```
LOCUS       HIVWEAU160       9720 bp ss-RNA                      22-FEB-1996
DEFINITION  Human immunodeficiency virus type 1, isolate WEAU 1.60, complete
            genome.
ACCESSION   U21135
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate WEAU,
            plasmid clone 1.60, replication competent.
REFERENCE   1  (bases 1 to 9720)
  AUTHORS   Ghosh,S.K., Hahn,B. and Shaw,G.
  TITLE     Unpublished
  JOURNAL   Unpublished (1995)
REFERENCE   2  (sites; protease cleavage)
  AUTHORS   Tozser,J., Blaha,I., Copeland,T.D., Wondrak,E.M. and Oroszlan,S.
  TITLE     Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide
            substrates representing cleavage sites in Gag and Gag-Pol
            polyproteins.
  JOURNAL   FEBS Letters 281: 77-80 (1991)
  MEDLINE   91200319
REFERENCE   3  (sites; gag-pol fusion protein frameshift)
   AUTHORS  Reil,H., Kollmus,H., Weidle,U.H. and Hauser,H.
   TITLE    A heptanucleotide sequence mediates ribosomal frameshifting
            in mammalian cells
   JOURNAL  J. Virol. 67: 5579-5584 (1993)
   MEDLINE  93353637
COMMENT     Sequence kindly provided prior to publication by Sajal K. Ghosh,
            UAB, Birmingham.  A cytopathic HIV-1 virus was cloned from an
            acutely infected patient.  The clone WEAU 1.60 is replication
            competent and upon transfection produces highly cytopathic T-cell
            tropic virus.  The clone and the viral isolate from which it was
            derived are syncytium-inducing (SI).

            The WEAU 1.60 clone was obtained from a coculture of the
            patient's PBMCs, first with normal donor PHA-stimulated lympho-
            cytes for 14 days, then with the H9 T-cell line for another
            14 days.  The patients' blood specimen was obtained 15 days after
            the onset of clinical symptoms of acute (primary) infection, and
            35 days after a single sexual encounter with a partner whose
            virus was proven phylogenetically to be responsible for the
            transmission event.  The patient is identified as "Patient #1"
            in N. Engl. J. Med. 324: 954-960 (1991) and as "WEAU 0575" in
            Science 259: 1749-1754 (1993).

            The WEAU 1.60 clone has been completely sequenced from a plasmid.
            It appears to group with subtype B viruses.  There is a deletion
            of a single T at position 9069, resulting in a frameshift
            mutation and premature termination of nef.  The frameshifting
            deletion in nef was NOT present in the patients' unclutured PBMCs
            where instead there is a "T".  The nef gene was not interrupted
            in 10 of 10 clones analyzed by PCR sequencing from the uncultured
            PBMCs.

            This sequence is the reference sequence for the Molecular
            Immunology database (http://hiv-web.lanl.gov/immuno/index.html).
            It has been sequenced in its entirety by two different labs
            (G. Shaw and L. Hood) with 100% concordance.

BASE COUNT     3467 a   1730 c   2334 g   2189 t
ORIGIN      Beginning of 5' LTR.
        1 tggaagggct aatttactcc caaaaaagac aagatatcct tgatctgtgg gtctaccaca
          -> 5' LTR U3
       61 cacaaggctt cttccctgat tggcagaact acacaccagg gccagggact agatacccac
```

```
 121 tgtgctttgg atggtgcttc aagctagtac cagtggaacc agagaaggta gaagaagcta
 181 atgaaggaga gaacaacagc ttgttacacc ctatgagcct gcatgggatg gatgaccatg
 241 agaaagaagt attaatgtgg aagtttgaca gcaaactagc atttcatcac gtagcccgag
 301 agctgcatcc ggagtacttt aaagactgct gacatcgagt tttctacaaG GGACTTTCCa
                                                      nfKb - ---------
 361 ctgGGGACTT TCCggggagg cgtgatctgg gcgggacagg ggagtggcga gccctcagat
   nfKb ------- ---
 421 gctgcaTATA Agcagctgct ttttgcctgt actgggtctc tctggttaga ccagatctga
     signal ---- -                     U3 /\ R
 481 gcctgggagc tctctggcta gctagggaac ccactgctta agcctcAATA AAgcttgcct
                                                 signal ---- --
 541 tgagtgcttc aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc
              R /\ U5
 601 agaccatttt agtcagtgtg gaaaatctct agcagtggcg cccgaacagg gacccgaaag
                                   U5 <- LTR end
 661 cgaaagggaa accagaggag ctctctcgac gcaggactcg gcttgctgaa gcgcgcacgg
 721 caagaggcga ggggcggcga ctgGTgagta cgccgaaatt ttgactagcg gaggctagaa
                 major 5’ sj /\
 781 ggagagagAT Gggtgcgaga gcgtcggtat taagcggggg agaattagat aggtgggaaa
gag cds start ->
