Mutation signature of SARS-CoV-2 variants raises questions to their natural origins.

Hiroshi Arakawa[1]*

[1]IFOM - FIRC Institute of Molecular Oncology Foundation, Milano, Italy

* Corresponding author:
IFOM - FIRC Institute of Molecular Oncology Foundation
IFOM-IEO Campus
Via Adamello 16
20139 Milano
Italy
tel: +39 02 574303306
fax: +39 02 574303231
e-mail: hiroshi.arakawa@ifom.eu

Keywords
SARS-CoV-2, Omicron, neutral evolution, synonymous substitution, nonsynonymous substitution

SARS-CoV-2 variants, such as Omicron, have acquired a number of mutations. These novel mutations have raised concerns regarding the continued efficacy of the antibodies generated in response to vaccination. In this study I show that the evolution of the SARS-CoV-2 variant Omicron, does not appear to follow a Darwinian trajectory. The mutations in the Omicron variant are heavily biased toward nonsynonymous substitutions rather than synonymous substitutions. Interestingly, mutations affecting the spike gene of the Omicron variant are almost exclusively nonsynonymous. Moreover, a nonsynonymous substitution bias within spike gene is a common feature of all of the SARS-CoV-2 variants assessed in this study. This mutational signature is a counter-evidence of neutral evolution, and shows that the spike genes of these SARS-CoV-2 variants have evolved without trial-and-error by mutation and selection. Thus I postulate that the spike genes of these SARS-CoV-2 variants are not the result of natural evolution but rather designed molecules.

Introduction

The SARS-CoV-2 Omicron variant (strain B.1.1.529) was first identified in Botswana[1], and reported to the World Health Organization (WHO) from South Africa on November 24, 2021[2]. The Omicron variant carries many novel mutations in its spike gene [3-5]. The large number of mutations, and their uniqueness led to concerns that the Omicron variant would be able to effectively evade the immune responses elicited by the existing COVID-19 vaccines[1].

Mutations in viral proteins are a normal part of viral evolution, and expected in the case of SARS-CoV-2 given its prevalence. However, the Omicron variant is unique as it acquired many mutations suddenly, a pattern of evolution more consistent with punctuated, rather than Darwinian evolution. Given this highly unusual evolutionary trajectory, the origins of the Omicron variant are an active topic of debate among scientists[1, 2]. The current hypotheses regarding the origins of the Omicron variant include: 1) It arose in immunosuppressed patients, chronically infected with COVID-19[6], 2) It slowly evolved over a period of months in a community with little or no viral surveillance infrastructure[1, 2], 3) the Omicron variant evolved in a non-human host[7] before spilling over into a human again with a new repertoire of mutations. These three central hypotheses have several problems.

If the Omicron variant evolved in immunosuppressed patients, one would expect it to bear other mutations as well, effectively rendering it less transmissible [2]. This does not appear to be the case; in fact the Omicron variant appears to be among the most transmissible known viruses. Similarly, if Omicron had evolved in a new animal host, it is unlikely that it would already be so transmissible from human to human. Finally, given the pandemic nature of COVID-19 and the global attention it has received it is unlikely that a sufficiently large, un-surveilled community exists wherein the Omicron could have evolved undetected over the course of months [2, 8]. Moreover, given Omicron's high rates of infectivity and transmissibility among COVID-19 vaccinated people, it appears that its mutations have enabled to circumvent vaccine-mediated immunity. Had the Omicron variant evolved in: animals, an un-surveilled community, or immunosuppressed patients there would have been little or no selective pressure on the virus to bypass vaccine mediated immunity. One recent paper suggests that the Omicron variant may have evolved in mice[7] however, there is no experimental evidence that it can be efficiently transmitted to mice.

To investigate the origins of the Omicron variant, it is necessary to verify whether it was established through a natural evolutionary process. The theory of neutral evolution states that

most changes are fixed by random genetic drift[9], and generally do not alter an organism's fitness. This theory applies to evolution at the molecular level, and is compatible with phenotypic evolution via natural selection. A synonymous substitution (also called a silent mutation) is a mutation on a codon that does not change the amino acid sequence. In contrast, a mutation that changes the amino acid sequence is called a nonsynonymous substitution, also known as a replacement mutation. The ratio of nonsynonymous to synonymous substitutions is used to estimate the balance between neutral, purifying selection, and beneficial mutations[10]. In this study, I investigate the origins of the SARS-CoV-2 variants by analyzing their molecular evolutionary trajectories.

Materials and Methods

Collection of genomes

Genomes of SARS-CoV-2 variants were retrieved from GISAID (https://www.gisaid.org/). The sequences of Wuhan SARS-CoV-2 were also retrieved from Genbank. The accession numbers are summarized in supplemental Table s1.

Genome analysis

SARS-CoV-2 genome sequences were aligned using CLC Genomics Workbench (QIAGEN, Aarhus, Denmark) and the alignment was further manually corrected. Phylogenetic trees were constructed using the neighbor-joining algorithm[11]. Jukes-Cantor was used for distance measurements. Bootstrap resampling was performed with 100 replications.

Mutation analysis

Each mutation was identified by aligning proto variant sequences to proto-Wuhan. Ka/Ks values were analyzed by Nei/Gojobori method[10] using KaKs calculator[12].

