BIO F242

Introduction To Bioinformatics

END-SEMESTER REPORT

Submitted by: Nishtha Pareek 2019B1A81044P

Submitted to: Professor Shibashish Chowdhury

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000040705/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), and envelope protein (E) genes, complete cds; membran…

Task 1: Detail Introduction of gene and its functions

Gene Name: surface glycoprotein UIT12537.1

Locus: OM250521 29849 bp RNA linear VRL 13-JAN-2022

Definition: Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000040705/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), and envelope protein (E) genes, complete cds; membrane glycoprotein (M) gene, partial cds; and ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds.

Accession: OM250521

Version: OM250521.1

DBLINK BioProject: PRJNA750736

BioSample: SAMN24854787

Source: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

Organism: Severe acute respiratory syndrome coronavirus 2

Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;

Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;

Betacoronavirus; Sarbecovirus

This gene is a severe acute respiratory syndrome coronavirus 2 isolate and is a genomic RNA. Its isolation source is clinical. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped, positive-sense, single-stranded RNA virus responsible for the Covid-19 pandemic. The positive-stranded RNA genome presents a 5’-cap and 3’-polyA tail, allowing its translation from host machinery. Similarly, to other CoVs, at the 5’-end, frameshift between Orfs, Orf1a and Orf1b leads to the production of two polypeptides that undergo proteolytic processing to produce 16 non-structural proteins, which direct various stages of the virus infection cycle such as virus assembly, transcription, replication, host control etc. The 3’-end encodes for structural proteins: S (spike glycoprotein), N (nucleocapsid protein), M (membrane protein) and E (Envelope

protein). The nucleocapsid protein binds to the viral genome and aids in the packing of the genome against the surface of the envelope. The other three structural proteins constitute the viral envelope. The 3'-end also encodes for the nine accessory proteins such as Orf3a, Orf3b, Orf6, Orf7a, Orf7b, Orf8, Orf9b, Orf9c, and Orf10. Accessory proteins play a key role in the virulence and host interaction of the SARS-CoV-2 virus.

The Spike Glycoprotein is a glycosylated type-I membrane protein existing in the trimeric pre-fusion form which is later cleaved into two sub-units S1 and S2, by a host Furin protease. The N-terminal S1 subunit consists of a Receptor-binding domain which helps in binding to the host cell receptor called angiotensin-converting enzyme 2 (ACE2). The binding of RBD to ACE2 and cleavage of the S2 subunit by host serine protease TMPRSS2 leads to dissociation between S1 and S2 subunits. The dissociation of subunits conformationally changes S2. The conformationally changed form of S2 is needed for the fusion of host and viral membranes and host entry. S protein works in tandem with other structural proteins N, E and M during membrane fusion in coronaviruses. The lower variability of the S2 sub fusion subunit makes it an ideal target for the development of inhibitors targeting viral entry.

The Envelope protein plays a vital role in viral morphogenesis and assembly. E protein acts as viroporins which assemble themselves into protein-lipid pores of the host responsible for forming membrane and are involved in ion transport. Interaction between the C-terminus of E and M proteins recruits E protein to the intermediary compartment of the Golgi-Endoplasmic Reticulum. This positioning of E-protein in the Golgi-ER intermediary compartment initiates the budding of the virus into the host cells. M proteins are the most abundant proteins in CoVs and attribute distinct shapes to the virus. M proteins also act as scaffolding platform to recruit other structural proteins and promotes membrane curvature during the budding of the virion. Nucleocapsid (N) protein is responsible for packaging viral RNA into ribonucleocapsid. N protein mediates viral assembly by interacting with the viral genome and M protein and boosting the replication and transcription of viral RNA. The N protein of SARS CoV-2 inhibits the activity of the cyclin-CDK complex, which hypo phosphorylates retinoblastoma protein and hence, inhibits the S phase progression of the cell cycle.

The accessory protein Orf3a forms an ion channel causing calcium influx. The ion channel is observed to be selective towards K^+ and Ca^{2+} over Na^+ in vitro. Orf3a channel-forming activity is associated with its pro-apoptotic activity in the SARS-CoV-2 virus. ORF3a also interacts with TRAF3, which activates ASC ubiquitination leading to caspase 1 activation and IL-1β maturation. ORF3a additionally interacts with structural proteins N, M and S to aid virus budding. ORF1ab contains overlapping open reading frames to encode polyproteins PP1ab and PP1a. ORF1a is the first open reading frame at the 5'-end of the genome. ORF1a, together with ORF2ab, occupies about 2/3 of the SARS-Cov-2 genome and is translated from 5'-capped RNA by cap-dependent translation. PP1ab is produced when a programmed ribosomal frameshift allows reading the stop codon, which terminates ORF-1a. The frameshift takes place in a slippery sequence followed by pseudoknot RNA secondary structure. These polyproteins are cleaved to yield 13 to 17 nonstructural proteins or NSPs. They release nonstructural proteins by auto-proteolysis and due to the action of internal cysteine protease domains. PP1a protein

contains nonstructural proteins nsp1 to nsp11, whereas pp1ab protein contains nsp1-10 and nsp12-16. Proteolytic processing is performed by two proteases: papin like protease domain located in multidomain protein nsp3 cleaving to nsp4. The second protease, 3CL protease, performs the remaining nsp5 cleavages through the C-terminus of the polyprotein. The C-terminal components of pp1ab polyprotein, that is, nsp12-16, have core enzymes for viral replication. After proteolytic processing is completed, non-structural proteins form a large protein complex called replicase-transcriptase complex (RTC), responsible for gene replication and transcription.

Task 2: Download Gene sequence, mRNA sequence, protein sequence and protein structure (if available)

Gene Sequence retrieved from FASTA (DNA):

>OM250521.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000040705/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), and envelope protein (E) genes, complete cds; membrane glycoprotein (M) gene, partial cds; and ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds

AGATCTGTTCTCTAAACGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACG

CAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCT

TACGGTTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGGTTTTGTCCGGGTGTGACCGAAAGGTAA

GATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACGTCCAACTCAGTTTGCCTGTTTTACAGGTT

CGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTA

AAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTT

CATCAAACGTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACTCGAA

GGCATTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGTGGGCGAAATACCAGTGG

CTTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCT

AAAGTCATTTGACTTAGGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACTGGAACACT

AAACATAGCAGTGGTGTTACCCGTGAACTCATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGTCG

ATAACAACTTCTGTGGCCCTGATGGCTACCCTCTTGAGTGCATTAAAGACCTTCTAGCACGTGCTGGTAA

AGCTTCATGCACTTTGTCCGAACAACTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAA

CATGAGCATGAAATTGCTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGAAA

TTAAATTGGCAAAGAAATTTGACACCTTCAATGGGGAATGTCCAAATTTTGTATTTCCCTTAAATTCCAT

AATCAAGACTATTCAACCAAGGGTTGAAAAGAAAAAGCTTGATGGCTTTATGGGTAGAATTCGATCTGTC

TATCCAGTTGCGTCACCAAATGAATGCAACCAAATGTGCCTTTCAACTCTCATGAAGTGTGATCATTGTG

GTGAAACTTCATGGCAGACGGGCGATTTTGTTAAAGCCACTTGCGAATTTTGTGGCACTGAGAATTTGAC

TAAAGAAGGTGCCACTACTTGTGGTTACTTACCCCAAAATGCTGTTGTTAAAATTTATTGTCCAGCATGT
CACAATTCAGAAGTAGGACCTGAGCATAGTCTTGCCGAATACCATAATGAATCTGGCTTGAAAACCATTC
TTCGTAAGGGTGGTCGCACTATTGCCTTTGGAGGCTGTGTGTTCTCTTATGTTGGTTGCCATAACAAGTG
TGCCTATTGGGTTCCACGTGCTAGCGCTAACATAGGTTGTAACCATACAGGTGTTGTTGGAGAAGGTTCC
GAAGGTCTTAATGACNNNNNNNNTGAAATACTCCAAAAAGAGAAAGTCAACATCAATATTGTTGGTGACT
TTAAACTTAATGAAGAGATCGCCATTATTTTGGCATCTTTTTCTGCTTCCACAAGTGCTTTTGTGGAAAC
TGTGAAAGGTTTGGATTATAAAGCATTCAAACAAATTGTTGAATCCTGTGGTAATTTTAAAGTTACAAAA
GGAAAAGCTAAAAAAGGTGCCTGGAATATTGGTGAACAGAAATCAATACTGAGTCCTCTTTATGCATTTG
CATCAGAGGCTGCTCGTGTTGTACGATCAATTTTCTCCCGCACTCTTGAAACTGCTCAAAATTCTGTGCG
TGTTTTACAGAAGGCCGCTATAACAATACTAGATGGAATTTCACAGTATTCACTGAGACTCATTGATGCT
ATGATGTTCACATCTGATTTGGCTACTAACAATCTAGTTGTAATGGCCTACATTACAGGTGGTGTTGTTC
AGTTGACTTCGCAGTGGCTAACTAACATCTTTGGCACTGTTTATGAAAAACTCAAACCCGTCCTTGATTG
GCTTGAAGAGAAGTTTAAGGAAGGTGTAGAGTTTCTTAGAGACGGTTGGGAAATTGTTAAATTTATCTCA
ACCTGTGCTTGTGAAATTGTCGGTGGACAAATTGTCACCTGTGCAAAGGAAATTAAGGAGAGTGTTCAGA
CATTCTTTAAGCTTGTAAATAAATTTTTGGCTTTGTGTGCTGACTCTATCATTATTGGTGGAGCTAAACT
TAAAGCCTTGAATTTAGGTGAAACATTTGTCACGCACTCAAAGGGATTGTACAGAAAGTGTGTTAAATCC
AGAGAAGAAACTGGCCTACTCATGCCTCTAAAAGCCCCAAAAGAAATTATCTTCTTAGAGGGAGAAACAC
TTCCCACAGAAGTGTTAACAGAGGAAGTTGTCTTGAAAACTGGTGATTTACAACCATTAGAACAACCTAC
TAGTGAAGCTGTTGAAGCTCCATTGGTTGGTACACCAGTTTGTATTAACGGGCTTATGTTGCTCGAAATC
AAAGACACAGAAAAGTACTGTGCCCTTGCACCTAATATGATGGTAACAAACAATACCTTCACACTCAAAG
GCGGTGCACCAACAAAGGTTACTTTTGGTGATGACACTGTGATAGAAGTGCAAGGTTACAAGAGTGTGAA
TATCACTTTTGAACTTGATGAAAGGATTGATAAAGTACTTAATGAGAAGTGCTCTGCCTATACAGTTGAA
CTCGGTACAGAAGTAAATGAGTTCGCCTGTGTTGTGGCAGATGCTGTCATAAAAACTTTGCAACCAGTAT
CTGAATTACTTACACCACTGGGCATTGATTTAGATGAGTGGAGTATGGCTACATACTACTTATTTGATGA
GTCTGGTGAGTTTAAATTGGCTTCACATATGTATTGTTCTTTTTACCCTCCAGATGAGGATGAAGAAGAA
GGTGATTGTGAAGAAGAAGAGTTTGAGCCATCAACTCAATATGAGTATGGTACTGAAGATGATTACCAAG
GTAAACCTTTGGAATTTGGTGCCACTTCTGCTGCTCTTCAACCTGAAGAAGAGCAAGAAGAAGATTGGTT
AGATGATGATAGTCAACAAACTGTTGGTCAACAAGACGGCAGTGAGGACAATCAGACAACTACTATTCAA
ACAATTGTTGAGGTTCAACCTCAATTAGAGATGGAACTTACACCAGTTGTTCAGACTATTGAAGTGAATA
GTTTTAGTGGTTATTTAAAACTTACTGACAATGTATACATTAAAAATGCAGACATTGTGGAAGAAGCTAA
AAAGGTAAAACCAACAGTGGTTGTTAATGCAGCCAATGTTTACCTTAAACATGGAGGAGGTGTTGCAGGA
GCCTTAAATNNNGCTACTAACAATGCCATGCAAGTTGAATCTGATGATTACATAGCTACTAATGGACCAC
TTAAAGTGGGTGGTAGTTGTGTTTTAAGCGGACACAATCTTGCTAAACACTGTCTTCATGTTGTCGGCCC
AAATGTTAACAAAGGTGAAGACATTCAACTTCTTAAGAGTGCTTATGAAAATTTTAATCAGCACGAAGTT
CTACTTGCACCATTATTATCAGCTGGTATTTTTGGTGCTGACCCTATACATTCTTTAAGAGTTTGTGTAG

ATACTGTTCGCACAAATGTCTACTTAGCTGTCTTTGATAAAAATCTCTATGACAAACTTGTTTCAAGCTT
TTTGGAAATGAAGAGTGAAAAGCAAGTTGAACAAAAGATCGCTGAGATTCCTAAAGAGGAAGTTAAGCCA
TTTATAACTGAAAGTAAACCTTCAGTTGAACAGAGAAAACAAGATGATAAGAAAATCAAAGCTTGTGTTG
AAGAAGTTACAACAACTCTGGAAGAAATTAAGTTCCTCACAGAAAACTTGTTACTTTATATTGACATTAA
TGGCAATCTTCATCCAGATTCTGCCACTCTTGTTAGTGACATTGACATCACTTTCTTAAAGAAAGATGCT
CCATATATAGTGGGTGATGTTGTTCAAGAGGGTGTTTTAACTGCTGTGGTTATACCTACTAAAAAGGCTG
GTGGCACTACTGAAATGCTAGCGAAAGCTTTGAGAAAAGTGCCAACAGACAATTATATAACCACTTACCC
GGGTCAGGGTTTAAATGGTTACACTGTAGAGGAGGCAAAGACAGTGCTTAAAAAGTGTAAAAGTGCCTTT
TACATTCTACCATCTATTATCTCTAATGAGAAGCAAGAAATTCTTGGAACTGTTTCTTGGAATTTGCGAG
AAATGCTTGCACATGCAGAAGAAACACGCAAATTAATGCCTGTCTGTGTGGAAACTAAAGCCATAGTTTC
AACTATACAGCGTAAATATAAGGGTATTAAAATACAAGAGGGTGTGGTTGATTATGGTGCTAGATTTTAC
TTTTACACCAGTAAAACAACTGTAGCGTCACTTATCAACACACTTAACGATCTAAATGAAACTCTTGTTA
CAATGCCACTTGGCTATGTAACACATGGCTTAAATTTGGAAGAAGCTGCTCGGTATATGAGATCTCTCAA
AGTGCCAGCTACAGTTTCTGTTTCTTCACCTGATGCTGTTACAGCGTATAATGGTTATCTTACTTCTTCT
TCTAAAACACCTGAAGAACATTTTATTGAAACCATCTCACTTGCTGGTTCCTATAAAGATTGGTCCTATT
CTGGACAATCTACACAACTAGGTATAGAATTTCTTAAGAGAGGTGATAAAAGTGTATATTACACTAGTAA
TCCTACCACATTCCACCTAGATGGTGAAGTTATCACCTTTGACAATCTTAAGACACTTCTTTCTTTGAGA
GAAGTGAGGACTATTAAGGTGNNNNNNACAGTAGACAACATTAACCTCCACACGCAAGTTGTGGACATGT
CAATGACATATGGACAACAGTTTGGTCCAACTTATTTGGATGGAGCTGATGTTACTAAAATAAAACCTCA
TAATTCACATGAAGGTAAAACATTTTATGTTTTACCTAATGATGACACTCTACGTGTTGAGGCTTTTGAG
TACTACCACACAACTGATCCTAGTTTTCTGGGTAGGTACATGTCAGCATTAAATCACACTAAAAAGTGGA
AATACCCACAAGTTAATGGTTTAACTTCTATTAAATGGGCAGATAACAACTGTTATCTTGCCACTGCATT
GTTAACACTCCAACAAATAGAGTTGAAGTTTAATCCACCTGCTCTACAAGATGCTTATTACAGAGCAAGG
GCTGGTGAAGCTGCTAACTTTTGTGCACTTATCTTAGCCTACTGTAATAAGACAGTAGGTGAGTTAGGTG
ATGTTAGAGAAACAATGAGTTACTTGTTTCAACATGCCAATTTAGATTCTTGCAAAAGAGTCTTGAACGT
GGTGTGTAAAACTTGTGGACAACAGCAGACAACCCTTAAGGGTGTAGAAGCTGTTATGTACATGGGCACA
CTTTCTTATGAACAATTTAAGAAAGGTGTTCAGATACCTTGTACGTGTGGTAAACAAGCTACAAAATATC
TAGTACAACAGGAGTCACCTTTTGTTATGATGTCAGCACCACCTGCTCAGTATGAACTTAAGCATGGTAC
ATTTACTTGTGCTAGTGAGTACACTGGTAATTACCAGTGTGGTCACTATAAACATATAACTTCTAAAGAA
ACTTTGTATTGCATAGACGGTGCTTTACTTACAAAGTCCTCAGAATACAAAGGTCCTATTACGGATGTTT
TCTACAAAGAAAACAGTTACACAACAACCATAAAACCAGTTACTTATAAATTGGATGGTGTTGTTTGTAC
AGAAATTGACCCTAAGTTGGACAATTATTATAAGAAAGACAATTCTTATTTCACAGAGCAACCAATTGAT
CTTGTACCAAACCAACCATATCCAAACGCAAGCTTCGATAATTTTAAGTTTGTATGTGATAATATCAAAT
TTGCTGATGATTTAAACCAGTTAACTGGTTATAAGAAACCTGCTTCAAGAGAGCTTAAAGTTACATTTTT
CCCTGACTTAAATGGTGATGTGGTGGCTATTGATTATAAACACTACACACCCTCTTTTAAGAAAGGAGCT

AAATTGTTACATAAACCTATTGTTTGGCATGTTAACAATGCAACTAATAAAGCCACGTATAAACCAAATA
CCTGGTGTATACGTTGTCTTTGGAGCACAAAACCAGTTGAAACATCAAATTCGTTTGATGTACTGAAGTC
AGAGGACGCGCAGGGAATGGATAATCTTGCCTGCGAAGATCTAAAACCAGTCTCTGAAGAAGTAGTGGAA
AATCCTACCATACAGAAAGACGTTCTTGAGTGTAATGTGAAAACTACCGAAGTTGTAGGAGACATTATAC
TTAAACCAGCAAATAATAGTTTAAAAATTACAGAAGAGGTTGGCCACACAGATCTAATGGCTGCTTATGT
AGACAATTCTAGTCTTACTATTAAGAAACCTAATGAATTATCTAGAGTATTAGGTTTGAAAACCCTTGCT
ACTCATGGTTTAGCTGCTGTTAATAGTGTCCCTTGGGATACTATAGCTAATTATGCTAAGCCTTTTCTTA
ACAAAGTTGTTAGTACAACTACTAACATAGTTACACGGTGTTTAAACCGTGTTTGTACTAATTATATGCC
TTATTTCTTTACTTTATTGCTACAATTGTGTACTTTTACTAGAAGTACAAATTCTAGAATTAAAGCATCT
ATGCCGACTACTATAGCAAAGAATACTGTTAAGAGTGTCGGTAAATTTTGTCTAGAGGCTTCATTTAATT
ATTTGAAGTCACCTAATTTTTCTAAACTGATAAATATTATAATTTGGTTTTTACTATTAAGTGTTTGCCT
AGGTTCTTTAATCTACTCAACCGCTGCTTTAGGTGTTTTAATGTCTAATTTAGGCATGCCTTCTTACTGT
ACTGGTTACAGAGAAGGCTATTTGAACTCTACTAATGTCACTATTGCAACCTACTGTACTGGTTCTATAT
CTTGTAGTGTTTGTCTTAGTGGTTTAGATTCTTTAGACACCTATCCTTCTTTAGAAACTATACAAATTAC
CATTTCATCTTTTAAATGGGATTTAACTGCTTTTGGCTTAGTTGCAGAGTGGTTTTTGGCATATATTCTT
TTCACTAGGTTTTTCTATGTACTTGGATTGGCTGCAATCATGCAATTGTTTTTCAGCTATTTTGCAGTAC
ATTTTATTAGTAATTCTTGGCTTATGTGGTTAATAATTAATCTTGTACAAATGGCCCCGATTTCAGCTAT
GGTTAGAATGTACATCTTCGTTGCATCATTTTATTATGTATGGAAAAGTTATGTGCATGTTGTAGACGGT
TGTAATTCATCAACTTGTATGATGTGTTACAAACGTAATAGAGCAACAAGAGTCGAATGTACAACTATTG
TTAATGGTGTTAGAAGGTCCTTTTATGTCTATGCTAATGGAGGTAAAGGCTTTTGCAAACTACACAATTG
GAATTGTGTTAATTGTGATACATTCTGTGCTGGTAGTACATTTATTAGTGATGAAGTTGCGAGAGACTTG
TCACTACAGTTTAAAAGACCAATAAATCCTACTGACCAGTCTTCTTACATCGTTGATAGTGTTACAGTGA
AGAATGGTTCCATCCATCTTTACTTTGATAAAGCTGGTCAAAAGACTTATGAAAGACATTCTCTCTCTCA
TTTTGTTAACTTAGACAACCTGAGAGCTAATAACACTAAAGGTTCATTGCCTATTAATGTTATAGTTTTT
GATGGTAAATCAAAATGTGAAGAATCATCTGCAAAATCAGCGTCTGTTTACTACAGTCAGCTTATGTGTC
AACCTATACTGTTNNNNNNNNAGGCATTAGTGTCTGATGTTGGTGATAGTGCGGAAGTTGCAGTTAAAAT
GTTTGATGCTTACGTTAATACGTTTTCATCAACTTTTAACGTACCAATGGAAAAACTCAAAACACTAGTT
GCAACTGCAGAAGCTGAACTTGCAAAGAATGTGTCCTTAGACAATGTCTTATCTACTTTTATTTCAGCAG
CTCGGCAAGGGTTTGTTGATTCAGATGTAGAAACTAAAGATGTTGTTGAATGTCTTAAATTGTCACATCA
ATCTGACATAGAAGTTACTGGCGATAGTTGTAATAACTATATGCTCACCTATAACAAAGTTGAAAACATG
ACACCCCGTGACCTTGGTGCTTGTATTGACTGTAGTGCGCGTCATATTAATGCGCAGGTAGCAAAAAGTC
ACAACATTGCTTTGATATGGAACGTTAAAGATTTCATGTCATTGTCTGAACAACTACGAAAACAAATACG
TAGTGCTGCTAAAAAGAATAACTTACCTTTTAAGTTGACATGTGCAACTACTAGACAAGTTGTTAATGTT
GTAACAACAAAGATAGCACTTAAGGGTGGTAAAATTGTTAATAATTGGTTGAAGCAGTTAATTAAAGTTA
CACTTGTGTTCCTTTTTGTTGCTGCTATTTTCTATTTAATAACACCTGTTCATGTCATGTCTAAACATAC

TGACTTTTCAAGTGAAATCATAGGATACAAGGCTATTGATGGTGGTGTCACTCGTGACATAGCATCTACA
GATACTTGTTTTGCTAACAAACATGCTGATTTTGACACATGGTTTAGCCAGCGTGGTGGTAGTTATACTA
ATGACAAAGCTTGCCCATTGATTGCTGCAGTCATAACAAGAGAAGTGGGTTTTGTCGTGCCTGGTTTGCC
TGGCACGATATTACGCACAACTAATGGTGACTTTTTGCATTTCTTACCTAGAGTTTTTAGTGCAGTTGGT
AACATCTGTTACACACCATCAAAACTTATAGAGTACACTGACTTTGCAACATCAGCTTGTGTTTTGGCTG
CTGAATGTACAATTTTTAAAGATGCTTCTGGTAAGCCAGTACCATATTGTTATGATACCAATGTACTAGA
AGGTTCTGTTGCTTATGAAAGTTTACGCCCTGACACACGTTATGTGCTCATGGATGGCTCTATTATTCAA
TTTCCTAACACCTACCTTGAAGGTTCTGTTAGAGTGGTAACAACTTTTGATTCTGAGTACTGTAGGCACG
GCACTTGTGAAAGATCAGAAGCTGGTGTTTGTGTATCTACTAGTGGTAGATGGGTACTTAACAATGATTA
TTACAGATCTTTACCAGGAGTTTTCTGTGGTGTAGATGCTGTAAATTTACTTACTAATATGTTTACACCA
CTAATTCAACCTATTGGTGCTTTGGACATATCAGCATCTATAGTAGCTGGTGGTATTGTAGCTATCGTAG
TAACATGCCTTGCCTACTATTTTATGAGGTTTAGAAGAGCTTTTGGTGAATACAGTCATGTAGTTGCCTT
TAATACTTTACTATTCCTTATGTCATTCACTGTACTCTGTTTAACACCAGTTTACTCATTCTTACCTGGT
GTTTATTCTGTTATTTACTTGTACTTGACATTTTATCTTACTAATGATGTTTCTTTTTTAGCACATATTC
AGTGGATGGTTATGTTCACACCTTTAGTACCTTTCTGGATAACAATTGCTTATATCATTTGTATTTCCAC
AAAGCATTTCTATTGGTTCTTTAGTAATTACCTAAAGAGACGTGTAGTCTTTAATGGTGTTTCCTTTAGT
ACTTTTGAAGAAGCTGCGCTGTGCACCTTTTTGTTNNNNAAAGAAATGTATCTAAAGTTGCGTAGTGATG
TGTTATTACCTCCTACGCAATATAATAGATACTTAGCTCTTTATAATAAGTACAAGTATTTTAGTGGAGC
AATGGATACAACTAGCTACAGAGAAGCTGCTTGTTGTCATCTCGCAAAGGCTCTCAATGACTTCAGTAAC
TCAGGTTCTGATGTTCTTTACCAACCACCACAAATCTCTATCACCTCAGCTGTTTTGCAGAGTGGTTTTA
GAAAAATGGCATTCCCATCTGGTAAAGTTGAGAGTTGTATGGTACAAGTAACTTGTGGTACAACTACACT
TAACGGTCTTTGGCTTGATGACGTAGTTTACTGTCCAAGACATGTGATCTGCACCTCTGAAGACATGCTT
AACCCTAATTATGAAGATTTACTCATTCGTAAGTCTAATCATAATTTCTTGGTACAGGCTGGTAATGTTC
AACTCAGGGTTATTGGACATTCTATGCAAAATTGTGTACTTAAGCTTAAGGTTGATACAGCCAATCCTAA
GACACCTAAGTATAAGTTTGTTCGCATTCAACCAGGACAGACTTTTTCAGTGTTAGCTTGTTACAATGGT
TCACCATCTGGTGTTTACCAATGTGCTATGAGGCCCAATTTCACTATTAAGGGTTCATTCCTTAATGGTT
CATGTGGTAGTGTTGGTTTTAACATAGATTATGACTGTGTCTCTTTTTGTTACATGCACCATATGGAATT
ACCAACTGGAGTTCATGCTGGCACAGACTTAGAAGGTAACTTTTATGGACCTTTTGTTGACAGGCAAACA
GCACAAGCAGCTGGTACGGACACAACTATTACAGTTAATGTTTTAGCTTGGTTGTACGCTGCTGTTATAA
ATGGAGACAGGTGGTTTCTCAATCGATTTACCACAACTCTTAATGACTTTAACCTTGTGGCTATGAAGTA
CAATTATGAACCTCTAACACAAGACCATGTTGACATACTAGGACCTCTTTCTGCTCAAACTGGAATTGCC
GTTTTAGATATGTGTGCTTCATTAAAAGAATTACTGCAAAATGGTATGAATGGACGTACCATATTGGGTA
GTGCTTTATTAGAAGATGAATTTACACCTTTTGATGTTGTTAGACAATGCTCAGGTGTTACTTTCCAAAG
TGCAGTGAAAAGAACAATCAAGGGTACACACCACTGGTTGTTACTCACAATTTTGACTTCNNNNNTAGTT
TTAGTCCAGAGTACTCAATGGTCTTTGTTCTTTTTTTTGTATGAAAATGCCTTTTTACCTTTTGCTATGG

GTATTATTGCTATGTCTGCTTTTGCAATGATGTTTGTCAAACATAAGCATGCATTTCTCTGTTTGTTTTT
GTTACCTTCTCTTGCCACTGTAGCTTATTTTAATATGGTCTATATGCCTGCTAGTTGGGTGATGCGTATT
ATGACATGGTTGGATATGGTTGATACTAGTTTGANNAGTTTGAAGCTAAAAGACTGTGTTATGTATGCAT
CAGCTGTAGTGTTACTAATCCTTATGACAGCAAGAACTGTGTATGATGATGGTGCTAGGAGAGTGTGGAC
ACTTATGAATGTCTTGACACTCGTTTATAAAGTTTATTATGGTAATGCTTTAGATCAAGCCATTTCCATG
TGGGCTCTTATAATCTCTGTTACTTCTAACTACTCAGGTGTAGTTACAACTGTCATGTTTTTGGCCAGAG
GTATTGTTTTTATGTGTGTTGAGTATTGCCCTATTTTCTTCATAACTGGTAATACACTTCAGTGTATAAT
GCTAGTTTATTGTTTCTTAGGCTATTTTTGTACTTGTTACTTTGGCCTCTTTTGTTTACTCAACCGCTAC
TTTAGACTGACTCTTGGTGTTTATGATTACTTAGTTTCTACACAGGAGTTTAGATATATGAATTCACAGG
GACTACTCCCACCCAAGAATAGCATAGATGCCTTCAAACTCAACATTAAATTGTTGGGTGTTGGTGGCAA
ACCTTGTATCAAAGTAGCCACTGTACAGTCTAAAATGTCAGATGTAAAGTGCACATCAGTAGTCTTACTC
TCAGTTTTGCAACAACTCAGAGTAGAATCATCATCTAAATTGTGGGCTCAATGTGTCCAGTTACACAATG
ACATTCTCTTAGCTAAAGATACTACTGAAGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTTGCTTTC
CATGCAGGGTGCTGTAGACATAAACAAGCTTTGTGAAGAAATGCTGGACAACAGGGCAACCTTACAAGCT
ATAGCCTCAGAGTTTAGTTCCCTTCCATCATATGCAGCTTTTGCTACTGCTCAAGAAGCTTATGAGCAGG
CTGTTGCTAATGGTGATTCTGAAGTTGTTCTTAAAAAGTTGAAGAAGTCTTTGAATGTGGCTAAATCTGA
ATTTGACCGTGATGCAGCCATGCAACGTAAGTTGGAAAAGATGGCTGATCAAGCTATGACCCAAATGTAT
AAACAGGCTAGATCTGAGGACAAGAGGGCAAAAGTTACTAGTGCTATGCAGACAATGCTTTTCACTATGC
TTAGAAAGTTGGATAATGATGCACTCAACAACATTATCAACAATGCAAGAGATGGTTGTGTTCCCTTGAA
CATAATACCNNNNACAACAGCAGCCAAACTAATGGTTGTCATACCAGACTATAACACATATAAAAATACG
TGTGATGGTACAACATTTACTTATGCATCAGCATTGTGGGAAATCCAACAGGTTGTAGATGCAGATAGTA
AAATTGTTCAACTTAGTGAAATTAGTATGGACAATTCACCTAATTTAGCATGGCCTCTTATTGTAACAGC
TTTAAGGGCCAATTCTGCTGTCAAATTACAGAATAATGAGCTTAGTCCTGTTGCACTACGACAGATGTCT
TGTGCTGCCGGTACTACACAAACTGCTTGCACTGATGACAATGCGTTAGCTTACTACAACACAACAAAGG
GAGGTAGGTTTGTACTTGCACTGTTATCCGATTTACAGGATTTGAAATGGGCTAGATTCCCTAAGAGTGA
TGGAACTGGTACTATCTATACAGAACTGGAACCACCTTGTAGGTTTGTTACAGACACACCTAAAGGTCCT
AAAGTGAAGTATTTATACTTTATTAAAGGATTAAACAACCTAAATAGAGGTATGGTACTTGGTAGTTTAG
CTGCCACAGTACGTCTACAAGCTGGTAATGCAACAGAAGTGCCTGCCAATTCAACTGTATTATCTTTCTG
TGCTTTTGCTGTAGATGCTGCTAAAGCTTACAAAGATTATCTAGCTAGTGGGGGACAACCAATCACTAAT
TGTGTTAAGATGTTGTGTACACACACTGGTACTGGTCAGGCAATAACAGTTACACCGGAAGCCAATATGG
ATCAAGAATCCTTTGGTGGTGCATCGTGTTGTCTGTACTGCCGTTGCCACATAGATCATCCAAATCCTAA
AGGATTTTGTGACTTAAAAGGTAAGTATGTACAAATACCTACAACTTGTGCTAATGACCNNNNNNNNNNN
NNNNNNNNNNNCACAGTCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGTTGTGATCAACTCC
GCGAACCCATGCTTCAGTCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCGGTGTAAGTGCAGCCCGT
CTTACACCGTGCGGCACAGGCACTAGTACTGATGTCGTATATAGGGCTTTTGACATCTACAATGATAAAG

