

1 **Unnaturalness in the evolution process of the SARS-CoV-2 variants and the possibility of deliberate**
2 **natural selection**

3 **Authors:** Atsushi Tanaka^{1,2,*}, Takayuki Miyazawa^{2,3,*}

4

5 ¹Division of Research Animal Laboratory and Translational Medicine, Research and Development Center
6 Osaka Medical and Pharmaceutical University, Takatsuki, Osaka 569-8686, Japan

7 ²Laboratory of Virus-Host Coevolution, Institute for Life and Medical Sciences, Kyoto University, Sakyo-ku,
8 Kyoto 606-8507, Japan

9 ³Resilience Research Unit, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8540, Japan

10

11 ***Corresponding authors:**

12 Dr. Atsushi Tanaka, Division of Research Animal Laboratory and Translational Medicine, Research and
13 Development Center Osaka Medical and Pharmaceutical University, Takatsuki, Osaka 569-8686, Japan

14 E-mail: atsushi.tanaka@ompu.ac.jp

15 Dr. Takayuki Miyazawa, Laboratory of Virus-Host Coevolution, Institute for Life and Medical Sciences, Kyoto
16 University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan.

17 E-mail: takavet@infront.kyoto-u.ac.jp

18 Phone: +81-(0)75-751-4814

19

20 Running head: Unnatural evolution of the SARS-CoV-2 variants

21 **Abstract**

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

37 **Introduction**

38 COVID-19, the coronavirus disease 2019, caused by severe acute respiratory syndrome coronavirus 2
39 (SARS-CoV-2), was first reported in December 2019 in Wuhan, China (1). This emerging infectious disease
40 has been unprecedentedly fast, diffusing worldwide, leading the World Health Organization (WHO) to declare
41 a global pandemic of COVID-19 on March 11, 2020 (<https://www.who.int/>). SARS-CoV-2, belonging to
42 betacoronavirus subgroup B, has a single-stranded positive-sense RNA genome that codes for ten genes,
43 ultimately producing 26 proteins according to an annotation of NCBI Reference Sequence: NC_045512.2. Its
44 genome size varies from 29.8 kb to 29.9 kb. It consists of four structural proteins: spike (S), envelope (E),
45 membrane (M), and nucleocapsid (N) proteins (2, 3). In the structural proteins, the S protein as the surface
46 glycoprotein is the largest protein, approximately 180 kDa, consisting of two subunits, S1 and S2. It mediates
47 membrane fusion and ultimately facilitates virus entry. The receptor binding domain (RBD) (amino acid
48 residues 319–541) of the S1 subunit interacts with the angiotensin-converting enzyme 2 (ACE2), binding to its
49 peptidase domain (4, 5).

50 Over the past three years (2019-2022), SARS-CoV-2 was re-accelerated by new variants that emerged
51 over several months in various geographical regions and then transferred over the world, then induced the
52 pandemic repeatedly.

53 In the early stage of the first pandemic, the most impactful mutation of SARS-CoV-2 was a non-
54 synonymous mutation D614G in the S protein. This mutation, which was not present in the ancestral lineage
55 that caused the Wuhan outbreak, became quickly dominant worldwide (6). Soon later, the variant of concern,

56 B.1.1.7 : 20I (Alpha, V1), the lineage B.1.1.7 (clade 501.YV1) or Alpha, characterized by 17 unique mutations
57 containing ten amino acids differences in the S protein, was first detected in southeastern England in late
58 September 2020 (7) and expanded rapidly across the United Kingdom to become predominant during early
59 2021, spreading to most European countries with similar success. By November 2021, local transmission of this
60 lineage had been reported in 175 countries (8). Shortly after that, the emergence of variant strains of SARS-
61 CoV-2 Alpha, variants B.1.351 : 20H (Beta, V2) was identified in October 2020, which was first detected in the
62 Eastern Cape province of South Africa from specimens collected in early August. This Beta variant was spread
63 within South Africa and appeared to have displaced the other SARS-CoV-2 lineages circulating there (9). Then
64 the variant P.1: 20J (Gamma, V3) was identified in Brazil in December 2020 that appeared to have evolved in
65 Brazil. Health officials in Japan first reported it publicly on January 10, 2021, after finding the Gamma variant
66 in four Brazilian travelers at Haneda Airport in Tokyo, Japan (10).