Results

Putative ancestral sequences of SARS-CoV-2 variants

Each SARS-CoV-2 variant, including Omicron, has accumulated diversity via: mutations, deletions, and insertions. It is expected that each variant sequence will be slightly different from its ancestral strain. To characterize the ancestor sequence of each variant, I downloaded ten of the earliest collected sequences of each variant from the GISAID database[3]. The ancestral prototypes of SARS-CoV-2 variants were identified by the shared and conserved sequences among those sequences of each variant and named proto "variant". The molecular phylogenetic tree of the ancestral sequences shows that they are distinct descendants of the ancestral Wuhan SARS-CoV-2, proto-Wuhan (Figure 1). The mutational spectrums of the

variants were identified by comparing the sequences of proto variants with proto-Wuhan (Figure s1).

Nonsynonymous substitution bias of SARS-CoV-2 variants

SARS-CoV-1 is the coronavirus that caused SARS[13], and the bat coronavirus RaTG13 is thought to be one of the proximal origins of the SARS-CoV-2[14]. In the evolution among SARS-CoV-1, RaTG13, and proto-Wuhan, synonymous substitutions were more frequent than nonsynonymous substitutions in most genes (Figure 2A). While the spike gene is 3.8 kb, ORF1ab is a large ORF, extending over 21.3 kb and accounting for 71.2% of the 30 kb of the SARS-CoV-2 genome. The size of other genes, M, ORF6, ORF7a, ORF7b, ORF8, and N, are small. Given the size of each gene, it makes sense that ORF1ab is more frequently mutated than other genes.

The accumulation of mutations in each variant from proto-Wuhan is limited (Figure 2B) compared to SARS-CoV1, RaTG13, and proto-Wuhan, reflecting their evolutionary distances. Proto-Omicron, harboring the most mutations, was the most divergent from proto-Wuhan (Figure 2B). In addition to one insertion, and five deletions (Table s2), proto-Omicron has 52 mutations, 30 of which are located in spike gene (Figure 2B). Interestingly, 29 of the 30 mutations in spike gene are nonsynonymous substitutions (Figure 2B). The synonymous substitutions of proto-Omicron are concentrated in ORF1ab, where four of the eleven mutations are synonymous. When ORF1ab is excluded from Omicron's total mutations, there are 36 nonsynonymous substitutions out of a total of 41 (87.8%).

I observed a similar trend in the other mutants (Figure 2B), with most of the proto variant's mutations concentrated in the spike gene and ORF1ab. The genes other than spike gene, ORF1ab, and N had very few mutations. Interestingly, only one synonymous substitution in the Omicron and Lambda variants was located on the spike gene. Most of each variant's synonymous mutations were located on ORF1ab. Therefore, I compared each variant's total mutational load, minus the mutations on ORF1ab. The resulting N (nonsynonymous)/S (synonymous) ratios were extremely high across variants. In particular, proto-Delta and proto-MuGH had no synonymous mutations outside of ORF1ab.

Lack of neutral evolution in the spike gene of SARS-CoV-2 variants

The number of synonymous and nonsynonymous loci varies with each codon. For example, a single nucleotide substitution at Leu codon TTA can generate nine different codons; seven nonsynonymous, and two synonymous substitutions. Thus, the three nucleotides of codon TTA consist of 7/3 nucleotides (=3 nucleotides x 7/9) of nonsynonymous sites and 2/3

nucleotides (=3 nucleotides x 2/9) of synonymous sites[10]. Thus, the mutation frequency can be normalized as the ratio of the number of nonsynonymous substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks).

The evolution among SARS-CoV-1, RaTG13, and proto-Wuhan shows that Ks is significantly higher than Ka in most genes (Figure 3A), indicating neutral evolution. These low Ka/Ks ratios are consistent with expected viral evolution, including SARS[15], which follows neutral evolution[16]. Mutational signatures from proto-Wuhan to the descendant proto variants are shown in Figure 3B. The Ka and Ks of: M, ORF6, ORF7a, ORF7b, ORF8, and N are over-represented reflecting their small sizes (Figure 3B). Depending on variant, the Ka was not always higher than Ks for: M, ORF6, ORF7a, ORF7b, ORF8, and N. The Ka of ORF1ab was consistently lower than the Ks in all the variants, except for proto-Delta. Conversely, the Ka of the spike gene is notably higher than its Ks across all of the proto variants. The Ka/Ks ratios of the spike genes of proto variants are in striking contrast with those observed during evolution among SARS-CoV-1, RaTG13, and proto-Wuhan. Thus, it appears that spike gene of these SARS-CoV-2 variants deviates from a neutral evolutionary trajectory.

Discussion

I identified the ancestral type sequence of the SARS-CoV-2 variants and analyzed their mutational differences from the Wuhan SARS-CoV-2. Next, I analyzed the ratio of synonymous and nonsynonymous substitutions across the variants in detail. The analysis of variants associated with Ka/Ks ratio was used to interpret the direction and magnitude of natural selection acting on the protein-coding genes. Ka/Ks ratio greater than one indicates Darwinian selection, while Ka/Ks ratio less than 1 indicates purifying selection[10]. Ka/Ks ratio of exactly one is indicative of no selection. When Ka/Ks values are significantly above one, the corresponding mutations are generally advantageous. The Ka of the spike gene was much higher than the Ks across the variants I examined. This would imply that each of these nonsynonymous substitutions in spike gene was the result of strong selective pressure.

Recently, several groups independently pointed out the nonsynonymous substitution bias in the Omicron variant [17-19]. Especially, Xi et al. compared the spike protein and ORF1ab of several SARS-CoV-2 variants and noticed that the lack of synonymous substitutions in the spike protein is unnatural[19], supporting my observation.