 841 aaattcggtt aaggccaggg ggaaataaaa aatataaact aaaacatata gtatgggcaa
 901 gcagggagct agaacgattc gcagtcaatc ctggcctgtt agaaacatca gaaggctgta
 961 gacaaatact gggacaatta caaccatccc ttcagacagg atcagaagaa cttagatcat
1021 tatataatac agtagctgtc ctctattgtg tgcatcaaaa gatagaggta aaagatacca
1081 aggaagcttt agataagata gaggaagagc aaaacaaaag caagaaaaag gcacagcaag
1141 cagcagctga cacagaaaac agaaaccagg tcagccaaaa ttaccctata gtgcagaacc
                                          p17MA /\ p24CA
1201 tacagggaca aatggtacat caggccctat cacctagaac tttaaatgca tgggtaaaag
1261 taatagaaga aaaggctttc agcccagaag taatacccat gttttcagca ttatcagaag
1321 gagccacccc acaagattta aacaccatgc taaacacagt ggggggacat caagcagcca
1381 tgcaaatgtt aaaagagacc atcaatgagg aagctgcaga atgggataga ttacatccag
1441 tgcatgcagg gcctattgca ccaggccaga tgagagaacc aaggggaagt gacatagcag
1501 gaactactag tacccttcag gaacaaatag catggatgac aaataatcca cccatcccag
1561 taggagaaat ctataaaaga tggataatcc taggattaaa taaaatagta agaatgtata
1621 gccctgtcag cattctggac ataagacaag gaccaaaaga acccttttaga gattatgtag
1681 accggttcta taaaactcta agagccgagc aagctacaca ggaggtaaaa aattggatga
1741 cagaaacctt gttggttcag aatgcgaacc cagattgtaa gactatttta aaagcattgg
1801 gaccaggagc tacactagaa gacatgatga cagcatgtca gggagtggga gggcccggcc
1861 ataaagcaag agttttggct gaagcaatga gccaagtaac aaatacagct accatgatga
                p24CA /\ p2X                                   p2X /\ p7NC
                                                                     (gag)
                                                               p2X /\ TF
                                                                  (gag-pol)
1921 tgcagagagg caattttagg agcccaagaa aaactattaa gtgtttcaat tgtggcaaag
1981 aaggacacat agccagaaat tgcagggccc ctaggaaaaa gggctgttgg aaatgtggac
2041 aggaaggaca ccaaatgaaa gattgtactg agagacaggc taatttttta gggaagatct
                                                     >    <          --
                                             gag-pol frameshift site
2101 ggtcttccca aaagggaagg ccagggaatt ttcctcagag cagactagaa ccaacagccc
     ---------- ---------- ----   p7NC /\ P6 (gag)
   stem-loop potentiates frameshift [3]
2161 caccagaaga gagcttcagg ttcagggaag agacaacaac tccctctcag aagcaggagc
2221 caatagacaa ggagctgtat cctttaactt ccctcaaatc actctttggc aacgacccct
                                   TF /\ Protease (pol)
2281 cgtcacaaTA Aagatagagg ggcaactgaa ggaagctcta ttagatacag gagcagatga
              < - gag P6 cds end
2341 tacagtatta gaagacatga atttgccagg gaaatggaaa ccaaaaatga tagggggaat
2401 tggaggtttt atcaaagtaa gacagtatga tcaggtaccc atagaaatct gtggacataa
2461 agctataggt acagtattag taggacctac acctgtcaac ataattggaa gaaatctgtt
```

```
 2521 gactcagatt ggctgcactt taaattttcc cattagtcct attgaaactg taccagtaaa
                         Protease /\ Reverse Transcriptase p51 (pol)
 2581 attaaagcca ggaatggatg gcccaaaagt taaacaatgg ccattgacag aagagaaaat
 2641 aaaagcatta gtagaaattt gtacagaaat ggaaaaggaa ggaaaaattt caaaaattgg
 2701 gcctgaaaat ccatataata ctccagtatt tgccataaag aaaaaagaca gtactaaatg
 2761 gagaaaatta gtagatttca gagaacttaa taagagaact caagacttct gggaagttca
 2821 attaggaata ccacatcctt cagggttaaa aaagaaaaaa tcagtaacag tactggatgt
 2881 gggtgatgca tatttttcag tacccttaga tgaagacttc aggaagtaca ctgcatttac
 2941 catacctagt ataaacaatg aaacaccagg gattagatat cagtacaatg tgcttccaca
 3001 gggatggaaa ggatcaccag caatattcca aagtagcatg acaaaaatat tagagccttt
 3061 tagaaaacaa aatccagaca tagttatcta tcaatacatg gatgatttgt atgtaggatc
 3121 tgacttagaa atagggcagc atagaacaaa aatagaggag ctgagacaac atctgttgag
 3181 gtggggattt accacaccag acaaaaaaca tcaaaaagac cctccattcc tttggatggg
 3241 ttatgaactc catcctgata aatggacagt acagcctata aagctgccag aaaaagaaag
 3301 ttggactgtc aatgacatac agaagttagt gggaaaattg aattgggcaa gtcagattta
 3361 cgcagggatt aaagtaaagc aactatgtaa actccttagg gggaccaaag cactaacaga
 3421 aataatacca ataacagaag aagcagagct agagctggca gaaaacaggg aaattctaaa
 3481 agaaccggta catggagtgt attatgaccc atcaaaagac ttaatagcag agctacagaa
 3541 gcaggggcaa ggccaatgga catatcagat ttatcaagag ccatttaaaa atctgaaaac
 3601 aggaaagtat gcaagagtga ggggtgccca cactaatgat gtaaaacaat taacagaggc
 3661 agtgcagaaa ataaccacag aaagcatagt aatatgggga aagactccta aatttaaact
 3721 acccatacaa aaagaaacat gggaaacatg gtggacagag tattggcaag ccacctggat
 3781 tcctgagtgg gagtttgtca ataccccctcc cttagtgaaa ttatggtatc agttagagaa
 3841 agaacccata gtaggagcag aaactttcta tgtagatggg gcagctaaca aggagactaa
         Reverse Transcriptase p51 /\ RNase H
 3901 attaggaaaa gcaggatatg ttactaacag aggaagacaa agagttgtct ccctaactga
