III

HIV-1/HIV-2/SIV Complete Genomes

Introduction 467

Table of HIV-1/HIV-2/SIV Complete Genome Alignments 469

Nucleotide Alignments of HIV-1/HIV-2/SIV Complete Genomes 470

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 mandrill, red-capped mangabey, Syke's monkey, and sun-tailed monkey.

Table 1 lists the sequence name, accession number, first author, and publication for each of the sequences in this alignment. See page 35 of last years 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, SMM-239, and SIV from a Mandrill. 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.

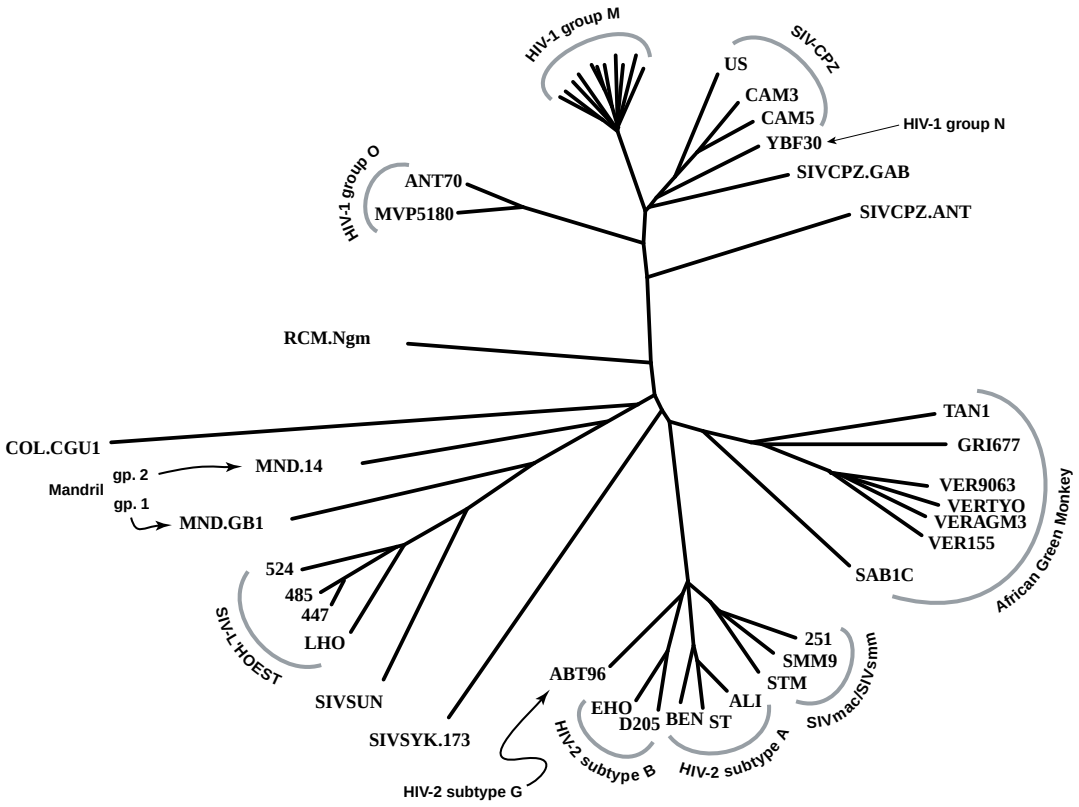

Figure 1. *This phylogenetic tree was constructed from the complete lentivirus genome alignment. Columns in the alignment for which any sequence is represented by a gap were first removed. The tree was constructed using PHYLIP DNADIST with maximum likelihood distance estimation and a transition/transversion 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
H1B.FR.HXB2	K03455	Wong-Staal, F	*Nature* **313**(6000):277–284 (1985)
CPZ.US.CPZUS	AF103818	Gao, F	*Nature* **397**(6718):436–441 (1999)
CPZ.GA.CPZGAB	X52154	Huet, T	*Nature* **345**(6273):356–359 (1990)
CPZ.CM.CAM3	AF115393	Corbet, S	*J Virol* **74**: 529–534 (2000)
CPZ.CM.CAM5	AJ271369	Souquiere, S	Unpublished
CPZ.CD.CPZANT	U42720	Vanden Haesevelde, MM	*Virology* **221**(2): 346–350 (1996)
H1A.UG.U455	M62320	Oram, JD	*ARHR* **6**(9):1073–1078 (1990)
H1C.ET.ETH2220	U46016	Salminen, MO	*ARHR* **12**(14):1329–1339 (1996)
H1D.CD.84ZR085	U88822	Gao, F	*J Virol* **72**(7):5680–98 (1998)
H1F1.BE.VI850	AF077336	Carr, JK	*Virology* **269**(1):95–104 (2000)
H1G.SE.SE6165	AF061642	Carr, JK	*Virology* **247**(1):22–31 (1998)
H1H.CF.90CF056	AF005496	Murphy, E	*ARHR* **9**(10):997–1006 (1993)
H1J.SE.SE7887	AF082394	Laukkanen, T	*ARHR* **15**(3):293–297 (1999)
H1K.CM.MP535C	AJ249239	Triques, K	*ARHR* **16**(2):139–151 (2000)
01_AE.TH.CM240	U54771	Carr, JK	*J Virol* **70**(9):5935–5943 (1996)
H1N.CM.YBF30	AJ006022	Simon, F	*Nature Med* **4**(9):1032–1037 (1998)
H1O.CM.MVP5180	L20571	Gurtler, LG	*J Virol* **68**:1581–1585 (1994)
H1O.CM.ANT70	L20587	Vanden Haesevelde, M	*J Virol* **68**(3):1586–1596 (1994)
MAC.US.SMM251	M19499	Franchini, G	*Nature* **328**(6130):539–543 (1987)
MAC.US.SMM239	M33262	Kestler, H	*Science* **248**:1109–1112 (1990)
STM.US.STM	M83293	Novembre, FJ	*Virology* **186**(2):783–787 (1992)
SMM.US.SMM9	M80194	Courgnaud, V	*J Virol* **66**:414–419 (1992)
H2A.GW.ALI	AF082339	Azevedo-Pereira, JM	Unpublished (1998)
H2A.SN.HIV2ST	M31113	Kumar, P	*J Virol* **64**(2):890–901 (1990)
H2A.DE.BEN	M30502	Kirchhoff, F	*Virology* **177**(1):305–311 (1990)
H2B.GH.D205	X61240	Kreutz, R	*ARHR* **8**(9):1619–1629 (1992)
H2B.CI.EHO	U27200	Rey-Cuille, MA	*Virology* **202**(1):471–476 (1994)
H2G.CI.ABT96	AF208027	Brennan, CA	*ARHR* **13**(5):401–404 (1997)
TANTALUS.UG.TAN1	U58991	Stivahtis, GL	*Virology* **228**:394–399 (1997)
GRIVET.ET.AGM_GRI-677	M66437	Fomsgaard, A	*Virology* **182**(1):397–402 (1991)
VERVET.KE.AGM_VER-9063	L40990	Hirsch, VM	*J Virol* **69**(2):955–967 (1995)
VERVET.KE.AGMTYO	X07805	Fukasawa, M	*Nature* **333**:457–461 (1988)
VERVET.KE.AGM155	M29975	Johnson, PR	*J Virol* **64**(3):1086–1092 (1990)
VERVET.DE.AGM3	M30931	Baier, M	*Virology* **176**(1):216–221 (1990)
SABAEUS.SN.SAB1C	U04005	Jin, MJ	*EMBO J* **13**(12):2935–2947 (1994)
SUN.GA.SIVSUN	AF131870	Beer, BE	*J Virol* **73**(9):7734–7744 (1999)
LHOEST.KE.SIVlhoest	AF075269	Hirsch, VM	*J Virol* **73**(2):1036–1045 (1999)
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)
MANDRIL.GA.MNDGB1	M27470	Tsujimoto, H	*Nature* **341**:539–541 (1989)
COL.-.SIV-COL-CGU1	AF301156	Courgnaud, V	*J Virol* **75**(2):857–866 (2001)
SYKES.KE.SYK173	L06042	Hirsch, VM	*J Virol* **67**(3):1517–1528 (1993)
RCM.NG.RCM-NGM-411	AF349680	Beer, B	*J Virol* **75**(24):12014–12027 (2001)
MANDRIL.GA.14CG	AF328295	Souquiere, S	*J Virol* **75**(15):7086–7096 (2001)

HIV-1/HIV-2/SIV
complete genomes

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

H1A.UG.U455                                                                                                                                                  0
H1C.ET.ETH2220                                                                                                                                               0
H1D.ZR.84ZR085                                                                                                                                               0
H1F.BE.VI850                                                                                                                                                 0
H1G.SE.SE6165                                                                                                                                                0
H1H.CF.90CF056                                                                                                                                               0
H1J.SE.SE92809                                                                                                                                               0
H1K.CM.MP535C                                                                                                                                                0
H101_AE.TH.CM240                                                                                                                                             0
H1N.CM.YBF30                                                                                                                                                 0
H1O.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
H1O.BE.ANT70       --------T-----T-------TAA----GC---A-----G-----T---G-G--TA----T--G--A-T-----------G------------------A-----A-C---G-TC--------A----    129

MAC.US.SMM251      --------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
MAC.US.SMM239      --------A-TTA-T--AGTGC-A-------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT----A-----A--T-----C---AA---A----    129

STM.US.STM                                                                                                                                                   0
SMM.US.SMM9                                                                                                                                                  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.ST                                                                                                                                                    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

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.GRI677   ----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.VER9063  ----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.VERTYO                                                                                                                                             0
VERVET.KE.VER155   ----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.VERAGM3                                                                                                                                            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
LHOEST.KE.LHOEST   -----------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.447                                                                                                                                                0
LHOEST.CD.485                                                                                                                                                0
LHOEST.CD.524                                                                                                                                                0
MANDRIL.GA.MNDGB1                                                                                                                                            0

COLOBUS.CM.CGU1                                                                                                                                              0
SYKES.KE.173                                                                                                  ---A-C-C-G--C--GT-A-T-----     26
RCM.NG.NGM411      --------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
MANDRIL.GA.14CG                                                                                                                                              0
```

```
H1B.FR.HXB2       GATGGTGCTACAAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAA...GCCAACAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGC......CTGCATGGAATGGATGACCCGGAGAGAGAAGT    250
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
H1A.UG.U455                                                                                                                                            0
H1C.ET.ETH2220                                                                                                                                         0
H1D.ZR.84ZR085                                                                                                                                         0
H1F.BE.VI850                                                                                                                                           0
H1G.SE.SE6165                                                                                                                                          0
H1H.CF.90CF056                                                                                                                                         0
H1J.SE.SE92809                                                                                                                                         0
H1K.CM.MP535C                                                                                                                                          0
H101_AE.TH.CM240                                                                                                                                       0
H1N.CM.YBF30                                                                                                                                           0
H1O.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
H1O.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
MAC.US.SMM251     -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
MAC.US.SMM239     -C---CTA-GG--AT----C--T--AA-T......GTA-C---T--G...--AC-GG-G-AT---G-GCATTAT--AA-G--T--A-CTCAA......ACTTCCCAGTG---------TTG-G----G--    244
STM.US.STM                                                                                                                                             0
SMM.US.SMM9                                                                                                                                            0
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.ST                                                                                                                                              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
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.GRI677  -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.VER9063 -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.VERTYO                                                                                                                                       0
VERVET.KE.VER155  -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.VERAGM3                                                                                                                                      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
LHOEST.KE.LHOEST  ------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.447                                                                                                                                          0
LHOEST.CD.485                                                                                                                                          0
LHOEST.CD.524                                                                                                                                          0
MANDRIL.GA.MNDGB1                                                                                                                                      0
COLOBUS.CM.CGU1                                                                                                                                        0
SYKES.KE.173      ------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
RCM.NG.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
MANDRIL.GA.14CG                                                                                                                                        0
```

HIV-1/HIV-2/SIV
complete genomes

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

CPZ.US.CPZUS      C---ATC---CA--------CTGAA--G--TCGGAGA----GA--TAA------------A------...---CG----...........................---.....................  329
CPZ.GA.CPZGAB     C--G-TC---C-C---------A-G--G--C--AAGA--T--T---A----A-AA------------...-A---AG--...........................----TTTAG...............  335
CPZ.CM.CAM3                                                                                                                                               0
CPZ.CM.CAM5                                                                                                                                               0
CPZ.CD.CPZANT                                                                                                                                             0

H1A.UG.U455                                                                                                                                               0
H1C.ET.ETH2220                                                                                                                                            0
H1D.ZR.84ZR085                                                                                                                                            0
H1F.BE.VI850                                                                                                                                              0
H1G.SE.SE6165                                                                                                                                             0
H1H.CF.90CF056                                                                                                                                            0
H1J.SE.SE92809                                                                                                                                            0
H1K.CM.MP535C                                                                                                                                             0
H101_AE.TH.CM240                                                                                                                                          0
H1N.CM.YBF30                                                                                                                                              0
H1O.CM.MVP5180    AC--A-A---CA------T--ATCAT---GC--AACA--T--A----TGC-AAA---C--A---CT-...---CCC--G.............................T-ACTGACACTGC........  343
H1O.BE.ANT70      AC-GAT-----A------T--ATCT----GCAACACC--TG-T--TATGATAACT--C--A---CT-...---C----GGA-..........................T-AAAACTGCTG.........  344

MAC.US.SMM251     TC---C-----A------TCCAACT-----C-ACAC-T-TGA---ATAT-CTAGAT-C--A--AG-G...--GG-AGCAA-T......CAGGCCTGTCAGAGGAAGA-GTTAGAAGAAGGCTAACCGC..  363
MAC.US.SMM239     TC---CA----A------TCCAACT--G--C-ACAC-T-TGA---ATAT-TTAGAT-C--A--AG-GT..--GG-AGCAA-T......CAGGCCTGTCAGAGGAAGA-GTTAGAAGAAGGCTAACCGC..  364

STM.US.STM                                                                                      ..TGGAAGGGATTTATTACAGTGAGA.GAAGGCTAACCGC..   37
SMM.US.SMM9                                                                                                                                               0

H2A.GW.ALI        A----TT-----A------CCTAAG--G--TCA-G--T---AA---TTTATT--A--C--A---G-A...--TGG-T--AAG...TCAGGCCTGCCAGAAGATGAG--GA-GGCAAGACTGAAAGCAAGA  368
H2A.SN.ST                                                                                                                                                 0
H2A.DE.BEN        A----T----CA-------TC-ATG--G--C-A-A-CT---A----TTCACT---T-C--A--AG-G...--TGG-C--AAG...TCAGGATTGCCAGAGAAAGAA--GA-GGCAAAACTGAAAGCAAGA  368
H2B.GH.D205       TC-TATC---CA------TTC--T---G----A-G--T-TG----TTTCA-TAG-TT---A--AG-G...---GG-T--CAG...TCAGGGCTACCAGAGGAGGAG--GA-GGCTAGACTAAAAGCAAGA  371
H2B.CI.EHO        CC-T-TC---CA-------TC--T--------A-G-CT-TG-----TTCAGCAG-TTC--A---G-G...--TGG-T-TCAG...TCAGGAATGCCAGAGAAAGAG--GA-GGCTAAACTGAGAGCAAGA  368
H2G.CI.ABT96                                                                                                                                              0

TANTALUS.UG.TAN1  A--G-T-----AA---...............AA--CA-T-C-A--AGTTC--TACA-C-----CAG-...--T...--GGA-...ATGCATGGTCTTGTTAAGAGGAAGT-GCCTAACCGCAGGCTTGCG  347
GRIVET.ET.GRI677  C--G-TC----A-------...............-CGA-GT----GGTG---TAC-GC-----CATG...--T...--GGA-...ATGCACGAACATGCAAAGCGC...T-GTTTCGC...TGGTTAGCC  350
VERVET.KE.VER9063 C--G-TC----A-------...............-C-ATGT-A--AGTGC--TATG-C--AA-CAGA...CAGT-TTTTGAAGACATGCATGCAATAGTGAAGAGGAAGT-GCTAACCGCAGGCTTGTG.  361
VERVET.KE.VERTYO                                                                                                                                          0
VERVET.KE.VER155  A--G-CA----A------T...............-CAATGT----TGTTC--TACG-C--CTCAAGG...GAGT-CTTTACAGACTTATATTCAACAGTTGGTACAG--A-CTAGCCGACCACAGGCTTG  362
VERVET.DE.VERAGM3                                                                                                                                         0
SABAEUS.SN.SAB1C  CC-G-T-----A-------CCA-TA------G-GG-CT-TG-T--TT-GAG------C--A---C-G...G-GCC-TCAGCA...CAAGGCTAACCACAAAACCACATC-TATG...............  352

SUN.GA.SIVSUN     AC-TATC----CA-----TCC-ACT--G--C-A-G--TTT-G----AT-C--AA---C--T--AG-G...-------C-TGTT...ACTAGTCTGCAGTGGGAGGTG-AG-TTAACCGCAAACCG......  365
LHOEST.KE.LHOEST  -C-TATC---CAC--CA-TCCAATGT-G---G-GG-CTTT--A----TCA-AAA---G--T-C-A-G...A-AC-A--TACT...GCTTTCGCTTTTGACTGCAAGA-GA-GTAGCCAAACCGC......  365
LHOEST.CD.447                                                                                                                                             0
LHOEST.CD.485                                                                                                                                             0
LHOEST.CD.524                                                                                                                                             0
MANDRIL.GA.MNDGB1                                                                                                                                         0

COLOBUS.CM.CGU1                                                                                                                                           0
SYKES.KE.173      CC-CAT-----AC--CA--CC--A---G--C-A-ACA-CAGGCTGGGAGAT-GCC-GCTT-C---TA...GAG-GAC-AACA...GGAAAACCACAAGAGCTGCAG-C-GCACTTAGTAAGAACATCTCC  262
RCM.NG.NGM        A----CA----A---CA-TCC--T-T-----G-AG--T--GCA-GGTATAGA--------T----T-...--TGGAG-GA-A...AAGAACAAGACCAAGTAACCACTT-CTCTGGAGTTGCCTTG....  364
MANDRIL.GA.14CG                                                                                                                                           0
```

```
  See Pohlmann, J Virol 72(7):5589-5598(1998) and similar publications for information on this enhancer region
                                                                                                   NF-kappa-B-II                   NF-kappa-B-I
H1B.FR.HXB2        ..................................................................TCGAGCTTGCT...ACA...AGGGACTTTCCG....................CTGGGGAC    368

CPZ.US.CPZUS       ...............................................TGGCTGTAACCGCGCAGGCGC.AATA-AAC----G..--TG...A--------TAA...GGGACTTTCCAA.......-----    396
CPZ.GA.CPZGAB      ....................................................ACTGGCGCATGCGC.ACAAGAAC----G..--TCTGC----------AA.......................-----    386
CPZ.CM.CAM3                                                                                                                                                  0
CPZ.CM.CAM5        ..........................................................GCGCCTGCGCAG-A--AAC----G..---A...----------TAA.......................-----     45
CPZ.CD.CPZANT                                                                                                                                                0

H1A.UG.U455                                                                                                                                                  0
H1C.ET.ETH2220                                                                                                                                               0
H1D.ZR.84ZR085                                                                                                                                               0
H1F.BE.VI850                                                                                                                                                 0
H1G.SE.SE6165                                                                                                                                                0
H1H.CF.90CF056                                                                                                                                               0
H1J.SE.SE92809                                                                                                                                               0
H1K.CM.MP535C                                                                                                                                                0
H101_AE.TH.CM240                                                                                                                                             0
H1N.CM.YBF30                                                                                                                                                 0
H1O.CM.MVP5180                                                                                                  ...----------A    ...GACTGCTGACACTGC------    375
H1O.BE.ANT70                                                                                   ...ACC-GA--A-----G..---CTGT--A-------A..GCAAAGACTGCTGACACTGC------    403

MAC.US.SMM251                                      AAGA...GGCCTTCTTAACATGGCTGACAAGAGGGAA-CTCGCTGA..GATAGC----------A.....................CAA--GG    431
MAC.US.SMM239                                      AAGA...GGCCTTCTTAACATGGCTGACAAGAAGGAA-CTCGCTGA..-ACAGC----------A.....................CAA--GG    432

STM.US.STM                                         AAGA...GGCCTTCTAAAGATGGCTGACAAGAAGGAA-CAAGCTGA..GACAGC----------A.....................CAA--GG    105
SMM.US.SMM9                                                                                                                                                  0

H2A.GW.ALI         GGGATACCATTTAGT.AAGAACAGGAACAGCT.GATTTGG...TCAGGGCAGGAAGTAACTACTG.....AAA-C.AGCTGA..GACTGC----------A.....................GAA--GG    464
H2A.SN.ST                                                                                                                                                    0
H2A.DE.BEN         GGGATACCATATAGTGA...ATAACAGGAACAACCATACT...TGGTCAAGGCAGGAAGTAGCTA.CTAAGAA-C.AGCTGA..GGCTGC----------A.....................GAA--GG    467
H2B.GH.D205        GGGATACCTACAGATTAGGCAAGAGACAGCAGCATAAACA...GGAACTAGCTGACACTGCACAA.GAAGGAA-CTAGCAGA..CACTGC----------A.....................AAA-AGG    474
H2B.CI.EHO         GGAATACCTACAGAGTAGACAGGAAATAACAGCACAAAGA...GGAACTAGCTAACACTGCATAGAGAAGGAA-CTAGCTGA..TACTGC---.-------A.....................GAA--GG    471
H2G.CI.ABT96                                                                                                                                                 0

TANTALUS.UG.TAN1   GT.............................TAAGCTAA...CTAGGCCGTTGCCTAGGAGATG.ACATAAAGAAG----G..-TCAGC----------A...........................    413
GRIVET.ET.GRI677   TAACCCGGTTG.......................................................................................................................    361
VERVET.KE.VER9063  ........................................GTTAAGCCGTTGCCAGGGAGATGACAT-T--AAC----G..---G...----------A.....................A-----    421
VERVET.KE.VERTYO   ...................................................................................................................................    0
VERVET.KE.VER155   ....................................CG...GTTTCCTGGTTGCCTAGGAGATGACAT-AAGAAC----G..-...----------A...................G-AC-----    425
VERVET.DE.VERAGM3  ...................................................................................................................................    0
SABAEUS.SN.SAB1C   ..............................................GAGTTGTCATGGTGATGACAT-AAGAAC----G..--TGAG----------A.........................GG    405

SUN.GA.SIVSUN      ............................CATCCTCTTGCA...TCGCCTAGGCAACGGGGCTAGC.GCAT......GC--GC..TAGAGT-----------...........................    427
LHOEST.KE.LHOEST   ........AAACCGCAGCACATCCTCTTGCAGCCCGGTTG...CTAAGGCAACCGGGCTAGCGCATGCGCAATG-------G..--GAGC----------A............GGACGGG-G-----G    469
LHOEST.CD.447                                                                                                                                                0
LHOEST.CD.485                                                                                                                                                0
LHOEST.CD.524                                                                                                                                                0
MANDRIL.GA.MNDGB1                                                                                                                                            0

COLOBUS.CM.CGU1                                                                                                                                              0
SYKES.KE.173       TAGGAGACTCCATGGTGACAAGCTCGGCCCAGGGGA.....................................................------C-GGGC..................GGGG-AA-G-    319
RCM.NG.NGM         ...............................................GAAACAGCGCATGCCTACTAAGA--A-GCTGA..CACAGA----------ACTGGCGCCTGCGCGCTGGTGTAA-----    442
MANDRIL.GA.14CG                                                                                                                                              0
```

HIV-1/HIV-2/SIV
complete genomes

```
HIV-2 and SIVs have extra TAR stem-loop(s) see Berkhout, Nucleic Acids REs 20(1):27-31(1992)
                                                                                                                                                                 /+1 mRNA start site
                                                                                                                                                   5' LTR U3 end \/ 3' LTR start
                                                                                                                                                            TAR element stem
                  NF-kappa-B-I              Sp1-III       Sp1-II      Sp1-I                              TATA Box                                  |   ||||^|||||||
H1B.FR.HXB2       T...TTCCAG............GG.AGGCGTGG.CCTGGGCGGGACTGGGGAGTGGCGA........GCCCTCAGATCCTGCATATAAGCAGCTGCT...TTTTGCCTGTACT...GGGTCTCTCTGG    465

CPZ.US.CPZUS      G...----..............AAGG--GTG--T-A.------A........ACA-G-CGTGGTTT.TA------A-GA-------A-------C---...C-C---T------..--------T.C-    491
CPZ.GA.CPZGAB     -...----G-............--AGA-GTG-TCGGGA------G.......T---G-TGTGGCTT.T--------.AG-------------------...--C----------..--------T-CA    485
CPZ.CM.CAM3                                                                                                                                              0
CPZ.CM.CAM5       -...-----A.............-GG--A-GCTCA..------........ACA-G-CGTGGTTT.CA--------AG--------------C---...--C---T-----G...T---------T-    141
CPZ.CD.CPZANT     ..............................................................................................................................      0

H1A.UG.U455                                                                                                                                              0
H1C.ET.ETH2220                                                                                                                                           0
H1D.ZR.84ZR085                                                                                                                                           0
H1F.BE.VI850                                                                                                                                             0
H1G.SE.SE6165                                                                                                                                            0
H1H.CF.90CF056                                                                                                                                           0
H1J.SE.SE92809                                                                                                                                           0
H1K.CM.MP535C                                                                                                                                            0
H101_AE.TH.CM240                                                                                                                          ----------T-    12
H1N.CM.YBF30      ................................................................................................T-C-C--T------..----------T-     26
H1O.CM.MVP5180    -...------............C-TG--A-G-ATAAG------TT-G---AGTG-C..........TAA---------G-------------------...--CC--T-----C...------TAG-TA    473
H1O.BE.ANT70      -...------............T-GGA-G-AC..AGG------TT-G---AGTG-C..........TAA--------AG--------------C---...--C---T-----C...-------GG-TA    499

MAC.US.SMM251     A...-GTT-T............--GGA-GAGCCG.G-C--GAAC-......................CC-A--TTCT-GA--T---A-TATCA----A.TT-CGCT-TGTATTC...A-TCGCTCTGC-    518
MAC.US.SMM239     A...-GTT-CGGG..GAGGTACT--GGA-GAGCCG.G-C--GAACG.....................CC-A--TTCT-GA--T---A-TATCA----A.TT-CGCT-TGTATTC...A-TCGCTCTGC-    530

STM.US.STM        C...-GTA-.............CAGG--A-GTAC.TG--AG-A-CTG-T--G-.AA-.........GCC-TT-T-CT-T---T---A-TA--A-C---.TA-AGCT-TGTATTC...A-TCACTCTGC-    203
SMM.US.SMM9                                                                                                          C...A-TCACTCTGC-     13

H2A.GW.ALI        C...-GTA-C............CAGG--AAG-AC.A----A--AG----T-G-GAA-.........GC--TCATAC-----T---A-T-T-C-C---.GC--GCATTGTATTC...A-TCGCTCTGC-    563
H2A.SN.ST         ...........................................................................................................A-TCGCTCTGC-     12
H2A.DE.BEN        C...-GTA-C............CAAG--A-G-AC.A----A--AG----T-G-GAA-.........GCC-T-AT-CT-A---T---A-T-T-C-C---.TC--GCATTGTATTC...A-TCGCTCTGC-    567
H2B.GH.D205       C...-GTA-C............A.TG--A-GA-C.AA---A---GT---TTG-GAA-.........GCC-AT-A-ATCT---T---A-T-T-C-C---.TC-CGCATTGTATTC...A-TCGCTCTGC-    573
H2B.CI.EHO        C...-GTA-C............A-TG--A-G-AC.A----A--AG-A---AG-.AA-.........GCC-ATCA---CT---T---A-T-T-C-C---.TTG-GCATTGTATTC...A-TCGCTCTGC-    570
H2G.CI.ABT96                                                                                                                                             0

TANTALUS.UG.TAN1  .........-............--TG-TGCGA-.GGC--TACTTGGGA-T-GC-T.............---CTCAGAG-------A------A----CGC-GGCTTG-AACTC...A-TCTCT-ACTA    503
GRIVET.ET.GRI677  ....................................-CCACC.GGG--TA--GAC.............T---TG-C-T...------GCTCAA-AAA...............................    405
VERVET.KE.VER9063 -...------............--CG---CAT-.GGC--T.AC-GGGA-T-GT-T.............TA--CTCAGAG-------A------A----CGC-GGCTTG-AACTC...A-TCTCT-ACTA    517
VERVET.KE.VERTYO  .....................................................................................................................C...A-TCTCT-ACTA     13
VERVET.KE.VER155  -...-----A............--CG--ACAT-.GGC--T.AC-GGGA-T-GC-T.............TA--CTCAGAG-------A------A----CGC-GGCTTG-AACTC...A-TCTCT-ACTA    521
VERVET.DE.VERAGM3 .....................................................................................................................C...A-TCTCT-ACTA     13
SABAEUS.SN.SAB1C  G...-GGAGA............CTGG-CG--ACTGGGA-TG-CTT.....................GCC--G-GCTGCT-AAT---AGCA..------.TTGCGCCTG-AA-AC...A-TCTCTG--T-    492

SUN.GA.SIVSUN     ....................................GACAG--AGGG--A-..............GCTTA-C--CT-A------------------.CT-CGCCTG-T.CGG...-A-.........    491
LHOEST.KE.LHOEST  G...--T...........................................................G-C--CT-G------------------..C-CCGCTTGT-CGG...-A-.........    518
LHOEST.CD.447                                                                                                                                            0
LHOEST.CD.485                                                                                                                                            0
LHOEST.CD.524                                                                                                                                            0
MANDRIL.GA.MNDGB1                                                                                      -GAGTC-C-ACTACAG-GG...CTAAGGG-TGTA     31

COLOBUS.CM.CGU1                                                                                                                                          0
SYKES.KE.173      -...CA--CT............-CTGC........................................................---A----C-A----......................CTGTAGCTA    355
RCM.NG.NGM        -...------............ACTGACGTG--AGGG---T-T-G--T-CCCTCTC-.................TGCTGCAT---AGCAGCTGCAT-TCGC-C--TAACC-G-...CTC--C--GG-A    535
MANDRIL.GA.14CG                                                                            GAGAGGCT-C-A-A-TGAGCC--GG-TGTTC...TCTGG-GAG-CC     43
```

```
                TAR element stem                                TAR element bulge  loop     TAR element stem
                ||||^                                           |||||^^^||||      |||||||||||| |||||||||||||
H1B.FR.HXB2     TTAGA.........................................CCAGATCTGAGCCTGGGAGCTCTCTGGCTA.ACTAGGGAACCCACTGCTTA...........AGCC..T.............   525

CPZ.US.CPZUS    C-----------------------------------------------AA-----C----------------GTG---..--------------.................--.-............   550
CPZ.GA.CPZGAB   C-G--.........................................------T-----------------A.-C-GTGA--.--------------.................--.-............   544
CPZ.CM.CAM3     ..............................................................................................................................   0
CPZ.CM.CAM5     --G--.........................................------TA------------------..GCC-GC-----------------.................--.-............   200
CPZ.CD.CPZANT                                                                                                                                   0

H1A.UG.U455                                                                                                                                     0
H1C.ET.ETH2220                                                                                                                                  0
H1D.ZR.84ZR085                                                                            ------------CG.G-------------------.................--.-............    39
H1F.BE.VI850                                                                                                                                    0
H1G.SE.SE6165                                                                                                                                   0
H1H.CF.90CF056                                                                                                                                  0
H1J.SE.SE92809                                                                                                                                  0
H1K.CM.MP535C                                                                                                                                   0
H101_AE.TH.CM240 -----.........................................----G--.-----C----------------.G-A-----------------A................--.-............    72
H1N.CM.YBF30    C-G--.........................................------TA----------------------.G-AG--A-C--ACTG.----................--.-............    85
H1O.CM.MVP5180  GAG--.........................................----G--------C--------C-----CT.CTAGCT------G------............CG-..-.............   533
H1O.BE.ANT70    GAG--.........................................----G--------C--------C-----CT.CTAGCT------G------............CG-..-.............   559

MAC.US.SMM251   GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAG----TGGCCAGTGCTGGGCAGAGT..GGCTCCACGCTTG..   639
MAC.US.SMM239   GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAG----TGGCCGGTGCTGGGCAGAGT..GACTCCACGCTTG..   651

STM.US.STM      GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A-------A-T-T--C---CTAG.---CTCACCAG----TGGCCAGTGCTGGGTAGAGT..GGCTCCACGCTTG..   324
SMM.US.SMM9     GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTGG.---CTCACCAG----TGGCCGGTGCTGGGCAGAGT..RGCTCCACGCTTG..   134

H2A.GW.ALI      GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAGTG--TGGCCGGCACTGGGCAGA-...GGCTCCACGCTTG..   683
H2A.SN.ST       GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAGTG--TGGCCGGCACTGGGCAGA-...GGCTCCACGCTTG..   132
H2A.DE.BEN      GAGAGGCTGGCAGATCGAGCCCTGAGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTGG.---CTCACCAGT---TGGCCGGTACTGGGCAGA-...GGCTCCACGCTTG..   687
H2B.GH.D205     GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAG----AGGCCAGTGTTGGGCAGA-...GGCTCCACGCTTG..   693
H2B.CI.EHO      GAGAGGCTGGCAGATCGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCACCAG----TGGCCGGTGCTGGGCAGA-...GGCTCCACGCTTG..   690
H2G.CI.ABT96                                                                                 ---CTAG.---CTCACCAG-G--TGGCCGGTGCTGGGCAGAGY..GGCTCCACGCTTG..    55

TANTALUS.UG.TAN1 GG---........................................----.CT------------T--G----TG-.G-CTAACCTGGTCTG--CATCCAGGGGTAAG-----..-C............   574
GRIVET.ET.GRI677 ..............................................................................................................................   405
VERVET.KE.VER9063 GG---.......................................----C-T.---------T-T--G----T--.G-CTAACCTGGT.TG--CA.CCAGGGGTAAGGA-T..C.............   585
VERVET.KE.VERTYO GG---........................................----C-T.---------T-T--G----T--.G-CTAACCTGGT.TG--CA.CCAGGGGTAAGGA-T..C.............    81
VERVET.KE.VER155 GG---........................................----C-T.---------T-T--G----T--.G-CTAACCTGGT.TG--CA.CCAGGGGTAAGGA-T..C.............   589
VERVET.DE.VERAGM3 GG---.......................................----C-T.---------T-T--G----T--.G-CTAACCTGGT.TG--CA.CCAGGGGTAAGGA-T..C.............    81
SABAEUS.SN.SAB1C GAG..GCTGCCAGATTGAGCC.TGGG.TGTTCTCTGGTAAGTCTGAA----C-T.---------T-T--G----GG-AGGC-A-CCTAAG-GCA--CTGGTGAGGTCTTG-TA..GACCCTCGCCTGGCG   615

SUN.GA.SIVSUN    ..............................TCTCT.GCT.CAGAGGCTAG-CGAC--C----A---G----ACCTAG.-GCT--CTTGT-G--C--......GAGTGAGCA-T..-GCTGAGCACCTCAC   580
LHOEST.KE.LHOEST ..............................TCCTT.GCTCTGGGGAC-AGC-GTC-CT------T.GAGCACACTGG.T-AGCTA-G-AGT-A--CCTGGGAGGCTACACTTA..GCACCCACCTGGGCC   613
LHOEST.CD.447                                                                                                                                   0
LHOEST.CD.485                                                                                                                                   0
LHOEST.CD.524                                                                                                                                   0
MANDRIL.GA.MNDGB1 -CTCTGAG.CAGATCCCCTTAGAGCAAGGACCAGAGTCCT.............GA-T-A-----TCTGAG-ACCTC-.CTCG---CTGATCACCTCG-GGTAGTGGAACTC-T..-GCTTGCTTGC...T   141

COLOBUS.CM.CGU1  ......................TGGAGTTCTGACT.CTTGCCTGGCTG-GCC-GA-CC-TGA-TG---TG-CA----.T-AGCCTTG--TGAGT--G-GTAGCTTG.........................   82
SYKES.KE.173     -CC-CGGAGTTGCAACTAGCTCTCTCCAGCGG.....................................C--GA-AGAC...-...-GTAGAG-G--CCGGGAGCCCTTGCCAGT..AAAGGCTAGTGAG.G   450
RCM.NG.NGM       GGCT-CCAGATTGAGCCTGGGTTGTTCTCTGGTAA.GTCTCTAGGAGA-TAGCT----------T-T--G--A-T-C.T--...---GGGCT----GGGAGCGCCTTGGCTGG..GTCTTGGTAGACCTT   658
MANDRIL.GA.14CG  --GA-GGAGTGCCTGCTTGTAGCCCTGGGCGGTTCGCAGGCCCCTGG-TT-TAGCTCTGGGTA-CT-GTCAG-TG-TCTGG-AA-GT-T-G--AAG-GAGCACTCTTGCTTGG..-TTTGGTAGACCTCT   171
```

HIV-1/HIV-2/SIV
complete genomes

```
                                                                                                        secondary structure in this region, see Rizvi, J. Virol 67:2681
                                                                                                               Poly_A signal  5'LTR R Repeat end \/ 5' LTR U5 start
H1B.FR.HXB2        .......................................................................................C.AATAAA.GCTTGCCTTG..AGT....GCTTCA.AGTAGTGT    559

CPZ.US.CPZUS       .......................................................................................-.------.----------.---....----T-.--------    584
CPZ.GA.CPZGAB      .......................................................................................-.------.----------.---...GTA--TG.--C-----    579
CPZ.CM.CAM3                                                                                                                                             0
CPZ.CM.CAM5        .......................................................................................-.------.----------.---....-AC-TG.--C-----    234
CPZ.CD.CPZANT                                                                                                                                           0

H1A.UG.U455        .............................................................................................--.----------.---....----G-.--------     29
H1C.ET.ETH2220                                                                                                                                          0
H1D.ZR.84ZR085     .......................................................................................-.------.----------.---....----T-.--------     73
H1F.BE.VI850                                                                                                                                            0
H1G.SE.SE6165                                                                                                                                           0
H1H.CF.90CF056                                                                                                                                          0
H1J.SE.SE92809                                                                                                                                          0
H1K.CM.MP535C                                                                                                                                           0
H101_AE.TH.CM240   .......................................................................................-.------.----------.---....----A-.---G----    106
H1N.CM.YBF30       .......................................................................................-.------.----------.---....---AA-..--G----    118
H1O.CM.MVP5180     .......................................................................................-.------.----------.---GAGAAGCA.......----    565
H1O.BE.ANT70       .......................................................................................-.------.----------.---GAGAAGCA.......----    591

MAC.US.SMM251      .................................................CTTG.CT.T.......AA...AGACCTCTT........-.------G.--.---A-T..TTAGAAGTAAG-C.-.GT----    693
MAC.US.SMM239      .................................................CTTG.CT.T.......AA...AGCCCTCTT........-.------G.--.---A-T..TTAGAAGTAAG-T.-.GT----    705

STM.US.STM         .................................................CTTG.CT.T.......AT...ATATCTCTT........-.------G.--.---AAT..TTAGAAGTAAG.C.-AGT----    378
SMM.US.SMM9        .................................................CTTG.CT.T.AAAAG........ACCTCTT........-.------G.--.---AGT..T.AGAAGTAAG-C.-.GT----    188

H2A.GW.ALI         .................................................CTTG.CT.T..AAAG........ACCTCTT.........------G.--.---AAT...TAGAAGCAGGTT.-AAG----    736
H2A.SN.ST          .................................................CTTG.CT.T.AAAAG........ACCTCTT.........------G.--.---AGT..TAGAAGCAAG-T-.-.GT----    186
H2A.DE.BEN         .................................................CTTG.CT.T..AAAG........ACCTCTT........-.------G.--.---A.-..TTAGAAGCAAGTT.-AGT----    741
H2B.GH.D205        .................................................CTTG.CT.T.AAAA.........CCCTCTT.........------..--.---A.A..TTAGAAGCAAGT..-AGT----    744
H2B.CI.EHO         .................................................CTTG.CT.T.AAAA.........CCCTCTT.........------A.--.----.A..TTAGAAGCAAGAC.-AGT----    743
H2G.CI.ABT96       .................................................CTTG.CT.TACAAAG........ACCTCTT.........------G.--GC-AAAT..TAGAAGTAAGCA-.GTGT----    112

TANTALUS.UG.TAN1   .................................................CTTG.CT.TAGAAGGCTA...TTG..............-T------.-.------GC..TTAAGAGT--ATC.T-AG.CAA    630
GRIVET.ET.GRI677   ..................................................................................................CC-GCTC----AG.TCGCTATA    426
VERVET.KE.VER9063  .................................................CTTGGCT.TGGAAAG.......................-T------ACT------GC..-T-AGAGCT-ATC.T-AGTCAA    639
VERVET.KE.VERTYO   .................................................CTTGGCT.TAGAAAG.......................-T------.-------GC..-T-AGAGCT-ATC.T-AGTCAA    133
VERVET.KE.VER155   .................................................CTTGGCT.TAGAAAG.......................-T------.-------GC..-T-AGAGCT-ATC.T-AGTCAA    641
VERVET.DE.VERAGM3  .................................................CTTGGCT.TAGAAAG.......................-T------.T---CG--GC..-T-AGAGCT-CTC.T-AGTCAA    134
SABAEUS.SN.SAB1C   AC.TGGCCATTGCCAGTAGCAGAGACTCCG...................CTTG.CT.T.GCTTG.......ATTGCCT.........-.------GT--------A..GT-TACAAG-AA-.GCA-----    702

SUN.GA.SIVSUN      CCGAGCCTGATC...............AGCTTGGGGAGCTAGAGGCTC.CTTG.CT.TTGCAGTACA...AG...CCTT........-.------G.-C------A..GT-TTACTGCAAG.CAA...--    670
LHOEST.KE.LHOEST   TGATCA..............GCCTAGGGAGC.TAAGGGCTCCTAG....CTTG.CT...ACTTTAAA...G....CCTT........-.------AAG----TCAT..T--TAGAAAG-AC.-AGCTA--    701
LHOEST.CD.447                                                                                                                                           0
LHOEST.CD.485                                                                                                                                           0
LHOEST.CD.524                                                                                                                                           0
MANDRIL.GA.MNDGB1  ATTGTCTTC..............................................................................------GTAACTTAGAA..TTAGAGCAAG-G-.GTA-----    188

COLOBUS.CM.CGU1    ..............CTTGGTTGAAGGGCTTGAGGGACTCCAGTCCCCTCGCGCTAAGGGTAGATAGG...TGCTTTGCATTATAAGAGC------GTGA-TTTGAA..GC-TCAA--AG-G.TT-GAGTA    192
SYKES.KE.173       TCCGGGTTTG.....GCCGCCCGGTCTTCGGCAGCTCTCCGTTG.....CTTGC.TTTGCTTACGCT...GATCAGCCCA........------GTG-G-TAAAC..GCAAGTAA-G--T.GC-CTCA-    554
RCM.NG.NGM         CAGCTGGGTA...CTTGGCCAGTCCCAGTGGGAGTAGGCTCCG......CTTGC....TTTCTGGTT...A..ATTGCT........-.------CTGCATG--CT..TAGTTGAA-CAG-.-TC-A.--    757
MANDRIL.GA.14CG    AGCAATCTC....TAGGCCAGGAGATTGTGGGTTAGACTACCG......CTTGC.....TTGCT.........ATTGATGCT.....-.------G----A---GA..TTAGAA.TAGCA-.GTA-----    267
```

```
                                                                                                                  see Zhang, J. Virol 72(7):5664-5471(1998)
                                                                                                                      5’ LTR U5 end \/Lys tRNA binding site
H1B.FR.HXB2          GTGCCCGT.CTGTT.GT........GTGACTCT.GGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGG...............................AAAATCTCTAGCAG..TGGCGCCCG    644

CPZ.US.CPZUS         -----TA-C-.T-CA.........--------.--------------------A....C-CT-A-AA--A..........................GGTG.-------------.---------    667
CPZ.GA.CPZGAB        -----TA-CTAT-CA..........-AC-----.-------------------T....-A-A-T-TA--CA.........................AGG..-------------.---------    663
CPZ.CM.CAM3          ..........................................................................................................----.---------     13
CPZ.CM.CAM5          -----TA-C-AT-CA.........--------.--------------------A....CA-T-ATAA--A-..........................GTG.-------------.---------    318
CPZ.CD.CPZANT        ...........................................................................................................................      0

H1A.UG.U455          --------.-----AT-........-------.----------------------AC-A---A-T----A..........................-------------.---------    113
H1C.ET.ETH2220       ..............................................................................................-------------.---------     22
H1D.ZR.84ZR085       --------.-----TT.........--------.--------------------AGC-------C-G----..........................-------------.---------    158
H1F.BE.VI850         ...........................................................................................................................      0
H1G.SE.SE6165        ..........................................................................................GGCTT.-------------.---------     27
H1H.CF.90CF056       ...........................................................................................................................      0
H1J.SE.SE92809       ...........................................................................................................................      0
H1K.CM.MP535C        ...........................................................................................................................      0
H101_AE.TH.CM240     --------.----G.T-........AG------.---------------------AC-C---A-T-A--A..........................---------C---.---------    190
H1N.CM.YBF30         ------A-.-CA--C-..........--A-----.----C----------------A-C-AGACTGAGTGA.........................-------------.---------    202
H1O.CM.MVP5180       ----T-A-CTG.--CAACC...........--.---GT---------------T-AC--AGAC.T-AAGCAG.........................-------------.---------    647
H1O.BE.ANT70         ----T-A-CTG.--CAACC...........--.---GT---------------T-AC--AGAC.T-AAGCAG.........................-------------.---------    673

MAC.US.SMM251        --T---A-CT.C-CCTAGTCGCCGCC--GTCAA.CTCGGTACTC-G-AATAAG...AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGAGAAACCGAAGCAGG------C-------AT---------    818
MAC.US.SMM239        --T---A-CT.C-CCTAGCCGCCGCC--GTCAA.CTCGGTACTCA--AATAAG...AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGGAAACCGAAGCAGG------C-------AT-------T-    830

STM.US.STM           --T---A-CT.C-CCTAGTCGCCGCC--GTCAA.CTCGGTACTC...........AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGGCTACTGAGGCAGA------C-------AT---------    494
SMM.US.SMM9          --T---A-CT.C-CCTAGMCGCCGCC--GTCA-.CTCGGTACTC--A-ATAAG...AAGSCCCTGGTC---TAGGACYCTTTCTGCTTTGGGAAACCGAGGCAGG------C-------GT---------    313

H2A.GW.ALI           --T---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGTGTTCACCTG.AGTA--AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGGAAACGGAGGCAGG------C-------GT---------    862
H2A.SN.ST            -CT---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGTGTTCATCTAAAGTA--AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGGAAACCAAGGCAGG------C-------GT---------    313
H2A.DE.BEN           --T---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGTGTTCATCTGAGT-AC.AAGACCCTGGTC---TAGGACCCTTCTCGCTTTGGGAATCCAAGGCAGG------C-------GT---------    867
H2B.GH.D205          --T---A-CT.C-CCTAGTCGC.GCC--GTCA..CTCGGTGCTCC-CTGA......TAGACCCTGGTC---TAGGACCCTTCTTGCTTTGGGAAACCGAAGCGGG------C-------AT---------    864
H2B.CI.EHO           --T---A-CT.C-CCTAGTCGCCGCC--GTCA-..TCGGTGCTCC-CTAG.......AAACCCTGGTC---TAGGACCCCTTCTGCTTTGGGAAACCAAGGCAGG------C-------AT---------    863
H2G.CI.ABT96         --T---A-CT.C-CCTAGTCGCCGCC--GTCAC.CTCGGTGTTC-CGTAGC.....AAGACCCTGGTC---TAGGACCCTTTCTGCTTTGGRGAACCAGAGCGGG------C-------AT---------    235

TANTALUS.UG.TAN1     -----.TC.A-TGC.............-CGC-C.TTCCT-.--G--A-----GTTA-TGG.....................GGTTTCTTAT.....CCAGGCTGAG-------C-----.---------    715
GRIVET.ET.GRI677     T--GAGTCAA.--G..CTCATTGCT-C-C-GAG.CC-CTAG--GTGAA----TC.TTAC-G....................GGTT..CTCCTGTACCCAGGTGGG-G--A---C-----.---------    527
VERVET.KE.VER9063    ------TCAT.TGAC-CCTCACTC..AAG-AGG.--A-C-GTTCCT-A-TG.............................GGTTCTCTCTCTGACCCAGGCGGG-G--A---C-----.---------    733
VERVET.KE.VERTYO     ------TCAT.TGAC-CCTCACTCTC-TGAA-G.--A-T-.TTCCT-A-TG.............................GGTTCTCTCTCTGACCCAGGCGAG-G--A---C-----.---------    228
VERVET.KE.VER155     ------TCAT.TAAT-CCTCACTC..-TGAA-G.--AG-AGTTCCT-A-TG.............................GGTTCTCTCTCAAACCCAGGCGAG-G--A---C----..---------    734
VERVET.DE.VERAGM3    ------TCAT.TGAC-CCTCACTCTTGA--GGG.TAA---.TTCCT-A-TG.............................GGTTCTCTCTCA.ACCCAGGCGAG-G--A---C-----.---------    228
SABAEUS.SN.SAB1C     ------A-TT.A--CC-CAGGACAACCCT...................................................GGTTACTAAGGATCCCTGATAGAA------CT------.---------    777

SUN.GA.SIVSUN        --T-----GT-C-C.TCTTTCTCCTTCT....A.AACCCTGTTC-............AC-GCTC-T-ACTA--GAGCAATTACTGGCCGCCTAGTGATCCGGGGCTG-CGGC.------.---------    779
LHOEST.KE.LHOEST     --CT-ATAGT.AG-CTCTTTCTCCTTC-C-CTG.-T-C-G-GATCTCT-ACT---GAGA--GGAG-CT---..........................GATCCGGG-C-GGG--TC---.---------    800
LHOEST.CD.447        ...........................................................................................................................      0
LHOEST.CD.485        ...........................................................................................................................      0
LHOEST.CD.524        ...........................................................................................................................      0
MANDRIL.GA.MNDGB1    TAT--AT-GTGCGCCTCTCTTCT................................AAACC-GT-GT--TCTCATTTAGAGAACAGAAGGACTTCTAGTTAACCCT-G-AGC--TT---.---------    283

COLOBUS.CM.CGU1      C-AGTTTCA-GTCCGCCTTCTC....................................-GCCCT...C....AGAGCGGGGCCTTTTTGCTG....TCCGGGAAGG---GGC--T----.---------    273
SYKES.KE.173         C--TT-TGCTG-G-CACCTCCCGAA-GTG---A.--G-G-GC-.....................................................................CTT---.-------G.    611
RCM.NG.NGM           ---TG-CCAT.T--ATCTCTCAGTGAAACGA--.CTGGGG-T..-GGAT-C-TC-GATA---GT................GGCAGAAGTACTGAAGGCTAAGAA------C-A-C---.---------    864
MANDRIL.GA.14CG      ACT-GTCCAT.T--GTCCTTGAT..............-GAA-CTCTGGTTA-T-GAGA-CCCTCAG-TC-GT-GCAGAGATTTGATATCTAGTGAGAGTAGAGAA--------C-----.---------    380
```

HIV-1/HIV-2/SIV complete genomes

```
                                                                                         Gag p7 nucleocapsid binds these loops, see DeGuzman, Science 279:384-388(1988)
                                                                                         packaging signal secondary structure, see Harrison J Virol 72:5886-5896(1998)
H1B.FR.HXB2          AACAGG.G.ACCTGA...A.........AGCGAAAGG..GA..AA........................CC.....AGA...GGAG.........CTCTCT.CGACGCA...GGACTCGGCTTGCTGAA    710

CPZ.US.CPZUS         ------.-.--T---...-.........-A------..TAC..--........................--AG...--GCTGAA-C.........T--.-GA--...-G..----------C.GCTG.    733
CPZ.GA.CPZGAB        ------.-.-------..-..........-T-----..TA-............................--GG...--GCT.-A-T.........---C-GA--...--C..-----------.GCTG.    727
CPZ.CM.CAM3          ------.-.--T---...-.........--TTG--A-.T-G..--........................--......----...--CTGA.......----TC.-------CG..-----------.GCTG.     82
CPZ.CM.CAM5          ------.-.--T---...-.........--T-----..AAG..--........................--......----...--CTGA.......----TC.-------C..------------.ACTG.    385
CPZ.CD.CPZANT        ....--.-.--T---..................GG--.................................-TAC...----....A-C..........--G-.GA--...-CA..--------TC-.GCTG.     49

H1A.UG.U455          ------.-.--TC--...-.........--------T..T..............................--......----...-A--..........T-----.-------..----------------.    175
H1C.ET.ETH2220       ------.-G------...-.........--------T..---..G-........................--......----...----..........A-----.-------..----------------.     88
H1D.ZR.84ZR085       ------.-.-------..-.........--------T..AG..--........................--......----...----..........A-----.-------..-----------------    224
H1F.BE.VI850         .....-.-.--T---...-.........--------T..AG..--........................--......----...-A--..........A-----.------GG..----------------.     60
H1G.SE.SE6165        ------.-.--T---...-.........--------T..T-..-CAGGGAC.TCGAAAGC...GAAAGTT--......----...-A--..........T-----.-------..----------------.    113
H1H.CF.90CF056       ...........T---...T.........--T-----T..A-..--........................--......----...-A--..........A-----.-------..----------------.     56
H1J.SE.SE92809       ......................................................................................................T-----.-------..----------------.     29
H1K.CM.MP535C        ...................................................................................................................................      0
H101_AE.TH.CM240     ------.-C--TC--...-.........--------T..T-..-TAGGGAC.TCGAAAGC...GAAAGTT--......----...-A--..........T-----.-------..----------------.    277
H1N.CM.YBF30         ------.-.--T---...-.........-A------T..AG..--........................--......G--...--CTGA.......A-----.-------..----------C-T--..    268
H1O.CM.MVP5180       ------.-.-.-GCG...-.........-AGTG--A-..TG..G-A.......................--......--G...-A--AA.......AA-CTC.-------AC.--G--------AGC-G.    716
H1O.BE.ANT70         ------.-.--T---...-.........GTGA--GT-..A-..-.........................--......--G...-A--AA.......AA-CTC.-------AC.--G--------AGC-G.    741

MAC.US.SMM251        ------.A..-T---...-.........G-A--GT-A..---..G.ACTCCT.GAGTACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA-CAA.CC--GAC-GAGTGCTCC    907
MAC.US.SMM239        ------GA..-T---...-.........G-A--GT-A..---..G.ACTCCT.GAGTACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA-CAA.CC--GAC-GAGTGCTCC    920

STM.US.STM           ------.-.--T---...-.........G-A--GT-A..---..G.ACTCCT.GAGTACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCC    584
SMM.US.SMM9          ------.-.--T---...-.........G-A--GT-A..---..G..CTCCT.GAGTACGG...CTGAGTGAA.....G-C...A-CAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCC    402

H2A.GW.ALI           ------.-.--T---...-.........GAA--CT-A..---..-GTCTAG..GAACACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCC    952
H2A.SN.ST            ------.-.--T---...-.........GAA--CT-A..---..-G.CCTTG.GAACACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCC    403
H2A.DE.BEN           ------.-.--T---...-.........GAG--CT-A..---..-GCCCTG..GAACTCGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAA--AA.CC--GAC-GAGTGCTCC    957
H2B.GH.D205          ------.-.-----G.............GAA--GT-A..---..GTCTTCAA.GAGCACAG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAATCAA.CC--GAC-GAGAGCTCC    954
H2B.CI.EHO           ------.-.-------...G........GAA--GT-A..---..GTCTTCG..GGACACGG...CTGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAATCAA.CC--GAC-GAGTGCTCC    953
H2G.CI.ABT96         ------.-.--T---...G.........-AG--GT-A..---..GTCCTTT..GAACACGG...CAGAGTGAA.....G-C...A-TAAG.......GG-GGC.A-GAT-CAA.CC--GAC-GAGTG-TCC    325

TANTALUS.UG.TAN1     ------.-.--T---...G.........TCA-----A..CT..TCACTGAA.GGGGAGGA...AGGGCAT-A.....CC-....-TC-AG.......T--C-C.G-----GGT.AAGTCG--ACGCGG-G-    806
GRIVET.ET.GRI677     ------.-.--T---...G.........-AGAGGCAT..CG..GCACCGAC.CGCTGAGT...TGCTGAG-G.....TCG...-AGAGG.......GA-GAC.TC-G-T-GG.-TGAGA-C--ACGA-TT    618
VERVET.KE.VER9063    ------.-.--T---...GGTCCGAGGACTT--GT-A..AG..GCACGTAC.AGCTGAGA...AGACGTCGG.....-CG...C-TAGG.......AA.C-G.--GG-TGCG.ACGTGACCAAGAAG-G.    831
VERVET.KE.VERTYO     ------.-.--T---...G.........T-A--GT-T..AG..GCACGTAC.AGCTGAGA...AGGCGTCGG.....-CG...C--AGG.......AA.G-G.--GG-TGCG.ACG-GACCAAGAAG--G    318
VERVET.KE.VER155     ------.-.--T-..................--GT-A..AG..GCACGTAC.AGCTGAGA...AGACGTCGG.....-CG...C--AGG.......AA.C-G.--GG-TGCG.ACGTGACCGAGAAG-G.    817
VERVET.DE.VERAGM3    ------.-.--T---...C.........T.T--GT-A..AG..GCACGTAC.AGCTGAGA...AGACGTCGG.....-CG...C--AGG.......AAGG-G.--GG-TGCG.ACGTGACCAAGAAG-G.    317
SABAEUS.SN.SAB1C     ------.-.--T---...C......TAA--A.CGTCCC.-G..CCGCAGAG.TAGACCAG...CCTGAGGA-.....TCC...TTT-AC.......TGAGTA.A---CGGAC.---AGAAAGGC-GACCG    871

SUN.GA.SIVSUN        ------.-.--A---...G.........G-A--G--A..-T..G-CGAAGA.GACGTGGA...GTGAAAC-T.....TC-...-A--AA.......G-GAGC.A-CT--TGA.-A-GAA--AAG-T-TCG    870
LHOEST.KE.LHOEST     ------.-.--T---...-.........G-A---GA-..-C..GCTGGCAA.GAACCTCT...GGAGGGAGG.....-A-...A-C-CG.......-AGAAG.A-G-CAGGA.A-CTAG-A-AGAGAA-G    891
LHOEST.CD.447        ...................................................................................................................................      0
LHOEST.CD.485        ...................................................................................................................................      0
LHOEST.CD.524        ...................................................................................................................................      0
MANDRIL.GA.MNDGB1    ----.-.-.--T---...-.........GAGAGGCAC..TG..-CACTTGA.GGCAGAGC...ACTCCGC-T.....G--...A--AGC.......AGG-TG.AAGGAG-GT.-----G-T--GAAGACG    373

COLOBUS.CM.CGU1      G-----.-.--T---...-..........--A-G.TCC..CT..G-GGGTCG.CGGAAGAA...GAAGAGGAG.....-CT...-ATCGC.......---CTC.A--A-AGAG.-T-AGTC-AAAAAG-GG    363
SYKES.KE.173         ------.-.--T---...C.........TAG-CC---..---..CCCGAGAG.GAGGTGAC...GCCAGTCAA.....--G...-AGTCC.......GGA--C.AAGA-GGAT.C-G--GTC-G-C--AGG    702
RCM.NG.NGM           ------.-.--T---...G..........--T--CT-A..---..G-CCTCCG.ACGCGGGT...CCGGA..--.........----...A---TG.......ACTGG-.A--ACGCGG.ACC-AGTC-AG-TAAG-    953
MANDRIL.GA.14CG      ------.-.--T---...C..........AA-GGTTC..---..GTCGCTCC.GACCTGTA...AGGGACAAGCGGCGGC-GCC--TCGGACAGACC-A-C-GA---A-AGAGTTA-TCAA-GAGC-CCG-    487
```

```
            Gag p7 nucleocapsid binds these loops, see DeGuzman, Science 279:384-388(1988)
            packaging signal secondary structure, see Harrison J Virol 72:5886-5896(1998)
H1B.FR.HXB2       G....CGCGCAC.G.GCAAGAGGC..GAGG.....GG...........CG...GCGAC.TG......GTGAGTACGCCA....AAA...............ATTTT...GACTAGC.GGAGG.....CT    777

CPZ.US.CPZUS      A....GTGCACACA.--G------.----------.C-A.....A...CT-....CA--...-................-A.............................-----...-.C---.A-T--------..    794
CPZ.GA.CPZGAB     A....GTGCACACA.----------.----------.C-G.....C...GA-....T.....-...............-A-T.......................-------...-------.--T--------..    789
CPZ.CM.CAM3       A....GTGCACACA.--G------.----------.C-A.....A...CT-....CA--...-................G.........................T-------...-.C---.--T--------..    142
CPZ.CM.CAM5       A....GTGCACGCA.-T-------.----------.C-A.....C...T-....CT--...-...............-.......AT.TT-...............T-------...-.C---.--T--------..    448
CPZ.CD.CPZANT     C...AGCGCGCGTA.---GA----.----------.C-G..A..A...........-....-...............--------CTT.CATTTTG.................-------...-.T-GCT.A-T-A.....GC    115

H1A.UG.U455       -....GTGCACACA.----------.---A.....-..............-.......-.....................TAA--T...............T-------..-------.-----------..    244
H1C.ET.ETH2220    A....GTGCACTC-.----------.---A.....-..............-.......-.....................AATTTT...............TA-------..-------.-----------..    157
H1D.ZR.84ZR085    A....GCGCGCTC-.----------.----------.----------.-A.......-.....................T..GAA--T...............T-------..-------.-----------..    294
H1F.BE.VI850      A....GTGCA--AC.-GC-AGA-G..CGA-.....A-..............-.......-.....................GAATTTTT...............T-------..-------.-----------..    132
H1G.SE.SE6165     -....GTGCACACA.----------.---A.....-..............-.......-.....................A---...............T-------..-------.-----------..    180
H1H.CF.90CF056    A....GTGCACACA.----------.---A.....-..............-.......-.....................ATTTT...............G-------..-------.-----------..    124
H1J.SE.SE92809    A....GTGCACAC-.----------.----------.----------.-.......-.....................AAAA-T...............T-------..-------.-----------..    100
H1K.CM.MP535C     ..............................................................................................................................    0
H101_AE.TH.CM240  -....GTGCACACA.----------.---A.....-..............-.......-.....................-------..-------.-----------..    342
H1N.CM.YBF30      .....GTGCACACA.--G------.----------.----------.-------GA.A.........---T-AGTA-GC.AATTTT...............GAC-G...-CGGT-G.CC--A.....AA    335
H1O.CM.MVP5180    A....GTGCAC-T-.CT-------.---A------A.....A...CT-A...CAAGA.G-.........------..........-------------A---...-.--G--.--T--------.-C    779
H1O.BE.ANT70      A....GTGCAC-C-.CT-------.---A------A.....A...CT-A...CA.GA.G-.....-------.....................-------...-.--G--.--T--------.-C    803

MAC.US.SMM251     T....ATAAAGGC-.CGGGTC..G..-TAC.....CAG.....A...CG......GCGT.GA......-GAGCGGGAGAGGAGG-GG...............CC-CC...-G...TT.-C--------.TA    975
MAC.US.SMM239     T....ATAAAGGC-.CGGGTC..G..-TAC.....CAG.....A...CG......GCGT.GA......-GAGCGGGAGAGGAAG-GG...............CC-CC...-G...TT.-C--------.TA    988

STM.US.STM        T....GTAAAGGC-.CGGGC...............-AA.....G...CG......GCGT.GA......-GAGCGGGAGGCGAAG-G-..............C-CC...-G...TT.-C-A-.....TG    645
SMM.US.SMM9       T....ATAAAGGC-.CGGGCC..G..-TAC.....CAG.....G...CG......GCGT.GA......-GAGCGGGAGTCGGAG-GG...............CC-CC...-G...TT.-C--------.TA    470

H2A.GW.ALI        T....A-AAAGGC-.CAGGCCAAG..-TAC.....CAA.....A...GG.C...CG-CG.--------.TG---CGG-AGTGAAG-GG...............CC-CC...-GG-G....A--------.TA   1024
H2A.SN.ST         T....A-AAA-GC-.CAGGCC-AG..-TAC.....CAA.....G...GG.C...-GCGT.GT......-GAGCGGGAGTGAAAG-GG...............CC-CC...-GG-G....A--------.TA    475
H2A.DE.BEN        T....A-AAAGGC-.CGGGCC-AG..-TAC.....CAA.....A...GG.C...-GCGT.GT......-GAGCGGGAGTGAAAG-GG...............CC-CC...-GG-G....A--------.TA   1029
H2B.GH.D205       T....GTAAA-GC-.CAGGCC..G..-TAC.....CAG.....G...CA......GCGT.GA......-GAGCGGGAGG-GAAG-GG...............C-CC...-GGAGTA.A..-------.TA   1022
H2B.CI.EHO        T....GTAAA-AC-.CAGGCC..G..-TAC.....CAG.....G...CA......GCGT.GA......-GAGCGGGAGG-GAAG-GG...............C-CC...-GGAGTA.AA--------.TA   1023
H2G.CI.ABT96      T....A-AAAGGC-.CGGGTC..G..-TAC.....CAG.....A...CA......GCGT.GA......-GA-CGGGAGT.GAAG-GG...............CC-CC...-GG-GCG.CAG-......TA    394

TANTALUS.UG.TAN1  -....GCGAAGAG-.ATCC-CCAG..--AA.....AAA.....G...TA-T...-C-G.G-......AG-CAGGG.................................G-G-...TGCA.....GG    862
GRIVET.ET.GRI677  T....TTT--TACC.TAGTC--CG..AGAA.....A-G.....C...TAG-...C--CG.AC......AG-G-CG--GGT.....................CCCA-...T-G-G--.AACCA.....AC    687
VERVET.KE.VER9063 .....-TT-GTGA-.TGGGCTTCT..CGA-.....T-C.....C...GA.A...AG.-A.GC......TC---CCTAGTTAGAGG-C...............TAGGA...AGGGCC..TAGCC.....A-    903
VERVET.KE.VERTYO  A....-TT-GTGA-.TAGGCTTCT..CGA-.....T-C.....C...GG.-...AAA-A.GC......TC---CCTAGTTAGAGG-C...............TAGGA...--GGCCG.TAGCC.....G-    393
VERVET.KE.VER155  .....-T--GTGA-.TAGGCTTCT..CGA-.....T-C.....C...GG.-...AAA-A.GC......TC---CCTAGTTAGAGG-C...............TAGGA...AGGGCCG.TAGCC.....G-    891
VERVET.DE.VERAGM3 .....-TT-GTGA-.TAGGCTTCT..CGA-.....T-C.....C...GG.-...AAA-A.GC......TC---CCTAGTTAGAGG-C...............TAGGA...AGGGCCG.TAGCC.....A-    391
SABAEUS.SN.SAB1C  T....GC--G-GT-.CAGGT-A-A..-GC-.....CCT.....G...CGGA...C-CGA.GA......CACCC-CG-G-TGTGTGTG...............TG-GA...--AG--A.---AAAAGAGAA    952

SUN.GA.SIVSUN     A....-CT-GTGGA.AGGC-CC-G..--A-.....TTA.....A...GGTA...AGTGA.G-.......-CACT---CAGGCACTGTC...............-GGGG...ACAGGC-.A--A.....AG-    946
LHOEST.KE.LHOEST  T....-TGCTCGG-.---G--A-A..--A-.....ACC.....T...GC.T...TGAGA.C-......AG--TA-GAAGGACTTG-C...............TGC-A...--AGGCG.-C--A.....AG    966
LHOEST.CD.447     ..............................................................................................................................    0
LHOEST.CD.485     ..............................................................................................................................    0
LHOEST.CD.524     ..............................................................................................................................    0
MANDRIL.GA.MNDGB1 C....-AG-AGGT-.AGTCAGT-G..--CT.....-AC.....T....TTA...CAAGA.AT......TA-TTGTAC--TAGTGT--...............GGGGC...AG-ATAG.TC--AGCA....    449

COLOBUS.CM.CGU1   A....-TTA-GGGT.-G-GTCAAG..-------.....ACT.....C...TA--...-TCTG.C-......CAA-AGTGAAAGTAAA-GT...............GCGCG...CT-GTCG.C-CA-TCAGGAA    444
SYKES.KE.173      C....--AC-.GGT.-.CT-GT-G..CTC-TG...-TG.....T...ACAC...A---G.AC......-A-GTA-GTGATCCAGCGC...............CAGAC...TG-CCA-.A-CA-TCGGGGA    783
RCM.NG.NGM        -....-AGC-TGGT.-GCTCGTAA..ATCT.....A-C.....G...AGAA...AA-TG.A-......-GCTAGTA-AGCAAGGCG-...............CCCGG...TCAGGTA.CTG--CCTCTAG   1034
MANDRIL.GA.14CG   CGCGCA-GA-T-AA.-GT-AGC-G..TGCACCGTATTGTGG..TAGTGAGT...C-AGG.ATCCGCTT-A-CAGG--AG-TCGCT--GGCAACCCCAGTAGGTAAAA...--AA--A.--G-AAGTAAGG    604
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
H1B.FR.HXB2        A.....GAAGGAG..AGAG.......................................................................................  789

CPZ.US.CPZUS       G.....--CCT---..G-GAAGGG.CGA..AGTCTCTAGG.G..GA...AGAGG.....................................................  831
CPZ.GA.CPZGAB      -......--CCT---..--GAAGGG.CGA..AGTCTCTAGG.G..GG...AGAAG.....................................................  826
CPZ.CM.CAM3        G.....--CCT---..G-GAAGGG.CGA..AGTCTCTAGG.G..GA...AGAGG.....................................................  179
CPZ.CM.CAM5        G.....--CCT---..--GAAGGG.CGA..AGTCTCTAGG.G..GA...AGAGG.....................................................  485
CPZ.CD.CPZANT      -.....ATCCT---..G-GAAGGT.CGA..AGTCTCTAGGAA..CA...GGAAG.....................................................  153

H1A.UG.U455        -...............---A.......................................................................................  256
H1C.ET.ETH2220     -...............----......................................................................................  169
H1D.ZR.84ZR085     -...............----......................................................................................  306
H1F.BE.VI850       -...............----......................................................................................  144
H1G.SE.SE6165      -...........G-..-----.....................................................................................  192
H1H.CF.90CF056     -...............----......................................................................................  136
H1J.SE.SE92809     -...............----......................................................................................  112
H1K.CM.MP535C      ........................................................................................................    0
H101_AE.TH.CM240   -...............----......................................................................................  354
H1N.CM.YBF30       G.....T-G-AGA..GAG-......................................................................................  347
H1O.CM.MVP5180     -......--CCT---..G-GAAGGG.CGA..AGTCCCTAGG.G..GA...GGAAG.....................................................  816
H1O.BE.ANT70       -......--CCT---..G-GAAGGG.CGA..AGTCTCTAGG.G..GA...GGAAG.....................................................  840

MAC.US.SMM251      -......-TGCA-C..-C-AAAAA.GAA..A........TAG..CT...GTCTTGTTATCCAGGAAGG...GATAATAAG...ATAGA..GTGGGAG.........  1040
MAC.US.SMM239      -......-TGCA-C..-C-AAAAA.GAA..A........TAG..CT...GTCTT.TTATCCAGGAAGG...GGTAATAAG...ATAGA..GTGGGAG.........  1052

STM.US.STM         -......-TGCA-C..-AGTAGCC.AGA............GG..GC...TGATTGTTCTACTCTTGAG...GGAAGAGGT...AGAAG..GTGGGAG.........  708
SMM.US.SMM9        -......-TGCA-C..-A-A............GAGTCATAGG..AC...TGAGTTCCCTACTTTTGAG...AAAAGAGTA...GGAGA..GTGGGAG.........  534

H2A.GW.ALI         -......-T-CCTA..CACCAAAA.T.T..GTAGCC.GAAAG..GG...CTTGTTATCCTACCTTTAG...ACAGGTAGA...AGATT..GTGGGAG.........  1095
H2A.SN.ST          -......-TGCCTA..CACCAAAT.ACA..GTAGCCAGAA.G..GG...CTTGTTATCCTACCTTTAG...ACGGGTAGA...AGATT..GTGGGAG.........  547
H2A.DE.BEN         -......-T-CCTA..CACCAAAA.ACT..GTAGCCAGAAAA..GG...CTTGTTATCCTACCTTTAG...ACAGGTAGA...AGATT..GTGGGAG.........  1102
H2B.GH.D205        -......-T-CTCC..-A-....................TTT..TG...CATTGTCTACTCTAAAGAG...GGAGTAGGG...CATAA..GTGGGAG.........  1078
H2B.CI.EHO         -......-T-CTCC..-A-TTT.....................GC...AATTCGTCTACTCTAGGAC...AGGGTACGG...AATAA..GTGGGAG.........  1079
H2G.CI.ABT96       -......-TGCACC..-A-AATTT.GTA..GCCAAAAAGGGC..TG...TGAATCCTACCTTGAATAG...TCAGGTAGA...AGT....GTGGGAG.........  465

TANTALUS.UG.TAN1   T.....AG-TCTA..G-T-GGCG.ACA..GATCAGCACCTG..GG...TGAGCA..........................................................  901
GRIVET.ET.GRI677   C.....C-GTTG-..-CGAAGGG.TTG..GTAGGGGACGGG..TC...GGAGCA..........................................................  726
VERVET.KE.VER9063  -.....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGAC...GG...TACGCA..........................................................  941
VERVET.KE.VERTYO   -.....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGT.G..GG...TACGCA..........................................................  431
VERVET.KE.VER155   -.....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGACG..GG...TACGTA..........................................................  930
VERVET.DE.VERAGM3  -.....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGACG..GG...TACGCA..........................................................  430
SABAEUS.SN.SAB1C   G.....TC--ACA..G-GCTGAC.AAG..GCTGGCTTAGCA..GT...GGCAGAAGGCTAGGTCGAG...CGGGAAAGG...AAA.....................  1016

SUN.GA.SIVSUN      C.....CT---CT..G-G-GGAG.GTC..CCGCTAGGTTAG..TA...CTGTCAAGTAGGCAGACTA...GTAGTACGT...GGAAT..AAAATAAGAAGTGTGTGTATAACTCTGTGTCTCGGAAGAGT  1053
LHOEST.KE.LHOEST   G.....T-------..--TAAGAC.CTC..ACCGGGAGAGTG..GC...AGCAGGCCGGGGGCAGCCT...TGTGAGAAG...TGTGT..AATTCTGATCTAGTGTAACTGGGACACTAGTGGTAGCTTCA  1073
LHOEST.CD.447      ........................................................................................................    0
LHOEST.CD.485      ........................................................................................................    0
LHOEST.CD.524      ........................................................................................................    0
MANDRIL.GA.MNDGB1  ........................................................................................................  449

COLOBUS.CM.CGU1    G.....TGGCAGA..C---AGAG.CAC..AGAGAGGGAGTG..CT...CGCCAGAGAAGGAGAAGAG...CTCATTAGG...AGGTC..TTCAGGCAGGGAGTAGCGCCCCTGGGCAGGTCGAGGGGGGC  551
SYKES.KE.173       -.....AT---G-..-CC.........................................................................................  794
RCM.NG.NGM         G.....-----GT..C--AGTCC.TAG..AGAAGGGAGGAA..A...........................................................  1066
MANDRIL.GA.14CG    CCGAAGC--A-T-AA--TAAAAG.GGC..CTTCCGAAGGGAAGGG...CAGGGGGCAAGAAA.........................................  660
```

```
                                                                 / Gag and Gag-Pol start (p17)
H1B.FR.HXB2        ..................................................ATGGGTGCGAGAGCGTCAGTA...TTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGA    864
H1B.FR.HXB2 Gag                                                      _M__G__A__R__A__S__V__.__L__S__G__G__E__L__D__R__W__E__K__I__R__L__R__P__G__G_    p17

CPZ.US.CPZUS       ..................................................--------------------T...C-G-CA------CG-------GCT--------G------C-T--A--T------    906
CPZ.GA.CPZGAB      ..................................................--------------------T...C---CA------A----------C---------G-----C-T-----C------    901
CPZ.CM.CAM3        ..................................................---------------------...----CA------CG-------GCT------------A--C--------------    254
CPZ.CM.CAM5        ..................................................-----------------------...----CG--A---CG---G---GCT-----G------A--C----M-----A---    560
CPZ.CD.CPZANT      ..................................................-----A---G-G-----T--T...--G--G--A-AGA-GC-----AC--------GT--CA--C-TC----C--T--C    228

H1A.UG.U455        ..................................................---------------------...---------AA-A--------TC------G-----------------------    331
H1C.ET.ETH2220     ..................................................-----------------A--...-----A--C-A-A--------GCC------------AA----------------    244
H1D.ZR.84ZR085     ..................................................---------------------...------------A----G---GCT------------------C-----------    381
H1F.BE.VI850       ..................................................-----------------A--...------------A--------GA--------------A---------G------    219
H1G.SE.SE6165      ..................................................-----------------------...----C-------A--------GCT----------------G------------    267
H1H.CF.90CF056     ..................................................---------------------...--------C---A--------GCT-----G---------C--------------    211
H1J.SE.SE92809     ..................................................-----------------A--...-----T------A--------GAT-----------------G-----------G    187
H1K.CM.MP535C      ..................................................---------------------...------------A--------GC---------G---------C------------     75
H101_AE.TH.CM240   ..................................................---------------------...-----T------A--------GC------------------GC-----------    429
H1N.CM.YBF30       ..................................................--------------------G..-----CA------A---------A-------TC----TAT--G--A---------    422
H1O.CM.MVP5180     ..................................................-----------------T--G...--G-CA---A-TA----G---GC-------CG----A-------------ATCT    891
H1O.BE.ANT70       ..................................................-----------T-----T--G...--G-CA--AA-CA----G---GC-------C-----A------A------ATCT    915

MAC.US.SMM251      ..................................................-----C------AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A--C--C-A--C--C---   1115
MAC.US.SMM239      ..................................................-----C-T----AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A--C--C-A--CAAC---   1127
MAC.US.SMM239 Gag                                                    _M__G__V__R__N__S__V__.__L__S__G__K__K__A__D__E__L__E__K__I__R__L__R__P__N__G_    p17

STM.US.STM         ..................................................-----C------AGC--C--C...--GTCA---AA-A--GC----GA--TA------G--A-----C----C--C--G    783
SMM.US.SMM9        ..................................................-------T----AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A-----C----C--C---    609

H2A.GW.ALI         ..................................................-----C------AAC--C--C...--G--A---A--A--GC---CGA--TA----G----A-----C----C--C---   1170
H2A.SN.ST          ..................................................-----C------AAC--C--C...--G--A---AA-A--GC---CGA--TA-----G---A-----C----C--C---    622
H2A.DE.BEN         ..................................................-----C------AAC--C--C...--G--A---AA-A--GC---CGA--TA------G--A-----C----C--C---   1177
H2B.GH.D205        ..................................................-----C-------GC--C--C...C-GTCA---AAGA--AC----GA--TA------G--A-----C----C--C---   1153
H2B.CI.EHO         ..................................................-----C-------GC--C--C...--GTCA---AAGA-GAC----GA--TA-----GG--A-----C----C--C---   1154
H2G.CI.ABT96       ..................................................-----C------AGC--C--C...C-GTCA---AA-A--GCT---GA--TA------G--A-----C----C--C---    540

TANTALUS.UG.TAN1   ..................................................-----A---G--CAC--G-C-...C-GTCA---A-GA-T--G--CAC--TT--G-----A--C---C-C--GAAC---    976
GRIVET.ET.GRI677   ..................................................-----C-G-G-TCAC----C-...C-GTCA---A--AGCC-C--CACG-TC--G--G---A--C--C-T--GAAC--G    801
VERVET.KE.VER9063  ..................................................-----G---GCTA-C----C-...C---AA---A--C-------C-A--TC--G-----A--A---C-C--GAAC---   1016
VERVET.KE.VERTYO   ..................................................-----G---GCTA-C----C-...C---ATA--A--C-------C-A--TT--G-----A--AC-TC-C--GAAC---    506
VERVET.KE.VER155   ..................................................-----G---GCTA-C----C-...C-G-ATA--A--C--------GA--TT--GC-T--A--AC-TC-C--GAAC---   1005
VERVET.DE.VERAGM3  ..................................................-----G---GCTA-C----C-...C---ATA--A--C-------CAA--TT--GC-T--A--AC-TC-C--GACC---    505
SABAEUS.SN.SAB1C   ..................................................-----------TAAC-----C...-----T--TA-GA--C-----GC--TC---TC-G-CA-----C----CAAC---   1091

SUN.GA.SIVSUN      GTCTCAAAC........................................-----A---G--AAC--CAC-...G-GGATA---A--TTG-GAGGA-C-TT----G-G-AGCTC-G-AA--T--A--G   1137
LHOEST.KE.LHOEST   CT...............................................-----AT--G-TAAC--T--C...C-C--TA--CA-ATTGA-A-AGAT-TTTGT-GTG-CA--C----A--T---A-T   1150
LHOEST.CD.447      ..................................................-----AT--G-GAAC--T--C...C----TA--CA-ATTGAG-GAGA--TCTGC-GT--CA--C----A--T-ACA-T     75
LHOEST.CD.485      ..................................................-----AT--G-GAAC--T--C...C----TA--CA-ATTGAG-GAGA--TCTGC-GT--CA--C----A--T--CA-T     75
LHOEST.CD.524      ..................................................-----AT--G-GAAC--C--C...C-C--TA--CA-ATTGAGA-AGAT-TTTGT--TG-CA--C----A--T---A-T     75
MANDRIL.GA.MNDGB1  ..................................................------AATG-GAAC--T-CC...--GTTA---ACT--T--G---AA--TT--G-----AA-A----A-AG---T--T    524

COLOBUS.CM.CGU1    CTGCAAGTGGAGTAAGCTGCGGCGCTCCAGGACCGGTAGAGGGAGGTCATTA-----CAACGAGCA-GG-C-G...---G-GAA-AA-AC------AGACCTAC-G---G-C---C-G-A-AA----AAG    678
SYKES.KE.173       ..................................................-----A--AGCG-G---G-C-ATA----CA--TA----GC----CA---AT--G---------C-CC-A--GAAA--G    872
RCM.NG.NGM         ..................................................----------G------A-T.....-GTCA---AAGA-G------GCT------C-GG--A-----C----C--C---   1141
MANDRIL.GA.14CG    ..................................................-----C-----C-------GT...C-T--G--A-AGA----G---GA-CT-----C----A-----C----CTCC---    735
MANDRIL.14CG Gag                                                     _M__G__A__S__A__S__G__.__L__R__G__E__K__L__D__E__L__E__T__I__R__L__R__P__S__G_    p17
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2        AAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTC   994
H1B.FR.HXB2 Gag    _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__Q__L__Q__P__S__L__   p17

CPZ.US.CPZUS       --------G----TGA-G------T----T-----------------G---A----T--TTG---C--A--T--AA-G------G----C------CTC---T--T-AAA----T--G-----G-T--AA  1036
CPZ.GA.CPZGAB      -GA----G-----TGA-G------T---------------------G---A--------TG-G-C--C--G--AA-G----GTAAG-----A----CTA--T-GT-ACA---AT--G-G---G-T--CA  1031
CPZ.CM.CAM3        -------------TGA-G------T-----------------A-----G--CA-------TTG---C-----T--AA-G-----GG-------A---GAG--GT--T-AA-G--AT--G-----G-T--CA   384
CPZ.CM.CAM5        -------------TGA-G------T----------------------G---A----T--TTG---C-----T--AA-G--G---G-T-----A---GA-A-TC----AAA----T--G-----G-T--CA   690
CPZ.CD.CPZANT      --------G--C-TGA--------C-G--T--------ATC------GC-G--T--T--GC-C-GCT-CTC---TC----------------T---GA-A-GGCTA-CCAT--AT-GAGC--T---A-AG   358

H1A.UG.U455        --C----------G-C-G------T----------------------G---AA----A--C----C--------T---------G-------A---CAG--------------AT--------G-T--C-   461
H1C.ET.ETH2220     ------C-C----TGC-G-----CC----C-------A---------G---AA---T---C----C----A---T-----T-------C--------A-------A-TAA-------------G-T----   374
H1D.ZR.84ZR085     --------G-----GC-------------------------A-----G--------T---C-----------T----------------------A-------A---A-------------G-T---A   511
H1F.BE.VI850       -------G-------A-G------C--A---------------------------T---C--G----------TC--------------------CA-A-----A-AA--------------------   349
H1G.SE.SE6165      -G----TC-------A--------C----------------------G--GA----T---C----C----A---T---------G-------T---CA-------A--A-----T-G--------T--C-   397
H1H.CF.90CF056     -------------GGC--------C----------------------G---A----T---C----C--C-----T---------C-----------CT---G---A-A-A----A----G---G-TA--A   341
H1J.SE.SE92809     -------------GGA----G---C----------------------G--CA----T---C----C--------TC----GT--G--A--------CA---------A-T------C------G-T--C-   317
H1K.CM.MP535C      ---------------C-G------C-----------------------------T---C-C--C--------T-----G---A-----------C-G------A-AAC---AA----G------A---   205
H101_AE.TH.CM240   -GA----------GGC-G------T----------------A---T-----A--------C----C---A--T-T---------G-------A---CA-------A-A-A----T----GT--A-T--CA   559
H1N.CM.YBF30       -----------C-G-A-G------T----------------------G---A-------TTG---C--A--T--CA-G--C---G-G--C------GCCA-GT----AAAT--AT--G-----G-T--CA   552
H1O.CM.MVP5180     --A--GGC-----GGC--------T----------------------G---A---A----TG---------T--A--------TG-------TAC-GAG---C-G--ACAG---T--G-G---G-T--CA  1021
H1O.BE.ANT70       --A--G-------G-C-------CT----------------A-----G---A--------TG--------AG--AC-------TG----G--TAA-GAGA-GC-G--ACAG---T--G-G---G-T--CA  1045

MAC.US.SMM251      --------G--C-TG--G--G---G-----------GCA-AT--AT----TA----T-G-T-AGCAGAAA-------G--G-ACAA------A---CA-A-------TTCGGTCT--GCT----TAG-G-  1245
MAC.US.SMM239      --------G--C-TG--G--G---G-----------GCA-AT--AT----TA----T-G-T-AGCAGAAA-------G--G-ACAA------A---CA-A-------TTCGGTCT--GCT----TAG-G-  1257
MAC.US.SMM239 Gag  _K__K__K__Y__M__L__K__H__V__V__W__A__A__N__E__L__D__R__F__G__L__G__E__S__L__L__E__N__K__E__G__C__Q__K__I__L__S__V__L__G__P__L__V__   p17

STM.US.STM         --------G--C-TG--G------G----G------GCA-AC--AT-G--CA----T-G-T-GGCAGAAA-------G--GT--AA------G---CA-A----TA-AACGGTCT--G------TAG-G-   913
SMM.US.SMM9        --------G--C-TGC-G--G---------------GCA-AT--AT-G--CA----T-G-T-AGCAGAAA-------G--G-ACAA------T---CA-A----T--TTC-GTTT--GCT----TAG---   739

H2A.GW.ALI         ------------C-GC----------T--G------GCG-AT--AT-G--CA------G-T-GGCAGAAA-------G--GT--AA------T--CCA-AG---T--TAA-GTTT--G------TAG-G-  1300
H2A.SN.ST          -------------GGC----------T--G------GCG-AT--AT-G--CA------G-T-GGCAGAGA-------G--GT--AA---G--T--CCA-A----T--TAC-GTTT--G-T----TAG-A-   752
H2A.DE.BEN         --------G--C-GG-----------T--G------GCG-AT--AT-G--TAA-----G-T-GGCAGAGA-------G--GT--AA------T--CCA-A-G--T--CA--GTTT--G-T----TAG-A-  1307
H2B.GH.D205        --------G----TGC--------G-----------GTG-AC--AT----CA----T-G-T-AGCAGAGA-------G---T--AA------A---CA-A----T--TAA-GTCT--GCT----TAG-A-  1283
H2B.CI.EHO         -------G---C-TGC--------------------GTG-AT--AT----GA----T-G-T-GGCAGAGA---G---G-G-T--AA------A--CC-GA----T-G-AA-GTCT--GG-----TAG-A-  1284
H2G.CI.ABT96       -----------C-TG---------G-T--G------GCG-AC--AT-G--CA------G-T-AGCAGAGA-------G---T--AA------T--CCA-A----T--TTC-GTAT--GCT----TAG-A-   670

TANTALUS.UG.TAN1   --------G---C----------CT--A--------G---AAA-AA-G--G--C--T-GGC-CC-CGAGAAA--C---------AA---G------CA-A-G--CA-A-A-GTAT-GAGC--GCTAGAA-  1106
GRIVET.ET.GRI677   --A--G--G--CC--A-T------T--A--------G-A-AA--AA-G--------T-GGT-AC--GAGAAA--T---------AA----------CA-A----CA-A-A-GTTT--ACC--G-TGGAA-   931
VERVET.KE.VER9063  ------------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-GTC--TTTT--T-TAGAA-  1146
VERVET.KE.VERTYO   --------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-GTC--CT-C--CCTAGAA-   636
VERVET.KE.VER155   --------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-GTT--CTCT--CCTAGAA-  1135
VERVET.DE.VERAGM3  --------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-GTA--CT-C--GCTAGAA-   635
SABAEUS.SN.SAB1C   -----G--G-----G----G----C---------------A---A-----T-----TAGCC--TCAG-AAAT--C-----G---AAG-----GGTCGTCA----T--TA-TGTA--C-TT--CCTAG---  1221

SUN.GA.SIVSUN      --A--G-CC--CC-GGC--G----G-----------G-T-AA--AT----TA-G--T-G-C----CAAG-AA------AG----GTG---------CAGA-G--T---TC-GTCTGTTGG--T-TGTA-G  1267
LHOEST.KE.LHOEST   --A--G-CT---C--AA--G----G---A--------CG-AA--AT----CA----T-G-T-GGGAT-ACAA--C--G------G-------G----AGA----C--ATCTGTCTGCTGG--T-TATA-G  1280
LHOEST.CD.447      --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---GTCTGTTGG---CTATACT   205
LHOEST.CD.485      --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---GTCTGTTGG---CTATACT   205
LHOEST.CD.524      --A--G-CT---C-GAA--G----G---AC-------CG-AA--A-----TA-G--T-GGC-AGGAG-AAA---TC--------G----G-------AGA-G--TT-----GTC-GTTGG--T-TATA--   205
MANDRIL.GA.MNDGB1  --A---TGT----G---G-----CC-CTGT---TGT-AAG-T--AT----TA----T-GCT-ATCGGA-AAA--CC-T------CAGC----A---GA-A----C--CTC-GTATGTTGG----TATA-G   654

COLOBUS.CM.CGU1    --AGG-TGT-----GA--------G-TAG----ATGT---CA---G--AGC--C-GT-TTT-A-TATT--AG------A-GT--G--AC---AGTGGCT------T---A-A--G-GACG----TAG-GG   808
SYKES.KE.173       -----GCGC--CCT-G-TCG----T----T-------AG-A----T-G--CA------GCC--TCGGACCAA--CA-G--GT-TAA----------GA-A----T--CAC-GTC----TT----TAGAAG  1002
RCM.NG.NGM         --------G----TG-----G---C-----------T-------A-----GA----T-GCT-G-G-GA-ACTT---------GTG-------AATCCAGA----TA-A--GGTTA-T-TG--TCTAG---  1271
MANDRIL.GA.14CG    -----G-----CC-G-----------A------T-----A---A-----T-----T-GCC--C--GAAAAA--A--------TCAG-----A--CGAGA----T--TA-CGTA--CTTT---CTAG---   865
MANDRIL.14CG Gag   _K__K__K__Y__Q__L__K__H__I__I__W__V__S__K__E__L__D__R__F__G__L__H__E__K__L__L__E__T__Q__E__G__C__E__K__I__L__S__V__L__F__P__L__V__   p17
```

```
H1B.FR.HXB2       AGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAA..........  1113
H1B.FR.HXB2 Gag   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__E__E__Q__.__.__.__.  p17

CPZ.US.CPZUS      -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--G-TA.............  1152
CPZ.GA.CPZGAB     -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       ---------------GC--GC----GC-T-T------TC-----GTG--T-GG-----T---A-G--A--TTCAG--G-G-----AC--C-G------ACA--AT-GA-A----CAGTT...........  503
CPZ.CM.CAM5       G------------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---ACAGTT...........  809
CPZ.CD.CPZANT     -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-AATGA-AGT-...........  477

H1A.UG.U455       --------A-----------------------------------GT------------A--------------TG----------------------A-T--A---------AT----------------  580
H1C.ET.ETH2220    --------A----G-------A--------T---------G-----T-----------A-------A--------------------------C--------------------A-------------  493
H1D.ZR.84ZR085    --------------------------A-----------------------------A---A-----------------------------------------------------A-------------  630
H1F.BE.VI850      ----------------G----AG-------T-------------GT--------A---A---------GC--G-G---C------------------------C----------A-------------  468
H1G.SE.SE6165     --------A----G---A---A-------------G---------------------A---------------G-------------A------C----AG-AG-G--AA-GATA-------------  516
H1H.CF.90CF056    --------A------------A--------T----CT-------GT---------C--A----G--AA-----TG-G--------------G--------T------------ATA-------------  460
H1J.SE.SE92809    --------A--C-----A---A------G-------------------------C--A-----G--------A-------------T------------G------------ATT-------------  436
H1K.CM.MP535C     -A--------------GA---A----C-----------A-----GT--------T---A-------A-------G-------------------C--------AC----------A-------------  324
H101_AE.TH.CM240  ---------T-----------A--------T--------------------GG--C--A--C------------G-------------------------T--A----------TA-------------  678
H1N.CM.YBF30      -------G-----------GC-C--T--------CG-TC-----GTT--T--------C---AGT------C-----C-CA----AC----------G---------A-A--GA-A--G...GAACAGCA  679
H1O.CM.MVP5180    -------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------TAATG...........  1140
H1O.BE.ANT70      -------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------TAATG...........  1164

MAC.US.SMM251     CA-----C------A-TT-A-A-AGCC-T--------T--CTGCGT-A---GG--CA-T--CGC-GAAGAGA-AG-G---C----TG-------AAA-C-G-TAG-G...C-GAGA--CCTAGTGGTGGA  1372
MAC.US.SMM239     CA-----C------A-TT-A-A-AGCC-T--------T--CTGCGT-A---GG--CA-T--CGC-GAAGAGA-AG-G---C----TG-------AAA-C-G-TAG-G...C-GAGA--CCTAGTGGTGGA  1384
MAC.US.SMM239 Gag P__T__G__S__E__N__L__K__S__L__Y__N__T__V__G__V__I__W__C__I__H__G__E__E__K__V__K__H__T__E__E__A__K__Q__I__V__.__Q__R__H__L__V__V__E  p17

STM.US.STM        CA-----T------A-TT-A-AGAGTC-T-T------T--CTGCGT-A-T-GG---A-C--CGC-GAAGAGA-AG-G---C-T--TG-------AAAGC-AGT-G--...A-GAGA--TCTTGTGGTAGA  1040
SMM.US.SMM9       C------T------A-TT-A-AGAGCC-T-----C--T--CCGCGT---T-GG--CA-T--CGC-GAAGAGA-AG-G---C-T--TG-A-----AAA-C-A-TAG-G...C-GAGA--TCTAGTGGTGGA  866

H2A.GW.ALI        CA------------A-TT-A-A-AGCC-T-T------T--CTGCGTAG-T-GG--C-----CGC-GAAGAGA-AG-G-----T--TG-A-G---AAA-C-A-TA---...C-GAGA--TCTAGCGGCAGA  1427
H2A.SN.ST         C------G------A-TT-A-A-AGCC-T-T------T--CTGCGT-A-T-GG---A-A--CGC-GAAGAGA-AGCG-----T--TG-A-----AAA-C-A---G--...C-GAGA--TCTAGTGGCAGA  879
H2A.DE.BEN        CA-----G------A-TT-A-A-AGCC-T-T------C--CTGCGT-A-T-GG--CT----CGC-GAAGAGA-AG-G-----T--TG-------AAAGA-ACTAGC-...C-GAGA--TCTAGTGGCAGA  1434
H2B.GH.D205       CA-----G------A-TT-A-A-AGCC-T-T-----TC--CTGCGT-A-T-T---CC----CGC-GAAGAGA-AG-G-----T--AG-------AAA-A-G-TAGC-...C-GAGA--TCTAGCGGCGGA  1410
H2B.CI.EHO        CA-----T------A-TT-A-A-AGTC-T--------C--CTGCGT-A---T---CC----CGC-GAAGAGA-AG-G-----T--TG-------AAA-A-G-TAGC-...C-GCGA--TCTAGCAGCGGA  1411
H2G.CI.ABT96      CA-----T------A-TT-A-AGAGCC-T--------T--CTGCGT-A-T-GG---T----CGCCGAACAGA-AG-----C-T--TG-A-----AAA-C-AGTAG-G...C-GAGA--TCTAGTGGTAGA  797

TANTALUS.UG.TAN1  CA------------G---A-A-AGTC-G-T----TTGTGCTGCGT-A-T-GG---A-T--CGC-GAACAGA-AG-G-----T--AG-------AG-G-TATTAG--A-AC--CGCTGCCATCTAGTGGA  1236
GRIVET.ET.GRI677  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--C-CT-CCATCTAGTGGA  1061
VERVET.KE.VER9063 CA-----C--G--G-GCT-A-A-AGTC-G-TC---CTT--GTGCGTG--A-----CA-C--C-GCGAACAGA-AG-G--------AG-------GG---CA-CAG--AGAC--TGCTGCCATCTAGTGGA  1276
VERVET.KE.VERTYO  CA--------G--G-GCT-A-A-AGTC-G-TC---CTT--GTGCGTA--A-----CT----CA-GGAACAGA-AG-G--------AG-------AG---CA-CAG--AGAC--C-CTGCCATCTAGTGGA  766
VERVET.KE.VER155  CA-----G--G----G-A-G-A-AGTC-G------CTG--GTGCGTAT-GCT---C--C--C---GAA-AGA-AG-G--------AG------------CA-TAG--AGAC--TGCTGCCACCTAGTGGA  1265
VERVET.DE.VERAGM3 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--TGCTGCCATCTAGTGGA  765
SABAEUS.SN.SAB1C  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--G--AGTGCCAGCAGAAAT  1351

SUN.GA.SIVSUN     CAT-T--T--TA-GA-C--G-AGG-C---GTAGG-------TGTGTG--TGCC---TGT------G-C--TCCAG----------AC----G---C--A-A---G-CAGACT-----C-GCAAAAAATGA  1397
LHOEST.KE.LHOEST  CA--T--T--TA-GA-C--A-AGG-C---GTAGGG------TGTGT-A---GC--CTGT--C-T-G-AG--AGA---TC------TC-------CA-CAGT--AG--A-AAT---A-C-GCACCAGAAGC  1410
LHOEST.CD.447     CC-----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-AAT---A-C-ACTCAGGAAGA  335
LHOEST.CD.485     CC-----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-AAT---A-C-ACTCAGGAAGA  335
LHOEST.CD.524     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-AATT-----GAACAAACCTGT  335
MANDRIL.GA.MNDGB1 -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-TCATAAC-AGAAAGGAAGA  784

COLOBUS.CM.CGU1   -C-------------TGT-G--G--CC-----GGA...TGTT-TGTGTGT--C--CC----CAG--AATGGA-T--TG-G-----TC-A--G--AGA-A-G--AG----A----CTT-TAAAAAGCAGGC  935
SYKES.KE.173      CT-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----CATGTAATTGGAAAGA  1132
RCM.NG.NGM        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-GC--TGCCATCTAGTGGA  1401
MANDRIL.GA.14CG   CT-----G-----GA-T---CTTG-GC-G--C--C--CTGTTGTTG-A---GG--C--A---GCG-AAG-GA-AG---CT--T--AG-A--G--AAA-A-G--AG----AC--TGCTTCCATCTAGCGGC  995
MANDRIL.14CG Gag  P__T__G__S__E__N__L__L__A__L__Y__N__T__C__G__G__I__W__C__V__H__G__K__V__K__V__T__D__T__E__E__A__K__K__K__V__E__Q__C__F__H__L__A__A  297
```

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2       .AACAAAAGTAAGAAAAAAGCACAGCAAGCA............GCAGCTGAC.....................ACAGGA..................................................  1158
H1B.FR.HXB2 Gag  __N__K__S__K__K__K__A__Q__Q__A__.__.__.__.__A__A__D__.__.__.__.__.__.__.__T__G__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   p17

CPZ.US.CPZUS      .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       .-CGGCCTCCCGAG--C-G-A-GT-GCCCA-............C-GCAGC-G.....................CAGCAGCAA.................................................   551
CPZ.CM.CAM5       .GC-GCCCAAG-A...C-G-A-GT-GCCCA-............C-GCAGC-A.....................CAGCAGCAG...............GAA..............................   857
CPZ.CD.CPZANT     .-TGC-G-CAC-AGC-G--A--GGAAGTAGCCAAACCGCAAGCAG--GCATGCTTCTGCGGCTGCTCCTGTTA-ACAA-CAG......TGGTGTCAGCGA..............................   570

H1A.UG.U455       .--T--G-AC---C---GGA-------G---............------A--.....................------..................................................   625
H1C.ET.ETH2220    .---G-----C--C-----A-------G---............-G---A-CT.....................GACA--..................................................   538
H1D.ZR.84ZR085    .-----C----------GG-----------.............-------G-.....................------..................................................   675
H1F.BE.VI850      .---------C--C----GA---------G.............---------.....................-A---G..................................................   513
H1G.SE.SE6165     .--G------C--G-----AT------G---............---ATG--T.....................-A----..................................................   561
H1H.CF.90CF056    .---------C--C-----A-----------............--------T.....................-AG-A-..................................................   505
H1J.SE.SE92809    .-----G-AC--AC-GC-G------A-----............-A-A-----.....................-A-AA-..................................................   481
H1K.CM.MP535C     .---------C--CG---GA----A---...............................................-A-GCAGCT..........................................   360
H101_AE.TH.CM240  .--T--G--CC--CG---GA-------G---............-------G-.....................------..................................................   723
H1N.CM.YBF30      C--GCCCGAGCCA-----CC--G-AGC--GG............-----G-CA.....................G--ACTGAT...............................................   728
H1O.CM.MVP5180    .GCA-GC--G---TCTGC--AGGCCGCTAAG............-A--AAACA.....................-GC.....................................................  1182
H1O.BE.ANT70      .GGG-GC--G---TCTGCG-ACGCCGCTAAG............-A--ACACA.....................-GC-C-..................................................  1209

MAC.US.SMM251     A..........................................A---GAACA.....................G---A-ACTATGCCAAAAACAAGTAGACCAACAGCA....................  1418
MAC.US.SMM239     A..........................................A---GAACA.....................----A-ACTATGCCAAAAACAAGTAGACCAACAGCA....................  1430
MAC.US.SMM239 Gag __.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__G__T__.__.__.__.__.__.__.__T__E__T__M__P__K__T__S__R__P__T__A__.__.__.__.__.__.__._   p17

STM.US.STM        G..........................................A-T-GAACT.....................G--AACAAAATGCCTGCCACAAGTAGACCAACAGCA....................  1086
SMM.US.SMM9       A..........................................A---GAACA.....................G---ACAAAATGCCAGCAACAAGTAGACCAACAGCA....................   912

H2A.GW.ALI        A..........................................AT--AAACA.....................G---AGAAAATGCCAAGCACAAGTAGACCAACAGCA....................  1473
H2A.SN.ST         A..........................................A--AAAACT.....................----A-AAAATGCCAAGTACAAGTAGACCAACAGCA....................   925
H2A.DE.BEN        A..........................................A-T-GAACT.....................G---AGAAAATGCCAAATACAAGTAGACCAACAGCA....................  1480
H2B.GH.D205       C..........................................A---AAA-A.....................-TGCC-GCTACAAATAAACCAACAGCA...........................  1447
H2B.CI.EHO        C..........................................A---AAA-A.....................-TGCC-GCTATGAGTAAACCAAGTAAA...........................  1448
H2G.CI.ABT96      A..........................................A-T-GAACT.....................----A-AAAGTGCCTGCCACAAGCAGACCAATAGCA....................   843

TANTALUS.UG.TAN1  A--AG---AA-CTGC-GC---G-CATCT-GT............-GCCAGC-G.....................CA-AATTACAACACAGCTGCG....................................  1297
GRIVET.ET.GRI677  C--A--TGAG--AGC-GCTAA-A--A--AAT............-AGA-AACA.....................G-G.....................................................  1104
VERVET.KE.VER9063 A--GG---AA-GAGC--C--AG-CATCTAGT............-G--AAA-G.....................-A-A-TAACAGGGAGACAACAGCG.................................  1340
VERVET.KE.VERTYO  A--AG---AA-GTGC--C--AGACATCTAGT............-G-CAAA-G.....................-A-AATGACAAGGGAATAGCAGCG.................................   830
VERVET.KE.VER155  C--AG---AA-CTGC-GTTA-G-CA-CT-GT............-G-CAGC-G.....................-A-AATAACACAGGAGGAACAGCG.................................  1329
VERVET.DE.VERAGM3 G--AG----A--TGC-G--AG-A-TAC-A--............-AGA-ATCT.....................-GT---CAAAAGAAAAATGACAAGGGAGTAACAGTG....................   841
SABAEUS.SN.SAB1C  G-CAG-----GCC-C-GCGA--TCTAGT-GC............CA-A--A-G.....................GA-CTGCAGGCAAAAAAGAAAAATGAGCCAACAGTG....................  1427

SUN.GA.SIVSUN     AG-A..............................................................................................................................  1401
LHOEST.KE.LHOEST  AGCAGGG...........................................................................................................................  1417
LHOEST.CD.447     A-GGC-G...........................................................................................................................   342
LHOEST.CD.485     C-GGC-G...........................................................................................................................   342
LHOEST.CD.524     G.................................................................................................................................   336
MANDRIL.GA.MNDGB1 A--GC-GGAGG-TG--..................................................................................................................   800

COLOBUS.CM.CGU1   C-TG-T-GAA-T-GCC..........................................................................................................AGCAAG   957
SYKES.KE.173      TG--CC-CCAGC--C-TCT-GTGGA---AGT............-A-AA-AG-.....................-GTCA-AACATGGCTAGTGAGACATCTAGTGGCCAAAAGGTAGTACAGCAGGAAAAA  1229
RCM.NG.NGM        C--AG-TGAG--TGC-GG--A---AG--AA-............-G---CACA.....................GTGAC-...................................................  1447
MANDRIL.GA.14CG   C--AGG-GAA-GTGC-GCCT--G-AA---A-GAGAAAGCAACA--GA-ACCTAGTGGCCGC.....................................................................  1056
MANDRIL.14CG Gag  __K__G__E__S__G__G__S__E__K__E__E__K__A__T__A__T__P__S__G__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   317
```

```
                                                Gag p17 \   / Gag p24
H1B.FR.HXB2        ......CACAGCAAT...CAGGTCAGC.........CAAAATTAC...CCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAG   1267
H1B.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__W__V__K__V__V__   p24

CPZ.US.CPZUS       ..........--T--C...ATT-GT---.........AGC---------...--AG--A-A------GCT-----A--G-------------A--G--T--C--------------C----------CG----   1315
CPZ.GA.CPZGAB      .........GCTGGC......A-T--T.........GG----------...--CC-C--A--A--TGCT--A-----G-----G-----------C-----C-----C--G----------------G----   1328
CPZ.CM.CAM3        .........GA--GC...GCT-----T.........AG---C---------...--AG----A-----TGCA--A--A--GC-A--G------C----G-----C--G--C--------T-----T------A---   657
CPZ.CM.CAM5        .........----GC......A-A--T.........AG---C---------...--A-----A-----TGCA--A--A------A-------------G--------G--CC-------T-----C--G---A---   960
CPZ.CD.CPZANT      .........CATCT-AGTGGC-AAG-G.........AG----------...--C--CA-AGT-G-TGCAGGA--AAT-GCAAGG------C-AC-GA----A-----C--------C--------GTGT----   679

H1A.UG.U455        ......................AG---T.........------------...--C--------A--TGCA--A------CCA-----C------T----------G--C--G-----------G----------   722
H1C.ET.ETH2220     ......GGA-AGG-C...........--T.........--------T...--------------T--G--------------------C--------G----------G----------------------   641
H1D.ZR.84ZR085     ......A------GC...---------.........--------T...--------A------C-A--A--------------C------C---------------G--C--------------A---   784
H1F.BE.VI850       ......................-----T.........----------...--------A-----TC-T-----A-----------C---T-TC-------------------------------G--GA---   610
H1G.SE.SE6165      ......A-------C....-A-----.........--------T...--------------TGCA--A--------------C---------A------------G----------------------   670
H1H.CF.90CF056     ......A-AGA---C...A-------T.........--------T...--------A-----TGCT--A-----G--------C-----------------G--C------------------------   614
H1J.SE.SE92809     ......G---A--G-...--------T.........--------T...--------------TC-G--A------CC------C------C--------------------------------GA---   590
H1K.CM.MP535C      .........GA---A...GG------T.........----------...---------A-----TC-G-----------------C------C----------------------------G--G---A---   466
H101_AE.TH.CM240   ......AG-----GC...A-A-----.........----------...-----------A--TGCA--A----------C-------C-TT---------------G-----------G----------   832
H1N.CM.YBF30       ............GC...A-TA----T.........AGG-----T...----C----C----CTGCT--A--A---------------C-GC-GA----C-----C--------T-----G-----GA---   831
H1O.CM.MVP5180     .........CCT-GG...---ACA--T.........----------...--------AACA--TGCA-----A--------------A-----C--C--C--G-------------------G-C-----   1288
H1O.BE.ANT70       ............-GG...----CGG-T.........----------...--------ATCA--TGCG-----A--------------------C--C--C--G-------------------G-C-----   1312

MAC.US.SMM251      ......-CATCT-GC...GGCAGAG-A.........GG----------...--AG--CAA--A-TAGGT...--TA-CTAT--C--C-T-C-AT--AGC--G-----A--------C---------T-GA---   1524
MAC.US.SMM239      ......-CATCT-GC...GGCAGAG-A.........GG----------...--AG--CAA--A-TAGGT...--TA-CTAT--C--C-T-C-AT--AGC--G-----A--------C---------T-GA---   1536
MAC.US.SMM239 Gag  _.__.__P__S__S__.__G__R__G__.__.__.__G__N__Y__.__P__V__Q__Q__I__G__.__G__N__Y__V__H__L__P__L__S__P__R__T__L__N__A__W__V__K__L__I__   p24

STM.US.STM         ......-CACCT-G-...GGCAGAG-A.........GG----------...--AG--CA---AGTAGG-....--CA-TTAT--C--C-TAC-AT--AGC--A-----C--G--------------GT----G-   1192
SMM.US.SMM9        ......-CACCT-G-...GGCAGAG-A.........GG----------...--AG-GCA---AGTAGGT...--CA-TTATACC--C-TAC-TT--AGT--A-----A--------T---------T-G----   1018

H2A.GW.ALI         ......-CACCT-G-...G-ACAGG-G.........GG---C-T-----...--CG--CAA--AGTAGC-....--CA-CTACACC---GT-C-GC-GAGC--CC----C---G----T---------T------   1579
H2A.SN.ST          ......-CACCT-GC...GG-AA-G-A.........GG---C-T-----...--CG--CAA--AGTGGC-....--CA-CTATACC---GT-C-AC--AGT--CC----CC-------T---------C------   1031
H2A.DE.BEN         ......-CACCT-G-...GG-AAA--A.........GG---C---------...--CG-GCAA--AGCGGGT...--CA-CTAT--C---GT-C-AC-GAGC--CC-----C----------------T----G-   1586
H2B.GH.D205        ......-CACCT-GC.........G--.........GG------T-----...--AG-GCA---ACTGGCT...--CA-CTAC--C--C-T-C-GC--AGC--CC----C--------T--------GT------   1547
H2B.CI.EHO         .........CCA-C-...AGC...C--.........-T-GC---T-----...--AG-GCA---A-TAGCT...--CA-TTATTCC--C-T-C-GC--AGC--CC----C--------C---------T------   1548
H2G.CI.ABT96       .........CCACC-...AGT...G--AGAGGA...GG---C---------...--AG-GCA---AGTAGG-....--CA-CTAT--T----T-C-AT--AGT---C----------------------T-G----   949

