**Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome**

NCBI Reference Sequence: NC\_045512.2

**Comparison to Bat-SARS-ZXC21 marked red. Region does not match Bat-SARS-ZC45.**

**Bat SARS-like coronavirus Isolate "bat-SL-CoVZXC21" Host "Rhinolophus pusillus"**

**Country "China" Collection date="Jul-2015" SOURCE: NIH GenBank: MG772934.1**

**HIV-1 analogs believed to be the binding sites of the AIDS gp120 antibodies obtained from the NIH AIDS Reagent program.** [**https://www.nature.com/articles/s41598-021-91746-7**](https://www.nature.com/articles/s41598-021-91746-7) **and** [**https://doi.org/10.1101/2020.01.30.927871**](https://doi.org/10.1101/2020.01.30.927871)

**HIV-1 sequences in question: GTNGTKR, YYHKNNKS, GDSSSG, and QTNSPRRA.**

**aagaa – suspected RNA modification precursors** **<https://doi.org/10.1016/j.cell.2020.04.011> and** [**https://doi.org/10.1016/j.bbadis.2020.165878**](https://doi.org/10.1016/j.bbadis.2020.165878)

**Spike protein coding region of the RNA is bolded: 21563 – 25384.**

**SARS-CoV-2 Wuhan 20712-20771 compared to**

**SARS-CoV CDC Patent US7220852 20701-20760**

**aagaatgctattagaaaagtgtgaccttcaaaattatggtgatagtgcaacattacctaa**

**aagaatgcttcttgaaaagtgtgaccttcagaattatggtgaaaatgctgttataccaaa**

[FASTA](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?report=fasta) [Graphics](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?report=graph)

[Go to:](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?report=genbank" \l "goto1798174254_0)

LOCUS NC\_045512 29903 bp ss-RNA linear VRL 18-JUL-2020

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1,

complete genome.

ACCESSION NC\_045512

VERSION NC\_045512.2

DBLINK BioProject: [PRJNA485481](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA485481)

KEYWORDS RefSeq.

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM [Severe acute respiratory syndrome coronavirus 2](https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2697049)

Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;

Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;

Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 29903)

AUTHORS Wu,F., Zhao,S., Yu,B., Chen,Y.M., Wang,W., Song,Z.G., Hu,Y.,

Tao,Z.W., Tian,J.H., Pei,Y.Y., Yuan,M.L., Zhang,Y.L., Dai,F.H.,

Liu,Y., Wang,Q.M., Zheng,J.J., Xu,L., Holmes,E.C. and Zhang,Y.Z.

TITLE A new coronavirus associated with human respiratory disease in

China

JOURNAL Nature 579 (7798), 265-269 (2020)

PUBMED [32015508](https://www.ncbi.nlm.nih.gov/pubmed/32015508)

REMARK Erratum:[Nature. 2020 Apr;580(7803):E7. PMID: 32296181]

REFERENCE 2 (bases 13476 to 13503)

AUTHORS Baranov,P.V., Henderson,C.M., Anderson,C.B., Gesteland,R.F.,

Atkins,J.F. and Howard,M.T.

TITLE Programmed ribosomal frameshifting in decoding the SARS-CoV genome

JOURNAL Virology 332 (2), 498-510 (2005)

PUBMED [15680415](https://www.ncbi.nlm.nih.gov/pubmed/15680415)

REFERENCE 3 (bases 29728 to 29768)

AUTHORS Robertson,M.P., Igel,H., Baertsch,R., Haussler,D., Ares,M. Jr. and

Scott,W.G.

TITLE The structure of a rigorously conserved RNA element within the SARS

virus genome

JOURNAL PLoS Biol. 3 (1), e5 (2005)

PUBMED [15630477](https://www.ncbi.nlm.nih.gov/pubmed/15630477)

REFERENCE 4 (bases 29609 to 29657)

AUTHORS Williams,G.D., Chang,R.Y. and Brian,D.A.

TITLE A phylogenetically conserved hairpin-type 3' untranslated region

pseudoknot functions in coronavirus RNA replication

JOURNAL J. Virol. 73 (10), 8349-8355 (1999)

PUBMED [10482585](https://www.ncbi.nlm.nih.gov/pubmed/10482585)

REFERENCE 5 (bases 1 to 29903)

CONSRTM NCBI Genome Project

TITLE Direct Submission

JOURNAL Submitted (17-JAN-2020) National Center for Biotechnology

Information, NIH, Bethesda, MD 20894, USA

REFERENCE 6 (bases 1 to 29903)

AUTHORS Wu,F., Zhao,S., Yu,B., Chen,Y.-M., Wang,W., Hu,Y., Song,Z.-G.,

Tao,Z.-W., Tian,J.-H., Pei,Y.-Y., Yuan,M.L., Zhang,Y.-L.,

Dai,F.-H., Liu,Y., Wang,Q.-M., Zheng,J.-J., Xu,L., Holmes,E.C. and

Zhang,Y.-Z.

TITLE Direct Submission

JOURNAL Submitted (05-JAN-2020) Shanghai Public Health Clinical Center &

School of Public Health, Fudan University, Shanghai, China

COMMENT REVIEWED [REFSEQ](https://www.ncbi.nlm.nih.gov/RefSeq/): This record has been curated by NCBI staff. The

reference sequence is identical to [MN908947](https://www.ncbi.nlm.nih.gov/nuccore/MN908947).

On Jan 17, 2020 this sequence version replaced [NC\_045512.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.1).

Annotation was added using homology to SARSr-CoV NC\_004718.3. ###

Formerly called 'Wuhan seafood market pneumonia virus.' If you have

questions or suggestions, please email us at info@ncbi.nlm.nih.gov

and include the accession number NC\_045512.### Protein structures

can be found at

<https://www.ncbi.nlm.nih.gov/structure/?term=sars-cov-2.###> Find

all other Severe acute respiratory syndrome coronavirus 2

(SARS-CoV-2) sequences at

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

##Assembly-Data-START##

Assembly Method :: Megahit v. V1.1.3

Sequencing Technology :: Illumina

##Assembly-Data-END##

COMPLETENESS: full length.

FEATURES Location/Qualifiers

source 1..29903

/organism="Severe acute respiratory syndrome coronavirus

2"

/mol\_type="genomic RNA"

/isolate="Wuhan-Hu-1"

/host="Homo sapiens"

/db\_xref="taxon:[2697049](https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2697049)"

/country="China"

/collection\_date="Dec-2019"

[5'UTR](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=1&to=265) 1..265

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=266&to=21555) 266..21555

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/db\_xref="GeneID:[43740578](https://www.ncbi.nlm.nih.gov/gene/43740578)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?location=266:13468,13468:21555) join(266..13468,13468..21555)

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/ribosomal\_slippage

/note="pp1ab; translated by -1 ribosomal frameshift"

/codon\_start=1

/product="ORF1ab polyprotein"

/protein\_id="[YP\_009724389.1](https://www.ncbi.nlm.nih.gov/protein/1796318597)"

/db\_xref="GeneID:[43740578](https://www.ncbi.nlm.nih.gov/gene/43740578)"

/translation="**MESLVPGFNEKTHVQLSLPVLQV**RDVLVRGFGDSVEEVLSEARQ

HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE

TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN

WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ

LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP

LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG

DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG

LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL

LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN

FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA

ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK

YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE

EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG

YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA

PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA

EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM

LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN

TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT

FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH

NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK

WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG

DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ

IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT

FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST

KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD

IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS

VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI

KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA

ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLET

IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL

MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE

CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP

INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI

NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN

TFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC

LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI

WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL

KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA

NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP

RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVA

YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSG

RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL

AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN

DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE

EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL

AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG

LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL

KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC

GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN

VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV

LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW

LLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL

LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR

TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR

GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL

VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL

LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL

CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK

SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN

NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA

DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA

CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP

KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK

AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH

PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA

DAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD

EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQR

LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYA

NLGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVV

DSYYSLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQ

TYHPNCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRE

LGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQ

TVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDI

RQLLFVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQ

DALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAAT

RGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLAR

KHTTCCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQ

AVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMM

ILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCS

QHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKH

PNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTV

LQAVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCD

VTDVTQLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDW

TNAGDYILANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKP

RPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSH

TVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGK

SHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKF

KVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQ

LPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLK

AHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNA

VASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSD

RDLYDKLQFTSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKT

EGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEG

CHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQFKHLIP

LMYKGLPWNVVRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPERTCCL

CDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHGNA

HVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK

FPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCL

FWNCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQ

LPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYN

MMISAGFSLWVYKQFDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNT

VYTKVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDIAANTVIWDY

KRDAPAHISTIGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEG

SVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFK

PRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESP

FELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKV

TIDYTEISFMLWCKDGHVETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGD

SATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP

TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEG

FFTYICGFIQQKLALGGSVAIKITEHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFL

IGCNYLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKE

GQINDMILSLLSKGRLIIRENNRVVISSDVLVNN"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=1&to=180) 266..805

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="leader protein"

/note="nsp1; produced by both pp1a and pp1ab"

/protein\_id="[YP\_009725297.1](https://www.ncbi.nlm.nih.gov/protein/1802476805)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=181&to=818) 806..2719

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp2"

/note="produced by both pp1a and pp1ab"

/protein\_id="[YP\_009725298.1](https://www.ncbi.nlm.nih.gov/protein/1802476806)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=819&to=2763) 2720..8554

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp3"

/note="former nsp1; conserved domains are: N-terminal

acidic (Ac), predicted phosphoesterase, papain-like

proteinase, Y-domain, transmembrane domain 1 (TM1),

adenosine diphosphate-ribose 1''-phosphatase (ADRP);

produced by both pp1a and pp1ab"

/protein\_id="[YP\_009725299.1](https://www.ncbi.nlm.nih.gov/protein/1802476807)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=2764&to=3263) 8555..10054

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp4"

/note="nsp4B\_TM; contains transmembrane domain 2 (TM2);

produced by both pp1a and pp1ab"

/protein\_id="[YP\_009725300.1](https://www.ncbi.nlm.nih.gov/protein/1802476808)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=3264&to=3569) 10055..10972

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="3C-like proteinase"

/note="nsp5A\_3CLpro and nsp5B\_3CLpro; main proteinase

(Mpro); mediates cleavages downstream of nsp4. 3D

structure of the SARSr-CoV homolog has been determined

(Yang et al., 2003); produced by both pp1a and pp1ab"

