Comparative analysis based on the spike glycoproteins of SARS-CoV2 isolated from COVID 19 patients of different countries.

Preeti Mangar[1], Smriti Pradhan[2], Subecha Rai[3], Khusboo Lepcha[4], Vivek Kumar Ranjan[2], Aditi Rai[4]*

[1]Department of Botany, University of North Bengal, Raja Rammohanpur, Darjeeling, West Bengal-734013, India.

[2]Department of Biotechnology, University of North Bengal, Raja Rammohanpur, Darjeeling, West Bengal-734013, India.

[3]Institute of Bioresources and Sustainable Development, Sikkim Centre, 5th Mile, Metro Point, Tadong, Gangtok, Sikkim-737102, India.

[4]Department of Microbiology, University of North Bengal, Raja Rammohanpur, Darjeeling, West Bengal-734013, India.

Abstract

SARS-CoV2 popularly known as (COVID-19) has presently received worldwide attention. It has been considered a pandemic by the World Health Organisation. Owing to its high transmittance factor the virus has brought about many deaths and spread to all the major countries of the world. Scientists and Researchers worldwide are giving their full efforts to develop a vaccine. In our present study, we have included the comparative analysis of the different spike glycoprotein sequences of the patients suffering from COVID-19 from different countries where this pandemic has occurred. Spike glycoproteins are the structural proteins that bring about the binding of the SARS-CoV-2 viral molecule to the ACE2 receptor of the host following which infection occurs. Through this data, we have shown the different point mutations in the spike glycoproteins that occurred over time in different countries as the disease progressed.

Keywords: SARS-CoV2, spike glycoprotein, mutation

*Corresponding Author- aditimicro009@nbu.ac.in

1. Introduction

Diseases since time immemorial has always cost mankind health and wealth, not surprisingly the quest for survival of mankind has been an everlasting battle.Epidemics and pandemics are not new to the history of mankind and records of diseases like Plague and Spanish Flu devastating lives have been an integral part of the epidemiological study of the human race. In December 2019, an incident occured in Wuhan, Southern China where a series of pneumonia cases, was reported. It wasnt long before the cases were classified as viral pneumonia and the virus was speculated to belong to β coronavirus. Primarily it was named as 2019- novel coronavirus (2019-nCoV) by World Health Organization (WHO), which later named the disease as coronavirus disease 2019 (COVID- 19). Also identified with the name SARS-CoV2, the epidemic COVID-19 progressed by leaps and bounds via human-to-human transmission, that made it basically impossible to contain it within a certain area thus leading to a pandemic that crossed borders spreading into the International community. The SARS-CoV2 being a β-coronavirus (Zhu et., 2019) like SARS-CoV and MERS-CoV is responsible for causing severe and potentially fatal respiratory tract infections (Yin *et al*., 2018) Sharing 96.2% identity to a bat CoV RaTG13, and 79.5% identity to SARS-CoV, it may be presumed that the SARS-CoV2 might be transmitted from bats to humans. However, recent studies comparing the receptors on host surfaces, suggest the possibility of alternative intermediate hosts (Liu *et al*., 2020).

Spike glycoprotein of SARS- CoV 2

The complete genome analysis of a strain of SARS-CoV2, obtained from Wuhan revealed that this enveloped virus has a positive stranded RNA genome with a size about 29.9 kb (Wu *et al*.,2020) The study of genomes of CoVs has revealed a variable number of (6–11) open reading frames (ORFs) (Song *et al*., 2019).The first ORF (ORF1a/b) encodes 16 non-structural proteins (NSP), and translates two polyproteins, pp1a and pp1ab, and the remaining ORFs encode accessory and structural proteins. However, four essential structural proteins are encoded by the viruses including spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins(Wang *et al*., 2020; Du *et al*., 2016). Among them, an envelope-anchored spike protein specifically recognizes its host receptor and it serves as a target for development of antibodies, entry inhibitors and vaccines (Du *et al*., 2016; He *et al*., 2006) The S protein is trimeric and each monomer is about 180 kDa, and contains two subunits, S1 and S2, The minimal RBD(Receptor Binding Domain) region is the fragment covering the residues 318-510 in S1 subunit (Xiao *et al*., 2003, Wong *et al*., 2004). The receptor-binding motif (RBM) is the RBD containing a loop region (residues 424-494), which makes complete contact with the receptor ACE2 (Angiotensin Converting Enzyme), the RBM region being tyrosine-rich, It was observed that six out of 14 residues of RBM were tyrosines that were in direct contact with ACE2 (Zhu *et al*., 2013). The S protein first binds to a host receptor through the receptor-binding domain (RBD) in the S1 subunit and then fuses the viral and host membranes through the S2 subunit (Liu *et al*., 2004). In the structure, N-terminal domain (NTD) and C-terminal domain (C-domain) portions of S1 fold as two independent domains, Depending on the virus, either NTD or C-domain (occasionally both) binds to a host receptor and functions as a receptor-binding domain (RBD) (Breslin *et al*., 2003 ; Lin *et al*., 2008). Recently, (Zhou *et al*.,2020) reported that SARS-CoV-2 uses ACE2 as the receptor which is similar to the S1 C-domain of SARS-CoV in the RBD that recognizes host angiontensin- converting enzyme 2 (ACE2) as its receptor (Babcock *et al*., 2004; Li *et al*., 2003). In one of the studies by (Wan *et al*., 2020) it was observed that among the ACE2-contacting residues in the RBD, 9 are fully conserved and 4 are partially conserved among 2019-nCoV and SARS-CoV from human, civet, and bat. ACE2 is a zinc-dependent peptidase that functions in the renin-angiotensin pathway and regulates blood pressure (Donoghue *et al*., 2000; Yagil and Yagil, 2003). However, the physiological function of ACE2 is not related to its role as the SARS-CoV receptor (Li *et al*., 2005b). ACE2 contains an N-terminal peptidase domain and a C-terminal collectrin domain. The enzymatic active site of ACE2 is buried in a claw-like structure with two lobes of the peptidase domain (Towler *et al*., 2004). The binding interactions between SARS-CoV, RBD and ACE2 largely determine the host range and cross-species infections of SARS-CoV(Lu *et al*., 2015).Using computer modeling, Xu *et al* (2020), found that the spike proteins of SARS-CoV-2 and SARS-CoV share 76.5% identity in amino acid sequences having almost identical 3-D structures in the receptor-binding domain that maintains Van der Waals forces. It has been reported that residue 394 (glutamine) in the SARS-CoV-2 receptor-binding domain (RBD), which corresponds to residue 479 in SARS-CoV, can be recognized by the critical lysine 31 on the human ACE2 receptor (Wu et., 2012). Through recent cryo-EM structure studies which further deciphered the S protein of SARS-CoV2 and ACE2 interaction at Angstrom resolution level it was revealed that the overall ACE2-binding mode of S protein of SARS-CoV2 is almost identical to the mode of S protein of SARS-CoV (Yan *et al* ., 2020; Lan *et al*., 2020). Yan *et al*., 2020 performed the bioinformatics analysis of the S protein sequences of coronaviruses and found an evolutionary mutation of K403R in S protein of SARS-CoV-2 compared with that in S protein of SARS-CoV, forming an adjacent RGD (R:Arginine, G:Glycine, D:Aspartic Acid) motif at the interaction surface. The RGD motif is the cell attachment site of a large amount of adhesive extracellular matrix and cell surface proteins and recognized by integrins. The evolutionary obtainment of the RGD (R:Arginine,G:Glycine,D:Aspartic Acid) motif in SARS-CoV2 may play an important role in promoting rapid human to human transmission. Yan *et al*., 2020 also suggested that the RGD motif on the S glycoprotein may bind to the integrin on the surface of host cells which can be a new potential mechanism for SARS-CoV2 infection, resulting in higher affinity with the host cells in comparison with SARS-CoV.