67 At about the same time, the Delta variant (Pango lineage designation B.1.617.2), which was first detected
68 in India in February 2021, and the Mu variant, also known as lineage B.1.621 first detected in Colombia in
69 January 2021 were reported (11, 12).

70 Almost one year later, these emergences of variants of concern, Omicron (phylogenetic assignment of
71 named global outbreak (Pango) lineage designation B.1.1.529; BA.1, Nextstrain clade 21K) is a variant of
72 SARS-CoV-2 first reported to the WHO by the Network for Genomics Surveillance in South Africa on
73 November 24, 2021 (13, 14) that has more than 50 amino acids changes when compared with first reported
74 strain Wuhan-Hu-H1 (NCBI Reference Sequence: NC_045512.2.), and 39 amino acids difference of them were

75 observed in the S protein. This variant was first detected in Botswana and has become the predominant
76 circulating variant worldwide (15).

77 In the United States, the San Francisco Department of Public Health has confirmed a case of COVID-19
78 among individuals in California was caused by the Omicron variant BA.1, which transferred from a traveler
79 who returned from South Africa on November 22, 2021 ([https://www.cdc.gov/media/releases/2021/s1201-
80 Omicron-variant.html](https://www.cdc.gov/media/releases/2021/s1201-
80 Omicron-variant.html)). Then, the first Omicron sub-lineage BA.1, expanded rapidly and replaced the Delta
81 variant (16).

82 Soon after (less than two weeks after), the first identification of the Omicron variant BA.1, the new
83 Omicron variant, BA.2 lineage, which has 31 amino acids changes in the S protein when compared with the
84 Wuhan-Hu-H1 in Denmark on December 5, 2021 (17). On February 22, 2022, WHO mentioned on the
85 Omicron sublineage BA.2 ([https://www.who.int/news/item/22-02-2022-statement-on-Omicron-sublineage-
86 ba.2](https://www.who.int/news/item/22-02-2022-statement-on-Omicron-sublineage-
86 ba.2)) that the Omicron variant of concern was currently the dominant variant circulating globally, replacing the
87 Delta variant (Pango lineage designation B.1.617.2) ([https://www.who.int/docs/default-
89 corr2.pdf?sfvrsn=918b09d_20](https://www.who.int/docs/default-
88 source/coronaviruse/2022-01-07-global-technical-brief-and-priority-action-on-Omicron---
89 corr2.pdf?sfvrsn=918b09d_20)), accounting for nearly all sequences reported to GISAID. Then as of March 16,
90 2023, WHO stated that the Omicron variants account for over 98% of the publicly available viral sequences
91 since February 2022 ([https://www.who.int/news/item/16-03-2023-statement-on-the-update-of-who-s-working-
92 definitions-and-tracking-system-for-SARS-CoV-2-variants-of-concern-and-variants-of-interest](https://www.who.int/news/item/16-03-2023-statement-on-the-update-of-who-s-working-
92 definitions-and-tracking-system-for-SARS-CoV-2-variants-of-concern-and-variants-of-interest)).

93 The Omicron variants BA.1 and BA.2 are suggested to have expanded and diverged around October to
94 December 2021, respectively. These mutants are estimated to diverge from the common ancestor around
95 February to March 2021 (18). Since the Omicron variants BA.1 and BA.2 share a common 14 amino acids
96 mutation in the S protein, the common ancestor of Omicron variants BA.1 and BA.2 may have already acquired
97 the 14 amino acid mutation in the S protein region around February to March 2021; however, no common
98 ancestral strain has been found in the international databases, and the strains may have acquired their mutations
99 through different pathways.