/protein\_id="[YP\_009725301.1](https://www.ncbi.nlm.nih.gov/protein/1802476809)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=3570&to=3859) 10973..11842

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp6"

/note="nsp6\_TM; putative transmembrane domain; produced by

both pp1a and pp1ab"

/protein\_id="[YP\_009725302.1](https://www.ncbi.nlm.nih.gov/protein/1802476810)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=3860&to=3942) 11843..12091

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp7"

/note="produced by both pp1a and pp1ab"

/protein\_id="[YP\_009725303.1](https://www.ncbi.nlm.nih.gov/protein/1802476811)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=3943&to=4140) 12092..12685

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp8"

/note="produced by both pp1a and pp1ab"

/protein\_id="[YP\_009725304.1](https://www.ncbi.nlm.nih.gov/protein/1802476812)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=4141&to=4253) 12686..13024

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp9"

/note="ssRNA-binding protein; produced by both pp1a and

pp1ab"

/protein\_id="[YP\_009725305.1](https://www.ncbi.nlm.nih.gov/protein/1802476813)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=4254&to=4392) 13025..13441

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp10"

/note="nsp10\_CysHis; formerly known as growth-factor-like

protein (GFL); produced by both pp1a and pp1ab"

/protein\_id="[YP\_009725306.1](https://www.ncbi.nlm.nih.gov/protein/1802476814)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=4393&to=5324) join(13442..13468,13468..16236)

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="RNA-dependent RNA polymerase"

/note="nsp12; NiRAN and RdRp; produced by pp1ab only"

/protein\_id="[YP\_009725307.1](https://www.ncbi.nlm.nih.gov/protein/1802476815)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=5325&to=5925) 16237..18039

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="helicase"

/note="nsp13\_ZBD, nsp13\_TB, and nsp\_HEL1core; zinc-binding

domain (ZD), NTPase/helicase domain (HEL), RNA

5'-triphosphatase; produced by pp1ab only"

/protein\_id="[YP\_009725308.1](https://www.ncbi.nlm.nih.gov/protein/1802476816)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=5926&to=6452) 18040..19620

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="3'-to-5' exonuclease"

/note="nsp14A2\_ExoN and nsp14B\_NMT; produced by pp1ab

only"

/protein\_id="[YP\_009725309.1](https://www.ncbi.nlm.nih.gov/protein/1802476817)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=6453&to=6798) 19621..20658

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="endoRNAse"

/note="nsp15-A1 and nsp15B-NendoU; produced by pp1ab only"

/protein\_id="[YP\_009725310.1](https://www.ncbi.nlm.nih.gov/protein/1802476818)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009724389.1?from=6799&to=7096) 20659..21552

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="2'-O-ribose methyltransferase"

/note="nsp16\_OMT; 2'-o-MT; produced by pp1ab only"

/protein\_id="[YP\_009725311.1](https://www.ncbi.nlm.nih.gov/protein/1802476819)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=266&to=13483) 266..13483

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/note="pp1a"

/codon\_start=1

/product="ORF1a polyprotein"

/protein\_id="[YP\_009725295.1](https://www.ncbi.nlm.nih.gov/protein/1802476803)"

/db\_xref="GeneID:[43740578](https://www.ncbi.nlm.nih.gov/gene/43740578)"

/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ

HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE

TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN

WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ

LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP

LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG

DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG

LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL

LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN

FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA

ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL

KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV

NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII

FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK

YCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEK

CSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEWSMATYYLFDESGEF

KLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPLEFGATSAALQPE

EEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQTIEVNSFSG

YLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKATNNAMQVESD

DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEVLLA

PLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIA

EIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGN

LHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREM

LAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLIN

TLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

KTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVIT

FDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPH

NSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK

WADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELG

DVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQ

IPCTCGKQATKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSK

ETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYY

KKDNSYFTEQPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVT

FFPDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWST

KPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGD

IILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLGLKTLATHGLAAVNS

VPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFTRSTNSRI

KASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA

ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLET

IQITISSFKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWL

MWLIINLVQMAPISAMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVE

CTTIVNGVRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRP

INPTDQSSYIVDSVTVKNGSIHLYFDKAGQKTYERHSLSHFVNLDNLRANNTKGSLPI

NVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLDQALVSDVGDSAEVAVKMFDAYVN

TFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSDVETKDVVEC

LKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALI

WNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKIVNNWL

KQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDTCFA

NKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFLP

RVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVA

YESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSG

RWVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCL

AYYFMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTN

DVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFE

EAALCTFLLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHL

AKALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNG

LWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVL

KLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSC

GSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVN

VLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAV

LDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQSAVKRTIKGTHHW

LLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFVKHKHAFLCLFL

LPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAVVLLILMTAR

TVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVTTVMFLAR

GIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLGVYDYL

VSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQSKMSDVKCTSVVL

LSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSMQGAVDINKL

CEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAK

SEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN

NIINNARDGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDA

DSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAAGTTQTA

CTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFVTDTP

KGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDAAK

AYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH

PNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQSA

DAQSFLNGFAV"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=1&to=180) 266..805

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="leader protein"

/note="nsp1; produced by both pp1a and pp1ab"

/protein\_id="[YP\_009742608.1](https://www.ncbi.nlm.nih.gov/protein/1826688918)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=181&to=818) 806..2719

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp2"

/note="produced by both pp1a and pp1ab"

/protein\_id="[YP\_009742609.1](https://www.ncbi.nlm.nih.gov/protein/1826688919)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=819&to=2763) 2720..8554

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp3"

/note="former nsp1; conserved domains are: N-terminal

acidic (Ac), predicted phosphoesterase, papain-like

proteinase, Y-domain, transmembrane domain 1 (TM1),

adenosine diphosphate-ribose 1''-phosphatase (ADRP);

produced by both pp1a and pp1ab"

/protein\_id="[YP\_009742610.1](https://www.ncbi.nlm.nih.gov/protein/1826688920)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=2764&to=3263) 8555..10054

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp4"

/note="nsp4B\_TM; contains transmembrane domain 2 (TM2);

produced by both pp1a and pp1ab"

/protein\_id="[YP\_009742611.1](https://www.ncbi.nlm.nih.gov/protein/1826688921)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=3264&to=3569) 10055..10972

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="3C-like proteinase"

/note="nsp5A\_3CLpro and nsp5B\_3CLpro; main proteinase

(Mpro); mediates cleavages downstream of nsp4. 3D

structure of the SARSr-CoV homolog has been determined

(Yang et al., 2003); produced by both pp1a and pp1ab"

/protein\_id="[YP\_009742612.1](https://www.ncbi.nlm.nih.gov/protein/1826688922)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=3570&to=3859) 10973..11842

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp6"

/note="nsp6\_TM; putative transmembrane domain; produced by

both pp1a and pp1ab"

/protein\_id="[YP\_009742613.1](https://www.ncbi.nlm.nih.gov/protein/1826688923)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=3860&to=3942) 11843..12091

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp7"

/note="produced by both pp1a and pp1ab"

/protein\_id="[YP\_009742614.1](https://www.ncbi.nlm.nih.gov/protein/1826688924)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=3943&to=4140) 12092..12685

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp8"

/note="produced by both pp1a and pp1ab"

/protein\_id="[YP\_009742615.1](https://www.ncbi.nlm.nih.gov/protein/1826688925)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=4141&to=4253) 12686..13024

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp9"

/note="ssRNA-binding protein; produced by both pp1a and

pp1ab"

/protein\_id="[YP\_009742616.1](https://www.ncbi.nlm.nih.gov/protein/1826688926)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=4254&to=4392) 13025..13441

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp10"

/note="nsp10\_CysHis; formerly known as growth-factor-like

protein (GFL); produced by both pp1a and pp1ab"

/protein\_id="[YP\_009742617.1](https://www.ncbi.nlm.nih.gov/protein/1826688927)"

[mat\_peptide](https://www.ncbi.nlm.nih.gov/protein/YP_009725295.1?from=4393&to=4405) 13442..13480

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/product="nsp11"

/note="produced by pp1a only"

/protein\_id="[YP\_009725312.1](https://www.ncbi.nlm.nih.gov/protein/1802476820)"

[stem\_loop](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=13476&to=13503) 13476..13503

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/inference="COORDINATES:

profile:Rfam-release-14.1:RF00507,Infernal:1.1.2"

/function="Coronavirus frameshifting stimulation element

stem-loop 1"

[stem\_loop](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=13488&to=13542) 13488..13542

/gene="ORF1ab"

/locus\_tag="GU280\_gp01"

/inference="COORDINATES:

profile:Rfam-release-14.1:RF00507,Infernal:1.1.2"

/function="Coronavirus frameshifting stimulation element

stem-loop 2"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=21563&to=25384) 21563..25384

/gene="S"

/locus\_tag="GU280\_gp02"

/gene\_synonym="spike glycoprotein"

/db\_xref="GeneID:[43740568](https://www.ncbi.nlm.nih.gov/gene/43740568)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=21563&to=25384) 21563..25384

/gene="S"

/locus\_tag="GU280\_gp02"

/gene\_synonym="spike glycoprotein"

/note="structural protein; spike protein"

/codon\_start=1

/product="surface glycoprotein"

/protein\_id="[YP\_009724390.1](https://www.ncbi.nlm.nih.gov/protein/1796318598)"

/db\_xref="GeneID:[43740568](https://www.ncbi.nlm.nih.gov/gene/43740568)"

/translation="MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFR

SSVLHSTQDLFLPFFSNVTWFHAIHVS**GTNGTKR**FDNPVLPFNDGVYFASTEKSNIIR

GWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGV**YYHKNNKS**WMESEFRVY

SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQ

GFSALEPLVDLPIGINITRFQTLLALHRSYLTP**GDSSSG**WTAGAAAYYVGYLQPRTFL

LKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN

LCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCF

TNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYN

YLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY

RVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFG

RDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAI

HADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQT**QTNSPR**

**RA**RSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTM

YICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG

GFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFN

GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQN

VLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGA

ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMS

ECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAH

FPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELD

SFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELG

KYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSE

PVLKGVKLHYT"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=25393&to=26220) 25393..26220

/gene="ORF3a"

/locus\_tag="GU280\_gp03"

/db\_xref="GeneID:[43740569](https://www.ncbi.nlm.nih.gov/gene/43740569)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=25393&to=26220) 25393..26220

/gene="ORF3a"

