

Human immunodeficiency virus 1, complete genome

NCBI Reference Sequence: NC_001802.1

[FASTA Graphics](#)

[Go to:](#)

LOCUS NC_001802 9181 bp ss-RNA linear VRL 13-AUG-2018

DEFINITION Human immunodeficiency virus 1, complete genome.

ACCESSION NC_001802

VERSION NC_001802.1

DBLINK BioProject: [PRJNA485481](#)

KEYWORDS RefSeq.

SOURCE Human immunodeficiency virus 1 (HIV-1)

ORGANISM [Human immunodeficiency virus 1](#)
Viruses; Riboviria; Pararnavirae; Artverviricota; Revtraviricetes;
Ortervirales; Retroviridae; Orthoretrovirinae; Lentivirus.

REFERENCE 1 (bases 1 to 9181)

AUTHORS Martoglio,B., Graf,R. and Dobberstein,B.

TITLE Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin

JOURNAL EMBO J. 16 (22), 6636-6645 (1997)

PUBMED [9362478](#)

REFERENCE 2 (bases 1 to 9181)

AUTHORS Petropoulos,C.J.

TITLE Appendix 2: Retroviral taxonomy, protein structure, sequences, and genetic maps

JOURNAL (in) Coffin,J.M. (Ed.); RETROVIRUSES: 757; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY, USA (1997)

REFERENCE 3 (bases 1 to 9181)

CONSRTM NCBI Genome Project

TITLE Direct Submission

JOURNAL Submitted (07-AUG-2002) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

REFERENCE 4 (bases 1 to 9181)

AUTHORS Chappey,C.

TITLE Direct Submission

JOURNAL Submitted (15-MAR-1999) NIH, NLM, Rockville Pike, Bethesda, MD 20894, USA

REMARK Sequence update by submitter

REFERENCE 5 (bases 1 to 9181)

AUTHORS Chappey,C.

TITLE Direct Submission

JOURNAL Submitted (12-NOV-1997) NIH, NLM, Rockville Pike, Bethesda, MD

20894, USA

COMMENT REVIEWED [REFSEQ](#): This record has been curated by NCBI staff. The reference sequence was derived from [AF033819](#).
The annotation of this sequence was corrected and updated with
the
kind help of Dr. Colombe Chappay (ViroLogic Inc., South San Francisco, CA USA) and Roger Ptak (Southern Research Institute, Frederick, MD USA).
COMPLETENESS: full length.

FEATURES

source	Location/Qualifiers
	1..9181
	/organism="Human immunodeficiency virus 1"
	/mol_type="genomic RNA"
	/db_xref="taxon: 11676 "
	/note="strain for reference annotation"
misc feature	1..96
	/note="repeat; positions of RNA transcription initialization and polyadenylation; Region: R"
regulatory	73..78
	/regulatory_class="polyA_signal_sequence"
	/note="both 5' and 3' poly A signals are transcribed
	RNA, but the 5' one is suppressed"
5'UTR	97..181
primer bind	182..199
gene	336..4642
	/gene="gag-pol"
	/locus_tag="HIV1gp1"
	/db_xref="GeneID: 155348 "
CDS	join(336..1637,1637..4642)
	/gene="gag-pol"
	/locus_tag="HIV1gp1"
	/ribosomal_slippage
	/note="fusion protein consisting of the viral structural proteins and enzymes; cleaved by the viral protease into individual mature proteins; The processing products of
the	Gag and Gag-Pol polyproteins were annotated with the
help	of Pettit et al., 2003 and references therein; Pr160"
	/codon_start=1
	/product="Gag-Pol"
	/protein_id=" NP_057849.4 "
	/db_xref="GeneID: 155348 "

/translation="MGARASVLSGGELDRWEKIRLRPGGKKKYKCLKHIVWASRELERF
AVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALD
KIEEEQNKSKKKAQQAADTGHSNQVSQNYPIVQNIQGQMVHQAI SPRTLNAWVKVVE
EKAFSPEVIPMFSALSEGATPQDLNLTMLNTVGGHQAMQMLKETINEEAAEWDRVHPV
HAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNPPPIPVGEIYKRWIILGLNKIVRM
YSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTIL

KALGPAATLEEMMTACQGVGGPGHKARVLAEMSQVTNSATIMMQRGNFRNQKIVKC
FNCGKEGHTARNCRAPRKKGCWKCCKEGHQMKDCTERQANFLREDLAFLLQGKAREFSS
EQTRANSPTRRELQVWGRDNNSPSEAGADRQGTVSFNFPQVTLWQRPLVTIKIGGQLK
EALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVG
PTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPMDGPKVKQWPLTEEKIKALVEI
CTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIP
HPAGLKKKKSVTVLVDVGDAYFSVPLDEDFRKYTAFTIIPSINNETPGIRYQYNVLPQGW
KGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDLIVGSDLEIGQHRTKIEELRQHLLR
WGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQ
IYPGIKVRQLCKLLRGTKALTEVIPLTEEALELAENREILKEPVHGVYYDPSKDLIA
EIQKQGQGWTYQIYQEPFKNLKTGKYARMGAHTNDVKQLTEAVQKITTESIWIWVK
TPKFKLPIQKETWETWWEYQATWIPEWEFVNTPLVWYQLEKEPIVGAETFYVD
GAANRETKLGKAGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQY
ALGIIQAQPDQSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAIRKV
LFLDGDIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCS
PGIWQLDCTHLEGKIVLVAHVVASGYIEAEVI PAETGQETAYFLLKLAGRWPVKTIHT
DNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVSMNKELKKIIGQVRDQAEHLK
TAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSR
NPLWKGPALLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMGDDCVASRQDED"

[mat peptide](#) join(1632..1637,1637..1798)
/gene="gag-pol"
/locus_tag="HIV1gp1"
/product="Gag-Pol Transframe peptide"
/experiment="DESCRIPTION:[PMID:15527852]"
/note="the Glu-Asp-Leu tripeptide (positions 4-6) is a specific inhibitor of the HIV-1 protease. Involved in regulation of the protease-mediated polyprotein processing; alternative p6 protein; p6*"
/protein_id="NP_787043.1"

[mat peptide](#) 1655..4639
/gene="gag-pol"
/locus_tag="HIV1gp1"
/product="Pol"
/note="unprocessed Pol polyprotein; includes part of the transframe peptide, protease, reverse transcriptase and integrase domains."
/protein_id="NP_789740.1"

[mat peptide](#) 1799..2095

polyprotein.

```

/gene="gag-pol"
/locus_tag="HIV1gp1"
/product="aspartic peptidase"
/experiment="DESCRIPTION:[PMID:2537531]"
/experiment="DESCRIPTION:[PMID:2548279]"
/experiment="DESCRIPTION:[PMID:3290901]"
/note="The proteinase domain of Gag-Pol (in the form of
homodimer) mediates all the cleavages in the

```

Cleaves itself from the polyprotein late in particle assembly; protease"

[mat peptide](#)

```

/protein_id="NP_705926.1"
2096..3775
/gene="gag-pol"
/locus_tag="HIV1gp1"
/product="p66 subunit"
/experiment="DESCRIPTION:[PMID:1374166]"
/experiment="EXISTENCE:[PMID:4316300]"
/note="transcribes single stranded viral RNA genome into
double stranded proviral DNA; HIV-1 reverse

```

transcriptase

is composed of the p66 subunit (this protein) and the

p51

subunit that lacks the RNase H domain of the larger subunit"

[mat peptide](#)

```

/protein_id="NP_705927.1"
2096..3415
/gene="gag-pol"
/locus_tag="HIV1gp1"
/product="reverse transcriptase p51 subunit"
/note="HIV-1 reverse transcriptase is composed of the

```

p66

subunit and the p51 subunit (this protein) that lacks

the

RNase H domain of the larger subunit"

[mat peptide](#)

```

/protein_id="NP_789739.1"
3776..4639
/gene="gag-pol"
/locus_tag="HIV1gp1"
/product="integrase"
/experiment="DESCRIPTION:[PMID:7983732]"
/experiment="DESCRIPTION:[PMID:8035478]"
/note="mediates integration of the viral DNA into the
infected cell chromosome"

```

[gene](#)

```

/protein_id="NP_705928.1"
336..1838
/gene="gag"
/locus_tag="HIV1gp2"
/db_xref="GeneID:155030"

```

[CDS](#)

```

336..1838
/gene="gag"
/locus_tag="HIV1gp2"
/note="The processing products of the Gag and Gag-Pol
polyproteins were annotated with the help of Pettit et
al., 2003 and references therein"
/codon_start=1
/product="Pr55 (Gag) "

```

/protein_id="[NP_057850.1](#)"
/db_xref="GeneID:[155030](#)"

/translation="MGARASVLSGGELDRWEKIRLRPGGKKKYKLGKLVWASRELERF
AVNPGLLETSEGCRQILGQLQPSLQGTGSEELRSLYNTVATLYCVHQRIEIKDTKEALD
KIEEEQNKSKKKAQQAADTGHNSQVSNQNYPIVQNIQGQMVHQAI SPRTLNAWVKVVE
EKAFSPEV I PMFSALSEGATPQDLN TMLNTVGGHQAMQMLKETINEEAAEWDRVHPV
HAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNPPPIPVGEIYKRWIILGLNKIVRM
YSPTSILDIRQGPKPEFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQANPDCKTIL
KALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQKIVKC
FNCGKEGHTARNCRAPRKKGCWKCCKGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFLQ
SRPEPTAPPEESFRSGVETTTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQ"