 3961 cacaacaaat cagaagactg agttacaagc tattcatcta gctttgcagg attcgggatt
 4021 agaagtaaac atagtgacag actcccaata tgcattagga atcattcaag cacaaccaga
 4081 tcagagtgaa tcagagttag tcagtcaaat aatagaacag ttaataaaaa aggaaaaggt
 4141 ctacctagca tgggtaccag cacacaaagg aattggagga aatgaacaag tagataaatt
 4201 agtcagttca ggaatcagga aagtactatt tttagatggg atagataagg cccaagaaga
                          RNase H /\ Integrase
 4261 acatgagaaa tatcacagta attggagagc aatggctagt gattttaacc tgccacctgt
 4321 agtagcaaaa gaaatagtag ccagctgtga taaatgtcag ctaaaaggag aagccatgca
 4381 tggacaagta gactgtagtc caggaatatg gcaactagat tgtacacatt tagagggaaa
 4441 aattatcctg gtagcagttc atgtagccag tggatatata gaagcagaag ttattccagc
 4501 agagacaggg caagaaacag catactttat cttaaaacta gcaggaagat ggccagtaaa
 4561 aacaatacat acagacaatg gcagcaattt caccagtact acggttaagg ccgcctgttg
 4621 gtgggcgggg atcaagcagg aatttggcat tccctacaat ccccaaagtc aaggagtgat
 4681 agaatctatg aataaagaat taaagaaaat tataggacag gtaagagatc aagctgaaca
 4741 tcttaagaca gcagtacaaa tggcagtatt catccacaat tttaaaagaa aagggggat
 4801 tgggggtac agtgcagggg aaagaataat agacataata gcaacagaca tacagactca
 4861 acaattacaa aaacaaatta caaaaattca aaattttcgg gtctattacA Gggacagcag
                                                        3′ sj /\
 4921 agatccactt tggaaaggac cagcaaagct tctctggaaa ggtgaagggg cagtagtaat
 4981 acaagataat agtgacataa aagtagtgcc aagaagaaaa gcaaagatca ttagggattA
                                                                      -
 5041 TGgaaaacag atggcaggtg atgattgtgt ggcaagtaga caggatgagg atTAGaacat
      >                                                          <-
vif cds start                                                 pol cds end
 5101 ggaaaagttt agtaaaacac catatgtata tttcaaagaa agctaaggga tggtcttata
 5161 gacatcacta cgaaagcact catccaagaa taagttcaga agtacatatc ccactagggg
 5221 aaggtaaatt agtaataaca acatattggg gtctgcatac aggagaaaga gactggcatt
 5281 tgggtcaagg agtctccata gaatggagga aacagagata tagcacacaa gtagaccctg
 5341 acctagcaga ccaactaatt catctgtatc attttgattg tttttcAGag tccgctataa
                                                          /\ 3′ sj
 5401 gaaatgccat attaggacat ttagttattc ctaggtgtga atatcaagca gggcataaca
 5461 agGTaggatc tctacaatac ttggcactaa cagcattaat aacaccaaaa aagataaagc
 5′ sj /\
```

```
 5521 cacctttgcc tagtgttaag aaactgacag aggatagATG gaacaagccc cagaagacca
                                vpr cds start ->
 5581 agggccacag agggagccat acaatgaatg gacacTAGag cttttagagg aacttaagag
                                            <- vif cds end
 5641 tgaagctgtt agacattttc ctagggtatg gctccatagc ttaggacaac atatctatga
 5701 aacttatggg gatacttgga caggagtgga agccataata agaatcctgc aacaactgct
 5761 gtttattcat ttcAGaattg ggtgtcgaca tagcagaata ggcattactc cacagagaag
                    /\ 3′ sj
 5821 agcaagaaAT Ggagccagta gatcaTAGac tagagccctg gaagcatcca ggaagtcagc
Tat cds start ->                  <- Vpr cds end
 5881 ctaagactgc ttgtaccaat tgctattgta aaagatgttg ctttcattgc caagtttgtt
 5941 tcatgacaaa AGgcttAGgc atctcctATG gcAGgaagaa gcggagacag cgacgaagat
                 /\    /\ rev cds ->     /\
              3′ sj  3′ sj  start        3′sj
 6001 ctcctcaaaa cagtcagact catcaagatt ctctatcaaa gcaGTaagta gtacatgtaA
                                   tat/rev 5′ sj /\                -
 6061 TGcaatcttt gcaaatatta gcaatagtag ccttagtagt agcaggaata atagcaatag
      >
Vpu cds start
 6121 ttgtgtggtc catagtactc atagaatata ggaaaatatt aagacaaaga aaaatagaca
 6181 ggttaattga tagaataaga gacagagcag aagacagtgg caATGagagt gaaggggatc
                                   env cds start ->
 6241 aggaagaatt atcagcactt gtggaaatgg ggcatcatgc tccttgggat attgatgatc
 6301 tgTAGtgctg cagaaaactt gtgggtcaca gtctattatg gggtacctgt gtggaaagaa
        <--/\ env signal pept end, gp120 start
    vpu cds end
 6361 gcaaccacca ctctattttg tgcatcagat gctaaagcat atgatacaga agtacataat
 6421 gtttgggcca cacatgcctg tgtacccaca gaccccaacc cacaagaagt agtattggaa
 6481 aatgtgacag aaaattttaa catgtggaaa aataacatgg tagaacagat gcatgaagat
 6541 ataattagtt tatgggatca aagcctaaaa ccatgtgtaa aattaacccc actctgtgtt
 6601 actttaaatt gcactaatgt gaatgtgact aatttgaaga atgagactaa taccaatagt
                  ------          --- ---             ----- -
                  imperfect repeat/duplication region
                    15 bp insertion compared to HXB2
 6661 agtagtggag gggaaaagat ggaggaggga gaaatgaaaa actgctcttt caatgtcacc
 6721 acactcataa gaaataagag aaagactgaa tatgcacttt tttataaact tgatgtaatg