Spike proteins has been found to be immunogenic and induce high IFN-γ-specific T-cell response (Janice *et al*., 2012). Mutations in the spike protein could change the tropism of a virus, including new hosts or increasing viral pathogenesis (Shang *et al*., 2020). SARS-CoV possess some residues in RBD that allows the interspecies infection, known as Y442, L472, N479, D480, and T487 (Lu et al., 2015). However, in SARS-CoV2, slight modification of some residues could improve the interaction with the human cellular receptor: L455, F486, Q493, and N501. In SARS-CoV, two main residues (479 and 487) have been associated to the recognition of the human ACE2 receptor (Lu *et al*., 2015). These residues suffered a punctual mutation from civet to human, K479N and S487T (Li, 2013). In the SARS-CoV2, the residues corresponding to N479 correspond to Q493 and T487 to N501. Moreover, a model shows the presence of the two capping loops in the binding domain which produces a stabilization effect over the interaction with the cellular receptor (Ortega *et al*., 2020). Thus the amino acid substitutions and the longer capping loops could explain the increase in binding affinities in SARS-CoV2 compared to SARS-CoV. Since mutations play a major role, our focus of the present study was to understand the mutations in the spike glycoproteins from different countries as it could provide us an idea about the constant shift in the structure of the spike glycoproteins and probably enabling it to be transmitted to different regions However, is the mutation dependant on the race or ethnicity of a person or the gene pool is an entire new story altogether.

2. Methodology

A total of 22 coding sequences for spike glycoproteins were retrieved from NCBI database (http://www.ncbi.nim.nih.gov). Multiple Sequence alignment of the CDS region was performed using CLUSTAL Omega (https://www.ebi.ac.uk/Tools/msa/clustalol) (Madeira *et al.*, 2019). The sequences were selected on the basis of their origin to 14 different countries affected by the COVID19 pandemic namely USA, India, China, Australia, Finland, South Korea, Brazil, Italy, Japan, Vietnam, Pakistan, Sweden, Taiwan and Spain (Accession numbers have been mentioned in Table 1). The phylogenetic analysis of SARS-CoV2 spike proteins of the different countries was done using MEGAX software **(**MEGA-X Version 7.0) (Kumar *et al.*, 2018). The phylogenetic analysis was accomplished through multiple comparisons using the neighbor- joining algorithm in the MEGA-X**.** Multiple comparisons were done by ClustalW multiple sequence alignment and the neighbor-joining phylogenies were estimated by p-distance method.

3. Result and Discussion

A total of 22 different amino acid sequences of the spike glycoproteins from different countries were analysed by Multiple Sequence alignment (Table 1, Fig 2). The spike glycoprotein sequence of India showed mutations in the S1 and S2 domains. The mutation Ala930Val in the spike protein of the Indian sequence (Accession Number Q1A985839) has been observed to be in the S2 domain. It is well understood that these point mutation enhances the surface area for interaction with the ACE2 receptor while conserving the physico-chemical property of the side chain. Additionally, increasing the chance of vander-walls interactions and contributing to the protein core stability another mutation of Arg408Ile is noted to in the RBD region of the spike protein of another Indian sequence (Accession number MT012098). It has been seen that the RBD regions are mostly tyrosine rich to ensure proper contact with the ACE2 receptor (Zhu et al., 2013). However in contradiction reducing the binding affinity, the same spike glycoprotein of India shows a deletion at 145 amino acid position whereas the rest of the spike sequences from the other countries have a tyrosine residue. The spike glycoprotein of China (Accession numbers QIA20044 and Q1004367) and Finland (Accession number QHU79173) showed mutations in Tyr28Asn, Asn74Lys and His49Tyr respectively. Our study is the first report that indicates a point mutation of His49Tyr in the sequence of Finland (QH QHU79173) sharing similarity with SARS-CoV S^{B}(s-domain) of Urbani sequence(Accession number AAP1344)(data not shown) isolated in the late phase of the 2002-2003 SARS-CoV epidemic and dissimilarity with the remaining SARS-CoV2 spike sequences indicating reversion of mutation at specific sites. Likewise the spike glycoproteins of Australia (Accession Number QHR84449) showed mutation at Ser247Arg. It is quite interesting to note that the four different sequences of United States from Cruise A(Accession numbers QII57278, QIJ96493, QIK50427) and Washington (Accession number QH1187830) show mutation at two different positions where one mutation is at Phe157Leu and the other mutation is at Gly181Val ,Asp614Gly and His655Tyr respectively. According to our study the spike glycoproteins of USA showed the maximum variations between the submitted sequences from United States itself. Though a drastic difference in the RBD region is not observed it is remarkable how small point mutations in other regions has shifted towards providing a favourable environment through hydrophobic interactions and hydrogen bonds as it binds to the host receptor. The spike glycoprotein of Sweden (Accession number Q1C53204) showed an amino acid deletion at the 910 position, however comparatively in all the other sequences of the spike glycoprotein of the other countries glycine was present. Interestingly, another spike glycoprotein sequence of Sweden (Accession number Q1C53204) showed a mutation in the Phe797Cys. The two Korean spike glycoprotein sequences (Accession number QH200379 and Accession number MT039890) showed mutation in the Ser221Trp however, it is interesting to note that the first spike glycoprotein sequence QH200379 was taken from a patient who had travelled to Wuhan from Korea and it is likely that he/she may have been infected in Wuhan, China. Probably these respective mutations in the SARS-CoV2 spike proteins of the different countries contribute to the occurrence of branches forming different clades in the phylogenetic tree (Fig 1). In the phylogenetic tree (Fig 1) SARS CoV-2 spike glycoprotein sequence from Sweden (QIC53204), China (YP009724390), Spain (QIQ08810) and Taiwan(MT06617) form a separate clade from the other sequences. Most of the SARS-CoV2 spike glycoprotein sequences of USA (QII57278, QII87830, QIJ96493, QIK50427), China(QIA20044), India(QIA98583), Finland(QHU79173, Australia(QHR84449) and South Korea(QHZ00379, MT039890) formed distinct cluster within the second clade. The remaining sequences of Brazil (MT126808), Italy(MT066156), Japan (LC528232) and China(NC_045512)corresponded to the second clade of the phylogenetic tree. However, the sequences of India (MT012098) and Vietnam (MT192772) formed a separate cluster within the phylogenetic tree. It is very important to analyse the spike glycoprotein sequences within different locations so as to monitor the spread and mobilization of the SARS-CoV2 to different countries across the globe.

