International Journal of Multidisciplinary Research and Development
www.allsubjectjournal.com
Online ISSN: 2349-4182, Print ISSN: 2349-5979
Received: 30-07-2021, Accepted: 15-08-2021, Published: 31-08-2021
Volume 8, Issue 8, 2021, Page No. 167-172

Insilico insights to mutational and evolutionary aspects of sars-COV2

Sudheer Menon
Deptartment of Bioinformatics, Bharathiar University, Tamil Nadu, India

Abstract
The in silico technology has various applications in biological sciences. As an example, the method was used in a drug repurposing study to look for potential COVID-19 treatments. The current coronavirus outbreak has been linked to severe acute respiratory syndrome 2 (SARS-CoV-2), and its phylogeny and taxonomy have been established. Several full-length sequences of viral genome have been made accessible since the onset of infections, and they have been used to acquire insights on viral dynamics. Gene-based sequence analysis, haplotype success and potential adaptability, numerous mutations at a site, and so on are examples of SARS-CoV-2 molecular evolution. The use of artificial intelligence in the mutational and evolutionary elements of SARS-Cov2 has aided scientists in comprehending the virus's nature and traits. Female immune systems were found to be more active than male immune systems in response to SARS-CoV-2 infections.

Keywords: sars-COV2, COVID-19 treatments, Insilico insights

Introduction

What exactly is in silico? An in silico experiment is one that is carried out on a computer or through computer simulation in biology and other experimental sciences. The term is pseudo-Latin for "in silicon" (in Latin, "in silicio"), and it refers to silicon in computer chips. It was coined in 1987 as a play on the Latin terms *in vivo*, *in vitro*, and in situ, which are frequently used in biology (especially systems biology). The latter terms refer to experiments conducted in living organisms, outside of living species, and where they can be found in nature, in that order. Originally, the word in silico referred only to computer simulations that modeled natural or laboratory processes (in all natural sciences), and did not relate to computer calculations in general. Christopher Langton used the phrase to characterize artificial life in an announcement of a symposium on the issue at the Center for Nonlinear Studies at Los Alamos National Laboratory in 1987. (Hameroff, 2014) [6]. the in silico technology has various applications in biological sciences. For example, the method was used in a drug repurposing study to look for potential COVID-19 (SARS-CoV-2) treatments (Lee *et al*., 2020) [8].

The World Health Organization (WHO) reported cases of pneumonia of unknown source in Wuhan City, Hubei Province of China, in early January 2020, and by 30 January 2020, WHO had upgraded the alert to a public health emergency of international significance. By March 12, 2020, the novel coronavirus (nCoV) outbreak had reached pandemic proportions and was designated as novel Covid-19 sickness (nCovid-19) (EDT, 2020). The current coronavirus outbreak is linked to severe acute respiratory syndrome 2 (SARS-CoV-2), which has its own phylogeny and classification (CSGICTV, 2020).

Coronaviruses are single-stranded, positive-stranded RNA viruses of the genus Coronavirus, family Coronaviridae, that can cause acute and chronic respiratory and central nervous system disorders in animals, including humans (To *et al*., 2013; Pillaiyar *et al*., 2016) [16, 14]. In certain people, the infection can also produce minor episodes of follicular conjunctivitis. In animal models, the infection has been found to cause symptoms similar to anterior uveitis, retinitis, and optic neuritis (Seah and Agrawal, 2020) [15]. Recent research has revealed the production of hyper-reflective lesions in the retina's ganglion cell and inner plexiform layers, notably near the papillomacular bundles (Marinho *et al*., 2020) [10]. Patients' sense of smell and taste bud sensitivity have also been demonstrated to be affected by the condition (Giacomelli *et al*., 2020) [5].

SARS-CoV-2 has recently emerged as a pandemic, killing about 2.4 million people worldwide. Several full-length sequences of viral genome have been made accessible since the onset of infections, and they have been used to acquire insights on viral dynamics (Periwal *et al*., 2021) [12]. Coronaviruses are enclosed viruses that contain single-strand RNA as their genetic material. SARS coronavirus triggered the first coronavirus pandemic in 2002, killing 919 people globally (Yang *et al*., 2020) [19]. MERS coronavirus produced another coronavirus epidemic in 2012, resulting in 2499 cases and 858 fatalities (Memish *et al*., 2020) [11]. Notably, MERS coronavirus was more virulent, with a fatality ratio of roughly 34.3 percent, compared to SARS coronavirus, which had a mortality ratio of around 9.6 percent. We are already experiencing another pandemic triggered by a new SARS coronavirus that originated in Wuhan, China (Zhou *et al*., 2020) [20].

