Repeated emergence of probabilistically and chronologically anomalous mutations in SARS-CoV-2 during the COVID-19 pandemic

Hideki Kakeya and Yoshihisa Matsumoto

Abstract

Mutations in 22 major SARS-CoV-2 variants before June 2023 are investigated. It is found that mutation spectra of spike (surface glycoprotein) in Alpha and Delta variants are deviant from the normal human SARS-CoV-2 mutation spectrum with statistical significance. Sudden surges of G614D mutation, a reversion of the earliest and the most dominant point mutation in SARS-CoV-2, are observed in the non-early periods of Delta and Omicron BA.2 prevalence, which does not agree with the expectation of natural evolution. Outstanding number of unique pure reversion mutants consisting only of reverse mutations are observed in the beginning of Omicron BA.1 prevalence, many of which can emerge only through a large amount of recombination. Community spread can hardly cause these probabilistically and chronologically anomalous mutation patterns.

Keywords: SARS-CoV-2, variant of concern, reverse mutation, mutation spectrum, surface glycoprotein

Introduction

Many variants of concern (VOCs) have emerged one after another since the first outbreak of SARS-CoV-2 in the end of 2019. One of the most conspicuous features of VOCs in the early stage of the pandemic is independence, where the surface glycoprotein (spike protein) of each VOC has few mutations in common. Another feature worth mentioning is the strong bias towards nonsynonymous (N) mutations over synonymous (S) mutations in the spike protein. Especially, the early strains of the Omicron variant had about 30 or more N mutations in the spike protein alone [1] while it had only one S mutation.

The independence of mutations among the VOCs have been discussed from various perspectives. The Alpha variant (B.1.1.7), which was the earliest VOCs of all, had many mutations that had not been observed before in the minor variants of the original Wuhan strain, which was speculated to have been caused by incubation in an immune-compromised patient [2]. Hassan et al. surveyed mutations in various VOCs across the continents to find that many mutations were specific to each location [3], which could cause emergence of independent mutations.

Phylogenetic analysis shows that the Omicron variant did not emerge from the other precedent VOCs [4]. Some discuss that the possible origins are either unknown human population under strong selective pressure to escape from vaccine-induced immune response, incubation in an immunocompromised patient, or evolution in a non-human host before spilling over back to human [5,6]. However, the highly vaccinated populations, who usually reside in medically advanced countries, are basically well-monitored, which means that evolution accompanying many mutations without being noticed is practically impossible. As for the incubation in an immunocompromised patient, the count of mutations observed so far is around 10 or fewer [7-9], which is not comparable to that observed in the Omicron variant.

The most wide-spread metric to evaluate selective pressure is dN/dS (Ka/Ks), which compares N mutations to S mutations [10,11]. Wei et al. argue that dN/dS as high as 6.64, which is observed in the spike protein of the Omicron variant, is extremely unlikely to emerge in an immunocompromised patient [12]. Mutation of virus can have a dN/dS value much higher than unity only when the virus spreads among multiple species [13]. Indeed, even the HIV-1 regulatory gene *tat*, which is known for its high selective pressure, has around 1.5 dN/dS ratio in the human population [14]. In SARS-CoV and SARS-CoV-2, dN/dS is usually smaller than 1 [15].

Wei et al. insist that the Omicron variant has evolved in mice [12], which is followed by Zhang et al. [16]. It is known, however, that the original strain of SARS-CoV-2 does not infect mice [17]. Therefore, it is unlikely that an early strain of SARS-CoV-2 infected from human to mice and back from mice to human under a natural environment. It should also be noted that a strain adapted to an animal with a long incubation period could not infect human better than the variants evolved in human-to-human transmission from the early stage of its emergence. Kakeya et al. indicate that a lab origin of the Omicron variant is likely [18], possibly caused by a spill-over from humanized mice [19]. Arakawa suggests that other variants can also have lab origins considering the consistently high dN/dS ratio [20].

Many lab-leak accidents have happened historically and the number of them has been increasing due to the recent spread of genetic engineering [21,22]. In the end of 2021, a researcher in Taiwan was bitten by a mouse in a biosafety level 3 laboratory and was infected with the Delta variant of SARS-CoV-2, spreading the disease around without noticing [23]. In this case, the incident was confirmed as a lab leak because the virus infection in Taiwan had been subdued due to a strict quarantine policy, which made it easier to identify the researcher as the source of infection. If a lab leak takes place in a city populated with many infected patients, it quite likely remains unnoticed.

After the emergence of No See'm technology [24], apparent traits of artificial genetic modification cannot be found even if a newly detected virus is a product in a laboratory. However, statistical analyses can show that an emergent virus is highly likely a product in a laboratory when the mutation pattern is significantly deviant from the expectation of natural occurrences.

As for the original strain of SARS-CoV-2, deviation from natural occurrence has been discussed based on the phylogenetic trees comparing SARS-CoV and SARS-CoV-2 [15], affinity of the receptor binding domain to the ACE2 receptors of humans compared with those of other animals [25], insertion of unique furin cleavage site at the S1/S2 junction not known in any other sarbecoviruses except for SARS-CoV-2 [26], analyses of pentapeptides presumed to be the principal targets of T-cell non-self recognition [27], and the allocations of BsaI/BsmBI restriction sites [28]. Though a couple of studies supporting the zoonotic origin of SARS-CoV-2 were published from the viewpoints of early mutations and locations at the onset of the outbreak [29,30], allegation of sampling bias was made against both studies in the following publications [31,32].

For fair evaluation on the origin of SARS-CoV-2 and its variants, large scale analyses are needed from various perspectives. In the present study, the authors apply statistical methods from multiple viewpoints to the mutations in the major variants of SARS-CoV-2 to evaluate the possibility of lab leaks during the COVID-19 pandemic.

Methods

We applied the following four analyses to the variants of SARS-CoV-2. For the analyses, the surface glycoprotein data of 22 VOCs were downloaded from NCBI (National Center for Biotechnology Information) GenBank in June 2023. To save computational cost, protein sequences including deletion and insertion were removed from the analyses. The reads were not complete in some of the sequences registered in the database. The missing part was filled with the consensus sequence in the following analyses. The names of the 22 VOCs and the numbers of sequences used for the following analyses are listed in Table 1.

Table 1. The VOCs analyzed in this paper and the count of protein sequences in the data set of each VOC.

B.1.1.7	B.1.351	P.1	B.1.617.2	C.37	B.1.621	BA.1	BA.1.1.18	BA.2	BA.2.3	BA.2.9
181878	3094	17172	34586	1047	3715	50522	43423	107198	30559	15197
BA.2.12.1	BA.4	BA4.1	BA4.6	BA.5	BA.5.1	BA.5.2	BA.5.6	BQ.1	BQ1.1	XBB.1.5
150965	5740	21191	22217	4036	30958	39589	19669	13686	28292	58409

In the first analysis, spectra of 12 kinds of point mutations (four nucleotides to the other three nucleotides) in the spike consensus sequence of Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Lambda (C.37), and Mu (B.1.621) variants from the original Wuhan spike sequence were obtained. The obtained mutation spectra were compared with that of SARS-CoV-2 mutations in humans [33].