Continuous analysis of the spike glycoprotein sequences of SARS-CoV2 obtained from different regions interacting with the ACE2 receptor is important. The various conserved domains as well as point mutations were noted across amino acid sequences pertaining to the spike glycoprotein of 14 different countries. Thus, one or a few seemingly trivial mutations at the receptor-binding surface of a virus may lead to dramatic epidemic outcomes by facilitating cross-species infections and human- to-human transmission of the virus. The changes in the amino acid residue of the Receptor binding domain may determine the host’s fate and role in a viral epidemic by presenting species barriers for viral infections. The findings can fill in an important missing link and lead to development of vaccines and therapeutics associated with the COVID-19 pandemic. Also, the present work gives a better insight to understand the positions of amino acids which may be susceptible to mutations and can drastically aid SARS-CoV2 to evolve in the near future to another potential pathogenic strain.

References-

Babcock GJ, Esshaki DJ, Thomas WD, Ambrosino DM. 2004. Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. Journal of Virology. 78, 4552–4560.

Breslin JJ, Mork I, Smith MK, Vogel LK, Hemmila EM, Bonavia A, Talbot PJ, Sjostrom H, Noren O, Holmes KV. 2003. Human coronavirus 229E: Receptorbinding domain and neutralization by soluble receptor at 37 degrees C. Journal of Virology.77, 4435–4438.

Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, Donovan M, Woolf B, Robison K, Jeyaseelan R, Breitbart RE, Acton S. 2000. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. Circulation Research. 87, E1–E9.

Du L, Zhao G, Kou Z, Ma C, Sun S, Poon VK, Lu L, Wang L, Debnath AK, Zheng BJ, Zhou Y, Jiang S. 2013. Identification of Receptor-Binding Domain in S protein of the Novel Human Coronavirus MERS-CoV as an Essential Targetfor Vaccine Development. J. Virology. 87, 9939–9942.

He YX, Li JJ, Li WH, Lustigman S, Farzan M, Jiang SB. 2006. Crossneutralization of human and palm civet severe acute respiratory syndrome coronaviruses by antibodies targeting the receptor-binding domain of spike protein. Journal of Immunology. 176, 6085–6092.

Janice Oh HL, Ken-En Gan S, Bertoletti A, Tan YJ. 2012. Understanding the T cell immune response in SARS coronavirus infection. Emerging microbes & infections 1.1: 1-6.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution. 2018 Jun 1; 35(6):1547-9.

Li F. 2013. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Research. 100: 246-54.

Li WD, Shi ZL, Yu M, Ren WZ, Smith C, Epstein JH, Wang HZ, Crameri G, Hu ZH, Zhang HJ, Zhang JH, McEachern J, Field H, Daszak P, Eaton BT, Zhang SY, Wang LF, 2005b. Bats are natural reservoirs of SARS-like coronaviruses. Science. 310, 676–679.

Li WH, Moore MJ, Vasilieva N, Sui JH, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M. 2003.Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426, 450–454.

Lin HX, Fen Y, Wong G, Wang LP, Li B, Zhao XS, Li Y, Smaill F, Zhang CS. 2008. Identification of residues in the receptor-binding domain (RBD) of the spike protein of human coronavirus NL63 that is critical for the RBD-ACE2 receptor interaction. Journal of General Virology. 89, 1015–1024.

Liu Y, O'Connor MB, Mandell KJ, Zen K, Ullrich A, Bühring HJ, Parkos CA. 2004. Peptide-mediated inhibition of neutrophil transmigration by blocking CD47 interactions with signal regulatory protein α. The Journal of Immunology. Feb 15;172(4):2578-85.

Liu Z, Xiao X, Wei X, Li J, Yang J, Tan H, Zhu J, Zhang Q, Wu J, Liu L. 2020.Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. Journal of medical virology.

Lu G, Wang Q, Gao GF. 2015. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiology. 23:468-78.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019 Jul; 47(W1) W636-W641. doi:10.1093/nar/gkz268. PMID: 30976793; PMCID: PMC6602479.

R Yan, Y Zhang, Y Guo, L Xia, Q Zhou. 2020. Structural basis for the recognition of the 2019-nCoV by human ACE2. BioRxiv.

Shang J, Wan Y, Liu C, Yount B, Gully K, Yang Y, Auerbach A, Peng G, Baric R, Li F. 2020. Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. PLoS Pathogens. 16(3):e1008392.

Shijia Y, Haixia S, Xianzhang B, Guohui W. 2020. An evolutionary RGD motif in the spike protein of SARS-CoV-2 may serve as a potential high risk factor for virus infection?

Song Z, Y, Bao L, Zhang L, Yu P, Qu Y, Zhu H, Zhao W, Han Y, Qin C. 2019. From SARS to MERS, thrusting coronaviruses into the spotlight. Viruses. 11(1):E59.

Towler P, Staker B, Prasad SG, Menon S, Tang J, Parsons T, Ryan D, Fisher M, Williams D, Dales NA, Patane MA, Pantoliano MW. 2004. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. Journal of Biological Chemistry. 279, 17996–18007.

Wan Y, Shang J, Graham R, Baric RS, Li F. 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus."Journal of virology 94.7.

Wang Q, Vlasova AN, Kenney SP, Saif LJ. 2019. Emerging and re-emerging coronaviruses in pigs. Current opinion in virology. Feb 1;34:39-49.

Wong SK, Li W, Moore MJ, Choe H, Farzan MA. 2004. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. Journal of Biological Chemistry. 279:3197-201.

Wu F, Zhao S, Yu B. 2020. A new coronavirus associated with human respiratory disease in China. Nature.

Wu KL, Peng GQ, Wilken M, Geraghty R.J, Li F. 2012. Mechanisms of Host Receptor Adaptation by Severe Acute Respiratory Syndrome Coronavirus. Journal of Biological Chemistry. 287, 8904–8911.

Xiao X, Chakraborti S, Dimitrov AS, Gramatikoff K, Dimitrov DS. 2003. The SARS-CoV S glycoprotein:expression and functional characterization. Biochemical and Biophysical Research Communications. 312:1159-64.

Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, Zhong W, Hao P. 2020. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its Spike protein for risk of human transmission. Science China Life Sciences. Mar;63(3):457-60.

Yagil Y, Yagil C. 2003. Hypothesis - ACE2 modulates blood pressure in the mammalian organism. Hypertension 41, 871–873.

Yan R, Zhang Y, Guo Y, Xia L, Zhou Q. 2020. Structural basis for the recognition of the 2019-nCoV by human ACE2. BioRxiv.

Yin Y, Wunderink RG. 2018. MERS, SARS and other coronaviruses as causes of pneumonia. Respirology. Feb;23(2):130-7.

