```
LOCUS       HIV93TH253        9720 bp    DNA             VRL       18-NOV-1996
DEFINITION  HIV-1 isolate 93th253 from Thailand, complete genome.
ACCESSION   U51189
KEYWORDS    .
SOURCE      Human immunodeficiency virus type 1.
  ORGANISM  Human immunodeficiency virus type 1 (HIV-1)
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9720)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     The heterosexual HIV-1 epidemic in Thailand is caused by an
            intersubtype (A/E) recombinant of African origin
  JOURNAL   J. Virology 70 (10), 7013-7029 (1996)
REFERENCE   2  (bases 1 to 9720)
  AUTHORS   Foley, Brian T.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1996) HIV Database, LANL, Mail Stop K710, Los
            Alamos, NM 87545, USA
COMMENT     Subtype E.  93TH253 has been previously named as CMU010 or
            302053.  It is from a Thai patient with end-stage AIDS.  The
            isolate has been adapted to growth in H9 cells.
            There are two frameshifts in 93TH253.   There is a G insertion at
            position 5798 which shifts the vpr gene.  The second is in vpu,
            where an A deletion occurs right after the ATG start codon.
FEATURES             Location/Qualifiers
     source          1..9720
                     /organism="Human immunodeficiency virus type 1"
                     /strain="93th253"
                     /chromosome="complete genome"
     LTR             1..633
                     /note="5’ LTR"
     misc_binding    363..372
                     /note="nf-kappa-b binding site"
     mRNA            455..9638
     repeat_region   455..551
     mRNA            455..9638
     misc_feature    763
                     /note="Description: major 5’ splice donor site"
     CDS             807..2279
                     /gene="gag"
                     /note="Description: Gag polyprotein gene; cleaved into
                     Capsid, Matrix and other proteins"
                     /codon_start=1
                     /product="Gag precursor polyprotein"
                     /translation="MGARASVLSGGKLDAWEKIRLRPGGKKKYKMKHLVWASRELERF
                     ALNPGLLETAEGCQQLIEQLQSTLKTGSEELKSLYNTIATLWCVHQRIEVKDTKEALD
                     KIEEVQKKSQQKKQQAAADTGSSSKVSQNYPIVQNAQGQMVHQPLSPRTLNAWVKVIE
                     EKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKETINEEAAEWDRVHPV
                     HAGPIPPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGDIYKRWIILGLNKIVRM
                     YQPVSILDIRQGPKEPFRDYVDRFYKVLRAEQATQEVKNWMTETLLVQNANPDCKSIL
                     KALGTGATLEEMMTACQGVGGPSHKARVLAEAMSQAQHANIMMQRGNFKGQTRIKCFN
                     CGKEGHLARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFPQSK
                     PEPTAPPAENWGMGEEQKDKEHPPPSVSLKSLFGNDPLSQ"
     CDS             join(807..2099,2098..5083)
                     /note="Gag-Pol fusion is caused by ribosomal slip on
                     ’tttttt’ slippery;sequence at 2096-2101, potentiated by
                     stem-loop structure at 2112-2135; Polyprotein cleaved into
```

```
                     protease, integrase, RNAse, reverse transcriptase etc"
                     /codon_start=1
                     /product="Gag-Pol fusion polyprotein"
                     /translation="MGARASVLSGGKLDAWEKIRLRPGGKKKYKMKHLVWASRELERF
                     ALNPGLLETAEGCQQLIEQLQSTLKTGSEELKSLYNTIATLWCVHQRIEVKDTKEALD
                     KIEEVQKKSQQKKQQAAADTGSSSKVSQNYPIVQNAQGQMVHQPLSPRTLNAWVKVIE
                     EKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKETINEEAAEWDRVHPV
                     HAGPIPPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGDIYKRWIILGLNKIVRM
                     YQPVSILDIRQGPKEPFRDYVDRFYKVLRAEQATQEVKNWMTETLLVQNANPDCKSIL
                     KALGTGATLEEMMTACQGVGGPSHKARVLAEAMSQAQHANIMMQRGNFKGQTRIKCFN
                     CGKEGHLARNCRAPRKKGCWKCGKEGHQMKDCTERQANFFFRENLAFQQREARKFSSE
                     QTRTNSPTSSRDLWDEGRDSLPSEAGAERQGPEPTFSFPQITLWQRPLVTVKIGGQLK
                     KALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVKQYDQILIEICGKKAIGTVLVG
                     PTPVNIIGRNMLTLIGCTLNFPISPISTVPVKLKPGMDGPRIKQWPLTEEKIKALTEI
                     CADMEREGRISKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIP
                     HPAGLKKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPSTNNETPGIRYQYNVLPQGW
                     KGSPAIFQASMTKILEPFRSKNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELREHLLK
                     WGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIELPEKESWTVNDIQKLVGKLNWASQ
                     IYAGIKVKQLCKLLRGTKALTDIVTLTEEAELELAENREILKDPVHGAYYDPSKDLIA
                     EIQKQGQDQWIYQIYQEPFKNLKTGKYARKRSAHTNDVKQLAEVVQKVVMESIVIWGK
                     TPKFKLPIQKETWETWWMDYWQAT$IPEWEFVNTPPLVKLWYQLEKDPIAGAETFYVD
                     GAANRETKLGKAGYVTDRGRQKVVSLTETTNQKTELHAIHLALQDSGSEVNIVTDSQY
                     ALGIIQAQPDRSESELVNQIIEKLIEKDKVYLSWVPAHKGIGGNEQVDKLVSSGIRKV
                     LFLDGIDKAQEEHERYHSNWRAMASDFNLPPIVAKEIVASCDKCQLKGEAMHGQVDCS
                     PGIWQLDCTHLEGKVILVAVHVASGYVEAEVIPAETGQETAYFLLKLAGRWPVKVVHT
                     DNGSNFTSAAVKAACWWANVKQEFGIPYNPQSQGVVESMNKELKKIIGQVREQAEHLK
                     TAVQMAVFIHNFKRKGGIGGYSAGERIIDIIASDLQTKELQKQITKIQKFRVCYRDSR
                     DPIWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKVKIIKDYGKQMAGDDCVAGRQDED"
stem_loop            2112..2135
                     /note="stem-loop potentiates ribosomal slip on tttttt
                     slippery sequence to produce Gag-Pol fusin polyprotein"
CDS                  5028..5606
                     /gene="vif"
                     /note="Description: Viral Infectivity factor"
                     /codon_start=1
                     /product="Vif protein"
                     /translation="MENRWQVMIVWQVDRMRIRTWNSLVKHHMYISKKAKQWFYRHHY
                     ESQHPKVSSEVHIPLGEARLVIRTYWGLQTGEKDWQLGHGVSIEWRQRKYSTQIDPDL
                     ADQLIHLQYFDCFSDSTIRRAILGQVVRRRCEYPSGHNKVGSLQYLALKALTTPKRIR
                     PPLPSVKKLTEDRWNKPQKIKGHRENPTMNGH"
misc_feature         5377
                     /note="3’ splice acceptor site"
misc_feature         5451
                     /note="5’ splice donor site"
CDS                  5546..5830
                     /gene="vpr"
                     /codon_start=1
                     /product="Vpr"
                     /note="G insertion at 5798 shifts frame for COOH-terminal
                     region of protein."
                     /translation="MEQAPEDQGSQREPYNEWTLELLEELKNEAVRHFPRPWLHGLGQ
                     YIYNNYGDTWEGVEAIIRILQQLLFVHFRIGCQHSRIGIFAREKRQEWSW"
CDS                  join(5818..6032,8363..8456)
                     /gene="tat"
                     /codon_start=1
                     /product="Tat"
                     /translation="MELVDPNLEPWNHPGSQPTTACSKCYCKKCCWHCQLCFLKKGLG
                     ISHGRKKRKHRRGTPQSRKDHQYPIPEQPLPIIRGGNPTDPKESKKEVASKAETDPCD"
CDS                  join(5957..6032,8363..8679)
```

```
                     /gene="rev"
                     /codon_start=1
                     /product="Rev"
                     /translation="MAGRSGSTDEELLRAVRIINILYQSNPYPSSEGGTRQTRKNRRR
                     RWRARQRQIRAISERILSTCLGRSTEPVPLQLPPLERLHLDCSEDCGTSGTQQSQGTE
                     TGVGRPQISGESSVILGPGTKN"
     intron          6033..8362
                      /note="Tat-Rev intron"
     CDS             join(6053..6057,6056..6297)
                     /gene="vpu"
                     /pseudo
                     /note="single base deletion, frameshift, near 6057 results
                     in premature termination"
                     /product="Vpu (defective)"
     CDS             6215..8803
                     /gene="env"
                     /pseudo
                     /note="premature stop codon at 8036-8038; tga here, tgg in
                     others"
                     /note="envelope glycoprotein, signal peptide, gp120 and
                     gp41"
                     /codon_start=1
                     /product="Envelope"
                     /translation="MRVKETQINWPNLWKWGTLILGLVIMCSASNNLWVTVYYGVPVW
                     RDADTTLFCASDAKAHETEVHNVWATHACIPTDPNPQEMHLENVTENFNMWKNNMVEQ
                     MQEDVISLWDQSLKPCVKLTPLCVTLNCTNANWTNANVTNVNNNVTNIVGNITEEVRN
                     CSFNMTTELIDKKQKVYALFYKLDIRQMNSNSSEYRLINCNTSVIKQACPKVSFDPIP
                     IHYCTPAGYAIIKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAEEKIII
                     RSENLTNNAKTIIVHLHESVEINCTRPFYNKRTRTSIGQGRVLYRTGDITGNIGKPYC
                     EINGTKWNKVLNQVTEKLKEHFNNRNISFQPPSGGDLEITMHHFICRGEFFYCNTTRL
                     FNNTCIGNKTMKECNDTIILPCKIKQIINMWQGVGQAMYNPPISGNINCVSNITGILL
                     TRDGGGGNGTNNEETFRPGGGNMKDNWRNELYKYKVVEIEPLGIAPTKAKRRVVEREK
                     RAVGIGAMIFGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHLLQL
                     TVWGIKQLQARVLAVERYLKDQKFLGLWGCSGKIICTTAVP"
     CDS             8805..9419
                     /gene="nef"
                     /codon_start=1
                     /product="Nef"
                     /translation="MGGKWSKSSIVGWPQVRERIKQTPPAAEGVGAVSQDLDKHGAVT
                     SSNMNNADCVWLRAQEEEGVGFPVRPQVPLRPMTYKGAFDLSFFLKEKGGLEGLVYSK
                     KRQEILDLWVYHTQGFFPDWHNYTPGPGIRYPLCFGWCFKLVPVDPREVEEDNKGENN
                     CLLHPMSQHGIEDEEREVLIWKFDSALARRHIARELRPEFYKDC"
     LTR             9088..9720
                     /note="3' LTR"
     repeat_region   9542..9638
     polyA_signal    9614..9619
     polyA_site      9638
     misc_feature    6053..6297
                     /gene="vpu (defective)"
                     /pseudo
                     /note="Description: vpu has 1 bp deletion frameshift near
                     6061"
BASE COUNT     3482 a   1711 c   2362 g   2165 t
ORIGIN
           1 tggaagggct agtttactcc aagaaaagac aagagatcct tgacttatgg gtctatcata
   5' LTR U3 ->
          61 cacaaggctt cttccctgat tggcataact acacaccagg gccagggatc agatatccac
         121 tgtgttttgg atggtgcttc aaactagtac cagttgaccc aagagaagta gaggaggaca
         181 acaaaggaga aaacaactgc ctgttacacc ccatgagcca gcatggaata gaggacgaag
```

```
 241 aaagagaagt gctgatatgg aagtttgaca gtgccctagc acgaagacac atagcccgag
 301 aactgcgtcc agagttctat aaagactgct gacaaagaag tttctaacta ggacttccgc
                                     <- nef cds end
 361 tggggacttt ccaggggagg tgtggccggg gcggagttgg ggagtggtta accctcagat
       -------- --- Nf Kappa B
 421 gctgcataaa agcagccgct tttcgcttgt actgggtctc tcttgttagg ccaggtcgag
         ------ --                LTR U3/\LTR R repeat
      TATA box signal
 481 cccgggagct ctctggctag caggggaacc cactgcttaa agcctcaata aagcttgcct
                                                          ---- --
                                                     poly-A signal
 541 tgagtgctta aagtggtgtg tgcccatctg tgttaggact ctggtaacta gagatccctc
    LTR R repeat/\LTR U5
 601 agatcactct agactgagta aaaatctcta gcagtggcgc ccgaacaggg acttgaaagc
                                      <- LTR U5
 661 gaaagttaat agggactcga aagcgaaagt tccagagaag ttctctcgac gcaggactcg
 721 gcttgctgag gtgcacacag caagaggcga gagcggcgac tggtgagtac gccaaatttt
                                                  /\ major 5’ sj
 781 gactagcgga ggctagaagg agagagatgg gtgcgagagc gtcagtatta agtgggggaa
                        gag cds ->
                   p17 MA start
 841 aattagatgc atgggaaaaa attcggttac ggccaggggg aaagaaaaaa tataaaatga
 901 aacatttagt atgggcaagc agagaattag aaagatttgc acttaaccct ggccttttag
 961 aaacagcaga aggatgtcaa caattaatag aacagttaca gtcaactctc aagacaggat
1021 cagaagaact taaatcatta tataatacaa tagcaaccct ctggtgcgta caccaaagga
1081 tagaggtaaa agacaccaag gaagctttag ataaaataga ggaagtacaa aagaaaagcc
1141 agcaaaagaa acagcaggca gcagctgaca caggaagcag cagcaaagtc agccaaaatt
1201 accctatagt gcaaaatgca caagggcaaa tggtacatca gcctttatca cctagaactt
gag p17 MA/\gag p24 CA
1261 tgaatgcatg ggtgaaagta atagaagaaa agggttttaa cccagaagta atacccatgt
1321 tctcagcatt atcagaggga gccaccccac aagatttaaa tatgatgcta aatatagtgg
1381 ggggacacca ggcagcaatg caaatgttaa aagaaaccat caatgaggaa gctgcagaat
1441 gggatagggt acacccagta catgcagggc ctattccacc aggtcagatg agggaaccaa
1501 ggggaagtga catagcagga actactagta cccttcaaga acaaatagga tggatgacaa
1561 acaatccacc tatcccagtg ggagacatct ataaaaggtg gataatcctg ggattaaata
1621 aaatagtaag aatgtatcaa cctgttagca ttttggacat aagacaaggg ccaaaagaac
1681 ccttcagaga ctatgtagat aggttctata aagttctcag agcggaacaa gccacacagg
1741 aggtaaaaaa ctggatgaca gaaaccttgc tagtccaaaa tgcgaatcca gactgtaagt
1801 ccattttaaa agcattagga acaggagcta cattagaaga aatgatgaca gcatgccagg
1861 gagtgggagg acctagccac aaggcaaggg ttttggctga ggcaatgagc caagcacaac
                                 gag p24 CA/\gag p2 X
1921 atgcaaatat aatgatgcag agaggcaatt ttaagggcca gacaagaatt aagtgcttca
           gag p2 X/\gag p7NC
                   \gag-pol TF
1981 actgtggcaa agaaggacac ctagccagaa attgcagggc ccctagaaaa aagggttgtt
2041 ggaaatgtgg aaaggaagga catcaaatga aagactgcac tgagagacag gctaattttt
                                                                  -----
                                                        gag-pol ribosomal
                                                        slip site
2101 tagggaaaat ttggccttcc aacaagggaa ggccagggaa ttttcctcag agcaaaccgg
     -           ^^^^^^^^^       ^^^^ ^^^^^   gag p7/\gag p6
                   stem    loop    stem
2161 agccaacagc tccgccagca gaaaactggg ggatggggga ggagcagaaa gacaaggaac
2221 atcctcctcc ttcagtttcc ctcaaatcac tctttggcaa cgacccccttg tcacaataaa
              gag-pol TF/\pol protease                    gag cds <-
                                                                  p6 end
2281 aataggagga caactgaaag aagctctatt agatacagga gcagatgata cagtattaga
2341 agatataaat ttgccaggga aatggaaacc aaaaatgata gggggaattg gaggttttat
2401 caaggtaagg caatatgatc agatacttat agaaatttgt ggaaaaaagg ctataggtac
```

```
