Genomic analysis and comparative multiple sequences of SARS-CoV2

Tai-Jay Chang[a,b,*], De-Ming Yang[c,d,e], Mong-Lien Wang[f,g], Kung-How Liang[g,h], Ping-Hsing Tsai[i], Shih-Hwa Chiou[i,j], Ta-Hsien Lin[k,l], Chin-Tien Wang[m,n]

[a]Laboratory of Genome Research, Basic Research Division, Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [b]School of Biomedical Science and Engineering, National Yang-Ming University, Taipei, Taiwan, ROC; [c]Microscopy Service Laboratory, Basic Research Division, Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [d]Institute of Biophotonics, School of Medical Technology and Engineering, National Yang-Ming University, Taipei, Taiwan, ROC; [e]Biophotonics and Molecular Imaging Research Center (BMIRC), National Yang-Ming University, Taipei, Taiwan, ROC; [f]Laboratory of Molecular Oncology, Basic Research Division, Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [g]Institute of Food Safety and Health Risk Assessment, National Yang-Ming University, Taipei, Taiwan, ROC; [h]Laboratory of Systems Biomedical Science, Basic Research Division, Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [i]Department of Medical Research, Cell Therapy Innovation Center, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [j]Institute of Pharmacology, School of Pharmaceutical Science, National Yang-Ming University, Taipei, Taiwan, ROC; [k]Laboratory of Nuclear Magnetic Resonance, Basic Research Division, Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [l]Institute of Biomedical Informatics, School of Medicine, National Yang-Ming University, Taipei, Taiwan, ROC; [m]Laboratory of Molecular Virology, Basic Research Division, Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [n]Institute of Clinical Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan, ROC

Abstract

Background: China announced an outbreak of new coronavirus in the city of Wuhan on December 31, 2019; lash to now, the virus transmission has become pandemic worldwide. Severe cases from the Huanan Seafood Wholesale market in Wuhan were confirmed pneumonia with a novel coronavirus (2019-nCoV). Understanding the molecular mechanisms of genome selection and packaging is critical for developing antiviral strategies. Thus, we defined the correlation in 10 severe acute respiratory syndrome coronavirus (SARS-CoV2) sequences from different countries to analyze the genomic patterns of disease origin and evolution aiming for developing new control pandemic processes.

Methods: We apply genomic analysis to observe SARS-CoV2 sequences from GenBank (http://www.ncbi.nim.nih.gov/genebank/): MN 908947 (China, C1), MN985325 (USA: WA, UW), MN996527 (China, C2), MT007544 (Australia: Victoria, A1), MT027064 (USA: CA, UC), MT039890 (South Korea, K1), MT066175 (Taiwan, T1), MT066176 (Taiwan, T2), LC528232 (Japan, J1), and LC528233 (Japan, J2) for genomic sequence alignment analysis. Multiple Sequence Alignment by Clustalw (https://www.genome.jp/tools-bin/clustalw) web service is applied as our alignment tool.

Results: We analyzed 10 sequences from the National Center for Biotechnology Information (NCBI) database by genome alignment and found no difference in amino acid sequences within M and N proteins. There are two amino acid variances in the spike (S) protein region. One mutation found from the South Korea sequence is verified. Two possible "L" and "S" SNPs found in ORF1ab and ORF8 regions are detected.

Conclusion: We performed genomic analysis and comparative multiple sequences of SARS-CoV2. Studies about the biological symptoms of SARS-CoV2 in clinic animals and humans will manipulate an understanding on the origin of pandemic crisis.

Keywords: Genomic analysis; Multiple sequences; Severe acute respiratory syndrome coronavirus 2

**Address correspondence: Dr. Tai-Jay Chang, Genome Research Laboratory, Basic Research Division, Department of Medical Research, Taipei Veterans General Hospital, 201, Section 2, Shi-Pai Road, Taipei 112, Taiwan, ROC. E-mail address: tjchang@vghtpe.gov.tw (T.-J.Chang)*

Conflicts of interest: The authors declare that they have no conflicts of interest related to the subject matter or materials discussed in this article.

Journal of Chinese Medical Association. (2020) 83: 537-543.

Received April 4, 2020; accepted April 8, 2020.

doi: 10.1097/JCMA.0000000000000335.