The Pfizer and Moderna mRNA based vaccines encode for a full-length spike gene, whose amino acid sequence is identical to that of the Wuhan SARS-CoV-2 spike gene, with the

exception of two proline substitutions. [20] In order to increase translation the spike genes of these mRNA vaccines have been heavily modified with synonymous substitutions[21]. This suggests that the spike genes of proto variants could potentially use synonymous substitutions to improve translation efficiency. Thus, codon usage cannot explain the strong selective pressure to eliminate synonymous substitutions in spike genes of proto variants.

Following the emergence of each proto variant, additional mutations accumulate during transmission among humans. The Ka/Ks ratios in the evolution of each variant did not exceed three[19]. Considering that the spike gene achieved high ratios of synonymous substitutions during evolution of SARS-CoV-1, RaTG13 and proto-Wuhan (Figure 3A), the lack of synonymous substitutions in spike gene of the proto variants (Figure 3B) is not a common feature of evolution of corona viruses. In fact, synonymous substitutions are also achieved during artificial evolution for gain-of-function using somatic hypermutations in a cell line[22].

The spike protein of coronaviruses mediates the transmission of the virus; it functions as a fusogen, which mediates membrane fusion after binding to the ACE2 receptor. Selective pressure on spike gene depends on human ACE2-mediated cell transmission. Generally speaking, the spike protein will not accept random amino acid changes in an effort to maintain the fusogen function.

Mutation and selection usually occur stepwise; a mutation is fixed after its functional selection, followed by the next mutation and selection. Nonsynonymous substitutions rarely improve protein function, and they are can be detrimental. Purifying selection removes nonsynonymous substitutions that damage enzymatic activity or protein structure, thus reducing the number of nonsynonymous substitutions in critical genes. On the other hand, synonymous substitutions rarely elicit phenotypic change, therefore they are not generally not subject to purifying selection, unless they drastically reduce translation efficiency. Synonymous substitutions tend to accumulate during evolution. Since the rate of synonymous substitution is usually similar among different genes, synonymous substitutions can be used as a molecular clock for dating the evolutionary time of closely related species[16]. Surprisingly, synonymous substitutions are significantly less prevalent than nonsynonymous substitutions among SARS-CoV-2 proto variants. Proto variants also exhibit a lack of synonymous substitutions in their spike gene, suggesting that the evolution of spike gene was occurred outside of a classic trial-and-error - mutation and selection scheme. Furthermore, the features of the molecular clock of synonymous substitutions suggest that the mutations in spike gene were acquired over a short period.

The analyzed SARS-CoV-2 variants have evolved novel spike proteins over a short period, while maintaining high transmission rates. The spike genes of these variants do not appear to have undergone neutral evolution, in contrast to the rest of the SARS-CoV-2 genome (Figure 4). This could be explained if one postulates that spike genes with specific mutations have been "artificially inserted" into the viral genome. As the mutations of proto-Delta and proto-MuGH have an overall biased for nonsynonymous substitutions (Figures 2 and 3), artificial genes may be present outside of the spike gene, depending on variants. Technically, site-directed mutagenesis can easily be used to introduce specific mutations using seamless cloning or genome editing. In fact, the evolution of SARS-CoV-2 remains a mystery, as is the controversy over how SARS-CoV-2 acquired the furin cleavage site[23].

The presence of many nonsynonymous substitutions and the lack of synonymous substitutions in the spike genes of the proto variants suggests that these genes specifically are not the byproduct of typical natural or artificial evolution. While it cannot be ruled out that the spike genes of these proto variants emerged stochastically, though the probability is exceedingly low, it is difficult to explain how proto variants have coincidentally achieved such high transmission rates among humans absent evidence for functional selection. I therefore postulate that SARS-CoV-2 variants are artificially designed viruses whose spike genes are made by site-directed mutagenesis.

Acknowledgments

I am grateful to B. Hershey, I. Psakhye, Y. Matsumoto, and H. Kakeya for critically reading the manuscript.

Figure 1. Phylogenetic tree. The ancestral type of each SARS-CoV-2 variant was named proto "variant". The evolutionary distance is shown by the scale.

Figure 2. **Synonymous (S) and nonsynonymous (N) substitutions. A. Synonymous and nonsynonymous substitutions on each gene among SARS-CoV-1, RaTG13, and proto Wuhan. B. Those from proto-Wuhan to the respective proto variants.** The sum of the synonymous and nonsynonymous substitutions with or without substitutions in ORF1ab, is also shown.

Figure 3. **Ka and Ks A. Ka and Ks on each gene among SARS-CoV-1, RaTG13, and proto-Wuhan. B. Ka and Ks from proto-Wuhan to the respective proto variants.** Ka is nonsynonymous substitutions per nonsynonymous sites, and Ks is synonymous substitutions per synonymous sites. When there is no selective pressure and mutations are introduced randomly, the ratio of Ka to Ks is 1.

Figure 4. Lack of neutral evolution in the spike gene of SARS-CoV-2 variants. The genome structure of the SARS-CoV-2 variant is shown.

Figure s1. Multiple alignments of SARS-CoV-2 proto variants. Green, synonymous substitutions; red, nonsynonymous substitutions; pink, ambiguous; blue, noncoding mutations; gray, stop codon-causing mutations; orange, insertions; purple, deletions.

Figure s2. Sequences of SARS-CoV-2 proto variants in the FASTA format.

