Unnatural evolutionary processes of SARS-CoV-2 variants and possibility of deliberate natural selection

1   **Atsushi Tanaka[1,2,*] †, Takayuki Miyazawa[2,3,*] †**

2   [1]Division of Research Animal Laboratory and Translational Medicine, Research and Development Center Osaka
3   Medical and Pharmaceutical University, Takatsuki, Osaka 569-8686, Japan

4   [2]Laboratory of Virus-Host Coevolution, Institute for Life and Medical Sciences, Kyoto University, Sakyo-ku,
5   Kyoto 606-8507, Japan

6   [3)] Resilience Research Unit, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8540, Japan

7
8   *** Correspondence:**
9   Atsushi Tanaka*****
10   atsushi.tanaka@ompu.ac.jp
11
12   Takayuki Miyazawa*****
13   takavet@infront.kyoto-u.ac.jp
14

16   **Abstract**

17   Over the past three years, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has repeatedly
18   caused pandemics, generating various mutated variants ranging from Alpha to Omicron. In this study, we
19   aimed to clarify the evolutionary processes leading to the formation of SARS-CoV-2 Omicron variants,
20   focusing on Omicron variants with many amino acid mutations in the spike protein among SARS-CoV-2
21   isolates. To determine the order of mutations leading to the formation of the SARS-CoV-2 Omicron variants,
22   we compared the sequences of 129 Omicron BA.1-related, 141 BA.1.1-related, and 122 BA.2-related isolates,
23   and attempted to clarify the evolutionary processes of SARS-CoV-2 Omicron variants, including the order of
24   mutations leading to their formation and the occurrence of homologous recombination. As a result, we
25   concluded that the formation of a part of Omicron isolates BA.1, BA.1.1, and BA.2 was not the product of
26   genome evolution, as is commonly observed in nature, such as the accumulation of mutations and homologous
27   recombinations. Furthermore, the study of 35 recombinant isolates of Omicron variants BA.1 and BA.2
28   confirmed that Omicron variants were already present in 2020. The analysis showed that Omicron variants
29   were formed by an entirely new mechanism that cannot be explained by previous biology, and knowing how
30   the SARS-CoV-2 variants were formed prompts a reconsideration of the SARS-CoV-2 pandemic.

31   **1   Introduction**
32
33   COVID-19, the coronavirus disease 2019, caused by severe acute respiratory syndrome coronavirus 2 (SARS-
34   CoV-2), was first reported in December 2019 in Wuhan, China (1). This emerging infectious disease was
35   unprecedentedly fast, spreading worldwide, leading the World Health Organization (WHO) to declare a global
36   pandemic of COVID-19 on March 11, 2020 (https://www.who.int/). SARS-CoV-2, belonging to
37   betacoronavirus subgroup B, has a single-stranded positive-sense RNA genome that codes for ten genes,
38   ultimately producing 26 proteins according to an annotation of NCBI Reference Sequence: NC_045512.2. Its
39   genome size varies from 29.8 to 29.9 kb. It consists of four structural proteins: spike (S), envelope (E),
40   membrane (M), and nucleocapsid (N) proteins (2, 3). In the structural proteins, the S protein as the surface
41   glycoprotein is the largest protein, being approximately 180 kDa, and consisting of two subunits, S1 and S2. It
42   mediates membrane fusion and ultimately facilitates virus entry. The receptor-binding domain (RBD) (amino

43 acid residues 319‑541) of the S1 subunit interacts with angiotensin‑converting enzyme 2 (ACE2), binding to
44 its peptidase domain (4, 5).
45
46 Over the three years from 2019 to 2022, SARS-CoV-2 was re-accelerated by new variants that emerged over
47 several months in various geographical regions and disseminated throughout the world, to induce the pandemic
48 repeatedly.
49
50 In the early stage of the first pandemic, the most impactful mutation of SARS-CoV-2 was the non-synonymous
51 mutation D614G in the S protein. This mutation, which was not present in the ancestral lineage that caused the
52 Wuhan outbreak, quickly became dominant worldwide (6). Soon after, the variant of concern, B.1.1.7 : 20I
53 (Alpha, V1), the lineage B.1.1.7 (clade 501.YV1), or Alpha, characterized by 17 unique mutations containing
54 ten amino acid differences in the S protein, was first detected in southeastern England in late September 2020
55 (7) and expanded rapidly across the United Kingdom to become predominant during early 2021, spreading to
56 most European countries with similar success. By November 2021, local transmission of this lineage had been
57 reported in 175 countries (8). Shortly after, the emergence of variant strains of SARS-CoV-2 Alpha, variants
58 B.1.351 : 20H (Beta, V2), was identified in October 2020, which was first detected in the Eastern Cape province
59 of South Africa from specimens collected in early August. This Beta variant spread within South Africa and was
60 considered to have displaced the other SARS-CoV-2 lineages circulating there (9). Then, the variant P.1: 20J
61 (Gamma, V3) was identified in Brazil in December 2020, thought to have evolved in Brazil. Health officials in
62 Japan first reported it publicly on January 10, 2021, after identifying the Gamma variant in four Brazilian
63 travelers at Haneda Airport in Tokyo, Japan (10).
64 At about the same time, the Delta variant (Pango lineage designation B.1.617.2), which was first detected in
65 India in February 2021, and the Mu variant, also known as lineage B.1.621 first detected in Colombia in January
66 2021, were reported (11, 12). While the lambda variant (Pango lineage designation C.37), was detected in Peru
67 in August 2020, but designated in June 15, 2021 by WHO (13, 14).
68
69 Almost one year later, regarding these emergences of variants of concern, Omicron (phylogenetic assignment
70 of named global outbreak (Pango) lineage designation B.1.1.529; BA.1, Nextstrain clade 21K) was a variant of
71 SARS-CoV-2 first reported to WHO by the Network for Genomics Surveillance in South Africa on November
72 24, 2021 (15, 16) with more than 50 amino acids changes when compared with the first reported strain Wuhan-
73 Hu-H1 (NCBI Reference Sequence: NC_045512.2.), and 39 of these amino acids difference were observed in
74 the S protein. This variant was first detected in Botswana and became the predominant circulating variant
75 worldwide (17).

76 In the United States, the San Francisco Department of Public Health confirmed that a case of COVID-19 among
77 individuals in California was caused by Omicron variant BA.1, carried by a traveler who returned from South
78 Africa on November 22, 2021 (https://www.cdc.gov/media/releases/2021/s1201-Omicron-variant.html). Then,
79 the first Omicron sub-lineage BA.1 expanded rapidly and replaced the Delta variant (18).

80 Less than two weeks later, the Omicron variant BA.1, the new Omicron variant, BA.2 lineage, showing 31
81 amino acids changes in the S protein when compared with the Wuhan-Hu-H1, was initially identified in
82 Denmark on December 5, 2021 (19).   On February 22, 2022, WHO mentioned the Omicron sublineage BA.2
83 (https://www.who.int/news/item/22-02-2022-statement-on-Omicron-sublineage-ba.2), whereby the Omicron
84 variant of concern was currently the dominant variant circulating globally, replacing the Delta variant (Pango
85 lineage designation B.1.617.2) (https://www.who.int/docs/default-source/coronaviruse/2022-01-07-global-
86 technical-brief-and-priority-action-on-Omicron---corr2.pdf?sfvrsn=918b09d_20), accounting for nearly all
87 sequences reported to GISAID. Then, as of March 16, 2023, WHO stated that the Omicron variants accounted
88 for over 98% of the publicly available viral sequences after February 2022 (https://www.who.int/news/item/16-
89 03-2023-statement-on-the-update-of-who-s-working-definitions-and-tracking-system-for-SARS-CoV-2-
90 variants-of-concern-and-variants-of-interest).

91 Omicron variants BA.1 and BA.2 were suggested to have expanded and diverged around October to December
92 2021, respectively. These mutants were estimated to have diverged from a common ancestor around February
93 to March 2021 (20). Since Omicron variants BA.1 and BA.2 share a common 14-amino acid mutation in the S

94 protein, the common ancestor of Omicron variants BA.1 and BA.2 may have already acquired the 14-amino
95 acid mutation in the S protein region around February to March 2021; however, no common ancestral strain has
96 been found in the international databases, and the strains may have acquired their mutations through different
97 pathways.

98 In this study, we attempted to clarify the evolutionary processes of the Omicron variant, which has two-times
99 more amino acid mutations in the S protein than other variants, by examining the rank order of introduced amino
100 acid mutations in the S protein.