In the second analysis, the ratio of reverse mutation G614D and the ratio of mutations to basic amino acids in the spike protein were calculated in the 22 VOCs. Increase of basic amino acids enhances the affinity of spike to cell membrane, which is charged negatively. It is known that D614 is stable in bats for long and in cell cultures for a certain period of time, while G614 is competent in human-to-human transmission [34], which makes extinction of D614 inevitable. Indeed, D614G is known to be the first major mutation observed in the original Wuhan strain [34,35]. Therefore, a surge of G614D reversion means something unnatural is taking place.

In the third analysis, unique mutants comprising only reverse point mutations to the original Wuhan strain and no other point mutations, which we call "pure reversion" in this paper, were counted in each VOC. The distributions in the number of reverse point mutations were compared among the VOCs and mutation patterns of pure reversion were analyzed.

In the fourth analysis, time sequences of emergence of G614D and unique pure reversion were compared with those of total registered sequences in each VOC. Sampled locations of G614D and unique pure reversion in the peak of emergence were extracted and mapped.

Results

Mutation spectra of the spike consensus sequences of Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Lambda (C.37), and Mu (B.1.621) variants, counting mutations from the original Wuhan spike sequence to the spike of each VOC, are compared with that of SARS-CoV-2 mutations in humans [33] in Fig. 1. The results of G-tests show the mutation spectra of Alpha and Delta variants are different from that of human SARS-CoV-2 with statistical significance ($p = 0.048$ and $p = 0.0089$ respectively), while those of Beta, Gamma, Lambda, and Mu variants are not significantly different from that of human SARS-CoV-2 ($p = 0.655$, $p = 0.739$, $p = 0.767$, and $p = 0.101$ respectively).

The counts of G614D reverse mutations divided by the counts of all mutations (ratio A) and those divided by the total numbers of sequences (ratio B) are summarized for the 22 VOCs in Fig. 2. With regards to ratio A, BA.2 has the highest value. This value or higher is attained with the probability of 1.4×10^{-3} under the normal distribution based on the data of 22 VOCs. With regards to ratio B, B.1.167.2 has the highest value. The probability of reaching this value or higher is 7.6×10^{-5} under a normal distribution. Thus, frequency of G614D mutations in B.1.167.2 and BA.2 has a statistically significant difference from the other VOCs.

The counts of mutations to basic amino acids divided by the counts of all mutations in the 22 VOCs are shown in

Fig. 3, where B.1.167.2 has the lowest value. This value or lower is attained with the probability of 4.9×10^{-3} under a normal distribution, meaning that a statistically significant difference exists between B.1.617.2 and the other VOCs.

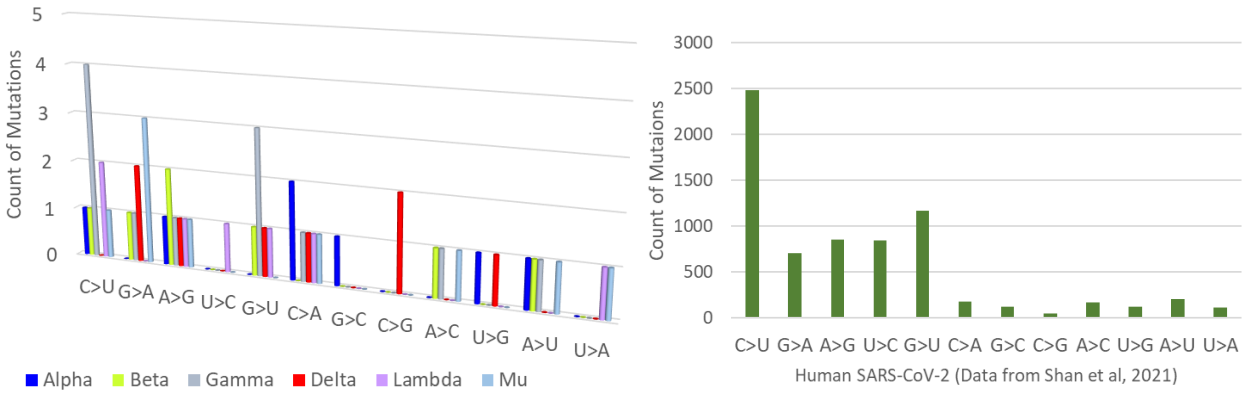


Figure 1. Mutation spectra of Alpha, Beta, Gamma, Delta, Lambda, and Mu variants (left) compared with the mutation spectrum of human SARS-CoV-2 (right) [33].

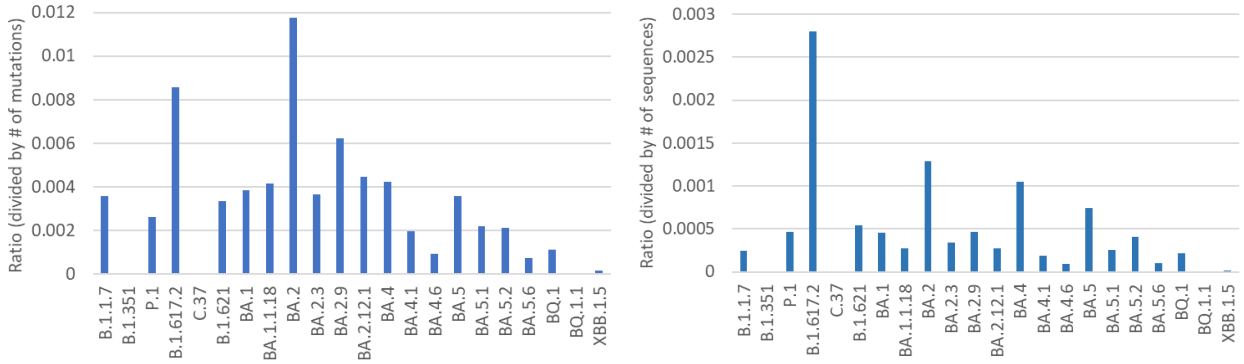


Figure 2. Counts of G614D reverse mutations divided by the counts of all mutations (left) and divided by the total numbers of sequences (right) in 22 SARS-CoV-2 VOCs.

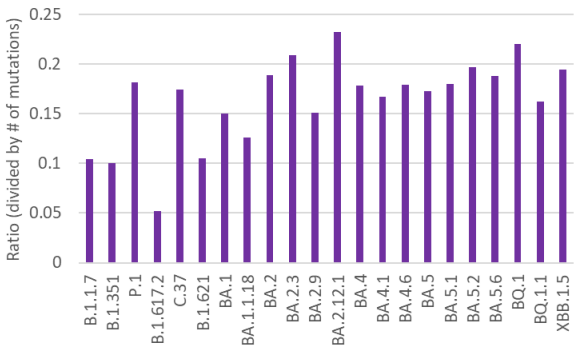


Figure 3. Counts of mutations to basic amino acids divided by the counts of all mutations in 22 SARS-CoV-2 VOCs.

