TABLE S1: Alignment of SL1 sequences

"Standard" SL1 sequences (total: 293)

Name												Count
A1.KE.86.ML170_1986_AF539405	A	CTCG	G	CTTGCTG	A-G	GTGCAC	A	CAGCAAG	AGG	CGAG	A	48
A1.SE.95.UGSE8131_AF107771	A	CTCG	G	CTTGCTG	AAG	GTGCAC	A	CAGCAAG	AGG	CGAG	A	9
A1.RU.03.03RU20_06_13_AY500393	G	CTCG	G	CTTGCTG	A-G	GTGCAC	A	CGGCAAG	AGG	CGGA	G	1
A2.CY.94.94CY017_41_AF286237	A	CTCG	G	CTTGCTG	A-G	GTGCAC	A	CGGCAAG	AGG	CGAG	G	2
A2.CD.97.97CDKTB48_AF286238	A	CTCG	G	CTTGCTG	A-G	GTGCAC	A	CGGCAAG	AGG	CGAG	A	1
B.AU.x.1181_AF538302	A	CTCG	G	CTTGCTG	A-A	GCGCGC	A	CGGCAAG	AGG	CGAG	G	35
B.AU.96.MBCD36_AF042105	A	CTCG	G	CTTGCTG	A-A	GCGCGC	A	CAGCAAG	AGG	CGAG	G	15
B.TW.94.TWCYS_AF086817	A	CTCG	G	CTTGCTG	A-A	GTGCGC	A	CGGCAAG	AGG	CGAG	G	4
B.CN.x.CNHN24_AY180905	A	CTCG	G	CTTGCTG	A-A	GCGCGC	A	CAGCAAG	AGG	CGAG	A	3
B.GB.86.GB8_AJ271445	A	CTCG	G	CTTGCTG	A-A	GCGCGC	G	CGGCAAG	AGG	CGAG	G	3
B.US.98.98USHVTN8229c6_AY56010	A	CTCG	G	CTTGCTG	C-A	GCGCGC	A	CGGCAAG	AGG	CGAG	G	2
B.AU.x.C42_AF538305	A	CTCG	G	CTTGCTG	A-A	GCGCGC	A	TAGCAAG	AGG	CGAG	G	1
B.AU.87.MBC925_AF042101	A	CTCG	G	CTTGCTG	T-A	GCGCGC	A	CGGCAAG	AGG	CGAG	G	1
B.MM.99.mSTD101_AB097870	A	CTCG	G	CTTGCTG	AAA	GCGCGC	G	CAGCAAG	AGG	CGAG	A	1
B.RU.04.04RU128005_AY682547	A	CTCG	G	CTTGCTG	A-A	GTGCGC	A	CAGCAAG	AGG	CGAG	G	1
B.US.84.SF33_AY352275	A	CTCG	G	CTTGCTG	AAA	GCGCGC	A	CGGCAAG	AGG	CGAG	G	1
B.US.98.98USHVTN3605c9_AY56010	A	CTCG	G	CTTGCTG	A--	GCGCGC	A	CGGCAAG	AGG	CGAG	A	1
C.BR.04.04BR013_AY727522	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CGGCAAG	AGG	CGAG	A	79
C.BW.00.00BW17956_AF443097	A	CTCG	G	CTTGCTG	A-A	GTGCAC	A	CGGCAAG	AGG	CGAG	A	11
C.BW.00.00BW192113_AF443101	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CGGCAAG	AGG	CGAG	G	11
C.BW.98.98BWMO18D5_AF443080	A	CTCG	G	CTTGCTG	AAA	GTGCAC	T	CGGCAAG	AGG	CGAG	A	5
C.BW.00.00BW18113_AF443098	G	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CGGCAAG	AGG	CGAG	A	3
C.BW.99.99BW4745_AF443085	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CGGCGAG	AGG	CGAG	A	3
C.BW.00.00BW20361_AF443102	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CGGCAGG	AGG	CGAG	A	2
C.BR.04.04BR038_AY727524	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CAGCAGG	AGG	CGAG	A	1
C.BW.00.00BW076820_AF443089	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	TGGCAAG	AGG	CGAG	A	1
C.BW.00.00BW3891_6_AF443113	A	CTCG	G	CTTGCTG	A-A	GTGTAC	T	CGGCAAG	AGG	CGAG	A	1
C.BW.00.00BW3970_2_AF443114	A	CTCG	G	CCTGCTG	A-A	GTGCAC	T	CGGCAAG	AGG	CGAG	A	1
