

The First Geographic Identification by Country of Sustainable Mutations of SARS-COV2 Sequence Samples: Worldwide Natural Selection Trends

Author information

Mohammadamin Mahmanzar¹⁺, Seyed Taleb Houseini²⁺, Karim Rahimian³⁺, Arsham Mikaeili Namini⁴, Amir Gholamzad⁵, Samaneh Tokhanbigli⁶, Mahsa Mollapour Sisakht⁷, Amin Farhadi⁸, Donna Lee Kuehu⁹, Youping Deng^{10*}

1. Department of Quantitative Health Sciences, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI 96813, USA; Mahmanza@hawaii.edu
2. Department of Biology, Faculty of Basic Sciences, Qaemshahr Branch, Islamic Azad University, Mazandaran, Iran; seyedtalebhouseini@gmail.com, seyedtalebhouseini@qaemiau.ac.ir
3. Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics. Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran. karim.rahimian@modares.ac.ir
4. Department of Animal Biology, Faculty of Biological Sciences, Kharazmi University, Tehran, Iran; arshammikaeili@gmail.com
5. Department of Laboratory Medicine, Faculty of Paramedical Sciences, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran; Amirgholamzad@gmail.com
6. Department of Molecular and Cellular Sciences, Faculty of Advanced Sciences and Technology, pharmaceutical Sciences Branch, Islamic Azad University, Tehran, Iran (IAUPS); Tokhanbigli.s@gmail.com
7. Department of Biochemistry, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands; m.molapoursisakht@erasmusmc.nl
8. Department of Biology, Payame Noor University, Tehran, Iran
9. Department of Quantitative Health Sciences, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI 96813, USA; dkuehu@hawaii.edu
10. Department of Quantitative Health Sciences, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI 96813, USA; dengy@hawaii.edu

+ These authors have the same contribution

***Corresponding Author:**

***Youping Deng**, PhD. Department of Quantitative Health Sciences, John A. Burns School of Medicine, University of Hawai'i at Mānoa, Tel: 808.692.1664(Office); Fax: 808.692.1970, Email: dengy@hawaii.edu

ABSTRACT The high mutation rates of RNA viruses, coupled with short generation times and large population sizes, allow viruses to evolve rapidly and adapt to the host environment. The rapidity of viral mutation also causes problems in developing successful vaccines and antiviral drugs. With the spread of SARS-CoV-2 worldwide, thousands of mutations have been identified, some of which have relatively high incidences, but their potential impacts on virus characteristics remain unknown. The present study analyzed mutation patterns, SARS-CoV-2 AASs retrieved from the GISAID database containing 10,500,000 samples. Python 3.8.0 programming language was utilized to pre-process FASTA data, align to the reference sequence, and analyze the sequences. Upon completion, all mutations discovered were categorized based on geographical regions and dates. The most stable mutations were found in nsp1(8% S135R), nsp12(99.3% P323L), nsp16 (1.2% R216C), envelope (30.6% T9I), spike (97.6% D614G), and Orf8 (3.5% S24L), and were identified in the United States on April 3, 2020, and England, Gibraltar, and, New Zealand, on January 1, 2020, respectively. The study of mutations is the key to improving understanding of the function of the SARS-CoV-2, and recent information on mutations helps provide strategic planning for the prevention and treatment of this disease. Viral mutation studies could improve the development of vaccines, antiviral drugs, and diagnostic assays designed with high accuracy, specifically useful during pandemics. This knowledge helps to be one step ahead of new emergence variants.

IMPORTANCE

More than two years into the global COVID-19 pandemic, the focus of attention is shifted to the emergence and spread of the SARS-CoV-2 variants that cause the evolutionary trend.

Here, we analyzed and compared about 10.5 million sequences of SARS-CoV-2 to extract the stable mutations, frequencies and the substitute amino acid that changed with the wild-type one in the evolutionary trend.

Also, developing and designing accurate vaccines could prepare long-term immunization against different local variants. In addition, according to the false negative results of the COVID-19 PCR test report in the diagnosis of new strains, investigating local mutation patterns could help to design local primer and vaccine.

KEYWORDS: SARS-CoV2, Natural selection, Mutation, Epidemiology, Amino Acid, region,

INTRODUCTION

Genetic variations, followed by changes in phenotype level in the Sars-Cov2 virus, lead to different virus behaviors. Sars-Cov2 is a single-stranded positive-sense RNA that is approximately 30kb in length, including nonstructural (NSP) and structural proteins. The structural viral substance, including the spike (S), envelope (E), and membrane (M), as well as the nucleocapsid (N) protein that bundles the viral genome, are translated from subgenomic RNAs(1). According to research, the SARS-CoV-2 lineage, the causal agent of the pandemic COVID-19, generates one to two mutations every month as it passes through consecutive hosts, increasing infection, pathogenicity, or immune evasion(2, 3). There are six main readout frames in the virus genome open reading frame (ORF). The ORF1ab and ORF1a genes are located at the 5' termini of the SARS-CoV-2 genome. ORF1ab contains 16 NSPs. The 3' termini of the genome are composed of four structural proteins: S glycoprotein, E protein, M glycoprotein, and N phosphoprotein(4). The angiotensin-converting enzyme 2 (ACE2) is the primary entry receptor for S protein in SARS-CoV-2, facilitating the virus entry(5, 6). Several S mutations occurred in the appearance of B.1.1.7, B.1.351, P.1, and B.1.617.2 variants. In most cases, these mutations do not affect pathogenicity and could only affect the dissemination(5-7). N501Y, E484K, and D614G are examples of common mutations between variants(5-7). SARS-CoV-2 has over 10,000 distinct mutations compared to the reference genome(8). From the 12,509 viral genomes of SARS-CoV-2 sequences analyzed so far in the pandemic COVID-19 outbreak, the CoV-GLUE project (<http://cov-glue.cvr.gla.ac.uk/#/home>) has identified 5,033 amino acid replacements, with 1,687 mutations found in the S glycoprotein, 334 in the N phosphoprotein, 95 in the M glycoprotein, and 45 in the E protein(4). The D614G (S glycoprotein) mutation was thought to have originated in January 2020 in Eastern China and spread throughout the world, with the most common amino acid (AA) changes occurring in 6,855 sequences(9). So far, stable mutants have been identified for Alpha (B.1.1.7 and Q lineages), Beta (B.1.351 and descendent lineages), Gamma (P.1 and descendent lineages), Epsilon (B.1.427 and B.1.429), Eta (B.1.525), Iota (B.1.526), Kappa (B.1.617.1, 1.617.3), Mu (B.1.621, B.1.621.1), and Zeta (B.1.1.529 and BA lineages)(<https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>)(10, 11). Geographically distinct SARS-CoV-2 etiological effects can be attributed to the genetic heterogeneity of SARS-CoV-2 specimens worldwide(12). The presence of so many different S protein mutations suggests numerous SARS-CoV-2 subtypes, each with its infectivity characteristics(8). In some cases, the Delta variant reduces the efficacy of therapies more than the other variants(13). Furthermore, viral mutation studies can help develop new vaccines, antiviral drugs, and diagnostic assays.

This study aimed to investigate and bioinformatically analyze stable mutations between January 2020 to May 2022 from different parts of the virus and in countries reported for the first time. By examining and analyzing with bioinformatic approaches, it has been possible to identify mutations in the coronavirus that remained stable over time and further identify each

AA conversion. This study reports the top 10 mutations of each protein and the earliest reporting from each country where samples with the stable mutations occurred.

RESULT

According to analysis of 10,500,000 samples, sustainability mutations have remained constant from January 1, 2020, to April 28, 2022, in France, Spain, Greece, Netherlands, England, New Zealand, Senegal, Russia and the USA are shown in Figure 1.

NSP1: The highest mutation is the conversion of Serine to Arginine, the most common mutation in North America, with 8% reported in 2020. Subsequent mutations in the conversion of Glutamic acid to Aspartic acid and Histidine to Tyrosine are 0.8% and 0.2% in England and the United States, respectively. Finally, the tenth mutation that has remained stable over time is the conversion of Proline to Serine at 0.08%, reported on January 1, 2020, in England.

NSP2: The conversion of Threonine to Isoleucine at 3.9% is the highest mutation in the Netherlands. The conversion of Valine to Isoleucine at 0.3% is the lowest mutation in North America in January 2020 for the nsp2 protein.

NSP3: Mutations in the nsp3 protein include conversion of Proline to Leucine at a rate of 50% (highest mutation), and subsequent mutations including conversion of Proline to Serine, Alanine to Serine, Threonine to Isoleucine (position 183), Alanine to Aspartic acid, and Isoleucine to Threonine (position 1412) from France and England reported on January 1, 2020, and the conversion of Proline to Leucine (tenth mutation) is 5.1% in England reported on January 1, 2020.

NSP4: The highest mutations in Europe were observed in England and the Netherlands, which include the mutations of Threonine to Isoleucine at 70.4% and the conversion of Valine to Leucine at 37.9%, respectively. The lowest mutation was detected in Northern Ireland with the replacement of Alanine to Valine at 0.5% for nsp4 protein reported on April 21, 2020.

NSP5, NSP6: Conversion of Leucine to Phenylalanine (position 89) was observed in nsp5 protein with 1.4% reported on January 1, 2020, in North America; however, in New Zealand, a 2.9% AA change (position 37, mutant 5) was observed in the same month for nsp6 protein. In the nsp5 protein, conversion of Alanine to Valine (position 260, mutant 6) was observed at 0.2% in North America, but this mutation occurred (position 2, mutant 7) in South America at 1.1% for nsp6 protein in January 2020.

NSP7: In the nsp7 protein, conversion of Leucine to Phenylalanine was observed in 2 different countries and positions, including England (position 71) at 0.2%, reported on January 1, 2020, and India (position 56) reported at 0.1%, reported on March 1, 2020 (Table1, Figure 2A).

NSP8, NSP9, NSP10: Conversion of Threonine to Isoleucine in North America for the nsp8 protein at 0.6% (position 145), for the nsp9 protein at 0.5% (position 35), and for the nsp10 protein at 0.1%, 0.1% and 0.08% (position 102, 12 and 51), respectively.

NSP11: In the nsp11 protein, conversion of Phenylalanine to Leucine was observed in 2 different positions, 11 and 7, in Italy, reported on April 9, 2020, and in Belgium, reported on January 29, 2021.

NSP12: The nsp12 protein mutations include the conversion of Proline to Leucine in 2 different positions, 323 and 227 at 99.3% and 1.8%, respectively, in England, and conversion of Arginine to Glutamine in position 197 at 0.5% and Glutamine to Arginine in position 875 at 0.5% in North America, all reported on April 6 and April 8, 2020.

NSP13, NSP14, NSP15, NSP16: The conversion of Proline to Leucine in position 77 (the first mutation) at 42.6% for the nsp13 protein, and position 46 (the fourth mutation), at 0.7% for the nsp14 protein, and in position 205 at 0.2% for the nsp15 protein reported from the Netherlands on January 1, 2020, and in position 215 (the fourth mutation) at 0.2% for the nsp16 protein from Norway, reported on March 9, 2020 (Table1, Figure2B).

ORF3a: In the Orf3a protein, the highest detected mutation is the conversion of Serine to Leucine at the rate of 43.1% reported in France, and the lowest mutation is the conversion of Tryptophan to Cysteine at the rate of 0.7% reported from Mexico, with both mutations occurring in early 2020. In North America, mutations were detected in the conversion of Threonine to Isoleucine at 10.2%, Glutamine to Histidine at 5%, and Glutamic acid to Glutamine at 2.7%, as reported in the USA. In South America, the conversion of Serine to Proline, at the rate of 1.2% in Brazil, was reported on March 9, 2020.

ORF6: Four mutations were detected in England in early 2020, the conversion of Aspartic acid to Leucine, Proline to Leucine at 0.1%, and Valine to Phenylalanine at 9.4%, 0.1% for two first and 0.03% for the last. Also, another There was mutation of the conversion of Arginine to Serine at 0.05% was reported in the Netherland.

ORF7a: The highest mutation of Orf7a was from the conversion of Threonine to Isoleucine at 41.4%, and Valine to Alanine at 41.1% found in the Netherlands and France, respectively, in early 2020. There was another mutation in England, the conversion of Threonine to Isoleucine at the rate of 0.2% found in England, the conversion of Leucine to Phenylalanine at the rate of 0.9% in France, and the conversion of Proline to Serine at the rate of 0.1% found in Wuhan. In Asia, there were two mutations: the conversion from Proline to Leucine at a rate of 1.5% in Saudi Arabia and Arginine to Glycine at a rate of 0.2% in India. The United States also had two mutations, the conversion of Valine to Isoleucine at the rate of 0.6% and Histidine to Tyrosine at 0.2%.

