IV

Primate Lentivirus Complete Genomes

Alignment of PLV Complete Genomes

Complete genomes of HIV-1, HIV-2, and SIV sequences are represented in the alignment. From non-human PLVs, viruses infecting the following hosts were included: Chimpanzees, *Pan troglodytes troglodytes* and *P.t. schweinfurtii* (cpz); the subspecies of African green monkeys, also known as Savannah monkeys, *Chlorocebus* (previously *Cercopithecus*) *aethiops aethiops* (gri), *C. a. pygerythrus* (ver), *C. a. tantalus* (tan) and *C. a. sabaeus* (sab); Drill-mangabeys including Sooty mangabeys, *Cercocebus atys* (smm) and Macaques infected with sooty mangabey virus (mac, stm, mne), and Red-capped mangabey, *C. torquatus* (rcm); Mandrills, *Madrillus sphinxs* (mnd-1, mnd-2) and Drills, *M. leucophaeus* (drl); L'Hoest monkeys, *Cercopithecus l'hoesti lhoesti* (lst) and Sun-tailed monkey, *C .l'h. solatus* (sun); Mona monkeys, *C. mona mona* (mon) and Dent's monkey, *C. m. denti* (den); De Brazza's monkey, *C. neglectus* (deb); Greater Spot-nosed monkey, *C. nictitans nictitans* (gsn); Moustached monkey, *C. cephus cephus* (mus); Syke's monkey, *C. albogularis* (syk); and Guereza colobus monkeys, *Colobus guereza* (col).

As mentioned in the general introduction, there are now 5642 non-human PLV sequences available in the database. During 2004, no new PLV species were published, and therefore the compendium alignment has not changed this year (the more complete web-alignment has been updated, however). As a result, and because a lot of work went into last years PLV section, the following discussion and the tree in figure 1 are republished this year.

For the Primate lentivirus complete genome DNA alignment, one of each of the major subtypes of the HIV-1 M group was chosen, plus one N group and 2 O group sequences. Six of the 7 SIV-CPZ were included, SIV-CPZ-CAM5 was not included because it is quite similar to CAM3. Two HIV-2 A group and 2 HIV-2 B group were included along with the single genomes sequenced to date for HIV-2 G and HIV-2 U. Five SIV(Mac-SMM-MNE-STM) genomes were included in this alignment and more are available in the HIV-2/SIV-SMM alignment. For the rest of the SIVs, all complete genomes were included. The principle in deciding which sequences to include in this alignment was to provide a representative sampling of all primate lentiviruses which have had complete or nearly complete genomes sequenced. Many HIV-1, HIV-2, SIVsmm and SIVmac genomes were not included, but they can be found in the HIV-1/SIVcpz and HIV-2/SIVsmm alignments.

Table 1 lists the sequence name, accession number, isolation country, first author, and publication for each of the sequences in this alignment. These genomes represent the full genome breadth of diversity discovered to date in the primate lentivirus lineage. There are, however, more sequences from shorter and partial genome and gene sequences available from some more simian species, but these are not included in this alignment because they are too short. They are available through our search interface and some are in the protein alignments in section VII. 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.

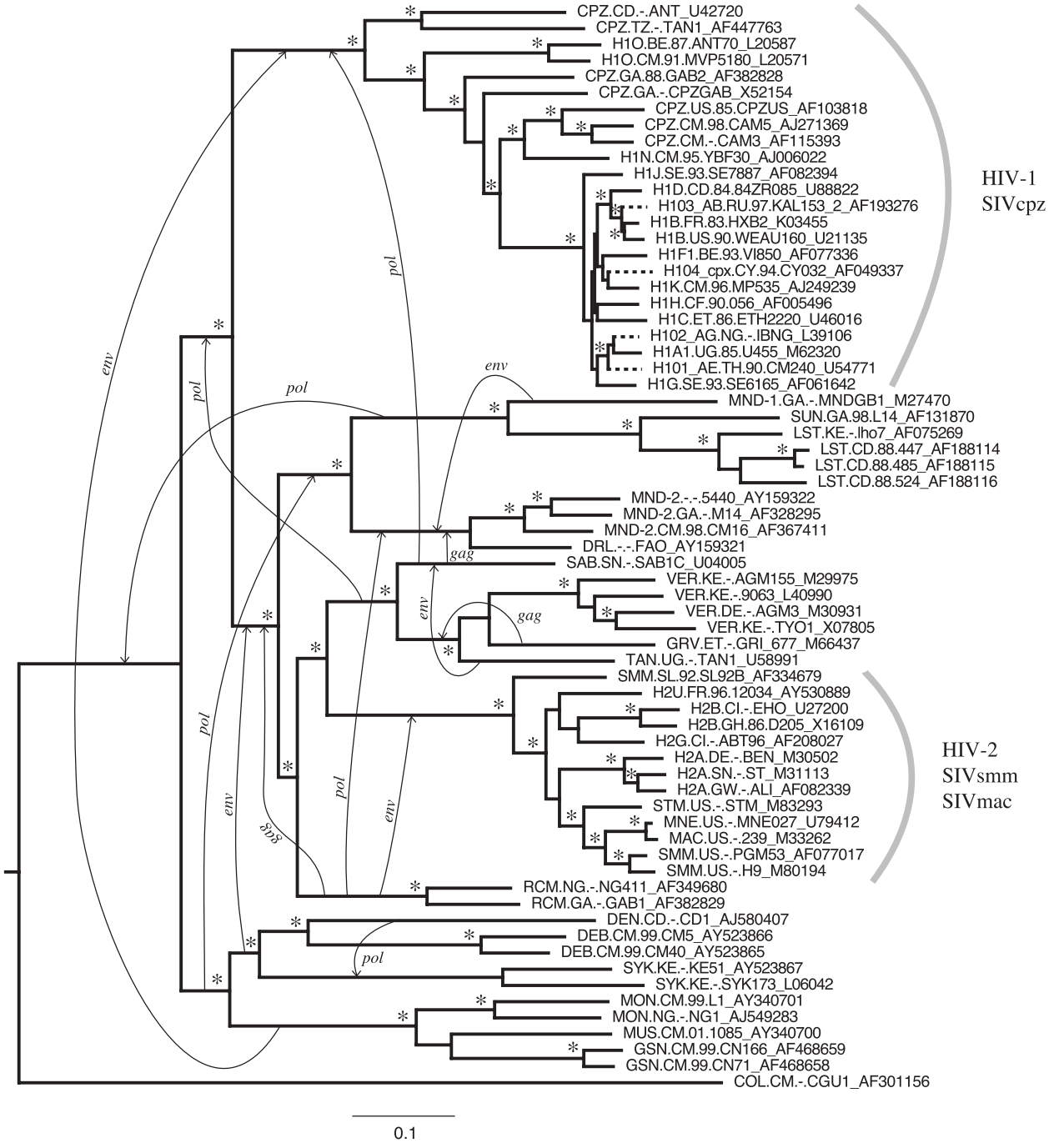
The tree in Figure 1 is a disclaimer of the apparent phylogenetic relationships between the PLVs. As has been pointed out by several authors, several of the PLV sequences appear to be recombinant. For instance, SIVrcm has been shown to display discordant clustering in different genomic regions (Beer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

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

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

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

Figure 1. Genetic relationships among Primate Lentiviruses (PLVs). The tree shows the inferred relationships in the following PLV complete genome alignment. The thin line arrows indicate the main differences inferred from the gag, pol and env gene fragments from the same alignment. Dashed tree branches indicate inferred positions for known recombinant HIV-1 circulating forms. The resulting network should therefore not be interpreted as a singular phylogenetic history of the PLVs. Asterisks indicate bootstrap support ≥70% in the complete alignment. The scale bar at the bottom shows genetic distance according to a F84 model with relative rates. The tree shown was constructed from a gapstripped version of the DNA alignment in this section of the compendium, leaving 5327 unambiguously aligned positions. A neighbor-joining tree, calculated using DNADIST (F84 model) and NEIGHBOR in the PHYLIP package, was used as a guide to calculate relative site rates. The relative site rates were estimated using a generalized version of DNArates (Korber et al., 2000) with eight categories (0.133; 0.287; 2.373; 6.826; 19.634; 56.472; and 115.653). Finally, these rates were given to fastDNAml (version 1.2.2) and a maximum likelihood tree was inferred from the gapstripped alignment (transition/



transversion ratio = 2). The tree space was searched by both Jumble and Global Rearrangements (G 3 3). Bootstrap support was calculated using SEQBOOT with the above 8 rate categories on 1000 resampled sets, and DNADIST and NEIGHBOR in PHYLIP (version 3.6.2). The separate gag, pol and env alignments were created using a new version of GeneCutter (see www.hiv.lanl.gov) and trees were calculated the same way as in the complete alignment with optimized site rates (not shown). The Colobus SIVcol sequence, which is the most divergent PLV sequence, was used as outgroup to visualize the tree using TreeEdit (version1.0) and Adobe Illustrator.

References

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

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

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

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

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

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

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

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

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

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

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

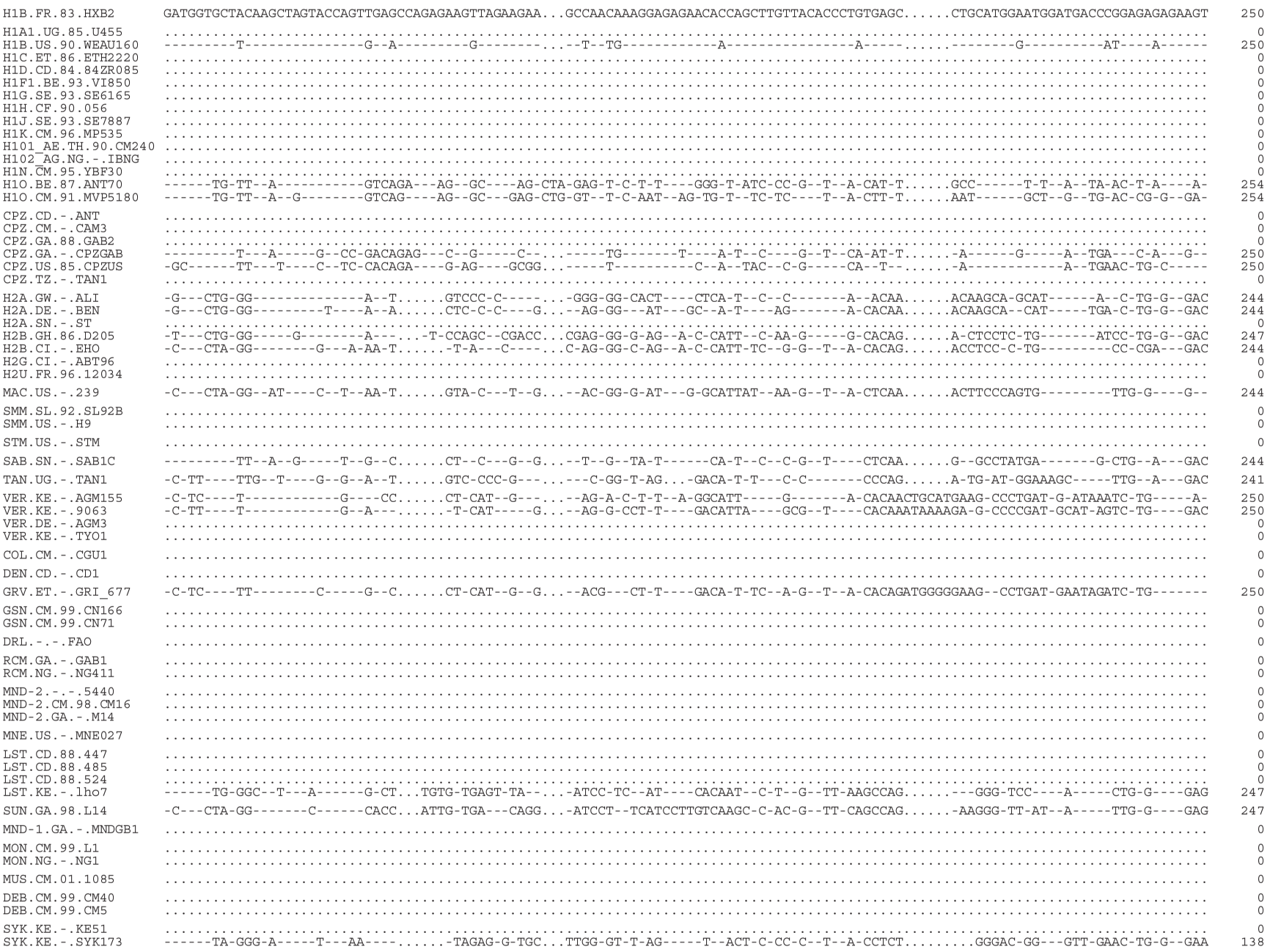
Table 1: Table of Primate Lentivirus (PLV) Sequences in the Complete Genome Alignments

Name	Accession	Country	Author	Reference
H1B.FR.83.HXB2	K03455	France	Wong-Staal, F	*Nature* **313**(6000):277–84 (1985)
H1A1.UG.85.U455	M62320	Uganda	Oram, JD	*ARHR* **6**(9):1073–8 (1990)
H1B.US.90.WEAU160	U21135	U.S.A.	Tozser, J	*FEBS Lett* **281**(1–2); 77–80 (1991)
H1C.ET.86.ETH2220	U46016	Ethiopia	Salminen, MO	*ARHR* **12**(14):1329–39 (1996)
H1D.CD.84.84ZR085	U88822	D.R.C.	Gao, F	*J Virol* **72**(7):5680–98 (1998)
H1F1.BE.93.VI850	AF077336	Belgium	Laukkanen, T	*Virology* **269**(1):95–104 (2000)
H1G.SE.93.SE6165	AF061642	Sweden	Laukkanen, T	*Virology* **247**(1):22–31 (1998)
H1H.CF.90.056	AF005496	C.A.R.	Murphy, E	*ARHR* **9**(10):997–1006 (1993)
H1J.SE.93.SE7887	AF082394	Sweden	Laukkanen, T	*ARHR* **15**(3):293–7 (1999)
H1K.CM.96.MP535	AJ249239	Cameroon	Peeters, M	*ARHR* **16**(2):139–51 (2000)
H101_AE.TH.90.CM240	U54771	Thailand	Laukkanen, T	*J Virol* **70**(9):5935–43 (1996)
H102_AG.NG.-.IBNG	L39106	Nigeria	Howard, TM	*ARHR* **10**(12):1755–7 (1994)
H1N.CM.95.YBF30	AJ006022	Cameroon	Simon, F	*Nat Med* **4**(9):1032–7 (1998)
H1O.BE.87.ANT70	L20587	Belgium	Vanden Haesevelde, M	*J Virol* **68**(3):1586–96 (1994)
H1O.CM.91.MVP5180	L20571	Cameroon	Gurtler, LG	*J Virol* **68**(3):1581–5 (1994)
CPZ.CD.-.ANT	U42720	D.R.C.	Vanden Haesevelde, M	*Virology* **221**(2):346–50 (1996)
CPZ.CM.-.CAM3	AF115393	Cameroon	Corbet, S	*J Virol* **74**(1):529–34 (2000)
CPZ.GA.88.GAB2	AF382828	Gabon	Gao, F	Unpublished
CPZ.GA.-.CPZGAB	X52154	Gabon	Huet, T	*Nature* **345**(6273):356–9 (1990)
CPZ.US.85.CPZUS	AF103818	U.S.A.	Gao, F	*Nature* **397**(6718):436–41 (1999)
CPZ.TZ.-.TAN1	AF447763	Tanzania	Santiago, ML	*J Virol* **77**(3):2233–2242 (2003)
H2A.GW.-.ALI	AF082339	Guinea-bissau	Azevedo-Pereira, JM	Unpublished
H2A.DE.-.BEN	M30502	Germany	Kirchhoff, F	*Virology* **177**(1):305–11 (1990)
H2A.SN.-.ST	M31113	Senegal	Kumar, P	*J Virol* **64**(2):890–901 (1990)
H2B.GH.86.D205	X16109	Ghana	Dietrich, U	*Nature* **342**(6252):948–50 (1989)
H2B.CI.-.EHO	U27200	Cote d'ivoire	Rey-Cuille, MA	*Virology* **202**(1):471–6 (1994)
H2G.CI.-.ABT96	AF208027	Cote d'ivoire	Brennan, CA	*ARHR* **13**(5):401–4 (1997)
H2U.FR.96.12034	AY530889	France	Damond, F	*ARHR* **20**(6):666–672 (2004)
MAC.US.-.239	M33262	U.S.A.	Kestler, H	*Science* **248**(4959):1109–12 (1990)
SMM.SL.92.SL92B	AF334679	Sierra Leone	Chen, Z	*J Virol* **70**(6):3617–27 (1996)
SMM.US.-.H9	M80194	U.S.A.	Courgnaud, V	*J Virol* **66**(1):414–9 (1992)
STM.US.-.STM	M83293	U.S.A.	Novembre, FJ	*Virology* **186**(2):783–7 (1992)
SAB.SN.-.SAB1C	U04005	Senegal	Jin, MJ	*EMBO J* **13**(12):2935–47 (1994)
TAN.UG.-.TAN1	U58991	Uganda	Stivahtis, GL	*Virology* **228**(2):394–9 (1997)
VER.KE.-.AGM155	M29975	Kenya	Johnson, PR	*J Virol* **64**(3):1086–92 (1990)
VER.KE.-.9063	L40990	Kenya	Hirsch, VM	*J Virol* **69**(2):955–67 (1995)
VER.DE.-.AGM3	M30931	Germany	Baier, M	*Virology* **176**(1):216–21 (1990)
VER.KE.-.TYO1	X07805	Kenya	Fukasawa, M	*Nature* **333**(6172):457–61 (1988)
COL.CM.-.CGU1	AF301156	Cameroon	Courgnaud, V	*J Virol* **75**(2):857–66 (2001)
DEN.CD.-.CD1	AJ580407	D.R.C.	Saragosti, S	Unpublished
GRV.ET.-.GRI_677	M66437	Ethiopia	Fomsgaard, A	*Virology* **182**(1):397–402 (1991)
GSN.CM.99.CN166	AF468659	Cameroon	Courgnaud, V	*J Virol* **76**(16):8298–309 (2002)
GSN.CM.99.CN71	AF468658	Cameroon	Courgnaud, V	*J Virol* **76**(16):8298–309 (2002)
DRL.-.x.FAO	AY159321		Hu, J	*J Virol* **77**(8):4867–4880 (2003)
RCM.GA.-.GAB1	AF382829	Gabon	Gao, F	*Science* **300**(5626); 1713 (2003)
RCM.NG.-.NG411	AF349680	Nigeria	Beer, BE	*J Virol* **75**(24):12014–27 (2001)
MND-2.-.-.5440	AY159322		Hu, J	*J Virol* **77**(8):4867–4880 (2003)
MND-2.CM.98.CM16	AF367411	Cameroon	Takehisa, J	*ARHR* **17**(12):1143–54 (2001)
MND-2.GA.-.M14	AF328295	Gabon	Souquiere, S	*J Virol* **75**(15):7086–96 (2001)
MNE.US.-.MNE027	U79412	U.S.A.	Kimata, JT	*J Virol* **72**(1):245–56 (1998)

LST.CD.88.447	AF188114	D.R.C.	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.CD.88.485	AF188115	D.R.C.	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.CD.88.524	AF188116	D.R.C.	Beer, BE	*J Virol* **74**(8):3892–8 (2000)
LST.KE.-.lho7	AF075269	Kenya	Hirsch, VM	*J Virol* **73**(2):1036–45 (1999)
SUN.GA.98.L14	AF131870	Gabon	Beer, BE	*J Virol* **73**(9):7734–44 (1999)
MND-1.GA.-.MNDGB1	M27470	Gabon	Tsujimoto, H	*Nature* **341**(6242):539–41 (1989)
MON.CM.99.L1	AY340701	Cameroon	Courgnaud, V	*J Virol* **77**(23):12523–12534 (2003)
MON.NG.-.NG1	AJ549283	Nigeria	Barlow, KL	*J Virol* **77**(12):6879–88 (2003)
MUS.CM.01.1085	AY340700	Cameroon	Courgnaud, V	*J Virol* **77**(23):12523–12534 (2003)
DEB.CM.99.CM40	AY523865	Cameroon	Bibollet-Ruche, F	*J Virol* **78**(14):7748–7762 (2004)
DEB.CM.99.CM5	AY523866	Cameroon	Bibollet-Ruche, F	*J Virol* **78**(14):7748–7762 (2004)
SYK.KE.-.KE51	AY523867	Kenya	Bibollet-Ruche, F	*J Virol* **78**(14):7748–7762 (2004)
SYK.KE.-.SYK173	L06042	Kenya	Hirsch, VM	*J Virol* **67**(3):1517–28 (1993)

```
                    / 5' LTR U3 start
H1B.FR.83.HXB2      .TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTG    129
H1A1.UG.85.U455     ..............................................................................................................................      0
H1B.US.90.WEAU160   .--------------T--------AA-------------------------G------------------T-----------G--------------------------CT-----C------TG-----    129
H1C.ET.86.ETH2220   ..............................................................................................................................      0
H1D.CD.84.84ZR085   ..............................................................................................................................      0
H1F1.BE.93.VI850    ..............................................................................................................................      0
H1G.SE.93.SE6165    ..............................................................................................................................      0
H1H.CF.90.056       ..............................................................................................................................      0
H1J.SE.93.SE7887    ..............................................................................................................................      0
H1K.CM.96.MP535     ..............................................................................................................................      0
H101_AE.TH.90.CM240 ..............................................................................................................................      0
H102_AG.NG.-.IBNG   ..............................................................................................................................      0
H1N.CM.95.YBF30     ..............................................................................................................................      0
H1O.BE.87.ANT70     C--------T-----T-------TAA----GC---A-----G-----T---G-G--TA----T--G--A-T----------G------------------A-----A-C---G-TC--------A----    130
H1O.CM.91.MVP5180   C----T---T-----T-------TAAG---GC---A-----G-----C-----A--T-----T--G--A-T----------G----TGT--------G--A-----ACCT----TC--------A----    130
CPZ.CD.-.ANT        ..............................................................................................................................      0
CPZ.CM.-.CAM3       ..............................................................................................................................      0
CPZ.GA.88.GAB2      ..............................................................................................................................      0
CPZ.GA.-.CPZGAB     .--------T--G--T-----AGGA---------G--------C--C---G----T------------T---------C-G------------A----A-----A-CA----TC-----CTGT----    129
CPZ.US.85.CPZUS     .--------T--G--T-T---AG-A----------G-----A-----C---G-T-----T--------AAT---------C-G------T-----C--------T--AG-A--------C--A--A-A--    129
CPZ.TZ.-.TAN1       ..............................................................................................................................      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.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
H2A.SN.-.ST         ..............................................................................................................................      0
H2B.GH.86.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
H2U.FR.96.12034     ..............................................................................................................................      0
MAC.US.-.239        .--------A-TTA-T--AGTGC-A-------TAGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT----A-----A--T-----C---AA---A----    129
SMM.SL.92.SL92B     ..............................................................................................................................      0
SMM.US.-.H9         ..............................................................................................................................      0
STM.US.-.STM        ..............................................................................................................................      0
SAB.SN.-.SAB1C      .--------A-TTA-T--AGTG--A-G---A-GA-G---T-G-----C-ATGCTCTTA-TGAGTGG---ATAG-AGA--GA-G------T--T---GAT--------T-C---G--C---AA-TG-----    129
TAN.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
VER.KE.-.AGM155     .----T---A-TTA-T-----G-TA-G---A-TA-G------A------ATGC-CT-A-TGA-TGG--AAT-A-TGA-----G-A-CGCA-GGT--AA---A--T-----A-----C--GAG-TG-----    129
VER.KE.-.9063       .----T---A-TTA-T-----G--A-G--GA-TAGA-----CA-C----ATGCTCTTA-TGA-TGG--AAT-A-TGA-----G-A-CGC--GGT--A----A-----CG-A--------C-GCTG-----    129
VER.DE.-.AGM3       ..............................................................................................................................      0
VER.KE.-.TYO1       ..............................................................................................................................      0
COL.CM.-.CGU1       ..............................................................................................................................      0
DEN.CD.-.CD1        ..............................................................................................................................      0
GRV.ET.-.GRI_677    .----T---A--TA-T----TG--A-----G--A-G-----GA--T---ATGC--TGA--GAGTGG--AATAA-AGA-----G---AGCT---T-------C--G-----A--G--C--GAGAGT-----    129
GSN.CM.99.CN166     ..............................................................................................................................      0
GSN.CM.99.CN71      ..............................................................................................................................      0
DRL.-.-.FAO         ..............................................................................................................................      0
RCM.GA.-.GAB1       ..............................................................................................................................      0
RCM.NG.-.NG411      ..............................................................................................................................      0
MND-2.-.-.5440      ..............................................................................................................................      0
MND-2.CM.98.CM16    ..............................................................................................................................      0
MND-2.GA.-.M14      ..............................................................................................................................      0
MNE.US.-.MNE027     ..............................................................................................................................      0
LST.CD.88.447       ..............................................................................................................................      0
LST.CD.88.485       ..............................................................................................................................      0
LST.CD.88.524       ..............................................................................................................................      0
LST.KE.-.lho7       .-----------TA-TGG-----GA-------G--C---T-AAT---C-ACTGTG-GA-TGAGTGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A-CA-----C--T-----A----    129
SUN.GA.98.L14       .----------CTGGT-----AG-ACC---G----A---T-G---T---ATGCAG-AA-TGAGTGG----T-A-AA---GC-G---AG-T-----CAA------T--AG-G---------AA-G------    129
MND-1.GA.-.MNDGB1   ..............................................................................................................................      0
MON.CM.99.L1        ..............................................................................................................................      0
MON.NG.-.NG1        ..............................................................................................................................      0
MUS.CM.01.1085      ..............................................................................................................................      0
DEB.CM.99.CM40      ..............................................................................................................................      0
DEB.CM.99.CM5       ..............................................................................................................................      0
SYK.KE.-.KE51       ..............................................................................................................................      0
SYK.KE.-.SYK173     ..........................................................................................---A-C-C-G--C--GT-A-T-----     26
```

PLV
complete genomes

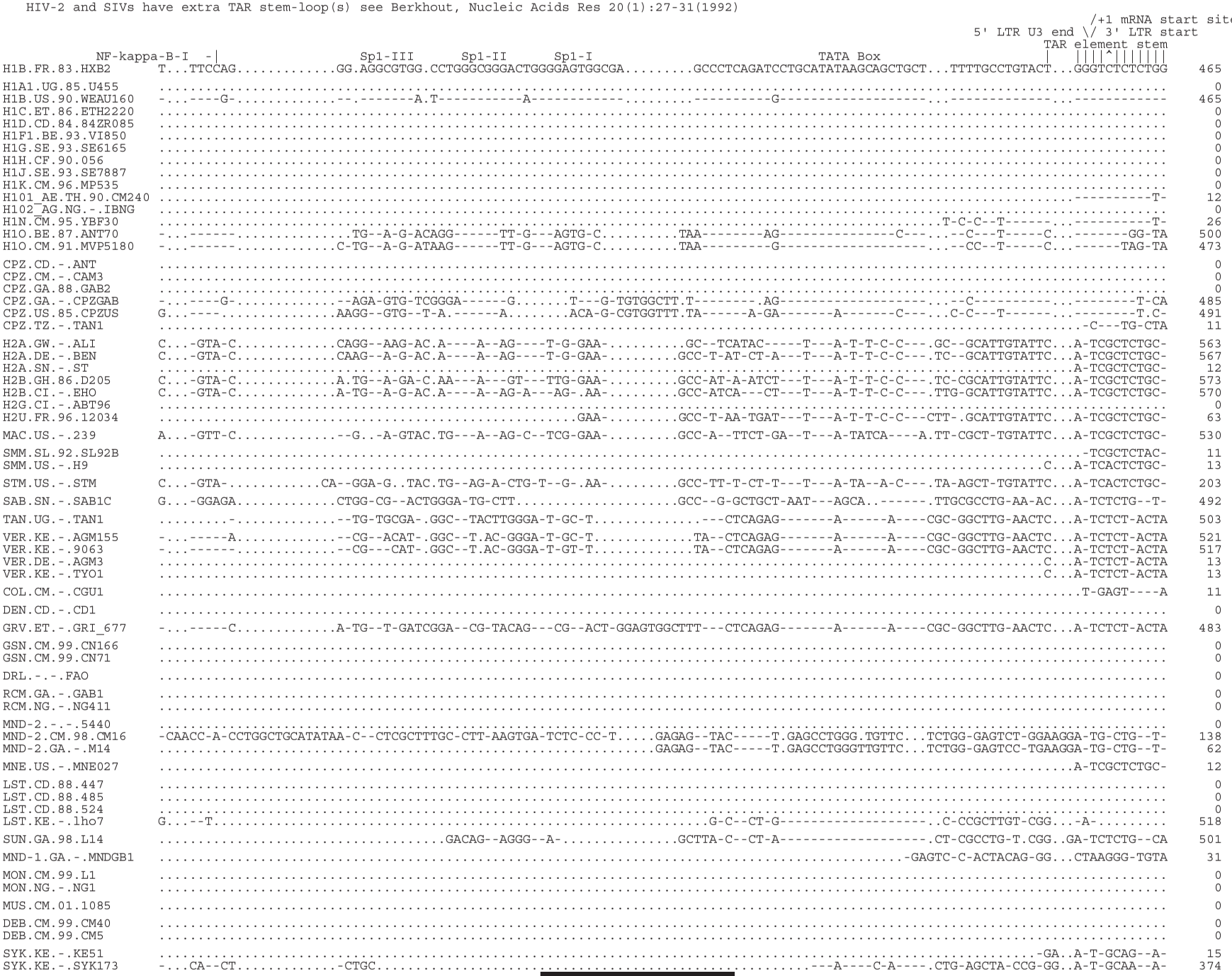

H1B.FR.83.HXB2 GATGGTGCTACAAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAA...GCCAACAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGC......CTGCATGGAATGGATGACCCGGAGAGAGAAGT 250
H1A1.UG.85.U455 0
H1B.US.90.WEAU160 250
H1C.ET.86.ETH2220 0
H1D.CD.84.84ZR085 0
H1F1.BE.93.VI850 0
H1G.SE.93.SE6165 0
H1H.CF.90.056 0
H1J.SE.93.SE7887 0
H1K.CM.96.MP535 0
H101_AE.TH.90.CM240 0
H102_AG.NG.-.IBNG 0
H1N.CM.95.YBF30 0
H1O.BE.87.ANT70 254
H1O.CM.91.MVP5180 254
CPZ.CD.-.ANT 0
CPZ.CM.-.CAM3 0
CPZ.GA.88.GAB2 0
CPZ.GA.-.CPZGAB 250
CPZ.US.85.CPZUS 250
CPZ.TZ.-.TAN1 0
H2A.GW.-.ALI 244
H2A.DE.-.BEN 244
H2A.SN.-.ST 0
H2B.GH.86.D205 247
H2B.CI.-.EHO 244
H2G.CI.-.ABT96 0
H2U.FR.96.12034 0
MAC.US.-.239 244
SMM.SL.92.SL92B 0
SMM.US.-.H9 0
STM.US.-.STM 0
SAB.SN.-.SAB1C 244
TAN.UG.-.TAN1 241
VER.KE.-.AGM155 250
VER.KE.-.9063 250
VER.DE.-.AGM3 0
VER.KE.-.TYO1 0
COL.CM.-.CGU1 0
DEN.CD.-.CD1 0
GRV.ET.-.GRI_677 250
GSN.CM.99.CN166 0
GSN.CM.99.CN71 0
DRL.-.-.FAO 0
RCM.GA.-.GAB1 0
RCM.NG.-.NG411 0
MND-2.-.-.5440 0
MND-2.CM.98.CM16 0
MND-2.GA.-.M14 0
MNE.US.-.MNE027 0
LST.CD.88.447 0
LST.CD.88.485 0
LST.CD.88.524 0
LST.KE.-.lho7 247
SUN.GA.98.L14 247
MND-1.GA.-.MNDGB1 0
MON.CM.99.L1 0
MON.NG.-.NG1 0
MUS.CM.01.1085 0
DEB.CM.99.CM40 0
DEB.CM.99.CM5 0
SYK.KE.-.KE51 0
SYK.KE.-.SYK173 138

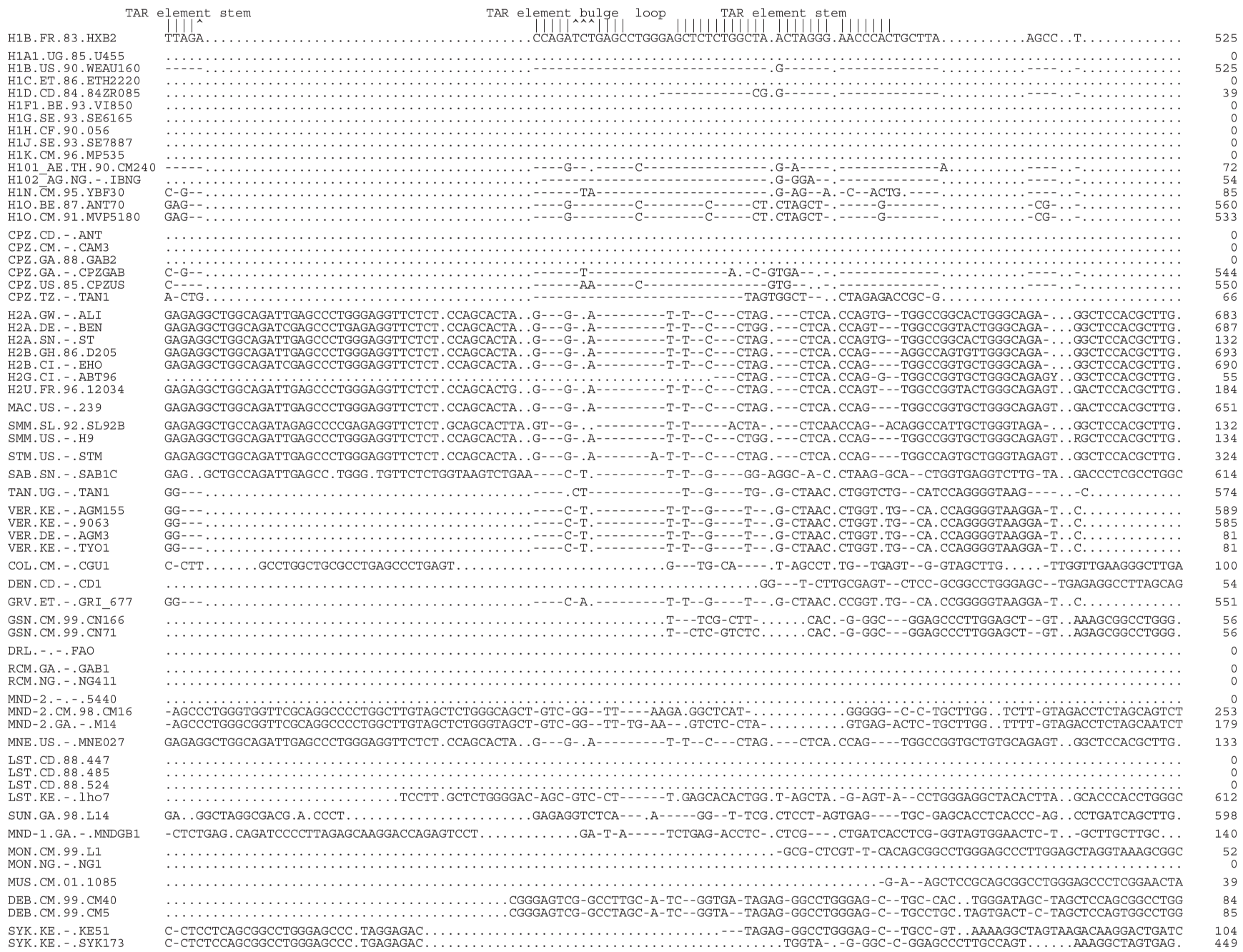
```
                                                                                                      |- TCF-1 alpha -|
H1B.FR.83.HXB2          GTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTAC...TTCAAGAACTGC........................TGACA...................   334

H1A1.UG.85.U455         ..................................................................................................................................     0
H1B.US.90.WEAU160       A---AT-----A----------AAA--------------G-A-----------------------..--T--AG-----...........................TGACA................   334
H1C.ET.86.ETH2220       ..................................................................................................................................     0
H1D.CD.84.84ZR085       ..................................................................................................................................     0
H1F1.BE.93.VI850        ..................................................................................................................................     0
H1G.SE.93.SE6165        ..................................................................................................................................     0
H1H.CF.90.056           ..................................................................................................................................     0
H1J.SE.93.SE7887        ..................................................................................................................................     0
H1K.CM.96.MP535         ..................................................................................................................................     0
H101_AE.TH.90.CM240     ..................................................................................................................................     0
H102_AG.NG.-.IBNG       ..................................................................................................................................     0
H1N.CM.95.YBF30         ..................................................................................................................................     0
H1O.BE.87.ANT70         AC-GAT-----A------T--ATCT----GCAACACC--TG-T--TATGATAACT--C--A---CT-...---C----G...........................G-CTAAAAACTGC........   343
H1O.CM.91.MVP5180       AC--A-A---CA------T--ATCAT---GC--AACA--T--A----TGC-AAA---C--A---CT-...---CCC--G...........................T-ACTGACACTGC........   343

CPZ.CD.-.ANT            ..................................................................................................................................     0
CPZ.CM.-.CAM3           ..................................................................................................................................     0
CPZ.GA.88.GAB2          ..................................................................................................................................     0
CPZ.GA.-.CPZGAB         C--G-TC---C-C---------A-G--G--C--AAGA--T--T---A----A-AA-----------...-A---AG--.........................----TTTAG...............   335
CPZ.US.85.CPZUS         C---ATC---CA--------CTGAA--G--TCGGAGA----GA--TAA-----------A------...---CG---..........................---.....................   329
CPZ.TZ.-.TAN1           ..................................................................................................................................     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.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
H2A.SN.-.ST             ..................................................................................................................................     0
H2B.GH.86.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
H2U.FR.96.12034         ..................................................................................................................................     0

MAC.US.-.239            TC---CA----A------TCCAACT--G--C-ACAC-T-TGA---ATAT-TTAGAT-C--A--AG-GT..--GG-AGCAA-T.....CAGGCCTGTCAGAGGAAGA-GTTAGAAGAAGGCTAACCGC..   364

SMM.SL.92.SL92B         ..................................................................................................................................     0
SMM.US.-.H9             ..................................................................................................................................     0

STM.US.-.STM            ........................................................................................TGGAAGGGATTTATTACAGTGAGA.GAAGGCTAACCGC..    37

SAB.SN.-.SAB1C          CC-G-T-----A-------CCA-TA------G-GG-CT-TG-T--TT-GAG------C--A---C-G...G-GCC-TCAGCA...CAAGGCTAACCACAAAACCACATC-TATG...............   352

TAN.UG.-.TAN1           A--G-T-----AA---...............AA--CA-T-C-A--AGTTC--TACA-C-----CAG-...--T...--GGA-...ATGCATGGTCTTGTTAAGAGGAAGT-GCCTAACCGCAGGCTTGCG   347

VER.KE.-.AGM155         A--G-CA----A------T...............-CAATGT----TGTTC--TACG-C--CTCAAGG...GAGT-CTTTACAGACTTATATTCAACAGTTGGTACAG--A-CTAGCCGACCACAGGCTTG   362
VER.KE.-.9063           C--G-TC----A-------...............-C-ATGT-A--AGTGC--TATG-C--AA-CAGA...CAGT-TTTTGAAGACATGCATGCAATAGTGAAGAGGAAGT-GCTAACCGCAGGCTTGTG.   361
VER.DE.-.AGM3           ..................................................................................................................................     0
VER.KE.-.TYO1           ..................................................................................................................................     0

COL.CM.-.CGU1           ..................................................................................................................................     0

DEN.CD.-.CD1            ..................................................................................................................................     0

GRV.ET.-.GRI_677        C--G-TC----A-------CCGAAGT-G--GG-GG-GT--CGC...................-----CATGTTTAAGG-C-TGCA-...GAACATGCAAAGCGCTAG.........................   335

GSN.CM.99.CN166         ..................................................................................................................................     0
GSN.CM.99.CN71          ..................................................................................................................................     0

DRL.-.-.FAO             ..................................................................................................................................     0

RCM.GA.-.GAB1           ..................................................................................................................................     0
RCM.NG.-.NG411          ..................................................................................................................................     0

MND-2.-.-.5440          ..................................................................................................................................     0
MND-2.CM.98.CM16        ..................................................................................................................................     0
MND-2.GA.-.M14          ..................................................................................................................................     0

MNE.US.-.MNE027         ..................................................................................................................................     0

LST.CD.88.447           ..................................................................................................................................     0
LST.CD.88.485           ..................................................................................................................................     0
LST.CD.88.524           ..................................................................................................................................     0
LST.KE.-.lho7           -C-TATC---CAC--CA-TCCAATGT-G---G-GG-CTTT--A----TCA-AAA---G--T-C-A-G...A-AC-A--TACT...GCTTTCGCTTTTGACTGCAAGA-GA-GTAGCCAAACCGC......   365

SUN.GA.98.L14           AC-TATC----CA-----TCC-ACT--G--C-A-G--TTT-G----AT-C--AA---C--T--AG-G...------C-TGTT...ACTAGTCTGCAGTGGGAGGTG-AG-TTAACCGCAAACCG......   365

MND-1.GA.-.MNDGB1       ..................................................................................................................................     0

MON.CM.99.L1            ..................................................................................................................................     0
MON.NG.-.NG1            ..................................................................................................................................     0

MUS.CM.01.1085          ..................................................................................................................................     0

DEB.CM.99.CM40          ..................................................................................................................................     0
DEB.CM.99.CM5           ..................................................................................................................................     0

SYK.KE.-.KE51           ..................................................................................................................................     0
SYK.KE.-.SYK173         CC-CAT-----AC--CA--CC--A---G--C-A-ACA-CAGGCTGGGAGAT-GCC-GCTT-C---TA...GAG-GAC-AACA...GGAAAACCACAAGAGCTGCAG-C-GCACTTAGTAAGAACATCTCC   262
```

PLV
complete genomes

See Pohlmann, J Virol 72(7):5589-5598(1998) and similar publications for information on this enhancer region

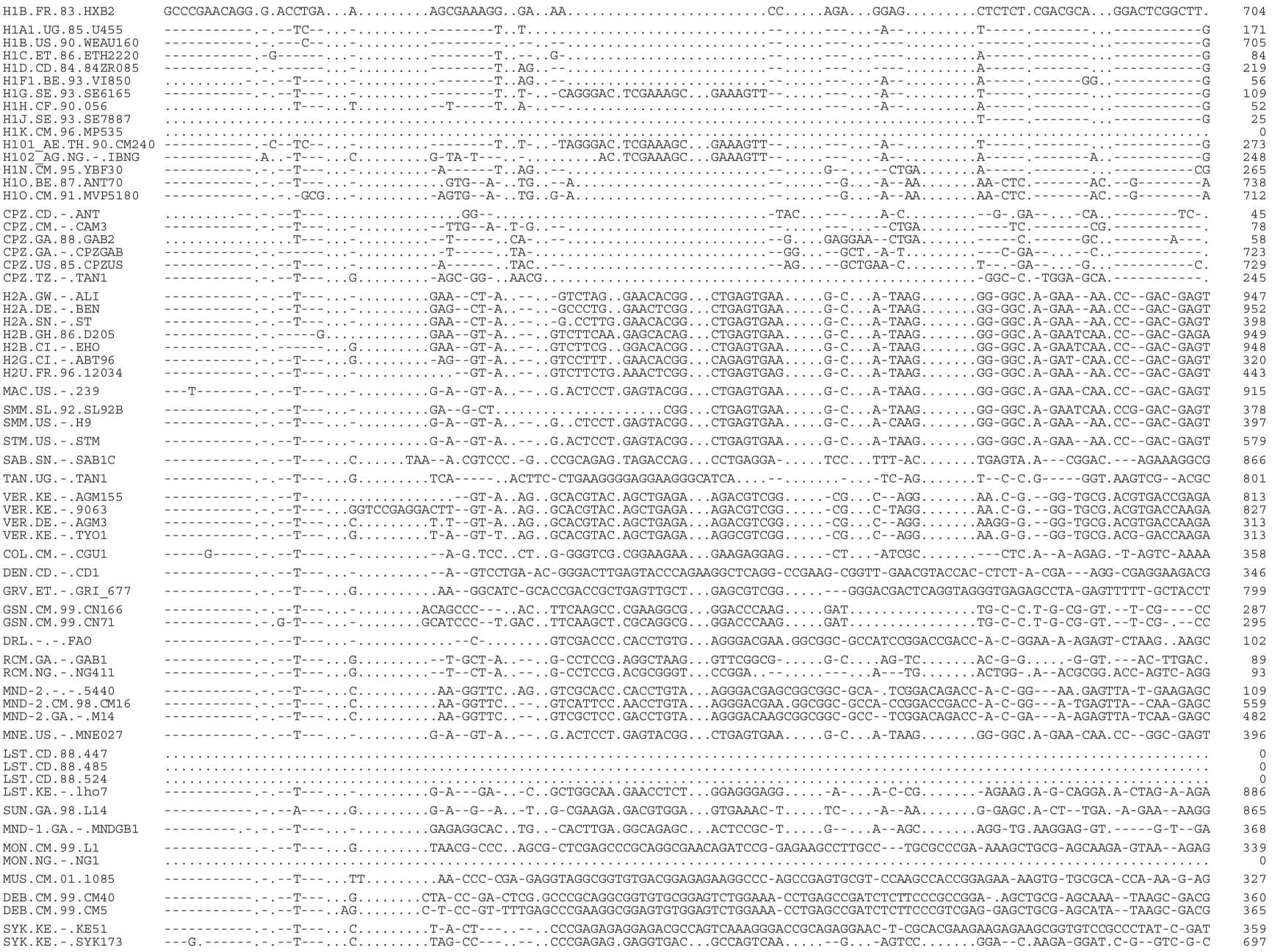
HIV-2 and SIVs have extra TAR stem-loop(s) see Berkhout, Nucleic Acids Res 20(1):27-31(1992)



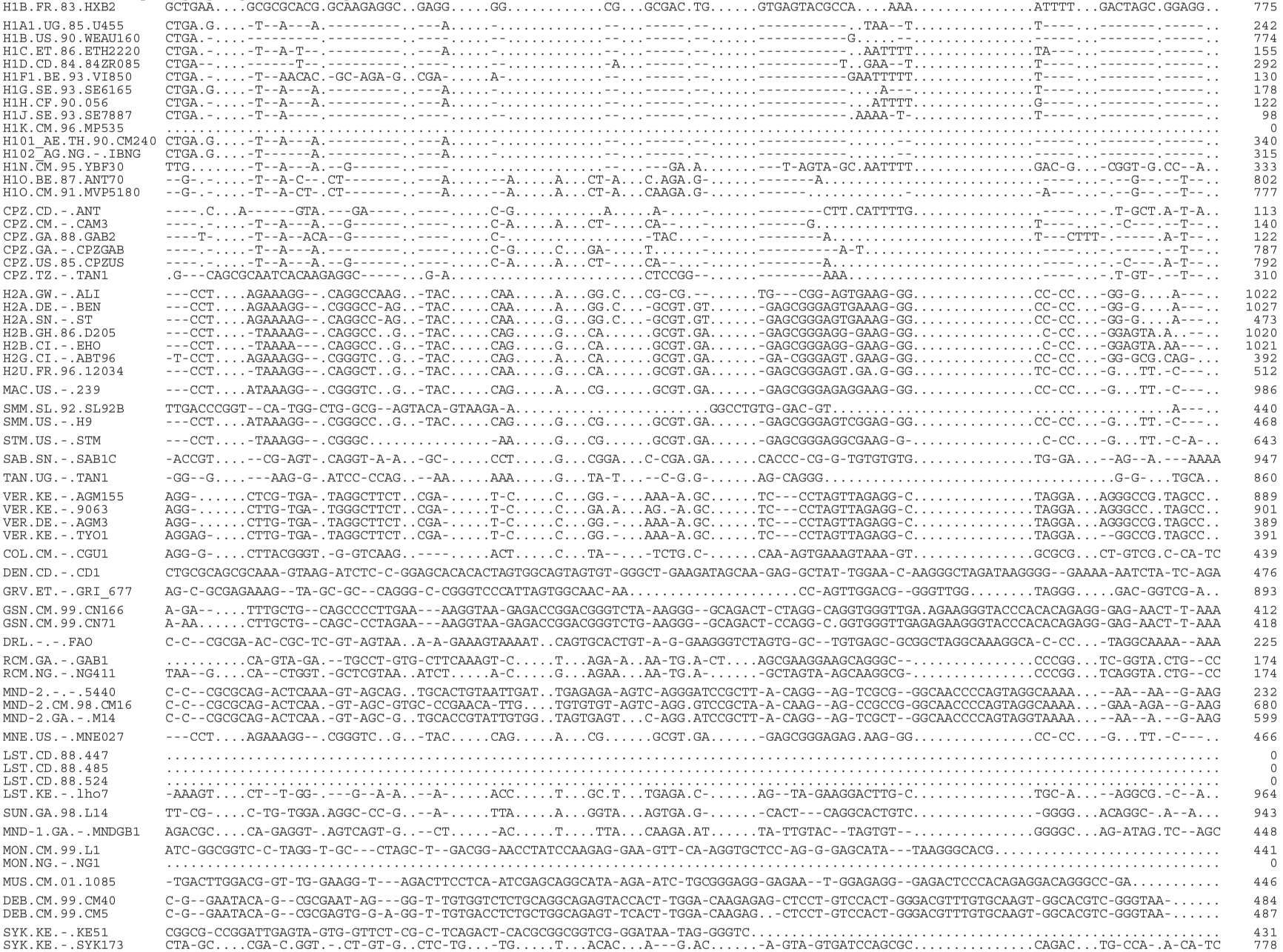

TAR element stem
TAR element bulge
loop
TAR element stem
H1B.FR.83.HXB2 TTAGA.....CCAGATCTGAGCCTGGGAGCTCTCTGGCTA.ACTAGGG.AACCCACTGCTTA.....AGCC..T..... 525
H1A1.UG.85.U455 0
H1B.US.90.WEAU160 -----.....-----.....G------.-----.....----..-..... 525
H1C.ET.86.ETH2220 0
H1D.CD.84.84ZR085 .....----------CG.G------.-----.....----..-..... 39
H1F1.BE.93.VI850 0
H1G.SE.93.SE6165 0
H1H.CF.90.056 0
H1J.SE.93.SE7887 0
H1K.CM.96.MP535 0
H101_AE.TH.90.CM240 -----.....----G--.-----C----------.G-A----.----------A.....----..-..... 72
H102_AG.NG.-.IBNG .....-----.....G-GGA--.-----.....----..-..... 54
H1N.CM.95.YBF30 C-G--.....------TA----------.G-AG--A.-C--ACTG.----.....----..-..... 85
H1O.BE.87.ANT70 GAG--.....----G--------C--------C-----CT.CTAGCT-.-----G-------.....-CG-..-..... 560
H1O.CM.91.MVP5180 GAG--.....----G--------C--------C-----CT.CTAGCT-.-----G-------.....-CG-..-..... 533
CPZ.CD.-.ANT 0
CPZ.CM.-.CAM3 0
CPZ.GA.88.GAB2 0
CPZ.GA.-.CPZGAB C-G--.....------T----------A.-C-GTGA--.-.----------.....----..-..... 544
CPZ.US.85.CPZUS C-----.....------AA-----C----------GTG--...-----.....----..-..... 550
CPZ.TZ.-.TAN1 A-CTG.....----------TAGTGGCT---..CTAGAGACCGC-G..... 66
H2A.GW.-.ALI GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAGTG--TGGCCGGCACTGGGCAGA-...GGCTCCACGCTTG. 683
H2A.DE.-.BEN GAGAGGCTGGCAGATCGAGCCCTGAGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTGG.---CTCA.CCAGT---TGGCCGGTACTGGGCAGA-...GGCTCCACGCTTG. 687
H2A.SN.-.ST GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAGTG--TGGCCGGCACTGGGCAGA-...GGCTCCACGCTTG. 132
H2B.GH.86.D205 GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAG----AGGCCAGTGTTGGGCAGA-...GGCTCCACGCTTG. 693
H2B.CI.-.EHO GAGAGGCTGGCAGATCGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAG----TGGCCGGTGCTGGGCAGA-...GGCTCCACGCTTG. 690
H2G.CI.-.ABT96 .....---CTAG.---CTCA.CCAG-G--TGGCCGGTGCTGGGCAGAGY..GGCTCCACGCTTG. 55
H2U.FR.96.12034 GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTG..G---G-.A---------T-T--C---CTAG.---CTCA.CCAGT---TGGCCGGTACTGGGCAGAGT..GACTCCACGCTTG. 184
MAC.US.-.239 GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAG----TGGCCGGTGCTGGGCAGAGT..GACTCCACGCTTG. 651
SMM.SL.92.SL92B GAGAGGCTGCCAGATAGAGCCCCGAGAGGTTCTCT.GCAGCACTTA.GT--G-..---------T-T-----ACTA-.---CTCAACCAG--ACAGGCCATTGCTGGGTAGA-...GGCTCCACGCTTG. 132
SMM.US.-.H9 GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTGG.---CTCA.CCAG----TGGCCGGTGCTGGGCAGAGT..RGCTCCACGCTTG. 134
STM.US.-.STM GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A-------A-T-T--C---CTAG.---CTCA.CCAG----TGGCCAGTGCTGGGTAGAGT..GGCTCCACGCTTG. 324
SAB.SN.-.SAB1C GAG..GCTGCCAGATTGAGCC.TGGG.TGTTCTCTGGTAAGTCTGAA----C-T.---------T-T--G----GG-AGGC-A-C.CTAAG-GCA--CTGGTGAGGTCTTG-TA..GACCCTCGCCTGGC 614
TAN.UG.-.TAN1 GG---.....-----.CT----------T--G----TG-.G-CTAAC.CTGGTCTG--CATCCAGGGGTAAG----..-C..... 574
VER.KE.-.AGM155 GG---.....----C-T.---------T-T--G----T--.G-CTAAC.CTGGT.TG--CA.CCAGGGGTAAGGA-T..C..... 589
VER.KE.-.9063 GG---.....----C-T.---------T-T--G----T--.G-CTAAC.CTGGT.TG--CA.CCAGGGGTAAGGA-T..C..... 585
VER.DE.-.AGM3 GG---.....----C-T.---------T-T--G----T--.G-CTAAC.CTGGT.TG--CA.CCAGGGGTAAGGA-T..C..... 81
VER.KE.-.TYO1 GG---.....----C-T.---------T-T--G----T--.G-CTAAC.CTGGT.TG--CA.CCAGGGGTAAGGA-T..C..... 81
COL.CM.-.CGU1 C-CTT.....GCCTGGCTGCGCCTGAGCCCTGAGT.....G---TG-CA----.T-AGCCT.TG--TGAGT--G-GTAGCTTG.....-TTGGTTGAAGGGCTTGA 100
DEN.CD.-.CD1 .....GG---T-CTTGCGAGT--CTCC-GCGGCCTGGGAGC--TGAGAGGCCTTAGCAG 54
GRV.ET.-.GRI_677 GG---.....----C-A.---------T-T--G----T--.G-CTAAC.CCGGT.TG--CA.CCGGGGGTAAGGA-T..C..... 551
GSN.CM.99.CN166 .....T---TCG-CTT-.....CAC.-G-GGC---GGAGCCCTTGGAGCT--GT..AAAGCGGCCTGGG. 56
GSN.CM.99.CN71 .....T--CTC-GTCTC.....CAC.-G-GGC---GGAGCCCTTGGAGCT--GT..AGAGCGGCCTGGG. 56
DRL.-.-.FAO 0
RCM.GA.-.GAB1 0
RCM.NG.-.NG411 0
MND-2.-.-.5440 0
MND-2.CM.98.CM16 -AGCCCTGGGTGGTTCGCAGGCCCCTGGCTTGTAGCTCTGGGCAGCT-GTC-GG--TT----AAGA.GGCTCAT-.....GGGGG--C-C-TGCTTGG..TCTT-GTAGACCTCTAGCAGTCT 253
MND-2.GA.-.M14 -AGCCCTGGGCGGTTCGCAGGCCCCTGGCTTGTAGCTCTGGGTAGCT-GTC-GG--TT-TG-AA---.GTCTC-CTA-.....GTGAG-ACTC-TGCTTGG..TTTT-GTAGACCTCTAGCAATCT 179
MNE.US.-.MNE027 GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCT.CCAGCACTA..G---G-.A---------T-T--C---CTAG.---CTCA.CCAG----TGGCCGGTGCTGTGCAGAGT..GGCTCCACGCTTG. 133
LST.CD.88.447 0
LST.CD.88.485 0
LST.CD.88.524 0
LST.KE.-.lho7 .....TCCTT.GCTCTGGGGAC-AGC-GTC-CT------T.GAGCACACTGG.T-AGCTA.-G-AGT-A--CCTGGGAGGCTACACTTA..GCACCCACCTGGGC 612
SUN.GA.98.L14 GA..GGCTAGGCGACG.A.CCCT.....GAGAGGTCTCA---.A-----GG--T-TCG.CTCCT-AGTGAG----TGC-GAGCACCTCACCC-AG..CCTGATCAGCTTG. 598
MND-1.GA.-.MNDGB1 -CTCTGAG.CAGATCCCCTTAGAGCAAGGACCAGAGTCCT.....GA-T-A-----TCTGAG-ACCTC-.CTCG---.CTGATCACCTCG-GGTAGTGGAACTC-T..-GCTTGCTTGC... 140
MON.CM.99.L1 .....-GCG-CTCGT-T-CACAGCGGCCTGGGAGCCCTTGGAGCTAGGTAAAGCGGC 52
MON.NG.-.NG1 0
MUS.CM.01.1085 .....-G-A--AGCTCCGCAGCGGCCTGGGAGCCCTCGGAACTA 39
DEB.CM.99.CM40 .....CGGGAGTCG-GCCTTGC-A-TC--GGTGA-TAGAG-GGCCTGGGAG-C--TGC-CAC..TGGGATAGC-TAGCTCCAGCGGCCTGG 84
DEB.CM.99.CM5 .....CGGGAGTCG-GCCTAGC-A-TC--GGTA--TAGAG-GGCCTGGGAG-C--TGCCTGC.TAGTGACT-C-TAGCTCCAGTGGCCTGG 85
SYK.KE.-.KE51 C-CTCCTCAGCGGCCTGGGAGCCC.TAGGAGAC.....---TAGAG-GGCCTGGGAG-C--TGCC-GT..AAAAGGCTAGTAAGACAAGGACTGATC 104
SYK.KE.-.SYK173 C-CTCTCCAGCGGCCTGGGAGCCC.TGAGAGAC.....TGGTA-.-G-GGC-C-GGAGCCCTTGCCAGT.....AAAGGCTAGTGAG. 449

```
                                                                                                        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.83.HXB2       ........................................................................................C.AATAAA.GCTTGCCTTG..AGT....GCTTCA.AGTAGTG   558

H1A1.UG.85.U455      ..........................................................................................--.----------.---...----G-.-------    28
H1B.US.90.WEAU160    ........................................................................................-.------.----------.---...------.-------   558
H1C.ET.86.ETH2220    ..................................................................................................................................     0
H1D.CD.84.84ZR085    ........................................................................................-.------.----------.---...----T-.-------    72
H1F1.BE.93.VI850     ..................................................................................................................................     0
H1G.SE.93.SE6165     ..................................................................................................................................     0
H1H.CF.90.056        ..................................................................................................................................     0
H1J.SE.93.SE7887     ..................................................................................................................................     0
H1K.CM.96.MP535      ..................................................................................................................................     0
H101_AE.TH.90.CM240  ........................................................................................-.------.----------.---...----A-.---G---   105
H102_AG.NG.-.IBNG    ........................................................................................-.------.----------.---...------.-------    87
H1N.CM.95.YBF30      ........................................................................................-.------.----------.---...---AA-..--G---   117
H1O.BE.87.ANT70      ........................................................................................-.------.----------.---GAGAAGCA.......---   591
H1O.CM.91.MVP5180    ........................................................................................-.------.----------.---GAGAAGCA.......---   564

CPZ.CD.-.ANT         ..................................................................................................................................     0
CPZ.CM.-.CAM3        ..................................................................................................................................     0
CPZ.GA.88.GAB2       ..................................................................................................................................     0
CPZ.GA.-.CPZGAB      ........................................................................................-.------.----------.---..GTA--TG.--C----   578
CPZ.US.85.CPZUS      ........................................................................................-.------.----------.---...----T-.-------   583
CPZ.TZ.-.TAN1        ..................................................CTTA.......ACGC.......................TC------.--C-----G..--A...-TG-T..-AC----   106

H2A.GW.-.ALI         ..................................................CTTG.CT.T..AAAG........ACCTCTT........------G.--.---AAT...TAGAAGCAGGTT.-AAG---   735
H2A.DE.-.BEN         ..................................................CTTG.CT.T..AAAG........ACCTCTT.......-.------G.--.---A.-..TTAGAAGCAAGTT.-AGT---   740
H2A.SN.-.ST          ..................................................CTTG.CT.T.AAAAG........ACCTCTT........------G.--.---AGT...TAGAAGCAAGTT.-AGT---   185
H2B.GH.86.D205       ..................................................CTTG.CT.T.AAAA.........CCCTCTT........------..--.---A.A..TTAGAAGCAAGT..-AGT---   743
H2B.CI.-.EHO         ..................................................CTTG.CT.T.AAAA.........CCCTCTT........------A.--.----.A..TTAGAAGCAAGAC.-AGT---   742
H2G.CI.-.ABT96       ..................................................CTTG.CT.TACAAAG........ACCTCTT........------G.--.---AAA..TTAGAAGTAAG--AGTGT---   111
H2U.FR.96.12034      ..................................................CTTG.CT.T..AAA.......ACCCTCT.......-.------G.--.---AGT...TAGAAGCAAG-TA-.GT---   236

MAC.US.-.239         ..................................................CTTG.CT.T..AAAG........CCCTCTT.......-.------G.--.---A-T..TTAGAAGTAAG-T.-.GT---   704

SMM.SL.92.SL92B      ..................................................CTTG.CT.T.....G.........ACTCT........-.------G---.--AACT...TAGAAGCAAGTC.-AG----   181
SMM.US.-.H9          ..................................................CTTG.CT.T.AAAAG........ACCTCTT.......-.------G.--.---AGT...TAGAAGTAAG-C.-.GT---   187

STM.US.-.STM         ..................................................CTTG.CT.T..ATAT........ATCTCTT.......-.------G.--.---AAT..TTAGAAGTAAG.C.-AGT---   377

SAB.SN.-.SAB1C       GAC.TGGCCATTGCCAGTAGCAGAGACTCCG...................CTTG.CT.T.GCTTGA.......TTGCCT.........-.------GT--------A..GT-TACAAG-AA-.GCA----   701

TAN.UG.-.TAN1        ..................................................CTTG.CT.TAGAAGGCTA...TTG..............-T------.-------GC..TTAAGAGT--ATC.T-AG.CA   629

VER.KE.-.AGM155      ..................................................CTTGGCT.TAGAAAG......................-T------.-------GC..-T-AGAGCT-ATC.T-AGTCA   640
VER.KE.-.9063        ..................................................CTTGGCT.TGGAAAG......................-T------ACT------GC..-T-AGAGCT-ATC.T-AGTCA   638
VER.DE.-.AGM3        ..................................................CTTGGCT.TAGAAAG......................-T------.T---CG--GC..-T-AGAGCT-CTC.T-AGTCA   133
VER.KE.-.TYO1        ..................................................CTTGGCT.TAGAAAG......................-T------.-------GC..-T-AGAGCT-ATC.T-AGTCA   132

COL.CM.-.CGU1        GGGACTC..........CAGTCCCCTCGCGCTAAGGGTAGATA.......GGTG.CTTTGCATTAT......AAGAG...........-.------GTGA-TTTGAA..GC-TCAA--AG-G.TT-GAGT   191

DEN.CD.-.CD1         CGGCCTGGGAGCCCTCTAAGGTATTGCAACAAGTACCCAGACAAG.....CTTG.CT.CGTGCTGC.......TTTTGATGACCAT..A.------GT-AGTTAACTGGC....................   147

GRV.ET.-.GRI_677     ..................................................CTTGGCT.TCATATAGC.....................TC------CCTGCT-GC-TAGTCGCTATAT-GG-GTCA----   609

GSN.CM.99.CN166      ................AGCCCTTTGCTAGAAGACGGTAGAATTTCG....CTTG.CT.TGCCTAAGGC...TTATGACT..........------.CT-CCAAAG-..-TAA.CGAG-CTGCTC-CA-T   146
GSN.CM.99.CN71       ................AGCCCTTGCCTAGAGGACGGTAGATTCTCGG...CTTG.CT.TGCCCTAGGGCTTTAGTGGCT.........-.------GCT-C-AAAG-..TACAGCAAG-CTGCTC-CA-T   153

DRL.-.-.FAO          ..................................................................................................................................     0

RCM.GA.-.GAB1        ..................................................................................................................................     0
RCM.NG.-.NG411       ..................................................................................................................................     0

MND-2.-.-.5440       ..................................................................................................................................     0
MND-2.CM.98.CM16     C...GGTGGCCAAGAGGCTGTGGGATTGACTACCG...............CTTGC.....TTGCT.......TTTAAG.TGCT.....-.------A----A--CAAG.TTAAGAAAAGCA-.GTG----   344
MND-2.GA.-.M14       C....TAGGCCAGGAGATTGTGGGTTAGACTACCG...............CTTGC.....TTGCT.........ATTGATGCT.....-.------G----A---GA..TTAGAA.TAGCA-.GTA----   266

MNE.US.-.MNE027      ..................................................CTTG.CT.T..AAAG........ACCTCTT.......-.------G.--.---A-T...TAGAAGTAAG-C.-.GC---   185

LST.CD.88.447        ..................................................................................................................................     0
LST.CD.88.485        ..................................................................................................................................     0
LST.CD.88.524        ..................................................................................................................................     0
LST.KE.-.lho7        CTGATCA..............GCCTAGGGAGC.TAAGGGCTCCTAG....CTTG.CT...ACTTTAAA...G....CCTT........-.------AAG----TCAT..T--TAGAAAG-AC.-AGCTA-   700

SUN.GA.98.L14        ..................................GGGAGCTAGAGGCTC.CTTG.CT.TTGCAGTACA...AG...CCTT........-.------G.-C------A..GT-TTACTGCAAG.CAA...-   669

MND-1.GA.-.MNDGB1    TATTGTCTTC..............................................................................------GTAACTTAGAA..TTAGAGCAAG-G-.GTA----   187

MON.CM.99.L1         CTGGGAGCCCTTGCCTAGAGAGTTGGTGAAGCTGACCCCGGCTCG.....CTTGC.......TTTAGTAAGGCCTTACAGCC......-.------GTGC-TGAAA-.......................   140
MON.NG.-.NG1         ..................................................................................................................................     0

MUS.CM.01.1085       GGTATAGCGGCCTGGGAGCCCTTGCCTAGGAGTACTGATTCTCGG.....CTTGC.......TTTGCTTCGAAGCTTATTGAAGCCCA-.------GTTG-AGAAC.......................   132

DEB.CM.99.CM40       GAGCCCTGAGCTGGGGTAGGCA.CTCCTCGGCACG..............CTTG.CT.TGCTACGCTT..CTATAGCCCA........G.------GTG-G-T.AA.........CTGC....-AGT---   171
DEB.CM.99.CM5        GAGCCCTGAGCTGGGGTTAGGTACTCCTCGGCACG..............CTTG.CT.TGCT.GCATT..TTATTGCCCA........G.------GTG-G-T.AA.........CTGC....-AGT---   172

SYK.KE.-.KE51        CACCTAGCCTCTACAATTCT.........CCGTTA..............CTTA.CT.TTACTACGCG..CTGATCAGCC........-.------GTG-GTT.AA.........CTAC....-AGTAAC   183
SYK.KE.-.SYK173      GTCCGGGTTTG.....GCCGCCCGGTCTTCGGCAGCTCTCCGTTG.....CTTGC.TTTGCTTACGCT...GATCAGCCCA.........------GTG-G-TAAAC..GCAAGTAA-G--T.GC-CTCA   553
```

see Zhang, *J. Virol* **72**(7):5664-5471(1998)
5' LTR U5 end \/ Lys tRNA binding site


H1B.FR.83.HXB2 704
H1A1.UG.85.U455 171
H1B.US.90.WEAU160 705
H1C.ET.86.ETH2220 84
H1D.CD.84.84ZR085 219
H1F1.BE.93.VI850 56
H1G.SE.93.SE6165 109
H1H.CF.90.056 52
H1J.SE.93.SE7887 25
H1K.CM.96.MP535 0
H101_AE.TH.90.CM240 273
H102_AG.NG.-.IBNG 248
H1N.CM.95.YBF30 265
H1O.BE.87.ANT70 738
H1O.CM.91.MVP5180 712
CPZ.CD.-.ANT 45
CPZ.CM.-.CAM3 78
CPZ.GA.88.GAB2 58
CPZ.GA.-.CPZGAB 723
CPZ.US.85.CPZUS 729
CPZ.TZ.-.TAN1 245
H2A.GW.-.ALI 947
H2A.DE.-.BEN 952
H2A.SN.-.ST 398
H2B.GH.86.D205 949
H2B.CI.-.EHO 948
H2G.CI.-.ABT96 320
H2U.FR.96.12034 443
MAC.US.-.239 915
SMM.SL.92.SL92B 378
SMM.US.-.H9 397
STM.US.-.STM 579
SAB.SN.-.SAB1C 866
TAN.UG.-.TAN1 801
VER.KE.-.AGM155 813
VER.KE.-.9063 827
VER.DE.-.AGM3 313
VER.KE.-.TYO1 313
COL.CM.-.CGU1 358
DEN.CD.-.CD1 346
GRV.ET.-.GRI_677 799
GSN.CM.99.CN166 287
GSN.CM.99.CN71 295
DRL.-.-.FAO 102
RCM.GA.-.GAB1 89
RCM.NG.-.NG411 93
MND-2.-.-.5440 109
MND-2.CM.98.CM16 559
MND-2.GA.-.M14 482
MNE.US.-.MNE027 396
LST.CD.88.447 0
LST.CD.88.485 0
LST.CD.88.524 0
LST.KE.-.lho7 886
SUN.GA.98.L14 865
MND-1.GA.-.MNDGB1 368
MON.CM.99.L1 339
MON.NG.-.NG1 0
MUS.CM.01.1085 327
DEB.CM.99.CM40 360
DEB.CM.99.CM5 365
SYK.KE.-.KE51 359
SYK.KE.-.SYK173 697
PLV
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)



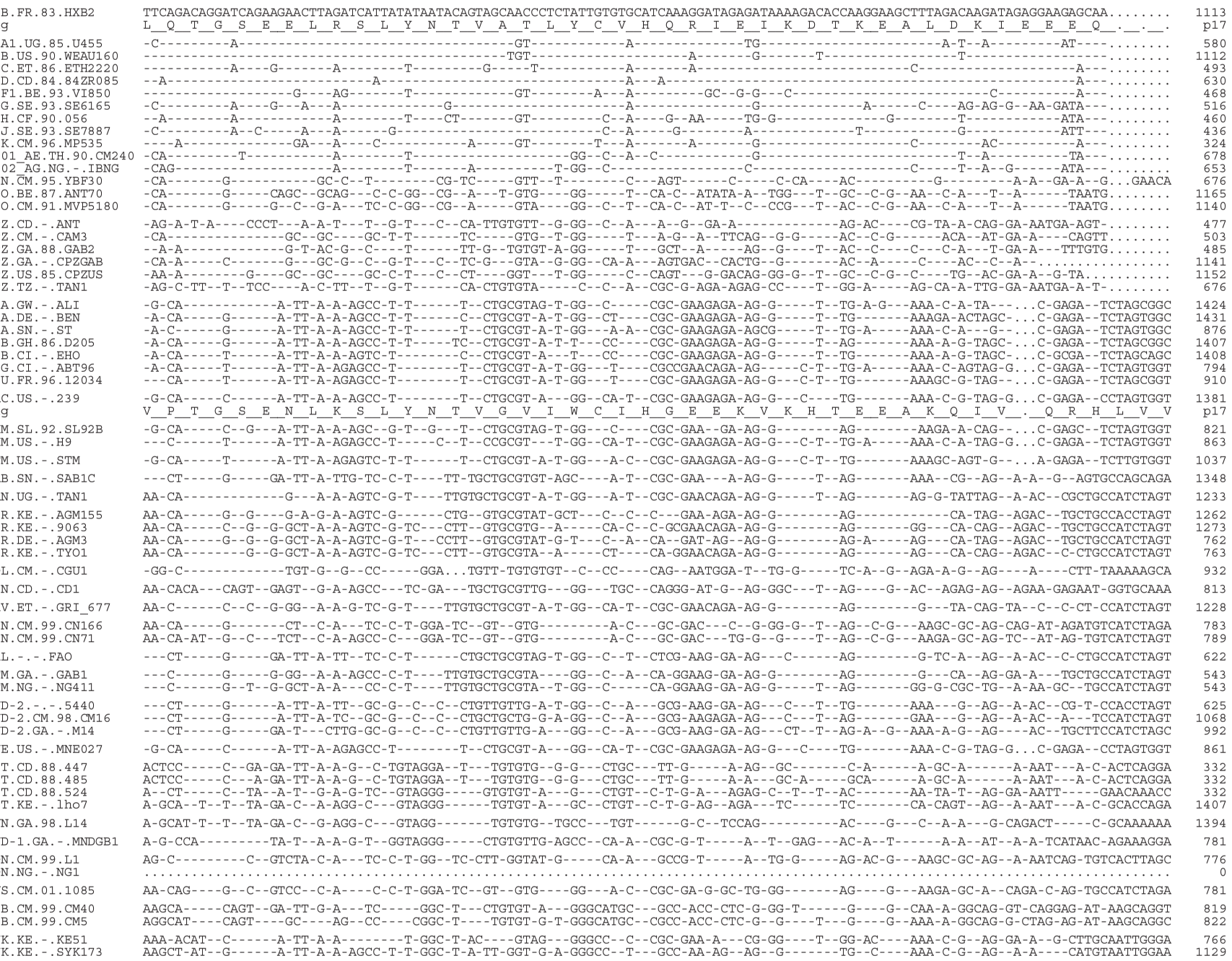
```
H1B.FR.83.HXB2        ...CTA....GAAGGAG..AGAG..........................................................................................................   789

H1A1.UG.85.U455       ...---....-------..---A..........................................................................................................   256
H1B.US.90.WEAU160     ...---....-------..----..........................................................................................................   788
H1C.ET.86.ETH2220     ...---....-------..----..........................................................................................................   169
H1D.CD.84.84ZR085     ...---....-------..----..........................................................................................................   306
H1F1.BE.93.VI850      ...---....-------..----..........................................................................................................   144
H1G.SE.93.SE6165      ...---....-----G-..----..........................................................................................................   192
H1H.CF.90.056         ...---....-------..----..........................................................................................................   136
H1J.SE.93.SE7887      ...---....-------..----..........................................................................................................   112
H1K.CM.96.MP535       .................................................................................................................................     0
H101_AE.TH.90.CM240   ...---....-------..----..........................................................................................................   354
H102_AG.NG.-.IBNG     ...---....-------..----..........................................................................................................   329
H1N.CM.95.YBF30       ...AAG....T-G-AGA..GAG-..........................................................................................................   347
H1O.BE.87.ANT70       ...-C-....--CCT--..G-GAAGGG.CGA..AGTCTCTAGG.G..GAGGAAG...........................................................................   841
H1O.CM.91.MVP5180     ...-C-....--CCT--..G-GAAGGG.CGA..AGTCCCTAGG.G..GAGGAAG...........................................................................   816

CPZ.CD.-.ANT          ...GC-....ATCCT--..G-GAAGGT.CGA..AGTCTCTAGGAA..CAGGAAG...........................................................................   153
CPZ.CM.-.CAM3         ...--G....--CCT--..G-GAAGGG.CGA..AGTCTCTAGG.G..GAAGAGG...........................................................................   179
CPZ.GA.88.GAB2        ...---....--CCT--..--GAAGGG.CGA..AGTCTCTAGGTG..GA.GAGG...........................................................................   161
CPZ.GA.-.CPZGAB       ...---....--CCT--..--GAAGGG.CGA..AGTCTCTAGG.G..GGAGAAG...........................................................................   826
CPZ.US.85.CPZUS       ...--G....--CCT--..G-GAAGGG.CGA..AGTCTCTAGG.G..GAAGAGG...........................................................................   831
CPZ.TZ.-.TAN1         ...GC-....ACCCT--..--GAAGGG.CGA..AGTCTCTAGGTAACAGGGGAA...........................................................................   352

H2A.GW.-.ALI          ...TA-....-T-CCTA..CACCAAAA.T.T..GTAGCC.GAAAG..GGCTTGTTATCCTACCTTTAG...ACAGGTAGA...AGATT..GTGGGAG................................  1095
H2A.DE.-.BEN          ...TA-....-T-CCTA..CACCAAAA.ACT..GTAGCCAGAAAA..GGCTTGTTATCCTACCTTTAG...ACAGGTAGA...AGATT..GTGGGAG................................  1102
H2A.SN.-.ST           ...TA-....-TGCCTA..CACCAAAT.ACA..GTAGCCAGAA.G..GGCTTGTTATCCTACCTTTAG...ACGGGTAGA...AGATT..GTGGGAG................................   547
H2B.GH.86.D205        ...TA-....-T-CTCC..-A-....................TTT..TGCATTGTCTACTCTAAAGAG...GGAGTAGGG...CATAA..GTGGGAG................................  1078
H2B.CI.-.EHO          ...TA-....-T-CTCC..-A-TTT......................GCAATTCGTCTACTCTAGGAC...AGGGTACGG...AATAA..GTGGGAG................................  1079
H2G.CI.-.ABT96        ...TA-....-TGCACC..-A-AATTT.GTA..GCCAAAAAGGGC..TGTGAATCCTACCTTGAATAG...TCAGGTAGA...AGT....GTGGGAG................................   465
H2U.FR.96.12034       ...TA-....-TGCA-C..-A-AAATA.GA...ACCAA..AGGGT..TGTCTGCCTCTGAGAAAGGAG....AGGTAGT...ATATT..GTGGGAG................................   582

MAC.US.-.239          ...TA-....-TGCA-C..-C-AAAAA.GAA..A........TAG..CTGTCTT.TTATCCAGGAAGG...GGTAATAAG...ATAGA..GTGGGAG................................  1052

SMM.SL.92.SL92B       ...T.................CCAAAA............GGTGTG..AATATCCTACCTACGGGATAG...GTAGAAGTA......AA..GCGGGAG................................   492
SMM.US.-.H9           ...TA-....-TGCA-C..-A-A............GAGTCATAGG..ACTGAGTTCCCTACTTTTGAG...AAAAGAGTA...GGAGA..GTGGGAG................................   534

STM.US.-.STM          ...TG-....-TGCA-C..-AGTAGCC.AGA............GG..GCTGATTGTTCTACTCTTGAG...GGAAGAGGT...AGAAG..GTGGGAG................................   708

SAB.SN.-.SAB1C        GAGAAG....TC--ACA..G-GCTGAC.AAG..GCTGGCTTAGCA..GTGGCAGAAGGCTAGGTCGAG...CGGGAAAGG...AAA...........................................  1016

TAN.UG.-.TAN1         ...GGT....AG-TCTA..G-T-GGCG.ACA..GATCAGCACCTG..GGTGAGCA..........................................................................   901

VER.KE.-.AGM155       ...G--....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGACG..GGTACGTA..........................................................................   930
VER.KE.-.9063         ...A--....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGAC...GGTACGCA..........................................................................   941
VER.DE.-.AGM3         ...A--....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGACG..GGTACGCA..........................................................................   430
VER.KE.-.TYO1         ...G--....ACTACTC..T-G-CAAG.TAG..GGCAGGCGGT.G..GGTACGCA..........................................................................   431

COL.CM.-.CGU1         AGGAAG....TGGCAGA..C---AGAG.CAC..AGAGAGGGAGTG..CTCGCCAGAGAAGGAGAAGAG...CTCATTAGG...AGGTC..TTCAGGCAGGGAGTAGCGCCCCTGGGCAGGTCGAGGGGG   549

DEN.CD.-.CD1          GG...............................................................................................................................   478

GRV.ET.-.GRI_677      ...GC-...........................................................................................................................   896

GSN.CM.99.CN166       GAGGAGTAACA--CTCCCACCAGAGGATAGGGCAGA.............................................................................................   448
GSN.CM.99.CN71        AAGAAGTAACA--CTCCCACCAGAGGATAGGGCAGA.............................................................................................   454

DRL.-.-.FAO           GAAAA-GGGAGA-GTA--GCCTGTTGCT.GAT..AAGTGAGGCAAAAGAAGCCTA.GGGGTAAGAAAAG.............................................................   290

RCM.GA.-.GAB1         TCTAGG....-------..C--AGTCC.TAG..AGAAGGGAGGAA..A.................................................................................   211
RCM.NG.-.NG411        TCTAGG....-----GT..C--AGTCC.TAG..AGAAGGGAGGAA..A.................................................................................   211

MND-2.-.-.5440        TAAGGCCGACGC--A-T-AA--TAAGGG.CTT..CCTCCGAAGGGAGAAGCAGGGGGCAAGAAA..................................................................   293
MND-2.CM.98.CM16      TAAGGCCGAGGC--A-T-AA-AGAGGTC.CTT..GTGAAAGAGGA.....CAGGGGGCAAGAAA..................................................................   736
MND-2.GA.-.M14        TAAGGCCGAAGC--A-T-AA--TAAAAG.GGC..CTTCCGAAGGGAAGGGCAGGGGGCAAGAAA..................................................................   660

MNE.US.-.MNE027       ...TA-....-TGCA-C..-C-AAAAA.GAA..A........TAG..CTGTCTT.TTATCCAGGAAGG...GATAATAAG...ATAGA..GTGGGAG................................   532

LST.CD.88.447         .................................................................................................................................     0
LST.CD.88.485         .................................................................................................................................     0
LST.CD.88.524         .................................................................................................................................     0
LST.KE.-.lho7         ...AGG....T------..--TAAGAC.CTC..ACCGGGAGAGTG..GCAGCAGGCCGGGGGCAGCCT...TGTGAGAAG...TGTGT..AATTCTGATCTAGTGTAACTGGGACACTAGTGGTAGCTT  1071

SUN.GA.98.L14         ..AG-C....CT---CT..G-G-GGAG.GTC..CCGCTAGGTTAG..TACTGTCAAGTAGGCAGACTA...GTAGTACGT...GGAAT..AAAATAAGAAGTGTGTGTATAACTCTGTGTCTCGGAAGA  1051

MND-1.GA.-.MNDGB1     A................................................................................................................................   449

MON.CM.99.L1          .................................................................................................................................   441
MON.NG.-.NG1          .................................................................................................................................     0

MUS.CM.01.1085        .................................................................................................................................   446

DEB.CM.99.CM40        .................................................................................................................................   484
DEB.CM.99.CM5         .................................................................................................................................   487

SYK.KE.-.KE51         .................................................................................................................................   431
SYK.KE.-.SYK173       GGGGA-....AT---G-..-CC...........................................................................................................   794
```

