG.U455           ----A--..................--------T-----------A------------............C-------C-----------------------G--------T-----C----CG---G-T   6059
C.ET.ETH2220        -------..................-----G--G---------C--------------............------------G-------------------G--G----------------C------C   6023
D.CD.84ZR085        ---CA--..................-----G--C--------------------T---............--------------T-----A----------------------G------C-C------T   6135
F1.BE.VI850         ----A--..................-----------A------ACA------------............--------------------------------G-----------------G-C------C   5960
G.SE.SE6165         ----A--..................-----G-----A---------------------............---------G----------------------GC-------T----------CC-----C   6017
H.CF.90CF056        ----A--..................-----G--G---------ACA--C---------............---------------T-G--A-------------------------------C-C----C   5954
J.SE.SE7887         -------..................--------C--------G--A------------............G--------G-----T----------------GA-------T--A-----C-C------T   5933
K.CM.MP535C         ----A--..................--------G---------ACA------------............---------G----------A--T------G-G--------T--------C-C------T   5813
01_AE.TH.CM240      ----A--..................--------G--------G------G--------............--------------T-----------------G-----T--T-----C----C------T   6182
02_AG.NG.IBNG       ----A--..................--------------------A--------T---............C-------C-----------------------G--------T-----C----C----G-T   6141
03_AB.RU.KAL153-2   ----A--..................---------------------------------............--------------------------------G-------------------C------T   6646
04_cpx.CY.94CY032-3 -------..................---------------------------------............------A--G--G-------A-------C---GC-----T-T----------C---T-CA   5984
N.CM.YBF30          ----A-A..................----C---C--A-----A--------T------............C-G-----A--G---T-------C-----T------------T-A-----A-C-A-GCTT   6201
O.CM.MVP5180        ----T--..................-----G--C--A--------A--C--C--T---............--G-----A--G---TT------T----AG---A-G--TTTCT-A-----ACAAA-G--C   6652
O.BE.ANT70          ----T--..................-----------A-----G--A--C--T--T---............--------C--G--T--T--A--T-----TC--A-G--TTTC--G-----ACAAA-GG--   6661
CPZ.US.CPZUS        ----A-T..................----C------A-----G--A-----T--T--C............C-------C--G---T----A--T---A----G---------T-G-------C-A-G-CT   6685
CPZ.GA.CPZGAB       ----A-T..................-----G--C--A--------A--C--T------............C-C--------G--T--C-----T--------G--G-----TT-A-----A-CC--GC--   6678
CPZ.CM.CAM3         C---AC-..................----C------A---A--ACT--C--T--T---............C-G-----------TT-G--A--------T--GC-G--G---A-A-----C-CCA-GG-A   6033
CPZ.CM.CAM5         ----A-T..................----C------A-----G--A-----T-GT---............C-G-----------T-----A--------T---C-G--A---A-A-----C-CCA-GC-A   6318
CPZ.CD.CPZANT       ----T-T..................---------AGT---ACA--A-----GN-ACAA............----TCC-A-------AT---------------C----A--TA-G---A-A-AAA-G--T   6037

H2A.GW.ALI Env      __.__.__.__.__.__.__N__T__V__T__E__Q__A__V__E__D__V__W__R__.__.__.__.__L__F__E__T__S__I__K__P__C__V__K__L__T__P__L__C__I__A__M__K_  gp120
H2A.GW.ALI          T..................AATACAG-AAC------AGCAGTA--A---G-CTGG--A............C---TT--GAC-TCAA----A--------C--G-----A--TT-A---A-AGCAA-G---   7029

H2A.SN.HIV2ST       T..................AATACAG-AAC------AGCAGTA--A---G-CTGG---............C---TT--GAC-TCAA----A-----C--C---C----A--CT-A-----AGCAA-GCGT   6481
H2A.DE.BEN          T..................AATACAG--AC------AGCAGTA--A---G-CTGGCA-............C---TT--GAC-TCAA----A--------C--GC----A--T--A-----GGCAA-G--T   7039
H2B.GH.D205         T..................AATACAG--AC-C-----GCAGTA--T---G-GTGG--A............C-C-TT--AACCTC-A----A--------C---C----------G-----GGCAA-G--C   7006
H2B.CI.EHO          T..................AATACAG--AC---T---GCAACAA-----G-GTGG---............C-C-TT--GACCTCAA----A--------T--------------G-----G-CAA-G---   7004
H2G.CI.ABT96        T..................AACACAG-AAC-------GCAGTA--A---G-TTGG-A-............C-G-TT--GAC-TCTAC---A--T-----T--------A--T--G--CA--GCCA-G-GA   6352
MAC.US.MM251        G..................AATACAG-CAC-------GCAATA-----CG--TGGCAA............C-C-TT--GACCTCAA-------T------------T-----T-A--CA---C-A-G-GA   6915
SMM.US.SIVSMMH9     T..................AATACAG-CAC------AGCAATA------G--TGG-AC............C-C-TT--AAC-TC-A-T-----C---------C-C--------A---A-A-CAA-G-G-   6407
STM.US.STM          C..................AATACAG-CAC------AGCAATA------G-CTGG-AC............C-T-TT--AAC-TC-A-T--A--T-----------G--T-----A---A--GCAA-G-G-   6581
TANTALUS.UG.TAN1    GG-CAGG............AATCCAT-A--G-C----GC-G-GAGTA-C---CATTTG............C-C-TT--AAGC-CT--G--------C------C-G--A--TA-G--CA-A-AGA-G--T   6614
GRIVET.ET.GRI-677   .G--A-T...............CCGC-----A----AGCAG-GAGTA-----CATCTA............C-C-TT--A----CGA-G-G---T-----T--GC-CT----CA-A---A---AAA-GTCC   6361
VERVET.KE.VER-9063  -G--AGG............AATCCAT-A--G-C----GCAGGAAGTA----TCATTTG............--G-TT--A----CAT-G--A-----C---------T-T--T--A---A-A-AAA-G-GC   6618
VERVET.KE.AGMTYO    TG--AGA............AACCCAT-AA---C---AGCAGCCAGTA-C--CCATTTA............C-C-TT--G----CTA-G--A--T-----G--G---T-A-----A--CA-C-AGA-G--T   6105
VERVET.KE.AGM155    -G-CAGA............AATCCCT-A----C---AGCAGGAAGTA-C--TCA-CTG............C-G-TT--A--G-CAT-------C--------GC--T-A--T--A---A-C-AAA-G--T   6616
VERVET.DE.AGM3      TG-CAGA............AACCCCT-A----C---AGCAGGAAGTA-----CA-CTG............C---TT--G--G-CT--G------------------T-A--TT-G--CA---AAA-GTCC   6116
SABAEUS.SN.SAB1C    TG---CTTGGAAAAATCGGAACCCTC-A----GC--AGCAG-GAGTA-C---CA-CT-............--G-TT--AAGT-CTT-------------T--GC-CT-A---A-G--CA-A-AAA-G--C   6825
SUN.GA.SIVSUN       -.....................ATA-GCAATA-T--A--AGTGC--C-AGC-TGGCAAGCTATGACCTCTA-GGTA---GCC-TTA-G-----T--------GA-C-AT--CTAT-----C-AAA-G---   6754
MANDRIL.GA.MNDGB1   T.....................ATA-G--A-A-T-----AGT-T-TC-AGC-TGGCA-GCTATGGGAAGTA-G-TA---ACC-TA--T--------------GA-T-A----TAT-------AGA-GC-A   6074
LHOEST.KE.hoest     -.....................GTA-AA-ACA-T---G-TATAC--C-AGCCTGGTC-GCTATGAATGCTA-GGTA--CAGT-TTA-G--A------------A-C-AT---TAT-----G--AA-GC--   6707
LHOEST.CD.hoest447  G.....................ATA-AA-ATA-T--A--TATAC--C-AGC-TGG---GCTATGACCAGTA-GGT----ACT-T-A-G-----T---------G---AT---TA------G--GA-G--A   5629
LHOEST.CD.hoest485  G.....................ATA-AA-ATA-T--A--CATAC--C-AGC-TGG---GCTATGACCAGTA-GGT----ACT-T-A-G-----T--------GA---AT---TA------G--AA-G--A   5626
LHOEST.CD.hoest524  -.....................ATT-A--ACA-T--A--AATAC--C-GGC-TGGG-AGCTATGACGAGCA-GGT----GCT-TTA-G---------AC----A---AT--TTAT------C-CA-G--A   5629
RCM.NG.SIV-NGM      -...............AATAGCTTGTAT-A-...--AGCAA-GTCTA--G-GTGGCA-CTT.............-AT---TCC-CTT-G--A----------GG-----A---A-G---A-A-CAA-GT-T   7149
COL.__.SIV-CGU1     T...........................AAGA-TAGCT-CAG--G---GG--TGG-TAATAACACAGACTACC-TA--A-----G----GA-------CT--G--G--AG--TA-----C-CCAA-G-TT   5973

SYKES.KE.SYK173     G............AATAGCAGTCAC---A--AGG---CAAATATTA--GGAC--GTCAGCTTTA.........-TCTTG---GC-AAT-GA--C-----G--GC--G-A--CA-G---A----GA-GTT-   6566
SYKES.KE.SYK173 Env __.__.__.__.__N__S__S__H__M__I__R__Q__Q__I__L__E__D__M__S__A__L__.__.__.__F__L__Q__A__N__R__P__C__V__K__L__A__P__M__C__I__R__M__L_  gp120
```

```
B.FR.HXB2 Env       _C__T__.__.__.__.__.__D__L__K__N__D__T__N__T__N__S__S__S__G__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._  gp120
B.FR.HXB2           TGCACT...............GATTTGAAGAATGATACTAATACCAATAGTAGTAGCGGGAGA...................................................................   6662

A.UG.U455           ---CA-AACATCACCATTAACA-CACC--T--CA-C--C---........................................................................................   6101
C.ET.ETH2220        --T-A-GCTATCAAAAACAATAC-AA-GTT-CAA---A--G-........................................................................................   6065
D.CD.84ZR085        ------GATGCATCGAGAAACAGCACTG-T---A-C-GC-CA........................................................................................   6177
F1.BE.VI850         --T---AATGCCACCAATAACAG-CA-G-A--GCCAGGGGCA........................................................................................   6002
G.SE.SE6165         --T---GATGTAACCAACAAA-GCAAC--A-GAA---A--GC--TG---A---C-CT-AA-CCAACAATAGCACTGTGGATAACCCA...........................................   6104
H.CF.90CF056        --T---AATGTCAGAAACAATACC-CT--C-GCAC--GC-G-........................................................................................   5996
J.SE.SE7887         --T-G-AACATCACTAGTAACAG-AAT-CA-C-AG--AC-G--GT.....................................................................................   5978
K.CM.MP535C         --T---GATTACAAAGGAACCA--AGC-CC---A--G-A-C--G--C-..................................................................................   5861
01_AE.TH.CM240      --T--CAATGCTAATTTGACCA--GGC-GT-GCA-A--C---GT-TC--AC...............................................................................   6233
02_AG.NG.IBNG       --TCA-AATTTCAACAACAGCT-CAGC--C-G-AGC-ACCTA--T-G-..................................................................................   6189
03_AB.RU.KAL153-2   ------GATTTGAAGAAGGAG-T-ACT-GT-CCA------G--G-.....................................................................................   6691
04_cpx.CY.94CY032-3 --T-T-AATGCAACTACTACTA--AGT-CC----GC---...........................................................................................   6023
N.CM.YBF30          --T-ACGATAGCTATGGGGAG--AAG---C---ACA-A-...........................................................................................   6240
O.CM.MVP5180        --TGTAGATCTGCAAACAAATA-AACAGGC....................................................................................................   6682
O.BE.ANT70          --T--AAAC.........................................................................................................................   6670
CPZ.US.CPZUS        --TCT-AACCCC.......................................GA---TAAT......................................................................   6706
CPZ.GA.CPZGAB       ----G-AAGGCTAACTTTAGC.....................CAGGCA-AA-ACCTAACA......................................................................   6717
CPZ.CM.CAM3         --T-GAAAGGTTACTTTCAATAGCACTTCT---AGG-A---A--TTC--C--TG-CTACT......................................................................   6093
CPZ.CM.CAM5         --T--GACCCTTAAGAACATTAGCACT-CT---AC--AC...-GT-CAGAGTTA-AAAAT......................................................................   6375
CPZ.CD.CPZANT       --T--AGGATACAATGGAACACC-ACA-CACCA.........-GT-CA-CA-CA--TACA......................................................................   6088

H2A.GW.ALI Env      _C__S__N__.__.__.__.__.__.__.__.__.__.__.__.__.__.__I__S__T__E__S__T__T__T__S__P__S__P__G__S__T__L__K__P__L__I__N__.__.__.__.__._  gp120
H2A.GW.ALI          --T-GCAAC..........................................-TA---ACAGAGAGCACAACCACATCCCCGAGCCCAGGGAGCACACTCAAACCCCTGATAAAT................   7101

H2A.SN.HIV2ST       --T-ACAGC..........................................-CA-CT-CA-A-AACACAACCTCCACACCAACAACCACCACAACAGCAAACACAACAATAGGA................   6553
H2A.DE.BEN          --T-GCAGG................................................-TTCA-GGGAATACCACGACCCCGAATCCCAGGACCTCGAGTTCCACAACCTCGAGACCACCCACATCCGCAG   7121
H2B.GH.D205         --T-G-AAAACCGAAACAAACCCAGG---TGCCAG--G--C-----CC-C--AGCCTACT-CCACCTCTCGTGGGCTGAAAACGATTAACGAAACAGACCCA............................   7108
H2B.CI.EHO          --T-A-AAGACGTGGAGCTCA-C-AGC--AG-GACC----CGT--TCCGCGTCCCT-A-ATCTTCTACTCAGACCCTACTCAACGAAGATAGCAAA..................................   7100
H2G.CI.ABT96        ----A-AAAACTGAGACAAACA-A-G-GGTCTAACAGGG-CC----C-GTA-CCCCT-CA-C-GTTCCCACAACTAAAATGGTGACAGCTGAACTAGTGAATAGCACTAGCCAG................   6466
MAC.US.MM251        ----A-AAAAGTGAGACAGATAGA-G-GGATTGACA-AATCAT-A-CA-CA-TA-CAACAGC-GCACCAACATCAGCACCAGTATCAGAAAAAATAGACATGGTCAATGAGACTAGTTCT..........   7035
SMM.US.SIVSMMH9     --T-A-AAAAGTGAGMCAGMCAGA-G-GGTTTAACAGGAGSAC-AGCACCA-CA-CAACACA-ACATCAACAACACCACCATCACCAATAATAGCAAAGGTTGTAAATGACAGTGATCCT..........   6527
STM.US.STM          --T-A-AAGAATGAAACAGACA-A-G-GGATTGACAGGG--A--AGTA-CA-CAGTGACACC-ACAGCAGCAGCAGCAGCAACAAAGCCAGAGTTAGTAAATGAA......ACAAGCTCC..........   6695
TANTALUS.UG.TAN1    --T---AAATTAACAAGTACC-CCCCT-CATCAAG---CCC---ATCA--C--C-CTAC-.................................GATCCCTGTCCAAATACCGACGAAAGTAGCTGTAACG   6711
GRIVET.ET.GRI-677   --TGTAGAGCTGAATGGTACA-CCAC--CA--G-CC--C-C---TGCA-C--CA-CAAT--CT..............................ACCCCCTGT...........................C   6434
VERVET.KE.VER-9063  --TGTAGAGTTGAACTCAAGTAGAAGA--CCCAAC-C-A-C-T-AGC-TCA-CA-CAAAAGCT...CCAAAAACAGGT...............GATCCCTGCATTAAATCTACCAATAATAATGTTAATT   6730
VERVET.KE.AGMTYO    --TGTAGAGTTAAATTCCACAAGAGAA-G-GCGACA--ACC---A-CG-CGCCG-AATCT-CCGGC...........................CTACCCTGTGTAGGCCCGACGTCAGGTGAA...AATC   6205
VERVET.KE.AGM155    --TGTAGAGTTAAAAGGCTCC-CAACCTCT-CCCCAG-A-CCT-T-C--CGGCAG-AACC-A-..............................CTACCCTGTGTTAGAAATAAAACAGACTCC...AACC   6713
VERVET.DE.AGM3      --TGTAGAATTGAACTCCTCT--GCCT-CC-CCAC-C----A-GT-CC-CGGCCTCAACA-CCAATATCACAGCCTCAACAACCACT......TTGCCGTGTGTCCAGAACAAGACAAGTACTGTG...T   6237
SABAEUS.SN.SAB1C    ---TA-CGCTTAGAAGGAGGA-CAGCA-CA-CAACAT-ACCAT-A-CATCA-CAGCAC-CCCCGAGGTAGTGAGTGTGGGCTTTAATGACTCAGTGATAGAACAAGAGATGGAAAAA.............   6942
SUN.GA.SIVSUN       --TC-ACCGAAACCGACAACTCCC-CCTCC--CTCA--AGTC-AA-G---CT--GATTATT-GACGACCACCACGGCTAAGACTACAACTCAGACTACAAGTAGTACCAGCTCCACTGCTAGTACAACTA   6884
MANDRIL.GA.MNDGB1   --TCAGGAAACAGAAAATGTATCAGCA-CA-CA-C--AGCC--TA-C--CACC--CTACT-C-TCTACAGTTGCAAGTAGTACAGAGATTTACTTAGATGTAGATAAAAAT...................   6185
LHOEST.KE.hoest     --TGGGGAAGTTACAAAAACACCAACA-CA-CACCA-AA..........................................ACTACTACACAGATGCCTTGTTTTATCAACGAACAGGTAACAGTTAAGA   6795
LHOEST.CD.hoest447  --TGAAGGGGAAGGAATAGGAACAACA--AGCAACA-AA-CA-AA-CG.................................TCAACAACACCCATGCCTTGCTTTATAACAGAAAAAACCACAACCACCA   5726
LHOEST.CD.hoest485  --T--AGGGGGAGGAATAGGACCAAAA-CA-CACCAT-ACCA-TAGGA-AA..............................ACAACAACACCCATGCCTTGCTTTATAGCAGAAACAACCACAACCACCA   5726
LHOEST.CD.hoest524  --T--AGGGGAAGGACTAGGTA-GAA--GA-CAACA--A-CA--A-CACCA..............................ACAACAACACCCATGCCTTGCTTTATAACAGAATCTACTACTACTACTA   5729
RCM.NG.SIV-NGM      --T---GCATTGAATGGTAGTTGGGATGGT-C--CC--CC-AT-ACC--CA-CA-CAACA---GCAACAACAAGAACAACAAAAGAAATAGGGAAAGATTGTTATGCGAATCAAACAGAGTGCAGTAGTG   7279
COL.__.SIV-CGU1     -----CAAAGTAAATAGGACT--AAATGGC-C-AG---AGTAG--CCA-CA-CA-CAAAT-ATAGTGCTTCAGATTGGGATGAGTCA...........................................   6060

SYKES.KE.SYK173     --T--ATTGGACAACAGCCCT-C-ACA-GT-CCCC-----CA-GTCCCCC--C--CACCTCC-AATGAAACGTGGTGGGGAGAC..............................................   6650
SYKES.KE.SYK173 Env _C__T__L__D__N__S__P__A__T__S__T__P__T__T__S__P__P__T__T__P__P__N__E__T__W__W__G__D__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._  gp120
```

```
B.FR.HXB2 Env       .__.__.__.__.__.__.__.__.__.__M__.__.__.__.__.__I__M__E__K__G__E__I__K__N__C__S__.__.__.__F__N__I__S__T__S__I__R__G  gp120
B.FR.HXB2           .............................ATG...............ATAATGGAGAAAGGAGAGATAAAAAACTGCTCT.........TTCAATATCAGCACAAGCATAAGAGG   6724

A.UG.U455           ..........................ATC-CA...............GATGGA-T--GG-A---A--G-----T------.........--------G-C----GAAC------A   6166
C.ET.ETH2220        .............................--C...............-AT-GT-CC--T-AT--A--G-----T--T---.........--------A-C----GAAC------A   6127
D.CD.84ZR085        ..........................CTGCCA...............-CGG--A--CC------A--G------------.........--------A-C----GTAG---CG-A   6242
F1.BE.VI850         .................................................................--GC-------T---.........--------G-C----GAAG------A   6043
G.SE.SE6165         ...........................................................-----A---------------.........------G-A-C----GAA-------A   6151
H.CF.90CF056        ...............................................--GGA--CAGG---G--AC---C---T------.........------G-A-CT---GTAC------A   6055
J.SE.SE7887         ...............................................G---GTAGTCCT-ATATC--G-C----------.........--------A-CT---GAA------AA   6037
K.CM.MP535C         ...............................................G-GG-AAGCCC--C---A--------T------.........--------A-C----GAA----A--A   5920
01_AE.TH.CM240      ..........................ATA--A...............GG--ATATA-C--AT--AG---G------TA--.........--T-----G-C----GAAC---C--A   6298
02_AG.NG.IBNG       ..................................................GACAT---T--G--A---------------.........--------A-C----GAAG------A   6245
03_AB.RU.KAL153-2   ..................................................--AA-A-TGATG--A--G------------.........-----C----C----GA-C-G----A   6747
04_cpx.CY.94CY032-3 ..................................................G--ATT----A--GA---------------.........---G----A-C----GAA-----G-A   6079
N.CM.YBF30          ..............ATGACAACAAGAGAACCA...............GAC--A-GAT-CAA-C-A--G-----T-----A.........------GCA-C---TGAGC---C--A   6317
O.CM.MVP5180        ..........................CTAT-A...............-ATGA-ACA-T-AAT-----G-G---T--TAG-.........--T---G-A-CT---GT-C-C-C--A   6747
O.BE.ANT70          .......................ATAGCTGGA...............-C--CAA-TG--AACCTT--G--G--G--TGAG.........--T---G-A-C---TGTT--C-A--A   6738
CPZ.US.CPZUS        ....................AGTAGTGCTG-AAATACTACTGATATA..................--G-G------T---.........--T-----A-CT--TGAAT------A   6774
CPZ.GA.CPZGAB       ....................AACCAGACATCTTCTCCGCCTCTCGAA..................--G--------T--A.........--T---G-G-CA---GAAT------A   6785
CPZ.CM.CAM3         ....................AACAGTCCT-AT...............GA--AAAYAG-CA-CACAG----G--T--TA-C.........--T-----G-CT--TGAGG------A   6164
CPZ.CM.CAM5         .....................................................ATA----A-ATA--G-----T--TA-A.........--T-----G-CT--TGAAG------A   6428
CPZ.CD.CPZANT       ....................GTAACACCA-A-ACAACAACACCAATA...GTAGATGGCATGAAGC--C--G-A--TAAC.........--T---CAG------G-AT-T-A--A   6171

H2A.GW.ALI Env      .__.__.__.__.__.__.__.__.__.__E__.__.__.__.__.__S__D__P__C__I__K__A__D__N__C__P__.__.__R__G__L__G__D__E__E__M__V__N  gp120
H2A.GW.ALI          .............................GA-...............-GCGATCCATGCAT-A--GC-G-C------C-C......AGGGGACTAGGGGATGA-GAG--GGTCAA   7166

H2A.SN.HIV2ST       .............................GA-...............-ATTCTTCATGCAT-CGC-C-G-C------A-A.........GGGTTGGGAGAAGA-GAG--GGTC-A   6615
H2A.DE.BEN          CCTCCATA...............ATAAATGAA...............-CTTCTA-CTGCAT---A-AC--C-CA---G-A.........GGATTAGGGTATGAGGAG--G-TGCA   7197
H2B.GH.D205         ........................................................TGCAT-A-A-ATG-C-G----A-A.........GGACTAGGAGAAGAGGAA----TGCA   7158
H2B.CI.EHO          ........................................................TGTATTC-A-ATG-C-G---TG-A.........GGG-TAGGACTAGA-GAA--G-T--A   7150
H2G.CI.ABT96        ........................................................TGCTT-AT-TACG-T--T--TA-A.........GGT-TACAGTCAGA-TC---GGTT--   6516
MAC.US.MM251        ........................................................TGTAT--CTCAG--T--T---A-A.........GG-TTGGAACAAGAGCAA--G-T-A-   7085
SMM.US.SIVSMMH9     ........................................................TGTAT-AGA-GT--T--T--TA-A.........GG-TTGGAACAGGAGCC---GGT-A-   6577
STM.US.STM          ........................................................TGT-T-AGT-AC--C------A-A.........GG-TTGGAAGAAGA-TCAC--GT---   6745
TANTALUS.UG.TAN1    CCACC...............TTAGTTACA-AT...............-GC-----TT-T-AGA-T-GTTCT-TA-----C.........--TGC---GGCAGG-TAT-GG----A   6787
GRIVET.ET.GRI-677   AGAATTGCAGTACAGAGCAG...ATAGAAGGA...............GA-----CAG-G-A-CCAGC-TCC------A--.........--TGCA--TGCAGG-TATCA-----A   6522
VERVET.KE.VER-9063  TACAGCCCTGTAATGCCTCCCTTATAGAAGA-...............GA-C-----G---A--CAGCTTC---T--TA-A.........--TGC---GGCAGGGTA---T----A   6821
VERVET.KE.AGMTYO    TACAGTCCTGTAATGCAAGCATTATAGAA-G-...............GAG------G-T-AGCCCGCCTCT--T--TA-A.........---GCA--GGCTGGCTATG------A   6296
VERVET.KE.AGM155    TACAGTCATGCAACGACACCATCATAGAA-A-...............GAG---A-TG-C-AG-CAGCGTC-------A-C.........--TGC---GGCTGGGTA---T--G-A   6804
VERVET.DE.AGM3      TAGAATCATGTAATGAAACAATCATAGAA-A-...............GA-T-AA-TG---AGCCTGCTTCT--T--TA-A.........--TGCA--GGCAGGGTATG------A   6328
SABAEUS.SN.SAB1C    ...........................................................-A-C--GCC-TG-----T--C.........--TGC---GGCAGG-TAT-G---G-A   6989
SUN.GA.SIVSUN       CTCCAATGCCTCTAGAT...................................................TGG--T--TA--...GACACAGAG--C--AGCGGAG----AT-A--T   6945
MANDRIL.GA.MNDGB1   ...................................................................................AATACAGAAG-A-AGGTAGAG--G-ATCAT-T   6217
LHOEST.KE.hoest     ACCCGGGGAATGAAACAAGA....................................TT--AG--AGATCTT--T---A-AAGAGGGCTTAATG-G-C--CAGAG--G-ATGC--A   6874
LHOEST.CD.hoest447  CCAAAAGATCTACAACTGGGGGACCAGAA...........................TT-AT----GAC-T---T--TAT-AAATGGGTAAATG-G-CT-CTGAG-TT-AT----A   5814
LHOEST.CD.hoest485  CCAAAGGGAATACAACTATGGAAAGT..............................TT-AT----GAC-C---T--TA--AGATGGTTAAATG-G-CT-CTGAG-TT-AT----A   5811
LHOEST.CD.hoest524  CAAAGAGTTCAGGGACTGAACAACCATCT...........................TT-AT---AGAT-C---T--TA-ACGATGGTTAAATG-G-CT-CAGA--TT-AT--G-A   5817
RCM.NG.SIV-NGM      TCCCT...................................................G-C-C---TG-C-TG--T--TGAA.........--T-CGG-GGCAGGGTTA-A-----A   7334
COL.__.SIV-CGU1     ...............................................-ATTG-A-AG--TATCC-TGGT-T--T---AGA.........A-G---TCA-CAG-TTT-T--CTTAA   6119

SYKES.KE.SYK173     ............................................................................................--C-GT-CAGAGCCA-G-TTCAA   6673
SYKES.KE.SYK173 Env .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__N__S__T__E__P__R__F__N  gp120
```

```
B.FR.HXB2 Env       __K__V__Q__K__E__Y__A__F__F__Y__K__L__D__I__I__P__I__D__N__.__.__D__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__S__Y_  gp120
B.FR.HXB2           TAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGATATAATACCAATAGATAAT......GATACT................................................ACCAGCTAT   6797

A.UG.U455           ----AAA-------T----T--C---------G----------G---A---CA----A......AC-GA-AAC.............................................-AT--T---   6242
C.ET.ETH2220        ----AAAAG---G-C-------C--------------------G-----C-TA----C......-G-T--................................................--TGAT---   6200
D.CD.84ZR085        ----AG-A--C---T-C-----C---------G-------G--G---A----------......--GGG-AAAAATGAAATCAAT........................GATACTTATGGT-CT---   6339
F1.BE.VI850         ----AA--T----TT--C---GC---------GG---------G---------GC---......A-C-A-AGC.............................................-GTGAA---   6119
G.SE.SE6165         ----AAAA----------C--G-----C----G-------G--G-------TA-----......-CA...................................................--T-AT---   6221
H.CF.90CF056        ----CA--------T-C-----C-C-------G-------G--G-------T-----C......A---G-................................................--TCAG---   6128
J.SE.SE7887         ---AAG-A--C----G--C--GC----C----G--AA---G--G-------T----G-......A-C-A-................................................-AA-AT---   6110
K.CM.MP535C         ----AA-A-A--G----C----C----C----G-------G--C-G---C-TA--GGC......--GGG-AATAACAGTAGT....................................--TGAA--C   6005
01_AE.TH.CM240      ----AA------G-TCC-----C-C--------G---------G---A---T--AG--......A-G-AGACTAGT..........................................-GTGAG---   6377
02_AG.NG.IBNG       ----AAAA----GATGC-----C---------G-------G--G---A---TA--G-A......A---A-GGT.............................................-GTCAG---   6321
03_AB.RU.KAL153-2   ---A---A-A------------C-C---------------G--G---A----------......---......................................................------   6814
04_cpx.CY.94CY032-3 ----AA-A-------------GC----C----G-A--------G-G-----TA--GC-......AGAGTGCCAATTAATGGTAGTAATAGGAATAAT...............AGTACAGAAGAG---   6185
N.CM.YBF30          ---AAA-A--C---TT--CT-TC-G------GT-GAA---G--G-------CA--GCC......................................................TATAAT-AA-CA---   6390
O.CM.MVP5180        C--AAA-G--C--A--C-G--TC-A--C---GT-TCA---C-G-GTAAGG-TA--G-C......TCA-A-GCAGTAAAT....................................GGA--A-CA---   6832
O.BE.ANT70          C--AAA-G-----A--C-G--TC-A--C---GT-TCA---T-G--GGA-C-GA--G-G......ACA-GCAGCACAAATAAGACAAAC...........................AGC-AA-TG---   6832
CPZ.US.CPZUS        C--AAA-A-AC---TG---T-C--A------GT-GA----C--GCT-AT--CA-----AAT............................................................-C----   6841
CPZ.GA.CPZGAB       ----AA-A-AC-G-TC---T--C-A------GTGGAG---G-GG--AACC---GG---......--G-ACAAC................................................-CA---   6858
CPZ.CM.CAM3         ---A-A-A-A----T----T-CC--------GT-GA----C--G--TTG--T-----C......-----CGAT................................................-CA---   6237
CPZ.CM.CAM5         ---AAA-A--G---T----T-TC-A------GT-GA-------G----T--T-GAGG-AATAAGAG--A-GAT................................................-C----   6507
CPZ.CD.CPZANT       ----AAA--A---ATGA-A--CA-A---------GGA---C-T--GAAGTGTC-GG-CAACAAT--G---AAC................................................T-----   6250

H2A.GW.ALI Env      __C__R__F__N__M__T__G__L__Q__R__D__K__P__K__Q__Y__N__E__T__.__.__W__Y__S__K__D__V__V__C__E__P__F__N__T__T__T__N__.__Q__T__R__C_  gp120
H2A.GW.ALI          -TGTCG-TTC--CATGACA-G---ACAGAGAG-TAAGCCA-A-CAGTAT-AT--A-CA......TGGTACTCAAAAGATGTGGTTTGTGAACCATTTAACACCACCACAAAC...CAG-----G-G-   7284

H2A.SN.HIV2ST       CTGTCA-TTC--TATGACA-G---AGAGAGGG-TAAGA-A-A-C--TAT-AT--A-CA......TGGTACTCAAAAGATGTAGTCTGTGAATCAAATGACACCAAGAAAGAG......-AA-CA-G-   6730
H2A.DE.BEN          ATGT-A-TTC--TATGA-G-GG--AGAAC-AG-TAAGA-A-GG-GGTAT-AG--C-CA......TGGTA-TTAGAAGATGTGGTTTGTGACAACACAACA...............GCTGG--CA-G-   7303
H2B.GH.D205         ATGTAATTTT-GTATGACG-G-C-AAGAAGAG-TGAGCTA-A-CA-TAT-A---C-CC......TGGTACTCAGAAGATTTAGAGTGTAATAATACCAGGAAGTATACC.........-G---A-GC   7270
H2B.CI.EHO          CTGTCAATTC---ATGACA-G---AAAAAGAG-TGAGTCA-A-CA-TAT-AG--C-CC......TGGTA-AAACAAGATTTAGTGTGTGAAAAGGGGACAAGGAGTAAT......GAA-G--AA-G-   7265
H2G.CI.ABT96        CTGCAAATTT--CATGACA-GGC-AAAAAGGG-T-AGAGA-AGGA-TAC-AT--G-C-......TGGTA-TCACAAGATTTGCAATGYGAACAGTCCAATAAGAGTGAGAAT...GAG-GT--A-GC   6634
MAC.US.MM251        CTGTAAATTC-CCATGACA-GG--AAAAAGAG-CAAGACA-AGGAGTAC-AT--A-C-......TGGTACTCTACAGATTTGGTTTGTGAACAAGGGAAT...AGCACTGATAATGAA-G---A-GC   7203
SMM.US.SIVSMMH9     -TGTAAATTT--CATGACA-GG--AAAAAGAG-CAAAA-G-G-GA-TAT-AT--A-CA......TGGTA-TCAAGAGATTTAGTGTGTGAACAAAATAGCAATGGAAATGAAACTGAC-GT-AA-GC   6698
STM.US.STM          -TGTAAATTC--TATGACA-G-C-AAAAAGAG-TAAAA-G-G-GAGTAC-AT--A-CG......TGGTACTCTAGCGATCTAATCTGTGAACAAAAT......GTCACTGGCGAAGAG-G---A-GC   6860
TANTALUS.UG.TAN1    -GTAAAAA-----T-TA--AGTAC--GG---G-T-AG--GT-GG--TGTGAGA-GG-A......A-C-ACACCACA....................................GGCACG-GAG-T-G-   6872
GRIVET.ET.GRI-677   -GTAAAAA----TT-T...AGCA-GACC-GGT-TGA-C-GGAGT--GTCTGCA-----......A-A--AGGAAGTGAAAAG..............................GGAAGT-AGGAT-G-   6610
VERVET.KE.VER-9063  -C--AA-A-A--TT-T...T--G-AG-A-GG--TGA--CAGAG--CTATTGTA-G--C......AGCT--AGCTCTAACTCT..............................AGCACA-AAGAA-GC   6909
VERVET.KE.AGMTYO    -C--AA-A-A--TT-T---T-TG-GG-G-GG--TGA--CAGA---CTATTGCA-A---......A-G--.......AATAGC..............................ACTAGC-AAGAG-G-   6381
VERVET.KE.AGM155    CC-AAA-A----TT-C...T--G-AG-A-GG--TGA--CAGA---CTTTTGTA-GCG-......AG---A...TCGCATAAT..............................GGGACA-AAGAG-GC   6889
VERVET.DE.AGM3      -C--AAAA----GT-T...T--G-GG-G-GG--TGA--CAGA---CATGTGTA-G--G......-G-......AACAAT.................................TCTAAC-GAGAA-G-   6407
SABAEUS.SN.SAB1C    -GTAAA-A-A--TT-C...T--ACAG-A-GGG-TGA-C-AGAGG-GGTTTGT--AG-G......-GA-GGGAAAAATCAAAC............GCAACA............CATACTGTGG-A-G-   7083
SUN.GA.SIVSUN       ATGTAAAT-T--T-TGACA-GGC---G-AGAG-TTG-A-A-CGGAGGTGGA-C-A--C......TT--GGGATACAGAAGTTACTTGCAATGGTAATGAC.....................-CA-G-   7045
MANDRIL.GA.MNDGB1   ATGTAG-T-T--CAT-ACA-G-C-A-GCAGGG-TTCGA-GGA-GA-AT-G--ACA---......TT--GAGGGGATGATGTGAAATGTGAAAATAAT........................-CT-GC   6314
LHOEST.KE.hoest     ATGCCA-T-T--T-T-ACA-GG--A-GCAGAG-CTGCAGG-CGGA-AT--A-C-A-GC......TT--GATATGATGATGTAACGTGCTCAGGAGAGAGGGAGAACAGA............-C--G-   6983
LHOEST.CD.hoest447  GTGTAGAT-T--T-TTACA-G-C-A-G-AGGG-CTG-AGG-C-GA-AT--AGC-G-G-......TT--GGTATGATGAGACAACATGTGATAAGAATGGA.....................-CT-G-   5914
LHOEST.CD.hoest485  ATGTAAAT-T--T-T-ACA-G-C-A-G-AGGG-CTG-A-AGA-GA-AT--AGC-G-G-......TT--GGTATGATGAGACAACCTGTGATAAGAATGGA.....................-C--G-   5911
LHOEST.CD.hoest524  ATGTTCAT-T--T-TGACA-GGC---G-AGAG-TTGCA-GGA-GAGAT--AGC-A---......TT--GGTATGATGAGACAACATGTAATGAGAATGGT.....................-CT-GC   5917
RCM.NG.SIV-NGM      -G-AAAAAT-----T-A---ACACA...-GGT-CTCAAGAGATC-CTGGTGTA-GGCA......-G--A-AATAATACTAGGCAATGT.......................................   7413
COL.__.SIV-CGU1     AG-TAGAA--G-GTT-G-ATT-GGA----CAGT-GAA---T---CTGTGC-T-GA---......A-A-A-GACAGTAATAGT....................................-T---GGCA   6204

SYKES.KE.SYK173     -TGCTCTTTT--CTTGACA-GGGGG--CA-GG-TAAAA-ACAGCA-TAT-G--C-TTC......TTCTA-AAGGATGACTTGATGAAAGAGGAAGGCAATAGTAGC...............TAT---   6779
SYKES.KE.SYK173 Env __C__S__F__N__L__T__G__G__F__K__D__K__K__Q__Q__Y__R__A__F__.__.__F__Y__K__D__D__L__M__K__E__E__G__N__S__S__.__.__.__.__.__Y__Y_  gp120
```

```
B.FR.HXB2 Env       _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__K__T__F__N__  gp120
B.FR.HXB2           AAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTAATAATAAGACGTTCAATG   6927

A.UG.U455           -GAC-A-T--A------T------AC------------T-----C-----G-----------------------------------A--------------------G-----GG--CCTGA--------   6372
C.ET.ETH2220        -GA--A-T--A------T------AC---A-----A--------------C--T--G--C--------T--------------T--A-------A------------G----GAG-------A----C--   6330
D.CD.84ZR085        -G---A-T--A------T-------C---------A--T--------A--------------------------------------A-----------A--------G-----CG----A-G--------   6469
F1.BE.VI850         -G-C-A-T--A------T------AC------------T--------------T-GG--T-----C--T--------------T--A-------A---A---------------G----A-GA-------   6249
G.SE.SE6165         -G-C-A---CA------TGT----AC---C-A------T------------A-T-----C-----------------------T--AA--------------T----G----GGG-----GA--------   6351
H.CF.90CF056        -G-C-A-T--A------T--------------------T--C--------G--------A--T--------------------T--A-----C--------------G-----C-----A--A-------   6258
J.SE.SE7887         -TTC-A-T--AC-----T-------------A---A--T-----------G--T---C-A--------------------------A---------------------------G------AT-------   6240
K.CM.MP535C         -G-C-A-T--A------T------AC----------A-T--------A---A----------------------------------A-----C--------------G-----GG----A-G--------   6135
01_AE.TH.CM240      -G---A-T--A------T--T----------AG-----T---------A----------T--------T------------A-T--A-------A-------T----G------G------AT-------   6507
02_AG.NG.IBNG       -G---A-T--A------T-------C------------T-----------------------------------------------A--------------------G-----GG----AGG--------   6451
03_AB.RU.KAL153-2   -G-----T---------T---------G-------A------------A-------------------------------------A--------------------G-----CG----A-A------C-   6944
04_cpx.CY.94CY032-3 -T---A-T--A-------G-----AC-----A------T--C--------G--T--------------------------------A-----------A---T----G------G-G--A-AT----C--   6315
N.CM.YBF30          -G-C-A-T--A------T---A---CTG-G-----A--T-----T---ACT-----------------A--------C-----A--AC-A--C-----C---A-G---------G-AGGA-AC--T-G--   6520
O.CM.MVP5180        -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-CAGAC--T----   6962
O.BE.ANT70          -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-CAGAA--T----   6962
CPZ.US.CPZUS        -GAC---T--A---C------A-C-CT--C--------T-----T---ACC-----------------A-----C--------A--AC-A--C-----C--C------------G-A--AGAT-----A-   6971
CPZ.GA.CPZGAB       -G-A-A-TT-A---C--T--TA---C---A-----A--A--------AACG-----------T-----A-----C--------A--A--A--G-----A-----T-----C---G----AGAC--TTCA-   6988
CPZ.CM.CAM3         -GAC-A-T--A---C--T---A---CT-----------G-----T---ACC-----------------A--------C-----AA-CC-A----A---A--CA-G-----C----TGCCA-AC--T----   6367
CPZ.CM.CAM5         -GA--A-T--A------T--TA--ACTG-A--------A-----T--AACT-----------------A--------C-----A--CC-A--------A--AA-G---------G--C-A-A---T----   6637
CPZ.CD.CPZANT       T-C--ATGGCAC--C------A--ACT--C-----AT-----GA----TCTA-T-----A--------T--------------T--A--A--A-A---T--AT-G-G---G-AG--G--GAT--T-CA-   6380

H2A.GW.ALI Env      _Y__M__N__H__C__N__T__S__V__I__T__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__D__D__I__N__Y__S__  gp120
H2A.GW.ALI          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---TC-AT-ATTCA-   7414

H2A.SN.HIV2ST       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---CC-AT-ATTCA-   6860
H2A.DE.BEN          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---CC-AC-ATTCA-   7433
H2B.GH.D205         T-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---CC-AC-ATTCA-   7400
H2B.CI.EHO          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---CT-AA-ATTCA-   7395
H2G.CI.ABT96        T-TA---GGCAC-----T--A---------CAGG-AT-A---GAT--ACAT-AT-GG--TG-T--AAGAT-TAGA--------T---C-A--G-----TT-G----G-------G---CT-AC-ATTCA-   6764
MAC.US.MM251        T-CA---ATCAC--------T--T--T--CCA-G-AT-T---GAC--ACAT-AT-GG--TA-T---AGAT-TAGG--------A--TC-A----A---TT-G--T-G-------G-C-CA-AT-ATTCA-   7333
SMM.US.SIVSMMH9     T-TA---ACCA----------AGT--T---CA-G-AT-----GAC---CAT-AT-GG--TG-T---AGAT-YAGA--------A---C-A----A---TT-G--T-GG------G---CC-AT-ATTCA-   6828
STM.US.STM          T-CA---G-CA---C--T---AGT--A---CA-G-AT-----GAC--ACAT-AT-GG--TG-CT-AAGGT-TAGG--------A--AC------A---TT-G--C-G---C---G---CA-AT-AT-CA-   6990
TANTALUS.UG.TAN1    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---CT-AT-AT-CA-   7002
GRIVET.ET.GRI-677   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-----GAT-AT-GA-   6740
VERVET.KE.VER-9063  T-TA---TTCAC-----TGAT-----T--A-A-G-A------GA---AAC--AT-GG--TGAGC-AAGGC--AGA--C--------A--A--A-----TC-GT-------CC--G--C-TGAT-AT----   7039
VERVET.KE.AGMTYO    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--G-AGAT-AT----   6511
VERVET.KE.AGM155    T-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--TG-GAT-ATGCA-   7019
VERVET.DE.AGM3      T-TA---TTCA------TGAT-----T--A-A-G-A------GAT--AAC--AT-GG--TGAGT-AAGAT--AGG--C-----------A--G-----TT-AT-------C--CG--T-TGAT-ATGCA-   6537
SABAEUS.SN.SAB1C    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---CAGAT-ATTCA-   7213
SUN.GA.SIVSUN       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--ATTAAATGC-A   7175
MANDRIL.GA.MNDGB1   T-TA---ATCA------TGAG-----T-A----G-A-A-----AG----G-CTT...TT-AT--GATGT--TTTAGG-----TG--TC-A--A-A--TC--GT---G--A----G-G---TTAAAT---A   6441
LHOEST.KE.hoest     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-AC-ATTAAA-TT-A   7113
LHOEST.CD.hoest447  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--ATT-AATTTCT   6044
LHOEST.CD.hoest485  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--ATT-AATTTCT   6041
LHOEST.CD.hoest524  T-CA----CCA---C--TGATAGTA-T---CTG--A-A---CAAT----G-GTTA-GTCAAATGC-TATT-TAGGCT------A--A-----C-ACAT-C--T-------C---G-A---CTAAA-TT-T   6047
RCM.NG.SIV-NGM      TTTA-A-GGCA---------AA--AGT--ACA----TT----GA-CCAAA--A--GG--A---T--AGGT--AGG--------A---C-A----A---C--A--G------CT-G--GCA-AT-AT-CA-   7543
COL.__.SIV-CGU1     -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----A-A--GGG---   6334

SYKES.KE.SYK173     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---CA-AT--TG-A-   6909
SYKES.KE.SYK173 Env _Y__L__L__H__C__N__T__S__V__I__S__A__A__C__E__K__Q__T__F__Q__P__F__P__I__Q__Y__C__A__P__P__G__Y__S__L__L__K__C__N__D__T__N__F__E__  gp120
```

```
B.FR.HXB2 Env       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__V  gp120
B.FR.HXB2           GAACAGGACCA......TGTACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAA...............GAAGAGGTAGT   7036

A.UG.U455           ---A---G---......--C-GG--------------------------------C-A------G---------------------------------------...............AG---AA--AG   6481
C.ET.ETH2220        ------------......--CCAT--------------------C------------AA-----G--------------AT---------T---A---------...............-GT---AC-A-   6439
D.CD.84ZR085        -------C---......--C-A----------T-----------------------------G--G--------------------------------------...............------A-C--   6578
F1.BE.VI850         -G-----G---......--C-AG-----------------------------------AA-----G------------T--T------------C---------...............----GTA----   6358
G.SE.SE6165         ------------......----A---------T-------------------------A------G---------------A--AC-G---------T--------...............-G-A-AA--AA   6460
H.CF.90CF056        --------TT-......---------------T--------------------------A-----G-------------------A--------A--C---------...............---C--A-CA-   6367
J.SE.SE7887         --------T--......--C-A----------T-----------------C-------A------G---------------A---C-----------A--------...............-G---CA--A-   6349
K.CM.MP535C         -G----------......----A-----------------------------------AA-----G-----------------------------C---------...............------A--A-   6244
01_AE.TH.CM240      -G-----G---......----A----------T----------C------------A------G------------T-------------------------...............------A--A-   6616
02_AG.NG.IBNG       -------G---......--C-A---------------------C----------C-A-----------------------------------------------...............-G---A-----   6560
03_AB.RU.KAL153-2   -------G---......---------------T-------------------------A------------------------------T------------...............-----------   7053
04_cpx.CY.94CY032-3 --TT---G---......--C------------T-G----G---C--T----------A------G------------T------------A--CT-----ACG...............-----------   6424
N.CM.YBF30          ---AT---AGC......------------G--T--T--------C-----------A-A------GA----C-----GT-AA-CC-------A--CT--AATAC-...............--T-GAA-T--   6629
O.CM.MVP5180        -------C-T-......--CCAC---A-TTCAGTG--TACT-----------C--C-A----AC----AGT---------AA-AC-G-----G-CA--CT-TAG-...............---A--A--AG   7071
O.BE.ANT70          -------CA--......--C-G---CA-A-CGGT---TACT-----------C----------AC----AGT-----G--AA-A---------G-CA--CT-TA--...............-G-A-AA--AG   7071
CPZ.US.CPZUS        ---AG---GAG......----A------T--T-----G--G-----T-----C--A-AA-----G--GA-T--A--G--CA-AA-----------T-----ACT.........AAAAAT-TTACT---AG   7086
CPZ.GA.CPZGAB       ---A---TAA-......-----T-----A--T-----G--T-----------------AA-----G--TA-C-----GT-A--TA-T-----A---T----T---GGAAACATAACTGTAAG--T--AGAA   7112
CPZ.CM.CAM3         --------A--GGAAGA----AC---A-A---------C---------T-----C--C--A-----G---A-T--A---T--A-CC-T-----A---G----C---...............A-CA--ACTA-   6482
CPZ.CM.CAM5         ------------......----------A-A----------------C--T-----C--A--A---------A-T--A---G--A-CC-T-----A--C-----T---GAGGGGCCCGCTGTT-CTAGAAGT-A   6761
CPZ.CD.CPZANT       --GT---GATG......----A-------TCAGT------T--C--T--------A--C---A-G--GG----ATGGT-A--A---------A-C-TACCA-AC-AACACTTCAGTAGTAATGA-T-GTCG   6504

H2A.GW.ALI Env      G__F__A__P__.__N__C__S__K__V__V__A__A__T__C__T__R__M__M__E__T__Q__T__S__T__W__F__G__F__N__G__T__R__A__E__.__.__.__.__.__N__R__T__Y  gp120
H2A.GW.ALI          -CTTT-C---C...AAT--CT-T--A--AGTAG-T-CTAC---C---AGGATG--GGA-A-GCA-ACT--T---TGGT-TGGC--T-------C-AGG------...............A-TAGAAC-TA   7526

H2A.SN.HIV2ST       -CTTT-AG--C...AAT---T-T--G--AGTAG-T-CTAC-------AGGATG--GGAAA-GCA-ACC--C---TGGT-TGGC--T-------C-AGG------...............A-TAGAAC-TA   6972
H2A.DE.BEN          -CTTT-A---T...AAG--C--T--A--AGTAG-T-CTTC---C---AGGATG--GGAAA-GCA-ACT--T---TGGT-TGGC--T-------C-AG-------...............A-TAGAAC-TA   7545
H2B.GH.D205         -CTTCATG--C...AAC--C-GT--G--AGTAG-GTCCTCC--C---AGAATG--GGAAA--CAGTCC--T--ATGGT-TGGC--C-----T-CAAGG-----G...............A-CAG-AC-TA   7512
H2B.CI.EHO          -CTTCATG--C...AAC----GT--G--AGTAGT-TCT-TG...TACAGAATG--GGAAA--CAGACC--T--ATGGT-TGGC--C-----T-CAAGG------...............A-CAG-AC-TA   7504
H2G.CI.ABT96        --TTT-C---T...AAT--CT-T--A--AGTGGTCTCTAC-------AGAATG--GGAAA--CA-ACC--T--ATGGT-CGG---T-----A-CAAGG------...............A-TAGAAC-TA   6876
MAC.US.MM251        -CTTTATG--T...AAA---T-T--G--GGTGGTCTCTTC---C---AGGATG--GGA-A--CAGACT--T---TGGT-TGGC--T-----A-C-AG-------...............A-TAGAACTTA   7445
SMM.US.SIVSMMH9     -CTTT-CT--T...AAC-----T--G--AGTAGTTTCTTC---C---AGAATG--GGAAA--CA-ACC--T---TGGT-TGGT--C-----T-C-AG-------...............A-TAGAAC-TA   6940
STM.US.STM          -CTTT-C---T...AAT--CT-T--G--AGTAGTCTCCTCG-----TAGAATG--GGA-A--CA-ACT------TGGT-TGGC--T-----A-C-AG-------...............A-TAGAAC-TA   7102
TANTALUS.UG.TAN1    --TTT--TGTT......----G------TTCAGT---TAGT-----TGGATTG--G-ATA-TAC---GAGCT-AGC-T-TGGCA------------AG------...............A-CAGAAC--A   7111
GRIVET.ET.GRI-677   -CTTT-CT----...AAG--C-AG-----TTCAGT---G--T-----TAGATT---C-ATA-TACTA--A-T--AGGGA-AGGA--------T---AG-T-----...............A-TAGAAC--A   6852
VERVET.KE.VER-9063  -TTATAAG-A-...AAT----GT-----TTCAGT------C-----TGGACTT--G-ATA--AC---T--T--AGG---A-----G-----A--CTATT-T--G...............A-TCGAAC-CA   7151
VERVET.KE.AGMTYO    -GTATAA--A-...AAT--CT-------ATCAGT---G--T------GGCTT---G-ATA--AC---GA----AGGGT--T--C-G-----A--CTATCAT--G...............A-TCGAACCCA   6623
VERVET.KE.AGM155    --TTTAAG---...GAA---T-T-----TTCAGT---G--T--C---AC-TT---G-ATA--AC----A-C---GGT---T-A--G-----A--CTATT-----...............A-TCGAACCCA   7131
VERVET.DE.AGM3      -GTTTAAGA--...AAC---T-T-----TTCAGTG--G--T-----TA-CTTG--A-ATA--AC---GA-T---GG----T----G-----G--CTACT----G...............A-TCGAACCCA   6649
SABAEUS.SN.SAB1C    --CATAAG......GCA----GG-----A-CAGT-TC-GCT--C--TAGGTT---C-ATA-TACT---AGC--AGG-A-AGGAA-T-----T---TAT-T--C-...............A-TAGAAC--A   7322
SUN.GA.SIVSUN       CT-AGAA-.........----A----A-A-CTG-TACC-CT-----TA--TAT--G-CAAGTAC---TAGT-GCTTTT-TGGC--T-----A-C-AG-CATA-G...............-----T-AGT-   7281
MANDRIL.GA.MNDGB1   AT-A-TTG.........----GC---A-ATCAG----G--G-----T--GCACT-AGTAG-CAC----AGT-GCTTTT-TGGC--T-----A-C-A-GCATA-G...............----GA-A-T-   6547
LHOEST.KE.hoest     -T-A-AA-.........---GA----A-A-CAG-TAC--C------TGGATAT--GCTTAGTAGT---AGT-G-TTCT-TGGC--T-----A-C-AATCATACC...............AGG--T-AGC-   7219
LHOEST.CD.hoest447  C-G-TAAT.........-----C---A-A-CAG--ACC-CT--C---G-CTAT--G-TTAGTTC---GAGT-G-TTCT-TGGC--T-----A-CCAATCATAC-...............--C--T-A-T-   6150
LHOEST.CD.hoest485  C-G-TAAT.........-----C---A-A-CAG--ACT-CT--C---G-CTAT--G-TTAGTAC---GAGT-G-TTCT-TGGC--T-----A-CCAATCATAC-...............--G--TAAGT-   6147
LHOEST.CD.hoest524  C-G-TAAT.........-----T---A-A-CAG-TAC--C---C---G--TAC--G-TAAGTTC---TAGT-GCTTCT-TGGC--T-----A-CCAATCATAC-...............-----C-AGT-   6153
RCM.NG.SIV-NGM      -CTTT-ATA-C......---CA------T-CAG--ACGTC---C-----CATG--A-ATA--AC------C--ATC-T-TGGA--------ATC-A-TAGT--G...............A-TAGAAC-TG   7652
COL.__.SIV-CGU1     -G------G--......--C-AC-----G-CAG----GTCC-----C----AGT---ATAT-AC---CATGT----TG-AT-AG-C----C-TCAAA--AGTTG...............TCT--TTGG-C   6443

SYKES.KE.SYK173     --GAC-ATGTC......-----------A-CTG-----TC---C-----G-A-T-C-ATA--T-G-CTAGC--ATGGT-C-AAC-G-----G-CCTACAA--C-...............A----TAA---   7018
SYKES.KE.SYK173 Env G__D__D__V__.__.__C__T__N__V__T__A__V__S__C__T__Q__E__F__N__T__L__A__S__T__W__F__Q__L__N__G__T__Y__K__A__.__.__.__.__.__K__D__K__V  gp120
```

```
                                                                                                                  |-> V3 loop start
B.FR.HXB2 Env       __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
B.FR.HXB2           AATTAGA.....................TCTGTCAATTTCACGGACAATGCTAAAACCATAATAGTACAGCTGAACACATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGAATC   7145

A.UG.U455           G------.....................----AA--------AA-------A--------------------TGT--ATC-----A----C------T-C-----TT----T...----------AT--A   6587
C.ET.ETH2220        -------.....................-T--AA---C-G--AA-------C----TA--------------T--TGA--------------C------G-------G---T---------G----T--A   6548
D.CD.84ZR085        ---A---.....................----AA---C----AA------------T------------T--T--TCA------------------------G---T----A--GGA----C----G-CA   6687
F1.BE.VI850         ---C---.....................---CAA---A--T-AA-T-----A-----------------C--T--TGA-------C-G--------------------------------G---G-T--A   6467
G.SE.SE6165         -G-----.....................----AA--------A------A-C---GT---------------T--T-A-A-------------------C-------------------TG-----G--A   6569
H.CF.90CF056        -------.....................A--AAA---A--T-A------A-C----A---------------T--G---C-A---A-C----CA-----C--G--T-----T-----G----C---T--A   6476
J.SE.SE7887         -------.....................----AA--CA--T-A--------C----A------------A--T--T-A-A-----------GTG---TAC-----T--T-----------G---G-T--A   6458
K.CM.MP535C         ------G.....................----AA---A-T--A--T---A-A----A---------------T--TGA-A-----C----------------G--A--------------------T--A   6353
01_AE.TH.CM240      ---C---.....................----AAG--C----AA-------C--------------G--C--T--T-A------------C--------C------TC--------------C---T--A   6725
02_AG.NG.IBNG       G--C---.....................----AA---A----AA-------C------------------T--GCT-ATC-----A-------------C-----T------------------G-TG-A   6669
03_AB.RU.KAL153-2   -------.....................---------------------A-------------------------AGA-C----G---------------------------------------G-T--T   7162
04_cpx.CY.94CY032-3 -------.....................---AAA---A----A------A-C----AT--------------TGCA-AGG-----A-------------C-----TGG------------------TG-A   6533
N.CM.YBF30          T------.................................-AT--T-G-CAC-GT-ATC-GT-G--G--ATG---TGAGA-A--GCC---A-----------G--AGGA--T------G--GG-CAGG-G   6726
O.CM.MVP5180        -----TG.....................GGAAAA---A-T--A--ATCA--A--G-AT--C------ACC--A-----TC--A--A-C--G-CC--C-T----GAAGGA-TTGCAGAGGT-C--GAT--A   7180
O.BE.ANT70          ---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--G   7177
CPZ.US.CPZUS        --G--A-....................................A--TT---AG-C-TT--TC-------AT-CTCAGAGGGA--CA-T--G-C-----TT-----AGGA---------GTGGG--ATG-A   7180
CPZ.GA.CPZGAB       T-AGTC-....................................A-A---A--G-TGT-TGG-----------AGTAGA-G-A--TAGTC-A------CAT-----AGGA-----------GGG-GAGG-G   7206
CPZ.CM.CAM3         TGC----....................................AGG--C-GGT-C-R-T-CC-GA-N--RT-TC-G-ANN-G--CAGC--A--------------AGGA--T-NC-GC---GG-CA---A   6576
CPZ.CM.CAM5         -GA-CT-....................................AGAC-CAAA-TC-TT------CA-...T-T--AGA-GGGA-TA-C--A-T-----TG-----AGGA--T--C-GC---GG-CA-C-A   6852
CPZ.CD.CPZANT       C-AA-AT....................................--ATC--TGCTTGTA-G-T-T-G-A-AGAATT-GA-AACT--AC-----CA----T------AGGA--T-GG---GT--G--ATC-A   6598

H2A.GW.ALI Env      __I__Y__.__.__.__.__.__.__.__.__W__H__G__R__D__N__R__T__I__I__S__L__N__K__H__Y__N__L__T__M__H__C__K__R__P__G__N__K__T__V__V__P__I_  gp120
H2A.GW.ALI          T--CTAT........................TGGC--GGT-GA--T---AGA-CT-T---C-GCT--A-CAAAC-TTATAA-C-TACT--GC------AG--G--AGGA--T--G---GTTGT-CC---A   7632

H2A.SN.HIV2ST       T--CTAT........................TGGC--GGT-G---T---AGA-CT-T---T-GCT--A-CAA-TTTTATAA-C-CACCG-AC------AG--G--AGGA-----G---GTTGT-CC---A   7078
H2A.DE.BEN          T--CTAT........................TGGC--GG--GA--T---AGG-CT-T---T-GCT--A-CAA-T-TTATAA-C-CAC---GCG-----AG-----AGGA--T--G---GTTTT-CC---A   7651
H2B.GH.D205         T--ATAT........................TGGC--GAA-AA------AGG-CC-T-----GCT--A-TACAT--TATAA-T-GTC---AC-C----AG--G--AGGA-----G--GGTTGT-CC---A   7618
H2B.CI.EHO          T--CTAT........................TGGC--GGT-AA------AGG-CT-T-----GCT--A-TTCAT-TTATAA-C-GAC---GC-C----A---G--AGGA-----A-TGGTTGTGCC---A   7610
H2G.CI.ABT96        T--CTAT........................TGGC--GGG-GA------AGG-CT-TA----GTT--A-TAA-T-TTATAA-T--AC---A-RA----GG-----AGGG--T--A---GT-TT-CC---T   6982
MAC.US.MM251        T---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--   7551
SMM.US.SIVSMMH9     C---TAT........................TGGC--RG--GAAG----AGA-CC-TA--T-GCT--A-TAA-T-TTRTAA-C--AC---G-GA----G------AGGA--T--G---GTCTT-CC-G--   7046
STM.US.STM          T--CTAT........................TGGC--GG--G---T---AGG-CT-TT--T-GCT-GA-CAA-T-TTACAA-T--AC---G-GC----G------AGGG-----A---GTCTT-CC-G--   7208
TANTALUS.UG.TAN1    ---AT-G..................CAAAAGCATGGAG-G-GCA-T---T--GTG-TA----A-C--A-TAAAC-TTATAAGT-GA-G--AGTG----G------AGGA--T--G---GTCTT-CC-G-A   7223
GRIVET.ET.GRI-677   G--AT-G..................CAGAAA-GAGGAAATGATA-TG--A-AGTT-TA----AGT-GA-TAA-TTTTACAACT-GAC-G-G-GA--CCG------TGGT--T--A---GTGTTGCC-G-A   6964
VERVET.KE.VER-9063  ---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-A   7263
VERVET.KE.AGMTYO    G--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-A   6735
VERVET.KE.AGM155    G--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-A   7243
VERVET.DE.AGM3      G--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-A   6761
SABAEUS.SN.SAB1C    ---CT-G..................CAGAAAAATGGCAATT-CA-TG--T--GTT-TT--T-GGC-CA-TAGAT-TTTTAA-T--ACT--C-GA----G------AGGA--T--G---GT-TTGCCTG-A   7434
SUN.GA.SIVSUN       ---ACC-ATCAAT.........AACAAGGTA-GAC-AGGGG-T--GGGA-AAT-TGTATGG-A----GCAGCT--GTGGGGGT---T----C-G--C-T---GAAAGGG--T-GGT--CA-GT---T-CT   7402
MANDRIL.GA.MNDGB1   G--ACCCATAGATGATAAATATAGGGGCC-A-AGG-A--TCATC-A-GGAAGTTTGT-TAT-AG--G-CAGGA--ATATGGCT--A-G--AG-A---CAC---AAAGGA--T-GGT--GT-GTG--T-CT   6677
LHOEST.KE.hoest     T---CC-CTAACT......CCAAATAAGATG-AAG--C-A-AT-GTGCAAAGTTTGTGTAT-A----GCAGGA--ATGGGG-T--ATT----GA--C-T----AAAGGA--T-GAT--GA-GT-TCC-CA   7343
LHOEST.CD.hoest447  ----CC-CTAACT......CCAAAAAAGATG-GA---C--CTA--AGCAAAGTTTGTATAT-A----GCAGGA--ATGGGGAC-CATT----GA----TT--GAAAGGA--T-GGT--GA-GT-TCT-C-   6274
LHOEST.CD.hoest485  ----CC-CTAACT......CCAAAAAAGATG-GA---C--CT----GCGAAGTTTGTATAT-A----TCAGGA--ATGGGGAC-CATT--C-GA----TT--GAAAGGA----GGT--GA-GTCTCT-C-   6271
LHOEST.CD.hoest524  ----CC-CTAACT......CCAAAAAAGATG-GAG--C-T-AT-GAGC-AAATTTGTTTAC-AG---GCAGGA--ATGGGGAC--ATC----GA----T----AAAGGA--T-GGT--GA-GTCTCC-CA   6277
RCM.NG.SIV-NGM      G--CTAT........................CAA-GAAAACAAAGT---AGG-C-GTT---GGGT--A-TAGAC-TTTTAA-T--AC-G-A-C-----GG-----A-GT----GA---GTG---G-----   7758
COL.__.SIV-CGU1     G-AAGAT...............AGAGAGGGA---TGGAAG-AT---TCA-GG-C--TTGAGTATTATTG-T-TCCA-A-GA-A---CTT-AGGG----T----AGAGGA--T-G-T-TCAC-G--ATC-A   6558

SYKES.KE.SYK173     T-GATTTATCAAA............CAAAAG-AT--AAAT......G-GT-AGT--TT--TT----T-CAGAAGCACTTAGAT-GC-G----TA---GA------AGGA--TG-GT-T-TC----AT--T   7130
SYKES.KE.SYK173 Env __R__F__I__K__.__.__.__.__Q__K__D__K__N__.__.__E__S__V__I__I__L__V__P__E__A__L__R__L__Q__I__I__C__E__R__P__G__N__E__S__I__K__N__I_  gp120
```

```
                                     loop tip                                                      V3 loop end <-|
B.FR.HXB2 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
B.FR.HXB2           CGTATCCAGAGA...GGACCAGGGAGAGCATTTGTTACAATA............GGAAAAATA...GGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGA   7257

A.UG.U455           ----GGT-T--TATA---T----ACA------CTA-GT--C-............--T------ATA--GG----A-----------------TG-C------AGGG-C-------GA---A--C-----G   6705
C.ET.ETH2220        A-G--A.........--------ACA-A----CTA-G---C-............---G-C---ATA---G----A------------------------GA--A------------A---C--C--A--G   6657
D.CD.84ZR085        -C---A.........----A----CA---GC-CTA-----C-............A-GT-T-C-ACAA-G-TA--AG--------T-------T------G---T---------------C---CG----G   6796
F1.BE.VI850         -A-T-A.........--------ACA-A----CTA-G---C-............---GC----ATA---G-C--A---A-G------------------G--A--C--------------C-GG-GT-TG   6576
G.SE.SE6165         A-A--G...G--ATT--------ACA-A-G--CTA-G---C-............--TGCC---ATA---G-C--A-----------------TG---C--A-AG---------AG-GG-C---C-GA-TG   6684
H.CF.90CF056        -A-T-A.........--G-----AC-------CTA-G---C-............--TG-C--CATA---G----A-----------------T---------A--G-C--------G------C-C---G   6585
J.SE.SE7887         -AC--G.........--------ACA--TGC-CTACG---C-............---G-----ATA--------A---G--A-----------------GA-AG-G-T----G----------CGTAGAG   6567
K.CM.MP535C         -A---G.........--------A-A------CTA-----C-............--TG-T---ATA---G----A-----G--------C---------G---A---------C-TG-------GCAGAG   6462
01_AE.TH.CM240      AC---A.........--------AC---T---CTA--G--C-............---G-T---ATA--------A---A-----T-----G-G----A-G--A-------------AGT----------G   6834
02_AG.NG.IBNG       -A---A.........---------CA------CTA-G---C-............--TG-C---ATA--GG----A-----------------TG-C----A-A--G----------A------C-T---G   6778
03_AB.RU.KAL153-2   -A---A.........-------------------TA-G---C-............---G-C---ACA---G----A-------------------------T-A----------------A----------   7271
04_cpx.CY.94CY032-3 -A---A.........--G-----ACT-A---GGTA-G---C-............--TG-----ATA---G----A------------------------G--AATG-T------G----C------GT--   6642
N.CM.YBF30          -AG--A.........-----T-CT-TGA-----TA--AC---............-A-------GTA---G-C--T---------T-C-----TG-CTC--A--A-CT----G-ACCA-TG-GG--TAGA-   6835
O.CM.MVP5180        TA--CA.........--T---AT----TGGCGCAG--TG-C-CTTAAAAGAAGTAAC--T-C-TCACC--GATCA--GGT---TT------CATA--A--AGA-TGT----G-A--TG-CC--C----A-   7301
O.BE.ANT70          A-A--A.........--T---AT-GCC...-GGTAC-GC--GGGAATAGGG...----C-GC-GGAAAC-GCTCA--GGC---TT----C--GTA--A-GCCA-TG-T---GGA--A-TA--------A-   7292
CPZ.US.CPZUS        -AAC-A.........--------A-TGA-T---TA--AC---............CC---G---GTA---G--G-A---G----------------CTCA-A-CTG-C----G-G--ACAGAG----T-C-   7289
CPZ.GA.CPZGAB       -AG--A.........--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-AG   7315
CPZ.CM.CAM3         -AG--A.........--------C-TGA-C---TA--AC---............-A---T--TGTA---G---CC----G----T-----C-G--C-A---GA--GTC---G-CG-A-GAC----TG-A-   6685
CPZ.CM.CAM5         -AG--A.........---------GT-T-C--CTA--AT---............-A---C---GTA--GG--CCC-----------------TG-C--C-AG--CC-G----TG--A-AG-----CG---   6961
CPZ.CD.CPZANT       -AA--A.........--------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-   6707

H2A.GW.ALI Env      _T__L__M__S__.__G__L__I__F__.__H__S__.__.__.__.__.__.__Q__P__I__N__K__R__P__R__Q__A__W__C__W__F__.__K__G__E__W__R__K__A__M__Q__E__  gp120
H2A.GW.ALI          ACAC-TAT-TC-...--GTT-ATATTT...CACTCC..................CAGCC---CAATAA--GACCC---------TGG--CTGGT-C...-A--GCG------GG--AG-CA-GC-GG--G   7735

H2A.SN.HIV2ST       ACAC--AT-TC-...--GTT--T-TTT...CACTCC..................CAGCC---CAATA---GACCC--G------TGG--CTGGT-C...-A--GCG-G-----GG-AG-CA-G--GG--G   7181
H2A.DE.BEN          ACAC-TAT-TC-...---TT--T-TTT...CACTC-..................CAGCC---CAACAC--GGCCT--G--G---TGG--CCGGT--...G---GC-G-----GGG-AG-CA-GC-GG--G   7754
H2B.GH.D205         A-A-C-GT-TC-...----T-CTTTTC...CA-TCA..................CAGCCT--CAATAAG-GACCC--------TTGG--CTGGT--...-AG-G---C----CAG-AG-CA----GG--G   7721
H2B.CI.EHO          A-A-C-GT-TC-...--TATTCTCTTC...CA-TCA..................CAGCCT--CAATAA--GACCT-A------TTGG--CTGGT--...-A--G---C----CAG-AG-CA--C-GG--G   7713
H2G.CI.ABT96        AC---TAT-TC-...--G-T--TCTTC...CACTC-..................CA-CC---CAAT-A--GACCC-----G--TTGG---TGGT--...G---GC--------GG-AG-CA-GC-GG--G   7085
MAC.US.MM251        ACC--TAT-TCT...---TTG-TTTTC...CACTCA..................CA-CC---CAAT-AT-GGCCA-AG--G---TGG---TGGT--...G---G---------GG-TG-AA-----G--G   7654
SMM.US.SIVSMMH9     ACC--TAT-TC-...--GTTG-TCTTC...CA-TCG..................CA-CCC---AAT-AG-GACCA-A---G--CTGG--CTGGT--...G---G-G-------AG-GG-CA-CCGGG-AG   7149
STM.US.STM          ACC--TAT-TC-...--CTTG-TCTTC...CA-TCG..................CA-CCC--TAAT-AG-GACCA-A------TTGG--CTGGT--...G---G-G------GAGGAG-AA-----G--G   7311
TANTALUS.UG.TAN1    ACC---AT-GC-...--T-TC-T-TTC...CACTCA..................CA-C--TATAATAC---AT-A---------TGG--CC-TT-C...CAG-GTG-T-----AGGAG-C-GG-GGG-AG   7326
GRIVET.ET.GRI-677   ACA---AT-GC-...--GTT--TATTT...CACTC-..................CAG---TATAATACC-GGT-A-A------GTGG--CC--T-C...CA--G-G-T-----AGGGG-A-GG---G-AG   7067
VERVET.KE.VER-9063  ACC--AAT-GC-...--TTT--TTTTT...CA-TCG..................CAG---TACAATACG--GC-T-----G--CTGG---C-TT-C...CA--GC-GG----GAG-AG-C-GG--GG--G   7366
VERVET.KE.AGMTYO    ACG--AAT-GCG...----TG-TTTTC...CACTC-..................CA----TACAACATG--GC-T-----G--TTGG---C--T-C...GA--GC--T----GAGGTG-C-GGCGGG-AG   6838
VERVET.KE.AGM155    ACG--AAT-GC-...---TT--TCTTC...CACTCA..................CAG--GTATAATAC--GAC-A--G--G--CTGG--CC--T-C...CAG-GC--T-----AGGAG---GG--GG-AG   7346
VERVET.DE.AGM3      ACA---AT-GC-...--G-T--T-TTT...CA-TC-..................CAG-GGTACAATAC--GGC----------TTGG---C--T-C...CAG-GC--C----GAGGAG-C-GG---G-AG   6864
SABAEUS.SN.SAB1C    ACC---AT-GCT...--CTTG-TCTTC...CA-AGC..................CAG---TATAACACC-GGT---A------TTGG--CTGGT--...G---GT--T----GAGGAG-C-GG--GG-AG   7537
SUN.GA.SIVSUN       ATA-GTAGT-C-...----TGCT-TTTTAT-A--GG..................CT-G-GCC-GGTTCT--GT-A----TG-----G-----AT--...GT--G-C-----GGA-GAG-C--TGCCACCT   7508
MANDRIL.GA.MNDGB1   -CATCAGCT-C-...---TT-TTATTTTATCA--GG..................TT-G--CCTGGAAAG---T-A-AGA---GCATG--C-C-T-C...-A--G-CGT---GGGTTAG-AC-TTGGAGTC   6783
LHOEST.KE.hoest     AT--G-TCT-C-...---TATTTATTTTAT-A--GC..................CT-G--CATGGCA-C-GAT-A---TT------G-----AT--...GA--GCC-----GGA-GA-TG--T--TA-TC   7449
LHOEST.CD.hoest447  AT--GTTCA-C-...----TTCT-TTTTAC-A--GC..................TTGG--CATGGTA-C-GGC-----TTG--T--G-----AT--...GA--G-C-G---GGA-GA-TG--TC-TA-TT   6380
LHOEST.CD.hoest485  ATC-GTTCA-C-...----TTCT-TTTTAC-A--GC..................TTGG--CATGGTA-C-GGC-----TTG--T--G-----AT--...GA--G-C-----GGA-GG-TG--T--TA-TT   6377
LHOEST.CD.hoest524  ATC-GTTCA-C-...----TTCT-TTTTAC-A--GG..................TT-G--CATGGGTCT-GAC-A---TTG-----A--C--AT--...GAG-GTC-----GGA-GA-TG--T--CA-TT   6383
RCM.NG.SIV-NGM      TCAT-AGCA-C-...---GT-TTC-TCT-TC-AAGGGTG...........................CAAAA--GGCCA-A-GG---TTGG----GAT--...GAT-G-G-T-----AGGTG-G-GG---G--G   7858
COL.__.SIV-CGU1     AAC-CTGCC-AT...---G--AAATTCTAC-A--AGCTT--TCCTTATTCAAAA--C-TTTATGGAA-GTG-CA-TTTGT-C--ATGAC-GGGCAG-A--AGAA---CAAA-CAC-ATTCGC--T-G-A-   6685

SYKES.KE.SYK173     -AAT-GGCTGC-...--CTATTTCCTCC--G-AA-A.........CAAGGGAAATTG----CCGGGA-GG-CGCA-A--GT--CTTC---CGAG-A...-C--G---T----CAG-ATT---T--G---G   7245
SYKES.KE.SYK173 Env _Q__L__A__A__.__G__Y__F__L__P__V__I__.__.__.__Q__G__K__L__K__T__G__R__D__A__K__R__A__F__C__R__V__.__T__G__N__W__T__E__F__F__K__Q__  gp120
```

```
B.FR.HXB2 Env       I__A__S__K__L__R__E__Q__F__G__N__N__K__T__I__I__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  gp120
B.FR.HXB2           TAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATC...................................................................................   7304

A.UG.U455           -----GAAC------AGA--A-G---AAT--C--A-C-............................................................................................   6743
C.ET.ETH2220        --AAGGAA------CA-A-G--C--CCCT---...------.........................................................................................   6695
D.CD.84ZR085        -------A------G--A-C-TT--GAACC-A-CA---............................................................................................   6834
F1.BE.VI850         --AAGGCAG------AGTCG--T--CCCT------...---GC-......................................................................................   6617
G.SE.SE6165         -C---GCAG--C--G-GA--ATC---AAT--G-GC-GCGAG-AC......................................................................................   6728
H.CF.90CF056        ---T--CAC-----G--AT---C--GAAC----GA-C-............................................................................................   6623
J.SE.SE7887         ------CA---C----------C---AAT--A-CA...............................................................................................   6602
K.CM.MP535C         --AAGGAA--GC---A------T---AAG---GGA-C--T-.........................................................................................   6503
01_AE.TH.CM240      --A--GAA-------A---G--C---AAT--G-CA...............................................................................................   6869
02_AG.NG.IBNG       ---T--CAC------AGACGT-C---AAG--C-CC-C-............................................................................................   6816
03_AB.RU.KAL153-2   ---T--T----------A----------G-----A-C--T-.........................................................................................   7312
04_cpx.CY.94CY032-3 --AG-GAAG----G-A-AG--TC--CCCT-----A-C--T--A-......................................................................................   6686
N.CM.YBF30          C-AGAGAGG--A---AGA--ATCC-G--G--A--C--C..........................................................................................AT   6875
O.CM.MVP5180        C-----TA-GG-ATTT-A-T-TTG-AAACC-A-CAG-G-AT.......................................................................................GT   7344
O.BE.ANT70          C----GAA-GG-ATTT-----T-G-AAAC----CAGGT-GT.......................................................................................AT   7335
CPZ.US.CPZUS        CTTTA......GAG-T-AT-A--AAG-A-GCA--CCTG----AGG-AGAGTTAATTCCA...AAT.................................................................   7345
CPZ.GA.CPZGAB       --AAA-AGGCT---GC-ACCTC--CAAATCGA-CAGC-G-C-AC--AACACTCAACAGA...GCC.................................................................   7377
CPZ.CM.CAM3         CT-G-CAAGCC--GC----G-TT---AC---CCT--C-CA-G-C-A-TTT...ACAGTA...AGT.................................................................   6744
CPZ.CM.CAM5         CC---TTAGCCA--...-C-A-G-A--AT--A-GC-T-G---A-G--ACCTTTACACCA...AAC.................................................................   7020
CPZ.CD.CPZANT       C--AGCA-GTTC-TGCG--G--T-GGAA---AGTAG-C-AC-A--C-AAT................................................................................   6757