Next, the spike protein sequences that have unique combinations of pure reversion, comprising only reverse point mutations to the original Wuhan strain and no other point mutations, were extracted for each VOC (Fig. 4). Here, the number of reverse point mutations included in each unique pure reversion strain is also surveyed and the counts of unique pure reversion mutants are plotted separately by the number of mutations.

The count of unique pure reversions divided by the total number of sequences has the highest value in BA.1, where the probability given by the normal distribution based on the data of 22 VOCs hitting this value or higher is 4.3×10^{-3}. When compared only among the series of Omicron variants (from BA.1 to XBB.1.5), all of which have similar or more mutations in the spike protein than BA.1 does, the high value of BA.1 is still significant, the probability hitting this value or higher being 6.4×10^{-3}. BA.1 has more unique pure reversion mutants that have many reverse point mutations, which increases the total number of unique pure reversions.

The positions of mutations in the pure reversion mutants of BA.1 that comprise two, three, four, and five point mutations are shown in Fig. 5. As this figure shows, the positions of mutations are scattered all across the spike sequence, where the positions of mutations are often apart from one another, meaning the pattern cannot be generated with a small number of recombination events.

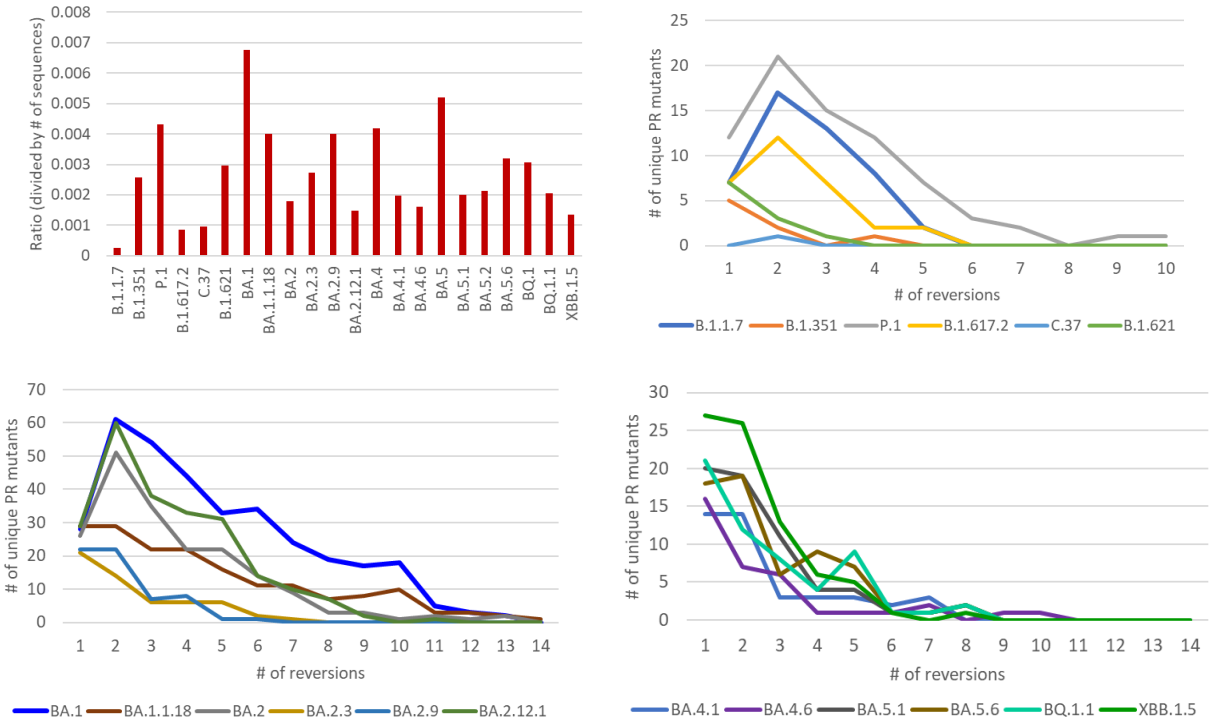


Figure 4. Counts of unique combination of pure reversions (PR) in spike divided by the total numbers of sequences in 22 SARS-CoV-2 VOCs (top-left) and counts of unique pure reversion mutants plotted separately by the number of reverse point mutations (top-right, bottom-left, bottom-right).

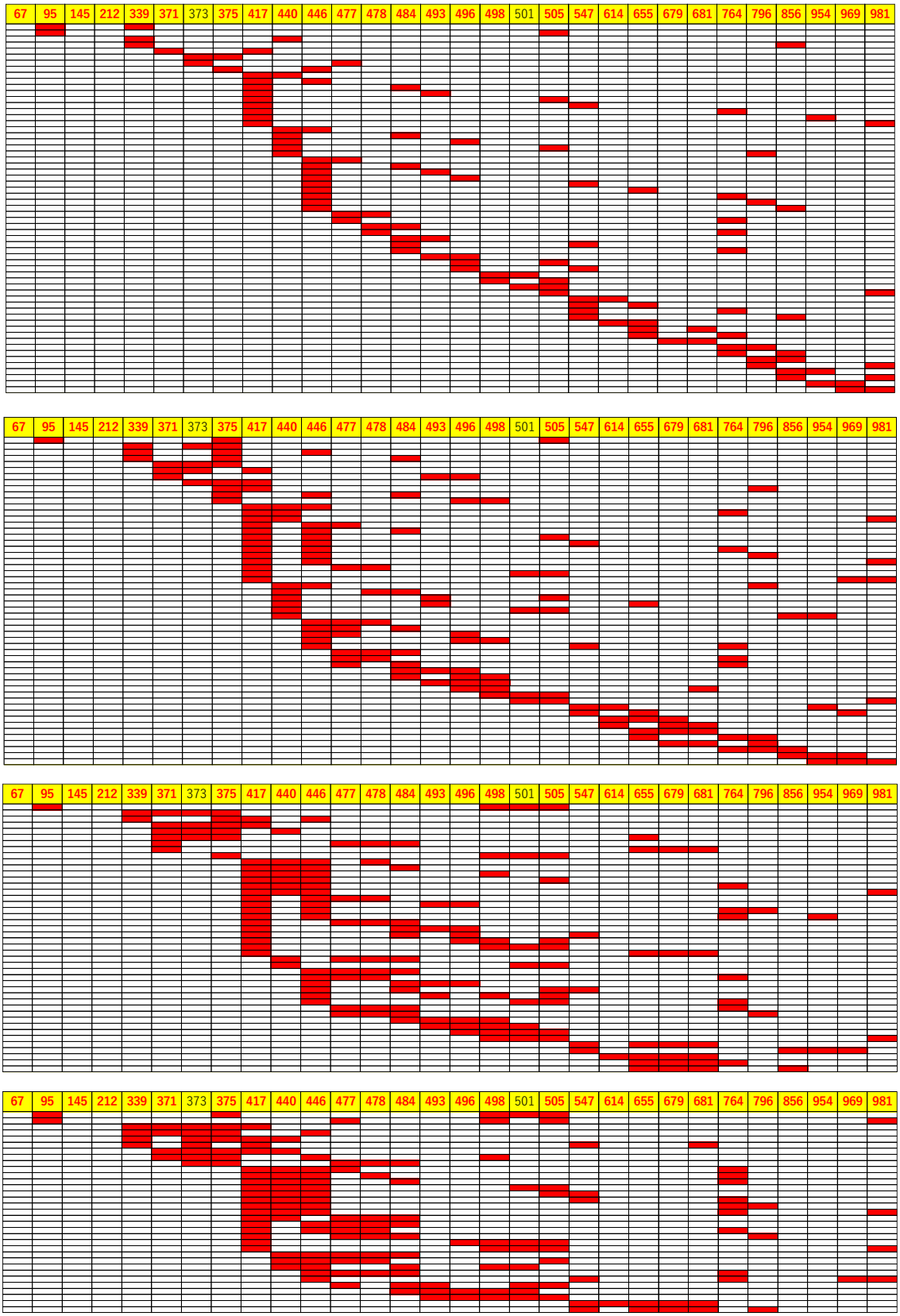
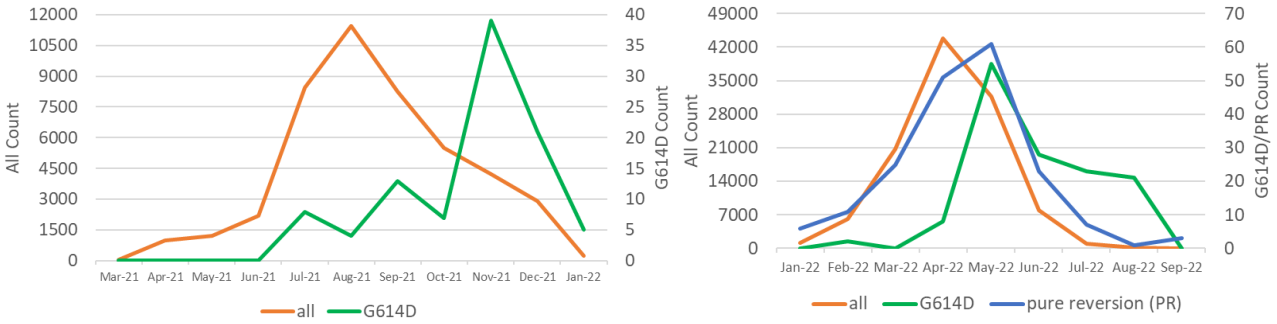