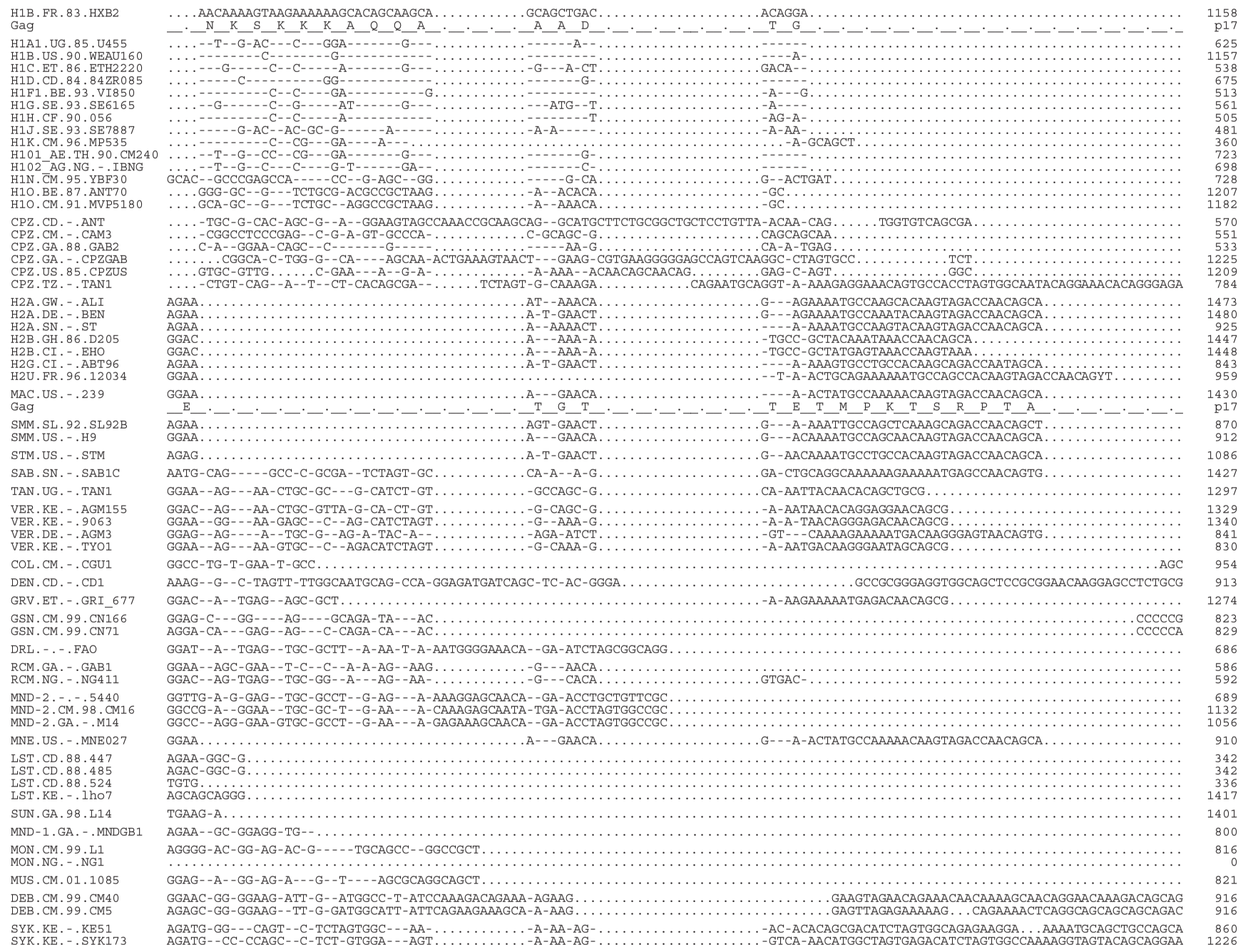
```
                                                                                      / Gag and Gag-Pol start (p17)
H1B.FR.83.HXB2          ..........................................................ATGGGTGCGAGAGCGTCAGTA...TTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAGGG   861
Gag                                                                               _M__G__A__R__A__S__V__.__L__S__G__G__E__L__D__R__W__E__K__I__R__L__R__P__G_   p17

H1A1.UG.85.U455         ..........................................................---------------------...---------AA-A--------TC-----G---------------------   328
H1B.US.90.WEAU160       ..........................................................-----------------G---...---------------------A-G---------------------------   860
H1C.ET.86.ETH2220       ..........................................................------------------A--...-----A--C-A-A--------GCC------------AA-------------   241
H1D.CD.84.84ZR085       ..........................................................---------------------...------------A----G---GCT------------------C--------   378
H1F1.BE.93.VI850        ..........................................................------------------A--...------------A--------GA--------------A---------G---   216
H1G.SE.93.SE6165        ..........................................................---------------------...----C-------A--------GCT-----------------G---------   264
H1H.CF.90.056           ..........................................................---------------------...--------C---A--------GCT-----G---------C-----------   208
H1J.SE.93.SE7887        ..........................................................------------------A--...-----T------A--------GAT-----------------G---------   184
H1K.CM.96.MP535         ..........................................................---------------------...------------A--------GC---------G---------C--------    72
H101_AE.TH.90.CM240     ..........................................................---------------------...-----T------A--------GC------------------GC--------   426
H102_AG.NG.-.IBNG       ..........................................................---------------------...-----T------A--------GC------G---------------------   401
H1N.CM.95.YBF30         ..........................................................--------------------G...----CA------A---------A-------TC----TAT--G--A------   419
H1O.BE.87.ANT70         ..........................................................-----------T-----T--G...--G-CA--AA-CA----G---GC-------C-----A------A------A   913
H1O.CM.91.MVP5180       ..........................................................-----------------T--G...--G-CA---A-TA----G---GC-------CG----A-------------A   888

CPZ.CD.-.ANT            ..........................................................-----A---G-G-----T--T...--G--G--A-AGA-GC-----AC-------GT--CA--C-TC----C--T   225
CPZ.CM.-.CAM3           ..........................................................---------------------...----CA------CG-------GCT-------------A--C-----------   251
CPZ.GA.88.GAB2          ..........................................................-----------------TA--...----C---C--GCG-------GCC-----G------A--C-C--A------   233
CPZ.GA.-.CPZGAB         ..........................................................--------------------T...C---CA------A----------C---------G-----C-T-----C---   898
CPZ.US.85.CPZUS         ..........................................................--------------------T...C-G-CA------CG-------GCT--------G------C-T--A--T---   903
CPZ.TZ.-.TAN1           ..........................................................--------------------G...--G--G--A-ATA-GC-G---AC-------TCC--AA--C-T-AAT-CA-A   424

H2A.GW.-.ALI            ..........................................................-----C------AAC--C--C...--G--A---A--A--GC---CGA--TA----G----A-----C----C--C  1167
H2A.DE.-.BEN            ..........................................................-----C------AAC--C--C...--G--A---AA-A--GC---CGA--TA------G--A-----C----C--C  1174
H2A.SN.-.ST             ..........................................................-----C------AAC--C--C...--G--A---AA-A--GC---CGA--TA-----G---A-----C----C--C   619
H2B.GH.86.D205          ..........................................................-----C-------GC--C--C...C-GTCA---AAGA--AC----GA--TA------G--A-----C----C--C  1150
H2B.CI.-.EHO            ..........................................................-----C-------GC--C--C...--GTCA---AAGA-GAC----GA--TA-----GG--A-----C----C--C  1151
H2G.CI.-.ABT96          ..........................................................-----C------AGC--C--C...C-GTCA---AA-A--GCT---GA--TA------G--A-----C----C--C   537
H2U.FR.96.12034         ..........................................................-----C------AAC--.--C...--GTCA---AAGA--G-----GA--TA------G--A-----C----CAAC   653

MAC.US.-.239            ..........................................................-----C-T----AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A--C--C-A--CAAC  1124
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_   p17

SMM.SL.92.SL92B         ..........................................................-----C-------GC--C--C...--GTCA---AA-A--GC----GA--TA------G--A-----C----C--C   564
SMM.US.-.H9             ..........................................................-------T----AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A-----C----C--C   606

STM.US.-.STM            ..........................................................-----C------AGC--C--C...--GTCA---AA-A--GC----GA--TA------G--A-----C----C--C   780

SAB.SN.-.SAB1C          ..........................................................-----------TAAC-----C...-----T--TA-GA--C-----GC--TC---TC-G-CA------C----CAAC  1088

TAN.UG.-.TAN1           ..........................................................-----A---G--CAC--G-C-...C-GTCA---A-GA-T--G--CAC--TT--G-----A--C---C-C--GAAC   973

VER.KE.-.AGM155         ..........................................................-----G---GCTA-C----C-...C-G-ATA--A--C--------GA--TT--GC-T--A--AC-TC-C--GAAC  1002
VER.KE.-.9063           ..........................................................-----G---GCTA-C----C-...C---AA---A--C-------C-A--TC--G-----A--A---C-C--GAAC  1013
VER.DE.-.AGM3           ..........................................................-----G---GCTA-C----C-...C---ATA--A--C-------CAA--TT--GC-T--A--AC-TC-C--GACC   502
VER.KE.-.TYO1           ..........................................................-----G---GCTA-C----C-...C---ATA--A--C-------C-A--TT--G-----A--AC-TC-C--GAAC   503

COL.CM.-.CGU1           GCCTGCAAGTGGAGTAAGCTGCGGCGCTCCAGGACCGGTAGAGGGAGGTCATTA.-----CAACGAGCA-GG-C-G...---G-GAA-AA-AC------AGACCTAC-G---G-C---C-G-A-AA----   675

DEN.CD.-.CD1            ..........................................................-----GAGCG-GAGC--G--TCCCA-G-CA--T-CG-CCC-G---A---TT------G-AA-A-----A------   553

GRV.ET.-.GRI_677        ..........................................................-----C-G-G-TCAC----C-...C-GTCA---A--AGCC-C--CACG-TC--G--G---A--C--C-T--GAAC   968

GSN.CM.99.CN166         ..........................................................-----A-----GCAC--G-CGATGC-G-CA--TACTA-G--G---A-G-AT--G---G-G--TC-C--A---AA-   523
GSN.CM.99.CN71          ..........................................................-----A-----GCAC--G-CGATGC-G-CG--TACTA-G--G---A-G-AT--G---G-GA-TC-C--A--GA-A   529

DRL.-.-.FAO             ..........................................................-----A-----C-------G-...C-T--G--A-A-A-GC-T---GCT-TT-----GG--A-----C----CTCC   362

RCM.GA.-.GAB1           ..........................................................--------------C---T-G...--GTCA---AAGA-GC----CGC-------TC-G--A-----C----C--C   283
RCM.NG.-.NG411          ..........................................................-----------G------A-T...--GTCA---AAGA-G-----GCT------C-GG--A-----C----C--C   283

MND-2.-.-.5440          ..........................................................-----C-----C-------G-...C-T--G--A-A-A----G---GA-CT------G---A-----C----CTCC   365
MND-2.CM.98.CM16        ..........................................................-----C-----C-------GT...C-T--G--A-A-A----G--CGAGCT------G---A-----C----CTCC   808
MND-2.GA.-.M14          ..........................................................-----C-----C-------GT...C-T--G--A-AGA----G--GA-CT-----C----A-----C----CTCC   732

MNE.US.-.MNE027         ..........................................................-----C------AAC--C--C...--GTCA---AAGA--GC----GA--TA---------A--C--C-A--C--C   604

LST.CD.88.447           ..........................................................-----AT--G-GAAC--T--C...C----TA--CA-ATTGAG-GAGA--TCTGC-GT--CA--C----A--T-AC    72
LST.CD.88.485           ..........................................................-----AT--G-GAAC--T--C...C----TA--CA-ATTGAG-GAGA--TCTGC-GT--CA--C----A--T--C    72
LST.CD.88.524           ..........................................................-----AT--G-GAAC--C--C...C-C--TA--CA-ATTGAGA-AGAT-TTTGT--TG-CA--C----A--T---    72
LST.KE.-.lho7           CACT......................................................-----AT--G-TAAC--T--C...C-C--TA--CA-ATTGA-A-AGAT-TTTGT-GTG-CA--C----A--T---  1147

SUN.GA.98.L14           GTGTCTCAAAC...............................................-----A---G--AAC--CAC-...G-GGATA---A--TTG-GAGGA-C-TT----G-G-AGCTC-G-AA--T--A  1134

MND-1.GA.-.MNDGB1       ..........................................................------AATG-GAAC--T-CC...--GTTA---ACT--T--G---AA--TT--G-----AA-A----A-AG---T   521

MON.CM.99.L1            ..........................................................-----G-----GCAC--G-CGATGC-T--T--TACTA-------CAAG-AT--G--GG-G---C-T--A--GC-C   516
MON.NG.-.NG1            ..........................................................................................................................................     0

MUS.CM.01.1085          ..........................................................-----A-----GCAC--G-CGATGC-G-CA--TAACA-G--G---A-G-AT--G--GG-GA--C-C--A---AAA   521

DEB.CM.99.CM40          ..........................................................-----G---G-GCG---G-C-CCA----C-----CG-CT-----CA---TT------G-AA-A--G--A---AAA   559
DEB.CM.99.CM5           ..........................................................-----G---G-GCGA--G-CGCCA----CG----C--CT-----CA---TT--G-----AA-A--G--A---AAA   562

SYK.KE.-.KE51           ..........................................................-----A--AG-G-GA--G-CGATCC---CA--TA--A-GC-T---AA--AT--G---------C-TC----GAA-   506
SYK.KE.-.SYK173         ..........................................................-----A--AGCG-G---G-C-ATA----CA--TA----GC----CA---AT--G---------C-CC-A--GAAA   869
```

```
H1B.FR.83.HXB2      GGAAAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCC   991
Gag                 _G__K__K__K__Y__K__L__K__H__I__V__W__A__S__R__E__L__E__R__F__A__V__N__P__G__L__L__E__T__S__E__G__C__R__Q__I__L__G__Q__L__Q__P__S__   p17

H1A1.UG.85.U455     -----C----------G-C-G------T---------------------G---AA----A--C----C--------T---------G-------A---CAG--------------AT--------G-T-   458
H1B.US.90.WEAU160   -----T-----------C---------------------------------------------C------------------------------------------------------AT-----------   990
H1C.ET.86.ETH2220   ---------C-C----TGC-G-----CC----C-------A---------G---AA---T---C----C----A---T-----T-------C--------A-------A-TAA------------G-T-   371
H1D.CD.84.84ZR085   -----------G-----GC-----------------------A-----G--------T---C----------T----------------------A-------A---A------------G-T-   508
H1F1.BE.93.VI850    ----------G-------A-G------C--A----------------------T---C--G---------TC-------------------CA-A-----A-AA-----------------   346
H1G.SE.93.SE6165    ----G----TC-------A--------C---------------------G--GA----T---C----C----A---T---------G-------T---CA-------A--A-----T-G--------T-   394
H1H.CF.90.056       ----------------GGC--------C---------------------G--A-----T---C----C--C-----T---------G-------A---CT---G---A-A-A----A----G---G-TA   338
H1J.SE.93.SE7887    --G-------------GGA----G---C---------------------G--CA----T---C----C--------TC----GT--G--A--------CA---------A-T------C------G-T-   314
H1K.CM.96.MP535     ----------------C-G------C-----------------------------T---C-C--C--------T-----G---A-----------C-G------A-AAC---AA----G------A   202
H101_AE.TH.90.CM240 ----GA----------GGC-G------T----------------A---T-----A--------C----C---A--T-T---------G-------A---CA-------A-A-A----T----GT--A-T-   556
H102_AG.NG.-.IBNG   ----------------G-C--------C---------------------G---A--------C----C--------T---------G----G--A---CA----C--A---A----T-----T--G-T-   531
H1N.CM.95.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-   549
H1O.BE.87.ANT70     TCT--A--G-------G-C-------CT---------------A-----G---A--------TG--------AG--AC-------TG---G--TAA-GAGA-GC-G--ACAG---T--G-G---G-T-  1043
H1O.CM.91.MVP5180   TCT--A--GGC-----GGC-------T----------------------G---A---A----TG---------T--A--------TG-------TAC-GAG---C-G--ACAG---T--G-G---G-T-  1018

CPZ.CD.-.ANT        --C--------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   355
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-   381
CPZ.GA.88.GAB2      ----G-----G---C-TGA-G-------------------------G---A-------TTG-G-C-----T--AA-G------AA---------TAT--GT--T--AA-------G-----G-T-   363
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-  1028
CPZ.US.85.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-  1033
CPZ.TZ.-.TAN1       --C-G-----------TT-A--------C---------CG-A--C--A---C-G--T-----GA-G-----C--T--CA-G--G-ACGT---------CT-GA----CA-CCTC-----G--G--T--GG   554

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  1297
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  1304
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   749
H2B.GH.86.D205      -----------G----TGC---------G----------GTG-AC--AT----CA----T-G-T-AGCAGAGA-------G---T--AA------A---CA-A----T--TAA-GTCT--GCT----TAG  1280
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  1281
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   667
H2U.FR.96.12034     ---------------C-TG---------G-T--G------GCG-AC--AT-G--CA-----T-G-T-AGCAGAAA-------G--GG-CAA------A---CA-A----TT-ATCGGT-T--GCT----TAG   783

MAC.US.-.239        -----------G--C-TG--G--G---G-----------GCA-AT--AT----TA----T-G-T-AGCAGAAA-------G--G-ACAA------A---CA-A-------TTCGGTCT--GCT----TAG  1254
Gag                 _G__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__   p17

SMM.SL.92.SL92B     ----G------G--C-TGC-T--G-----TA--------GC---A--AT----CA----T-G-T-TCGGCAGAAA-------G----GCAA------A---CA-AG----T-A-C-GTA---GCT----TAA   694
SMM.US.-.H9         -----------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   736

STM.US.-.STM        --G--------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   910

SAB.SN.-.SAB1C      --------G--G-----G----G----C---------------A---A-----T-----TAGCC--TCAG-AAAT--C-----G---AAG-----GGTCGTCA----T--TA-TGTA--C-TT--CCTAG  1218

TAN.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--GCTAG  1103

VER.KE.-.AGM155     -----------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--CCTAG  1132
VER.KE.-.9063       ---------------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-TAG  1143
VER.DE.-.AGM3       -----------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--GCTAG   632
VER.KE.-.TYO1       -----------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--CCTAG   633

COL.CM.-.CGU1       AAG--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   805

DEN.CD.-.CD1        A-T-----G--G--C-TGA-T--G--CC-GA-T---T---AG-A----A-G--G-----T-GCT-AGGAGAACAG---C-------C-AT-----A--CCA-ACT--CA-A-AGACCT--GCT----TAG   683

GRV.ET.-.GRI_677    --G--A--G--G--CC--A-T------T--A--------G-A-AA--AA-G--------T-GGT-AC--GAGAAA--T---------AA-----------CA-A----CA-A-A-GTTT--ACC--G-TGG  1098

GSN.CM.99.CN166     -----------G--C-TGA-C--G--CT----------CTC--AA-----T--GA---AT--GC-GTCCGAC-CT--TC-T-----ATCA---G--G---C---GG---T-A-A-ATTG-CTTC--CCTAG   653
GSN.CM.99.CN71      --------G--G--CGCGA-C--G--CT----------CTC--A------T--GA---A---GC-ATCAGGCA-T--TC-T---T-TAA------G---CA---GG---T---A-GTCACCTCC----TAG   659

DRL.-.-.FAO         -----A--G--------GC-C--G---C-T--C----T-GCG-A---A-----T-------GCC-GC--GAGA-AT-A-------GTCAG-----G--CCA-A----TT-A---GTA--CTTT--TCTAG   492

RCM.GA.-.GAB1       --G--A--G-----C-TGC-G--G---T-G---------T---AAA-A---A-TAA---T-GCT-G-G-GA-CATT-----------G--AC---A---GA-A-----T-A---GTC--G-TG--TCTAG   413
RCM.NG.-.NG411      -----------G----TG-----G---C-----------T-------A-----GA----T-GCT-G-G-GA-ACTT---------GTG-------AATCCAGA----TA-A--GGTTA-T-TG--TCTAG   413

MND-2.-.-.5440      -----------G--CC-G---------G-TA------T-----A---A-----CA----T-GCC-AC--GAGAAGT---------GTCAG-----A--CGA-A----T--TA-CGTA--CTTT---CTAG   495
MND-2.CM.98.CM16    -----A--G-----CC-GC-G------C---A------T-----A---A-----CA----T-GCC--C--GAAAAG----------GTAAG-----A--CGAGA----T--TA-CGTA--CTTT---CTAG   938
MND-2.GA.-.M14      --------G-----CC-G-------------A------T-----A---A-----T-----T-GCC--C--GAAAAA--A--------TCAG-----A--CGAGA----T--TA-CGTA--CTTT---CTAG   862

MNE.US.-.MNE027     --G--------G--C-TG--G--G---G-----------GCA-AT--AT----TA----T-G-T-AGCAGACA-------G--G-ACAA------A---CA-A-------TTCGGCCT--GCT----TAG   734

LST.CD.88.447       A-T--A--G-CT---C-GAAG-G----T----G--------CA--A--AT----TA-G--T-G-T-AGGAG-ACA-T--C----G---G----T-----C-AGA----T--T---GTCTGTTGG---CTAT   202
LST.CD.88.485       A-T--A--G-CG---C-GAA--G------T--G--------CA-AA--AT----TA-G--T-G-TT-AGGAG-ACA-T--C-G--G---G----T-----C-AGAG---T--T---GTCTGTTGG---CTAT   202
LST.CD.88.524       A-T--A--G-CT---C-GAA--G----G---AC--------CG-AA--A-----TA-G--T-GGC-AGGAG-AAA---TC--------G-----G-------AGA-G--TT------GTC-GTTGG--T-TAT   202
LST.KE.-.lho7       A-T--A--G-CT---C--AA--G----G---A---------CG-AA--AT----CA----T-G-T-GGGAT-ACAA--C--G------G-------G-----AGA----C---ATCTGTCTGCTGG--T-TAT  1277

SUN.GA.98.L14       --G--A--G-CC---CC-GGC--G---G------------G-T-AA--AT----TA-G--T-G-C-----CAAG-AA------AG----GTG----------CAGA-G---T----TC-GTCTGTTGG--T-TGT  1264

MND-1.GA.-.MNDGB1   --T--A---TGT----G---G-----CC-CTGT---TGT-AAG-T--AT----TA----T-GCT-ATCGGA-AAA--CC-T------CAGC----A---GA-A-----C--CTC-GTATGTTGG----TAT   651

MON.CM.99.L1        --T-----G--G--CCTGA-T--G---------------TGC--A------T--TA------GTC-CTCGGACTCG--CC-G--G---CAG--C-------AGA-G--TT---A-GTTA-T-TT--TCTA-   646
MON.NG.-.NG1        .......................................................................................................................................     0

MUS.CM.01.1085      --------G-----CCTGA-C--G--C--TA-------TTC---------T--TA----T-G-C-CTC-GA-TC---T--G---T--AGG--T--G---CA-A-G--CA-A-AGGTCA-C-TC--C-TAG   651

DEB.CM.99.CM40      --C--A--G--G----TGA----G--CC-C--C------G-T-A---A-----CA----T-GCC--GCAG-A-CAT-A-------ATAAG---------CAGG----TT--ACTAG-T----G--T-TAG   689
DEB.CM.99.CM5       --C--------G----TTGA----G--CC-C--C--------T-----A-----CA----T-GCC--GCAGGA-CAT-A--G------CAA------A---CA-G-G--CT-AACTAGAT-------T-TAG   692

SYK.KE.-.KE51       --G-----GCGC--CCT-G-TCG----T----T------G-A-A------G--CA------GCC-CTCAGACCAG--TC-T----GTAA------T---GA------CA-AAA-ACCA-CTT-----TAG   636
SYK.KE.-.SYK173     --G-----GCGC--CCT-G-TCG----T----T-------AG-A----T-G--CA------GCC--TCGGACCAA--CA-G--GT-TAA----------GA-A----T--CAC-GTC----TT----TAG   999
```

PLV
complete genomes

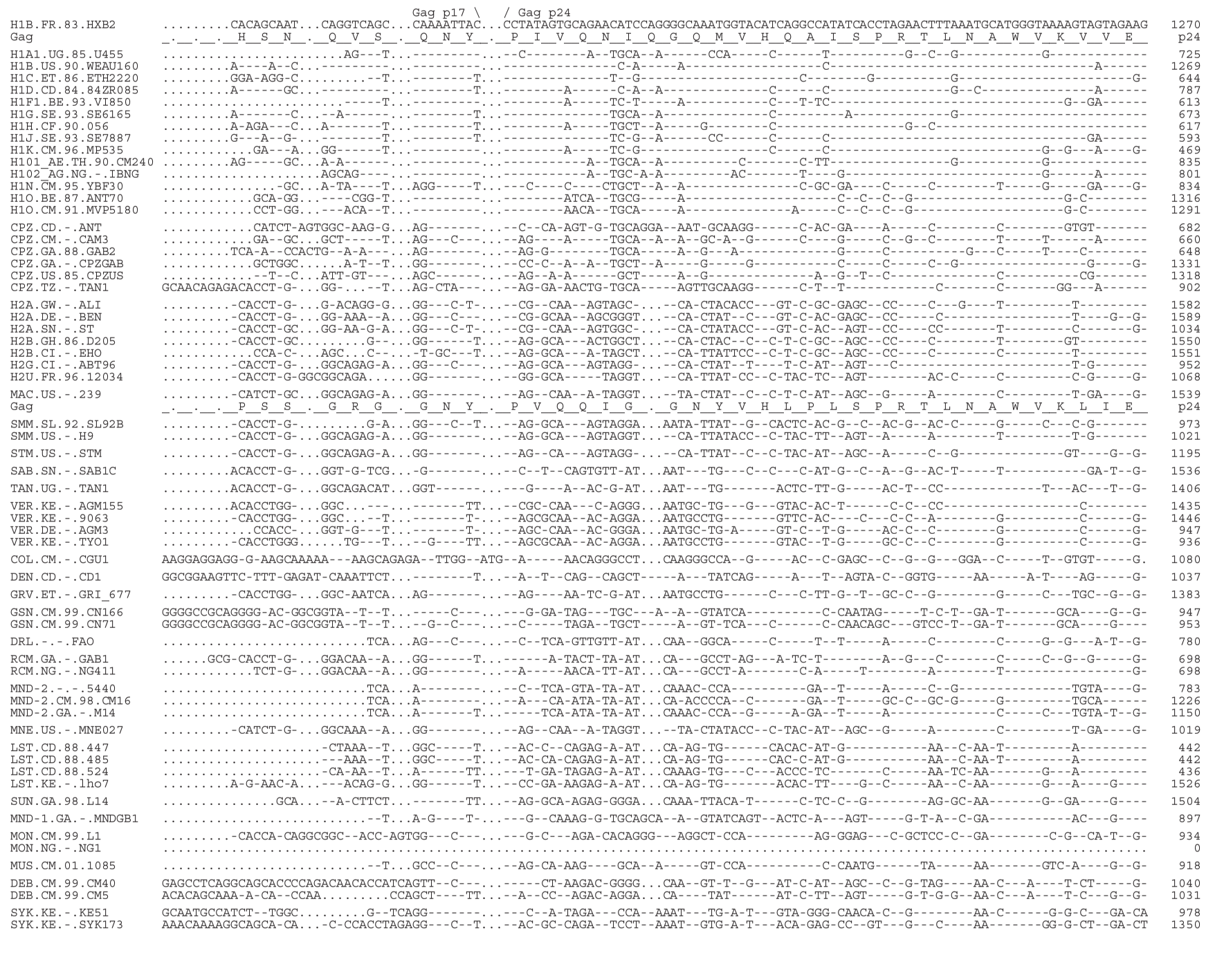
PLV
complete genomes

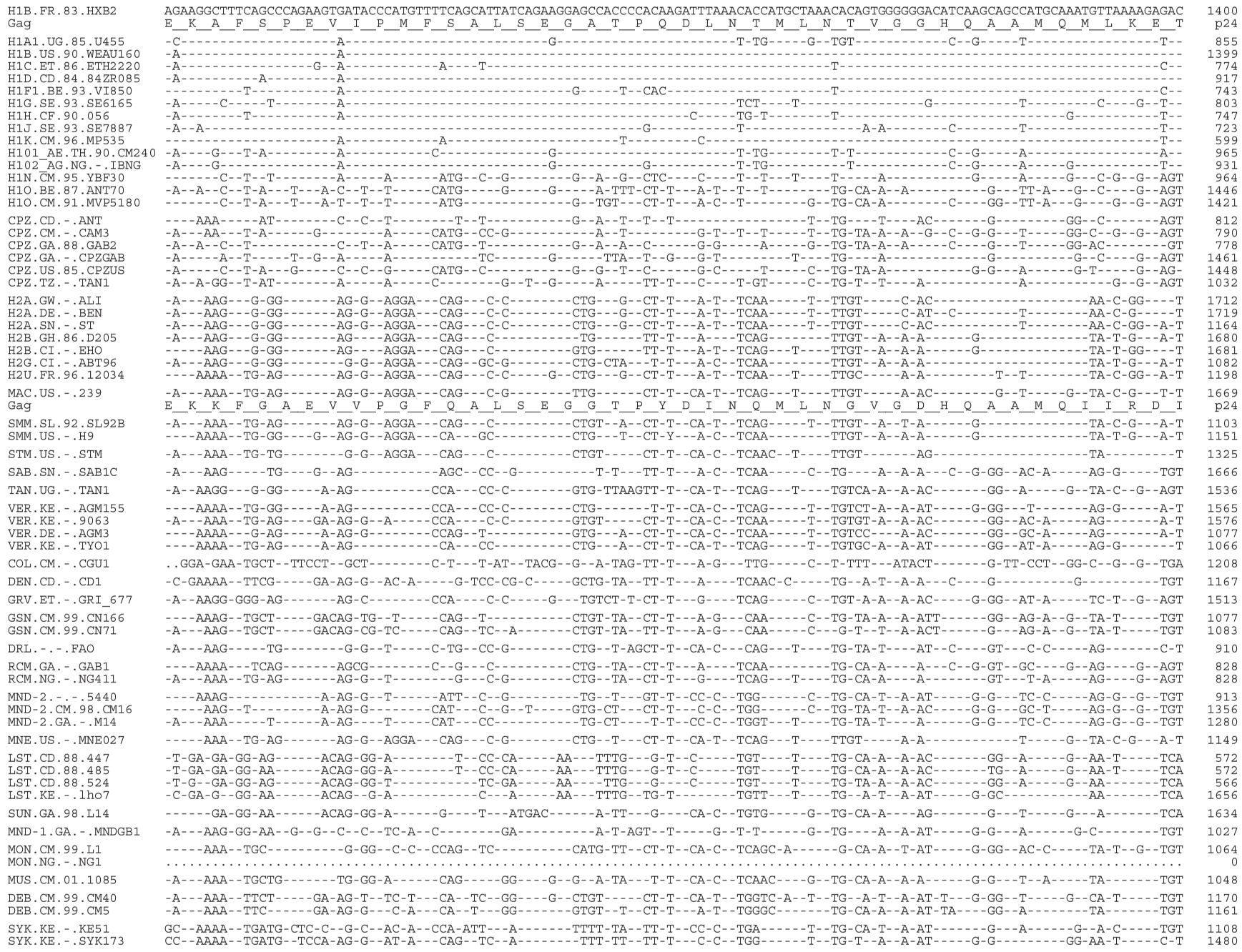




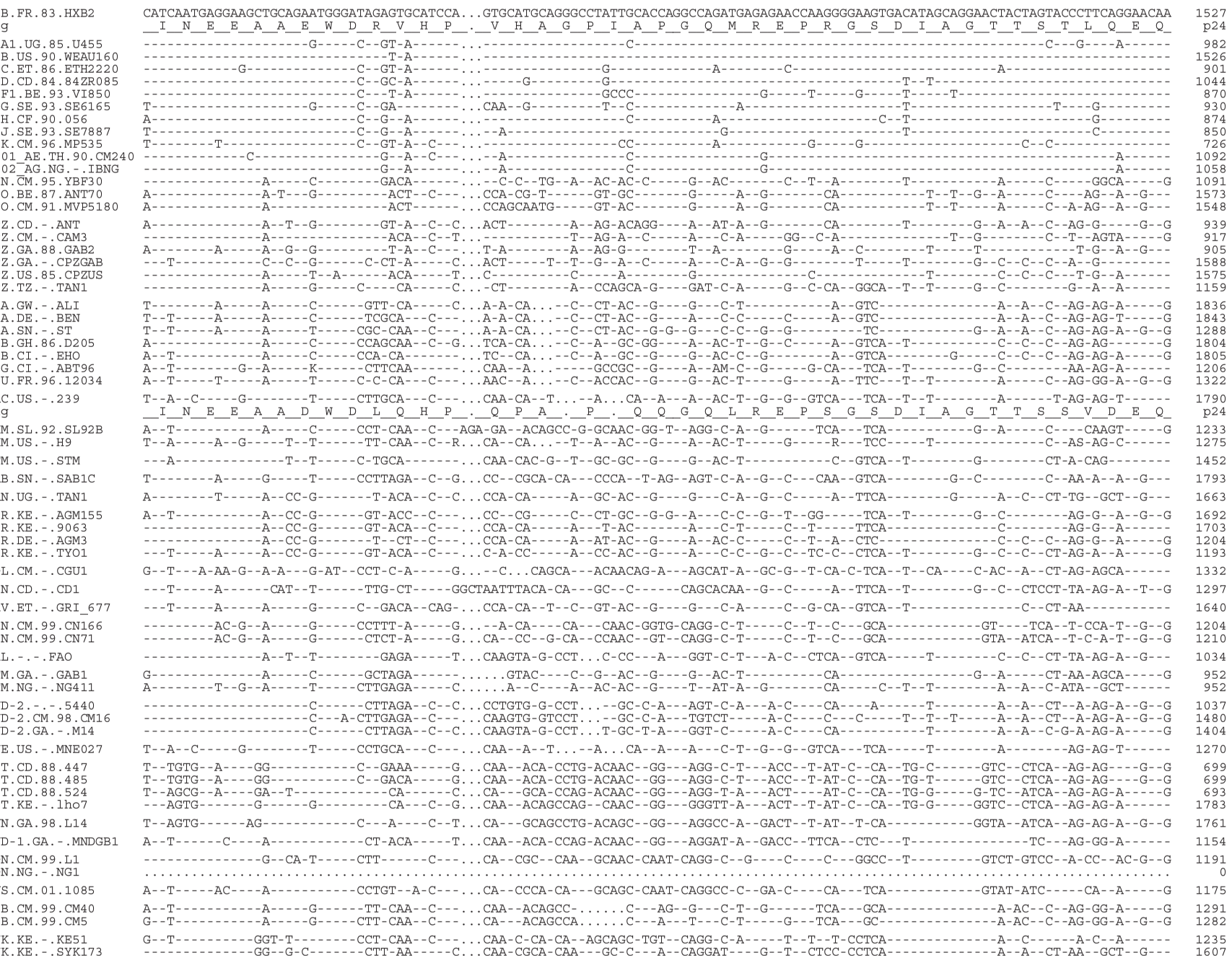
PLV
complete genomes

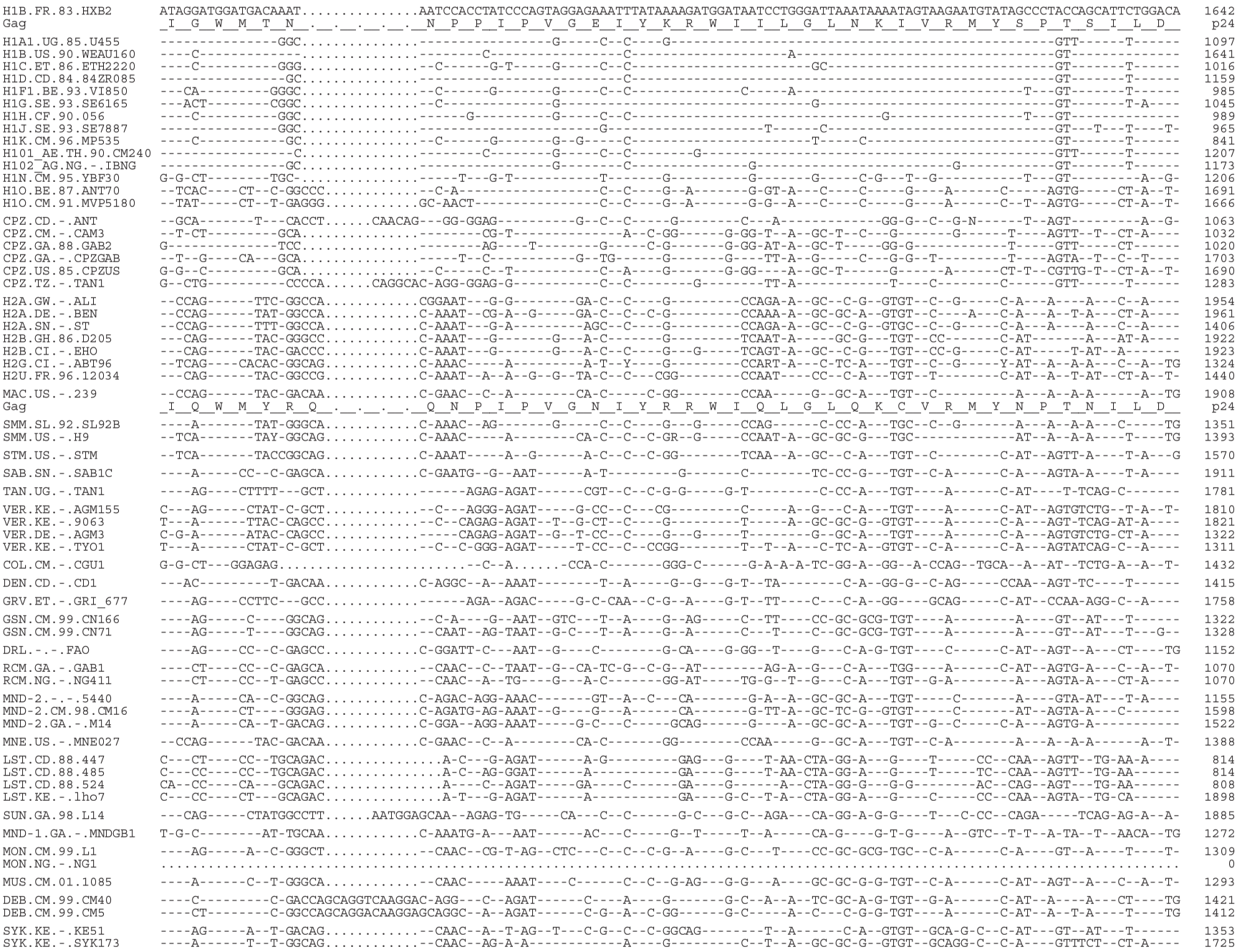
PLV
complete genomes



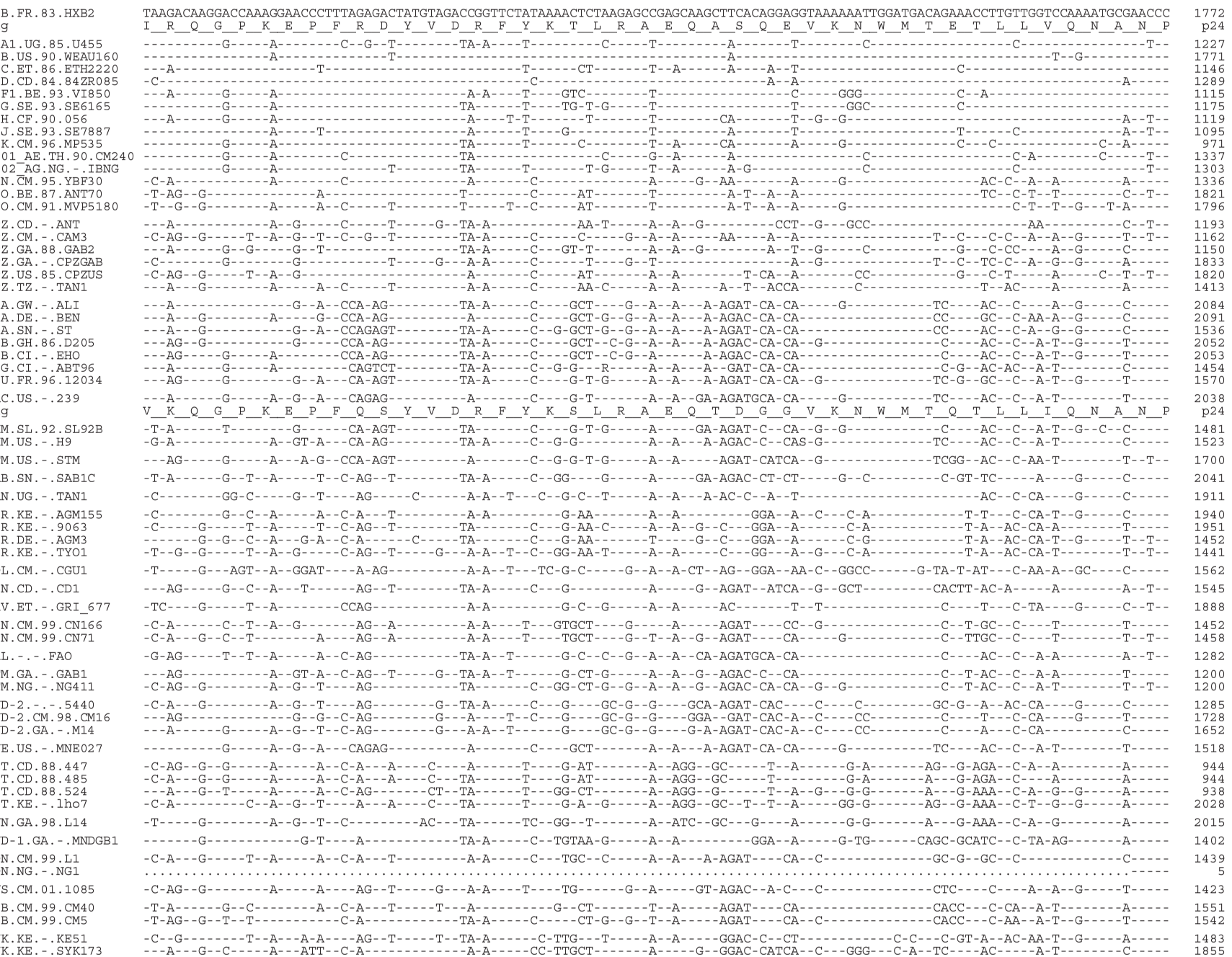


PLV
complete genomes

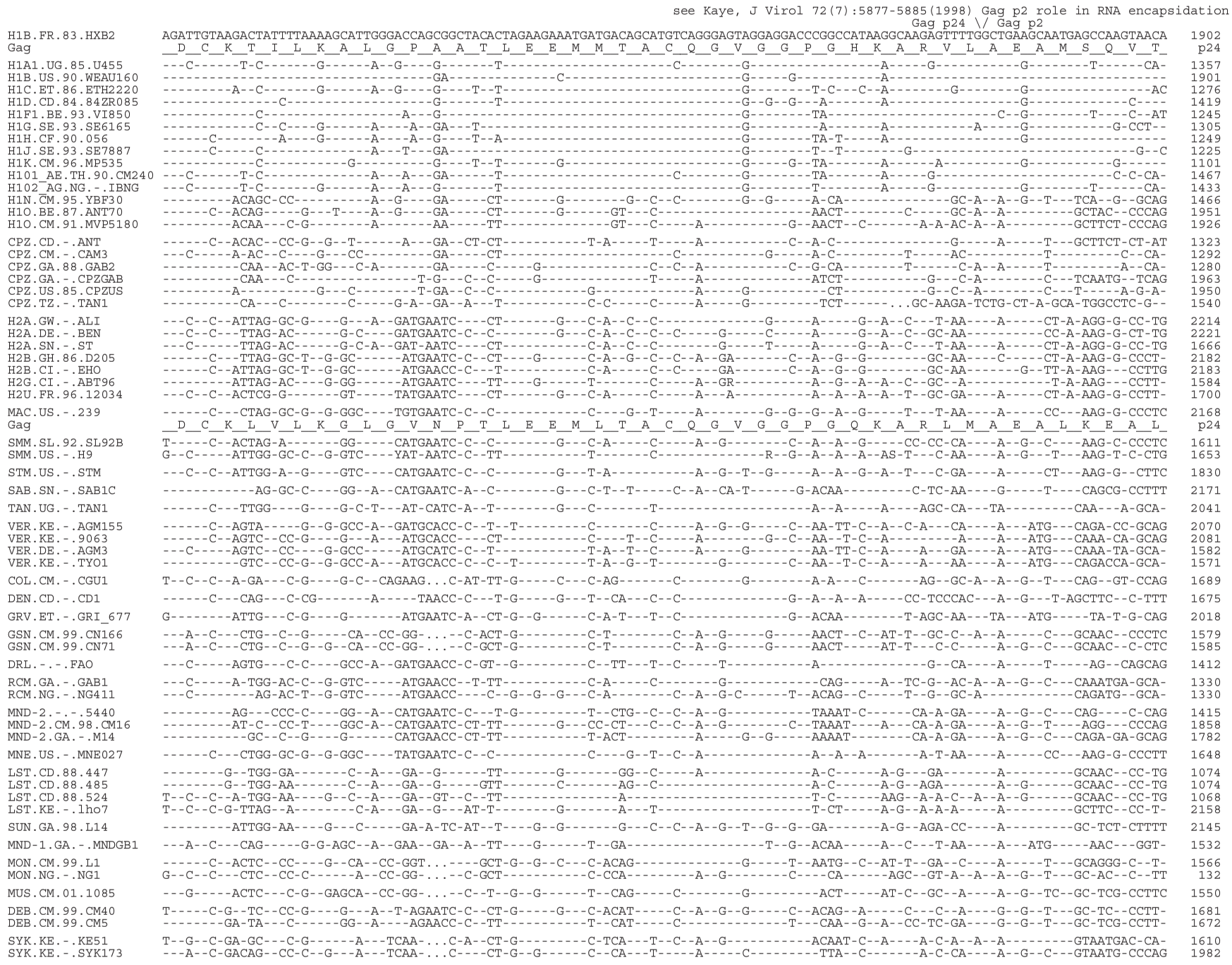

H1B.FR.83.HXB2
Gag
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Gag
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173



PLV
complete genomes


H1B.FR.83.HXB2 TAAGACAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCGAACCC 1772
Gag I R 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 p24
H1A1.UG.85.U455 1227
H1B.US.90.WEAU160 1771
H1C.ET.86.ETH2220 1146
H1D.CD.84.84ZR085 1289
H1F1.BE.93.VI850 1115
H1G.SE.93.SE6165 1175
H1H.CF.90.056 1119
H1J.SE.93.SE7887 1095
H1K.CM.96.MP535 971
H101_AE.TH.90.CM240 1337
H102_AG.NG.-.IBNG 1303
H1N.CM.95.YBF30 1336
H1O.BE.87.ANT70 1821
H1O.CM.91.MVP5180 1796
CPZ.CD.-.ANT 1193
CPZ.CM.-.CAM3 1162
CPZ.GA.88.GAB2 1150
CPZ.GA.-.CPZGAB 1833
CPZ.US.85.CPZUS 1820
CPZ.TZ.-.TAN1 1413
H2A.GW.-.ALI 2084
H2A.DE.-.BEN 2091
H2A.SN.-.ST 1536
H2B.GH.86.D205 2052
H2B.CI.-.EHO 2053
H2G.CI.-.ABT96 1454
H2U.FR.96.12034 1570
MAC.US.-.239 2038
Gag V K Q G P K E P F Q S Y V D R F Y K S L R A E Q T D G G V K N W M T Q T L L I Q N A N P p24
SMM.SL.92.SL92B 1481
SMM.US.-.H9 1523
STM.US.-.STM 1700
SAB.SN.-.SAB1C 2041
TAN.UG.-.TAN1 1911
VER.KE.-.AGM155 1940
VER.KE.-.9063 1951
VER.DE.-.AGM3 1452
VER.KE.-.TYO1 1441
COL.CM.-.CGU1 1562
DEN.CD.-.CD1 1545
GRV.ET.-.GRI_677 1888
GSN.CM.99.CN166 1452
GSN.CM.99.CN71 1458
DRL.-.-.FAO 1282
RCM.GA.-.GAB1 1200
RCM.NG.-.NG411 1200
MND-2.-.-.5440 1285
MND-2.CM.98.CM16 1728
MND-2.GA.-.M14 1652
MNE.US.-.MNE027 1518
LST.CD.88.447 944
LST.CD.88.485 944
LST.CD.88.524 938
LST.KE.-.lho7 2028
SUN.GA.98.L14 2015
MND-1.GA.-.MNDGB1 1402
MON.CM.99.L1 1439
MON.NG.-.NG1 5
MUS.CM.01.1085 1423
DEB.CM.99.CM40 1551
DEB.CM.99.CM5 1542
SYK.KE.-.KE51 1483
SYK.KE.-.SYK173 1855

see Kaye, J Virol 72(7):5877-5885(1998) Gag p2 role in RNA encapsidation



PLV
complete genomes

PLV complete genomes

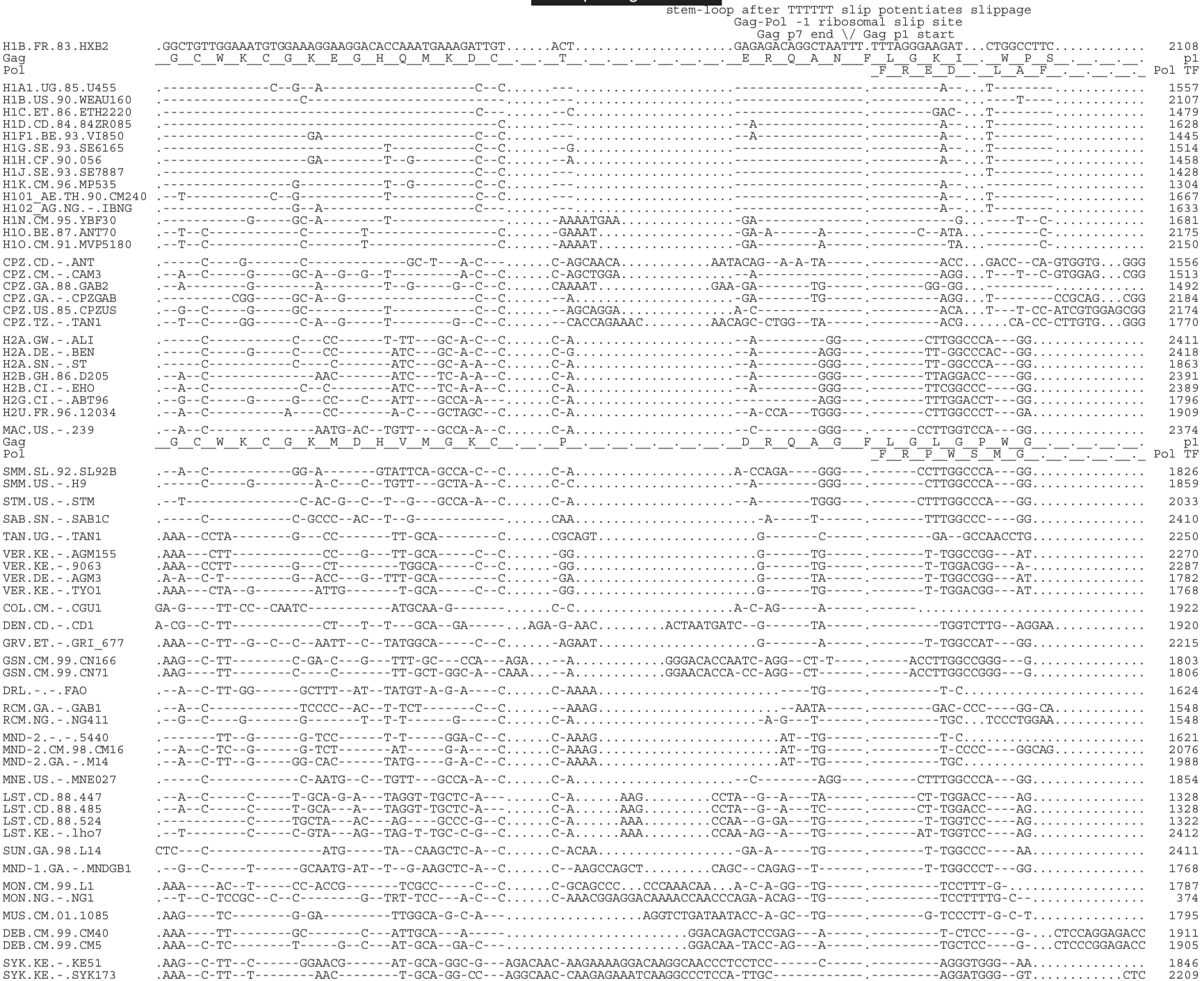
see Kaye, J Virol 72(7):5877-5885(1998) Gag p2 role in RNA encapsidation
/ Gag-Pol fusion TF protein start
Gag p2 \/ Gag p7 NC

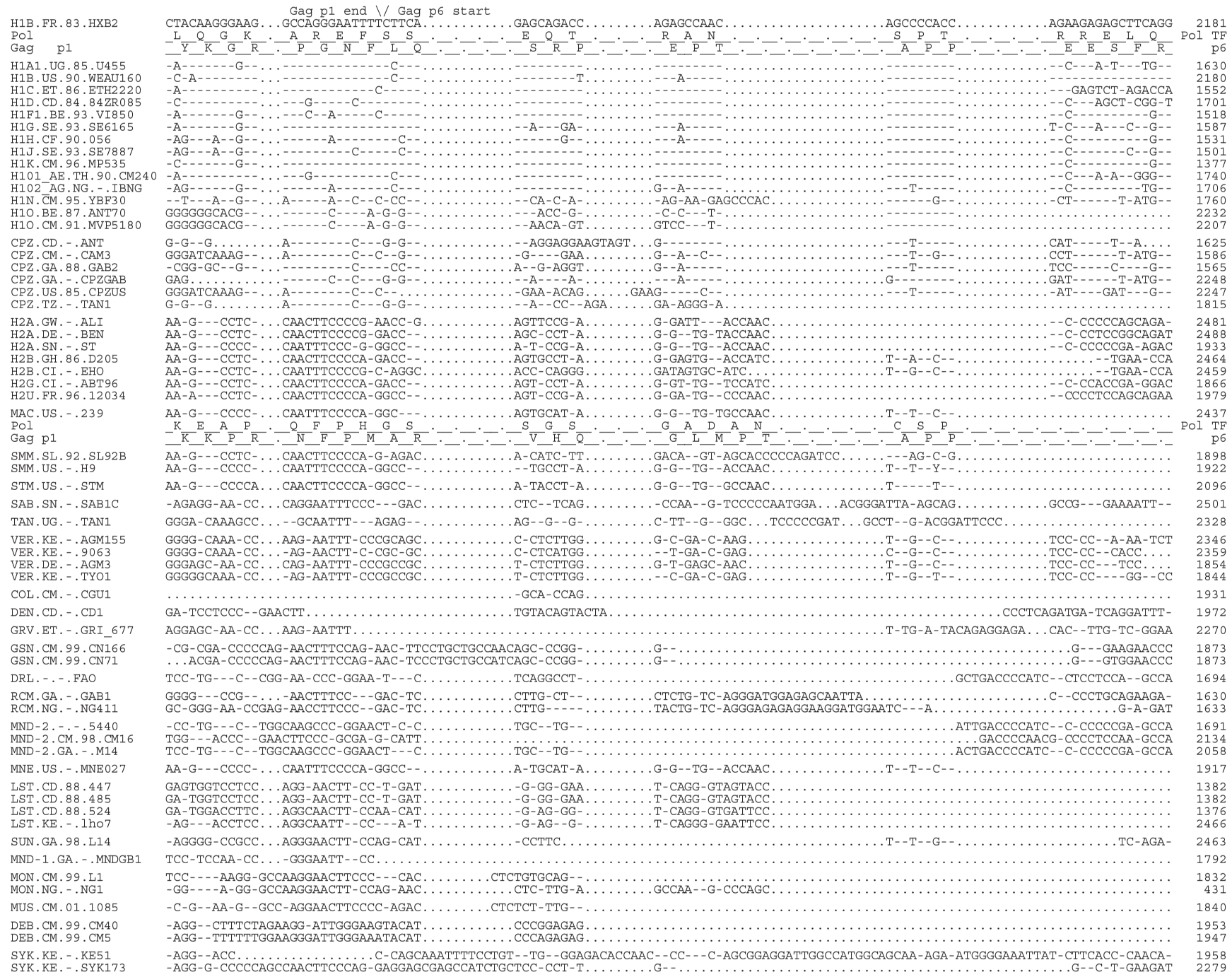
```
H1B.FR.83.HXB2      AATTCA......GCTACC..............ATAATGATGCAGAGAGGCAATTTT...AGGAACCAA.....................................AGAAAG.........A  1954
Gag                 _N__S__.__.__A__T__.__.__.__.__.__I__M__M__Q__R__G__N__F__._R__N__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__K__.__.__.__  p7

H1A1.UG.85.U455     C-GA--.......AGC.....................------------------------.----GG--CG.....................................-GA..........-  1403
H1B.US.90.WEAU160   ---A--.......------..................--G----------------------.-----G--C-.....................................-----A.........-  1953
H1C.ET.86.ETH2220   ---A--.......A-C.....................-------------A-A----------.-A-GG--CT.....................................-A--GA.........G  1325
H1D.CD.84.84ZR085   -G-G-C......T--G-TGCC.....................-----------A-A-A---------.-A-GG-AC-.....................................-----A.........-  1474
H1F1.BE.93.VI850    TCAG-C.....................................-------------A-A-T---------.-A-GG----.....................................----GA.........G  1291
H1G.SE.93.SE6165    GG-G--.......--AG-AGCC.....................-----------------A---------.-A-GG--C-.....................................----GA.........-  1360
H1H.CF.90.056       ---A--.......AA---AGCC.....................-------------A------C------.-A-GG----.....................................-----A.........T  1304
H1J.SE.93.SE7887    ---A-C......AAC.....................----------A-----T--C---.---G----T.....................................-A--GA.........-  1274
H1K.CM.96.MP535     ---C--.......-T-.....................G------------A------C---.-A-GG---T.....................................-----A.........-  1150
H101_AE.TH.90.CM240 C--G--.......A--.....................------------------------C...-A-GG---G.....................................--A.........-  1513
H102_AG.NG.-.IBNG   C-GG-C......AAC.....................G-------------------------.---GG---G.....................................-GA..........-  1479
H1N.CM.95.YBF30     C-GC--.......A-A--TAGT.............G-CT-TGCA--A--G--A--C------.-AAGG-AT-.....................................--G--A.........C  1521
H1O.BE.87.ANT70     C-AGAT......TTG-AAGGAGGATACACAGCAG--T-C-----A-----GC-AAA-...CCA-TTAGG.....................................-A-GGA.........-  2018
H1O.CM.91.MVP5180   C-AGAT......TTA-AAGGAGGATACACAGCAG--T-C-----A-----GC-GAA-...CCA--TAG-.....................................-A-GG-.........C  1993

CPZ.CD.-.ANT        ---G--CAGGGAA-C.............GCAG-CT-TC--------------GGA...-AT-GAGG-GGA.....................................-A--GA.........C  1384
CPZ.CM.-.CAM3       --CC-T......A-A.............GGTG-GT-CC----A--G--A---AA-...G----A-C-.....................................-CT-G-.........-  1344
CPZ.GA.88.GAB2      C-GA--.....................AGCTCA---T-C-----A----CAGGGGGA...--A-CT-CT.....................................CCT-G-.........-  1332
CPZ.GA.-.CPZGAB     ---CA-GGGAGA--A.............GATG-TT-CT-C----AG--AC-AGG-...GC-GG--C-.....................................-A--GA.........-  2021
CPZ.US.85.CPZUS     --CC-C......T-A.............AGCG--T-TC-A----A---A--CGC-...G----G-C-.....................................G-G-GA.........-  2002
CPZ.TZ.-.TAN1       -CAG--GGGGGA-TA.................-AT---C-----G----A--AAGA...CCACC-TT-.....................................-A----------GGTC  1598

H2A.GW.-.ALI        -CAC--.......---C-TATC......CCATTTGCGGCAGCC--ACA-.....................................................----G-.........-  2257
H2A.DE.-.BEN        GGAC--......AGCC-TATC......CCATTTGC-GCAGCC--ACA-.....................................................------.........G  2264
H2A.SN.-.ST         GCAC--.......--CC--ATC......CCATTTGC-GCAGCC--ACAG....................................................----G-.........-  1709
H2B.GH.86.D205      -CAC-T......--AC--ATA......CCGTTTGCTGCCG-T--ACA-AAAGCAGGG...-A-.....................................---GG-.........-  2237
H2B.CI.-.EHO        -CAC-T......T-C---AAT......CCGTTTGCCGCCGCT--ACC-A-AGCAGGG.....................................-AG-G-.........-  2235
H2G.CI.-.ABT96      --CC-T......A-AG--CTA......CCATTTGC-GCTGCT--ACA-AAA-CAGGA...G-A.....................................-A--G-......AGC-  1642
H2U.FR.96.12034     C-AC-T......ATGC-TATA......CCCTTTGC-GCAG-C--ACAGA-AGGCGGCCCA.....................................--G-G-.........C  1755

MAC.US.-.239        GCAC--.......-TGC-AATC......CCTTTTGC-GCAGCC--ACAGA-GGGACCA.....................................------.........C  2220
Gag                 _G__P__.__.__V__P__I__._.__P__F__A__A__A__Q__Q__R__G__P__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__R__K__.__.__.__  p8

SMM.SL.92.SL92B     -CGGGGTCTTTA-TGG-AGCACAGTTTAGGGGAGC-GC--AAGGCCA---G.....................................................-AT---CCT.....-  1672
SMM.US.-.H9         -CGCA-.......-GGC-ACTC......CCATTTGCRGCAG-C--ACAGAAAGGACAA.....................................................R----------.-  1705

STM.US.-.STM        C-AC--.......-ACC-ACTG......CCCTTCGC-GCAGCC--ACAGCAGGGA.....................................................C---G-.........-  1879

SAB.SN.-.SAB1C      C-GCA-......CAA--AGTT......GGGAAC--CT-TG-----CA---GGCAAGA...CCC-GGGGTCCCTTAGGGGGAAGAGGTAGACCTCTG.....................-ACCCA.........-  2259

TAN.UG.-.TAN1       GGAGTT.........................-AC---G----AG---CACCGCGA...G-AGGTAGGGGCAGAGGC.....................................---GGC.........C  2096

VER.KE.-.AGM155     -G-CAG.........................-AC---G-A---CAG--AGG-GGA...---GGAAG-CCA.....................................---CCC.........C  2119
VER.KE.-.9063       -C-CA-.........................-AC---G-----CA---GGCAGCA...G-AGGAGT-AGGAGACAG.....................................---CCC.........C  2136
VER.DE.-.AGM3       -GCCAG.........................-AC--------ACAG---GG-CAG...--AGGAAG-CCA.....................................---CCC.........C  1631
VER.KE.-.TYO1       ---CA-.........................-AC---G-----CAG--AGG-CCA...-AA-GA---.....................................---CCC.........C  1617

COL.CM.-.CGU1       C-AGAG.....................AGAACC-AC-----TG-AGT-AAG-CAGCA.....................................................................  1725

DEN.CD.-.CD1        C-GAAG...CAA-GGTTATGCCATGGTTCAAGG-A-GG-GCC--ACCCCCAGTAGAA.....................................................-GA......AGGG  1736

GRV.ET.-.GRI_677    ..............................AT---G-C--AGTG--ACC-CAG...-AA--GGGCCCC.....................................C--GG-.........C  2061

GSN.CM.99.CN166     --AGGC......A---GTTCA.........TAC-AC---G-----.....................--AGGCCCCCCTGGAAGG.....................CAAG----A.........-  1637
GSN.CM.99.CN71      --AGGC.........-G-TCA.........TAC-AT---G----A.....................-GAGGCCCCCCAGGAAGG.....................CAAG----A.........-  1640

DRL.-.-.FAO         GCAG-T......ATA-TGATGCAGCAGAAT......CCCCCCAGAG-TCCACCGCGGGGACCC-GGGG-CCCCCTCCA.....................................--A.........-  1482

RCM.GA.-.GAB1       -G-AAT.........................--C---GCT---CA-TCAGCAAAC...---GGG-CTCCAAGAAGATCAGGAGGA.....................AATCC---T.........  1396
RCM.NG.-.NG411      -G-AAT.........................--C---GCC--ACAG--AGGACC-AGA-AAGGA-CCCCTAAACTAGGGGGAGGA.....................CCT-G-.........T  1397

MND-2.-.-.5440      GCAG-T......-TA-TGATGCAGAACTCGGGGGG-CC-CC--G-G-TCC-CCGAGACAACCCCCT.....................................-GA.........-  1479
MND-2.CM.98.CM16    TCAG-T......-TA-TG.........ATGCAG-ATTC-GGAGG-CC-CCGCGGGGCCCCCC--GA---CCTCCT.....................................-GA.........-  1922
MND-2.GA.-.M14      TCAGA-......-TC-TGATGCAGAACTCGGGGGG-CCACC--G-G-TCCTCCGAGACAACCCCCT.....................................-GA.........-  1846

MNE.US.-.MNE027     GCAC--.......-GGC-ACTC......CCTTTTGC-GCAGCC--ACAGAAGGGACCA.....................................------.........C  1700

LST.CD.88.447       C-AAGT......CAA-TGAGA.........CAG-AT---G-A--A...-TG-CACCA...CCA-GGA-TGCACAGGGAAGGTTTGTAAGAACAGGAGGAGGAGGC......CCT-------.....C  1168
LST.CD.88.485       C-AAGT......CAA-TGAGA.........CAG-AT---G-A--A...-TG-CACCA...CCA-GGA-TGAGCAGGGAAGGTTTGTAAGAACAGGAGGAGGAGGC......CCT-------.....C  1168
LST.CD.88.524       C-GGGG......CAG-TGAAA.........CATC-C---G-----...CAA-CACCA...CCT-GAA-TGCACAAGGGAGATTTGTAAGAACGGGGGGAGGTGGC......CCA----GA.....C  1162
LST.KE.-.lho7       C-GCA-......CAAGTAAGA.........CAA-AC---G-----...-T--C-CCG...CTT-GAA-TGCGCAAGGCAGGTTTGTGAGAACAGGAGGAGGAGGT......CCC--G--A.....C  2252

SUN.GA.98.L14       -GGCAG......CAAGTAGGA......AGACAAGCT---G-----CA-AA-TTACCA...CCC-GAA-CAGCCAGGGGAGGTTTGTAAGAATAGGAGGAGGAGGA......CCT--G-------.....C  2245

MND-1.GA.-.MNDGB1   GGACA-......T-A.............CAA-ATT-TG------...CAG-GAGGGCCTCAA-GAGG-CCAGTTAGACAACCTACTGGA.....................-G-.........-  1602

MON.CM.99.L1        GCC.........AAC-TGCCC.........ATG-AT---G-A-----.-C--GAGGACCA.....................................CCCCA--G-AAGGGA...C  1621
MON.NG.-.NG1        ---AGC......AAC-TGCCA.........ATG-AC---G-A--A.....................................GGCAGAGGAGGTRGTCAGCCAC-CCGACAAGGCRWGC  199

MUS.CM.01.1085      --ACAG......--AGG-ACA.........CTG-AC---G----A.....................GGAGCCAAGGGAGCCAGGGGAAGGGGCAACGGCAGCCAGGGAAACCGAGGA-  1632

DEB.CM.99.CM40      --AGAG..............GCAGGGTCCCTAGGC---G-A--ACA-A-G-GAGGA.....................................CCCCCTGGTTCT----GA.....C  1742
DEB.CM.99.CM5       --AGA-..............GCAGGCTCCCTGGGG---G-----CA-A-AGGAGGA.....................................CAAGGAGGAGGTCCT----GA.....C  1736

SYK.KE.-.KE51       C-CAGC.....................TTGGGC-----A--A...--GCCAAGA.....................................CAAGGAAGCAATCCC--G-GA......GGGC  1665
SYK.KE.-.SYK173     C-G---.......-TA.................-AT---G-A--AG-GCCATCAAAA...G--.....................................---TCA.........-  2025
```

```
H1B.FR.83.HXB2      TTGTTAAGTGTTTCAATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGC...AGGGCCCCTAGGAAAAAG.....  2022
Gag                 I__V__K__C__F__N__C__G__K__E__G__H__T__A__R__N__C__._R__A__P__R__K__K__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  p7
H1A1.UG.85.U455     GAA-----------C-------------A---CT-----A------T...-----------------.....  1471
H1B.US.90.WEAU160   C-A-----------------------------A----T----------...-----------------.....  2021
H1C.ET.86.ETH2220   CAA----A--------C--------G---------CT-----------...----------------A.....  1393
H1D.CD.84.84ZR085   ---------------C-----------------T---A-------T...-----------------.....  1542
H1F1.BE.93.VI850    -------A-----T-----------------A----T----------...----------A------.....  1359
H1G.SE.93.SE6165    C-A-C--A-----------------G-----A--TCT----------...-----------------.....  1428
H1H.CF.90.056       -------A--C-----C-----------G--A----T----------...-----------------.....  1372
H1J.SE.93.SE7887    -------------------------C----A----T---A-A---C---...----------A------.....  1342
H1K.CM.96.MP535     ----------C-----C-----------------T----------...----------A------.....  1218
H101_AE.TH.90.CM240 GAA-------C-----C-----T-G------A---CT----------...----------A---C--.....  1581
H102_AG.NG.-.IBNG   CAA-A----------C-------------A---CT----------...-A----------G-.....  1547
H1N.CM.95.YBF30     CCA----A--------------------G--C--TTTG--A-----C--T...-A--------A-G-GGA.....  1589
H1O.BE.87.ANT70     C-A-A--A--------C-----A-----G--A--T-T---A--------T...C-A--A------------.....  2086
H1O.CM.91.MVP5180   CCA-A--A--C----------A-----G--A--T-T---A-A---C--T...C-A--A-----A----G-.....  2061
CPZ.CD.-.ANT        C-C-C--A-----T--C--C--T-----G--C--T--T--A---------...-A------A--A-GG--A.....  1452
CPZ.CM.-.CAM3       AAA----A--C--T--C--------G-----A--TTT---------C---...--A-----C--A-GG--.....  1412
CPZ.GA.88.GAB2      GA--C-----C----------A--G-----T--T-T---A--------T...-AA--------A-G----.....  1400
CPZ.GA.-.CPZGAB     AAA-A--A--C--T--------G--------T---CT---T--------T...-AA--A--A--A-G---A.....  2089
CPZ.US.85.CPZUS     AAA-C--A-----T--------------G--A--TCT-------------...-----T--A--A-G---A.....  2070
CPZ.TZ.-.TAN1       AGC-GC-------T--C-----G----T---C--T-----A--------T...-----T--A--A--G--A.....  1666
H2A.GW.-.ALI        CAA-------C-GG--------A--G---------T-G--A---C-A---...C-A--A--C--A---C--.....  2325
H2A.DE.-.BEN        CAA---G--A--GG--C-----A--G--G--A---T-G--A---C-G---...C-A--A-----A-G-C--.....  2332
H2A.SN.-.ST         CAA-------C-GG-----C--A--G---------T-G--A---C-A---...C-A--A-----A-G-C-A.....  1777
H2B.GH.86.D205      CA--G-CA--C-GG--C---------C-G--A----------GC-A---...-----------A-G-C--.....  2305
H2B.CI.-.EHO        CA--G-CA--C-GG--C--C-----G-CG--A--T--------GC-G---...-A---------A-GGC--.....  2303
H2G.CI.-.ABT96      CAA-A--A----GG----------G--G--A--T----TT---C-G---...--A-----C--A-G-C--.....  1710
H2U.FR.96.12034     CA--C-GA----GG--C--------G-----A-----T--A--GC-G--T...-AA--------A-G-C--.....  1823
MAC.US.-.239        CAA---------GG--------G-----G--A---T-T--A--GC-A---...--A-----A--A-G-C--.....  2288
Gag                 P__I__K__C__W__N__C__G__K__E__G__H__S__A__R__Q__C__._R__A__P__R__R__Q__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.  p8
SMM.SL.92.SL92B     -CA-C-G------T--C-----A--GAC---A---T-T--G---C-G---...--A--T--C--A-GG--.....  1740
SMM.US.-.H9         -AA--W------RG--C---R-A--G-----A---T-T--A---C-A-T-...--A--------A-G-C-R.....  1773
STM.US.-.STM        CA--A--A--C-GG--------G----C---------T--A-A-C-G--T...-AA-G---G--A-G-C-A.....  1947
SAB.SN.-.SAB1C      A-A----A----A---------A--GCCT--T---CTG--T---TT---T...-A----------G-C-A.....  2327
TAN.UG.-.TAN1       CACCCCGC--C-----A-----AC--ATT-----T-T-CAA-A-G----T...CCAAGAG-AG-ACC---C.....  2164
VER.KE.-.AGM155     CGCCA-------A---C-----A---TTT--C----TGCAG--GC-G--T...CCT-AG--A--A----TA.....  2187
VER.KE.-.9063       CAC-A--A----A------C--A---TTC--A----TGCAA--GC-A---...-CA-AA--A--A--G-TA.....  2204
VER.DE.-.AGM3       CA--A-------A---------A---TTT--C--T-TGCAA---C-A---...CCT-AA--A--A--G-T-.....  1699
VER.KE.-.TYO1       CAC-A-GA----AT--------A---TTT--C--T-TGCAA---C-A--T...CC--AA--A-------CA.....  1685
COL.CM.-.CGU1       ......--A--C---------CAGGG-AT-------TT--------TG--TCCC-AAAGA--A-TAGG-GGAGCCGGTCGAGGAAGAGGCCGAGGACGGGGAGGCTTTAGAGGAGCTCCCAGAAGGCCAGT  1850
DEN.CD.-.CD1        GAC---GA----------C--ATT--A---CT-----AGG----T...--AAAG--C-A--G-GT-.....ACTCAACCAGG  1815
GRV.ET.-.GRI_677    CGC-A--A--C--T-----------TTT--A--T-TGCAA--GG-A---...-A---A--A--AC-G-TC.....  2129
GSN.CM.99.CN166     CCCCA--A----AT--------AC--TTC--A---CT---A---G----T...CCTAAA--C-AAG---GA.....  1705
GSN.CM.99.CN71      CCCCA--A----A---------A--GTTT--A--T-TC--A---G-C--T...CCCAAA--C-A-G---GA.....  1708
DRL.-.-.FAO         ACCC--GA---CCA--C--C--GC-GTTT--A--T--CCTG---C-A--T...-CAAAG-----A------.....  1550
RCM.GA.-.GAB1       ..T-A-GA----A---------T--GCC---A----TTT-T---T----T...-AA--------A-GG---.....  1462
RCM.NG.-.NG411      -CC-C--A--C-AT----------AC---T--T--T--T---T------...-AAA-AAG---A--G--A.....  1465
MND-2.-.-.5440      ACCC--GA--CCC------C--A--GTTT--A--TGT-CTG--GG-----...--A-----A--A--G-GA.....  1547
MND-2.CM.98.CM16    ACA-C-GA--CCCT--C-----A---TTT--T--TGGGCTG--GG-C--T...-TAAG----C-A------.....  1990
MND-2.GA.-.M14      ACCCC-GA---CC---C-----A-GGT-T--A--TGTTCTG---G-C--T...--ACT---C--A--G---.....  1914
MNE.US.-.MNE027     CAA---------GG--------G-----G--A---T-T--A--GC-A---...--A-----A--A-G-C--.....  1768
LST.CD.88.447       C-T-A-CT--C--T--------A--GCC---T--T-----A----TG--T...--ACAG-----AC-GG--.....  1236
LST.CD.88.485       C-T-A-CT--C--T--------A--GCC---T--T-----A----TG--T...--ACAG-----AC-GG--.....  1236
LST.CD.88.524       C-C-C-CT----------G--GCC---A--TT----A-A--TG---...---CAG-----AC-GG-A.....  1230
LST.KE.-.lho7       CCC-C-CT--C----------A--GCC---T--------A----TG---...--ACAG--C--AC-GG-A.....  2320
SUN.GA.98.L14       CCA-G-CT-----T--------T---CC---A--TCT---A-ATC-G---...--A-AG--A-A----GGA.....CCACCCGG  2321
MND-1.GA.-.MNDGB1   AACC--TC--C-----C---AAT----------TGT---A--GTTC-T-...-A---------A-GG--A.....  1670
MON.CM.99.L1        AACCC--A--C----------A---TTT-------TG--G-A-------...-AA-----CCAA-GG-G-.....  1689
MON.NG.-.NG1        AAA-C-G-----AT--------G---TTT--A--TGTG--A-AG--C--T...-CT--T--C--A--G-CA.....  267
MUS.CM.01.1085      A-CCC-GA--C-A---C-----GC--T-T--C---GT------GG------...CCCAAA--C-----C--A.....  1700
DEB.CM.99.CM40      AGA---GA--C----------AC--ATT--A--TCTCCAA-AGG-C---...-AAAGA-----A--G-CC.....  1810
DEB.CM.99.CM5       AGC---GA--C--T--------AC-GATT-----TGTGCAA--GG----T...-AAAAG-----A---GTT.....  1804
SYK.KE.-.KE51       CAACA-GA--C--T--------GC-GTT---A--TCTTCAA-A-G-C---...CCTAGG--A-AA--GCTT.....  1733
SYK.KE.-.SYK173     -GA----A--C-A---------GC--ATT-------TGCAA-A-G-C---...-AAAAG--ACTT---GCT.....  2093
```