H2A.GW.ALI Env      V__K__E__T__L__V__K__H__P__R__Y__K__G__T__N__D__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__N__Q  gp120
H2A.GW.ALI          -GAAGGAA-CCC-TGT-A----TCCCA-GT----AGG---C-ATGA-ACA...........................................................................AACCA   7790

H2A.SN.HIV2ST       -GAAGCTA-CCC-TGC-A----TCCCA-GT----AGG---C-ACGA-ACA...........................................................................GAAAA   7236
H2A.DE.BEN          -GAAGCAA-CCC-TGT-C----TCCCA--T-C--AGG--TC-ATGATACA...........................................................................GGGAA   7809
H2B.GH.D205         -GAAA--G-CCA-C-T-A----TCCCA-GT----AGG-GGTGC--AAAAT...ATCACA..................................................................AGCGT   7782
H2B.CI.EHO          -GAAAGAG-CCA-T-A-A-T--TCCCA--T--TCAGG-----C--ATATC...........................................................................TCACA   7768
H2G.CI.ABT96        --AAGGAA-CCG-GGTCA-G--CCCCA-CT----AGG-----ATGA-ACA...........................................................................AAGAA   7140
MAC.US.MM251        -GAAACAG-CCA-TGTCA----TCCCA-GT---C-GG---T-AC-ATACT..............................................................................GA   7706
SMM.US.SIVSMMH9     -GAAGGMAMCC--GGTCA----TCCCA-GT----GGG---T-ATRAKAYT..............................................................................RA   7201
STM.US.STM          -TAAGGAA-CC--GGTCA----TCCCA-GT---CAGG---C-ATGA-ACG..............................................................................GC   7363
TANTALUS.UG.TAN1    -GAGA-AA-C-A--GTG--G-TTCCAAA-G-G--AT-TCG-GGG-CAAAT...........................................................................AACAC   7381
GRIVET.ET.GRI-677   -CAGAGAAG--G---AGA--GTGAAAAATCT--CAG--GT-...-G-ATA..............................................................................GA   7116
VERVET.KE.VER-9063  -TAAAGAAG--A---TCA--TT-CCAAA-G---GGT-TCGGGG--C-AAT...........................................................................GATAC   7421
VERVET.KE.AGMTYO    --AAACAA---A--GT---GTT-CCAAA-G-C-GGT-T-A-GG--C-AAT...........................................................................AATAC   6893
VERVET.KE.AGM155    --CAAGAGG--A--GT-A--TT-CCAAA-G-ACGGT-CCA-GGC-C-AAT...........................................................................GATAC   7401
VERVET.DE.AGM3      --AAA-ATG--A--GT-A--TT-CCAAA-G---GAT-CCA-GG--C-AAT...........................................................................GATAC   6919
SABAEUS.SN.SAB1C    -CAAAGAA-C-A--GTGAG--T-CCACC---A--GT-T-GTGGC-CAAAT...........................................................................GACAC   7592
SUN.GA.SIVSUN       ---GG-AACT---G--GC-GGTGGAACCTTTAGC---C-TG-GT-ATGGCTGT...ACATTT...GATAATATAACAAAGACATGCAGATTTACC...AATGGTACAGATTTTAAAAAGATGATTAAGTT   7629
MANDRIL.GA.MNDGB1   ------AAG--C---ATA--TT-AA--ACTCC-TC---GTG-ACCAGACCTGTAAAAATTTTACTAGCACTGGAGAGGAGAACAAACAAAACACGGACAAGCAAAAGGAGTTTGCCAAATGCATAAAGAC   6913
LHOEST.KE.hoest     ---GC-AA-TGC---AG--GTTGAA--C-G-AGCC-TG-ATTAT-CAGAAGGGACAGGGACATGTGATTCTAAGAAAACAACTTGT........................GGACGGAAATTAAAAGGCTT   7555
LHOEST.CD.hoest447  ---GA-A--TGC-T-A-AGG-TTAA--C--CAGCG-TG-ATTAT-C-CATGGAAAC......TGTATAGATAACACAAAGCCTTGT........................GGTAGACAACTAAAGGGATT   6480
LHOEST.CD.hoest485  ---GG-AA-TGC-TCA-A-G-TTAA--C--CAGCA-TG-ATTAT-CAAAT...GGACAA...TGTCCTGAGGGAAAACAGCCTTGT........................GGTAGACATCTAAAGGGATT   6477
LHOEST.CD.hoest524  ---GA--G-TG----AGA-G-TTAA--AG-C-GC--TG-ATTAT-AAGAA...GGAAAT...TGTTCCACACCAGGAAAACCATGT........................GGTAGACAATTAAAGGGTCT   6483
RCM.NG.SIV-NGM      --AAAGAGGC-A-T-ACA-GAC-AAA---T----AGGT---CG-CAGCCT..............................................................................GA   7910
COL.__.SIV-CGU1     -CAAA-AG--T----C--CTTGG...CT-G-GCGC-TTT--CGC-AAAATATA...........................................................................AC   6737

SYKES.KE.SYK173     --CA-GAAC-GGCC-CGA--ACC...T-G--A---GT---C-AC......................................................................................   7286
SYKES.KE.SYK173 Env V__H__E__Q__A__T__K__T__.__W__K__N__V__T__N__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  gp120
```

```
B.FR.HXB2 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
B.FR.HXB2           ..........TTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGG   7424

A.UG.U455           ....ATAATC---GCTAGC--------------TAT-------AC---A--T--C-------------A----------T--C--CA--T--GGC----------C-T----AA-GGC--C-TG-CA   6869
C.ET.ETH2220        ....ATAGAA-------C-----------------T-------AC---A--T--C-------------A----------T--C---A--T--A-C----------C--AAAAC-AG-ACTGTT-AAT   6821
D.CD.84ZR085        ....ATAATT-----A-C---------------T---------AC---A---------------------------------C---A--T--GG-------------G-A---AA--TATCAGGACAT   6960
F1.BE.VI850         ....ATAAAA-----C-------------------T-------AC--T---T------------A---A-------------CG--A--T--GG---------GAC--AG-C-CC----A-GGCACT   6743
G.SE.SE6165         ....ATAACC-----CTC---TG------------T-------AC---A--T-----C-T---A---A--G--------T------A--T--GGC---------TCA-GCCT---G-GATCA-A-A-T   6854
H.CF.90CF056        ....ATAAGC-------C-AA-------------ATG---G--AG---A--T------------A---A-----------T--C---A--T--GGG-------------G----GAA-TGCA----AAT   6749
J.SE.SE7887         ....ATAAAC----CATC-C-------------TAT---------G--A--T----------------A--------A--T-----A--T--A-------C-------G----GA---G-A--GCATT   6728
K.CM.MP535C         ....ACATTTAAACCA-C-AA-C-------A-----------C---------TG----------C---A--------------C---A-----A----------GAG--AG--GAG---G-----...   6626
01_AE.TH.CM240      ....ATAATC---C-A-C-C----------A--T-T-------AC--T---TCA----------A----------------T--C---A-----A-------------A------CC-AGGA-A-GAAACC   6995
02_AG.NG.IBNG       ....ATAATC---GCTA-CC---T---------TGT-------AC---A--T----------------A----------T--C---A--T--A-------------C------GA------C-A-A-C   6942
03_AB.RU.KAL153-2   .......GTC-----T--------------------------------T-------------------------------------A-----A-------------C------......-A-GG-ACT   7429
04_cpx.CY.94CY032-3 ..........---GCT-C-C--GT-----------T-------AC---A--T--C---------AA---A----------T--C---A------C-----------------ACACA-GC-A-A-GG-ACA   6806
N.CM.YBF30          AACCTTCAGGGC-CGAG-GAGGAAT-A---A----T----G-GAC-CACTTA-TG--C------A---A--G--------T-----CA-TT-CA--T-A------GAGGAA-TAC----CGAG--AG-T   7005
O.CM.MVP5180        TACCATAATA--C-GCAG-A-TAGT--T--A--TG-----G-AAGCCATTTACA------C---CAT--A-----C--T--T-----CA--T-TGGGA-------CTA-----TTA-C--CT----AAA-   7474
O.BE.ANT70          TAACATGACA--C--T--CAG-AGC--T--A--T-T---GG-AACCCATTTACAC-----C---CAT--A-----C--T--T-----CA--G-TA-GA--------TA---C-TT-CA...T---ACG-A   7462
CPZ.US.CPZUS        .....................G-------A---------G-G----ATATG-TGC-------------A--------T--T------A---TT-CC--A-----C-TG--C-ACAAC...........   7441
CPZ.GA.CPZGAB       ......................--T-----A--T-----GG-CAC-CAT--T-TG-------------A-----------T------A-T-GC---A-A---.........---GACAAC...........   7467
CPZ.CM.CAM3         .....................C-------A---------G-AAC--ATATG-TG--------------A-----C--T--T--C---A-T--TAC-T---------TA---G---AAA--C-AC-A-ATA   6852
CPZ.CM.CAM5         .....................C-C-----A--T---------AC--ATATG-TG--------------A-----C--T--T------A-C--T-C---C--C----AGTT-AAT-A--C-T--GA-AAC   7128
CPZ.CD.CPZANT       .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......GGAAAC   6880

H2A.GW.ALI Env      __I__N__F__T__K__P__G__R__G__S__D__A__E__V__V__Y__M__W__T__N__C__R__G__E__F__L__H__C__N__M__T__W__F__L__N__W__V__E__N__K__T__G__Q_  gp120
H2A.GW.ALI          AATTAACTTTACA--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-   7920

H2A.SN.HIV2ST       AATTCGTTTTA-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-   7366
H2A.DE.BEN          AATTAACTTTACG--A-CGGGAG----CTCA-----G---G-G-C-TTTATGT-GAC---C--CA----A------C-----------CATG--TTGGT-CC-C---T-GGTAGAAGAC--G-ACCAAACA   7939
H2B.GH.D205         AAAGTTAGTA-CAG-A--TGGAAA---TTCA--T-------CAAC-TATATGT-GACC------A----A-----C--A---------ATG--CTGGT-C--A--CT-GGTAGA-AAC--A-CA-A-ACA   7912
H2B.CI.EHO          GATAAGGTTAGCAG----CG-GAG-A-CTCA--T------G-AAG-TATATGT-GAC---C---A-G--A--G---C-----------ATG--TTTTT-C--A--CT-GGTAGAAAA--GA-C-GGGCTC   7898
H2G.CI.ABT96        MATAACCTTTACA-CA-C-GGAGA---TTCA--T------G-AAAGTTCATGT-GAC------CA----A-----C--G--T-----AATG--TTGGT-TC-A---T-G-TAGA-GAC...........   7258
MAC.US.MM251        TAAAATCAAT--A-C-GCTC-TGG------A--T--G---G--ACCTTCATGT-GACA-----CA----A--G--CC-----------AATG-ATTGGT-TC-A---T-GGTAGA-GA--GGGA-GTAACT   7836
SMM.US.SIVSMMH9     ACAAATTAAGC-A-CAGCTC-AGG------A--T------G--ACTTTCATGT-GACA------C----A-----C--A--T--C--AATG-ATTGGT-TC-----T-GGTAGAAGY--TA-AA-A-G-T   7331
STM.US.STM          AAAAATAAGGA-AGT-GCTC-TGGG-----A--T-----GG-CACCTTCATGT-GACA-----CC-T--A------C-T--T--C--AATG-ATTGGT-C--A---T-G-TAGAAAAC-GA-----A-CA   7493
TANTALUS.UG.TAN1    AAGGCAGATT-GGCTAAGTAGACA-T-G--A--T------GCA-CT-ACATTT-GC-C-----CCA---A--G--------T----C-C-TGATTGGT-TG-----T-GCTGAATAA-G-GTC--A-A-T   7511
GRIVET.ET.GRI-677   AAATATACATC-G-GAAGG...AT-T-G--A--T------TCA-CG-ATTTTT-G--C------CA---T-----------T-----GATGGACTGGT-TA-C---TA-CTAAACAA-CGA-CAGAAGAT   7243
VERVET.KE.VER-9063  AGAGAAAATA-A-CT---GAGACAGTTT--A--T------GCA-C--ATTTGT-G--------CCAG--A--G--C--------C--GATGGACTGGT-TC-A---TA-TTGAATAA-C-ATCAGTAGAT   7551
VERVET.KE.AGMTYO    AGAACACATA-ACCT----AGACA-T-G--A---------GCATCC-ACTTGT-G---------CA---A-----C--T--T-----GATGGATTGGT-T--A---TACTTAAATAA---A-CATGGGAT   7023
VERVET.KE.AGM155    AAACAAAATC---TT----AGACA-TTT--A---------GCA-C--AT-TAT-G--C--C---CA---------C-----------AATGGACTGGT-T--A---TA-CTGAATAA-TTA-CAGTGGAT   7531
VERVET.DE.AGM3      TGAAGAGATT-A-CT---GAGACT-TTT--A--T------GCA-C--ATTTAT-G---------CAG--------C-----T-----AATGGATTGGT-TC-A---TACCTGAATAA-CG--CAGTAGAT   7049
SABAEUS.SN.SAB1C    AAACAAGATT---CTA---AGACA-T-G--A--T------TC--A-TTCTT-TT----------CA---T-----C-----T-----AATGGATTGGT-CC-C--CTA-TTGAATAA---GTCAGTAGAC   7722
SUN.GA.SIVSUN       TAAA......CCA-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-CAAT   7753
MANDRIL.GA.MNDGB1   TCTTAAGATAGA---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-   7043
LHOEST.KE.hoest     ACCAATAGCTAA--T-ACTAGGCAT----CA---TT--C--CA-AG-T-TTA-TGCA--C--------A----A-G----TT------GT---TAGGA-A--CC-AGAATGGAACAA---A-A-T-AGAT   7685
LHOEST.CD.hoest447  ACCTATAGCAAA--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-AGAT   6610
LHOEST.CD.hoest485  ACCTATAGCAAA--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-AGAT   6607
LHOEST.CD.hoest524  ACCCATAGCAAA--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-AGAT   6613
RCM.NG.SIV-NGM      GAATATAACTA-A-GATC-GT--AT-----A--TGAT--GGCAAA-TGTTTTT-G--A-------AT--A------C-T--T--C--AATG-ATTGGT-TC-A--CTA-CTAAATAA---A-CAGAAG-C   8040
COL.__.SIV-CGU1     TATCACTCCAAGG--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

SYKES.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-T   7412
SYKES.KE.SYK173 Env __.__T__T__W__R__S__Q__P__G__G__D__L__E__V__R__T__H__W__F__Q__C__G__G__E__F__F__Y__C__N__V__S__K__L__F__A__N__I__T__N__G__N__A__S_  gp120
```

```
B.FR.HXB2 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__A__M__  gp120
B.FR.HXB2           AGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACA...........................ATCACCCTCCCATGCAGAATAAAACAAATTATAAACATGTGGCAGAAAGTAGGAAAAGCAATGT   7527

A.UG.U455           -A-GACATG---C-----GG----..........................................--A--T----A---T--------G-----------T----------G-------C---------   6957
C.ET.ETH2220        --C-G-AC-AAT-T---C................................................-----A----A----------------------------------GG-------CG--------   6903
D.CD.84ZR085        -----A-GGTTAAATG---CA-TC..........................................--A---A--------------------------------------G------------------   7048
F1.BE.VI850         ..................................................................-----T--------TC-G------------G--------------GG-------CG--------   6807
G.SE.SE6165         ---GAGA-T---A-T...................................................-----A--------T-AG------------G-G-GA----------GG--G---C---------   6933
H.CF.90CF056        TAC--ATC-AATGAC-CA--GGGAA-C-A--AC.................................--T--A--G---------------------G---------------G-------CG--------   6846
J.SE.SE7887         GAAG--AC-AATGAC-CA-GTNAC-C-ACT....................................-----AA-A-----T-A-------------G-G-GA--------A-GGAC----C-------C-   6822
K.CM.MP535C         ..................................................................-----A--------T--------G--G-----------------------G-----------A-   6690
01_AE.TH.CM240      -TGG-G-GGT-TAATG-C-CT.............................................-----A--T-------AG-----G---------------------GG--C----C---------   7080
02_AG.NG.IBNG       -C-G-CA-CCACA--GGGTCA-A---CACT....................................--A--T----A------------G------G----T------------------C---------   7036
03_AB.RU.KAL153-2   GAAGAGTT-AATAAC-C-G-AGGA--TAT-....................................G---A----------------------------------------G---------------G--   7523
04_cpx.CY.94CY032-3 -AC-T-AC-A-TA--G--TCT--AA-TTC--CC.................................-----A----A-------C--------T--G---GG---------G----G--GC---------   6903
N.CM.YBF30          GAGC--............................................................--T--T--G--T--T-------G---G---G----TT-----ACA-GG----------G---T-   7075
O.CM.MVP5180        TCCGGATGCCA-GAG-TC--AGGGAGCAATGAGAC--ATAAAAATGGT.....................--TA-A--T----AGT---G---GC-AG---GATCA---AT---G-G--AGTCGAG---C-   7583
O.BE.ANT70          -CC--CTGTA-TGTT-G---TGT-AGT......C-AGGTAACAATGGC.....................--T--A--T----A-C-G-G---GG-GG---GGTCA---ATA-GG-G-CAGTCG-G-C-C-   7565
CPZ.US.CPZUS        -A---C--CAACA-C-C--T-...................................................--A--TAAG--T-------G-------G----TCA----AT--G-----------G---C-   7523
CPZ.GA.CPZGAB       ...-T-AC-AATGGC-T--TA...................................................-TA--G-----------T-G---G--AG-T-GTTCA---AT--GG-------G--G---A-   7546
CPZ.CM.CAM3         -C--AG-G--ATAAC-CC-C-...................................................TT-T-------T-------GG------G----TCC----AT--G---G--G--G-G---T-   6934
CPZ.CM.CAM5         CAA-ACA-CT-CA-C-GA--T-ACAC-AATGACACA-ACTCAGACAATAGTACACTA.............--AA-AT----T-------GG------G----TCC----AT--G---G--G----G---C-   7246
CPZ.CD.CPZANT       CTC-TCAC-AACGG-GCCCT-...................................................-TAGCA-AT--------T--G--G--AG-T--TCAT---GGC-T---TTC-----GC--T-   6962

H2A.GW.ALI Env      _E__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__H__N__Y__A__P__C__H__I__K__Q__I__I__N__I__W__H__K__A__G__K__N__V__  gp120
H2A.GW.ALI          GAACAG.........................................................CAC-ATTATGCA--G---CAT-----G-----A--T--T--C-----C----C---G---AATG-A-   7993

H2A.SN.HIV2ST       -CACAG.........................................................CAC-ATTATG-G------CAT-----G-----A--T--T-CC-----C--G-----G---AATG-A-   7439
H2A.DE.BEN          C-GCGC............................................................-A-TAT......---CAT-----G--G--A--T--T-CC-----T--------G---AATG-A-   8003
H2B.GH.D205         -C-CGG.........................................................AGA-ATTATGCG-----TCAT--C-G---G--A--C----C------C--G--T--G---AAT--A-   7985
H2B.CI.EHO          -AG-GA............................................................-ATTATGCGT-----CAC--C-G---G--AG-C----C------C--GA-T----G-AATG---   7968
H2G.CI.ABT96        ........................AG-AAT-TGACTCTCCTAAAGCCCCAGGAACGACAAAGACGC-ATTATG-A--C---CAT--T-GG--G--A--T----CA-----C------------AATG---   7364
MAC.US.MM251        -CC....................................CAGAGGCCAAAGGAACGGCATAGAAGG-ATTA-G-G--G--TCAT--T-G------A--C----CT-----T--------C---AATG-T-   7930
SMM.US.SIVSMMH9     TC-........................A--T-GACA-GTCAAAACCAGAAAGAACGATACCAGAAA-ATTATG-G-----TCAT--TRG---G--A--C----C------C--------C---AATG---   7437
STM.US.STM          GAA.....................ATGA--GA-TGG-ACAAAAACAAAAAGGAACAACAAAAGAGG-ATTATG-A------CAC--C-G----G-A--T--T-CA-----T--------G---AATG-T-   7602
TANTALUS.UG.TAN1    G-A-GAA-T-TAGATGTAG-AGG-A-TAATT-CAC---TGGAAAGGATAAACCGTGCTACAAAAGG-CATATG----C---CAT--T-GGTC----G-C--TGAC---T-C-CGC-CA-T--GAA--CC-   7641
GRIVET.ET.GRI-677   GCAGAA...--TA-T----GG--CTGT-AC-AA-GG-AGCCAGGACCAGGACCATGTGTTCAGAGA-CTTATG-TG-C---CAT---CG----G-AG----TGAT---T-C-CT--CTCT---AAGG-A-   7370
VERVET.KE.VER-9063  GCA-A-C-TAATCA-TG--GT-A----AC--AAA-AGGGAATGCCCCAGGACCATGTGTACAAAGG-CTTATG-TG-T---CAC--C-G-TCTG-C-----TGAT---T-T-CCA--TC---GAA-G---   7681
VERVET.KE.AGMTYO    GCATACC-TAAT-TTTG--G--GCA--AAG...A-AGG-CACGCACCAGGACCATGTGTACAAAGG-CGTATG-TG-T-A-CAT--C-GGTCTG-A-----TGAT-CCT-T-CCC--TC---GAA--CT-   7150
VERVET.KE.AGM155    GC-GA-C-TAATCATTG---A-ACA-C-C-G-GA-AGGTCGAAGTCCAGGTCCCTGTGTACAGAGA-CTTATG-TG-C---CAT--CCG-TCTG-C-----TGAT---T-T-CTA--TC---GAA--CA-   7661
VERVET.DE.AGM3      CCGGACC-TAATC-GTG---TGG-ACGAAGG-AA-AGGTAAGGCACCAGGACCCTGTGCACAAAGA-CATATG-TG-T---CAT---CG-TCTG-C--T--TGAT---T-C-C-C--TC--GGAA--CC-   7179
SABAEUS.SN.SAB1C    CCA...............G--CA-A-CAACT---CA-A-AATAATACAAAGCCATGTTGGCAGAGA-C-TATG-T--C---CAC----GG---G-AG----TGAT---T-C-CCT-GTCT---AAG-CA-   7837
SUN.GA.SIVSUN       -AATGG.........................................................TATCCTTGGGC-AAT---CAC------AGTG-G---G-TGAT---GC-TC---G--C--GAAG--T-   7826
MANDRIL.GA.MNDGB1   -AGTGG.........................................................TATCCTTATGC-T----TCA---T-GG-----AG--G-TGAC---AT-C---------G-AAG--A-   7116
LHOEST.KE.hoest     -AATGG.........................................................TATCCTTGGGC-AAC---CAC------AGC--C--TG-TGAT---GCT-C-A-------GAAG--T-   7758
LHOEST.CD.hoest447  -AATGG.........................................................TATCC-TGGGC-AAT---CAC------TC-G-----G-TGAT---GCATCT-----G--GAA---T-   6683
LHOEST.CD.hoest485  -AATGG.........................................................TATCC-TGGGC-AAT---CAC------TC-G-C---G-TGAT---GCA-CT-----G--GAA---C-   6680
LHOEST.CD.hoest524  -AATGG.........................................................TATCC-TGGGC-AAT---CAC------TC-------G-TGAT---GCA-CT--T--G--GAA---T-   6686
RCM.NG.SIV-NGM      -A-..........................................AAAAATGAAGCTAGGCAAGCC--GTTTG-T--T--T-TC-CT---ATGG--G-T--TGAT---T-T-C---GTCT-G-AA-G-A-   8128
COL.__.SIV-CGU1     ..................................................................CCAGT-A-GAAC--T-CC--T-G-A-GC-GG-G--T-GT---GTC-CTCAT-CC-GGATCC---   6916

SYKES.KE.SYK173     -AG-A-A-T.............................................TATGCCAGTAACC-TCGT--GT----TGCC--TCGC-----A--T--TGAT---AGAT-T---A-----TTG--A-   7497
SYKES.KE.SYK173 Env _K__N__N__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__Y__A__S__N__L__R__L__S__C__A__I__R__Q__I__I__N__D__W__R__Y__V__R__K__L__I__  gp120
```

```
B.FR.HXB2 Env       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__I__F__R__P  gp120
B.FR.HXB2           ATGCCCCTCCCATCAGTGGA.........CAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAAT...............AGCAACAATGAGTCCGAGATCTTCAGACC   7633

A.UG.U455           --------------CAA---.........GT---A--G---GA------C---------A--A-----------------G----C-AAC...............--T-CAA-AAAT----C---------   7063
C.ET.ETH2220        -------------TGAA---.........AT---A-TG---AG---------C-----A--A---C-G---C-T-----A-CC--A............GAGCCAC-T-GCACAAAA-----A--------   7012
D.CD.84ZR085        -------------TGAA--G.........-------AT----------------------A-----G--------------GC-...............-AT--T-C-C--AAT--C-C---T-----   7154
F1.BE.VI850         --A--T-------TGCA---.........A-C----CC---AAC--------------TT-A-----G------------------..................G-G-G-A-CATT----C---------   6910
G.SE.SE6165         --------------GCA---.........A-C---GA----AAC----GC---------A--AA-------------------G---AAT...AACAATACA-AT-CA-G----AGT--------------   7051
H.CF.90CF056        --------------CAA---.........A-C----TG---GT----------------A--AA----G----TT--C-AG-----C...............GCGTCTGCA--AAATT-T-C---------   6952
J.SE.SE7887         -------------TGCA---.........A-C----CG--CA-----------------AT-A-----G---C-T-----A--A--CAGG...GGGAATGGC--TG-G----GCA-T----C---------   6940
K.CM.MP535C         -------------TGCA---.........AGT----AC---AGC---------------AA--A----G---------------...................-AT--T-C-C-TAAT----C---------   6793
01_AE.TH.CM240      ----T-----------------.......AG-----AT---GT----------------AA-A-----G---------------GT-AAT...............--T-CG--TAA-----C---------   7186
02_AG.NG.IBNG       --------------CAA---.........AT---A------GAT-----C---------A--A-----------------A--G---..................--T-G-ACAAAT--A-C------G--   7139
03_AB.RU.KAL153-2   --------------GCA---.........----------------------------A-------------------------C...............CAG-G-----TTA-------T--T---A-   7629
04_cpx.CY.94CY032-3 -C---T-C-----TGCA--G.........AGC----AC--CAGC---G----------AA-AA----G----------------C-......................---A-TA-T-------------   7003
N.CM.YBF30          ----A--A--A--TC-G---.........GTTC---AC---A-C--C--------T--A---G-TC--GA-TATAG------GCC-..................G---CCA--GAAACA--AG-ATAT--   7178
O.CM.MVP5180        ----A---------CCC--C.........A-CT-A-C----CAT--C--C--A--T--AA--A-TC--CAGTT----CAACCATGG...............-ATTC--CA-GTGAAA-T-CAC-T-----   7689
O.BE.ANT70          ----A----------AA--T.........A-TC-A-C----ATG-----C--A--T--AA--A-CC--CA--TG---AACACATGG............AAC----G---CA-CAAT-TA-CA--T-----   7674
CPZ.US.CPZUS        T------A--A----AA--T.........GTGC-A--T---AAC---------A--G--AA--A-TC-TGAC-T-AGCATAA-CGCAGTC......AATAAC...G-T-G-AG-AATATA-CAG-G-TG--   7635
CPZ.GA.CPZGAB       ----A--C--A-----A---.........A-C--C-CT--CAAT--T--C-----T--A--T--C-----TTC---CAC-CC-GTGACT.........AAC...----G--GTAA-CT--CA--T-----   7655
CPZ.CM.CAM3         -------A--A--T--A---.........GTT--A--T--CA-------CG----T---A-TA--C-TGAG-C--GGCA---C-TC............AAC-A--G--TCACCAA-ATC-CGC-ATAC--   7043
CPZ.CM.CAM5         -------A--A--A--A--G.........A-C--A-C---CAT---------A--T--A--C--CC-CGA-GC--GA--AAAATGGGAA.........AAC..........A-TA-TATA-CAG-GTAC--   7349
CPZ.CD.CPZANT       -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-AAG-GTC---   7059

H2A.GW.ALI Env      Y__L__P__P__R__E__G__.__.__.__E__L__T__C__N__S__T__V__T__S__L__I__A__N__I__D__T__D__G__.__.__.__.__.__.__.__.__.__N__Q__T__N__I__T  gp120
H2A.GW.ALI          --TTG-----T-GGGAA---.........G-GT-G-CC--CAAC----CAG-A--CA-CT--A-TGCT-AC-TT--CACG-A-GGC...........................AA-C---CAAAT-TTA-   8087

H2A.SN.HIV2ST       --TTG-----T-GGGAA---..........--GT-A-CC--CAAC--T-CAG-G--CA-CA-AA-TGCT-AC-TT--C--A--AG-G...........................AA-C---CAAAT-TTA-   7533
H2A.DE.BEN          --TTG-----T-GGGAA--G.........G-GT-GGCC---GA-----CAG-A--CA-CA-AA-TGCT-AC-TT--CATA-A---AAAT.....................CG-A-TC-T-C-AA--TTA-   8103
H2B.GH.D205         --TTG-----A-GAGAA--T.........G-GC-CTC----AAT--C-C-G-G--TA-C--CA--GCC-AC-TCA-CTCA-AC..........................A-CAGTACC-C-AAT-TTT-   8079
H2B.CI.EHO          --TTG-----A-GAGAG--T.........G--C-CTCC---AAT--C-C-G----CA-C--CA--GCC-AC-TT--CT-GATAG--AAG.........................AA-CTT-CTAAT-TTA-   8065
H2G.CI.ABT96        -CTTG--C----GAGAA--T.........G-TC-G-CG---AAT----CAG-C--CA-C--AA--GCG-AT-T---A-AGAA--GGTCTAAC......................AA-C-T-CAAAT-TCAT   7464
MAC.US.MM251        --TTG-----A-GAGAG---.........G-CC-C-CG---AAC--C-CAG-G--CA-T--CA--GC--AC-T----T-GAC-G--GGA.........................AA-C-A-CTAGT-TCA-   8027
SMM.US.SIVSMMH9     R-TTG-----T-GGGAA---.........G-CC-G-C----AAC--C-C-G-A--TA-C--CA--GC-GA--T----T-GATC---GGC.........................AATA---C-AAT-TCA-   7534
STM.US.STM          --TTG-----A-GACAA---.........G-CC---CT---AAT--C-C-G-G--CA-TA-AA--GC--AC-TT--CT-GAC---CAAT........................AAT----CTAA--TCA-   7699
TANTALUS.UG.TAN1    ----A--A--A-GGGAA---..........--CT-AGA---CA------CAG-G---TCTA--A-GG--T--CTG---TA-AAC-GCAAA........................GAAAG--CTAATGTGA-   7738
GRIVET.ET.GRI-677   ----T--A--A-GGGAA--T..........--TT-GGAG---AAC---TCAG-C--G-CA--ATACG-GG---T----TA-AAC--CAAG.....................TCTGG-CCA--AAATGTGA-   7470
VERVET.KE.VER-9063  ----T--A--A-GGGAA---..........--TC-GGA---CA----C-CGG-C--T--TA--AC-G--GA-CT-A--TA-AAC---AAC.......................-ATAGA-C-AATGTGA-   7778
VERVET.KE.AGMTYO    ----T--G--A-GAGAA---..........--TT-GCA---CAGG--C-CAG-C--T---A--AC-G-TGAGCTTA--TA-AA--G-AAA........................AA-AGA-C-AA-GTGA-   7247
VERVET.KE.AGM155    ----T--A--A-GAGAA---..........--TT-GCAG--CA-G--C-CAG----T---A--AC-G--GAGCT-A-CTA-AA---CCAG........................AA-AG--CAAATGT-A-   7758
VERVET.DE.AGM3      ----A--G--A-GAGAA--G..........--CT-GCA---CA----C-CGG-A--G--TA--TC-G-GGAGCT-A--TACAA--G-AAG........................AA-AG--CTAATGT-A-   7276
SABAEUS.SN.SAB1C    -C----------GAGAA--C..........--CT-GGAG---AAC--G-CAGCC----C----TATG--GA-TTGA-CTA-AAC-G-AAA........................AA-AG--C-AATGTGA-   7934
SUN.GA.SIVSUN       -CTTG--C--A-CATCA--GTTTAATAACAG---A------G--AAC-GAG-G----AAGCATGG--C---CTG--AA-G-TGG-AGACTGGAAGGTAAATG-ATCA--CATC--A-TTG-AGC-TTC-A   7956
MANDRIL.GA.MNDGB1   --TTA--A--T-CATCA---TTTAATAAT--C--A--G---A--CAT-GGG-A----AAA--TAC--TGA--TGC-AAAGATAG--AGTAATGAAACAAAA............ATGC-A--TAAATTCTT   7234
LHOEST.KE.hoest     --CTA--A--T-CTTCA--TTTCAACAATAG---C------A-TCAT-GAG-A----AAA--T-C--TGAG-TG--AAAAT-GG-ACCACATGAGGACTTAG-AGG---CCTAAGT...--AAAGTTCTT   7885
LHOEST.CD.hoest447  --CT------A-CATCA--CTTTAACAACAG---A------A-TCAT-GAG-A----AAA--TGG--TCA--TG--AAAGTT-G-ACCAGAAGAAAACCTGG-AGGA---TT---T...G--AGATTCTT   6810
LHOEST.CD.hoest485  --AT------A-CATCA--CTTTAATAACAG---A------A-TCAT-GAG-A----AAA--TGG--TGA-CTG--AAAGT-GG-GCCAGAGGAAAACCTGG-GGGA---TT---T...G--AGATTCTT   6807
LHOEST.CD.hoest524  --CTT-----A-CATCA---TTTAACAATAG---A------A-TCAC-GAG-G----AAA--TGG--TGA--TG--AAAAT-GG-ACCACATGAAGACTTAG-AGGA---CT---T...--TAAGTTTTT   6813
RCM.NG.SIV-NGM      -CA-T--A--A-GGCCA-AT.........GCCT-G-----CAACG-C-C-G-G---TACT--T--GC-GAT-TT--CTA-ACAG--AAT.........................ATGACA-CAAATGTCA-   8225
COL.__.SIV-CGU1     ---GG--G--TCCAG-A--T..........--TC-ACA----AAC-GGG-G-AACAGCCAG-TA-TGCGTTC-TG-GCACCATAG-AGGA.........GAC-ATG-TGG-A-TGGGTGTGCT-ATCC-G-   7028

SYKES.KE.SYK173     -CCTG--C--T-CGGCA--G..........--T----A----A-T--C---G-C--T-CAG-AT-GACGGAC-T---ATACTACCCAGGA........................--TAC-T-GAA-TTTA-   7594
SYKES.KE.SYK173 Env Y__L__P__P__T__A__G__.__.__.__H__I__K__C__T__S__N__V__T__A__V__L__T__D__I__E__Y__Y__P__G__.__.__.__.__.__.__.__.__.__S__T__L__N__F__T  gp120
```

```
                                              Rev responsive element secondary structure in RNA, see Charpentier, J Mol Evol 266:950-956(1997)
                                              see HIV-1 complete genome annotation in Compendium p. 318–325 for stem-loop positions

                                                                                                 |-> Rev responsive element start
B.FR.HXB2 Env       __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__.__.__.__.__.__R_  gp120
B.FR.HXB2           TGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAG...............AGA   7748

A.UG.U455           -----------------------------A---------------G------------------------C------------------G-------------------G--...............---   7178
C.ET.ETH2220        --A---------------------------------------------------G---G-----A-G---C----------------T---C----A--G---------G--...............---   7127
D.CD.84ZR085        ------------------A-----C---------------------------------------------C--------------------------------------G-A...............---   7269
F1.BE.VI850         --AG------A------AA------------------C-------G------------G-------------------------------------A---CA----------...............---   7025
G.SE.SE6165         ---------------------------------------------G------AC---------CA--T--C------------------G-----G-------------G--...............---   7166
H.CF.90CF056        ---------------------------------------G------------------------------C-G---A--------------A---G-------------G--...............---   7067
J.SE.SE7887         -AC-------A------A----------------------------------------G-G-----G---C-------------------------A------------G--...............---   7055
K.CM.MP535C         ------------C------------------------C-G------------------C-G---------C-----A------------G---C-G----------------...............---   6908
01_AE.TH.CM240      ----------A-C--A-A----------------------------------------C-----------C-----A------------G-------------------G--...............---   7301
02_AG.NG.IBNG       ---------------------------------------------G------------------------C----T------------CGT-----A------------G--...............---   7254
03_AB.RU.KAL153-2   ---------------------------------------------------------T-------------------------------G----------------------...............---   7744
04_cpx.CY.94CY032-3 ------------C------------------------C----------------------G---------A---------------AT-------G----C-----------...............---   7118
N.CM.YBF30          CTC-------A-C---GTTA-TCTC------CAA--G--G-----G--C--------T-GC--A-----CA-------------AGGT--A--T--A---C-CACA---AGT...............---   7293
O.CM.MVP5180        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-CATTCACACCCCTCAC--G   7819
O.BE.ANT70          AAT---G-----C----AA--T-TA-------C------G-TC--C--C----------GGG-AA----T--TA-T--G--------ACGTATTGCA--GCC---CA-AAGCACTAGAACTCAT...---   7801
CPZ.US.CPZUS        -AC--------------C--CTTTA----A--A-------C----G-----G--G--C-GC--A-----TA-------G-----AGGT-----C--A--GCATACA---A-A...............---   7750
CPZ.GA.CPZGAB       -AC-------A-C----AA--T-T------------GC-C-----G-----------TCG---A--G--T---TC---G-----A--A--A----GA--GCATACA---GCAAGACAGAAAGAC...---   7782
CPZ.CM.CAM3         -AC-------A-C---GTA---TTG-----GCT------GC-------C--G-----T-GC--A-----CA-------------A-GT--A--T--A--GCACACA--ATCA...............---   7158
CPZ.CM.CAM5         -TC---G-----A-----A---TTG-----GCT-------A-C-GG-----------T-GC--A-----CA-------------A--T-----T--A--GCACAC----GCA...............---   7464
CPZ.CD.CPZANT       --CT-C-A---TAGCA...---C-A------GCG--G--G-CC-GG--CC-G--G---G-G----RT--C--GTC------C--A--A-CA-GN--A--GCC--AAA-AA-ACAACACTCC......---   7180

H2A.GW.ALI Env      __F__S__A__E__V__A__E__L__Y__R__L__E__L__G__D__Y__K__L__V__E__I__T__P__I__G__F__A__P__T__S__E__R__R__Y__S__S__.__T__.__.__P__R__R_  gp120
H2A.GW.ALI          CTTTA-T-C---GG--GCA--ACTA-ACC--TTA-----GGGGG-T------T-----G-G--AAC----A-T--CT-C-----T--ATCA-A--G----TACTCCTCT...ACT......CCAAGG--G   8208