[1] E. Callaway, Heavily mutated Omicron variant puts scientists on alert, Nature, 600 (2021) 21.
[2] K. Kupferschmidt, Where did 'weird' Omicron come from?, Science, 374 (2021) 1179.
[3] Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality, Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 22 (2017).
[4] J. Hadfield, C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R.A. Neher, Nextstrain: real-time tracking of pathogen evolution, Bioinformatics (Oxford, England), 34 (2018) 4121-4123.
[5] M. Kandeel, M.E.M. Mohamed, H.M. Abd El-Lateef, K.N. Venugopala, H.S. El-Beltagi, Omicron variant genome evolution and phylogenetics, Journal of medical virology, (2021).
[6] B. Choi, M.C. Choudhary, J. Regan, J.A. Sparks, R.F. Padera, X. Qiu, I.H. Solomon, H.H. Kuo, J. Boucau, K. Bowman, U.D. Adhikari, M.L. Winkler, A.A. Mueller, T.Y. Hsu, M. Desjardins, L.R. Baden, B.T. Chan, B.D. Walker, M. Lichterfeld, M. Brigl, D.S. Kwon, S. Kanjilal, E.T. Richardson, A.H. Jonsson, G. Alter, A.K. Barczak, W.P. Hanage, X.G. Yu, G.D. Gaiha, M.S. Seaman, M. Cernadas, J.Z. Li, Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host, The New England journal of medicine, 383 (2020) 2291-2293.
[7] C. Wei, K.J. Shan, W. Wang, S. Zhang, Q. Huan, W. Qian, Evidence for a mouse origin of the SARS-CoV-2 Omicron variant, Journal of genetics and genomics = Yi chuan xue bao, (2021).
[8] E. Wilkinson, M. Giovanetti, H. Tegally, J.E. San, R. Lessells, D. Cuadros, D.P. Martin, D.A. Rasmussen, A.N. Zekri, A.K. Sangare, A.S. Ouedraogo, A.K. Sesay, A. Priscilla, A.S. Kemi, A.M. Olubusuyi, A.O.O. Oluwapelumi, A. Hammami, A.A. Amuri, A. Sayed, A.E.O. Ouma, A. Elargoubi, N.A. Ajayi, A.F. Victoria, A. Kazeem, A. George, A.J. Trotter, A.A. Yahaya, A.K. Keita, A. Diallo, A. Kone, A. Souissi, A. Chtourou, A.V. Gutierrez, A.J. Page, A. Vinze, A. Iranzadeh, A. Lambisia, A. Ismail, A. Rosemary, A. Sylverken, A. Femi, A. Ibrahimi, B. Marycelin, B.S. Oderinde, B. Bolajoko, B. Dhaala, B.L. Herring, B.M. Njanpop-Lafourcade, B. Kleinhans, B. McInnis, B. Tegomoh, C. Brook, C.B. Pratt, C. Scheepers, C.G. Akoua-Koffi, C.N. Agoti, C. Peyrefitte, C. Daubenberger, C.M. Morang'a, D.J. Nokes, D.G. Amoako, D.L. Bugembe, D. Park, D. Baker, D. Doolabh, D. Ssemwanga, D. Tshiabuila, D.