1. INTRODUCTION

The severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV) transmitted from animals to humans have caused severe pneumonia in the world.[1] SARS emerged in 2002 in Guangdong, China, and its subsequent global spread resulted in 8096 infected cases and 774 deaths.[2] After China announced an outbreak of a new coronavirus in the city of Wuhan on December 31, 2019, lash to now, the world has become pandemic. Severe cases from the Huanan Seafood Wholesale market in Wuhan were confirmed pneumonia with the infection of a novel coronavirus

(2019-nCoV),[3] named SARS-CoV2 (International Committee on Taxonomy of Viruses).[4,5]

Classification of coronavirus divided into subfamily of the Coronaviridae: four genera (α, β, ρ and δ). Both SARS-CoV2 (2019) and SARS-CoV1 (2003) belong to coronavirus.[6] Structural biology analysis of Wuhan coronavirus SARS-CoV2 divided it into ORF1ab, S, ORF3, E, M, ORF6, 7a, 8, and ORF10. The spike (S) protein allows the virus to attach the membrane of the host cell and the N (nucleocapsid) protein holds the virus RNA genome. The E (envelope) and M (membrane) alone with S protein form a viral envelope.[7] The nonstructural RNA genome of ORF1ab, ORF3, ORF6, 7a, 8, and ORF10 contains highly conserved information for genome replication.[8]

The coronavirus transmission starts from attaching host cell membrane receptors before endocytosis to enter host cells. RNA gene1 of virus genome then begins its replication and synthesizes the subgenomic RNAs with new transcription afterward. Since then, N protein and new genomic RNA assemble to form helical nucleocapsids, which will interact with M protein inserted in endoplasmic reticulum and anchored in host Golgi. E and M proteins then begin to trigger budding processes. S together with helical N translates on membrane-bound polysomes rough endoplasmic reticulum and is transported to Golgi. Finally, virions are released by exocytosis to finish the life cycle and replication of the virus.[9]

SARS-CoV1 transmits possibly through bat and civet as intermediate hosts, and finally to human with the symptoms of severe respiratory impacts and 10% mortality rate. However, Wuhan SARS-CoV2 is suspected to be transmitted from bat (RaTG13) to pangolin as intermediate host before it is transmitted to humans by some unknown mechanisms with symptoms of severe respiratory impacts and unclear mortality rate.[10] The genomic sequence of RaTG13 cited the 96% similarity with the Wuhan coronavirus.[11] Although intermediate host is not clear, genomic sequence comparison obviously demonstrates S receptor-binding domain (RBD) of Wuhan SARS-CoV2 with the similarity in 90% homolog of pangolin. Then, pangolin might contribute the S protein region to be cross-transmitted to RATG13 to a recombinant new mutant Wuhan SARS-CoV2 to transmit to human finally.[12]

The S protein of SARS-CoV1 and SARS-CoV2 responsible for viral entry mediates the binding to host cell membrane of angiotensin-converting enzyme 2 (ACE2) through its RBD.[13] The surface glycoprotein of spike of SARS-CoV comprises two components: S1 and S2. The S protein of SARS-CoV2 binds to the host receptor ACE2 through its S1 subunit, which contains RBD, followed by fusing the viral and host membranes through the S2 subunit, which contains the fusion peptide primed by host protease. To prime the fusion of viral membrane with host membrane, SARS-S is cleaved by a cellular protease called cathepsin L, thereby exposing the S2 domain of the S protein for membrane fusion, followed by endocytosis and forms low pH endosomes.[14]

Understanding the molecular mechanisms of genome selection and packaging is critical for developing antiviral strategies. Thus, in this report, we compared SARS-CoV2 sequences from different countries to analyze the genomic patterns of disease origin and evolution, providing genomic information for the development of new control methods against the worldwide SARS-CoV2 pandemic.