101 **2   Results**

102 Each variant is considered to have arisen through an independent evolutionary pathway from isolates with the
103 D614G mutation in the S protein. Concerning the genetic variation in the S protein of these variants, most of the
104 mutations were non-synonymous (Fig. 1). There were no synonymous mutations in the Alpha, Beta, Gamma,
105 Delta, or Mu variants, but only one each in the Lambda and Omicron variants. Among these variants, the
106 Omicron variant (BA.1 lineage), which shows the greatest accumulation of mutations in the S protein, is
107 primarily non-synonymous in the S protein and has only one synonymous mutation, at c25000u. The
108 synonymous/non-synonymous ratio is abnormal, considering how human coronaviruses have mutated (See
109 Supplemental Figure 1).

110 At first, we considered the existence of the isolate of SARS-CoV-2, whose amino acid sequence in the S protein
111 contains the Omicron-BA.1-type mutation subsets, but one mutation position was not mutated and comprised
112 the original Wuhan-type amino acid sequence. We designated these isolates as BA.1-01. Each amino acid
113 sequence of the S protein region was named BA.1-0.1_S: amino acids of the Omicron-BA.1 type (Oaa) and
114 Wuhan type (Waa) and its position number (XXX) (Ex., BA.1-0.1_S:OaaXXXWaa), as described in Methods.
115 Then, the putative isolates bearing BA.1-0.1_S:OaaXXXWaa were searched for using the BLAST program
116 based on their amino acid sequences. In this search, we obtained the isolates whose identities showed 100%
117 matches   with   the   query   amino   acid   sequence   except   for   SARS-CoV-2_human_USA_NY-
118 PV55373_2022(GenBank: ON246090.1), whose identity was 99.92%.

119 Surprisingly, we found that Omicron BA.1-0.1 isolates were detected at all mutation sites except N501Y (Fig.
120 2A). In the BA.1 lineage of the Omicron variant, there are Omicron isolates (BA.1.1) with the R346K mutation
121 seen in the Mu(m) variant (termed B.1.621), *i.e.*, BA.1_S can be defined as BA.1.1_S:K346R. We also
122 performed a BLAST search for isolates with amino acid sequences of BA.1-0.1.1_S:OaaXXXWaa, as described
123 in Methods. As a result, Omicron BA.1.1-subset-0.1 isolates were detected at all mutation sites except S373P
124 (Fig. 2B). Similar to the BA.1 lineage of the Omicron variant, in the BA.2 lineage of the Omicron variant,
125 isolates of BA.2-0.1 were found at all mutant sites except T478K and P681H in the S protein (Supplemental
126 Figure 2). The presence of these isolates refutes the establishment of Omicron strains through a continuous
127 evolutionary process by accumulating mutations. So, we could not determine which mutation occurred first or
128 last to form the Omicron variants. As shown in Fig. 2B, over half of the BA.1.1-0.1 isolates have the synonymous
129 mutation c21595u detected in the S protein. However, this does not help explain the order in which the c21595u
130 mutation arose. Curiously, in BA.1 strain isolates, this c21595u mutation was only detected in SARS-CoV-
131 2_human_USA_ID-CDC-LC0481844_2022 (GenBank: OM409228.1) and SARS-CoV-2_human_USA_MI-
132 CDC- ASC210597972_2022 (GenBank: OM396816.1). These isolates commonly lack the G339D mutation.
133 This synonymous mutation is in a mutation-prone hotspot location. Still, if the same mutation occurred
134 independently in different isolates, it is highly unnatural for the proportion of c21595u occurrences to be
135 significantly biased in the Omicron variants BA.1.1-0.1.

136 It has been reported that two different variants were infected in a single cell while establishing various SARS-
137 CoV-2 variants, causing homologous recombination in the process of viral RNA synthesis, resulting in multiple
138 variants. On considering that homologous recombination caused the isolates shown in Fig. 2, some of the
139 breakpoints at which strand changes occur by homologous recombination are too short (1nt, 2nt, 3nt, etc.) (Fig.
140 3 and Supplemental Figure 3). Therefore, it is unreasonable to employ homologous recombination as the basis
141 for establishing these isolates. Also, most of these isolates were found in the USA between 2021 and 2022;

142 however, considering that the most prevalent variant in the USA in August 2021 was the Delta variant, it is most
143 unlikely that it did not cause mutations such as T478K and D614G, which were already prevalent. It is
144 inconceivable that the oldest variants (such as T478K and D614G), which were no longer prevalent, were
145 sufficiently present to cause superinfection and be involved in homologous recombination. Also, most of these
146 isolates were isolated in the USA between 2021 and 2022. Still, given that the isolates primarily prevalent in the
147 USA in August 2021 were Delta variants, it is unlikely that an older type of variant without the T478K or D614G
148 mutation was present to cause superinfection and be involved in homologous recombination. This consideration
149 is supported by the fact that all of these BA.1-0.1 and BA.1.1-0.1 isolates retained the sequence of the BA.1
150 lineage in all regions except the S protein (Fig. 4). In addition, the fact that all of these BA.1-0.1 and BA.1.1-
151 0.1 strains retained the sequence of Omicron strain BA.1 except for the S protein also makes it unreasonable to
152 consider that these isolates arose by homologous recombination with an older type of mutant without the T478K
153 or D614G mutations (Fig. 4).

154 Furthermore, some of the BA.1 and BA.1-0.1 isolates have mutation subsets (synonymous: u10135c, ORF3:
155 L106F, N: D343G) up- and downstream of the S gene, and the u10135c and L106F (ORF3) mutations
156 correspond perfectly. Therefore, it is considered that homologous recombination between the BA.1 variant and
157 variants without these mutations did not occur during the mutants' formation processes (Fig. 4). The synonymous
158 mutation c2470u occurred in BA.1.1 compared with BA.1, and this c2470u mutation was retained by most,
159 excluding a few of the BA.1.1-0.1 isolates (SARS-CoV-2_human_USA_IL-CDC-ASC210695497_2022 :
160 GenBank: OM770362.1; SARS-CoV-2_human_USA_NY-CDC-LC0450936_2021: GenBank: OM228453.1）.
161 The synonymous mutation c2470u has also only been observed in a minimal number of BA.1-0.1 isolates
162 （ SARS-CoV-2_human_USA_OR-CDC-LC0470830_2022: GenBank: OM367679.1; SARS-CoV-
163 2_human_USA_ID-CDC-LC0481844_2022: GenBank: OM409228.1; SARS-CoV-2_human_USA_MI-CDC-
164 ASC210597972_2022: GenBank:OM396816.1; SARS-CoV-2_human_USA_WI-CDC-LC0494047_2022:
165 GenBank: OM500517.1）. These results suggest that the establishment of BA.1-0.1 and BA.1.1-0.1 isolates
166 occurred independently. On the other hand, if reversion mutations caused each of these isolates with one amino
167 acid different to the Wuhan-type, these isolates could be detected by examining an astronomical number of
168 isolates. However, these virus strains were detected in the number of sequenced whole genomes (a limited
169 number), rather than in astronomical numbers examined. The fact that most of these mutations occurred without
170 synonymous mutations (Fig. 2) suggests that none of them arose as a result of trial-and-error random mutations
171 in nature. Few synonymous mutations are detected in some BA.1-0.1, BA.1.1-0.1, and BA.2-0.1 isolates (Fig.
172 2 and Supplemental Figure 2), as seen in other viruses (Supplemental Figure 1). The c25000u is the only
173 synonymous mutation that did not occur until BA.1, BA.1.1, BA.2, BA.1-0.1 BA.1.1-0.1, and BA.2-0.1 isolates
174 were formed and was not observed in previous variants such as alpha, beta, gamma, delta, etc. Nevertheless, it
175 is curious to find the occurrence of mutants with synonymous mutations such as c22120u, c24034u, c23635u,
176 c24448u, c21811u, a23884g, c22987u, c23609a, c23413u, c23896u, c22879u, u24097a, c23893u, c24442u,
177 u24847c, c24382u, c22264u, c22879u, c22480u, u21976c, c22480u, g24577a, and u23101c in BA.1.1, BA.1-
178 0.1, and BA.1.1-0.1 isolates (Fig. 2 Synonymous Others), and a22948g, c23635u, c21859u, c22945u, c23701u,
179 c22987u, a24433g, c23347u, u24640c, a24619g, c24865u, a23989g, u23047c, u24346c, c21811u, c21952u,
180 a22753u, c23635u, c24023u, c24382u, and c22572u in BA.2-0.1 isolates (Supplemental Figure 2 Synonymous
181 Others) after the formation of mutants with these subsets.

182 Although the only bias in our isolates collection, was only selection of isolates whose identities showed 100%
183 matches with the query amino acid sequence in the BLAST search, these curious tendencies were observed is
184 very interesting.