Figure 5. The positions of mutations in the pure reversion mutants of BA.1 that comprise two, three, four, and five point mutations (shown with red rectangles). The yellow row shows the numbers of amino acids in spike, where the existence of single reversion of the amino acid is expressed with bold red numbers.

In Fig.2, G614D reversion is observed in B.1.617.2 and BA.2 more often than other strains with statistical siginificance. Monthly counts of G614D reversions in the first emergence and all sequences in B.1.617.2 and BA.2 are compared in Fig. 6. As for BA.2, monthly counts of unique pure reversions are also compared. The figure shows that the peaks of G614D sampling come after the peaks of whole sampling both in B.1.617.2 and BA.2. The peak of pure reversions in BA.2 comes at the same timing as that of G614D, though the rise of the former precedes that of the latter.

The positions of mutations in the pure reversion mutants of BA.2 comprising two mutations are shown in Fig. 7. As this figure shows, more G614D reversions are observed compared with BA.1, while reversions at the 440th amino acid also stands out.



(a) B.1.617.2 (b) BA.2

Figure 6. Monthly counts of G614D reversions and all sequences in B.1.617.2 (left) and BA.2 (right). As for BA.2, monthly counts of unique pure reversions are also plotted.

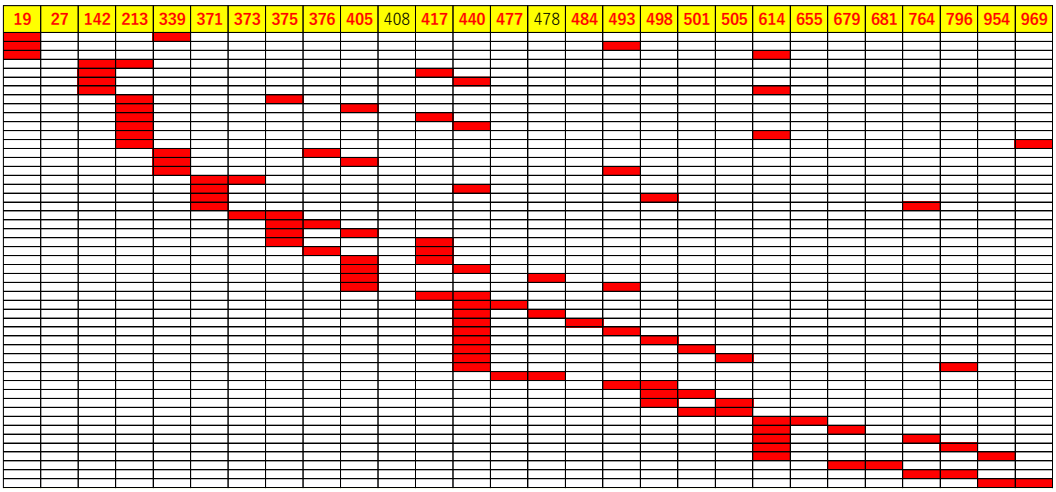
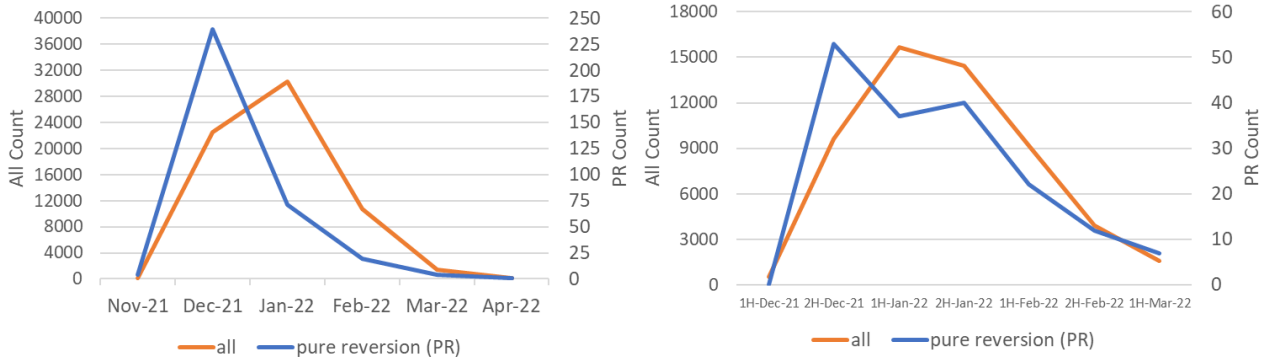


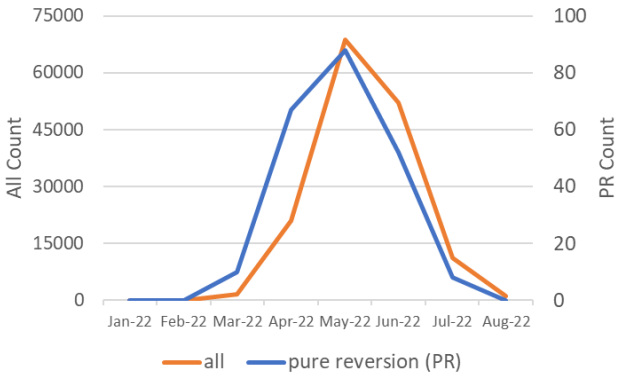
Figure 7. The positions of mutations in the pure reversion mutants of BA.2 that comprise two point mutations (shown with red rectangles).