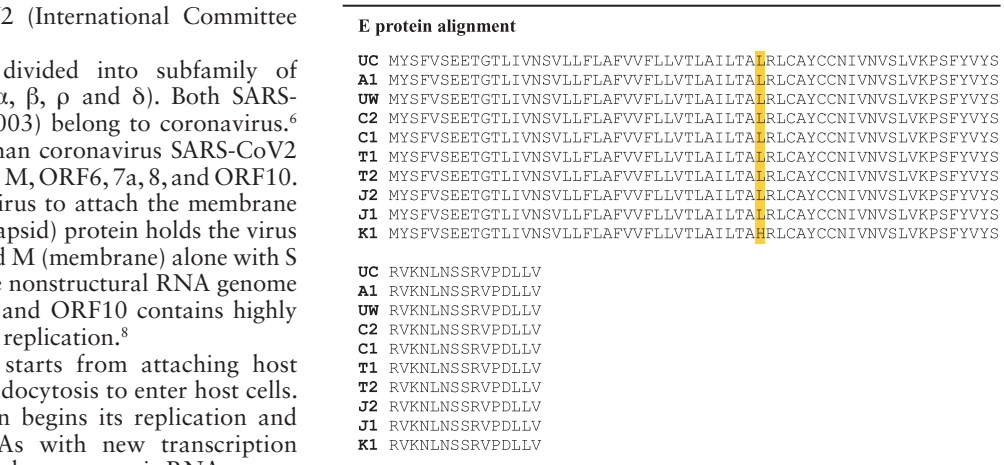

Fig. 1 Genomic analysis of E protein amino acid sequence. We found one amino acid mutation at "H" from South Korea comparing the "L" from other nine sequences. Yellow line indicates the difference in 10 sequence alignments.

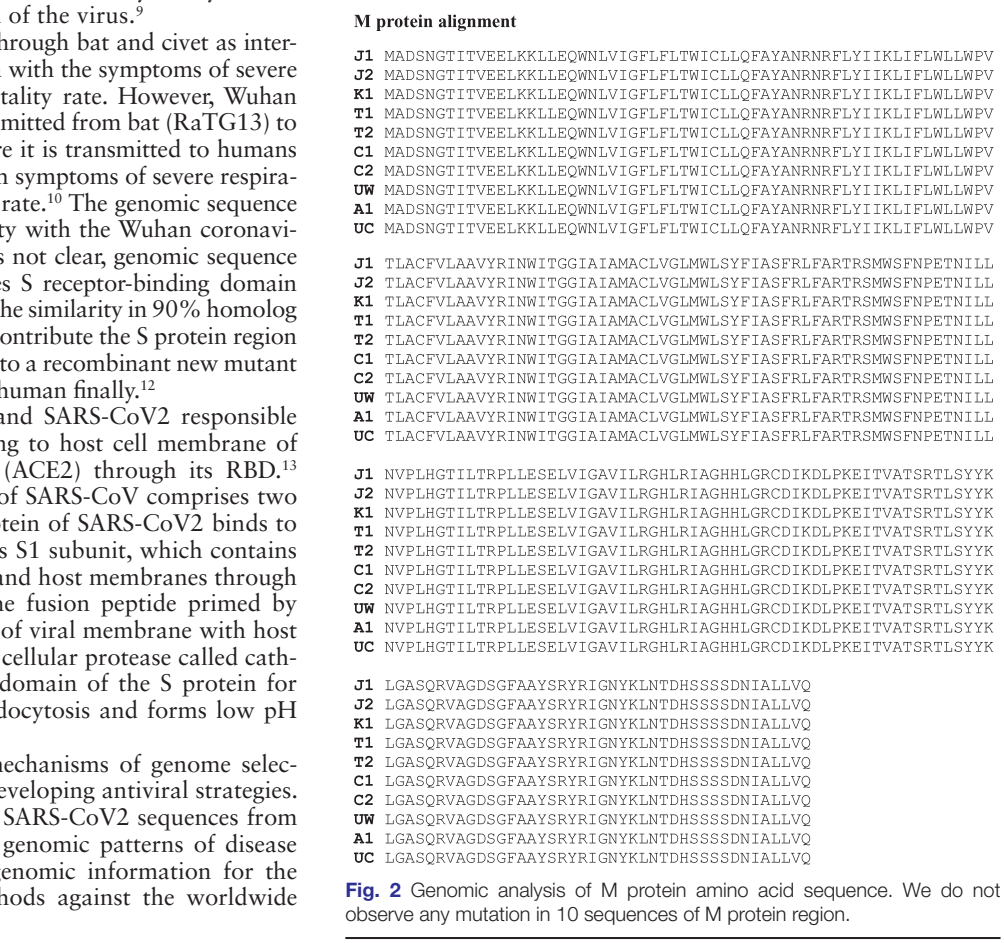

Fig. 2 Genomic analysis of M protein amino acid sequence. We do not observe any mutation in 10 sequences of M protein region.