185 If two different viral variants infect a single cell simultaneously in the process of establishing various SARS-
186 CoV-2 variants, and if homologous recombination occurs during viral RNA synthesis between the Omicron
187 variant BA.1 lineage and BA.2 lineage, it is expected that there are variants caused by homologous
188 recombination between the BA.1 and BA.2 lineages.

189 Therefore, we also performed BLAST searches for isolates with mutant amino acid subsets of both the Omicron
190 variant BA.1 and BA.2 strains. We detected recombinant isolates of Omicron BA.1 and BA.2 lineages.

191 Surprisingly, the recombinant Omicron BA.1 and BA.2 lineages, SARS-CoV-2/human/PRI/PR-PR-UPRRP-
192 582/2020 (GenBank: ON928946.1), were already present in Puerto Rico in 2020. Omicron (B.1.1.529) is a
193 variant of SARS-CoV-2 first reported to WHO by the Network for Genomics Surveillance in South Africa on
194 November 24, 2021 (15, 16). It was first detected in Botswana and spread to become the predominant variant in
195 circulation worldwide (17). Following the appearance of the original B.1.1.529 variant, several subvariants of
196 Omicron emerged, including BA.1, BA.2, BA.3, BA.4, and BA.5 (21). After October 2022, two subvariants of
197 BA.5 called BQ.1 and BQ.1.1 emerged.

198 The question then arose about why a recombinant strain, SARS-CoV-2/human/PRI/PR-UPRRP-582/2020
199 (GenBank: ON928946.1), already existed in 2020. We searched for SARS-CoV-2 isolates prevalent in Puerto
200 Rico using the keywords "PRI", "PR-UPRRP", and "2020" in the NCBI search; nucleotide
201 (https://www.ncbi.nlm.nih.gov/). Consequently, we found 29 Omicron-associated sequences in the 127 hits
202 obtained (Fig. 5B). These results suggest that the SARS-CoV-2 variants bearing the amino acid sequences of
203 the S protein are identical to Omicron BA.1 and Omicron BA.2 variants, which were already prevalent in Puerto
204 Rico in 2020, with 15 isolates showing the complete Omicron BA.1+ R346K_mut-subset (BA1.1) , and 14
205 isolates showing a synonymous substitution of c21595u. Four isolates had an amino acid sequence of the S
206 protein that perfectly matched that of Omicron BA2 (BA.2_S), four isolates were Omicron BA.2-0.1 (BA.2-
207 S:K440N) and four isolates were Omicron BA.2-0.1 (BA.2-S:K440N)+F79S, BA.2-0.1 (BA.2-
208 S:K440N)+Q613H, BA.2-0.1 (BA.2-S:K440N)+212S+D215E and BA.2-0.1 (BA.2-S:K440N)+212S (Fig. 5B).

209

210 **3   Discussion**

211 Several hypotheses have been proposed in which the original SARS-CoV-2 virus resulted from an accidental
212 laboratory spill. With recent developments in biotechnology, many viruses, including coronaviruses, have been
213 artificially synthesized and used in various experiments (22-24). The artificial generation of mutant viruses in
214 laboratories and study of viral phenotypes by introducing mutations is called "reverse genetics", being a common
215 technique in virology. It has been claimed that SARS-CoV-2 must have been artificially generated because of
216 the unnatural presence of a codon (CGG) encoding a contiguous arginine at the furin cleavage site of SARS-
217 CoV-2. This claim has been refuted based on the following facts: 1) there is no logical reason for a genetically
218 engineered virus to utilize such a suboptimal furin cleavage site; 2) The only previous study on artificial insertion
219 of furin cleavage sites at the S1/S2 boundary of the S protein of SARS-CoV-1 using the pseudotype virus
220 experimental system utilized the optimal "RRSRR" sequence, which is different from the furin cleavage site's
221 sequence present in SARS-CoV-2; 3) There is no evidence of previous studies at the Wuhan Institute of Virology
222 that artificially inserted a complete furin cleavage site in coronaviruses; 4) Unnatural CGG codons adjacent to
223 arginine at the furin cleavage site are rare in coronaviruses but are observed at a particular frequency in SARS-
224 CoV-1, SARS-CoV-2, and other human coronaviruses. However, these are only declarations and are not logical.
225 No one has offered an explanation why a naturally occurring virus would utilize a suboptimal furin cleavage
226 site. There has been no mention of the technical possibility of inserting this furin cleavage site or a CGG codon
227 artificially. The insertion of a polybasic furin cleavage site into the S protein makes it impossible to conclude
228 whether SARS-CoV-2 is a naturally occurring or an artificial virus.

229 Despite the accumulation of many mutations in the S protein of Omicron mutants, most of the mutations are
230 non-synonymous, with only one synonymous mutation of c25000u, which is highly unnatural, leading to the
231 hypothesis that the Omicron mutants were artificially synthesized. The following results presented in this study
232 may support the hypothesis that the Omicron variants were artificially synthesized rather than naturally
233 occurring: 1) the presence of Omicron variant-associated isolates with one mutation site being the Wuhan-type;
234 2) the almost complete absence of synonymous mutations in the S protein in these isolates; 3) the Omicron
235 variant, which should have been first reported to WHO from South Africa on November 24, 2021, was already
236 endemic in Puerto Rico in 2020, and there were isolates that were recombinants between Omicron strains BA1
237 and BA2. In addition, the Omicron mutant-related isolates (BA.1-0.1, BA.1.1-0.1, and BA.2-0.1 isolates) with
238 a Wuhan-type mutation at one of the mutation sites were established. Some had synonymous mutations after

239 establishing the Omicron mutant-related isolates (Fig. 2 and Supplemental Figure 2 Synonymous Others). It is
240 reasonable to assume that viruses with the reversion amino acid mutations caused by non-synonymous mutations
241 in the S protein were artificially synthesized and then acquired further synonymous mutations in the natural
242 environment.

243 Assuming that artificially synthesized mutants with only non-synonymous mutations are spread globally, this
244 would explain how mutants with non-synonymous mutations without previous synonymous mutations develop
245 synonymous mutations under natural circumstances. Considering the current epidemic situation of SARS-CoV-
246 2, it is unlikely that these viruses arose spontaneously. Based on our efforts to explain the formation of the
247 SARS-CoV-2 isolates, they were formed by a completely new mechanism that cannot be explained by previous
248 biology.

249 One idea, the hypothesis that these viruses were artificially generated, is more reasonable than proposing a novel
250 mutation acquisition mechanism. However, is there any reason to artificially create these mutants, which are
251 unlikely to have occurred naturally, given the current SARS-CoV-2 epidemic?

252 It is known that the pathogenicity, host specificity, cell tropism, and immunogenicity of numerous viruses can
253 be altered by mutation of a single (or several) amino acid(s) of a viral protein on the viral envelope (envelope
254 protein, HA protein, spike protein, etc.). A single-amino-acid substitution in the HA protein of the 2009
255 pandemic A (H1N1) influenza viruses changed their replication and pathogenicity (25). In the Chikungunya
256 virus, single amino acid changes in the E2 glycoprotein influenced glycosaminoglycan utilization for target-cell
257 binding (26), and a single amino acid change in the E1 glycoprotein affected mosquito vector specificity and
258 epidemic potential (27). In previous coronaviruses such as MERS-CoV and SARS-CoV-1, point mutations have
259 been demonstrated to confer resistance to neutralizing antibodies (28-30).

260 Consider that the SARS-CoV-2 Omicron variant and its one-amino-acid reversion mutants were artificially and
261 systematically generated. In that case, we should suspect that the other variants (Alpha to Delta) were also
262 artificially generated viruses. Indeed, the lack of findings to date that many of the various mutations seen,
263 especially in the early variants, are indeed associated with increased viral infection (31) supports the hypothesis
264 that each variant was artificially synthesized to identify the amino acids of the S protein responsible for
265 infectivity and pathogenicity. The possibility that the set of mutants was artificially generated to identify amino
266 acids of the S protein involved in the infectivity and virulence is supported.

267 Reverse genetics experiments are an essential part of virus research, and it is inimical to virus research to
268 consider that artificially synthesized viruses were deliberately spread throughout the world. However, now that
269 reverse genetics has become common in virus research, we believe it is not scientific to discuss the mutation
270 process of SARS-CoV-2 without excluding the possibility of artificially synthesized viruses.