Bassirou, D.S.Y. Amuzu, D. Goedhals, D.O. Omuoyo, D. Maruapula, E. Foster-Nyarko, E.K. Lusamaki, E. Simulundu, E.M. Ong'era, E.N. Ngabana, E. Shumba, E. El Fahime, E. Lokilo, E. Mukantwari, E. Philomena, E. Belarbi, E. Simon-Loriere, E.A. Anoh, F. Leendertz, F. Ajili, F.O. Enoch, F. Wasfi, F. Abdelmoula, F.S. Mosha, F.T. Takawira, F. Derrar, F. Bouzid, F. Onikepe, F. Adeola, F.M. Muyembe, F. Tanser, F.A. Dratibi, G.K. Mbunsu, G. Thilliez, G.L. Kay, G. Githinji, G. van Zyl, G.A. Awandare, G. Schubert, G.P. Maphalala, H.C. Ranaivoson, H. Lemriss, H. Anise, H. Abe, H.H. Karray, H. Nansumba, H.A. Elgahzaly, H. Gumbo, I. Smeti, I.B. Ayed, I. Odia, I.B. Ben Boubaker, I. Gaaloul, I. Gazy, I. Mudau, I. Ssewanyana, I. Konstantinus, J.B. Lekana-Douk, J.C. Makangara, J.M. Tamfum, J.M. Heraud, J.G. Shaffer, J. Giandhari, J. Li, J. Yasuda, J.Q. Mends, J. Kiconco, J.M. Morobe, J.O. Gyapong, J.C. Okolie, J.T. Kayiwa, J.A. Edwards, J. Gyamfi, J. Farah, J. Nakaseegu, J.M. Ngoi, J. Namulondo, J.C. Andeko, J.J. Lutwama, J. O'Grady, K. Siddle, K.T. Adeyemi, K.A. Tumedi, K.M. Said, K. Hae-Young, K.O. Duedu, L. Belyamani, L. Fki-Berrajah, L. Singh, L.O. Martins, L. Tyers, M. Ramuth, M. Mastouri, M. Aouni, M. El Hefnawi, M.I. Matsheka, M. Kebabonye, M. Diop, M. Turki, M. Paye, M.M. Nyaga, M. Mareka, M.M. Damaris, M.W. Mburu, M. Mpina, M. Nwando, M. Owusu, M.R. Wiley, M.T. Youtchou, M.O. Ayekaba, M. Abouelhoda, M.G. Seadawy, M.K. Khalifa, M. Sekhele, M. Ouadghiri, M.M. Diagne, M. Mwenda, M. Allam, M.V.T. Phan, N. Abid, N. Touil, N. Rujeni, N. Kharrat, N. Ismael, N. Dia, N. Mabunda, N.Y. Hsiao, N.B. Silochi, N. Nsenga, N. Gumede, N. Mulder, N. Ndodo, N.H. Razanajatovo, N. Iguosadolo, O. Judith, O.C. Kingsley, O. Sylvanus, O. Peter, O. Femi, O. Idowu, O. Testimony, O.E. Chukwuma, O.E. Ogah, C.K. Onwuamah, O. Cyril, O. Faye, O. Tomori, P. Ondoa, P. Combe, P. Semanda, P.E. Oluniyi, P. Arnaldo, P.K. Quashie, P. Dussart, P.A. Bester, P.K. Mbala, R. Ayivor-Djanie, R. Njouom, R.O. Phillips, R. Gorman, R.A. Kingsley, R.A.A. Carr, S. El Kabbaj, S. Gargouri, S. Masmoudi, S. Sankhe, S.B. Lawal, S. Kassim, S. Trabelsi, S. Metha, S. Kammoun, S. Lemriss, S.H.A. Agwa, S. Calvignac-Spencer, S.F. Schaffner, S. Doumbia, S.M. Mandanda, S. Aryeetey, S.S. Ahmed, S. Elhamoumi, S. Andriamandimby, S. Tope, S. Lekana-Douki, S. Prosolek, S. Ouangraoua, S.A. Mundeke, S. Rudder, S. Panji, S. Pillay, S. Engelbrecht, S. Nabadda, S. Behillil, S.L. Budiaki, S. van der Werf, T. Mashe, T. Aanniz, T. Mohale, T. Le-Viet, T. Schindler, U.J. Anyaneji, U. Chinedu, U. Ramphal, U. Jessica, U. George, V. Fonseca, V. Enouf, V. Gorova, W.H. Roshdy, W.K. Ampofo, W. Preiser, W.T. Choga, Y. Bediako, Y. Naidoo, Y. Butera, Z.R. de Laurent, A.A. Sall, A. Rebai, A. von Gottberg, B. Kouriba, C. Williamson, D.J. Bridges, I. Chikwe, J.N. Bhiman, M. Mine, M. Cotten, S. Moyo, S. Gaseitsiwe, N. Saasa, P.C. Sabeti, P. Kaleebu, Y.K. Tebeje, S.K. Tessema, C. Happi, J. Nkengasong, T. de Oliveira, A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa, Science, 374 (2021) 423-431.

[9] M. Kimura, Evolutionary rate at the molecular level, Nature, 217 (1968) 624-626.

[10] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, Molecular biology and evolution, 3 (1986) 418-426.

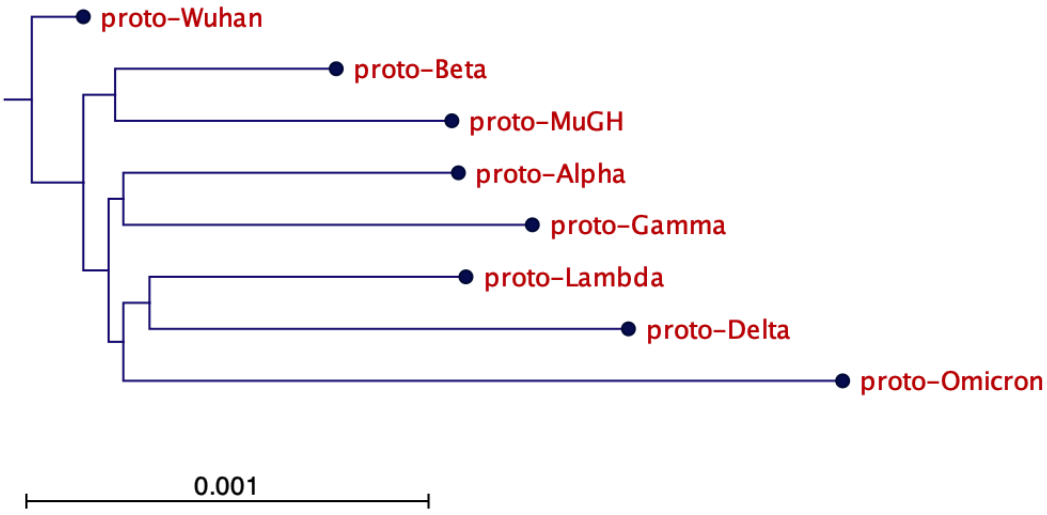
[11] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Molecular biology and evolution, 4 (1987) 406-425.