TANTALUS.UG.TAN1   ......ACACCT-G-...GGCAGACAT.........GGT---------...---G----A--AC-G-AT...AAT---TG-------ACTC-TT-G-----AC-T--CC-------------T---AC---T-   1403
GRIVET.ET.GRI677   ......-CACCTGG-...GGC-AATCA.........AG----------...--AG----AA-TC-G-AT...AATGCCTG------C---C-TT-G--T--GC-C--G--------G-----C---TGC--G-   1210
VERVET.KE.VER9063  ......-CACCTGG-...GGC...--T.........-------T-----...-AGCGCAA--AC-AGGA...AATGCCTG-------GTTC-AC----C---C-C--A--------G----------C-----   1443
VERVET.KE.VERTYO   ......-CACCTGGG......TG---T.........--G----TT-----...-AGCGCAA--AC-AGGA...AATGCCTG-------GTAC--T-G-----GC-C--C--------G----------C-----   933
VERVET.KE.VER155   ......ACACCTGG-...GGC...---.........-------TT-----...-CGC-CAA---C-AGGG...AATGC-TG---G---GTAC-AC-T------C-C--CC------------------C-----   1432
VERVET.DE.VERAGM3  .........CCACC-...GGT-G---T.........-------T-----...-AGC-CAA--AC-GGGA...AATGC-TG-A-----GT-C--T-G-----AC-C--C--------G----------C-----   944
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-----------GA-T-   1533

SUN.GA.SIVSUN      ............GCA....-A-CTTCT.........-------TT-----...-AG-GCA-AGAG-GGGA...CAAA-TTACA-T------C-TC-C--G--------AG-GC-AA-------G--GA----G-   1501
LHOEST.KE.LHOEST   ......A-G-AAC-A....---ACAG-G.........GG------T-----...-CC-GA-AAGAG-A-AT...CA-AG-TG-------ACAC-TT----G--C-----AA--C-AA-------G---A----G-   1523
LHOEST.CD.447      ..................-CTAAA--T.........GGC-----T-----...-AC-C--CAGAG-A-AT...CA-AG-TG------CACAC-AT-G-----------AA--C-AA-T---------A------   439
LHOEST.CD.485      ..................---AAA--T.........GGC-----T-----...-AC-CA-CAGAG-A-AT...CA-AG-TG------CAC-C-AT-G-----------AA--C-AA-T---------A------   439
LHOEST.CD.524      ..................-CA-AA--T.........A------TT-----...-T-GA-TAGAG-A-AT...CAAAG-TG---C---ACCC-TC-------C-----AA-TC-AA-------G---A------   433
MANDRIL.GA.MNDGB1  ......................--T.........A-G----T-------...--G--CAAAG-G-TGCAGCA--A--GTATCAGT--ACTC-A---AGT-----G-T-A--C-GA-----------AC---G-   894

COLOBUS.CM.CGU1    GAGGAGG-G-AAGCAAAAA---AAGCAGAG......A--TTGG--ATG--A-----AACAGGGCCT...CAAGGGCCA--G-----AC--C-GAGC--C--G--G---GGA--C-----T--GTGT----   1078
SYKES.KE.173       CAAAAGGCAGCA-CA...-C-CCACCTAGA......GG---C--T-----...-AC-GC-CAGA--TCCT--AAAT--GTG-A-T---ACA-GAG-CC--GT---G---C----AA-------GG-G-CT--GA   1347
RCM.NG.NGM         ......................TCT--TGGACAAAGAGG----------...--A-----AACA-TT-AT...CA---GCCT-A-------C-A-----T---------A--------T---------------   1550
MANDRIL.GA.14CG    ......................TCA.........A-------T-------...-----TCA-ATA-TA-AT...CAAAC-CCA--G-----A-GA--T-----A--------------C-----C---TGTA-T-   1147
MANDRIL.14CG Gag   _.__.__.__.__.__.__.__S__.__.__.__K__N__Y__.__P__I__Q__I__I__N__.__Q__T__P__V__H__Q__G__I__S__P__R__T__L__N__A__W__V__K__C__I__   p24
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2        AAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGA
H1B.FR.HXB2 Gag    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__M__Q__M__L__K__E    p24

CPZ.US.CPZUS       ----A-----C--T-A---G-----C--C--G-----CATG--C-----------G--G--T--C------G-C-----T-----C--TG-TA-A-----------G-G---A------G-T-----G--   1445
CPZ.GA.CPZGAB      -G--A-----A--T-----T--G--A-----A-----------TC-------G------TTA--T--G---G-T---------T----TG----A-------------G----------G--C----G--   1458
CPZ.CM.CAM3        ----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-----GG-C--G--G--    787
CPZ.CM.CAM5        ----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-----GG-T-----G--   1090
CPZ.CD.CPZANT      -------AAA-----AT--------C--C--T--------T--T-----------G--A--T--T--T-----------------T--TG----T----AC-------G---------GG--C-------    809

H1A.UG.U455        ----C--------------------A-------------------------G-----------------------T-TG-----G--TGT------------C--G-----T-----------------    852
H1C.ET.ETH2220     -G--A-----------------G--A------------A----T-------------------------------------------T-----------------------------------------    771
H1D.ZR.84ZR085     ----A----------A---------A-----------------------------------------------------------------------------------------------------    914
H1F.BE.VI850       -------------T-----------A----------------------------G-----T--CAC---------------------T-----------------------------------------    740
H1G.SE.SE6165      ----A-----C-----T--------A-----------------------------------------------------TCT----T----------------G-----------T---------C----G--    800
H1H.CF.90CF056     ----A--------T-----------A-----------------------------------------------C-----TG-T---------T-----------------------------G-----------    744
H1J.SE.SE92809     ----A--A-------------------------------------------------------G-----------T---------------A-A--------C--------T-----------------    720
H1K.CM.MP535C      -G----------------------A------------A---------------------T---------C-----------------------------------------------------------    596
H101_AE.TH.CM240   ----A----G---T-A---------A-----------C--------------G-----------------------T-TG--------T-T------------C--G-----A-----------------    962
H1N.CM.YBF30       -G--------C--T--T--------A-----A------ATG--C--G--------G--A--G--CTC----C----T--T---T----T-----A-----------G-----A-----G---C-G--G--    961
H1O.CM.MVP5180     ----------C--T-A---T---A-T--T--T------ATG--------------G--TGT---CT-T---A-C--T--------G--TG-CA-A--------C----GG--TT-A---G----G--G--   1418
H1O.BE.ANT70       ----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-A---G--C----G--   1442

MAC.US.SMM251      -G--A---AAA--TG-AG-------AG-G--AGGA---CAG---C-G--------CTG------CT-T--CA-T--TCAG---T----TTGT-----A-AC--------G--T-----G--TA-C-G---   1654
MAC.US.SMM239      -G--A---AAA--TG-AG-------AG-G--AGGA---CAG---C-G--------TTG------CT-T--CA-T--TCAG---T----TTGT-----A-AC--------G--T-----G--TA-C-G---   1666
MAC.US.SMM239 Gag  E__E__K__K__F__G__A__E__V__V__P__G__F__Q__A__L__S__E__G__G__T__P__Y__D__I__N__Q__M__L__N__G__V__G__D__H__Q__A__A__M__Q__I__I__R__D    p24

STM.US.STM         -G--A---AAA--TG-TG--------G-G--AGGA---CAG---C----------CTGT-----CT-T--CA-C--TCAAC--T----TTGT-------AG--------------------TA-------   1322
SMM.US.SMM9        -------AAAA--TG-GG----G--AG-G--AGGA---CA---GC----------CTG---T--CT-Y---A-C--TCAA--------TTGT--A--A-A------G--------------TA-T-G---   1148

H2A.GW.ALI         ----A---AAG---G-GG-------AG-G--AGGA---CAG---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AC--------------------AA-C-GG--   1709
H2A.SN.ST          -G--A---AAG---G-GG-------AG-G--AGGA---CAG---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AC-----------T--------AA-C-GG--   1161
H2A.DE.BEN         -G--A---AAG---G-GG-------AG-G--AGGA---CAG---C-C--------CTG---G--CT-T---A-T--TCAA-----T--TTGT-----C-AT--C--------T--------AA-C-G---   1716
H2B.GH.D205        ----A---AAG---G-GG-------AG----AGGA---CAG---C----------TG------TT-T---A----TCAG--------TTGT--A--A-A------G--------------TA-T-G---   1677
H2B.CI.EHO         --------AAG---G-GG-------AG----AGGA---CAG---C----------GTG------TT-T---A-T--TCAG---T----TTGT--A--A-A------G--------------TA-T-GG--   1678
H2G.CI.ABT96       ----A---AAG---G-GG--------G-G--AGGA---CAG--GC-G--------CTG-CTA---T-T---A-C--TCAA--------TTGT--A--A-A--------------------TA-T-G---   1079

TANTALUS.UG.TAN1   -G--A--AAGG---G-GG-----A-AG----------CCA---CC-C--------GTG-TTAAGTT-T--CA-T--TCAG---T----TGTCA-A--A-AC-------GG--A-----G--TA-C--G--   1533
GRIVET.ET.GRI677   -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-A-----TC-T--G--   1340
VERVET.KE.VER9063  -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-A-----AG-------   1573
VERVET.KE.VERTYO   -G-----AAAA--TG-AG-----A-AG----------CA---CC----------CTG---A--CT-T--CA-T--TCAG-----T--TGTGC-A--A-AT-------GG--AT-A-----AG-G-----   1063
VERVET.KE.VER155   -------AAAA--TG-GG-----A-AG----------CCA---CC-C--------CTG-------T-T--CA-C--TCAG-----T--TGTCT-A--A-AT-----G-GG---T-------AG-G-----   1562
VERVET.DE.VERAGM3  -G-----AAAA---G-AG-----A-AG-G--------CCAG--T-----------GTG---A--CT-T--CA-C--TCAA-----T--TGTCC----A-AC-------GG--GC-A-----AG-------   1074
SABAEUS.SN.SAB1C   -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-A-----AG-G-----   1663

SUN.GA.SIVSUN      ---------GA-GG-AA------ACAG-GG-A------G----T---ATGAC------A-TT---G----CA-C--TGTG-----G--TG-CA-A----AG-------G---T-----G---A-------   1631
LHOEST.KE.LHOEST   ----C-GA-G--GG-AA------ACAG--G-A-----------C---A-----AA---TTTG--TG-T--------TGTT---T----TG--A-T--A-AT-----G-GC-----------AA-------   1653
LHOEST.CD.447      ----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-----G--AA-T-----    569
LHOEST.CD.485      ----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-----G--AA-T-----    569
LHOEST.CD.524      ----T-G--GA-GG-AG------ACAG-GG-T-----------TC-GA-----AA----TTG---G----C-----TGT----T----TG-TA-A--A-AC-------GG--A-----G--AA-------    563
MANDRIL.GA.MNDGB1  ----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------G-C--------   1024

COLOBUS.CM.CGU1    -G...GGA-GAA-TGCT--TTCCT--GCT--------C-T---T-AT--TACG--G--A-TAG-TT-T---A-G---TTG-----C--T-TTT---ATACT-------G-TT-CCT--GG-C--G--G--   1205
SYKES.KE.173       -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---------GG-AA-T-----   1477
RCM.NG.NGM         -G--A---AAA--TG-AG-------AG-G--T--------G---C-G--------CTG--TA--CT-T---G----TCAG---T----TG-CA-A----A--------GT---T-A-----AG----G--   1680
MANDRIL.GA.14CG    -G--A---AAA-----T------A-AG----T-----CAT---CC----------TG-CT----T-T--CC-C--TGGT---T----TG-TA-T----A------G-G---TC-C-----AG-G--G--   1277
MANDRIL.14CG Gag   E__E__K__K__F__S__P__E__I__V__P__M__F__I__A__L__S__E__G__G__L__P__Y__D__L__N__G__M__L__N__A__I__G__E__H__Q__G__A__L__Q__I__V__K__D    p24
```

```
H1B.FR.HXB2       GACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAA  1527
H1B.FR.HXB2 Gag   __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__T__L__Q__E__Q_   p24

CPZ.US.CPZUS      AG----------------A-----T--A------ACA-----TC----------C-----A--------G--------------C--------------T--------C--C--C---T-G--A------  1575
CPZ.GA.CPZGAB     AGT---T-----------C--C--G------C-CT-A-----CACT-----T--T--G--A--C-----A---C-A--G--G--------T-----T--------G--C--C--C--A--G---------  1588
CPZ.CM.CAM3       AGT---------------A---------------ACA--C--T-----------T--AG-A--C-----A---C-A-----GG--C-A--------------T--G--------C--T--AGTA-----G  917
CPZ.CM.CAM5       AGT---------A-----A-----G--------GC-A--C--T--------------A--A--C-----A--AC-------GG----A--G-----------T-----------C--T--A--A------  1220
CPZ.CD.CPZANT     AGTA--------------A--T--G--------GT-A--C--CACT--------A--AG-ACAGG---A--AT-A--G-------CA--------T--------G--A--A--C--AG-G-----G--G  939

H1A.UG.U455       T----------------------G-----C--GT-A----------------------C-----------------------------------------------C---G----A------  982
H1C.ET.ETH2220    C--------------G--------------C--GT-A---------------------G------------A--------C--------------------------A--------------------  901
H1D.ZR.84ZR085    ----------------------------C--GC-A----------G---------G-----------------------------T--T-----------------------------  1044
H1F.BE.VI850      C----------------------------C---T-A--------------------GCCC----------------G-----T-----G-----T-----T--------------------------  870
H1G.SE.SE6165     T--T--------------------G-----C--GA--------CAA--G---------T--C-------------A--------------------T------------------T----G---------  930
H1H.CF.90CF056    T--A--------------------------C--G--A-----------------------C----------A--------------------C--T----------------------G---------  874
H1J.SE.SE92809    T--T--------------------------C--G--A--------A---------------------------G----------------------T----------------------C---------  850
H1K.CM.MP535C     T--T--------T----------------C--GT-A--C--------------------CC----------A----------G-----G--------------------C--C--------------  726
H101_AE.TH.CM240  A---------------C----------------G--A--C-----A---------------C----------------G-----------------------------------------A------  1092
H1N.CM.YBF30      AGT---------------A-----C--------GACA--------C-C--TG--A--AC-AC-C-----G--AC-------C--T--A--------T-----------A-----C-----GGCA-----G  1091
H1O.CM.MVP5180    AGTA--------------A---------------ACT------CCAGCAATG-----GT-AC-------G-----A--G-------CA-----------T--T-----A-----C--A-AG--A--G---  1548
H1O.BE.ANT70      AGTA--------------A-T---G---------ACT--C---CCA-CG-T------GT-GC-------G-----A--G-------CA-----------T--T--G--A-----C----AG--A--G---  1572

MAC.US.SMM251     T-TT--A--------G--------T-----CTTGCA---C---CAA-CA--T...--A...CA--A---A---C-T--G--G--GTCA---TCA--T--T--------A------T-AG-AG-T------  1778
MAC.US.SMM239     T-TT--A--C-----G--------T-----CTTGCA---C---CAA-CA--T...--A...CA--A---A--AC-T--G--G--GTCA---TCA--T--T--------A------T-AG-AG-T------  1790
MAC.US.SMM239 Gag __I__I__N__E__E__A__A__D__W__D__L__Q__H__P__Q__P__A__.__P__.__Q__Q__G__Q__L__R__E__P__S__G__S__D__I__A__G__T__T__S__S__V__D__E__Q_   p24

STM.US.STM        --T---A--------------T--T-----C-TGCA-------CAA-CAC-G--T--GC-GC--G----G--AC-T-----------C--GTCA--T--------G--------CT-A-CAG--------  1452
SMM.US.SMM9       A-TT--A-----A--G-----T--T------TT-CAA--C--RCA--CA...--T--A--AC--G----A--AC-T-----G-----R---TCC-----T--------------C--AS-AG-C------  1275

H2A.GW.ALI        --TT--------A-----A-----C------GTT-CA-----CA-A-CA...--C--CT-AC--G----G---C-T-----------A--GTC---------------A--A--C--AG-AG-A-----G  1836
H2A.SN.ST         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--C--AG-AG-A--G--G  1288
H2A.DE.BEN        --TT--T-----A-----A-----C------TCGCA---C---A-A-CA...--C--CT-AC--G----A---C-C-----C-----A--GTC---------------A--A--C--AG-AG-T-----G  1843
H2B.GH.D205       A-TA--------------A-----C-----CCAGCAA--C--GTCA-CA...--C--A--GC-GG----A--AC-T--G--C-----A--GTCA--T-----------C--C--C--AG-AG-------G  1804
H2B.CI.EHO        --TA--T-----------A-----C-----CCA-CA-------TC--CA...--C--A--GC--G----G--AC-C--G-------A--GTCA--------G-----C--C--C--AG-AG-A-----G  1805
H2G.CI.ABT96      A-TA--T--------G--A-----K------CTTCAA------CAA--A...-----GCCGC--G----A--AM-C--G-----G-CA---TCA--T--------G--C--------AA-AG-A------  1206

TANTALUS.UG.TAN1  AGTA--------T-----A--CC-G-------T-ACA--C--CCCA-CA-----A--GC-AC--G----G---C-A--G--C-----A--TTCA--------G-----A--C--CT-TG--GCT--G---  1663
GRIVET.ET.GRI677  AGT---T-----A-----A-----G-----C--GACA--CAG-CCA-CA--T--C--GT-AC--G----G---C-A-----C--G-CA--GTCA--T--------------C--CT-AA-----------  1470
VERVET.KE.VER9063 A-T---------------A--CC-G------GT-ACA--C---CCA-CA-----A---T-AC-------A---C-T-----C--T-----TTCA--------------C--------AG-G--A------  1703
VERVET.KE.VERTYO  --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--CT-AG-A--A-----G  1193
VERVET.KE.VER155  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--------AG-G--A--G--G  1692
VERVET.DE.VERAGM3 A-T---------------A--CC-G-------T--CT--C---CCA-CA-----A--AT-AC--G----A--AC-C-----C--T--A--CTC---------------C--C--C--AG-G--A-----G  1204
SABAEUS.SN.SAB1C  TGTT--------A-----G-----T-----CCTTAGA--C--GCC--CGCA-CA---CCCA--T-AG--AGT-C-A--G--C---CAA--GTCA--------G--C--------C--AA-A--A--G---  1793

SUN.GA.SIVSUN     TCAT--AGTG------AG------------C---A----C--TCA---GCAGCCTG-ACAGC--GG---AGGCC-A--GACT--T-AT--T-CA-----------GGTA--ATCA--AG-AG-A--G--G  1761
LHOEST.KE.LHOEST  TCA---AGTG-------G------G---------CA---C--GCAA--ACAGCCAG--CAAC--GG---GGGTT-A---ACT--T-AT-C--CA--TG-G-----GGTC--CTCA--AG-AG-A------  1783
LHOEST.CD.447     TCAT--TGTG--A----GG-----------C--GAAA-----GCAA--ACA-CCTG-ACAAC--GG---AGG-C-T---ACC--T-AT-C--CA--TG-C------GTC--CTCA--AG-AG----G--G  699
LHOEST.CD.485     TCAT--TGTG--A----GG-----------C--GACA-----GCAA--ACA-CCTG-ACAAC--GG---AGG-C-T---ACC--T-AT-C--CA--TG-T------GTC--CTCA--AG-AG----G--G  699
LHOEST.CD.524     TCAT--AGCG--A----GA--T------------CA------CCA---GCA-CCAG-ACAAC--GG---AGG-T-A---ACT----AT-C--CA--TG-G-----G-TC--ATCA--AG-AG-A-----G  693
MANDRIL.GA.MNDGB1 TGTA--T------C----A------------CT-ACA-----TCAA--ACA-CCAG-ACAAC--GG---AGGAT-A--GACC--TTCA--CTC---T---------------TC---AG-GG-A------  1154

COLOBUS.CM.CGU1   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--CT-AG-AGCA------  1332
SYKES.KE.173      C-T--------------GG--G-C-------CTT-AA-----CCAA-CGCA-CAA---GC-C---A--CAGGAT----G--T--CTCC-CCTCA--------------A--A--CT-AA--GCT--G---  1607
RCM.NG.NGM        AGTA--------T--G--A-----T-----CTTGAGA-----CAAT-CA-G-CCA...C-AC--G----T--AT-A--G-------CA-----C--T--T--------A--A--C-ATA--GCT------  1807
MANDRIL.GA.14CG   TGT---------------------C------CTTAGA--C--CCAAGTA-G-CCT...T-GC-T-A---GGT-C-----A-C----CA----------T--------A--A--C-G-A-AG-A-----G  1404
MANDRIL.14CG Gag  __V__I__N__E__E__A__A__D__W__D__L__R__H__P__Q__V__G__P__.__L__P__Q__G__V__L__R__N__P__T__G__S__D__I__A__G__T__T__S__S__I__E__E__Q_   p24
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
H1B.FR.HXB2        ATAGGATGGATGACAAAT.........AATCCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGAC   1648
H1B.FR.HXB2 Gag    _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__I__L__D__I__R__   p24

CPZ.US.CPZUS       G-G--C---------.........GCA--C-----C--T-----------C--A----G-------G-GG----A--GC-T------G------A------CT-T--CGTTG-T--CT-A--T--C-AG-   1696
CPZ.GA.CPZGAB      --T--G----CA---.........GCA-----T--C----------G--TG------G-------G----TT-A--G-----C---G-G--T---------T----AGTA--T--C--T-----C----   1709
CPZ.CM.CAM3        --T-CT---------.........GCA--------CG-T--------------A--C-GG------G-GG-T--A--GC-T--C---G-------G------T----AGTT--T--CT-A-----C-AG-   1038
CPZ.CM.CAM5        G-T-CT---------.........GCA--------C-----------G-----C----G-------G-GG-T--A--GC-C--C---G----G-A------CT----CGT---T-----A--T--C-A--   1341
CPZ.CD.CPZANT      --GCA-------T---CACCTCAACAG---GG-GGAG---------G--C--C-----G--------C---A-------------GG-G--C--G-N------T--AGT---------A--G----A--   1069

H1A.UG.U455        ---------------GGC.........-----------------G-----C--C----G-------------------------------------------GTT------T------------   1103
H1C.ET.ETH2220     ----C----------GGG..........--C------G-T-----G-----C--C-----------------------GC-----------------------------GT-------T---------A--   1022
H1D.ZR.84ZR085     ----------------GC.........-------------------------C-----------------------------------------------------GT-------T-------C----   1165
H1F.BE.VI850       ---CA---------GGGC.........--C------G-------G-----C--C--------------C-----A----------------------------T---GT-------T---------A--    991
H1G.SE.SE6165      ---ACT--------CGGC.........--C--------------G-------------------------------G-----------------------------GT-------T-A----------   1051
H1H.CF.90CF056     ----C----------GGC.........------G----------G-----C--C-------------------------------G-----------------T---GT----------------A--    995
H1J.SE.SE92809     ---------------GGC.........-----------------------G--------------------T------C---------------------------GT---T---T----T-------    971
H1K.CM.MP535C      ----C----------.........AGC--C------G-------G--G-----C----------------------T-----C----------------------GT-------T------------    847
H101_AE.TH.CM240   -----------------C.........--------C--------G-----C--C--------G-----------------------------------------GTT------T------------   1213
H1N.CM.YBF30       G-G-CT--------TGC-.........-----T---G-T-----------T-------G----------G-------G-----C-G---T--G-----------T---GT---------A--G--C-A--   1212
H1O.CM.MVP5180     --TAT-----CT--T-GAGGG......GC-AACT----------------C--C----G--A-------G-G--A---C----C-----G----A------C--T--AGTG-----CT-A--T--T--G-   1672
H1O.BE.ANT70       --TCAC----CT--C-GGCCC.......--C-A-----------------C--C----G--A-------G-GT-A---C----C-----G----A------C-----AGTG-----CT-A--T--T-AG-   1696

MAC.US.SMM251      --CCAG------TAC-GACAA......C-GAAC--C--A--------CA-C-----C-GG--------CCAA-----G--GC-A---TGT--C----------A---A--A-A------A--TG---A--   1902
MAC.US.SMM239      --CCAG------TAC-GACAA......C-GAAC--C--A--------CA-C-----C-GG--------CCAA-----G--GC-A---TGT--C----------A---A--A-A------A--TG---A--   1914
MAC.US.SMM239 Gag  _I__Q__W__M__Y__R__Q__.__.__Q__N__P__I__P__V__G__N__I__Y__R__R__W__I__Q__L__G__L__Q__K__C__V__R__M__Y__N__P__T__N__I__L__D__V__K__    p24

STM.US.STM         --TCA-------TACCGGCAG......C-AAAT-----A--G------A-C--C--C-GG--------TCAA--A--GC--C-A---TGT--C--------C-AT--AGTT-A----T-A---G---AG-   1576
SMM.US.SMM9        --TCA-------TAY-GGCAG......C-AAAC-----A--------CA-C--C--C-GR--G-----CCAAT-A--GC-GC-G---TGC-------------AT--A--A-A----T----TG-G-A--   1399

H2A.GW.ALI         --CCAG------TTC-GGCCA......CGGAAT---G-G--------GA-C--C----G---------CCAGA-A--GC--C-G--GTGT--C--G-----C-A---A----A---C--A-------A--   1960
H2A.SN.ST          --CCAG------TTT-GGCCA......C-AAAT---G-A---------AGC--C----G---------CCAGA-A--GC--C-G--GTGC--C--G-----C-A---A----A---C--A-------A--   1412
H2A.DE.BEN         --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---CT-A-------A--   1967
H2B.GH.D205        ---CAG------TAC-GGGCC......C-AAAT---G-------G---A-C-------G---------TCAAT-A-----GC-G---TGT--CC-------C-AT-------A---AT-A-------AG-   1928
H2B.CI.EHO         ---CAG------TAC-GACCC......C-AAAT---G-------G---A-C--C----G---G-----TCAGT-A--GC--C-G---TGT--CC-G-----C-AT-----T-AT--A----------AG-   1929
H2G.CI.ABT96       --TCAG----CACAC-GGCAG......C-AAAC-----A---------A-T--Y----G---------CCART-A---C-TC-A---TGT--C--G-----Y-AT--A--A-A---C--A--TG---A--   1330

TANTALUS.UG.TAN1   ----AG----CTTTT---GCT.....------AGAG-AGAT------CGT--C--C-G-G-----G-T---------C-CC-A---TGT----A------C-AT----T-TCAG-C--------C----   1787
GRIVET.ET.GRI677   ----AG----CCTTC---GCC.....------AGA--AGAC-----G-C-CAA--C-G--A----G-T--TT----C---C-A--GG----GCAG-----C-AT--CCAA-AGG-C--A-----TC---   1594
VERVET.KE.VER9063  T---A------TTAC-CAGCC.......--C--CAGAG-AGAT--T--G-CT--C----G---------T-----A--GC-GC-G--GTGT----A------C-A---AGT-TCAG-AT-A-----C----   1827
VERVET.KE.VERTYO   T---A------CTAT-C-GCT.......--C--C-GGG-AGAT-----T-CC--C--CCGG--------T--T--A---C-TC-A--GTGT--C-A------C-A---AGTATCAG-C--A-----T--G-   1317
VERVET.KE.VER155   C---AG-----CTAT-C-GCT.......--C---AGGG-AGAT-----G-CC--C---CG---------C-----A--G---C-A---TGT----A------C-AT--AGTGTCTG--T-A--T--C----   1816
VERVET.DE.VERAGM3  C-G-A------ATAC-CAGCC......-----CAGAG-AGAT--G--T-CC--C----G---G-----T--------G--GC-A---TGT----A------C-A---AGTGTCTG-CT-A----------   1328
SABAEUS.SN.SAB1C   ----A-----CC--C-GAGCA......C-GAATG--G--AAT------A-T---------G-------C--------TC-CC-G---TGT--C-A------C-A---AGTA-A----T-A-----T-A--   1917

SUN.GA.SIVSUN      ---CAG-----CTATGGCCTTAATGGAGCAA--AGAG-TG-----CA---C--C------G-C---G-C--AGA---CA-GG-A-G-G----T---C-CC--CAGA-----TCAG-AG-A--A--T----   1891
LHOEST.KE.LHOEST   C---CC----CT---GCAGAC.........A-T---G-AGAT------A-----------GA----G-C--T-A--CTA-GG-A---G----C-----CC--CAA--AGTA--TG-CA-------C-A--   1904
LHOEST.CD.447      C---CT----CC--TGCAGAC.........A-C---G-AGAT------A-G---------GAG---G----T-AA-CTA-GG-A---G----T-----CC--CAA--AGTT--TG-AA-A-----C-AG-    820
LHOEST.CD.485      C---CC----CC--TGCAGAC.........A-C--AG-GGAT------A-----------GA----G----T-AA-CTA-GG-A---G----T-----TC--CAA--AGTT--TG-AA-------C-A--    820
LHOEST.CD.524      CA--CC----CA---GCAGAC.........A----C--AGAT-----GA----C------GA----G-----A--CTA-GG-G---G-G--------AC--CAG--AGT---TG-AA---------A--    814
MANDRIL.GA.MNDGB1  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--AACA--TG------   1278

COLOBUS.CM.CGU1    G-G-CT---GGAGAG...................--C--A......-CCA-C-------GGG-C------G-A-A-A-TC-GG-A--GG--A-CCAG--TGCA--A--AT--TCTG-A--A--T--T----   1438
SYKES.KE.173       ----A------T--T-GGCAG.......--CAAC--AG-A-A------------A----G---------C--T--A--GC-GC-G--GTGT--GCAGG-C--C-A----GTTTCT--CT-A-------A--   1731
RCM.NG.NGM         ----CT----CC--T-GAGCC.......--CAAC--A--TG----G---A-C-------GG-AT-----TG-G--T--G---C-A---TGT--G-A--------A---AGTA-A---CT-A-----C-AG-   1931
MANDRIL.GA.14CG    ----A-----CA--T-GACAG......C-GGA--AGG-AAAT-----G-C---C-----GCAG------G----A--GC-GC-A---TGT--G--C-----C-A---AGTG-A--------------A--   1528
MANDRIL.14CG Gag   _I__E__W__G__T__R__Q__.__.__Q__E__Q__V__N__V__G__G__I__Y__K__Q__W__I__V__L__G__L__Q__K__C__V__S__M__Y__N__P__V__N__I__L__D__I__K__    p24
```

```
H1B.FR.HXB2        AAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCCAGATTG   1778
H1B.FR.HXB2 Gag    Q__G__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__N__A__N__P__D__C    p24

CPZ.US.CPZUS       -G-----T--A--G---------------------A-------C-----AT-------A--A--------T-CA--A-------CC-----------G---C-T-----A-----C--T--T--------   1826
CPZ.GA.CPZGAB      ----G--------G-----------T-----G---A-A-----C------T----G--T-----------------A--G--------------T--C--TC-C--A--G--G-----A-----------   1839
CPZ.CM.CAM3        -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-----T--T-----C--   1168
CPZ.CM.CAM5        -G-----T--A--G--A--C--G--T--------TA----T---------T-------A--A--G--AA-------A----------------T--C---C-T--A--T--------T--T-----C--   1471
CPZ.CD.CPZANT      ----------A--G-----C-----T-----G--TA-A-----------AA-T-----A--A--G---------CCT--G---GCC--------------------AA----------C--T--------   1199

H1A.UG.U455        ----G-----A--------C--G--T--------TA-A----T---------C-----T---------A-------T--------C------------------C----------------T-----C--   1233
H1C.ET.ETH2220     ----------------T------------------------T------CT-------T--A------A----A--T--------------------C------------------------------   1152
H1D.ZR.84ZR085     ----------------------------------------C-------------------------------A--A-----------------------------------A-----------   1295
H1F.BE.VI850       ----G-----A-----------------------A-A----T----GTC--------T-----------------C-----GGG-----------C--A----------------------------   1121
H1G.SE.SE6165      ----G-----A-----------------------TA------T----TG-T-G-----T-----------------T------GGC-----------C------------------------------   1181
H1H.CF.90CF056     ----G-----A-----------------------A----T-T-------T-------T--------CA-------T--G--G-----------------------------------A--T--------   1125
H1J.SE.SE92809     ----------A-----T-----------------A------T----G----------T---------A-------T-------------------T------C-------------A--T--------   1101
H1K.CM.MP535C      ----G-----A-----------------------TA------T------C--------T--A-----CA-------A-----G--------------C---C-------------C--A-----------   977
H101_AE.TH.CM240   ----G-----A--------C--------------TA----------------C-----G--A------A----------------C------------------C-A--------C-----T-----C--   1343
H1N.CM.YBF30       ----------A--------C---------------A-------C--------------A-----G--AA-------A-----G-----------------AC-C--A--A--------A-----------   1342
H1O.CM.MVP5180     -G--------A-----A--C-----T--------T-----T--C-----AT-------T---------A-T--A--A-----G---------------------C-T--T--G---T-A-----------   1802
H1O.BE.ANT70       -G--------------A-----------------T--------C-----AT-------T---------A-T--A--A----------------------TC--C-T--T--------C--T--------   1826

MAC.US.SMM251      ----G-----A--G--A---CAGAG----------A-------C----GCT-------A--A---A-AGATGCA-CA-----G-----------TC----AC--C--A-T--------T-----------   2032
MAC.US.SMM239      ----G-----A--G--A---CAGAG----------A-------C----G-T-------A--A--GA-AGATGCA-CA-----G-----------TC----AC--C--A-T--------T-----------   2044
MAC.US.SMM239 Gag  Q__G__P__K__E__P__F__Q__S__Y__V__D__R__F__Y__K__S__L__R__A__E__Q__T__D__G__G__V__K__N__W__M__T__Q__T__L__L__I__Q__N__A__N__P__D__C    p24

STM.US.STM         ----G-----A---A-G--CCA-AGT---------A-------C--G-G-T-G-----A--A-----AGAT-CATCA--G--------------TCGG--AC--C-AA-T--------T--T-----C--   1706
SMM.US.SMM9        ----------A--GT-A---CA-AG---------TA-A-----C--G-G---------A--A---A-AGAC-C--CAS-G--------------TC----AC--C--A-T--------C-----G--C--   1529

H2A.GW.ALI         -------------G--A--CCA-AG---------TA-A-----C----GCT----G--A--A---A-AGAT-CA-CA-----G-----------TC----AC--C----A--G-----C--------C--   2090
H2A.SN.ST          -G----------G--A--CCAGAGT--------TA-A-----C--G-GCT-G--G--A--A---A-AGAT-CA-CA----------------CC----AC--C-A--G--G-----C--------C--   1542
H2A.DE.BEN         -G--------A-----G--CCA-AG----------A-------C----GCT-G--G--A--A---A-AGAC-CA-CA----------------CC----GC--C-AA-A--G-----C--------C--   2097
H2B.GH.D205        -G----------G-----CCA-AG---------TA-A-----C----GCT--C-G--A--A---A-AGAC-CA-CA--G--------------C----AC--C--A-T--G-----T-----------   2058
H2B.CI.EHO         ----G-----A--------CCA-AG---------TA-A-----C----GCT--C-G--A------A-AGAC-CA-CA----------------C----AC--C--A-T--G-----T-----------   2059
H2G.CI.ABT96       ----G-----A---------CAGTCT--------TA-A-----C--G-G---R-----A--A---A--GAT-CA-CA----------------C-G--AC-AC--A-T--------C-----------   1460

TANTALUS.UG.TAN1   ----GG-C-----G--T----AG-----C------A-A--T---C--G-C--T-----A--A---A-AC-C--A--T----------------------AC--C-CA----G-----C-----------   1917
GRIVET.ET.GRI677   -G-----T--A--------CCAG------------A-A---------G-C--G-----A--A-----AC-------T--T--------------C----T---C-TA----G-----C--T--G-----   1724
VERVET.KE.VER9063  -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--------T-----------   1957
VERVET.KE.VERTYO   -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--------T--T--------   1447
VERVET.KE.VER155   ----G--C--A-----A--C-A---T---------A-A---------G-AA-------A--A---------GGA--A--C---C-A------------T-T---C-CA-T--G-----C-----------   1946
VERVET.DE.VERAGM3  -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-----T--T-----C--   1458
SABAEUS.SN.SAB1C   ----G--T--A-----T--C-AG--T--------TA-A-----C--GG----G-----A-----GA-AGAC-CT-CT-----G--C---------C-GT-TC-----A----G-----C-----------   2047

SUN.GA.SIVSUN      -G--------A--G--T--C---------AC---TA------TC--GG--T-------A--ATC--GC---G----A------G-G-------A---G-AAA--C-A--G--------A-----------   2021
LHOEST.KE.LHOEST   -------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-----A-----T--C--   2034
LHOEST.CD.447      -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--------A-----------   950
LHOEST.CD.485      -G--G-----A-----A--C-A---A----C---TA------T----G-AT-------A--AGG--GC-----T---------G-A-------A---G-AGA--C----A--------A-----------   950
LHOEST.CD.524      -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-----A-----T--C--   944
MANDRIL.GA.MNDGB1  -G------------G-T----A------------TA-A-----C--TGTAA-G-----A--A---------GGA--A-----G-TG------CAGC-GCATC--C-TA-AG-------A--------A--   1408

COLOBUS.CM.CGU1    -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----C-----T--C--   1568
SYKES.KE.173       -G--C-----A---ATT--C-A-------------A-A-----CC-TTGCT-------A-----G--GGAC-CATCA--C---GGG---C-A--TC----AC-----A-T--------C--------A--   1861
RCM.NG.NGM         -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--------T--T-----C--   2061
MANDRIL.GA.14CG    ----G-----A--G--T--C-AG--------G---A-A--T------G----GC-G--G----G-A-AGAT-CAC-A--C----CC---------C----A---C-CA----------C-----------   1658
MANDRIL.14CG Gag   Q__G__P__K__E__P__F__K__D__Y__V__D__R__F__Y__K__A__L__R__A__E__R__T__D__P__Q__V__K__T__W__M__T__Q__T__L__L__I__Q__N__A__N__P__D__C    p24
```

HIV-1/HIV-2/SIV
complete genomes

```
                                                                 see Kaye, J Virol 72(7):5877-5885(1998) Gag p2 role in RNA encapsidation
                                                                                                   Gag p24 \/ Gag p2
H1B.FR.HXB2 Gag   TAAGACTATTTTAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCA   1908
H1B.FR.HXB2       __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__Q__V__T__N__S_    p2

CPZ.US.CPZUS      ---A---------G---C--------T-GA--C--C-----------G------------A-----G----------CT-----------G--C--A--------C---T-----A-G-A---CC-C   1956
CPZ.GA.CPZGAB     ----CAA---C---------------T-G---C--C-----G------------T-----A-------------ATCT-----------G---C-A--------C---TCAATG--TCAG---CA-   1969
CPZ.CM.CAM3       ---A-AC--C----G---CC---------GA-----CT--------------------C-----------G-----A-C---------T-----AC----------T---------C--CA---CC-T   1298
CPZ.CM.CAM5       C--A-GC--C--G-G---C--A-------GA-----CT------------------C--C--------G--------A-C---------T------C----------C------------CA------C   1601
CPZ.CD.CPZANT     C--ACAC--CC-G--G--T------A---GA--CT-CT----------T-A-----T-----A-----------C--A-C---------------G-----A-----T---GCTTCT-CT-AT---G--   1329

H1A.UG.U455       ----T-C-------G------A--G----G-------T--------------------C--------G----------------A-----G-----------G--------T------CA-C-GA--   1363
H1C.ET.ETH2220    ---A--C-------G------A--G----G----T--T---------------------------G--------T-C---C--A--------G---------G----------------AC---A--   1282
H1D.ZR.84ZR085    ---------C-----------------G-------T-----------------------------G--G--G---A-------A-----------------G---------------C-----G-G-C   1425
H1F.BE.VI850      ------C---------------A---G--------------------------------------G--------TA-----------------------C--G--------T----C--ATTCAG-C   1251
H1G.SE.SE6165     ------C--C----G------A----A--GA---T------------------------------G---------A-------A-----------A-----G-----------G-CCT--GG-G--   1311
H1H.CF.90CF056    C--------A----G------A----A--G----T--A---------------------------G--------TA-T-----A-----------------G--------------------A--   1255
H1J.SE.SE92809    C-----C--------------A---T---GA----------------------------------G--------T--T------G---------------------------------G--C---A-C   1231
H1K.CM.MP535C     ------C-----------G----------G----T--T-------G-------------------G-----G--TA-------A------A----------G------------G---------C--   1107
H101_AE.TH.CM240  ----T-C--------------A---A---GA------T--------------------C--------G--------TA-------A-----G------------------------C-C-CA-C--G--   1473
H1N.CM.YBF30      ---ACAGC-CC----------A--G----GA-----CT-------G--------G--C--C--------G--G-----A-CA-------------GC-A--A--G--T---TCA--G--GCAGC-GC--   1472
H1O.CM.MVP5180    ---ACAA---C-G--------A-------AA-----TT--------------GT---C-----A-----------G--AACT--C-------A-A-AC-A--A---------GCTTCT-CCCAGC-AGAT   1932
H1O.BE.ANT70      C--ACAG-----G---T----A--G----GA-----CT-------G------GT---C---------------------AACT--------C-----GC-A--A---------GCTAC--CCCAGC-AGAT   1956

MAC.US.SMM251     C---CTAG-GC-G--G-GGC----TGTGAATC-C--C-----------C----G--T-----A--------G-----A--A--G-----T---T-AA----A-----CC---AAG-G-CCCTCGCAC--   2162
MAC.US.SMM239     C---CTAG-GC-G--G-GGC----TGTGAATC-C--C-----------C----G--T-----A--------G--G--G--A--G-----T---T-AA----A-----CC---AAG-G-CCCTCGCAC--   2174
MAC.US.SMM239 Gag __K__L__V__L__K__G__L__G__V__N__P__T__L__E__E__M__L__T__A__C__Q__G__V__G__G__P__G__Q__K__A__R__L__M__A__E__A__L__K__E__A__L__G__P_    p2

STM.US.STM        C--ATTGG-A--G----GTC----CATGAATC-C--C--------G---T-A-----------A--G--T--G-----A--A--G--A--T---C-GA----A-----CT---AAG-G--CTTCC-AC--   1836
SMM.US.SMM9       ---ATTGG-GC-C--G-GTC----YAT-AATC-C--TT----------T-------C---------------R--G--A--A--A--AS-T---C-AA----A--G---T---AAG-T-C-CTG-CGCA-   1659

H2A.GW.ALI        C--ATTAG-GC-G----G---A--GATGAATC----CT-------G---C-A--C--C--------------G-----A-----G--A--C---T-AA----A-----CT-A-AGG-G-CC-TG-CAC--   2220
H2A.SN.ST         ----TTAG-AC------G-C-A--GAT-AATC----CT-----------C-A--C--C--------G-----T-----A-----G--A--C---T-AA----A-----CT-A-AGG-G-CC-TGGCAC--   1672
H2A.DE.BEN        C---TTAG-AC------G-C----GATGAATC-C--C--------G---C-A--C--C--C-----G-----C-----A-----G--A--C--GC-AA-----------CC-A-AAG-G-CT-TGGGAC--   2227
H2B.GH.D205       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-AAG-G-CCCT--CAC-T   2188
H2B.CI.EHO        C--ATTAG-GC-T--G-GC------ATGAACC-C---T-----------C-A-----C--C-----GA-------C--A--G--G--------GC-AA-------G--TT-A-AAG---CCTTG-CAC-T   2189
H2G.CI.ABT96      ---ATTAG-AC----G-GG------ATGAATC----TT----G------T-------C-----A--GR-----------A--G--A--A--C--GC--A-----------T-A-AAG---CCTT---CC-T   1590

TANTALUS.UG.TAN1  C---TTGG----G----G-C-T---AT-CATC-A--T--------G---C-------T---------------------A--A-----A----AGC-CA---TA---------CAA---A-GCA-GGAGTT   2047
GRIVET.ET.GRI677  ---ATTG---C-G----G-------ATGAATC-A--CT-G--G------C-A-T---T--C--------------G--ACAA--------T-AGC-AA---TA---ATG-----TA-T-G-CAG......   1848
VERVET.KE.VER9063 C--AGTC--CC-G----G---A---ATGCACC----CT-----------C----T--C-----A--------G--C--AA--T-C--A---------A----A---ATG---CAAA-CA-GCAG-C-CA-   2087
VERVET.KE.VERTYO  ----GTC--CC-G--G-GCC-A---ATGCACC-C--C--T---------T-A--G--T--------G--------C--AA--T-C--A----A---AA----A---ATG---CAGACCA-GCA----CA-   1577
VERVET.KE.VER155  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---CAGA-CC-GCAG-G-CAG   2076
VERVET.DE.VERAGM3 ---AGTC--CC----G-GCC-----ATGCATC-C--T------------T-A--T--C-----A-----G--------AA-TT-C--A----A---GA----A---ATG---CAAA-TA-GCA--GCCAG   1588
SABAEUS.SN.SAB1C  ------AG-GC-C----GG--A--CATGAATC-A--C--------G---C-T--T-----C--A--CA-T------G-ACAA---------C-TC-AA----G-----T----CAGCG-CCTTTC-GCA-   2177

SUN.GA.SIVSUN     ---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-TCT-CTTTT-GGCAG   2151
LHOEST.KE.LHOEST  C-G-TTAG--A--------C-A---GA--G---AT-T--------G-------A---T--------------------T-CT-----A-G--A-A-A-----A---------GCTTC--CC-T-C-GCA-   2164
LHOEST.CD.447     --G--TGG-GA-------C--A---GA--G------TT-------G-------GG--C-----A--------------A-C------A-G---GA-------A---------GCAAC--CC-TGC-AAGT   1080
LHOEST.CD.485     --G--TGG-AA-------C--A---GA--G-----GTT-------C-------AG--C--------------------A-C------A-G--AGA-------A--G------GCAAC--CC-TGC-AAGT   1080
LHOEST.CD.524     C--A-TGG-AA----G--C--A---GA--GT--C--TT---------------A------------------------T-C------AAG--A-A-C--A--A--G------GCAAC--CC-TGC-GGGG   1074
MANDRIL.GA.MNDGB1 C---CAG-----G-G-AGC--A--GAA--GA--A--TT----G------T--GA------------------T--G--ACAA-----A--C---T-AA----A---ATG-----AAC---GGT-GGACA-   1538

COLOBUS.CM.CGU1   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---CAG--GT-CCAGC-AGAG   1695
SYKES.KE.173      C-GACAG--CC-C--G---A---TCAA-...C----CT-G--G------C--CA---T-----A-----C----------TTA--C-------A-C-CA----A--G--C---GTAATG-CCCAGC-G---   1988
RCM.NG.NGM        ------AG-AC-T--G-GTC-----ATGAACC----C--G--G--G---C-A--------C--A--G-C------T--ACAG--C------T--G--GC-A------------CAGATG--GCA--G-AAT   2191
MANDRIL.GA.14CG   -----GC--C--G----G------CATGAACC-CT-TT-----------T-ACT---------A-----G--G-----AAAAT--------CA-A-GA----A--G--C---CAGA-GA-GCAGTCAGA-   1788
MANDRIL.14CG Gag  __K__S__I__L__K__G__L__G__M__N__P__S__L__E__E__M__L__L__A__C__Q__G__V__G__G__P__K__Y__K__A__Q__M__M__A__E__A__M__Q__K__M__Q__S__E_    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
H1B.FR.HXB2        ......GCTACC..............ATAATGATGCAGAGAGGCAATTTT...AGGAACCAA.....................................AGAAAG...ATTGTTAAGTGTTTCAATT   1972
H1B.FR.HXB2 Gag    _.__.__A__T__.__.__.__.__.__I__M__M__Q__R__G__N__F__.__R__N__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__K__.__I__V__K__C__F__N__    p7

CPZ.US.CPZUS       ......T-A...............AGCG--T-TC-A----A---A--CGC-...G----G-C-.....................................G-G-GA...-AAA-C--A-----T----   2020
CPZ.GA.CPZGAB      GGGAGA--A...............GATG-TT-CT-C----AG--AC-AGG-...GC-GG--C-.....................................-A--GA...-AAA-A--A--C--T----   2039
CPZ.CM.CAM3        ......A-A...............GGTG-GT-CC----A--G--A---AA-...G----A-C-.....................................-CT-G-...-AAA----A--C--T--C-   1362
CPZ.CM.CAM5        ......T-A...............AATG-GT-T-----------A---AA-...G---GAGT-.....................................-AT------.-AAA----A--C-AT----   1665
CPZ.CD.CPZANT      CAGGGAA-C...............GCAG-CT-TC--------------GGA...-AT-GAGG-GGA.....................................-A--GA...CC-C-C--A-----T--C-   1402

H1A.UG.U455        ......AGC...............-----------------------...---GG--CG.......................................-GA...-GAA------------C-   1421
H1C.ET.ETH2220     ......A-C...............-------------A-A--------...-A-GG--CT.......................................-A--GA...GCAA----A-------C-   1343
H1D.ZR.84ZR085     ......T--G-TGCC.........-----------A-A-A--------...-A-GG-AC-.......................................-----A...----------------C-   1492
H1F.BE.VI850       ........................-------------A-A-T------...-A-GG----.......................................----GA...G-------A-----T----   1309
H1G.SE.SE6165      .......--AG-AGCC........---------------A--------...-A-GG--C-.......................................----GA...-C-A-C--A----------   1378
H1H.CF.90CF056     ......AA---AGCC.........-------------A------C---...-A-GG----.......................................-----A...T-------A--C-----C-   1322
H1J.SE.SE92809     ......AAC...............-----------A-----T--C---.....-G----T.......................................-A--GA...-------------------   1292
H1K.CM.MP535C      ......-T................G------------A------C---...-A-GG---T.......................................-----A...-----------C-----C-   1168
H101_AE.TH.CM240   ......A--................-----------------------C...-A-GG---G.......................................--A...-GAA-------C-----C-   1531
H1N.CM.YBF30       ......A-A--TAGT.........G-CT-TGCA--A--G--A--C---...-AAGG-AT-.......................................--G--A...CCCA----A----------   1539
H1O.CM.MVP5180     ......TTA-AAGGAGGATACACAGCAG--T-C-----A-----GC-GAA-...CCA--TAG-.....................................-A-GG-...CCCA-A--A--C-------   2011
H1O.BE.ANT70       ......TTG-AAGGAGGATACACAGCAG--T-C-----A-----GC-AAA-...CCA-TTAGG.....................................-A-GGA...-C-A-A--A-------C-   2035

MAC.US.SMM251      ......-TGC-AATC......CCTTTTGC-GCAGCC----AGA-GGGACCA.....................................--------...CCAA---------GG----   2226
MAC.US.SMM239      ......-TGC-AATC......CCTTTTGC-GCAGCC--ACAGA-GGGACCA.....................................--------...CCAA---------GG----   2238
MAC.US.SMM239 Gag  _.__.__V__P__I__.__.__P__F__A__A__A__Q__Q__R__G__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__K__.__P__I__K__C__W__N__    p7

STM.US.STM         ......-ACC-ACTG......CCCTTCGC-GCAGCC--ACAGCAGGGA.......................................C---G-...-CA--A--A--C-GG----   1897
SMM.US.SMM9        ......-GGC-ACTC......CCATTTGCRGCAG-C--ACAGAAAGGACAA.....................................R--------..--AA--W------RG--C-   1723

H2A.GW.ALI         .......---C-TATC.....CCATTTGCGGCAGCC--ACA-.............................................----G-...-CAA-------C-GG----   2275
H2A.SN.ST          .......--CC--ATC.....CCATTTGC-GCAGCC--ACAG.............................................----G-...-CAA-------C-GG----   1727
H2A.DE.BEN         ......AGCC-TATC......CCATTTGC-GCAGCC--ACA-.............................................--------...GCAA---G--A--GG--C-   2282
H2B.GH.D205        .......--AC--ATA.....CCGTTTGCTGCCG-T--ACA-AAAGCAGGG...-A-..............................---GG-...-CA--G-CA--C-GG--C-   2255
H2B.CI.EHO         ......T-C---AAT......CCGTTTGCCGCCGCT--ACC-A-AGCAGGG....................................-AG-G-...-CA--G-CA--C-GG--C-   2253
H2G.CI.ABT96       ......A-AG--CTA......CCATTTGC-GCTGCT--ACA-AAA-CAGGA...G-A..............................-A--G-AGC-CAA-A--A----GG----   1660

TANTALUS.UG.TAN1   ............................-AC---G----AG---CACCGCGA...G-AGGTAGGGGCAGAGGC.............................---GGC...CCACCCCGC--C-----A-   2114
GRIVET.ET.GRI677   ............................-AT---G-C--AGTG--ACC-CAG...-AA--GGGCCCC...................................C--GG-...CCGC-A--A--C--T----   1909
VERVET.KE.VER9063  ............................-AC---G-----CA---GGCAGCA...G-AGGAGT-AGGAGACAG.............................---CCC...CCAC-A--A----A-----   2154
VERVET.KE.VERTYO   ............................-AC---G-----CAG--AGG-CCA...-AA-GA---.....................................---CCC...CCAC-A-GA----AT----   1635
VERVET.KE.VER155   ............................-AC---G-A---CAG--AGG-GGA....---GGAAG-CCA.................................---CCC...CCGCCA-------A---C-   2137
VERVET.DE.VERAGM3  ............................-AC--------ACAG---GG-CAG....--AGGAAG-CCA.................................---CCC...CCA--A-------A-----   1649
SABAEUS.SN.SAB1C   ......CAA--AGTT......GGGAAC--CT-TG-----CA---GGCAAGA...CCC-GGGGTCCCTTAGGGGGAAGAGGTAGACCTCTG............ACCCA...-A-A----A----A-----   2277

SUN.GA.SIVSUN      ......CAAGTAGGA......AGACAAGCT---G-----CA-AA-TTACCA...CCC-GAA-CAGCCAGGGGAGGTTTGTAAGAATAGGAGGAGGAGGACCT--G-----.CCCA-G-CT-----T----   2263
LHOEST.KE.LHOEST   ......CAAGTAAGA.........CAA-AC---G-------...-T--C-CCG...CTT-GAA-TGCGCAAGGCAGGTTTGTGAGAACAGGAGGAGGAGGTCCC--G--A...CCCC-C-CT--C-------   2270
LHOEST.CD.447      ......CAA-TGAGA.........CAG-AT---G-A--A...-TG-CACCA...CCA-GGA-TGCACAGGGAAGGTTTGTAAGAACAGGAGGAGGAGGCCCT--------..CC-T-A-CT--C--T----   1186
LHOEST.CD.485      ......CAA-TGAGA.........CAG-AT---G-A--A...-TG-CACCA...CCA-GGA-TGAGCAGGGAAGGTTTGTAAGAACAGGAGGAGGAGGCCCT--------..CC-T-A-CT--C--T----   1186
LHOEST.CD.524      ......CAG-TGAAA.........CATC-C---G-------...CAA-CACCA...CCT-GAA-TGCACAAGGGAGATTTGTAAGAACGGGGGGAGGTGGCCCA----GA...CC-C-C-CT----------   1180
MANDRIL.GA.MNDGB1  ......T-A...............CAA-ATT-TG-----CAGA-AGGGCC-...CAA-GAGG-CCAGTTAGACAACCTACTGGA.....................-G-...-AACC--TC--C-----C-   1620

COLOBUS.CM.CGU1    ....................AGAACC-AC-----TG-AGT-AAG-CAGCA.....................................................--A--C-------   1738
SYKES.KE.173       .......-TA..................-AT---G-A--AG-GCCATCAAAA...G--.........................................---TCA...--GA----A--C-A-----   2043
RCM.NG.NGM         ............................--C---GCC--ACAG--AGGACC-....-A--AGG-CCCCCTAAACTAGGGGGAGGA..................CCT-G-...T-CC-C--A--C-AT----   2270
MANDRIL.GA.14CG    ......-TC-TGATGCAGAACTCGGGGGG-CCACC--G-G-TCCTCCGAGACAACCCCCT.....................................-GA...-ACCCC-GA---CC---C-   1864
MANDRIL.14CG Gag   _.__.__V__M__M__Q__N__S__G__G__P__P__R__G__P__P__R__Q__P__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__.__N__P__R__C__P__N__    p7
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2        GTGGCAAAGAAGGGCACACAGCCAGAAATTGC...AGGGCCCCTAGGAAAAAG.....................................................GGCTGTTGGAAATGTGG   2039
H1B.FR.HXB2 Gag    C__G__K__E__G__H__T__A__R__N__C__.__R__A__P__R__K__K__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G__C__W__K__C__G    p7

CPZ.US.CPZUS       ----------G--A--TCT-------------...-----T--A--A-G---A.....................................................--G--C-----G-----   2087
CPZ.GA.CPZGAB      ----G--------T---CT---T--------T...-AA--A--A--A-G---A.....................................................---------CGG-----   2106
CPZ.CM.CAM3        -------G-----A--TTT---------C---...--A-----C--A-GG---.....................................................--A--C-----G-----   1429
CPZ.CM.CAM5        ----------------TGT---T-----C---...--A-----A--A-G----.....................................................--A--C--------C--   1732
CPZ.CD.CPZANT      -C--T-----G--C--T--T--A---------...-A------A--A-GG--A.....................................................-----C----G------   1469

H1A.UG.U455        -------------A---CT-----A------T...-----------------.....................................................--------------C--   1488
H1C.ET.ETH2220     -------G---------CT-------------...-----------------A.....................................................----------------   1410
H1D.ZR.84ZR085     -----------------T---A--------T...-----------------.....................................................----------------   1559
H1F.BE.VI850       -------------A----T-------------...-----------A------.....................................................----------------   1376
H1G.SE.SE6165      -------G-----A--TCT-------------...-----------------.....................................................----------------   1445
H1H.CF.90CF056     ----------G--A----T-------------...-----------------.....................................................----------------   1389
H1J.SE.SE92809     --------C----A----T--A-A---C---...-----------A------.....................................................----------------   1359
H1K.CM.MP535C      -----------------T-------------...-----------A------.....................................................----------------   1235
H101_AE.TH.CM240   ----T-G------A---CT-------------...-----------A---C--.....................................................--T----------C--   1598
H1N.CM.YBF30       ----------G--C--TTTG--A-----C--T...-A---------A-G-GGA.....................................................----------G-----   1606
H1O.CM.MVP5180     ----A-----G--A--T-T---A-A---C--T...C-A--A-----A----G-.....................................................--T--C-----------   2078
H1O.BE.ANT70       ----A-----G--A--T-T---A--------T...C-A--A-----------.....................................................--T--C-----------   2102

MAC.US.SMM251      ----G--G--G--A---T-T--A--GC-A---...--A-----A--A-G-C--.....................................................--A--C-----------   2293
MAC.US.SMM239      ----G-----G--A---T-T--A--GC-A---...--A-----A--A-G-C--.....................................................--A--C-----------   2305
MAC.US.SMM239 Gag  C__G__K__E__G__H__S__A__R__Q__C__.__R__A__P__R__R__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G__C__W__K__C__G    p7

STM.US.STM         ----G----C---------T--A-A-C-G--T...AA-G---G--A-G-C-A.....................................................--T--------------   1964
SMM.US.SMM9        --R-A--G-----A---T-T--A---C-A-T-...--A--------A-G-C-R.....................................................-----C-----G-----   1790

H2A.GW.ALI         ----A--G---------T-G--A---C-A---...C-A--A--C--A---C--.....................................................-----C-----------   2342
H2A.SN.ST          -C--A--G---------T-G--A---C-A---...C-A--A-----A-G-C-A.....................................................-----C-----------   1794
H2A.DE.BEN         ----A--G--G--A---T-G--A---C-G---...C-A--A-----A-G-C--.....................................................-----C-----G-----   2349
H2B.GH.D205        --------C-G--A-----------GC-A---...-----------A-G-C--.....................................................--A--C-----------   2322
H2B.CI.EHO         -C-----G-CG--A--T--------GC-G---...-A---------A-GGC--.....................................................--A--C-----------   2320
H2G.CI.ABT96       -------G--G--A--T----TT---C-G---...--A-----C--A-G-C--.....................................................--G--C-----G-----   1727

TANTALUS.UG.TAN1   ----AC--ATT-----T-T-CAA-A-G----T...CCAAGAG-AG-ACC---C.....................................AAA--CCTA--------   2181
GRIVET.ET.GRI677   --------TTT--A--T-TGCAA--GG-A---...-A---A--A--AC-G-TC.....................................AAA--C-TT--G--C--   1976
VERVET.KE.VER9063  -C--A---TTC--A----TGCAA--GC-A---...-CA-AA--A--A--G-TA.....................................AAA--CCTT--------   2221
VERVET.KE.VERTYO   ----A---TTT--C--T-TGCAA---C-A--T...CC--AA--A-------CA.....................................AAA---CTA--G-----   1702
VERVET.KE.VER155   ----A---TTT--C----TGCAG--GC-G--T...CCT-AG--A--A----TA.....................................AAA---CTT--------   2204
VERVET.DE.VERAGM3  ----A---TTT--C--T-TGCAA---C-A---...CCT-AA--A--A--G-T-.....................................A-A--C-T---------   1716
SABAEUS.SN.SAB1C   ----A--GCCT--T---CTG--T---TT---T...-A-----------G-C-A.....................................-----C-----------   2344

SUN.GA.SIVSUN      ----T---CC---A--TCT---A-ATC-G---...--A-AG--A-A----GGA.....................CCACCCGGCTC---C-----------   2339
LHOEST.KE.LHOEST   ----A--GCC---T--------A----TG---...--ACAG--C--AC-GG-A.....................................--T-------C-----   2337
LHOEST.CD.447      ----A--GCC---T--T-----A----TG--T...--ACAG-----AC-GG--.....................................--A--C-----C-----   1253
LHOEST.CD.485      ----A--GCC---T--T-----A----TG--T...--ACAG-----AC-GG--.....................................--A--C-----C-----   1253
LHOEST.CD.524      ----G--GCC---A--TT----A-A--TG---...---CAG-----AC-GG-A.....................................----------C-----   1247
MANDRIL.GA.MNDGB1  --AAT-----------TGT---A--GTTC-T-...-A---------A-GG--A.....................................-G--C-----T-----   1687

COLOBUS.CM.CGU1    --CAGGG-AT-------TT--------TG--TCCC-AAAGA--A-TAGG-GGAGCCGGTCGAGGAAGAGGCCGAGGACGGGGAGGCTTTAGAGGAGCTCCCAGAAGGCCAGTGA-G----TT-CC--CAA   1868
SYKES.KE.173       ----GC--ATT-------TGCAA-A-G-C---...-AAAAG--ACTT---GCT.....................................AAA--C-TT--T-----   2110
RCM.NG.NGM         --------AC---T--T--T--T---T-----...-AAA-AAG---A--G--A.....................................--G--C----G------   2337
MANDRIL.GA.14CG    ----A-GGT-T--A--TGTTCTG---G-C--T...--ACT---C--A--G---.....................................--A--C-TT--G-----   1931
MANDRIL.14CG Gag   C__G__R__Y__G__H__V__L__R__D__C__.__R__L__P__R__K__K__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__G__C__F__K__C__G    p7
```

```
                                                                  stem-loop after TTTTTT slip potentiates slippage
                                                                          Gag-Pol –1 ribosomal slip site
                                                                            Gag p7 end \/ Gag p1 start
H1B.FR.HXB2       AAAGGAAGGACACCAAATGAAAGATTGT......ACT....................GAGAGACAGGCTAATTTTTTAGGGAAGAT...CTGGCCTTC...........CTACAAGGGAAG...GCC   2123
H1B.FR.HXB2  Gag  __K__E__G__H__Q__M__K__D__C__.__.__T__.__.__.__.__.__.__.__E__R__Q__A__N__F__L__G__K__I__.__W__P__S__.__.__.__.__Y__K__G__R__.__P_    p6
H1B.FR.HXB2  Pol                                                                                _F__R__E__D__.__L__A__F__.__.__.__.__L__Q__G__K__.__A_ Pol TF

CPZ.US.CPZUS      GC----------T-----------C--C.......--AGCAGGA..............A-C-----------------------ACA...T---T-CC-ATCGTGGAGCGGGGGATCAAAG-...A--   2189
CPZ.GA.CPZGAB     GC-A--G-----------------C--C.......--A......................-GA-------TG--------------AGG...T--------CCGCAG...CGGGAG......GCC.....   2190
CPZ.CM.CAM3       GC-A--G--G--T---------A-C--C......C-AGCTGGA................--A----------------------AGG...T---T--C-GTGGAG...CGGGGGATCAAAG-...A--   1528
CPZ.CM.CAM5       -C-------G--TA-C--------A---.......--AACTGGA...............-GC-----------------------GA...T---T--C-GTGGAG...CGGGAGATCAAA-...A--   1831
CPZ.CD.CPZANT     -C-------------GC-T---A-C---......C-AGCAACA...........AATACAG--A-A-TA--------------ACC...GACC--CA-GTGGTG...GGGG-G--G...-CC.....   1565

H1A.UG.U455       G--A--------------------C--C.......---.....................------------------------A--...T-------------------A-------G-------   1572
H1C.ET.ETH2220    ------------------------C---.......--C.....................------------------------GAC-...T-------------------A---------------   1494
H1D.ZR.84ZR085    --------------------------C.......---.....................--A----------------------A--...T-------------------C---------------   1643
H1F.BE.VI850      --GA--------------------C--C.......---.....................--A----------------------A--...T-------------------A-------G-------   1460
H1G.SE.SE6165     ------------T-----------C--C.......--G.....................------------------------A--...T-------------------A-------G-------   1529
H1H.CF.90CF056    --GA--------T--G--------C--C.......--A.....................------------------------A--...T-------------------AG---A--G-------   1473
H1J.SE.SE92809    ------------------------C--C.......---.....................------------------------T--...T-------------------AG---A--G-------   1443
H1K.CM.MP535C     G-----------T--G--------C--C.......---.....................------------------------A--...--------------------C-------G-------   1319
H101_AE.TH.CM240  G-----------T-----------C--C.......---.....................------------------------A--...T-------------------A---------------   1682
H1N.CM.YBF30      GC-A--------T---------------......-AAAATGAA...............-GA------------------------G..----T--C-............--T---A--G----.A--   1696
H1O.CM.MVP5180    -C-------T----------------C......-AAAAT..................-GA-------A---------------TA..--------C-............GGGGGGCACG-------   2165
H1O.BE.ANT70      -C-------T----------------C......-GAAAT..................-GA-A------A-----------C--ATA..--------C-............GGGGGGCACG-------   2189

MAC.US.SMM251     ---AATG-AC--TGTT---GCCA-A--C.....C-A......................--C--------GGG---------CCTTGGTCCA---GG..............AA-G---CCCC-...CAA   2377
MAC.US.SMM239     ---AATG-AC--TGTT---GCCA-A--C.....C-A......................--C--------GGG---------CCTTGGTCCA---GG..............AA-G---CCCC-...CAA   2389
MAC.US.SMM239 Gag __K__M__D__H__V__M__G__K__C__.__.__P__.__.__.__.__.__.__.__D__R__Q__A__G__F__L__G__L__G__P__W__G__.__.__.__.__.__K__K__P__R__.__N    p6
MAC.US.SMM239 Pol                                                                               _F__R__P__W__S__M__G__.__.__.__.__.__K__E__A__P__.__Q_ Pol TF

STM.US.STM        -C-AC-G--C--T--G---GCCA-A--C.....C-A......................--A-------TGGG---------CTTTGGCCCA---GG..............AA-G---CCCCA...CAA   2048
SMM.US.SMM9       ---A-C---C--TGTT---GCTA-A--C.....C-C......................--A--------GGG----------CTTGGCCCA---GG..............AA-G---CCCC-...CAA   1874

H2A.GW.ALI        C---CC------T-TT---GC-A-C--C.....C-A......................--A---------GG----------CTTGGCCCA---GG..............AA-G---CCTC-...CAA   2426
H2A.SN.ST         C----C-------ATC---GC-A-A--C.....C-A......................--A--------GGG----------TT-GGCCCA---GG..............AA-G---CCCC-...CAA   1878
H2A.DE.BEN        C---CC-------ATC---GC-A-C--C.....C-G......................--A--------AGG----------TT-GGCCCAC--GG..............AA-G---CCTC-...CAA   2433
H2B.GH.D205       ---AAC-------ATC---TC-A-A--C.....C-A......................--A--------GGG----------TTAGGACC----GG..............AA-G---CCTC-...CAA   2406
H2B.CI.EHO        -C--C--------ATC---TC-A-A--C.....C-A......................--A--------GGG----------TTCGGCCC----GG..............AA-G---CCTC-...CAA   2404
H2G.CI.ABT96      G---CC---C---ATT---GCCA-A---.....C-A.......................----------AGG----------TTTGGACCT---GG..............AA-G---CCTC-...CAA   1811

TANTALUS.UG.TAN1  G---CC-------TT-GCA--------C.....CGCAGT.....................G-------C-------------GA--GCCAACCTG..............GGGA-CAAAGCC....-G   2265
GRIVET.ET.GRI677  C--AATT--C--TATGGCA-----C--C.....-AGAAT.....................G-------A------------T-TGGCCAT---GG..............AGGAGC-AA-CC...AAG   2060
VERVET.KE.VER9063 G---CT--------TGGCA-----C--C.....-GG.......................G------TG------------T-TGGACGG---A-..............GGGG-CAAA-CC...-AG   2302
VERVET.KE.VERTYO  ---ATTG-------T-GCA-----C--C.....-GG.......................G------TG------------T-TGGACGG---AT.............GGGGGCAAA-CC...-AG   1783
VERVET.KE.VER155  ----CC---G---TT-GCA-----C--C.....-GG.......................G------TG------------T-TGGCCGG---AT.............GGGG-CAAA-CC...AAG   2285
VERVET.DE.VERAGM3 G--ACC---G--TTT-GCA--------C.....-GA.......................G------TA------------T-TGGCCGG---AT.............GGGAGC-AA-CC...CAG   1797
SABAEUS.SN.SAB1C  C-GCCC--AC--T--G------------.....CAA.......................-A-----T-------------TTTGGCCC----GG..............-AGAGG-AA-CC...CAG   2425

SUN.GA.SIVSUN     ----ATG-----TA--CAAGCTC-A--C.....C-ACAA...................-GA-A-----TG------------T-TGGCCC----AA..............-AGGGG-CCGCC...AGG   2426
LHOEST.KE.LHOEST  C-GTA---AG--TAG-T-TGC-C-G--C.....C-A......AAA.........CCAA-AG--A---TG-----------AT-TGGTCC----AG..............-AG---ACCTCC...AGG   2427
LHOEST.CD.447     T-GCA-G-A---TAGGT-TGCTC-A---.....C-A......AAG.........CCTA--G--A---TA-----------CT-TGGACC----AG..............GAGTGGTCCTCC...AGG   1343
LHOEST.CD.485     T-GCA---A---TAGGT-TGCTC-A---.....C-A......AAG.........CCTA--G--A---TC-----------CT-TGGACC----AG..............GA-TGGTCCTCC...AGG   1343
LHOEST.CD.524     TGCTA---AC---AG----GCCC-G--C.....C-A......AAA.........CCAA--G-GA---TG------------T-TGGTCC----AG..............GA-TGGACCTTC...AGG   1337
MANDRIL.GA.MNDGB1 -GCAATG-AT--T--G-AAGCTC-A--C.....C--AAGCCAGCT.........CAGC--CAGAG--T-------------T-TGGCCCT---GG..............TCC-TCCAA-CC...-GG   1783

COLOBUS.CM.CGU1   TC-----------ATGCAA-G-------.....C-C.....................A-C-AG-----A-----------.......................................   1922
SYKES.KE.173      ---AAC--------T-GCA-GG-CC---AGGCAAC-CAAGAGAAATCAAGGCCCTCCA-TTGC---------------------AGGATGGG--GT............CTC-AGG-G-CCCCCAGCCAA   2227
RCM.NG.NGM        -G-------T--T-T------G-----C.....C-AAAA.......................G---T---------------TGCTCC----AAGCA............GGGG--ACCG-------AA   2421
MANDRIL.GA.14CG   GG-CAC------TATG----G-A-C--C.....C-AAAA.......................AT--TG-------------TGC.......................TCC-TG---C--TGGCAA   2006
MANDRIL.14CG Gag  __D__T__G__H__M__M__R__N__C__.__.__P__K__.__.__.__.__.__.__.__.__M__V__N__F__L__G__L__A__.__.__.__.__.__.__.__.__.__P__W__G__S__G__K    p6
MANDRIL.14CG Pol                                                                                _F__R__E__C__.__.__.__.__.__.__.__.__S__L__G__Q__W__Q_ Pol TF
```

HIV-1/HIV-2/SIV
complete genomes

```
                    Gag p1 end \/ Gag p6 start
H1B.FR.HXB2         AGGGAATTTTCTTCA...........GAGCAGACC.........AGAGCCAAC....................AGCCCCACC...........AGAAGAGAGCTTCAGGTCTGGGGTAGAGACAAC.   2198
H1B.FR.HXB2 Gag     __G__N__F__L__Q__.__.__.__.__S__R__P__.__.__.__E__P__T__.__.__.__.__.__.__.__.__A__P__P__.__.__.__.__E__E__S__F__R__S__G__V__E__T__R_    p6
H1B.FR.HXB2 Pol     _R__E__F__S__S__.__.__.__.__E__Q__T__.__.__.__R__A__N__.__.__.__.__.__.__.__.__S__P__T__.__.__.__.__R__R__E__L__Q__V__W__G__R__D__N__ Pol TF

CPZ.US.CPZUS        ------C--C--.................GGA-AACAG...GAAG-----C--....................---T-----...........-AT----GAT---G---A-CAA-A-------GT.   2264
CPZ.GA.CPZGAB       ---C--C---G-G--............--A----A-.........---A-----....................G---------..........GAT------T-ATG---ACCA--AG--AGA....   2262
CPZ.CM.CAM3         ------C--C--G--............-G----GAA.........G--A--C--....................---T--G--...........CCT------T-ATG---A-CA--A--CA---GG.   1603
CPZ.CM.CAM5         ------C----CG--............-G--G-GAA.........G--A--C--....................---T-----...........CATG-----T-ATG---T-CA--A-ACA----T.   1906
CPZ.CD.CPZANT       ------C---G-G--............--AGGAGGAAGTAGT...G---------....................---T-----...........CAT-----T--A......-CA--AG---CA....   1637

H1A.UG.U455         -----------C---............---------.........---------....................---------...........--C---A-T---TG--ATG----A-A---T....   1644
H1C.ET.ETH2220      ---------C-----............---------.........---------....................---------...........---GAGTCT-AGACCAGAGCCAAC--CCC--C-A   1570
H1D.ZR.84ZR085      G-----C--------............---------.........---------....................---------...........--C---AGCT-CGG-T-TG---AGGAGATA-C-.   1718
H1F.BE.VI850        C--A-----C-----............---------.........---------....................---------...........--C----------G---TCA-A-AG----T---.   1535
H1G.SE.SE6165       ---------------............--A---GA-.........---A-----....................---------...........T-C---A---C--G---TC--A-AG----T-G-.   1604
H1H.CF.90CF056      ---A--------C--............------G--.........---A-----....................---------...........--C----------G---TC--A-AG----TG--.   1548
H1J.SE.SE92809      ------C-----C--............---------.........---------....................---------...........--C-------C--G--CTC--A-AG----T....   1515
H1K.CM.MP535C       ---------------............---------.........---------....................---------...........--C----------G---TC----AG----T---.   1394
H101_AE.TH.CM240    G----------C---............---------.........---------....................---------...........--C---A-A--GGG--ATG----A-----T---.   1757
H1N.CM.YBF30        ---A--C--C-CC--............--CA-C-A-.........-AG-AA-GAGCCCAC..............------G--...........-CT------T-ATG---T-CA--AG--.......   1771
H1O.CM.MVP5180      ---C----A-G-G--............--AACA-GT.........GTCC---T-....................---------..........................................   2207
H1O.BE.ANT70        ---C----A-G-G--............---ACC-G-.........-C-C---T-....................---------..........................................   2231

MAC.US.SMM251       TTTCCCCA-GGC---............AGTGCAT-A.........G-G--TG--GCCAAC..............T--T--C--..........................................   2425
MAC.US.SMM239       TTTCCCCA-GGC---............AGTGCAT-A.........G-G--TG-TGCCAAC..............T--T--C--..........................................   2437
MAC.US.SMM239 Gag   __F__P__M__A__Q__.__.__.__.__V__H__Q__.__.__.__G__L__M__P__T__.__.__.__.__.__A__P__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._    p6
MAC.US.SMM239 Pol   _F__P__H__G__S__.__.__.__.__S__G__S__.__.__.__G__A__D__A__N__.__.__.__.__.__C__S__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__ Pol TF

STM.US.STM          CTTCCCCA-GGCC--............A-TACCT-A.........G-G--TG--GCCAAC..............T-----T--..........................................   2096
SMM.US.SMM9         TTTCCCCA-GGCC--............--TGCCT-A.........G-G--TG--ACCAAC..............T--T--Y--..........................................   1922

H2A.GW.ALI          CTTCCCCG-AACC-G............AGTTCCG-A.........G-GATT---ACCAAC.......................................--C-CCCCCAGCAGA-C-A-CA-CG--CCTGCTG   2499
H2A.SN.ST           TTTCCC-G-GGCC--............A-T-CCG-A.........G-G--TG--ACCAAC.......................................--C-CCCCCGA-AGACC-A-TA-AG--CCT-CTA   1951
H2A.DE.BEN          CTTCCCCG-GACC--............AGC-CCT-A.........G-G--TG-TACCAAC.......................................--C-CCTCCGGCAGATC-A-CA-CG--ACTGTTG   2506
H2B.GH.D205         CTTCCCCA-GACC--............AGTGCCT-A.........G-GAGTG--ACCATC..............T--A--C--..................--TGAA-CCAG-A-A--GCAT----C-T   2482
H2B.CI.EHO          TTTCCCCG-C-AGGC............ACC-CAGGG.........GATAGTGC-ATC.................T--G--C--..................--TGAA-CCAG-ATTC-GCAT----C-T   2477
H2G.CI.ABT96        CTTCCCCA-GACC--............AGT-CCT-A.........G-GT-TG--TCCATC.......................................--C-CCACCGA-GGACC-A-CA-----CCT-CTG   1884

TANTALUS.UG.TAN1    CAATTT---AGAG--............AG--G--G-.........C-TT--G--GGC.................TC----GAT...........GCCT-C-CA-GGATTCC-CACA-GTTC-C--G-T   2344
GRIVET.ET.GRI677    -AATTT-G-G-AAT-............C--AG--GA.........CACAGTTGGTCTGGAACCAAC.........------C--...........-ATG--A-CAGC.......................   2126
VERVET.KE.VER9063   -AACTT-C-CGC-GC............C-CTCATGG.........--T-GA-C-GAG.................C--G--C--...........TCC-CC--CACC.......................   2359
VERVET.KE.VERTYO    -AATTT-CCCGCCGC............T-CTCTTGG.........--C-GA-C-GAG.................T--G--T--...........TCC-CC----GG--CCA-CCC.............   1849
VERVET.KE.VER155    -AATTT-CCCGCAGC............C-CTCTTGG.........G-C-GA-C-AAG.................T--G--C--...........TCC-CC--A-AA-TCTA-ACC.............   2351
VERVET.DE.VERAGM3   -AATTT-CCCGCCGC............T-CTCTTGG.........G-T-GAGC-AAC.................T--G--C--...........TCC-CC---TCC......................   1854
SABAEUS.SN.SAB1C    GAATTTCCC---GAC............CTC--TCAG.........-CCAA--G-TCCCCCAATGGA...ACGGGATTA-AGCAG...........GCCG---GAAAATT---A--CA-ACAGAC-TC-G   2519

SUN.GA.SIVSUN       GAACTT-CCAG-CAT............-CCTTC.................................T--T--G--.....................TC-AGA-GA-CT-AC-TT-GGG-A.   2480
LHOEST.KE.LHOEST    CAATT--CC---A-T............-G-AG--G-.........T-CAGGG-GAATTCC.....................................................................A   2467
LHOEST.CD.447       -AACTT-CC-T-GAT............-G-GG-GAA.........T-CAGG-GTAGTACC.....................................................................A   1383
LHOEST.CD.485       -AACTT-CC-T-GAT............-G-GG-GAA.........T-CAGG-GTAGTACC.....................................................................A   1383
LHOEST.CD.524       CAACTT-CCAA-CAT............-G-AG-GG-.........T-CAGG-GTGATTCC.....................................................................T   1377
MANDRIL.GA.MNDGB1   GAATT--CC......................................................................................................................GG-A   1796
MANDRIL.GA.14CG     GCCC-GGAACT---C............TGC--TG--.....................................ACTGACCCCATC--C-CCCCGA-GCCAGGGAT--AG--TC--G-A   2076

COLOBUS.CM.CGU1     ...........................-GCA-CCAG.......................................................................................   1931
SYKES.KE.173        CTTCCCAG-GAGGAGCGAGCCATCTGCTCC-CCT-T.........G--...................................................G--C-T-GAAGATGGGCCATGGCTC---TGG   2297
RCM.NG.NGM          CCTTCCC--GAC-TC............CTTG-----.........TACTG-TC-AGGGAGAGAGGAAGGATGGAATC---AGTA............C--TCCAGCAGAAGA-ATGCTCAGGA-.......   2511
MANDRIL.14CG Gag    __P__R__N__F__P__.__.__.__.__A__M__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__L__T__P__S__G__P__P__M__P__G__M__E__D__P__A_    p6
MANDRIL.14CG Pol    _A__Q__E__L__S__.__.__.__.__C__H__A__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__D__P__I__S__T__P__D__G__R__D__G__G__S__S__ Pol TF
```

```
H1B.FR.HXB2        ........................................................................AACTCCCCC.....................TCAGAAGCAGGAGCCGATAGACA   2230
H1B.FR.HXB2 Gag    _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__P__P__.__.__.__.__.__.__.__Q__K__Q__E__P__I__D__     p6
H1B.FR.HXB2 Pol    .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__N__S__P__.__.__.__.__.__.__.__S__E__A__G__A__D__R__Q Pol TF

CPZ.US.CPZUS       ..........................................................GACACA......GGAGAAACAGGGAAA.........................GG-A---G-..............   2294
CPZ.GA.CPZGAB      ..............................................................GAAGAGCCAGGAGAAGAAAGAGGGAGA............GAGC-G.................   2295
CPZ.CM.CAM3        ..........................................................AAAGAGACAGGAGGAGGAGAGGGAGAAGGA.....................AA--G-A-C..............   1642
CPZ.CM.CAM5        ..............................................................GGAGAAAAAG-AGGAGGA.....................GA--G-A-C..............   1933
CPZ.CD.CPZANT      ..............................................CAAGAGGACTCAGAAGGG--T-AA....................GGG-G--G-..............   1670

H1A.UG.U455        ........................................................................G--CT----------------------GC---A-------T--A-----   1676
H1C.ET.ETH2220     CCAGA.....................................................GAGCTTCAGGTTCGAGGAAG-AA-.....................A-CTTCT-C-A---A-GAGCTG-   1622
H1D.ZR.84ZR085     ..............................................................CCCT-C-AGAA.....................A---G-A---A--GAC-A-----   1752
H1F.BE.VI850       ........................................................................CC-CT-T--....................GA--C--G--C--AAAGAC-GGG   1567
H1G.SE.SE6165      ..............................................................................................C-CCTCC-C-A---A-GAGATG-   1627
H1H.CF.90CF056     ..............................................................................C---T-.....................--C-----------A-C-GA-GG   1577
H1J.SE.SE92809     ..............................................................................C---T-.....................C-C---A----------AG----   1544
H1K.CM.MP535C      ..............................................................CC-CT-T--.....................GAG-C--G--ACCAAAGACA-GG   1426
H101_AE.TH.CM240   .....................................................GGGGGAAGAGATAACCT-CTTA--.....................GA--C--G--C--AAAGACA-GG   1804
H1N.CM.YBF30       ..........................................................GAAGAGCACACAGGGGAAGGAG-TG-AGGA.....................GA-CC--G--AG-A-AGAGA--T   1824
H1O.CM.MVP5180     ..........................................................AATGGAGGAGGCAGTGAAGGA-CAAGAGAA.....................-----GT---A--GG-GATC-GG   2260
H1O.BE.ANT70       ..........................................................GATGGAGGAGGAAGTGAAGGG-CAGGAGAA.....................----G-A---A-AGG-GGCCCG-   2284

MAC.US.SMM251      ........................AGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAGA.....................AAGC-GAG--A----TTACA-GG   2508
MAC.US.SMM239      ........................AGAGGACCCAGCTGTGGATCTGCTAAAGAACTACATGCAGTTGGGCAAGCAGCAGAGAGA.........AAAGCAGAGAGAAAGC-GAG--A----TTACA-GG   2532
MAC.US.SMM239 Gag  _.__.__.__.__.__.__.__.__.__E__D__P__A__V__D__L__L__K__N__Y__M__Q__L__G__K__Q__Q__R__E__.__.__.__K__Q__R__E__S__R__E__K__P__Y__K__     p6
MAC.US.SMM239 Pol  .__.__.__.__.__.__.__.__.__R__G__P__S__C__G__S__A__K__E__L__H__A__V__G__Q__A__G__E__R__.__.__.__K__A__E__R__K__Q__R__E__A__L__Q__G Pol TF

STM.US.STM         ........................AGAAGACCCAGCTGCGGATCTGCTGAGAAGTTACATGCAGCTGGGCAAGAAGCAGAGAGA.....................GAGC-G-A--ACA--CTACA-GG   2179
SMM.US.SMM9        ........................AGAGGATCCAGCTGTGGATCTACTGAAGAATTACATGAAGGTGGGCAGAAGGCAGAGGGA.....................GA-C-GAG--AGA--TTACA-GG   2005

H2A.GW.ALI         GAGCAGTATATGCAGCAGGGGAGAAAGCAAAGAGAGCAGAGAGAGAGGCCATACAAAGAGGTGACGGAGGACTTACTGCAC-T-GA.....................G---GGAG--ACA--ACACA-AG   2608
H2A.SN.ST          GAGAAGTACATGCAGCAAGGGAAAAGGCAGAGAGAGCAGAGAGAGAGGCCATACAAAGAAGTGACAGAGGACTTCCTGCAG-T-GA.....................GA-AC-AG--ACA--ATGCAGAG   2060
H2A.DE.BEN         GAGAGATATATGCAGCAAGGGAGAAAGCAGAGGGAGCAGAGGGAGAGACCATACAAAGAGGTGACGGAGGACTTGCTGCAC-T-GA.....................G----GAG--ACA--TCACAGAG   2615
H2B.GH.D205        CGGGGGGCGACACCATCTGCGCCCCCTGCAGATCCAGCAGTGGAGATGCTGAAAAGTTACATGCAGATGGGGAGACA-CAGAGAGA.....................GAGCCGAG--AGA--CTACA-GG   2591
H2B.CI.EHO         CAGGGTGCGATTCCATCTGCACCCCCTGCAGATCCAGCAGAGGAGATGCTAAAGAACTACATGCAACTAGGGAAGAAGCAGAAGGA.....................GA-C-GAG--AGA--CTACA-GG   2586
H2G.CI.ABT96       AAGAATTACATGCARCTGGGGAGGAAGCAAAAGGAGCA.....................................................--GAA-CA-A--CTACA-GG   1942

TANTALUS.UG.TAN1   GCGGGAGC...............................ATACGACCCAGCCAGGAAACTGCTGGAGCAGTATGCC-AAAAGGG.....................AG-TC-ATT-AGAAA-CAGA-GG   2420
GRIVET.ET.GRI677   ................TTACGATCCAGCAAAGAAGCTCCTCCAGCAGTATGCAGAGAAGGGACAGCGCCTGAGAGAGGAGAGAGA.....................A----CAAG-A-A-A--AG--G-   2218
VERVET.KE.VER9063  ................ATACGACCCAGCAAAGAAGCTCCTGCAGCAGTATGCAGAGAAGGGAAAACAGTTGAGGGA-CAAGGGAA.....................GAGA-CA-CACCAA-A-ATCCAG   2451
VERVET.KE.VERTYO   ................ATACGACCCAGCAAAGAAGCTCCTGCAGCAATATGCAGAGAAAGGGAAACAACTGAGGGAGCAAAAGAG.....................GA-TCCA-C--CAAT--ATCCGG   1941
VERVET.KE.VER155   ................TTACGACCCAGCAAAGAAGCTCCTGCAGCAGTATGCAGAGAAAGGGAAACAAATGAGAAATCAGAA-AG.....................AA-CCCC-CA-C-AAC-ATCCAG   2443
VERVET.DE.VERAGM3  ................ATACGACCCTGCAAAGAAGCTCCTGCAGCAATATGCAGACAAGGGGAAGCAGTTGAGGGA-CAAAGGAA.....................AA-ACCA-CA-CAGT--ATCC-G   1946
SABAEUS.SN.SAB1C   ACCAGAGGACCAGGGCCAGACGATCCAGCAACAGCCCTGTTAAAGCAGTATGCTGTTCAGGGGAAACGGCAGAAACAGCAGTGG-A.....................AA-CC-CTC-CCA-AACAGAG-C   2628

SUN.GA.SIVSUN      .............CAGAATGACACCACCCCAGTCAAAGGCAGAGAGAGCTCTGGAGACCTACAGACTGTTAGGACAGGGC-T-AG.....................AGCAC-----A--AG--AG--A-   2575
LHOEST.KE.LHOEST   TCAGCTCCCCCAATGGAGAGTGCTCCAACAAAAGCAGAGAGAGCACTGGA...GACTTACAGGACTTTAGGACAACAGCTGAAAAG.....................G---C-----C-AGTCCC-C-G-   2573
LHOEST.CD.447      TCAGCTCCCCCAATGGAGAGGAGCCCAACACAAGCAGAGAGAGCCTTAGA...AACATACAGGAACCTCGGACAACAGCTGAGGAG.....................G--AC----...-ATCCC-C-G-   1486
LHOEST.CD.485      TCAGCTCCCCCAATGGAAAGGAACTCAACAAAAGCAGAGAGAGCCTTAGA...AACATACAGGAACCTCGGACAACA-CTGAGGAA.....................G--AC----...-GTTCC-C-G-   1486
LHOEST.CD.524      TCAGCGCCACCCATGGAGAAAGTGCCTCCAACAAGAGCAGAGAGGGCAATAGAGACATACAGGAACCTAGGACAACA-CTGAGGAA.....................G--AC-A--...AGTTCC-C-G-   1483
MANDRIL.GA.MNDGB1  CAAGAGGTGACTCCAACAGCTCCACCATTAGAGGAGAAACCTCTGCAGAAAACTCTGAGCACTTATCAGAAATTAGGG-GAGGG-T.....................CAG-C--A--AT-AA-GAG--G-   1905
MANDRIL.GA.14CG    GAGAAGATGTTATTAGACTATATGAGAAAGGGGCAACAACAGAGGGCAGCAAAGGAGAC.....................................CA--C--G--A--GAC-AG-GTC   2158

COLOBUS.CM.CGU1    ...............................GGGGGTGGAGCTACAGACAGCCATCTT...CCCCCCGAAGATGAGC-AAGA--T.....................G-CACC......CAG--GG--G-   1999
SYKES.KE.173       TCAGCACAGATGAGCCAACAAGCACAAGCGAAGGCACAGAACTCTCCCTCGAAGAAACCCCCCACCAACAGGGAAGTGCTCT-G--.....................GA--G--AGCAGTGG--AG--GG   2406
RCM.NG.NGM         ...........................................ATACCTAGCATTGGGGAGACAGCACAAACAGGAGCAG-GGAAAGA.....................GA-C---G--A-AGT-GGGAGAG   2579
MANDRIL.14CG Gag   _E__K__M__L__L__D__Y__M__R__K__G__Q__Q__Q__R__A__A__K__E__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__K__Q__E__K__D__K__G__     p6
MANDRIL.14CG Pol   R__E__D__V__I__R__L__Y__E__K__G__A__T__T__E__G__S__K__G__D__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__Q__G__G__E__G__Q__G__S Pol TF
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
                                Pol TF \/ Pol Protease p10          Gag p6 end \
H1B.FR.HXB2        AGGAACTGTATCCTTTAACTTCCCTCAGGTCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATG   2360
H1B.FR.HXB2 Gag    K__E__L__Y__P__L__T__S__L__R__S__L__F__G__N__D__P__S__S__Q__$_                                                                          p6
H1B.FR.HXB2 Pol    __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__L__K__E__A__L__L__D__T__G__A__D__D__T__V__L__E__E__M_    p10

CPZ.US.CPZUS       .-CCCT-...--AGC-----------AA-------------G---------C---G------AG---CA-----GA--GTA--------T-----C--------T------------C----CA-C--A   2420
CPZ.GA.CPZGAB      .TCT-TA...---ACC---C-------AA-------------G---------A--C--G----AG---AA---------TGT------T-GC-------------T------------A----GAG---A   2421
CPZ.CM.CAM3        .CCTGTATCC---G-----C-------AA-------------G---------C--GG-G----A----AA-----G----TA--------GC-------------T------------A----TA-TG-A   1771
CPZ.CM.CAM5        .CCTGTATCC--AA-----C-------AA-------------G---------C-TGG-G----AG---AA-----GT---TA--------TC-G--------G--T-----C------A----TA-TG-A   2062
CPZ.CD.CPZANT      .------ACC---C-CGTA------G-AA--C-----------A----AA-GA-GGA-G-TCTC---CA---A---AA-TGTC-G--------------------T-----C------G----G-G---T   1799

H1A.UG.U455        G-----A-AC-------GT--------AA---------------------T------G----A-----A--A--G--G-TA--------------------------------C--------C--A   1806
H1C.ET.ETH2220     -A--CAG-G-AG-C-------------AA-------------------A--T--T--------A--------A--G--------G-----C-----C-------------------------------A   1752
H1D.ZR.84ZR085     -------------C---G---------AA--------------------G-T--------------------A--G-----A-----------------------------------------C--A   1882
H1F.BE.VI850       -ACTGTACCC---C---G---------AA--------------------T-A-----------A--------A--GA------G-----------------------------------------C--A   1697
H1G.SE.SE6165      -----AA-G-G-TA-ATC-C-------AA-------------G--------A------G----A--------A--G----TA-----C--------C--------------C----------------A   1757
H1H.CF.90CF056     -CA-GGAACC---C---G----------A-------------G--------T--T---G----A----A---A--GT---G-------------------------------------------G--A   1707
H1J.SE.SE92809     --------------C------------AA-------------G--------T----------GA-----------G--G-G-----------------------------------------C--A   1674
H1K.CM.MP535C      -AC-GAGCCC-----------------AA--------------------A-T-----------AG-------A--GT---GA-----------------------------------------------A   1556
H101_AE.TH.CM240   -AC-T-CTCC-------GT--------AA---------------------T------G----A-----A--A--G--G--A-----------------------------------------T--A   1934
H1N.CM.YBF30       CTCTGTACCCA-----------------A--------------------G--A-----G----A-----AAAG--G---GA--------T--------------T------------A-------GC-A   1954
H1O.CM.MVP5180     -A--G-----C--A--TG-C-------AA--C--------G-CA----AA-A------GC----G-T-----T--T---TGT--G---T--C-G--------G------------------A-TA-C--A   2390
H1O.BE.ANT70       -C--G--C-----G--TG-C-------AA--C--------G-CA----AA-A--T---GC--GAG-T-----C--C---TGT----T-T-GC-G-----------------C------C--A-CA-C--A   2414

MAC.US.SMM251      ---TGACAG-GGA---GCTGCA-----AT--T--------AGGA----AG-A-----TGCTC-T--T-AA--A--G-CTGTA----TAT----G--------G--T------T-TA-TG--AC--G---A   2638
MAC.US.SMM239      ---TGACAG-GGA---GCTGCA-----AT--T--------AGGA----AG-A-----TGCTC-T--T-AA--A--G-CTGTA----TAT--C-G--------G--T------T-TA-TG--AC--G---A   2662
MAC.US.SMM239 Gag  E__V__G__E__D__L__L__H__L__N__S__L__F__G__G__D__Q__$_                                                                                  p6
MAC.US.SMM239 Pol  __G__D__R__G__F__G__A__P__Q__F__S__L__W__R__R__P__V__V__T__A__H__I__E__G__Q__P__V__E__V__L__L__D__T__G__A__D__D__S__I__V__T__G__I_    p10

STM.US.STM         ---TGACAG-GGA---GCTGCA-----AT--T--------AG-A----AG-A------GCCC-T--T-A---T--G-CTGTA----TA-----G--C-----G--T--C--CT--A--G-G-CG-GG--A   2309
SMM.US.SMM9        ---T-AC-G-GGA---GCTGCA-----AT--T--------AG-A----AG-A-----TGCCT-C--T-AA-AA--G-CC-TA----TAT-------------G--T------T--A-TG---C--GG--A   2135

H2A.GW.ALI         ---TGACAG-GGAC--GCTGCA-----AT--T--------A--A----AG-A------GCCT-C--T-A---C--G-C-GT-----T-T--C----C-----G--T--C--CT--A--G---C--G---A   2738
H2A.SN.ST          --ACGACAG-GGAC--GCTGCA-----AT--T--------A--A----AG-A------GC-C-TG-T-A---C--G-C-GTA----T-T-GC----C-----G--T--C--CT--A--G---C--GCG-A   2190
H2A.DE.BEN         ----GACAG-GGAC--GCTGCA-----AT--T--------A--A----AG-A------GCGT-C--C-A--AT--G-CGGTA----TCT--C----C-----G--T-----CT--A--G---C--G---A   2745
H2B.GH.D205        ---TGACAG-GGA---GCTGCA-----AT--T--------AG-A----AG-A----A-GC-TGT--C-A---T--GTC-GTA----TAT--C----C-------TT--C--CT--A--G---C--GG--A   2721
H2B.CI.EHO         ---TGACAG-GGA---GCTGCA-----AT--T--------AG-A----AG-A----A-GC--CT--T-A---T---TC-GTA----TAT--C----C--------T-----CT--A--G---C--GG--A   2716
H2G.CI.ABT96       ---TGACAG-RGAH--GCTGCA------C--T--------AG------AG-A-----TGC-T-C----AAA-T--G-C-GTA----TA--------C-----G--T-----CT-TA--G---C--G---A   2072

TANTALUS.UG.TAN1   --A-GGAACTGGAGGACTA----T-G--T--C--------AGGA----AA-GAGT-AG--CT-C---CA---C--G-C-GTAC---T---G-------------T--------CA--A--A-G---GAA   2550
GRIVET.ET.GRI677   -A---G--G-GGA-G-TT-C-TGAG-TC...C--------AGGA----AA-GAAACG-G-C-TC----AA--AACG-C-GT-C----CT-G--------------------C--TA--A-TC------A-   2345
VERVET.KE.VER9063  -CTGGGC-G-AGGA-ATT---TGAA-TC...C--------AG-A----AA-AAAA---G-C-TT--T-AA--AGTC-CTGTAA---TC--G-----------G--------C--CA-CA--A------CA   2578
VERVET.KE.VERTYO   -TTGGACCG-GGGA-ATT---TGAA-TC...C--------AG-A----AA-AAAG---G-GT-T----AA---GTC-CC-TTA-G--A--GC----C-----G--------C--CA--A-TA------AT   2068
VERVET.KE.VER155   -TTGGAACG-GGGA-ATT---TGAA-TC...C--------AG-A----AA-AAGG--CTGT-TA-----A--AACTGCCGTTA-G--AT-------------G--------C--TA--A--A-G--T-CA   2570
VERVET.DE.VERAGM3  -TTGGACAG-GGGA-ATT---TGAA-TC...C--------AG-A----AA-AAAA---G-TT-C----AA---GTC-CC-TCAG---AT----------G--G-----------CA-TA--A-----GCA   2073
SABAEUS.SN.SAB1C   CCT-CGA-G-GG---AC-GC-----------T--------AG-A----AA-GAAA--TG-GT-C----A---A--GAA-GTCACT-----------C--------------CT-----A--C-G-GT--A   2758

SUN.GA.SIVSUN      G--GGGAA-G--AGGAGC-A-G-T-G-ATATG------CCAG-A-----T-A-AG-AGG-CTTC--TAAT-------CTGT-ACT---T----G--C--G-----------C---A--G-GTCT---GAT   2705
LHOEST.KE.LHOEST   -ATGTG--G-CGAGCCCTGCCTGAATTCACTCT-...-CCAG------AA-A-AA-GT---TT-G-CAAT-------C-GT-TC---A--G--G--C--T--G--T-----C---A-T--TTC-----AT   2700
LHOEST.CD.447      -ATGTG--G-CGAGCC-TGCCTGAGTTTC-TCT-----CCAG------AA-A-AA-AC--TCT-G-CAAT-----G-C-GT-TCT-----TC-T--C--T--G-----------TA-C--TTC-----AT   1616
LHOEST.CD.485      -ATGTG--G-CGAGCC-TGCCTGAGTTTC-TCT-----CCAG------AA-A-AA-AC--TCT-G-CAATA----G-C-GT-TCT-----TC-T--C--TA-G-----------CA-C--TTC-----AT   1616
LHOEST.CD.524      --TGTG--G-CGAGCCTTGCCTGAATTTC-TCT-----CCAG------AA-A-AG-A-G--CT-G-CAAT------GA-GTAACT--C-----G--C--T--------------A-C--TTCT----AT   1613
MANDRIL.GA.MNDGB1  --AG-GA-G-GGA---TCA------G--TA-T------CAAG-A----AA-A-AAGAGG-CTCAG-G-AT--TGTCACT-TAAG-------C-------------T--------CA-C--TA-T----GA   2035

COLOBUS.CM.CGU1    ---GGGA-AG--TC-ATC-C---T-G-A---T-----C--AG------AA-A--T-T-G--G-AG---AA--A---AAGTGCC----G--G-------------------------T--TA-----GAA   2129
SYKES.KE.173       --ACCAA-AG--TC-ATC-C--T-----C--T--------AG-A----AA-GA-AGA-G--G-TG---AA--TG-C--GGT-C--ATGT--G----------G-----------TA--A-TAG----GAA   2536
RCM.NG.NGM         CCT-CGA-G--G-G--GT-C-------AT--T--------A-G-----TG-AACATGGG--G-A----A-------AAGGT---G--CT--------------------G------TA--C--A-TT-A   2709
MANDRIL.GA.14CG    CCT-CGA-GCAG---AC-AC--------T--T--------A-CA----AACTACAGT-G--G-A----A---A---AAGGTA-----C---C-G--C--T-----------------A-CA----TG-A   2288
MANDRIL.14CG Gag   P__Y__E__G__A__Y__N__S__L__S__S__L__F__G__T__D__Q__L__Q__$_                                                                                  p6
MANDRIL.14CG Pol   __L__R__G__S__L__Q__L__P__Q__F__S__L__W__N__R__P__G__G__V__V__E__I__E__G__Q__K__V__E__A__L__L__D__T__G__A__D__D__T__V__I__K__D__V_    p10
```

```
H1B.FR.HXB2       AGTTTG...........CCAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATA.........   2468
H1B.FR.HXB2 Pol   _S__L__.__.__.__.__P__G__R__W__K__P__K__M__I__G__G__I__G__G__F__I__K__V__R__Q__Y__D__Q__I__L__I__E__I__C__G__H__K__A__I__.__.__.__   p10

CPZ.US.CPZUS      CAAA-T...........GA---G-C-----G------------G--------------------A-------A---A--------CG-CAAT--------AGAG---AGA-----ACAG.........   2528
CPZ.GA.CPZGAB     CAA--A...........-A----CTT--------------------A-----------------------C-A---A-T----A-TG---A------G--AGAA--GAGA----TAG--.........   2529
CPZ.CM.CAM3       CAAC-A...........A------A-----G-----------------------------------T-A---A-----CA-C---GCA--------TGAG---AGA---A-A-C-.........   1879
CPZ.CM.CAM5       CAA--------------A------AG----G-----G---------------------------T-----T-A---A------A-C---G-A--------AGAG--CAGA---A-A-C-.........   2170
CPZ.CD.CPZANT     CA----------------A-----AC-----G-------CA--T--T--------G--------TTCC---CA-------A-CA-AG---C----C----TG-A-ACAGG-C--TAC--.........   1907

H1A.UG.U455       -A----------------------A---------------A-----------------------------------------------T--------T------A-A--GA-----.........   1914
H1C.ET.ETH2220    -A----------------------A--------------------A-----------------T--------------------A---A----------T------A-A--G------.........   1860
H1D.ZR.84ZR085    -A----------------------AG-------------------------------------------------------------C-------------------------G------.........   1990
H1F.BE.VI850      -A----------------------A-----------------------------------------------A----------A-C--------------T--------C--G------.........   1805
H1G.SE.SE6165     -A---A...........-------------------------------------------------------------A--------AG---CT--------TA-----A-A--G------.........   1865
H1H.CF.90CF056    -A----------------G----A-------------------------------------------------------G--AG--GC----------------A-A--G------.........   1815
H1J.SE.SE92809    GAC--------------C---A-------------------------------------------------------A-CG--G---CG--------TGAG---A-A--G------.........   1782
H1K.CM.MP535C     -A----------------------A-----------------------------------------------A--------AG----T--------T--------A--G------.........   1664
H101_AE.TH.CM240  -A----------------------A-----------------------------------------G---AG--A--------------T--------------A-A--G------.........   2042
H1N.CM.YBF30      CAA--A...........GAG----A--------------------A-----------A-----------G-----A------A-T---ACAG----C--ACAG---AGA-----AG-T.........   2062
H1O.CM.MVP5180    CAA--A...........GA-----------C------------------T--A-----C-----A-------A-G-----A-CA-TG-GACAG-----G-ACAA---A-GG---TACAG.........   2498
H1O.BE.ANT70      CAA--------------GA-----A-----------------------T--A-----------A-------A-G-A------A-TG-GACAG-------AGAA---AGGG-G-TACAG.........   2522

MAC.US.SMM251     GAG--A...........GGTCC-CAT-AT-CC--------AG----A-----A-----------T--TACT-A-G-A--CA-AA-TG--AAA------G-T-TA--CA-A-GGAT--A-.........   2746
MAC.US.SMM239     GAG--A...........GGTCC-CAT-AT-CC--------AG----A-----A-----------T--TACT-A-G-A--CA-AA-TG--GAA------G-T-TA--CA-A-GGAT--A-.........   2770
MAC.US.SMM239 Pol _E__L__.__.__.__.__G__P__H__Y__T__P__K__I__V__G__G__I__G__G__F__I__N__G__K__E__Y__K__N__V__E__I__E__V__L__G__K__R__I__K__.__.__.__   p10

STM.US.STM        GAGC-A...........GGGCT-CA--AC-CC-----GG-AG----T-----A-----A-----A--TACT-AGG-A-T-A-GA-TG--AA------GG-A-TA--CA-G---ATA-A-.........   2417
SMM.US.SMM9       GAA--------------GGTCC--AT-AT-CC--T-----AG----A-----A-----C--C--T--CACC-A-G-A---A-AG-TG--AAA---A--G-A-TA--CA-AGT-AT--AG.........   2243

H2A.GW.ALI        GAG--A...........GGGA-C-AC-AT-CC--------AG----------A--G--A--C--A--TACC-A-G-A-----AG-TG--GAA---A--G-ACTAAATA-A-G--TA-A-.........   2846
H2A.SN.ST         GAG--A...........GGGA-C-AT-AT-GT-----G--AG----------A--G--A--C--A--TACC-A-G-A---A-AA-TG--GAA---AG-G-A-TAAATA-A-G--TA-G-.........   2298
H2A.DE.BEN        GAA--A...........GGG-AC-AT-AC-CT--------AG----------A--G--A-----A--CACC-A-G-A--CA-AA-TG--GAA---A--G-ACTAAATA-A-G--TA-G-.........   2853
H2B.GH.D205       GAA--A...........GGTA-C-AT-AC-CC--------AG----A--G--A-----G--C--A--TACC-A-G-A--CA-AG-TG--GAA------G-AGTG---A-A-G--TA-GG........   2829
H2B.CI.EHO        GAA--A...........GGCA-C-AT-AC-CC--------AG----T--G--A-----A-----A--TACC-ATG-A--CA-AA-TG--GAA------G-AGTA---A-A-G--TA-G-.........   2824
H2G.CI.ABT96      GAA--A...........GG-A-T-AT-AT-GT--------AG----A-----A-----A-----T--TACC-A-G-A--CA-AA-TG--GAA--TA--G-GGTA---A-A-G-ATACAG.........   2180

TANTALUS.UG.TAN1  GA-A-C......CATTTAGA----C-C---CGG--T---G-AG----A--T--A-----GGGA---TC-----A---A-----CA-T--TT-AG-GA-T--AGA----A-AG--A---G-.........   2664
GRIVET.ET.GRI677  GAC---CACTTTCCCCCA-ATAA-CC----CGTT-C--GG-AG----A--T--A-----AGGG--TC-T--C-A-G-A---C-GGG-G---AAG--C--T-GGAG-ATA-A-TCATC-CC.........   2465
VERVET.KE.VER9063 GACA-A......CAACTAA-----CCT----G-------G-A-----A-----------AGGAC-T--T----A-G-----AG-A-TG--GAAG--A----AGAA-ATA-A-TTTTG-G-.........   2692
VERVET.KE.VERTYO  GA---A......CAATTAT----TCC-----G---C-----T--------C--A-----AGGCC-T--T----A-G-A---A-CG-C-GGGAAG--A----AGAA-ATA-A-TTTTG-G-.........   2182
VERVET.KE.VER155  GA---A......CAATTAAGG---TC-----G---------AG----A--------G--AGGGT-A--C----A-G-A------A-TG--GAAG--C--T-GGAA-ACA-G-T-TTA-G-.........   2684
VERVET.DE.VERAGM3 GA---A......CAATTAT------C---------------A--------C--------GGGAC----T----A-G-----AG-G-T-GGGAAG--AG-T-GGAA-ACA-A-TTTTG-G-.........   2187
SABAEUS.SN.SAB1C  GAA--A...........GG--AC-AT----------G---A--T--A-----A-----A-G---T--T----A-GCA--CC-CA-TCAGGAGG--A----AGAA-ATA-A-CTTG--A-.........   2866

SUN.GA.SIVSUN     TC-G--.......TCCATAGA---TCAC-----G--C-GGG-AG-G--T--T--A--G--ACAA--T-G-----A-G-A----CAG-TG-CT-TG-------TA-G-ATA-A-C-----C-.........   2819
LHOEST.KE.LHOEST  TCAG-A......AGAATAGA----CCC-AT-CT----G-G-AG-------C--A--G--GCAA--T-GG----A-G-A---AGAG-TG-CT-T--------AGCA---A-A-CTA-AG--.........   2814
LHOEST.CD.447     TCAG-T......AGAATAGAG---CCT-AT-C---T-G-G-AG----------A--G--ACAA--T-G-----A-G-A--CA-GG-TG-CT-T--------AGCA---A-A-C-A-AG--.........   1730
LHOEST.CD.485     TCAG-T......AAAATAGAG---CCT-AT-C---T-G-G-AG----------A--G--ACAA--T-G-----A-G-A--CA-GG-TG-CT-T--------AGCA--GA-G-C-A-AG--.........   1730
LHOEST.CD.524     TCAG-T......AGAATAGA----CCT-AT-CT----G-G-AG-------T--A--G--ACAA--T-------A-G-A-TCAGGG-TG-TT-T---C----AGCA--GA-G-CCA-AG--.........   1727
MANDRIL.GA.MNDGB1 -A-A-A......AAATTAAA-----AT---C-G--------T-----------A--T--AAACT-A-G-----A----------A-TG--TATG-------AA-A--GA-GGG-A-AT-T.........   2149

COLOBUS.CM.CGU1   GA-G-T......CAATTAA----CGCT----C-G--GT-CAA--TCA------A--G--AGCA----G-----A-------AGAG-T-A-A-TC--A-G--GG-A---A-AG--TAC-G-.........   2243
SYKES.KE.173      GA-A-T......CAATTA-ACCA-CC----TC---------AG----A---T-A-----AAAC--T-C---T--G--A---AGAA-T--TAGAT-CACTG-TGT-AA--CCTC--GA-A-AGAAAGCAAG   2660
RCM.NG.NGM        GA---A...........GG--A--AT-----G--------T-----A-----A--G--A-A---A--T--T-A---A-T-A-C--TTGTAAA---ACT--AGCA--GA-G-CTA-ACAT.........   2817
MANDRIL.GA.14CG   GA---A...........A------AT--------CC----T-----A----------AGCA--T-GG--G-A-------TTCA-TTGTAAA---AC-G-AGCA---A-A-GTA--CAT.........   2396
MANDRIL.14CG Pol  _D__L__.__.__.__.__T__G__N__W__K__P__Q__I__I__G__G__I__G__G__A__I__R__V__K__Q__Y__F__N__C__K__I__T__V__A__G__K__S__T__H__.__.__.__   763
```

HIV-1/HIV-2/SIV
complete genomes

```
                                                                                     Pol p10 protease end \/ Pol p51 and p66 RT and RT-RNaseH start
H1B.FR.HXB2         .....GGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTTCCCATTAGCCCTATTGAGACTGTA......CCAGTAAAATTAAA   2587
H1B.FR.HXB2 Pol    .__.__G__T__V__L__V__G__P__T__P__V__N__I__I__G__R__N__L__L__T__Q__I__G__C__T__L__N__F__P__I__S__P__I__E__T__V__.__.__P__V__K__L__K    p51