ORF7b: The highest rate of mutation of Orf7b is the conversion of Threonine to Isoleucine at a rate of 39.5% found in France. There were two more mutations in England, the conversion

of Serine to Leucine and Methionine to Isoleucine at the rate of 0.1% and 0.05%, respectively. Mutations in Europe also included the conversion of Histidine to Tyrosine at the rate of 0.1% found in France and Alanine to Serine at 0.08% found in Greece. Three mutations were found in Asia, the conversion of Alanine to Valine at the rate of 0.04% in Guangdong, Serine to Leucine at 0.09% in Japan, and Cysteine to Phenylalanine at the rate of 0.05% in Thailand.

ORF8: The highest mutation rate in Oceania was the conversion of Serine to Leucine at 3.5% found in New Zealand. In Europe, England had the most mutations with the conversion of Glutamic Acid to Lysine at the rate of 2.6%, Alanine to valine at the rate of 0.5%, and Valine to Leucine at the rate of 0.4%. In Italy, there was a mutation with the conversion of Serine to Phenylalanine at the rate of 0.5%, and in Gibraltar, a conversion of Alanine to Serine at the rate of 0.3% (Table2, Figure 3A).

ORF9b: The highest mutation rate in Orf9b is the conversion of Threonine to Alanine at the rate of 61.4% found in France, and additional mutations in Europe, were the conversion of Valine to Leucine at the rate of 0.6% in Russia, and Glycine to Aspartic acid at the rate of 0.4% in England. Two mutations were found in Canada, the conversion of Serine to Cysteine at the rate of 4.2% and Serine to Leucine at the rate of 0.8%. There was one mutation in Asia with the conversion of Proline to Leucine at the rate of 0.2%, found in Foshan.

ORF9c: Europe had the most mutations in the Orf9c protein. There were seven mutations in England, the conversions of Glycine to Asparagine at the rate of 48.8%, Valine to Leucine at the rate of 2%, Leucine to Phenylalanine at the rate of 1.8%, Glutamic Acid to Glutamine at the rate of 0.8%, Alanine to Valine at the rate of 0.1%, Serine to Leucine at the rate of 0.1% and Histidine to Tyrosine at the rate of 0.1%. (Table2, Figure3B).

Envelope, Membrane, Spike, Nucleoprotein: The conversion of Leucine to Phenylalanine was observed in 2 positions, 21 and 73, at 0.1% and 0.08% for the envelope protein, and the conversion of Serine to Phenylalanine was observed in mutant7 and mutant8 at 0.08% and 0.07% in England, reported on January 1, 2020. For the nucleoprotein, the conversion of Arginine to Methionine at position 203 at 62.8% was the highest mutation in France. The conversion of Alanine to Valine at position 220 at a rate of 2.5% was the lowest mutation found in England. Both were reported on January 1, 2020. For the membrane, the conversion of Isoleucine to Threonine at position 82, at a rate of 47.1%, was the highest mutation found in Senegal. The conversion of Serine to Asparagine in position 197 at a rate of 0.2% was the lowest mutation found in Spain. Both were reported on January 1, 2020. Finally, for the Spike protein, the conversion of Aspartic acid to Glycine at a rate of 97.6% was the highest mutation in Russia. The conversion of Aspartic acid to Tyrosine at a rate of 10% was the lowest mutation found in the Netherlands. Both were reported on January 1, 2020. (Table 3, Figure4).

DISCUSSION

During the early stages of the pandemic, the virus swiftly moved throughout the world and continued to spread globally. There was a more diversified mix of strains among the samples submitted to GISAID(14-16). Type G strains became more prominent in nations with higher resistance, resulting in a strain shift. Based on previous reports, the G strains may be a simpler target for vaccination because they contain larger quantities of surface proteins that target antibodies that bind to vaccines. Spike protein enables virus attachment to the human cell surface ACE2 receptor, allowing for viral entrance during infection(17-19). It is divided into two components called S1 and S2. The S1 unit has a receptor-binding domain (RBD) that may directly connect to the ACE2 receptor and the primary target of SARS-CoV-2 neutralizing antibodies (Ab). As a result, S1 is a hotspot for mutations with significant clinical significance in virulence, transmissibility, and host immune evasion(5, 20-22). The Alpha version features an N501Y mutation, which means that N asparagine has been replaced with Y tyrosine at the 501 residues, and K417N—lysine K has been replaced with asparagine N(23). A developing variety originating from B.1.1.7 also has the E484K mutation, which replaces glutamic acid E with lysine K. Other than N501Y, Beta and Gamma versions feature additional substitutions(23). The E484K mutation is present in the Beta variation, whereas the E484K and K417T mutations are present in the Gamma variant [GISAID-hCov19 Variants [Internet]. [cited 2021 June 6]. <https://www.gisaid.org/hcov19-variants/>]. The AA change D614G has been documented as a result of a nucleotide A to G mutation at position 23403 in the Wuhan reference strain. In early March 2020, this mutation was the first known alteration in the Spike analysis. During that period, the G614 form was uncommon but essential in Europe developments. The D614G mutation is frequently seen in conjunction with two additional mutations; C to T mutation at position 241 in the '5 UTR compared to the Wuhan reference sequence (silent C to T mutation at viral RdRp site) (P323L RdRp)(24). The V483A and G476S mutations were discovered in US samples initially, whereas the V367F mutant was found in China, Hong Kong, France and the Netherlands. The mutations V367F and D364Y have been shown to improve the stability of the spike protein structure and promote effective binding to the ACE2 receptor(11). A passenger returning to Taiwan from Wuhan reported an identical nucleotide-382 deletion in Orf8. However, the clinical consequence of this deletion has not been recorded. Orf8 is a good location for viral coronavirus alterations, according to research by Young E Barnaby et al. Deletion in this area appears to result in a milder illness with less systemic proinflammatory cytokine production and a more effective immune response to SARS-Cov2(25). Further research into this mutation might help to further therapy and vaccine development(25). NSP1 of Orf1a/Orf1ab and Orf8 have been identified as mutation hotspots linked to virulence and transmissibility. NSP1 is a critical protein that inhibits type I interferon activation in the host while promoting viral replication(23, 26). Yuan-Qin Min et al., 2021, concluded that in the nsp1 protein, the conversion of Aspartic Acid to Glutamic Acid occurred at position 75 in China, but in our study, the conversion of Glutamic Acid to Aspartic Acid occurred at position 87 in England. The present research is not consistent with our analyses(27). Orf8 is an immune-evasive protein that reduces the

expression of major histocompatibility complex class I (MHC-I) in host cells(25). Recently, it was discovered that the Alpha variation found from a single immunocompromised person had a premature stop codon at position 27 of Orf8(10). S-protein mutation D614G has influenced the SARS-CoV-2 transmissibility rate due to increased affinity for olfactory epithelium. It has been shown to have higher transmissibility in animal models(28, 29). The Alpha variation has been reported to increase hospitalizations and death, which may be attributable to their resistance to neutralizing antibodies owing to RBD mutations(30). Yuan and colleagues discovered that the CR3022 neutralizing antibody derived from a SARS patient binds to the RBD region of the spike in the SARS-Cov-2 virus by producing the crystal structure of this complex. This crystal structure enables researchers to comprehend how the SARS-CoV-2 virus is discovered. Identifying and obtaining these epitopes enables designing a vaccine against the SARS-CoV-2 virus and the detection of antibodies to additional coronaviruses in the future(31). Wang and colleagues have identified 13 mutation sites in the SARS-CoV-2 Orf1ab, Orf3a, Orf8 and N regions, with 28144 in Orf8 and 8782 Orf1a showing a mutation rate of 30.53% and 29.47%, respectively(32). The SARS-CoV-2 RdRp (also known as nsp12) is a crucial component of the replication/transcription machinery. Compared to SARS-CoV, SARS-CoV-2 has a high homology for nsp12, suggesting that its function and mode of action may be well preserved(33). This is supported by a recent cryo-EM structural investigation of SARS-CoV-2 nsp12(34). RdRps are thought to be one of the most important targets for antiviral medication development against a wide range of viruses. Favipiravir, Galidesivir, Remdesivir, and Ribavirin are several RdRp inhibitors evaluated for SARS-CoV-2 treatment(35). Interestingly, the docking location is not near the RdRp catalytic domain(36). Furthermore, additional prospective medications like filibuvir, cepharanthine, simeprevir, and tegobuvir are expected to be RdRp inhibitors(37). Natural mutations in key residues for pharmacological effectiveness can result in drug resistance, with a considerable reduction in the binding affinity of these compounds to the RdRp(38-40). We found 10 mutations for Orf7a protein which, according to a recent study, showed that ectodomain Orf7a binds to monocytes in the human peripheral blood, reducing its ability to deliver antigens, inducing dramatic expression of proinflammatory cytokines, and since the lungs are the site of SARS-CoV-2 proliferation causing the accumulation of monocytes in the lungs(41). Also, the Orf8 signaling pathway promotes proinflammatory by activation of IL-17, which contributes to cytokine storm in SARS-CoV-2 infection(42), includes apoptosis(43) and antagonizes the IFN signaling pathway(44). It has been shown that Orf9b could cause the inflammasome's activation, and significant antibody responses have been found to Orf9b. On the other hand, Orf9c expression impaired interferon signaling, antigen processing and presentation, complement signaling, and induced IL-6 signaling(45). Our study reports that the T60A mutation, which is correlated to Orf9b, is the more frequent mutation found in France, although Gopika et al. found this mutation in Thailand as well(46). Gupta S et al., concluded that in the Orf3a protein, Serine to Leucine conversion occurred at position 171 in India. However, we observed this AA conversion occurrence in France at a rate of 43.1%(47). Also, Michael D Sacco et al. concluded that mutations in Proline to Histidine at position 132, conversion of Leucine to Phenylalanine at position 89, and Lysine to Arginine at position 88 for the nsp5

protein in the omicron strain was consistent with our studies(48). In addition, Abdullah Shah et al. demonstrated that in the nsp2 protein, the AA change of Arginine to Cysteine occurred at position 197. The conversion of Valine to Isoleucine at position 378 occurred in Pakistan. However, in our study, the R27C mutation and the V485I mutation were detected in India and United States(49).

Conclusion

The development of efficient therapy and early identification of variants rely on a better understanding of virus behavior, and forecast mutation jumps to control the worldwide pandemic. A geographical view and a survey on mutations are the keys to developing a strategic plan to prevent burst events likely to increase transmissibility, morbidity and mortality from the COVID19 pandemic. Recent findings of SARS-CoV-2 mutations call for a massive effort in the scientific community to identify new variants that may increase viral spreading and allow escape from natural or vaccine-induced neutralizing immunity. This study of SARS-CoV-2 variations may improve therapies, vaccine design, and diagnostic methods.

MATERIALS AND METHODS

Source of the sequences: The research analyzed data from the genome of SARS-CoV-2 amino-acid sequences (AASs). The AAS of the Wuhan-2019 virus, accession number 'EPI_ISL_402124', was used as the reference sequence. This genetic sequence was applied to explore tightly correlated viruses against the whole database with BLASTn with a default parameter of 100. However, we obtained more than 100 results. This sample was used to compare against all 10,500,000 AASs samples obtained from the GISAID(14-16) database between 2020/1/1 - 2022/04/25. This database was provided by consent through Erasmus Medical Center's permission.

Analysis and processing of sequences: Python 3.8.0 was utilized to pre-process FASTA data and analyze sequence alignments and mutations. Each change between the sample and the reference marked a mutation, and the placement and substitution of the AAs were documented. Non-human samples, as well as samples containing unspecified AAs (reported as X), were excluded. The complete process was optimized using the 'Numpy' and 'Pandas' libraries.

Detection of mutations: The algorithm design was based on all sequences having the same length. The terms 'Refseq' and 'seq' refers to reference sequence and sample sequence. The algorithm specification was:

for refitem,

seqitem in zip (refseq, seq)

If (refitem! = seqitem)

Report of novel mutants: Following data extraction, the continent's name and geographic locations of each sample were retrieved and presented using the pycountry-convert 0.5.8 program using Python 'Titlecase' package to create worldwide maps of mutation prevalence.

Find candidate mutation: A high-frequency rate of at least 0.03% mutations was selected for plotting samples based on timelines.

Statistical analysis: R 4.0.3 and Microsoft Power BI used data standardization and comparison chart outlining. Each region's normalized frequency was presented to properly compare the data from each continent. For this purpose, the number of mutations on each continent was divided by the number of similar sequences in equal proportions.

List Of Abbreviations

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2),

Open reading frames (ORF),

Nonstructural protein 1 (NSP1),

Coronavirus disease 2019 (COVID-19),

Amino acid sequence (AAS),

Amino acid (AA),

Interferon (IFN).

Supplementary

- Supplementary file1 (first mutant).xlsx

Declaration of competing interest

The authors declare that they have no conflicts of interest that might be relevant to the contents of this manuscript, and the research was carried out regardless of commercial or financial relationships that may cause any conflict of interest.