[mat peptide](#) 336..731
/gene="gag"
/locus_tag="HIV1gp2"
/product="matrix"
/experiment="DESCRIPTION: [PMID:[12032547](#)]"
/experiment="DESCRIPTION: [PMID:[1710290](#)]"
/experiment="DESCRIPTION: [PMID:[8610175](#)]"
/note="viral structural protein; forms the outer structural shell of HIV-1 virions; involved in the

nuclear

import of the HIV-1 preintegration complex; p17"

[mat peptide](#) 732..1424
/gene="gag"
/locus_tag="HIV1gp2"
/product="capsid"
/experiment="DESCRIPTION: [PMID:[15208690](#)]"
/experiment="DESCRIPTION: [PMID:[16041386](#)]"
/experiment="DESCRIPTION: [PMID:[21248851](#)]"
/note="viral structural protein; forms the core of HIV-1 virions; p24"

[mat peptide](#) 1425..1466
/gene="gag"
/locus_tag="HIV1gp2"
/product="p2"
/note="Processing of Gag-Pol by the protease domain

dimer

starts with cleavage between the p2 and nucleocapsid proteins."

[mat peptide](#) 1467..1631
/gene="gag"
/locus_tag="HIV1gp2"
/product="nucleocapsid"
/experiment="DESCRIPTION: [PMID:[1639074](#)]"
/experiment="DESCRIPTION: [PMID:[7666546](#)]"
/note="viral structural protein; coats the genomic RNA

inside the virion core; binds and delivers full-length viral RNAs into assembling HIV-1 virions; p7"

[mat peptide](#) /protein_id="NP_579881.1"
1632..1679
/gene="gag"
/locus_tag="HIV1gp2"
/product="p1"
/note="important for virus infectivity, protein processing, and genomic RNA dimer stability"

[mat peptide](#) /protein_id="NP_787042.1"
1680..1835
/gene="gag"
/locus_tag="HIV1gp2"
/product="p6"
/experiment="DESCRIPTION:[PMID:10085158]"
/experiment="DESCRIPTION:[PMID:15527852]"
/note="important for incorporation of Vpr into assembling HIV-1 virions; helps mediate efficient virus particle release from infected cells"

[gene](#) /protein_id="NP_579883.1"
4587..5165
/gene="vif"
/locus_tag="HIV1gp3"
/db_xref="GeneID:155459"

[CDS](#) 4587..5165
/gene="vif"
/locus_tag="HIV1gp3"
/note="p23; viral infectivity factor; viral accessory protein important for virus replication in vivo"
/codon_start=1
/product="Vif"
/protein_id="NP_057851.1"
/db_xref="GeneID:155459"

/translation="MENRWQVMIVWQVDRMRIRTWKSLVKHHMYVSGKARGWFYRHHY
ESPHPRISSEVHIPLGDARLVITTYWGLHTGERDWHLGQGVSIWRKKRYSTQVDPEL
ADQLIHLYYFDCFSDSAIRKALLGHIVSPRCEYQAGHNKVGSLQYLALAAALITPKKIK
PPLPSVTKLTEDRWNKPKTKGHRGSHTMNGH"

[gene](#) 5105..5396
/gene="vpr"
/locus_tag="HIV1gp4"
/db_xref="GeneID:155807"

[CDS](#) join(5105..5319,5321..5396)
/gene="vpr"
/locus_tag="HIV1gp4"
/exception="artificial frameshift"
/note="p15; viral protein R; viral accessory protein important for virus replication in vivo; involved in the nuclear import of the HIV-1 preintegration complex; induces G2 cell cycle arrest; influences mutation rates during viral DNA synthesis; An artificial frameshift eliminating the orf-disrupting nucleotide at position 5320 is introduced to obtain the typical HIV-1 Vpr protein"

sequence. For this particular HIV-1 strain, HXB2, only a short (78 amino acid long) variant of the Vpr sequence can

be obtained by translation of nucleotides 5105 through 5341 without the frameshift"

```
/codon_start=1
/product="Vpr"
/protein_id="NP_057852.2"
/db_xref="GeneID:155807"
```

```
/translation="MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQ
HIYETYGDTWAGVEAIIRILQQLLFIFRIGCRHSRIGVTRQRRRANGASRS"
```

[gene](#) 5377..7970

```
/gene="tat"
/locus_tag="HIV1gp5"
/db_xref="GeneID:155871"
```