As SARS-CoV-2 infections proceed, the virus may accumulate multiple additional mutations, such as the observed Delta variant. Virus evolution is heavily influenced by nucleotide substitution, and various organizations are examining viral genomic sequences to uncover changes driving SARS-CoV-2 evolution and pathogenesis. Several nucleotide changes and deletions were discovered in SARS-CoV-2 genomic sequences from throughout the world in a previous study (Phan, 2020) [13]. Phylogenetic analysis of SARS-CoV-2 data revealed three clusters of the virus circulating around the world. Clusters A and C were found to be more frequent in Europe and America, while Cluster B was found only in East Asia

(Forster *et al*., 2020) [3].

SARS-CoV-2 infections are expanding rapidly, and the virus is accumulating new mutations that may aid in the virus's infectivity. As a result, we used meta-analysis techniques to detect extremely important alterations in viral genomic sequences from throughout the world. We used a suite of algorithms, including normal mode analysis, discrete molecular dynamics, and all-atom molecular dynamic simulations, to predict the possible effect of these mutations on parent proteins in order to better understand the possible role of these mutations in viral structural proteins. Our detailed examination of many SARS-CoV-2 structural protein mutations revealed that L37H mutation in E protein, G204R and P344S in N protein, and D614G in S protein were destabilizing the parent protein, whereas P13L, S197L, and R203K in N protein were stabilizing the parent protein (Periwal *et al*., 2021) [12].

The molecular development of SARS-CoV-2

Sequence analysis based on genes

The ratio of non-synonymous to synonymous (dN/dS) substitutions among a set of sequences is a simple and widely used metric for detecting selection. According to this paradigm, a ratio of one suggests neutral evolution, a ratio less than one shows purifying selection, and a ratio greater than one indicates positive selection (Kryazhimskiy and Plotkin, 2008) [7]. In a recent study, the 385 haplotypes discovered from the 15,789 full-length sequences of SARS-CoV-2 downloaded from the EpiCov data portal at gisaid.org were used to compute dN/dS for the entire genome and then for each gene. Given the dN/dS model assumptions (i.e., synonymous mutations are neutral), the findings imply that multiple proteins may be under positive selection, with nsp2 showing the strongest signal (Figure 1). (Garvin *et al*., 2020) [4].