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD. 2020. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. BioRxiv.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P. 2020. A novel coronavirus from patients with pneumonia in China, 2019. New England Journal of Medicine.

Zhu X, Liu Q, Du L, Lu L, Jiang S. 2013. Receptor-binding domain as a target for developing SARS vaccines. Journal of Thoracic Disease 5.Suppl : S142.

Table 1: It represents the Accession Number, Country and the Date of submission of the sequences in the NCBI database.

Serial No.	Accession No.	Place	Date of submission
1	QHR84449	Australia	30.01.2020
2	QHU79173	Finland	17.03.2020
3	QH200379	South Korea	11.02.2020
4	QIA20044	Yunnan(China)	09.02.2020
5	QIA98583	India	11.02.2020
6	QIC53204	Sweden	20.02.2020
7	Q1157278	Cruise A (U.S.A.)	07.03.2020
8	Q1187830	WA/ (U.S.A.)	06.03.2020
9	QIJ96493	Cruise A (U.S.A.)	12.03.2020
10	QIK50427	Cruise A (U.S.A.)	16.03.2020
11	Q1004367	Yunnan (China)	20.03.2020
12	Q1Q08810	Valencia (Spain)	23.03.2020
13	MT012098	India	01.02.2020
14	YP009724390	Wuhan (China)	17.01.2020
15	MT240479	Pakistan	25.03.2020
16	MT126808	Brazil	02.03.2020
17	MT192772	Vietnam	17.3.2020
18	LC528232	Japan	29.02.2020
19	MT039890	South Korea	11.02.2020
20	MT066175	Taiwan	14.2.2020
21	NC045512	Wuhan (China)	3.3.2020
22	MT066156	Italy	09.03.2020

Fig 1. Phylogenetic analysis of SARS cov2 spike proteins obtained from different countries. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.00942794 is shown. The evolutionary distances were computed using the p-distance method and are in the

units of the number of amino acid differences per site. The analysis involved 22 amino acid sequences. Evolutionary analyses were conducted in MEGAX software.