[CDS](#) join(5377..5591,7925..7970)

```
/gene="tat"
/locus_tag="HIV1gp5"
/note="p14; transcriptional activator; viral regulatory
protein required for virus replication; transactivates
```

the viral LTR promoter through interactions with cellular transcription factors; associated with pathogenic

effects of the virus; the length of Tat varies depending on

virus strain or clade"

```
/codon_start=1
/product="Tat"
/protein_id="NP_057853.1"
/db_xref="GeneID:155871"
```

```
/translation="MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALG
ISYGRKKRRQRRRAHQNSQTHQASLSKQPTSQPRGDPTGPKE"
```

[gene](#) 5516..8199

```
/gene="rev"
/locus_tag="HIV1gp6"
/db_xref="GeneID:155908"
```

[CDS](#) join(5516..5591,7925..8199)

```
/gene="rev"
/locus_tag="HIV1gp6"
/note="p19; regulator of expression of virion proteins;
prevents splicing of viral RNA; shuttles unspliced viral
RNA to the cytoplasm for expression of viral proteins
```

and incorporation of full length viral genomic RNA into virions"

```
/codon_start=1
/product="Rev"
/protein_id="NP_057854.1"
/db_xref="GeneID:155908"
```

```
/translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPNPEGTRQARRNRRR
```

```
WRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDNCNEDCGTSGTQGVGSPQI
LVESPTVLESGTKE"
```

[gene](#) 5608..5856
 /gene="vpu"
 /locus_tag="HIV1gp7"
 /db_xref="GeneID:[155945](#)"

[CDS](#) 5608..5856
 /gene="vpu"
 /locus_tag="HIV1gp7"
 /note="p16; viral protein U; viral accessory protein important for virus replication in vivo; promotes degradation of CD4 and down-regulates cell surface expression of MHC class I proteins; helps mediate efficient virus particle release from infected cells; reported to induce apoptosis by suppressing the nuclear factor kappaB-dependent expression of antiapoptotic factors; may attenuate the level of Env precursor(gp160) biosynthesis; Vpu and gp160 are translated from

different

reading frames of the same bicistronic mRNA"
 /codon_start=1
 /product="Vpu"
 /protein_id="[NP_057855.1](#)"
 /db_xref="GeneID:[155945](#)"

/translation="MQPIPIVAIVALVVAIIIIAIVVWSIVIIIEYRKILRQRKIDRLID
 RLIERAEDSGNESEGEISALVEMGVEMGHAPWDVDDL"

[gene](#) 5771..8341
 /gene="env"
 /locus_tag="HIV1gp8"
 /db_xref="GeneID:[155971](#)"

[CDS](#) 5771..8341
 /gene="env"
 /locus_tag="HIV1gp8"
 /note="gp160; envelope glycoprotein; envelope

polyprotein;

cleaved by cellular proteases into mature proteins gp120
 and gp41"
 /codon_start=1
 /product="Envelope surface glycoprotein gp160,

precursor"

/protein_id="[NP_057856.1](#)"
 /db_xref="GeneID:[155971](#)"

/translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
 WKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVVTENFNMWKNMVE
 QMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
 ISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKVSFEPIPIHYC
 APAGFAILKCNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEDEVVIRSVN
 FTDNAKTIIVQLNTSVEINCTRPNNNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNIS
 RAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
 STWFNSTWSTEGSNTEGSDTITLPCRKIQIINMWQKVGKAMYAPPISGQIRCSSNIT

GLLLTRDGGNSNNESEIFRPGGGDMRDNRSELYKYKVVVKIEPLGVAPTAKARRVVQR
EKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQNNLLRAIEAQHLL
QLTVWGIKQLQARILAVERYLKDQQLLGIWGC SGKLICTTAVPWNASWSNKSLEQIWN
HTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYI
KLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGER
DRDRSIRLVNGSLALIWDDLRSCLFSYHRLRDLILLIVTRIVELLGRRGWEALKYWWN
LLQYWSQELKNSAVSLLNATAIAVAEGTDREVIEVVQGACRAIRHIPRRIRQGLERILL

"
[sig peptide](#) 5771..5854
/gene="env"
/locus_tag="HIV1gp8"
/product="hypothetical protein"
/protein_id="[NP_579893.2](#)"
[mat peptide](#) 5855..7303
/gene="env"
/locus_tag="HIV1gp8"
/product="Envelope surface glycoprotein gp120"
/experiment="DESCRIPTION: [PMID:[24179160](#)]"
/note="mediates binding of HIV-1 to CD4 and cellular
co-receptors; cooperates with gp41 to mediate fusion of
viral membrane with cellular membrane during virus entry
into cells; Envelope surface unit; SU"
/protein_id="[NP_579894.2](#)"
[mat peptide](#) 7304..8338
/gene="env"
/locus_tag="HIV1gp8"
/product="Envelope transmembrane domain"
/note="cooperates with gp120 to mediate fusion of viral
membrane with cellular membrane during virus entry into
cells; Envelope transmembrane glycoprotein gp41; TM"
/protein_id="[NP_579895.1](#)"
[gene](#) complement (6919..7488)
/gene="asp"
/locus_tag="HIV1gp10"
/db_xref="GeneID:[19424028](#)"
[CDS](#) complement (6919..7488)
/gene="asp"
/locus_tag="HIV1gp10"
/note="minus-strand-encoded antisense protein of unknown
function; region is highly conserved and likely
analogous
leucine
to the HTLV-1 encoded, HBZ, a nuclear basic region
zipper (b-ZIP) protein"
/codon_start=1
/product="Asp"
/protein_id="[YP_009028572.1](#)"
/db_xref="GeneID:[19424028](#)"