C.BW.98.98BWMC14A3_AF443078	A	CTCG	G	CTTGCTG	AAA	GTGCAC	T	CGGCAAG	AGG	CGAG	G	1
C.BW.98.98BWMO36A5_AF443081	A	CTCG	G	CTTGCTG	GAA	GTGCAC	T	CGGCAAG	AGG	CGAG	G	1
C.BW.99.99BW393212_AF443083	A	CCCG	G	CTTGCTG	A-A	GTGCAC	T	CGGCAAG	AGG	CGAG	A	1
C.IL.98.98IS002_AF286233	A	CTCG	G	CTTGCTA	A-A	GTGCAC	T	TGGCAAG	AGG	CGAG	A	1
C.IN.94.94IN476_AF286223	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CAGCAAG	AGG	CGAG	G	1
C.ZA.00.1197MB_AY463234	A	CTCG	G	CTTGCTG	AAC	GTGCAC	T	CGGCAAG	AGG	CGAG	A	1
C.ZA.03.SK029B2_AY772691	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CAGTAAG	AGG	CGAG	A	1
C.ZA.03.SK065B1_AY772694	A	CTCG	G	CTTGCTG	A-A	GTGCAC	T	CGGCAGG	AGG	CGAG	G	1
D.UG.92.92UG001_AJ320484	A	CTCG	G	CTTGCTG	A--	GCGCGC	A	CGGCAAG	AGG	CGAG	G	2
D.CD.84.84ZR085_U88822	A	CTCG	G	CTTGCTG	AAA	GCGCGC	T	CGGCAAG	AGG	CGAG	G	1
F1.BR.89.BZ163_AY173958	A	CTCG	G	CTTGCTG	A-A	GTGCAC	A	CGGCAAG	AGG	CGAG	G	5
G.KE.93.HH8793_12_1_AF061641	A	CTCG	G	CTTGCTG	A-G	GTGCAC	G	CAGCAAG	AGG	CGAG	A	4
G.NG.92.92NG083_U88826	A	CTCG	G	CTTGCTG	A-A	GTGCAC	A	CAGCAAG	AGG	CGAG	A	2
G.ES.00.X558_AF423760	A	CTCG	G	CTTGCTG	AAG	GTGCAC	C	CAGCAAG	AGG	CGAG	A	1
H.BE.93.VI991_AF190127	A	CTCG	G	CTTGCTG	A-A	GTGCAC	A	CAGCAAG	AGG	CGAG	G	1
01_AE.TH.93.93TH9021_AF164485	A	CTCG	G	CTTGCTG	C-G	GTGCAC	A	CAGCAAG	AGG	CGAG	A	2
02_AG.NG.x.IBNG_L39106	A	CTCG	G	CTTGCTG	A-G	GTGCGC	A	CAGCAAG	AGG	CGAG	A	1
11_cpx.CM.96.4496_AF492623	A	CTCG	G	CTTGTTG	A-G	GTGCAC	A	CAGCAAG	AGG	CGAG	A	1
11_cpx.FR.99.MP1307_AJ291720	A	CTCG	G	CTTGCTG	AAA	GTGCAC	G	CAGCAAG	AGG	CGAG	A	1
12_BF.AR.99.ARMA159_AF385936	A	CTCG	G	CTTGCTG	A-A	GCGCGC	A	CGGCAAG	AGG	CGAG	A	1
13_cpx.CM.96.1849_AF460972	A	CTCG	G	CTTGCTG	A--	GCGCGC	A	CAGCAAG	AGG	CGAG	G	2
01B.MM.00.mIDU502_AB097865	A	CTCG	G	CTTGCTG	A-A	GCGCGC	G	CAGCAAG	AGG	CGAG	A	1
06A1.BJ.x.B76_AJ293865	A	CTCG	G	CTTGCTG	A--	GTGCAC	G	CAGCAAG	AGG	CGAG	G	1
A1D.TZ.96.TZBFL0088_AF442570	A	CTCG	G	CTTGCTG	A-A	GCGCGC	A	CAGCAAG	AGG	CGAG	A	1
A2C.ZM.90.ZAM184_U86780	A	CTCG	G	CTTGCTG	A-A	GTGCAC	C	CGGCAAG	AGG	CGAG	A	1
N.CM.97.YBF106_AJ271370	A	CTCG	G	CTCGTTG	---	GTGCAC	A	CAGCGAG	AGG	CGAG	G	3
O.BE.87.ANT70_L20587	G	CTCG	G	CTTAGCG	G-A	GTGCAC	C	CGCTAAG	AGG	CGAG	A	3
O.CM.91.MVP5180_L20571	G	CTCG	G	CTTAGCG	G-A	GTGCAC	C	TGCTAAG	AGG	CGAG	A	1
CPZ.CD.x.ANT_U42720	A	CTCG	G	TCTGCTG	C-A	GCGCGC	G	TAGCAGA	AGG	CGAG	A	1
CPZ.GA.x.CPZGAB_X52154	A	CTCG	G	CTTGCTG	A--	GTGCAC	A	CAGCAAG	AGG	CGAG	G	1
CPZ.US.85.CPZUS_AF103818	A	CTCG	G	CTCGCTG	A--	GTGCAC	A	CAGCGAG	AGG	CGAG	G	1