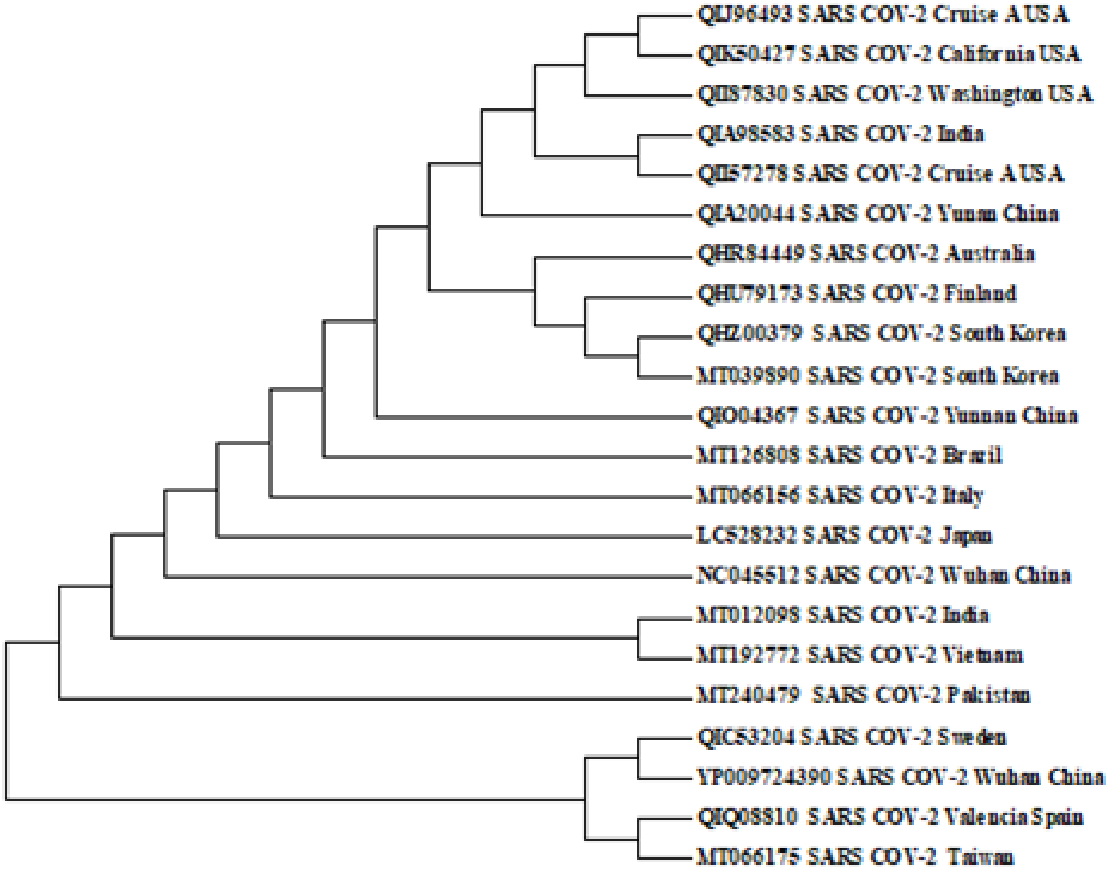
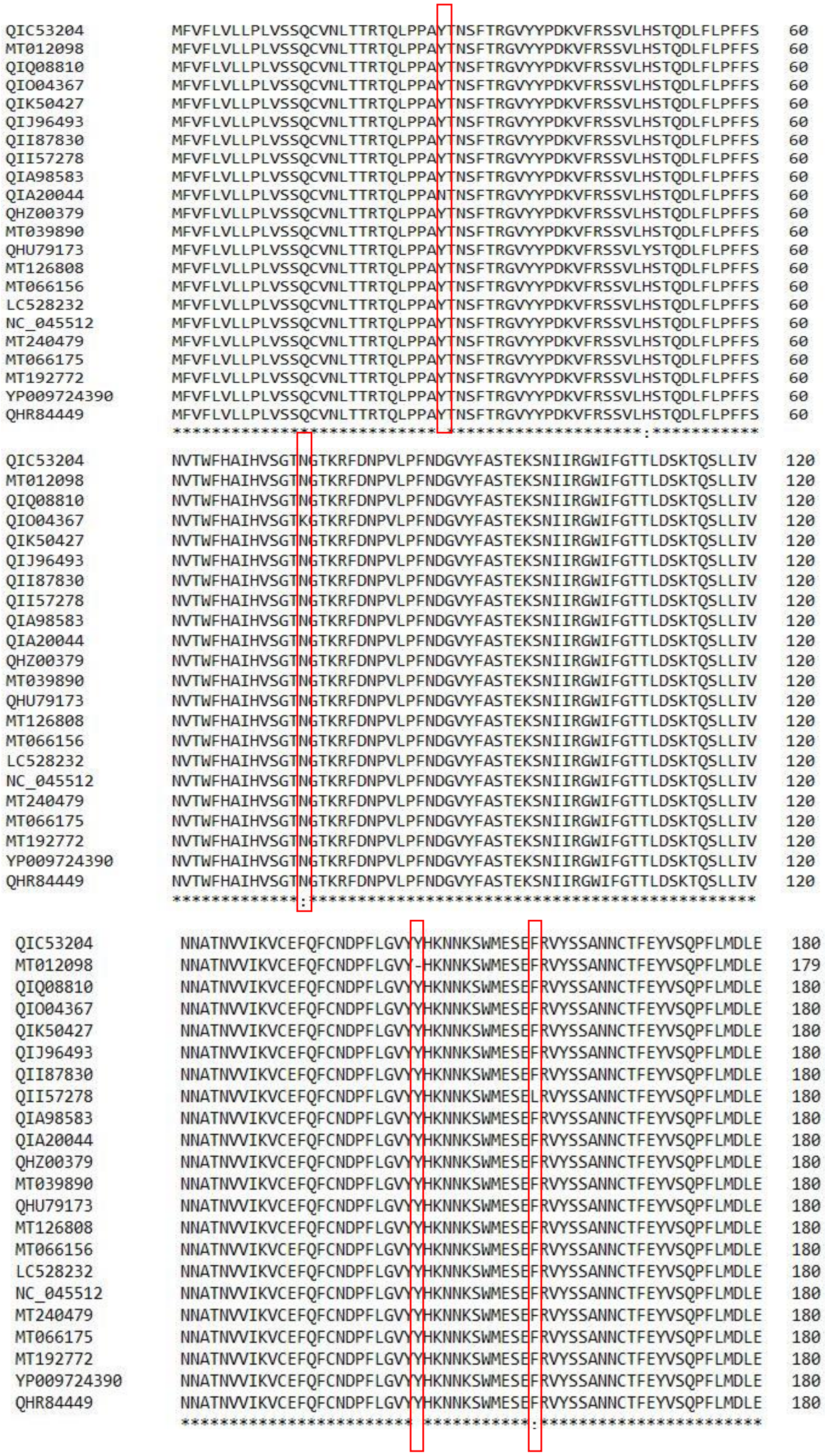
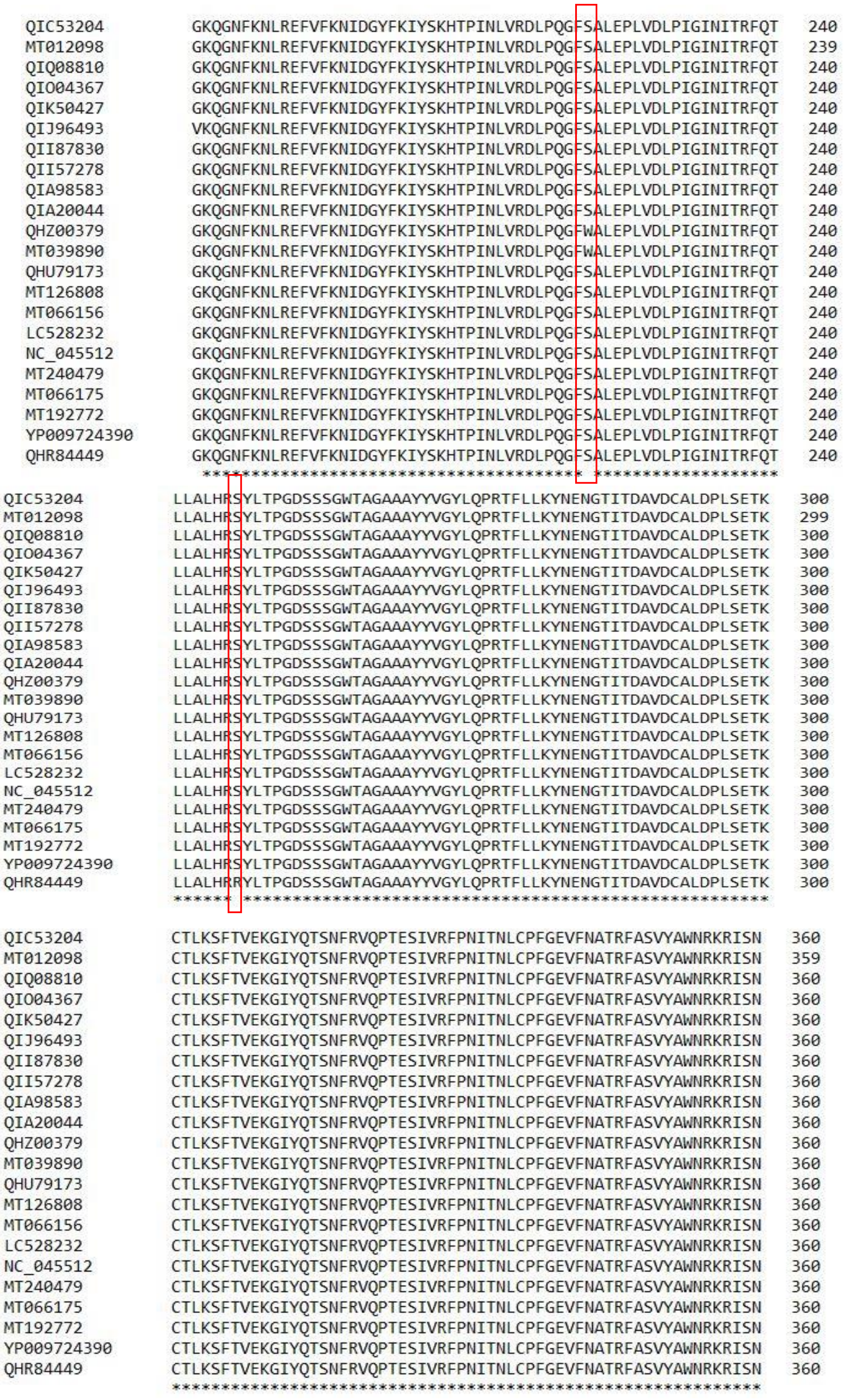