/locus\_tag="GU280\_gp03"

/codon\_start=1

/product="ORF3a protein"

/protein\_id="[YP\_009724391.1](https://www.ncbi.nlm.nih.gov/protein/1796318599)"

/db\_xref="GeneID:[43740569](https://www.ncbi.nlm.nih.gov/gene/43740569)"

/translation="MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFG

WLIVGVALLAVFQSASKIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLE

APFLYLYALVYFLQSINFVRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPY

NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQ

LSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=26245&to=26472) 26245..26472

/gene="E"

/locus\_tag="GU280\_gp04"

/db\_xref="GeneID:[43740570](https://www.ncbi.nlm.nih.gov/gene/43740570)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=26245&to=26472) 26245..26472

/gene="E"

/locus\_tag="GU280\_gp04"

/note="ORF4; structural protein; E protein"

/codon\_start=1

/product="envelope protein"

/protein\_id="[YP\_009724392.1](https://www.ncbi.nlm.nih.gov/protein/1796318600)"

/db\_xref="GeneID:[43740570](https://www.ncbi.nlm.nih.gov/gene/43740570)"

/translation="MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCC

NIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=26523&to=27191) 26523..27191

/gene="M"

/locus\_tag="GU280\_gp05"

/db\_xref="GeneID:[43740571](https://www.ncbi.nlm.nih.gov/gene/43740571)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=26523&to=27191) 26523..27191

/gene="M"

/locus\_tag="GU280\_gp05"

/note="ORF5; structural protein"

/codon\_start=1

/product="membrane glycoprotein"

/protein\_id="[YP\_009724393.1](https://www.ncbi.nlm.nih.gov/protein/1796318601)"

/db\_xref="GeneID:[43740571](https://www.ncbi.nlm.nih.gov/gene/43740571)"

/translation="MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNR

FLYIIKLIFLWLLWPVTLACFVLAAVYRINWITGGIAIAMACLVGLMWLSYFIASFRL

FARTRSMWSFNPETNILLNVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCD

IKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIA

LLVQ"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=27202&to=27387) 27202..27387

/gene="ORF6"

/locus\_tag="GU280\_gp06"

/db\_xref="GeneID:[43740572](https://www.ncbi.nlm.nih.gov/gene/43740572)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=27202&to=27387) 27202..27387

/gene="ORF6"

/locus\_tag="GU280\_gp06"

/codon\_start=1

/product="ORF6 protein"

/protein\_id="[YP\_009724394.1](https://www.ncbi.nlm.nih.gov/protein/1796318602)"

/db\_xref="GeneID:[43740572](https://www.ncbi.nlm.nih.gov/gene/43740572)"

/translation="MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSL

TENKYSQLDEEQPMEID"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=27394&to=27759) 27394..27759

/gene="ORF7a"

/locus\_tag="GU280\_gp07"

/db\_xref="GeneID:[43740573](https://www.ncbi.nlm.nih.gov/gene/43740573)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=27394&to=27759) 27394..27759

/gene="ORF7a"

/locus\_tag="GU280\_gp07"

/codon\_start=1

/product="ORF7a protein"

/protein\_id="[YP\_009724395.1](https://www.ncbi.nlm.nih.gov/protein/1796318603)"

/db\_xref="GeneID:[43740573](https://www.ncbi.nlm.nih.gov/gene/43740573)"

/translation="MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNS

PFHPLADNKFALTCFSTQFAFACPDGVKHVYQLRARSVSPKLFIRQEEVQELYSPIFL

IVAAIVFITLCFTLKRKTE"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=27756&to=27887) 27756..27887

/gene="ORF7b"

/locus\_tag="GU280\_gp08"

/db\_xref="GeneID:[43740574](https://www.ncbi.nlm.nih.gov/gene/43740574)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=27756&to=27887) 27756..27887

/gene="ORF7b"

/locus\_tag="GU280\_gp08"

/codon\_start=1

/product="ORF7b"

/protein\_id="[YP\_009725318.1](https://www.ncbi.nlm.nih.gov/protein/1820616061)"

/db\_xref="GeneID:[43740574](https://www.ncbi.nlm.nih.gov/gene/43740574)"

/translation="MIELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNETCHA"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=27894&to=28259) 27894..28259

/gene="ORF8"

/locus\_tag="GU280\_gp09"

/db\_xref="GeneID:[43740577](https://www.ncbi.nlm.nih.gov/gene/43740577)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=27894&to=28259) 27894..28259

/gene="ORF8"

/locus\_tag="GU280\_gp09"

/codon\_start=1

/product="ORF8 protein"

/protein\_id="[YP\_009724396.1](https://www.ncbi.nlm.nih.gov/protein/1796318604)"

/db\_xref="GeneID:[43740577](https://www.ncbi.nlm.nih.gov/gene/43740577)"

/translation="MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSK

WYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGSLVVRC

SFYEDFLEYHDVRVVLDFI"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=28274&to=29533) 28274..29533

/gene="N"

/locus\_tag="GU280\_gp10"

/db\_xref="GeneID:[43740575](https://www.ncbi.nlm.nih.gov/gene/43740575)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=28274&to=29533) 28274..29533

/gene="N"

/locus\_tag="GU280\_gp10"

/note="ORF9; structural protein"

/codon\_start=1

/product="nucleocapsid phosphoprotein"

/protein\_id="[YP\_009724397.2](https://www.ncbi.nlm.nih.gov/protein/1798174255)"

/db\_xref="GeneID:[43740575](https://www.ncbi.nlm.nih.gov/gene/43740575)"

/translation="MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQG

LPNNTASWFTALTQHGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMK

DLSPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQ

LPQGTTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAA

LALLLLDRLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGR

RGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYT

GAIKLDDKDPNFKDQVILLNKHIDAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTV

TLLPAADLDDFSKQLQQSMSSADSTQA"

[gene](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=29558&to=29674) 29558..29674

/gene="ORF10"

/locus\_tag="GU280\_gp11"

/db\_xref="GeneID:[43740576](https://www.ncbi.nlm.nih.gov/gene/43740576)"

[CDS](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=29558&to=29674) 29558..29674

/gene="ORF10"

/locus\_tag="GU280\_gp11"

/codon\_start=1

/product="ORF10 protein"

/protein\_id="[YP\_009725255.1](https://www.ncbi.nlm.nih.gov/protein/1798174256)"

/db\_xref="GeneID:[43740576](https://www.ncbi.nlm.nih.gov/gene/43740576)"

/translation="MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNFNLT"

[stem\_loop](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=29609&to=29644) 29609..29644

/gene="ORF10"

/locus\_tag="GU280\_gp11"

/inference="COORDINATES:

profile::Rfam-release-14.1:RF00165,Infernal:1.1.2"

/function="Coronavirus 3' UTR pseudoknot stem-loop 1"

[stem\_loop](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=29629&to=29657) 29629..29657

/gene="ORF10"

/locus\_tag="GU280\_gp11"

/inference="COORDINATES:

profile::Rfam-release-14.1:RF00165,Infernal:1.1.2"

/function="Coronavirus 3' UTR pseudoknot stem-loop 2"

[3'UTR](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=29675&to=29903) 29675..29903

[stem\_loop](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?from=29728&to=29768) 29728..29768

/inference="COORDINATES:

profile:Rfam-release-14.1:RF00164,Infernal:1.1.2"

/note="basepair exception: alignment to the Rfam model

implies coordinates 29740:29758 form a noncanonical C:T

basepair, but the homologous positions form a highly

conserved C:G basepair in other viruses, including SARS

(NC\_004718.3)"

/function="Coronavirus 3' stem-loop II-like motif (s2m)"