100 In this study, we tried to dissolve the evolution process of the Omicron variant, which has twice more
101 amino acid mutations in the S protein than other variants, by examining the rank order of introduction of amino
102 acid mutations in the S protein.

103

104 **Results**

105 Each variant is thought to have arisen through an independent evolutionary pathway from isolates with
106 the D614G mutation in the S protein. In the genetic variation in the S protein in these variants, most of the
107 mutations were non-synonymous (Fig. 1). There were no synonymous mutations in the Alpha, Beta, Gamma,
108 Delta, or Mu variants, but only one each in the Lambda and Omicron variants. Among these variants, the
109 Omicron variant (BA.1 lineage), which shows the greatest accumulation of mutations in the S protein, is
110 primarily non-synonymous in the S protein and has only one synonymous mutation at c25000u. The

111 synonymous/non-synonymous ratio is abnormal, given how human coronaviruses have mutated (See
112 Supplemental Figure 1).

113 At first, we presumed the existence of the isolate of SARS-CoV-2, whose amino acid sequence of S
114 protein contains the Omicron-BA.1-type mutation subsets, but one mutation position was not mutated and
115 constituted with the original Wuhan-type amino acid sequence. We designated these isolates as BA.1-01. Each
116 amino acid sequence of the S protein region was named the BA.1-0.1_S: amino acids of Omicron-BA.1 type
117 (Oaa) and Wuhan type (Waa) and its position number (XXX) (Ex., BA.1-0.1_S:OaaXXXWaa), as described
118 in the Methods section. And then, the putative isolates bearing the BA.1-0.1_S:OaaXXXWaa, were searched
119 using the BLAST program by their amino acid sequences. In this search, we obtained the isolates whose
120 identities is the 100% matches with the query amino acid sequence except for the SARS-CoV-
121 2_human_USA_NY-PV55373_2022(GenBank: ON246090.1) whose percent identities was 99.92%.

122 Surprisingly, we found that the Omicron BA.1-0.1 isolates were detected at all mutation sites except
123 N501Y (Fig. 2A). In the BA.1 lineage of the Omicron variant, there are Omicron isolates (BA.1.1) with the
124 R346K mutation seen in the Mu(m) variant (termed B.1.621), *i.e.*, BA.1_S can be defined as BA.1.1_S:K346R.
125 We also performed a BLAST search for isolates with amino acid sequences of BA.1-0.1.1_S:OaaXXXWaa, as
126 described in the Methods section. As a result, Omicron BA.1.1-subset-0.1 isolates were detected at all mutation
127 sites except S373P (Fig. 2B). Similar to the BA.1 lineage of the Omicron variant, in the BA.2 lineage of the
128 Omicron variant, isolates of BA.2-0.1 were found at all mutant sites except T478K and P681H in the S protein
129 (Supplemental Figure 2). The presence of these isolates refutes the establishment of Omicron strains through a

130 continuous evolutionary process by accumulating mutations. So, we could not determine which mutation
131 occurred first or last in forming the Omicron variants. As shown in Fig. 2B, over half of the BA.1.1-0.1 isolates
132 have the synonymous mutation c21595u detected in the S protein. However, this does not explain in what order
133 the c21595u mutation arose. Curiously, in the BA.1 strain isolates, this c21595u mutation was only detected in
134 SARS-CoV-2_human_USA_ID-CDC-LC0481844_2022 (GenBank: OM409228.1) and SARS-CoV-
135 2_human_USA_MI-CDC-ASC210597972_2022 (GenBank: OM396816.1). These isolates commonly lack the
136 G339D mutation. This synonymous mutation is in a mutation-prone hotspot location. Still, if the same mutation
137 occurred independently in different isolates, it is highly unnatural that the proportion of c21595u occurrences is
138 significantly biased in the Omicron variants BA.1.1-0.1.