TAGCTGGTTTTGCTAAATTCCTAAAAACTAATTGTTGTCGCTTCCAAGAAAAGGACGAAGATGACAATTT
AATTGATTCTTACTTTGTAGTTAAGANNNNCACTTTCTCTAACTACCAACATGAAGAAACAATTTATAAT
TTACTTAAGGATTGTCCAGCTGTTGCTAAACATGACTTCTTTAAGTTTAGAATAGACGGTGACATGGTAC
CACATATATCACGTCAACGTCTTACTAAATACACAATGGCAGACCTCGTCTATGCTTTAAGGCATTTTGA
TGAAGGTAATTGTGACACATTAAAAGAAATACTTGTCACATACAATTGTTGTGATGATGATTATTTCAAT
AAAAAGGACTGGTATGATTTTGTAGAAAACCCAGATATATTACGCGTATACGCCAACTTAGGTGAACGTG
TACGCCAAGCTTTGTTAAAAACAGTACAATTCTGTGATGCCATGCGAAATGCTGGTATTGTTGGTGTACT
GACATTAGATAATCAAGATCTCAATGGTAACTGGTATGATTTCGGTGATTTCATACAAACCACGCCAGGT
AGTGGAGTTCCTGTTGTAGATTCTTATTATTCATTGTTAATGCCTATATTAACCTTGACCAGGGCTTTAA
CTGCAGAGTCACATGTTGACACTGACTTAACAAAGCCTTACATTAAGTGGGATTTGTTAAAATATGACTT
CACGGAAGAGAGGTTAAAACTCTTTGACCGTTATTTTAAATATTGGGATCAGACATACCACCCAAATTGT
GTTAACTGTTTGGATGACAGATGCATTCTGCATTGTGCAAACTTTAATGTTTTATTCTCTACAGTGTTCC
CACTTACAAGTTTTGGACCACTAGTGAGAAAAATATTTGTTGATGGTGTTCCATTTGTAGTTTCAACTGG
ATACCACTTCAGAGAGCTAGGTGTTGTACATAATCAGGATGTAAACTTACATAGCTCTAGACTTAGTTTT
AAGGAATTACTTGTGTATGCTGCTGACCCTGCTATGCACGCTGCTTCTGGTAATCTATTACTAGATAAAC
GCACTACGTGCTTTTCAGTAGCTGCACTTACTAACAATGTTGCTTTTCAAACTGTCAAACCCGGTAATTT
TAACAAAGACTTCTATGACTTTGCTGTGTCTAAGGGTTTCTTTAAGGAAGGAAGTTCTGTTGAATTAAAA
CACTTCTTCTTTGCTCAGGATGGTAATGCTGCTATCAGCGATTATGACTACTATCGTTATAATCTACCAA
CAATGTGTGATATCAGACAACTACTATTTGTAGTTGAAGTTGTTGATAAGTACTTTGATTGTTACGATGG
TGGCTGTATTAATGCTAACCAAGTCATCGTCAACAACCTAGACAAATCAGCTGGTTTTCCATTTAATAAA
TGGGGTAAGGCTAGACTTTATTATGATTCAATGAGTTATGAGGATCAAGATGCACTTTTCGCATATACAA
AACGTAATGTCATCCCTACTATAACTCAAATGAATCTTAAGTATGCCATTAGTGCAAAGAATAGAGCTCG
CACCGTAGCTGGTGTCTCTATCTGTAGTACTATGACCAATAGACAGTTTCATCAAAAATTATTGAAATCA
ATAGCCGCCACTAGAGGAGCTACTGTAGTAATTGGAACAAGCAAATTCTATGGTGGTTGGCACAACATGT
TAAAAACTGTTTATAGTGATGTAGAAAACCCTCACCTTATGGGTTGGGATTATCCTAAATGTGATAGAGC
CATGCCTAACATGCTTAGAATTATGGCCTCACTTGTTCTTGCTCGCAAACATACAACGTGTTGTAGCTTG
TCACACCGTTTCTATAGATTAGCTAATGAGTGTGCTCAAGTATTGAGTGAAATGGTCATGTGTGGCGGTT
CACTATATGTTAAACCAGGTGGAACCTCATCAGGAGATGCCACAACTGCTTATGCTAATAGTGTTTTTAA
CATTTGTCAAGCTGTCACGGCCAATGTTAATGCACTTTTATCTACTGATGGTAACAAAATTGCCGATAAG
TATGTCCGCAATTTACAACACAGACTTTATGAGTGTCTCTATAGAAATAGAGATGTTGACACAGACTTTG
TGAATGAGTTTTACGCATATTTGCGTAAACATTTCTCAATGATGATACTCTCTGACGATGCTGTTGTGTG
TTTCAATAGCACTTATGCATCTCAAGGTCTAGTGGCTAGCATAAAGAACTTTAAGTCAGTTCTTTATTAT
CAAAACAATGTTTTTATGTCTGAAGCAAAATGTTGGACTGAGACTGACCTTACTAAAGGACCTCATGAAT
TTTGCTCTCAACATACAATGCTAGTTAAACAGGGTGATGATTATGTGTACCTTCCTTACCCAGATCCATC
AAGAATCCTAGGGGCCGGCTGTTTTGTAGATGATATCGTAAAAACAGATGGTACACTTATGATTGAACGG

TTCGTGTCTTTAGCTATAGATGCTTACCCACTTACTAAACATCCTAATCAGGAGTATGCTGATGTCTTTC
ATTTGTACTTACAATACATAAGAAAGCTACATGATGAGTTAACAGGACACATGTTAGACATGTATTCTGT
TATGCTTACTAATGATAACACTTCAAGGTATTGGGAACCTGAGTTTTATGAGGCTATGTACACACCGCAT
ACAGTCTTACAGGCTGTTGGGGCTTGTGTTCTTTGCAATTCACAGACTTCATTAAGATGTGGTGCTTGCA
TACGTAGACCATTCTTATGTTGTAAATGCTGTTACGACCATGTCATATCAACATCACATAAATTAGTCTT
GTCTGTTAATCCGTATGTTTGCAATGCTCCAGGTTGTGATGTCACAGATGTGACTCAACTTTACTTAGGA
GGTATGAGCTATTATTGTAAATCACATAAACCACCCATTAGTTTTCCATTGTGTGCTAATGGACAAGTTT
TTGGTTTATATAAAAATACATGTGTTGGTAGCGATAATGTTACTGACTTTAATGCAATTGCAACATGTGA
CTGGACAAATGCTGGTGATTACATTTTAGCTAACACCTGTACTGAAAGACTCAAGCTTTTTGCAGCAGAA
ACGCTCAAAGCTACTGAGGAGACATTTAAACTGTCTTATGGTATTGCTACTGTACGTGAAGTGCTGTCTG
ACAGAGAATTACATCTTTCATGGGAAGTTGGTAAACCTAGACCACCACTTAACCGAAATTATGTCTTTAC
TGGTTATCGTGTAACTAAAAACAGTAAAGTACAAATAGGAGAGTACACCTTTGAAAAAGGTGACTATGGT
GATGCTGTTGTTTACCGAGGTACAACAACTTACAAATTAAATGTTGGTGATTATTTTGTGCTGACATCAC
ATACAGTAATGCCATTAAGTGCACCTACACTAGTGCCACAAGAGCACTATGTTAGAATTACTGGCTTATA
CCCAACACTCAATATCTCAGATGAGTTTTCTAGCAATGTTGCAAATTATCAAAAGGTTGGTATGCAAAAG
TATTCTACACTCCAGGGACCACCTGGTACTGGTAAGAGTCATTTTGCTATTGGCCTAGCTCTCTACTACC
CTTCTGCTCGCATAGTGTATACAGCTTGCTCTCATGCCGCTGTTGATGCACTATGTGAGAAGGCATTAAA
ATATTTGCCTATAGATAAATGTAGTAGAATTATACCTGCACGTGCTCGTGTAGAGTGTTTTGATAAATTC
AAAGTGAATTCAACATTAGAACAGTATGTCTTTTGTACTGTAAATGCATTGCCTGAGACGACAGCAGATA
TAGTTGTCTTTGATGAAATTTCAATGGCCACAAATTATGATTTGAGTGTTGTCAATGCCAGATTACGTGC
TNAGCACTATGTGTACATTGGCGACCCTGCTCAATTACCTGCACCACGCACATTGCTAACTAAGGGCACA
CTAGAACCAGAATATTTCAATTCAGTGTGTAGACTTATGAAAACTATAGGTCCAGACATGTTCCTCGGAA
CTTGTCGGCGTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAGCTTAAAGC
ACATAAAGACAAATCAGCTCAATGCTTTAAAATGTTTTATAAGGGTGTTATCACGCATGATGTTTCATCT
GCAATTAACAGGCCACAAATAGGCGTGGTAAGAGAATTCCTTACACGTAACCCTGCTTGGAGAAAAGCTG
TCTTTATTTCACCTTATAATTCACAGAATGCTGTAGCCTCAAAGATTTTGGGACTACCAACTCAAACTGT
TGATTCATCACAGGGCTCAGAATATGACTATGTCATATTCACTCAAACCACTGAAACAGCTCACTCTTGT
AATGTAAACAGATTTAATGTTGCTATTACCAGAGCAAAAGTAGGCATACTTTGCATAATGTCTGATAGAG
ACCTTTATGACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTACAAGCTGA
AAATGTAACAGGACTCTTTAAAGATTGTAGTAAGGTAATCACTGGGTTACATCCTACACAGGCACCTACA
CACCTCAGTGTTGACACTAAATTCAAAACTGAAGGTTTATGTGTTGACATACCTGGCATACCTAAGGACA
TGACCTATAGAAGACTCATCTCTATGATGGGTTTTAAAATGAATTATCAAGTTAATGGTTACCCTAACAT
GTTTATCACCCGCGAAGAAGCTATAAGACATGTACGTGCATGGATTGGCTTCGATGTCGAGGGGTGTCAT
GCTACTAGAGAAGCTGTTGGTACCAATTTACCTTTACAGCTAGGTTTTTCTACAGGTGTTAACCTAGTTG
CTGTACCTACAGGTTATGTTGATACACCTAATAATACAGATTTTTCCAGAGTTAGTGCTAAACCACCGCC

TGGAGATCAATTTAAACACCTCATACCACTTATGTACAAAGGACTTCCTTGGAATGTAGTGCGTATAAAG
ATTGTACAAATGTTAAGTGACACACTTAAAAATCTCTCTGACAGAGTCGTATTTGTCTTATGGGCACATG
GCTTTGAGTTGACATCTATGAAGTATTTTGTGAAAATAGGACCTGAGCGCACCTGTTGTCTATGTGATAG
ACGTGCCACATGCTTTTCCACTGCTTCAGACACTTATGCCTGTTGGCATCATTCTATTGGATTTGATTAC
GTCTATAATCCGTTTATGATTGATGTTCAACAATGGGGTTTTACAGGTAACCTACAAAGCAACCATGATC
TGTATTGTCAAGTCCATGGTAATGCACATGTAGCTAGTTGTGATGCAATCATGACTAGGTGTCTAGCTGT
CCACGAGTGCTTTGTTAAGCGTGTTGACTGGACTATTGAATATCCTATAATTGGTGATGAACTGAAGATT
AATGCGGCTTGTAGAAAGGTTCAACACATGGTTGTTAAAGCTGCATTATTAGCAGACAAATTCCCAGTTC
TTCACGACATTGGTAACCCTAAAGCTATTAAGTGTGTACCTCAAGCTGATGTAGAATGGAAGTTCTATGA
TGCACAGCCTTGTAGTGACAAAGCTTATAAAATAGAAGAATTATTCTATTCTTATGCCACACATTCTGAC
AAATTCACAGATGGTGTATGCCTATTTTGGAATTGCAATGTCGATAGATATCCTGCTAATTCCATTGTTT
GTAGATTTGACACTAGAGTGCTATCTAACCTTAACTTGCCTGGTTGTGATGGTGGCAGTTTGTATGTAAA
TAAACATGCATTCCACACACCAGCTTTTGATAAAAGTGCTTTTGTTAATTTAAAACAATTACCATTTTTC
TATTACTCTGACAGTCCATGTGAGTCTCATGGAAAACAAGTAGTGTCAGATATAGATTATGTACCACTAA
AGTCTGCTACGTGTATAACACGTTGCAATTTAGGTGGTGCTGTCTGTAGACATCATGCTAATGAGTACAG
ATTGTATCTCGATGCTTATAACATGATGATCTCAGCTGGCTTTAGCTTGTGGGTTTACNNNNNNNNNGAT
ACTTATAACCTCTGGAACACTTTTACAAGACTTCAGAGTTTAGAAAATGTGGCTTTTAATGTTGTAAATA
AGGGACACTTTGATGGACAACAGGGTGAAGTACCAGTTTCTATCATTAATAACACTGTTTACACAAAAGT
TGATGGTGTTGATGTAGAATTGTTTGAAAATAAAACAACATTACCTGTTAATGTAGCATTTGAGCTTTGG
GCTAAGCGCAACATTAAACCAGTACCAGAGGTGAAAATACTCAATAATTTGGGTGTGGACATTGCTGCTA
ATACTGTGATCTGGGACTACAAAAGAGATGCTCCAGCACATATATCTACTATTGGTGTTTGTTCTATGAC
TGACATAGCCAAGAAACCAACTGAAACGATTTGTGCACCACTCACTGTCTTTTTTGATGGTAGAGTTGAT
GGTCAAGTAGACTTATTTAGAAATGCCCGTAATGGTGTTCTTATTACAGAAGGTAGTGTTAAAGGTTTAC
AACCATCTGTAGGTCCCAAACAAGCTAGTCTTAATGGAGTCACATTAATTGGAGAAGCCGTAAAAACACA
GTTCAATTATTATAAGAAAGTTGATGGTGTTGTCCAACAATTACCTGAAACTTACTTTACTCAGAGTAGA
AATTTACAAGAATTTAAACCCAGGAGTCAAATGGAAATTGATTTCTTAGAATTAGCTATGGATGAATTCA
TTGAACGGTATAAATTAGAAGGCTATGCCTTCGAACATATCGTTTATGGAGATTTTAGTCATAGTCAGTT
AGGTGGTTTACATCTACTGATTGGACTAGCTAAACGTTTTAAGGAATCACCTTTTGAATTAGAAGATTTT
ATTCCTATGGACAGTACAGTTAAAAACTATTTCATAACAGATGCGCAAACAGGTTCATCTAAGTGTGTGT
GTTCTGTTATTGATTTATTACTTGATGATTTTGTTGAAATAATAAAATCCCAAGATTTATCTGTAGTTTC
TAAGGTTGTCAAAGTGACTATTGACTATACAGAAATTTCATTTATGCTTTGGTGTAAAGATGGCCATGTA
GAAACATTTTACCCAAAATTACAATCTAGTCAAGCGTGGCAACCGGGTGTTGCTATGCCTAATCTTTACA
AAATGCAAAGAATGCTATTAGAAAAGTGTGACCTTCAAAATTATGGTGATAGTGCAACATTACCTAAAGG
CATAATGATGAATGTCGCAAAATATACTCAACTGTGTCAATATTTAAACACATTAACATTAGCTGTACCC
TATAATATGAGAGTTATACATTTTGGTGCTGGTTCTGATAAAGGAGTTGCACCAGGTACAGCTGTTTTAA

GACAGTGGTTGCCTACGGGTACGCTGCTTGTCGATTCAGATCTTAATGACTTTGTCTCTGATGCAGATTC
AACTTTGATTGGTGATTGTGCAACTGTACATACAGCTAATAAATGGGATCTCATTATTAGTGATATGTAC
GACCCTAAGACTAAAAATGTTACAAAAGAAAATGACTCTAAAGAGGGTTTTTTCACTTACATTTGTGGGT
TTATACAACAAAAGCTAGCTCTTGGAGGTTCCGTGGCTATAAAGATAACAGAACATTCTTGGAATGCTGA
TCTTTATAAGCTCATGGGACACTTCGCATGGTGGACAGCCTTTGTTACTAATGTGAATGCGTCATCATCT
GAAGCATTTTTAATTGGATGTAATTATCTTGGCAAACCACGCGAACAAATAGATGGTTATGTCATGCATG
CAAATTACATATTTTGGAGGAATACAAATCCAATTCAGTTGTCTTCCTATTCTTTATTTGACATGAGTAA
ATTTCCCCTTAAATTAAGGGGTACTGCTGTTATGTCTTTAAAAGAAGGTCAAATCAATGATATGATTTTA
TCTCTTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGTGATGTTCTTG
TTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATC
TTACAACCAGAACTCAATTACCCCCTACATACACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAA
AGTTTTCAGATCCTCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTCCAATGTTACTTGG
TTCCATGCTATACATGTCTCTGGGACCAATGTTATTAAGAGGTTTGATAACCCTGTCCTACCATTTAATG
ATGGTGTTTATTTTGCTTCCACTGAGAAGTCTAACATAATAAGAGGCTGGATTTTTGGTACTACTTTAGA
TTCGAAGACCCAGTCCCTACTTATTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAA
TTTTGTAATGATCCATTTTTGGGTGTTTATTANNNNNNAAACAACAAAAGTTGGATGGAAAGTGAGTTCA
GAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCAGCCTTTTCTTATGGACCTTGAAGG
AAAACAGGGTAATTTCAAAAATCTTAGGGAATTTGTGTTTAAGAATATTGATGGTTATTTTAAAATATAT
TCTAAGCACACGCCTATTAATTTAGTGCGTGATCTCCCTCAGGGTTTTTCGGCTTTAGAACCATTGGTAG
ATTTGCCAATAGGTATTAACATCACTAGGTTTCAAACTTTACTTGCTTTACATNNNNNNNNNNNNNNNN
NNNNAATTCTTCTTCAGGTTGGACAGCTGGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGACT
TTTCTATTAAAATATAATGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTCTCAG
AAACAAAGTGTACGTTGAAATCCTTCACTGTAGAAAAAGGAATCTATCAAACTTCTAACTTTAGAGTCCA
ACCAACAGAATCTATTGTTAGATTTCCTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCC
ACCAGATTTGCATCTGTTTATGCTTGGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCC
TATATAATTCCGCATCATTTTCCACTTTTAAGTGTTATGGAGTGTCTCCTACTAAATTAAATGATCTCTG
CTTTACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAA
ACTGGAAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAATT
CTAACAATCTTGATTCTAAGGTTGGTGGTAATTATAATTACCAGTATAGATTGTTTAGGAAGTCTAATCT
CAAACCTTTTGAGAGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGTAATGGTGTTGAA
GGTTTTAATTGTTACTCTCCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTTGGTTACCAACCAT
ACAGAGTAGTAGTACTTTCTTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTAC
TAATTTGGTTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAACAGGCACAGGTGTTCTTACTGAG
TCTAACAAAAAGTTTCTGCCTTTCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTG
ATCCACAGACACTTGAGATTCTTGACATTACACCATGTTCTTTTGGTGGTGTCAGTGTTATAACACCAGG

AACAAATACTTCTAACCAGGTTGCTGTTCTTTATCAGGGTGTTAACTGCACAGAAGTCCCTGTTGCTATT
CATGCAGATCAACTTACTCCTACTTGGCGTGTTTATTCTACAGGTTCTAATGTTTTTCAAACACGTGCAG
GCTGTTTAATAGGGGCTGAACATGTCAACAACTCATATGAGTGTGACATACCCATTGGTGCAGGTATATG
CGCTAGTTATCAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCATCATTGCC
TACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGCCATACCCACAAATT
TTACTATTAGTGTTACTACAGAAATTCTACCAGTGTCTATGACCAAGACATCAGTAGATTGTACAATGTA
CATTTGTGGTGATTCAACTGAATGCAGCAATCTTTTGTTGCAATATGGCAGTTTTTGTACACAATTAAAC
CGTGCTTTAACTGGAATAGCTGTTGAACAAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAAA
TTTACAAAACACCACCAATTAAAGATTTTGGTGGTTTTAATTTTTCACAAATATTACCAGATCCATCAAA
ACCAAGCAAGAGGTCATTTATTGAAGATCTACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATC
AAACAATATGGTGATTGCCTTGGTGATATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTTAACGGCC
TTAATGTTTTGCCACCTTTGCTCACAGATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGTAC
AATCACTTCTGGTTGGACCTTTGGTGCAGGTGCTGCATTACAAATACCATTTGCTATGCAAATGGCTTAT
AGGTTTAATGGTATTGGAGTTACACAGAATGTTCTCTATGAGAACCAAAAATTGATTGCCAACCAATTTA
ATAGTGCTATTGGCAAAATTCAAGACTCACTTTCTTCCACAGCAAGTGCACTTGGAAAACTTCAAGATGT
GGTCAACCAANNNNNNNNAGCTTTAAACACGCTTGTTAAACAACTTAGCTCCAATTTTGGTGCAATTTCA
AGTGTTTTAAATGATATCCTTTCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGATAGGTTGATCA
CAGGCAGACTTCAAAGTTTGCAGACATATGTGACTCAACAATTAATTAGAGCTGCAGAAATCAGAGCTTC
TGCTAATCTTGCTGCTACTAAAATGTCAGAGTGTGTACTTGGACAATCAAAAAGAGTTGATTTTTGTGGA
AAGGGCTATCATCTTATGTCCTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCATGTGACTTATG
TCCCTGCACAAGAAAAGAACTTCACAACTGCTCCTGCCATTTGTCATGATGGAAAAGCACACTTTCCTCG
TGAAGGTGTCTTTGTTTCAAATGGCACACACTGGTTTGTAACACAAAGGAATTTTTATGAACCACAAATC
ATTACTACAGACAACACATTTGTGTCTGGTAACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTT
ATGATCCTTTGCAACCTGAATTAGACTCATTCAAGGAGGAGTTAGATAAATATTTTAAGAATCATACATC
ACCAGATGTTGATTTAGGTGACATCTCTGGCATTAATGCTTCAGTTGTAAACATTCAAAAAGAAATTGAC
CGCCTCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCGATCTCCAAGAACTTGGAAAGTATGAGC
AGTATATAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTTGATTGCCATAGTAATGGTGAC
AATTATGCTTTGCTGTATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGATCCTGCTGC
AAATTTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAAACGAACTTAT
GGATTTGTTTATGAGAATCTTCACAATTGGAACTGTAACTTTGAAGCAAGGTGAAATCAAGGATGCTACT
CCTTCAGATTTTGTTCGCGCTACTGCAACGATACCGATACAAGCCTCACTCCCTTTCGGATGGCTTATTG
TTGGCGTTGCACTTCTTGCTGTTTTTCAGAGCGCTTCCAAAATCATAACCCTCAAAAAGAGATGGCAACT
AGCACTCTCCAAGGGTGTTCACTTTGTTTGCANCTTGCTGTTGTTGTTTGTAACAGTTTACTCACACCTT
TTGCTCGTTGCTGCTGGCCTTGAAGCCCCTTTTCTCTATCTTTATGCTTTAGTCTACTTCTTGCAGAGTA
TAAACTTTGTAAGAATAATAATGAGGCTTTGGCTTTGCTGGAAATGCCGTTCCAAAAACCCATTACTTTA

TGATGCCAACTATTTTCTTTGCTGGCATACTAATTGTTACGACTATTGTATACCTTACAATAGTGTAACT
TCTTCAATTGTCATTACTTCAGGTGATGGCACAACAAGTCCTATTTCTGAACATGACTACCAGATTGGTG
GTTATACTGAAAAATGGGAATCTGGAGTAAAAGACTGTGTTGTATTACACAGTTACTTCACTTCAGACTA
TTACCAGCTGTACTCAACTCAATTGAGTACAGACACTGGTGTTGAACATGTTACCTTCTTCATCTACAAT
AAAATTGTTGATGAGCCTGAAGAACATGTCCAAATTCACACAATCGACGGTTCATCCGGAGTTGTTAATC
CAGTAATGGAACCAATTTATGATGAACCGACGACGACTACTAGCGTGCCTTTGTAAGCACAAGCTGATGA
GTACGAACTTATGTACTCATTCGTTTCGGAAGAGACAGGTACGTTAATAGTTAATAGCGTACTTCTTTTT
CTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCCATCCTTACTGCGCTTCGATTGTGTGCGTACTGCT
GCAATATTGTTAACGTGAGTCTTGTAAAACCTTCTTTTTACGTTTACTCTCGTTTTAAAAATCTGAATTC
TTCTAGAGTTCCTGATCTTCTGGTCTAAACGAACTAAATATTATATTAGTTTTTCTGTTTGGAACTTTAA
TTTTAGCCATGGCAGATTCCAACGGTACTATTACCGTTGAAGAGCTTAAAAAGCTCCTTGAACAATGGAA
CCTAGTAATAGGTTTCCTATTCCTTACATGGATTTGTCTTCTACAATTTGCCTATGCCAACAGGAATAGG
TTTTTGTATATAATTAAGTTAATTTTCCTCTGGCTGTTATGGCCAGTAACTTTAGCTTGTTTTGTGCTTG
CTGCTGTTTACAGAATAAATTGGATCACCGGTGGAATTGCTATCGCAATGGCTTGTCTTGTAGGCTTGAT
GTGGCTCAGCTACTTCATTGCTTCTTTCAGACTGTTTGCGCGTACGCGTTCCATGTGGTCATTCAATCCA
GAAACTAACATTCTTCTCAACGTGCCACTTCATGGCACTATTCTGACCAGACCGCTTCTAGAAAGTGAAC
TCGTAATCGGAGCTGTGATCCTTCGTGGACATCTTCGTATTGCTGGACACCATCTAGGACGCTGTGACAT
CAAGGACCTGCCTAAAGAAATCACTGTTGCTACATCACGAACGCTTTCTTATTACAAATTGGGAGCTTCG
CAGCGTGTAGCAGGTGACTCAGGTTTTGCTGCATACAGTCGCTACAGGATTGGCAACTATAAATTAAACA
CAGACCATTCCAGTAGCAGTGACAATATTGCTTNNNNNNNNNNNNNNGTGACAACAGATGTTTCATCTCG
TTGACTTTCAGGTTACTATAGCAGAGATATTACTAATTATTATGAGGACTTTTAAAGTTTCCATTTGGAA
TCTTGATTACATCATAAACCTCATAATTAAAAATTTATCTAAGTCACTAACTGAGAATAAATATTCTCAA
TTAGATGAAGAGCAACCAATGGAGATTGATTAAACGAACATGAAAATTATTCTTTTCTTGGCACTGATAA
CACTCGCTACTTGTGAGCTTTATCACTACCAAGAGTGTGTTAGAGGTACAACAGTACTTTTAAAAGAACC
TTGCTCTTCTGGAACATACGAGGGCAATTCACCATTTCATCCTCTAGCTGATAACAAATTTGCACTGACT
TGCTTTAGCACTCAATTTGCTTTTGCTTGTCCTGACGGCGTAAAACACGTCTATCAGTTACGTGCCAGAT
CAGTTTCACCTAAACTTTTCATCAGACAAGAGGAAGTTCAAGAACTTTACTCTCCAATTTTTCTTATTGT
TGCGGCAATAGTGTTTATAACACTTTGCTTCACACTCAAAAGAAAGACAGAATGATTGAACTTTCATTAA
TTGACTTCTATTTGTGCTTTTTAGCCTTTCTGCTATTCCTTGTTNTAATTATGCTTATTATCTTTTGGTT
CTCACTTGAACTGCAAGATCATAATGAAACTTGTCACGCCTAAACGAACATGAAATTTCTTGTTTTCTTA
GGAATCATCACAACTGTAGCTGCATTTCACCAAGAATGTAGTTTACAGTCATGTACTCAACATCAACCAT
ATGTAGTTGATGACCCGTGTCCTATTCACTTCTATTCTAAATGGTATATTAGAGTAGGAGCTAGAAAATC
AGCACCTTTAATTGAATTGTGCGTGGATGAGGCTGGTTCTAAATCACCCATTCAGTACATCGATATCGGT
AATTATACAGTTTCCTGTTTACCTTTTACAATTAATTGCCAGGAACCTAAATTGGGTAGTCTTGTAGTGC
GTTGTTCGTTCTATGAAGACTTTTTAGAGTATCATGACGTTCGTGTTGTTTTAGATTTTATCTAAACGAA

CAAACTTAAATGTCTGATAATGGACCCCAAAATCAGCGAAATGCACTCCGCATTACGTTTGGTGGACCCT
CAGATTCAACTGGCAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAACAACGTCGGCCCCAAGG
TTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTTAAATTCCCT
CGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTCCAGATGACCAAATTGGCTACTACCGAAGAGCTA
CCAGACGAATTCGTGGTGGTGACGGTAAAATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGG
AACTGGGCCAGAAGCTGGACTTCCCTATGGTGCTAACAAAGACGGCATCATATGGGTTGCAACTGAGGGA
GCCTTGAATACACCAAAAGATCACATTGGCACCCGCAATCCTGCTAACAATGCTGCAATCGTGCTACAAC
TTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCGGCAGTCAAGCCTCTTC
TCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATTCAACTCCAGGCAGCAGTAANCGAACTTCTCCT
GCTAGAATGGCTGGCAATTGCGGTGATGCTGCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTG
AGAGCAAAATGTCTGGTAAAGGCCAACAACAACAAGGCCAAACTGTCACTAAGAAATCTGCTGCTGAGGC
TTCTAAGAAGCCTCGGCAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGT
GGTCCAGAACAAACCCAAGGAAATTTTGGGGACCAGGAACTAATCAGACAAGGAACTGATTACAAACATT
GGCCGCAAATTGCACAATTTGCCCCCAGCGCTTCAGCGTTCTTCGGAATGTCGCGCATTGGCATGGAAGT
CACACCTTCGGGAACGTGGTTGACCTACACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTTTAAA
GATCAAGTCATTTTGCTGAATAAGCATATTGACGCATACAAAACATTCCCACCAAYAGAGCCTAAAAAGG
ACAAAAAGAAGAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACTGTGACTCT
TCTTCCTGCTGCAGATTTGGATGATTTCTCCAAACAATTGCAACAATCCATGAGCAGTGCTGACTCAACT
CAGGCCTAAACTCATGCAGACCACACAAGGCAGATGGGCTATATAAACGTTTTCGCTTTTCCGTTTACGA
TATATAGTCTACTCTTGTGCAGAATGAATTCTCGTAACTACATAGCACAAGTAGATGTAGTTAACTTTAA
TCTCACATAGCAATCTTTAATCAGTGTGTAACATTAGGGAGGACTTGAAAGAGCCACCACATTTTCACCG
AGGCCACGCGGAGTACGATCGAGTGTACAGTGAACAATGCTAGGGAGAGCTGCCTATATGGAAGAGCCCT
AATGTGTAAAATTAATTTTAGTAGTGCTATCCCCATGTGATTTTAATAGCTTCTTAGGAGAATGAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAA

3. mRNA Sequence

mRNA Sequence (Obtained from Python sequence)

CAGUAUAAUUAAUAACUAAUUACUGUCGUUGACAGGACACGAGUAACUCGUCUAUCUUC
UGCAGGCUGCUUACGGUUUCGUCCGUGUUGCAGCCGAUCAUCAGCACAUCUAGGUUUUGU
CCGGGUGUGACCGAAAGGUAAGAUGGAGAGCCUUGUCCCUGGUUUCAACGAGAAAACAC
ACGUCCAACUCAGUUUGCCUGUUUUACAGGUUCGCGACGUGCUCGUACGUGGCUUUGGAG
ACUCCGUGGAGGAGGUCUUAUCAGAGGCACGUCAACAUCUUAAAGAUGGCACUUGUGGC
UUAGUAGAAGUUGAAAAAGGCGUUUUGCCUCAACUUGAACAGCCCUAUGUGUUCAUCAA
ACGUUCGGAUGCUCGAACUGCACCUCAUGGUCAUGUUAUGGUUGAGCUGGUAGCAGAAC
UCGAAGGCAUUCAGUACGGUCGUAGUGGUGAGACACUUGGUGUCCUUGUCCCUCAUGUG
GGCGAAAUACCAGUGGCUUACCGCAAGGUUCUUCUUCGUAAGAACGGUAAUAAAGGAGC

UGGUGGCCAUAGUUACGGCGCCGAUCUAAAGUCAUUUGACUUAGGCGACGAGCUUGGCA
CUGAUCCUUAUGAAGAUUUUCAAGAAAACUGGAACACUAAACAUAGCAGUGGUGUUACC
CGUGAACUCAUGCGUGAGCUUAACGGAGGGGCAUACACUCGCUAUGUCGAUAACAACUUC
UGUGGCCCUGAUGGCUACCCUCUUGAGUGCAUUAAAGACCUUCUAGCACGUGCUGGUAAA
GCUUCAUGCACUUUGUCCGAACAACUGGACUUUAUUGACACUAAGAGGGGUGUAUACUG
CUGCCGUGAACAUGAGCAUGAAAUUGCUUGGUACACGGAACGUUCUGAAAAGAGCUAUG
AAUUGCAGACACCUUUUGAAAUUAAAUUGGCAAAGAAAUUUGACACCUUCAAUGGGGAA
UGUCCAAAUUUUGUAUUUCCCUUAAAUUCCAUAAUCAAGACUAUUCAACCAAGGGUUGA
AAAGAAAAAGCUUGAUGGCUUUAUGGGUAGAAUUCGAUCUGUCUAUCCAGUUGCGUCAC
CAAAUGAAUGCAACCAAAUGUGCCUUUCAACUCUCAUGAAGUGUGAUCAUUGUGGUGAA
ACUUCAUGGCAGACGGGCGAUUUUGUUAAAGCCACUUGCGAAUUUUGUGGCACUGAGAA
UUUGACUAAAGAAGGUGCCACUACUUGUGGUUACUUACCCCAAAAUGCUGUUGUUAAAA
UUUAUUGUCCAGCAUGUCACAAUUCAGAAGUAGGACCUGAGCAUAGUCUUGCCGAAUAC
CAUAAUGAAUCUGGCUUGAAAACCAUUCUUCGUAAGGGUGGUCGCACUAUUGCCUUUGG
AGGCUGUGUGUUCUCUUAUGUUGGUUGCCAUAACAAGUGUGCCUAUUGGGUUCCACGUG
CUAGCGCUAACAUAGGUUGUAACCAUACAGGUGUUGUUGGAGAAGGUUCCGAAGGUCUU
AAUGACNNNNNNNNUGAAAUACUCCAAAAAGAGAAAGUCAACAUCAAUAUUGUUGGUGA
CUUUAAACUUAAUGAAGAGAUCGCCAUUAUUUUGGCAUCUUUUUCUGCUUCCACAAGUG
CUUUUGUGGAAACUGUGAAAGGUUUGGAUUAUAAAGCAUUCAAACAAAUUGUUGAAUCC
UGUGGUAAUUUUAAAGUUACAAAAGGAAAAGCUAAAAAAGGUGCCUGGAAUAUUGGUGA
ACAGAAAUCAAUACUGAGUCCUCUUUAUGCAUUUGCAUCAGAGGCUGCUCGUGUUGUAC
GAUCAAUUUUCUCCCGCACUCUUGAAACUGCUCAAAAUUCUGUGCGUGUUUUACAGAAG
GCCGCUAUAACAAUACUAGAUGGAAUUUCACAGUAUUCACUGAGACUCAUUGAUGCUAU
GAUGUUCACAUCUGAUUUGGCUACUAACAAUCUAGUUGUAAUGGCCUACAUUACAGGUG
GUGUUGUUCAGUUGACUUCGCAGUGGCUAACUAACAUCUUUGGCACUGUUUAUGAAAAA
CUCAAACCCGUCCUUGAUUGGCUUGAAGAGAAGUUUAAGGAAGGUGUAGAGUUUCUUAG
AGACGGUUGGGAAAUUGUUAAAUUUAUCUCAACCUGUGCUUGUGAAAUUGUCGGUGGAC
AAAUUGUCACCUGUGCAAAGGAAAUUAAGGAGAGUGUUCAGACAUUCUUUAAGCUUGUA
AAUAAAUUUUUGGCUUUGUGUGCUGACUCUAUCAUUAUUGGUGGAGCUAAACUUAAAGC
CUUGAAUUUAGGUGAAACAUUUGUCACGCACUCAAAGGGAUUGUACAGAAAGUGUGUUA
AAUCCAGAGAAGAAACUGGCCUACUCAUGCCUCUAAAAGCCCCAAAAGAAAUUAUCUUCU
UAGAGGGAGAAACACUUCCCACAGAAGUGUUAACAGAGGAAGUUGUCUUGAAAACUGGU
GAUUUACAACCAUUAGAACAACCUACUAGUGAAGCUGUUGAAGCUCCAUUGGUUGGUAC
ACCAGUUUGUAUUAACGGGCUUAUGUUGCUCGAAAUCAAAGACACAGAAAAGUACUGUG
CCCUUGCACCUAAUAUGAUGGUAACAAACAAUACCUUCACACUCAAAGGCGGUGCACCAA
CAAAGGUUACUUUUGGUGAUGACACUGUGAUAGAAGUGCAAGGUUACAAGAGUGUGAAU
AUCACUUUUGAACUUGAUGAAAGGAUUGAUAAAGUACUUAAUGAGAAGUGCUCUGCCUA
UACAGUUGAACUCGGUACAGAAGUAAAUGAGUUCGCCUGUGUUGUGGCAGAUGCUGUCA
UAAAAACUUUGCAACCAGUAUCUGAAUUACUUACACCACUGGGCAUUGAUUUAGAUGAG
UGGAGUAUGGCUACAUACUACUUAUUUGAUGAGUCUGGUGAGUUUAAAUUGGCUUCACA
UAUGUAUUGUUCUUUUUACCCUCCAGAUGAGGAUGAAGAAGAAGGUGAUUGUGAAGAAG
AAGAGUUUGAGCCAUCAACUCAAUAUGAGUAUGGUACUGAAGAUGAUUACCAAGGUAAA
CCUUUGGAAUUUGGUGCCACUUCUGCUGCUCUUCAACCUGAAGAAGAGCAAGAAGAAGA
UUGGUUAGAUGAUGAUAGUCAACAAACUGUUGGUCAACAAGACGGCAGUGAGGACAAUC
AGACAACUACUAUUCAAACAAUUGUUGAGGUUCAACCUCAAUUAGAGAUGGAACUUACA
CCAGUUGUUCAGACUAUUGAAGUGAAUAGUUUUAGUGGUUAUUUAAAACUUACUGACAA
UGUAUACAUUAAAAAUGCAGACAUUGUGGAAGAAGCUAAAAAGGUAAAACCAACAGUGG