    2461 agtattagta ggacctacac ctgtcaacat aattggacga aatatgttga ctcagattgg
    2521 ttgtactcta aatttcccaa ttagtcctat tgacactgta ccagtaaaat taaagccagg
             pol protease/\pol reverse transcriptase
    2581 aatggatgga ccaaaggtta aacagtggcc attgacagaa gaaaaaataa aagcattaac
    2641 agaaatttgt aaagagatgg aagaggaagg aaaaatctca aaaattggac ctgaaaatcc
    2701 atacaatact ccagtatttg ctataaagaa aaaggacagc accaaatgga ggaaattagt
    2761 agatttcaga gaactcaata aaagaactca ggacttttgg gaagttcaat taggaatacc
    2821 gcatccagca ggtttaaaaa agaaaaaatc agtaacagtg ctagatgtgg gagatgcata
    2881 tttttcagtt cctttagatg aaagctttag aaagtatact gcattcacca tacctagtat
    2941 aaacaatgag acaccaggaa tcagatatca gtacaatgtg ctgccacagg gatggaaagg
    3001 atcaccggca atattccaga gtagcatgac aaaaatctta gagcccttta gaataaaaaa
    3061 tccagaaatg gttatctatc aatacatgga tgacttgtat gtaggatctg atttagaaat
    3121 agggcagcac agaacaaaaa tagaggagct aagagctcat ctattgagct ggggatttac
    3181 tacaccagac aaaaagcatc agaaggaacc tccattcctt tggatgggat atgaactcca
    3241 tcctgacaga tggacagtcc agcctataga actgccagaa aaagacagct ggactgtcaa
    3301 tgatatacag aaattagtgg gaaaactgaa ttgggcaagt caaatttatg cagggattaa
    3361 ggtaaagcaa ctgtgtaaac tcctcagggg aactaaagca ctaacagata tagtaccact
    3421 gactgaagaa gcagaattag agttggaaga gaacagggag attctaagaa tccctgtgca
    3481 tggagtatat tatgacccat caaaagactt agtagcagaa gtacagaaac aagggcagga
    3541 ccaatggaca tatcaaattt atcaagagcc atttaaaaat ctaaaaacag gaaaatattc
    3601 cagaaaaagg tctgctcaca ctaatgatgt aagacaatta acagaagtgg tacaaaaaat
    3661 agccacagaa agcatagtaa tatggggaaa gacccctaaa tttagactac ccatacaaag
    3721 agaaacatgg gaaacatggt ggatggagta ttggcaggct acctggattc ctgaatggga
    3781 gtttgttaat acccctcctc tagtaaaatt atggtaccaa ttagaaaaag accccatagt
    3841 gggagcagag actttctatg tagatggggc agctagtagg gagactaagc taggaaaagc
pol reverse transcriptase/\pol RNAse
    3901 agggtatgtc actgacaggg gaagacaaaa ggtaatttcc ctaactgaga caacaaatca
    3961 aaagactgaa ttacatgcga tccatttagc cttgcaggat tcaggatcag aagtaaatat
    4021 agtaacagac tcacaatatg cattaggaat cattcaggca caaccagaca ggagtgaatc
    4081 agaagtagtc agccaaataa tagaggagct aataaaaaag gaaaaagtct acctgtcatg
    4141 ggtaccagca cacaagggga ttggaggaaa tgaacaagta gataaattag tcatttcagg
    4201 aatcaggaag gtactatttt tagatgggat aaataaggct caagaagaac atgaaagata
                pol RNAse/\pol integrase
    4261 tcacagcaat tggagaacaa tggctagtga ctttaatttg ccacctatag tagcaaagga
    4321 aatagtagcc aactgtgata aatgtcaact aaaaggggaa gctatgcatg gacaagtaga
    4381 ctgtagtcca gggatatggc aattagattg cacacatcta gaaggaaaag tcatcctggt
    4441 agcagtccac gtggccagtg gatatataga agcagaagtt atcccagcag aaacaggaca
    4501 ggagacagca tactttctgc taaaattagc aggaaggtgg ccagtaaaag taatacacac
    4561 agacaacggt agcaatttca ccagcgctgc agttaaagca gcctgttggt gggccaatgt
    4621 ccgacaggaa tttgggatcc cctacaatcc ccaaagtcaa ggagtagtag aatcaatgaa
    4681 taaggaatta aagaaaatca tagggcaggt aagggagcaa gctgaacacc ttaagacagc
    4741 agtacaaatg gcagtattca ttcacaattt taaaagaaaa ggggggattg gggggtacag
    4801 tgcaggggaa agaataatag acataatagc aacagacata caaactaaag aattacaaaa
    4861 acaaattaca aaaattcaaa attttcgggt ttattacagg gacagcagag acccaatttg
    4921 gaaaggacca gcaaaactac tctggaaagg tgaaggggca gtagtaatac aagacaatag
    4981 tgatataaaa gtagtaccaa gaagaaaagc aaagatcatt agggattatg gaaaacagat
                                                vif cds start ->
    5041 ggcaggtgat gattgtgtgg caggtagaca ggatgaggat tagaacatgg aacagtttag
    5101 taaaacacca catgtatatc tcaaagaaag ctaaacagtg gttttataga catcattatg
                                                              <- pol cds
                                                                 integrase end
    5161 aaagccagca tccaaaggtg agttcagaag tacatatccc actaggagaa gctagattag
    5221 taataagaac atattggggt ctgcagacag gagaaaagga ctggcaattg ggtcatggag
    5281 tctccataga atggaggcag agaaaatata gcacacaaat agatcctgac ctagcagacc
    5341 aactgattca tctacaatac tttgactgtt tttcagactc taccataagg agagccatat
    5401 taggacaagt agttagacgt aggtgtgaat atccatcagg acataacaag gtaggatccc
    5461 tacaatattt ggcactgaaa gcattaacaa caccaaaaag gataaggcca cctctgccta
    5521 gtgtaaagaa attaacagaa gatagatgga acaagcccca gaagatcaag ggtcacagag
                      vpr cds start ->
```

```
5581 agaaccctac aatgaatgga cattagaact gttagaggag cttaaaaatg aagctgttag
                                <- vif cds end
5641 acattttcct aggccctggc tccatggctt aggacagtac atctataaca attatgggga
5701 tacttgggaa ggggttgaag ctataataag aattttgcaa caactactgt ttgttcattt
5761 cagaattggg tgtcaacata gcagaatagg catttttgcc agggagaaga ggcaggaatg
                                                     tat cds start ->
5821 gagctggtag atcctaacct agagccctgg aatcatccgg gaagtcagcc tacaactgct
             <- vpr cds end
5881 tgtagcaagt gttactgtaa aaaatgttgc tggcattgcc aactatgctt tctgaaaaaa
5941 ggcttaggca tctcccatgg caggaagaag cggaagcacc gacgaggaac tcctcagagc
        rev cds start ->
6001 cgtaaggatc atcaatatcc tataccagag cagtaagtaa taagtatatg taatgcacct
                      tat, rev exon end/\intron start
                                                vpu cds start ->
6061 ttggaaatta gtgcaatagt aggactgtta gtagcgctaa tcttagcaat agtagtgtgg
        /\               <- premature termination of vpu
         1 bp deletion          1 bp deletion, frameshift
6121 actatagtag ctatagaatt taagaaaata ctaaggcaga gaaaaataga caggttagtt
6181 aagagaataa gagaaagaga agaagacagt ggaaatgaga gtgaaggaga cacagataaa
                                  env cds ->
                     signal peptide start
6241 ttggccaaac ttgtggaaat gggggacttt gatccttggg ttggtgataa tgtgtagtgc
                                                 env signal peptide/\gp120
6301 ctcaaacaac ttgtgggtta cagtttacta tggggttcct gtgtggagag atgcagatac
6361 caccctattt tgtgcatcag atgccaaagc acatgagaca gaagtgcaca atgtctgggc
6421 cacacatgcc tgtataccca cagaccccaa cccacaagaa atgcacctgg aaaatgtaac
6481 agaaaatttt aacatgtgga aaaataacat ggtagagcag atgcaggagg atgtaatcag
6541 tttatgggat caaagtctaa agccatgtgt aaagttaact cctctctgcg ttactttaaa
6601 ttgtaccaat gctaattgga ccaatgctaa tgtgaccaat gtcaataaca atgtcactaa
6661 catagtagga aatataacag aggaagtaag aaactgctct tttaatatga ccacagaact
6721 aatagataag aagcagaagg tctatgcact tttttataag cttgatataa gacaaatgaa
6781 tagtaatagt agtgagtata ggttaataaa ttgtaatact tcagtcatta agcaggcttg
6841 tccaaaggta tcctttgacc caattcctat acattattgt actccagctg gttatgcgat
6901 tataaagtgt aatgataaga atttcaatgg gacagggcca tgtaaaaatg tcagctcagt
6961 acaatgcaca catggaatta agccagtagt atcaactcaa ttgctgttaa atggcagtct
7021 agcagaagaa aagataataa tcagatctga aaatctcaca aacaatgcca aaaccataat
7081 agtgcacctt catgaatctg tagaaatcaa ttgtaccaga cccttctaca ataaaagaac
                         V3 loop start ->
7141 aaggacatct ataggacaag gacgagtact ctatagaaca ggagacataa caggaaatat
7201 aggaaaacca tattgtgaga ttaatggaac aaaatggaat aaagttttaa atcaggtaac
                    <- V3 loop end
7261 tgaaaaatta aaagagcact ttaataatag gaacataagc tttcagccac catcaggagg
7321 agatctagaa attacaatgc atcattttat ttgtagaggg gaatttttct attgcaatac
7381 aacacgactg tttaataata cttgcatagg aaataaaacc atgaaggagt gtaatgacac
7441 tatcatactt ccatgcaaga taaagcaaat tataaacatg tggcagggag taggacaagc
7501 aatgtataat cctcccatca gtggaaacat taattgtgta tcaaatatta caggaatact
7561 attgacaaga gatggtggtg gtggtaatgg tacgaataat gaggagacct tcagacctgg
7621 aggagggaac atgaaggaca attggagaaa tgaattatat aaatataaag tagtagaaat
7681 tgaaccacta ggaatagcac ccaccaaggc aaagagaaga gtggtggaga gagaaaaaag
7741 agcagtggga ataggagcta tgatctttgg gttcttagga gcagcaggaa gcactatggg
env gp120/\env gp41
7801 cgcggcgtca ataacgctga cggtacaggc cagacaatta ttgtctggta tagtgcaaca
7861 gcaaagcaat ttgctgaggg ctatagaggc gcagcagcat ctgttgcaac tcacagtctg
7921 gggcattaaa cagctccagg caagagtcct ggctgtggaa agatacctaa aagatcaaaa
7981 gttcctagga ctttggggct gctctggaaa aatcatctgc accactgctg tgccctgaaa
8041 ctccacttgg agtaatagat cttatgaaga gatttgggac aacatgacat ggatagaatg
8101 ggagagagaa attagcaatt acacaaacca aatatatgag atacttacag aatcgcagaa
8161 ccagcaggaa agaaatgaaa aggacttgtt agaattggat aaatgggcaa gtctgtggaa
8221 ttggtttggc ataacaaaat ggctgtggta tataaaaata tttataatga tagtaggagg
```

```
        8281 tttaataggt ttaagaataa tttttgctgt gctttctata gtaaatagag ttaggcaggg
        8341 atactcacct ttgtctttcc agacccctac ccatcatcag aggggggaac ccgacagacc
                 tat, rev intron end/\exon start
        8401 cgaaagaatc gaagaaggag gtggcgagca aggcagagac agatccgtgc gattagtgag
                                                                      <- tat
   cds end
        8461 cggattctta gcacttgcct gggacgatct acggagcctg tgcctcttca gctaccaccg
        8521 cttgagagac ttcatcttga ttgcagcgag gactgtggaa cttctgggac acagcagtct
        8581 caagggactg agacgggggt gggaaggcct caaatatctg gggaatcttc tgttatattg
        8641 gggccaggaa ctaaaaatta gtgctatttc tttgtttgat gctttagcag tagtggtagc
                                <- rev cds end
        8701 ggggtggaca gatagggtta tagaagtagc acaaggagct tggagagcca ttctccacat
        8761 acctagaaga atcagacagg gcttagaaag ggctttgcta taacatggga ggcaagtggt
                                                        <- env cds
                                                           gp41 end
                                                 nef cds start ->
        8821 caaaaagtag catagtggga tggcctcagg tcagggaaag aataaagcaa actcctccag
        8881 cagcagaagg agtaggagca gtatctcaag atctagataa acatggagca gtaacaagta
        8941 gtaatatgaa taatgctgat tgtgtctggc tgagagcaca agaggaagag ggggtaggct
        9001 ttccagtcag gccgcaggta cctctaagac caatgactta taagggagct tttgatctta
        9061 gcttcttttt aaaagaaaag gggggactgg aagggctagt ttactccaag aaaagacaag
                  U3 region start 3’ LTR ->
        9121 agatccttga cttatgggtc tatcatacac aaggcttctt ccctgattgg cataactaca
        9181 caccagggcc agggatcaga tatccactgt gttttggatg gtgcttcaaa ctagtaccag
        9241 ttgacccaag agaagtagag gaggacaaca aaggagaaaa caactgcctg ttacacccca
        9301 tgagccagca tggaatagag gacgaagaaa gagaagtgct gatatggaag tttgacagtg
        9361 ccctagcacg aagacacata gcccgagaac tgcgtccaga gttctataaa gactgctgac
                                                                        <-
                                                                        nef
   cds end
        9421 aaagaagttt ctaactagga cttccgctgg ggactttcca ggggaggtgt ggccggggcg
                                            -- --------- Nf-Kappa B
        9481 gagttgggga gtggttaacc ctcagatgct gcataaaagc agccgctttt cgcttgtact
        9541 gggtctctct tgttaggcca ggtcgagccc gggagctctc tggctagcag gggaacccac
LTR U3 region/\LTR repeat region
        9601 tgcttaaagc ctcaataaag cttgccttga gtgcttaaag tggtgtgtgc ccatctgtgt
                            ------        LTR R repeat/\LTR U5 region
                         poly-A signal
        9661 taggactctg gtaactagag atccctcaga tcactctaga ctgagtaaaa atctctagca
                                                                         <-
                                                                         3’LTR
                                                                         U5 end
```

CM240

```
LOCUS       HIVCM240    9203 bp    DNA             VRL       13-SEP-1996
DEFINITION  Human immunodeficiency virus type 1, strain CM240, complete
            proviral genome.
ACCESSION   U54771
SOURCE      Human immunodeficiency virus type 1 (HIV-1), strain CM240; PBMCs
            extracted from male patient from Thailand; proviral DNA amplified;
            complete genomic sequence determined from three clones of viral genome.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9203)
  AUTHORS   Carr,J.K., Salminen,M.O., Koch,C., Gotte,D., Artenstein,A.W.,
            Hegerich,P.A., St. Louis,D., Burke,D.S. and McCutchan,F.E.
  TITLE     Full-length sequence and mosaic structure of a human
            immunodeficiency virus type 1 isolate from Thailand
  JOURNAL   J. Virol. 70 (9), 5935-5943 (1996)
REFERENCE   2  (bases 1 to 9203)
  AUTHORS   Carr,J.K., Salminen,M.O., Gotte,D.R., Artenstein,A.W.,
            Hegerich,P.A., St., Louis,D., Burke,D.S. and McCutchan,F.E.
   TITLE    Direct Submission
 JOURNAL    Submitted (11-APR-1996) J.K. Carr, Molecular Epidemiology, Henry M.
            Jackson, 1600 E. Gude Dr., Rockville, MD 20850, USA
 COMMENT    Blood from an asymptomatic 21-year-old Thai man was transported
            from Chiang Mai to the USA where PBMCs were separated and
            co-cultivated with PHA-stimulated donor PBMCs.  DNA from p24
            antigen-positive culture was used to amplify the proviral DNA.
            The complete genomic sequence of the provirus was determined by the
            compilation of three clones containing different parts of the viral
            genome.  CM240 is an example of a Thai E virus, which is a mosaic of
            a clade A virus and a clade E virus.
FEATURES             Location/Qualifiers
     source          1..9203
                     /organism="Human immunodeficiency virus type 1"
                     /proviral
                     /strain="CM240"
                     /note="Blood from an asymptomatic young Thai man was
                     transported from Chiang Mai to the USA where PBMCs were
                     separated and co-cultivated with PHA-stimulated donor
                     PBMCs.  DNA from p24 antigen-positive culture was used to
                     amplify the proviral DNA.  The complete genomic sequence
                     of the provirus was determined by the compilation of three
                     clones containing different parts of the viral genome.
                     Once the sequence of the provirus was determined the
                     genome was permuted to the organizational structure of the
                     viral genomic RNA; CM240 is an example of a Thai E virus,
                     which is a mosaic of a clade A virus and a clade E virus.
                     See also GenBank accession numbers L14572 and L11761."