139 It has been reported that two different variants were infected in a single cell while establishing various
140 SARS-CoV-2 variants, causing homologous recombination in the viral RNA synthesis process, resulting in
141 multiple variants. Assuming that homologous recombination caused the isolates shown in Fig. 2, some of the
142 breakpoints at which strand changes occur by homologous recombination are too short (1nt, 2nt, 3nt, etc.) (Fig.
143 3 and Supplemental Figure 3). Therefore, it is unreasonable to use homologous recombination as the basis for
144 establishing these isolates. Also, most of these isolates were found in the USA between 2021 and 2022; however,
145 considering that the major prevalent variant in the USA in August 2021 was the Delta variant, it is most unlikely
146 that it did not cause mutations such as T478K and D614G, which are already prevalent. It is inconceivable that
147 the oldest variants (such as T478K and D614G), which are no longer prevalent, were present enough to cause
148 superinfection and to have been involved in homologous recombination. Also, most of these isolates were

149 isolated in the USA between 2021 and 2022. Still, given that the isolates primarily prevalent in the USA in
150 August 2021 were Delta variants, it is unlikely that an older type of variant without the T478K or D614G
151 mutation was present to cause superinfection and involved in homologous recombination. This assumption is
152 supported by the fact that all of these BA.1-0.1 isolates and BA.1.1-0.1 isolates retained the sequence of the
153 BA.1 lineage in all regions except the S protein (Fig. 4). In addition, the fact that all of these BA.1-0.1 and
154 BA.1.1-0.1 strains retain the sequence of Omicron strain BA.1 except for the S protein also makes it
155 unreasonable to assume that these isolates arose by homologous recombination with an older type of mutant
156 without the T478K or D614G mutations (Fig. 4).

157 Furthermore, some of the BA.1 isolates and BA.1-0.1 isolates have mutation subsets (synonymous:
158 u10135c, ORF3: L106F, N: D343G) upstream and downstream of the S gene and the u10135c and L106F
159 (ORF3) mutations correspond perfectly. Therefore, it is considered homologous recombination between the
160 BA.1 variant and the variants without these mutations did not occur during the mutants' formation processes
161 (Fig. 4). The synonymous mutation c2470u occurred in BA.1.1 compared to BA.1.1, and this c2470u mutation
162 was retained by most but a few isolates of the BA.1.1-0.1 isolates (SARS-CoV-2_human_USA_IL-CDC-
163 ASC210695497_2022: GenBank: OM770362.1; SARS-CoV-2_human_USA_NY-CDC-LC0450936_2021:
164 GenBank: OM228453.1). The synonymous mutation c2470u has also only been observed in a minimal
165 number of BA.1-0.1 isolates (SARS-CoV-2_human_USA_OR-CDC-LC0470830_2022: GenBank:
166 OM367679.1; SARS-CoV-2_human_USA_ID-CDC-LC0481844_2022: GenBank: OM409228.1; SARS-
167 CoV-2_human_USA_MI-CDC-ASC210597972_2022: GenBank:OM396816.1; SARS-CoV-

168 2_human_USA_WI-CDC-LC0494047_2022: GenBank: OM500517.1). These results suggest that the
169 establishment of BA.1-0.1 isolates and BA.1.1-0.1 isolates occur independently. On the other hand, if
170 reversion mutations caused each of these isolates with one amino acid to the Wuhan-type, these isolates could
171 be detected by examining an astronomical number of isolates. However, these virus strains were detected in
172 the number of sequenced whole genome (a limited number), rather than in the astronomical numbers
173 examined. The fact that most of these mutations occurred without synonymous mutations (Fig. 2) suggests
174 that none of these mutations arose as a result of trial-and-error random mutations in nature. Few synonymous
175 mutations are detected in some BA.1-0.1, BA.1.1-0.1, and BA.2-0.1 isolates (Fig. 2 and Supplemental Figure
176 2), as seen in other viruses (Supplemental Figure 1). The c25000u is the only synonymous mutation that did
177 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 were formed and was not
178 observed in previous variants such as alpha, beta, gamma, delta, etc. Nevertheless, it is very puzzling to find
179 the occurrence of mutants with synonymous mutations such as c22120u, c24034u, c23635u, c24448u,
180 c21811u, a23884g, c22987u, c23609a, c23413u, c23896u, c22879u, u24097a, c23893u, c24442u, u24847c,
181 c24382u, c22264u, c22879u, c22480u, u21976c, c22480u, g24577a and u23101c in the BA.1.1, BA.1-0.1 and
182 BA.1.1-0.1 isolates (Fig. 2 Synonymous Others), and a22948g, c23635u, c21859u, c22945u, c23701u,
183 c22987u, a24433g, c23347u, u24640c, a24619g, c24865u, a23989g, u23047c, u24346c, c21811u, c21952u,
184 a22753u, c23635u, c24023u, c24382u and c22572u in the BA.2-0.1 isolates (Supplemental Figure 2
185 Synonymous Others) after the formation of mutants with these subsets.