ORIGIN

1 attaaaggtt tataccttcc caggtaacaa accaaccaac tttcgatctc ttgtagatct

61 gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcatgc ttagtgcact

121 cacgcagtat aattaataac taattactgt cgttgacagg acacgagtaa ctcgtctatc

181 ttctgcaggc tgcttacggt ttcgtccgtg ttgcagccga tcatcagcac atctaggttt

241 cgtccgggtg tgaccgaaag gtaagatgga gagccttgtc cctggtttca acgagaaaac

301 acacgtccaa ctcagtttgc ctgttttaca ggttcgcgac gtgctcgtac gtggctttgg

361 agactccgtg gaggaggtct tatcagaggc acgtcaacat cttaaagatg gcacttgtgg

421 cttagtagaa gttgaaaaag gcgttttgcc tcaacttgaa cagccctatg tgttcatcaa

481 acgttcggat gctcgaactg cacctcatgg tcatgttatg gttgagctgg tagcagaact

541 cgaaggcatt cagtacggtc gtagtggtga gacacttggt gtccttgtcc ctcatgtggg

601 cgaaatacca gtggcttacc gcaaggttct tcttcgtaag aacggtaata aaggagctgg

661 tggccatagt tacggcgccg atctaaagtc atttgactta ggcgacgagc ttggcactga

721 tccttatgaa gattttcaag aaaactggaa cactaaacat agcagtggtg ttacccgtga

781 actcatgcgt gagcttaacg gaggggcata cactcgctat gtcgataaca acttctgtgg

841 ccctgatggc taccctcttg agtgcattaa agaccttcta gcacgtgctg gtaaagcttc

901 atgcactttg tccgaacaac tggactttat tgacactaag aggggtgtat actgctgccg

961 tgaacatgag catgaaattg cttggtacac ggaacgttct gaaaagagct atgaattgca

1021 gacacctttt gaaattaaat tggca**aagaa** atttgacacc ttcaatgggg aatgtccaaa

1081 ttttgtattt cccttaaatt ccataatcaa gactattcaa ccaagggttg aa**aagaa**aaa

1141 gcttgatggc tttatgggta gaattcgatc tgtctatcca gttgcgtcac caaatgaatg

1201 caaccaaatg tgcctttcaa ctctcatgaa gtgtgatcat tgtggtgaaa cttcatggca

1261 gacgggcgat tttgttaaag ccacttgcga attttgtggc actgagaatt tgactaaaga

1321 aggtgccact acttgtggtt acttacccca aaatgctgtt gttaaaattt attgtccagc

1381 atgtcacaat tcagaagtag gacctgagca tagtcttgcc gaataccata atgaatctgg

1441 cttgaaaacc attcttcgta agggtggtcg cactattgcc tttggaggct gtgtgttctc

1501 ttatgttggt tgccataaca agtgtgccta ttgggttcca cgtgctagcg ctaacatagg

1561 ttgtaaccat acaggtgttg ttggagaagg ttccgaaggt cttaatgaca accttcttga

1621 aatactccaa aaagagaaag tcaacatcaa tattgttggt gactttaaac ttaatgaaga

1681 gatcgccatt attttggcat ctttttctgc ttccacaagt gcttttgtgg aaactgtgaa

1741 aggtttggat tataaagcat tcaaacaaat tgttgaatcc tgtggtaatt ttaaagttac

1801 aaaaggaaaa gctaaaaaag gtgcctggaa tattggtgaa cagaaatcaa tactgagtcc

1861 tctttatgca tttgcatcag aggctgctcg tgttgtacga tcaattttct cccgcactct

1921 tgaaactgct caaaattctg tgcgtgtttt acagaaggcc gctataacaa tactagatgg

1981 aatttcacag tattcactga gactcattga tgctatgatg ttcacatctg atttggctac

2041 taacaatcta gttgtaatgg cctacattac aggtggtgtt gttcagttga cttcgcagtg

2101 gctaactaac atctttggca ctgtttatga aaaactcaaa cccgtccttg attggcttga

2161 agagaagttt aaggaaggtg tagagtttct tagagacggt tgggaaattg ttaaatttat

2221 ctcaacctgt gcttgtgaaa ttgtcggtgg acaaattgtc acctgtgcaa aggaaattaa

2281 ggagagtgtt cagacattct ttaagcttgt aaataaattt ttggctttgt gtgctgactc

2341 tatcattatt ggtggagcta aacttaaagc cttgaattta ggtgaaacat ttgtcacgca

2401 ctcaaaggga ttgtacagaa agtgtgttaa atccagagaa gaaactggcc tactcatgcc

2461 tctaaaagcc ccaa**aagaa**a ttatcttctt agagggagaa acacttccca cagaagtgtt

2521 aacagaggaa gttgtcttga aaactggtga tttacaacca ttagaacaac ctactagtga

2581 agctgttgaa gctccattgg ttggtacacc agtttgtatt aacgggctta tgttgctcga

2641 aatcaaagac acagaaaagt actgtgccct tgcacctaat atgatggtaa caaacaatac

2701 cttcacactc aaaggcggtg caccaacaaa ggttactttt ggtgatgaca ctgtgataga

2761 agtgcaaggt tacaagagtg tgaatatcac ttttgaactt gatgaaagga ttgataaagt

2821 acttaatgag aagtgctctg cctatacagt tgaactcggt acagaagtaa atgagttcgc

2881 ctgtgttgtg gcagatgctg tcataaaaac tttgcaacca gtatctgaat tacttacacc

2941 actgggcatt gatttagatg agtggagtat ggctacatac tacttatttg atgagtctgg

3001 tgagtttaaa ttggcttcac atatgtattg ttctttctac cctccagatg aggatgaaga

3061 agaaggtgat tgtg**aagaa**g aagagtttga gccatcaact caatatgagt atggtactga

3121 agatgattac caaggtaaac ctttggaatt tggtgccact tctgctgctc ttcaacctga

3181 agaagagcaa g**aagaa**gatt ggttagatga tgatagtcaa caaactgttg gtcaacaaga

3241 cggcagtgag gacaatcaga caactactat tcaaacaatt gttgaggttc aacctcaatt

3301 agagatggaa cttacaccag ttgttcagac tattgaagtg aatagtttta gtggttattt

3361 aaaacttact gacaatgtat acattaaaaa tgcagacatt gtgg**aagaa**g ctaaaaaggt

3421 aaaaccaaca gtggttgtta atgcagccaa tgtttacctt aaacatggag gaggtgttgc

3481 aggagcctta aataaggcta ctaacaatgc catgcaagtt gaatctgatg attacatagc

3541 tactaatgga ccacttaaag tgggtggtag ttgtgtttta agcggacaca atcttgctaa

3601 acactgtctt catgttgtcg gcccaaatgt taacaaaggt gaagacattc aacttcttaa

3661 gagtgcttat gaaaatttta atcagcacga agttctactt gcaccattat tatcagctgg

3721 tatttttggt gctgacccta tacattcttt aagagtttgt gtagatactg ttcgcacaaa

3781 tgtctactta gctgtctttg ataaaaatct ctatgacaaa cttgtttcaa gctttttgga

3841 aatgaagagt gaaaagcaag ttgaacaaaa gatcgctgag attcctaaag aggaagttaa

3901 gccatttata actgaaagta aaccttcagt tgaacagaga aaacaagatg at**aagaa**aat

3961 caaagcttgt gttg**aagaa**g ttacaacaac tctgg**aagaa** actaagttcc tcacagaaaa

4021 cttgttactt tatattgaca ttaatggcaa tcttcatcca gattctgcca ctcttgttag

4081 tgacattgac atcactttct ta**aagaa**aga tgctccatat atagtgggtg atgttgttca

4141 agagggtgtt ttaactgctg tggttatacc tactaaaaag gctggtggca ctactgaaat

4201 gctagcgaaa gctttgagaa aagtgccaac agacaattat ataaccactt acccgggtca

4261 gggtttaaat ggttacactg tagaggaggc aaagacagtg cttaaaaagt gtaaaagtgc

4321 cttttacatt ctaccatcta ttatctctaa tgagaagcaa gaaattcttg gaactgtttc

4381 ttggaatttg cgagaaatgc ttgcacatgc ag**aagaa**aca cgcaaattaa tgcctgtctg

4441 tgtggaaact aaagccatag tttcaactat acagcgtaaa tataagggta ttaaaataca

4501 agagggtgtg gttgattatg gtgctagatt ttacttttac accagtaaaa caactgtagc

4561 gtcacttatc aacacactta acgatctaaa tgaaactctt gttacaatgc cacttggcta

4621 tgtaacacat ggcttaaatt tgg**aagaa**gc tgctcggtat atgagatctc tcaaagtgcc

4681 agctacagtt tctgtttctt cacctgatgc tgttacagcg tataatggtt atcttacttc

4741 ttcttctaaa acacctgaag aacattttat tgaaaccatc tcacttgctg gttcctataa

4801 agattggtcc tattctggac aatctacaca actaggtata gaatttctta agagaggtga

4861 taaaagtgta tattacacta gtaatcctac cacattccac ctagatggtg aagttatcac

4921 ctttgacaat cttaagacac ttctttcttt gagagaagtg aggactatta aggtgtttac

4981 aacagtagac aacattaacc tccacacgca agttgtggac atgtcaatga catatggaca

5041 acagtttggt ccaacttatt tggatggagc tgatgttact aaaataaaac ctcataattc

5101 acatgaaggt aaaacatttt atgttttacc taatgatgac actctacgtg ttgaggcttt

5161 tgagtactac cacacaactg atcctagttt tctgggtagg tacatgtcag cattaaatca

5221 cactaaaaag tggaaatacc cacaagttaa tggtttaact tctattaaat gggcagataa

5281 caactgttat cttgccactg cattgttaac actccaacaa atagagttga agtttaatcc

5341 acctgctcta caagatgctt attacagagc aagggctggt gaagctgcta acttttgtgc

5401 acttatctta gcctactgta ataagacagt aggtgagtta ggtgatgtta gagaaacaat

5461 gagttacttg tttcaacatg ccaatttaga ttcttgcaaa agagtcttga acgtggtgtg

5521 taaaacttgt ggacaacagc agacaaccct taagggtgta gaagctgtta tgtacatggg

5581 cacactttct tatgaacaat tt**aagaa**agg tgttcagata ccttgtacgt gtggtaaaca

5641 agctacaaaa tatctagtac aacaggagtc accttttgtt atgatgtcag caccacctgc

5701 tcagtatgaa cttaagcatg gtacatttac ttgtgctagt gagtacactg gtaattacca

5761 gtgtggtcac tataaacata taacttctaa agaaactttg tattgcatag acggtgcttt

5821 acttacaaag tcctcagaat acaaaggtcc tattacggat gttttctaca **aagaa**aacag

5881 ttacacaaca accataaaac cagttactta taaattggat ggtgttgttt gtacagaaat

5941 tgaccctaag ttggacaatt attat**aagaa** agacaattct tatttcacag agcaaccaat

6001 tgatcttgta ccaaaccaac catatccaaa cgcaagcttc gataatttta agtttgtatg