UUGUUAAUGCAGCCAAUGUUUACCUUAAACAUGGAGGAGGUGUUGCAGAGCCUUAAAUN
NNGCUACUAACAAUGCCAUGCAAGUUGAAUCUGAUGAUUACAUAGCUACUAAUGGACCA
CUUAAAGUGGGUGGUAGUUGUGUUUUAAGCGGACACAAUCUUGCUAAACACUGUCUUCA
UGUUGUCGGCCCAAAUGUUAACAAAGGUGAAGACAUUCAACUUCUUAAGAGUGCUUAUG
AAAAUUUUAAUCAGCACGAAGUUCUACUUGCACCAUUAUUAUCAGCUGGUAUUUUUGGU
GCUGACCCUAUACAUUCUUUAAGAGUUUGUGUAGAUACUGUUCGCACAAAUGUCUACUU
AGCUGUCUUUGAUAAAAAUCUCUAUGACAAACUUGUUUCAAGCUUUUUGGAAAUGAAGA
GUGAAAAGCAAGUUGAACAAAAGAUCGCUGAGAUUCCUAAAGAGGAAGUUAAGCCAUUU
AUAACUGAAAGUAAACCUUCAGUUGAACAGAGAAAACAAGAUGAUAAGAAAAUCAAAGC
UUGUGUUGAAGAAGUUACAACAACUCUGGAAGAAAUUAAGUUCCUCACAGAAAACUUGU
UACUUUAUAUUGACAUUAAUGGCAAUCUUCAUCCAGAUUCUGCCACUCUUGUUAGUGAC
AUUGACAUCACUUUCUUAAAGAAAGAUGCUCCAUAUAUAGUGGGUGAUGUUGUUCAAGA
GGGUGUUUUAACUGCUGUGGUUAUACCUACUAAAAAGGCUGGUGGCACUACUGAAAUGC
UAGCGAAAGCUUUGAGAAAAGUGCCAACAGACAAUUAUAUAACCACUUACCCGGGUCAG
GGUUUAAAUGGUUACACUGUAGAGGAGGCAAAGACAGUGCUUAAAAAGUGUAAAAGUGC
CUUUUACAUUCUACCAUCUAUUAUCUCUAAUGAGAAGCAAGAAAUUCUUGGAACUGUUU
CUUGGAAUUUGCGAGAAAUGCUUGCACAUGCAGAAGAAACACGCAAAUUAAUGCCUGUC
UGUGUGGAAACUAAAGCCAUAGUUUCAACUAUACAGCGUAAAUAUAAGGGUAUUAAAAU
ACAAGAGGGUGUGGUUGAUUAUGGUGCUAGAUUUUACUUUUACACCAGUAAAACAACUG
UAGCGUCACUUAUCAACACACUUAACGAUCUAAAUGAAACUCUUGUUACAAUGCCACUUG
GCUAUGUAACACAUGGCUUAAAUUUGGAAGAAGCUGCUCGGUAUAUGAGAUCUCUCAAA
GUGCCAGCUACAGUUUCUGUUUCUUCACCUGAUGCUGUUACAGCGUAUAAUGGUUAUCU
UACUUCUUCUUCUAAAACACCUGAAGAACAUUUUAUUGAAACCAUCUCACUUGCUGGUUC
CUAUAAAGAUUGGUCCUAUUCUGGACAAUCUACACAACUAGGUAUAGAAUUUCUUAAGA
GAGGUGAUAAAAGUGUAUAUUACACUAGUAAUCCUACCACAUUCCACCUAGAUGGUGAA
GUUAUCACCUUUGACAAUCUUAAGACACUUCUUUCUUUGAGAGAAGUGAGGACUAUUAA
GGUGNNNNNNACAGUAGACAACAUUAACCUCCACACGCAAGUUGUGGACAUGUCAAUGA
CAUAUGGACAACAGUUUGGUCCAACUUAUUUGGAUGGAGCUGAUGUUACUAAAAUAAAA
CCUCAUAAUUCACAUGAAGGUAAAACAUUUUAUGUUUUACCUAAUGAUGACACUCUACG
UGUUGAGGCUUUUGAGUACUACCACACAACUGAUCCUAGUUUUCUGGGUAGGUACAUGU
CAGCAUUAAAUCACACUAAAAAGUGGAAAUACCCACAAGUUAAUGGUUUAACUUCUAUU
AAAUGGGCAGAUAACAACUGUUAUCUUGCCACUGCAUUGUUAACACUCCAACAAAUAGA
GUUGAAGUUUAAUCCACCUGCUCUACAAGAUGCUUAUUACAGAGCAAGGGCUGGUGAAG
CUGCUAACUUUUGUGCACUUAUCUUAGCCUACUGUAAUAAGACAGUAGGUGAGUUAGGU
GAUGUUAGAGAAACAAUGAGUUACUUGUUUCAACAUGCCAAUUUAGAUUCUUGCAAAAG
AGUCUUGAACGUGGUGUGUAAAACUUGUGGACAACAGCAGACAACCCUUAAGGGUGUAG
AAGCUGUUAUGUACAUGGGCACACUUUCUUAUGAACAAUUUAAGAAAGGUGUUCAGAUA
CCUUGUACGUGUGGUAAACAAGCUACAAAAUAUCUAGUACAACAGGAGUCACCUUUUGU
UAUGAUGUCAGCACCACCUGCUCAGUAUGAACUUAAGCAUGGUACAUUUACUUGUGCUA
GUGAGUACACUGGUAAUUACCAGUGUGGUCACUAUAAACAUAUAACUUCUAAAGAAACU
UUGUAUUGCAUAGACGGUGCUUUACUUACAAAGUCCUCAGAAUACAAAGGUCCUAUUAC
GGAUGUUUUCUACAAAGAAAACAGUUACACAACAACCAUAAAACCAGUUACUUAUAAAU
UGGAUGGUGUUGUUUGUACAGAAAUUGACCCUAAGUUGGACAAUUAUUAUAAGAAAGAC
AAUUCUUAUUUCACAGAGCAACCAAUUGAUCUUGUACCAAACCAACCAUAUCCAAACGCA
AGCUUCGAUAAUUUUAAGUUUGUAUGUGAUAAUAUCAAAUUUGCUGAUGAUUUAAACCA
GUUAACUGGUUAUAAGAAACCUGCUUCAAGAGAGCUUAAAGUUACAUUUUUCCCUGACU
UAAAUGGUGAUGUGGUGGCUAUUGAUUAUAAACACUACACACCCUCUUUUAAGAAAGGA

GCUAAAUUGUUACAUAAACCUAUUGUUUGGCAUGUUAACAAUGCAACUAAUAAAGCCAC
GUAUAAACCAAAUACCUGGUGUAUACGUUGUCUUUGGAGCACAAAACCAGUUGAAACAU
CAAAUUCGUUUGAUGUACUGAAGUCAGAGGACGCGCAGGGAAUGGAUAAUCUUGCCUGC
GAAGAUCUAAAACCAGUCUCUGAAGAAGUAGUGGAAAAUCCUACCAUACAGAAAGACGU
UCUUGAGUGUAAUGUGAAAACUACCGAAGUUGUAGGAGACAUUAUACUUAAACCAGCAA
AUAAUAGUUUAAAAAUUACAGAAGAGGUUGGCCACACAGAUCUAAUGGCUGCUUAUGUA
GACAAUUCUAGUCUUACUAUUAAGAAACCUAAUGAAUUAUCUAGAGUAUUAGGUUUGAA
AACCCUUGCUACUCAUGGUUUAGCUGCUGUUAAUAGUGUCCCUUGGGAUACUAUAGCUA
AUUAUGCUAAGCCUUUUCUUAACAAAGUUGUUAGUACAACUACUAACAUAGUUACACGG
UGUUUAAACCGUGUUUGUACUAAUUAUAUGCCUUAUUUCUUUACUUUAUUGCUACAAUU
GUGUACUUUUACUAGAAGUACAAAUUCUAGAAUUAAAGCAUCUAUGCCGACUACUAUAG
CAAAGAAUACUGUUAAGAGUGUCGGUAAAUUUUGUCUAGAGGCUUCAUUUAAUUAUUUG
AAGUCACCUAAUUUUUCUAAACUGAUAAAUAUUAUAAUUUGGUUUUUACUAUUAAGUGU
UUGCCUAGGUUCUUUAAUCUACUCAACCGCUGCUUUAGGUGUUUUAAUGUCUAAUUUAG
GCAUGCCUUCUUACUGUACUGGUUACAGAGAAGGCUAUUUGAACUCUACUAAUGUCACU
AUUGCAACCUACUGUACUGGUUCUAUAUCUUGUAGUGUUUGUCUUAGUGGUUUAGAUUC
UUUAGACACCUAUCCUUCUUUAGAAACUAUACAAAUUACCAUUUCAUCUUUUAAAUGGG
AUUUAACUGCUUUUGGCUUAGUUGCAGAGUGGUUUUUGGCAUAUAUUCUUUUCACUAGG
UUUUUCUAUGUACUUGGAUUGGCUGCAAUCAUGCAAUUGUUUUUCAGCUAUUUUGCAGU
ACAUUUUAUUAGUAAUUCUUGGCUUAUGUGGUUAAUAAUUAAUCUUGUACAAAUGGCCC
CGAUUUCAGCUAUGGUUAGAAUGUACAUCUUCGUUGCAUCAUUUUAUUAUGUAUGGAAA
AGUUAUGUGCAUGUUGUAGACGGUUGUAAUUCAUCAACUUGUAUGAUGUGUUACAAACG
UAAUAGAGCAACAAGAGUCGAAUGUACAACUAUUGUUAAUGGUGUUAGAAGGUCCUUUU
AUGUCUAUGCUAAUGGAGGUAAAGGCUUUUGCAAACUACACAAUUGGAAUUGUGUUAAU
UGUGAUACAUUCUGUGCUGGUAGUACAUUUAUUAGUGAUGAAGUUGCGAGAGACUUGUC
ACUACAGUUUAAAAGACCAAUAAAUCCUACUGACCAGUCUUCUUACAUCGUUGAUAGUG
UUACAGUGAAGAAUGGUUCCAUCCAUCUUUACUUUGAUAAAGCUGGUCAAAAGACUUAU
GAAAGACAUUCUCUCUCUCAUUUUGUUAACUUAGACAACCUGAGAGCUAAUAACACUAA
AGGUUCAUUGCCUAUUAAUGUUAUAGUUUUUGAUGGUAAAUCAAAAUGUGAAGAAUCAU
CUGCAAAAUCAGCGUCUGUUUACUACAGUCAGCUUAUGUGUCAACCUAUACUGUUNNNN
NNNNAGGCAUUAGUGUCUGAUGUUGGUGAUAGUGCGGAAGUUGCAGUUAAAAUGUUUGA
UGCUUACGUUAAUACGUUUUCAUCAACUUUUAACGUACCAAUGGAAAAACUCAAAACAC
UAGUUGCAACUGCAGAAGCUGAACUUGCAAAGAAUGUGUCCUUAGACAAUGUCUUAUCU
ACUUUUAUUUCAGCAGCUCGGCAAGGGUUUGUUGAUUCAGAUGUAGAAACUAAAGAUGU
UGUUGAAUGUCUUAAAUUGUCACAUCAAUCUGACAUAGAAGUUACUGGCGAUAGUUGUA
AUAACUAUAUGCUCACCUAUAACAAAGUUGAAAACAUGACACCCCGUGACCUUGGUGCUU
GUAUUGACUGUAGUGCGCGUCAUAUUAAUGCGCAGGUAGCAAAAAGUCACAACAUUGCU
UUGAUAUGGAACGUUAAAGAUUUCAUGUCAUUGUCUGAACAACUACGAAAACAAAUACG
UAGUGCUGCUAAAAAGAAUAACUUACCUUUUAAGUUGACAUGUGCAACUACUAGACAAG
UUGUUAAUGUUGUAACAACAAAGAUAGCACUUAAGGGUGGUAAAAUUGUUAAUAAUUGG
UUGAAGCAGUUAAUUAAAGUUACACUUGUGUUCCUUUUUGUUGCUGCUAUUUUCUAUUU
AAUAACACCUGUUCAUGUCAUGUCUAAACAUACUGACUUUUCAAGUGAAAUCAUAGGAU
ACAAGGCUAUUGAUGGUGGUGUCACUCGUGACAUAGCAUCUACAGAUACUUGUUUUGCU
AACAAACAUGCUGAUUUUGACACAUGGUUUAGCCAGCGUGGUGGUAGUUAUACUAAUGA
CAAAGCUUGCCCAUUGAUUGCUGCAGUCAUAACAAGAGAAGUGGGUUUUGUCGUGCCUG
GUUUGCCUGGCACGAUAUUACGCACAACUAAUGGUGACUUUUUGCAUUUCUUACCUAGA
GUUUUUAGUGCAGUUGGUAACAUCUGUUACACACCAUCAAAACUUAUAGAGUACACUGA

CUUUGCAACAUCAGCUUGUGUUUUGGCUGCUGAAUGUACAAUUUUUAAAGAUGCUUCUG
GUAAGCCAGUACCAUAUUGUUAUGAUACCAAUGUACUAGAAGGUUCUGUUGCUUAUGAA
AGUUUACGCCCUGACACACGUUAUGUGCUCAUGGAUGGCUCUAUUAUUCAAUUUCCUAAC
ACCUACCUUGAAGGUUCUGUUAGAGUGGUAACAACUUUUGAUUCUGAGUACUGUAGGCA
CGGCACUUGUGAAAGAUCAGAAGCUGGUGUUUGUGUAUCUACUAGUGGUAGAUGGGUAC
UUAACAAUGAUUAUUACAGAUCUUUACCAGGAGUUUUCUGUGGUGUAGAUGCUGUAAAU
UUACUUACUAAUAUGUUUACACCACUAAUUCAACCUAUUGGUGCUUUGGACAUAUCAGC
AUCUAUAGUAGCUGGUGGUAUUGUAGCUAUCGUAGUAACAUGCCUUGCCUACUAUUUUA
UGAGGUUUAGAAGAGCUUUUGGUGAAUACAGUCAUGUAGUUGCCUUUAAUACUUUACUA
UUCCUUAUGUCAUUCACUGUACUCUGUUUAACACCAGUUUACUCAUUCUUACCUGGUGUU
UAUUCUGUUAUUUACUUGUACUUGACAUUUUAUCUUACUAAUGAUGUUUCUUUUUUAGC
ACAUAUUCAGUGGAUGGUUAUGUUCACACCUUUAGUACCUUUCUGGAUAACAAUUGCUU
AUAUCAUUUGUAUUUCCACAAAGCAUUUCUAUUGGUUCUUUAGUAAUUACCUAAAGAGA
CGUGUAGUCUUUAAUGGUGUUUCCUUUAGUACUUUUGAAGAAGCUGCGCUGUGCACCUU
UUUGUUNNNNAAAGAAAUGUAUCUAAAGUUGCGUAGUGAUGUGUUAUUACCUCCUACGC
AAUAUAAUAGAUACUUAGCUCUUUAUAAUAAGUACAAGUAUUUUAGUGGAGCAAUGGAU
ACAACUAGCUACAGAGAAGCUGCUUGUUGUCAUCUCGCAAAGGCUCUCAAUGACUUCAGU
AACUCAGGUUCUGAUGUUCUUUACCAACCACCACAAAUCUCUAUCACCUCAGCUGUUUUG
CAGAGUGGUUUUAGAAAAAUGGCAUUCCCAUCUGGUAAAGUUGAGAGUUGUAUGGUACA
AGUAACUUGUGGUACAACUACACUUAACGGUCUUUGGCUUGAUGACGUAGUUUACUGUC
CAAGACAUGUGAUCUGCACCUCUGAAGACAUGCUUAACCCUAAUUAUGAAGAUUUACUC
AUUCGUAAGUCUAAUCAUAAUUUCUUGGUACAGGCUGGUAAUGUUCAACUCAGGGUUAU
UGGACAUUCUAUGCAAAAUUGUGUACUUAAGCUUAAGGUUGAUACAGCCAAUCCUAAGA
CACCUAAGUAUAAGUUUGUUCGCAUUCAACCAGGACAGACUUUUUCAGUGUUAGCUUGU
UACAAUGGUUCACCAUCUGGUGUUUACCAAUGUGCUAUGAGGCCCAAUUUCACUAUUAA
GGGUUCAUUCCUUAAUGGUUCAUGUGGUAGUGUUGGUUUUAACAUAGAUUAUGACUGUG
UCUCUUUUUGUUACAUGCACCAUAUGGAAUUACCAACUGGAGUUCAUGCUGGCACAGAC
UUAGAAGGUAACUUUUAUGGACCUUUUGUUGACAGGCAAACAGCACAAGCAGCUGGUAC
GGACACAACUAUUACAGUUAAUGUUUUAGCUUGGUUGUACGCUGCUGUUAUAAAUGGAG
ACAGGUGGUUUCUCAAUCGAUUUACCACAACUCUUAAUGACUUUAACCUUGUGGCUAUG
AAGUACAAUUAUGAACCUCUAACACAAGACCAUGUUGACAUACUAGGACCUCUUUCUGCU
CAAACUGGAAUUGCCGUUUUAGAUAUGUGUGCUUCAUUAAAAGAAUUACUGCAAAAUGG
UAUGAAUGGACGUACCAUAUUGGGUAGUGCUUUAUUAGAAGAUGAAUUUACACCUUUUG
AUGUUGUUAGACAAUGCUCAGGUGUUACUUUCCAAAGUGCAGUGAAAAGAACAAUCAAG
GGUACACACCACUGGUUGUUACUCACAAUUUUGACUUCNNNNNUAGUUUUAGUCCAGAG
UACUCAAUGGUCUUUGUUCUUUUUUUUGUAUGAAAAUGCCUUUUUACCUUUUGCUAUGG
GUAUUAUUGCUAUGUCUGCUUUUGCAAUGAUGUUUGUCAAACAUAAGCAUGCAUUUCUC
UGUUUGUUUUUGUUACCUUCUCUUGCCACUGUAGCUUAUUUUAAUAUGGUCUAUAUGCC
UGCUAGUUGGGUGAUGCGUAUUAUGACAUGGUUGGAUAUGGUUGAUACUAGUUUGANNA
GUUUGAAGCUAAAAGACUGUGUUAUGUAUGCAUCAGCUGUAGUGUUACUAAUCCUUAUG
ACAGCAAGAACUGUGUAUGAUGAUGGUGCUAGGAGAGUGUGGACACUUAUGAAUGUCUU
GACACUCGUUUAUAAAGUUUAUUAUGGUAAUGCUUUAGAUCAAGCCAUUUCCAUGUGGG
CUCUUAUAAUCUCUGUUACUUCUAACUACUCAGGUGUAGUUACAACUGUCAUGUUUUUG
GCCAGAGGUAUUGUUUUUAUGUGUGUUGAGUAUUGCCCUAUUUUCUUCAUAACUGGUAA
UACACUUCAGUGUAUAAUGCUAGUUUAUUGUUUCUUAGGCUAUUUUUGUACUUGUUACU
UUGGCCUCUUUUGUUUACUCAACCGCUACUUUAGACUGACUCUUGGUGUUUAUGAUUAC
UUAGUUUCUACACAGGAGUUUAGAUAUAUGAAUUCACAGGGACUACUCCCACCCAAGAA

UAGCAUAGAUGCCUUCAAACUCAACAUUAAAUUGUUGGGUGUUGGUGGCAAACCUUGUA
UCAAAGUAGCCACUGUACAGUCUAAAAUGUCAGAUGUAAAGUGCACAUCAGUAGUCUUA
CUCUCAGUUUUGCAACAACUCAGAGUAGAAUCAUCAUCUAAAUUGUGGGCUCAAUGUGU
CCAGUUACACAAUGACAUUCUCUUAGCUAAAGAUACUACUGAAGCCUUUGAAAAAAUGG
UUUCACUACUUUCUGUUUUGCUUUCCAUGCAGGGUGCUGUAGACAUAAACAAGCUUUGU
GAAGAAAUGCUGGACAACAGGGCAACCUUACAAGCUAUAGCCUCAGAGUUUAGUUCCCU
UCCAUCAUAUGCAGCUUUUGCUACUGCUCAAGAAGCUUAUGAGCAGGCUGUUGCUAAUG
GUGAUUCUGAAGUUGUUCUUAAAAAGUUGAAGAAGUCUUUGAAUGUGGCUAAAUCUGAA
UUUGACCGUGAUGCAGCCAUGCAACGUAAGUUGGAAAAGAUGGCUGAUCAAGCUAUGAC
CCAAAUGUAUAAACAGGCUAGAUCUGAGGACAAGAGGGCAAAAGUUACUAGUGCUAUGC
AGACAAUGCUUUUCACUAUGCUUAGAAAGUUGGAUAAUGAUGCACUCAACAACAUUAUC
AACAAUGCAAGAGAUGGUUGUGUUCCCUUGAACAUAAUACCNNNNACAACAGCAGCCAA
ACUAAUGGUUGUCAUACCAGACUAUAACACAUAUAAAAAUACGUGUGAUGGUACAACAU
UUACUUAUGCAUCAGCAUUGUGGGAAAUCCAACAGGUUGUAGAUGCAGAUAGUAAAAUU
GUUCAACUUAGUGAAAUUAGUAUGGACAAUUCACCUAAUUUAGCAUGGCCUCUUAUUGU
AACAGCUUUAAGGGCCAAUUCUGCUGUCAAAUUACAGAAUAAUGAGCUUAGUCCUGUUG
CACUACGACAGAUGUCUUGUGCUGCCGGUACUACACAAACUGCUUGCACUGAUGACAAUG
CGUUAGCUUACUACAACACAACAAAGGGAGGUAGGUUUGUACUUGCACUGUUAUCCGAU
UUACAGGAUUUGAAAUGGGCUAGAUUCCCUAAGAGUGAUGGAACUGGUACUAUCUAUAC
AGAACUGGAACCACCUUGUAGGUUUGUUACAGACACACCUAAAGGUCCUAAAGUGAAGU
AUUUAUACUUUAUUAAAGGAUUAAACAACCUAAAUAGAGGUAUGGUACUUGGUAGUUUA
GCUGCCACAGUACGUCUACAAGCUGGUAAUGCAACAGAAGUGCCUGCCAAUUCAACUGUA
UUAUCUUUCUGUGCUUUUGCUGUAGAUGCUGCUAAAGCUUACAAAGAUUAUCUAGCUAG
UGGGGGACAACCAAUCACUAAUUGUGUUAAGAUGUUGUGUACACACACUGGUACUGGUC
AGGCAAUAACAGUUACACCGGAAGCCAAUAUGGAUCAAGAAUCCUUUGGUGGUGCAUCG
UGUUGUCUGUACUGCCGUUGCCACAUAGAUCAUCCAAAUCCUAAAGGAUUUUGUGACUU
AAAAGGUAAGUAUGUACAAAUACCUACAACUUGUGCUAAUGACCNNNNNNNNNNNNNNN
NNNNNNNCACAGUCUGUACCGUCUGCGGUAUGUGGAAAGGUUAUGGCUGUAGUUGUGAU
CAACUCCGCGAACCCAUGCUUCAGUCAGCUGAUGCACAAUCGUUUUUAAACGGGUUUGCG
GUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUAUA
GGGCUUUUGACAUCUACAAUGAUAAAGUAGCUGGUUUUGCUAAAUUCCUAAAAACUAAU
UGUUGUCGCUUCCAAGAAAAGGACGAAGAUGACAAUUUAAUUGAUUCUUACUUUGUAGU
UAAGANNNNCACUUUCUCUAACUACCAACAUGAAGAAACAAUUUAUAAUUUACUUAAGG
AUUGUCCAGCUGUUGCUAAACAUGACUUCUUUAAGUUUAGAAUAGACGGUGACAUGGUA
CCACAUAUAUCACGUCAACGUCUUACUAAAUACACAAUGGCAGACCUCGUCUAUGCUUUA
AGGCAUUUUGAUGAAGGUAAUUGUGACACAUUAAAAGAAAUACUUGUCACAUACAAUUG
UUGUGAUGAUGAUUAUUUCAAUAAAAAGGACUGGUAUGAUUUUGUAGAAAACCCAGAUA
UAUUACGCGUAUACGCCAACUUAGGUGAACGUGUACGCCAAGCUUUGUUAAAAACAGUA
CAAUUCUGUGAUGCCAUGCGAAAUGCUGGUAUUGUUGGUGUACUGACAUUAGAUAAUCA
AGAUCUCAAUGGUAACUGGUAUGAUUUCGGUGAUUUCAUACAAACCACGCCAGGUAGUG
GAGUUCCUGUUGUAGAUUCUUAUUAUUCAUUGUUAAUGCCUAUAUUAACCUUGACCAGG
GCUUUAACUGCAGAGUCACAUGUUGACACUGACUUAACAAAGCCUUACAUUAAGUGGGA
UUUGUUAAAAUAUGACUUCACGGAAGAGAGGUUAAAACUCUUUGACCGUUAUUUUAAAU
AUUGGGAUCAGACAUACCACCCAAAUUGUGUUAACUGUUUGGAUGACAGAUGCAUUCUG
CAUUGUGCAAACUUUAAUGUUUUAUUCUCUACAGUGUUCCCACUUACAAGUUUUGGACC
ACUAGUGAGAAAAAUAUUUGUUGAUGGUGUUCCAUUUGUAGUUUCAACUGGAUACCACU
UCAGAGAGCUAGGUGUUGUACAUAAUCAGGAUGUAAACUUACAUAGCUCUAGACUUAGU

UUUAAGGAAUUACUUGUGUAUGCUGCUGACCCUGCUAUGCACGCUGCUUCUGGUAAUCU
AUUACUAGAUAAACGCACUACGUGCUUUUCAGUAGCUGCACUUACUAACAAUGUUGCUU
UUCAAACUGUCAAACCCGGUAAUUUUAACAAAGACUUCUAUGACUUUGCUGUGUCUAAG
GGUUUCUUUAAGGAAGGAAGUUCUGUUGAAUUAAAACACUUCUUCUUUGCUCAGGAUGG
UAAUGCUGCUAUCAGCGAUUAUGACUACUAUCGUUAUAAUCUACCAACAAUGUGUGAUA
UCAGACAACUACUAUUUGUAGUUGAAGUUGUUGAUAAGUACUUUGAUUGUUACGAUGGU
GGCUGUAUUAAUGCUAACCAAGUCAUCGUCAACAACCUAGACAAAUCAGCUGGUUUUCCA
UUUAAUAAAUGGGGUAAGGCUAGACUUUAUUAUGAUUCAAUGAGUUAUGAGGAUCAAGA
UGCACUUUUCGCAUAUACAAAACGUAAUGUCAUCCCUACUAUAACUCAAAUGAAUCUUA
AGUAUGCCAUUAGUGCAAAGAAUAGAGCUCGCACCGUAGCUGGUGUCUCUAUCUGUAGU
ACUAUGACCAAUAGACAGUUUCAUCAAAAAUUAUUGAAAUCAAUAGCCGCCACUAGAGG
AGCUACUGUAGUAAUUGGAACAAGCAAAUUCUAUGGUGGUUGGCACAACAUGUUAAAAA
CUGUUUAUAGUGAUGUAGAAAACCCUCACCUUAUGGGUUGGGAUUAUCCUAAAUGUGAU
AGAGCCAUGCCUAACAUGCUUAGAAUUAUGGCCUCACUUGUUCUUGCUCGCAAACAUACA
ACGUGUUGUAGCUUGUCACACCGUUUCUAUAGAUUAGCUAAUGAGUGUGCUCAAGUAUU
GAGUGAAAUGGUCAUGUGUGGCGGUUCACUAUAUGUUAAACCAGGUGGAACCUCAUCAG
GAGAUGCCACAACUGCUUAUGCUAAUAGUGUUUUUAACAUUUGUCAAGCUGUCACGGCC
AAUGUUAAUGCACUUUUAUCUACUGAUGGUAACAAAAUUGCCGAUAAGUAUGUCCGCAA
UUUACAACACAGACUUUAUGAGUGUCUCUAUAGAAAUAGAGAUGUUGACACAGACUUUG
UGAAUGAGUUUUACGCAUAUUUGCGUAAACAUUUCUCAAUGAUGAUACUCUCUGACGAU
GCUGUUGUGUGUUUCAAUAGCACUUAUGCAUCUCAAGGUCUAGUGGCUAGCAUAAAGAA
CUUUAAGUCAGUUCUUUAUUAUCAAAACAAUGUUUUUAUGUCUGAAGCAAAAUGUUGGA
CUGAGACUGACCUUACUAAAGGACCUCAUGAAUUUUGCUCUCAACAUACAAUGCUAGUU
AAACAGGGUGAUGAUUAUGUGUACCUUCCUUACCCAGAUCCAUCAAGAAUCCUAGGGGCC
GGCUGUUUUGUAGAUGAUAUCGUAAAAACAGAUGGUACACUUAUGAUUGAACGGUUCGU
GUCUUUAGCUAUAGAUGCUUACCCACUUACUAAACAUCCUAAUCAGGAGUAUGCUGAUG
UCUUUCAUUUGUACUUACAAUACAUAAGAAAGCUACAUGAUGAGUUAACAGGACACAUG
UUAGACAUGUAUUCUGUUAUGCUUACUAAUGAUAACACUUCAAGGUAUUGGGAACCUGA
GUUUUAUGAGGCUAUGUACACACCGCAUACAGUCUUACAGGCUGUUGGGGCUUGUGUUC
UUUGCAAUUCACAGACUUCAUUAAGAUGUGGUGCUUGCAUACGUAGACCAUUCUUAUGU
UGUAAAUGCUGUUACGACCAUGUCAUAUCAACAUCACAUAAAUUAGUCUUGUCUGUUAA
UCCGUAUGUUUGCAAUGCUCCAGGUUGUGAUGUCACAGAUGUGACUCAACUUUACUUAG
GAGGUAUGAGCUAUUAUUGUAAAUCACAUAAACCACCCAUUAGUUUUCCAUUGUGUGCU
AAUGGACAAGUUUUUGGUUUAUAUAAAAAUACAUGUGUUGGUAGCGAUAAUGUUACUGA
CUUUAAUGCAAUUGCAACAUGUGACUGGACAAAUGCUGGUGAUUACAUUUUAGCUAACA
CCUGUACUGAAAGACUCAAGCUUUUUGCAGCAGAAACGCUCAAAGCUACUGAGGAGACA
UUUAAACUGUCUUAUGGUAUUGCUACUGUACGUGAAGUGCUGUCUGACAGAGAAUUACA
UCUUUCAUGGGAAGUUGGUAAACCUAGACCACCACUUAACCGAAAUUAUGUCUUUACUG
GUUAUCGUGUAACUAAAAACAGUAAAGUACAAAUAGGAGAGUACACCUUUGAAAAAGGU
GACUAUGGUGAUGCUGUUGUUUACCGAGGUACAACAACUUACAAAUUAAAUGUUGGUGA
UUAUUUUGUGCUGACAUCACAUACAGUAAUGCCAUUAAGUGCACCUACACUAGUGCCACA
AGAGCACUAUGUUAGAAUUACUGGCUUAUACCCAACACUCAAUAUCUCAGAUGAGUUUU
CUAGCAAUGUUGCAAAUUAUCAAAAGGUUGGUAUGCAAAAGUAUUCUACACUCCAGGGA
CCACCUGGUACUGGUAAGAGUCAUUUUGCUAUUGGCCUAGCUCUCUACUACCCUUCUGCU
CGCAUAGUGUAUACAGCUUGCUCUCAUGCCGCUGUUGAUGCACUAUGUGAGAAGGCAUU
AAAAUAUUUGCCUAUAGAUAAAUGUAGUAGAAUUAUACCUGCACGUGCUCGUGUAGAGU
GUUUUGAUAAAUUCAAAGUGAAUUCAACAUUAGAACAGUAUGUCUUUUGUACUGUAAAU