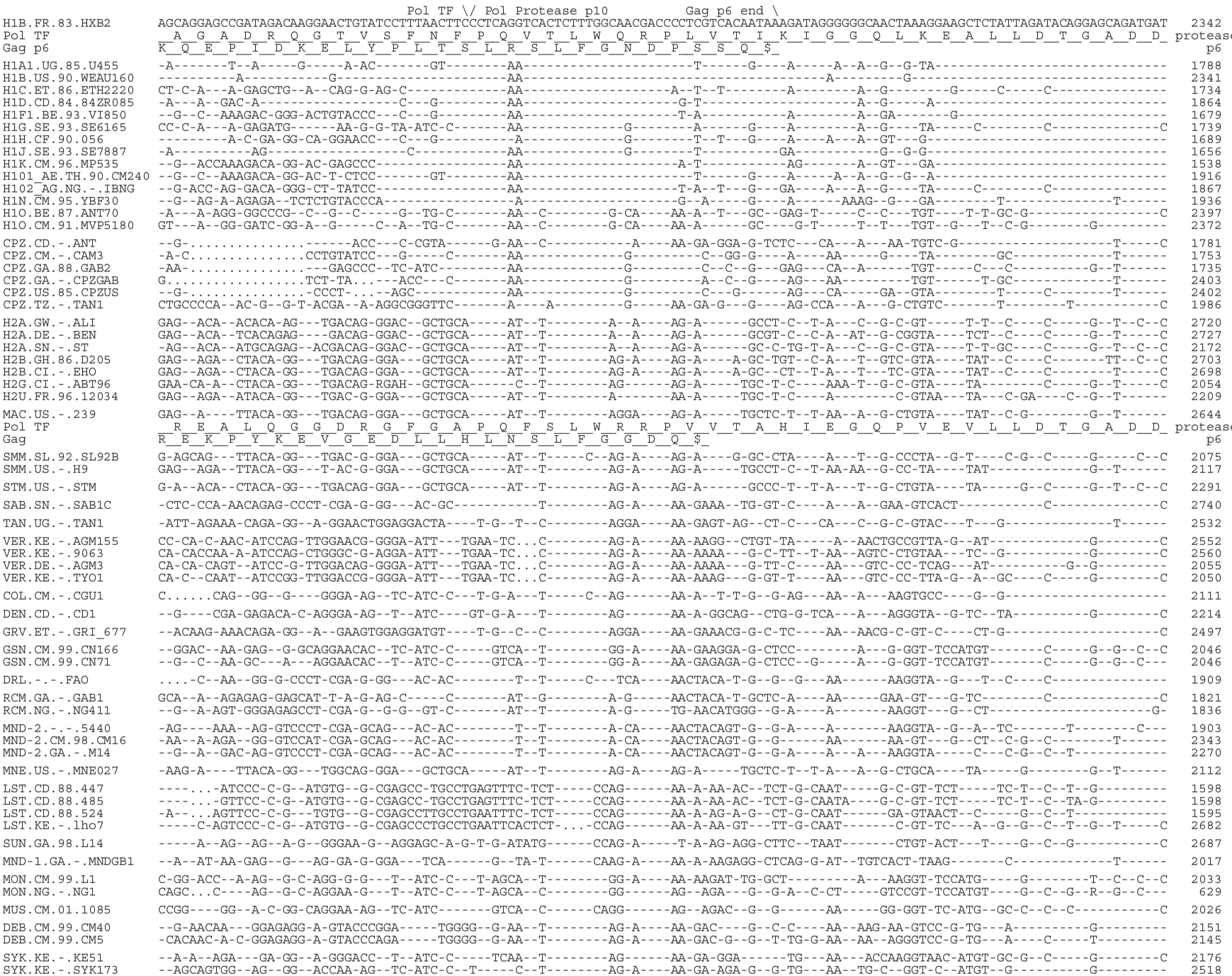
PLV
complete genomes

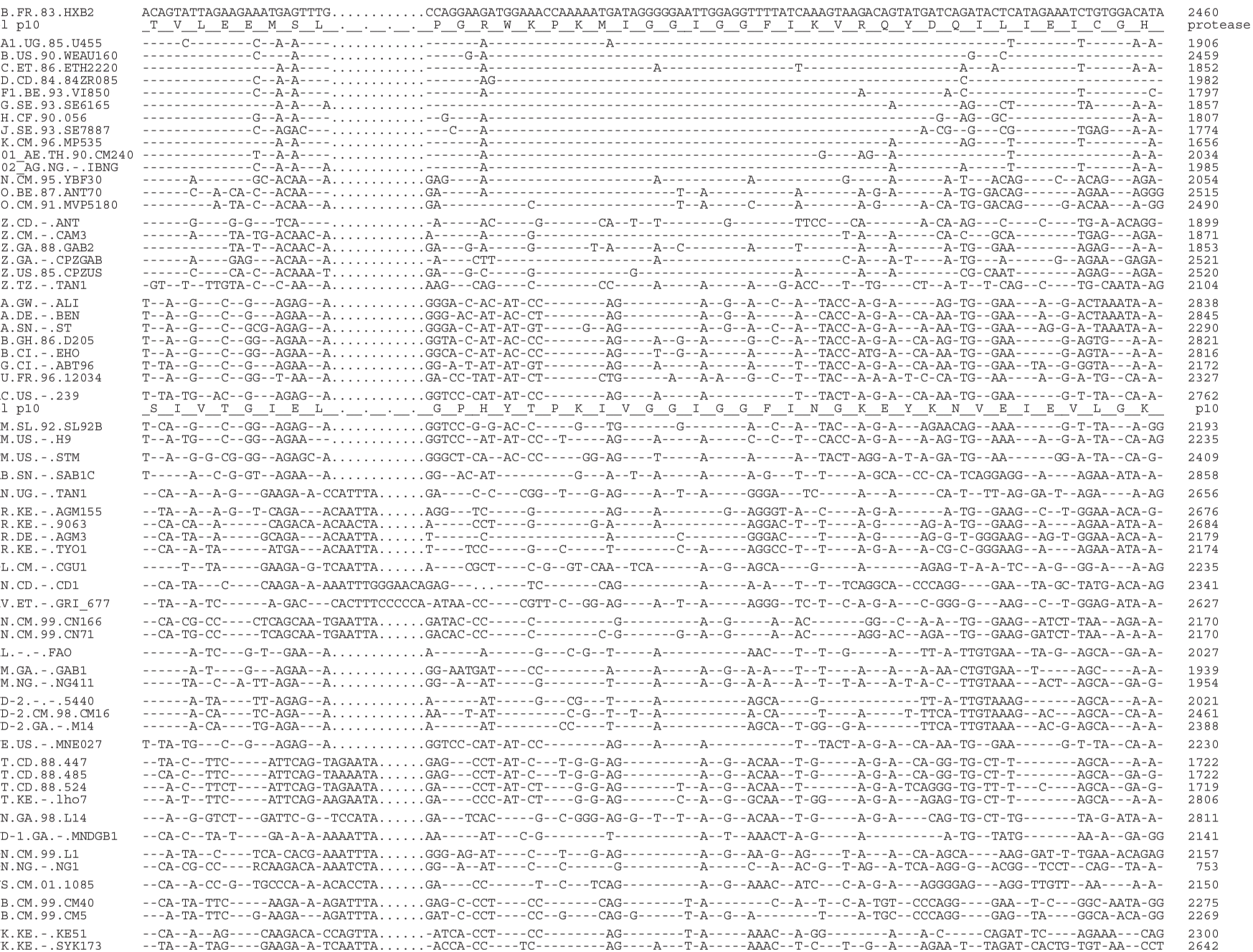
PLV
complete genomes

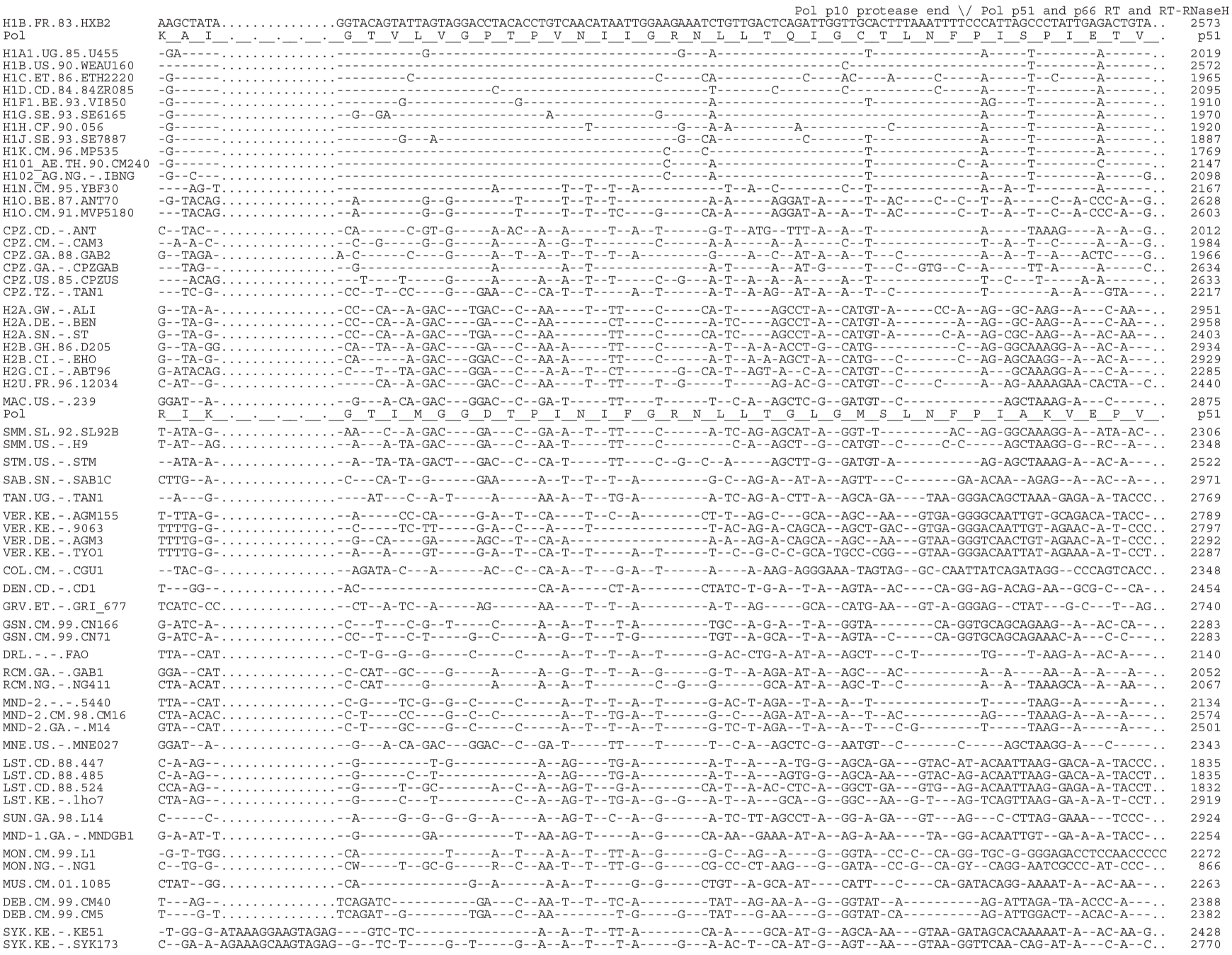



Gag p1 end \/ Gag p6 start
H1B.FR.83.HXB2
Pol
Gag p1
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Pol
Gag p1
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes


H1B.FR.83.HXB2
Pol TF
Gag p6
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Pol TF
Gag p6
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173

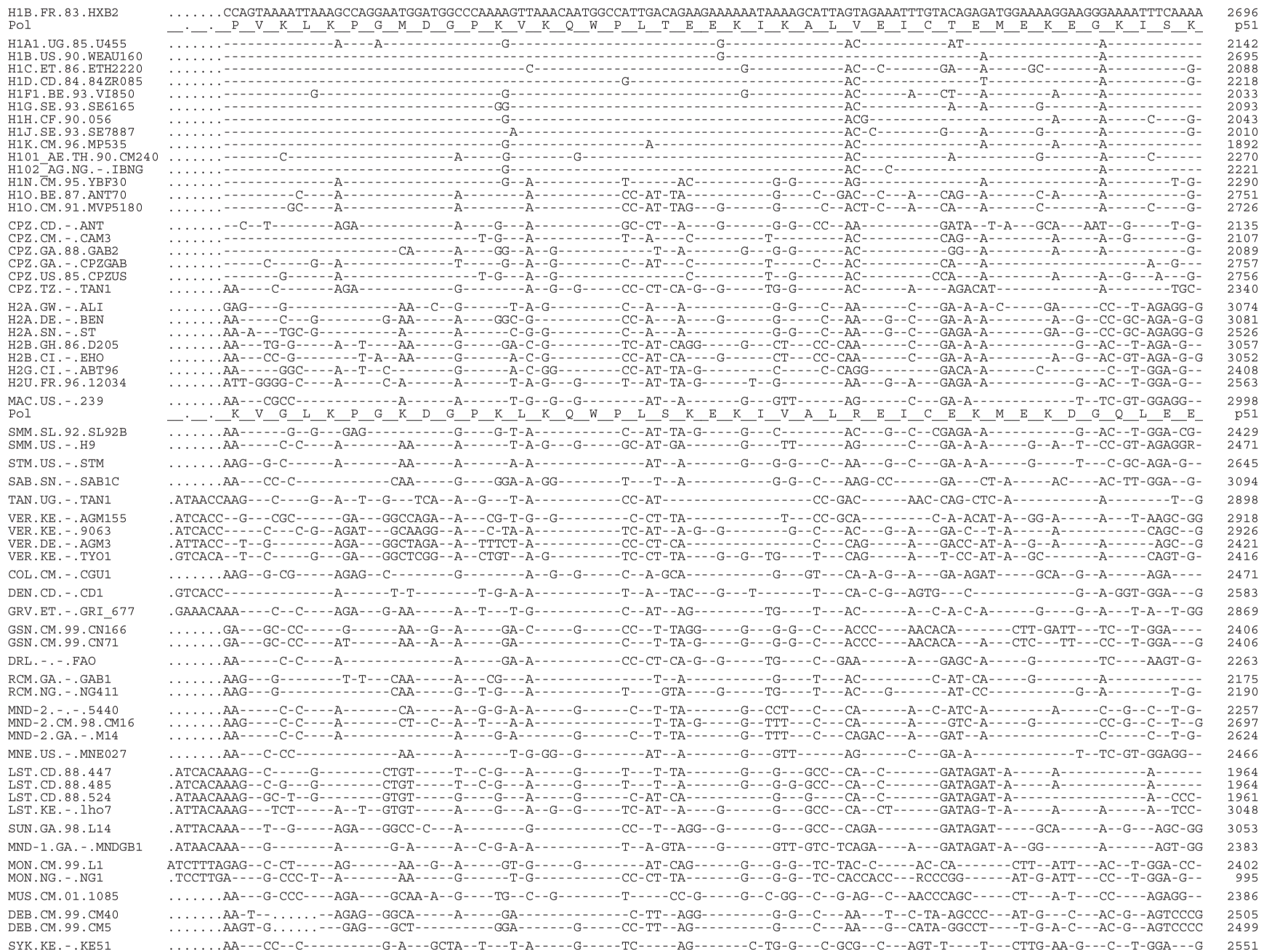

Pol TF \/ Pol Protease p10
Gag p6 end \
H1B.FR.83.HXB2
Pol TF
Gag p6
protease
p6
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Pol TF
Gag
protease
p6
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes

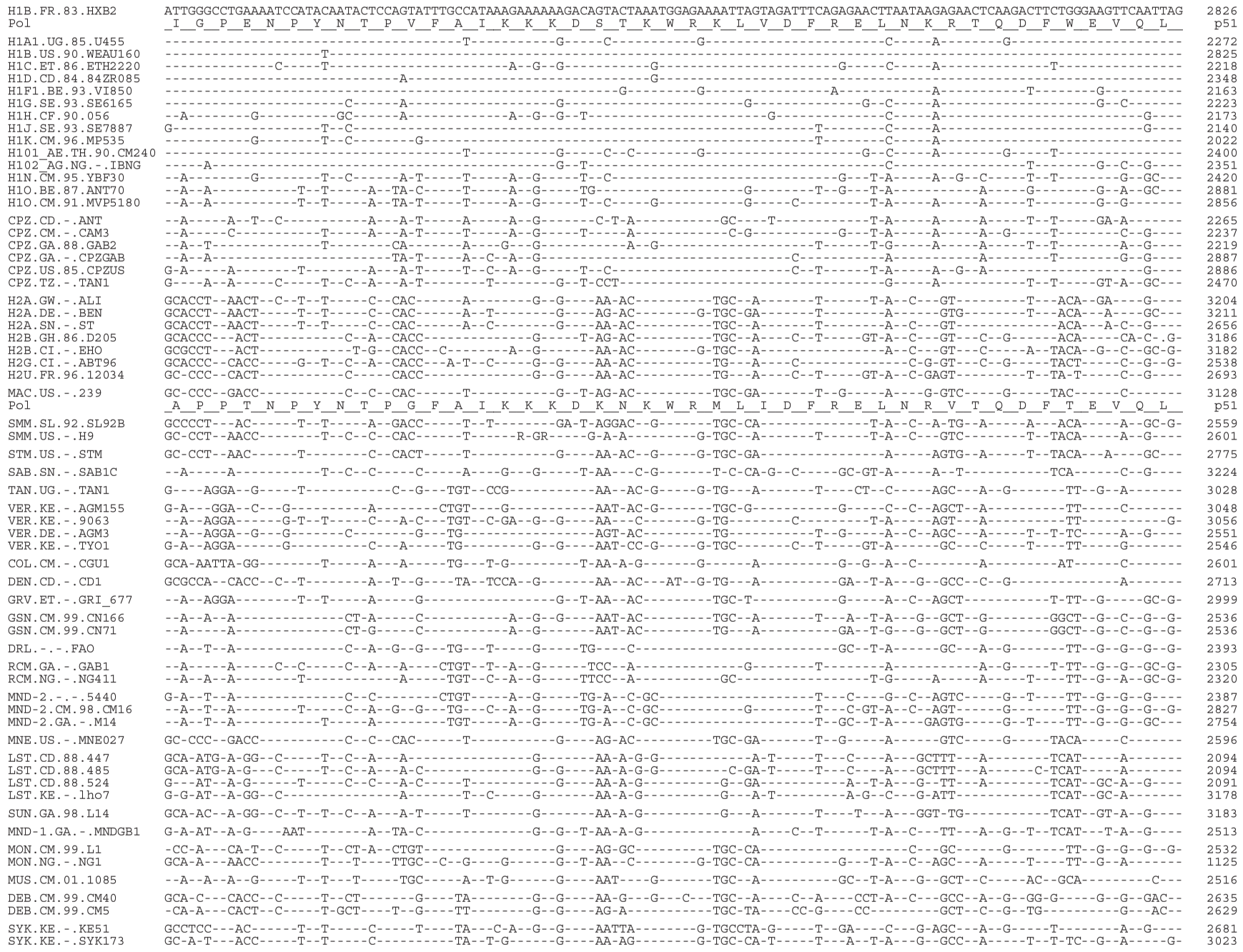

H1B.FR.83.HXB2 2460
Pol p10 protease
H1A1.UG.85.U455 1906
H1B.US.90.WEAU160 2459
H1C.ET.86.ETH2220 1852
H1D.CD.84.84ZR085 1982
H1F1.BE.93.VI850 1797
H1G.SE.93.SE6165 1857
H1H.CF.90.056 1807
H1J.SE.93.SE7887 1774
H1K.CM.96.MP535 1656
H101_AE.TH.90.CM240 2034
H102_AG.NG.-.IBNG 1985
H1N.CM.95.YBF30 2054
H1O.BE.87.ANT70 2515
H1O.CM.91.MVP5180 2490
CPZ.CD.-.ANT 1899
CPZ.CM.-.CAM3 1871
CPZ.GA.88.GAB2 1853
CPZ.GA.-.CPZGAB 2521
CPZ.US.85.CPZUS 2520
CPZ.TZ.-.TAN1 2104
H2A.GW.-.ALI 2838
H2A.DE.-.BEN 2845
H2A.SN.-.ST 2290
H2B.GH.86.D205 2821
H2B.CI.-.EHO 2816
H2G.CI.-.ABT96 2172
H2U.FR.96.12034 2327
MAC.US.-.239 2762
Pol p10 p10
SMM.SL.92.SL92B 2193
SMM.US.-.H9 2235
STM.US.-.STM 2409
SAB.SN.-.SAB1C 2858
TAN.UG.-.TAN1 2656
VER.KE.-.AGM155 2676
VER.KE.-.9063 2684
VER.DE.-.AGM3 2179
VER.KE.-.TYO1 2174
COL.CM.-.CGU1 2235
DEN.CD.-.CD1 2341
GRV.ET.-.GRI_677 2627
GSN.CM.99.CN166 2170
GSN.CM.99.CN71 2170
DRL.-.-.FAO 2027
RCM.GA.-.GAB1 1939
RCM.NG.-.NG411 1954
MND-2.-.-.5440 2021
MND-2.CM.98.CM16 2461
MND-2.GA.-.M14 2388
MNE.US.-.MNE027 2230
LST.CD.88.447 1722
LST.CD.88.485 1722
LST.CD.88.524 1719
LST.KE.-.lho7 2806
SUN.GA.98.L14 2811
MND-1.GA.-.MNDGB1 2141
MON.CM.99.L1 2157
MON.NG.-.NG1 753
MUS.CM.01.1085 2150
DEB.CM.99.CM40 2275
DEB.CM.99.CM5 2269
SYK.KE.-.KE51 2300
SYK.KE.-.SYK173 2642

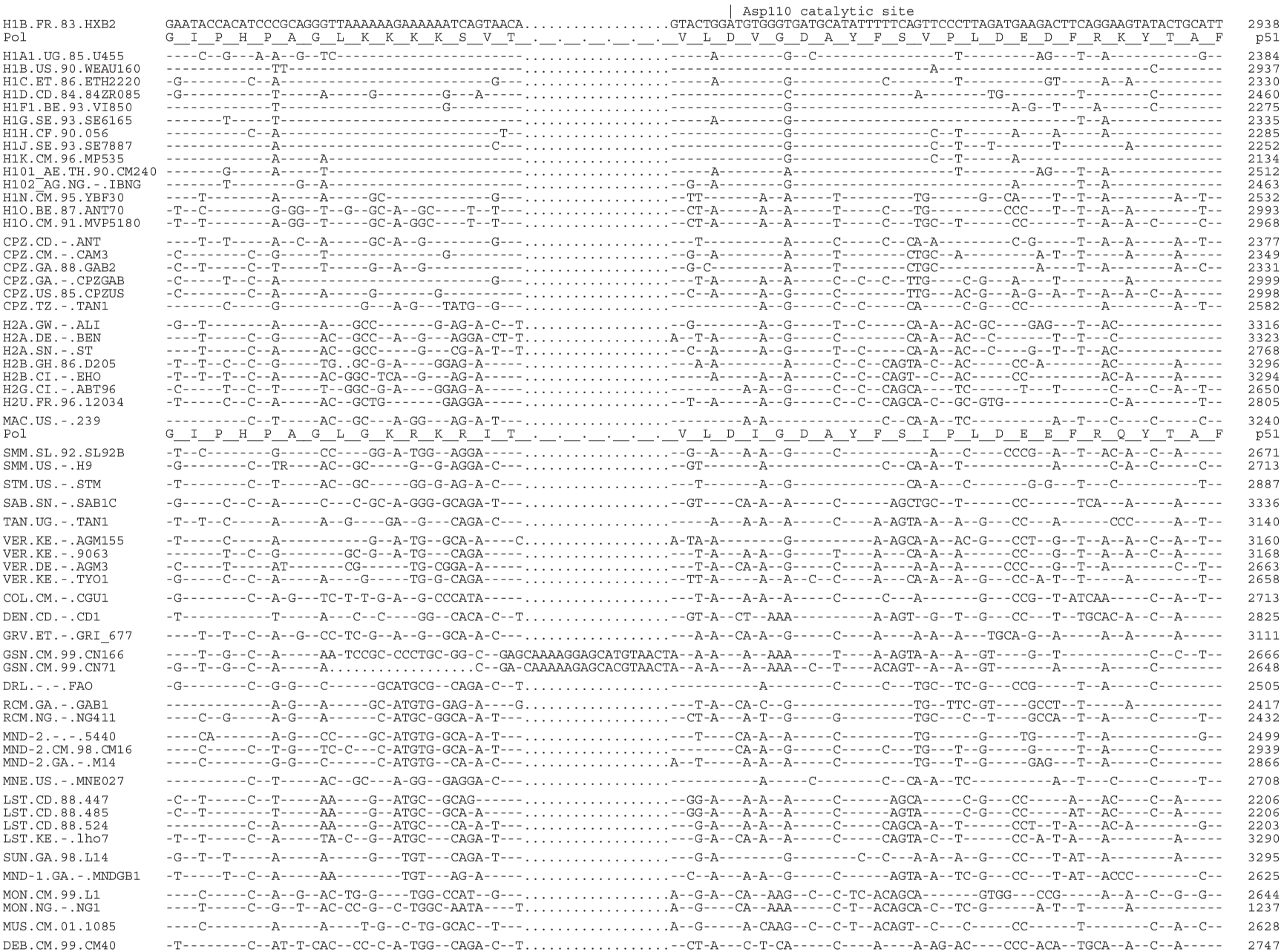


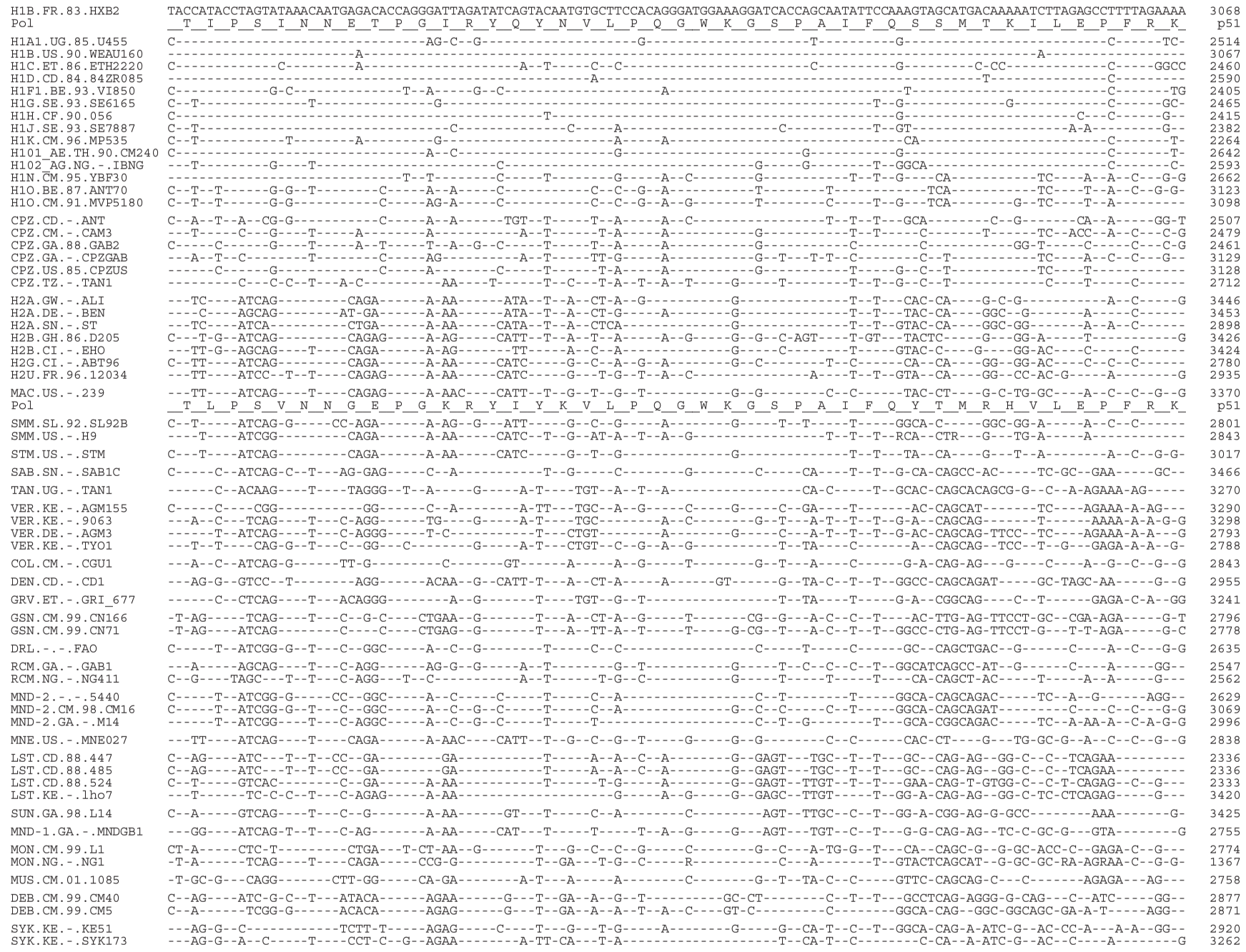
PLV
complete genomes

PLV
complete genomes

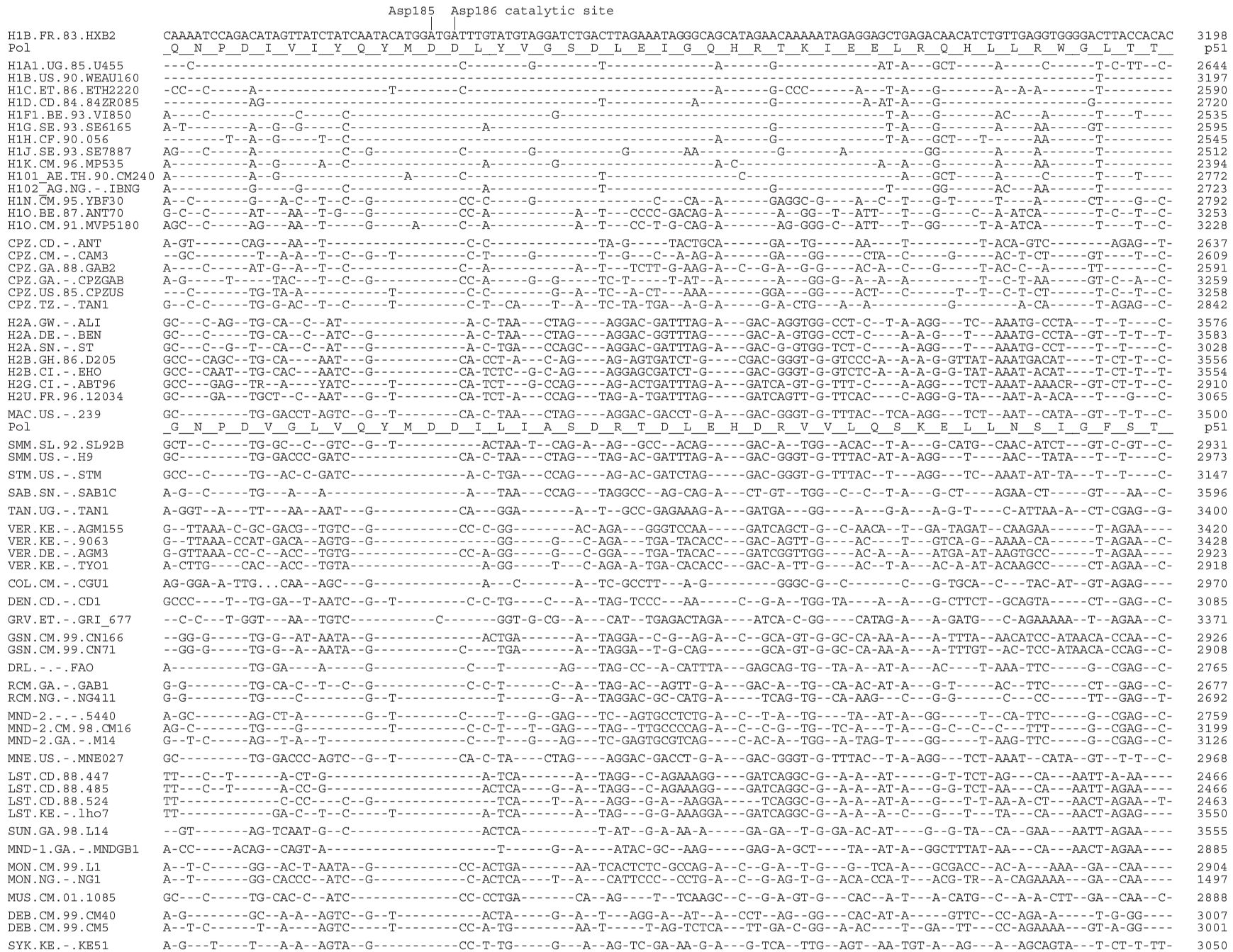



H1B.FR.83.HXB2 ATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAACTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAG 2826
Pol _I__G__P__E__N__P__Y__N__T__P__V__F__A__I__K__K__K__D__S__T__K__W__R__K__L__V__D__F__R__E__L__N__K__R__T__Q__D__F__W__E__V__Q__L__ p51
H1A1.UG.85.U455 2272
H1B.US.90.WEAU160 2825
H1C.ET.86.ETH2220 2218
H1D.CD.84.84ZR085 2348
H1F1.BE.93.VI850 2163
H1G.SE.93.SE6165 2223
H1H.CF.90.056 2173
H1J.SE.93.SE7887 2140
H1K.CM.96.MP535 2022
H101_AE.TH.90.CM240 2400
H102_AG.NG.-.IBNG 2351
H1N.CM.95.YBF30 2420
H1O.BE.87.ANT70 2881
H1O.CM.91.MVP5180 2856
CPZ.CD.-.ANT 2265
CPZ.CM.-.CAM3 2237
CPZ.GA.88.GAB2 2219
CPZ.GA.-.CPZGAB 2887
CPZ.US.85.CPZUS 2886
CPZ.TZ.-.TAN1 2470
H2A.GW.-.ALI 3204
H2A.DE.-.BEN 3211
H2A.SN.-.ST 2656
H2B.GH.86.D205 3186
H2B.CI.-.EHO 3182
H2G.CI.-.ABT96 2538
H2U.FR.96.12034 2693
MAC.US.-.239 3128
Pol _A__P__P__T__N__P__Y__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__ p51
SMM.SL.92.SL92B 2559
SMM.US.-.H9 2601
STM.US.-.STM 2775
SAB.SN.-.SAB1C 3224
TAN.UG.-.TAN1 3028
VER.KE.-.AGM155 3048
VER.KE.-.9063 3056
VER.DE.-.AGM3 2551
VER.KE.-.TYO1 2546
COL.CM.-.CGU1 2601
DEN.CD.-.CD1 2713
GRV.ET.-.GRI_677 2999
GSN.CM.99.CN166 2536
GSN.CM.99.CN71 2536
DRL.-.-.FAO 2393
RCM.GA.-.GAB1 2305
RCM.NG.-.NG411 2320
MND-2.-.-.5440 2387
MND-2.CM.98.CM16 2827
MND-2.GA.-.M14 2754
MNE.US.-.MNE027 2596
LST.CD.88.447 2094
LST.CD.88.485 2094
LST.CD.88.524 2091
LST.KE.-.lho7 3178
SUN.GA.98.L14 3183
MND-1.GA.-.MNDGB1 2513
MON.CM.99.L1 2532
MON.NG.-.NG1 1125
MUS.CM.01.1085 2516
DEB.CM.99.CM40 2635
DEB.CM.99.CM5 2629
SYK.KE.-.KE51 2681
SYK.KE.-.SYK173 3023
PLV
complete genomes


Asp110 catalytic site
H1B.FR.83.HXB2
Pol
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Pol
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173

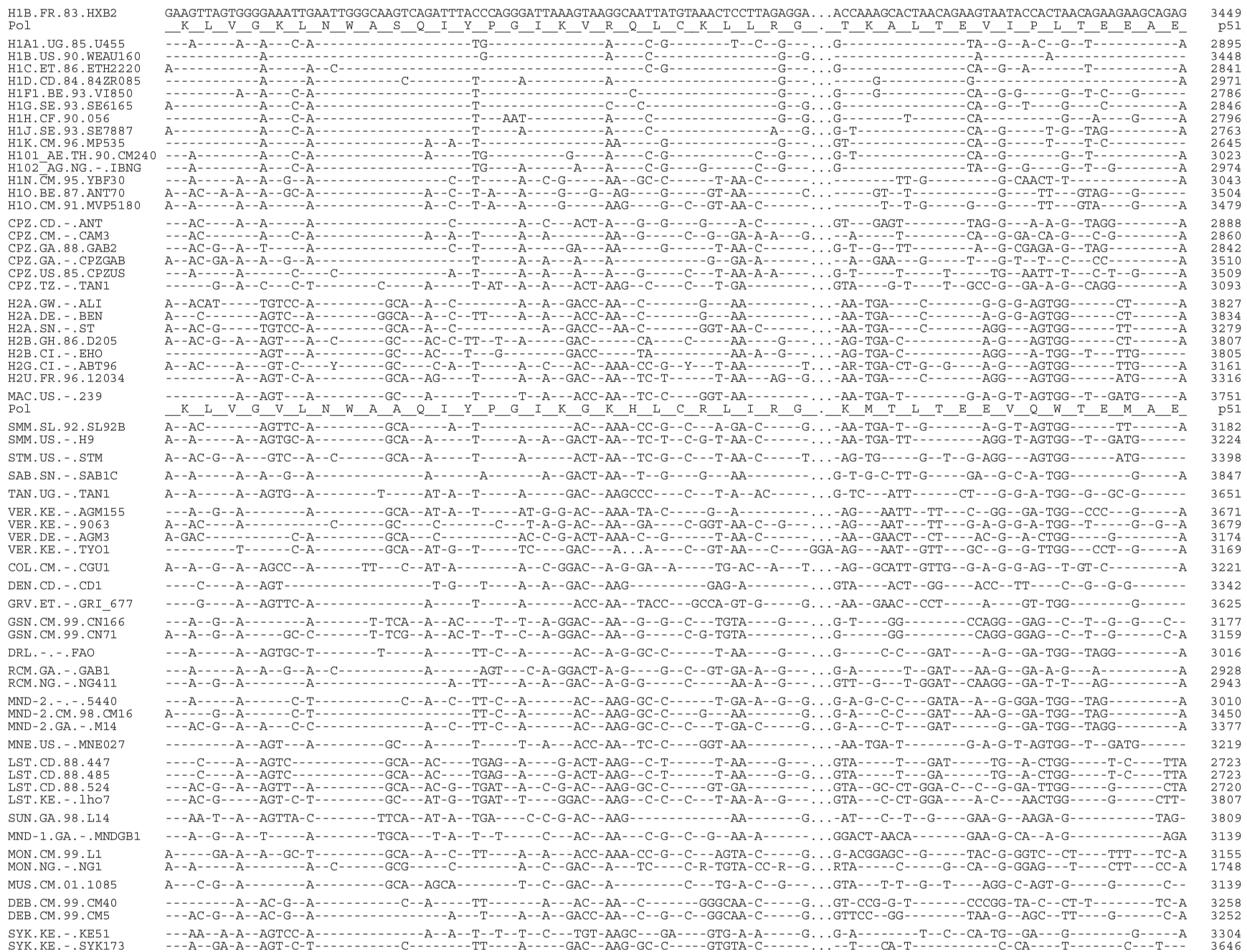


PLV
complete genomes


Asp185
Asp186 catalytic site
H1B.FR.83.HXB2 CAAAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAGCTGAGACAACATCTGTTGAGGTGGGGACTTACCACAC 3198
Pol _Q__N__P__D__I__V__I__Y__Q__Y__M__D__D__L__Y__V__G__S__D__L__E__I__G__Q__H__R__T__K__I__E__E__L__R__Q__H__L__L__R__W__G__L__T__T__ p51
H1A1.UG.85.U455 ---C-----------------------------------C--------G--------T-------------A------G-------------AT-A---GCT-----A-----C------T-C-TT--C- 2644
H1B.US.90.WEAU160 ---T--------- 3197
H1C.ET.86.ETH2220 -CC--C-----A-----------------T--------C-------------------------------A------G-CCC------A---T-A---G-------A--A-A-------T--------- 2590
H1D.CD.84.84ZR085 ----------AG---T----------A--------G----------A-AT-A---G-------------------G--------- 2720
H1F1.BE.93.VI850 A----C----------C-----C-------------------------G-------------------------------------T-A---G-------AC----A------T----T---- 2535
H1G.SE.93.SE6165 A-T--------A--G--G----C----------------A-------------------------------G-------------T-A---G-------A----AA-----GT--------- 2595
H1H.CF.90.056 --------T--A--G-----T--C---A------G-------------T-A---GCT---T------AA-----T--------- 2545
H1J.SE.93.SE7887 AG---C-----A----------C--G----------C--------G----------G------AA---------G------A----------GG-------A----A-------T--------- 2512
H1K.CM.96.MP535 A----------A--G----A--C-----------------A--------G-------------------A-C-----------------A--A---G-------A----AA------T--------- 2394
H101_AE.TH.90.CM240 A----------A--G----------------A------C----------------T-----------------C-----------------A---GCT-----A-----C------T----T---- 2772
H102_AG.NG.-.IBNG A----------G-----G-----C----------------A----------T----------G-------------T-----GG------AC---AA------T--------- 2723
H1N.CM.95.YBF30 A--C-------G---A-C--T--C--G----------CC-C-----G-------------C---CA--A------GAGGC-G----A--C--T---G-T-----T----A------CT----G--C- 2792
H1O.BE.87.ANT70 G-C--C-----AT---AA--T-G---G----------CC-A----------A--T---CCCC-GACAG-A-------A--GG--T--ATT---T---G----C--A-ATCA-------T-C--T--C- 3253
H1O.CM.91.MVP5180 AGC--C-----AG---AA--T-----G-----A-----C--A----------A--T---CC-T-G-CAG-A-------AG-GGG-C--ATT---T--GG-----T-A-ATCA-------T----T--C- 3228
CPZ.CD.-.ANT A-GT------CAG---AA--T----------------C-C--------TA-G-----TACTGCA------GA--TG-----AA----T---------T-ACA-GTC--------AGAG--T- 2637
CPZ.CM.-.CAM3 --GC--------T--A-A--T--C--G--T---------C-T-----G--------T--G---C--A-AG-A------GA---GG-----CTA--C-----G-----AC-T-CT-----GT----T--C- 2609
CPZ.GA.88.GAB2 A----C-----AT-G--A--T--C----------C--A----------A--T---TCTT-G-AAG-A--C--G-A--G-G----AC-A--C---G-----T-AC-C--A-----TT-------C- 2591
CPZ.GA.-.CPZGAB A-G-----T-----TAC---T--C--G----------CC-A-----G-G-----TC-T-----T-AT--A-------A---GG-G--A--A-A---------T--C-T-AA-----GT-C--A--C- 3259
CPZ.US.85.CPZUS ---C-------TG-TA-A----------T---------C-C--------G--A--TC--A-CT---AAA-------GGA---GG----ACT---C------T--T--C-T-CT------T-C--T--C- 3258
CPZ.TZ.-.TAN1 G--C-C-----TG-G-AC--T--C-----T---------C-T--CA----T--A--TC-TA-TGA--A-G-A-----G-A-CTG---A--A----------G--------A-CA------T-AGAG--C- 2842
H2A.GW.-.ALI GC---C-AG--TG-CA--C--AT----------------A-C-TAA---CTAG----AGGAC-GATTTAG-A---GAC-AGGTGG-CCT-C--T-A-AGG---TC--AAATG-CCTA---T--T----C- 3576
H2A.DE.-.BEN GC---C-----TG-CA--C--ATC--G----------A-C-TAA---CTAG----AGGAC-GGTTTAG-----GAC-A-GTGG-CCT-C----A-A-G---T---AAATG-CCTA--GT--T-T--T- 3583
H2A.SN.-.ST GC---C--G--T--CA--C--AT---G----------A-C-TGA---CCAGC---AGGAC-GATTTAG-A---GAC-G-GTGG-TCT-C----A-AGG---T---AAATG-CCT----T--T----C- 3028
H2B.GH.86.D205 GCC--CAGC--TG-CA----AAT---G----------CA-CCT-A---C-AG----AG-AGTGATCT-G----CGAC-GGGT-G-GTCCC-A--A-A-G-GTTAT-AAATGACAT----T-CT-T--C- 3556
H2B.CI.-.EHO GCC--CAAT--TG-CAC---AATC--G----------CA-TCTC--G-C-AG----AGGAGCGATCT-G-----GAC-GGGT-G-GTCTC-A--A-A-G-G-TAT-AAAT-ACAT----T-CT-T--T- 3554
H2G.CI.-.ABT96 GCC---GAG--TR--A----YATC----T---------CA-TCT---G-CCAG----AG-ACTGATTTAG-A---GATCA-GT-G-TTT-C----A-AGG---TCT-AAAT-AAACR--GT-CT-T--C- 2910
H2U.FR.96.12034 GC----GA---TGCT--C--AAT---G--T---------CA-TCT-A---CCAG---TAG-A-TGATTTAG-----GATCAGTT-G-TTCAC-----C-AGG-G-TA---AAT-A-ACA---T---G---C- 3065
MAC.US.-.239 GC---------TG-GACCT-AGTC--G--T--------CA-C-TAA---CTAG----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC--TCA-AGG---TCT--AAT--CATA--GT--T-T--C- 3500
Pol _G__N__P__D__V__G__L__V__Q__Y__M__D__D__I__L__I__A__S__D__R__T__D__L__E__H__D__R__V__V__L__Q__S__K__E__L__L__N__S__I__G__F__S__T__ p51
SMM.SL.92.SL92B GCT--C-----TG-GC--C--GTC--G--T----------ACTAA-T--CAG-A--AG--GCC--ACAG-----GAC-A--TGG--ACAC--T-A---G-CATG--CAAC-ATCT---GT-C-GT--C- 2931
SMM.US.-.H9 GC---------TG-GACCC-GATC-------------CA-C-TAA---CTAG---TAG-AC-GATTTAG-A---GAC-GGGT-G-TTTAC-AT-A-AGG---T----AAC--TATA---T--T----C- 2973
STM.US.-.STM GCC--C-----TG--AC-C-GATC----------A-C-TGA---CCAG----AG-AC-GATCTAG-----GAC-GGGT-G-TTTAC--T---AGG---TC--AAAT-AT-TA---T--T----C- 3147
SAB.SN.-.SAB1C A-G--C-----TG---A---A-----------------A---TAA---CCAG---TAGGCC--AG-CAG-A---CT-GT--TGG--C--C--T-A---G-CT----AGAA-CT-----GT---AA--C- 3596
TAN.UG.-.TAN1 A-GGT--A---TT---AA--AAT---G----------CA---GGA-------A--T--GCC-GAGAAAG-A---GATGA---GG----A---G-A---A-G-T----C-ATTAA-A--CT-CGAG--G- 3400
VER.KE.-.AGM155 G--TTAAA-C-GC-GACG--TGTC--G----------CC-C-GG--------AC-AGA---GGGTCCAA-----GATCAGCT-G--C-AACA--T--GA-TAGAT--CAAGAA------T-AGAA---- 3420
VER.KE.-.9063 G--TTAAA-CCAT-GACA--AGTG--G----------------GG-----G---C-AGA---TGA-TACACC---GAC-AGTT-G----AC----T---GTCA-G--AAAA-CA------T-AGAA--C- 3428
VER.DE.-.AGM3 G-GTTAAA-CC-C--ACC--TGTG----------CC-A-GG----G---C-GGA---TGA-TACAC----GATCGGTTGG----AC-A--A---ATGA-AT-AAGTGCC------T-AGAA---- 2923
VER.KE.-.TYO1 A-CTTG----CAC--ACC--TGTA----------------A-GG-----T---C-AGA-A-TGA-CACACC---GAC-A-TT-G----AC--T-A---AC-A-AT-ACAAGCC-----CT-AGAA--C- 2918
COL.CM.-.CGU1 AG-GGA-A-TTG...CAA--AGC---G----------A---C-------A--TC-GCCTT---A-G---------GGGC-G--C-------C---G-TGCA--C---TAC-AT--GT-AGAG---- 2970
DEN.CD.-.CD1 GCCC----T--TG-GA--T-AATC--G--T---------C-C-TG----C---A--TAG-TCCC---AA-----C--G-A--TGG-TA----A--A---G-CTTCT--GCAGTA-----CT--GAG--C- 3085
GRV.ET.-.GRI_677 --C-C---T-GGT---AA--TGTC----------C-------GGT-G-CG--A----CAT--TGAGACTAGA----ATCA-C-GG----CATAG-A---A-GATG---C-AGAAAAA--T--AGAA--C- 3371
GSN.CM.99.CN166 --GG-G-----TG-G--AT-AATA--G----------ACTGA-------A--TAGGA--C-G--AG-A--C--GCA-GT-G-GC-CA-AA-A---A-TTTA--AACATCC-ATAACA-CCAA--C- 2926
GSN.CM.99.CN71 --GG-G-----TG-G--A--AATA--G----------C----TGA-------A--TAGGA--T-G-CAG-------GCA-GT-G-GC-CA-AA-A---A-TTTGT--AC-TCC-ATAACA-CCAG--C- 2908
DRL.-.-.FAO A----------TG-GA----A-----G----------C---T-------AG---TAG-CC--A-CATTTA---GAGCAG-TG--TA-A--AT-A---AC----T-AAA-TTC-----G--CGAG--C- 2765
RCM.GA.-.GAB1 G-G--------TG-CA-C--T--C--G----------C-C-T------C--A--TAG-AC--AGTT-G-A---GAC-A--TG--CA-AC-AT-A---G-T-----AC--TTC-----CT--GAG--C- 2677
RCM.NG.-.NG411 G-G--------TG----C--------G--T----------T------G--A--TAGGAC-GC-CATG-A-----TCAG-TG--CA-AAG---C---G-G------C---CC-----TT--GAG--T- 2692
MND-2.-.-.5440 A-GC-------AG-CT-A--------G--T--------C----T---G--GAG---TC--AGTGCCTCTG-A--C--T-A--TG----TA--AT-A--GG-----T--CA-TTC-----G--CGAG--C- 2759
MND-2.CM.98.CM16 AG-C-------TG----G----------T---------C-C-T---T--GAG---TAG--TTGCCCCAG-A--C--C-G--TG--TC-A---T-A---G-C--C--C---TTT-----G--CGAG--C- 3199
MND-2.GA.-.M14 G--T-C-----AG--T-A--T----------------C---T---G---AG---TC-GAGTGCGTCAG----C-AC-A--TGG--A-TAG-T----GG-----T-AAG-TTC-----G--CGAG--C- 3126
MNE.US.-.MNE027 GC---------TG-GACCC-AGTC--G--T--------CA-C-TA----CTAG----AGGAC-GACCT-G-A---GAC-GGGT-G-TTTAC--T-A-AGG---TCT-AAAT--CATA--GT--T-T--C- 2968
LST.CD.88.447 TT---C--T------A-CT-G----------------A-TCA-------A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-----G-T-TCT-AG---CA---AATT-A-AA---- 2466
LST.CD.88.485 TT---C--T------A-CC-G----------------ACTCA----G--A--TAGG--C-AGAAAGG----GATCAGGC-G--A-A--AT-A---G-G-TCT-AA---CA---AATT-AGAA---- 2466
LST.CD.88.524 TT-------------C-CC---C--G----------------TCA----T--A---AGG--G-A--AAGGA-----TCAGGC-G--A-A--AT-A---G---T-T-AA-A-CT---AACT-AGAA--T- 2463
LST.KE.-.lho7 TT-------------GA-C--T--C-----T----------A-TCA-------A--TAG---G-G-AAAGGA---GATCAGGC-G--A-A--A--C---G---T---TA---CA--AACT-AGAG---- 3550
SUN.GA.98.L14 --GT-------AG-TCAAT-G--C----------------ACTCA----------T-AT--G-A-AA-A-----GA-GA--T-G-GA-AC-AT----G--G-TA--CA--GAA---AATT-AGAA---- 3555
MND-1.GA.-.MNDGB1 A-CC-----ACAG--CAGT-A----------------------T------G--A----ATAC-GC--AAG-----GAG-A-GCT----TA--AT-A--GGCTTTAT-AA---CA---AACT-AGAA---- 2885
MON.CM.99.L1 A--T-C-----GG--AC-T-AATA--G----------CC-ACTGA-------AA-TCACTCTC-GCCAG-A--C--G-A--T-G---G--TCA-A---GCGACC--AC-A---AAA--GA--CAA---- 2904
MON.NG.-.NG1 A--T-------GG-CACCC--ATC--G----------C-ACTCA----T--A---CATTCCC--CCTG-A--C--G-AG-T-G--ACA-CCA-T---ACG-TR--A-CAGAAAA---GA--CAA---- 1497
MUS.CM.01.1085 GC---C-----TG-CAC-C--ATC----------CC-CCTGA-------CA--AG----T--TCAAGC--C--G-A-GT-G---CAC--A-T---A-CATG--C--A-A-CTT--GA--CAA--C- 2888
DEB.CM.99.CM40 A-G--------GC--A-A--AGTC--G--T----------ACTA-----G--A--T---AGG-A--AT--A--CCT--AG--GG---CAC-AT-A-----GTTC--CC-AGA-A-----T-G-GG---- 3007
DEB.CM.99.CM5 A--T-C------T--A----AGTC-----T---------CC-A-TG--------AA-T-----T-AG-TCTCA---TT-GA-C-GG--A-AC----T--GA--TT---CC-AGAAAA---GT-A-GG---- 3001
SYK.KE.-.KE51 A-G---T----T---A-A--AGTA--G----------CC-T-TG-----G---A--AG-TC-GA-AA-G-A---G-TCA--TTG--AGT--AA-TGT-A--AG---A--AGCAGTA---T-CT-T-TT- 3050
SYK.KE.-.SYK173 A-GT--T----TC-CAC-C-AATA--G----------C--A-TGA-T--CA----TAG-TC-GAGAA-GCA---CAGGAG-T-G--C-AC-AA-AGTGAC-GCA--A--A-AAGT----T---AGGTC- 3392

```
H1B.FR.83.HXB2      CAGACAAAAAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAGTGCTGCCAGAAAAAGACAGC......TGGACTGTCAATGACATACA   3322
Pol                 P__D__K__K__H__Q__K__E__P__P__F__L__W__M__G__Y__E__L__H__P__D__K__W__T__V__Q__P__I__V__L__P__E__K__D__S__.__.__W__T__V__N__D__I__Q    p51

H1A1.UG.85.U455     ----------G-------------------T----------G--------T--------C----------T---------CA------------------.....---------------T-----   2768
H1B.US.90.WEAU160   ----------------A-----C-----------------------------------------------------------AA--------------A--T.....---------------------   3321
H1C.ET.86.ETH2220   -------G----------------------T----------G--------T--------C---------------------CA-----------G--T---.....---------------T-----   2714
H1D.CD.84.84ZR085   ------------------------------T-------------------------------G-----------T-----ACA--------G-----A---.....---------------T-----   2844
H1F1.BE.93.VI850    -------------------------------------GC-------------------------------G---------CAAT-----A-C--G------.....---------------T-----   2659
H1G.SE.93.SE6165    ----T--------------------------------A-----G-----------C--------G-----A------CA---------C--G--A---.....---------------T-----   2719
H1H.CF.90.056       -----------------------C-----T----------A----------------C--------------A--G--AAA-----------------.....---------------T-----   2669
H1J.SE.93.SE7887    ----T-------------------------T----------A-----G--T--------C----------C--A------CAA-------------AGA-.....---------------T-----   2636
H1K.CM.96.MP535     ----------G-------------C-----T----------A----------------C----------G---------CAA--------C--G--T---.....---------------T-----   2518
H101_AE.TH.90.CM240 ----------G--------G--------------------A----------------C-G---------C---------AA-----------------.....---------------T-----   2896
H102_AG.NG.-.IBNG   ----------G-------------------T----------------------------------------C------G---AA--A--------------.....---------------T-----   2847
H1N.CM.95.YBF30     -T----------------G-----C--G-----C--------A-------------A--C----------C-----A---AA-T-A---------G--TGTA.....---------------T-----   2916
H1O.BE.87.ANT70     -T--------G-----A--G--------C--T--A--------G-----G-------A--C---------------T-C--CCAAT----TA-C--G--TGTG.....------A--A-----T-----   3377
H1O.CM.91.MVP5180   -T--T-----G-----------G--------C--TT-A--------A-----G-----C--A--C--G-------------C--CCAAT----T--C-----AGTG.....------A--A-----T-----   3352

CPZ.CD.-.ANT        -T--------G-------------------T-AA--------A-----GT-A------A--C--------T------AAA---AA---A-----GCC---TGAT.....------A--T--------C--   2761
CPZ.CM.-.CAM3       -T--T-----G--------G--G--A--T--TT-G--------C-----G-------------G---------G-----A---CA-T-A---C-G-----G-C-.....-----------------T--T--   2733
CPZ.GA.88.GAB2      -T-----G--G--C--A--G--------T--T--C--------G-----G--T------A--C----------G-----A---CA------TC-G-----G-AT.....----------A--------T--   2715
CPZ.GA.-.CPZGAB     ----------------A--G--G--A--C--TT-A--------A-----G-----C--A--C----------C--------TCAAT-A-----------GGTA.....---------------T-----   3383
CPZ.US.85.CPZUS     -T--T--G--G------G-G-----A------T-A--------C---------------A------G--------G-----C---CA-T-A---C-------A-TT.....------A--------T--T--   3382
CPZ.TZ.-.TAN1       -T--------GT-----G--A-----------A-G--------C-----G--A------AA------------T--AAA---CACAT-A------CC---GCAG.....------A--G---C-T--C--   2966

H2A.GW.-.ALI        ----TG-G--GTTC--A------C--------A-AAA--------C----G---GTGG--AAC--------AGC-G--AAAA---CAAT----CC-G-----AGTA.....------A-----------C--   3700
H2A.DE.-.BEN        ----TG-G--GTTC--A------C---------T-AA--------C-G------ATGG--AAC--------AGC-G---AAAC--CAA-----CC-G-------TA.....------A-----------C--   3707
H2A.SN.-.ST         ----TG-G--GTTC--A------C--------A--AA--------C--------GTGG--AAC--------AGC-G--AAGA---CAAT----CC----G--AGTA.....------A-----------C--   3152
H2B.GH.86.D205      ----AG----GTTC--A------C-----G---AAA--------------G---TGG--AA-A--G----A-C-G--AAAA---CAA------------AGTT.....------A--G----CA--T--   3680
H2B.CI.-.EHO        ----AG----GTTC--A------C---------AAA--------G-----G---TGG--AA-G-------A-C-G--AAAA---CA---A-----------GGTT.....------A--A--------T--   3678
H2G.CI.-.ABT96      -T--TG----GTTC--A------C--------AT-AA--------A-------T-RTGG--AA-A--G----A-C---AAAG--TAAC-----M--------G-CT.....------A--A-----T--T--   3034
H2U.FR.96.12034     -----G-----TT---------C--A---CA-AAA--------C---------ATGG---A-A--G----A-T-G---AAA--T-A---C-------G---TGTT.....------A--A--------T--   3189

MAC.US.-.239        ----AG-G---TTC--A-----T--C-----T-AA--------G--C---T-GTGG--AACA-------AGT-G--AAAG----A-T-----C---G---G-C-.....------A--G-----T-----   3624
Pol                 P__E__E__K__F__Q__K__D__P__P__F__Q__W__M__G__Y__E__L__W__P__T__K__W__K__L__Q__K__I__E__L__P__Q__R__E__T__.__.__W__T__V__N__D__I__Q    p51
SMM.SL.92.SL92B     ----AG-C--GTT---A--GA-T--A------A-AG--------A---TTGT-AT-C--CA-G--G----A-T-----AAA----AA--C--T-----G--GC-A.....------A--A-----------   3055
SMM.US.-.H9         ----AG-G--GTTC--A------T--C--------AG--------A-------T-GTGG--CACC--------A-C-G--AAAA----A-T----GC---G---A-C-.....------A--A---A-T-----   3097

STM.US.-.STM        -T--AG-G---TT-------G--T---------T-AA--------A------GT-ATGG--AACA--G----A-C-G--AAAA----AAT-A--TC-G-G----GTT.....------A--G-----T--C--   3271

SAB.SN.-.SAB1C      -T--A--G--GTT---A--G--T--A----A---C--------G-------T-GT----AA-G------CA-C-G---GAG--CACC--A--------GG--AGAA.....------A--G-----T--T--   3720

TAN.UG.-.TAN1       ----T--G--GGTG------AG---------ATGA---------A--CA-GT-GT-C---TC-C-C----GTC-CA-CAAA----A-A-A-----GC----ACA-.....------A--G-----T--T--   3524

VER.KE.-.AGM155     ----G-----GGTG--A-G----------C--TGAG--------A---A--T-ATGG---C---------AGT----AAG-----AAT-AGAGA-G------ACAA.....------A--G-----TC-T--   3544
VER.KE.-.9063       ----------GGTG--A--GA-G--A----ATGAA--------A--CA--T-ATGG--AC-A-----CA-C--A-TAG----AACT-AGA------G--ACAA.....------------T-----T--C--   3552
VER.DE.-.AGM3       -------G---GTA-------A-----A--T-ATGAG--------A--CA--T-GTGG--AC-C--G----CAGA--AGCAGC----AAT-AGA---C------AGAA.....------A--------T-----   3047
VER.KE.-.TYO1       ----A--G--GATG--A---------A--T-ATGAG--------A--CA-----TTGG---C-C-------GA-C--AGCAGA---CAA----GAG---------TGAA.....-----------------C--   3042

COL.CM.-.CGU1       ----A--G---T-C---GC---T---------A-AAG--------C-------T-A-------A-GC-------A-T----AAAGG-CAAAA-T-------C-G--T-AG.....------A--A-----T-----   3094

DEN.CD.-.CD1        -T--TGCC---TT---AG--AC---A--CA-A-AA--------G-------T-A--C--CA-AC-C---CA-A-----AAA----AC-------TTGCC---GGATGGGCAG------G--G----CA--T--   3215

GRV.ET.-.GRI_677    -------G---GTC--A-G-------G----GGGAA--------G---A--T-G------GA-----------CA-TA-CAAA----AAT-A--CCCCTT---AG-A...GAA-----A--A--CA-A-----   3498

GSN.CM.99.CN166     -C--GGCC------C--ACC---TTA---GG--AAG---C-C--G------GT-G-------A-GGG------GGC-C------GG-G-AA--C------CC-G--T-C-.....------A--G--------C--   3050
GSN.CM.99.CN71      -C--GGCC------C--AC----TTA----G--AGA---C-C--G------GT-G--C----A-GGG------GGC-C--A--AG-G-A---C------TC------T-.....------A--A--------C--   3032

DRL.-.-.FAO         -G---------TT----G-CA-T------T-ATGAA--------G--CC-G--G------AA-G--G---------G--AGAA--TCGCT-----T--T------A-TT.....-----------T-----T--T--   2889

RCM.GA.-.GAB1       -----------TT----G-TA-----------ATT-G--------G------G--G--C--AA-A-GT------------GAG--CAA-T-A------G------AGAA.....------A--T--C--T--T--   2801
RCM.NG.-.NG411      ----T--G---TT----G-CA-G------C--TGAA--------C-----TG----------A--C-------------AGAG--CAA-T-A-T----------AGAA.....------A--A-----T-----   2816

MND-2.-.-.5440      ----------GTTC--A--G---------C--TGAA--------A-----TG--G--C--AA-G------------T---AAA---CA---A------------A---.....------A--A-----T--T--   2883
MND-2.CM.98.CM16    -------G--GTT---A--G--------A--C--TGAG--------A--CAT---G--C---A-G-----------G---AAAG--CA---T------G-----AGAA.....------A--A--------C--   3323
MND-2.GA.-.M14      ----------GTTC--A-----------------T--TGAG--------G-----TG--G--C--AC-G-----------C--AAAG---CA-----------------G-CA.....------A--A-----T--T--   3250

MNE.US.-.MNE027     ----AG-G--GTTC--A------T--C-----T-AA--------G-------T-ATGG--AACA-------AGT-G--AAAG----A-T-----C-------G-C-.....------A--G-----T-----   3092

LST.CD.88.447       ----G--G---TT---AGC---------C-AT-A---------G-----T---T------A----GG---GA-A--G--AAA-CTAAA---------C-T----CTTCAGAAGACA--A--T--CA-------   2596
LST.CD.88.485       ----A--G---TT---AGC---G------C-AT-A---------G-----T---T------A----GG---GA-A--G--AAAG-TAAA---------C-T----CTTCAGAAGACA--A--A--CA-------   2596
LST.CD.88.524       ----G--G---TT---AG-CA----A--C-A--A----------G-----TC--G--C--A----G-----GA-A--G--AGAG--AAA---------G-T-----CAAGGAAAACT--A--A-----A--T--   2593
LST.KE.-.lho7       ----A--G--GTTC--AGC---G--A--C-AT-A--------------T---G-----------GG---GA-A--G-AAAG--TAAAT-A------G-TG--TCTAACAAAAACT--A--A---C-A-----   3680

SUN.GA.98.L14       ----G------T-----GG---G--A--C-ATAAG--------G--CAT-T-A--C--A--C-G----GAGA--G--AAG---AAAT-A--CCCTCT---AGAAGAG...CCA--A--G-----------   3682

MND-1.GA.-.MNDGB1   -T--A--G---T-----------------C--T-A--------G-----GT-A--C--A-------G----AGA--G-AAAGG-TCAA--A------TT--CAGAACAG...CCA--A--A------A------   3012

MON.CM.99.L1        -CCCAG-G---TT----G-T--CTAC--GG-G-AA---T-A--G------G--A--C--ATCCC-T----A-A-TGCCAAA----A---A------CC-G--ACAA.....----------------G--C--   3028
MON.NG.-.NG1        --ATAG-G---T-----G-CA-GTAC------G-AG---T----G------------C--CACCC-T---------C--AAAG----AA-----------C------CAG.....------A--------GG--C--   1621

MUS.CM.01.1085      -TC-AG-----TT----G----CTAC---G---AA---T-A--G-----G--A--C--CCGGGGC----A-A--------CG---CAT------C--C----AGTT.....------A--T-----TG-G--   3012

DEB.CM.99.CM40      --CCTG-G--GT-----G----C--A--C--T-AC---------G------GT-A--C---A-A------CGGT-GAT---AG-TACC--A------TG-----AGAA.....------A--GC--A-A--T--   3131
DEB.CM.99.CM5       -CCCGG-G---T-C---G-G--T--C-----A--A---------G------GT-A------AA-G------CACT-GTT---A-----A---A------G-G--GGAG.....------A----C-CA-A--C--   3125

SYK.KE.-.KE51       --CCAG-----TGG---G-CA--TTC--C--G-AA---T-A--G--CACG--A--C---------G----GTC----AAAGG-TCAA--A-------TTACAGAATCA...CCA--A--T-----GT----   3177
SYK.KE.-.SYK173     --A-AG-----TGG---G-CC--TA----A-G-AG---T-A-----CACT--T-----A--C--G----CAGT-G---AAA----AA--C--TA-T-T---TGA-GAA...ATC--A--G----C-GT----   3519
```

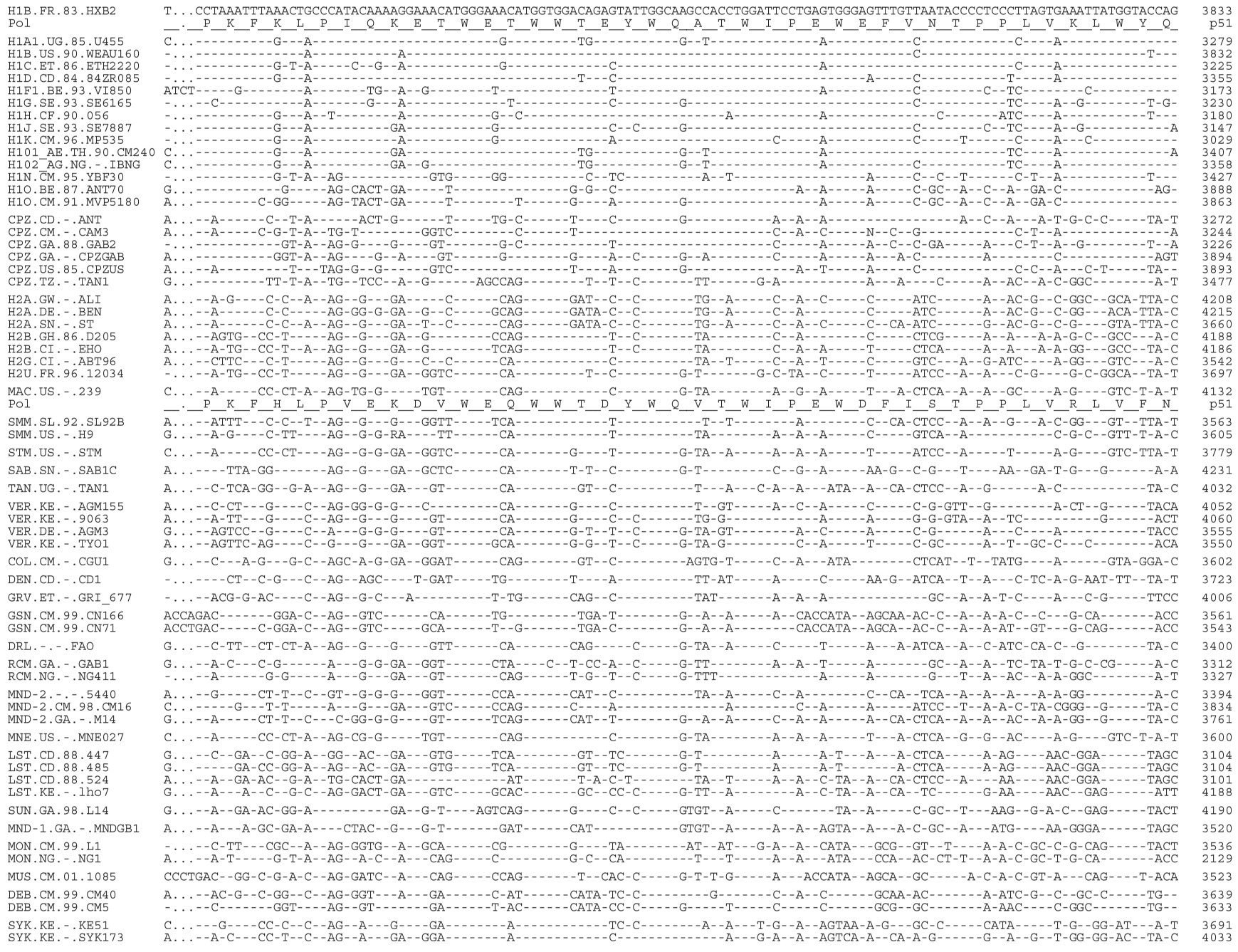
PLV
complete genomes

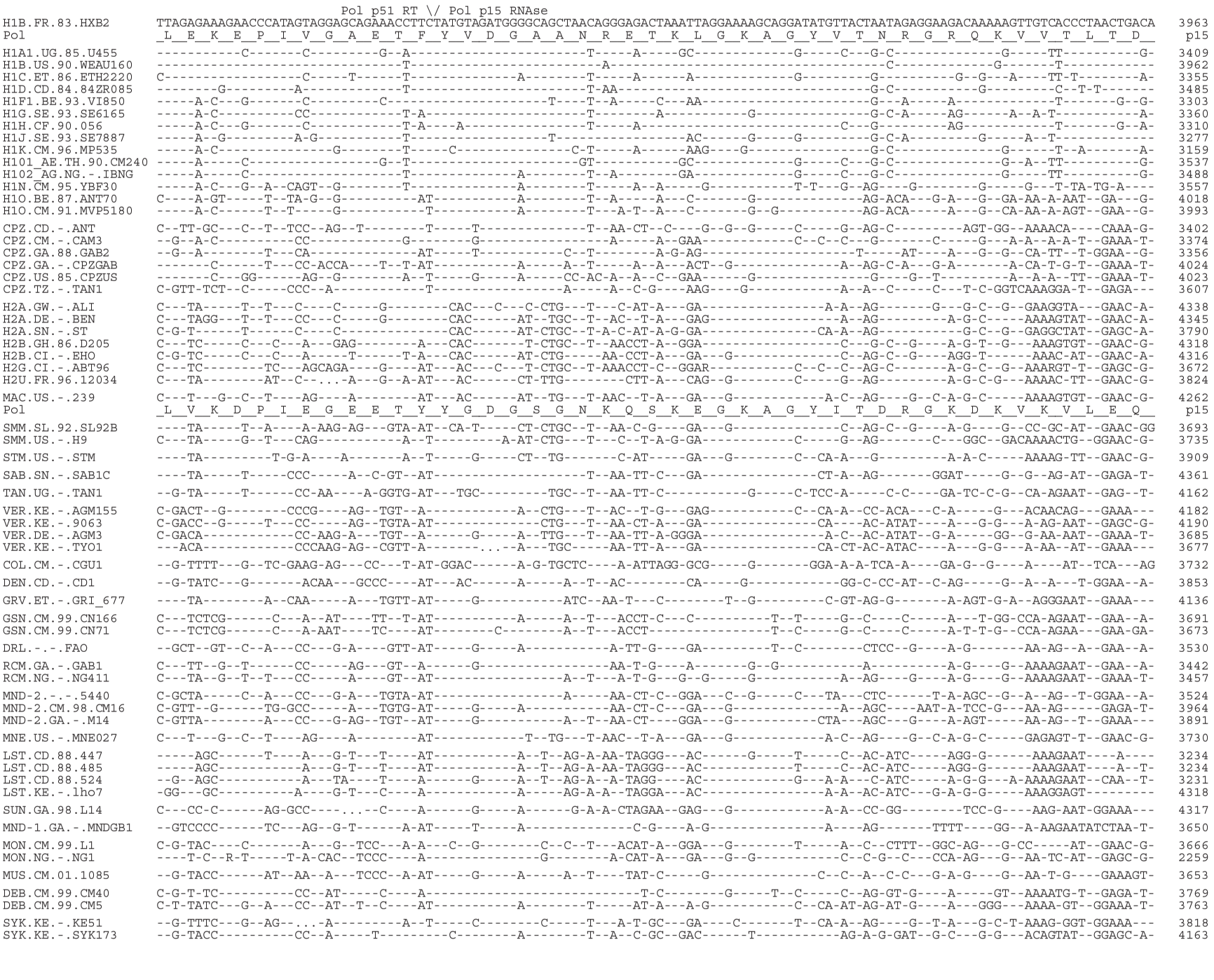


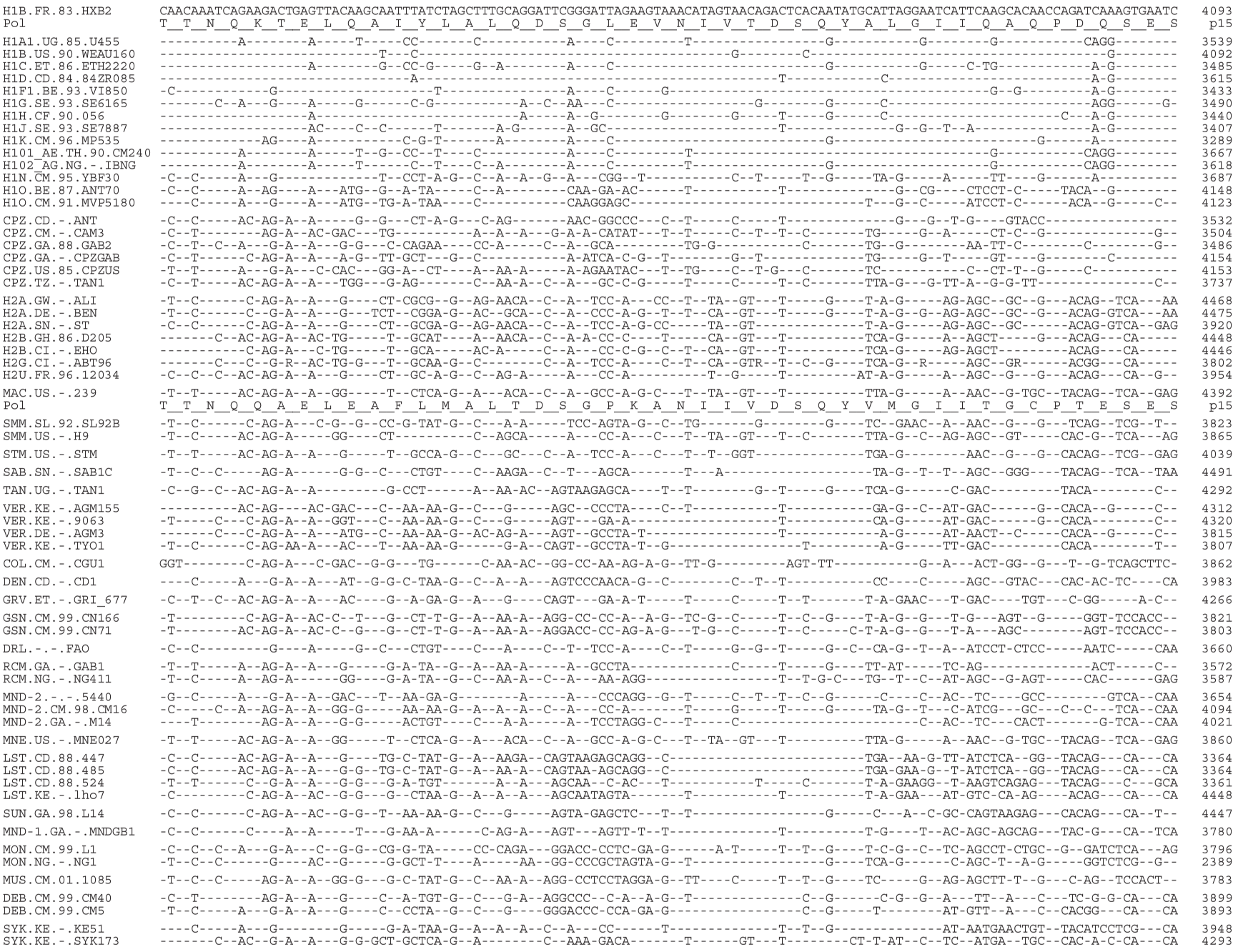
```
H1B.FR.83.HXB2      CTAGAACTGGCAGAAAACAGAGAGATTCTAAAAGAACCAGTACATGGAGTGTATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAG  3579
Pol                 _L__E__L__A__E__N__R__E__I__L__K__E__P__V__H__G__V__Y__Y__D__P__S__K__D__L__I__A__E__I__Q__K__Q__G__Q__G__Q__W__T__Y__Q__I__Y__Q__   p51

H1A1.UG.85.U455     T-----T-------G-----G-------------C--T--G--------A---------------------C--G---------------A--A-------A-------------------------  3025
H1B.US.90.WEAU160   -----G-------------G--A-------------G-------------------------------------------------GC-------------------------------G----------  3578
H1C.ET.86.ETH2220   T-----T-------G-----G--A--------G-----------------A-T-------------------------T----------A------A-T-A----------T----T----C----  2971
H1D.CD.84.84ZR085   T-------------------G----------G------A-G-------------------------------------------T-------A--A-------T-------------------------  3101
H1F1.BE.93.VI850    T-----T-----A----T--G----------G--------------G--A-------------------------------------A--A--AG-C--G-----------------------GA  2916
H1G.SE.93.SE6165    A-G--GT-------G-----G--A--------G------T-----------C-------------------A-----------G----A--A--A----T--A-------------------------  2976
H1H.CF.90.056       T-G---T-------------G--------G-G-------A----------A--------T-------------------------G------A-------------------------------G-  2926
H1J.SE.93.SE7887    T-G----------------A-------------------------G--A---------T--G-------A-----------G-G-----A--A----TG-A-----------------------G-  2893
H1K.CM.96.MP535     T-----T-A-----G-------------------G-----G-----G--A-------------------------------------A--A---A-T-A-----------------A-------  2775
H101_AE.TH.90.CM240 T----GT-------G-----G-----------ACC--T--G--------A-------------------------G--------G-------A--A-----G-A-------------------------  3153
H102_AG.NG.-.IBNG   T-----T-------G-----G--A-----------T--------G--A-----------A-----------G-------T-------A--A-------A-------------------------  3104
H1N.CM.95.YBF30     T-------A----------G-----AT----------CC-G--------C-----------GG------A---G----------T--A-----A--A-----T--G----------G--------G-  3173
H1O.BE.87.ANT70     --G---T-A-AG-----------A-GGT-----C-------G-----G--A--C---C-A--TGAT--G--TC--TGG-TTA-T--T--------A--AGGG-AG--------T--C--G--A-----G-  3634
H1O.CM.91.MVP5180   ------T-A-A------------A-AG--------G---------------A-----CC-G--TGAC--------GTGG-TTAGT--T--------T--AG----G--------T--C--GG-A-----G-  3609

CPZ.CD.-.ANT        T-----T-A-A------T-A-C-A------C-GC-GAA-A--G-G---TAC--C---C-G--TGGTCT-CCAC---A----ACC---------A--A---TC---A--------C--------A------A  3018
CPZ.CM.-.CAM3       A-G---T-A-A---G-----R------T---RG--T-----G-----G-----C---------GA---G--A------------N-------A-----AA-TA-T--G-----T----------TC--G-  2990
CPZ.GA.88.GAB2      T----GT-A-A------T--G--A---T-------------G-----CACC--C---------GA------AC-----------G-T--A------AATA-T--G--------T-----G----T--G-  2972
CPZ.GA.-.CPZGAB     T-----T-A-------T--G-----AG---GCAC-----------G--A--C-----T---GAC-----GC-T-----------------A-----CA-CT----G-----T-----G--A-T--G-  3640
CPZ.US.85.CPZUS     A-G--GT-A-A-----------A--------G-------------G--C-----------GA-----A---G--------G----A--A--A--AAGGA-T--G--------C-------T-----  3639
CPZ.TZ.-.TAN1       T-G--GT-A-A------T-AGC-------------AAG--T--A----CA--C-----T--TAA-TT-CCTC-GCA-----C-----------------------A--G----------G--A-----G-  3223

H2A.GW.-.ALI        GC------A-AG--G------ATT--CT---GTC-GGA-CA-G-G--GCAC-----CC-AGA-GA---G--G---GA----AC-G-C--A--AG-TCAAG-CAAT-----------A----AC-C--G-  3957
H2A.DE.-.BEN        GC---G--A-A--------A-ATT--CT-G-GCC-GGA-CA-G-A---TAT-----CC-AGA-GA------A---GAG---AC---C--A--AAGCCAAGG-CAT-----------CA----AC-C--G-  3964
H2A.SN.-.ST         GCG-----A-A--------A-ATC--CT---GCC-GGA-CA-G-A---TGC-----CC-AGAGGA---G--GC--GA----AC-G-C--A--AG-TCAAG-CAAT--G--------A-G--AC-C--G-  3409
H2B.GH.86.D205      GC---G--ACAG--G--T-A-ATC--CT--G--C-GGA-CA-G-A---TCC--C--CA-GGA-AGGGT-CCGC--GA----AC-G-------AA-CCTAGC-AAT--G--------CA-----C----G-  3937
H2B.CI.-.EHO        GC----T-CCAG-------A-ATC--C---G--C--GAGCAGG-A---TCC-----CA-AGA-GGGGT-CCT---GA----AC-G-G-----AA-TCTAGC-AAT--G--------CA-G---C----G-  3935
H2G.CI.-.ABT96      GC----T--CA--------AGATT--C---R--C-GGARCAGG-A--C-CC-----CA-AGA-GR-G-GCCT--RGA---CACTG-------AA-TTTAG-CAAT--G--------CA--G--C------  3291
H2U.FR.96.12034     GC----T-T-----G--T-AGATT---T-G-GTC-GGA-CA-G-G--G-CT-----CAGGGA-GA-G--A-TC--GA----AC-G-G-TT--AA-TCAAG-TAAT--------C--CA-G--CC----G-  3446

MAC.US.-.239        GC----TAT-AG-----T-A-ATA-----C-GTC-GGA-CA-G-A---TGT-----CC-AGA-GGC--GCCA---GA---CACGG--ATA---AGTCA-G-CAAT--G---T-T---A-----C-C----  3881
Pol                 _A__E__Y__E__E__N__K__I__I__L__S__Q__E__Q__E__G__C__Y__Y__Q__E__G__K__P__L__E__A__T__V__I__K__S__Q__D__N__Q__W__S__Y__K__I__H__Q__   p51

SMM.SL.92.SL92B     GC----T-------G--T---ATT--------TC--GAGCAGG-A---AGA-----CAGAGA-GATG--CCGC--GA----ACGG-T-TA---A-TCAAG-CAAT--G----GT---A-G---C------  3312
SMM.US.-.H9         GC----TAT-A--------AGATA----CC-GTC--GA-CC-G-G---TGT--C---C-GGAGGG---GCCAC--GAG---CC-G--ATA---AGTCA-G-TAAT------T-----A-----C-C----  3354

STM.US.-.STM        GC----TAT--------T-A-ATT-----C-GTC-GGAGCAGG-A--GCAA--C--CCGAGA-GAC--GCCTC--GAG---AC-G--GT---AG-TCA-G-CAAT------T-C---A-G--AC-C----  3528

SAB.SN.-.SAB1C      ------T-A-A------T---C----CT---G-C--AAGCAG--G---CAA-----------TG-TCTTCCT----GG--TA-GG-T-TT--A-T---AG-T---------GG------G--A--C--GC  3977

TAN.UG.-.TAN1       G----GT-CAG-C----TCA------CT--------GAGCA---G----CT--C---------AGT---TCT--G-G---CACT--TACC----T--ATG-T--G------GG-------T-CA-A--G-  3781

VER.KE.-.AGM155     GC---GTAC-A--------AG-----C------AC-GAGCA-G-A--TACT-------CA---GA----CC-C-T-GG----C-G-------ATTA--AG-T--G------T----C---T-CA-G--G-  3801
VER.KE.-.9063       GC---GTAT-----G----A---A---T-G---AC-GA-CA-G-A---ACT-------CA---GG---GCCTC-T-G-----C-G-C-----ATTA-AAG-T--G-------GT-----GT--A-A--G-  3809
VER.DE.-.AGM3       GC----TAT----------A------CT-----ACGGA-CAGG-A--GACC--C--CA-A---GG--G-CC-A-C-G-----C-G-G--A--A-TA-AAGG---T-------GT--C---T-CA-G----  3304
VER.KE.-.TYO1       GCT---TAT--------TGC---------T---AC-GA-CAGG-A---ACC-----CA-A---GG--T-CCTA-T-GG----C-G-------ATT--AAGG---A--G----GT--C---T-CA-A----  3299

COL.CM.-.CGU1       GAG--G--ACA---T----A-C----AA-C------GA---T--G---------C---ATA---CATG-G---A-TTGG-T---C--TTTC-G-TTA---A----A------GGC---GC-G-C-GC---A  3351

DEN.CD.-.CD1        GC---G--TAA----GCACA---A--AT--------AGTA-TTCA---AGT------AGG---GAG---CCT--G--------G--CTCT--ATTA--AG--AAT--G---GG----AC---CA-A--G-  3472

GRV.ET.-.GRI_677    GC----TATAAGA-C--TCA--G---AG-GC-G---A--CA-G-A---ACA--C--------TCTC-----A---------AC-G-T--A-----A--AG----G----------C---T-CACC----  3755

GSN.CM.99.CN166     A-----T-A-------T--G-----CT-----C--T-TTC-G-G--T-GC---------G--GA---GCCGC-CG-G-TG-----TACCTCC-TA--AG--CAG------GGG--CATGT-CAGC----  3307
GSN.CM.99.CN71      A----GT-A-----------------------C--T-CTC-G-G--CAGC--C--C---G--GA---GCCAC-TG-G-T------TACCTCC-TA--AG--CAG------GG---CATGT-CAG---G-  3289

DRL.-.-.FAO         ------TAT-A------T-A-CTA--AA-------G-A-A-G-----G--A------C-G---GA----CC-C---A---CA-G-----A--A-T---TA-T--A------T-------G--AG-A----  3146

RCM.GA.-.GAB1       A-----TAT-A--------AGAT-------------AAGT-G-----G----------TGA-AAG---CC----G-G---A-C--T------TTA-AAGGT--A--G----T-C--------AG-A--G-  3058
RCM.NG.-.NG411      T-----TAT-A--------A--CTG-C--G--G--GAA-A-------G--T------A-----GA----CCA--------TA--G-T-----ATTA-AAGG---G-------GC--C--G--AG-G--G-  3073

MND-2.-.-.5440      ------TAT-A--------AGCTA--AG-GC-G--GGA----------------C--CC-G---GA----CCT-----G---A-GG-G--A---TTAACA-----A--G----GT--C-----AG-A--G-  3140
MND-2.CM.98.CM16    T-----TAT-A--------A-CTC--AG-GC-G--GGA---G--------A--C---C-A---GAG--GCCT-----G---A-GG-T--A--ATT-ACA-----G-------GT--C-----AG-G--G-  3580
MND-2.GA.-.M14      T-----TAT-A------T-AGTTA--AG-GC-G--GGA-----------T------C-G---GG----CCTC-G--G--CA-GG------A--T-ACA--G--A-------GC-----G--AG-A--G-  3507

MNE.US.-.MNE027     GC----TAT-AG-----T-A-ATA-----C-GTC-GGA-CA-G-A---TGT-----CC-AGA-GGC--GCCA---GAG--CACGG--ATA---A-TCA-G-CAAT--G---T-T---A-G---C-C----  3349

LST.CD.88.447       GA----TAT-A-C-----CA-C----A---------AA-C-G--G----CT--------T--TA-T-----A--GG---T-AG-G-----C--AGTAA-A-G--AATTATT--C-TC--GTGGAGA----  2853
LST.CD.88.485       GA----TAT-A-C-----CA-C----A---------AA-C-G--G----CT----------TA-T-----A--GG---T-AG-G-----C--A-TAAAA-G--GGTCATT--C-TC--GTGGAGG----  2853
LST.CD.88.524       GA----TAT-AGACC--T-AGACA------------AA------G--G-CT--C-----T---AGT-----AC-TG---T-AGGG----AC--A-CAA-A----GGTTATA--T-TC---TGGAGG----  2850
LST.KE.-.lho7       GA----TAT-GGC-G----A----G-A--T------AAGA-G--G----CC----------GA---G--AC-T--T-TCAGGG-----C-AA-CAAAA-G--GAT-ATT--T-TC--GTGGAGA----  3937

SUN.GA.98.L14       GA----TACCAGAGT--TCAG---G-AT--C-G--GT-----TCA--CAGT--C-----A---GAT-----A-----TTGCAG-G-G--A---GTAAA------AAT--TA--C-T----TGG-TG--G-  3939

MND-1.GA.-.MNDGB1   --G---TAT-A-C----T-A------CT-GGCT---GA-CA-G-A--GTCT--------T--TAAT--G--A---TAT-T-AG-T-T-----AACAACAGG---AG-TATAT---T----TGGA-G----  3269

MON.CM.99.L1        GCT--G--A--G-----T--G--------C--G--GGA-TCCACC--G-GA-----C-----CAACG-GCCA----AG-TG--GT--ACCTCA-TA-CAG--AA-------GGC--CAG-T-C-TC----  3285
MON.NG.-.NG1        GC----T-A--------T--G--A-------GTC---AGTCGAC---G-GA-----C----CCAGCTC-----G-AG-T---G--TACCTCC-T--YAG----A--G---GGG--CRGGTA--TC--G-  1878

MUS.CM.01.1085      GC----T----------TCAG-----C--TC--C-GA----C--G--G-CA--C---------GA-C-GCCA---G-G--------TGTCTCA-TA--AG-CTCA--G---GG----A-CT-CACC--GA  3269

DEB.CM.99.CM40      GC------TAG---G----AG--A--C--G--GC--GA---GTCA-----T--C---A-AGAGCAGG--CCA------------T--TCC--ATTA---A----A------GG----GTG--CAGA---C  3388
DEB.CM.99.CM5       GC-------AGG--G--T-A---A-C--CC---C-GGA---CTCA--G--A--C--CA-AGA-GATG-GCC-C-----T----G--TTCC--AATA--AAG---A--G---GG----GT----AGA---A  3382

SYK.KE.-.KE51       GC----A----T-----T--------A--C----T-GA-CA-ACA--GTCT------C-G---GA----CCA---GA----C-C--TAGC---TTA--A---CAA--G---GGG--CAT----A-A----  3434
SYK.KE.-.SYK173     GC----T-A-A---G--T---C-A--CT----G---GA-CAGGCA--GTCT----------TAA----CCT---GA----C-C--CACT--A-TA---TCCCAG------GG---CATG---A-G----  3776
```