H2A.SN.HIV2ST       CTTTA-T-C---GG--GCA--ACTA-ACC--TTA-----GGGGG-T------T-GA--G--G-AAC----A-T--CT-T-----T--ACCA-T---A---TACTCCTCT...GCT......CCAGTG--G   7654
H2A.DE.BEN          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...ACT......CCAGTG--G   8224
H2B.GH.D205         --TGA-T-C---AG--TCA--ACTG-AT---TTG---C-GGGGG-C------T-----G----AAC----A-T--CT-C-----T--AG-T-T--GA---TACTCCTCA...GTA......AAACCG--G   8200
H2B.CI.EHO          --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...GTG......ACACCG--G   8186
H2G.CI.ABT96        -TTCA-T-C---GG--GCA--ACTG-ATC--TTA-----GGGGG-T------T-----G----AACT---A-T--CT-T--------A--T-T-------TATTCCTCT...GTG......ACACCG-A-   7585
MAC.US.MM251        CATGA-T-C---GG--GCA--ACTG-ATC--TTG--G--GGGAG-T------T-----G-G--CACT--GA-T--CT-G--C-----AG-T-TG-----GTACACTACTGGTGGC......ACCTCA---   8151
SMM.US.SIVSMMH9     CATGA-T-C---GG--GCA--ACTG-ATC--TTG--G--GGGAG-T--C---T-----G-G---ACT---A-T--CT-C--C-----A-GT-T---A--GTACACCACAACTGGT......GCCTCA---   7658
STM.US.STM          --C-A-T-C---GG--GCA--ACTG-ATC--TTG-----GGGAG-T--C---T-----G----AACT---A-T--CT-G--C--A--G--T-TG------TACACAACT...AGT......ACCTCA---   7820
TANTALUS.UG.TAN1    ATTGAC--CTA--T--GA-A---TA---GCTTA------GGGA-G------GC-CA--G-----------A-C--TT-C--G--A--AG--AT--GA---TAT--T-G--CAACT............C--   7856
GRIVET.ET.GRI-677   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-CCACA............---   7588
VERVET.KE.VER-9063  ATT-A-TCCCC-G--AGAA-GA-TC---GC-GCA-----GGGACG-------T-G---G----CAC----A-C---T-T--------G--T-T--GAC-TTATAC--GA...GGA......CAAGAC---   7899
VERVET.KE.AGMTYO    ACT-A-TCCCC-G--AGAATCT-TC---GCGGC------GGGC-G---C---T----GG----CAC----A-T--CT-T--------AG-A-T--G-C-TTATAC--GA...GGA......CATGAG--G   7368
VERVET.KE.AGM155    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...GGC......CAAGAG--G   7879
VERVET.DE.AGM3      ATT-A-TCCCC-G--AGAAAC--TC---GC-GCA-----GGGC-GG--C---T-----G-----AC----A-T--CT-C--------AG-A-T--GA--GTATAC--GA...GGT......CATGAC---   7397
SABAEUS.SN.SAB1C    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............CCCGAA---   8052
SUN.GA.SIVSUN       GCC-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...............AGG-AG   8071
MANDRIL.GA.MNDGB1   GCCTCCCA-T--A-CCTCCA-TC-A-TTGTTGC-T-TGG-GC-C-T------T----C-----AATG---A-T--CA-------T--AG-T-TG--A---CACACTT-A-CTGAA......CATCAT-A-   7358
LHOEST.KE.hoest     GCCTCC-TC-TGGGAA-CTA-TC-A-TTGTGGCA---GGG-C----------T-GA-----T-GA-T---A-T--GT-T--T--T--AG-T-AGC-C--GTAC-CACCC............AGGGGA--G   8003
LHOEST.CD.hoest447  GCCTCCCTCCTGGGAA-CAA-TC-G-TTGTGGCAAC-GG-GC-C-C-------T--A---GGC--AG----A-T--CT-T--T--A--AG-T-TGC-TC-GTAC-CACCC............AGTAGT-AG   6928
LHOEST.CD.hoest485  GCCTCCCTC-TGGGA--CAA-TC-A-TTGTGGCAACGGG-GC-C-T------T--A---GG---AG----A-T--CT-T--T--A--AGCT-AGC-CCAGTAC-CACCC............AGTAGT-AG   6925
LHOEST.CD.hoest524  GCCTCCCTC-TGGGA--CAA-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............AGGGGA--G   6931
RCM.NG.SIV-NGM      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.....................   8334
COL.__.SIV-CGU1     A-C-CC-AATTT--A-CAT-CATTAA-T-CTTTG-----GGGA-G---C---T-G--------GAGGA--ACTACTTAT-TC--A--AG-TAT---A---TC----AATGTAAATTGGCATCATGGG---   7158

SYKES.KE.SYK173     -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.........AATACC--G   7715
SYKES.KE.SYK173 Env __P__T__A__N__V__E__D__V__W__R__A__D__L__F__N__Y__K__L__I__Q__I__K__P__I__G__F__A__P__T__D__Q__R__R__Y__E__L__P__.__.__.__N__T__R_  gp120
```

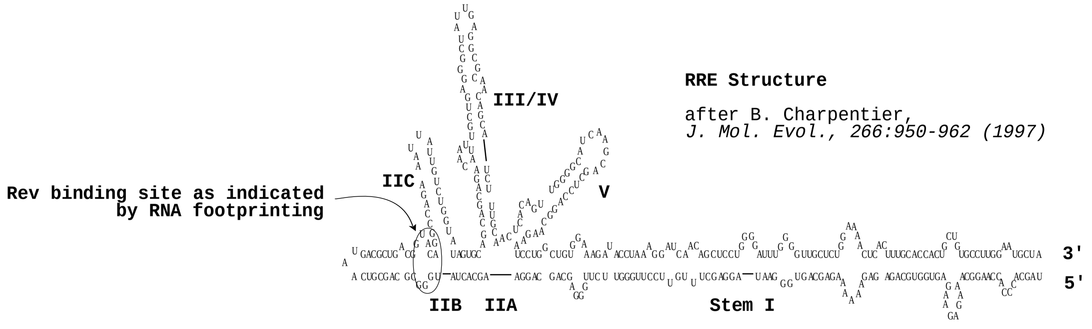
```
                          Rev responsive element secondary structure in RNA, see Charpentier, J Mol Evol 266:950-956(1997)
                                        see HIV-1 complete genome annotation in Compendium p. 318–325 for stem-loop positions
                      gp120 \/ gp41 cleavage after R__E__K__R
B.FR.HXB2 Env       _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__Q__L__L__S__G__I__V__   gp41
B.FR.HXB2           GAAAAAAGAGCA...GTGGGA...ATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGC   7872

A.UG.U455           ------------...--T---...C--------A-C--------------A--------T--------A--------G--G-----A-------------------------------------------   7302
C.ET.ETH2220        ------------GCA.........C-------------------------------------------------------A-----A--------------------------------G----------   7248
D.CD.84ZR085        ------------...A-A---...C--------A----------------------------------G-----------G------------------------------G------------------   7393
F1.BE.VI850         ------------...-C----...C----------------------TC-T--G-ACAGCA-GGAGCA---------G--G--G--A-----------------------------------A-------   7149
G.SE.SE6165         ------------...--T---...C-G------G-C--------------A--------------------------G--G-----A----------------T------------------C-------   7290
H.CF.90CF056        ------------...------...--G-------CT-----------------------------------------G--G-----A--------------------G----------------------   7191
J.SE.SE7887         ------------...------...---------G--------------------A----------T--------------G-----A----------------T---------G--------C-------   7179
K.CM.MP535C         ------------...--A---...T--------G-----T---A--------------------------------G--G-----A-------------------------------------------   7032
01_AE.TH.CM240      --------G---...------...---------A--A--T----------A--------------------------G--G-----A-------------------------------------------   7425
02_AG.NG.IBNG       -----------G...--T---...C-G------G-C--------------A--------------------------GCGG-----A--------------------------------C--C-------   7378
03_AB.RU.KAL153-2   --G---------...------...---------G-------------------------------------------G--G-----A-------------------------------------------   7868
04_cpx.CY.94CY032-3 ------T-----...------...-----G--CA----------------------------------------------G--------------------------------------C--C-------   7242
N.CM.YBF30          ------------GCCT-T---...C----T--GC----T--------TC-T-----------G-----------------G-----A------------------C-GAC------A-----G-------   7420
O.CM.MVP5180        ------------...--A---...T-G---ATGC-A---T-G---G-GC-AA-T--------T-----------------GG---CAG------------G-A--CACAGTG--C--AAG----------   7943
O.BE.ANT70          ------------...--A---...T-G---ATGC-A---T-G---G-TC-AA-T--------T-----------------GG---CA------G--------A--CACACT--GC--AAG----------   7925
CPZ.US.CPZUS        ------------GCCT-C---...C----T--AC-------G-----TC-T-----------------------------A---G-AGT----------------C-------------A--G-------   7877
CPZ.GA.CPZGAB       C-G---------GCCT-C--T...C-G-----AC--------------C----T--------------------------GG--G-A------------------C----G--GC----G----------   7909
CPZ.CM.CAM3         ------------GCCT-T---...C-G--T--GC----T-----------A-----------------------------A---G--GT----------------C-G-----------A--G-------   7285
CPZ.CM.CAM5         --G---------GCCT-T---...C-------AC-------C-----TC-A-----------------------------A---G--GT----------------C-G--C--G-----A--G-------   7591
CPZ.CD.CPZANT       C---------GCATT...---...-----GCTG--C---T-G--TC-TC-CA-T-----T--C--T--A-----------G-----AG---------C----A----GA-T--GY-CCA------T--A-   7304

H2A.GW.ALI Env      _N__K__R__G__V__F__.__V__L__G__.__.__F__L__A__F__L__A__T__A__G__S__A__M__G__T__A__A__L__T__L__S__A__Q__S__R__T__L__L__A__G__I__V__   gp41
H2A.GW.ALI          A-T-------GTGTGT-C...GTGC----G......---T-A--T--TC-C-CGA-------TTCTG-A------A-G--AG-TT-A------T-T-CT---T-TC-GACT------G-C--G-------   8329

H2A.SN.HIV2ST       A-T-------GTGTAT-C...GTGC----G......---T-A--T--TC-CACGA--------GCTG-A--------G--G--CT--------T---CT---T-TC-GACT------G-C--G-------   7775
H2A.DE.BEN          A-C-------GTGTGT-C...GTGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------GCGG--CC--------T-A-CC---T--C-GACT---C--G-C--G-------   8345
H2B.GH.D205         A-T-------GTGTAA--...GTGC-----......---T-G--A---C-T-C-ATG-----TTCTG-A---------A-G--TT--------T-A-CT---T--C-GACT---C--G-C--G-------   8321
H2B.CI.EHO          A-T-------GTGTAC-T...GTGC----G......--TT-G--A---C-T-CGA-G-----TTCTG-A--------G--G--CT-----T--T---CT---T--C-GACT---C--G----G-------   8307
H2G.CI.ABT96        A-T-------GTGTAT-T...GTGC----G......--TT-G--A--TC-T-C-ATG-----TTCTG-A------A----G--GT--------T---CT---T--C-GACT--G---G----G-------   7706
MAC.US.MM251        A-T-------GGGTCT-T...GTGC----G......---T-G--T--TC-C-C-A-G-----TTCTG-A--------G--G--GT----------C-CT---T--C-GACT------G----G-------   8272
SMM.US.SIVSMMH9     A-T-------GGGTCT-T...GTGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------G--G--GT--------T---CT---T--C-GACT--G---G----G-------   7779
STM.US.STM          ACT--G----GGGTCT-C...GTGC----G......---T-G--T--TC-C-CGA-------TTCTG-A-----------G--GC------------CT---T--C-GACT------A----G-------   7941
TANTALUS.UG.TAN1    -----G--G-TGCCCT-C...GTGT-G--G......-----A--C--T-----G-----T---GCTG-A-----T---A-AG-G-CAG--T-------C---T--CAG------C-TG-A--G---T---   7977
GRIVET.ET.GRI-677   --------G-TGCCAT-C...GTGC----G......--T--A--C-----------T--T----CTG-A-----------GG---CA--------A--C---T-TC-G--T--GC-TG----G---T---   7709
VERVET.KE.VER-9063  TC----C---TCCCGT-T...GTGC----G......-----A--C-----A-----T--T----CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG--T---C-TG----G---A---   8020
VERVET.KE.AGMTYO    C----G----TCCCGT-T...GTGC----G......---T-A--C--------G--T--T----CTG-A-----A-----GG-G-GC-GC--------C---T--C-G--T--GC-TG----G---C---   7489
VERVET.KE.AGM155    C-----C---TCCCGT-C...GTGC----G......-----A--C-----------T--T--G-CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG--T---C-TG----G---T---   8000
VERVET.DE.AGM3      AC---GC---TCCCGT-C...GTGC----G......-----A--C-----A-----T--T--G-CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG--T---C-TG----G---C---   7518
SABAEUS.SN.SAB1C    C-------G-TGCCAT-C...GTGC----G......-----A--T--------------T---GCTG-A-----T-----GG-G-CAG-C--------C---T-TCAG--G---C-CG----G---C---   8173
SUN.GA.SIVSUN       -----G--G---GCACCT-TTGCCC-----......GCTT-G-CAC-TC-CA-T--T--G--T-CTG-A------TT--TA--G-C--TA--A--T--C------CA-GC-G-C---CAA--G---T---   8195
MANDRIL.GA.MNDGB1   --G--G----G-GCA...-T-ATAC-T--T......A-------TC-GC-CTCGCTG------TC-G-G------T---TG--GG--G-A-----T--C--AT-TCAGTCT--GG--A----G------G   7479
LHOEST.KE.hoest     C---CC------GCGCCTCTGGCTT-----......GC-T-G--AC-TC-CA-T--T-----C-CTG-A------TT--TA--G-C--TA--A--T--C------CA-GC-G-C---CAA--G---T---   8127
LHOEST.CD.hoest447  C-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-GT-G--A-ACAA--G---T---   7052
LHOEST.CD.hoest485  C-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-GT-G--A-ACAA--G---T---   7049
LHOEST.CD.hoest524  C---C-------GCTCCACT-GCTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C------CA-GT-G--A-ACAA--G---T---   7055
RCM.NG.SIV-NGM      A--C---AGAG-GTACCTTTGGTGC-G--T......--T--A--T---C-CTCG--------T-CTG-A-----------GG-G-CAG-C-----A--C---T--C-G--T--GC-TG----G---T---   8458
COL.__.SIV-CGU1     C----G----GCATCT-T-C-TTCTCCATC.........T-A-CCC-----A-T-GG--C--TGCTG-A------T----G--TG--G--------A-C-----TCA-TCCC-TAATGGAA-GGCTAGTG   7279

SYKES.KE.SYK173     --------G---GCACCATTGGCAC-T--G......-----G--A--AC-CTCG--T--G--C-CTG-A-----G-GC--TG-G-CAG----A--AC-C---T-TCAGACT--G---G----G-------   7839
SYKES.KE.SYK173 Env _E__K__R__A__A__P__L__A__L__G__.__.__F__L__G__L__L__S__A__A__G__T__A__M__G__G__A__A__T__A__L__T__L__Q__S__Q__T__L__L__A__G__I__V__   gp41
```



RRE Structure

after B. Charpentier,
J. Mol. Evol., 266:950-962 (1997)

```
                        Rev responsive element secondary structure in RNA, see Charpentier, J Mol Evol 266:950-956(1997)
                             see  HIV-1 complete genome annotation in Compendium p. 318–325 for stem-loop positions
B.FR.HXB2 Env      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__R__Y__L__K__D__Q__Q__L   gp41
B.FR.HXB2          AGCAGCAGAACAATTTGCTG...AGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCT   7999

A.UG.U455          -A-------G----C-----...--------A-----T---------------A-------T-----------T--A---------------G-----------------------C--------------   7429
C.ET.ETH2220       -A-----A-G----------...-AA-----A---------------A-------G-----G-----------T------------A-----G--------A-A------C------G--------------   7375
D.CD.84ZR085       -A-----A------------...--------A-----------------------------G-----------T--A-------------------------A-----------G---------------   7520
F1.BE.VI850        -A-----A------------...-----------A-------------------G-----------------T--A---------------G--------------------------------------   7276
G.SE.SE6165        -A-----AGG---C------...--------A--------G------------------------------T--A---------------C----------A---------------------------   7417
H.CF.90CF056       -------A-G----------...--A-----AC------G-------A-------G-----G-----------T--A---------------G---C------------------G-------------   7318
J.SE.SE7887        -A-----A-G----------...-A------A-----------------------A-------T-----------T-----------------G-----------------------------------A--   7306
K.CM.MP535C        -A-----A-G----------...--A-----A--A--------A--------A--G-----G-----------T--A-------G-----------------A------------------------   7159
01_AE.TH.CM240     -A-----A-G----------...--------A--------G-------------------------------T--A---------------G--------------------------------A--T-   7552
02_AG.NG.IBNG      -A-------G----------...-AA-----A-----T-----A---------A-------G-----------T--A---------------G--------C---------------GA-----------   7505
03_AB.RU.KAL153-2  -A--------T---C-----...---------------------A-------------------------------------------G---G-------------G-----------------------   7995
04_cpx.CY.94CY032-3 ---------G----------...--------A-----T-----A---------AG-----G-----------T--A---------------G--------C-------C--------------------   7369
N.CM.YBF30         -A--------T-T-C--T--...--A--A--A-----------A---T-----------T--A-----------T--A------------A-G----T---A-A-----------T-G------G--AA-   7547
O.CM.MVP5180       -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----CCCTTA--C--A----G--A-G   8070
O.BE.ANT70         -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----CC-TA---C--A----G--A--   8052
CPZ.US.CPZUS       -A--------T---C----A...--A--A--A-----C---------T-A--------AT----T--------T--A-----T--A--C---G-A--T-----------------T--------G---A-   8004
CPZ.GA.CPZGAB      ----A-----T---C--T--...-AA--A--A-----T---------C--A-------AT--A-------AG-A--A--A--A--A--C---T-G--T-----A-----G-----GC--------G---A-   8036
CPZ.CM.CAM3        -A-----------CC-----...C-A--A--A------------------T-----------T----T---------T--A---------A-------G-T--T-----A------------C-GA-----G--AA-   7412
CPZ.CM.CAM5        -A-----------CC-----...C-A--A--A------------A---T-----------T----T-------------A---------A-------G-T--T-----A--------------------G---A-   7718
CPZ.CD.CPZANT      -A------GC----C-----...CAA--C--A---A-A--G--A------C-A--G---T-G--------AG-A--A--A---------------G--T--A--C--G-AG-------GA--------A--   7431

H2A.GW.ALI Env     Q__Q__Q__Q__Q__L__L__.__D__V__V__K__R__Q__Q__E__M__L__R__L__T__V__W__G__T__K__N__L__Q__A__R__V__T__A__I__E__K__Y__L__K__D__Q__A__R   gp41
H2A.GW.ALI         ----A---C-AC-GC--T--...GAC-TGG-CA--AGA------G-AA------G---G--C---------A-CG--AA-T-------------G--ACT---A-C--G-A----T-------C--GGC--G   8456

H2A.SN.HIV2ST      ----A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AA------G---G--C---------A-CA--AA-T-------------G--ACT---A-C--G-A----T-------C--GGC--A   7902
H2A.DE.BEN         ----A---C-AC-GC--T--...GAC-TAG-CA--AGA-----AG-AA------G---G--C---------A-CG--AA-C-------------G--ACT---A-C--G-AG---------C----GGCA-A   8472
H2B.GH.D205        --------C-AC-GCCTG--...GAC-TGG-CA-AAGA------G-A-------GG--G--C---------G-CG--AA-C-------------G--ACT--CA-C--G-A------------------GGCA-A   8448
H2B.CI.EHO         --------C-AC-GC-CG--...GAC-TGG-CA-AAGA-----AG-A-------GG--G--C---------G-CG--AA-C-------------G--ACT--CA-C--G-A---T--C--A--C--GGCA-A   8434
H2G.CI.ABT96       ----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-A---T---------C--GGCA-G   7833
MAC.US.MM251       ----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-AG---T-------C--GGC--A   8399
SMM.US.SIVSMMH9    ----A---C-AC-GC--T--...GAT-TGG-CA--AGA-----AG-AT------G---G--C---------A-CT---A-C------A-T---G--ACT---A-C--G-AG----------------GGC--C   7906
STM.US.STM         ----A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AT------G---G--C---------A-C----A-C------A-T---G--ACT---A-C--G-A-----------------GGCA-A   8068
TANTALUS.UG.TAN1   ----------G---C-----...GC---AG-----CA---G-----GA--C-AA-G-----CA-T-----TG-G--AA-C---A-T--CC-CG--ACT--CC-T----A-------G-------GACA-G   8104
GRIVET.ET.GRI-677  ----------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-AG------G-------GGCA-G   7836
VERVET.KE.VER-9063 ----------A---C-----...GC----G-G-----T--------AA-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-AG------G-------GGCA-G   8147
VERVET.KE.AGMTYO   ----------G---C-----...GC----G-G-----T--------GA-----A-G--G--CA-T-----TG-T--AA-C---A-T--CC-CG--ACA--CC-T--G-AG------G-------GGCA-G   7616
VERVET.KE.AGM155   ----------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-AG-----GGC------GGCA-G   8127
VERVET.DE.AGM3     ----------G---C-----...GC----G-G-----T--------GA-----A-G--G--CA-T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-AG------G----C--GGC--G   7645
SABAEUS.SN.SAB1C   ----------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-A-------G-------GGCA-G   8300
SUN.GA.SIVSUN      ----------GC-G------...GT-CT-G-----AA---G--AG-G--AC-AAG------CA-A-----AG-G---A-T-----A---C-CC-GACA--CC-A--GGAG--TG-GC-A--C---AGC--   8322
MANDRIL.GA.MNDGB1  -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-AT--TA-C-----C---GCTT-   7606
LHOEST.KE.hoest    ----------GC-G------...GT-CT-G-----AAA--G--AG-A-----AAG------CA-A-----AG-G---A-T-----A---C-CC-GACA--CC-T--GGAG--TG-G--AC-----GC---   8254
LHOEST.CD.hoest447 ----------GC-G------...GT-CT-G-----AAA--G--AG-G--C--AAG------CA-------AG-G---A-T-----A---C-CC-GACA---A-T--GGAG-----------C---ACT--   7179
LHOEST.CD.hoest485 ----------GC-G------...GT-CT-G-----AAA--G--AG-G--C--AAG------CA-A-----AG-G---A-T-----A---C-CC-GACA---A-T--GGAG-----------C---GCT--   7176
LHOEST.CD.hoest524 ----------GC-G------...GT-CT-G-----AAA--G--AG-G---C-AAG------CA-------AG-G---A-T-----A---C-CC-GACA---A-T--GGAG---T-G--A--C---GCA--   7182
RCM.NG.SIV-NGM     ----------G--G------...GACATAG----ACA---G---G-A---C-AA-G-----C--------G-CT--AA-C---------GC-TG--ACT---A-T--G-A-------GC---C---TCT--   8585
COL.__.SIV-CGU1    CCAGCAGC---CGAA--T--CTG-AACTGG-C---A----GTCTGCA---C-A--G-----G--------GG-T--AA-C-------T-C--G-TGCCA--A-A---G-T--TT-GG----A---GCAAA   7409

SYKES.KE.SYK173    --------C-A--G------...GA---AG-------A--G--A--C--C--AGGCT-A--T--------TG----AA-C--TA-C--CC-GC--ACA---T-A--G-CG--T--C-GA--C--GGCCA-   7966
SYKES.KE.SYK173 Env Q__Q__Q__Q__K__L__L__.__E__A__V__E__A__Q__Q__H__L__L__G__L__T__V__W__G__V__K__N__L__N__A__R__L__T__A__L__E__T__Y__L__R__D__Q__A__I   gp41
```

```
                                                    Rev responsive element end <-|
B.FR.HXB2 Env       __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__.__.__.__.__.__.__.__.__.__L__E__Q__I__W__N__H__T_   gp41
B.FR.HXB2           CCTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCT...........................CTGGAACAGATTTGGAATCACACG   8102

A.UG.U455           ---A--A--------C-----------G-----C---------A-------C-----CT--------------------...........................-A---GG-C--A------A---T-   7532
C.ET.ETH2220        ---A-----------C-----------------C-----------------------CT----------------G---...........................-AA--GG--------G--A---T-   7478
D.CD.84ZR085        ---A--A----------------------A-------------A----A--C-----CT---------------G----...........................G-A--GT-T------GGGA---T-   7623
F1.BE.VI850         ---A-----C-----C-----------------C---------AA------C-----CT--------------------...........................-A----G-----------A---T-   7379
G.SE.SE6165         ---A-----------C-----------------C---------AA------C-----C-T-------------------...........................TATA-TG-A------G--A---T-   7520
H.CF.90CF056        ---------------C-----------------C---------AA------------CT-------------------A...........................-A-AGTG-A--C---G-CA---T-   7421
J.SE.SE7887         ---A-----------C-----------------C---------AA------C---------------------------...........................TAT--GG-C------G-GA---T-   7409
K.CM.MP535C         ---A-----------C-----------------C---------AA------C-----CT----C------------...........................TCTTG----G----------CA---T-   7262
01_AE.TH.CM240      ---A--AC-------C-----------A----C-----------------C-----CT-C-C-----------G----...........................T-T---G----------CA---T-   7655
02_AG.NG.IBNG       ---A--A--------C-----------------C---------A----A--C-----CT----C------------A--...........................T-TA-TG-C--A---G--A---T-   7608
03_AB.RU.KAL153-2   ----------------------------------------------------------A-----------------C--...........................--A--TG-------A---A---T-   8098
04_cpx.CY.94CY032-3 ---A--A--------C-----------------C---------AA---A--------CT--------------------...........................TATA-TG-T--A---G-CA-T-T-   7472
N.CM.YBF30          ---AA-TC-A-----C-----A------ACA--A---TAT--CA---------------AG-C------C--C--TA-CTCT........................TAT--TACA--C------A-TTTA   7653
O.CM.MVP5180        ---AAACC-A-----C--TAAA--------A--C--TTA---AT-A--AAAA-----CA-ATCA---TCAGGA........................AGATATAATGAT--CAGT------G-CA--CTT   8176
O.BE.ANT70          ---AA-CC-A-----C--TAAA-----G--AG-C---TA---AT-A--AAAA------AGA-CA----TAGGA.................................AAC---AGC------G-CAC-TTA   8149
CPZ.US.CPZUS        ---A--TC-G-----C-----A------ACA-----TTAT--CA---------------A--CC--------C--C........................CTCTCCTAT--TGCT------GGCA-TCTA   8110
CPZ.GA.CPZGAB       T-----CC-G-----C-----------GGCTG-C--TTAT--CA-G-----------CAACTC----CC-GGG...AGC.....................AATTCCACA--TG-C------GGGA-TCTA   8142
CPZ.CM.CAM3         ---A--CC-A-----C-----A------GC------TTAT---A--------------AA--C-------GC---T........................ACATCCT-T--TG----C------A-TTTA   7518
CPZ.CM.CAM5         ---AA-TC-A-----C-----A------GC---C--TTAT--CA-G----------------C-------GC---C........................ACATCCTAT--TG----C------A-TTTA   7824
CPZ.CD.CPZANT       AT--A-CC-C--------TG---AC--GG-G-CC--TCA----A-G------------AA-TCC---GTA--CTTCA-GCAAACATGTGCAAAGAACAGCAGTGATA-AC--TGT------G-AA-T-T-   7561

H2A.GW.ALI Env      __L__N__S__W__G__C__A__F__R__Q__V__C__H__T__T__V__P__W__V__N__N__S__.__.__.__.__.__.__.__.__.__.__.__.__.__L__K__P__D__W__D__N__M_   gp41
H2A.GW.ALI          G--AAATTCA-----A--TG-ATTT-G--AAG-C---CA----A----A--A---GTAAA--AC-CC.......................................T-AA---CTGA----G-CA---T-   8547

H2A.SN.HIV2ST       A--AAATTCA-----A--TG-GTTT-G--AAG-C---CA----A----A--A---GTAAA-GACACC.......................................T-AACG-CTGA------CA---T-   7993
H2A.DE.BEN          G--AAATTCA-----A--TG-GTTT-G--AGG-C---CA----A----A--G---GTAAA-GAC-CT.......................................T-ATCG-CTGAC-----AA-T-T-   8563
H2B.GH.D205         A--AAATTCA-----A--TG--TTC-G--AGG-C--TCA---GA----A--A---CCAAA-GAAACC.......................................--CAC--CA-A-------A---T-   8539
H2B.CI.EHO          A--AAATTCA-----A--TG--TTC-G--AGG-C--TCA---GA----A--A---GTAAA-GAA-CC.......................................--TA-G-CAGAC------A---T-   8525
H2G.CI.ABT96        G--AAATTCA-----A--TG-GTTT-G--AGG-C---CA----A----A------G----CCT-G-TGC-------A-A...........................T----G-CACAA------A-T-T-   7936
MAC.US.MM251        G--AAATGC------A--TG-GTTT-G--AAG-C---CA----A----A--A---CCAAA-GCAA-T.......................................--AAC--CAGAC-----CA-TGAT   8490
SMM.US.SIVSMMH9     RN-AAATTCA-----A--TG--TTT-GG-AGG-C---CA----A-K--A--A---CCAAA-GACACA.......................................T--AC--CT-RC-----CA-T-T-   7997
STM.US.STM          G--AAATTC------A--TG-GTTT-G--AGG-C---CA---CA----A--A---CCAAA-GA--CT.......................................T---T--C-GA------CA-T-T-   8159
TANTALUS.UG.TAN1    G--AAATT-G-----A--TG-ATTC----AGG-G--TCA----A-G-----G----C-TTC-ACAAT..........................................AC--CAGAC---G-CA-T-T-   8192
GRIVET.ET.GRI-677   G--AAATTCA-----A--TG-GT-G----AAG-A--TCA---CA-A-----A-----GTA--A-AAC..........................................ACT-CT-AG---G-CA-T-T-   7924
VERVET.KE.VER-9063  AT--AACG-C-----G---G-AT-G----AGG-C--TCAT--CA-A--A--A---C-ATGGCAAAAC.......................................A--ACG-CC-A----C-AA---T-   8238
VERVET.KE.AGMTYO    A--AAACTCC-----G---G-AT-G----AAG-A--TCAT--CA-A---GAG---CCCTGG-CAAAT.......................................-G-ACT-C-GA----C-AA-T-T-   7707
VERVET.KE.AGM155    GT-AAACGC------G---G-GT-G----AAG-A--TCAT--AA-A--A--C----CGTGG-A-AAT..........................................AC--CAGAG------A-T-T-   8215
VERVET.DE.AGM3      GT--AATGC------G---G-AT-G--G-AAG-C--TCAT--AA-C--A--G---C-GTGG-A-AAT........................................AG-ACC-CTGA------A-T-T-   7736
SABAEUS.SN.SAB1C    GT--AAC--------A--TG--TTC-GG-AGG-G--TCA---GA-G--C-T------GTA--A-AAC..........................................ACT-CCGAC---G-AA---T-   8388
SUN.GA.SIVSUN       TT-A-CTTC------G--TCAAT-G----AAG-A---CAT---AA------C------TA--A-..........................................A-CACG-CC-A-----C-A-GGAC   8410
MANDRIL.GA.MNDGB1   G---TCTCAA-----G--T--AT-GGC--AGG-G--TCAT---T----AGAG----C-AA--CAA-C.......................................A-CACT-CA-A-----CATCAGAA   7697
LHOEST.KE.hoest     ---T-CCTC------G--TCAAT-G----AGG----TCAT---AA----GAG----CATA--A-..........................................A-CAC--CC-A-----CCA-GGAT   8342
LHOEST.CD.hoest447  G--T-CTTCC-----G--TCAAT-G----AAG-A--TCAT---AA---AGAA-----CTA--A-..........................................A-AACT-CT-A-----CAAGGGAT   7267
LHOEST.CD.hoest485  G--T-CTTCC-----G--TCAAT-G----AA--A--TCAT---AA----GAA------TA--A-..........................................A-AACT-CT-A-----CAAGGGAT   7264
LHOEST.CD.hoest524  A--T-CCTC------G--TCAGT-G----AAG-G--TCAT---AA---A--A------TA--A-..........................................G-TACT-CT-A-----CAAGGGAT   7270
RCM.NG.SIV-NGM      AT-AAAT-CA-TT--C--TG-AT-G-G--AAG-G--TCA---AA-G---GAG----TATA----..........................................-AAAC--CAGAG-----CA-ACA-   8673
COL.__.SIV-CGU1     G--T-CC-GCATT--G--TG-AAAC-TG-AA-----T-GA---AT------C-----CAAG-CC---....................................GGAGA---GG-TCCA---C-GA---T-   7503

SYKES.KE.SYK173     -T--TCA-A------A---G-ATTC--G-AG--C---CAT--A--A---A-G---G-GAAAGCA--TG-A--C--C--CAAC..................TTCTGT-CCA---CACAG-----AA-T-T-   8078
SYKES.KE.SYK173 Env __L__S__N__W__G__C__A__F__K__Q__I__C__H__T__A__V__T__W__E__K__A__C__G__N__N__S__N__.__.__.__.__.__.__F__C__P__K__P__Q__W__K__N__M_   gp41
```

```
B.FR.HXB2 Env       _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__E__L__D__K__W__A__S__L__   gp41
B.FR.HXB2           ACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGT   8232

A.UG.U455           ------C--C-A-----G-A--------GT-G-------G--A-----T-TCAAC-------G--------G--------G------------T--C-------C---G--C--G-------A-C---   7662
C.ET.ETH2220        ---------C-------T----------GT-----T---GA-A-----T--AAT--GC------TC--------------G--C--A------A----T---C-A-C---G--C-------A--A-C---   7608
D.CD.84ZR085        ---------C-------A---------G----------GG--------T--AA---G--------------T------------------A-----------A-----G--C--G-----G-------   7753
F1.BE.VI850         -----------------A-AG--G----G-------T---A-A-----T--AAA--G--------G--A-----------G-------------------------C------C------------C---   7509
G.SE.SE6165         --T-----A--A-----A--G-----C------------TA-CA----T--AG-C--C---------A--G--------G----------------C-------CC--G--CC-G------------   7650
H.CF.90CF056        --T--------A-----T-A-C------G----------GAGGA----T--AGG--GC------TC-------C------G--------------G--C-----A-C---G--C------------C---   7551
J.SE.SE7887         --------AC-A-----A--G-------------------G-AA-----T--AGTC------------G-A-----------G-----T------A---C-------C---G--C--G---A--A-----   7539
K.CM.MP535C         -----------------A-A---G---GG-------T--GA-AC----T-TAAG-------------A---C-----G----------------T-------C---G--C------------C---   7392
01_AE.TH.CM240      --A-----A--A-----G----------G-----------A-CA----T-TGAGA--C--AC---------G-----A--G--C-G-------A-G--T--G--A-----G----------------C---   7785
02_AG.NG.IBNG       --------AC-A-----G-A--------G----------GA-A-----T-TAATC----------------G----G---G------------------C--G----C---G--C--G---------C-A-   7738
03_AB.RU.KAL153-2   -----------------A-----------T---------G-T------T--AAT-----A-----------G-----A---------------------A-------C----------------------   8228
04_cpx.CY.94CY032-3 ------T--C-A-----T-A-------------------CAAA-----T-TGGG---C----------A--G--------G-----------G-----C-------CC--G--C--G--------CC---   7602
N.CM.YBF30          ------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--------G---C------AC---C---   7783
O.CM.MVP5180        --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-----GC-----G-------CTC-C-T-   8306
O.BE.ANT70          --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----G------G-------CTC-A-T-   8279
CPZ.US.CPZUS        --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--------G---C-----T-----C-A-   8240
CPZ.GA.CPZGAB       --A---CA-C-A-----T-A-TT-G-GTCT--C------G-GAA---TTTTGGTC-GT-A--G---G-A---TCA--A-----G--A------AGG--CC----------G---C-------C--C----   8272
CPZ.CM.CAM3         --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----------G---C-----T-------A-   7648
CPZ.CM.CAM5         --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--------G---C-----T-------A-   7954
CPZ.CD.CPZANT       --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---C----C------AGCTCA--A-   7691

H2A.GW.ALI Env      _T__W__Q__E__W__E__Q__Q__V__R__Y__L__E__A__N__I__S__E__Q__L__E__R__A__Q__I__Q__Q__E__K__N__T__Y__E__L__Q__K__L__N__S__W__D__V__F__   gp41
H2A.GW.ALI          --G---CAA--------ACA-C--G-CCGTT-CCTAGAGGCAAAT--CAGTGAACAGT-A---CGGG-A----TT---------------ACGT-T---C--CAAA-----A---GC----ATGT---TA   8677

H2A.SN.HIV2ST       --A---CA---A-----GCA-CG---CCG---CCTAGAGGCAAAT--CAGTGAAAGTT-A---C-GG-A----T---------------CATGT-T---C--CAAA-----A---GC----ATGT---TG   8123
H2A.DE.BEN          --A---CA---------G-A-C--G-CCG-T-CCTAGAGGCAAAT--CAGTCAAAGTT-A------G-C----TT--A-------------ATGT-T------CAAA-----A---GC----AT-T-C-TG   8693
H2B.GH.D205         --A---CAAC-------A-AGC--G-CC-TTTC-TGGATGCAAAT---ACGG-TC-GT-A-----GG-T--G-TA--A--G---------CATGT-T------CA-A--A--A---GC----ATGT---TG   8669
H2B.CI.EHO          --A---CAAC-A-----G--GC--G-CCG-TTC-TGGATGCAAAT---ACAAAA---C-A-----GG-A--G-TA--A-------------CATGT-T--G---CA-A-----A--C------AT-T---CA   8655
H2G.CI.ABT96        --A---CAA--A-----A-AGC-G--C---TT--TGGAGGA-AAC--CACACGGC--T-A------G-A----TT-----------A---ATGT-T--G---CA-A-----A---GC----ATGTA--TG   8066
MAC.US.MM251        --T---CAA--------GC--A-GG--G--TTC-TGGAGGAAAAT---ACAG--C-CC-A-----GG-A----TT--A-----G-----CATGT-T------CAAA-G--GA---GC----ATGTG--TG   8620
SMM.US.SIVSMMH9     -YT---CAA--------A-A-C-GG-G---TTCCTAGAGGCAAAT---ACTSAA-Y-T-G------G-A----TT---------M-----ATGT-T-----GCAAA-----A---GSK---ATNT---YG   8127
STM.US.STM          --A---CA---A-----G---A-GG--G--TT-CTTGAGGCAAAT---ACACAAC-GT-A------G-T---GTT--A--G---------ATGT-T--G--GCAAA--C--A---GC----ATGTA--TG   8289
TANTALUS.UG.TAN1    ------CA---A-----G--CC----A-CTGCC-TGGA-G-AAAT--TAGTA-TACTC---TCA--G-AT-TG-G--AG-GC-G--A---ATGG-TACT-ATCAAA-G----G-G-T---A-TTC--G--   8322
GRIVET.ET.GRI-677   --T---T----------G---C-------TGCC-TGGA-G--AAC---ACTCAAC--T-G------G-A-----T---G--TC-------CTGG-TCTG-ACCA-A----G---G-T---T--G-G--C-   8054
VERVET.KE.VER-9063  ------T----------A---C-G---GGAG-G-TGGA-G--AAC---ACAGAACAGT-A-T-A--G-AAG-G-A--AG----G--A---TT-G-T-CC-ATCAAAG---GACATCT---T---AC--C-   8368
VERVET.KE.AGMTYO    --T---T----------A---C----AGCTG---TGGA----AAC--TACGGGACA-T-A-TGA-GG-TAG-G-A---G-G---------CT-G-T-CC-ATCA-A-G---AC--GT---T--GA---C-   7837
VERVET.KE.AGM155    ------T----------A-A-C-G--AG-AGGA-TGGAGG--AAC---ACAAAACA-T-G---C-GG-AAGGG-A--AG-G---------TTGG-T-CT-ATCAAA-G--GTCAG-C---T-G-----T-   8345
VERVET.DE.AGM3      --T---C----A-----A---C-G--ATCGT---TGGA-G-TAAC---ACAA-ACA-T-A--G---G-CAG-GCA---G-G--G------TTGG-T-C--ACCAAA-----AG--GT---T--GA---C-   7866
SABAEUS.SN.SAB1C    --G---CAA--A-----G--GC----AG-A--G--TGA-GCTAAC--CAGTAGGA-TT-A--GC--G-A--TG-A---G--C--------TTGG-TAGT-ATCAAA----G-TGTCT---T--GAC--T-   8518
SUN.GA.SIVSUN       -----------A-----T--GC--G-A--A-TG--TGATGATAAT---ACAG-TC-TC-CC-G--GG-AT-TGTAACAG-GTT-G-----C--A-CA-G--CAAAC-GC--C-AG---TT...-AC--C-   8537
MANDRIL.GA.MNDGB1   --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-A-G----GAG-CCTTA--TC--G-G   7827
LHOEST.KE.hoest     -------GA--A-----ATC-A-GG-AGCA-T---TGAC-AGAAC--TACT--TC-GT-AC-----G-AT-C-C-ACAG-GCT-G-----C-GA-TA----TAA-A-G---C-AG---TT...-AC--T-   8469
LHOEST.CD.hoest447  --------T------------C--G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC-----G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA-A-GC--C-GG-G-TC...-A---T-   7394
LHOEST.CD.hoest485  --------T------------CG-G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC-----G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA-A-GC--C-GG-G-TT...-AC--T-   7391
LHOEST.CD.hoest524  --------A--A-----A---C--G--GGATC--TAGA-GCAAAC--CAC-A-AC-GT-GC----GG-TT-C-CAACAG-GTT-G-----AGGA-TA-C--TAAAA-G---C-AG-T-TT...-A---T-   7397
RCM.NG.SIV-NGM      --A---C-T--------A---A-T---TCA-GA-TAGA-G-AAAC--CTCAGTAGCTT-AC-G--TG-A---G-A-----T---CGT----TG--T--C---GAAA----GA--TC-----G-GACA--C   8803
COL.__.SIV-CGU1     -------AAC-----C-TGAGCG-G-A-GA---------GATA-----G-GG-AGACT-A-T---GG-AT-TG--TTA-----GG-A------A-GA----GGCA-------GGG-T---A-T-A--G--   7633