     source          join(1..180,8986..9203)
                     /organism="Human immunodeficiency virus type 1"
                     /clone="1"
     source          181..8369
                     /organism="Human immunodeficiency virus type 1"
                     /clone="2"
     CDS             355..1866
                     /gene="gag"
                     /codon_start=1
                     /product="gag"
                     /db_xref="PID:g1537051"
                     /translation="MGARASVLSGGKLDAWEKIRLRPGGRKKYRLKHLVWASRELERF
```

```
                ALNPSFLETAEGCQQIIEQLQSTLKTGLEELKSLFNTVATLWCVHQRIEVKDTKEALD
                KIEEVQNKSQRKTQQAAAGTGSSSKVSQNYPIVQNAQGQMAHQPLSPRTLNAWVKVVE
                EKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKETINEEPAEWDRVHPV
                HAGPIPPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGDIYKRWIILGLNKIVRM
                YSPVSILDIRQGPKEPFRDYVDRFYKTLRAEQATQEVKNWMTETLLVQNANPDCKSIL
                KALGTGATLEEMMTACQGVGGPSHKARVLAEAMSHAQHATIMMQRGNFKGQKRIKCFN
                CGREGHLARNCRAPRKQGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFPQSR
                PEPTAPPAENWGMGEEITGEEITSLPKQEQKDKEHPPPLVSLKSLFGNDPLSQ"
CDS             <1644..4670
                /gene="pol"
                /note="Method: conceptual translation supplied by author"
                /codon_start=1
                /product="pol"
                /db_xref="PID:g1537052"
                /translation="FFRENLAFQQGKAGEFSSEQTRANSPTSRKLGDGGRDNGGRDNL
                LTEAGAERQGTSSSFSFPQITLWQRPLVTVKIGGQLKEALLDTGADDTVLEDINLPGK
                WKPKMIGGIGGFIKVKQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNF
                PISPIDTVPVTLKPGMDGPKVKQWPLTEEKIKALTEICKEMEEEGKISKIGPENPYNT
                PVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYF
                SVPLDESFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRIK
                NPEMVIYQYKDDLYVGSDLEIGQHRTKIEELRAHLLSWGFTTPDKKHQKEPPFLWMGY
                ELHPDRWTVQPIELPEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGAKALT
                DIVPLTEEAELELAENREILKTPVHGVYYDPSKDLVAEVQKQGQDQWTYQIYQEPFKN
                LKTGKYARRGSAHTNDVRQLTEVVQKVATESIVIWGKTPKFRLPIQRETWETWWMEYW
                QATWIPEWEFVNTPPLVKLWYQLEKDPIVGAETFYVDGAASRETKLGKAGYVTDRGRQ
                KVVSLTETTNQKTELHAIHLALQDSGSEVNIVTDSQYALGIIQAQPDRSESEVVNQII
                EELIKKEKVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHERYHSNWR
                TMASDFNLPPIVAKEIVTNCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVH
                VASGYIEAEVIPAETGQETAYFLLKLAGRWPVKVIHTDNGSNFTSAAVKAACWWANVQ
                QEFGIPYNPQSQGVVESMNKELKKIIGQVREQAEHLKTAVQMAVFIHNFKRKGGIGGY
                SAGERIIDIIATDIQTKELQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQ
                DNSDIKVVPRRKAKIIRDYGKQMAGDDCVAGRQDED"
CDS             4615..5187
                /gene="vif"
                /codon_start=1
                /product="vif"
                /db_xref="PID:g1537053"
                /translation="MENRWQVMIVWQVDRMRIRTWNSLVKHHMYISKKAKKWFYRHHY
                ESQHPKVSSEVHIPLGEARLVIRTYWGLQTGEKDWHLGHGVSIEWRQRKYSTQIDPDL
                ADKLIHLQYFGCFSDSAIRKAILGQVVRRRCEYPSGHNKVGSLQYLALKALTTPKRIR
                PPLPSVEITEDRWNKPQKRGHRENPTMNGH"
CDS             5130..5417
                /gene="vpr"
                /codon_start=1
                /product="vpr"
                /db_xref="PID:g1537054"
                /translation="MEQAPEEGPQREPYNEWTLELLEELKNEAVRHFPRPWLQGLGQY
                IYNNYGDTWEGVEAIIRMLQQLLFVHFRIGCQHSRIGIMPGRRGRNGTGRS"
CDS             join(5398..5612,7932..8022)
                /gene="tat"
                /codon_start=1
                /product="tat"
                /db_xref="PID:g1537055"
                /translation="MELVDPNLEPWNHPGSQPTTACSKCYCKKCCWHCQLCFLKKGLG
                ISYGRKKRKHRRGTPQSSKDHQNPIPKQPLPIIRRNPTDPKESKKEVASKAETDQCD"
CDS             join(5537..5612,7932..8227)
                /gene="rev"
                /codon_start=1
                /product="rev"
```

```
                     /db_xref="PID:g1537056"
                     /translation="MAGRSGSTDEELLRAVRIIKILYQSNPFPSSEGTRQTRRNRRRR
                     WRARQRQISAISERILSTCLGRSTEPVPLQLPPVERLNLDCSEDGGTSGTQQSQGTET
                     GVGRPQISGESSVILGPGTKN"
     CDS             5633..5701
                     /gene="vpu"
                     /note="this clone has a premature translation stop in vpu"
                     /codon_start=1
                     /product="truncated vpu"
                     /db_xref="PID:g1537057"
                     /translation="MTPLEISAIVGLIVALILAIVV"
     CDS             5796..8369
                     /gene="env"
                     /codon_start=1
                     /product="env"
                     /db_xref="PID:g1537058"
                     /translation="MRVKETQMNWPNLWKWGTLILGLVIICSASDNLWVTVYYGVPVW
                     RDADTTLFCASDAKAHETEVHNVWATHACVPTDPNPQEIHLENVTENFNMWKNNMVEQ
                     MQEDVISLWDQSLKPCVKLTPLCVTLNCTNANLTNGSSKTNVSNIIGNITDEVRNCTF
                     NMTTELTDKKQKVHALFYKLDIVQIEDKKTSSEYRLINCNTSVIKQACPKISFDPIPI
                     HYCTPAGYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAEEEIIIR
                     SEDLTNNAKTIIVHLNKSVEINCTRPSNNTRTSITIGPGRVFYRTGDIIGNIRKAYCE
                     INGTKWNKVLKQVTEKLKEHFNKTIIFQPPSGGDLEITMHHFNCRGEFFYCNTTKLFN
                     NTCLGNETMAGCNDTITLPCKIKQIINMWQGAGQAMYAPPISGRINCVSNITGILLTR
                     DGGVNNTDNETFRPGGGNIKDNWRSELYKYKVVQIEPLGIAPTRAKRRVVEREKRAVG
                     IGAMIFGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHLLQLTVWG
                     IKQLQARVLAVERYLKDQKFLGLWGCSGKIICTTAVPWNSTWSNRSFEEIWNNMTWIE
                     WEREISNYTNQIYEILTESQNQQDRNEKDLLELDKWASLWNWFDITNWLWYIKIFIMI
                     VGGLIGLRIIFAVLSIVNRVRQGYSPLSFQTPSHHQKEPDRPEGIEEGGGEQGRDRSV
                     RLVSGFLALAWDDLRSLCLFSYHRLRDLTLIAARTVELLGHSSLKGLRRGWEGLKYLG
                     NLLLYWGQELKISAISLLDATAIAAAGWTDRVIEVAQGAWRAILHIPRRIRQGLERTL
                     L"
     source          8370..8985
                     /organism="Human immunodeficiency virus type 1"
                     /clone="3"
     CDS             8371..8985
                     /gene="nef"
                     /codon_start=1
                     /product="nef"
                     /db_xref="PID:g1537059"
                     /translation="MGSKWSKSSIVGWPQVREKIKQTPPATEGVGAVSQDLDKHGAIT
                     SSNIDNADCVWLRAQEDEEVGFPVMPQVPLRPMTYKGAFDLSFFLKEKGGLDGLIYSK
                     KRQEILDLWVYNTQGFFPDWQNYTPGPGIRFPLCFGWCFKLVPVDQREVEEDNKGENN
                     CLLHPMSQHGIEDEEREVLMWKFDSALARKHVAREQHPEYYKDC"
     misc_signal     9014..9024
                     /note="NFkappaB"
     misc_signal     9029..9038
                     /note="sp1.3"
     misc_signal     9040..9049
                     /note="sp1.2"
     misc_signal     9051..9061
                     /note="sp1.1"
     TATA_signal     9080..9084
BASE COUNT     3336 a   1622 c   2220 g   2025 t
ORIGIN
        1 GGgtctctct tgttagacca ggtcgagccc gggagctctc tggctagcaa gggaacccac
R rpt start->
 mRNA start->
```

```
  61 tgcttaaagc ctcaataaag cttgccttga gtgcttaaag tggtgtgtgc ccgtctgtgt
                   ------            R repeat/\LTR U5
               poly-A signal
 121 taggactctg gtaactagag atccctcaga ccactctaga ctgagtaaaa atctctacca
                                                                  <- 5′ LTR
 181 gtggcgcccg aacagggcac tcgaaagcga aagttaatag ggactcgaaa gcgaaagttc
 241 cagagaagtt ctctcgacgc aggactcggc ttgctgaggt gcacacagca agaggcgaga
 301 gcggcgactg gtgagtacgc caaattttga ctagcggagg ctagaaggag agagatgggt
               /\major 5′ sj                           gag cds ->
                                          p17 MA (matrix) start
 361 gcgagagcgt cagtattaag tgggggaaaa ttagatgcat gggaaaaaat tcggttgcgg
 421 ccagggggaa gaaaaaaata taggctgaaa catttagtat gggcaagcag agagttagaa
 481 agattcgcac ttaaccctag ctttttagaa acagcagaag gatgtcaaca aataatagaa
 541 cagttacagt caactctcaa gacaggatta gaagaactta aatcattatt taatacagta
 601 gcaaccctct ggtgcgtaca ccaaaggata gaggtaaaag acaccaagga agctttagat
 661 aaaatagagg aagtacaaaa taagagccag cgaaagacac agcaggcagc agctggcaca
 721 ggaagcagca gcaaagtcag ccaaaattac cctatagtgc aaaatgcaca agggcaaatg
                         gag p17 MA/ \gag p24 CA
 781 gcacatcagc ctttatcacc tagaactttg aatgcatggg tgaaagtagt agaagaaaag
 841 ggttttaacc cagaagtaat acccatgttc tcagcattat cagagggagc caccccacaa
 901 gatttaaata tgatgctaaa tatagtgggg ggacaccagg cagcaatgca aatgttaaaa
 961 gaaaccatca atgaggaacc tgcagaatgg gatagggtac acccagtaca tgcagggcct
1021 attccaccag gccagatgag ggaaccaagg ggaagtgaca tagcaggaac tactagtacc
1081 cttcaagaac aaataggatg gatgacaaac aatccaccca tcccagtggg agacatctat
1141 aaaaggtgga taatcctggg attaaataaa atagtaagaa tgtatagccc tgttagcatt
1201 ttggacataa gacaagggcc aaaagaaccc ttcagagact atgtagatag gttctataaa
1261 actctcagag cggaacaagc tacacaggag gtaaaaaact ggatgacaga aaccttgcta
1321 gtccaaaacg cgaatccaga ctgtaagtcc attttaaaag cattaggaac aggagctaca
1381 ttagaagaaa tgatgacagc atgccaggga gtgggaggac ctagccataa agcaagggtt
1441 ttggctgaag caatgagcca cgcacaacat gcaactataa tgatgcagag aggcaatttc
gag p24 CA/\gag p2x                          gag p2x/\gag p7NC
                                                        \gag-pol TF
1501 aagggccaga aaagaattaa gtgcttcaac tgtggtagag aaggacacct agccagaaat
1561 tgcagggccc ctagaaaaca gggttgttgg aaatgcggga aggaaggaca tcaaatgaaa
1621 gactgcactg agagacaggc taatttttta gggaaaattt ggccttccaa caagggaagg
                              ------             ^^^^^^^^       ^^^^^^
                  gag-pol  ribosomal slip site      stem   loop    stem
1681 ccggggaatt ttcctcagag cagaccagag ccaacagccc caccagcaga aaactggggg
     ^^  gag p7NC/\gag p6
1741 atgggggaag agataacggg ggaagagata acctccttac cgaagcagga gcagaaagac
1801 aaggaacatc ctcctccttt agtttccctc aaatcactct ttggcaacga ccccttgtca
               gag-pol fusion TF/\pol protease
1861 cagtaaaaat aggaggacag ctgaaagaag ctctattaga tacaggagca gatgatacag
         <- gag cds
            p6 end
1921 tattagaaga tataaatttg ccaggaaaat ggaaaccaaa aatgataggg ggaattggag
1981 gttttatcaa ggtaaagcaa tatgatcaga tacttataga aatctgtgga aaaaaggcta
2041 taggtacagt attagtagga cctacacctg tcaacataat tggacgaaat atgttgactc
2101 agattggttg tactttaaat ttcccaatta gtcctattga cactgtacca gtaacattaa
                  pol protease/\pol reverse transcriptase
2161 agccaggaat ggatggacca aaggttaaac agtggccatt gacagaagaa aaaataaaag
2221 cattaacaga aatttgtaaa gagatggaag aggaaggaaa aatctcaaaa attgggcctg
2281 aaaatccata caatactcca gtatttgcta taaagaaaaa ggacagcacc aaatggagga
2341 aattagtaga tttcagagag ctcaataaaa gaactcagga cttttgggaa gttcaattag
2401 gaataccgca tccagcaggt ttaaaaaaga aaaaatcagt aacagtacta gatgtgggag
2461 atgcatattt ttcagttcct ttagatgaaa gctttagaaa gtatactgca ttcaccatac
2521 ctagtataaa caatgagaca ccaggaatca gatatcagta caatgtgctg ccacagggat
2581 ggaaaggatc accggcaata ttccagagta gcatgacaaa aatcttagag ccctttagaa
2641 taaaaaatcc agaaatggtt atctatcaat acaaggatga cttgtatgta ggatctgatt
```

```
2701 tagaaatagg gcagcacaga acaaaaatag aggagctaag agctcatcta ttgagctggg
2761 gatttactac accagacaaa aagcatcaga aggaacctcc attcctttgg atgggatatg
2821 aactccatcc tgacagatgg acagtccagc ctatagaact gccagaaaaa gacagctgga
2881 ctgtcaatga tatacagaaa ttagtgggaa aactaaattg ggcaagtcaa atttatgcag
2941 ggattaaggt aaagcaactg tgtaaactcc tcaggggagc taaagcacta acagacatag
3001 taccactgac tgaagaagca gaattagagt tggcagagaa cagggagatt ctaaaaaccc
3061 ctgtgcatgg agtatattat gacccatcaa aagacttagt agcagaagta cagaaacaag
3121 ggcaggacca atggacatat caaatttatc aagagccatt taaaaatcta aaaacaggaa
3181 aatatgccag aagagggtct gctcacacta atgatgtaag acaattaaca gaagtggtgc
3241 aaaaagtagc cacagaaagc atagtaatat ggggaaagac ccctaaattt agactaccca
3301 tacaaagaga aacatgggaa acatggtgga tggagtattg gcaggctacc tggattcctg
3361 aatgggagtt tgttaatacc cctcctctag taaaattatg gtaccaatta gaaaaagacc
3421 ccatagtagg agcagagact ttctatgtag atggggcagc tagtagggag actaagctag
    pol reverse transcriptase/\pol RNAse
3481 gaaaagcagg gtatgtcact gacagaggaa gacaaaaggt agtttcccta actgagacaa
3541 caaatcaaaa gactgaatta catgcgatcc atttagcctt gcaggattca ggatcagaag
3601 taaatatagt aacagactca caatatgcat taggaatcat tcaggcacaa ccagacagga
3661 gtgaatcaga agtagtcaac caaataatag aggagctaat aaaaaaggag aaagtctacc
3721 tgtcatgggt accagcacac aaggggattg gaggaaatga acaagtagat aaattagtca
3781 gttcaggaat caggaaggtg ctatttttag atgggataga taaggctcaa gaagaacatg
                    pol RNAse/\pol integrase
3841 aaagatatca cagcaattgg agaacaatgg ctagtgattt taatttgcca cctatagtag
3901 caaaggaaat agtaaccaac tgtgataaat gtcaactaaa aggggaagct atgcatggac
3961 aagtagactg tagtccaggg atatggcaat tagattgcac acatctagaa ggaaaagtca
4021 tcctggtagc agtccacgtg gccagtggat atatagaagc agaagttatc ccagcagaaa
4081 caggacagga gacagcatac tttctgctaa aactagcagg aagatggcca gtaaaagtaa
4141 tacacacaga caacggtagc aatttcacca gcgctgcagt taaagcagcc tgttggtggg
4201 ccaatgtcca acaggaattt gggatcccct acaatcccca aagtcaagga gtagtagaat
4261 ctatgaataa ggaattaaag aaaatcatag ggcaggtaag agagcaagct gaacacctta
4321 aaacagcagt acaaatggca gtattcattc acaattttaa aagaaaaggg gggattgggg
4381 ggtacagtgc aggggaaaga ataatagaca taatagcaac agacatacaa actaaagaat
4441 tacaaaaaca aattacaaaa attcaaaatt ttcgggttta ttacagggac agcagagacc
                                                     /\ 3′ sj
4501 caatttggaa aggaccagca aaactactct ggaaaggtga aggggcagta gtaatacaag
                                          /\ 5′ sj
4561 acaatagtga tataaaagta gtaccaagaa gaaaagcaaa gatcattagg gattatggaa
                                                 vif cds start ->
4621 aacagatggc aggtgatgat tgtgtggcag gtagacagga tgaggattag aacatggaac
                                                        <- pol cds
                                                           integrase end
4681 agtttagtaa aacatcatat gtatatctca aagaaagcta aaaagtggtt ttatagacat
4741 cattatgaaa gccagcatcc aaaggtaagc tcagaagtac atatcccact aggagaggct
4801 agattagtaa taagaacata ttggggtctg caaacaggag aaaaggactg gcacttgggt
4861 catggagtct ccatagaatg gaggcagaga aaatatagca cacaaataga tcctgaccta
4921 gcagacaaac tgattcatct acaatatttt ggctgttttt cagactctgc cataaggaaa
                                                   /\ 3′ sj
4981 gccatattag gacaagtagt tagacgtagg tgtgaatatc catcaggaca taacaaggta
                                                                 /\ 5′sj
5041 ggatccctac aatatttggc actgaaagca ttaacaacac caaaaaggat aaggccacct
5101 ctgcctagtg tagaaataac agaagataga tggaacaagc cccagaagag gggccacaga
                     vpr cds start - >
5161 gagaacccta caatgaatgg acattagaac tgttagagga gcttaaaaat gaagctgtta
                                <- vif cds end
5221 gacattttcc taggccctgg ctccaaggct taggacagta catctataac aattatgggg
5281 atacttggga aggggttgaa gctataataa gaatgttgca acaactactg tttgttcatt
5341 tcagaattgg gtgtcaacat agcagaatag gcattatgcc agggagaaga ggcaggaatg
        /\ 3′ sj                                     tat cds start ->
5401 gaactggtag atcctaacct agagccctgg aatcatccgg gaagtcagcc tacaactgct
                    <- vpr cds end
```

```
5461 tgtagcaagt gttactgtaa aaaatgttgc tggcattgcc aactatgctt tctgaaaaaa
5521 ggcttaggca tctcctatgg caggaagaag cggaagcacc gacgaggaac tcctcagagc
        rev cds start ->     /\ 3' sj
5581 agtaaggatc atcaaaatcc tataccaaag cagtaagtaa taagtatatg taatgacacc
                  (tat,rev,nef) 5' sj /\       vpu cds start ->
5641 tttggaaatt agtgcaatag taggactgat agtagcgcta atcttagcaa tagtagtgtg
5701 aactatagta gctatagaag ttaagaaaat actaaggcaa agaaaaatag acaggttagt
    <- vpu cds premature end
5761 taagagaata agagaaagag cagaagacag tggaaatgag agtgaaggag acacagatga
                                  env cds ->
                      signal peptide start
5821 attggccaaa cttgtggaaa tgggggactt tgatccttgg gttggtgata atttgtagtg
                                                 env signal peptide/\gp120
                                                                <- vpu cds
                                                                   normal end
5881 cctcagacaa cttgtgggtt acagtttatt atggggtgcc tgtgtggaga gatgcagata
5941 ccaccctatt ttgtgcatca gatgccaagg cacatgagac agaagtgcac aatgtctggg
6001 ccacacatgc ctgtgtaccc acagacccca acccacaaga aatacacctg gaaaatgtaa
6061 cagaaaattt taacatgtgg aaaaataaca tggtagagca gatgcaggag gatgtaatca
6121 gtttatggga tcaaagtcta aagccatgtg taaagttaac tcctctctgc gttactttaa
6181 attgtaccaa tgctaatttg accaatggca gtagcaaaac caatgtctct aacataatag
6241 gaaatataac agatgaagta agaaactgta cttttaatat gaccacagaa ctaacagata
6301 agaagcagaa ggtccatgca ctcttttata agcttgatat agtacaaatt gaagataaga
6361 agactagtag tgagtatagg ttaataaatt gtaatacttc agtcattaag caggcttgtc
6421 caaagatatc ctttgatcca attcctatac attattgtac tccagctggt tatgcgattt
6481 taaagtgtaa tgataagaat ttcaatggga cagggccatg taaaaatgtc agctcagtac
6541 aatgcacaca tggaattaag ccagtggtat caactcaatt gctgttaaat ggcagtctag
6601 cagaagaaga gataataatc agatctgaag atctcacaaa caatgccaaa accataatag
6661 tgcaccttaa taaatctgta gaaatcaatt gtaccagacc ctccaacaat acaagaacaa
                      V3 loop start - >
6721 gtataactat aggaccagga cgagtattct atagaacagg agatataata ggaaatataa
6781 gaaaagcata ttgtgagatt aatggaacaa aatggaataa agttttaaaa caggtaactg
                  <- V3 loop end
6841 aaaaattaaa agagcacttt aataagacaa taatctttca accaccctca ggaggagatc
6901 tagaaattac aatgcatcat tttaattgta gaggggaatt tttctattgc aatacaacaa
6961 aactgtttaa taatacttgc ctaggaaatg aaaccatggc ggggtgtaat gacactatca
7021 cacttccatg caagataaag caaattataa acatgtggca gggagcagga caagcaatgt
7081 atgctcctcc catcagtgga agaattaatt gtgtatcaaa tattacagga atactattga
7141 caagagatgg tggtgttaat aatacggata acgagacctt cagacctgga ggaggaaaca
7201 taaaggacaa ttggagaagt gaattatata aatataaagt agtacaaatt gaaccactag
7261 gaatagcacc caccagggca aagagaagag tggtggagag agaaaaaagg gcagtgggaa
                                                 env gp120/ \env gp41
7321 taggagctat gatctttggg ttcttaggag cagcaggaag cactatgggc gcggcgtcaa
7381 taacgctgac ggtacaggcc agacaattat tgtctggtat agtgcaacag caaagcaatt
7441 tgctgagggc tatagaggcg cagcagcatc tgttgcaact cacagtctgg ggcattaaac
7501 agctccaggc aagagtcctg gctgtggaaa gatacctaaa ggatcaaaag ttcctaggac
7561 tttggggctg ctctggaaaa atcatctgca ccactgctgt gccctggaac tccacttgga
7621 gtaatagatc ttttgaagag atttggaaca acatgacatg gatagaatgg gagagagaaa
7681 ttagcaatta cacaaaccaa atatatgaga tacttacaga atcgcagaac caacaggaca
7741 ggaatgaaaa ggatttgtta gaattggata aatgggcaag tctgtggaat tggtttgaca
7801 taacaaattg gctgtggtat ataaaaatat ttataatgat agtaggaggt ttaataggtt
7861 taagaataat ttttgctgtg ctttctatag taaatagagt taggcaggga tactcacctt
7921 tgtctttcca gacccttcc catcatcaga aggaacccga cagacccgaa ggaatcgaag
                /\ (tat, rev, nef) 3' sj
7981 aaggaggtgg cgagcaaggc agagacagat cagtgcgatt agtgagcgga ttcttagcac
                                                   <- tat cds end
8041 ttgcctggga cgatctacgg agcctgtgcc tcttcagcta ccaccggttg agagacttaa
8101 ccttgattgc agcgaggacg gtggaacttc tgggacacag cagtctcaag ggactgagac
8161 gggggtggga aggcctcaaa tatctgggga atcttctgtt atattggggc caggaactaa
```

```
8221 aaattagtgc tatttctttg cttgatgcta cagcaatagc agcagcgggg tggacagaca
          <-rev cds end
8281 gggttataga agtagcacaa ggagcttgga gagccattct ccacatacct agaagaatca
8341 gacagggctt agaaaggact ttgctataac atgggaagta agtggtcaaa aagtagcata
                                  <- env cds
                                     gp41 end
                        nef cds start ->
8401 gtgggatggc ctcaggtcag ggaaaaaata aagcaaactc ctccagcaac agaaggagta
8461 ggagcagtat ctcaagatct agataaacat ggagcaataa caagtagtaa tatagataat
8521 gctgattgtg tctggctgag agcacaagag gacgaggagg taggctttcc agtcatgccg
8581 caggtacctc taagaccaat gacttataag ggagcttttg atcttagctt ctttttaaaa
8641 gaaaaggggg gactggatgg gctaatttac tccaagaaaa gacaagagat ccttgactta
                       \3'LTR U3 start
8701 tgggtctata atacacaagg cttcttccct gattggcaaa actacacacc agggccaggg
8761 atcagattcc cactgtgttt tggatggtgc ttcaagctag taccagttga ccaaagagaa
8821 gtagaggagg acaacaaagg agaaaacaac tgcctgttac accccatgag ccagcatgga
8881 atagaggacg aagaaagaga agtgctgatg tggaagtttg acagtgccct agcacgaaaa
8941 cacgtagccc gagaacagca tccagagtac tataaagact gctgacaagg aagtttctac
                                                   <- nef cds end
9001 tagaacttcc gctggggact ttccagggga ggtgtggccg gggcggagtt ggggagtagc
                  ------- ---- Nf-Kappa B
9061 taaccctcag atgctgcata aaagcagccg cttttcgctt gtactgggtc tctcttgtta
                                                LTR U3/\R rpt
9121 gaccaggtcg agcccgggag ctctctggct agcaagggaa cccactgctt aaagcctcaa
                                                                     --
9181 taaagcttgc cttgagtgct taa
     ----                   <- mRNA end
 signal poly-A      R rpt end/
```

```
LOCUS       HIVETH2220   9031 bp    DNA             VRL       19-JUN-1996
DEFINITION  HIV-1 isolate C2220 from Ethiopia, subtype C, complete genome.