271 Finally, we would like to add that while artificially synthesized viruses may have spread, we are not criticizing
272 reverse genetics technology, as such technology has led to marked advances in virology. In addition, our analysis
273 employed databases with a limited number of viral sequences, and we cannot deny the possibility that unreliable
274 data may have been registered due to technical problems in sequencing or some other issues. Furthermore, we
275 do not conclude that these viruses were artificially synthesized and distributed based on malicious intent. This
276 study aims to point out that SARS-CoV-2 has undergone unthinkable mutations based on conventional
277 coronavirus mutation mechanisms, and we hope that the possibility of artificial creation is included in serious
278 discussions on the formation of SARS-CoV-2 variants.

279 Nonetheless, the analysis we have shown here concludes that the Omicron variants were formed by a completely
280 new mechanism that cannot be explained by previous biology. The process of how SARS-CoV-2 mutations
281 occurred should prompt a reconsideration of the SARS-CoV-2 pandemic. If the SARS-CoV-2 epidemic strain
282 is an artificially mutated virus and if the corona disaster (corona hoopla) was a well-designed global experiment
283 in human inoculation and a social experiment, then the design of this experiment and the nature of the virus used
284 make it likely that this experiment (corona hoopla) is a preliminary experiment.

285 **4   Methods**

286 **4. 1   Data acquisition**

287 The SARS-CoV-2 RNA genome, genes, and proteins according to an annotation of SARS-CoV-2 Wuhan-Hu-
288 H1 (COVID-19/Wuhan-Hu-1CHN/2019/First Isolate) NCBI Reference Sequence: NC_045512.2 were used as
289 references for the definition of mutations, and provided a basis for the numbering of nucleotides and amino acids
290 of each protein. Genome data of SARS-CoV-2 isolates described in this article were obtained from the NCBI
291 Nucleotide database (https://www.ncbi.nlm.nih.gov/) on 25/11/2022 to 17/03/2023.

292 **4. 2   Query of representative SARS-CoV-2 variant genome**

293 Amino acid sequences of spike protein of SARS-CoV-2 variants（Alpha:B.1.1.7, Beta:B.1.351, Gamma:P1,
294 Delta:B.1.617.2.63, Lambda:C.37, Mu:B.1.621, Omicron:BA.1, BA.1.1, and BA.2）were obtained from an
295 Internet site, Stanford Coronavirus Antiviral & Resistance Database (https://covdb.stanford.edu/) or Covariant
296 (https://covariants.org/), and used as a query sequence for an NCBI protein BLAST search (blastp: protein-
297 protein   BLAST,
298 https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasth
299 ome). Then, the whole genome sequence of each isolate bearing the query spike sequence was derived from the
300 BLAST search result, identified with query amino acid sequences of 100%. The nucleotide sequences of the
301 detected SARS-CoV-2 genome were as follows: GenBank Accession No.: GenBank: MW423686.2;
302 MW430966.1; MW430967.1; MW422256.1; MW598419.1; MW667552.1; MW667553.1; MW721502.1;
303 MW721504.1; MW520923.1; MW642248.1; MW642249.1; MW642250.1; MZ182066.1; MZ155303.1;
304 MZ155230.1; MZ170364.1; MZ179869.1; MW666666.1; MW696612.1; MW699217.1; MW644498.1;
305 MZ727706.1; MZ620161.1; MZ637380.1; MZ637401.1; MZ780550.1; OL672836.1; OL677199.1;
306 OP769083.1; OL764360.1; OL815447.1; ON762438.1; OL849989.1; OL897126.1; OL896945.1;
307 OL896936.1; OL896931.1; OM233931.1; OM072551.1; OM072822.1; OM296922.1.
308

309 **4. 3   Query of SARS-CoV-2 Omicron variant genome bearing an S protein amino acid sequence in**
310 **which one of the Omicron-type nucleotide mutation subsets was not mutated and retains the original**
311 **SARS-CoV-2 Wuhan-Hu-H1-type arrangement.**

312 For each of the Omicron variants, BA.1, BA.1.1, and BA.2, the isolate series bearing an S protein amino acid
313 sequence in which one of the Omicron-type nucleotide mutation subsets is not mutated and retains the original
314 SARS-CoV-2 Wuhan-Hu-H1-type arrangement were named BA.1-0.1, BA.1.1-0.1 and BA.2-0.1, respectively.
315 In addition, in this article, we named the amino acid sequences of spike protein of BA.1, BA.1.1, and BA.2 as
316 BA.1_S, BA.1.1_S, and BA.2_S, respectively, and then the series of amino acid sequences of spike protein of
317 BA.1-0.1, BA.1.1-0.1, and BA.2-0.1 were named, respectively, as follows: Omicron BA.1-0.1 spike series
318 (BA.1-0.1_Ss)   were   named   as   BA.1_S:V67A;   BA.1_S:69H_70V;   BA.1_S:I95T;
319 BA.1_S:D142G_143V_144Y_145Y; BA.1_S:I211N_212L; BA.1_S:ΔEPE; BA.1_S:D339G; BA.1_S:L371S;
320 BA.1_S:P373S;   BA.1_S:F375S;   BA.1_S:N417K;   BA.1_S:K440N;   BA.1_S:S446G;   BA.1_S:N477S;
321 BA.1_S:K478T;   BA.1_S:A484E;   BA.1_S:R493Q;   BA.1_S:S496G;   BA.1_S:R498Q;   BA.1_S:Y501N;
322 BA.1_S:H505Y;   BA.1_S:K547T;   BA.1_S:G614D;   BA.1_S:Y655H;   BA.1_S:K679N;   BA.1_S:H681P;
323 BA.1_S:K764N; BA.1_S:Y796D; BA.1_S:K856N; BA.1_S:H954Q; BA.1_S:K969N and BA.1_S:F981L /
324 Omicron BA.1.1-0.1 spike series (BA.1.1-0.1_Ss) were named as BA.1.1_S:V67A; BA.1.1_S:69H_70V;
325 BA.1.1_S:I95T;   BA.1.1_S:D142G_143V_144Y_145Y;   BA.1.1_S:I211N_212L;   BA.1.1_S:ΔEPE;
326 BA.1.1_S:D339G;   BA.1.1_S:L371S;   BA.1.1_S:P373S;   BA.1.1_S:F375S;   BA.1.1_S:N417K;
327 BA.1.1_S:K440N;   BA.1.1_S:S446G;   BA.1.1_S:N477S;   BA.1.1_S:K478T;   BA.1.1_S:A484E;
328 BA.1.1_S:R493Q;   BA.1.1_S:S496G;   BA.1.1_S:R498Q;   BA.1.1_S:Y501N;   BA.1.1_S:H505Y;
329 BA.1.1_S:K547T;   BA.1.1_S:G614D;   BA.1.1_S:Y655H;   BA.1.1_S:K679N;   BA.1.1_S:H681P;
330 BA.1.1_S:K764N;   BA.1.1_S:Y796D;   BA.1.1_S:K856N;   BA.1.1_S:H954Q;   BA.1.1_S:K969N;
331 BA.1.1_S:F981L / Omicron BA.2-0.1 spike series (BA.2-0.1_Ss) were named as BA.2_S:I19T;