Funding

This work was partially supported by the NIH grants and 2U54CA143727, 5P30GM114737, 5P20GM103466, 5U54MD007601, 5P30CA071789.

Acknowledgments

The authors thank all of the researchers who have shared genome data openly via the Global Initiative on Sharing All Influenza Data (GISAID).

Authors' contributions

Conceptualization, M.M. and K.R. and Y.D; methodology, M.M and K.R; software, M.M and K.R; validation, M.M., K.R., S.T and D.L.K.; formal analysis, M.M, and K.R; investigation, M.M, S.T.H, A.G, M.M.S and A.F; resources, K.R, S.T, and A.M.N; data curation, M.M, K.R, A.M.N, S.T, and D.L.K; writing—original draft preparation, S.T.H, and A.G.; writing—review and editing, S.T, D.L.K.; visualization, A.M.N.; supervision, Y.D.; project administration, Y.D, M.M; funding acquisition, Y.D.

Data availability

- The raw data supporting the conclusions of this article is available in supplementary file(s).

References

1. Pyrc K, Dijkman R, Deng L, Jebbink MF, Ross HA, Berkhout B, van der Hoek L. 2006. Mosaic structure of human coronavirus NL63, one thousand years of evolution. *J Mol Biol* 364:964-73.
2. Nagy A, Pongor S, Gyorffy B. 2021. Different mutations in SARS-CoV-2 associate with severe and mild outcome. *Int J Antimicrob Agents* 57:106272.
3. Hodcroft EB, Domman DB, Snyder DJ, Oguntuyo KY, Van Diest M, Densmore KH, Schwalm KC, Femling J, Carroll JL, Scott RS, Whyte MM, Edwards MW, Hull NC, Kevill CG, Vanchiere JA, Lee B, Dinwiddie DL, Cooper VS, Kamil JP. 2021. Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike protein variants affecting amino acid position 677. *medRxiv* doi:10.1101/2021.02.12.21251658.
4. Ortiz-Prado E, Simbana-Rivera K, Gomez-Barreno L, Rubio-Neira M, Guaman LP, Kyriakidis NC, Muslin C, Jaramillo AMG, Barba-Ostria C, Cevallos-Robalino D, Sanches-SanMiguel H, Unigarro L, Zalakeviciute R, Gadian N, Lopez-Cortes A. 2020. Clinical, molecular, and epidemiological characterization of the SARS-CoV-2 virus and the Coronavirus Disease 2019 (COVID-19), a comprehensive literature review. *Diagn Microbiol Infect Dis* 98:115094.
5. Huang Y, Yang C, Xu XF, Xu W, Liu SW. 2020. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin* 41:1141-1149.
6. Bai C, Warshel A. 2020. Critical Differences between the Binding Features of the Spike Proteins of SARS-CoV-2 and SARS-CoV. *J Phys Chem B* 124:5907-5912.
7. Hatmal MM, Alshaer W, Al-Hatamleh MAI, Hatmal M, Smadi O, Taha MO, Oweida AJ, Boer JC, Mohamud R, Plebanski M. 2020. Comprehensive Structural and Molecular Comparison of Spike Proteins of SARS-CoV-2, SARS-CoV and MERS-CoV, and Their Interactions with ACE2. *Cells* 9.
8. Chen J, Wang R, Wang M, Wei GW. 2020. Mutations Strengthened SARS-CoV-2 Infectivity. *J Mol Biol* 432:5212-5226.
9. Weimer L, Cattari G, Binelli A, Belasio EF, Piras S, Proddighe A, Sensi F. 2021. The First Year of Sars-Cov-2: Which Mutations Spread Rapidly Around the World? - Minireview. *Living with Sars-Cov-2 Variants An Update in the Pandemic Era*:2-5.
10. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, Connor T, Peacock T, Robertson D, Volz E. 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations [WWW Document]. *Virological* URL <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (accessed 14 21).
11. Ou J, Zhou Z, Dai R, Zhang J, Lan W, Zhao S, Wu J, Seto D, Cui L, Zhang G. 2020. Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein. *BioRxiv* 2020.
12. Mercatelli D, Giorgi FM. 2020. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front Microbiol* 11:1800.
13. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Perez Marc G, Moreira ED, Zerbini C, Bailey R, Swanson KA, Roychoudhury S, Koury K, Li P, Kalina WV, Cooper D, Frenck RW, Jr., Hammitt LL, Tureci O, Nell H, Schaefer A, Unal S, Tresnan DB, Mather S, Dormitzer PR, Sahin U, Jansen KU, Gruber WC, Group

- CCT. 2020. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* 383:2603-2615.
14. Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 22.
15. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, Akite N, Ho J, Lee RT, Yeo W, Curation Team GC, Maurer-Stroh S. 2021. GISAID's Role in Pandemic Response. *China CDC Wkly* 3:1049-1051.
16. Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 1:33-46.
17. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367:1444-1448.
18. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426:450-4.
19. Li F, Li W, Farzan M, Harrison SC. 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309:1864-8.
20. Yi C, Sun X, Ye J, Ding L, Liu M, Yang Z, Lu X, Zhang Y, Ma L, Gu W, Qu A, Xu J, Shi Z, Ling Z, Sun B. 2020. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol Immunol* 17:621-630.
21. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. 2020. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581:221-224.
22. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581:215-220.
23. Lin JW, Tang C, Wei HC, Du B, Chen C, Wang M, Zhou Y, Yu MX, Cheng L, Kuivanen S, Ogando NS, Levanov L, Zhao Y, Li CL, Zhou R, Li Z, Zhang Y, Sun K, Wang C, Chen L, Xiao X, Zheng X, Chen SS, Zhou Z, Yang R, Zhang D, Xu M, Song J, Wang D, Li Y, Lei S, Zeng W, Yang Q, He P, Zhang Y, Zhou L, Cao L, Luo F, Liu H, Wang L, Ye F, Zhang M, Li M, Fan W, Li X, Li K, Ke B, Xu J, Yang H, He S, et al. 2021. Genomic monitoring of SARS-CoV-2 uncovers an Nsp1 deletion variant that modulates type I interferon response. *Cell Host Microbe* 29:489-502 e8.
24. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI, Sheffield C-GG, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182:812-827 e19.
25. Young BE, Fong SW, Chan YH, Mak TM, Ang LW, Anderson DE, Lee CY, Amrun SN, Lee B, Goh YS, Su YCF, Wei WE, Kalimuddin S, Chai LYA, Pada S, Tan SY, Sun L, Parthasarathy P, Chen YYC, Barkham T, Lin RTP, Maurer-Stroh S, Leo YS, Wang LF, Renia L, Lee VJ, Smith GJD, Lye DC, Ng LFP. 2020. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* 396:603-611.
26. Xia H, Cao Z, Xie X, Zhang X, Chen JY, Wang H, Menachery VD, Rajsbaum R, Shi PY. 2020. Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep* 33:108234.

27. Min YQ, Mo Q, Wang J, Deng F, Wang H, Ning YJ. 2020. SARS-CoV-2 nsp1: Bioinformatics, Potential Structural and Functional Features, and Implications for Drug/Vaccine Designs. *Front Microbiol* 11:587317.
28. Zhou B, Thao TTN, Hoffmann D, Taddeo A, Ebert N, Labroussaa F, Pohlmann A, King J, Steiner S, Kelly JN, Portmann J, Halwe NJ, Ulrich L, Trueb BS, Fan X, Hoffmann B, Wang L, Thomann L, Lin X, Stalder H, Pozzi B, de Brot S, Jiang N, Cui D, Hossain J, Wilson MM, Keller MW, Stark TJ, Barnes JR, Dijkman R, Jores J, Benarafa C, Wentworth DE, Thiel V, Beer M. 2021. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* 592:122-127.
29. Butowt R, Bilinska K, Von Bartheld CS. 2020. Chemosensory Dysfunction in COVID-19: Integration of Genetic and Epidemiological Data Points to D614G Spike Protein Variant as a Contributing Factor. *ACS Chem Neurosci* 11:3180-3184.
30. Ho D, Wang P, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong P, Graham B, Mascola J, Chang J, Yin M, Sobieszczyk M, Kyratsous C, Shapiro L, Sheng Z, Nair M, Huang Y. 2021. Increased Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7 to Antibody Neutralization. *Res Sq* doi:10.21203/rs.3.rs-155394/v1.
31. Yuan M, Wu NC, Zhu X, Lee CD, So RTY, Lv H, Mok CKP, Wilson IA. 2020. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* 368:630-633.
32. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, Zhang Z. 2020. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol* 92:667-674.
33. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579:265-269.
34. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z. 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368:779-782.
35. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC, Zella D, Ippodrino R. 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med* 18:179.
36. Chang Y-C, Tung Y-A, Lee K-H, Chen T-F, Hsiao Y-C, Chang H-C, Hsieh T-T, Su C-H, Wang S-S, Yu J-Y. 2020. Potential therapeutic agents for COVID-19 based on the analysis of protease and RNA polymerase docking.
37. Ruan Z, Liu C, Guo Y, He Z, Huang X, Jia X, Yang T. 2021. SARS-CoV-2 and SARS-CoV: Virtual screening of potential inhibitors targeting RNA-dependent RNA polymerase activity (NSP12). *Journal of medical virology* 93:389-400.
38. Shahidi M, Mahmanzar M, Mahdavi B, Tokhanbigli S, Sisakht M, Moradi B, Ganjalikhany M, Rahimian K, Ganjalikhani-Hakemi M. 2022. SARS-CoV-2 NSP3, NSP4 and NSP6 mutations and Epistasis during the pandemic in the world: Evolutionary Trends and Natural Selections in Six Continents.
39. Rahimian K, Mahmanzar M, Mahdavi B, Arefian E, Kuehu DL, Deng Y. 2022. SARS2Mutant: SARS-CoV-2 Amino-Acid Mutation Atlas Database. *bioRxiv*.
40. Mazhari S, Alavifard H, Rahimian K, Karimi Z, Mahmanzar M, Sisakht MM, Bitaraf M, Arefian E. 2021. SARS-CoV-2 NSP-12 mutations survey during the pandemic in the world.

41. Zhou Z, Huang C, Zhou Z, Huang Z, Su L, Kang S, Chen X, Chen Q, He S, Rong X, Xiao F, Chen J, Chen S. 2021. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14(+) monocytes. *iScience* 24:102187.
42. Lin X, Fu B, Yin S, Li Z, Liu H, Zhang H, Xing N, Wang Y, Xue W, Xiong Y, Zhang S, Zhao Q, Xu S, Zhang J, Wang P, Nian W, Wang X, Wu H. 2021. ORF8 contributes to cytokine storm during SARS-CoV-2 infection by activating IL-17 pathway. *iScience* 24:102293.
43. Chen CY, Ping YH, Lee HC, Chen KH, Lee YM, Chan YJ, Lien TC, Jap TS, Lin CH, Kao LS, Chen YM. 2007. Open reading frame 8a of the human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induces apoptosis. *J Infect Dis* 196:405-15.
44. Wong HH, Fung TS, Fang S, Huang M, Le MT, Liu DX. 2018. Accessory proteins 8b and 8ab of severe acute respiratory syndrome coronavirus suppress the interferon signaling pathway by mediating ubiquitin-dependent rapid degradation of interferon regulatory factor 3. *Virology* 515:165-175.
45. Dominguez Andres A, Feng Y, Campos AR, Yin J, Yang CC, James B, Murad R, Kim H, Deshpande AJ, Gordon DE, Krogan N, Pippa R, Ronai ZA. 2020. SARS-CoV-2 ORF9c Is a Membrane-Associated Protein that Suppresses Antiviral Responses in Cells. *bioRxiv* doi:10.1101/2020.08.18.256776.
46. Trieu G, Trieu VN. 2021. Mutational analysis of SARS-CoV-2. ORF8 and the evolution of the Delta and Omicron variants. *medRxiv*.
47. Gupta S, Mallick D, Banerjee K, Mukherjee S, Sarkar S, Lee ST, Basuchowdhuri P, Jana SS. 2022. D155Y substitution of SARS-CoV-2 ORF3a weakens binding with Caveolin-1. *Comput Struct Biotechnol J* 20:766-778.
48. Sacco MD, Hu Y, Gongora MV, Meilleur F, Kemp MT, Zhang X, Wang J, Chen Y. 2022. The P132H mutation in the main protease of Omicron SARS-CoV-2 decreases thermal stability without compromising catalysis or small-molecule drug inhibition. *Cell Res* doi:10.1038/s41422-022-00640-y.
49. Shah A, Rehmat S, Aslam I, Suleman M, Batool F, Aziz A, Rashid F, Midrarullah, Nawaz MA, Ali SS, Junaid M, Khan A, Wei DQ. 2022. Comparative mutational analysis of SARS-CoV-2 isolates from Pakistan and structural-functional implications using computational modelling and simulation approaches. *Comput Biol Med* 141:105170.