 6781 ccaatagatc atgataatac aagctatacg ttgataaatt gtaaatcctc aaccattaca
 6841 caggcctgtc caaaggtatc ctttgaacca attcccatac attattgtgc cccggctggt
 6901 tttgcgattc taaagtgtaa tgataagaag ttcaatggaa agggaccctg taaaaatgtc
 6961 agcacagtac aatgtacaca cggaattaga ccagtagtgt caacccaatt gctgttaaat
 7021 ggcagtctag cagaagaaga catagtaatt agatctgaaa atttcacgga caatgctaaa
 7081 aacataatag tacagctgaa tgtatccata gaaattaatt gtacaagacc caacaacaat
 7141 acaagaaaaa aaataacttt aggaccaggg agagtacttt atacaacagg agaaataata
 7201 ggagatataa gacgagcaca ttgtaacctt agtagaacaa gttggaataa cactttaaaa
 7261 cagatagttg aaaaattaag agaaataaaa caatttaaga ataaaacaat agtttttaaa
 7321 caatcctcag gaggggaccc agaaattgta atgcacagtt ttaattgtgg aggggaattt
 7381 ttctactgta attcaacaca gctgtttaat agtacttggc atgctaatgg tacttggaag
 7441 aatactgaag gggcagataa caatatcaca ctcccatgca gaataaaaca aattataaac
 7501 aggtggcagg aagtaggaaa agcaatgtat gccccaccca tcgaaggaca aattagatgt
 7561 ttatcaaata ttacagggtt actattaaca agagatggtg gtagtagtga agagaaccag
 7621 accgagatct tcagacctgg aggaggaaat atgaaggata attggagaag tgaattatat
 7681 aaatataaag tagtaaaaat tgaaccatta ggagtagcac ccaccaaggc aaagagaaga
 7741 gtggtgcaga gagaaaaaag agcagtggga atgctaggag ctatgttcct tgggttcttg
                  env gp120 /\ env gp41
 7801 ggagcagcag gaagcactat gggcgcagcg tcaatgacgc tgacggtaca ggccagacta
 7861 ttattgtctg gtatagtgca acagcagaac aatctgctga gggctattga ggcgcaacag
 7921 catctgttcg aactcacagt ctggggcatc aaacagctcc aggcaagagt cctggctgtg
 7981 gaaagatacc taaaggatca gcagctcctg gggatttggg gttgctctgg aaaactcatc
 8041 tgcaccacta ctgtgccttg gaatgctagt tggagtaata gatctcagga ttacatttgg
```

```
8101 aataacatga cctggatgga gtgggagaga gaaattaaca attacacagg cttaatatac
8161 aacttaattg aagaatcgca gaaccaacaa gaaaaaaatg agcaagaatt attggaattg
8221 gataaatggg caagtttgtg gacttggttt gacatatcaa actggctgtg gtatataaaa
8281 atcttcataa tgatagtagg aggcttgata ggtttaagaa tagtttttac tgtactttcc
8341 atagtaaata gagttaggca gggatactca ccattgtcat ttcAGaccca cctcccagcc
                                      tat/rev 3' sj /\
8401 ccgaggggac ccgacaggcc cgaaggaatc gaagaagaag gtggagagag agacagagac
8461 agatccggaa gatTAGtgga tggattctta acacttatct gggtcgacct acggagcctg
                    <- tat cds end
8521 tgcctcttcc tctaccaccg cttgatcgac ttactcttga ttgcaaagag gattgtggaa
8581 cttctgggac gcagggggtg ggaagctctc aaatattgtt ggaatctcct gcagtattgg
8641 agccaggaac taaagaaTAG tgctgttagt ttgcttaatg ccacagctat agcagtagct
                        <- rev cds end
8701 gaggggacag atagggttat agaaatagtg caaagaactt gtagagctat tctccacata
8761 cctagaagaa taagacaggg cttagaaagg gctttgctaT AAgATGggtg gcatatggtc
                                                <- -> nef cds
                                      env gp41 end      start
8821 aaaacgtagt ggtagtggat ggcctgctat aagggaaaga atgaaacgag ctgagccagc
8881 agcagaaggg gtgggagcag tatctcgaga cttggcaaaa catggagcaa tcacaagtag
8941 caatgagaat aatgctgatt gcgtctggtt aaaagcacaa gaggatgagg aggtgggttt
9001 tccagttaga cctcaggtac ctttaagacc aatgacttac aaggcagcgc atgatcttag
9061 ccactttTAA aagaaaaggg gggactggaa gggctaattt actcccaaaa aagacaagat
        /  \ <- nef cds         -> 3' LTR U3
       /    \   premature end
      frameshift ("t" missing)
9121 atccttgatc tgtgggtcta ccacacacaa ggcttcttcc ctgattggca gaactacaca
9181 ccagggccag ggactagata cccactgtgc tttggatggt gcttcaagct agtaccagtg
9241 gaaccagaga aggtagaaga agctaatgaa ggagagaaca acagcttgtt acaccctatg
9301 agcctgcatg ggatggatga ccatgagaaa gaagtattaa tgtggaagtt tgacagcaaa
9361 ctagcatttc atcacgtagc ccgagagctg catccggagt actttaaaga ctgcTGAcat
                                                                  <- nef cds
                                                                normal end
                                                             (not this clone)
9421 cgagttttct acaagggact ttccactggg gactttccgg ggaggcgtga tctgggcggg
9481 acaggggagt ggcgagccct cagatgctgc aTATAAgcag ctgctttttg cctgtactgg
                                signal -----                     U3 /\ R
9541 gtctctctgg ttagaccaga tctgagcctg ggagctctct ggctagctag ggaacccact
9601 gcttaagcct cAATAAAgct tgccttgagt gcttcaagta gtgtgtgccc gtctgttgtg
          signal ------                 R /\ U5
9661 tgactctggt aactagagat ccctcagacc attttagtca gtgtggaaaa tctctagcag
                                                              3' LTR <-
//
```

```
LOCUS       HIVIBNG         9201 bp     ss-RNA                 22-FEB-1996
DEFINITION  Human immunodeficiency type 1, Nigerian isolate IbNg, complete
            genome.