ACCESSION   U46016
SOURCE      Human immunodeficiency virus type 1 (HIV-1), strain C2220 (ETH2220),
            clone pNOTA/C2220/C3.6.4; subtype C from Ethiopia; isolate by
            short-term co-culture on patient's PBMCs.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9031)
  AUTHORS   Salminen,M.O., Johansson,B., Sonnerborg,A., Ayehunie,S., Gotte,D.,
            Leinikki,P., Burke,D.S. and McCutchan,F.E.
  TITLE     Full-length sequence of an Ethiopian human immunodeficiency virus
            type 1 (HIV-1) isolate of genetic subtype C
  JOURNAL   AIDS Research and Human Retroviruses 12 (14), 1329-1339 (1996)
REFERENCE   2  (bases 1 to 9031)
  AUTHORS   Salminen,M.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1996) Mika O. Salminen, Global Molecular
            Epidemiology, Henry M. Jackson Foundation, 1600 East Gude Drive,
            Rockville, MD 20850, USA
 COMMENT    U46016 is the first reported almost full length subtype C
            sequence from Ethiopia.  In its genomic organization, this clone
            closely resembles subtype A, B, and D isolates except that the core
            promoter contains three potential binding sites for the
            transcription factor NF-kB instead of containing two.
            See also GenBank accession number U15061.
FEATURES             Location/Qualifiers
     source          1..9031
                     /organism="Human immunodeficiency virus type 1"
                     /strain="C2220 (ETH2220)"
                     /note="subtype C from Ethiopia; isolated by short term
                     co-culture on normal human PBMC"
                     /clone="pNOTA/C2220/C3.6.4"
     CDS             170..1684
                     /gene="gag"
                     /note="GAG p17/p24/NCp7/p6"
                     /codon_start=1
                     /product="matrix, capsid and core proteins"
                     /db_xref="PID:g1353861"
                     /translation="MGARASILRGEKLDAWEKIKLRPGGKKHYMLKHLVWANRELEKF
                     ALNPDLLDTSAGCKQIIKQLQPALQTGTEELKSLFNTVATLYCVHQKIEIKDTKEALD
                     KIEEEQNESQQKTQQAGAADRGKDSQNYPIVQNMQGQMVHQPISARTLNAWVKVVEEK
                     AFSPEVIPMFTALSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHA
                     GPVAPGQMRDPRGSDIAGTTSTLQEQIAWMTGNPPVPVGDIYKRWIILGLNKIVRMYS
                     PVSILDIKQGPKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRA
                     LGPGASLEEMMTACQGVGGPAHKARVLAEAMSQVNNTTIMMQKSNFKGPKRAIKCFNC
                     GKEGHLARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGRLWPSNKGRPGNFLQSRP
                     EPTAPPESLRPEPTAPPPESFRFEEATPSPKQELKDREALTSLKSLFGNDHLLQ"
     CDS             <1456..4488
                     /gene="pol"
                     /note="initiation codon is unknown"
                     /codon_start=1
                     /product="protease/RT/RNAseH/integrase"
                     /db_xref="PID:g1353862"
                     /translation="FFRETLAFQQGKAREFPSEQTRANSPTRESQTRANSPTTRELQV
                     RGSNTFSEAGAERQGSLNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEINLP
                     GKWKPKMIGGIGGFIKVRQYDQIIIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGRTL
                     NFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTAICEEMEQEGKISRIGPENPY
```

```
                     NTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDA
                     YFSVPLDEGFRKYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPPIFQSSMPQILEPFR
                     APNPEIVIYQYMDDLYVGSDLEIGQHRAPIEELREHLLKWGFTTPDKKHQKEPPFLWM
                     GYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKA
                     LTDIVTLTEEAELELAENREILKEPVHGVFYDPSKDLIAEIQKQGNDQWTFQFYQEPF
                     KNLKTGKFAKRGTAHTNDVKQLTAVVQKIALESIVIWGKTPKFRLPIQKETWEAWWTD
                     YWQATWIPEWEFVNTPPLVKLWYQLEKEPIAGVETFYVDGAANRETKIGKAGYVTDRG
                     RQKIVSLTETTNQKTELQAIQLALQDSGSEVNIVTDSQYALGIILAQPDKSESEIVNQ
                     IIEQLISKERVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSN
                     WRAMANEFNIPPVVPKEIVACCDKCQLKGEAIHGQVNCSPGIWQLDCTHLEGKIILVA
                     VHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVRVIHTDNGSNFTSNAVKAACWWAG
                     IQQEFGIPYNPQSQGVVESMNKELKKIIGQVREQAEHLKTAVQMAVFIHNFKRRGGIG
                     GYSAGERIIDIIASDIQTKELQNQILKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVV
                     IQDNSDIKVVPRRKAKIIRDYGKQMAGADCVAGRQDED"
CDS                  4433..5011
                     /gene="vif"
                     /note="infectivity factor"
                     /codon_start=1
                     /product="VIF"
                     /db_xref="PID:g1353863"
                     /translation="MENRWQVLIVWQVDRMKIRTWNSLVKHHMHISRRANGWVYRHHY
                     DSRHPKVSSEVHIPLGEARLIIKTYWGLQTGERDWHLGHGVSIEWRLRSYNTQVDPGL
                     ADHLIHMHYFDCFAESAIRKAILGYRVSPRCDYQAGHNKVGSLQYLALTALIKPKKAK
                     PPLPSVSKLVEDKWNKPQKTRGRRGNHTMNGH"
CDS                  4951..5241
                     /gene="vpr"
                     /codon_start=1
                     /product="VPR"
                     /db_xref="PID:g1353864"
                     /translation="MEQAPEDQGPQREPYNEWALELLEELKQEAVRHFPRPWLHNLGQ
                     YIYETYGDTWSGVEALIRTLQQLMFIHFRIGCQHSRIGILRQRRARNGASRS"
CDS                  5222..5350
                     /gene="tat"
                     /note="transactivating factor; prematurely truncated at nt
                     5350 due to a deletion at nt 5344; second exon not
                     expressed"
                     /codon_start=1
                     /product="TAT"
                     /db_xref="PID:g1353865"
                     /translation="MEPVDPNLEPWNHPGSQPKTACNQCYCKKCSYHCLVCFLTKA"
CDS                  join(5360..5435,7755..8002)
                     /gene="rev"
                     /note="regulator of protein expression; prematurely
                     truncated at nt 8002"
                     /codon_start=1
                     /product="REV"
                     /db_xref="PID:g1353866"
                     /translation="MAGRSGDSDEELLKAVRIIKILYQSNPYPTPEGTRQARRNRRRR
                     WRARQRQIHTLSERILSNFLGRPAEPVPLQLPPLERLNLDCSEDSGTSGTQQSQGTTE
                     GVGNP"
CDS                  5459..5719
                     /gene="vpu"
                     /codon_start=1
                     /function="promotes extracellular release of virions"
                     /product="VPU"
                     /db_xref="PID:g1353867"
                     /translation="MVDLLAKVDYRIVIVAFIVALIIAIVVWTIAYIEYRKLLRQRRI
                     DRLIKRTRERAEDSGNESDGDTEELSTMVDMGNLRLLDVNDL"
```

```
     CDS             5637..8192
                     /gene="env"
                     /note="viral envelope gene;  contains external and
                     transmembrane glycoproteins; cleaved to gp120 & gp41"
                     /codon_start=1
                     /product="gp160"
                     /db_xref="PID:g1353868"
                     /translation="MKVMGIQRNCQQWWIWGILGFWMLMICNGMGNLWVTVYYGVPVW
                     KDASPTLFCASDAKAYDTEVHNVWGTFACVPTDPSPQELGLENVTENFNMWKNDMVEQ
                     MHQDIISLWDQGLKPCVKLTPLCVTLNCNAIKNNTKVTNNSINSANDEMKNCSFNITT
                     ELRDKKRKAYALFYKLDIVPLNNGSTDYRLINCNTSTITQACPKVSLDPIPIHYCAPA
                     GYAILKCRDKTFTGTGPCHNVSTVQCTHGIKPVVSTQLLLNGSIAEGETIIRFENLTN
                     NAKIIIVQLNESVEITCTRPSNNTRESIRIGPGQTFYATGDIIGDIRQAHCNISEEKW
                     NKTLQKVKEKLQKHFPNKTIEFKPSSGGDLEITTHSFNCGGEFFYCNTSNLFNSTKLE
                     LFNSSTNLNITLQCRIKQIINMWQGVGRAMYAPPIEGIIMCRSNITGLLLTRDGAKEP
                     HSTKEIFRPEGGDMRDNWRSELYKYKVVEIKPLGVAPTKPKRRVVEREKRAALGALFL
                     GFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLKAIEAQQHMLQLTVWGIKQLQT
                     RVLAIERHLRDQQLLGIWGCSGKLICTTAVPWNSSWSNKSQEEIWDNMTWMQWDREIS
                     NYTDIIYNLLEVSQNQQDKNEKDLLALDKWENLWNWFNITNWLWYIKIFIMIVGGVIG
                     LRIIFAVLSIVNRVRQGYSPLSFQTLIPHPRGPDRLGGIEEEGGEQGRDRSIRLVNGF
                     LAIFWDDLRSLCLFSYHRLRDLILIAARTVELLGRSSLKGLQRGWETLKYLGSLVQYW
                     GLELKKSAINLLNTTAIVVGEGTDRFIELIQRIWRAFCNIPRRIRQGLEAALQ"
     misc_RNA        7145..7378
                     /gene="env"
                     /note="RRE region of mRNA; viral envelope gene"
     CDS             8194..8817
                     /gene="nef"
                     /note="regulatory protein"
                     /codon_start=1
                     /product="NEF"
                     /db_xref="PID:g1353869"
                     /translation="MGGTMSKCSPVGWPAIRERIRRAAPAAEGVGAASRDLDKYGALT
                     SSNTPANNPDCAWLEAQEEEEEVGFPVRPQVPLRPMTYKAAFDLSLFLKEKGGLEGLI
                     YSKKRQEILDLWVYNTQGFFPDWQNYTPGPGVRYPLTFGWCFKLVPVDPSEVEEINEG
                     ENNCLLHPASLHGMEDEDREVLKWKFDSHLARRHMARELHPEYYKDC"
     LTR             8486..>9031
                     /note="3′, U3 and R regions only"
     protein_bind    8825..8834
                     /note="extra NFkB site III"
                     /bound_moiety="NFkb"
     protein_bind    8838..8847
                     /note="NFkB site II"
                     /bound_moiety="NFkb"
     protein_bind    8851..8860
                     /note="NFkB site I"
                     /bound_moiety="NFkb"
     misc_RNA        8939..8997
                     /note="TAR region of mRNA"
BASE COUNT     3270 a   1616 c   2141 g   2004 t
ORIGIN
        1 aaatctctag cagtggcgcc cgaacagggg acctgaaagc gaaagtgaga ccagaggaga
                     <- 5′ LTR
       61 tctctcgacg caggactcgg cttgctgaag tgcactcggc aagaggcgag agcggcgact
      121 ggtgagtacg ccaattttta tttgactagc ggaggctaga aggagagaga tgggtgcgag
          /\ major 5′ sj                                gag cds - >
                                          p17 MA (matrix) start
      181 agcgtcaata ttaagaggcg aaaaattaga tgcctgggaa aaaattaagt taaggccagg
      241 gggaaagaaa cactatatgc tgaaacacct agtctgggca aacagggagc tggaaaaatt
      301 tgcacttaac cctgaccttt tagatacatc agcaggctgt aaacaaataa ttaaacagct
```

```
 361 acaaccagct cttcagacag gaacagagga acttaaatca ttatttaata cagtggcaac
 421 tctctattgt gtacatcaaa agatagagat aaaagacacc aaggaagcct tagacaagat
 481 agaggaagaa caaaacgaaa gtcagcaaaa aacacagcag gcaggagcag ctgacagagg
 541 aaaggacagt caaaattatc ctatagtgca gaatatgcag gggcaaatgg tacatcagcc
             gag p17 MA/\gag p24 CA
 601 catatcagct agaactttga atgcatgggt aaaagtagta gaggaaaagg ctttcagccc
 661 agaggtaata cccatgttta cagctttatc agaaggagcc accccacaag atttaaacac
 721 catgctaaat acagtggggg gacatcaagc agccatgcaa atgttaaaag acaccatcaa
 781 tgaggaggct gcagaatggg acaggttaca tccagtgcat gcagggcctg ttgcaccagg
 841 ccaaatgaga gacccaaggg gaagtgacat agcaggaaca actagtaccc ttcaggaaca
 901 aatagcatgg atgacaggga acccacctgt tccagtggga gacatctata aaagatggat
 961 aatcctgggg ctaaataaaa tagtaagaat gtatagccct gtcagcattt tggacataaa
1021 acaaggacca aaggaacctt ttagagacta tgtagaccgg ttctttaaaa ccttaagagc
1081 tgaacaagct acacaagatg taaaaaattg gatgacagac accttgttgg tccaaaatgc
1141 gaacccagat tgtaaaacca ttttaagagc attagggcca ggggcttcat tagaagaaat
1201 gatgacagca tgtcagggag tgggaggacc tgcccacaaa gcaagagtgt tggctgaggc
                                                   gag p24 CA/\gag p2x
1261 aatgagccaa gtaaacaata caaccataat gatgcagaaa agcaatttta agggccctaa
                              gag p2x/\gag p7NC
                                       \gag-pol TF
1321 aagagcaatt aaatgtttca actgtggcaa ggaagggcac ctagccagaa attgcagggc
1381 ccctaggaaa aaaggctgtt ggaaatgtgg aaaggaagga caccaaatga aagactgtac
1441 cgagagacag gctaattttt tagggagact ttggccttcc aacaagggaa ggccagggaa
                     ----- - gag-pol ribosomal slip site
1501 tttccttcag agcagaccag agccaacagc cccaccagag agtctcagac cagagccaac
gag p7NC/\gag p6
1561 agccccacca ccagagagct tcaggttcga ggaagcaaca ccttctccga agcaggagct
1621 gaaagacagg gaagccttaa cttccctcaa atcactcttt ggcaacgacc acttgttaca
            gag-pol fusion TF/\pol protease
1681 ataaaaatag ggggacagct aaaggaggct ctcttagaca caggagcaga tgatacagta
       <- gag cds
          p6 end
1741 ttagaagaaa taaatttgcc aggaaaatgg aaaccaaaaa tgataggagg aattggaggt
1801 tttattaaag taagacagta tgatcaaata atcatagaaa tttgtggaaa aaaggctata
1861 ggtacagtac tagtaggacc tacacctgtc aacataattg gcagaaacat gttgactcag
1921 cttggacgca cattaaactt tccaattagt cccattgaaa ctgtaccagt aaaattaaag
               pol protease/\pol reverse transcriptase
1981 ccaggaatgg atggcccaaa agtcaaacaa tggccattga cagaagaaaa gataaaagca
2041 ttaacagcaa tttgtgaaga aatggagcag gaaggaaaaa tttcaagaat tgggcctgaa
2101 aacccatata atactccagt atttgccata aaaaagaagg acagtactaa gtggagaaaa
2161 ttagtagatt tcagggaact caataaaaga actcaagact tttgggaagt tcaattaggg
2221 ataccacacc cagcagggtt aaaaaagaaa aaatcagtga cagtactaga tgtgggggat