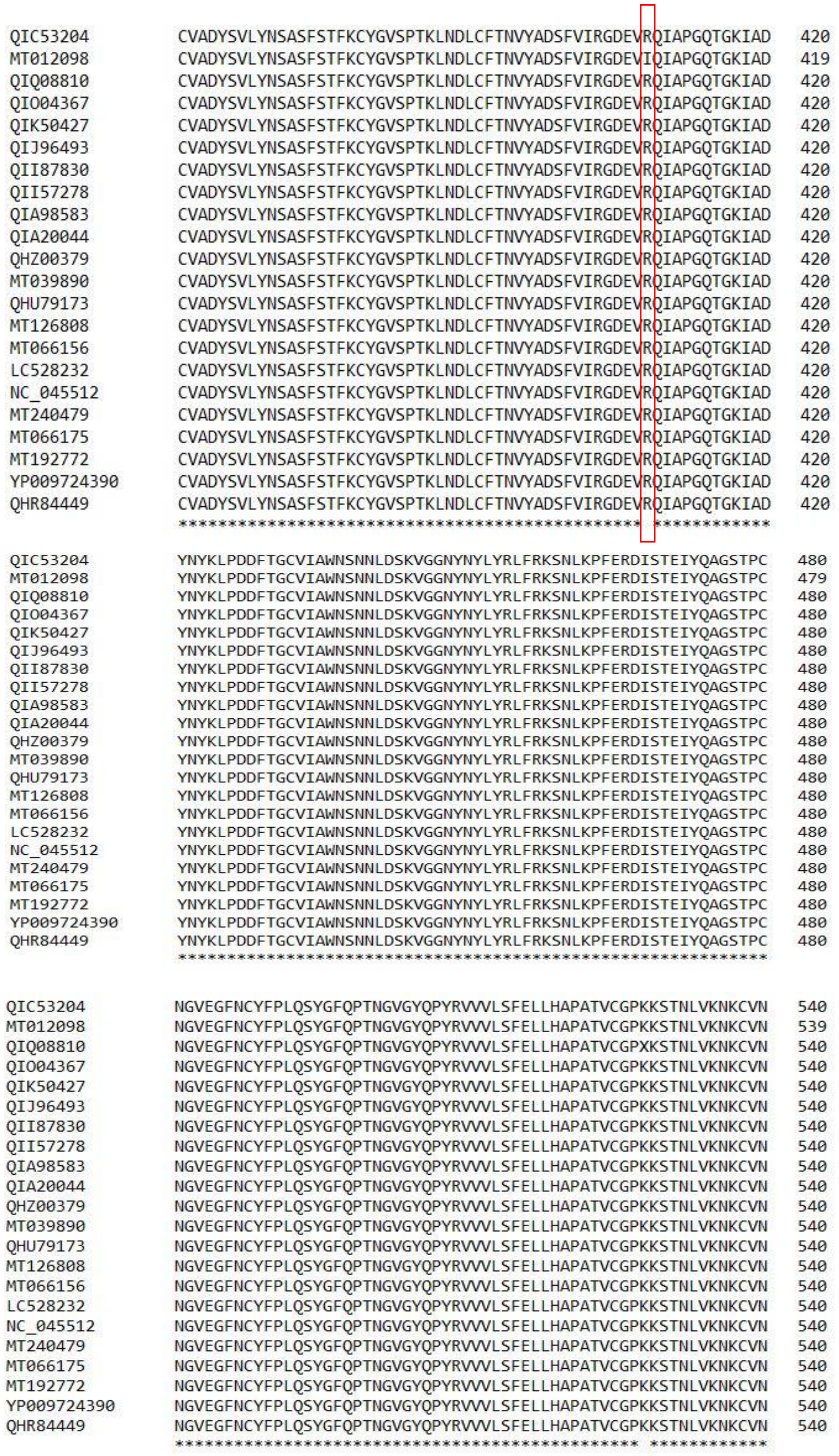
Fig 2. Multiple sequence alignment of 22 SARS-CoV2 glycoprotein encoding sequences using CLUSTAL Omega (https://www.ebi.ac.uk/Tools/msa/clustalol). The “*” means the identical and fully conserved amino acid residue. “:” means conservation within strong group of amino acid residue. “.” means conservation within weaker group of amino acid residues. The red bordered boxes indicate the point mutations .