Monthly counts of unique pure reversions in the first emergence and all sequences in BA.1, BA.1.1.18, BA.2.12.1, BA.5.2, BQ.1.1, and P.1 are compared in Fig. 8. The peak of pure reversion in BA.1 is peculiar, for it precedes the peak of the whole sampling, where the peaks of pure reversion in the other Omicron variants shift rightwards (later) as time passes from the emergence of BA.1. In BQ.1.1, the surge of pure reversion almost matches the surge of whole sampling, which also holds true in P.1.

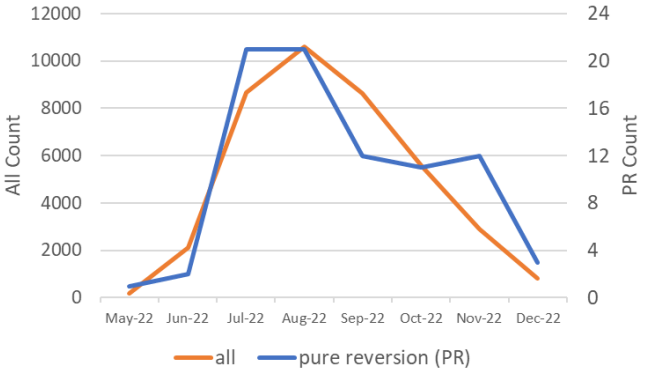


(a) BA. 1

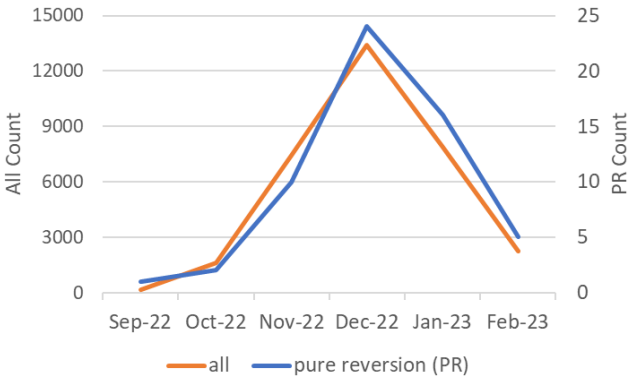
(b) BA.1.1.18



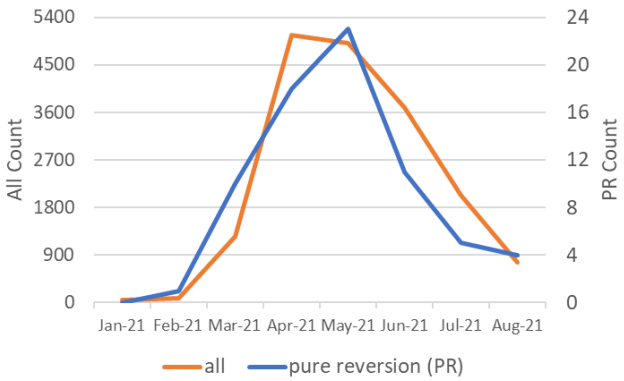
(c) BA. 2.12.1



(d) BA.5.2



(e) BQ.1.1

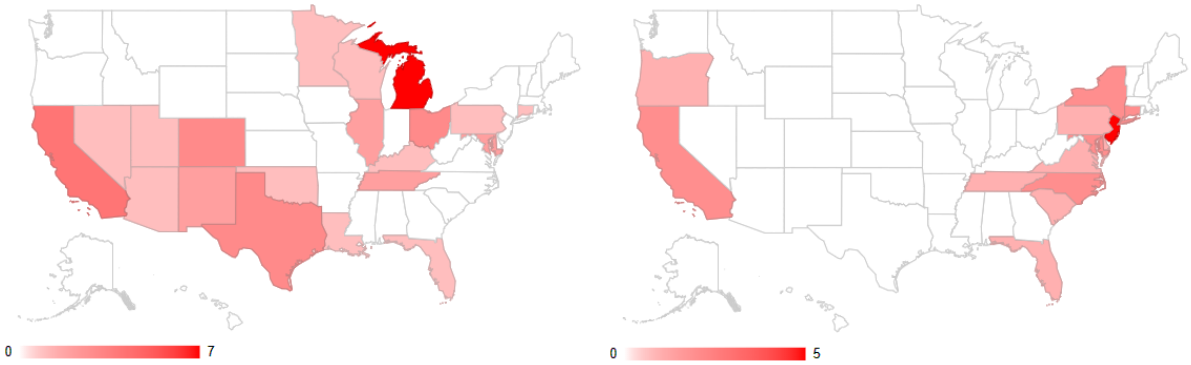


(f) P.1

Figure 8. Monthly counts of unique pure reversions and all sequences in BA.1, BA.1.1.18, BA.2.12.1, BA.5.2, BQ.1.1, and P.1 As for BA.1.1.18, monthly counts are divided into the first half and the second half of the month.

Locations of G614D detection at the peak of its emergence in B.1.617.2 (November 2021) and in BA.2 (May 2022) are shown in Fig. 9. The counts of sampling are shown in the map of the United States, for most of the G614D sequences were sampled in the United States (all 39 in B.1.617.2 and 48 out of 55 in BA.2, where 23 BA.2 data lack registration of state name). The heatmap shows that B.1.617.2 has its epicenters around Michigan, while BA.2 has its epicenters around New Jersey.

Locations where unique pure reversion strains were detected at the peak of its emergence in BA.1 (December 2021) are shown in Fig. 10. Here again, most of pure reversion sequences of BA.1 were sampled in the United States (235 out of 239, where 28 data lack registration of state name) and the heatmap shows the epicenter is around New York and New Jersey.



(a) B.1.617.2 (b) BA.2

Figure 9. Heatmaps of G614D detection at the peaks of its emergence for B.1.617.2 (November 2021) and BA.2 (May 2022).

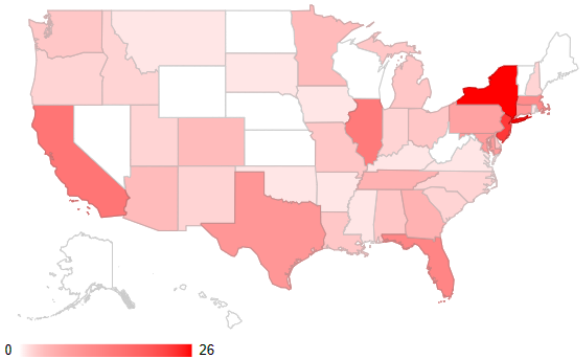


Figure 10. Heatmap expressing detection of BA.1 unique pure reversion at the peak of its emergence (December 2021).