SYKES.KE.SYK173     --A---CACAGA-----ACA---GG-AG------TG--TGA-CAT--TG-TGGA--GC-CAG----G-A---G-A--A--G--G-GA----TG--T--TC--ACAA-G---C-GG------ACTCG----   8208
SYKES.KE.SYK173 Env _T__W__H__R__W__E__Q__E__V__D__N__L__T__D__H__I__D__G__L__L__R__E__A__Q__E__Q__Q__E__R__N__V__H__D__L__T__K__L__Q__E__W__D__S__L__   gp41
```

```
B.FR.HXB2 Env       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__I__V__N__R__V__R__Q__G__Y__S   gp41
B.FR.HXB2           GGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTC   8362

A.UG.U455           -A--------------CT-------------------T-G-C----TG----A--------------AA----A---------------A----G---------A-C--------------------C--   7792
C.ET.ETH2220        -----------------------------------------A----------------G-----TG--A--------------A----------G--------------------------------C--   7738
D.CD.84ZR085        ------------G-------C-A------------------A--------------------------A----------------------A--------------A-----------------G--C--   7883
F1.BE.VI850         -----------G-----T----C------------------A----------------G---------A----C--------------------------------A------------A-------C--   7639
G.SE.SE6165         -----------GG--------GA-----A------------A----T-----------------T--AA-------------------------G-----------A--------------------C--   7780
H.CF.90CF056        ---C-------G-----T--C--------------------A----------------------T--AA--------------A----------G-----------A--------------------C--   7681
J.SE.SE7887         -----------------T----C-----------C--C---A--------------A-------T--AA--------------A----------G---G-------A--------------------C--   7669
K.CM.MP535C         -----------G----------A-----A------------A--------------A-----------A---------------C---------G------G----A--------C-----------C--   7522
01_AE.TH.CM240      -----------G-----------------------------A----T-----------------T--AA--------------A----------G-----------A--------------------C--   7915
02_AG.NG.IBNG       -----------G----------------A----------G-A----T-----------------T--AA-------------------------G---A-----A-A--------------------C--   7868
03_AB.RU.KAL153-2   -----------G-----T----A------------------A-----------------------------------------A------------------------------G---------------   8358
04_cpx.CY.94CY032-3 ------------G---------A-----A------------A----T---------------------A----C---------A----------G-----------A--------------------C--   7732
N.CM.YBF30          ---GC------GGT--T-----A------------------A--GCT-------------C----A-T-----CA-------CA-AAG-ATA---A-AA-----A-AGCA--------------------   7913
O.CM.MVP5180        -----------G-------T--A---T--------------A--GCT-----C-----G----CAC-AA-----A-----G-TA--A-GATAA-----AA-C-------G-ACA--------------CA   8436
O.BE.ANT70          --------C--G-------T--A---T--------------A--GCA-----C----------CAC-A-----GG-G---G-TA-CA-GATA------AA---------A-ACA--------------CA   8409
CPZ.US.CPZUS        ----C------G-T--T--C--C------------------A----TT----AG------C-A----A-----AA-C-----T--AGG--TGA--T----AC----AGCA-A------------T-----   8370
CPZ.GA.CPZGAB       -----------G----------A------------------A-T---C-T---GC---------AA-CA----G--------TA--A-GA-A--TT-C--AG----C-GG--------------C--C--   8402
CPZ.CM.CAM3         -----------G-T--T--CCGC-----------C--------G--T-------------C------A-----CA----------AGGG---A-CA-A--AT-T--AGCA-AG--------------C--   7778
CPZ.CM.CAM5         ---G------C--T-----TC-G---T--------------A----TT-G--AG------C---T--AA-------T---C--A-AGGAAT---TA-G--AG-TA-AGCA-AG--------------C--   8084
CPZ.CD.CPZANT       -----------G--------C-A-----A------------A----T--T-------------CTA-T-----AC-------TT-GC---TAT--G--AG-TGCT-A-GA-AG---------------CA   7821

H2A.GW.ALI Env      T__N__W__L__D__L__T__A__W__V__K__Y__I__Q__Y__G__V__Y__I__I__V__G__I__V__A__L__R__I__V__I__Y__V__V__Q__M__L__S__R__L__R__K__G__Y__R   gp41
H2A.GW.ALI          CC--C---C--G--T----CGCC---G-CAA------TC---ATGGAG-TTAT---A---T---AA-A----C-C-T--------AA-ATA----G-GCAA--GT-A-G----C-C---A----C---AG   8807

H2A.SN.HIV2ST       -C--C------G-TT----CTCC---A-CAAA-----TC-G-ATGGAG-TTAT-------T---AA-AA---T------------AA-ATA----G-ACAA--GT-A-G----C----AA----C---AG   8253
H2A.DE.BEN          -C--C------G--T----CTCC---G-CAA------TC---ATGGAG-GCAT-------TG--AA-AA---C-----------CAA-CTA----G-GCAAT-GT-A-G----T----AA----C---AG   8823
H2B.GH.D205         -T---------G--C-C--CTCC---A-AAAA-----CC-T---GGAC--TAT-------C---G--A----TG-----------AG-ATACA-TG-GCAG--GC-AGCA---C-----A----C---AG   8799
H2B.CI.EHO          -T---------G--T-C--CTCC---A--GCA--C--C-GG---GGAT--TAT------AT---AA-A----TA----------CAA-ATACA-TA-ACAG--GC-AGCA--GC-----A----C---AG   8785
H2G.CI.ABT96        -C---------G--C----CTCC---G-TAAA--CG-TT-C---GG-C-TTATG----G-C---AG-AA---TG-----------CA-ATA----G-ACAA---C-AGGA---C-------A--G---AG   8196
MAC.US.MM251        -C---------G--C-TG-TTC----A-AAA-------C---ATGGA--TTATG----T-T---AG-AA--CTG-----------GA-CTA-A--G-ACAA--GC-AGC--AGT-A--------G---AG   8750
SMM.US.SIVSMMH9     -C------K--G--C-T--TTY----A-AAAA------C---ATGGTG--C-------TCT---AG-AA---A------------GA-ATA----G-GCAG--GT-AGC---GT-A--A-----T---AG   8257
STM.US.STM          -C---------G--C-T--CTC----G-AA-A--C---C---ATGGAG--TATT----TAT---A--A---ATG------G---CAA-CTA-A--A-GCAAT-GT-AGC---GC-----A-A--T---CG   8419
TANTALUS.UG.TAN1    ----CATC---G--G-CT--TCA---T-C----GG--T----GGGGAT-TTAT------AT---AC-TA--TTG--T--G--G-C--GGCT-A-TTGGGGATGCA-AGC---G-----------T--C-T   8452
GRIVET.ET.GRI-677   --TCA-----CTCAC-GT---C----T-AG-C---G-T---A--GGAT-TT-AG-GA---TGATTA-TC----A------T-T-CA-GG-TAT--TGGGGATGTA-C-GA-ATA--------------AA   8184
VERVET.KE.VER-9063  --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-TGCA-CGG---G-----------C-----   8498
VERVET.KE.AGMTYO    --TC------CG-TT-CT----A-----TAACAT-T-----A-GGGAT-TT-AG--A---T---AA-AA----G------T--C---ACA-A---TA-GGATGTA-AGTG--G---------------GT   7967
VERVET.KE.AGM155    --TC------CG-TT-TT----A------AACAT-T----GA--GG-T-TT--GC----AT----G-TA----G------T-GC---ACA-AT--TA-A--TGCA-AGC---G-----------T--C--   8475
VERVET.DE.AGM3      --TC------CG-TT-CT----G------AACAT-C-----A--GGAT-TT--GAT---CT---TA-TA----A------T-GC---A-A-A---TA----TGCA-AGC---G-----------T--C--   7996
SABAEUS.SN.SAB1C    --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--TC-T-GG-AGT----AA-A--G---G-   8648
SUN.GA.SIVSUN       ---G------GG--C-GT-TC-A---T-TCTA--C-------ATGCAG-CC----TA----CATTA-AA---CAGCT-----TC--AG-TT-A-CA-ACAACAGA-CT----GA-GT-T---------AG   8667
MANDRIL.GA.MNDGB1   CA-G-------G--T-T--TTGG---G-TCAA--CT------GGGGAG-TT-CT----GTT---AA-TA----A--------TT-G--A--CT--TGGAA--C-A-A-G---GT----------C---CG   7957
LHOEST.KE.hoest     ---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-AT--AGTA-A-A-GT-TAG---G---AG   8599
LHOEST.CD.hoest447  ----------AG----CTT-TCA---T-CCAA-----T--G-ATGCAG--C-A---A--AT---TA-TA---T-------G----AAGCTT-A-TG-GCAA-ATA-AGTA-A-A-GT-TAG-------AG   7524
LHOEST.CD.hoest485  ---G------GG-T--TTT-TCA---T-CCAA-----T--G-ATGCAG--C-A---A--AT---TA-TA---TG-----------AAGCTT-A-TG-GCAG-ATA-AGTA-A-A-GT-TAG-------AG   7521
LHOEST.CD.hoest524  ---G----A-GG--C-C--T-C----T-CCAA-----T--G-ATGCAG-GC-T---A-CAT---TA-AA--ATAC-------TC-AAGCTT-A-CA-ACAA-GT--AGTG-AGA-GT-TAG---C---AG   7527
RCM.NG.SIV-NGM      TC-G----C-A-----GGATTGG---T-AAA--------GGA--GGA--TT-C--TA-TTT---AA-AA----C---------A-C---TTAC-CTGGAGCTGT--T-G--AGC-----GCA--T--CA-   8933
COL.__.SIV-CGU1     TTTCAG-C---GG-C--TTC--CATCT-TAA----G--CTC-ATGCTGC-TATG-T--------T--AA----AC-T-----TA--A-G-TA---A-AG--TGCA-A-GAG---C-TTCAGA-TTA-GGG   7763

SYKES.KE.SYK173     ---GC------G--T-GT-G--A---T-C-TT---C----GA--GGAT-TTATG--A------CT--A--GCT------GC-----AG-TT-AGTG-AGGA--CA-A--G-ATT-GTT-GGA--C---GT   8338
SYKES.KE.SYK173 Env W__S__W__F__D__L__S__K__W__F__F__Y__L__K__I__G__F__Y__V__I__G__A__L__V__L__L__R__L__V__S__F__S__V__G__I__I__K__N__L__L__G__G__Y__V   gp41
```

```
                                            |-> Tat exon 2 start              Tat exon 2 ends in HXB2 <-|
B.RE.HXB2 Tat                               P__T__S__Q__P__.__.__R__G__D__P__T__G__.__.__.__.__P__K__E__$                                               exon2
                  Tat and Rev intron end \/ Tat and Rev exon start
                                          |-> Rev exon 2 start
B.RE.HXB2 Rev                             D__P__P__P__N__P__.__.__E__G__T__R__Q__A__.__.__.__.__R__R__N__R__R__R__R__W__R__E__R__Q__R__Q__I__H__S__I__  exon2
B.FR.HXB2 Env       __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__R__D__R__D__R__S__I__R__L_   gp41
B.FR.HXB2           ACCA......TTATCGTTTCAGACCCACCTCCCAACC......CCGAGGGGACCCGACAGG............CCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTA   8468

A.UG.U455           ---T......C-G-------------TTGC-----T-......--AGA---T-T--G----............--A-G-A----C------------------CA--G--A---------G-----C---   7898
C.ET.ETH2220        ----......----------------TTA-----CA-......------------------............-T--G------C------------------CA--G---------------A--C---   7844
D.CD.84ZR085        ---T......----------------T-------G--......------------------............--A---------------------------CA--------GT-C------------G   7989
F1.BE.VI850         ---T......--G--A--A-------TTA------G-......------------------............-----------C-------G----------CA--G--A----------G-G--C---   7745
G.SE.SE6165         ---T......--G--A----------TTAC--ACCA-......-A-----A----------............-----------C-------G------C---CA--G-----G------AG-G------   7886
H.CF.90CF056        ---T......--G--T----------TTG----G-A-......--AC--------------............----------C--------G------C---CA----------------G-GA-----   7787
J.SE.SE7887         ---T......--G--A----------TTA------A-......--A-C--A-G--------............----G------C-------G----------CA--G----AC------G---------   7775
K.CM.MP535C         ---T......--G--A----------TTA--------......T--------G-A-----A............-----------C------------------CA-----A-A-------AG--------   7628
01_AE.TH.CM240      ---T......--G--T--C-------CTTC--ATCAT......-A--A--A---------A............-----------C-------G------C---CA--G------------AG-G------   8021
02_AG.NG.IBNG       ---T......--G-----C-------TTAC--ACCA-......-A-----A---------A............------A----C-------G----------CA-----A----------G-G------   7974
03_AB.RU.KAL153-2   ----......-----A----------G-------G--......-A----------------............----------------------A--------------------C-------------   8464
04_cpx.CY.94CY032-3 ---T......--G--T--G-------TTA-------A...ACC-AAC------T-------............----G-----C---------------C---CA-------AG------------C---   7841
N.CM.YBF30          T--C......C-T-----G-------TTA-------A......G-A--------A------............--A----A--C-----G--GC-T-------CA--------G-------G-G------   8019
O.CM.MVP5180        ---C......C-C-----G----T--CTG-----CA-......-G-CA--A-G-A--A-C-............--A-G-A---C--G---------------AG-------GCC--AG-GG-CAGCC--G   8542
O.BE.ANT70          ---C......C-C-----A----T--C-AA--ATCA-......-AAGA--A-G-A-GA-C-............--A-G-A---C--G--G--G---------AGA--G---GCC---G-GG--A-CC-CG   8515
CPZ.US.CPZUS        T--C......C-C-----A-------TTT-------A......A-TC---A---A------............--A-------C--------GC-C----A-A-C-----AC-T------A-CGA-----   8476
CPZ.GA.CPZGAB       ---T......C-C-----G-------TTA-----GT-......-A-----A--AA-GA---............-T--G--A---C--C----G------C---CA-------AG------AG-GA-----   8508
CPZ.CM.CAM3         T--C......C-C-----G-------TTA----G--A......A-A----A---A------............--A-G-----C---------C-T-------CC--G--A--G------------C---   7884
CPZ.CM.CAM5         T--C......C-C-----G-------TTA----G--A......A-A----A---A------............-GA-G-----C---------C---------CC--G-----G---------CGA---G   8190
CPZ.CD.CPZANT       T--T......C-G--A------......A----T---CAAAAC-A-CA--AT--A--GCA-............--A----A----AG-------------AGA-A------GAT---G-GG-GGGCC--G   7927

                                                           |-> Tat exon 2 start
H2A.GW.ALI Tat                                             S__I__S__T__R__.__.__T__R__N__S__.__Q__P__E__E__K__Q__K__K__T__L__E__T__T__L__G__T__D__C__G  exon2
                                                            |-> Rev exon 2 start
H2A.GW.ALI Rev                                              P__Y__P__Q__G__.__.__P__G__T__A__.__N__Q__R__R__N__R__R__R__R__W__R__Q__R__W__G__Q__I__V__  exon2
H2A.GW.ALI Env      __P__V__F__S__S__P__P__G__Y__I__Q__Q__I__H__I__H__K__.__.__D__Q__E__Q__.__P__T__R__G__E__T__E__E__D__V__G__D__N__V__G__D__R__L__W_   gp41
H2A.GW.ALI          G--TGTTTTC-CC--CCC--CCGGTT--A---A-CAGATCCATATCCACAAG......GACCAGGAACAG...--AACCA--GG----AC-----AA-AC-TTG------AC-TTG-GGA--GATTG-GG   8928

H2A.SN.HIV2ST       G--TGTTTTC-CT--CCCC-CCG-TT--T---A-CAGATCCATATCCACAAG......GACCGGGAACAG...--A-CCA--GA----AC-----AA-AC-TTG--A----C-TTG--GA--A-T-G-GG   8374
H2A.DE.BEN          G--TGTTTTC-CT--CCCC-CCGGTT-T----A-CAGATCCATATCCACAAG......GACCGGGGACAG...--A-CCAACGA--G-AC-----AA-AC-TCG---G-GACAGTG-T-A-GACTTG-GG   8944
H2B.GH.D205         G---GTATTC-CC--CCC--CCT-TT-TACT-A-CAGATCCCTATCC-CAAG......GACCGGGGACAG...--A-CCAACGA----AC-----AA----GTG--A--GAC-GGGAC-A--GATCC-GG   8920
H2B.CI.EHO          G---GTATTC-CC--CCC--CCT-TT-TACT-A-CAGATCCCTATCC-CAAG......GACCGGGGACAG...--A-CCAACGA----AC-----AA----GTG--A---AC--GG-C-A--GATCT-GG   8906
H2G.CI.ABT96        G--TGTTTYT-CT--CCC--CCT-TT-TG-T-AGCAGATCCATATCC-CAAG......GACCAGGAACWG...--AACCAA-GA----AT------A-A-AGTG--A---A--GTG-C-A--GATCT-GG   8317
MAC.US.MM251        G---GTGTTC-CT--CCCA-CCT-TT-TT---AGTAGACTCATA-CCAACAG......GACCCGGCACTG...--AACCA--GA--GCA-------A-AC-GTG----AG-C-GTG-CAA--GCTCC-GG   8871
SMM.US.SIVSMMH9     G---GTGTTC-CT--CCC--CCG-TTRTG-TYAGCAGATCCCTATCYARACG......G-CCAGGAACTG...--AACCAA-GA--G---------A-AC-GTG---R-----GTG-CAA--GATCT-GG   8378
STM.US.STM          G--TGTGTTC-CT--CCC--CCT-TTGT-GT-AGCAGATCCCTATCCACAAG......G-CCAGGAACAG...--AACCAA-GA--G-AC-----AA----GTG---------GTG--AT--ACTCC-GG   8540
TANTALUS.UG.TAN1    T--T......--G--TCC-----T-A-TA---GCCTG..................-GG---GAA......CAG--A--CAACGC--G--G-----ACAA---TTCCAG---CAG----GA--AGTCGCCG   8552
GRIVET.ET.GRI-677   T--T......C-C...CCC----T---TA---AC-GTTCA...G--GAAC-G--A----ACGGAGGAGGG...-AA--CA--GGT-G----AGCA-CA-CAGC-A-TTG-T-AGATTGCAGGAAGAG-CC   8299
VERVET.KE.VER-9063  T--C......-----TCCA----T---TA---ACC-GTGG...AA-G-AC-G--A-G--ACGCAGGAGAG...--A----A-GGT-G---GAG--ACAAC--C--CTCAGATTCTT-GCAG-AAG---CT   8616
VERVET.KE.AGMTYO    T--T......C----TCCA----T---TA---ACCAAGTGGGGAA-G-AC-G--A----ACGCCGACGAG...--A-G--A-GGT-G---CA-CA-CA-GATC-A-TTAGAGTCTT-G-AG-AAGAC-CC   8088
VERVET.KE.AGM155    T--T......-----TCC-----T---TA---ATC-GTGG...AA-G-ACAG--A----ACGCAGGAGAG...--A----A-GGT-G-AG-AC---CAA-AGC-A-TCT-CGC-TTAGCAG-AAGA---T   8593
VERVET.DE.AGM3      T--T......C-T--TCCA----T---TA---ACC-GTGG...AA-G-ACAG--A----ACGCAGAAGGG...--A-G--A-GGT-G---CA-GC-CAAGA-C--CTC-GAGCCTT-GCAG-AAGA--CT   8114
SABAEUS.SN.SAB1C    ---TCTCTCC-CCCTCCCCTCTT-T--TTAA............-A--TCCATAT-C--CT-GACAAGGGACAG--A--CAA-GA------GC----CA-C-GTG-GCTA-ACAG-GACAG--GGTC--AC   8766
SUN.GA.SIVSUN       GGTG......--G--CCC-TCTG-TT-TG-TGA-CAGGATTGG-TCCA--A-.........ACCTGCCCCAAA---AC--AC-A-----------AA-AGACAGAGA-AGA-AGA-TC-A---CAAC---   8782
MANDRIL.GA.MNDGB1   ---TGTCTTT-C-.....................CAGGACTGC-A-CA-AAC......CTATACCGCAAACGG--A--CAACGG----------A-CAACAGCTT---ACT--GAGAGCA--ACTCCGAG   8060
LHOEST.KE.hoest     GGTG......C-C--CCCCTCTGTTT-TA-TGA-CAGGACTACAA-T---AGAAA--AGA-AACCAAGAACAG--A--CA--GA------GA----A-C---T-C---G-CTAT-...-A---CAAC---   8720
LHOEST.CD.hoest447  GGTT......C-TG-CCCCTCTG-TT-TG-TGA-CAGGACTGCAA-T---AGAAAAGAGA-AACCAAGAACAG--A--CAA-GA--G---GAT--AGAA---C-A---A-ACAT-...-A---CAAC---   7645
LHOEST.CD.hoest485  GGTT......C-TG-CCCCTCTG-TT-TG-TGA-CAGGACTACAA-T---AGAAA-GAGA-AACCCAGAACAG--A--CAA-GA--G---GAT--AGAA---C-A---AGACAT-...-A---CAAC---   7642
LHOEST.CD.hoest524  GGTT......C-CG-CCCCTCTG-TT-TG-TGA-CAGGACTTCAA-T---AGAA--GAGA-AGCCAAGAACAG--A--CAA-GA--G---GAC--AGAA---C-----A-ACAT-...-A---CAAC---   7648
RCM.NG.SIV-NGM      G--TTTGCTC-CTC-CCCA-CTTGTT-T-AT-AGCAGATCCCTATCC-CAAT-GG--GGAACAGCCAATCAAAGAA-G--A--A--G--GCAGT--A-A--GCG-T-GACACA-ATAT-A--ACT-GCAG   9063
COL.__.SIV-CGU1     -TTT......CAGCA-A-AGGC-GGACTAATGTGT-TTCACAGATCC-T--C-AA---C--GAGTGGGAACAG--A--CAACGC-------AT-CACA-----CA---A-ACAG---CAG--GATTT-GG   7887

SYKES.KE.SYK173     T--TATTTTGCAGAACCC-ACCCAG...........................GG-CGG-A-GACCCAGGCAAA----C--ACGA-------GCA-------C-----AG--CT--ACGT-TC-ACC--C   8441
SYKES.KE.SYK173 Env __P__I__L__Q__N__P__T__Q__.__.__.__.__.__.__.__.__.__G__R__K__D__P__G__K__P__A__D__E__E__E__G__S__G__D__R__E__G__L__N__V__S__T__F_   gp41
                                   |-> Tat exon 2 start
SYKES.KE.SYK173 Tat                T__L__P__R__.__.__.__.__.__.__.__.__.__A__G__R__T__Q__A__N__P__Q__T__K__K__K__A__V__E__T__E__K__D__S__T__S__L__P__S  exon2
                                 |-> Rev exon 2 start
SYKES.KE.SYK173 Rev              E__P__Y__P__G__.__.__.__.__.__.__.__.__.__P__E__G__P__R__Q__T__R__R__R__R__R__R__Q__W__R__Q__R__R__T__Q__R__L__Y__L__  exon2
```

```
B.FR.HXB2 Rev       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__L__D__C__N__E__D__C__G__T__S  exon2
B.FR.HXB2 Env       _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__L__I__V__T__R__I__V__E__L__   gp41
B.FR.HXB2           GTGAACGGATCCTTGGCACTTATC.....................TGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTC  8577

A.UG.U455           ----G-----T---A--GA--GC-.....................--------C-------A-----------------------A-----------CGC---------G----AGC-----------  8007
C.ET.ETH2220        ----------T---A---A--T-T.....................--------C---------------------T--------------------A---------C-G------CA----------  7953
D.CD.84ZR085        ----------T--CA---------.....................------------A---A------------------------------A---A----A----C-G------------------  8098
F1.BE.VI850         ----C-----T---A--T---GC-.....................------------A---A---------------------G--A------------CA-A--A----C-G------------CAGGG  7854
G.SE.SE6165         ----G-----T---AC-G------.....................--------C---------------------------------------CCA-----------G------CA----------  7995
H.CF.90CF056        ----------T---AC--G--G--.....................--------C--C----------CA-------------G--T-----------------A------GT-----C----------  7896
J.SE.SE7887         ----------T-C-A--T---GC-.....................--------C-------A---------------T--------A-----------CG---------C-G------C-----G-AC--  7884
K.CM.MP535C         ----G---C-T---A--G---GC-.....................----------------A-------------------G--AA------A-----A---A-----G---------CC-----AGGG  7737
01_AE.TH.CM240      ----G-----T---A------GC-.....................-----------A-----------------------------G------------AC---------C-G------CG----------  8130
02_AG.NG.IBNG       ----G-----T---A------GC-.....................---------------------------------------A------------A---------C-G------C-----------  8083
03_AB.RU.KAL153-2   ------A---T---A---------.....................--------C---A-----------T------T--------A-----------------------C-G------------------  8573
04_cpx.CY.94CY032-3 ----------T----C--------.....................--------C-------A--------------------G--A-------A----------A------G------C----------  7950
N.CM.YBF30          ----G-----T--CA--T---G--.....................-----G--C--C----A-----TGA-----CT-------------C-----C---------AC-G-G-----C-C-------C-  8128
O.CM.MVP5180        CCACCA----T----CA--AGT-G.....................-ACACG-----CA---CAA-AAT-T-G-GG-CT-------T------CA-----A-A-CAGGGA-CCG----C-GA-C--CTAC-  8651
O.BE.ANT70          CC-C-A--G-T----C----GT-G.....................-ACACG--C--CA-A-CAA-AATAT-G-GG-CT-------T------CA-----GCA-CAGGGA-CCA--A-G-GA-CAGCTA--  8624
CPZ.US.CPZUS        --C-G-----T---A------G--.....................-----A-----T----A---CCTGA-----CT---------A---GAG--------A--A---C-G-G-----CA--AC--A-C-  8585
CPZ.GA.CPZGAB       ---G-A--G-G----C--------.....................--------C--C--A-A----G-GA-T-GG--------GA------C-AG----GC--GC-A---GTG---ACAAC-CA-GACGT  8617
CPZ.CM.CAM3         ----G-----T---A--T------.....................-----A-----C----A---CCTG------T-T---------A--------------A-----AC-GGG-----CAT-----AAC-  7993
CPZ.CM.CAM5         --C-G---G-T---A--T---G--.....................-----A-----T----A---CCTGA-----TT----------A----T-----------CA--CC-GTG-----CA---C--A---  8299
CPZ.CD.CPZANT       CA-C----G-T---C-----CT-G.....................----TG--C---AC----A-AAT--AG-GG-T------GAT--GC---AC--GT-------ACC-TTG-GCAG-CC-CC---AC-  8036

                           Tat exon 2 end ->|
H2A.GW.ALI Tat      __P__G__R__S__H__I__Y__I__S__$                                                                                                    exon2
H2A.GW.ALI Rev      A__L__A__D__R__I__F__T__F__P__D__P__P__A__S__S__P__L__D__R__A__V__Q__H__L__Q__G__L__T__I__Q__D__L__P__D__.__.__.__.__.__.__.__.__.  exon2
H2A.GW.ALI Env      _P__W__P__I__A__Y__L__H__F__L__I__H__L__L__A__R__L__L__I__G__L__Y__S__I__C__R__D__L__L__S__R__I__S__P__.__.__.__.__.__.__.__.__.__   gp41
H2A.GW.ALI          CCCTGGCCGAT-GCATATT-ACATTTCCTGATCCACCTGCTAGCTC-CCT-TTGA-CG--CTGTACA--A---G---GG--TTA-TA-CC--GAT--CC-CG...........................  9030

H2A.SN.HIV2ST       CCCTGGCCGATAAGATATA-ACATTTCCTGATCCGCCAGCTGATTC-CCT-TTGAACA-ACTATACAA-A---G---GG--TTA-TA-CC--GAG---C-AG...........................  8476
H2A.DE.BEN          CCTTGGCC-ATAAACTATG-GCAGTTCCTGATCCACCTACTGACTC-CCT-TTGA-CG--CTATACAA-A---G---AG--TTA-TA-CC-AGA---CC-CG...........................  9046
H2B.GH.D205         CCCTGGCAGATAGAATACA--CATTTTCTACTGCGCCAGCTGAGGAACCT-TTAA-TT--CTATACAA-GG--G---AAC-TTA-TG----AGAC---C-AA...........................  9022
H2B.CI.EHO          CCCTGGCAGAT-GAATACA-CCA-TTCCCGATTCGCCAACTGAGGGACCT-TTGA-TT--CTATACA--GG--G---AAC-TTATTA-CA-AGAC---C-AA...........................  9008
H2G.CI.ABT96        CCTTGGCAGATAGAATATA--CATTTCCTGMTCCGCCAGCTAGGGAACCT-TTGACTT--CTATACA--AA--G---AGC-TTA-TA-CC--GAT-Y-C-AG...........................  8419
MAC.US.MM251        CCTTGGCAGATAGAATATA--CATTTCCTGATCCGCCAACTGATAC-CCT-TTGACTT--CTAT-CA--AA--G---AAC-TTG-TA-C-----CA-AC-AG...........................  8973
SMM.US.SIVSMMH9     CCTTRGCAGATAGAATATA--CATTTCCTGATCCGCCAGCTGATAC-CCT-TTGACTT--CTAT-CA--AGT-G---GG-TTGG-TA-----GAN--GC-AA...........................  8480
STM.US.STM          CCTTGGCA-ATAGAATATAC-CATTTCCTGATCCGCCAGCTAGTAC-CCT-TTGACTT--CTGTACAA-AA------GGCTTGT-TA-CC--GAT--AC-AG...........................  8642
TANTALUS.UG.TAN1    CCCTCA-TGAAAGAAT-TT-AT-GCCCAACAGAGGAGGGATCCAAGC---G--GAGAGCTT-G-G-CAG-AT--G-C-A-TTGGT------C-AT-AGCAGT-G-T-AT-G-GACTTTACCA-ATC--C-  8682
GRIVET.ET.GRI-677   TCA-CACCT--GAG-ATCAAC-A-TGGTGGCTCAACTTCAAGAGC--CAG-TTGAGAATA--GACT--GTGT-A--A-AT-TG--TGACCCTCCT-A--T--...........................  8401
VERVET.KE.VER-9063  -GA-CGA--CAAAA-AG-GGC-GTTGGAGCAAACGCTTGACCAAT---CT-TCGAC-TTC---ACC--G--A-A--A-AC-TT--TGACCCTCCTAA-T-AT...........................  8718
VERVET.KE.AGMTYO    AA--G-A-G-GTA--CAGT-G-C-GCCTGGCTGACGAGGCTCAAC..................ACT--G--A-A--A-AGTTG--TGACCCTCCT-A-T-AG...........................  8172
VERVET.KE.AGM155    -G-GGAC--GA-AA-AGGACC-GTTGGTGCAGGCAATTGACCAAT---TT-TTGACA-TC---A-C--GT-ACA--A-AG-TG--TGACCCTCCTTC-CAAG...........................  8695
VERVET.DE.AGM3      -GC-CA-C-GAG-G-AAGAGC-A-TGGTGCAAGCGATTGACCAAT---TG-TCGA-CA-C---A-C--G--A-A--A-AGTTG--TGACCCTCCTAG-T-AT...........................  8216
SABAEUS.SN.SAB1C    ACTTGGCAGAGAGAATTCT-G-GACACCTGTGTCACAGATTGATCACTTGGCTCAG-AATTTGACCA-TTGG---TCGACAAT-TT-AGC-ACCT-CAT-G-...........................  8868
SUN.GA.SIVSUN       -A-C-GTCCAAGAAA-A-TC-T-GCCACCACCTTGGACAGTAGAC--------G-CTTT---AGAC-CT--G--AGTGACT-TT-T-AA-T-GCTGAA-GCT-G-GGAA--GTCCTAGCACAAA-CA-CT  8912
MANDRIL.GA.MNDGB1   AACTTGAAGGAAGAAT-CT-A-A-AGATCCTTGATAGAGGACCTGACCAG-TT-GCCA--GAGTGACCAA-T-GGCTT-GG-TGAAAAATCTGAGAGC-GCA...........................  8162
LHOEST.KE.hoest     -A-C-GT-CAAGAAA-A-TC-TC-AGGCCACTGTGGAACGTGGAC---A----G-CTTTACAAGAC-CG--G--AGTGACT-T--TGAA-T-GCT-AAGGAAGG-GGAA--CTCCTCC-ATCACT-G-GT  8850
LHOEST.CD.hoest447  -AAC-GTACAAGAAA-A-TC-TC-ACTCCACCGTGGAACGTGGAC---AG---G-CTTTACAAGAC-CA--G--TGTGACT-T--TGAA-T-GAT-AAGGC---TGGAA-TCTTCTCC-CAGCCTTG-GT  7775
LHOEST.CD.hoest485  -AAC-GT-CAAGAAA-A-TC-TC-ATTCCACCGTGGAACGTGGAC---AG---G-CTTTACAGGAC-CA--G--TGTGACT-T--TGAA-T-GAT-AAGGCA--TGGAA-TCTTCTCC-CAGCCTTG-GT  7772
LHOEST.CD.hoest524  -A-C-GT-CAAGAAA-A-TC-T--AGGCCACCGTGGAACGTGGAC---A----G-CTTTACAGGAC-CA--G--TGTGACT-T--TGAA-T-GCTGAAGGCG-G-GGAA--CT-CTCC-CACCCTCG-GT  7778
RCM.NG.SIV-NGM      AGAGGATAT-TT-ATATC--G--ACTCCGCCCGATAGAGATCCTT--CAGGAGG--CTATCAAA-C---AG-AA-TCACT-TCAGTGA-CTACC-GAGC---CA-TCAA-CC-TTCTCGAA-CCTT----  9193
COL.__.SIV-CGU1     CGATTGATGGAAGA-TGGTGG-G-..............................................................................................................  7911

SYKES.KE.SYK173     AGC-GA-A---T---AGG-AATCTTTGGAAGCAGGACAGCAGCTT---AGGACAG-CT-C---AGC-T-AGA-CTCTGATTAGA-AGC-T-CTAT-ACCTGG...........................  8543
SYKES.KE.SYK173 Env _S__R__E__S__L__R__Q__S__L__E__A__G__Q__Q__L__W__R__T__V__C__S__S__F__R__S__L__I__R__Q__L__T__I__T__W__.__.__.__.__.__.__.__.__.__   gp41
SYKES.KE.SYK173 Rev Q__Q__R__I__F__E__A__I__F__G__S__R__T__A__A__L__E__D__S__L__Q__Q__L__Q__I__S__D__$                                                  exon2
                                                                                   end Rev exon 2 <-|
SYKES.KE.SYK173 Tat __A__E__N__L__$                                                                                                                   exon2
                  end Tat exon 2 <-|
```

```
B.FR.HXB2 Rev       __G__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__Q__G__V__G__.__.__.__.__.__.__.__.__.__.__.__.__.__S__P__Q__I__L__.__.__._  exon2
B.FR.HXB2 Env       L__G__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__R__G__W__.__.__.__.__.__.__.__.__.__.__.__.__E__A__L__K__Y__W__.__.__.   gp41
B.FR.HXB2           TGGGA...................................................CGCAGGGGGTGG....................................GAAGCCCTCAAATATTGG........   8612

A.UG.U455           -----..............................CGCAGCAGCCTCAAGGGACTGA-ACT-------....................................--G-G----------CT-........   8063
C.ET.ETH2220        -----..............................CGCAGCAGTCTCAAGGGACTA-AG---------....................................---A----T------CT-........   8009
D.CD.84ZR085        -----...................................................------------....................................---------------CT-........   8133
F1.BE.VI850         GCCTG...................................................A-G---------....................................-----T---------CT-........   7889
G.SE.SE6165         -----..............................CGCAGCAGCCTCAAGGGACTGA-ACT-------....................................----GT--T-------T-........   8051
H.CF.90CF056        ----G...................................................A-A------A--....................................---------------CTC........   7931
J.SE.SE7887         -----...................................................-T----------....................................--GAT----------CT-........   7919
K.CM.MP535C         GACTG...................................................A-GG----T---....................................---------------CT-........   7772
01_AE.TH.CM240      -----..............................CACAGCAGTCTCAAGGGACTGA-AC--------....................................----G----------CT-........   8186
02_AG.NG.IBNG       -----..............................CACAACTGTCTCAAGGGACTGA-ACT-------....................................-G-------------CT-........   8139
03_AB.RU.KAL153-2   -----...................................................------------....................................------------------........   8608
04_cpx.CY.94CY032-3 ----G...................................................ATA---------....................................-----------G---CT-........   7985
N.CM.YBF30          -----..............................CAGAGTCTCAGCAGGGGACTG-AACTACT-AAT....................................---CT-AGA-C-C-C-T-........   8184
O.CM.MVP5180        -----..............................CTGGGACTGTGGATCCTGGGA-AA-A-ACAATT....................................-----TTGT-G-CT---T........   8707
O.BE.ANT70          --A-G..............................CTTGGACTGTGGATCCTAGGG-AG-A-ATAATT....................................A-T-TTTG--G-AT---T........   8680
CPZ.US.CPZUS        -----..............................CAGAACATCAACAAGGGACTG...CAACT--T-....................................A-T-AA----G-GCAC-C...TGC..   8641
CPZ.GA.CPZGAB       -----..........................................CATCTAATA-TTCACA-CCTC....................................AG-CTG--ACGGG-GA-ACTGTGCCT   8669
CPZ.CM.CAM3         -----..............................CAGAGCCTCAACAAGGGACTG-AACA-CT-A--....................................A-CTT-AG--G---CCT-........   8049
CPZ.CM.CAM5         -----..............................CAGAGCATCAACAAGGGACTG-AACA-TT-AA-....................................--CTTTAGT-C---CCTA........   8355
CPZ.CD.CPZANT       -CT-C..............................AGAATTACTTTCAGACTGTGCAA-CATCT-GA-....................................A-CAAT----GCACCCTC........   8092