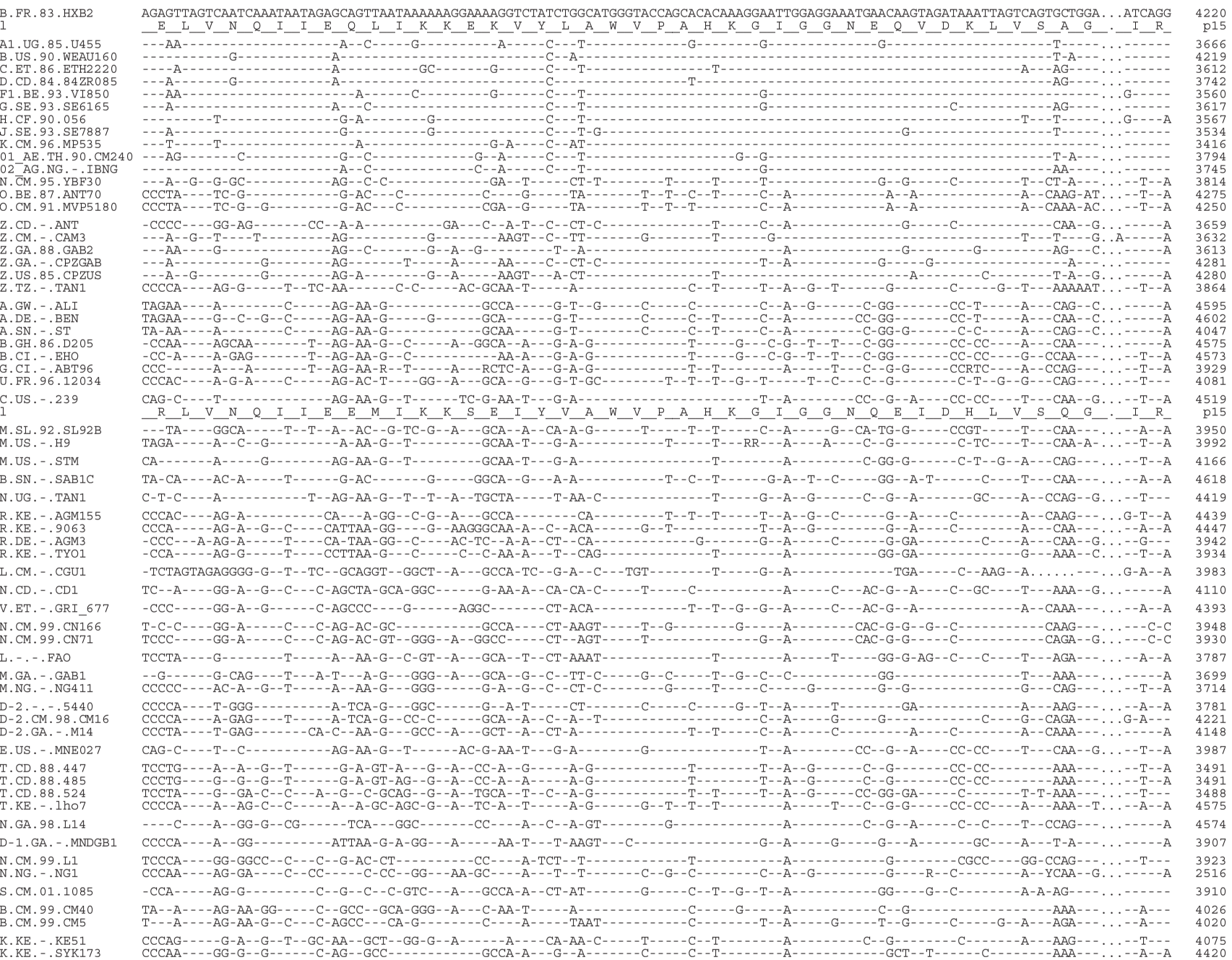
PLV
complete genomes

```
H1B.FR.83.HXB2         AGCCA...TTTAAAAATCTGAAAACAGGAAAATATGCAAGAATGAGGGGTGCCCACACTAATGATGTAAAACAATTAACAGAGGCAGTGCAAAAAATAACCACAGAAAGCATAGTAATATGGGGAAAGAC  3706
Pol                    E__P__.__F__K__N__L__K__T__G__K__Y__A__R__M__R__G__A__H__T__N__D__V__K__Q__L__T__E__A__V__Q__K__I__T__T__E__S__I__V__I__W__G__K__T   p51

H1A1.UG.85.U455        -----...---------------------------------AA---TC---T---------------------------A-TG---------G-GT----------------------------T  3152
H1B.US.90.WEAU160      -----...----------------------G---------G----------------------------------------------G--------------------------------------  3705
H1C.ET.86.ETH2220      -A---...--C-----------------G--G-T-----A--GAG--AC-------------------G--G-------C--T---------G--G--CTG-------------------------  3098
H1D.CD.84.84ZR085      -A---...----------------------G--------------------------------------G-------------------------G---T-----------G--------------  3228
H1F1.BE.93.VI850       -C---...-----------A----------G-------A-G-----TCG-------------------------------A-----A-----G---G-TTT---------------------G----G  3043
H1G.SE.93.SE6165       -----...-AC-------------------G-------A--G-G--TC---------------------G--------A-T-------------G-------G------------------------  3103
H1H.CF.90.056          -----...-------------G---------------A-----AAC------------------A----------------A-------------G---T-T-------------------------A-T  3053
H1J.SE.93.SE7887       -----...--------C-----------G---------A--G----A-------------------------G----A-TG-----------G--TTG---GC-----------------A--  3020
H1K.CM.96.MP535        -----...CA------------------G--G-------------ATC-------------------G-----------A-------------G--------G---------------------A--  2902
H101_AE.TH.90.CM240    -----...-----------A----------------C----GAG--TC---T----------------G------------A-TG--------G--G-------------------------------  3280
H102_AG.NG.-.IBNG      -----...---------T-A------------------A--A----TC-------------------G-----------T---------G--G-T-TG----------------------------  3231
H1N.CM.95.YBF30        --TT-...CA-------T-A----------G-------A-----ATC------T--------A-------G---GTT--A-TG--AAGG---G-GG-A--------T--------T----------  3300
H1O.BE.87.ANT70        -AGA-...CA---G--C--C--------G------A-T--GCAA-A--CCT-------A------A---G-------G----A-T-A-C--G--GG-GT-TCA----TCT---A-T--C--------ATT  3761
H1O.CM.91.MVP5180      -TGA-...CA---G--C--T----------------T--GCAA-A--CCT-------A------A---G------GG----A-T---C--G--GG-GT-TCA---GCT-----T--------G--ATT  3736

CPZ.CD.-.ANT           -TGA-...GGA---CTGT-A---G-------------T--GCCT-CA--AA-T-----------G--T-GG------G-T-GA-T-------------GGATT------T-----T--C------G--GT  3145
CPZ.CM.-.CAM3          -A-AG...CA---G--C--A--------C--------T--GCAA--ATCA--A--T---------CA---G-------G------T---A---------G-T--T-----T--------C-------A--  3117
CPZ.GA.88.GAB2         -AA--...GGA-GG-----A-----------------C---CAA--A-CA--T-CT--------CA---G----C--G-T--A-TG----C----G-TT--C-T-----T---A-CC-T-------CTGT  3099
CPZ.GA.-.CPZGAB        -A---...CA---G---T-----G-------------C---CAA---TCA--A-----A-----CA-C-GG---C-GG-G--A-----A--------TG-T--T-------------T--------A--  3767
CPZ.US.85.CPZUS        -A-GG...CA---G-----A--G--------------C---CAA--ATCA--A--T---------A-C-G---GC--GTT--A-TG-----------G-T--T--G-----T--C--T--------GT  3766
CPZ.TZ.-.TAN1          -AGA-...GGG------T-A-------------------A-TCACCA---A-------C-----GA---G-------G---GACTGA-A--G------GG--AT--G------A-----T-----T-TTGT  3350

H2A.GW.-.ALI           GA......GAA----T---A---GTG-A---G-------A-----AAAA-A----T--C--C-GG--C-G-TTG---G--C---T---T---------GGA-A----GCAC-G--C--T------CGA-T  4081
H2A.DE.-.BEN           -A......GAG----TC--A---GT------G-------AG--A-AAAA-A----T--C----GG--C-G-TT-C--G--C---T---T--G------GGA-A---GGCAC----C--T------CG--T  4088
H2A.SN.-.ST            GA......GGA----T---A---GT-------------AGG-A-AAAA-A-------C--C-GA--C-G--TCC--G--C-A-T---T---------GGA-A----GCAC----C--T------CGA-T  3533
H2B.GH.86.D205         GA......AA---GTC--A---GT-------------AGG-T-AAAACA-G------C--C-GG---G-T-C-GG--C-T-T---T-G-------GG--A----GC-C----C--C------G---T  4061
H2B.CI.-.EHO           GA......GA-----TC--A---GT-------------AGG-T-AAAACA-T------C----GA----G--T---GG-TC-T-T---C----------GGA-AG---GCAT-G--C--C------G---T  4059
H2G.CI.-.ABT96         GA......AAC-G--TC--A---GT---T----------AG--A-AAAA-A----T--A----GA----G-TT---GG-TC-T-TG--C------R---GGA-AG---GCAC---------------ATT  3415
H2U.FR.96.12034        GT......GAC-GG-CAT-A---GT---------T---C-AGG-A-A-AA-A-T--T--A----GA--TCGC-TGC--G--C-T-TG--A--G--G---GGA-A----GC----------------A-T  3570

MAC.US.-.239           -A......GAC----TA------GT---------T-----AG--A-A-AA-A-A--T--C----GA--G-G--T-----G--C-T-T-A-A--G------GGA-AG---GCA-----G--C------C--GT  4005
Pol                    E__.__.__D__K__I__L__K__V__G__K__F__A__K__I__K__N__T__H__T__N__G__V__R__L__L__A__H__V__I__Q__K__I__G__K__E__A__I__V__I__W__G__Q__V   p51

SMM.SL.92.SL92B        GT......GAC-GG-T---C--GGT---G--G-T-----A---A-AAAA-A-A--T--A----GAA---GGTT-C-GG-TA-T-T---T-----G---GGA-A------TC----G-------------  3436
SMM.US.-.H9            --......GAC----TA--A---GT---C----T-----AG--T-AAAA-A-A--T--A----GA--C-GCTT----G--C-T-T-------G------GGA-A----GCA---A----T------C--GT  3478

STM.US.-.STM           -A......GAC----TAT------GT---C--G-T-----A---A-AAAACA-A--T--A----GA--T-G-TT---GG--C-T-TGA-T--G--G---GGG-A----GCA-----G--C------C---T  3652

SAB.SN.-.SAB1C         CAGA-...AA-----TA-----GGTT--C----------A---C-AAACA--A------C-----AT---G-ATG---G---GCCT---A--G--G---GGA-AG-----------C--C------C---T  4104

TAN.UG.-.TAN1          --......GGA---GTA--A---------------A---AG-GC-AAAA-A-T--T--A-----GT-T-G-GT---GG---GAATG--C----GGG-GTG--AG---GCG--CACCT-T--------ATT  3905

VER.KE.-.AGM155        -A......GGA----TCT-A--GGT---G--G-TC--C-A-CA--AA-C-A-T-----C-----GT-GCGTGT-C--G---GA-T---A--G------GGG-A---GGC-C-------T------C-ATT  3925
VER.KE.-.9063          -A......GGA----TCT-----GT------G-------A-CAA-A-AACA----T--A-----AC-TCGCAC---GG-C-GCCT---A--------CTGT-A---GGCAC-G--C--T------G-ACT  3933
VER.DE.-.AGM3          --......GGAC--GTAT-A---GT---T-----CA---AGCA--AAAACA-T--T--C-----GT-CCGTGT---GG---GATT---A------C-TTGT-A----TCTT----T----------G--TT  3428
VER.KE.-.TYO1          -A......GGAC--GTCT-----GT----------CA-C-AGCAA-A-AACA----T--A-----AC-TCGCAC----G-T-GTTT------G--G--TTG--A----GCTC----T--T-----G-TATT  3423

COL.CM.-.CGU1          -A......CA-GG--CG--A-G--G--------C--AAC-C-GGC-AAAA-CAG---TT----AG-A-GC-GG-G---G-TTCA-TTA-A------G---GGT-GG---GC-------C-CC----------GT  3475

DEN.CD.-.CD1           -A......AAAC-TCTA--C---------C------CCAT-AGGAC-AA--AAG---TTA---CCCATATC---------G-CA-A-TGA--GC-------GGA-G----GC-T-------C------C-GACT  3596

GRV.ET.-.GRI_677       -A......GGGGC-GTAT-A--GGTG----G-------C-AGCAA--A-AAA-T--T-----------C---GGACTC--G--C-CCTT--C---------CTGT-AG---GCAC-TACC--T--------GACT  3879

GSN.CM.99.CN166        -AGGC......-G-TGT-A---T---CC----TC----AG-CC--AAA---A----TA--CTCCTATC-G------TG----T--TT--ACC---G--GGT-GG--G--TCAG--T-CC----------GT  3431
GSN.CM.99.CN71         -CAGC......-G-TG--A-G-T----C-----T-----A--CC--AAC---A--T-TA--CTC-TATC-G------GG----T--TA--ACC---G--GGT-GG-----TCAG--T-CT---------AGT  3413

DRL.-.-.FAO            -TGAT...AAC--GCCA--A----------------T-AGG-A-A-AA-AGT------C-------A-G-G-ATG---G---GACT---A------G---G-A-A---GGCAC-G------C---------G-TT  3273

RCM.GA.-.GAB1          --T--...GGA---CCA-----------C---------T-AGCA--AAACA--A------C-----AA-T-G-ATGC-GG---G-TT---A--------TG-A-A----GC------T--T-------G-CT  3185
RCM.NG.-.NG411         GAGAC...-CA---CC---A--G-----G--G--------A-CAA-A-ACA--A--T--A-----AA-T-G-ATGC-GG---GCTT---A--G-------TG--A----GCT---------T----------CT  3200

MND-2.-.-.5440         -AGAC...AAC---CC---C--GGT------G------C--G-CA-A-AA-----------A------AT-G-GGGT-C-TG---G-TTG--A------GG--G-T-AG--GT-TC--------C---------CT  3267
MND-2.CM.98.CM16       --GAT...AA---GCCG-----GGT---G---------C----CA-A-AA---T--T--A-----GC-T-G-GTGC-TG---GACT---A--G--G--TG-A-A----GCTT----G--C---------ACT  3707
MND-2.GA.-.M14         -AGAG...AA---GCC---T--GGT---G--G------T--G-CA-AACA-----------A------GT---GGGTGC-TG---GATT---A--------G---A---GT-------G--T------G-ATT  3634

MNE.US.-.MNE027        -A......GAC----TA------GT---------T-----AG--A-AAAA-A-A--T--C----GA--T-G--T-----G--C-T-T-A-A--G------GGA-AG---GCA-----G--C------C--GT  3473

LST.CD.88.447          GA......AA-TCT-T-----G-G-------GG---CA-----CAA-AATCA--T-----A---CCCC--C--A----GG-----------A----------GGA-A---G--T------C--T------TC-TGT  2977
LST.CD.88.485          GA......AA-TCT-T-----G-G-------GG---CA----GCAA-AA-CA--T-----A---CCCC--C--A-G--GG-----------A----------GGA-A---G--T------C--T------TC-TGT  2977
LST.CD.88.524          GA......AA-GCC-T---A-G-G------G----CA-----CAA-A-A-----T----TG----CCCC-TC-G-------G-----A--T--A--G-------GGA-A-------------C--T------TTTTGT  2974
LST.KE.-.lho7          GA......AA---C-TCT-A-G-G-T--G-GG---CA-----CA--A--CA--A-----A---CCCC--C-GA-----GT----A--TA-T--G--G---GGA-A------------C--C------CTTTGT  4061

SUN.GA.98.L14          GA......AAAC--GTAT-A-GGGT-----G-----CA--AG-GAG---CA--A--TGAG---CC-TGTC-----G---G---CT---CC-A------G---GGA-G-------------T----------TTTGT  4063

MND-1.GA.-.MNDGB1      GA......AA---GGT-T-A-G-G----G--------GG-A-CA--AAACA--A--T-G-------CC-C-TGA----GG---GT--TAC---G--GG--GGA-G-------T----------C------TTTTGT  3393

MON.CM.99.L1           -A......AAATC-GTAT-----G-----------T---C-A---T--AA--A------T-C---ACCTATC-G--G---G-----T---GC-AGCC---T-GGGA-A---GGC-------C-C--------CGACT  3409
MON.NG.-.NG1           -A......RA--GGGT---A--G-----C-----TC-----AG--T-AAAA-------T-C---CACATACC-G------GG-----T--T--AACC---C--GGG-AG---GGC-T-G--C--T---------CT  2002

MUS.CM.01.1085         GC......AAGGGGGTAT-A-GG-----------T---C-A-G-A----TCA-TA--------AGCTATC-G------GG-----C---A--GCC-------GGA-G--------C------C--T------G---C-  3393

DEB.CM.99.CM40         CT......AAGGG--TAT-A--G-----------TCT-T---GAT-AA--A--T---TA--------T-CC-T-------G-TA-----TA--T-T-------GGA-------------TT-T--------GA-T  3512
DEB.CM.99.CM5          CA......AAGGG-TTG--A---------G-----TCT---AGGAC-A---AA-T--TTA-------CT-TC-T---C--G-TA------A--T-T-------GG-------G--T------CTAT---------AGT  3506

SYK.KE.-.KE51          G----CAAGAA---CC------TC----------ACA-G--A-GCATAT-CGA----TCAC-------TACC-GGC---GG-TC--CTGA--A-T-------GGA-T-C--GC-C--TGG--T---------A-T  3564
SYK.KE.-.SYK173        ---A-AAAGGCCC-CCC--A-T-----------GACA-----A--CCTTT-CA--A--T-G-------CTATC--TCTC-TG-TC-ATTGT-AA-T-------GGG-TCC--TCAC--TGGTAT----------GT  3906
```

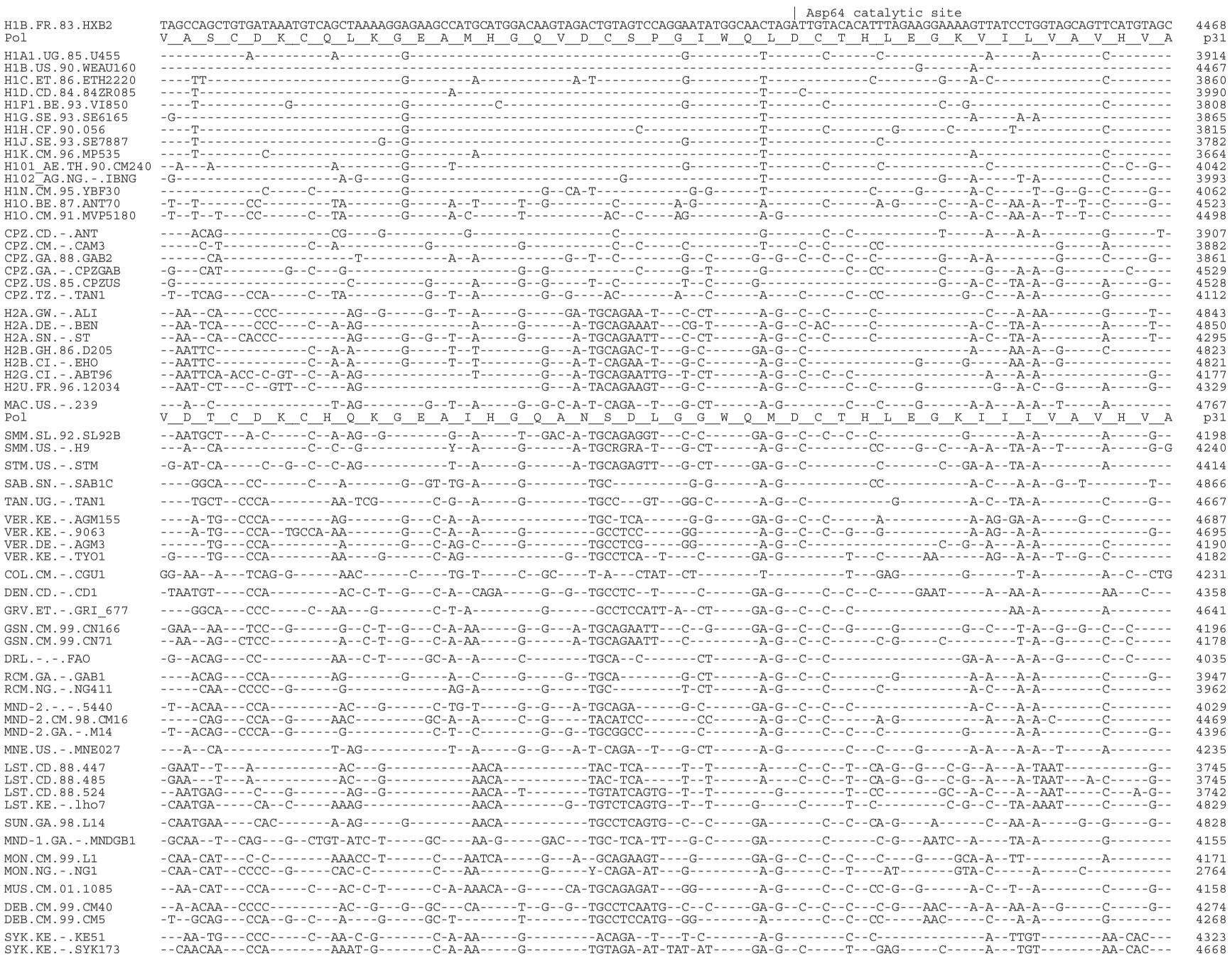

H1B.FR.83.HXB2 3833
Pol p51
H1A1.UG.85.U455 3279
H1B.US.90.WEAU160 3832
H1C.ET.86.ETH2220 3225
H1D.CD.84.84ZR085 3355
H1F1.BE.93.VI850 3173
H1G.SE.93.SE6165 3230
H1H.CF.90.056 3180
H1J.SE.93.SE7887 3147
H1K.CM.96.MP535 3029
H101_AE.TH.90.CM240 3407
H102_AG.NG.-.IBNG 3358
H1N.CM.95.YBF30 3427
H1O.BE.87.ANT70 3888
H1O.CM.91.MVP5180 3863
CPZ.CD.-.ANT 3272
CPZ.CM.-.CAM3 3244
CPZ.GA.88.GAB2 3226
CPZ.GA.-.CPZGAB 3894
CPZ.US.85.CPZUS 3893
CPZ.TZ.-.TAN1 3477
H2A.GW.-.ALI 4208
H2A.DE.-.BEN 4215
H2A.SN.-.ST 3660
H2B.GH.86.D205 4188
H2B.CI.-.EHO 4186
H2G.CI.-.ABT96 3542
H2U.FR.96.12034 3697
MAC.US.-.239 4132
Pol p51
SMM.SL.92.SL92B 3563
SMM.US.-.H9 3605
STM.US.-.STM 3779
SAB.SN.-.SAB1C 4231
TAN.UG.-.TAN1 4032
VER.KE.-.AGM155 4052
VER.KE.-.9063 4060
VER.DE.-.AGM3 3555
VER.KE.-.TYO1 3550
COL.CM.-.CGU1 3602
DEN.CD.-.CD1 3723
GRV.ET.-.GRI_677 4006
GSN.CM.99.CN166 3561
GSN.CM.99.CN71 3543
DRL.-.-.FAO 3400
RCM.GA.-.GAB1 3312
RCM.NG.-.NG411 3327
MND-2.-.-.5440 3394
MND-2.CM.98.CM16 3834
MND-2.GA.-.M14 3761
MNE.US.-.MNE027 3600
LST.CD.88.447 3104
LST.CD.88.485 3104
LST.CD.88.524 3101
LST.KE.-.lho7 4188
SUN.GA.98.L14 4190
MND-1.GA.-.MNDGB1 3520
MON.CM.99.L1 3536
MON.NG.-.NG1 2129
MUS.CM.01.1085 3523
DEB.CM.99.CM40 3639
DEB.CM.99.CM5 3633
SYK.KE.-.KE51 3691
SYK.KE.-.SYK173 4033
PLV
complete genomes

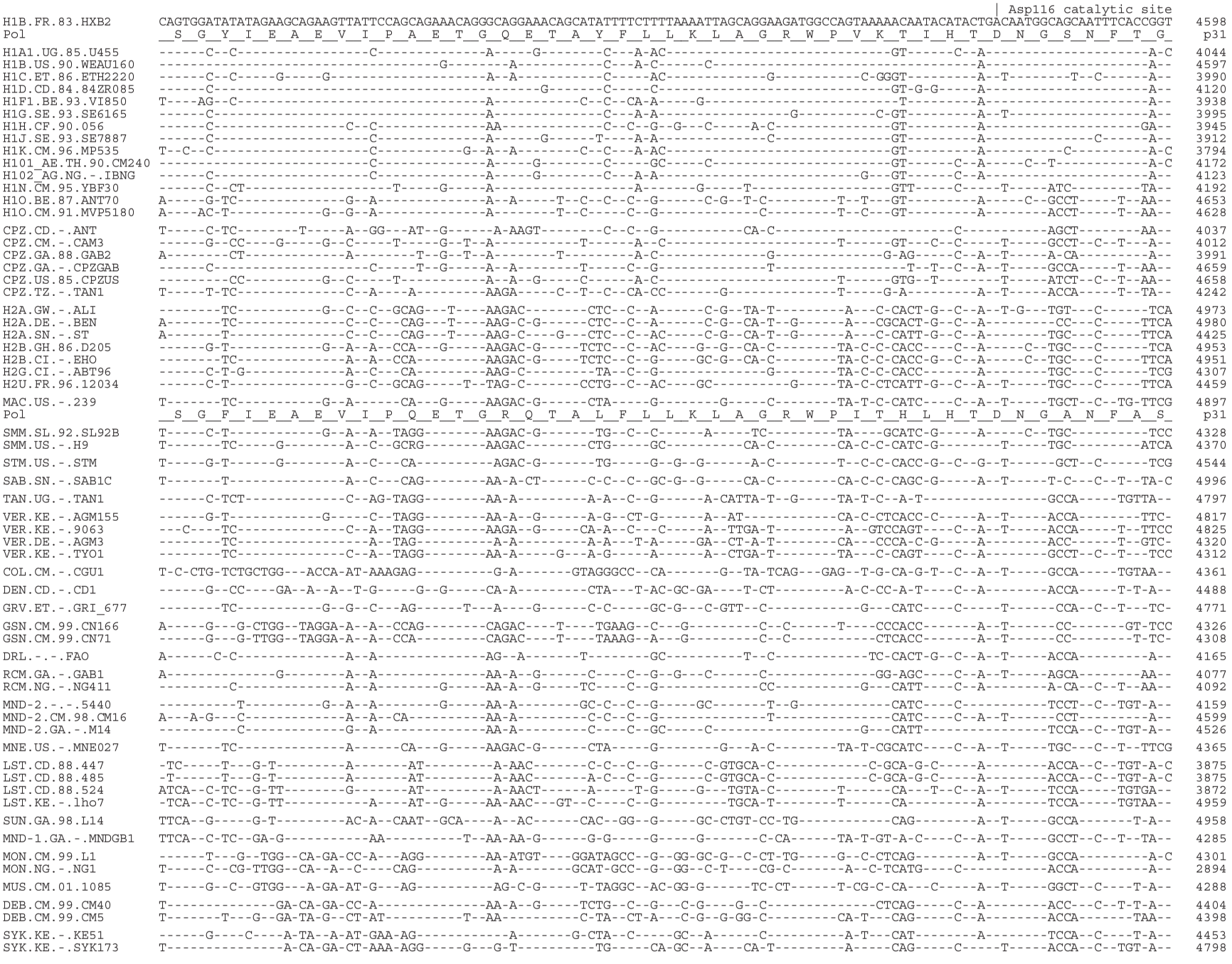

Pol p51 RT \/ Pol p15 RNAse
H1B.FR.83.HXB2 TTAGAGAAAGAACCCATAGTAGGAGCAGAAACCTTCTATGTAGATGGGGCAGCTAACAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACTAATAGAGGAAGACAAAAAGTTGTCACCCTAACTGACA 3963
Pol _L__E__K__E__P__I__V__G__A__E__T__F__Y__V__D__G__A__A__N__R__E__T__K__L__G__K__A__G__Y__V__T__N__R__G__R__Q__K__V__V__T__L__T__D__ p15
H1A1.UG.85.U455 3409
H1B.US.90.WEAU160 3962
H1C.ET.86.ETH2220 3355
H1D.CD.84.84ZR085 3485
H1F1.BE.93.VI850 3303
H1G.SE.93.SE6165 3360
H1H.CF.90.056 3310
H1J.SE.93.SE7887 3277
H1K.CM.96.MP535 3159
H101_AE.TH.90.CM240 3537
H102_AG.NG.-.IBNG 3488
H1N.CM.95.YBF30 3557
H1O.BE.87.ANT70 4018
H1O.CM.91.MVP5180 3993
CPZ.CD.-.ANT 3402
CPZ.CM.-.CAM3 3374
CPZ.GA.88.GAB2 3356
CPZ.GA.-.CPZGAB 4024
CPZ.US.85.CPZUS 4023
CPZ.TZ.-.TAN1 3607
H2A.GW.-.ALI 4338
H2A.DE.-.BEN 4345
H2A.SN.-.ST 3790
H2B.GH.86.D205 4318
H2B.CI.-.EHO 4316
H2G.CI.-.ABT96 3672
H2U.FR.96.12034 3824
MAC.US.-.239 4262
Pol _L__V__K__D__P__I__E__G__E__E__T__Y__Y__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__ p15
SMM.SL.92.SL92B 3693
SMM.US.-.H9 3735
STM.US.-.STM 3909
SAB.SN.-.SAB1C 4361
TAN.UG.-.TAN1 4162
VER.KE.-.AGM155 4182
VER.KE.-.9063 4190
VER.DE.-.AGM3 3685
VER.KE.-.TYO1 3677
COL.CM.-.CGU1 3732
DEN.CD.-.CD1 3853
GRV.ET.-.GRI_677 4136
GSN.CM.99.CN166 3691
GSN.CM.99.CN71 3673
DRL.-.-.FAO 3530
RCM.GA.-.GAB1 3442
RCM.NG.-.NG411 3457
MND-2.-.-.5440 3524
MND-2.CM.98.CM16 3964
MND-2.GA.-.M14 3891
MNE.US.-.MNE027 3730
LST.CD.88.447 3234
LST.CD.88.485 3234
LST.CD.88.524 3231
LST.KE.-.lho7 4318
SUN.GA.98.L14 4317
MND-1.GA.-.MNDGB1 3650
MON.CM.99.L1 3666
MON.NG.-.NG1 2259
MUS.CM.01.1085 3653
DEB.CM.99.CM40 3769
DEB.CM.99.CM5 3763
SYK.KE.-.KE51 3818
SYK.KE.-.SYK173 4163


H1B.FR.83.HXB2 CAACAAATCAGAAGACTGAGTTACAAGCAATTTATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCACAACCAGATCAAAGTGAATC 4093
Pol T__T__N__Q__K__T__E__L__Q__A__I__Y__L__A__L__Q__D__S__G__L__E__V__N__I__V__T__D__S__Q__Y__A__L__G__I__I__Q__A__Q__P__D__Q__S__E__S p15
H1A1.UG.85.U455 4539
H1B.US.90.WEAU160 4092
H1C.ET.86.ETH2220 3485
H1D.CD.84.84ZR085 3615
H1F1.BE.93.VI850 3433
H1G.SE.93.SE6165 3490
H1H.CF.90.056 3440
H1J.SE.93.SE7887 3407
H1K.CM.96.MP535 3289
H101_AE.TH.90.CM240 3667
H102_AG.NG.-.IBNG 3618
H1N.CM.95.YBF30 3687
H1O.BE.87.ANT70 4148
H1O.CM.91.MVP5180 4123
CPZ.CD.-.ANT 3532
CPZ.CM.-.CAM3 3504
CPZ.GA.88.GAB2 3486
CPZ.GA.-.CPZGAB 4154
CPZ.US.85.CPZUS 4153
CPZ.TZ.-.TAN1 3737
H2A.GW.-.ALI 4468
H2A.DE.-.BEN 4475
H2A.SN.-.ST 3920
H2B.GH.86.D205 4448
H2B.CI.-.EHO 4446
H2G.CI.-.ABT96 3802
H2U.FR.96.12034 3954
MAC.US.-.239 4392
Pol T__T__N__Q__Q__A__E__L__E__A__F__L__M__A__L__T__D__S__G__P__K__A__N__I__I__V__D__S__Q__Y__V__M__G__I__I__T__G__C__P__T__E__S__E__S p15
SMM.SL.92.SL92B 3823
SMM.US.-.H9 3865
STM.US.-.STM 4039
SAB.SN.-.SAB1C 4491
TAN.UG.-.TAN1 4292
VER.KE.-.AGM155 4312
VER.KE.-.9063 4320
VER.DE.-.AGM3 3815
VER.KE.-.TYO1 3807
COL.CM.-.CGU1 3862
DEN.CD.-.CD1 3983
GRV.ET.-.GRI_677 4266
GSN.CM.99.CN166 3821
GSN.CM.99.CN71 3803
DRL.-.-.FAO 3660
RCM.GA.-.GAB1 3572
RCM.NG.-.NG411 3587
MND-2.-.-.5440 3654
MND-2.CM.98.CM16 4094
MND-2.GA.-.M14 4021
MNE.US.-.MNE027 3860
LST.CD.88.447 3364
LST.CD.88.485 3364
LST.CD.88.524 3361
LST.KE.-.lho7 4448
SUN.GA.98.L14 4447
MND-1.GA.-.MNDGB1 3780
MON.CM.99.L1 3796
MON.NG.-.NG1 2389
MUS.CM.01.1085 3783
DEB.CM.99.CM40 3899
DEB.CM.99.CM5 3893
SYK.KE.-.KE51 3948
SYK.KE.-.SYK173 4293
PLV
complete genomes

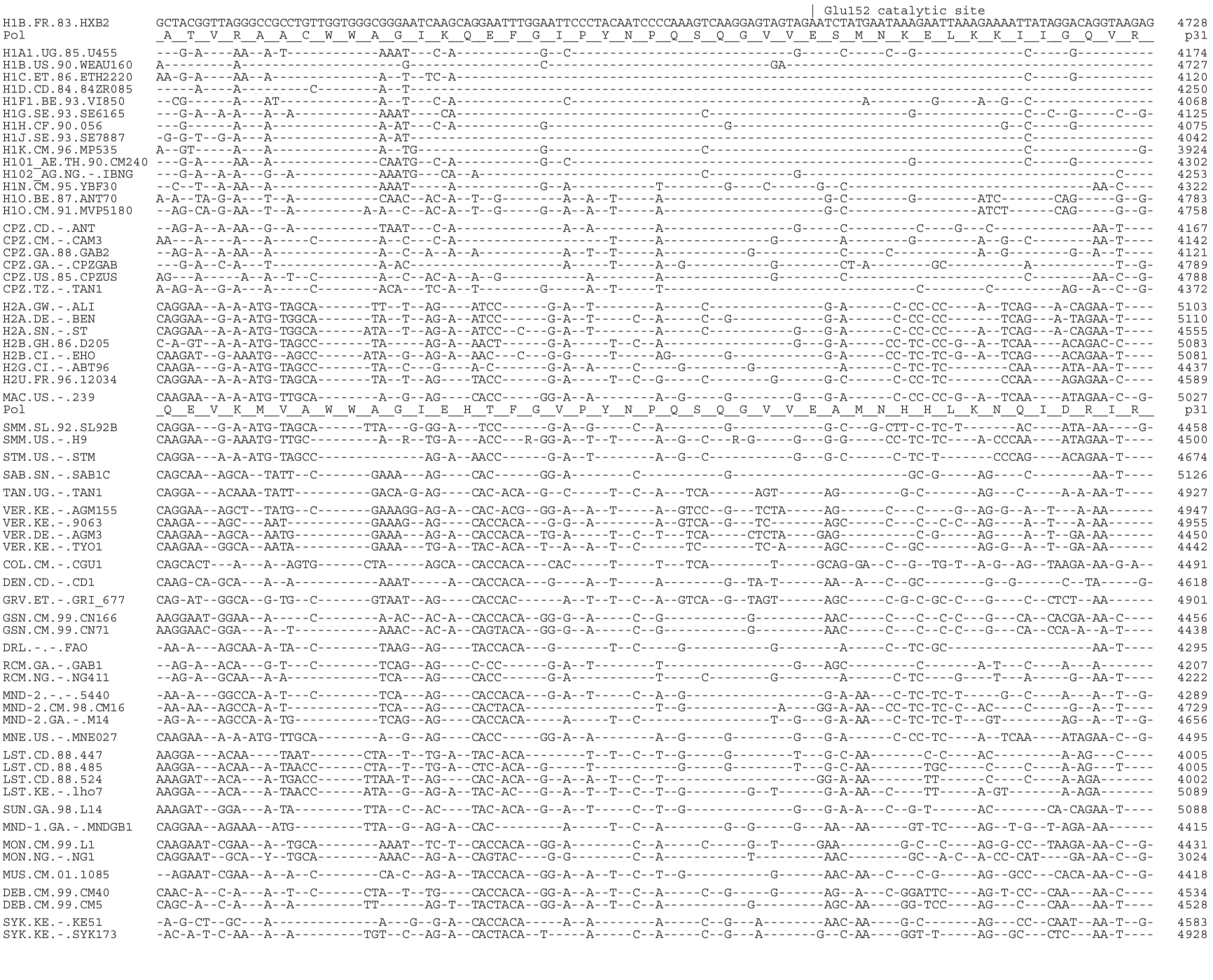

H1B.FR.83.HXB2 AGAGTTAGTCAATCAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTATCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGA...ATCAGG 4220
Pol __E__L__V__N__Q__I__I__E__Q__L__I__K__K__E__K__V__Y__L__A__W__V__P__A__H__K__G__I__G__G__N__E__Q__V__D__K__L__V__S__A__G__._I__R_ p15
H1A1.UG.85.U455 3666
H1B.US.90.WEAU160 4219
H1C.ET.86.ETH2220 3612
H1D.CD.84.84ZR085 3742
H1F1.BE.93.VI850 3560
H1G.SE.93.SE6165 3617
H1H.CF.90.056 3567
H1J.SE.93.SE7887 3534
H1K.CM.96.MP535 3416
H101_AE.TH.90.CM240 3794
H102_AG.NG.-.IBNG 3745
H1N.CM.95.YBF30 3814
H1O.BE.87.ANT70 4275
H1O.CM.91.MVP5180 4250
CPZ.CD.-.ANT 3659
CPZ.CM.-.CAM3 3632
CPZ.GA.88.GAB2 3613
CPZ.GA.-.CPZGAB 4281
CPZ.US.85.CPZUS 4280
CPZ.TZ.-.TAN1 3864
H2A.GW.-.ALI 4595
H2A.DE.-.BEN 4602
H2A.SN.-.ST 4047
H2B.GH.86.D205 4575
H2B.CI.-.EHO 4573
H2G.CI.-.ABT96 3929
H2U.FR.96.12034 4081
MAC.US.-.239 4519
Pol __R__L__V__N__Q__I__I__E__E__M__I__K__K__S__E__I__Y__V__A__W__V__P__A__H__K__G__I__G__G__N__Q__E__I__D__H__L__V__S__Q__G__._I__R_ p15
SMM.SL.92.SL92B 3950
SMM.US.-.H9 3992
STM.US.-.STM 4166
SAB.SN.-.SAB1C 4618
TAN.UG.-.TAN1 4419
VER.KE.-.AGM155 4439
VER.KE.-.9063 4447
VER.DE.-.AGM3 3942
VER.KE.-.TYO1 3934
COL.CM.-.CGU1 3983
DEN.CD.-.CD1 4110
GRV.ET.-.GRI_677 4393
GSN.CM.99.CN166 3948
GSN.CM.99.CN71 3930
DRL.-.-.FAO 3787
RCM.GA.-.GAB1 3699
RCM.NG.-.NG411 3714
MND-2.-.-.5440 3781
MND-2.CM.98.CM16 4221
MND-2.GA.-.M14 4148
MNE.US.-.MNE027 3987
LST.CD.88.447 3491
LST.CD.88.485 3491
LST.CD.88.524 3488
LST.KE.-.lho7 4575
SUN.GA.98.L14 4574
MND-1.GA.-.MNDGB1 3907
MON.CM.99.L1 3923
MON.NG.-.NG1 2516
MUS.CM.01.1085 3910
DEB.CM.99.CM40 4026
DEB.CM.99.CM5 4020
SYK.KE.-.KE51 4075
SYK.KE.-.SYK173 4420

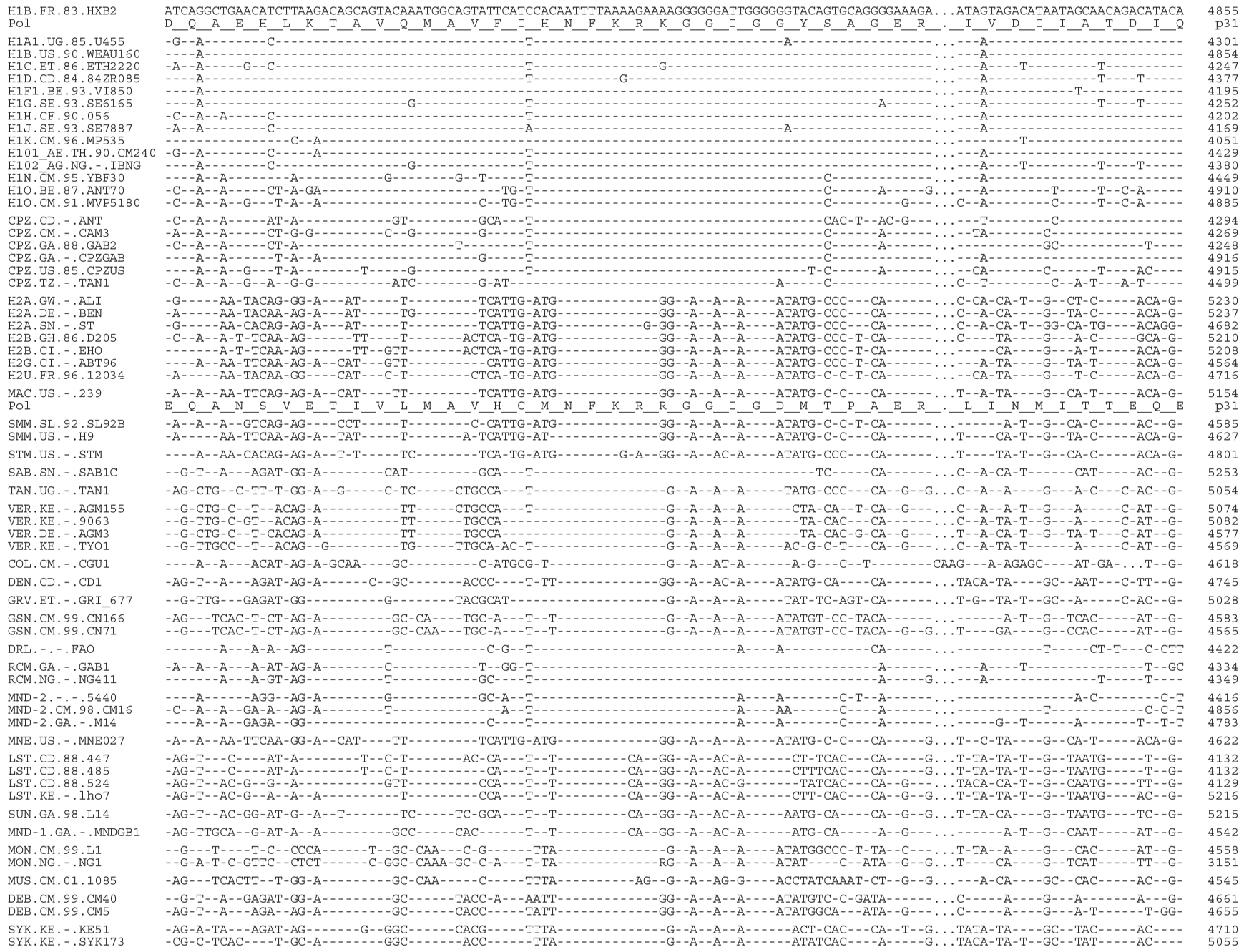
```
 Pol p66 RT & Pol p15 RNAse H end \/ Pol p31 Integrase start
H1B.FR.83.HXB2        ......AAAGTACTATTTTTAGATGGAATAGAT...AAGGCCCAAGATGAACATGAGAAA...TATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAG  4338
                      _.__.__K__V__L__F__L__D__G__I__D__.__K__A__Q__D__E__H__E__K__.__Y__H__S__N__W__R__A__M__A__S__D__F__N__L__P__P__V__V__A__K__E__I__  integrase

H1A1.UG.85.U455       ......--G--G--------------G------..-----T-----G--C-----A---..------T-C--C-------------------------T-----------G-----G--G-------  3784
H1B.US.90.WEAU160     ......--------------------G------..-----------A------------..--------------------------------------------------------------------  4337
H1C.ET.86.ETH2220     ......-----G--G---C--------------..-----T-----A--G-----A---..--------C------------------A---A-----TA-C-----C------C-C----------  3730
H1D.CD.84.84ZR085     ......--------------G------------..-----T-----G------------..--------C-------------------------------------G-----------------  3860
H1F1.BE.93.VI850      ......---A----G-----------G------..-----A-----A--------A---..-------AC-----------G-----------------T---------A-----------------  3678
H1G.SE.93.SE6165      ......-------------C-------C-----..--A--------A--G-------G-..-------AC--C---------------C----------TT--------A-----------------  3735
H1H.CF.90.056         ......-----G-------C-------G-----..--A--T-----A--------A-GG..-----T-AC-----------G-----------------T--A------A-----------------  3685
H1J.SE.93.SE7887      ......-----G--G---C-------G------..--A--T-----A--T-----A---..-----T-C---------------------------------T----------------------G----  3652
H1K.CM.96.MP535       ......-----G------C--------------..-----T-----A--------A---..-----T-AC--C-----------------------------T---------A-------C----------  3534
H101_AE.TH.90.CM240   ......--G--G--------------G------..-----T-----A--------A-G-..--------C---------A------------------TT--------A----------G-------  3912
H102_AG.NG.-.IBNG     ......--G-----------------C------..--A--------A--G-----A-G-..--------C------AG-------------------T---------A-----------------  3863
H1N.CM.95.YBF30       ......---A--T----CC-------T-----A..--A--------A--T-----C-G-..--------C------A---------C-----------T-A--C--CA----G--------------  3932
H1O.BE.87.ANT70       ......-G---C--G--CC----A---------C..C----A-----A--T-----A---..-----T-----------A----C-A--------A---GGA--A-----A--G--G--C--G-----CA  4393
H1O.CM.91.MVP5180     ......-G---C--G--CC-G--A---------..C----A-----A--T-----A---..-----T---------------T-A--------C---GGAT-A-----AA-------C--G-----CA  4368

CPZ.CD.-.ANT          ......C----C--C---C-G--A--C------..--A--T-----G--C-----T---..-----T--C---------T--T-A--AGA---A-AC--T--T--C---A-T--G--T--------TA  3777
CPZ.CM.-.CAM3         ......-----T--C-----------T------..T---CA-A-G--A---------G--A..-----T-AC-----------T-----A--------C--TA-A----A-A----G--C------C---  3752
CPZ.GA.88.GAB2        ......--G--G--------------T------..-----A-----A--C-------G-..-----T--C--C---------C------GA---A--CC-A--C-----G-------C----------  3731
CPZ.GA.-.CPZGAB       ......-----G-----CC-------------C..-G---------A--------A-GG..-----T----------A---T-----------------TT-A-----CA-----------------  4399
CPZ.US.85.CPZUS       ......-----T--T--CC-------C------..--A--A--G--A--C-----C---..-----T----------C---T-----C-----------------------A--G--C--T--G--G----  4398
CPZ.TZ.-.TAN1         ......---A--T----CC-G--------TA--..G-A--A--G--A--C-----T---..-----------------A---TT-A---GA---A-A---T-----C--A--T--G--T--------TA  3982

H2A.GW.-.ALI          ......C-----T-G---C----AAA------G..CCA--T--G--A--------A--G..-----T--C---GT--A--A-C-AT-CCA-A-A---GGAT-A--CAA-C-G--G----G-C-G----  4713
H2A.DE.-.BEN          ......C-----T----CC-G--AAAG-----G..CCC--T-----G--------A---..-----T--C-T-ATA-A--A-C-AA-CCA-A-A---GGAA-A--C-T-C---------G-C-G----  4720
H2A.SN.-.ST           ......C-----T----CC----GAA------A..CCC--T--G--G--------A---..-----T--C---GTA-A--A-C-AT-CCA-A-A---GGA-----CAAAT----G----G-C------  4165
H2B.GH.86.D205        ......C-G--CT-G--CC----AAA------A..CCA-----G--A--G-----A---..-----TG-C---GTA-A--A-C---TCCA-A-A--CGGAA-T----AAT----G------C-G----  4693
H2B.CI.-.EHO          ......C--A-C------C----AAA------A..CCA--T-----A--------A---..-----T-A----GTA-A--A-C-A-TCCA-A-A---GGGA-T----AAT----G----G-C------  4691
H2G.CI.-.ABT96        ......C----------C--GR-AAA------A..CCT--T-----A--------A--R..-T---T------GTA-A--AGT-AA--CA-A-G---GG-A-T--T-AAT-----------C-G----  4047
H2U.FR.96.12034       ......C-------G---C----AAA------A..CCA--T--G--A------------..--C--T--C---GTT-A--AGC-ACAGCA-A-A---GGTA-A----AAT-------C---C------  4199

MAC.US.-.239          ......C----T--C--C--G--AAAG-----G..CCA--A-----A--------T---..--C--T------GTA-A--A-T---TATTCA-A---GGAT-A--CAGAA----G--C-G-C-G----  4637
Pol                   _.__.__Q__V__L__F__L__E__K__I__E__.__P__A__Q__E__E__H__D__K__.__Y__H__S__N__V__K__E__L__V__F__K__F__G__L__P__R__I__V__A__R__Q__I__  p31

SMM.SL.92.SL92B       ......C-G--C--G--CC----AA-T-----G..CCA--A-----A--C-----C---..-----T--C---GTA-AG-AGT-A---CAGA-A-A----A-C----AAT-------T---C-G----  4068
SMM.US.-.H9           ......C----C-----C-----AAA------A..CCA--G--G--A--G-----A---..--C--T--C---GTA-A-AA-T---TATTCA-A---RGTT-A--TAGGC-----------C------  4110

STM.US.-.STM          ......C-G--C-----CC----AAA------A..CCA--T-----A--G-----A---..-----T--C---GT--A--AGC-A-TCTTCA-G---GG-A-A---AGGC----------GC------  4284

SAB.SN.-.SAB1C        ......C----------CC-G--CA----T--A..G-A--A-----A--G-----T---..--C--TGCC---------GAGT---CAACAA--A---GG-T-A--TG--A------------------  4736

TAN.UG.-.TAN1         ......C----G-----C-----CA----G--A..G-A-----G--ATC------T---..-----T-C-------CA-TTT--AAGAGA--CA--CGGTA-A---G-AC-G---------------  4537

VER.KE.-.AGM155       ......-G-A--T-G--CA-T-GCA-G-----A..G-A--A-----A--------T-GG..-----------C------AATC-A--AGACACA---GGAT-----AAA-------T--------T-  4557
VER.KE.-.9063         ......-G---T------A---GAA-------A..G-A--T-----A--------C-G-..-----T----------A-AATT-A--GGACACA---GGAT-A----AGA----T--C-----G----  4565
VER.DE.-.AGM3         ......---A-TT----------AAA------A..G-A-----G--G--------A-GG..--C--T-A---------GAACT-A--AGACAC--A-GGG--A----AAA-T--G-------------  4060
VER.KE.-.TYO1         ......-G---TT----C-----AAA------A..G-A--T-----AA-G-----A-G-..-----T-A--------A-AACC-A--AGA-ACA-A-GGG--T----AAA-------------G----  4052

COL.CM.-.CGU1         ......C-G---A-G-GGA-----AA------A..GCA--AG----A--C---C-A--G..-T---T------GT-CAGTATC-AAAAGAAC-A---GGGT-A--CA------------------TT  4101

DEN.CD.-.CD1          ......C--A-C-----C-----ACA---TCCA..C-A--T-----A--------T---..--C--T--C-------A--AG--CAGAGA-A-G-A-GG---A---A--A-G-----T--G-----CA  4228

GRV.ET.-.GRI_677      ......C-GA-C--C--C------A-------A..G-A--A---------C----CA--G..--C--T-AC---------AGT----TACAG--A---GGAT-A--TAA-A-------------G----  4511

GSN.CM.99.CN166       ......C----G--T--CC-G--AAAC-----G..CCT--AGT---A--C-----A--G..-----T----------AGTACC-AAGAGA-C-A-AC--AA-C--CG-AC--T-G--T----------  4066
GSN.CM.99.CN71        ......C-------C--CA-G--GAAC-----G..CCC--AGT---G--C-----A---..-----T----------AGTATT-AAGGGA-C-A-AC--AA-C--GG-AC--T-G--T----------  4048

DRL.-.-.FAO           ......C----GT----CC-----AAC--G--A..--A--A--G--A--------TTT-..--C--T-A----------T-CT-A--ACAA--A---GGTT-A---GGGA----G-------------  3905

RCM.GA.-.GAB1         ......C-G-----G--CC-------G-----A..--A--A-----A--------A--G..-----T-A-------------T----AGAA------C-AA-T----AAA----G-------------  3817
RCM.NG.-.NG411        ......C-G--------CC-----C--G----GA..--A--A--G--A--------A---..-----T--C-----------------AGAA--A---C-AA-A--T-AAA-----------------  3832

MND-2.-.-.5440        ......C-G--------------G---------..--A-----G--A--------C---..--------A------------GT--T--CAG--C--------A--C--CA----G--------G----  3899
MND-2.CM.98.CM16      ......C-------G--------A--C------..--A--A-----A--G-----T---..--------A---------G---T-A--ACAA--C--CTGTA-A---AACA----G-------------  4339
MND-2.GA.-.M14        ......C----------CC-G--A---------..--A--T--G--G--------C---..--C--T-A----------T--T-AAG-CAG--A----G-A-A--C---A----G-------------  4266

MNE.US.-.MNE027       ......C----C--C--C--G--AAAG-----G..CCA--A-----A--------T---..--C--T------GTA-A--A-T---TATTCA-A---GGAT-A--CAGAC----G--C---C-G----  4105

LST.CD.88.447         CAAAAAC-------G---C---GAAAG-----G..CCC--A--G--A--GT--AGC---..-T---T-A----GCT-A--ATT-A-AAGAAA-A---C--A-A--C---A-G-----T--GC------  3615
LST.CD.88.485         CAAATAC-------G---C----AAA------G..CCA--A--G--A--G---AGC---..-T---T-A----GCT-A--ATC-A-AAGAAA-A---C----A--C---A-G-----T--GC------  3615
LST.CD.88.524         ACAAAAC-G--CA-G---C-G--GAAG-----A..CCA--TGT---A--G----GA--G..-T-----A----GCA-A--ACT-A-AAGAAA-G-A-GGGA-A--C--CA-G-----T--GC------  3612
LST.KE.-.lho7         CAGAAAC-G--CA-G---C----AAAG-----A..CCT--AGT---A-------AGC---..-TC--T-AC---GCA-AG-ATC-A-AAGAAA-A--------A--C--CA-G-----C---C----T-  4699

SUN.GA.98.L14         AAGAGAC-G---A-G---A----AAAG-----A..CCA--TGT---A-------G---G..-TC--T-AC---GCAGCCT-TT--CAGGAGATG---G--A-A--C-TA--G--------GC-G----  4698

MND-1.GA.-.MNDGB1     ......-----CT----CC--C-AAAT-----A..CCA--A--G--A------------..-----T--C---GAAGC-CA-T-AAGAGAGA-A--CC--T-A---G--C-------C---C-G--T-  4025

MON.CM.99.L1          ......C----C--G---A-G--AAATT-G--A..CCC--AAGG--A--T-----A--G..-----T----------A-TACT-AAGAGA-ACC-ACC--A-T--TA--C-GT-G--C----------  4041
MON.NG.-.NG1          ......C-G--C--C--CA-G--GAAC--T--A..CCR----G---G--T-----A---..--------C--------A-TATC-CAGGGA-AC--A-GG-A-T---A-AC-CC----C-----G----  2634

MUS.CM.01.1085        ......C----C-----CA-G--ACAG-----A..CCT--AA-G--A--T-----A---..-----------------AGTACC-CAGGGA-A-A-ACC-AA-C---G--T-GT----C----------  4028

DEB.CM.99.CM40        ......-GGA-C--C--C--G--AA-GG----G..G-A--------G--C-----A---..--C--TG-G-------A-CA-T-AAGGGA---A-ACC-A--A--TA-AC-CA----T--GC----TA  4144
DEB.CM.99.CM5         ......-G-A-C-----C--G--AA--G----A..G----T-----G--T---------..-----TG---------A-CA-T--AGAGA---A---C-GT-A--TA-CT--A----C--GC-G--TA  4138

SYK.KE.-.KE51         ......---A----G---C-G---AAC--CCCA..--A--A--G--A------------..--------C----GTAGA-TAC--AAGGCAA--A---C-T--A--CAGACAG-----C----CT----  4193
SYK.KE.-.SYK173       ......---A-C-----CC----AA----TCCC..C-A--A--G--A--T-----A-GG..--C---TCA---AT-GAGTATC-TAGGCAG--G---C-T--A--C-GACAG----------C---TA  4538
```

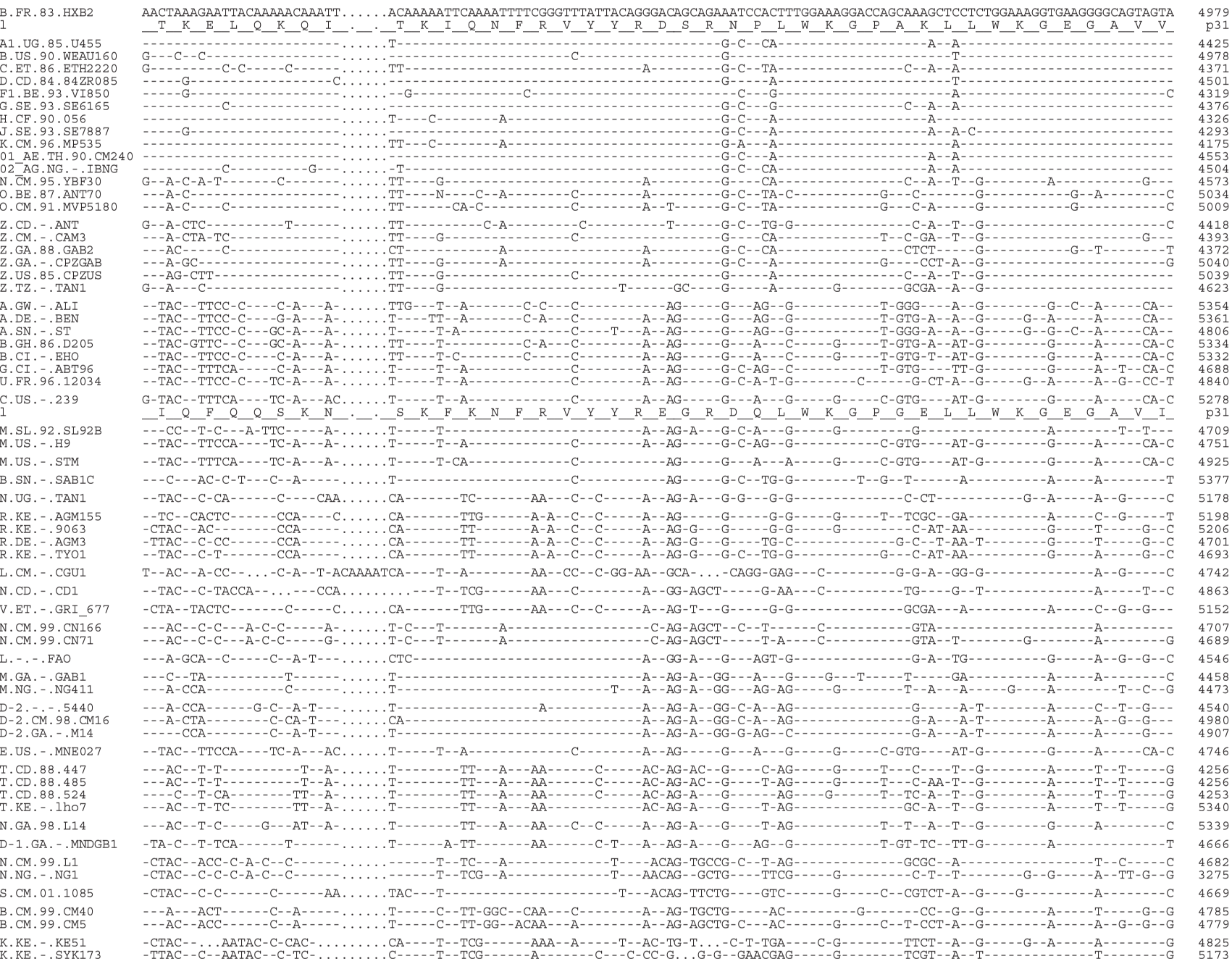

Asp64 catalytic site
H1B.FR.83.HXB2 TAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATTTAGAAGGAAAAGTTATCCTGGTAGCAGTTCATGTAGC 4468
Pol V A S C D K C Q L K G E A M H G Q V D C S P G I W Q L D C T H L E G K V I L V A V H V A p31
H1A1.UG.85.U455 3914
H1B.US.90.WEAU160 4467
H1C.ET.86.ETH2220 3860
H1D.CD.84.84ZR085 3990
H1F1.BE.93.VI850 3808
H1G.SE.93.SE6165 3865
H1H.CF.90.056 3815
H1J.SE.93.SE7887 3782
H1K.CM.96.MP535 3664
H101_AE.TH.90.CM240 4042
H102_AG.NG.-.IBNG 3993
H1N.CM.95.YBF30 4062
H1O.BE.87.ANT70 4523
H1O.CM.91.MVP5180 4498
CPZ.CD.-.ANT 3907
CPZ.CM.-.CAM3 3882
CPZ.GA.88.GAB2 3861
CPZ.GA.-.CPZGAB 4529
CPZ.US.85.CPZUS 4528
CPZ.TZ.-.TAN1 4112
H2A.GW.-.ALI 4843
H2A.DE.-.BEN 4850
H2A.SN.-.ST 4295
H2B.GH.86.D205 4823
H2B.CI.-.EHO 4821
H2G.CI.-.ABT96 4177
H2U.FR.96.12034 4329
MAC.US.-.239 4767
Pol V D T C D K C H Q K G E A 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 p31
SMM.SL.92.SL92B 4198
SMM.US.-.H9 4240
STM.US.-.STM 4414
SAB.SN.-.SAB1C 4866
TAN.UG.-.TAN1 4667
VER.KE.-.AGM155 4687
VER.KE.-.9063 4695
VER.DE.-.AGM3 4190
VER.KE.-.TYO1 4182
COL.CM.-.CGU1 4231
DEN.CD.-.CD1 4358
GRV.ET.-.GRI_677 4641
GSN.CM.99.CN166 4196
GSN.CM.99.CN71 4178
DRL.-.-.FAO 4035
RCM.GA.-.GAB1 3947
RCM.NG.-.NG411 3962
MND-2.-.-.5440 4029
MND-2.CM.98.CM16 4469
MND-2.GA.-.M14 4396
MNE.US.-.MNE027 4235
LST.CD.88.447 3745
LST.CD.88.485 3745
LST.CD.88.524 3742
LST.KE.-.lho7 4829
SUN.GA.98.L14 4828
MND-1.GA.-.MNDGB1 4155
MON.CM.99.L1 4171
MON.NG.-.NG1 2764
MUS.CM.01.1085 4158
DEB.CM.99.CM40 4274
DEB.CM.99.CM5 4268
SYK.KE.-.KE51 4323
SYK.KE.-.SYK173 4668

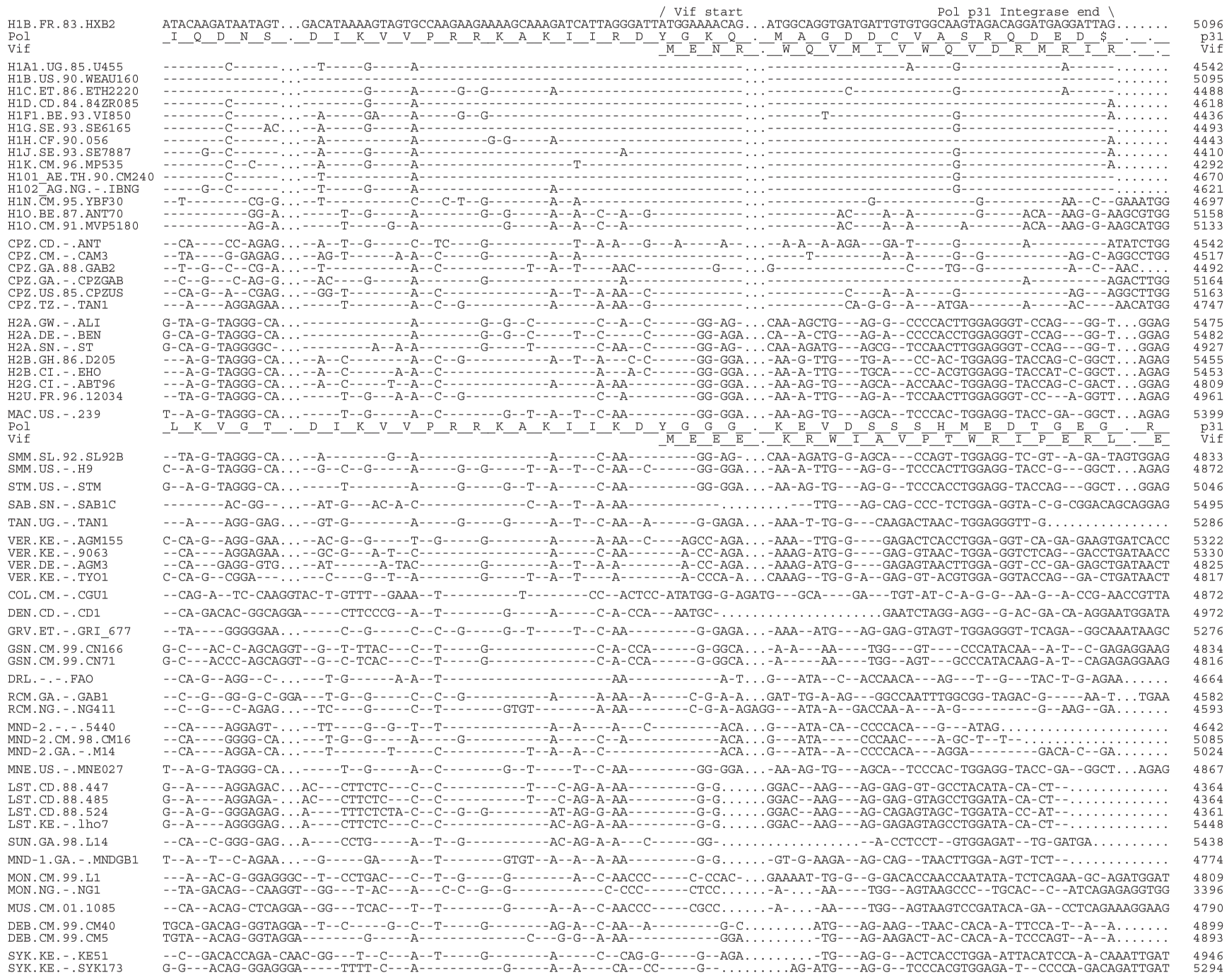


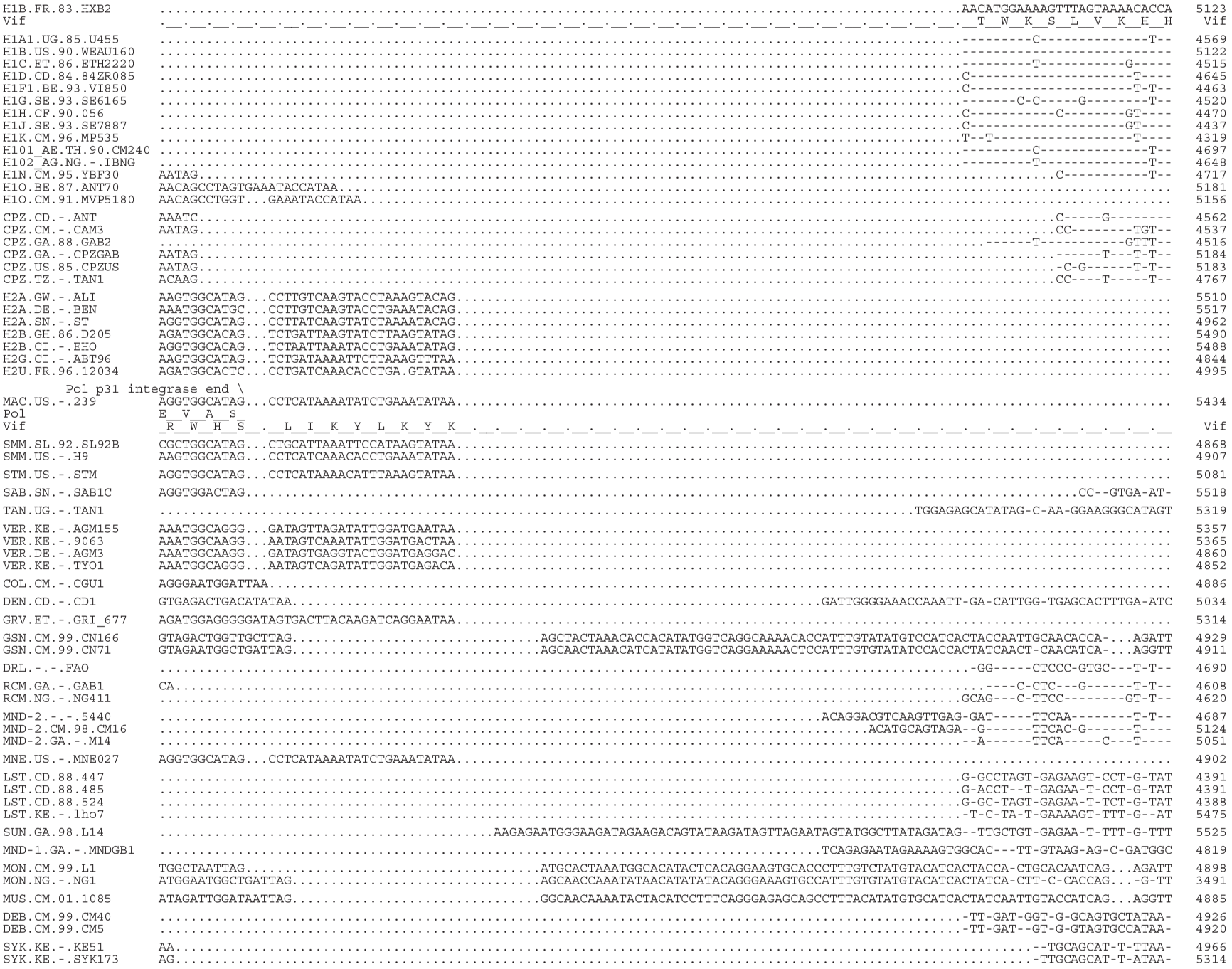
PLV
complete genomes

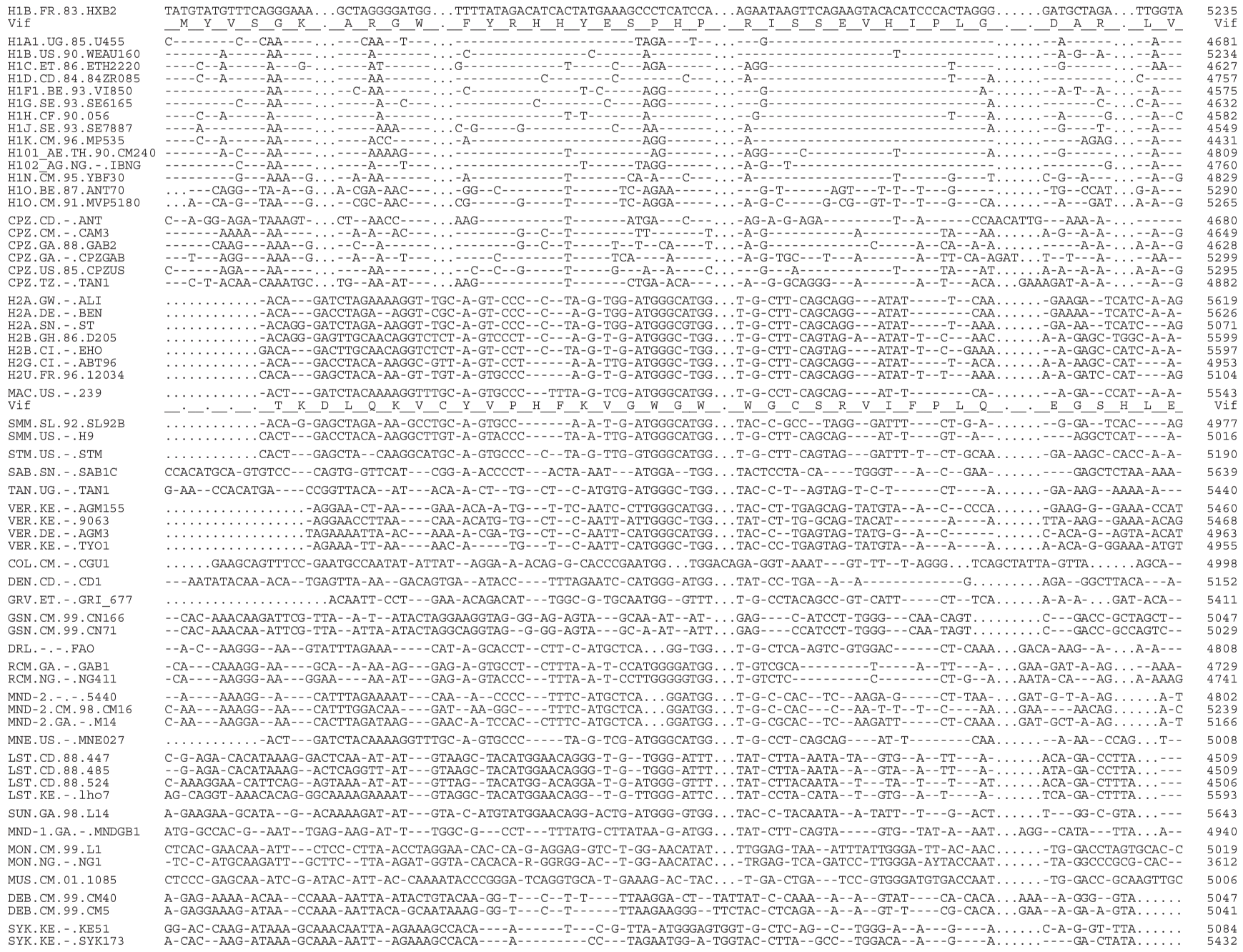

Glu152 catalytic site
H1B.FR.83.HXB2 GCTACGGTTAGGGCCGCCTGTTGGTGGGCGGGAATCAAGCAGGAATTTGGAATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATTAAAGAAAATTATAGGACAGGTAAGAG 4728
Pol _A__T__V__R__A__A__C__W__W__A__G__I__K__Q__E__F__G__I__P__Y__N__P__Q__S__Q__G__V__V__E__S__M__N__K__E__L__K__K__I__I__G__Q__V__R__ p31
H1A1.UG.85.U455 4174
H1B.US.90.WEAU160 4727
H1C.ET.86.ETH2220 4120
H1D.CD.84.84ZR085 4250
H1F1.BE.93.VI850 4068
H1G.SE.93.SE6165 4125
H1H.CF.90.056 4075
H1J.SE.93.SE7887 4042
H1K.CM.96.MP535 3924
H101_AE.TH.90.CM240 4302
H102_AG.NG.-.IBNG 4253
H1N.CM.95.YBF30 4322
H1O.BE.87.ANT70 4783
H1O.CM.91.MVP5180 4758
CPZ.CD.-.ANT 4167
CPZ.CM.-.CAM3 4142
CPZ.GA.88.GAB2 4121
CPZ.GA.-.CPZGAB 4789
CPZ.US.85.CPZUS 4788
CPZ.TZ.-.TAN1 4372
H2A.GW.-.ALI 5103
H2A.DE.-.BEN 5110
H2A.SN.-.ST 4555
H2B.GH.86.D205 5083
H2B.CI.-.EHO 5081
H2G.CI.-.ABT96 4437
H2U.FR.96.12034 4589
MAC.US.-.239 5027
Pol _Q__E__V__K__M__V__A__W__W__A__G__I__E__H__T__F__G__V__P__Y__N__P__Q__S__Q__G__V__V__E__A__M__N__H__H__L__K__N__Q__I__D__R__I__R__ p31
SMM.SL.92.SL92B 4458
SMM.US.-.H9 4500
STM.US.-.STM 4674
SAB.SN.-.SAB1C 5126
TAN.UG.-.TAN1 4927
VER.KE.-.AGM155 4947
VER.KE.-.9063 4955
VER.DE.-.AGM3 4450
VER.KE.-.TYO1 4442
COL.CM.-.CGU1 4491
DEN.CD.-.CD1 4618
GRV.ET.-.GRI_677 4901
GSN.CM.99.CN166 4456
GSN.CM.99.CN71 4438
DRL.-.-.FAO 4295
RCM.GA.-.GAB1 4207
RCM.NG.-.NG411 4222
MND-2.-.-.5440 4289
MND-2.CM.98.CM16 4729
MND-2.GA.-.M14 4656
MNE.US.-.MNE027 4495
LST.CD.88.447 4005
LST.CD.88.485 4005
LST.CD.88.524 4002
LST.KE.-.lho7 5089
SUN.GA.98.L14 5088
MND-1.GA.-.MNDGB1 4415
MON.CM.99.L1 4431
MON.NG.-.NG1 3024
MUS.CM.01.1085 4418
DEB.CM.99.CM40 4534
DEB.CM.99.CM5 4528
SYK.KE.-.KE51 4583
SYK.KE.-.SYK173 4928