Discussion

From the mutation spectra (Fig. 1), Alpha and Delta variants are deviant from the normal human mutation, which means that they are not likely the products of incubation in humans. While a previous study speculates that Alpha variant may have originated from an immunocompromised patient [2], it bears the spike mutation N501Y, which enables SARS-CoV-2 to infect mice and is widely used for experimental purposes in laboratories to test the

effectiveness of medicines and vaccines against SARS-CoV-2 [36]. Though N501Y is also known to enhance ability to bind to human ACE2 receptors [37], which can be selected positively in humans, a possibility that a mutant grown to infect mice for experimental purposes escaped from a laboratory is worth investigating considering the mutation spectrum obtained in this study.

The mutation spectrum of Delta variant is farther away from the normal human mutation ($p = 8.9 \times 10^{-3}$). Delta variant is notable with four mutations to Arginine, the strongest basic amino acid, out of eight point mutations in its spike protein. One of the four mutations is P681R in the furin cleavage site, which is regarded as the key mutation that enhances infectivity and virulence of Delta variant [38]. Considering these features as well as the anomaly of the mutation spectrum, possibility of lab origin is considerably high for the Delta variant.

Delta is also irregular in its significantly low rate of mutations to basic amino acids from the consensus Delta sequence, as shown in Fig. 3. One possible reason is that Delta variant is under selective pressure not to bear more basic amino acids due to the four prior mutations to Arginine. Subsequent dominant VOCs are significant in their large numbers of point mutations to basic amino acids, Omicron BA.1 with 14 out of 30 and BA.2 with 12 out of 28. Mutations to basic amino acids are considered to be under positive selective pressure, for a positively charged spike increases affinity to cell membrane, which is charged negatively. Inability to include more basic amino acids in the spike protein may be one of the reasons why Delta was not successful in generating dominant progenies.

Figure 2 shows that Delta and Omicron BA.2 have notably more numbers of mutants including G614D. As Fig. 6 shows, these mutants emerged after the surge of each VOC. Since D614 is not stable in vivo, not only in humans but also in hamsters [39], emergence of this reverse mutation in the non-early phase of the prevalence is extremely strange. One possible scenario is that the original VOC was first sampled from an early patient, which was kept in cell cultures for research, where D614G reverse mutation was obtained possibly through recombination with lab-kept bat sarbecovirus or wild-type SARS-CoV-2, and escaped from the laboratory by accident to spread into human populations.

Figure 4 shows the significantly large number of unique pure reversions in Omicron BA.1, which can be a strong counterevidence against the immunocompromised patient origin together with the anomaly of mutation spectrum of Omicron BA.1 given by the previous studies [12,19].

Emergence of Omicron through strong selective pressure in an unknown human population is also extremely unlikely considering that the variation of unique pure reversion decreases as time goes by, as shown in Fig. 8. It is true that the series of Omicron variants have emerged to escape from the immune system responding the precedented strains [40], we do not observe in those variants the variations that BA.1 had in the beginning of its prevalence. It is also noteworthy that Omicron variant BA.1 has 30 point mutations in the spike protein, while XBB.1.15 has 38. Gamma had 12 point mutations in spike, which was the largest number before Omicron. Sudden increase in number of point mutations as many as 18 is not comparable to the gradual increase of eight mutations

during the 1.5 years of Omicron dominance.

Variations of unique pure reversions requiring many times of repeated recombination, as shown in Fig. 5, indicate existence of certain non-natural process to enhance its generation, such as serial passages in a laboratory or manipulations of genomic sequences, as shown in Fig. 11. Though recombination often happens in nature, the rises of pure reversions coincide with those of sampled numbers in Gamma variant, which is independent from the precedented VOCs, and in the late versions of Omicron variant (Fig. 8), suggesting the sudden rise of pure reversions preceding the prevalence of the variant cannot be explained by natural recombination.

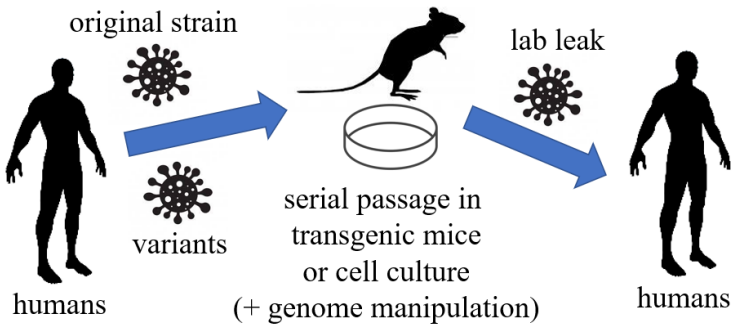


Figure 11. Possible scenarios of mutation and infection.

From the discussion above, it is plausible that lab leaks like the case in Taiwan in late 2021 have happened multiple times, likely at the onset of Delta and Omicron BA.1, amid the prevalence of Delta and Omicron BA.2, and possibly at the onset of Alpha variant, while they have remained unnoticed since the virus was widespread among human populations and could not be distinguished from the community transmission.

The limitation of this study is that the analyses are all based on the data registered in GenBank, where possibilities of data errors, including common bias intrinsic to RT-PCR or DNA sequencing, are not taken into account. Some may claim that the anomalies found in this paper are all caused by data errors. Most of the anomalies observed above, however, cannot be caused by random errors in data registered by various submitters, while a common bias that runs across submitters cannot cause variation of unique reversions, which is why possibilities of lab leak should be taken seriously.

Many researchers have suspected that the original strain of SARS-CoV-2 may have leaked from the Wuhan Institute of Virology (WIV) [41-48]. Even the authors of the paper that wrote “Our analyses clearly show that SARS-CoV-2 is not a laboratory construct or a purposefully manipulated virus” [49] raised the possibility of lab origin repeatedly in their private communications both before and after submission of the paper [50,51].

As for the variants, it is reported that most of the mutations in the spike protein of Omicron are known to affect infectivity through previous variants or past experiments, and the introduction of these mutations to the virus can be used to develop pan-variant vaccine [52]. Indeed, quite a large number of laboratories have kept SARS-CoV-2

for experimental purposes, often accompanying directed mutations, since the beginning of the COVID-19 pandemic. It should be noted that some of the experiments manipulating the genome of SARS-CoV-2 and its variants are extremely risky. One study reports that a chimeric virus with the Omicron spike spliced into the backbone of the original Wuhan strain regains lethality lost by the Omicron variant [53]. Should this synthetic virus be leaked from a laboratory, it could claim millions of lives with its lethality and immune-escape ability combined.