332 BA.2_S:24L_25P_26P_S27A; BA.2_S:D142G; BA.2_S:V213G; BA.2_S:D339G; BA.2_S:F371S;
333 BA.2_S:P373S; BA.2_S:F375S; BA.2_S:A376T; BA.2_S:N405D; BA.2_S:S408R; BA.2_S:N417K;
334 BA.2_S:K440N; BA.2_S:N477S; BA.2_S:K478T; BA.2_S:A484E; BA.2_S:R493Q; BA.2_S:R498Q;
335 BA.2_S:Y501N; BA.2_S:H505Y; BA.2_S:G614D; BA.2_S:Y655H; BA.2_S:K679N; BA.2_S:H681P;
336 BA.2_S:K764N; BA.2_S:Y796D; BA.2_S:H954Q; BA.2_S:K969N, and these constructs are shown in Figs. 2,
337 4 and supplemental Figure 1. These amino acids sequences of spike protein of SARS-CoV-2 Omicron variants,
338 BA.1-0.1, BA.1.1-0.1, and BA.2-0.1, were used as query sequences for an NCBI protein BLAST search. Then,
339 the whole genome sequences of BA.1-0.1, BA.1.1-0.1, and BA.2-0.1 isolates bearing the query spike sequence
340 were derived from the BLAST search results, identified with a query amino acid sequence of 100%. The
341 nucleotide sequences of the detected SARS-CoV-2 genome were as follows: GenBank Accession No.:
342 OM117411.1; OP797378.1; OM789835.1; OP928789.1; OP928803.1; OP929381.1; OP929396.1;
343 OP929417.1; OM173977.1; OM518459.1; OM566981.1; ON019560.1; OM097227.1; OM096937.1;
344 OM099902.1; OM117114.1; OM096685.1; OM354436.1; OM646886.1; OM472901.1; OM364511.1;
345 OM131858.1; OL815451.1; OL896986.1; OL897116.1; OL897118.1; OL896964.1; OM367679.1;
346 OM343778.1; OM409228.1; OM396816.1; OM134162.1; OM075886.1; OM123427.1; OM122677.1;
347 OM121681.1; OM224850.1; ON246090.1; OM931599.1; OM864873.1; OM906519.1; OM906587.1;
348 OM464776.1; OM015999.1; OM015958.1; OM015597.1; OM016329.1; OL898806.1; OL898861.1;
349 OM016937.1; OM016186.1; OM036549.1; OM051171.1; OM126493.1; OM079115.1; OM099199.1;
350 OM134489.1; OM098796.1; ON618279.1; ON618009.1; OM627701.1; OM356511.1; OM295457.1;
351 ON700063.1; OM033824.1; ON368355.1; OM084700.1; ON208126.1; OM566593.1; OM945690.2;
352 ON030252.1; ON019844.1; OM890075.1; ON020044.1; OM833954.1; ON376082.1; OM084604.1;
353 OP795273.1; ON066609.1; OM352882.1; OM290510.1; OM369978.1; OM199342.1; OM011974.1;
354 OM090274.1; OM043984.1; OM121683.1; OM121624.1; OM175506.1; OM360429.1; OM360221.1;
355 OM358058.1; OM500517.1; OM135027.1; OM742858.1; OM521685.1; OM896558.1; ON694155.1;
356 OM686755.1; OM484260.1; OM332056.1; OM156397.1; OM079447.1; OM134645.1; OM173298.1;
357 OM123082.1; OM116023.1; OM652943.1; OL994299.1; OL994920.1; OM122027.1; OM121015.1;
358 OL898817.1; OM527504.1; OM225320.1; OM931491.1; OM931575.1; OM931587.1; OM034409.1;
359 OM036283.1; OL996129.1; OM035680.1; OM096996.1; ON065532.1; OM968098.1; OM816604.1;
360 ON235452.1; ON334146.1; OP024162.1; OP209732.1; OM354578.1; OM099080.1; OM297301.1;
361 OM297438.1; OM365368.1; OM449159.1; OM078863.1; OM096959.1; OM117155.1; OM133880.1;
362 OM077358.1; OM372005.1; OM770362.1; OM897488.1; OM918459.1; OM918478.1; OL897115.1;
363 OL897114.1; OL986779.1; OL986696.1; OL987046.1; ON831866.1; OM864099.1; OM863888.1;
364 OP745925.1; ON831672.1; OM043643.1; OM176192.1; OM226685.1; OM343689.1; OM295527.1;
365 OM894975.1; OM846676.1; OM822024.1; OM846844.1; OM906550.1; OM015933.1; OM016323.1;
366 OM016331.1; OM035685.1; OM022498.1; OM156115.1; OM036875.1; OM099560.1; OM199246.1;
367 OM067048.1; OM079299.1; OM099911.1; OM116588.1; OM097010.1; OM173300.1; OM805961.1;
368 OM983266.1; OM983325.1; ON618010.1; OM084691.1; ON021265.1; ON039239.1; ON056981.1;
369 ON144127.1; OM770527.1; OM156164.1; OM155119.1; OM199353.1; OM084630.1; OM084605.1;
370 OM084621.1; OM359369.1; OM411574.1; OM584789.1; OM720486.1; OM429777.1; ON047062.1;
371 ON065416.1; OP415118.1; OM954373.1; ON042406.1; OM335528.1; OM332335.1; OM353626.1;
372 OM332813.1; OM197398.1; OM226919.1; OM228399.1; OM225859.1; OM271353.1; OM159454.1;
373 OM224473.1; OM358278.1; OM361030.1; OM412141.1; OM496298.1; OM044048.1; OM121864.1;
374 OM224477.1; OM227379.1; OM228453.1; OM622156.1; OM906370.1; OM970683.1; ON117965.1;
375 OM198667.1; OM357800.1; OM357161.1; OM335230.1; OM261124.1; OM077578.1; OM497172.1;
376 OM625194.1; OM907131.1; ON047464.1; OM911851.1; OM042846.1; OM155337.1; OM097339.1;
377 OM116805.1; OM134409.1; OM686782.1; OM695863.1; OM724725.1; OM174366.1; OM822132.1;
378 OM822106.1; OM822105.1; OM822485.1; OM135143.1; OM125829.1; OM098855.1; OM156118.1;
379 OM155114.1; OM863926.1; OP359104.1; ON209298.1; ON232806.1; ON421981.1; ON811217.1;
380 OM698275.1; ON052769.1; ON060006.1; ON060007.1; ON060009.1; OM843171.1; OM843276.1;
381 OM843550.1; OM843316.1; OM843340.1; ON049267.1; ON450720.1; ON250163.1; ON256603.1;
382 ON480422.1; OM888844.1; OM890089.1; ON009425.2; ON082904.1; OM901275.1; OM877094.2;
383 OM877095.2; OM877096.2; OM877097.2; ON378542.1; ON389858.1; ON389889.1; ON390359.1;
384 OM936703.1; ON352711.1; ON378000.1; ON177702.1; ON205494.1; ON378633.1; ON617689.1;

385   ON619375.1; OM567618.1; OM659585.1; OM770913.1; OM781641.1; OM533441.1; OM533458.1;
386   OM570235.1; OM570252.1; OM570249.1; OM283361.1; OM283362.1; OM283320.1; OM283343.1;
387   ON618014.1; ON618018.1; ON618019.1; ON618363.1; ON311615.1; ON383919.1; OP579158.1;
388   OP054411.1; ON633107.1; ON414693.1; ON422887.1; OP364296.1; OP629673.1; ON363097.1;
389   OP633561.1; ON458445.1; ON592247.1; ON549687.1; ON067040.1; ON321116.1; ON199452.1;
390   ON200331.1; OM861064.1; OM969592.1; ON019120.1; ON221861.1; OM861619.1; ON091288.1;
391   ON151370.1; ON233850.1; ON236456.1; ON296711.1; ON535443.1; ON624524.1; ON377450.1;
392   ON397268.1; ON239032.1; ON373649.1; ON481637.1; ON701163.1; ON312677.1; ON349263.1;
393   ON377487.1; ON377609.1; OM638574.1; OM911616.1; OM988767.1; ON019770.1; OM988769.1;
394   ON468158.1; ON608924.1; ON604965.1; ON535763.1; ON378227.1; ON378238.1; ON728470.1.

4.4 Query of recombinant SARS-CoV-2 Omicron variant between BA.1 and BA.2 genome

Deduced recombinant spike protein between Omicron variants, BA.1 and BA.2 shown as BA.1_S:T19I_L24-_P25-_P26-_A27S_V213G_S371F_T376A_D405N_R408S was used as a query sequence for an NCBI protein BLAST search. The whole genome sequence of BA.1 and BA.2 recombinant-Omicron isolates showed some of the specific amino acid mutations observed in variant BA.1 and BA.2 in the S protein. The nucleotide sequences of the detected SARS-CoV-2 genome were as follows: GenBank Accession No.: OM360636.1; OM410816.1; OM429902.1; OM497964.1; OM565587.1; OM628132.1; ON549899.1; ON449685.1; ON176765.1; OM628094.1; ON099844.1; OM942313.1; ON395480.1; ON171854.1; ON172005.1; ON076710.1; ON928946.1; OM932113.1; OM942438.1; OM989528.1; OM499181.1; ON414822.1; OM878325.1; ON103067.1; ON103153.1; ON419036.1; ON928719.1; ON337887.1; ON420444.1; ON146520.1; OM469541.1; OM904085.1; ON254531.1; OM881098.1; ON373310.1.

4.5 Query of SARS-CoV-2 Omicron variant genome detected in Puerto Rico in 2020

Nucleotide sequences were searched using the keywords PRI PR-UPRRP 2020 (Search details: PRI[All Fields] AND (PR[All Fields] AND UPRRP[All Fields]) AND 2020[All Fields]). The search results were all SARS-CoV-2 isolate genome sequences. Among these sequences, SARS-CoV-2 Omicron variant-related sequences were picked up as follows: GenBank Accession No.: ON928761.1; ON928660.1; ON928794.1; ON928762.1; ON928848.1; ON928741.1; ON928918.1; ON928680.1; ON928975.1; ON928949.1; ON928673.1; ON928865.1; ON928716.1; ON928663.1; ON928779.1; ON928896.1; ON928946.1; ON928912.1; ON928704.1; ON928873.1; ON928813.1; ON928898.1; ON928765.1; ON928912.1; ON928883.1; ON928957.1; ON928880.1; ON928699.1; ON928724.1; ON928941.1.