CPZ.US.CPZUS        .....---T----T-----G-----A-----A--A--T--T-----C--G---A-T--A--A--A-----A--T-------------T--C---T-----A-A------......-------G------   2647
CPZ.GA.CPZGAB       .....--G----------------A--------A--T--------------A-T-----A--AT-G-----T---C--GTG--C--A-----TT-A-----A-----C......-----C-----G--   2648
CPZ.CM.CAM3         .....--C--G-----G--G-----A-----A--G--T--------C------A-A-----A--A-----C--T-------------T--A--T--C-----A--A--G......--------------   1998
CPZ.CM.CAM5         .....--C----------------A-----A--A--T--------C------A-A--A--A--A--------T---C----------T--A--T--CG-------A--G......--------------   2289
CPZ.CD.CPZANT       .....-CA------C-GT-G-----A-AC--A--A-----T--A--T------G-T--ATG--TTT-A--A--T-------------A-----TAAAG---A--A--G......--C--T--------   2026

H1A.UG.U455         .....----------G------------------------------G---A------------------T-------------A-----T--------A------......--------------   2033
H1C.ET.ETH2220      .....---------C---------------------------C-----CA----------C----AC----A-----C-----A-----T--C-----A------......--------------   1979
H1D.ZR.84ZR085      .....-------------------C---------------------T-------C-------C----------C-----A-----T--------A------......--------------   2109
H1F.BE.VI850        .....--------G-------------G-----------------A-------------------T--------------AG----T--------A------......-----------G--   1924
H1G.SE.SE6165       .....--G--GA------------------A-------------G------A-------------------------------A-----T--------A------......--------------   1984
H1H.CF.90CF056      .....------------------------------T-----------G---A-A--------A-----------C-----------A-----T--------A------......--------------   1934
H1J.SE.SE92809      .....--------G---A----------------------------G--CA-----------C-------T-------------A-----T--------A------......--------------   1901
H1K.CM.MP535C       .....----------------------------------------C----C--------------------T-------------A-----T--------A------......--------------   1783
H101_AE.TH.CM240    .....----------------------------------------C-----A-------------------T----------C--A-----T-------C-------......-------C------   2161
H1N.CM.YBF30        .....-------------------A--------T--T--T--A-----------T--A--C--------C--T-------------A--A--T--------A------......--------------   2181
H1O.CM.MVP5180      .....--A--------G--G--------T-----T--T--TC----G-----CA-A-----AGGAT-A--A--T--AC-------C--T--A--T--C--A-CCC-A--G......---------GC----   2617
H1O.BE.ANT70        .....--A--------G--G--------T-----T--T--T-----------A-A-----AGGAT-A-----T--AC----C--C--T--A-----C--A-CCC-A--G......---------C----   2641

MAC.US.SMM251       .....--G---A-CA-GAC---GGAC--T--GA-T-----TT----T--G---T--C-A--AGCTC-G--GATGT--C-----C-------AGCTAAGG-A---C-----......AA---C-CC-----   2865
MAC.US.SMM239       .....--G---A-CA-GAC---GGAC--C--GA-T-----TT----T------T--C-A--AGCTC-G--GATGT--C-------------AGCTAAAG-A---C-----......AA---CGCC-----   2889
MAC.US.SMM239 Pol   .__.__G__T__I__M__G__G__D__T__P__I__N__I__F__G__R__N__L__L__T__G__L__G__M__S__L__N__F__P__I__A__K__V__E__P__V__.__.__K__V__G__L__K    p51

STM.US.STM          .....--A--TA-TA-GACT---GAC--C--CA-T-----TT----C--G--C--A-----AGCTT-G--GATGT-A-----------AG-AGCTAAAG-A--AC-A---......AAG--G-C------   2536
SMM.US.SMM9         .....--A---A-TA-GAC----GA---C--AA-T-----TT----C---------C-A--AGCT--G--CATGT--C-----C-C-----AGCTAAGG-G--RC--A--......AA-----C-C----   2362

H2A.GW.ALI          .....-CC--CA--A-GAC---TGAC--C--AA----T--TT----C-----CA-T-----AGCCT-A--CATGT-A-----CC-A--AG--GC-AAG--A---C-AA--......GAG----G------   2965
H2A.SN.ST           .....-CC--CA--A-GAC---TGA---C--AA-------TT----C-----CA-TC----AGCCT-A--CATGT-A------C-A--AG-CGC-AAG--A--AC-AA--......AA-A---TGC-G--   2417
H2A.DE.BEN          .....-CC--CA--A-GAC----GA---C--AA-------CT----C------A-TC----AGCCT-A--CATGT-A--------A--AG--GC-AAG--A---C-AA--......AA-----C---G--   2972
H2B.GH.D205         .....-CA--TA--A-GAC----GA---C--AA-A-----TT----C------A-T--A-A-ACCT-G--CATG-----------C--AG-GGCAAAGG-A--AC-A---......AA---TG-G-----   2948
H2B.CI.EHO          .....-CA------A-GAC---GGAC--C--AA-A-----TT----C------A-T--A-A-AGCT-A--CATG---C-------C--AG-AGCAAGG--A--AC-A---......AA---CC-G-----   2943
H2G.CI.ABT96        .....-C---T--TA-GAC---GGA---C---A-A--T--CT-------G--CA-T--AGT-A--C-A--CATGT--C----------A---GCAAAGG-A---C-A---......AA----GGC----   2299

TANTALUS.UG.TAN1    .....----AT---C--A-T-----A-----AA-A--T--TG-A---------A-TC-AG-A-CTT-A--AGCA-GA----TAA-GGGACAGCTAAA-GAGA-A-TACCCATAACCAAG---C----G--   2789
GRIVET.ET.GRI677    .....--CT--A-TC--A-----AG------AA----T--T--A---------A-T--AG-----GCA--CATG-AA---GT-A-GGGAG--CTAT---G-C---T--AGGAAACAAA----C--C----   2590
VERVET.KE.VER9063   .....--C-----TC-TT-----G-A--C---A----T---------------T-AC-AG-A-CAGCA--AGCT-GAC--GTGA-GGGACAATTGT-AGAAC-A-T-CCCATCACC-----C---C-G--   2817
VERVET.KE.VERTYO    .....--A---A----GT-----G-A--T--CA-T--T-----A--T------T--C--G-C-C-GCA-TGCC-CGG---GTAA-GGGACAATTAT-AGAAA-A-T-CCTGTCACA--T--C-----G--   2307
VERVET.KE.VER155    .....--A-----CC-CA-----G-A--T--CA----T--C--A--------CT-T--AG-C---GCA--AGC--AA---GTGA-GGGGCAATTGT-GCAGACA-TACC-ATCACC--G---CGC-----   2809
VERVET.DE.VERAGM3   .....--G--CA----GA-----AGC--T--CA-A------------------A-A--AG-A-CAGCA--AGC--AA---GTAA-GGGTCAACTGT-AGAAC-A-T-CCCATTACC--T--G--------   2312
SABAEUS.SN.SAB1C    .....-C---CA-T--G------GAA-----A--T--T--T--A---------G-C--AG-A--AT-A--AGTT---C-------GA-ACAA--AGAG--A--AC--A--......AA---CC-C-----   2985

SUN.GA.SIVSUN       .....--A-----G--G--G--G--A-----A--AG-T--C--A--G------A-TC-TT-AGCCT-A--GG-A-GA---GT---AG---C-CTTAG-GAAA---TCCC-ATTACAAA---T--G-----   2944
LHOEST.KE.LHOEST    .....--G-----C---T----------C--A--AG-T--TG-A--G--G---A-T--A---GCA--G--GGC--AA--G-T---AG-TCAGTTAAG-GA-A-A-T-CCTATTACAAAG---TCT-----   2939
LHOEST.CD.447       .....--G---------T-G----------A--AG----TG-A---------A-T--A---A-TG-G--AGCA-GA---GTAC-AT-ACAATTAAG-GACA-A-TACCCATCACAAAG--C-----G--   1855
LHOEST.CD.485       .....--G------C--T------------A--AG----TG-A---------A-T--A---AGTG-G--AGCA-AA---GTAC-AG-ACAATTAAG-GACA-A-TACCTATCACAAAG--C-G---G--   1855
LHOEST.CD.524       .....--G-----T--GC--------A--C--A--AG-T---G-A--------CA-T--A-AC-CTC-A--GGCT-GA---GTG--AG-ACAATTAAG-GAGA-A-TACCTATAACAAAG--GC-T--G--   1852
MANDRIL.GA.MNDGB1   .....--G--------GA----------T--AA-AG-T-----A--G-----CA-AA--GAAA-AT-A--AG-A-AA----TA--GG-ACAATTGT--GA-A-A-TACC-ATAACAAA---G--------   2274

COLOBUS.CM.CGU1     .....--AGATA-C---A------AC--C--CA-A--T---G-A--T------A----A-AAG-AGGGAAA-TAGTAG--GC-CAATTATCAGATAGG--CCCAGTCACC......AAG--G-CG-----   2362
SYKES.KE.173        TAGAG--G--TC-T-----G--T-----G--A--A--T---T-A-----G---A-AC-T--CA-AT-G--AGT--AA---GTAA-GGTTCAA-CAG-AT-A---C-A--C......AA-----GTC-G--   2784
RCM.NG.NGM          .....-C-CAT-----G--------A-----A--A--T--------C--G---G------GCA-AT-A--AGC-T--C----------A--A--TAAAGCA--A--AA--......AAG---G-------   2936
MANDRIL.GA.14CG     .....-C-T----GC----G--C-----C-----A--T--T--A--T------G-TC-T-AGA--T-A--A--T---C-G---------T-----TAAG--A-----A---......AA-----C-C----   2515
MANDRIL.14CG Pol    .__.__A__S__V__L__V__G__P__T__P__V__N__I__I__G__R__N__V__L__K__K__L__G__C__T__L__N__F__P__I__S__K__I__E__T__V__.__.__K__V__T__L__K    p51
```

```
H1B.FR.HXB2        GCCAGGAATGGATGGCCCAAAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACAGAGATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATAC   2717
H1B.FR.HXB2 Pol    __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__T__E__M__E__K__E__G__K__I__S__K__I__G__P__E__N__P__Y_   p51

CPZ.US.CPZUS       A-----------------T-G---A--G----------------C---------T--G------AC---------CCA---A--------A-----A--G---A---G-G-A-----A--------T---   2777
CPZ.GA.CPZGAB      A---------------T-----G--A--G--------C--AT---C---------T-----C---AC----------CA---A--------------------A--G-----A--A---------------   2778
CPZ.CM.CAM3        -----------------T-G---A-----------T--A----C---------T---------AC----------CAG--A--------A-----A--G-------G---A-----C-----------T   2128
CPZ.CM.CAM5        --------------------G---A-----------T--A----C---------C--G--C---AC---------CCAG--A-----------G--A-----------G--A--C--A--G--C--T--T   2419
CPZ.CD.CPZANT      AGA------------A----G---A-----------GC-CT--A----G------G-G--CC--AA----------GATA--T-A---GCA---AAT--G-----T-G---A-----A--T--C------   2156

H1A.UG.U455        A----A---------------G--------------------------G---------------AC-----------AT-----------------A-----------------------------   2163
H1C.ET.ETH2220     ------------------------C--------------------------G------------AC--C-------GA---A-----GC-------A----------G---------------C-----T   2109
H1D.ZR.84ZR085     ------------------------------------G--------------------------AC---------------T--------------A----------G----------------------   2239
H1F.BE.VI850       --------------------G----------------------------G------------AC------A---CT---A--------A-----A-------------------------------   2054
H1G.SE.SE6165      --------------------GG-----------------------------------------AC-----------A---A------G-------A--------------------------------   2114
H1H.CF.90CF056     ---------------------G-----------------------------------------ACG-----------------------A-----A-----C----G---A--------G---------   2064
H1J.SE.SE92809     ----------------------A-----------------------------------------AC-C--------G----A------G-------A----------G-G-------------------T   2031
H1K.CM.MP535C      --------------------G-----------------A-----------------------AC---------------A--------A-----A----------------------G--------T   1913
H101_AE.TH.CM240   ---------------A-----G--------G-------------------------------AC-----------A----------G-------A-----C---------------------------   2291
H1N.CM.YBF30       A--------------------G--A-----------T------AC----------G-G------AG---------------A--------------A--------T-G---A--------G--------T   2311
H1O.CM.MVP5180     A--------------A--------A-----------CC-AT-TAG---G------G-----C--ACT-C---A---CA---A------C-------A-----C----G---A--A----------T--T   2747
H1O.BE.ANT70       A--------------A--------A-----------CC-AT-TA-----------G----C--GAC--C---A---CAG--A------C-A-----A----------G---A--A----------T--T   2771

MAC.US.SMM251      --------A--T---A------T-G--G--G--------AT--A-------G---GTT------AG------C---GA-A-------------T--TC-GT-GGAGG--GC-CCC--GACC---------   2995
MAC.US.SMM239      --------A------A------T-G--G--G--------AT--A-------G---GTT------AG------C---GA-A-------------T--TC-GT-GGAGG--GC-CCC--GACC---------   3019
MAC.US.SMM239 Pol  __P__G__K__D__G__P__K__L__K__Q__W__P__L__S__K__E__K__I__V__A__L__R__E__I__C__E__K__M__E__K__D__G__Q__L__E__E__A__P__P__T__N__P__Y_   p51

STM.US.STM         A-------AA-----A------A-A--------------AT--A-------G---G-G---C--AA---G--C---GA-A-A-----G-----T---C-GC-AGA-G--GC-CCT--AAC------T---   2666
SMM.US.SMM9        A-------AA-----A------T-A-G---G-----GC-AT-GA-------G----TT------AG------C---GA-A-A-----G--A--T--CC-GT-AGAGGR-GC-CCT--AACC--------T   2492

H2A.GW.ALI         --------AA--C--G------T-A-G---------C--A---A-----------G-G---C--AA------C---GA-A-A-C-----GA-----CC--T-AGAGG-GGCACCT--AACT--C--T--T   3095
H2A.SN.ST          --------A------A------C-G-G---------C--A---A-----------G-G---C--AA---G--C---GAGA-A-------GA--G--CC-GC-AGAGG-GGCACCT--AACT-----T--T   2547
H2A.DE.BEN         ------G-AA-----A----GGC-G-----------CC-A---A----G------G-----C--AA---G--C---GA-A-A--------A--G--CC-GC-AGA-G-GGCACCT--AACT-----T--T   3102
H2B.GH.D205        A--T----AA-----G-----GA-C-G---------TC-AT-CAGG-----G---CT---CC-CAA------C---GA-A-A-----------G--AC--T-AGA-G--GCACCC---ACT---------   3078
H2B.CI.EHO         ---T-A--AA-----G------A-C-G---------CC-AT-CA----G------CT---CC-CAA------C---GA-A-A--------A--G--AC-GT-AGA-G-GGCGCCT---ACT---------   3073
H2G.CI.ABT96       A--T--C--------G------A-C-GG--------CC-AT-TA-G---------C-----C-CAGG---------GACA-A------C--------C--T-GGA-G--GCACCC--CACC-----G--T   2429

TANTALUS.UG.TAN1   A--T--G---TCA--A--G---T-A-----------CC-AT-------------------CC-GAC------AAC-CAG-CTC-A-----------A--------T--GG----AGGA--G-----T---   2919
GRIVET.ET.GRI677   AGA---G-AA-----A--T---T-G-----------C--AT--AG---------TG----T---AC------A--C-A-C-A------G----G--A---T-A--T-GG--A--AGGA--------T--T   2720
VERVET.KE.VER9063  AGAT---GCAAGG--A---C-TA-A-----------TC-AT--A-G--G--------G--C---AC---G--A---GA-C--T-A-----A-----------CAGC--G--A--AGGA------G-T--T   2947
VERVET.KE.VERTYO   -GA---GGCTCGG--A--CTGT--A-G---------TC-CT-TA----G--G--TG----T---CAG-----A---T-CC-AT-A--GC-------A-----CAGT-G-G-A--AGGA------G-----   2437
VERVET.KE.VER155   -GA---GGCCAGA--A---CG-T-G--G---------C-CT-TA-----------T---CC-GCA---------C-A-ACAT-A--GG-A-----A---T-AAGC-GGG-A---GGA--C---G-----   2939
VERVET.DE.VERAGM3  AGA---GGCTAGA--A--TTTCT-A-----------CC-CT-CA----------------C---CAG-----A---GACC-AT-A--G--A-----A------AGC--G--A--AGGA--G---G-----   2442
SABAEUS.SN.SAB1C   -------CAA-----G----GGA-A-GG--------T---T--A-----------G-G--C---AAG-CC------GA---CT-A-----AC----AC-TT-GGA--G---A-----A-----------T   3115

SUN.GA.SIVSUN      AGA---GGCC-C---A--------G-----------CC--T--AGG--G------G---GCC--CAGA--------GATAGAT-----GCA-----A--G---AGC-GGGCA-AC--A-GG--C--T--T   3074
LHOEST.KE.LHOEST   A--T--GTGT-----A----G---A--G--G-----TC-AT--A----G------G---GCC--CA--CT------GATAG-T-A-----A-----A-----A--TCC-G-G-AT--A-GG--C------   3069
LHOEST.CD.447      ------CTGT-----T--C-G---A-----G-----T---T-TA-------G---G---GCC--CA--C-------GATAGAT-A-----A-----------A------GCA-ATG-A-GG--C-----T   1985
LHOEST.CD.485      ------CTGT-----T--C-G---A-----G-----T---T--A-------G---G-G-GCC--CA--C-------GATAGAT-A-----A-----------A------GCA-ATG-A-G---C-----T   1985
LHOEST.CD.524      ------GTGT-----G----G---A-----G------C-AT-CA-----------G---G----CA--C-------GATAGAT-A-----------------A--CCC-G---AT--A-G------T---   1982
MANDRIL.GA.MNDGB1  A------G-A-----A--C-G-A-A-----------T--A-GTA----G------GTT-GTC-TCAGA----A---GATAGAT-A--GG-------A------AGT-GGG-A-AT--A-G----AAT---   2404

COLOBUS.CM.CGU1    AGAG--C--------G--------A--G--G-----C--A-GCA-----------G---GT---CA-A-G--A---GA-AGAT-----GCA--G--A-----AGA----GCA-AATTA-GG--------T   2492
SYKES.KE.173       A----AC-AA--ATTG--CCGGC-A--------------AT-G-T---G---T--G-G--CC-CAAG-C---AGTAGA---C---CTC----C---AC--T-GGAG---GC-A-T---ACC-----T--T   2914
RCM.NG.NGM         -------CAA-----G--T-G---A-----------T----GTA----G-----TG----T---AC---G-------AT-CC-----------G--A--------T-G---A--A--A--------T---   3066
MANDRIL.GA.14CG    A--------------A----G-A-------G------C--T-TA-------G--TTT---C---CAGAC---A---GAT--A--------------C-----C--T-G---A--T--A-----------T   2645
MANDRIL.14CG Pol   __P__G__M__D__G__P__R__I__K__Q__W__P__L__S__K__E__K__I__L__A__L__Q__T__I__C__D__E__M__E__K__E__G__K__I__S__R__I__G__P__E__N__P__Y_   p51
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2        AATACTCCAGTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAACTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCACATCCCGCAGGGT   2847
H1B.FR.HXB2  Pol   _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__R__T__Q__D__F__W__E__V__Q__L__G__I__P__H__P__A__G__    p51

CPZ.US.CPZUS       -----A---A-T-----T--C--A--G-----T--C----------------------C--T------T-A-----A--G--A--------------------G-----C--------C--A-----A-   2907
CPZ.GA.CPZGAB      --------TA-T-----A--C--A--G-------------------------------C-----------A-----------A--------T--------G--G-----C------T--C--A-------   2908
CPZ.CM.CAM3        -----A---A-T-----T-----A--G-----T-----A------------C-G-----------G---T-A-----A-----A--G-----T--------C--G-----C--------C--G-----T-   2258
CPZ.CM.CAM5        --C--A---A-T-----T-----A--G-----------A------------C-G--------T------T-A--C--A-----A--G--A--------G--C--------C--------C--A-----AC   2549
CPZ.CD.CPZANT      -----A---A-T-----A-----A--G-------C-T-A-----------GC----T------------T-A-----A-----A-----T--T-----GA-A-----------T--T-----A--C--A-   2286

H1A.UG.U455        ----------------T-----------G-----C-----------G----------------------C-----A--------G--------------------------C--G---A-A--G--TC   2293
H1C.ET.ETH2220     ----------------------A--G--G-----------G----------------------G-----C-----A--------------T-----------------G--------C--A-------   2239
H1D.ZR.84ZR085     ---------A-------------------------------G-----------------------------------------------------------G----------T-----A-   2369
H1F.BE.VI850       ----------------------------------G---------G----------------A------------A-----------T--------G----------------------T-------   2184
H1G.SE.SE6165      --C------A-------------------G----------------------G--------------G--C-----A--------------------G--C--------------T-----T-------   2244
H1H.CF.90CF056     -GC------A-------------A--G--G--T----------------------G--------------C-----A--------------------------G--------------C--A-------   2194
H1J.SE.SE92809     --C-------------------------------------------------------T--------C-----A--------------------------G----------------A-------   2161
H1K.CM.MP535C      --C--------G-----------------------------------------------T--------C-----A-----------------------------------------A-----A-   2043
H101_AE.TH.CM240   ----------------T-----------G-----C--C--------G--------------------G--C-----A--------G-----T----------------------G-----A-----T-   2421
H1N.CM.YBF30       --C------A-T-----T-----A--G-----T--C----------------------------G---T-A-----A--G--C-----T--T--------G--GC-------T--------A-----A-   2441
H1O.CM.MVP5180     -----A--TA-T-----T-----A--G-----T--C-----G-----------G-----C---------T-A-----A-----A-----T--------G--G--------T--T--------A-GG--T-   2877
H1O.BE.ANT70       -----A--TA-C-----T-----A--G-----TG------------------G--------T--G---T-A-----------A-----G--------G--A--GC----T--C--------G-GG--T-   2901

MAC.US.SMM251      --C--C--CAC--------------------T-AG-AC----------TGC-GA-------T--G-----A----G-GTC-----G-----TACA-----C----------------C--T-----AC   3125
MAC.US.SMM239      --C--C--CAC------T----------G--T-AG-AC----------TGC-GA-------T--G-----A----G-GTC-----G-----TAC------C----------------C--T-----AC   3149
MAC.US.SMM239 Pol  _N__T__P__G__F__A__I__K__K__K__D__K__N__K__W__R__M__L__I__D__F__R__E__L__N__R__V__T__Q__D__F__T__E__V__Q__L__G__I__P__H__P__A__G__    p51

STM.US.STM         -----C--CACT-----T-----------G----AA-AC--G-----G-TGC-GA----------------A-----AGTG--A-----T--TACA---A----GC----T--------C--T-----AC   2796
SMM.US.SMM9        --C--C--CAC------T------R-GR-----G-A-A---------G-TGC--A-------T------T-A--C---GTC--------T--TACA-----A--G-----G--------C--TR----AC   2622

H2A.GW.ALI         -----C--CAC------A--------G--G----AA-AC----------TGC--A-------T------T-A--C---GT---------T---ACA--GA----G-----G--T--------A-----A-   3225
H2A.SN.ST          -----C--CAC------A--C--------G----AA-AC----------TGC--A-------T--------A--C---GT-------------ACA---A-C--G-------T-----C--A-----AC   2677
H2A.DE.BEN         -----C--CAC------A--T--------G----AG-AC--------G-TGC-GA-------T--------A------GTG--------T---ACA---A----GC-------T-----C--G-----AC   3232
H2B.GH.D205        --C--A--CACC--------------G-----T-AG-AC----------TGC--A----C--T------GT-A--C---GT---C--G------ACA-----CA-C-.G--T--T--C--C--G-----TG   3207
H2B.CI.EHO         ---T-G--CACC--C--------A--G-------AA-AC--------G-TGC--A----------------A--C---GT---C-----A--TACA--G--C--GC-G--T--T--T--C--A-----AC   3203
H2G.CI.ABT96       --C--A--CACC---A-T--C-----G--G----AA-AC----------TG---A----C--------------C-G-GT---C--G-----TACT-----C--G--G--C-----T--C--T-----T-   2559

TANTALUS.UG.TAN1   --------C--G---TGT--CCG-----------AA--AC-G-----G-TG---A-------T----CT--C-----AGC---A--G-------TT--G--A--------T--T--C-----A-----A-   3049
GRIVET.ET.GRI677   -----A-----G-----------------G--T-AA--AC---------TGC-T-----------G-----A--C--AGCT------------T-TT--G-----GC-G-----T--T--C--A--G--CC   2850
VERVET.KE.VER9063  -----C---A-C---TGT--C-GA--G--G----AA---C-------G-TG--------C---------T-A-----AGT---A----------TT-----------G--------T--C--G-------   3077
VERVET.KE.VERTYO   -----C---A-----TG---------G--G----AAT-CC-G-----G-TGC-------C--T-----GT-A------GC---C-----T----TT-----G--------G-----C--C--A-----A-   2567
VERVET.KE.VER155   -----A--------CTGT----G-----------AAT-AC-G-------TGC-G----------G-----C--C--AGCT--A----------TT-----C--------T-----C-----A-------   3069
VERVET.DE.VERAGM3  --C--------G---TG----------------AGT-AC---------TG-----------T--G-----A--C--AGC---A-----T--T-TC-----A--G-----C-----T-----AT------   2572
SABAEUS.SN.SAB1C   --C--C-----C-----A----G---G-----T-AA---C-G-------T-C-CA-G--C-----GC-GT-A-----A--T-----------TCA------C--G-----G-----C--C--A-----C-   3245

SUN.GA.SIVSUN      --C--A---A-T-----T--------G-------AA-A-G----------G---A-------T----T---A---GGT-TG-----------TCAT--GT-A--G-----G--T--T-----A-----A-   3204
LHOEST.KE.LHOEST   ---------A-------T--C-----G-------AA-A-G----------G---A-T---------A-G--C---G-ATT------------TCAT--GC-A--------T--T-----C--A-----TA   3199
LHOEST.CD.447      --C--A---A----------------G-------AA-A-G-G------------A-T-----T---C----A---GCTTT---A--------TCAT-----A--------C--T-----C--T-----AA   2115
LHOEST.CD.485      --C--A---A-C--------------G--G----AA-A-G-G---------C-GA-T-----T---C----A---GCTTT---A------C-TCAT-----A--------C--T--------T-----AA   2115
LHOEST.CD.524      --C--C---A-C-----T-----------G----AA-A-G----------G--GA-----------A--T-A---G--TT---A--------TCAT--GC-A--G--------------C--A-----AA   2112
MANDRIL.GA.MNDGB1  -----A--TA-C--------------G--G--T-AA-A-G--------------A----C--T------T-A--C---TT---A--G--T--TCAT---T-A--G-----T-----T--C--A-----AA   2534

COLOBUS.CM.CGU1    -----A---A-----TG---T-G---------T-AA-A-G-------G------A----------G--G--A--C--------A---------AT------C--------G--------C--A-G---TC   2622
SYKES.KE.173       -----C----------TA--T-G------G----AA-AG--------G-TGC-CA-T-----T---A--T-A---G--GCC--A-----T--T-TC--G--A-----G--------C--C--A-G---AC   3044
RCM.NG.NGM         ---------A-----TGT--C--A--G-----TTCC--A-----------GC-----------------T-G-----A-----A-----T----TT--G--A--GC-G-----C--G-----A-G---A-   3196
MANDRIL.GA.14CG    -----A---------TGT-----A--G-----TG-A--C-GC--------------------T--GC--T-A----GAGTG-----G--T----TT--G--G--GC-------C--------G-G---C-   2775
MANDRIL.14CG Pol   _N__T__P__V__F__G__I__K__K__K__D__G__T__S__W__R__K__L__V__D__F__R__Q__L__N__R__V__T__Q__D__F__F__E__V__Q__L__G__I__P__H__P__G__G__    p51
```

```
H1B.FR.HXB2        TAAAAAAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATGAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAG   2977
H1B.FR.HXB2  Pol   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__K__Y__T__A__F__T__I__P__S__I__N__N__E__T__P__G__I__R   p51

CPZ.US.CPZUS       ----G-------G------------C--A-----A--G-----C--------TTG---AC-G---A-G--A--T--A--A--C--A-----------C------G-------------C-----A-----   3037
CPZ.GA.CPZGAB      -------------------G------T-A-----A--A-----C--C--C--TTG----C-G---A----T--------A---------------A--T--C--------T--------C-----AG----   3038
CPZ.CM.CAM3        -------------G----------G--A--------A-----T--------CTGC--A---------A-T--T-----A-----A--------T-----C---G----T-----A--------A-----   2388
CPZ.CM.CAM5        ----G--A--G--G--------------A--------A-----T-----C--CTGC--A---------A-T--T--A--A--C--A--T-----T-----C---G----T--------------AG----   2679
CPZ.CD.CPZANT      ----GC-A--G--------G------T----------A-----C-----C--CA-A---------C-G-----T--A--A-----A--T--C--A--T--A--CG-G-----------C-----A--A--   2416

H1A.UG.U455        ---------------------------A--------G--C-----------------T---------AG---T--A----------G--C-------------------------------AG-C--   2423
H1C.ET.ETH2220     -------------------G--------A--------G-----------C--------T----------GT-----A--A-----------C-------------C---------A--------------   2369
H1D.ZR.84ZR085     ----G--------G---A-------------------C---------------A--------TG---------T--------C-------------------------------------------   2499
H1F.BE.VI850       -------------G----------------------G---------------------------A-G--T----A------C--------C------------G-C--------------T--A-----   2314
H1G.SE.SE6165      ---------------------------A--------G-----------------------------------T--A--------------C--T--------------T---------------G----   2374
H1H.CF.90CF056     -------------------T----------------G-----------------C--T------A----A-----A--------------C-------------------------------------   2324
H1J.SE.SE92809     -------------------C----------------G-----------------C--T---T-------T--------A-----------C--T--------------------------------C--   2291
H1K.CM.MP535C      ------------------------------------G-----------------C--T------A------------------------C--T-----------T--------A---------G----   2173
H101_AE.TH.CM240   ---------------------------A--------A-------------------T---------AG---T--A--------------C-------------------------------A--C--   2551
H1N.CM.YBF30       ----GC-------------G-----TT-------A--A-----T---------TG------G--CA----T--T--A--------A--T--------------------------------T--T-----   2571
H1O.CM.MVP5180     ----GC-A-GGC----T--T-----CT-A-----A--A-----T-----C---TGC--T------CC------T--A--A--C-----C--C--T--T------G-G-----------C-----AG-A--   3007
H1O.BE.ANT70       -G--GC-A--GC----T--T-----CT-A-----A--A-----T-----C---TG---------CCC---T--T--A--A--------T--C--T--T------G-G--T--------C-----A--A--   3031

MAC.US.SMM251      --GC---A-GG---AGGA-T------------A-A--------------C--CA-A--TC----------A--T---C----C-----C-----TT----ATCAG----T----CAGAG-----A-AAC-   3255
MAC.US.SMM239      --GC---A-GG---AG-A-T------------A-A--------------C--CA-A--TC----------A--T---C----C-----C-----TT----ATCAG----T----CAGAG-----A-AAC-   3279
MAC.US.SMM239 Pol  L__G__K__R__K__R__I__T__V__L__D__I__G__D__A__Y__F__S__I__P__L__D__E__E__F__R__Q__Y__T__A__F__T__L__P__S__V__N__N__G__E__P__G__K__R   p51

STM.US.STM         --GC-----GG-G-AG-A-C------T-------A--G--------------CA-A---C-------G-G---T---C----------T--C---T----ATCAG---------CAGA------A-AA--   2926
SMM.US.SMM9        --GC-----G--G-AGGA-C-----GT-------A--------------C--CA-A--T-----------A------C-A--C-----C------T----ATCGG---------CAGA------A-AA--   2752

H2A.GW.ALI         --GCC-------G-AG-A-C--T--G--------A--G-----T--C-----CA-A--AC-GC----GAG---T--AC----------------TC----ATCAG---------CAGA------A-AA--   3355
H2A.SN.ST          --GCC-----G---CG-A-T--T--C--A-----A--G-----T--C-----CA-A--AC--C----G--T--T--AC----------------TC----ATCA----------CTGA------A-AA--   2807
H2A.DE.BEN         --GCC--A--G---AGGA-CT-TA--T-A-----A--G-----C--------CA-A--AC--C-------T--T---C----------------C----AGCAG--------AT-GA------A-AA--   3362
H2B.GH.D205        ..GC-G-A----GGAG-A--------A-A--------A-----C--C--CAGTA-C--AC-----CC-A-------AC-------A-----C---T-G--ATCAG---------CAGAG-----A-AG--   3335
H2B.CI.EHO         -GGC-TCA--G---AG-A----------A-----A--A-----C--C--CAGT--C--AC-----CC---------AC-A-----A--------TT-G--AGCAG----T----CAGA------A-AG--   3333
H2G.CI.ABT96       -GGC-G-A----GGAG-A--------------A-A--G-----C--C--CAGCA----TC------T---T------C----C--A--T--C--TT----ATCAG---------CAGA------A-AA--   2689

TANTALUS.UG.TAN1   -G----GA--G---CAGA-C--------A---A-A--A-----C----A-AGTA-A--A--G---CC---A------CCC-----A--T--------C--ACAAG----T----TAGGG--T--A-----   3179
GRIVET.ET.GRI677   -TC-G--A--G--GCA-A-C------A-A--CA-A--G-----C----A----A-A--A---TGCA-G--A-----A--A-----A-----------C--CTCAG----T---ACAGGG--------A--   2980
VERVET.KE.VER9063  -GC-G--A-TG---CAGA--------T-A---A-A--G-----T----A---CA-A--A------CC---G--T--A--A--C--A--------A--C---TCAG----T---C-AGG------TG----   3207
VERVET.KE.VERTYO   ---G-----TG-G-CAGA-------TT-A-----A--A--C--C----A---CA-A--A--G---CC-A-T--T-----A--------T-----T--T----CAG-G--T---C--GG---C--------   2697
VERVET.KE.VER155   ----G--A-TG--GCA-A----CA-TA-A--------G----------A-AGCA-A--AC-G---CCT--G--T--A--A--C--A--T--C-----C----CGG-----------GG------C--A--   3199
VERVET.DE.VERAGM3  -CG------TG-CGGA-A--------T-A--CA-A--G-----C----A----A-A--A-----CCC---G--T--A--------C--T--------T--ATCAG----T---C-AGGG-----T-C---   2702
SABAEUS.SN.SAB1C   --C-GC-A-GGG-GCAGA-T-----GT----CA-A--A-----C------AGCTGC--T------CC------TCA---A-----A-----C-----C--ATCAG-C--T---AG-GAG-----C--A--   3375

SUN.GA.SIVSUN      ----G---TGT---CAGA-T------G-A--------G--------C--C---A-A--A--G---CC---T-AT--A--------A-----C--A-----GTCAG----T---C--G-------A-AA--   3334
LHOEST.KE.LHOEST   -C--G--ATGC---CAGA-T------G-A--CA-A--A-----C-----CAGTA-C--T------CC-A-T-A---A--------A--------T------TC-C-C--T---C-AGAG-----A-AA--   3329
LHOEST.CD.447      ----G--ATGC--GCAG--------GG-A---A-A--A-----C------AGCA-----C-G---CC-----A---AC----C--A-----C--AG----ATC---T--T--CC--GA--------GA--   2245
LHOEST.CD.485      ----G--ATGC--GCA-A-------GG-A---A-A--A-----C------AGTA-----C-G---CC-----AT--AC----C--A-----C--AG----ATC---T--T--CC--GA--------GA--   2245
LHOEST.CD.524      ----G--ATGC---CA-A-T------G-A---A-A--A-----C-----CAGCA-A--T------CCT--T-A---AC-A--------G--C--T-----GTCAC--------C--GA------A-AA--   2242
MANDRIL.GA.MNDGB1  --------TGT---AG-A-------C--A---A-A--G-----C------AGTA-A--TC-G---CC---T-AT--ACCC--------C-----GG----ATCAG-T--T---C-AG-------A-AA--   2664

COLOBUS.CM.CGU1    -T-T-G-A--G-CCCATA--------T-A---A-A--A-----C-----C---A-------G---CCG--T-ATCAA-----C--A--T-----A--C--ATCAG-G------TT-G---------C---   2752
SYKES.KE.173       -----CG-C-----...C-----A-TA-A---T-AAAA----------AC-----C---C-G--CA----G--T---CCA--C--G--------AG-G--A--C-----T----CCT-C--G--AGAA--   3171
RCM.NG.NGM         -----C-ATGC-GGCA-A-T-----CT-A---A-T--G-----G----------TGC---C--T----GCCA--T--A-----C-----T--C--G----TAGC---T--T---C-AGG---T--C-----   3326
MANDRIL.GA.14CG    -----C-ATGTG--CA-A-C---A--T-----A-A--A-----C---------TG---T--G-----GAG---T--A-----C--------------T--ATCGG----T---C-AGGC-----A--C--   2905
MANDRIL.14CG  Pol  L__K__Q__C__E__Q__I__T__I__L__D__I__G__D__A__Y__F__S__C__P__L__D__E__S__F__R__K__Y__T__A__F__T__I__P__S__V__N__N__Q__G__P__G__I__R   p51
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2       ATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTTATCTATCAATACATGGATGAT   3107
H1B.FR.HXB2  Pol  __Y__Q__Y__N__V__L__P__Q__G__W__K__G__S__P__A__I__F__Q__S__S__M__T__K__I__L__E__P__F__R__K__Q__N__P__D__I__V__I__Y__Q__Y__M__D__D_   p51

CPZ.US.CPZUS      ---C-----T------T-A-----A-----------G-----------T-----G--C--T-----------TC----T---------------C-------TG-TA-A-----------T---------   3167
CPZ.GA.CPZGAB     ------A--T-----TT-G-----A-----------G--T---T-T--C--------C--T-----------TC----A--C--C---G--A-G-----T-----TAC---T--C--G-----------C   3168
CPZ.CM.CAM3       ------A--T------T-A-----A-----------G-----------T--T-----C-------T------TC--ACC--A--C---C-G--GC--------T--A-A--T--C--G--T---------   2518
CPZ.CM.CAM5       ---------T------T-A--R--A-----------G-----G-----C--T-----C--------------A-----T--A--C--------GC-------TC--A-A-----C-----T--------C   2809
CPZ.CD.CPZANT     ----TGT--T-----T--A-----A--C-----------------T--T--T---GCA--------C--G-------CA--A-----GG-TA-GT------CAG---AA--T-----------------   2546

H1A.UG.U455       G--------------------G---------------------T----------G---------------------------C------TC----C-----------------------------C   2553
H1C.ET.ETH2220    ------A--T-----C--C-----------------------C----------G---------C-CC-------------C-----GGCC-CC--C-----A-----------------T--------C   2499
H1D.ZR.84ZR085    ---------------A-----------------------------------------------T---------------C---------------------AG----------------------   2629
H1F.BE.VI850      G--C--------------------A-----------------------------T-----------------------C-------TGA----C-----------C-----C---------------   2444
H1G.SE.SE6165     -------------------------------------------------T--G-------------G------------C------GC-A-T--------A--G--G-----C---------------   2504
H1H.CF.90CF056    ---------T-------------------------------------------G----------------------C---C------G----------T--A--G-----T--C-------------   2454
H1J.SE.SE92809    ------------C-----A--------------------C-----------T--GT--------------------A-A---------G--AG---C-----A-----------C--G-----------C   2421
H1K.CM.MP535C     ------------------A-----A-----------------------------CA-----------------------C-------T-A----------A--G-----A--C---------------   2303
H101_AE.TH.CM240  ------------------G----------------------G-----------G--------------------------C-------T-A----------A--G----------------A------C   2681
H1N.CM.YBF30      ---C-----T--------G-----A--C--------G-----------T--T--G----CA-----------TC----A--A--C---G-GA--C-------G---A-C--T--C--G-----------C   2701
H1O.CM.MVP5180    ---C-----------C--C--G--A--G--------T--------C-----T--G---TCA--------G--TC----T--A---------AGC--C-----AG---AA--T-----G-----A-----C   3137
H1O.BE.ANT70      ---C-----------C--C--G--A-----------T--------T------------TCA-----------TC----T--A--C--G-G-G-C--C-----AT---AA--T-G---G-----------C   3161

MAC.US.SMM251     ---CATT--T--G--T--G--T-----------G--G--------C--C------TAC-CT----G-C-TG-GC----A--C--C--G--GGC---------TG-GACCT-AGTC--G--T--------C   3385
MAC.US.SMM239     ---CATT--T--G--T--G--T-----------G--G--------C--C------TAC-CT----G-C-TG-GC----A--C--C--G--GGC---------TG-GACCT-AGTC--G--T--------C   3409
MAC.US.SMM239 Pol __Y__I__Y__K__V__L__P__Q__G__W__K__G__S__P__A__I__F__Q__Y__T__M__R__H__V__L__E__P__F__R__K__G__N__P__D__V__G__L__V__Q__Y__M__D__D_   p51

STM.US.STM        ---CATC-----G--T--G-------------G--------------T--T---TA--CA----G---T--A--------A--C--G-G-GCC--C-----TG--AC-C-GATC---------------   3056
SMM.US.SMM9       ---CATC--T--G--AT-A--T--A--G-----------------T--T--T--RCA--CTR---G---TG-A-----A------------GC---------TG-GACCC-GATC--------------C   2882

H2A.GW.ALI        ----ATA--T--A--CT-A--G-----------G--------------T--T---CAC-CA----G-C-G-----------A--C-----GGC---C-AG--TG-CA--C--AT----------------   3485
H2A.SN.ST         ---CATA--T--A--CTCA-------------G--------------T--T--GTAC-CA----GGC-GG-------A--A--C------GC---C--G--T--CA--C--AT---G-----------   2937
H2A.DE.BEN        ----ATA--T--A--CT-G-----A-------G--------------T--T---TAC-CA----GGC--G-------A-----C------GC---C-----TG-CA--C--ATC--G------------   3492
H2B.GH.D205       ---CATT--T--A--T--A-----A--G-----G--G--C-AGT----T-GT---TACTC-----G---GG-A-----T----------GGCC--CAGC--TG-CA----AAT---G-----------C   3465
H2B.CI.EHO        -----TT-----A--C--A-------------G-----C--------T-----GTAC-C----G----GG-AC----C-----C------GCC--CAAT--TG-CAC---AATC--G-----------C   3463
H2G.CI.ABT96      ---CATC-----G--C--A--G--A-------G--C--------T--C------CA--CA----GG--GG-AC----C--C--C------GCC---GAG--TR--A----YATC-----T--------C   2819

TANTALUS.UG.TAN1  G-----A-T----TGT--A--T--A-----------------CA-C-----T--GCAC-CAGCACAGCG-G--C--A-AGAAA-AG-----A-GGT--A---TT---AA--AAT---G-----------C   3309
GRIVET.ET.GRI677  G-------T----TGT--G--T-----------------T--TA----T-----G-A--CGGCAG----C--T------GAGA-C-A--GG--C-C---T-GGT---AA--TGTC-----------C---   3110
VERVET.KE.VER9063 G-----A-T----TGC--------A--C--------G--T---A-T--T--T--G-A--CAGCAG-------T------AAAA-A-A-G-GG--TTAAA-CCAT-GACA--AGTG--G------------   3337
VERVET.KE.VERTYO  G-----A-T---CTGT--C--G--A--G-----------T--TA----C-------A--CAGCAG--TCC--T--G---GAGA-A-A--G-A-CTTG----CAC--ACC--TGTA---------------   2827
VERVET.KE.VER155  ------A-TT---TGC--A--G-----C-----G-----C--GA----T-------AC-CAGCAT-------TC----AGAAA-A-AG---G--TTAAA-C-GC-GACG--TGTC--G-----------C   3329
VERVET.DE.VERAGM3 --------T---CTGT--------A--------G-----C---A-T--T--T--G-AC-CAGCAG-TTCC--TC----AGAAA-A-A---GG-GTTAAA-CC-C--ACC--TGTG--------------C   2832
SABAEUS.SN.SAB1C  ---------T--G-----C--------G--------C-----CA----T--T--G-CA-CAGCC-AC-----TC-GC--GAA-----GC--A-G--C-----TG---A---A-----------------   3505

SUN.GA.SIVSUN     ----GT---T-----C--A--------C---------AGT--TTGC--C--T--GG-A-CGG-AG-G-GCC--------AAA-------G---GT-------AG-TCAAT-G--C---------------   3464
LHOEST.KE.LHOEST  ------------------G-----A--G-----G--GAGC--TTGT-----T--GG-A-CAG-AG--GG-C-TC-CTCAGAG-----G---TT------------GA-C--T--C-----T---------   3459
LHOEST.CD.447     ---------T-----A--A--C--A--------G--GAGT---TGC--T--T---GC--CAG-AG--GG-C--C--TCAGAA---------TT---C--T------A-CT-G-----------------   2375
LHOEST.CD.485     ---------T-----A--A--C--A--------G--GAGT---TGC--T--T---GC--CAG-AG--GG-C--C--TCAGAA---------TT---C--T------A-CC-G-----------------   2375
LHOEST.CD.524     ---------T------T-G-----A--------G--GAGT--TTGT--T--T---GAA-CAG-T-GTGG-C--C-T-CAGAG--C--G---TT-------------C-CC----C--G------------   2372
MANDRIL.GA.MNDGB1 ---CAT---T-----T-----T--A--G-----G---AGT---TGT--C--T---G-G-CAG-AG--TC-C-GC-G---GTA--------GA-CC-----ACAG--CAGT-A-----------------   2794

COLOBUS.CM.CGU1   ----GT------A-----A--G-----T--------G--T---A----C------G-A-CAG-AG---G----C----A--G--C--G--GAG-GGA-A-TTG...CAA--AGC---G------------   2879
SYKES.KE.173      ------A-TT-CA---T-A--------------------T--CA-T--C--------C-CA--A-ATC-G--AC--C----A--------GA-GT--T----TC-CAC-C-AATA--G-----------C   3301
RCM.NG.NGM        ------A--T------T-G--C-----------G-----T-----T-----T----CA-CAGCT-AC-----T-----A--A-----G---G-G--------TG----C--------G--T----------   3456
MANDRIL.GA.14CG   G--C-----T-----T--------------------C--T--G-------T---GCA-CGGCAGAC-----TC--A-AA-A--C-A-G-GG--T-C-----AG--T-A--T-----------------   3035
MANDRIL.14CG  Pol __Y__Q__Y__N__V__L__P__Q__G__W__K__G__S__P__A__I__F__Q__A__T__A__D__K__I__L__K__T__F__K__E__E__Y__P__E__V__L__I__Y__Q__Y__M__D__D_   p51
```

```
H1B.FR.HXB2        TTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTTGGA   3237
H1B.FR.HXB2  Pol   _L__Y__V__G__S__D__L__E__I__G__Q__H__R__T__K__I__E__E__L__R__Q__H__L__L__R__W__G__L__T__T__P__D__K__K__H__Q__K__E__P__P__F__L__W__   p51

CPZ.US.CPZUS       C-C--------G--A--TC--A-CT---AAA-------GGA---GG----ACT---C------T--T--C-T-CT------T-C--T--C--T--T--G--G------G-G-----A------T-A----   3297
CPZ.GA.CPZGAB      C-A-----G--G-----TC-T-----T-AT--A-------A---GG-G--A--A--A---------T--C-T-AA-----GT-C--A--C-----------------A--G--G--A--C--TT-A----   3298
CPZ.CM.CAM3        C-T-----G--------T--G---C--A-AG-A------GA---GG-----CTA--C-----G-----AC-T-CT-----GT----T--C--T--T-----G--------G--G--A--T--TT-G----   2648
CPZ.CM.CAM5        C-T---------------C-G--G-----AA-A------GA---GG----ATTA--C------------C-T-CT-----TT----T--C--T--------G-----A--G-----A--C--TT-G----   2939
CPZ.CD.CPZANT      C-C--------------TA-G-----TACTGCA------GA--TG-----AA----T---------T-ACA-GTC--------AGAG--T--T--------G--------------------T-AA----   2676

H1A.UG.U455        --------G--------T-------------A------G-------------AT-A---GCT-----A-----C------T-C-TT--C----------G--------------------T-------   2683
H1C.ET.ETH2220     -------------------------------A------G-CCC------A---T-A---G-------A--A-A-------T---------------G----------------------T-------   2629
H1D.ZR.84ZR085     ----------------T----------A---------G-----------A-AT-A---G-------------------G--------------------------------------T-------   2759
H1F.BE.VI850       -----------G-------------------------------------T-A---G-------AC----A------T----T--------------------------------------   2574
H1G.SE.SE6165      --A-----------------------------------G--------------T-A---G-------A----AA-----GT-------------T--------------------------------   2634
H1H.CF.90CF056     -------------------------------A------G--------------T-A---GCT---T------AA------T--------------------------------C-----T-------   2584
H1J.SE.SE92809     --------G-----------G-------AA----------G-------A----------GG-------A----A-------T-------------T-------------------------T-------   2551
H1K.CM.MP535C      --A--------G--------------------A-C-----------------A--A---G-------A----AA------T------------------G--------------C-----T-------   2433
H101_AE.TH.CM240   -----------------T-----------------C-------------------A---GCT-----A-----C------T----T--------------G--------G-------------------   2811
H1N.CM.YBF30       C-C-----G---------------C---CA--A------GAGGC-G----A--C--T---G-T-----T----A------CT----G--C--T-----------------G-----C--G-----C----   2831
H1O.CM.MVP5180     --A-----------A--T---CC-T-G-CAG-A-------AG-GGG-C--ATT---T--GG-----T-A-ATCA-------T----T--C--T--T-----G--------G--------C--TT-A----   3267
H1O.BE.ANT70       C-A-----------A--T---CCCC-GACAG-A-------A--GG--T--ATT---T---G----C--A-ATCA-------T-C--T--C--T--------G-----A--G--------C--T--A----   3291

MAC.US.SMM251      A-C-TAA---CTAG----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC----A-AGG---TCT-AAAT--CATA--GT-CT-T--C-----AG-G---TTC--A-----T--C-----T-AA----   3515
MAC.US.SMM239      A-C-TAA---CTAG----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC--TCA-AGG---TCT--AAT--CATA--GT--T-T--C-----AG-G---TTC--A-----T--C-----T-AA----   3539
MAC.US.SMM239 Pol  _I__L__I__A__S__D__R__T__D__L__E__H__D__R__V__V__L__Q__S__K__E__L__L__N__S__I__G__F__S__T__P__E__E__K__F__Q__K__D__P__P__F__Q__W__   p51

STM.US.STM         A-C-TGA---CCAG----AG-AC-GATCTAG-----GAC-GGGT-G-TTTAC--T---AGG---TC--AAAT-AT-TA---T--T----C--T--AG-G---TT------G--T--------T-AA----   3186
SMM.US.SMM9        A-C-TAA---CTAG---TAG-AC-GATTTAG-A---GAC-GGGT-G-TTTAC-AT-A-AGG---T----AAC--TATA---T--T----C-----AG-G--GTTC--A-----T--C-------AG----   3012

H2A.GW.ALI         A-C-TAA---CTAG----AGGAC-GATTTAG-A---GAC-AGGTGG-CCT-C--T-A-AGG---TC--AAATG-CCTA---T--T----C-----TG-G--GTTC--A-----C-------A-AAA----   3615
H2A.SN.ST          A-C-TGA---CCAGC---AGGAC-GATTTAG-A---GAC-G-GTGG-TCT-C----A-AGG---T---AAATG-CCT----T--T----C-----TG-G--GTTC--A-----C-------A--AA----   3067
H2A.DE.BEN         A-C-TAA---CTAG----AGGAC-GGTTTAG-----GAC-A-GTGG-CCT-C----A-A-G---T---AAATG-CCTA--GT--T-T--T-----TG-G--GTTC--A-----C--------T-AA----   3622
H2B.GH.D205        A-CCT-A---C-AG----AG-AGTGATCT-G----CGAC-GGGT-G-GTCCC-A--A-A-G-GTTAT-AAATGACAT----T-CT-T--C-----AG----GTTC--A-----C-----G---AAA----   3595
H2B.CI.EHO         A-TCTC--G-C-AG----AGGAGCGATCT-G-----GAC-GGGT-G-GTCTC-A--A-A-G-G-TAT-AAAT-ACAT----T-CT-T--T-----AG----GTTC--A-----C---------AAA----   3593
H2G.CI.ABT96       A-TCT---G-CCAG----AG-ACTGATTTAG-A---GATCA-GT-G-TTT-C----A-AGG---TCT-AAAT-AAACR--GT-CT-T--C--T--TG----GTTC--A-----C-------AT-AA----   2949

TANTALUS.UG.TAN1   A---GGA-------A--T--GCC-GAGAAAG-A---GATGA---GG----A---G-A---A-G-T----C-ATTAA-A--CT-CGAG--G-----T--G--GGTG------AG--------ATGA-----   3439
GRIVET.ET.GRI677   ----GGT-G-CG--A---CAT--TGAGACTAGA----ATCA-C-GG----CATAG-A---A-GATG---C-AGAAAAA--T--AGAA--C--------G---GTC--A-G------G----GGGAA----   3240
VERVET.KE.VER9063  ----GG-----G---C-AGA---TGA-TACACC---GAC-AGTT-G----AC----T---GTCA-G--AAAA-CA------T-AGAA--C-----------GGTG--A--GA-G--A----ATGAA----   3467
VERVET.KE.VERTYO   --A-GG-----T---C-AGA-A-TGA-CACACC---GAC-A-TT-G----AC--T-A---AC-A-AT-ACAAGCC-----CT-AGAA--C-----A--G--GATG--A--------A--T-ATGAG----   2957
VERVET.KE.VER155   C-C-GG--------AC-AGA---GGGTCCAA-----GATCAGCT-G--C-AACA--T--GA-TAGAT--CAAGAA------T-AGAA--------G-----GGTG--A-G---------C--TGAG----   3459
VERVET.DE.VERAGM3  C-A-GG-----G---C-GGA---TGA-TACAC----GATCGGTTGG----AC-A--A---ATGA-AT-AAGTGCC------T-AGAA-----------G---GTA------A----A--T-ATGAG----   2962
SABAEUS.SN.SAB1C   A---TAA---CCAG---TAGGCC--AG-CAG-A---CT-GT--TGG--C--C--T-A---G-CT----AGAA-CT-----GT---AA--C--T--A--G--GTT---A--G--T--A----A---C----   3635

SUN.GA.SIVSUN      --ACTCA----------T-AT--G-A-AA-A-----GA-GA--T-G-GA-AC-AT----G--G-TA--CA--GAA---AATT-AGAA--------G------T-----GG---G--A--C-ATAAG----   3594
LHOEST.KE.LHOEST   --A-TCA-------A--TAG---G-G-AAAGGA---GATCAGGC-G--A-A--A--C---G---T---TA---CA---AACT-AGAG--------A--G--GTTC--AGC---G--A--C-AT-A-----   3589
LHOEST.CD.447      --A-TCA-------A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-----G-T-TCT-AG---CA---AATT-A-AA--------G--G---TT---AGC---------C-AT-A-----   2505
LHOEST.CD.485      --ACTCA----G--A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-A---G-G-TCT-AA---CA---AATT-AGAA--------A--G---TT---AGC---G-----C-AT-A-----   2505
LHOEST.CD.524      ----TCA----T--A---AGG--G-A--AAGGA-----TCAGGC-G--A-A--AT-A---G---T-T-AA-A-CT---AACT-AGAA--T-----G--G---TT---AG-CA----A--C-A--A-----   2502
MANDRIL.GA.MNDGB1  ----T------G--A----ATAC-GC--AAG-----GAG-A-GCT----TA--AT-A--GGCTTTAT-AA---CA---AACT-AGAA-----T--A--G---T----------------C--T-A-----   2924

COLOBUS.CM.CGU1    --A---C-------A--TC-GCCTT---A-G---------GGGC-G--C-------C---G-TGCA--C---TAC-AT--GT-AGAG--------A--G---T-C---GC---T-------A-AAG----   3009
SYKES.KE.173       --A-TGA-T--CA----TAG-TC-GAGAA-GCA---CAGGAG-T-G--C-AC-AA-AGTGAC-GCA--A--A-AAGT----T---AGGTC---A-AG-----TGG---G-CC--TA----A-G-AG---T   3431
RCM.NG.NGM         ----T------G--A--TAGGAC-GC-CATG-A-----TCAG-TG--CA-AAG---C---G-G------C---CC-----TT--GAG--T-----T--G---TT----G-CA-G-----C--TGAA----   3586
MANDRIL.GA.14CG    C---T---G---AG---TC-GAGTGCGTCAG----C-AC-A--TGG--A-TAG-T----GG-----T-AAG-TTC-----G--CGAG--C----------GTTC--A----------T--TGAG----   3165
MANDRIL.14CG Pol   _L__F__V__G__S__D__L__S__A__S__E__H__N__K__M__V__N__R__L__R__E__H__L__R__F__W__G__L__E__T__P__D__K__K__F__Q__K__E__P__P__F__E__W__   p51
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2       TGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGC......TGGACTGTCAATGACATACAGAAGTTAGTGGGGAAATTGAATTGGGCAAGTCAGATTTA   3361
H1B.FR.HXB2  Pol  M__G__Y__E__L__H__P__D__K__W__T__V__Q__P__I__V__L__P__E__K__D__S__.__.__W__T__V__N__D__I__Q__K__L__V__G__K__L__N__W__A__S__Q__I__Y   p51

CPZ.US.CPZUS      ----C--------------A-----G--------G-----C---CA-T-A---C-------A-TT.....-----A--------T--T-----A-----A------C----C--------------A--   3421
CPZ.GA.CPZGAB     ----A-----G-----C--A--C-----------C--------TCAAT-A-----------GGTA.....--------------T-----A--AC-GA-A--A--G--A-------------------   3422
CPZ.CM.CAM3       ----C-----G--------------G--------G-----A---CA-T-A---C-G-----G-C-.....--------------T--T-----AC-------A---C-A--------------A--A--   2772
CPZ.CM.CAM5       ----A-----G--------A--C-----------G-----A---CAAT-A---C-G-----G-T-.....--------------T--T-----A-----A--A-----A--------------A-----   3063
CPZ.CD.CPZANT     ----A-----GT-A-----A--C--------T------AAA---AA---A-----GCC---TGAT.....------A--T--------C-----AC----A--A-----A------------------C--   2800

H1A.UG.U455       ----G--------T--------C-----------T---------CA-------------------.....--------------T--------A-----A--A---C-A-------------------   2807
H1C.ET.ETH2220    ----G--------T--------C--------------------CA------------G--T---.....--------------T-----A-----------A-----A--C-----------------   2753
H1D.ZR.84ZR085    -------------------------G------------T-----ACA--------G-----A---.....--------------T----------------A-----A-----------C--------   2883
H1F.BE.VI850      ----GC---------------------------G---------CAAT-----A-C--G-----.....--------------------------A--A---C-A-------------------   2698
H1G.SE.SE6165     ----A-----G-----------C--------G-----A------CA---------C--G--A---.....--------------T-----A-----------A---C-A-------------------   2758
H1H.CF.90CF056    ----A-----------------C--------------A--G--AAA------------------.....--------------T----------------A---C-A-------------------   2708
H1J.SE.SE92809    ----A-----G--T--------C-----------C--A------CAA--------------AGA-.....--------------T-----A-----------A---C-A-------------------   2675
H1K.CM.MP535C     ----A-----------------C-----------G---------CAA--------C--G--T---.....--------------T----------------A-----A--------------A--A--   2557
H101_AE.TH.CM240  ----A-----------------C-G---------C----------AA------------------.....--------------T--------A--------A---C-A--------------A-----   2935
H1N.CM.YBF30      ----A--------------A--C-----------C-----A---AA-T-A--------G--TGTA.....--------------T--------A-----A--A--G--A-----------------C--   2955
H1O.CM.MVP5180    ----A-----G-----C--A--C--G--------------C--CCAAT----T--C-----AGTG.....------A--A-----T-----A--A-----A--A-----A--------------A--C--   3391
H1O.BE.ANT70      ----G-----G--------A--C--------------T-C--CCAAT----TA-C--G--TGTG.....------A--A-----T-----A--AC--A-A--A--GC-A--------------A--C--   3415

MAC.US.SMM251     ----G--C---T-GTGG--GACA-------AGT-G--AAAG----A-T-----C---G---G-C-.....-----A--G-----T--------------A--AGT---A---------GC---A-----   3639
MAC.US.SMM239     ----G--C---T-GTGG--AACA-------AGT-G--AAAG----A-T-----C---G---G-C-.....-----A--G-----T--------------A--AGT---A---------GC---A-----   3663
MAC.US.SMM239 Pol M__G__Y__E__L__W__P__T__K__W__K__L__Q__K__I__E__L__P__Q__R__E__T__.__.__W__T__V__N__D__I__Q__K__L__V__G__V__L__N__W__A__A__Q__I__Y   p51

STM.US.STM        ----A-----GT-ATGG--AACA--G----A-C-G--AAAA----AAT-A--TC-G-G----GTT.....-----A--G-----T--C--A--AC-G--A---GTC--A--C------GCA--A-----   3310
SMM.US.SMM9       ----A------T-GTGG--CACC-------A-C-G--AAAA----A-T----GC---G---A-C-.....-----A--A---A-T-----A--A-----A--AGTGC-A---------GCA--A-----   3136

H2A.GW.ALI        ----C----G---GTGG--AAC--------AGC-G--AAAA---CAAT----CC-G-----AGTA.....-----A-----------C--A--ACAT-----TGTCC-A---------GCA--A--C--   3739
H2A.SN.ST         ----C--------GTGG--AAC--------AGC-G--AAGA---CAAT----CC----G--AGTA.....-----A-----------C--A--AC-G-----TGTCC-A---------GCA--A--C--   3191
H2A.DE.BEN        ----C-G------ATGG--AAC--------AGC-G---AAAC--CAA-----CC-G-------TA.....-----A-----------C--A---C-------AGTC--A--------GGCA--A--C--   3746
H2B.GH.D205       ----------G---TGG--AA-A--G----A-C-G--AAAA---CAA--------------AGTT.....-----A--G----CA--T--A--AC-G--A--AGT---A--C------GC---AC-C-T   3719
H2B.CI.EHO        ----G-----G---TGG--AA-G-------A-C-G--AAAA---CA---A-----------GGTT.....-----A--A--------T--------------AGT---A---------GC---AC---T   3717
H2G.CI.ABT96      ----A------T-RTGG--AA-A--G----A-C----AAAG--TAAC-----M--------G-CT.....-----A--A-----T--T--A--AC----A---GT-C----Y------GC----C-A--   3073

TANTALUS.UG.TAN1  ----A--CA-GT-GT-C---TC-C-C----GTC-CA-CAAA----A-A-A-----GC----ACA-.....-----A--G-----T--T--A--A-----A--AGTG--A--------T-----AT-A--   3563
GRIVET.ET.GRI677  ----G---A--T-G-----GA----------CA-TA-CAAA----AAT-A--CCCCTT---AG-AGAA..------A--A--CA-A---------G----A--AGTTC-A--------------A-----   3367
VERVET.KE.VER9063 ----A--CA--T-ATGG--AC-A------CA-C--A-TAG----AACT-AGA------G--ACAA.....---------T-----T--C--A--AC----A-----------C------GC----C----   3591
VERVET.KE.VERTYO  ----A--CA----TTGG---C-C------GA-C--AGCAGA---CAA---GAG--------TGAA.....----------------C-----------T------C-A---------GCA--AT-G--   3081
VERVET.KE.VER155  ----A---A--T-ATGG---C---------AGT----AAG-----AAT-AGAGA-G-----ACAA.....------A--G-----TC-T-----A--G--A--------A---------GCA--AT-A--   3583
VERVET.DE.VERAGM3 ----A--CA--T-GTGG--AC-C--G---CAGA--AGCAGC----AAT-AGA---C-----AGAA.....---------A-----T-----A-GAC-----------C-A---------GCA---C----   3086
SABAEUS.SN.SAB1C  ----G------T-GT----AA-G------CA-C-G---GAG--CACC--A-------GG--AGAA.....------A--G-----T--T--A--A-----A--A--G--A--------------A-----   3759

SUN.GA.SIVSUN     ----G--CAT-T-A--C--A--C-G----GAGA--G--AAG---AAAT-A--CCCTCT---AGAAGAG...CCA--A--G--------------AA-T--A--AGTTA-C---------TTCA--AT-A--   3721
LHOEST.KE.LHOEST  ---------T---G----------GG---GA-A--G-AAAG--TAAAT-A-----G-TG--TCTAACAAAAACT--A--A---C-A--------AC-G-----AGT-C-T---------GC---AT-G--   3719
LHOEST.CD.447     ----G----T---T-----A----GG---GA-A--G--AAA-CTAAA--------C-T----CTTCAGAAGACA--A--T--CA-----------C----A--AGTC-----------GCA--AC----   2635
LHOEST.CD.485     ----G----T---T-----A----GG---GA-A--G--AAAG-TAAA--------C-T----CTTCAGAAGACA--A--A--CA-----------C----A--AGTC-----------GCA--AC----   2635
LHOEST.CD.524     ----G----TC--G--C--A----G----GA-A--G--AGAG--AAA--------G-T-----CAAGGAAAACT--A--A-----A--T-----AC-G--A--AGTT--A---------GCA--AC-G--   2632
MANDRIL.GA.MNDGB1 ----G-----GT-A--C--A-----G----AGA--G-AAAGG-TCAA--A------TT--CAGAACAG...CCA--A--A-----A--------A--G--A--T-----A--------TGCA---T-A--   3051

COLOBUS.CM.CGU1   ----C------T-A------A-GC------A-T----AAAGG-CAAAA-T------C-G--T-AG.....-----A--A-----T-----A--A--G--A--AGCC--A------TT---C--AT-A--   3133
SYKES.KE.173      -A-----CACT--T-----A--C--G---CAGT-G---AAA----AA--C--TA-T-T---TGA-GAA...ATC--A--G---C-GT-------A--GA-A--AGT-C-T-----------C--------   3558
RCM.NG.NGM        ----C----TG--------A--C--------------AGAG--CAA-T-A-T---------AGAA.....-----A--A-----T--------A--G--A--------A----------------A--   3710
MANDRIL.GA.14CG   ----G----TG--G--C--AC-G-----------C--AAAG---CA--------------G-CA.....-----A--A-----T--T-----AC-G--A--A---C-C--------------A--C--   3289
MANDRIL.14CG  Pol M__G__Y__V__L__H__P__Q__K__W__T__V__Q__K__I__Q__L__P__E__K__E__T__.__.__W__T__V__N__D__I__Q__K__L__V__G__K__L__N__W__A__S__Q__I__Y   p51
```

```
H1B.FR.HXB2        CCCAGGGATTAAAGTAAGGCAATTATGTAAACTCCTTAGAGGA...ACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGAGAGATTCTAAAAGAACCA   3488
H1B.FR.HXB2  Pol   __P__G__I__K__V__R__Q__L__C__K__L__L__R__G__.__T__K__A__L__T__E__V__I__P__L__T__E__E__A__E__L__E__L__A__E__N__R__E__I__L__K__E__P_   p51

CPZ.US.CPZUS       T-----A--A---A---A---G-----C---T-AA-A-A----...G-T-----T-----T-----TG--AATT-T---C-T--G-----AA-G--GT-A-A------------A--------G------   3548
CPZ.GA.CPZGAB      T-----A--A---A---A------------G--GA-A------...--A--GAA---G-----T---G-T--T--C---CC---------AT-----T-A--------T--G-----AG---GCAC----   3549
CPZ.CM.CAM3        T-----A--A-------AA--G-----C--G--GA-A-A---G...--A-----T--------CA--G-GA-CA-G---C-G--------AA-G---T-A-A---G-----R------T---RG--T---   2899
CPZ.CM.CAM5        TT-------A-------AA--G-----C------A-A-A---G...--A-----CT-G--T------G--G-CA-G---C-G--------AA-G---T-A-A------T-----A---T----G--T--T   3190
CPZ.CD.CPZANT      ------A--C---ACT-A---G--G----G----A-C------...GT---GAGT--------TAG-G---A-A-G--TAGG--------AT-----T-A-A------T-A-C-A------C-GC-GAA-   2927

H1A.UG.U455        TG---------------A----C-G--------T--C--G---...G----------------TA--G--A-C--G--T-----------AT-----T-------G-----G--------------C--T   2934
H1C.ET.ETH2220     ----------------------C-G--------------G---...G----------------CA--G--A-------T-----------AT-----T-------G-----G--A--------G------   2880
H1D.ZR.84ZR085     T-----A----------A--------------------G---...-----G-----------G-----------------------AT-----------------G-----------G------   3010
H1F.BE.VI850       T-------------------C------------------G---...G----G-----------CA--G-G----G--T-C---G------T-----T-----A----T--G----------G-------   2825
H1G.SE.SE6165      T----------------C---C-----------------G--G...G----------------CA--G--T----G----C---------AA-G--GT-------G-----G--A-------G------T   2885
H1H.CF.90CF056     T---AAT----------A----C----------------G--G...G--------T-------CA---------G---A----G-----AT-G---T-------------G--------G-G-------   2835
H1J.SE.SE92809     T-----A----------A----C---------------A---G...G-T--------------CA--G-----T-G--TAG---------AT-G----------------A------------------   2802
H1K.CM.MP535C      T----------------AA-----G--------------G---...GT---------------CA--G-----T----T-C----------T-----T-A-----G--------------------G---   2684
H101_AE.TH.CM240   TG----------G----A----C-G-----------C--G---...G-T--------------CA--G-------G--T-----------AT----GT-------G-----G------------ACC--T   3062
H1N.CM.YBF30       T-----A--C-G-----AA--GC-C------T-AA-C------...--------TT-G---------G-CAACT-T--------------AT-------A-----------G-----AT----------C   3082
H1O.CM.MVP5180     T-A---A----G-----AAG----G--C--GT-AA-C------...------T--T-G-----G---G----TT---GTA----G-----A------T-A-A------------A-AG--------G---   3518
H1O.BE.ANT70       T-A---A----G---G--AG----G-----GT-AA-------C...-----GT--T-----------G----TT---GTAG---G--------G---T-A-AG-----------A-GGT-----C-----   3542

MAC.US.SMM251      T-----T--A---ACC-AA--TC-C----GGT-AA--------...-AA-TGA-T--------G-A-G-T-AGTGG--T--GATG-----AGC----TAT-AG-----T-AGATA-----C-GTC-GGA-   3766
MAC.US.SMM239      T-----T--A---ACC-AA--TC-C----GGT-AA--------...-AA-TGA-T--------G-A-G-T-AGTGG--T--GATG-----AGC----TAT-AG-----T-A-ATA-----C-GTC-GGA-   3790
MAC.US.SMM239 Pol  __P__G__I__K__G__K__H__L__C__R__L__I__R__G__.__K__M__T__L__T__E__E__V__Q__W__T__E__M__A__E__A__E__Y__E__E__N__K__I__I__L__S__Q__E_   p51

STM.US.STM         T-----A------ACT-AA--TC-G--C---T-AA-C-----T...-AG-TG-----G--T--G-AGG---AGTGG------ATG------GC----TAT--------T-A-ATT-----C-GTC-GGAG   3437
SMM.US.SMM9        T-----A--A--GACT-AA--TC-T--C-G-T-AA-C------...-AA-TGA-TT---------AGG-T-AGTGG--T--GATG------GC----TAT-A--------AGATA----CC-GTC--GA-   3263

H2A.GW.ALI         ------A--A--GACC-AA--C-------G---AA--------...-AA-TGA----C-------G-G-G-AGTGG------CT------AGC------A-AG--G------ATT--CT---GTC-GGA-   3866
H2A.SN.ST          ---------A--GACC--AA-C-------GGT-AA-C------...-AA-TGA----C-------AGG---AGTGG------TT------AGCG-----A-A--------A-ATC--CT---GCC-GGA-   3318
H2A.DE.BEN         TT----A--A---ACC-AA--C-------G---AA--------...-AA-TGA----C-------A-G-G-AGTGG------CT------AGC---G--A-A--------A-ATT--CT-G-GCC-GGA-   3873
H2B.GH.D205        T--T--A-----GAC------CA----C----AA----G---...-AG-TGA-C----------A-G---AGTGG------CT------AGC---G--ACAG--G--T-A-ATC--CT--G--C-GGA-   3846
H2B.CI.EHO         ---G--------GACC-----TA----------AA-A--G---...-AG-TGA-C----------AGG---A-TGG--T---TTG------GC----T-CCAG-------A-ATC--C---G--C--GAG   3844
H2G.CI.ABT96       T-----A--C---AC--AAA-CC-G--Y---T-AA-------T...-AR-TGA-CT-G--G----A-G---AGTGG------TTG-----AGC----T--CA--------AGATT--C---R--C-GGAR   3200

TANTALUS.UG.TAN1   T-----A-----GAC--AAGCCC----C---T-A--AC-----...G-TC---ATT------CT---G-G-A-TGG--G--GC-G------G----GT-CAG-C----TCA------CT--------GAG   3690
GRIVET.ET.GRI677   T-----A------ACC-AA--TACC---GCCA-GT-G-----G...-AA--GAAC--CCT-----A----GT-TGG--------G------GC----TATAAGA-C--TCA--G---AG-GC-G---A--   3494
VERVET.KE.VER9063  ---C---T-A-G-AC--AA--GA----C-GGT-AA-C--G---...-AG---AAT---TT---G-A-G-G-A-TGG--T-----G--G--AGC---GTAT-----G----A---A---T-G---AC-GA-   3718
VERVET.KE.VERTYO   T-----TC----GAC---A...A----C--GT-AA---C----GGA-AG---AAT--GTT---GC--G--G-TTGG---CCT--G-----AGCT---TAT--------TGC---------T---AC-GA-   3208
VERVET.KE.VER155   T-----AT-G-G-AC--AAA-TA-C-----G--A---------...-AG---AATT--TT---C--GG--GA-TGG--CCC---G-----AGC---GTAC-A--------AG-----C------AC-GAG   3710
VERVET.DE.VERAGM3  ------AC-C-G-ACT-AAA-C--G------T-AA-C------...-AA--GAACT--CT----AC-G--A-CTGG-----G--------AGC----TAT----------A------CT-----ACGGA-   3213
SABAEUS.SN.SAB1C   -A-------A--GACT-AA--T--G----G---AA---------...G-T-G-C-TT-G-----GA--G-C-A-TGG--------G-----A------T-A-A------T---C----CT---G-C--AAG   3886

SUN.GA.SIVSUN      TGA----C-C-G-AC--AAG-------------AA-------GAAT-AACC-TT-GC-GAGA------AGATGAC-GA-----CTAGG--A...---TACCAGAGT--TCAG---G-AT--C-G--GT--   3848
LHOEST.KE.LHOEST   TGAT--T----GGAC--AAG--C-C--C---T-AA-A--G---...GTA---C-CT-GGA----A-C---AACTGG-----G-----CTT-GA----TAT-GGC-G----A----G-A--T------AAG   3846
LHOEST.CD.447      TGAG--A----G-ACT-AAG--C-T------T-AA----G---...GTA-----T---GAT-----TG--A-CTGG-----T-C----TTAGA----TAT-A-C-----CA-C----A---------AA-   2762
LHOEST.CD.485      TGAG--A----G-ACT-AAG--C-T------T-AA----G--G...GTA-----T---GA------TG--A-CTGG-----T-C----TTAGA----TAT-A-C-----CA-C----A---------AA-   2762
LHOEST.CD.524      TGAT--A--C-G-AC--AAG-GC-C-----GT-GA--------...GTA--GC-CT-GGA-C--C--G-GA-TTGG-----G------CTAGA----TAT-AGACC--T-AGACA-----------AA-   2759
MANDRIL.GA.MNDGB1  T--T-----C---AC--AA---C-G--C--G--AA-A-------...GGACT-AACA-------GAA-G-CA--A-G------------AGA--G---TAT-A-C----T-A------CT-GGCT---GA-   3178

COLOBUS.CM.CGU1    ------A--C-GGAC--A-G-GA--A-----TG-AC--A---T...-AG--GCATT-GTTG--G-A-G-G-AG--T-GT-C---------AGAG--G--ACA---T----A-C----AA-C------GA-   3260
SYKES.KE.173       TT----A-----GAC--AAG-GC-C-----GTGTA-C------...--T---CA-T-----------C-CA----T-------C---T---GC----T-A-A---G--T---C-A--CT----G---GA-   3685
RCM.NG.NGM         TT----A--A--GAC--A-G-G-----C-----AA-A--G---...GTT--G---T-GGAT--CAAGG--GA-T-T---AG---------AT-----TAT-A--------A--CTG-C--G--G--GAA-   3837
MANDRIL.GA.14CG    TT-C--A------AC--AAG-GC-C--C---T-GA-C-----G...G-A---C-T---GAT------G--GA-TGG--TAGG--------AT-----TAT-A------T-AGTTA--AG-GC-G--GGA-   3416
MANDRIL.14CG .Pol  __S__G__I__K__G__K__E__L__C__K__L__I__R__G__.__A__K__P__L__D__E__V__V__E__W__T__R__E__A__E__L__E__Y__E__E__N__K__L__I__V__Q__E__E_   p51
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
H1B.FR.HXB2        GTACATGGAGTGTATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCA...TTTAAAAATCTGAAAACAGGAAAATATGCAA   3615
H1B.FR.HXB2  Pol   _V__H__G__V__Y__Y__D__P__S__K__D__L__I__A__E__I__Q__K__Q__G__Q__G__Q__W__T__Y__Q__I__Y__Q__E__P__.__F__K__N__L__K__T__G__K__Y__A__   p51

CPZ.US.CPZUS       --------G--C------------GA------A---G--------G----A--A--A--AAGGA-T--G--------C-------T------A-GG...CA---G-----A--G--------------C-   3675
CPZ.GA.CPZGAB      --------G--A--C-----T---GAC-----GC-T----------------A-----CA-CT----G-----T-----G--A-T---G--A---...CA---G---T----G-----G--------C-   3676
CPZ.CM.CAM3        --G-----G-----C---------GA---G--A------------N-------A-----AA-TA-T--G-----T----------TC--G--A-AG...CA---G--C--A--------C--------T-   3026
CPZ.CM.CAM5        --G--------T--C-----T---GA---G--G-----------GG----A--A-----AA-TA-T--G-----T--C--G----TC--G--A-AG...CA---G--CT-A--------G--------T-   3317
CPZ.CD.CPZANT      A--G-G---TAC--C---C-G--TGGTCT-CCAC---A----ACC--------A--A---TC---A--------C--------A------A-TGA-...GGA---CTGT-A---G-------------T-   3054

H1A.UG.U455        --G--------A---------------------C--G--------------A--A-------A-------------------------------...------------------------------   3061
H1C.ET.ETH2220     -----------A-T----------------------------T-----------A------A-T-A-----------T----T----C-----A---...--C-----------------G--G-T-----   3007
H1D.ZR.84ZR085     A-G-----------------------------------------T-------A--A--------T-----------------------A---...-----------------------G-------   3137
H1F.BE.VI850       --------G--A-----------------------------------------A--A--AG-C--G-----------------------GA-C---...-----------A-----------G-------   2952
H1G.SE.SE6165      -----------C--------------------A------------G----A--A--A----T--A-----------------------------...-AC--------------------G-------   3012
H1H.CF.90CF056     A----------A--------T-----------------------------G------A-----------------------------G------...---------------G---------------   2962
H1J.SE.SE92809     --------G--A---------T--G-------A------------G-G-----A--A----TG-A-----------------------G------...--------C-----------G----------   2929
H1K.CM.MP535C      --G-----G--A-----------------------------------------A--A---A-T-A------------------A------------...CA------------------G--G-------   2811
H101_AE.TH.CM240   --G--------A-----------------------G--------G-------A--A-----G-A-----------------------------...-----------A-----------------C-   3189
H1N.CM.YBF30       C-G--------C------------GG------A---G----------T--A-----A--A-----T--G-----------G--------G---TT-...CA-------T-A-----------G-------   3209
H1O.CM.MVP5180     -----------A-----CC-G--TGAC--------GTGG-TTAGT--T--------T--AG----G--------T--C--GG-A-----G--TGA-...CA---G--C--T----------------T-   3645
H1O.BE.ANT70       --G-----G--A--C---C-A--TGAT--G--TC--TGG-TTA-T--T--------A--AGGG-AG--------T--C--G--A-----G--AGA-...CA---G--C--C--------G------A-T-   3669

MAC.US.SMM251      CA-G-A---TGT-----CC-AGA-GGC--GCCA---GA---CACGG--ATA---AGTCA-G-CAAT--G---T-T---A-----C-C-----A......GAC----TA------GT--------T-----   3890
MAC.US.SMM239      CA-G-A---TGT-----CC-AGA-GGC--GCCA---GA---CACGG--ATA---AGTCA-G-CAAT--G---T-T---A-----C-C-----A......GAC----TA------GT--------T-----   3914
MAC.US.SMM239 Pol  _Q__E__G__C__Y__Y__Q__E__G__K__P__L__E__A__T__V__I__K__S__Q__D__N__Q__W__S__Y__K__I__H__Q__E__.__.__D__K__I__L__K__V__G__K__F__A__   p51

STM.US.STM         CAGG-A--GCAA--C--CCGAGA-GAC--GCCTC--GAG---AC-G--GT---AG-TCA-G-CAAT------T-C---A-G--AC-C-----A......GAC----TAT-----GT---C--G-T-----   3561
SMM.US.SMM9        CC-G-G---TGT--C---C-GGAGGG---GCCAC--GAG---CC-G--ATA---AGTCA-G-TAAT------T-----A-----C-C------......GAC----TA--A---GT---C----T-----   3387

H2A.GW.ALI         CA-G-G--GCAC-----CC-AGA-GA---G--G---GA----AC-G-C--A--AG-TCAAG-CAAT------------A----AC-C--G-GA......GAA----T---A---GTG-A---G-------   3990
H2A.SN.ST          CA-G-A---TGC-----CC-AGAGGA---G--GC--GA----AC-G-C--A--AG-TCAAG-CAAT--G---------A-G--AC-C--G-GA......GGA----T---A---GT--------------   3442
H2A.DE.BEN         CA-G-A---TAT-----CC-AGA-GA------A---GAG---AC---C--A--AAGCCAAGG-CAT-----------CA----AC-C--G--A......GAG----TC--A---GT------G-------   3997
H2B.GH.D205        CA-G-A---TCC--C--CA-GGA-AGGGT-CCGC--GA----AC-G--------AA-CCTAGC-AAT--G--------CA-----C----G-GA......AA----GTC--A---GT--------------   3970
H2B.CI.EHO         CAGG-A---TCC-----CA-AGA-GGGGT-CCT---GA----AC-G-G-----AA-TCTAGC-AAT--G--------CA-G---C----G-GA......GA-----TC--A---GT--------------   3968
H2G.CI.ABT96       CAGG-A--C-CC-----CA-AGA-GR-G-GCCT--RGA---CACTG-------AA-TTTAG-CAAT--G--------CA--G--C------GA......AAC-G--TC--A---GT---T----------   3324

TANTALUS.UG.TAN1   CA---G----CT--C---------AGT---TCT--G-G---CACT--TACC----T--ATG-T--G------GG-------T-CA-A--G---......GGA---GTA--A---------------A---   3814
GRIVET.ET.GRI677   CA-G-A---ACA--C--------TCTC-----A---------AC-G-T--A-----A--AG----G-----------C---T-CACC-----A......GGGGC-GTAT-A--GGTG----G------C-   3618
VERVET.KE.VER9063  CA-G-A---ACT-------CA---GG---GCCTC-T-G-----C-G-C-----ATTA-AAG-T--G-------GT-----GT--A-A--G--A......GGA----TCT-----GT------G-------   3842
VERVET.KE.VERTYO   CAGG-A---ACC-----CA-A---GG--T-CCTA-T-GG----C-G-------ATT--AAGG---A--G----GT--C---T-CA-A-----A......GGAC--GTCT-----GT---------CA-C-   3332
VERVET.KE.VER155   CA-G-A--TACT-------CA---GA----CC-C-T-GG----C-G-------ATTA--AG-T--G------T----C---T-CA-G--G--A......GGA----TCT-A--GGT---G--G-TC--C-   3834
VERVET.DE.VERAGM3  CAGG-A--GACC--C--CA-A---GG--G-CC-A-C-G-----C-G-G--A--A-TA-AAGG---T-------GT--C---T-CA-G------.....GGAC--GTAT-A---GT---T-----CA---   3337
SABAEUS.SN.SAB1C   CAG--G---CAA-----------TG-TCTTCCT----GG--TA-GG-T-TT--A-T---AG-T---------GG------G--A--C--GCCAGA-...AA-----TA-----GGTT--C----------   4013

SUN.GA.SIVSUN      ---TCA--CAGT--C-----A---GAT-----A-----TTGCAG-G-G--A---GTAAA------AAT--TA--C-T----TGG-TG--G-GA......AAAC--GTAT-A-GGGT-----G----CA--   3972
LHOEST.KE.LHOEST   A-G--G----CC------------GA---G--AC-T--T-TCAGGG-----C-AA-CAAAA-G--GAT-ATT--T-TC--GTGGAGA----GA......AA---C-TCT-A-G-G-T--G-GG---CA--   3970
LHOEST.CD.447      C-G--G----CT--------T--TA-T-----A--GG---T-AG-G-----C--AGTAA-A-G--AATTATT--C-TC--GTGGAGA----GA......AA-TCT-T-----G-G------GG---CA--   2886
LHOEST.CD.485      C-G--G----CT-----------TA-T-----A--GG---T-AG-G-----C--A-TAAAA-G--GGTCATT--C-TC--GTGGAGG----GA......AA-TCT-T-----G-G------GG---CA--   2886
LHOEST.CD.524      -----G--G-CT--C-----T---AGT-----AC-TG---T-AGGG----AC--A-CAA-A----GGTTATA--T-TC---TGGAGG----GA......AA-GCC-T---A-G-G------G----CA--   2883
MANDRIL.GA.MNDGB1  CA-G-A--GTCT--------T--TAAT--G--A---TAT-T-AG-T-T-----AACAACAGG---AG-TATAT---T----TGGA-G----GA......AA---GGT-T-A-G-G----G-------GG-   3302

COLOBUS.CM.CGU1    --T--G--------C---ATA---CATG-G---A-TTGG-T---C--TTTC-G-TTA---A----A------GGC---GC-G-C-GC---A-A......CA-GG--CG--A-G--G-------C--AAC-   3384
SYKES.KE.173       CAGGCA--GTCT-----------TAA----CCT---GA----C-C--CACT--A-TA---TCCCAG------GG---CATG---A-G-------A-AAAGGCCC-CCC--A-T---------GACA----   3815
RCM.NG.NGM         A-------G--T------A-----GA----CCA--------TA--G-T-----ATTA-AAGG---G-------GC--C--G--AG-G--G-GAGAC...-CA---CC---A--G-----G--G-------   3964
MANDRIL.GA.14CG    -----------T------C-G---GG----CCTC-G--G--CA-GG-------A-T-ACA--G--A-------GC-----G--AG-A--G--AGAG...AA---GCC---T--GGT---G--G-----T-   3543
MANDRIL.14CG  Pol  _V__H__G__V__Y__Y__Q__P__G__K__P__L__M__A__K__V__Q__K__L__T__Q__G__Q__W__S__Y__Q__I__E__Q__E__E__.__N__K__P__L__K__V__G__K__Y__A__   p51
```

```
H1B.FR.HXB2        GAATGAGGGGTGCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGCATAGTAATATGGGGAAAGACT...CCTAAATTTAAACTGCCCATACAAAAGGAAACATG   3742
H1B.FR.HXB2  Pol   R__M__R__G__A__H__T__N__D__V__K__Q__L__T__E__A__V__Q__K__I__T__T__E__S__I__V__I__W__G__K__T__._P__K__F__K__L__P__I__Q__K__E__T__W    p51

CPZ.US.CPZUS       --CAA--ATCA--A--T---------A-C-G---GC--GTT--A-TG------------G-T--T--G-----T--C--T---------GTA...--A---------T---TAG-G--G------GTC--   3802
CPZ.GA.CPZGAB      --CAA---TCA--A-----A-----CA-C-GG---C-GG-G--A-----A--------TG-T--T--------------T--------A--A...----------GGT-A--AG----G--A----GT--   3803
CPZ.CM.CAM3        -GCAA--ATCA--A--T--------CA---G-------G------T---A---------G-T--T-----T--------C--------A--A...--A-----C-G-T-A--TG-T--------GGTC--   3153
CPZ.CM.CAM5        -GCA---ATCA--A--T--------CA---G-------G------TG------------G----T-----T--------C--------A--A...--A-----C-G-T-A--TG-T---------GTC--   3444
CPZ.CD.CPZANT      -GCCT-CA--AA-T----------G--T-GG------G-T-GA-T-------------GGATT-----T-----T--C------G--GTA...--A------C--T-A------ACT-G------T--   3181

H1A.UG.U455        ---AA---TC---T-----------------------------A-TG---------G-GT-------------------------------TC...----------G---A--------------------   3188
H1C.ET.ETH2220     A--GAG--AC---------------------G--G-------C--T---------G---G--CTG----------------------------...----------G-T-A-----C--G--A--------   3134
H1D.ZR.84ZR085     -------------------------------------------------------G-----------------------G---T-----------...----------G---A--------------------   3264
H1F.BE.VI850       A-G-----TCG-----------------------------A-----A-----G---G-TTT---------------------G----GATCT-----G--------A-------TG--A--G-----   3082
H1G.SE.SE6165      A--G-G--TC-----------------------G--------A-T-------------G-------G-----------------------...--C-----------A-------G---A--------   3139
H1H.CF.90CF056     A------AAC----------------A----------------A-----------G---T-T-------------------------A-T-...----------G---A--T--------A--------   3089
H1J.SE.SE92809     A--G----A-----------------------------G----A-TG------------G--TTG---GC-----------------A---...----------G---A----------GA--------   3056
H1K.CM.MP535C      -------ATC---------------------G-----------A---------------G--------G-------------------A---...----------G---A-----------A--------   2938
H101_AE.TH.CM240   ---GAG--TC---T----------------G------------A-TG---------G--G-------------------------------C...----------G---A----------GA--------   3316
H1N.CM.YBF30       A------ATC------T---------A-------G---GTT--A-TG--AAGG---G-GG-A--------T--------T------------...----------G-T-A--AG-----------GTG--   3336
H1O.CM.MVP5180     -GCAA-A--CCT-------A------A---G------GG----A-T---C--G--GG-GT-TCA----GCT-----T--------G--ATTA...---------C-GG-----AG-TACT-GA-----T--   3772
H1O.BE.ANT70       -GCAA-A--CCT-------A------A---G-------G----A-T-A-C--G--GG-GT-TCA----TCT---A-T--C--------ATTG...-----------G-----AG-CACT-GA-----T--   3796

MAC.US.SMM251      AG--A-A-AA-A-A--T--C----GA--T-G--T----G--C-T-T-A-A--G------GGA-AG---GCA-----G--C------C--GTC...--A-----CC-CT-A--AG-TG-G-G---TGT---   4017
MAC.US.SMM239      AG--A-A-AA-A-A--T--C----GA--G-G--T----G--C-T-T-A-A--G------GGA-AG---GCA-----G--C------C--GTC...--A-----CC-CT-A--AG-TG-G-----TGT---   4041
MAC.US.SMM239 Pol  K__I__K__N__T__H__T__N__G__V__R__L__L__A__H__V__I__Q__K__I__G__K__E__A__I__V__I__W__G__Q__V__.__P__K__F__H__L__P__V__E__K__D__V__W    p51

STM.US.STM         A---A-AAAACA-A--T--A----GA--T-G-TT---GG--C-T-TGA-T--G--G---GGG-A----GCA-----G--C------C---TC...--A-----CC-CT----AG--G-G-GA--GGTC--   3688
SMM.US.SMM9        AG--T-AAAA-A-A--T--A----GA--C-GCTT----G--C-T-T------G------GGA-A----GCA---A----T------C--GTG...--A-G----C-TT----AG--G-G-RA----TT--   3514

H2A.GW.ALI         A-----AAAA-A----T--C--C-GG--C-G-TTG---G--C---T---T---------GGA-A----GCAC-G--C--T------CGA-TA...--A-G----C-C--A--AG--G---GA-----C--   4117
H2A.SN.ST          AGG-A-AAAA-A-------C--C-GA--C-G--TCC--G--C-A-T---T---------GGA-A----GCAC----C--T------CGA-TA...--A------C-C--A--AG--G---GA--T--C--   3569
H2A.DE.BEN         AG--A-AAAA-A----T--C----GG--C-G-TT-C--G--C---T---T--G------GGA-A---GGCAC----C--T------CG--TA...--A------C-C-----AG-GG-G-GA--G--C--   4124
H2B.GH.D205        AGG-T-AAAACA-G-----C--C-GG----G--T-C-GG--C-T-T---T--G------GG--A----GC-C----C--C------G---TA...--AGTG--CC-T-----AG--G---GA--G-----   4097
H2B.CI.EHO         AGG-T-AAAACA-T-----C----GA----G--T---GG-TC-T-T---C---------GGA-AG---GCAT-G--C--C------G---TA...--A-TG--CC-T--A--AG--G---GA--G-----   4095
H2G.CI.ABT96       AG--A-AAAA-A----T--A----GA----G-TT---GG-TC-T-TG--C-----R---GGA-AG---GCAC----------------ATTA...--CTTC---C-T-----AG--G---G---C--C--   3451

TANTALUS.UG.TAN1   AG-GC-AAAA-A-T--T--A-----GT-T-G-GT---GG---GAATG--C----GGG-GTG--AG---GCG--CACCT-T--------ATTA...--C-TCA-GG--G-A--AG--G---GA---GT---   3941
GRIVET.ET.GRI677   AGCAA--A-AAA-T--T---------C---GGACTC--G--C-CCTT--C--------CTGT-AG---GCAC-TACC--T-------GACT-...--ACG-G-AC----C--AG--G-C---A-------   3745
VERVET.KE.VER9063  A-CAA-A-AACA----T--A-----AC-TCGCAC---GG-C-GCCT---A--------CTGT-A---GGCAC-G--C--T------G-ACTA...--A-TT---G----C--AG--G---G----GT---   3969
VERVET.KE.VERTYO   AGCAA-A-AACA----T--A-----AC-TCGCAC----G-T-GTTT------G--G--TTG--A----GCTC----T--T-----G-TATTA...--AGTTC-AG----C--G---G---GA--GGT---   3459
VERVET.KE.VER155   A-CA--AA-C-A-T-----C-----GT-GCGTGT-C--G---GA-T---A--G------GGG-A---GGC-C-------T------C-ATTA...--C-CT---G----C--AG-GG-G-G---C-----   3961
VERVET.DE.VERAGM3  AGCA--AAAACA-T--T--C-----GT-CCGTGT---GG---GATT---A------C-TTGT-A----TCTT----T---------G--TTG...--AGTCC--G----C--A---G-G-G----GT---   3464
SABAEUS.SN.SAB1C   A---C-AAACA--A-----C-----AT---G-ATG---G---GCCT---A--G--G---GGA-AG-----------C--C------C---TA...----TTA-GG-------AG--G---GA--GCTC--   4140

SUN.GA.SIVSUN      AG-GAG---CA--A--TGAG---CC-TGTC----G---G---CT--CC-A-----G---GGA-G------------T---------TTTGTG...--A--GA-AC-GG-A----------GA--G-T---   4099
LHOEST.KE.LHOEST   --CA--A--CA--A-----A---CCCC--C-GA-----GT---A--TA-T--G--G---GGA-A------------C--C-----CTTTGTG...--A---A--C--G-C--AG-GACT-GA---GTC--   4097
LHOEST.CD.447      --CAA-AATCA--T-----A---CCCC--C--A----GG----------A---------GGA-A---G--T-----C--T-----TC-TGTG...--C--GA--C-GG-A--GG--AC--GA---GTG--   3013
LHOEST.CD.485      -GCAA-AA-CA--T-----A---CCCC--C--A-G--GG----------A---------GGA-A---G--T-----C--T-----TC-TGTG...------GA-CC-GG-A--AG--AC--GA---GTG--   3013
LHOEST.CD.524      --CAA-A-A----T----TG---CCCC-TC-G------G----A--T--A--G------GGA-A------------C--T-----TTTTGTA...--A--GA-AC--G-A--TG-CACT-GA--------   3010
MANDRIL.GA.MNDGB1  A-CA--AAACA--A--T-G------CC-C-TGA----GG---GT--TAC---G--GG--GGA-G------T--------C-----TTTTGTA...--A---A-GC-GA-A----CTAC--G---G-T---   3429

COLOBUS.CM.CGU1    C-GGC-AAAA-CAG---TT----AG-A-GC-GG-G---G-TTCA-TTA-A-----G---GGT-GG---GC------C-CC---------GTG...--C---A-G---G-C--AGC-A-G-GA--GGAT--   3511
SYKES.KE.173       A--CCTTT-CA--A--T-G------CTATC--TCTC-TG-TC-ATTGT-AA-T------GGG-TCC--TCAC--TGGTAT---------GTA...--A-C---CC-T--C--AG--A---GA--GGA---   3942
RCM.NG.NGM         A-CAA-A-ACA--A--T--A-----AA-T-G-ATGC-GG---GCTT---A--G------TG--A----GCT--------T---------CT-...--A------G-G-----A---G---GA---GT---   4091
MANDRIL.GA.14CG    -G-CA-AACA---------A-----GT---GGGTGC-TG---GATT---A---------G---A---GT-------G--T------G-ATTG...--A-----CT-T--C---C-GG-G-G----GT---   3670
MANDRIL.14CG Pol   R__G__K__H__A__H__T__N__E__L__R__V__L__A__G__L__V__Q__K__I__A__K__E__C__I__V__I__W__G__E__L__.__P__K__F__Y__L__P__L__E__R__E__V__W    p51
```

HIV-1/HIV-2/SIV
complete genomes

```
                                                                                      Pol p66 Reverse Transcriptase continued (RT is dimer of p51 and p66)
                                                                                                                  Pol p51 Reverse Transcriptase end \/ Pol p15
H1B.FR.HXB2        GGAAACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCTGAGTGGGAGTTTGTTAATACCCCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACCTTCTAT   3872
H1B.FR.HXB2  Pol   __E__T__W__W__T__E__Y__W__Q__A__T__W__I__P__E__W__E__F__V__N__T__P__P__L__V__K__L__W__Y__Q__L__E__K__E__P__I__V__G__A__E__T__F__Y_    p15

CPZ.US.CPZUS       -------------T----A--------G--------------A--T-----A-----C------------C-C--A---C-T-----TA---------C---GG------AG--G---------A--T---   3932
CPZ.GA.CPZGAB      ----G--------G---A--C-----G--A--------C-----A-----A---A-C--------A---------C------------AGT-------C------T---CC-ACCA----T--T-AT---   3933
CPZ.CM.CAM3        ------C------T-------------------------C--A--C-----N--C--G------------C-T--A---------------A--G--A-C----------CC------------G------   3283
CPZ.CM.CAM5        ------C------T----A--------------------A--A--C-----A-----A------------C-T--A---------------A--G--ACC----------CC---G--G-----A------   3574
CPZ.CD.CPZANT      ---TG-C------T----C--------G-----------A--A--A-----------------A--C--A---A-T-G-C-C-----TA-TC--TT-GC---C--T--TCC--AG--T--------T---   3311

H1A.UG.U455        ----G---------TG-----------G--T--------------A-----------C------------C---A-------------------------C-------C---------G--A------   3318
H1C.ET.ETH2220     ----G------------C---------------------------A------------------------C---A----------------C-----------------C-----T------T------   3264
H1D.ZR.84ZR085     --------------T---C------------------------------A-----C------------T----A-----------------------G---------A-------------T------   3394
H1F.BE.VI850       ---T-------------T---------------------------A-----------C--------C--TC--------C-----------------A-C---G-------C---------C---------   3212
H1G.SE.SE6165      -----T-----------C--------G------------------------------C------------TC----A--G--------T-G------A-C-----------CC------------T-A----   3269
H1H.CF.90CF056     ---G--C-----------------------A-----------A------------C-------ATC----A------------T--------A-C---G-------C------------T-A----   3219
H1J.SE.SE92809     ---G-------------C--C-----G-----------------------C--------C--TC----A--G------------A-----A--G---------A-G------------T------   3186
H1K.CM.MP535C      ---G-------------A-----------------C-----A-----------C-----T------C----A---C----------------A-C----------------G--------T-----C   3068
H101_AE.TH.CM240   --------------TG-----------G--T-------------A--------------------TC----A-------------A-----A-----C-----------------G--T------   3446
H1N.CM.YBF30       ---GG----------C--TC----------A--T-----------------A-----C--C--T------C-T--A------------T---------A-C---G--A--CAGT--G--------T------   3466
H1O.CM.MVP5180     ------T------G----A--------G-----------------A-----A-----C-GC--A--C--A--GA-C-----------------------A-C------T--T-----G-----------T---   3902
H1O.BE.ANT70       -------------G-G--C-----------------------A--A-----A-----C-GC--A--C--A--GA-C------------AG-C----A-GT-----T--TA-G--G----------AT---   3926

MAC.US.SMM251      ----CAG-----------C--------G-TA--------A--G--------T---A-CTCA--G--A--AC----A-G----GTC-T-A-TC---T---G--C--T----AG----A---------AT---   4147
MAC.US.SMM239      ----CAG-----------C--------G-TA--------A--G--A-----T---A-CTCA--A--A--GC----A-G----GTC-T-A-TC---T---G--C--T----AG----A---------AT---   4171
MAC.US.SMM239 Pol  __E__Q__W__W__T__D__Y__W__Q__V__T__W__I__P__E__W__D__F__I__S__T__P__P__L__V__R__L__V__F__N__L__V__K__D__P__I__E__G__E__E__T__Y__Y_    p15

STM.US.STM         ----CA-------G----T---------G-TA--A-----A--A--A-----T-----ATCC--A-----T-----A-G----GTC-TTA-T----TA---------T-G-A----A--------A--T---   3818
SMM.US.SMM9        ----CA------------T----------TA--------A--A--A-----C-----GTCA--A-----------C-G-C--GTT-T-A-CC---TA-----G--T---CAG-----------A--T---   3644

H2A.GW.ALI         ----CAG------GAT--C--C-------TG--A-----C--A--C-----C-----ATC------A--AC-G--C-GGC--GCA-TTA-CC---TA-----T--T---C----C-----G--------C   4247
H2A.SN.ST          ----CAG------GATA-C--C-------TG--A-----C--A--C-----C--CA-ATC------G--AC-G--C-G----GTA-TTA-CC-G-T------T------C----C---------------C   3699
H2A.DE.BEN         ---GCAG------GATA-C--C-------TG--A-----C--A--------C-----ATC------A--AC-G--C-GG---ACA-TTA-CC---TAGG---T--T---CC---C-----G--------C   4254
H2B.GH.D205        ---CCAG-----------T--C-------TA--------C--A--------C-----CTCG-----A--A---A-A-G-C--GCC---A-CC---TC-----C---C---A---GAG--------A---C   4227
H2B.CI.EHO         ---TCAG-----------T--C-------TA--------C--A--A-----T-----CTCA-----A--A---A-A-GG---GCC--TA-CC-G-TC-----C---C---A-----T-------T-A---C   4225
H2G.CI.ABT96       ----CA------------C----------TA--T-----C--A--T-----C-----GTC---A--G-ATC----A-GG---GTC---A-CC---TC---------TC---AGCAGA----G----AT---   3581

TANTALUS.UG.TAN1   ----CA--------GT--C----------T---A---C-A--A---ATA--A--CA-CTCC--A--G------A-C-----------TA-C--G-TA-----T------CC-AA----A-GGTG-AT---   4071
GRIVET.ET.GRI677   ---T-TG------CAG--C----------TAT-------A--A--A-------------GC--A--A-T-C----A---C-G-----TTCC----TA--------A--CAA-----A---TGTT-AT---   3875
VERVET.KE.VER9063  ----CA-------G----C--C-------TG-G------------A-----A-----G-G-GTA--A--TC----------G------ACTC-GACC--G-----T---CC-----AG--TGTA-AT---   4099
VERVET.KE.VERTYO   ---GCA-------G-G--T--C-----G-TA-G---------C--A-----T-----C-GC-----A--T--GC-C---C--------ACA---ACA------------CCCAAG-AG--CGTT-A----   3589
VERVET.KE.VER155   ----CA-------G----C----------T--GT-----A--C--A-----C-----C-G-GTT--G---------A-CT--G-----TACAC-GACT--G---------CCCG----AG--TGT--A----   4091
VERVET.DE.VERAGM3  ----CA-------G-T--T--C-----G-TA-GT--------A--C-----A-----C-G------A---C----A-----------TACCC-GACA------------CC-AAG-A---TGT--A----   3594
SABAEUS.SN.SAB1C   ----CA-------T-T--C--------G-T---------C--G--A-----AA-G--C-G---T----AA--GA-T-G---G------A-A----TA-----T-----CCC-----A--C-GT--AT---   4270

SUN.GA.SIVSUN      -AGTCAG------G----C--C-----GTGT--A-----C-------TA--A-----C-GC--T---AAG--G-A-C--GAG-----TACTC---CC-C------AG-GCC----...--C----A----   4226
LHOEST.KE.LHOEST   ---GCAC-------GC--CC-C-----G-TT--A--------A--C-TA--A--CA--TC------G-AA----AAC--GAG------ATT-GG---GC-----------A----G-T---C----A----   4227
LHOEST.CD.447      ---TCA--------GT--TC-------G-T---------A-----A-T---A---A-CTCA-----A-AG----AAC-GGA------TAGC-----AGC------T----A---G-T---T----AT---   3143
LHOEST.CD.485      ---TCA--------GT--TC-------G-T---------A-----A-T-------A-CTCA-----A-AG----AAC-GGA------TAGC-----AGC-----------A---G-T---T----AT---   3143
LHOEST.CD.524      -----AT-------T-A-C-T---------TA--T-----A--A--C-TA--A--CA-CTCC--A----AA----AAC-GGA------TAGC--G--AGC-----------A---TA----T----A----   3140
MANDRIL.GA.MNDGB1  ----GAT------CAT------------GTGT--A-----A--A--AGTA--A---A-C-GC--A---ATG----AA-GGGA------TAGC--GTCCCC------TC---AG--G-T------A-AT---   3559

COLOBUS.CM.CGU1    ----CAG-------GT--C---------AGTG-T-----C--A---ATA--------CTCAT--T--TATG----A------GTA-GGA-C--G-TTTT---G--TC-GAAG-AG---CC---T-AT-GG   3641
SYKES.KE.173       -----A------------C-----------------A---G-A--A--AGTCA-A--CA-A-G------G--A--G--T-GG-GG-AC--TA-C--G-TACC------------CC--A-----T---------C   4072
RCM.NG.NGM         ----CAG------T-G--T--C-----G-TTT-------------A-----A------G---T--A--AC--A-A-GGC--------A-TC---TA--G--T--T---CC-----A----GT--AT---   4221
MANDRIL.GA.14CG    ---TCAG------CAT--T---------G--A--A-----C--A--A-----A--CA-CTCA--A--A--AC--A-A-GG--G-----TA-CC-GTTA--------A---CC---G-AG--TGT--AT---   3800
MANDRIL.14CG  Pol  __D__Q__W__W__H__D__Y__W__Q__A__T__W__I__P__E__W__E__F__I__S__T__P__P__L__I__R__L__W__Y__N__L__L__K__E__P__I__P__G__E__D__V__Y__Y_    p15
```

```
H1B.FR.HXB2        GTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACTAATAGAGGAAGACAAAAAGTTGTCACCCTAACTGACACAACAAATCAGAAGACTGAGTTACAAGCAATTTATCTAG   4002
H1B.FR.HXB2  Pol   _V__D__G__A__A__N__R__E__T__K__L__G__K__A__G__Y__V__T__N__R__G__R__Q__K__V__V__T__L__T__D__T__T__N__Q__K__T__E__L__Q__A__I__Y__L__   p15

CPZ.US.CPZUS       --G-----A-----CC-AC-A--A--C--GAA------G---------------G----G--T---------A--A-A--TT--GAAA-T--T--T-----A---G-A---C-CAC---GG-A--CT---   4062
CPZ.GA.CPZGAB      --------A-----A--T-----A--A---ACT--G--------------A--AG-C-A---G-A-------A-CA-T-G-T--GAAA-T--C--T------C-AG-A--A---A-G--TT-GCT---G-   4063
CPZ.CM.CAM3        --G-----------------A-----A--GAA------------C--C--C---G-------C-----G---A-A--A-A-T--GAAA-T--C--T--------AG-A--AC-GAC---TG---------   3413
CPZ.CM.CAM5        -------------------A--A--A--GAA------------T--CA----AG----G--C-----G---A-AA-A-ATT--GAAA-T--T--T--------AG-A--AC--AC---TG-------G-   3704
CPZ.CD.CPZANT      --T--------------T--AA-CT--C----G--G--G-----C-----G--AG-C----------AGT-GG--AAAACA----CAAA-G--C--C-----AC-AG-A--A-----G--G---CT-A-G-   3441

H1A.UG.U455        -----------------T-----A-----GC-------------G-----C---G-C--------------G-----TT----------G-----------A--------A-----T-----CC------   3448
H1C.ET.ETH2220     --------A--------T-----A------A-------------G---------G----------G--G---A----TT-T--------A--------------------A--------G--CC-G----   3394
H1D.ZR.84ZR085     -----------------T-AA-------------------------------G-C----------G---------C--T-T-------------------------------------A------   3524
H1F.BE.VI850       ------------T----T--A-----C---AA----------------------G---A-----A-------------T-------G--G--C------------G--------------------T---   3342
H1G.SE.SE6165      -----------------T--------A-----------G---------------G-C-A-----AG------A--A-T-----------A--------C--A---G----A-----G------C-G----   3399
H1H.CF.90CF056     A----------------T-----A-------------------------C---G---------AG------------T-------G--A--------------------A------------------   3349
H1J.SE.SE92809     ------------T-----------------AC------G-----G---------G-C-A---------G-----A--T-------------------------------AC----C--C------T---   3316
H1K.CM.MP535C      ---------------C-T-----A------AAG----G----------------G-C--------------G------T--A-------A--------------AG----A-----------C-G-T---   3198
H101_AE.TH.CM240   ----------------GT-----------GC-------------G-----C---G-C--------------G--A--TT----------G-----------A--------A-----T--G--CC--T---   3576
H1N.CM.YBF30       --------A--------T-----A--A-----G-----------T-T---G--AG----G--------G-----G---T-TA-TG-A-----C--C-----A---G-------------T--CCT-A-G-   3596
H1O.CM.MVP5180     --------A--------T---A-T--A---C-------G--G-----------AG-ACA-----A---G--CA-AA-A-AGT--GAA--G-----C-----A---G----A---ATG--TG-A-TAA---   4032
H1O.BE.ANT70       --------A--------T--A-----A---C-------G--------------AG-ACA---G-A---G--GA-AA-A-AAT--GA---G--C--C-----A--AG----A---ATG--G--A-TA----   4056

MAC.US.SMM251      AC------AT--TG---T-AAC--T-A---GA---G-----------A-C--AG----G--C-A-G-C-----AAAAGTGT--GAAC-G--T--T-----AC-AG-A--A---G-----T--CTCA-G-   4277
MAC.US.SMM239      AC------AT--TG---T-AAC--T-A---GA---G-----------A-C--AG----G--C-A-G-C-----AAAAGTGT--GAAC-G--T--T-----AC-AG-A--A--GG-----T--CTCA-G-   4301
MAC.US.SMM239 Pol  _G__D__G__S__G__N__K__Q__S__K__E__G__K__A__G__Y__I__T__D__R__G__K__D__K__V__K__V__L__E__Q__T__T__N__Q__Q__A__E__L__E__A__F__L__M__   p15

STM.US.STM         --G-----CT--TG-------C-AT-----GA---G--------C--CA-A---G---------A-A-C-----AAAAG-TT--GAAC-G--T--T-----AC-AG-A--A---G-----T--GCCA-G-   3948
SMM.US.SMM9        ------A-AT-CTG---T---C--T-A-G-GA------------C-----G--AG-------C---GGC--GACAAAACTG--GGAAC-G--T--C-----AC-AG-A-----GG----CT---------   3774

H2A.GW.ALI         AC---C---C-CTG---T---C-AT-A---GA----------------A-A--AG-------G---G-C--G--GAAGGTA---GAAC-A--T--C------C-AG-A--A---G----CT-CGCG--G-   4377
H2A.SN.ST          AC------AT-CTGC--T-A-C-AT-A-G-GA----------------CA-A--AG----------G-C--G--GAGGCTAT--GAGC-A--C--C------C-AG-A--A---G----CT--GCGA-G-   3829
H2A.DE.BEN         AC------AT--TGC--T--AC--T-A---GAG-----------------A--AG---------A-G-C-----AAAAGTAT--GAAC-A--T--C------C--G-A--A---G---TCT--CGGA-G-   4384
H2B.GH.D205        AC-------T-CTGC--T--AACCT-A--GGA-----------------C---G-C--G----A-G-T--G---AAAGTGT--GAAC-G--------C--AC-AG-A--AC-TG-----T--GCAT---   4357
H2B.CI.EHO         AC------AT-CTG-----AA-CCT-A---GA---G-------------C--AG-C--G----AGG-T------AAAC-AT--GAAC-A-----------C-AG-A---C-TG-----T--GCA----   4355
H2G.CI.ABT96       AC---C---T-CTGC--T-AAACCT-C--GGAR-----------C--C--C--AG-C-------A-G-C--G--AAARGT-T--GAGC-G--------C---C--G-R--AC-TG-G---T--GCAA-G-   3711