2. METHODS

2.1. Sequence resource

Studies focus on evolutionary and phylogenetic analyses have applied in disease progression for Wuhan lung pneumonia treatment. Herein, we apply genomic analysis to observe SARS-CoV2 sequences from GenBank (http://www.ncbi.nim.nih.gov/genebank/): MN 908947 (China, C1), MN985325 (USA: WA, UW), MN996527 (China, C2), MT007544 (Australia: Victoria, A1), MT027064 (USA: CA, UC), MT039890 (South Korea, K1), MT066175 (Taiwan, T1), MT066176 (Taiwan, T2), LC528232

N protein alignment

```
J1 MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
J2 MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
K1 MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
T1 MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
T2 MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
C1 MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
C2 MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
UW MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
A1 MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
UC MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG

J1 KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
J2 KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
K1 KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
T1 KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
T2 KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
C1 KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
C2 KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
UW KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
A1 KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
UC KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG

J1 LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
J2 LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
K1 LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
T1 LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
T2 LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
C1 LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
C2 LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
UW LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
A1 LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
UC LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS

J1 QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
J2 QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
K1 QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
T1 QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
T2 QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
C1 QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
C2 QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
UW QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
A1 QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
UC QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ

J1 QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
J2 QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
K1 QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
T1 QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
T2 QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
C1 QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
C2 QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
UW QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
A1 QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
UC QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH

J1 WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
J2 WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
K1 WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
T1 WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
T2 WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
C1 WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
C2 WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
UW WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
A1 WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
UC WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY

J1 KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
J2 KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
K1 KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
T1 KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
T2 KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
C1 KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
C2 KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
UW KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
A1 KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
UC KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
```

Fig. 3 Genomic analysis of N protein amino acid sequence. We do not observe any mutation in 10 sequences of N protein region.

(Japan, J1), and LC528233 (Japan, J2) for genomic sequence alignment analysis.

2.2. Method applied

Multiple Sequence Alignment by Clustalw (https://www.genome.jp/tools-bin/clustalw) web service is applied as our alignment tool.

S protein alignment

```
J1 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
J2 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
UW MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
UC MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
A1 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
T2 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
C1 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
C2 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
T1 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
K1 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS

J1 NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
J2 NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
UW NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
UC NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
A1 NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
T2 NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
C1 NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
C2 NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
T1 NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
K1 NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV

J1 NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
J2 NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
UW NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
UC NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
A1 NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
T2 NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
C1 NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
C2 NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
T1 NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
K1 NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE

J1 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
J2 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
UW GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
UC GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
A1 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
T2 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
C1 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
C2 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
T1 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGININTRFQT
K1 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFWALEPLVDLPIGININTRFQT

J1 LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
J2 LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
UW LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
UC LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
A1 LLALHRRYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
T2 LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
C1 LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
C2 LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
T1 LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
K1 LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK

J1 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
J2 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
UW CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
UC CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
A1 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
T2 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
C1 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
C2 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
T1 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
K1 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN

J1 CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
J2 CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
UW CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
UC CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
A1 CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
T2 CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
C1 CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
C2 CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
T1 CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
K1 CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD

J1 YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
J2 YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
UW YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
UC YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
A1 YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
T2 YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
C1 YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
C2 YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
T1 YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
K1 YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
```

Fig. 4 Genomic analysis of S protein amino acid sequence. One amino acid mutation at "W" from South Korea comparing "S" in other nine sequences. One amino acid mutation at "R" from Australia was observed comparing "S" from another nine sequences. Two yellow lines indicate the difference in 10 sequence alignments.

```
J1 NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
J2 NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
UW NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
UC NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
A1 NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
T2 NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
C1 NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
C2 NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
T1 NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
K1 NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN

J1 FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
J2 FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
UW FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
UC FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
A1 FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
T2 FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
C1 FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
C2 FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
T1 FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
K1 FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP

J1 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
J2 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
UW GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
UC GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
A1 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
T2 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
C1 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
C2 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
T1 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
K1 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY

J1 ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
J2 ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
UW ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
UC ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
A1 ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
T2 ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
C1 ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
C2 ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
T1 ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
K1 ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI

J1 SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
J2 SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
UW SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
UC SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
A1 SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
T2 SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
C1 SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
C2 SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
T1 SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
K1 SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE

J1 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
J2 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
UW VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
UC VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
A1 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
T2 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
C1 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
C2 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
T1 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
K1 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC

J1 LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
J2 LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
UW LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
UC LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
A1 LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
T2 LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
C1 LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
C2 LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
T1 LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
K1 LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM

J1 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
J2 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
UW QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
UC QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
A1 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
T2 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
C1 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
C2 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
T1 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
K1 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN

J1 TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
J2 TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
UW TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
UC TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
A1 TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
T2 TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
C1 TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
C2 TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
T1 TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
K1 TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA

J1 SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
J2 SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
UW SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
UC SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
A1 SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
T2 SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
C1 SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
C2 SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
T1 SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
K1 SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA

J1 ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
J2 ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
UW ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
UC ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
A1 ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
T2 ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
C1 ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
C2 ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
T1 ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
K1 ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP

J1 LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
J2 LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
UW LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
UC LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
A1 LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
T2 LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
C1 LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
C2 LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
T1 LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
K1 LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL

J1 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
J2 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
UW QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
UC QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
A1 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
T2 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
C1 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
C2 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
T1 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
K1 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD

J1 SEPVLKGVKLHYT
J2 SEPVLKGVKLHYT
UW SEPVLKGVKLHYT
UC SEPVLKGVKLHYT
A1 SEPVLKGVKLHYT
T2 SEPVLKGVKLHYT
C1 SEPVLKGVKLHYT
C2 SEPVLKGVKLHYT
T1 SEPVLKGVKLHYT
K1 SEPVLKGVKLHYT
```