Genomes were aligned using SnapGene software or GENETYX software. Numbering of nucleotides and amino acids of each protein was determined using Wuhan-Hu-1 (NC_045512.2; COVID-19/Wuhan-Hu-1CHN/2019/First Isolate) as a reference strain for the definition of mutations.

5. References

1.   F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J. J. Zheng, L. Xu, E. C. Holmes and Y. Z. Zhang: A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265-269 (2020) doi:10.1038/s41586-020-2008-3

2.   R. A. Khailany, M. Safdar and M. Ozaslan: Genomic characterization of a novel SARS-CoV-2. *Gene Rep*, 19, 100682 (2020) doi:10.1016/j.genrep.2020.100682

3.   Y. M. Bar-On, A. Flamholz, R. Phillips and R. Milo: SARS-CoV-2 (COVID-19) by the numbers. *Elife*, 9 (2020) doi:10.7554/eLife.57309

4.   J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang and X. Wang: Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215-220 (2020) doi:10.1038/s41586-020-2180-5

5.   D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham and J. S. McLellan: Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260-1263 (2020) doi:10.1126/science.abb2507

6.   B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, K. M. Hastie, M. D. Parker, D. G. Partridge, C. M. Evans, T. M. Freeman, T. I. de Silva, C.-G. G. Sheffield, C. McDanal, L. G. Perez, H. Tang, A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Saphire and D. C. Montefiori: Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4), 812-827 e19 (2020) doi:10.1016/j.cell.2020.06.043

7.   A. Carrazco-Montalvo, A. Bruno, D. de Mora, M. Olmedo, J. Garces, M. Paez, M. Regato-Arrata, M. Gonzalez, J. Romero, O. Mestanza, B. Freire-Paspuel, A. Gaviria, S. A. Orlando, M. A. Garcia-Bereguiain and L. Patino: First Report of SARS-CoV-2 Lineage B.1.1.7 (Alpha Variant) in Ecuador, January 2021. *Infect Drug Resist*, 14, 5183-5188 (2021) doi:10.2147/IDR.S319439

8.   S. Zarate, B. Taboada, J. E. Munoz-Medina, P. Isa, A. Sanchez-Flores, C. Boukadida, A. Herrera-Estrella, N. Selem Mojica, M. Rosales-Rivera, B. Gomez-Gil, A. G. Salas-Lais, C. E. Santacruz-Tinoco, H. Montoya-Fuentes, J. E. Alvarado-Yaah, G. M. Molina-Salinas, G. E. Espinoza-Ayala, J. A. Enciso-Moreno, R. M. Gutierrez-Rios, A. Loza, J. Moreno-Contreras, R. Garcia-Lopez, X. Rivera-Gutierrez, A. Comas-Garcia, R. M. Wong-Chew, M. E. Jimenez-Corona, R. M. Del Angel, J. A. Vazquez-Perez, M. Matias-Florentino, M. Perez-Garcia, S. Avila-Rios, H. G. Castelan-Sanchez, L. Delaye, L. P. Martinez-Castilla, M. Escalera-Zamudio, S. Lopez and C. F. Arias: The Alpha Variant (B.1.1.7) of SARS-CoV-2 Failed to Become Dominant in Mexico. *Microbiol Spectr*, 10(2), e0224021 (2022) doi:10.1128/spectrum.02240-21

9.   H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E. J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A. J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, N.-y. Hsiao, S. Korsman, M.-A. Davies, L. Tyers, I. Mudau, D. York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A. Alisoltani-Dehkordi, A. Godzik, C. K. Wibmer, B. T. Sewell, J. Lourenço, L. C. J. Alcantara, S. L. Kosakovsky Pond, S. Weaver, D. Martin, R. J. Lessells, J. N. Bhiman, C. Williamson and T. de Oliveira: Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*, 592(7854), 438-443 (2021) doi:10.1038/s41586-021-03402-9

10.   T. Fujino, H. Nomoto, S. Kutsuna, M. Ujiie, T. Suzuki, R. Sato, T. Fujimoto, M. Kuroda, T. Wakita and N. Ohmagari: Novel SARS-CoV-2 Variant in Travelers from Brazil to Japan. *Emerg Infect Dis*, 27(4), 1243-5 (2021) doi:10.3201/eid2704.210138

11.   M. Joshi, M. Kumar, V. Srivastava, D. Kumar, D. S. Rathore, R. Pandit, D. W. Graham and C. G. Joshi: Genetic sequencing detected the SARS-CoV-2 delta variant in wastewater a month prior to the first COVID-19 case in Ahmedabad (India). *Environ Pollut*, 310, 119757 (2022) doi:10.1016/j.envpol.2022.119757

12.   K. Laiton-Donato, C. Franco-Munoz, D. A. Alvarez-Diaz, H. A. Ruiz-Moreno, J. A. Usme-Ciro, D. A. Prada, J. Reales-Gonzalez, S. Corchuelo, M. T. Herrera-Sepulveda, J. Naizaque, G. Santamaria, J. Rivera, P. Rojas, J. H. Ortiz, A. Cardona, D. Malo, F. Prieto-Alvarado, F. R. Gomez, M. Wiesner, M. L. O. Martinez and M. Mercado-Reyes: Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. *Infect Genet Evol*, 95, 105038 (2021) doi:10.1016/j.meegid.2021.105038