TANTALUS.UG.TAN1   TGC---------TGC--T--AA-TT-C-----------G-----C-TCC-A-----C-C----GA-TC-C-G--CA-AGAAT--GAG--T--C--G--C--AC-AG-A--A---------G-CCT-----   4201
GRIVET.ET.GRI677   --G-----------ATC--AA-T---C--------T--G---------C-GT-AG-G-------A-AGT-G-A--AGGGAAT--GAAA----C--T--C--AC-AG-A--A---AC----G--A-GA-G-   4005
VERVET.KE.VER9063  -----------CTG---T--AA-CT-A---GA----------------CA----AC-ATAT----A---G-G---A-AG-AAT--GAGC-G--T-----C---C-AG-A--A--GGT---C--AA-AA-G-   4229
VERVET.KE.VERTYO   ---...--A---TGC-----AA-TT-A---GA----------------CA-CT-AC-ATAC----A---G-G---A-AA--AT--GAAA----T--C------C-AG-AA-A---AC---T--AA-AA-G-   3716
VERVET.KE.VER155   --------A--CTG---T--AC--T-G---GAG-----------C--CA-A--CC-ACA---C-A-----G---ACAACAG---GAAA-------------AC-AG----AC-GAC---C--AA-AA-G-   4221
VERVET.DE.VERAGM3  --G-----A--TTG---T--AA-TT-A-GGGA----------------A-C--AC-ATAT--G-A-----GG--G-AA-AAT--GAAA-T--------C---C-AG-A--A---ATG--C--AA-AA-G-   3724
SABAEUS.SN.SAB1C   -----------------T--AA-TT-C---GA----------------CT-A--AG-------GGAT-----G--G--AG-AT--GAGA-T--T--C--C-----AG-A------G-G--C---CTGT---   4400

SUN.GA.SIVSUN      --G-----A------G-A-A-CTAGAA--GAG---G-----------A-A--CC-GG--------TCC-G----AAG-AAT-GGAAA----C--C------C-AG-A--AC--G-G--T--AA-AA-G-   4356
LHOEST.KE.LHOEST   --------A-----AG-A-A--TAGGA---AC----------------A-A--AC-ATC---G-A-G-G-----AAAGGAGT----------C---------C-AG-A--AC--G-G---G--CTAA-G-   4357
LHOEST.CD.447      --------A--T--AG-A-AA-TAGGG---AC------G-----T-----C--AC-ATC-----AGG-G------AAAGAAT----A-----C--C-----AC-AG-A--A---G----TG-C-TAT-G-   3273
LHOEST.CD.485      --------A--T--AG-A-AA-TAGGG---AC------------T-----C--AC-ATC-----AGG-G------AAAGAAT----A--T--C--C-----AC-AG-A--A---G-G--TG-C-TAT-G-   3273
LHOEST.CD.524      --G-----A--T--AG-A--A-TAGG----AC------------G---A-A---C-ATC-----A-G-G---A-AAAAGAAT--CAA--T--T--T------C--G-A--A---G-G---G-A-TGT---   3270
MANDRIL.GA.MNDGB1  --T-----A--------------C-G----A-G----------------A----AG-------TTTT----GG--A-AAGAATATCTAA-T--C--C------C----A--A-----T---G-AA-A----   3689

COLOBUS.CM.CGU1    AC------A-G-TGCTC----A-ATTAGG-GCG-----G-------GGA-A-A-TCA-A----GA-G--G----A----AT--TCA---AGGGT--------C-AG-A---C-GAC--G-G---TG----   3771
SYKES.KE.173       --------A--------T--A--C-GC--GAC------T-----------AG-A-G-GAT--G-C---G-G---ACAGTAT--GGAGC-A--------C--AC--G-A--A---G-G-GCT-GCTCA-G-   4202
RCM.NG.NGM         -------------A--T---A-T-G---G--G-----G-----------A--AG----G----A-G----G---AAAAGAAT--GAAA-T--T--C-----A--AG-A------GG----G-A-TA--G-   4351
MANDRIL.GA.14CG    --G-----------A--T--AA-CT----GGA---G-----------CTA---AGC---G----A-AGT-----AA-AG--T--GAAA-------T--------AG-A--A---G-G-----ACTGT---   3930
MANDRIL.14CG  Pol  _V__D__G__A__A__N__R__N__S__K__E__G__K__A__G__Y__Y__T__G__R__G__K__S__K__V__I__A__L__E__N__T__T__N__Q__K__A__E__L__E__A__I__L__L__   p15
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2       CTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCACAACCAGATCAAAGTGAATCAGAGTTAGTCAATCAAATAATAGAGCAGTTAATAAAAAA   4132
H1B.FR.HXB2  Pol  A__L__Q__D__S__G__L__E__V__N__I__V__T__D__S__Q__Y__A__L__G__I__I__Q__A__Q__P__D__Q__S__E__S__E__L__V__N__Q__I__I__E__Q__L__I__K__K   p15

CPZ.US.CPZUS      -A--AA-A-----A-AGAATAC---T--TG----C--T-G---C-------TC-----------C--CT-T--G-----C--------------A--G--------G--------AG-A--------G--   4192
CPZ.GA.CPZGAB     -C-----------A-ATCA-C-G--T--------G--T-------------TG-----G--T-----GT----G--------C-----------A-----------G--------AG-------T-----   4193
CPZ.CM.CAM3       -A--A--A---G-A-A-CATAT---T--T-----C--T--T--C-------TG-----G--A-----CT-C--G--------------G-----A--G--T----T---------AG----------G--   3543
CPZ.CM.CAM5       -A---------G-A-A-CATAT---C--T-----T--T--T--T-------TGC-------A-----CT-T--G--------------------A-----AGGGTCT--------AG-------------   3834
CPZ.CD.CPZANT     --C-AG-------AAC-GGCCC---C--T-----C-----T-------------G---G--T-G----GTACC------------------CCCC----GG-AG-------CC--A-A----------G   3571

H1A.UG.U455       -C-----------A----C---------T---------------G-----------G--------G-----------CAGG-----------AA-------------------A--C-----G----   3578
H1C.ET.ETH2220    -G--A--------A----C--------------------------G-----------G-----C-TG------------A-G------------A------------------A----------GC--   3524
H1D.ZR.84ZR085    ------------------------------T-------------C-------------------------A-G-----------A-------G-------------A--------------   3654
H1F.BE.VI850      -------------A----C------G-------------------------------------G--G---------A-G-----------AA---------------------A------C----   3472
H1G.SE.SE6165     -------A--C--AA---C------------------G--------G------C--------------------------AGG------G----A--------------------A---C----------   3529
H1H.CF.90CF056    -------A-----A--G--------G--------G-----T-----G------C---------------------C---A-G-----------------T---------------G-A--------G--   3479
H1J.SE.SE92809    ----A-G------A--GC----------------------T--------------G--G--T--A---------------A-G-----------A--------------------G----------G--   3446
H1K.CM.MP535C     ----A--------A----C--------------------------G--------------------------G------A-------------T-----T-----------------A-----------   3328
H101_AE.TH.CM240  -C-----------A----C---------T-------------------------------------G-----------CAGG-----------AG-------C------------G--C----------   3706
H1N.CM.YBF30      -C--A--A--G--A---CGG--T-----------C--T-----T--G-----TA-G-----A-----TT----G------A-------------A--G--G-GC-----------AG--C-C--------   3726
H1O.CM.MVP5180    -C-----------CAAGGAGC-----------------------------T---G--C-----ATCCT-C------ACA--G-----C--CCCTA----TC-G--G---------G-AC---C------   4162
H1O.BE.ANT70      -C--A--------CAA-GA-AC------T-----------T--------------G--CG----CTCCT-C-----TACA--G--------CCCTA----TC-G------------G-AC---C------   4186

MAC.US.SMM251     -A---ACA--C--A--GCC-A-GAC---T--TA--GT---T----------TTA-G-----A--AAC--G-TGC--TACAG--TCA--GAGCAG-C----T--C-----------AG-AA-G--T-----   4407
MAC.US.SMM239     -A---ACA--C--A--GCC-A-G-C---T--TA--GT---T----------TTA-G-----A--AAC--G-TGC--TACAG--TCA--GAGCAG-C----T-------------AG-AA-G--T-----   4431
MAC.US.SMM239 Pol A__L__T__D__S__G__P__K__A__N__I__I__V__D__S__Q__Y__V__M__G__I__I__T__G__C__P__T__E__S__E__S__R__L__V__N__Q__I__I__E__E__M__I__K__K   p15

STM.US.STM        -C---GC---C--A--TCC-A---C---T--T--GGT--------------TGA-G--------AAC--G---G--CACAG--TCG--GAGCA-------A-----G--------AG-AA-G--T-----   4078
SMM.US.SMM9       -C--AGCA-----A---CC-A---C---T--TA--GT---T--C-------TTA-G--C--AG-AGC--GT-----CAC-G--TCA---AGTAGA-----A--C--G---------A-AA-G--T-----   3904

H2A.GW.ALI        -AG-AACA--C--A--TCC-A---CC--T--TA--GT---T-----G----T-A-G-----AG-AGC--GC--G---ACAG--TCA---AATAGAA----A--------C-----AG-AA-G--------   4507
H2A.SN.ST         -AG-AACA--C--A--TCC-A-G-CC-----TA--GT--------------T-A-G-----AG-AGC--GC------ACAG-GTCA--GAGTA-AA----A--------C-----AG-AA-G--------   3959
H2A.DE.BEN        -AC--GCA--C--A--CCC-A-G--T--T--CA--GT---T-----G----T-A-G--G--AG-AGC--GC--G---ACAG-GTCA---AATAGAA----G--C--G--C-----AG-AA-G-----G--   4514
H2B.GH.D205       -A--AACA--C--A-A-CC-C----T-----CA--GT---T----------TCA-G-----A--AGCT-----G---ACAG---CA------CCAA----AGCAA-------T--AG-AA-G--C-----   4487
H2B.CI.EHO        -AC-A-----C--A---CC-C-G--C--T--CA--GT---T----------TCA-G-----AG-AGCT---------ACAG---CA------CC-A----A-GAG-------T--AG-AA-G--C-----   4485
H2G.CI.ABT96      -C--------C--A--TCC-A----C--T--CA--GTR--T--C--G----TCA-G--R-----AGC--GR------ACGG---CA-----CCC------A---A-------T--AG-AA-R--T-----   3841

TANTALUS.UG.TAN1  -A--AA-AC--AGTAAGAGCA----T--T--------G--T-----G----TCA-G------C-GAC---------TACA--------C--C-T-C----A-----------T--AG-AA-G--T--T--   4331
GRIVET.ET.GRI677  -A---G----CAGT---GA-A-T-----T-----C-----T--T-------T-A-GAAC---T-GAC----TGT---C-GG-----A-C---CCC-----GG-A--G------C-AGCCC----G-----   4135
VERVET.KE.VER9063 -C---G-----AGT---GA-A------------------T-----------CA-G-----AT-GAC------G--CACA--------C--CCCA-----AG-A--G--C----CATTAA-GG----G--   4359
VERVET.KE.VERTYO  -----G-A--CAGT--GCCTA-T--G----------------T---------A-G-----TT-GAC---------CACA--------T---CCA-----AG-G-----T----CCTTAA-G---C----   3846
VERVET.KE.VER155  -C---G-----AGC--CCCTA----C--T-----------T-----------GA-G--C--AT-GAC------G--CACA--G-----C--CCCAC----AG-A----------CA---A-GG--C-G--   4351
VERVET.DE.VERAGM3 -AC-AG-A---AGT--GCCTA-T-----------------T------------A-G-----AT-AACT--C-----CACA--G-----C---CCC---A-AG-A-----T----CA-TAA-GG--C----   3854
SABAEUS.SN.SAB1C  -C--AAGA--C--T---AGCA-------T---A------------------TA-G--T--T--AGC--GGG----TACAG--TCA--TAATA-CA----AC-A-----T------G-AC-------G--   4530

SUN.GA.SIVSUN     -C---G-----AGTA-GAGCTC---T--T-----------------G-----C---A--C-GC-CAGTAAGAG---CACAG---CA--T------C----A--GG-G--CG------TCA---GGC----   4486
LHOEST.KE.LHOEST  -A--A--A---AGCAATAGTA-------T-----------T----------T-A-GAA---AT-GTC-CA-AG----ACAG---CA---CACCCCA----A--AG-C--C-----A--A-GC-AGC-G--   4487
LHOEST.CD.447     -A--AAGA--CAGTAAGAGCAGG--C------------------------TGA--AA-G-TT-ATCTCA--GG--TACAG---CA---CATCCTG----A--A--G--T------G-A-GT-A---G--   3403
LHOEST.CD.485     -A--AA-A--CAGTAA-AGCAGG--C------------------------TGA-GAA-G--T-ATCTCA--GG--TACAG---CA---CACCCTG----G--G--G--T------G-A-GT-AG--G--   3403
LHOEST.CD.524     ----A--A---AGCAA--C-AC---T-----------T-----C-------T-A-GAAGG--T-AAGTCAGAG---TACAG---C---GCATCCTA----G--GA-C--C---A--G--C-GCAG--G--   3400
MANDRIL.GA.MNDGB1 --C-AG-A---AGT---AGTT-T--T--------------T----------T-G----T--AC-AGC-AGCAG---TAC-G---CA--TCACCCA----A--GG----------ATTAA-G-A-GG---   3819

COLOBUS.CM.CGU1   --C-AA-AC--GG-CC-AA-AG-A-G--TT-G---------AGT-TT-----------G--A---ACT-GG--G---T--G-TCAGCTTC--TCTAGTAGAGGGG-G--T--TC--GCAGGT--GGCT--   3901
SYKES.KE.173      -A--------C--AAA-GACA-------T------GT---T--------CT-T-AT--C--TC--ATGA--TGC--CAC-A-C-CA---CACCCAA----GG-G--G------C-AG--GCC--------   4332
RCM.NG.NGM        -C--AA-A--C--A---AA-AGG-----------------T--T-G-C---TG--T--C--AT-AGC--G-AGT-----CAC------GAGCCCCC----AC-A--G--T-----A--AA-G---GGG--   4481
MANDRIL.GA.14CG   -C---A-A-----A--TCCTAGG-C---T-----C---------------------C--AC--TC---CACT-------G-TCA--CAACCCTA----T-GAG-------CA-C--AA-G---GCC--   4060
MANDRIL.14CG  Pol A__L__K__D__S__G__P__R__A__N__I__V__T__D__S__Q__Y__A__L__G__I__L__S__A__T__P__D__R__S__D__N__P__I__V__R__E__I__I__N__Q__M__I__G__K   p15
```

```
                                                                                                     Pol p66 RT and Pol p15 RNAse H end \/ Pol p31 Integrase start
H1B.FR.HXB2          GGAAAAGGTCTATCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGA...ATCAGG......AAAGTACTATTTTTAGATGGAATAGAT...AAG   4250
H1B.FR.HXB2          __E__K__V__Y__L__A__W__V__P__A__H__K__G__I__G__G__N__E__Q__V__D__K__L__V__S__A__G__.__I__R__.__.__K__V__L__F__L__D__G__I__D__.__K_    p31

CPZ.US.CPZUS         A-----AAGT---A-CT---------------T-----------------------A--------C---------T-A--G..-----A.....-----T--T--CC-------C-------...--A   4310
CPZ.GA.CPZGAB        A-----AA----C--CT-C--------------------T--A-----------G-----G-----------------A---..------.....-----G-----CC------------C...-G-   4311
CPZ.CM.CAM3          ------AAGT--C--TT-------G--------T------G-----------------------------T---T----G..A-----A.....-----T--C-----------T-------..T---   3663
CPZ.CM.CAM5          ------AAGT--C--CT-------G----------------------------------------C----T---T-A--C..--T--A.....-----C--C--C--------T--C---...---   3952
CPZ.CD.CPZANT        A---C--A-T--C--CT-C-------------T-----C--A--------------G-----C------------CAA--G..-----A.....C----C--C---C-G--A--C-------...--A   3689

H1A.UG.U455          ------A-----C---T------------G--------G-------------G------------------T-----..------.....--G--G--------------G-------...---   3696
H1C.ET.ETH2220       -----G------C---T---------------T------------------------------------A---AG---..------.....-----G--G---C-----------------...---   3642
H1D.ZR.84ZR085       ------------C-----------------T-----------------------------------------AG---..------.....---------------G---------------...---   3772
H1F.BE.VI850         -----G------C---T---------------------G----------------------------------------..G----.....---A----G----------G-------...---   3590
H1G.SE.SE6165        ------------C---T---------------------G-----------------C-----------AG---..------.....------------C--------C-------...--A   3647
H1H.CF.90CF056       ------------C---T--------------------------------------------T---T-----..G---A.....-----G-----C-------G-------...--A   3597
H1J.SE.SE92809       ------------C---T-G------------------------------G---------------T-----..------.....-----G--G---C-------G-------...--A   3564
H1K.CM.MP535C        -----G-A----C--AT------------------------------------------------------..------.....-----G-----C-----------------...---   3446
H101_AE.TH.CM240     ---G--A-----C---T-------------------G--G------------------------------T-A---..------.....--G--G--------------G-------...---   3824
H1N.CM.YBF30         -----GA--T-----CT-T--------T-----T-----T-------------G--G-----C--------T--CT-A---..--T--A.....---A--T----CC-------T-----A...--A   3844
H1O.CM.MVP5180       ----CGA--G-----TA-------T--T--T--------C--A------------A--A-------------A--CAAA-AC..--T--A......-G---C--G--CC-G--A-----------...C--   4280
H1O.BE.ANT70         ----C----G-----TA-------T--T--C--T-----C--A------------A--A-------------A--CAAG-AT..--T--A......-G---C--G--CC----A--------C...C--   4304

MAC.US.SMM251        -TC-G-AA-T---G-A----------------------T--A--------CC--G--A----CC-CC----T---CAG--G..--T--A......C----T--C--C--G--AAAG-----G...CCA   4525
MAC.US.SMM239        -TC-G-AA-T---G-A----------------------T--A--------CC--G--A----CC-CC----T---CAA--G..--T--A......C----T--C--C--G--AAAG-----G...CCA   4549
MAC.US.SMM239 Pol    __S__E__I__Y__V__A__W__V__P__A__H__K__G__I__G__G__N__Q__E__I__D__H__L__V__S__Q__G__.__I__R__.__.__Q__V__L__F__L__E__K__I__E__.__P_    p31

STM.US.STM           ----GCAA-T---G-A-----------------T--------A---------C-GG-G------C-T--G--A---CAG---..--T--A......C-G--C-----CC----AAA------A...CCA   4196
SMM.US.SMM9          ----GCAA-T---G-A------------T-----T---RR---A----A----C--G--------C-TC----T---CAA-A-..--T--A......C----C-----C-----AAA------A...CCA   4022

H2A.GW.ALI           ----GCCA-----G-T--G-----C-----C--------C--A--G------C-GG-------CC-T-----A---CAG--C..-----A......C-----T-G---C----AAA------G...CCA   4625
H2A.SN.ST            ----GCAA-----G-T--------C-----C--T-----C--A---------C-GG-G------C-C-----A---CAG--C..-----A......C-----T----CC----GAA------A...CCC   4077
H2A.DE.BEN           ----GCA------G-T--------C-----C--T-----C--A--------CC-GG-------CC-T-----A---CAA--C..-----A......C-----T----CC-G--AAAG-----G...CCC   4632
H2B.GH.D205          A--GGCA--A---G-A-G------------T-----G---C-G--T--T---C-GG-------CC-CC----A---CAA---..-----A......C-G--CT-G--CC----AAA------A...CCA   4605
H2B.CI.EHO           ------AA-A---G-A-G------------T-----G---C-G--T--T---C-GG-------CC-CC----G--CCAA---..--T--A......C--A-C------C----AAA------A...CCA   4603
H2G.CI.ABT96         A---RCTC-A---G-A-G------------T--T---------A------T---C--G-G-----CCRTC----A--CCAG---..--T--A......C----------C--GR-AAA------A...CCT   3959

TANTALUS.UG.TAN1     A--TGCTA-----T-AA-C-------------T--------G--A--G------C--G--A-------GC----A--CCAG--G..--T--.......C----G-----C-----CA----G--A...G-A   4449
GRIVET.ET.GRI677     -AGGC-------CT-ACA-----------T--T--G--G--A-----C---AC-G--A-------------A--CAAA---..--A--A......C-GA-C--C--C------A-------A...G-A   4253
VERVET.KE.VER9063    AAGGGCAA-A--C--ACA------G--T-----------T--A--G---------G--A----C--------A---CAA---..--A--A......-G---T------A---GAA-------A...G-A   4477
VERVET.KE.VERTYO     -C--C-AA-A---T--CAG--------------T--------A-----------GG-GA-----------G---AAA--C..--T--A......-G---TT----C----AAA-------A...G-A   3964
VERVET.KE.VER155     A---GCCA--------CA-----------T--T--T-----T--A--G--C------G--A----C--------A--CAAG---..G-T--A......-G-A--T-G--CA-T-GCA-G-----A...G-A   4469
VERVET.DE.VERAGM3    AC-TC--A-A--CT--CA-------------G-------G--A-----C------G-GA--------C----A---CAA--G..--G---........---A-TT----------AAA------A...G-A   3972
SABAEUS.SN.SAB1C     ---GGCA--G---A-A-----------T--C--T------G-A--T--C-----GG--A-T------C----T---CAA---..--A--A.....C-------------CC-G--CA----T--A...G-A   4648

SUN.GA.SIVSUN        ---CC----A--C--A-GT--------G-------------A---------C--G--A-----C--C----T--CCAG---..-----AAAGAGAC-G---A-G---A----AAAG-----A...CCA   4610
LHOEST.KE.LHOEST     A--TC-A--T-----A-G------G--T--T--T--------A-----T---C--G-G-----CC-CC----A---AAA--T..--A--ACAGAAAC-G--CA-G---C----AAAG-----A...CCT   4611
LHOEST.CD.447        A--CC-A--G-----A-G-----------T--------T--A--G------C--G-------CC-CC--------AAA---..--T--ACAAAAAC-------G---C---GAAAG-----G...CCC   3527
LHOEST.CD.485        A--CC-A--A-----A-G-----------T--------T--A--G------C--G-------CC-CC--------AAA---..--T--ACAAATAC-------G---C----AAA------G...CCA   3527
LHOEST.CD.524        A--TGCA--T--C--A-G-----------T--T-----T--A--G-----CC-GG-GA-----C-------T-T-AAA---..--T---ACAAAAC-G--CA-G---C-G--GAAG-----A...CCA   3524
MANDRIL.GA.MNDGB1    A-----AA-T---T-AAGT---C----------------G--A-----G-----G---A-------GC----A---T-A---..-----A..........-----CT----CC--C-AAAT-----A...CCA   3937

COLOBUS.CM.CGU1      A---GCCA-TC--G-A--C---TGT--------T-----G--A-------------TGA-----C--AAG--A......---..G-A--A......C-G---A-G-GGA-----AA------A...GCA   4013
SYKES.KE.173         ----GCCA-A---G--A-------C-----C--T--------A-----------GCT--T------C--------AAA---..--A--A..........--A-C-----CC----AA----TCCC...C-A   4450
RCM.NG.NGM           ----G-A--G--C--CT-C--------G-----T-----C-----G--------G--G-------------G---CAG---..--T--A......C-G--------CC----C--G----GA...--A   4599
MANDRIL.GA.14CG      A---GCT--A--CT-A--------------T--T-----C--A-----C---------A----C--------A--CAAA---..-----A......C----------CC-G--A-----------...--A   4178
MANDRIL.14CG  Pol    __E__G__V__Y__L__A__W__V__P__A__H__K__G__I__G__G__N__E__Q__I__D__K__L__V__S__K__G__.__I__R__.__.__Q__V__L__F__L__E__G__I__D__.__K_    p31
```

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2        GCCCAAGATGAACATGAGAAA...TATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCA   4377
H1B.FR.HXB2  Pol   _A__Q__D__E__H__E__K__.__Y__H__S__N__W__R__A__M__A__S__D__F__N__L__P__P__V__V__A__K__E__I__V__A__S__C__D__K__C__Q__L__K__G__E__A__   p31

CPZ.US.CPZUS       --A--G--A--C-----C---...-----T----------C---T-----C--------------------A--G--C--T--G--G-----G----------------------C--------G----   4437
CPZ.GA.CPZGAB      --------A--------A-GG...-----T----------A---T-----------------TT-A-----CA--------------------G---CAT--------G--C---G---------------   4438
CPZ.CM.CAM3        CA-A-G--A--------..GGAAA-----T-AC-----------T-----A--------C--TA-A----A-A----G--C------C--------C-T-----------C--A----------G----   3791
CPZ.CM.CAM5        --A--G--A--------A--G...-----T-AC--C----A---T-----C--------C--T--A-----------G--C-----------------T-----C--G--C---T-------G-------   4079
CPZ.CD.CPZANT      --T-----G--C-----T---...-----T--C---------T--T-A--AGA---A-AC--T--T--C---A-T--G--T--------TA----ACAG--------------CG----G----------   3816

H1A.UG.U455        --T-----G--C-----A---...------T-C--C-----------------------------T-----------G-----G--G-----------------A----------A--------G-------   3823
H1C.ET.ETH2220     --T-----A--G-----A---...--------C-------------------A---A-----TA-C-----C------C-C--------------TT------------------------G-------   3769
H1D.ZR.84ZR085     --T-----G-----------...--------C-----------------------------------------G--------------------T-----------------------------A-   3899
H1F.BE.VI850       --A-----A--------A---...-------AC-----------G-----------------T---------A---------------------T-----------G--------------G-------   3717
H1G.SE.SE6165      --------A--G-------G-...-------AC--C--------------C-----------TT--------A-------------------G-----------------------------G-------   3774
H1H.CF.90CF056     --T-----A--------A-GG...-----T-AC------------G----------------T--A------A---------------------T-------------------------G-------   3724
H1J.SE.SE92809     --T-----A--T-----A---...-----T--C----------------------------T----------------------G--------T-----------------------G--G-------   3691
H1K.CM.MP535C      --T-----A--------A---...-----T-AC--C-------------------------T---------A-------C--------------T--------C-----------------G-------   3573
H101_AE.TH.CM240   --T-----A--------A-G-...--------C---------A-------------------TT--------A----------G---------A---A--------------A--------G-----T-   3951
H1N.CM.YBF30       --------A--T-----C-G-...--------C-------A---------C------------T-A--C--CA----G---------------------------C-----C-----------G-------   3971
H1O.CM.MVP5180     --A-----A--T-----A---...-----T---------------T-A--------C---GGAT-A-----AA-------C--G-----CA-T--T--T---CC------C--TA-------G-----A-   4407
H1O.BE.ANT70       --A-----A--T-----A---...-----T----------A----C-A--------A---GGA--A-----A--G--G--C--G-----CA-T--T------CC---------TA-------G-----A-   4431

MAC.US.SMM251      --A-----A--------T---...--C--T------GTA-A--A-T---TATTCA-A---GGAT-A--CAGAA----G--C-G-C-G-------A--C---------------T-AG-----------T-   4652
MAC.US.SMM239      --A-----A--------T---...--C--T------GTA-A--A-T---TATTCA-A---GGAT-A--CAGAA----G--C-G-C-G-------A--C---------------T-AG--------G--T-   4676
MAC.US.SMM239 Pol  _A__Q__E__E__H__D__K__.__Y__H__S__N__V__K__E__L__V__F__K__F__G__L__P__R__I__V__A__R__Q__I__V__D__T__C__D__K__C__H__Q__K__G__E__A__   p31

STM.US.STM         --T-----A--G-----A---...-----T--C---GT--A--AGC-A-TCTTCA-G---GG-A-A---AGGC----------GC-------G-AT-CA-----C--G--C--C-AG-----------T-   4323
SMM.US.SMM9        --G--G--A--G-----A---...--C--T--C---GTA-A-AA-T---TATTCA-A---RGTT-A--TAGGC-----------C---------A--CA-----------C--C--G-----------Y-   4149

H2A.GW.ALI         --T--G--A--------A--G...-----T--C---GT--A--A-C-AT-CCA-A-A---GGAT-A--CAA-C-G--G----G-C-G------AA--CA----CCC---------AG--G-----G--T-   4752
H2A.SN.ST          --T--G--G--------A---...-----T--C---GTA-A--A-C-AT-CCA-A-A---GGA-----CAAAT----G----G-C--------AA--CA--CACCC---------AG-----G--G--T-   4204
H2A.DE.BEN         --T-----G--------A---...-----T--C-T-ATA-A--A-C-AA-CCA-A-A---GGAA-A--C-T-C---------G-C-G------AA-TCA----CCC----C--A-AG-------------   4759
H2B.GH.D205        -----G--A--G-----A---...-----TG-C---GTA-A--A-C---TCCA-A-A--CGGAA-T----AAT----G------C-G------AATTC------------C--A-A------G-----T-   4732
H2B.CI.EHO         --T-----A--------A---...-----T-A----GTA-A--A-C-A-TCCA-A-A---GGGA-T----AAT----G----G-C--------AATTC------------C--A-A------G-----T-   4730
H2G.CI.ABT96       --T-----A--------A--R...-T---T------GTA-A--AGT-AA--CA-A-G---GG-A-T--T-AAT-----------C-G------AATTCA-ACC-C-GT--C--A-AG-------------   4086

TANTALUS.UG.TAN1   -----G--ATC------T---...-----T-C-------CA-TTT--AAGAGA--CA--CGGTA-A---G-AC-G--------------------TGCT--CCCA--------AA-TCG-------C-G-   4576
GRIVET.ET.GRI677   --A--------C----CA--G...--C--T-AC---------AGT----TACAG--A---GGAT-A--TAA-A-------------G--------GGCA---CCC-----C--AA---G-------C-T-   4380
VERVET.KE.VER9063  --T-----A--------C-G-...-----T----------A-AATT-A--GGACACA---GGAT-A----AGA----T--C-----G--------A-TG---CCA--TGCCA-AA-------G---C-A-   4604
VERVET.KE.VERTYO   --T-----AA-G-----A-G-...-----T-A--------A-AACC-A--AGA-ACA-A-GGG--T----AAA-------------G-----G----TG---CCA--------AA----G------C-AG   4091
VERVET.KE.VER155   --A-----A--------T-GG...-----------C------AATC-A--AGACACA---GGAT------AAA-------T--------T-----A-TG--CCCA--------AG-------G---C-A-   4596
VERVET.DE.VERAGM3  -----G--G--------A-GG...--C--T-A---------GAACT-A--AGACAC--A-GGG--A----AAA-T--G--------------------TG---CCA---------A-------G---C-AG   4099
SABAEUS.SN.SAB1C   --A-----A--G-----T---...--C--TGCC--------GAGT---CAACAA--A---GG-T-A--TG--A---------------------GGCA---CC------C---A-------G--GT-TG   4775

SUN.GA.SIVSUN      --TGT---A-------G---G...-TC--T-AC---GCAGCCT-TT--CAGGAGATG---G--A-A--C-TA--G--------GC-G-----CAATGAA----CAC-------A-AG-----G-------   4737
LHOEST.KE.LHOEST   --AGT---A------AGC---...-TC--T-AC---GCA-AG-ATC-A-AAGAAA-A--------A--C--CA-G-----C---C----T--CAATGA-----CA--C-----AAAG-------------   4738
LHOEST.CD.447      --A--G--A--GT--AGC---...-T---T-A----GCT-A--ATT-A-AAGAAA-A---C--A-A--C---A-G-----T--GC-------GAAT--T---A-----------AC---G----------   3654
LHOEST.CD.485      --A--G--A--G---AGC---...-T---T-A----GCT-A--ATC-A-AAGAAA-A---C----A--C---A-G-----T--GC-------GAA---T---A-----------AC---G----------   3654
LHOEST.CD.524      --TGT---A--G----GA--G...-T-----A----GCA-A--ACT-A-AAGAAA-G-A-GGGA-A--C--CA-G-----T--GC--------AATGAG----C---G-------AG--G----------   3651
MANDRIL.GA.MNDGB1  --A--G--A-------------...-----T--C---GAAGC-CA-T-AAGAGAGA-A--CC--T-A---G--C-------C---C-G--T--GCAA--T--CAG---G--CTGT-ATC-T-----GC---   4064

COLOBUS.CM.CGU1    --AG----A--C---C-A--G...-T---T------GT-CAGTATC-AAAAGAAC-A---GGGT-A--CA------------------TTGG-AA--A---TCAG-G------AAC------C----TG   4140
SYKES.KE.173       --A--G--A--T-----A-GG...--C---TCA---AT-GAGTATC-TAGGCAG--G---C-T--A--C-GACAG----------C---TA--CAACAA---CCA--------AAAT-G-------C-A-   4577
RCM.NG.NGM         --A--G--A--------A---...-----T--C----------------AGAA--A---C-AA-A--T-AAA----------------------CAA--CCCC--G------G-------------AG   4726
MANDRIL.GA.14CG    --T--G--G--------C---...--C--T-A----------T--T-AAG-CAG--A----G-A-A--C---A----G--------------T--ACAG--CCCA--G------G-----------C-T-   4305
MANDRIL.14CG  Pol  _A__Q__E__E__H__D__K__.__Y__H__N__N__W__R__S__L__S__Q__E__F__S__I__P__P__I__V__A__K__E__I__V__A__Q__C__P__K__C__Q__V__K__G__E__P__   p31
```

```
H1B.FR.HXB2        TGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAAAC   4507
H1B.FR.HXB2  Pol   M__H__G__Q__V__D__C__S__P__G__I__W__Q__L__D__C__T__H__L__E__G__K__V__I__L__V__A__V__H__V__A__S__G__Y__I__E__A__E__V__I__P__A__E__T    p31

CPZ.US.CPZUS       -A-----G--G-----T--C--------T--C-----G----------------------G---A-C--TA-A--G-----A-----G-----------CC----------G--C-----T--------   4567
CPZ.GA.CPZGAB      -------G--G-----------C-----G--T------G----------C--CC-------C-----G---A-A--G--------C-----------C-------------------C-----T--G--   4568
CPZ.CM.CAM3        -------G-----------C--C-----C---------T----C--C-----CC---------------------G-----A--------------G--CC----G-----G--C-----T-----G--   3921
CPZ.CM.CAM5        ---------------------------C--T------T----C---------C-----------A-C-----A--G-----A-----G-----------CC----G-----G--C-----T--------   4209
CPZ.CD.CPZANT      G-----------------C-----------------G-------C--C-----------T-----A---A-A--------G------T-T-----C-TC--------T-----A--GG---AT--G--   3946

H1A.UG.U455        ---------------------------G---------T-------C------C-------------A-----A--------C--------------C--C-----------------C-----------   3953
H1C.ET.ETH2220     -A------------A-T-----------G---------T-------------C--------G---A-C--------------C--------------C--C-----G-----G----------------   3899
H1D.ZR.84ZR085     ------------------------------------T----C-----------------------------------------------C-----------------------------------   4029
H1F.BE.VI850       ----C----------------------G---------T-------C-------------C--G----------------C--------T----AG--C-----------------------------   3847
H1G.SE.SE6165      ------------------------------------T--------------------------A-----A-A-----------------------C-------------------------------   3904
H1H.CF.90CF056     ----------------------C---------------T-------C--------G------C-------T-----------C--------------C-----------------C--C-----------   3854
H1J.SE.SE92809     ---------------------------G---------T--------------C-------------------------------------------C-------------------C-----------   3821
H1K.CM.MP535C      -A----------------------------------T--------------------------A---------------C--------T--C--C--------------------C-----------   3703
H101_AE.TH.CM240   ----------------------------G---------T-------C------C-------------C--------------C--C--G-----------------------------C-----------   4081
H1N.CM.YBF30       ----------G--CA-T-----------G-G------T-------------C-----G------A-C-----T--G--G--C-----G--------C--CT-------------------T-----G--   4101
H1O.CM.MVP5180     C------T----------AC--C----AG---------A-G-------------------C---A-C--AA-A--T--T--C--------A----AC-T-----------G--G--A-----------   4537
H1O.BE.ANT70       -T-----T--G--------C-------A-G--------A-------C------A-G-----C---A-C--AA-A--T--T--C-----G--A-----G-TC--------------G--A-----------   4561

MAC.US.SMM251      -A-----G--G---A-T-CAGA--T---G-CT------A-G--C-----C---C-----------A-AG--A-A--T-----A--------T-------TC--------------A------CA---G--   4782
MAC.US.SMM239      -A-----G--G-C-A-T-CAGA--T---G-CT------A-G--------C---C----G------A-A---A-A--T-----A--------T-------TC-----------G--A------CA---G--   4806
MAC.US.SMM239 Pol  I__H__G__Q__A__N__S__D__L__G__G__W__Q__M__D__C__T__H__L__E__G__K__I__I__I__V__A__V__H__V__A__S__G__F__I__E__A__E__V__I__P__Q__E__T    p31

STM.US.STM         -A-----G------A-TGCAGAGTT---G-CT-----GA-G--C---------C-------C--GA-A--TA-A--------A--------T-----G-T------G--------A--C---CA------   4453
SMM.US.SMM9        -A-----G------A-TGCRGRA-T---G-CT------A-G--C--------CC----G--C---A-A--TA-A--T-----A-----G-GT-------TC-----G--------A--C--GCRG-----   4279

H2A.GW.ALI         -A-----G-----GA-TGCAGAA-T---C-CT------A-G--C--C-----C--------------C---A-AA-------G-----T----------TC-----------G--C--C--GCAG---T-   4882
H2A.SN.ST          -A-----G------A-TGCAGAATT---C-CT------A-G--C--C-----C------------A-C--TA-A--------A-----T--A-------T---------------C--C---CAG---T-   4334
H2A.DE.BEN         -A-----G------A-TGCAGAAAT---CG-T------A-G--C-AC-----C------------A-C--TA-A--------A-----T--A-------TC--------------C--C---CAG---T-   4889
H2B.GH.D205        -T--------G---A-TGCAGAC-T---G-C------GA-G--C----------------------A----AA-A--G-----C--------------G-T------------G--A--A--CCA---G--   4862
H2B.CI.EHO         -T--------G---A-T-CAGAA-T---G-C-------A-G--C--------------G-----G-----AA-A--G---------------------TC--------------A--A--CCA------   4860
H2G.CI.ABT96       -T--------G---A-TGCAGAATTG--T-CT------A-G--C--C--C-----------------A---A-A--------------G--------C-T-G-------------A--C---CA------   4216

TANTALUS.UG.TAN1   -A-----G--------TGCC---GT---GG-C------A-G--C-----------G---------A-C--TA-A--------C-----G--------C-TCT-------------C--AG-TAGG-----   4706
GRIVET.ET.GRI677   A---------G------GCCTCCATT-A--CT-----GA-G--C--C--C--------------------AA-A--------A----------------TC-----------G--G--C---AG------   4510
VERVET.KE.VER9063  -A-----G---------GCCTCC-----GG--------A-G--C--C--G---G-----------A-AG--A-A--------------G-----C----TC--------------C--A--TAGG-----   4734
VERVET.KE.VERTYO   -------------G--TGCCTCA--T----C------GA-G--------T---C-----AA------AG--A-A--T--G--C----------------TC--------------C--A--TAGG-----   4221
VERVET.KE.VER155   -A--------------TGC-TCA------G-G-----GA-G--C--C------A-----------A-AG-GA-A-----G--C--------------G-T------------G-----C--TAGG-----   4726
VERVET.DE.VERAGM3  -C-----G--------TGCCTCG-----GG--------A-G--C-----------------C--G--A---A-A--------C----------------TC-----------------A--TAG------   4229
SABAEUS.SN.SAB1C   -A-----G--------TGC----------G-G------A-G-----------CC-----------A-C---A-A--G--T--------T--T-------T--------------A--C---CAG-----   4905

SUN.GA.SIVSUN      -AACA-----------TGCCTCAGTG--C--C-----GA-------C--C--CA-G----A------C--AA-A-----G--G-----G--TTCA--G-----G-T---------AC-A--CAAT--GCA   4867
LHOEST.KE.LHOEST   -AACA--------G--TGTCTCAGTG--T--T-----G-----C-----T--C---------C-G--C--TA-AAAT-----C-----G---TCA--C-TC--G-TT-----------A----AT--G--   4868
LHOEST.CD.447      -AACA-----------TAC-TCA-----T--T------A----C--C--T--CA-G--G---C-G--A---A-TAAT-----------G---TC-----T---G-T---------A-------AT-----   3784
LHOEST.CD.485      -AACA-----------TAC-TCA-----T--T------A----C--C--T--CA-G--G---C-G--A---A-TAAT---A-C-----G---T------T---G-T---------A-------AT-----   3784
LHOEST.CD.524      -AACA--T--------TGTATCAGTG--T--T--------G--------T--CC-------GC--A-C---A--AAT-----C---A-G--ATCA--C-TC--G-TT--------G-------AT-----   3781
MANDRIL.GA.MNDGB1  -AA-G-----GAC---TGC-TCA-TT---G-C-----GA-------C------C-G---AATC--A----TA-A--------G--------TTCA--C-TC--GA-G-----------AA----------   4194

COLOBUS.CM.CGU1    -T-----C--GC----T-A---CTAT--CT--------T----------T---GAG--------G------T-A--------A--C--CTGT-C-CTG-TCTGCTGG---ACCA-AT-AAAGAG------   4270
SYKES.KE.173       AA-----G--------TGTAGA-AT-TAT-AT-----GA-G--C-----T---GAG-----C-----A---TGT--------AA-CAC---T--------------A-CA-GA-CT-AAA-AGG-----   4707
RCM.NG.NGM         -A--------G-----TGC----------T-CT------A-G-----C------C-----------A-C---A-A--------C----------------C--------------A--A--------G--   4856
MANDRIL.GA.14CG    -C--------G--G--TGCGGCC-------C-------A-G-----C--C---C----------G------A-A--------C-----G-----------C-------------G--A-----------   4435
MANDRIL.14CG  Pol  I__H__G__Q__V__D__A__A__P__G__G__W__Q__M__D__C__T__H__L__E__G__K__V__I__I__V__A__V__H__V__A__S__G__Y__L__E__A__E__V__I__P__A__E__T    p31
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
H1B.FR.HXB2       AGGGCAGGAAACAGCATATTTTCTTTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACTGACAATGGCAGCAATTTCACCGGTGCTACGGTTAGGGCCGCCTGTTGGTGGGCGGGAATCAAG   4637
H1B.FR.HXB2  Pol  __G__Q__E__T__A__Y__F__L__L__K__L__A__G__R__W__P__V__K__T__I__H__T__D__N__G__S__N__F__T__G__A__T__V__R__A__A__C__W__W__A__G__I__K_   p31

CPZ.US.CPZUS      ---A--A--G-----C------A-C-----------------------T------GTG--T--------T-----ATCT--C--T--AA--AG---A-----A--A--T--C--------A--C--AC-A   4697
CPZ.GA.CPZGAB     ---A--A--------T-----C--G--------------T-----------------T--T--C--A--T-----GCCA-----T--AA-----G-A--C-A---T--------------A-AC------   4698
CPZ.CM.CAM3       T--A-----------T------A--C---------------------T------GT---C--C-----T-----GCCT--C--T---A--AA---A-----A---A-----C--------A--C---C-A   4051
CPZ.CM.CAM5       ---A--A--------T------A-----------------------T------GT------------T-----GCCT--C--T--TA--AAC--A--C-AA--A--------------T--C---C-A   4339
CPZ.CD.CPZANT     ----A-AAGT--------C--C--G-----------CA-C-----------------------C-----------AGCT--------AA----AG-A--A-AA--G--A-----------TAAT---C-A   4076

H1A.UG.U455       ---A-----G--------C---A-AC----------------------------GT------C--A---------------------A-C---G-A----AA--A-T------------AAAT---C-A   4083
H1C.ET.ETH2220    ---A--A-----------C-----AC-------------G-----------C-GGGT---------A--T--------T--C------A--AA-G-A----AA--A--------------A--T--TC-A   4029
H1D.ZR.84ZR085    ----------G-------C-----C----------------------------GT-G-G-----A---------------------A-------A----A---------C--------A--T------   4159
H1F.BE.VI850      ---A-----------C--C--CA-A-----G------------------------T---------A---------------------A----CG------A---AT------------A--T---C-A   3977
H1G.SE.SE6165     ---A--------------C---A-A-----------------G----------C-GT---------A--T------------------A-----G-A--A-A---A--A-----------AAAT----CA   4034
H1H.CF.90CF056    ---AA-------------C--C--G--G---C-----A-C--------------GT---------A-------------------GA-----G------A---A--------------A-AT---C-A   3984
H1J.SE.SE92809    ---A------G------T----A-A--------------C---------------GT---------A--------------C------A---G-G-T--G-A---A--------------A-AT------   3951
H1K.CM.MP535C     G--A--A--------C--C---A-AC----------------------------GT---------A----------C----------A-CA--GT-----A---A--------------A--TG-----   3833
H101_AE.TH.CM240  ---A-----G--------C-----GC-----C----------------------GT------C--A-----C--T------------A-C---G-A----AA--A--------------CAATG--C-A   4211
H1N.CM.YBF30      ---A------------------A-------G-----T------------------GTT-----C-----T-----ATC---------TA----C--T--A-AA--A--------------AAAT-----A   4231
H1O.CM.MVP5180    ---A--------T--C-----C--G------------C---------T--C---GT---------A--------ACCT-----T--AA----AG-CA-G-AA--T--A---------A-A--C--AC-A   4667
H1O.BE.ANT70      ---A--A-----T--C--C--C--G------C-G--T-C---------T--T---GT---------A-----C--GCCT-----T--AA--A-A--TA-G-A---T--A-----------CAAC--AC-A   4691

MAC.US.SMM251     ---AAGAC-G------CTA-----G--------G-----C--------TA-T-C-CATC----C--A--T-----TGCT--C--TG--TCGCAAGAA--A-A-ATG-TTGCA--------A--G--AG--   4912
MAC.US.SMM239     ---AAGAC-G------CTA-----G--------G-----C--------TA-T-C-CATC----C--A--T-----TGCT--C--TG-TTCGCAAGAA--A-A-ATG-TTGCA--------A--G--AG--   4936
MAC.US.SMM239 Pol __G__R__Q__T__A__L__F__L__L__K__L__A__G__R__W__P__I__T__H__L__H__T__D__N__G__A__N__F__A__S__Q__E__V__K__M__V__A__W__W__A__G__I__E_   p31

STM.US.STM        ----AGAC-G-------TG-----G--G--G------A-C--------T--C-C-CACC-G--C--G--T------GCT--C------TCGCAGGA---A-A-ATG-TAGCC--------------AG-A   4583
SMM.US.SMM9       ---AAGAC--------CTG-----GC----------CA-C--------CA-C-C-CATC-G--------T-----TGC---------ATCACAAGAA--G-AAATG-TTGC---------A--R--TG-A   4409

H2A.GW.ALI        ---AAGAC--------CTC--C--A------C-G--TA-T---------A---C-CACT-G--C--A--T-G---TGT---C------TCACAGGAA--A-A-ATG-TAGCA-------TT--T--AG--   5012
H2A.SN.ST         ---AAG-C----G---CTC--C--AC-----C-G--CA-T--G------A---C-CATT-G--C--A--------TGC---C-----TTCACAGGAA--A-A-ATG-TGGCA------ATA--T--AG-A   4464
H2A.DE.BEN        ---AAG-C-G------CTC--C--A------C-G--CA-T--G------A---CGCACT-G--C--A---------CC---C-----TTCACAGGAA--G-A-ATG-TGGCA-------TA--T--AG-A   5019
H2B.GH.D205       ---AAGAC-G-----TCTC--C--AC----G--G--CA-C--------TA-C-C-CACC----C--A-----C--TGC---C------TCAC-A-GT--A-A-ATG-TAGCC-------TA-----AG-A   4992
H2B.CI.EHO        ---AAGAC-G-----TCTC--C--G-----GC-G--CA-C--------TA-C-C-CACC-G--C--A-----C--TGC---C-----TTCACAAGAT--G-AAATG--AGCC------ATA--G--AG-A   4990
H2G.CI.ABT96      ---AAG-C---------TA--C--G--------------G--------TA-C-C-CACC-------A--------TGC---C------TCGCAAGA---G-A-ATG-TAGCC-------TA--C---G--   4346

TANTALUS.UG.TAN1  ---AA-A---------A-A--C--G------A-CATTA-T--G-----TA-T-C--A-T----------------GCCA-----TGTTA--CAGGA---ACAAA-TATT----------GACA-G-AG--   4836
GRIVET.ET.GRI677  T---A----G------C-C-----GC-G---C-GTT--C------------G---CATC----C-----T------CCA--C--T---TC-CAG-AT--GGCA--G-TG--C-------GTAAT--AG--   4640
VERVET.KE.VER9063 ---AAGA--G-----CA-A--C---C-----A--TTGA-T---------A--GTCCAGT----C--A--T-----ACCA-----T--TTCCCAAGA---AGC---AAT-----------GAAAG--AG--   4864
VERVET.KE.VERTYO  ---AA-A-----G---A-G-----A------A--CTGA-T--------TA---C-CAGT----C--A--------GCCT--C--T---TCCCAAGAA--GGCA--AATA----------GAAA---TG-A   4351
VERVET.KE.VER155  ---AA-A--G------A-G--CT-G------A--AT------------CA-C-CTCACC-C-----A--T-----ACCA--------TTC-CAGGAA--AGCT--TATG--C-------GAAAGG-AG-A   4856
VERVET.DE.VERAGM3 ----A-A---------A-A---T-A-----GA--CT-A-T--------CA---CCCA-C-G-----A--------ACC------T--GTC-CAAGAA--AGCA--AATG----------GAAA---AG-A   4359
SABAEUS.SN.SAB1C  ---AA-A-CT------C-C--C--GC-G--G-----CA-C--------CA-C-C-CAGC-G-----A--T-----T-C---C--T--TA-CCAGCAA--AGCA--TATT--C-------GAAA---AG--   5035

SUN.GA.SIVSUN     ----A--AC------CAC--GG--G-----GC-CTGT-CC-TG------------CAG--------A--T-----GCCA-------T-A--AAAGAT--GGA---A-TA---------TTA--C--AC--   4997
LHOEST.KE.LHOEST  ---AA-AAC---GT--C----C--G---------TGCA-T--------T------CA---------A--------TCCA-----TGTAA--AAGGA---ACA---A-TAACC------ATA--G--AG-A   4998
LHOEST.CD.447     ----A-AAC-------C-C--C--G------C-GTGCA-C------------C-GCA-G-C-----A--------ACCA--C--TGT-A-CAAGGA---ACAA----TAAT-------CTA--T--TG-A   3914
LHOEST.CD.485     ----A-AAC-------A-C--C--G------C-GTGCA-C------------C-GCA-G-C-----A--------ACCA--C--TGT-A-CAAGGA---ACAA--A-TAACC------CTA--T--TG-A   3914
LHOEST.CD.524     ----A-AACT------A-----T-G-----G---TGTA-C--------T------CA---T--C--A--T-----TCCA-----TGTGA--AAAGAT--ACA---A-TGACC------TTAA-T--AG--   3911
MANDRIL.GA.MNDGB1 T--AA-AA-G------G-G-----G-----G-------C-CA------TA-T-GT-A-C----C--A--T-----GCCT--C--T--TA--CAGGAA--AGAAA--ATG---------TTA--G--AG-A   4324

COLOBUS.CM.CGU1   ----G-A-------GTAGGGCC--CA-------G--TA-TCAG---GAG--T-G-CA-G-T--C--A--T-----GCCA-----TGTAA--CAGCACT---A---A--AGTG------CTA-----AGCA   4400
SYKES.KE.173      G---G-T----------TG-----CA-GC--A----CA-T---------A-----CAG-----C-----T-----ACCA--C--TGT-A---AC-A-T-C-AA--A--A---------TGT--C--AG-A   4837
RCM.NG.NGM        ---AA-A--G-----TC----C--G-------------CC-----------G---CATT----C--A--------A-CA--C--T--AA----AG-A--GCAA--A-A------------TCA---AG--   4986
MANDRIL.GA.14CG   ---AA-A---------C----C--GC-----C------------------G---CATT----------------TCCA--C--TGT-A---AG-A---AGCCA-A-TG-----------TCAG--AG--   4565
MANDRIL.14CG  Pol __G__K__E__T__A__H__F__L__L__K__L__A__G__R__W__P__V__K__H__L__H__T__D__N__G__P__N__F__V__S__E__K__V__A__T__V__C__W__W__A__Q__I__E_   p31
```

```
H1B.FR.HXB2        CAGGAATTTGGAATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAATGGCAG   4767
H1B.FR.HXB2  Pol   _Q__E__F__G__I__P__Y__N__P__Q__S__Q__G__V__V__E__S__M__N__K__E__L__K__K__I__I__G__Q__V__R__D__Q__A__E__H__L__K__T__A__V__Q__M__A__   p31

CPZ.US.CPZUS       --A--G-----------A--------A--------------G--------C-----------------------C--------AA-C--G-----A--A--G---T-A--------T-----G-------   4827
CPZ.GA.CPZGAB      --------------A-----T-----A--G-----------G--------CT-A--------GC----------A-----------T--G-----A--A------T-A--A-------------------   4828
CPZ.CM.CAM3        --------------------T-----A--------------G--------A--------G--------A--G--C--------AA-T-----A--A--A-----CT-G-G---------C--G-------   4181
CPZ.CM.CAM5        -----G--------C-----------A-----------------G-----C-----------------------C---------A-T-----C--A--A--G--G--A--------T--C----------   4469
CPZ.CD.CPZANT      --------------A--A--------A-----------------G-----C---------C----G---C-------------AA-T-----C--A--A-----AT-A------------GT--------   4206

H1A.UG.U455        -----------G--C-----------------------------G-----C-----C--G--------------C-----G-----------G--A--------C-----------------------   4213
H1C.ET.ETH2220     ------------------------------------------------------------------------C-----G-----------A--A-----G--C-----------------------   4159
H1D.ZR.84ZR085     ------------------------------------------------------------------------------------------------A-----------------------------   4289
H1F.BE.VI850       --------------C-------------------------------------A--------G-----A--G--C-------------------A-----------------------------   4107
H1G.SE.SE6165      ------------------------------C-------------------------G--------------C--C--G-----C--G-----A--------------------------G-------   4164
H1H.CF.90CF056     -----------G----------------------G-----------------------------------G--C-----G-----------C--A--A-----C-----------------------   4114
H1J.SE.SE92809     ------------------------------------------------------------------------C-----------------A--A--------C-----------------------   4081
H1K.CM.MP535C      -----------G--------------------C--------------------------------------C--------------G-----------------C--A-----------------   3963
H101_AE.TH.CM240   -----------G--C-----------------------------------------------G--------------C-----G-----------G--A--------C-----A-------------------   4341
H1N.CM.YBF30       -----------G--A-----------T--------G----C------G--C-------------------------------AA-C--------A--A--------A-----------G--------G-   4361
H1O.CM.MVP5180     --T--G-----G--A--A--T-----A---------------------G-C-----------------ATCT------CAG-----G--G--C--A--A--G---T-A--A-------------------   4797
H1O.BE.ANT70       --T--G--------A--A--T-----A---------------------G-C--------G--------ATC-------CAG-----G--G--C--A--A-----CT-A-GA------------------   4821

MAC.US.SMM251      --CACC-----GG-A--A--------A--G-----G--------G---G-A------C-CC-CC-G--A--TCAA----ATAGAA-C--G--A--A--AA-TTCAG-AG-A--CAT----TT--------   5042
MAC.US.SMM239      --CACC-----GG-A--A--------A--G-----G--------G---G-A------C-CC-CC-G--A--TCAA----ATAGAA-C--G--A--A--AA-TTCAG-AG-A--CAT----TT--------   5066
MAC.US.SMM239 Pol  _H__T__F__G__V__P__Y__N__P__Q__S__Q__G__V__V__E__A__M__N__H__H__L__K__N__Q__I__D__R__I__R__E__Q__A__N__S__V__E__T__I__V__L__M__A__   p31

STM.US.STM         --AACC------G-A--T--------A--G--C-----------G---G-C------C-TC-T-------CCCAG----ACAGAA-T--------A--AA-CACAG-AG-A--T-T-----TC-------   4713
SMM.US.SMM9        ---ACC---R-GG-A--T--T-----A--G--C---R-G-----G---G-G-----CC-TC-TC----A-CCCAA----ATAGAA-T-----A-----AA-TTCAA-AG-A--TAT-----T-------A   4539

H2A.GW.ALI         --ATCC------G-A--T--------A-----C--------------G-A------C-CC-CC----A--TCAG---A-CAGAA-T-----G-----AA-TACAG-GG-A---AT-----T--------   5142
H2A.SN.ST          --ATCC--C---G-A--T--------A-----C-----------G---G-A------C-CC-CC----A--TCAG---A-CAGAA-T-----G-----AA-CACAG-AG-A---AT-----T--------   4594
H2A.DE.BEN         --ATCC------G-A--T-----C--A-----C--G-----------G-A------C-CC-CC-------TCAG---A-TAGAA-T-----A-----AA-TACAA-AG-A---AT-----TG-------   5149
H2B.GH.D205        --AACT------G-A-----T--C--A-----------------G---G-A-----CC-TC-CC-G--A--TCAA----ACAGAC-C-----C--A--A-T-TCAA-AG------TT----T-------A   5122
H2B.CI.EHO         --AAC---C---G-G-----T-----AG-------G-----------G-A-----CC-TC-TC-G--A--TCAG----ACAGAA-T-----------A-T-TCAA-AG------TT--GTT-------A   5120
H2G.CI.ABT96       --A-C-------G-A--A-----C-----G--------G-----------A------C-TC-TC--------CAA----ATA-AA-T--------A--AA-TTCAA-AG-A--CAT---GTT--------   4476

TANTALUS.UG.TAN1   --CAC-ACA--G--C-----T--C--A---TCA------AGT------AG--------G-C-------AG---C-----A-A-AA-T-----AG-CTG--C-TT-T-GG-A--G-----C-TC-----CT   4966
GRIVET.ET.GRI677   --CACCAC------A--T--T--C--A--GTCA--G--TAGT------AGC-----C-G-C-GC-C---G----C--CTCT--AA--------G-TTG---GAGAT-GG----------G--------TA   4770
VERVET.KE.VER9063  --CACCACA---G-G--A--------A--GTCA--G---TC-------AGC-----C---C--C-C--AG----A--T---A-AA--------G-TTG-C-GT--ACAG-A----------TT------T   4994
VERVET.KE.VERTYO   --TAC-ACA--T--A--A--T--C------TC-------TC-A-----AGC-----C--GC-------AG-G--A--T--GA-AA--------G-TTGCC--T--ACAG--G---------TG-----TT   4481
VERVET.KE.VER155   --CAC-ACG--GG-A--A--T-----A--GTCC--G---TCTA-----AG------C---C----G--AG-G--A--T---A-AA--------G-CTG-C--T--ACAG-A----------TT-----CT   4986
VERVET.DE.VERAGM3  --CACCACA--TG-A-----T--C--T---TCA-----CTCTA----GAG----------C-G-----AG----A--T--GA-AA--------G-CTG-C--T-CACAG-A----------TT------T   4489
SABAEUS.SN.SAB1C   --CAC------GG-A--------C-----------G----------------------GC-G-----AG----C--------AA-T------G-T--A---AGAT-GG-A--------CAT--------   5165

SUN.GA.SIVSUN      --TAC-ACA--G--A--T-----C--T--G--------------G---G-A-A---C--G-T------AC-------CA-CAGAA-T-----AG-T--AC-GG-AT-G--A--T-------TC-----TC   5127
LHOEST.KE.LHOEST   --TAC-AC---G--A--T--T--C--T--G--------G--T------G-A-AA--C----TT-----A-GT-------A-AGA--------AG-T--AC-G--A--A--A----------T--------   5128
LHOEST.CD.447      --TAC-ACA--------T--T--C--T--G-----G--------T---G-C-AA-------C-C----AC---------A-AG---C-----AG-T---C----AT-A--------T--C-T-------A   4044
LHOEST.CD.485      --CTC-ACA--G-----T----------G-----G--------T---G-C-AA-------TGC-----C----C----A-AG---T-----AG-T---C----AT-A--------T--C-T--------   4044
LHOEST.CD.524      --CAC-ACA--G--A--A--T--C--T--------------------GG-A-AA-------TT------C----C----A-AGA--------AG-T--AC-G--G--A-----------GTT--------   4041
MANDRIL.GA.MNDGB1  --CAC---------A-----T--C--A--------G--G-----G---AA--AA-----GT-TC----AG--T-G--T-AGA-AA-------AG-TTGCA--G-AT-A--A---------GCC-----CA   4454

COLOBUS.CM.CGU1    --CACCACA---CAC-----T-----T---TCA--------T-----GCAG-GA--C--G--TG-T--A-G--AG--TAAGA-AA-G-A------A--A---ACAT-AG-A-GCAA----GC--------   4530
SYKES.KE.173       --CACTACA--T-----A-----C--A-----C--G---A-------G--C-AA----GGT-T-----AG--GC---CTC---AA-T-----CG-C-TCAC----T-GC-A--------GGC-------A   4967
RCM.NG.NGM         --CACC------G-A-----------T-----C--------G--------A------C-TC----G----T---A-----G--AA-T-----------A---A-GT-AG----------T---------   5116
MANDRIL.GA.14CG    --CACCACA-----A-----T--C----------------T--G---G-A-AA---C-TC-TC-T---GT--------AG--A--T--G-----A--A--GAGA--GG--------------------   4695
MANDRIL.14CG  Pol  _H__T__G__G__I__P__Y__N__P__Q__S__Q__G__V__V__E__A__K__N__H__H__L__K__V__I__I__E__Q__V__R__D__Q__A__E__R__L__E__T__A__V__Q__M__A__   p31
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
H1B.FR.HXB2        TATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTT   4897
H1B.FR.HXB2  Pol   V__F__I__H__N__F__K__R__K__G__G__I__G__G__Y__S__A__G__E__R__I__V__D__I__I__A__T__D__I__Q__T__K__E__L__Q__K__Q__I__T__K__I__Q__N__F    p31

CPZ.US.CPZUS       -------T----------------------------------T-C------A--------CA-------C------T----AC-------AG-CTT----------------TT----G----------   4957
CPZ.GA.CPZGAB      -G-----T-------------------------------------C----------------A-------------------------A-GC-----------------TT----G-------A--   4958
CPZ.CM.CAM3        -G-----T-----------------------------------C------A--------TA-------C--------------------A-CTA-TC--------------TT----G----------   4311
CPZ.CM.CAM5        -C-----T----------------------------------T-C------A-----G--TA-------C--------------------A-CTA-T---------------TT----G----------   4599
CPZ.CD.CPZANT      -GCA---T-------------------------------------CAC-T--AC-G------T--------C----------------G--A-CTC----------T------TT----------C-A--   4336

H1A.UG.U455        -------T--------------------------------A--------------------A-------------------------------------------------T----------------   4343
H1C.ET.ETH2220     -------T---------------G-----------------------------------A----T---------T----------G---------C-C-----C------TT----------------   4289
H1D.ZR.84ZR085     -------T-----------G-------------------------------------A--------------T----T----------G-----------------C----------------   4419
H1F.BE.VI850       -------------------------------------------------------A-----------T-----------------G--------------------G--------------   4237
H1G.SE.SE6165      -------T-----------------------------------------A---------A--------------T----T--------------C-------------------------------   4294
H1H.CF.90CF056     -------T-----------------------------------------------------A-------------------------------------------------T----C--------A--   4244
H1J.SE.SE92809     -------A-------------------------------A--------------------A-----------------------------G-----------------------------------   4211
H1K.CM.MP535C      -------------------------------------------------------------T-----------------------------------------TT---C--------A--   4093
H101_AE.TH.CM240   -------T-----------------------------------------------------A-------------------------------------------------------------   4471
H1N.CM.YBF30       -T-----T-----------------------------------C----------------A-------------------------G--A-C-A-T-------C-------TT----G----------   4491
H1O.CM.MVP5180     -C--TG-T-----------------------------------C---------G---C--A--------C-----T--C-A--------A-C----C--------------TT------CA-C-----   4927
H1O.BE.ANT70       ----TG-T-----------------------------------C------A-----G---A--------T-----T--C-A--------A-C------------------TT----N----C--A--   4951

MAC.US.SMM251      -TCATTG-ATG-------------GG--A--A--A----ATATG-C-C---CA------T--A-TA----G--CA-T-----ACA-G---TAC--TTTCA----TC-A---ACT-----T--A-------   5172
MAC.US.SMM239      -TCATTG-ATG-------------GG--A--A--A----ATATG-C-C---CA------T--A-TA----G--CA-T-----ACA-G-G-TAC--TTTCA----TC-A---ACT-----T--A-------   5196
MAC.US.SMM239 Pol  V__H__C__M__N__F__K__R__R__G__G__I__G__D__M__T__P__A__E__R__L__I__N__M__I__T__T__E__Q__E__I__Q__F__Q__Q__S__K__N__S__K__F__K__N__F    p31

STM.US.STM         -TCA-TG-ATG--------G-A--GG--A--AC-A----ATATG-CCC---CA------T----TA-T--G--CA-C-----ACA-G---TAC--TTTCA----TC-A---A-T-----T-CA-------   4843
SMM.US.SMM9        -TCATTG-AT--------------GG--A--A--A----ATATG-C-C---CA------T----CA-T--G--TA-C-----ACA-G---TAC--TTCCA----TC-A---A-T-----T--A-------   4669

H2A.GW.ALI         -TCATTG-ATG-------------GG--A--A--A----ATATG-CCC---CA------C-CA-CA-T--G--CT-C-----ACA-G---TAC--TTCC-C----C-A---A-TTG---T--A-------   5272
H2A.SN.ST          -TCATTG-ATG-----------G-GG--A--A--A----ATATG-CCC---CA------C--A-CA-T--GG-CA-TG----ACAGG---TAC--TTCC-C---GC-A---A-T-----T-A--------   4724
H2A.DE.BEN         -TCATTG-ATG-------------GG--A--A--A----ATATG-CCC---CA------C--A-CA----G--TA-C-----ACA-G---TAC--TTCC-C----G-A---A-T----TT--A-------   5279
H2B.GH.D205        CTCA-TG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA------C----TA----G---A-C-----GCA-G---TAC-GTTC--C---GC-A---A-TT----T----------   5252
H2B.CI.EHO         CTCA-TG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA-----------CA----G---A-T-----ACA-G---TAC--TTCC-C----C-A---A-TT----T-C--------   5250
H2G.CI.ABT96       --CATTG-ATG-------------GG--A--A--A----ATATG-CCC-T-CA---------A-TA----G--TA-T-----ACA-G---TAC--TTTCA-----C-A---A-T-----T--A-------   4606

TANTALUS.UG.TAN1   GCCA---T-----------------G--A--A--A-----TATG-CCC---CA--G--GC--A--A----G---A-C---C-AC--G---TAC--C-CA------C----CAACA-------TC------   5096
GRIVET.ET.GRI677   CGCAT--------------------G--A--A--A-----TAT-TC-AGT-CA------T-G--TA-T--GC--A-----C-AC--G--CTA--TACTC-------C-----CCA-------TTG-----   4900
VERVET.KE.VER9063  GCCA--------------------G--A--A--A-------TA-CAC---CA------C--A-TA-T--G---A-----C-AT--G--CTAC--AC--------CCA-----CA-------TT------   5124
VERVET.KE.VERTYO   GCA-AC-T-----------------G--A--A--A-----AC-G-C-T---CA--G---C--A-TA-T------A-----C-AT--G---TAC--C-T-------CCA-----CA-------TT------   4611
VERVET.KE.VER155   GCCA---T-----------------G--A--A--A------CTA-CA--T-CA--G---C--A--A-T--G---A-----C-AT--G---TC--CACTC------CCA----CCA-------TTG-----   5116
VERVET.DE.VERAGM3  GCCA--------------------G--A--A--A-------TA-CAC-G-CA--G---T--A-CA-T--G--TA-T---C-AT--G--TTAC--C-CC------CCA-----CA-------TT------   4619
SABAEUS.SN.SAB1C   -GCA---T-----------------------------------TC-----CA------C--A-CA-T------CAT-----AC--G----C---AC-C-T---C--A-----T----------------   5295

SUN.GA.SIVSUN      -GCA---T--T---------CA--GG--A--AC-A-----AATG-CA----CA--G--GT-TA-CA----G--TAATG----TC--G----AC--T-C-----G---AT--A-T--------TT---A--   5257
LHOEST.KE.LHOEST   -CCA---T--T---------CA--GG--A--AC-A------CTT-CAC---CA--G--GT-TA-TA-T--G--TAATG----AC--G----AC--T-TC--------TT--A-T--------TT---A--   5258
LHOEST.CD.447      C-CA---T--T---------CA--GG--A--AC-A------CT-TCAC---CA-----GT-TA-TA-T--G--TAATG-----T--G----AC--T-T----------T--A-T--------TT---A--   4174
LHOEST.CD.485      --CA---T--T---------CA--GG--A--AC-A------CTTTCAC---CA-----GT-TA-TA-T--G--TAATG-----T--G----AC--T-T----------T--A-T--------TT---A--   4174
LHOEST.CD.524      -CCA---T--T---------CA--GG--A--AC-G-------TATCAC---CA--G---TACA-CA-T--G--CAATG----TT--G-----C--T-CA--------TT--A-T--------TT---A--   4171
MANDRIL.GA.MNDGB1  C------T--T---------CA--GG--A--AC-A------ATG-CA-----A--G---------A-T--G--CAAT-----AT--G--TA-C--T-TCA------T------T------A-TT------   4584

COLOBUS.CM.CGU1    -C-ATGCG-T---------------G--A--AT-A------A-G---C--T--------CA------A-G-GC-AT--T---AT--G-T--AC--A-CC--AC----TT-CAA-ATC---AATTT--AAA   4660
SYKES.KE.173       CC------TTA-------------G--A--A--A------ATATCAC----A-----GTACA-TA-T--GC--TAT-----AC-----TTAC--C--AATAC--C-TC-...C-----T--TCG-----   5094
RCM.NG.NGM         -GC----T-------------------------------------------A-----G---A--------------T-----T-------A-CCA----------C------T----------------   5246
MANDRIL.GA.14CG    --C----T-------------------------A----A-------C----A----------G--T------A-------T--T-T-----CCA--------C--A-T---T----------------   4825
MANDRIL.14CG  Pol  V__L__I__H__N__F__K__R__K__G__G__I__G__E__Y__S__P__G__E__R__I__V__D__I__I__T__T__D__I__L__T__T__K__L__Q__Q__N__I__S__K__I__Q__N__F    p31
```

```
H1B.FR.HXB2       TCGGGTTTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGT...GACATAAAAGTAGTGCCAAGAAGAAAAGCA   5024
H1B.FR.HXB2       __R__V__Y__Y__R__D__S__R__N__P__L__W__K__G__P__A__K__L__L__W__K__G__E__G__A__V__V__I__Q__D__N__S__.__D__I__K__V__V__P__R__R__K__A_   p31

CPZ.US.CPZUS      ------C------------------G-----A----------------C--A--T--G-------------------------CA-G--A--CGAG...--GG-T--------A--C------------   5084
CPZ.GA.CPZGAB     ---------------A---------G-C---A-------------G----CCT-A--G----------------------G--C--G---C-AG-G...--AC----G-----A--------G------   5085
CPZ.CM.CAM3       ------C------------------G----CA-------------T--C-GA--T--G--------------------G-----TA----G-GAGAG...--AG-T------A-A--C-----G------   4438
CPZ.CM.CAM5       ------------------------G-----A-------------T--C-GA--T--G----------------------G--TA----G--CGAG...--AG-T--------A--C--------G--T   4726
CPZ.CD.CPZANT     -------C----------T------G-C--TG-G----------------C-A--T--G----------------------C--CA----CC-AGAG...--A--T--G-----C--TC----G------   4463

H1A.UG.U455       ------------------------G-C--CA-----------------A--A-----------------------------------C------...--T-----G-----A---------------   4470
H1C.ET.ETH2220    ---------------A---------G-C--TA----------------C--A--A-----------------------------------------...--------G-----A-----G--G------   4416
H1D.ZR.84ZR085    ------------------------G-----A----------------------T-----------------------------------C------...--------G-----A---------------   4546
H1F.BE.VI850      C-------------------------C---G---------------------A--------------------------C--------C------...--A-----GA----A-----G--G------   4364
H1G.SE.SE6165     ------------------------G-C---G----------------C--A--A-----------------------------------C----AC...--A-----G-----A---------------   4421
H1H.CF.90CF056    ------------------------G-C---A------------------A-----------------------------------------C------...--A-----------A---------G-G---   4371
H1J.SE.SE92809    ------------------------G-C---A------------------A--A-C-------------------------------G--C------...--A-----G-----A---------------   4338
H1K.CM.MP535C     ------------------------G-A---A---------------------A-----------------------------------C--C---...--A-----G-----A---------------   4220
H101_AE.TH.CM240  ------------------------G-C---A------------------A--A-----------------------------------C------...--T-----------A---------------   4598
H1N.CM.YBF30      ---------------A---------G----CA----------------C--A--T--G--------A----------G-----T--------CG-G...--T-----------C---C-T--G------   4618
H1O.CM.MVP5180    ------C--------A--T------G-C--TA-------------G---C-A-----G----------G----------C-----------AG-A...-----T-----G--A----------G---   5054
H1O.BE.ANT70      ------C--------A---------G-C--TA-C-----------G---C-------G----------G--A--------C-----------GG-A...-----T--G-----A--------G--G---   5078

MAC.US.SMM251     ------C--------A--AG-----G---A---G-----G-----C-GTG----AT-G--------G-----A-----CA-CT--A-G-TAGGG-CA...-----T--G-----A--C--------G--T   5299
MAC.US.SMM239     ------C--------A--AG-----G---A---G-----G-----C-GTG----AT-G--------G-----A-----CA-CT--A-G-TAGGG-CA...-----T--G-----A--C--------G--T   5323
MAC.US.SMM239     __R__V__Y__Y__R__E__G__R__D__Q__L__W__K__G__P__G__E__L__L__W__K__G__E__G__A__V__I__L__K__V__G__T__.__D__I__K__V__V__P__R__R__K__A_   p31

STM.US.STM        ------C-----------AG-----G---A---A-----G-----C-GTG----AT-G--------G-----A-----CA-CG--A-G-TAGGG-CA...-----T--------A-----G-----G--T   4970
SMM.US.SMM9       ------C--------A--AG-----G-C-AG--G-----------C-GTG----AT-G--------G-----A-----CA-CC--A-G-TAGGG-CA...--G--C--G-----A-----G--G-----T   4796

H2A.GW.ALI        C-C---C-----------AG-----G---AG--G-----------T-GGG----A--G--------G--C--A-----CA--G-TA-G-TAGGG-CA...--------------A--------G--G--C   5399
H2A.SN.ST         ------C----T---A--AG-----G---AG--G-----------T-GGG-A--A--G-----G--G--C--A-----CA--G-CA-G-TAGGGGC-...----------A--A-A--------G-----T   4851
H2A.DE.BEN        C-A---C--------A--AG-----G---AG--G-----------T-GTG-A--A--G-----G--A-----A-----CA--G-CA-G-TAGGG-CA...--------------A--------G--G--C   5406
H2B.GH.D205       C-A---C--------A--AG-----G---A---C-----G-----T-GTG-A--AT-G--------G-----A-----CA-C---A-G-TAGGG-CA...--A--C--------A--C--G---------   5379
H2B.CI.EHO        C-----C--------A--AG-----G---A---C-----G-----T-GTG-T--AT-G--------G-----A-----CA-C---A-G-TAGGG-CA...--A--C------A-A--C------------   5377
H2G.CI.ABT96      ------C--------A--AG-----G-C-AG--C-----------T-GTG----TT-G--------G-----A--T--CA-C---A-G-TAGGG-CA...--A--C-----T--A--C------------   4733

TANTALUS.UG.TAN1  -AA---C--C-----A--AG-A---G-G---G-G--------------C-CT-----------G--A-----A--G-----C---A----AGG-GAG...--GT-G--------A-----G-----G---   5223
GRIVET.ET.GRI677  -AA---C--C-----A--AG-T---G-----G-G--------------GCGA---A----------A-----C--G--G-----TA----GGGGGAA...-----C--G-----C--C--G-----G--T   5027
VERVET.KE.VER9063 -A-A--C--C-----A--AG-G---G-----G-G-----G---------C-AT-AA----------G-----T-----G--C--CA----AGGAGAA...--GC-G---A-T--C---------------   5251
VERVET.KE.VERTYO  -A-A--C--C-----A--AG-G---G-C--TG-G-----------G---C-AT-AA----------G-----A-----G--CC-CA-G--CGGA---...---C----G--T--A--------G-----T   4738
VERVET.KE.VER155  -A-A--C--C-----A--AG-----G-----G-G-----G-----T--TCGC--GA----------A-----C--G-----TC-CA-G--AGG-GAA...--AC-G--G-----T--G-----G------   5243
VERVET.DE.VERAGM3 -A-A--C--C-----A--AG-G---G----TG-C-------------G-C--T-AA-T--------G-----T-----G--C--CA---GAGG-GTG...--AT------A-TAC--------G------   4746
SABAEUS.SN.SAB1C  ------C-----------AG-----G-C--TG-G--------T--G--T------A----------A-----A--------T--------AC-GG--...--AT-G---AC-A-C----------C   5422

SUN.GA.SIVSUN     -AA---C--C-----A--AG-A---G----T-AG-----------T--T--A--T--G--------G-----A--------C--CA--C-GGG-GAG...A----CCTG-----A--T--G---------   5384
LHOEST.KE.LHOEST  -AA------------AC-AG-A---G----T-AG--------------GC-A--T--G--------A-----T--T-----GG--A----AGGGGAG...A----CTTCTC---C--C------------   5385
LHOEST.CD.447     -AA------C-----AC-AG-AC--G----C-AG-----G-----T---C----T--G--------A-----T--T-----GG--A----AGGAGAC...AC---CTTCTC---C--C-----------T   4301
LHOEST.CD.485     -AA------C-----AC-AG-AC--G----T-AG-----G-----T---C-AA-T--G--------A-----T--T-----GG--A----AGGAGA-...AC---CTTCTC---C--C-----------T   4301
LHOEST.CD.524     -AA------C-----AC-AG-A---G------AG----G------T--TC-A--T--G--------A-----T--T-----GG--A-G--GGGAGAG...A----TTTCTCTA-C--C--G--G------   4298
MANDRIL.GA.MNDGB1 -AA------C-T---A--AG-A---G---AG--G-----------T-GT-TC--TT-G--------A--------------TT--A--T--C-AGAA...--G-----GA----A--T--------GTGT   4711

COLOBUS.CM.CGU1   -TTTAAGGCC---T--A-AGAGCAC-CAGG-GAG---C---------G-G-A--GG-G--------------A--G-----C--CAG-A--TC-CAAGGTAC-T-GTTT--GAAA--T----------T-   4790
SYKES.KE.173      -A-------C---C-CC-G...G-G--GAACGAG-----G--------TCGT--A--T--------------T--------GG-G---ACAG-GGAGGGA-----TTTT-C---A----------G---   5221
RCM.NG.NGM        ----------T---A--AG-A--GG---AG-AG-----G--------T--A---A-----G----A--------T--C--G--C--G---C-AGAG...--TC----G-----C--T--------GTGT   5373
MANDRIL.GA.14CG   --------------A--AG-A--GG-G-AG--C--------------G-A---A-T--------A-----A-----G-----CA----AGGA-CA...----T-------T-----C----------T   4952
MANDRIL.14CG  Pol __R__V__Y__Y__R__E__G__R__E__Q__L__W__K__G__P__A__E__L__I__W__K__G__E__G__A__V__V__I__K__E__G__T__.__D__L__K__V__V__P__R__R__K__A_   p31
```

HIV-1/HIV-2/SIV
complete genomes

```
                                                                      Pol p31 Integrase end \
H1B.FR.HXB2        _K__I__I__R__D__Y__G__K__Q__M__A__G__D__D__C__V__A__S__R__Q__D__E__D__$_                                             p31
                                  / Vif start
H1B.FR.HXB2                       _M__E__N__R__W__Q__V__M__I__V__W__Q__V__D__R__M__R__I__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   Vif
H1B.FR.HXB2        AAGATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAG.........................................................   5096

CPZ.US.CPZUS       --A--T--A-AA--C---------------------C----A--A-----G-------------AG---A.GGCTTGGAATAG............................................   5168
CPZ.GA.CPZGAB      -----------A-----------------------------------------------A----------A.GACTTGGAATAG............................................   5169
CPZ.CM.CAM3        --A--A-------------------------T---------A--A-----G--G-----------AG-C-A.GGCCTGGAATAG............................................   4522
CPZ.CM.CAM5        -----A--C--A---------------------------G-A--A-----G--------A--------C-A.AACCTGGAATAG............................................   4810
CPZ.CD.CPZANT      -----T--A-A---G---A-----A----A-A--A-AGA---GA-T-----G--------A----------A.TATCTGGAAATC............................................   4547

H1A.UG.U455        ---------------------------------------------A-----G------------A------.........................................................   4542
H1C.ET.ETH2220     --A---------------------------------C------------G------------A------.........................................................   4488
H1D.ZR.84ZR085     ----------------------------------------------------------------------A.........................................................   4618
H1F.BE.VI850       ---------------------------------T--------------G------------------A.........................................................   4436
H1G.SE.SE6165      -------------------------------------------------G-------------------.........................................................   4493
H1H.CF.90CF056     --A-------------------------------------------------------------------A.........................................................   4443
H1J.SE.SE92809     -----------A------------------------------------G------------------A.........................................................   4410
H1K.CM.MP535C      -----T------------------------------------------G------------------A.........................................................   4292
H101_AE.TH.CM240   -------------------------------------------------G-------------------.........................................................   4670
H1N.CM.YBF30       --A--A--------------------------------G------------G----------AA--C--.GAAATGGAATAG............................................   4702
H1O.CM.MVP5180     --A--A--C--A-----------------------AC----A--A------A-------ACA--AAG-G-A.AGCATGGAACAGCCTGGT..GAAATACCATAA.........................   5156
H1O.BE.ANT70       --A--A--C--A--G--------------------AC----A--A--------G-----ACA--AAG-G-A.AGCGTGGAACAGCCTAGT..GAAATACCATAA.........................   5180

MAC.US.SMM251      --A--T--C-AA---------GG-GGA-AA-AG-TG---AGCA--TCCCAC-TGGAGG-TACC-GA--GGCT...AGAGAGGTGGCATAG...CCTCATAAAATATCTGAAATATAA.............   5410
MAC.US.SMM239      --A--T--C-AA---------GG-GGA-AA-AG-TG---AGCA--TCCCAC-TGGAGG-TACC-GA--GGCT...AGAGAGGTGGCATAG...CCTCATAAAATATCTGAAATATAA.............   5434
MAC.US.SMM239 Pol _K__I__I__K__D__Y__G__G__G__K__E__V__D__S__S__S__H__M__E__D__T__G__E__G__.__R__E__V__A__$_                          p31
MAC.US.SMM239 Vif                 _M__E__E__E__K__R__W__I__A__V__P__T__W__R__I__P__E__R__L__.__E__R__W__H__S__.__L__I__K__Y__L__K__Y__K__.__.__.__.__._   Vif

STM.US.STM         --A-----C-AA---------GG-GGA-AA-AG-TG---AG-G--TCCCACCTGGAGG-TACCAG---GGCT...GGAGAGGTGGCATAG...CCTCATAAAACATTTAAAGTATAA.............   5081
SMM.US.SMM9        --A--T--C-AA--C------GG-GGA-AA-A-TTG---AG-G--TCCCACTTGGAGG-TACC-G---GGCT...AGAGAAGTGGCATAG...CCTCATCAAACACCTGAAATATAA.............   4907

H2A.GW.ALI         --------C--A--C------GG-AG-CAA-AGCTG---AG-G--CCCCACTTGGAGGGT-CCAG---GG-T...GGAGAAGTGGCATAG...CCTTGTCAAGTACCTAAAGTACAG.............   5510
H2A.SN.ST          --------C-AA--C------GG-AG-CAA-AGATG---AGCG--TCCAACTTGGAGGGT-CCAG---GG-T...GGAGAGGTGGCATAG...CCTTATCAAGTATCTAAAATACAG.............   4962
H2A.DE.BEN         -----T--C-----C------GG-AGACA--A-CTG---AG-A--CCCCACCTGGAGGGT-CCAG---GG-T...GGAGAAATGGCATGC...CCTTGTCAAGTACCTGAAATACAG.............   5517
H2B.GH.D205        --A--T--A---C-C------GG-GGA-AA-G-TTG---TG-A---CC-AC-TGGAGG-TACCAG-C-GGCT...AGAGAGATGGCACAG...TCTGATTAAGTATCTTAAGTATAG.............   5490
H2B.CI.EHO         --------A---A-C------GG-GGA-AA-A-TTG---TGCA---CC-ACGTGGAGG-TACCAT-C-GGCT...AGAGAGGTGGCACAG...TCTAATTAAATACCTGAAATATAG.............   5488
H2G.CI.ABT96       --A-----A-AA---------GG-GGA-AA-AG-TG---AGCA--ACCAAC-TGGAGG-TACCAG-C-GACT...GGAGAAGTGGCATAG...TCTGATAAAATTCTTAAAGTTTAA.............   4844

TANTALUS.UG.TAN1   --A--T--C-AA--A------G-GAGA-AAA-T-TG-G---CAAGACTAAC-TGGAGGGTT-G.................................................................   5286
GRIVET.ET.GRI677   -----T--C-AA---------G-GAGA-AAA--ATG---AG-GAG-GTAGT-TGGAGGGT-TCAGA--GGCA.AATAAGCAGATGGAGGGG..GATAGTGACTTACAAGATCAGGAATAA..........   5144
VERVET.KE.VER9063  --A-----C-AA-------A-CC-AGA-AAAG-ATG-G----GAG-GTAAC-TGGA-GGTCTCAG--GACCT.GATAACCAAATGGCAAGG..AATAGTCAAATATTGGATGACTAA.............   5365
VERVET.KE.VERTYO   --A--T----A--------A-CCCA-ACAAAG--TG-G-A--GAG-GT-ACGTGGA-GGTACCAG--GA-CT.GATAACTAAATGGCAGGG..AATAGTCAGATATTGGATGAGACA.............   4852
VERVET.KE.VER155   --A-----A-AA--C----AGCC-AGA-AAA--TTG-G----GAGACTCACCTGGA-GGT-CA-GA-GAAGT.GATCACCAAATGGCAGGG..GATAGTTAGATATTGGATGAATAA.............   5357
VERVET.DE.VERAGM3  --A--T--A-A--------A-CC-AGA-AAAG-ATG-G----GAGAGTAACTTGGA-GGT-CC-GA-GAGCT.GATAACTAAATGGCAAGG..GATAGTGAGGTACTGGATGAGGAC.............   4860
SABAEUS.SN.SAB1C   --A--T--A-AA-----------......---TTG---AG-CAG-CCC-TCTGGA-GGTA-C-G-CGGACA.GCAGGAGAGGTGGACTAG.........................................   5506

SUN.GA.SIVSUN      --AC-AG-A-A---C------GG-................--A-CCTCCT--GTGGAGAT--TG-GATGA...............................................AAGAGAAT   5446
LHOEST.KE.LHOEST   --AC-AG-A-AA---------G-G...GGAC--AAG---AG-GAGAGTAGCCTGGATA-CA-CT--...............................................................   5448
LHOEST.CD.447      ---C-AG-A-AA---------G-G...GGAC--AAG---AG-GAG-GT-GCCTACATA-CA-CT--...............................................................   4364
LHOEST.CD.485      ---C-AG-A-AA---------G-G...GGAC--AAG---AG-GAG-GTAGCCTGGATA-CA-CT--...............................................................   4364
LHOEST.CD.524      --AT-AG-G-AA---------G-G...GGAC--AAG---AG-CAGAGTAGC-TGGATA-CC-AT--...............................................................   4361
MANDRIL.GA.MNDGB1  --A--A--A-AA---------G-G...-GT-G-AAGA--AG-CAG--TAACTTGGA-AGT-TCT--...............................................................   4774

COLOBUS.CM.CGU1    -------CC--ACTCC-ATATGG-G-AGAT-TG--CAG---GAA-T--TT-TCC-ATGG-C-AA----AGCC.GAAACCGTTAAGGGAATG..GATTAA..............................   4886
SYKES.KE.173       --A--A----CA--CC-----G--......AG-ATG---AG-G--TCCCACGTGGAGA-T--CCCCA-GACA.GATTGATAG...............................................   5296
RCM.NG.NGM         --------A-AA--------C-G-A-AGA--TG-A-AGA--GACCAATATGGAAG-TAGAC-G-----AAGT.TGA......................................................   5448
MANDRIL.GA.14CG    --A--A--C---------------ACAG---ATA--A--CCCCACA---AG-ATAGACA-GA--CAGCTTGA.........................................................   5024
MANDRIL.14CG  Pol _K__I__I__R__D__Y__G__K__T__V__D__S__N__P__H__M__E__N__R__Q__D__A__A__$_                                             p31
MANDRIL.14CG  Vif                 _M__E__K__Q__W__I__V__I__P__G__W__R__I__D__R__M__Q__L__E__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   Vif
```

```
H1B.FR.HXB2        ...............................................AACATGGAAAAGTTTAGTAAAACACCATATGTATGTTTCAGGGAAA...GCTAGGGGATGG...TTTTATAGACATCA   5168
H1B.FR.HXB2  Vif   .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__W__K__S__L__V__K__H__H__M__Y__V__S__G__K__.__A__R__G__W__.__F__Y__R__H__H    Vif

CPZ.US.CPZUS       ..............................................................-C-G------T-T--C------AGA---AA----...----AA------...--C--C--G-----   5228
CPZ.GA.CPZGAB      ..............................................................------T---T-T-----T---AGG---AAA--G....--A--A--T---...-----C--------   5229
CPZ.CM.CAM3        ..............................................................CC--------TGT---------AAAA--AA----...--A-A--AC---...--------G--C--   4582
CPZ.CM.CAM5        ..............................................................CC--------AC----------AGAA--AA---G...--A-AAAAG---...--------G-----   4870
CPZ.CD.CPZANT      ..............................................................C-----G--------C--A-GG-AGA-TAAAGT-...CT--AACC----...AAG----------   4607

H1A.UG.U455        ...........................................------C--------------T--C--------C--CAA----...---CAA--T---...--------------   4614
H1C.ET.ETH2220     ...........................................------T-----------G---------C--A-----A---G-...----AT-------...G-------------   4560
H1D.ZR.84ZR085     ...........................................C---------------------T--------C--A-----AA----...----AA------...-----------C--   4690
H1F.BE.VI850       ...........................................C---------------------T-T--------------AA----...--C-AA-------...C------------   4508
H1G.SE.SE6165      ...........................................-------C-C-----G--------T-----------C---AA----...-----A--C---...----------C---   4565
H1H.CF.90CF056     ...........................................C-----------C--------GT--------C--A-----A-----...-----A------...--------------   4515
H1J.SE.SE92809     ...........................................C---------------------GT--------A--------AA----...-----AAA----...C-G-----G-----   4482
H1K.CM.MP535C      ...........................................T--T------------------T--------C--A-----AA----...----ACC------...-A------------   4364
H101_AE.TH.CM240   ...........................................---------C--------------T---------A-C---AA----...----AAAAG---...--------------   4742
H1N.CM.YBF30       ...........................................C-----------T-----------G---AAA--G....--A-AA-------...-A------------   4762
H1O.CM.MVP5180     ...........................................A--CA-G--TAA---G....--CGC-AAC---...CG------G-----   5198
H1O.BE.ANT70       ...........................................---CAGG--TA-A--G...A-CGA-AAC---...-GG--C--------   5222

MAC.US.SMM251      ...........................................-ACT---GATCTACAAAAGGTTTGC-A-GTGCCC---TT   5449
MAC.US.SMM239      ...........................................-ACT---GATCTACAAAAGGTTTGC-A-GTGCCC---TT   5473
MAC.US.SMM239 Vif  .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__K__D__L__Q__K__V__C__Y__V__P__H__F    Vif

STM.US.STM         ...........................................CACT---GAGCTA--CAAGGCATGC-A-GTGCCC--C--   5120
SMM.US.SMM9        ...........................................CACT---GACCTACA-AAGGCTTGT-A-GTACCC-----   4946

H2A.GW.ALI         ...........................................-ACA---GATCTAGAAAAGGT-TGC-A-GT-CCC--C--   5549
H2A.SN.ST          ...........................................-ACAGG-GATCTAGA-AAGGT-TGC-A-GT-CCC--C--   5001
H2A.DE.BEN         ...........................................-ACA---GACCTAGA--AGGT-CGC-A-GT-CCC--C--   5556
H2B.GH.D205        ...........................................-ACAGG-GAGTTGCAACAGGTCTCT-A-GTCCCT--C--   5529
H2B.CI.EHO         ...........................................GACA---GACTTGCAACAGGTCTCT-A-GT-CCT--C--   5527
H2G.CI.ABT96       ...........................................-ACA---GACCTACA-AAGGC-GTT-A-GT-CCT-----   4883

TANTALUS.UG.TAN1   ...................................TGGAGAGCATATAG-C-AA-GGAAGGGCATAGTG-AA--CCACATGA----CCGGTTACA--AT---ACA-A-CT--TG--CT-   5370
GRIVET.ET.GRI677   ...........................................ACAATT-CCT---GAA-ACAGACAT---TG   5174
VERVET.KE.VER9063  ...........................................-AGGAACCTTAA----CAA-ACATG-TG--CT-   5398
VERVET.KE.VERTYO   ...........................................-AGAAA-TT-AA----AAC-A-----TG---T-   4885
VERVET.KE.VER155   ...........................................-AGGAA-CT-AA----GAA-ACA-A-TG---T-   5390
VERVET.DE.VERAGM3  ...........................................TAGAAAATTA-AC---AAA-A-CGA-TG--CT-   4893
SABAEUS.SN.SAB1C   ...........................................CC--GTGA-AT-CCACATGCA-GTGTCC---CAGTG-GTTCAT---CGG-A-ACCCCT---AC   5569

SUN.GA.SIVSUN      GGGAAGATAGAAGACAGTATAAGATAGTTAGAATAGTATGGCTTATAGATAG--TTGCTGT-GAGAA-T-TTT-G-TTTA-GAAGAA-GCATA--G--ACAAAAGAT-AT---GTA-C-ATGTATGGAAC   5576
LHOEST.KE.LHOEST   ...........................................-T-C-TA-T-GAAAAGT-TTT-G--ATAG-CAGGT-AAACACAG-GGCAAAAGAAAAT---GTAGGC-TACATGGAAC   5526
LHOEST.CD.447      ...........................................G-GCCTAGT-GAGAAGT-CCT-G-TATC-G-AGA-CACATAAAG-GACTCAA-AT-AT---GTAAGC-TACATGGAAC   4442
LHOEST.CD.485      ...........................................G-ACCT--T-GAGAA-T-CCT-G-TAT--G-AGA-CACATAAAG--ACTCAGGTT-AT---GTAAGC-TACATGGAAC   4442
LHOEST.CD.524      ...........................................G-GC-TAGT-GAGAA-T-TCT-G-TATC-AAAGGAA-CATTCAG--AGTAAA-AT-AT---GTTAG--TACATGG-AC   4439
MANDRIL.GA.MNDGB1  ...............................TCAGAGAATAGAAAAGTGGCAC---TT-GTAAG-AG-C-GATGGCATG-GCCAC-G--AAT--TGAG-AAG-AT-T---TGGC-G---CCT---TT   4870

COLOBUS.CM.CGU1    ...........................................GAA-C-GTTTC-AGAAT-CCAATAT-A----TT-GGAA-   4925
SYKES.KE.173       ...........................................ATTGCAGCATA---T-A-AACACATA-GTA-AA-A-C---GAATTGGA-AA-GCCACA-A-A-ACAT--CT-   5368
RCM.NG.NGM         ...............................GCAGTGGCATTCCTT-GT-AA-T-TCA-CAGTAT--GGGAA-G-A-G-A-C-AA--ATTGGGAGTA-GTACCCCATTTTAAAGTACCT.....   5536
MANDRIL.GA.14CG    ...........................................--A------TTCA-----C---T----C-AA---AAGGA--AA---CACTTAGATAAG---GAAC-A-TCCAC--CTT   5102
MANDRIL.14CG  Vif  .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__K__W__N__S__L__V__K__Y__H__K__Y__K__E__E__K__H__L__D__K__W__E__L__F__H__H__F    Vif
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
H1B.FR.HXB2       CTATGAAAGCCCTCATCCAAGAATAAGTTCAGAAGTACACATCCCACTAGGG......GATGCTAGA......TTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGACTGGCAT......   5280
H1B.FR.HXB2  Vif  __Y__E__S__P__H__P__R__I__S__S__E__V__H__I__P__L__G__.__.__D__A__R__.__.__L__V__I__T__T__Y__W__G__L__H__T__G__E__R__D__W__H__.__._   Vif

CPZ.US.CPZUS      T-----G---A--A----C---G---A------------T-----TA---AT......A-A--A-A-......A-A--G--T-T---G--------GT--A-A---------CA--C------C......   5340
CPZ.GA.CPZGAB     T-----TCA----A------A-G-TGC---T---A-------A--TT-CA-AGAT...T--T---A-.......--AA--G-G-----C-------CCT-AAGCC----------G-C-------.....   5344
CPZ.CM.CAM3       T--------TT-------T-A-G-----------A----------TA---AA......A-G--A-A-.......--A--GG-------C--------AT-AATGC-----C-GCAT-C-------.....   4694
CPZ.CM.CAM5       T-----G--TGA-A----T-A-G-G--------------T-----TT---AT......ACA--A-A-.......--A--GG-C--------------A---ACA-----GC--CAT-CT-----C......   4982
CPZ.CD.CPZANT     T-------ATGA---C----AG-A-G-AGA---------T--A------CCAACATTG---AAA-A-.......------G-G---GTG-TC-----AT-A--ATGT------C-CCCA-----C......   4725

H1A.UG.U455       ---------TAGA---T-----G-----------------------------.......--A------.......--A---G---G-----------------------------A---------C......   4726
H1C.ET.ETH2220    T-----C---AGA-------AGG-----------------------T-----.......--G------.......--AA------A-------------T----A--------------T------.....   4672
H1D.ZR.84ZR085    ------C--------C----A-------------------------T----A.....----------.....C-----G-----------------------------------A------.....   4802
H1F.BE.VI850      --T-C-----AGG---------G---------------------------A-.......--A-T--A-.......--A--------------------G-----CC-------------A------.....   4620
H1G.SE.SE6165     -----C----AGG---------G---------------------------A.............C-......C-A---G----------------G----------------A---------A......   4677
H1H.CF.90CF056    T-T-------A----------G-------------------------T----A.......--A-----G.......--A--C-----C-----C---------A----------------A------.....   4627
H1J.SE.SE92809    ------C---AA--------A-----------------------------A.......--G----T-.......--A-------------------------G--------------T-----C......   4594
H1K.CM.MP535C     ----------AGG-------A-----------------------------------------AGAG.......--A---G---------------------T------------G---------.....   4476
H101_AE.TH.CM240  T----------AG-------AGG----C-----------T-----------A.......--G------.......--A-------G-----------------A----------AG--------C......   4854
H1N.CM.YBF30      T-------CA-A---C----A------------------T------G----T......C-G--A-.........--A--G-C-GTC--T--------G--AACA---------CAGTCT------.....   4874
H1O.CM.MVP5180    T------TC-AGGA------A-G-C-----G-CG--GT-T--T---G---CA........--A---GAT......A-A--GG-C--C-----------AT-AATGC----G--------GGAA--C......   5310
H1O.BE.ANT70      T------TC-AGAA--------G-T------AGT---T-T--T---G-----.......-TG--CCAT......G-A---G----C-----------AT--ATGC----G--------TGAA---.....   5334

MAC.US.SMM251     TA-G-TCG-ATGGGCATGGT-G-CCT-CAGCAG----AT-T----C---CA-........--A-GA--CCAT...--A-A-G--CA-GGG------CA-T--...---CC--------GG----TCAGTACT   5567
MAC.US.SMM239     TA-G-TCG-ATGGGCATGGT-G-CCT-CAGCAG----AT-T--------CA-........--A-GA--CCAT...--A-A-G--CA-GGG------CA-T--...---CC-----A--GG----TCAGTACT   5591
MAC.US.SMM239 Vif __K__V__G__W__G__W__W__G__C__S__R__V__I__F__P__L__Q__.__.__E__G__S__H__.__L__E__V__Q__G__Y__W__H__L__.__T__P__E__K__G__W__L__S__T_   1801

STM.US.STM        TA-G-TTG-GTGGGCATGGT-G-CTT-CAGTAG---GATTT-T--CT-GCAA......-GA-AAGC-CAC...C-A-A-G-TCA-GGG------AA-T-A...---CC-----A--GA----TCAGTGAA   Vif
SMM.US.SMM9       TA-A-TTG-ATGGGCATGGT-G-CTT-CAGCAG----AT-T----GT--A-.........----AGGCTCAT...----A-G--CA-GG-------AA-T--...---CC---GRA--GG----TCAGTACT   5064

H2A.GW.ALI        TA-G-TGG-ATGGGCATGGT-G-CTT-CAGCAGG---ATAT-----T--CAA......-GAAGA--TCAT...C-A-AG---CAGG--------AAC--A...---CC-----A--GA----TCTCCTCT   5667
H2A.SN.ST         TA-G-TGG-ATGGGCGTGGT-G-CTT-CAGCAGG---ATAT-----T--AAA......-GA-AA--TCAT...C---AG---CAGG----C---AAC--A...---CC-----A--GA----TCTCCTCC   5119
H2A.DE.BEN        -A-G-TGG-ATGGGCATGGT-G-CTT-CAGCAGG---ATAT--------CAA......-GAAAA--TCAT...C-A-A----CAGG--------AAC--A...---CC-----A--GA----TCTCCTCT   5674
H2B.GH.D205       -A-G-T-G-ATGGGC-TGGT-G-CTT-CAGTAG-A--ATAT-T--C---AAC......A-A-GAGC-TGG...C-A-A-G-CCA-GG-------AAC--A...---CCC------G-GA-TCTTGAGCTCC   5647
H2B.CI.EHO        TA-G-T-G-ATGGGC-TGGT-G-CTT-CAGTAG----ATAT-T--C--GAAA........-A-GAGC-CAT...C-A-A-G-CCA-GG---C---AAC---...--CCC------G-GA-TCTTGAGTTCC   5645
H2G.CI.ABT96      -A-A-TTG-ATGGGC-TGGT-G-CTT-CAGCAGG---ATAT-----T--ACA......A-A-AAGC-CAT...----A----CA-GGG------AACT--...--CCCT----AG-GA---TTAAGCCAG   5001

TANTALUS.UG.TAN1  -C-ATGTG-ATGGGC-TGGTAC-C-T--AGTAG-T-C-T------CT----A.......-GA-AAG--AAA...A-A---G--GATTGT--C---CAC--A...--GCC----CAG-GA----TCTCTACC   5488
GRIVET.ET.GRI677  GC-G-TGCAATGG--GTTTT-G-CCTACAGCC-GT-CATT-----CT--TCA......A-A-A-GATTAC...A-A-A-G-G-AT-TT---CACAAC--C...--CCC--------GA----TCTCAAGT   5292
VERVET.KE.VER9063 -C-AATT-ATTGGGC-TGGTAT-CT-TG-GCAG-TACAT-------A----A......TTA-AAG--GAA...A-ACAG---GACTGC------CA-T-A...--CC--------G-----TCTCTACT   5516
VERVET.KE.VERTYO  -C-AATT-CATGGGC-TGGTAC-CC-TGAGTAG-TATGTA--A---A----A......A-ACA-G-GGAA...A-ATGTG--GACCT-------CA-T-A...--CC---GCA--GA----TATCCACA   5003
VERVET.KE.VER155  TC-AATC-CTTGGGCATGGTAC-CT-TGAGCAG-TATGTA--A--C--CCCA......-GAAG-G--GAA...A-CCATG-GGAT-TC------CA-T-A...G-TCC-A--CA--GA----TCTCAACT   5508
VERVET.DE.VERAGM3 -C-AATT-CATGGGCATGGTAC-C--TGAGTAG-TATG-G--A--C------.......C-ACA-G--AGT...A-ACATG--GATCT-------CA-----...---CC-----AG-GA----TATCAACA   5011
SABAEUS.SN.SAB1C  TA-AAT---ATGGA--TGGTACTCCTA-CA----TGGGT---A--C--GAA-.........----GAGCTCTA...A-AAA-G-T--T-AT------CA-T-A...--CCC---G-A--GA---TTAGAAACC   5687

SUN.GA.SIVSUN     AGG-ACTG-ATGGG-GTGGTAC-C-TACAATA--A-TATT--T---G--ACT......T---GG-C-GTA...G-A----GG...-TC---G-ACAC--C...---CC--C--A--GG-T-AT-AACCAG   5691
LHOEST.KE.LHOEST  AGG--T-G-TTGGG-ATTCTAT-CCTA-CATA--T--GTG--A--T----A-......TCA-GA-CTTTA...G-A--C-G-...-T----GCACA-T-...G---C--G---G-GA---ATCTCACAA   5641
LHOEST.CD.447     AGGG-T-G--TGGG-ATTTTAT-CTTA-AATA-TA--GTG--A--TT---A-......ACA-GA-CCTTA...G-A--G-G-...TGC---GCACACT-...G---C--G-----GA---ATCTCTCAG   4557
LHOEST.CD.485     AGGG-T-G--TGGG-ATTTTAT-CTTA-AATA--A--GTA--A--TT---A-......ATA-GA-CCTTA...G-A----G-...TGC---GCACACT-...G---C--G-----GA---ATCTCTCAG   4557
LHOEST.CD.524     AGGA-T-G-ATGGG-GTTTTAT-CTTACAATA--T---TA--T---T---AT......ACA-GA-CTTTA...G-A----G-...-TT---GCTCA------...G---C--GG----GA---AT-TCACAG   4554
MANDRIL.GA.MNDGB1 TATG-CTTATAA-G-ATGGTAT-CTT-CAGTA-----GTG--TAT-A--...AATAGG--CATA---TTA...A-A--T-G-...-GC------CA-T-...CA-AT---GGT--GA--CTTAAGTACT   4988

COLOBUS.CM.CGU1   AACA--GTCA--CG-ATGGT-G-C-GAAGGTA--A--G-GT-TTACT--..........-GGT-AGCTATTAAG--AAGC----T-TAC-TGG-CAC-T-A...--CCC---T---T-TCATAGAGGACCA   5043
SYKES.KE.173      -C-AAT-GAATGG--ATGGTAC-CTTA--GCC--TGGACA--A---G----A..........----GA-CTATA...-G-A-C-CG...TTC---CACAA-T-...G-CCC--------GA---TTACATATG   5483
RCM.NGM           ..........TGGGGGTGGT-GTCTC----------C--------CT-G--A......A--AAC-CTAAG...A-AAAGG-T------------A----C...---AT-----AG-GA---TTAGGAACC   5644
MANDRIL.GA.14CG   T.....................CA-T-C----G-TGGTGG-CG-ACTCTCAAAAGATTATCC-CTTCAAAGATGGCTC--AG-TT-TTAT-ACT-CC--ATGG-ACTTGACCCC---AAAAGGATGGTTA   5211
MANDRIL.14CG  Vif __.__.__.__.__.__.__.__Q__C__S__G__W__W__G__H__S__Q__K__I__I__P__F__K__D__G__S__K__I__I__I__T__A__L__W__N__L__T__P__E__K__G__W__L_   Vif
```

```
H1B.FR.HXB2   Vif _.__.__.__.__L__G__Q__G__V__S__I__E__W__R__K__K__R__Y__S__T__Q__V__D__P__E__L__A__D__Q__L__I__H__L__Y__Y__F__D__C__F__S__D__S__A_      Vif
H1B.FR.HXB2       ............TTGGGTCAGGGAGTCTCCATAGAATGGAGGAAAAAGAGATATAGCACACAAGTAGACCCTGAACTAGCAGACCAACTAATTCATCTGTATTACTTTGACTGTTTTTCAGACTCTGCT.    5397

CPZ.US.CPZUS      .............--A--G--T-----A--A-----------AC--GGA-CT--C-AG------A----T-----GA-G------A-G-----------------T-----T------A---C------C.    5457
CPZ.GA.CPZGAB     .............--A--C--T-----------TC--------CT-GGATCT---GT------G--G--T--CTTCAC------TAG-T-G--C--CTCTC-A--T-----T--C---G----AA----C.    5461
CPZ.CM.CAM3       ............C----G--T-----G--------------ATGTGG-TC------T-----------------GA-G-----TA----------C--TC-A--T-----T------G---CT-----C.    4811
CPZ.CM.CAM5       ............C----A--T-----G--------------ATGTGG-TC---C-AA-----------T------ACG-----TAG---------C--TC----T-----T------G---CT-----C.    5099
CPZ.CD.CPZANT     .............--A--A--T--T--GAG------------ATGTGGA-AG----T---------------A---AC---T--T---A-G--A--C-AA-----T--CCC------------TCAA--A.    4842

H1A.UG.U455       ............--------T--G------------------CTG--A--------------------T-----C-----------T--------C---C----T-----------------A-----C.    4843
H1C.ET.ETH2220    ............--------T--------------------ATTG-GA--C----A-----------------GC--G--------C---------A--C----T-----T------G----A-----C.    4789
H1D.ZR.84ZR085    ............C-----------------------------G-------------------------GC--G------------------A-------T-----T------G-----------.    4919
H1F.BE.VI850      ............C----------------------------C-GGG--AG-----G------A---------GC--G--G--------G--C---A-A-----T-----T-----------A-----C.    4737
H1G.SE.SE6165     ............--------T--G----------G------C-G-GA--------A----------G-----C-----------T--G---------C----T-----T-----------------C.    4794
H1H.CF.90CF056    .............--A--C-----------------------ACTG--A--G-----------------G----GC--G------------------A--C----T-----T-----------A-----C.    4744
H1J.SE.SE92809    ............-----------------------------C-G-GA--G--C--A----------------GC--G------------------A---G---------T-----------------C.    4711
H1K.CM.MP535C     ............-----------------------------CTG--A--G--C--A-----------A-----C--------------------A-A-----T-----T-----------A-----A.    4593
H101_AE.TH.CM240  ............--------T---------------------C-G-GA-A-------------A----T-----C---------A----G--------AC-A--T----G------------------C.    4971
H1N.CM.YBF30      ............C-A--A--T-----A--------------ACT--GA-A---C-AG--------T--T------A-G------A-G-----A-----TC----T-----T------A---C------C.    4991
H1O.CM.MVP5180    ..............-----A--T--G--TAGT---------CAAT-C---GAG----AA-----GA-T--T------AC-------AGGA-G--A------C----T--CACA------A----A--A--A.    5427
H1O.BE.ANT70      ..............-----A--T--G--TAGT---------C-TT-T----AG----AA-----GA-T---------AC-------AGGA-G--A------C----T---AC-------A---CA--A--A.    5451

MAC.US.SMM251     TATGCAGTG.................AGG---ACC---TACTC--G--AC-T-T-G---G-T---ACA--A--CTAT------ATTT--C-G---AGCAC---T--CCCT--C---A---CGGGA-AA.    5678
MAC.US.SMM239     TATGCAGTG.................AGG---ACC---TACTC-----AC-T-T-G---G-T---ACA--AA-CTAT------ATTT--C-G---AGCAC---T--CCCT--C---A---CGGGA-AA.    5702
MAC.SMM239    Vif _Y__A__V__.__.__.__.__.__.__R__I__T__W__Y__S__K__N__F__W__T__D__V__T__P__N__Y__A__D__I__L__L__H__S__T__Y__F__P__C__F__T__G__G__E_      Vif

STM.US.STM        TATGCTGTG.................AGG---ACT---TAT-CC-GA-AT-TCT-GT--G-T---ACA--A--CTGT-----T--G---C-G---GGCAC---T--CCCT--C---A---CGGGA-AA.    5349
SMM.US.SMM9       TATGCAGTG.................AGA---ACC---TACTCC-GA-AT-TCT-G---G-T-C-ACA--A--TTAT------ACTT--C-G---GGCAC---T--CCCT--C--------GGGA-AA.    5175

H2A.GW.ALI        TATGCAGTA.................AGA---ACT---TAT-C-G-A-AG-TCT-G---G-T--TAC---A--CTGT------TCC-----A---GGCAC---T--CTCT--C---A-G-CAGG--AA.    5778
H2A.SN.ST         TATTCAGTA.................AGAC--ACT---TAT-C-G-A-A--TCT-G---G-T--TAC---A--CTGT--G---TCC-----A---AGCAC---T--CTCT--C---A-G-CAGGC-AA.    5230
H2A.DE.BEN        CATGCAGTA.................AGAT--ACC---TAT-C-G-A-AG-TCT-G---G-T--TAC---A--CTGT------ATC-----A---AGCAC---T--CTCT--C---A-G-CAGG--AA.    5785
H2B.GH.D205       TATGCTGTA.................AGAC--AC----TATG-G-G--AC-T-TAT---G-T---ACA-----TG-G--------G---C-G---GG--C---T--CTCT--C-------C-AA--AA.    5758
H2B.CI.EHO        TATGCTGTA.................AGAC--AC----TATG-G-G---C-T-TAT---G-T---ACT-----TG---------G-T--C-G---GG--C---T--CTCT-CG---A---CTAA--AA.    5756
H2G.CI.ABT96      TATGCAGTA.................AGAT--ACC---TAC-C--GA-A--T-TAT---G-T--GACA--A--GAC------T---T--T-G---GGA-C---T-----T--C---A---CTGG--AA.    5112

TANTALUS.UG.TAN1  TATGCAGTAGCAA-A.........AGT-TTGAGA-T---CAA--C...-CG--C-AA---G-----ACT-----TG-G--G-----TA-G------TGCC-C-----CCCT--C---A----TAGA--C.    5605
GRIVET.ET.GRI677  CATGGAGTAGGG--A.........TC--ATTACC-TCAA-A-GG-......----AG---G-------T--A-G-AC-------AG-A-G--A--C--A-----T---A--------A----TAGA--C.    5406
VERVET.KE.VER9063 TACGCCATAGGAA-A.........CAA-A-T-G-G-GCAGTAGG-...-AT-TC--A---G--T-------A-C-AC------TGG-A-G--A--C--AC-C--T---TCA------A----AAGG--C.    5633
VERVET.KE.VERTYO  TATGCAGTAGGTA-A.........CAA-ATG--AGCAATTTAG--TCT-A------A---G--T---------CTAC------TAGTA----A---GGTC-C--T---A-T------AA---AAGA--C.    5123
VERVET.KE.VER155  TATGCAGTAGGAA-A.........CAA-ATG-TAGCCTAGTA--TG-T-A------A---G--T----T--CA-TAC-------TCCA-G--A---TGTC----T---AC-------A----TAGA--C.    5628
VERVET.DE.VERAGM3 TATGCTGAGGGGA-A.........CAG-ATC--AGCAAT---G-TCCTT-G-----G---G--T-G--T----C-AC------TAGC--G--A---ACCC----T---ACT------A----AAGG--C.    5131
SABAEUS.SN.SAB1C  TATGCAACA.................GGA----G--AC--T---GG-GA--GGTTT---G-GC-T--T---TGGAC---G-----TA-T------TG--CC--T--CCC-------A----TAGA---.    5798

SUN.GA.SIVSUN     TGGGGATGT.................AGT--G-------TTT-T--C--T---CAA---G--A----T---CTTG-G-----T--GA-G--C--CT-TA-G--T-----T----GGA-CTCTAGGTG-.    5802
LHOEST.KE.LHOEST  TGGGCAGTT.................------------GTCT-T-GAGAC---CAA---G--A-------CAT-------------A-G--C--CTATA-A--T---A----C--CAT-C-AAGA-AC.    5752
LHOEST.CD.447     TGGGCAGTA.................--A----------TTT-TGG--AC---CAA---G--A----T--CATT------A-T---A-G--A--CTGTA-A--T---A----C-----TACAAGG-AC.    4668
LHOEST.CD.485     TGGGCAGTA.................--A----------TTT-TGG--AC---CAA---G--A----T--CATT--------T---A-G--A--CTGTA-A--T---A----C-----TACAAGG-A-.    4668
LHOEST.CD.524     TGGGCAGTA.................--A----------TCT-TGGA-AC---CAG---G----G--T--CATT-----T--T---A-GC-----G--A-G--T----------GGGTGC-TAGG-AC.    4665
MANDRIL.GA.MNDGB1 TATGCAGTAAGCA-A-AAGCA-T---TAGACCGCC-CCCTTTG-G--AGAG-GGT-T---G-GA--ACT--A--GG-------T--T-----A---T-AC---TT-A------C--CATG---AG---A.    5117

COLOBUS.CM.CGU1   CCAGTC.....................--TTAT----TACA--T-GGT-AG-GGGAA---G-TT--AG-TGGC---A-----TAGGGTAC--G--CA-CATAAGGGGGA-G---G-ACGCATAGAG-AAG    5152
SYKES.KE.173      CAGGGAATT.................AGG--TC---ACCAATGG--CCAG-GG-A---TG-CT--ACA--A-C-G---------GC--T--A---AAT-T---T--CCCT--C---A---CAAGA--G.    5594
RCM.NG.NGM        TATGGAGTAGGAA-A......-C-TA-ATAGATA-CA-AT-TG-TCCCT-T--CTT----G-TA-T-----AAC-G-------TA--T--G-A---AATC-G--T---CCT------A----TAAA--A.    5767
MANDRIL.GA.14CG   TCTCAATATGCAA-C...........A-AGC-----AC-AC---GG-GAC---TAT--T--CA-------A-T-AC-------AG-A-G------TG-G-A--T--CCCA------A---CTG-C---.    5328
MANDRIL.14CG  Vif _S__Q__Y__A__I__.__.__.__.__.__T__A__E__Y__N__K__G__D__Y__Y__T__H__I__D__P__V__G__A__D__R__M__I__H__W__E__Y__F__P__C__F__T__G__A__A__    Vif
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2   Vif .__I__R__K__A__L__L__G__H__I__V__S__P__R__C__E__Y__Q__A__G__H__N__.__.__.__.__.__.__K__V__G__S__L__Q__Y__L__A__L__A__A__L__I__.__.     Vif
H1B.FR.HXB2       ..ATAAGAAAGGCCTTATTAGGACACATAGTTAGCCCTAGGTGTGAATATCAAGCAGGACATAAC.................AAGGTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATA.....    5502

CPZ.US.CPZUS      ..--C-----A--TG-T-----GAGACC---A-TT--C-AA--------C-CTA----G--C--T.................--------C---T-G----T------TT--AAG--C---G--.....    5562
CPZ.GA.CPZGAB     ..--C----GA--TA-------G--GC-T--AGCG---C-T--------CA---A------C-GG.................C----------CT----G-T------C---AA---T------.....    5566
CPZ.CM.CAM3       ..--C--G--A--TG-T-----GA-ACA---ATAT----A---------C-C------G---C-A.................C-------G--AT-------T--A--C---AG---C-GGG--.....    4916
CPZ.CM.CAM5       ..--C--G--A--TG-T-----GA-ACA---ATAT----A---------C-C-A----G--CC-A.................C---K---G--A--------T--A--C---AG---T-GGG-G.....    5204
CPZ.CD.CPZANT     ..G-C---C-A--AA-CC-----G-A-G-A-ATTGA-ATAC--CC-C--CA-GAAG-------GT.................C-------GA-A--G--G--T--A--TT-TTGTAAGA-CC--GAATT    4952

H1A.UG.U455       ..-----G-GA---A----------A--------A-----------------------------.................--------------------T-----T--GAA-------G--.....    4948
H1C.ET.ETH2220    ..-----G--A---A--------T-T-G------------------C-----------------T.................-----------------------------GA-------G---.....    4894
H1D.ZR.84ZR085    ..--C-----A---A----------T--------T-----------------------------.................------------CT----G--T--A------A------------.....    5024
H1F.BE.VI850      ..-----G--A---A----------T-G-A--------------A-C-----------------T.................-----------C--------T---------A-----------.....    4842
H1G.SE.SE6165     ..-----G--A---A----------A-------------------------------------T.................C-------------------T------T-GAA--T----G--.....    4899
H1H.CF.90CF056    ..-----G--A---A---------GTG-------A---------A-C----C------------A.................C--------A-----------------T--A--------G-G.....    4849
H1J.SE.SE92809    ..-----G------A----------A--------------------C--C---------------.................--------------------TC--------A-----------.....    4816
H1K.CM.MP535C     ..G----G------A----------T-G------T-------------G---------------.................-----------CT-------T--------CA--------G--.....    4698
H101_AE.TH.CM240  ..-----G--A---A----------AG-------A-G--------------C-T-----------.................-----------C--------T--------GAA--------C-.....    5076
H1N.CM.YBF30      ..-----GC-A--GG-C-----GAGACC---ATTA---------------C------G--C--A.................C-------CA-C--------TC-A------A----C-GGG-G.....    5096
H1O.CM.MVP5180    ..--C--G------A-TC----G--G-G---GCTGA-C-A---------C-TG----------GT.................C-------GA-A-------T---A--CT-GAA----G--G-G.....    5532
H1O.BE.ANT70      ..G-C--G--A---A-CC-------G-G---ACTGA-C-A---------C-CTA---------GT.................C-------GA-A------CTGC-A--T---AG----G--G--.....    5556

MAC.US.SMM251     ..G-G----G----A-CAGG---G-ACA-C-GCTGT--T-C--CA-G-TC-CGAG--CT----GGTAC..............C-----CC-AGC-----G-----A------AA--T-G---GC.....    5787
MAC.US.SMM239     ..G-G----G----A-CAGG---G-ACA-C-GCTGT--T-C--CAGG-TC-CGAG--CT-----GTAC..............C-----CC-AGC-----G-----A-----GAA--T-G---GC.....    5811
MAC.US.SMM239 Vif .__V__R__R__A__I__R__G__E__Q__L__L__S__C__C__R__F__P__R__A__H__K__Y__.__.__.__.__.__Q__V__P__S__L__Q__Y__L__A__L__K__V__V__S__.__.     Vif

STM.US.STM        ..G-G----GA---A-CAGG---G-A-AGT-ACTGT--T-C--CAGG-TCACGAA--CT-----GAAC..............C-----CC-AGCT----G-----A--TT--AA--T-G-GGA-C....    5458
SMM.US.SMM9       ..G--C---G----A-CAGG---G-A-A-T-GCTGT--T-C--CA-G-TC-CGAA--CT-----GAAT..............C-----CC-AGC-----G--TC-A------A---T-G---GTC....    5284

H2A.GW.ALI        ..G------GA---A-CAG---GG-A-AGT-ATTGT-CT-C--CA-T----CCCAG-CC-----GTCA..............C-----CCG--A--C----TTC----CT---TG-T-G-G........    5883
H2A.SN.ST         ..G------GA---A-CAG---GG-A-AGC-ATTAT-CT-C--CA-C--C-CCCA--CC-----GTAC..............C-----CCG--A--C--G-TTC----CT---TG-T-G-G........    5335
H2A.DE.BEN        ..G------GA---A-CAG---GG-A-AGT-ATTGT-CT-C--CA-C----CCCA--CT-----AGCA..............C-----CC---A--T------C-A--C----T--T-G-.........    5890
H2B.GH.D205       ..G----G-GA---A-CAGG---G-A-AGA-ATTGT-CTAC--CA-C----C-T---CT--CG-AGGG..............C-----CC-AGCT----G-TTC-A--C---AGG-TCG--CAGGAAGG    5871
H2B.CI.EHO        ..G----G-GA---A-CAGG---G-A-AGA-ATTGT-CCAC--CA-C--C-C-T---CT----CAGGC..............C-----CC-AG-T----G-TTC-A--C---AG--TTG--CA-GAAGG    5869
H2G.CI.ABT96      ..G------GA---A-CAGG---G-GCAGA-ATTGT-CT-C--CA-C----C-A---CT-----AAGG..............C----GCCTAGCT-G----TTC-A--CT--CA--T-G-GCA-.....    5220

TANTALUS.UG.TAN1  ..--CCA-C-----A-CAG---GG-A-GCT--TTGTGG...--CAC---CA-G-A------CGTAGCAGAAAACCATTGGGGTC-----C-G--C--G--G-TT--A--C--GA---T--AT-C-GACTT    5730
GRIVET.ET.GRI677  ..--CCA-C----TA-CAG---GG-G-AGTA--CGTGG...--CAC--TCA-G-A---------A..............GGTC-----CA---A--G---CTT----------TT----AT-C-A....    5512
VERVET.KE.VER9063 ..--TCA---A--TA--AGG---G-G-G-T-CGT-TTC...---C---TC-CC-AG--------AACCACA.........GGA------CCC--A--G--G-----A--C--TCTG--C...........    5738
VERVET.KE.VERTYO  ..--CCA-C-A--TC-GAGG--C----G-T--GT-TTC...---C-G-T--C--A---G-----AAGCACA.........GGAC-----CC----T-G--G---C-A--T--GCTC--............    5228
VERVET.KE.VER155  ..--CCA-C----AC--AGG---A---GGT-C-T-TTC...---C---T--C--G---------ACTAACA.........GGTC-----CCC--CT-G-----T--A---T--CT---C...........    5733
VERVET.DE.VERAGM3 ..--C--G--A---C----G-----G-GGT-C-C-TTC...---C-G-TC-CC-AG-----C--GAAAACA.........GGAC-----CCC---T-G--------A--T--CCTT--............    5236
SABAEUS.SN.SAB1C  ..G-CCA-C-A---A-CAGG---G-A-AGTA-CTTTGG...---A--C-C--G-TG------C-GCCAACA.........GGGC-A---CC---AT----G--T--A-----TAGG-T--AT-C-A....    5910

SUN.GA.SIVSUN     ..--C----GA--TA-G-----GG-A-A-A-ACTA-A-GA---CAG-A------T--CT--C--AGGC..............CT---TCTC---T-G--G-TTC-CTGCT-GAG--TTC-CC.......    5908
LHOEST.KE.LHOEST  ..--T----G---AC-TC-G---G-A-G-A-A-CTATGT-T--CC-G...-CTTT---------AGGT..............TGT--TCTC---T-G--G-TTA-CTGCT-GAG-CA-C--CA-C....    5858
LHOEST.CD.447     ..--T--G-----T----GG---G-A-AGA-A-TGG-ATTC----CTATG-CCA-T--------AGGT..............TGT--TCTC--AT-G--G-T-C-CTGCT-GAG-CA-C--CA-C....    4777
LHOEST.CD.485     ..--T--G-----T----GG---G-A-AGA-A-TGG-A-TC--C-CTATG-CCA-T--------AGGT..............TGT--TCTC--AT-G--G-TTC-CTGCT-GAG-CA-C--CA-C....    4777
LHOEST.CD.524     ..---GTT-G----A--AGG---G-A-AGA--CTG-AC--C---A--ATG-CTA-T--------AGGT..............TGT--TCTC--AT-G--G---C-CTGCT-GAG-CA-C--CA-C....    4774
MANDRIL.GA.MNDGB1 ..G-T-TG--A---A-CAGG---G-AGA---GTTAAAAGTT---AG--T--C---T--C-----AGCA............CAAGGT--TCTC---T-G--G-TTC-CTGCT-GAG--TCA-CT.......    5226

COLOBUS.CM.CGU1   AAGC------A--TA--AGG--CTTACCCTGG-AT---...-----T-TC--G-T------CCTA.................GGA--GTTGAAGT-G---G-AC-ATGT--TCTG--CC---G-GAGAA    5261
SYKES.KE.173      ..G-C-ACC----AG-CAG---GG-AC--T-G-CAT-ACAC--CTGGACC-C-.........----CTGAC..............C-----CCC--A--T--G--T--A---T--CA--TG-ACC--AAAG.    5700
RCM.NG.NGM        ..--T---C-A---A-C------G-GCA---GCTG-T-...----GG-T-G-GAA--------GAGAT..............C-------GAGC--------TC-----A--AA---T-GGGCTAGG..    5875
MANDRIL.GA.14CG   ..G-G-------TGC-G-AC---G-A-G-A-AGTAG-G...---T-CAGC-CCTGG--------AGGT..............C-------GA--------CT------TT--AG---C-AC........    5430
MANDRIL.14CG  Vif .__V__R__K__V__L__Y__G__E__R__I__V__A__.__C__Y__S__P__W__G__H__K__G__.__.__.__.__.__Q__V__G__T__L__Q__L__L__A__L__R__A__Y__.__.__.     Vif
```

```
                                                                                             / Vpr start
H1B.FR.HXB2   Vpr                                                                           _M__E__Q__A__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   Vpr
H1B.FR.HXB2   Vif __.__.__.__T__P__K__K__I__.__K__P__P__L__P__S__V__T__K__L__T__E__D__R__W__N__K__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.   Vif
H1B.FR.HXB2       ..........ACACCAAAAAAGATA...AAGCCACCTTTGCCTAGTGTTACGAAACTGACAGAGGATAGATGGAACAAGCC................................................   5570

CPZ.US.CPZUS      ..........GG--AG-GC--A-G-...-GA-----C--A--------G--T---T-------A--------------TG................................................   5630
CPZ.GA.CPZGAB     ..........T--GAG-G--G-CAT...-GA--------A--------CG-C---T-A----------------------A................................................   5634
CPZ.CM.CAM3       ..........-G-GT-GG-----A-...--A-----C-----------C--T--GT-A-----A-------------G--A................................................   4984
CPZ.CM.CAM5       ..........GG-GT-C------A-...-CA-----C-----C---A--GSC--GT-AG----A--------------ATA................................................   5272
CPZ.CD.CPZANT     CAGA......GG-TACCCC---GGTCCA-GAAG--AG--C------C-GT-T-TCT-A-----A-----------------................................................   5027

H1A.UG.U455       ..........--C----C--G-GC-...---------------------A----T-A-----------------------................................................   5016
H1C.ET.ETH2220    ..........-AG---------GC-...---------C-----------GT---T-AGT--------A------------................................................   4962
H1D.ZR.84ZR085    ..........G-----------G-...--------------C-------A---G--A-----A----------------................................................   5092
H1F.BE.VI850      ..........G-T---G-G----C-...--------GC----------CCA------AGT------C-------------................................................   4910
H1G.SE.SE6165     ..........---T------G--GT...-G-----------C---------G--T-AG---A-----------------................................................   4967
H1H.CF.90CF056    ..........G-------------...--------------------GA--G--AGT-----------------------................................................   4917
H1J.SE.SE92809    ..........-G--------G--G-...-------------------CA------AGT-----------------------................................................   4884
H1K.CM.MP535C     ..........G-T----G--G-CC-...---------G-----------A------AGT----------------------................................................   4766
H101_AE.TH.CM240  ..........----------G-----...-G-------C----------A...G--A-A-----A-----------------................................................   5141
H1N.CM.YBF30      ..........GG-G----G----G-...--------C--A--------G--T--G--A-----A------------G---A................................................   5164
H1O.CM.MVP5180    ..........-A-GT-----GA-AT...-----T--CC-A--C-----CCA--G-T-A-----A-----------------................................................   5600
H1O.BE.ANT70      ..........-A-G---G--GC-GG...-----T--CC-A--C-----CCA----T-A-----A----------------A................................................   562

MAC.US.SMM251     ...............................................................----TC-................................................   5793
MAC.US.SMM239     ...............................................................----TC-................................................   5817
MAC.US.SMM239 Vif .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__D__V__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__   Vif
MAC.US.SMM239 Vpx                                                                           _M__S__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   Vpx
                                                                                            \ Vpx start

STM.US.STM        ...............................................................---TC-................................................   5464
SMM.US.SMM9       ...............................................................---TCR................................................   5290

H2A.GW.ALI        ..............-A-C---..........................................----C-................................................   5896
H2A.SN.ST         ..............-A-C---..........................................----C-................................................   5348
H2A.DE.BEN        ..............-A-C---..........................................---AC-................................................   5903
H2B.GH.D205       A...AAAA.......................................................-----T................................................   5882
H2B.CI.EHO        A...AAAG.......................................................-----T................................................   5880
H2G.CI.ABT96      .AAAGGAC.......................................................-----T................................................   5233

TANTALUS.UG.TAN1  C...CTAAGA-....................................................----C-G---GAAGA........................................   5753
GRIVET.ET.GRI677  ...............................................................----C-TC--GAAGA........................................   5527
VERVET.KE.VER9063 ..............-ATC-G-..........................................----C-TC--GGAGA........................................   5760
VERVET.KE.VERTYO  ..............-ATC---..........................................----CCTC--GGAGA........................................   5250
VERVET.KE.VER155  ..............-ATC---..........................................----CCTC--GAAGA........................................   5755
VERVET.DE.VERAGM3 ..............-ACC---..........................................----CCTC--G-AGA........................................   5258
SABAEUS.SN.SAB1C  ...............................................................----CCTC--GAGGGTGGCTCCCACCAGTAGGAGGGGATCCTCCCAAGGATCCCCCCAAG   5970

SUN.GA.SIVSUN     ...............................................................----CC-GC.............................................A   5918
LHOEST.KE.LHOEST  ...............................................................---TCC-GGCAGAGG.......................................C   5874
LHOEST.CD.447     ...............................................................---TCC-G--CAAGG.......................................C   4793
LHOEST.CD.485     ...............................................................---TCC-G-GCAAGG.......................................C   4793
LHOEST.CD.524     ...............................................................---TCC-G-CAAAGA.......................................C   4790
MANDRIL.GA.MNDGB1 ...............................................................---...................................................G   5230

COLOBUS.CM.CGU1   A...GAGAGG--C-ACC-G---.........................................---AG-................................................   5286
SYKES.KE.173      ...............................................................----CGG-G--TTTC.........................TTCAATCCCTCCCAGCAT   5733
RCM.NG.NGM        ..............-AGC-GCTA-AG...G.................................----C-G-G-AGA..........................................   5903
MANDRIL.GA.14CG   ....................................................-TCA-GTTTT-CAGAA----C-G--...AGA..........................................   5458
MANDRIL.14CG  Vif __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__I__Q__V__L__Q__K__W__Q__.__K__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.   Vif
MANDRIL.14CG  Vpx                                                                           _M__A__E__.__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._   Vpx
                                                                                            \ Vpx start
```