TABLES

Table1. Top 10 of the First stable mutation in sars-cov2, including nsp1 to nsp16

Locus	mutant 1	mutant 2	mutant 3	mutant 4	mutant 5	mutant 6	mutant 7	mutant 8	mutant 9	mutant 10
nsp1 Rate Country Continent Date	S(135)->R	E(87)->D	H(110)->Y	R(24)->C	R(124)->C	V(28)->I	S(100)->N	E(37)->D	K(120)->N	P(62)->S
	8%	0.8%	0.2%	0.2%	0.1%	0.1%	0.08%	0.08%	0.09%	0.08%
	United States	England	United States	Sudan	Netherlands	England	France	England		
	North America	Europe	North America	Africa	Europe					
	2020-04-03	2020-01-01					2020-04-01	2020-01-01		
nsp2 Rate Country Continent Date	T(85)->I	K(81)->N	P(129)->L	L(550)->F	G(339)->S	A(419)->V	A(318)->V	E(587)->K	R(27)->C	V(485)->I
	3.9%	2.1%	1.1%	0.9%	0.8%	0.5%	0.4%	0.3%	0.3%	0.3%
	Netherlands	DRC	Russia	England		United States	England	Canada	India	United States
	Europe	Africa	Europe			North America	Europe	North America	Asia	North America
	2020-01-01	2020-05-07	2020-01-01			2020-03-12	2020-01-01	2020-04-20	2020-01-01	
nsp3 Rate Country Continent Date	P(1228)->L	P(1469)->S	A(488)->S	T(183)->I	A(890)->D	I(1412)->T	A(1711)->V	G(489)->S	T(24)->I	P(822)->L
	50%	49.9%	49.4%	16.4%	16.3%	16.2%	14.2%	11.2%	11.2%	5.1%
	England		France	England			India	Guangdong	United States	England
	Europe					Asia			North America	Europe
	2020-01-01								2020-01-26	2020-04-03
nsp4 Rate Country Continent Date	T(492)->I	V(167)->L	L(438)->F	L(264)->F	T(327)->I	A(446)->V	V(94)->A	A(146)->V	F(17)->L	A(92)->V
	70.4%	37.9%	9.4%	9.7%	9.6%	4.1%	0.7%	0.5%	0.4%	0.5%
	England	Netherlands		Germany	Wuhan	England	Australia	Wales	England	Northern Ireland
	Europe				Asia	Europe	Oceania	North America	Europe	
	2020-01-01				2019-12-30	2020-01-01	2020-08-07	2020-03-26	2020-01-01	2020-04-21

Table1. Continue

nsp5 Rate Country Continent Date	P(132)->H	K(90)->R	L(89)->F	T(169)->S	P(108)->S	A(260)->V	G(15)->S	K(88)->R	T(21)->I	P(96)->S
	30.8%	1.6%	1.4%	0.5%	0.2%	0.2%	0.2%	0.1%	0.1%	0.06%
	United States			Japan		Canada	Ireland	Netherlands	United States	France
	North America			Asia		North America	Europe		North America	Europe
	2020-01-01			2021-01-01	2020-01-01					
nsp6 Rate Country Continent Date	T(77)->A	V(149)->A	T(181)->I	I(162)->V	L(37)->F	C(197)->F	A(2)->V	I(49)->V	H(11)->Q	Q(160)->R
	71.1%	6.8%	5.5%	3.3%	2.9%	2.5%	1.1%	0.6%	0.4%	0.4%
	France	England		United States	New Zealand	Lithuania	Ecuador	United States	Italy	Singapore
	Europe			North America	Oceania	Europe	South America	North America	Europe	Asia
	2020-01-01			2020-06-24	2020-01-01	2020-04-18	2020-01-01		2020-04-11	2020-01-27
nsp7 Rate Country Continent Date	L(71)->F	S(25)->L	L(56)->F	D(77)->N	M(75)->I	M(3)->I	S(26)->F	T(81)->I	Q(63)->H	V(58)->I
	0.2%	0.1%	0.1%	0.08%	0.09%	0.05%	0.06%	0.05%	0.02%	0.02%
	England	Austria	India	Netherlands	Gibraltar	Belgium	England	Colombia	United States	
	Europe		Asia	Europe				South America	North America	
	2020-01-01		2020-03-01	2020-01-01						2020-05-12
nsp8 Rate Country Continent Date	Q(24)->R	T(145)->I	T(141)->M	T(148)->I	T(187)->I	T(123)->I	P(10)->S	T(89)->I	V(34)->F	Q(198)->H
	0.7%	0.6%	0.3%	0.1%	0.1%	0.1%	0.1%	0.08%	0.07	0.07%
	England	United States	England		United States		England	Netherlands	England	United States
	Europe	North America	Europe		North America		Europe			North America
	2020-01-01				2020-03-05	2020-01-01				

Table1. Continue

nsp9 Rate Country Continent Date	T(35)->I	I(65)->V	R(39)->K	L(42)->F	M(101)->I	T(24)->I	Q(11)->H	T(34)->I	T(109)->I	T(21)->I
	0.5%	0.4%	0.4%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%
	United States		England			Austria	United States	England	New Zealand	Austria
	North America		Europe				North America	Europe	Oceania	Europe
	2020-01-01		2020-04-05	2020-01-01			2020-04-14	2020-01-01		
nsp10 Rate Country Continent Date	T(102)->I	T(12)->I	R(134)->C	A(104)->V	A(32)->V	T(51)->I	T(111)->I	P(84)->S	T(101)->I	I(55)->T
	0.1%	0.1%	0.7%	0.1%	0.08%	0.08%	0.08%	0.04%	0.07%	0.06%
	England		Malaysia	Ecuador	Kazakhstan	England	India	Senegal	United States	
	Europe		Asia	South America	Asia	Europe	Asia	Africa	North America	
	2020-01-01		2020-03-14	2020-01-01				2020-03-25	2020-01-01	2020-03-18
nsp11 Rate Country Continent Date	S(6)->L	A(4)->V	A(2)->V	F(11)->L	A(12)->V	S(1)->L	Q(5)->R	D(3)->G	G(10)->R	F(7)->L
	0.1%	0.03%	0.02%	0.005%	0.003%	0.002%	0.001%	0.001%	1/20E-06	4/99e-06
	Greece	United States		Italy	United States	Netherlands	United States	India		Belgium
	Europe	North America		Europe	North America	Europe	North America	Asia		Europe
	2020-01-01		2020-03-26	2020-04-09	2020-03-09	2020-03-27	2020-10-11	2020-05-05	2020-09-10	2021-01-29
nsp12 Rate Country Continent Date	P(323)->L	G(671)->S	L(838)->I	F(192)->V	F(694)->Y	P(227)->L	Q(875)->R	A(185)->S	R(197)->Q	V(776)->L
	99.3%	42.8%	3%	2.5%	2%	1.8%	0.5%	0.4%	0.5%	0.4%
	England		Mexico	United States		England	United States	England	United States	England
	Europe		North America			Europe	North America	Europe	North America	Europe
	2020-01-01		2020-05-01	2020-06-23	2020-06-12	2020-01-01	2020-04-08	2020-01-01	2020-04-06	2020-01-01

Table1. Continue

nsp13 Rate Country Continent Date	P(77)->L	R(392)->C	I(334)->V	K(460)->R	H(164)->Y	E(341)->D	A(296)->S	E(261)->D	D(260)->Y	S(36)->P
	42.6%	9.8%	3.1%	2.5%	1.8%	1.4%	1%	0.9%	0.6%	0.5%
	Netherlands		England		Russia	Netherlands	England		Senegal	Hong Kong
	Europe								Africa	Asia
	2020-01-01									2020-02-22
nsp14 Rate Country Continent Date	A(394)->V	I(42)->V	N(129)->D	P(46)->L	T(16)->I	P(451)->S	A(323)->S	M(72)->I	P(43)->L	S(374)->F
	35.1%	33.9%	1.2%	0.7%	0.7%	0.5%	0.4%	0.4%	0.3%	0.3%
	Greece	United States	United States	Netherlands	England	United States	Gambia	Russia	Japan	France
	Europe	North America		Europe		North America	Africa	Europe	Asia	Europe
	2020-01-01									2020-03-01
msp15 Rate Country Continent Date	T(112)->I	P(205)->L	A(80)->V	K(259)->R	P(262)->S	H(234)->Y	V(127)->F	V(320)->L	A(171)->V	T(33)->I
	8.9%	0.2%	0.3%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%
	England	Netherlands	Oman	England	France	Sweden	Niger	England		
	Europe		Asia	Europe			Africa	Europe		
	2020-01-01		2020-04-25	2020-01-01	2020-03-05	2020-03-09	2020-01-01			
nsp16 Rate Country Continent Date	R(216)->C	K(160)->R	Q(238)->H	P(215)->L	T(35)->I	D(179)->G	R(287)->I	P(236)->S	P(134)->S	T(140)->I
	1.2%	1.3%	0.9%	0.2%	0.2%	0.2%	0.2%	0.1%	0.2%	0.1%
	Gibraltar	Wuhan	Japan	Norway	Netherlands	Saudi Arabia	Japan	France	New Zealand	Netherlands
	Europe	Asia		Europe		Asia		Europe	Oceania	Europe
	2020-01-01	2019-12-30	2020-01-01	2020-03-09	2020-01-01	2020-03-24	2020-01-01	2020-03-01	2020-01-01	

Table2. Top 10 of the First stable mutation in sars-cov2, including Orf3a to Orf9b

Locus	mutant 1	mutant 2	mutant 3	mutant 4	mutant 5	mutant 6	mutant 7	mutant 8	mutant 9	mutant 10
Orf3a	S(26)->L	T(223)->I	Q(57)->H	E(239)->Q	L(106)->F	G(172)->V	H(78)->Y	S(253)->P	D(155)->Y	W(131)->C
Rate	43.1%	10.2%	5%	2.7%	2.1%	1.3%	1.4%	1.2%	0.8%	0.7%
Country	France	United States				India	England	Brazil	Germany	Mexico
Continent	Europe	North America				Asia	Europe	South America	Europe	North America
Date	2020-01-01			2020-06-27	2020-01-01			2020-03-09	2020-01-01	
Orf6	D(61)->L	K(48)->N	E(46)->V	P(57)->L	T(21)->I	I(33)->T	W(27)->L	R(20)->S	Y(49)->H	V(24)->F
Rate	9.4%	0.4%	0.2%	0.1%	0.07%	0.1%	0.09%	0.05%	0.04%	0.03%
Country	England		United States	England	Austria	Venezuela	India	Netherlands	India	England
Continent	Europe		North America	Europe		South America	Asia	Europe	Asia	Europe
Date	2020-01-01	2020-04-06	2020-07-21	2020-01-01						
Orf7a	T(120)->I	V(82)->A	P(45)->L	L(116)->F	V(71)->I	R(118)->G	H(73)->Y	P(34)->S	P(84)->S	T(14)->I
Rate	41.4%	41.1%	1.5%	0.9%	0.6%	0.2%	0.2%	0.1%	0.1%	0.2%
Country	Netherlands	France	Saudi Arabia	France	United States	India	United States	Wuhan	Wales	England
Continent	Europe		Asia	Europe	North America	Asia	North America	Asia	North America	Europe
Date	2020-01-01		2020-02-17	2020-01-01	2020-03-20	2020-06-05	2020-01-01			
Orf7b	T(40)->I	S(5)->L	H(42)->Y	A(43)->V	S(31)->L	L(4)->F	A(15)->S	L(32)->F	M(1)->I	C(41)->F
Rate	39.5%	0.1%	0.1%	0.04%	0.09%	0.07%	0.08%	0.07%	0.05%	0.05
Country	France	England	France	Guangdong	Japan	Gambia	Greece	United States	England	Thailand
Continent	Europe			Asia		Africa	Europe	North America	Europe	Asia
Date	2020-01-01		2020-03-01	2020-01-28	2020-01-01					

Table2. Continue

Orf8 Rate Country Continent Date	S(24)->L	E(92)->K	T(11)->I	A(65)->V	F(120)->L	S(67)->F	A(51)->S	I(121)->L	V(100)->L	L(84)->S
	3.5%	2.6%	1.3%	0.5%	0.4%	0.5%	0.3%	0.4%	0.4%	0.4%
	New Zealand	England	Armenia	England	Thailand	Italy	Gibraltar	France	England	Wuhan
	Oceania	Europe	Asia	Europe	Asia	Europe				Asia
	2020-01-01									2019-12-30
Orf9b Rate Country Continent Date	T(60)->A	S(6)->C	Q(77)->E	P(10)->S	R(32)->L	S(50)->L	V(15)->L	G(38)->D	P(3)->L	H(9)->D
	61.4%	4.2%	1.8%	0.8%	0.6%	0.8%	0.6%	0.4%	0.2%	0.1%
	France	Canada	Brazil	England		Canada	Russia	England	Foshan	Niger
	Europe	North America		Europe	Europe	North America	Europe	Europe	Asia	Africa
	2020-01-01	2020-04-06	2020-03-20	2020-01-01		2020-03-01	2020-01-01		2020-01-22	2020-01-01
Orf9c Rate Country Continent Date	G(50)->N	V(49)->L	L(52)->F	L(67)->F	E(51)->Q	M(63)->T	A(59)->V	S(4)->L	H(34)->Y	V(40)->F
	48.8%	2%	1.8%	1.8%	0.8%	0.1%	0.1%	0.1%	0.1%	0.1%
	England		United States	England	Austria	United States	England			Greece
	Europe		North America	Europe		North America	Europe			
	2020-01-01					2020-03-12	2020-03-27	2020-01-01		