Fig. 4 (Continued)

3. RESULTS

3.1. E protein

The structures reveal that E has a short and hydrophilic N-amino terminus consisting of 7–12 amino acids, followed by a large hydrophobic transmembrane domain of 25 amino acids, and ends with a long, hydrophilic C-carboxyl terminus (C-terminal), which comprises the majority of the protein.[15] By analyzing E protein alignment, one amino acid mutation "H" was observed from South Korea (K1) comparing the "L" from other nine sequences in E protein sequence alignment.

3.2.S protein

S protein mediates attachment of SARS-CoV1 to the host cell surface receptors and subsequent fusion between them to

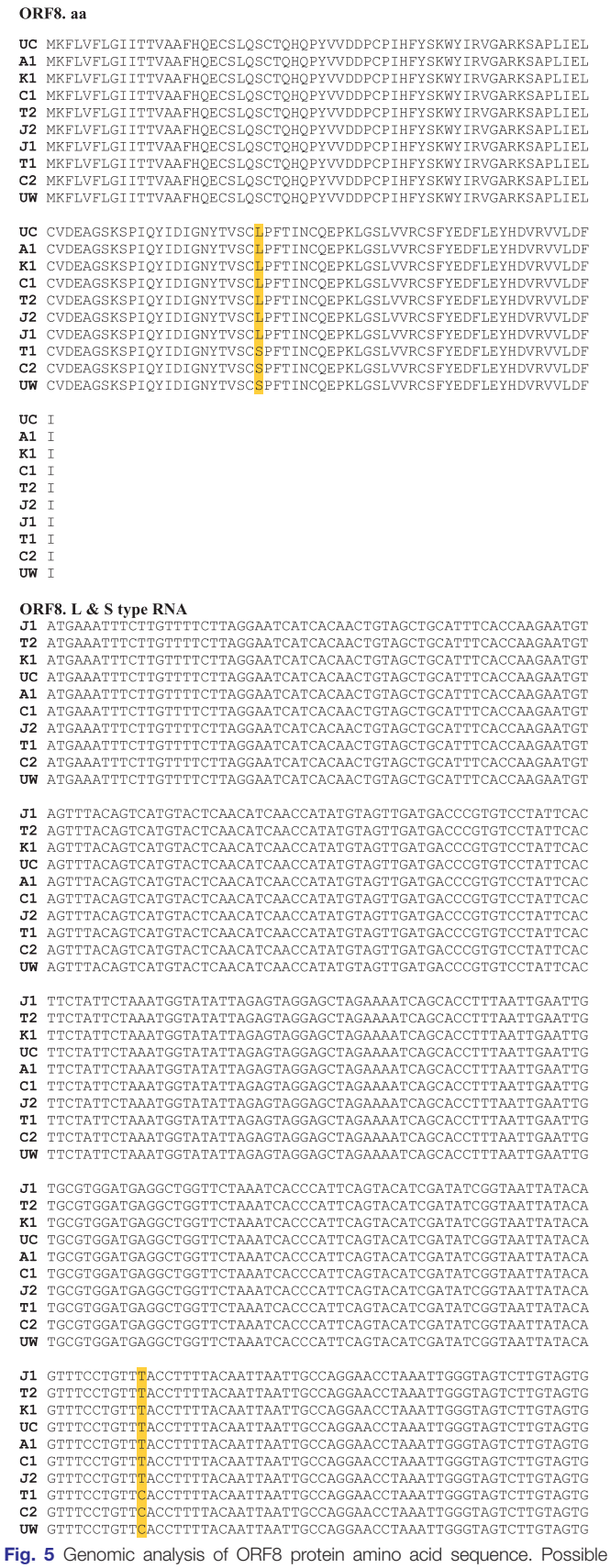

Fig. 5 Genomic analysis of ORF8 protein amino acid sequence. Possible subtypes were found in OFR8 with L and S subtypes. "L" type appeared in UC, A1, K1, C1, T2, J2, and T1. "S" type was observed in T1, C2, and UW. Yellow line indicates the difference in 10 sequence alignments.

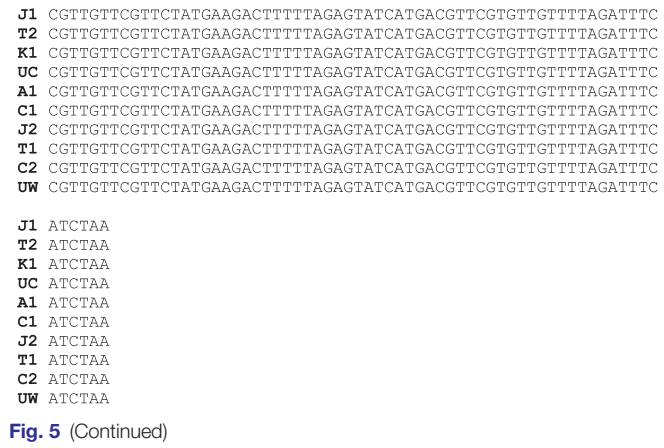

Fig. 5 (Continued)

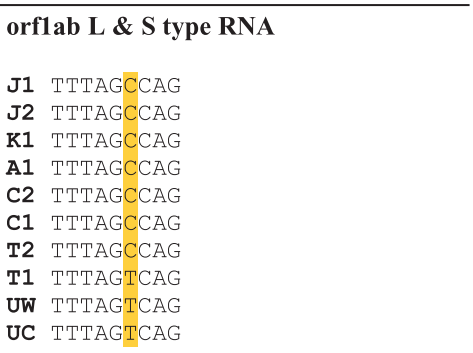

Fig. 6 Genomic analysis of orf1ab protein amino acid sequence. Genomic analysis regions of orf1ab RNA sequence. Yellow line indicates the difference in 10 RNA sequence alignment.

facilitate viral entry into the host cell.[15] The expression of S protein at the cell membrane can mediate cell–cell fusion. This formation supports to offer a strategy to let spread the virus between cells to subvert function of virus-neutralizing antibodies mechanisms.