HIV-1/HIV-2/SIV
complete genomes

```
                                                                                                  Vif end in HIV-1 \
H1B.FR.HXB2   Vif __.__.__.__.__.__.__.__.__.__.__.__Q__K__T__K__G__H__R__G__S__H__T__M__N__G__H__$_                                               Vif
H1B.FR.HXB2   Vpr _.__.__.__.__.__.__.__.__.__.__.__P__E__D__Q__G__P__Q__R__E__P__H__N__E__W__T__L__E__L__L__E__E__L__K__N__E__A__V__R__H__F__.__P__     Vpr
H1B.FR.HXB2       .................................CCAGAAGACCAAGGGCCACAGAGGGAGCCACACAATGAATGGACACTAGAGCTTTTAGAGGAGCTTAAGAATGAAGCTGTTAGACATTTT...CCTA    5664

CPZ.US.CPZUS      .................................-----------------GCC--A--A---T-------G---G--T----TG-C-----A--A--A--AG-A--G---T-GC-C------...----    5724
CPZ.GA.CPZGAB     .................................-----G--------T----CA--A--A--TT--C-G------G--T-----ACC--G--A--AT-A--A--------A--A--------C...----    5728
CPZ.CM.CAM3       .................................--------T--G--T----CA--A------T-----------G--T----A--GC-G-----A--A--------------AC-C------...---C    5078
CPZ.CM.CAM5       .................................-----------G-------CA--A--A---T-----------R-------AA-A-----A-----A--A-C---G---C-A---------...---C    5366
CPZ.CD.CPZANT     .................................--GC-G--TG-G-----------A--A---ATG-----G---TTGT----AAC------A--AA-A--A-----G--A--A--------C...--AC    5121

H1A.UG.U455       .................................-----------G-----------------GT--GC-----A-G--T-----T-G-----A---------C------A--------------...----    5110
H1C.ET.ETH2220    .................................-----------G-----G--------A---T-----------G----------------------C---C-G--------C-----C---...----    5056
H1D.ZR.84ZR085    .................................--G-C-----------G------------T--------------T-----------------------G---------C---------...----    5186
H1F.BE.VI850      .................................----G------G-----------------T-----------G-------AA------------------------------G-------...----    5004
H1G.SE.SE6165     .................................-----------G-----------A--A--CT-----------G--T----A--G-----A--------A----------------------...----    5061
H1H.CF.90CF056    .................................-----------G-------------------------------------------------A----------G--------G-------...----    5011
H1J.SE.SE92809    .................................-----------C-----------A------T--C---------------------G-----------A-----------------C---...----    4978
H1K.CM.MP535C     .................................-----------G-------------------A--------------T-----A----G-------------GA------------------...----    4860
H101_AE.TH.CM240  .................................--------...-G-----------A--A--CT-------------T----A--G-------------A----------------------...----    5232
H1N.CM.YBF30      .................................--------TGC---------------A--CT-T---------G-------AT-A-----A--AT-A--A-----------GC-C------...--A-    5258
H1O.CM.MVP5180    .................................-TG---A-T--G--A---GCT---------TT----------------------CC-G--A-----G--AG-A-----A--A--------C...----    5694
H1O.BE.ANT70      .................................--T--GA-T--G--A---GCT-AA-----CTT----------G----------C-----A-----A--AGCA-----A--A--------C...----    5718

MAC.US.SMM251     GATCCCAGGGAGAGAATCCCACCTGGA.........A-CAGTGG--AAGAGACA-TA-GAGAGGC-TTC------CT-AACAGAACAG--------A-A--C-GA--G--A--A-AC--CC-A...--A-    5911
MAC.US.SMM239     GATCCCAGGGAGAGAATCCCACCTGGA.........A-CAGTGG--AAGAGACA-TA-GAGAGGC-TTC------CT-AACAGAACAG--------A-A--C-GA--G--G--A-AC--CC-A...--A-    5935
MAC.US.SMM239 Vif R__S__Q__G__E__N__P__T__W__.__.__.__K__Q__W__R__R__D__N__R__R__G__L__R__M__G__K__Q__N__S__R__G__D__K__Q__R__G__G__K__P__P__.__T__K     Vif
MAC.US.SMM239 Vpx _D__P__R__E__R__I__P__P__G__.__.__.__N__S__G__E__E__T__I__G__E__G__F__E__W__L__N__R__G__V__E__E__I__N__R__E__A__V__N__H__L__.__P__     Vpx

STM.US.STM        GATCCCAGAGAGAGAATACCGCCCGGA.........A-CAGTGG--AAGAGGCA-TA---GA-GCATTC--G---CT--ACAGAACGG-------CA-A--CCGA--G--A--A-AC---C-A...--G-    5582
SMM.US.SMM9       SRTCCCAGGGAGAGRATCCCACCTGGA.........A-CAGTGG--AAGA-ACRRTA-GCGAGGCYTTC--C---CT--ACAGAACAS----A--AA-A--C-GG-C---A--G-AT-----G...--R-    5408

H2A.GW.ALI        AACCCCAGAGAAACAGTACCACCAGGA.........A-CAGTGGC-AAGAGACT-TC-GAGAGGC-TTC--G---CT-GACAG-ACAG------CCT-A--C-GA--G--G--G-AT--CC-G...--CC    6014
H2A.SN.ST         GGCCCCAGAGAGACAATACCACCAGGA.........A-CAGTGGC-AAGA-ACT-TC-GAGAGGC-TTC--G---CT-GACAG-ACGG----A-CCA-A--C-GA--G--A--G-AC--CC-G...--CC    5466
H2A.DE.BEN        GACCCCAGAGAAAGGGTACCGCCAGGA.........A-CAGTGG--AAGAGACC-TT-GAGAGGC-TTC--G---CT-GAGAG-ACCA----A-CCT-A--C-GG--G--A--G-AC---C-G...--CC    6021
H2B.GH.D205       ...CCCAGGGAGAGAGTGCCACCAGGA.........A-CAG-G-C-AAGA-ACAGTA-GAGA-GCATTC-CT---CT-GA-AGAACAA--ACA-----C--C-GG-T---G--C-AC-----G...--CC    5997
H2B.CI.EHO        ...CCCAGGGAGAGAGTACCACCAGGA.........A-CAG-GGC-AAGA-ACAGTA-GAGAGGCATTC------CT-GAGACAACA-----AC-T--C--C-GA-T---A--C-AT--C--G...--C-    5995
H2G.CI.ABT96      ...CCAAGGGAGAGAGTMCCACCAGAA.........A-CAG-G-C-A-GAGACA-TC-GAGAGGCATTC------CT-GA-AGAACAG-MACA---A-C--C-GGAT---A--C-AT-----A...--C-    5348

TANTALUS.UG.TAN1  GATTCCAGGGAAAGAAGACCCGGATGGTTAGAAATTTGG--T-TCA-C......--------GTGGG-C-----ATT-AG---CA-GG---CA--A--C--TC-G-----ACAAC-C-------...GGG-    5874
GRIVET.ET.GRI677  GATCCAAGAGAACCTTTACCAGGATGG......-TG---AT-TGG-AT-T-G-C---------TGGG-C------CT--A---CA-GC-CAG---T--A--CG-A-----CAGA--G--C----...GGA-    5648
VERVET.KE.VER9063 GGTCCCAGAGAAAACAGGCCAGGAGAA......GT---GAT-TGG-AT-T-A-C--A--A---TGGG-C------CT-AGG--CA-G-----A--T--C--TC-G-----CAAATT--------...GGC-    5881
VERVET.KE.VERTYO  GATCCAAGAGAGGCAAGACCAGGAGAA......GT----ATTTGG-AT-T-AGC--------GTGGG-CA-----CT-AG---TA-G---C-A--T--C--CC-G--G--GAGGCTG-------...GGG-    5371
VERVET.KE.VER155  GATCCCAGAGAGGAGAGACCAGGAGGA......-T----ATTTGG-AT-TCAGC--------GTGGG-C------CTCAG---TA-GG----A--AA-C--C--C-----CAAACTG-------...GGCC    5876
VERVET.DE.VERAGM3 GATCCCAGAGAAGCAAGACCGGGGGAA......-T----ATATGG-ATTTGAGC--------GTGGG-C------CT-AG---CA-GC----A--TA-C--TC-G-----CAAG-TG-------...GGGC    5379
SABAEUS.SN.SAB1C  AATCCCAGAGAAGAGATACCAGGATGG......-T----ACATGG-ATTTG-CC--A-----GTT-G-C------CTC-G---CA-G---C----C--C--TTCA--G--CCAGT-C--C--C...--A-    6091

SUN.GA.SIVSUN     GGAGAGAGCCAGTAGAACAGCCACCA.........GA-GACGAGG--C-AC-GAGA-AGC-ATTTGC-CA-TG-CTGG-AGATA-CA-GGA--A-ATCAGG--CGA-GCGGAAAG-CAT--C...C-AGC    6036
LHOEST.KE.LHOEST  AGCAAAGAAGTCCAGAGAGACCACCA.........GAGGACTTTG-AC-AC--AG--AGC-ATGGGGTCAGTG-TTAG-AGACA-CA-GGAG-A-ATAA----GGAGGCAA-ACGCCAT--C...C-A..    5990
LHOEST.CD.447     AGCAAAGGAGTCCAGAAAAACCTCCA.........GAGGACTTTG--C-AC--AG--AGC-ATGGGGTCA-TG-CTAG-AGACA-CA-GGAG-A-ATAA----GGAGGC-A-ACGCCAT--C...C-A..    4909
LHOEST.CD.485     AGCAAAGGAGTCCAGAAAGACCTCCA.........GAGGACTTTG-AC-AC--AGA-AGC-ATGGGGTCA-TG-TTAG-AGACA-CA-GGAG-A-ATAA-G--GGAGGC-A-ACGCCAT--C...C-A..    4909
LHOEST.CD.524     AGCAGCGCAGCCCAGAAAGGCCCCCA.........GAGGACTTTG-TC-TC--AGA-AGC-ATGGGCTCA-TG-CTAG-AGACA-AA-GGAG-A-ATCA-G---GAGGCAA-ACGCCAT--C...C-A..    4906
MANDRIL.GA.MNDGB1 GCCAGAAGAGAGATGAGCAAGTATCA.........GA-GAT-A-G-AC-TC-CAGA-AGC-ATACA-TCAGTG-CTAG-AGATA--A-GGAG-A-ATAA-G---GA-GCAA-AA--CAC--C...C-TCT    5348

COLOBUS.CM.CGU1   ...GTCCTGAAAGGGACTTCACA.............................................TG-TTGCAGGACCAGG-AGAGAACATAACCCGGGAGGC-A-AAG-CAC--C...C-T-T    5362
SYKES.KE.173      GTGCAAGGAACACCATGGTTCTTCATTCCA............AG-AATGT-G--TT-ACC---A--GTGATCAATGTCACT-TAAAGGC---ACT-G-AGT--CA-----ATCC-AG------ACT---C    5851
RCM.NG.NGM        ...GAAAGAGTGCCAGAGGCCCCCACTGGG............GCT--AGATGTA......GAGTTTGCCCCC---CTT-ACAGAA-GC--ACA--AG-C--CTTA-----CAGGTTG--C---CAT--AG    6012
MANDRIL.GA.14CG   .........GCACCAGAG..............G--...CCA--G---G--GGAGAA-T-GG--TAG-AC-G---CTGGA-AGATCA--G--AC-AG-G--C-GA--G---CAATT---C--CCAC--AG    5561
MANDRIL.14CG  Vif .__.__.__S__T__R__.__.__.__.__.__G__.__T__T__G__G__R__R__S__G__T__R__T__V__G__G__K__I__I__G__T__S__E__Q__R__G__S__I__T__L__P__P__R     Vif
MANDRIL.14CG  Vpx _.__.__.__A__P__E__.__.__.__.__.__A__.__P__Q__G__A__G__E__V__G__L__E__Q__W__L__E__R__S__L__E__Q__V__N__R__E__A__Q__L__H__F__H__P__     Vpx
```

```
H1B.FR.HXB2        GGATTTGGCTCCATGGCTTAGGGCAACATATCTATGAAACTTATGGGGAT.......................ACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGCAACAACTGCTGTTTATCCA    5770
H1B.FR.HXB2   Vpr R__I__W__L__H__G__L__G__Q__H__I__Y__E__T__Y__G__D__.__.__.__.__.__.__.__.__T__W__A__G__V__E__A__I__I__R__I__L__Q__Q__L__L__F__I__H     Vpr

CPZ.US.CPZUS       -ACCC-----A--A--------A--GT-C-----CAGT--------A--C.......................-------AG-----A-----A--T-----G--CT-A--------TT----------    5830
CPZ.GA.CPZGAB      -ACCA---T-G---CAA-----A---TTC--T-----C--------A---.......................--A----T------A--G-----C--------CT-A-----CT------------    5834
CPZ.CM.CAM3        -AGCA-----A-----A-----A--GTT---T-----C--------A---.......................--A--------T--A-----A--C--T-----A--------G--T----------    5184
CPZ.CM.CAM5        -CCC------A---A-AC----A--GT-C----------------A---.......................--A---AT---G--A-----A--T--T-----AT-------T--T------G----    5472
CPZ.CD.CPZANT      A-CC-ACAT-A--ACA-C-T--AA-TTGGG-A----C--A---C--A---.......................T-A---AG-------C-G-TTT-G----TTC-C--T---A--GCT-------CA--    5227

H1A.UG.U455        --CAG-----------A-----A------------A-C--C--------.......................-------AG-----T-----T------------T--------T-A--------T--    5216
H1C.ET.ETH2220     -ACCA---------AA------A---T----------C--------.......................------T-G-----A-----TT---------C------------AA----------    5162
H1D.ZR.84ZR085     ----A-----T---A-------A---T----------------------.......................--C----------T-----T----------------------A--------T--    5292
H1F.BE.VI850       --CCA-----T--C--------C------------A-C--C--------C.......................-------AG-----T-----T--------G--AT----------A----------    5110
H1G.SE.SE6165      --C-------T-----------A--G---------A-C-----------.......................-------A------T-----------------A--A--------A----------    5167
H1H.CF.90CF056     -AG-A---------CAA-----A--G---------A-C--C-----A---.......................-------T------T-----TT---------CG-----------A--------T--    5117
H1J.SE.SE92809     --CCA-----T---A-T-----A--GT--------AGC--------A---.......................--A----A------C-----------------A-----------AT-------T--    5084
H1K.CM.MP535C      --CCA---------AATC-G--A--G---------AC------------.......................-------A----C-T-----T-----------AT----------A--------T--    4966
H101_AE.TH.CM240   --CCC--------A--------A--GT-C------A-C-A----------.......................-------A---G--T-----T-----------GT----------A------G-T--    5338
H1N.CM.YBF30       ----------A-----G-----A-----C------A-C--A-----A--C.......................--C----AG--G--A--G--A--T--C--G--A--A------T-A----------    5364
H1O.CM.MVP5180     --CC----T-A--A-C--GT-----GT-C--T-----G--------A--C.......................-------A------TATG--A--T--------CT-A--------A-------C---    5800
H1O.BE.ANT70       --CC------A--C-C---G--A---T-C--T-----G-----------C.......................-------T------TATG--A--T--------CT-A------T---------C---    5824

MAC.US.SMM251      --GAGCT-A-TTTCCAGG-TT-----AGGTCT-GG---TAC-GGCAT---..........GAACAAGGGATGTCACAAA-CTAT-T-AAAT-CAGATACT-GT-TT-AA-----A-GGCTT-A-----G--    6032
MAC.US.SMM239      --GAGCTAA-TTTCCAGG-TT-----AGGTCT-GG---TAC-GGCAT---..........GAACAAGGGATGTCAC-AA-CTAT-T-AAAT-CAGATACT-GT-TT-AA-A---A-GGCTT-A-----G--    6056
MAC.US.SMM239 Vpx R__E__L__I__F__Q__V__W__Q__R__S__W__E__Y__W__H__D__.__.__.__E__Q__G__M__S__P__S__Y__V__K__Y__R__Y__L__C__L__I__Q__K__A__L__F__M__H     Vpx
MAC.US.SMM239 Vif __G__A__N__F__P__G__L__A__K__V__L__G__I__L__A__$_                                                                                                  Vif
                            Vif end in HIV-2 and SMM sequences /

STM.US.STM         --GAACT-A-TTTCCAGG-CT---G-AGGTCT-GG---TAC-GGCAT---..........GAGCAAGGGATGCCCGGGA-CTAT-TTAAAT-TAGATACT-GT-CT-GA-A--GA--GCTT-------G--    5703
SMM.US.SMM9        --GAGCTAA-TTTCCAGG-CT---G-AGGKC--GG---TAC-GGCAT---..........GAAATGGGGATGTCAGTCA-CTACACTAAATRCAGATACT-GT-CT-GA-A--GA--GCT--------G--    5529

H2A.GW.ALI         --GAGCTCA-TTTCCAGG-GT-----AGGTC--GGAG-TAC-GGCAT---..........GAACAAGGAATGTCACAAA-TTACAC-AAAT-TAGATACT-GT-CT-AA----GA--GCTA---AC-CA--    6135
H2A.SN.ST          -AGAGCTTA-TTTCCAGG-GT-----AGGTC--GGAG-TAC-GGCAT---..........GAACAAGGAATGTCA-TAA-TTACAC-AA-T-TAGATATT-GT-CC-AA----GA--GCTA----C--A--    5587
H2A.DE.BEN         -AGAGCTCA-TTTCCAGG-GT-----AGGTC--GGAG-TA--GGCAT---..........GAACAAGGGATGTCAG-AA-CTACAC-AA-T-TAGATATT-GT-CC-AA-----A--GCTA-A----CA--    6142
H2B.GH.D205        -AGAACTTA-TTTCCAGG-CT----GAGGTCT-GG-C-TAC-GGC-T--G........GAACAGGGCATGTCA-T-A-CTATACCAAAT-TAGATACT-GTTGC-AA----GA--GCAA-----G-G--    6118
H2B.CI.EHO         --GAACTTA-TTTCCAGG-CT----GA-GTCT-GG-C-TA--GGC-T--G........GAACAGGGCATGTCA-T-A-TTATACCAAAT-TAGGTATT-GT-TC-AA-----A--GCAA-------A--    6116
H2G.CI.ABT96       --GAGCTTA-TTTCCAGG-TT---G-AGGTC--GG---TA--GGCAT---..........GAACAGGGCATGTCAG--A-TTACACTAA-T-TAGATATT-GTTGC-GA-----A-GGCTA-----G-A--    5469

TANTALUS.UG.TAN1   -AGAACTC---TTCCAAG--T--A-C...T---G-C-GGAAG-A-----AAGG.........AATGGGGCACCA-TGATA-A-A-G-CTT-TAGATATTAT--GT-GG-A--GA-GGCTT-A---G-G--    5992
GRIVET.ET.GRI677   T--ACAT---AATCC-AG--T--A-T...TA--G--T-GAGG-G--AAGGAGA.........CATAATACCCCATGGAAT-AGAT--GCT-CAAGTACTAT------G-T---A-GTCTA-----G-A--    5766
VERVET.KE.VER9063  -AGAGCTC---TTCCAAG-TT--A-T...TAT-G-C-GGAGG-A--A--GAGA.........CGGGGAGCCCCA-TGAC--AGA-G-CCT-TAGATACTAT--GT-GG-T--GA--GCA--C---G-T--    5999
VERVET.KE.VERTYO   -AGAACTC--GTTCCAAG--T--A-C...TA--G-C-GGAGG-A--A--AAGA.........CATGGTACTCCC-TGAT--A-A-G-CCT-CAAATATTAT--GC-AG-A---A-GGCT--C---G-G--    5489
VERVET.KE.VER155   -AGAACTC--AT-CCAAG--T--A-T...TAT-G-C-GGAGG-A-----GAGA.........CAGGGAAGACCC-TAGC--A-A-G-CAT-TAAGTATTATC-CT-AG-T--GA--GCT--C---G-G--    5994
VERVET.DE.VERAGM3  -CGAGCTC--GTTCCAAG--T--A-C...TAT-G-C-GGAGG-G--A--AAGG.........AATCGCACTCCC-TGCTA-A-A-G-CTT-TAAATATTAT-A-T-GG-----A--GCT--C---G-G--    5497
SABAEUS.SN.SAB1C   ---A-CTC--TTTCC-TC-TT--TGGA----TGTG---GAACCA-CCAT-.........GATCATGGGCAAACC-GGCT--A---CTG-T-TAAATATTGT------------A-GGCTT-A---G-G--    6212

SUN.GA.SIVSUN      A-CAA-A-T-G-C...-AAGTCTTGCA---CTGTGCG......................GACTCTACAGGGTCAGAAAA--A---GTGTATGAGAGC---C-CTT-G--CA-TAG-GCCT----C--T--    6141
LHOEST.KE.LHOEST   .--AG-AATGTTGGCC-AAGTCT-GCA----TGTGT-......................GACAGTACAGGGTCAGAAGA--AG-CCTGCATGAAAT-C--C-CCC-AG-CA-TAG-GCTT-----C-T--    6097
LHOEST.CD.447      .--AACAAT-TTGGCC-AAGTCT--GA---CTGTGTC......................AGCAGTACCGGGTCAGAGGA-AA--CCTGCATGAAAT-C--T-CC--AG-AA-TAG-GCA--C-----T--    5016
LHOEST.CD.485      .--AGCCAT-TTGGCC-AAGTCT-GGA---CTGTGTC......................AGCAGTACCGGGTCAGAGGA--A--CCTGCATGAGAT-C--T-CC--AG-AA-TAG-GCA--C-----T--    5016
LHOEST.CD.524      .--AGCAAT-TTGGCC-AAGTCT-GGA---CTGTGTC......................AGCAGTACCGGGTCAGAAGA--AG-CTTGCATGAGAT-C--C-CTT-AG-TA-TAG-GCT--C-A-T-A--    5013
MANDRIL.GA.MNDGB1  CAT-A-CCTAAATGCAG-ATCA-A-TAT-G-G-GCA----........................ACAGGGAGTGAGGAA-AG-CCTGT--GAAAT-T--T-CCT-AA--A-TAG-GCCA-T-GGG----    5453

COLOBUS.CM.CGU1    -AC-CACATCT--GT-A-A-T-TGGGGT-G-ACTA---GG......................CACTGGGGGAGAC-G---A-G-CT-GTCTG--TTATG-G-----CA--G-GTGTGCC--C-ACT-G--    5470
SYKES.KE.173       AAGAAATATAT...--AG-TT--A-T--ATC--TAA-CGAA...-A--CAGGGACTGATTCCCCGACAATGGCTTGGGAAAG-ACCA--TTG-A---GG-----GCA--CA-T-TTA----C---GAA--    5975
RCM.NG.NGM         A-T-CATTT---G-CTG-GGC-TACTTG-G-GG-AC-C...-GGCAT---..........AGGCTTGGAAGAAGCCT-GA-TAT-C--GCT-TAGATATC-GCTTC-GA-----A--GCT--------T--    6130
MANDRIL.GA.14CG    A-T-CCTTT---G-CTT-GGAACACC...TGTGTA---CAC-GGCAT---..........AGACTTAGAAGAACCTTAAATTAT-CCAAAT-TAGGTATC-TCTGC-GA----GA--GCCA-----G-T--    5679
MANDRIL.14CG  Vpx  E__F__L__F__R__L__W__N__T__.__C__V__E__H__W__H__D__.__.__.__R__L__R__R__T__L__N__Y__A__K__Y__R__Y__L__L__L__M__Q__K__A__M__F__V__H     Vpx
MANDRIL.14CG. Vif  __V__P__F__P__S__L__E__H__L__.__C__R__T__L__A__$_                                                                                                  Vif
                            Vif end in HIV-2 and mandril sequences /
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
 HXB2 Vpr premature end (without frameshift)\                                                                  / Tat exon 1 start
 H1B.FR.HXB2   Vpr _#._F__Q__N__W__V__S__T__$_                                                                  _M__E__P__V__D__P__R__.__.__.__.      Tat
      frameshift insert in HXB2                                                                          HIV-1 Vpr normal end \
H1B.FR.HXB2   Vpr _..__F__R__I__G__C__R__H__S__R__.__I__G__V__.__.__.__.__.__T__R__.__.__.__.__.__.__Q__R__R__A__R__N__G__A__S__R__S__$_                  Vpr
H1B.FR.HXB2       T..TTTCAGAATTGGGTGTCGACATAGCAGA...ATAGGCGTT..............ACTCGA................CAGAGGAGAGCAAGAAATGGAGCCAGTAGATCCTAG...........   5850

CPZ.US.CPZUS      -...-AT--G-----C---AAT--------G...-----GA--TCCTTGGCAAGGAGA----CT............CAA......G-----AGG-ATGGATCCAGT--ATCC-AA..............   5921
CPZ.GA.CPZGAB     -...--T---C-------C-AG----------..------A-C...............CTC-C-............CAA...AGA--A---T-C-ATGGATCCAAT--ATCC-GA..............   5913
CPZ.CM.CAM3       -...--T--GC----C----A----------..-----GA-C...............-TC-CT................GG------GAG--ATGGA-CCAGT--ATCC-GA..............   5260
CPZ.CM.CAM5       -...--T--GC----C----A----------..-----GA--..............CT--YT................G----A--GAG--ATGGAACCAGT--ATCY-AA..............   5548
CPZ.CD.CPZANT     -...-AT---CAC--A---GCT--------G...---.......................................-G-CATGGACCCCGT--AC-C-GA..............   5282

H1A.UG.U455       -...--------------C-A----------..------A--................-T--C..................GG---A----GC--G--------------------A...........   5295
H1C.ET.ETH2220    -...------------A--C-AG----------..------A--................TTA---..................-----A----------------------------A...........   5241
H1D.ZR.84ZR085    -...----------------A----------..------A--................----C..................-----A---------------T-------------A...........   5371
H1F.BE.VI850      -...--------------C--T----------..------A--................GT--C..................-----A----T---------------------------.........   5189
H1G.SE.SE6165     -...--------------C-A----------..------A--................--A-C..................-G---A--G-T---GG-----C--G----------A...........   5246
H1H.CF.90CF056    -...--------------C-A----------..-----AA--.......................................-----A----T----------C-------------A...........   5196
H1J.SE.SE92809    -...---------C-----C-AT----------..------A--................-TA-C..................--A--A----GG--G--------------------A...........   5163
H1K.CM.MP535C     -...--------------C-AC----------..------A--................-TA-C..................-----A----GG---------T-------------A...........   5045
H101_AE.TH.CM240  -...----------------A----------..------A--................-TG-C..................GG---A----GC--G------A-TG----------A...........   5417
H1N.CM.YBF30      -...-AT--G-----C--C-AG--C-------..-----GA-C..................----CT..................--A------AGG-ATGGAACCAGT--ATCC-AG..............   5440
H1O.CM.MVP5180    -...-AT--------A--C-A------T---..------AA--................CTC-C..................TCT-AC-C-AG-G---GA---AGA--A-ATGGA-CCAGTAGATCCTGA   5891
H1O.BE.ANT70      -...-AT--------A--C-A------T---..------AA--................-AC-C..................AGAG-A----G-----GAAAT-GATCC--TAGA-CCTGA.........   5906

MAC.US.SMM251     -...-G--AG-AA--C---A--............TGTCTA-GG...............GAAG--...................--CG--GC--G-G--TGG...AGACCAG--C-TCCTCCTCCTCCCCCT   6111
MAC.US.SMM239     -...-G--AG-AA--C---A--............TGTCTA-GG...............GAAG--...................--TG--GC--GGG--TGG...AGACCAG--C-TCCTCCTCCTCCCCCT   6135
MAC.US.SMM239 Vpx __._.__C__K__K__G__C__R__.__.__.__.__C__L__G__.__.__.__.__E__G__.__.__.__.__.__.__H__G__A__G__G__W__.__R__P__G__P__P__P__P__P__P_    Vpx

STM.US.STM        -...AG--AG-GG--A--CA--............TGCTTG-GG...............GAAG--...................--TG--GC--G-G--TGG...AGACCAG-CC-TCCTCCTCCTCCCCCT   5782
SMM.US.SMM9       -...-G--AG-AA------A-G...........TGCTTA-GA...............GGAGAG..................--TG--GC--GGG--TGG...AGACCAG-GC-TCCTCCTCCTCCCCCT   5608

H2A.GW.ALI        -...----TG-AA-----CACC............TGCCTG-GG...............GGAG-G..................--TG--CC--G-G--TGG...AGATCAG--C-TCCCCCTCCTCCCCCT   6214
H2A.SN.ST         -...-CT-AG-GA-----CACT............TGCCTG-GG...............GGAG--...................--TG--CCG-G-G--TGG...AGATCAG--C-TCCCCCTCCTCCCCCT   5666
H2A.DE.BEN        -...----AG-GA-----CACT............TGCT-G-GG...............GAGGAC.................AT-G-CC-G-A-G--TTG-A--A-CAAG--C-TCCCCCTCCTCCCCCT   6224
H2B.GH.D205       C...-AT-C--AG--C---A-G............TGCCTGCAG...............GAGG-C..................--TG--CC--GGG--TGG...AGATCAG--C-TCCTCCTCCTCCTCCC   6197
H2B.CI.EHO        C...--TGC--AG--T---G--............TGTCTGCGG...............GAAG-C..................--TG--CC--GGG--TGG...AGATCAG--C-TCCTCCTCCTCCTCCC   6195
H2G.CI.ABT96      -...-GT-AG-AA--A--CAC-............TGCTTGCAG...............-AGG-C..................--TG--CCT-G-G--TGG...AGACAAG--C-TCCTCCTCCTCCTCCT   5548

TANTALUS.UG.TAN1  -...--TC--TG---C--CA-GAGA--A-CT...CCTTTT-AG...............C-CTAC.................G-AGAA--GAG--ATGGA-TC-GAG-GG--CGGG-TGGC.........   6074
GRIVET.ET.GRI677  -...------TG---T---A--AGG--AG-----CCTTTTTCC...............C--TAC.................G-AGA---GAG--ATGGACA--GAG-AG--G--...CCACCCCCTCCT   5854
VERVET.KE.VER9063 -...---C-TTG---T-----TAGA--ACAG...CCCTT--AA...............C--TAC.................G--GAA---AG-GATGGACA--GGG-AG--AGAGC-AACCGTGCTCCA   6090
VERVET.KE.VERTYO  -...--TC--TG------CA-GAGA--GCAG...CCCTTT-AA...............C-ATAC.................G--GA---GAG-GATGGACA--GGG-AG-CAGAGCAAATCGTGTCCCA   5580
VERVET.KE.VER155  -...---C-GTG---A-----CAGG--ACA-...CCCTTT-AG...............C-ATAC.................G--GA---GAG--ATGGACA--GGG-AG--AGACCAGGACGTGTCCCA   6085
VERVET.DE.VERAGM3 -...---C-GTG---A--C--CAGA--ACA-...CCCTTT-AA...............C-ATAC.................G-AGAA--GAGGGATGGACA--GGG-AG--CGAGCAGGGCGCGTACCA   5588
SABAEUS.SN.SAB1C  -...A-G-A-GG-A-A---T-TA-G.........CC-AAGAC-........CACCCTG-GTATGGACCAGGAGCAGGAGGC-C-CCCCC--GTCTGGGA-----TTCAG--GGAGCTGCATCGGCCGCT   6321

SUN.GA.SIVSUN     -...C-GCC-...AT-----CTTC-TA--TCAGA--CA-AAG-..............GGATC-..................GGA-ATGTCAACCC--GGCC-C-A-CA---C-AAGACCAGGGCAAGGG   6232
LHOEST.KE.LHOEST  C...C--CC-...ACT----CCAGCTATGC-...GC--C-AGA..............G-AG--..................ACAGCCCATC---C-G-A-C---ACCA---C-TA-TGAAGGAGACATA   6185
LHOEST.CD.447     -...--GCCT...ACC----CCAGCTATGC-...GC--CAAGA..............G-AG--..................ACAGCTCACC-G-C-G-A-C---ACCA---C--ACTAAAAGAGACATA   5104
LHOEST.CD.485     -...--GCCT...ACC----CCAGCTATGC-...GC-ACAAGA..............G-AG--..................ACAGCTCACC-G-C-G-A-C---ACCA-A-C--ACTGAAAGAGATATA   5104
LHOEST.CD.524     -...C-GCCT...CAA----CCAG-TATGC-...GC--C-AGA..............G-AG--..................ACAGCACACC---C-G-A-C---ACC----C-AACTGGAAGAGATATA   5101
MANDRIL.GA.MNDGB1 C...C-AGCTCAA------GATGGA-C-TTC...-GG-AAAGA..............-GA-C..................--ACT-CCCC-CTC-GGATTCAGGCCA---GGAG-TAGATTATAA...   5541

COLOBUS.CM.CGU1   C.........................................................GTT.......................CAT...GCGGGGCA-TGTC-CTTTC---CTACGCAAAACCCT   5513
SYKES.KE.173      -...--TGCTGCA--A----CT--AC-G-CC...-G-TAT-CC...............-GA-AC.................AGAG--TACC-CCATCC-TCTTAG.......................   6042
RCM.NG.NGM        -...-G-CA-TCA------TCT--G...............AGA.......................................--TG-ACAG-G-CA-GCAA-G-AAGCAG--GAAAGAATCCAGATTCTT   6203
MANDRIL.GA.14CG   -...A-GCA-CAA--A--C-C-TG---A--TGGACATTCAAGA.................G--.....................................................CCTCCTCCT   5731
MANDRIL.14CG  Vpx __.__M__Q__Q__G__C__P__C__R__S__G__H__S__R__.__.__.__.__.__.__G__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__P__P__P_    Vpx
```

```
H1B.FR.HXB2          ..................................................................................................... 5850
H1B.FR.HXB2    Tat   _.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__ Tat

CPZ.US.CPZUS         ..................................................................................................... 5921
CPZ.GA.CPZGAB        ..................................................................................................... 5913
CPZ.CM.CAM3          ..................................................................................................... 5260
CPZ.CM.CAM5          ..................................................................................................... 5548
CPZ.CD.CPZANT        ..................................................................................................... 5282

H1A.UG.U455          ..................................................................................................... 5295
H1C.ET.ETH2220       ..................................................................................................... 5241
H1D.ZR.84ZR085       ..................................................................................................... 5371
H1F.BE.VI850         ..................................................................................................... 5189
H1G.SE.SE6165        ..................................................................................................... 5246
H1H.CF.90CF056       ..................................................................................................... 5196
H1J.SE.SE92809       ..................................................................................................... 5163
H1K.CM.MP535C        ..................................................................................................... 5045
H101_AE.TH.CM240     ..................................................................................................... 5417
H1N.CM.YBF30         ..................................................................................................... 5440
H1O.CM.MVP5180       ..................................................................................................... 5891
H1O.BE.ANT70         ..................................................................................................... 5906

MAC.US.SMM251        CCAGGACTAGCATA...A...............................ATGGAAGAAAGACCTCCAGAAAATGAAGGCCCACAAAGGGAACCATGGGATGAATGGGTAGTGGAGGTTCTGGAAGAAT 6205
MAC.US.SMM239        CCAGGACTAGCATA...A...............................ATGGAAGAAAGACCTCCAGAAAATGAAGGACCACAAAGGGAACCATGGGATGAATGGGTAGTGGAGGTTCTGGAAGAAC 6229
                           Vpx end \                                 / Vpr start
MAC.SMM239 Vpx       _P__G__L__A__$_                                 _M__E__E__R__P__P__E__N__E__G__P__Q__R__E__P__W__D__E__W__V__V__E__V__L__E__E__  Vpr

STM.US.STM           CCAGGCTTAGCATA...G...............................A.ATGACACACAGACCTCCAGAGGATGAAGGCCCACAGAGAGAACCATGGGATGAGTGGGTAGTGGAAGTTCTGGAGGAAA 5877
SMM.US.SMM9          CCAGGACTAGCATA...A...............................A.ATGACAGAAAGACCTCCAGAAGATGAAGCCCCACAGAGGGAACCGTGGGATGAATGGGTRGTAGAAGTTCTGGAGGAAC 5703

H2A.GW.ALI           CCAGGTCTAGTCTA...A......................TGACTGAAGCACCAACAGAGTTTCCCCCGGCCGGAATGGGACCCCACCAGGGAGCCAGGGATGAGTGGGTAATAGAAGTCCTGAGAGAAA 6319
H2A.SN.ST            CCAGGTCTAGTCTA...A......................TGACTGAAGCACCAACAGAGTCTCCCCCGGAGGATAGGACCCCACCGAGGGAGCCAGGGGATGAGTGGGTAATAGAAACCCTGAGAGAGA 5771
H2A.DE.BEN           CCAGGTCTAGTCTA...A......................TGACTGAAGCACCAACAGAGTTTCCCCCAGAAGATGGGACCCCACGGAGGGACTTAGGGAGTGACTGGGTAATAGAAACTCTGAGGGAAA 6329
H2B.GH.D205          CCAGGCCTGGCCTA...A......................TGGCAGAAGCAGCCCCAGAGATCCCTCCAGAGAACGAGAACCCACAAAGAGAACCGTGGGAAGAGTGGATAGGGGAGATCCTGGAGGAAA 6302
H2B.CI.EHO           CCAGGTCTAGCCTA...A......................TGGCAGAAGCAGTCCCAGAGATTCCTCCAGAGGATAAAAACCCACAAAGAGAACCGTGGGAACAGTGGGTAGTGGACGTCCTGGAGGAAA 6300
H2G.CI.ABT96         CCAGGTCTAGCATAG.AT......................GGCAGAAGAAATCCCCCCA..........GAAGATGAGGCCCCCCAGAGAGAGCCATGGGATGAATGGGTAGTAGAAGTTCTGGAAGAAA 5645

TANTALUS.UG.TAN1     ..................................................................................................... 6074
GRIVET.ET.GRI677     CCAGGACTTGCATA...G...................................................................................... 5869
VERVET.KE.VER9063    CCAGGACTTGATTA...G...................................................................................... 6105
VERVET.KE.VERTYO     CCAGGACTTGAGTG...A...................................................................................... 5595
VERVET.KE.VER155     CCAGGACTTGATTA...A...................................................................................... 6100
VERVET.DE.VERAGM3    CCAGGACTTGATTG...A...................................................................................... 5603
SABAEUS.SN.SAB1C     CCAGGCCTGTGA............................................................................................ 6333

SUN.GA.SIVSUN        GACATTAGAAGAGG...CCTACAAGACTAA.......................................................................... 6259
LHOEST.KE.LHOEST     CAACGAGGCAGTGA...CTAA................................................................................ 6203
LHOEST.CD.447        CAAAGAGGCCGTTA...CTAA................................................................................ 5122
LHOEST.CD.485        CAAAGAGGCCGTTA...CTAA................................................................................ 5122
LHOEST.CD.524        CAAAGAGGCCATTA...CTGA................................................................................ 5119
MANDRIL.GA.MNDGB1    ..................................................................................................... 5541

COLOBUS.CM.CGU1      TAG..................................................................................................... 5516
SYKES.KE.173         ..................................................................................................... 6042
RCM.NG.NGM           CCGGGAATGTAA.................TCAAGCACTCCTCTCATTGCAGAGATGGAGATGCTCCCGGAAGATGAAGAACCACAGAGAGAACCCTATGATGAATGGCTAATGGATACCCTTATAGAGA 6315
MANDRIL.GA.14CG      CCAGGATTGGCCTAA......................CTTCTGTCTTGCAG..ATGGAACAGCCGCCAGAGGATGAGGCTCCACAAAGAGAACCATATAATGAGTGGCTGATAGATACATTAGCAGAAA 5836
MANDRIL.14CG   Vpx   _P__G__L__A__$_                                   _M__E__Q__P__P__E__D__E__A__P__Q__R__E__P__Y__N__E__W__L__I__D__T__L__A__E__  Vpr
                           Vpx end /                                   \ Vpr start
```

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2          .................................................................................................................................  5850
H1B.FR.HXB2   Tat    .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.   Tat

CPZ.US.CPZUS         .................................................................................................................................  5921
CPZ.GA.CPZGAB        .................................................................................................................................  5913
CPZ.CM.CAM3          .................................................................................................................................  5260
CPZ.CM.CAM5          .................................................................................................................................  5548
CPZ.CD.CPZANT        .................................................................................................................................  5282

H1A.UG.U455          .................................................................................................................................  5295
H1C.ET.ETH2220       .................................................................................................................................  5241
H1D.ZR.84ZR085       .................................................................................................................................  5371
H1F.BE.VI850         .................................................................................................................................  5189
H1G.SE.SE6165        .................................................................................................................................  5246
H1H.CF.90CF056       .................................................................................................................................  5196
H1J.SE.SE92809       .................................................................................................................................  5163
H1K.CM.MP535C        .................................................................................................................................  5045
H101_AE.TH.CM240     .................................................................................................................................  5417
H1N.CM.YBF30         .................................................................................................................................  5440
H1O.CM.MVP5180       .................................................................................................................................  5891
H1O.BE.ANT70         .................................................................................................................................  5906

MAC.US.SMM251        TGAAAGAAGAAGCTTTAAAACATTTTGATCCTCGCTTGCTAACTGCACTTGGTAATCATATCTATAATAGACATGGAGACACCCTTGAGGGAGCAGGAGAACTCATTAGAATCCTCCAACGAGCGCTCTT  6335
MAC.US.SMM239        TGAAAGAAGAAGCTTTAAAACATTTTGATCCTCGCTTGCTAACTGCACTTGGTAATCATATCTATAATAGACATGGAGACACCCTTGAGGGAGCAGGAGAACTCATTAGAATCCTCCAACGAGCGCTCTT  6359
MAC.US.SMM239 Vpr    L__K__E__E__A__L__K__H__F__D__P__R__L__L__T__A__L__G__N__H__I__Y__N__R__H__G__D__T__L__E__G__A__G__E__L__I__R__I__L__Q__R__A__L__F   Vpr
MAC.US.SMM239 Tat                                                                                   _M__E__T__P__L__R__E__Q__E__N__S__L__E__S__S__N__E__R__S__   Tat
                                                                                                    \ Tat exon 1 start

STM.US.STM           TAAAACAAGAAGCCCTGCGGCATTTTGATCCTCGCTTGCTAAGCGCGCTTGGTAACTATATCTATGATAGACATGGAGACACCCTTGAAGGAGCAGGAGAGCTCATTAAGATCCTCCAGCGAGCCCTCTT  6007
SMM.US.SMM9          KAAAAGAAGAAGCCCTGAACCAKTTTGATCCTCGCTTGCTAATYGCGCTTGGTAACTATATCTATGATAGACATGGAGACACCCTTGAAGGAGCAGGAGAGCTCATTAGAATCCTCCAGGGAGCACTCTT  5833

H2A.GW.ALI           TAAAAGAAGAAGCTTTAAGGCATTTTGACCCTCGCATGCTAATTGCTCTTGGCGGCTATATCTATACTAGACATGGAGACACCCTTGAAAGAGCCAGGGAGCTCATTAATGCCCTACAACGAGCCCTCTT  6449
H2A.SN.ST            TAAAATAAGAAGCTTTAAAGCACTTTGACCCTCGCTTGCTAATTACTCTTGGCAACTATATCTATGCTAGACATGGAGACACCCTTGAAGGCGCCAGAGGGCTCATTAGGATCCTACAACGAGCCCTCCT  5901
H2A.DE.BEN           TAAAGGAAGAAGCCTTAAGACATTTTGATCCCCGCTTGCTAATTGCTCTTGGCTACTATATCCATAATAGACATGGAGACACCCTTGAAGGCGCCAGAGAGCTCATTAAAACCCTACAACGAGCCCTCTT  6459
H2B.GH.D205          TAAAGCAAGAAGCCTTAAAGCATTTTGATCCTCGCTTGCTAACTGCGCTTGGTAACTTTATCTACAGTAGGCATGGAGATACCCTTGCAGGAGCAGGAGAGCTCATTAAAATCCTCCAACGAGCCCTCTT  6432
H2B.CI.EHO           TAAAACAAGAAGCCTTAAAGCATTTTGACCCCCGCTTATTAACTGCACTTGGCAATTTTATCTACAATAGGCATGGAAATACCCTTGAAGGAGCAGGAGAGCTCATTAAACTCCTCCAGCGGGCACTCTT  6430
H2G.CI.ABT96         TCAAAGAAGAAGCCTTAAAACATTTTGACCCTCGCTTGCTAACTGCGCTTGGTAACTATATCTATGATAGGCATGGAGACACCCTCGAGGGAGCAGGAGAACTCATTAAAATCCTGCAGAGAGCTCTCTT  5775

TANTALUS.UG.TAN1     .................................................................................................................................  6074
GRIVET.ET.GRI677     .................................................................................................................................  5869
VERVET.KE.VER9063    .................................................................................................................................  6105
VERVET.KE.VERTYO     .................................................................................................................................  5595
VERVET.KE.VER155     .................................................................................................................................  6100
VERVET.DE.VERAGM3    .................................................................................................................................  5603
SABAEUS.SN.SAB1C     .................................................................................................................................  6333

SUN.GA.SIVSUN        .................................................................................................................................  6259
LHOEST.KE.LHOEST     .................................................................................................................................  6203
LHOEST.CD.447        .................................................................................................................................  5122
LHOEST.CD.485        .................................................................................................................................  5122
LHOEST.CD.524        .................................................................................................................................  5119
MANDRIL.GA.MNDGB1    .................................................................................................................................  5541
MANDRIL.GA.14CG      TTCAAGAAGAAGCTTTGAAGCATTTTGATAGGCGCTTATTGCATGCAGTAGGAAGCTGGATTTATGAAACATATGGGGACACCTTGGAGGGGGTGCAGAAGCTTATCACCCTCTTACAGAGAGCTCTATT  5966

COLOBUS.CM.CGU1      .................................................................................................................................  5516
SYKES.KE.173         .................................................................................................................................  6042
RCM.NG.NGM           TACAAGAAGAAGCTAAGAAACATTTTACACCTGAATTGCTTACACAAGAAGGAAATTATATATATGAGCAGCATGGAGATAGTTTAGAGGGAGTAAAAGCTATGATAACCTTGCTTAATAAAGCACTTTT  6445
MANDRIL.14CG  Vpr    I__Q__E__E__A__L__K__H__F__D__R__R__L__L__H__A__V__G__S__W__I__Y__E__T__Y__G__D__T__L__E__G__V__Q__K__L__I__T__L__L__Q__R__A__L__F   Vpr
MANDRIL.GA.14CG                                                                            _M__K__H__M__G__T__P__W__R__G__C__R__S__L__S__P__S__Y__R__E__L__Y__   Tat
                                                                                           \ Tat exon 1 star
```

```
H1B.FR.HXB2   Tat __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__L__E__P__W__K__H__P__G_   Tat
H1B.FR.HXB2       .........................................................................................................ACTAGAGCCCTGGAAGCATCCAGGA   5875

CPZ.US.CPZUS      .........................................................................................................CA---------------G------   5946
CPZ.GA.CPZGAB     .........................................................................................................C--G-----------A--C-----C   5938
CPZ.CM.CAM3       .........................................................................................................T-----------CTA---------   5285
CPZ.CM.CAM5       .........................................................................................................T-----------CTA---------   5573
CPZ.CD.CPZANT     .........................................................................................................-AC-CCT--A---TTA--C--TCCT   5307

H1A.UG.U455       .........................................................................................................C--------------A--C--G---   5320
H1C.ET.ETH2220    .........................................................................................................C-----A--------C---------   5266
H1D.ZR.84ZR085    .........................................................................................................CA----T--------C---------   5396
H1F.BE.VI850      .........................................................................................................CT----T--------T---------   5214
H1G.SE.SE6165     .........................................................................................................C--------------T-----G--G   5271
H1H.CF.90CF056    .........................................................................................................---------------C---------   5221
H1J.SE.SE92809    .........................................................................................................CAG------------T---------   5188
H1K.CM.MP535C     .........................................................................................................CA-------------C--A------   5070
H101_AE.TH.CM240  .........................................................................................................C--------------T-----G---   5442
H1N.CM.YBF30      .........................................................................................................-T-------------T---------   5465
H1O.CM.MVP5180    .........................................................................................................GA-GCCC--T---C-T--C--T--G   5916
H1O.BE.ANT70      .........................................................................................................GG-GCCC--T---C-C--C--T---   5931

MAC.US.SMM251     CATGCATTTTAGAGGCGGATGCAACCACTCCAGAATCGGCCAACCTGGGGGAGGAAATCCTCTCTCAACTATACCGCCCTCTTGAGGCGTGCTATAA................................   6432
MAC.US.SMM239     CATGCATTTCAGAGGCGGATGCATCCACTCCAGAATCGGCCAACCTGGGGGAGGAAATCCTCTCTCAGCTATACCGCCCTCTAGAAGCATGCTATAA................................   6456
MAC.US.SMM239 Tat S__C__I__S__E__A__D__A__S__T__P__E__S__A__N__L__G__E__E__I__L__S__Q__L__Y__R__P__L__E__A__C__Y__.__.__.__.__.__.__.__.__.__.__.   Tat
MAC.US.SMM239 Vpr __M__H__F__R__G__G__C__I__H__S__R__I__G__Q__P__G__G__G__N__P__L__S__A__I__P__P__S__R__S__M__L__$_                                   Vpr
                                                                                                 Vpr end /

STM.US.STM        CATGCACTTCAGAGGCGGTTGCCGCCACTCCAGGATTGGCCAACCAGGAGGAGGAAATCCTTTGGCAACTATACCGCCCACTAGAGGAGTGTTGTAA................................   6104
SMM.US.SMM9       CATCCATTTCAGAGGTGGATGCAGACACTCCAGAATCGGCCAGTCTGGAGGARGAAATCCTCTCTCAACTATACCGCCCKGTAGAGGCGTGTTATAA................................   5930

H2A.GW.ALI        CATGCACTTCAGAGCAGGATGTGGCCGTTCAAGAGTTGGCCAAACAAGGGGAAGAAATCCTCTCTCAGCTATACCGACCCCTAGAAACATGCAATAA................................   6546
H2A.SN.ST         CTTGCACTTCAGAGCAGGATGCGGCCGCTCAAGGATTGGTCAGCCCAGGGGACGAAATCCTTTATCAGCTATACCAACCCCTAGAGGCATGCGATAA................................   5998
H2A.DE.BEN        CGTGCACTTCAGAGCGGGATGTAACCGCTCAAGAATTGGCTAAACAAGGAGAAGAACTCCTTGCCCAGCTGCACCGACCCCTAGAGCCATGCACTAA................................   6556
H2B.GH.D205       CCTCCACTTCAGAGCCGGTTGTCAACACTCAAGGATTGGACAATCAGGGGGAGGAAATCCTCTCTCAACTATACCGCCCCCTTAAGGCATGCGATAA................................   6529
H2B.CI.EHO        CCTCCACTTCAGAGGGGGTTGCCAACACTCAAGGATTGGACAACCGGGGGGAGGAAATCCTCTCTCAGCTATACCGCCCTCTTAAGGCATGCAGTAA................................   6527
H2G.CI.ABT96      CCTGCACTTTCGAGGAGACTGTAGATGCTCCAGGATTGGAGGGAACGCAGGCGGAAATCCTCTATCAACTATACCGCCCTCTAGAGGCGTGTTCTAA................................   5872

TANTALUS.UG.TAN1  ............................................................................................................----AG-CTCCTC--GGACTTG   6096
GRIVET.ET.GRI677  ..................................................................................................................................   5869
VERVET.KE.VER9063 ...........................................................................................................................A-A-TAT   6112
VERVET.KE.VERTYO  ...........................................................................................................................AGACTAT   5602
VERVET.KE.VER155  ...........................................................................................................................A-A-TAC   6107
VERVET.DE.VERAGM3 ...........................................................................................................................A-A-CTC   5610
SABAEUS.SN.SAB1C  ..................................................................................................................................   6333

SUN.GA.SIVSUN     ..................................................................................................................................   6259
LHOEST.KE.LHOEST  ..................................................................................................................................   6203
LHOEST.CD.447     ..................................................................................................................................   5122
LHOEST.CD.485     ..................................................................................................................................   5122
LHOEST.CD.524     ..................................................................................................................................   5119
MANDRIL.GA.MNDGB1 ...................................................................................................................--AACT-CTAG-AGAGTAT   5560

COLOBUS.CM.CGU1   ..................................................................................................................................   5516
SYKES.KE.173      ........................................................................................................................AAGGAACATTCCT-   6056
RCM.NG.NGM        CTTGCACTTCAGGCATGGATGTGAAGGGAGCAGGATTGGAGGAGCGAGAGGAGGAAATCCTCTTAGATCCATTCTTCACTCGAGAAACATCTTGTAA................................   6542
MANDRIL.GA.14CG   TCTGCATTTTAGACATGGATGCAGGGAAAGCCGAATTGGACAAGAAGGAGGCAAATATAACCCCCTTAGATCCTTTCCAAGGCCGAACAACCCCTTGTAA.............................   6066
MANDRIL.14CG  Tat F__C__I__L__D__M__D__A__G__K__A__E__L__D__K__K__E__A__N__I__T__P__L__D__P__F__Q__G__R__T__T__P__C__N__.__.__.__.__.__.__.__.__.__.   Tat
MANDRIL.14CG  Vpr __L__H__F__R__H__G__C__R__E__S__R__I__G__Q__E__G__G__K__Y__N__P__L__R__S__F__P__R__P__N__N__P__L__$_                                Vpr
                                                                                                    Vpr end /
```

HIV-1/HIV-2/SIV complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
                                                                                                                                                / Rev exon 1 start
H1B.FR.HXB2   Rev                                                                                                                            _M__A__G__R__S__G__D__.__.__S__D_  Rev ex1
H1B.FR.HXB2   Tat _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__G__I__S__Y__G__R__K__K__R__R__.__.__Q__R__  Tat ex1
H1B.FR.HXB2       AGTCAGCCTAAAACTGCTTGT...ACCAATTGCTATTGTAAAAAGTGTTGCTTTCATTGCCAAGTTTGTTTCATAACAAAAGCCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGA......CAGCGAC   5996

CPZ.US.CPZUS      --C--------G--------C...-AT--------C--------------------C--TGTG--------T-CG-AG--G-G------A--T----------------------G-GCGA...-GCA---   6070
CPZ.GA.CPZGAB     --------A-G---A-T-----...-AT-----------C---GCT-----T-A---C--TAT-TA---C----C--A-----G------A-----------------------A-CCACA...AGA----   6062
CPZ.CM.CAM3       --C-------G---------C...-AT--C-------------T---C------------AT---------TGAG-A-----G------A-----------------------A--G......GGA----   5406
CPZ.CM.CAM5       --------------------C...-AT--C--T---------TGT--C------------ATG-----C--TGAG-A-----G------A-----A-----------------A--G......GGA---A   5694
CPZ.CD.CPZANT     GC-ACA--AGC---GC-A--C...-AT--C--T--C------TGC-----T-AC-------CTC-C--C----C--A-----G---G--G--T----------------GAA-AGCTCGA...AG-AACA   5431

H1A.UG.U455       ----------C----------...-G---C--T---------GT--------GG---------C----C--TC-G-A-----G---------T----------A-----------A-.......-CC----   5441
H1C.ET.ETH2220    -----------G---------...-ATC-A--T-----------A---A---A------T-T------C--TC-G--.----G---------T---------------------C--.............   5386
H1D.ZR.84ZR085    ----------GG---------...-A------T-----------------A---------------C--------G----G-----------------------------------.............   5517
H1F.BE.VI850      --C-------C----C-----...-----A--T---------CGA------------------TGG--C--T-C---G--G-G-----------------------------------.............   5335
H1G.SE.SE6165     ---------------C-C---...-A---A--T-T-------GT--------GG--------------C--TC-G-AC----G-------------------------------AG.......--C----   5392
H1H.CF.90CF056    ---------C-----------...-A------T--------------C----A----------A-G--C--TT---AG----G------A--T-----------------------C.............   5342
H1J.SE.SE92809    -----------------------...--T-----T------------------A------T-----G--C--TC--CAG--G-G---------T-------------------------.............   5309
H1K.CM.MP535C     ----------------------...-A-C-G------------CG--------A----------A-A--C--TT---AG--G-G---------T---A--------------A-------.......-CC----   5191
H101_AE.TH.CM240  ----------C----------...-G--G--T--C--------A-------GG---------C-A--C--TC-G-A-----G-------------------------------AG.......--C----   5563
H1N.CM.YBF30      --C--A--------A-----C...-AT----------------GA-------A---C---TT-TA---C----C--AG----G------------A--------------------T........-A----   5586
H1O.CM.MVP5180    --CA----CC----CC-----...-AT-----------C----GA--C----A-------T-T----------C--AG--G-GT--G--A------C----------------A---.......AGA-C-G   6037
H1O.BE.ANT70      --------CC-G-TCC-----...-AT-----------C----GA--C----A-------T-T---------G---G---G-GT--G--A----------------------AG--.......-GA-C-G   6052

MAC.US.SMM251     ..........................--CA----------------------AC--------GT-------TC-T-A---G-GA--G--G--A-GT----AGCA-TCACGAAA---......AGAA--A   6530
MAC.US.SMM239     ..........................--CA----------------------AC--------GT-------TC-T-A-----G---G--G--A-GT----AGCAATCACGAAA---......AGAA--A   6554
MAC.US.SMM239 Tat _.__.__.__.__.__.__.__.__N__T__C__Y__C__K__K__C__C__Y__H__C__Q__F__C__F__L__K__K__G__L__G__I__G__Y__E__Q__S__R__K__R__.__.__R__R__  Tat ex1
MAC.US.SMM239 Rev                                                                                                                            _M__S__N__H__E__R__D__.__.__E__E_  Rev ex1
                                                                                                                                             \ Rev exon 1 start

STM.US.STM        ............................G----T---C-----A-------A---C-----GC-G-----TG-T-------GTC-T--A--AA------AGC-ATCA-GAA-A---......GTTAAGA   6202
SMM.US.SMM9       ............................G--Y--C-----G--A-------A------T--GCA------TC-Y-A---G-G---G--A--A-G-----AGCA-C-ACGAA-A---......ACT-CGA   6028

H2A.GW.ALI        ..........................--CA-----------GG-A--C----AC--------GC-G-----TT---AT--G-GAC-C--G--A-GG----A-C-A---GGCA--C--......AGAA--T   6644
H2A.SN.ST         ..........................---A--T--C-----------C----AC--------GA-G-----TT---AC--G-GGC-C--G--A-GG----AAC-A---GGCA-A---......AGAA--A   6096
H2A.DE.BEN        ..........................---A-----------GCGA--CA-T--C--------GC-G------TCG-A---G-GGC-C--A--A--A----AGC-A---GGCA-AC--......AGAA-GA   6654
H2B.GH.D205       .........................T-CA-----C-----G--A--C----AC--------GC-------TC-T-A---G-GTC-T--G--A-GT----A-C-CTC--GAA----------.TCTGC-A   6627
H2B.CI.EHO        .........................T-CA-----C--------A----C--AC--------GC----C--TC-T-A---G-G-C-G--G--A-GT----AAC-CTC--GAAA---------.TCTTC-A   6625
H2G.CI.ABT96      .........................T--A--T-----C-----A--C----A------T--GT-A--C--TC-T-AC--G-GAC-C--AG-A-GT----AGCAACCT-GAA-A---...AGAACT-CGA   5973

TANTALUS.UG.TAN1  CA-AG----TT---AC-A---ACT-A---A----T---C----GA--C----A---C--T--------C--TC-GCAG--G-G----------A-----CATGTCTCT-G-ATC---......AGA-CTA   6220
GRIVET.ET.GRI677  .....---ATT-CAAC-A---ACT-AT--A--------C-----A--C--T-A---C--TG-GC----C---C-GCAG--G-G------AG-GCGT---CATGTCTCT-G-AAA---......AGAAA-A   5988
VERVET.KE.VER9063 -AGA----GCT---AA-----AGA-AT--G----T---C-------------A---C--T--GC-------TC-GCA-----G-C----TG-TA-----CATGCCCCT-G-ACC---......AGAAAGA   6236
VERVET.KE.VERTYO  CAGA-----CTGCAGA-----AAA-AT--A--T-T---C-----A-----T-AC--C--T--GC----C--TC-GCA-----G-C----TG-TA-----CATGCCCCT-G-ACC---......AGAAAGA   5726
VERVET.KE.VER155  --GA-A--CCTTGAGA-A---ACA-AT--A----T---C-----A--C--T-A----------T-C--C---T--CGG----GAC----T--TA-----CATGCCTTT-G-ACC---......AGAAAGA   6231
VERVET.DE.VERAGM3 -AAGCA--CCTG-AGCGG---ACA-A---G------------TGT--C--T-A---C--T--GC----C--TT--CA---G-GA-----TG-TA-----CATGCCCCT-G-ATC---......AGAAAGA   5734
SABAEUS.SN.SAB1C  ......................T-AT-C-----TC--C---GTT--C-----------TATTC-C--C---CAT-A---G---C-------TCG----TATGTCCCT-G--CA---......AGAGCTT   6434

SUN.GA.SIVSUN     ........CTTGGAG--C---GAT-A---A----GG----GG-GA--C--------C--T---C----C--TC-GCAG----GG-----G---CA----TATGTCT-C-G-G-AC--.....----AT-   6375
LHOEST.KE.LHOEST  .....AG-CCTGCAA--G---CAG-AT-CC--T-GC--------A--------------TATTC----C--TCAGCAG----GG-----GC-GCA----TATGTCC-C-G-AAAC-G......AGA---G   6322
LHOEST.CD.447     .....---ACTG-G---C---AAC-A---G----GC----------------------TATTC----C--TCAGCAG--G-GA-----G--ACGA---TATGTCC-C-G-A----------A-AA--G   5241
LHOEST.CD.485     .....---ACTG-G--TC---GAC-A---G----GC--C----G--------------TATTC----C--TCAGCAG--G-GA-----G--ACGA---TATGTCC-C-G-A----------AGAA--G   5241
LHOEST.CD.524     .....T--ACTCCAAT-C--CCAG-AT--G---C-C--C---GTT-----------C--TAT-C----C---CAGCAG--G--G-----G--ACGG---TATGTCC-C-G-ACC---......AG-A--G   5238
MANDRIL.GA.MNDGB1 TA-------TTGCAA------GAG-AT--A----GG--C--G--A--C-----------TATGC----C--TCA--AG--G-GT-----A--AAGG--CCATGTCT-C-G-AAAC-T......GTA-C-G   5684

COLOBUS.CM.CGU1   .............AG--G---AATGA---A----GG-----GGCC-----------C--T--GC-A--C--TC-TT------GGC-G--A---CG----.....-------AAT-AG.........A-TG   5618
SYKES.KE.173      GAGA-AGGACC-C-AA-ACCATGT-AT--A----T---C-----C-------A---------GC--------C-TCA---G-GA--G--A---A-T----C---ACCA-G-AA-C--......GCAGCT-   6180
RCM.NG.NGM        .........................T-GC--T-----C----GA---GTG-A---C--T--GC----C--TC-TCAG--G--T--------AAA-----CTTCT-G--GAA-A--------.AGAG-CA   6640
MANDRIL.GA.14CG   ...........................---A--T-----C--G--A-------A---C-----GTG---C---T-GCAG-----T-----G--ACAT---CATGTCT-C-GAACC--G......AGAACT-   6164
MANDRIL.14CG  Tat _.__.__.__.__.__.__.__.__.__K__C__Y__C__K__K__C__C__Y__H__C__Q__C__C__F__L__Q__K__A__L__G__I__H__Y__H__V__Y__R__T__R__.__.__R__T__  Tat ex1
MANDRIL.14CG  Rev                                                                                                                            _M__S__T__E__P__G__.__.__E__L_  Rev ex1
                                                                                                                                             \ Rev exon 1 start
```

```
                                                Tat, Rev  exon 1 \/ intron
H1B.FR.HXB2   Tat R__R__A__H__Q__.__.__N__S__Q__T__H__Q__A__S__L__S__K__Q                                  / Vpu start (ACG start codon in HXB2)
H1B.FR.HXB2   Rev _E__E__L__I__R__.__.__T__V__R__L__I__K__L__L__Y__Q__S__                                  _T__Q__P__I__P__.__I__V__A__I__V__A     Vpu
H1B.FR.HXB2       GAAGAGCTCATCAG......AACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAAGTA.......................GTACATGTAACGCAACCT..ATACCA.ATAGTAGCAATAGTAGC    6093

CPZ.US.CPZUS      ---C-A--GC-A--AGC......---GC-GA-A------A-C-----CGC-----------GCCCTTTTG.......................-CT-T--TTAA-..TGGTTT.GA-A---GTT--A----    6170
CPZ.GA.CPZGAB     -C-CG----C-GCAGGC......---A-A-A-A------A----A--C--------------------------------------T-T-ACT-TG..T--GTC.GGCT---TTC-CA--TT    6151
CPZ.CM.CAM3       --C-CTTGTC-GCTGCA...-G----AC-CA-A------G-------CAGC-----------....................ACTTC-----CT-T--T-A---.TGGGA-.CA-A---GCT--A-T-G    5509
CPZ.CM.CAM5       --CACT-G-T-GCTTCA...-G----AC-CA-A------A----A--CG-C-----------....................ATTT-G----CT-T--T-AT-..TGGGAG.CA-A---GTT--A----    5797
CPZ.CD.CPZANT     --C--A-AAC-GCTGAA...-G----G-A-A-A------A-C--G--------T--------...........................-TCT---T-ACTAA-..---TTT.................    5511

H1A.UG.U455       --G--C---C---A......GG----A--GA-------A--CT-A--C--------------A.....................TTA-CTT------T-AC-----.T-GGA-.--CTGG------AC--G    5542
H1C.ET.ETH2220    ---------C---A......-G----A--GA-------AA-CT-A-----------------CCAAATAATAGATGTAATG....GT-GAT-TACTA----AAGTAGATTAT.-G-A---T---------    5505
H1D.ZR.84ZR085    -----C---C---C......-G-----------------A-C--A--C--------------..................................T----T--..T---A-.---T------------    5614
H1F.BE.VI850      AC---A---C---A......-G------AGTG------AA-------C--------------GTGTTA..............AATA---T------T-TC-TA-..T-GTT-.GC-A---GG----C---    5442
H1G.SE.SE6165     --G--A---C----......-G----A-AGGC-------A-C--G--C---------G----GAAAAA..............ATTA---T------T---GT-A..T--GT-.---T------------G    5499
H1H.CF.90CF056    AC---A---C-GCA......-GTTTG--AGA-------AA----A----------------.......................T--TCATA--T-T-TATA..T--GG-.T---G-AT-GG--CGCT    5439
H1J.SE.SE92809    --TC-----C--CT......GG----A------------A-CT-A--C--------------AATCAA...............GTA-C-T------T-AT-----.T-G-A-.----C------------    5415
H1K.CM.MP535C     ---C-A---C-T-C.....------G---AC-------A-C-----CG--------G----GTGCTA..............AGTA----------T-GTGT-C..T--G--.--TTC-AT-G---C-TT    5298
H101_AE.TH.CM240  --G--A---C----......-G----A--GA-------AA-C--A--C--------------.....................ATAA---T------T-AC-----..T-GGA-.--TAGT----------G    5664
H1N.CM.YBF30      -----A---C----......-G----A---G--------A-CT-A--C--G-----------.....................AAACCT---T--...............--G-TG.TC-T-G-G-T-CA----    5677
H1O.CM.MVP5180    C-GCT---GCAAGC........TA--CAGA-A--A---A-C--G--C--G-----------.......................ACGC-G-T---T-AA..GAGAAC.C-GC----CT--A----    6128
H1O.BE.ANT70      CTGCT--AAGC--T............-CAGA----A---A-C--G--C--------------.......................ACGC---T---T-A...-GGGAC.C-GC--------A--AT    6140

MAC.US.SMM251     CTCCGAAAA-G...........GC-A--G--A--AC-T-----GC------CA-------.......................................................................    6579
                                                Tat, Rev  exon 1 \/ intron
MAC.US.SMM239     CTCCGAAAA-G...........GC-A--G--A--AC-T-----GC------CA-------.......................................................................    6603
MAC.US.SMM239 Tat T__P__K__K__.__.__.__.__A__K__A__N__T__S__S__A__S__N__K                                                                      Tat ex1
MAC.US.SMM239 Rev _L__R__K__R__.__.__.__.__L__R__L__I__H__L__L__H__Q__T__                                                                      Rev ex1

STM.US.STM        A--AG---A-G..................---T---C-AT----GC------CA-------.......................................................................    6245
SMM.US.SMM9       AG-AGA--A-G.................G--A--AC-TT----GC----G-CA-------.......................................................................    6071

H2A.GW.ALI        CTCC-AAGA-A............-T-A--G-----TCGT-----GC----G-CA---G---.......................................................................    6693
H2A.SN.ST         CTCCGAAGA-A............-C-A--G-----TCGT-----GC----G-CA---G---.......................................................................    6145
H2A.DE.BEN        CTCC-AGGA-A............-C-A------C-TCGC-----GC-C--G-CA---G---.......................................................................    6703
H2B.GH.D205       A-------A-G....................---AC-GC-C-----GC-C--G-CA---G---.......................................................................    6670
H2B.CI.EHO        A---G---A-G....................---AC-AC-T-----GC-C----CG---G---.......................................................................    6668
H2G.CI.ABT96      A--AGA--A-G.................G--A--TC-TT----GC------CG-------.......................................................................    6016

TANTALUS.UG.TAN1  AG-AGAA----................TCA-A------GAA-CT-G-----C----------.......................................................................    6266
GRIVET.ET.GRI677  C---CA----A................GA-A------ACC--A--CGGC----------.......................................................................    6031
VERVET.KE.VER9063 A---T-T---G................C--A---GCCTA----A-CAGG-C--------.......................................................................    6279
VERVET.KE.VERTYO  AG-TTCG-TCG................CT-A--TTG---C--T--CAGC-C--------.......................................................................    5769
VERVET.KE.VER155  AG-TT---TCG................G--G---GCAT-C--G--C-GC----------.......................................................................    6274
VERVET.DE.VERAGM3 AG-TT----CG................CT-G---GCTT-C--GA-CA---A----G---.......................................................................    5777
SABAEUS.SN.SAB1C  CT-AGAAG...............ATTT-G---A-TCAAGT-TCTCT-C-CA-C--GTA.......................................................................    6477

SUN.GA.SIVSUN     A-TC-ATATC-..................-AGG--A-GCAAGAGGCTGT-CGAAG------.......................................................................    6418
LHOEST.KE.LHOEST  TT-CCT-GATA-TT..............GAGAC---GCAGGA--CTGTGG-CAG-----G.......................................................................    6368
LHOEST.CD.447     TCGCCCAGGTATTT..............GAAG--A-GCAAGA--CTGTGG-CGG-----A.......................................................................    5287
LHOEST.CD.485     TT-CCCAGGTATTT..............GAAG--A-GCAAGA--CTGTGG-CGG-----G.......................................................................    5287
LHOEST.CD.524     TT-CCTACATA-TT..............GAAAC-G-GCAGGA--CTGTGG-CAG----AA.......................................................................    5284
MANDRIL.GA.MNDGB1 ---CTAA-A-GA--.........-TA-CTGG-AG-GGT-AAGAAGCTAT-CGAAG------.......................................................................    5736
COLOBUS.CM.CGU1   CT-C-CAA.............................................-CA--------.......................................................................    5637
SYKES.KE.173      -C--TAT-....................T--G-GG-T-A-AGCGC-C-T-CAGG-------.......................................................................    6221
RCM.NG.NGM        --CC--GAA--...AAGAAG--G-A-----T-A---C-TT-GA-AC---TG-----------.....................................TGG-A-GT-CTGGTCT-T    6717
MANDRIL.GA.14CG   -G-AGAGA-T--T-...GGA--A-TATCACAGG--AGT-AAG-AGCTGTGGGAAG------.......................................................................    6222
MANDRIL.14CG  Tat R__K__R__L__L__.__G__K__I__S__Q__D__S__E__A__A__V__G__R                                                                      Tat ex1
MANDRIL.14CG  Rev _G__R__D__F__W__.__E__K__Y__H__R__I__V__K__Q__L__W__E__                                                                      Rev ex1
                                                Tat, Rev  exon 1 /\ intron
```

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2   Vpu __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__I__D__R__L__I__E_    Vpu
H1B.FR.HXB2       ATTAGTAGTAGCAATAATAATAGCA...ATAGTTGTGTGGTCCATAGTAATCATAGAA......TAT.........AGGAAAATATTAAGACAAAGAAAAATAGACAGGTTAATTGATAGACTAATAGAA    6205

CPZ.US.CPZUS      -----G-A---A-GGT--TT---T-...G-GA-AA-C---GGAC----TGCT-G-TTG......-GGAGACAAATA-AA-T--A-GA--ATACTCA-C--................--G-TT-A--ATCTT    6276
CPZ.GA.CPZGAB     -G---GCT--...--TGC-TGGAAC...--TTG-A-A---GGATACA-T--T-A-TGG...GGA---AGAAGGTAT-AA-G-CATAGGCTTG-G-C-G--............-----G-----T-ACCT-    6260
CPZ.CM.CAM3       TA---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-T-GGAC-TT-A--GA-AG-    5624
CPZ.CM.CAM5       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-TTACA--TT-GC-GA-AG-    5912
CPZ.CD.CPZANT     .GAGTAT-CTTTTC-TGCCT-TAGT...-----AC-A---ATT--CTGT--ACCTAT-CTCTATA-GCTTTATAAA-TAT-T-AGCAGCAG----T-G-T-ATA-G--AAATCAAAGA-TCA--GA--TG    5637

H1A.UG.U455       GC-GA--------C----CT-----..-----A------A-T------GGT------......--AAA.......-AA---T-GC---AG----A---------------T-AA-C---A---G----    5656
H1C.ET.ETH2220    ---CA--------C----C------..------------A------C-TAT------......---.........------T-G------------G------T---------A-A---ACT-GG---    5617
H1D.ZR.84ZR085    -------T----GC-----------..------------A--------T--------......---.........C-T-G---AA---G---------------T--------------A---G----    5726
H1F.BE.VI850      ----A--------C----------C..------------A-T------TAT------......---.........-A----C-GG----G------------A-T-AA---TA-A-A---A---G----    5554
H1G.SE.SE6165     ----A--------T-C---GC---C..------------A--------T-T------......---.........---G-----AG--A---G-A--G-----GA-A----C--------A---G----    5611
H1H.CF.90CF056    -G-----AC-TTT--C---GCC......G--A--------A--------TAT------......---.........-AA---T-GG----G----AG-----------C-------A---A--GG----    5548
H1J.SE.SE92809    ---TA----------TT-CC-T---..----G-A----A-T----GTAT------......---.........-A----C-G----------G-------T-A------------A---G----    5527
H1K.CM.MP535C     -G------C-CT----T--GC-.........---A----A---A--------TAT------......---.........------T-GG---A-----A--G----A--T-------A------A---G----    5407
H101_AE.TH.CM240  -C-GA-------GC----CT-----..-----A-----AA-T------GCT------......GT-.........-A-------C----G---------------------G--A-G---A---G----    5776
H1N.CM.YBF30      G----G--C----G-T-GC-------..G--A-A--C---G-AT--C--TAT-G----......---.........-A-------AA-TTG--GGA-------A-ACACA---GAC-G---A---G----    5789
H1O.CM.MVP5180    T---AGT-CTTTGTGTC-T---AAT...G--C--A-A----TGT-TAACC-T-G-ATT......---.........TTAGTGCA-AG--A----GAT-G--GG--GCA-GA---ACT-GA-AG-T--AGG    6240
H1O.BE.ANT70      TA-TAGT-CTTTGC-GT-T---AAT...G--A--T-A---GGGT-TA-TC-T-G-A--......---.........TTAG--CA-AAGGA----GAC-G--AG--G--AGA---ACT-GAGAGGT--AG-    6252

MAC.US.SMM251     ...................................................................................................................................    6579
MAC.US.SMM239     ...................................................................................................................................    6603
MAC.US.SMM239     __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._    5924

STM.US.STM        ...................................................................................................................................    6245
SMM.US.SMM9       ...................................................................................................................................    6071

H2A.GW.ALI        ...................................................................................................................................    6693
H2A.SN.ST         ...................................................................................................................................    6145
H2A.DE.BEN        ...................................................................................................................................    6703
H2B.GH.D205       ...................................................................................................................................    6670
H2B.CI.EHO        ...................................................................................................................................    6668
H2G.CI.ABT96      ...................................................................................................................................    6016

TANTALUS.UG.TAN1  ...................................................................................................................................    6266
GRIVET.ET.GRI677  ...................................................................................................................................    6031
VERVET.KE.VER9063 ...................................................................................................................................    6279
VERVET.KE.VERTYO  ...................................................................................................................................    5769
VERVET.KE.VER155  ...................................................................................................................................    6274
VERVET.DE.VERAGM3 ...................................................................................................................................    5777
SABAEUS.SN.SAB1C  ...................................................................................................................................    6477

SUN.GA.SIVSUN     ...................................................................................................................................    6418
LHOEST.KE.LHOEST  ...................................................................................................................................    6368
LHOEST.CD.447     ...................................................................................................................................    5287
LHOEST.CD.485     ...................................................................................................................................    5287
LHOEST.CD.524     ...................................................................................................................................    5284
MANDRIL.GA.MNDGB1 ...................................................................................................................................    5736

COLOBUS.CM.CGU1   ...................................................................................................................................    5637
SYKES.KE.173      ...................................................................................................................................    6221
RCM.NG.NGM        -CCT-CTT-TAG-TGCT--GCTTATAAGG-GA-A......GGAGATA----T-G----GTTTTTAT-.........T-T-G-G-T-AT---A--TAG--GG--A-TCTT-AG--................    6816
MANDRIL.GA.14CG   ...................................................................................................................................    6222
MANDRIL.GA.14CG   __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__._    5542
```

```
                                                   / Env gp160 start in HIV-1
H1B.FR.HXB2   Env                                  _M__R__V__.__.__K__E__K__Y__Q__H__L__W__.__.__.__.__.__.__.__R__W__G__W__R__W__G__T__M__L__   gp120
H1B.FR.HXB2   Vpu _.__.__.__.__.__.__.__R__A__E__D__S__G__N__E__S__.__.__E__G__E__I__S__A__L__V__.__.__.__.__.__.__.__E__M__G__V__E__M__G__H__H__A__P    Vpu
H1B.FR.HXB2       ....................AGAGCAGAAGACAGTGGCAATGAGAGTG......AAGGAGAAATATCAGCACTTGTGG.....................AGATGGGGGTGGAGATGGGGCACCATGCTCC   6288

CPZ.US.CPZUS      TTAGAAAGA............-TCAG-AT-AGAGAA-AAG--AGTG-AA......-T-A-AGTGATGG--A-GAA-AA-AGACTCTGGCTAAGT............TAT--TCTTCTCTCG-G-T--A-AA   6377
CPZ.GA.CPZGAB     ATATTGAGA............GA-AG--C---AGA-A-TGGAA............-T-A-AGT-ATGG--A-GAA-AA-AGAGACTGGAACAGCTTATCC-T-ATTACAATC-T-ACAAT--TTT----AA   6367
CPZ.CM.CAM3       ATAATATCA............---CAG-----T-----A-----A----------T------GATCA-AGG-AGC-T......................-T-GA-AAA-AC-TG-TTATGGGTT--A-AA   5716
CPZ.CM.CAM5       ATAAGAATA............---CAG-----T-----A-----A--CA......-T-----GGA-GA---T-AAT---.....................-T-ATTT-C-AC-TGCTAATGGGTT--A-AA   6004
CPZ.CD.CPZANT     TTAAGTAGA............---CTTAGTATAGA-A-TGCAATAGAA-AAG...-T-AG---GCCGATA--TA-TATT......TGGGGT.........CTGGCTTT-CTA-TCCA-TTT-TAGA-AAGG   5738

H1A.UG.U455       ....................-----------------------------.....-T--G--T-C-GAG-A-T-ATCCT.....................T-C...TT------------A--T---A--T   5736
H1C.ET.ETH2220    ....................----------------------A----------T--G--T-C-GAG-A-T-GTCAA.....................CA-...T------T------A-T-T-AGG-T   5697
H1D.ZR.84ZR085    .....................----A------------------------.....----G--T-A-GAG-A---ATCAA.....................CCC...TT-----A--------T--------   5806
H1F.BE.VI850      ....................----------------------------.....-G--G--TGC-GAG-A-T-G-CA-......................CAC...TT-G---A-------CTTT-AT---   5634
H1G.SE.SE6165     ....................----------------A----------.....-C--G--T-C-GAG-A-T-G--AA.....................CAC...TT-----A------G--TT--A---   5691
H1H.CF.90CF056    ....................--------------------C------------.....-T-----C-C-GAG-A-T-ATCCA.....................-C...TTA---------------T-AA--T   5628
H1J.SE.SE92809    ....................----------------T--C------------.....-T-----C-C-GA--A-T-G-CA-......................-CC...TT-------G-----CT----A--T   5607
H1K.CM.MP535C     .....................AG--C---AGACA-TGGC......A......-T-AGAGTGAGGGG-ATGCA-A--.......................-AT---CA-ACATTGG--AA-TGGGGCA--T   5481
H101_AE.TH.CM240  ....................----------------A----------.....------C-C-GAT-A-T-G-CCA.....................-AC...TT-----A------G--TT--A---   5856
H1N.CM.YBF30      .....................----A------T-----------A----------.....-T--G--TGC-GAGTGGT-G-AT-.....................G-G...AT-AA---TG-TT-GTTAC-CT--T   5869
H1O.CM.MVP5180    AGAATAAAGGAAATCAGGGATGACAGT--CT-TGAAA-T-----AGAA-.......--CA-C-GGA-GTCATGGAGC-TA.....................TAC...ATAGCC-TG......G-TT----AA   6335
H1O.BE.ANT70      AGAATAAGAGAAATTAGGGATGATAGT--CT-TGAAA-----GAGAA-.......-G-A-C--GA-GTTATGGACC-T-.....................TAC...TTAGCC-TG......G-TT--A-AA   6347

MAC.US.SMM251     .....................................................-T--GATGTCT-GG-A-TCA-CT-.....................CTTATC-CCATCTTGCTTTTA-GTG-CTATG   6634
MAC.US.SMM239     .....................................................-T--GATGTCT-GG-A-TCA-CT-.....................CTTATC-CCATCTTGCTTTTA-GTG-CTATG   6658
MAC.US.SMM239                                                          _M__G__C__L__G__N__Q__L__.__.__.__.__.__.__.__L__I__A__I__L__L__L__S__V__Y__   gp120
                                                                       \ Env gp160 start

STM.US.STM        .....................................................-T--CCTGCCC-GGAA-TCAACT-.....................CTTATC-CTATCTTGCTATTA-GTGCTTG-T   6300
SMM.US.SMM9       .....................................................-T--GATGTCT-GG-A-TCA-CT-.....................CTTATC-C-CTCTCGCTATYA-GTGCTTCAG   6126

H2A.GW.ALI        .....................................................-T-AT-TCTAG-AGAA-TCA-CT-.....................CTTGTTACTATCTT-CTA-CT-GTGCTTG-T   6748
H2A.SN.ST         ...................................................AAG-T-...TGTGG-AG-A-TCAACTA.....................TTTGTT-CCA-CTTGCTA-CT-GTGCTTG-T   6200
H2A.DE.BEN        .....................................................-T---CCTGG-AG-A-TCA-CT-.....................TTTGTT-TCATTTT-CTAACA-GTGCTTG-T   6758
H2B.GH.D205       .....................................................-T--CAT-T-T-AGCAG-CGCCT-...................CCT-TTGC-CTCCT-CTTATA--T...--CAGTG   6725
H2B.CI.EHO        .....................................................-T--CAC-TGT-A-TA-T-ACCTA.....................CTTGTTACACTCCTGCTTATA-GT--CTATG   6723
H2G.CI.ABT96      .....................................................-T--CAT-TCT-GGCA-TCA-CT-.....................CTTATC-CAACCTTRCTCTTA-GT-CTTATG   6071

TANTALUS.UG.TAN1  .....................................................-T--GACC--TAAG-GGAAAAG--.....................GT--TATTAGTA-TT-T---ATTA-GCT-AA   6321
GRIVET.ET.GRI677  .....................................................-T--G--G--TG-TTATAAAAATA.....................CT-ATAATAGCA-T-....-G-TA-GTA-A.   6082
VERVET.KE.VER9063 .....................................................-T-A-ACCTCT-TTAGGTA-AGTT.....................CTTATA-CTATC-T-....-A-TA-GTTC-.   6330
VERVET.KE.VERTYO  .....................................................-T-AG-T-TACAATAATAACC-TA.....................G--ATAATAGT--T-....-A-TAGG-A-A.   5820
VERVET.KE.VER155  .....................................................-T-ACA--G-TCTTAGGAA-T-TT.....................-T-GTATTAG-A-T-....-G-TAGGAA-A.   6325
VERVET.DE.VERAGM3 .....................................................-T-A--CTGACATTA-TGA-AG--.....................-T-CTATTAATAG-GATA--AGTAG----TA   5832
SABAEUS.SN.SAB1C  .....................................................-GTAT---GCT--TTACAG-ACT-.....................CTT---CTAA-TG-G--TT-G-G-T-AG-GT   6532

SUN.GA.SIVSUN     ...................................................AAG-T-...CGG-GC-CTG-AA-A-T-...................ATAG-T-TCTCT-T-CTG-TA--T-GT...A-AG   6473
LHOEST.KE.LHOEST  ...................................................AGG-T-...GC--G--CAGGA--AG-A.....................-TTCT-CTT-T-CTTCTT----TA--A....   6419
LHOEST.CD.447     .....................................................-T-A-CTGCCC-GGA-TAA-ACTT...............CTCTTG-TCATTTTAA-AGT-CAA--G..........   5338
LHOEST.CD.485     .....................................................-T---CTGTCT-A---T-C-ACTT...............TTCTTGCTC-TTTTAG-AGC-CAA--G..........   5338
LHOEST.CD.524     .....................................................-T--CATGTCCAGGTA--A-A-T-.....................CT--T-CT-CTCT-C-T---ACTA-CAGCAG   5339
MANDRIL.GA.MNDGB1 .....................................................-T-A--TGTCCAGGT-TT--TCTT...............ATACTAT-T-TA-CC.TACT-GTA--T-TT--AGGA.   5795

COLOBUS.CM.CGU1   .....................................................-T-CTT-G-----TCTTTA-AAAC.........................TT-CTAGT--TA--T-TAG--T-A.   5685
SYKES.KE.173      ...................................................AAA-T--CAGCT-T-AGAACT-AC.....................ATAGTT--TCTT-TT--C-TAATA-GT...T-AG   6276
RCM.NG.NGM        .....................................................-T--CT--G.........--AATA.....................--TATA--AATT-T-ATA--ATTGT-AGGTA   6862
MANDRIL.GA.14CG   .....................................................-TAATAGT-AG-AT--TTAGACAT.........TTAATCACAGCA-T-GTAATAATAG-TATA--T-TA...GGA.   6285
MANDRIL.14CG  Env                                                                  _M__L__R__H__.__.__.__L__I__T__A__I__V__I__I__G__I__G__I__.__G__   gp120
                                                                                   \ Env gp160 start
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
                                 Vpu end \
H1B.FR.HXB2   Vpu __W__D__V__D__D__.__.__L__$_
H1B.FR.HXB2   Env 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__C__A__S__D__A__K__A  gp120
H1B.FR.HXB2       TTGGGATGTTGATG......ATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACC.........ACCACTCTATTTTGTGCATCAGATGCTAAAGC   6403

CPZ.US.CPZUS      -CCCAGGA---........................TCT-GC--A----CT-----A-----------------A----GA--T-T-GAG...........--A--CT----C-----C--T-----A--G--   6474
CPZ.GA.CPZGAB     CCCCATGT----CC.....................TCTG-G--A-----A-----A--------A--------T---C-T--T--TGA-.........C-GGTA--C--------C-----C-----G--   6467
CPZ.CM.CAM3       -CCCATGT----CTGGTAGT................--TC-G-------CA-----G--C--------------------A--G---GA-...........--A--A--C--------C-----G--GTC---   5822
CPZ.CM.CAM5       CCCCATGT----CTGGC..................C-TG-G-G-----CA-----G--C-----A--------A-----A--G---GAA...........--A--A--T--C-----C-----G--CTC---   6107
CPZ.CD.CPZANT     GGAC-..............................A-TG--GAC-AT--A-----A-TC-----A--C-----C----GAA-T--G--A.........C-T--------------CA--A----CTCCAT   5829

H1A.UG.U455       -G---T--A-A--A.........AT-T--AA-G--C--C-------------G-----C--------------------A--T---GTT..............-----C--------------------------   5848
H1C.ET.ETH2220    --T-------A---......--T----A--GA-TG-GG--C--------------------------------A-----A--T--T-G-.........C-T----------------------------   5812
H1D.ZR.84ZR085    --------------...ATG-C--A-------G----T--C--------------T-----------------A-----A--------.........-----G-----------------------T-   5924
H1F.BE.VI850      -G-----A--A--A.....-------A----G----C--C-------------------------------------A---------..........--T--------C--------------------   5749
H1G.SE.SE6165     -----T--G----A.......--T--------CT--A-T--T--------T---------------------------G-A--T---GAT.......----------------------C---------T-   5806
H1H.CF.90CF056    -G--T------C--.....-------------G--C----C--------T-----A----------------------A--G----AA..........-----------C----------------G--   5743
H1J.SE.SE92809    ----A-----A---......--T----AG--A-A----G-C-------------------------------------A--T----AG..........------T-------C-----T-----------   5722
H1K.CM.MP535C     -ATTCT--GGA--ATTGATA--------CAA-G----C--------------------------------------A--------A.........C----------C------------------   5602
H101_AE.TH.CM240  -----T--G----A.......--T--------CT----C--C--------T-----T-----------G----------GA--T---GAT.........-----C--------------------C--G--   5971
H1N.CM.YBF30      A-CTTC-AG-A-GCTTGATC-AGGTA-T--GGT-T---C--CAT-----A-----G--C-----------A--A----GA------GAG...........--A-----T--C-----T--------------   5990
H1O.CM.MVP5180    -C......CCATGT.........-TG---TA--GTA--C------AT-C------T----C-----------A---G-A--G---G-A.........C-AGTA-----C-----T-----------CCT   6441
H1O.BE.ANT70      CC......CCATGT............TTGAGCCTTAG-C-GC-A-AT-CA---------GC------G-----A---G-A--T-----A.........C-AGTA-----C-----T-----------CCT   6455

MAC.US.SMM251     GGATCTAT...............----C-............CAA-AT----------T------T-----A-CT----G-A-T--G--A..........-TTC-C--C--C------.............   6714
MAC.US.SMM239     GGATCTAT...............----C-............C-A-AT----------T------T-----A-CT----G-A-T--G--A..........-TTC-C--C--------.............   6738
MAC.US.SMM239 Env G__I__Y__.__.__.__.__.__C__T__.__.__.__.__L__Y__V__T__V__F__Y__G__V__P__A__W__R__N__A__T__.__.__.__I__P__L__F__C__A__.__.__.__.__.  gp120

STM.US.STM        -GACCTAT...............--C-C-............CA--AT-----G--A-T------T-----A-CA----G-A-T--G--G..........-TTC-C--C--C------.............   6380
SMM.US.SMM9       GGATTTAT...............---GT-............CAA-AT--A-----A-TC-----T-----A-CA----G-A-T--G--A.........GTTC-C--C--C------.............   6206