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

190 Therefore, we also performed BLAST searches for isolates with mutant amino acid subsets of both the
191 Omicron variants' BA.1 and BA.2 strains. We detected recombinant isolates of Omicron BA.1 and BA.2
192 lineages. Surprisingly, the recombinant of Omicron BA.1 and BA.2 lineages, SARS-CoV-2/human/PRI/PR-
193 PR-UPRRP-582/2020 (GenBank: ON928946.1), was already present in Puerto Rico in 2020. Omicron
194 (B.1.1.529) is a variant of SARS-CoV-2 first reported to the WHO by the Network for Genomics Surveillance
195 in South Africa on November 24, 2021 (13, 14). It was first detected in Botswana and has spread to become the
196 predominant variant in circulation worldwide (15). Following the appearance of the original B.1.1.529 variant,
197 several subvariants of Omicron have emerged, including BA.1, BA.2, BA.3, BA.4, and BA.5 (19). Since
198 October 2022, two subvariants of BA.5 called BQ.1 and BQ.1.1 have emerged.

199 The question then arose about why a recombinant strain, SARS-CoV-2/human/PRI/PR-UPRRP-
200 582/2020 (GenBank: ON928946.1), already existed in 2020. We searched for SARS-CoV-2 isolates prevalent
201 in Puerto Rico using the keywords "PRI", "PR-UPRRP", and "2020". Consequently, we found 29 Omicron-
202 associated sequences in the 127 hits obtained (Fig. 5B). These results suggest that the SARS-CoV-2 lineages
203 bearing the amino acid sequences of the S protein are identical to that of Omicron BA.1 and Omicron BA.2
204 lineages were already prevalent in Puerto Rico in 2020, with 15 isolates having the complete Omicron BA.1+

205 R346K_mut-subset (BA1.1) , 14 isolates had a synonymous substitution of c21595u. Four isolates had an amino
206 acid sequence of the S protein perfectly matched that of Omicron BA2 (BA.2_S), and eight were Omicron
207 BA.2-0.1 (BA.2-S:K440N).

208

209 **Discussion**

210 Several hypotheses have been proposed that the original SARS-CoV-2 virus resulted from an accidental
211 laboratory spill. With recent developments in biotechnology, many viruses, including coronaviruses, have been
212 artificially synthesized and used in various experiments (20-22). The artificial generation of mutant viruses in
213 the laboratory and the study of viral phenotypes by introducing mutations is called "reverse genetics" and is a
214 common technique in virology. It has been claimed that SARS-CoV-2 is artificially generated because of the
215 unnatural presence of a codon (CGG) encoding a contiguous arginine at the furin cleavage site of SARS-CoV-
216 2. This claim is refuted based on the following facts; 1) there is no logical reason for a genetically engineered
217 virus to utilize such a suboptimal furin cleavage site. 2) The only previous study on artificial insertion of furin
218 cleavage sites at the S1/S2 boundary of the S protein of SARS-CoV-1 using the pseudotype virus experimental
219 system utilized the optimal "RRSRR" sequence, which is different from the furin cleavage site's sequence
220 present in SARS-CoV-2. 3) There is also no evidence of previous studies at the Wuhan Institute of Virology
221 that artificially inserted a complete furin cleavage site in coronaviruses. 4) Unnatural CGG codons adjacent to
222 the arginine at the furin cleavage site are rare in coronaviruses but are observed at a particular frequency in
223 SARS-CoV-1, SARS-CoV-2, and other human coronaviruses. However, these are only declarations and are