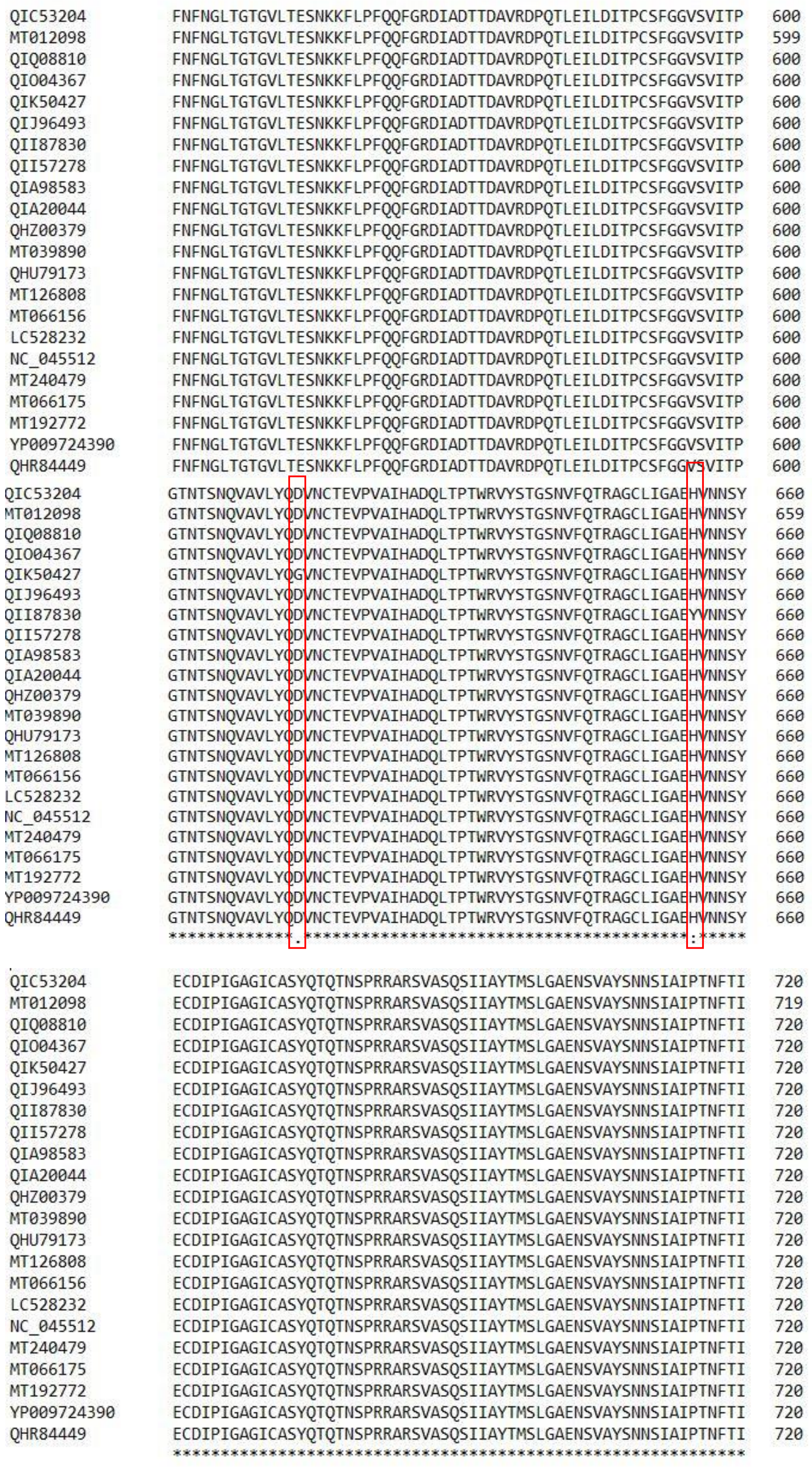
```
QIC53204        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
MT012098        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QIQ08810        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QIO04367        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QIK50427        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QIJ96493        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QII87830        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QII57278        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QIA98583        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QIA20044        MFVFLVLLPLVSSQCVNLTTRTQLPPANTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QHZ00379        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
MT039890        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QHU79173        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLYSTQDLFLPFFS   60
MT126808        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
MT066156        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
LC528232        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
NC_045512       MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
MT240479        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
MT066175        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
MT192772        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
YP009724390     MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
QHR84449        MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS   60
                *************************** ********************:***********

QIC53204        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
MT012098        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QIQ08810        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QIO04367        NVTWFHAIHVSGTKGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QIK50427        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QIJ96493        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QII87830        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QII57278        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QIA98583        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QIA20044        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QHZ00379        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
MT039890        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QHU79173        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
MT126808        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
MT066156        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
LC528232        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
NC_045512       NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
MT240479        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
MT066175        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
MT192772        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
YP009724390     NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
QHR84449        NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV   120
                *************:**********************************************

QIC53204        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
MT012098        NNATNVVIKVCEFQFCNDPFLGVY-HKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   179
QIQ08810        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QIO04367        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QIK50427        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QIJ96493        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QII87830        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QII57278        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESELRVYSSANNCTFEYVSQPFLMDLE   180
QIA98583        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QIA20044        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QHZ00379        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
MT039890        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QHU79173        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
MT126808        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
MT066156        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
LC528232        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
NC_045512       NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
MT240479        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
MT066175        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
MT192772        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
YP009724390     NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
QHR84449        NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE   180
                ************************ ***********:***********************
```



```
QIC53204         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
MT012098         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   239
QIQ08810         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QIO04367         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QIK50427         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QIJ96493         VKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QII87830         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QII57278         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QIA98583         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QIA20044         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QHZ00379         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFWALEPLVDLPIGINITRFQT   240
MT039890         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFWALEPLVDLPIGINITRFQT   240
QHU79173         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
MT126808         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
MT066156         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
LC528232         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
NC_045512        GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
MT240479         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
MT066175         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
MT192772         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
YP009724390      GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
QHR84449         GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT   240
                  *************************************** *******************

QIC53204         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
MT012098         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   299
QIQ08810         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QIO04367         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QIK50427         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QIJ96493         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QII87830         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QII57278         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QIA98583         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QIA20044         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QHZ00379         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
MT039890         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QHU79173         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
MT126808         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
MT066156         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
LC528232         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
NC_045512        LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
MT240479         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
MT066175         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
MT192772         LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
YP009724390      LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
QHR84449         LLALHRRYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK   300
                 ****** *****************************************************

QIC53204         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
MT012098         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   359
QIQ08810         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QIO04367         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QIK50427         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QIJ96493         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QII87830         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QII57278         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QIA98583         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QIA20044         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QHZ00379         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
MT039890         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QHU79173         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
MT126808         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
MT066156         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
LC528232         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
NC_045512        CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
MT240479         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
MT066175         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
MT192772         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
YP009724390      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
QHR84449         CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN   360
                 ************************************************************
```



```
QIC53204        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
MT012098        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVIQIAPGQTGKIAD   419
QIQ08810        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QIO04367        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QIK50427        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QIJ96493        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QII87830        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QII57278        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QIA98583        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QIA20044        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QHZ00379        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
MT039890        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QHU79173        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
MT126808        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
MT066156        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
LC528232        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
NC_045512       CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
MT240479        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
MT066175        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
MT192772        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
YP009724390     CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
QHR84449        CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD   420
                *********************************************** ************

QIC53204        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
MT012098        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   479
QIQ08810        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QIO04367        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QIK50427        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QIJ96493        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QII87830        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QII57278        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QIA98583        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QIA20044        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QHZ00379        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
MT039890        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QHU79173        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
MT126808        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
MT066156        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
LC528232        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
NC_045512       YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
MT240479        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
MT066175        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
MT192772        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
YP009724390     YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
QHR84449        YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC   480
                ************************************************************

QIC53204        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
MT012098        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   539
QIQ08810        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPXKSTNLVKNKCVN   540
QIO04367        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QIK50427        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QIJ96493        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QII87830        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QII57278        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QIA98583        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QIA20044        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QHZ00379        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
MT039890        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QHU79173        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
MT126808        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
MT066156        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
LC528232        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
NC_045512       NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
MT240479        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
MT066175        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
MT192772        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
YP009724390     NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
QHR84449        NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN   540
                *********************************************** ************
```



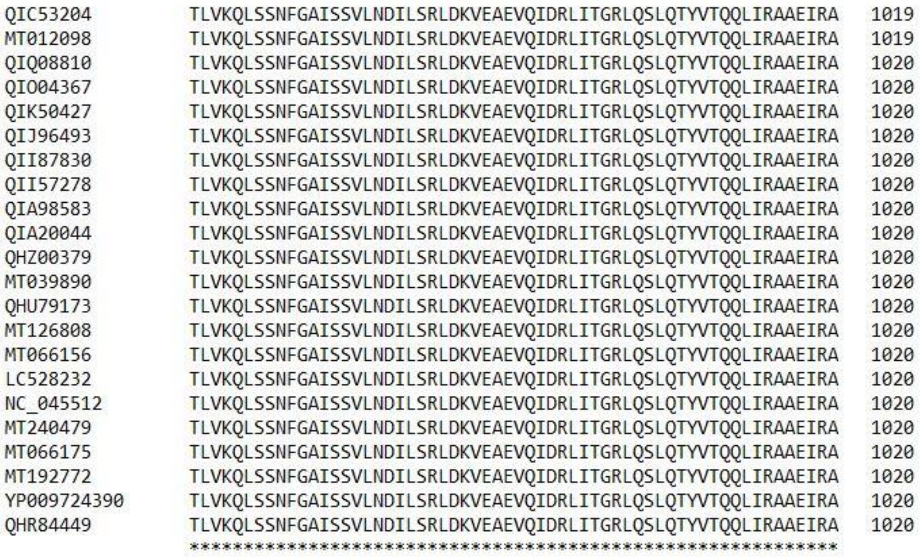
```
QIC53204        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
MT012098        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   599
QIQ08810        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QIO04367        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QIK50427        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QIJ96493        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QII87830        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QII57278        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QIA98583        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QIA20044        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QHZ00379        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
MT039890        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QHU79173        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
MT126808        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
MT066156        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
LC528232        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
NC_045512       FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
MT240479        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
MT066175        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
MT192772        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
YP009724390     FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QHR84449        FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP   600
QIC53204        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
MT012098        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   659
QIQ08810        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QIO04367        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QIK50427        GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QIJ96493        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QII87830        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEYVNNSY   660
QII57278        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QIA98583        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QIA20044        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QHZ00379        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
MT039890        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QHU79173        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
MT126808        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
MT066156        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
LC528232        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
NC_045512       GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
MT240479        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
MT066175        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
MT192772        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
YP009724390     GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
QHR84449        GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY   660
                *************.****************************************:*****

QIC53204        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
MT012098        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   719
QIQ08810        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QIO04367        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QIK50427        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QIJ96493        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QII87830        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QII57278        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QIA98583        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QIA20044        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QHZ00379        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
MT039890        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QHU79173        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
MT126808        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
MT066156        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
LC528232        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
NC_045512       ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
MT240479        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
MT066175        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
MT192772        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
YP009724390     ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
QHR84449        ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI   720
                ************************************************************
```

```
QIC53204        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
MT012098        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   779
QIQ08810        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QIO04367        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QIK50427        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QIJ96493        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QII87830        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QII57278        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QIA98583        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QIA20044        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QHZ00379        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
MT039890        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QHU79173        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
MT126808        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
MT066156        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
LC528232        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
NC_045512       SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
MT240479        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
MT066175        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
MT192772        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
YP009724390     SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
QHR84449        SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE   780
                ************************************************************
```