GCAUUGCCUGAGACGACAGCAGAUAUAGUUGUCUUUGAUGAAAUUUCAAUGGCCACAAA
UUAUGAUUUGAGUGUUGUCAAUGCCAGAUUACGUGCUNAGCACUAUGUGUACAUUGGCG
ACCCUGCUCAAUUACCUGCACCACGCACAUUGCUAACUAAGGGCACACUAGAACCAGAAU
AUUUCAAUUCAGUGUGUAGACUUAUGAAAACUAUAGGUCCAGACAUGUUCCUCGGAACU
UGUCGGCGUUGUCCUGCUGAAAUUGUUGACACUGUGAGUGCUUUGGUUUAUGAUAAUAA
GCUUAAAGCACAUAAAGACAAAUCAGCUCAAUGCUUUAAAAUGUUUUAUAAGGGUGUUA
UCACGCAUGAUGUUUCAUCUGCAAUUAACAGGCCACAAAUAGGCGUGGUAAGAGAAUUC
CUUACACGUAACCCUGCUUGGAGAAAAGCUGUCUUUAUUUCACCUUAUAAUUCACAGAA
UGCUGUAGCCUCAAAGAUUUUGGGACUACCAACUCAAACUGUUGAUUCAUCACAGGGCUC
AGAAUAUGACUAUGUCAUAUUCACUCAAACCACUGAAACAGCUCACUCUUGUAAUGUAA
ACAGAUUUAAUGUUGCUAUUACCAGAGCAAAAGUAGGCAUACUUUGCAUAAUGUCUGAU
AGAGACCUUUAUGACAAGUUGCAAUUUACAAGUCUUGAAAUUCCACGUAGGAAUGUGGC
AACUUUACAAGCUGAAAAUGUAACAGGACUCUUUAAAGAUUGUAGUAAGGUAAUCACUG
GGUUACAUCCUACACAGGCACCUACACACCUCAGUGUUGACACUAAAUUCAAAACUGAAG
GUUUAUGUGUUGACAUACCUGGCAUACCUAAGGACAUGACCUAUAGAAGACUCAUCUCU
AUGAUGGGUUUUAAAAUGAAUUAUCAAGUUAAUGGUUACCCUAACAUGUUUAUCACCCG
CGAAGAAGCUAUAAGACAUGUACGUGCAUGGAUUGGCUUCGAUGUCGAGGGGUGUCAUG
CUACUAGAGAAGCUGUUGGUACCAAUUUACCUUUACAGCUAGGUUUUUCUACAGGUGUU
AACCUAGUUGCUGUACCUACAGGUUAUGUUGAUACACCUAAUAAUACAGAUUUUUCCAG
AGUUAGUGCUAAACCACCGCCUGGAGAUCAAUUUAAACACCUCAUACCACUUAUGUACAA
AGGACUUCCUUGGAAUGUAGUGCGUAUAAAGAUUGUACAAAUGUUAAGUGACACACUUA
AAAAUCUCUCUGACAGAGUCGUAUUUGUCUUAUGGGCACAUGGCUUUGAGUUGACAUCU
AUGAAGUAUUUUGUGAAAAUAGGACCUGAGCGCACCUGUUGUCUAUGUGAUAGACGUGC
CACAUGCUUUUCCACUGCUUCAGACACUUAUGCCUGUUGGCAUCAUUCUAUUGGAUUUGA
UUACGUCUAUAAUCCGUUUAUGAUUGAUGUUCAACAAUGGGGUUUUACAGGUAACCUAC
AAAGCAACCAUGAUCUGUAUUGUCAAGUCCAUGGUAAUGCACAUGUAGCUAGUUGUGAU
GCAAUCAUGACUAGGUGUCUAGCUGUCCACGAGUGCUUUGUUAAGCGUGUUGACUGGAC
UAUUGAAUAUCCUAUAAUUGGUGAUGAACUGAAGAUUAAUGCGGCUUGUAGAAAGGUUC
AACACAUGGUUGUUAAAGCUGCAUUAUUAGCAGACAAAUUCCCAGUUCUUCACGACAUU
GGUAACCCUAAAGCUAUUAAGUGUGUACCUCAAGCUGAUGUAGAAUGGAAGUUCUAUGA
UGCACAGCCUUGUAGUGACAAAGCUUAUAAAAUAGAAGAAUUAUUCUAUUCUUAUGCCA
CACAUUCUGACAAAUUCACAGAUGGUGUAUGCCUAUUUUGGAAUUGCAAUGUCGAUAGA
UAUCCUGCUAAUUCCAUUGUUUGUAGAUUUGACACUAGAGUGCUAUCUAACCUUAACUU
GCCUGGUUGUGAUGGUGGCAGUUUGUAUGUAAAUAAACAUGCAUUCCACACACCAGCUU
UUGAUAAAAGUGCUUUUGUUAAUUUAAAACAAUUACCAUUUUUCUAUUACUCUGACAGU
CCAUGUGAGUCUCAUGGAAAACAAGUAGUGUCAGAUAUAGAUUAUGUACCACUAAAGUC
UGCUACGUGUAUAACACGUUGCAAUUUAGGUGGUGCUGUCUGUAGACAUCAUGCUAAUG
AGUACAGAUUGUAUCUCGAUGCUUAUAACAUGAUGAUCUCAGCUGGCUUUAGCUUGUGG
GUUUACNNNNNNNNNGAUACUUAUAACCUCUGGAACACUUUUACAAGACUUCAGAGUUU
AGAAAAUGUGGCUUUUAAUGUUGUAAAUAAGGGACACUUUGAUGGACAACAGGGUGAAG
UACCAGUUUCUAUCAUUAAUAACACUGUUUACACAAAAGUUGAUGGUGUUGAUGUAGAA
UUGUUUGAAAAUAAAACAACAUUACCUGUUAAUGUAGCAUUUGAGCUUUGGGCUAAGCG
CAACAUUAAACCAGUACCAGAGGUGAAAAUACUCAAUAAUUUGGGUGUGGACAUUGCUG
CUAAUACUGUGAUCUGGGACUACAAAAGAGAUGCUCCAGCACAUAUAUCUACUAUUGGU
GUUUGUUCUAUGACUGACAUAGCCAAGAAACCAACUGAAACGAUUUGUGCACCACUCACU
GUCUUUUUUGAUGGUAGAGUUGAUGGUCAAGUAGACUUAUUUAGAAAUGCCCGUAAUGG
UGUUCUUAUUACAGAAGGUAGUGUUAAAGGUUUACAACCAUCUGUAGGUCCCAAACAAG

CUAGUCUUAAUGGAGUCACAUUAAUUGGAGAAGCCGUAAAAACACAGUUCAAUUAUUAU
AAGAAAGUUGAUGGUGUUGUCCAACAAUUACCUGAAACUUACUUUACUCAGAGUAGAAA
UUUACAAGAAUUUAAACCCAGGAGUCAAAUGGAAAUUGAUUUCUUAGAAUUAGCUAUGG
AUGAAUUCAUUGAACGGUAUAAAUUAGAAGGCUAUGCCUUCGAACAUAUCGUUUAUGGA
GAUUUUAGUCAUAGUCAGUUAGGUGGUUUACAUCUACUGAUUGGACUAGCUAAACGUUU
UAAGGAAUCACCUUUUGAAUUAGAAGAUUUUAUUCCUAUGGACAGUACAGUUAAAAACU
AUUUCAUAACAGAUGCGCAAACAGGUUCAUCUAAGUGUGUGUGUUCUGUUAUUGAUUUA
UUACUUGAUGAUUUUGUUGAAAUAAUAAAAUCCCAAGAUUUAUCUGUAGUUUCUAAGGU
UGUCAAAGUGACUAUUGACUAUACAGAAAUUUCAUUUAUGCUUUGGUGUAAAGAUGGCC
AUGUAGAAACAUUUUACCCAAAAUUACAAUCUAGUCAAGCGUGGCAACCGGGUGUUGCU
AUGCCUAAUCUUUACAAAAUGCAAAGAAUGCUAUUAGAAAAGUGUGACCUUCAAAAUUA
UGGUGAUAGUGCAACAUUACCUAAAGGCAUAAUGAUGAAUGUCGCAAAAUAUACUCAAC
UGUGUCAAUAUUUAAACACAUUAACAUUAGCUGUACCCUAUAAUAUGAGAGUUAUACAU
UUUGGUGCUGGUUCUGAUAAAGGAGUUGCACCAGGUACAGCUGUUUUAAGACAGUGGUU
GCCUACGGGUACGCUGCUUGUCGAUUCAGAUCUUAAUGACUUUGUCUCUGAUGCAGAUU
CAACUUUGAUUGGUGAUUGUGCAACUGUACAUACAGCUAAUAAAUGGGAUCUCAUUAUU
AGUGAUAUGUACGACCCUAAGACUAAAAAUGUUACAAAAGAAAAUGACUCUAAAGAGGG
UUUUUUCACUUACAUUUGUGGGUUUAUACAACAAAAGCUAGCUCUUGGAGGUUCCGUGG
CUAUAAAGAUAACAGAACAUUCUUGGAAUGCUGAUCUUUAUAAGCUCAUGGGACACUUC
GCAUGGUGGACAGCCUUUGUUACUAAUGUGAAUGCGUCAUCAUCUGAAGCAUUUUUAAU
UGGAUGUAAUUAUCUUGGCAAACCACGCGAACAAAUAGAUGGUUAUGUCAUGCAUGCAA
AUUACAUAUUUUGGAGGAAUACAAAUCCAAUUCAGUUGUCUUCCUAUUCUUUAUUUGAC
AUGAGUAAAUUUCCCCUUAAAUUAAGGGGUACUGCUGUUAUGUCUUUAAAAGAAGGUCA
AAUCAAUGAUAUGAUUUUAUCUCUUCUUAGUAAAGGUAGACUUAUAAUUAGAGAAAACA
ACAGAGUUGUUAUUUCUAGUGAUGUUCUUGUUAACAACUAAACGAACAAUGUUUGUUUU
UCUUGUUUUAUUGCCACUAGUCUCUAGUCAGUGUGUUAAUCUUACAACCAGAACUCAAU
UACCCCCUACAUACACUAAUUCUUUCACACGUGGUGUUUAUUACCCUGACAAAGUUUUCA
GAUCCUCAGUUUUACAUUCAACUCAGGACUUGUUCUUACCUUUCUUUUCCAAUGUUACUU
GGUUCCAUGCUAUACAUGUCUCUGGGACCAAUGUUAUUAAGAGGUUUGAUAACCCUGUC
CUACCAUUUAAUGAUGGUGUUUAUUUUGCUUCCACUGAGAAGUCUAACAUAAUAAGAGG
CUGGAUUUUUGGUACUACUUUAGAUUCGAAGACCCAGUCCCUACUUAUUGUUAAUAACG
CUACUAAUGUUGUUAUUAAAGUCUGUGAAUUUCAAUUUUGUAAUGAUCCAUUUUUGGGU
GUUUAUUANNNNNNAAACAACAAAAGUUGGAUGGAAAGUGAGUUCAGAGUUUAUUCUAG
UGCGAAUAAUUGCACUUUUGAAUAUGUCUCUCAGCCUUUUCUUAUGGACCUUGAAGGAA
AACAGGGUAAUUUCAAAAAUCUUAGGGAAUUUGUGUUUAAGAAUAUUGAUGGUUAUUUU
AAAAUAUAUUCUAAGCACACGCCUAUUAAUUUAGUGCGUGAUCUCCCUCAGGGUUUUUC
GGCUUUAGAACCAUUGGUAGAUUUGCCAAUAGGUAUUAACAUCACUAGGUUUCAAACUU
UACUUGCUUUACAUNNNNNNNNNNNNNNNNNNNNNAAUUCUUCUUCAGGUUGGACAGCU
GGUGCUGCAGCUUAUUAUGUGGGUUAUCUUCAACCUAGGACUUUUCUAUUAAAAUAUAA
UGAAAAUGGAACCAUUACAGAUGCUGUAGACUGUGCACUUGACCCUCUCUCAGAAACAA
AGUGUACGUUGAAAUCCUUCACUGUAGAAAAAGGAAUCUAUCAAACUUCUAACUUUAGA
GUCCAACCAACAGAAUCUAUUGUUAGAUUUCCUAAUAUUACAAACUUGUGCCCUUUUGG
UGAAGUUUUUAACGCCACCAGAUUUGCAUCUGUUUAUGCUUGGAACAGGAAGAGAAUCA
GCAACUGUGUUGCUGAUUAUUCUGUCCUAUAUAAUUCCGCAUCAUUUUCCACUUUUAAG
UGUUAUGGAGUGUCUCCUACUAAAUUAAAUGAUCUCUGCUUUACUAAUGUCUAUGCAGA
UUCAUUUGUAAUUAGAGGUGAUGAAGUCAGACAAAUCGCUCCAGGGCAAACUGGAAAGA
UUGCUGAUUAUAAUUAUAAAUUACCAGAUGAUUUUACAGGCUGCGUUAUAGCUUGGAAU

UCUAACAAUCUUGAUUCUAAGGUUGGUGGUAAUUAUAAUUACCAGUAUAGAUUGUUUAG
GAAGUCUAAUCUCAAACCUUUUGAGAGAGAUAUUUCAACUGAAAUCUAUCAGGCCGGUA
GCACACCUUGUAAUGGUGUUGAAGGUUUUAAUUGUUACUCUCCUUUACAAUCAUAUGGU
UUCCAACCCACUAAUGGUGUUGGUUACCAACCAUACAGAGUAGUAGUACUUUCUUUUGA
ACUUCUACAUGCACCAGCAACUGUUUGUGGACCUAAAAAGUCUACUAAUUUGGUUAAAA
ACAAAUGUGUCAAUUUCAACUUCAAUGGUUUAACAGGCACAGGUGUUCUUACUGAGUCU
AACAAAAAGUUUCUGCCUUUCCAACAAUUUGGCAGAGACAUUGCUGACACUACUGAUGC
UGUCCGUGAUCCACAGACACUUGAGAUUCUUGACAUUACACCAUGUUCUUUUGGUGGUG
UCAGUGUUAUAACACCAGGAACAAAUACUUCUAACCAGGUUGCUGUUCUUUAUCAGGGU
GUUAACUGCACAGAAGUCCCUGUUGCUAUUCAUGCAGAUCAACUUACUCCUACUUGGCGU
GUUUAUUCUACAGGUUCUAAUGUUUUUCAAACACGUGCAGGCUGUUUAAUAGGGGCUGA
ACAUGUCAACAACUCAUAUGAGUGUGACAUACCCAUUGGUGCAGGUAUAUGCGCUAGUU
AUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGUAGUGUAGCUAGUCAAUCCAUCAUUG
CCUACACUAUGUCACUUGGUGCAGAAAAUUCAGUUGCUUACUCUAAUAACUCUAUUGCCA
UACCCACAAAUUUUACUAUUAGUGUUACUACAGAAAUUCUACCAGUGUCUAUGACCAAG
ACAUCAGUAGAUUGUACAAUGUACAUUUGUGGUGAUUCAACUGAAUGCAGCAAUCUUUU
GUUGCAAUAUGGCAGUUUUUGUACACAAUUAAACCGUGCUUUAACUGGAAUAGCUGUUG
AACAAGACAAAAACACCCAAGAAGUUUUUGCACAAGUCAAACAAAUUUACAAAACACCAC
CAAUUAAAGAUUUUGGUGGUUUUAAUUUUUCACAAAUAUUACCAGAUCCAUCAAAACCA
AGCAAGAGGUCAUUUAUUGAAGAUCUACUUUUCAACAAAGUGACACUUGCAGAUGCUGG
CUUCAUCAAACAAUAUGGUGAUUGCCUUGGUGAUAUUGCUGCUAGAGACCUCAUUUGUG
CACAAAAGUUUAACGGCCUUAAUGUUUUGCCACCUUUGCUCACAGAUGAAAUGAUUGCU
CAAUACACUUCUGCACUGUUAGCGGGUACAAUCACUUCUGGUUGGACCUUUGGUGCAGG
UGCUGCAUUACAAAUACCAUUUGCUAUGCAAAUGGCUUAUAGGUUUAAUGGUAUUGGAG
UUACACAGAAUGUUCUCUAUGAGAACCAAAAAUUGAUUGCCAACCAAUUUAAUAGUGCU
AUUGGCAAAAUUCAAGACUCACUUUCUUCCACAGCAAGUGCACUUGGAAAACUUCAAGA
UGUGGUCAACCAANNNNNNNNAGCUUUAAACACGCUUGUUAAACAACUUAGCUCCAAUU
UUGGUGCAAUUUCAAGUGUUUUAAAUGAUAUCCUUUCACGUCUUGACAAAGUUGAGGCU
GAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGAC
UCAACAAUUAAUUAGAGCUGCAGAAAUCAGAGCUUCUGCUAAUCUUGCUGCUACUAAAA
UGUCAGAGUGUGUACUUGGACAAUCAAAAAGAGUUGAUUUUUGUGGAAAGGGCUAUCAU
CUUAUGUCCUUCCCUCAGUCAGCACCUCAUGGUGUAGUCUUCUUGCAUGUGACUUAUGUC
CCUGCACAAGAAAAGAACUUCACAACUGCUCCUGCCAUUUGUCAUGAUGGAAAAGCACAC
UUUCCUCGUGAAGGUGUCUUUGUUUCAAAUGGCACACACUGGUUUGUAACACAAAGGAA
UUUUUAUGAACCACAAAUCAUUACUACAGACAACACAUUUGUGUCUGGUAACUGUGAUG
UUGUAAUAGGAAUUGUCAACAACACAGUUUAUGAUCCUUUGCAACCUGAAUUAGACUCA
UUCAAGGAGGAGUUAGAUAAAUAUUUUAAGAAUCAUACAUCACCAGAUGUUGAUUUAGG
UGACAUCUCUGGCAUUAAUGCUUCAGUUGUAAACAUUCAAAAAGAAAUUGACCGCCUCA
AUGAGGUUGCCAAGAAUUUAAAUGAAUCUCUCAUCGAUCUCCAAGAACUUGGAAAGUAU
GAGCAGUAUAUAAAAUGGCCAUGGUACAUUUGGCUAGGUUUUAUAGCUGGCUUGAUUGC
CAUAGUAAUGGUGACAAUUAUGCUUUGCUGUAUGACCAGUUGCUGUAGUUGUCUCAAGG
GCUGUUGUUCUUGUGGAUCCUGCUGCAAAUUUGAUGAAGACGACUCUGAGCCAGUGCUC
AAAGGAGUCAAAUUACAUUACACAUAAACGAACUUAUGGAUUUGUUUAUGAGAAUCUUC
ACAAUUGGAACUGUAACUUUGAAGCAAGGUGAAAUCAAGGAUGCUACUCCUUCAGAUUU
UGUUCGCGCUACUGCAACGAUACCGAUACAAGCCUCACUCCCUUUCGGAUGGCUUAUUGU
UGGCGUUGCACUUCUUGCUGUUUUUCAGAGCGCUUCCAAAAUCAUAACCCUCAAAAAGAG
AUGGCAACUAGCACUCUCCAAGGGUGUUCACUUUGUUUGCANCUUGCUGUUGUUGUUUG

UAACAGUUUACUCACACCUUUUGCUCGUUGCUGCUGGCCUUGAAGCCCCUUUUCUCUAUC
UUUAUGCUUUAGUCUACUUCUUGCAGAGUAUAAACUUUGUAAGAAUAAUAAUGAGGCUU
UGGCUUUGCUGGAAAUGCCGUUCCAAAAACCCAUUACUUUAUGAUGCCAACUAUUUUCU
UUGCUGGCAUACUAAUUGUUACGACUAUUGUAUACCUUACAAUAGUGUAACUUCUUCAA
UUGUCAUUACUUCAGGUGAUGGCACAACAAGUCCUAUUUCUGAACAUGACUACCAGAUU
GGUGGUUAUACUGAAAAAUGGGAAUCUGGAGUAAAAGACUGUGUUGUAUUACACAGUUA
CUUCACUUCAGACUAUUACCAGCUGUACUCAACUCAAUUGAGUACAGACACUGGUGUUGA
ACAUGUUACCUUCUUCAUCUACAAUAAAAUUGUUGAUGAGCCUGAAGAACAUGUCCAAA
UUCACACAAUCGACGGUUCAUCCGGAGUUGUUAAUCCAGUAAUGGAACCAAUUUAUGAU
GAACCGACGACGACUACUAGCGUGCCUUUGUAAGCACAAGCUGAUGAGUACGAACUUAU
GUACUCAUUCGUUUCGGAAGAGACAGGUACGUUAAUAGUUAAUAGCGUACUUCUUUUUC
UUGCUUUCGUGGUAUUCUUGCUAGUUACACUAGCCAUCCUUACUGCGCUUCGAUUGUGU
GCGUACUGCUGCAAUAUUGUUAACGUGAGUCUUGUAAAACCUUCUUUUUACGUUUACUC
UCGUUUUAAAAAUCUGAAUUCUUCUAGAGUUCCUGAUCUUCUGGUCUAAACGAACUAAA
UAUUAUAUUAGUUUUUCUGUUUGGAACUUUAAUUUUAGCCAUGGCAGAUUCCAACGGUA
CUAUUACCGUUGAAGAGCUUAAAAAGCUCCUUGAACAAUGGAACCUAGUAAUAGGUUUC
CUAUUCCUUACAUGGAUUUGUCUUCUACAAUUUGCCUAUGCCAACAGGAAUAGGUUUUU
GUAUAUAAUUAAGUUAAUUUUCCUCUGGCUGUUAUGGCCAGUAACUUUAGCUUGUUUUG
UGCUUGCUGCUGUUUACAGAAUAAAUUGGAUCACCGGUGGAAUUGCUAUCGCAAUGGCU
UGUCUUGUAGGCUUGAUGUGGCUCAGCUACUUCAUUGCUUCUUUCAGACUGUUUGCGCG
UACGCGUUCCAUGUGGUCAUUCAAUCCAGAAACUAACAUUCUUCUCAACGUGCCACUUCA
UGGCACUAUUCUGACCAGACCGCUUCUAGAAAGUGAACUCGUAAUCGGAGCUGUGAUCCU
UCGUGGACAUCUUCGUAUUGCUGGACACCAUCUAGGACGCUGUGACAUCAAGGACCUGCC
UAAAGAAAUCACUGUUGCUACAUCACGAACGCUUUCUUAUUACAAAUUGGGAGCUUCGC
AGCGUGUAGCAGGUGACUCAGGUUUUGCUGCAUACAGUCGCUACAGGAUUGGCAACUAU
AAAUUAAACACAGACCAUUCCAGUAGCAGUGACAAUAUUGCUUNNNNNNNNNNNNNNGU
GACAACAGAUGUUUCAUCUCGUUGACUUUCAGGUUACUAUAGCAGAGAUAUUACUAAUU
AUUAUGAGGACUUUUAAAGUUUCCAUUUGGAAUCUUGAUUACAUCAUAAACCUCAUAAU
UAAAAAUUUAUCUAAGUCACUAACUGAGAAUAAAUAUUCUCAAUUAGAUGAAGAGCAAC
CAAUGGAGAUUGAUUAAACGAACAUGAAAAUUAUUCUUUUCUUGGCACUGAUAACACUC
GCUACUUGUGAGCUUUAUCACUACCAAGAGUGUGUUAGAGGUACAACAGUACUUUUAAA
AGAACCUUGCUCUUCUGGAACAUACGAGGGCAAUUCACCAUUUCAUCCUCUAGCUGAUAA
CAAAUUUGCACUGACUUGCUUUAGCACUCAAUUUGCUUUUGCUUGUCCUGACGGCGUAA
AACACGUCUAUCAGUUACGUGCCAGAUCAGUUUCACCUAAACUUUUCAUCAGACAAGAGG
AAGUUCAAGAACUUUACUCUCCAAUUUUUCUUAUUGUUGCGGCAAUAGUGUUUAUAACA
CUUUGCUUCACACUCAAAAGAAAGACAGAAUGAUUGAACUUUCAUUAAUUGACUUCUAU
UUGUGCUUUUUAGCCUUUCUGCUAUUCCUUGUUNUAAUUAUGCUUAUUAUCUUUUGGUU
CUCACUUGAACUGCAAGAUCAUAAUGAAACUUGUCACGCCUAAACGAACAUGAAAUUUC
UUGUUUUCUUAGGAAUCAUCACAACUGUAGCUGCAUUUCACCAAGAAUGUAGUUUACAG
UCAUGUACUCAACAUCAACCAUAUGUAGUUGAUGACCCGUGUCCUAUUCACUUCUAUUCU
AAAUGGUAUAUUAGAGUAGGAGCUAGAAAAUCAGCACCUUUAAUUGAAUUGUGCGUGGA
UGAGGCUGGUUCUAAAUCACCCAUUCAGUACAUCGAUAUCGGUAAUUAUACAGUUUCCU
GUUUACCUUUUACAAUUAAUUGCCAGGAACCUAAAUUGGGUAGUCUUGUAGUGCGUUGU
UCGUUCUAUGAAGACUUUUUAGAGUAUCAUGACGUUCGUGUUGUUUUAGAUUUUAUCUA
AACGAACAAACUUAAAUGUCUGAUAAUGGACCCCAAAAUCAGCGAAAUGCACUCCGCAUU
ACGUUUGGUGGACCCUCAGAUUCAACUGGCAGUAACCAGAAUGGAGAACGCAGUGGGGC
GCGAUCAAAACAACGUCGGCCCCAAGGUUUACCCAAUAAUACUGCGUCUUGGUUCACCGC

UCUCACUCAACAUGGCAAGGAAGACCUUAAAUUCCCUCGAGGACAAGGCGUUCCAAUUAA
CACCAAUAGCAGUCCAGAUGACCAAAUUGGCUACUACCGAAGAGCUACCAGACGAAUUCG
UGGUGGUGACGGUAAAAUGAAAGAUCUCAGUCCAAGAUGGUAUUUCUACUACCUAGGAA
CUGGGCCAGAAGCUGGACUUCCCUAUGGUGCUAACAAAGACGGCAUCAUAUGGGUUGCA
ACUGAGGGAGCCUUGAAUACACCAAAAGAUCACAUUGGCACCCGCAAUCCUGCUAACAAU
GCUGCAAUCGUGCUACAACUUCCUCAAGGAACAACAUUGCCAAAAGGCUUCUACGCAGAA
GGGAGCAGAGGCGGCAGUCAAGCCUCUUCUCGUUCCUCAUCACGUAGUCGCAACAGUUCA
AGAAAUUCAACUCCAGGCAGCAGUAANCGAACUUCUCCUGCUAGAAUGGCUGGCAAUUGC
GGUGAUGCUGCUCUUGCUUUGCUGCUGCUUGACAGAUUGAACCAGCUUGAGAGCAAAAU
GUCUGGUAAAGGCCAACAACAACAAGGCCAAACUGUCACUAAGAAAUCUGCUGCUGAGGC
UUCUAAGAAGCCUCGGCAAAAACGUACUGCCACUAAAGCAUACAAUGUAACACAAGCUUU
CGGCAGACGUGGUCCAGAACAAACCCAAGGAAAUUUUGGGGACCAGGAACUAAUCAGAC
AAGGAACUGAUUACAAACAUUGGCCGCAAAUUGCACAAUUUGCCCCCAGCGCUUCAGCGU
UCUUCGGAAUGUCGCGCAUUGGCAUGGAAGUCACACCUUCGGGAACGUGGUUGACCUACA
CAGGUGCCAUCAAAUUGGAUGACAAAGAUCCAAAUUUUAAAGAUCAAGUCAUUUUGCUG
AAUAAGCAUAUUGACGCAUACAAAACAUUCCCACCAAYAGAGCCUAAAAAGGACAAAAA
GAAGAAGGCUGAUGAAACUCAAGCCUUACCGCAGAGACAGAAGAAACAGCAAACUGUGA
CUCUUCUUCCUGCUGCAGAUUUGGAUGAUUUCUCCAAACAAUUGCAACAAUCCAUGAGCA
GUGCUGACUCAACUCAGGCCUAAACUCAUGCAGACCACACAAGGCAGAUGGGCUAUAUAA
ACGUUUUCGCUUUUCCGUUUACGAUAUAUAGUCUACUCUUGUGCAGAAUGAAUUCUCGU
AACUACAUAGCACAAGUAGAUGUAGUUAACUUUAAUCUCACAUAGCAAUCUUUAAUCAG
UGUGUAACAUUAGGGAGGACUUGAAAGAGCCACCACAUUUUCACCGAGGCCACGCGGAG
UACGAUCGAGUGUACAGUGAACAAUGCUAGGGAGAGCUGCCUAUAUGGAAGAGCCCUAA
UGUGUAAAAUUAAUUUUAGUAGUGCUAUCCCCAUGUGAUUUUAAUAGCUUCUUAGGAGA
AUGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Protein Sequence (Using NCBI)

```
  1 mfvflvllpl vssqcvnltt rtqlpptytn sftrgvyypd kvfrssvlhs tqdlflpffs
 61 nvtwfhaihv sgtnvikrfd npvlpfndgv yfasteksni irgwifgttl dsktqslliv
121 nnatnvvikv cefqfcndpf lgvyxxxnnk swmesefrvy ssannctfey vsqpflmdle
181 gkqgnfknlr efvfknidgy fkiyskhtpi nlvrdlpqgf saleplvdlp iginitrfqt
241 llalhxxxxx xxnsssgwta gaaayyvgyl qprtfllkyn engtitdavd caldplsetk
301 ctlksftvek giyqtsnfrv qptesivrfp nitnlcpfge vfnatrfasv yawnrkrisn
361 cvadysvlyn sasfstfkcy gvsptklndl cftnvyadsf virgdevrqi apgqtgkiad
421 ynyklpddft gcviawnsnn ldskvggnyn yqyrlfrksn lkpferdist eiyqagstpc
481 ngvegfncys plqsygfqpt ngvgyqpyrv vvlsfellha patvcgpkks tnlvknkcvn
541 fnfngltgtg vltesnkkfl pfqqfgrdia dttdavrdpq tleilditpc sfggvsvitp
601 gtntsnqvav lyqgvnctev pvaihadqlt ptwrvystgs nvfqtragcl igaehvnnsy
661 ecdipigagi casyqtqtns prrarsvasq siiaytmslg aensvaysnn siaiptnfti
721 svtteilpvs mtktsvdctm yicgdstecs nlllqygsfc tqlnraltgi aveqdkntqe
781 vfaqvkqiyk tppikdfggf nfsqilpdps kpskrsfied llfnkvtlad agfikqygdc
841 lgdiaardli caqkfnglnv lpplltdemi aqytsallag titsgwtfga gaalqipfam
901 qmayrfngig vtqnvlyenq klianqfnsa igkiqdslss tasalgklqd vvnqxxxaln
961 tlvkqlssnf gaissvlndi lsrldkveae vqidrlitgr lqslqtyvtq qliraaeira
1021 sanlaatkms ecvlgqskrv dfcgkgyhlm sfpqsaphgv vflhvtyvpa qeknfttapa
```

1081 ichdgkahfp regvfvsngt hwfvtqrnfy epqiittdnt fvsgncdvvi givnntvydp
1141 lqpeldsfke eldkyfknht spdvdlgdis ginasvvniq keidrlneva knlneslidl
1201 qelgkyeqyi kwpwyiwlgf iagliaivmv timlccmtsc csclkgccsc gscckfdedd
1261 sepvlkgvkl hyt

Longest Amino Acid Length as obtained in FGENESV

1. Paste the Sequence of Longest Amino Acid:

MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQHLKDGTCGLVEVEKGV

LPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGETLGVLVPHVGEIPVAYRK

VLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSGVTRELMRELNGG

AYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQLDFIDTKRGVYCCREHEHEIAW

YTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFPLNSIIKTIQPRVEKKKLDGFMGRI

RSVYPVASPNECNQMCLSTLMKCDHCGETSWQTGDFVKATCEFCGTENLTKEGATTCGYL

PQNAVVKIYCPACHNSEVGPEHSLAEYHNESGLKTILRKGGRTIAFGGCVFSYVGCHNKC

AYWVPRASANIGCNHTGVVGEGSEGLNDXXXEILQKEKVNINIVGDFKLNEEIAIILASF

SASTSAFVETVKGLDYKAFKQIVESCGNFKVTKGKAKKGAWNIGEQKSILSPLYAFASEA

ARVVRSIFSRTLETAQNSVRVLQKAAITILDGISQYSLRLIDAMMFTSDLATNNLVVMAY

ITGGVVQLTSQWLTNIFGTVYEKLKPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIV

GGQIVTCAKEIKESVQTFFKLVNKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKC

VKSREETGLLMPLKAPKEIIFLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVG

TPVCINGLMLLEIKDTEKYCALAPNMMVTNNTFTLKGGAPTKVTFGDDTVIEVQGYKSVN

ITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVADAVIKTLQPVSELLTPLGIDLDEW

SMATYYLFDESGEFKLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTEDDYQGKPL

EFGATSAALQPEEEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVV

QTIEVNSFSGYLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNXATN

NAMQVESDDYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQ

HEVLLAPLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVE

QKIAEIPKEEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEEIKFLTENLLLYIDIN

GNLHPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKV

PTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREMLA

HAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLINTLND

LNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSSKTPEEH

FIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVITFDNLKTLL

SLREVRTIKVXXTVDNINLHTQVVDMSMTYGQQFGPTYLDGADVTKIKPHNSHEGKTFYV

LPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIKWADNNCYLATAL

LTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELGDVRETMSYLFQHAN

LDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQIPCTCGKQATKYLVQQ

ESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSKETLYCIDGALLTKSSEYK

GPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYYKKDNSYFTEQPIDLVPNQPY

PNASFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVTFFPDLNGDVVAIDYKHYTPSFK

KGAKLLHKPIVWHVNNATNKATYKPNTWCIRCLWSTKPVETSNSFDVLKSEDAQGMDNLA

CEDLKPVSEEVVENPTIQKDVLECNVKTTEVVGDIILKPANNSLKITEEVGHTDLMAAYV

DNSSLTIKKPNELSRVLGLKTLATHGLAAVNSVPWDTIANYAKPFLNKVALVSDVGDSAE

VAVKMFDAYVNTFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQGFVDSD

VETKDVVECLKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAK

SHNIALIWNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGGKI

VNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIASTDT

CFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLHFL

PRVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSVAY

ESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGRWV

LNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGAVLQSGFRKMAFPSG

KVESCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLVQAG

NVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMR

PNFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTA

QAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILG

PLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGYFCTCYFGL

FCLLNRYFRLTLGVYDYLVSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVA

TVQSKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLS

VLLSMQGAVDINKLCEEMLDNRATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVV

LKKLKKSLNVAKSEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTM

LRKLDNDALNNIINNARDGCVPLNIIXXTTAAKLMVVIPDYNTYKNTCDGTTFTYASALW

EIQQVVDADSKIVQLSEISMDNSPNLAWPLIVTALRANSAVKLQNNELSPVALRQMSCAA

GTTQTACTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDGTGTIYTELEPPCRFV

TDTPKGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQAGNATEVPANSTVLSFCAFAVDA

AKAYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDH

PNPKGFCDLKVCTVCGMWKGYGCSCDQLREPMLQSADAQSFLNGFAVIDGDMVPHISRQR

LTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANL

GERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSYY

SLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHPNC

VNCLDDRCILHCANFNVLFSTVFPLTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQD

VNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNKD

FYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVDK

YFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIPT

ITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGW

HNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLARKHTTCCSLSHRFYRLANE

CAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIADK

YVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMMILSDDAVVCFNSTYASQGLVAS

IKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRIL

GAGCFVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNQEYADVFHLYLQYIRKLHDELTGH

MLDMYSVMLTNDNTSRYWEPEFYEAMYTPHTVLQAVGACVLCNSQTSLRCGACIRRPFLC

CKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDVTQLYLGGMSYYCKSHKPPISFPLCAN

GQVFGLYKNTCVGSDNVTDFNAIATCDWTNAGDYILANTCTERLKLFAAETLKATEETFK

LSYGIATVREVLSDRELHLSWEVGKPRPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYG

DAVVYRGTTTYKLNVGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNV

ANYQKVGMQKYSTLQGPPGTGKSHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLP

IDKCSRIIPARARVECFDKFKVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSV

VNARLRAXHYVYIGDPAQLPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPA

EIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAW

RKAVFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAIT

RAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATLQAENVTGLFKDCSKVITGLHPTQAPT

HLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRA

WIGFDVEGCHATREAVGTNLPLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPPPGDQ

FKHLIPLMYKGLPWNVVRIKIVQMLSDTLKNLSDRVVFVLWAHGFELTSMKYFVKIGPER

TCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQWGFTGNLQSNHDLYCQVHG

NAHVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHMVVKAALLADK

FPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEELFYSYATHSDKFTDGVCLFW

NCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFHTPAFDKSAFVNLKQLPFF

YYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRLYLDAYNMMISAG

FSLWVYXXXDTYNLWNTFTRLQSLENVAFNVVNKGHFDGQQGEVPVSIINNTVYTKVDGV

DVELFENKTTLPVNVAFELWAKRNIKPVPEVKILNNLGVDIAANTVIWDYKRDAPAHIST

IGVCSMTDIAKKPTETICAPLTVFFDGRVDGQVDLFRNARNGVLITEGSVKGLQPSVGPK

QASLNGVTLIGEAVKTQFNYYKKVDGVVQQLPETYFTQSRNLQEFKPRSQMEIDFLELAM

DEFIERYKLEGYAFEHIVYGDFSHSQLGGLHLLIGLAKRFKESPFELEDFIPMDSTVKNY

FITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKVVKVTIDYTEISFMLWCKDGHV

ETFYPKLQSSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDSATLPKGIMMNVAKYTQLCQ

YLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTLLVDSDLNDFVSDADSTLI

GDCATVHTANKWDLIISDMYDPKTKNVTKENDSKEGFFTYICGFIQQKLALGGSVAIKIT

EHSWNADLYKLMGHFAWWTAFVTNVNASSSEAFLIGCNYLGKPREQIDGYVMHANYIFWR

NTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKEGQINDMILSLLSKGWTAGAAAYYVGYLQ

PRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPN

ITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC

FTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNY

QYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYSPLQSYGFQPTNGVGYQPYRVV

VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIAD

TTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTP

TWRVYSTGSNVFQTRAGCLIGAEHNNSYECDIPIGAGICASYQTQTNSPRRARSVASQS

IIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSN

LLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSK

PSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLNVLPPLLTDEMIA

QYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAI

GKIQDSLSSTASALGKLQDVVNQXXXALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEV

QIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMS

FPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYE

PQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISG

INASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWAVVLVDPAANLMKTTLS

QCSKESNYITHKRTYGFVYENLHNWNCNFEASVTSSIVITSGDGTTSPISEHDYQIGGYT

EKWESGVKDCVVLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEPEEHVQIHTID

GSSGVVNPVMEPIYDEPTTTTSVPLLFARTRSMWSFNPETNILLNVPLHGTILTRPLLES

ELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSYYKLGASQRVAGIITTVAA

FHQECSLQSCTQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIELCVDEAGSKSPIQYID

IGNYTVSCLPFTINCQEPKLGSLVRNALRITFGGPSDSTGSNQNGERSGARSKQRRPQGL

PNNTASWFTALTQHGKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLS

PRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGT

TLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTPGSSXRTSPARMAGNCGDAALALLLLD

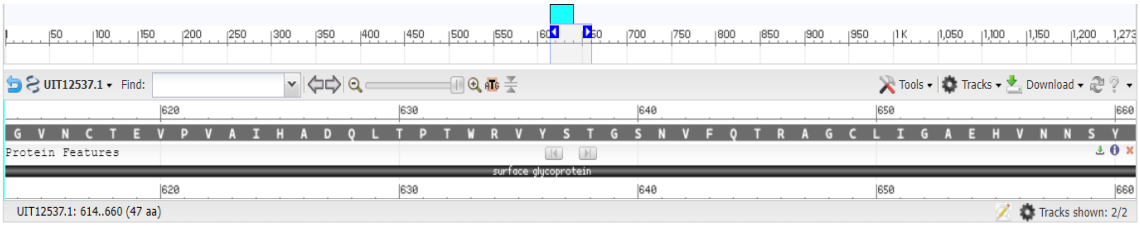
RLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGN

FGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPN

FKDQVILLNKHIDAYKTFPPXEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSK

QLQQSMSSADSTQA

Since protein structure is not available, graphical summary of the protein sequence is as follows:



Conclusion and Inference:

1. **Genes and their Location:** This nucleotide sequence contains 11 different genes which are located at the following locations and following CDS

Gene Location	CDS	CDS Length
212..21501	join(212..13414,13414..21501)	join(212..13414,13414..21501) = 21289 bp
21509..25330	21509..25330	21509..25330 = 3821 bp
25339..26166	25339..26166	25339..26166 = 827 bp
26191..26418	26191..26418	26191..26418 = 227 bp

26469..>27123	26469..>27123	26469..>27123 = 654 bp
27148..27333	27148..27333	27148..27333 = 185 bp
27340..27705	27340..27705	27340..27705 = 365 bp
27702..27833	27702..27833	27702..27833 = 131 bp
27840..28205	27840..28205	27840..28205 = 365 bp
28220..29479	28220..29479	28220..29479 = 1259 bp
29504..29620	29504..29620	29504..29620 = 116 bp

2. **Exon Count and its Location:** Since viral genome, so no exons are present
3. **A, T, G. C Composition:** The given gene sequence of surface glycoprotein UIT12537.1 gene consists of:
    i. **A:** 8899
    ii. **T:** 9548
    iii. **G:** 5833
    iv. **C:** 5440
4. **Number of Start Codons (ATG):** 723
5. **Number of Stop Codons (TAG+ TGA + TAA)**: 1766
6. **Percentage GC Content**- 37.766759355422295%
7. **The Length of mRNA Transcript is** …29849…….. bp
8. **Total number of EcoRI are:** 9
9. **Total number of BamHI are:** 1
10. **Total HindIII are**: 17
11. **Intron-Exon Boundary at positions:** 212, 21501, 21509, 25330, 25339, 2616626191, 26418, 26469, 27123, 27148, 27333, 27340, 27705, 27702, 27833, 27840, 28205, 28220, 29479, 29504, 29620

Using FGENESV and GENSCAN

12. **Number of Isochore:** 1(0 - 43 C+G%)
13. **Number of genes:** 1 gene (12 coding sequences) in GENSCAN, 11 genes in FGENESV0, 12 genes in FGENESV
14. **Number of Initial exon** (ATG to 5' splice site): 1
15. **Number of Internal exon** (3' splice site to 5' splice site): 10
16. **Number of Terminal exon** (3' splice site to stop codon): 1
17. **Number of Single-exon gene (ATG to stop)**: 0
18. **Length of longest Promoter (TATA box / initation site):** 0 (no promoter region mentioned
19. **Length of longest poly-A signal** (consensus: AATAAA): 6
20. **Maximum Coding region score (tenth bit units):** 4112
21. **(No exon found) Maximum P: probability of exon (sum over all parses containing exon):** 0.975
22. **Length of Longest Amino Acid**: 7814 aa

By ORF Finder:

23. **The ORF that has Maximum Length:** ORF4
24. **The length of ORF that has Maximum Length:** 13218 nucleotides

25. **The location of START Codon in Maximum Length of ORF:** 212
26. **The location of STOP Codon in Maximum Length of ORF:** 13429
27. **The Direction of STRAND in Maximum Length of ORF:** Positive
28. **The Length of Amino Acid encoded by BIGGEST ORF:** 4405

Task 3: Self dot plot of gene, mRNA, protein sequence, and draw inference from your dot plots.

Self Dot-Plot for gene

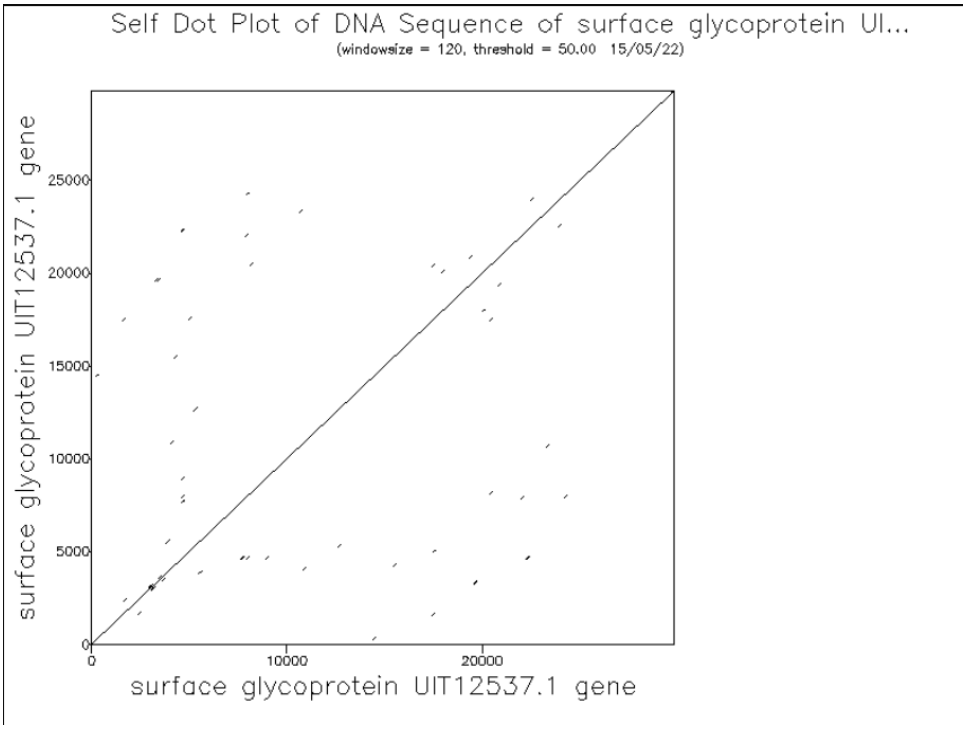


Figure 1: Self dot-match of Surface Glycoprotein UIT12537.1 gene

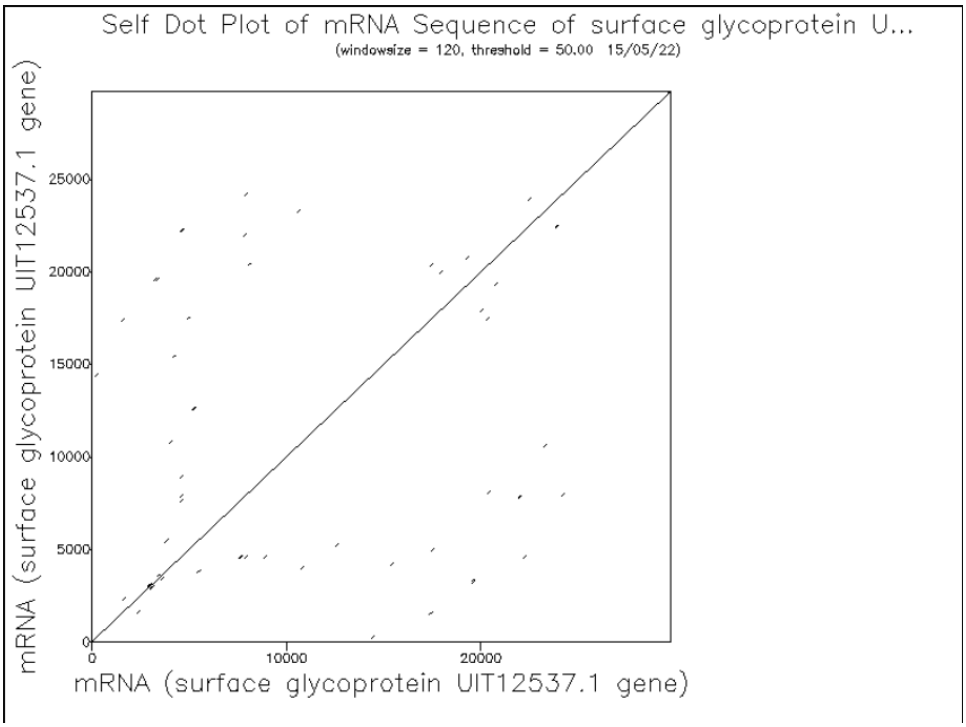


Figure 2: Self dot-match of Surface Glycoprotein UIT12537.1 mRNA Sequence

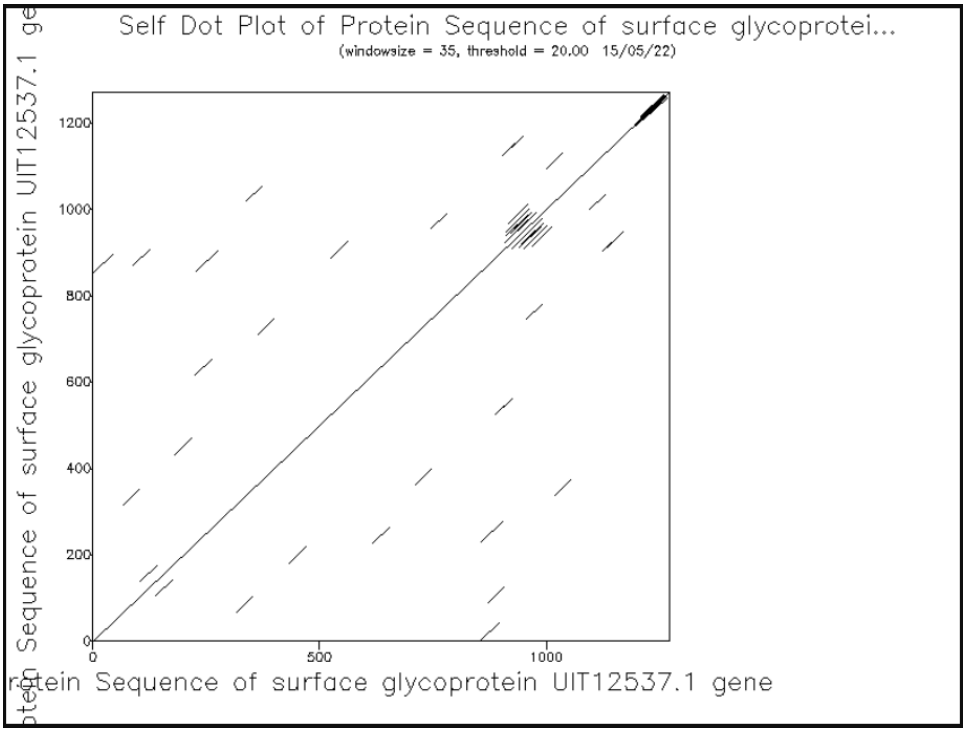


Figure 2: Self dot-match of Surface Glycoprotein UIT12537.1 Protein Sequence

Conclusion and Inference:

The main diagonal represents the sequence's alignment with itself. Since, here we have plotted self dot-plot of DNA, mRNA and protein, we observed a diagonal here.

1. The Pairwise Sequence Alignment (PSA) of Surface Glycoprotein UIT12537.1 gene with itself shows maximum alignment by selecting window size 120 and threshold 50.00.
2. The Pairwise Sequence Alignment (PSA) of Surface Glycoprotein UIT12537.1 gene with itself shows maximum alignment by selecting window size 120 and threshold 50.00.
3. The Pairwise Sequence Alignment (PSA) of Surface Glycoprotein UIT12537.1 gene with itself shows maximum alignment by selecting window size 35 and threshold 20.00. The repeats being parallel to the main diagonal means that the repeats are occurring in different places throughout the protein sequence but are not inverted.

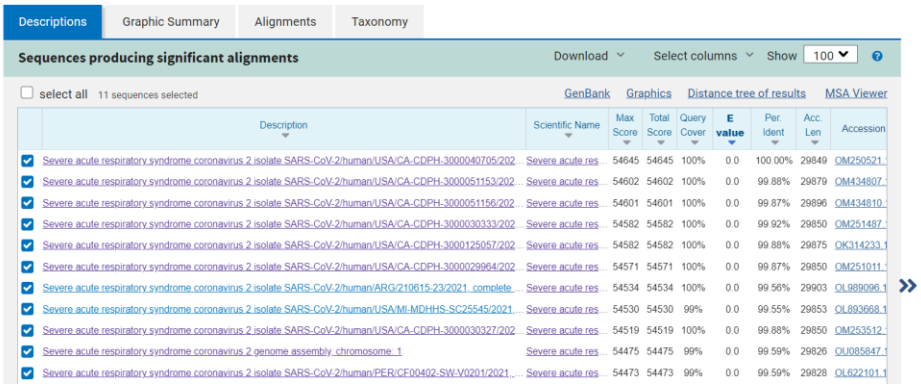
Task 4: BLAST your gene and protein sequence and find the top 5 eukaryotic and 5 prokaryotic genes and protein (from different species) sequences. Tabulate results mentioning score, E-value, % of identity, query coverage, etc.

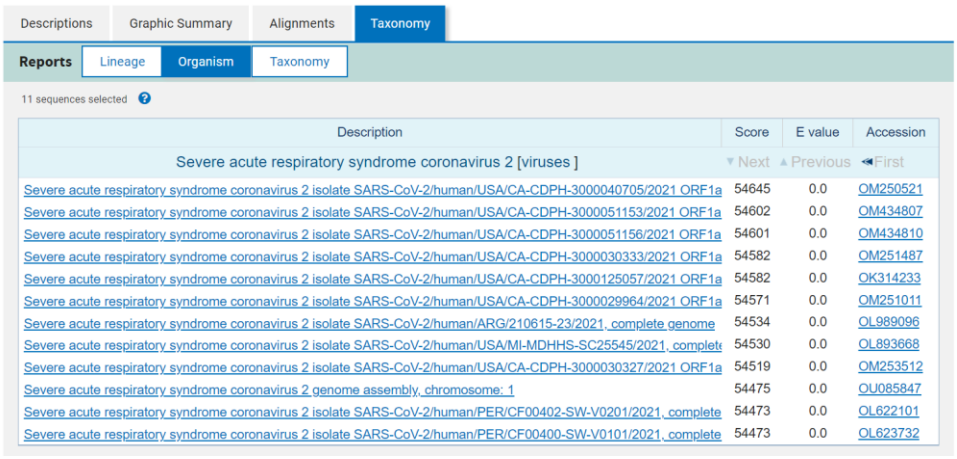
Since the gene belongs to SARS-CoV-2 belong to a virus, it does not share similarity with any prokaryote or eukaryote gene. Therefore, instead of any prokaryote or eukaryote, I have taken the top ten sequences most similar to the surface glycoprotein UIT12537.1 that appeared in BLAST (Megablast for similar sequences).

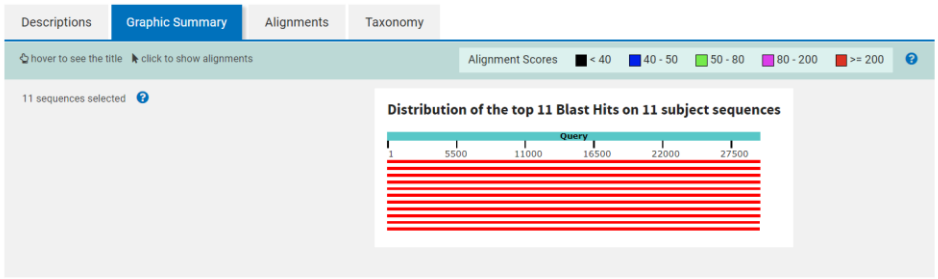
BLASTN

S.No	Accession Number	Max Score	Total Score	E-Value	Query Coverage	% Identity
1	OM434807.1	54602	54602	0.0	100%	99.88%
2	OM434810.1	54601	54601	0.0	100%	99.87%
3	OM251487.1	54582	54582	0.0	100%	99.92%
4	OK314233.1	54582	54582	0.0	100%	99.88%
5	OM251011.1	54571	54571	0.0	100%	99.87%
6	OL989096.1	54534	54534	0.0	100%	99.56%
7	OL893668.1	54530	54530	0.0	99%	99.55%
8	OM253512.1	54519	54519	0.0	100%	99.88%
9	OU085847.1	54475	54475	0.0	99%	99.59%
10	OL622101.1	54473	54473	0.0	99%	99.59%

Screenshots of BLASTN Interface and Results





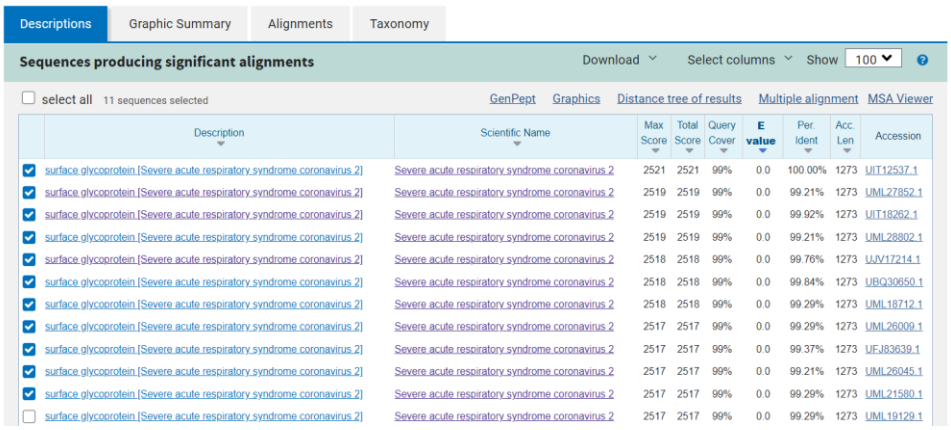


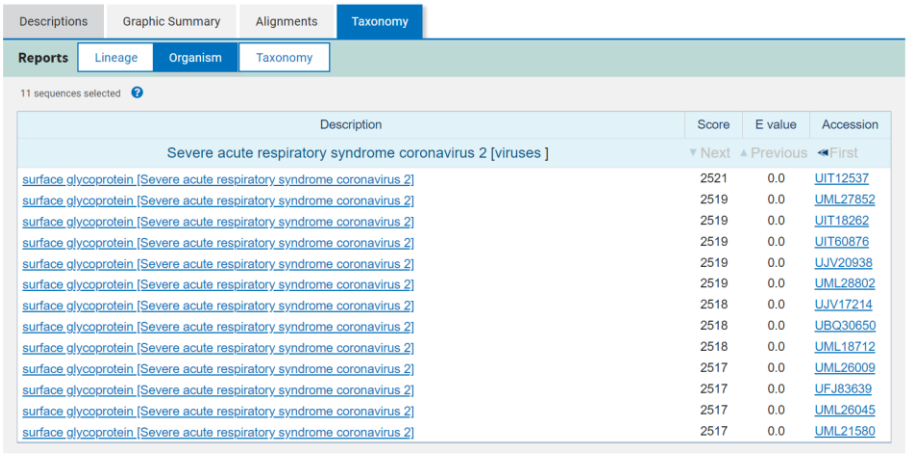
BLASTP

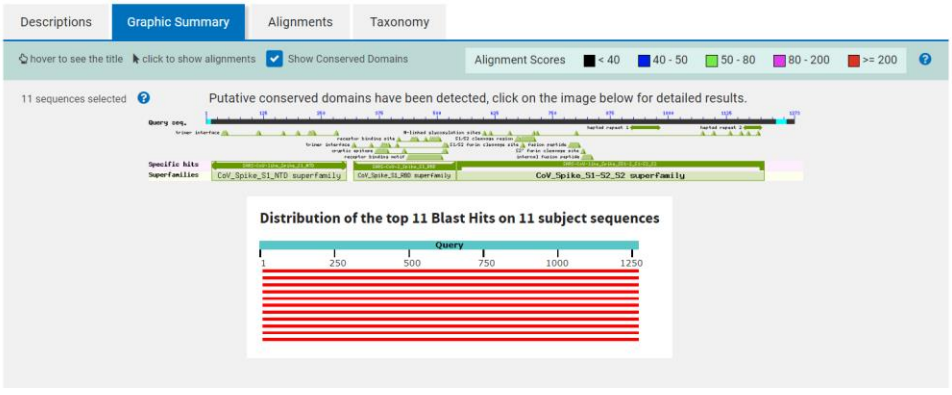
S.No	Accession Number	Max Score	Total Score	E-Value	Query Coverage	% Identity
1	UML27852.1	2519	2519	0.0	99.0%	99.21%
2	UIT18262.1	2519	2519	0.0	99.0%	99.92%
3	UML28802.1	2519	2519	0.0	99.0%	99.21%
4	UJV17214.1	2518	2518	0.0	99.0%	99.76%
5	UBQ30650.1	2518	2518	0.0	99.0%	99.84%
6	UML18712.1	2518	2518	0.0	99.0%	99.29%
7	UML26009.1	2517	2517	0.0	99.0%	99.29%
8	UFJ83639.1	2517	2517	0.0	99.0%	99.37%

9	UML26045.1	2517	2517	0.0	99.0%	99.21%
10	UML21580.1	2517	2517	0.0	99.0%	99.29%

Screenshots of BLASTP and Interface







Conclusion and Inference

As the highest score indicates highest similarity in BLAST, The Pairwise Sequence Alignment (PSA) using BLASTn of surface glycoprotein UIT12537.1 gene nucleotide sequence shows maximum alignment with **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000051153/2021 (OM434807.1) having Score 54562, E-value 0 and percentage similarity of 99.88%,** respectively.

The top ten organisms that are showing maximum similarity with surface glycoprotein UIT12537.1 gene

1. **SARS-CoV-2/human/USA/CA-CDPH-3000051153/2021 (OM434807.1)** having **E-value 0 and percentage similarity of 99.88%**, respectively.
2. **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000051156/2021 (OM434810.1)** having E-value...**0.0**....and Percentage similarity...**99.87%**..... respectively.
3. **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000030333/2021 (OM251487.1)** having E-value...**0.0**....and Percentage similarity...**99.92%**..... respectively.
4. **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000125057/2021 (OK314233.1)** having E-value...**0.0**....and Percentage similarity...**99.88%**..... respectively.
5. **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000029964/2021 (OM251011.1)** having E-value...**0.0**....and Percentage similarity...**99.87%**..... respectively.
6. **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ARG/210615-23/2021, complete genome (OL989096.1)** having E-value...**0.0**....and Percentage similarity...**99.56%**..... respectively.
7. **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MI-MDHHS-SC25545/2021, complete genome (OL893668.1)** having E-value...**0.0**....and Percentage similarity...**99.55%**..... respectively.
8. **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDPH-3000040705/2021 (OM253512)** having E-value...**0.0**....and Percentage similarity...**99.88%**..... respectively.
9. **Severe acute respiratory syndrome coronavirus 2 genome assembly, chromosome: 1 (OU085847.1)** having E-value...**0.0**....and Percentage similarity...**99.59%**..... respectively.
10. **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/PER/CF00402-SW-V0201/2021 (OL622101.1)** having E-value...**0.0**....and Percentage similarity...**99.59%**..... respectively.

B) The Pairwise Sequence Alignment (PSA) using BLASTp of surface glycoprotein UIT12537.1 gene amino acid sequence shows maximum alignment with **Surface Glycoprotein UML27852.1,**

Surface Glycoprotein UIT18262.1 and Surface Glycoprotein UML28802.1 with score 2519 E-value 0 and percentage similarity of 79.37%, respectively.

The top protein sequences that are showing maximum similarity with Surface Glycoprotein UIT12537.1

11. **Surface Glycoprotein UML27852.1** having E-value...**0.0....** and Percentage similarity.....99.21**%**.... respectively
12. **Surface Glycoprotein UIT18262.1** having E-value...**0.0..**.... and Percentage similarity.....**99.92%**....respectively
13. **Surface Glycorotein UML28802.1** having E-value...**0.0...**. and Percentage similarity...**..99.21%**.... respectively.
14. **Surface Glycoprotein UJV17214.1** having E-value...**0.0..**. and Percentage similarity**....99.76%.**.....respectively
15. **Surface Glycoprotein UBQ30650.1** having E-value**....0.0.**.... and percentage similarity.....**99.84%**...respectively
16. **Surface Glycoprotein UBQ30650.1** having E-value**....0.0.**.... and percentage similarity.....**99.29%**...respectively
17. **Surface Glycoprotein UML18712.1** having E-value**....0.0.**.... and percentage similarity.....**99.29%**...respectively
18. **Surface Glycoprotein UML26009.1** having E-value**....0.0.**.... and percentage similarity.....**99.37%**...respectively
19. **Surface Glycoprotein UFJ83639.1** having E-value**....0.0.**.... and percentage similarity.....**99.21%**...respectively
20. **Surface Glycoprotein UML21580.1** having E-value**....0.0.**.... and percentage similarity.....**99.29%**...respectively

Task 5: Perform pairwise global alignment (Needleman-Wunch, EMBOSS) of your query sequence (both DNA and protein) with all 10 sequences which you retrieved by BLAST search. Tabulate your result by mentioning score and % of identity etc.

Pairwise global alignment (Needleman-Wunch using NCBI and EMBOSS) of DNA sequence tabulation

SR. No	Accession No.	Needleman-Wunch Score	EMBOSS Needle Score	Needleman-Wunch Percent Identity	EMBOSS Needle Percent Identity
1.	OM434807.1	58943	60423.0	100.00%	99.8%

2.	OM434810.1	58909	60598.5	100.00%	99.7%
3.	OM251487.1	58981	60530.5	100.00%	99.9%
4.	OK314233.1	58930	60587.0	100.00%	99.8%
5.	OM251011.1	58971	60786.5	100.00%	99.9%
6.	OL989096.1	58825	59992.5	99.00%	99.4%
7.	OL893668.1	58887	61011.5	100.00%	99.5%
8.	OM253512.1	58896	60255.5	100.00%	99.9%
9.	OU085847.1	58720	60568.0	99.00%	99.3%
10.	OL622101.1	58710	59543.5	99.00%	99.3%

Pairwise global alignment (Needleman-Wunch using NCBI and EMBOSS) of Protein sequence tabulation

SR.No	**Accession No.**	**Needleman-Wunch Score**	**EMBOSS Needle Score**	**Needleman-Wunch Percent Identity**	**EMBOSS Needle Percent Identity**
1.	UML27852.1	6629	6631.0	99.0%	99.2%
2.	UIT18262.1	6629	6629.0	100.0%	99.9%
3.	UML28802.1	6629	6632.0	99.00%	99.2%
4.	UJV17214.1	6629	6629.0	100.00%	99.8%
5.	UBQ30650.1	6629	6629.0	100.00%	99.8%
6.	UML18712.1	6629	6631.0	99.00%	99.3%
7.	UML26009.1	6621	6622.0	99.00%	99.3%
8.	UFJ83639.1	6624	6625.0	99.00%	99.4%
9.	UML26045.1	6621	6622.0	99.00%	99.2%

10.	UML21580.1	6624	6625.0	99.00%	99.3%

Conclusion/Inference

As the per cent identity gives the percentage of identical matches between the two sequences over the reported aligned region (including any gaps in the length), therefore following conclusions are derived.

1. For Nucleotide: The Pairwise Sequence Alignment (PSA) using **EMBOSS NEEDLE** between set of the DNA sequences as retrieved by BlastN shows maximum similarity of **Surface Glycoprotein UIT12537.**1 (OM250521) with **SARS-CoV-2 Isolate (Accession Number: OM253512.1) showing 99.9%** identity as shown in the table.

   If PSA is performed using **NCBI Needle-Wunch Method**, the maximum similarity of **Surface Glycoprotein UIT12537.1 (OM250521.1) is with genes with accession number: OM434807.1, OM434810.1, OM251487.1, OK314233.1, OM251011.1, OL893688.1, OM253512.1 each showing 100% percentage identity** as shown in the table.

2. For Protein: The Pairwise Sequence Alignment (PSA) using **EMBOSS NEEDLE** between set of protein sequences as retrieved by BlastP shows maximum similarity of **Surface Glycoprotein UIT12537.**1 with **surface glycoprotein UIT18262.1** with **99.9%** identity as shown in the table.