H2A.GW.ALI Rev      __.__.__.__.__.__.__P__P__T__D__L__P__E__S__S__E__S__A__D__N__N__Q__G__.__.__.__.__.__.__.__.__.__.__.__.__L__A__E__T__$           exon2
H2A.GW.ALI Env      .__.__.__.__.__.__I__L__Q__P__I__F__R__S__L__Q__R__A__L__T__T__I__R__.__.__.__.__.__.__.__.__.__.__.__.__D__W__L__R__L__K__.__.__.   gp41
H2A.GW.ALI          .................ATCCTCCAACCGATCTTCCGGAGTCTTCAGAGAGCGCTGACA-CAATCA--....................................--CTGG--G-G-CT-AAA........   9099

H2A.SN.HIV2ST       .................ACCCTCCAACTAATCTCCCAGAGTCTTCGGAGAGCATTGACAGCA-TCA-A....................................--CTGG--G-G--T-AAC........   8545
H2A.DE.BEN          .............................ACCCGCCGACTGATCTCCCAGAGTCTAACAGCAATCA--....................................--CTGG--G-G-CT-AA-........   9103
H2B.GH.D205         .........................................ATCCTCCACCAAATCTCA-CCAATCTC....................................C--C-A----GGCTCCC-........   9067
H2B.CI.EHO          ..................................................ACCCTC-AACCA-TACTC....................................C--C-G----GGCT-CCA........   9044
H2G.CI.ABT96        ..................................................ACTCTC-AACC-...CT-TTCCAAAGAATCTCCAGGACCCTCCAAGCAATTAGA---CAT---CGGCTAGAA........   8488
MAC.US.MM251        .................ATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTA--A----TTC-A....................................----T-----GGAC-GAA........   9042
SMM.US.SIVSMMH9     .................ATCCTCCAGCCAGTGCTCCAGAGCCTCTCAAGGACGCTG-AG--A-CCC-T....................................----T-A---G-GT-GA-........   8549
STM.US.STM          .................ACCCTCCACCCAACCTTCCAAAGGATCTCCAGGATACTG-AG--AATTA-A....................................----T-G---G-CTAG-A........   8711
TANTALUS.UG.TAN1    GCA-G.............................................AGCCTCACG-CTCTTCTT....................................C--CTG--T-G-C-GGA-........   8723
GRIVET.ET.GRI-677   ..............................................................................................................A---GGACAGCA........   8413
VERVET.KE.VER-9063  ..............................................................................................................--T-G-GGCGCT........   8730
VERVET.KE.AGMTYO    ..............................................................................................................--T-GGA-AGCT........   8184
VERVET.KE.AGM155    ..............................................................................................................--T-G-AGCGCC........   8707
VERVET.DE.AGM3      ..............................................................................................................--T-GGAGCGCT........   8228
SABAEUS.SN.SAB1C    ..................................................TGCCTC-AGGACATCCAA....................................C-GAGAAC--G-C-GCTA........   8904
SUN.GA.SIVSUN       ACCACCTGCTAAGTTTCTTGTGGCACCTACTTACGACTTCCTTCCATCATGGGCAA---CTTT--CA-....................................ACCCT-AGAGGT-GGCT-........   8998
MANDRIL.GA.MNDGB1   .....................................................ATTGAATAT----TC....................................CTC--AG--C--G-AGC-........   8195
LHOEST.KE.hoest     G-CAGAGCCTCAGTTGGCTGTGGCATTTGTTGATCCTTTTCTTCCAAAATGGGCAA---CTTT--CA-....................................ACCAG-AG--G--GGAT-........   8936
LHOEST.CD.hoest447  G-CAGTTCCTTTCGTGGCTAGGCCATTTAGTGATCCTTTTCTTCCAACATGGGCAA-ATCTTC--AAA....................................-CCTG-AG--G-CGGGT-........   7861
LHOEST.CD.hoest485  G-CAGTTCCTTTCGTGGCTAGGCCACTTAGTGATCCTTTTCTTCCAACATGGGCAA-ATCTTT--AAA....................................-CCAG-AG-GG--GGAT-........   7858
LHOEST.CD.hoest524  G-CA-TTCCTCACTTGGCTGGGACATATAGTGATCCTTTCTTACCACCATGGGCAA-ATCTTT--CAA....................................ACCCT-AG--G--TCCT-........   7864
RCM.NG.SIV-NGM      ATTTGCGGTGGATAGGAGCTAAGCTACAAGATGGGTGGCAA.........................................................................................   9234
COL.__.SIV-CGU1     .............................AGCTTGTTGCGTCATTGCAGCAATGGGAT-CTTACTGTC....................................CTGC-T-AAC-TCCGCCA........   7968

SYKES.KE.SYK173     ..................................................................................................................................   8543
SYKES.KE.SYK173 Env .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.    Env
```

```
                                                     Rev end <-|
B.FR.HXB2 Rev       _.__V__E__S__P__T__V__L__E__S__G__T__K__E__$                                                                         exon2
B.FR.HXB2 Env       __.__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__V__I__E__V__V__Q__.__G__A__C_   gp41
B.FR.HXB2           ....TGGAATCTCCTACAGTATTGGAGTCAGGAACTAAAGAATAGTGCTGTTAGCTTGCTCAATGCCACAGCCATAGCAGTAGCTGAGGGGACAGATAGGGTTATAGAAGTAGTACAA...GGAGCTTGT   8735

A.UG.U455           ....--------T--G-T-------G---G---------A-T-------A---CT-----TG-C--TGT---AG------------G-T---T----------------A---G----...AC-AT-G--   8186
C.ET.ETH2220        ....G-A-GC--TG-G---------G---T---G-----A--G------A---ATC----T---A-------A----T-----G---A--A--------AT--------T--A-----...A--AT---G   8132
D.CD.84ZR085        ....-----------G--A----------GA-----G------------A--------G-TG----A-----A---------------------------A-------CA-----AGG...A--------   8256
F1.BE.VI850         ....G--------ACG-G-------------------------------A-C---C--T-T---A-------A----T---------A-----------AA----------TT-G---...A-----G--   8012
G.SE.SE6165         ....--------T--GTT-------G-CAG-------------------A----T-----TG--A-AGT---A------------A-TT------------------------C----...A--------   8174
H.CF.90CF056        ....--------T-----A--C---G-A---------------------A--GAT-----T-CA--------A-----------------A------G--A------T-A---G----...A-------G   8054
J.SE.SE7887         ....GT---C--TG--TG-------G-G---------------------A----T-----T---A-------A--------------A--A--------AA-C------A---C----...A------T-   8042
K.CM.MP535C         ....-----C--TG----------------------------------CA----------T---A-------A-------------GA--A--------AA--------A---G----...A-----T-   7895
01_AE.TH.CM240      ....G-------T--GTTA------G-C-----------A-T-------A--TCT-----TG----T-----A-------C---G-G-T-------C----------------C----...--------G   8309
02_AG.NG.IBNG       ....--------TA--TCA--C---GT----------------------A---AT-----T---A-A-T---A----T-------A-CT----------A-C-------A---G----...A---TGG--   8262
03_AB.RU.KAL153-2   ....----------------------T--------------G-------A---ATC--A-AGG-A---T---A-------------G-T--------------------A---G----...A--TT----   8731
04_cpx.CY.94CY032-3 ....-----CT----G-T-------G-A-----G---------------A---AT---T-T---A-------A-----------------A---------A---------C------G...A--------   8108
N.CM.YBF30          ....---GGAA-A--TGCA------G-AA-A--GT---G-G--------A-C--------T---A-A-----T--T-T------A--A--A---------A--------T---C----...A--ATAG-A   8307
O.CM.MVP5180        ....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-CA-C--CTT--GTC-----...A--ATAG-A   8830
O.BE.ANT70          ....GCAGC-G-AAC---A--C---CTA--A---T-GC-----------ACA--------AG-CA-ACTT--AG-G--------CA-TT----T--CG-CA-A--C-C--GGA-----...A--ATAG-A   8803
CPZ.US.CPZUS        ....---GGGG-AA-CGCT--C---GCAAG---G-----AGT-------ACA-----AT-AG--A-A-----A--------------A--A--T-----AA--------T--AC-AG-...A-GCTC-T-   8764
CPZ.GA.CPZGAB       GCTGG--GGAA-TA-T--A------G-AA----G-----A-T-------A-A---C----TG----A-----T--T--------A--A--T--------AA-A-------CTT-T---...-TTA-ACT-   8796
CPZ.CM.CAM3         ....---GGAG-GA--ACC--C---G-AAG---G---C-A-C------CA-A-----AT-AG----A--------T-----------A--A-----C--AA-C--------T-C----...AT-ATAG--   8172
CPZ.CM.CAM5         ....AA-GGAG-A---ACA--C---G-AAGA--GA--C-A-T-------ACA-----AT-AG--A-A-----A--T---------A-T--A--------AA-------GT-T------...AT--TAG-G   8478
CPZ.CD.CPZANT       ......................----CAATAATC...-GA-C-GAGATCA---AGAACA-TG-CAGACTT--T--TTGG----GG--AAAA--------CA-AC-TCT--CTC-C---...ACTATAGTC   8194

                                         |-> Nef start
H2A.GW.ALI Nef                           M__G__A__S__G__S__K__K__R__S__G__P__L__Q__G__L__R__E__R__L__L__Q__T__P__G__E__T__C__G__G__Q__C__S__.__G__S__G     Nef
H2A.GW.ALI Env      __.__A__A__Y__L__Q__Y__G__C__E__W__I__Q__E__A__F__R__A__F__A__R__I__A__R__E__T__L__T__N__T__W__R__D__L__W__G__A__V__Q__.__W__V__G_   gp41
H2A.GW.ALI          ....GCAGCCTA---G------G--T-CG--TGGA-CC-AG-AGCGTTCCGGGC---TGCA-GGATTG-GAGAGAGA-TC-TA-AA-CACCTGGAGAGACT-GTGG-GG-C---G--G...T-G-TCG-G   9222

H2A.SN.HIV2ST       ....ACAGCCTA---G--A---G--G-CG--TGGA-CC-AG-AGCGTTCCGAGC---CGCG-GG--T--G-GAGAGA-TC-TA-AA-C-CCTGGAGAG-CT-CTGG-GGAC-C-GGG-...CA-AT-G-G   8668
H2A.DE.BEN          ....GC-GCC-AA--G--A---G--T-CG--TGGA-CC-AG-AGC-TTCCAAGCA--CGCG-GGA-T--GAGAGAGA-TC-T--G-GC-C-TGG-GAT--T-ATGG----C--CG-G-...C-CATCG-G   9226
H2B.GH.D205         ....GTTGCGTA---C------G---TCAGCTGGT-CC-AG-AGCGCTCCGAGCGGCAGCA-GG--T--G-GAGA-A-TC-T--GAGT-CCGG---G-CCT-ATGG--G-CTC-CAGG...A----GGCG   9190
H2B.CI.EHO          ....CCTGCATA---C-GA---G---TCAGCTGGT-CC-AG-AGCAATCCAAGCAGCAGC--GG--TG-G-GAGAGA-TC-T--GAGC-C-G-GAGG-CCTCGTGG-G---CC-CAGG...A----GGCA   9167
H2G.CI.ABT96        ....GCAGC-TATT-CAGC---G--TTC-GCTGG--CC-AG-AGCGTGCACCGCAGCAACT-GG--TG-GCAAGAGA-TC-TA-AAGCAC-TGGAGAGCCT-ATGGA--ACTC-GGG-...A-G-T-G-A   8611
MAC.US.MM251        ....CT--CCTA------A---G--T-GAGCT-TT-CC-TG-GGCG-TCCAAGC-GGCTGG-GAT-TG-GA-AGA-A-TC-T--G-GC-C-TGGAGAGACT-ATGG--GACTC-TAGG...A---G-G-A   9165
SMM.US.SIVSMMH9     ....ATA-CCTA----------G--T-G-GTT-C--CC-AG-AGCA--GCARR---G-TGG--ATTTGYRCGAGAGA-TC-T--AAGC-C-TGGAGAGACT-ATGG--GACTC-GGG-...A-G-T-G-A   8672
STM.US.STM          ....GCAGCCTATT--------G--T-CATCTGGA-CC-AG-AGCA--GCAAGCAGCATGG-GA--TG-G-GAGAGA-TC-T--GAGC-C-GGGAGAGACT-ATGG---ACTC-TGG-...A-G-T-G-G   8834
TANTALUS.UG.TAN1    ....--CC-ATA-A-T------G--T-G---C--T-C--AG-GG-A--A-CA--G-CTT-TG-G--TTTG--G-GC--T-CGCAAAGC-CC-GTCG--C-C-ATGGA-T-CTTGCAG-...TCC----A-   8846
GRIVET.ET.GRI-677   ....GT-GGATA---G------G--CTC---C----CC-AG-GGCA--AACAG-GC-TGCTC-A--TCTG--G-GG--T-CGAGG--A-CCTGG-GC--AC-GGGT-CTA-T--C-G-...TCC----A-   8536
VERVET.KE.VER-9063  ....---C-GTA-A-C--A---G--CT-GG------C--A-CCGCA--ACAAGAAGCA-GTCTC--TGTG--ACGCTTT-C-CAAA-C-CAGGTC-CCAAA-ATGGCTT-CTTGCAG-...TCCA---A-   8853
VERVET.KE.AGMTYO    ....-TTC-ATA---G--A---G--CTCGCA-----C--A-CCG-C--ACAAGAAA-A---C-AA-TCTG--AGGC-TT-C-CAAA-C-CATGTC-CCA-A-ATGGCTT-CTTGCAG-...TCC----A-   8307
VERVET.KE.AGM155    ....---C-GTA-T----A---G--CT-GG---G--C--AGCCGCA--ACAAGAAGCAG-T--CCATCTG--G-GCTTT-C-CGCA-C-C-G--C-CCA-A-ATGGCTT-CTTGCAG-...TCC----A-   8830
VERVET.DE.AGM3      ....-TCC-GTA-A----A---G--CT-GG------C--AGCCGCA--ACAAGAAGCAG-TGTC--TTTG--ACGCCTT-C-CAAA-C-C-GGCT-CCA-A-ATGGCTT-CTTGCAG-...TCC----A-   8351
SABAEUS.SN.SAB1C    ....ACAGC--AT--TG-A---G--T-G--A--G--C--AGCAGCA--A-CGTCAC-CACTGTG--TGTG-T-CA---T-CGA-AAGC-CC-GT---TCA--ACG-C-T-CTTGCAG-...TCCAT-GT-   9027
SUN.GA.SIVSUN       ....G-TTCG-A---TATTC-GGCT-CGAGCAGGA-C-G---CGCGTGCAGA-C-AGCAGGG-GAGAGT-T-ATCGT-TCAGAAG-CAA-A-GTAGA-CCT---GTCTC-GCAG-A--...T-GAGGCCC   9121
MANDRIL.GA.MNDGB1   ....ATC-GAAG-T-GGGT-CGCTACTCGTCAGCT-TGC-GCA-T-AG-TGG--GGCCGGTT-CA--GG-TGGC--CA--C-AA-AGAATCTTCACAG--TGC-G-----C-A-CATT...-CT--GG-G   8318
LHOEST.KE.hoest     ....GT-G-AAA-GCT---A-GATTCAGAGCTGG--C-G-G-G-AATGCAGA--GAACAGAGGACAGTT-T-G-GCA-T.........-AT-G-A-A-ACA-CCA-TTG-GCAA-A-G...AAGAGA--G   9050
LHOEST.CD.hoest447  ....GT-G-AAA-GCT---A-GATTGCGAGCTGGAGCTG---C-AGATCAGA--GAACAGA-GA-A--T-T-AGC---T.........CATG--AGA-ACC-C-G-CTG-G-CA-A-G...AAGAGA--G   7975
LHOEST.CD.hoest485  ....GT-G-AAA-GCT---A-GATTGCG-GCTGGAGCTG---C-AGATCAGA--GAACAGA-GA-A--T-T-AGC---T.........CATG-GAGAGACA-C-G-CTG-G-CA-A-G...AA-AGA--G   7972
LHOEST.CD.hoest524  ....GT-G-GAGTGCT---A-AACTGCGAGCTGGA-C-G---C-AGAGCAGA--GAACAGAGGA-AGCT-T-A-A-A-T.........AATGG-ACA-ACA---GGTTG-G-CA-A-CAGAA--AGA--G   7981
RCM.NG.SIV-NGM      ...............................--GT-C--AGGCTTCAGCAGCT-GC-TGCGG-GATGG-GAGG-CTAAT-CTTACT-CACCTGGAGAG-AC-ATGC-CG--TTGCAG-...-ACC--GCA   9330
COL.__.SIV-CGU1     ....ACA-CCTCAGACA--AGGA-----GCTTGG-AG-CTCT-C-GTA-C-ATTGG-A-C--GG-GTC--AG-C-TCAGA-TCT----ACACTTC-G---TGGC--CG-TCT-CGGCG...C-T-GC--G   8091

SYKES.KE.SYK173     .......GGAT--A-CAGC---G--TTCA-C--G--C--A-TCGCA--A-C-TCGC-CGGA-GG-AA-T-CG-GATTGG--G--G-CTATTTGGC-AGCAA-ATAC-CC-CTACCAG-...C---T-GTC   8663
SYKES.KE.SYK173 Env __.__.__G__F__I__S__Y__G__F__N__E__L__K__I__A__A__A__S__L__G__R__E__I__R__D__W__V__A__A__I__W__Q__A__I__Y__A__A__T__R__.__R__V__V_   gp41
SYKES.KE.SYK173 Nef                      M__G__S__T__S__S__K__S__Q__Q__L__R__S__E__G__K__Y__A__I__G__W__R__L__F__G__K__Q__Y__T__P__L__P__D__.__E__L__S     Nef
                                         |-> Nef start
```

```
                                                     gp160 and gp41 Env end <-|                |-> Nef starts
B.FR.HXB2 Env       _R__A__I__R__H__I__P__R__R__I__R__Q__G__L__E__R__I__L__L__$                 M__G__G__K__W__S__K__S__S__V__I__G__W__P__T__V__R__E__    Nef
B.FR.HXB2           AGAGCTATTCGCCACATACCTAGAAGAATAAGACAGGGCTTGGAAAGGATTTTGCTATAA..G............ATGGGTGGCAAGTGGTCAAAAAGTAGTGTGATTGGATGGCCTACTGTAAGGGAAA   8851

A.UG.U455           ----------T-A----------------C-----A-----A------GC----------..C............----------------------AG--CAGAG-G-A-------GAG--T---A---   8302
C.ET.ETH2220        ------T-CT--A--------------------------------GCAGC-----A----..A............-----G----C-AT-------T-----CCAG-A---------G--A----A----   8248
D.CD.84ZR085        -A----G-G-TT--------C-C---------------T--A------GC----------..A............-----------A---------------A-AG-----------G--A---------   8372
F1.BE.VI850         ------G---T-A--------------------------GCA------GC------G---..A............-----G-----------------C---A-AG-----------G-----G------   8128
G.SE.SE6165         --------C-T-A---------C---G--------A-----------AGC----------..A............-----A--------------------CA-AG----------CGAG----------   8290
H.CF.90CF056        ----------T------------------------------T-----A-GC---------.TA............-----A-----A--------------GA--GG---G---T-----A---------   8171
J.SE.SE7887         ----------TT--------------------------T--A-----AGC----------..A............-----GAA---A------------............------CAG----------   8146
K.CM.MP535C         ------T-A-TT--------------------------TC-C-----AGC---A------..A............---------------------------A-AG-----------G--A---------   8011
01_AE.TH.CM240      -----C----T------------------C-----------A-------C----------..C............-----AA-T-----------------CA-AG-G---------CAG--C-------   8425
02_AG.NG.IBNG       ------------A----------------C-----------C------GC----------..C............-----------------------C--CA-AG-G----------AG--T-T-A---   8378
03_AB.RU.KAL153-2   --------G--TA----------G-----C---------GCA----A-GC-----A----..C............-----G-----A--------------CA-AG-G---------CAG--T-------   8847
04_cpx.CY.94CY032-3 ---------T--A----------------C---------C-T-----AGC------T---..A............-----A-----A--------------CA-AG-----------GAGA---------   8224
N.CM.YBF30          --G-GA--ATTA-----------------C-----A---C-A-----AGCAC--A-----..-............-----AAAG-TT--------G--C--CC-AG-A--------AGAAA-C--A----   8423
O.CM.MVP5180        CA--GAT-C-TT-----C--A--------T-----A--TGCA-----A--C--AG-G---..C............-----GAATGCA---AGC-----C-AAT-TGCA------T-AGAA-----A--T-   8946
O.BE.ANT70          -C--GA-----TA----C--A--G-----T-----------A-----A-G---AT-G---..C............-----AAA-GCA-T-AG----G---AAT-TGAG------G-AG-A-----A----   8919
CPZ.US.CPZUS        CT--G----ATA--------A--------T--G--A---C-A--G--A-GC--AT----................------AA---------T--G-----CA-AG-A--------AGAA--C--AA-C-   8878
CPZ.GA.CPZGAB       ---AT----A-AA----C-----GC-C--------A--GC-------AGC---A--T---..-CACTGCATTAAG-----AAC---A-----T---------C--G-A---------GAG--C--AAG--   8924
CPZ.CM.CAM3         ----GC----TG--T-----A--------T--------T--A-----A-GC---------..C............-----CAA--G---------------CC--G-A--G-----AGA---C--A----   8288
CPZ.CM.CAM5         ----GC----TA--T--------G-----T-----A--T--A-----A-G----T-----..C............------AA-------------------C--G-G------G-AG----T--AC---   8594
CPZ.CD.CPZANT       ---ATC--AA-GG-AG-------GC-C--C-----A--G-------TTGCA--AAAT---..AC...........------TCTGCA-----T----TC-AGTG-G-----GCAAGACAA-C--TTAGG-   8311

H2A.GW.ALI Env      _R__R__I__L__A__V__P__R__R__I__R__Q__G__A__E__I__A__L__L__$                                                                          gp41
H2A.GW.ALI Nef      __G__G__Y__S__Q__S__Q__G__G__S__G__R__G__Q__K__L__P__S__C__E__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._    Nef
H2A.GW.ALI          --GAGG--A-T-GCAG-C--A--G--G--C--G-----GGCA----TTGCCC-C--G-G-......................................................................   9282

H2A.SN.HIV2ST       --G-GA--A-TTGCAG-C--A-----G--C--G-----GGCA----TCGCCC-C--G-G-......................................................................   8728
H2A.DE.BEN          --G-GA--A-T-GCAG-T--A--------C--G-----AGCA---CTCGCCC-C--G-G-......................................................................   9286
H2B.GH.D205         -----A--CAT-GCG--C--C--G-----C-----A--AC-T---CTC-CCC-CT-G-G-......................................................................   9250
H2B.CI.EHO          G---AG--CATTGCA-----C--G--G--C-----A--GGCT---CTCGCCC-CT-G-G-......................................................................   9227
H2G.CI.ABT96        ----GG--A-T-GCA-----G--------C--G--A--GC-T--GCTC-C-C-CT-G-G-......................................................................   8671
MAC.US.MM251        ---TGG--C-T-GCA--C-----G--G--T--G--A--GC-T--GCTC-CGC-CT-G-G-......................................................................   9225
SMM.US.SIVSMMH9     --G-GA--A-T-GCA--C--MR--C-C--C--G--A--GC-T--GCTC-C-C-CT-G-G-......................................................................   8732
STM.US.STM          ---AGG--CG-GGC---C--A--GC-C--C-----A--GC-T---CTC-CAC-CT-G-G-......................................................................   8894
TANTALUS.UG.TAN1    C-G--A--C-T-G-ACAT--A--------GC----A-AAC-----C--TGG--CAAC--G......................................................................   8906
GRIVET.ET.GRI-677   C-G--AG-CAT-A---GT--A------G-GC-G--A---C-T----AAG-CC--GGG---......................................................................   8596
VERVET.KE.VER-9063  --GCAC--CAT-AG--GT--A------G--C----A--GC-T--GGAA--CC-TAAT--G......................................................................   8913
VERVET.KE.AGMTYO    --GAAC--CGT-A---GT--A------G-GC----A---C-T--GGAA--CC-TAAT--G......................................................................   8367
VERVET.KE.AGM155    C-G--A--CAT-A--TCT--A------G-GC----A--GC-T--GGAAG-CC-TAAT--G......................................................................   8890
VERVET.DE.AGM3      --G--A--CAT-A--TCT--A------G-GC----A---C-T---G-A--CC-TAAT--G......................................................................   8411
SABAEUS.SN.SAB1C    --G--AG-CAT-GCACAT--A--------GC-C--A-AGC-T--G--ATGG--TAAT--G......................................................................   9087
SUN.GA.SIVSUN       -A-TGG-A-A-AACA-G-GGA--T-----TCCCAGT-AGACCAC-GA--CCACTT-G---......................................................................   9181
MANDRIL.GA.MNDGB1   -CGTGC--AT-G-------TC-----G--T-----AA-TGCA---C-ACCCC-TAAC---......................................................................   8378
LHOEST.KE.hoest     ---CTC-GATT-GGAGG-AGA--TG-G--TTCCAGT-AAGCCAC-GA--CCGCTT-G--G......................................................................   9110
LHOEST.CD.hoest447  ---TTC-GATT-AGAGG-AGA--TG-GT-TCCCAGT-AGACCAC-GA--CCGCT--G--G......................................................................   8035
LHOEST.CD.hoest485  ---TTC-GATT-AGAGGGAGA--TG-GT-TCCCAGT-AGACCAC-GA--CCGCT--G--G......................................................................   8032
LHOEST.CD.hoest524  ---TTC-GATT-AGA......--TG-GC-TCCCAGT-AGACCAC-GA--CCACT--G--G......................................................................   8035
RCM.NG.SIV-NGM      G--TGGCCAGC-ACGT-GTGC-----G--C--G--------C--G--AT--C-TAAC--G......................................................................   9390
COL.__.SIV-CGU1     C-TCGAGCG-CTG-ATAT-TGC--G--TGG-T-T-T-A-CGTCCCCAAGG-CCTGCC---......................................................................   8151

SYKES.KE.SYK173     GAG--CG--GCAGC-T-G--GC-G--GC-TC----A--------G-TCTACC-TAAC-G-......................................................................   8723
SYKES.KE.SYK173 Env _E__A__V__A__A__L__P__R__R__L__R__Q__G__L__E__I__Y__L__N__$                                                                          gp41
SYKES.KE.SYK173 Nef __R__P__L__Q__P__C__R__G__G__F__D__K__A__W__R__S__T__L__T__E__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._    Nef
```

```
B.FR.HXB2 Nef       R__M__R__R__.__.__.__.__.__.__.__.__.__.__.__.__A__.__.__.__E__P__A__A__D__R__V__G__A__A__S__R__D__L__E__K__H__G__A__I__T__.__S__S   Nef
B.FR.HXB2           GAATGAGACGA....................................GCT.........GAGCCAGCAGCAGATAGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATGGAGCAATCACA...AGTAG  8933

A.UG.U455           --------....................................GAAA--.........CCTG--------A-AG-A--A-------T-----A---TT-A--T---T--------G-----...--C--  8384
C.ET.ETH2220        ----A------....................................---.........-CT-----------GG-A--A--------G---------T-A--C---T----G---C-T---...--C--  8330
D.CD.84ZR085        ----A---AA-............ACTGATCCAAGGGAAAGAAGAAGAC--.........---------------G------------T-----------------G------G---------...-----  8478
F1.BE.VI850         ---------A-....................................A-C.........CCTA----------AG------------TG---------T-A--C-G--GG--------T---...--C--  8210
G.SE.SE6165         ----A---AAC....................................A-C.........CCTA----------AG-A--A-------T-----A---TT-A--T-GG---------------...--C--  8372
H.CF.90CF056        -------G---....................................---.........--A----T------AG----A-------TG--------TT----T-G--GC--G---G-----...-T--A  8253
J.SE.SE7887         -------G---.................................GCTC-C.........-CT------------G-A----------TG----A---TT---CT--G-----G---------...--C--  8231
K.CM.MP535C         -----------.........GCACGACCAGCAGCAGACAGGGTGGGAA-A.........C-A-----------CG------------T-----A--------CT-G------G---G-----...--C--  8120
01_AE.TH.CM240      A---A-AG....................................CAAA--.........CCT------A----AG-A--A-------T-----A---T--A--T--------------A---...-----  8507
02_AG.NG.IBNG       --------....................................CAAA-C.........CCTA--------ACAG-A--A-------------A---TT-A--T-G----------------...--C--  8460
03_AB.RU.KAL153-2   ----A------.................................GCTC--.........-CT-----G---AGAG-A--A---C---T-----A---TT----T--GT--------G-----...--C--  8932
04_cpx.CY.94CY032-3 -------G---GCTCGAGCTGAGCCAGAAAGAATGAGGCGAGCTCAA---.........-------------CAG-A--A-------TG----A----T----C--------G---------...-T--A  8342
N.CM.YBF30          --------A--.....................CAAACGCAAGAACCA--A.........-TAGAGC------TAG-A-CA--------T----A---T--A-CT--T-GA--G--C-----C...-TA--  8520
O.CM.MVP5180        -----------.....................TCCTCCTCTGATCCTCAA.........C-A---TGT---CC-G-A--A-----T-TC--CA-G--GT-A-C--CTAGA--G-G---AT--...---TC  9043
O.BE.ANT70          --------A--..................ACTAGAACTTTCCCTGAGT--.........--A---TGC---CC-G-A--A---CAGATC--CA-G--AT-A-C-GCTAGA----GG--AC--...---TC  9019
CPZ.US.CPZUS        --T--....................................AGA...CAAACTCAGACAACAG----------AG-A--A---C-T-TC----A----T-A-C-G-------G--T-----T...-C---  8966
CPZ.GA.CPZGAB       ----A....................................AGGGAA---.........CCAA----------GG-A--A----A--TT--GAAG-----A----G---C-----T-----T...-----  9006
CPZ.CM.CAM3         --C-A....................................AGCGAACTC.........A-C-TGCTT---CCAG-A--A---C---TC--A-A----T-A--GGC----------G----C...-C---  8370
CPZ.CM.CAM5         -G--A....................................AGACAAA--.........CCA---------TCAG-A--A---C---TC--A-A---T--A--GGC--G------------C...-C---  8676
CPZ.CD.CPZANT       A---A....................................CATGAAA-A.........A-T---......---GATA-A--GC-CTGTGGAAAC--AT---CC-GTAGA--G---C-A--C...TCCTC  8387

H2A.GW.ALI Nef      .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G__Q__R__Y__Q__Q__G__D__F__M__N__T__P__W__R__T__P__A__T__.__E__R   Nef
H2A.GW.ALI          ................................................................--G-CAGCGGTATCA-CAG-G--AT-T-ATGA-TACCCC-TGGAGAACCC--GCA---...GAA--  9345

H2A.SN.HIV2ST       ................................................................G-G-CGGCGGTATCAACAG-G--AT-T-ATGA-TACCCC-TGGAGA-CCC--GCAGA-...G-GGA  8791
H2A.DE.BEN          ................................................................G-G-CAGC-GTATCA-CAG-G--AG-ACATGA--AGCCC-TGGAGAAACC--GCA---...GAA--  9349
H2B.GH.D205         ................................................................................................................G-G-CAA-AT...GCCC-  9265
H2B.CI.EHO          ................................................................................................................T-G-CAA-AG...-C-TT  9242
H2G.CI.ABT96        ................................................................--GGCA-C-ATACAC-CAG-GTCA--T-ATGA-TACCCC-TGGAGAAACC--GCA---...GAA--  8734
MAC.US.MM251        ................................................................G-G-CAGA-ATACAATCAG-GGCAG-A-ATGA-TACTCC-TGGAGAAACC--GCTGA-...GAA-A  9288
SMM.US.SIVSMMH9     ................................................................GC-TCAGA-ATATTGT-AG-GTCA--TCATGA-TACCCCTTGGAGGAACC--GCA---...GAAG-  8795
STM.US.STM          ................................................................GC-TCAGAGATATAAT-A--GTCA--T-ATGA-TACCCCTTGGA-AAACC-GGCAG--...GAA--  8957
TANTALUS.UG.TAN1    ..........................................................................................................GAGG---AAG-G---C...CC-GA  8927
GRIVET.ET.GRI-677   .............................................................................................................A--TAA---G--T...CCAGA  8614
VERVET.KE.VER-9063  ....................................................................................................-A-G-GGA----CAAC--G--T...CCAGA  8940
VERVET.KE.AGMTYO    ....................................................................................................-A--C--A----CAAC--G--G...GAAGA  8394
VERVET.KE.AGM155    ....................................................................................................-A-G-G-A----CAAC--G--T...CCAGA  8917
VERVET.DE.AGM3      ....................................................................................................-A-G-G-A----TAA---G--T...CCAGA  8438
SABAEUS.SN.SAB1C    .................................................................................................G-AA-G-G-CTCCTCTAG--CAGA-...GAAG-  9117
SUN.GA.SIVSUN       ..................................................................................................................................  9181
MANDRIL.GA.MNDGB1   ..................................................................................................................................  8378
LHOEST.KE.hoest     ..................................................................................................................................  9110
LHOEST.CD.hoest447  ..................................................................................................................................  8035
LHOEST.CD.hoest485  ..................................................................................................................................  8032
LHOEST.CD.hoest524  ..................................................................................................................................  8035
RCM.NG.SIV-NGM      ..................................................................................................................................  9390
COL.__.SIV-CGU1     ..................................................................................................................................  8151

SYKES.KE.SYK173     ...................................................................................................................GCC--TTGACCCACA  8738
SYKES.KE.SYK173 Nef .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__P__I__D__P__H   Nef
```

```
B.FR.HXB2 Nef       __N__T__A__A__T__N__A__A__.__.__.__.__.__.__.__.__C__A__W__L__E__A__Q__E__E__E__E__.__.__V__G__F__.__P__V__T__P__Q__V__P__L__R__P_   Nef
B.FR.HXB2           CAATACAGCAGCTACCAATGCTGCT.......................TGTGCCTGGCTAGAAGCACAAGAGGAGGAGGAG......GTGGGTTTT...CCAGTCACACCTCAGGTACCTTTAAGACCA  9030

A.UG.U455           T------T--T----T-----CAG-.......................-----------G-----G--------A-GA--C......--A--C--C...-----T-GG--A------------------  8481
C.ET.ETH2220        ---C---C-C--C-AT---C---A-.......................-----------G-----G-----------A--AGAG...--A--C---...-------G----------------------  8430
D.CD.84ZR085        ----------A----T--------C.......................---------G----------G--A-----C---GAG...-----C---...-------G----------------------  8578
F1.BE.VI850         ------TAG-A----T---C---AC.......................-TG--------G--------------------A......--A--C---...-------G-------------G--------  8307
G.SE.SE6165         ------------C-AT---C---A-.......................-----------G-----------------CTCAGAA...--A--C---...-------G---A--A--------G------  8472
H.CF.90CF056        T---------T----T---CG--A-.......................GCC--------G--------------C-G----GAA...--A--C---...-------GG---------------------  8353
J.SE.SE7887         ---------------A-------AC.......................-----------G---------ACA---------......--A--A---...-------A-------A----A---------  8328
K.CM.MP535C         ------TT--CAC-AT---C---A-.......................-----------G--------G-----A-----A......--A------...-------G------------A---------  8217
01_AE.TH.CM240      T---......ATAGAT-------A-.......................----T------GAG------------C------......--A--C---...-------TG--G---------C--------  8598
02_AG.NG.IBNG       T---------CAA--T--CC---A-.......................-----------G--------------T---A-T......--A--C---...-------GG--A-----G-----G------  8557
03_AB.RU.KAL153-2   T-------------AT-------A-.......................--C--------G--G--G--------A------......--A--C---...-------GG--G-----------G-----G  9029
04_cpx.CY.94CY032-3 ---------------T---C---AC.......................AAAA-------G-----------A-----A--AGAG...--A------...-------GG--A--------------G---  8442
N.CM.YBF30          A-----TAG--AC-AT----AAAG-.......................ATA--T-----------------A--A--A---GAA...--A--C---...-----ACGC-----------A-----G---  8620
O.CM.MVP5180        -C-C--TC-TCAA-A------A--C.......................CT---A-TC-----CAGC--CA-A--T-----T......--A--C--C...-----A-G------A--G---C----G---  9140
O.BE.ANT70          -C----TC-TCAA-A------A--C.......................CT---A-TC------AGT--CC-A-----A--A......--A------...-----AG-------A--G---C----G---  9116
CPZ.US.CPZUS        A------C--CAG-A---CCAAA--.......................CTG--A---T----T-AGATGACTA-TC-T--AAGTGAA--A--C--C...--T--TCGC-----A-----A-----G--C  9069
CPZ.GA.CPZGAB       G--C--CC---AG--T---CAAA--.......................CTA--T--------G-A-ATG--CA-T--A......GAA--A--A--C...-----A-G---A-----T--AACG------  9103
CPZ.CM.CAM3         GC-C---C--CAG-AT---CAGA--.......................CTG------T------A-ATGC-A--AC-CA-TGAAGAA--A--C---...-----T-G------A------C-T-----T  8473
CPZ.CM.CAM5         A---CAC...A-A-AAT--CAGA-C.......................CTG------T------A-ATG--AA-TC-T--AGAAGAA--A--C---...-----T-G---C--------CC--------  8776
CPZ.CD.CPZANT       T-C--T-...-G---AG-AAAA-A-.......................GTAAT-ACCTATTC--A-G-TC-TACA-----A...GGGACA--C---...--T--G-G------A--C--CA-G-----C  8484

H2A.GW.ALI Nef      __E__K__E__L__Y__K__Q__Q__.__.__.__.__.__.__.__.__N__M__D__D__V__D__L__D__D__D__D__S__L__V__G__V__.__S__V__T__P__R__V__Q__L__R__T_   Nef
H2A.GW.ALI          AG-G-A--A-TTGTA---GCAACAG.......................AA-ATGGATGAT-T--ATTT---T--T--T--TAGCCTA--A--GG-C...T-T--------AAGA----AA------A--  9448