469   13.   P. L. Wink, F. C. Z. Volpato, F. L. Monteiro, J. B. Willig, A. P. Zavascki, A. L. Barth and A. F. Martins:
470   First identification of SARS-CoV-2 lambda (C.37) variant in Southern Brazil. *Infect Control Hosp Epidemiol*,
471   43(12), 1996-1997 (2022) doi:10.1017/ice.2021.390
472   14.   COVID-19 Weekly Epidemiological Update Edition 44, published 15 June 2021. *World Health*
473   *Organization* (2021)
474   15.   J. Quarleri, V. Galvan and M. V. Delpino: Omicron variant of the SARS-CoV-2: a quest to define the
475   consequences of its high mutational load. *Geroscience*, 44(1), 53-56 (2022) doi:10.1007/s11357-021-00500-4
476   16.   A. Gowrisankar, T. M. C. Priyanka and S. Banerjee: Omicron: a mysterious variant of concern. *Eur*
477   *Phys J Plus*, 137(1), 100 (2022) doi:10.1140/epjp/s13360-021-02321-y
478   17.   A. Vitiello, F. Ferrara, A. M. Auti, M. Di Domenico and M. Boccellino: Advances in the Omicron
479   variant development. *J Intern Med*, 292(1), 81-90 (2022) doi:10.1111/joim.13478
480   18.   Y. Fan, X. Li, L. Zhang, S. Wan, L. Zhang and F. Zhou: SARS-CoV-2 Omicron variant: recent progress
481   and future perspectives. *Signal Transduct Target Ther*, 7(1), 141 (2022) doi:10.1038/s41392-022-00997-x
482   19.   J. Fonager, M. Bennedbaek, P. Bager, J. Wohlfahrt, K. M. Ellegaard, A. C. Ingham, S. M. Edslev, M.
483   Stegger, R. N. Sieber, R. Lassauniere, A. Fomsgaard, T. Lillebaek, C. W. Svarrer, F. T. Moller, C. H. Moller,
484   R. Legarth, T. V. Sydenham, K. Steinke, S. J. Paulsen, J. A. S. Castruita, U. V. Schneider, C. H. Schouw, X. C.
485   Nielsen, M. Overvad, R. T. Nielsen, R. L. Marvig, M. S. Pedersen, L. Nielsen, L. L. Nilsson, J. Bybjerg-
486   Grauholm, I. H. Tarpgaard, T. S. Ebsen, J. U. H. Lam, V. Gunalan and M. Rasmussen: Molecular epidemiology
487   of the SARS-CoV-2 variant Omicron BA.2 sub-lineage in Denmark, 29 November 2021 to 2 January 2022.
488   *Euro Surveill*, 27(10) (2022) doi:10.2807/1560-7917.ES.2022.27.10.2200181
489   20.   L. B. Shrestha, C. Foster, W. Rawlinson, N. Tedla and R. A. Bull: Evolution of the SARS-CoV-2
490   omicron variants BA.1 to BA.5: Implications for immune escape and transmission. *Rev Med Virol*, 32(5), e2381
491   (2022) doi:10.1002/rmv.2381
492   21.   L. Yao, K. L. Zhu, X. L. Jiang, X. J. Wang, B. D. Zhan, H. X. Gao, X. Y. Geng, L. J. Duan, E. H. Dai
493   and M. J. Ma: Omicron subvariants escape antibodies elicited by vaccination and BA.2.2 infection. *Lancet Infect*
494   *Dis*, 22(8), 1116-1117 (2022) doi:10.1016/S1473-3099(22)00410-8
495   22.   S. Torii, C. Ono, R. Suzuki, Y. Morioka, I. Anzai, Y. Fauzyah, Y. Maeda, W. Kamitani, T. Fukuhara
496   and Y. Matsuura: Establishment of a reverse genetics system for SARS-CoV-2 using circular polymerase
497   extension reaction. *Cell Rep*, 35(3), 109014 (2021) doi:10.1016/j.celrep.2021.109014
498   23.   T. Y. Taha, I. P. Chen, J. M. Hayashi, T. Tabata, K. Walcott, G. R. Kimmerly, A. M. Syed, A. Ciling,
499   R. K. Suryawanshi, H. S. Martin, B. H. Bach, C. L. Tsou, M. Montano, M. M. Khalid, B. K. Sreekumar, G.
500   Renuka Kumar, S. Wyman, J. A. Doudna and M. Ott: Rapid assembly of SARS-CoV-2 genomes reveals
501   attenuation of the Omicron BA.1 variant through NSP6. *Nat Commun*, 14(1), 2308 (2023) doi:10.1038/s41467-
502   023-37787-0
503   24.   W. Wang, X. Peng, Y. Jin, J. A. Pan and D. Guo: Reverse genetics systems for SARS-CoV-2. *J Med*
504   *Virol*, 94(7), 3017-3031 (2022) doi:10.1002/jmv.27738
505   25.   L. Xu, L. Bao, Q. Lv, W. Deng, Y. Ma, F. Li, L. Zhan, H. Zhu, C. Ma and C. Qin: A single-amino-acid
506   substitution in the HA protein changes the replication and pathogenicity of the 2009 pandemic A (H1N1)
507   influenza viruses in vitro and in vivo. *Virol J*, 7, 325 (2010) doi:10.1186/1743-422X-7-325
508   26.   L. A. Silva, S. Khomandiak, A. W. Ashbrook, R. Weller, M. T. Heise, T. E. Morrison and T. S.
509   Dermody: A single-amino-acid polymorphism in Chikungunya virus E2 glycoprotein influences
510   glycosaminoglycan utilization. *J Virol*, 88(5), 2385-97 (2014) doi:10.1128/JVI.03116-13
511   27.   K. A. Tsetsarkin, D. L. Vanlandingham, C. E. McGee and S. Higgs: A single mutation in chikungunya
512   virus   affects   vector   specificity   and   epidemic   potential. *PLoS Pathog*, 3(12), e201 (2007)
513   doi:10.1371/journal.ppat.0030201
514   28.   X. C. Tang, S. S. Agnihothram, Y. Jiao, J. Stanhope, R. L. Graham, E. C. Peterson, Y. Avnir, A. S.
515   Tallarico, J. Sheehan, Q. Zhu, R. S. Baric and W. A. Marasco: Identification of human neutralizing antibodies
516   against MERS-CoV and their role in virus adaptive evolution. *Proc Natl Acad Sci U S A*, 111(19), E2018-26
517   (2014) doi:10.1073/pnas.1402074111
518   29.   J. Sui, D. R. Aird, A. Tamin, A. Murakami, M. Yan, A. Yammanuru, H. Jing, B. Kan, X. Liu, Q. Zhu,
519   Q. A. Yuan, G. P. Adams, W. J. Bellini, J. Xu, L. J. Anderson and W. A. Marasco: Broadening of neutralization
520   activity to directly block a dominant antibody-driven SARS-coronavirus evolution pathway. *PLoS Pathog*, 4(11),
521   e1000197 (2008) doi:10.1371/journal.ppat.1000197

522 30.   J. ter Meulen, E. N. van den Brink, L. L. Poon, W. E. Marissen, C. S. Leung, F. Cox, C. Y. Cheung, A.
523 Q. Bakker, J. A. Bogaards, E. van Deventer, W. Preiser, H. W. Doerr, V. T. Chow, J. de Kruif, J. S. Peiris and
524 J. Goudsmit: Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of
525 escape mutants. *PLoS Med*, 3(7), e237 (2006) doi:10.1371/journal.pmed.0030237
526 31.   L. van Dorp, D. Richard, C. C. S. Tan, L. P. Shaw, M. Acman and F. Balloux: No evidence for increased
527 transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun*, 11(1), 5986 (2020)
528 doi:10.1038/s41467-020-19818-2

529

530 **Conflict of Interest**

531 The authors declare that the research was conducted in the absence of any commercial or financial
532 relationships that could be construed as a potential conflict of interest.

Figure legends

534 **Fig. 1. Mutation subsets of S protein of SARS-CoV-2 variants**.

535 Sequences of S protein of SARS-CoV-2 variants（variants of concern, VOCs: Alpha:B.1.1.7, Beta:B.1.351,
536 Gamma: Delta:B.1.617.2.63, and Omicron:BA.1; BA.2 and variants of interest, VOIs: Lambda:C.37,
537 Mu:B.1.621) are compared with the SARS-CoV-2 Wuhan-Hu-H1 reference sequence, and different amino acids
538 (amino acid change, deletion, and insertion) and synonymous changes of nucleotides are shown. Non-
539 synonymous changes are shown by amino acid changes (capital letters), and synonymous changes are shown by
540 nucleotide changes (small letters). Amino acids different from Wuhan-Hu-H1 found in each variant: Alpha:
541 B.1.1.7, Beta: B.1.351, Gamma: P1, Delta: B.1.617.2.63, Lambda: C.37, Mu: B.1.621, and Omicron: BA.1,
542 BA.2 are highlighted with red, orange, green, yellow, aquamarine, lime, deep sky blue, and blue violet,
543 respectively. Amino acid changes common to Omicron: BA.1 and BA.2 are shown in purple.

544

545 **Fig. 2. Mutations of S proteins of SARS-CoV-2 Omicron isolates.**

546 **(A)** Different amino acids and synonymous changes of nucleotides in **S** proteins of SARS-CoV-2 Omicron BA.1,
547 BA.1.1 isolates, and BA.1-0.1s compared with SARS-CoV-2 Wuhan-Hu-H1. Nucleotide deletions and
548 insertions were "deletion[1]" (deletion: nt 21,766-21,771), "deletion[2]" (deletion: nt 21,987-21,995), "deletion[3]"
549 (deletion: nt 22,194-22,196), and "insertion[4]" (insertion between 22,206-22,196), and introduced amino acid
550 changes were H69-_V70-, G142D_V143-_Y144-_Y145-, N211I_L212-, and 215ins.EPE, respectively. (B)
551 Different amino acids and synonymous nucleotide changes in S proteins of SARS-CoV-2 Omicron BA.1.1-0.1
552 isolates. Amino acids different from Wuhan-Hu-H1 found in each variant: Alpha: B.1.1.7, Beta: B.1.351,
553 Gamma: P1, Delta: B.1.617.2.63, Mu: B.1.621, and Omicron: BA.1, BA.2 are highlighted with red, orange,
554 green, yellow, lime, deep sky blue, and blue violet, respectively. Amino acid changes common to Omicron:BA.1
555 and BA.2 are highlighted with purple.

556

557 **Fig. 3. Estimated homologous recombination breakpoints of the SARS-CoV-2 S gene of Omicron BA.1-**
558 **0.1 or BA.1.1-0.1.**

559 Sequence alignment of amino acids and their coding nucleotides (nt.21,746-21,787; nt.22,658-22,702;
560 nt.22,976-23,011, and nt.23,582-23,620) containing the mutation point of the SARS-CoV-2 S gene of the
561 Omicron BA.1 variant compared with SARS-CoV-2 Wuhan-Hu-H1. Changed nucleotides and amino acids of
562 Omicron BA.1 are shown in red letters. Estimated homologous recombination breakpoints of the SARS-CoV-
563 2 S gene of Omicron BA.1-0.1 or BA.1.1-0.1 are shown by asterisks.

564

565 **Fig. 4. Representative mutations of SARS-CoV-2 Omicron isolates other than S protein.**

566 **(A)** Representative amino acids and synonymous changes of nucleotides of SARS-CoV-2 Omicron BA.1,
567 BA.1.1 isolates, and BA.1-0.1 compared with SARS-CoV-2 Wuhan-Hu-H1. **(B)** Representative amino acids
568 and synonymous changes of nucleotides of SARS-CoV-2 Omicron BA.1.1-0.1 compared with SARS-CoV-2
569 Wuhan-Hu-H1. Amino acids different from Wuhan-Hu-H1 found in each variant: Alpha: B.1.1.7, Lambda: C.37,
570 Mu: B.1.621, and Omicron: BA.1 are highlighted with red, aquamarine, deep sky blue, and blue violet,
571 respectively.