2281 gcatatttct cagttccttt agatgaaggt ttcagaaaat atactgcatt caccatacct
2341 agtacaaaca atgaaacacc agggattaga tatcaatata atgtcctccc acagggatgg
2401 aaaggatcac caccaatatt ccagagtagc atgccccaaa tcttagagcc ctttagggcc
2461 cccaacccag aaatagttat ctatcaatat atggatgact tgtatgtagg atctgactta
2521 gaaatagggc aacatagagc cccaatagaa gagttaagag aacatctatt aaagtgggga
2581 tttaccacac cagacaagaa acatcagaaa gaacctccat ttctttggat ggggtatgaa
2641 cttcatcctg acaaatggac agtacagcct atacagctgc cagaaaagga tagctggact
2701 gtcaatgata tacaaaagtt agtgggaaaa ttaaactggg caagtcagat ttacccaggg
2761 attaaagtaa ggcaactgtg taaactcctt aggggagcca aagcactaac agacatagta
2821 acactaactg aagaagcaga attagaattg gcagagaaca gggaaattct aaaggaacca
2881 gtacatggag tattttatga cccatcaaaa gacttaatag ctgaaataca gaaacagggg
2941 aatgaccaat ggacatttca attttaccaa gaaccattca aaaatctgaa aacagggaag
3001 tttgcaaaaa gagggactgc ccacactaat gatgtaaagc agttaacagc ggtagtgcaa
3061 aagatagccc tggaaagcat agtaatatgg ggaaagactc ctaaatttag attacccatc
3121 cagaaagaaa catgggaagc atggtggaca gactattggc aagccacctg gattcctgaa
3181 tgggagtttg ttaatacccc tcccctagta aaattatggt accagctaga gaaagaaccc
```

```
    3241 atagcaggag tagaaacttt ctatgtagat ggagcagcta atagggaaac taaaatagga
      pol reverse transcriptase/\pol RNAse
    3301 aaagcagggt atgttactga tagaggaagg cagaaaattg tttctctaac tgaaacaaca
    3361 aatcagaaga ctgaattaca agcgatccag ctagcgttac aggattcagg atcagaagta
    3421 aacatagtaa cagactcaca gtatgcatta gggatcatcc tggcacaacc agataagagt
    3481 gaatcagaga tagtcaatca aataatagaa cagttaataa gcaaggaaag ggtctacctg
    3541 tcatgggtac cagcacataa aggaattgga ggaaatgaac aagtagataa attagtaagt
    3601 agtggaatca ggaaagtgct gtttctagat ggaatagata aggctcaaga agagcatgaa
                      pol RNAse/\pol integrase
    3661 aaatatcaca gcaattggag agcaatggct aatgaattta atatcccacc cgtagtaccc
    3721 aaagaaatag tagcttgctg tgataaatgt cagctaaaag gggaagccat acatggacaa
    3781 gtaaattgta gtccagggat atggcaatta gattgtacac acttagaagg gaaaatcatc
    3841 ctggtagcag tccatgtagc cagtggctac atagaggcag aggttattcc agcagaaaca
    3901 ggacaagaaa cagcatactt tctactaaaa ttagcaggga gatggccagt cagggtaata
    3961 catacagata atggcagtaa cttcaccagt aatgcagtta aagcagcctg ttggtgggca
    4021 ggtattcaac aggaatttgg aattccctac aatccccaaa gtcaaggagt agtagaatct
    4081 atgaataaag aattaaagaa aatcataggg caggtaagag aacaagctga gcaccttaag
    4141 acagcagtac aaatggcagt attcattcac aattttaaaa gaagaggggg gattgggggg
    4201 tacagtgcag gggaaagaat aatagatata atagcatcag acatacagac taaagaactc
    4261 caaaaccaaa ttttaaaaat tcaaaatttt cgggtttatt acagagacag cagagaccct
                                                        /\ 3′ sj
    4321 atttggaaag gaccagccaa actactctgg aaaggtgaag gggcagtagt aatacaagat
                                              /\ 5′ sj
    4381 aatagtgaca taaaggtagt accaaggagg aaagcaaaaa tcattaggga ttatggaaaa
                                                   vif cds start ->
    4441 cagatggcag gtgctgattg tgtggcaggt agacaggatg aagattagaa catggaatag
                                                          <- pol cds
                                                             integrase end
    4501 tttagtaaag caccatatgc atatttcaag gagagctaat ggatgggttt atagacatca
    4561 ttatgacagc agacatccaa aggtaagttc agaagtacac atcccattag gggaggctag
    4621 attaataata aaaacatatt ggggtttgca aacaggagaa agagattggc atttgggtca
    4681 tggagtctcc atagaatgga gattgagaag ctataacaca caagtagacc ctggcctggc
    4741 agaccaccta attcatatgc attattttga ttgttttgca gaatctgcca taaggaaagc
                                                   /\ 3′ sj
    4801 catattagga tatagagtta gccctaggtg tgactatcaa gcaggacata ataaggtagg
                                                                  /\ 5′ sj
    4861 atctctacaa tacttggcac tgacagcatt gataaagcca aaaaaggcaa agccacctct
    4921 gcctagtgtt agtaaattag tagaggataa atggaacaag ccccagaaga ccaggggccg
                             vpr cds start ->
    4981 cagagggaac catacaatga atgggcacta gagcttttag aggagctcaa gcaggaagct
                                        < - vif cds end
    5041 gtcagacact ttcctagacc atggctccat aacttaggac aatatatcta tgaaacctat
    5101 ggggatactt ggtcgggagt agaagcttta ataagaactc tgcaacaact aatgtttatc
    5161 catttcagaa ttggatgcca gcatagcaga ataggcattt tacgacagag aagagcaaga
                /\ 3′ sj
    5221 aatggagcca gtagatccta acctagaacc ctggaaccat ccaggaagtc agcctaagac
      tat ->                 < - vpr cds end
start cds
    5281 tgcttgtaat caatgttatt gtaaaaaatg tagctatcat tgtctagttt gctttctgac
    5341 aaaggcttag gcatttccta tggcaggaag aagcggcgac agcgacgaag agctcctcaa
            /\ rev cds start - >    /\ 3′ sj
      1 pt deletion
                <- tat cds premature end
                   due to deletion
    5401 agcagtaagg atcatcaaaa tcttatatca aagcagtaag taccaaataa tagatgtaat
                         (tat, rev, nef) 5′ sj /\          vpu cds start ->
    5461 ggttgattta ctagcaaaag tagattatag aatagtaata gtagcattca tagtagcact
    5521 aatcatagca atagttgtgt ggaccatagc atatatagaa tataggaaat tgttaagaca
```

```
5581 aagaagaata gataggttaa ttaaaagaac tagggaaaga gcagaagaca gtggcaatga
                                                          env cds ->
                                                   signal peptide
5641 aagtgatggg gatacagagg aattgtcaac aatggtggat atggggaatc ttaggctttt
5701 ggatgttaat gatttgtaat ggaatgggga acttgtgggt cacagtctat tatggggtac
                      <- vpu cds end
       env signal peptide/ \gp120
5761 ctgtatggaa agatgctagc cctactctat tttgtgcatc agatgctaaa gcatatgata
5821 cagaagtaca taatgtctgg ggtacatttg cctgtgtccc cacagacccc agcccacaag
5881 aattgggttt ggaaaatgtg acagaaaatt ttaacatgtg gaaaaatgac atggtggagc
5941 agatgcatca ggatataatc agtttatggg atcaaggcct aaagccatgt gtaaagttga
6001 ccccactctg tgttacttta aactgtaatg ctatcaaaaa caatactaag gttacaaata
6061 atagtatcaa tagtgccaat gatgaaatga aaaattgttc tttcaatata accacagaac
6121 taagagataa gaaaaggaag gcatatgcac ttttttataa acttgatata gtaccactta
6181 ataacggttc tactgattat agattaataa attgtaatac ctcaaccata acacaagcct
6241 gtccaaaggt ctctttggac ccaattccta tacattattg tgctccagct ggttatgcga
6301 ttctaaagtg tagagataag acattcactg gaacaggacc atgccataat gtcagcacag
6361 tacaatgcac acatggaatt aaaccagtgg tatcaactca actattgtta aatggtagta
6421 tagcagaagg tgagacaata attagatttg aaaatctgac aaacaatgcc aaaataataa
6481 tagtacagct taatgaatct gtagaaatta cttgtacgag acccagcaat aatacaagag
                         V3 loop start ->
6541 aaagtataag gataggacca ggacaaacat tctatgcaac aggagacata ataggagata
6601 taagacaagc acattgtaac attagtgaag aaaaatggaa taaaactcta caaaaggtaa
                     <- V3 loop end
6661 aggaaaaatt acaaaagcac ttccctaata aaacaataga atttaagcca tcctcaggag
6721 gggacctaga aattacaaca catagcttta attgtggagg agaatttttc tattgcaata
6781 catcaaacct gtttaatagc acaaaactag aactgtttaa tagcagtaca aatttaaaca
6841 tcacactcca atgcagaata aaacaaatta taaacatgtg gcagggagta ggacgagcaa
6901 tgtatgcccc tcccattgaa ggaataataa tgtgtagatc aaatatcaca ggactactac
6961 tgacacgtga tggagccaaa gagccacata gcacaaaaga gatattcaga cctgaaggag
7021 gagatatgag ggacaattgg agaagtgaat tatataaata taaagtggta gaaattaagc
7081 cactaggagt agcacccact aagccaaaaa ggagagtggt ggagagagaa aaaagagcag
                                                          env gp120/\env gp41
7141 cactaggagc tttgttcctt gggttcttgg gagcagcagg aagcactatg ggcgcagcat
7201 caataacgct gacggtacag gccagacaat tattgtcggg tatagtgcaa cagcaaagca
7261 atttgctgaa agctatagag gcgcaacagc atatgttgca gctcacggtc tggggcatta
7321 agcagctcca gacaagagtc ctggctatag aaagacacct aagggatcaa cagctcctag
7381 ggatttgggg ctgctctgga aaactcatct gcaccactgc tgtgccttgg aactctagtt
7441 ggagtaataa gtctcaagag gagatttggg ataacatgac ctggatgcag tgggatagag
7501 aaattagtaa ttatacagac ataatataca atttgcttga agtctcgcaa aaccagcagg
7561 acaaaaatga aaaagattta ctagcattgg acaaatggga aaatctgtgg aattggttta
7621 acataacaaa ttggctgtgg tatataaaaa tattcataat gatagtggga ggtgtgatag
7681 gtttaagaat aatttttgct gtgctttcta tagtgaatag agttaggcag ggatactcac
7741 cattatcgtt tcagaccctt atcccacacc cgaggggacc cgacaggctc ggaggaatcg
                  /\ (tat, rev, nef) 3’ sj
7801 aagaagaagg tggagagcaa ggcagagaca gatccatacg cttagtgaac ggattcttag
                                                        <- tat cds end
7861 caattttttg ggacgacctg cggagcctgt gcctcttcag ttaccaccgc ttgagagact
7921 taatcttgat tgcagcgagg acagtggaac ttctgggacg cagcagtctc aagggactac
7981 agagggggtg ggaaaccctt aaatatctgg gaagccttgt gcagtattgg ggtctggagc
                           <- premature rev end
8041 taaaaaagag tgctattaat ctgcttaata ccacagcaat agtagtaggt gaaggaacag
            <- rev cds normal end
8101 atagatttat agaattaata caaagaattt ggagagcttt ctgcaacata cctagaagaa
8161 taagacaggg cttggaagca gctttgcaat aaaatggggg gcacgatgtc aaaatgtagt
                                     <- env cds
                                        gp41 end
                           nef cds start ->
8221 ccagtaggat ggcctgctat aagagaaaga ataagacgag ctgctccagc agcagaggga
```

```
8281 gtaggagcag cgtctcgaga cttagacaaa tatggggcac ttacaagcag caacacaccc
8341 gccaataatc ctgattgtgc ctggctggaa gcgcaagagg aggaagaaga ggtaggcttt
8401 ccagtcagac ctcaggtacc tttaagacca atgacttata aggcagcatt cgatctcagc
8461 ctctttttaa aagaaaaggg gggactggaa gggttaattt actccaagaa aaggcaggag
                    3′ LTR U3 ->
8521 attcttgatt tgtgggtcta taacacacaa ggcttcttcc ctgattggca aaactacaca
8581 ccaggaccag gggtcagata tccactgacc tttggatggt gcttcaagct agtaccagtt
8641 gacccaagtg aagtggaaga aatcaatgaa ggagagaaca actgcttgct gcaccctgcg
8701 agcctgcatg gaatggagga tgaagacaga gaagtattaa agtggaagtt tgacagtcac
8761 ctagcacgca gacacatggc ccgcgagcta catccggagt attacaaaga ctgctgacac
                                                                 <- nef cds
                                                                    end
8821 agacgggact ttccgccggg actttccact ggggcgttcc aggaggaggg gtctgggcgg
         ------ ----   --- -------    ----------
       NFkB site III   NFkB site II   NFkB site I
8881 gactgggagt ggccaaccct cagatgctgc atataagcag ctgcttttcg cttgtaccgg
                                                            LTR U3/\R rpt
8941 gtctctctag gtagaccaga tctgagcctg ggagctctct ggctatctgg ggaacccact
9001 gcttaagcct caataaagct tgccttgagt g
                ------
             poly A signal
```

```
LOCUS       HIVU51188    9843 bp    DNA             VRL       18-JUL-1996
DEFINITION  Human immunodeficiency virus type 1 (HIV-1), clone 90cr402 from the
            Central African Republic, complete genome.
ACCESSION   U51188
KEYWORDS    .
SOURCE      Human immunodeficiency virus type 1 (HIV-1), clone 90cr402; PBMCs
            extracted from symptomatic male patient from Central African
            Republic; expanded in chimpanzee cells and re-expanded in human
            PBMCs; complete genome lambda cloned and sequenced.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9843)
  AUTHORS   Gao,F., Morrison,S.G., Robertson,D.L., Thornton,C.L., Craig,S.,
            Karlsson,G., Sodroski,J., Morgado,M., Galvao-Castro,B., von
            Briesen,H., Beddows,S., Weber,J., Sharp,P.M., Shaw,G.M. and
            Hahn,B.H.
  TITLE     Molecular cloning and analysis of functional envelope genes from
            human immunodeficiency virus type 1 sequence subtypes A through G.
            The WHO and NIAID Networks for HIV Isolation and Characterization
  JOURNAL   J. Virol. 70 (3), 1651-1657 (1996)
  MEDLINE   96190564
REFERENCE   2  (bases 1 to 9843)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     The heterosexual HIV-1 epidemic in Thailand is caused by an
            intersubtype (A/E) recombinant of African origin
  JOURNAL   J. Virology 70 (10), 7013-7029 (1996)
REFERENCE   3  (bases 1 to 9843)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1996) John Blouin, HIV Database, Los Alamos
            National Laboratory, Eniwetok Dr., Los Alamos, NM 87545, USA
  COMMENT   A set of three complete genomes from a study linking the HIV-1
            epidemic in the heterosexual population in Thailand to an A/E
            recombinant. 90cr402, previously named CAR-E 4002, was obtained
            from a man from Bangui, Central African Republic, who had
            lymphadeopathy, diarrhea, severe weight loss and recurrent
            respiratory infections. He was infected through heterosexual
            contact, but the year of infection is unknown. 90cr402 was first
            adapted to growth in chimpanzee cells, expanded in chimpanzee
            cells, and then re-expanded in human PBMCs before lamba cloning
            and sequencing. 90cr402 and another sequence in the study, 93th253,
            are subtype A/E recombinants, and comparison of the two strains
            shows that they were derived from a common A/E recombinant ancestor,
            presumably from Central Africa.