ACCESSION   L39106
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate IbNg
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1 (bases 1 to 9201)
  AUTHORS   Howard, T.M. and Rasheed, S.
  TITLE     Genomic Structure and Nucleotide Sequence Analysis of a New HIV
            Type 1 Subtype A Strain from Nigeria.
  JOURNAL   AIDS Res. Hum. Retroviruses In Press (1996)
REFERENCE   2 (bases 5758 to 7492)
  AUTHORS   Howard, T.M., Olaylele, D.O., and Rasheed, S.
  TITLE     Sequence analysis of the glycoprotein 120 coding region of a new
            HIV type 1 subtype A strain (HIV-1IbNg) from Nigeria.
  JOURNAL   Aids Res Hum Retroviruses 10, 1755-1757 (1994)
  MEDLINE   95194745
COMMENT     HIV-1 IbNg was isolated from the PBMCs of an apparently healthy
            23 year old man from Nigeria.  The patients' PBMCs were cocultured
            with PHA-stimulated donor PBMCs from an HIV sero-negative donor.
            After confirming HIV infectio in the culture, a mixture of cells
            and culture supernatant was used to infect a second culture of
            donor PBMCs, with fresh PHA-stimulated PBMCs added on days 4 and
            6.  The cultured cells were harvested on day 8 and cytoplasmic
            RNA was harvested.  RT-PCR was used to amplify the complete
            HIV-1 genome in 5 overlapping segments.

            The IbNg sequence has a 16 bp insertion within the Lys-tRNA
            primer binding site, just 3' of the 5' LTR.  It also has a
            single nucleotide deletion in the tat cds at position 5450.
FEATURES             Location/Qualifiers
BASE COUNT     3299 a   1623 c   2218 g   2061 t
ORIGIN      1 bp upstream of BglII site, in 5' LTR.
        1 cagatctgag cctgggagct ctctggctag cggaggaacc cactgcttaa gcctcaataa
                                                                     -----
                                                                poly-A signal
       61 agcttgcctt gagtgcttca agtagtgtgt gcccgtctga ttgtgtgact ctggtaacta
          -                  R /\ U5
      121 gagatccctc agaccactct agactgtgta aaaatctcta gCAgtggcgc ccgaacagga
                                                   LTR <-
      181 cttgacggta atagggactc gaaagcgaaa gttccagaga agatctctcg acgcaaggac
      241 tcggcttgct gaggtgcgca cagcaagagg cgagagcggc gactgGTgag tacgccaaat
                                             major 5' sj /\
      301 tttgactagc ggaggctaga aggagagagA TGggtgcgag agcgtcagta ttaagtgggg
                           gag cds start ->
      361 gaaaattaga tgcatgggag aaaattcggt taaggccagg gggaaagaaa aaatatagac
      421 taaaacatct agtatgggca agcagggagc tggaaagatt cgcacttaac cctggccttt
      481 tagaaacagc agagggatgt caacaactaa tggaacagtt acaatcagct ctcaggacag
      541 gatcagaaga acttaaatca ttatttaata caatagcaac cctttggtgc gtgcatcaaa
      601 ggatagacat aaaagacacc aaggaagcct tagataaaat ggaggaaata caaaataaga
      661 gcaagcaaaa ggtacagcag acagcagctg ccacaggaag cagcagccaa aattacccta
                                                                p17MA /\ p24CA
      721 tagtgcaaaa tgccaaaggg caaatgacac atcagtccat gtcacctaga actttaaatg
      781 catgggtgaa agtaatagaa gaaaagggtt tcagcccaga agtaataccc atgttttcag
      841 cattatcaga gggagccacc ccgcaagatt taaatatgat gctaaacata gtgggggggac
      901 accaggcagc aatgcagatg ttaaaagata ccatcaatga ggaagctgca gaatgggata
      961 gggtacatcc agtacatgca gggcctattc caccaggcca gatgagggaa ccaaggggaa
```

```
 1021 gtgacatagc aggaactact agtacccttc aagaacaaat aggatggatg acaagcaatc
 1081 cacctatccc agtgggagaa atctataaaa gatggatagt cctgggatta aataaaatag
 1141 taaggatgta tagccctgtc agcattttgg acataagaca agggccaaaa gaacccttta
 1201 gagattatgt agataggttc tttaaaactt tgagagctga acaagctacg caggaggtaa
 1261 aaaactggat gacagaaacc ttgctggtcc aaaatgcgaa tccagactgt aagtccattt
 1321 taagagcatt aggaacaggg gctacattag aagaaatgat gacagcatgt cagggagtgg
 1381 gaggacctgg ccataaagca agggttttgg ctgaggcaat gagtcaagta caacaggcca