H2A.GW.ALI        -A-TATAT...............----AA............CAA-AT--G--T--T-T------C--G--A-CA-----AA-T---T--..........-TTC-C--C---------.............   6828
H2A.SN.ST         -AATATAT...............--CGTC............CAA-AT--G--T--T-TC-----C--G--C-------GAA-T---T--..........-TTC-C--C---------.............   6280
H2A.DE.BEN        -A-TATAT...............-----C............CA--AT--G--T--T-TC-----CA----C-C------AA-T---T-T..........-TTC-CT----------.............   6838
H2B.GH.D205       GGTTTG-A...............----AA............CAA-AT--T--T----TC-----CA----C-CA----G-A-C-----A.........GTTC-C--CA-------.............   6805
H2B.CI.EHO        GGTAT---...............G-C-AG............AAC-TT-----T----TC-----TA----C-CA-----AA-T---T-A..........-TTC-C--C--------T............   6803
H2G.CI.ABT96      GGTTCTAT...............----CC............AAA-AT-----T--T-T------CA----C-CA----G-A-C--G-G-..........GTAC----C--C------.............   6151

TANTALUS.UG.TAN1  -A--AC--......................TTATAT-GG-C-CA--ATA-T-----G-T------TA-C--A--A-----AA-CAGCT-A.........GTGCAAGCT--C---ATG..............   6407
GRIVET.ET.GRI677  ..--A--A......................-G--AC......C---AT--G-----G-T------AA-C--A--A-----AA-TT----A.........GTTCAGGC------CATG..............   6160
VERVET.KE.VER9063 ..TTAT-A......................AG-GA-AG-...CAA-----------G-TC------A-------A------A-CAGCT-A.........GTACAAGCC--C---ATG..............   6411
VERVET.KE.VERTYO  ..-T-T-A......................AG--AG......CAA---A-A-------T------AA----A--A-----AA-CAGCT-G.........GTGCAGGCT--C--CATG..............   5898
VERVET.KE.VER155  ..------A.....................AG----A--C-GCA----A-A-----G-TC-----A-----A--A-----AA-CAGCT-A.........GT-CAAGCT-----CATG..............   6409
VERVET.DE.VERAGM3 A-ACA-G-......................CAA.........CAA-----------A-T------A-----A--A-----AA-CAGCT-A.........GTACAGGCT--C--CATG..............   5909
SABAEUS.SN.SAB1C  GGTTAG-C......................CA--AT--T-----G-T------TA-C--A--A-----AA-TAGTT-A.........GT-CAAGCC------AAG..............   6606

SUN.GA.SIVSUN     CAATA...........................CA--AT--A-----A-TC------ACT---AA----G-ACC----GTA.........GTGC-AT--A-A-----T..............   6544
LHOEST.KE.LHOEST  ..T--.................G-A-AA...CAA-AT--A-----A-T---------G--AAAT---G-T--TAATGTTTCA......GTGC-CT-GA-------T..............   6497
LHOEST.CD.447     ......................GAG-AAA-A-AGA-G...CA--AT--A-----G-TC--------G---AAA---G-A------GTA.........GTGC-AT--A-------T..............   5419
LHOEST.CD.485     ......................GAGGCAAAG-AG......CA--AT--A-----A-T---------G---AA----G-A--T--TGTA.........GTGC-AT--A-------T..............   5416
LHOEST.CD.524     GA.....................GAA-AG-AG-A-......CAA-AT--A-----A-TC-----A-----AAAA---G-A--T--TGTA.........GTAC--T--A-C-----T..............   5419
MANDRIL.GA.MNDGB1 ........................-AA...........CAA-AT--G-------TC-----A-----A--A-----------T-AA..........--ACA-T-GA-------T..............   5864

COLOBUS.CM.CGU1   ............................AG--AT......AAA-----A-----A---C-A--A-------CA---G-A------GATGTAAATGATCAGCAGT-C--C--CTTT..............   5766
SYKES.KE.173      GATTT---...............GAA-AA......C--...CAA-AT--A-----A-T-------A----ACAT---G-A--T---TAT.........G-AC----C------A-C..............   6359
RCM.NG.NGM        -AATT--A.........GTAGGAGAG-AA...........CAA-AT--G-----A-T------T--C--A--A------A-CAGT--A.........GTTC--A-G-------T..............   6948
MANDRIL.GA.14CG   ........................-AA...........CAA-----A-----A--------CAC----AAA---C-CCC-----GA..........--GCA---T--------.............   6354
MANDRIL.14CG  Env .__.__.__.__.__.__.__.__.__K__.__.__.__.__Q__W__V__T__V__Y__Y__G__G__P__K__W__H__P__A__R__.__.__.__T__H__L__F__C__A__.__.__.__.__.  gp120
```

```
H1B.FR.HXB2   Env __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__N__D__.__.__.__._   gp120
H1B.FR.HXB2       ATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGGAAAAATGAC............   6521

CPZ.US.CPZUS      ---CA-GCAG----CC--C--CA-------------G--A-----C--T--T--------T------------CAT---CC---------T-----G---G--------G-----A-T............   6592
CPZ.GA.CPZGAB     -C--AG--------CT------A-------------G--A--------T-----T----GT--T--G------T-TC-TCC------A-T----TCA-----------------A-T............   6585
CPZ.CM.CAM3       TCT-A-C-A-----CC------A--------T----A--------T--T--------T--------------GCAGA-CCC------A--T-----C-----T--------C---AC-............   5940
CPZ.CM.CAM5       -CT-A-C-A---A-CC--C---A--------T----A--------C--TGT---T--T-----C-----G--GTATA--CC------AT-T----GC-----------------A-T............   6225
CPZ.CD.CPZANT     GACAAG---------G--C-----A-----A--TACCAGT-----G--A-T---T--AG-T--TATT...-----TAG-C-C---ACCT---TCTGG-----TGCT-AT------T-T............   5944

H1A.UG.U455       -------G----A--G--------C-----T--------------------------------------A---AT-----C------------G-A-----------------A--............   5966
H1C.ET.ETH2220    ------------A----------C----GT---TT---------C-------------G----------T-G-GT----A-------------------------------------............   5930
H1D.ZR.84ZR085    ----A-A-------C-------A-C-----T--------------------------------------A---A-C--AA------C--------------------------CA--............   6042
H1F.BE.VI850      ------A-G-----C------C-------------------------------------------------T-TC--AA------A------------G-T------------A--............   5867
H1G.SE.SE6165     ----AG-G-T--AAGC--------C-----T--------------------------------------A--A-TA---A------A------T-------------------A--............   5924
H1H.CF.90CF056    ------G-----AAAG--------C-----T--------A-----------------------------GA-G--CA---AG-----A-----G-GC-----T------G-----A--............   5861
H1J.SE.SE92809    ----AG---T--AAA---------C-----T--------T--------------T--A-G---------GA-GAATC--CC------A--------C----------------------............   5840
H1K.CM.MP535C     ------G-AG--------------C-----T--C-----------------G----------------------AGA---A------A--------C-----------------A--............   5720
H101_AE.TH.CM240  -C----G-----A--G--C-----C-------------------------------------------------A--CACC---A------A--------------------------A--............   6089
H1N.CM.YBF30      CC--AG--------CT--C--CA-C-----------A--A-----T--T--T--T-----T-----------GC----ACCC-----A--T-----A-----T------G-----A-A............   6108
H1O.CM.MVP5180    -ACAAGC--T--ACAG------A-------AT----A-----C--T--T-----T-----T-----T---T-TCC-C-A-GC-----------T--C---G-T--A---------T--............   6559
H1O.BE.ANT70      -ACAAGC--T--AAAG------A-------AT----A--------T--T----------CT---T-T---TATCC----CAC-----------TG-C-----T--A---------T--............   6568

MAC.US.SMM251     ........--CA--AATAGGG--AC-----GA---AC-CAG--CC----AGATA-TGGTG-TTATTC----...T-GGCCC-T-----T-------GC---G-TGCT---G-G..................   6816
MAC.US.SMM239     ........--CA--AATAGGG--AC-----GA---AC-CAG--CC----AGATA-TGGTG-TTATTC----...--GGCCC-T-----T-------GC---G-TGCC-----T..................   6840
MAC.US.SMM239 Env __.__.__T__K__N__R__D__G__W__G__T__T__Q__C__L__P__D__N__G__D__Y__S__E__.__V__A__L__N__V__T__E__S__F__D__A__W__N__.__.__.__.__.__._   gp120

STM.US.STM        ........--CA--AATAGGG--AC-----GA---AC-CAA---C----AGACA-TGGTG-TTACTC----...T-GGCAA-------C-----GGC----G-TGCC---G-C..................   6482
SMM.US.SMM9       ........--CAGAAATAGGG--AC-----GA--CACACAA--CT-G--AGATA-TGGTG-TTACTC----...T-GGCAA-T-----C-----GGC----G-TGCT---G-T..................   6308

H2A.GW.ALI        ........--CA-AAATAGAG--AC-----GA--CATACAG--CT----AGACA-TGATG-TTAT--G---...A--GCTT-G-----------GGC---CG-TGCA---G-T..................   6930
H2A.SN.ST         ........--TA-AAATAGAG--AC-----GA--CATACAG--CT-G--AGACA-TGATG--TAT--G---...A--GCTT-------------GGCC--CG--GCA-----T..................   6382
H2A.DE.BEN        ........--TA-AAATAGAG-CAC-----GG--CATACAG--CT-G--AGACA-TGATG-TTAT--G---...A--A-TT-------------GGC----G-TGCA-----T..................   6940
H2B.GH.D205       ........--CACAAACAGAG-CACC----GA--TGTACAG---C-C--AGACA-TGGTG--TACACT--G...A-CAG-C----CA-A-----GGC----G-TGCA---G-T..................   6907
H2B.CI.EHO        ........--CAGAAACAGAG--ACC----GA--TGTACAA--CC-C--AGATA-TGATG--TATACC---...A-CCAAT-----A-A-----GGC----G-TGCA---G-T..................   6905
H2G.CI.ABT96      ........--CA-AAATAGGG-CACA----G---TACACAG--CC-T--TGATA-TGATG-TTACAGT---...C--GCAT-G-----A------GC----G-TGCC---G-T..................   6253

TANTALUS.UG.TAN1  .........---CCTAATACC--CC-------A---AC-AAT---A----AGAT--T-ATG-TTATACT---...---CA-T-------CT-T--G--A---G-AGCA-----GG-CAGG............   6515
GRIVET.ET.GRI677  ........--GCCCAATACC---A-G-----A--CACCAA---CA----AGAT--T-ATG-TAATAC---G...--GCCTC----CA-T------GC---CG-GGCT---...G--A-T............   6265
VERVET.KE.VER9063 ........--TCCTACTACC-GGC-A-----T---AC-AAT--CA----AGAT--T-ATG-TTATAC---G...--GCCAT----CA-T--T--GCCC---G-AGCC---GC-G--AGG............   6519
VERVET.KE.VERTYO  ........--TCCCACCACA-GCT-A-----T--TAC-AAT--CA----AGAT--T-ATG--TATAC----...---CCTC-----A-C--T---CCA---G-GGCA---GGTG--AGA............   6006
VERVET.KE.VER155  .........---CCTACTAC--GGT-G-----A--TAC-AAT--CA----AGAT--T-ATG--TATAC----...---CCAC-G---A-A-----GCCA---G-AGCA---GC-G-CAGA............   6517
VERVET.DE.VERAGM3 .........---CCCACCACC-GAC-A-----A--TAC-AA--CGA----AGAT--T-ATG--TACAC---G...---CCAT----CA-C--T---CCA---G-AGCA---GCTG-CAGA............   6017
SABAEUS.SN.SAB1C  ........--CCCCAATACG--CC-A------T-TAC-AAT--CA-C---GAT--TGAACCAGA-GG-AC-...A--GCA-A-GTACCC-T-CCT---A-C-CAGAAAAATTTG---CTTGGAAAAATCGG   6726

SUN.GA.SIVSUN     .......T---CTAATA---GCT-G----TG--TAC-T-A--CT-G--AGATTTA-AA-CTTATGC----...--GCCCA---CA-G-TT----G-GAA-TTT-CAGA-GG-..................   6646
LHOEST.KE.LHOEST  .......T---CCAATAC--GCC-C----T---GACAT-T--CT-G--AGATTTG-AATCTTATGC---G...---CCTA-CT--AAT-TTTC-G-AAACTTT-CAATACC-..................   6599
LHOEST.CD.447     .......T---CTAATA---G-T-A----T----ACAT-A--CA----AGACTTG-AGTCTTATGC-C-G...A-TCCTA-TC-CAAC-T-AGTTTAAA-TT--CTAAAG-G..................   5521
LHOEST.CD.485     .......T---CTAATA---GCT-A----T----ACAT-A--CC----AGATTTA-AGTCTTATGC-C-G...A-TCCTA-TC-CAAC-T-AGTTTAAA-TT--CTCAAG-G..................   5518
LHOEST.CD.524     .......T---CTAATA---GCT-G----T---TACAT-T---C-C--AGATCTG-AATCTTATGCTC-G...A-CCCTA-TC--AAT-TTACTCTAAA-TTT-CTCAAG--..................   5521
MANDRIL.GA.MNDGB1 .............-----TAATTCA-G-C-C----TA--CAC-AAT--CA----TT--TTG--AG-TTATG-T--G...---GAAA-TCC--AT-T-A-GG-AAA-TTT-CAG-ACTT..................   5966

COLOBUS.CM.CGU1   .......T-TTCCTCC-CAG-GA--CAACAGGTGTTG-GG---T----AC-TCCG--TGGAAAG-C--T--A-CA-AACA-GCCAAATGT-AC-G-AGC-TTTGATCTTTTT..................   5871
SYKES.KE.173      ........--CTCCCATA-AGGA-GA-----T--TA-GAAT-----G--GT---CA......GAT---AT--A----AG---C--CA-T----G-G-A-ACTTTCCAGCATGG............AATAGC   6464
RCM.NG.NGM        ........--T--TAACGCAC-ATCA----GA--GTTAAAT---A----AGA--GAGG--TAT---C----...---CCA---------T----G--A---G-TGCT---G--................AAT   7053
MANDRIL.GA.14CG   ........--T--CAACA-CTCCC------TA--CACAAGT--------A-GCTTATTGC-TTATG-G--GCA-C-CA-CCCC---A-A-------------CAG-CCCT-T-................   6459
MANDRIL.14CG  Env __.__.__T__D__N__N__S__L__W__V__T__T__S__C__V__P__S__L__L__H__Y__E__E__Q__L__I__P__N__I__T__E__N__F__T__V__P__I__.__.__.__.__.__._   gp120
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2   Env _.__.__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__C__T__.__.__.__.__  gp120
H1B.FR.HXB2       ......ATGGTAGAACAGATGCATGAGGATATAATCAGT............TTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCACT.............  6620

CPZ.US.CPZUS      .....----C------A-----G--A-----T--T--C............C-------C--G---T----A--T---A----G---------T-G-------C-A-G-CT--TCT-AACCCC.......  6697
CPZ.GA.CPZGAB     .....-----G--C--A--------A--C--T-----.............C-C--------G--T--C-----T--------G--G-----TT-A-----A-CC--GC------G-AAGGCTAACTTTA  6697
CPZ.CM.CAM3       .....----C------A---A--ACT--C--T--T---.............C-G-----------TT-G--A--------T--GC-G--G---A-A-----C-CCA-GG-A--T-GAAAGGTTACTTTCA  6052
CPZ.CM.CAM5       .....----C------A-----G--A-----T-GT---.............C-G-----------T-----A--------T---C-G--A---A-A-----C-CCA-GC-A--T--GACCCTTAAGAACA  6337
CPZ.CD.CPZANT     .....----------AGT---ACA--A-----GN-ACAA............----TCC-A-------AT---------------C----A--TA-G---A-A-AAA-G--T--T--AGGATACAATGGAA  6056

H1A.UG.U455       .....--------T-----------A------------.............C-------C----------------------G--------T-----C----CG---G-T---CA-AACATCACCATTA  6078
H1C.ET.ETH2220    .....-----G--G---------C--------------.............------------G------------------G--G----------------C------C--T-A-GCTATCAAAAACA  6042
H1D.ZR.84ZR085    .....-----G--C--------------------T---.............--------------T-----A----------------------G------C-C------T------GATGCATCGAGAA  6154
H1F.BE.VI850      .....------------A------ACA------------.............--------------------------------G-----------------G-C------C--T---AATGCCACCAATA  5979
H1G.SE.SE6165     .....-----G-----A----------------------.............---------G----------------------GC-------T----------CC-----C--T---GATGTAACCAACA  6036
H1H.CF.90CF056    .....-----G--G---------ACA--C---------.............---------------T-G--A------------------------------C-C----C--T---AATGTCAGAAACA  5973
H1J.SE.SE92809    .....--------C--------G--A------------.............G--------G-----T----------------GA-------T--A-----C-C------T--T-G-AACATCACTAGTA  5952
H1K.CM.MP535C     .....--------G---------ACA------------.............---------G----------A--T------G-G--------T--------C-C------T--T---GATTACAAAGGAA  5832
H101_AE.TH.CM240  .....--------G--------G------G--------.............--------------T-----------------G-----T--T-----C----C------T--T--CAATGCTAATTTGA  6201
H1N.CM.YBF30      .....----C---C--A-----A--------T------.............C-G-----A--G---T-------C-----T------------T-A-----A-C-A-GCTT--T-ACGATAGCTATGGGG  6220
H1O.CM.MVP5180    .....-----G--C--A--------A--C--C--T---.............--G-----A--G--TT-------T----AG---A-G--TTTCT-A-----ACAAA-G--C--TGTAGATCTGCAAACAA  6671
H1O.BE.ANT70      .....-----------A-----G--A--C--T--T---.............--------C--G--T--T--A--T-----TC--A-G--TTTC--G-----ACAAA-GG----T--AAAC..........  6670

MAC.US.SMM251     AATACAG-CAC-------GCAATA-----CG--TGGCAA............C-C-TT--GACCTCAA-------T------------T-----T-A--CA---C-A-G-GA----A-AAAAGTGAGACAG  6934
MAC.US.SMM239     AATACAG-CAC-------GCAATA------G--TGGCAA............C-C-TT--GACCTCAA-------T------------T-----T-A--CA---C-A-G-GA----A-AAAAGTGAGACAG  6958
MAC.US.SMM239 Env _N__T__V__G__E__Q__A__I__E__D__V__W__Q__.__.__.__.__L__F__E__T__S__I__K__P__C__V__K__L__S__P__L__C__I__T__M__R__C__N__K__S__E__T__  gp120

STM.US.STM        AATACAG-CAC------AGCAATA------G-CTGG-AC............C-T-TT--AAC-TC-A-T--A--T-----------G--T-----A---A--GCAA-G-G---T-A-AAGAATGAAACAG  6600
SMM.US.SMM9       AATACAG-CAC------AGCAATA------G--TGG-AC............C-C-TT--AAC-TC-A-T-----C---------C-C--------A---A-A-CAA-G-G---T-A-AAAAGTGAGMCAG  6426

H2A.GW.ALI        AATACAG-AAC------AGCAGTA--A---G-CTGG--A............C---TT--GAC-TCAA----A--------C--G-----A--TT-A---A-AGCAA-G-----T-GCAAC...ATAAGCA  7045
H2A.SN.ST         AATACAG-AAC------AGCAGTA--A---G-CTGG---............C---TT--GAC-TCAA----A-----C--C---C----A--CT-A-----AGCAA-GCGT--T-ACAGC...ACAACTG  6497
H2A.DE.BEN        AATACAG--AC------AGCAGTA--A---G-CTGGCA-...........C---TT--GAC-TCAA----A--------C--GC----A--T--A-----GGCAA-G--T--T-GCAGG...GTTCAAG  7055
H2B.GH.D205       AATACAG--AC-C-----GCAGTA--T---G-GTGG--A............C-C-TT--AACCTC-A----A--------C---C----------G-----GGCAA-G--C--T-G-AAAACCGAAACAA  7025
H2B.CI.EHO        AATACAG--AC---T---GCAACAA-----G-GTGG---............C-C-TT--GACCTCAA----A--------T--------------G-----G-CAA-G-----T-A-AAGACGTGGAGCT  7023
H2G.CI.ABT96      AACACAG-AAC-------GCAGTA--A---G-TTGG-A-...........C-G-TT--GAC-TCTAC---A--T-----T--------A--T--G--CA--GCCA-G-GA----A-AAAACTGAGACAA  6371

TANTALUS.UG.TAN1  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--T---AAATTAACAAGTA  6633
GRIVET.ET.GRI677  ...CCGC-----A----AGCAG-GAGTA-----CATCTA...........C-C-TT--A----CGA-G-G---T-----T--GC-CT----CA-A---A---AAA-GTCC--TGTAGAGCTGAATGGTA  6380
VERVET.KE.VER9063 AATCCAT-A--G-C----GCAGGAAGTA----TCATTTG...........--G-TT--A----CAT-G--A-----C---------T-T--T--A---A-A-AAA-G-GC--TGTAGAGTTGAACTCAA  6637
VERVET.KE.VERTYO  AACCCAT-AA---C---AGCAGCCAGTA-C--CCATTTA...........C-C-TT--G----CTA-G--A--T-----G--G---T-A-----A--CA-C-AGA-G--T--TGTAGAGTTAAATTCCA  6124
VERVET.KE.VER155  AATCCCT-A----C---AGCAGGAAGTA-C--TCA-CTG...........C-G-TT--A--G-CAT-------C--------GC--T-A--T--A---A-C-AAA-G--T--TGTAGAGTTAAAAGGCT  6635
VERVET.DE.VERAGM3 AACCCCT-A----C---AGCAGGAAGTA-----CA-CTG...........C---TT--G--G-CT--G------------------T-A--TT-G--CA---AAA-GTCC--TGTAGAATTGAACTCCT  6135
SABAEUS.SN.SAB1C  AACCCTC-A----GC--AGCAG-GAGTA-C---CA-CT-...........--G-TT--AAGT-CTT-------------T--GC-CT-A---A-G--CA-A-AAA-G--C---TA-CGCTTAGAAGGAG  6844

SUN.GA.SIVSUN     ...ATA-GCAATA-T--A--AGTGC--C-AGC-TGGCAAGCTATGACCTCTA-GGTA---GCC-TTA-G-----T--------GA-C-AT--CTAT-----C-AAA-G-----TC-ACCGAAACCGACAA  6773
LHOEST.KE.LHOEST  ...GTA-AA-ACA-T---G-TATAC--C-AGCCTGGTC-GCTATGAATGCTA-GGTA--CAGT-TTA-G--A------------A-C-AT---TAT-----G--AA-GC----TGGGGAAGTTACAAAAA  6726
LHOEST.CD.447     ...ATA-AA-ATA-T--A--TATAC--C-AGC-TGG---GCTATGACCAGTA-GGT----ACT-T-A-G-----T---------G---AT---TA------G--GA-G--A--TGAAGGGGAAGGAATAG  5648
LHOEST.CD.485     ...ATA-AA-ATA-T--A--CATAC--C-AGC-TGG---GCTATGACCAGTA-GGT----ACT-T-A-G-----T--------GA---AT---TA------G--AA-G--A--T--AGGGGGAGGAATAG  5645
LHOEST.CD.524     ...ATT-A--ACA-T--A--AATAC--C-GGC-TGGG-AGCTATGACGAGCA-GGT----GCT-TTA-G---------AC----A---AT--TTAT------C-CA-G--A--T--AGGGGAAGGACTAG  5648
MANDRIL.GA.MNDGB1 ...ATA-G--A-A-T-----AGT-T-TC-AGC-TGGCA-GCTATGGGAAGTA-G-TA---ACC-TA--T--------------GA-T-A----TAT-------AGA-GC-A--TCAGGAAACAGAAAATG  6093

COLOBUS.CM.CGU1   .........AAGA-TAGCT-CAG--G---GG--TGG-TAATAACACAGACTACC-TA--A-----G----GA-------CT--G--G--AG--TA-----C-CCAA-G-TT-----CAAAGTAAATAGGA  5992
SYKES.KE.173      AGTCAC---A--AGG---CAAATATTA--GGAC--GTCAGCTTTA.........-TCTTG---GC-AAT-GA--C-----G--GC--G-A--CA-G---A----GA-GTT---T--ATTGGACAACAGCC  6585
RCM.NG.NGM        AGCTTGTAT-A-...--AGCAA-GTCTA--G-GTGGCA-CTT............-AT---TCC-CTT-G--A----------GG-----A---A-G---A-A-CAA-GT-T--T---GCATTGAATGGTA  7168
MANDRIL.GA.14CG   ......CAA-AGA-TG-AG--GTAAG-C-AGC-TGGG-AGCTATTTCCTCTA-GATA---GC-GTT-----A--G-----------G--A---TA------C-AGA-GC-A--T--AAAGGGAGAAAACA  6583
MANDRIL.14CG  Env _.__.__Q__E__N__E__V__V__R__Q__A__W__G__A__I__S__S__M__I__D__G__V__L__K__P__C__V__K__L__T__P__Y__C__V__K__M__Q__C__T__K__G__E__N__  gp120
```

```
H1B.FR.HXB2   Env .__D__L__K__N__D__T__N__T__N__S__S__S__G__R__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  gp120
H1B.FR.HXB2       ..GATTTGAAGAATGATACTAATACCAATAGTAGTAGCGGGAGA..........................................................................................  6662

CPZ.US.CPZUS      ................................GA---TAAT..............................................................AGTAGTGCTGTAAATAC  6723
CPZ.GA.CPZGAB     GC.....................CAGGCA-AA-ACCTAACA.............................................................AACCAGACATCTTCTCC  6734
CPZ.CM.CAM3       ATAGCACTTCT---AGG-A---A--TTC--C--TG-CTACT.......................................................................AACAGTCC  6101
CPZ.CM.CAM5       TTAGCACT-CT---AC--AC...-GT-CAGAGTTA-AAAAT...................................................................................  6375
CPZ.CD.CPZANT     CACC-ACA-CACCA.........-GT-CA-CA-CA--TACA.............................................................GTAACACCAAAGACAAC  6105

H1A.UG.U455       ACA-CACC--T--CA-C--C---.......................................................................................................  6101
H1C.ET.ETH2220    ATAC-AA-GTT-CAA---A--G-.......................................................................................................  6065
H1D.ZR.84ZR085    ACAGCACTG-T---A-C-GC-CA.......................................................................................................  6177
H1F.BE.VI850      ACAG-CA-G-A--GCCAGGGGCA.......................................................................................................  6002
H1G.SE.SE6165     AA-GCAAC--A-GAA---A--GC--TG---A---C-CT-AA-CCAACAATAGCACTGTGGATAACCCA..........................................................  6104
H1H.CF.90CF056    ATACC-CT--C-GCAC--GC-G-.......................................................................................................  5996
H1J.SE.SE92809    ACAG-AAT-CA-C-AG--AC-G--GT....................................................................................................  5978
H1K.CM.MP535C     CCA--AGC-CC---A--G-A-C--G--C-.................................................................................................  5861
H101_AE.TH.CM240  CCA--GGC-GT-GCA-A--C---GT-TC--AC..............................................................................................  6233
H1N.CM.YBF30      AG--AAG---C---ACA-A-......................................................................................................ATGAC  6245
H1O.CM.MVP5180    ATA-AACAGGC...................................................................................................................  6682
H1O.BE.ANT70      ..............................................................................................................................  6670

MAC.US.SMM251     ATAGA-G-GGATTGACA-AATCAT-A-CA-CA-TA-CAACAGC-GCACCAACATCAGCACCAGTA.....................................................TCAGAAAAAATAGA  7013
MAC.US.SMM239     ATAGA-G-GGATTGACA-AATCA......-TA-CA-CAACAGC-TCAACAACATCAACGACAGCA.....................................................TCAGCAAAAGTAGA  7031
MAC.SMM239    Env D__R__W__G__L__T__K__S__.__.__I__T__T__T__A__S__T__T__S__T__T__A__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__S__A__K__V__D  gp120

STM.US.STM        ACA-A-G-GGATTGACAGGG--A--AGTA-CA-CAGTGACACC-ACAGCAGCAGCAGCAGCAACA.........................................................AAGCCAGA  6673
SMM.US.SMM9       MCAGA-G-GGTTTAACAGGAGSAC-AGCACCA-CA-CAACACA-ACATCAACAACACCACCATCACCAATAATAGCA.............................................AAG.....  6506

H2A.GW.ALI        CA--GAGC-CA-CCACAT-CCCG-G-CCAG-G--C-CACTC.................................................................................AAACCC..  7092
H2A.SN.ST         CAA-AAAC-CA-CCTCC--ACCA...-CA-CC-CC-CAACAGC-..............................................................................AACACAAC  6546
H2A.DE.BEN        GGA--ACC-C--CCCCG-A-CCC-GG-CCTCG---TC-ACA-CCTCGAGACCACCCACATCC............................................................GCAGCCTC  7125
H2B.GH.D205       ACCCAGG---TGCCAG--G--C-----CC-C--AGCCTACT-CCACCTCT.....................................................................CGTGGGCTGAA  7086
H2B.CI.EHO        CA-C-AGC--AG-GACC----CGT--TCCGCGTCCCT-A-ATCTTCTACTCAG...........................................................................AC  7078
H2G.CI.ABT96      ACA-A-G-GGTCTAACAGGG-CC----C-GTA-CCCCT-CA-C-GTTCCCACAACTAAAATGGTGACAGCT.........................................................GA  6444

TANTALUS.UG.TAN1  CC-CCCCT-CATCAAG---CCC---ATCA--C--C-CTAC-.................................GATCCCTGTCCAAATACCGACGAAAGTAGCTGTAACGCCACC..............  6716
GRIVET.ET.GRI677  CA-CCAC--CA--G-CC--C-C---TGCA-C--CA-CAAT--CT..............................ACCCCCTGT..............................CAGAATTGCAGTACAGAGCA  6453
VERVET.KE.VER9063 GTAGAAGA--CCCAAC-C-A-C-T-AGC-TCA-CA-CAAAAGCT...CCAAAAACAGGT...............GATCCCTGCATTAAATCTACCAATAATAATGTTAATTTACAGCCCTGTAATGCCTC  6749
VERVET.KE.VERTYO  CAAGAGAA-G-GCGACA--ACC---A-CG-CGCCG-AATCT-CCGGC...........................CTACCCTGTGTAGGCCCGACGTCAGGTGAA...AATCTACAGTCCTGTAATGCAAG  6224
VERVET.KE.VER155  CC-CAACCTCT-CCCCAG-A-CCT-T-C--CGGCAG-AACC-A-..............................CTACCCTGTGTTAGAAATAAAACAGACTCC...AACCTACAGTCATGCAACGACAC  6732
VERVET.DE.VERAGM3 CT--GCCT-CC-CCAC-C----A-GT-CC-CGGCCTCAACA-CCAATATCACAGCCTCAACAACCACT......TTGCCGTGTGTCCAGAACAAGACAAGTACTGTG...TTAGAATCATGTAATGAAAC  6256
SABAEUS.SN.SAB1C  GA-CAGCA-CA-CAACAT-ACCAT-A-CATCA-CAGCAC-CCCCGAGGTAGTGAGT..............................................................GTGGGCTTTAATGACTC  6917

SUN.GA.SIVSUN     CTCCC-CCTCC--CTCA--AGTC-AA-G---CT--GATTATT-GACGACCACCACGGCTAAGACTACAACTCAGACTACAAGTAGTACC.....................................AGCTC  6867
LHOEST.KE.LHOEST  CACCAACA-CA-CACCA-AA.......................................ACTACTACACAGATGCCTTGTTTTATCAACGAACAGGTA..................ACAGTTAAGAA  6796
LHOEST.CD.447     GAACAACA--AGCAACA-AA-CA-AA-CG..............................TCAACAACACCCATGCCTTGCTTTATAACAGAAAAAACCACAACCACCACCAAAAGATCTACAACTGG  5745
LHOEST.CD.485     GACCAAAA-CA-CACCAT-ACCA-TAGGA-AA...........................ACAACAACACCCATGCCTTGCTTTATAGCAGAAACAACCACAACCACCACCAAAGGGAATACAACTAT  5745
LHOEST.CD.524     GTA-GAA--GA-CAACA--A-CA--A-CACCA...........................ACAACAACACCCATGCCTTGCTTTATAACAGAATCTACTACTACTACTACAAAGAGTTCAGGGACTGA  5748
MANDRIL.GA.MNDGB1 TATCAGCA-CA-CA-C--AGCC--TA-C--CACC--CTACT-C-TCTACAGTTGCAAGTAGTACAGAGATTTACTTAGATGTAGATAAAAAT........................................  6185

COLOBUS.CM.CGU1   CT--AAATGGC-C-AG---AGTAG--CCA-CA-CA-CAAAT-ATAGTGCTTCAGATTGGGATGAGTCA..................................................................  6060
SYKES.KE.173      CT-C-ACA-GT-CCCC-----CA-GT.........................................................................................CCCCCTACTAC  6622
RCM.NG.NGM        GTTGGGATGGT-C--CC--CC-AT-ACC--CA-CA-CAACA---GCAACAACAAGAACAACAAAAGAAATAGGGAAAGATTGTTATGCGAATCAAACAGAGTGCAGTAGTGTCCCT..............  7284
MANDRIL.GA.14CG   GTAG-ACC-GC...........-T-CC-C-CCC-TGCCC-CTACCACCACTCCCACCACTCCA........................GCAGTGACGGGGCTGGAGGTGACGGTACAG...........  6666
MANDRIL.14CG  Env S__S__G__S__.__.__.__.__T__T__T__P__M__P__T__T__T__T__P__T__T__P__.__.__.__.__.__.__.__.__.__A__V__T__G__L__E__V__T__V__Q__.__.__.__.  gp120
```

HIV-1/HIV-2/SIV complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2   Env __.__.__M__.__I__M__E__K__G__E__I__K__N__.__C__S__.__.__.__F__N__I__S__T__S__I__R__G__K__V__Q__K__E__Y__A__F__F__Y__K__L__D__I__I_  gp120
H1B.FR.HXB2       .......ATG...ATAATGGAGAAAGGAGAGATAAAAAAC...TGCTCT.........TTCAATATCAGCACAAGCATAAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGATATAATA  6770

CPZ.US.CPZUS      TACTGAT--A.......................G-G---------T-------------T-----A-CT--TGAAT------AC--AAA-A-AC---TG---T-C--A------GT-GA----C--GCT  6820
CPZ.GA.CPZGAB     GCCTCTCGAA.......................G--------------T--A..........T---G-G-CA---GAAT------A----AA-A-AC-G-TC---T--C-A------GTGGAG---G-GG--  6831
CPZ.CM.CAM3       TAAT.........GA--AAAYAG-CA-CACAG----G--T....--TA-C...........T-----G-CT--TGAGG------A---A-A-A-A----T----T-CC--------GT-GA----C--G--  6210
CPZ.CM.CAM5       ...................ATA----A-ATA--G-----T....-TA-A..........T-----G-CT--TGAAG------A---AAA-A--G---T----T-TC-A------GT-GA-------G--  6474
CPZ.CD.CPZANT     AACACCA--AGTAGATGGCAT---G......C--C--G-A....-TAAC..........T---CAG------G-AT-T-A--A----AAA--A---ATGA-A--CA-A---------GGA---C-T--G  6217

H1A.UG.U455       ....ATC-CA...GATGGA-T--GG-A---A--G-----T...-----------------------G-C----GAAC------A----AAA-------T----T--C---------G----------G--  6212
H1C.ET.ETH2220    ........--C....-AT-GT-CC--T-AT--A--G-----T....-T-----------------A-C----GAAC------A----AAAAG---G-C-------C-------------------G--  6173
H1D.ZR.84ZR085    .......C--CCA-CGG--A--CC------A--G-----------------------------A-C----GTAG---CG-A----AG-A--C---T-C-----C---------G-------G--G--  6288
H1F.BE.VI850      ..............................GC---------T-------------------G-C----GAAG------A----AA--T----TT--C---GC---------GG---------G--  6089
H1G.SE.SE6165     ......................-----A----------------------------------G-A-C----GAA-------A----AAAA----------C--G-----C----G-------G--G--  6197
H1H.CF.90CF056    .............--GGA--CAGG---G--AC---C---T-----------------------G-A-CT---GTAC------A----CA--------T-C-----C-C-------G-------G--G--  6101
H1J.SE.SE92809    .............G---GTAGTCCT-ATATC--G-C-----------------------------A-CT---GAA------AA---AAG-A--C----G--C--GC----C----G--AA---G--G--  6083
H1K.CM.MP535C     .............G-GG-AAGCCC--C---A--------T-----------------------A-C----GAA----A--A----AA-A-A--G----C----C----C----G-------G--C-G  5966
H101_AE.TH.CM240  ........--AATAGG--ATATA-C--AT--AG---G---------TA-----------T-----G-C----GAAC---C--A----AA------G-TCC-----C-C--------G---------G--  6344
H1N.CM.YBF30      AACAAGAGAACCAGAC--A-GAT-CAA-C-A--G-----T...-----A..................GCA-C---TGAGC---C--A---AAA-A--C---TT--CT-TC-G------GT-GAA---G--G--  6363
H1O.CM.MVP5180    .......C-ATTA-ATGA-ACA-T-AAT-----G-G---T....-TAG-----------T---G-A-CT---GT-C-C-C--AC--AAA-G--C--A--C-G--TC-A--C---GT-TCA---C-G-GT  6793
H1O.BE.ANT70      ....ATAGCTGGA-C--CAA-TG--AACCTT--G--G--G....-TGAG..........T---G-A-C---TGTT--C-A--AC--AAA-G-----A--C-G--TC-A--C---GT-TCA---T-G--G  6784

MAC.US.SMM251     CATGGTC-ATGAG-CT-GTTCTTGTAT--CTCAG--T--T.....--A-A.........GG-TTGGAACAAGAGCAA--G-T-A-CTGTAAATTC-CCATGACA-GG--AAAAAGAG-CAAGACA-AGGAG  7131
MAC.US.SMM239     CATGGTC-ATGAG-CT-GTTCTTGTAT--CCCAGG-T--T.....-A-A.........GG-TTGGAACAAGAGCAA--G-T-A-CTGTAAATTC--CATGACA-GG--AAAAAGAG-CAAGA-A-A-GAG  7149
MAC.US.SMM239 Env __M__V__N__E__G__S__S__C__I__G__Q__D__N__.__C__T__.__.__.__G__L__E__Q__E__Q__M__I__S__C__K__F__N__M__T__G__L__K__R__D__K__K__K__E_  gp120

STM.US.STM        GTTAGTA-ATGAA-C--GCTCCTGT-T-AGT-AC--C---.....--A-A.........GG-TTGGAAGAAGA-TCAC--GT----TGTAAATTC--TATGACA-G-C-AAAAAGAG-TAAAA-G-G-GAG  6791
SMM.US.SMM9       .GTTGTA-ATGAC-GTGATCCTTGTAT-AGA-GT--T--T....-TA-A.........GG-TTGGAACAGGAGCC---GGT-A--TGTAAATTT--CATGACA-GG--AAAAAGAG-CAAAA-G-G-GA-  6623

H2A.GW.ALI        .CTGATA-ATGAG-GCGATCCATGCAT-A--GC-G-C---.....---C-C......AGGGGACTAGGGGATGA-GAG--GGTCAA-TGTCG-TTC--CATGACA-G---ACAGAGAG-TAAGCCA-A-CAG  7212
H2A.SN.ST         AATAGGA...GAG-ATTCTTCATGCAT-CGC-C-G-C---.....---A-A.........GGGTTGGGAGAAGA-GAG--GGTC-ACTGTCA-TTC--TATGACA-G---AGAGAGGG-TAAGA-A-A-C--  6661
H2A.DE.BEN        CATAATA-ATGAA-CTTCTA-CTGCAT---A-AC--C-CA.....-G-A.........GGATTAGGGTATGAGGAG--G-TGCAATGT-A-TTC--TATGA-G-GG--AGAAC-AG-TAAGA-A-GG-GG  7243
H2B.GH.D205       AACGATT-ACGAA-C-GACCCATGCAT-A-A-ATG-C-G-....--A-A.........GGACTAGGAGAAGAGGAA----TGCAATGTAATTTT-GTATGACG-G-C-AAGAAGAG-TGAGCTA-A-CA-  7204
H2B.CI.EHO        CCTACTC-ACGAAGAT-GCA-ATGTATTC-A-ATG-C-G-....-TG-A.........GGG-TAGGACTAGA-GAA--G-T--ACTGTCAATTC---ATGACA-G---AAAAAGAG-TGAGTCA-A-CA-  7196
H2G.CI.ABT96      ACTAGTG-ATAGC-CT-GCC--TGCTT-AT-TACG-T--T....-TA-A.........GGT-TACAGTCAGA-TC---GGTT--CTGCAAATTT--CATGACA-GGC-AAAAAGGG-T-AGAGA-AGGA-  6562

TANTALUS.UG.TAN1  .TTAGTT-CAAAT-GC-----TT-T-AGA-T-GTTCT-TA.........-------C...........TGC---GGCAGG-TAT-GG----A-GTAAAAA-----T-T...AATAG-AC--GGT-TGA-C-GGAGT-G  6830
GRIVET.ET.GRI677  G...ATAGAAGGAGA-----CAG-G-A-CCAGC-TCC---.....---A--..........TGCA--TGCAGG-TATCA-----A-GTAAAAA----TT-T...AGCA-GACC-GGT-TGA-C-GGAGT--  6565
VERVET.KE.VER9063 CCTTATAGAAGAGGA-C-----G---A--CAGCTTC---T....-TA-A..........TGC---GGCAGGGTA---T----A-C--AA-A-A--TT-T...T--G-AG-A-GG--TGA--CAGAG--C  6864
VERVET.KE.VERTYO  CATTATAGAAAGGGAG------G-T-AGCCCGCCTCT--T....-TA-A...........---GCA--GGCTGGCTATG------A-C--AA-A-A--TT-T---T-TG-GG-G-GG--TGA--CAGA---C  6342
VERVET.KE.VER155  CATCATAGAAAAGGAG---A-TG-C-AG-CAGCGTC---.....---A-C..........TGC---GGCTGGGTA---T--G-ACC-AAA-A----TT-C...T--G-AG-A-GG--TGA--CAGA---C  6847
VERVET.DE.VERAGM3 AATCATAGAAAAGGA-T-AA-TG---AGCCTGCTTCT--T....-TA-A..........TGCA--GGCAGGGTATG------A-C--AAAA----GT-T...T--G-GG-G-GG--TGA--CAGA---C  6371
SABAEUS.SN.SAB1C  AGTGATAGAACAAGAG-----A----A-C--GCC-TG---.....--T--C..........TGC---GGCAGG-TAT-G---G-A-GTAAA-A-A--TT-C...T--ACAG-A-GGG-TGA-C-AGAGG-G  7032

SUN.GA.SIVSUN     CACTGCT-GTACA-CT-CTCCA-TGCCTCTAGATTGG--T....-TA--GACACA...GAG--C--AGCGGAG----AT-A--TATGTAAAT-T--T-TGACA-GGC---G-AGAG-TTG-A-A-CGGAG  6991
LHOEST.KE.LHOEST  CCCGGGG-ATGAA-C-...AGATT--AG--AGATCTT--T....--A-AAGAGGGCTTAATG-G-C--CAGAG--G-ATGC--AATGCCA-T-T--T-T-ACA-GG--A-GCAGAG-CTGCAGG-CGGA-  6920
LHOEST.CD.447     GGGACCAGAA...........TT-AT----GAC-T---T....-TAT-AAATGGGTAAATG-G-CT-CTGAG-TT-AT----AGTGTAGAT-T--T-TTACA-G-C-A-G-AGGG-CTG-AGG-C-GA-  5860
LHOEST.CD.485     GGAAAGT..............TT-AT----GAC-C---T....-TA--AGATGGTTAAATG-G-CT-CTGAG-TT-AT----AATGTAAAT-T--T-T-ACA-G-C-A-G-AGGG-CTG-A-AGA-GA-  5857
LHOEST.CD.524     ACAACCATCT...........TT-AT---AGAT-C---T....-TA-ACGATGGTTAAATG-G-CT-CAGA--TT-AT--G-AATGTTCAT-T--T-TGACA-GGC---G-AGAG-TTGCA-GGA-GAG  5863
MANDRIL.GA.MNDGB1 ...................................................AATACAGAAG-A-AGGTAGAG--G-ATCAT-TATGTAG-T-T--CAT-ACA-G-C-A-GCAGGG-TTCGA-GGA-GA-  6263

COLOBUS.CM.CGU1   .............-ATTG-A-AG--TATCC-TGGT-T--T....--AGA......ATGAATTCA-CAGCTTTCTTAC-T-A--A--GAAA-G--TT-----TA-G-----CAGTAG--GA-TTA-CTG-G  6168
SYKES.KE.173      ACCTCCA-ATGAA-CG...............................TGGTGGGGAGA---C-GT-CAGAGCCA-G-TTCAA-TGCTCTTTT--CTTGACA-GGGGG--CA-GG-TAAAA-ACAGCA-  6719
RCM.NG.NGM        ......................G-C-C---TG-C-TG--T....--TGAA......TTTACGGTGGCAG-GTT--AA-G-GAT-AA--AA--A-AGT-A-TG-CA--....-GG--CTC-AGA---C-CTGG  7380
MANDRIL.GA.14CG   .......-ATAATGAG-GTACC-C-CA---A...--T-GGGTA--TAAATTTAACACA..................-C-G--TTATGTAG-G-TTGTA--ATA-AGA--AAAG-G-GCT--AGATATGAG  6768
MANDRIL.14CG  Env __.__.__N__N__E__S__T__T__Q__E__.__N__R__V__C__K__F__N__T__.__.__.__.__.__.__G__G__L__C__R__D__C__K__I__E__I__K__E__S__F__R__Y__E_  gp120
```

```
H1B.FR.HXB2   Env _P__I__D__N__.__D__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__S__Y__K__L__T__S__C__N__T__S__V__I__T__Q__A__C__P__K__V__   gp120
H1B.FR.HXB2       CCAATAGATAAT...GATACT.....................................................ACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTAT   6849

CPZ.US.CPZUS      -AT--CA-----.........AAT.................................................-C-----GAC---T--A---C------A-C-CT--C--------T-----T---ACC-   6893
CPZ.GA.CPZGAB     AACC---GG---...--G-ACAAC.................................................-CA----G-A-A-TT-A---C--T--TA---C---A-----A--A--------AACG-   6910
CPZ.CM.CAM3       TTG--T-----C...-----CGAT.................................................-CA----GAC-A-T--A---C--T---A---CT----------G-----T---ACC-   6289
CPZ.CM.CAM5       --T--T-GAGG-AATA-G-G-AATGAT..............................................-C-----GA--A-T--A------T--TA--ACTG-A--------A-----T--AACT-   6559
CPZ.CD.CPZANT     AAGTGTC-GG-CAACA--GAGACTAAC..............................................T-----T-C--ATGGCAC--C------A--ACT--C-----AT-----GA----TCTA   6302

H1A.UG.U455       -A---CA----A...AC-GA-AAC.................................................-AT--T----GAC-A-T--A------T------AC------------T-----C-----G-   6294
H1C.ET.ETH2220    ---C-TA----C...-G-T--....................................................--TGAT----GA--A-T--A------T------AC---A-----A-------------C-   6252
H1D.ZR.84ZR085    -A----------...--GGG-AAAAATGAAATCAAT......................GATACTTATGGT-CT----G---A-T--A------T-------C---------A--T--------A----   6391
H1F.BE.VI850      -------GC---...A-C-A-AGC.................................................-GTGAA----G-C-A-T--A------T------AC------------T-------------   6171
H1G.SE.SE6165     -----TA------...-CA......................................................--T-AT----G-C-A---CA------TGT----AC---C-A------T------------A   6273
H1H.CF.90CF056    -----T-----C...A---G-....................................................--TCAG----G-C-A-T--A------T--------------------T--C--------G-   6180
H1J.SE.SE92809    -----T----G-...A-C-A-....................................................-AA-AT----TTC-A-T--AC-----T-------------A---A--T-----------G-   6162
H1K.CM.MP535C     ---C-TA--GGC...--GGG-AATAACAGTAGT........................................--TGAA--C-G-C-A-T--A------T------AC----------A-T--------A---A   6057
H101_AE.TH.CM240  -A---T--AG--...A-G-AGACTAGT.............................................-GTGAG----G---A-T--A------T--T----------AG-----T---------A---   6429
H1N.CM.YBF30      -----CA--GCC.............................................TATAAT-AA-CA----G-C-A-T--A------T---A---CTG-G-----A--T-----T---ACT-   6442
H1O.CM.MVP5180    AAGG-TA--G-C...TCA-A-GCAGTAAAT...........................GGA--A-CA----T---A--T-A-------T--A--A-T--C-AG----------G------A   6884
H1O.BE.ANT70      GA-C-GA--G-G...ACA-GCAGCACAAATAAGACAAAC..................AGC-AA-TG----CA--A--T-A-------T--A--AC---C--G--A----------------   6884

MAC.US.SMM251     TAC-AT--A-C-...TGGTACTCTACAGATTTGGTTTGTGAACAAGGGAAT...AGCACTGATAATGAA-G---A-GCT-CA---ATCAC--------T--T--T--CCA-G-AT-T---GAC--ACAT-   7255
MAC.US.SMM239     TAC-AT--A-C-...TGGTACTCTGCAGATTTGGTATGTGAACAAGGGAAT...AACACTGGTAATGAA-GT--A-G-T-CA---ACCAC--------T--T--T--CCA-G--T-T---GAC--ACAT-   7273
MAC.US.SMM239 Env _Y__N__E__T__.__W__Y__S__A__D__L__V__C__E__Q__G__N__.__N__T__G__N__E__S__R__C__Y__M__N__H__C__N__T__S__V__I__Q__E__S__C__D__K__H__   gp120

STM.US.STM        TAC-AT--A-CG...TGGTACTCTAGCGATCTAATCTGTGAACAAAAT......GTCACTGGCGAAGAG-G---A-GCT-CA---G-CA---C--T---AGT--A---CA-G-AT-----GAC--ACAT-   6912
SMM.US.SMM9       TAT-AT--A-CA...TGGTA-TCAAGAGATTTAGTGTGTGAACAAAATAGCAATGGAAATGAAACTGAC-GT-AA-GCT-TA---ACCA----------AGT--T---CA-G-AT-----GAC---CAT-   6750

H2A.GW.ALI        TAT-AT--A-CA...TGGTACTCAAAAGATGTGGTTTGTGAACCATTTAACACCACCACAAAC...CAG-----G-G-T-CA---ACCA---C-----A--------C---G--T-A---GAT---CAC-   7336
H2A.SN.ST         TAT-AT--A-CA...TGGTACTCAAAAGATGTAGTCTGTGAATCAAATGACACCAAGAAAGAG......-AA-CA-G-T-CA---ACCAC--C-----A--------C---G--T-A---GAC---CAC-   6782
H2A.DE.BEN        TAT-AG--C-CA...TGGTA-TTAGAAGATGTGGTTTGTGACAACACAACA...............GCTGG--CA-G-T-CA---G-CA---C-----A---A----C-A-G--T-A---GAT---CAC-   7355
H2B.GH.D205       TAT-A---C-CC...TGGTACTCAGAAGATTTAGAGTGTAATAATACCAGGAAGTATACC.........-G---A-GCT-TA-A-G--CC--C-----AA--A-T--CCA-G--T-A---GAC--ACAT-   7322
H2B.CI.EHO        TAT-AG--C-CC...TGGTA-AAACAAGATTTAGTGTGTGAAAAGGGGACAAGGAGTAAT......GAA-G--AA-G-T-TA-A-A--CC-----T--A---A-T--CCA-G-AT-A---GAC--ACAT-   7317
H2G.CI.ABT96      TAC-AT--G-C-...TGGTA-TCACAAGATTTGCAATGYGAACAGTCCAATAAGAGTGAGAAT...GAG-GT--A-GCT-TA---GGCAC-----T--A---------CAGG-AT-A---GAT--ACAT-   6686

TANTALUS.UG.TAN1  GT-TGT--G--GGAAA-C-ACACCACA...................................GGCACG-GAG-T-G-T-CA---TTCAC--C---GA---T--A--A-A-G-A--T---GAG--AACT-   6924
GRIVET.ET.GRI677  GTCTGCA------...A-A--AGGAAGTGAAAAG............................GGAAGT-AGGAT-G-T-CA---T-CA------TGAT-----G--A-A-G-A--T---GAT--AAC--   6662
VERVET.KE.VER9063 TATTGTA-G--C...AGCT--AGCTCTAACTCT.............................AGCACA-AAGAA-GCT-TA---TTCAC-----TGAT-----T--A-A-G-A------GA---AAC--   6961
VERVET.KE.VERTYO  TATTGCA-A---...A-G---......AATAGC..............................ACTAGC-AAGAG-G-T-CA---TTCA------TGA------T--A-A-G-A--A---GAC--AAC--   6433
VERVET.KE.VER155  TTTTGTA-GCG-...AG---A...TCGCATAAT..............................GGGACA-AAGAG-GCT-TA---TCCAC-----TGAT-----T--A-AGG-A--T---GAT---AC--   6941
VERVET.DE.VERAGM3 ATGTGTA-G--G...-G-......AACAAT.................................TCTAAC-GAGAA-G-T-TA---TTCA------TGAT-----T--A-A-G-A------GAT--AAC--   6459
SABAEUS.SN.SAB1C  GTTTGT--AG-G...-GA-GGGAAAAATCAAAC............GCAACA............CATACTGTGG-A-G-T-CA---TCCA------T--------A--C-A-G----T---GAT--AACT-   7135

SUN.GA.SIVSUN     GTGGA-C-A--C...TT--GGGATACAGAAGTTACTTGCAATGGTAATGAC......................CA-G-T-TA-----CA---C--TGAT---A-T--------A-A---C-AT----GTA   7097
LHOEST.KE.LHOEST  AT--A-C-A-GC...TT--GATATGATGATGTAACGTGCTCAGGAGAGAGGGAGAACAGA............C--G-T-CA----TCA------TGAT---A-A--A-----A-AT---AAT----GGG   7035
LHOEST.CD.447     AT--AGC-G-G-...TT--GGTATGATGAGACAACATGTGATAAGAATGGA......................CT-G-T-TA----CCAC-----TGATAGTA-T---CAG--A-AT--CAAT----G-G   5966
LHOEST.CD.485     AT--AGC-G-G-...TT--GGTATGATGAGACAACCTGTGATAAGAATGGA......................C--G-T-TA----CCAC-----TGATAGTA-T---CAG--A-AT---AAT----G-G   5963
LHOEST.CD.524     AT--AGC-A-----...TT--GGTATGATGAGACAACATGTAATGAGAATGGT....................CT-GCT-CA----CCA---C--TGATAGTA-T---CTG--A-A---CAAT----G-G   5969
MANDRIL.GA.MNDGB1 AT-G--ACA-----...TT--GAGGGGATGATGTGAAATGTGAAAATAAT......................CT-GCT-TA---ATCA------TGAG-----T-A----G-A-A-----AG----G-C   6366

COLOBUS.CM.CGU1   -TTGG-A-.......A-A-A-GACAGTAATAGT...................................-T---GGCA-CAA---A-GA----GCA-AT-ATACAG----C--A-TA---GAT-T-ACTA   6256
SYKES.KE.173      TAT-G--C-TTC...TTCTA-AAGGATGACTTGATGAAAGAGGAAGGCAATAGTAGC...............TAT---T-CC--TT-CA------T--AAGT--T--AT-CGCA--A---GAG--ACAGA   6831
RCM.NG.NGM        TGT-AG-CA......-G--A-AATAATACTAGGCAATGT...................................TTTA-A-GGCA---------AA--AGT--ACA----TT----GA-CCAAA--   7465
MANDRIL.GA.14CG   GATG-GACCTG-.........ACTAAAAAGGAAAATGACACTGAAACT....................GAAGAA-G-T-TA-----CAC-----T--A-----A--A-------AT---AAT----CC-   6868
MANDRIL.14CG  Env _D__V__T__C__.__.__.__T__K__K__E__N__D__T__E__T__.__.__.__.__.__.__.__E__E__C__Y__M__T__H__C__N__T__S__V__I__T__Q__D__C__N__K__A__   gp120
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2   Env 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__G__T__G__P__.__.__C__T__N__V__S__T__V__Q__C__T__H__G  gp120
H1B.FR.HXB2       CCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCA......TGTACAAATGTCAGCACAGTACAATGTACACATGG   6973

CPZ.US.CPZUS      ----------------A-----C--------A--AC-A--C-----C--C------------G-A--AGAT-----A----AG---GAG.........----A------T--T-----G--G-----T-----   7017
CPZ.GA.CPZGAB     ----------T-----A-----C--------A--A--A--G-----A-----T-----C---G----AGAC--TTCA----A---TAA-.........-----T-----A--T-----G--T-----------   7034
CPZ.CM.CAM3       ----------------A--------C-----AA-CC-A----A---A--CA-G-----C----TGCCA-AC--T------------A--GGAAGA----AC---A-A--------C--------T-----   6419
CPZ.CM.CAM5       ----------------A--------C-----A--CC-A--------A--AA-G---------G--C-A-A---T--------------.........----------A-A--------------C--T-----   6683
CPZ.CD.CPZANT     -T-----A--------T--------------T--A--A--A-A---T--AT-G-G----G-AG--G--GAT--T-CA---GT---GATG.........----A-------TCAGT------T--C--T-----   6426

H1A.UG.U455       ----------------------------------A--------------------G-----GG--CCTGA-----------A---G---........-C-GG----------------------------   6418
H1C.ET.ETH2220    -T--G--C--------T--------------T--A-------A------------G----GAG-------A----C--------------........-CCAT-------------------C--------   6376
H1D.ZR.84ZR085    ----------------------------------A-----------A--------G-----CG----A-G---------------C---........-C-A----------T------------------   6515
H1F.BE.VI850      -T-GG--T-----C--T--------------T--A-------A---A---------------G----A-GA--------G-----G---........-C-AG----------------------------   6295
H1G.SE.SE6165     -T-----C----------------------T--AA--------------T----G----GGG-----GA-------------------.........----A---------T--------------------   6397
H1H.CF.90CF056    -------A--T--------------------T--A-----C--------------G-----C-----A--A---------------TT-.......--------------T--------------------   6304
H1J.SE.SE92809    -T---C-A--------------------------A--------------------------G------AT---------------T--........-C-A---------T-----------------C--   6286
H1K.CM.MP535C     ----------------------------------A-----C--------------G-----GG----A-G---------G---------.........----A-----------------------------   6181
H101_AE.TH.CM240  -------T--------T------------A-T--A-------A-------T----G------G------AT--------G-----G---.........----A----------T----------C--------   6553
H1N.CM.YBF30      ----------------A--------C-----A--AC-A--C-----C---A-G---------G-AGGA-AC--T-G-----AT---AGC.....----------G--T--T--------C--------   6566
H1O.CM.MVP5180    GT--------C-----------C--------T--AA-A--A-A---C--CT-T--G------G-C-CAGAC--T-----------C-T-........-CCAC---A-TTCAGTG--TACT-----------   7008
H1O.BE.ANT70      -T-----A--------A-----C--------T--A--A--A-A---T--CT-T--G-----C-GC-CAGAA--T-----------CA--........-C-G---CA-A-CGGT---TACT-----------   7008

MAC.US.SMM251     AT-GG--TA-T---AGAT-TAGG--------A--TC-A----A---TT-G--T-G-------G-C-CA-AT-ATTCA--CTTTATG--T...AAA---T-T--G--GGTGGTCTCTTC---C---AGGAT   7382
MAC.US.SMM239     AT-GG--TG-T---AGAT-TAGG--------A--TC-A----A---TT-G--T-G-------G-C-CA-AT-ATTCA--CTTTATG--T...AAA---T-T--G--GGTGGTCTCTTC---C---AGGAT   7400
MAC.US.SMM239 Env Y__W__D__A__I__R__F__R__Y__C__A__P__P__G__Y__A__L__L__R__C__N__D__T__N__Y__S__G__F__M__P__.__K__C__S__K__V__V__V__S__S__C__T__R__M  gp120

STM.US.STM        AT-GG--TG-CT-AAGGT-TAGG--------A--AC------A---TT-G--C-G---C---G---CA-AT-AT-CA--CTTT-C---T...AAT--CT-T--G--AGTAGTCTCCTCG-----TAGAAT   7039
SMM.US.SMM9       AT-GG--TG-T---AGAT-YAGA--------A---C-A----A---TT-G--T-GG------G---CC-AT-ATTCA--CTTT-CT--T...AAC-----T--G--AGTAGTTTCTTC---C---AGAAT   6877

H2A.GW.ALI        AT-GG--TG-T--AAGGT-TAGA--C-----A--AC------AC--CC-A----G---CG--G---TC-AT-ATTCA--CTTT-C---C...AAT--CT-T--A--AGTAG-T-CTAC---C---AGGAT   7463
H2A.SN.ST         AT-GG--TA-T--GAGGT-TAGA--------A--AC-G--------CC-G----G---C---G---CC-AT-ATTCA--CTTT-AG--C...AAT---T-T--G--AGTAG-T-CTAC-------AGGAT   6909
H2A.DE.BEN        AT-GG--TG-T--GAGGT-TAGA--C-----A--AC-G--C-----CC-AT---G-------G---CC-AC-ATTCA--CTTT-A---T...AAG--C--T--A--AGTAG-T-CTTC---C---AGGAT   7482
H2B.GH.D205       AT-GG--CAGCT-AAGGT-TAGG--------T--CC-G--G---TTTC-A----G-------G---CC-AC-ATTCA--CTTCATG--C...AAC--C-GT--G--AGTAG-GTCCTCC--C---AGAAT   7449
H2B.CI.EHO        AT-GG--TAGTT-AAGAT-TAGG--------T--CC-A--A-----TT-G----G---C---G---CT-AA-ATTCA--CTTCATG--C...AAC----GT--G--AGTAGT-TCT-TG...TACAGAAT   7441
H2G.CI.ABT96      AT-GG--TG-T--AAGAT-TAGA--------T---C-A--G-----TT-G----G-------G---CT-AC-ATTCA---TTT-C---T...AAT--CT-T--A--AGTGGTCTCTAC-------AGAAT   6813

TANTALUS.UG.TAN1  AT-GG--TA-CT-AAGAT--AGA--C-----A--A--A--C-----T--CT----------GG---CT-AT-AT-CA---TTT--TGTT.........----G------TTCAGT---TAGT-----TGGATT   7048
GRIVET.ET.GRI677  AT-GG--TA-TT-AAGAG--AGA--C-----A--A--A--G-A---TT-G------------G-----GAT-AT-GA--CTTT-CT---...AAG--C-AG-----TTCAGT---G--T-----TAGATT   6789
VERVET.KE.VER9063 AT-GG--TGAGC-AAGGC--AGA--C--------A--A--A-----TC-GT-------CC--G--C-TGAT-AT-----TTATAAG-A-...AAT----GT-----TTCAGT------C-----TGGACT   7088
VERVET.KE.VERTYO  AT-GG--T-AGT-GAGGT--AGG--------T--A--A----A---TT-G------------G--G-AGAT-AT-----GTATAA--A-...AAT--CT-------ATCAGT---G--T------GGCTT   6560
VERVET.KE.VER155  AT-GG--TGA-T-AAGAC--AGA--------T--A--A--A-AC--TT-G--T---------G--TG-GAT-ATGCA---TTTAAG---...GAA---T-T-----TTCAGT---G--T--C---AC-TT   7068
VERVET.DE.VERAGM3 AT-GG--TGAGT-AAGAT--AGG--C-----------A--G-----TT-AT-------C--CG--T-TGAT-ATGCA--GTTTAAGA--...AAC---T-T-----TTCAGTG--G--T-----TA-CTT   6586
SABAEUS.SN.SAB1C  AT-GG--TA-CT--AGGC-TAGG--------T--A--A----A---TT-G----G----GCAG---CAGAT-ATTCA---CATAAG......GCA----GG-----A-CAGT-TC-GCT--C--TAGGTT   7259

SUN.GA.SIVSUN     TTA-GC-AAATGC-TA-T-CAGGCT------A--T--G--G-A-AT-T-AT---G----G--G-A--ATTAAATGC-ACT-AGAA-.............----A----A-A-CTG-TACC-CT-----TA--TA   7218
LHOEST.KE.LHOEST  TAA-GC-AAATGC-TATT-TAGGCT------T--A--A--C-A-AT-C-AT-G-G---C---G-AC-ATTAAA-TT-A-T-A-AA-.............---GA----A-A-CAG-TAC--C------TGGATA   7156
LHOEST.CD.447     TAA-ATCTAATGC-TATT-TAGGCT------A--A--A--C-A-AT-C--T----G--C--CG-A--ATT-AATTTCTC-G-TAAT.............-----C---A-A-CAG--ACC-CT--C---G-CTA   6087
LHOEST.CD.485     TAA-ATCTAATGC-TATT-TAGGCT------A--A--A--C-A-AT-C--T-------C---G-A--ATT-AATTTCTC-G-TAAT.............-----C---A-A-CAG--ACT-CT--C---G-CTA   6084
LHOEST.CD.524     TTA-GTCAAATGC-TATT-TAGGCT------A--A-----C-ACAT-C--T-------C---G-A---CTAAA-TT-TC-G-TAAT.............-----T---A-A-CAG-TAC--C---C---G--TA   6090
MANDRIL.GA.MNDGB1 TT...TT-AT--GATGT--TTTAGG-----TG--TC-A--A-A--TC--GT---G--A----G-G---TTAAAT---AAT-A-TTG.............----GC---A-ATCAG----G--G-----T--GCA   6484

COLOBUS.CM.CGU1   TAG-G--T---G--AGG-C-GGG-TC-----AG-TC-A--A-ACAT-C--T---G----G--G----A-A--GGG----G------G--........-C-AC-----G-CAG----GTCC-----C----A   6380
SYKES.KE.173      -----C-A--CT----G--T--A--C-----T--AC-G----ACT-AT-GT-G-----C---G---CA-AT--TG-A---GAC-ATGTC.............----------A-CTG-----TC---C-----G-A   6955
RCM.NG.NGM        A--GG--A---T--AGGT--AGG--------A---C-A----A---C--A--G------CT-G--GCA-AT-AT-CA--CTTT-ATA-C.........----CA------T-CAG--ACGTC---C-----CAT   7589
MANDRIL.GA.14CG   -AACA--CAA---GA-AT-TAGA-TA--------TC-A--G-ACATTC-G----G----CG-G-G--ATT-AAT--AACT-A-TT-............----G-C------TCAG-------G-----TG-CCC   6989
MANDRIL.14CG  Env S__G__D__K__M__T__F__R__L__C__A__P__P__G__Y__I__L__L__R__C__R__E__K__L__N__K__T__K__L__.__.__.__C__A__N__V__S__A__V__Q__C__T__D__P  gp120
```

```
H1B.FR.HXB2   Env __I__R__P__V__V__S__T__Q__L__L__L__N__G__S__L__A__E__.__E__E__V__V__I__R__.__.__.__.__.__.__.__S__V__N__F__T__D__N__A__K__T__I__I_  gp120
H1B.FR.HXB2       AATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAA...GAAGAGGTAGTAATTAGA.....................TCTGTCAATTTCACGGACAATGCTAAAACCATAATA   7079

CPZ.US.CPZUS      C--A-AA-----G--GA-T--A--G--CA-AA-----------T-----ACT...A--A-T--TACTG-A---.......................AGT...AAA...A--TT---AG-C-TT--TC--   7114
CPZ.GA.CPZGAB     -----AA-----G--TA-C-----GT-A--TA-T-----A---T----T---...-G-A-CA--ACTG-A---......................GTG-AG---AAGT-AA-A---A--G-TGT-TGG---   7140
CPZ.CM.CAM3       C--C--A-----G---A-T--A---T--A-CC-T-----A---G----C---...A-CA--ACTA-TGC----.................................AGG--C-GGT-C-R-T-CC-G   6510
CPZ.CM.CAM5       C--A--A---------A-T--A---G--A-CC-T-----A--C-----T---GAG-GGCCC-CT--TGC----....................AG--AAG--C-A...AGAC-CAAA-TC-TT------   6789
CPZ.CD.CPZANT     ---A--C---A-G--GG----ATGGT-A--A--------A-C-TACCA-AC-AACACTTCA--------G-AT.................GGTCGCAAA---.........-ATC--TGCTTGTA-G-T-T   6532

H1A.UG.U455       ---C-A------G---------------------------------------...AG---AA--AGG------.....................----AA--------AA-------A------------   6524
H1C.ET.ETH2220    -----AA-----G--------------AT----------T---A---------...-GT---AC-A--------.....................-T--AA---C-G--AA-------C----TA------   6482
H1D.ZR.84ZR085    ------------G--G-----------------------------------...------A-C-----A---......................----AA---C----AA------------T-------   6621
H1F.BE.VI850      -----AA-----G------------T--T-------------C---------...----GTA-------C---......................---CAA---A--T-AA-T-----A------------   6401
H1G.SE.SE6165     -----A------G--------------A--AC-G---------T--------...-G-A-AA--AA-G-----......................----AA--------A------A-C---GT-------   6503
H1H.CF.90CF056    ------A-----G----------------A--------A--C---------...---C--A-CA--------.....................A--AAA---A--T-A------A-C----A-------   6410
H1J.SE.SE92809    -----A------G--------------A---C-----------A---------...-G---CA--A--------.....................----AA--CA--T-A--------C----A-------   6392
H1K.CM.MP535C     -----AA-----G----------------------------C---------...------A--A-------G.....................----AA---A-T--A--T---A-A----A-------   6287
H101_AE.TH.CM240  -----A------G------------T-------------------------...------A--A----C---......................----AAG--C----AA-------C------------   6659
H1N.CM.YBF30      ---A-A------GA----C-----GT-AA-CC-------A--CT--AATAC-...--T-GAA-T--T------...................................-AT--T-G-CAC-GT-ATC-GT-G   6660
H1O.CM.MVP5180    C--C-A----AC----AGT--------AA-AC-G-----G-CA--CT-TAG-...---A--A--AG-----TG...................GGAAAA---A-T--A--ATCA--A--G-AT--C---   7114
H1O.BE.ANT70      C---------AC----AGT-----G--AA-A--------G-CA--CT-TA--...-G-A-AA--AG---G-TG...................G-AAAAG--A-TTT---AGG--GA----AT--C---   7114

MAC.US.SMM251     G--GGA-A--CAGACT--T---TGGT-TGGC--T-----A-C-AG-------...A-TAGAACTTAT---TAC.......................TGGC--GGT-G---T---AGG-CT-TA--T-GT   7485
MAC.US.SMM239     G--GGA-A--CAGACT--T---TGGT-TGGC--T-----A-C-AG-------...A-TAGAACTTAT---TAC.......................TGGC--GGT-G---T---AGG-CT-TA--T-GT   7503
MAC.US.SMM239 Env __M__E__T__Q__G__S__T__W__F__G__F__N__G__T__R__A__E__.__N__R__G__Y__I__Y__.__.__.__.__.__.__.__.__W__H__G__R__D__N__R__T__I__I__S_  gp120

STM.US.STM        G--GGA-A--CA-ACT------TGGT-TGGC--T-----A-C-AG-------...A-TAGAAC-TAT--CTAT.......................TGGC--GG--G---T---AGG-CT-TT--T-GC   7142
SMM.US.SMM9       G--GGAAA--CA-ACC--T---TGGT-TGGT--C-----T-C-AG-------...A-TAGAAC-TAC---TAT.......................TGGC--RG--GAAG----AGA-CC-TA--T-GC   6980

H2A.GW.ALI        G--GGA-A-GCA-ACT--T---TGGT-TGGC--T-------C-AGG------...A-TAGAAC-TAT--CTAT.......................TGGC--GGT-GA--T---AGA-CT-T---C-GC   7566
H2A.SN.ST         G--GGAAA-GCA-ACC--C---TGGT-TGGC--T-------C-AGG------...A-TAGAAC-TAT--CTAT.......................TGGC--GGT-G---T---AGA-CT-T---T-GC   7012
H2A.DE.BEN        G--GGAAA-GCA-ACT--T---TGGT-TGGC--T-------C-AG-------...A-TAGAAC-TAT--CTAT.......................TGGC--GG--GA--T---AGG-CT-T---T-GC   7585
H2B.GH.D205       G--GGAAA--CAGTCC--T--ATGGT-TGGC--C-----T-CAAGG-----G...A-CAG-AC-TAT--ATAT.......................TGGC--GAA-AA------AGG-CC-T-----GC   7552
H2B.CI.EHO        G--GGAAA--CAGACC--T--ATGGT-TGGC--C-----T-CAAGG------...A-CAG-AC-TAT--CTAT.......................TGGC--GGT-AA------AGG-CT-T-----GC   7544
H2G.CI.ABT96      G--GGAAA--CA-ACC--T--ATGGT-CGG---T-----A-CAAGG------...A-TAGAAC-TAT--CTAT.......................TGGC--GGG-GA------AGG-CT-TA----GT   6916

TANTALUS.UG.TAN1  G--G-ATA-TAC---GAGCT-AGC-T-TGGCA------------AG-------...A-CAGAAC--A---AT-G..................CAAAAGCATGGAG-G-GCA-T---T--GTG-TA----A-   7157
GRIVET.ET.GRI677  ---C-ATA-TACTA--A-T--AGGGA-AGGA--------T---AG-T------...A-TAGAAC--AG--AT-G..................CAGAAA-GAGGAAATGATA-TG--A-AGTT-TA----AG   6898
VERVET.KE.VER9063 T--G-ATA--AC---T--T--AGG---A-----G-----A--CTATT-T--G...A-TCGAAC-CA---AT-G..................CAGAAGCAT-GAG-G-ATA-TG--T--GT--TAG-TC-T   7197
VERVET.KE.VERTYO  ---G-ATA--AC---GA----AGGGT--T--C-G-----A--CTATCAT--G...A-TCGAACCCAG--AT-G..................CAGAAACAT-GGG-A-ATA-T--CA-AGT-TTG--CT-G   6669
VERVET.KE.VER155  ---G-ATA--AC----A-C---GGT---T-A--G-----A--CTATT------...A-TCGAACCCAG--CT-G..................CAAAAACATGGAG-G-GCA-TG-CT-AGTGTTA--CT-G   7177
VERVET.DE.VERAGM3 G--A-ATA--AC---GA-T---GG----T----G-----G--CTACT----G...A-TCGAACCCAG--AT-G..................CAGAAACAT-GAG-A-GCA-TG-CT-AGTGTTAG-GT--   6695
SABAEUS.SN.SAB1C  ---C-ATA-TACT---AGC--AGG-A-AGGAA-T-----T---TAT-T--C-...A-TAGAAC--A---CT-G..................CAGAAAAATGGCAATT-CA-TG--T--GTT-TT--T-GG   7368

SUN.GA.SIVSUN     T--G-CAAGTAC---TAGT-GCTTTT-TGGC--T-----A-C-AG-CATA-G...-----T-AGT----ACC-ATCAAT.........AACAAGGTA-GAC-AGGGG-T--GGGA-AAT-TGTATGG-A-   7336
LHOEST.KE.LHOEST  T--GCTTAGTAGT---AGT-G-TTCT-TGGC--T-----A-C-AATCATACC...AGG--T-AGC-T---CC-CTAACT......CCAAATAAGATG-AAG--C-A-AT-GTGCAAAGTTTGTGTAT-A-   7277
LHOEST.CD.447     T--G-TTAGTTC---GAGT-G-TTCT-TGGC--T-----A-CCAATCATAC-...--C--T-A-T-----CC-CTAACT......CCAAAAAAGATG-GA---C--CTA--AGCAAAGTTTGTATAT-A-   6208
LHOEST.CD.485     T--G-TTAGTAC---GAGT-G-TTCT-TGGC--T-----A-CCAATCATAC-...--G--TAAGT-----CC-CTAACT......CCAAAAAAGATG-GA---C--CT----GCGAAGTTTGTATAT-A-   6205
LHOEST.CD.524     C--G-TAAGTTC---TAGT-GCTTCT-TGGC--T-----A-CCAATCATAC-...-----C-AGT-----CC-CTAACT......CCAAAAAAGATG-GAG--C-T-AT-GAGC-AAATTTGTTTAC-AG   6211
MANDRIL.GA.MNDGB1 CT-AGTAG-CAC---AGT-GCTTTT-TGGC--T-----A-C-A-GCATA-G...----GA-A-T-G--ACCCATAGATGATAAATATAGGGGCC-A-AGG-A--TCATC-A-GGAAGTTTGT-TAT-AG   6611

COLOBUS.CM.CGU1   GT---ATAT-AC---CATGT----TG-AT-AG-C----C-TCAAA--AGTTG...TCT--TTGG-CG-AAGAT..............AGAGAGGGA---TGGAAG-AT---TCA-GG-C--TTGAGTAT   6492
SYKES.KE.173      -T-C-ATA--T-G-CTAGC--ATGGT-C-AAC-G-----G-CCTACAA--C-...A----TAA---T-GATTTATCAAA............CAAAAG-AT--AAAT......G-GT-AGT--TT--TT--   7064
RCM.NG.NGM        G--A-ATA--AC------C--ATC-T-TGGA--------ATC-A-TAGT--G...A-TAGAAC-TGG--CTAT.......................CAA-GAAAACAAAGT---AGG-C-GTT---GGG   7692
MANDRIL.GA.14CG   ---GCCTG-TACTA--------ATGT-TGGC--T-----A-CAAA-CATA-T...T-T--T-AGT-----TTG...........GTGAACCCTCAAAAGG-G--TCAT---C--AAAT-TGTTTAT-GG   7104
MANDRIL.14CG  Env __M__P__A__G__I__S__T__M__F__G__F__N__G__T__K__H__N__.__Y__D__E__L__I__L__.__.__.__.__V__N__P__Q__K__E__F__H__D__H__K__Y__V__Y__R_  gp120
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
                                                |- V3 loop                                                  loop tip
H1B.FR.HXB2        GTACAGCTGAACACATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGAATCCGTATCCAGAGA...GGACCAGGGAGAGCATTTGTTACAATA............GGAAAAATA.   7193
H1B.FR.HXB2   Env  _V__Q__L__N__T__S__V__E__I__N__C__T__R__P__N__N__N__T__R__K__R__I__R__I__Q__R__.__G__P__G__R__A__F__V__T__I__.__.__.__.__G__K__I__   gp120

CPZ.US.CPZUS       -----AT-CTCAGAGGGA--CA-T--G-C-----TT-----AGGA---------GTGGG--ATG-A-AAC-A.....................A-TGA-T---TA--AC--------------CC---G---G   7223
CPZ.GA.CPZGAB      --------AGTAGA-G-A--TAGTC-A------CAT-----AGGA-----------GGG-GAGG-G-AG--A...........G-----A-TGA-C---TA--AT--------------A---TG--G   7249
CPZ.CM.CAM3        A-N--RT-TC-G-ANN-G--CAGC--A--------------AGGA--T-NC-GC---GG-CA---A-AG--A.....................C-TGA-C---TA--AC--------------A---T--TG   6619
CPZ.CM.CAM5        CA-...T-T--AGA-GGGA-TA-C--A-T-----TG-----AGGA--T--C-GC---GG-CA-C-A-AG--A......................GT-T-C--CTA--AT--------------A---C---G   6895
CPZ.CD.CPZANT      -G-A-AGAATT-GA-AACT--AC-----CA----T------AGGA--T-GG---GT--G--ATC-A-AA--A.....................A-TGA-T--CTA--ACG-------------A--T-GC-A   6641

H1A.UG.U455        --------TGT--ATC-----A----C------T-C-----TT----T...----------AT--A----GGT-T--TATA---T----ACA------CTA-GT--C-------------T------A   6639
H1C.ET.ETH2220     --------T--TGA--------------C------G-------G---T---------G----T--AA-G--A.....................ACA-A----CTA-G---C-------------G-C---A   6591
H1D.ZR.84ZR085     -----T--T--TCA------------------------G---T----A--GGA----C----G-CA-C---A..............A----CA---GC-CTA-----C-------------A-GT-T-C-A   6730
H1F.BE.VI850       -----C--T--TGA-------C-G-------------------------------G---G-T--A-A-T-A.....................ACA-A----CTA-G---C-------------GC----A   6510
H1G.SE.SE6165      --------T--T-A-A-------------------C------------------TG-----G--AA-A--G...G--ATT--------ACA-A-G--CTA-G---C-------------TGCC---A   6618
H1H.CF.90CF056     --------T--G---C-A---A-C----CA-----C--G--T-----T-----G----C---T--A-A-T-A...........G-----AC-------CTA-G---C-------------TG-C--CA   6519
H1J.SE.SE92809     -----A--T--T-A-A-----------GTG---TAC-----T--T----------G---G-T--A-AC--G.....................ACA--TGC-CTACG---C-------------G-----A   6501
H1K.CM.MP535C      --------T--TGA-A-----C----------------G--A--------------------T--A-A---G.....................A-A------CTA-----C-------------TG-T---A   6396
H101_AE.TH.CM240   --G--C--T--T-A------------C--------C------TC-------------C---T--AAC---A.....................AC---T---CTA--G--C-------------G-T---A   6768
H1N.CM.YBF30       --G--ATG---TGAGA-A--GCC---A-----------G--AGGA--T------G--GG-CAGG-G-AG--A...............T-CT-TGA-----TA--AC--------------A-------G   6769
H1O.CM.MVP5180     ---ACC--A-----TC--A--A-C--G-CC--C-T----GAAGGA-TTGCAGAGGT-C--GAT--ATA--CA..........--T---AT----TGGCGCAG--TG-C-CTTAAAAGAAGTAAC--T-C-T   7235
H1O.BE.ANT70       --GACC--A---T-TA-CC--A-C--G-CC---GA------AC-A-TA...GAC-T-C--GAG--GA-A--A..........--T---AT-GCC...-GGTAC-GC--GGGAATAGGG...----C-GC-G   7226

MAC.US.SMM251      T--A-TAA-T-TTATAA-C--AC---G--A----G------AGGA--T--G---GTTTT-CC-G--ACC--TAT-TCT.....-TTG-TTTTC...CACTCA..................CA-CC---CA   7591
MAC.US.SMM239      T--A-TAA-T-TTATAA-C--AC---G--A----G------AGGA--T--G---GTTTT-CC-G--ACC--TAT-TCT.....-TTG-TTTTC...CACTCA..................CA-CC---CA   7609
MAC.US.SMM239 Env  _L__N__K__Y__Y__N__L__T__M__K__C__R__R__P__G__N__K__T__V__L__P__V__T__I__M__S__.__G__L__V__F__.__H__S__.__.__.__.__.__.__Q__P__I__   gp120

STM.US.STM         T-GA-CAA-T-TTACAA-T--AC---G-GC----G------AGGG-----A---GTCTT-CC-G--ACC--TAT-TC-.....-CTTG-TCTTC...CA-TCG..................CA-CCC--TA   7248
SMM.US.SMM9        T--A-TAA-T-TTRTAA-C--AC---G-GA----G------AGGA--T--G---GTCTT-CC-G--ACC--TAT-TC-.....-GTTG-TCTTC...CA-TCG..................CA-CCC---A   7086

H2A.GW.ALI         T--A-CAAAC-TTATAA-C-TACT--GC------AG--G--AGGA--T--G---GTTGT-CC---AACAC-TAT-TC-.....-GTT-ATATTT...CACTCC..................CAGCC---CA   7672
H2A.SN.ST          T--A-CAA-TTTTATAA-C-CACCG-AC------AG--G--AGGA-----G---GTTGT-CC---AACAC--AT-TC-.....-GTT--T-TTT...CACTCC..................CAGCC---CA   7118
H2A.DE.BEN         T--A-CAA-T-TTATAA-C-CAC---GCG-----AG-----AGGA--T--G---GTTTT-CC---AACAC-TAT-TC-.......--TT--T-TTT...CACTC-..................CAGCC---CA   7691
H2B.GH.D205        T--A-TACAT--TATAA-T-GTC---AC-C----AG--G--AGGA-----G--GGTTGT-CC---AA-A-C-GT-TC-.......-T-CTTTTC...CA-TCA..................CAGCCT--CA   7658
H2B.CI.EHO         T--A-TTCAT-TTATAA-C-GAC---GC-C----A---G--AGGA-----A-TGGTTGTGCC---AA-A-C-GT-TC-.....-TATTCTCTTC...CA-TCA..................CAGCCT--CA   7650
H2G.CI.ABT96       T--A-TAA-T-TTATAA-T--AC---A-RA----GG-----AGGG--T--A---GT-TT-CC---TAC---TAT-TC-.....-G-T--TCTTC...CACTC-..................CA-CC---CA   7022

TANTALUS.UG.TAN1   C--A-TAAAC-TTATAAGT-GA-G--AGTG----G------AGGA--T--G---GTCTT-CC-G-AACC---AT-GC-.....-T-TC-T-TTC...CACTCA..................CA-C--TATA   7263
GRIVET.ET.GRI677   T-GA-TAA-TTTTACAACT-GAC-G-G-GA--CCG------TGGT--T--A---GTGTTGCC-G-AACA---AT-GC-.....-GTT--TATTT...CACTC-..................CAG---TATA   7004
VERVET.KE.VER9063  T-TA-CAA-T-TTACAA-T-GAC-G-C-C---C-A---G--AGGA--T--G---GTCTT-CC-G-AACC--AAT-GC-.....-TTT--TTTTT...CA-TCG..................CAG---TACA   7303
VERVET.KE.VERTYO   T-CA-CAA-C--TATAA-C--TC-G-C-CC----GG-----AGGA-----G---GTCCT-CCGG-AACG--AAT-GCG.......-TG-TTTTC...CACTC-..................CA----TACA   6775
VERVET.KE.VER155   C-CA-TAA-C-TTATAACC-GAC-G---CA--C-A---G--AGGG--T--G---GTCTTGCC-G-AACG--AAT-GC-.....-TT--TCTTC...CACTCA..................CAG--GTATA   7283
VERVET.DE.VERAGM3  T-TA-TAAAC-TTACAA-C--AC-G---C---C-A------AGGA-----A---GTCTT-CC-G-AACA---AT-GC-.....-G-T--T-TTT...CA-TC-..................CAG-GGTACA   6801
SABAEUS.SN.SAB1C   C-CA-TAGAT-TTTTAA-T--ACT--C-GA----G------AGGA--T--G---GT-TTGCCTG-AACC---AT-GCT.....-CTTG-TCTTC...CA-AGC..................CAG---TATA   7474

SUN.GA.SIVSUN      ---GCAGCT--GTGGGGGT---T----C-G--C-T---GAAAGGG--T-GGT--CA-GT---T-CTATA-GTAGT-C-.......-TGCT-TTTTAT-A--GG..................CT-G-GCC-G   7445
LHOEST.KE.LHOEST   ---GCAGGA--ATGGGG-T--ATT----GA--C-T----AAAGGA--T-GAT--GA-GT-TCC-CAAT--G-TCT-C-.......-TATTTATTTTAT-A--GC..................CT-G--CATG   7386
LHOEST.CD.447      ---GCAGGA--ATGGGGAC-CATT----GA----TT--GAAAGGA--T-GGT--GA-GT-TCT-C-AT--GTTCA-C-.......-TTCT-TTTTAC-A--GC..................TTGG--CATG   6317
LHOEST.CD.485      ---TCAGGA--ATGGGGAC-CATT--C-GA----TT--GAAAGGA----GGT--GA-GTCTCT-C-ATC-GTTCA-C-.......-TTCT-TTTTAC-A--GC..................TTGG--CATG   6314
LHOEST.CD.524      ---GCAGGA--ATGGGGAC--ATC----GA----T----AAAGGA--T-GGT--GA-GTCTCC-CAATC-GTTCA-C-.......-TTCT-TTTTAC-A--GG..................TT-G--CATG   6320
MANDRIL.GA.MNDGB1  --G-CAGGA--ATATGGCT--A-G--AG-A---CAC---AAAGGA--T-GGT--GT-GTG--T-CT-CATCAGCT-C-.......-TT-TTATTTTATCA--GG..................TT-G--CCTG   6720

COLOBUS.CM.CGU1    TATTG-T-TCCA-A-GA-A---CTT-AGGG----T----AGAGGA--T-G-T-TCAC-G--ATC-AAAC-CTGCC-AT.....--G--AAATTCTAC-A--AG..................CTT-TTCCTT   6601
SYKES.KE.173       --T-CAGAAGCACTTAGAT-GC-G----TA---GA------AGGA--TG-GT-T-TC----AT--T-AAT-GGCTGC-.....-CTATTTCCTCC--G-AA-A.........CAAGGGAAATTG----CCG   7182
RCM.NG.NGM         T--A-TAGAC-TTTTAA-T--AC-G-A-C-----GG-----A-GT----GA---GTG---G-----TCAT-AGCA-C-.......-GT-TTC-TCT-T.......................CT--GGG-GC   7795
MANDRIL.GA.14CG    --GA-CAAA--ATGGGGGT-GA-GG-GGTA---GAT--GAAGGGA----GGT-T-T-GTTTC-G---CA-GTGCATC-.......-TT-ATTTTCTATCA--GG.................TT-G--CC-G   7213
MANDRIL.14CG  Env  _V__N__K__K__W__G__L__K__V__V__C__D__R__K__G__N__R__S__I__V__S__V__P__S__G__S__.__G__L__I__F__Y__H__G__.__.__.__.__.__.__L__E__P__   gp120
```

```
                                V3 end -|
H1B.FR.HXB2        ..GGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATC.................   7304
H1B.FR.HXB2   Env  .__G__N__M__R__Q__A__H__C__N__I__S__R__A__K__W__N__N__T__L__K__Q__I__A__S__K__L__R__E__Q__F__G__N__N__K__T__I__I__.__.__.__.__.__.  gp120

CPZ.US.CPZUS       TA---G--G-A---G----------------CTCA-A-CTG-C----G-G--ACAGAG----T-C-CTTTA......GAG-T-AT-A--AAG-A-GCA--CCTG----AGG-AGAGTTAATTCCA...AA   7344
CPZ.GA.CPZGAB      TA---G---CC---TCT--CT-C-----G--A-A-G-GA-C-C-------CGA--GG-GG--G-AG--AAA-AGGCT---GC-ACCTC--CAAATCGA-CAGC-G-C-AC--AACACTCAACAGA...GC   7376
CPZ.CM.CAM3        TA---G---CC----G----T-----C-G--C-A---GA--GTC---G-CG-A-GAC----TG-A-CT-G-CAAGCC--GC----G-TT---AC---CCT--C-CA-G-C-A-TTT...ACAGTA...AG   6743
CPZ.CM.CAM5        TA--GG--CCC-----------------TG-C--C-AG--CC-G----TG--A-AG-----CG---CC---TTAGCCA-....-C-A-G-A--AT--A-GC-T-G---A-G--ACCTTTACACCA...AA   7019
CPZ.CD.CPZANT      CA---G-C-CT--GA----GTTC----CAG-C-A--AGA-GCT----G-AC-AG-ACGT--CA-A-C--AGCA-GTTC-TGCG--G--T-GGAA---AGTAG-C-AC-A--C-AAT.............   6757

H1A.UG.U455        TA--GG----A-----------------TG-C------AGGG-C-------GA---A--C-----G-----GAAC------AGA--A-G---AAT--C--A-C-..........................   6743
H1C.ET.ETH2220     TA---G----A-----------------------GA--A-----------A---C--C--A--G--AAGGAA------CA-A-G--C--CCCT---...-------.......................   6695
H1D.ZR.84ZR085     CAA-G-TA--AG--------T-------T------G---T---------------C---CG----G-------A------G--A-C-TT--GAACC-A-CA---.........................   6834
H1F.BE.VI850       TA---G-C--A---A-G-----------------G--A--C-------------C-GG-GT-TG--AAGGCAG------AGTCG--T--CCCT------.....---GC...................   6617
H1G.SE.SE6165      TA---G-C--A-----------------TG---C--A-AG---------AG-GG-C---C-GA-TG-C---GCAG--C--G-GA--ATC---AAT--G-GC-GCGAG-AC...................   6728
H1H.CF.90CF056     TA---G----A-----------------T---------A--G-C--------G------C-C---G---T--CAC-----G--AT---C--GAAC----GA-C-..........................   6623
H1J.SE.SE92809     TA--------A---G--A-----------------GA-AG-G-T----G----------CGTAGAG------CA---C----------C---AAT--A-CA...........................   6602
H1K.CM.MP535C      TA---G----A-----G--------C---------G---A---------C-TG-------GCAGAG--AAGGAA--GC---A------T---AAG---GGA-C--T-.......................   6503
H101_AE.TH.CM240   TA--------A---A-----T-----G-G----A-G--A------------AGT----------G--A--GAA-------A---G--C---AAT--G-CA...........................   6869
H1N.CM.YBF30       TA---G-C--T---------T-C-----TG-CTC--A--A-CT----G-ACCA-TG-GG--TAGA-C-AGAGAGG--A---AGA--ATCC-G--G--A--C--C.......................   6873
H1O.CM.MVP5180     CACC--GATCA--GGT---TT------CATA--A--AGA-TGT----G-A--TG-CC--C----A-C-----TA-GG-ATTT-A-T-TTG-AAACC-A-CAG-G-AT.....................   7342
H1O.BE.ANT70       GAAAC-GCTCA--GGC---TT----C--GTA--A-GCCA-TG-T---GGA--A-TA--------A-C----GAA-GG-ATTT-----T-G-AAAC----CAGGT-GT.....................   7333

MAC.US.SMM251      AT-AT-GGCCA-AG--G---TGG---TGGT--...G---G---------GG-TG-AA-----G--G-GAAACAG-CCA-TGTCA----TCCCA-GT---C-GG---T-AC-ATACT.............   7704
MAC.US.SMM239      AT-AT-GGCCA-AG--G---TGG---TGGT--...G---G---------GG-TG-AA-----G--G-GAAGCAG-CCA-TGTCA----TCCCA-GT---C-GG---T-AC-ATACT.............   7722
MAC.US.SMM239 Env  N__D__R__P__K__Q__A__W__C__W__F__.__G__G__K__W__K__D__A__I__K__E__V__K__Q__T__I__V__K__H__P__R__Y__T__G__T__N__N__T__.__.__.__.__.__.  gp120

STM.US.STM         AT-AG-GACCA-A------TTGG--CTGGT--...G---G-G------GAGGAG-AA-----G--G-TAAGGAA-CC--GGTCA----TCCCA-GT---CAGG---C-ATGA-ACG.............   7361
SMM.US.SMM9        AT-AG-GACCA-A---G--CTGG--CTGGT--...G---G-G-------AG-GG-CA-CCGGG-AG-GAAGGMAMCC--GGTCA----TCCCA-GT----GGG---T-ATRAKAYT.............   7199

H2A.GW.ALI         ATAA--GACCC---------TGG--CTGGT-C...-A--GCG------GG--AG-CA-GC-GG--G-GAAGGAA-CCC-TGT-A----TCCCA-GT----AGG---C-ATGA-ACA.............   7785
H2A.SN.ST          ATA---GACCC--G------TGG--CTGGT-C...-A--GCG-G-----GG-AG-CA-G--GG--G-GAAGCTA-CCC-TGC-A----TCCCA-GT----AGG---C-ACGA-ACA.............   7231
H2A.DE.BEN         ACAC--GGCCT--G--G---TGG--CCGGT--...G---GC-G-----GGG-AG-CA-GC-GG--G-GAAGCAA-CCC-TGT-C----TCCCA--T-C--AGG--TC-ATGATACA.............   7804
H2B.GH.D205        ATAAG-GACCC--------TTGG--CTGGT--...-AG-G---C----CAG-AG-CA----GG--G-GAAA--G-CCA-C-T-A----TCCCA-GT----AGG-GGTGC--AAAAT...ATCACA.....   7777
H2B.CI.EHO         ATAA--GACCT-A------TTGG--CTGGT--...-A--G---C----CAG-AG-CA--C-GG--G-GAAAGAG-CCA-T-A-A-T--TCCCA--T--TCAGG-----C--ATATC.............   7763
H2G.CI.ABT96       AT-A--GACCC-----G--TTGG---TGGT--...G---GC--------GG-AG-CA-GC-GG--G--AAGGAA-CCG-GGTCA-G--CCCCA-CT----AGG-----ATGA-ACA.............   7135

TANTALUS.UG.TAN1   ATAC---AT-A---------TGG--CC-TT-C...CAG-GTG-T-----AGGAG-C-GG-GGG-AG-GAGA-AA-C-A--GTG--G-TTCCAAA-G-G--AT-TCG-GGG-CAAAT.............   7376
GRIVET.ET.GRI677   ATACC-GGT-A-A------GTGG--CC--T-C...CA--G-G-T-----AGGGG-A-GG---G-AG-CAGAGAAG--G---AGA--GTGAAAAATCT--CAG--GT-...-G-ATA.............   7114
VERVET.KE.VER9063  ATACG--GC-T-----G--CTGG---C-TT-C...CA--GC-GG----GAG-AG-C-GG--GG--G-TAAAGAAG--A---TCA--TT-CCAAA-G---GGT-TCGGGG--C-AAT.............   7416
VERVET.KE.VERTYO   ACATG--GC-T-----G--TTGG---C--T-C...GA--GC--T----GAGGTG-C-GGCGGG-AG--AAACAA---A--GT---GTT-CCAAA-G-C-GGT-T-A-GG--C-AAT.............   6888
VERVET.KE.VER155   ATAC--GAC-A--G--G--CTGG--CC--T-C...CAG-GC--T-----AGGAG---GG--GG-AG--CAAGAGG--A--GT-A--TT-CCAAA-G-ACGGT-CCA-GGC-C-AAT.............   7396
VERVET.DE.VERAGM3  ATAC--GGC----------TTGG---C--T-C...CAG-GC--C----GAGGAG-C-GG---G-AG--AAA-ATG--A--GT-A--TT-CCAAA-G---GAT-CCA-GG--C-AAT.............   6914
SABAEUS.SN.SAB1C   ACACC-GGT---A------TTGG--CTGGT--...G---GT--T----GAGGAG-C-GG--GG-AG-CAAAGAA-C-A--GTGAG--T-CCACC---A--GT-T-GTGGC-CAAAT.............   7587

SUN.GA.SIVSUN      GTTCT--GT-A----TG-----G-----AT--...GT--G-C-----GGA-GAG-C--TGCCACCT---GG-AACT---G--GC-GGTGGAACCTTTAGC---C-TG-GT-ATGGCTGT...ACATTT..   7567
LHOEST.KE.LHOEST   GCA-C-GAT-A---TT------G-----AT--...GA--GCC-----GGA-GA-TG--T--TA-TC---GC-AA-TGC---AG--GTTGAA--C-G-AGCC-TG-ATTAT-CAGAAGGGACAGGGACATG   7513
LHOEST.CD.447      GTA-C-GGC-----TTG--T--G-----AT--...GA--G-C-G---GGA-GA-TG--TC-TA-TT---GA-A--TGC-T-A-AGG-TTAA--C--CAGCG-TG-ATTAT-C-CATGGAAAC......TG   6438
LHOEST.CD.485      GTA-C-GGC-----TTG--T--G-----AT--...GA--G-C-----GGA-GG-TG--T--TA-TT---GG-AA-TGC-TCA-A-G-TTAA--C--CAGCA-TG-ATTAT-CAAAT...GGACAA...TG   6435
LHOEST.CD.524      GGTCT-GAC-A---TTG-----A--C--AT--...GAG-GTC-----GGA-GA-TG--T--CA-TT---GA--G-TG----AGA-G-TTAA--AG-C-GC--TG-ATTAT-AAGAA...GGAAAT...TG   6441
MANDRIL.GA.MNDGB1  GAAAG---T-A-AGA---GCATG--C-C-T-C...-A--G-CGT---GGGTTAG-AC-TTGGAGTC------AAG--C---ATA--TT-AA--ACTCC-TC---GTG-ACCAGACCTGTAAAAATTTTAC   6847

COLOBUS.CM.CGU1    ATTC---AGGC-TTT-T-G-AGG---C-GT--......-T-CC-AT--C-GGGCAGAAT--GA-A-ACAAA-CAC----CGC-AT-G--A-CAA---G---TT----GCTTGGCTAGAGCGCATTTCACG   6725
SYKES.KE.173       GGA-GG-CGCA-A--GT--CTTC---CGAG-A...-C--G---T----CAG-ATT---T--G---G--CA-GAAC-GGCC-CGA--ACC...T-G--A---GT---C-AC...................   7286
RCM.NG.NGM         AAAA--GGCCA-A-GG---TTGG----GAT--...GAT-G-G-T-----AGGTG-G-GG---G--G--AAAGAGGC-A-T-ACA-GAC-AAA---T----AGGT---CG-CAGCCT.............   7908
MANDRIL.GA.14CG    GCA-G---C---A-A---G-ATG--C-GGT-A...-T--G-C-----GGA-GAG----G--CGCAT--T-AC-TG------AGA--GT-AA--C--G--T-T-T-G-GGTT-TAAT...........GC   7328
MANDRIL.14CG  Env  G__R__N__L__K__K__G__M__C__R__L__.__I__G__Q__W__G__R__A__L__N__G__L__S__R__E__L__K__K__V__N__A__S__I__Y__R__G__F__N__.__.__.__.__A  gp120
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
H1B.FR.HXB2   Env __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__F__K__Q__S__S__G__G__D__P__E__I__V__T__H__S__F__N__C_  gp120
H1B.FR.HXB2       ...............................................................................TTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGT   7358

CPZ.US.CPZUS      T.....................................................................................G-------A---------G-G----ATATG-TGC--------   7387
CPZ.GA.CPZGAB     C.....................................................................................--T-----A--T-----GG-CAC-CAT--T-TG---------   7419
CPZ.CM.CAM3       T.....................................................................................C-------A---------G-AAC--ATATG-TG---------   6786
CPZ.CM.CAM5       C.....................................................................................C-C-----A--T---------AC--ATATG-TG---------   7062
CPZ.CD.CPZANT     ......................................................................GCGAAAACAA-ATG-AC--T-CA--AT--A--T--T---G-AAA-GT---TT-G--------C   6820

H1A.UG.U455       ......................................................................ATAATC---GCTAGC--------------TAT-------AC---A--T--C---------   6803
H1C.ET.ETH2220    ......................................................................ATAGAA-------C----------------T-------AC---A--T--C---------   6755
H1D.ZR.84ZR085    ......................................................................ATAATT-----A-C---------------T---------AC---A---------------   6894
H1F.BE.VI850      ......................................................................ATAAAA-----C-------------------T-------AC--T---T------------   6677
H1G.SE.SE6165     ......................................................................ATAACC-----CTC---TG------------T-------AC---A--T-----C-T----   6788
H1H.CF.90CF056    ......................................................................ATAAGC-------C-AA-------------ATG---G--AG---A--T------------   6683
H1J.SE.SE92809    ......................................................................ATAAAC----CATC-C-------------TAT---------G--A--T------------   6662
H1K.CM.MP535C     ......................................................................ACATTTAAACCA-C-AA-C-------A------------C---------TG---------   6563
H101_AE.TH.CM240  ......................................................................ATAATC---C-A-C-C----------A--T-T-------AC--T---TCA----------   6929
H1N.CM.YBF30      ....................................................................ATAACCTTCAGGGC-CGAG-GAGGAAT-A---A----T----G-GAC-CACTTA-TG--C------   6939
H1O.CM.MVP5180    ....................................................................GTTACCATAATA--C-GCAG-A-TAGT--T--A--TG-----G-AAGCCATTTACA------C---   7408
H1O.BE.ANT70      ....................................................................ATTAACATGACA--C--T--CAG-AGC--T--A--T-T---GG-AACCCATTTACAC-----C---   7399

MAC.US.SMM251     ....................................................................GATAAAATCAAT--A-C-GCTC-TGG------A--T--G---G--ACCTTCATGT-GACA-----C   7770
MAC.US.SMM239     ....................................................................GATAAAATCAAT--G-C-GCTC-TGG------A--T--G---G--ACCTTCATGT-GACA-----C   7788
MAC.US.SMM239 Env __.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__D__K__I__N__L__T__G__P__G__G__G__D__P__E__V__G__F__M__W__G__N__C_  gp120

STM.US.STM        ....................................................................GCAAAAATAAGGA-AGT-GCTC-TGGG-----A--T-----GG-CACCTTCATGT-GACA-----C   7427
SMM.US.SMM9       ....................................................................RAACAAATTAAGC-A-CAGCTC-AGG------A--T------G--ACTTTCATGT-GACA------   7265

H2A.GW.ALI        ..................................................................AACCAAATTAACTTTACA--A-C-GGAAG---CTCA--TG-----G-G---TATATGT-GAC---C--C   7854
H2A.SN.ST         ..................................................................GAAAAAATTCGTTTTA-AGC--C-GGAGA-C-CTCA---------G-G-C-TACATGT-GAC---C--C   7300
H2A.DE.BEN        ..................................................................GGGAAAATTAACTTTACG--A-CGGGAG----CTCA-----G---G-G-C-TTTATGT-GAC---C--C   7873
H2B.GH.D205       ..................................................................AGCGTAAAGTTAGTA-CAG-A--TGGAAA---TTCA--T-------CAAC-TATATGT-GACC------   7846
H2B.CI.EHO        ..................................................................TCACAGATAAGGTTAGCAG----CG-GAG-A-CTCA--T------G-AAG-TATATGT-GAC---C---   7832
H2G.CI.ABT96      ..................................................................AAGAAMATAACCTTTACA-CA-C-GGAGA---TTCA--T------G-AAAGTTCATGT-GAC------C   7204

TANTALUS.UG.TAN1  ..................................................................AACACAAGGCAGATT-GGCTAAGTAGACA-T-G--A--T------GCA-CT-ACATTT-GC-C-----C   7445
GRIVET.ET.GRI677  ....................................................................GAAAATATACATC-G-GAAGG...AT-T-G--A--T------TCA-CG-ATTTTT-G--C------   7177
VERVET.KE.VER9063 ..................................................................GATACAGAGAAAATA-A-CT---GAGACAGTTT--A--T------GCA-C--ATTTGT-G--------C   7485
VERVET.KE.VERTYO  ..................................................................AATACAGAACACATA-ACCT----AGACA-T-G--A---------GCATCC-ACTTGT-G---------   6957
VERVET.KE.VER155  ..................................................................GATACAAACAAAATC---TT----AGACA-TTT--A---------GCA-C--AT-TAT-G--C--C---   7465
VERVET.DE.VERAGM3 ..................................................................GATACTGAAGAGATT-A-CT---GAGACT-TTT--A--T------GCA-C--ATTTAT-G---------   6983
SABAEUS.SN.SAB1C  ..................................................................GACACAAACAAGATT---CTA---AGACA-T-G--A--T------TC--A-TTCTT-TT----------   7656

SUN.GA.SIVSUN     .GATAATATAACAAAGACATGCAGATTTACC...AATGGTACAGATTTTAAAAAGATGATTAAGTTTAAA......CCA-T-G-GAAGCAT----CA--TG---C--CA-AG-T-TTG-TGA-G-C----   7687
LHOEST.KE.LHOEST  TGATTCTAAGAAAACAACTTGT.......................GGACGGAAATTAAAAGGCTTACCAATAGCTAA--T-ACTAGGCAT----CA---TT--C--CA-AG-T-TTA-TGCA--C----   7619
LHOEST.CD.447     TATAGATAACACAAAGCCTTGT.......................GGTAGACAACTAAAGGGATTACCTATAGCAAA--T-AC-AGAAG---G--AA-T-TT-C--CA-AG-T-TTG-TGCA--CA---   6544
LHOEST.CD.485     TCCTGAGGGAAAACAGCCTTGT.......................GGTAGACATCTAAAGGGATTACCTATAGCAAA--T-AC-AGACGT--G-AA-GT-TT-C--CA-AG-T-TTG-TGCA--CA---   6541
LHOEST.CD.524     TTCCACACCAGGAAAACCATGT........................GGTAGACAATTAAAGGGTCTACCCATAGCAAA--T-ACGAGAAG-----TA--T-TC-C--CA-AG-T--T--TGCAC-C----   6547
MANDRIL.GA.MNDGB1 TAGCACTGGAGAGGAGAACAAACAAAACACGGACAAGCAAAAGGAGTTTGCCAAATGCATAAAGACTCTTAAGATAGA---TT-TA-TA--TC---A--TAG--C-GCA-A--T-ATG-TGA-G-CA---   6977

COLOBUS.CM.CGU1   CAAAAATATA.......................................................ACTATCACTCCAAGG--TGG-AATAG-AC-A-T--T-----GGC-ACTTTCAC-TT-G-G-T----   6801
SYKES.KE.173      ....................................................................ACCACA-GG-G-TCCCAGC----G--A--T-T---GG-ACG---A--TT-G--CC-A---   7346
RCM.NG.NGM        ....................................................................GAGAATATAACTA-A-GATC-GT--AT-----A--TGAT--GGCAAA-TGTTTTT-G--A------   7974
MANDRIL.GA.14CG   AAGTGGCCCTTGCAGGGGGATAAACAATAGAGGAGTGGTAAACAAAACAGGGTGTGCCTTGAAAACAATAGAAGTAAG---CT-CA--A-CATT--C--A----G-GCA-A---AATT-TGA-CCTC---   7458
MANDRIL.14CG  Env __S__G__P__C__R__G__I__N__N__R__G__V__V__N__K__T__G__C__A__L__K__T__I__E__V__S__N__Y__T__T__I__G__E__P__G__A__E__T__I__M__I__L__C_  gp120
```

```
H1B.FR.HXB2   Env _G__G__E__F__F__Y__C__N__S__T__Q__L__F__N__S__T__W__F__N__S__T__W__S__T__E__G__S__N__N__T__E__G__S__D__T__.__.__.__.__.__.__.__.__  gp120
H1B.FR.HXB2       GGAGGGGAATTTTTCTACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACA.......................  7463

CPZ.US.CPZUS      -----A--------T--T------A---TT-CC--A-----C-TG--C-ACAAC............-A---C--CAACA-C-C--T-......................................  7462
CPZ.GA.CPZGAB     -----A-----------T------A-T-GC---A-A---........---GACAAC...............-T-AC-AATGGC-T--TA.......................................  7485
CPZ.CM.CAM3       -----A-----C--T--T--C---A-T--TAC-T--------TA---G---AAA--C-AC-A-ATA-C--AG-G--ATAAC-CC-C-.........................................  6873
CPZ.CM.CAM5       -----A-----C--T--T------A-C--T-C---C--C----AGTT-AAT-A--C-T--GA-AACCAA-ACA-CT-CA-C-GA--T-ACAC-AATGACACA-ACTCAGACAATAGTACACTA.......  7185
CPZ.CD.CPZANT     CA---A-----C--T--T---G--AT-----CTTG---C---GCC--A-ACACG......GGAAACCTC-TCAC-AACGG-GCCCT-......................................  6901

H1A.UG.U455       -----A-----------T--C--CA--T--GGC-----------C-T----AA-GGC--C-TG-CA-A-GACATG---C-----GG----.....................................  6893
H1C.ET.ETH2220    -----A-----------T--C---A--T--A-C-----------C--AAAAC-AG-ACTGTT-AAT--C-G-AC-AAT-T---C.........................................  6839
H1D.ZR.84ZR085    --------------------C---A--T--GG-------------G-A---AA--TATCAGGACAT-----A-GGTTAAATG---CA-TC....................................  6984
H1F.BE.VI850      A----A--------------CG--A--T--GG----------GAC--AG-C-CC----A-GGCACT.....................................................  6743
H1G.SE.SE6165     A----A--G--------T------A--T--GGC---------TCA-GCCT---G-GATCA-A-A-T---GAGA-T---A-T.....................................  6869
H1H.CF.90CF056    A----A-----------T--C---A--T--GGG-------------G----GAA-TGCA----AATTAC--ATC-AATGAC-CA--GGGAA-C-A--AC...............  6782
H1J.SE.SE92809    -----A--------A--T------A--T--A-------C-------G----GA---G-A--GCATTGAAG--AC-AATGAC-CA-GTNAC-C-ACT...............  6758
H1K.CM.MP535C     -C---A--------------C---A-----A-----------GAG--AG--GAG---G-----.....................................................  6626
H101_AE.TH.CM240  A----------------T--C---A-----A------------A------CC-AGGA-A-GAAACC-TGG-G-GGT-TAATG-C-CT.....................  7016
H1N.CM.YBF30      A----A--G--------T-----CA-TT-CA--T-A------GAGGAA-TAC----CGAG--AG-TGAGC--.....................................  7011
H1O.CM.MVP5180    CAT--A-----C--T--T-----CA--T-TGGGA-------CTA-----TTA-C--CT----AAA-TCCGGATGCCA-GAG-TC--AGGGAGCAATGAGAC--ATAAAAATGGT.....  7522
H1O.BE.ANT70      CAT--A-----C--T--T-----CA--G-TA-GA--------TA---C-TT-CA...T---ACG-A-CC--CTGTA-TGTT-G---TGT-AGT......C-AGGTAACAATGGC.....  7504

MAC.US.SMM251     A----A--G--CC----------AATG-ATTGGT-TC-A---T-GGTAGA-GA--GGGA-GTAACT-CC.....................................CAGAGGCCAAAGGAACGGCATAGAA  7864
MAC.US.SMM239     A----A--G--CC----------AATG-ATTGGT-TC-A---T-GGTAGAAGA--GG-A---AGCT-AC.....................................CAGAAGCCAAAGGAACAGCATAAAA  7882
MAC.US.SMM239 Env _R__G__E__F__L__Y__C__K__M__N__W__F__L__N__W__V__E__D__R__N__T__A__N__.__.__.__.__.__.__.__.__.__.__.__.__.__Q__K__P__K__E__Q__H__K__  gp120

STM.US.STM        C-T--A------C-T--T--C--AATG-ATTGGT-C--A---T-G-TAGAAAAC-GA-----A-CAGAA.....................ATGA--GA-TGG-ACAAAAACAAAAAGGAACAACAAAAGA  7536
SMM.US.SMM9       C----A-----C--A--T--C--AATG-ATTGGT-TC-----T-GGTAGAAGY--TA-AA-A-G-TTC-.......................A--T-GACA-GTCAAAACCAGAAAGAACGATACCAGA  7371

H2A.GW.ALI        A----A------C--C-T--C--CATG--TTGGT-CC-C---T-GGTGGAAAAC--A-CGGG-CA-GAACAG.....................................................C  7927
H2A.SN.ST         A----A------C-------C---ATG--TTGGT-CC-C---T-GGTAGAAAAC-GA-CG-A-CA--CACAG.....................................................C  7373
H2A.DE.BEN        A----A------C----------CATG--TTGGT-CC-C---T-GGTAGAAGAC--G-ACCAAACAC-GCGC.....................................................  7945
H2B.GH.D205       A----A-----C--A---------ATG--CTGGT-C--A--CT-GGTAGA-AAC--A-CA-A-ACA-C-CGG.....................................................A  7919
H2B.CI.EHO        A-G--A--G---C-----------ATG--TTTTT-C--A--CT-GGTAGAAAA--GA-C-GGGCTC-AG-GA.....................................................  7904
H2G.CI.ABT96      A----A-----C--G--T-----AATG--TTGGT-TC-A---T-G-TAGA-GAC-GA-A--TGACTCTCCTA.................................AAGCCCCAGGAACGACAAAGAC  7298

TANTALUS.UG.TAN1  CA---A--G--------T----C-C-TGATTGGT-TG-----T-GCTGAATAA-G-GTC--A-A-TG-A-GAA-T-TAGATGTAG-AGG-A-TAATT-CAC---TGGAAAGGATAAACCGTGCTACAAAA  7575
GRIVET.ET.GRI677  CA---T-----------T-----GATGGACTGGT-TA-C---TA-CTAAACAA-CGA-CAGAAGATGCAGAA...--TA-T----GG--CTGT-AC-AA-GG-AGCCAGGACCAGGACCATGTGTTCAGA  7304
VERVET.KE.VER9063 CAG--A--G--C--------C--GATGGACTGGT-TC-A---TA-TTGAATAA-C-ATCAGTAGATGCA-A-C-TAATCA-TG--GT-A----AC--AAA-AGGGAATGCCCCAGGACCATGTGTACAAA  7615
VERVET.KE.VERTYO  CA---A-----C--T--T-----GATGGATTGGT-T--A---TACTTAAATAA---A-CATGGGATGCATACC-TAAT-TTTG--G--GCA--AAG...A-AGG-CACGCACCAGGACCATGTGTACAAA  7084
VERVET.KE.VER155  CA---------C-----------AATGGACTGGT-T--A---TA-CTGAATAA-TTA-CAGTGGATGC-GA-C-TAATCATTG---A-ACA-C-C-G-GA-AGGTCGAAGTCCAGGTCCCTGTGTACAGA  7595
VERVET.DE.VERAGM3 CAG--------C-----T-----AATGGATTGGT-TC-A---TACCTGAATAA-CG--CAGTAGATCCGGACC-TAATC-GTG---TGG-ACGAAGG-AA-AGGTAAGGCACCAGGACCCTGTGCACAAA  7113
SABAEUS.SN.SAB1C  CA---T-----C-----T-----AATGGATTGGT-CC-C--CTA-TTGAATAA---GTCAGTAGACCCA...............G--CA-A-CAACT---CA-A-AATAATACAAAGCCATGTTGGCAGA  7771

SUN.GA.SIVSUN     ----AA--GA-G----TT-------T----AG-A-T--C--AGTATGGAATGAC-CA-CCT-CAAT-AATGG.....................................................T  7760
LHOEST.KE.LHOEST  ----A----A-G----TT------GT---TAGGA-A--CC-AGAATGGAACAA---A-A-T-AGAT-AATGG.....................................................T  7692
LHOEST.CD.447     ----AA---A-G----T---C---GT----AG-A-A--CC-GGAATGGAATAA---A-ACT-AGAT-AATGG.....................................................T  6617
LHOEST.CD.485     ----A----A-G--T-T---C---AT----AG-A-AC-CC-AGAATGGAATAA---A-ACT-AGAT-AATGG.....................................................T  6614
LHOEST.CD.524     --T-A----A-G----TT------GTC---AG-A-C--CC-GGAATGGAACAA---A-ACT-AGAT-AATGG.....................................................T  6620
MANDRIL.GA.MNDGB1 CA---T---A-G----T-------GT----AG-A-CA-G-GGGCATGGAATGA-CC--A-GAGAA--AGTGG.....................................................T  7050

COLOBUS.CM.CGU1   CATA--TT---C--T---------G--T-TAGTT-A-GG--ACA-GA--CC.......................................................................  6852
SYKES.KE.173      --G--A--G--C-----------GT-T-TA-G--C---GC--AC-T-AC-AA-GGA-ACG-AA-T-AG-A-A-T.............................TATGCCAGTA  7431
RCM.NG.NGM        -AT--A------C-T--T--C--AATG-ATTGGT-TC-A--CTA-CTAAATAA---A-CAGAAG-C-A-.....................................AAAAATGAAGCTAGGCAAG  8062
MANDRIL.GA.14CG   -----A----A-----TT-------GG---AG-A-T-GG--A-CCTGGAATAACC-A-ACT-AAATGT-TGG.....................................................T  7531
MANDRIL.14CG  Env _G__G__E__Y__F__F__C__N__W__T__R__I__W__K__T__W__N__N__Q__N__S__N__V__W__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__  gp120
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
H1B.FR.HXB2   Env .__I__T__L__P__C__R__I__K__Q__I__I__N__M__W__Q__K__V__G__K__A__M__Y__A__P__P__I__S__G__.__.__.__Q__I__R__C__S__S__N__I__T__G__L__L  gp120
H1B.FR.HXB2       ..ATCACCCTCCCATGCAGAATAAAACAAATTATAAACATGTGGCAGAAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGA........CAAATTAGATGTTCATCAAATATTACAGGGCTGCT   7582

CPZ.US.CPZUS      .....--A--TAAG--T-------G-------G----TCA----AT--G-----------G---C-T------A--A----AA--T.........GTGC-A--T---AAC--------A--G--AA--A-   7578
CPZ.GA.CPZGAB     .....-TA--G-----------T-G---G--AG-T-GTTCA---AT--GG-------G--G---A-----A--C--A-----A----------.A-C--C-CT--CAAT--T--C-----T--A--T--   7601
CPZ.CM.CAM3       .....TT-T-------T-------GG------G-----TCC---AT--G---G--G--G-G---T--------A--A--T--A----------.GTT--A--T--CA-------CG----T---A-TA-   6989
CPZ.CM.CAM5       .....--AA-AT----T-------GG------G-----TCC---AT--G---G--G----G---C-------A--A--A--A--G.........A-C--A-C---CAT---------A--T--A--C--   7301
CPZ.CD.CPZANT     .....-TAGCA-AT--------T--G--G--AG-T--TCAT---GGC-T---TTC-----GC--T--CTTAG-C--T-GG--A--G.........A-TG--TCC---A-T--C-GC--A--T--AA-TA-   7017

H1A.UG.U455       ..--A--T----A---T--------G-----------T----------G-------C----------------------CAA----------.GT---A--G---GA------C--------A--A--   7012
H1C.ET.ETH2220    ..-----A----A-----------------------------------GG-------CG---------------------TGAA----------.AT---A-TG---AG---------C-----A--A--   6958
H1D.ZR.84ZR085    ..--A---A-----------------------------------G----------------------------TGAA--G.........-------AT----------------------A--   7103
H1F.BE.VI850      ..-----T--------TC-G-----------G--------------GG-------CG----------A--T-------TGCA----------.A-C----CC---AAC--------------TT-A--   6862
H1G.SE.SE6165     ..-----A--------T-AG-----------G-G-GA----------GG--G---C----------------------GCA----------.A-C---GA----AAC----GC--------A--AA-   6988
H1H.CF.90CF056    ..--T--A--G---------------------G--------------G-------CG----------------------CAA----------.A-C----TG---GT--------------A--AA-   6901
H1J.SE.SE92809    ..-----AA-A-----T-A-------------G-G-GA--------A-GGAC----C-------C--------------TGCA----------.A-C----CG--CA----------------AT-A--   6877
H1K.CM.MP535C     ..-----A--------T--------G--G----------------------G-----------A--------------TGCA----------.AGT----AC---AGC--------------AA--A-   6745
H101_AE.TH.CM240  ..-----A--T-------AG-----G---------------------GG--C----C-------------T-----------------------.AG-----AT---GT--------------AA-A--   7135
H1N.CM.YBF30      ..--T--T--G--T--T-------G---G---G----TT-----ACA-GG----------G---T-----A--A--A--TC-G----------.GTTC---AC---A-C--C--------T--A---G-   7130
H1O.CM.MVP5180    .....--TA-A--T----AGT---G---GC-AG---GATCA---AT---G-G--AGTCGAG---C-----A---------CCC--C.........A-CT-A-C----CAT--C--C--A--T--AA--A-   7638
H1O.BE.ANT70      .....--T--A--T----A-C-G-G---GG-GG---GGTCA---ATA-GG-G-CAGTCG-G-C-C-----A----------AA--T.........A-TC-A-C----ATG-----C--A--T--AA--A-   7620

MAC.US.SMM251     GG-ATTA-G-G--G--TCAT--T-G------A--C----CT-----T--------C---AATG-T---TTG-----A-GAGAG----------.G-CC-C-CG---AAC--C-CAG-G--CA-T--CA-   7985
MAC.US.SMM239     GG-ATTA-G-G-----TCAT--T-G------A--C----CT-----T--------C---AATG-T---TTG-----A-GAGAG----------.G-CC-C-CG---AAC--C-CAG-G--CA-T--CA-   8003
MAC.US.SMM239 Env R__N__Y__V__P__C__H__I__R__Q__I__I__N__G__W__H__K__V__G__K__N__V__Y__L__P__P__R__E__G__.__.__.__D__L__T__C__N__S__T__V__T__S__L__I  gp120

STM.US.STM        GG-ATTATG-A------CAC--C-G----G-A--T--T-CA-----T--------G---AATG-T---TTG-----A-GACAA----------.G-CC---CT---AAT--C-C-G-G--CA-TA-AA-   7657
SMM.US.SMM9       AA-ATTATG-G-----TCAT--TRG---G--A--C----C------C--------C---AATG---R-TTG-----T-GGGAA----------.G-CC-G-C----AAC--C-C-G-A--TA-C--CA-   7492

H2A.GW.ALI        AC-ATTATGCA--G---CAT-----G-----A--T--T--C-----C----C---G---AATG-A---TTG-----T-GGGAA----------.G-GT-G-CC--CAAC----CAG-A--CA-CT--A-   8048
H2A.SN.ST         AC-ATTATG-G------CAT-----G-----A--T--T-CC-----C--G-----G---AATG-A---TTG-----T-GGGAA----------.--GT-A-CC--CAAC--T-CAG-G--CA-CA-AA-   7494
H2A.DE.BEN        ..-A-TAT.......---CAT-----G--G--A--T--T-CC-----T--------G---AATG-A---TTG-----T-GGGAA--G.........G-GT-GGCC---GA-----CAG-A--CA-CA-AA-   8058
H2B.GH.D205       GA-ATTATGCG-----TCAT--C-G---G--A--C----C------C--G--T--G---AAT--A---TTG-----A-GAGAA--T.........G-GC-CTC----AAT--C-C-G-G--TA-C--CA-   8040
H2B.CI.EHO        ..-ATTATGCGT-----CAC--C-G---G--AG-C----C------C--GA-T----G-AATG-----TTG-----A-GAGAG--T.........G--C-CTCC---AAT--C-C-G----CA-C--CA-   8023
H2G.CI.ABT96      GC-ATTATG-A--C---CAT--T-GG--G--A--T----CA-----C------------AATG----CTTG--C----GAGAA--T.........G-TC-G-CG---AAT----CAG-C--CA-C--AA-   7419

TANTALUS.UG.TAN1  GG-CATATG----C---CAT--T-GGTC----G-C--TGAC---T-C-CGC-CA-T--GAA--CC-----A--A--A-GGGAA----------.--CT-AGA---CA------CAG-G---TCTA--A-   7696
GRIVET.ET.GRI677  GA-CTTATG-TG-C---CAT---CG----G-AG----TGAT---T-C-CT--CTCT---AAGG-A-----T--A--A-GGGAA--T.........--TT-GGAG---AAC---TCAG-C--G-CA--ATA   7425
VERVET.KE.VER9063 GG-CTTATG-TG-T---CAC--C-G-TCTG-C-----TGAT---T-T-CCA--TC---GAA-G-------T--A--A-GGGAA----------.--TC-GGA---CA----C-CGG-C--T--TA--AC   7736
VERVET.KE.VERTYO  GG-CGTATG-TG-T-A-CAT--C-GGTCTG-A-----TGAT-CCT-T-CCC--TC---GAA--CT-----T--G--A-GAGAA----------.--TT-GCA---CAGG--C-CAG-C--T---A--AC   7205
VERVET.KE.VER155  GA-CTTATG-TG-C---CAT--CCG-TCTG-C-----TGAT---T-T-CTA--TC---GAA--CA-----T--A--A-GAGAA----------.--TT-GCAG--CA-G--C-CAG----T---A--AC   7716
VERVET.DE.VERAGM3 GA-CATATG-TG-T---CAT---CG-TCTG-C--T--TGAT---T-C-C-C--TC--GGAA--CC-----A--G--A-GAGAA--G.........--CT-GCA---CA----C-CGG-A--G--TA--TC   7234
SABAEUS.SN.SAB1C  GA-C-TATG-T--C---CAC----GG---G-AG----TGAT---T-C-CCT-GTCT---AAG-CA--C----------GAGAA--C.........--CT-GGAG---AAC--G-CAGCC----C----TA   7892

SUN.GA.SIVSUN     ATCCTTGGGC-AAT---CAC------AGTG-G---G-TGAT---GC-TC---G--C--GAAG--T--CTTG--C--A-CATCA--GTTTAATAACAG---A------G--AAC-GAG-G----AAGCATG   7890
LHOEST.KE.LHOEST  ATCCTTGGGC-AAC---CAC------AGC--C--TG-TGAT---GCT-C-A-------GAAG--T---CTA--A--T-CTTCA--TTTCAACAATAG---C------A-TCAT-GAG-A----AAA--T-   7822
LHOEST.CD.447     ATCC-TGGGC-AAT---CAC------TC-G-----G-TGAT---GCATCT-----G--GAA---T---CT------A-CATCA--CTTTAACAACAG---A------A-TCAT-GAG-A----AAA--TG   6747
LHOEST.CD.485     ATCC-TGGGC-AAT---CAC------TC-G-C---G-TGAT---GCA-CT-----G--GAA---C---AT------A-CATCA--CTTTAATAACAG---A------A-TCAT-GAG-A----AAA--TG   6744
LHOEST.CD.524     ATCC-TGGGC-AAT---CAC------TC-------G-TGAT---GCA-CT--T--G--GAA---T---CTT-----A-CATCA---TTTAACAATAG---A------A-TCAC-GAG-G----AAA--TG   6750
MANDRIL.GA.MNDGB1 ATCCTTATGC-T----TCA---T-GG-----AG--G-TGAC---AT-C---------G-AAG--A---TTA--A--T-CATCA---TTTAATAAT--C--A--G---A--CAT-GGG-A----AAA--TA   7180

COLOBUS.CM.CGU1   ..CCAGT-A-GAAC--T-CC--T-G-A-GC-GG-G--T-GT---GTC-CTCAT-CC-GGATCC------GG--G--TCCAG-A--T..........--TC-ACA----AAC-GGG-G-AACAGCCAG-TA-   6971
SYKES.KE.173      ACC-TCGT--GT----TGCC--TCGC-----A--T--TGAT---AGAT-T---A-----TTG--A--CCTG--C--T-CGGCA--G..........--T----A----A-T--C---G-C--T-CAG-AT-   7552
RCM.NG.NGM        CC--GTTTG-T--T--T-TC-CT---ATGG--G-T--TGAT---T-T-C---GTCT-G-AA-G-A--CA-T--A--A-GGCCA-AT.........GCCT-G-----CAACG-C-C-G-G---TACT--T-   8183
MANDRIL.GA.14CG   ATCCATATA-GT----T-A---C-G---G--AG-GG-TGAT-----T-----G----GGAAG--T---ATG-----TG-GTC----TTTAATAAT--T--A-----CA-TAATG--G-C----A-A--T-   7661
MANDRIL.14CG  Env Y__P__Y__M__S__C__K__I__R__Q__I__V__D__D__W__H__K__V__G__R__K__I__Y__M__P__P__V__S__G__F__N__N__H__I__R__C__T__N__D__V__T__E__M__F  gp120
```

```
H1B.FR.HXB2   Env __L__T__R__D__G__G__N__.__.__.__.__.__S__N__N__E__S__E__I__F__R__P__G__G__G__D__M__R__D__N__W__R__S__E__L__Y__K__Y__K__V__V__K__I_  gp120
H1B.FR.HXB2       ATTAACAAGAGATGGTGGTAAT...............AGCAACAATGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATT  7697

CPZ.US.CPZUS      TC-TGAC-T-AGCATAA-CGCAGTC......AATAAC...G-T-G-AG-AATATA-CAG-G-TG---AC--------------C--CTTTA----A--A-------C----G-----G--G--C-GC--A  7699
CPZ.GA.CPZGAB     C-----TTC---CAC-CC-GTGACT.........AAC.......-----G--GTAA-CT--CA--T------AC-------A-C----AA--T-T------------GC-C-----G-----------TCG---A  7719
CPZ.CM.CAM3       -C-TGAG-C--GGCA---C-TC............AAC-A--G--TCACCAA-ATC-CGC-ATAC---AC-------A-C---GTA---TTG-----GCT------GC-------C--G-----T-GC--A  7107
CPZ.CM.CAM5       CC-CGA-GC--GA--AAAATGGGAA.........AAC.........A-TA-TATA-CAG-GTAC---TC---G-----A-----A---TTG-----GCT-------A-C-GG-----------T-GC--A  7413
CPZ.CD.CPZANT     G--G...GA--G-CAAATAT--.......................-----AA-T-TT-AAG-GTC-----CT-C-A---TAGCA..---C-A------GCG--G--G-CC-GG--CC-G--G---G-G---  7120

H1A.UG.U455       ---------------G----C-AAC................--T-CAA-AAAT----C-------------------------------------A---------------G------------------  7127
H1C.ET.ETH2220    -C-G---C-T-----A-CC--A............GAGCCAC-T-GCACAAAA-----A----------A-----------------------------------------------G---G-----  7076
H1D.ZR.84ZR085    ---G---------------GC-...............-AT--T-C-C--AAT--C-C---T-----------------------A-----C-------------------------------------  7218
H1F.BE.VI850      ---G-----------------................G-G-G-A-CATT----C-----------AG------A------AA------------------C-------G------------G-----  6974
H1G.SE.SE6165     -----------------G---AAT...AACAATACA-AT-CA-G----AGT-------------------------------------------------------G------AC---------C  7115
H1H.CF.90CF056    ---G----TT--C-AG-----C..............GCGTCTGCA--AAATT-T-C-------------------------------------------------G-----------------------  7016
H1J.SE.SE92809    ---G---C-T-----A--A--CAGG...GGGAATGGC--TG-G----GCA-T----C----------AC-------A------A-----------------------------------------G-G---  7004
H1K.CM.MP535C     ---G---------------.................-AT--T-C-C-TAAT----C---------------------C-----------------------C-G-----------------C-G---  6857
H101_AE.TH.CM240  ---G---------------GT-AAT................-T-CG--TAA-----C-------------------A-C--A-A-----------------------------------------C-----  7250
H1N.CM.YBF30      TC--GA-TATAG------GCC-.................G---CCA--GAAACA--AG-ATAT--CTC-------A-C---GTTA-TCTC------CAA--G--G-----G--C--------T-GC--A  7242
H1O.CM.MVP5180    TC--CAGTT----CAACCATGG...............-ATTC--CA-GTGAAA-T-CAC-T-----A-T---G----------AA--T-TA-------C-A----G--C--C--C---------C-G--A  7753
H1O.BE.ANT70      CC--CA--TG---AACACATGG............AAC----G---CA-CAAT-TA-CA--T-----AAT---G-----C----AA--T-TA-------C------G-TC--C--C----------GGG-A  7738

MAC.US.SMM251     -GC--AC-T----T-GAC-G--GGA........................AA-C-A-CTAGT-TCA-CATGA-T-C---GG--GCA--ACTG-ATC--TTG--G--GGGAG-T------T-----G-G--C  8091
MAC.US.SMM239     -GC--AC-T----T-GAT-G--GGA........................AA-C-A-CTAAT-TCA-CATGA-T-C---GG--GCA--ACTG-ATC--TTG-----GGGAG-T------T-----G-G--C  8109
MAC.US.SMM239 Env __A__N__I__D__W__I__D__G__.__.__.__.__.__.__.__.__N__Q__G__N__I__T__M__S__A__E__V__A__E__L__Y__R__L__E__L__G__D__Y__K__L__V__E__I_  gp120

STM.US.STM        -GC--AC-TT--CT-GAC---CAAT........................AAT----CTAA--TCA---C-A-T-C---GG--GCA--ACTG-ATC--TTG-----GGGAG-T--C---T-----G----A  7763
SMM.US.SMM9       -GC-GA--T----T-GATC---GGC........................AATA---C-AAT-TCA-CATGA-T-C---GG--GCA--ACTG-ATC--TTG--G--GGGAG-T--C---T-----G-G---  7598

H2A.GW.ALI        TGCT-AC-TT--CACG-A-GGC...........................AA-C---CAAAT-TTA-CTTTA-T-C---GG--GCA--ACTA-ACC--TTA-----GGGGG-T------T-----G-G--A  8151
H2A.SN.ST         TGCT-AC-TT--C--A--AG-G...........................AA-C---CAAAT-TTA-CTTTA-T-C---GG--GCA--ACTA-ACC--TTA-----GGGGG-T------T-GA--G--G-A  7597
H2A.DE.BEN        TGCT-AC-TT--CATA-A---AAAT......................CG-A-TC-T-C-AA--TTA-CTTTA-T-C---AG--GCA--ACTG-ACC--TTA---C-GGGAG-C--C---T--A--G----A  8167
H2B.GH.D205       -GCC-AC-TCA-CTCA-AC..........................A-CAGTACC-C-AAT-TTT---TGA-T-C---AG--TCA--ACTG-AT---TTG---C-GGGGG-C------T-----G----A  8143
H2B.CI.EHO        -GCC-AC-TT--CT-GATAG--AAG........................AA-CTT-CTAAT-TTA---TGA-T-C---AG--TCA--ACTG-AT-A-TTG---C-GGGGG-C--C---T-----G-G--A  8129
H2G.CI.ABT96      -GCG-AT-T---A-AGAA--GGTCTAAC.....................AA-C-T-CAAAT-TCAT-TTCA-T-C---GG--GCA--ACTG-ATC--TTA-----GGGGG-T------T-----G----A  7528

TANTALUS.UG.TAN1  GG--T--CTG---TA-AAC-GCAAA........................GAAAG--CTAATGTGA-ATTGAC--CTA--T--GA-A---TA---GCTTA------GGGA-G------GC-CA--G-----  7802
GRIVET.ET.GRI677  CG-GG---T----TA-AAC--CAAG.....................TCTGG-CCA--AAATGTGA-CCT-A-TCCTC-GG-AC-CAG--TA---GCGTAC---C-GGGAG-C------T-----G-G--A  7534
VERVET.KE.VER9063 -G--GA-CT-A--TA-AAC---AAC......................-ATAGA-C-AATGTGA-ATT-A-TCCCC-G--AGAA-GA-TC---GC-GCA-----GGGACG-------T-G---G----C  7842
VERVET.KE.VERTYO  -G-TGAGCTTA--TA-AA--G-AAA......................AA-AGA-C-AA-GTGA-ACT-A-TCCCC-G--AGAATCT-TC---GCGGC-----GGGC-G---C---T----GG----C  7311
VERVET.KE.VER155  -G--GAGCT-A-CTA-AA---CCAG......................AA-AG--CAAATGT-A-ATTGA-TCCCC-G--AGAAAC--TC---GCGGCA-----GGGC-G---C---T-G---G-G---  7822
VERVET.DE.VERAGM3 -G-GGAGCT-A--TACAA--G-AAG......................AA-AG--CTAATGT-A-ATT-A-TCCCC-G--AGAAAC--TC---GC-GCA-----GGGC-GG--C---T-----G-----  7340
SABAEUS.SN.SAB1C  TG--GA-TTGA-CTA-AAC-G-AAA......................AA-AG--C-AATGTGA-GTT-TC-CCTC-G--TGAAAG--T----GCT-A------GGGAG-T--C---T----GG----C  7998

SUN.GA.SIVSUN     G--C---CTG--AA-G-TGG-AGACTGGAAGGTAAATG-ATCA--CATC--A-TTG-AGC-TTC-AGCC-CCTACCA-C-CCTTAA-TC-A-TTGTG--CAC-GG-GCAC-T--C---C------G----  8020
LHOEST.KE.LHOEST  C--TGAG-TG--AAAAT-GG-ACCACATGAGGACTTAG-AGG---CCTAAGT...--AAAGTTCTTGCCTCC-TC-TGGGAA-CTA-TC-A-TTGTGGCA---GGG-C----------T-GA-----T-G  7949
LHOEST.CD.447     G--TCA--TG--AAAGTT-G-ACCAGAAGAAAACCTGG-AGGA---TT---T...G--AGATTCTTGCCTCCCTCCTGGGAA-CAA-TC-G-TTGTGGCAAC-GG-GC-C-C------T--A---GGC--  6874
LHOEST.CD.485     G--TGA-CTG--AAAGT-GG-GCCAGAGGAAAACCTGG-GGGA---TT---T...G--AGATTCTTGCCTCCCTC-TGGGA--CAA-TC-A-TTGTGGCAACGGG-GC-C-T------T--A---GG---  6871
LHOEST.CD.524     G--TGA--TG--AAAAT-GG-ACCACATGAAGACTTAG-AGGA---CT---T...--TAAGTTTTTGCCTCCCTC-TGGGA--CAA-TC-A-TTGTGGCAAC-GG-GC-C-T--C---T-GA-C-G-C-G  6877
MANDRIL.GA.MNDGB1 C--TGA--TGC-AAAGATAG--AGTAATGAAACAAAA...........ATGC-A--TAAATTCTTGCCTCCCA-T--A-CCTCCA-TC-A-TTGTTGC-T-TGG-GC-C-T------T----C-----A  7298

COLOBUS.CM.CGU1   TGCGTTC-TG-GCACCATAG-AGGA......GACAATGATGGT----G--GT...GCT-ATCC-G-A-C-CC-AATTT--A-CAT-CATTAA-T-CTTTG-----GGGA-G---C---T-G--------G  7092
SYKES.KE.173      GACGGAC-T---ATACTACCCAGGA........................--TAC-T-GAA-TTTA--CC-AC--C-A--G--GA---TGT-------GCA--TC-G-TC--T------T-GA-TC-----  7658
RCM.NG.NGM        -GC-GAT-TT--CTA-ACAG--AAT.......................ATGACA-CAAATGTCA-CTTGTCT-CT--AG-AG-A--T-TA---GC-GCA-----GGGC-G---C--------GG-G---  8289
MANDRIL.GA.14CG   T--TGAGGTCC-ACTA-T-G-AGGA..................--AAGA-A-CTA--AAAGTTT-TGCC-CA--AT--GG-ACAAA-TC-A-TC-C-GCA-TGGG-GCAC-T------T-G------G-G  7773
MANDRIL.14CG  Env __F__E__V__Q__L__V__E__G__.__.__.__.__.__.__K__R__Y__L__I__K__F__L__P__Q__D__E__V__Q__N__Q__F__T__A__V__G__G__H__Y__K__L__V__K__V_  gp120
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
                    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                    gp120 \/ gp41
H1B.FR.HXB2       GAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAG..............AGAGAAAAAAGAGCA...GTGGGA...ATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAA   7806
H1B.FR.HXB2   Env _E__P__L__G__V__A__P__T__K__A__K__R__R__V__V__Q__.__.__.__.__.__R__E__K__R__A__.__V__G__.__I__G__A__L__F__L__G__F__L__G__A__A__G__   gp41

CPZ.US.CPZUS      -----TA-------G-----AGGT-----C--A--GCATACA---A-A...............---------------GCCT-C--...C----T--AC-------G-----TC-T-------------   7811
CPZ.GA.CPZGAB     --G--T---TC---G-----A--A--A----GA--GCATACA---GCAAGACAGAAAGAC...----C-G---------GCCT-C--T...C-G-----AC--------------C----T----------   7843
CPZ.CM.CAM3       -----CA-------------A-GT--A--T--A--GCACACA--ATCA...............------------------GCCT-T---...C-G--T--GC----T-----------A-------------   7219
CPZ.CM.CAM5       -----CA-------------A--T-----T--A--GCACAC----GCA...................G---------GCCT-T---...C-------AC-------C-----TC-A-------------   7525
CPZ.CD.CPZANT     -RT--C--GTC------C--A--A-CA-GN--A--GCC--AAA-AA-ACAACACTCC.........---C---------GCATT...----...-----GCTG--C---T-G--TC-TC-CA-T-----T--C-   7238

H1A.UG.U455       ------C------------------G------------------G--...............................---T---...C--------A-C--------------A--------T----   7236
H1C.ET.ETH2220    A-G---C----------------T---C----A--G---------G--...............................GCA.........C-------------------------------------   7182
H1D.ZR.84ZR085    ------C-------------------------------------G-A...............................A-A----...C--------A-----------------------------   7327
H1F.BE.VI850      ------------------------------A---CA---------...............................---C-----...C----------------------TC-T--G-ACAGCA-GG   7083
H1G.SE.SE6165     A--T--C------------------G-----G------------G--...............................---T---...C-G------G-C--------------A-------------   7224
H1H.CF.90CF056    ------C-G---A--------------A---G------------G--...............................----------...--G-------CT----------------------------   7125
H1J.SE.SE92809    --G---C-----------------------A------------G--...............................----------...---------G-------------------------A---------   7113
H1K.CM.MP535C     ------C-----A------------G---C-G-----------------------------------------....A---...T--------G-----T----A-------------------   6966
H101_AE.TH.CM240  ------C-----A------------G------------------G--...........................G---..........---------A--A--T----------A-------------   7359
H1N.CM.YBF30      -----CA-------------AGGT--A--T--A---C-CACA---AGT...............................GCCT-T---...C----T--GC----T--------TC-T-----------G-   7354
H1O.CM.MVP5180    A----T--TA-T--------T--A--AATGTCA---CC-A-AA-AA-CATTCACACCCCTCAC--G------------....A-----T-G---ATGC-A---T-G---G-GC-AA-T--------T-   7877
H1O.BE.ANT70      A----T--TA-T--G--------ACGTATTGCA--GCC---CA-AAGCACTAGAACTCAT...................-----A---...T-G---ATGC-A---T-G---G-TC-AA-T--------T-   7859

MAC.US.SMM251     ACT--GA-T--CT-G--C-----AG-T-TG-----GTACACTACTGGTGGC......ACCTCA---A-T-------GGGTCT-T...GTGC----G.........---T-G--T--TC-C-C-A-G-----TT   8206
MAC.US.SMM239     ACT---A-T--CT-G--C-----AG-T-TG-----GTACACTACTGGTGGC......ACCTCA---A-T-------GGGTCT-T...GTGC----G.........---T-G--T--TC-C-C-A-G-----TT   8224
MAC.US.SMM239 Env _T__P__I__G__L__A__P__T__D__V__K__R__Y__G__G__G__G__.__.__T__S__R__N__K__R__G__V__F__.__V__L__G__.__.__F__L__G__F__L__A__T__A__G__   gp41

STM.US.STM        ACT---A-T--CT-G--C--A--G--T-TG------TACACAACT...AGT......ACCTCA---ACT--G----GGGTCT-C...GTGC----G.........---T-G--T--TC-C-CGA-------TT   7875
SMM.US.SMM9       ACT---A-T--CT-C--C-----A-GT-T---A--GTACACCACAACTGGT......GCCTCA---A-T-------GGGTCT-T...GTGC----G.........---T-G--T--TC-C-CGA-------TT   7713

H2A.GW.ALI        AC----A-T--CT-C-----T--ATCA-A--G----TACTCCTCT...ACT......CCAAGG--GA-T-------GTGTGT-C...GTGC----G.........---T-A--T--TC-C-CGA-------TT   8263
H2A.SN.ST         AC----A-T--CT-T-----T--ACCA-T---A---TACTCCTCT...GCT......CCAGTG--GA-T-------GTGTAT-C...GTGC----G.........---T-A--T--TC-CACGA--------G   7709
H2A.DE.BEN        AC----A-T--CT-C-----T--AG-TCAG-GA--GTACTCCTCA...ACT......CCAGTG--GA-C-------GTGTGT-C...GTGC----G.........---T-G--T--TC-C-CGA-------TT   8279
H2B.GH.D205       AC----A-T--CT-C-----T--AG-T-T--GA---TACTCCTCA...GTA......AAACCG--GA-T-------GTGTAA--...GTGC-----.........---T-G--A---C-T-C-ATG-----TT   8255
H2B.CI.EHO        AC----A-T--CT-T-----T--A-GTAT---A---TATTCCTCA...GTG......ACACCG--GA-T-------GTGTAC-T...GTGC----G.......--TT-G--A---C-T-CGA-G-----TT   8241
H2G.CI.ABT96      ACT---A-T--CT-T--------A--T-T-------TATTCCTCT...GTG......ACACCG-A-A-T-------GTGTAT-T...GTGC----G.......--TT-G--A--TC-T-C-ATG-----TT   7640

TANTALUS.UG.TAN1  ------A-C--TT-C--G--A--AG--AT--GA---TAT--T-G--CAACT............C-------G--G-TGCCCT-C...GTGT-G--G.........-----A--C--T-----G-----T---G   7911
GRIVET.ET.GRI677  AC----A-T--CT-T--T--T--AG-T-T--GA---TATACT-GC-CCACA...........................G-TGCCAT-C...GTGC----G.......--T--A--C-----------T--T----   7643
VERVET.KE.VER9063 AC----A-C---T-T--------G--T-T--GAC-TTATAC--GA...GGA......CAAGAC---TC----C---TCCCGT-T...GTGC----G.........-----A--C-----A-----T--T----   7954
VERVET.KE.VERTYO  AC----A-T--CT-T--------AG-A-T--G-C-TTATAC--GA...GGA......CATGAG--GC----G----TCCCGT-T...GTGC----G.........---T-A--C--------G--T--T----   7423
VERVET.KE.VER155  AC----A-T---T-T--------AG-A-TC-G-C--TACAC--GA...GGC......CAAGAG--GC-----C---TCCCGT-C...GTGC----G.........-----A--C-----------T--T--G-   7934
VERVET.DE.VERAGM3 AC----A-T--CT-C--------AG-A-T--GA--GTATAC--GA...GGT......CATGAC---AC---GC---TCCCGT-C...GTGC----G.........-----A--C-----A-----T--T--G-   7452
SABAEUS.SN.SAB1C  A-G---A-T---T-T--T--A-----A-T--GA---TACACT-GC...........CCCGAA---C-------G-TGCCAT-C...GTGC----G.........-----A--T--------------T---G   8107

SUN.GA.SIVSUN     AGG---A-T--GT-T--T--A--AG-T-AGC-CC-GTAC-CACCC.............AGG-AG-----G--G---GCACCT-TTGCCC-----.......GCTT-G-CAC-TC-CA-T--T--G--T-   8129
LHOEST.KE.LHOEST  A-T---A-T--GT-T--T--T--AG-T-AGC-C--GTAC-CACCC............AGGGGA--GC---CC------GCGCCTCTGGCTT-----.......GC-T-G--AC-TC-CA-T--T-----C-   8061
LHOEST.CD.447     AG----A-T--CT-T--T--A--AG-T-TGC-TC-GTAC-CACCC............AGTAGT-AGC-G--G------GCACC-CTGGCTC----T......GC-T-G--AC-TC-CA-T--T-----T-   6986
LHOEST.CD.485     AG----A-T--CT-T--T--A--AGCT-AGC-CCAGTAC-CACCC............AGTAGT-AGC-G--G------GCACC-CTGGCTC----T......GC-T-G--AC-TC-CA-T--T-----T-   6983
LHOEST.CD.524     AG----A-T--CT-T--T--A--ATCA-AGC-TC-GTAC-CACCC............AGGGGA--GC---C-------GCTCCACT-GCTC----T......GC-T-G--AC-TC-CA-T--T-----T-   6989
MANDRIL.GA.MNDGB1 ATG---A-T--CA-------T--AG-T-TG--A---CACACTT-A-CTGAA......CATCAT-A---G--G----G-GCA...-T-ATAC-T--T......A-------TC-GC-CTCGCTG------T   7413

COLOBUS.CM.CGU1   AGGA--ACTACTTAT-TC--A--AG-TAT---A---TC----AATGTAAATTGGCATCATGGG---C----G----GCATCT-T-C-TTCTCCATC.........T-A-CCC-----A-T-GG--C--TG   7213
SYKES.KE.173      A-----A-T--CT-T--G-----TG-TCA-CGTC-GTAT-A-T---CT.........AATACC--G--------G---GCACCATTGGCAC-T--G......-----G--A--AC-CTCG--T--G--C-   7773
RCM.NG.NGM        ATG---A-T---TAT-----A--A--T-TC-GA--GTAC-A-ACA.....................A--C---AGAG-GTACCTTTGGTGC-G--T.......--T--A--T---C-CTCG--------T-   8392
MANDRIL.GA.14CG   --T--GA-----T-C--C--T--TG-T-T-C-C---TATCATC-A-CAGAT.........GCA-AGC-G--G----G-GCA...-TCTTGC-T--.......A-G--C--CC-------CTTG-----TT   7885
MANDRIL.14CG  Env _D__P__I__G__F__A__P__T__D__V__H__R__Y__H__L__P__D__.__.__.__A__K__Q__K__R__G__A__.__V__L__L__G__.__.__M__L__G__L__L__G__L__A__G__   gp41
```

```
H1B.FR.HXB2        GCACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTG...AGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGT   7933
H1B.FR.HXB2   Env  S__T__M__G__A__A__S__M__T__L__T__V__Q__A__R__Q__L__L__S__G__I__V__Q__Q__Q__N__N__L__L__.__R__A__I__E__A__Q__Q__H__L__L__Q__L__T__V   gp41

CPZ.US.CPZUS       ----------------A---G-AGT----------------C-------------A--G--------A--------T---C----A...--A--A--A-----C---------T-A--------AT----   7938
CPZ.GA.CPZGAB      ----------------GG--G-A------------------C----G--GC----G--------------A-----T---C--T--...-AA--A--A-----T--------C--A--------AT--A-   7970
CPZ.CM.CAM3        ----------------A---G--GT----------------C-G-----------A--G--------A-----------CC------...C-A--A--A--A---------------T-----------T----   7346
CPZ.CM.CAM5        ----------------A---G--GT----------------C-G--C--G-----A--G--------A-----------CC------...C-A--A--A--A-----------A---T----------T----   7652
CPZ.CD.CPZANT      -T--A-----------G-----AG---------C----A----GA-T--GY-CCA------T--A--A------GC----C------...CAA--C--A---A-A--G--A------C-A--G---T-G--   7365

H1A.UG.U455        ----A--------G--G-----A-----------------------------------------------A-------G----C------...--------A-----T---------------A-------T--   7363
H1C.ET.ETH2220     ----------------A-----A--------------------------------G-----------A-----A-G----------...-AA-----A---------------A-------G-----G--   7309
H1D.ZR.84ZR085     ----G-----------G-----------------------------G-------------------A-----A------------...--------A-----------------------------G--   7454
H1F.BE.VI850       AGCA---------G--G--G--A-----------------------------------A--------A-----A------------...------------A--------------------G--------   7210
H1G.SE.SE6165      -------------G--G-----A----------------T------------------C--------A-----AGG---C------...--------A--------G----------------------   7351
H1H.CF.90CF056     -------------G--G-----A-------------------G-----------------------------A-G----------...--A-----AC------G-------A-------G-----G--   7252
H1J.SE.SE92809     -T--------------G-----A----------------T---------G--------C--------A-----A-G----------...-A-----A---------------------A-------T--   7240
H1K.CM.MP535C      -------------G--G-----A-----------------------------------------------A-----A-G----------...--A-----A--A--------A--------A--G-----G--   7093
H101_AE.TH.CM240   -------------G--G-----A-----------------------------------------------A-----A-G----------...--------A--------G----------------------   7486
H1N.CM.YBF30       ----------------G-----A-----------------C-GAC------A-----G--------A--------T-T-C--T--...--A--A--A-----------A---T----------T--A-   7481
H1O.CM.MVP5180     ----------------GG---CAG------------G-A--CACAGTG--C--AAG-----------A------G----CC------...--A--G--AC----C--G--A--CT--C--AGGT-AT-T--   8004
H1O.BE.ANT70       ----------------GG---CA------G--------A--CACACT--GC--AAG-----------A------G----CC----A...--A--A--AC----C--G-----AT--C--AGG--AT-TR-   7986

MAC.US.SMM251      CTG-A--------G--G--GT----------C-CT---T--C-GACT------G----G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AT------G---G--C--   8333
MAC.US.SMM239      CTG-A--------G--G--GT----------C-CT---T--C--ACT------G----G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AT------G---G--C--   8351
MAC.US.SMM239 Env  S__A__M__G__A__A__S__L__T__L__T__A__Q__S__R__T__L__L__A__G__I__V__Q__Q__Q__Q__Q__L__L__.__D__V__V__K__R__Q__Q__E__L__L__R__L__T__V   gp41

STM.US.STM         CTG-A-----------G--GC------------CT---T--C-GACT------A----G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AT------G---G--C--   8002
SMM.US.SMM9        CTG-A--------G--G--GT--------T---CT---T--C-GACT--G---G----G-----------A---C-AC-GC--T--...GAT-TGG-CA--AGA-----AG-AT------G---G--C--   7840

H2A.GW.ALI         CTG-A------A-G--AG-TT-A------T-T-CT---T-TC-GACT------G-C--G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA------G-AA------G---G--C--   8390
H2A.SN.ST          CTG-A--------G--G--CT--------T---CT---T--C-GACT------G-C--G-----------A---C-AC-GC--T--...GAC-TGG-CA--AGA-----AG-AA------G---G--C--   7836
H2A.DE.BEN         CTG-A--------GCGG--CC--------T-A-CC---T--C-GACT---C--G-C--G-----------A---C-AC-GC--T--...GAC-TAG-CA--AGA-----AG-AA------G---G--C--   8406
H2B.GH.D205        CTG-A---------A-G--TT--------T-A-CT---T--C-GACT---C--G-C--G---------------C-AC-GCCTG--...GAC-TGG-CA-AAGA------G-A-------GG--G--C--   8382
H2B.CI.EHO         CTG-A--------G--G--CT-----T--T---CT---T--C-GACT---C--G----G---------------C-AC-GC-CG--...GAC-TGG-CA-AAGA-----AG-A-------GG--G--C--   8368
H2G.CI.ABT96       CTG-A------A----G--GT--------T---CT---T--C-GACT--G---G----G-----------A---C-AC-GC--T--...GAT-TGG-CA--AGA------G-AA------G---G--C--   7767

TANTALUS.UG.TAN1   CTG-A-----T---A-AG-G-CAG--T-------C---T--CAG------C-TG-A--G---T-------------G---C------...GC---AG-----CA---G-----GA--C-AA-G-----CA-   8038
GRIVET.ET.GRI677   CTG-A-----------GG---CA--------A--C---T-TC-G--T--GC-TG----G---T-------------G--C-------...GC----G-G--ACA------A--GT-----A-G--G--CA-   7770
VERVET.KE.VER9063  CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG--T---C-TG----G---A-------------A---C------...GC----G-G-----T--------AA-----A-G--G--CA-   8081
VERVET.KE.VERTYO   CTG-A-----A-----GG-G-GC-GC--------C---T--C-G--T--GC-TG----G---C-------------G---C------...GC----G-G-----T--------GA-----A-G--G--CA-   7550
VERVET.KE.VER155   CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG--T---C-TG----G---T-------------G---C------...GC----G-G-GA--T--------GA-----A-G--G--CA-   8061
VERVET.DE.VERAGM3  CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG--T---C-TG----G---C-------------G---C------...GC----G-G-----T--------GA-----A-G--G--CA-   7579
SABAEUS.SN.SAB1C   CTG-A-----T-----GG-G-CAG-C--------C---T-TCAG--G---C-CG----G---C-------------A---C----T...GC----G-A---CA---G-----AA--C-AA-G-----TA-   8234

SUN.GA.SIVSUN      CTG-A------TT--TA--G-C--TA--A--T--C------CA-GC-G-C---CAA--G---T-------------GC-G------...GT-CT-G-----AA---G--AG-G--AC-AAG------CA-   8256
LHOEST.KE.LHOEST   CTG-A------TT--TA--G-C--TA--A--T--C------CA-GC-G-C---CAA--G---T-------------GC-G------...GT-CT-G-----AAA--G--AG-A-----AAG------CA-   8188
LHOEST.CD.447      CTG-A------TT--TA--G-C--TA--A--T--C--A---CA-GT-G--A-ACAA--G---T-------------GC-G------...GT-CT-G-----AAA--G--AG-G--C--AAG------CA-   7113
LHOEST.CD.485      CTG-A------TT--TA--G-C--TA--A--T--C------CA-GT-G--A-ACAA--G---T-------------GC-G------...GT-CT-G-----AAA--G--AG-G--C--AAG------CA-   7110
LHOEST.CD.524      CTG-A------TT--TA--G-C--TA--A--T--C------CA-GT-G--A-ACAA--G---T-------------GC-G------...GT-CT-G-----AAA--G--AG-G---C-AAG------CA-   7116
MANDRIL.GA.MNDGB1  C-G-G------T---TG--GG--G-A-----T--C--AT-TCAGTCT--GG--A----G------G-A--A--A--AC-G---T--...-A-CTC--A---CAA--GTCTG-A--C--AA-------CA-   7540

COLOBUS.CM.CGU1    CTG-A------T----G--TG--G--------A-C-----TCA-TCCC-TAATGGAA-GGCTAGTGCCAGCAGC---CGAA--T--CTG-AACTGG-C---A----GTCTGCA---C-A--G-----G--   7343
SYKES.KE.173       CTG-A-----G-GC--TG-G-CAG----A--AC-C---T-TCAGACT--G---G----G--------------C-A--G------...GA---AG-------A--G--A--C--C--AGGCT-A--T--   7900
RCM.NG.NGM         CTG-A-----------GG-G-CAG-C-----A--C---T--C-G--T--GC-TG----G---T-------------G--G------...GACATAG----ACA---G---G-A---C-AA-G-----C--   8519
MANDRIL.GA.14CG    C-G-G------T---TGG-GG--G-C-----T--C---T--CAGGCT--G---AA---G--T---G----------G-T-C------...--CCTG--A--TCA---CTCCG-GT-A--AA----A--TA-   8012
MANDRIL.14CG  Env  S__A__M__G__S__V__A__V__A__L__T__V__Q__S__Q__G__L__L__N__G__I__V__E__Q__Q__K__I__L__L__.__S__L__I__D__Q__H__S__E__L__L__K__L__T__I   gp41
```