FEATURES             Location/Qualifiers
     source          1..9843
                     /organism="Human immunodeficiency virus type 1"
                     /note="from the Central African Republic"
                     /clone="90cr402"
     LTR             1..656
     CDS             832..2328
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
```

```
                /translation="MGARASILSGGKLDAWEKIRLRPGGKKQYRMKHLVWASRELER
                FALNPGLLETAEGCQQLIEQLQSTIKTGSEELKSLFNTIATLWCVHQRIDVKDTKEA
                LDKIEEVQNKSKQKAQQAAAGTGSSSKVSQNYPIVQNAQGQMVHQALSPRTLNAWVK
                VVEEKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQILKDTINEEAAEWDR
                VHPVHAGPIPPGQMREPRGADIAGTTSTLHEQIGWMTSNPPIPVGEIYKKWIILGLN
                KIVRMYSPVSILDIRQGPKEPFRDYVDRFFKTLRAEQATQEVKNWMTETLLVQNANP
                DCKSILKALGTGATLEEMMTACQGVGGPGHKARVLAEAMSQVHHTNIMMQKGNFKGQ
                RKIKCFNCGKEGHLARNCRAPRKKGCWKCGREGHQMKDCTERQANFLGKIWPSSKGR
                PGNFPQSRPEPTAPPMESLGMGEEITSFPKQEQKDKKQPPPLVSLKSLFGNDPLSQ"
CDS             join(832..2120,2120..5132)
                /codon_start=1
                /product="gag-pol polyprotein"
                /translation="MGARASILSGGKLDAWEKIRLRPGGKKQYRMKHLVWASRELER
                FALNPGLLETAEGCQQLIEQLQSTIKTGSEELKSLFNTIATLWCVHQRIDVKDTKEA
                LDKIEEVQNKSKQKAQQAAAGTGSSSKVSQNYPIVQNAQGQMVHQALSPRTLNAWVK
                VVEEKGFNPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQILKDTINEEAAEWDR
                VHPVHAGPIPPGQMREPRGADIAGTTSTLHEQIGWMTSNPPIPVGEIYKKWIILGLN
                KIVRMYSPVSILDIRQGPKEPFRDYVDRFFKTLRAEQATQEVKNWMTETLLVQNANP
                DCKSILKALGTGATLEEMMTACQGVGGPGHKARVLAEAMSQVHHTNIMMQKGNFKGQ
                RKIKCFNCGKEGHLARNCRAPRKKGCWKCGREGHQMKDCTERQANFFRENLAFQQGE
                ARKFPSEQTRANSPTNGELGDGGRDNLLPEAGAERQETASSFSFPQITLWQRPLVTV
                KVGGQLKEALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVRQYDQILIEICGKK
                AIGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPIDTVPVTLKPGMDGPKVKQWPLTE
                EKIKALTEICKEMEEEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQ
                DFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEGFRKYTAFTIPSINNETPGV
                RYQYNVLPQGWKGSPAIFQSSMTKILEPFRARNPEIVIYQYMDDLYVGSDLEIGQHR
                TKVEDLRAHLLSWGFTTPDKKHQKEPPFLWMGYELHPDRWTVQPIVLPEKDSWTVND
                IQKLVGKLNWASQIYAGIKVRQLCKLLRGAKALTDIVTLTEEAELELAENREILKTP
                VHGVYYDPSKDLVAEIQKQGQDQWTYQIYQEPFKNLKTGKYARKRSAHTNDVRQLAE
                VVQKVATESIVIWGKTPKFRLPIQRETWETWWMEYWQATWIPEWEFVNTPPLVKLWY
                QLEKDPIMGAETFYVDGAASRETKQGKAGYVTDRGRQKVVSLTETTNQKTELHAIHL
                ALQDSGPEVNIVTDSQYALGIIQAQPDRSESDIVNQIIEKLIEKEKVYLSWVPAHKG
                IGGNEQVDKLVSSGIRKVLFLDGIDKAQEDHERYHSNWRAMASDFNLPPIVAKEIVA
                SCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQ
                ETAYFLLKLAGRWPVRVIHTDNGSNFTSAAVKAACWWANVQQEFGIPYNPQSQGVVE
                SMNKELKKIIGQVREQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIIDIIATDIQ
                TKALQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAK
                IIRDYGKQMAGDDCVAGRQNED"
CDS             5077..5355
                /gene="vif"
                /note="premature end at 5355. Normal end would be at 5655."
                /codon_start=1
                /product="vif protein"
                /translation="MENRWQVMIVWQVDRMRIRAWNSLVKHHMYSSKKAAKWFYRHHY
                ESQHPKVSSEVHIPLGDARLIIRTYWGLHTGEKDWHLGHGVSIEWRQRK"
CDS             5084..5374
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /translation="MEQAPEDQGPQREPYNEWTLELLEELKTEAVRHFPRPWLHGLGK
                HIYNTYGDTWEGVEAIIRILQQLLFVHFRIGCQHSRIGIIRGTRGRNGAGRS"
CDS             join(5866..6080,8442..8526)
                /gene="tat"
                /codon_start=1
                /product="tat protein"
                /translation="MEPVDPNLEPWKHPGSQPTTACSKCYCKMCCWHCQLCFLKKGLG
                ISYGRKKRKHRRGPSQDSKDHQNSIPKQPLPTSRGNPTGPKESKKKVESKAETDP"
```

```
     CDS             join(6005..6080,8442..8737)
                     /gene="rev"
                     /codon_start=1
                     /product="rev protein"
                     /translation="MAGRSGSTDEDLLRTVRIIKILYQSNPYPPAEGTRQARKNRRRK
                     WRARQRQIHKIGERILSTCLGRSPEPVPLQLPPLERLHLDCSEDCGTSGTQQSTGTET
                     EVGRPQISGESSVILGSGTKN"
     CDS             6101..6346
                     /gene="vpu"
                     /codon_start=1
                     /product="vpu protein"
                     /translation="RLPLHICAIVGLIVALIIAIVVWTIVAIEYKKLRRQRKIDRLVQ
                     RISERAEDSGNESEGDTEELAKLVEMGDFDPWVGDNL"
     CDS             6264..8879
                     /gene="env"
                     /codon_start=1
                     /product="env polyprotein"
                     /translation="MRVKGTRRNWPNLWKWGTLILGLVIICSASDNLWVTVYYGVPVW
                     RDADTILFCASDAKAHVTEVHNVWATHACVPTDPNPQEIYLENVTENFDMWKNNMVEQ
                     MQEDVISLWDQSLQPCVKLTPLCVTLHCTKASFTNATSDRIKMEDAVRNCSFNMTTEL
                     QDKKQEVHALFYTSDVVQISSSVQNNNNSNTSGQNNSHKFRLIHCNTSVIKQACPKVS
                     FDPIPIHYCAPAGYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
                     EEIIIRSEDLTDNAKTIIVHLNKSIEINCTRPFKKVRISARIGPGRVFHTTGNINGDI
                     RKAYCEINKTKWKETLKQVTRKLREHLNGTMTISFRPSSGGDPEITMHHFNCRGEFFY
                     CNTTALFNSTWINGTMQEVNGTNSGNITLPCRIKQIVNMWQEVGRAMYAPPISEVINC
                     VSNITGILLTRDGGINQNQTNKNETFRPGGGNIKDNWRSELYKYKVVQIEPLGIAPTK
                     ARRRVVEREKRAVGIGAMIFGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRA
                     IEAQQHLLQLTVWGIKQLQARVLAVERYLKDQKFLGLWGCSGKIICTTSVPWNSSWSN
                     KSYEAIWNNMTWIEWDREINEYTNQIYELLTESQDQQERNEKDLLALDKWASLWNWFD
                     ITRWLWYIKIFIIIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLTHQQREPDRPER
                     IEEESGEQGRDRSIRLVSGFLALAWDDLRSLCLFSYHRLRDFILIAARTVELLGHSSL
                     QGLRRRWEGLKYLGNLLSYWVQELRISAITLLDATAITVAGWTDRVIEIVQRAWRAIL
                     HIPRRIRQGLERALL"
     CDS             8881..9516
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /translation="MGGKWSKSCIVGWPQVRERIRQTPVAEERQTPAAAEGVGAVSQD
                     LDKHGAVTSSNINNADNVWLEAQEEEEVGFPVRPQVPLRPMTYKGAFDLSFFLKEKGG
                     LDGLIHSKRRQEILDLWVHNTQGYFPDWQNYTPGPGVRYPLCFGWCFKLVPVDPREVE
                     EDNKGENNCLLHPMSQHGIDDDEREVLMWKFDSSLARRHIARELHPEYYKD"
     LTR             9188..9843
     misc_signal     386..395
                     /note="NFkappaB"
     misc_signal     9573..9582
                     /note="NFkappaB"
BASE COUNT     3501 a   1764 c   2400 g   2178 t
ORIGIN
        1 tggatgggct aattcactcc aagagaagac aagagatcct tgacttatgg gtccacaata
5' LTR U3 ->
       61 cacaaggcta cttccctgat tggcaaaact acacaccagg gccgggggtc agatacccac
      121 tgtgttttgg atggtgcttc aagctagtac cagttgatcc aagagaagta gaggaggaca
      181 ataaaggaga aaacaactgc ctgttacacc ccatgagcca gcatggaata gatgatgatg
      241 aaagagaagt gctaatgtgg aagtttgaca gttccctagc acgaagacac atagctcgag
      301 agctgcaccc agagtactac aaagactgat gacaaagaag ttccaagact gctgacaaag
                                                                  <- nef
                                                                     cds end
      361 gagtttctaa ctaggacttc cgctggggac tttccagggg aggtgtggcc tgggcggggt
                                     ------ ------ Nf Kappa B
```

```
 421 tggggagtgg ctaaccctca gatgctgcat aaaagcagcc gcttttcgct tgtactgggt
                                                          LTR U3/\R rpt
 481 ctctcttgtt agaccaggtc gagcccggga gctctctggc tagcagggga acccactgct
 541 taaagcctca ataaagcttg ccttgagtgc ttaaagtggt gtgtgcccgt ctgtgttagg
             - -----              R rpt/\LTR U5
          poly-A signal
 601 actctggtaa ctagagatcc ctcagaccac tctagactga gtaaaaatct ctagcagtgg
                                                                <- LTR
 661 cgcccgaaca gggacttgaa agcgaaagtt aatagggact cgaaagcgaa agttccagag
 721 aagttctctc gacgcaggac tcggcttgct gaggtgcaca cagcaagagg cgagagcggc
 781 gactggtgag tacgccaaaa attttgacta gcggaggcta gaaggagaga gatgggtgcg
        /\ major 5′ sj                              gag cds ->
                                                p17 MA start
 841 agagcgtcaa tattaagtgg gggaaaatta gatgcatggg agaaaattcg gctaaggcca
 901 gggggaaaga aacaatatag aatgaaacat ttagtatggg caagcagaga gttagaaaga
 961 ttcgcactta accctggcct tttagaaaca gcagagggat gtcaacaact aatagaacag
1021 ttacaatcaa ctatcaagac aggatcagaa gaacttaaat cactatttaa tacaatagca
1081 actctctggt gcgtacatca aaggatagat gtaaaagaca ccaaggaagc tttagataaa
1141 atagaggaag tacaaaataa gagcaagcaa aaagcacagc aggcagcagc tggcacagga
1201 agcagcagca aagtcagcca aaattaccct atagtgcaaa atgcacaagg gcaaatggta
                      gag p17 MA/\gag p24 CA
1261 catcaggcct tatcacctag aactttgaat gcatgggtga aagtagtaga agagaagggt
1321 tttaacccag aagtaatacc catgttctca gcattatcag aaggagccac cccacaagat
1381 ttaaatatga tgctcaacat agtgggggga caccaggcag caatgcaaat attaaaagat
1441 accatcaatg aggaagctgc agaatgggat agggtacacc cagtacatgc agggcctatt
1501 ccaccaggcc agatgagaga accaagggga gctgacatag caggaactac tagtaccctt
1561 catgaacaga taggatggat gacaagcaac ccacctattc cagtgggaga aatctataaa
1621 aagtggataa tcctgggatt aaataaaata gtaagaatgt atagccctgt tagcattttg
1681 gacataagac aagggccaaa agaacccttc agagactatg tagacaggtt ctttaaaact
1741 ctcagagcgg aacaagctac acaggaggta aaaaactgga tgacagaaac tttgctagtc
1801 caaaatgcga acccagactg taagtccatt ttaaaagcac taggaacagg ggctacatta
1861 gaagaaatga tgacagcatg ccagggagtg ggaggacctg gacataaagc aagggttttg
                                                         gag p24 CA/
1921 gctgaggcaa tgagtcaagt acaccataca aatataatga tgcagaaagg caattttaag
     \gag p2x                          gag p2x/\p7NC
                                              \gag-pol TF
1981 ggccagagaa aaattaagtg tttcaactgt ggcaaagaag gacacctagc cagaaattgc
2041 agggccccta gaaaaaaggg ttgttggaaa tgtgggaggg aaggacacca aatgaaagac
2101 tgcactgaaa gacaggctaa ttttttaggg aaaatctggc cttccagcaa ggggaggcca
                           ------          ^^^^^ ^^ ^^       ^^ ^^^^^^
       gag-pol ribosomal frameshift site      stem      loop     stem
2161 ggaaatttcc ctcagagcag accagagcca acagccccac caatggagag cttggggatg
     ^
     gag p7NC/\gag P6
2221 ggggaagaga taacctcctt cccgaagcag gagcagaaag acaagaaaca gcctcctcct
2281 ttagtttccc tcaaatcact ctttggcaac gaccccttgt cacagtaaaa gtaggaggac
  gag-pol TF/\pol protease                                <- gag cds
                                                             P6 end
2341 agctaaaaga agctctatta gacacaggag cagatgatac agtattagaa gacataaatt
2401 tgccaggaaa atggaaacca aaaatgatag ggggaattgg aggttttatc aaagtaagac
2461 aatatgatca gatacttata gaaatttgtg ggaaaaaggc tataggtaca gtgttagtag
2521 gacctacccc cgtcaacata attggacgga acatgttgac tcagattggt tgtactttaa
2581 atttcccaat tagtcctatt gacactgtac cagtaacatt aaagccagga atggatggac
pol protease/\pol RT
2641 caaaggttaa acaatggcca ttgacagaag aaaaaataaa agcattaaca gaaatttgta
2701 aagaaatgga agaggaagga aaaatctcaa aaattgggcc tgaaaatcca tacaatactc
2761 cagtatttgc tataaagaaa aaggacagca ctaaatggag aaaattagta gatttcagag
2821 agctcaataa aagaactcag gacttttggg aagttcaatt aggaatacca catccagcag
2881 gtctaaaaaa gaaaaaatca gtaacagtac tagatgtggg agacgcatat ttttcagttc
```

```
2941 ctttagatga aggctttaga aagtatactg cattcaccat acctagtata aacaatgaga
3001 caccaggagt cagatatcag tacaatgtgc tgccgcaggg atggaaagga tcaccagcaa
3061 tattccagag tagcatgaca aaaatcttag agccctttag agcaagaaat ccagaaatag
3121 ttatctatca atacatggat gacttgtatg taggatctga tttagaaata gggcagcata
3181 gaacaaaagt agaggatcta agagctcatc tattgagctg gggatttact acaccagaca
3241 aaaagcatca gaaagaaccg ccattccttt ggatgggata tgaactccat cctgacagat
3301 ggacagtcca gcctatagtg ctgccagaaa aagacagctg gactgtcaat gatatacaga
3361 agttagtggg aaaactaaat tgggcaagtc aaatttatgc agggattaaa gtgaggcaac
3421 tgtgtaaact cctcagggga gctaaagcac taacagacat agtaacactg actgaggaag
3481 cagaattaga attggcagag aacagggaaa ttctaaaaac ccctgtgcat ggggtatatt
3541 atgacccatc aaaagactta gtagcagaaa tacagaaaca agggcaagac caatggacat
3601 atcaaattta tcaagagcca tttaaaaatc taaaaacagg aaaatatgca agaaaaaggt
3661 ctgctcacac taatgatgta agacaattag cagaagtggt gcaaaaagtg gccacagaaa
3721 gcatagtaat atggggaaag acccctaaat ttagactacc catacaaaga gaaacatggg
3781 agacatggtg gatggagtat tggcaggcta cctggattcc tgaatgggag tttgtcaata
3841 cccctcctct agtaaaatta tggtaccaat tagaaaaaga ccccataatg ggagcagaga
3901 ctttctatgt agatggggca gctagtaggg agactaagca aggaaaagca gggtatgtca
   pol RT/\pol RNAse
3961 ctgacagagg aagacaaaaa gtagtttccc taactgagac aacaaatcaa aagactgaat
4021 tacatgcgat ccatttagcc ttgcaggatt caggaccaga agtaaacata gtaacagact
4081 cacaatatgc attaggaatt attcaggcac aaccagacag gagtgagtca gatatagtca
4141 atcaaataat agagaaacta atagaaaagg aaaaagtcta cctgtcatgg gtcccagcac
4201 ataaggggat tggaggaaat gaacaagtag ataaattagt cagttctgga atcaggaagg
4261 tgctattttt agatgggata gataaggctc aagaagacca tgaaagatat cacagcaatt
pol RNAse/\pol integrase
4321 ggagagcaat ggctagtgac tttaatttgc cacctatagt agcaaaggaa atagtagcca
4381 gctgtgataa atgtcagcta aaaggggaag ccatgcatgg acaagtagac tgtagtccag
4441 ggatatggca actagattgc acgcatctag aaggaaaagt catcctggta gcagtccacg
4501 tggccagtgg atatatagaa gcagaagtta tcccagcaga aacaggacag gagacagcat
4561 actttctgct aaaattagca ggaagatggc cagtaagggt aatacacaca gacaatggca
4621 gcaatttcac cagcgctgca gttaaagcag cctgttggtg ggccaatgtc caacaggaat
4681 ttggaattcc ctacaatccc caaagccaag gagtagtgga atctatgaat aaggaattaa
4741 agaaaatcat agggcaggta agagagcaag ctgaacacct taagacagca gtacaaatgg
4801 cagtattcat tcacaatttt aaaagaaaag gggggattgg ggagtacagt gcaggggaaa
4861 gaataataga cataatagca acagacatac aaactaaagc attacaaaaa caaattacaa
4921 aaattcaaaa ttttcgggtt tattacaggg acagcagaga tccaatttgg aaaggaccag
4981 caaaactact ctggaaaggt gaaggggcag tagtaataca ggacaatagt gatataaaag
5041 tagtaccaag aagaaaagca aagatcatta gggattatgg aaaacagatg gcaggtgatg
                                vif cds start ->
5101 attgtgtggc aggtagacag aatgaggatt agagcatgga acagtttagt aaaacatcat
                                      <- pol cds
                                         integrase end
5161 atgtatagct caaagaaagc tgccaaatgg ttttatagac atcattatga aagccagcat
5221 ccaaaagtaa gttcagaagt acacatccca ctaggggatg ctagattaat aataagaaca
5281 tattggggtc tgcatacagg agaaaaagac tggcacttgg gtcatggagt ctccatagaa
5341 tggaggcaga gaaaatagag cacacaaata gatcctgacc tagcagacca actgattcat
                   <- vif premature end
5401 ctgcaatatt ttgactgttt ttcagactct gccataagga aagccttatt aggacaaata
                               /\ 3′ sj