Table3. envelope, membrane, nucleoprotein, spike

Locus	mutant 1	mutant 2	mutant 3	mutant 4	mutant 5	mutant 6	mutant 7	mutant 8	mutant 9	mutant 10
envelope Rate Country Continent Date	T(9)->I	P(71)->L	V(62)->F	L(21)->F	V(58)->F	L(73)->F	S(55)->F	S(68)->F	V(49)->L	V(24)->M
	30.6%	0.4%	0.2%	0.1%	0.09%	0.08	0.08%	0.07%	0.03%	0.02%
	England	Sweden	France	England						
	Europe 2020-01-01									
membrane Rate Country Continent Date	I(82)->T	A(63)->T	Q(19)->E	D(3)->G	A(2)->S	V(70)->L	A(81)->S	F(28)->L	T(30)->I	S(197)->N
	47.1%	24.8%	24.6%	16%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%
	Senegal	United States		Netherlands	Greece	England	United States		Denmark	Spain
	Africa	North America		Europe			North America		Europe	
	2020-01-01						2020-01-19	2020-01-01	2020-04-27	2020-03-15
nucleoprotein Rate Country Continent Date	R(203)->M	D(377)->Y	D(63)->G	G(215)->C	G(204)->R	D(3)->L	S(235)->F	Q(9)->L	P(199)->L	A(220)->V
	62.8%	63.2%	62%	56.8%	22.3%	17.2%	17.3%	4.3%	2.5%	2.5%
	France	Netherlands	France		England			Canada	Netherlands	England
	Europe							North America	Europe	
	2020-01-01							2020-04-06	2020-01-01	
spike Rate Country Continent Date	D(614)->G	E(484)->K	L(18)->F	A(222)->V	N(501)->Y	H(655)->Y	K(417)->T	P(681)->H	V(1176)->F	D(138)->Y
	97.6%	14%	16.2%	14.9%	13.5%	12.8%	9.4%	7.7%	10.7%	10%
	Russia	England			Australia	England	United States	England		Netherlands
	Europe				Oceania	Europe	North America	Europe		
	2020-01-01						2020-04-07	2020-01-01		

FIGURES

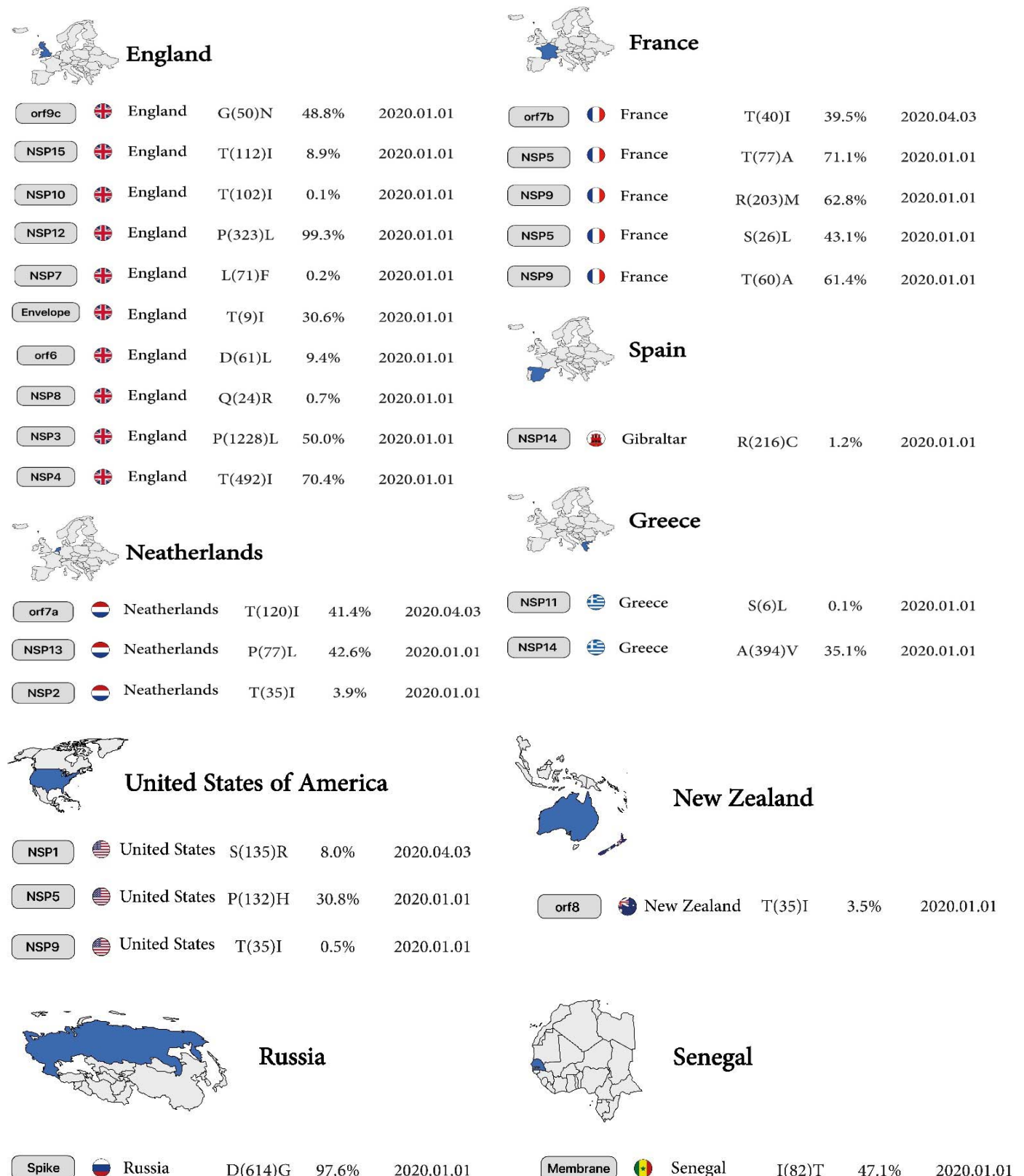


Figure 1. The first country reported sustainable mutation in each gene. From left to right, gene name, country name, Amino acid substitutions and positions, the Mutation frequency of all mutations occurring in that gene, and date of the first sample include mentioned mutation.

NSP1					
1		United States	S(135)R	8.01%	2020.04.03
2		England	E(87)D	0.85%	2020.01.01
3		United States	H(110)Y	0.24%	2020.01.01
10		England	P(62)S	0.08%	2020.01.01
NSP2					
1		Netherlands	T(85)I	3.94%	2020.01.01
10		United States	V(485)I	0.33%	2020.01.01
NSP3					
1		England	P(1228)L	50.02%	2020.01.01
2		England	P(1469)S	49.96%	2020.01.01
3		France	A(488)S	49.49%	2020.01.01
4		England	T(183)I	16.40%	2020.01.01
5		England	A(890)D	16.30%	2020.01.01
6		England	P(822)L	5.17%	2020.01.01
NSP4					
1		England	T(492)I	70.48%	2020.01.01
2		Netherlands	V(167)L	37.98%	2020.01.01
10		Northern Ireland	A(92)V	0.50%	2020.04.21
NSP5					
3		United States	L(89)F	1.4%	2020.01.01
6		Canada	A(260)V	0.2%	2020.01.01
NSP6					
5		New Zealand	L(37)F	2.99%	2020.01.01
7		Ecuador	A(2)V	1.16%	2020.01.01
NSP7					
1		England	L(71)F	0.22%	2020.01.01
3		India	L(56)F	0.10%	2020.03.01

Figure 2A: First stable mutations in SARS-CoV-2, including nsp1-nsp7

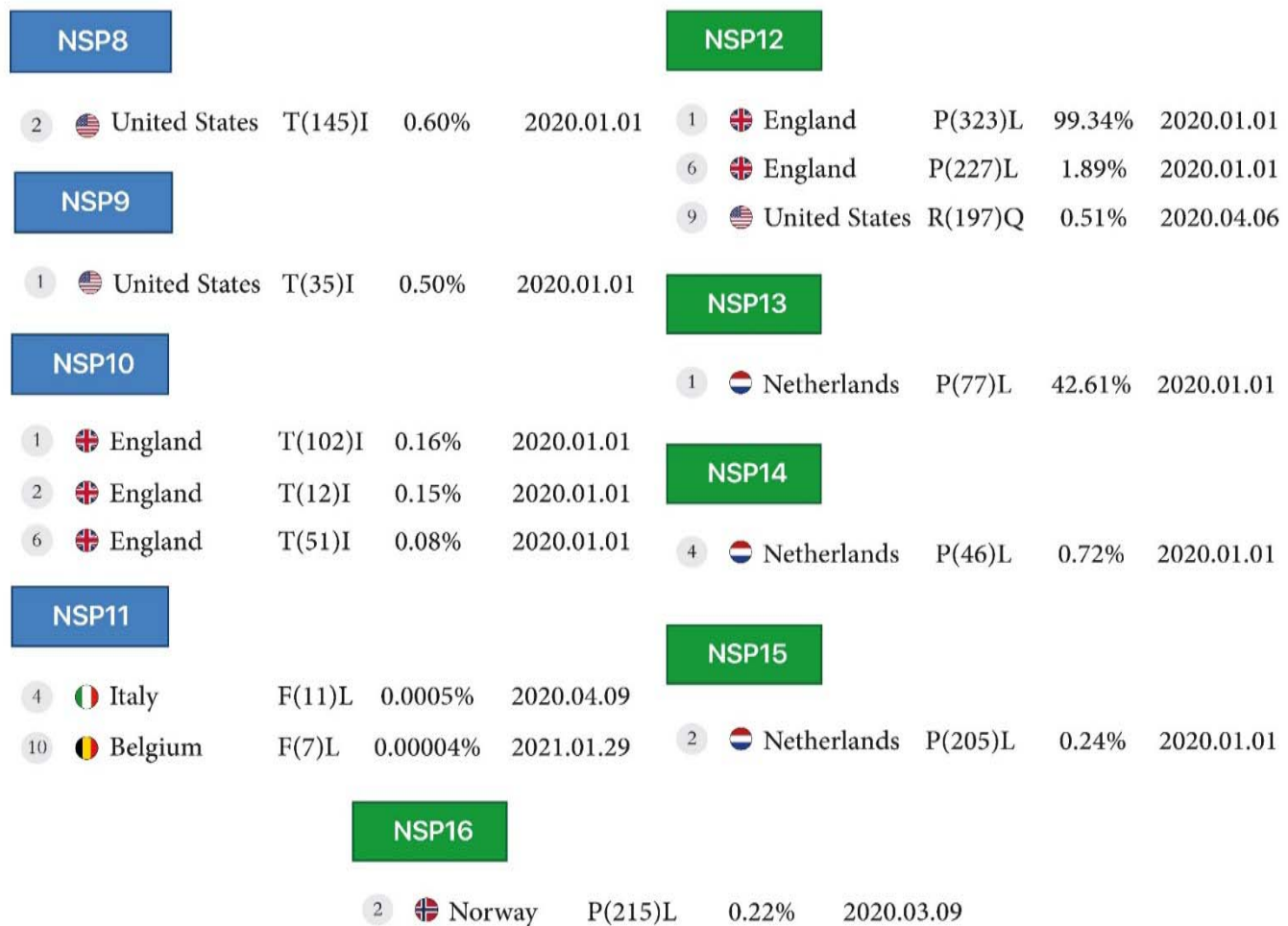


Figure 2B: First stable mutations in SARS-CoV-2, including nsp8-nsp16

Orf3a					Orf7a				
1	 France	S(26)L	43.1%	2020.01.01	1	 Netherlands	T(120)I	41.4%	2020.01.01
2	 United States	T(223)I	10.2%	2020.01.01	2	 France	V(82)A	41.1%	2020.01.01
3	 United States	Q(57)H	5.0%	2020.01.01	3	 Saudi Arabia	P(45)L	1.5%	2020.02.17
4	 United States	E(239)Q	2.7%	2020.06.27	4	 France	L(116)F	0.9%	2020.01.01
5	 United States	L(106)F	2.1%	2020.01.01	5	 United States	V(71)I	0.6%	2020.03.20
6	 India	G(172)V	1.3%	2020.01.01	6	 India	R(118)G	0.2%	2020.06.05
7	 England	H(78)Y	1.4%	2020.01.01	7	 United States	H(73)Y	0.2%	2020.01.01
8	 Brazil	S(253)P	1.2%	2020.03.09	8	 Wuhan	P(34)S	0.1%	2020.01.01
9	 Germany	D(155)Y	0.8%	2020.01.01	9	 Wales	P(84)S	0.1%	2020.01.01
10	 Mexico	W(131)C	0.7%	2020.01.01	10	 England	T(14)I	0.2%	2020.01.01
Orf6					Orf8				
1	 England	D(61)L	9.4%	2020.01.01	1	 New Zealand	S(24)L	3.53%	2020.01.01
2	 England	K(48)N	0.4%	2020.04.06	2	 England	E(92)K	2.65%	2020.01.01
3	 United States	E(46)V	0.2%	2020.07.21	3	 Armenia	T(11)I	1.34%	2020.01.01
4	 England	P(57)L	0.1%	2020.01.01	4	 England	A(65)V	0.55%	2020.01.01
5	 Austria	T(21)I	0.07%	2020.01.01	5	 Thailand	F(120)L	0.45%	2020.01.01
6	 Venezuela	I(33)T	0.1%	2020.01.01	6	 Italy	S(67)F	0.58%	2020.01.01
7	 India	W(27)L	0.09%	2020.01.01	7	 Gibraltar	A(51)S	0.32%	2020.01.01
8	 Netherlands	R(20)S	0.05%	2020.01.01	8	 France	I(121)L	0.44%	2020.01.01
9	 India	Y(49)H	0.04%	2020.01.01	9	 England	V(100)L	0.44%	2020.01.01
10	 England	V(24)F	0.03%	2020.01.01	10	 Wuhan	L(84)S	0.40%	2019.12.30