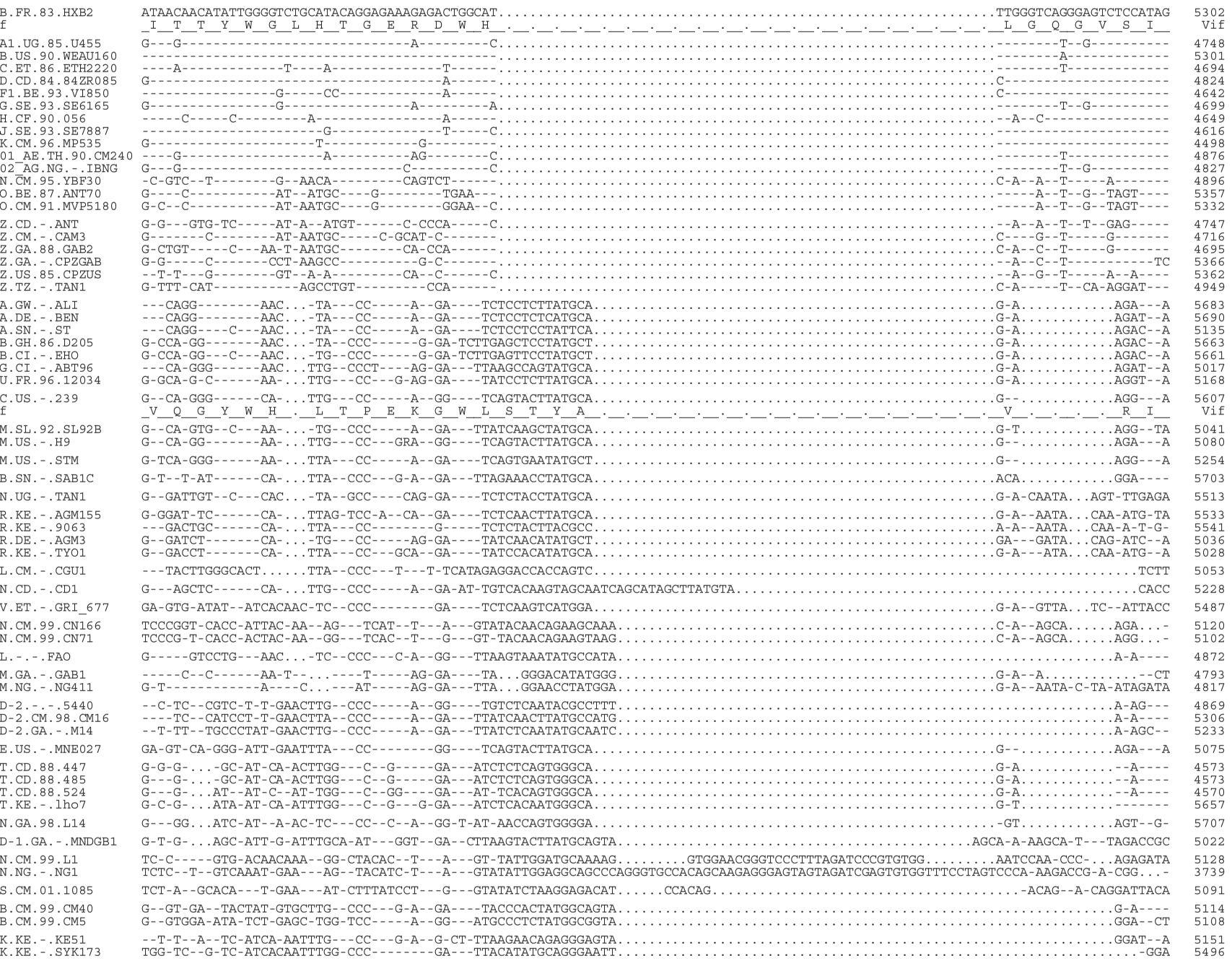

H1B.FR.83.HXB2
Pol
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Pol
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes

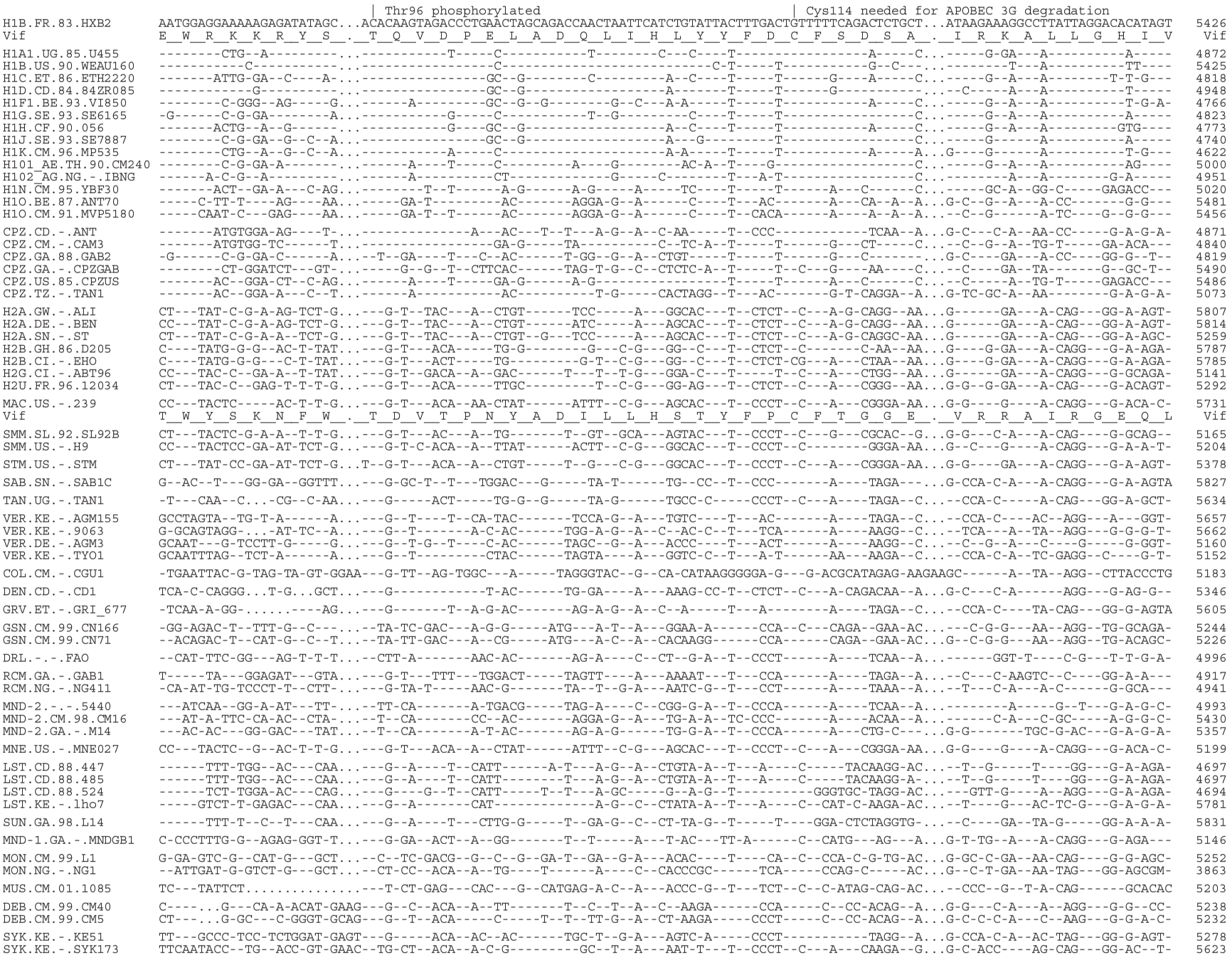

H1B.FR.83.HXB2 AACTAAAGAATTACAAAAACAAATT.....ACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAAGGTGAAGGGGCAGTAGTA 4979
Pol __T__K__E__L__Q__K__Q__I__.__.__T__K__I__Q__N__F__R__V__Y__Y__R__D__S__R__N__P__L__W__K__G__P__A__K__L__L__W__K__G__E__G__A__V__V_ p31
H1A1.UG.85.U455 4425
H1B.US.90.WEAU160 4978
H1C.ET.86.ETH2220 4371
H1D.CD.84.84ZR085 4501
H1F1.BE.93.VI850 4319
H1G.SE.93.SE6165 4376
H1H.CF.90.056 4326
H1J.SE.93.SE7887 4293
H1K.CM.96.MP535 4175
H101_AE.TH.90.CM240 4553
H102_AG.NG.-.IBNG 4504
H1N.CM.95.YBF30 4573
H1O.BE.87.ANT70 5034
H1O.CM.91.MVP5180 5009
CPZ.CD.-.ANT 4418
CPZ.CM.-.CAM3 4393
CPZ.GA.88.GAB2 4372
CPZ.GA.-.CPZGAB 5040
CPZ.US.85.CPZUS 5039
CPZ.TZ.-.TAN1 4623
H2A.GW.-.ALI 5354
H2A.DE.-.BEN 5361
H2A.SN.-.ST 4806
H2B.GH.86.D205 5334
H2B.CI.-.EHO 5332
H2G.CI.-.ABT96 4688
H2U.FR.96.12034 4840
MAC.US.-.239 5278
Pol __I__Q__F__Q__Q__S__K__N__.__.__S__K__F__K__N__F__R__V__Y__Y__R__E__G__R__D__Q__L__W__K__G__P__G__E__L__L__W__K__G__E__G__A__V__I_ p31
SMM.SL.92.SL92B 4709
SMM.US.-.H9 4751
STM.US.-.STM 4925
SAB.SN.-.SAB1C 5377
TAN.UG.-.TAN1 5178
VER.KE.-.AGM155 5198
VER.KE.-.9063 5206
VER.DE.-.AGM3 4701
VER.KE.-.TYO1 4693
COL.CM.-.CGU1 4742
DEN.CD.-.CD1 4863
GRV.ET.-.GRI_677 5152
GSN.CM.99.CN166 4707
GSN.CM.99.CN71 4689
DRL.-.-.FAO 4546
RCM.GA.-.GAB1 4458
RCM.NG.-.NG411 4473
MND-2.-.-.5440 4540
MND-2.CM.98.CM16 4980
MND-2.GA.-.M14 4907
MNE.US.-.MNE027 4746
LST.CD.88.447 4256
LST.CD.88.485 4256
LST.CD.88.524 4253
LST.KE.-.lho7 5340
SUN.GA.98.L14 5339
MND-1.GA.-.MNDGB1 4666
MON.CM.99.L1 4682
MON.NG.-.NG1 3275
MUS.CM.01.1085 4669
DEB.CM.99.CM40 4785
DEB.CM.99.CM5 4779
SYK.KE.-.KE51 4825
SYK.KE.-.SYK173 5173

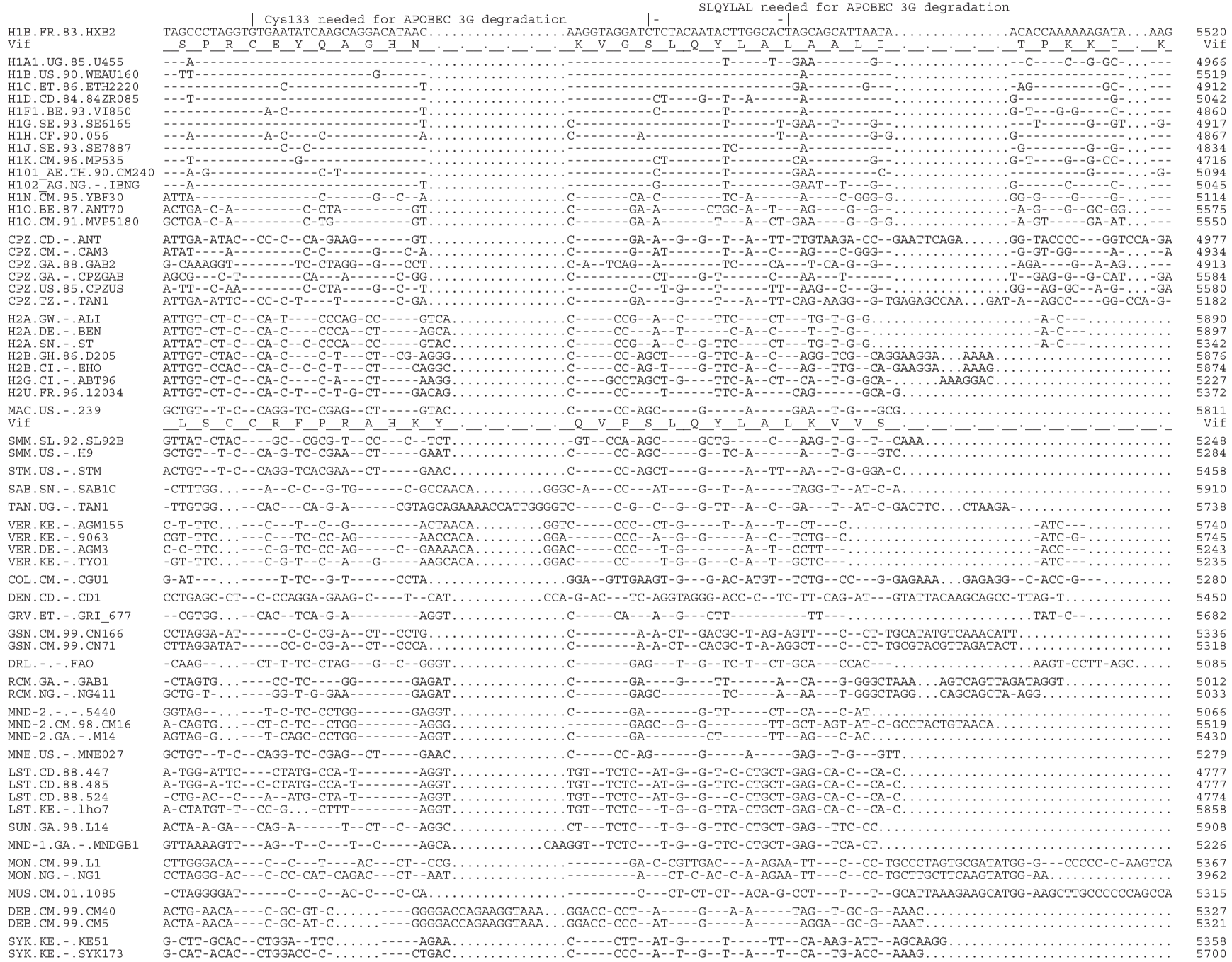

/ Vif start
Pol p31 Integrase end \
H1B.FR.83.HXB2
Pol
Vif
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Pol
Vif
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes

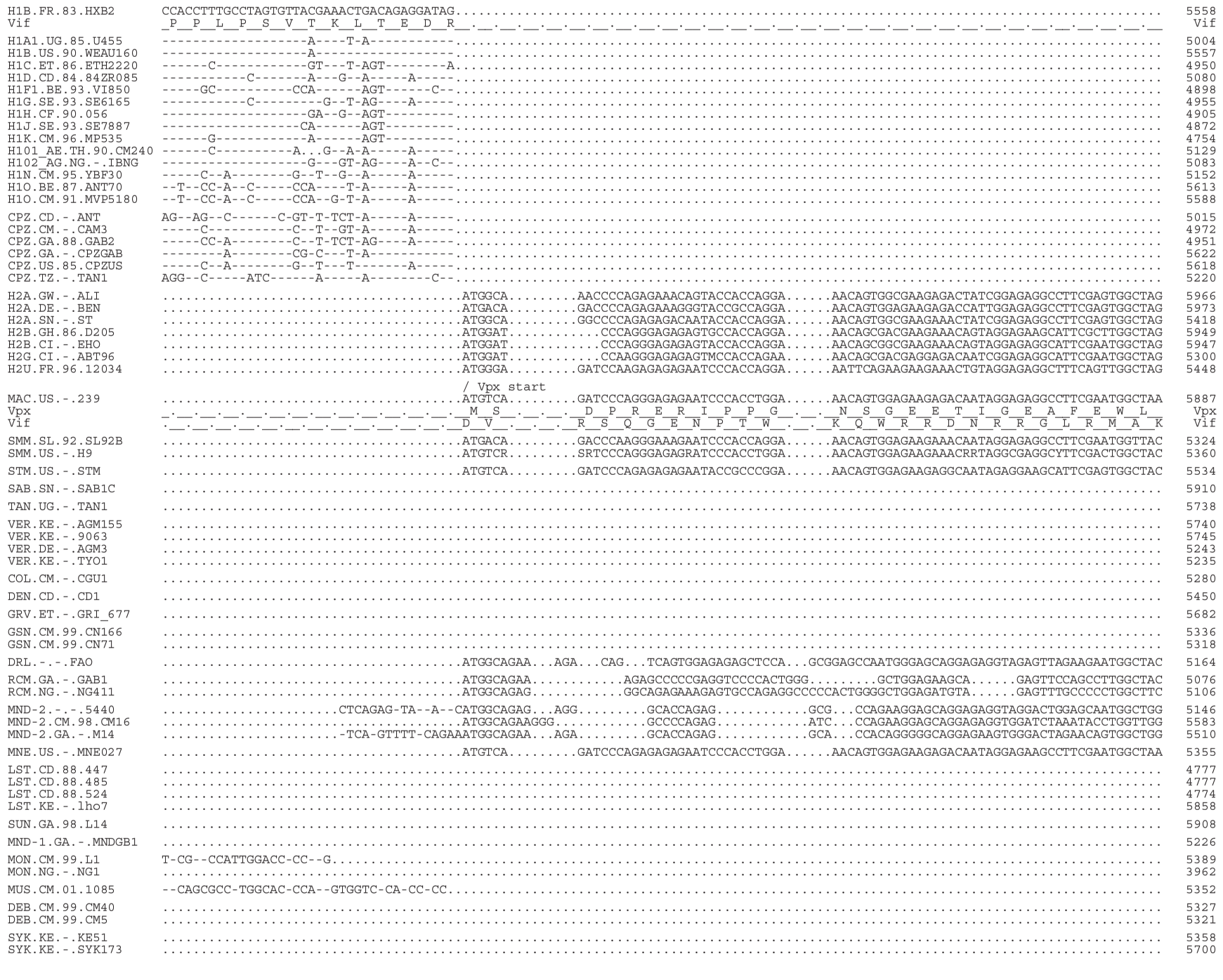

H1B.FR.83.HXB2 AACATGGAAAAGTTTAGTAAAACACCA 5123
Vif T W K S L V K H H Vif
H1A1.UG.85.U455 C T-- 4569
H1B.US.90.WEAU160 5122
H1C.ET.86.ETH2220 T G----- 4515
H1D.CD.84.84ZR085 C T---- 4645
H1F1.BE.93.VI850 C T-T-- 4463
H1G.SE.93.SE6165 C-C-----G-------T-- 4520
H1H.CF.90.056 C---------C--------GT---- 4470
H1J.SE.93.SE7887 C-------------------GT---- 4437
H1K.CM.96.MP535 T--T----------------T---- 4319
H101_AE.TH.90.CM240 C---------T-- 4697
H102_AG.NG.-.IBNG T---------T-- 4648
H1N.CM.95.YBF30 AATAG C----------T-- 4717
H1O.BE.87.ANT70 AACAGCCTAGTGAAATACCATAA 5181
H1O.CM.91.MVP5180 AACAGCCTGGT...GAAATACCATAA 5156
CPZ.CD.-.ANT AAATC C-----G-------- 4562
CPZ.CM.-.CAM3 AATAG CC--------TGT-- 4537
CPZ.GA.88.GAB2 ------T----------GTTT-- 4516
CPZ.GA.-.CPZGAB AATAG ------T---T-T-- 5184
CPZ.US.85.CPZUS AATAG -C-G------T-T-- 5183
CPZ.TZ.-.TAN1 ACAAG CC----T-----T-- 4767
H2A.GW.-.ALI AAGTGGCATAG...CCTTGTCAAGTACCTAAAGTACAG 5510
H2A.DE.-.BEN AAATGGCATGC...CCTTGTCAAGTACCTGAAATACAG 5517
H2A.SN.-.ST AGGTGGCATAG...CCTTATCAAGTATCTAAAATACAG 4962
H2B.GH.86.D205 AGATGGCACAG...TCTGATTAAGTATCTTAAGTATAG 5490
H2B.CI.-.EHO AGGTGGCACAG...TCTAATTAAATACCTGAAATATAG 5488
H2G.CI.-.ABT96 AAGTGGCATAG...TCTGATAAAATTCTTAAAGTTTAA 4844
H2U.FR.96.12034 AGATGGCACTC...CCTGATCAAACACCTGA.GTATAA 4995
Pol p31 integrase end \
MAC.US.-.239 AGGTGGCATAG...CCTCATAAAATATCTGAAATATAA 5434
Pol E V A $
Vif R W H S L I K Y L K Y K Vif
SMM.SL.92.SL92B CGCTGGCATAG...CTGCATTAAATTCCATAAGTATAA 4868
SMM.US.-.H9 AAGTGGCATAG...CCTCATCAAACACCTGAAATATAA 4907
STM.US.-.STM AGGTGGCATAG...CCTCATAAAACATTTAAAGTATAA 5081
SAB.SN.-.SAB1C AGGTGGACTAG CC--GTGA-AT- 5518
TAN.UG.-.TAN1 TGGAGAGCATATAG-C-AA-GGAAGGGCATAGT 5319
VER.KE.-.AGM155 AAATGGCAGGG...GATAGTTAGATATTGGATGAATAA 5357
VER.KE.-.9063 AAATGGCAAGG...AATAGTCAAATATTGGATGACTAA 5365
VER.DE.-.AGM3 AAATGGCAAGG...GATAGTGAGGTACTGGATGAGGAC 4860
VER.KE.-.TYO1 AAATGGCAGGG...AATAGTCAGATATTGGATGAGACA 4852
COL.CM.-.CGU1 AGGGAATGGATTAA 4886
DEN.CD.-.CD1 GTGAGACTGACATATAA GATTGGGGAAACCAAATT-GA-CATTGG-TGAGCACTTTGA-ATC 5034
GRV.ET.-.GRI_677 AGATGGAGGGGGATAGTGACTTACAAGATCAGGAATAA 5314
GSN.CM.99.CN166 GTAGACTGGTTGCTTAG...AGCTACTAAACACCACATATGGTCAGGCAAAACACCATTTGTATATGTCCATCACTACCAATTGCAACACCA-...AGATT 4929
GSN.CM.99.CN71 GTAGAATGGCTGATTAG...AGCAACTAAACATCATATATGGTCAGGAAAAACTCCATTTGTGTATATCCACCACTATCAACT-CAACATCA-...AGGTT 4911
DRL.-.-.FAO -GG-----CTCCC-GTGC---T-T-- 4690
RCM.GA.-.GAB1 CA-----C-CTC---G------T-T-- 4608
RCM.NG.-.NG411 GCAG---C-TTCC--------GT-T-- 4620
MND-2.-.-.5440 ACAGGACGTCAAGTTGAG-GAT-----TTCAA--------T-T-- 4687
MND-2.CM.98.CM16 ACATGCAGTAGA--G------TTCAC-G------T---- 5124
MND-2.GA.-.M14 --A------TTCA-----C---T---- 5051
MNE.US.-.MNE027 AGGTGGCATAG...CCTCATAAAATATCTGAAATATAA 4902
LST.CD.88.447 G-GCCTAGT-GAGAAGT-CCT-G-TAT 4391
LST.CD.88.485 G-ACCT--T-GAGAA-T-CCT-G-TAT 4391
LST.CD.88.524 G-GC-TAGT-GAGAA-T-TCT-G-TAT 4388
LST.KE.-.lho7 -T-C-TA-T-GAAAAGT-TTT-G--AT 5475
SUN.GA.98.L14 AAGAGAATGGGAAGATAGAAGACAGTATAAGATAGTTAGAATAGTATGGCTTATAGATAG--TTGCTGT-GAGAA-T-TTT-G-TTT 5525
MND-1.GA.-.MNDGB1 TCAGAGAATAGAAAAGTGGCAC---TT-GTAAG-AG-C-GATGGC 4819
MON.CM.99.L1 TGGCTAATTAG...ATGCACTAAATGGCACATACTCACAGGAAGTGCACCCTTTGTCTATGTACATCACTACCA-CTGCACAATCAG...AGATT 4898
MON.NG.-.NG1 ATGGAATGGCTGATTAG...AGCAACCAAATATAACATATATACAGGGAAAGTGCCATTTGTGTATGTACATCACTATCA-CTT-C-CACCAG...-G-TT 3491
MUS.CM.01.1085 ATAGATTGGATAATTAG...GGCAACAAAATACTACATCCTTTCAGGGAGAGCAGCCTTTACATATGTGCATCACTATCAATTGTACCATCAG...AGGTT 4885
DEB.CM.99.CM40 -TT-GAT-GGT-G-GCAGTGCTATAA- 4926
DEB.CM.99.CM5 -TT-GAT--GT-G-GTAGTGCCATAA- 4920
SYK.KE.-.KE51 AA--TGCAGCAT-T-TTAA- 4966
SYK.KE.-.SYK173 AG-TTGCAGCAT-T-ATAA- 5314


H1B.FR.83.HXB2
Vif
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Vif
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes


H1B.FR.83.HXB2 ATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGACTGGCAT TTGGGTCAGGGAGTCTCCATAG 5302
Vif _I__T__T__Y__W__G__L__H__T__G__E__R__D__W__H__ _L__G__Q__G__V__S__I__ Vif
H1A1.UG.85.U455 G---G A C 4748
H1B.US.90.WEAU160 A 5301
H1C.ET.86.ETH2220 ----A T----A T 4694
H1D.CD.84.84ZR085 G A C 4824
H1F1.BE.93.VI850 G-----CC A C 4642
H1G.SE.93.SE6165 G G A A T--G 4699
H1H.CF.90.056 -----C-----C---------A A --A--C 4649
H1J.SE.93.SE7887 G T-----C 4616
H1K.CM.96.MP535 G T G 4498
H101_AE.TH.90.CM240 ----G A-----AG--------C T 4876
H102_AG.NG.-.IBNG G---G C T--G 4827
H1N.CM.95.YBF30 -C-GTC--T--------G--AACA---------CAGTCT C-A--A--T-----A------- 4896
H1O.BE.87.ANT70 G----C-----AT--ATGC----G--------TGAA--- -----A--T--G--TAGT---- 5357
H1O.CM.91.MVP5180 G-C--C-----AT-AATGC----G--------GGAA--C -----A--T--G--TAGT---- 5332
CPZ.CD.-.ANT G-G---GTG-TC-----AT-A--ATGT------C-CCCA-----C --A--A--T--T--GAG----- 4747
CPZ.CM.-.CAM3 G-------C--------AT-AATGC-----C-GCAT-C------- C----G--T-----G------- 4716
CPZ.GA.88.GAB2 G-CTGT-----C---AA-T-AATGC--------CA-CCA------ C-A--C--T-----G------- 4695
CPZ.GA.-.CPZGAB G-G-----C-------CCT-AAGCC---------G-C------- --A--C--T----------TC 5366
CPZ.US.85.CPZUS --T-T---G--------GT--A-A---------CA--C------C --A--G--T-----A--A---- 5362
CPZ.TZ.-.TAN1 G-TTT-CAT----------AGCCTGT---------CCA------ C-A-----T--CA-AGGAT--- 4949
H2A.GW.-.ALI ---CAGG--------AAC...-TA---CC-----A--GA----TCTCCTCTTATGCA G-A.....AGA---A 5683
H2A.DE.-.BEN ---CAGG--------AAC...-TA---CC-----A--GA----TCTCCTCTCATGCA G-A.....AGAT--A 5690
H2A.SN.-.ST ---CAGG----C---AAC...-TA---CC-----A--GA----TCTCCTCCTATTCA G-A.....AGAC--A 5135
H2B.GH.86.D205 G-CCA-GG-------AAC...-TA--CCC------G-GA-TCTTGAGCTCCTATGCT G-A.....AGAC--A 5663
H2B.CI.-.EHO G-CCA-GG---C---AAC...-TG--CCC------G-GA-TCTTGAGTTCCTATGCT G-A.....AGAC--A 5661
H2G.CI.-.ABT96 ---CA-GGG------AAC...TTG--CCCT----AG-GA---TTAAGCCAGTATGCA G-A.....AGAT--A 5017
H2U.FR.96.12034 G-GCA-G-C------AA-...TTG---CC---G-AG-GA----TATCCTCTTATGCA G-A.....AGGT--A 5168
MAC.US.-.239 G--CA-GGG------CA-...TTG---CC-----A--GG----TCAGTACTTATGCA G--.....AGG---A 5607
Vif _V__Q__G__Y__W__H__._L__T__P__E__K__G__W__L__S__T__Y__A__ _R__I__ Vif
SMM.SL.92.SL92B G--CA-GTG--C---AA-...-TG--CCC-----A--GA---TTATCAAGCTATGCA G-T.....AGG--TA 5041
SMM.US.-.H9 G--CA-GG-------AA-...TTG---CC---GRA--GG----TCAGTACTTATGCA G--.....AGA---A 5080
STM.US.-.STM G-TCA-GGG------AA-...TTA---CC-----A--GA----TCAGTGAATATGCT G--.....AGG---A 5254
SAB.SN.-.SAB1C G-T--T-AT------CA-...TTA--CCC---G-A--GA---TTAGAAACCTATGCA ACA.....GGA---- 5703
TAN.UG.-.TAN1 G--GATTGT--C---CAC...-TA--GCC----CAG-GA----TCTCTACCTATGCA G-A-CAATA...AGT-TTGAGA 5513
VER.KE.-.AGM155 G-GGAT-TC------CA-...TTAG-TCC-A--CA--GA----TCTCAACTTATGCA G-A--AATA...CAA-ATG-TA 5533
VER.KE.-.9063 ---GACTGC------CA-...TTA---CC--------G-----TCTCTACTTACGCC A-A--AATA...CAA-A-T-G- 5541
VER.DE.-.AGM3 G--GATCT-------CA-...-TG---CC-----AG-GA----TATCAACATATGCT GA---GATA...CAG-ATC--A 5036
VER.KE.-.TYO1 G--GACCT-------CA-...TTA---CC---GCA--GA----TATCCACATATGCA G-A---ATA...CAA-ATG--A 5028
COL.CM.-.CGU1 ---TACTTGGGCACT......TTA---CCC---T---T-TCATAGAGGACCACCAGTC TCTT 5053
DEN.CD.-.CD1 G---AGCTC------CA-...TTG--CCC-----A--GA-AT-TGTCACAAGTAGCAATCAGCATAGCTTATGTA CACC 5228
GRV.ET.-.GRI_677 GA-GTG-ATAT--ATCACAAC-TC--CCC--------GA----TCTCAAGTCATGGA G-A--GTTA...TC--ATTACC 5487
GSN.CM.99.CN166 TCCCGGT-CACC-ATTAC-AA--AG---TCAT--T---A---GTATACAACAGAAGCAAA C-A--AGCA......AGA...- 5120
GSN.CM.99.CN71 TCCCG-T-CACC-ACTAC-AA--GG---TCAC--T---G---GT-TACAACAGAAGTAAG C-A--AGCA......AGG...- 5102
DRL.-.-.FAO G-----GTCCTG---AAC...-TC--CCC---C-A--GG---TTAAGTAAATATGCCATA A-A---- 4872
RCM.GA.-.GAB1 -----C--C------AA-T--..----T-----AG-GA----TA...GGGACATATGGG G-A--A.....--CT 4793
RCM.NG.-.NG411 G-T----------A----C.....AT-----AG-GA---TTA...GGAACCTATGGA G-A--AATA-C-TA-ATAGATA 4817
MND-2.-.-.5440 --C-TC--CGTC-T-T-GAACTTG--CCC-----A--GG----TGTCTCAATACGCCTTT A-AG--- 4869
MND-2.CM.98.CM16 ----TC--CATCCT-T-GAACTTA--CCC-----A--GA---TTATCAACTTATGCCATG A-A---- 5306
MND-2.GA.-.M14 --T-TT--TGCCCTAT-GAACTTG--CCC-----A--GA---TTATCTCAATATGCAATC A-AGC-- 5233
MNE.US.-.MNE027 GA-GT-CA-GGG-ATT-GAATTTA---CC--------GG----TCAGTACTTATGCA G--.....AGA---A 5075
LST.CD.88.447 G-G-G-...-GC-AT-CA-ACTTGG---C--G-----GA---ATCTCTCAGTGGGCA G-A.....--A---- 4573
LST.CD.88.485 G---G-...-GC-AT-CA-ACTTGG---C--G-----GA---ATCTCTCAGTGGGCA G-A.....--A---- 4573
LST.CD.88.524 G---G-...AT--AT-C--AT-TGG---C--GG----GA---AT-TCACAGTGGGCA G-A.....--A---- 4570
LST.KE.-.lho7 G-C-G-...ATA-AT-CA-ATTTGG---C--G---G-GA---ATCTCACAATGGGCA G-T.....------- 5657
SUN.GA.98.L14 G---GG...ATC-AT--A-AC-TC---CC--C--A--GG-T-AT-AACCAGTGGGGA -GT.....AGT--G- 5707
MND-1.GA.-.MNDGB1 G-T-G-...AGC-ATT-G-ATTTGCA-AT---GGT--GA--CTTAAGTACTTATGCAGTA AGCA-A-AAGCA-T---TAGACCGC 5022
MON.CM.99.L1 TC-C-----GTG-ACAACAAA--GG-CTACAC--T---A---GT-TATTGGATGCAAAAG.....GTGGAACGGGTCCCTTTAGATCCCGTGTGG.....AATCCAA-CCC-...AGAGATA 5128
MON.NG.-.NG1 TCTC--T--GTCAAAT-GAA---AG--TACATC-T---A---GTATATTGGAGGCAGCCCAGGGTGCCACAGCAAGAGGGAGTAGTAGATCGAGTGTGGTTTCCTAGTCCCA-AAGACCG-A-CGG...- 3739
MUS.CM.01.1085 TCT-A--GCACA---T-GAA---AT-CTTTATCCT---G---GTATATCTAAGGAGACAT......CCACAG -ACAG--A-CAGGATTACA 5091
DEB.CM.99.CM40 G--GT-GA--TACTAT-GTGCTTG--CCC---G-A--GA----TACCCACTATGGCAGTA G-A---- 5114
DEB.CM.99.CM5 G--GTGGA-ATA-TCT-GAGC-TGG-TCC-----A--GG---ATGCCCTCTATGGCGGTA GGA--CT 5108
SYK.KE.-.KE51 --T-T--A--TC-ATCA-AATTTG---CC---G-A--G-CT-TTAAGAACAGAGGGAGTA GGAT--A 5151
SYK.KE.-.SYK173 TGG-TC--G-TC-ATCACAATTTGG-CCC--------GA---TTACATATGCAGGGAATT -GGA 5496

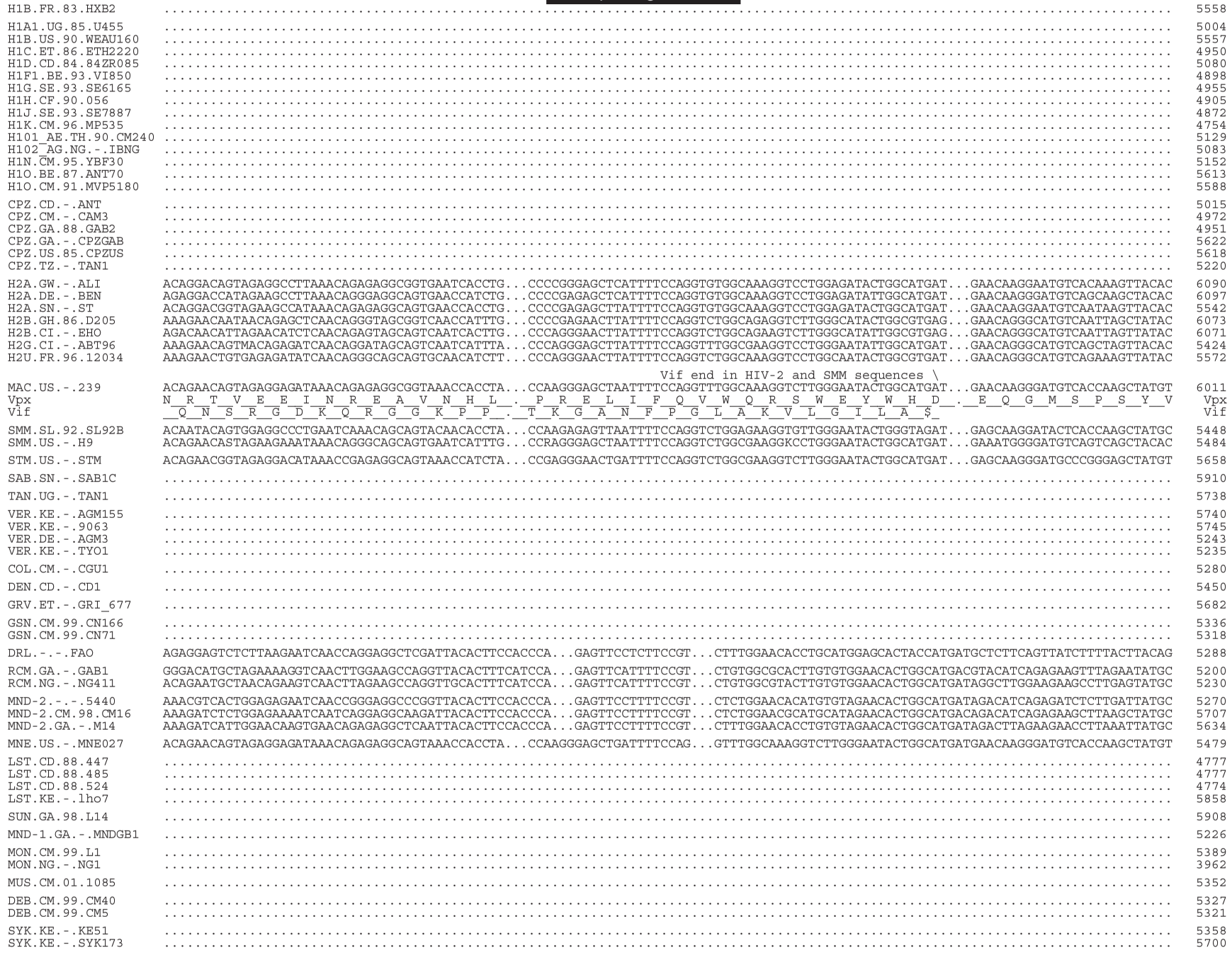

Thr96 phosphorylated
Cys114 needed for APOBEC 3G degradation
H1B.FR.83.HXB2
Vif
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Vif
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes

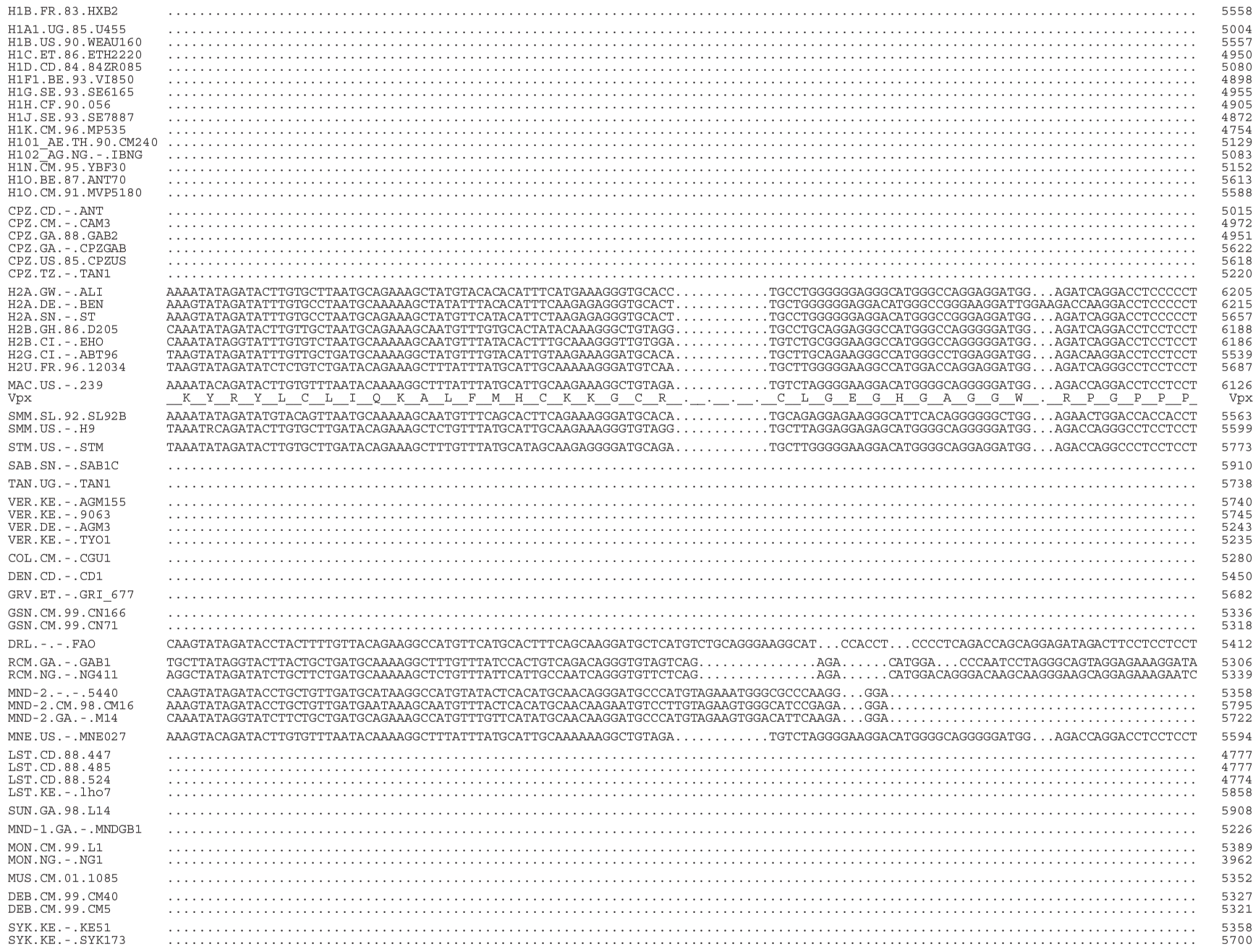

Cys133 needed for APOBEC 3G degradation
SLQYLAL needed for APOBEC 3G degradation
H1B.FR.83.HXB2
Vif
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Vif
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173



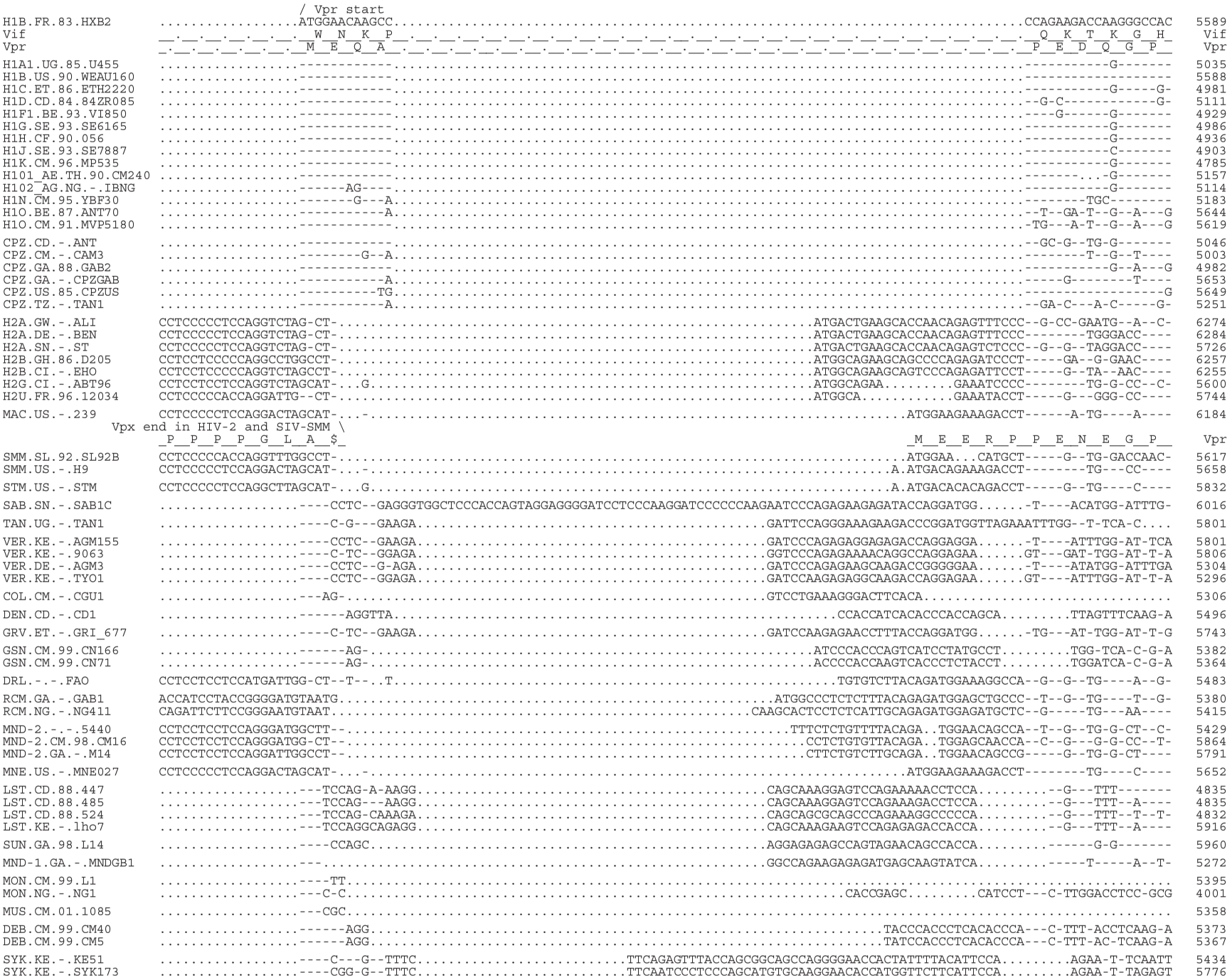
PLV
complete genomes

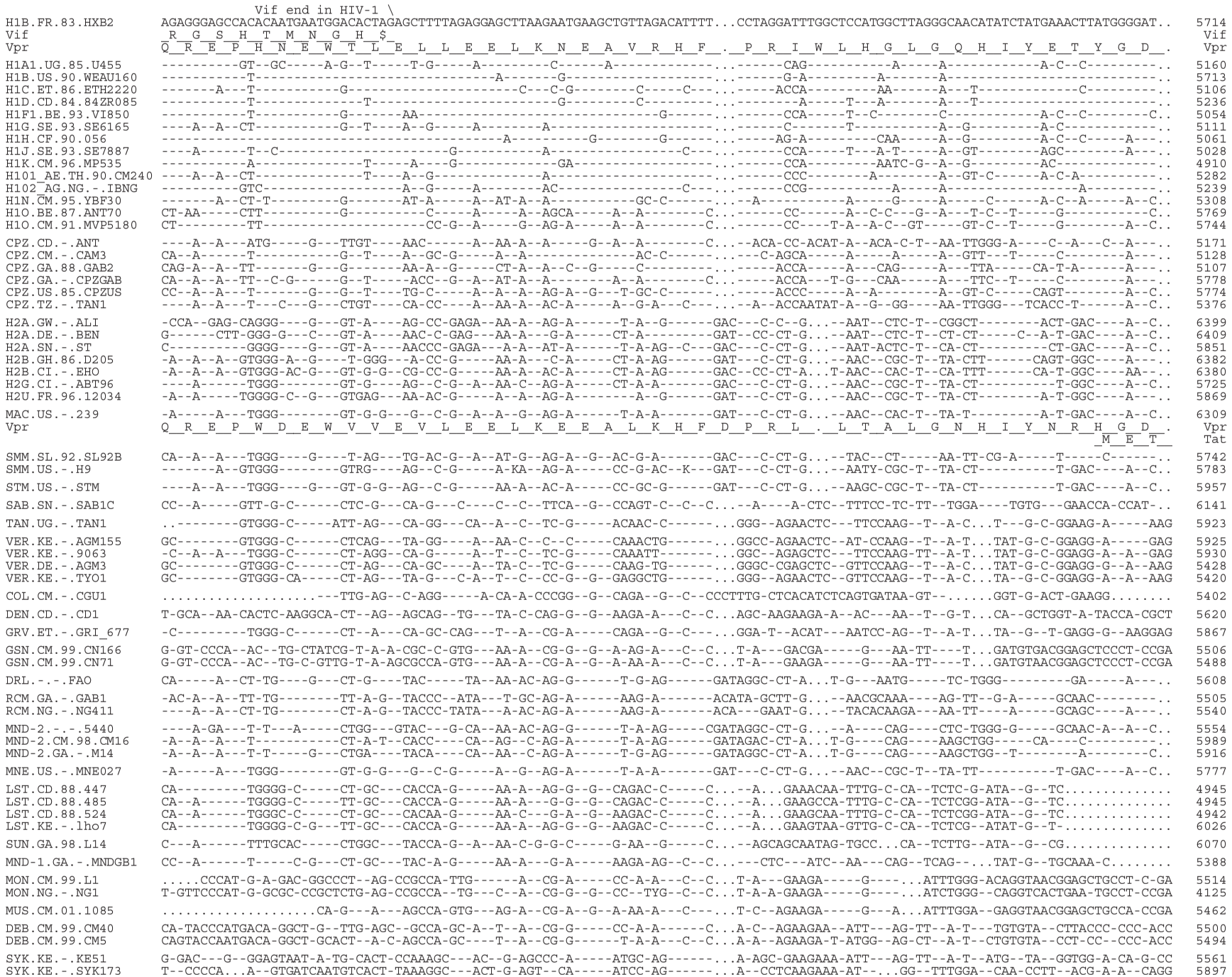
PLV
complete genomes



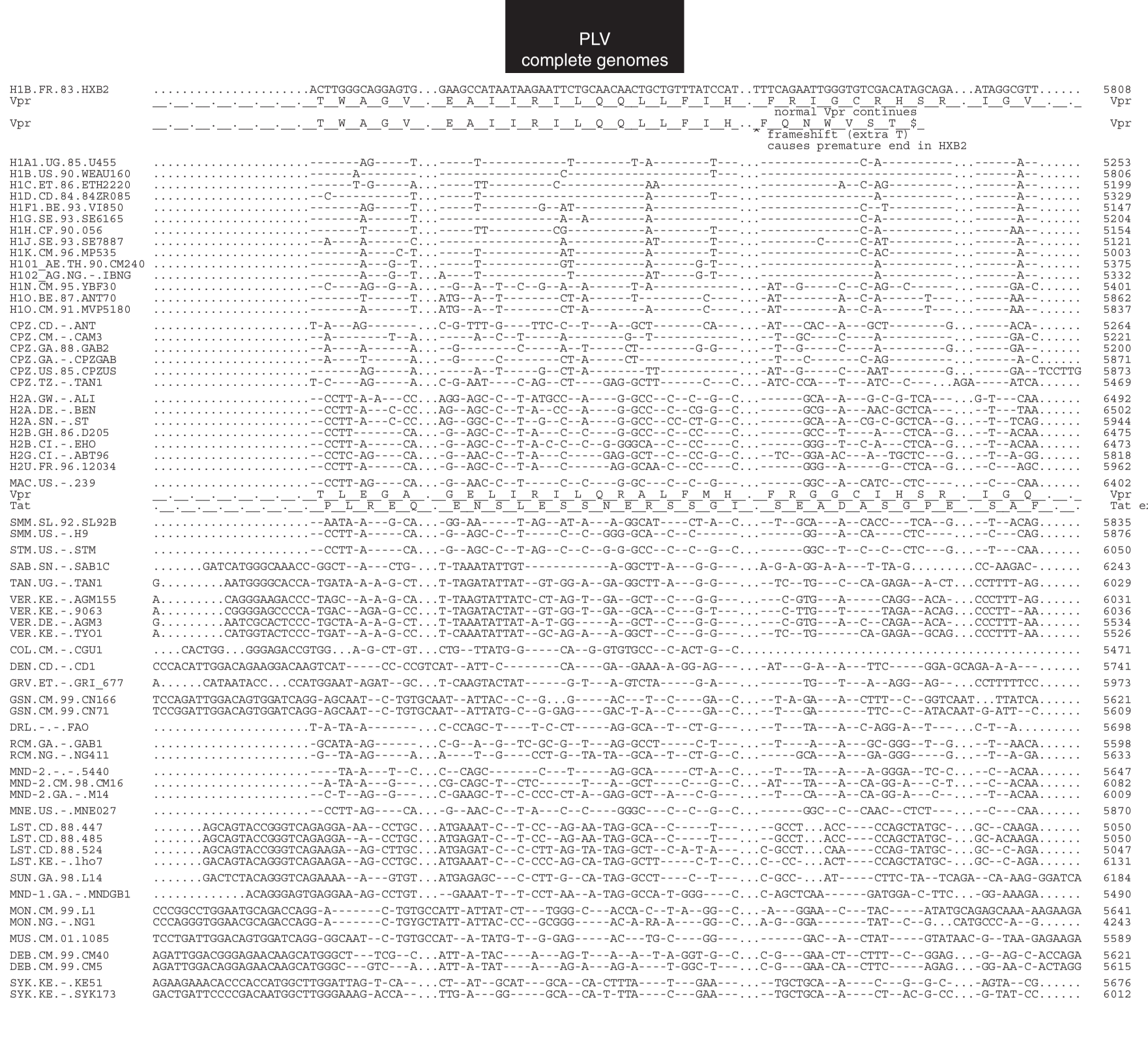


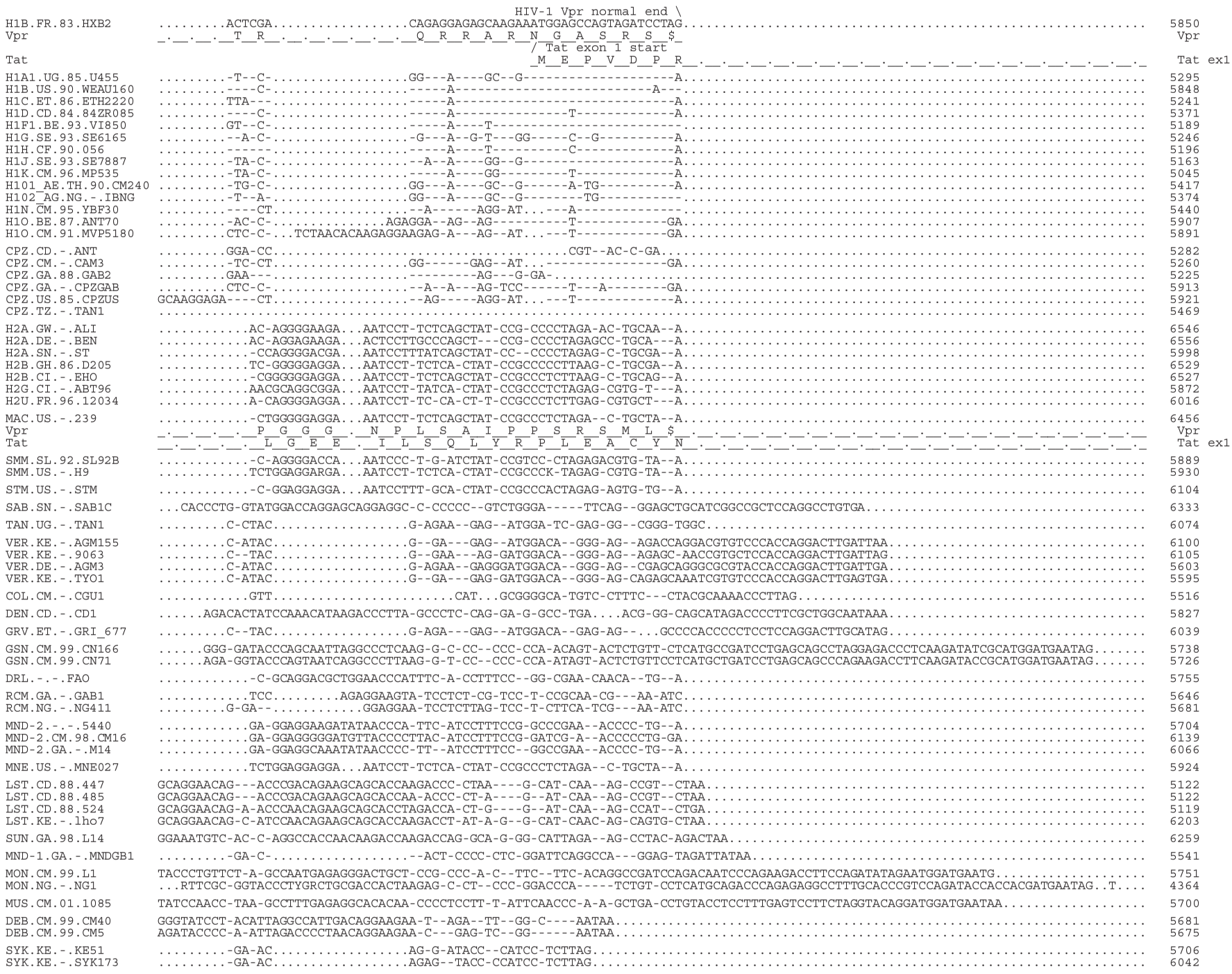
PLV
complete genomes


/ Vpr start
Vpx end in HIV-2 and SIV-SMM \

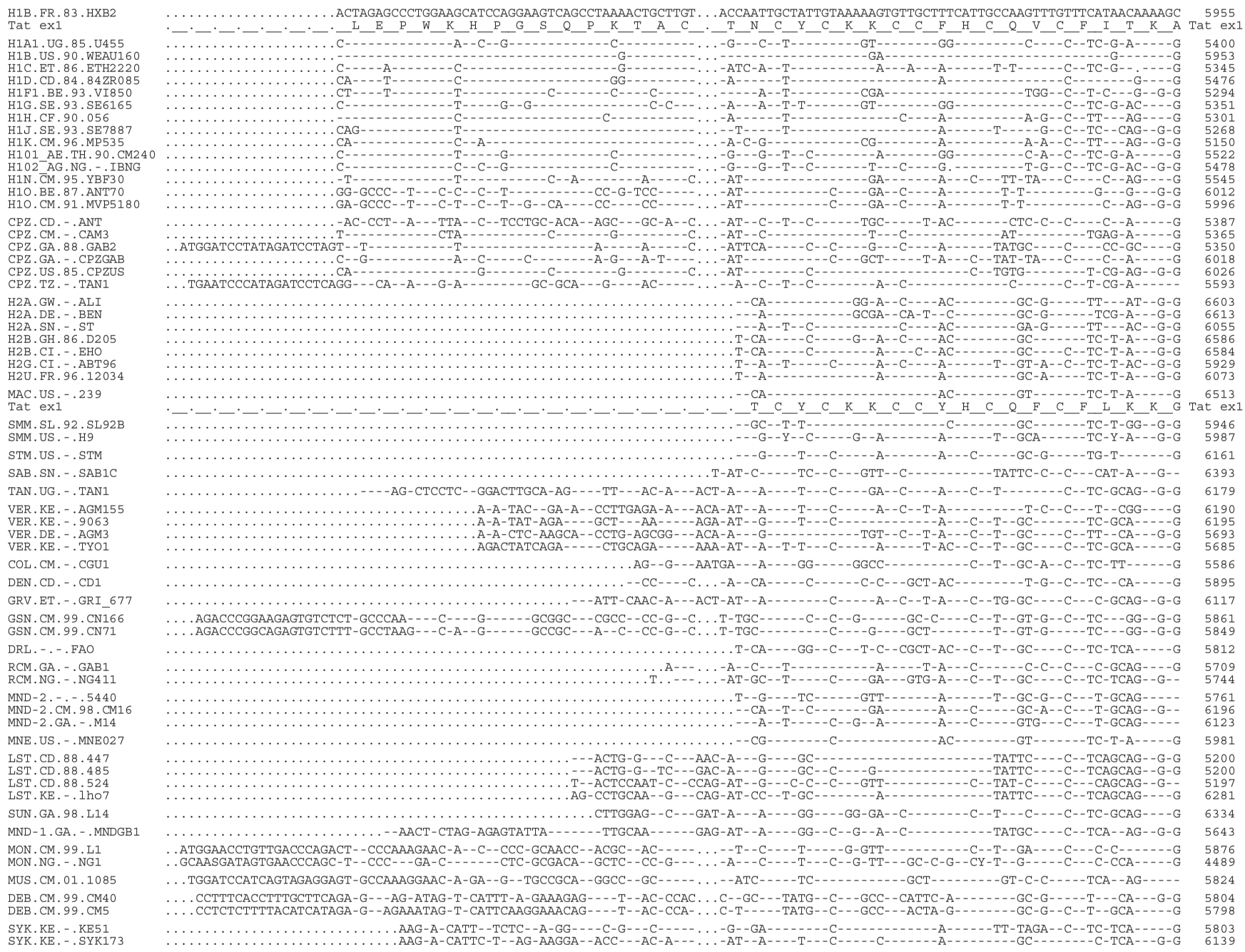


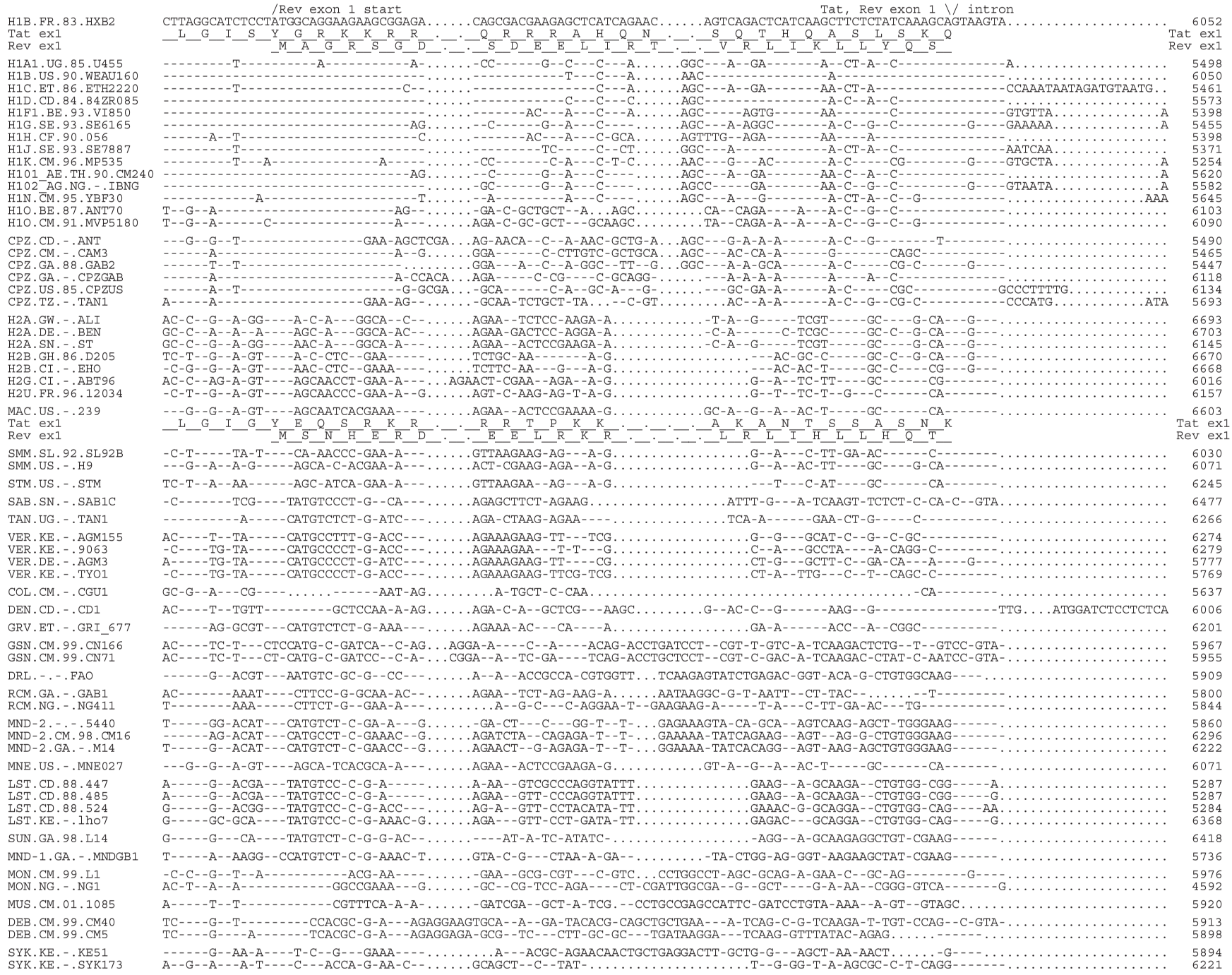
PLV
complete genomes


PLV
complete genomes
normal Vpr continues
frameshift (extra T)
causes premature end in HXB2