5461 gttagaccta ggtgtgaata tccagcagga cataacaagg taggatctct acaatatttg
                                             /\ 5′ sj
5521 gcactgaaag cattaacaaa aacaaaaaag acaaagccac ctctgcctag tgttaggaaa
5581 ttaacagaag atagatggaa caagccccag aagaccaagg gccacagaga gagccctaca
      vpr cds start ->
5641 atgaatggac attagaactg ttagaggagc ttaaaactga agctgttaga cattttccta
                   <- vif cds end
5701 ggccctggct ccatggctta ggaaagcata tctataacac ttatggggat acttgggaag
5761 gggttgaagc tataataaga attttgcaac aactactgtt tgttcatttc agaattgggt
```

```
    5821 gccaacatag cagaataggc attatacgag ggacaagagg caggaatgga gccggtagat
                                            tat cds start ->          <- vpr
                                                                         cds end
    5881 cctaacctag agccctggaa acatccggga agtcagccta caactgcttg tagcaagtgt
    5941 tattgtaaaa tgtgttgctg gcattgccaa ctatgctttc tgaaaaaagg cttaggcatc
    6001 tcctatggca ggaagaagcg gaagcaccga cgaggacctt ctcaggacag taaggatcat
rev cds start->
    6061 caaaattcta taccaaagca gtaagtatta agtatatata aggttgcctt tgcatatctg
                 tat-rev exon/ \intron
                                vpu cds normal start ->
                                ATG in others, AGG here
    6121 tgcaatagta ggactgatag tagcgctaat catagcaata gtagtgtgga ctatagtggc
    6181 tatagaatat aagaaattaa gaaggcaaag aaaaatagat aggttagttc aaagaataag
    6241 tgaaagagca gaagacagtg gaaatgagag tgaaggggac acggaggaat tggccaaact
                          env cds ->
             signal peptide start
    6301 tgtggaaatg ggggactttg atccttgggt tggtgataat ttgtagtgcc tcagacaact
                                                        <- vpu cds end
                                            env signal pept/\gp120
    6361 tgtgggttac agtttattat ggggttcctg tgtggagaga tgcagatacc atcctatttt
    6421 gtgcatcaga tgccaaagca catgtgacag aagtgcacaa tgtttgggct acacatgcct
    6481 gtgtacccac agaccccaac ccacaagaaa tatacctgga aaatgtaaca gaaaattttg
    6541 acatgtggaa aaataacatg gtagagcaga tgcaggagga tgtaatcagt ttatgggatc
    6601 aaagtctaca gccatgtgta aagttaactc ctctctgtgt tactttacat tgtaccaagg
    6661 ctagttttac taatgccacc agtgacagaa taaaaatgga agatgcagta agaaactgtt
    6721 cttttaatat gaccacagaa ctacaagata aaaagcagga ggtccatgca ctttttttata
    6781 cgtctgatgt agtacaaatt agtagtagtg tacaaaataa taataacagt aatactagtg
    6841 gacaaaataa tagtcataag tttagattaa tacattgtaa tacttcagtc attaagcagg
    6901 cttgtccaaa ggtatccttt gatccaattc ccatacatta ttgtgctcca gctggttatg
    6961 cgattctaaa gtgtaatgat aagaatttta atggaacagg gccatgcaaa aatgtcagct
    7021 cagtacaatg cacacatgga attaagccag tggtatcaac tcaattgctg ttaaatggca
    7081 gtctagcaga agaagagata ataatcagat ctgaagatct cacagacaat gccaaaacca
    7141 taatagtgca ccttaataaa tctatagaaa tcaattgtac cagacccttc aagaaagtaa
    7201 gaataagtgc aaggatagga ccaggacgag tattccatac aacaggaaac ataaatggtg
    7261 atataagaaa agcatattgt gaaattaata aaacaaaatg gaaagaaact ttaaaacagg
    7321 taacaagaaa attaagagag caccttaatg ggacaatgac aataagcttt cgaccatcct
    7381 caggaggaga tccagaaatc acaatgcatc attttaattg tagaggggaa tttttctatt
    7441 gcaatacaac agcactgttt aatagtactt ggataaatgg aaccatgcag gaggttaatg
    7501 gcacaaactc aggcaatatc acacttccat gcaggataaa gcaaattgta aacatgtggc
    7561 aggaagtagg acgagcaatg tatgctcctc ccatcagtga agtaattaat tgtgtatcaa
    7621 atattacagg aatactattg acaagagatg ggggtattaa tcaaaatcag actaacaaaa
    7681 atgagacctt cagacctgga ggaggaaata taaaggacaa ttggagaagt gaattataca
    7741 aatataaagt agtacaaatt gaaccactag gaatagcacc caccaaggca aggagaagag
    7801 tggtggagag agaaaaaaga gcagtgggaa taggagctat gatctttggg ttcttaggag
                   env gp120/ \gp41
    7861 cagcaggaag cactatgggc gcggcgtcaa taacgctgac ggtacaggcc agacaattat
    7921 tgtctggtat agtgcaacag caaagcaatt tgctgagggc tatagaggcg cagcagcatc
    7981 tgttgcaact cacagtctgg ggcattaaac agctccaggc aagagtcctg gctgtggaaa
    8041 gatacctaaa agatcaaaag ttcctaggac tttggggctg ctctggaaaa atcatctgca
    8101 ccacttctgt gccttggaac tccagttgga gtaataaatc ttatgaagct atttggaaca
    8161 acatgacatg gatagaatgg gatagagaaa ttaacgaata cacaaaccaa atatatgagc
    8221 tacttacaga atcgcaggac cagcaggaaa ggaatgaaaa ggatttgtta gcattggaca
    8281 agtgggcaag tctgtggaat tggtttgaca taacaaggtg gctgtggtat ataaaaatat
    8341 ttataataat agtaggaggt ttaataggtt taagaatagt ttttgctgtg ctttctatag
    8401 taaatagagt taggcaggga tactcacctt tgtcgttcca gacccttacc caccagcaga
                                     tat-rev intron/\exon
    8461 gggaacccga caggcccgaa agaatcgaag aagaaagtgg agagcaaggc agagacagat
    8521 ccataagatt ggtgagcgga ttcttagcac ttgcctggga cgatctccgg agcctgtgcc
             <- tat cds end
    8581 tcttcagcta ccaccgcttg agagacttca tcttgattgc agcgaggact gtggaacttc
    8641 tgggacacag cagtctacag ggactgagac ggaggtggga aggcctcaaa tatctgggga
```

```
   8701 atcttctgtc atattgggtt caggaactaa gaattagtgc tattactttg cttgatgcta
                                                <- rev cds end
   8761 cagcaataac agtagcgggg tggacagata gggttataga aatagtacaa agagcttgga
   8821 gagctatcct ccatatacct agaagaatca gacagggctt ggaaagggct ttgctataac
                                                                  <- env cds
                                                                     gp41 end
   8881 atgggtggca aatggtcaaa aagttgcata gtgggatggc ctcaggtcag ggaaagaata
nef cds ->
   8941 aggcaaactc ctgtagcaga agaaaggcaa actcctgcag cagcagaagg agtaggagca
   9001 gtatctcaag atctagataa acatggagca gtaacaagca gtaatataaa taatgctgat
   9061 aatgtctggc tggaagcgca agaggaagag gaggtaggct ttccagtcag gccgcaggta
   9121 cctctaagac caatgactta taagggagct tttgatctta gcttcttttt aaaagaaaag
   9181 gggggactgg atgggctaat tcactccaag agaagacaag agatccttga cttatgggtc
           LTR ->
   9241 cacaatacac aaggctactt ccctgattgg caaaactaca caccagggcc gggggtcaga
   9301 tacccactgt gttttggatg gtgcttcaag ctagtaccag ttgatccaag agaagtagag
   9361 gaggacaata aaggagaaaa caactgcctg ttacacccca tgagccagca tggaatagat
   9421 gatgatgaaa gagaagtgct aatgtggaag tttgacagtt ccctagcacg aagacacata
   9481 gctcgagagc tgcacccaga gtactacaaa gactgatgac aaagaagttc caagactgct
                                            <- nef cds end
   9541 gacaaaggag tttctaacta ggacttccgc tggggacttt ccaggggagg tgtggcctgg
                                          --------- --
                                           nf kappa B
   9601 gcggggttgg ggagtggcta accctcagat gctgcataaa agcagccgct tttcgcttgt
   9661 actgggtctc tcttgttaga ccaggtcgag cccgggagct ctctggctag caggggaacc
         U3/\R rpt
   9721 cactgcttaa agcctcaata aagcttgcct tgagtgctta aagtggtgtg tgcccgtctg
                        ---- --               R rpt/\U5
                        poly-A signal
   9781 tgttaggact ctggtaacta gagatccctc agaccactct agactgagta aaaatctcta
   9841 gca
         <- 3’ LTR U5 end
```

```
LOCUS       HIVU51190    8980 bp    DNA             VRL       18-JUL-1996
DEFINITION  HIV-1 clone 92ug037 from Uganda, complete genome
ACCESSION   U51190
KEYWORDS    .
SOURCE      Human immunodeficiency virus type 1 (HIV-1), clone 92ug037 from
            Uganda; genomic DNA isolated from PBMCs; genome extended by nested
            PCR amplification.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 8980)
  AUTHORS   Gao,F., Morrison,S.G., Robertson,D.L., Thornton,C.L., Craig,S.,
            Karlsson,G., Sodroski,J., Morgado,M., Galvao-Castro,B., von
            Briesen,H., Beddows,S., Weber,J., Sharp,P.M., Shaw,G.M. and
            Hahn,B.H.
  TITLE     Molecular cloning and analysis of functional envelope genes from
            human immunodeficiency virus type 1 sequence subtypes A through G.
            The WHO and NIAID Networks for HIV Isolation and Characterization
  JOURNAL   J. Virol. 70 (3), 1651-1657 (1996)
  MEDLINE   96190564
REFERENCE   2  (bases 1 to 8980)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     The heterosexual HIV-1 epidemic in Thailand is caused by an
            intersubtype (A/E) recombinant of African origin
  JOURNAL   J. Virology 70 (10), 7013-7029 (1996)
REFERENCE   3  (bases 1 to 8980)
  AUTHORS   Gao,F., Robertson,D.L., Morrison,S.G., Hui,H., Craig,S.,
            Fultz,P.N., Decker,J., Girard,M., Shaw,G.M., Hahn,B.H. and
            Sharp,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1996) John Blouin, HIV Database, Los Alamos
            National Laboratory, Enitwetok Dr., Los Alamos, NM 87545, USA
 COMMENT    A set of three complete genomes from a study linking the HIV-1
            epidemic in the heterosexual population in Thailand to an A/E
            recombinant. 92ug037 was obtained from WHO-NIAIDS and comes from
            an asymptomatic 31-year-old female from Entebbe, Uganda, early in
            infection. The year of infection is unknown and the mode of
            infection was heterosexual contact. 92ug037, U51190, was
            established and propagated by cocultivation with normal donor
            lymphocytes, and then PCR amplified and sequenced. 92ug037
            is subtype A.
FEATURES             Location/Qualifiers
     source          1..8980
                     /organism="Human immunodeficiency virus type 1"
                     /note="from Uganda"
                     /clone="92ug037"
     CDS             160..1659
                     /gene="gag"
                     /codon_start=1
                     /product="gag polyprotein"
                     /translation="MGARASVLSGGKLDAWEKIRLRPGGKKKYRLKHLVWASRELERF
                     ALNPSLLETTEGCQQIMEQLQSALRTGTEELRSLYNTVATLYCVHQRIEVKDTKEALD
                     KIEEIQKKSKQKTQQAAADTGSSSKVSQNYPIVQNAQGQMIHQSLSPRTLNAWVKVIE
                     EKALSPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKDTINEEAAEWDRLHPV
                     HAGPVAPGQMREPRGSDIAGTTSTPQEQIAWMTGNPPIPVGDIYKRWMILGLNKIVRM
                     YSPVSILDIKQGPKEPFRDYVDRFFKTLRAEQATQEVKGWMTETLLIQNANPDCKSIL
                     RALGAGATLEEMMTACQGVGGPGHKARVLAEAMSQVQHTNIMMQRGNFKGQKRIKCFN
                     CGKEGHLAKNCRAPRKKGCWKCGREGHQMKDCTERQANFLGKIWPSSKGRPGNFPQSR
                     PEPTAPPAAEIFGMREEIVSPPKQEQNDRDQNPPSVSLKSLFGNDLLSQ"
```

```
CDS             join(160..1450,1450..3146)
                /note="Premature stop codon at 3146, normal end would be
                at 1642."
                /codon_start=1
                /product="gag-pol polyprotein"
                /note="Gag-Pol fusion polpyprotein is produced by
                ribosomal frameshift slip on tttttt slippery site
                at 1450-1455 potentiated by an mRNA stem-loop structure
                at 1450-1455"
                /translation="MGARASVLSGGKLDAWEKIRLRPGGKKKYRLKHLVWASRELERFA
                LNPSLLETTEGCQQIMEQLQSALRTGTEELRSLYNTVATLYCVHQRIEVKDTKEALDKI
                EEIQKKSKQKTQQAAADTGSSSKVSQNYPIVQNAQGQMIHQSLSPRTLNAWVKVIEEKA
                LSPEVIPMFSALSEGATPQDLNMMLNIVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGP
                VAPGQMREPRGSDIAGTTSTPQEQIAWMTGNPPIPVGDIYKRWMILGLNKIVRMYSPVS
                ILDIKQGPKEPFRDYVDRFFKTLRAEQATQEVKGWMTETLLIQNANPDCKSILRALGAG
                ATLEEMMTACQGVGGPGHKARVLAEAMSQVQHTNIMMQRGNFKGQKRIKCFNCGKEGHL
                AKNCRAPRKKGCWKCGREGHQMKDCTERQANFFRENLAFQQREARKFSSEQTRTNSPTS
                SRDLWDEGRDSLPSEAGAERQGPEPTFSFPQITLWQRPLVTVKIGGQLKKALLDTGADD
                TVLEDINLPGKWKPKMIGGIGGFIKVKQYDQILIEICGKKAIGTVLVGPTPVNIIGRNM
                LTLIGCTLNFPISPISTVPVKLKPGMDGPRIKQWPLTEEKIKALTEICADMEREGRISK
                IGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTV
                LDVGDAYFSVPLDESFRKYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQASMTKI
                LEPFRSKNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELREHLLKWGFTTPDKKHQKEPP
                FLWMGYELHPDKWTVQPIELPEKESWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRG
                TKALTDIVTLTEEAELELAENREILKDPVHGAYYDPSKDLIAEIQKQGQDQWIYQIYQE
                PFKNLKTGKYARKRSAHTNDVKQLAEVVQKVVMESIVIWGKTPKFKLPIQKETWETWWM
                DYWQAT"
CDS             4408..4986
                /gene="vif"
                /codon_start=1
                /product="vif protein"
                /translation="MENRWQVMIVWQVDRMRIRTWNSLVKHHMYISRRAKGWFYRHHYE
                SRHPKVSSEVHIPIGDARIVVRTYWGLQTGEKDWHLGHGVSIEWRLKRYSTQIDPDLAD
                QLIHLHYFNCFSDSAIRKAILGQVVSPRCDYQTGHNKVGSLQYLALKALVTPSRMKPPL
                PSVKKLAEDRWNKPQKTRGRRESHTMNGC"
CDS             4926..5219
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /translation="MEQAPEDQGPQREPYNEWMLDLLEDLKHEAVRHFPRPWLHGLGQH
                IYHTYGDTWEGVEAIIRILQQLLFVHFRIGCQHSRIGINIRGRRVRDGSGRS"
CDS             join(5200..5414,7739..7830)
                /gene="tat"
                /codon_start=1
                /product="tat protein"
                /translation="MDPVDPSLEPWNHPGSQPKTPCNKCYCKVCCYHCQCCFLNKGLGI
                SYGRKKRKPRRGTPQSNKDHQNPIPKQPIPRTQGDSTGPEESKKKVESKTEADRYA"
CDS             join(5339..5414,7739..8035)
                /gene="rev"
                /codon_start=1
                /product="rev protein"
                /translation="MAGRSGNPDEELLRAIRIIKILYQSNPYPEPKGTRQARKNRRRRW
                RARQRQIDTLSERVLSTCLGRPAEPVPLQLPPIERLHLDCSEDCGTSGTQQSQGVETGV
                GRTQVSGESPVVLGSGTKN"
CDS             5441..5686
                /gene="vpu"
                /codon_start=1
                /product="vpu protein"
                /translation="MQLLEICAVVGLVVALIIAIVVWTIVGIEYKKLLKQRKIDRLVDR
                IRERAEDSGNESDGDREELSLLVDMGDYDLGDDNNL"
CDS             5604..8177
```

```
                     /gene="env"
                     /codon_start=1
                     /product="env polyprotein"
                     /translation="MRVMGIERNYPCWWTWGIMILGMIIICNTAENLWVTVYYGVPIWK
                     DANTTLFCASDAKAYDTEVHNVWATHACVPTDPSPQELKMENVTEEFNMWKNNMVEQMH
                     TDIISLWDQSLKPCVQLTPLCVTLDCSYNITNNITNSITNSSVNMREEIKNCSFNMTTE
                     LRDKNRKVYSLFYKLDVVQINNGNNSSNLYRLINCNTSALTQARPKVTFEPIPIHYCAP
                     AGYAILKCNDKEFNGTGLCKNVSTVQCTHGIRPVVSTQLLLNGSLAEGKVMIRSENITN
                     NVKNIIVQLNESVTINCTRPNNNTRRSVRIGPGQTFYATGDIIGDIRQAHCNVSGSQWN
                     KTLHQVVEQLRKYWNNNTIIFNSSSGGDLEITTHSFNCAGEFFYCNTSGLFNSTWVNGT
                     TSSTSNGTITLPCRIKQIINMWQRVGQAMYAPPIQGVIKCESNITGLILTRDGGVNSSD
                     SETFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKARRRVVEREKRAVTLGAVFIGFL
                     GTAGSTMGAASITLTVQARKLLSGIVQQQSNLLRAIEAQQHLLKLTVWGIKQLQARVLA
                     VERYLRDQQLLGIWGCSGKLICPTNVPWNSSWSNKSLDEIWENMTWLQWDKEISNYTIK
                     IYELIEESQIQQERNEKDLLELDKWASLWNWFDISKWLWYIKIFIMIVGGLIGLRIVFA
                     VLSVINRVRQGYSPLSFQTHTPNPRGLDRPGRIEEEGGEQDRGRSIRLVSGFLALAWDD