Imperfect palindrome

```
B.AU.x.C24_AF538304              A|CTCG|G|CTTGCTG|AA|ACGCGC-|G|CGGCAAG|AGG|CGAG|G   2
C.BW.96.96BW15C02_AF110974       A|CTCG|G|CTTGCTG|AA|TTGCAC-|T|CGGCAAG|AGG|CGAA|A   1
C.ZA.00.1228MB_AY463222          A|CTCG|G|CTTGCTG|AA|GCGCAC-|T|CGGCAAG|AGG|CGAG|A   1
F1.BE.93.VI850_AF077336          A|CTCG|G|CTTGCTG|AA|GTGCAAC|A|CGGCAAG|AGG|CGAG|A   1
01_AE.CF.90.90CF11697_AF197340   A|CTCG|G|CTTGCTG|AG|GTGCTC-|A|CAGCAAG|AGG|CGAG|A   1
12_BF.UY.99.URTR23_AF385934      A|CTCG|G|CTTGCTG|AA|GCGCAC-|G|--GCAAG|AGG|CGAG|G   1
12_BF.UY.99.URTR35_AF385935      A|CTCG|G|CTTGCTG|AG|GCGCAC-|G|--GCAAG|AGG|CGAG|G   1
16_A2D.KR.97.97KR004_AF286239    A|CTCG|G|CTTGCTG|AA|GCGC---|A|CGGCAAG|AGG|CGAG|G   1
CPZ.TZ.x.TAN1_AF447763           A|CTCG|G|CTTGTGA|CA|GCGCA--|A|TCACAAG|AGG|CGAG|G   1
```

Non-standard flanks of the palindrome

```
B.NL.00.671_00T36_AY423387       A|CTCG|G|CTTGCTG|AAGAAGAA|GCGCGC|G-|CGGCAAG|AGG|CGAG|G   1
12_BF.BR.99.BREPM108_AY771589    A|CTCG|G|CTTGCTG|AAGAA---|GTGCGC|A-|CGGCAAG|AGG|CGAG|G   1
C.BW.00.00BW16162_AF443092       A|CTCG|G|CTTGCTG|AA------|GTGCAC|TC|CGGCAAG|AGG|CGAG|A   1
```