                              p24CA /\ p2X
 1441 acgtaatgat gcagagaggc aattttaggg gccagagaac aataaagtgt ttcaactgtg
            p2X /\ p7NC (gag)
                   TF (gag-pol)
 1501 gcaaagaagg acacctagcc agaaattgca aggcccctag gaaaaggggc tgttggaaat
 1561 gtgggaaaga aggacaccaa atgaaagact gtactgagag acaggctaat tttttaggga
                                                                - -----
                                                     gag-pol -1 frameshift site
 1621 aaatttggcc ttccagcaag gggaggccag gaaattttcc tcagagcaga ccggaaccaa
          ----- ----        --------    p7NC /\ p6 (gag)
           stem      loop     stem
     stem-loop ptentiates gag-pol frameshift
 1681 cagctccacc agcagagagc tttgggatgg gggaagagat acctccctct ccacagcagg
 1741 aaccgaggga caaggggcta tatcctcctt taacttccct caaatcactc tttggcaacg
                                          TF /\ Protease (pol)
 1801 accctTAGtt acagtaagaa tagagggaca gctaatagaa gccctattag acacaggagc
gag cds end <-
 1861 agatgataca gtattagaag acataaattt accaggaaaa tggaaaccaa aaatgatagg
 1921 gggaattgga ggttttatca aagtaagaca atatgatcaa atacttatag aaatttgtgg
 1981 aaaaaaggcc ataggtacag tattagtagg acctacacct gtcaacataa ttggacgaaa
 2041 tatgttgact cagattggtt gtactttaaa ttttccaatt agtcctattg aaactgtgcc
                                  Protease /\ p51 Reverse Transcriptase
 2101 agtaaaatta aagccaggaa tggatggccc aaaggttaaa caatggccat tgacagaaga
 2161 aaaaataaaa gcattaacag acatttgtac agagatggaa aaggaaggaa aaatttcaaa
 2221 aattggacct gaaaatccat acaatactcc agtatttgcc ataaagaaaa aggatagtac
 2281 taaatggaga aaattagtag atttcagaga actcaataag agaactcaag atttctggga
 2341 ggtccagtta ggaatacctc atcccgcggg attaaaaaag aaaaaatcag taacagtgct
 2401 agatgtgggg gatgcatatt tttcagttcc cttagataaa gactttagaa agtatactgc
 2461 atttactata cctagtgtaa ataatgagac accagggatt agataccagt acaatgtgct
 2521 tccacaggga tggaaagggt caccggcaat atttcaggca agcatgacaa aaatcttaga
 2581 gccctttaga acaaaaaatc cagagatagt gatctaccaa tacatggatg atttatatgt
 2641 aggatctgat ttagaaatag ggcagcatag agcaaaaata gaggagttga gaggacatct
 2701 actgaaatgg ggatttacca caccagacaa aaagcatcag aaagaacctc catttctttg
 2761 gatgggttat gaactccatc ctgataaatg gacagtccag cctgtagaac taccagaaaa
 2821 agacagctgg actgtcaatg atatacagaa attagtggga aaactaaatt gggcaagtca
 2881 gatttatgca ggaattaaaa taaagcaact gtgtagactc ctcaggggag ccaaagcact
 2941 aacagatata gtagcactga ctgaggaagc agaattagaa ttggcagaga acagggaaat
 3001 tctaaaagaa cctgtacatg gggtatatta tgacccaaca aaagacttag tagcagaatt
 3061 acagaaacaa gggcaagacc aatggacata tcaaatttat caagagccat ttaaaaattt
 3121 aaaaacagga aaatatgcaa aaaagaggtc tgcccacact aatgatgtaa agcaattaac
 3181 agaggtagtg caaaaagtag ctatggaaag catagtaata tggggaaaga cccctaaatt
 3241 tagactaccc atacaaagag agacatggga aacatggtgg atggagtatt ggcaggctac
 3301 ctggattcct gaatgggagt ttgtcaatac ccctcctcta gtaaaattat ggtaccagtt
 3361 agaaaaagac cccatagtag gagcagaaac tttctatgta gatggagcag ctaatagaga
                p51 Reverse Transcriptase /\ RNase H
 3421 gactaagata ggaaaagcag ggtatgtcac tgacagagga agacaaaagg ttgtttccct
 3481 aactgagaca acaaatcaaa agactgaatt acatgcaatt catttagcct tacaggattc
 3541 aggatcagaa gtaaatatag taacagactc acagtatgca ttaggaatca ttcaggcaca
 3601 accagacagg agtgaatcag agttagtcaa tcaaataata gagaagctaa taaaaaagga
 3661 caaagtctac ctgtcatggg taccagcaca caaagggatt ggaggaaatg aacaagtaga
 3721 taaattagtc agtaatggaa tcaggaaggt actattttta gatggcatag ataaagccca
                                    RNase H /\ Integrase
```

```
    3781 agaagagcat gaaagatatc acagcaattg gaaggcaatg gctagtgatt ttaatctgcc
    3841 acctatagta gcaaaagaaa tagtggccag ctgtgataaa tgtcagatga aaggggaagc
    3901 catgcatgga caagtagact gtggtccagg aatatggcaa ttagattgta cacatttaga
    3961 agggaaaatt atcttagtag cagtccatgt agccagtggc tatatagaag cagaagttat
    4021 cccagcagaa acaggacagg agacagcata ctttatatta aaattagcag gaagatggcc
    4081 agtgaaagta atacacacag acaatggcag caatttcacc agtgctgcag taaaggcggc
    4141 atgttggtgg gcaaatgtca cacaagaatt tggaattccc tacaatcccc aaagccaagg