   If PSA is performed using **NCBI Needle-Wunch Method**, the maximum similarity of **Surface Glycoprotein UIT12537.1 is with Surface Glycoprotein UIT18262.1, Surface Glycoprotein UJV17214.1 and UBQ30650.1 with 100% identity.**

Task 6: Make multiple sequence alignment (CLUSTALW, or any other popular algorithm or software) of your query with the other 20 sequences that you retrieved through blast (both DNA and protein). Analyze your result by mentioning consensus sequences and motifs that you may identify through MSA

Results of Multiple Sequence Alignment of DNA Sequences using CLUSTALW

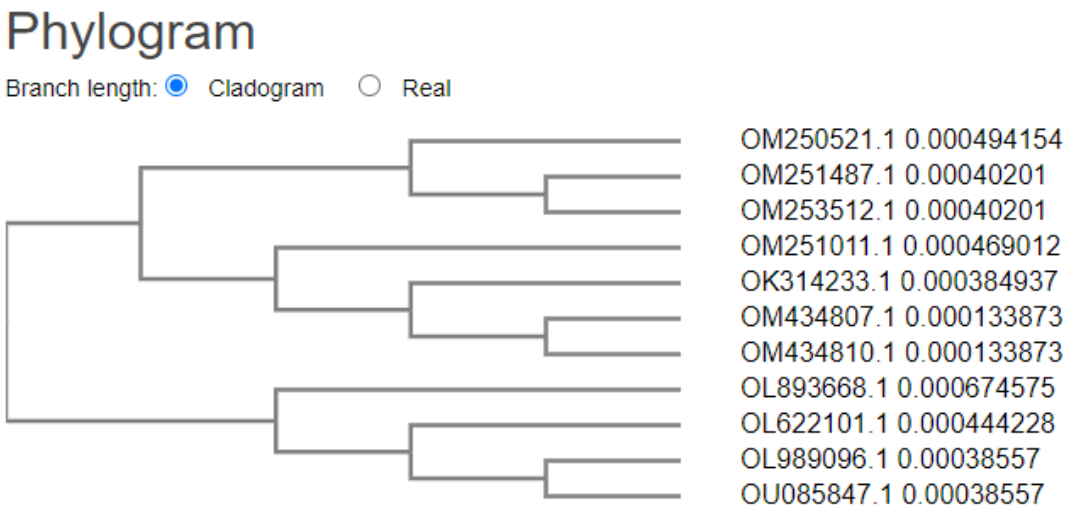


Fig: Phylogram or Guide Tree of Nucleotide Sequence obtained in ClustalW

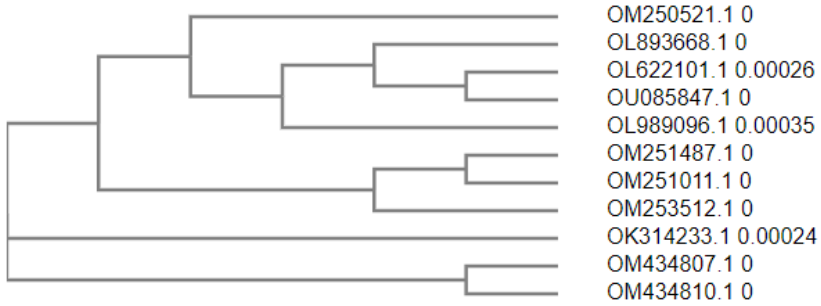


Fig: Phylogenetic Tree of Nucleotide Sequence obtained in ClustalW

Consensus Sequence generated by EMBOSS Cons for given set of nucleotides

>EMBOSS0001

AnATCnnnnnTCTAnnnnnnCTTTnAAnTnnnTnTGnCTnnnAnnCGnnTnnAnnCTTnG
TGnnnnCnnnCAGTATnnTTnATnAnTAnTnnCnnnCnnTnnCAGnAnnCAnnnAnCnCG
TCTnTCGTCTGCAGGCTnCTTACnGnTTnnTCnTnnTnnGnnnCnnTCnTCnGnnCTnCT
nGGTGTTGnCnGnGTnCnnCnGAAnnnnnAGnTGnAGnGCnnTGnCCCnnGTnTnAnnnn
GAnAACAnnCnTCCnACTnnGTTnGnCnnTnnnACAGGTTnnCGACGTnnTCnTnCnTnn
CTnnnGnGnCTCCnnGTnGnAGGTCTTATCnGAGGCnCnTCnnnnGnnnAnnGATnnnAn
nTGnGnCnTnnTnnAAnTnAAAAnAGnnGTTnTGCCTnAnnTnnAAnAGnnnnATGTGTT
CnTCAAnnnnTnGGAnnCTCGnAnTTCAnCnnAnGnTCnTnTGATGGTTGAGCTGnTnnn
nGnnnTCGAAnGnnTnnnGTnCnGnCnnAGnnnnnAGACAnTnGGTnnnnTnGTCnnnnn
nGTGnGCnAnnTGnCAGTGnCnTnCCGCAnGGnGCnnnnTCnnnAGnnnnnTAAnAnAnG
nnnnnGnGnCnnnAGnnnnGnCnnnGAnCTnnnGnCnnTnGTCnTnGnnGnnGnGnTnnG
nAnTnATnCTTATGAnGnnnnTCnAGnAnnnTnGAACATTAAnnATAnCAnnnnnnnTAC
nnGTnAACTCATnnnTGAGGTnAnnnnAGnGGnnTACnCnnGCnnnnTCGnnAACnAnTn

nnnTnnCCCTGnTnnnTnnnCnnnTnCnTnnATnnnnnAnnTTnnnnCnCnnnnTGGTnn
AnCTnCnTGCTnnTnGnnCGTnCnAnTGnAnnTTnnTGnnnnTnnGAnGnnTnnnTnnTn
nTGnCnnnnACnnGGGnnnGAAAnTGCTnnnnnCAnGGAnnGnnnTGAAAAGnGnnATGn
AnnGCAGAnnCnTnTnAAAAnTAnnATGnnnAnnAAAnTTGAnAnnTTnAATnGGnAAnG
nnnAAATTTnnnATTTnCnnTnAATnnCATAATnAAGAnTnTTCAnCCAAnnnTTnAAnA
nAnAAnACTnnAnnnCTnnnTGnnnAGAAnnAGATnTnTGnnTnCAnTnGnGnnAnnAAA
TGAnTnnAnCCAnnTGTGnCnTnnAACTnTCAnnAnnnnTGATnnTTnTnnTnAAnCnTC
ATGnnnGATGnnCGTTTTnGTnAAAnnnAnTnGCGAAnTGnGTnTnnnnnAGAAnTTGnC
TAAnGAnGnTGCnnCnAnnnGTnGTTnnnTAnnnCAnAnnnCTnTTGTTnnnATTTATTn
nCCAnnAnnnCnnAnTTnAnAAnTAGnnCnnnAGCATAnnnTTnnnnAnTAnCATnAnnn
ATnTGnCTTGnnGACnATTATnCnTAAGnnTGnTnGnACnAnAnCCTTTGnAGGnTGnGT
GnTCTCnTTnnTTnnTTGnnATGnnnnGTGTnCTnTTTnGGTTGCnnGTnCTAGnGnTnn
nATTGGTTGTAACCATnnAGnnGTTGTnnnAGAnGTTnCCnnAnnnCTnnnGnnnnnnnn
nnnTGAnAnnnTCnnAAnAnnGAnAnnCnAnATCAAnnnTnnnnATnAnTnnAAACTTAA
AnnAnAGnTnGnnATTATTTTnnnATTnnTnAnnnCTTCnnCATTTnnTTnTnTnnnnnn
TnTGnnAnGTnTGGTTTTTnnnnnATTnnnAnAAATTGTTnnnTnnnnnnnTAATTTnAA
AnTTTTAAAnnnTAAAGnnAAAAnAnnnnnnTnnAAnATnGnTnAACAGAAAnnAnTnCn
nAGnnnnnTnnAnnnAnnTGnATnATAGGnnGnnCGTnTTnTnCnnTnAATTnTCnnnCT
CnTTnTGnAnnnTGnTCAAnnTTCTGnnCnnGTTTTACAGnnGnCCGnTATnnnnATnCT
nnATnGAAnTnCGCAGTAnTnACnnnnnnTnATnGATnnTAnnATnTTCAnATnnGATTT
GnnTATTnACTATTnAnnTnTnATGnnnTAnnTTACAnnTnnnnTnTTnnAGTTGACTTC
nCAGTnGnTAnTGnnCATCTnTnGnnnTnnTnnTGAAnAACTAAAAnCnGTnCTnnATTn
GnTnnAAnAnAAnnTnnnGnnnnnTGTnGnGTTnCnnAGAnnnnGnnGGGAGnTGnnTAA
nTTnATnnCnACnnGTGnTTnnTnAATTnTnnnTGGACAAnTTGTnAnnTnTnnAnnGGn
nnTnAAnnAGnnnnnTCAGnnAnnCnnnAAnCTnnnAAnnnnnTTnnnnnCTTTnTGnnn
nGAnTnnATnATTTTTnGTnGTGnTnTnCTTAAAnCnTTnTnnTTnGnGGAAAnATTnnT
nACGCnnTnAAAGGnnnTGnAnnnnnnGTGnnnTAAATnnAGnGAnnnAACnGGnnTnnT
nATnCCTCnAAAnnnnCnAAnnGAnnnnAnCTTCTTAnAGGnnnAAAnAnnnCnCACAnC
nnnGTnAAnnGAGnAAnTnnTCnnGAAnAnTGGTnAnTnnnAACnnTTnnnAnAACnnAn
nAnTnAnnAnnnTnAnnnnnnAnTnnTnGnTAnnCnGnTnTnnATnCACnGGnTnnTGnT
nCACGAnnTnnnAnnnACAGnAnAnTnCTGnGCnnnnnCAnCnAAnAnGAnGnnnACnnA
CAnnAnCnTCAnAnTnnAAnGCnnnnCACCAACAnCnnnnACnTnnGnnGnnGACnnnnn
nAnAnnAGTnCnAGGnnnnAnGAnnnTGnAnATnAnnTTTGAnnTTnnAGnAnGGATAnA
nnnAnTAnTnAAnnnGAAGTGATnTnTnAAnAnnnTnnAAnTnnGTACnGnnnTnAnTnA
GTTnnnCTnTGTnGTGnnnnAnnnnATnAnnnnAnCTTTGnAACnnGTAnCnnnATTAnn
nAnACnnnTGnnCnnTGATnTnnATTAnTnnnnnnnGGCnnCATnCTAnTTAnTTGnTGn
ATnnnGnnAnnnTAnATnTGnTTnGnnTATGnnTnGTTnTTTTTAnCnTCnnGAnnnnnn
TGAnGnnnnAnnTnnnTGnnAnnnGnnnnAGAnnnAGnnnnnnACTnAnAAGnAGAATGG
TAnnGAnnATnAnnACnAAnGTAnnnnTnTGnAnTTTAnTACCACnnnnGnnnnTCTTnA
nnCTGAnGnAnAnCnnnnAnnnnTnTnnnTAnATGAnAAnAGnnAnnAnAnTGTTGGTnA
nnAGnAnGnCAnnGAnnAnnnGnTGAnAAnnACnAnnCAAAnnAnnATnAAnnnnCAACn
nCnATTAnAnATGnAACTnAnAnCnnTnnTTnAGAnnAnnGAnnTnnAnAATTTnnnnnG
nnAnTTAAAACTTAnAGnnnnTGTAnAnATTAAAAAnGnAnAnATnnTGnAnnAAnCnAA
AAAGnTAAAAnCAAnAnTGAAnnTTAAnAnAnnnAAnnnnnAnCnnnAAnnTnnAnnnnn
nnTnnnnnnnnCCTTAAAnAnGGnnnnnAnCnAnGACnTGCAAnTTGnAnnnnnnGnnnn
nATAnCnAnnAAnnnnCCACTGAnnATnnnnnnnnnnnTnGTnTTnnnCnnAnAnnnnnT
nGnTnAnnnnTnTnnnnAnGTnGTCGGCCnAAAnGnTAACAnnnnnnAnnnnnnnCAACT

nnTnAnnAnTGnnnAnnAAAnnTTnAAnnTnnAnGAAGTTCTAATnnnAnnATTAnTnAC
nACTnnnATTTTnnnTnnTnAnnATnnnCAnTTTTTAnTnGTTnnnnTnnnnnnnnTTCn
nnnAnAnGTCTnnTTnnnTnTnTTnGnTnAAAAnnTnTnTnAnnAnCTTnTTTnAAnnAT
TTTGnnnATGAAnAnnnAAnAnCAAGTTnnAnAAAAnAnAGnnGAGAnnnnnAAnnAGnA
AATnAAGCCnnTnATAAnnnAAAGnAAAnnnTnAnnTnAAnAnnGAAAAnAAGAnnATAA
GnAAATCAAAGnTnAnAnTnAGnnAnTAAnAACAACnCTGGAnGnAATnAAGTTnCTAnn
AnAAnnCTnGAAAnTATATAnTnnCnnnAAnnnnAATnnTnATnCAnnnTCnnCAAnnnT
nGTnATnnTCATnnAnATnnCnnTCTTnATGnTAGATnCnnCATnnnTAnTnnnTnATnn
nnTnCnnnAnnnTnTnTnnnnnGTGnTGnTTATnnnnAnTATnnAGnnnnnnGGnnnnnn
nnAAnTnCTAnnGAnnGnnGTGnnAnTAnTGnnAAnnGAnAAnnATnTAAnAAnTTACCn
nnnnCAGnGTnTAAAnnGnnAnACTnnAGnGGnGGnAnAnAnAGTGnTTAnAnAGTGTAA
AAGnnnAnnnTACTTnnTAnnAnnTATnATnTCTnnTnnnAnnnAAnAAnTTCTAnnAnC
TnTnTnnnGnAAnTTGnnAnTnATGCTnnnnCATGnAGnAnnnAnACnnAAATTnnTGCn
TnTnnAnnnGnAAnCnAAAnnnnnAnTnTCAnCTnTAnAGCnnAAAnAnnAGnnTnTTnA
AnTAnAAGnnGnTnTnnTnGnTTnnnGnnnnAAnnnnnAGTTTTnnAnnATTAAnAnAnn
TGTnnCnTCTnTTATnnnnAnACnTAAnnATCnACATnAAACAnnAnnnnCAATGnCnnT
nnnnAATnTAnTAnnnnnCTnAnAnTTGGAnGnnnnnnnnCGGnnTnTGAnATnTnnnAA
nnTGCCAGnTAnAnTnTnTGTTTCnTCAnnTnAnnnnnTTnnnnnGTATnnTnnTnAnCT
GATTTCnTnTTnTAnAATAnnTGnnnnACnTnTTATTnTAAnnAnCnnnCTTnnTAnTTn
nTnnAAAnAATnnTCCTAnTTnGTnCnnnCTACAAAnnTnGGTATTnnnnTTnnTAAnnn
AnATnAnnAAnnnnAnTATTACAnnAGnnAnnnnAnnAnATTnnAnnTAnnnGnAGnAGT
nnTAACnTnTnACAAnnTnAnGACnCTTnTnTCTTTGAnAnnAGTGAGAAnnnTTnTnnT
nnnnnnnAnnnnAGACAnnnTTAnnnnnnAnnnnCAAnTnnnGAnnATnTCAAnnACAnA
nnGAnAATnGTnnnGTCnAnnnTATnTGGAnGnnGnTnATGTnnnnAAAATnAAnnnTnA
nnAnTnAnnTnTAnnnAAAAnAnTnnnnnTTnTACnTAATGAnGAAAnTCTTCnTnTTnA
GnnTnTnGnGTnnnnCCACnnnnTnnnnCnTAnTTTTnnnnnTAnnnAnnTGnTnnnAnT
AAnTnnCnnnnnAAnnTGnnAnnAnCnAnAAnTnAnAnGTTnAAnTTnTAnTAAAnAGnn
AnnTAnnAAnTnTCATCTnnnnAnTGnnnTGnTAACAnTTnAnnnnnTAnnGnTGAATTT
TAAnTCnnCnnnnnnAnAnnnTnnTTnTTnCAnAnnACnGnnnnAnnAnnnTnnTAnnnn
nTAnGnACTGnTnnTnnCnnAnnnTAnTnTGnnnnTAGnTnAGTnnnnTGnTGTnAGAnA
AnAGATGAGTTAnTTGTTnCAAnAnGCnAAATnAGATTCTTGnnAnAnnnTCnTGAnnnT
GnTGTnTnnAACnnAnGnnTnAnnGnnGnCAnCnnTnAAnGGTGTnGAnnnTGTnAnnTA
CnTnGGnAnnnTnTAnnAnnnATnATnTAnGAAAGnnGTTCnnATnnCnnnTACnTGnnG
nAAnnTnnnnAnAnAAnATCTAGTnnnAnAnnAGnnACnAnnTnTTATGTnGTAnnnACn
ACnTnCTCAGTAnnnAnTTnAGCAnnAnnCATTnAnTnGTnnnAnnnnnnnCAnnGnTnA
TTAnCAGnnnGTnnnnnATAnACTnATAACTTCTAnnGnnACTnTnAATnnCATAnAnCn
TnnTnnACTTnnTAAnnCnTCAnAnnnnAnAnnTnnTnTnAnnnnTnTTnnnTAnAAAGA
AnnCAnATAnAnnACAnCnAnAAAAnnAnnnACTnnTAAAnTAnAnnnnnnnGnnTnnnn
AnAAATTnAnnnnAnGTTGnnnnTTnATTAnAAGAnnnACAnTTnTTAnTTnATAnnnCn
AAnAAnnAnTCTTnTnnCAnAnCAnAnAnAnCnAATCnCnnnCTnCnAnAAnnnTAAnTT
TnnAnnTGATAnTATnAAATTTnnTnnTnATTTAAnnnTnTTAACTnnTTAnATGATAnn
TGCnACnAGAnAGCTTAnAGTTAnAnTTTTCCnTGAnnnnAnnnnTnAnnTnnTnGCTnT
nnnTTATAAAnAnTTnACACnnTnnTTnATGATAAnAGCTAnAnTnTTnCnTATAnnnAn
TGnnnnGCATnTTnACAATnnAAnnnnnAAnGnCAnGTTnAAACCAAAnnCnnnnnnnAT
ACGTnGTATnnnnAnAACAnAnCnAGTnGAnnnnTnAnnTnCnnnAnAnGTnCTGAAGnn
AGAnnAnnnnCAGnnnATGGATAATCnTnnnTGnnnnnnnnnAAAACnAATCTCnnnnGn
nGnAnTGnAAAAnCnnACnnTnCAGAAnnAnGnnCnnnAnnnnnnnnnGAAAAnnACnnn

nnTnnTAGGAGnnATnAnACTnnAACCnnnAAAGnAnAnnTTAnAAATTAnAnAnGAGnT
AnGAnACnnAnAACTAATGnCnAnnTnTGTnGnCAAnnnnAnnnnnnnGnTTAnGAAnnC
nAATnnAnTATTnATAnTATTAAGTnTnAnAAnnnnTnnTACTnATnATTTAGnTnnTnn
nnnnnnTnTCCTnTGGGATnnGnTAnnnATTnnnGCTnAnCnTTTnCTnnAnnAAnTTnT
TAGnAnAAnTnnTAnCATAnnnAnnnGGTnTnTAnAnCTTnTnTnnTnnnAnnATnTnnn
TTATnTCTTTnTTnTAnTnnTAnnATTnnnTACTTTTnnTAnTTGTTnnAATTCTnGAnT
TnAAnnATnAATnCnnAnTATTnTAGnAAAGAATAnTnTTnnnnGTnTAnnnnAAnnATG
AnTAnTGnCnnnATTnAnnnATTTGnTnTnnnnnAnTnTTTnTATAnTnATAAGTATTnT
AATTTnTTTTTTAnnATTAAnnnTTTnnnTAGGTTCTTTAnTCTAnTnAnTnGCTGCTTT
nGnTnTTTTAATTnCTnATnTnCTnnTGnCTTnTTACTGTnnTnTTTAnnnAGnAnGCnn
TnTnAnnTnTnnTAnTGnCAnnAnTGCnAnnTnnnnTnnnnnnTCnnTATnnTnTnnTnT
TTnTCTTnTTGTTTTAnATnCTnTAGTCnTCTnnnnTnTTTTAnAnATTATnnAAAnTAC
CATTnnTTnTTAnAAATnnnAnTTAnCTnCTTTTnnnTTAnTTGnAGnGTnnTTnnTGnT
ATnTnnTnTnTTnAnTAGTTTTTTnTATnTACTnnGnTTnnCnGnAnTTTTnnnnTTGTT
nTTnnGnTnTnTnnCATnAnATTnTnTTnnTnATTCTTnGCTnnTGTnnTTAnTAnTTAn
TCTTGTnnnATTnGnnnnnATTTnAGnTATGTTnAnAAnGTACnTCTTnnTTnnATnATT
nTAnTATnTATnGnnnTnGTATnAGCATnTTnTnnAnGGTnGnAATTTnTnAAnnTnTnT
GATGTnTnAnAnACnTnATnnAGnAnnAnGAGTTGAAAnTAnAATAATnGnTnATGnnGn
nAnnAnnnnnnTnTnnGTnTnnGnTnTnGnAGGTnnnnnnTTTTnnAnnCTACnnAATnn
GAAnTnTGTTnnTTnTGAnACAnTnnGTGnTGnTnnnAnAnTTATnAnTnATGnnGTnGn
GGTAGAnTTnTnAnTAnTGnTTAAAnGACnAAnAAnTCnTAAnnACnAnTnnnCnnACnA
CnTnnnTCTnnTnACnnTnnAnAnTnnnTCnATnnnTCTTTAnTnTnnnAnAGTTGnTCn
AAnGAnTTnTGnnATAnAnnnnCTCTnnnATnTTnTTAAnnnAGACAnCCTnnnAnnTAA
TnACACTAAAnnTTCATTnnnnnTTAnTnTTATAnTnTTTGATnGTAnATCAAAATnnnA
AnAnTnnTnnGTAAAATnAnAnnnTGTnnAnTATnnTCAGCATAnGnGTCAnCnTATAnC
nnnnnnnnnnnnnnGnnTTnGTnTnnnATnTTnGnnnTAGnnnGnAnTnTnnAnTnAnnAT
GnTTGAnGnTTAnnTGnAnnnGTTATnATnnAnnnTTAAnnTAnnAATnnnnAAnCTnAA
AnCAnTAnnTnnAAnTnnAnAAnCnnAACnnnAAnAGnAnnTGnnnnAAnACAAnnTCTT
AnnTAnnTTTnnTTCnnCAnCnnGGnAnnnGnTTnTTnnTTnnnAnnTnGnAAnTAAAGn
TTTnnTTnAnnGnnTnAAATTnTnAAATnTnTnnnAnATnGTAnTTAnTnGCnAnAnTnn
TAAnAAnTATATGCTnnCnnnTAAnnnAGTnnnAAnnATnACACCnCnTAnnnnTnnnnn
nTnnAnnnACnnnnnnGCGnnnnnnnnTTnATGnGnnGGTAGnAnnnnGTnnCAnnATTnn
nnnGAnnTnGAAnGTnAAAnnTTTnnTGTnATnnTnTGnAnnAnTAAnnAnAnnAnnACG
TnGTnnnnnTAAAnAnAAnAACTnAnnnnTTnAGnTGAnAnGAnnAACTACTnnnCnAnT
TnnTAATGTTGTAnnAnTAAAnAnnnnnnTnAnGnnTnnnnAAAnnnnAAAnnATTnnTT
GAnGGTGnTAATnnAAnTAAnAnTTGnnTnCnGTTTTnTTnnTnnTnTnTTCTnTTnnnT
ATTnnTTnTnCTnnTnTTnTCTAAAnAnnnnnACTTTTCnAnTTnAnTnATAAnAnnnAn
GnCTnTTGAnGnnnnAnnnnnTnnTnACATAnnATnTAnnGTnACnTGnTTnGCnAAnnA
nCATnnnnAnnTTGnnnnnTnGnTAAnCnAGCnnGnTnnTAGnTATnnTAnTGACAAnGn
nnnnnnAGTnATTnnnnnAnTCAnAAnnnnnGAnnnnnGnTTTnTGnTGnCnnnTTnGnn
GnGGnnGnTnTTAnGCAnnACTnnTGGnGnnTnnTnnCATTTCnnACnTATAGTnTTTAn
TnnAnTTnGTAACnTCnnTTAnnnAnTATnnnAnnTnATAnCnTnnACTnACnnTGnAAn
ATnAnnnTGTnnTTTGnCTnCTnAATATnnnATTnTTAAnTnTnnnnnTnnTAAnCnAnT
ATnATATnnGnnTGTTnnnAnTnnACAnnAnnnTnnTGTTnnnnnTnAAAnnTTAnnnnn
TGnnnCnCnTnATATGnTnnTnGAnGnCnCTATnnTTnAnTTnnCnnnnACCnAnCTTnT
AnGTTnTnTTnTAnTnnnAACnACTTTTGATnTTnnGTAnTGnnGGCAnnnnnnTTGTnA
nAGnTCAGAAGnnGGnGnTTnTGTnTCTAnTnnTnnnAnGnnGGTnnTTnAnnATnAAnn

nGnnnnAGnTTnnnnAnnAnTnATnAGTGGTGnAGATnnTnTnAAnTTnnTTnCTnnTnT
GTTnAnnCnAnTAATTnAACnTATnnGTTnTTTnnACnTATnAnnnnnTATAGTnGnTnG
TnGnAnTnTnnATnTnnTATnnnCnTnnnTnGTnTAnTAnTTnnTGnnGTnTnnnnnnnC
TnnTnnTnTATnnnGTnnTGTAGnTGnnTTTnnnAnTTTAnTATTCnTTATGTCnTTnAn
nGTACTCTGTTnnnCATnAnnTTAnTCATTCTTnnCTGnTnnTTAnTnTnTTAnTTATTn
nTACnTGnTATTTnATCTTTnTAnTnnTnnTTCTTTTTTnnTnnTTATTCnGTnGATGnT
TATnTTnnAnCnTTTAnTnnnTTTnTGGTTAnCAnnnnnTTATTnCnTTTnTnnTTnnnn
AnAGCATnTnTnATnnTTnTTTnnTCATnACnTAAnTAnnnGTGnnnTnTTnAnTnnTnn
TTnnTnTnnnnnnTTTnnAnAAGnnGnGCTGTGnnCnnTTTTGTTnnnnAAnnnAnnnTn
TCnAAnnTTTnGTnnnGATGnGTnATTnnnTnnTAnnnnAnAnAnnAGATAnTAnnnnCT
nnnnAATAAnTAnAAnnATTTTnnnnnnnnAATAnATACAACTAGCTTnnnAGAAGATnn
nTGnTnnCAnCTnGCAAAGGnnCTnnnTGnTnTCAnnAnCTCAnGnnnnGnTGTnnTTTn
nCAAnCACCACAnATnnCnATCnTCTCnnCnnTCnnnCAnAnTGnTTTnnnAnAnCTGnn
ATTCCnAnnnGnTATAGTnGAAAGTTGTnTGnTnCAnGTAAnnTGnnGnAnAnnTnnAnn
nAACnnTCTTnTGGTnGAnnAnnTAnTnTnnnnnnTAnGnnnnnTGATnTnnnnnTCTGn
nGACnnGnTnAAnnCTnnTnnTnAAnAnnnAnnCTTTnnnAnnTnTAATnAnAATTTCTn
nnTAnTnGnTnGTAnTnTnnAAnTnnGGnnTAnnnGnCnTnnTATGnAnAAnTGnGTAnT
nnAGnTTAAnGTTnAnAnAGnnnAnnnTAnnAnAnnnAnGTnnAAGTTnnTnCTCATnCA
ACnAnGAnnnAnTTTTTnnGTnTTAnnTnnTnACAAnnGTTnACnnTnTnGTnTTnnnnn
ATnnGCTnTnnnnCnnnTnTTnACTATTAAGGnTTnATnCCTTAnTnnTTCATnAnGnAG
TGTTnnTTTnAnnATnGATnATGAnTnnGTnTTTTTTTnTTAnATGnTnnnTATGGnATT
TCnnATTGnAnnTnnnnCTnGnAnnnACTnAnAAGnnAACTTnnATGnACCnTTTnnTnA
nAGGnnAACnGCnnAnGnnnnTnTnACGGnCAnAAnnCATAnnnnTnAnGTTnTAnCnnn
nnTnTnnnnnGnnnTnATnnATnnnGnnnGGTGGnTGCTnnATnnATTnACnACAACTCn
TnnTTnCTTnAACnTTGTAGCTnTnnnnTTnAATTnTnnAnCnnTnACACnAnnnAnTAT
TnTnATnCTAnnACCTCnnACnnCnnAnACnnnAnAnGnCGTnTTnnnTnTGnnnnnnnn
ATTAnnnnnnTTnnTGnnAnATGnTnTnnATnAAnnTAnCnnnTTnnnTAnTGnnTTATT
nGAAGnAnAATnTnnAnTTnTTnATnnTnnTAGnnAATnnTnAnnTnTTAnGTTnnAAnn
TGAAGTTnAnAnnnTAnTnnAnnnTAnACAGCACTGAnnnnTACnnACAATnnTnACnnC
nnnnnnAnnTTTAnnnCAnAGTnnTCAnTnnTCTTTGTTCTnnTTTnTnnnTnAnAATGn
nTTTTnAnCTTTTnnTATGnnnATTATTnCTTTnTnTnTTTTnnnAnTGnTnTTnnTnnn
ACnTnnGCnTnnATnnnTnTGTTTGTTnTTnnTAnnTnnTCTTTCCnCTGTnnnTTnTTn
TnATnTnnTCTnTnTGCnnnnTATTTTnnTGATGCGTATTATnnnnTnGnTnGnTnnGnT
nGnnnnTnGATnnnnnnGTTTGAnGTTnAnAnnnnnTGTTnTGTnTGnnTnAnnnGTAnn
GTTAnTnnTnnnTATGAnnnnnnnnnnnTnnTAnnATnAnnGnnnnnGGnnAGTGTnGAn
AnTnATGAnnGTnnTGnGAnTnGTTTATnAAnTnTnTTnTnGTnnnnnnTTAnnTnTAnn
nATTnnCATGTTnnnTCTTATnATCTnTnTTnnTTCTnACTnCTnAnnTnTAnTTnnnnn
TnTCATnTnTnTGGnnnTAnnTnTnnnTTTTnnnTGTTTTGnGnAnTnnnCnTTTTTCTT
nnTnnnTGnTAnTAnnnTTCnnTnTTTAnTGnTAGTnnATTGTTTCTTAnGnTnnTTnTG
TnnTTTTTnnTTTnnnnTnTTTTTTTTAnTnnnTnnnTnCTTTnnnnTTnnTnTTnnTnn
TTnTGnTTnCTTAGTnTnTAnnnTGGAGTTTAnnTnTATGnAnTnnCAnnGnnTAnnnnn
AnnnAnnAATAGnAnAnAnGACTTCnnACTCAAnAnTAAATnGTnGnnTnTTnGnnnCAn
AnCTTGTAnnnnnGTnGnnACnGTACAnTnTAnAnTGTCAGATGTAAAnTGCAnAnCAnT
AnTCTnACTnTnnnnTnTnCAACAnCnnAnnTnAnAnTCATnnTnTAAnTnnTnnGCTnA
nTnTnnnCAnTTnCAnAATnnCAnTCTCnnnGCTAAnnnTAnnACnnAnGCCTTnnnnnA
nATGnTnnnACTACTTTCTnTTTTGnnTTnnnTGnAnnnTGnnGTnnnCTTnAnCnAnCn
nTGnGnnGTnnTGnTGnAnAACAGnGnnnnnnTnnAAnnnAnAnCCTCAGAGnTTAGTTC

CnnnnnAnCATAnnCnnnTTTTnCnnTTnCnCAnnAnnCTTnnGAGCTnnnnnTnnnTAA
nnnnnATTnTnnAnTnnTnCTTAnnnnGnTGAnGAAGnnGTTGnTnnTnGnTAAAnGTnA
ATTTnnnCTTnnnnnnnnnATGnAnCnTAAGTTGnnnAAGATGnCnnATnnnGnnnTGAn
CCAAAnnTnnAAnCTGnnnAGAnnnnAGnACAAGAGnnnAAAAGnnnCTAGTnCnAnGnn
GAnAAnGCAnnTnAnnnTnCTnnGAAAnTnGnAnnnTGnTnCACTCnnCAAnnnnATnAn
nATTnnnAnAnAnnnTTGAnTnnCnTTGAACATAAnACnnnnnACnACnnnnnCCAAAnn
AnTnGTTnTCnTACnACAnTAnAACAnATAnnAAnAnnCGTnnnAnGnTAnAACATTTAn
nnATnnAnnAGnAnTGnnnnAAnTnCnACAGnTnnnAGATGnTnnTAGnAAAAnTnnnnA
nTTTnnnnAAnTnAnAATGnAnnAnTnACnnAAnnTAGnATGGnnnnTTATTnTAAnAnn
TnTAAGGnnCAnnTnnnnnnTnnAATnnCAnnAnAnnnnGCTTnnnCCTnnnGCnnTAAG
AnnGATGTCnTnTGTnnnnnnnnnnAnnnAAAnTnnTnnnnCnGnnGACnnAnnnCTAnn
nnnnnnnnnnACAAnAAAnnnnnGTAnGnTTnnAnnTnCACnnTnAnCnGATnTAnAnTA
nTnnAnATnnGnTAGnTTCnnTAAGAGnnATGGAAnTGGTnCnnnnTnnnnAGAAnTGnA
AnnACCnnnnAnnTTTnTnAnnnACnnACnTnAnGnTnnnAAnnTGAAnnAnnTAnACnT
TATnAnAnnATnAAnnAAnATAnATAnAnnnnTGnnAnTTnnTAnTTTAnnnGnnAnnGT
AnGnnTAnAAGnTGnTAAnnnAnCAnAnGTnnnTnnCAAnTnAnnnnTnnnATnnnTnTG
nnnTnTTnnTnTAnATnnTnnTnnnnCTTAnAnnTAnTATCTAnTnnnTnnnnnnCnnnC
AATnAnTAATTnnGTTAAGAnGTAnnnnnnACnnnTnGnTAnnnGnnAnnnAnTAAnAnT
TACACnGGnAnnCAnnAnGnAnCAAGAATCnnnnnGnnnTnCATnGTGnAnnCTGTAnTn
nnTnnnnCAnATnGATCATnnAnnTnCTnAnGGnTnnTnnnACTTnAnAnnTAAnnATnn
nnAnATnnCTACnACnnnTGnTAATGACnnnnnnnnnnnnnnnnnTnnnnnnnAnAnnCnn
nnnnGTnTnnnGTATnnnGAAAGnTTnTGGnTnTnnTnnTGAnnAnnTnnnnGnTnCnGT
nnTTnnnnnnnCnGAnGCnnAATnnTTnTTnAACnGGnTnnnGnTGTAnnTnCnnCnnGn
nTTnCnnCGTnCGnnnnAGnCnnnAnnnnnGATnnCGnAnAnnGnGCnTTnGnCnTnTnn
AATnAnnAAnTAGnnnnTTTTnCTAAnnTCnTAAAAAnGAATTnTTnTnTnTTCnAnnAA
AAAGAnGnAGnTGnnnnTTnAAnnnAnAnTnnnnTnnTAnTTAAnAnnnnnAnTATnnnT
AAnTAnnAAnAnnnAnnAACAnTTTATAnTnTAnTnAAGAATTAnCnAnnTnTTAnnAAn
nAnnAnnTnTTTAnGnTTnnAATAnAnAnTnnnAnGnTACnAnnTATAnCAnnTnnnnnn
nnnnnnAAATACAnAATnGnCGACnTCGTCTnTnCTnnAAGnCAnTnnnAnnnnnnTAnT
nnnnACAnATTnAnnnAnATAnTTGTnACnTnnAAnnnTTAnGAnnAnGAnTATTnnAAT
AnAAAnGAnnnGnnnnTTnTTnTAnAAAAnAnAnnTATnnTnnnnnTnTAnnnnAAnnnn
nnnnAAnnTGnnnnnnAAnCnnnnTnnnAnAnnnTAnAnTnCnnTnATnnnAnnnGAAAn
GCTnnTATTnTTGnTnTACnGAnAnTAGnnnTnnTAnnnnTCnAnnnTnAnTnGTAnnnn
nTCnnnnnTTTnATAnnAACnAnGAnnnnTAnTGnAnTTnCTnnnnnAnATnnTnAnTnn
TnnTTGTTAnTGnnTATnnTnAnTnTTAnnnnnnnTTTAnnTnnAnAnTnACnTGnnnnn
nnTnAnTnAAnAnAGCCTTACATnAnGnnGnAnnnnTnAAAATnTnAnnTnAnnnnnGAG
AGnTnAAAAnTnTTTnnnCnTTGnnnTAAATATTnnnATnnGAnnnnCnAnnnAAATTnT
nTTAnCTGTnnnnATnACAnAnGnATTnTGCATTnTGCAAAnnTTAAnGTTTnATTCTnT
AnnGTGTnnnCAnTTAnnAnTTTTnnACnACnAnnnnnnAAAnTATTnGTTnATGnnnnn
TnATnAGTAnTTTnnAnTGnnTnCCAnTTCnnAnAGCTAnnTGTTGTnnATAnTnAGnAT
GnAAnCTTnnAnAnnTCnAnAnTnAGnnTTnAGnAnTTAnTnGTnnATGnTGnTGnCnnn
nnTATnnACTnTGnTnTnnGnAnTCTnTTACTAnnTAnnCnnnCTACGnnCTTTTCnnTn
nnnGCAnTnnCTnnnnATGTTnnTTTnCAnnnnnTCAnACCnnnnAnTTTnnACnnAnAA
nTnTnnnACnTTGTTnTnTnnAAnnnTTTnTTTAnGnnnnnAnGTTnTnTTGnnTTnTnA
nAnTnnnTnnTTGTTCAGnATnnTAATnCTGCTnTnnGCnATnAnGAnTTnTnnCnnTTn
nnnnTAnCnACnATGnnTnATATCAnAnAACTnCnATnTGTAGTTGnAnTTnnnGATAAn
TACTTTnTnTnTTAnnAnnnnnGnTGTATTAnTGnTAnnnnnGTnATnGTnnAnAnnCTA