/translation="MPQTVSCNRCCCASIALSKLFCCTIPDNNCLACTVSVIEAAPI

VLPAAPKNPRNKAPIPTALFSLCTTLLFALVGATPNGSIFTTLYLYNSLLQLSLISPP

PGLKISDSLLLLPPSLVNSSPVIFDEHLICPLMGGAYIAFPTFCHMFIICFILHGRVI

VSLPSVLFDPVSVLQVLLNQVLLNSCVLQ"

[gene](#)

8343..8963

/gene="nef"

/locus_tag="HIV1gp9"

/db_xref="GeneID:[156110](#)"

[CDS](#)

8343..8963

/gene="nef"

/locus_tag="HIV1gp9"

/note="p27; negative factor; viral accessory protein; important for virus replication in vivo; determinant of HIV-1 pathogenesis; down-regulates cell surface CD4 and MHC class I molecules; enhances virus infectivity

through;

interactions with multiple cellular signaling proteins; This particular nucleotide sequence has a premature stop codon in place of a well-conserved tryptophan codon at position 8712-8714 that truncates the HIV1 Nef protein sequence to a 123 amino acids-long N-terminal portion

(not

shown)"

/codon_start=1

/transl_except=(pos:8712..8714,aa:Trp)

/product="Nef"

/protein_id="NP_057857.2"

/db_xref="GeneID:[156110](#)"

/translation="MGGKWSKSSVIGWPTVRERMRAEPAADRVGAAASRDLEKHGAIT

SSNTAATNAACAWLEAQEEEEVGFVPTQVPLRPMTYKAAVDLSHFLKEKGGLEGLIH

SQRRQDILDLDWIYHTQGYFPDQNYTPGPGVRYPLTFGWICYKLVPEPDKIEEANKGE

NTSLLHPVSLHGMDPPEREVLEWRFDLSRLAFHHVARELHPEYFKNC"

[3'UTR](#)

8631..9085

[misc feature](#)

9086..9181

/note="repeat; positions of RNA transcription initialization and polyadenylation; Region: R"

[regulatory](#)

9158..9163

/regulatory_class="polyA_signal_sequence"

/note="both 5' and 3' poly A signals are transcribed

into

RNA, but the 5' one is suppressed"

ORIGIN

```
1 ggtctctctg gttagaccag atctgagcct gggagctctc tggctaacta gggaaaccac
61 tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt
121 gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca
181 gtggcgcccg aacagggacc tgaaagcgaa agggaaacca gaggagctct ctcgacgcag
241 gactcggtt gctgaagcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc
301 aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa
361 gcggggggaga attagatcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat
421 ataaattaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg
481 gcctgttaga aacatcagaa ggctgttagc aaatactggg acagctacaa ccatcccttc
541 agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc
601 atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa
```