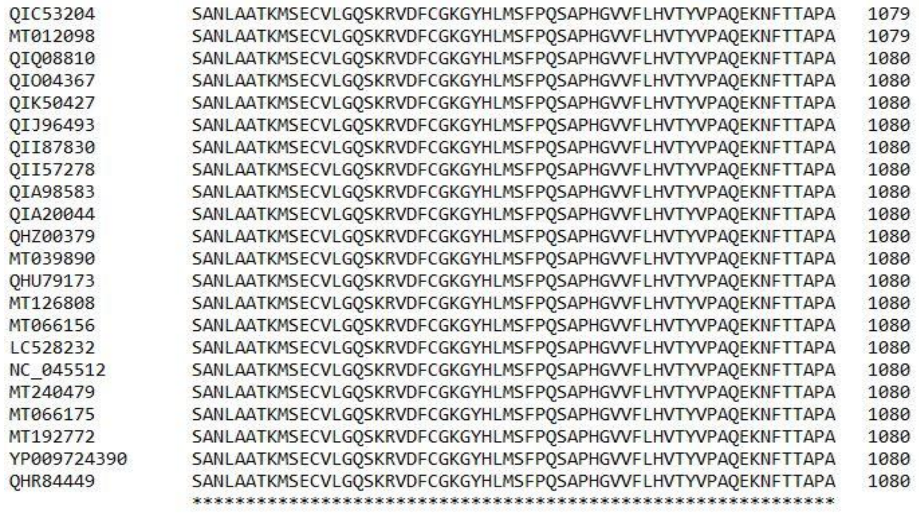
```
QIC53204        VFAQVKQIYKTPPIKDCGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
MT012098        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   839
QIQ08810        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QIO04367        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QIK50427        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QIJ96493        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QII87830        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QII57278        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QIA98583        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QIA20044        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QHZ00379        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
MT039890        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QHU79173        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
MT126808        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
MT066156        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
LC528232        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
NC_045512       VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
MT240479        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
MT066175        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
MT192772        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
YP009724390     VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
QHR84449        VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC   840
                **************** *******************************************
```

```
QIC53204        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
MT012098        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   899
QIQ08810        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QIO04367        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QIK50427        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QIJ96493        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QII87830        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QII57278        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QIA98583        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QIA20044        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QHZ00379        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
MT039890        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QHU79173        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
MT126808        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
MT066156        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
LC528232        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
NC_045512       LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
MT240479        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
MT066175        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
MT192772        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
YP009724390     LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
QHR84449        LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM   900
                ************************************************************
```

```
QIC53204        QMAYRFNGI-VTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   959
MT012098        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   959
QIQ08810        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QIO04367        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QIK50427        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QIJ96493        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QII87830        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QII57278        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QIA98583        QMAYRFNGIGVTQNVLYENQKLIANQFNSVIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QIA20044        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QHZ00379        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
MT039890        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QHU79173        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
MT126808        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
MT066156        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
LC528232        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
NC_045512       QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
MT240479        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
MT066175        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
MT192772        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
YP009724390     QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
QHR84449        QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN   960
                ********* ******************* ******************************
```



```
QIC53204        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1019
MT012098        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1019
QIQ08810        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QIO04367        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QIK50427        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QIJ96493        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QII87830        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QII57278        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QIA98583        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QIA20044        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QHZ00379        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
MT039890        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QHU79173        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
MT126808        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
MT066156        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
LC528232        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
NC_045512       TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
MT240479        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
MT066175        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
MT192772        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
YP009724390     TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
QHR84449        TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA   1020
                ************************************************************
```



```
QIC53204        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1079
MT012098        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1079
QIQ08810        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QIO04367        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QIK50427        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QIJ96493        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QII87830        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QII57278        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QIA98583        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QIA20044        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QHZ00379        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
MT039890        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QHU79173        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
MT126808        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
MT066156        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
LC528232        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
NC_045512       SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
MT240479        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
MT066175        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
MT192772        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
YP009724390     SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
QHR84449        SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA   1080
                ************************************************************
```

```
QIC53204        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1199
MT012098        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1199
QIQ08810        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QIO04367        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QIK50427        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QIJ96493        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QII87830        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QII57278        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QIA98583        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QIA20044        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QHZ00379        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
MT039890        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QHU79173        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
MT126808        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
MT066156        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
LC528232        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
NC_045512       LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
MT240479        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
MT066175        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
MT192772        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
YP009724390     LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
QHR84449        LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL   1200
                ************************************************************
```

```
QIC53204        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1259
MT012098        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1259
QIQ08810        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QIO04367        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QIK50427        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QIJ96493        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QII87830        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QII57278        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QIA98583        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QIA20044        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QHZ00379        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
MT039890        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QHU79173        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
MT126808        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
MT066156        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
LC528232        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
NC_045512       QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
MT240479        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
MT066175        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
MT192772        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
YP009724390     QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
QHR84449        QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD   1260
                ************************************************************
```

```
QIC53204        SEPVLKGVKLHYT  1272
MT012098        SEPVLKGVKLHYT  1272
QIQ08810        SEPVLKGVKLHYT  1273
QIO04367        SEPVLKGVKLHYT  1273
QIK50427        SEPVLKGVKLHYT  1273
QIJ96493        SEPVLKGVKLHYT  1273
QII87830        SEPVLKGVKLHYT  1273
QII57278        SEPVLKGVKLHYT  1273
QIA98583        SEPVLKGVKLHYT  1273
QIA20044        SEPVLKGVKLHYT  1273
QHZ00379        SEPVLKGVKLHYT  1273
MT039890        SEPVLKGVKLHYT  1273
QHU79173        SEPVLKGVKLHYT  1273
MT126808        SEPVLKGVKLHYT  1273
MT066156        SEPVLKGVKLHYT  1273
LC528232        SEPVLKGVKLHYT  1273
NC_045512       SEPVLKGVKLHYT  1273
MT240479        SEPVLKGVKLHYT  1273
MT066175        SEPVLKGVKLHYT  1273
MT192772        SEPVLKGVKLHYT  1273
YP009724390     SEPVLKGVKLHYT  1273
QHR84449        SEPVLKGVKLHYT  1273
                *************
```