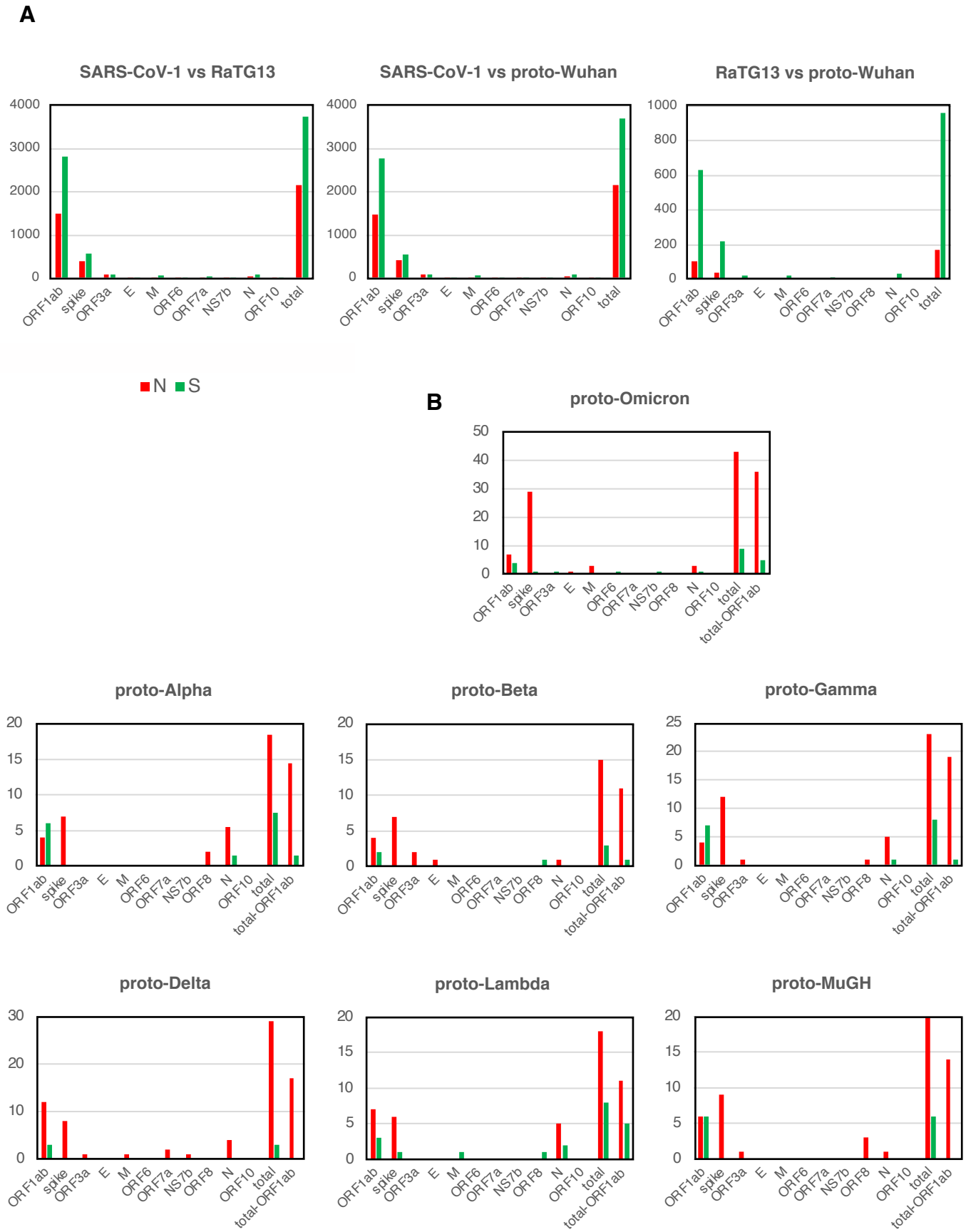
[12] Z. Zhang, J. Li, X.Q. Zhao, J. Wang, G.K. Wong, J. Yu, KaKs_Calculator: calculating Ka and Ks through model selection and model averaging, Genomics, proteomics & bioinformatics, 4 (2006) 259-263.