6061 tgataatatc aaatttgctg atgatttaaa ccagttaact ggttataaga aacctgcttc

6121 aagagagctt aaagttacat ttttccctga cttaaatggt gatgtggtgg ctattgatta

6181 taaacactac acaccctctt tt**aagaa**agg agctaaattg ttacataaac ctattgtttg

6241 gcatgttaac aatgcaacta ataaagccac gtataaacca aatacctggt gtatacgttg

6301 tctttggagc acaaaaccag ttgaaacatc aaattcgttt gatgtactga agtcagagga

6361 cgcgcaggga atggataatc ttgcctgcga agatctaaaa ccagtctctg **aagaa**gtagt

6421 ggaaaatcct accatacaga aagacgttct tgagtgtaat gtgaaaacta ccgaagttgt

6481 aggagacatt atacttaaac cagcaaataa tagtttaaaa attacagaag aggttggcca

6541 cacagatcta atggctgctt atgtagacaa ttctagtctt actattaaga aacctaatga

6601 attatctaga gtattaggtt tgaaaaccct tgctactcat ggtttagctg ctgttaatag

6661 tgtcccttgg gatactatag ctaattatgc taagcctttt cttaacaaag ttgttagtac

6721 aactactaac atagttacac ggtgtttaaa ccgtgtttgt actaattata tgccttattt

6781 ctttacttta ttgctacaat tgtgtacttt tactagaagt acaaattcta gaattaaagc

6841 atctatgccg actactatag ca**aagaa**tac tgttaagagt gtcggtaaat tttgtctaga

6901 ggcttcattt aattatttga agtcacctaa tttttctaaa ctgataaata ttataatttg

6961 gtttttacta ttaagtgttt gcctaggttc tttaatctac tcaaccgctg ctttaggtgt

7021 tttaatgtct aatttaggca tgccttctta ctgtactggt tacagagaag gctatttgaa

7081 ctctactaat gtcactattg caacctactg tactggttct ataccttgta gtgtttgtct

7141 tagtggttta gattctttag acacctatcc ttctttagaa actatacaaa ttaccatttc

7201 atcttttaaa tgggatttaa ctgcttttgg cttagttgca gagtggtttt tggcatatat

7261 tcttttcact aggtttttct atgtacttgg attggctgca atcatgcaat tgtttttcag

7321 ctattttgca gtacatttta ttagtaattc ttggcttatg tggttaataa ttaatcttgt

7381 acaaatggcc ccgatttcag ctatggttag aatgtacatc ttctttgcat cattttatta

7441 tgtatggaaa agttatgtgc atgttgtaga cggttgtaat tcatcaactt gtatgatgtg

7501 ttacaaacgt aatagagcaa caagagtcga atgtacaact attgttaatg gtgttagaag

7561 gtccttttat gtctatgcta atggaggtaa aggcttttgc aaactacaca attggaattg

7621 tgttaattgt gatacattct gtgctggtag tacatttatt agtgatgaag ttgcgagaga

7681 cttgtcacta cagtttaaaa gaccaataaa tcctactgac cagtcttctt acatcgttga

7741 tagtgttaca gtg**aagaa**tg gttccatcca tctttacttt gataaagctg gtcaaaagac

7801 ttatgaaaga cattctctct ctcattttgt taacttagac aacctgagag ctaataacac

7861 taaaggttca ttgcctatta atgttatagt ttttgatggt aaatcaaaat gtg**aagaa**tc

7921 atctgcaaaa tcagcgtctg tttactacag tcagcttatg tgtcaaccta tactgttact

7981 agatcaggca ttagtgtctg atgttggtga tagtgcggaa gttgcagtta aaatgtttga

8041 tgcttacgtt aatacgtttt catcaacttt taacgtacca atggaaaaac tcaaaacact

8101 agttgcaact gcagaagctg aacttgcaaa gaatgtgtcc ttagacaatg tcttatctac

8161 ttttatttca gcagctcggc aagggtttgt tgattcagat gtagaaacta aagatgttgt

8221 tgaatgtctt aaattgtcac atcaatctga catagaagtt actggcgata gttgtaataa

8281 ctatatgctc acctataaca aagttgaaaa catgacaccc cgtgaccttg gtgcttgtat

8341 tgactgtagt gcgcgtcata ttaatgcgca ggtagcaaaa agtcacaaca ttgctttgat

8401 atggaacgtt aaagatttca tgtcattgtc tgaacaacta cgaaaacaaa tacgtagtgc

8461 tgctaaaaag aataacttac cttttaagtt gacatgtgca actactagac aagttgttaa

8521 tgttgtaaca acaaagatag cacttaaggg tggtaaaatt gttaataatt ggttgaagca

8581 gttaattaaa gttacacttg tgttcctttt tgttgctgct attttctatt taataacacc

8641 tgttcatgtc atgtctaaac atactgactt ttcaagtgaa atcataggat acaaggctat

8701 tgatggtggt gtcactcgtg acatagcatc tacagatact tgttttgcta acaaacatgc

8761 tgattttgac acatggttta gccagcgtgg tggtagttat actaatgaca aagcttgccc

8821 attgattgct gcagtcataa caagagaagt gggttttgtc gtgcctggtt tgcctggcac

8881 gatattacgc acaactaatg gtgacttttt gcatttctta cctagagttt ttagtgcagt

8941 tggtaacatc tgttacacac catcaaaact tatagagtac actgactttg caacatcagc

9001 ttgtgttttg gctgctgaat gtacaatttt taaagatgct tctggtaagc cagtaccata

9061 ttgttatgat accaatgtac tagaaggttc tgttgcttat gaaagtttac gccctgacac

9121 acgttatgtg ctcatggatg gctctattat tcaatttcct aacacctacc ttgaaggttc

9181 tgttagagtg gtaacaactt ttgattctga gtactgtagg cacggcactt gtgaaagatc

9241 agaagctggt gtttgtgtat ctactagtgg tagatgggta cttaacaatg attattacag

9301 atctttacca ggagttttct gtggtgtaga tgctgtaaat ttacttacta atatgtttac

9361 accactaatt caacctattg gtgctttgga catatcagca tctatagtag ctggtggtat

9421 tgtagctatc gtagtaacat gccttgccta ctattttatg aggtttagaa gagcttttgg

9481 tgaatacagt catgtagttg cctttaatac tttactattc cttatgtcat tcactgtact

9541 ctgtttaaca ccagtttact cattcttacc tggtgtttat tctgttattt acttgtactt

9601 gacattttat cttactaatg atgtttcttt tttagcacat attcagtgga tggttatgtt

9661 cacaccttta gtacctttct ggataacaat tgcttatatc atttgtattt ccacaaagca

9721 tttctattgg ttctttagta attacctaaa gagacgtgta gtctttaatg gtgtttcctt

9781 tagtactttt g**aagaa**gctg cgctgtgcac ctttttgtta aata**aagaa**a tgtatctaaa

9841 gttgcgtagt gatgtgctat tacctcttac gcaatataat agatacttag ctctttataa

9901 taagtacaag tattttagtg gagcaatgga tacaactagc tacagagaag ctgcttgttg

9961 tcatctcgca aaggctctca atgacttcag taactcaggt tctgatgttc tttaccaacc

10021 accacaaacc tctatcacct cagctgtttt gcagagtggt tttagaaaaa tggcattccc

10081 atctggtaaa gttgagggtt gtatggtaca agtaacttgt ggtacaacta cacttaacgg

10141 tctttggctt gatgacgtag tttactgtcc aagacatgtg atctgcacct ctgaagacat

10201 gcttaaccct aattatgaag atttactcat tcgtaagtct aatcataatt tcttggtaca

10261 ggctggtaat gttcaactca gggttattgg acattctatg caaaattgtg tacttaagct

10321 taaggttgat acagccaatc ctaagacacc taagtataag tttgttcgca ttcaaccagg

10381 acagactttt tcagtgttag cttgttacaa tggttcacca tctggtgttt accaatgtgc

10441 tatgaggccc aatttcacta ttaagggttc attccttaat ggttcatgtg gtagtgttgg

10501 ttttaacata gattatgact gtgtctcttt ttgttacatg caccatatgg aattaccaac

10561 tggagttcat gctggcacag acttagaagg taacttttat ggaccttttg ttgacaggca

10621 aacagcacaa gcagctggta cggacacaac tattacagtt aatgttttag cttggttgta

10681 cgctgctgtt ataaatggag acaggtggtt tctcaatcga tttaccacaa ctcttaatga

10741 ctttaacctt gtggctatga agtacaatta tgaacctcta acacaagacc atgttgacat

10801 actaggacct ctttctgctc aaactggaat tgccgtttta gatatgtgtg cttcattaaa

10861 agaattactg caaaatggta tgaatggacg taccatattg ggtagtgctt tattagaaga

10921 tgaatttaca ccttttgatg ttgttagaca atgctcaggt gttactttcc aaagtgcagt

10981 gaa**aagaa**ca atcaagggta cacaccactg gttgttactc acaattttga cttcactttt

11041 agttttagtc cagagtactc aatggtcttt gttctttttt ttgtatgaaa atgccttttt

11101 accttttgct atgggtatta ttgctatgtc tgcttttgca atgatgtttg tcaaacataa

11161 gcatgcattt ctctgtttgt ttttgttacc ttctcttgcc actgtagctt attttaatat

11221 ggtctatatg cctgctagtt gggtgatgcg tattatgaca tggttggata tggttgatac

11281 tagtttgtct ggttttaagc taaaagactg tgttatgtat gcatcagctg tagtgttact

11341 aatccttatg acagc**aagaa** ctgtgtatga tgatggtgct aggagagtgt ggacacttat

11401 gaatgtcttg acactcgttt ataaagttta ttatggtaat gctttagatc aagccatttc

11461 catgtgggct cttataatct ctgttacttc taactactca ggtgtagtta caactgtcat

11521 gtttttggcc agaggtattg tttttatgtg tgttgagtat tgccctattt tcttcataac

11581 tggtaataca cttcagtgta taatgctagt ttattgtttc ttaggctatt tttgtacttg

11641 ttactttggc ctcttttgtt tactcaaccg ctactttaga ctgactcttg gtgtttatga

11701 ttacttagtt tctacacagg agtttagata tatgaattca cagggactac tcccacccaa

11761 gaatagcata gatgccttca aactcaacat taaattgttg ggtgttggtg gcaaaccttg

11821 tatcaaagta gccactgtac agtctaaaat gtcagatgta aagtgcacat cagtagtctt

11881 actctcagtt ttgcaacaac tcagagtaga atcatcatct aaattgtggg ctcaatgtgt

11941 ccagttacac aatgacattc tcttagctaa agatactact gaagcctttg aaaaaatggt

12001 ttcactactt tctgttttgc tttccatgca gggtgctgta gacataaaca agctttgtga

12061 agaaatgctg gacaacaggg caaccttaca agctatagcc tcagagttta gttcccttcc