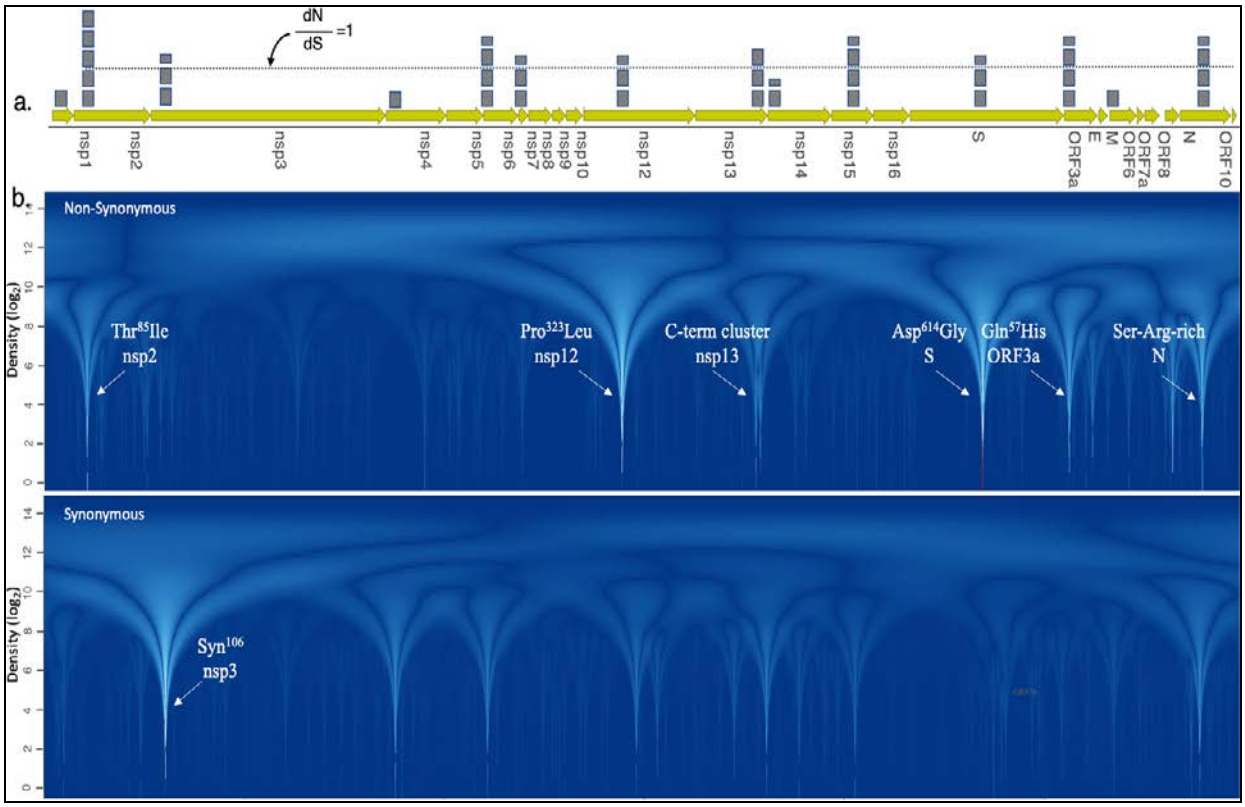

Fig 1: Several proteins may be under positive selection with the highest signal in nsp2 (Garvin *et al*., 2020) [4].

a) Non-synonymous to synonymous mutation ratio (dN/dS) per gene (barplots). We estimated ratios from the 385 identified haplotypes using full-length sequences containing 395 variable coding sites from GISAID. Genes with less than 10 mutations across the population were removed, as were haplotypes with fewer than five people. Given the rarity of synonymous and non-synonymous mutations, ten genes (E, nsp7, nsp8, nsp10, nsp16, ORF6, ORF7A, ORF7B, and ORF8) are likely under high purifying selection at the nucleotide level. dN/dS was calculated using all modifications in a gene. Barplots are centered on the gene with the greatest signal. b) Wavelet analysis of SARS-CoV-2 genome non-synonymous (top) and synonymous (bottom) mutations. The arrows point to mutation sites mentioned in the text. The density of the wavelet across the genome is represented as a log-scale on the y-axis. Higher values suggest a wider wavelet and, as a result, coarser granularity.

The success of haplotypes and their capacity for adaptability

Because of the rapid reporting of full-length genomes and their haploid (and mostly non-recombining) character, it is possible to establish a mutational genealogy of the virus. The timeline of mutations as the virus spreads around the world can be represented as a median-joining network. It is thus feasible to identify mutations that happened before and after the appearance (or removal) of a haplotype from the sampled population. Furthermore, knowing when a virus was sampled provides a temporal approximation of a specific haplotype's half-life. A definition of the term "success" in the context of haplotypes Researchers calculated SC (i.e., viral fitness) as the ratio of the number of individuals (N) with a specific variant to the number of days that variant was sampled (T) and then by the number of geographic regions (G) out of six (Figure 2) in which it is present [SC = (N/T) 1/G]. Under this scenario, the most

successful haplotypes are those that persist for an extended period of time before mutating and infect a large number of people across many geographic regions. Scientists are particularly interested in viral types that are unsuccessful according to preset models but enhance fitness with a subsequent mutational event, as these may represent adaptive or compensatory responses. These ineffective haplotypes will be those that remain over several days and are distributed in a wide range of geographical areas but with a low frequency (Garvin *et al*., 2020) [4].