H2A.SN.HIV2ST       G...-A--GCT-GTA---GCAACAA.......................AA-ATGGATGAT-T--ATTC---T--T--T--C...CTA--A--GG-C...--T--------AAGA-----A------GA-  8888
H2A.DE.BEN          AC-G-A--ATTTGTAT-GGCAGCAA.......................AA-ATGGATGAT-T--ATTCT--T--T--T--C...CTAA-A--AG--...--T--T-----AAGA-----ACGG---GA-  9449
H2B.GH.D205         GGGGG---A--GGGGAGGGCAACAA.............................GATGC---T-A-AGT--T-----C--TGAA...-----AGCCATATGTAAG--C--AATA-----AC-G------  9362
H2B.CI.EHO          GGGAG---A--GGGGAGGAAAACAA................................GATTC--ATG-G--T--T-----CAATGAA--A--GG-C...-GT--A-G---CGG---C--AC-G--G---  9336
H2G.CI.ABT96        AGCA-AGTT---ATA--GGYAGCAA..........................AATAATGAT--T-T-G-TAGT--T--TA-TGACTTA--A--AG-A...------TAC--CA-A--C--GC-----GT-  8834
MAC.US.MM251        AG-A-A-TT---ATA--GAAAACAA.......................AA-ATGGATGATAT--ATG-G--A--T--T--C...TTG--A--GG-A...T----G-GG--AA-A--T--CC-----G--  9388
SMM.US.SIVSMMH9     AGCA-A-TT--G-TAT-GACAACAA.......................AACATGGATGAT-TG-ATA-T--A--T.........TTA--A----G-...-----G-GC--AAGA--C--CG-----ATC  8889
STM.US.STM          -GC--AGCT--AATAT-GACAGCAA.......................AACATGGATGAT-T--ATG-G--A--T--TA-T...CTA--A--AG-A...G----ACAT--AAGA--C--AC-T--GGA-  9057
TANTALUS.UG.TAN1    AGGA-G-AA-TTGG-AGC--A--AC.......................ACATGGGATGACTGG-A--C------A--A--A......-----C---...-----A-G---A-GA--G----------A-  9024
GRIVET.ET.GRI-677   -GGCCGCCGCCTGCAAG-A-GA-AC.......................ACCTTTGATGAGTGG-ATG-T--T--A--A--A......--A--C--C...--T--GCA-----GA--C--C-------AG  8711
VERVET.KE.VER-9063  TGGA-G-CGCCTGCAAG-A-GA-AC.......................GAATGGGATGA-TGGT--G-T--T--A--T--A......-----A---...--T--A-A---AAGA-----A--G--G-A-  9037
VERVET.KE.AGMTYO    GG-A-GGAGGCT-CAAG-A-GA-AC.......................ACCTGGGAAGAGTGGT-GG-T--T-----A--A......-----A---...-----G-G---AAGA-----C-----G-A-  8491
VERVET.KE.AGM155    AGGA-G-CGTCTACAGG-C-GG-AC.......................CAATGGGATGA-TGGT--G-T--A--A--T--A......-----A---...-----A-G---AAGA--G--AC------A-  9014
VERVET.DE.AGM3      AGGA-G-AA-TTACAAG-A-GA-A-.......................AAATGGGATGA-TGGT-TG-T--A--A--T--A......--A--A---...-----A-G---AAGA--G--GC------A-  8535
SABAEUS.SN.SAB1C    TGTA-A-ATGAAACATC-A-GAAGA.......................CAGC---C-TGGT-T-ATG-G--T--A--A--A......--A--C---...-----A-GG---TGCT----AC--C--G--  9214
SUN.GA.SIVSUN       .................................................................................................................................  9181
MANDRIL.GA.MNDGB1   ...............................AGATGAGAA..ACTTGAC-TAA-ACA--AG--TC--G---------A--A......--T--A---...--T--GTGT-GC--A--TT-C------GTG  8466
LHOEST.KE.hoest     .................................................................................................................................  9110
LHOEST.CD.hoest447  .................................................................................................................................  8035
LHOEST.CD.hoest485  .................................................................................................................................  8032
LHOEST.CD.hoest524  .................................................................................................................................  8035
RCM.NG.SIV-NGM      ............AGGAGGGTT-A-AATAGAAACTCAGCAGTCAGTGGATGCAATAGATTGGT-T-A-G-TACA--T--TACC...TTG--A--G---...-----G-G---A--A--C--CA---G---  9502
COL.__.SIV-CGU1     ...-TGCC-TTGCC-GCT-TGCTA-AGACATCCAGATGTATATTACCACAACC-AGAA-C---G-ACGT---A--T-TA--A...GCA-GAT---A-......--G-TGT---GAC-G--AG--C---AG  8269

SYKES.KE.SYK173     -GGGC-G-ACAGAGA-TGG-GACA-TCGGGAGGACAAAAATTCTCACCTG-A-ATATAG-GC-G-ATG---GA--CACA-GT...CTC--A-----C...--T--GTGT-----AAC---AC-C---A--  8862
SYKES.KE.SYK173 Nef __G__P__D__R__D__W__G__H__S__G__G__Q__K__F__S__P__G__D__I__V__Q__D__E__G__D__T__G__.__L__V__G__F__.__P__V__C__P__Q__T__P__L__R__T_   Nef
```

```
B.FR.HXB2 Nef       _.__.__M__T__Y__K__A__A__V__D__L__S__H__F__L__K__E__K__G__G__L__E__G__L__I__H__S__Q__R__R__Q__D__I__L__D__L__W__I__Y__H__T__Q__G__   Nef
                                                                                 |-> 3′ LTR U3 region
B.FR.HXB2           ......ATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCT   9154

A.UG.U455           ......--------T---------T-T-----C---TT---------------------------T---T----------A--G-A---------A---------------G----T-----------A-   8605
C.ET.ETH2220        ......--------T--------AT-C-----C----T-------------------------------T-----T-----A-G-A---G--G--G--T------T-----G----TA------------   8554
D.CD.84ZR085        ......----------------------------------------G-A-----------------------G--T-------G-A---------------------T---G------------------   8702
F1.BE.VI850         ......-----C--T-----------------C--T---------------------------------T-----T-----A-G-A----GG-----C-------------G-T----------------   8431
G.SE.SE6165         ......-------TT----GT---T-T-----C---TT---------------------------T---------T-----A-G-A---------G--------C------G-----A-T--------A-   8596
H.CF.90CF056        ......-----C--T----G----T-T-----C-----T--------------------------T---T-----T-----A-GCA------G--C---------T-A---G----TA------------   8477
J.SE.SE7887         ......-----A-------G------T---------TT---------------------------T---T-----T-----A-G-A---------G---------------G-TC-TA----------T-   8452
K.CM.MP535C         ......--------T---------T-T-----CG--TT---------------------------T---------T-----A-G-A---------A---------------G-T--T-----------T-   8341
01_AE.TH.CM240      ......--------T----G----T-T---------TT---------------------------T---------T-----A-G-A---------G--------CT-A---G----TA-T----------   8722
02_AG.NG.IBNG       ......--------T----G------C-----C--------------------------------T---T-----T-----A-G-A---------A---------------G----TA----------A-   8681
03_AB.RU.KAL153-2   ......-----A--T----G----T-C-----C-------------------G----.-------T---T-----T-----A-G-A---------A---------------G----T-------------   9152
04_cpx.CY.94CY032-3 ......-----C-TT--A-G----T-------C--------------------------------T---------T-----A-G-A---------G---------------G----T-----------T-   8566
N.CM.YBF30          ......--A--C--T--ACAG---T-T------TC-TT------------T---------------------G--TGG---AG--A---G--------T--A--C--C-----G--T------------A   8744
O.CM.MVP5180        ......-----C-TT--A-----CT-T--C--C---TT------------------A--------T---T-----T-------T-AG---GC---A-----G-----C-----A--T-----T--G--A-   9264
O.BE.ANT70          ......-----C--T--A-G---AT-T--C--C---TT------------------A------------T-----T-------T-A----GC---A-----G-----T---G-G--TA----T--G--A-   9240
CPZ.US.CPZUS        ......-----------ACA---AT-C---T-AG--TT-------------------------------T--G--T-T---AG------------G-----A-----C---G-T-----T--------AA   9193
CPZ.GA.CPZGAB       ......--------T--A------T-T------TCA---------------------------------T--G--T-----AGG-----------G--------C--C---G----T-------------   9227
CPZ.CM.CAM3         ......-----C-TT--A------T-T---T-A---------------A--------------------------T-----AT--NNNNNNNNNNNNNNN-G----ACG--TA---T--------G----   8597
CPZ.CM.CAM5         ......-----A------CA----T-T---T-A------------------------------------T--G--T-----AC--A---G-----G-----A-----C---G-T--T--------G--AA   8900
CPZ.CD.CPZANT       ......-----AG-A--ATT---------C--GTCATGN--------------------------TN--------T-NN-A-TT-A----GC--C------A---A-------G-TTA----C--G--AA   8608

H2A.GW.ALI Nef      _.__.__M__T__Y__K__L__A__V__D__M__S__H__L__I__K__E__R__G__G__L__E__G__M__F__Y__S__E__R__R__H__R__I__L__D__I__Y__L__E__K__E__E__G__   Nef
H2A.GW.ALI          ......-----A-----ATTG--A------A-GTCA--T--AA---------G----------------A-GT--T--AGTG-G--------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G----GA   9572

H2A.SN.HIV2ST       ......-----A--T-G-TTG--AAG----A-GTCA--T--GA----------------------------GTA-T--AGTG-T--G----GTAGAG----A--CA-A-ACT-AG-AA-GGA-G-G--AA   9012
H2A.DE.BEN          ......-----C--T--ATTG--AA-----A-GTCA--T---A----------A---------C-----A-GT--T--AGTAGG--G-----TAGA-----A--CA-A-ACC-AG-AA-AGAGG----GA   9573
H2B.GH.D205         ......-----A-----ACT---A-----CA-GTCT--T---A--------C-----------------A-GTA-T-TAGTG-G--------CAGA--AT-A--CACA-ATT-TG-GA-TGA-G-----A   9486
H2B.CI.EHO          ......-----A-T---ACT---A-----CA-GTCT--T-------------------A----------A-TT-CT-TAGTG-G-----G--TA-A--A--A--CACA-ACT-AG-AA-TGA-G-----A   9460
H2G.CI.ABT96        ......----G------ATTG--AA----CA-GTCA--T---A--------------------------A-TTA-T--AGTG----------TAGA-----A--CA-A-AC--GG-AA-GGAG-----TA   8958
MAC.US.MM251        ......-----------ATTG--AA-----A-GTCT--T---A--------------------------A-TTA-T--AGTGC---------TAGA---T-A--CA---ACT-AG-AA-GGA-G-----A   9512
SMM.US.SIVSMMH9     ......-----A-----ATTG--AA----CA-GTCT--T---A--------------------------G-TTA-T--AGTATT--------TA-A--AT-A-----R-ACT-AG-AA-RGA-G----AA   9013
STM.US.STM          ......------------TTG--AA----C--GTCA--T---A-----AGC------------------A-TTA-T--AGTG-G--------TAGA--AT-G--CA---ACT-AG-AA-AGA-G----GA   9181
TANTALUS.UG.TAN1    ......-----A--T---TTG--A------T-CTC---------------------A--------G--CA--TA-TGG--A-C--AG---G-GC-A-----GA----C-ATGC-CTAA-TGA-TGG---A   9148
GRIVET.ET.GRI-677   ......-----C--T--ATT---A--G--CT--TC-------------TC---------------T---A--TA-T----TG--------G--A-G-----GA--T---ATGC--TGA--GAGTGG--AA   8835
VERVET.KE.VER-9063  ......-----------ATT---A--G--CT--TCG--------------------A--------T---A-TTA-T-----G----G--GA-TAGA-----CA-C----ATGCTCTTA-TGA-TGG--AA   9161
VERVET.KE.AGMTYO    ......--------T--ACTT--A--G---T--TCG--------------------A--------T---A-TTA-T-----G-T--G---A-TA-G-----GA------ATGCTCTTA-TGA-TGG--GA   8615
VERVET.KE.AGM155    ......--A--A-----ACTT--A------T--TCG--------------------A--------T---A-TTA-T-----G-T--G---A-TA-G------A------ATGC-CT-A-TGA-TGG--AA   9138
VERVET.DE.AGM3      ......-----C--T--ATT---G--G--CT--TCG-----------------------------T---A-TTA-T-----G-C--G--GA-TC-G-----AA-C----ACGC-CT-A-TGAGTGG--AA   8659
SABAEUS.SN.SAB1C    ......-----A--T---TTG--AA----CT--G-----------------------------------A-TTA-T--AGTG----G---A-GA-G---T-G-----C-ATGCTCTTA-TGAGTGG---A   9338
SUN.GA.SIVSUN       .....GCCA--A--T---CAGCTAA-------CTCA--T---A-------------T--------------CTGGT-----AG--CC---G----A---T-G---T---ATGCAG-AA-TGAGTGG----   9306
MANDRIL.GA.MNDGB1   CCA......T-A-----A-ATCTGA----CT-CTCT--T---A---------------------G----A--TA-T-TAG-AGG------G----A-----A-----C-ATGCAG-GA-TGAGTGG--A-   8590
LHOEST.KE.hoest     .....CCCT--C--T---A-TCT-A----C--CTCA--T---A-------------T---------------TA-TGG-----G--------G--C---T-AAT---C-ACTGTG-GA-TGAGTGG--TC   9235
LHOEST.CD.hoest447  .....CCCT--C--T---A-TCTGA----C--CTCA--T---A-------------T---------------TA-TGG-----G--------G--C---T-ATG---T-ACTGTG-GA-TGA-TGG--TC   8160
LHOEST.CD.hoest485  .....CCCT-----T---A-TCTGA----C--CTCA--T---A-------------T---------------TA-TGG-----G--------G--C---T-GTG---T-ACTGTG-GA-TGA-TGG--TC   8157
LHOEST.CD.hoest524  .....CTCT--C--T--AA-CTTGA----C--CTCA--T---A-------------T---------------TA-TGG-----G-----GA-T--C---T-AACA--C-ACTGTG-GA-TGA-TGG--TC   8160
RCM.NG.SIV-NGM      ......---T-------ATT---AA-------ATCT---------------------------------A-TTA-TT---AG----G---AGGCGC--TT-A--C--C-AC--GG-GA-TGA---T--AA   9626
COL.__.SIV-CGU1     ......GC---CATT---CT-CTG--G--CA-CTCTTG------G-----------T--------G---A-T---AGAA---CTGAG---G-T--CC-GA-A--G-A--ATGC----ATTGAGTGG----   8393

SYKES.KE.SYK173     ......T-A--C-----ATTG--AA----C--CTCT--T---A-----A-C-----A------C-G--CA-G-A-T-T-GTG---AG---G-T--A--T---C----C-ACC-AC-GA-TGAG--T--AA   8986
SYKES.KE.SYK173 Nef _.__.__L__T__Y__K__L__A__I__D__L__S__H__F__I__K__N__K__G__G__L__Q__G__M__N__Y__C__E__K__R__D__E__I__L__H__L__Y__L__Q__N__E__H__G__   Nef
```

```
    Nef end premature in HXB2   \/
B.FR.HXB2 Nef       Y__F__P__D__$__Q__N__Y__T__P__G__P__G__.__.__.__.__.__.__.__V__R__Y__P__L__T__F__G__W__C__Y__K__L__V__P__V__E__P__D__K__I__E__E__A   Nef
B.FR.HXB2           ACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGG.....................GTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGATAGAAGAGGC   9263

A.UG.U455           T-----------G------T-----------------A.....................A-------C-----A--A--------------------------------T----C-G-AG----G--A--   8714
C.ET.ETH2220        T-----------G---A--------------A------.....................-------------------------------T------------------C---AG-G-AG-G-----AAT   8663
D.CD.84ZR085        T-----------G---A---------------------.....................A-----------T-A--------G-------T-G----------------T-----GGT-G-----A----   8811
F1.BE.VI850         ------------G-------------------------.....................A-------------------G----------T------------------T-----AG--G-G---A----   8540
G.SE.SE6165         -------A----G------------G------------.....................ACT----TC--C-------------------T---A---------A-G--T----CAG--G----G--A--   8705
H.CF.90CF056        ----------C-G-------------------------.....................-AG----T---C-----------G-------T---------------AA-T---C-GG--G-----C----   8586
J.SE.SE7887         ----------C-G---A---------------------.....................A-------C---T----A--------------------------------T---AGCG-AG----G--A--   8561
K.CM.MP535C         T-----------G---A---------------------.....................A-------C--------------G--------------------------T----CAG-AG--------A-   8450
01_AE.TH.CM240      T-----------G---A---------------------.....................A------TC------TGT-------------T------------------C-A-AGAG-AG----G----A   8831
02_AG.NG.IBNG       T------A----G-------------------------.....................ACT----TC--------------G-------T---A---------A-G--T----CAG--G----G--A--   8790
03_AB.RU.KAL153-2   ------------G---...............-------.....................A-T----TC-----A--A-----G-----------------------C--C----C-G-AG----------   9246
04_cpx.CY.94CY032-3 T-----------G-G-T--------------A------.....................-AG----TC------TG--------------T---------------A--T---C-GG--G-G--------   8675
N.CM.YBF30          T-C-------C-G---T--------------------A.....................A-T-----C--CG-A----------------T---A---------T-GTCAG-T--AG-AG----------   8853
O.CM.MVP5180        T-----------G----TGT--------G--A-----A.....................CCT----TC--------A----------TG-TT--A--G--------GTCAG----AG--GC---GAGACT   9373
O.BE.ANT70          T-----------G------------------A-----A.....................AC---G-TC--------A----------TG-TT--A-----------GTCAGA---AG--GC----AGACT   9349
CPZ.US.CPZUS        T---------C-G------T-----C--------T--A.....................--A--------C--A--A-A---G-------TT---T----C--TC-CACAGA---GG-AG----GC----   9302
CPZ.GA.CPZGAB       T---------C-G-------------A----A-----A.....................ACA----TC-----CTGT-------------T---A-----G--CC-GACAGAG--GC--G-----C-A--   9336
CPZ.CM.CAM3         T-----------G------------------------A.....................--------C-----A--A-A---GG------TT--A--G--T---C-GACA-----AG-AG-G---C----   8706
CPZ.CM.CAM5         TT--------C-G------T-----------------A.....................--------C-----A--A-------------T------G-----TC-GACA-----GG-AG-G--GA-A--   9009
CPZ.CD.CPZANT       T-----------G------T------GA--A------A.....................--T-----------CTGTAGA--------T-TT---T----------A--AT--CCACCTGAT........   8709

H2A.GW.ALI Nef      I__I__P__D__W__Q__N__Y__T__H__G__P__G__.__.__.__.__.__.__.__I__R__Y__P__M__F__F__G__W__L__W__K__L__V__P__V__D__V__P__Q__E__G__E__D   Nef
H2A.GW.ALI          TAA-T--A----G---------T--T-AT--------A.....................A-A--G--C--GA--TT------G---CTG-GG--------------A--TGTCCCAC-AGA--GG----A   9681

H2A.SN.HIV2ST       TAA-TGGA--C-G---------T--T-AT--A-----A.....................--A--G------AA-TT------G----TA-GG--------------A--TGTCCCAC-AGAG-G---T-A   9121
H2A.DE.BEN          TAA-A--A----G------T--T--T-AT--------A.....................--A--G--C---A--TA---C--G---CTG-GG---------T----A--A-TCTCAC-AGAG-C-----A   9682
H2B.GH.D205         TTG-GT---GC-G---A-----T--T-AT---------.....................A-A--G-----CAAATA------T---CTG-GG-----G--------A---GTGCCAGCAGCGACCCGA-A   9595
H2B.CI.EHO          TTG-GT---GA-G--------------AT--------A.....................--A--------CAA-TT------C---CTA-GG--------G---A-AA-TATGATAGCAGA-CC-----A   9569
H2G.CI.ABT96        TAA-T--A----G------T------G----A-----T.....................A-A---------R----G-----C---CTA-GG--A-----------CAC-GTCTC-G-TGAG-C-C---A   9067
MAC.US.MM251        T-A-A--A----G----G-T-----CT----A-----A.....................A-T-----C---AA---A-----C---CTA-GG--AT----C--T--AA-TGT-TCAG-TGAG-C-C---A   9621
SMM.US.SIVSMMH9     TAA-A--A----G---A--T------G----------A.....................A-------C--TA--TT----------CT--GG--AT---WC--G--AA-TGTCTCAG-TGA--CTC---A   9122
STM.US.STM          TAG-T--A----G------T------G----A------.....................A-A-----C---AA-CAG---------CTA-GG--A-----------A--TATGTCA--TGAG-C-C---A   9290
TANTALUS.UG.TAN1    TTA-AGA-----G---AGCA--TT-C--T--------C.....................AC---GA-G--C-GATG---C--C-TT----TTG--T----G--G--A--TGTCTCCC--GA--CCC---A   9257
GRIVET.ET.GRI-677   TAA-AGA-----G---AGCT---T-------C--G---.....................A-A--G--C--GAGAGT------C-TC----TT--------C-----G--C-TGC--G--GAG-C-CGCAA   8944
VERVET.KE.VER-9063  T-A-TGA-----G-A-CGC--GGT--A----A-----C.....................--A--------C-GCTG------C-TT----T------------G--A---ATGC--G-AGAG-C----A-   9270
VERVET.KE.AGMTYO    TAA-TGA-----G-A-TGC--GGT-GAAG--A-----A.....................A-A----TC--TAAATG------G-TC----TT--------G-----G--CTT-C--G--GA--C-C-AA-   8724
VERVET.KE.AGM155    T-A-TGA-----G-A-CGCA-GGT--AA---A--T---.....................A-A-----C--GAG-TG------C-TC----T------------G----CC-TGC--G--GA--C---AA-   9247
VERVET.DE.AGM3      T-A-TGA-----G-A-TGCT-GGT--GA---A-----A.....................A-------C---AGATG---C--C-TC----TT--AT-G--------A--C-TGC--G--GA--C----A-   8768
SABAEUS.SN.SAB1C    TAG-AGA--GA-G------T--T---GAT--------T.....................AC---G--C---AA-TG--------------TT--A--G-----T--G--C-TGTCAG--GAG-CTA--AA   9447
SUN.GA.SIVSUN       T-A-AA---GC-G---AG-T-----CAA------T--A.....................--G---------AA-G-------C---CTA-GG-------C------CACCATT---G-AGACAGG--T..   9413
MANDRIL.GA.MNDGB1   TTGAA----GA-G---AC-G--T--GA----T-----A.....................AC---------TAA---A-------TCCTG-TT-----G-A------GAGCAG--C--TAGG---T---..   8697
LHOEST.KE.hoest     T-A-TGGA--C-TCAT----------GAT--A-----A.....................ACA-----C--T-----A-----------TG-GGC--T---A------G-CTTGT---G--TATA----TC-   9344
LHOEST.CD.hoest447  T-A-TGGA--C-TCAT--CT------GAT---------.....................ACGC-T-TC---T-AG-A----------TG-GGC----G-A------G-CCTGT---G--TACAC---TC-   8269
LHOEST.CD.hoest485  T-A-TGGA--C-TCAT--C-------GAT--A------.....................ACAC-T--C---T-A--A----------TG-GGC----G-A------G-CCTGT---G--TACAC---TC-   8266
LHOEST.CD.hoest524  T-A-TGGA--C-TCAT----------GAT--A-----A.....................ACTC-T--C-----CG-----------CTG-GGC--T-G-AG-----A-CATGT---G--T-TA----TC-   8269
RCM.NG.SIV-NGM      TAA-A--A----G------T-----T------------.....................A-A---------ACA-TG-----C---CTA-GGC-AT----------A--TGTCAG-G-TGA--C-ACA-A   9735
COL.__.SIV-CGU1     G---AAAA-G--G-TT-G-G--TGA-GAT-A--TG--AGAGGATGGAGCATTAAAAGAG-A----A-G--T---GTTGC-------CTT-GG--AT----CTACA-A----A-CT-GGAGAGT-T-CATA   8523

SYKES.KE.SYK173     T-A-A...--CAG-ATC--T-----CT----------A.....................AC-C-G--C--GT-A-T-------------TA-GGG-A-----T---AA---AT---GGATGCTTG---TA   9092
SYKES.KE.SYK173 Nef I__I__.__D__R__I__N__Y__T__S__G__P__G__.__.__.__.__.__.__.__T__R__Y__P__L__I__F__G__W__L__W__E__L__V__P__N__E__I__E__G__C__L__E__Y   Nef
```

```
B.FR.HXB2 Nef       __N__K__G__E__N__T__S__.__L__L__H__P__V__S__L__H__G__M__D__D__.__.__P__E__R__E__V__L__E__W__R__F__D__S__R__L__A__F__H__H__.__V__A_   Nef
B.FR.HXB2           CAATAAAGGAGAGAACACCAGC...TTGTTACACCCTGTGAGCCTGCATGGGATGGATGAC......CCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCAC...GTGGCC   9381

A.UG.U455           T-C-GG-----------A---T...C-AC--------A-AT---AA-----AG-A-----T......GAA----A----------AT-----A---------TAC---G-----AA-A---...AGA--T   8832
C.ET.ETH2220        ----G------------A-T--...---C-G-------C------------A-----G--T......GAA--C--------A---A------A---------T-A-------CGCAGA---...A-----   8781
D.CD.84ZR085        --C-G-------AG---A-T--...---C--------A-CT---A------A-----G---......GA-----AG---------T------A---A-----A-A---------G-G---...AA----   8929
F1.BE.VI850         ----G------------A-T--...-----------CA------AA-----A-----G--T......GAA--C--------AC-GAG-----A----------A----G---C-GAGA---...A-A---   8658
G.SE.SE6165         -----------------A---T...C-A--------CA-CT---A------A-----G---......GAA--C---------C-G-T------A---------TA--------CGGAGA---...A-A---   8823
H.CF.90CF056        ----G------------A----...---C-------CA-------------A-----G--T......GAC-G----------C-GAT-----AA--------T--A--------GACA---...T-----   8704
J.SE.SE7887         T---G-------A----A-T--...C----------CA-AT---A------A--A--G--T......GAA--A------------C------A----------TC-------CGGAGA---...A-A---   8679
K.CM.MP535C         --CAG--------G---A-T--...C----------CA-A-A--A------A-----G--T......GAAC-C------A-A---AT-----A---------TTC-------CGAAGA---...--A---   8568
01_AE.TH.CM240      ---C--------A----A-T--...C----------CA------A------A--A--G---......GAA--A---------C-GAT-----A---------TGC-------CGAA-A---...--A---   8949
02_AG.NG.IBNG       ----G------------A---T...--A--------CA-CT-T-AA-----A-----G---......GAC--T---------C-GATC-----A---------TA-A--G-----CAGA---...ACA--T   8908
03_AB.RU.KAL153-2   --CAG-G----------A----...C-A---------A-AT---AA--C--A--------T......GA---A-A----------AT-----A---------T-----G--TC-AACA---...AGA---   9364
04_cpx.CY.94CY032-3 --C-G---------------T-T...---C-G------A-A----A------A-----G--T......GAA------------A---A------A---------T-----G----ACA-G---...--A---   8793
N.CM.YBF30          T---G-------C----ATGC-...C-C--------CA-AT-T-AA-----AGCA-----T......GATC-T-A---------G-T----C-A---------TC-------AGAAGA--T...--A--A   8971
O.CM.MVP5180        GGG---TAC-A-TG-AGATGCTAGTC-TC----T--A-CTT-TAAT-----AGCT--G--T......G-AC-CG-G--GA-AC--A-A---CA------T--ATCAT---GC--AACA--T...A-A---   9494
O.BE.ANT70          AGGA--TAC-TGTG-G-GGGCTAATC-CC-G--T--A-CAT-TGCC-----AT-T--A--T......A-AC-T-A----A-AC-GAT-----A------T--ATCT----GCAACACC--T...--T--T   9470
CPZ.US.CPZUS        ---------------C--A--TA...C--C-G-----CA--T---A------A-----A--T......GAAC-TG-C-----C---ATC---CA---------CTGAA--G--TCGGAGA---...AGA--T   9420
CPZ.GA.CPZGAB       ----G-------T----A-T--...C----G--T--CA-AT-T-A------------A--T......GA---C-A---G--C--G-TC---C-C----------A-G--G--C--AAGA--T...A-T---   9454
CPZ.CM.CAM3         ----G-------T--T-A--T-...C-AC-G-----CA-CT---A------A-----A--T......G-AC-TG-------C--G-TC-----A----------A-G--G--C--AAGA---...A-A---   8824
CPZ.CM.CAM5         ---CG-------C-C--AT-TA...C--C-T-----CA-AT---AA-----A-----G--T......GAAC-TG-------C----TC-----A----------A-G--G--TC-GA-A---...C-A---   9127
CPZ.CD.CPZANT       .......-AT----GG-A--TA...C-CC-C-------CTT-TACA-----AGAT-GA---......--AC-T-A---GA-C---AGA---GAA-----TGCAA----GATGAGAAGA---...A-T--A   8820


H2A.GW.ALI Nef      __T__E__T__.__.__H__C__.__L__L__H__P__V__Q__T__S__R__H__D__D__.__.__T__H__G__E__T__L__V__W__R__F__D__P__K__L__A__H__D__Y__.__K__A_   Nef
H2A.GW.ALI          --C-G-GACT......CA-T--...C--C-------A--ACAAACAAGCA--CAT------......A--C-TG-G--GACA----TT-----A------CCTAAG--G--TCA-G--T--...AAA---   9793

H2A.SN.HIV2ST       --G-G-GACT......CA-T--...--AG-G--T--A-CACAAACAAGCA--T-T------......---C-TG-----ACA----TT------------CC-ACG-----T---AGCT--...-A----   9233
H2A.DE.BEN          AG--G-G-CC......-A-T--...--AG-------A-CACAAACAAGCA-ACAT-----T......GA-C-TG-G--GACA----T----CA-------TC-ATG--G--C-A-A-CT--...AA----   9794
H2B.GH.D205         GG-GG-G-A-ACC...CATT--...C-AA-G-----G-CACAGA-CTCCTCATG-------......ATCC-TG-G--GACTC-TATC---CA------TTC--T---G----A-G--T-T...------T   9710
H2B.CI.EHO          -G-GG--ACC......CATT-T...C--G-G--T--A-CACAGACCTCC-CATG-------......--CC-CGA---GACCC-T-TC---CA-------TC--T---------A-G-CT-T...-------   9681
H2G.CI.ABT96        GG-CG-GACT......CATT--...---G----T--A-CACAAACTAGCCCATG-------......-------CC---------G-C-----A-------CCTACAT----C-A-G-CT--...AG---A   9179
MAC.US.MM251        GG--G-GA-G......CATTAT...--AA-G--G--A-CTCAAACTTCCAA-TG-------......--TTG-G----G--TC---C-----A------TCCAACT-----C-ACAC-T-T...-A---A   9733
SMM.US.SIVSMMH9     AG-CG-GAC-......CATT--...---A-G--T--A-CACAAACTTC-CA-TG-------......--CTG-G----G--AC-G-CA----A------TCCARAGT----T-A-R-CT-T...AA---A   9234
STM.US.STM          GG--G-T--GACA...CATTAT...C--G-G--T--A-CACAGACA---CA-TG-------......--TTG-G-------AC-G-TT----AA-----TCCA-TAT----CCA-AC-T-T...-A----   9405
TANTALUS.UG.TAN1    TG-G-G-.........CA-T-T...---C-C-------CCCAGA-TG-AT--GAAAGC---......--TTG--A---GACA--G-T-----AA---A-TCCA-T--------G----GT--...AACC-G   9366
GRIVET.ET.GRI-677   -TG-G-GA--......CA-T-T...C--A-G--T--A-CACAGA--GGG-AAGATCC---TGGAATAGATC-TG--------C--G-TC----A--------CCGAAGT-G--GG-GG-GT--...CGCC-G   9062
VERVET.KE.VER-9063  -TG-G-GA--......CATTA-...---GCG--T----CACAAA-AA-A-A-GACCCC--TGGCATCAGTC-TG----GACC--G-TC----A--------CCTATGT-----G-G--GT-T...-ACC-A   9388
VERVET.KE.AGMTYO    ATG-G--A--......CATT--...C-AG-C--T--A-C-CAGA--GGA-AAGATCCA--TGGTATCAGCC-TG----GA-C--G-T-----A------TCCTATGT-G---A-A--GT--...-ACC--   8842
VERVET.KE.AGM155    ATG-G--A-G......CATT--...---G-------A-CACAA-------AAGACCC---TGGTATAAATC-TG-----A-A--G-CA----A------TCCAATGT-G--TG----GT--...-ACC--   9365
VERVET.DE.AGM3      TTG-G-GA--......CATT--...C--G-G--T--A-CACAAG--AGG-AAGACCC---TGGAATCAACC-TG--------C--G-TC----A------TCC-ATGT-----G-C--AT-T...-ACC-T   8886
SABAEUS.SN.SAB1C    T-G-G-GAAC......CA-T--...C--C-G--T----CTCAAG--GCCTATGA-------......G-CTG--A---GACCC-G-T-----A-------CCA-TA------G-GG-CT-T...--T--T   9559
SUN.GA.SIVSUN       ........................CCTAAT--T---TGTCAAGCC-TACT-CATTCCAG-...CAG-AA-G-GTTA-T-AAGACCCT---G-AGAGAGACTTAT-TGGA-----G---C-ACTC-----   9515
MANDRIL.GA.MNDGB1   .T--GC--CTA-C--TCAT......C--------T-CTCCCAGT-ATG-CCTCA---A--T......--A--AG----GACCC-CAT----TCTGGGAC-CT-AT-T-GC-TA-GAC-TTG...CA-CAT   8811
LHOEST.KE.hoest     -TCAG-T-AGACA...CAAT--...C-T--G--TT-AAGCCAG---GGG-TCC----A---......--CTG-G----GAG-C-TATC---CAC--CA-TCCAATGT-G---G-GG-CTTT...A-A---   9459
LHOEST.CD.hoest447  -TCAG-TTACAGA...CAGT--...C--C-G---T-AAGTCAG---GGG-TCCA---G---......--CTG-G-----AG-C-TATC---CAC---A-TCCTATG------G-GG-CT-T...--A---   8384
LHOEST.CD.hoest485  -TCAG-T-CCAGC...CAAT--...C--C-G---T-AAGCCAG---GGG-TCCA---G---......--CTG-G----GAG-C-TATC---CAC--CAGTCCTATG------G-GG-CTTT...--A---   8381
LHOEST.CD.hoest524  -TCAG-T-AGAGA...CAAT-T...---C----TT-AAGCCAGT--GGG-TCC-A--A---......--CTG-G-----AG-C-CATC---CAC---A-TCC-ATGT-----G-GG-CT-T...--A---   8384
RCM.NG.SIV-NGM      G---G-GA-G......CATT-T...C-CC-G--T--A-CAGAAACTAG---------C---......--CTG-G-G--G--A----CA----A----A-ACC--T-T-----G-AG--T--...-CA-GG   9847
COL.__.SIV-CGU1     --GCT-T...........................................................--CTTGAGTC----GAGCGTT-CAAGCAGG-AGAAGAAGAA-CCA...............   8575

SYKES.KE.SYK173     TG-GG-G.........CAT-CACTCC-CC-C--T--A......GCCTC---ACA--GGAGTTCGAGCAT--GAGA-CCTCATG-G--ACTTCAACCCCCACCTG--TATA--CCAGG-TGG...-A-ATG   9204
SYKES.KE.SYK173 Nef __E__E__.__.__.__H__T__L__L__L__H__P__.__.__A__S__G__Q__G__S__S__S__M__G__E__P__H__V__E__L__Q__P__P__P__G__Y__T__P__G__W__.__E__M_   Nef
```

```
                                                              Nef end <-|
B.FR.HXB2 Nef       _.__.__.__.__.__R__E__L__H__P__E__Y__F__K__N__C__.__$                                                                                Nef
                                                        TCF-1 alpha
B.FR.HXB2           ...............CGAGAGCTGCATCCGGAGTACTTCAAGAACTGC...TGA............................................................................   9417

A.UG.U455           ...............TAT----------------T--AT--AG----A..................................................................................   8865
C.ET.ETH2220        ...............--C-----A-----------T-A---AG-----...---............................................................................   8817
D.CD.84ZR085        ...............AA---AAA-T-----------------------...---CATCGAGCTC..................................................................   8975
F1.BE.VI850         ...............A-----AGA----------T--A-C-AG----A..................................................................................   8691
G.SE.SE6165         ...............----------------------A---AG-----...---............................................................................   8859
H.CF.90CF056        ...............----TAAA-------------...--AG-----...---............................................................................   8737
J.SE.SE7887         ...............-------------------TT-A---AG-----...---............................................................................   8715
K.CM.MP535C         ...............-----------------C----A---AG--.....................................................................................   8598
01_AE.TH.CM240      ...............-----A-A------A-------AT--AG-----...---............................................................................   8985
02_AG.NG.IBNG       ...............A---------------------A---AG-----...---............................................................................   8944
03_AB.RU.KAL153-2   ...............-------------------T--A---AG-----...---............................................................................   9400
04_cpx.CY.94CY032-3 ...............-------------------TT-A---AG-----...---............................................................................   8829
N.CM.YBF30          ...............A------------------TT-A----------...---............................................................................   9007
O.CM.MVP5180        ...............-TGC-AAA---C--A---CT----CCC--G-AA..................................................................................   9527
O.BE.ANT70          ...............ATGATAACT--C--A---CT----C----GGA-...-A-.....................................................................AAACTGC   9513
CPZ.US.CPZUS        ...............AA------------A---------CG------ATGGC..............................................................................   9457
CPZ.GA.CPZGAB       ...............A----A-AA-------------A---AG----ACTT-AGACTGGCGCATGCGCACAAGAACTGC...................................................   9518
CPZ.CM.CAM3         ...............A----A-AA-----------T-A-C-------A..................................................................................   8857
CPZ.CM.CAM5         ...............A----A-AA-----------T-ATGGAG--CA-AACCC-GCGCCTGCGCAGTAGAACTGC.......................................................   9187
CPZ.CD.CPZANT       ...............A----AAGA--C--T--A--T----GAG---AAGAC--C............................................................................   8859

                                                                                                                                     Nef end <-|
H2A.GW.ALI Nef      _.__.__.__.__.__F__I__L__H__P__E__E__F__G__Y__K__.__S__G__L__P__E__D__E__W__K__A__R__L__K__A__R__G__I__P__F__S__K__N__R__N__S__$    Nef
H2A.GW.ALI          ...............TTTATT--A--C--A---G-A--TGG-T--AAG...-C-GGCCTGCCAGAAGATGAGTGGAAGGCAAGACTGAAAGCAAGAGGGATACCATTTAGTAAGAACAGGAACAGCTGAT   9905