Imperfect upper stem

```
B.US.90.US1_AY173952             A|CTCG|G|CTTGCTG|AA|GCGCGC|G|--GCAAG|AGG|CGAG|G   1
B.GA.88.OYI_M26727               A|CTCG|G|CTTGCTT|A-|GCGCGC|A|CGGCAAG|AGG|CGAG|G   1
B.NL.86.3202A21_U34604           A|CTCG|G|CTTGCTG|AA|GCGCGC|A|CGCCAAG|AGG|CGAG|G   1
C.ZA.01.01ZATM45_AY228557        A|CTCG|G|CTTGCTT|AA|GTGCAC|T|CGGCAAG|AGG|CGAG|A   1
01BC.MM.99.mIDU107_AB097868      A|CTCG|G|CTTGCTG|AA|GTGCAC|T|CGGC-AG|AGG|CGAG|A   1
05_DF.BE.x.VI1310_AF193253       A|CTCG|G|CTTGCTG|AA|GCGCGC|A|AGGCAAG|AGG|CGAG|G   1
```

Imperfect lower stem

```
B.AU.x.C76_AF538306              A|CTCT|G|CTTGCTG|A-A|GCGCGC|A|CAGCAAG|AGG|CGAG|G   1
C.BW.99.99BW46424_AF443084       G|ATCG|G|CTTGCTG|AAA|GTGCAC|T|CGGCAAG|AGG|CGAG|A   1
C.BW.00.00BW22767_AF443107       A|CTCG|G|CTTGCTG|A-A|GTGCAC|T|CGGCAAG|AGG|CGAA|A   2
J.SE.94.SE7022_AF082395          A|CTCG|G|CTTGCTG|A-A|GTGCAC|A|CGGCAAG|AGG|CGAA|G   1
01_AE.CF.90.90CF4071_AF197341    A|CTCG|G|CTTGCTG|A-G|GTGCAC|G|CAGCAAG|AGG|CGAA|G   1
01_AE.TH.93.93TH057_AF197338     A|CTCG|G|CTTGCTG|A-G|GTGCAC|A|CAGCAAG|AGG|CAAG|A   1
11_cpx.CM.97.MP818_AJ291718      A|CTCG|G|CTTGCTG|A-G|GTGCAC|A|CAGCAAG|AGG|CGAA|A   1
```

Variations in the internal loop

```
A1.KE.97.ML013_2_AY322185        A|CTCG|G|CTTGCTG|AG|GTGCAC|A|CAGCAAG|AAG|CGAG|A   1
C.BW.98.98BWMC122_AF443076       A|CTCG|G|CTTGCTG|AA|GTGCAC|T|CGGCAAG|GGG|CGAG|A   1
                                               Total "non-standard" sequences:   29
```

We retrieved the alignments of complete genomes of HIV-1 available on September 2006 at the Los Alamos HIV Sequence Database (1). Only the sequences without gaps in the SL1 region were used; alignments were manually modified to take into account the structural features of SL1 stem-loop. Note that some of the isolates may not correspond to fully infectious viruses, so the sequence variability presented in the table may be somewhat overestimated. On the other hand, multiple isolates from the same patients are not presented in these alignments, which may somewhat decrease the sequence variability. Only unique sequence variants of SL1 are shown in the table; the number in the right column shows the number of isolates with this particular sequence. The left column contains the database sequence ID of an arbitrary representative isolate. The ID contains the subtype or the recombinant number, country of origin, year of isolation and the GenBank ID (1). Entry A1.KE.86.ML170_1986_AF539405 contains SL1 sequence representative of the Mal isolate, and entry B.AU.x.1181_AF538302 contains sequence representative of the Lai isolate. Residues of the 6-nt palindromic loop are shown in red, upper stem in blue, lower stem in green, G-rich asymmetric internal loop in magenta, and residues flanking the palindrome are shown in cyan.

(1) Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J.W., Wolinsky, S. and Korber, B. (2005) HIV Sequence Compendium 2005, Los Alamos National Laboratory, Los Alamos, NM.