Figure 3A: First stable mutations in SARS-CoV-2 including ORF3a, ORF6, ORF7a, ORF8

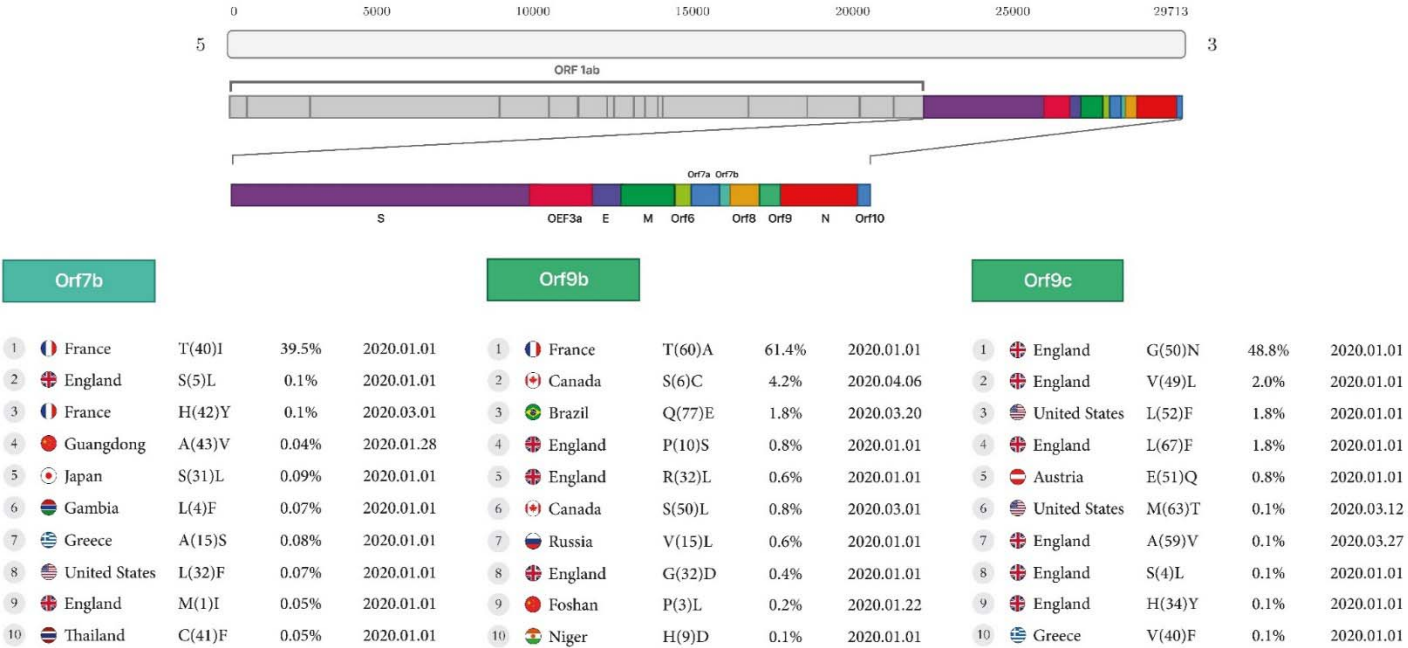


Figure 3B: First stable mutations in SARS-CoV-2 including (ORF7b, ORF9b, ORF9c)

Spike					Envelope				
1	 Russia	D(614)G	97.6%	2020.01.01	1	 England	T(9)I	30.6%	2020.01.01
2	 England	E(484)K	14.0%	2020.01.01	2	 Sweden	P(71)L	0.4%	2020.01.01
3	 England	L(87)F	16.2%	2020.01.01	3	 France	V(62)F	0.2%	2020.01.01
4	 England	A(222)V	14.9%	2020.01.01	4	 England	L(21)F	0.1%	2020.01.01
5	 Australia	N(501)Y	13.5%	2020.01.01	5	 England	V(58)F	0.09%	2020.01.01
6	 England	H(655)Y	12.8%	2020.01.01	6	 England	L(73)F	0.08%	2020.01.01
7	 United States	K(417)T	9.4%	2020.04.07	7	 England	S(55)F	0.08%	2020.01.01
8	 England	P(681)H	7.7%	2020.01.01	8	 England	S(68)F	0.07%	2020.01.01
9	 England	V(1176)F	10.7%	2020.01.01	9	 England	V(49)L	0.03%	2020.01.01
10	 Netherlands	D(138)Y	10.0%	2020.01.01	10	 England	V(24)M	0.02%	2020.01.01
Nucleoprotein					Membrane				
1	 France	R(203)M	62.8%	2020.01.01	1	 Senegal	I(82)T	47.1%	2020.01.01
2	 Netherlands	D(377)Y	63.2%	2020.01.01	2	 United States	A(63)T	24.8%	2020.01.01
3	 France	D(63)G	62.0%	2020.01.01	3	 United States	Q(19)E	24.6%	2020.01.01
4	 France	G(215)C	56.8%	2020.01.01	4	 Netherlands	D(3)G	16.0%	2020.01.01
5	 England	G(204)R	22.3%	2020.01.01	5	 Greece	A(2)S	0.2%	2020.01.01
6	 England	D(3)L	17.2%	2020.01.01	6	 England	V(70)L	0.2%	2020.01.01
7	 England	S(235)F	17.3%	2020.01.01	7	 United States	A(81)S	0.2%	2020.01.19
8	 Canada	Q(9)L	4.3%	2020.04.06	8	 United States	F(28)L	0.2%	2020.01.01
9	 Netherlands	P(199)L	2.5%	2020.01.01	9	 Denmark	T(30)I	0.2%	2020.04.27
10	 England	A(220)V	2.5%	2020.01.01	10	 Spain	S(197)N	0.2%	2020.03.15

Figure 4. First stable mutation in sars-cov2 including Envelope, Membrane, Nucleoprotein, and Spike.

NSP1

1		United States	S(135)R	8.0%	2020.04.03
2		England	E(87)D	0.8%	2020.01.01
3		United States	H(110)Y	0.2%	2020.01.01
10		England	P(62)S	0.08%	2020.01.01

bioRxiv preprint doi: <https://doi.org/10.1101/2022.07.18.500565>; this version posted July 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

NSP2

1		Netherlands	T(85)I	3.9%	2020.01.01
10		United States	V(485)I	0.3%	2020.01.01

NSP3

1		England	P(1228)L	50.0%	2020.01.01
2		England	P(1469)S	49.9%	2020.01.01
3		France	A(488)S	49.4%	2020.01.01
4		England	T(183)I	16.4%	2020.01.01
5		England	A(890)D	16.3%	2020.01.01
6		England	P(822)L	5.1%	2020.01.01

NSP4

1		England	T(492)I	70.4%	2020.01.01
2		Netherlands	V(167)L	37.9%	2020.01.01
10		Northern Ireland	A(92)V	0.5%	2020.04.21

NSP5

3		United States	L(89)F	1.4%	2020.01.01
6		Canada	A(260)V	0.2%	2020.01.01

NSP6

5		New Zealand	L(37)F	2.9%	2020.01.01
7		Ecuador	A(2)V	1.1%	2020.01.01

NSP7

1		England	L(71)F	0.2%	2020.01.01
3		India	L(56)F	0.1%	2020.03.01

NSP8

2		United States	T(145)I	0.6%	2020.01.01
---	---	---------------	---------	------	------------

NSP9

1		United States	T(35)I	0.5%	2020.01.01
---	---	---------------	--------	------	------------

NSP10

1		England	T(102)I	0.1%	2020.01.01
2		England	T(12)I	0.1%	2020.01.01
6		England	T(51)I	0.08%	2020.01.01

NSP11

4		Italy	F(11)L	0.0005%	2020.04.09
10		Belgium	F(7)L	0.00004%	2021.01.29

NSP12

1		England	P(323)L	99.3%	2020.01.01
6		England	P(227)L	1.8%	2020.01.01
9		United States	R(197)Q	0.5%	2020.04.06

NSP13

1		Netherlands	P(77)L	42.6%	2020.01.01
---	---	-------------	--------	-------	------------

NSP14

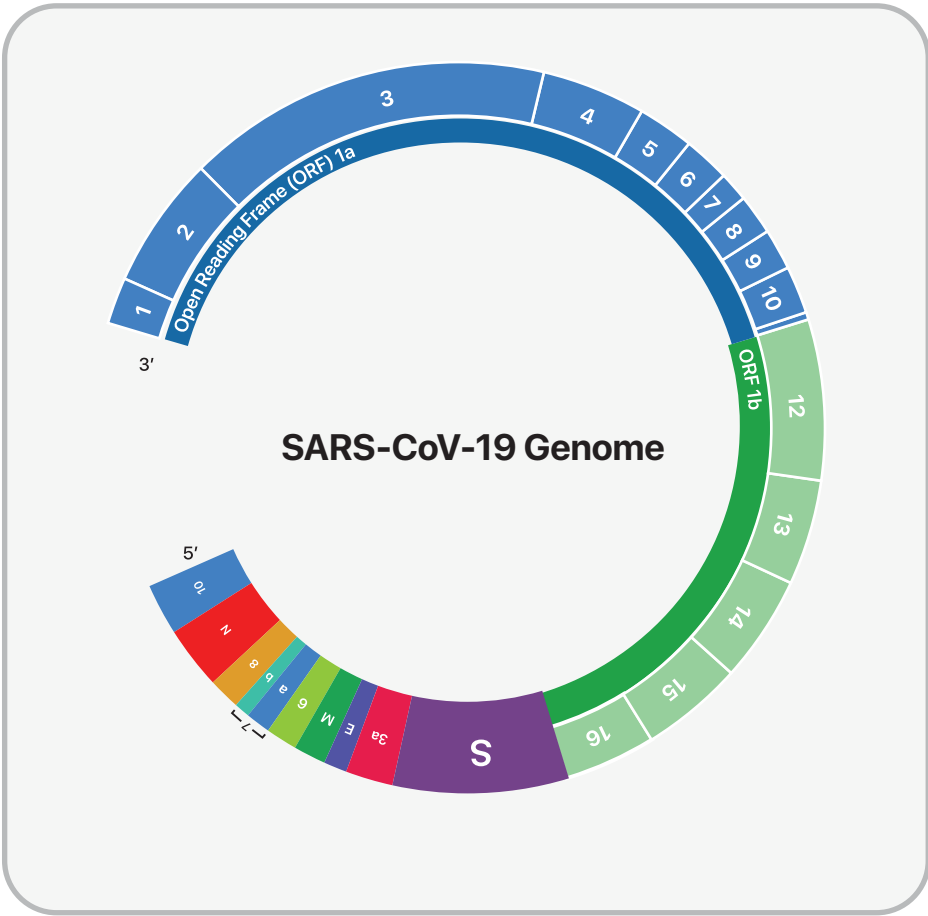
4		Netherlands	P(46)L	0.7%	2020.01.01
---	---	-------------	--------	------	------------