    4201 agtagtggaa tctatgaata aagaattaaa gaaaattata ggacaggtca gagatcaagc
    4261 tgaacacctt aagacagcag tacagatggc agtattcatt cacaatttta aaagaaaagg
    4321 ggggattggg gggtacagtg caggggaaag aataatagat ataatagcat cagatataca
    4381 aactaaagaa ctacaaaaac agattataaa aattcaaaat tttcgggttt attacAGgga
                                                                3′ sj /\
    4441 cagcagagac cccatttgga aaggaccagc aaaactactc tggaaaggtg aaggggcagt
    4501 agtaatacag gacaatagtg atataaaggt agtaccaaga agaaaagcaa aaatcattag
    4561 ggattATGga aaacagatgg caggtgatga ttgtgtggca ggtagacagg atgaggatTA
vif cds start ->                                                        <
    4621 Gaacatggaa tagtttagta aaacatcata tgtatgtctc aaagaaagct aagggttggt
         - pol cds end
    4681 tttatagaca tcactttgaa agtaggcatc caaaagtatg ttcagaagta cacatcccac
    4741 taggggatgc taggttagta gtaagaacat attggggtct gcatacagga gaacgagact
    4801 ggcacttggg tcatggggtc tccatagaat ggaagcagaa aagatatagc acacaaatag
    4861 accctgactt agcagaccaa ctgattcacc tgtattattt taattgtttt tcAGactctg
                                                            3′ sj /\
    4921 ccataagaaa agccatatta ggagaaatag ttagacctag gtgtgaatat caagcaggac
    4981 ataataagGT aggatcgcta caatatttgg cactgaatgc tttagtagca ccaacaaaga
          5′ sj /\
    5041 caaagccacc tttgcctagt gttaggaagt tagcagaaga cagATGgaaa gagccccaga
                                         vpr cds start ->
    5101 agaccagggg ccacagaggg agccgtccaa tgaatggaca cTAGaactgt tagaagagct
                                                       <- vif cds end
    5161 taaacatgaa gctgttagac actttcctag gccgtggctc catggattag gacaacatat
    5221 ctatgaaaca tatggggata cttgggaagg ggttaaagct ataataagaa ttttgcaaca
    5281 actattgttt gttcatttcA Gaattgggtg tcaacatagc agaataggca ttattcaagg
                        3′ sj /\
    5341 gagaagaggc aggaATGgag ctggtagatc cTAGcctaga gccctggaac cacccgggaa
          tat cds start ->                 <- vpr cds end
    5401 gtcagcctac aactgcttgt agcaagtgtt actgtaaaat gtgctgctgc attgccaatt
                                                             /\
                                             single bp deletion in tat cds
    5461 gtgctttctg aacaAGggct tAGgcatctc ctATGgcAGg aagaagcgga gacgccgacg
                        /\      /\         ->    /\ 3′ sj
                      3′ sj   3′ sj   rev cds start
    5521 aggaactcct cagagccgtc aggatcatca aaatcctgta ccaaagcaGT gagtagtaat
                                              tat/rev 5′ sj /\
    5581 aattagcata tgtgATGcaa cctttaacaa taactgcaat agtaggacta gtagtagcat
          vpu cds start ->
    5641 tcatagcagc catagttgtg tggagcatag tatatataga atataggaaa ataaggaaac
    5701 agaagaaaat agacaggtta cttgatagaa taagagaaag agcagaagat agtggcaATG
                                                         env cds start ->
    5761 agagtgatgg ggatacagaa gaattatcca ctcttatgga gatggggtac gaatatattt
    5821 tggataatga taatttgTAA tgctgaacaa ttgtgggtca cggtctacta tggggtacct
                           <- vpu cds end
    5881 gtgtggaaga cggcagagac caccctattt tgtgcatcag atgctaaagc atatgataca
    5941 gaagtacata atgtctgggc tacacatgcc tgtgtaccca cagaccctaa cccacaagaa
    6001 atacatttgg aaaatgtaac agaaaagttt aacatgtgga aaaataacat ggtagaacag
    6061 atgcatgaag atataattag tctatgggac caaagcctaa agccatgtgt aaagttaacc
    6121 cctctctgcg ttactttaga ttgtcataat ttcaacaaca gctacagcaa cagtagcaac
    6181 ctaactagtg acatgaatgg ggaaataaaa aactgctctt tcaatataac cacagaagta
    6241 agagataaga aaaagaagat gcatgcactt ttttatagac ttgatgtagt acaaattaat
```

```
   6301 gaaaataatg gtagtcagta taggttaata aattgtaata cctcagccat tacacaggct
   6361 tgtccaaagg tatcctttga gccaattccc atacattatt gtgccccagc tggttttgcg
   6421 attctaaagt gtaaggataa agggttcaat ggaacagggc catgcaaaaa tgtcagcaca
   6481 gtacaatgca cacatggaat caagccagta gtatcaactc aactgctgtt aaatggcagt
   6541 ctagcagaag gagaagtagt gatcagatct gaaaatatca caaacaatgc caaaaccata