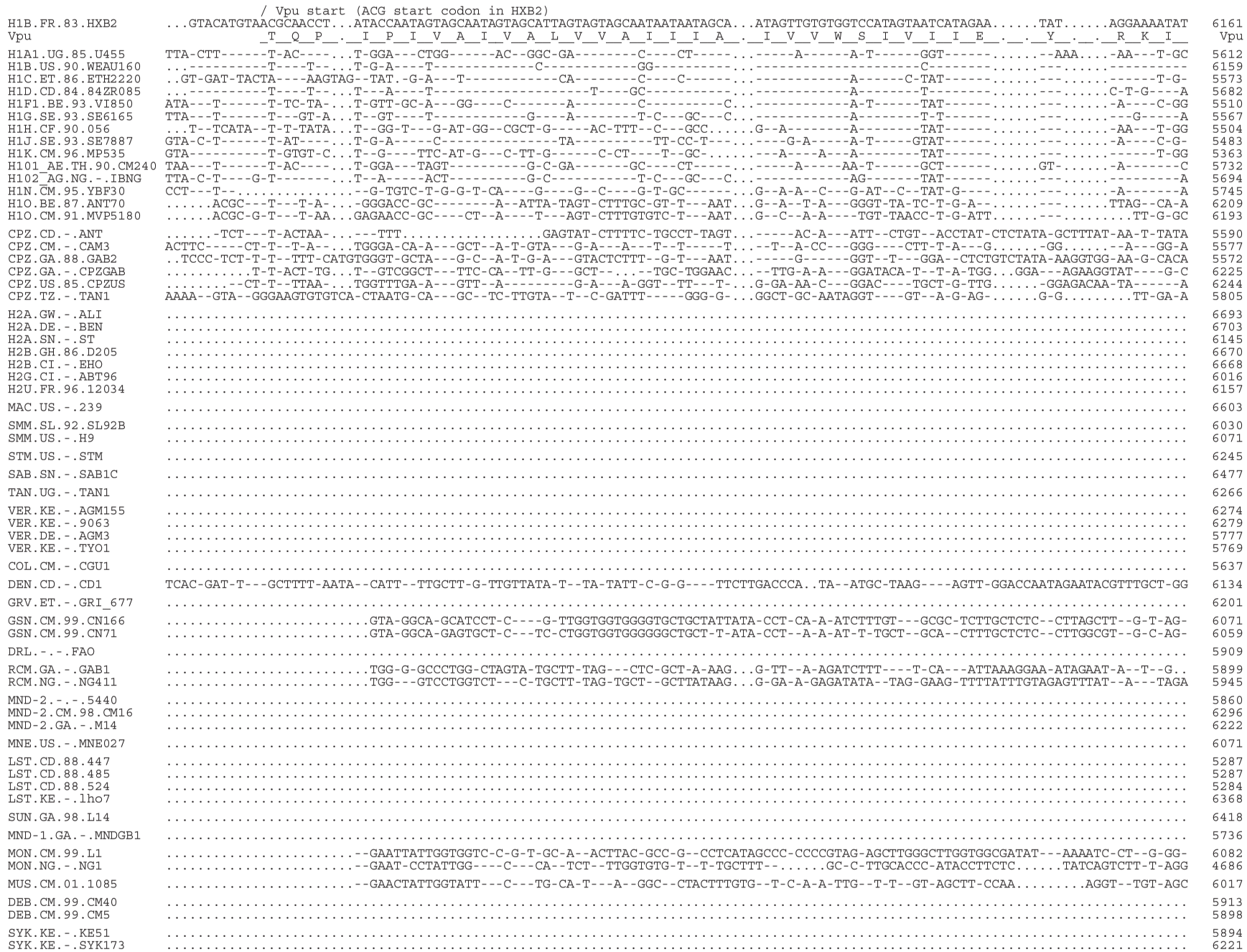


PLV
complete genomes

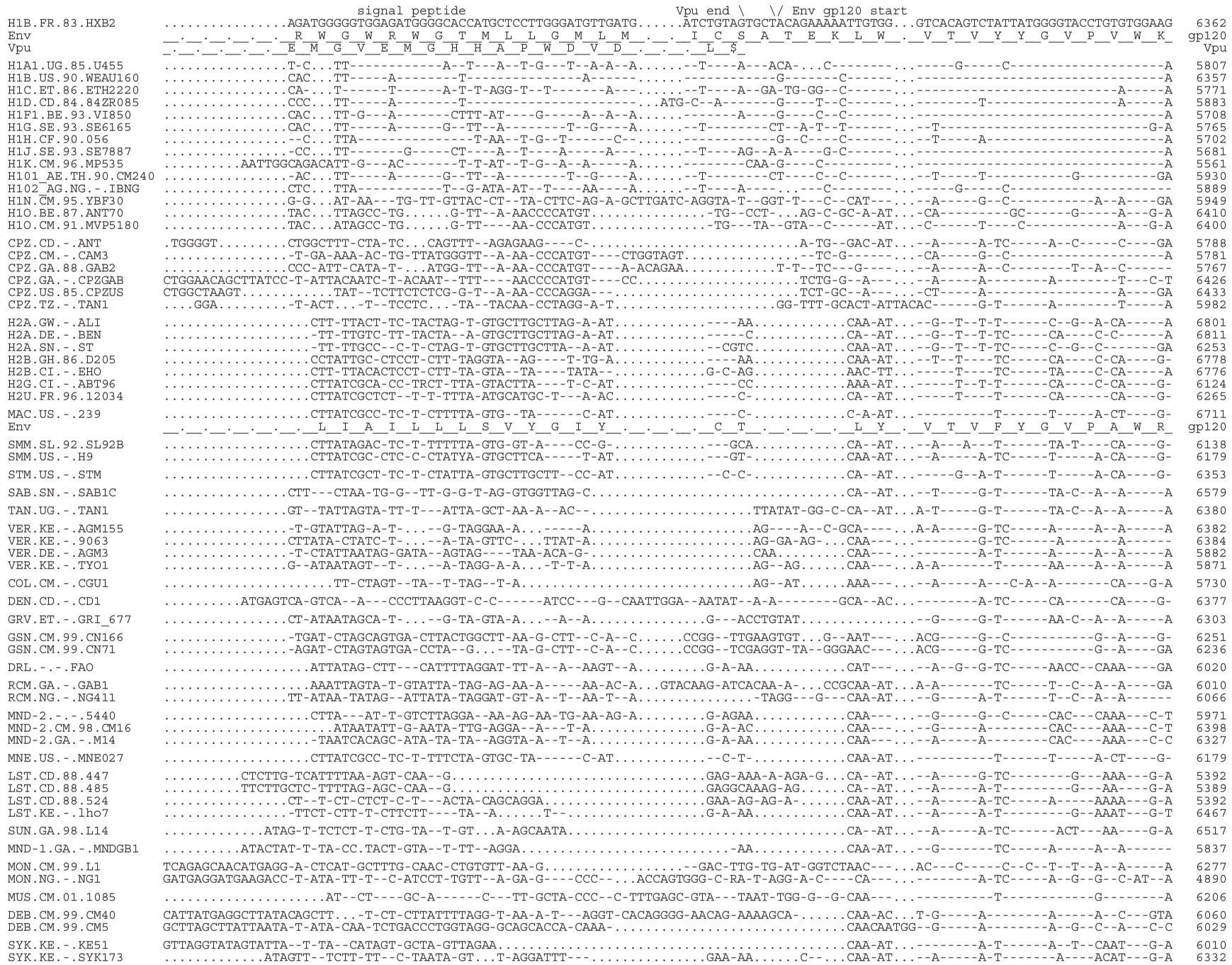


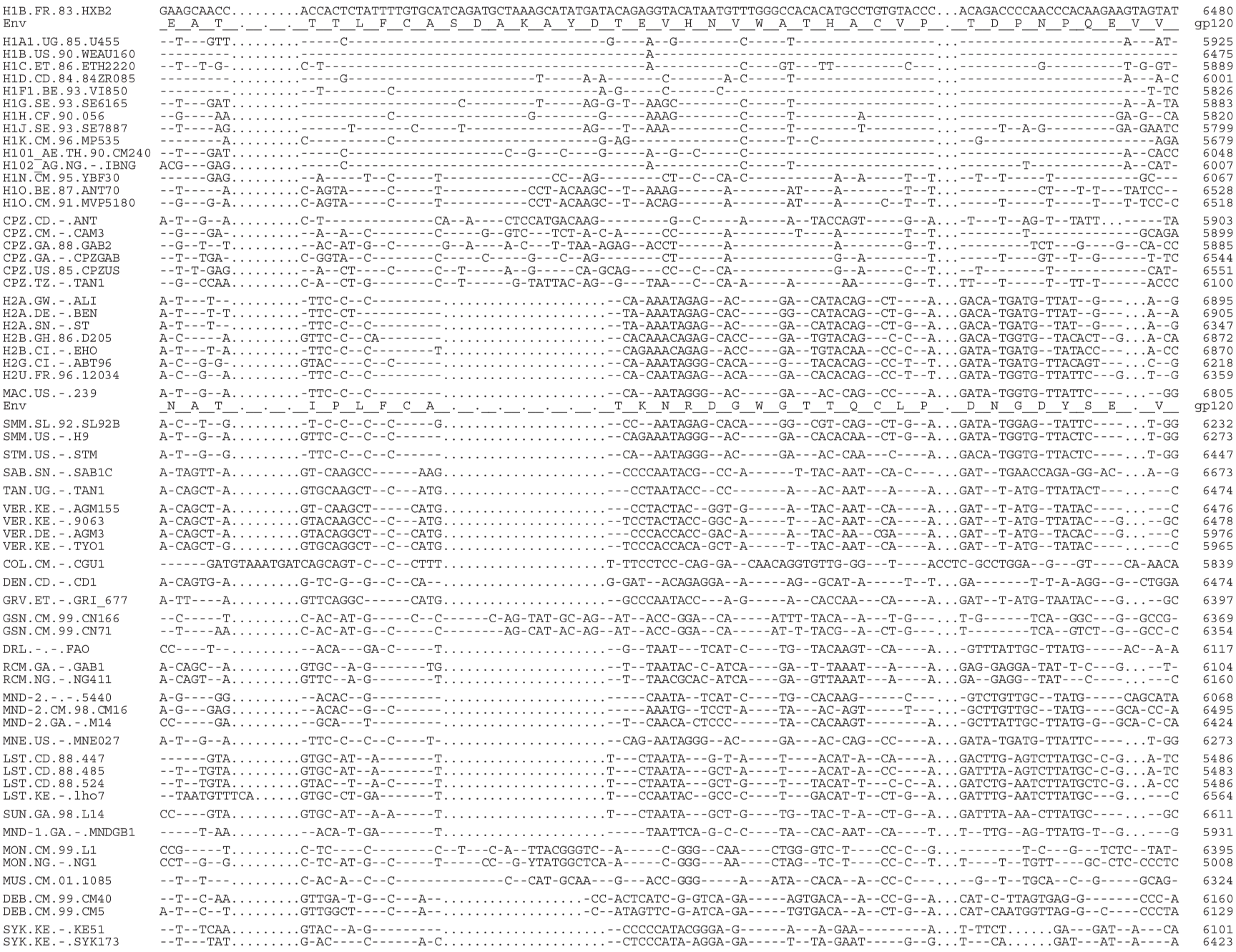


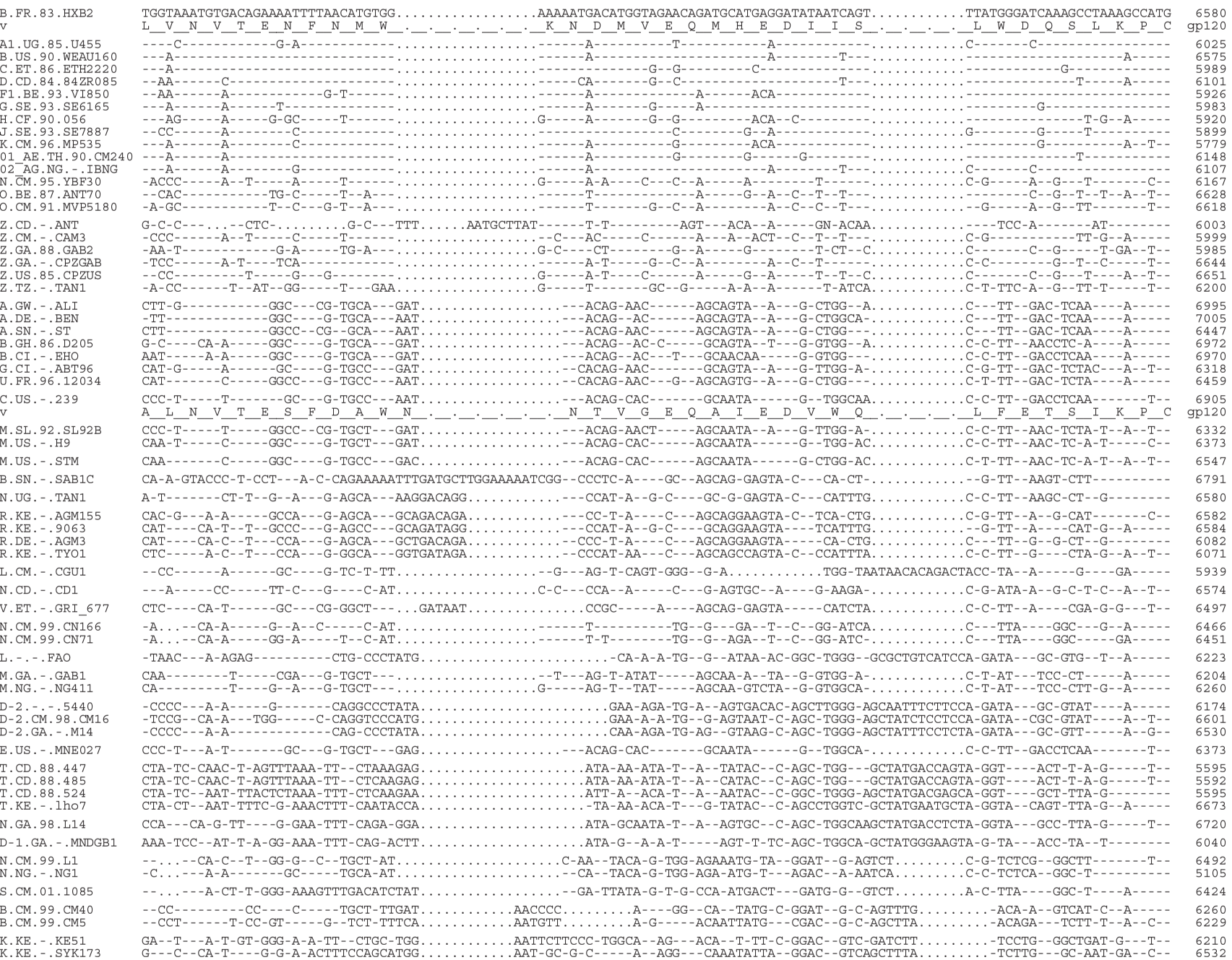
PLV
complete genomes

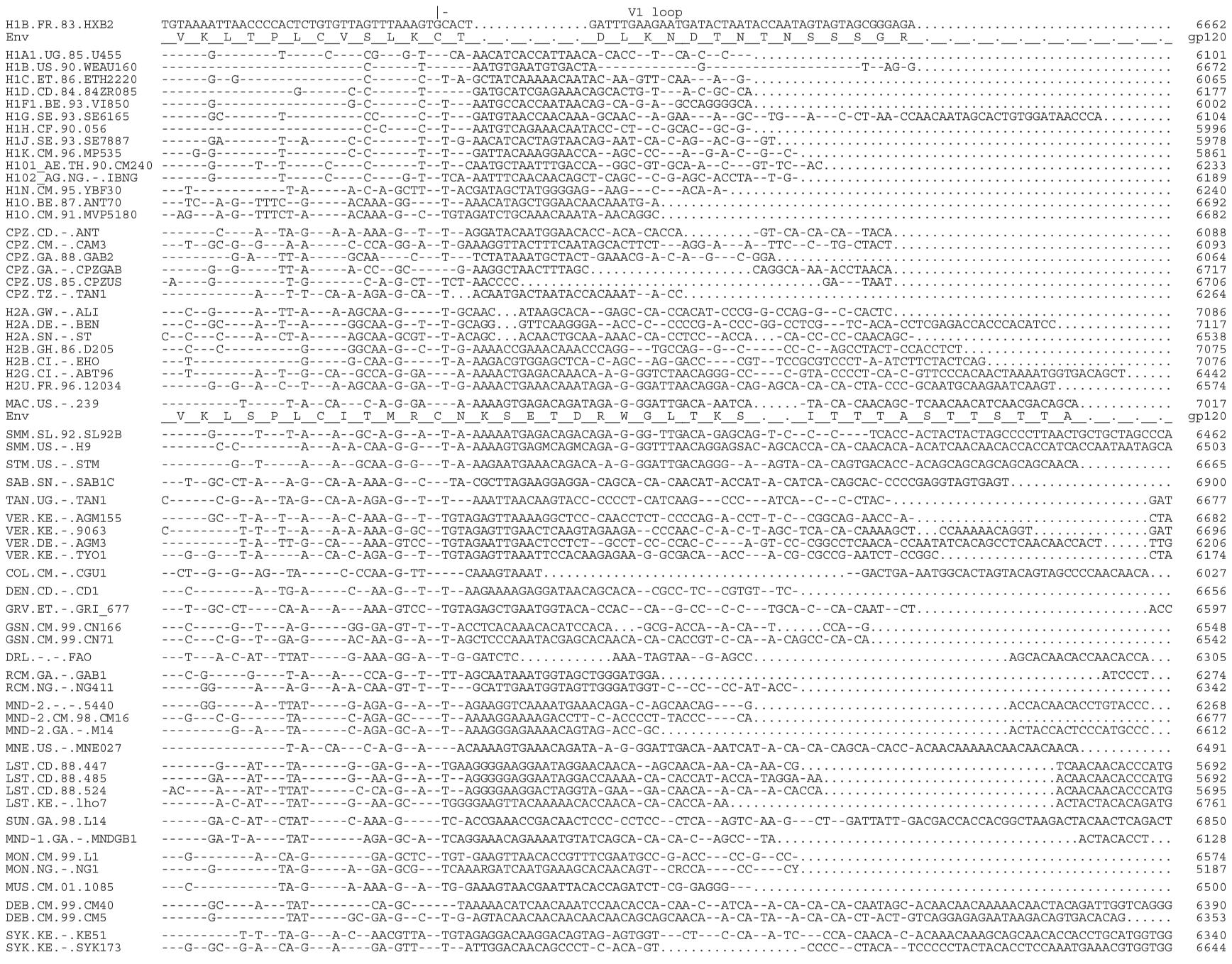

/ Vpu start (ACG start codon in HXB2)
H1B.FR.83.HXB2 ...GTACATGTAACGCAACCT...ATACCAATAGTAGCAATAGTAGCATTAGTAGTAGCAATAATAATAGCA...ATAGTTGTGTGGTCCATAGTAATCATAGAA.....TAT.....AGGAAAATAT 6161
Vpu _T__Q__P__.__I__P__I__V__A__I__V__A__L__V__V__A__I__I__I__A__.__I__V__V__W__S__I__V__I__I__E__.__.__Y__.__.__R__K__I__ Vpu
H1A1.UG.85.U455 TTA-CTT------T-AC----...T-GGA---CTGG------AC--GGC-GA--------C----CT-----...-----A------A-T------GGT------.....--AAA....-AA---T-GC 5612
H1B.US.90.WEAU160 ...----------T----T-...T-G-A----T-----------C-----------GG----------...----------C--------.....---.....---------- 6159
H1C.ET.86.ETH2220 ..GT-GAT-TACTA----AAGTAG--TAT.-G-A---T----------CA--------C----C------...----------A------C-TAT------.....---.....-----T-G- 5573
H1D.CD.84.84ZR085 ...----------T----T--...T---A----T-------------------T----GC----------...----------A--------T--------.....---.....C-T-G----A 5682
H1F1.BE.93.VI850 ATA---T------T-TC-TA-...T-GTT-GC-A---GG----C-------A--------C----------C...----------A-T------TAT------.....---.....-A----C-GG 5510
H1G.SE.93.SE6165 TTA---T------T---GT-A...T--GT----T----------G----A--------T-C---GC---C...----------A--------T-T------.....---.....---G-----A 5567
H1H.CF.90.056 ...T--TCATA--T-T-TATA...T--GG-T---G-AT-GG--CGCT-G-----AC-TTT--C---GCC.....G--A--------A--------TAT------.....---.....-AA---T-GG 5504
H1J.SE.93.SE7887 GTA-C-T------T-AT----...T-G-A-----C----------TA---------TT-CC-T---...----G-A-----A-T-----GTAT------.....---.....-A----C-G- 5483
H1K.CM.96.MP535 GTA----------T-GTGT-C...T--G----TTC-AT-G---C-TT-G------C-CT----T--GC-.....---A----A---A--------TAT------.....---.....------T-GG 5363
H101_AE.TH.90.CM240 TAA---T------T-AC----...T-GGA---TAGT---------G-C-GA-------GC----CT-----...-----A-----AA-T------GCT------.....GT-.....-A-------C 5732
H102_AG.NG.-.IBNG TTA-C-T----G-T-------...T--A-----ACT---------G-C----------T-C---GC---C...----------AG-------TAT------.....---.....---------A 5694
H1N.CM.95.YBF30 CCT---T--.....--G-TGTC-T-G-G-T-CA----G----G--C----G-T-GC------...G--A-A--C---G-AT--C--TAT-G----.....---.....-A-------A 5745
H1O.BE.87.ANT70 .....ACGC---T---T-A-...-GGGACC-GC--------A--ATTA-TAGT-CTTTGC-GT-T---AAT..G--A--T-A---GGGT-TA-TC-T-G-A--.....---.....TTAG--CA-A 6209
H1O.CM.91.MVP5180 .....ACGC-G-T---T-AA...GAGAACC-GC----CT--A----T---AGT-CTTTGTGTC-T---AAT..G--C--A-A----TGT-TAACC-T-G-ATT.....---.....TT-G-GC 6193
CPZ.CD.-.ANT .....-TCT---T-ACTAA-...---TTT.....GAGTAT-CTTTTC-TGCCT-TAGT..-----AC-A---ATT--CTGT--ACCTAT-CTCTATA-GCTTTAT-AA-T-TATA 5590
CPZ.CM.-.CAM3 ACTTC-----CT-T--T-A-...TGGGA-CA-A---GCT--A-T-GTA---G-A---A---T--T-----T...--T--A-CC---GGG----CTT-T-A---G.....-GG.....--A---GG-A 5577
CPZ.GA.88.GAB2 ..TCCC-TCT-T-T--TTT-CATGTGGGT-GCTA---G-C--A-T-G-A---GTACTCTTT--G-T---AAT..-----G------GGT--T---GGA--CTCTGTCTATA-AAGGTGG-AA-G-CACA 5572
CPZ.GA.-.CPZGAB ..........T-T-ACT-TG...T--GTCGGCT---TTC-CA--TT-G---GCT--....--TGC-TGGAAC....--TTG-A-A---GGATACA-T--T-A-TGG...GGA---AGAAGGTAT----G-C 6225
CPZ.US.85.CPZUS .....--CT-T--TTAA-...TGGTTTGA-A---GTT--A---------G-A---A-GGT--TT---T-...G-GA-AA-C---GGAC----TGCT-G-TTG.....-GGAGACAA-TA------A 6244
CPZ.TZ.-.TAN1 AAAA--GTA--GGGAAGTGTGTCA-CTAATG-CA---GC--TC-TTGTA--T--C-GATTT-----GGG-G-...GGCT-GC-AATAGGT----GT--A-G-AG-.....G-G.....TT-GA-A 5805
H2A.GW.-.ALI 6693
H2A.DE.-.BEN 6703
H2A.SN.-.ST 6145
H2B.GH.86.D205 6670
H2B.CI.-.EHO 6668
H2G.CI.-.ABT96 6016
H2U.FR.96.12034 6157
MAC.US.-.239 6603
SMM.SL.92.SL92B 6030
SMM.US.-.H9 6071
STM.US.-.STM 6245
SAB.SN.-.SAB1C 6477
TAN.UG.-.TAN1 6266
VER.KE.-.AGM155 6274
VER.KE.-.9063 6279
VER.DE.-.AGM3 5777
VER.KE.-.TYO1 5769
COL.CM.-.CGU1 5637
DEN.CD.-.CD1 TCAC-GAT-T---GCTTTT-AATA--CATT--TTGCTT-G-TTGTTATA-T--TA-TATT-C-G-G----TTCTTGACCCA..TA--ATGC-TAAG----AGTT-GGACCAATAGAATACGTTTGCT-GG 6134
GRV.ET.-.GRI_677 6201
GSN.CM.99.CN166 .....GTA-GGCA-GCATCCT-C----G-TTGGTGGTGGGGTGCTGCTATTATA-CCT-CA-A-ATCTTTGT---GCGC-TCTTGCTCTC--CTTAGCTT--G-T-AG- 6071
GSN.CM.99.CN71 .....GTA-GGCA-GAGTGCT-C---TC-CTGGTGGTGGGGGGCTGCT-T-ATA-CCT--A-A-AT-T-TGCT--GCA--CTTTGCTCTC--CTTGGCGT--G-C-AG- 6059
DRL.-.-.FAO 5909
RCM.GA.-.GAB1 .....TGG-G-GCCCTGG-CTAGTA-TGCTT-TAG---CTC-GCT-A-AAG...G-TT--A-AGATCTTT----T-CA---ATTAAAGGAA-ATAGAAT-A--T--G.. 5899
RCM.NG.-.NG411 .....TGG---GTCCTGGTCT---C-TGCTT-TAG-TGCT--GCTTATAAG...G-GA-A-GAGATATA--TAG-GAAG-TTTTATTTGTAGAGTTTAT--A---TAGA 5945
MND-2.-.-.5440 5860
MND-2.CM.98.CM16 6296
MND-2.GA.-.M14 6222
MNE.US.-.MNE027 6071
LST.CD.88.447 5287
LST.CD.88.485 5287
LST.CD.88.524 5284
LST.KE.-.lho7 6368
SUN.GA.98.L14 6418
MND-1.GA.-.MNDGB1 5736
MON.CM.99.L1 .....--GAATTATTGGTGGTC-C-G-T-GC-A--ACTTAC-GCC-G--CCTCATAGCCC-CCCCGTAG-AGCTTGGGCTTGGTGGCGATAT---AAAATC-CT--G-GG- 6082
MON.NG.-.NG1 .....--GAAT-CCTATTGG----C---CA--TCT--TTGGTGTG-T--T-TGCTTT--.....GC-C-TTGCACCC-ATACCTTCTC.....TATCAGTCTT-T-AGG 4686
MUS.CM.01.1085 .....--GAACTATTGGTATT---C---TG-CA-T---A--GGC--CTACTTTGTG--T-C-A-A-TTG--T-T--GT-AGCTT-CCAA.....AGGT--TGT-AGC 6017
DEB.CM.99.CM40 5913
DEB.CM.99.CM5 5898
SYK.KE.-.KE51 5894
SYK.KE.-.SYK173 6221

PLV
complete genomes

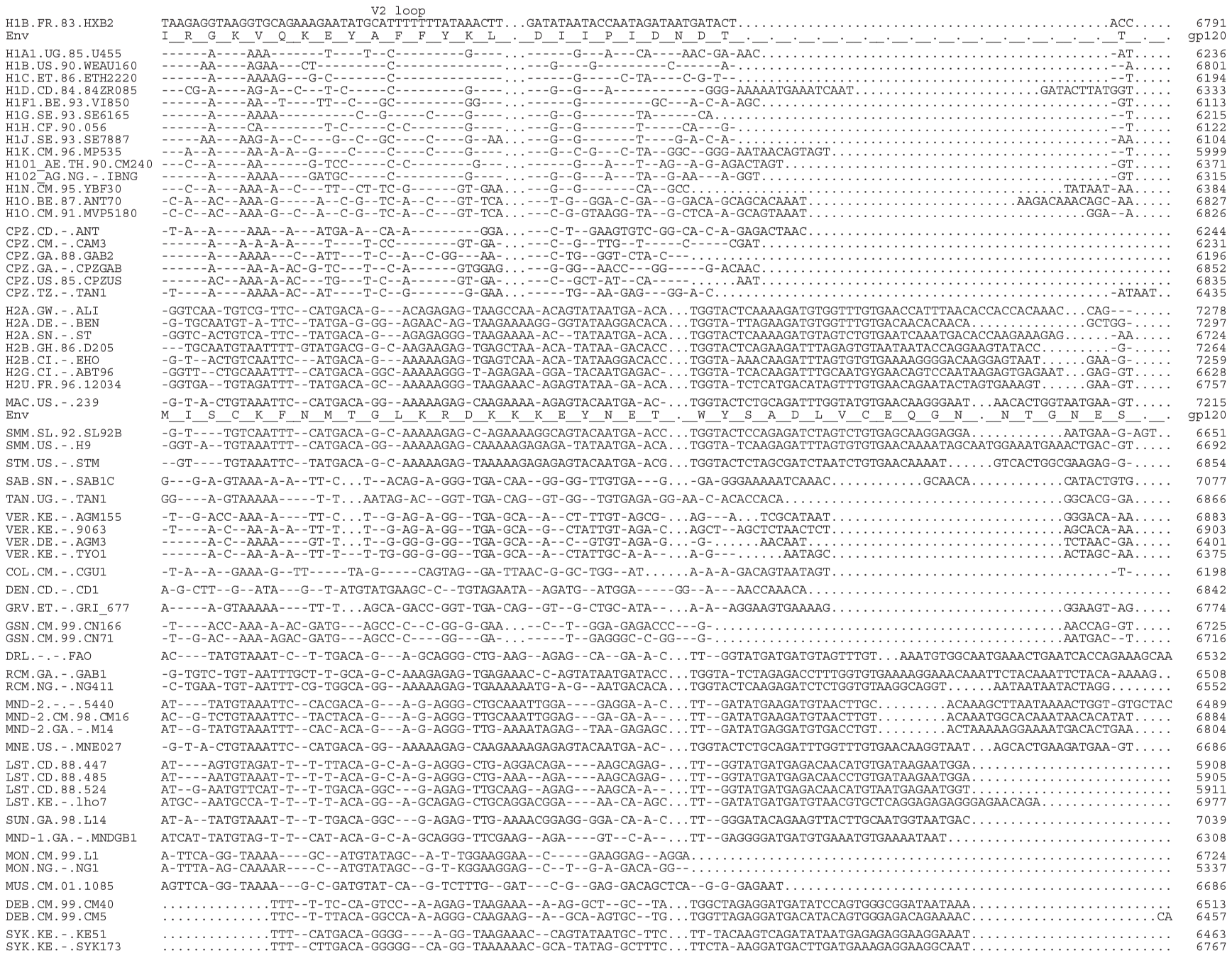

signal peptide
Vpu end \
\/ Env gp120 start
H1B.FR.83.HXB2
Env
Vpu
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Env
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173


H1B.FR.83.HXB2 GAAGCAACC.....ACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCC...ACAGACCCCAACCCACAAGAAGTAGTAT 6480
Env _E__A__T__.__.__.__T__T__L__F__C__A__S__D__A__K__A__Y__D__T__E__V__H__N__V__W__A__T__H__A__C__V__P__.__T__D__P__N__P__Q__E__V__V__ gp120
H1A1.UG.85.U455 --T---GTT.....------C-----------------------------------G----A--G--------C-----T-----------------...---------------------A---AT- 5925
H1B.US.90.WEAU160 ---------.....---A-----------------------------------...------------------------------ 6475
H1C.ET.86.ETH2220 --T--T-G-.....C-T-----------------------------------A----------C----GT---TT---------C---...----------G----------T-G-GT- 5889
H1D.CD.84.84ZR085 ---------.....-----G-----------------T-----A-A-------C-------A-C-----T-----------------...---------------------A---A-C 6001
H1F1.BE.93.VI850 ---------.....--T--------C-----------------------A-G-----C------C-----------------------...-----------------------T-TC 5826
H1G.SE.93.SE6165 --T---GAT.....-----------------------C---------T-----AG-G-T--AAGC--------C-----T-----------------...---------------------A--A-TA 5883
H1H.CF.90.056 --G----AA.....----------C-----------------G--------G-----AAAG--------C-----T--------A---------...--------------------GA-G--CA 5820
H1J.SE.93.SE7887 --T----AG.....------T-------C-----T---------------AG---T--AAA---------C-----T--------T---------...-----T--A-G---------GA-GAATC 5799
H1K.CM.96.MP535 --------A.....C----------C-----------------------G-AG----------C-----T--C-----------------...--G---------------------AGA 5679
H101_AE.TH.90.CM240 --T---GAT.....-----C-------------------C--G---C----G-----A--G--C-----C-----------------...---------------------A--CACC 6048
H102_AG.NG.-.IBNG ACG---GAG.....-----C-----------------------------------A----------C-----T----------------...--------T----------A--CAT- 6007
H1N.CM.95.YBF30 ------GAG.....--A-----T--C-----T-------------CC--AG--------CT--C--CA-C----------A--A-----T--T...--T--T-----T----------GC--- 6067
H1O.BE.87.ANT70 --T-----A.....C-AGTA-----C-----T----------CCT-ACAAGC--T--AAAG------A-------AT----A--------T--T...----------CT---T-T---TATCC-- 6528
H1O.CM.91.MVP5180 --G---G-A.....C-AGTA-----C-----T----------CCT-ACAAGC--T--ACAG------A-------AT----A-----C--T--T...------T-----T-----T---T-TCC-C 6518
CPZ.CD.-.ANT A-T--G--A.....C-T-------------CA--A---CTCCATGACAAG---------G--C----A-----A--TACCAGT----G--A...-T---T--AG-T--TATT...-----TA 5903
CPZ.CM.-.CAM3 --G---GA-.....--A--A--C--------C-----G--GTC---TCT-A-C-A-----CC------A--------T----A--------T--T...--------T--------------GCAGA 5899
CPZ.GA.88.GAB2 --G--T--T.....--AC-AT-G--C------G--A----A-C---T-TAA-AGAG--ACCT------A----------A--------G--T...----------TCT---G----G---CA-CC 5885
CPZ.GA.-.CPZGAB --T--TGA-.....C-GGTA--C--------C-----C-----G---C--AG--------CT------A----------G--A--------T...------T----GT--T--G------T-TC 6544
CPZ.US.85.CPZUS --T-T-GAG.....--A--CT----C-----C--T-----A--G-----CA-GCAG----CC--C--CA----------G--A-----C--T...--T--------T----------CAT- 6551
CPZ.TZ.-.TAN1 --G--CCAA.....C-A--CT-G--------C--T------G-TATTAC-AG--G---TAA---C--CA-A-----A------AA------G--T...TT---T-----T--TT-T------ACCC 6100
H2A.GW.-.ALI A-T---T-.....-TTC-C--C---------.....--CA-AAATAGAG--AC-----GA--CATACAG--CT----A...GACA-TGATG-TTAT--G---...A--G 6895
H2A.DE.-.BEN A-T---T-T.....-TTC-CT----------.....--TA-AAATAGAG-CAC-----GG--CATACAG--CT-G--A...GACA-TGATG-TTAT--G---...A--A 6905
H2A.SN.-.ST A-T---T--.....-TTC-C--C---------.....--TA-AAATAGAG--AC-----GA--CATACAG--CT-G--A...GACA-TGATG--TAT--G----...A--G 6347
H2B.GH.86.D205 A-C-----A.....GTTC-C--CA--------.....--CACAAACAGAG-CACC----GA--TGTACAG---C-C--A...GACA-TGGTG--TACACT--G...A-CA 6872
H2B.CI.-.EHO A-T---T-A.....-TTC-C--C--------T.....--CAGAAACAGAG--ACC----GA--TGTACAA--CC-C--A...GATA-TGATG--TATACC---...A-CC 6870
H2G.CI.-.ABT96 A-C--G-G-.....GTAC----C--C------.....--CA-AAATAGGG-CACA----G---TACACAG--CC-T--T...GATA-TGATG-TTACAGT---...C--G 6218
H2U.FR.96.12034 A-C--G--A.....-TTC-C--C---------.....--CA-CAATAGAG--ACA----GA--CACACAG--CC-T--T...GATA-TGGTG-TTATTC--G...T--G 6359
MAC.US.-.239 A-T--G--A.....-TTC-C--C---------.....--CA--AATAGGG--AC-----GA---AC-CAG--CC----A...GATA-TGGTG-TTATTC----...--GG 6805
Env _N__A__T__.__.__.__I__P__L__F__C__A__.__.__.__.__.__.__T__K__N__R__D__G__W__G__T__T__Q__C__L__P__.__D__N__G__D__Y__S__E__.__V__ gp120
SMM.SL.92.SL92B A-C--T--G.....-T-C-C--C--C-----G.....--CC--AATAGAG-CACA----GG--CGT-CAG--CT-G--A...GATA-TGGAG--TATTC----...T-GG 6232
SMM.US.-.H9 A-T--G--A.....GTTC-C--C--C------.....--CAGAAATAGGG--AC-----GA--CACACAA--CT-G--A...GATA-TGGTG-TTACTC----...T-GG 6273
STM.US.-.STM A-T--G--G.....-TTC-C--C--C------.....--CA--AATAGGG--AC-----GA---AC-CAA---C----A...GACA-TGGTG-TTACTC----...T-GG 6447
SAB.SN.-.SAB1C A-TAGTT-A.....GT-CAAGCC------AAG.....--CCCCAATACG--CC-A------T-TAC-AAT--CA-C---...GAT--TGAACCAGA-GG-AC-...A--G 6673
TAN.UG.-.TAN1 A-CAGCT-A.....GTGCAAGCT--C---ATG.....---CCTAATACC--CC-------A---AC-AAT---A----A...GAT--T-ATG-TTATACT---...---C 6474
VER.KE.-.AGM155 A-CAGCT-A.....GT-CAAGCT-----CATG.....---CCTACTAC--GGT-G-----A--TAC-AAT--CA----A...GAT--T-ATG--TATAC----...---C 6476
VER.KE.-.9063 A-CAGCT-A.....GTACAAGCC--C---ATG.....--TCCTACTACC-GGC-A-----T---AC-AAT--CA----A...GAT--T-ATG-TTATAC---G...--GC 6478
VER.DE.-.AGM3 A-CAGCT-A.....GTACAGGCT--C--CATG.....---CCCACCACC-GAC-A-----A--TAC-AA--CGA----A...GAT--T-ATG--TACAC---G...---C 5976
VER.KE.-.TYO1 A-CAGCT-G.....GTGCAGGCT--C--CATG.....--TCCCACCACA-GCT-A-----T--TAC-AAT--CA----A...GAT--T-ATG--TATAC----...---C 5965
COL.CM.-.CGU1 ------GATGTAAATGATCAGCAGT-C--C--CTTT.....T-TTCCTCC-CAG-GA--CAACAGGTGTTG-GG---T----ACCTC-GCCTGGA--G---GT----CA-AACA 5839
DEN.CD.-.CD1 A-CAGTG-A.....G-TC-G--G--C--CA--.....G-GAT--ACAGAGGA--A-----AG--GCAT-A-----T--T...GA-------T-T-A-AGG--G--CTGGA 6474
GRV.ET.-.GRI_677 A-TT----A.....GTTCAGGC------CATG.....--GCCCAATACC---A-G-----A--CACCAA---CA----A...GAT--T-ATG-TAATAC---G...--GC 6397
GSN.CM.99.CN166 --C-----T.....C-AC-AT-G-----C--C------C-AG-TAT-GC-AG--AT--ACC-GGA--CA-----ATTT-TACA--A---T-G---...T-G------TCA--GGC---G--GCCG- 6369
GSN.CM.99.CN71 --T----AA.....C-AC-AT-G--C-----C--------AG-CAT-AC-AG--AT--ACC-GGA--CA-----AT-T-TACG--A--CT-G---...T--------TCA--GTCT--G--GCC-C 6354
DRL.-.-.FAO CC----T--.....--ACA---GA-C-----T.....--G--TAAT---TCAT-C----TG--TACAAGT--CA----A...-GTTTATTGC-TTATG-----AC--A-A 6117
RCM.GA.-.GAB1 A-CAGC--A.....GTGC--A-G-------TG.....--T--TAATAC-C-ATCA----GA--T-TAAAT---A----A...GAG-GAGGA-TAT-T-C---G...--T- 6104
RCM.NG.-.NG411 A-CAGT--A.....GTTC--A-G--------T.....--T--TAACGCAC-ATCA----GA--GTTAAAT---A----A...GA--GAGG--TAT---C----...---C 6160
MND-2.-.-.5440 A-G----GG.....--ACAC--G---------.....-----CAATA--TCAT-C----TG--CACAAG------C---...-GTCTGTTGC--TATG-----CAGCATA 6068
MND-2.CM.98.CM16 A-G---GAG.....--ACAC--G--C------.....-----AAATG--TCCT-A----TA---AC-AGT-----T---...-GCTTGTTGC--TATG----GCA-CC-A 6495
MND-2.GA.-.M14 CC-----GA.....--GCA---T---------.....--T--CAACA-CTCCC------TA--CACAAGT--------A...-GCTTATTGC-TTATG-G--GCA-C-CA 6424
MNE.US.-.MNE027 A-T--G--A.....-TTC-C--C--C----T-.....--CAG-AATAGGG--AC-----GA---AC-CAG--CC----A...GATA-TGATG-TTATTC------...T-GG 6273
LST.CD.88.447 ------GTA.....GTGC-AT--A-------T.....T---CTAATA---G-T-A----T----ACAT-A--CA----A...GACTTG-AGTCTTATGC-C-G...A-TC 5486
LST.CD.88.485 --T--TGTA.....GTGC-AT--A-------T.....T---CTAATA---GCT-A----T----ACAT-A--CC----A...GATTTA-AGTCTTATGC-C-G...A-TC 5483
LST.CD.88.524 --T--TGTA.....GTAC--T--A-C-----T.....T---CTAATA---GCT-G----T---TACAT-T---C-C--A...GATCTG-AATCTTATGCTC-G...A-CC 5486
LST.KE.-.lho7 --TAATGTTTCA.....GTGC-CT-GA-------T.....T---CCAATAC--GCC-C----T---GACAT-T--CT-G--A...GATTTG-AATCTTATGC---G...---C 6564
SUN.GA.98.L14 CC----GTA.....GTGC-AT--A-A-----T.....T---CTAATA---GCT-G----TG--TAC-T-A--CT-G--A...GATTTA-AA-CTTATGC----...--GC 6611
MND-1.GA.-.MNDGB1 -----T-AA.....--ACA-T-GA-------T.....-----TAATTCA-G-C-C----TA--CAC-AAT--CA----T...T--TTG--AG-TTATG-T--G...---G 5931
MON.CM.99.L1 CCG-----T.....C-TC----C--------C--T---C-A--TTACGGGTC--A-----C-GGG--CAA----CTGG-GTC-T----CC-C--G...---------T-C---G---TCTC--TAT- 6395
MON.NG.-.NG1 CCT--G--G.....C-TC-AT-G--C-----T-----CC--G-YTATGGCTCA-A-----C-GGG---AA----CTAG--TC-T----CC-C--T...T----T--TGTT----GC-CTC-CCCTC 5008
MUS.CM.01.1085 --T--T-----.....C-AC-A--C--C-------------C--CAT-GCAA---G---ACC-GGG-----A---ATA--CACA--A--CC-C----...--G--T--TGCA--C--G-----GCAG- 6324
DEB.CM.99.CM40 --T--C-AA.....GTTGA-T-G--C---A--.....-CC-ACTCATC-G-GTCA-GA-----AGTGACA--A--CC-G--A...CAT-C-TTAGTGAG-G-------CCC-A 6160
DEB.CM.99.CM5 A-T--C--T.....GTTGGCT----C---A--.....C--ATAGTTC-G-ATCA-GA-----TGTGACA--A--CT-G--A...CAT-CAATGGTTAG-G--C----CCCTA 6129
SYK.KE.-.KE51 --T--TCAA.....GTAC--A-G---------.....--CCCCCATACGGGA-G------A---A-GAA---------A...T-TTCT......GA---GAT--A---CA 6101
SYK.KE.-.SYK173 --T---TAT.....G-AC----C------A-C.....--CTCCCATA-AGGA-GA-----T--TA-GAAT-----G--G...T---CA......GAT---AT--A----A 6423
PLV
complete genomes

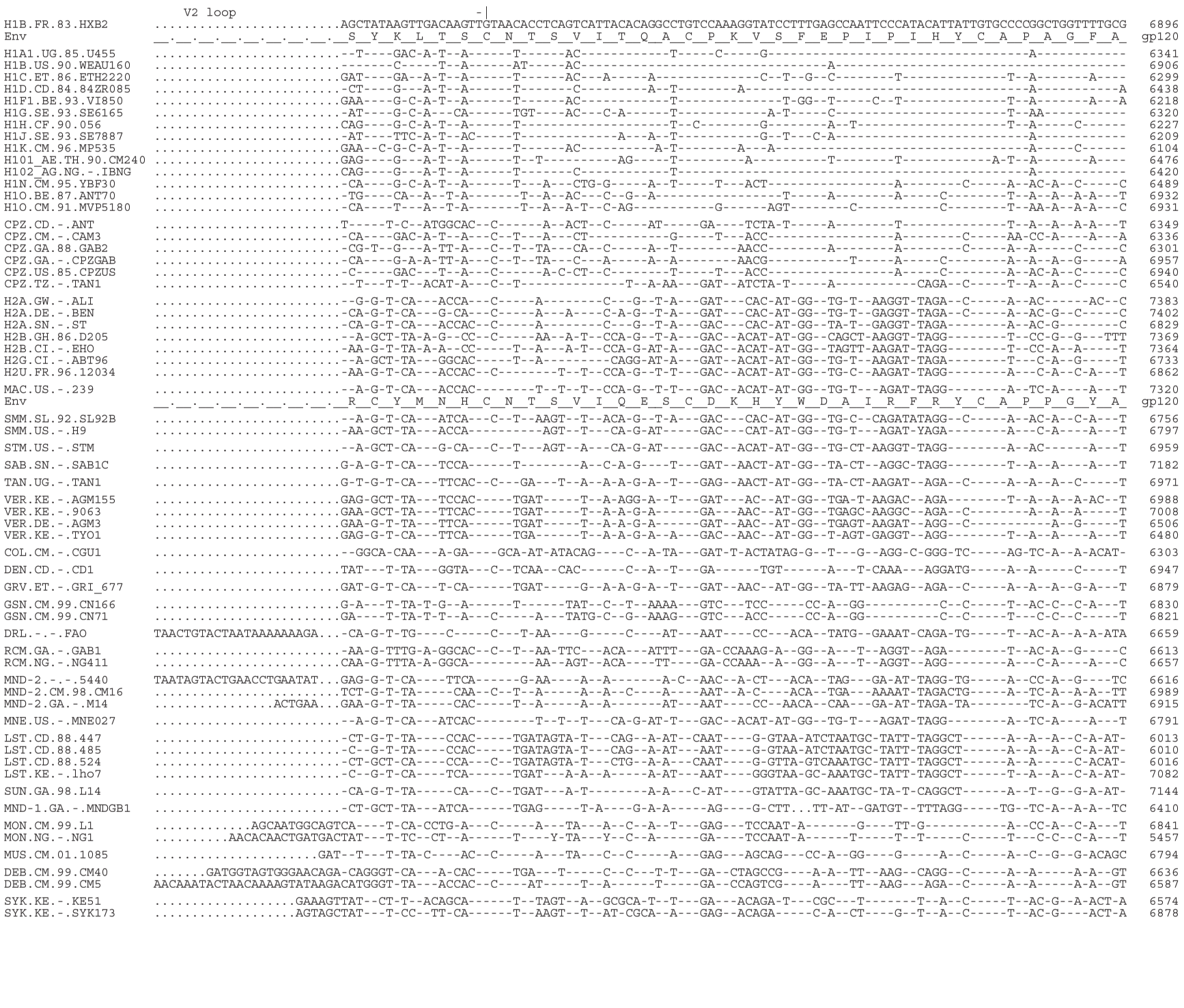

H1B.FR.83.HXB2 TGGTAAATGTGACAGAAAATTTTAACATGTGG.....AAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGT.....TTATGGGATCAAAGCCTAAAGCCATG 6580
Env L__V__N__V__T__E__N__F__N__M__W__.__.__.__.__.__K__N__D__M__V__E__Q__M__H__E__D__I__I__S__.__.__.__.__L__W__D__Q__S__L__K__P__C gp120
H1A1.UG.85.U455 ----C----------G-A----------.....A---------T----------A----------.....C-------C---------- 6025
H1B.US.90.WEAU160 ---A----------.....A----------A---------T---.....----------A----- 6575
H1C.ET.86.ETH2220 ---A----------.....G--G---------C----------.....----------G---------- 5989
H1D.CD.84.84ZR085 --AA------C----------.....CA------G--C----------T---.....----------T-----A----- 6101
H1F1.BE.93.VI850 --AA------A----------G-T------.....A----------A------ACA----------.....---------- 5926
H1G.SE.93.SE6165 ---A------A------T----------.....A------G----A----------.....---------G---------- 5983
H1H.CF.90.056 ---AG-----A-----G-GC-----T------.....G-----A------G--G---------ACA--C---------.....----------T-G--A----- 5920
H1J.SE.93.SE7887 --CC------A--------C----------.....C--------G--A----------.....G--------G-----T---------- 5899
H1K.CM.96.MP535 ---A------A--------C----------.....A---------G---------ACA----------.....---------G---------A--T-- 5779
H101_AE.TH.90.CM240 ---A------A----------.....A---------G---------G------G----------.....----------T---------- 6148
H102_AG.NG.-.IBNG ---A------A--------G----------.....A----------A---------T---.....C-------C---------- 6107
H1N.CM.95.YBF30 -ACCC-----A--T-----A-----T------.....G-----A-A----C---C--A-----A--------T-----.....C-G-----A--G---T-------C-- 6167
H1O.BE.87.ANT70 --CAC----------TG-C-----T--A---.....------T----------A-----G--A--C--T--T---.....--------C--G--T--T--A--T-- 6628
H1O.CM.91.MVP5180 -A-GC----------T--C---G-T--A---.....------T-------G--C--A--------A--C--C--T---.....--G-----A--G--TT-------T-- 6618
CPZ.CD.-.ANT G-C-C---...--CTC-.....G-C---TTT.....AATGCTTAT------T-T---------AGT---ACA--A-----GN-ACAA.....----TCC-A-------AT-------- 6003
CPZ.CM.-.CAM3 -CCC------A--T-----C-----T------.....---C---AC-----C------A---A--ACT--C--T--T---.....C-G----------TT-G--A----- 5999
CPZ.GA.88.GAB2 -AA-T----------G-A----TG-A---.....G-C----CT--------G--A-----G--------T-CT--C.....C--------C--G-----T-GA--T-- 5985
CPZ.GA.-.CPZGAB -TCC------A-T----TCA----------.....------A-T-----G--C--A--------A--C--T----------.....C-C--------G--T--C-----T-- 6644
CPZ.US.85.CPZUS --CC---------T-----G---G--------.....G-----A-T----C------A-----G--A-----T--T--C.....C-------C--G---T----A--T-- 6651
CPZ.TZ.-.TAN1 -A-CC------T--AT--GG-----T---GAA.....G-----T-------GC--G-----A-A--A------T-ATCA.....C-T-TTC-A--G--TT-T-----T-- 6200
H2A.GW.-.ALI CTT-G----------GGC---CG-TGCA---GAT.....---ACAG-AAC------AGCAGTA--A---G-CTGG--A.....C---TT--GAC-TCAA----A----- 6995
H2A.DE.-.BEN -TT----------GGC----G-TGCA---AAT.....---ACAG--AC------AGCAGTA--A---G-CTGGCA-.....C---TT--GAC-TCAA----A----- 7005
H2A.SN.-.ST CTT----------GGCC--CG--GCA---AAT.....---ACAG-AAC------AGCAGTA--A---G-CTGG---.....C---TT--GAC-TCAA----A----- 6447
H2B.GH.86.D205 G-C----CA-A-----GGC----G-TGCA---GAT.....---ACAG--AC-C-----GCAGTA--T---G-GTGG--A.....C-C-TT--AACCTC-A----A----- 6972
H2B.CI.-.EHO AAT-----A-A-----GGC----G-TGCA---GAT.....---ACAG--AC---T---GCAACAA-----G-GTGG---.....C-C-TT--GACCTCAA----A----- 6970
H2G.CI.-.ABT96 CAT-G-----A------GC----G-TGCC---GAT.....--CACAG-AAC------GCAGTA--A---G-TTGG-A-.....C-G-TT--GAC-TCTAC---A--T-- 6318
H2U.FR.96.12034 CAT-------C-----GGCC---G-TGCC---AAT.....--CACAG-AAC---G--AGCAGTG--A---G-CTGG---.....C-T-TT--GAC-TCTA----A----- 6459
MAC.US.-.239 CCC-T-----T-------GC---G-TGCC---AAT.....---ACAG-CAC-------GCAATA------G--TGGCAA.....C-C-TT--GACCTCAA-------T-- 6905
Env A__L__N__V__T__E__S__F__D__A__W__N__.__.__.__.__.__N__T__V__G__E__Q__A__I__E__D__V__W__Q__.__.__.__.__L__F__E__T__S__I__K__P__C gp120
SMM.SL.92.SL92B CCC-T-----T------GGCC--CG-TGCT---GAT.....---ACAG-AACT-----AGCAATA--A---G-TTGG-A-.....C-C-TT--AAC-TCTA-T--A--T-- 6332
SMM.US.-.H9 CAA-T-----C-----GGC----G-TGCT---GAT.....---ACAG-CAC------AGCAATA------G--TGG-AC.....C-C-TT--AAC-TC-A-T-----C-- 6373
STM.US.-.STM CAA-------C-----GGC----G-TGCC---GAC.....---ACAG-CAC------AGCAATA------G-CTGG-AC.....C-T-TT--AAC-TC-A-T--A--T-- 6547
SAB.SN.-.SAB1C CA-A-GTACCC-T-CCT---A-C-CAGAAAAATTTGATGCTTGGAAAAATCGG--CCCTC-A----GC--AGCAG-GAGTA-C---CA-CT-.....--G-TT--AAGT-CTT---------- 6791
TAN.UG.-.TAN1 A-T-------CT-T--G--A---G-AGCA---AAGGACAGG.....---CCAT-A--G-C----GC-G-GAGTA-C---CATTTG.....C-C-TT--AAGC-CT--G-------- 6580
VER.KE.-.AGM155 CAC-G---A-A-----GCCA---G-AGCA---GCAGACAGA.....---CC-T-A----C---AGCAGGAAGTA-C--TCA-CTG.....C-G-TT--A--G-CAT-------C-- 6582
VER.KE.-.9063 CAT----CA-T--T--GCCC---G-AGCC---GCAGATAGG.....---CCAT-A--G-C----GCAGGAAGTA----TCATTTG.....--G-TT--A----CAT-G--A----- 6584
VER.DE.-.AGM3 CAT----CA-C--T---CCA---G-AGCA---GCTGACAGA.....--CCC-T-A----C---AGCAGGAAGTA-----CA-CTG.....C---TT--G--G-CT--G-------- 6082
VER.KE.-.TYO1 CTC-----A-C--T---CCA---G-GGCA---GGTGATAGA.....--CCCAT-AA---C---AGCAGCCAGTA-C--CCATTTA.....C-C-TT--G----CTA-G--A--T-- 6071
COL.CM.-.CGU1 --CC------A------GC----G-TC-T-TT.....---G---AG-T-CAGT-GGG--G-A.....TGG-TAATAACACAGACTACC-TA--A-----G----GA----- 5939
DEN.CD.-.CD1 ---A-----CC-----TT-C---G----C-AT.....C-C---CCA--A-----C---G-AGTGC--A----G-AAGA-.....C-G-ATA-A--G-C-T-C--A--T-- 6574
GRV.ET.-.GRI_677 CTC----CA-T------GC---CG-GGCT-----..GATAAT.....CCGC-----A----AGCAG-GAGTA-----CATCTA.....C-C-TT--A----CGA-G-G---T-- 6497
GSN.CM.99.CN166 -A.....CA-A-----G--A--C-----C-AT.....------T----------TG--G---GA--T--C--GG-ATCA.....C---TTA----GGC---G--A----- 6466
GSN.CM.99.CN71 -A.....CA-A-----GG-A-----T--C-AT.....------T-T--------TG--G--AGA--T--C--GG-ATC-.....C---TTA----GGC-----GA----- 6451
DRL.-.-.FAO -TAAC---A-AGAG----------CTG-CCCTATG.....-CA-A-A-TG--G--ATAA-AC-GGC-TGGG--GCGCTGTCATCCA-GATA---GC-GTG--T--A----- 6223
RCM.GA.-.GAB1 CAA--------T-----CGA---G-TGCT---.....--T---AG-T-ATAT-----AGCAA-A--TA--G-GTGG-A-.....C-T-AT---TCC-CT-----A---- 6204
RCM.NG.-.NG411 CA---------T----G--A---G-TGCT---.....G-----AG-T--TAT-----AGCAA-GTCTA--G-GTGGCA-.....C-T-AT---TCC-CTT-G--A----- 6260
MND-2.-.-.5440 -CCCC---A-A-----G-------CAGGCCCTATA.....GAA-AGA-TG-A--AGTGACAC-AGCTTGGG-AGCAATTTCTTCCA-GATA---GC-GTAT----A----- 6174
MND-2.CM.98.CM16 -TCCG--CA-A---TGG-----C-CAGGTCCCATG.....GAA-A-A-TG--G-AGTAAT-C-AGC-TGGG-AGCTATCTCCTCCA-GATA--CGC-GTAT----A--T-- 6601
MND-2.GA.-.M14 -CCCC---A-A----------CAG-CCCTATA.....CAA-AGA-TG-AG--GTAAG-C-AGC-TGGG-AGCTATTTCCTCTA-GATA---GC-GTT-----A--G-- 6530
MNE.US.-.MNE027 CCC-T---A-T-------GC---G-TGCT---GAG.....---ACAG-CAC-------GCAATA------G--TGGCA-.....C-C-TT--GACCTCAA-------T-- 6373
LST.CD.88.447 CTA-TC-CAAC-T-AGTTTAAA-TT--CTAAAGAG.....ATA-AA-ATA-T--A--TATAC--C-AGC-TGG---GCTATGACCAGTA-GGT----ACT-T-A-G-----T-- 5595
LST.CD.88.485 CTA-TC-CAAC-T-AGTTTAAA-TT--CTCAAGAG.....ATA-AA-ATA-T--A--CATAC--C-AGC-TGG---GCTATGACCAGTA-GGT----ACT-T-A-G-----T-- 5592
LST.CD.88.524 CTA-TC--AAT-TTACTCTAAA-TTT-CTCAAGAA.....ATT-A--ACA-T--A--AATAC--C-GGC-TGGG-AGCTATGACGAGCA-GGT----GCT-TTA-G-------- 5595
LST.KE.-.lho7 CTA-CT--AAT-TTTC-G-AAACTTT-CAATACCA.....-TA-AA-ACA-T---G-TATAC--C-AGCCTGGTC-GCTATGAATGCTA-GGTA--CAGT-TTA-G--A----- 6673
SUN.GA.98.L14 CCA---CA-G-TT----G-GAA-TTT-CAGA-GGA.....ATA-GCAATA-T--A--AGTGC--C-AGC-TGGCAAGCTATGACCTCTA-GGTA---GCC-TTA-G-----T-- 6720
MND-1.GA.-.MNDGB1 AAA-TCC--AT-T-A-GG-AAA-TTT-CAG-ACTT.....ATA-G--A-A-T-----AGT-T-TC-AGC-TGGCA-GCTATGGGAAGTA-G-TA---ACC-TA--T-------- 6040
MON.CM.99.L1 --.....CA-C--T--GG-G--C--TGCT-AT.....C-AA--TACA-G-TGG-AGAAATG-TA--GGAT--G-AGTCT.....C-G-TCTCG--GGCTT-------T-- 6492
MON.NG.-.NG1 -C.....-A-A-------GC-----TGCA-AT.....--CA--TACA-G-TGG-AGA-ATG-T---AGAC--A-AATCA.....C-C-TCTCA--GGC-T---------- 5105
MUS.CM.01.1085 --.....-A-CT-T-GGG-AAAGTTTGACATCTAT.....--GA-TTATA-G-T-G-CCA-ATGACT---GATG-G--GTCT.....A-C-TTA----GGC-T----A----- 6424
DEB.CM.99.CM40 --CC---------CC----C-----TGCT-TTGAT.....AACCCC.....A----GG--CA--TATG-C-GGAT--G-C-AGTTTG.....-ACA-A--GTCAT-C--A----- 6260
DEB.CM.99.CM5 --CCT------T-CC-GT-----G--TCT-TTTCA.....AATGTT.....A-G-----ACAATTATG---CGAC--G-C-AGCTTA.....-ACAGA---TCTT-T--A--C-- 6229
SYK.KE.-.KE51 GA--T---A-T-GT-GGG-A-A-TT--CTGC-TGG.....AATTCTTCCC-TGGCA--AG---ACA--T-TT-C-GGAC--GTC-GATCTT.....-TCCTG--GGCTGAT-G---T-- 6210
SYK.KE.-.SYK173 G---C--CA-T----G-G-A-ACTTTCCAGCATGG.....AAT-GC-G-C-----A--AGG---CAAATATTA--GGAC--GTCAGCTTTA.....-TCTTG---GC-AAT-GA--C-- 6532

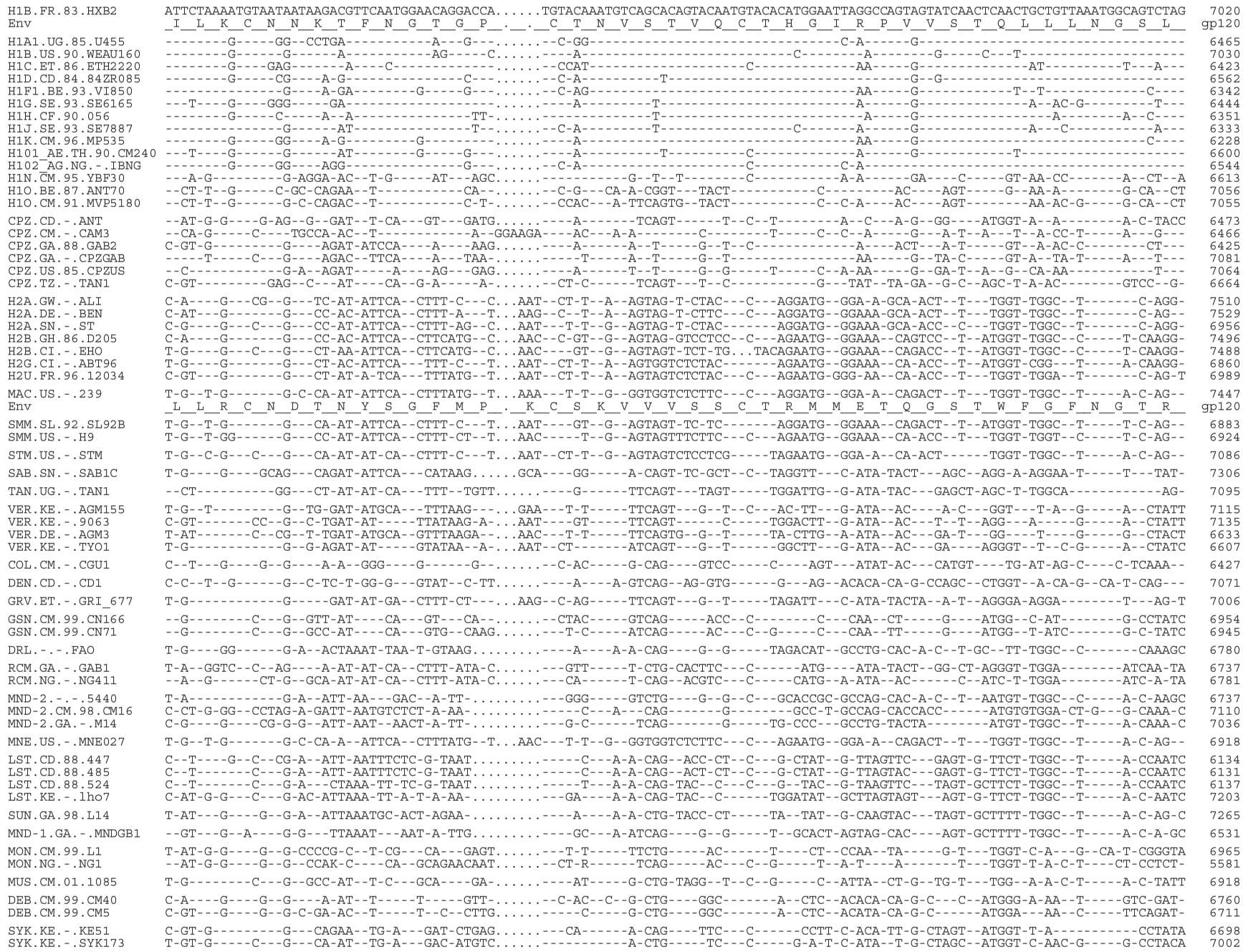

V1 loop
H1B.FR.83.HXB2
Env
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Env
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes

PLV
complete genomes


V2 loop
H1B.FR.83.HXB2
Env
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Env
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes

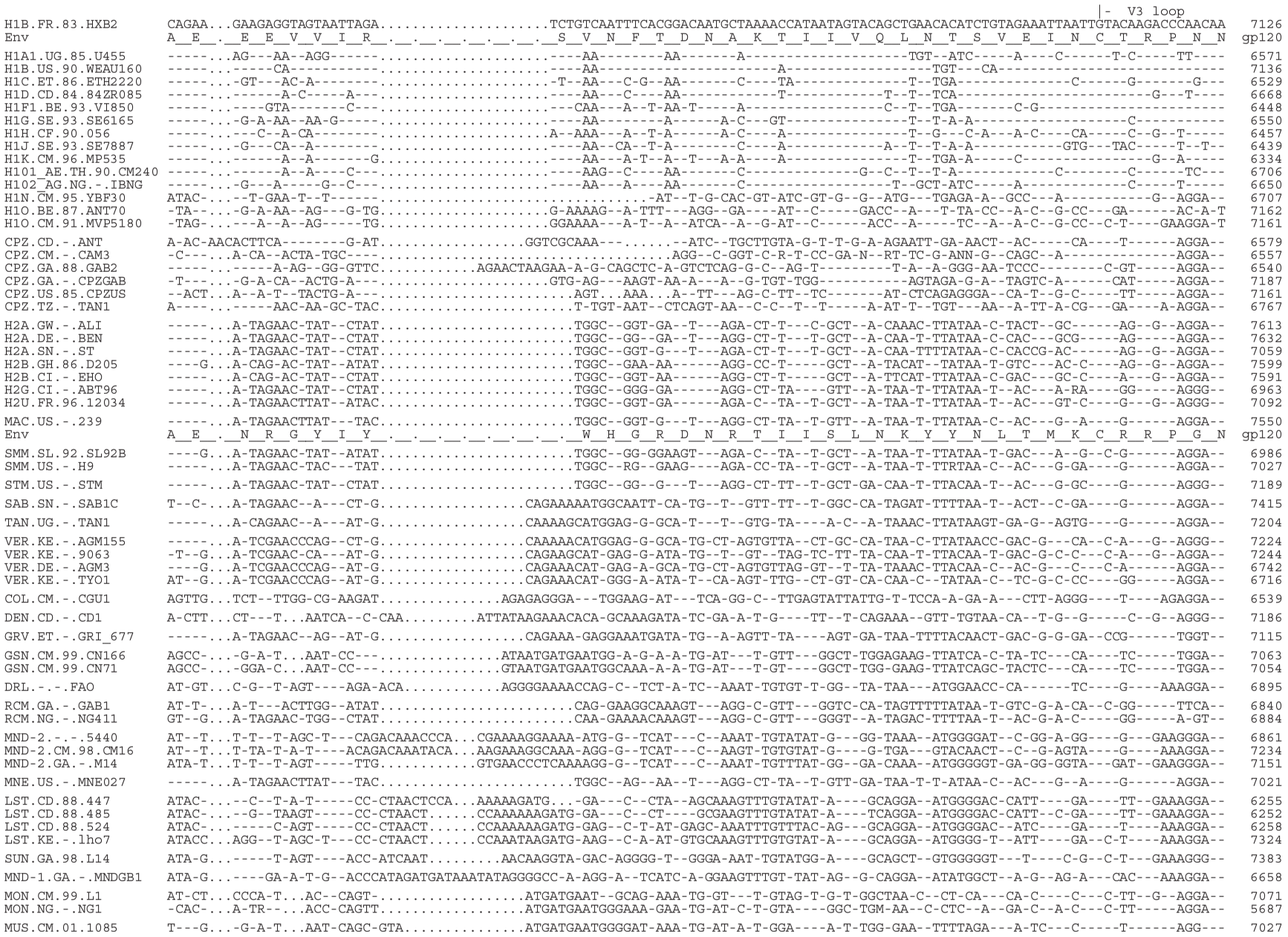
PLV
complete genomes

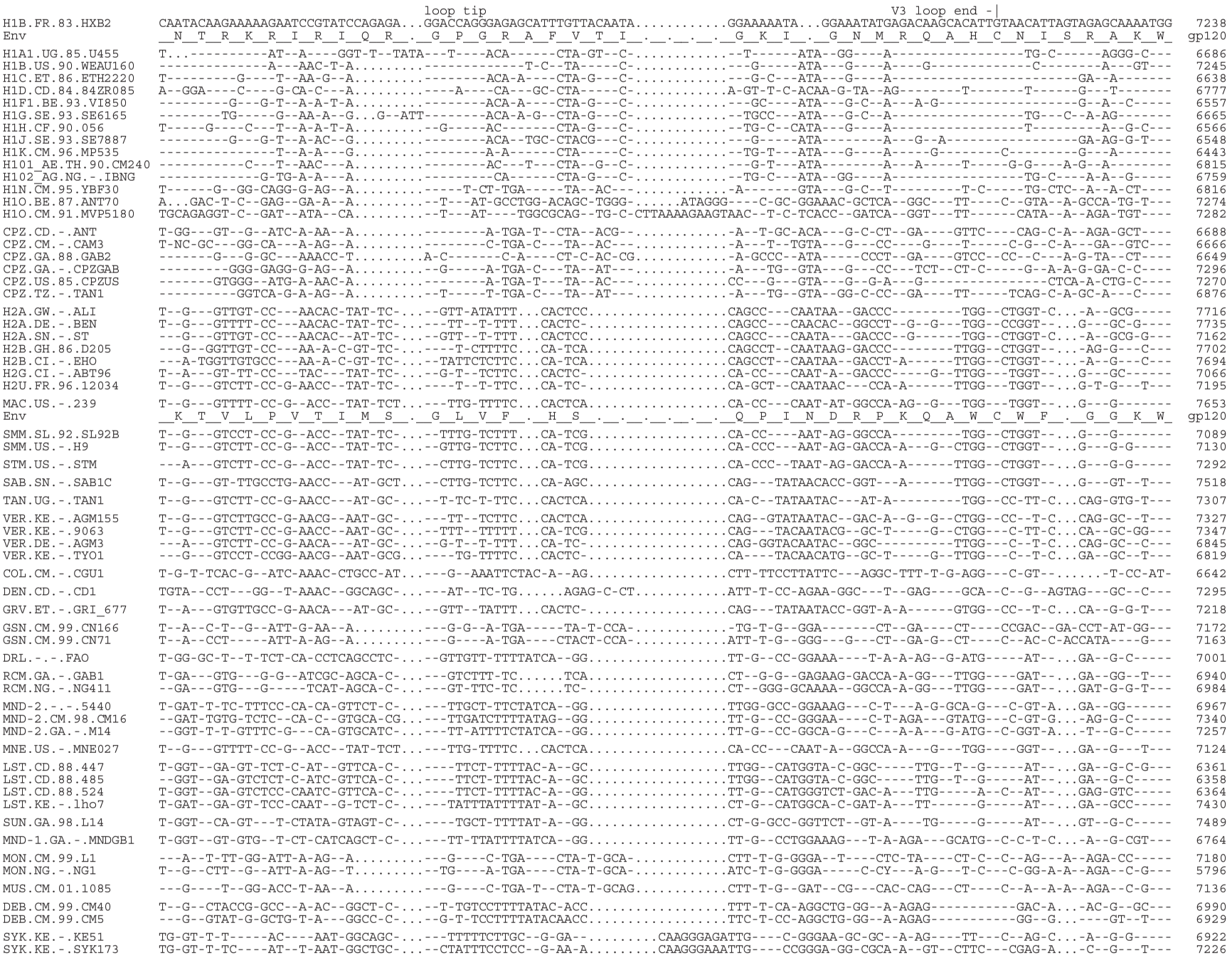




PLV
complete genomes

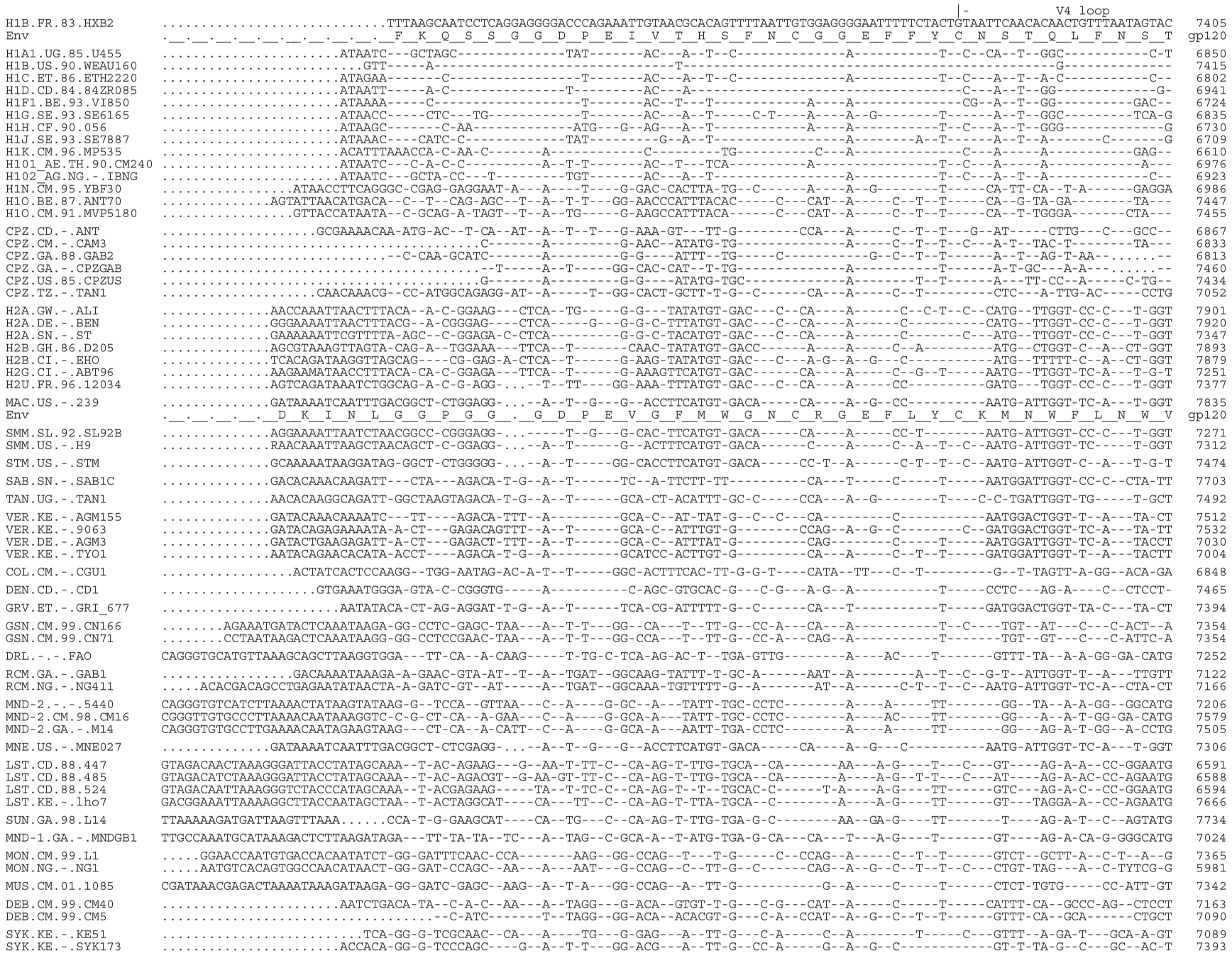
PLV
complete genomes



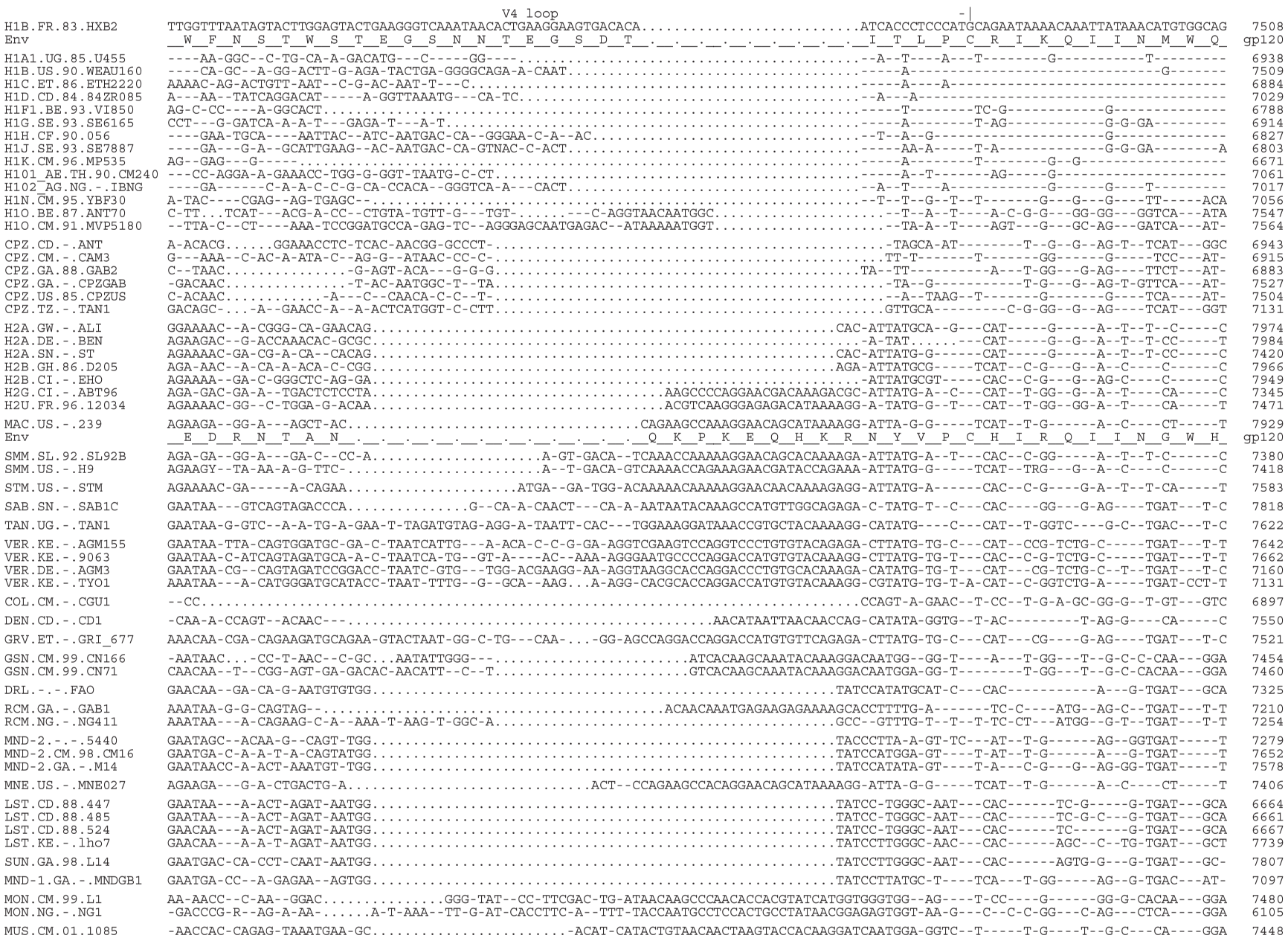


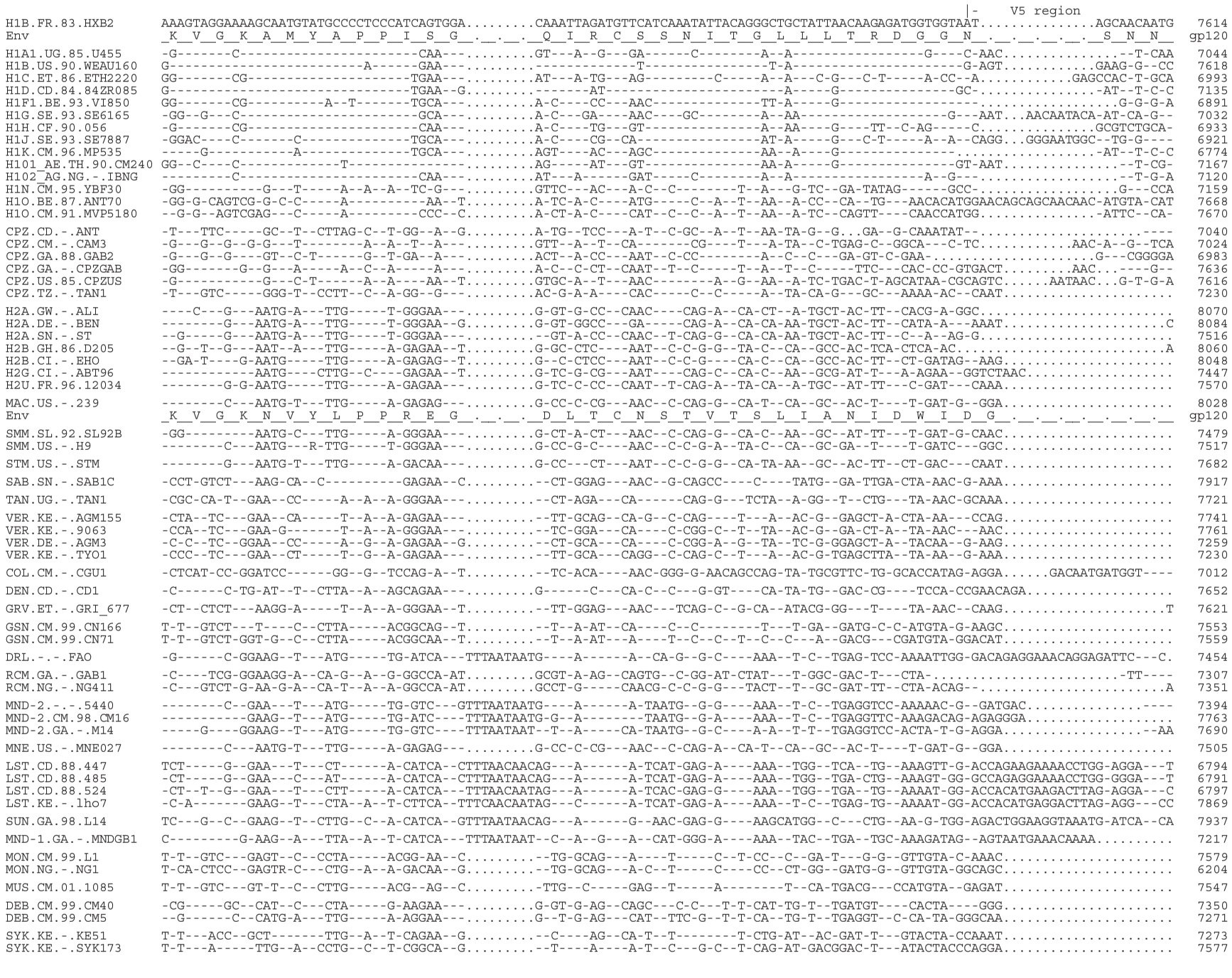
PLV
complete genomes

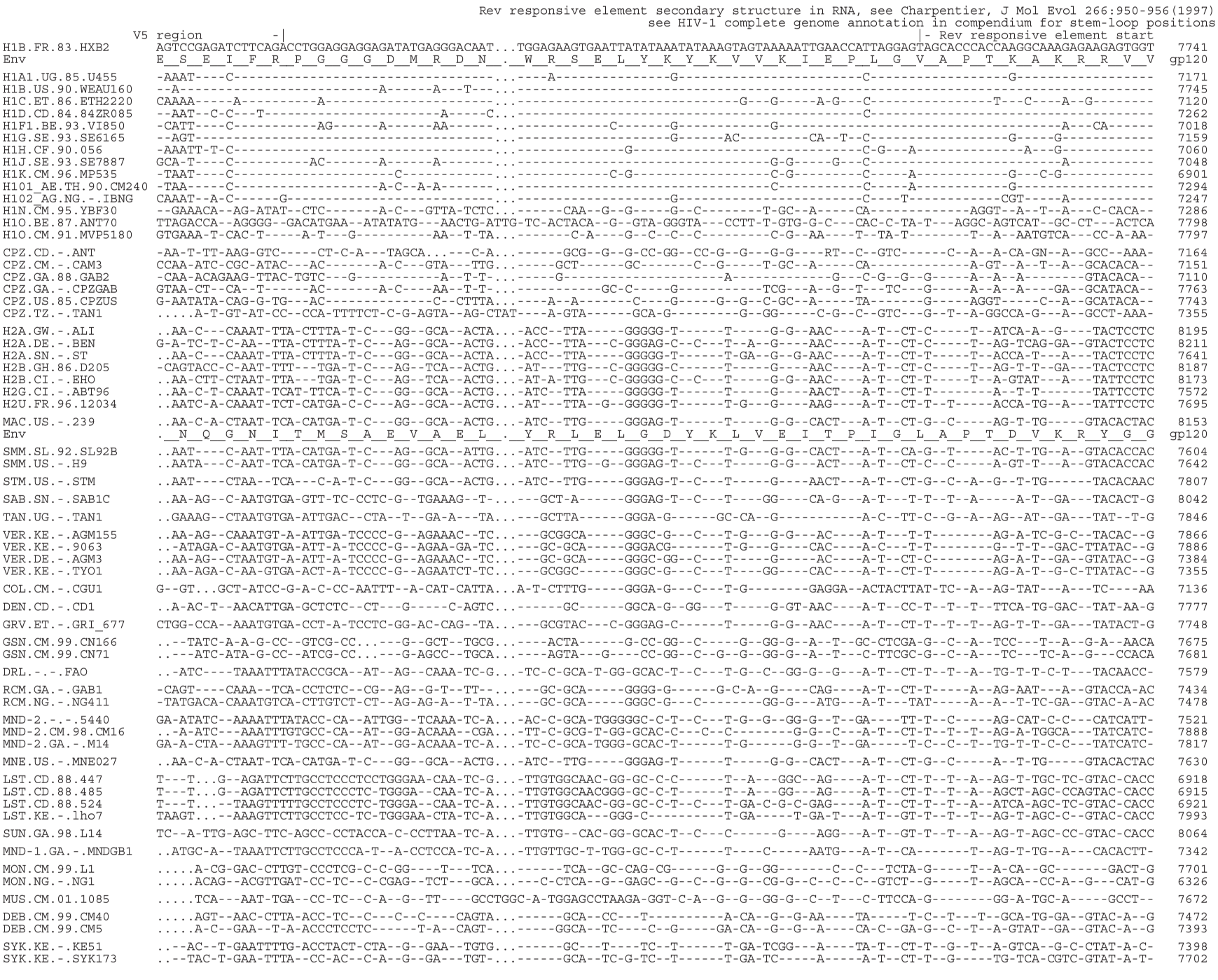
```
H1B.FR.83.HXB2       AATAACACTTTAAAACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATC.....  7304
Env                  _N__N__T__L__K__Q__I__A__S__K__L__R__E__Q__F__G__N__N__K__T__I__I__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__  gp120
H1A1.UG.85.U455      ----GA---A--C-----G-----GAAC------AGA--A-G---AAT--C--A-C-.....  6743
H1B.US.90.WEAU160    ----------------------T-GAA-----------AT-AAACA-TT---G--T-A--C---A.....  7311
H1C.ET.86.ETH2220    -----A---C--C--A--G--AAGGAA------CA-A-G--C--CCCT---...------.....  6695
H1D.CD.84.84ZR085    --------C---CG----G-------A------G--A-C-TT--GAACC-A-CA---.....  6834
H1F1.BE.93.VI850     ---------C-GG-GT-TG--AAGGCAG------AGTCG--T--CCCT------...---GC-.....  6617
H1G.SE.93.SE6165     --AG-GG-C---C-GA-TG-C---GCAG--C--G-GA--ATC---AAT--G-GC-GCGAG-AC.....  6728
H1H.CF.90.056        -----G------C-C---G---T--CAC-----G--AT---C--GAAC----GA-C-.....  6623
H1J.SE.93.SE7887     -G----------CGTAGAG------CA---C----------C---AAT--A-CA.....  6602
H1K.CM.96.MP535      --C-TG-------GCAGAG--AAGGAA--GC---A------T---AAG---GGA-C--T-.....  6503
H101_AE.TH.90.CM240  -----AGT----------G--A--GAA-------A---G--C---AAT--G-CA.....  6869
H102_AG.NG.-.IBNG    -----A-----C-T---G---T--CAC------AGACGT-C---AAG--C-CC-C-.....  6816
H1N.CM.95.YBF30      G-ACCA-TG-GG--TAGA-C-AGAGAGG--A---AGA--ATCC-G--G--A--C--C.....  6873
H1O.BE.87.ANT70      GGA--A-TA--------A-C----GAA-GG-ATTT-----T-G-AAAC----CAGGT.....  7331
H1O.CM.91.MVP5180    G-A--TG-CC--C----A-C-----TA-GG-ATTT-A-T-TTG-AAACC-A-CAG-G-AT.....  7342

CPZ.CD.-.ANT         G-AC-AG-ACGT--CA-A-C--AGCA-GTTC-TGCG--G--T-GGAA---AGTAG-C-AC-A--C-AAT.....  6757
CPZ.CM.-.CAM3        G-CG-A-GAC----TG-A-CT-G-CAAGCC--GC----G-TT---AC---CCT--C-CA-G-C-A-TTTACAGTAAGT.....  6744
CPZ.GA.88.GAB2       -------T-C----CAGA---AAACAA--GA-T-A-A-CAGCACAACGTGGC-C-GGGGGGAC---ACA.....  6718
CPZ.GA.-.CPZGAB      ---CGA--GG-GG--G-AG--AAA-AGGCT---GC-ACCTC--CAAATCGA-CAGC-G-C-AC--AACACTCAACAGAGCC.....  7377
CPZ.US.85.CPZUS      G-G--ACAGAG----T-C-CTTTAGAG-T-A---A-A-GG--GCAAACCTG-CA--GGT-GAGT-AATTCCAAAT.....  7345

H2A.GW.-.ALI         -GG--AG-CA-GC-GG--G-GAAGGAA-CCC-TGT-A----TCCCA-GT----AGG---C-ATGA-ACA.....  7785
H2A.DE.-.BEN         -GGG-AG-CA-GC-GG--G-GAAGCAA-CCC-TGT-C----TCCCA--T-C--AGG--TC-ATGATACA.....  7804
H2A.SN.-.ST          --GG-AG-CA-G--GG--G-GAAGCTA-CCC-TGC-A----TCCCA-GT----AGG---C-ACGA-ACA.....  7231
H2B.GH.86.D205       -CAG-AG-CA----GG--G-GAAA--G-CCA-C-T-A----TCCCA-GT----AGG-GGTGC--AAAAT...ATCACA.....  7777
H2B.CI.-.EHO         -CAG-AG-CA--C-GG--G-GAAAGAG-CCA-T-A-A-T--TCCCA--T--TCAGG-----C--ATATC.....  7763
H2G.CI.-.ABT96       --GG-AG-CA-GC-GG--G--AAGGAA-CCG-GGTCA-G--CCCCA-CT----AGG-----ATGA-ACA.....  7135
H2U.FR.96.12034      --AGGGG-AA-CC-GG--G-GAAGGAG-CCA-TGCCA-C--TCCCA-GT--C--GGT--C-AG-ATATT.....  7264

MAC.US.-.239         --GG-TG-AA-----G--G-GAAGCAG-CCA-TGTCA----TCCCA-GT---C-GG---T-AC-ATACT.....  7722
Env                  _K__D__A__I__K__E__V__K__Q__T__I__V__K__H__P__R__Y__T__G__T__N__N__T__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__  gp120
SMM.SL.92.SL92B      -GAG-AG--A-GC-GG--G-TAAG-AA-CCA--GTCA----CCCCA-GT---C-GG---C-ATGATACT.....  7158
SMM.US.-.H9          --AG-GG-CA-CCGGG-AG-GAAGGMAMCC--GGTCA----TCCCA-GT----GGG---T-ATRAKAYT.....  7199

STM.US.-.STM         -GAGGAG-AA-----G--G-TAAGGAA-CC--GGTCA----TCCCA-GT---CAGG---C-ATGA-ACG.....  7361

SAB.SN.-.SAB1C       -GAGGAG-C-GG--GG-AG-CAAAGAA-C-A--GTGAG--T-CCACC---A--GT-T-GTGGC-CAAAT.....  7587

TAN.UG.-.TAN1        --AGGAG-C-GG-GGG-AG-GAGA-AA-C-A--GTG--G-TTCCAAA-G-G--AT-TCG-GGG-CAAAT.....  7376

VER.KE.-.AGM155      --AGGAG---GG--GG-AG--CAAGAGG--A--GT-A--TT-CCAAA-G-ACGGT-CCA-GGC-C-AAT.....  7396
VER.KE.-.9063        -GAG-AG-C-GG--GG--G-TAAAGAAG--A---TCA--TT-CCAAA-G---GGT-TCGGGG--C-AAT.....  7416
VER.DE.-.AGM3        -GAGGAG-C-GG---G-AG--AAA-ATG--A--GT-A--TT-CCAAA-G---GAT-CCA-GG--C-AAT.....  6914
VER.KE.-.TYO1        -GAGGTG-C-GGCGGG-AG--AAACAA---A--GT---GTT-CCAAA-G-C-GGT-T-A-GG--C-AAT.....  6888

COL.CM.-.CGU1        -C-GGGCAGAAT--GA-A-ACAAA-CAC----CGC-AT-G--A-CAA---G---TT----GCTTGGCTAGAGCGCATTTCACGCAAAAATATA.....  6735

DEN.CD.-.CD1         -C--GA--AAGGC-GG-A----AA-AAGC-G-T-A---GG-G---AAC-TC---GGC---.....  7355

GRV.ET.-.GRI_677     --AGGGG-A-GG---G-AG-CAGAGAAG--G---AGA--GTGAAAAATCT--CAG--GT--GC--AGAA.....  7287

GSN.CM.99.CN166      ----GGG-AA---G-G-AGCTATG-AAGC-A-G-AT-----T-GG-A------A-CTGGG.....  7232
GSN.CM.99.CN71       G---TAG-AA---GG--AGCCATG-AGGCCA-G-AGA-C--T-GG-A---G-TC--C-ATGAT-C-ACC.....  7232

DRL.-.-.FAO          GGA--GG-----GG-AGTC----C-AAG------A-TC-GT-AG----TCC-T-TGG-G--AT-CTACT.........AATAGCACATGCATAAGTCTCAAAGGCAAAAAGATCCAGGACAGAAACGTGA  7122

RCM.GA.-.GAB1        -CGG-TG-A-GG---G-AG--AAAGAG-G-G-G-A-AC-AC-AAA--GT--CGAGGT--T-GT-A-ACA.....  7009
RCM.NG.-.NG411       --AGGTG-G-GG---G--G--AAAGAGGC-A-T-ACA-GAC-AAA---T----AGGT.....  7041

MND-2.-.-.5440       GGAC-GG-----C-TAGTC--T-ATTAG-GC-----A-GAT-AA--AT-GC-T-T-C-A-GAC-A-CAC.........AATATGACATGTAAAAGCAGTAACAACAAA...........AAGAACACAA  7076
MND-2.CM.98.CM16     GGA-TAG-CA-GC-TG-TC----A-TAG--------A--AT-GACA-CTCC-TATGG-G--ATG-AACA.........AAGGGCTATTGCAAGGGGTTAAAAAGAAAG...........GAAAATAGAA  7449
MND-2.GA.-.M14       GGA-GAG----G--CGCAT--T-AC-TG------AGA--GT-AA--C--G--T-T-T-G-GGTT-TAAT......GCAAGTGGCCCTTGCAGGGGGATAAACAATAGAGGA......GTGGTAAACAAAA  7375

MNE.US.-.MNE027      --GG-GG-AA-----G--G-GAAGCAG-CCA-TGTCA----TCCCA-GT---C-GG---T-AC-ATACT.....  7193

LST.CD.88.447        GGA-GA-TG--TC-TA-TT---GA-A--TGC-T-A-AGG-TTAA--C--CAGCG-TG-ATTAT-C-CATGGAAAC......TGTATAGATAACACAAAGCCTTGT.....................G  6461
LST.CD.88.485        GGA-GG-TG--T--TA-TT---GG-AA-TGC-TCA-A-G-TTAA--C--CAGCA-TG-ATTAT-CAAAT...GGACAA...TGTCCTGAGGGAAAACAGCCTTGT.....................G  6458
LST.CD.88.524        GGA-GA-TG--T--CA-TT---GA--G-TG----AGA-G-TTAA--AG-C-GC--TG-ATTAT-AAGAA...GGAAAT...TGTTCCACACCAGGAAAACCATGT.....................G  6464
LST.KE.-.lho7        GGA-GA-TG--T--TA-TC---GC-AA-TGC---AG--GTTGAA--C-G-AGCC-TG-ATTAT-CAGAAGGGACAGGGACATGTGATTCTAAGAAAACAACTTGT.....................G  7536

SUN.GA.98.L14        GGA-GAG-C--TGCCACCT---GG-AACT---G--GC-GGTGGAACCTTTAGC---C-TG-GT-ATGGCTGT...ACATTT...GATAATATAACAAAGACATGCAGATTTACC...AATGGTACAGATT  7610

MND-1.GA.-.MNDGB1    GGGTTAG-AC-TTGGAGTC------AAG--C---ATA--TT-AA--ACTCC-TC---GTG-ACCAGACCTGTAAAAATTTTACTAGCACTGGAGAGGAGAACAAACAAAACACGGACAAGCAAAAGGAGT  6894

MON.CM.99.L1         G-G-CTG-CC--C-GGCAG-TCACGAGGC-G---A-ACTG-G-GG-A---G--A--C-AT.....  7240
MON.NG.-.NG1         G-CGCTG-C---C-TGCAGCCCA--TRGCTA-C-RGR-GG---GG-A--GC--A--T-AT.....  5856

MUS.CM.01.1085       G--CCAG-A--G-G---AGCTATGCTAGCCA-G---A----T-GGCATGG--GATTGCA-.....AAACAGAACCACAGCT  7212

DEB.CM.99.CM40       GCA---G-AAC---GG-A--GCGACAA--GA--TTG---AT---------GGCA--T-GG-C--AT.....  7056
DEB.CM.99.CM5        -CAGGAG-AG-T--GG-----AGG----G-----AG--GAT--A-AATGTG-CA----T--C-G-A.....  6995

SYK.KE.-.KE51        GGA--GTT---TG-G---G-CCA--ATG-GAGT-T-A-GGTG...T-G--A---GTG--CTC--CA.....  6985
SYK.KE.-.SYK173      -CAG-ATT---T--G---G--CA-GAAC-GGCC-CGA--ACC...T-G--A---GT---C-AC.....  7286
```