nAnAnATCAGCTGnnTnnnCATTTAAnAAnTnnnnTnAGGnnnnnnTnTATTATGAnnnA
nTGnnnnATGAnGATnnnGnTnnAnTTTTnGnATnTAnnAnnnnTAAnnTnnTCnnTAnT
ATAAnnCAAATnnnTnTnnAnTnnnnAnTnAnnnnAAAnAnnAnAnnnCGnACCnTAnnn
GGnnTnnnTATnnnnnnTnCnnTGAnnnnTnnnCAnTTnnnTCnAnnATTATnnAnAnnA
nTAnnCnnCAnTAnAnnAnnTAnTnnAnnAnnTnnAnCAnGnAAAnTnTAnnnnnnnTnG
nAnnAnAnAnnAAAnACTGTnGAnAGTGnnnnAnnAAAnnnnnAnCTTATnnnnTnGnAn
TAnCnnAnATnnGnnAnAGnnATGnnnAACATnCTTAnnnTTATGGnnTCACTnnTTnTT
nnnnnnnnAAnnnnCAACGTnnTnnnTCTTGTnACnnCnnnTnTATnGnnTAGCTAATnnn
nnnnCTnnAGTATTnnGTnAnATGGTnATnnGTnnCnnTTnAnTATnTnnTnnAnCAGGT
GnAnCnTnnTnnnnAnnnGnCAnnnnTGnnnATnCTnATAnnGTnnTnnAnnTTnnTCAn
nnnnTnAnnGnnnnTnTTAnTGnnnTnTTnTCTnCTnnnnnTAnCAAAATTnnCnATnAn
nnnGTnnnnAATTTACAACTnAnAATTnnTGnGTnTnTCnATAnAAAnAnAGAnnTTGnC
AnnGACTTTGTAnATnnGnTTTnnGCnTnTnTGnnTAnAnnTTTnnCAnTGATGATnnTC
TnTGnnnATnCTnnTnnnTGnTnnnAnAnnACnnAnnnnTnnnnnnGTnTAnTGnCTAnn
nTnAAGnAnnTnAAGTnnnTTnTnnATTATnAAAnnAATnTTnTTnTGTnTnnnAnAAAA
AGnTnGATnTnnnnTnnnCTnAnTAAAGnnnnTnnTnnnnTnTnnnCTnnACAnAnnATG
CTnnTTAATCnnGnTnAnnAnTnnnnnTACnnTnnnnACnCnGAnCnnTnAAnnATCnTn
nnnnCnnGCnnTnTTnnAGAnGATAnCnnnnAnnCnGATGnTACAnTnATGATnGAAnnn
TTnnTGnTTnTAGnnTTAnAnGnTnACnCACTTnnTAnACAnCnnAATnnGnAnTnnGnn
nnnnTCnTnCTTnTnnnCTTnCnAnTnATnAnAAAnnTnAATnATnAGTnAACAGnAnAC
AnnnnnnACATGTnnTnTGAnATGnnTACTAnTnATAAnnnTTnAAnGTnnTGGnTACTn
GnGnTnTATGnnGnTATGnnCAnnnnnCATAnnGTnTTAnGGnnTnnTGGnnnTnGnnTT
CTTTGnAATTnGCAGAnnnCnTTnnnnTnTnnTGCnnGnnTnCGTnnnCnnTTnnTnnnT
nnTnnATnnTnnnAnGACCATnTTnTATnnnnAnnnCATAnnnnAnTCTTGTCTnnTnAT
CnnTATnTnTnnTATnCTCnAnnnTnnnnTnTnACAnnTGTGAnnCnnCTTnnnTTnnnA
nnTATGAGnTATnAnnnTTnATCnnAnAAAnCAnnnATTAnTnTTnnATTnTGnGnTAAn
nGACAAnTTTTTnGTTTnTATAnAAnnAnnTnTnTTnTTnnnGnTAATGTnAATGnnTTT
nTTGnnATnGCAAnATnnnAnTnGACAnnnGnnnnnnATnAnATTTTAGCTnAnAnnnnn
AnTGnAAnAnnnnAGnTnnTnGCAGCnGnAAnGnnCAAAnCTACTnAGnAGAnAnnnAAA
CnnTCnnAnnnnATnnnnAnTnnAnnnnAnnTGCTGnnTGAnnTAnAATTnnAnnTnTCA
TGGnnAnTnGnTAAnnnTnGnnnnCnnCnTAAnnnnAAnnAnnTCnTTnnTnnnTAnnnn
GTAAnTAAnAACnnTnnAnTnnAAATnGnAnAnTAnAnnnTTGAAAAnTAnnAnTAnGGT
nnTnCnnnTnTnTAnnnnGnTACAAnGnnnnACAAnTTAAATGTTGnnGATnATTTTnTn
nTnAnnTnACAnAnnGTAATnnnnTTAAGnnnACnTAnnnTAnnGnnAnAAGAGCACnnn
nnnnnAnnnnnnnGCnnnTACCnnACACTnAAnnTnTnAnATGAnTnTTCnnnnnAnnTn
nCAAATnAnnAAAAnnTTnnAnTGCAnAAGnATnnTAnAnTnnnGnGnCnAnnTAnTAnT
nGTnAnAGnnAnnnnGnnnTTnnnCTAnCnnTnnnCTACCTTTTnnCnnnCAnnGTGnAT
ACnnCTTnCnCTCnnGnCGCTnnTnnTnnACTnnnTGnnAnGnnnnTAnnGTATnTGCnT
ATnGnnAAATGnAnTAnAAnTATACnnnnAnnTnnnnnTGTAGAnTnTTTnnAnAAnnTC
nAAGTGAATTnnAnnTTAnAnnnnnATnnnnTTTGTAnTnTAAATnnnAnGnnTnnGnCn
nnAGnAnAnATAnTnnnnTTTGnnnnAnnnnnAATGnnnnCnnnnTnTnAnTTGAnAnTT
GnnnnTnCnAnATTnnnTGnTnnGnnnTAnGTnnnnAnTGnnGACnnnnCTnnnTTnnCT
nCnnCAnnCACnnTGCTnAnTAAnnnCnnnCnnGAnCCAnAATAnTTnAnnnCnGnnTnn
AGnnnnATGAAAACTAnAnGTCnAGAnnTnTTCnTAnGAAnnTGTnnGnnnTGTnCnnCT
nnAnnnGTTGACnnTGTnnnTGnnnTGnTnTAnGnTnnTAnGCTnnnAGCnCnTAnnnnn
ATAnnAnnnnnnTGCTTTAAnATnTnTnATAAnnnnnTTATnACnnATnATGTTTnATnT
GCnnTnnnCAGnCnAnTnATnnnnnTGnTAnGAGnAnnCCnTAAnCnTnAnnCnGCnTnn

AnnAAAnCnGTnnTnnTnnCnCnTTnTAAnTnACAGAnnnCTnTAnnCnnAAnGATnnTn
GnAnTACnnACTCAAACTnnTnATTnnTnnnAnnnCnnAGnATAnnnnTATnTCnnAnTC
ACTCAnAnnACTnAnACnnCnnAnTnTTnnAnnnTAAAnnnATATnAnnTTnCTATTnnn
AnAnCAAAAGTAnnnATAnTnTnnATAnTGnAnnAnAnnnAnnTTTATGAnnAnTnGCAA
nTnACAAnTnTnnAAATTnCAAGTnGnAATnTnGnAAnTTTnnAAnnTGAAAAnGTAAnA
nGACTCnTTAAnnAnnnnAGnAAGGTAnnCACTnnGTTAnATnnTAnAnAnnCAnCAnCA
CnCnTCAnnnTTnAnAnnAnAnTCAAAACnnnAnnnTTATnnnTTnAnAnAnCAnnnATn
nnTnAnGACATGAnnTAnAnAnnAnnnATCTCnAnGATnnGTnTTAAAnTnAAnTnTnAA
nTnAnnGGTTAnnnAAAnATnTTTATCAnnnTnGAnGnAnCnATAAGATATnTAnnTnCn
nnGAnnnnnTTCnAnGTCGnGnnGnGnCATGnnnCTAnnnAnnnGGnTGGTnnnnnnTTA
CnnTTAnAGnTnGnnTTnnnTnCAGnTnnTAnCnnnGTTGCTGTTCnTnnAnnnTnnnTn
nAnnnnCnnnAnnAnnnAGATTTTTnnnTAnTnnnnnnnAAAnnACnGnCTnnnGAnnAA
TTnAAACACCnCnnACnnCTnATGnnCAAnGnACnnCCnTnGnAnnTAGTnnGTAnAAAG
nTTnTACnAnnGTTAAnnGACAnAnTnAAAAATTnCTnTGnnnnAGTnnTATnnnTnTTA
TnnnCnCnnGGCTnnnAGTTnTCATCTnnnAAGTATTTTGTGnnAAnTGGACnTnnGnGn
AnnTGTTGnCnATnnnAnAGnCGnnCCnCATnCTTTnnCAnTGnnnCAGACnnTnnTnnT
TnTTnnCnTCATnCTnnTGnATTTnnTnnnGTCnnTAnTCTnTTnnTGnTTGnTnTTnnA
CnATnnnnTTTTnnAnGTnACnTnnAnAGCAAnnnnGATCTGTAnTnTnAAGnCnAnGGn
AnnnnnnnTGTnAnTnnnnnnnnnnnTnnTnATGAnnAnGTnTCTAGTTGTnCACnAnTnA
TTTGTTnnGnGTnnnGnnTGnACTATTGnnTAnnnnATAATTGGTGATGnACTnAnGATT
AAnnnnGnnTnnAGnAAGnTnnAAnAnnTGnTnnTTnAAGnTGnATTnTTnnnnnnnnnA
TTnCnAnTTCTTCAnnAnnnnnnnAAnCnTnAAnCTnTnAnGTnnGTAnCnCAAnCnnnn
nTAnnATGGAAnnTCTAnnAnnnACAGnnTTnTAGTGnnAAnnCTnAnAAnATnGnAGAA
TTAnTnnnTnnTnATnnnAnAnATTnTnAnnAnTTnAnAnATnnnnTATnCnTAnTTTnn
AATTnnnAnGTnGAnnnnTATnnTnnTAnnnCnnnTGTnTnTAGAnTTnTCAnTAnAnTG
CTATnTnAnnTnnAnTTGCCnnnTTnTnnTGnTnGnAGnTTnnATnTnnnTnTnnnnGCA
nTnnnCAnnnnAnnnTTnnnnAAnAGTnCTTnnnnnnATTTnAAAnAAnTAnCATTTTTC
TTTTnnnnTnAnAnTCnATnTnAnTTTnnTnnnAnACAAnnAGTnnCnGnTATnnAnTAn
GnAnAnnnnAAGTnTnnnAnnTnTATnAnnCATnnCAAnTTAnnTnTTGnTGTCTnnnGA
CTTnAnnCTnAnGnGTAnnnATTnnATnTCGTnnTnnnTAnnATnAnnATCTnnnnTnnn
nTTAnnnTnTGnnnnTnnnnnCnnnnnGCTnnTTnTnnnnTnTnnAAnAnTnTTnnnAGA
CTTnnGAnTTnAGAAAnTnTnGnnTnTnATGTTnTAnATAAGnnAnAnTTnnnTnTAnAA
nAGnnnGAAGnnnnAnTTTnnATCATnAAnAAnAnnnnnTACnnnnAAnTnnAAnnnnTn
nAnnTAnnATnnTTnAAnATTAnnAnnACAnTnnATnTTAATGTAnnAnTTnAnCATTGn
GnnATGnnnAAnnnTAAAnnnnTAnnAGnGGTnAAAnTAnnCAATAAnCTGnGnGnGnAn
nnnnnnnnTAAnAnnGTGnnnTGGGAnTnnATAATAnAnnnnCnAGCAnnTAnnACTACT
ATnGnnnTTnGnnnTATGAnnnAnATAnnnnAGnnnCnAnnTnnAAnGnTnnnnnnACCA
CTCAnnnTnnTnTnTGnTnnTnnnnnnnAnnnTCnATTnnnnTTnnTnnnAAAnnnCnnT
nAnnnTnTTnnTnTTAnAGAAGnnAGnGTnAAnGnnTTAnnACnAnnnGTnGTnnnnAnA
nAnGnTAnTCTnnATGnnnnnnnnnnAAnnnnAnAAGnCGTAAnAnnnCAnTTnnAnTAT
nAnAAGAAAnAnGAnnGTnTTnTnnnAnAAnTAnnnAAnACnTnnTTnAnnnAnAATAnn
AAnnTnnAAnnnnTTAnACnCAGAAnnCAAATGnAAATTnATTTnnnAGnAnTAGnnATn
nAnnAATTnATTnAnnAGTATAAATTAnAAnnnTATnCCTTnGAAnnTATnAnTTATGnn
GnTTTTAnTnATAnnnAGTTnGnTGnTTTnnnTnTAnTnATTGnnnTnGnTnnAnnTTTn
AnnnAnTnAnnnTTnnAnnTAnAAGnTTTnAnnCnTnTnnACnnTAnAnnTAAnAACTAn
TTCATAnnAGAnGnnCAAAnAnnTTCnTCTnAGTnnnTGnnnnnTGTTnTTGATTTnnTA
nTTGATGnTTTTnTnnTAATnATnAAnTnCnTnnTTTTnTnTnTAnAnTnTAAnnTTnTC

AAnnTnnnTnTTnAnTnTAnAnnAnTnTCATTnTTnCTTTnGTnTAAAGATnnnCATnTn
nnAAnnTTTnAnnnAAAAnTnnAATnTAnTCAAnnGTnnnAACnGGnTnnTGnTnTnnCn
AATCnTTACAAnnTGCAAAnnnTnCTAnTAnnnAAnTGnnAnnTnnAAAATTATnATnAT
AGnnnAAnnTnAnnTAAAnnnATAATGATGnATGACnnAAAAnAnAnnnnAnnnTGTnnA
nnnTnAAAnAnAnTAAnAnTnnnTGTnnnCTnTAAnAnnnnAnATnTnnATTTTnnTnCT
nGnTnnTAnnAATnAnTTGTACnnGGTAnnGTTnTnnTAAnnCAGTnGnTnCnnnnGnnT
nnGnTGnnnGnCGnTTnAnnnCTnnnnGnCTTnnnCTnTnnTnnnnAnTnnAnTTTGATT
nTTGATnnnnnAAnTnTnCATnnnnnnATnAnnnnnGAnCTnnTnATTAnnnATATGnAn
nAnnCnAnnnCnnAAAnnGTTAnAAnnnnAAAnnnAnnnAnAnAnnnnTTnnnnAnnnAn
ATnnnTnnGnTTATnnnAnAAAnnCTAGCnnTTGGAGGTTnCnTnnnnAnAAAGnTnnnA
GAnnATnnTnGGnAnnnnnAAnnnTATAAGnTCATnGnACnnnTnnnnnTnTGnAnAGnC
TnnnnnACTAnTGnGnATGnnnnATnnTCTnnAnnATTnnTAATTnGnnGTnATnATCTT
nnnAAnnnnCAnGAnnAAATAnnTGnTTAnGTCATnCATGnAAAnTAnATAnTTnnnAnG
AnnAnAAAnCnAATTnAnTnnTnTnnnnAnnnTnTATTTnACATnAGTnnnnTTCCnCTT
nAnTnAAGGnnTAnTGnTnTTTTnTCTTTAAAAGAAGGnCnAATnnnTnATnTnATTTTA
nnnnTTnTnAnTAnAnnnAnAnTTnTnATTnnAnAAAAnAAnAnAGTnGTnTTnTnTnnT
GAnnnnnnnGTTnAnTTnTAAAnnAnnAAnGTnTnnTTnTCTnnTnTTATnnnnACTATT
CTnTnTTnnnTnTnTTnATnnnnnnTCnnnnAnTCAnTTnnnCnTnnCATnAnnnAnTnn
TTTCAnnCnTCnTnTTnATnnCCnTnnnCAAnTnnTCAnnTTnTnnnTTnTACATTCAAC
nnnnGAnnnnTTnTTnnnTTTCnTTnnnAnTnnTACTTGnTTCCATnnTnnAnnTnTnTC
TnGnnnCnnTnTTATnAnGTnnTTTGnTnnnnnnnTnCTAnnATTTAnTGnTnnnnTnTn
nTTTnCTTnCAnTnTGAnGTnTAACnTnATnnnAnnnnnGATTnTTGGTnnTACTTnAGA
TTCGAnnnnnnAGTnnCTACTnATTnnTAnTnnnnnTnCTAnTnnnnTTTTnAnAnTnnn
nnnAnTnCnATTnTTnATTnnTnnATTTTTGnnTnTTTATTnnnTnnnAAACnAnAAAnG
TnnGATGnnnAnTGAnnnnAnAGTTnATnnTnGnnnGAAnAnnnGnnCTTTTnnnTnTGT
nnnnCnnnnnTTTnTTnTnnAnnnTnnAnnnAnACnnnnTAnTTTnnAAAATnTnAnGGA
ATTnGTnnTTAnGAAnATnGAnGGTTAnTTTTnAATATnnTnnAnGnTnAnGnnTATTAA
nTTAnTGnnTnnTnnCCCnCAnnnTTTTTnGnnTTTnGnnCnnTTGnTnnnTTTnCnAnT
nnnTAATAACnnnnnTnGGTTTTAnnnTTTAnTnnCnTTAnnnnnnnnnnnnnnnnnnnn
nnnnnnTTnnnnnTnnnnnTnnnnnnnnnGnTnnTnnnnnTnnTnAnGTnGGnnATCTnCA
nCCTnnGACTTTTnTnTTnnnnnATAATGAAnnTGGAnnnnTnAnnnATnnTnTAnnnTG
nnCACTnnAnnnnCTnTnnnAnnnAAAnTnnACGTTGAnnTnCTnnAnnnnAnAAAnAGG
AAnnTATCAAnCTTCnAAnTnnAGAnnnnAAnCAACAnAAnCTATnGTTnnATnnnCnAA
TnnnACAAAnTTGTnnnnTnTTnnnnnAnTATTnAAnGCTnnCnGATTTnnnnnTnTTTn
TnCnTnGnnCnnGnAGnnnnTnAnnAnnnnnGTTGCnnnTTnnTCAGTnnTATnTAnTTn
nnnAnnnnTTnCCACTTTnAAnTnTnATGnnnTnTCTCCTnCnAnATTAAATGATnTnTG
nTTTnTnAnTnTnTATnnAnnTTCnTTTGTAATTAnAAnTnATnAnGnnAnACAAnTnnC
TnnAGnGnAnACnGnAAAnnTnGnnGATnnnAAnTAnAAAnTAnnAnnnnnnnTTAnAnT
nnnnATTATnnnnnnGAnTTnTnAnAATnnTGAnTnTnAGGTTGGTGnTAnnTnnAAnTT
nnnnTnTAGATTGTTTATnATTnnTAATnTCAAAnnTTTTGnnAGAGnTATTTnAAnTnA
AATCTnTnAGnnnGnnAnnAnnCnnTnnAATnnTnTnnnAGnTnTnAAnTnTnAnnTnnn
nnTnnnnTnnnnTGTTnTnnnnCCnnnTAATnATnTTGnTTnnnnnCCATACAnAnTAnT
GGTnnTTnnTnnTnAACTTnnAnATGnnnCnnnnnnTnTTTnnnnnnnnnnAnAAGnnnnn
nnATTTGGTTnTnnACAAAnnAnTnATTTTnAAnTTnAAAnnTnTAAnAGnnACAnnTnn
nnTTACTGAGTnTnAnnAAAAGTTnnTGnTTTTnCnnnnAnTAGGnAnAnnCnTTnnTnA
CACnAnTnATGnnGnnnnTnnnCCnCnGACACTnGAGATnCTnGACATTAnACnnTGnnC
nTTnGGnnTnnTnAGnnTTnTnAnACTnGnnACnnnnnnnTCTnAnCAnnnnGCnGTTnn

nnATCTnnnTnTnAnCTGnnnAnTAnTCnnTATTnCnATTCnnGCnGATCAnCTnnCnCn
nnnTTnnCnnnTnnAnTnnAnAnTnTTTAnTGTTTTTnnnnTACnTnnAnnnnGTnTAAT
nnnnnnTnAACATGTnAAnnnTTnATAnnAGnGnnAnATnnCnAnTnnTnCnnGTATnnG
nGCnAnnTAnCAGAnnnnGnnTnAnTnTCCnCTnnnnnnnnGnnGnGTAnTnAGTCnCnn
CnnCATnGnCnAnnnnnTGnnACTTnnTnCAGnAnnTnCAGTnnnTTnnTnnnnnAACnn
TATnnnCnTnnnnnCAAnTnTTACTATTnnnnTTnCTnCAnAAATTCTAnnAnTnTnnnT
GnnnAAnAnATnAnTnnATTGTAnAnTGTAnAnTnGnnATnATTnAAnTGnAnGnAGnAA
nnTTnTGnTGCAATATGnnAnTTTnnnTAnnnnnTnnAACnnTGCnTTAnnTnnnATAnn
TnTTGAAnnnGnnAnAnAnnnnnAAnAnnTnTTnnnAnAAnnAAAACnAATnnAnAAnnn
ACnACCAAnnAAAAATTTTnnTnnTTnnnAnTTnTnAnAAATAnTAnnnGATnCATnAAA
nnnnAACAAnnnGTnAnTTATTnAnnATnnACnAnnnAnnnAAnTnAnAnTnGnAnnTnn
TnnnTnnAnnAAnCAATAnnnTnAnnnnCTnGnTnAnATnnnnnnTAnnGnCnTnnTnTn
nGnnnTAnnGnTTAACGnnnTnnnTGTTnTGCnAnnTnnnnTCACnGnnnAAATGnTTnn
TCAnTnCnCnTCnnnnnTGTTAnnGnnTACAATnACnnnTnGTTnGnnnTTTGGnGnAnn
nnCTnnnTTACnnnTACCATTTGCnATGCnnnTGGCTTAnAnnTTTnATGnnnTTnGnnn
TnnACAnAnnGTTnTnTnTGnGnAnnAAnAATTnATnnTnAACnnnnTnAAnAnTnnnAT
TnGnAnnAnnnAAnAnTnACTATTnTCnACAGCAAnnnnAnTTnnnAAnCTTCAAnAnnn
GnTnnACCAnnnnnnnnnnnnnnnTAAnCACGnnAnTnAAAnAnCTTAnnTCnAAnnTnnG
nnnAATnTnAAnTGTTnTAAnTnTTATnnTTTnACnTCTTGnnTnnnTnnAGnnTnAAGT
nCAAATTGnnAnGTnGnTnAnnGnCAnAnnnCAAAnTTTnnnGnnATnnGAGnCnnAAnA
nTnAATTAGnnnTnnAnnnAnCAnnnATTnTnnnnAnCnnnnnnCnACTnAAnTGTnAnT
GnnnnTACTnGnACnATnAAAAAGnGTTnATTTTTnnGnnAAnnAnnAnnAnnTnTTGTn
nTTnnnnCAnnCnGCAnTnnnTGnTnnAnnCTnCnTGCATnTGnnTnnTGTCnCnTCnCn
nnAnAnnAnnTTnnnnACTnCnCnnnnCAnnnGTnnTnnTnnAnAAnnnCACnTnCCTnn
TGAnnnnnTnnnnnnTTCnAnTnTCAnnnAnnnGTTnnTnnnACAnAnGAAnnnTnATGn
nnnACAAATCAnnAnTAnAnnnAACAnAnAnnnnTCAnnTnnnnnnnAnnnTnnAnnAnG
AATTnTnAnnAnnnnnnTnTAnnATnnnTnnnnAnCnnnATnAnAnTTAnTnAnnnnnnA
nnnnnATAAnTATTnTAAnAnTnnnAnATCAnnAnATnTTnnTTTAAGnnACATnTnTnn
CATnAnTnnTTnAnnTGTnAnnnTTCnAAAAnnAATTnnnCnnnTnAAnnnnnTnnnnAA
GAAnTTAAnnGnATCnnTnATnnnnCnnnnAnAACTnnGnAAnTnnnnnnnGTAnAnnAA
ATnnnnATnGnnCATnnGnCTAnnnTnTAnnGnTnGnTnnATnnnnnnnGTAnTnnTGAC
AnTTnnnnTTTnnTnnTnGnnnnGTTGnTGnTnnTGTnTnnnGGGATGnnnTnCTTnTGT
ATTnTGCTGnAnnnTTGnTGTnGnnGnnnCTGnnnCnGTnnTnAnnGnnnnnAnnnnnCA
nTnCACATAAACGAAnnnnTnnATTTnnnTAnnAnAAnCnnnAnAnTTnGnACTTTAAnn
TnnAAnnAAnnTGAAnTnnAGnATnnTAnTnnnTnnGAnTTnnnAnnnnnnACnnnAnnG
ATACnGATnnnnGnCnCACTnCnnnnnnGATnnCnnATTnnTnnCnnTnCnCTTnnnGCT
GTTTTnnnGnnnGCnTnCnACnTCnTnnTnnTnAnnAnnnnAnnnCAAnTAGCAnnnTCC
AAnnGnGnTnACTnnnnnTnnnnCTTGCnGnTGnTGTTTnTnnCAGTTnnCTCnTnCnnT
TTGCTnnTTGCTnnTnGCnnTnAnnCCTnTTTnCTnTnTCTTTnnGCTTTnnTCTnCTTC
TTnnnnATTATnAAnTTTnnnnCTATnATnATnAGnnTTTnGnTTTnnTnnnAATnnnGn
TnnnnnnACnnATTAnnTnnTGATGCCAnnTnTnnTCTTTnCTnGnTTACnAnTTnTTAn
nACTnTTGTATACnnnAnAnTAnTGTnACTnCTTnAnTnnTCATTACTTnAGnTGATnnn
nnAACAnnnnCnATTnnTGAnnnnnnnTAACAnnTnGnTGnTTnTAnnnAAAAnnGnnnA
TnTGnnGTnAnnnAnnnTGnnnTnnTnnAnAnTnAnTAnACTnCnGAnTAnTAnnAnnTn
nnnnCAnCnCAnnTGAnTAnnnACACTnGTnTTnAACnTGnTAnnTTnnnnATnTTCAAT
AnAATnnTnnnTnnnnCTnnAGnAnnnnTACAAnTTnnCAnnATCnACAAnTnnTnnnnn
nTnnTnAAnCCAGTAAnGGAnCnnnnnTAnnAnnnAnnnnnGACGnnnAnnnnnnTGnnn

nTGnAAnnACnAnTnnnnnnGTACnnACTTnnGnAnnnnnTnnTTAnGnnnnAGAnAnGT
nCnnnnnnnGTTAATAGnnnACTTCTTTTTCnTTnnTTnnTGGTATTCTTnnTnnTTnTA
nTAGCnAnnnnTnnTnCnTTnCnnTTGTnnnCGTnnTGCTnCnTnnTnGTTnAnnTnnnT
CTTGnnAnAnCTTnTTTnTnnnTTTnnTnnnnnnTnAAnAnTCTnAATTCTTnTAnnGTn
nnnnnTnnnnTnnTnnTnnnGnnnnAnATnTTATATTnGTTTTnnTnTnTnnAnnTTTAn
TTTTnnnnnTnnnnGATTnnAAnnnTnnTnTTnnnnTTnAnnAGnTTAnAAnGCTACTnG
nACnnTGGAAnnTAnTAAnAnGTTTCnTnnTCnnTAnATnnTAnTnnnTTnTnnAATnTn
CCTATnnnnnnnnnnATnnnTTnTTGTnTATAATnAnnTnAATTTTnCnTTTnCnnTTAT
GnnnAnTnnCTTTnnCTTGTTnTnTnnTTGnTnCTnTTTnCnGnnTnnnTTnGnTCnTnG
nTnnnnTTGCnAnnnnAATGGCnTnTCnnGTAGnnTTGATGTnGCnnnGCnnCTTCnTnn
CTTCTTGCnGAnTnTnTGCnnnTnCnCGTTnCnTGTGnTnnTTCnnnCnnGnAnCTnnnn
nTnnTCnnnAnnnGCnACnnCATnnCACnTTnnTnnCCAnnCCnCTnnTnGnAnnTnnAC
TCnnAnTCGGnnnTnTGATCnnTnnTnnnnnnnnnCnGnTTnCnnGAnnCnAnnTnGnAn
TCTGnGnnnTCAnGGACCTnnnnnAnGAnnTnACnnnTnCnAnnTCnCGAAnGnnnTCnT
nnTTCnAAnTnnAAnnTTnnCnnnTTnTAGCnnGTGAnTnnGGTnTTnnTnCnTAnnGnC
nnnAnAGGnTTGnnnAnnATnAAnTAnACACAnnnCnnTCnACTAnCAnTnAnnAnAnnn
CnnnnnnnnnnnnnnnnGTGACAAnAGAnnTTTCATnTnnnnGAnnnTnAnnTTAnTnTn
nnnnAGAnnTTnnnAATTATnATGnnGnnTnnTAAAnnTTnCnTnTnnnATnTnGATTAn
ATnATAnnnCTnATAATTAnnAATTTATCAnnnTCAnTATCnnAnAATAAATATTnTCnA
TnAnAnnAAnnGnAAAnAATnnAnATTnATTAAAnGAACAnnAAAAnTnnnnTTTTnnTn
nnACnnAnAACAnTnnnTACTTGTnAnnnnTAnnAnTACCAAnnGnnTnTTnGnnGTAnn
ACnnnACTTnTnAnnGnnnnnnnnnCnTAnnnnnnATAnnAGGnnAATnnnCCATTTCAn
CnTCnAnCnGnTAnCAAnTnTnnnnTnAnnTnCTTnAnCAnTnAAnTTnnnTTTnCTTTT
nCTGAnGnCnTAAAnCnnnTnTnTCnnnTnCnnGCnAnATnAnnTTCnnCTAnnnnnTnC
ATCnnACAnnAGnnAnnnCAAnAAnnTTACTCTnnAATnTnnnnTATnnTTnnnnnnAnA
nTnTTTATAnTnTTTTnnTTnnnACTnnAAAGAnAnnnAnAnTnATnGnnnTTnnAAnAA
TTnACTnnTAnTTnTnCTTTTTAnnnTTTnTGnTnTTnnTTGTTTTnTTTnTGCTTnTTT
nCTnTTnGTTnTnATTnnnATTnYnAnnTnnTnnTGnATCTTnnnACnnCTnnAnGATCA
TGAAAnTTCTTGTTnTnnnAnnAAnnAnnAnAACTnTnnnTnCnTnTnAnnAAnnATnnn
nTTTnCAGTnnTnTACnAAnnAnnAAnCATATnnnGTTnATGACCnnnnnCnTATTCAnT
TnTAnTnTAnnnnnnATATnnnATTnnnnnCTAGnAAAnnAGnAnCTTTAAnnnAAnAGT
GnGTnnATnAGnnTnGTTnnnAATCnnnnATnnnGnAnATnGTTATTGnnAATTAnACAn
TTTCnnnTTTnCCnTTTACAATTAATTnnnAnnnnCnTAAAnTnnnnAnTnTTGTnGnnC
GTTnTnCGTTnTnnGTAGACTTnTTAGAnTnTnnTGnnnTTnGTnTTGTTTTnnATTTTA
TCTnnACGnTCnAnnTTAAATnTnTnATAnnnGAnCnnAAAATCnnnGAnnTnCnnTCnn
CAnTACnnnTnnnGGAnnnTnnnAnTCAACnnGCAGTnAGCAGnATGnAnAACGnAnnGn
GGCnnGAnCAAnAnAAnnnnnnnCnnnnGGTnnACAnAATAAnAnnGnnnnnnGGnTCnn
CnCnCTCAnnCnACATGGnnAnGnAnnnCTnAnATnCCnTnnnnnAnnAnGCnnnnCnAT
TAACnCCnAnAACAnnnnAGAnnnCCAAnTnnnnnAnnnnCnAnnAnCnACnAnAnnAAT
TnnTnnTGnnGAnGnnnAAnnnAAAGAnnTCnnnCCAAnAnnGnAnTTnTnnTAnCTAGn
AnnTGGnnnnGAnnCTnnAnnTCCCTATGGnnnnnACAAnnACnnCnnCnTnnGGnTTGn
AACnGAGnnnGCnTTGnnnnnnnnAnAnnAnCAnAnnGGnACnnGnAAnCnnnnnAACAn
nGCAGCAnnCnnnnTACnACTnCCnCnAnnnnCAACATnGnnnAAnGnCTTCnACGCAGA
AnGnAGCAGAGnnnnnAnTCAAnCnTCnnCnnGnnCnTCAnCnCnnAnnCnnnAnnnTnC
nnnAnATnnnAnTCnAGnCnnnAnTnAnCnAnnAnCTCnnnCnnnAnnnGCnGnnAAnnn
CCnTnnnGCnGnnnnnGnnnTGCTnnTGnTTGnnAnnTTGnnCnnGCTTnnnnGCnnAnT
GTCTGnTAnAnGCnnAnnACnAnAAnGnnnAAnnGTCAnnAAnAAnTnnGnnGnnGAGGC

nTnnAnnAAnCnnCnGCAAnnGCnnAnTnnnAnTAnAnnnTACAAnGTAAnACnAnCnnT
nnGCAnACGTnnnCnAnAACAnACnCnAGnAnnnTTnGnGGACCAGnAACnAnnCAnnnA
AnGnACnGnnnACAAAnAnTnnnCGnAnATnGnACAnTTTnCnnnCAGCGnnnnAnnnTT
CnnCnnnnTGCCGCnnnTnCnnAnGnnnnnnnCAnCTTCnnGnnCnTGGTnGnnnTACnC
AnnnnnnnnnnAAnnnGnTnAnnnAnnnCCAnnnTnTAAAnATnnAATnATnnnGCTnAA
TAnGnAnATnGACnnnTAnnnnAnATnnAnAAnAATAnAGnCTAAAAAnnACAAAnAGCA
GAAnGnnGAnnAAACnnAAnCCTnAAGnCnGAGnCnGAAGAAACAnnAAACTnnGAnnnn
TnnnnnnnCTnnAnnnTTnnnnnnnTTCTnCAnnCAACTGCAAnnnTnnATGAnnnnTnC
TnACTnnAnTCAnnCnTnAnnnCnnGnnnAnCACAnAnGnCAGATGnnCnATATnAACnn
nnnnnnnTTnnnGTTTACnATATAnAnTnTAnTnTTnTnCnGAnTGAATTnTnnTnnCTA
CTTAGCAnAAGTAnnnnTnnTTAnnnnTAAnCnCAnATAGnAnTCTTnAATCAnTnTnTA
ACATnAnnGAnnACTTnAAnnAGnnAnCAnnnTTnnAnCGnnGnnnnnnGnAnnACnATn
nAGTGnACAGnGAACAnnGnnnnnnnnnnnnnGCnTAnAnnGAAGAnnGCTAATGnGnnAT
ATnnATnTnAGnnnTnCTnTnnnnnTnnnATTTnnATAGnTnnTnAnnTGnATnAnnAAA
AnnAAAAnnnAAAAAAAAAAAAAAAAAAAnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnn

Results of Multiple Sequence Alignment of Protein Sequences using CLUSTALW

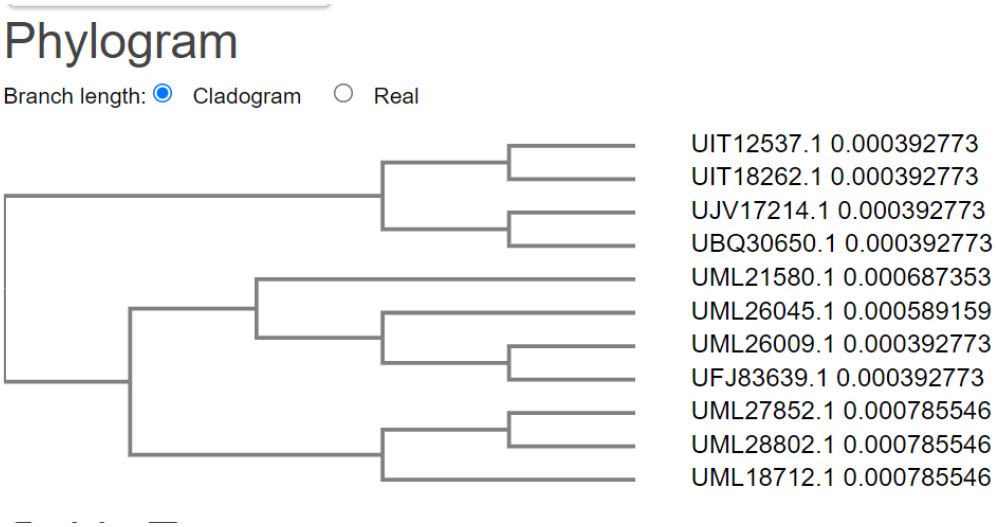


Fig: Phylogram or Guide Tree of Protein Sequence obtained in ClustalW

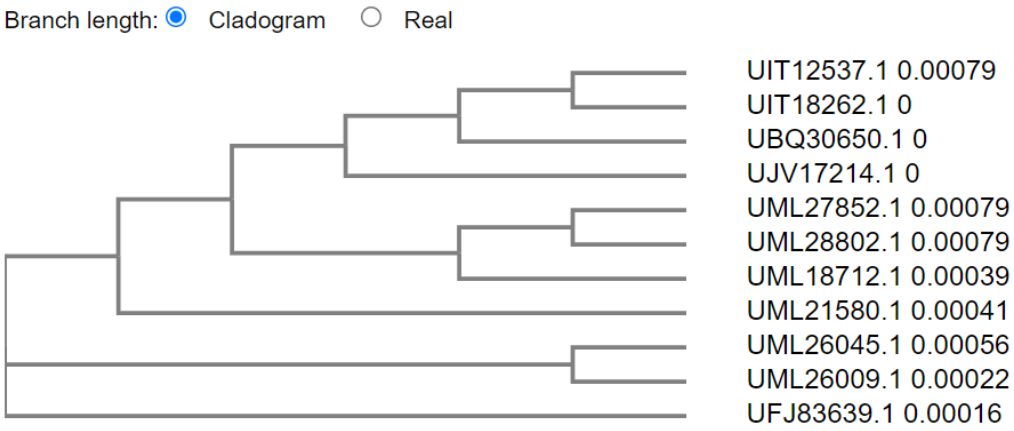


Fig: Phylogenetic Tree of Protein Sequence obtained in ClustalW

Consensus Sequence of given set of proteins as generated in EMBOSS Cons

>EMBOSS0001

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS

NVTWFHAIHVSGTNVIKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV

NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE

GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT

LLALHxxxxxxxNSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK

CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN

CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD

YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYQYRLFRKSNLKPFERDISTEIYQAGSTPC

NGVEGFNCYSPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN

FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP

GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY

ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI

SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE

VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC

LGDIAARDLICAQKFNGLNVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM

QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN

TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA

SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA

ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP

LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL

QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD

SEPVLKGVKLHYT

Conclusion/Inference

A) For DNA Sequence: The Multiple Sequence Alignment (MSA) using Clustal Omega of surface glycoprotein UIT12537.1 gene sequence with sequences obtained in the Blastn shows Presence of **Conserved regions (*)**. The guided tree in the form of cladogram (using distance matrices) showed that are closely related with a minimum distance of **0.000133873**.