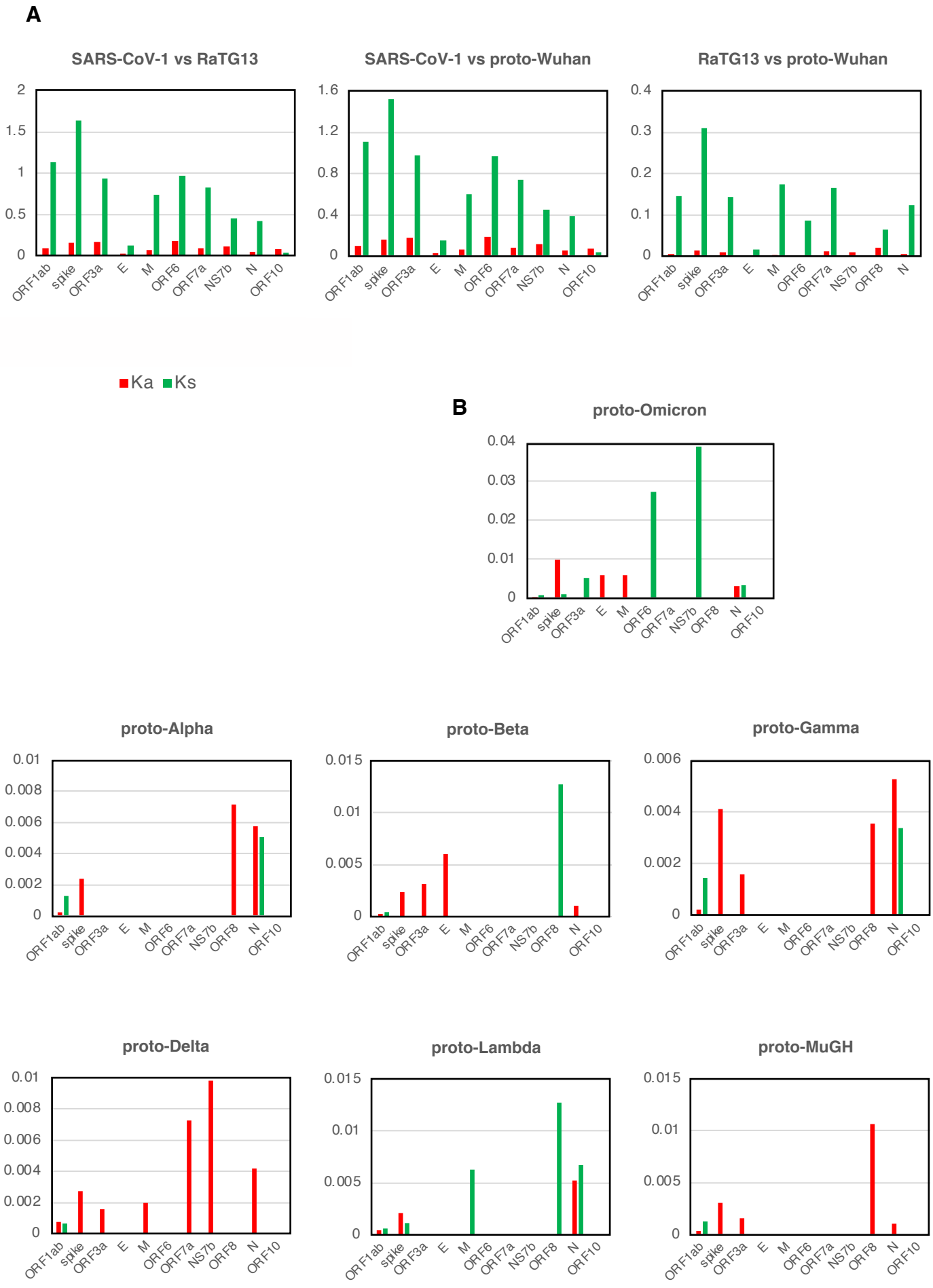
[13] M.A. Marra, S.J. Jones, C.R. Astell, R.A. Holt, A. Brooks-Wilson, Y.S. Butterfield, J. Khattra, J.K. Asano, S.A. Barber, S.Y. Chan, A. Cloutier, S.M. Coughlin, D. Freeman, N. Girn, O.L. Griffith, S.R. Leach, M. Mayo, H. McDonald, S.B. Montgomery, P.K. Pandoh, A.S. Petrescu, A.G. Robertson, J.E. Schein, A. Siddiqui, D.E. Smailus, J.M. Stott, G.S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T.F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G.A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R.C. Brunham, M. Krajden, M. Petric, D.M. Skowronski, C. Upton, R.L. Roper, The Genome sequence of the SARS-associated coronavirus, Science, 300 (2003) 1399-1404.