12121 atcatatgca gcttttgcta ctgctcaaga agcttatgag caggctgttg ctaatggtga

12181 ttctgaagtt gttcttaaaa agttg**aagaa** gtctttgaat gtggctaaat ctgaatttga

12241 ccgtgatgca gccatgcaac gtaagttgga aaagatggct gatcaagcta tgacccaaat

12301 gtataaacag gctagatctg aggacaagag ggcaaaagtt actagtgcta tgcagacaat

12361 gcttttcact atgcttagaa agttggataa tgatgcactc aacaacatta tcaacaatgc

12421 aagagatggt tgtgttccct tgaacataat acctcttaca acagcagcca aactaatggt

12481 tgtcatacca gactataaca catataaaaa tacgtgtgat ggtacaacat ttacttatgc

12541 atcagcattg tgggaaatcc aacaggttgt agatgcagat agtaaaattg ttcaacttag

12601 tgaaattagt atggacaatt cacctaattt agcatggcct cttattgtaa cagctttaag

12661 ggccaattct gctgtcaaat tacagaataa tgagcttagt cctgttgcac tacgacagat

12721 gtcttgtgct gccggtacta cacaaactgc ttgcactgat gacaatgcgt tagcttacta

12781 caacacaaca aagggaggta ggtttgtact tgcactgtta tccgatttac aggatttgaa

12841 atgggctaga ttccctaaga gtgatggaac tggtactatc tatacagaac tggaaccacc

12901 ttgtaggttt gttacagaca cacctaaagg tcctaaagtg aagtatttat actttattaa

12961 aggattaaac aacctaaata gaggtatggt acttggtagt ttagctgcca cagtacgtct

13021 acaagctggt aatgcaacag aagtgcctgc caattcaact gtattatctt tctgtgcttt

13081 tgctgtagat gctgctaaag cttacaaaga ttatctagct agtgggggac aaccaatcac

13141 taattgtgtt aagatgttgt gtacacacac tggtactggt caggcaataa cagttacacc

13201 ggaagccaat atggatcaag aatcctttgg tggtgcatcg tgttgtctgt actgccgttg

13261 ccacatagat catccaaatc ctaaaggatt ttgtgactta aaaggtaagt atgtacaaat

13321 acctacaact tgtgctaatg accctgtggg ttttacactt aaaaacacag tctgtaccgt

13381 ctgcggtatg tggaaaggtt atggctgtag ttgtgatcaa ctccgcgaac ccatgcttca

13441 gtcagctgat gcacaatcgt ttttaaacgg gtttgcggtg taagtgcagc ccgtcttaca

13501 ccgtgcggca caggcactag tactgatgtc gtatacaggg cttttgacat ctacaatgat

13561 aaagtagctg gttttgctaa attcctaaaa actaattgtt gtcgcttcca agaaaaggac

13621 gaagatgaca atttaattga ttcttacttt gtagttaaga gacacacttt ctctaactac

13681 caacatgaag aaacaattta taatttactt aaggattgtc cagctgttgc taaacatgac

13741 ttctttaagt ttagaataga cggtgacatg gtaccacata tatcacgtca acgtcttact

13801 aaatacacaa tggcagacct cgtctatgct ttaaggcatt ttgatgaagg taattgtgac

13861 acattaaaag aaatacttgt cacatacaat tgttgtgatg atgattattt caataaaaag

13921 gactggtatg attttgtaga aaacccagat atattacgcg tatacgccaa cttaggtgaa

13981 cgtgtacgcc aagctttgtt aaaaacagta caattctgtg atgccatgcg aaatgctggt

14041 attgttggtg tactgacatt agataatcaa gatctcaatg gtaactggta tgatttcggt

14101 gatttcatac aaaccacgcc aggtagtgga gttcctgttg tagattctta ttattcattg

14161 ttaatgccta tattaacctt gaccagggct ttaactgcag agtcacatgt tgacactgac

14221 ttaacaaagc cttacattaa gtgggatttg ttaaaatatg acttcacgga agagaggtta

14281 aaactctttg accgttattt taaatattgg gatcagacat accacccaaa ttgtgttaac

14341 tgtttggatg acagatgcat tctgcattgt gcaaacttta atgttttatt ctctacagtg

14401 ttcccaccta caagttttgg accactagtg agaaaaatat ttgttgatgg tgttccattt

14461 gtagtttcaa ctggatacca cttcagagag ctaggtgttg tacataatca ggatgtaaac

14521 ttacatagct ctagacttag ttttaaggaa ttacttgtgt atgctgctga ccctgctatg

14581 cacgctgctt ctggtaatct attactagat aaacgcacta cgtgcttttc agtagctgca

14641 cttactaaca atgttgcttt tcaaactgtc aaacccggta attttaacaa agacttctat

14701 gactttgctg tgtctaaggg tttctttaag gaaggaagtt ctgttgaatt aaaacacttc

14761 ttctttgctc aggatggtaa tgctgctatc agcgattatg actactatcg ttataatcta

14821 ccaacaatgt gtgatatcag acaactacta tttgtagttg aagttgttga taagtacttt

14881 gattgttacg atggtggctg tattaatgct aaccaagtca tcgtcaacaa cctagacaaa

14941 tcagctggtt ttccatttaa taaatggggt aaggctagac tttattatga ttcaatgagt

15001 tatgaggatc aagatgcact tttcgcatat acaaaacgta atgtcatccc tactataact

15061 caaatgaatc ttaagtatgc cattagtgca **aagaa**tagag ctcgcaccgt agctggtgtc

15121 tctatctgta gtactatgac caatagacag tttcatcaaa aattattgaa atcaatagcc

15181 gccactagag gagctactgt agtaattgga acaagcaaat tctatggtgg ttggcacaac

15241 atgttaaaaa ctgtttatag tgatgtagaa aaccctcacc ttatgggttg ggattatcct

15301 aaatgtgata gagccatgcc taacatgctt agaattatgg cctcacttgt tcttgctcgc

15361 aaacatacaa cgtgttgtag cttgtcacac cgtttctata gattagctaa tgagtgtgct

15421 caagtattga gtgaaatggt catgtgtggc ggttcactat atgttaaacc aggtggaacc

15481 tcatcaggag atgccacaac tgcttatgct aatagtgttt ttaacatttg tcaagctgtc

15541 acggccaatg ttaatgcact tttatctact gatggtaaca aaattgccga taagtatgtc

15601 cgcaatttac aacacagact ttatgagtgt ctctatagaa atagagatgt tgacacagac

15661 tttgtgaatg agttttacgc atatttgcgt aaacatttct caatgatgat actctctgac

15721 gatgctgttg tgtgtttcaa tagcacttat gcatctcaag gtctagtggc tagcataaag

15781 aactttaagt cagttcttta ttatcaaaac aatgttttta tgtctgaagc aaaatgttgg

15841 actgagactg accttactaa aggacctcat gaattttgct ctcaacatac aatgctagtt

15901 aaacagggtg atgattatgt gtaccttcct tacccagatc catc**aagaa**t cctaggggcc

15961 ggctgttttg tagatgatat cgtaaaaaca gatggtacac ttatgattga acggttcgtg

16021 tctttagcta tagatgctta cccacttact aaacatccta atcaggagta tgctgatgtc

16081 tttcatttgt acttacaata cat**aagaa**ag ctacatgatg agttaacagg acacatgtta

16141 gacatgtatt ctgttatgct tactaatgat aacacttcaa ggtattggga acctgagttt

16201 tatgaggcta tgtacacacc gcatacagtc ttacaggctg ttggggcttg tgttctttgc

16261 aattcacaga cttcattaag atgtggtgct tgcatacgta gaccattctt atgttgtaaa

16321 tgctgttacg accatgtcat atcaacatca cataaattag tcttgtctgt taatccgtat

16381 gtttgcaatg ctccaggttg tgatgtcaca gatgtgactc aactttactt aggaggtatg

16441 agctattatt gtaaatcaca taaaccaccc attagttttc cattgtgtgc taatggacaa

16501 gtttttggtt tatataaaaa tacatgtgtt ggtagcgata atgttactga ctttaatgca

16561 attgcaacat gtgactggac aaatgctggt gattacattt tagctaacac ctgtactgaa

16621 agactcaagc tttttgcagc agaaacgctc aaagctactg aggagacatt taaactgtct

16681 tatggtattg ctactgtacg tgaagtgctg tctgacagag aattacatct ttcatgggaa

16741 gttggtaaac ctagaccacc acttaaccga aattatgtct ttactggtta tcgtgtaact

16801 aaaaacagta aagtacaaat aggagagtac acctttgaaa aaggtgacta tggtgatgct

16861 gttgtttacc gaggtacaac aacttacaaa ttaaatgttg gtgattattt tgtgctgaca

16921 tcacatacag taatgccatt aagtgcacct acactagtgc cacaagagca ctatgttaga

16981 attactggct tatacccaac actcaatatc tcagatgagt tttctagcaa tgttgcaaat

17041 tatcaaaagg ttggtatgca aaagtattct acactccagg gaccacctgg tactggtaag

17101 agtcattttg ctattggcct agctctctac tacccttctg ctcgcatagt gtatacagct

17161 tgctctcatg ccgctgttga tgcactatgt gagaaggcat taaaatattt gcctatagat

17221 aaatgtagta gaattatacc tgcacgtgct cgtgtagagt gttttgataa attcaaagtg

17281 aattcaacat tagaacagta tgtcttttgt actgtaaatg cattgcctga gacgacagca

17341 gatatagttg tctttgatga aatttcaatg gccacaaatt atgatttgag tgttgtcaat

17401 gccagattac gtgctaagca ctatgtgtac attggcgacc ctgctcaatt acctgcacca

17461 cgcacattgc taactaaggg cacactagaa ccagaatatt tcaattcagt gtgtagactt

17521 atgaaaacta taggtccaga catgttcctc ggaacttgtc ggcgttgtcc tgctgaaatt

17581 gttgacactg tgagtgcttt ggtttatgat aataagctta aagcacataa agacaaatca

17641 gctcaatgct ttaaaatgtt ttataagggt gttatcacgc atgatgtttc atctgcaatt

17701 aacaggccac aaataggcgt ggtaagagaa ttccttacac gtaaccctgc ttggagaaaa

17761 gctgtcttta tttcacctta taattcacag aatgctgtag cctcaaagat tttgggacta

17821 ccaactcaaa ctgttgattc atcacagggc tcagaatatg actatgtcat attcactcaa

17881 accactgaaa cagctcactc ttgtaatgta aacagattta atgttgctat taccagagca

17941 aaagtaggca tactttgcat aatgtctgat agagaccttt atgacaagtt gcaatttaca

18001 agtcttgaaa ttccacgtag gaatgtggca actttacaag ctgaaaatgt aacaggactc

18061 tttaaagatt gtagtaaggt aatcactggg ttacatccta cacaggcacc tacacacctc