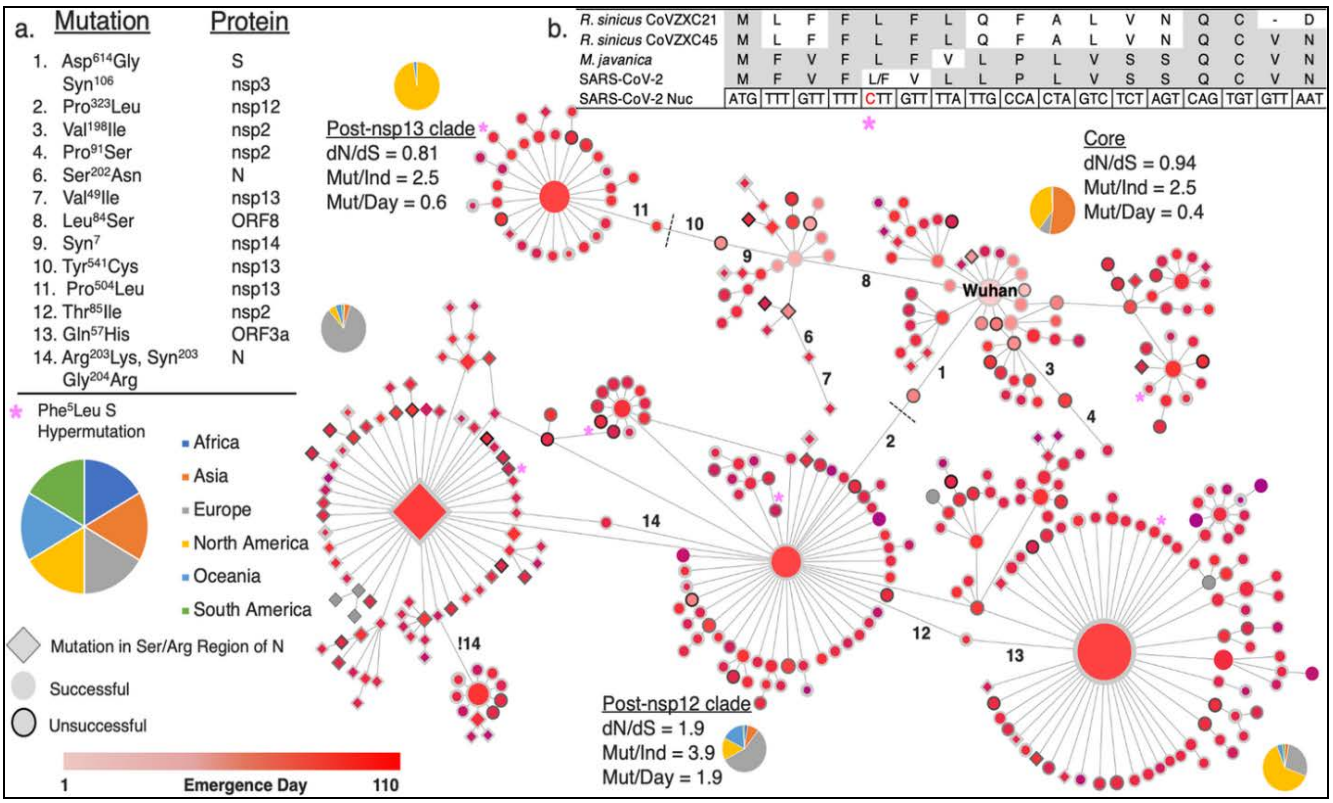

Fig 2: A definition of haplotype "success" SC (i.e., viral fitness) as the ratio of the number of individuals (N) having a particular variant to the number of days that variant was sampled (T), followed by the number of geographic regions (G) out of six (Garvin *et al*., 2020) [4].

These researchers discovered five edges on the haplotype network that may indicate harmful mutations, followed by compensatory modifications that boosted the modeled fitness (Figure 2).

Multiple mutations at a single location

Garvin *et al*. (2020) [4] classified 395 sites as variable (found in ten or more individuals), 64 of which undergo more than one mutation, and 62 of which are in coding areas. The most common alternative allele is a synonymous alteration at 20 of the 62 locations (32%). Almost half (48%) of the second alternate alleles are either synonymous with the reference allele or synonymous with the first alternate allele. The majority of multiple mutations at a site are uncommon, however others are common and locally unique. Beginning in late March, an AGC codon (coding for a serine) at location 1197 in the nsp3 protein mutated at the third position of AGA (arginine) in 27 people from Southeast Asia and one person from Australia. Previously, this same codon altered to AGT, resulting in a synonymous mutation found mostly in Washington State (14 out of 18 individuals). Notably, the N protein contains 13 sites, half of which are in the serine/arginine-rich region, demonstrating the high mutation rate at this functionally critical position (Figure 1). (Garvin *et al*., 2020) [4].