V4 loop
H1B.FR.83.HXB2 7405
Env gp120
H1A1.UG.85.U455 6850
H1B.US.90.WEAU160 7415
H1C.ET.86.ETH2220 6802
H1D.CD.84.84ZR085 6941
H1F1.BE.93.VI850 6724
H1G.SE.93.SE6165 6835
H1H.CF.90.056 6730
H1J.SE.93.SE7887 6709
H1K.CM.96.MP535 6610
H101_AE.TH.90.CM240 6976
H102_AG.NG.-.IBNG 6923
H1N.CM.95.YBF30 6986
H1O.BE.87.ANT70 7447
H1O.CM.91.MVP5180 7455
CPZ.CD.-.ANT 6867
CPZ.CM.-.CAM3 6833
CPZ.GA.88.GAB2 6813
CPZ.GA.-.CPZGAB 7460
CPZ.US.85.CPZUS 7434
CPZ.TZ.-.TAN1 7052
H2A.GW.-.ALI 7901
H2A.DE.-.BEN 7920
H2A.SN.-.ST 7347
H2B.GH.86.D205 7893
H2B.CI.-.EHO 7879
H2G.CI.-.ABT96 7251
H2U.FR.96.12034 7377
MAC.US.-.239 7835
Env gp120
SMM.SL.92.SL92B 7271
SMM.US.-.H9 7312
STM.US.-.STM 7474
SAB.SN.-.SAB1C 7703
TAN.UG.-.TAN1 7492
VER.KE.-.AGM155 7512
VER.KE.-.9063 7532
VER.DE.-.AGM3 7030
VER.KE.-.TYO1 7004
COL.CM.-.CGU1 6848
DEN.CD.-.CD1 7465
GRV.ET.-.GRI_677 7394
GSN.CM.99.CN166 7354
GSN.CM.99.CN71 7354
DRL.-.-.FAO 7252
RCM.GA.-.GAB1 7122
RCM.NG.-.NG411 7166
MND-2.-.-.5440 7206
MND-2.CM.98.CM16 7579
MND-2.GA.-.M14 7505
MNE.US.-.MNE027 7306
LST.CD.88.447 6591
LST.CD.88.485 6588
LST.CD.88.524 6594
LST.KE.-.lho7 7666
SUN.GA.98.L14 7734
MND-1.GA.-.MNDGB1 7024
MON.CM.99.L1 7365
MON.NG.-.NG1 5981
MUS.CM.01.1085 7342
DEB.CM.99.CM40 7163
DEB.CM.99.CM5 7090
SYK.KE.-.KE51 7089
SYK.KE.-.SYK173 7393
PLV
complete genomes

PLV
complete genomes

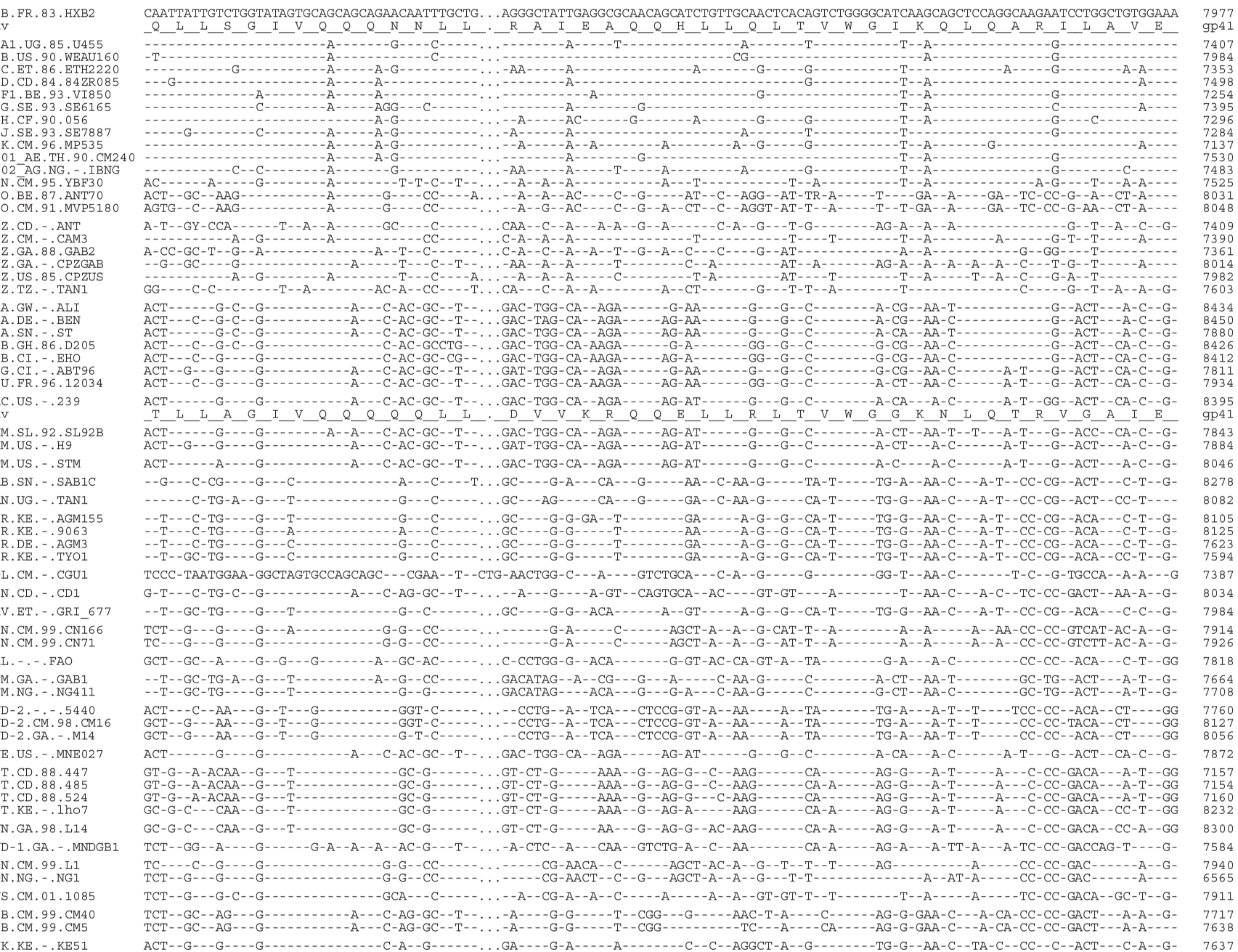


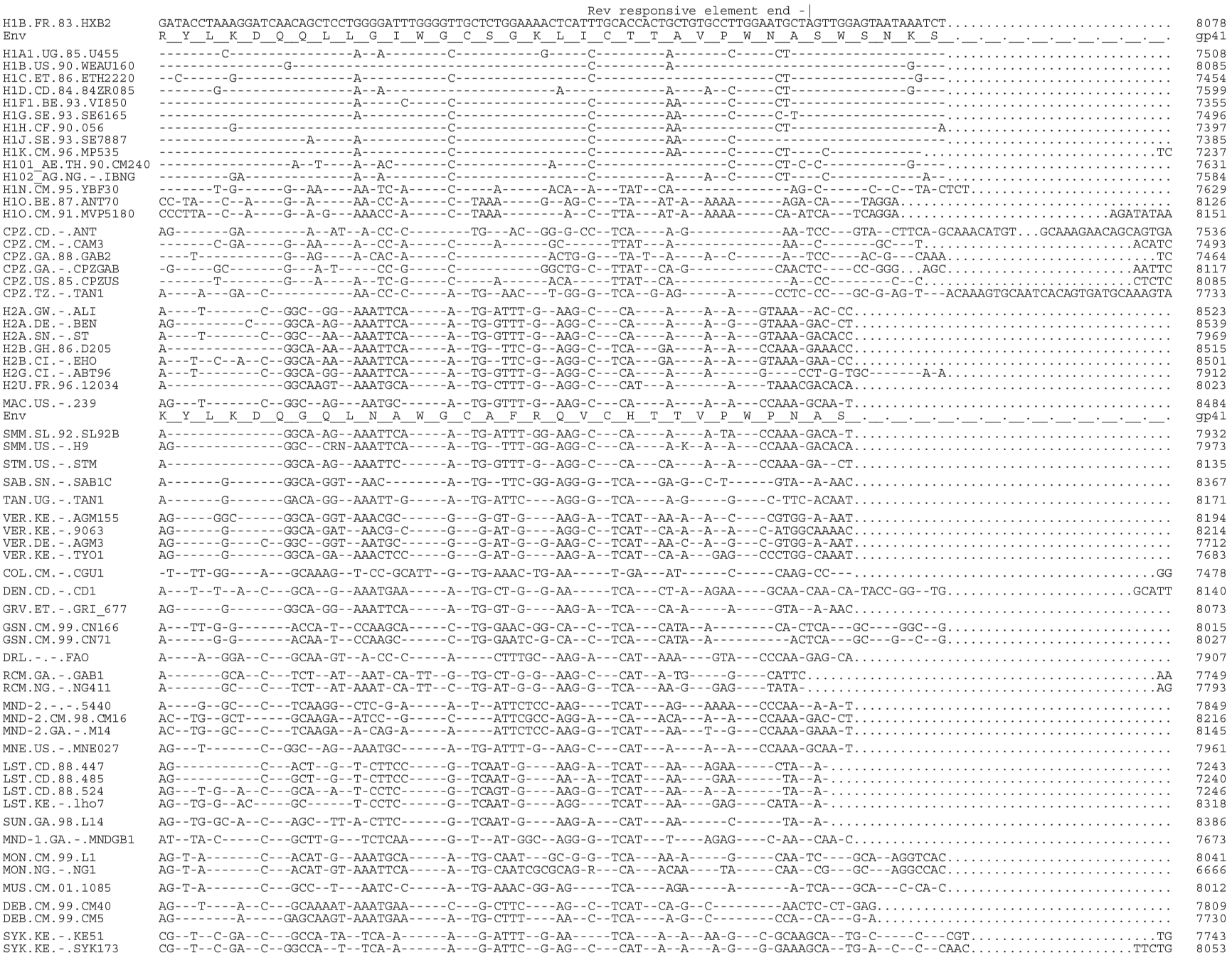

V5 region
H1B.FR.83.HXB2 AAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGA.....CAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAAT.....AGCAACAATG 7614
Env _K__V__G__K__A__M__Y__A__P__P__I__S__G__._.__._.__Q__I__R__C__S__S__N__I__T__G__L__L__L__T__R__D__G__G__N__._.__._.__._.__S__N__N__ gp120
H1A1.UG.85.U455 -G-------C----------CAA---.....GT---A--G---GA------C--------A--A----------G----C-AAC.....--T-CAA 7044
H1B.US.90.WEAU160 G----------A------GAA---.....T----------T-A----------G-AGT.....GAAG-G--CC 7618
H1C.ET.86.ETH2220 GG-------CG----------TGAA---.....AT---A-TG---AG---------C-----A--A---C-G---C-T-----A-CC--A.....GAGCCAC-T-GCA 6993
H1D.CD.84.84ZR085 G----------TGAA--G.....-------AT----------A-----G----------GC-.....-AT--T-C-C 7135
H1F1.BE.93.VI850 GG-------CG----------A--T-------TGCA---.....A-C----CC---AAC----------TT-A-----G----------.....G-G-G-A 6891
H1G.SE.93.SE6165 -GG--G---C----------GCA---.....A-C---GA----AAC----GC--------A--AA----------G---AAT...AACAATACA-AT-CA-G-- 7032
H1H.CF.90.056 -G-------CG----------CAA---.....A-C----TG---GT----------A--AA----G----TT--C-AG-----C.....GCGTCTGCA- 6933
H1J.SE.93.SE7887 -GGAC----C-------C----------TGCA---.....A-C----CG--CA----------AT-A-----G---C-T-----A--A--CAGG...GGGAATGGC--TG-G---- 6921
H1K.CM.96.MP535 -----G----------A----------TGCA---.....AGT----AC---AGC----------AA--A----G----------.....-AT--T-C-C 6774
H101_AE.TH.90.CM240 GG--C----C----------T----------.....AG-----AT---GT----------AA-A-----G----------GT-AAT.....--T-CG- 7167
H102_AG.NG.-.IBNG --------C----------CAA---.....AT---A------GAT-----C--------A--A----------A--G---.....--T-G-A 7120
H1N.CM.95.YBF30 -GG----------G---T-----A--A--A--TC-G---.....GTTC---AC---A-C--C--------T--A---G-TC--GA-TATAG------GCC-.....G---CCA 7159
H1O.BE.87.ANT70 -GG-G-CAGTCG-G-C-C-----A----------AA--T.....A-TC-A-C----ATG-----C--A--T--AA--A-CC--CA--TG---AACACATGGAACAGCAGCAACAAC-ATGTA-CAT 7668
H1O.CM.91.MVP5180 --G-G--AGTCGAG---C-----A----------CCC--C.....A-CT-A-C----CAT--C--C--A--T--AA--A-TC--CAGTT----CAACCATGG.....-ATTC--CA- 7670
CPZ.CD.-.ANT -T---TTC-----GC--T--CTTAG-C--T-GG--A--G.....A-TG--TCC---A-T--C-GC--A--T--AA-TA-G--G...GA--G-CAAATAT--.....---- 7040
CPZ.CM.-.CAM3 -G---G--G--G-G---T--------A--A--T--A---.....GTT--A--T--CA-------CG----T---A-TA--C-TGAG-C--GGCA---C-TC.....AAC-A--G--TCA 7024
CPZ.GA.88.GAB2 -G---G--G----GT--C-T------G--T-GA--A---.....ACT--A-CC---AAT--C-CC--------A--C--C---GA-GT-C-GAA-.....G---CGGGGA 6983
CPZ.GA.-.CPZGAB -GG-------G--G---A-----A--C--A-----A---.....A-C--C-CT--CAAT--T--C-----T--A--T--C-----TTC---CAC-CC-GTGACT.....AAC..----G-- 7636
CPZ.US.85.CPZUS -G----------G---C-T------A--A----AA--T.....GTGC-A--T---AAC--------A--G--AA--A-TC-TGAC-T-AGCATAA-CGCAGTC.....AATAAC...G-T-G-A 7616
CPZ.TZ.-.TAN1 -T---GTC-----GGG-T--CCTT--C--A-GG--G---.....AC-G-A-A----CAC-----C--C-----A--TA-CA-G---GC---AAAA-AC--CAAT..... 7230
H2A.GW.-.ALI ----C---G---AATG-A---TTG-----T-GGGAA---.....G-GT-G-CC--CAAC----CAG-A--CA-CT--A-TGCT-AC-TT--CACG-A-GGC..... 8070
H2A.DE.-.BEN --------G---AATG-A---TTG-----T-GGGAA--G.....G-GT-GGCC---GA-----CAG-A--CA-CA-AA-TGCT-AC-TT--CATA-A---AAAT.....C 8084
H2A.SN.-.ST --G-----G---AATG-A---TTG-----T-GGGAA---.....--GT-A-CC--CAAC--T-CAG-G--CA-CA-AA-TGCT-AC-TT--C--A--AG-G..... 7516
H2B.GH.86.D205 --G--T--G---AAT--A---TTG-----A-GAGAA--T.....G-GC-CTC----AAT--C-C-G-G--TA-C--CA--GCC-AC-TCA-CTCA-AC.....A 8060
H2B.CI.-.EHO --GA-T----G-AATG-----TTG-----A-GAGAG--T.....G--C-CTCC---AAT--C-C-G----CA-C--CA--GCC-AC-TT--CT-GATAG--AAG..... 8048
H2G.CI.-.ABT96 ----------AATG----CTTG--C----GAGAA--T.....G-TC-G-CG---AAT----CAG-C--CA-C--AA--GCG-AT-T---A-AGAA--GGTCTAAC..... 7447
H2U.FR.96.12034 --------G-G-AATG-----TTG-----A-GAGAA---.....G-TC-C-CC--CAAT--T-CAG-A--TA-CA--A-TGC--AT-TT---C-GAT---CAAA..... 7570
MAC.US.-.239 --------C---AATG-T---TTG-----A-GAGAG---.....G-CC-C-CG---AAC--C-CAG-G--CA-T--CA--GC--AC-T----T-GAT-G--GGA..... 8028
Env _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__A__N__I__D__W__I__D__G__._.__._.__._.__._.__ gp120
SMM.SL.92.SL92B -GG---------AATG-C---TTG-----A-GGGAA---.....G-CT-A-CT---AAC--C-CAG-G--CA-C--AA--GC--AT-TT---T-GAT-G-CAAC..... 7479
SMM.US.-.H9 --------C---AATG---R-TTG-----T-GGGAA---.....G-CC-G-C----AAC--C-C-G-A--TA-C--CA--GC-GA--T----T-GATC---GGC..... 7517
STM.US.-.STM --------G---AATG-T---TTG-----A-GACAA---.....G-CC---CT---AAT--C-C-G-G--CA-TA-AA--GC--AC-TT--CT-GAC---CAAT..... 7682
SAB.SN.-.SAB1C -CCT-GTCT---AAG-CA--C----------GAGAA--C.....--CT-GGAG---AAC--G-CAGCC----C----TATG--GA-TTGA-CTA-AAC-G-AAA..... 7917
TAN.UG.-.TAN1 -CGC-CA-T--GAA--CC-----A--A--A-GGGAA---.....--CT-AGA---CA------CAG-G---TCTA--A-GG--T--CTG---TA-AAC-GCAAA..... 7721
VER.KE.-.AGM155 -CTA--TC---GAA--CA-----T--A--A-GAGAA---.....--TT-GCAG--CA-G--C-CAG----T---A--AC-G--GAGCT-A-CTA-AA---CCAG..... 7741
VER.KE.-.9063 -CCA--TC---GAA-G-------T--A--A-GGGAA---.....--TC-GGA---CA----C-CGG-C--T--TA--AC-G--GA-CT-A--TA-AAC---AAC..... 7761
VER.DE.-.AGM3 -C-C--TC--GGAA--CC-----A--G--A-GAGAA--G.....--CT-GCA---CA----C-CGG-A--G--TA--TC-G-GGAGCT-A--TACAA--G-AAG..... 7259
VER.KE.-.TYO1 -CCC--TC---GAA--CT-----T--G--A-GAGAA---.....--TT-GCA---CAGG--C-CAG-C--T---A--AC-G-TGAGCTTA--TA-AA--G-AAA..... 7230
COL.CM.-.CGU1 -CTCAT-CC-GGATCC------GG--G--TCCAG-A--T.....--TC-ACA----AAC-GGG-G-AACAGCCAG-TA-TGCGTTC-TG-GCACCATAG-AGGA.....GACAATGATGGT---- 7012
DEN.CD.-.CD1 -C------C-TG-AT--T--CTTA--A--AGCAGAA---.....G------C---CA-C--C---G-GT----CA-TA-TG--GAC-CG----TCCA-CCGAACAGA..... 7652
GRV.ET.-.GRI_677 -CT--CTCT---AAGG-A-----T--A--A-GGGAA--T.....--TT-GGAG---AAC---TCAG-C--G-CA--ATACG-GG---T----TA-AAC--CAAG.....T 7621
GSN.CM.99.CN166 T-T--GTCT---T----C--CTTA-----ACGGCAG--T.....--T--A-AT--CA----C--C--------T---T-GA--GATG-C-C-ATGTA-G-AAGC..... 7553
GSN.CM.99.CN71 T-T--GTCT-GGT-G--C--CTTA-----ACGGCAA--T.....--T--A-AT---A----T--C--C--T--C--C---A--GACG---CGATGTA-GGACAT..... 7559
DRL.-.-.FAO -G------C-GGAAG--T---ATG-----TG-ATCA---TTTAATAATG----A------A--CA-G--G-C----AAA--T-C--TGAG-TCC-AAAATTGG-GACAGAGGAAACAGGAGATTC---C. 7454
RCM.GA.-.GAB1 -C----TCG-GGAAGG-A--CA-G--A--G-GGCCA-AT.....GCGT-A-AG--CAGTG--C-GG-AT-CTAT---T-GGC-GAC-T---CTA-.....-TT---- 7307
RCM.NG.-.NG411 -C---GTCT-G-AA-G-A--CA-T--A--A-GGCCA-AT.....GCCT-G-----CAACG-C-C-G-G---TACT--T--GC-GAT-TT--CTA-ACAG--.....A 7351
MND-2.-.-.5440 --------C--GAA---T---ATG-----TG-GTC---GTTTAATAATG----A------A-TAATG--G-G----AAA--T-C--TGAGGTCC-AAAAAC-G--GATGAC.....---- 7394
MND-2.CM.98.CM16 ----------GAAG--T---ATG-----TG-ATC----TTTAATAATG-G--A--------TAATG--G-A----AAA--T-C--TGAGGTTC-AAAGACAG-AGAGGGA..... 7763
MND-2.GA.-.M14 -----G----GGAAG--T---ATG-----TG-GTC----TTTAATAAT--T--A-----CA-TAATG--G-C----A-A--T-T--TGAGGTCC-ACTA-T-G-AGGA.....--AA 7690
MNE.US.-.MNE027 --------C---AATG-T---TTG-----A-GAGAG---.....G-CC-C-CG---AAC--C-CAG-A--CA-T--CA--GC--AC-T----T-GAT-G--GGA..... 7505
LST.CD.88.447 TCT-----G--GAA---T---CT------A-CATCA--CTTTAACAACAG---A------A-TCAT-GAG-A----AAA--TGG--TCA--TG--AAAGTT-G-ACCAGAAGAAAACCTGG-AGGA---T 6794
LST.CD.88.485 -CT-----G--GAA---C---AT------A-CATCA--CTTTAATAACAG---A------A-TCAT-GAG-A----AAA--TGG--TGA-CTG--AAAGT-GG-GCCAGAGGAAAACCTGG-GGGA---T 6791
LST.CD.88.524 -CT--T--G--GAA---T---CTT-----A-CATCA---TTTAACAATAG---A------A-TCAC-GAG-G----AAA--TGG--TGA--TG--AAAAT-GG-ACCACATGAAGACTTAG-AGGA---C 6797
LST.KE.-.lho7 -C-A-------GAAG--T---CTA--A--T-CTTCA--TTTCAACAATAG---C------A-TCAT-GAG-A----AAA--T-C--TGAG-TG--AAAAT-GG-ACCACATGAGGACTTAG-AGG---CC 7869
SUN.GA.98.L14 TC---G--C--GAAG--T--CTTG--C--A-CATCA--GTTTAATAACAG---A------G--AAC-GAG-G----AAGCATGG--C---CTG--AA-G-TGG-AGACTGGAAGGTAAATG-ATCA--CA 7937
MND-1.GA.-.MNDGB1 C----------G-AAG--A---TTA--A--T-CATCA---TTTAATAAT--C--A--G---A--CAT-GGG-A----AAA--TAC--TGA--TGC-AAAGATAG--AGTAATGAAACAAAA..... 7217
MON.CM.99.L1 T-T--GTC---GAGT--C--CCTA-----ACGG-AA--C.....--TG-GCAG---A----T-----C--T-CC--C--GA-T---G-G--GTTGTA-C-AAAC..... 7579
MON.NG.-.NG1 T-CA-CTCC--GAGTR-C---CTG--A--A-GACAA--G.....--TG-GCAG---A-C--T-----C----CC--CT-GG--GATG-G--GTTGTA-GGCAGC..... 6204
MUS.CM.01.1085 T-T--GTC---GT-T--C--CTTG-----ACG--AG--C.....-TTG--C-----GAG--T-----A--------T--CA-TGACG---CCATGTA--GAGAT..... 7547
DEB.CM.99.CM40 -CG-----GC--CAT--C---CTA-----G-AAGAA---.....G-GT-G-AG--CAGC---C-C---T-T-CAT-TG-T--TGATGT----CACTA----GGG..... 7350
DEB.CM.99.CM5 --G------C--CATG-A---TTG-----A-AGGAA---.....G--T-G-AG---CAT--TTC-G--T-T-CA--TG-T--TGAGGT--C-CA-TA-GGGCAA..... 7271
SYK.KE.-.KE51 T-T---ACC--GCT-------TTG--A--T-CAGAA--G.....--C----AG--CA-T--T--------T-CTG-AT--AC-GAT-T---GTACTA-CCAAAT..... 7273
SYK.KE.-.SYK173 T-T---A-----TTG--A--CCTG--C--T-CGGCA--G.....--T----A----A-T--C---G-C--T-CAG-AT-GACGGAC-T---ATACTACCCAGGA..... 7577
PLV
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 for stem-loop positions
V5 region
Rev responsive element start
H1B.FR.83.HXB2
Env
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Env
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173

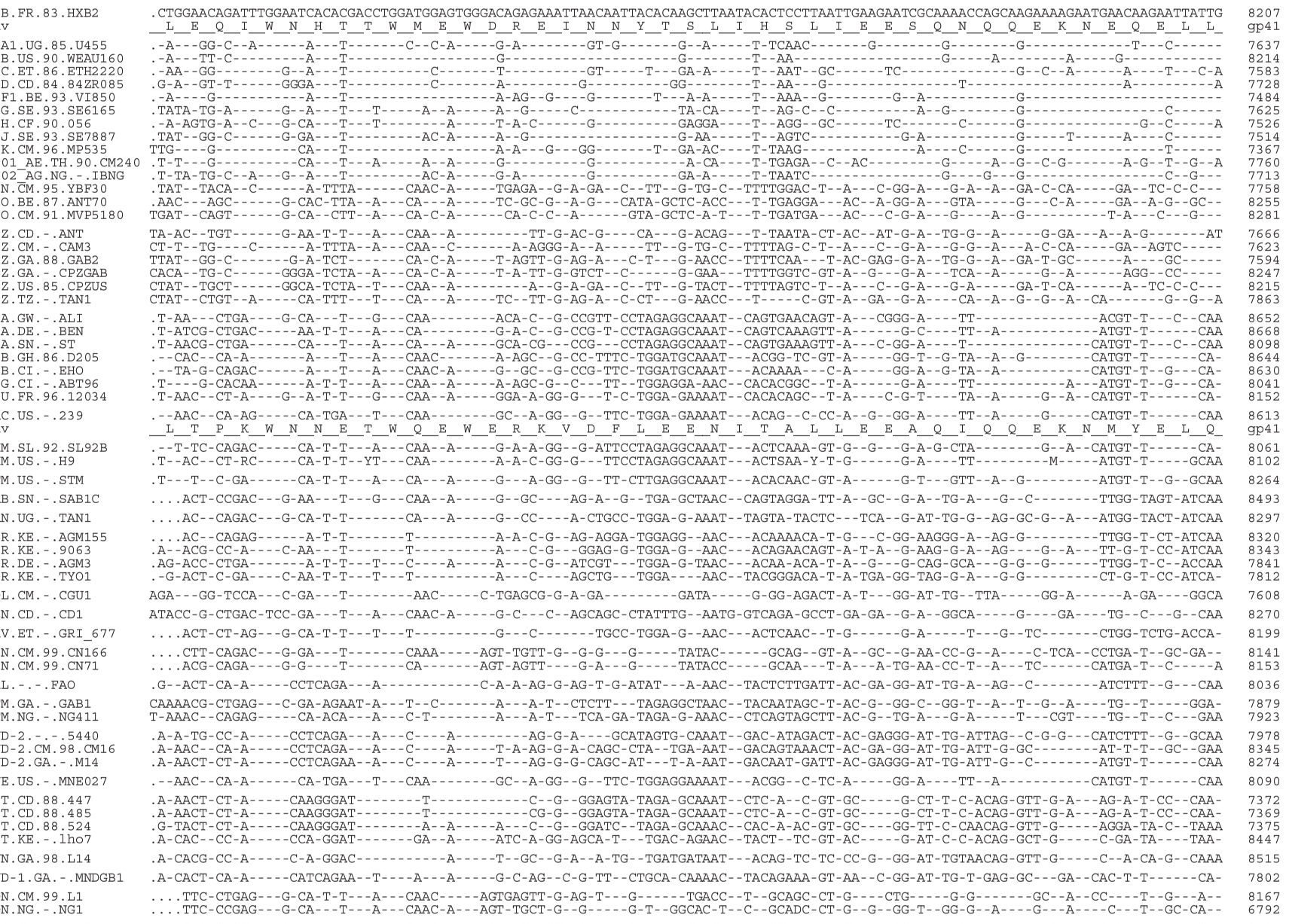
```
                Rev responsive element secondary structure in RNA, see Charpentier, J Mol Evol 266:950-956(1997)
                see HIV-1 complete genome annotation in compendium for stem-loop positions
                                        gp120 \/ gp41
H1B.FR.83.HXB2      GCAG..............AGAGAAAAAAGAGCA...GTGGGA...ATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAG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

H1A1.UG.85.U455     -G--..............----------------....T----...C--------A-C--------------A--------T--------A--------G--G-----A----------------------  7280
H1B.US.90.WEAU160   ----..............----------------..........-------ATGC--------A-------------------------------------------G------------------------  7857
H1C.ET.86.ETH2220   -G--..............----------------GCA.........C-------------------------------------------------------A-----A----------------------  7226
H1D.CD.84.84ZR085   -G-A..............----------------...A-A----...C--------A-------------------------------G-----------G------------------------------  7371
H1F1.BE.93.VI850    ----..............----------------...C------...C----------------------TC-T--G-ACAGCA-GGAGCA---------G--G--G--A----------------------  7127
H1G.SE.93.SE6165    -G--..............----------------....--T-----.C-G------G-C--------------A-------------------------G--G-----A----------------T----  7268
H1H.CF.90.056       -G--..............----------------.....---------.---CT-------------------------------------------G--G-----A------------------G  7169
H1J.SE.93.SE7887    -G--..............----------------...--------...---------G--------------------A----------T--------------G-----A----------------T----  7157
H1K.CM.96.MP535     ----..............----------------....-A----...T--------G----T----A-------------------------------G--G-----A----------------------  7010
H101_AE.TH.90.CM240 -G--..............----------G-----...----------...---------A--A--T----------A-------------------------G--G-----A----------------------  7403
H102_AG.NG.-.IBNG   -G--..............----------------G....-T------.C-G------G-C--------------A-------------------------GCGG-----A----------------------  7356
H1N.CM.95.YBF30     -AGT..............----------------GCCT-T------.C----T--GC----T--------TC-T----------G-----------------G----A------------------C-G  7398
H1O.BE.87.ANT70     T.................----------------....A------...T-G---ATGC-A---T-G---G-TC-AA-T--------T-----------------GG---CA------G--------A--CAC  7904
H1O.CM.91.MVP5180   AA-CATTCACACCCCTCAC--G----------------....A-----...T-G---ATGC-A---T-G---G-GC-AA-T--------T-----------------GG---CAG-----------G-A--CAC  7921

CPZ.CD.-.ANT        AA-ACAACACTCC.....---C---------GCATT...---.....-----GCTG--C---T-G--TC-TC-CA-T-----T--C--T--A----------G-----AG---------C----A----G  7282
CPZ.CM.-.CAM3       ATCA..............----------GCCT-T------.C-G--T--GC----T----------A-------------------------------A---G--GT---------------C-G  7263
CPZ.GA.88.GAB2      -GCAAAAGCACTTGAT...---TCC---------GCCT-T------.C-G--G--GC----T-----A---C-T-------------------------A--TG---T----------A--TC--  7234
CPZ.GA.-.CPZGAB     -GCAAGACAGAAAGAC...---C-G---------GCCT-C--T...C-G-----AC----------------C----T-------------------GG--G-A----------------C--  7887
CPZ.US.85.CPZUS     -A-A..............----------GCCT-C-----...C----T--AC-------G-----TC-T-------------------------------A---G-AGT----------------C--  7855
CPZ.TZ.-.TAN1       TA--GCCAATCATACT...--GTC--G-----ATGTG...--C...------CTG-----T-----A--TC-TA-T----------T--A----------G-----AG---------C----------  7476

H2A.GW.-.ALI        T...ACT......CCAAGG--GA-T-------GTGTGT-C...GTGC----G......---T-A--T--TC-C-CGA-------TTCTG-A------A-G--AG-TT-A------T-T-CT---T-TC-G  8307
H2A.DE.-.BEN        A...ACT......CCAGTG--GA-C-------GTGTGT-C...GTGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------GCGG--CC--------T-A-CC---T--C-G  8323
H2A.SN.-.ST         T...GCT......CCAGTG--GA-T-------GTGTAT-C...GTGC----G......---T-A--T--TC-CACGA-------GCTG-A--------G--G--CT--------T---CT---T-TC-G  7753
H2B.GH.86.D205      A...GTA......AAACCG--GA-T-------GTGTAA---...GTGC-------.....---T-G--A---C-T-C-ATG-----TTCTG-A---------A-G--TT--------T-A-CT---T--C-G  8299
H2B.CI.-.EHO        A...GTG......ACACCG--GA-T-------GTGTAC-T...GTGC----G......--TT-G--A---C-T-CGA-G-----TTCTG-A--------G--G--CT-----T--T---CT---T--C-G  8285
H2G.CI.-.ABT96      T...GTG......ACACCG-A-A-T-------GTGTAT-T...GTGC----G......--TT-G--A--TC-T-C-ATG-----TTCTG-A------A----G--GT--------T---CT---T--C-G  7684
H2U.FR.96.12034     A...GTG......ACACCG--GA-T-----G-G-GTAT-T...GTGC-T--G......---T-G--T--TC-C-CCA-G-----TTCTG-A-----T--G--G--GT--------T-A-CT---T--C-G  7807

MAC.US.-.239        TGGTGGC......ACCTCA---A-T-------GGGTCT-T...GTGC----G......---T-G--T--TC-C-C-A-G-----TTCTG-A--------G--G--GT---------C-CT---T--C--  8268
Env gp120           __G__G__.__._T__S__R__N__K__R__G__V__F__._V__L__G__._.__F__L__G__F__L__A__T__A__G__S__A__M__G__A__A__S__L__T__L__T__A__Q__S__R_  gp41

SMM.SL.92.SL92B     A...AGT......ACCTCA-AGA-T-------G-GTCT-T...GTGC----G......---T-G--T--TC-C-CGA-G-----TTCTG-A----------G--GC-A--------T-CT---T--C-G  7716
SMM.US.-.H9         AACTGGT......GCCTCA---A-T-------GGGTCT-T...GTGC----G......---T-G--T--TC-C-CGA-------TTCTG-A--------G--G--GT--------T---CT---T--C-G  7757

STM.US.-.STM        T...AGT......ACCTCA---ACT--G----GGGTCT-C...GTGC----G......---T-G--T--TC-C-CGA-------TTCTG-A----------G--GC----------CT---T--C-G  7919

SAB.SN.-.SAB1C      C...........CCCGAA---C-------G-TGCCAT-C...GTGC----G......-----A--T-------------T---GCTG-A-----T-----GG-G-CAG-C--------C---T-TCAG  8151

TAN.UG.-.TAN1       --CAACT.............C-------G--G-TGCCCT-C...GTGT-G--G......-----A--C--T-----G-----T---GCTG-A-----T---A-AG-G-CAG--T-------C---T--CAG  7955

VER.KE.-.AGM155     A...GGC......CAAGAG--GC-----C---TCCCGT-C...GTGC----G......-----A--C----------T--T--G-CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG  7978
VER.KE.-.9063       A...GGA......CAAGAC---TC----C---TCCCGT-T...GTGC----G......-----A--C-----A-----T--T----CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG  7998
VER.DE.-.AGM3       A...GGT......CATGAC---AC---GC---TCCCGT-C...GTGC----G......-----A--C-----A-----T--T--G-CTG-A-----A-----GG-G-CAG-C--------C---T-TCAG  7496
VER.KE.-.TYO1       A...GGA......CATGAG--GC----G----TCCCGT-T...GTGC----G......---T-A--C--------G--T--T----CTG-A-----A-----GG-G-GC-GC--------C---T--C-G  7467

COL.CM.-.CGU1       TGTAAATTGGCATCATGGG---C----G----GCATCT-T-C-TTCTCCATC.........T-A-CCC-----A-T-GG--C--TGCTG-A------T----G--TG--G--------A-C-----TCA-  7257

DEN.CD.-.CD1        A-CAGAATCGGCAGAGTCTGTTTC-CGCC--AGGCGG-AA-TGTCTT---TGCT-GG-C--A-A--A--TC-CA-C----------CTGTC--------G-T-GG--CAGGCT-AG--------T--C-G  7907

GRV.ET.-.GRI_677    C-CCACA.....................G-TGCCAT-C...GTGC----G......--T--A--C----------T--T----CTG-A----------GG---CA--------A--C---T-TC-G  7857

GSN.CM.99.CN166     C-CC............GCTGTG--G---------C...C-TTC-...T-G---ATA-CC---T-GAC---TC-CA-T--G-----T-C---G----------GG---CAG--------A----T--C-G  7787
GSN.CM.99.CN71      T-CAGGGGTG......CATGTGA----G--G--C...A-TTC-...T-G--CATA-CC---T-GAC---TC-CA-T--G-----TGC---G----------GG---CAG-----------AT--C-G  7799

DRL.-.-.FAO         --C-GGA.........CAC--GC-G--G----GTGCA...-TCGTGC-T--G......A--------CC-----A-CTTG-----TTC-G-A---T--T---TGG-G-C-G-CA----T--C---T--CAG  7691

RCM.GA.-.GAB1       -...................-A-C-G-----G-T-...CCTTTG...G-GCTG-G-...--T--A--T---C-CTC---------T-CTG-A----------GG-G-CAG-C-----T--C---T--C-G  7537
RCM.NG.-.NG411      A...................-A-C----G----T-...CCTTTG...G-GCTG-G-...--T--A--T---C-CTCG--------T-CTG-A----------GG-G-CAG-C-----A--C---T--C-G  7581

MND-2.-.-.5440      --CTAAT.........ACA--GC----------G-GCG...-TCTTAC-T---.....A-G--C--CC--C-A--TTTG-----TTC-G-G------T---TGG-GG--G-C-----T--C---T--CAG  7633
MND-2.CM.98.CM16    --CAGAA.........GCT--GC-G--G--G-G-GCG...-TCTTGC-T--G.....A-GT-C--CC--C-A--TTTG-----TTC---G------T---TGG-GG--G-C-----T--C---T-TCAG  8000
MND-2.GA.-.M14      A-CAGAT.........GCA-AGC-G--G----G-GCA...-TCTTGC-T---.....A-G--C--CC-------CTTG-----TTC-G-G------T---TGG-GG--G-C-----T--C---T--CAG  7929

MNE.US.-.MNE027     TGGTGGC......ACCCCA---A-T-------GGGTCT-T...GTGC----G......---T-A--T--TC-C-C-A-G-----TTCTG-A--------G--G--GT----------C-CT---T--C-G  7745

LST.CD.88.447       C............AGTAGT-AGC-G--G------GCACC-CTGGCTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C--A---CA-  7030
LST.CD.88.485       C............AGTAGT-AGC-G--G------GCACC-CTGGCTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C------CA-  7027
LST.CD.88.524       C............AGGGGA--GC---C-------GCTCCACT-GCTC----T......GC-T-G--AC-TC-CA-T--T-----T-CTG-A------TT--TA--G-C--TA--A--T--C------CA-  7033
LST.KE.-.lho7       C............AGGGGA--GC---CC------GCGCCTCTGGCTT----------.GC-T-G--AC-TC-CA-T--T-----C-CTG-A------TT--TA--G-C--TA--A--T--C------CA-  8105

SUN.GA.98.L14       C..............AGG-AG-----G--G---GCACCT-TTGCCC----------.GCTT-G-CAC-TC-CA-T--T--G--T-CTG-A------TT--TA--G-C--TA--A--T--C------CA-  8173

MND-1.GA.-.MNDGB1   A-CTGAA......CATCAT-A---G--G----G-GCA...-T-ATAC-T--T......A-------TC-GC-CTCGCTG------TC-G-G------T---TG--GG--G-A-----T--C--AT-TCAG  7457

MON.CM.99.L1        -GTA............CAC--G-----G------ATT......ACTT-G--TATGGCC--T--G--C---C-CA-CA-G------G----A----------GG-G-CAG-C----------T-TC-G  7813
MON.NG.-.NG1        --TA............ATG-------G-C----TATC......ACAT-G---ATGGCC--T--G--C--TC-CA-CA-G------G----G----------GG-G-CAG-C----------T--C-G  6438

MUS.CM.01.1085      C-CA.........CATGTGCAT-C--G-GCCATCTCA...TTGGGC---ACC......--TT-A--A--TC-CA-C---------G----A----------GG-G-CAG-C----------T-TC-G  7784

DEB.CM.99.CM40      A-C-.......GAGTCGGTA---C-C--G--G---GCA-GAATTGCTT-T--T......C-GG-A-CA--TC-CA-TA-------TGCTG----------G----CAG----------T-TC-C  7590
DEB.CM.99.CM5       A-C-.......GAGTCTGCA---C-------G--GGCG-GAATTGCTT-T--T......C-AG-G-CT--TC--A-TA-------TGCTG----------G----CAG----------T--C-C  7511

SYK.KE.-.KE51       --CT............ACG-A-C-G---------GCACCATTGGCAC-T--G......-----G--A--AC-CTCG-----G--T-CTG-A-----GAGC--TG-G-CAG----T--AC-C---T--CAG  7510
SYK.KE.-.SYK173     --CT.........AATACC--G--------G---GCACCATTGGCAC-T--G......-----G--A--AC-CTCG--T--G--C-CTG-A-----G-GC--TG-G-CAG----A--AC-C---T-TCAG  7817
```

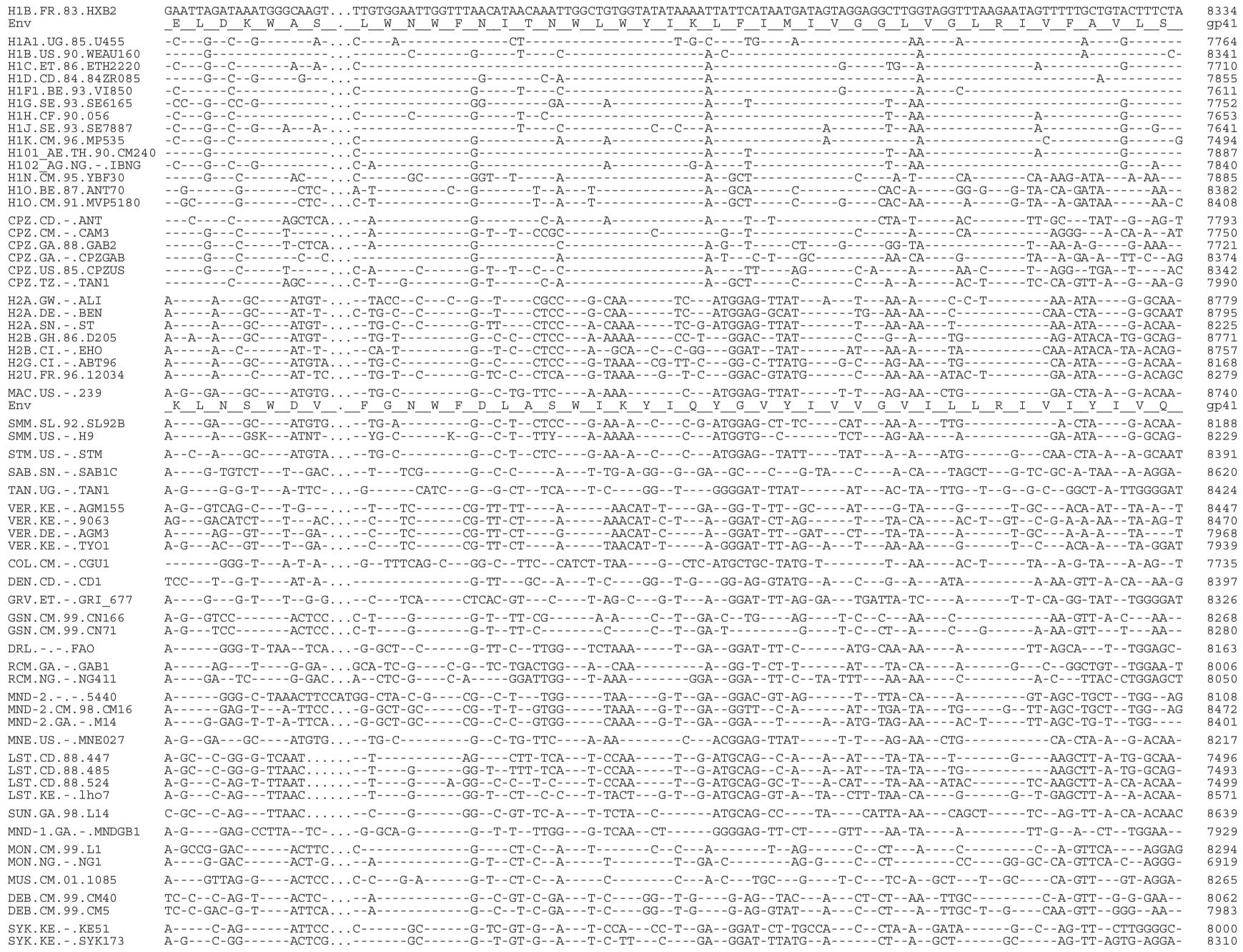
PLV
complete genomes

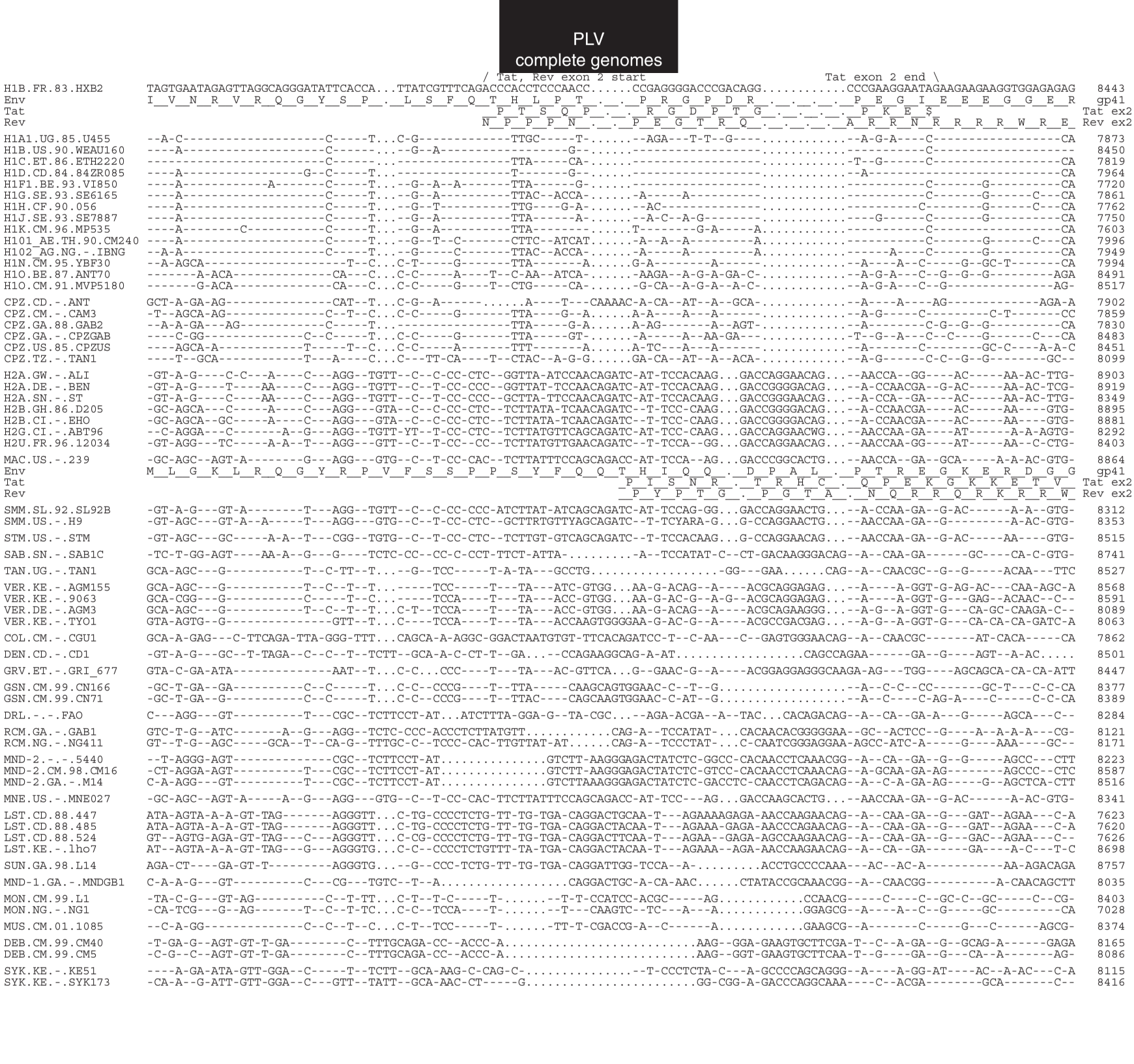


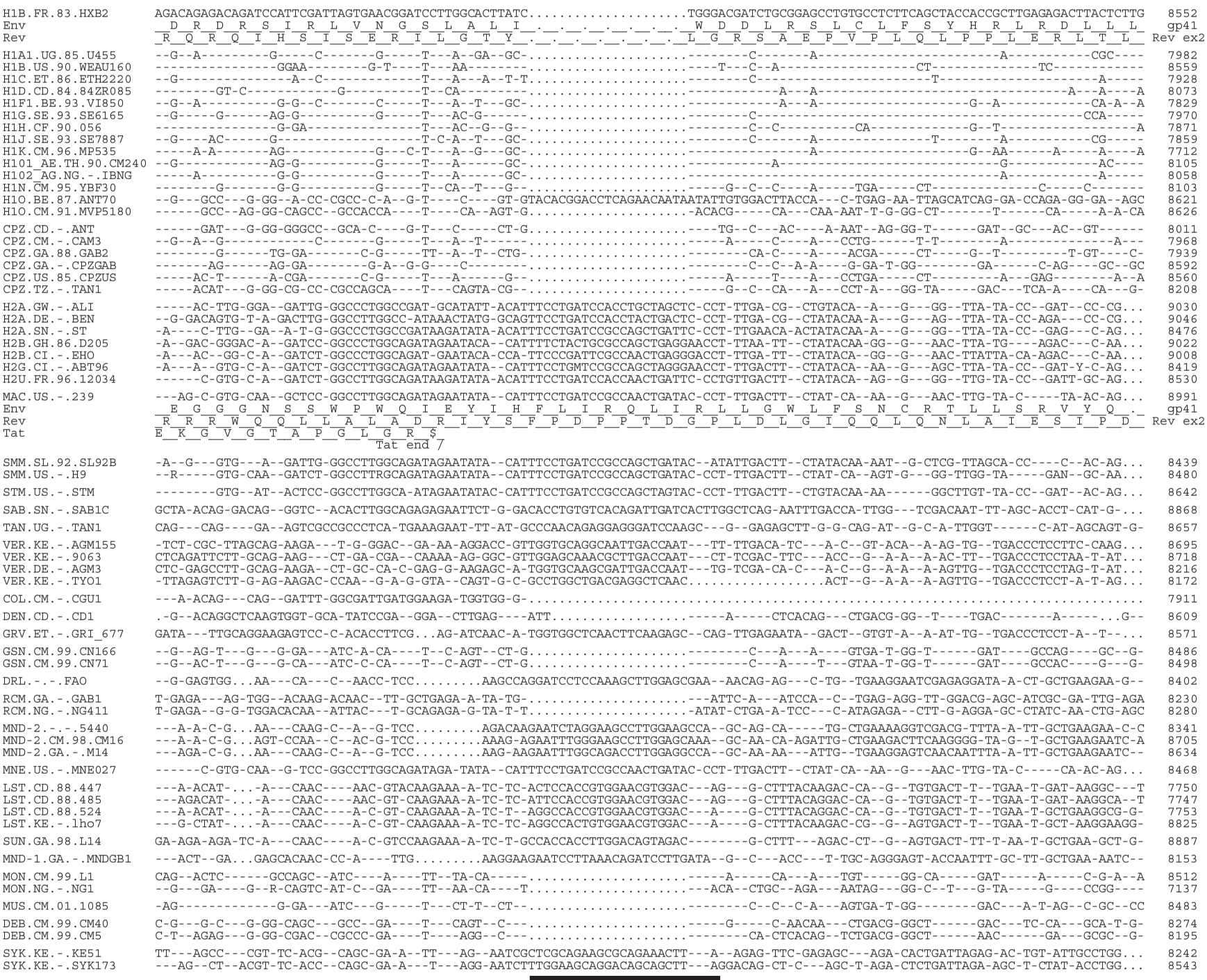


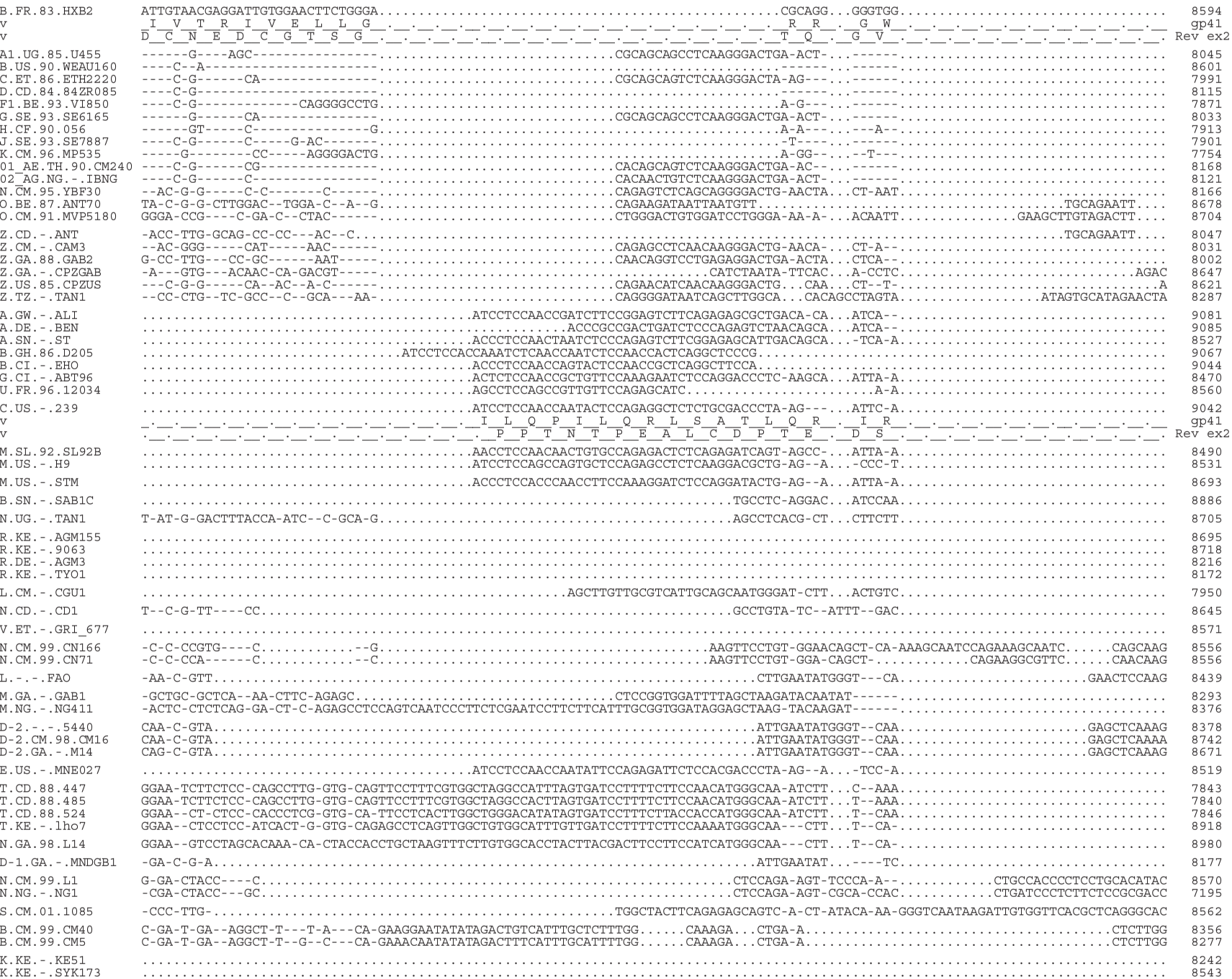
PLV
complete genomes

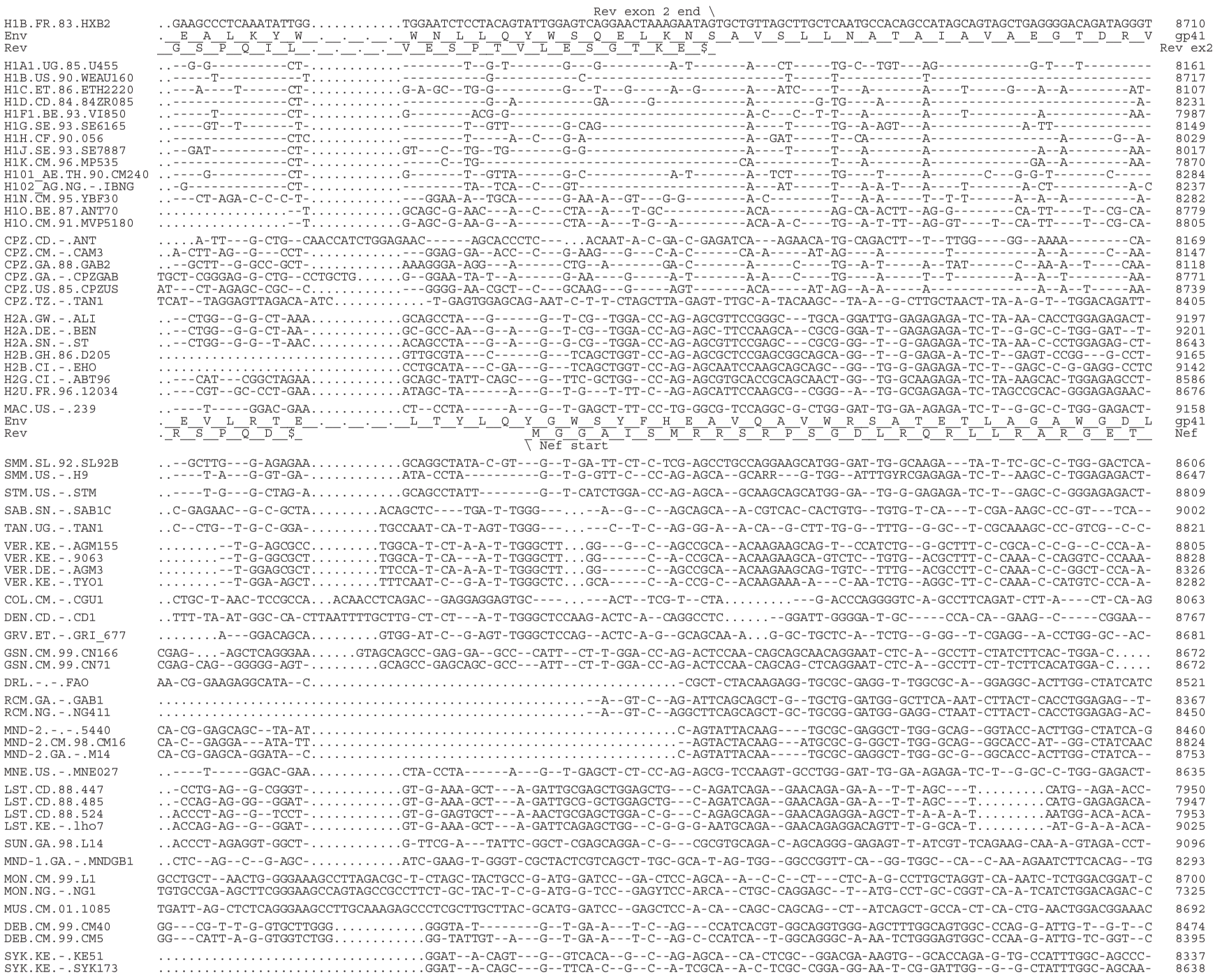

H1B.FR.83.HXB2 .CTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTG 8207
Env __L__E__Q__I__W__N__H__T__T__W__M__E__W__D__R__E__I__N__N__Y__T__S__L__I__H__S__L__I__E__E__S__Q__N__Q__Q__E__K__N__E__Q__E__L__L_ gp41
H1A1.UG.85.U455 .-A---GG-C--A------A---T-------C--C-A-----G-A--------GT-G-------G--A-----T-TCAAC-------G--------G-------G-----------T---C------ 7637
H1B.US.90.WEAU160 .-A---TT-C----------A---T-----------------G-----------------G--------T--AA-----------------G-----A--------A-----G---------- 8214
H1C.ET.86.ETH2220 .-AA--GG---------G--A---T----------C-------T----------GT-----T---GA-A-----T--AAT--GC------TC-------------G--C--A------A---T---C-A 7583
H1D.CD.84.84ZR085 .G-A--GT-T------GGGA---T----------C-------A----------G---------GG--------T--AA---G---------------T-----------------A---------A 7728
H1F1.BE.93.VI850 .-A----G----------A---T-----------------A-AG--G----G-------T---A-A-----T--AAA--G--------G--A----------G---------------------- 7484
H1G.SE.93.SE6165 .TATA-TG-A------G--A---T---T-----A--A-----A--G-----C-----------TA-CA----T--AG-C--C----------A--G--------G---------------C------ 7625
H1H.CF.90.056 .-A-AGTG-A--C---G-CA---T---T--------A-----T-A-C------G----------GAGGA----T--AGG--GC------TC-------C------G--------------G--C-----A 7526
H1J.SE.93.SE7887 .TAT--GG-C------G-GA---T---------AC-A-----A--G-----------------G-AA-----T--AGTC----------G-A----------G-----T------A----C------ 7514
H1K.CM.96.MP535 TTG----G----------CA---T-----------------A-A---G---GG-------T--GA-AC----T-TAAG--------------A----C------G--------------T------ 7367
H101_AE.TH.90.CM240 .T-T---G----------CA---T---A-----A--A-----G----------G----------A-CA----T-TGAGA--C--AC---------G-----A--G--C-G-------A-G--T--G--A 7760
H102_AG.NG.-.IBNG .T-TA-TG-C--A---G--A---T---------AC-A-----G-A--------G---------GA-A----T-TAATC----------G----G--G-------------C--G--- 7713
H1N.CM.95.YBF30 .TAT--TACA--C------A-TTTA------CAAC-A-----TGAGA--G-A-GA--C--TT--G-TG-C--TTTTGGAC-T--A---C-GG-A---G-A--A--GA-C-CA-----GA--TC-C-C--- 7758
H1O.BE.87.ANT70 .AAC---AGC------G-CAC-TTA--A---CA---A-----TC-GC-G--A-G---CATA-GCTC-ACC---T-TGAGGA---AC--A-GG-A---GTA-----G---C-A-----GA--A-G--GC-- 8255
H1O.CM.91.MVP5180 TGAT--CAGT------G-CA--CTT--A---CA-C-A------CA-C-C--A------GTA-GCTC-A-T---T-TGATGA---AC---C-G-A---G-----A-G----------T-A---C---G--- 8281
CPZ.CD.-.ANT TA-AC--TGT------G-AA-T-T---A---CAA--A---------TT-G-AC-G----CA---G-ACAG---T-TAATA-CT-AC--AT-G-A--TG-G--A-----G-GA---A--A-G-------AT 7666
CPZ.CM.-.CAM3 CT-T--TG----C------A-TTTA--A---CAA--C-------A-AGGG-A--A-----TT--G-TG-C--TTTTAG-C-T--A---C--G-A---G-G--A---A-C-CA-----GA--AGTC----- 7623
CPZ.GA.88.GAB2 TTAT--GG-C-------G-A-TCT-------CA-C-A----T-AGTT-G-AG-A---C-T---G-AACC--TTTTCAA---T-AC-GAG-G-A--TG-G--A--GA-T-GC-----A----GC----- 7594
CPZ.GA.-.CPZGAB CACA--TG-C------GGGA-TCTA--A---CA-C-A-----T-A-TT-G-GTCT--C------G-GAA---TTTTGGTC-GT-A--G---G-A---TCA--A-----G--A------AGG--CC----- 8247
CPZ.US.85.CPZUS CTAT--TGCT------GGCA-TCTA--T---CAA--A---------A--G-A-GA--C--TT--G-TACT--TTTTAGTC-T--A---C--G-A---G-A-----GA-T-CA------A--TC-C-C--- 8215
CPZ.TZ.-.TAN1 CTAT--CTGT--A-----CA-TTT---T---CA---A-----TC--TT-G-AG-A--C-CT---G-AACC---T-----C-GT-A--GA--G-A----CA--A-G--G--A--CA-------G--G--A 7863
H2A.GW.-.ALI .T-AA---CTGA----G-CA---T---G---CAA--------ACA-C--G-CCGTT-CCTAGAGGCAAAT--CAGTGAACAGT-A---CGGG-A----TT---------------ACGT-T---C--CAA 8652
H2A.DE.-.BEN .T-ATCG-CTGAC-----AA-T-T---A---CA---------G-A-C--G-CCG-T-CCTAGAGGCAAAT--CAGTCAAAGTT-A------G-C----TT--A------------ATGT-T------CAA 8668
H2A.SN.-.ST .T-AACG-CTGA------CA---T---A---CA---A----GCA-CG---CCG---CCTAGAGGCAAAT--CAGTGAAAGTT-A---C-GG-A----T---------------CATGT-T---C--CAA 8098
H2B.GH.86.D205 .--CAC--CA-A-------A---T---A---CAAC-------A-AGC--G-CC-TTTC-TGGATGCAAAT---ACGG-TC-GT-A-----GG-T--G-TA--A--G-------CATGT-T------CA- 8644
H2B.CI.-.EHO .--TA-G-CAGAC------A---T---A---CAAC-A-----G--GC--G-CCG-TTC-TGGATGCAAAT---ACAAAA---C-A-----GG-A--G-TA--A----------CATGT-T--G---CA- 8630
H2G.CI.-.ABT96 .T----G-CACAA------A-T-T---A---CAA--A-----A-AGC-G--C---TT--TGGAGGA-AAC--CACACGGC--T-A------G-A----TT----------A---ATGT-T--G---CA- 8041
H2U.FR.96.12034 .T-AAC--CT-A----G--A-T-T---G---CAA--A-----GGA-A-GG-G---T-C-TGGA-GAAAAT--CACACAGC--T-A----C-G-T----TA--A-----G--A--CATGT-T------CA- 8152
MAC.US.-.239 .--AAC--CA-AG-----CA-TGA---T---CAA--------GC--A-GG--G--TTC-TGGA-GAAAAT---ACAG--C-CC-A--G--GG-A----TT--A-----G-----CATGT-T------CAA 8613
Env __L__T__P__K__W__N__N__E__T__W__Q__E__W__E__R__K__V__D__F__L__E__E__N__I__T__A__L__L__E__E__A__Q__I__Q__Q__E__K__N__M__Y__E__L__Q_ gp41
SMM.SL.92.SL92B .--T-TC-CAGAC-----CA-T-T---A---CAA--A-----G-A-A-GG--G-ATTCCTAGAGGCAAAT---ACTCAAA-GT-G--G---G-A-G-CTA-------G--A--CATGT-T-----CA- 8061
SMM.US.-.H9 .T--AC--CT-RC-----CA-T-T--YT---CAA--------A-A-C-GG-G---TTCCTAGAGGCAAAT---ACTSAA-Y-T-G------G-A----TT---------M-----ATGT-T-----GCAA 8102
STM.US.-.STM .T---T--C-GA------CA-T-T---A---CA---A-----G---A-GG--G--TT-CTTGAGGCAAAT---ACACAAC-GT-A------G-T---GTT--A--G---------ATGT-T--G--GCAA 8264
SAB.SN.-.SAB1C ....ACT-CCGAC---G-AA---T---G---CAA--A-----G--GC----AG-A--G--TGA-GCTAAC--CAGTAGGA-TT-A--GC--G-A--TG-A---G--C--------TTGG-TAGT-ATCAA 8493
TAN.UG.-.TAN1 ....AC--CAGAC---G-CA-T-T-------CA---A-----G--CC----A-CTGCC-TGGA-G-AAAT--TAGTA-TACTC---TCA--G-AT-TG-G--AG-GC-G--A---ATGG-TACT-ATCAA 8297
VER.KE.-.AGM155 ....AC--CAGAG------A-T-T-------T----------A-A-C-G--AG-AGGA-TGGAGG--AAC---ACAAAACA-T-G---C-GG-AAGGG-A--AG-G---------TTGG-T-CT-ATCAA 8320
VER.KE.-.9063 .A--ACG-CC-A----C-AA---T-------T----------A---C-G---GGAG-G-TGGA-G--AAC---ACAGAACAGT-A-T-A--G-AAG-G-A--AG----G--A---TT-G-T-CC-ATCAA 8343
VER.DE.-.AGM3 .AG-ACC-CTGA-------A-T-T---T---C----A-----A----C-G--ATCGT---TGGA-G-TAAC---ACAA-ACA-T-A--G---G-CAG-GCA---G-G--G------TTGG-T-C--ACCAA 7841
VER.KE.-.TYO1 .-G-ACT-C-GA----C-AA-T-T---T---T----------A---C----AGCTG---TGGA----AAC--TACGGGACA-T-A-TGA-GG-TAG-G-A---G-G---------CT-G-T-CC-ATCA- 7812
COL.CM.-.CGU1 AGA---GG-TCCA---C-GA---T---------AAC-----C-TGAGCG-G-A-GA----------GATA-----G-GG-AGACT-A-T---GG-AT-TG--TTA-----GG-A------A-GA----GGCA 7608
DEN.CD.-.CD1 ATACC-G-CTGAC-TCC-GA---T---A---CAAC-A-----G-C---C--AGCAGC-CTATTTG--AATG-GTCAGA-GCCT-GA-GA--G-A--GGCA-----G----GA----TG--C---G--CAA 8270
GRV.ET.-.GRI_677 ....ACT-CT-AG---G-CA-T-T---T---T----------G---C-------TGCC-TGGA-G--AAC---ACTCAAC--T-G------G-A-----T---G--TC-------CTGG-TCTG-ACCA- 8199
GSN.CM.99.CN166 ....CTT-CAGAC---G-GA---T-------CAAA-----AGT-TGTT-G--G-G---G-----TATAC------GCAG--GT-A--GC--G-AA-CC-G--A----C-TCA--CCTGA-T--GC-GA-- 8141
GSN.CM.99.CN71 ....ACG-CAGA----G-G----T-------CA-------AGT-AGTT----G-A---G-----TATACC-----GCAA---T-A---A-TG-AA-CC-T--A---TC------CATGA-T--C-----A 8153
DRL.-.-.FAO .G--ACT-CA-A-----CCTCAGA---A-----------C-A-A-AG-G-AG-T-G-ATAT---A-AAC--TACTCTTGATT-AC-GA-GG-AT-TG-A--AG--C--------ATCTTT--G---CAA 8036
RCM.GA.-.GAB1 CAAAACG-CTGAG---C-GA-AGAAT-A---T--C-------A---A-T--CTCTT---TAGAGGCTAAC--TACAATAGC-T-AC-G--GG-C--GG-T--A--T--G--A----TG--T-----GGA- 7879
RCM.NG.-.NG411 T-AAAC--CAGAG-----CA-ACA---A---C-T--------A---A-T---TCA-GA-TAGA-G-AAAC--CTCAGTAGCTT-AC-G--TG-A---G-A-----T---CGT----TG--T--C---GAA 7923
MND-2.-.-.5440 .A-A-TG-CC-A-----CCTCAGA---A---C----A---------AG-G-A----GCATAGTG-CAAAT--GAC-ATAGACT-AC-GAGGG-AT-TG-ATTAG--C-G-G---CATCTTT--G--GCAA 7978
MND-2.CM.98.CM16 .A-AAC--CA-A-----CCTCAGA---A---C----A-----T-A-AG-G-A-CAGC-CTA--TGA-AAT--GACAGTAAACT-AC-GA-GG-AT-TG-ATT-G-GC--------AT-T-T--GC--GAA 8345
MND-2.GA.-.M14 .A-AACT-CT-A-----CCTCAGAA--A---C----A-----T---AG-G-G-CAGC-AT---T-A-AAT--GACAAT-GATT-AC-GAGGG-AT-TG-ATT-G--C--------ATGT-T------CAA 8274
MNE.US.-.MNE027 .--AAC--CA-A------CA-TGA---T---CAA--------GC--A-GG--G--TTC-TGGAGGAAAAT---ACGG--C-TC-A-----GG-A----TT--A----------CATGT-T------CAA 8090
LST.CD.88.447 .A-AACT-CT-A-----CAAGGGAT--------T----------C--G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC-----G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA- 7372
LST.CD.88.485 .A-AACT-CT-A-----CAAGGGAT--------T----------CG-G--GGAGTA-TAGA-GCAAAT--CTC-A--C-GT-GC-----G-CT-T-C-ACAG-GTT-G-A---AG-A-T-CC--CAA- 7369
LST.CD.88.524 .G-TACT-CT-A-----CAAGGGAT--------A--A-----A---C--G--GGATC--TAGA-GCAAAC--CAC-A-AC-GT-GC----GG-TT-C-CAACAG-GTT-G-----AGGA-TA-C--TAAA 7375
LST.KE.-.lho7 .A-CAC--CC-A-----CCA-GGAT-------GA--A-----ATC-A-GG-AGCA-T---TGAC-AGAAC--TACT--TC-GT-AC-----G-AT-C-C-ACAG-GCT-G-----C-GA-TA----TAA- 8447
SUN.GA.98.L14 .A-CACG-CC-A-----C-A-GGAC----------A-----T--GC--G-A--A-TG--TGATGATAAT---ACAG-TC-TC-CC-G--GG-AT-TGTAACAG-GTT-G-----C--A-CA-G--CAAA 8515
MND-1.GA.-.MNDGB1 .A-CACT-CA-A-----CATCAGAA--T----A---A-----G-C-AG--C-G-TT--CTGCA-CAAAAC--TACAGAAA-GT-AA--C-GG-AT-TG-T-GAG-GC---GA--CAC-T-T------CA- 7802
MON.CM.99.L1 ....TTC-CTGAG---G-CA-T-T---A---CAAC-----AGTGAGTT-G-AG-T---G------TGACC--T--GCAGC-CT-G----CTG-----G-G-------GC--A--CC----T--G---A-- 8167
MON.NG.-.NG1 ....TTC-CCGAG---G-CA----T---A---CAAC-A---AGT-TGCT-G--G-----G-T--GGCAC-T--C--GCADC-CT-G--G--GG-T--GG-G--A----G----C---T--GC-CA-- 6792
MUS.CM.01.1085 ....C-G-CAGG----G-AA-T-T---A---CAAC-----AG---GTTGG-AG-----G-T--T-CAACT-----GGAG---T-G---TTGG-A----GA-----G---G----C-G--TA-----CAA 8138
DEB.CM.99.CM40 .....-T-CAGAC---G-CA-T-T-------CA---------A-TGA-GG-AGCA-----T---GATGAGTGGG-AGGAGCTT-AC-GAG-G-A---G-G-----G--G-G-----T---T-CCC--CAA 7935
DEB.CM.99.CM5 ....----CTGA-------A-T-T-------CAA--------A--GA--G-AGCA--C--T---GATGA-TGGG-AGGAGCCT-AC-GAG-G-A--GG-A--A-----G-GA----TG--T-CC---CA- 7856
SYK.KE.-.KE51 C-CAAC--CCCAA---G-AA---T---A---CATACA-----AC-GC-GG-AG-T----TG---GA-CAC---G-TAAT---C-GAG----G-T--GG-G--------G-------TG--T--TC-TACT 7873
SYK.KE.-.SYK173 T-CCA---CACAG-----AA-T-T---A---CACAGA-----ACA---GG-AG------TG--TGA-CAT--TG-TGGA--GC-CAG----G-A---G-A--A--G--G-GA----TG--T--TC--ACA 8183