   6601 atagtacagt tggctaatcc tgtaaaaatt aattgtacca gacctaacaa caatacaaga
   6661 aaaggtgtac atataggacc agggcaagca ttctatgcaa caggtgacat aataggggat
   6721 ataagacaag cacattgtaa tgtcagtaaa acagaatgga ataaaacttt acatcaggta
   6781 gttacacaat taaagacgta ctttaagaac accacaataa tctttgctaa ccccttagga
   6841 ggggatgtag aaattacaac acatagtttt aattgtggag gagaattttt ctattgcaat
   6901 acatcaaaac tgtttaatag cacttgggat aatagcaata gcactgccaa ccacacaggg
   6961 tcaaatgaca ctataactct ccaatgcaga ataaagcaaa ttgtaaatat gtggcagaaa
   7021 gtaggacaag caatgtatgc ccctcccatc caaggaataa taagatgtga ttcaaacatt
   7081 acaggactac tattaacaag agatggaggg aataatagta caaatgaaac cttcaggcct
   7141 ggaggaggag atatgaggga caattggaga agtgaattat ataagtataa agtagtaaaa
   7201 attgaaccac taggtgtagc acccacccgt gcaaaaagaa gagtggtgga gagagaaaaa
   7261 agagcggttg gactgggagc tgtcttcctt gggttcttag gagcagcagg aagcactatg
env gp120 /\ env gp 41
   7321 ggcgcgcggt caataacgct gacggtacag gccagacaat tattgtccgg catagtgcaa
   7381 cagcagagca atttgctgaa agctatagag gctcaacaac atctgttgaa actcacggtc
   7441 tggggcatta aacagctcca ggcaagagtc ctggctctgg aaagatacct aagagatcaa
   7501 cagctcctag gaatttgggg ctgctctgga aaactcatct gcaccactac tgtaccctgg
   7561 aactctagct ggagtaataa aacttttaat gacatatggg ataacatgac ctggatacaa
   7621 tgggagaaag aaattagcaa ttacacagac ataatatata atctaattga agaatcgcag
   7681 aaccggcagg aaaagaatga acaagacttg ttggcattgg acaagtgggc aagtctatgg
   7741 aattggtttg acataacaaa ttggctatgg tatataagaa tatttataat gatagtagga
   7801 ggtttaatag gtttaagaat agtttttgct gtgcttacta taataaatag agttaggcag
   7861 ggatactcac ctttgtcgtt ccAGaccctt acccaccacc agagggaacc cgacagaccc
                  tat/rev 5′ sj /\
   7921 gaaagaatcg aagaaggagg tggagagcaa gacaaagaca gatccgtgcg atTAGtgagc
                                                 tat cds normal end <-
                                     (not in this isolate due to deletion)
   7981 ggattcttag cacttgcctg ggacgatctg cggagcctgt gcctcttcag ctaccaccga
   8041 ttgagagact taatcttgat tgcagcgagg actgtggaac ttctgggaca caactgtctc
   8101 aagggactga gactggggtg gggagccctc aaatatctgt ggaatcttat atcatactgg
   8161 gttcaggaac taaagaaTAG tgctattaat ttgcttaata caatagcaat agtagtagct
                         <- rev cds end
   8221 aactggacag atagagctat agaaatagga caaagagtgg gtagagctat tcgcaacata
   8281 cctagaagaa tcagacaggg cttcgaaagg gctttgctaT AAcATGggtg gcaagtggtc
                                      env cds end <- -> nef cds start
   8341 aaaaagcagc atagtgggat ggcctaaggt tatgaaaaga atgagacaaa cccctacagc
   8401 agcaacagga gtaggagcag catctcaaga tttagataga catggagcaa tcacaagcag
   8461 taatacagca caaactaacc ctgattgtgc ctggctggaa gcacaagagg atgagaatgt
   8521 aggctttcca gtcaggccac aggtgccttt gagaccaatg acttataagg gagctgtcga
   8581 tctcagccac tttttaaaag aaaagggggg acTGgatggg ttaatttact ccaagaaaag
                                          -> 3′ LTR U3
   8641 acaagaaatc cttgatctgt gggtctataa cacacaagga ttcttcccag attggcagaa
   8701 ctacacacca gggccaggga ctagattccc actgaccttt gggtggtgct tcaaactagt
   8761 accaatggat ccagcagagg tagaggaagc caatgaagga gagaacaaca gtttattaca
   8821 ccccatctgt caacatggaa tggaggacga cgatagagaa gtgctgatct ggagatttga
   8881 cagtagactg gcattcagac acacagctag agagctgcat ccggagtact acaaagactg
   8941 cTGAcacaga agttgctgac aGGGACTTTC CactgGGGAC TTTCCgggga ggtgtggttt
          <- nef cds end       --------- -    ----- -----
                                   nfKb          nfKb
   9001 gggaggagtt ggggagtggc taaccctcag accgtgcaTA TAAgcagctg cttctcgcct
                                                 -- ---
                                              TATA signal
   9061 gtactgggtc tctcttgcta gaccagatct gagcctggga gctctctggc tagcggagga
          U3 /\ R
```

```
     9121 acccactgct taagcctcAA TAAAgcttgc cttgagtgct tcaagtagtg tgtgcccgtc
                                -- ----                  R /\ U5
                              poly-A signal
     9181 tgattgtgtg actctggtaa c
//
```