SARS-Cov2 Mutational and Evolutionary Aspects of Artificial Intelligence

As the COVID-19 epidemic proceeds, scientists have discovered additional SARS-CoV-2 variants with potentially deadly characteristics. Variant B.1.1.7 lineage clade GR from the Global Initiative for Sharing All Influenza Data (GISAID) was discovered in the United Kingdom, and it appears to be more transmissible. Simultaneously, South African authorities found variation B.1.351, which contains multiple changes with B.1.1.7 and may be highly transmissible. In Brazil, a variation known as P.1 with 17 non-synonymous mutations was discovered. The World Health Organization recently expressed alarm about the variety B.1.617.2, which was initially discovered in India but is now being exported worldwide. The importance of rapid development of precise molecular assays for uniquely identifying novel variations cannot be overstated. Some workers' in-silico studies revealed that the sequences in the tested primer sets have good precision and are based on two or more mutations. The researchers also published an analysis of critical mutations for SARS-CoV-2 variants (B.1.1.7, B.1.351, P.1, B.1.617.2 and B.1.1.519). Other scientists can use their given methods to quickly create primer sets for each new variant, which can then be used as part of a multiplexed approach for the first diagnosis of COVID-19 patients (Figure 3). (Lopez-Rincon *et al*., 2021) [9].

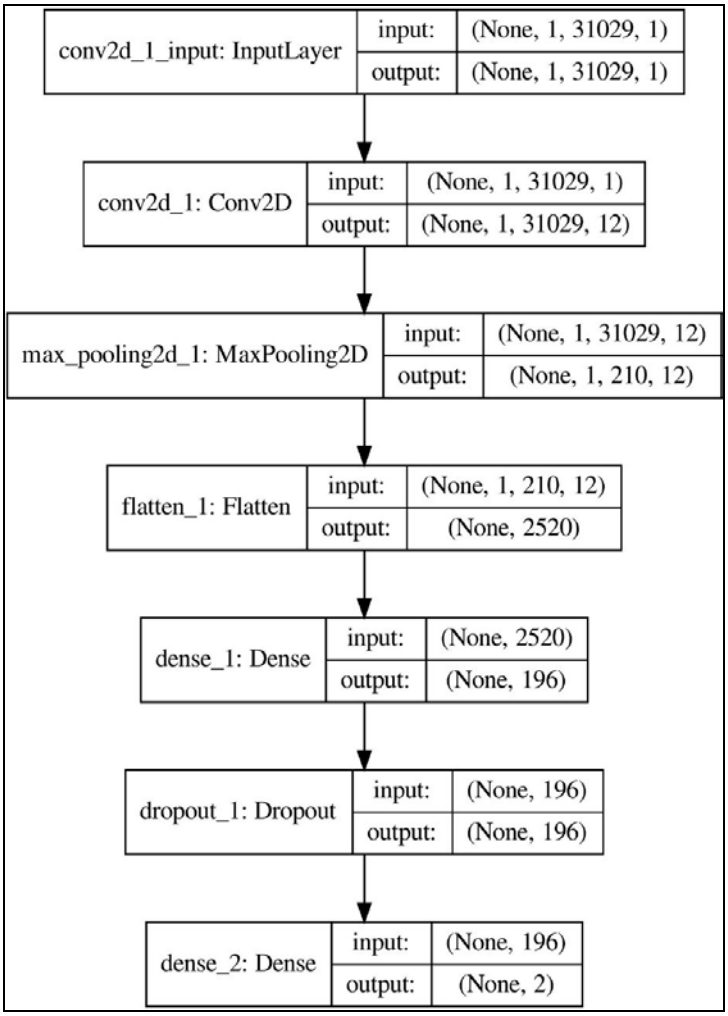

Fig 3: CNN architecture used to classify variants B.1.1.7, B.1.351, and P.1. Used in a study by Lopez-Rincon *et al.* (2021) [9].