During analysis of S protein, one amino acid mutation at "W' was observed from South Korea (K1) comparing "S" from other nine sequences. One amino acid mutation at "R" from Australia (A1) was observed comparing "S" from another nine sequences.

3.3. M and N proteins

The M protein is abundant which defines the shape of the viral envelope. N functions primarily to bind to RNA genome of SARS-CoV, making up the nucleocapsid.[15] Although N protein is most involved in processes viral genome signaling, it is also involved RNA replication cycle with host cellular response to viral infection.

Although the sequence difference between SARS-CoV1 and SARS-CoV2 within M and N proteins, there is no SNP variant observed between M and N protein sequence alignments from different patients.

3.4. L and S subtypes

Possible subtypes suggested in reference to two subtypes were found in OFR8 with L and S subtypes. By alignment, leucine of "L" type appeared in UC, A1, K1, C1, T2, J2, and T1, while as, Serine of "S" type was observed in T1, C2, and UW.

Possible subtypes suggested in reference to two subtypes were found in OFR1ab with L and S subtypes. By alignment, RNA sequence "C" type appeared in J1, J2, K1, A1, C1, C2, and T2 "T" type was observed in T1, UW, and UC.

4. DISCUSSION

4.1. Point mutation

Although SARS-CoV1 and SARS-CoV2 share the sequence similarity with 80% homolog, their functions are various. In comparison to 10 strains from 10 patients in structural protein regions, one mutation is observed by analyzing E protein alignment, the amino acid mutation "H" was observed from South Korea (K1) comparing the "L" from other nine sequences in E protein sequence alignment.