18121 agtgttgaca ctaaattcaa aactgaaggt ttatgtgttg acatacctgg catacctaag

18181 gacatgacct atagaagact catctctatg atgggtttta aaatgaatta tcaagttaat

18241 ggttacccta acatgtttat cacccgcgaa gaagctataa gacatgtacg tgcatggatt

18301 ggcttcgatg tcgaggggtg tcatgctact agagaagctg ttggtaccaa tttaccttta

18361 cagctaggtt tttctacagg tgttaaccta gttgctgtac ctacaggtta tgttgataca

18421 cctaataata cagatttttc cagagttagt gctaaaccac cgcctggaga tcaatttaaa

18481 cacctcatac cacttatgta caaaggactt ccttggaatg tagtgcgtat aaagattgta

18541 caaatgttaa gtgacacact taaaaatctc tctgacagag tcgtatttgt cttatgggca

18601 catggctttg agttgacatc tatgaagtat tttgtgaaaa taggacctga gcgcacctgt

18661 tgtctatgtg atagacgtgc cacatgcttt tccactgctt cagacactta tgcctgttgg

18721 catcattcta ttggatttga ttacgtctat aatccgttta tgattgatgt tcaacaatgg

18781 ggttttacag gtaacctaca aagcaaccat gatctgtatt gtcaagtcca tggtaatgca

18841 catgtagcta gttgtgatgc aatcatgact aggtgtctag ctgtccacga gtgctttgtt

18901 aagcgtgttg actggactat tgaatatcct ataattggtg atgaactgaa gattaatgcg

18961 gcttgtagaa aggttcaaca catggttgtt aaagctgcat tattagcaga caaattccca

19021 gttcttcacg acattggtaa ccctaaagct attaagtgtg tacctcaagc tgatgtagaa

19081 tggaagttct atgatgcaca gccttgtagt gacaaagctt ataaaataga agaattattc

19141 tattcttatg ccacacattc tgacaaattc acagatggtg tatgcctatt ttggaattgc

19201 aatgtcgata gatatcctgc taattccatt gtttgtagat ttgacactag agtgctatct

19261 aaccttaact tgcctggttg tgatggtggc agtttgtatg taaataaaca tgcattccac

19321 acaccagctt ttgataaaag tgcttttgtt aatttaaaac aattaccatt tttctattac

19381 tctgacagtc catgtgagtc tcatggaaaa caagtagtgt cagatataga ttatgtacca

19441 ctaaagtctg ctacgtgtat aacacgttgc aatttaggtg gtgctgtctg tagacatcat

19501 gctaatgagt acagattgta tctcgatgct tataacatga tgatctcagc tggctttagc

19561 ttgtgggttt acaaacaatt tgatacttat aacctctgga acacttttac aagacttcag

19621 agtttagaaa atgtggcttt taatgttgta aataagggac actttgatgg acaacagggt

19681 gaagtaccag tttctatcat taataacact gtttacacaa aagttgatgg tgttgatgta

19741 gaattgtttg aaaataaaac aacattacct gttaatgtag catttgagct ttgggctaag

19801 cgcaacatta aaccagtacc agaggtgaaa atactcaata atttgggtgt ggacattgct

19861 gctaatactg tgatctggga ctacaaaaga gatgctccag cacatatatc tactattggt

19921 gtttgttcta tgactgacat agcc**aagaa**a ccaactgaaa cgatttgtgc accactcact

19981 gtcttttttg atggtagagt tgatggtcaa gtagacttat ttagaaatgc ccgtaatggt

20041 gttcttatta cagaaggtag tgttaaaggt ttacaaccat ctgtaggtcc caaacaagct

20101 agtcttaatg gagtcacatt aattggagaa gccgtaaaaa cacagttcaa ttattataag

20161 aaagttgatg gtgttgtcca acaattacct gaaacttact ttactcagag tagaaattta

20221 c**aagaa**ttta aacccaggag tcaaatggaa attgatttct tagaattagc tatggatgaa

20281 ttcattgaac ggtataaatt agaaggctat gccttcgaac atatcgttta tggagatttt

20341 agtcatagtc agttaggtgg tttacatcta ctgattggac tagctaaacg ttttaaggaa

20401 tcaccttttg aattagaaga ttttattcct atggacagta cagttaaaaa ctatttcata

20461 acagatgcgc aaacaggttc atctaagtgt gtgtgttctg ttattgattt attacttgat

20521 gattttgttg aaataataaa atcccaagat ttatctgtag tttctaaggt tgtcaaagtg

20581 actattgact atacagaaat ttcatttatg ctttggtgta aagatggcca tgtagaaaca

20641 ttttacccaa aattacaatc tagtcaagcg tggcaaccgg gtgttgctat gcctaatctt

20701 tacaaaatgc a**aagaatgct** **attagaaaag tgtgaccttc aaaattatgg tgatagtgca**

20761 **acattaccta a**aggcataat gatgaatgtc gcaaaatata ctcaactgtg tcaatattta

20821 aacacattaa cattagctgt accctataat atgagagtta tacattttgg tgctggttct

20881 gataaaggag ttgcaccagg tacagctgtt ttaagacagt ggttgcctac gggtacgctg

20941 cttgtcgatt cagatcttaa tgactttgtc tctgatgcag attcaacttt gattggtgat

21001 tgtgcaactg tacatacagc taataaatgg gatctcatta ttagtgatat gtacgaccct

21061 aagactaaaa atgttacaaa agaaaatgac tctaaagagg gttttttcac ttacatttgt

21121 gggtttatac aacaaaagct agctcttgga ggttccgtgg ctataaagat aacagaacat

21181 tcttggaatg ctgatcttta taagctcatg ggacacttcg catggtggac agcctttgtt

21241 actaatgtga atgcgtcatc atctgaagca tttttaattg gatgtaatta tcttggcaaa

21301 ccacgcgaac aaatagatgg ttatgtcatg catgcaaatt acatattttg gaggaataca

21361 aatccaattc agttgtcttc ctattcttta tttgacatga gtaaatttcc ccttaaatta

21421 aggggtactg ctgttatgtc tttaa**aagaa** ggtcaaatca atgatatgat tttatctctt

21481 cttagtaaag gtagacttat aattagagaa aacaacagag ttgttatttc tagtgatgtt

**21541** cttgttaaca actaaacgaa ca**atgtttgt** **ttttcttgtt ttattgccac tagtctctag**

**21601 tcagtgtgtt aatcttacaa ccagaactca attaccccct gcatacacta attctttcac**

**21661 acgtggtgtt tattaccctg acaaagtttt cagatcctca gttttacatt caactcagga**

**21721 cttgttctta cctttctttt ccaatgttac ttggttccat gctatacatg tctctgggac**

**21781 caatggtact aagaggtttg ataaccctgt cctaccattt aatgatggtg tttattttgc**

**21841 ttccactgag aagtctaaca taataagagg ctggattttt ggtactactt tagattcgaa**

**21901 gacccagtcc ctacttattg ttaataacgc tactaatgtt gttattaaag tctgtgaatt**

**21961 tcaattttgt aatgatccat ttttgggtgt ttattaccac aaaaacaaca aaagttggat**

**22021 ggaaagtgag ttcagagttt attctagtgc gaataattgc acttttgaat atgtctctca**

**22081 gccttttctt atggaccttg aaggaaaaca gggtaatttc aaaaatctta gggaatttgt**

**22141 gtttaagaat attgatggtt attttaaaat atattctaag cacacgccta ttaatttagt**

**22201 gcgtgatctc cctcagggtt tttcggcttt agaaccattg gtagatttgc caataggtat**

**22261 taacatcact aggtttcaaa ctttacttgc tttacataga agttatttga ctcctggtga**

**22321 ttcttcttca ggttggacag ctggtgctgc agcttattat gtgggttatc ttcaacctag**

**22381 gacttttcta ttaaaatata atgaaaatgg aaccattaca gatgctgtag actgtgcact**

**22441 tgaccctctc tcagaaacaa agtgtacgtt gaaatccttc actgtagaaa aaggaatcta**

**22501 tcaaacttct aactttagag tccaaccaac agaatctatt gttagatttc ctaatattac**

**22561 aaacttgtgc ccttttggtg aagtttttaa cgccaccaga tttgcatctg tttatgcttg**

**22621 gaacaggaag agaatcagca actgtgttgc tgattattct gtcctatata attccgcatc**

**22681 attttccact tttaagtgtt atggagtgtc tcctactaaa ttaaatgatc tctgctttac**

**22741 taatgtctat gcagattcat ttgtaattag aggtgatgaa gtcagacaaa tcgctccagg**

**22801 gcaaactgga aagattgctg attataatta taaattacca gatgatttta caggctgcgt**

**22861 tatagcttgg aattctaaca atcttgattc taaggttggt ggtaattata attacctgta**

**22921 tagattgttt aggaagtcta atctcaaacc ttttgagaga gatatttcaa ctgaaatcta**

**22981 tcaggccggt agcacacctt gtaatggtgt tgaaggtttt aattgttact ttcctttaca**

**23041 atcatatggt ttccaaccca ctaatggtgt tggttaccaa ccatacagag tagtagtact**

**23101 ttcttttgaa cttctacatg caccagcaac tgtttgtgga cctaaaaagt ctactaattt**

**23161 ggttaaaaac aaatgtgtca atttcaactt caatggttta acaggcacag gtgttcttac**

**23221 tgagtctaac aaaaagtttc tgcctttcca acaatttggc agagacattg ctgacactac**

**23281 tgatgctgtc cgtgatccac agacacttga gattcttgac attacaccat gttcttttgg**

**23341 tggtgtcagt gttataacac caggaacaaa tacttctaac caggttgctg ttctttatca**

**23401 ggatgttaac tgcacagaag tccctgttgc tattcatgca gatcaactta ctcctacttg**

**23461 gcgtgtttat tctacaggtt ctaatgtttt tcaaacacgt gcaggctgtt taataggggc**

**23521 tgaacatgtc aacaactcat atgagtgtga catacccatt ggtgcaggta tatgcgctag**

**23581 ttatcagact cagactaatt ctcctcggcg ggcacgtagt gtagctagtc aatccatcat**

**23641 tgcctacact atgtcacttg gtgcagaaaa ttcagttgct tactctaata actctattgc**

**23701 catacccaca aattttacta ttagtgttac cacagaaatt ctaccagtgt ctatgaccaa**

**23761 gacatcagta gattgtacaa tgtacatttg tggtgattca actgaatgca gcaatctttt**

**23821 gttgcaatat ggcagttttt gtacacaatt aaaccgtgct ttaactggaa tagctgttga**

**23881 acaagacaaa aacacccaag aagtttttgc acaagtcaaa caaatttaca aaacaccacc**

**23941 aattaaagat tttggtggtt ttaatttttc acaaatatta ccagatccat caaaaccaag**