Characterizing SARS-CoV-2 mutations in the United States

Since it was originally sequenced in early January 2020, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been changing. The genetic variants have evolved into a few unique clusters, each with its own set of characteristics. Because the United States (US) has the most virally infected patients in the world, it is critical to understand the US SARS-CoV-2. Some researchers discovered that the US SARS-CoV-2 has four substrains and five top US SARS-CoV-2 mutations were first detected in China (2 cases), Singapore (2 cases), and the United Kingdom using genotyping, sequence alignment, time evolution, k-means clustering, protein-folding stability, algebraic topology, and network theory (1 case). The following three top US SARS-CoV-2 variants were discovered in the United States. These top eight mutations are divided into two distinct groups. The first group, which consists of five concurrent mutations, is dominant, whereas the second group, which consists of three concurrent mutations, eventually fades out. According to the findings of these researchers, female immune systems are more active than male immune systems in reacting to SARS-CoV-2 infections. They discovered that one of the major mutations on ORF8, 27964C>T-(S24L), had an abnormally strong gender dependence. Based on the study of all spike protein alterations, scientists discovered that three of the four US SASR-CoV-2 substrains became more infectious. Their research advocates for better virus control and containment techniques in the United States (Wang *et al.*, 2020) [17, 19, 20].

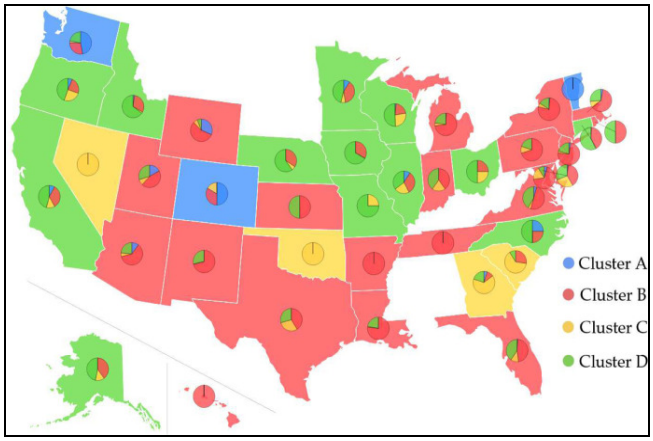

Fig 4: depicts a pie chart representation of four clusters in the United States on July 14, 2020. Clusters A, B, C, and D are represented by the blue, red, yellow, and green colors, respectively. The dominant cluster determines the basic color of each state. Some states do not provide GISAID with whole genomic sequences. As a result, the basic color of these states was not established. (Wang *et al.*, 2020) [17, 19, 20].

Since it was first sequenced in early January 2020, SARS-CoV-2 has been mutating. Wang *et al.* (2021) [18] investigated the mutations in 45,494 full SARS-CoV-2 genome sequences from around the world. There are 12,754

sequences from the United States among them. According to their findings, there are four substrains and eleven top mutations in the United States. These top eleven mutations are divided into three distinct groupings. The first and second groups, which have 5 and 8 concurrent mutations, are dominant, whereas the third group, which has three concurrent mutations, is rapidly fading out. Wang *et al.* (2021) [18] also discovered that female immune systems are more active in reacting to SARS-CoV-2 infections than male immune systems. One of the most common mutations in ORF8, 27964C > T-(S24L), exhibits an extremely strong gender dependence (Figure 5). They discover that two of four SARS-CoV-2 substrains in the United States become possibly more contagious after analyzing all changes on the spike protein.