Inside the envelope, there is the nucleocapsid, which is formed from multiple copies of the nucleocapsid (N) protein, which are bound to the positive-sense single-stranded RNA genome in a continuous beads-on-a-string type conformation.[16] The lipid bilayer envelope, membrane proteins, and nucleocapsid protect the virus when it is outside the host cell.[17]

Although the N protein holds the RNA genome, and M protein with E and S proteins together creates the viral envelope to protect the virus when it is outside the host cell, we do not find any point mutation of M and E proteins within 10 sequences.

The M protein is the most abundant structural protein and defines the shape of the viral envelope. Binding of M to N stabilizes the nucleocapsid (N protein–RNA complex), as well as the internal core of virions, and ultimately, promotes completion of viral assembly.[18]

During analysis of S protein, one amino acid mutation at "W" was observed from South Korea (K1) comparing "S" from other nine sequences. One amino acid mutation at "R" from Australia (A1) was observed comparing "S" from another nine sequences. Report[19] mentioned a single amino acid reversion (L294-to-Q) in the S protein is sufficient to abrogate the phenotype and grows well at and below 32°C.

4.2. Spike protein and receptor (ACE)

A novel and pathogenic SARS-CoV2 was found in Wuhan, China in 2019, and its rapid national and international spread poses a global health emergency. The S protein mediates viral entry into host cells by first binding to a host receptor through the RBD in the S1 subunit and then fusing the viral and host membranes through the S2 subunit priming by host cell proteases.[20–23] Unraveling which cellular factors are used by SARS-CoV-2 for entry might provide insights into viral transmission and reveals therapeutic targets. SARS-CoV and MERS-CoV RBDs recognize different receptors. SARS-CoV recognizes ACE2 as its receptor, whereas MERS-CoV recognizes dipeptidyl peptidase 4 as its receptor.[13,24] Since SARS-CoV2 recognizes ACE2 as its host receptor binding to viral S protein,[25] it is critical to define the RBD in SARS-CoV2 S protein in the most likely target for the mechanism of virus attachment such as new developing inhibitors, neutralizing antibodies, and vaccines.

Authors from Tai group[26] demonstrate by characterizing SARS-CoV2 RBD to display a multiple sequence alignment of RBDs of SARS-CoV2, SARS-CoV, and MERS-CoV S proteins.

They identified the RBD in SARS-CoV2 S protein, and found that the RBD protein bound strongly to human and bat ACE2 receptors. SARS-CoV2 RBD displayed a significantly higher binding affinity to ACE2 receptor than SARS-CoV RBD. Subsequently, SARS-CoV RBD-specific antibodies could cross-react with SARS-CoV2 RBD protein. Meanwhile, SARS-CoV RBD-induced antisera which could cross-neutralize SARS-CoV2 suggests the potentials to develop SARS-CoV RBD-based vaccines for prevention of SARS-CoV2 and SARS-CoV infection.[26]

Hoffmann group mentions SARS-CoV1 and SARS-CoV2 share 76% amino acid identity in S protein region. By the amino acid alignment, they observe the receptor-binding motif of SARS-CoV1 corresponding to the sequences of bat-associated beta-coronavirus S proteins. Demonstration of high or low similarity by taking advantage of ACE2 as cellular receptor reveals that SARS-CoV2 possesses crucial amino acid residues for ACE2 binding.

They also found similarity signal to points out between SARS-CoV2 and SARS-CoV1 during transmitting host cells stage and then identified a potential target for antiviral intervention. Inspecting conserved amino acids within ACE2 domain, Hoffmann group performed SARS-CoV2 to transmit cell entry depends on ACE2 and transmembrane serine protease 2 two proteins and is blocked by applied clinically-proven protease inhibitor.[27]

4.3. SNP or subtype

We found that SNPs at locations 8782 (orf1ab: T8517C, synonymous) and 28 144 (ORF8: C251T, S84L) showed possible linkage in 10 sequences from different countries. As report "On the origin and continuing evolution of SARS-CoV-2" emphasized two subtypes of "L" and "S" from their data as they exhibited a "CT" haplotype (defined as "L" type because T28 144 is in the codon of leucine) and other "TC" haplotype (defined as "S" type because C28 144 is in the codon of serine) at these two sites.[28] The authors show S is ancestral, related viruses like bat (RaTG13). They also depict that L is more prevalent with progressive, especially in Wuhan.[29]

However, according to data, it is too early to speculate on such consequences because there is no evidence whether it will affect some strategies such as vaccination. As mutation doesn't occur within the S1 spike protein domain to influence the antigen targeting for vaccine production.