The Real tree shows that the number of mutational steps is highest in genes **with accession number OM434807.1 and OM434810.1** both have values as **.000133873.**

The Phylogenetic tree (neighbour-joining method) shows that **genes with accession numbers OL622101.1 and OU085847.1** are recent descendants with values of **0.00026 and 0**, respectively while genes with **accession number OK314233.1 and ancestor of genes (OM434807.1 and OM434810.1) have diverged** during early phage of evolution having **0.00024 and 0 value** respectively.

B) For Protein Sequence: The Multiple Sequence Alignment (MSA) using Clustal Omega of surface glycoprotein UIT12537.1 gene sequence with sequences obtained in the Blastp shows Presence of **Conserved regions (*)**. The guided tree in the form of cladogram (using distance matrices) showed that (**surface glycoprotein UIT12537.1 and surface glycoprotein UIT18262.1),** (**surface glycoprotein UJV17214.1 and surface glycoprotein UBQ30650.1),** (**surface glycoprotein UML26009.1 and surface glycoprotein UFJ83639.1),** are closely related with a minimum distance of **0.000392773**.

The Real tree shows that the number of mutational steps are highest in **surface glycoprotein UML26009.1** and **surface glycoprotein UFJ83639.1** having **0.000392773 and 0.000392773 value respectively.**

The Phylogenetic tree (Neighbor Joining method) shows that **surface glycoprotein UIT12537.1 and UIT18262.1** are recent descendants with values 0.00079 and 0 respectively while **surface glycoprotein UFJ83639.1** has been diverged during early phage of evolution having **0.00016** value.

Task 7: Draw MP, NJ, ML trees (with 500 bootstrap) with your multiple sequence alignment and compare your trees and draw inference regarding the evolution of your gene/protein.

MP, NJ, ML Trees (with 500 bootstrap) for nucleotide multiple sequence alignment

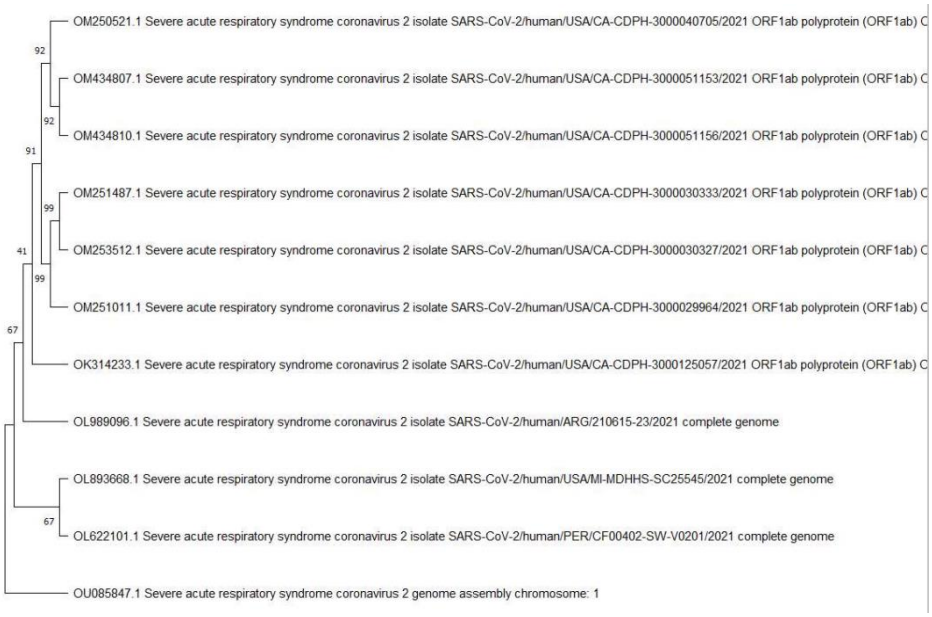


Fig: Phylogenetic Tree of given nucleotide sequence using Maximum Parsimony (MP)

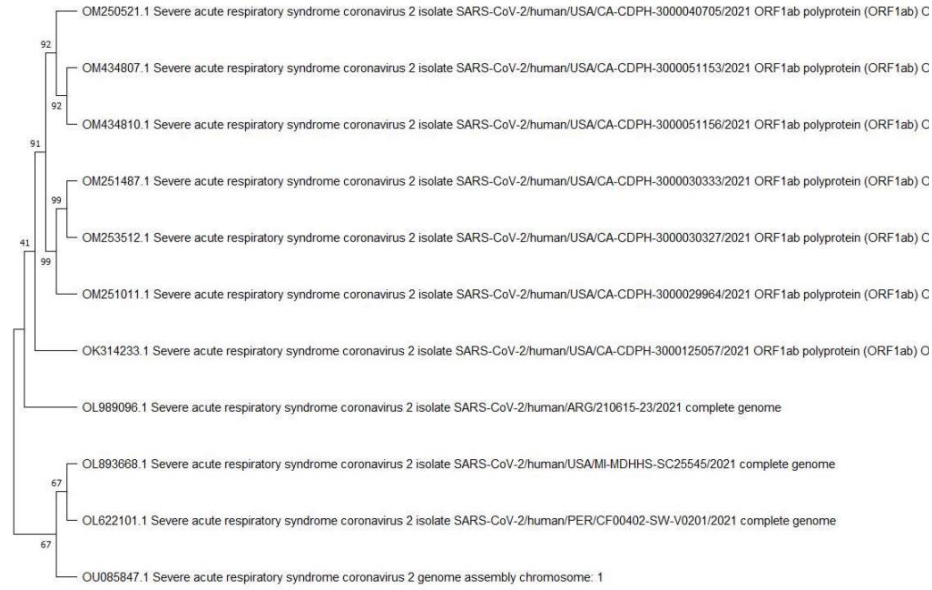


Fig: Phylogenetic Tree of given nucleotide sequence using Maximum Parsimony (MP) with bootstrap

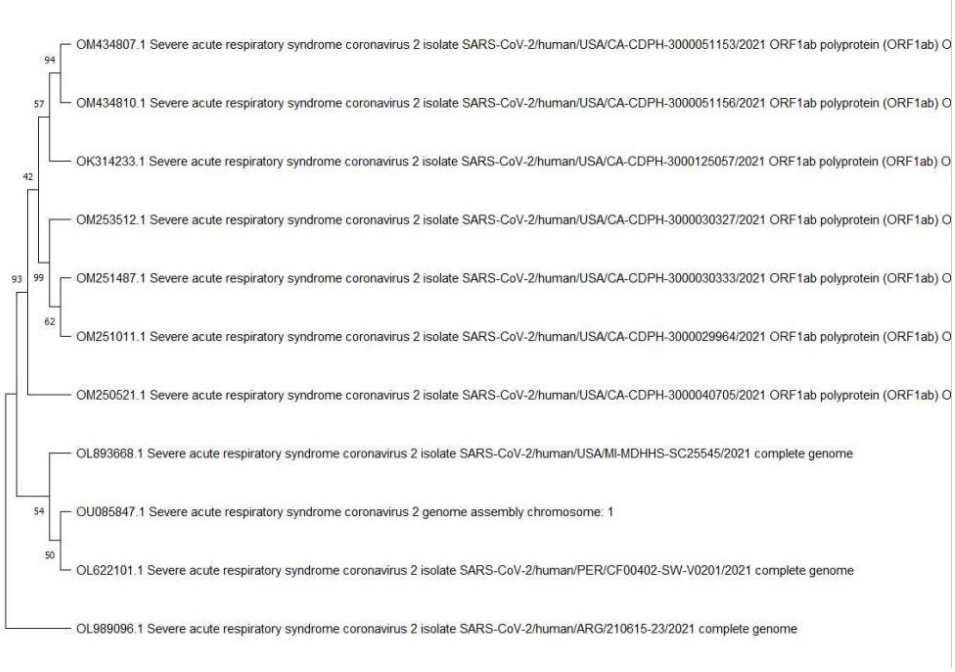


Fig: Phylogenetic Tree of given nucleotide sequence using Neighborhood-Joining Method

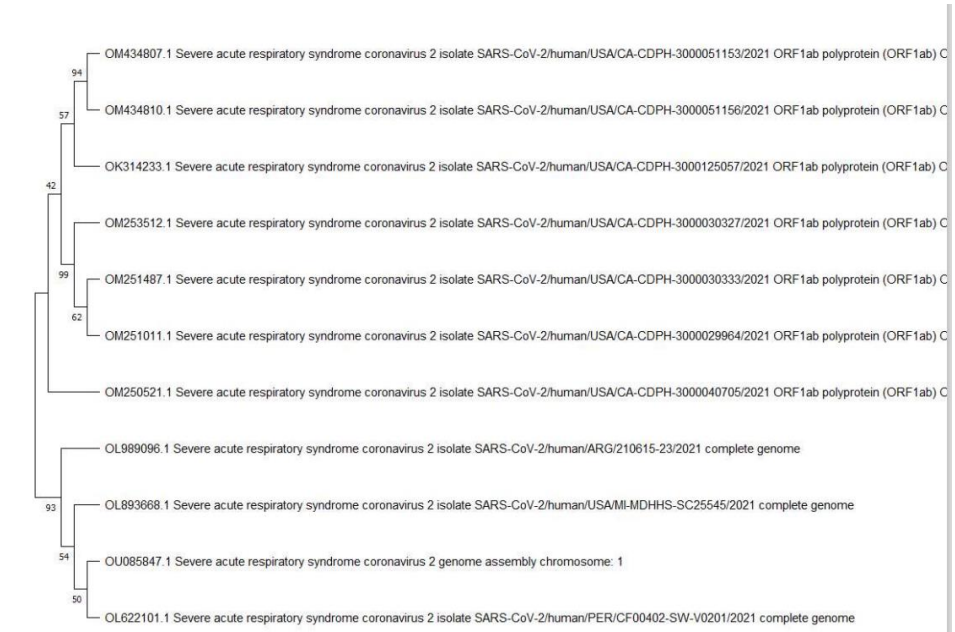


Fig: Phylogenetic Tree of given nucleotide sequence using Neighborhood-Joining Method (with bootstrap)

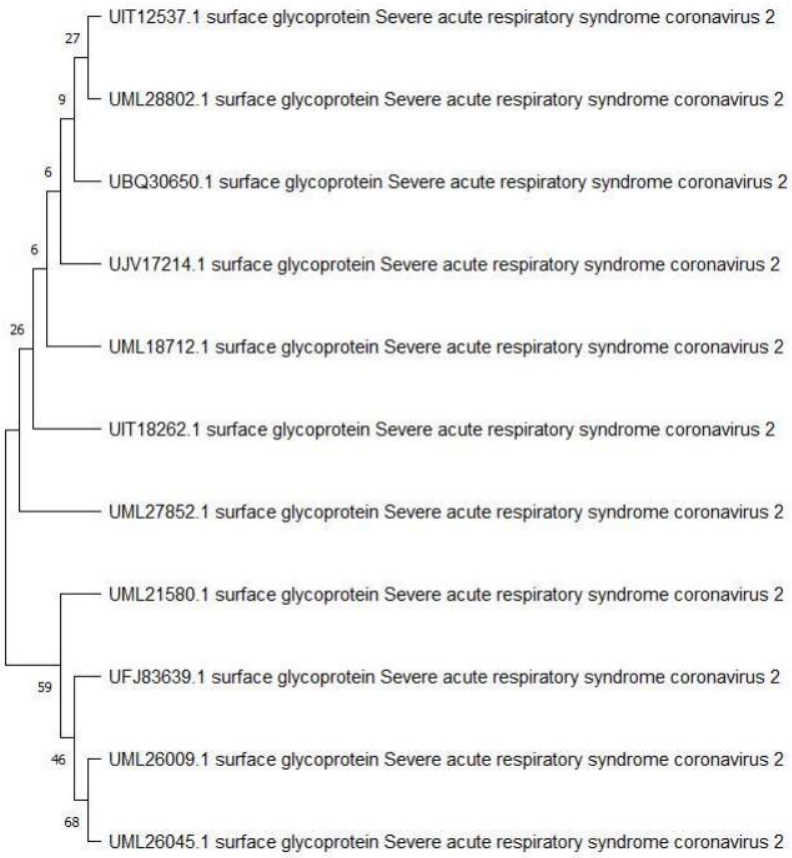


Fig: Phylogenetic Tree of given nucleotide sequence using Maximum-Likelihood (ML) method

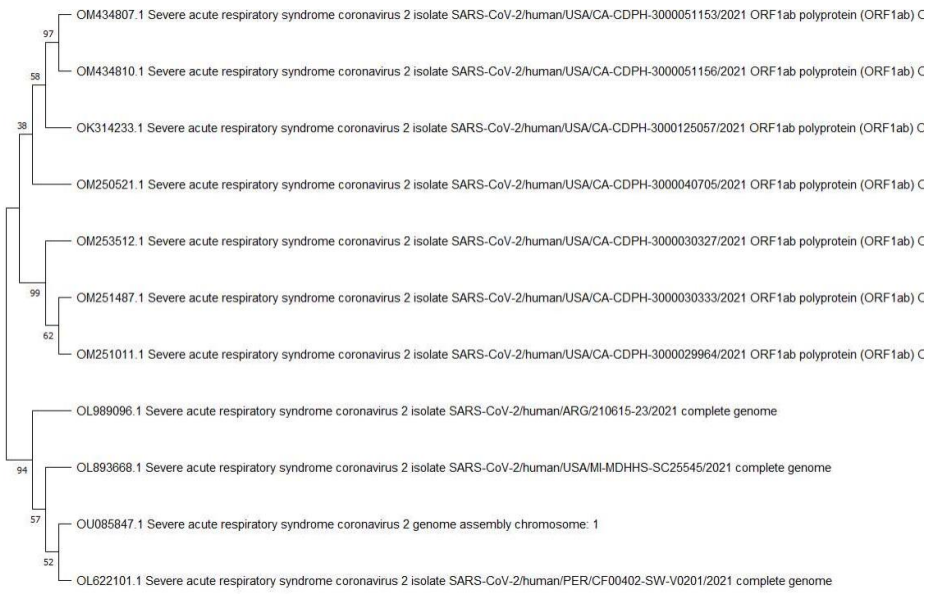


Fig: Phylogenetic Tree of given nucleotide sequence using Maximum-Likelihood (ML) method (with bootstrap)

MP, NJ, ML Trees (with 500 bootstrap) for protein multiple sequence alignment

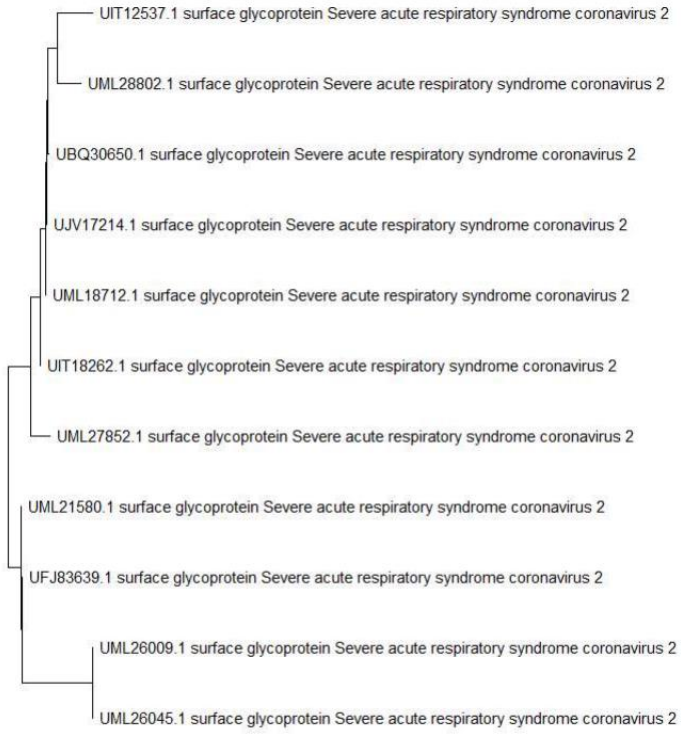


Fig: Phylogenetic Tree of given protein sequence using Maximum-Parsimony (MP) method

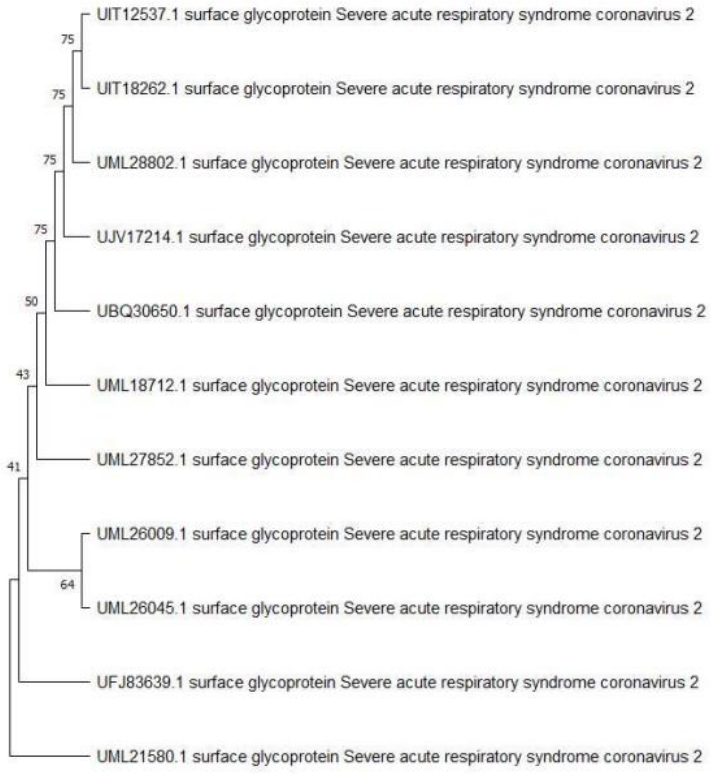


Fig: Phylogenetic Tree of given protein sequence using Maximum-Parsimony (MP) method (with bootstrap)

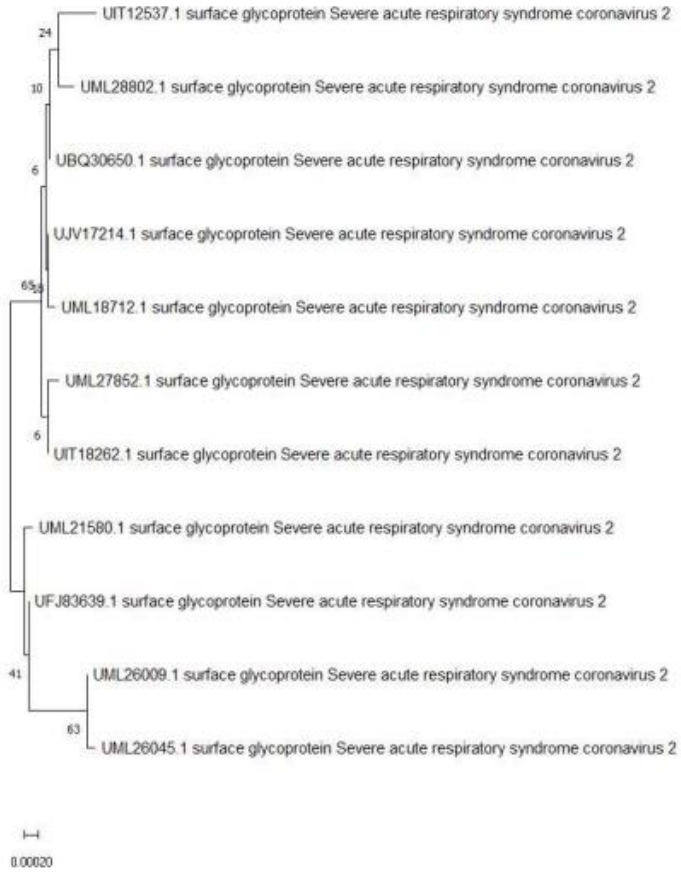


Fig: Phylogenetic Tree of given set of protein sequence using Neighborhood Joining (NJ) method

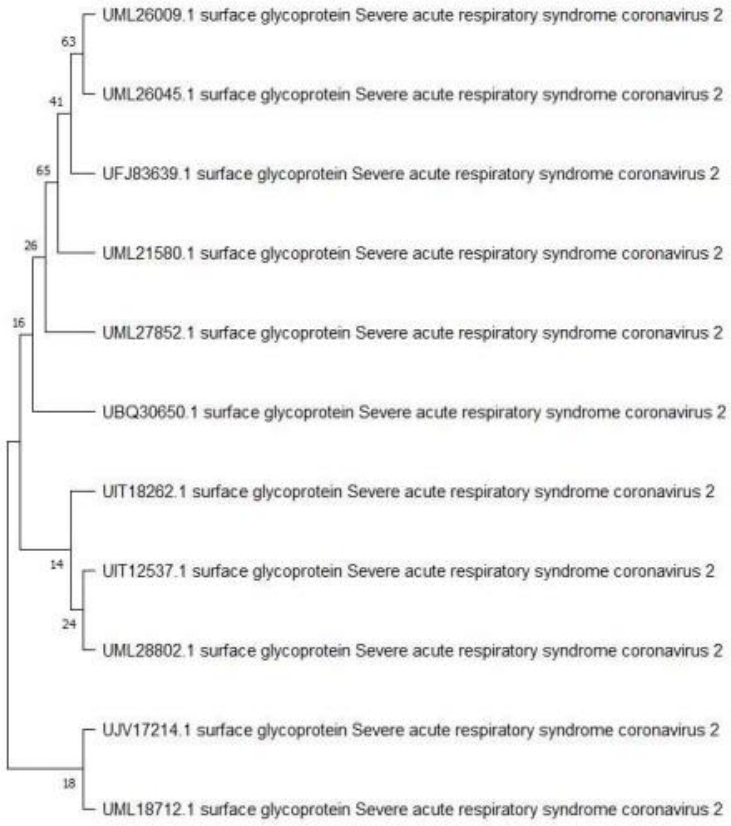


Fig: Phylogenetic Tree of given set of protein sequence using Neighborhood Joining (NJ) method (with bootstrap)

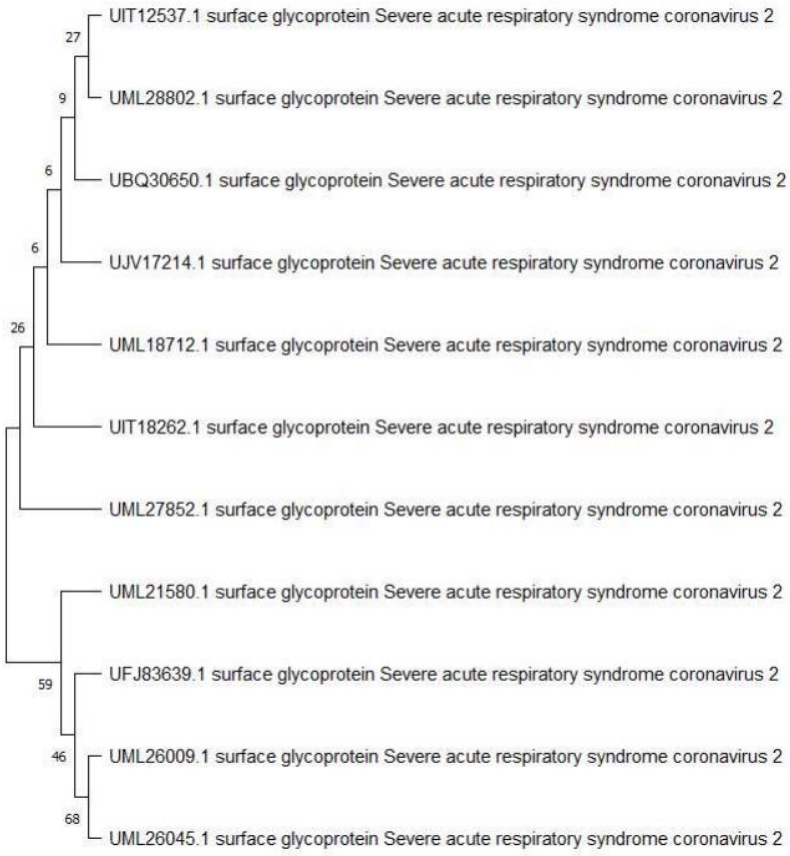


Fig: Phylogenetic Tree of given set of protein sequence using Maximum Likelihood (ML) method

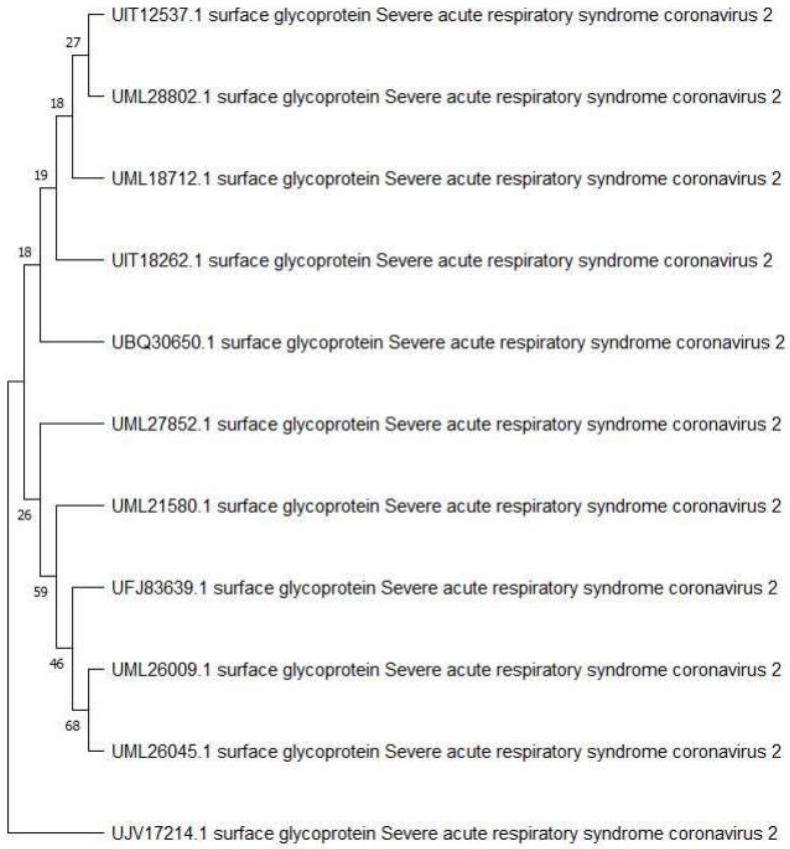


Fig: Phylogenetic Tree of given set of protein sequence using Maximum Likelihood (ML) method with bootstrap

Conclusion/Inference

As we know, in the maximum parsimony method, we choose trees minimizing the number of changes required to explain the data. Here, since the sequences are highly similar as shown by the BlastN, BlastP and Needle analysis, the maximum parsimony method can be used to construct the phylogenetic tree. Distance-based methods such as Neighbourhood Joining method build trees using evolutionary distances between operational taxonomic units. In Maximum Likelihood method, under a model of sequence evolution, we find a tree which gives highest likelihood of the observed data

1. <u>For nucleotide:</u>

   The Phylogenetic analysis of nucleotide sequences of Surface Glycoprotein UIT12537.1 (OM250521.1) using distance and character based methods shows that the most recent common ancestor (MRCA) is genes with accession number OM251487.1 and OM253512.1 along with OM251011.1 with a Confidence value of 99.

2. For Protein:

   The Phylogenetic analysis of protein sequences of Surface Glycoprotein UIT12537.1 (OM250521.1) using Maximum Parsimony shows that the most recent common ancestor (MRCA) are Surface Glycoprotein UIT12537.1 and Surface Glycoprotein UIT18262.1 with a Confidence value of 75.

   The Phylogenetic analysis of protein sequences of Surface Glycoprotein UIT12537.1 (OM250521.1) using Neighbourhood Joining shows that the most recent common ancestor (MRCA) are Surface Glycoprotein UML26009.1 and Surface Glycoprotein UML26045.1 with a Confidence value of 63.

   The Phylogenetic analysis of protein sequences of Surface Glycoprotein UIT12537.1 (OM250521.1) using Maximum Likelihood shows that the most recent common ancestor (MRCA) are Surface Glycoprotein UIT12537.1 and Surface Glycoprotein UML28802.1 with a confidence value of 27, along with Surface Glycoprotein UML26009.1 and Surface Glycoprotein UML26045.1 with a confidence value of 68.

References:

1. **https://www.ncbi.nlm.nih.gov/protein/UIT12537**
2. **https://www.ncbi.nlm.nih.gov/nuccore/OM250521**
3. **https://www.ncbi.nlm.nih.gov/orffinder/**
4. **https://www.ncbi.nlm.nih.gov/blast/**
5. **https://www.ebi.ac.uk/Tools/msa/clustalo/**
6. **https://www.ebi.ac.uk/Tools/psa/emboss_needle/**
7. **https://www.megasoftware.net/**
8. **http://hollywood.mit.edu/GENSCAN.html**
9. http://www.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgr
10. oup=gfind) respectively and analyze the given gene sequence
11. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&BLAST_SPEC=GlobalAln&LINK_LOC=BlastHomeLink
12. Yoshimoto, F.K. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *Protein J* **39,** 198–216 (2020). https://doi.org/10.1007/s10930-020-09901-4
13. Ahmad Abu Turab Naqvi, Kisa Fatima, Taj Mohammad, Urooj Fatima, Indrakant K. Singh, Archana Singh, Shaikh Muhammad Atif, Gururao Hariprasad, Gulam Mustafa Hasan, Md. Imtaiyaz Hassan, Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach, Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, Volume 1866, Issue 10, 2020, 165878, SSN 0925-4439

https://doi.org/10.1016/j.bbadis.2020.16587