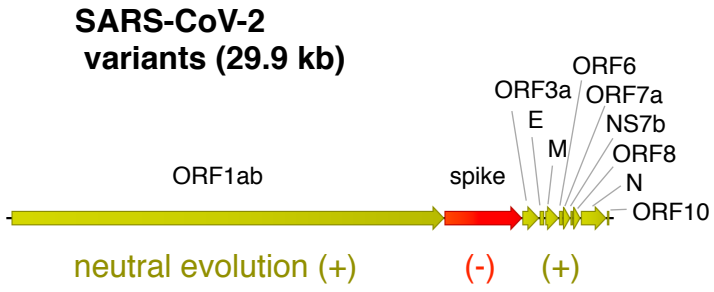
[14] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R.D. Jiang, M.Q. Liu, Y. Chen, X.R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F.X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature, 579 (2020) 270-273.
[15] L.D. Hu, G.Y. Zheng, H.S. Jiang, Y. Xia, Y. Zhang, X.Y. Kong, Mutation analysis of 20 SARS virus genome sequences: evidence for negative selection in replicase ORF1b and spike gene, Acta pharmacologica Sinica, 24 (2003) 741-745.
[16] T. Gojobori, E.N. Moriyama, M. Kimura, Molecular clock of viral evolution, and the neutral theory, Proceedings of the National Academy of Sciences of the United States of America, 87 (1990) 10015-10018.
[17] R. Viana, S. Moyo, D.G. Amoako, H. Tegally, C. Scheepers, C.L. Althaus, U.J. Anyaneji, P.A. Bester, M.F. Boni, M. Chand, W.T. Choga, R. Colquhoun, M. Davids, K. Deforche, D. Doolabh, L. du Plessis, S. Engelbrecht, J. Everatt, J. Giandhari, M. Giovanetti, D. Hardie, V. Hill, N.Y. Hsiao, A. Iranzadeh, A. Ismail, C. Joseph, R. Joseph, L. Koopile, S.L. Kosakovsky Pond, M.U.G. Kraemer, L. Kuate-Lere, O. Laguda-Akingba, O. Lesetedi-Mafoko, R.J. Lessells, S. Lockman, A.G. Lucaci, A. Maharaj, B. Mahlangu, T. Maponga, K. Mahlakwane, Z. Makatini, G. Marais, D. Maruapula, K. Masupu, M. Matshaba, S. Mayaphi, N. Mbhele, M.B. Mbulawa, A. Mendes, K. Mlisana, A. Mnguni, T. Mohale, M. Moir, K. Moruisi, M. Mosepele, G. Motsatsi, M.S. Motswaledi, T. Mphoyakgosi, N. Msomi, P.N. Mwangi, Y. Naidoo, N. Ntuli, M. Nyaga, L. Olubayo, S. Pillay, B. Radibe, Y. Ramphal, U. Ramphal, J.E. San, L. Scott, R. Shapiro, L. Singh, P. Smith-Lawrence, W. Stevens, A. Strydom, K. Subramoney, N. Tebeila, D. Tshiabuila, J. Tsui, S. van Wyk, S. Weaver, C.K. Wibmer, E. Wilkinson, N. Wolter, A.E. Zarebski, B. Zuze, D. Goedhals, W. Preiser, F. Treurnicht, M. Venter, C. Williamson, O.G. Pybus, J. Bhiman, A. Glass, D.P. Martin, A. Rambaut, S. Gaseitsiwe, A. von Gottberg, T. de Oliveira, Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa, Nature, 603 (2022) 679-686.
[18] D.P. Martin, S. Lytras, A.G. Lucaci, W. Maier, B. Grüning, S.D. Shank, S. Weaver, O.A. MacLean, R.J. Orton, P. Lemey, M.F. Boni, H. Tegally, G.W. Harkins, C. Scheepers, J.N. Bhiman, J. Everatt, D.G. Amoako, J.E. San, J. Giandhari, A. Sigal, C. Williamson, N.Y. Hsiao, A. von Gottberg, A. De Klerk, R.W. Shafer, D.L. Robertson, R.J. Wilkinson, B.T. Sewell, R. Lessells, A. Nekrutenko, A.J. Greaney, T.N. Starr, J.D. Bloom, B. Murrell, E. Wilkinson, R.K. Gupta, T. de Oliveira, S.L. Kosakovsky Pond, Selection Analysis Identifies Clusters of Unusual Mutational Changes in Omicron Lineage BA.1 That Likely Impact Spike Function, Molecular biology and evolution, 39 (2022).
[19] B. Xi, Y. Meng, D. Jiang, Y. Bai, Z. Chen, Y. Qu, S. Li, J. Wei, L. Huang, H. Du, Analyses of Long-Term Epidemic Trends and Evolution Characteristics of Haplotype Subtypes Reveal the Dynamic Selection on SARS-CoV-2, viruses, 14 (2022) 454.
[20] L.A. Jackson, E.J. Anderson, N.G. Rouphael, P.C. Roberts, M. Makhene, R.N. Coler, M.P. McCullough, J.D. Chappell, M.R. Denison, L.J. Stevens, A.J. Pruijssers, A. McDermott, B. Flach, N.A. Doria-Rose, K.S. Corbett, K.M. Morabito, S. O'Dell, S.D. Schmidt, P.A. Swanson, 2nd, M. Padilla, J.R. Mascola, K.M. Neuzil, H. Bennett, W. Sun, E. Peters, M. Makowski, J. Albert, K. Cross, W. Buchanan, R. Pikaart-Tautges, J.E. Ledgerwood, B.S. Graham, J.H. Beigel, An mRNA Vaccine against SARS-CoV-2 - Preliminary Report, The New England journal of medicine, 383 (2020) 1920-1931.
[21] D.E. Jeong, M.J. McCoy, K.L. Artiles, O. Ilbay, A.Z. Fire, K.C. Nadeau, H.R. Park, B.E. Betts, S.D. Boyd, R.A. Hoh, M.J. Shoura, Assemblies of putative SARS-CoV2-spike-encoding mRNA sequences for vaccines BNT-162b2 and mRNA-1273, Available online: https://virological.org/t/assemblies-of-putative-sars-cov2-spike-encoding-mrna-sequences-for-vaccines-bnt-162b2-andmrna-1273/663, (2021).
[22] H. Arakawa, H. Kudo, V. Batrak, R.B. Caldwell, M.A. Rieger, J.W. Ellwart, J.M. Buerstedde, Protein evolution by hypermutation and selection in the B cell line DT40, Nucleic Acids Res, 36 (2008) e1.

[23] B.K. Ambati, A. Varshney, K. Lundstrom, G. Palú, B.D. Uhal, V.N. Uversky, A.M. Brufsky, MSH3 Homology and Potential Recombination Link to SARS-CoV-2 Furin Cleavage Site, Frontiers Virol, 2 (2022) 834808.


proto-Wuhan
proto-Beta
proto-MuGH
proto-Alpha
proto-Gamma
proto-Lambda
proto-Delta
proto-Omicron
0.001


A
SARS-CoV-1 vs RaTG13
SARS-CoV-1 vs proto-Wuhan
RaTG13 vs proto-Wuhan
N S
B
proto-Omicron
proto-Alpha
proto-Beta
proto-Gamma
proto-Delta
proto-Lambda
proto-MuGH
ORF1ab
spike
ORF3a
E
M
ORF6
ORF7a
NS7b
ORF8
N
ORF10
total
total-ORF1ab


A
SARS-CoV-1 vs RaTG13
SARS-CoV-1 vs proto-Wuhan
RaTG13 vs proto-Wuhan
Ka
Ks
B
proto-Omicron
proto-Alpha
proto-Beta
proto-Gamma
proto-Delta
proto-Lambda
proto-MuGH
ORF1ab
spike
ORF3a
E
M
ORF6
ORF7a
NS7b
ORF8
N
ORF10


SARS-CoV-2
variants (29.9 kb)
ORF6
ORF3a
ORF7a
E
NS7b
M
ORF8
ORF1ab
spike
N
ORF10
neutral evolution (+)
(-)
(+)