As speculated, SARS-CoV2 RNA viruses cross species barriers into humans; they would not be well-adapted to host cells. They should modify and allow them to adapt and become able to replicate within, and broad transmit humans. However, we do not catch data about testing the relative replication cycle in human cells. It would be difficult to verify human interference with the impact of co-strains relatively. Thus, a balance between SARS-CoV2 virulence and patient's genetics personal phenomena with environmental factors would be important to confirm the subtypes of origin and evolution.

Due to study limitations, we cannot handle the SARS-CoV2 biological study directly from patient specimens, which will not observe the correlation from clinical to laboratory analysis directly.

In conclusion, we analyzed 10 sequences from the NCBI database by genome alignment and found no difference in amino acid sequences within M and N proteins. There are two amino acid variances in S protein region. One mutation found from South Korea sequence is verified. Two possible "L" and "S" SNPs found in ORF1ab and ORF8 regions are detected. Since our data are limited to a small population, more studies about the biological symptoms of SARS-CoV2 in clinic animals and humans will manipulate an understanding on the origin of pandemic crisis.

ACKNOWLEDGMENTS

This research was funded by Taipei Veterans General Hospital; grant numbers V107E-002-2, V108D46-004-MY2-1, V108E-006-4, 108E-006-5, and 109VACS-003.

REFERENCES

1. Fehr AR, Channappanavar R, Perlman S. Middle East respiratory syndrome: emergence of a pathogenic human coronavirus. *Annu Rev Med* 2017;**68**:387–99.
2. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 2016;**14**:523–34.
3. WHO (2020). Novel Coronavirus (2019-nCoV) Situation Report 23. Available at https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200212-sitrep-23-ncov.pdf?sfvrsn=41e9fb78_4
4. "Coronavirus Disease 2019". World Health Organization. Retrieved 15 March 2020. 2020. Available at https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen
5. Naming the 2019 Coronavirus Feb 5, 2020. Available at https://talk.ictvonline.org
6. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, et al. The genome sequence of the SARS-associated coronavirus. *Science* 2003; **5624**:1399–404.
7. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003;**300**:1394–9.
8. Anderson KG, Rambaut A, Lipkin WI, Holes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nature Medicine* 2020 17 March.
9. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* 2015;**1282**:1–23.
10. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;**395**:497–506.
11. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020;**395**:514–23.
12. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* 2020;**9**:221–36.
13. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;**426**:450–4.
14. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. Available at https://doi.org/10.1101/2020.01.26.919985.
15. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virol J* 2019;**16**:69.
16. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein – forms and functions. *Antiviral Res* 2014;**103**:39–50.
17. Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, et al. A structural analysis of M protein in coronavirus assembly and morphology. *Journal Struct Biol* 2011;**174**: 11–22.
18. Escors D, Ortego J, Laude H, Enjuanes L. The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. *J Virol* 2001;**75**:1312–24.
19. Shen S, Law YC, Liu DX. A single amino acid mutation in the spike protein of coronavirus infectious bronchitis virus hampers its maturation and incorporation into virions at the nonpermissive temperature. *Virology* 2004;**326**:288–98.
20. Liu S, Xiao G, Chen Y, He Y, Niu J, Escalante CR, et al. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* 2004;**363**:938–47.
21. Wang Q, Wong G, Lu G, Yan J, Gao GF. MERS-CoV spike protein: Targets for vaccines and therapeutics. *Antiviral Res* 2016;**133**:165–77.
22. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005;**309**:1864–8.
23. Lu G, Hu Y, Wang Q, Qi J, Gao F, Li Y, et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 2013;**500**:227–31.
24. Raj VS, Mou H, Smits SL, Dekkers DH, Müller MA, Dijkman R, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 2013;**495**:251–4.
25. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;**579**:270–3.
26. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine *Cell mole Immunol* 2020 March 19. doi.org/10.1038/s41423-020-0400-4.
27. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler TErichsen S, et al., SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020 Mar 4. **pii**: S0092–8674(20)30229-4
28. Lam TTY, Shum MHH, Zhu HC, Tong YG, Ni XB, Liao YS, et al. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in Southern China. *Nature* 2020. Doi: 10.1038/s41586-020-2169-0
29. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2 *NSR* 2020 nwaa036, https://doi.org/10.1093/nsr/nwaa036