224 illogical. No one explained why a naturally occurring virus would utilize a suboptimal furin cleavage site. It has
225 not been mentioned about technically possible to insert this furin cleavage site or a CGG codon artificially.
226 Inserting a polybasic furin cleavage site on the S protein makes it impossible to conclude whether SARS-CoV-
227 2 is a naturally occurring or artificial virus by discussing whether it is natural or artificial.

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

242 Assuming that artificially synthesized mutants with only non-synonymous mutations are spread globally,
243 this would explain how mutants with non-synonymous mutations without previous synonymous mutations
244 develop synonymous mutations under natural circumstances. Considering the current epidemic situation of
245 SARS-CoV-2, it is unlikely that these viruses arose spontaneously. In explaining the formation of the SARS-
246 CoV-2 isolate, as shown here, the SARS-CoV-2 isolates are formed by a completely new mechanism that cannot
247 be explained by previous biology.

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

251 It has been known that the pathogenicity, host specificity, cell tropism, and immunogenicity of numerous
252 viruses can be altered by mutation of a single (or several) amino acid(s) of a viral protein on the viral envelope
253 (envelope protein, HA protein, spike protein, etc.). A single-amino-acid substitution in the HA protein of the
254 2009 pandemic A (H1N1) influenza viruses changes their replication and pathogenicity (23). In the
255 Chikungunya virus, single amino acid changes in the E2 glycoprotein influence glycosaminoglycan utilization
256 for target cell binding (24) , and single amino acid change in the E1 glycoprotein affects mosquito vector
257 specificity and epidemic potential (25). In previous coronaviruses such as MERS-CoV and SARS-CoV-1, point
258 mutations have been demonstrated to confer resistance to neutralizing antibodies (26-28).

259 Suppose the SARS-CoV-2 Omicron variant and its one amino acid reversion mutants were artificially
260 and systematically generated. In that case, we should suspect that the other variants (Alpha to Delta) may also

261 be artificially generated viruses. Indeed, the lack of findings to date that many of the various mutations seen,
262 especially in the early variants, are indeed associated with increased viral infection (29) supports the hypothesis
263 that each variant was artificially synthesized to identify the amino acids of the S protein responsible for
264 infectivity and pathogenicity. The possibility that the set of mutants was artificially generated to identify amino
265 acids of the S protein involved in the infectivity and virulence is supported.

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

270 Finally, we would like to add that while artificially synthesized viruses may have spread, we are not
271 criticizing reverse genetics technology, as reverse genetics technology has made tremendous advances in
272 virology. In addition, our analysis uses databases with a limited number of viral sequences, and we cannot deny
273 the possibility that unreliable data may have been registered due to technical problems in sequencing or some
274 malicious intent. Furthermore, we do not conclude that these viruses were artificially synthesized and distributed
275 based on malicious intent. This paper aims to point out that SARS-CoV-2 has undergone unthinkable mutations
276 under conventional coronavirus mutation mechanisms, and we hope that the possibility of artificial theory
277 should be included in the discussion seriously as to the formation of SARS-CoV-2 variants.

278 Nonetheless, the analysis we have shown here concludes that the Omicron variants are formed by a
279 completely new mechanism that cannot be explained by previous biology. The way how the SARS-CoV-2
280 mutations occurred should prompt a reconsideration of the SARS-CoV-2 pandemic.