NSP15

2		Netherlands	P(205)L	0.2%	2020.01.01
---	--	-------------	---------	------	------------

NSP16

2		Norway	P(215)L	0.2%	2020.03.09
---	---	--------	---------	------	------------



Orf3a

bioRxiv preprint doi: <https://doi.org/10.1101/2022.07.18.500565>; this version posted July 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1		France	S(26)L	43.1%	2020.01.01
2		United States	T(223)I	10.2%	2020.01.01
3		United States	Q(57)H	5.0%	2020.01.01
4		United States	E(239)Q	2.7%	2020.06.27
5		United States	L(106)F	2.1%	2020.01.01
6		India	G(172)V	1.3%	2020.01.01
7		England	H(78)Y	1.4%	2020.01.01
8		Brazil	S(253)P	1.2%	2020.03.09
9		Germany	D(155)Y	0.8%	2020.01.01
10		Mexico	W(131)C	0.7%	2020.01.01

Orf6

1		England	D(61)L	9.4%	2020.01.01
2		England	K(48)N	0.4%	2020.04.06
3		United States	E(46)V	0.2%	2020.07.21
4		England	P(57)L	0.1%	2020.01.01
5		Austria	T(21)I	0.07%	2020.01.01
6		Venezuela	I(33)T	0.1%	2020.01.01
7		India	W(27)L	0.09%	2020.01.01
8		Netherlands	R(20)S	0.05%	2020.01.01
9		India	Y(49)H	0.04%	2020.01.01
10		England	V(24)F	0.03%	2020.01.01

Orf8

1		New Zealand	S(24)L	3.53%	2020.01.01
2		England	E(92)K	2.65%	2020.01.01
3		Armenia	T(11)I	1.34%	2020.01.01
4		England	A(65)V	0.55%	2020.01.01
5		Thailand	F(120)L	0.45%	2020.01.01
6		Italy	S(67)F	0.58%	2020.01.01
7		Gibraltar	A(51)S	0.32%	2020.01.01
8		France	I(121)L	0.44%	2020.01.01
9		England	V(100)L	0.44%	2020.01.01
10		Wuhan	L(84)S	0.40%	2019.12.30

Orf7a

1		Netherlands	T(120)I	41.4%	2020.01.01
2		France	V(82)A	41.1%	2020.01.01
3		Saudi Arabia	P(45)L	1.5%	2020.02.17
4		Farnce	L(116)F	0.9%	2020.01.01
5		United States	V(71)I	0.6%	2020.03.20
6		India	R(118)G	0.2%	2020.06.05
7		United States	H(73)Y	0.2%	2020.01.01
8		Wuhan	P(34)S	0.1%	2020.01.01
9		Wales	P(84)S	0.1%	2020.01.01
10		England	T(14)I	0.2%	2020.01.01

Orf7b

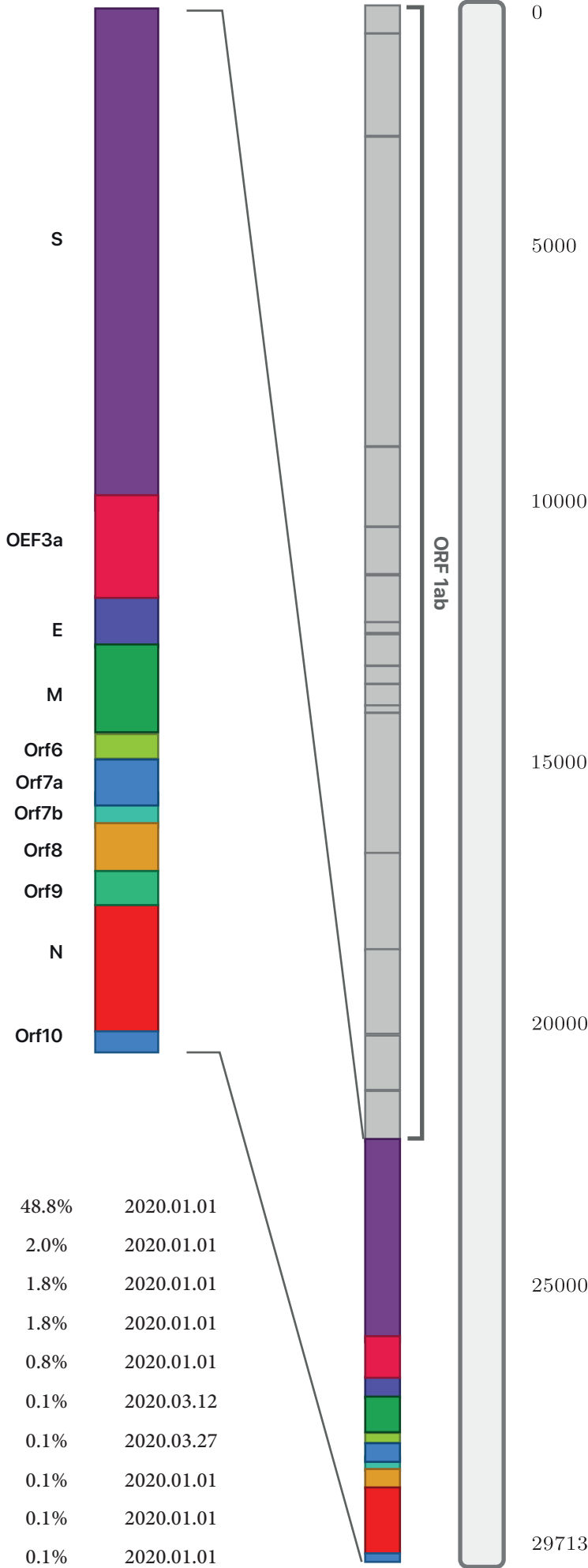
1		France	T(40)I	39.5%	2020.01.01
2		England	S(5)L	0.1%	2020.01.01
3		France	H(42)Y	0.1%	2020.03.01
4		Guangdong	A(43)V	0.04%	2020.01.28
5		Japan	S(31)L	0.09%	2020.01.01
6		Gambia	L(4)F	0.07%	2020.01.01
7		Greece	A(15)S	0.08%	2020.01.01
8		United States	L(32)F	0.07%	2020.01.01
9		England	M(1)I	0.05%	2020.01.01
10		Thailand	C(41)F	0.05%	2020.01.01

Orf9b

1		France	T(60)A	61.4%	2020.01.01
2		Canada	S(6)C	4.2%	2020.04.06
3		Brazil	Q(77)E	1.8%	2020.03.20
4		England	P(10)S	0.8%	2020.01.01
5		England	R(32)L	0.6%	2020.01.01
6		Canada	S(50)L	0.8%	2020.03.01
7		Russia	V(15)L	0.6%	2020.01.01
8		England	G(32)D	0.4%	2020.01.01
9		Foshan	P(3)L	0.2%	2020.01.22
10		Niger	H(9)D	0.1%	2020.01.01

Orf9c

1		England	G(50)N	48.8%	2020.01.01
2		England	V(49)L	2.0%	2020.01.01
3		United States	L(52)F	1.8%	2020.01.01
4		England	L(67)F	1.8%	2020.01.01
5		Austria	E(51)Q	0.8%	2020.01.01
6		United States	M(63)T	0.1%	2020.03.12
7		England	A(59)V	0.1%	2020.03.27
8		England	S(4)L	0.1%	2020.01.01
9		England	H(34)Y	0.1%	2020.01.01
10		Greece	V(40)F	0.1%	2020.01.01

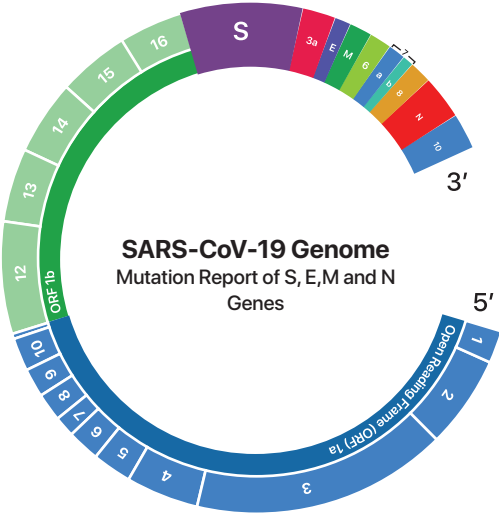


Spike

1		Russia	D(614)G	97.6%	2020.01.01
2		England	E(484)K	14.0%	2020.01.01
3		England	L(87)F	16.2%	2020.01.01
4		England	A(222)V	14.9%	2020.01.01
5		Australia	N(501)Y	13.5%	2020.01.01
6		England	H(655)Y	12.8%	2020.01.01
7		United States	K(417)T	9.4%	2020.04.07
8		England	P(681)H	7.7%	2020.01.01
9		England	V(1176)F	10.7%	2020.01.01
10		Netherlands	D(138)Y	10.0%	2020.01.01

Nucleoprotein

1		France	R(203)M	62.8%	2020.01.01
2		Netherlands	D(377)Y	63.2%	2020.01.01
3		France	D(63)G	62.0%	2020.01.01
4		France	G(215)C	56.8%	2020.01.01
5		England	G(204)R	22.3%	2020.01.01
6		England	D(3)L	17.2%	2020.01.01
7		England	S(235)F	17.3%	2020.01.01
8		Canada	Q(9)L	4.3%	2020.04.06
9		Netherlands	P(199)L	2.5%	2020.01.01
10		England	A(220)V	2.5%	2020.01.01



Envelope

1		England	T(9)I	30.6%	2020.01.01
2		Sweden	P(71)L	0.4%	2020.01.01
3		France	V(62)F	0.2%	2020.01.01
4		England	L(21)F	0.1%	2020.01.01
5		England	V(58)F	0.09%	2020.01.01
6		England	L(73)F	0.08%	2020.01.01
7		England	S(55)F	0.08%	2020.01.01
8		England	S(68)F	0.07%	2020.01.01
9		England	V(49)L	0.03%	2020.01.01
10		England	V(24)M	0.02%	2020.01.01

Membrane

1		Senegal	I(82)T	47.1%	2020.01.01
2		United States	A(63)T	24.8%	2020.01.01
3		United States	Q(19)E	24.6%	2020.01.01
4		Netherlands	D(3)G	16.0%	2020.01.01
5		Greece	A(2)S	0.2%	2020.01.01
6		England	V(70)L	0.2%	2020.01.01
7		United States	A(81)S	0.2%	2020.01.19
8		United States	F(28)L	0.2%	2020.01.01
9		Denmark	T(30)I	0.2%	2020.04.27
10		Spain	S(197)N	0.2%	2020.03.15