572 Amino acid changes common to Omicron: BA.1 and BA.2 are shown in purple. Synonymous nucleotide
573 changes: c2470u observed in Omicron:BA.1.1 mainly shown with blue. Synonymous and non-synonymous
574 changes: u10135c of nsp5, L106F in ORF3, and D343G in N protein subset observed in ~40% of Omicron;

575 BA.1-0.1 are highlighted with emerald-green. Undetermined nucleotides or amino acids are shown as UD or X,
576 respectively.

577

578 **Fig. 5. Mutations of S proteins of SARS-CoV-2 Omicron BA.1-BA.2 recombinant isolates and SARS-CoV-**
579 **2 Omicron BA.1 and BA.2 isolates detected in Puerto Rico in 2020.**

580 **(A)** Different amino acids and synonymous change of nucleotides in S proteins of SARS-CoV-2 Omicron BA.1-
581 BA.2 recombinant isolates compared with SARS-CoV-2 Wuhan-Hu-H1. Nucleotide deletions, "deletion[5]"
582 (deletion: nt 21,633-21,641), introduced the amino acids changes L24-_P25-_P26-_A27S. **(B)** Different amino
583 acids and synonymous change of nucleotides in S proteins of SARS-CoV-2 Omicron BA.1.1 and Omicron
584 BA.1-BA.2 recombinant isolate, highlighted with magenta (GenBank: ON928946.1), Omicron BA.2, and
585 Omicron 2-0.1(K440N), detected in Puerto Rico in 2020. Amino acids different from Wuhan-Hu-H1 found in
586 each variant: Alpha: B.1.1.7, Beta: B.1.351, Gamma: P1, Delta: B.1.617.2.63, Mu: B.1.621, and Omicron: BA.1,
587 BA.2 are highlighted with red, orange, green, yellow, lime, deep sky blue, and blue-violet, respectively. Amino
588 acid changes common to Omicron: BA.1 and BA.2 are highlighted with purple.

589

590 **Supplemental Figure 1**

591 **Human coronavirus 229E strains detected in Seattle, USA, in 2015 and 2019.**

592 Alignment of nucleotide **(A)** and amino acid **(B)** sequences of the S protein of Human coronavirus 229E strains,
593 HCoV_229E/Seattle/USA/SC3112/2015 (GenBank: KY983587.1), and CoV_229E/Seattle/USA/SC0865/2019
594 (GenBank: MN306046.1). The number of nucleotide substitutions observed between them was 32, amino acid
595 substitutions numbered 18 between them, and the synonymous (14: 32-18)–non-synonymous mutation (18) rate
596 between them was 1.285

597

598 **Supplemental Figure 2**

599 **Different amino acids and synonymous changes of nucleotides in S proteins of SARS-CoV-2 Omicron**
600 **BA.2 isolates and BA.2-0.1s compared with SARS-CoV-2 Wuhan-Hu-H1.**

601 Nucleotide deletions, "deletion[5]" (deletion: nt 21,633-21,641), introduced the amino acid changes L24-_P25-
602 _P26-_A27S. Amino acids different from Wuhan-Hu-H1 found in each variant: Alpha: B.1.1.7, Beta: B.1.351,
603 Gamma: P1, Delta: B.1.617.2.63, Mu: B.1.621, and Omicron: BA.1, BA.2 are highlighted with red, orange,
604 green, yellow, lime, deep sky blue, and blue-violet, respectively. Amino acid changes common to Omicron:
605 BA.1 and BA.2 are highlighted with purple.

606

607 **Supplemental Figure 3**

608 **Estimated homologous recombination breakpoints of the SARS-CoV-2 S gene of the Omicron BA.2-0.1**
609 **or BA.1-BA.2 recombinant.**

610 **(A)** Sequence alignment of the amino acids and coding nucleotides (nt. 22,658-22,702) containing the mutation
611 point of the SARS-CoV-2 S gene of Omicron BA.2 variants compared with SARS-CoV-2 Wuhan-Hu-H1. **(B)**
612 Sequence alignment of the amino acids and coding nucleotides (nt. 22,178-22,222) containing the mutation point
613 of the SARS-CoV-2 S gene of Omicron BA,1, BA.2 variant and BA.1-BA.2 recombinant isolate compared with
614 SARS-CoV-2 Wuhan-Hu-H1. Changed nucleotides and amino acids of Omicron variants BA.1, BA.2, and

615   BA.1-BA.2 recombinant isolates compared with SARS-CoV-2 Wuhan-Hu-H1 sequences are shown in red
616   letters. Asterisks show an estimated homologous recombination breakpoint of the SARS-CoV-2 S gene of
617   Omicron BA.2-0.1.

Fig. 1

Fig. 2 A

Fig. 2 B

Fig. 3

```
                      21,750     21,760     21,770     21,780
SARS-CoV-2_Wuhan-Hu-1  GUUACUUGGUUCCAUGCUAUACAUGUCUCUGGGACCAAUGGU
SARS-CoV-2_Omicron_BA.1 GUUACUUGGUUCCAUGUUAU------CUCUGGGACCAAUGGU
break point                     ***
                       V U W F H A I H V S G U N  G
                       V U W F H V I - - S G U N  G
                               A67V  H69- V70-


                      22,660    22,670    22,680    22,690    22,700
                       UCUGUCCUAUAUAAUUCCGCAUCAUUUUCCACUUUUAAGUGUUAU
                       UCUGUCCUAUAUAAUCUCGCACCAUUUUUCACUUUUAAGUGUUAU
                                      **** ******
                       S V L Y N S A S F S T F K C Y
                       S V L Y N L A P F F T F K C Y
                               S371L S373P S375F


                      22,980    22,990    23,000    23,010
                       AUCUAUCAGGCCGGUAGCACACCUUGUAAUGGUGUU
                       AUCUAUCAGGCCGGUAACAAACCUUGUAAUGGUGUU
                                       **
                       I Y Q A G S T P C N G V
                       I Y Q A G N K P C N G V
                               S477N  T478K


                       23,590    23,600    23,610    23,620
                       UAUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGUAGU
                       UAUCAGACUCAGACUAAGUCUCAUCGGCGGGCACGUAGU
                                       ****
                       Y Q T Q T N S P R R A R S
                       Y Q T Q T K S H R R A R S
                               N679K   P681H
```

Fig. 4 A

Fig. 4 B

Fig. 5

A
SARS-CoV-2_Omicron_BA1-BA2 recombinant (BA1/2)

B
SARS-CoV-2_isolated in Puerto Rico in 2020

Supplemental Figure 1

A

B

Supplemental Figure 2

Supplemental Figure 3

A

```
                                  22,660    22,670    22,680    22,690    22,700
SARS-CoV-2_Wuhan-Hu-1.      UCUGUCCUAUAUAAUUCCGCAUCAUUUUCCACUUUUAAGUGUUAU
SARS-CoV-2_Omicron_BA.2     UCUGUCCUAUAUAAUUUCGCACCAUUUUUCGCUUUUAAGUGUUAU
Omicron_BA.2-0.1 break point.                    **** ****** *
                                  S V L Y N S A S F S T F K C Y
                                  S V L Y N F A P F F A F K C Y
                                         S371F S373P S375F T376A
```

B

```
                            22,180    22,190    22,200              22,210    22,220
SARS-CoV-2_Wuhan-Hu-1       AAGCACACGCCUAUUAAUUUAGUGCGUGA---------UCUCCCUCAGGGUUUU
SARS-CoV-2_Omicron_BA.1     AAGCACACGCCUAUUAU---AGUGCGUGAGCCAGAAGAUCUCCCUCAGGGUUUU
SARS-CoV-2_Omicron_BA.2     AAGCACACGCCUAUUAAUUUAGGGCGUGA---------UCUCCCUCAGGGUUUU
Omicron_BA.1-BA.2_rec       AAGCACACGCCUAUUAU---AGGGCGUGAGCCAGAAGAUCUCCCUCAGGGUUUU
Omicron_BA.1-BA.2_rec break point          ** ******
                            K H U P I N L V R - - - D L P Q G F
                            K H U P I - I V R E P E D L P Q G F
                            K H U P I N L G R - - - D L P Q G F
                            K H U P I - I G R E P E D L P Q G F
                                N211- L212I V213G insertion
```