HIV-1/HIV-2/SIV
complete genomes

```
                                                                                                                         Rev responsive element end -|
H1B.FR.HXB2   Env __W__G__I__K__Q__L__Q__A__R__I__L__A__V__E__R__Y__L__K__D__Q__Q__L__L__G__I__W__G__C__S__G__K__L__I__C__T__T__A__V__P__W__N__A__S_   gp41
H1B.FR.HXB2       CTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGT   8063

CPZ.US.CPZUS      T--------T--A-----T--A--C---G-A--T-----------------T--------G---A----A--TC-G-----C-----A------ACA-----TTAT--CA----------------A--CC   8068
CPZ.GA.CPZGAB     ------AG-A--A--A--A--A--C---T-G--T-----A-----G-----GC-------G---A-T-----CC-G-----C-----------GGCTG-C--TTAT--CA-G-----------CAACTC-   8100
CPZ.CM.CAM3       T--------T--A--------A------G-T--T-----A-----------C-GA-----G--AA----A--CC-A-----C-----A------GC------TTAT---A--------------AA--C-   7476
CPZ.CM.CAM5       T-----------A--------A------G-T--T-----A--------------------G---A----AA-TC-A-----C-----A------GC---C--TTAT--CA-G----------------C-   7782
CPZ.CD.CPZANT     ------AG-A--A--A--------------G--T--A--C--G-AG-------GA--------A--AT--A-CC-C---------TG---AC--GG-G-CC--TCA----A-G-----------AA-TCC   7495

H1A.UG.U455       ---------T--A---------------G----------------------C----------------A--A-------C-----------G-----C---------A-------C-----CT-----   7493
H1C.ET.ETH2220    ---------T-----------A-----G--------A-A------C------G---------------A-----------C-------------------C-----------------------CT-----   7439
H1D.ZR.84ZR085    ---------T--A--------------------------A-----------G----------------A--A-----------------------A-------------A---A--C-----CT-----   7584
H1F.BE.VI850      ---------T--A---------------G-------------------------------------A-----C-----C-------------------C---------AA------C-----CT-----   7340
H1G.SE.SE6165     ---------T--A---------------C----------A------------------------------A----------C-------------------C---------AA------C-----C-T----   7481
H1H.CF.90CF056    ---------T--A---------------G----C-------------------G--------------------------C-------------------C---------AA-----------CT-----   7382
H1J.SE.SE92809    ---------T------------------G-------------------------------------A-----A----------C-------------------C---------AA------C-----------   7370
H1K.CM.MP535C     ---------T--A-------G------------------A-----------------------------A----------C-------------------C---------AA------C-----CT----C   7223
H101_AE.TH.CM240  ---------T--A---------------G-------------------------------A--T----A--AC-------C------------A----C------------------C-----CT-C-C-   7616
H1N.CM.YBF30      ---------T--A-------------A-G----T---A-A-----------T-G------G--AA----AA-TC-A-----C-----A------ACA--A---TAT--CA----------------AG-C-   7611
H1O.CM.MVP5180    A-----T--T-GA--A----GA--TC-CC-G-AA--CT-A----CCCTTA--C--A----G--A-G---AAACC-A-----C--TAAA--------A--C--TTA---AT-A--AAAA-----CA-ATCA   8134
H1O.BE.ANT70      A-----T----GA--A----GA--TC-CC-G--A--CT-A----CC-TA---C--A----G--A-----AA-CC-A-----C--TAAA-----G--AG-C---TA---AT-A--AAAA------AGA-CA   8116

MAC.US.SMM251     ------A-CA---A-C------A-T--GG--ACT--CA-C--G-AG---T-------C--GGC--AG--AAATGC------A--TG-GTTT-G--AAG-C---CA----A----A--A---CCAAA-GCA   8463
MAC.US.SMM239     ------A-CA---A-C------A-T--GG--ACT--CA-C--G-AG---T-------C--GGC--AG---AATGC------A--TG-GTTT-G--AAG-C---CA----A----A--A---CCAAA-GCA   8481
MAC.US.SMM239 Env __W__G__G__K__N__L__Q__T__R__V__G__A__I__E__K__Y__L__K__D__Q__G__Q__L__N__A__W__G__C__A__F__R__Q__V__C__H__T__T__V__P__W__P__N__A_   gp41

STM.US.STM        ------A-C----A-C------A-T---G--ACT---A-C--G-A---------------GGCA-AG--AAATTC------A--TG-GTTT-G--AGG-C---CA---CA----A--A---CCAAA-GA-   8132
SMM.US.SMM9       ------A-CT---A-C------A-T---G--ACT---A-C--G-AG--------------GGC--CRN-AAATTCA-----A--TG--TTT-GG-AGG-C---CA----A-K--A--A---CCAAA-GAC   7970

H2A.GW.ALI        ------A-CG--AA-T------------G--ACT---A-C--G-A----T-------C--GGC--GG--AAATTCA-----A--TG-ATTT-G--AAG-C---CA----A----A--A---GTAAA--AC   8520
H2A.SN.ST         ------A-CA--AA-T------------G--ACT---A-C--G-A----T-------C--GGC--AA--AAATTCA-----A--TG-GTTT-G--AAG-C---CA----A----A--A---GTAAA-GAC   7966
H2A.DE.BEN        ------A-CG--AA-C------------G--ACT---A-C--G-AG---------C----GGCA-AG--AAATTCA-----A--TG-GTTT-G--AGG-C---CA----A----A--G---GTAAA-GAC   8536
H2B.GH.D205       ------G-CG--AA-C------------G--ACT--CA-C--G-A---------------GGCA-AA--AAATTCA-----A--TG--TTC-G--AGG-C--TCA---GA----A--A---CCAAA-GAA   8512
H2B.CI.EHO        ------G-CG--AA-C------------G--ACT--CA-C--G-A---T--C--A--C--GGCA-AA--AAATTCA-----A--TG--TTC-G--AGG-C--TCA---GA----A--A---GTAAA-GAA   8498
H2G.CI.ABT96      ------G-CG--AA-C------A-T---G--ACT--CA-C--G-A---T--------C--GGCA-GG--AAATTCA-----A--TG-GTTT-G--AGG-C---CA----A----A------G----CCT-   7897

TANTALUS.UG.TAN1  T-----TG-G--AA-C---A-T--CC-CG--ACT--CC-T----A-------G-------GACA-GG--AAATT-G-----A--TG-ATTC----AGG-G--TCA----A-G-----G----C-TTC-AC   8168
GRIVET.ET.GRI677  T-----TG-G--AA-C---A-T--CC-CG--ACA---C-C--G-AG------G-------GGCA-GG--AAATTCA-----A--TG-GT-G----AAG-A--TCA---CA-A-----A-----GTA--A-   7900
VERVET.KE.VER9063 T-----TG-G--AA-C---A-T--CC-CG--ACA--CC-T--G-AG------G-------GGCA-GAT--AACG-C-----G---G-AT-G----AGG-C--TCAT--CA-A--A--A---C-ATGGCAA   8211
VERVET.KE.VERTYO  T-----TG-T--AA-C---A-T--CC-CG--ACA--CC-T--G-AG------G-------GGCA-GA--AAACTCC-----G---G-AT-G----AAG-A--TCAT--CA-A---GAG---CCCTGG-CA   7680
VERVET.KE.VER155  T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-AG-----GGC------GGCA-GGT-AAACGC------G---G-GT-G----AAG-A--TCAT--AA-A--A--C----CGTGG-A-   8191
VERVET.DE.VERAGM3 T-----TG-G--AA-C---A-T--CC-CG--ACA---C-T--G-AG------G----C--GGC--GGT--AATGC------G---G-AT-G--G-AAG-C--TCAT--AA-C--A--G---C-GTGG-A-   7709
SABAEUS.SN.SAB1C  T-----TG-A--AA-C---A-T--CC-CG--ACT---C-T--G-A-------G-------GGCA-GGT--AAC--------A--TG--TTC-GG-AGG-G--TCA---GA-G--C-T------GTA--A-   8364

SUN.GA.SIVSUN     A-----AG-G---A-T-----A---C-CC-GACA--CC-A--GGAG--TG-GC-A--C---AGC--TT-A-CTTC------G--TCAAT-G----AAG-A---CAT---AA------C------TA--A-   8386
LHOEST.KE.LHOEST  A-----AG-G---A-T-----A---C-CC-GACA--CC-T--GGAG--TG-G--AC-----GC------T-CCTC------G--TCAAT-G----AGG----TCAT---AA----GAG----CATA--A-   8318
LHOEST.CD.447     ------AG-G---A-T-----A---C-CC-GACA---A-T--GGAG----------C---ACT--G--T-CTTCC-----G--TCAAT-G----AAG-A--TCAT---AA---AGAA-----CTA--A-   7243
LHOEST.CD.485     A-----AG-G---A-T-----A---C-CC-GACA---A-T--GGAG----------C---GCT--G--T-CTTCC-----G--TCAAT-G----AA--A--TCAT---AA----GAA------TA--A-   7240
LHOEST.CD.524     ------AG-G---A-T-----A---C-CC-GACA---A-T--GGAG---T-G--A--C---GCA--A--T-CCTC------G--TCAGT-G----AAG-G--TCAT---AA---A--A------TA--A-   7246
MANDRIL.GA.MNDGB1 A-----AG-A---A-TT-A---A-TC-CC-GACCAG-T----G-AT--TA-C-----C---GCTT-G---TCTCAA-----G--T--AT-GGC--AGG-G--TCAT---T----AGAG----C-AA--CA   7670

COLOBUS.CM.CGU1   ------GG-T--AA-C-------T-C--G-TGCCA--A-A---G-T--TT-GG----A---GCAAAG--T-CC-GCATT--G--TG-AAAC-TG-AA-----T-GA---AT------C-----CAAG-CC   7473
SYKES.KE.173      ------TG----AA-C--TA-C--CC-GC--ACA---T-A--G-CG--T--C-GA--C--GGCCA--T--TCA-A------A---G-ATTC--G-AG--C---CAT--A--A---A-G---G-GAAAGCA   8030
RCM.NG.NGM        ------G-CT--AA-C--------GC-TG--ACT---A-T--G-A-------GC---C---TCT--AT-AAAT-CA-TT--C--TG-AT-G-G--AAG-G--TCA---AA-G---GAG----TATA----   8649
MANDRIL.GA.14CG   ------TG-A--AA-T--T-----CC-CC--ACA--CT----GGAC--TG--GC---C---TCAAGA--A-CAG-A-----A-----ATTCTCC-AAG-G--TCAT---AA---T--G---CCAAA-GAA   8142
MANDRIL.14CG  Env __W__G__V__K__N__L__Q__A__R__L__G__A__L__E__D__Y__V__G__D__Q__S__R__L__A__V__W__G__C__S__F__S__Q__V__C__H__T__N__V__P__W__P__N__E_   gp41
```

```
H1B.FR.HXB2        _W__S__N__K__S__.__.__.__.__.__.__.__.__.__L__E__Q__I__W__N__H__T__T__W__M__E__W__D__R__E__I__N__N__Y__T__S__L__I__H__S__L__I__E__    gp41
H1B.FR.HXB2   Env TGGAGTAATAAATCT.........................CTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAG   8166

CPZ.US.CPZUS       --------C--C.......................CTCTCCTAT--TGCT------GGCA-TCTA--T---CAA--A---------A--G-A-GA--C--TT--G-TACT--TTTTAGTC-T--A---C   8174
CPZ.GA.CPZGAB      ---CC-GGG...AGC....................AATTCCACA--TG-C------GGGA-TCTA--A---CA-C-A-----T-A-TT-G-GTCT--C------G-GAA---TTTTGGTC-GT-A--G-   8206
CPZ.CM.CAM3        ------GC---T.......................ACATCCT-T--TG----C------A-TTTA--A---CAA--C-------A-AGGG-A--A-----TT--G-TG-C--TTTTAG-C-T--A---C   7582
CPZ.CM.CAM5        ------GC---C.......................ACATCCTAT--TG----C------A-TTTA--A---CAA--T-----T-A-A-GG-A--A------T--G-AG-C--TTTTAG-C-T--A--GC   7888
CPZ.CD.CPZANT      ---GTA--CTTCA-GCAAACATGTGCAAAGAACAGCAGTGATA-AC--TGT------G-AA-T-T---A---CAA--A---------TT-G-AC-G----CA---G-ACAG---T-TAATA-CT-AC--A   7625

H1A.UG.U455        ---------------...........................-A---GG-C--A------A---T-------C--C-A-----G-A--------GT-G-------G--A-----T-TCAAC-------G-   7596
H1C.ET.ETH2220     -----------G---...........................-AA--GG--------G--A---T----------C-------T----------GT-----T---GA-A-----T--AAT--GC------   7542
H1D.ZR.84ZR085     ----------G----...........................G-A--GT-T------GGGA---T----------C-------A---------G----------GG--------T--AA---G-------   7687
H1F.BE.VI850       ---------------...........................-A----G-----------A---T------------------A-AG--G----G-------T---A-A-----T--AAA--G-------   7443
H1G.SE.SE6165      ---------------...........................TATA-TG-A------G--A---T---T-----A--A-----A--G-----C------------TA-CA----T--AG-C--C------   7584
H1H.CF.90CF056     --------------A..........................-A-AGTG-A--C---G-CA---T---T--------A-----T-A-C------G----------GAGGA----T--AGG--GC------   7485
H1J.SE.SE92809     ---------------..........................TAT--GG-C------G-GA---T---------AC-A-----A--G------------------G-AA-----T--AGTC---------   7473
H1K.CM.MP535C      ------------------------------------TCTTG----G----------CA---T------------------A-A---G---GG-------T--GA-AC----T-TAAG----------   7326
H101_AE.TH.CM240   ----------G----........................T-T---G----------CA---T---A-----A--A-----G----------G-----------A-CA----T-TGAGA--C--AC--   7719
H1N.CM.YBF30       -----C--C--TA-CTCT......................TAT--TACA--C------A-TTTA------CAAC-A-----TGAGA--G-A-GA--C--TT--G-TG-C--TTTTGGAC-T--A---C   7717
H1O.CM.MVP5180     ---TCAGGA......................AGATATAATGAT--CAGT------G-CA--CTT--A---CA-C-A------CA-C-C--A------GTA-GCTC-A-T---T-TGATGA---AC---   8240
H1O.BE.ANT70       ----TAGGA..............................AAC---AGC------G-CAC-TTA--A---CA---A-----TC-GC-G--A-G---CATA-GCTC-ACC---T-TGAGGA---AC--A   8213

MAC.US.SMM251      A-T.....................................--AAC--CAGAC-----CA-TGAT--T---CAA--------GC--A-GG--G--TTC-TGGAGGAAAAT---ACAG--C-CC-A----   8554
MAC.US.SMM239      A-T.....................................--AAC--CA-AG-----CA-TGA---T---CAA--------GC--A-GG--G--TTC-TGGA-GAAAAT---ACAG--C-CC-A--G-   8572
MAC.US.SMM239 Env  _S__.__.__.__.__.__.__.__.__.__.__.__.__.__L__T__P__K__W__N__N__E__T__W__Q__E__W__E__R__K__V__D__F__L__E__E__N__I__T__A__L__L__E__   gp41

STM.US.STM         -CT.....................................T---T--C-GA------CA-T-T---A---CA---A-----G---A-GG--G--TT-CTTGAGGCAAAT---ACACAAC-GT-A----   8223
SMM.US.SMM9        ACA.....................................T--AC--CT-RC-----CA-T-T--YT---CAA--------A-A-C-GG-G---TTCCTAGAGGCAAAT---ACTSAA-Y-T-G----   8061

H2A.GW.ALI         -CC.....................................T-AA---CTGA----G-CA---T---G---CAA--------ACA-C--G-CCGTT-CCTAGAGGCAAAT--CAGTGAACAGT-A---C   8611
H2A.SN.ST          ACC.....................................T-AACG-CTGA------CA---T---A---CA---A-----GCA-CG---CCG---CCTAGAGGCAAAT--CAGTGAAAGTT-A---C   8057
H2A.DE.BEN         -CT.....................................T-ATCG-CTGAC-----AA-T-T---A---CA---------G-A-C--G-CCG-T-CCTAGAGGCAAAT--CAGTCAAAGTT-A----   8627
H2B.GH.D205        ACC.....................................--CAC--CA-A-------A---T---A---CAAC-------A-AGC--G-CC-TTTC-TGGATGCAAAT---ACGG-TC-GT-A----   8603
H2B.CI.EHO         -CC.....................................--TA-G-CAGAC------A---T---A---CAAC-A-----G--GC--G-CCG-TTC-TGGATGCAAAT---ACAAAA---C-A----   8589
H2G.CI.ABT96       G-TGC-------A-A.........................T----G-CACAA------A-T-T---A---CAA--A-----A-AGC-G--C---TT--TGGAGGA-AAC--CACACGGC--T-A----   8000

TANTALUS.UG.TAN1   AAT.........................................AC--CAGAC---G-CA-T-T-------CA---A-----G--CC----A-CTGCC-TGGA-G-AAAT--TAGTA-TACTC---TCA   8256
GRIVET.ET.GRI677   AAC.........................................ACT-CT-AG---G-CA-T-T---T---T----------G---C-------TGCC-TGGA-G--AAC---ACTCAAC--T-G----   7988
VERVET.KE.VER9063  AAC.....................................A--ACG-CC-A----C-AA---T-------T----------A---C-G---GGAG-G-TGGA-G--AAC---ACAGAACAGT-A-T-A   8302
VERVET.KE.VERTYO   AAT.....................................-G-ACT-C-GA----C-AA-T-T---T---T----------A---C----AGCTG---TGGA----AAC--TACGGGACA-T-A-TGA   7771
VERVET.KE.VER155   AAT.........................................AC--CAGAG------A-T-T-------T----------A-A-C-G--AG-AGGA-TGGAGG--AAC---ACAAAACA-T-G---C   8279
VERVET.DE.VERAGM3  AAT.....................................AG-ACC-CTGA-------A-T-T---T---C----A-----A---C-G--ATCGT---TGGA-G-TAAC---ACAA-ACA-T-A--G-   7800
SABAEUS.SN.SAB1C   AAC.........................................ACT-CCGAC---G-AA---T---G---CAA--A-----G--GC----AG-A--G--TGA-GCTAAC--CAGTAGGA-TT-A--GC   8452

SUN.GA.SIVSUN      ......................................A-CACG-CC-A-----C-A-GGAC-----------A-----T--GC--G-A--A-TG--TGATGATAAT---ACAG-TC-TC-CC-G-   8474
LHOEST.KE.LHOEST   ......................................A-CAC--CC-A-----CCA-GGAT-------GA--A-----ATC-A-GG-AGCA-T---TGAC-AGAAC--TACT--TC-GT-AC---   8406
LHOEST.CD.447      ......................................A-AACT-CT-A-----CAAGGGAT--------T------------C--G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC---   7331
LHOEST.CD.485      ......................................A-AACT-CT-A-----CAAGGGAT--------T------------CG-G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC---   7328
LHOEST.CD.524      ......................................G-TACT-CT-A-----CAAGGGAT--------A--A-----A---C--G--GGATC--TAGA-GCAAAC--CAC-A-AC-GT-GC---   7334
MANDRIL.GA.MNDGB1  A-C.....................................A-CACT-CA-A-----CATCAGAA--T----A---A-----G-C-AG--C-G-TT--CTGCA-CAAAAC--TACAGAAA-GT-AA--C   7761

COLOBUS.CM.CGU1    ---....................................GGAGA---GG-TCCA---C-GA---T--------AAC-----C-TGAGCG-G-A-GA---------GATA-----G-GG-AGACT-A-T--   7567
SYKES.KE.173       --TG-A--C--C--CAAC..................TTCTGT-CCA---CACAG-----AA-T-T---A---CACAGA-----ACA---GG-AG------TG--TGA-CAT--TG-TGGA--GC-CAG--   8142
RCM.NG.NGM         ......................................-AAAC--CAGAG-----CA-ACA---A---C-T--------A---A-T---TCA-GA-TAGA-G-AAAC--CTCAGTAGCTT-AC-G-   8737
MANDRIL.GA.14CG    A-T.....................................A-AACT-CT-A-----CCTCAGAA--A---C----A-----T---AG-G-G-CAGC-AT---T-A-AAT--GACAAT-GATT-AC-GA   8233
MANDRIL.14CG  Env  _S__.__.__.__.__.__.__.__.__.__.__.__.__.__I__T__P__N__W__T__S__E__T__W__L__E__W__D__R__R__V__T__G__I__T__N__N__M__T__I__D__L__Q__   gp41
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2   Env E__S__Q__N__Q__Q__E__K__N__E__Q__E__L__L__E__L__D__K__W__A__S__._L__W__N__W__F__N__I__T__N__W__L__W__Y__I__K__L__F__I__M__I__V__-   gp41
H1B.FR.HXB2       AATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGT...TTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGG   8293

CPZ.US.CPZUS      --G-A---G-A-----GA-T-CA------A--TC-C-C--------G---C-----T------...C...TATGGAAC-GGTTTGAT-TT-CCAACTG-CT--GGTAT-T-AA-A-AT-TT-A-------T   8298
CPZ.GA.CPZGAB     --G-A---TCA--A-----G--A------AGG--CC----------G---C-------C--C..----------------G----------A------------------A-T---C-T---GC------   8333
CPZ.CM.CAM3       --G-A---G-G--A---A-C-CA-----GA--AGTC----------G---C-----T------...--A------------G-T--T--CCGC------------C--------G--T-------------C   7709
CPZ.CM.CAM5       --G-A---G-G--A--GA-C-CA------A----TC-T--------G---C-----T------...--A----G------C--T-----TC-G---T--------------A----TT-G--AG------C   8015
CPZ.CD.CPZANT     T-G-A--TG-G--A-----G-GA---A--A-G-------AT---C----C------AGCTCA...--A------------G--------C-A-----A------------A----T--T-----------   7752

H1A.UG.U455       -------G--------G-------------T---C-------C---G--C--G-------A-...C----A--------------CT-------------------T-G-C----TG----A--------   7723
H1C.ET.ETH2220    TC--------------G--C--A------A----T---C-A-C---G--C-------A--A-...C-----------------------------------------A----------------G--   7669
H1D.ZR.84ZR085    ---------T------------------A----------A-----G--C--G-----G---...----------------G-------C-A------------------A------------------   7814
H1F.BE.VI850      -G--A-----------G-------------------------C------C------------...C--------------G-----T----C------------------A----------------G--   7570
H1G.SE.SE6165     ----A--G--------G-----------------C-------CC--G--CC-G---------...---------------GG--------GA-----A------------A----T--------------   7711
H1H.CF.90CF056    TC-------C------G--------------G--C-----A-C---G--C-------------...C------C-------G-----T--C--------------------A------------------   7612
H1J.SE.SE92809    --G-A-----------G-----T------A----C-------C---G--C--G---A---A-...--------------------T----C-----------C--C---A--------------A----   7600
H1K.CM.MP535C     ----A----C------G-----------------T-------C---G--C-------------...C--------------G----------A-----A------------A--------------A----   7453
H101_AE.TH.CM240  -------G-----A--G--C-G-------A-G--T--G--A-----G---------------...C--------------G-----------------------------A----T--------------   7846
H1N.CM.YBF30      -GG-A---G-A--A--GA-C-CA-----GA--TC-C-C--------G---C------AC---...C------GC------GGT--T-----A------------------A--GCT------------C   7844
H1O.CM.MVP5180    C-G-A---G----A--G----------T-A---C---G-----GC-----G-------CTC-...C-T------------G-------T--A---T--------------A--GCT-----C-----G--   8367
H1O.BE.ANT70      -GG-A---GTA-----G---C-A-----GA--A-G--GC----G------G-------CTC-...A-T---------C--G-------T--A---T--------------A--GCA-----C--------   8340

MAC.US.SMM251     -GG-A----TT--A-----G-----CATGT-T------CAAA-G--GA---GC----ATGTG...--TG-C---------G--C-TG-TTC----A-AAA-------C---ATGGA--TTATG----T-T   8681
MAC.US.SMM239     -GG-A----TT--A-----G-----CATGT-T------CAAA-G--GA---GC----ATGTG...--TG-C---------G--C-TG-TTC----A-AAA-------C---ATGGAG-TTAT-----T-T   8699
MAC.US.SMM239 Env E__A__Q__I__Q__Q__E__K__N__M__Y__E__L__Q__K__L__N__S__W__D__V__._F__G__N__W__F__D__L__A__S__W__I__K__Y__I__Q__Y__G__V__Y__I__V__V   gp41

STM.US.STM        --G-T---GTT--A--G---------ATGT-T--G--GCAAA--C--A---GC----ATGTA...--TG-C---------G--C-T--CTC----G-AA-A--C---C---ATGGAG--TATT----TAT   8350
SMM.US.SMM9       --G-A----TT---------M-----ATGT-T-----GCAAA-----A---GSK---ATNT-...--YG-C------K--G--C-T--TTY----A-AAAA------C---ATGGTG--C-------TCT   8188

H2A.GW.ALI        GGG-A----TT---------------ACGT-T---C--CAAA-----A---GC----ATGT-...--TACC--C---C--G--T----CGCC---G-CAA------TC---ATGGAG-TTAT---A---T   8738
H2A.SN.ST         -GG-A----T--------------CATGT-T---C--CAAA-----A---GC----ATGT-...--TG-C--C------G-TT----CTCC---A-CAAA-----TC-G-ATGGAG-TTAT-------T   8184
H2A.DE.BEN        --G-C----TT--A----------ATGT-T------CAAA-----A---GC----AT-T-...C-TG-C--C------G--T----CTCC---G-CAA------TC---ATGGAG-GCAT-------T   8754
H2B.GH.D205       -GG-T--G-TA--A--G--------CATGT-T------CA-A--A--A---GC----ATGT-...--TG-T---------G--C-C--CTCC---A-AAAA-----CC-T---GGAC--TAT-------C   8730
H2B.CI.EHO        -GG-A--G-TA--A----------CATGT-T--G---CA-A-----A--C------AT-T-...--CA-T---------G--T-C--CTCC---A--GCA--C--C-GG---GGAT--TAT------AT   8716
H2G.CI.ABT96      --G-A----TT-----------A---ATGT-T--G---CA-A-----A---GC----ATGTA...--TG-C---------G--C----CTCC---G-TAAA--CG-TT-C---GG-C-TTATG----G-C   8127

TANTALUS.UG.TAN1  --G-AT-TG-G--AG-GC-G--A---ATGG-TACT-ATCAAA-G----G-G-T---A-TTC-...-G------CATC---G--G-CT--TCA---T-C----GG--T----GGGGAT-TTAT------AT   8383
GRIVET.ET.GRI677  --G-A-----T---G--TC-------CTGG-TCTG-ACCA-A----G---G-T---T--G-G...--C---TCA-----CTCAC-GT---C----T-AG-C---G-T---A--GGAT-TT-AG-GA---T   8115
VERVET.KE.VER9063 --G-AAG-G-A--AG----G--A---TT-G-T-CC-ATCAAAG---GACATCT---T---AC...--C---TC------CG-TT-CT----A-----AAACAT-C-T---A--GGAT-CT-AG------T   8429
VERVET.KE.VERTYO  -GG-TAG-G-A---G-G---------CT-G-T-CC-ATCA-A-G---AC--GT---T--GA-...--C---TC------CG-TT-CT----A-----TAACAT-T-----A-GGGAT-TT-AG--A---T   7898
VERVET.KE.VER155  -GG-AAGGG-A--AG-G---------TTGG-T-CT-ATCAAA-G--GTCAG-C---T-G-...--T---TC------CG-TT-TT----A------AACAT-T----GA--GG-T-TT--GC----AT   8406
VERVET.DE.VERAGM3 --G-CAG-GCA---G-G--G------TTGG-T-C--ACCAAA-----AG--GT---T--GA-...--C---TC------CG-TT-CT----G------AACAT-C-----A--GGAT-TT--GAT---CT   7927
SABAEUS.SN.SAB1C  --G-A--TG-A---G--C--------TTGG-TAGT-ATCAAA----G-TGTCT---T--GAC...--T---TCG------G--C-C-----A---T-TG-A-GG--G--GA--GC---C---G-TA---C   8579

SUN.GA.SIVSUN     -GG-AT-TGTAACAG-GTT-G-----C--A-CA-G--CAAAC-GC--C-AG---TTAAC.......--C----G------GG--C-GT-TC-A---T-TCTA--C-------ATGCAG-CC----TA----   8598
LHOEST.KE.LHOEST  --G-AT-C-C-ACAG-GCT-G-----C-GA-TA----TAA-A-G---C-AG---TTAAC.......--T----G------GG-T--CT--C-C---T-TACT---G-T--G-ATGCAG-GT-A--TA--CT   8530
LHOEST.CD.447     --G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA-A-GC--C-GG-G-TCAAT.......--T-----------AG----CTT-TCA---T-CCAA-----T--G-ATGCAG--C-A---A--AT   7455
LHOEST.CD.485     --G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA-A-GC--C-GG-G-TTAAC.......--T----G------GG-T--TTT-TCA---T-CCAA-----T--G-ATGCAG--C-A---A--AT   7452
LHOEST.CD.524     -GG-TT-C-CAACAG-GTT-G-----AGGA-TA-C--TAAAA-G---C-AG-T-TTAAT.......--T----G----A-GG--C-C--T-C----T-CCAA-----T--G-ATGCAG-GC-T---A-CAT   7458
MANDRIL.GA.MNDGB1 -GG-AT-TG-T-GAG-GC---GA--CAC-T-T------CA-A-G----GAG-CCTTA--TC-...-G-GCA-G-------G--T-T--TTGG---G-TCAA--CT------GGGGAG-TT-CT----GTT   7888

COLOBUS.CM.CGU1   -GG-AT-TG--TTA-----GG-A------A-GA----GGCA-------GGG-T---A-T-A-...-G--TTTCAG-C---GG-C--TTC--CATCT-TAA----G--CTC-ATGCTGC-TATG-T-----   7694
SYKES.KE.173      --G-A---G-A--A--G--G-GA----TG--T--TC--ACAA-G---C-GG------ACTCG...-------GC------G--T-GT-G--A---T-C-TT---C----GA--GGAT-TTATG--A----   8269
RCM.NG.NGM        -TG-A---G-A-----T---CGT----TG--T--C---GAAA----GA--TC-----G-GAC...A--CTC-G----C-A-----GGATTGG---T-AAA--------GGA--GGA--TT-C--TA-TTT   8864
MANDRIL.GA.14CG   GGG-AT-TG-ATT-G--C--------ATGT-T------CAAA----G-GAG-T-T-A-TTCA...-G-GCT-GC-----CG--C-C--GTGG-----CAAA---G-T--GA--GGA---T-----A--AT   8360
MANDRIL.14CG  Env R__A__Y__E__L__E__Q__K__N__M__Y__E__L__Q__K__L__G__D__L__T__S__._W__A__S__W__F__D__L__T__W__W__L__K__Y__V__K__I__G__I__L__I__I__M   gp41
```

```
                                                                                                                    / Tat, Rev exon 2 start
H1B.FR.HXB2   Tat                                                                                                   __P__T__S__Q__P__.__.__R__G__D__P__T__G__.  Tat ex2
H1B.FR.HXB2   Rev                                                                                                   D__P__P__P__N__P__.__.__E__G__T__R__Q__.__  Rev ex2
H1B.FR.HXB2   Env __G__L__V__G__L__R__I__V__F__A__V__L__S__I__V__N__R__V__R__Q__G__Y__S__P__.__L__S__F__Q__T__H__L__P__T__.__.__P__R__G__P__D__R__._  gp41
H1B.FR.HXB2       AGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCA...TTATCGTTTCAGACCCACCTCCCAACC.....CCGAGGGGACCCGACAGG...  8411

CPZ.US.CPZUS      --CAAGCT---TAGG--TC-G-A--G-A-G---GA-A-T-TC-C-AGTAGC-AAAGTTAG-CAGGG--ATT-TCCCC-C-----A-------TTT-------A......A-TC---A---A-------...  8419
CPZ.GA.CPZGAB     ---AA-CA----G--------TA--A-GA-A--TT-C--AG----C-GG-------------C--C-----T...C-C-----G-------TTA-----GT-......-A-----A--AA-GA---...  8451
CPZ.CM.CAM3       ------A-----CA---------AGGG---A-CA-A--AT-T--AGCA-AG-------------C--T--C...C-C-----G-------TTA----G--A......A-A----A---A-------...  7827
CPZ.CM.CAM5       ---T--AA------T---C--A-AGGAAT---TA-G--AG-TA-AGCA-AG-------------C--T--C...C-C-----G-------TTA----G--A......A-A----A---A-------...  8133
CPZ.CD.CPZANT     --CTA-T-----AC-------TT-GC---TAT--G--AG-TGCT-A-GA-AG--------------CAT--T...C-G--A------......A----T---CAAAAC-A-CA--AT--A--GCA-...  7870

H1A.UG.U455       ------AA----A---------------A----G---------A-C-------------------C-----T...C-G------------TTGC-----T-......--AGA---T-T--G-----...  7841
H1C.ET.ETH2220    ---TG--A-------------A---------G------------------------------C------..---------------TTA-----CA-......------------------...  7787
H1D.ZR.84ZR085    -------A--------------------A-------------A----------------G--C-----T...---------------T-------G-......------------------...  7932
H1F.BE.VI850      -------A----C-----------------------------A-----------A------C-----T...-G--A--A-------TTA------G-......------------------...  7688
H1G.SE.SE6165     ---T--AA----------------------G-----------A------------------C-----T...-G--A----------TTAC--ACCA-......-A-----A----------...  7829
H1H.CF.90CF056    ---T--AA-------------A---------G-----------A------------------C-----T...-G--T----------TTG-----G-A-......--AC--------------...  7730
H1J.SE.SE92809    ---T--AA-------------A---------G---G-------A------------------C-----T...-G--A----------TTA------A-......--A-C--A-G--------...  7718
H1K.CM.MP535C     -------A--------------C---------G------G----A--------C----------C-----T...-G--A----------TTA------------.....T---------G-A-----A...  7571
H101_AE.TH.CM240  ---T--AA-------------A---------G-----------A------------------C-----T...-G--T--C-------CTTC--ATCAT......-A--A--A---------A...  7964
H1N.CM.YBF30      ----A-T-----CA-------CA-AAG-ATA---A-AA-----A-AGCA-------------------T--C...C-T-----G-------TTA-------A......G-A--------A-------...  7962
H1O.CM.MVP5180    --CAC-AA-----A-----G-TA--A-GATAA-----AA-C-------G-ACA-------------CA---C...C-C-----G----T--CTG-----CA-......-G-CA--A-G-A--A-C-...  8485
H1O.BE.ANT70      --CAC-A-----GG-G---G-TA-CA-GATA------AA---------A-ACA-------------CA---C...C-C-----A----T--C-AA--ATCA-......-AAGA--A-G-A-GA-C-...  8458

MAC.US.SMM251     ---AG-AA--CTG----------GA-CTA-A--G-ACAA--GC-AGC--AGT-A--------G---AGG---GTG--C--T-CC-CAC--TCTTATTTCCAGTAGACT-AT-CCCA--AG...GACCCG  8808
MAC.US.SMM239     ---AG-AA--CTG----------GA-CTA-A--G-ACAA--GC-AGC--AGT-A--------G---AGG---GTG--C--T-CC-CAC--TCTTATTTCCAGCAGACC-AT-TCCA--AG...GACCCG  8826
MAC.US.SMM239 Tat                                                                                                         __P__I__S__N__R__.__T__R  Tat ex2
MAC.US.SMM239 Rev                                                                                                         D__P__Y__P__T__.__G__P__  Rev ex2
MAC.US.SMM239 Env __G__V__I__L__L__R__I__V__I__Y__I__V__Q__M__L__G__K__L__R__Q__G__Y__R__P__V__F__S__S__P__P__S__Y__F__Q__Q__T__H__I__Q__Q__.__D__P_  gp41
                                                                                                                          \ Tat, Rev exon 2 start

STM.US.STM        ---A--A---ATG------G---CAA-CTA-A--A-GCAAT-GT-AGC---GC-----A-A--T---CGG--TGTG--C--T-CC-CTC--TCTTGT-GTCAGCAGATC--T-TCCACAAG...G-CCAG  8477
SMM.US.SMM9       ---AG-AA----A-----------GA-ATA----G-GCAG--GT-AGC---GT-A--A-----T---AGG---GTG--C--T-CC-CTC--GCTTRTGTTYAGCAGATC--T-TCYARA-G...G-CCAG  8315

H2A.GW.ALI        ---AA-A----C-C-T--------AA-ATA----G-GCAA--GT-A-G----C-C---A----C---AGG--TGTT--C--C-CC-CTC--GGTTA-ATCCAACAGATC-AT-TCCACAAG...GACCAG  8865
H2A.SN.ST         ---AA-AA---T-----------AA-ATA----G-ACAA--GT-A-G----C----AA----C---AGG--TGTT--C--T-CC-CCC--GCTTA-TTCCAACAGATC-AT-TCCACAAG...GACCGG  8311
H2A.DE.BEN        G--AA-AA---C-----------CAA-CTA----G-GCAAT-GT-A-G----T----AA----C---AGG--TGTT--C--T-CC-CCC--GGTTAT-TCCAACAGATC-AT-TCCACAAG...GACCGG  8881
H2B.GH.D205       ---G--A----TG----------AG-ATACA-TG-GCAG--GC-AGCA---C-----A----C---AGG---GTA--C--C-CC-CTC--TCTTATA-TCAACAGATC--T-TCC-CAAG...GACCGG  8857
H2B.CI.EHO        ---AA-A----TA----------CAA-ATACA-TA-ACAG--GC-AGCA--GC-----A----C---AGG---GTA--C--C-CC-CTC--TCTTATA-TCAACAGATC--T-TCC-CAAG...GACCGG  8843
H2G.CI.ABT96      ---AG-AA---TG----------CA-ATA----G-ACAA---C-AGGA---C-------A--G---AGG--TGTT-YT--T-CC-CTC--TCTTATGTTCAGCAGATC-AT-TCC-CAAG...GACCAG  8254

TANTALUS.UG.TAN1  ---AC-TA--TTG--T--G--G-C--GGCT-A-TTGGGGATGCA-AGC---G-----------T--C-TT--T...--G--TCC-----T-A-TA---GCCTG..................-GG---GAA  8492
GRIVET.ET.GRI677  GATTA-TC----A------T-T-CA-GG-TAT--TGGGGATGTA-C-GA-ATA-------------AAT--T...C-C...CCC----T---TA---AC-GTTCA...G--GAAC-G--A----ACGGA  8236
VERVET.KE.VER9063 ---TA-CA----AC-T--GT--C-G-A-A-AA--TA-AG-TGCA-CGG---G-----------C-----T--C...-----TCCA----T---TA---ACC-GTGG...AA-G-AC-G--A-G--ACGCA  8553
VERVET.KE.VERTYO  ---AA-AA----G------T--C---ACA-A---TA-GGATGTA-AGTG--G---------------GTT--T...C----TCCA----T---TA---ACCAAGTGGGGAA-G-AC-G--A----ACGCC  8025
VERVET.KE.VER155  ----G-TA----G------T-GC---ACA-AT--TA-A--TGCA-AGC---G-----------T--C--T--T...-----TCC-----T---TA---ATC-GTGG...AA-G-ACAG--A----ACGCA  8530
VERVET.DE.VERAGM3 ---TA-TA----A------T-GC---A-A-A---TA----TGCA-AGC---G-----------T--C--T--T...C-T--TCCA----T---TA---ACC-GTGG...AA-G-ACAG--A----ACGCA  8051
SABAEUS.SN.SAB1C  ----A-CA---TAGCT---G-TC-GC-A-TAA--A-AGGA--TC-T-GG-AGT----AA-A--G---G----TCTC-CC--CC-C-CCT-TTCT-ATTA-.........-A--TCCATAT-C--CT-GAC  8700

SUN.GA.SIVSUN     CATTA-AA---CAGCT-----TC--AG-TT-A-CA-ACAACAGA-CT----GA-GT-T---------AGGGTG...-G--CCC-TCTG-TT-TG-TGA-CAGGATTGG-TCCA--A-.........ACC  8716
LHOEST.KE.LHOEST  T-TAA-CA----G------G--T-GAGCTT-A--A-ACAA-AT--AGTA-A-A-GT-TAG---G---AGGGTG...C-C--CCCCTCTGTTT-TA-TGA-CAGGACTACAA-T---AGAAA--AGA-AAC  8657
LHOEST.CD.447     ---TA-TA---T-------G----AAGCTT-A-TG-GCAA-ATA-AGTA-A-A-GT-TAG-------AGGGTT...C-TG-CCCCTCTG-TT-TG-TGA-CAGGACTGCAA-T---AGAAAAGAGA-AAC  7582
LHOEST.CD.485     ---TA-TA---TG----------AAGCTT-A-TG-GCAG-ATA-AGTA-A-A-GT-TAG-------AGGGTT...C-TG-CCCCTCTG-TT-TG-TGA-CAGGACTACAA-T---AGAAA-GAGA-AAC  7579
LHOEST.CD.524     ---TA-AA--ATAC-------TC-AAGCTT-A-CA-ACAA-GT--AGTG-AGA-GT-TAG---C---AGGGTT...C-CG-CCCCTCTG-TT-TG-TGA-CAGGACTTCAA-T---AGAA--GAGA-AGC  7585
MANDRIL.GA.MNDGB1 ---AA-TA----A--------TT-G--A--CT--TGGAA--C-A-A-G---GT---------C---CG---TGTC--T--A.................CAGGACTGC-A-CA-AAC......CTATAC  7994

COLOBUS.CM.CGU1   ---T--AA----AC-T-----TA--A-G-TA---A-AG--TGCA-A-GAG---C-TTCAGA-TTA-GGG-TTT...CAGCA-A-AGGC-GGACTAATGTGT-TTCACAGATCC-T--C-AA---C--GAG  7821
SYKES.KE.173      --CT--A--GCT------GC-----AG-TT-AGTG-AGGA--CA-A--G-ATT-GTT-GGA--C---GTT--TATT--GCA-AAC-CT-----G.......................GG-CGG-A-GAC  8375
RCM.NG.NGM        ---AA-AA----C---------A-C---TTAC-CTGGAGCTGT--T-G--AGC-----GCA--T--CA-G--TTTGC-C--TCCC-CAC-TTGTTAT-ATCAGCAGATC--T-TCC-CAATCGGGA-GAA  8994
MANDRIL.GA.14CG   G-TAG-AA----AC-T-----TT-AGC-TG-T--TGG----C-A-AGG---GT----------T---CGC--TCTTCCT-AT..............GTCTTAAAGGGAGACTATCTC-GACCTC-CAAC  8475
MANDRIL.14CG  Tat                                                              __R__V__I__G__L__F__L__M__.__.__.__.__.__S__$_                            Tat
MANDRIL.14CG  Rev                                                              A__G__L__S__P__S__S__L__.__.__.__.__.__C__L__K__G__R__L__S__P__T__S__Q__  Rev ex2
MANDRIL.14CG  Env __V__V__I__G__L__R__I__L__A__G__L__W__S__G__I__G__R__F__R__Q__G__Y__R__P__L__P__Y__.__.__.__.__.__V__L__K__G__D__Y__L__R__P__R__N_  gp41
                                                                   \ Tat, Rev exon 2 start      Tat exon2 end
```

HIV-1/HIV-2/SIV complete genomes

```
                        Tat exon 2 end \
H1B.FR.HXB2   Tat  __.__.__.__P__K__E__$_                                                                                             Tat
H1B.FR.HXB2   Rev  .__.__.__A__R__R__N__R__R__R__R__W__R__E__R__Q__R__Q__I__H__S__I__S__E__R__I__L__G__T__Y__.__.__.__.__.__.__.__L__G__R__S__A__E__P  Rev ex2
H1B.FR.HXB2   Env  _.__.__.__P__E__G__I__E__E__E__G__G__E__R__D__R__D__R__S__I__R__L__V__N__G__S__L__A__L__I__.__.__.__.__.__.__.__W__D__D__L__R__S__   gp41
H1B.FR.HXB2        .........CCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCCTTGGCACTTATC.....................TGGGACGATCTGCGGAGCC  8511

CPZ.US.CPZUS       .........--A-------C--------GC-C----A-A-C-----AC-T------A-CGA-------C-G-----T---A------G--.....................-----A-----T----A--  8519
CPZ.GA.CPZGAB      .........-T--G--A---C--C----G------C---CA-------AG------AG-GA--------G-A--G-G----C--------.....................--------C--C--A-A--  8551
CPZ.CM.CAM3        .........--A-G-----C---------C-T-------CC--G--A--G-----------C-------G-----T---A--T------.....................-----A-----C----A--  7927
CPZ.CM.CAM5        .........-GA-G-----C---------C---------CC--G-----G--------CGA----G--C-G---G-T---A--T---G--.....................-----A-----T----A--  8233
CPZ.CD.CPZANT      .........--A----A----AG------------AGA-A------GAT---G-GG-GGGCC--GCA-C----G-T---C-----CT-G.....................----TG--C---AC----A  7970

H1A.UG.U455        .........--A-G-A----C-----------------CA--G--A---------G-----C-------G-----T---A--GA--GC-.....................--------C-------A--  7941
H1C.ET.ETH2220     .........-T--G------C-----------------CA--G---------------A--C------------T---A---A--T-T.....................--------C----------  7887
H1D.ZR.84ZR085     .........--A--------------------------CA--------GT-C-----------G---------T--CA---------.....................------------A---A--  8032
H1F.BE.VI850       .........----------C-------G----------CA--G--A---------G-G--C-------C-----T---A--T---GC-.....................------------A---A--  7788
H1G.SE.SE6165      .........----------C-------G------C---CA--G-----G------AG-G---------G-----T---AC-G------.....................--------C----------  7929
H1H.CF.90CF056     .........----------C-------G------C---CA---------------G-GA---------------T---AC--G--G--.....................--------C--C-------  7830
H1J.SE.SE92809     .........----G------C-------G----------CA--G----AC------G-----------------T-C-A--T---GC-.....................--------C-------A--  7818
H1K.CM.MP535C      .........----------C-----------------CA-----A-A-------AG-----------G---C-T---A--G---GC-.....................---------------A--  7671
H101_AE.TH.CM240   .........----------C-------G------C---CA--G------------AG-G---------G-----T---A------GC-.....................-----------A-------  8064
H1N.CM.YBF30       .........--A----A--C-----G--GC-T-------CA--------G-------G-G---------G-----T--CA--T---G--.....................-----G--C--C----A--  8062
H1O.CM.MVP5180     .........--A-G-A---C--G--------------AG-------GCC--AG-GG-CAGCC--GCCACCA----T----CA--AGT-G.....................ACACG-----CA---CAA  8585
H1O.BE.ANT70       .........--A-G-A---C--G--G--G---------AGA--G---GCC---G-GG--A-CC-CGCC-C-A--G-T----C----GT-G.....................ACACG--C--CA-A-CAA  8558

MAC.US.SMM251      GCACTG...--AACCA--GA--GCA-------A-AC-GTG----AG-C-GTG-CAA--GCTCC-GGCCTTGGCAGATAGAATATA--CATTTCCTGATCCGCCAACTGATAC-CCT-TTGACTT--CTAT  8935
MAC.US.SMM239      GCACTG...--AACCA--GA--GCA-----A-A-AC-GTG----AG-C-GTG-CAA--GCTCC-GGCCTTGGCAGATAGAATATA--CATTTCCTGATCCGCCAACTGATAC-CCT-TTGACTT--CTAT  8953
MAC.US.SMM239 Tat  __H__$_                                                                                                                  Tat
MAC.US.SMM239 Rev  G__T__.__A__N__Q__R__R__Q__R__K__R__R__W__R__R__R__W__Q__Q__L__L__A__L__A__D__R__I__Y__S__F__P__D__P__P__T__D__G__P__L__D__L__G__I  Rev ex2
MAC.US.SMM239 Env  _A__L__.__P__T__R__E__G__K__E__R__D__G__G__E__G__G__G__N__S__S__W__P__W__Q__I__E__Y__I__H__F__L__I__R__Q__L__I__R__L__L__G__W__L__  gp41
         Tat exon 2 end /

STM.US.STM         GAACAG...--AACCAA-GA--G-AC-----AA----GTG---------GTG--AT--ACTCC-GGCCTTGGCA-ATAGAATATAC-CATTTCCTGATCCGCCAGCTAGTAC-CCT-TTGACTT--CTGT  8604
SMM.US.SMM9        GAACTG...--AACCAA-GA--G---------A-AC-GTG---R-----GTG-CAA--GATCT-GGCCTTRGCAGATAGAATATA--CATTTCCTGATCCGCCAGCTGATAC-CCT-TTGACTT--CTAT  8442

H2A.GW.ALI         GAACAG...--AACCA--GG----AC-----AA-AC-TTG------AC-TTG-GGA--GATTG-GGCCCTGGCCGAT-GCATATT-ACATTTCCTGATCCACCTGCTAGCTC-CCT-TTGA-CG--CTGT  8992
H2A.SN.ST          GAACAG...--A-CCA--GA----AC-----AA-AC-TTG--A----C-TTG--GA--A-T-G-GGCCCTGGCCGATAAGATATA-ACATTTCCTGATCCGCCAGCTGATTC-CCT-TTGAACA-ACTAT  8438
H2A.DE.BEN         GGACAG...--A-CCAACGA--G-AC-----AA-AC-TCG---G-GACAGTG-T-A-GACTTG-GGCCTTGGCC-ATAAACTATG-GCAGTTCCTGATCCACCTACTGACTC-CCT-TTGA-CG--CTAT  9008
H2B.GH.D205        GGACAG...--A-CCAACGA----AC-----AA----GTG--A--GAC-GGGAC-A--GATCC-GGCCCTGGCAGATAGAATACA--CATTTTCTACTGCGCCAGCTGAGGAACCT-TTAA-TT--CTAT  8984
H2B.CI.EHO         GGACAG...--A-CCAACGA----AC-----AA----GTG--A---AC--GG-C-A--GATCT-GGCCCTGGCAGAT-GAATACA-CCA-TTCCCGATTCGCCAACTGAGGGACCT-TTGA-TT--CTAT  8970
H2G.CI.ABT96       GAACWG...--AACCAA-GA----AT------A-A-AGTG--A---A--GTG-C-A--GATCT-GGCCTTGGCAGATAGAATATA--CATTTCCTGMTCCGCCAGCTAGGGAACCT-TTGACTT--CTAT  8381

TANTALUS.UG.TAN1   ......CAG--A--CAACGC--G--G-----ACAA---TTCCAG---CAG----GA--AGTCGCCGCCCTCA-TGAAAGAAT-TT-AT-GCCCAACAGAGGAGGGATCCAAGC---G--GAGAGCTT-G-  8616
GRIVET.ET.GRI677   GGAGGG...-AA--CA--GGT-G----AGCA-CA-CAGC-A-TTG-T-AGATTGCAGGAAGAG-CCTCA-CACCT--GAG-ATCAAC-A-TGGTGGCTCAACTTCAAGAGC--CAG-TTGAGAATA--GA  8363
VERVET.KE.VER9063  GGAGAG...--A----A-GGT-G---GAG--ACAAC--C--CTCAGATTCTT-GCAG-AAG---CT-GA-CGA--CAAAA-AG-GGC-GTTGGAGCAAACGCTTGACCAAT---CT-TCGAC-TTC---A  8680
VERVET.KE.VERTYO   GACGAG...--A-G--A-GGT-G---CA-CA-CA-GATC-A-TTAGAGTCTT-G-AG-AAGAC-CCAA--G-A-G-GTA--CAGT-G-C-GCCTGGCTGACGAGGCTCAAC..................A  8134
VERVET.KE.VER155   GGAGAG...--A----A-GGT-G-AG-AC---CAA-AGC-A-TCT-CGC-TTAGCAG-AAGA---T-G-GGAC--GA-AA-AGGACC-GTTGGTGCAGGCAATTGACCAAT---TT-TTGACA-TC---A  8657
VERVET.DE.VERAGM3  GAAGGG...--A-G--A-GGT-G---CA-GC-CAAGA-C--CTC-GAGCCTT-GCAG-AAGA--CT-GC-CA-C-GAG-G-AAGAGC-A-TGGTGCAAGCGATTGACCAAT---TG-TCGA-CA-C---A  8178
SABAEUS.SN.SAB1C   AAGGGACAG--A--CAA-GA------GC----CA-C-GTG-GCTA-ACAG-GACAG--GGTC--ACACTTGGCAGAGAGAATTCT-G-GACACCTGTGTCACAGATTGATCACTTGGCTCAG-AATTTGA  8830

SUN.GA.SIVSUN      TGCCCCAAA---AC--AC-A-----------AA-AGACAGAGA-AGA-AGA-TC-A---CAAC----A-C-GTCCAAGAAA-A-TC-T-GCCACCACCTTGGACAGTAGAC--------G-CTTT---AG  8846
LHOEST.KE.LHOEST   CAAGAACAG--A--CA--GA------GA----A-C---T-C---G-CTAT-...-A---CAAC----A-C-GT-CAAGAAA-A-TC-TC-AGGCCACTGTGGAACGTGGAC---A----G-CTTTACAAG  8784
LHOEST.CD.447      CAAGAACAG--A--CAA-GA--G---GAT--AGAA---C-A---A-ACAT-...-A---CAAC----AAC-GTACAAGAAA-A-TC-TC-ACTCCACCGTGGAACGTGGAC---AG---G-CTTTACAAG  7709
LHOEST.CD.485      CCAGAACAG--A--CAA-GA--G---GAT--AGAA---C-A---AGACAT-...-A---CAAC----AAC-GT-CAAGAAA-A-TC-TC-ATTCCACCGTGGAACGTGGAC---AG---G-CTTTACAGG  7706
LHOEST.CD.524      CAAGAACAG--A--CAA-GA--G---GAC--AGAA---C-----A-ACAT-...-A---CAAC----A-C-GT-CAAGAAA-A-TC-T--AGGCCACCGTGGAACGTGGAC---A----G-CTTTACAGG  7712
MANDRIL.GA.MNDGB1  CGCAAACGG--A--CAACGG----------A-CAACAGCTT---ACT--GA...GAGCACAAC-CC-A----TTG.........AAGGAAGAATCCTTAAACAGATCCTTGATA--G--C---ACC---T  8112

COLOBUS.CM.CGU1    TGGGAACAG--A--CAACGC-------AT-CACA-----CA---A-ACAG---CAG--GATTT-GGCGATTGATGGAAGA-TGGTGG-G-.........................................  7911
SYKES.KE.173       CCAGGCAAA----C--ACGA--------GCA-------C-----AG--CT--ACGT-TC-ACC--CAGC-GA-A---T---AGG-AATCTTTGGAAGCAGGACAGCAGCTT---AGGACAG-CT-C---A  8505
RCM.NG.NGM         CAGCCAATCAAA------GA-A---G--GCA-------AG-C-GTG--C---A--ATTACAAC-GGCA--GA----AT--TTATA-CC-GATACTCCGCCCGATAGAGATCCTTTG-AGGAG--TCTAT-  9124
MANDRIL.GA.14CG    CTCAGACAG--A--C-A-GA-AG-----G--AGCTCA-CTT---AGA-C-G...AA---CAAG-C--A--G-TCC.........AAG-AAGAATTTGGCAGACCTTGGAGGCCA--GC-AA-AA---ATT  8593
MANDRIL.14CG  Rev  P__Q__T__A__R__R__R__K__R__R__R__A__Q__L__R__R__T__.__E__H__Q__V__R__E__L__.__.__.__Q__E__R__I__W__Q__T__L__E__A__R__A__N__K__E__L  Rev ex2
MANDRIL.14CG  Env  _L__R__Q__P__D__E__E__R__E__G__E__L__N__L__E__E__Q__.__N__I__K__S__E__S__S__.__.__.__K__K__E__F__G__R__P__W__R__P__E__Q__I__R__N__  gp41
```

```
H1B.FR.HXB2   Rev __V__P__L__Q__L__P__P__L__E__R__L__T__L__D__C__N__E__D__C__G__T__S__G__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__T__Q__._  Rev ex2
H1B.FR.HXB2   Env L__C__L__F__S__Y__H__R__L__R__D__L__L__L__I__V__T__R__I__V__E__L__L__G__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__R__.  gp41
H1B.FR.HXB2       TGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGA.....................................................CGCAGG..  8588

CPZ.US.CPZUS      -CCTGA-----CT---------A---GAG--------A--A---C-G-G-----CA--AC--A-C------.............................CAGAACATCAACAAGGGACTG...CAA..  8614
CPZ.GA.CPZGAB     --G-GA-T-GG--------GA------C-AG----GC--GC-A---GTG---ACAAC-CA-GACGT-----.......................................CATCTAATA-TTCAC..  8637
CPZ.CM.CAM3       -CCTG------T-T--------A--------------A-----AC-GGG-----CAT-----AAC------.............................CAGAGCCTCAACAAGGGACTG-AACA-..  8025
CPZ.CM.CAM5       -CCTGA-----TT---------A----T-----------CA--CC-GTG-----CA---C--A--------.............................CAGAGCATCAACAAGGGACTG-AACA-..  8331
CPZ.CD.CPZANT     -AAT--AG-GG-T------GAT--GC---AC--GT-------ACC-TTG-GCAG-CC-CC---AC--CT-C.............................AGAATTACTTTCAGACTGTGCAA-CAT..  8068

H1A.UG.U455       ---------------------A-----------CGC---------G----AGC---------------.............................CGCAGCAGCCTCAAGGGACTGA-ACT-..  8039
H1C.ET.ETH2220    -------------T--------------------A---------C-G------CA---------------.............................CGCAGCAGTCTCAAGGGACTA-AG----..  7985
H1D.ZR.84ZR085    ------------------------------A---A----A----C-G-----------------------.............................................------..  8109
H1F.BE.VI850      ------------------G--A------------CA-A--A----C-G-------------CAGGGGCCTG.......................................A-G----..  7865
H1G.SE.SE6165     --------------------------------CCA----------G------CA---------------.............................CGCAGCAGCCTCAAGGGACTGA-ACT-..  8027
H1H.CF.90CF056    ---CA-------------G--T------------------A------GT-----C---------------G.......................................A-A----..  7907
H1J.SE.SE92809    -------------T--------A-----------CG---------C-G------C-----G-AC-------.......................................-T-----..  7895
H1K.CM.MP535C     ------------------G--AA------A-----A----A-----G--------CC-----AGGGGACTG.......................................A-GG---..  7748
H101_AE.TH.CM240  ---------------------G------------AC--------C-G------CG---------------.............................CACAGCAGTCTCAAGGGACTGA-AC--..  8162
H1N.CM.YBF30      ---TGA-----CT--------------C-----C---------AC-G-G-----C-C-------C------.............................CAGAGTCTCAGCAGGGGACTG-AACTA..  8160
H1O.CM.MVP5180    -AAT-T-G-GG-CT-------T------CA-----A-A-CAGGGA-CCG----C-GA-C--CTAC------.............................CTGGGACTGTGGATCCTGGGA-AA-A-..  8683
H1O.BE.ANT70      -AATAT-G-GG-CT-------T------CA-----GCA-CAGGGA-CCA--A-G-GA-CAGCTA----A-G.............................CTTGGACTGTGGATCCTAGGG-AG-A-..  8656

MAC.US.SMM251     -CA--AA--G---AAC-TTG-TA-C-----CA-AC-AG.........................................ATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTA--A---..  9018
MAC.US.SMM239     -CA--AA--G---AAC-TTG-TA-C-----TA-AC-AG.........................................ATCCTCCAACCAATACTCCAGAGGCTCTCTGCGACCCTA-AG----..  9036
MAC.US.SMM239 Rev __Q__Q__L__Q__N__L__A__I__E__S__I__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__D__P__P__T__N__T__P__E__A__L__C__D__P__T__E__._  Rev ex2
MAC.US.SMM239 Env F__S__N__C__R__T__L__L__S__R__V__Y__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__I__L__Q__P__I__L__Q__R__L__S__A__T__L__Q__R__.  gp41

STM.US.STM        ACAA-AA------GGCTTGT-TA-CC--GAT--AC-AG.........................................ACCCTCCACCCAACCTTCCAAAGGATCTCCAGGATACTG-AG--A..  8687
SMM.US.SMM9       -CA--AGT-G---GG-TTGG-TA-----GAN--GC-AA.........................................ATCCTCCAGCCAGTGCTCCAGAGCCTCTCAAGGACGCTG-AG--A..  8525

H2A.GW.ALI        ACA--A---G---GG--TTA-TA-CC--GAT--CC-CG.........................................ATCCTCCAACCGATCTTCCGGAGTCTTCAGAGAGCGCTGACA-CA..  9075
H2A.SN.ST         ACAA-A---G---GG--TTA-TA-CC--GAG---C-AG.........................................ACCCTCCAACTAATCTCCCAGAGTCTTCGGAGAGCATTGACAGCA..  8521
H2A.DE.BEN        ACAA-A---G---AG--TTA-TA-CC-AGA---CC-CG.......................................ACCCGCCGACTGATCTCCCAGAGTCTAACAGCA..  9079
H2B.GH.D205       ACAA-GG--G---AAC-TTA-TG----AGAC---C-AA.......................................ATCCTCCACCAAATCTCA-CC..  9043
H2B.CI.EHO        ACA--GG--G---AAC-TTATTA-CA-AGAC---C-AA.....................................ACCCTC-AACCA..  9020
H2G.CI.ABT96      ACA--AA--G---AGC-TTA-TA-CC--GAT-Y-C-AG.....................................ACTCTC-AACC-..  8431

TANTALUS.UG.TAN1  G-CAG-AT--G-C-A-TTGGT------C-AT-AGCAGT-G-T-AT-G-GACTTTACCA-ATC--C-GCA-G.........................................AGCCTCACG-CT..  8699
GRIVET.ET.GRI677  CT--GTGT-A--A-AT-TG--TGACCCTCCT-A--T--.......................................................................................  8401
VERVET.KE.VER9063 CC--G--A-A--A-AC-TT--TGACCCTCCTAA-T-AT.......................................................................................  8718
VERVET.KE.VERTYO  CT--G--A-A--A-AGTTG--TGACCCTCCT-A-T-AG.......................................................................................  8172
VERVET.KE.VER155  -C--GT-ACA--A-AG-TG--TGACCCTCCTTC-CAAG.......................................................................................  8695
VERVET.DE.VERAGM3 -C--G--A-A--A-AGTTG--TGACCCTCCTAG-T-AT.......................................................................................  8216
SABAEUS.SN.SAB1C  CCA-TTGG---TCGACAAT-TT-AGC-ACCT-CAT-G-.......................................................TGCCTC-AGGAC..  8880

SUN.GA.SIVSUN     AC-CT--G--AGTGACT-TT-T-AA-T-GCTGAA-GCT-G-GGAA--GTCCTAGCACAAA-CA-CTACCACCTGCTAAGTTTCTTGTGGCACCTACTTACGACTTCCTTCCATCATGGGCAA---CTT..  8974
LHOEST.KE.LHOEST  AC-CG--G--AGTGACT-T--TGAA-T-GCT-AAGGAAGG-GGAA--CTCCTCC-ATCACT-G-GTG-CAGAGCCTCAGTTGGCTGTGGCATTTGTTGATCCTTTTCTTCCAAAATGGGCAA---CTT..  8912
LHOEST.CD.447     AC-CA--G--TGTGACT-T--TGAA-T-GAT-AAGGC---TGGAA-TCTTCTCC-CAGCCTTG-GTG-CAGTTCCTTTCGTGGCTAGGCCATTTAGTGATCCTTTTCTTCCAACATGGGCAA-ATCTT..  7837
LHOEST.CD.485     AC-CA--G--TGTGACT-T--TGAA-T-GAT-AAGGCA--TGGAA-TCTTCTCC-CAGCCTTG-GTG-CAGTTCCTTTCGTGGCTAGGCCACTTAGTGATCCTTTTCTTCCAACATGGGCAA-ATCTT..  7834
LHOEST.CD.524     AC-CA--G--TGTGACT-T--TGAA-T-GCTGAAGGCG-G-GGAA--CT-CTCC-CACCCTCG-GTG-CA-TTCCTCACTTGGCTGGGACATATAGTGATCCTTTCTTACCACCATGGGCAA-ATCTT..  7840
MANDRIL.GA.MNDGB1 -TGC-AGGGAGT-ACCAATTT-GC-TT-GCTGAA-AATC---GA-C-G-A.....................................ATTGAATAT..  8171

COLOBUS.CM.CGU1   ...........................................................................AGCTTGTTGCGTCATTGCAGCAATGGGAT-CTT..  7944
SYKES.KE.173      GC-T-AGA-CTCTGATTAGA-AGC-T-CTAT-ACCTGG.........................................................................  8543
RCM.NG.NGM        AAAT-TG-AG--A--CA-T-TCAG---TCT--CAGAG-C-CCAGTC--TCCCTTC-CGAATC-----TCATTTGCGGTGGATAGGAGCTAAGCTACAAGATGGGTGGCAA.....................  9234
MANDRIL.GA.14CG   G--TGAAGGAGTCAACAATTTA-A-TT-GCTGAAGAATC--CAG-C-GTA...................................................ATTGAATATGG  8654
MANDRIL.14CG  Env W__L__K__E__S__T__I__Y__I__W__L__K__N__L__Q__A__V__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__I__E__Y__G  gp41
MANDRIL.14CG  Rev __V__E__G__V__N__N__L__H__L__A__E__E__S__G__G__S__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__N__$_   Rev
                                                                                                                                  Rev exon 2 end /
```