661	acaaaagtaa	gaaaaaagca	cagcaagcag	cagctgacac	aggacacagc	aatcaggtca
721	gccaaaatta	ccctatagtg	cagaacatcc	agggggcaaat	ggtacatcag	gccatatcac
781	ctagaacttt	aaatgcatgg	gtaaaagtag	tagaagagaa	ggcttttcagc	ccagaagtga
841	tacccatggt	ttcagcatta	tcagaaggag	ccaccccaca	agatttaaac	accatgctaa
901	acacagtggg	gggacatcaa	gcagccatgc	aaatgttaaa	agagaccatc	aatgaggaag
961	ctgcagaatg	ggatagagtg	catccagtg	atgcagggcc	tattgcacca	ggccagatga
1021	gagaaccaag	gggaagtgc	atagcaggaa	ctactagtac	ccttcaggaa	caaataggat
1081	ggatgacaaa	taatccacct	atcccagtag	gagaaattta	taaaagatgg	ataatcctgg
1141	gattaaataa	aatagtaaga	atgtatagcc	ctaccagcat	tctggacata	agacaaggac
1201	caaaggaacc	cttttagagac	tatgtagacc	ggttctataa	aactctaaga	gccgagcaag
1261	cttcacagga	ggtaaaaaat	tggatgacag	aaaccttggt	ggtccaaaaat	gcgaaccag
1321	attgtaagac	tattttaaaa	gcattgggac	cagcggctac	actagaagaa	atgatgacag
1381	atgtcaggg	agtaggagga	cccggccata	aggcaagagt	tttggctgaa	gcaatgagcc
1441	aagtaacaaa	ttcagctacc	ataatgatgc	agagaggcaa	ttttaggaac	caaagaaaga
1501	ttgttaagt	tttcaattgt	ggcaagaag	ggcacacagc	cagaaattgc	agggccccta
1561	ggaaaaaggg	ctgttggaaa	tgtggaaagg	aaggacacca	aatgaaagat	tgtactgaga
1621	gacaggctaa	ttttttaggg	aagatctggc	cttcctacaa	gggaaggcca	gggaattttc
1681	ttcagagcag	accagagcca	acagccccac	cagaagagag	ccttcaggtct	ggggtagaga
1741	caacaactcc	ccctcagaag	caggagccga	tagacaagga	actgtatcct	ttaacttccc
1801	tcaggtcact	ctttggcaac	gacccctcgt	cacaataaag	ataggggggc	aactaaagga
1861	agctctatta	gatacaggag	cagatgatac	agtattagaa	gaaatgagtt	tgccaggaag
1921	atggaaacca	aaaatgatag	ggggaattgg	aggttttatc	aaagtaagac	agtatgatca
1981	gatactcata	gaaatctgtg	gacataaagc	tataggtaca	gtattagtag	gacctacacc
2041	tgtcaacata	attggaagaa	atctgttgac	tcagattggt	tgcaacttaa	attttcccat
2101	tagccctatt	gagactgtac	cagtaaaatt	aaagccagga	atggatggcc	caaaagttaa
2161	acaatggcca	ttgacagaag	aaaaataaaa	agcattagta	gaaatttgta	cagagatgga
2221	aaaggaaggg	aaaatttcaa	aaattggggc	tgaaaatcca	tacaatactc	cagtatttgc
2281	cataaagaaa	aaagacagta	ctaaatggag	aaaattagta	gatttcagag	aacttaataa
2341	gagaactcaa	gacttctggg	aagttcaatt	aggaatacca	catcccgcag	ggttaaaaaa
2401	gaaaaaatca	gtaacagtac	tggatgtggg	tgatgcata	ttttcagttc	ccttagatga
2461	agacttcagg	aagtatactg	catttaccat	acctagtata	aacaatgaga	caccagggat
2521	tagatatcag	tacaatgtgc	ttccacaggg	atggaaagga	tcaccagcaa	tattccaaag
2581	tagcatgaca	aaaatcttag	agccttttag	aaaacaaaat	ccagacatag	ttatctatca
2641	atacatggat	gatttgtatg	taggatctga	cttagaaaata	gggcagcata	gaacaaaaat
2701	agaggagctg	agacaacatc	tgttgaggtg	gggacttacc	acaccagaca	aaaaacatca
2761	gaaagaacct	ccattccttt	ggatgggtta	tgaactccat	cctgataaat	ggacagtaca
2821	gcctatagtg	ctgccagaaa	aagacagctg	gactgtcaat	gacatacaga	agttagtggg
2881	gaaattgaat	tgggcaagtc	agatttacc	agggattaaa	gtaaggcaat	tatgtaaact
2941	ccttagagga	accaaagcac	taacagaagt	aataccacta	acagaagaag	cagagctaga
3001	actggcagaa	aacagagaga	ttctaaaaga	accagtacat	ggagtgtatt	atgacccatc
3061	aaaagactta	atagcagaaa	tacagaagca	ggggcaaggc	caatggacat	atcaaattta
3121	tcaagagcca	tttaaaaaatc	tgaaaacagg	aaaatatgca	agaatgaggg	gtgcccacac
3181	taatgatgta	aaacaattaa	cagaggcagt	gcaaaaaata	accacagaaa	gcatagtaat
3241	atggggaaaag	actcctaaat	ttaaactgcc	catacaaaaag	gaaacatggg	aaacatgggtg
3301	gacagagtat	tggcaagcca	cctggattcc	tgagtgggag	tttgtttaata	cccctccctt
3361	agtgaaatta	tggtagcagt	tagagaaaga	acccatagta	ggagcagaaa	ccttctatgt
3421	agatggggca	gctaacaggg	agactaaatt	aggaaaagca	ggatatgtta	ctaatagagg
3481	aagacaaaaa	gttgtcacc	taactgacac	aacaaatcag	aagactgagt	tacaagcaat
3541	ttatctagct	ttgcaggatt	cgggattaga	agtaaacata	gtaacagact	cacaatatgc
3601	attaggaatc	attcaagcac	aaccagatca	aagtgaatca	gagttagtca	atcaaataat
3661	agagcagtta	ataaaaaagg	aaaaggctca	tctggcatgg	gtaccagcac	acaaaggaat
3721	tggaggaaat	gaacaagtag	ataaattagt	cagtgtctgga	atcaggaaaag	tactatTTTT
3781	agatggaata	gataaggccc	aagatgaaca	tgagaaatat	cacagtaatt	ggagagcaat
3841	ggctagtgat	tttaacctgc	cacctgtagt	agcaaaaagaa	atagtagcca	gctgtgataa
3901	atgtcagcta	aaaggagaag	ccatgcatgg	acaagtagac	tgtagtccag	gaatatggca
3961	actagattgt	acacatttag	aaggaaaagt	tatcctggta	gcagttcatg	tagccagtg
4021	atatatagaa	gcagaagtta	ttccagcaga	aacagggcag	gaaacagcat	atTTTctTTT