Historically, lab leaks have been covered up repeatedly in the field of microbiology [54]. A typical example is the Sverdlovsk anthrax leak in 1979 [55], which took 15 years to be accepted officially as a lab-leak event. Also, it took about 30 years to reach a consensus among virologists that the 1977 Russian influenza H1N1 originated from a frozen virus in a laboratory [56]. Recently, the swine flu outbreak in 2009 [57] and the Ebola outbreak in West Africa in 2014 [58] are also suspected to have research related origins, which has not been investigated enough because of the opacity of the related laboratories.

To prevent the next pandemic, the origin of the virus needs to be unraveled [59], which is attainable only through transparent, objective, and data-driven investigations [60]. To pursue what has really happened in laboratories during the pandemic, thorough investigation should be carried out by an independent organization without any conflict of interest. An agency like the IAEA (International Atomic Energy Agency) for atomic engineering, being free from reputation risk even when they find something wrong, should be established for the oversight of virology research dealing and tinkering with deadly viruses. Unfortunately, some universities have declined to share data on the research activities related to SARS-CoV-2 in their virology laboratories upon FOIA (Freedom of Information Act) requests.

It is true that all the anomalies shown in this paper are just statistical biases, not the definitive proofs of lab leaks. Therefore, laboratories of concerns should be granted the benefit of the doubt. In the case of forensic investigation, however, prosecutors are allowed to raid the suspect when an investigation warrant is issued by the judiciary. A compulsory investigation is admitted when circumstantial evidence strongly supports the prosecutor's claim. In the same manner, compulsory investigations into laboratories of concerns should be executed by an independent oversight agency when the likelihood of natural emergence is below a certain threshold. Probabilistic and chronological evaluation as shown in this paper can work as a prerequisite assessment tool to warrant such investigations.

Conflicts of interest

The authors declare no conflict of interests exist.

Data availability statement

The data are available from the corresponding author upon reasonable request.

Author contributions

HK carried out the whole data analysis. YM supervised the work from the aspect of molecular biology and genetics.

Acknowledgement

The authors thank Dr. Hiroshi Tauchi and Dr. Takeshi Nitta for comments and suggestions on the manuscript.

References

[1] Callaway E. Heavily mutated Omicron variant puts scientists on alert. Nature 2021;600(7887):21. DOI: 10.1038/d41586-021-03552-w

[2] Hill V, Plessis LD, Peacock TP, et al. The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. Virus Evolution 2022;8(2):1-13. DOI: 10.1093/ve/veac080

[3] Hassan SS, Kodakandla V, Redwan EM, et al. Non-uniform aspects of the SARS-CoV-2 intraspecies evolution reopen question of its origin. International Journal of Biological Macromolecules 2022; 222:972-993. DOI: 10.1016/j.ijbiomac.2022.09.184

[4] Kupferschmidt K. Where did 'weird' Omicron come from? Science 2021;374(6572):1179. DOI: 10.1126/science.acx9738

[5] Mallapaty C. The hunt for the origin of Omicron. Nature 2022;602(7898):26-28. DOI: 10.1038/d41586-022-00215-2

[6] Jung C, Kmiec, D, Koepke L, et al. Omicron: what makes the latest SARS-CoV-2 variant of concern so concerning? J Virology 2022;96(6): e02077-21. DOI: 10.1128/jvi.02077-21

[7] Choi B, Choudhary MC, Regan J, et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. The New England Journal of Medicine 2020;383(23):2291-2293. DOI: 10.1056/NEJMc2031364

[8] Kemp SA, Collier DA, Datier RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature 2021;592:277-282. DOI: 10.1038/s41586-021-03291-y

[9] Truong TT, Ryutov A, Pandey U, et al. Increased viral variants in children and young adults with impaired humoral immunity and persistent SARS-CoV-2 infection: A consecutive case series. EBioMedicine 2021 May;67:103355. DOI: 10.1016/j.ebiom.2021.103355

[10] Miyata T, Yasunaga T. Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Molecular Evolution 1980;16(1):23–36. DOI: 10.1007/BF01732067

[11] Li WH, Wu CI, Luo CC. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Molecular Biology and Evolution 1985;2(2):150–174.
DOI: 10.1093/oxfordjournals.molbev.a040343

[12] Wei C, Shan KJ, Wang W, et al. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. J Genet Genomics 2021;48(12):1111-1121.
DOI: 10.1016/j.jgg.2021.12.003

[13] Kryazhimskiy S, Plotkin JB, The population genetics of dN/dS. PLOS Genetics 2008;4(12):e1000304.
DOI: 10.1371/journal.pgen.1000304

[14] Hasan Z, Hasan M, Ashik AI, et al. Prediction of immune pressure on HIV-1 regulatory gene tat by human host through bioinformatics tools. J Adv Biotechnol Exp Ther. 2020 Sep;3(3):233-240.
DOI: 10.5455/jabet.2020.d129

[15] Zhan SH, Deverman BE, Chan YA. SARS-CoV-2 is well adapted for humans. What does this mean for re-emergence? bioRxiv 2020.
DOI: 10.1101/2020.05.01.073262

[16] Zhang W, Shi K, Geng Q, et al. Structural basis for mouse receptor recognition by SARS-CoV-2 omicron variant. PNAS 2022; 119(44): e2206509119.
DOI: 10.1073/pnas.2206509119

[17] Piplani S, Singh PK, Winkler DA, et al. In silico comparison of SARS-CoV-2 spike protein-ACE2 binding affinities across species and implications for virus origin. Science Report 2021;11:13063.
DOI: 10.1038/s41598-021-92388-5

[18] Kakeya H, Matsumoto Y. A probabilistic approach to evaluate the likelihood of artificial genetic modification and its application to SARS-CoV-2 Omicron variant. ISPJ Trans. Bioinformatics 2022;15:22-29.
DOI: 10.2197/ipsjtbio.15.22

[19] Kakeya H, Arakawa H, Matsumoto Y. Multiple probabilistic analyses suggest non-natural origin of SARS-CoV-2 Omicron variant. Zenodo 2023.
DOI: 10.5281/zenodo.7470652

[20] Arakawa H. Mutation signature of SARS-CoV-2 variants raises questions to their natural origins. Zenodo 2022.
DOI: 10.5281/zenodo.6601991

[21] Butler D. Fears grow over lab-bred flu. Nature 2011;480(7378):421–422.
DOI: 10.1038/480421a

[22] Biosafety in the balance. Nature 2014;510(7506):443.
DOI: 10.1038/510443a

[23] Silver A. Taiwan’s science academy fined for biosafety lapses after lab worker contracts COVID-19. Science 2022.
DOI: 10.1126/science.ada0525

[24] Yount B, Denison MR, Weiss SR, et al. Systematic assembly of a full-length Infectious cDNA of mouse hepatitis virus strain A59. J Virology 2002;76(21):11065–11078.
DOI: 10.1128/JVI.76.21.11065-11078.2002

[25] Piplani S, Singh PK, Winkler DA, et al. In silico comparison of SARS-CoV-2 spike protein-ACE2 binding affinities across species and implications for virus origin. Science Report 2021;11:13063.
DOI: 10.1038/s41598-021-92388-5