                     LRNLCLFSYHRLRDFILIAARTVELPGHSSLKGLRLGWEGLKYLGNLLLYWGRELKISA
                     INLLDTIAIAVAGWTDRVIETVQRLGRAILNIPRRIRQGFERALL"
     CDS             8179..8826
                     /gene="nef"
                     /codon_start=1
                     /product="nef protein"
                     /translation="MGNKWSKSCIVGWPEVRERIRQTPTAARERTRQAPTAAKGVGAVS
                     QDLDKHGAVTSSNVNHPSCVWLEAQEEEEVGFPVRPQVPLRPMTYKAAFDLGFFLKEKG
                     GLDGLIYSKKRQEILDLWVYHTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVPVDEDEVE
                     EATGGENNSLLHPICQHGMDDEEKETLRWKFDSSLARVHKARELHPEFYKDC"
     misc_signal     8844..8853
                     /note="NFkappaB"
     misc_signal     8858..8867
                     /note="NFkappaB"
BASE COUNT     3238 a   1569 c   2174 g   1999 t
ORIGIN
        1 gacttgaaag cgaaagtaat agggactcga aaacgaaagt tccagagaag aagctctctt
       61 gacgcaggac tcggcttgct gaggtgcaca cagcaagagg cgagagcggc gactggtgag
                                                                  /\major 5’ sj
      121 tacgccattt tttgactagc ggaggctaga aggagagaga tgggtgcgag agcgtcagta
                                          gag cds ->
                                      p17 MA start
      181 ttaagtgggg gaaaattaga tgcatgggag aaaattcggt taaggccagg gggaaagaaa
      241 aaatatagat taaaacatct agtatgggca agcagggagc tggaaagatt tgcacttaac
      301 cctagccttt tagaaacaac agaaggatgt caacaaataa tggaacaatt acaatcagct
      361 ctcagaacag gaacagaaga acttagatca ttatataata cagtagcaac cctctattgc
      421 gtacatcaac ggatagaggt aaaagacacc aaggaagctc tagataaaat agaggagata
      481 caaaagaaaa gcaagcaaaa gacacagcag gcagcagctg acacaggaag tagcagcaag
      541 gtcagccaaa attaccctat agtgcaaaat gcacaagggc aaatgatcca ccagtccttg
               gag p17 MA/\gag p24 CA
      601 tcacctagga ctttgaatgc atgggtgaag gtaatagaag aaaaggctct cagcccagaa
      661 gtaataccca tgttctcagc attatcagaa ggagccaccc cacaagattt aaatatgatg
      721 ctgaacatag tgggggggaca ccaggcagct atgcaaatgt taaaagatac catcaatgag
      781 gaagctgcag aatgggacag gctacatcca gtacatgcag ggcctgttgc accaggccag
      841 atgagagaac caaggggaag tgatatagca ggaactacta gtacccctca agaacaaata
      901 gcatggatga caggcaaccc acccatccca gtgggagaca tctataaaag atggatgatc
      961 ctgggattaa ataaaatagt aagaatgtat agccctgtta gcattttgga tataaaacaa
     1021 gggccaaaag aacccttcag agactatgta gataggtttt ttaaaactct cagagctgag
     1081 caagctacac aggaggtaaa aggttggatg acagaaacgt tactgatcca aaatgcaaat
     1141 ccagattgta aatccatcct aagagcatta ggagcagggg ctacattaga agaaatgatg
     1201 acagcatgcc agggagtggg aggacccggc cataaagcaa gagttttggc tgaggcaatg
                                                   gag p24 CA/\gag p2x
     1261 agtcaagtac aacatacaaa cataatgatg cagagaggca attttaaggg ccagaaaagg
                               gag p2x/\gag p7NC
                                       \gag-pol fusion TF
```

```
1321 attaagtgtt tcaactgtgg caaagaagga catctagcca aaaattgcag ggctcctaga
1381 aaaaagggct gttggaaatg tggaagggaa gggcaccaaa tgaaggactg cactgagaga
1441 caggctaatt ttttagggaa aatctggcct tccagcaaag ggaggccagg aaattttcct
                -- ----          ^^^^^^^  ^^      ^ ^ ^^^^^^^
gag-pol ribosomal frame shift       stem-loop inverted repeats
                                                        gag p7NC/\p6
1501 cagagcagac cagaaccaac agccccacca gcagcagaga tctttgggat gagggaagag
1561 atagtctccc ctccgaagca ggagcagaac gacagggacc agaacccacc ttcagtttcc
                                                         gag-pol TF/\pol
                                                                 protease
1621 ctcaaatcac tctttggcaa cgacctcttg tcacagtaaa gataggggga cagctaaaaa
                                            <- gag cds
                                               p6 end
1681 aagctctatt agatacagga gcagatgata cagtattaga agacataaat ttgccaggaa
1741 aatggaaacc aaaaatgata gggggaattg gaggtttcat caaggtaaag cagtatgatc
1801 agatacttat agaaatttgt ggaaaaaagg ctataggtac agtattggta ggacctacac
1861 ctgtcaacat aattggaaga aatatgttga ccctgattgg ttgtacttta aatttcccaa
                                                   pol protease/\pol RT
1921 ttagtcctat tagtactgta ccagtaaaat taaaaccagg aatggatggc ccaaggatta
1981 aacaatggcc attgacagaa gaaaaaataa aagcattaac agaaatttgt gcagatatgg
2041 aaagagaagg aagaatttca aaaattgggc ctgaaaatcc atacaatact ccaatatttg
2101 ctataaagaa aaaggacagc actaaatgga gaaaattagt agattttaga gagctcaata
2161 aaagaactca agacttttgg gaagttcaat taggaatacc gcatccagcg ggcttaaaaa
2221 agaaaaaatc agtaacagta ctagatgtgg gggacgcata tttttcagtt cccttagatg
2281 aaagctttag aaagtatact gcattcacca tacctagtac aaacaatgag acaccaggaa
2341 tcaggtatca gtacaatgtg cttccacagg gatggaaagg atcaccggca atattccagg
2401 ccagcatgac aaaaatctta gagcccttta gatcaaaaaa tccagacata gttatctatc
2461 aatacatgga tgacttgtat gtaggatctg atttagaaat agggcagcat agaacaaaaa
2521 tagaagaatt aagagaacat ctattaaaat ggggatttac tacaccagac aaaaagcatc
2581 agaaagaacc tccatttctt tggatgggat atgaactcca tcctgataag tggacagtcc
2641 aaccgataga gctgccagaa aaggaaagct ggactgtcaa tgatatacag aaattagtag
2701 ggaaactaaa ttgggcaagt caaatttatg caggaattaa agtaaaacaa ttgtgtaaac
2761 tcctcagggg aaccaaagca ttaacagata tagtaacatt gactgaggaa gcagaattag
2821 aattggcaga gaacagggag attttaaaag accctgtgca tggagcatat tatgacccat
2881 caaaagactt aatagcagag atacagaaac aagggcaaga ccaatggata tatcaaattt
2941 atcaagagcc atttaaaaat ctaaaaacag gaaaatatgc aagaaaaagg tctgctcaca
3001 ctaatgatgt aaaacaattg gcagaagtgg tgcaaaaggt ggtcatggaa agcatagtaa
3061 tatggggaaa gactcctaaa tttaaactac ccatacaaaa agaaacatgg gaaacatggt
3121 ggatggacta ttggcaggct acctgaattc ctgaatggga gtttgtcaat acccctcctc
                                 <- pol cds premature end
3181 tagtaaaatt atggtaccag ttagagaaag accccatagc aggagcagag actttctatg
                                                         pol RT/\pol
                                                                RNAse
3241 tagatggggc agccaatagg gagactaagc taggaaaagc agggtatgtc actgacagag
3301 gaaggcaaaa ggttgtttcc ctaactgaga caacaaatca aaagactgaa ttacatgcaa
3361 tccatctagc cttgcaggat tcaggatcag aagtaaacat agtaacagac tcacagtatg
3421 cattaggaat cattcaggcc cagccagaca ggagtgaatc agagttagtc aatcaaataa
3481 tagagaagct aatagaaaag gacaaagtct acctgtcatg ggtaccagca cacaaaggaa
3541 ttggaggaaa tgaacaagta gataaattag tcagttctgg aatcaggaag gtactatttt
                                                       pol RNAse/\pol
                                                                integrase
3601 tagatgggat agataaagct caagaagaac atgaaagata tcacagcaat tggagagcaa
3661 tggctagtga ttttaatctg ccacctatag tagcaaagga aatagtagcc agctgtgata
3721 aatgtcagct aaaaggggaa gccatgcatg gacaagtaga ctgtagccca gggatatggc
3781 aattagattg tacacatcta gaaggaaaag taattctggt agcagtccat gtggctagtg
3841 gctacgtaga agcagaagtt attccagcag aaacaggaca ggagacagca tactttctac
3901 taaagctagc aggaagatgg ccagtaaaag tagtacacac agacaatggc agcaatttca
3961 ccagcgctgc ggttaaagca gcctgttggt gggcaaatgt taaacaggaa tttggtattc
4021 cctacaatcc ccaaagtcaa ggagtagtag aatctatgaa taaggaatta aagaaaatca
4081 tagggcaggt aagggagcaa gctgaacatc ttaagacagc agtacaaatg gcagtgttca
4141 ttcacaattt taaaagaaaa ggggggattg gggggtacag tgcaggggaa agaataatag
```

```
     4201 acataatagc atcagactta caaactaaag aattacaaaa acaaattaca aaaattcaaa
     4261 aatttcgggt ttgttacagg gacagcagag atccaatttg gaaaggacca gcaaaactac
     4321 tctggaaagg tgaaggggca gtggtaatac aggacaatag tgatataaaa gtagtaccaa
     4381 gaagaaaagt aaagatcatt aaggattatg gaaaacagat ggcaggtgat gattgtgtgg
                         vif cds start ->
     4441 caggtagaca ggatgaggat tagaacatgg aacagtctag taaaacatca tatgtatatc
                                <- pol cds normal
                                   integrase end
     4501 tcacggagag ctaaaggttg gttttataga catcactatg aaagcaggca tccaaaagta
     4561 agctcagaag tacacatccc aataggggat gctagaatag tagtaagaac atattggggt
     4621 ctgcagacag gagaaaaaga ctggcacttg ggtcatgggg tctccataga atggaggcta
     4681 aaaagatata gcacacaaat agaccctgac ctggcagacc aactaattca tctgcattat
     4741 tttaactgtt tttcagactc tgccataagg aaagccatat tagggcaagt agttagccct
                     3'sj /\
     4801 aggtgtgatt atcaaacagg acacaacaag gtaggatctt tacaatattt agcactaaaa
                                    5'sj / \
     4861 gcattagtaa caccatcaag gatgaagcca cctttgccta gtgttaagaa attagcagag
     4921 gatagatgga acaagcccca gaagaccagg ggccgcagag agagccatac aatgaatgga
 vpr cds start ->
     4981 tgttagatct gttagaagat cttaagcatg aagctgttag acattttcct aggccatggc
              <- vif cds end
     5041 tccatggatt agggcaacat atctatcaca cctatgggga tacttgggaa ggagttgaag
     5101 ctataataag aattttgcag caactactgt ttgtccattt cagaatcggg tgccaacaca
     5161 gcagaatagg cattaatatt cgagggagaa gagtcaggga tggatccggt agatcctagc
                                     tat cds start - >                 <-
                                                                  vpr cds end
     5221 ctagagccct ggaaccatcc gggaagtcag cctaaaactc cttgtaacaa gtgttactgt
     5281 aaagtgtgtt gctatcattg ccaatgctgc tttctgaaca agggcttagg catctcctat
                                                          rev cds start ->
     5341 ggcaggaaga agcggaaacc ccgacgagga actcctcaga gcaataagga tcatcaaaat
     5401 cctataccaa agcagtaagt atcagtaatt agtatatgta atgcagcttt tggaaatctg
tat-rev exon/intron 5'sj/\              vpu cds start ->
     5461 tgcagtagta ggactggtag tagcgctaat catagcaata gttgtgtgga ctatagtagg
     5521 tatagaatat aagaaattgc taaagcaaag aaaaatagac aggttagttg atagaataag
     5581 agaaagagca gaagacagtg gcaatgagag tgatggggat agagaggaat tatccttgct
                           env cds ->
              signal peptide start
     5641 ggtggacatg ggggattatg atcttgggga tgataataat ttgtaatact gcagaaaact
                                         env signal peptide/\env gp120
                                                        <- vpu cds end
     5701 tgtgggttac tgtctactat ggggtaccta tatggaagga tgcaaatacc accttatttt
     5761 gtgcatcaga tgcgaaagca tatgatacag aagtgcataa tgtctgggct acgcatgcct
     5821 gtgtacctac agaccccagc ccacaagaac taaagatgga aaatgtgaca gaagagttta
     5881 acatgtggaa aaataacatg gtagagcaga tgcatacaga tataatcagt ctatgggacc
     5941 aaagcctaaa accatgtgta cagttaaccc ctctctgcgt tactttagat tgtagctata
     6001 acatcaccaa taatatcacc aatagcatca ccaatagctc agttaacatg agagaagaaa
     6061 taaaaaactg ctctttcaat atgaccacag aattaaggga taagaatcgg aaggtatatt
     6121 cactttttta taaacttgat gtagtacaaa ttaataatgg taataacagt agtaatctgt
     6181 atagattaat aaattgtaat acctcagccc ttacacaggc tcgtccaaag gtaacctttg
     6241 agccaattcc catacattat tgtgccccag ctggttatgc gattctaaaa tgtaatgata
     6301 aggagttcaa tggaacaggg ctatgcaaga atgtcagcac agtacaatgc acacatggaa

     6361 tcaggccagt agtgtcaact caactgctgt taaatggcag tttagcagaa ggaaaggtaa
     6421 tgattagatc tgaaaatatc acaaacaatg tcaaaaacat aatagtacaa cttaacgagt
     6481 ctgtaacaat taattgtacc agacctaaca ataatacaag aagaagtgta cgtataggac
           V3 loop start ->
     6541 caggacaaac attctatgca acaggtgata taatagggga tataagacaa gcacattgta
                                                                   <- V3 loop
                                                                      end
     6601 atgtcagtgg gtcacaatgg aataaaactt tacaccaggt agttgaacaa ttaagaaaat
     6661 attggaacaa caatacaata atctttaata gctcctcagg aggggattta gaaattacaa
```

```
6721 cacatagttt taattgtgca ggagaatttt tctattgtaa tacatcaggc ctgtttaata
6781 gtacttgggt aaatggcact accagcagca cgtcaaatgg cactataact ctcccatgca
6841 gaataaagca aattataaat atgtggcaga gagtaggaca agcaatgtat gcccctccca
6901 tccaaggagt aataaagtgt gaatcaaaca ttacaggact aatattaaca agagatggtg
6961 gggttaatag cagtgacagt gaaaccttca gacctggagg aggagatatg agggataatt
7021 ggagaagtga attatataag tataaggtag taaaaattga accactagga gtagcaccca
7081 ccaaggcaag gagaagagtg gtggagagag aaaaaagagc agttacactg ggagctgtat
                                   env gp120/\env gp41
7141 tcattgggtt cttaggaaca gctggaagca caatgggcgc ggcgtcaata acgctgacgg
7201 tacaggccag aaaattattg tctggcatag tgcaacagca aagcaatttg ctgagggcta
7261 tagaggctca acagcatctg ttgaaactca ctgtctgggg cattaaacag ctccaggcaa
7321 gagtcctggc tgtggaaaga tacctaaggg atcaacagct cctaggaatt tggggctgct
7381 ctggaaaact catctgcccc actaatgtgc cctggaactc tagttggagt aataaatctc
7441 tagatgaaat atgggaaaac atgacctggc tgcaatggga taaagaaatt agcaattaca
7501 caatcaaaat atatgagcta attgaagaat cgcagatcca gcaggaaagg aatgaaaaag
7561 acttactgga gttggacaag tgggcaagtc tgtggaattg gtttgacata tcaaaatggc
7621 tgtggtatat aaaaatattt ataatgatag taggaggcct aataggatta agaatagttt
7681 ttgctgtgct ttctgtaata aatagagtta ggcagggata ctcaccccta tcgtttcaga
                                   tat and rev cds resume (exon 2) /\
7741 cccatacccc gaacccaagg ggactcgaca ggcccggaag aatcgaagaa gaaggtggag
7801 agcaagacag aggcagatcg atacgcttag tgagcgggtt cttagcactt gcctgggacg
                                  <- tat cds end
7861 acctgcggaa cctgtgcctc ttcagctacc accgattgag agacttcatc ttgattgcag
7921 cgaggactgt ggaacttccg ggacacagca gtctcaaggg gttgagactg gggtgggaag
7981 gactcaagta tctggggaat ctcctgttgt attggggtcg ggaactaaaa attagtgcta
                                                                <- rev cds end
8041 ttaatttgct tgataccata gcaatagcag tagctggctg gacagatagg gttatagaaa
8101 cagtacaaag gcttggtaga gctattctca acatacctag aagaatcagg cagggcttcg
8161 aaagggcttt actataacat gggtaacaag tggtcaaaga gttgcatagt gggatggcct
                     <- env cds
                        gp41 end
          nef cds start ->
8221 gaggttaggg aaagaataag acaaactcct acagcagcaa gggaaagaac aagacaagcc
           ---- ---------- ---------- ---------- ---------- -------
                  direct repeat copy 1           direct repeat copy 2
8281 cctacagcag caaaaggagt aggagcagta tctcaagatt tagataaaca tggagcagtc
8341 acaagcagca atgtaaatca ccctagttgc gtctggctgg aagcgcagga ggaagaagag
8401 gtaggctttc cagtcagacc acaggtacct ctaaggccaa tgacttacaa ggcagctttc
8461 gatctcggct tctttttaaa agaaaagggg ggactggatg ggttaattta ctccaagaaa
                   3′ LTR U3 region start ->
8521 agacaagaaa tccttgatct gtgggtctac cacacacaag gctacttccc tgattggcag
8581 aattacacac cagggccagg gatcagatac ccactaacat ttggatggtg cttcaagcta
8641 gtaccagtgg atgaagatga agtagaggaa gctactggag gagagaacaa tagcctatta
8701 caccctatat gccaacatgg aatggatgat gaggagaaag aaacattaag gtggaagttt
8761 gacagcagcc tggcacgagt acacaaagca agagagctgc atccggagtt ttacaaagac
8821 tgctgacaca ggagttgctg actgggactt tccgctgggg actttccagg ggaggtgtgg
         <- nef cds end       ------- ---   ------------
                                 NF kappa B sites
8881 tttgggcgga gttggggagt ggctaaccct cagatgctgc atataagcag ctgcttctcg
8941 cttgtactgg gtctctcttg gtagaccaga tcgagcctgg
LTR U3 region/\R rpt
```