H2A.SN.HIV2ST       ...............TTTATT-GAT-C--A---G-G--TGG-T--AAG...-C-GGCCTGCCAGAGGATGAATGGAAGGCAAGACTGAAAGCAAGAGGGATACCGTTTAGCTAAAAACAGGAACAGCTAT   9345
H2A.DE.BEN          ...............TTCACT---T-C--A--AG-G--TGG-C--AAG...-C-GGATTGCCAGAGAAAGAATGGAAGGCAAAACTGAAAGCAAGAGGGATACCATATAGTGAATAACAGGAACAACCAT   9906
H2B.GH.D205         ...............TTCA-TAG-TT---A--AG-G---GG-T--CAG...-C-GGGCTACCAGAGGAGGAGTGGAAGGCTAGACTAAAAGCAAGAGGGATACCTACAGATTAGGCAAGAGACAGCAGCA   9822
H2B.CI.EHO          ...............TTCAGCAG-TTC--A---G-G--TGG-T-TCAG...-C-GGAATGCCAGAGAAAGAGTGGAAGGCTAAACTGAGAGCAAGAGGAATACCTACAGAGTAGACAGGAAATAACAGCA   9793
H2G.CI.ABT96        ...............TTTATAT-------T---G-G----G-TGGAAGTCAG-GCTACCAGAAGCAGTATGGAAAGAGAAACTAAAACAGAGAGGACTGCCTATAGAATAAGCAGAAAGAGG........   9286
MAC.US.MM251        ...............TAT-CTAGAT-C--A--AG-G--GG-AGCAA-T...CAGGCCTGTCAGAGGAAGAGGTTAGAAGAAGGCTAA...........................................   9802
SMM.US.SIVSMMH9     ...............TTT-TTGA---C--A--AG-G--TGGT-GTCAG...-C-GGCTTGTCAAAGGAAGAGGTGCAGAGAAGGCTAACCGCAAGAGGCCTTTTAAAAATGGCTGACAAGAAGAAAACAA   9346
STM.US.STM          ...............TTT-TTAGA--C--A--AG-G--TGGTTCTAAG...-C-GGCTTGCCAAAGGAAGAGGTTGAGAGAAGGCTAACCGCAAGAGGCCTTCTAAAGATGGCTGACAAGAAGGAAACAA   9517
TANTALUS.UG.TAN1    ...............GACAGCT-TA-GGACAT-C-TGGTCTTGTTAAG...A-GAAGTAG....................................................................CC   9410
GRIVET.ET.GRI-677   ...............GACAT-T-TA-GGACAT-C--GAAC-TGCAAAG...C-CTAG.........................................................................   9101
VERVET.KE.VER-9063  ...............AACAGA-A-T--TTT--AG--A-GC-TGCAATA...GTGAAGAGGAAGTAG................................................................   9436
VERVET.KE.AGMTYO    ...............AATCG-GA-T-CTTTACTG--A-GC-TGGGCTG...GTGAAGAGGAAGTAG................................................................   8890
VERVET.KE.AGM155    ...............TC-AG-GA-T-CTTTACAG----AT-TTCAACA...GTTGGTACAGGAAACTAG.............................................................   9416
VERVET.DE.AGM3      ...............AACAGAAAAT---TCACTG--A-GC-TG-TCTT...G-CAAGAGGAAGTAG................................................................   8934
SABAEUS.SN.SAB1C    ...............T-GAG------C--A---C-GG-GCC-TCAGCA...CA-GGCTAA......................................................................   9601
SUN.GA.SIVSUN       TATGATTTTAGGGCCAT-C-AAA---C--T--AG-G------C-TGTTACTA-TCTGCAGTGGGAGGTGTAGCTTAACCGCAAACCGCATCCTCTTGCATCGCCTAGGCAACG.................   9628
MANDRIL.GA.MNDGB1   ...............TA-C-TA-CACC-T-AGTGTT-CA-TA-GGCTAAGAGT-TTGAACATCTGCCATTTTGGAAGAGGAAGTAGCCTAACCGCAAAACCACATCCTACTGCAGAACTGTAG.TTGCTT   8925
LHOEST.KE.hoest     ...............-TCA-AAA---G--T-C-A-GA-AC-A--TACTGCT-TCGCTTTT...............GACTGCAAGAGGAAGTAGCCAAACCGCAAACCGCAGCACATCCTCTTGCAGCCCG   9559
LHOEST.CD.hoest447  ...............-TCAGAAAA--A--TA-AAG-A-AC-AGCA-CAGCTATTGGCTTG.....................CAGAGGAAGTAGCCAAACCGCAGACCGCAACACATCCTCTTGCAGCCTG   8478
LHOEST.CD.hoest485  ...............-TCAGAAAA--A--TA--AG-A-AC-AGCA-CATCTCTTAGCTTG...............AACTGTAAGAGGAAGTAGCCAAACCGCAGACCGCAGCACATCCTCTTGCAGCCTG   8481
LHOEST.CD.hoest524  ...............-TGAGAAA---A--A-CAAGTG-TC-AGCAACTGCTATGAGATTTAACTGCGAGAGGAAGTAGCCAAACCGCAAACCGCAGAACATCCTCTTGCAGCCCGGTTGCCCTAGCAACC   8499
RCM.NG.SIV-NGM      ...............TATAGA--------T----T---TGGA........................GAGAGAAAGAACAAGACCAAGTAACCA...CTTCCTCTGGAGTTGCCTTGGAAACAGCGCATGC   9935
COL.__.SIV-CGU1     ...............-A-C--G-TGCCATA---ATGG-GG-CTGA.....................................................................................   8605

SYKES.KE.SYK173     ...............GCCCGCT----GTTA---AGACAA-CAGGAAAACCACA-GAGCTGCAGTCAGCACTTAGTAAGAACATCTCCTAGGAG...ACTCCATGGTGACAAGCTCGGCCCAGGGGA....   9312
SYKES.KE.SYK173 Nef _.__.__.__.__.__A__R__L__Q__L__E__R__Q__T__G__K__P__Q__E__L__Q__S__A__L__S__K__N__I__S__$                                         Nef
                                                                                          Nef end <-|
```

```
                                                                                       NF-kappa-B-II                     NF-kappa-B-I  |<Sp1-III >|  |->Sp1-II
B.FR.HXB2           ..........................................CATCGAGCTTGCTACAA.........GGGACTTTCCG.CTG...................GGGACTTTCCAGGGAGGCG.TG..GCCT    9473

A.UG.U455           .............................AAG.ACTG.CTGA--CA--AG-----GAC-.........----------A.---...................----------GA--GA-G..--T.-GT-    8932
C.ET.ETH2220        .......................................CACAGA--G-AC-TTCCGCC.........----------A.---...................---.-G-------AG-AG-G....-T--    8874
D.CD.84ZR085        ..................................................................................................................................    8975
F1.BE.VI850         .............GACTGCTGACACAGAGATT.GCTG.ACACAGAA--ATC-..A-A-..........----------A.---...................------------A-G-CG-TCC..AGAG    8772
G.SE.SE6165         .......................................CACAGAA-TTGC--AC-AG..........----------AC--....................----------G--------CG-...---    8917
H.CF.90CF056        .......................................CACAGAA--T.C-..G-A-..........-----------.--....................----------G-------.G--A..T--    8791
J.SE.SE7887         .......................................CAAAGAA-TTTC-..AG-G..........-----------.---...................--------------GA-GAG--..----    8772
K.CM.MP535C         ..................................................................................................................................    8598
01_AE.TH.CM240      .......................................CAAGGAA-TTTC-.ACT............A-A----.---.---...................--------------GA-GTG--..---G    9040
02_AG.NG.IBNG       .......................................CACAGAA-TTGC--AC-............----------A.---...................----------.---GA-GTG-...-GT-    8999
03_AB.RU.KAL153-2   .........................CACAGAC.TGCT.GACATGAA-TTGC--AC-............-----------.-.-...................--------------GA-GTG-...-GT-    9467
04_cpx.CY.94CY032-3 .......................................CACAGAA-TTGC--AC-A-..........-----------C-C....................--------------GA-GCGC-..----    8888
N.CM.YBF30          .......................................CAAGGGACTTTACTGCTG-CAA.......--------ATACT--....GGGACTTTCCGCCA.--------------.A-GTG-...-GT-    9082
O.CM.MVP5180        ...................................................CTGACACTGC.......----------AGAC.TGCTGACACTGCG......------------C-T--GAGG-ATAAGG    9592
O.BE.ANT70          TGACCTGAAGATTGCTGACACTGT............................................--A-------AG-AAAGACTGCTGACACTGCG..------------..T--GAGG-ACAGGG    9595
CPZ.US.CPZUS        .......................TGTAACCGC.GCAGGCGCAATA.A-A--GCTG--TG.........A--------TAA.......GGGACTTTCCAA...-----G-----A--G.-G..--..-GTC    9538
CPZ.GA.CPZGAB       .......................TGACTCTG..................-..................----------AA......................----------G-----AG-.-...-GTC    9563
CPZ.CM.CAM3         .......................TGCCCAGCG.CCTG.CGCAGTAGA-A--GCTG----.........---------TAA......................-----------A--G--G..--..-TTC    8927
CPZ.CM.CAM5         .......................TGACAA.......................................---------TAA......................-----------A--G--A..G-...-TC    9228
CPZ.CD.CPZANT       .......................TAACAGGCG.GATATAAAA-CG-TGA-AGAGC-G-CTTGTGGTCT----------ACTAT...................----G-GA-T-.................    8929

H2A.GW.ALI          TTGGTCAGGGCAGGAAGTAACTACTGAA..........AACAGC-GAG.AC---A.............----------AGAA-....................--G--G-.A-CCAG--GAAG-ACATGG    9990
H2A.SN.HIV2ST       ACTTGGTCAGGGCAGGAAGTAACTA.............ACAGA-AAC-.GC--AG--TGCA.......----------AGAA....................---G--G-T.-CCAG--GAGG...-A-A    9430
H2A.DE.BEN          ACTTGGTCAAGGCAGGAAGTAGCTA.............CTAAG-AAC-.GC--AGG-TGCA.......----------AGAA....................---G--G-.A-CCA---GAGG...-A-A    9991
H2B.GH.D205         TAAACAGGAACTAGCTGACACTGCACAA..........GAAGG-AACTAGCA-AC--TGCA.......----------AAAA....................-A-G--G-.A-CA.T--GAGG...AG-A    9910
H2B.CI.EHO          CAAAGAGGAACTAGCTAACACTGCATAGA.........GAAGG-AACTAGC--A---TGCA........---------AGAA....................---G--G-.A-CA-T--GAG....-GAC    9881
H2G.CI.ABT96        .............................................AC-AGC--AC---GCA.......----------AGTA....................T--G--G-.A-CA-.--GAG.....GTA    9337
MAC.US.MM251        ..........CCGCAAGAGGCCTTCTTAACAT.GGCT.GACAAGA.......................----AAC--GCTGA-..ATAGCA...........-----------CAAG--GATGTTATGGG    9884
SMM.US.SIVSMMH9     GCTAG...........................................-ACA--A.............----------ACAAA...................---GA-G-T.-C.-G--A-G-TCT-GGG    9399
STM.US.STM          GCTGA...........................................-ACA--A.............----------ACAA-....................--G--G-.A-CA-G--A-G-ACT-GGA    9570
TANTALUS.UG.TAN1    TAACCGCAGGCTTGCGGTTAAGCTAACTAGGC.CGTT.GCCTAGGA--TGACATA-AG-AGT.........G--GAT-AG-A....................--------------T--T-CG...AGGG    9506
GRIVET.ET.GRI-677   .........................TGTCAGC.ACTT.TGCGGT-.......................-----------C-A....................-----------CA-T--GTGGAT.CGGA    9160
VERVET.KE.VER-9063  ....CTAACCGCAGGCTTGTGGTTAAGCCGTTGCCAG.GGAGATGAC-TT-GAAACTGCTGACAG...----------AA......................-----------.............-GGC    9523
VERVET.KE.AGMTYO    ....CCAGACCGCAAGCCTGCGGTTAGAACAT.CACC.ATGGAGAT--CA--AAA-ACTGCTGAC...----------AG-GAA..................-----------A-...........--GG    8982
VERVET.KE.AGM155    ......CCGACCACAGGCTTGCGGTTTCCTGGTTGCC.TAGGAGAT--CA--AAG-ACTGCTGAC...----------AG-AC...................-----------A--C--...........    9506
VERVET.DE.AGM3      ......CTAACCGCAGGCTTGTGGTTAAGCACATCACCATGGTGAT--CA--AAG-ACTGCTGAC...----------AG-AA...................--------------C--G.TCATG-G-G    9035
SABAEUS.SN.SAB1C    ...................CCACAAAACCACA.TCCT.ATGGAG-T-TCA-G-TG-TG-CATTAA...-AACTGC-GACTGA-A..................-------------...........-TGG    9678
SUN.GA.SIVSUN       ......................................GGGCT-G--CATGC--GCT-GAGTA.....-----------GACAGGG................................A-G--GAG.-CTTA..    9684
MANDRIL.GA.MNDGB1   GGCAACCTGCTTAGCAACCTGGACTGGCGCTTGCGCG.CTAGG-A.......................----------..................................................AA    8981
LHOEST.KE.hoest     GTTGCTAAGGCAACCGGGCTAGCGCATGCGC.......AATGGC-T-CTGA.....-G-GCA......----------AGGACGGGCGG............G--AGG---G...................    9640
LHOEST.CD.hoest447  GTTGCCAGGGCAACCAGGCTAGCGCATGCGC.......AGGA-T-T-CTGA.....TG-GCA......----------AGGACGGGCGG............G--AGG---G...................    8559
LHOEST.CD.hoest485  GTTGCCAGGGCAACCAGGCTAGCGCATGCGC.......AGGA-T-T-CTGA.....T--GCA......----------AGGACGGGCGG............G--AGG--CA...................    8562
LHOEST.CD.hoest524  GGGCTAGCGCATGCGC......................ACAAGC-T-CTGA.....AC-GCA......----------AGGACGGGCGG............G--AGG---G...................    8565
RCM.NG.SIV-NGM      CTACT.................................GAGAAGAT-CTGA....CAC-GAA......----------ACTG-CGCCTGCGCGCTGGTGTAA------------ACT-A--.....TGGG   10017
COL.__.SIV-CGU1     ...........................................G-T-CTGA....GAC-GCA......-----------TTC....................-----------A-TT--GAGG.......    8655
SYKES.KE.SYK173     ....................................................................--AG-C-GGGCGGG-G..............................................    9328
```

```
                                                                                                        3’LTR U3 end \/ 3’ LTR R repeat start
                                                                                                                     TAR element stem          bulge
                  ->Sp1-II  <-|   |->Sp1-I <-|                                  TATA Box                            |||| |^||||  |||||||^  |||||^^^|||
B.FR.HXB2           GGG...CGGGACT.GGGG.AGTGGCGA......G......CCCTCAG.ATCCT....GCATATAAGCAG..CTGC..TTTTTG.CCTGT...AC..TGGG.TCTCTC..TGGTTAGA..CCAGATCTGAG   9566

A.UG.U455           T--G..---AGT-.----.GA.-TG-C......TAG....-------.--G--....------------..----..----C-.-----...--..----.------..-T------..------.C---   9025
C.ET.ETH2220        ---...-------.----.------C-......A......-------.--G--....------------..----..----C-.-T---...--..C---.------..-A-G----..-----------   8966
D.CD.84ZR085        ..................................................................................................................................   8975
F1.BE.VI850         --....-------.----.------TC......A......-------.--G--....------------..-C--..----C-.-----...--..----.------..-A------..------T----   8864
G.SE.SE6165         ---...A--AGT-.----.------T-......A......-------.-AG--....------------..-C--..--C-C-.-----...--..----.------..-T------..------T----   9010
H.CF.90CF056        ---...----GA-.----.------C-......A......-------.--G--....------------..----..----C-.-T---...--..----.-----T..--------..-----------   8884
J.SE.SE7887         ---...----GT-.----.------T-......A......-------.--G--....------------..----..----C-.-T---...--..----.------..-T------..-----------   8865
K.CM.MP535C         ..................................................................................................................................   8598
01_AE.TH.CM240      ---...---AGT-.----.---A--T-......A......-------.--G--....-----A------..-C--..----C-.-T---...--..----.------..-T------..----G--.---   9132
02_AG.NG.IBNG       T--G..A--AGT-.----.------T-......A......-------.-C-G-....------------..----..--C-C-.-----...--..----.------..-T-C----..-----------   9093
03_AB.RU.KAL153-2   T--G..---AGT-.----.------T-......A......-------.--G--....------------..----..----C-.-----...--..----.------..--------..-------.---   9560
04_cpx.CY.94CY032-3 ---...A---GT-.----.------T-......A......-------.--G--....-----A------..-C--..--C-C-.-----...--..----.------..--------..-----------   8981
N.CM.YBF30          ---GGAGT--CT-.-.........................-------.-.G--....-----A------..-C--..--C-C-.-T---...--..----.------..-T-C-G--..------TA---   9165
O.CM.MVP5180        --CGG......T-.C---G------T-......A......-------.--G--....------------..----..---CC-.-T---...--..C---.---TAG..-TAGAG--..----G------   9683
O.BE.ANT70          --CGG......T-.C---G------T-......A......-------.-AG--....------------..-C--..---C--.-T---...--..C---.----GG..-TAGAG--..----G------   9686
CPZ.US.CPZUS        .A-G..GC---ACA---C....-TG-T......TTTA...------A.-GA--....-----A------..-C--.T.C-C--.-T---...--..----.-----...-C-C----..------AA---   9631
CPZ.GA.CPZGAB       ---AGG----GT-G---T....-TG-C......TTTG...-------.-.G--....------------..----.T.--C--.-----...--..----.-----T..-CAC-G--..------T----   9659
CPZ.CM.CAM3         A--G..-------.---C....-TG-T......TTTA...-------.-AG--....------------..-C--.T.C-C--.-T---...--..----.------..-T-C----..-----------   9021
CPZ.CM.CAM5         A--G..------A.---C....-TG-T.....TTCAG...TTT----.-.........................................................................   9261
CPZ.CD.CPZANT       .......T--G-G.--CA....-TG-GAGTGGCTTTC...----G-CT..G--....------------..----.T.--GC-.-T---...TA..A---C-----G..CCTAGGAG..-------A---   9024

H2A.GW.ALI          -A-GAG-T--TGG.--AA.C-...................---.TCAT-.-TC....CTG-----ATGTAC-C--TGC--.GCAT.---..ATTC.A-TC.G----GC.G-AGAG-CTGG-----.----  10085
H2A.SN.HIV2ST       T--GA...---GC.C--T.G-G-AACG.............------TAC.TT-....CTG-----ATGTAC-C--.TAC-CGCAT.---..ATTC.A-TC.G----GC.G-AGAG-CTGG-----.----   9529
H2A.DE.BEN          T--GA...---GC.T--T.G-G-AACG.............------TAC.TTA....CTG-----ATGTAC-C--.T-C--GCAT.---..ATTC.A-TC.G----GC.G-AGAG-CTGG-----.C---  10090
H2B.GH.D205         A--GA...--GG-.T--T.T-G-AACG.............---ATTAA-.T--....CTG-----ATGTAC-C--.T-C-CGCAT.---..ATTC...................................   9978
H2B.CI.EHO          AT-GG...A-GAG.CA--.GAG-AACG.............---ATCAAG.A--....CTG-----ATGTAC-C--.T--G-GCAT.---..ATTC.A-TC.G----GC.G-AGAG-CTGG-----.C---   9980
H2G.CI.ABT96        CTATG...---GG.-ACT.G--T-G--......ACG....---C.-AA-.T--....CTG-----AT-TAG----.A---CGCT-.---..ATTC.A-TC.G----GC.G-AGAG-CTGG-----.----   9438
MAC.US.MM251        -A-GAG-C--T-G.--AAC-....................---A-TTTC.TTG....ATG-----AT-TCA----.A---.C-CT----..ATTC.A-TC.G----GC.G-AGAG-CTGG-----.----   9980
SMM.US.SIVSMMH9     A--AGCT.--CTG.-AAC.G....................---A-TTAT.T--....CTG-----AT-CAA----.A---.C-CT----..ATTC.A-TC.A----GC.G-AGAG-CTGG-----.----   9493
STM.US.STM          --AACTG.-TGGG.-AAC.G....................----TTTAC.TT-....CTG-----AT-CAA-C--.T-A-.A-CT----..ATTC.A-TC.A----GC.G-AGAG-CTGG-----.----   9664
TANTALUS.UG.TAN1    C--TA.....CT-..---.------TT............C-------.-.G--....-----A------..A---..-CGC--G-T---..A--..-CA-.-----T.ACTAGG---..----C-.----   9597
GRIVET.ET.GRI-677   --CGGTACA-GGG.C--TACTG--A-T......-GCTTTC-------.-.G--....-----A------..A---..-CGC--G-T---..A--..-CA-.-----T.ACTAGG---..----C-.A---   9264
VERVET.KE.VER-9063  ---CCAT.--G-G.-TAC.G-G-AGTG......-TT.TTA-------.-.G--....-----A------..A---..-CGC--G-T---..A--..-CA-.-----T.ACTAGG---..----C-.----   9624
VERVET.KE.AGMTYO    -ACAT...--G-G.-TCC.G-G-AGTG......-CT.TTA-------.-.G--....-----A------..A---..-CGC--G-T---..A--..-CA-.-----T.ACTAGG---..----C-.----   9081
VERVET.KE.AGM155    .....GACATGGG.C--TAC-G--A-T......-GCTTTA-------.-.G--....-----A------..A---..-CGC--G-T---..A--..-CA-.-----T.ACTAGG---..----C-.----   9605
VERVET.DE.AGM3      -TACGGG-A-TGG.CTTTA.....................-------.-.G--....-----A------..A---..-CGC--G-T---..A--..-CA-.-----T.ACTAGG---..----C-.----   9124
SABAEUS.SN.SAB1C    A-ACT.G--CGG-.ACT-GGAGT-GCT......TGC....---.GT-CT.G--....CA----------..----..---.GCG-----..A-A..ACA-.-----GC.-T-GAG-CTG------.----   9778
SUN.GA.SIVSUN       ........................................---A-..TT.A--....------------..----TC.--CGC.-.----...TCGG-A-.-----GC.-CAGAG--TAGGC--C...GA   9753
MANDRIL.GA.MNDGB1   ACAGG...-AGGG.--A-.GC-C-................---CA..TG.-TG....CT----------..----AT.--CGC.T.----...TCGG-A-.-----ACTACAGAG-CTAA..........   9063
LHOEST.KE.hoest     ........................................---A-..TT.G--....------------..----TC.-CCGC.T.----...TCGG-A-.--CT-GC.-CTGGG--CCAGCAG-.C.GC   9710
LHOEST.CD.hoest447  ........................................---A-..TT.G--....------------..----TC.-...................................................   8588
LHOEST.CD.hoest485  ........................................---A-..TT.G--....-----A..----..----TC.-...................................................   8589
LHOEST.CD.hoest524  ........................................---A-..TT.G--....------------..----TC.-...................................................   8594
RCM.NG.SIV-NGM      A--GG...--TG-.--CT.T-...................-----TCCT.G--....------------..----AT.--CGCT-.---..A--C.A-TC.---.C-C.---GAG-CTA------.----  10107
COL.__.SIV-CGU1     ...........GACCT--GC-GA-G--AGGGCGTG.....----GCTA-A.--GCTT-T---A---TCAG.AGC-TCGAGAGCTT.C-AGGC......................................   8728
SYKES.KE.SYK173     ...........AA.--CT.CA...................----......A--....-----A--.....G-C-A.....-GCT-.----.AG-T.ATCCGCGGAGTTGCAAC---CTCT-TCCA..-C-   9400
```

```
        TAR element   loop        stem
                    |       | |||||||||    |||                           ||||                  |||||||    |||
B.FR.HXB2           C.CTGGGAG.CTCTCTGGC....TAA...........................CTAG..................GGAACCC....ACTGCTTAA...................................   9606

A.UG.U455           -.-------.---------....--G...........................-G--..................-------....---------...................................   9065
C.ET.ETH2220        -.-------.---------....--T...........................--G-..................-------....---------...................................   9006
D.CD.84ZR085        ..................................................................................................................................   8975
F1.BE.VI850         -.-----G-.---------....--G...........................--G-..................-------....--------....................................   8903
G.SE.SE6165         -.-------.---------....--G...........................-AG-..................-------....---------A..................................   9051
H.CF.90CF056        -.-------.---------....--G...........................---A..................-------....---------...................................   8924
J.SE.SE7887         -.-------.---------....---...........................----..................-------....---------...................................   8905
K.CM.MP535C         ..................................................................................................................................   8598
01_AE.TH.CM240      -.-C-----.---------....--G...........................-A--..................-------....---------A..................................   9173
02_AG.NG.IBNG       -.-------.---------....--G...........................-GGA..................-------....---------...................................   9133
03_AB.RU.KAL153-2   -.-------.---------....--G...........................---A..................-------....---------...................................   9600
04_cpx.CY.94CY032-3 -.-------.---------....--G...........................----..................-------....---------...................................   9021
N.CM.YBF30          T.-------.-ATAT---G...............................................................................................................   9182
O.CM.MVP5180        -.-C-----.---C-----....CT...........................CTAGC..................T------....G--------...................................   9723
O.BE.ANT70          -.-C-----.---C-----....CT...........................CTAGC..................T------....G--CG----...................................   9726
CPZ.US.CPZUS        -.-C-----.---------....--G...........................TGTA...................------....---------...................................   9670
CPZ.GA.CPZGAB       -.-------.------AC-.....-G...........................TG-A..................-------....---------...................................   9698
CPZ.CM.CAM3         -.-------.--------T....C-G...........................G---..................A------....---------...................................   9061
CPZ.CM.CAM5         ..................................................................................................................................   9261
CPZ.CD.CPZANT       -.-------.--------TA.AT-TTT..........................GGCAG.................-AG----......-----.....................................   9065

H2A.GW.ALI          -C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGTTCC--.................GCTA--CT-TCACC-G-----GGCCGGCACTGGGCAGA....................  10172
H2A.SN.HIV2ST       -C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGTTCC--.................GCTA--CT-TCACC-G-----GGCCGGCACTGGGCAGA....................   9616
H2A.DE.BEN          -C---A---GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGTTCC--.................GCT---CT-TCACC-G-A---GGCCGGTACTGGGCAGA....................  10177
H2B.GH.D205         ..................................................................................................................................   9978
H2B.CI.EHO          -C-------GT-----CCAG.CACT-GC..AGGTAGAGCCTGGGTGTTC.CCTGC...................TA--CT-TCACC-GCA---GGCCGGTGCTGGGCAGAC...................  10068
H2G.CI.ABT96        -C-C-----GT-----CCAG.CACT-GC..AGGTAGAGCCTGGGTGTTC.CCTGC...................TA--CT-TCACC-GCA---GGCCGGTGCTGGGCAGAGC..................   9527
MAC.US.MM251        -C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGTTCC--.................GCTA--CT-TCACC-GCA---GGCCAGTGCTGGGCAGAG...................  10068
SMM.US.SIVSMMH9     -C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGTTCC--.................GCT---CT-TCACC-GCA---GGCCGGTGCTGGGCAGAG...................   9581
STM.US.STM          -C-------GT-----CC...AGC-CTA...GCAGGTAGAGCCTGGGTGTTCC--.................GCTA--CT-TCACC-GCA---GGCCAGTGCTGGGTAGAG...................   9752
TANTALUS.UG.TAN1    -.-------.T--G----T....G-GCCTA..ACCTGGTCTGGCCATCCAGGGG--A..................-AGC-T-....CT------GAAGGCTAT...........................   9670
GRIVET.ET.GRI-677   -.-----T-.T--G----T....--GCCTA..ACCCGGT.TGGCCA.CCGGGGG--A..................---CT--....TTG----C-TATA...............................   9331
VERVET.KE.VER-9063  -.-----T-.T--G----T....--GCCTA..ACCTGGT.TGGCCA.CCAGGGG--A..................---CT--....TTG----GGAA.................................   9689
VERVET.KE.AGMTYO    -.-----T-.T--G----T....--GCCTA..ACCTGGT.TGGCCA.CCAGGGG--A..................---CT--....TTG-----GAA.................................   9146
VERVET.KE.AGM155    -.-----T-.T--G----T....--GCCTA..ACCTGGT.TGGCCA.CCAGGGG--A..................---CT--....TTG-----GAA.................................   9670
VERVET.DE.AGM3      -.-----T-.T--G----T....--GCCTA..ACCTGGT.TGGCCA.CCAGGGG--A..................---CT--....TTG-----GAA.................................   9189
SABAEUS.SN.SAB1C    -.-----T-.T-------T..AAGTCTG.AACCAGCTTGAGCCTGGGTGTTCG--.-GGGAAGGCAAGCCTAAGA-C-CT-TGGTG-GGT---GGTAGACCCTCGCC............TGGCGACTGGC   9890
SUN.GA.SIVSUN       -C---A---G.----ACCTA.GAGCTGG........CTTGTCGCTCCTGAGT.GAGCACTTGCTAAGCACC.TCACCCGAG-CTGAT-A----GG..GGAGCTAGA........................   9845
MANDRIL.GA.MNDGB1   ...G--TT-TA------AGC.AGATCCCCTTAGAGCAAGGACCAGAGT....C--GA............GTGACT.-GGT-TGAGC-..C--C-CTCGGGGCTG....................ATCACC   9150
LHOEST.KE.hoest     TC-----T-A.G-A-ACTGGTCAGCTAA........GCAGTCAGCCCTGGGA.GGCTACA..........CTTAGCACC-A-CTGGG-CTGA-C-GCCTAGGGAGCTAAG....................   9800
LHOEST.CD.hoest447  ..................................................................................................................................   8588
LHOEST.CD.hoest485  ..................................................................................................................................   8589
LHOEST.CD.hoest524  ..................................................................................................................................   8594
RCM.NG.SIV-NGM      -.-----TTGT-------TAAGTCTCTAG..GAGACTAG..CTTGAGCCTGGGTGT.TCGCTAGTTCTCTGAAGG-CTTG-TGGGAG-GC---GGCTGGGTCTTGGTAGACCTTCAGCTGGGTACTTGGC  10231
COL.__.SIV-CGU1     ..................................................................................................................................   8728
SYKES.KE.SYK173     GC------...G-C---AGA.GACTGGTAGAGCGGCCCG..GGAGCCCTTGCCAGTAAAGGCTAGTGAGGTCCGG-TTTGG-CGCCCGGT---CGGCAGCTCTCCGTTGCTT..................   9506
```

```
                          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
B.FR.HXB2           ...............................................GCC.TC.AATAAA..G.CTTGCC.....TTGAG..TG.CTTCAAGTAGTG...TGTGCCCGTCTG.TT..GTG.TGACTCTGG   9669

A.UG.U455           ...............................................---.--.------..-.------.....-----..GTG---G--------...------------.--..A-T.---------   9129
C.ET.ETH2220        ...............................................---.--.------..-.------.....-----..--..............................................   9031
D.CD.84ZR085        ..................................................................................................................................   8975
F1.BE.VI850         ..................................................................................................................................   8903
G.SE.SE6165         ...............................................---.--.------..-.------.....AA-CC..................................................   9074
H.CF.90CF056        ...............................................---.--.------..-.------.....-----..--.---T.........................................   8953
J.SE.SE7887         ...............................................---.--.------..-.------.....-----..G-CGCATGCAAGCC-.................................   8943
K.CM.MP535C         ..................................................................................................................................   8598
01_AE.TH.CM240      ...............................................---.--.------..-.------.....-----..--.---A-........................................   9203
02_AG.NG.IBNG       ...............................................---.--.------..-.------.....-----..--.------------...------------A--..---.---------   9197
03_AB.RU.KAL153-2   ...............................................---.--.------..-.------.....-----..--.-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN.NNNNNNNNN   9669
04_cpx.CY.94CY032-3 ...............................................---.--.------..-.------.....-----..--.----.........................................   9050
N.CM.YBF30          ..................................................................................................................................   9182
O.CM.MVP5180        ...............................................CG-.--.------..-.------.....-----..--.AGAAGCAGT---...--CT-A..----.--CAACC.CTGG-G-CT   9786
O.BE.ANT70          ...............................................CG-.--.------..-.------.....-----..--.AGCTTG--CTG-AAG-T--AT-CCA--T-AGCAGT.CC-A-A--A   9795
CPZ.US.CPZUS        ...............................................---.--.------..-.------.....-----..--CT--A.-------...------TA--C..--C..A-.---------   9732
CPZ.GA.CPZGAB       ...............................................---.--.------..-.------.....-----.TGTAT--G.--C----...------TA---A.--C..A-.AC-------   9762
CPZ.CM.CAM3         ...............................................---.--.------..-.------.....-----..--A---G.--C----...------TA--CA.--C..A-.-A-------   9124
CPZ.CM.CAM5         ..................................................................................................................................   9261
CPZ.CD.CPZANT       ................................................-T.-..............................................................................   9068

H2A.GW.ALI          ................CGGCTCCACGCT.TGCTTGC...TTAA.AGAC-TC-T.------..-.--.---AA...--AGA..A-CAGGTT-AAG---TGT-CCCAT-TCTCCTAGTC-CC.GCCTGG-CA  10270
H2A.SN.HIV2ST       ................CGGCTCCACGCT.TGCTTGC...TTAAAAGAC-TC-T.------.GCT.GCCAG....T-A--..AGCA.............................................   9672
H2A.DE.BEN          ................CGGCTCCACGCT.TGCTTGC...TTAA.AGAC-TC-TC------..-.--.---AG...--AGA..A-CAAGTT-AGT---TGT-CCCAT-TCTCCTAGTC-CC.GCCTGG-CA  10276
H2B.GH.D205         ..................................................................................................................................   9978
H2B.CI.EHO          .................GGCTCCACGCT.TGCTTGCTTAA.AA..CCCT-T-..------..A.--.---TA...--AGA..A-CAAGAC-AGT---TGT-CCCAT-TCTCCTAGTC-CC.GCCTGG-CA  10165
H2G.CI.ABT96        .................GGCTCCACGCT.TGCTTGCTTACAAC.GACCT-T-..------..-CT.GC-AAA...--AGA..A-TAAG.C-AGT---TGTGT............................   9599
MAC.US.MM251        ................TGGCTCCACGCT.TGCTTGC...TTAA.AGAC-TC-TC------..-.--.---ATT..--AGA..A-TAAG-C---GTGTGTTCCCAT-TC--CTAG-CGCC-.CCTGGTCAA  10168
SMM.US.SIVSMMH9     ................TRGCTCCACGCT.TGCTTGC...TTAA.AGAC-TC-TC------..-G--.---AG...--AGA..A-TA............................................   9638
STM.US.STM          ................TGGCTCCACGCT.TGCTTGC...TTAT.ATAT-TC-TC------..-.--.---AAT..--AGA..A-TAAG-.-AGT-G-TGT-CCCAT-TCTCCTAGTC-CC.GC.--GGTC   9850
TANTALUS.UG.TAN1    ...............................................TG-.-..------..-..-----TGC..--A--..A-T--ATCT-.--CAAGTGCCT-ATTG-GCGCCCTTCC.-C-...G--   9737
GRIVET.ET.GRI-677   ................................................G-.--.------....-C---TCGC..--AGT..C-CTA-ATT-G---CAAG--CT-ATTG---CGCCGAGC.CTCTAGA--   9402
VERVET.KE.VER-9063  ...............................................AG-.-..------..ACT-----TGCA.--AGA..GCTTA--TGAGTCAAGTGCCCT-ATTGAC-CC-CAC-C.AAG-AGG--   9763
VERVET.KE.AGMTYO    ...............................................AG-.-..------....------TGCA.--AG...................................................   9170
VERVET.KE.AGM155    ...............................................AG-.-..------....------TGCA.--AGA..GCTTA--TGAGTCAAGTGCCCT-ATTAA--CC-CAC-C.-TGAA-G--   9742
VERVET.DE.AGM3      ...............................................AG-.-..------...T---CG-TGC.A--AG-..CAGAGA-TT--GAGAGACAT----T-GTGCA-CCA-CA.CA-G-GA--   9262
SABAEUS.SN.SAB1C    CATTGCCAGTAGCAG.AGACTCC.......GCTTGCTTGCTTG..ATTG-C--.------..-T------.....--AGT..-TA-AAGT-AAGCAAGTG-------A-T-AT-CCTCA-.GACAA-CCT   9999
SUN.GA.SIVSUN       .................GGCTCC....T.TGCTTTGCAGT.AC.AAGC-TTC..------..-C--.---TTAGT--T-C..--CAAG.C-AGT--TCCCGT................GT.-CT----TT   9926
MANDRIL.GA.MNDGB1   TCGAGGTAGTGGAA.....CTCCTTGCT.TGCTTGCTAT......TGT-TTC..------..-TAACTTAGAA..--AGA..GCA.............................................   9215
LHOEST.KE.hoest     .................GGCTCCTAGCT.TGCTACTTTAAA.....GC-TTC..------..AAG----TCATTG--AGA..AAGTACA-GC-----TCTCA-AGTA.GTCTC--TCTCC.-TCGC-CT-   9899
LHOEST.CD.hoest447  ..................................................................................................................................   8588
LHOEST.CD.hoest485  ..................................................................................................................................   8589
LHOEST.CD.hoest524  ..................................................................................................................................   8594
RCM.NG.SIV-NGM      CAGTCCCAGTGGGAGTAGGCTCC..GCT.TGCTTTCTGGTTAA..TT--TC...------..CT.GCATGCTCTTAGTT-AACCAGAATC-AGT---TGCCCATTTTA---CTCAGT-AAAC--------  10350
COL.__.SIV-CGU1     ..................................................................................................................................   8728
SYKES.KE.SYK173     .........................GCTTTGCTTACG..CTGA..TCAG-CC.A------..-TG-G-TAAACGCAA-TA..ACGTC-GCTC-CA-CTGT-C---TG-GTCACC-CCCGA.A-GTG....   9597
```

```
                                                                                           3′ LTR end <-|
B.FR.HXB2           TAA.CTAGAGATCC................................CTCAGAC..CCTTTTAGTCAGTGTGGAAAATCTCTAGCA.       9719

A.UG.U455           ---.----------................................-------..-AC-A---A-T.GTGTA-------------.       9178
C.ET.ETH2220        ......................................................................................       9031
D.CD.84ZR085        ......................................................................................       8975
F1.BE.VI850         ......................................................................................       8903
G.SE.SE6165         ......................................................................................       9074
H.CF.90CF056        ......................................................................................       8953
J.SE.SE7887         ......................................................................................       8943
K.CM.MP535C         ......................................................................................       8598
01_AE.TH.CM240      ......................................................................................       9203
02_AG.NG.IBNG       ---.-.................................................................................       9201
03_AB.RU.KAL153-2   NNNNNNNNNNNNNNNNN.............................NNNNNNN..NNNNNNNNNNNNNNNNNNNNNN.........       9715
04_cpx.CY.94CY032-3 ......................................................................................       9050
N.CM.YBF30          ......................................................................................       9182
O.CM.MVP5180        AG-......----.........................................................................       9793
O.BE.ANT70          CCCTAAC...............................................................................       9802
CPZ.US.CPZUS        ---.----------................................------ACT-.--AA-A..GTA-GT--------------.       9781
CPZ.GA.CPZGAB       ---.----------................................------.TTA.AA---TAGTCAAG-A---.---------.       9811
CPZ.CM.CAM3         ---.----------................................------A..-.-A---TAAGTA-GT-----------....       9170
CPZ.CM.CAM5         ......................................................................................       9261
CPZ.CD.CPZANT       ......................................................................................       9068

H2A.GW.ALI          -TCGG-GTTC-C-TGAGTAACAAGACCCTGGTCTGTTAGGACCCTTTCTGCTT..TGGGAA-CGG--GCA--------C------.      10353
H2A.SN.HIV2ST       ......................................................................................       9672
H2A.DE.BEN          -TCGG-GTTC---TGAGTAACAAGACCCTGGTCTGTTAGGACCCTT---GCTT..TGGGAATCCA--GCA--------C------.      10359
H2B.GH.D205         ......................................................................................       9978
H2B.CI.EHO          -TCGG-G......-TCCACTAGAAACCCTGGTCTGTTAGGACCCCTTCTGCTT..TGGGAA-CCA--GCA--------C------.      10242
H2G.CI.ABT96        ......................................................................................       9599
MAC.US.MM251        CTCGG-..-CTCGGTAATAAGAAGACCCTGGTCTGTTAGGACCCTTTCTGCTT..TGAGAA-CCG-AGCA--------C------.      10249
SMM.US.SIVSMMH9     ......................................................................................       9638
STM.US.STM          A-CT-.........GGTA.CTCAA...GACCCTGGTCTGTTAGGAC-CTT.....T--GC..........................       9892
TANTALUS.UG.TAN1    .--.-CCTC-T-A-TGG.....................................................................       9752
GRIVET.ET.GRI-677   -G-A-CTCTCT-A-TG......................................................................       9418
VERVET.KE.VER-9063  .--C-GTTCCT-A-TGG.....................................................................       9779
VERVET.KE.AGMTYO    ......................................................................................       9170
VERVET.KE.AGM155    .-GAAGTTCCT-A-TGG.....................................................................       9758
VERVET.DE.AGM3      G--GACCCT---GGAAT.............................-AACC-T..GGAGAAGT.......................       9294
SABAEUS.SN.SAB1C    GGTTACTA--............................................................................      10009
SUN.GA.SIVSUN       CTCCT...TCTAAACCCTGTTCGACTGCTCTCAGCTAGGGAGCAATTA-T-G-..-GCC-AGTGATCC-G--CTG-CGG------.      10006
MANDRIL.GA.MNDGB1   ......................................................................................       9215
LHOEST.KE.hoest     GTTCAGT--TC--TCACTAGAGAGATTGGAGCCTTGTGATCCGGGA-AGG-CT..T-CAG..........................       9957
LHOEST.CD.hoest447  ......................................................................................       8588
LHOEST.CD.hoest485  ......................................................................................       8589
LHOEST.CD.hoest524  ......................................................................................       8594
RCM.NG.SIV-NGM      GGTTGGGA...................TCCCTCAGATATTTGTGGCAGA-..GTA-TGAAGGC-A--AAAAA-TCCCAA-C--....      10412
COL.___.SIV-CGU1    ......................................................................................       8728
SYKES.KE.SYK173     ......................................................................................       9597
```