[26] Chan AY, Zhan SH. The emergence of the spike furin cleavage site in SARS-CoV-2. Mol. Biol. Evol. 2022;39(1):msab327
DOI: 10.1093/molbev/msab327

[27] Flegr J, Zahradník D, Zemková M. Thus spoke peptides: SARS-CoV-2 spike gene evolved in humans and then shortly in rats while the rest of its genome in horseshoe bats and then in treeshrews. Commun Integr Biol. 2022; 5(1):96–104.
DOI: 10.1080/19420889.2022.2057010

[28] Bruttel V, Washburne A, VanDongen A, Endonuclease fingerprint indicates a synthetic origin of SARS-CoV-2. bioRxiv 2022.
DOI: 10.1101/2022.10.18.512756

[29] Pekar JE, Magee A, Parker E, et al. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science 2022; 377(6609):960-966.
DOI:10.1126/science.abp8337

[30] Worobey M, Levy J, Serrano LM, et al. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. Science 2022;377(6609):951-959.
DOI:10.1126/science.abp8715

[31] Massey SE, Jones A, Zhang D, Deigin Y, Quay SC. Unwarranted exclusion of intermediate lineage A-B SARS-CoV-2 genomes is inconsistent with the two-spillover hypothesis of the origin of COVID-19. Microbiology Research 2023; 14(1):448-453.
DOI: 10.3390/microbiolres14010033

[32] Daoyu Z, Gilles D, Adrian J, et al. Zoonosis at the Huanan Seafood Market: a critique. Zenodo 2022.
DOI: 10.5281/zenodo.7169296

[33] Shan KJ, Wei C, Wang Y, et al. Host-specific asymmetric accumulation of mutation types reveals that the origin of SARS-CoV-2 is consistent with a natural process. Innovation 2021;2(4), 100159.
DOI: 10.1016/j.xinn.2021.100159

[34] Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 2020; 182(4), 812-827.
DOI: 10.1016/j.cell.2020.06.043

[35] Volz E, Hill V, McCrone JT, et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity, Cell 2021; 184(1), 64-75.
DOI: 10.1016/j.cell.2020.11.020

[36] Gu H, Chen Q, Yang G, et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy, Science 2020;369(6511), 1603-1607.
DOI: 10.1126/science.abc4730

[37] Starr TN, Greaney AJ, Hilton SK, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell 2020;182(5), 1295-1310.
DOI: 10.1016/j.cell.2020.08.012

[38] Liu Y, Liu J, Johnson BA, et al. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. Cell Rep 2022;39(7):110829.
DOI: 10.1016/j.celrep.2022.110829

[39] Hou YJ, Chiba S, Halfmann P, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo, Science 2020;370(6523):1464-1468
DOI: 10.1126/science.abe8499

[40] Wilks SH, Mühlemannet B, Shen X, et al. Mapping SARS-CoV-2 antigenic relationships and serological responses. bioRxiv 2023.
DOI: 10.1101/2022.01.28.477987

[41] Sallard E, Halloy J, Casane D, et al. Tracing the origins of SARS-COV-2 in coronavirus phylogenies: a review. Environ Chem Lett 2021;19(4), 769-785.
DOI: 10.1007/s10311-020-01151-1

[42] Segreto R, Deigin Y, McCairn K, et al. Should we discount the laboratory origin of COVID-19? Environ Chem Lett 2021;Mar 15, 1-15
DOI: 10.1007/s10311-021-01211-0

[43] Quay SC. A Bayesian analysis concludes beyond a reasonable doubt that SARS-CoV-2 is not a natural zoonosis but instead is laboratory derived, Zenodo 2021.
DOI: 10.5281/zenodo.4477081

[44] Wiesendanger R. Studie zum Ursprung der Coronavirus-Pandemie. ResearchGate 2021.
DOI:10.13140/RG.2.2.31754.80323

[45] Markson S. What really happened in Wuhan. HarperCollins 2021.
ISBN-13: 978-1460761083

[46] Chan AJ and Ridley M. Viral: the search for the origin of Covid-19. Fourth Estate 2021.
ISBN-13: 978-0008487492

[47] Harrison NL, Sachs JD. A call for an independent inquiry into the origin of the SARS-CoV-2 virus, PNAS 2022;119(21):e2202769119
DOI: 10.1073/pnas.2202769119

[48] Coccia M. Meta-analysis to explain unknown causes of the origins of SARS-CoV-2, Environmental Research 2022; 211:113062
DOI: 10.1016/j.envres.2022.113062

[49] Andersen KG, Rambaut A, Lipkin WI, et al. The proximal origin of SARS-CoV-2. Nature Medicine 2020;26(4):450-452.
DOI: 10.1038/s41591-020-0820-9

[50] Leopold J. NIH FOIA Anthony Fauci emails.
https://www.documentcloud.org/documents/20793561-leopold-nih-foia-anthony-fauci-emails [cited July 24, 2023]

[51] Selective Subcommittee on the Coronavirus Pandemic. The proximal origin of the cover-up: Did the "Bethesda boys" downplay a lab leak? Interim Major Staff Report 2023.
https://oversight.house.gov/wp-content/uploads/2023/07/Final-Report-6.pdf [cited July 24, 2023]

[52] Bruttel V, Molecular evidence indicates a synthetic origin of SARS-CoV2 Omicron,
https://www.stopgof.com/english/omicron-origin/ [cited July 24, 2023]

[53] Chen DY, Kenney D, Chin CV, et al. Role of spike in the pathogenic and antigenic behavior of SARS-CoV-2 BA.1 Omicron, bioRxiv 2020.
DOI: 10.1101/2022.10.13.512134

[54] Young A. Pandora's gamble: lab leaks, pandemics, and a world at risk. Center Street 2023.
ISBN-13: 978-1546002932

[55] Meselson M, Guillemin J, Hugh-Jones M. The Sverdlovsk anthrax outbreak of 1979. Science 1994;266(5188):1202-1208.
DOI: 10.1126/science.7973702

[56] Kransnitz M, Levine AJ, Rabadan R. Anomalies in the influenza virus genome database: new biology or laboratory errors? J Virol. 2008 Sep;82(17):8947-50.
DOI: 10.1128/JVI.00101-08.

[57] Gibbs AJ, Armstrong JS, Downie JC. From where did the 2009 'swine-origin' influenza A virus (H1N1) emerge? Virology Journal 2009;6:207.
DOI: 10.1186/1743-422X-6-207

[58] Latham J. Did West Africa's Ebola outbreak of 2014 have a lab origin?
https://jonathanlatham.net/did-west-africas-ebola-outbreak-of-2014-have-a-lab-origin/ [cited July 24, 2023]

[59] Relman DA. To stop the next pandemic, we need to unravel the origins of COVID-19. PNAS 2020;117 (47):29246-29248.
DOI: 10.1073/pnas.2021133117

[60] Bloom JD, Chan YA, Baric RS, et al. Investigate the origins of COVID-19, Science 2021; 372(6543):694.
DOI: 10.1126/science.abj0016