H1B.FR.83.HXB2 GAATTAGATAAATGGGCAAGT...TTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTA 8334
Env _E__L__D__K__W__A__S__._L__W__N__W__F__N__I__T__N__W__L__W__Y__I__K__L__F__I__M__I__V__G__G__L__V__G__L__R__I__V__F__A__V__L__S__ gp41
H1A1.UG.85.U455 7764
H1B.US.90.WEAU160 8341
H1C.ET.86.ETH2220 7710
H1D.CD.84.84ZR085 7855
H1F1.BE.93.VI850 7611
H1G.SE.93.SE6165 7752
H1H.CF.90.056 7653
H1J.SE.93.SE7887 7641
H1K.CM.96.MP535 7494
H101_AE.TH.90.CM240 7887
H102_AG.NG.-.IBNG 7840
H1N.CM.95.YBF30 7885
H1O.BE.87.ANT70 8382
H1O.CM.91.MVP5180 8408
CPZ.CD.-.ANT 7793
CPZ.CM.-.CAM3 7750
CPZ.GA.88.GAB2 7721
CPZ.GA.-.CPZGAB 8374
CPZ.US.85.CPZUS 8342
CPZ.TZ.-.TAN1 7990
H2A.GW.-.ALI 8779
H2A.DE.-.BEN 8795
H2A.SN.-.ST 8225
H2B.GH.86.D205 8771
H2B.CI.-.EHO 8757
H2G.CI.-.ABT96 8168
H2U.FR.96.12034 8279
MAC.US.-.239 8740
Env _K__L__N__S__W__D__V__._F__G__N__W__F__D__L__A__S__W__I__K__Y__I__Q__Y__G__V__Y__I__V__V__G__V__I__L__L__R__I__V__I__Y__I__V__Q__ gp41
SMM.SL.92.SL92B 8188
SMM.US.-.H9 8229
STM.US.-.STM 8391
SAB.SN.-.SAB1C 8620
TAN.UG.-.TAN1 8424
VER.KE.-.AGM155 8447
VER.KE.-.9063 8470
VER.DE.-.AGM3 7968
VER.KE.-.TYO1 7939
COL.CM.-.CGU1 7735
DEN.CD.-.CD1 8397
GRV.ET.-.GRI_677 8326
GSN.CM.99.CN166 8268
GSN.CM.99.CN71 8280
DRL.-.-.FAO 8163
RCM.GA.-.GAB1 8006
RCM.NG.-.NG411 8050
MND-2.-.-.5440 8108
MND-2.CM.98.CM16 8472
MND-2.GA.-.M14 8401
MNE.US.-.MNE027 8217
LST.CD.88.447 7496
LST.CD.88.485 7493
LST.CD.88.524 7499
LST.KE.-.lho7 8571
SUN.GA.98.L14 8639
MND-1.GA.-.MNDGB1 7929
MON.CM.99.L1 8294
MON.NG.-.NG1 6919
MUS.CM.01.1085 8265
DEB.CM.99.CM40 8062
DEB.CM.99.CM5 7983
SYK.KE.-.KE51 8000
SYK.KE.-.SYK173 8310
PLV
complete genomes

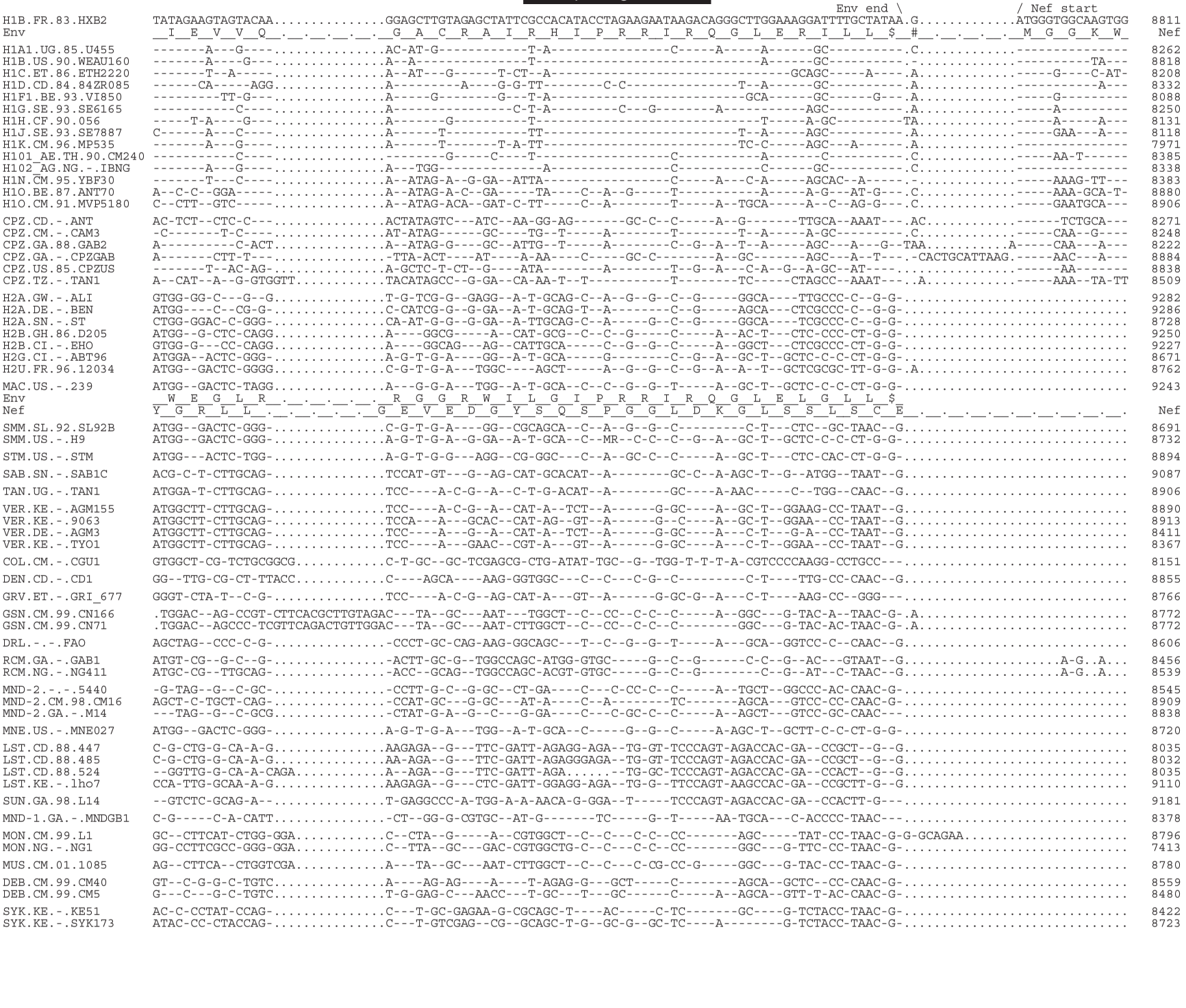

PLV
complete genomes
/ Tat, Rev exon 2 start
Tat exon 2 end \
H1B.FR.83.HXB2
Env
Tat
Rev
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Env
Tat
Rev
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173

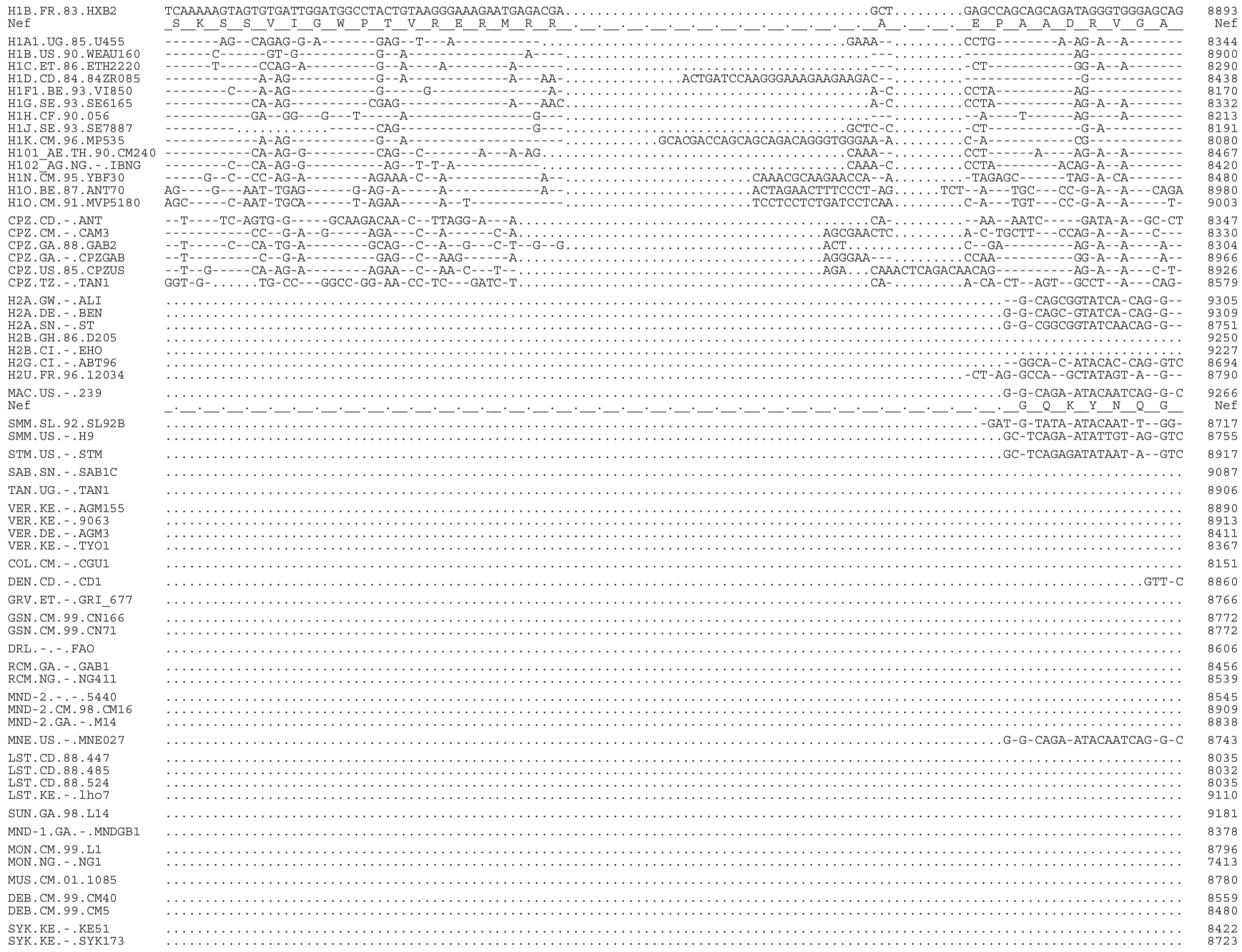

H1B.FR.83.HXB2
Env
Rev
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Env
Rev
Tat
Tat end 7
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes


H1B.FR.83.HXB2
Env
Rev
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Env
Rev
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
8594
gp41
Rev ex2
8045
8601
7991
8115
7871
8033
7913
7901
7754
8168
8121
8166
8678
8704
8047
8031
8002
8647
8621
8287
9081
9085
8527
9067
9044
8470
8560
9042
gp41
Rev ex2
8490
8531
8693
8886
8705
8695
8718
8216
8172
7950
8645
8571
8556
8556
8439
8293
8376
8378
8742
8671
8519
7843
7840
7846
8918
8980
8177
8570
7195
8562
8356
8277
8242
8543


Rev exon 2 end \
Env
Rev
Rev ex2
gp41
\ Nef start
Nef
H1B.FR.83.HXB2
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes

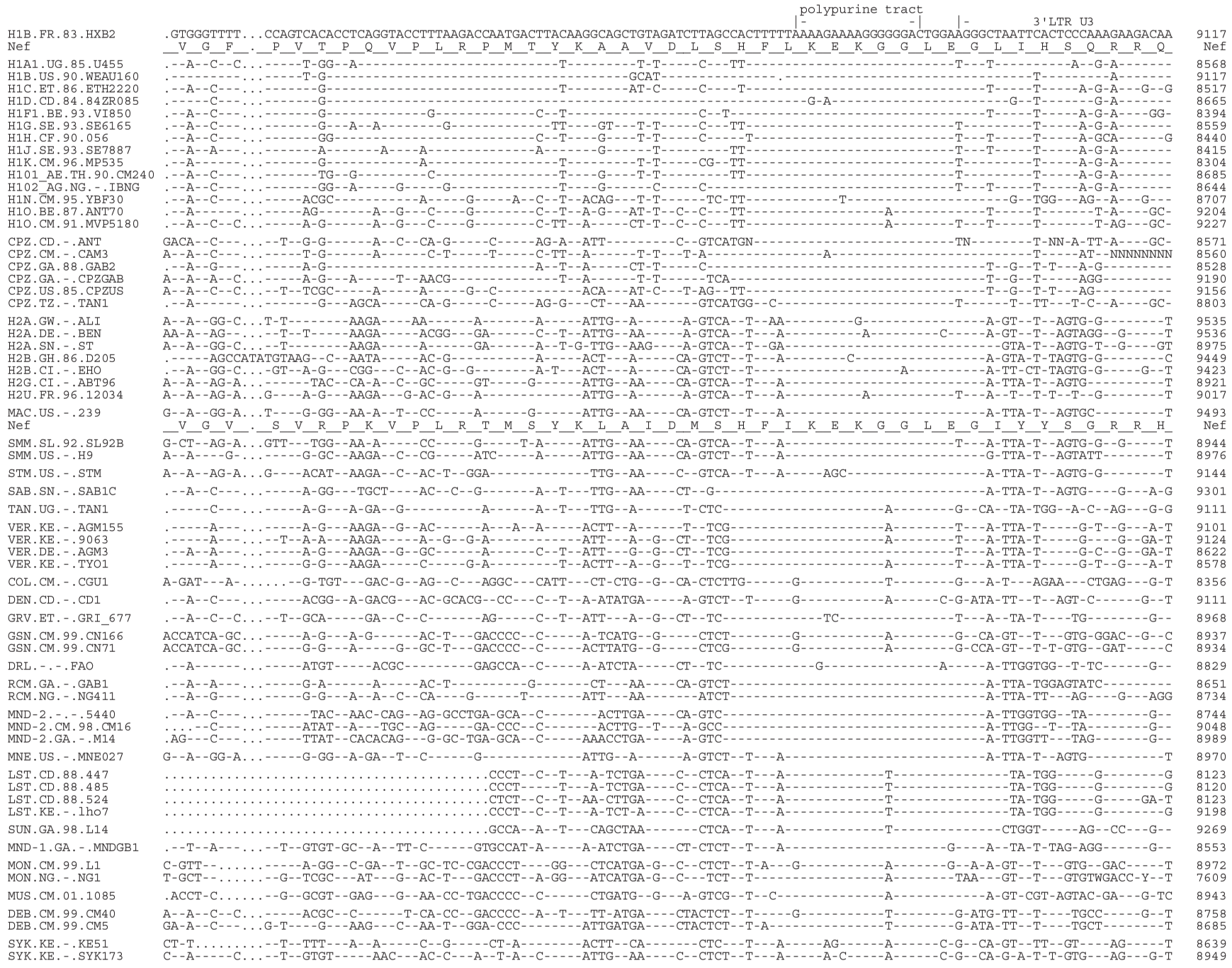
PLV
complete genomes

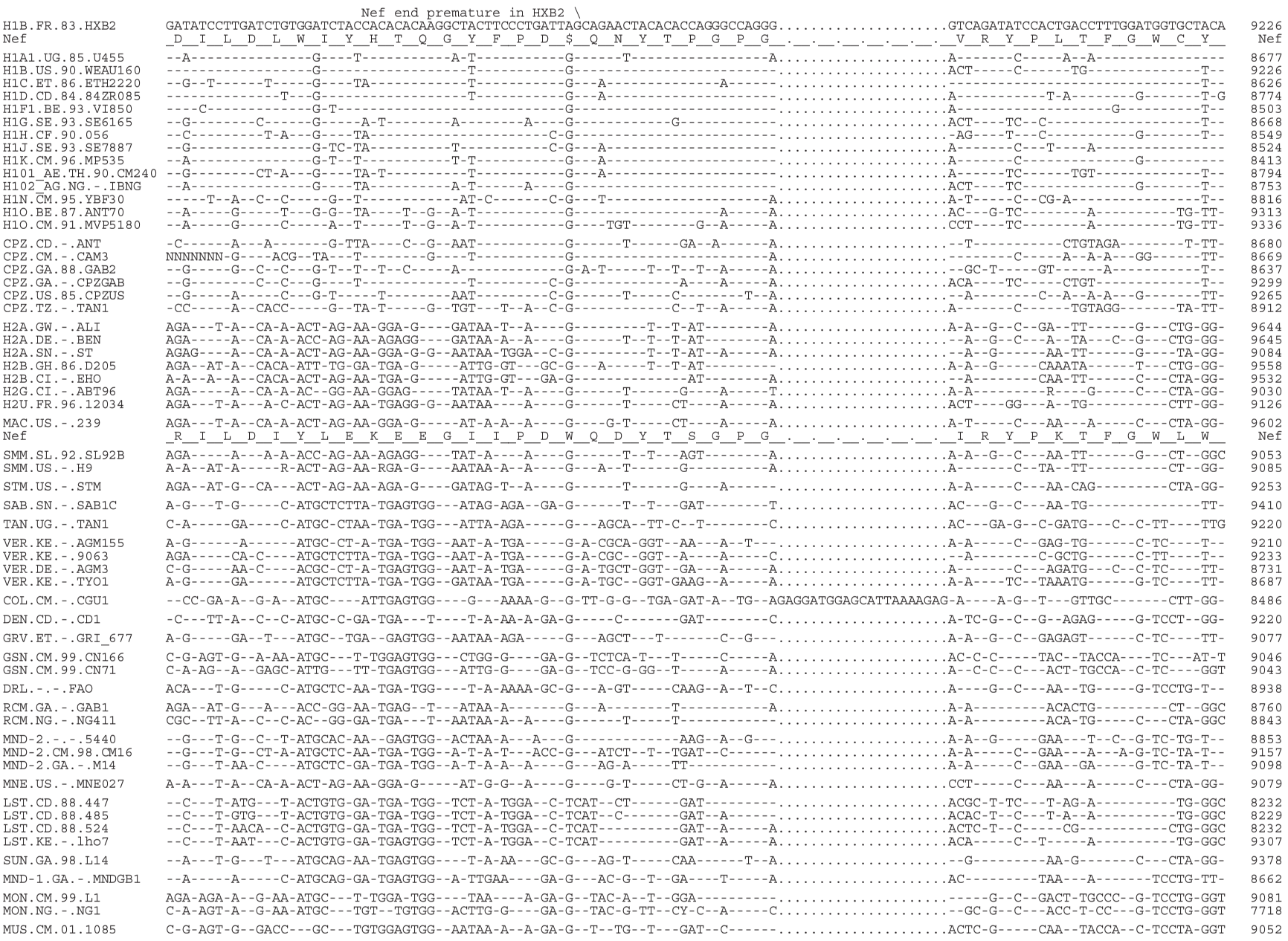


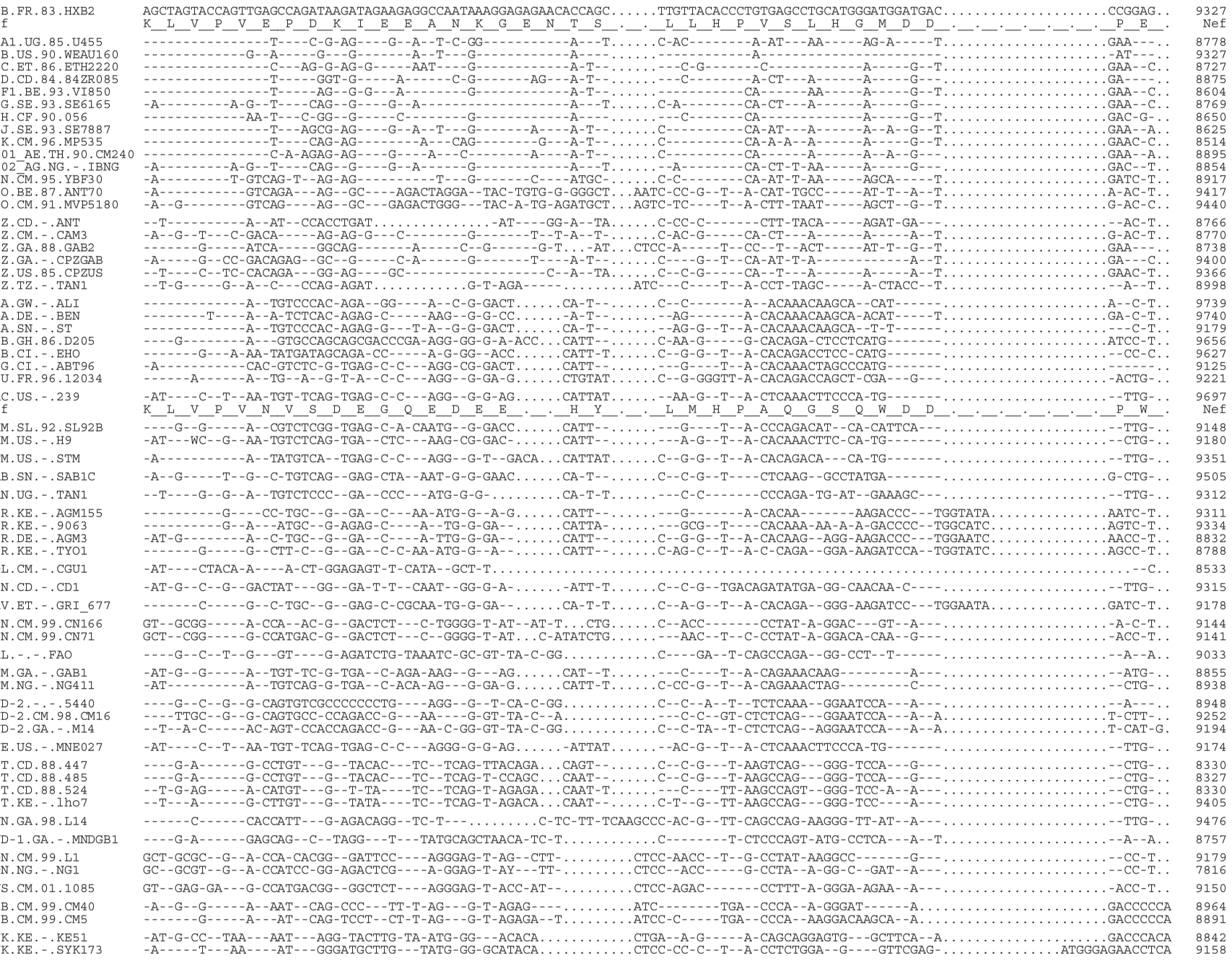


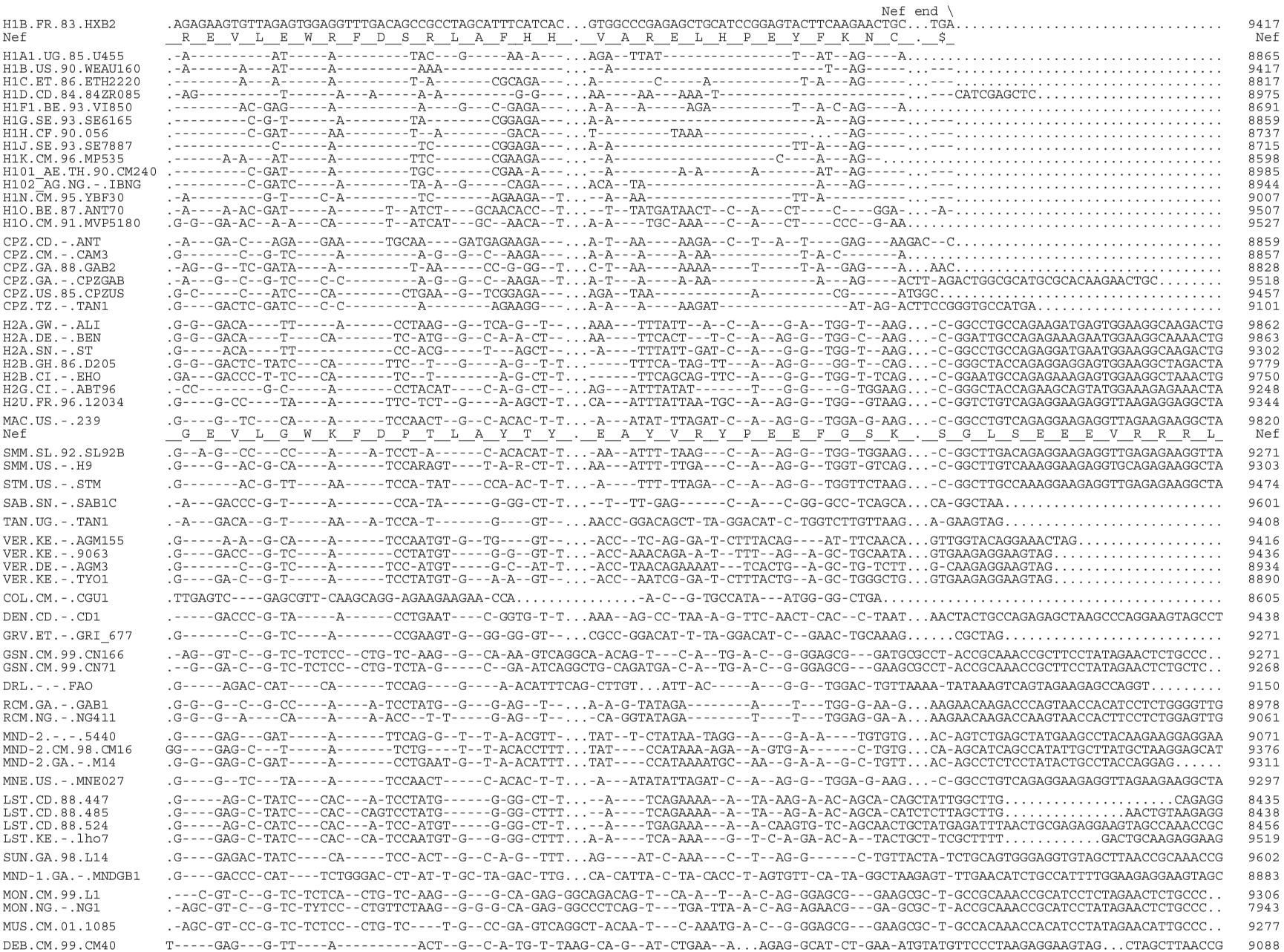
PLV
complete genomes


H1B.FR.83.HXB2
Nef
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
Premature stop
MAC.US.-.239
Nef
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173


polypurine tract
3'LTR U3
H1B.FR.83.HXB2
Nef
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Nef
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes

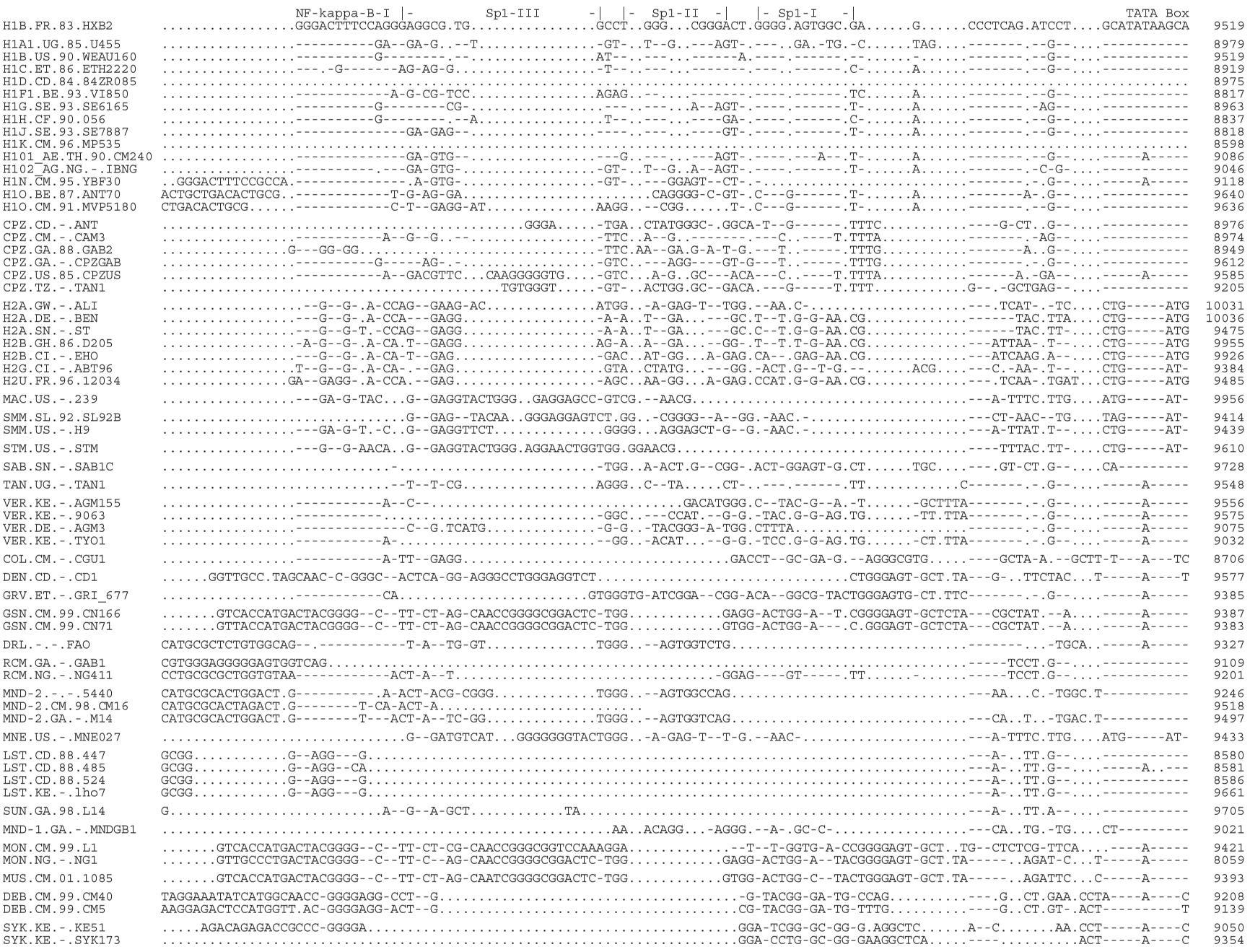

Nef end premature in HXB2 \
H1B.FR.83.HXB2 GATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGG.....GTCAGATATCCACTGACCTTTGGATGGTGCTACA 9226
Nef _D__I__L__D__L__W__I__Y__H__T__Q__G__Y__F__P__D__$__Q__N__Y__T__P__G__P__G__.__.__.__.__.__.__.__V__R__Y__P__L__T__F__G__W__C__Y__ Nef
H1A1.UG.85.U455 --A---------G----T----------A-T----------G------T----------A.....A-------C-----A--A---------- 8677
H1B.US.90.WEAU160 ----------G----------T----------G----------.....ACT-----C------TG----------T-- 9226
H1C.ET.86.ETH2220 --G--T------T-----G----TA----------T----------G---A----------A------.....----------T-- 8626
H1D.CD.84.84ZR085 ----------T---G----------T----------G---A----------.....A----------T-A--------G-------T-G 8774
H1F1.BE.93.VI850 ----C----------G-T----------G----------.....A----------G---------T-- 8503
H1G.SE.93.SE6165 --G--------C------G-----A-T--------A--------A----G----------G----------.....ACT----TC--C----------T-- 8668
H1H.CF.90.056 --C---------T-A---G----TA----------C-G----------.....-AG----T---C----------G-------T-- 8549
H1J.SE.93.SE7887 --G----------G-TC-TA----------T----------C-G---A----------.....A-------C---T----A---------- 8524
H1K.CM.96.MP535 --A----------G-T--T----------T-T----------G---A----------.....A-------C----------G---------- 8413
H101_AE.TH.90.CM240 --G--------CT-A---G----TA-T----------T----------G---A----------.....A------TC------TGT----------T-- 8794
H102_AG.NG.-.IBNG --A----------G----TA----------A-T------A----G----------.....ACT----TC----------G-------T-- 8753
H1N.CM.95.YBF30 -----T--A--C--C-----G--T----------AT-C-------C-G---T----------A.....A-T-----C--CG-A----------T-- 8816
H1O.BE.87.ANT70 --A-----G-----T---G-G--TA----T--G--A-T----------G----------A----A.....AC---G-TC--------A----------TG-TT- 9313
H1O.CM.91.MVP5180 --A-----G-----C-----A--T-----T--G--A-T----------G----TGT--------G--A----A.....CCT----TC--------A----------TG-TT- 9336
CPZ.CD.-.ANT -C------A---A-------G-TTA----C--G--AAT----------G------T------GA--A------A.....--T----------CTGTAGA--------T-TT- 8680
CPZ.CM.-.CAM3 NNNNNNN-G----ACG--TA---T--------G----T----------G----------A.....--------C-----A--A-A---GG------TT- 8669
CPZ.GA.88.GAB2 --G-----G--C--C---G-T--T--T--C-----A----------G-A-T-----T--T--T--A-----A.....--GC-T-----GT------A----------T-- 8637
CPZ.GA.-.CPZGAB --G--------C--C---G----T----------T---------C-G----------A----A----A.....ACA----TC-----CTGT----------T-- 9299
CPZ.US.85.CPZUS --G-----A-----C---G-T-----T--------AAT----------C-G------T-----C-------T--A.....--A--------C--A--A-A---G------TT- 9265
CPZ.TZ.-.TAN1 -CC-----A--CACC-----G--TA-T-----G--TGT---T--A--C-G----------C--T--A----A.....A-------C------TGTAGG-------TA-TT- 8912
H2A.GW.-.ALI AGA---T-A--CA-A-ACT-AG-AA-GGA-G----GATAA-T--A----G---------T--T-AT--------A.....A-A--G--C--GA--TT------G---CTG-GG- 9644
H2A.DE.-.BEN AGA-----A--CA-A-ACC-AG-AA-AGAGG----GATAA-A--A----G---------T--T--T-AT--------A.....--A--G--C---A--TA---C--G---CTG-GG- 9645
H2A.SN.-.ST AGAG----A--CA-A-ACT-AG-AA-GGA-G-G--AATAA-TGGA--C-G---------T--T-AT--A----A.....--A--G------AA-TT------G----TA-GG- 9084
H2B.GH.86.D205 AGA--AT-A--CACA-ATT-TG-GA-TGA-G-----ATTG-GT---GC-G---A-----T--T-AT---------.....A-A--G-----CAAATA------T---CTG-GG- 9558
H2B.CI.-.EHO A-A--A--A--CACA-ACT-AG-AA-TGA-G-----ATTG-GT---GA-G----------AT---------.....--A---------CAA-TT------C---CTA-GG- 9532
H2G.CI.-.ABT96 AGA-----A--CA-A-AC--GG-AA-GGAG-----TATAA-T--A----G------T------G----A-----T.....A-A---------R----G-----C---CTA-GG- 9030
H2U.FR.96.12034 AGA---T-A---A-C-ACT-AG-AA-TGAGG-G--AATAA----A----G------T-----CT---A-----A.....ACT----GG---A--TG---------CTT-GG- 9126
MAC.US.-.239 AGA---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-A-A--A----G----G-T-----CT---A-----A.....A-T-----C--AA---A-----C---CTA-GG- 9602
Nef _R__I__L__D__I__Y__L__E__K__E__E__G__I__I__P__D__W__Q__D__Y__T__S__G__P__G__.__.__.__.__.__.__.__I__R__Y__P__K__T__F__G__W__L__W__ Nef
SMM.SL.92.SL92B AGA-----A---A-A-ACC-AG-AA-AGAGG----TAT-A----A----G------T--T--AGT--------A.....A-A--G--C---AA-TT------G---CT--GGC 9053
SMM.US.-.H9 A-A--AT-A-----R-ACT-AG-AA-RGA-G----AATAA-A--A----G---A--T------G---------A.....A-------C--TA--TT----------CT--GG- 9085
STM.US.-.STM AGA--AT-G--CA---ACT-AG-AA-AGA-G----GATAG-T--A----G------T-----G----A------.....A-A-----C---AA-CAG---------CTA-GG- 9253
SAB.SN.-.SAB1C A-G---T-G-----C-ATGCTCTTA-TGAGTGG---ATAG-AGA--GA-G------T--T---GAT--------T.....AC---G--C---AA-TG----------TT- 9410
TAN.UG.-.TAN1 C-A-----GA----C-ATGC-CTAA-TGA-TGG---ATTA-AGA-----G---AGCA--TT-C--T--------C.....AC---GA-G--C-GATG---C--C-TT----TTG 9220
VER.KE.-.AGM155 A-G------A------ATGC-CT-A-TGA-TGG--AAT-A-TGA-----G-A-CGCA-GGT--AA---A--T---.....A-A-----C--GAG-TG------C-TC----T-- 9210
VER.KE.-.9063 AGA-----CA-C----ATGCTCTTA-TGA-TGG--AAT-A-TGA-----G-A-CGC--GGT--A----A-----C.....--A---------C-GCTG------C-TT----T-- 9233
VER.DE.-.AGM3 C-G-----AA-C----ACGC-CT-A-TGAGTGG--AAT-A-TGA-----G-A-TGCT-GGT--GA---A-----A.....A-------C---AGATG---C--C-TC----TT- 8731
VER.KE.-.TYO1 A-G-----GA------ATGCTCTTA-TGA-TGG--GATAA-TGA-----G-A-TGC--GGT-GAAG--A-----A.....A-A----TC--TAAATG------G-TC----TT- 8687
COL.CM.-.CGU1 --CC-GA-A--G-A--ATGC----ATTGAGTGG----G---AAAA-G--G-TT-G-G--TGA-GAT-A--TG--AGAGGATGGAGCATTAAAAGAG-A----A-G--T---GTTGC-------CTT-GG- 8486
DEN.CD.-.CD1 -C---TT-A--C--C-ATGC-C-GA-TGA---T----T-A-AA-A-GA-G-----C-------GAT--------C.....A-TC-G--C--G--AGAG-----G-TCCT--GG- 9220
GRV.ET.-.GRI_677 A-G-----GA--T---ATGC--TGA--GAGTGG--AATAA-AGA-----G---AGCT---T-------C--G---.....A-A--G--C--GAGAGT------C-TC----TT- 9077
GSN.CM.99.CN166 C-G-AGT-G--A-AA-ATGC---T-TGGAGTGG---CTGG-G----GA-G--TCTCA-T---T-----C-----A.....AC-C-C-----TAC--TACCA----TC---AT-T 9046
GSN.CM.99.CN71 C-A-AG--A--GAGC-ATTG---TT-TGAGTGG---ATTG-G----GA-G--TCC-G-GG--T------A-----A.....A--C-C--C---ACT-TGCCA--C-TC----GGT 9043
DRL.-.-.FAO ACA---T-G-----C-ATGCTC-AA-TGA-TGG----T-A-AAAA-GC-G---A-GT-----CAAG--A--T--C.....A----G--C---AA--TG-----G-TCCTG-T-- 8938
RCM.GA.-.GAB1 AGA--AT-G---A---ACC-GG-AA-TGAG--T---ATAA-A-------G---A--------T----------A.....A-A---------ACACTG---------CT--GGC 8760
RCM.NG.-.NG411 CGC--TT-A--C--C-AC--GG-GA-TGA---T--AATAA-A--A----G------T-----T----------.....A-A---------ACA-TG-----C---CTA-GGC 8843
MND-2.-.-.5440 --G---T-G--C--T-ATGCAC-AA--GAGTGG--ACTAA-A---A---G----------AAG--A--G---.....A-A--G-----GAA---T--C--G-TC-TG-T-- 8853
MND-2.CM.98.CM16 --G---T-G--CT-A-ATGCTC-AA-TGA-TGG--A-T-A-T---ACC-G---ATCT--T--TGAT--C------.....A-A-----C--GAA---A---A-G-TC-TA-T-- 9157
MND-2.GA.-.M14 --G---T-AA-C----ATGCTC-GA-TGA-TGG--A-T-A-A--A----G---AG-A-----TT----------.....A-A-----C--GAA--GA-----G-TC-TA-T-- 9098
MNE.US.-.MNE027 A-A---T-A--CA-A-ACT-AG-AA-GGA-G-----AT-G-G--A----G----G-T-----CT-G--A-----A.....CCT-----C---AA---A-----C---CTA-GG- 9079
LST.CD.88.447 --C---T-ATG---T-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT--CT------GAT---------.....ACGC-T-TC---T-AG-A----------TG-GGC 8232
LST.CD.88.485 --C---T-GTG---T-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT--C-------GAT--A-----.....ACAC-T--C---T-A--A----------TG-GGC 8229
LST.CD.88.524 --C---T-AACA--C-ACTGTG-GA-TGA-TGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A.....ACTC-T--C------CG----------CTG-GGC 8232
LST.KE.-.lho7 --C---T-AAT---C-ACTGTG-GA-TGAGTGG--TCT-A-TGGA--C-TCAT----------GAT--A-----A.....ACA-----C--T-----A----------TG-GGC 9307
SUN.GA.98.L14 --A---T-G---T---ATGCAG-AA-TGAGTGG----T-A-AA---GC-G---AG-T-----CAA------T--A.....--G---------AA-G-------C---CTA-GG- 9378
MND-1.GA.-.MNDGB1 --A-----A-----C-ATGCAG-GA-TGAGTGG--A-TTGAA----GA-G---AC-G--T--GA----T-----A.....AC----------TAA---A-------TCCTG-TT- 8662
MON.CM.99.L1 AGA-AGA-A--G-AA-ATGC---T-TGGA-TGG----TAA----A-GA-G--TAC-A--T--GGA----------.....-----G--C--GACT-TGCCC--G-TCCTG-GGT 9081
MON.NG.-.NG1 C-A-AGT-A--G-AA-ATGC---TGT--TGTGG--ACTTG-G----GA-G--TAC-G-TT--CY-C--A-----C.....--GC-G--C---ACC-T-CC---G-TCCTG-GGT 7718
MUS.CM.01.1085 C-G-AGT-G--GACC---GC---TGTGGAGTGG--AATAA-A--A-GA-G--T--TG--T---GAT--C------.....ACTC-G-----CAA--TACCA--C-TCCTA-GGT 9052
DEB.CM.99.CM40 -C---TT-A--G--TCATGC-C-GA-TGA-TGG--AATTA-A--A-GC-G--TAC-G------AAG--A--T--A.....--T--G------A-TTA---C--G-TCCT--TT- 8867
DEB.CM.99.CM5 -C---AT-A--AA-CCATGC-C----TGA-TGG--AATAA-A--A-G--G-TT---G------GA---C-----A.....CCA--G------ACATA------G-TCCTA-TT- 8794
SYK.KE.-.KE51 --G--T--AC-G----ATT-AA-GA-TGAG--T--TGTGA-A...---G-CATC-C---T--GAGT--------A.....ACA-----C--G----T---C--G----TA-GGG 8745
SYK.KE.-.SYK173 --A--T---C----C-ACC-AC-GA-TGAG--T--AAT-A-A...--CAG-ATC--T-----CT---------A.....AC-C-G--C--GT-A-T----------TA-GGG 9055

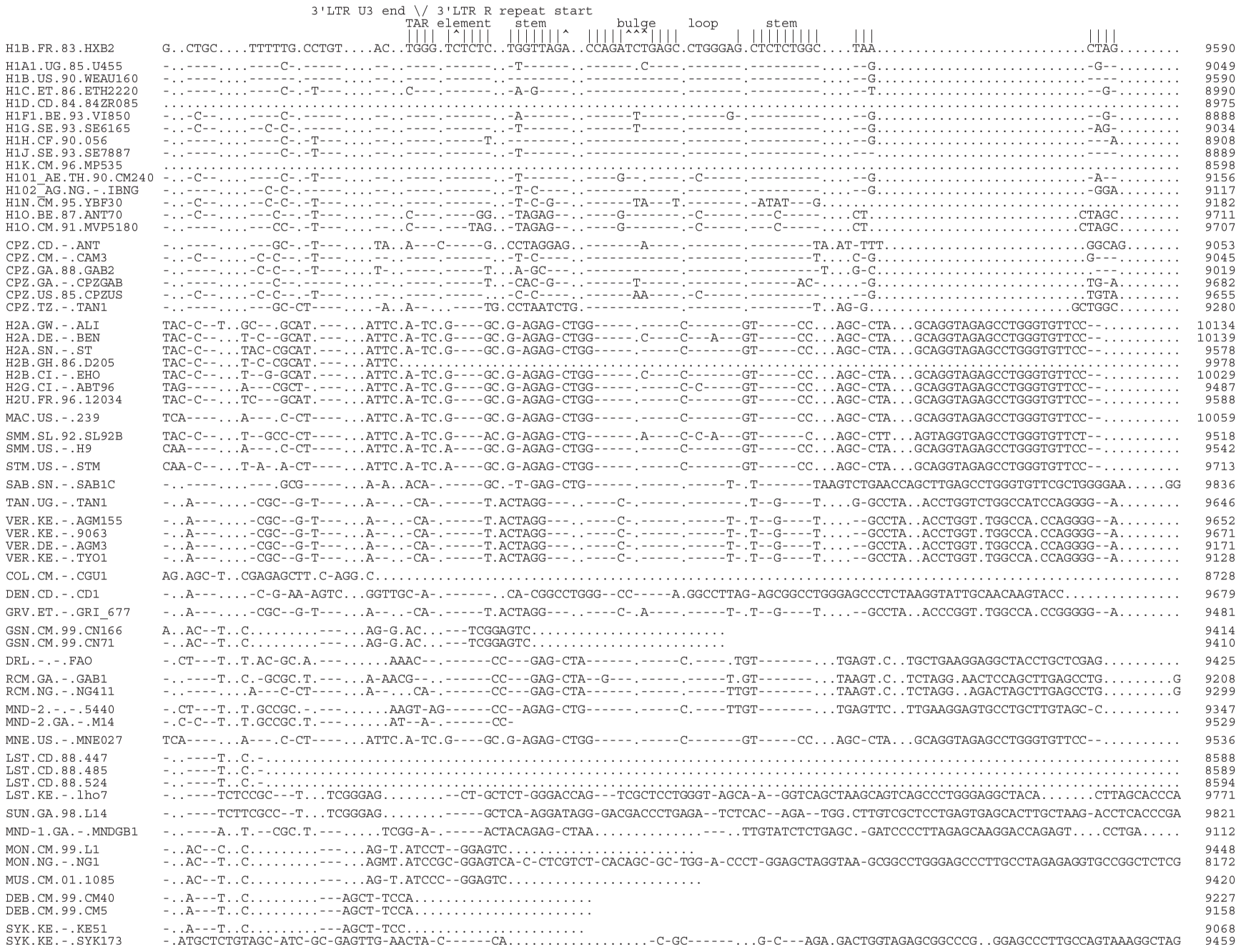

H1B.FR.83.HXB2
Nef
H1A1.UG.85.U455
H1B.US.90.WEAU160
H1C.ET.86.ETH2220
H1D.CD.84.84ZR085
H1F1.BE.93.VI850
H1G.SE.93.SE6165
H1H.CF.90.056
H1J.SE.93.SE7887
H1K.CM.96.MP535
H101_AE.TH.90.CM240
H102_AG.NG.-.IBNG
H1N.CM.95.YBF30
H1O.BE.87.ANT70
H1O.CM.91.MVP5180
CPZ.CD.-.ANT
CPZ.CM.-.CAM3
CPZ.GA.88.GAB2
CPZ.GA.-.CPZGAB
CPZ.US.85.CPZUS
CPZ.TZ.-.TAN1
H2A.GW.-.ALI
H2A.DE.-.BEN
H2A.SN.-.ST
H2B.GH.86.D205
H2B.CI.-.EHO
H2G.CI.-.ABT96
H2U.FR.96.12034
MAC.US.-.239
Nef
SMM.SL.92.SL92B
SMM.US.-.H9
STM.US.-.STM
SAB.SN.-.SAB1C
TAN.UG.-.TAN1
VER.KE.-.AGM155
VER.KE.-.9063
VER.DE.-.AGM3
VER.KE.-.TYO1
COL.CM.-.CGU1
DEN.CD.-.CD1
GRV.ET.-.GRI_677
GSN.CM.99.CN166
GSN.CM.99.CN71
DRL.-.-.FAO
RCM.GA.-.GAB1
RCM.NG.-.NG411
MND-2.-.-.5440
MND-2.CM.98.CM16
MND-2.GA.-.M14
MNE.US.-.MNE027
LST.CD.88.447
LST.CD.88.485
LST.CD.88.524
LST.KE.-.lho7
SUN.GA.98.L14
MND-1.GA.-.MNDGB1
MON.CM.99.L1
MON.NG.-.NG1
MUS.CM.01.1085
DEB.CM.99.CM40
DEB.CM.99.CM5
SYK.KE.-.KE51
SYK.KE.-.SYK173
PLV
complete genomes


Nef end \
H1B.FR.83.HXB2 9417
Nef Nef
H1A1.UG.85.U455 8865
H1B.US.90.WEAU160 9417
H1C.ET.86.ETH2220 8817
H1D.CD.84.84ZR085 8975
H1F1.BE.93.VI850 8691
H1G.SE.93.SE6165 8859
H1H.CF.90.056 8737
H1J.SE.93.SE7887 8715
H1K.CM.96.MP535 8598
H101_AE.TH.90.CM240 8985
H102_AG.NG.-.IBNG 8944
H1N.CM.95.YBF30 9007
H1O.BE.87.ANT70 9507
H1O.CM.91.MVP5180 9527
CPZ.CD.-.ANT 8859
CPZ.CM.-.CAM3 8857
CPZ.GA.88.GAB2 8828
CPZ.GA.-.CPZGAB 9518
CPZ.US.85.CPZUS 9457
CPZ.TZ.-.TAN1 9101
H2A.GW.-.ALI 9862
H2A.DE.-.BEN 9863
H2A.SN.-.ST 9302
H2B.GH.86.D205 9779
H2B.CI.-.EHO 9750
H2G.CI.-.ABT96 9248
H2U.FR.96.12034 9344
MAC.US.-.239 9820
Nef Nef
SMM.SL.92.SL92B 9271
SMM.US.-.H9 9303
STM.US.-.STM 9474
SAB.SN.-.SAB1C 9601
TAN.UG.-.TAN1 9408
VER.KE.-.AGM155 9416
VER.KE.-.9063 9436
VER.DE.-.AGM3 8934
VER.KE.-.TYO1 8890
COL.CM.-.CGU1 8605
DEN.CD.-.CD1 9438
GRV.ET.-.GRI_677 9271
GSN.CM.99.CN166 9271
GSN.CM.99.CN71 9268
DRL.-.-.FAO 9150
RCM.GA.-.GAB1 8978
RCM.NG.-.NG411 9061
MND-2.-.-.5440 9071
MND-2.CM.98.CM16 9376
MND-2.GA.-.M14 9311
MNE.US.-.MNE027 9297
LST.CD.88.447 8435
LST.CD.88.485 8438
LST.CD.88.524 8456
LST.KE.-.lho7 9519
SUN.GA.98.L14 9602
MND-1.GA.-.MNDGB1 8883
MON.CM.99.L1 9306
MON.NG.-.NG1 7943
MUS.CM.01.1085 9277
DEB.CM.99.CM40 9088
DEB.CM.99.CM5 9018
SYK.KE.-.KE51 8954
SYK.KE.-.SYK173 9282

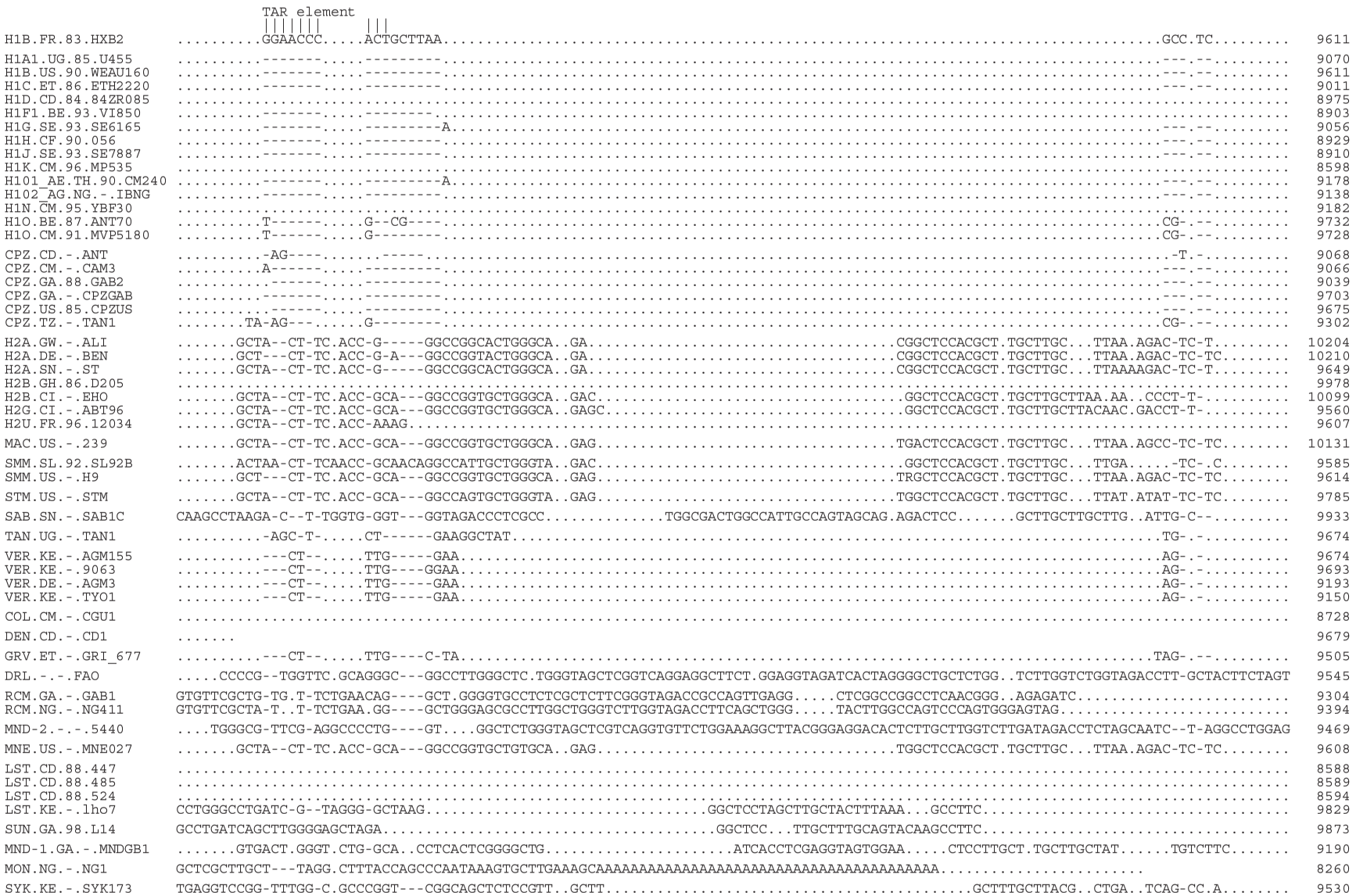
```
                                                                                                                                             NF-kappa-B-II
H1B.FR.83.HXB2        ..................................................................................CATCGAGCTTGCTACAA..........GGGACTTTCCG.CTG..   9448

H1A1.UG.85.U455       .....................................................................AAG.ACTG.CTGA--CA--AG-----GAC-..........----------A.---..   8907
H1B.US.90.WEAU160     ....................................................................................-------T--T-----.........----------A.---..   9448
H1C.ET.86.ETH2220     ...............................................................................CACAGA--G-AC-TTCCGCC..........----------A.---..   8851
H1D.CD.84.84ZR085     ................................................................................................................................   8975
H1F1.BE.93.VI850      ......................................................GACTGCTGACACAGAGATT.GCTG.ACACAGAA--ATC-..A-A-..........----------A.---..   8746
H1G.SE.93.SE6165      ...............................................................................CACAGAA-TTGC--AC-AG..........----------AC---..   8892
H1H.CF.90.056         ...............................................................................CACAGAA--T.C-..G-A-..........-----------.---..   8766
H1J.SE.93.SE7887      ...............................................................................CAAAGAA-TTTC-..AG-G..........-----------.---..   8746
H1K.CM.96.MP535       ................................................................................................................................   8598
H101_AE.TH.90.CM240   ...............................................................................CAAGGAA-TTTC-.ACT............A-A----.---.---..   9014
H102_AG.NG.-.IBNG     ...............................................................................CACAGAA-TTGC--AC-..........----------A.---..   8975
H1N.CM.95.YBF30       ............................................................................CAAGGGAC-TT-CTGCTGACA-..........--------ATACT--..   9044
H1O.BE.87.ANT70       ......................................................................AAACTGCTGACCTGAA--TTGCTGACACTGT.........--A-------AG-AAAG   9555
H1O.CM.91.MVP5180     ...........................................................................................CTGACACTGC.........----------AGAC.TG   9553

CPZ.CD.-.ANT          ..................................................................TAACAGGCGGATATAAAAC-GCT--CAGA--AGACTTGTGGTCT...----------ACTAT..   8918
CPZ.CM.-.CAM3         ..................................................................TGCCCAGCG.CCTG.CGCAGTAGA-A--GCTG----..........---------TAA.....   8903
CPZ.GA.88.GAB2        .............................................................CGCAACTGGCGCATGCGCTCTAGAA-TG-TGAGCCTGA--TTTTACAA.....----------AA.....   8888
CPZ.GA.-.CPZGAB       ..................................................................TGACTCTG..................-........................----------AA.....   9539
CPZ.US.85.CPZUS       ..................................................................TGTAACCGC.GCAGGCGCAATA.A-A--GCTG--TG..........A---------TAA.....   9503
CPZ.TZ.-.TAN1         ......................................CTCAGAACTGCTGACAG.AGGAC.TTTT......GGACTC...........................................----------AA.....   9145

H2A.GW.-.ALI          AAAGCAAGAGGGATACCATTTAGTAAGAACAGGAACAGCTGATTTGGTCAGGGCAGGAAGTAACTACTGAA..........AACAGC-GAG.AC---A..............----------AGAA-..   9964
H2A.DE.-.BEN          AAAGCAAGAGGGATACCATATAGTGAATAACAGGAACAACCATACTTGGTCAAGGCAGGAAGTAGCTA............CTAAG-AAC-.GC--AGG-TGCA.........----------AGAA...   9967
H2A.SN.-.ST           AAAGCAAGAGGGATACCGTTTAGCTAAAAACAGGAACAGCTATACTTGGTCAGGGCAGGAAGTAACTA............ACAGA-AAC-.GC--AG--TGCA.........----------AGAA...   9406
H2B.GH.86.D205        AAAGCAAGAGGGATACCTACAGATTAGGCAAGAGACAGCAGCATAAACAGGAACTAGCTGACACTGCACAA..........GAAGG-AACTAGCA-AC--TGCA.........----------AAAA...   9887
H2B.CI.-.EHO          AGAGCAAGAGGAATACCTACAGAGTAGACAGGAAATAACAGCACAAAGAGGAACTAGCTAACACTGCATAGA.........GAAGG-AACTAGC--A---TGCA.........----------AGAA...   9858
H2G.CI.-.ABT96        AAACAGAGAGGACTGCCTATAGAATAAGC...................................................AGAAAGA.GG..AC-AGC--AC---GCA.........----------AGTA...   9316
H2U.FR.96.12034       ACCGCAAGAGGCCTCCCAGTAAAGAATTGCTGA...............................................CAGAAAGG-AAC-AGC--AG---GCA.........----------AC-A...   9417

                                                                                                      Nef end \
MAC.US.-.239          ACCGCAAGAGGCCTTCTTAACATGGCT.....................................................GACAAGAAGG-AACTCGC--AA---GCA.........----------ACAA...   9889
Nef                   _T__A__R__G__L__L__N__M__A__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__.__D__K__K__E__T__R__$_                                  Nef

SMM.SL.92.SL92B       GCTAACAAGCCAAAACCGCAGAAGAAGATGGCG...............................................GATAAGAAGG-AAC-AGC--AG---GCA.........----------ACAA...   9346
SMM.US.-.H9           ACCGCAAGAGGCCTTTTAAAAATGGCTGACAAGAAGAAAACAAGCTAG...................................-ACA--A..............----------ACAAA..   9373

STM.US.-.STM          ACCGCAAGAGGCCTTCTAAAGATGGCT....................................................GACAAGAAGGAAA--AGCT-AG..AC-GC-..........----------ACAA...   9543

SAB.SN.-.SAB1C        ..............................................CCACAAAACCACATCCTATGGAGTTGTC.ATGGTGAT--CA--AAG-ACTGCTGACTGAGA----------AG.....   9673

TAN.UG.-.TAN1         ......................................CCTAACCGCAGGCTTGCGGTTAAGCTAACTAGGCCGTTGCC.TAGGAGAT--CA-AAAG-AGTGCTGATCAGCA----------AG.....   9493

VER.KE.-.AGM155       ...............................................CCGACCACAGGCTTGCGGTTTCCTGGTTGCC.TAGGAGAT--CA--AAG-ACTGCTGAC.....----------AG-AC..   9489
VER.KE.-.9063         ...............................................CTAACCGCAGGCTTGTGGTTAAGCCGTTGCC.AGGGAGAT--CA--TGA-ACTGCTGACAG.----------AA.....   9508
VER.DE.-.AGM3         ...............................................CTAACCGCAGGCTTGTGGTTAAGCACATCACCATGGTGAT--CA--AAG-ACTGCTGAC.....----------AG-AA..   9008
VER.KE.-.TYO1         ...............................................CCAGACCGCAAGCCTGCGGTTAGAACAT.CACC.ATGGAGAT--CA--AAA-ACTGCTGAC.....----------AG-GAA.   8965

COL.CM.-.CGU1         .................................................................................G-T-CTGA....GAC-GCA........-----------TTC...   8634

DEN.CD.-.CD1          AACCGCAAACCACATCCTACTGCAGAACTTGCTGAGAGGCTCA.........................................................................----------A......   9492

GRV.ET.-.GRI_677      ...........................................................................................T-TCA--ACTTTGCGGTT.....-----------C-A...   9303

GSN.CM.99.CN166       ....................................................................................................................----------A......   9282
GSN.CM.99.CN71        ....................................................................................................................----------A......   9279

DRL.-.-.FAO           .........TGTTGCAAGAGGAAGTGGTGGCAGTTCAAGCCAACCACAGAAGGTTGTCATGGAGACAACCTTCAGAAGTGCTGACTGAGA-.....................----------ACAG-AG   9249

RCM.GA.-.GAB1         CCTTGGTAACCAGGCAGAAGAATCTGCTGATGCAAAAGGGACTTTCCACTGGTGCATGCGCACTGGGGAA.............................................-----------GGATGA   9065
RCM.NG.-.NG411        CCTTGGAAACAGCGCATGCCTACTGAGAAGATGCTGACACAGAA.......................................................................----------ACTG-CG   9122

MND-2.-.-.5440        AAGCCTGACTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCTAACCACAAAGGGTTGCTATAAGCAACTGCTAACTCTGGA..................................----------ACTG-TG   9168
MND-2.CM.98.CM16      AAGCCAGAGTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCCGACCACAGAAGGTTGCCATGGAGATGACCTGCAGAAATGCTGACTC-G--......................----------ACTA-TG   9484
MND-2.GA.-.M14        AAATCAGACTGCTGCAAGAGGAAGTGGTGGCAGTTCTAGCTAACCACAGAAGGTTGCTATGGAGATGCTCTTT.........AATTCA-CTGA....CTCTGGA.......----------ACTA-TG   9419

MNE.US.-.MNE027       ACCGCAAGAGGCCTCTTAAAAATGGCT....................................................GACAAGAAGGAAA-TAGCT-AG..AC-GC-..........----------ATAA...   9366

LST.CD.88.447         AAGTAGCCAAACCGCAGACCGCAACACATCCTCTTGCAGCCTGGTTGCCAGGGCAACCAGGCTAGCGCATGCGC.......AGGA-T-T-CTGA.....TG-GCA........----------AGGACGG   8545
LST.CD.88.485         AAGTAGCCAAACCGCAGACCGCAGCACATCCTCTTGCAGCCTGGTTGCCAGGGCAACCAGGCTAGCGCATGCGC.......AGGA-T-T-CTGA.....T--GCA........----------AGGACGG   8548
LST.CD.88.524         AAACCGCAGAACATCCTCTTGCAGCCCGGTTGCCCTAGCAACCGGGCTAGCGCATGCGC.......................ACAAGC-T-CTGA.....AC-GCA........----------AGGACGG   8551
LST.KE.-.lho7         TAGCCAAACCGCAAACCGCAGCACATCCTCTTGCAGCCCGGTTGCTAAGGCAAC...CGGGCTAGCGCATGCGC.......AATGGC-T-CTGA.....-G-GCA........----------AGGACGG   9626

SUN.GA.98.L14         CATCCTCTTGCATCGCCTAGGCAACG.........................................................GGGC-A-C--A---GCGCTAGAGTA.....-----------GACAGG   9670

MND-1.GA.-.MNDGB1     CTAACCGCAAAACCACATCCTACTGCAGAACTGTAG.TTGCTTGGCAACCTGCTTAGCAACCTGGACTGGCGCTTGCGCG.CTAGG-A.........................----------.......   8979

MON.CM.99.L1          ....................................................................................................................----------A......   9317
MON.NG.-.NG1          ....................................................................................................................----------A......   7954

MUS.CM.01.1085        ....................................................................................................................----------A......   9288

DEB.CM.99.CM40        CAAACCACATCCTACAGCAG.CGCCC..........................................................................................G---------AGT--CC   9131
DEB.CM.99.CM5         CAAACCGCATCCTCTTGCAGAGAGCC..........................................................................................C--T-------AGTA-TC   9062

SYK.KE.-.KE51         ACCACAGGGTACATCTCCTAGGAGACACCTAGG...................................................................................................   8987
SYK.KE.-.SYK173       ACTCCATGGTGACAAGCTCGGCCCAGGGGA......................................................................................................   9312
```

PLV
complete genomes





PLV
complete genomes


TAR element
H1B.FR.83.HXB2 9611
H1A1.UG.85.U455 9070
H1B.US.90.WEAU160 9611
H1C.ET.86.ETH2220 9011
H1D.CD.84.84ZR085 8975
H1F1.BE.93.VI850 8903
H1G.SE.93.SE6165 9056
H1H.CF.90.056 8929
H1J.SE.93.SE7887 8910
H1K.CM.96.MP535 8598
H101_AE.TH.90.CM240 9178
H102_AG.NG.-.IBNG 9138
H1N.CM.95.YBF30 9182
H1O.BE.87.ANT70 9732
H1O.CM.91.MVP5180 9728
CPZ.CD.-.ANT 9068
CPZ.CM.-.CAM3 9066
CPZ.GA.88.GAB2 9039
CPZ.GA.-.CPZGAB 9703
CPZ.US.85.CPZUS 9675
CPZ.TZ.-.TAN1 9302
H2A.GW.-.ALI 10204
H2A.DE.-.BEN 10210
H2A.SN.-.ST 9649
H2B.GH.86.D205 9978
H2B.CI.-.EHO 10099
H2G.CI.-.ABT96 9560
H2U.FR.96.12034 9607
MAC.US.-.239 10131
SMM.SL.92.SL92B 9585
SMM.US.-.H9 9614
STM.US.-.STM 9785
SAB.SN.-.SAB1C 9933
TAN.UG.-.TAN1 9674
VER.KE.-.AGM155 9674
VER.KE.-.9063 9693
VER.DE.-.AGM3 9193
VER.KE.-.TYO1 9150
COL.CM.-.CGU1 8728
DEN.CD.-.CD1 9679
GRV.ET.-.GRI_677 9505
DRL.-.-.FAO 9545
RCM.GA.-.GAB1 9304
RCM.NG.-.NG411 9394
MND-2.-.-.5440 9469
MNE.US.-.MNE027 9608
LST.CD.88.447 8588
LST.CD.88.485 8589
LST.CD.88.524 8594
LST.KE.-.lho7 9829
SUN.GA.98.L14 9873
MND-1.GA.-.MNDGB1 9190
MON.NG.-.NG1 8260
SYK.KE.-.SYK173 9530

```
                                                  TAR elements of HIV-2 and SIV have extra stem-loop, see Berkhout, Nucleic Acids Res 20(1):27-31(1992)
                                                                                Poly-A signal     3'LTR R repeat end \/ 3'LTR U5 start
H1B.FR.83.HXB2          ..................................................AATAAAG...CTTGCC.......TTGAG..TG.CTTCAA.GTAGTG...TGTGCCCGTC.TG.TT..GTG.TGACTCT...G   9668

H1A1.UG.85.U455         ..................................................-------...------.......------..GTG---G--.-------...----------.--.--..A-T.----------   9128
H1B.US.90.WEAU160       ..................................................-------...------.......------..--.------.------...----------.--.--..---.----------   9668
H1C.ET.86.ETH2220       ..................................................-------...------.......------..--.....................................................   9031
H1D.CD.84.84ZR085       ..................................................................................................................................   8975
H1F1.BE.93.VI850        ..................................................................................................................................   8903
H1G.SE.93.SE6165        ..................................................-------...------.....AA-CC.....................................................   9074
H1H.CF.90.056           ..................................................-------...------.......-----..---.---T.....................................................   8953
H1J.SE.93.SE7887        ..................................................-------...------.......------..G-CGCATGC.AAGCC-.....................................   8943
H1K.CM.96.MP535         ..................................................................................................................................   8598
H101_AE.TH.90.CM240     ..................................................-------...------.......------..--.---A-.....................................   9203
H102_AG.NG.-.IBNG       ..................................................-------...------.......-----..---.------.------...----------.--A--.---.----------   9196
H1N.CM.95.YBF30         ..................................................................................................................................   9182
H1O.BE.87.ANT70         ..................................................-------...------.......------..--.AG.....................................   9754
H1O.CM.91.MVP5180       ..................................................-------...------.......-----..---.AGAAGC.AGT---...--CT-A..--.--.--CAACC.CTGG-G-...C   9785

CPZ.CD.-.ANT            ..................................................................................................................................   9068
CPZ.CM.-.CAM3           ..................................................-------...------.......------..--A---G.-.-C----...------TA---CA.--C..A-.-A--------...-   9123
CPZ.GA.88.GAB2          ..................................................-------...------.......------..--CTC-A.-.------...------T-A--.--.--C..A-.-A--------...-   9096
CPZ.GA.-.CPZGAB         ..................................................-------...------.......------.TGTAT--G.-.-C----...------TA---.A.--C..A-.AC--------...-   9761
CPZ.US.85.CPZUS         ..................................................-------...------.......------..--CT--A.-.-------...------TA---C.--C..A-.----------...-   9731
CPZ.TZ.-.TAN1           ..................................................-------...-C----.......-----..A-.TG-T.-.                                              9326

H2A.GW.-.ALI            ..................................................-------...--.---AA.....--AGA..A-CAGGTT-.AAG---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...C  10269
H2A.DE.-.BEN            ..................................................-------...--.---AG.....--AGA..A-CAAGTT-.AGT---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...C  10275
H2A.SN.-.ST             ..................................................-------...--.---AG.....--AGA..A-CA.....................................................   9672
H2B.GH.86.D205          ..................................................................................................................................   9978
H2B.CI.-.EHO            ..................................................------A...--.---TA.....--AGA..A-CAAGAC-.AGT---TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...C  10164
H2G.CI.-.ABT96          ..................................................-------C..T.GC-AAA.....--AGA..A-TAAG.C-.AGT---TGTGT.....................................   9599
H2U.FR.96.12034         ........................................                                                                                           9607

MAC.US.-.239            ..................................................-------...--.---AT..T..--AGA..A-TAAG-T-..GT---TGT-CCCAT-TCT.CCTAGCC-CC.GCCTGG-...C  10196

SMM.SL.92.SL92B         ..................................................-------G..--.--AAC.....--AGA..A-CAAGTC..                                          9613
SMM.US.-.H9             ..................................................-------G..--.---AG.....--AGA..A-TA.....................................................   9638

STM.US.-.STM            ..................................................-------...--.---AA..T..--AGA..A-TAAG-.-A.GT-G-TGT-CCCAT-TCT.CCTAGTC-CC.GCCTGG-...C   9850

SAB.SN.-.SAB1C          ..................................................-------T..------.......--.--TT-A.-AAGT-AAGCA..AGTGTGTG--CATT-AT-CCTCA-GAC-AC-CTGGT  10002

TAN.UG.-.TAN1           ..................................................--------...-----TG..C..--A--..A-T--ATCT.-.--CAAGTGCCT-ATTG-.GCGCCCTTCC.-C-...G...-   9736

VER.KE.-.AGM155         ..................................................------....------TG..CA.--AGA..GCTTA--TG.AGTCAAGTGCCCT-ATTAA.--CC-CAC-C.-TGAA-G...-   9741
VER.KE.-.9063           ..................................................------AC..T-----TG..CA.--AGA..GCTTA--TG.AGTCAAGTGCCCT-ATTGA.C-CC-CAC-C.AAG-AGG...-   9762
VER.DE.-.AGM3           ..................................................------.T..---CG-TG..C.A--AG-..CAGAGA-TT.--GAGAGACAT----T-GT.GCA-CCA-CA.CA-G-GA...-   9261
VER.KE.-.TYO1           ..................................................------....------TG..CA.--AG.....................................................   9170

COL.CM.-.CGU1           ..................................................................................................................................   8728

GRV.ET.-.GRI_677        ..................................................------CC..TGCT-GC.....--AGT..C-CTA-ATT...G-A-TCAA----T-A-T.GCT..GC-CC.GAG-CTCTAGA   9570

DRL.-.-.FAO             GGCCATAGAAGTAGGGGAGTAGGCACCGCTTGCTTATCCTAATTGCTC-------C..TACC-ATT..TAGA--GATGA-TAAGTGT.ACTC-TGTC-T-TGTTC-T.A-TGAGA..CTCTGG-TA...C   9664

RCM.GA.-.GAB1           ..............ACCGCTTGCTT.ATAGCCTTG....AA..GCTC-------C........ATGCCAG--AGTTTACTG-AAGCAAGT---TGCCTGTTTTAC-.-CTCAGCAGTTAACGA--CTG-   9403
RCM.NG.-.NG411          ..............GCT.CCGCTTGCTTTCTGG...TTAATTGCTC------C...TGCATGCTCTTAG----A...CCAGAATCAAGT---TGCCCATTTTA---.-CTCAGT-AAAC-------.G-   9495

MND-2.-.-.5440          ATTGTG....GGATTGACTACCGCTT.GCTTGCT....ATTATTGCCC-------C..T-AC-TGA.TTAGAAT--CAA-TAAGAGT.ACTCGTCCA-T-TG.TC-TCAAT.AGAAACTCTGG-TA...C   9581

MNE.US.-.MNE027         ..................................................-------...--.---AT.....--AGA..A-TAAG-C-..GT---TGC-CCCAT-TCT.CCTAGTC-CC              9664

LST.CD.88.447           ..................................................................................................................................   8588
LST.CD.88.485           ..................................................................................................................................   8589
LST.CD.88.524           ..................................................................................................................................   8594
LST.KE.-.lho7           ..................................................------....AA-TTGCTCATTG-T--AAA-TACAAGCTAGT--CTCA-AGTAGTC--T-TC-CCTTC-CCC.........   9897

SUN.GA.98.L14           ..........................................................AATAAAGCCTGCCT----TTTACT.....................GCAAGCAA   9905

MND-1.GA.-.MNDGB1       ..................................................-------T..AACTTAGA..A..--AGA..GCA.....................................................   9215

SYK.KE.-.SYK173         ..................................................-------T..G-G-TAAA..CGCAA-TA..ACGTC-GCT.C-CA-CTGT-C---TG-GT.CACC-CCCGA.A-GTG......   9597
```

```
                                                                                                      3'LTR end \
H1B.FR.83.HXB2       GT..AA.CTAGAGATCC....................................CTCAGAC..CCTTTTAGTCAGTGTGGAAAATCTCTAGCA                                  9719

H1A1.UG.85.U455      --..--.----------....................................-------..-AC-A---A-T.GTGTA-------------                                  9178
H1B.US.90.WEAU160    --..--.----------....................................-------..-A--------------------------                                  9719
H102_AG.NG.-.IBNG    --..--.-                                                                                                                        9201
H1O.CM.91.MVP5180    TA..G-......----                                                                                                                9793
CPZ.CM.-.CAM3        --..--.----------....................................------A.CTA--A--A..GTA-GT-----------                                     9170
CPZ.GA.88.GAB2       --..GT.----------....................................------..CTAAACATAA.G-A-----------------                                  9146
CPZ.GA.-.CPZGAB      --..--.----------....................................------TTAAA--A---..TCAAG-A---.---------                                  9811
CPZ.US.85.CPZUS      --..--.----------....................................------A.CT---AA-A..GTA-GT-------------                                  9781

H2A.GW.-.ALI         A-..TCGG-GTTC-C-TGA..GTAACAAGACCCT...GGTCTGTTAGGACCCTTTCTGCTT..TGGGAA-CGG--GCA--------C------                                10353
H2A.DE.-.BEN         A-..TCGG-GTTC---TGA..GTAACAAGACCCT...GGTCTGTTAGGACCCTT---GCTT..TGGGAATCCA--GCA--------C------                                10359
H2B.CI.-.EHO         A-..TCGG-G......-TC..CACTAGAAACCCT...GGTCTGTTAGGACCCCTTCTGCTT..TGGGAA-CCA--GCA--------C------                                10242

MAC.US.-.239         AAC.TCGG-..-CTCAATA..ATAAGAAGACCCT...GGTCTGTTAGGACCCTTTCTGCTT..TGGGAA-CCG-AGCA--------C------                                10279

STM.US. .STM         AAC.TCGG .. CTCAA..........GACCCT...GGTCTGTTAGGACCCTT.....T TG                                                                 9892

SAB.SN.-.SAB1C       .........TACT-AG                                                                                                               10009

TAN.UG.-.TAN1        -...--.-CCTC-T-A-TG..G                                                                                                          9752

VER.KE.-.AGM155      -...-GAAGTTCCT-A-TG..G                                                                                                          9758
VER.KE.-.9063        -...--C-GTTCCT-A-TG..G                                                                                                          9779
VER.DE.-.AGM3        -G..--GACCCT---GGAA..T...............................-AACC-T..GGAGAAGT                                                         9294

GRV.ET.-.GRI_677     -GTG--....................................CCTCTCTTACTGGGTT-..T-C-G--CC---GTG---G--A---C----                                    9623

DRL.-.-.FAO          TGGAG-T-CCTCAGA-TAGGTGAGAGCTATCCTGAGCTTTAGTGAGAGTAAAGAAAA.AT-TC-AGCAG                                                           9732

RCM.GA.-.GAB1        -GTAGGGA-CCCTCA..GATTCTTGTGGCAGAAGAGCCTTGGGCTAAGAA.AATTC-CT--CAGT                                                               9465
RCM.NG.-.NG411       -GTTGGGA-CCCTCA..GATATTTGTGGCAGAAGTACTGAAGGCTAAGAAAAAATC-CA--CAG                                                                9557

MND-2.-.-.5440       CGGAG-T-CCTCAGAT-TGTGGCAGAGATTTGATATCTA..GTGAGAGTAGAGAAAA-AT-TC-AGCAGT-G-GCCCGAAC-GGGACT-GA--AGGGTTCAGGTCGCACCCACCTGTAAGGGAC   9703
LST.KE.-.lho7        ...........TGGTTCAGTGATCTCTCACTAGAGAGATTGGAGCCTTGTGAT-CGG---AGGGC--CCAG                                                         9957

SUN.GA.98.L14        --GTTCC-GT..............GTTCTCTCT...TTCTCCT.TCTAAACCCTGTTCGACTG--C-C--CT--G-A-C--TTA--GGCCGCCTAGTGATCCGGGGCTGACGGCTAGCA      10006
```