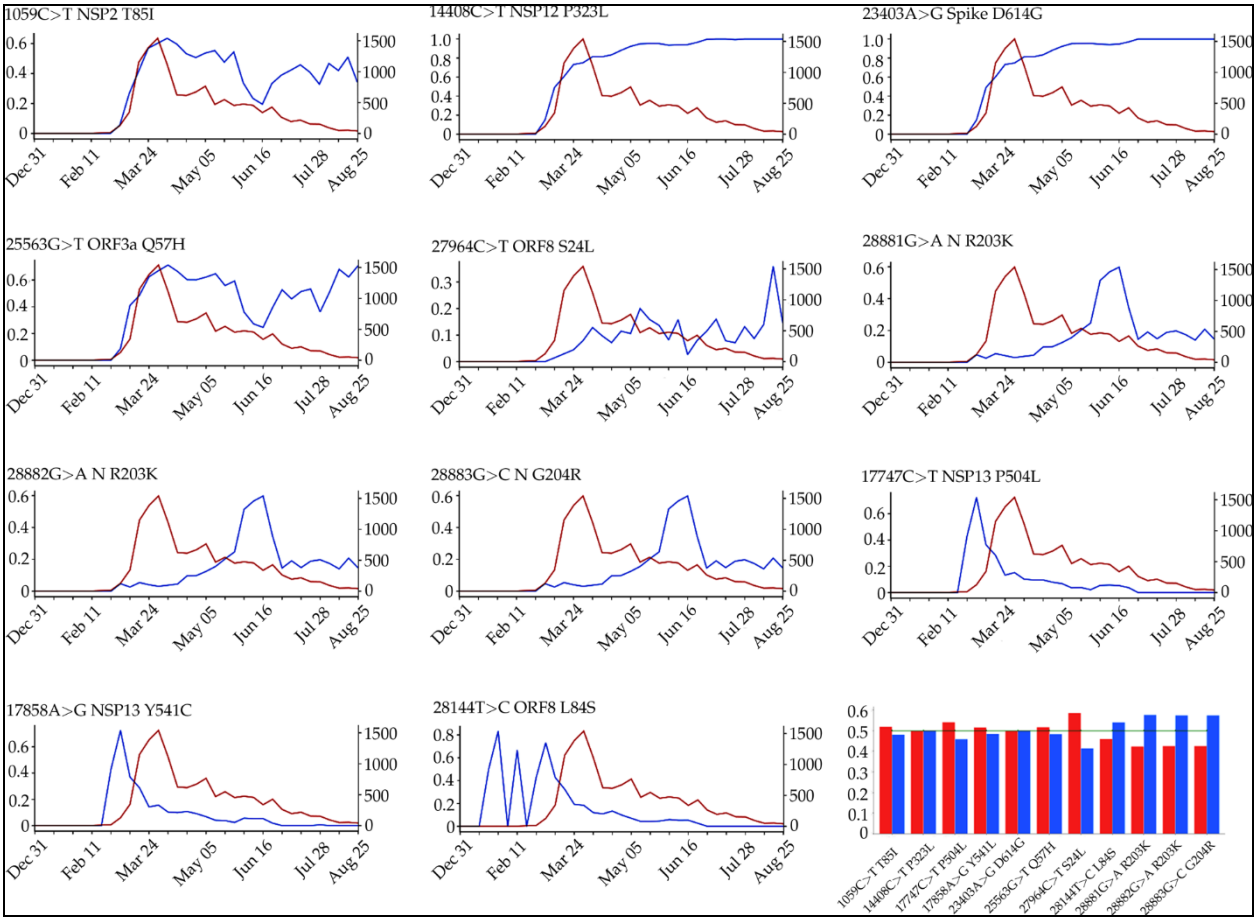

Fig 5: The evolution and the gender distribution of the top 11 missense mutation ratios.

The blue lines depict the evolution of the top 11 missense mutation ratios (the y-axis on the left), which are calculated as the number of genome sequences with a specific mutation divided by the total number of genome sequences. The evolution of the total number of genomic sequences is represented by the red lines (the y-axis on the right). The gender distribution of the ratio of the number of samples with the top 11 missense mutations over the total number of samples with age and/or gender labels is represented by the bar plot. In the United States, red bars reflect female ratios and blue bars show male ratios.

References

1. Coronaviridae Study Group of the International Committee on Taxonomy of V. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol,2020:5:536-44.
2. Eurosurveillance Editorial T. Note from the editors: World Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern. Euro Surveill,2020:25:200131e.
3. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proceedings of the National Academy of Sciences of the United States of America, 2020:117(17):9241-9243. 10.1073/pnas.2004999117
4. Garvin MR, Prates ET, Pavicic M, Jones P, Amos BK, Geiger A *et al.* Potentially adaptive SARS-CoV-2 mutations discovered with novel spatiotemporal and explainable AI models. Genome Biol,2020:21:304.
5. Giacomelli A, Pezzati L, Conti F, Bernacchia D, Siano M, Oreni L *et al.* Self-reported olfactory and taste disorders in SARS-CoV-2 patients: a cross-sectional study. Clin Infect Dis, 2020.
6. Hameroff SR. Ultimate Computing Biomolecular Consciousness and Nanotechnology. Elsevier Publishers, 2014.
7. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. Plos Genet,2008:4(12):e1000304.
8. Lee VS, Chong WL, Sukumaran SD, Nimmanpipug P, Letchumanan V, Goh BH *et al.* Computational screening and identifying binding interaction of anti-viral and anti-malarial drugs: toward the potential cure for SARS-CoV-2. Progress in Drug Discovery & Biomedical Science,2008:3:1-9.