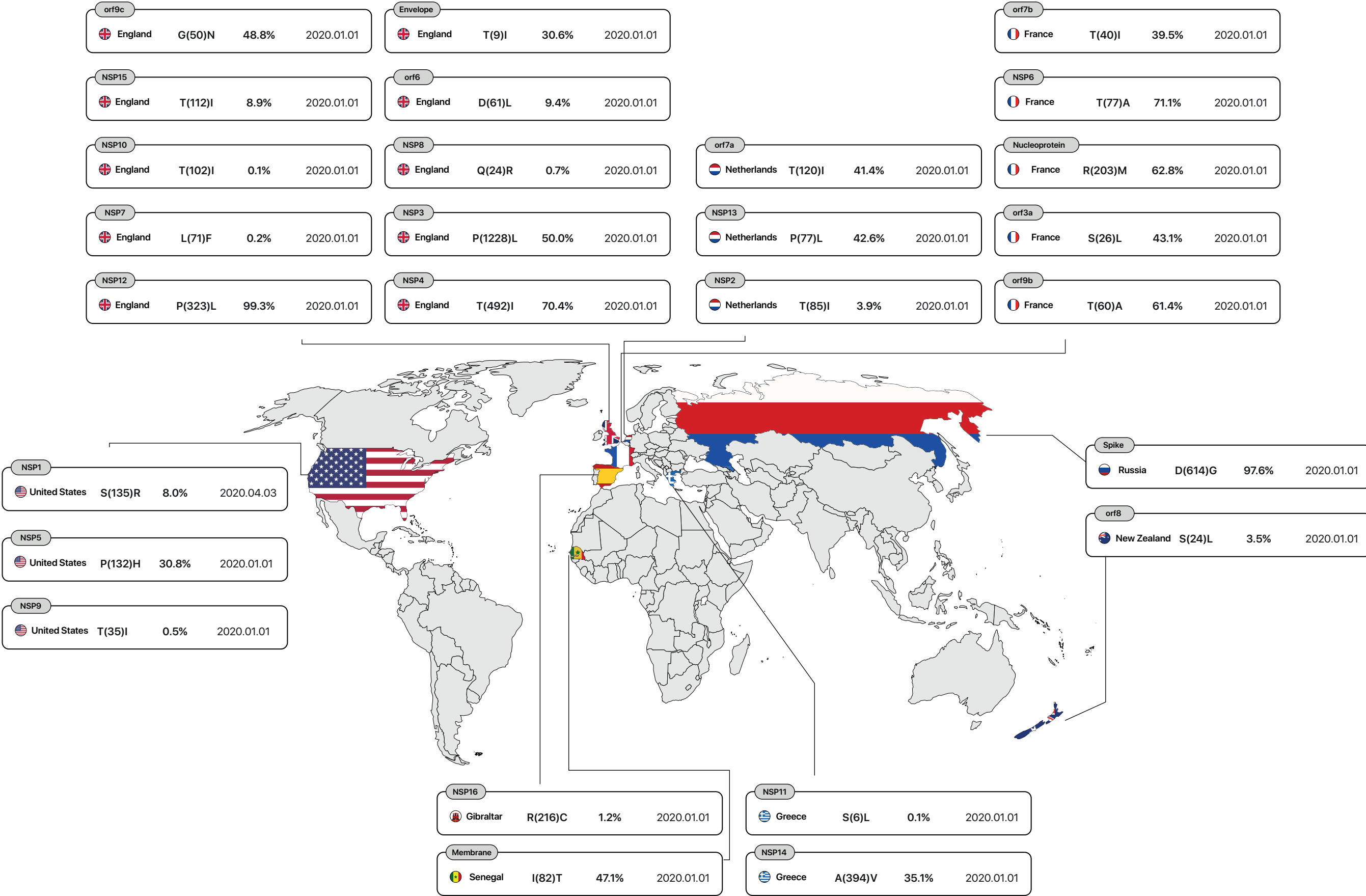


Table1. Top 10 of the First stable mutation in sars-cov2 including: nsp1 to nsp16										
Locus	mutant 1	mutant 2	mutant 3	mutant 4	mutant 5	mutant 6	mutant 7	mutant 8	mutant 9	mutant 10
nsp1	S(135)->R	E(87)->D	H(110)->Y	R(24)->C	R(124)->C	V(28)->I	S(100)->N	E(37)->D	K(120)->N	P(62)->S
Rate	8%	0.8%	0.2%	0.2%	0.1%	0.1%	0.08%	0.08%	0.09%	0.08%
Country	United States	England	United States	Sudan	Netherlands	England	France		England	
Continent	North America	Europe	North America	Africa			Europe			
Date	2020-04-03			2020-01-01			2020-04-01		2020-01-01	
nsp2	T(85)->I	K(81)->N	P(129)->L	L(550)->F	G(339)->S	A(419)->V	A(318)->V	E(587)->K	R(27)->C	V(485)->I
Rate	3.9%	2.1%	1.1%	0.9%	0.8%	0.5%	0.4%	0.3%	0.3%	0.3%
Country	Netherlands	DRC	Russia	England		United States	England	Canada	India	United States
Continent	Europe	Africa		Europe		North America	Europe	North America	Asia	North America
Date	2020-01-01	2020-05-07		2020-01-01		2020-03-12	2020-01-01	2020-04-20	2020-01-01	
nsp3	P(1228)->L	P(1469)->S	A(488)->S	T(183)->I	A(890)->D	I(1412)->T	A(1711)->V	G(489)->S	T(24)->I	P(822)->L
Rate	50%	49.9%	49.4%	16.4%	16.3%	16.2%	14.2%	11.2%	11.2%	5.1%
Country	England		France		England		India	Guangdong	United States	England
Continent			Europe				Asia		North America	Europe
Date				2020-01-01				2020-01-26	2020-04-03	2020-01-01
nsp4	T(492)->I	V(167)->L	L(438)->F	L(264)->F	T(327)->I	A(446)->V	V(94)->A	A(146)->V	F(17)->L	A(92)->V
Rate	70.4%	37.9%	9.4%	9.7%	9.6%	4.1%	0.7%	0.5%	0.4%	0.5%
Country	England	Netherlands		Germany	Wuhan	England	Australia	Wales	England	Northern Ireland
Continent		Europe			Asia	Europe	Oceania	North America	Europe	
Date		2020-01-01			2019-12-30	2020-01-01	2020-08-07	2020-03-26	2020-01-01	2020-04-21

Table1. Continue											
nsp5 Rate Country Continent Date	P(132)->H	K(90)->R	L(89)->F	T(169)->S	P(108)->S	A(260)->V	G(15)->S	K(88)->R	T(21)->I	P(96)->S	
	30.8%	1.6%	1.4%	0.5%	0.2%	0.2%	0.2%	0.1%	0.1%	0.06%	
	United States			Japan		Canada	Ireland	Netherlands	United States	France	
	North America			Asia		North America	Europe		North America	Europe	
	2020-01-01			2021-01-01		2020-01-01			2020-03-01		
nsp6 Rate Country Continent Date	T(77)->A	V(149)->A	T(181)->I	I(162)->V	L(37)->F	C(197)->F	A(2)->V	I(49)->V	H(11)->Q	Q(160)->R	
	71.1%	6.8%	5.5%	3.3%	2.9%	2.5%	1.1%	0.6%	0.4%	0.4%	
	France	England		United States	New Zealand	Lithuania	Ecuador	United States	Italy	Singapore	
	Europe			North America	Oceania	Europe	South America	North America	Europe	Asia	
	2020-01-01			2020-06-24	2020-01-01	2020-04-18	2020-01-01		2020-04-11	2020-01-27	
nsp7 Rate Country Continent Date	L(71)->F	S(25)->L	L(56)->F	D(77)->N	M(75)->I	M(3)->I	S(26)->F	T(81)->I	Q(63)->H	V(58)->I	
	0.2%	0.1%	0.1%	0.08%	0.09%	0.05%	0.06%	0.05%	0.02%	0.02%	
	England	Austria	India	Netherlands	Gibraltar	Belgium	England	Colombia	United States		
	Europe		Asia	Europe			South America		North America		
	2020-01-01		2020-03-01		2020-01-01			2020-05-12		2020-03-17	
nsp8 Rate Country Continent Date	Q(24)->R	T(145)->I	T(141)->M	T(148)->I	T(187)->I	T(123)->I	P(10)->S	T(89)->I	V(34)->F	Q(198)->H	
	0.7%	0.6%	0.3%	0.1%	0.1%	0.1%	0.1%	0.08%	0.07	0.07%	
	England	United States		England		United States		England	Netherlands	England	United States
	Europe	North America		Europe		North America		Europe		North America	
	2020-01-01			2020-03-05			2020-01-01				

Table1. Continue											
nsp9	T(35)->I	I(65)->V	R(39)->K	L(42)->F	M(101)->I	T(24)->I	Q(11)->H	T(34)->I	T(109)->I	T(21)->I	
	Rate	0.5%	0.4%	0.4%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	
	Country	United States			England		Austria	United States	England	New Zealand	Austria
	Continent	North America			Europe			North America	Europe	Oceania	Europe
	Date	2020-01-01		2020-04-05		2020-01-01		2020-04-14		2020-01-01	
nsp10	T(102)->I	T(12)->I	R(134)->C	A(104)->V	A(32)->V	T(51)->I	T(111)->I	P(84)->S	T(101)->I	I(55)->T	
	Rate	0.1%	0.1%	0.7%	0.1%	0.08%	0.08%	0.08%	0.04%	0.07%	0.06%
	Country	England		Malaysia	Ecuador	Kazakhstan	England	India	Senegal	United States	
	Continent	Europe		Asia	South America	Asia	Europe	Asia	Africa	North America	
	Date	2020-01-01		2020-03-14		2020-01-01			2020-03-25	2020-01-01	2020-03-18
nsp11	S(6)->L	A(4)->V	A(2)->V	F(11)->L	A(12)->V	S(1)->L	Q(5)->R	D(3)->G	G(10)->R	F(7)->L	
	Rate	0.1%	0.03%	0.02%	0.005%	0.003%	0.002%	0.001%	0.001%	1/20E-06	4/99e-06
	Country	Greece	United States		Italy	United States	Netherlands	United States	India		Belgium
	Continent	Europe	north America		Europe	North America	Europe	North America	Asia		Europe
	Date	2020-01-01		2020-03-26		2020-04-09	2020-03-09	2020-03-27	2020-10-11	2020-05-05	2020-09-10
nsp12	P(323)->L	G(671)->S	L(838)->I	F(192)->V	F(694)->Y	P(227)->L	Q(875)->R	A(185)->S	R(197)->Q	V(776)->L	
	Rate	99.3%	42.8%	3%	2.5%	2%	1.8%	0.5%	0.4%	0.5%	0.4%
	Country	England		Mexico	United States		England	United States	England	United States	England
	Continent	Europe		north America			Europe	North America	Europe	North America	Europe
	Date	2020-01-01		2020-05-01		2020-06-23	2020-06-12	2020-01-01	2020-04-08	2020-01-01	2020-04-06

Table2. Top 10 of the First stable mutation in sars-cov2 including: Orf3a to Orf9b										
Locus	mutant 1	mutant 2	mutant 3	mutant 4	mutant 5	mutant 6	mutant 7	mutant 8	mutant 9	mutant 10
Orf3a	S(26)->L	T(223)->I	Q(57)->H	E(239)->Q	L(106)->F	G(172)->V	H(78)->Y	S(253)->P	D(155)->Y	W(131)->C
Rate	43.1%	10.2%	5%	2.7%	2.1%	1.3%	1.4%	1.2%	0.8%	0.7%
Country	France	United States				India	England	Brazil	Germany	Mexico
Continent	Europe	north America				Asia	Europe	South America	Europe	North America
Date	2020-01-01		2020-06-27			2020-01-01		2020-03-09		2020-01-01
Orf6	D(61)->L	K(48)->N	E(46)->V	P(57)->L	T(21)->I	I(33)->T	W(27)->L	R(20)->S	Y(49)->H	V(24)->F
Rate	9.4%	0.4%	0.2%	0.1%	0.07%	0.1%	0.09%	0.05%	0.04%	0.03%
Country	England		United States	England	Austria	Venezuela	India	Netherlands	India	England
Continent	Europe		north America	Europe		South America	Asia	Europe	Asia	Europe
Date	2020-01-01	2020-04-06	2020-07-21				2020-01-01			
Orf7a	T(120)->I	V(82)->A	P(45)->L	L(116)->F	V(71)->I	R(118)->G	H(73)->Y	P(34)->S	P(84)->S	T(14)->I
Rate	41.4%	41.1%	1.5%	0.9%	0.6%	0.2%	0.2%	0.1%	0.1%	0.2%
Country	Netherlands	France	Saudi Arabia	France	United States	India	United States	Wuhan	Wales	England
Continent	Europe		Asia	Europe	North America	Asia	North America	Asia	North America	Europe
Date	2020-01-01		2020-02-17	2020-01-01	2020-03-20	2020-06-05		2020-01-01		
Orf7b	T(40)->I	S(5)->L	H(42)->Y	A(43)->V	S(31)->L	L(4)->F	A(15)->S	L(32)->F	M(1)->I	C(41)->F
Rate	39.5%	0.1%	0.1%	0.04%	0.09%	0.07%	0.08%	0.07%	0.05%	0.05
Country	France	England	France	Guangdong	Japan	Gambia	Greece	United States	England	Thailand
Continent	Europe			Asia		Africa	Europe	North America	Europe	Asia
Date	2020-01-01		2020-03-01	2020-01-28			2020-01-01			

Table2. Continue										
Orf8	S(24)->L	E(92)->K	T(11)->I	A(65)->V	F(120)->L	S(67)->F	A(51)->S	I(121)->L	V(100)->L	L(84)->S
Rate	3.5%	2.6%	1.3%	0.5%	0.4%	0.5%	0.3%	0.4%	0.4%	0.4%
Country	New Zealand	England	Armenia	England	Thailand	Italy	Gibraltar	France	England	Wuhan
Continent	Oceania	Europe	Asia	Europe	Asia	Europe				Asia
Date	2020-01-01									2019-12-30
Orf9b	T(60)->A	S(6)->C	Q(77)->E	P(10)->S	R(32)->L	S(50)->L	V(15)->L	G(38)->D	P(3)->L	H(9)->D
Rate	61.4%	4.2%	1.8%	0.8%	0.6%	0.8%	0.6%	0.4%	0.2%	0.1%
Country	France	Canada	Brazil	England		Canada	Russia	England	Foshan	Niger
Continent	Europe	north America		Europe	Europe	North America	Europe	Europe	Asia	Africa
Date	2020-01-01	2020-04-06	2020-03-20	2020-01-01		2020-03-01	2020-01-01		2020-01-22	2020-01-01
Orf9c	G(50)->N	V(49)->L	L(52)->F	L(67)->F	E(51)->Q	M(63)->T	A(59)->V	S(4)->L	H(34)->Y	V(40)->F
Rate	48.8%	2%	1.8%	1.8%	0.8%	0.1%	0.1%	0.1%	0.1%	0.1%
Country	England		United States	England	Austria	United States		England		Greece
Continent	Europe		North America	Europe		North America		Europe		
Date	2020-01-01					2020-03-12	2020-03-27	2020-01-01		