**24001 caagaggtca tttattgaag atctactttt caacaaagtg acacttgcag atgctggctt**

**24061 catcaaacaa tatggtgatt gccttggtga tattgctgct agagacctca tttgtgcaca**

**24121 aaagtttaac ggccttactg ttttgccacc tttgctcaca gatgaaatga ttgctcaata**

**24181 cacttctgca ctgttagcgg gtacaatcac ttctggttgg acctttggtg caggtgctgc**

**24241 attacaaata ccatttgcta tgcaaatggc ttataggttt aatggtattg gagttacaca**

**24301 gaatgttctc tatgagaacc aaaaattgat tgccaaccaa tttaatagtg ctattggcaa**

**24361 aattcaagac tcactttctt ccacagcaag tgcacttgga aaacttcaag atgtggtcaa**

**24421 ccaaaatgca caagctttaa acacgcttgt taaacaactt agctccaatt ttggtgcaat**

**24481 ttcaagtgtt ttaaatgata tcctttcacg tcttgacaaa gttgaggctg aagtgcaaat**

**24541 tgataggttg atcacaggca gacttcaaag tttgcagaca tatgtgactc aacaattaat**

**24601 tagagctgca gaaatcagag cttctgctaa tcttgctgct actaaaatgt cagagtgtgt**

**24661 acttggacaa tcaaaaagag ttgatttttg tggaaagggc tatcatctta tgtccttccc**

**24721 tcagtcagca cctcatggtg tagtcttctt gcatgtgact tatgtccctg cacaagaaaa**

**24781 gaacttcaca actgctcctg ccatttgtca tgatggaaaa gcacactttc ctcgtgaagg**

**24841 tgtctttgtt tcaaatggca cacactggtt tgtaacacaa aggaattttt atgaaccaca**

**24901 aatcattact acagacaaca catttgtgtc tggtaactgt gatgttgtaa taggaattgt**

**24961 caacaacaca gtttatgatc ctttgcaacc tgaattagac tcattcaagg aggagttaga**

**25021 taaatatttt aagaatcata catcaccaga tgttgattta ggtgacatct ctggcattaa**

**25081 tgcttcagtt gtaaacattc aaaaagaaat tgaccgcctc aatgaggttg ccaagaattt**

**25141 aaatgaatct ctcatcgatc tccaagaact tggaaagtat gagcagtata taaaatggcc**

**25201 atggtacatt tggctaggtt ttatagctgg cttgattgcc atagtaatgg tgacaattat**

**25261 gctttgctgt atgaccagtt gctgtagttg tctcaagggc tgttgttctt gtggatcctg**

**25321 ctgcaaattt gatgaagacg actctgagcc agtgctcaaa ggagtcaaat tacattacac**

**25381** **ataa**acgaac ttatggattt gtttatgaga atcttcacaa ttggaactgt aactttgaag

25441 caaggtgaaa tcaaggatgc tactccttca gattttgttc gcgctactgc aacgataccg

25501 atacaagcct cactcccttt cggatggctt attgttggcg ttgcacttct tgctgttttt

25561 cagagcgctt ccaaaatcat aaccctcaaa aagagatggc aactagcact ctccaagggt

25621 gttcactttg tttgcaactt gctgttgttg tttgtaacag tttactcaca ccttttgctc

25681 gttgctgctg gccttgaagc cccttttctc tatctttatg ctttagtcta cttcttgcag

25741 agtataaact ttgt**aagaa**t aataatgagg ctttggcttt gctggaaatg ccgttccaaa

25801 aacccattac tttatgatgc caactatttt ctttgctggc atactaattg ttacgactat

25861 tgtatacctt acaatagtgt aacttcttca attgtcatta cttcaggtga tggcacaaca

25921 agtcctattt ctgaacatga ctaccagatt ggtggttata ctgaaaaatg ggaatctgga

25981 gtaaaagact gtgttgtatt acacagttac ttcacttcag actattacca gctgtactca

26041 actcaattga gtacagacac tggtgttgaa catgttacct tcttcatcta caataaaatt

26101 gttgatgagc ctg**aagaa**ca tgtccaaatt cacacaatcg acggttcatc cggagttgtt

26161 aatccagtaa tggaaccaat ttatgatgaa ccgacgacga ctactagcgt gcctttgtaa

26221 gcacaagctg atgagtacga acttatgtac tcattcgttt cggaagagac aggtacgtta

26281 atagttaata gcgtacttct ttttcttgct ttcgtggtat tcttgctagt tacactagcc

26341 atccttactg cgcttcgatt gtgtgcgtac tgctgcaata ttgttaacgt gagtcttgta

26401 aaaccttctt tttacgttta ctctcgtgtt aaaaatctga attcttctag agttcctgat

26461 cttctggtct aaacgaacta aatattatat tagtttttct gtttggaact ttaattttag

26521 ccatggcaga ttccaacggt actattaccg ttgaagagct taaaaagctc cttgaacaat

26581 ggaacctagt aataggtttc ctattcctta catggatttg tcttctacaa tttgcctatg

26641 ccaacaggaa taggtttttg tatataatta agttaatttt cctctggctg ttatggccag

26701 taactttagc ttgttttgtg cttgctgctg tttacagaat aaattggatc accggtggaa

26761 ttgctatcgc aatggcttgt cttgtaggct tgatgtggct cagctacttc attgcttctt

26821 tcagactgtt tgcgcgtacg cgttccatgt ggtcattcaa tccagaaact aacattcttc

26881 tcaacgtgcc actccatggc actattctga ccagaccgct tctagaaagt gaactcgtaa

26941 tcggagctgt gatccttcgt ggacatcttc gtattgctgg acaccatcta ggacgctgtg

27001 acatcaagga cctgcctaaa gaaatcactg ttgctacatc acgaacgctt tcttattaca

27061 aattgggagc ttcgcagcgt gtagcaggtg actcaggttt tgctgcatac agtcgctaca

27121 ggattggcaa ctataaatta aacacagacc attccagtag cagtgacaat attgctttgc

27181 ttgtacagta agtgacaaca gatgtttcat ctcgttgact ttcaggttac tatagcagag

27241 atattactaa ttattatgag gacttttaaa gtttccattt ggaatcttga ttacatcata

27301 aacctcataa ttaaaaattt atctaagtca ctaactgaga ataaatattc tcaattagat

27361 gaagagcaac caatggagat tgattaaacg aacatgaaaa ttattctttt cttggcactg

27421 ataacactcg ctacttgtga gctttatcac taccaagagt gtgttagagg tacaacagta

27481 cttttaaaag aaccttgctc ttctggaaca tacgagggca attcaccatt tcatcctcta

27541 gctgataaca aatttgcact gacttgcttt agcactcaat ttgcttttgc ttgtcctgac

27601 ggcgtaaaac acgtctatca gttacgtgcc agatcagttt cacctaaact gttcatcaga

27661 caagaggaag ttc**aagaa**ct ttactctcca atttttctta ttgttgcggc aatagtgttt

27721 ataacacttt gcttcacact caa**aagaa**ag acagaatgat tgaactttca ttaattgact

27781 tctatttgtg ctttttagcc tttctgctat tccttgtttt aattatgctt attatctttt

27841 ggttctcact tgaactgcaa gatcataatg aaacttgtca cgcctaaacg aacatgaaat

27901 ttcttgtttt cttaggaatc atcacaactg tagctgcatt tcacc**aagaa** tgtagtttac

27961 agtcatgtac tcaacatcaa ccatatgtag ttgatgaccc gtgtcctatt cacttctatt

28021 ctaaatggta tattagagta ggagctagaa aatcagcacc tttaattgaa ttgtgcgtgg

28081 atgaggctgg ttctaaatca cccattcagt acatcgatat cggtaattat acagtttcct

28141 gtttaccttt tacaattaat tgccaggaac ctaaattggg tagtcttgta gtgcgttgtt

28201 cgttctatga agacttttta gagtatcatg acgttcgtgt tgttttagat ttcatctaaa

28261 cgaacaaact aaaatgtctg ataatggacc ccaaaatcag cgaaatgcac cccgcattac

28321 gtttggtgga ccctcagatt caactggcag taaccagaat ggagaacgca gtggggcgcg

28381 atcaaaacaa cgtcggcccc aaggtttacc caataatact gcgtcttggt tcaccgctct

28441 cactcaacat ggcaaggaag accttaaatt ccctcgagga caaggcgttc caattaacac

28501 caatagcagt ccagatgacc aaattggcta ctaccgaaga gctaccagac gaattcgtgg

28561 tggtgacggt aaaatgaaag atctcagtcc aagatggtat ttctactacc taggaactgg

28621 gccagaagct ggacttccct atggtgctaa caaagacggc atcatatggg ttgcaactga

28681 gggagccttg aatacaccaa aagatcacat tggcacccgc aatcctgcta acaatgctgc

28741 aatcgtgcta caacttcctc aaggaacaac attgccaaaa ggcttctacg cagaagggag

28801 cagaggcggc agtcaagcct cttctcgttc ctcatcacgt agtcgcaaca gttc**aagaa**a

28861 ttcaactcca ggcagcagta ggggaacttc tcctgctaga atggctggca atggcggtga

28921 tgctgctctt gctttgctgc tgcttgacag attgaaccag cttgagagca aaatgtctgg

28981 taaaggccaa caacaacaag gccaaactgt cact**aagaa**a tctgctgctg aggcttctaa

29041 gaagcctcgg caaaaacgta ctgccactaa agcatacaat gtaacacaag ctttcggcag

29101 acgtggtcca gaacaaaccc aaggaaattt tggggaccag gaactaatca gacaaggaac

29161 tgattacaaa cattggccgc aaattgcaca atttgccccc agcgcttcag cgttcttcgg

29221 aatgtcgcgc attggcatgg aagtcacacc ttcgggaacg tggttgacct acacaggtgc

29281 catcaaattg gatgacaaag atccaaattt caaagatcaa gtcattttgc tgaataagca

29341 tattgacgca tacaaaacat tcccaccaac agagcctaaa aaggacaaaa ag**aagaa**ggc

29401 tgatgaaact caagccttac cgcagagaca g**aagaa**acag caaactgtga ctcttcttcc

29461 tgctgcagat ttggatgatt tctccaaaca attgcaacaa tccatgagca gtgctgactc

29521 aactcaggcc taaactcatg cagaccacac aaggcagatg ggctatataa acgttttcgc

29581 ttttccgttt acgatatata gtctactctt gtgcagaatg aattctcgta actacatagc

29641 acaagtagat gtagttaact ttaatctcac atagcaatct ttaatcagtg tgtaacatta

29701 gggaggactt gaaagagcca ccacattttc accgaggcca cgcggagtac gatcgagtgt

29761 acagtgaaca atgctaggga gagctgccta tatggaagag ccctaatgtg taaaattaat

29821 tttagtagtg ctatccccat gtgattttaa tagcttctta ggagaatgac aaaaaaaaaa

29881 aaaaaaaaaa aaaaaaaaaa aaa

//