9. Lopez-Rincon A, Perez-Romero CA, Tonda A, Mendoza-Maldonado L, Claassen E, Garssen J *et al*. Design of Specific Primer Sets for the Detection of B.1.1.7, B.1.351, P.1, B.1.617.2 and B.1.1.519 Variants of SARS-CoV-2 using Artificial Intelligence. Bio Rxiv 01.20.427043, 2021.
10. Marinho PM, Marcos AAA, Romano AC, Nascimento H, Belfort R Jr. Retinal findings in patients with COVID-19. Lancet,2020:395:1610.
11. Memish ZA, Perlman S, Van Kerkhove MD, Zumla A. Middle East respiratory syndrome. The Lancet, 2020:395(10229):1063-1077. 10.1016/S0140 6736(19)33221-0.
12. Periwal N, Rathod SB, Pal R, Sharma P, Nebhnani L, Barnwal RP *et al*. In silico characterization of mutations circulating in SARS-CoV-2 structural proteins. J Biomol. Struct Dyn, 2021, 1-16.
13. Phan T. Genetic diversity and evolution of SARS-CoV-2. Infection, Genetics and Evolution,2020:81:104260. 10.1016/j.meegid.2020.104260.
14. Pillaiyar T, Manickam M, Namasivayam V, Hayashi Y, Jung SH. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. J Med Chem,2016:59:6595-628.
15. Seah I, Agrawal R. Can the coronavirus disease 2019 (COVID-19) affect the eyes? A review of coronaviruses and ocular implications in humans and animals. Ocul Immunol Inflamm,2020:28:391-5.
16. To KK, Hung IF, Chan JF, Yuen KY. From SARS coronavirus to novel animal and human coronaviruses. J Thorac Dis,2013:5(Suppl 2):S103-8.
17. Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei G. Characterizing SARS-CoV-2 mutations in the United States. Research square, rs.3.rs-49671, 2020.
18. Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei G. Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. Commun Biol,2021:4:228.
19. Yang Y, Peng F, Wang R, Yange M, Guan K, Jiang T *et al*. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. Journal of Autoimmunity,2020:109:102434. 10.1016/j.jaut.2020.102434
20. 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(7798):270-273. 10.1038/s41586-020-2012-7.
21. Sudheer Menon. “Preparation and computational analysis of Bisulphite sequencing in Germfree Mice” International Journal for Science and Advance Research in Technology,2020:6(9):557-565.
22. Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. “Computational identification of promoter regions in prokaryotes and Eukaryotes” EPRA International Journal of Agriculture and Rural Economic Research (ARER),2021:9(7):21-28.
23. Sudheer Menon. “Bioinformatics approaches to understand gene looping in human genome” EPRA International Journal of Research & Development (IJRD),2021:6(7):170-173.
24. Sudheer Menon. “Insilico analysis of terpenoids in Saccharomyces Cerevisiae”international Journal of Engineering Applied Sciences and Technology, ISSN No. 2455-2143,2021:6(1):43-52.
25. Sudheer Menon. “Computational analysis of Histone modification and TFBs that mediates gene looping” Bioinformatics, Pharmaceutical, and Chemical Sciences (RJLBPCS),2021:7(3):53-70.
26. Sudheer Menon Shanmughavel piramanayakam, Gopal Prasad Agarwal. “FPMD-Fungal promoter motif database: A database for the Promoter motifs regions in fungal genomes” EPRA International Journal of Multidisciplinary research,2021:7(7):620-623.
27. Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. Computational Identification of promoter regions in fungal genomes, International Journal of Advance Research, Ideas and Innovations in Technology,2021:7(4):908-914.
28. Sudheer Menon, Vincent Chi Hang Lui, Paul Kwong Hang Tam. Bioinformatics methods for identifying hirschsprung disease genes, International Journal for Research in Applied Science & Engineering Technology (IJRASET),2021:9(7):2974-2978.