281 **Methods**

282 **Data acquisition**

283 SARS-CoV-2 RNA genome and genes and proteins according to an annotation of the SARS-CoV-2
284 Wuhan-Hu-H1 (COVID-19/Wuhan-Hu-1CHN/2019/First Isolate) NCBI Reference Sequence: NC_045512.2
285 were used as a reference strain for the definition of mutations and provided base of the numbering of
286 nucleotide and amino acids number of each protein. Genome data of SARS-CoV-2 isolates which showed in
287 this article were obtained by coping from NCBI Nucleotide database (<https://www.ncbi.nlm.nih.gov/>) on
288 25/11/2022 to 17/03/2023.

289 **Query of representative SARS-CoV-2 variant genome**

290 Amino acids sequences of spike protein of SARS-CoV-2 variants (Alpha:B.1.1.7, Beta:B.1.351,
291 Gamma:P1, Delta:B.1.617.2.63, Lambda:C.37, Mu:B.1.621, Omicron:BA.1, BA.1.1 and BA.2) were
292 obtained from internet site, Stanford Coronavirus Antiviral & Resistance Database
293 (<https://covdb.stanford.edu/>) or Covariant (<https://covariants.org/>) and used as a query sequence for NCBI
294 protein BLAST search (blastp: protein-protein BLAST,
295 https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasth
296 ome). Then the whole genome sequence of each variant isolates bearing the query spike sequence were
297 derived from the BLAST search resultant which identified with query amino acids sequence with 100%. And
298 then the nucleotide sequences of detected SARS-CoV-2 genome were as follows, GenBank Accession No.:
299 GenBank: MW423686.2; MW430966.1; MW430967.1; MW422256.1; MW598419.1;

300 MW667552.1; MW667553.1; MW721502.1; MW721504.1; MW520923.1; MW642248.1;
301 MW642249.1; MW642250.1; MZ182066.1; MZ155303.1; MZ155230.1; MZ170364.1;
302 MZ179869.1; MW666666.1; MW696612.1; MW699217.1; MW644498.1; MZ727706.1;
303 MZ620161.1; MZ637380.1; MZ637401.1; MZ780550.1; OL672836.1; OL677199.1; OP769083.1;
304 OL764360.1; OL815447.1; ON762438.1; OL849989.1; OL897126.1; OL896945.1; OL896936.1;
305 OL896931.1; OM233931.1; OM072551.1; OM072822.1; OM296922.1.

306

307 **Query of SARS-CoV-2 Omicron variant genome which bearing a S protein amino acids sequence in**
308 **which one of the Omicron-type nucleotide mutation subsets is not mutated and remains the original**
309 **SARS-CoV-2 Wuhan-Hu-H1-type arrangement.** For each of the Omicron variants, BA.1, BA.1.1 and BA.2,
310 the isolates series which bearing a S protein amino acids sequence in which one of the Omicron-type nucleotide
311 mutation subsets is not mutated and remains the original SARS-CoV-2 Wuhan-Hu-H1-type arrangement were
312 named with BA.1-0.1, BA.1.1-0.1 and BA.2-0.1, respectively. In addition, in this article, we named the amino
313 acid sequences of spike protein of BA.1, BA.1.1 and BA.2 as the BA.1_S, BA.1.1_S and BA.2_S, respectively,
314 and then the series of amino acid sequences of spike protein of BA.1-0.1, BA.1.1-0.1 and BA.2-0.1 were named,
315 respectively, as follows; Each Omicron BA.1-0.1 spike series (BA.1-0.1_Ss) were named as BA.1_S:V67A;
316 BA.1_S:69H_70V; BA.1_S:I95T; BA.1_S:D142G_143V_144Y_145Y; BA.1_S:I211N_212L; BA.1_S:ΔEPE;
317 BA.1_S:D339G; BA.1_S:L371S; BA.1_S:P373S; BA.1_S:F375S; BA.1_S:N417K; BA.1_S:K440N;
318 BA.1_S:S446G; BA.1_S:N477S; BA.1_S:K478T; BA.1_S:A484E; BA.1_S:R493Q; BA.1_S:S496G;