HIV-1/HIV-2/SIV
complete genomes

```
                                                                                                                    Rev exon 2 end \
H1B.FR.HXB2   Rev _G__V__.__.__.__.__.__.__.__.__.__.__.__.__.__G__S__P__Q__I__L__.__.__.__.__V__E__S__P__T__V__L__E__S__G__T__K__E__$_             Rev ex2
H1B.FR.HXB2   Env __G__W__.__.__.__.__.__.__.__.__.__.__.__.__.__E__A__L__K__Y__W__.__.__.__.__W__N__L__L__Q__Y__W__S__Q__E__L__K__N__S__A__V__S__L__L_   gp41
H1B.FR.HXB2       .GGGTGG...................................GAAGCCCTCAAATATTGG...........TGGAATCTCCTACAGTATTGGAGTCAGGAACTAAAGAATAGTGCTGTTAGCTTGCTC   8669

CPZ.US.CPZUS      .CT--T-...................................A-T-AA----G-GCAC-C...TGC......---GGGG-AA-CGCT--C---GCAAG---G-----AGT-------ACA-----AT-A   8698
CPZ.GA.CPZGAB     .A-CCTC...................................AG-CTG--ACGGG-GA-ACTGTGCCTGCTGG--GGAA-TA-T--A------G-AA----G-----A-T-------A-A---C----T   8730
CPZ.CM.CAM3       .CT-A--...................................A-CTT-AG--G---CCT-...........---GGAG-GA--ACC--C---G-AAG---G---C-A-C------CA-A-----AT-A   8106
CPZ.CM.CAM5       .TT-AA-...................................--CTTTAGT-C---CCTA...........AA-GGAG-A---ACA--C---G-AAGA--GA--C-A-T-------ACA-----AT-A   8412
CPZ.CD.CPZANT     .CT-GA-...................................A-CAAT----GCACCCTC..........................----CAATAATC...-GA-C-GAGATCA---AGAACA-T   8128

H1A.UG.U455       .-------..................................--G-G---------CT-...........--------T--G-T-------G---G---------A-T-------A---CT-----T   8120
H1C.ET.ETH2220    .-------..................................---A----T------CT-...........G-A-GC--TG-G--------G---T---G-----A--G------A---ATC----T   8066
H1D.ZR.84ZR085    .-------..................................---------------CT-...........----------G--A---------GA-----G------------A--------G-T   8190
H1F.BE.VI850      .-------..................................-----T---------CT-...........G--------ACG-G-------------------------------A-C---C--T-T   7946
H1G.SE.SE6165     .-------..................................----GT--T-------T-...........--------T--GTT-------G-CAG------------------A----T-----T   8108
H1H.CF.90CF056    .---A--...................................---------------CTC...........--------T-----A--C---G-A--------------------A--GAT-----T   7988
H1J.SE.SE92809    .-------..................................--GAT----------CT-...........GT---C--TG--TG-------G-G-------------------A----T-----T   7976
H1K.CM.MP535C     .--T---...................................---------------CT-...........-----C--TG-----------------------------------CA----------T   7829
H101_AE.TH.CM240  .-------..................................----G----------CT-...........G-------T--GTTA------G-C-----------A-T-------A--TCT-----T   8243
H1N.CM.YBF30      .CT-AAT...................................---CT-AGA-C-C-C-T-...........---GGAA-A--TGCA------G-AA-A--GT---G-G--------A-C--------T   8241
H1O.CM.MVP5180    .ACAATT...................................-----TTGT-G-CT---T...........G-AGC-G-AA-G--A------CTA--A---T-G--A---------ACA-A-C----T   8764
H1O.BE.ANT70      .ATAATT...................................A-T-TTTG--G-AT---T...........GCAGC-G-AAC---A--C---CTA--A---T-GC-----------ACA--------A   8737

MAC.US.SMM251     .-TTC-A...................................----T-----GGAC-GAA...........CT--CCTA------A---G--T-GAGCT-TT-CC-TG-GGCG-TCCAAGC-GGCTGG   9099
MAC.US.SMM239     .ATTC-A...................................----T-----GGAC-GAA...........CT--CCTA------A---G--T-GAGCT-TT-CC-TG-GGCG-TCCAGGC-G-CTGG   9117
MAC.US.SMM239 Env __I__R__.__.__.__.__.__.__.__.__.__.__.__.__.__E__V__L__R__T__E__.__.__.__.__L__T__Y__L__Q__Y__G__W__S__Y__F__H__E__A__V__Q__A__V__W_   gp41
MAC.US.SMM239 Rev _D__S__.__.__.__.__.__.__.__.__.__.__.__.__.__R__S__P__Q__D__$_                                                                  Rev ex 2
                                                             Rev exon 2 end /

STM.US.STM        .ATTA-A...................................----T-G---G-CTAG-A...........GCAGCCTATT--------G--T-CATCTGGA-CC-AG-AGCA--GCAAGCAGCATGG   8768
SMM.US.SMM9       .-CCC-T...................................----T-A---G-GT-GA-...........ATA-CCTA----------G--T-G-GTT-C--CC-AG-AGCA--GCARR---G-TGG   8606

H2A.GW.ALI        .ATCA--...................................--CTGG--G-G-CT-AAA...........GCAGCCTA---G------G--T-CG--TGGA-CC-AG-AGCGTTCCGGGC---TGCA   9156
H2A.SN.ST         .-TCA-A...................................--CTGG--G-G--T-AAC...........ACAGCCTA---G--A---G--G-CG--TGGA-CC-AG-AGCGTTCCGAGC---CGCG   8602
H2A.DE.BEN        .ATCA--...................................--CTGG--G-G-CT-AA-...........GC-GCC-AA--G--A---G--T-CG--TGGA-CC-AG-AGC-TTCCAAGCA--CGCG   9160
H2B.GH.D205       .AATCTC...................................C--C-A----GGCTCCC-...........GTTGCGTA---C------G---TCAGCTGGT-CC-AG-AGCGCTCCGAGCGGCAGCA   9124
H2B.CI.EHO        .-TACTC...................................C--C-G----GGCT-CCA...........CCTGCATA---C-GA---G---TCAGCTGGT-CC-AG-AGCAATCCAAGCAGCAGC-   9101
H2G.CI.ABT96      ....CT-TTCCAAAGAATCTCCAGGACCCTCCAAGCAATTAGA---CAT---CGGCTAGAA...........GCAGC-TATT-CAGC---G--TTC-GCTGG--CC-AG-AGCGTGCACCGCAGCAACT   8545

TANTALUS.UG.TAN1  .CTTCTT...................................C--CTG--T-G-C-GGA-...........--CC-ATA-A-T------G--T-G---C--T-C--AG-GG-A--A-CA--G-CTT-T   8780
GRIVET.ET.GRI677  ..........................................A---GGACAGCA...........GT-GGATA---G------G--CTC---C----CC-AG-GGCA--AACAG-GC-TGCT   8470
VERVET.KE.VER9063 ..........................................--T-G-GGCGCT...........---C-GTA-A-C--A---G--CT-GG------C--A-CCGCA--ACAAGAAGCA-GT   8787
VERVET.KE.VERTYO  ..........................................--T-GGA-AGCT...........-TTC-ATA---G--A---G--CTCGCA-----C--A-CCG-C--ACAAGAAA-A---   8241
VERVET.KE.VER155  ..........................................--T-G-AGCGCC...........---C-GTA-T----A---G--CT-GG---G--C--AGCCGCA--ACAAGAAGCAG-T   8764
VERVET.DE.VERAGM3 ..........................................--T-GGAGCGCT...........-TCC-GTA-A----A---G--CT-GG------C--AGCCGCA--ACAAGAAGCAG-T   8285
SABAEUS.SN.SAB1C  .ATCCAA...................................C-GAGAAC--G-C-GCTA...........ACAGC--AT--TG-A---G--T-G--A--G--C--AGCAGCA--A-CGTCAC-CACT   8961

SUN.GA.SIVSUN     .T--CA-...................................ACCCT-AGAGGT-GGCT-.................................................................   8998
LHOEST.KE.LHOEST  .T--CA-...................................ACCAG-AG--G--GGAT-.................................................................   8936
LHOEST.CD.447     .C--AAA...................................-CCTG-AG--G-CGGGT-...........GT-G-AAA-GCT---A-GATTGCGAGCTGGAGCTG---C-AGATCAGA--GAACAGA   7918
LHOEST.CD.485     .T--AAA...................................-CCAG-AG-GG--GGAT-...........GT-G-AAA-GCT---A-GATTGCG-GCTGGAGCTG---C-AGATCAGA--GAACAGA   7915
LHOEST.CD.524     .T--CAA...................................ACCCT-AG--G--TCCT-...........GT-G-GAGTGCT---A-AACTGCGAGCTGGA-C-G---C-AGAGCAGA--GAACAGA   7921
MANDRIL.GA.MNDGB1 .----TC...................................CTC--AG--C--G-AGC-...........ATC-GAAG-T-GGGT-CGCTACTCGTCAGCT-TGC-GCA-T-AG-TGG--GGCCGGT   8252

COLOBUS.CM.CGU1   .ACTGTC...................................CTGC-T-AAC-TCCGCCA...........ACA-CCTCAGACA--AGGA-----GCTTGG-AG-CTCT-C-GTA-C-ATTGG-A-C-   8025
SYKES.KE.173      ........................................................................GGAT--A-CAGC---G--TTCA-C--G--C--A-TCGCA--A-C-TCGC-CGGA   8597
RCM.NG.NGM        ........................................................................................--GT-C--AGGCTTCAGCAGCT-GC-TGCG   9264
MANDRIL.GA.14CG   GT--CAA........................GAGCTCAAAGCA-CG-GAGCA-GGATA--C.............................................C-AGTATTACAA-----TGCG   8712
MANDRIL.14CG  Env __W__Q__.__.__.__.__.__.__.__.__E__L__K__A__G__G__A__R__I__C__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__Q__V__L__Q__S__F__A_   gp41
```

```
                                                                                                                                                                                  Env end \
H1B.FR.HXB2   Env _N__A__T__A__I__A__V__A__E__G__T__D__R__V__I__E__V__V__Q__.__G__A__C__R__A__I__R__H__I__P__R__R__I__R__Q__G__L__E__R__I__L__L__$_    gp41
H1B.FR.HXB2       AATGCCACAGCCATAGCAGTAGCTGAGGGGACAGATAGGGTTATAGAAGTAGTACAA...GGAGCTTGTAGAGCTATTCGCCACATACCTAGAAGAATAAGACAGGGCTTGGAAAGGATTTTGCTATAA.   8795

CPZ.US.CPZUS      G--A-A-----A--------------A--A--T-----AA--------T--AC-AG-...A-GCTC-T-CT--G----ATA--------A--------T--G--A---C-A--G--A-GC--AT----..   8823
CPZ.GA.CPZGAB     G----A-----T--T--------A--A--T--------AA-A-------CTT-T---...-TTA-ACT----AT----A-AA----C-----GC-C--------A--GC-------AGC---A--T---.   8856
CPZ.CM.CAM3       G----A--------T-----------A--A-----C--AA--C-------T-C----...AT-ATAG------GC----TG--T-----A--------T--------T--A-----A-GC---------.   8232
CPZ.CM.CAM5       G--A-A-----A--T---------A-T--A--------AA-------GT-T------...AT--TAG-G----GC----TA--T--------G-----T-----A--T--A-----A-G----T-----.   8538
CPZ.CD.CPZANT     G-CAGACTT--T--TTGG----GG--AAAA--------CA-AC-TCT--CTC-C---...ACTATAGTC---ATC--AA-GG-AG-------GC-C--C-----A--G-------TTGCA--AAAT---.   8254

H1A.UG.U455       G-C--TGT---AG------------G-T---T----------------A---G----...AC-AT-G------------T-A----------------C-----A-----A------GC----------.   8246
H1C.ET.ETH2220    ---A-------A----T-----G---A--A--------AT--------T--A-----...A--AT---G------T-CT--A---------------------------------GCAGC-----A----.   8192
H1D.ZR.84ZR085    G----A-----A--------------------------A-------CA-----AGG...A---------A---G-G-TT--------C-C--------------T--A------GC----------.   8316
H1F.BE.VI850      ---A-------A----T---------A-----------AA-----------TT-G---...A-----G--------G---T-A----------------------------GCA------GC------G---.   8072
H1G.SE.SE6165     G--A-AGT---A------------A-TT-----------------------C----...A----------------C-T-A---------C---G--------A-----------AGC----------.   8234
H1H.CF.90CF056    --CA-------A-----------------A------G--A------T-A----G---...A-------G----------T-----------------------------T-----A-GC---------T   8115
H1J.SE.SE92809    ---A-------A--------------A--A--------AA-C------A---C----...A------T-----------TT---------------------------T--A-----AGC----------.   8102
H1K.CM.MP535C     ---A-------A-------------GA--A--------AA--------A---G----...A------T-------T-A-TT--------------------------TC-C-----AGC---A------.   7955
H101_AE.TH.CM240  G----T-----A-------C---G-G-T-------C----------------C----...--------G-----C----T-----------------C-----------A-------C----------.   8369
H1N.CM.YBF30      ---A-A-----T--T-T------A--A--A----------A--------T---C----...A--ATAG-A--G-GA--ATTA-----------------C-----A---C-A-----AGCAC--A-----.   8367
H1O.CM.MVP5180    G--A-T-TT--AG-GT----T--CA-TT----T--CG-CA-C--CTT--GTC-----...A--ATAG-ACA--GAT-C-TT-----C--A--------T-----A--TGCA-----A--C--AG-G---.   8890
H1O.BE.ANT70      G-CA-ACTT--AG-G--------CA-TT----T--CG-CA-A--C-C--GGA-----...A--ATAG-A-C--GA-----TA----C--A--G-----T-----------A-----A-G---AT-G---.   8863

MAC.US.SMM251     -GAT-TG-GA-AGA-A-TC-T--G-GC-C-TGGAGAGACT-ATGG--GACTC-TAGG...A---G-G-A---TGG--C-T-GCA--C-----G--G--T--G--A--GC-T--GCTC-CGC-CT-G-G-.   9225
MAC.US.SMM239     -GAT-TG-GA-AGAGA-TC-T--G-GC-C-TGG-GAGACT-ATGG--GACTC-TAGG...A---G-G-A---TGG--A-T-GCA--C--C--G--G--T-----A--GC-T--GCTC-C-C-CT-G-G-.   9243
MAC.US.SMM239 Env _R__S__A__T__E__T__L__A__G__A__W__G__D__L__W__E__G__L__R__.__R__G__G__R__W__I__L__G__I__P__R__R__I__R__Q__G__L__E__L__G__L__L__$_    gp41
                                                                                                                                                                                  Env end /

STM.US.STM        -GA--TG-G-GAGAGA-TC-T--GAGC-C-GGGAGAGACT-ATGG---ACTC-TGG-...A-G-T-G-G---AGG--CG-GGC---C--A--GC-C--C-----A--GC-T---CTC-CAC-CT-G-G-.   8894
SMM.US.SMM9       --ATTTGYRCGAGAGA-TC-T--AAGC-C-TGGAGAGACT-ATGG--GACTC-GGG-...A-G-T-G-A--G-GA--A-T-GCA--C--MR--C-C--C--G--A--GC-T--GCTC-C-C-CT-G-G-.   8732

H2A.GW.ALI        -GGATTG-GAGAGAGA-TC-TA-AA-CACCTGGAGAGACT-GTGG-GG-C---G--G...T-G-TCG-G--GAGG--A-T-GCAG-C--A--G--G--C--G-----GGCA----TTGCCC-C--G-G-.   9282
H2A.SN.ST         -GG--T--G-GAGAGA-TC-TA-AA-C-CCTGGAGAG-CT-CTGG-GGAC-C-GGG-...CA-AT-G-G--G-GA--A-TTGCAG-C--A-----G--C--G-----GGCA----TCGCCC-C--G-G-.   8728
H2A.DE.BEN        -GGA-T--GAGAGAGA-TC-T--G-GC-C-TGG-GAT--T-ATGG----C--CG-G-...C-CATCG-G--G-GA--A-T-GCAG-T--A--------C--G-----AGCA---CTCGCCC-C--G-G-.   9286
H2B.GH.D205       -GG--T--G-GAGA-A-TC-T--GAGT-CCGG---G-CCT-ATGG--G-CTC-CAGG...A----GGCG-----A--CAT-GCG--C--C--G-----C-----A--AC-T---CTC-CCC-CT-G-G-.   9250
H2B.CI.EHO        -GG--TG-G-GAGAGA-TC-T--GAGC-C-G-GAGG-CCTCGTGG-G---CC-CAGG...A----GGCAG---AG--CATTGCA-----C--G--G--C-----A--GGCT---CTCGCCC-CT-G-G-.   9227
H2G.CI.ABT96      -GG--TG-GCAAGAGA-TC-TA-AAGCAC-TGGAGAGCCT-ATGGA--ACTC-GGG-...A-G-T-G-A----GG--A-T-GCA-----G--------C--G--A--GC-T--GCTC-C-C-CT-G-G-.   8671

TANTALUS.UG.TAN1  G-G--TTTG--G-GC--T-CGCAAAGC-CC-GTCG--C-C-ATGGA-T-CTTGCAG-...TCC----A-C-G--A--C-T-G-ACAT--A--------GC----A-AAC-----C--TGG--CAAC--G.   8906
GRIVET.ET.GRI677  C-A--TCTG--G-GG--T-CGAGG--A-CCTGG-GC--AC-GGGT-CTA-T--C-G-...TCC----A-C-G--AG-CAT-A---GT--A------G-GC-G--A---C-T----AAG-CC--GGG---.   8596
VERVET.KE.VER9063 CTC--TGTG--ACGCTTT-C-CAAA-C-CAGGTC-CCAAA-ATGGCTT-CTTGCAG-...TCCA---A---GCAC--CAT-AG--GT--A------G--C----A--GC-T--GGAA--CC-TAAT--G.   8913
VERVET.KE.VERTYO  C-AA-TCTG--AGGC-TT-C-CAAA-C-CATGTC-CCA-A-ATGGCTT-CTTGCAG-...TCC----A---GAAC--CGT-A---GT--A------G-GC----A---C-T--GGAA--CC-TAAT--G.   8367
VERVET.KE.VER155  --CCATCTG--G-GCTTT-C-CGCA-C-C-G--C-CCA-A-ATGGCTT-CTTGCAG-...TCC----A-C-G--A--CAT-A--TCT--A------G-GC----A--GC-T--GGAAG-CC-TAAT--G.   8890
VERVET.DE.VERAGM3 GTC--TTTG--ACGCCTT-C-CAAA-C-C-GGCT-CCA-A-ATGGCTT-CTTGCAG-...TCC----A---G--A--CAT-A--TCT--A------G-GC----A---C-T---G-A--CC-TAAT--G.   8411
SABAEUS.SN.SAB1C  GTG--TGTG-T-CA---T-CGA-AAGC-CC-GT---TCA--ACG-C-T-CTTGCAG-...TCCAT-GT---G--AG-CAT-GCACAT--A--------GC-C--A-AGC-T--G--ATGG--TAAT--G.   9087

SUN.GA.SIVSUN     ..........................................................................................................................   8998
LHOEST.KE.LHOEST  ..........................................................................................................................   8936
LHOEST.CD.447     -GA-A--T-T-AGC---T.........CATG--AGA-ACC-C-G-CTG-G-CA-A-G...AAGAGA--G---TTC-GATT-AGAGG-AGA--TG-GT-TCCCAGT-AGACCAC-GA--CCGCT--G--G.   8035
LHOEST.CD.485     -GA-A--T-T-AGC---T.........CATG-GAGAGACA-C-G-CTG-G-CA-A-G...AA-AGA--G---TTC-GATT-AGAGGGAGA--TG-GT-TCCCAGT-AGACCAC-GA--CCGCT--G--G.   8032
LHOEST.CD.524     GGA-AGCT-T-A-A-A-T.........AATGG-ACA-ACA---GGTTG-G-CA-A-CAGAA--AGA--G---TTC-GATT-AGA......--TG-GC-TCCCAGT-AGACCAC-GA--CCACT--G--G.   8035
MANDRIL.GA.MNDGB1 T-CA--GG-TGGC--CA--C-AA-AGAATCTTCACAG--TGC-G-----C-A-CATT...-CT--GG-G-CGTGC--AT-G-------TC-----G--T-----AA-TGCA---C-ACCCC-TAAC---.   8378

COLOBUS.CM.CGU1   -GG-GTC--AG-C-TCAGA-TCT----ACACTTC-G---TGGC--CG-TCT-CGGCG...C-T-GC--GC-TCGAGCG-CTG-ATAT-TGC--G--TGG-T-T-T-A-CGTCCCCAAGG-CCTGCC---.   8151
SYKES.KE.173      -GG-AA-T-CG-GATTGG--G--G-CTATTTGGC-AGCAA-ATAC-CC-CTACCAG-...C---T-GTCGAG--CG--GCAGC-T-G--GC-G--GC-TC----A--------G-TCTACC-TAAC-G-.   8723
RCM.NG.NGM        G-GATGG-GAGG-CTAAT-CTTACT-CACCTGGAGAG-AC-ATGC-CG--TTGCAG-...-ACC--GCAG--TGGCCAGC-ACGT-GTGC-----G--C--G--------C--G--AT--C-TAAC--G.   9390
MANDRIL.GA.14CG   C-GAGGCT-TGG-GC-G--GGCACC-ACTTGG-CTATCA-----TAG--G--C-GCG...-CTAT-G-A--G--C---G-GA----C---C-GC-C--C-----A--AGCT---GTCC-GC-CAAC---.   8838
MANDRIL.GA.14CG   _Q__R__L__W__S__G__G__H__Q__L__G__L__S__V__I__R__G__A__G__.__A__I__G__R__A__I__G__N__I__P__R__R__I__R__Q__G__A__E__V__L__L__N__$_    gp41
                                                                                                                                                                                  Env end /
```

HIV-1/HIV-2/SIV complete genomes

HIV-1/HIV-2/SIV complete genomes

```
                                 / Nef start
H1B.FR.HXB2                      _M__G__G__K__W__S__K__S__S__V__I__G__W__P__T__V__R__E__R__M__R__R__.__.__.__.__.__.__.__.__.__.__.__.__.__A__.__.__.    Nef
H1B.FR.HXB2   Nef G............ATGGGTGGCAAGTGGTCAAAAAGTAGTGTGATTGGATGGCCTACTGTAAGGGAAAGAATGAGACGA.....................................GCT........   8865

CPZ.US.CPZUS      ..............------AA---------T--G-----CA-AG-A--------AGAA--C--AA-C---T--.....................................AGA...CAAACTCAGAC   8897
CPZ.GA.CPZGAB     -CACTGCATTAAG.-----AAC---A-----T---------C--G-A---------GAG--C--AAG------A.....................................AGGGAA---.........   8938
CPZ.CM.CAM3       C.............-----CAA--G---------------CC--G-A--G-----AGA---C--A------C-A.....................................AGCGAACTC.........   8302
CPZ.CM.CAM5       C.............------AA-------------------C--G-G------G-AG----T--AC----G--A.....................................AGACAAA-..........   8608
CPZ.CD.CPZANT     AC............------TCTGCA-----T----TC-AGTG-G-----GCAAGACAA-C--TTAGG-A---A.....................................CATGAAA-A........   8325

H1A.UG.U455       C.............---------------------AG--CAGAG-G-A-------GAG--T---A-----------.......................................GAAA--..........   8316
H1C.ET.ETH2220    A.............-----G----C-AT-------T-----CCAG-A---------G--A----A--------A------.......................................---.........   8262
H1D.ZR.84ZR085    A.............-----------A---------------A-AG-----------G--A-------------A---AA-...............ACTGATCCAAGGGAAAGAAGAAGAC--.........   8410
H1F.BE.VI850      A.............-----G----------------C---A-AG----------G-----G--------------A-.....................................A-C.........   8142
H1G.SE.SE6165     A.............-----A--------------------CA-AG----------CGAG--------------A---AAC...................................A-C.........   8304
H1H.CF.90CF056    A.............-----A-----A--------------GA--GG---G---T-----A----------------G---.......................................---.........   8185
H1J.SE.SE92809    A.............-----GAA---A------------.............------CAG----------------G---.....................................GCTC-C.......   8163
H1K.CM.MP535C     A.............--------------------------A-AG-----------G--A--------------------.............GCACGACCAGCAGCAGACAGGGTGGGAA-A.......   8052
H101_AE.TH.CM240  C.............-----AA-T----------------CA-AG-G---------CAG--C-------A---A-AG.....................................CAAA--..........   8439
H1N.CM.YBF30      -.............-----AAAG-TT--------G--C--CC-AG-A--------AGAAA-C--A------------A--.......................CAAACGCAAGAACCA--A........   8452
H1O.CM.MVP5180    C.............-----GAATGCA---AGC-----C-AAT-TGCA------T-AGAA-----A--T------------.......................TCCTCCTCTGATCCTCAA........   8975
H1O.BE.ANT70      C.............-----AAA-GCA-T-AG----G---AAT-TGAG------G-AG-A-----A------------A--.....................ACTAGAACTTTCCCTGAGT--........   8951

MAC.US.SMM251     ..............................................................................................................................   9225
MAC.US.SMM239     ..............................................................................................................................   9243
MAC.US.SMM239     .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.   6803

STM.US.STM        ..............................................................................................................................   8894
SMM.US.SMM9       ..............................................................................................................................   8732

H2A.GW.ALI        ..............................................................................................................................   9282
H2A.SN.ST         ..............................................................................................................................   8728
H2A.DE.BEN        ..............................................................................................................................   9286
H2B.GH.D205       ..............................................................................................................................   9250
H2B.CI.EHO        ..............................................................................................................................   9227
H2G.CI.ABT96      ..............................................................................................................................   8671

TANTALUS.UG.TAN1  ..............................................................................................................................   8906
GRIVET.ET.GRI677  ..............................................................................................................................   8596
VERVET.KE.VER9063 ..............................................................................................................................   8913
VERVET.KE.VERTYO  ..............................................................................................................................   8367
VERVET.KE.VER155  ..............................................................................................................................   8890
VERVET.DE.VERAGM3 ..............................................................................................................................   8411
SABAEUS.SN.SAB1C  ..............................................................................................................................   9087

SUN.GA.SIVSUN     ........................-GTTCGC-CCTTATTCAG-C-AC--GCAGGAT-A-G--C-CGTGC-GAACC-GC-GGGAGAGAGTATCATCGTCTC.............................   9074
LHOEST.KE.LHOEST  ........................----AA--CGCTC--AAGAT-CA--GCTGGCT-A-GG--AA-TGC-GAAG--AC-GAGGACAGTTATCGAGCACTG.............................   9012
LHOEST.CD.447     ..............................................................................................................................   8035
LHOEST.CD.485     ..............................................................................................................................   8032
LHOEST.CD.524     ..............................................................................................................................   8035
MANDRIL.GA.MNDGB1 ..............................................................................................................................   8378

COLOBUS.CM.CGU1   ..............................................................................................................................   8151
SYKES.KE.173      ..............................................................................................................................   8723
RCM.NG.NGM        ..............................................................................................................................   9390
MANDRIL.GA.14CG   ..............................................................................................................................   8838
MANDRIL.GA.14CG   .__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.   6413
```

```
H1B.FR.HXB2        .GAGCCAGCAGCAGATAGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATGGAGCAATCACA...AGTAGCAATACAGCAGCTACCAATGCTGCT.......................TGTGCCTGG   8967
H1B.FR.HXB2   Nef __E__P__A__A__D__R__V__G__A__A__S__R__D__L__E__K__H__G__A__I__T__.__S__S__N__T__A__A__T__N__A__A__.__.__.__.__.__.__.__.__C__A__W_   Nef

CPZ.US.CPZUS       AACAG----------AG-A--A---C-T-TC----A----T-A-C-G-------G--T-----T...-C---A------C--CAG-A---CCAAA--.......................CTG--A---   9000
CPZ.GA.CPZGAB      .CCAA----------GG-A--A----A--TT--GAAG-----A----G---C-----T-----T..-----G--C--CC---AG--T---CAAA--.......................CTA--T---   9040
CPZ.CM.CAM3        .A-C-TGCTT---CCAG-A--A---C---TC--A-A----T-A--GGC----------G----C...-C---GC-C---C--CAG-AT---CAGA--.......................CTG------   8404
CPZ.CM.CAM5        .CCA---------TCAG-A--A---C---TC--A-A---T--A--GGC--G------------C...-C---A---CAC...A-A-AAT--CAGA-C.......................CTG------   8707
CPZ.CD.CPZANT      .A-T---.......---GATA-A--GC-CTGTGGAAAC--AT---CC-GTAGA--G---C-A--C...TCCTCT-C--T-...-G---AG-AAAA-A-.......................GTAAT-ACC   8418

H1A.UG.U455        .CCTG--------A-AG-A--A-------T-----A---TT-A--T---T--------G-----...--C--T------T--T----T-----CAG-.........................---------   8418
H1C.ET.ETH2220     .-CT-----------GG-A--A--------G---------T-A--C---T----G---C-T---...--C-----C---C-C--C-AT---C---A-.........................---------   8364
H1D.ZR.84ZR085     .---------------G------------T----------------G------G---------...---------------A----T--------C.........................---------   8512
H1F.BE.VI850       .CCTA----------AG------------TG---------T-A--C-G--GG--------T---...--C--------TAG-A----T---C---AC.......................-TG------   8244
H1G.SE.SE6165      .CCTA----------AG-A--A-------T-----A---TT-A--T-GG----------------...--C--------------C-AT---C---A-.......................---------   8406
H1H.CF.90CF056     .--A----T------AG----A-------TG--------TT----T-G--GC--G---G-------...-T--AT---------T----T---CG--A-.......................GCC------   8287
H1J.SE.SE92809     .-CT------------G-A----------TG----A---TT---CT--G-----G---------...--C-----------------A-------AC.........................---------   8265
H1K.CM.MP535C      .C-A-----------CG------------T-----A--------CT-G------G---G-----...--C--------TT--CAC-AT---C---A-.........................---------   8154
H101_AE.TH.CM240   .CCT------A----AG-A--A-------T-----A---T--A--T--------------A---...------T---......ATAGAT-------A-.........................----T----   8535
H1N.CM.YBF30       .-TAGAGC------TAG-A-CA--------T----A---T--A-CT--T-GA--G--C-----C...-TA--A-----TAG--AC-AT----AAAG-.......................ATA--T---   8554
H1O.CM.MVP5180     .C-A---TGT---CC-G-A--A-----T-TC--CA-G--GT-A-C--CTAGA--G-G---AT--...---TC-C-C--TC-TCAA-A------A--C.......................CT---A-TC   9077
H1O.BE.ANT70       .--A---TGC---CC-G-A--A---CAGATC--CA-G--AT-A-C-GCTAGA----GG--AC--...---TC-C----TC-TCAA-A------A--C.......................CT---A-TC   9053

MAC.US.SMM251      ......G-G-CAGA-ATACAATCAG-GGCAG-A-ATGA-TACTCC-TGGAGAAACC--GCTGA-...GAA-AAG-A-A-TT---ATA--GAAAACAA.......................AA-ATGGAT   9322
MAC.US.SMM239      ......G-G-CAGA-ATACAATCAG-G-CAG-A-ATGA-TACTCC-TGGAGAAACC--GCTGA-...GAG--AG-A-A-TT---ATA--GAAAACAA.......................AA-ATGGAT   9340
MAC.US.SMM239 Nef                                   _M__N__G__P__W__R__N__P__A__E__.__E__R__E__K__L__A__Y__R__K__Q__.__.__.__.__.__.__.__N__M__D_   Nef
                                                    \ Nef start

STM.US.STM         ......GC-TCAGAGATATAAT-A--GTCA--T-ATGA-TACCCCTTGGA-AAACC-GGCAG--...GAA---GC--AGCT--AATAT-GACAGCAA.......................AACATGGAT   8991
SMM.US.SMM9        ......GC-TCAGA-ATATTGT-AG-GTCA--TCATGA-TACCCCTTGGAGGAACC--GCA-----...GAAG-AGCA-A-TT--G-TAT-GACAACAA.......................AACATGGAT   8829

H2A.GW.ALI         ......--G-CAGCGGTATCA-CAG-G--AT-T-ATGA-TACCCC-TGGAGAACCC--GCA-----...GAA--AG-G-A--A-TTGTA---GCAACAG.......................AA-ATGGAT   9379
H2A.SN.ST          ......G-G-CGGCGGTATCAACAG-G--AT-T-ATGA-TACCCC-TGGAGA-CCC--GCAGA-...G-GGAG...-A--GCT-GTA---GCAACAA.......................AA-ATGGAT   8822
H2A.DE.BEN         ......G-G-CAGC-GTATCA-CAG-G--AG-ACATGA--AGCCC-TGGAGAAACC--GCA-----...GAA--AC-G-A--ATTTGTAT-GGCAGCAA.......................AA-ATGGAT   9383
H2B.GH.D205        ..................................................G-G-CAA-AT...GCCC-GGGGG---A--GGGGAGGGCAACAA.............................GAT   9293
H2B.CI.EHO         ..................................................T-G-CAA-AG...-C-TTGGGAG---A--GGGGAGGAAAACAA.............................GAT   9270
H2G.CI.ABT96       ......--GGCA-C-ATACAC-CAG-GTCA--T-ATGA-TACCCC-TGGAGAAACC--GCA-----...GAA--AGCA-AGTT---ATA--GGYAGCAA.........................AATAAT   8765

TANTALUS.UG.TAN1   ..............................................GAGG---AAG-G---C...CC-GAAGGA-G-AA-TTGG-AGC--A--AC.......................ACATGGGAT   8961
GRIVET.ET.GRI677   ...................................................A--TAA---G--T...CCAGA-GGCCGCCGCCTGCAAG-A-GA-AC.......................ACCTTTGAT   8648
VERVET.KE.VER9063  ......................................-A-G-GGA----CAAC--G--T...CCAGATGGA-G-CGCCTGCAAG-A-GA-AC.......................GAATGGGAT   8974
VERVET.KE.VERTYO   ......................................-A--C--A----CAAC--G--G...GAAGAGG-A-GGAGGCT-CAAG-A-GA-AC.......................ACCTGGGAA   8428
VERVET.KE.VER155   ......................................-A-G-G-A----CAAC--G--T...CCAGAAGGA-G-CGTCTACAGG-C-GG-AC.......................CAATGGGAT   8951
VERVET.DE.VERAGM3  ......................................-A-G-G-A----TAA---G--T...CCAGAAGGA-G-AA-TTACAAG-A-GA-A-.......................AAATGGGAT   8472
SABAEUS.SN.SAB1C   ....................................G-AA-G-G-CTCCTCTAG--CAGA-...GAAG-TGTA-A-ATGAAACATC-A-GAAGA.......................CAGC---C-   9151

SUN.GA.SIVSUN      ........................AG-AGGCAAGA--TAGAACCT.........TT-G--T-G...GCAGAA---GG--GCC-A-ATGGA.....................................   9128
LHOEST.KE.LHOEST   ............................AT-GAAAA-ACATCCA-TTGG.........GCA.........-AA-GA-G-......GATGGA.....................................   9051
LHOEST.CD.447      ......................................................................................................................   8035
LHOEST.CD.485      ......................................................................................................................   8032
LHOEST.CD.524      ......................................................................................................................   8035
MANDRIL.GA.MNDGB1  ..................................................................................AGATGAGAA..ACTTGAC-TAA-ACA-...............   8403

COLOBUS.CM.CGU1    ..........................................................................-TGCC-TTGCC-GCT-TGCTA-AGACATCCAGATGTATATTACCACAACC-AGAA   8206
SYKES.KE.173       ...................................................GCC--TTGACCCACA-GGGC-G-ACAGAGA-TGG-GACA-TCGGGAGGACAAAAATTCTCACCTG-A-ATATA   8796
RCM.NG.NGM         ...........................................................AGGAGGGTT-A-AATAGAAACTCAGCAGTCAGTGGATGCAATAGAT   9436
MANDRIL.GA.14CG    ......................................................................................................................A   8839
MANDRIL.GA.14CG    __.__.__.__.__.__.__.__.__
```

HIV-1/HIV-2/SIV complete genomes

```
                                                                                                                                                                                                                                |- 3'LTR U3
H1B.FR.HXB2        CTAGAAGCACAAGAGGAGGAGGAG......GTGGGTTTT...CCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGG   9088
H1B.FR.HXB2   Nef  _L__E__A__Q__E__E__E__E__.__.__V__G__F__.__P__V__T__P__Q__V__P__L__R__P__M__T__Y__K__A__A__V__D__L__S__H__F__L__K__E__K__G__G__L__    Nef

CPZ.US.CPZUS       T----T-AGATGACTA-TC-T--AAGTGAA--A--C--C...--T--TCGC-----A-----A-----G--C-----------ACA---AT-C---T-AG--TT--------------------------   9127
CPZ.GA.CPZGAB      -----G-A-ATG--CA-T--A......GAA--A--A--C...-----A-G---A-----T--AACG--------------T--A------T-T------TCA--------------------------   9161
CPZ.CM.CAM3        T------A-ATGC-A--AC-CA-TGAAGAA--A--C---...-----T-G------A------C-T-----T-----C-TT--A------T-T---T-A---------------A--------------   8531
CPZ.CM.CAM5        T------A-ATG--AA-TC-T--AGAAGAA--A--C---...-----T-G---C--------CC-------------A------CA----T-T---T-A------------------------------   8834
CPZ.CD.CPZANT      TATTC--A-G-TC-TACA-----A...GGGACA--C---...--T--G-G------A--C--CA-G-----C-----AG-A--ATT---------C--GTCATGN------------------------   8542

H1A.UG.U455        --G-----G--------A-GA--C........--A--C--C...-----T-GG--A--------------------------T---------T-T-----C---TT--------------------------   8539
H1C.ET.ETH2220     --G-----G-----------A--AGAG...--A--C---...-------G-----------------------------T--------AT-C-----C----T--------------------------   8488
H1D.ZR.84ZR085     G----------G--A-----C---GAG...------C---...-------G-------------------------------------------------------------G-A---------------   8636
H1F.BE.VI850       --G--------------------A........--A--C---...-------G-------------G-------------C--T-----------------C--T---------------------------   8365
H1G.SE.SE6165      --G-----------------CTCAGAA...--A--C---...-------G---A--A--------G-------------TT----GT---T-T-----C---TT--------------------------   8530
H1H.CF.90CF056     --G--------------C-G---GAA...--A--C---...-------GG--------------------------C--T----G----T-T-----C-----T------------------------   8411
H1J.SE.SE92809     --G---------ACA---------.......--A--A---...-------A-------A----A--------------A-------G------T---------TT--------------------------   8386
H1K.CM.MP535C      --G--------G-----A-----A........--A-------...-------G------------A-----------------T---------T-T-----CG--TT--------------------------   8275
H101_AE.TH.CM240   --GAG------------C------.......--A--C---...-------TG--G---------C----------------T----G----T-T---------TT--------------------------   8656
H1N.CM.YBF30       --------------A--A--A---GAA...--A--C---...-----ACGC-----------A-----G-----A--C--T--ACAG---T-T------TC-TT------------T-------------   8678
H1O.CM.MVP5180     -----CAGC--CA-A--T-----T........--A--C--C...-----A-G------A--G---C----G--------C-TT--A-----CT-T--C--C---TT------------------A-------   9198
H1O.BE.ANT70       ------AGT--CC-A-----A--A........--A-------...-----AG-------A--G---C----G--------C--T--A-G---AT-T--C--C---TT------------------A-------   9174

MAC.US.SMM251      GATAT--ATG-G--A--T--T--C...TTG--A--GG-A...T----G-GG--AA-A--T--CC-----G-------------ATTG--AA-----A-GTCT--T---A--------------------   9446
MAC.US.SMM239      GATAT--ATG-GT-A--T--T--C...TTG--A--GG-A...T----G-GG--AA-A--T--CC-----A------G------ATTG--AA----CA-GTCT--T---A--------------------   9464
MAC.US.SMM239 Nef  _D__I__D__E__$__D__D__D__._L__V__G__V__._S__V__R__P__K__V__P__L__R__T__M__S__Y__K__L__A__I__D__M__S__H__F__I__K__E__K__G__G__L__    Nef

STM.US.STM         GAT-T--ATG-G--A--T--TA-T...CTA--A--AG-A...G----ACAT--AAGA--C--AC-T--GGA-------------TTG--AA----C--GTCA--T---A-----AGC-------------   9115
SMM.US.SMM9        GAT-TG-ATA-T--A--T.........TTA--A----G-...-----G-GC--AAGA--C--CG-----ATC-----A-----ATTG--AA----CA-GTCT--T---A--------------------   8947

H2A.GW.ALI         GAT-T--ATTT---T--T--T--TAGCCTA--A--GG-C...T-T--------AAGA----AA------A-------A-----ATTG--A------A-GTCA--T--AA---------G-----------   9506
H2A.SN.ST          GAT-T--ATTC---T--T--T--C...CTA--A--GG-C...--T--------AAGA-----A------GA------A--T-G-TTG--AAG----A-GTCA--T--GA---------------------   8946
H2A.DE.BEN         GAT-T--ATTCT--T--T--T--C...CTAA-A--AG--...--T--T-----AAGA-----ACGG---GA------C--T--ATTG--AA-----A-GTCA--T---A----------A---------C   9507
H2B.GH.D205        GC---T-A-AGT--T-----C--TGAA...------AGCCATATGTAAG--C--AATA-----AC-G-----------A-----ACT---A-----CA-GTCT--T---A--------C------------   9420
H2B.CI.EHO         TC---T-AGG-T--T-----CA-TGAA...--A--GG-C...-GT--A-G---CGG---C--AC-G--G--------A-T---ACT---A-----CA-GTCT--T------------------A-----   9394
H2G.CI.ABT96       GAT--T-T-G-TAGT--T--TA-TGACTTA--A--AG-A...------TAC--CA-A--C--GC-----GT-----G------ATTG--AA----CA-GTCA--T---A--------------------   8892

TANTALUS.UG.TAN1   GACTGG-A--C------A--A--A.........-----C---...-----A-G---A-GA--G----------A------A--T---TTG--A------T-CTC---------------------A-------   9082
GRIVET.ET.GRI677   GAGTGG-ATG-T--T--A--A--A.......--A--C--C...--T--GCA-----GA--C--C-------AG-----C--T--ATT---A--G--CT--TC-------------TC--------------   8769
VERVET.KE.VER9063  GA-TGGT--G-T--T--A--T--A.........-----A---...--T--A-A---AAGA-----A--G--G-A-----------ATT---A--G--CT--TCG--------------------A-------   9095
VERVET.KE.VERTYO   GAGTGGT-GG-T--T-----A--A.........-----A---...-----G-G---AAGA-----C-----G-A---------T--ACTT--A--G---T--TCG--------------------A-------   8549
VERVET.KE.VER155   GA-TGGT--G-T--A--A--T--A.........-----A---...-----A-G---AAGA--G--AC------A---A--A-----ACTT--A------T--TCG--------------------A-------   9072
VERVET.DE.VERAGM3  GA-TGGT-TG-T--A--A--T--A.......--A--A---...-----A-G---AAGA--G--GC------A------C--T--ATT---G--G--CT--TCG----------------------------   8593
SABAEUS.SN.SAB1C   TGGT-T-ATG-G--T--A--A--A.......--A--C---...-----A-GG---TGCT----AC--C--G-------A--T---TTG--AA----CT--G----------------------------   9272

SUN.GA.SIVSUN      ......AT-G--C-A-----A-TAGAA..........--C...-----G-G---A---AG---AC-TT-TAAGCCA--A--T---CAGCTAA-------CTCA--T---A-------------T-------   9240
LHOEST.KE.LHOEST   ......-ACTC---TTC------AGAA...------A---...-----G-AG--A---AG---GC-TT-TAGCCCT--C--T---A-TCT-A----C--CTCA--T---A-------------T-------   9169
LHOEST.CD.447      ......................................................................CCCT--C--T---A-TCTGA----C--CTCA--T---A-------------T-------   8094
LHOEST.CD.485      ......................................................................CCCT-----T---A-TCTGA----C--CTCA--T---A-------------T-------   8091
LHOEST.CD.524      ......................................................................CTCT--C--T--AA-CTTGA----C--CTCA--T---A-------------T-------   8094
MANDRIL.GA.MNDGB1  -AG--TC--G----------A--A.......--T--A---...--T--GTGT-GC--A--TT-C------GTGCCAT-A-----A-ATCTGA----CT-CTCT--T---A--------------------   8524

COLOBUS.CM.CGU1    -C---G-ACGT---A--T-TA--A...GCA-GAT---A-......--G-TGT---GAC-G--AG--C---AGGC---CATT---CT-CTG--G--CA-CTCTTG------G-----------T-------   8327
SYKES.KE.173       G-GC-G-ATG---GA--CACA-GT...CTC--A-----C...--T--GTGT-----AAC---AC-C---A--T-A--C-----ATTG--AA----C--CTCT--T---A-----A-C-----A------C   8920
RCM.NG.NGM         TGGT-T-A-G-TACA--T--TACC...TTG--A--G---...-----G-G---A--A--C--CA----G------T-------ATT---AA-------ATCT---------------------------   9560
MANDRIL.GA.14CG    G----T-ATG-GC-AA-A--A--A......AG---C---...-----TTAT--CACACAG---G-GC-TGA-GCA--C-----AAACCTGA-----A-GTC---------------------------   8960
MANDRIL.GA.14CG                                                                                                                _M__S__H__F__L__K__E__K__G__G__L__    Nef
                                                                                                                               \ Nef start
```

```
                                                                Nef end premature in HXB2   \
H1B.FR.HXB2       AAGGGCTAATTCACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGG.....................GTCAG   9197
H1B.FR.HXB2   Nef E__G__L__I__H__S__Q__R__R__Q__D__I__L__D__L__W__I__Y__H__T__Q__G__Y__F__P__D__$__Q__N__Y__T__P__G__P__G__.__.__.__.__.__.__.__V__R   Nef

CPZ.US.CPZUS      -----T--G--T-T---AG------------G-----A-----C---G-T-----T--------AAT---------C-G------T-----C--------T--A......................--A--   9236
CPZ.GA.CPZGAB     -----T--G--T-----AGG-----------G--------C--C---G----T-------------T---------C-G-------------A----A-----A......................ACA--   9270
CPZ.CM.CAM3       -----------T-----AT--NNNNNNNNNNNNNNN-G----ACG--TA---T--------G----T-----------G-----------------------A......................-----   8640
CPZ.CM.CAM5       -----T--G--T-----AC--A---G-----G-----A-----C---G-T--T--------G--AATT--------C-G------T-----------------A......................-----   8943
CPZ.CD.CPZANT     -TN--------T-NN-A-TT-A----GC--C------A---A-------G-TTA----C--G--AAT-----------G------T------GA--A------A......................--T--   8651

H1A.UG.U455       -T---T---------A--G-A---------A--------------G----T-----------A-T-----------G------T-----------------A......................A----   8648
H1C.ET.ETH2220    -----T-----T-----A-G-A---G--G--G--T------T-----G----TA------------T-----------G---A--------------A------.......................-----   8597
H1D.ZR.84ZR085    --------G--T-------G-A---------------------T---G------------------T-----------G---A---------------------.......................A----   8745
H1F.BE.VI850      -----T-----T-----A-G-A----GG-----C-------------G-T---------------------------G-----------------------.......................A----   8474
H1G.SE.SE6165     -T---------T-----A-G-A---------G--------C------G-----A-T--------A--------A----G------------G------------.......................ACT--   8639
H1H.CF.90CF056    -T---T-----T-----A-GCA------G--C---------T-A---G----TA----------------------C-G-------------------------.......................-AG--   8520
H1J.SE.SE92809    -T---T-----T-----A-G-A---------G---------------G-TC-TA----------T-----------C-G---A---------------------.......................A----   8495
H1K.CM.MP535C     -T---------T-----A-G-A---------A---------------G-T--T-----------T-T-----------G---A---------------------.......................A----   8384
H101_AE.TH.CM240  -T---------T-----A-G-A---------G--------CT-A---G----TA-T---------T-----------G---A---------------------.......................A----   8765
H1N.CM.YBF30      --------G--TGG---AG--A---G--------T--A--C--C-----G--T------------AT-C-------C-G---T--------------------A......................A-T--   8787
H1O.CM.MVP5180    -T---T-----T-------T-AG---GC---A-----G-----C-----A--T-----T--G--A-T-----------G----TGT--------G--A-----A......................CCT--   9307
H1O.BE.ANT70      -----T-----T-------T-A----GC---A-----G-----T---G-G--TA----T--G--A-T-----------G-----------------A-----A......................AC---   9283

MAC.US.SMM251     -----A-TTA-T--AGTGC---------TAGA---T-A--CA---ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT----A-----A......................A-T--   9555
MAC.US.SMM239     -----A-TTA-T--AGTGC---------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT----A-----A......................A-T--   9573
MAC.US.SMM239 Nef E__G__I__Y__Y__S__G__R__R__H__R__I__L__D__I__Y__L__E__K__E__E__G__I__I__P__D__W__Q__D__Y__T__S__G__P__G__.__.__.__.__.__.__.__I__R   Nef

STM.US.STM        -----A-TTA-T--AGTG-G--------TAGA--AT-G--CA---ACT-AG-AA-AGA-G----GATAG-T--A----G------T------G----A------.......................A-A--   9224
SMM.US.SMM9       -----G-TTA-T--AGTATT--------TA-A--AT-A-----R-ACT-AG-AA-RGA-G----AATAA-A--A----G---A--T------G----------A......................A----   9056

H2A.GW.ALI        -----A-GT--T--AGTG-G--------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G----GATAA-T--A----G---------T--T-AT--------A......................A-A--   9615
H2A.SN.ST         -------GTA-T--AGTG-T--G----GTAGAG----A--CA-A-ACT-AG-AA-GGA-G-G--AATAA-TGGA--C-G---------T--T-AT--A-----A......................--A--   9055
H2A.DE.BEN        -----A-GT--T--AGTAGG--G-----TAGA-----A--CA-A-ACC-AG-AA-AGAGG----GATAA-A--A----G------T--T--T-AT--------A......................--A--   9616
H2B.GH.D205       -----A-GTA-T-TAGTG-G--------CAGA--AT-A--CACA-ATT-TG-GA-TGA-G-----ATTG-GT---GC-G---A-----T--T-AT---------.......................A-A--   9529
H2B.CI.EHO        -----A-TT-CT-TAGTG-G-----G--TA-A--A--A--CACA-ACT-AG-AA-TGA-G-----ATTG-GT---GA-G--------------AT--------A......................--A--   9503
H2G.CI.ABT96      -----A-TTA-T--AGTG----------TAGA-----A--CA-A-AC--GG-AA-GGAG-----TATAA-T--A----G------T------G----A-----T......................A-A--   9001

TANTALUS.UG.TAN1  -G--CA--TA-TGG--A-C--AG---G-GC-A-----GA----C-ATGC-CTAA-TGA-TGG---ATTA-AGA-----G---AGCA--TT-C--T--------C......................AC---   9191
GRIVET.ET.GRI677  -T---A--TA-T----TG--------G--A-G-----GA--T---ATGC--TGA--GAGTGG--AATAA-AGA-----G---AGCT---T-------C--G---.......................A-A--   8878
VERVET.KE.VER9063 -T---A-TTA-T-----G----G--GA-TAGA-----CA-C----ATGCTCTTA-TGA-TGG--AAT-A-TGA-----G-A-CGC--GGT--A----A-----C......................--A--   9204
VERVET.KE.VERTYO  -T---A-TTA-T-----G-T--G---A-TA-G-----GA------ATGCTCTTA-TGA-TGG--GATAA-TGA-----G-A-TGC--GGT-GAAG--A-----A......................A-A--   8658
VERVET.KE.VER155  -T---A-TTA-T-----G-T--G---A-TA-G------A------ATGC-CT-A-TGA-TGG--AAT-A-TGA-----G-A-CGCA-GGT--AA---A--T---.......................A-A--   9181
VERVET.DE.VERAGM3 -T---A-TTA-T-----G-C--G--GA-TC-G-----AA-C----ACGC-CT-A-TGAGTGG--AAT-A-TGA-----G-A-TGCT-GGT--GA---A-----A......................A----   8702
SABAEUS.SN.SAB1C  -----A-TTA-T--AGTG----G---A-GA-G---T-G-----C-ATGCTCTTA-TGAGTGG---ATAG-AGA--GA-G------T--T---GAT--------T......................AC---   9381

SUN.GA.SIVSUN     -------CTGGT-----AG--CC---G----A---T-G---T---ATGCAG-AA-TGAGTGG----T-A-AA---GC-G---AG-T-----CAA------T--A......................--G--   9349
LHOEST.KE.LHOEST  --------TA-TGG-----G--------G--C---T-AAT---C-ACTGTG-GA-TGAGTGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A......................ACA--   9278
LHOEST.CD.447     --------TA-TGG-----G--------G--C---T-ATG---T-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT--CT------GAT---------.......................ACGC-   8203
LHOEST.CD.485     --------TA-TGG-----G--------G--C---T-GTG---T-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT--C-------GAT--A------.......................ACAC-   8200
LHOEST.CD.524     --------TA-TGG-----G-----GA-T--C---T-AACA--C-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A......................ACTC-   8203
MANDRIL.GA.MNDGB1 G----A--TA-T-TAG-AGG------G----A-----A-----C-ATGCAG-GA-TGAGTGG--A-TTGAA----GA-G---AC-G--T--GA----T-----A......................AC---   8633

COLOBUS.CM.CGU1   -G---A-T---AGAA---CTGAG---G-T--CC-GA-A--G-A--ATGC----ATTGAGTGG----G---AAAA-G--G-TT-G-G--TGA-GAT-A--TG--AGAGGATGGAGCATTAAAAGAG-A---   8457
SYKES.KE.173      -G--CA-G-A-T-T-GTG---AG---G-T--A--T---C----C-ACC-AC-GA-TGAG--T--AAT-A-A...--CAG-ATC--T-----CT----------A......................AC-C-   9026
RCM.NG.NGM        -----A-TTA-TT---AG----G---AGGCGC--TT-A--C--C-AC--GG-GA-TGA---T--AATAA-A--A----G------T-----T-----------.......................A-A--   9669
MANDRIL.GA.14CG   -----A-TTGGTT---TAG-------G----G---T-AA-C----ATGCTC-GA-TGA-TGG--A-T-A-A--A----G---AG-A-----TT----------.......................A-A--   9069
MANDRIL.14CG  Nef E__G__I__W__F__S__R__R__R__E__E__I__L__N__L__Y__A__Q__N__E__W__G__F__I__P__D__W__Q__E__Y__T__S__G__P__G__.__.__.__.__.__.__.__I__R   Nef
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2   Nef __Y__P__L__T__F__G__W__C__Y__K__L__V__P__V__E__P__D__K__I__E__E__A__N__K__G__E__N__T__S__.__L__L__H__P__V__S__L__H__G__M__D__D__._   Nef
H1B.FR.HXB2       ATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGATAGAAGAGGCCAATAAAGGAGAGAACACCAGC...TTGTTACACCCTGTGAGCCTGCATGGGATGGATGAC...  9321

CPZ.US.CPZUS      ------C--A--A-A---G-------TT---T----C--TC-CACAGA---GG-AG----GC------------------C--A--TA...C--C-G-----CA--T---A------A-----A--T...  9360
CPZ.GA.CPZGAB     --TC-----CTGT-------------T---A-----G--CC-GACAGAG--GC--G-----C-A------G-------T----A-T--...C----G--T--CA-AT-T-A------------A--T...  9394
CPZ.CM.CAM3       ---C-----A--A-A---GG------TT--A--G--T---C-GACA-----AG-AG-G---C--------G-------T--T-A--T-...C-AC-G------CA-CT---A------A-----A--T...  8764
CPZ.CM.CAM5       ---C-----A--A-------------T------G-----TC-GACA-----GG-AG-G--GA-A-----CG-------C-C--AT-TA...C--C-T-----CA-AT---AA-----A-----G--T...  9067
CPZ.CD.CPZANT     ---------CTGTAGA--------T-TT---T----------A--AT--CCACCTGAT...............-AT----GG-A--TA...C-CC-C-------CTT-TACA-----AGAT-GA---...  8760

H1A.UG.U455       ---C-----A--A--------------------------------T----C-G-AG----G--A--T-C-GG-----------A---T...C-AC--------A-AT---AA-----AG-A-----T...  8772
H1C.ET.ETH2220    -------------------------T-----------------C---AG-G-AG-G-----AAT----G-----------A-T--...---C-G-------C------------A-----G--T...  8721
H1D.ZR.84ZR085    -------T-A--------G-------T-G----------------T-----GGT-G-----A------C-G-------AG---A-T--...---C--------A-CT---A------A-----G---...  8869
H1F.BE.VI850      ---------------G----------T-----------------T-----AG--G-G---A--------G------------A-T--...------------CA------AA-----A-----G--T...  8598
H1G.SE.SE6165     --TC--C-------------------T---A---------A-G--T----CAG--G----G--A-------------------A---T...C-A--------CA-CT---A------A-----G---...  8763
H1H.CF.90CF056    --T---C-----------G-------T---------------AA-T---C-GG--G-----C--------G------------A----...---C-------CA-------------A-----G--T...  8644
H1J.SE.SE92809    ---C---T----A--------------------------------T---AGCG-AG----G--A--T---G-------A----A-T--...C----------CA-AT---A------A--A--G--T...  8619
H1K.CM.MP535C     ---C--------------G-------------------------T----CAG-AG--------A---CAG--------G---A-T--...C----------CA-A-A--A------A-----G--T...  8508
H101_AE.TH.CM240  --TC------TGT-------------T-----------------C-A-AGAG-AG----G----A---C--------A----A-T--...C----------CA------A------A--A--G---...  8889
H1N.CM.YBF30      ---C--CG-A----------------T---A---------T-GTCAG-T--AG-AG----------T---G-------C----ATGC-...C-C--------CA-AT-T-AA-----AGCA-----T...  8911
H1O.CM.MVP5180    --TC--------A----------TG-TT--A--G--------GTCAG----AG--GC---GAGACTGGG---TAC-A-TG-AGATGCTAGTC-TC----T--A-CTT-TAAT-----AGCT--G--T...  9434
H1O.BE.ANT70      G-TC--------A----------TG-TT--A-----------GTCAGA---AG--GC----AGACTAGGA--TAC-TGTG-G-GGGCTAATC-CC-G--T--A-CAT-TGCC-----AT-T--A--T...  9410

MAC.US.SMM251     ---C---AA---A-----C---CTA-GG--AT----C--T--AA-TGT-TCAG-TGAG-C-C---AGG--G-GA-G......CATTAT....--AA-G--G--A-CTCAAACTTCCAA-TG-------...  9673
MAC.US.SMM239     ---C---AA---A-----C---CTA-GG--AT----C--T--AA-TGT-TCAG-TGAG-C-C---AGG--G-G-AG......CATTAT....--AA-G--T--A-CTCAAACTTCCCA-TG--------...  9691
MAC.US.SMM239 Nef __Y__P__K__T__F__G__W__L__W__K__L__V__P__V__N__V__S__D__E__G__Q__E__D__E__E__.__.__H__Y__.__L__M__H__P__A__Q__G__S__Q__W__D__D__._   Nef

STM.US.STM        ---C---AA-CAG---------CTA-GG--A-----------A--TATGTCA--TGAG-C-C---AGG--G-T--GACA...CATTAT...C--G-G--T--A-CACAGACA---CA-TG-------...  9345
SMM.US.SMM9       ---C--TA--TT----------CT--GG--AT---WC--G--AA-TGTCTCAG-TGA--CTC---AAG-CG-GAC-......CATT--......--A-G--T--A-CACAAACTTC-CA-TG--------...  9174

H2A.GW.ALI        G--C--GA--TT------G---CTG-GG--------------A--TGTCCCAC-AGA--GG----A--C-G-GACT......CA-T--...C--C-------A--ACAAACAAGCA--CAT-------...  9733
H2A.SN.ST         G------AA-TT------G----TA-GG--------------A--TGTCCCAC-AGAG-G---T-A--G-G-GACT......CA-T--.....--AG-G--T--A-CACAAACAAGCA--T-T-------...  9173
H2A.DE.BEN        G--C---A--TA---C--G---CTG-GG---------T----A--A-TCTCAC-AGAG-C-----AAG--G-G-CC.......-A-T--.....--AG-------A-CACAAACAAGCA-ACAT-----T...  9734
H2B.GH.D205       G-----CAAATA------T---CTG-GG-----G--------A---GTGCCAGCAGCGACCCGA-AGG-GG-G-A-ACC...CATT--...C-AA-G-----G-CACAGA-CTCCTCATG-------...  9650
H2B.CI.EHO        ------CAA-TT------C---CTA-GG--------G---A-AA-TATGATAGCAGA-CC-----A-G-GG--ACC......CATT-T...C--G-G--T--A-CACAGACCTCC-CATG--------...  9621
H2G.CI.ABT96      -------R----G-----C---CTA-GG--A-----------CAC-GTCTC-G-TGAG-C-C---AGG-CG-GACT......CATT--.....--G----T--A-CACAAACTAGCCCATG--------...  9119

TANTALUS.UG.TAN1  GA-G--C-GATG---C--C-TT----TTG--T----G--G--A--TGTCTCCC--GA--CCC---ATG-G-G-.........CA-T-T.....-C-C-------CCCAGA-TG-AT--GAAAGC---...  9306
GRIVET.ET.GRI677  G--C--GAGAGT------C-TC----TT--------C-----G--C-TGC--G--GAG-C-CGCAA-TG-G-GA--......CA-T-T...C--A-G--T--A-CACAGA--GGG-AAGATCC---TGGA  8999
VERVET.KE.VER9063 ------C-GCTG------C-TT----T------------G--A---ATGC--G-AGAG-C----A--TG-G-GA--......CATTA-.....-GCG--T----CACAAA-AA-A-A-GACCCC--TGGC  9325
VERVET.KE.VERTYO  --TC--TAAATG------G-TC----TT--------G-----G--CTT-C--G--GA--C-C-AA-ATG-G--A--......CATT--...C-AG-C--T--A-C-CAGA--GGA-AAGATCCA--TGGT  8779
VERVET.KE.VER155  ---C--GAG-TG------C-TC----T------------G----CC-TGC--G--GA--C---AA-ATG-G--A-G......CATT--......-G-------A-CACAA-------AAGACCC---TGGT  9302
VERVET.DE.VERAGM3 ---C---AGATG---C--C-TC----TT--AT-G--------A--C-TGC--G--GA--C----A-TTG-G-GA--......CATT--...C--G-G--T--A-CACAAG--AGG-AAGACCC---TGGA  8823
SABAEUS.SN.SAB1C  G--C---AA-TG--------------TT--A--G-----T--G--C-TGTCAG--GAG-CTA--AAT-G-G-GAAC......CA-T--...C--C-G--T----CTCAAG--GCCTATGA--------...  9499

SUN.GA.SIVSUN     -------AA-G-------C---CTA-GG-------C------CACCATT---G-AGACAGG--T...........................CCTAAT--T---TGTCAAGCC-TACT-CATTCCAG-CAG  9452
LHOEST.KE.LHOEST  ---C--T-----A----------TG-GGC--T---A------G-CTTGT---G--TATA----TC--TCAG-T-AGACA...CAAT--...C-T--G--TT-AAGCCAG---GGG-TCC----A---...  9399
LHOEST.CD.447     T-TC---T-AG-A----------TG-GGC----G-A------G-CCTGT---G--TACAC---TC--TCAG-TTACAGA...CAGT--...C--C-G---T-AAGTCAG---GGG-TCCA---G---...  8324
LHOEST.CD.485     T--C---T-A--A----------TG-GGC----G-A------G-CCTGT---G--TACAC---TC--TCAG-T-CCAGC...CAAT--...C--C-G---T-AAGCCAG---GGG-TCCA---G---...  8321
LHOEST.CD.524     T--C-----CG-----------CTG-GGC--T-G-AG-----A-CATGT---G--T-TA----TC--TCAG-T-AGAGA...CAAT-T.....-C----TT-AAGCCAGT--GGG-TCC-A--A---...  8324
MANDRIL.GA.MNDGB1 ------TAA---A-------TCCTG-TT-----G-A------GAGCAG--C--TAGG---T---...T--GC--CTA-C--TCAT......C--------T-CTCCCAGT-ATG-CCTCA---A--T...  8751

COLOBUS.CM.CGU1   -A-G--T---GTTGC-------CTT-GG--AT----CTACA-A----A-CT-GGAGAGT-T-CATA--GCT-T.............................................................  8530
SYKES.KE.173      G--C--GT-A-T-----------TA-GGG-A-----T---AA----AT---GGGATGCTTG---TATG-GG-G.........CAT-CACTCC-CC-C--T--A......GCCTC---ACA--GGAGTTCG  9141
RCM.NG.NGM        -------ACA-TG-----C---CTA-GGC-AT----------A--TGTCAG-G-TGA--C-ACA-AG---G-GA-G......CATT-T...C-CC-G--T--A-CAGAAACTAG---------C------...  9787
MANDRIL.GA.14CG   ---C--GAA--GA-----G-TC-TA-T----T--A-C-----AC-AGT-CCACCAGACC-G---AA-C-GG-GT-TA-C-GG.........C--C-TA--T-CTCTCAG--AGGAATCCA---A--A...  9187
MANDRIL.14CG  Nef __Y__P__K__R__F__G__F__L__F__K__L__I__P__V__Q__V__P__P__D__Q__E__N__Q__E__C__N__R__.__.__.__L__L__N__S__S__Q__L__G__I__Q__E__E__._   Nef
```

```
                                                                                                                                                                 Nef end \
H1B.FR.HXB2   Nef _.__.__.__.__.__.__P__E__.__R__E__V__L__E__W__R__F__D__S__R__L__A__F__H__H__.__V__A__R__E__L__H__P__E__Y__.__F__K__N__C__.__$__.__    Nef
H1B.FR.HXB2       .................CCGGAG...AGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCAC...GTGGCCCGAGAGCTGCATCCGGAGTAC...TTCAAGAACTGC...TGA....   9417
                                                                                                                                         TCF-1 alpha
CPZ.US.CPZUS      .................GAAC-T...G-C-----C---ATC---CA--------CTGAA--G--TCGGAGA---...AGA--TAA------------A------...---CG------ATGGC......   9457
CPZ.GA.CPZGAB     .................GA---C...-A---G--C--G-TC---C-C---------A-G--G--C--AAGA--T...A-T---A----A-AA-------------...-A---AG----ACTT-AGACTG   9497
CPZ.CM.CAM3       .................G-AC-T...G-------C--G-TC-----A---------A-G--G--C--AAGA---...A-A---A----A-AA----------T...-A-C-------A..........   8857
CPZ.CM.CAM5       .................GAAC-T...G-------C----TC-----A---------A-G--G--TC-GA-A---...C-A---A----A-AA----------T...-ATGGAG--CA-AACCC-GCGC   9170
CPZ.CD.CPZANT     .................--AC-T...-A---GA-C---AGA---GAA-----TGCAA----GATGAGAAGA---...A-T--AA----AAGA--C--T--A--T...----GAG---AAGAC--C....   8859

H1A.UG.U455       .................GAA---...-A----------AT-----A---------TAC---G-----AA-A---...AGA--TTAT---------------T-...-AT--AG----A..........   8865
H1C.ET.ETH2220    .................GAA--C...---------A---A------A---------T-A-------CGCAGA---...A-------C-----A----------T...-A---AG-----..----....   8817
H1D.ZR.84ZR085    .................GA-----...-AG----------T------A---A-----A-A---------G-G---...AA----AA---AAA-T-----------...----------------...---CATC   8969
H1F.BE.VI850      .................GAA--C...---------AC-GAG-----A----------A----G---C-GAGA---...A-A---A-----AGA----------T-...-A-C-AG----A..........   8691
H1G.SE.SE6165     .................GAA--C...---------C-G-T------A--------TA--------CGGAGA---...A-A-------------------------...-A---AG-----..----....   8859
H1H.CF.90CF056    .................GAC-G-...---------C-GAT-----AA--------T--A--------GACA---...T---------TAAA-------------.....--AG-----..----....   8737
H1J.SE.SE92809    .................GAA--A...------------C------A----------TC-------CGGAGA---...A-A-----------------------TT...-A---AG-----..----....   8715
H1K.CM.MP535C     .................GAAC-C...------A-A---AT-----A----------TTC-------CGAAGA---...--A--------------------C---...-A---AG--.............   8598
H101_AE.TH.CM240  .................GAA--A...---------C-GAT-----A----------TGC-------CGAA-A---...--A--------A-A------A------...-AT--AG-----..----....   8985
H1N.CM.YBF30      .................GATC-T...-A---------G-T----C-A---------TC-------AGAAGA--T...--A--AA-----------------TT...-A-----------..----....   9007
H1O.CM.MVP5180    .................G-AC-C...G-G--GA-AC--A-A---CA------T--ATCAT---GC--AACA--T...A-A----TGC-AAA---C--A---CT-...----CCC--G-AA..........   9527
H1O.BE.ANT70      .................A-AC-T...-A----A-AC-GAT-----A------T--ATCT----GCAACACC--T...--T--TATGATAACT--C--A---CT-...----C----GGA-...-A-....   9506

MAC.US.SMM251     .................--TTG-...G----G--TC---C-----A------TCCAACT-----C-ACAC-T-T...-A---ATAT-CTAGAT-C--A--AG-G...--GG-AGCAA-T...CAGGCCT   9773
MAC.US.SMM239     .................--TTG-...G----G--TC---CA----A------TCCAACT--G--C-ACAC-T-T...-A---ATAT-TTAGAT-C--A--AG-G..T--GG-AGCAA-T...CAGGCCT   9792
MAC.US.SMM239 Nef _.__.__.__.__.__.__P__W__.__G__E__V__L__G__W__K__F__D__P__T__L__A__Y__T__Y__.__E__A__Y__V__R__Y__P__E__E____F__G__S__K__.__S__G__L    Nef

STM.US.STM        .................--TTG-...G-------AC-G-TT----AA-----TCCA-TAT----CCA-AC-T-T...-A----TTT-TTAGA--C--A--AG-G...--TGGTTCTAAG...-C-GGCT   9445
SMM.US.SMM9       .................--CTG-...G----G--AC-G-CA----A------TCCARAGT----T-A-R-CT-T...AA---ATTT-TTGA---C--A--AG-G...--TGGT-GTCAG...-C-GGCT   9274

H2A.GW.ALI        .................A--C-T...G-G--GACA----TT-----A------CCTAAG--G--TCA-G--T--...AAA---TTTATT--A--C--A---G-A...--TGG-T--AAG...-C-GGCC   9833
H2A.SN.ST         .................---C-T...G-----ACA----TT------------CC-ACG-----T---AGCT--...-A----TTTATT-GAT-C--A---G-G...--TGG-T--AAG...-C-GGCC   9273
H2A.DE.BEN        .................GA-C-T...G-G--GACA----T----CA-------TC-ATG--G--C-A-A-CT--...AA----TTCACT---T-C--A--AG-G...--TGG-C--AAG...-C-GGAT   9834
H2B.GH.D205       .................ATCC-T...G-G--GACTC-TATC---CA------TTC--T---G----A-G--T-T...------TTTCA-TAG-TT---A--AG-G...---GG-T--CAG...-C-GGGC   9750
H2B.CI.EHO        .................--CC-C...GA---GACCC-T-TC---CA-------TC--T--------A-G-CT-T...------TTCAGCAG-TTC--A---G-G...--TGG-T-TCAG...-C-GGAA   9721
H2G.CI.ABT96      .................-------...-CC--------G-C-----A-------CCTACAT----C-A-G-CT--...AG---ATTTATAT-------T---G-G...----G-TGGAAG...-C-GGGC   9219

TANTALUS.UG.TAN1  .................--TTG-...-A---GACA--G-T-----AA---A-TCCA-T-------G----GT--...AACC-GGACAGCT-TA-GGACAT-C-T...GGTCTTGTTAAG...A-GAAGT   9406
GRIVET.ET.GRI677  ATA..............GATC-T...G-------C--G-TC----A-------CCGAAGT-G--GG-GG-GT--...CGCC-GGACAT-T-TA-GGACAT-C--...GAAC-TGCAAAG...C-CTAG.   9101
VERVET.KE.VER9063 ATC..............AGTC-T...G----GACC--G-TC----A-------CCTATGT-----G-G--GT-T...-ACC-AAACAGA-A-T--TTT--AG--...A-GC-TGCAATA...GTGAAGA   9428
VERVET.KE.VERTYO  ATC..............AGCC-T...G----GA-C--G-T-----A------TCCTATGT-G---A-A--GT--...-ACC--AATCG-GA-T-CTTTACTG--...A-GC-TGGGCTG...GTGAAGA   8882
VERVET.KE.VER155  ATA..............AATC-T...G-----A-A--G-CA----A------TCCAATGT-G--TG----GT--...-ACC--TC-AG-GA-T-CTTTACAG--...--AT-TTCAACA...GTTGGTA   9405
VERVET.DE.VERAGM3 ATC..............AACC-T...G-------C--G-TC----A------TCC-ATGT-----G-C--AT-T...-ACC-TAACAGAAAAT---TCACTG--...A-GC-TG-TCTT...G-CAAGA   8926
SABAEUS.SN.SAB1C  .................G-CTG-...-A---GACCC-G-T-----A-------CCA-TA------G-GG-CT-T...--T--TT-GAG------C--A---C-G...G-GCC-TCAGCA...CA-GGCT   9599

SUN.GA.SIVSUN     CAAGGGGTTAATGAAGAC--TTG-...G----GAGAC-TATC----CA-----TCC-ACT--G--C-A-G--TTT...AG----AT-C-AAA---C--T--AG-G...------C-TGTT..........   9563
LHOEST.KE.LHOEST  .................--CTG-...G----GAG-C-TATC---CAC--CA-TCCAATGT-G---G-GG-CTTT...A-A----TCA-AAA---G--T-C-A-G...A-AC-A--TACTGCT-TCGCTT   9502
LHOEST.CD.447     .................--CTG-...G-----AG-C-TATC---CAC---A-TCCTATG------G-GG-CT-T...--A----TCAGAAAA--A--TA-AAG-...A-AC-AGCA-CAGCTATTGGCT   8427
LHOEST.CD.485     .................--CTG-...G----GAG-C-TATC---CAC--CAGTCCTATG------G-GG-CTTT...--A----TCAGAAAA--A--TA--AG-...A-AC-AGCA-CATCTCTTAGCT   8424
LHOEST.CD.524     .................--CTG-...G-----AG-C-CATC---CAC---A-TCC-ATGT-----G-GG-CT-T...--A----TGAGAAA---A--A-CAAGT...G-TC-AGCAACTGCTATGAGAT   8427
MANDRIL.GA.MNDGB1 .................--A--A...G----GACCC-CAT----TCTGGGAC-CT-AT-T-GC-TA-GAC-TTG...CA-CATTA-C-TA-CACC-T-AGTGTT...-CA-TA-GGCTAAGAGT-TTGA   8854

COLOBUS.CM.CGU1   ....................--C...TTGAGTC----GAGCGTT-CAAGCAGG-AGAAGAAGAA-CCA...............-A-C--G-TGCCATA---ATG...G-GG-CTGA.............   8605
SYKES.KE.173      AGC..............AT--GA...GA-CCTCATG-G--ACTTCAACCCCCACCTG--TATA--CCAGG-TGG...-A-ATGGCCCGCT----GTTA---AGA...CAA-CAGGAAAACCACA-GAGC   9247
RCM.NG.NGM        .................--CTG-...G-G--G--A----CA----A----A-ACC--T-T-----G-AG--T--...-CA-GGTATAGA--------T----T-...--TGGA................   9874
MANDRIL.GA.14CG   .................T-CAT-G..G-G--GAG-C-GAT-----A-------CCTGAAT-G--T-A-ACATTT...TAT----CCATAAAATGC--AA--G-A...-A--G-C-TGTT...AC-AGCC   9288
MANDRIL.14CG  Nef _.__.__.__.__.__.__S__M_  _G__G__E__A__D__V__E__V__$_                                                                        Nef
                                                      Nef end /
```

HIV-1/HIV-2/SIV complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
H1B.FR.HXB2        ..........................................................................................CATCGAGCTTGCTACA   9433

CPZ.US.CPZUS       ......................................................................TGTAACCGC.GCAGGCGCAATA.A-A--GCTG--T   9490
CPZ.GA.CPZGAB      GCGCATGCGCACAAGAACTGC.................................................TGACTCTG...................-........   9527
CPZ.CM.CAM3        ......................................................................TGCCCAGCG.CCTG.CGCAGTAGA-A--GCTG---   8890
CPZ.CM.CAM5        CTGCGCAGTAGAACTGC.....................................................TGACAA.............................   9193
CPZ.CD.CPZANT      ......................................................................TAACAGGCG.GATATAAAA-CG-TGA-AGAGC-G-   8893

H1A.UG.U455        ..........................................................................AAG.ACTG.CTGA--CA--AG-----GAC   8892
H1C.ET.ETH2220     ..........................................................................CACAGA--G-AC-TTCCGC   8836
H1D.ZR.84ZR085     GAGCTC.................................................................................................   8975
H1F.BE.VI850       ......................................................GACTGCTGACACAGAGATT.GCTG.ACACAGAA--ATC-..A-A-   8732
H1G.SE.SE6165      ..........................................................................CACAGAA-TTGC--AC-AG   8878
H1H.CF.90CF056     ..........................................................................CACAGAA--T.C-..G-A-   8753
H1J.SE.SE92809     ..........................................................................CAAAGAA-TTTC-..AG-G   8732
H1K.CM.MP535C      ......................................................................................................   8598
H101_AE.TH.CM240   ..........................................................................CAAGGAA-TTTC-.ACT..   9001
H1N.CM.YBF30       ..........................................................................CAAGGGACTTTACTGCTG-   9026
H1O.CM.MVP5180     ..........................................................................................CTGACAC   9534
H1O.BE.ANT70       ..............................................AAACTGCTGACCTGAAGATTGCTGACACTGT.............................   9537

MAC.US.SMM251      GTCAGAGGAAGAGGTTAGAAGAAGGCTAA.........................................CCGCAAGAGGCCTTCTTAACAT.GGCT.GACAAGA............   9835
MAC.US.SMM239      GTCAGAGGAAGAGGTTAGAAGAAGGCTAA.........................................CCGCAAGAGGCCTTCTTAACAT.GGCT.GACAAGA............   9854
MAC.US.SMM239 Nef  _V__R__G__R__G__$_                                                                                          Nef
                              Nef end /

STM.US.STM         TGCCAAAGGAAGAGGTTGAGAGAAGGCTAACCGCAAGAGGCCTTCTAAAGATGGCTGACAAGAAGGAAACAAGCTGA.......................-ACA--A...   9529
SMM.US.SMM9        TGTCAAAGGAAGAGGTGCAGAGAAGGCTAACCGCAAGAGGCCTTTTAAAAATGGCTGACAAGAAGAAAACAAGCTAG.......................-ACA--A...   9358

H2A.GW.ALI         TGCCAGAAGATGAGTGGAAGGCAAGACTGAAAGCAAGAGGGATACCATTTAGTAAGAACAGGAACAGCTGATTTGGTCAGGGCAGGAAGTAACTACTGAA..........AACAGC-GAG.AC---A...   9949
H2A.SN.ST          TGCCAGAGGATGAATGGAAGGCAAGACTGAAAGCAAGAGGGATACCGTTTAGCTAAAAACAGGAACAGCTATACTTGGTCAGGGCAGGAAGTAACTA............ACAGA-AAC-.GC--AG--T   9389
H2A.DE.BEN         TGCCAGAGAAAGAATGGAAGGCAAAACTGAAAGCAAGAGGGATACCATATAGTGAATAACAGGAACAACCATACTTGGTCAAGGCAGGAAGTAGCTA.............CTAAG-AAC-.GC--AGG-T   9950
H2B.GH.D205        TACCAGAGGAGGAGTGGAAGGCTAGACTAAAAGCAAGAGGGATACCTACAGATTAGGCAAGAGACAGCAGCATAAACAGGAACTAGCTGACACTGCACAA.........GAAGG-AACTAGCA-AC--T   9870
H2B.CI.EHO         TGCCAGAGAAAGAGTGGAAGGCTAAACTGAGAGCAAGAGGAATACCTACAGAGTAGACAGGAAATAACAGCACAAAGAGGAACTAGCTAACACTGCATAGA.........GAAGG-AACTAGC--A---T   9842
H2G.CI.ABT96       TACCAGAAGCAGTATGGAAAGAGAAACTAAAACAGAGAGGACTGCCTATAGAATAAGCAGAAAGAGG..............................AC-AGC--AC---   9299

TANTALUS.UG.TAN1   AG.....................................................................CCTAACCGCAGGCTTGCGGTTAAGCTAACTAGGC.CGTT.GCCTAGGA--TGACATA-AG   9466
GRIVET.ET.GRI677   ......................................................................TGTCAGC.ACTT.TGCGGT-.............   9119
VERVET.KE.VER9063  GGAAGTAG...........................................................CTAACCGCAGGCTTGTGGTTAAGCCGTTGCCAG.GGAGATGAC-TT-GAAACTG   9489
VERVET.KE.VERTYO   GGAAGTAG...........................................................CCAGACCGCAAGCCTGCGGTTAGAACAT.CACC.ATGGAGAT--CA--AAA-AC   8942
VERVET.KE.VER155   CAGGAAACTAG........................................................CCGACCACAGGCTTGCGGTTTCCTGGTTGCC.TAGGAGAT--CA--AAG-AC   9467
VERVET.DE.VERAGM3  GGAAGTAG...........................................................CTAACCGCAGGCTTGTGGTTAAGCACATCACCATGGTGAT--CA--AAG-AC   8986
SABAEUS.SN.SAB1C   AA.....................................................................CCACAAAACCACA.TCCT.ATGGAG-T-TCA-G-TG-TG   9638

SUN.GA.SIVSUN      ..ACTAGTCTGCAGTGGGAGGTGTAGCTTAACCGCAAACCGCATCCTCTTGCATCGCCTAGGCAACG.....................................GGGCT-G--CATGC--GCT-   9648
LHOEST.KE.LHOEST   TT...............GACTGCAAGAGGAAGTAGCCAAACCGCAAACCGCAGCACATCCTCTTGCAGCCCGGTTGCTAAGGCAACCGGGCTAGCGCATGCGC.......AATGGC-T-CTGA.....-G   9605
LHOEST.CD.447      TG....................CAGAGGAAGTAGCCAAACCGCAGACCGCAACACATCCTCTTGCAGCCTGGTTGCCAGGGCAACCAGGCTAGCGCATGCGC.......AGGA-T-T-CTGA.....TG   8524
LHOEST.CD.485      TG...............AACTGTAAGAGGAAGTAGCCAAACCGCAGACCGCAGCACATCCTCTTGCAGCCTGGTTGCCAGGGCAACCAGGCTAGCGCATGCGC.......AGGA-T-T-CTGA.....T-   8527
LHOEST.CD.524      TTAACTGCGAGAGGAAGTAGCCAAACCGCAAACCGCAGAACATCCTCTTGCAGCCCGGTTGCCCTAGCAACCGGGCTAGCGCATGCGC.......................ACAAGC-T-CTGA.....AC   8530
MANDRIL.GA.MNDGB1  ACATCTGCCATTTTGGAAGAGGAAGTAGCCTAACCGCAAAACCACATCCTACTGCAGAACTGTAG.TTGCTTGGCAACCTGCTTAGCAACCTGGACTGGCGCTTGCGCG.CTAGG-A............   8969

COLOBUS.CM.CGU1    ..........................................................................G-T-CTGA....GAC   8616
SYKES.KE.173       TGCAGTCAGCACTTAGTAAGAACATCTCCTAGGAG...ACTCCATGGTGACAAGCTCGGCCCAGGGGA.....................................................   9312
RCM.NG.NGM         ........GAGAGAAAGAACAAGACCAAGTAACCA...CTTCCTCTGGAGTTGCCTTGGAAACAGCGCATGCCTACT..............................GAGAAGAT-CTGA....CAC   9956
MANDRIL.GA.14CG    TCTCCTATACTGCCTACCAGGAG......AAATCAGACTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCTAACCACAGAAGGTTGCTATGGAGATGCTCTTT..........AATTCA-CTGA....CTC   9398
```

```
                                NF-kappa-B-II                    NF-kappa-B-I |-       Sp1-III        -|  |-  Sp1-II  -|   |-  Sp1-I -|
H1B.FR.HXB2        A...........GGGACTTTCCG.CTG..................GGGACTTTCCAGGGAGGCG.TG...............GCCT..GGG...CGGGACT.GGGG.AGTGGCGA......G......   9496

CPZ.US.CPZUS       G...........A--------TAA.......GGGACTTTCCAA...-----G-----A--G.-G..-..................-GTC...A-G..GC---ACA---C.....-TG-T......TTTA...   9562
CPZ.GA.CPZGAB      ......................-----AA.....................----------G-----AG-.-..................-GTC..---AGG----GT-G---T.....-TG-C......TTTG...   9590
CPZ.CM.CAM3        -.....................-----TAA.....................-----------A--G--G..-..................-TTC..A--G..--------.---C.....-TG-T......TTTA...   9851
CPZ.CM.CAM5        ......................-----TAA.....................-----------A--G--A..G-.................-TC..A--G..-------A.---C.....-TG-T.....TTCAG...   9253
CPZ.CD.CPZANT      CTTGTGGTCT..----------ACTAT....................----G-GA-T-.....................................T--G-G.--CA.....-TG-GAGTGGCTTTC...   8954

H1A.UG.U455        -.....................-----A.---....................----------GA--GA-G..--T...............-GT-..T--G..---AGT-.----..GA.-TG-C......TAG....   8956
H1C.ET.ETH2220     C.....................-----A.---.....................----.-G-------AG-AG-G..................-T--..---...--------.---.------C-......A......   8896
H1D.ZR.84ZR085     ......................................................................................................................   8975
H1F.BE.VI850       ......................-----A.---....................------------A-G-CG-TCC...............AGAG..--...-------.----.------TC......A......   8794
H1G.SE.SE6165      ......................----------AC--...................----------G--------CG-..................---.---...A--AGT-.----.------T-......A......   8940
H1H.CF.90CF056     ......................----------.--...................----------G--------.--A................T--.---...----GA-.----.------C-......A......   8814
H1J.SE.SE92809     ......................----------.---..................--------------GA-GAG--.................---.---...----GT-.----.------T-......A......   8795
H1K.CM.MP535C      ......................................................................................................................   8598
H101_AE.TH.CM240   ............A-A----.---.----------------------------------------GA-GTG--.................---G..---...---AGT-.----.---A--T-......A......   9063
H1N.CM.YBF30       CAA.....................-------ATACT--....GGGACTTTCCGCCA.--------------.A-GTG-.................-GT-..---GGAGT--CT-.-.......................   9096
H1O.CM.MVP5180     TGC.....................----------AGAC.TGCTGACACTGCG......------------C-T--GAGG-AT..............AAGG..--CGG......T-.C---G------T-......A......   9613
H1O.BE.ANT70       .............--A-------AG-AAAGACTGCTGACACTGCG..------------..T--GAGG-AC..............AGGG..--CGG......T-.C---G------T-......A......   9616

MAC.US.SMM251      ..............----AAC--GCTGA-..ATAGCA...........-----------CAAG--GATGTTA..............TGGG..-A-GAG-C--T-G.--AAC-....................   9903
MAC.US.SMM239      ............A---AAC--GCTGAA..ACAGCA...........-----------CAAG--GATGTTAC...GGGGAGGTACTGGG..-A-GAG-C--T-G.--AACG...................   9933

STM.US.STM         ......................----------ACAA-....................--G--G-.A-CA-G--A-G-ACT..............-GGA..--AACTG.-TGGG.-AAC.G...................   9587
SMM.US.SMM9        ......................----------ACAAA...................---GA-G-T.-C.-G--A-G-TCT..............-GGG..A--AGCT.--CTG.-AAC.G...................   9416

H2A.GW.ALI         ......................----------AGAA-....................--G--G-.A-CCAG--GAAG-AC..............ATGG..-A-GAG-T--TGG.--AA.C-...................  10009
H2A.SN.ST          GCA.....................----------AGAA...................---G--G-T.-CCAG--GAGG................-A-A..T--GA..----GC.C--T.G-G-AACG.............   9452
H2A.DE.BEN         GCA.....................----------AGAA...................---G--G-.A-CCA---GAGG................-A-A..T--GA..----GC.T--T.G-G-AACG.............  10013
H2B.GH.D205        GCA.....................----------AAAA...................-A-G--G-.A-CA.T--GAGG................AG-A..A--GA..--GG-.T--T.T-G-AACG.............   9932
H2B.CI.EHO         GCA.....................---------AGAA....................---G--G-.A-CA-T--GAG................-GAC..AT-GG...A-GAG.CA--.GAG-AACG.............   9903
H2G.CI.ABT96       GCA.....................----------AGTA...................T--G--G-.A-CA-.--GAG................GTA..CTATG..----GG.-ACT.G--T-G--......ACG....   9362

TANTALUS.UG.TAN1   -AGT...........G--GAT-AG-A...................--------------T--T-CG................AGGG..C--TA.....CT-..---.------TT............C   9526
GRIVET.ET.GRI677   ......................-----C-A...................-----------CA-T--GTGGAT..............CGGA..--CGGTACA-GGG.C--TACTG--A-T......-GCTTTC   9193
VERVET.KE.VER9063  CTGACAG...----------AA.....................----------.........................-GGC..---CCAT.--G-G.-TAC.G-G-AGTG......-TT.TTA   9553
VERVET.KE.VERTYO   TGCTGAC...----------AG-GAA.................----------A-........................--GG..-ACAT...--G-G.-TCC.G-G-AGTG......-CT.TTA   9010
VERVET.KE.VER155   TGCTGAC...----------AG-AC...................----------A--C-.................................GACATGGG.C--TAC-G--A-T......-GCTTTA   9534
VERVET.DE.VERAGM3  TGCTGAC...----------AG-AA...................--------------C--G.TCATG..............-G-G..-TACGGG-A-TGG.CTTTA....................   9053
SABAEUS.SN.SAB1C   -CATTAA.....-AACTGC-GACTGA-A.................----------------------------------------TGG..A-ACT.G--CGG-.ACT-GGAGT-GCT......TGC....   9706

SUN.GA.SIVSUN      GAGTA......----------GACAGGG...........................A-G--GAG.-CT..............TA.........................................   9684
LHOEST.KE.LHOEST   -GCA.....----------AGGACGGGCGG............G--AGG---G.....................................................................   9640
LHOEST.CD.447      -GCA.....----------AGGACGGGCGG............G--AGG---G.....................................................................   8559
LHOEST.CD.485      -GCA.....----------AGGACGGGCGG............G--AGG--CA.....................................................................   8562
LHOEST.CD.524      -GCA.....----------AGGACGGGCGG............G--AGG---G.....................................................................   8565
MANDRIL.GA.MNDGB1  ......................................................................AA..ACAGG..-AGGG.--A-.GC-C-..................   9000

COLOBUS.CM.CGU1    -GCA.....----------TTC.....................----------A-TT--GAGG...............................GACCT--GC-GA-G--AGGGCGTG.....   8679
SYKES.KE.173       .............--AG-C-GGGCGGG-G.....................................................................AA.--CT.CA...................   9336
RCM.NG.NGM         -GAA.....----------ACTG-CGCCTGCGCGCTGGTGTAA------------ACT-A--...................TGGG..A--GG.....-TG-.--CT.T-..................  10033
MANDRIL.GA.14CG    TGGA.....----------ACTA-TGCATGCGCACTGGACT.G--------T---ACT-A--TC-GG..............TGGG..--AGTGGTCAG..........................   9474
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV
complete genomes

```
                                                                 3'LTR U3 end \/ 3'LTR R repeat start
                                                                                TAR element   stem         bulge    loop      stem
                                   TATA Box                                     |||| |^||||  |||||||^  |||||^^^||||        | |||||||||    |||
H1B.FR.HXB2        CCCTCAG.ATCCT....GCATATAAGCAG..CTGC....TTTTTG.CCTGT....AC..TGGG.TCTCTC..TGGTTAGA..CCAGATCTGAGC.CTGGGAG.CTCTCTGGC....TAA..........   9586

CPZ.US.CPZUS       ------A.-GA--...-----A-------..-C--...T.C-C--.-T---....--..----.-----...-C-C-----..------AA-----.-C-----.----------....--G..........   9651
CPZ.GA.CPZGAB      -------.-.G--...-------------..-------...T.--C--.-----....--..----.-----T..-CAC-G--..------T-----.-------.------AC-......-G..........   9678
CPZ.CM.CAM3        -------.-AG--...-------------..-C--...T.C-C--.-T---....--..----.------..-T-C----..------------.-------.--------T....C-G..........   9041
CPZ.CM.CAM5        TTT----.-......................................................................................................................   9261
CPZ.CD.CPZANT      ----G-CT..G--...-------------..-----...T.--GC-.-T---....TA..A---C-----G..CCTAGGAG..-------A----.-------.--------TA.AT-TTT.........   9048

H1A.UG.U455        -------.--G--...-------------..----...----C-.-----....--..----.------..-T------..-------.C----.-------.----------....--G..........   9045
H1C.ET.ETH2220     -------.--G--...-------------..----...----C-.-T---....--..C---.------..-A-G----..------------.-------.----------....--T..........   8986
H1D.ZR.84ZR085     ......................................................................................................................   8975
H1F.BE.VI850       -------.--G--...-------------..-C--...----C-.-----....--..----.------..-A------..------T-----.-----G-.----------....--G..........   8884
H1G.SE.SE6165      -------.-AG--...-------------..-C--...--C-C-.-----....--..----.------..-T------..------T-----.-------.----------....--G..........   9030
H1H.CF.90CF056     -------.--G--...-------------..----...----C-.-T---....--..----.-----T..---------..------------.-------.----------....--G..........   8904
H1J.SE.SE92809     -------.--G--...-------------..----...----C-.-T---....--..----.------..-T------..------------.-------.----------....----------...   8885
H1K.CM.MP535C      ......................................................................................................................   8598
H101_AE.TH.CM240   -------.--G--...-------A-------..-C--...----C-.-T---....--..----.------..-T-------..----G--.-----.-C-----.----------....--G..........   9152
H1N.CM.YBF30       -------.-.G--...-----A-------..-C--...--C-C-.-T---....--..----.------..-T-C-G--..------TA---T.--------.-ATAT---G................   9182
H1O.CM.MVP5180     -------.--G--...-------------..----...----CC-.-T---....--..C---.---TAG..-TAGAG--..----G-------.-C-----.---C-----....CT...........   9702
H1O.BE.ANT70       -------.-AG--...-------------..-C--...----C--.-T---....--..C---.----GG..-TAGAG--..----G-------.-C-----.---C-----....CT...........   9705

MAC.US.SMM251      ---A-TTTC.TTG....ATG-----AT-TCA----...A---.C-CT----...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGT  10011
MAC.US.SMM239      ---A-TTTC.TTG....ATG-----AT-TCA----...A---.C-CT----...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGT  10041

STM.US.STM         ----TTTAC.TT-....CTG-----AT-CAA-C--...T-A-.A-CT----...ATTC.A-TC.A----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGT   9695
SMM.US.SMM9        ---A-TTAT.T--...CTG-----AT-CAA----...A---.C-CT----...ATTC.A-TC.A----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGT   9524

H2A.GW.ALI         ---.TCAT-.-TC....CTG-----ATGTAC-C--T..GC--.GCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGT  10116
H2A.SN.ST          ------TAC.TT-....CTG-----ATGTAC-C--...TAC-CGCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-------GT-----CC...AGC-CTA...GCAGGT   9560
H2A.DE.BEN         ------TAC.TTA....CTG-----ATGTAC-C--...T-C--GCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG------.C----C---A---GT-----CC...AGC-CTA...GCAGGT  10121
H2B.GH.D205        ---ATTAA-.T--...CTG-----ATGTAC-C--...T-C-CGCAT.---...ATTC...........................................................................   9978
H2B.CI.EHO         ---ATCAAG.A-...CTG-----ATGTAC-C--...T--G-GCAT.---...ATTC.A-TC.G----GC.G-AGAG-CTGG------.C----C-------GT-----CCAG.CACT-GC..AGGTAGA  10014
H2G.CI.ABT96       ---C.-AA-.T--...CTG-----AT-TAG----...A---CGCT-.-----...ATTC.A-TC.G----GC.G-AGAG-CTGG-----.-----C-C-----GT-----CCAG.CACT-GC..AGGTAGA   9472

TANTALUS.UG.TAN1   -------.-.G--...-----A-------..A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---..----C-.-----.-------.T--G----T....G-GCCTA..ACCTG   9626
GRIVET.ET.GRI677   -------.-.G--...-----A-------..A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---..----C-.A----.-----T-.T--G----T....--GCCTA..ACCCG   9293
VERVET.KE.VER9063  -------.-.G--...-----A-------..A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---..----C-.-----.-----T-.T--G----T....--GCCTA..ACCTG   9653
VERVET.KE.VERTYO   -------.-.G--...-----A-------..A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---..----C-.-----.-----T-.T--G----T....--GCCTA..ACCTG   9110
VERVET.KE.VER155   -------.-.G--...-----A-------..A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---..----C-.-----.-----T-.T--G----T....--GCCTA..ACCTG   9634
VERVET.DE.VERAGM3  -------.-.G--...-----A-------..A---....-CGC--G-T---...A--..-CA-.-----T.ACTAGG---..----C-.-----.-----T-.T--G----T....--GCCTA..ACCTG   9153
SABAEUS.SN.SAB1C   ---.GT-CT.G--...CA----------..----...---.GCG------...A-A..ACA-.-----GC.-T-GAG-CTG------.-----.-----T-.T-------T..AAGTCTG.AACCAGCT   9810

SUN.GA.SIVSUN      ---A-..TT.A--...-------------..----T..C--CGC.-.---.....TCGG-A-.-----GC.-CAGAG--TAGGC--C...GA-C---A---G.----ACCTA.GAGCTGG........C   9780
LHOEST.KE.LHOEST   ---A-..TT.G--...-------------..----T..C.-CCGC.T.---.....TCGG-A-.--CT-GC.-CTGGG--CCAGCAG-.C.GCTC-----T-A.G-A-ACTGGTCAGCTAA........G   9738
LHOEST.CD.447      ---A-..TT.G--...-------------..----T..C.-..............................................................................   8588
LHOEST.CD.485      ---A-..TT.G--...-----A..-----..----T..C.-..............................................................................   8589
LHOEST.CD.524      ---A-..TT.G--...-------------..----T..C.-..............................................................................   8594
MANDRIL.GA.MNDGB1  ---CA..TG.-TG....CT----------..----A..T.--CGC.T.---.....TCGG-A-.-----ACTACAGAG-CTAA............G--TT-TA------AGC.AGATCCCCTTAGAGCA   9096

COLOBUS.CM.CGU1    ----GCTA-A.--GCTT-T---A---TCAG.AGC-T..CGAGAGCTT.C-AGG.C...............................................................   8728
SYKES.KE.173       ----......A--...-----A--.....G-C-A.......-GCT-.---...AG-T.ATCCGCGGAGTTGCAAC---CTCT-TCCA..-C-GC------...G-C---AGA.GACTGGTAGAGCGGCC   9433
RCM.NG.NGM         -----TCCT.G--...-------------.-----A..T.--CGCT-.---...A--C.A-TC.----.C-C.---GAG-CTA------.-----.-----TTGT-------TAAGTCTCTAG..GAGACT  10141
MANDRIL.GA.14CG    ---CA..T..-TGAC.T------------.C-C--T..T.GCCGC.T.---.....AT--A-.------CC...................................................   9528
```

```
                                        TAR element
                                  ||||                 |||||||    |||
H1B.FR.HXB2        ................CTAG..................GGAACCC....ACTGCTTAA.......................................................................  9606

CPZ.US.CPZUS       ................TGTA..................-------------------.......................................................................  9670
CPZ.GA.CPZGAB      ................TG-A..................-------------------.......................................................................  9698
CPZ.CM.CAM3        ................G---..................A------------------.......................................................................  9061
CPZ.CM.CAM5        ................................................................................................................................  9261
CPZ.CD.CPZANT      ................GGCAG.................-AG----------.----.......................................................................  9065

H1A.UG.U455        ................-G--..................-------------------.......................................................................  9065
H1C.ET.ETH2220     ................--G-..................-------------------.......................................................................  9006
H1D.ZR.84ZR085     ................................................................................................................................  8975
H1F.BE.VI850       ................--G-..................-------------------.......................................................................  8903
H1G.SE.SE6165      ................-AG-..................-------------------A......................................................................  9051
H1H.CF.90CF056     ................---A..................-------------------.......................................................................  8924
H1J.SE.SE92809     ................----..................-------------------.......................................................................  8905
H1K.CM.MP535C      ................................................................................................................................  8598
H101_AE.TH.CM240   ................-A--..................-------------------A......................................................................  9173
H1N.CM.YBF30       ................................................................................................................................  9182
H1O.CM.MVP5180     ...............CTAGC.................T---------.G-------.......................................................................  9723
H1O.BE.ANT70       ...............CTAGC.................T---------.G--CG----.......................................................................  9726

MAC.US.SMM251      AGAGCCTGGGTGTTCC--................GCTA--CT-TCACC-GCA---GGCCAGTGCTGGGCA..GAG..................................TGGCTCCACGCT.TGCTTG  10086
MAC.US.SMM239      AGAGCCTGGGTGTTCC--................GCTA--CT-TCACC-GCA---GGCCGGTGCTGGGCA..GAG..................................TGACTCCACGCT.TGCTTG  10116

STM.US.STM         AGAGCCTGGGTGTTCC--................GCTA--CT-TCACC-GCA---GGCCAGTGCTGGGTA..GAG..................................TGGCTCCACGCT.TGCTTG   9770
SMM.US.SMM9        AGAGCCTGGGTGTTCC--................GCT---CT-TCACC-GCA---GGCCGGTGCTGGGCA..GAG..................................TRGCTCCACGCT.TGCTTG   9599

H2A.GW.ALI         AGAGCCTGGGTGTTCC--................GCTA--CT-TCACC-G-----GGCCGGCACTGGGCA..GA...................................CGGCTCCACGCT.TGCTTG  10190
H2A.SN.ST          AGAGCCTGGGTGTTCC--................GCTA--CT-TCACC-G-----GGCCGGCACTGGGCA..GA...................................CGGCTCCACGCT.TGCTTG   9634
H2A.DE.BEN         AGAGCCTGGGTGTTCC--................GCT---CT-TCACC-G-A---GGCCGGTACTGGGCA..GA...................................CGGCTCCACGCT.TGCTTG  10195
H2B.GH.D205        ................................................................................................................................  9978
H2B.CI.EHO         GCCTGGGTGTTC.CCTGC................TA--CT-TCACC-GCA---GGCCGGTGCTGGGCA..GAC..................................GGCTCCACGCT.TGCTTG  10085
H2G.CI.ABT96       GCCTGGGTGTTC.CCTGC................TA--CT-TCACC-GCA---GGCCGGTGCTGGGCA..GAGC.................................GGCTCCACGCT.TGCTTG   9544

TANTALUS.UG.TAN1   GTCTGGCCATCCAGGGG--A..................-AGC-T-....CT------GAAGGCTAT.......................................................  9670
GRIVET.ET.GRI677   GT.TGGCCA.CCGGGGG--A...................---CT-.....TTG----C-TATA.........................................................  9331
VERVET.KE.VER9063  GT.TGGCCA.CCAGGGG--A...................---CT-.....TTG----GGAA..........................................................  9689
VERVET.KE.VERTYO   GT.TGGCCA.CCAGGGG--A...................---CT-.....TTG-----GAA.........................................................  9146
VERVET.KE.VER155   GT.TGGCCA.CCAGGGG--A...................---CT-.....TTG-----GAA.........................................................  9670
VERVET.DE.VERAGM3  GT.TGGCCA.CCAGGGG--A...................---CT-.....TTG-----GAA.........................................................  9189
SABAEUS.SN.SAB1C   TGAGCCTGGGTGTTCG--.-GGGAAGGCAAGCCTAAGA-C-CT-TGGTG-GGT---GGTAGACCCTCGCC.............TGGCGACTGGCCATTGCCAGTAGCAG.AGACTCC.......GCTTG   9917

SUN.GA.SIVSUN      TTGTCGCTCCTGAGT.GAGCACTTGCTAAGCACC.TCACCCGAG-CTGAT-A----GG..GGAGCTAGA.....................................GGCTCC....T.TGCTTT   9858
LHOEST.KE.LHOEST   CAGTCAGCCCTGGGA.GGCTACA..........CTTAGCACC-A-CTGGG-CTGA-C-GCCTAGGGAGCTA..AG...................................GGCTCCTAGCT.TGCTAC   9817
LHOEST.CD.447      ................................................................................................................................  8588
LHOEST.CD.485      ................................................................................................................................  8589
LHOEST.CD.524      ................................................................................................................................  8594
MANDRIL.GA.MNDGB1  AGGACCAGAGT....C--GA...........GTGACT.-GGT-TGAGC-..C--C-CTCGGGGCTG.....................ATCACCTCGAGGTAGTGGAA.....CTCCTTGCT.TGCTTG   9179

COLOBUS.CM.CGU1    ................................................................................................................................  8728
SYKES.KE.173       CG..GGAGCCCTTGCCAGTAAAGGCTAGTGAGGTCCGG-TTTGG-CGCCCGGT---CGGCAGCTCTCCGTT..GCTT.............................GCTTTGCTTA   9516
RCM.NG.NGM         AG..CTTGAGCCTGGGTGT.TCGCTAGTTCTCTGAAGG-CTTG-TGGGAG-GC---GGCTGGGTCTTGGTA..GACCTTCAGCTGGGTACTTGGCCAGTCCCAGTGGGAGTAGGCTCC..GCT.TGCTTT  10263
MANDRIL.GA.14CG    ................................................................................................................................  9528
```

HIV-1/HIV-2/SIV
complete genomes

HIV-1/HIV-2/SIV complete genomes

```
                    TAR elements of HIV-2 and SIV have extra stem-loop, see Berkhout, Nucleic Acids Res 20(1):27-31(1992)
                                                        Poly-A signal        3'LTR R repeat end \/ 3'LTR U5 start
H1B.FR.HXB2         ............GCC.TC...............................AATAAAG...CTTGCC.......TTGAG..TG.CTTCAA.GTAGTG...TGTGCCCGTCTG.TT..GTG.TGACTCT.   9667

CPZ.US.CPZUS        ...............---.--.......................................-------..----------.....--CT--A.-.------...------TA--C.--C..A-.-------.   9730
CPZ.GA.CPZGAB       ...............---.--.......................................-------..----------.....TGTAT--G.-.-C----...------TA---A.--C..A-.AC-----.   9760
CPZ.CM.CAM3         ...............---.--.......................................-------..----------.....--A---G.-.-C----...------TA--CA.--C..A-.-A-----.   9122
CPZ.CM.CAM5         ........................................................................................................................................   9261
CPZ.CD.CPZANT       .............-T.-..........................................................................................................................   9068

H1A.UG.U455         ...............---.--.......................................-------..----------.....GTG---G--.------...-------------.--..A-T.-------.   9127
H1C.ET.ETH2220      ...............---.--.......................................-------..----------.....--.......................................................   9031
H1D.ZR.84ZR085      ........................................................................................................................................   8975
H1F.BE.VI850        ........................................................................................................................................   8903
H1G.SE.SE6165       ...............---.--.......................................-------..----------.....AA-CC...................................................   9074
H1H.CF.90CF056      ...............---.--.......................................-------..----------.....--.---T.............................................   8953
H1J.SE.SE92809      ...............---.--.......................................-------..----------.....G-CGCATGC.AAGCC-.....................................   8943
H1K.CM.MP535C       ........................................................................................................................................   8598
H101_AE.TH.CM240    ...............---.--.......................................-------..----------.....--.---A-.............................................   9203
H1N.CM.YBF30        ........................................................................................................................................   9182
H1O.CM.MVP5180      ............CG-.--.........................................-------..----------.....--.AGAAGC.AGT---...--CT-A.-----.--CAACC.CTGG-G-.   9784
H1O.BE.ANT70        ............CG-.--.........................................-------..----------.....--.AGCTTG.--CTG-AAG-T--AT-CCA--T-AGCAGT.CC-A-A-.   9793

MAC.US.SMM251       C...TTAA.AGAC-TC-TC.........................................-------...--.---AT..T..--AGA..A-TAAG-C-.--GTGTGTTCCCAT-TC--CTAG-CGCC-.CCTGGTC.  10166
MAC.US.SMM239       C...TTAA.AGCC-TC-TC.........................................-------...--.---AT..T..--AGA..A-TAAG-T-.--GTGTGTTCCCAT-TC--CTAGCCGCC-.CCTGGTC.  10196

STM.US.STM          C...TTAT.ATAT-TC-TC.........................................-------...--.---AA..T..--AGA..A-TAAG-.-.AGT-G-TGT-CCCAT-TCTCCTAGTC-CC.GC.--GG.   9848
SMM.US.SMM9         C...TTAA.AGAC-TC-TC.........................................-------G..--.---AG......--AGA..A-TA.........................................   9638

H2A.GW.ALI          C...TTAA.AGAC-TC-T..........................................-------...--.---AA.....--AGA..A-CAGGTT-.AAG---TGT-CCCAT-TCTCCTAGTC-CC.GCCTGG-.  10268
H2A.SN.ST           C...TTAAAAGAC-TC-T..........................................-------C..TGCCAG......T-A-..AGCA.............................................   9672
H2A.DE.BEN          C...TTAA.AGAC-TC-TC.........................................-------...--.---AG......--AGA..A-CAAGTT-.AGT---TGT-CCCAT-TCTCCTAGTC-CC.GCCTGG-.  10274
H2B.GH.D205         ........................................................................................................................................   9978
H2B.CI.EHO          CTTAA.AA..CCCT-T-...........................................------A...--.---TA.....--AGA..A-CAAGAC-.AGT---TGT-CCCAT-TCTCCTAGTC-CC.GCCTGG-.  10163
H2G.CI.ABT96        CTTACAAC.GACCT-T-...........................................-------C..T.GC-AAA.....--AGA..A-TAAG.C-.AGT---TGTGT.........................   9599

TANTALUS.UG.TAN1    ............TG-.-...........................................-------...-----TG..C..--A--..A-T--ATCT.-.--CAAGTGCCT-ATTG-GCGCCCTTCC.-C-...G.   9735
GRIVET.ET.GRI677    ............G-.--...........................................------....-C---TCG..C..--AGT..C-CTA-ATT.-G---CAAG--CT-ATTG---CGCCGAGC.CTCTAGA.   9400
VERVET.KE.VER9063   ............AG-.-...........................................------AC..T-----TG..CA.--AGA..GCTTA--TG.AGTCAAGTGCCCT-ATTGAC-CC-CAC-C.AAG-AGG.   9761
VERVET.KE.VERTYO    ............AG-.-...........................................------...------TG..CA.--AG..................................................   9170
VERVET.KE.VER155    ............AG-.-...........................................------...------TG..CA.--AGA..GCTTA--TG.AGTCAAGTGCCCT-ATTAA--CC-CAC-C.-TGAA-G.   9740
VERVET.DE.VERAGM3   ............AG-.-...........................................------.T..---CG-TG..C.A--AG-..CAGAGA-TT.--GAGAGACAT----T-GTGCA-CCA-CA.CA-G-GA.   9260
SABAEUS.SN.SAB1C    CTTGCTTG..ATTG-C--..........................................-------T..-------------AGT..-TA-AAGT-.AAGCAAGTG-------A-T-AT-CCTCA-.GACAA-C.   9997

SUN.GA.SIVSUN       GCAGT.AC.AAGC-TTC...........................................-------C..--.---TT..AGT--T-C..--CAAG.C-.AGT--TCCCGT................GT.-CT----.   9924
LHOEST.KE.LHOEST    TTTAAA.....GC-TTC...........................................------AA..G----TCA..TTG--AGA..AAGTACA-G.C-----TCTCA-AGTA.GTCTC--TCTCC.-TCGC-C.   9897
LHOEST.CD.447       ........................................................................................................................................   8588
LHOEST.CD.485       ........................................................................................................................................   8589
LHOEST.CD.524       ........................................................................................................................................   8594
MANDRIL.GA.MNDGB1   CTAT......TGT-TTC...........................................-------T..AACTTAGA..A..--AGA..GCA.........................................   9215

COLOBUS.CM.CGU1     ........................................................................................................................................   8728
SYKES.KE.173        CG..CTGA..TCAG-CC.A.........................................-------T..G-G-TAAA..CGCAA-TA..ACGTC-GCT.C-CA-CTGT-C---TG-GTCACC-CCCGA.A-GTG...   9597
RCM.NG.NGM          CTGGTTAA..TT--TC............................................------CT...GCATGCT..CTTAGTT-AACCAGAATC-.AGT---TGCCCATTTTA---CTCAGT-AAAC------.  10348
MANDRIL.GA.14CG     ........................................................................................................................................   9528
```

```
                                                
```