4081 aaaattagca ggaagatggc cagtaaaaaac aatacatact gacaatggca gcaatttcac
4141 cgggtgctacg gttaggggccg cctgttgggtg ggcgggaatc aagcaggaat ttggaattcc
4201 ctacaatccc caaagtcaag gagtagtaga atctatgaat aaagaattaa agaaaattat
4261 aggacaggta agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat
4321 ccacaattttt aaaagaaaag gggggattgg ggggtacagt gcaggggaaa gaatagtaga
4381 cataatagca acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaa
4441 ttttcggggtt tattacaggg acagcagaaa tccacttttg aaaggaccag caaagctcct
4501 ctggaaaggt gaaggggcag tagtaataca agataatagt gacataaaaag tagtgccaag
4561 aagaaaagca aagatcatta gggattatgg aaaacagatg gcaggtgatg attgtgtggc
4621 aagtagacag gatgaggatt agaacatgga aaagtttagt aaaacaccat atgtatgttt
4681 cagggaaagc taggggatgg ttttatagac atcactatga aagccctcat ccaagaataa
4741 gttcagaagt acacatccca ctaggggatg ctagattgggt aataacaaca tattgggggtc
4801 tgcatacag agaaagagac tggcatttgg gtcagggagt ctccatagaa tggaggaaaa
4861 agagatatag cacacaagta gaccctgaac tagcagacca actaattcat ctgtattact
4921 ttgactgttt ttcagactct gctataagaa aggccttatt aggacacata gttagcccta
4981 ggtgtgaata tcaagcagga cataacaagg taggatctct acaatacttg gcactagcag
5041 cattaataac accaaaaaag ataaagccac ctttgcttag tgttacgaaa ctgacagagg
5101 atagatggaa caagccccag aagaccaagg gccacagagg gagccacaca atgaatggac
5161 actagagctt ttagaggagc ttaagaatga agctgttaga cattttccta ggatttggct
5221 ccatggctta gggcaacata tctatgaaac ttatggggat acttgggcag gagtgggaagc
5281 cataataaga attctgcaac aactgctgtt tatccatttt cagaattggg tgtcgacata
5341 gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tccatagacta
5401 gagccctgga agcatccagg aagtcagcct aaaactgctt gtaccaattg ctattgtaaa
5461 aagtgttgct ttcattgcca agtttgtttc ataacaaaag ccttaggcat ctccatggc
5521 aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct
5581 ctatcaaagc agtaagtagt acatgtaatg caacctatac caatagtagc aatagtagca
5641 ttagtagtag caataataat agcaatagtt gtgtgggtcca tagtaatcat agaataatagg
5701 aaaatattaa gacaaaagaa aatagacaggg ttaattgata gactaataga aagagcagaa
5761 gacagtggca atgagagtga aggagaaata tcagcacttg tggagatggg ggtggagatg
5821 gggcaccatg ctccctggga tgttgatgat ctgtagtgtc acagaaaaat tgtgggtcac
5881 agtctattat ggggtacctg tgtggaagga agcaaccacc actctatttt gtgcatcaga
5941 tgctaaagca tatgatacag aggtacataa tgtttgggcc acacatgcct gttaccac
6001 agacccaac ccacaagaag tagtattgggt aaatgtgaca gaaaatttta acatgtggaa
6061 aatgacatg gtagaacaga tgcattgagga tataatcagt ttatgggatc aaagcctaaa
6121 gccatgtgta aaattaaccc cactctgtgt tagtttaaaag tgcactgatt tgaagaatga
6181 tactaatacc aatagtagta gggggagaat gataatggag aaaggagaga taaaaactg
6241 ctctttcaat atcagcaca gcataagagg taaggtgcag aaagaatatg cattttttta
6301 taaacttgat ataataccaa tagataatga tactaccagc tataagttga caagttgtaa
6361 cacctcagtc attacacagg cctgtccaaa ggtatccttt gagccaattc ccatacatta
6421 ttgtgccccg gctgggtttt cgattctaaa atgtaataat aagacgttca atggaacagg
6481 accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac
6541 tcaactgctg ttaaatggca gtctagcaga agaagaggta gtaattagat ctgtcaattt
6601 cacggacaat gctaaaacca taatagtaca gctgaacaca tctgtagaaa ttaattgtac
6661 aagacccaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt
6721 tgttacaata ggaaaaatag gaaatatgag acaagcatat tgtaacatta gtagagcaaa
6781 atggaataac acttttaaac agatagctag caaattaaga gaacaatttg gaaataataa
6841 aacaataatc tttaagcaat cctcaggagg ggaccagaa attgtaacgc acagttttta
6901 ttgtggaggg gaatttttct actgtaattc aacacaactg tttaatagta ctgtgtttta
6961 tagtacttgg agtactgaag ggtcaaataa cactgaagga agtgacacaa tccacctccc
7021 atgcagaata aaacaaatta taaacatgtg gcagaaagta ggaaaagcaa tgtatgcccc
7081 tcccatcagt ggacaaatta gatgttcac aaatattaca gggctgctat taacaagaga
7141 tgggtgtaat agcaacaatg agtccgagat ctccagacct ggaggaggag atatgaggga
7201 caattggaga agtgaattat ataaatataa agtagtaaaa attgaacct taggagtagc
7261 acccacciaag gcaaagagaa gagtgggtgca gagagaaaaa agagcagtgga gaataggagc
7321 tttgttctct gggttcttgg gagcagcagg aagcactatg ggcgcagcct caatgacgct
7381 gacggtagac gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag
7441 ggctatttag gcgcaacagc atctgtttgca actcacagtc tggggcatca agcagctcca

7501 ggcaagaatc ctggctgtgg aaagatacct aaaggatcaa cagctcctgg ggatttgggg
7561 ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgctagtt ggagtaataa
7621 atctctggaa cagatttgga atcacacgac ctggatggag tgggacagag aaattaacaa
7681 ttacacaagc ttaatacact ccttaattga agaatcgcaa aaccagcaag aaaagaatga
7741 acaagaatta ttggaattag ataaatgggc aagtttgtgg aattggttta acataacaaa
7801 ttggctgtgg tatataaaat tattcataat gatagtagga ggcttggtag gtttaagaat
7861 agtttttgcg gtactttcta tagtgaatag agttaggcag ggatattcac cattatcgtt
7921 tcagaccac ctcccaacce cgaggggacc cgacaggccc gaaggaatag aagaagaagg
7981 tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg
8041 ggacgatctg cggagcctgt gcctcttcag ctaccaccgc ttgagagact tactcttgat
8101 tgtaacgagg attgtggaac ttctgggacg caggggggtgg gaagccctca aatattgggtg
8161 gaatctccta cagtattgga gtcaggaact aaagaatagt gctgttagct tgctcaatgc
8221 cacagccata gcagtagctg aggggacaga tagggttata gaagtagtac aaggagcttg
8281 tagagctatt cgccacatac ctagaagaat aagacagggc ttggaaagga ttttgctata
8341 agatgggtgg caagtggcca aaaagttagt tgattggatg gcctactgta agggaaagaa
8401 tgagacgagc tgagccagca gcagataggg tgggagcagc atctcgagac ctggaaaaac
8461 atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttggtgc tggctagaag
8521 cacaagagga ggaggaggtg ggttttccag tcacacctca ggtaccttta agaccaatga
8581 cttacaaggc agctgtagat cttagccact ttttaaaaga aaagggggga ctggaagggc
8641 taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct
8701 acttccctga ttagcagaac tacacaccag ggccaggggt cagatatcca ctgaccttg
8761 gatggtgcta caagctagta ccagttgagc cagataagat agaagaggcc aataaaggag
8821 agaacaccag cttgttacac cctgtgagcc tgcattggat ggatgacctg gagagagaag
8881 tgtagagtg gaggtttgac agccgcctag catttcatca cgtggcccga gagctgcatc
8941 cggagtactt caagaactgc tgacatcgag cttgctacaa gggactttcc gctggggact
9001 ttccagggag gcgtggcctg ggcgggactg gggagtggcg agccctcaga tctgcatat
9061 aagcagctgc tttttgctg tactgggtct ctctggttag accagatctg agcctgggag
9121 ctctctggct aactagggaa cccactgctt aagcctcaat aaagcttggc ttgagtgctt
9181 c

//