319 BA.1_S:R498Q; BA.1_S:Y501N; BA.1_S:H505Y; BA.1_S:K547T; BA.1_S:G614D; BA.1_S:Y655H;
320 BA.1_S:K679N; BA.1_S:H681P; BA.1_S:K764N; BA.1_S:Y796D; BA.1_S:K856N; BA.1_S:H954Q;
321 BA.1_S:K969N and BA.1_S:F981L / Omicron BA.1.1-0.1 spike series (BA.1.1-0.1_Ss) were named as
322 BA.1.1_S:V67A; BA.1.1_S:69H_70V; BA.1.1_S:I95T; BA.1.1_S:D142G_143V_144Y_145Y;
323 BA.1.1_S:I211N_212L; BA.1.1_S:ΔEPE; BA.1.1_S:D339G; BA.1.1_S:L371S; BA.1.1_S:P373S;
324 BA.1.1_S:F375S; BA.1.1_S:N417K; BA.1.1_S:K440N; BA.1.1_S:S446G; BA.1.1_S:N477S;
325 BA.1.1_S:K478T; BA.1.1_S:A484E; BA.1.1_S:R493Q; BA.1.1_S:S496G; BA.1.1_S:R498Q;
326 BA.1.1_S:Y501N; BA.1.1_S:H505Y; BA.1.1_S:K547T; BA.1.1_S:G614D; BA.1.1_S:Y655H;
327 BA.1.1_S:K679N; BA.1.1_S:H681P; BA.1.1_S:K764N; BA.1.1_S:Y796D; BA.1.1_S:K856N;
328 BA.1.1_S:H954Q; BA.1.1_S:K969N; BA.1.1_S:F981L / Omicron BA.2-0.1 spike series (BA.2-0.1_Ss) were
329 named as BA.2_S:I19T; BA.2_S:24L_25P_26P_S27A; BA.2_S:D142G; BA.2_S:V213G; BA.2_S:D339G;
330 BA.2_S:F371S; BA.2_S:P373S; BA.2_S:F375S; BA.2_S:A376T; BA.2_S:N405D; BA.2_S:S408R;
331 BA.2_S:N417K; BA.2_S:K440N; BA.2_S:N477S; BA.2_S:K478T; BA.2_S:A484E; BA.2_S:R493Q;
332 BA.2_S:R498Q; BA.2_S:Y501N; BA.2_S:H505Y; BA.2_S:G614D; BA.2_S:Y655H; BA.2_S:K679N;
333 BA.2_S:H681P; BA.2_S:K764N; BA.2_S:Y796D; BA.2_S:H954Q; BA.2_S:K969N, and these constructs were
334 showed in Figs. 2, 4 and supplemental Fig 1. These amino acids sequences of spike protein of SARS-CoV-2
335 Omicron variants, BA.1-0.1, BA.1.1-0.1 and BA.2-0.1 were used as a query sequence for NCBI protein BLAST
336 search. Then the whole genome sequence of BA.1-0.1, BA.1.1-0.1 and BA.2-0.1 isolates bearing the query
337 spike sequence were derived from the BLAST search resultant which identified with query amino acids

338 sequence with 100%. And then the nucleotide sequences of detected SARS-CoV-2 genome were as follows,

339 GenBank Accession No.: OM117411.1; OP797378.1; OM789835.1; OP928789.1; OP928803.1;

340 OP929381.1; OP929396.1; OP929417.1; OM173977.1; OM518459.1; OM566981.1;

341 ON019560.1; OM097227.1; OM096937.1; OM099902.1; OM117114.1; OM096685.1;

342 OM354436.1; OM646886.1; OM472901.1; OM364511.1; OM131858.1; OL815451.1;

343 OL896986.1; OL897116.1; OL897118.1; OL896964.1; OM367679.1; OM343778.1;

344 OM409228.1; OM396816.1; OM134162.1; OM075886.1; OM123427.1; OM122677.1;

345 OM121681.1; OM224850.1; ON246090.1; OM931599.1; OM864873.1; OM906519.1;

346 OM906587.1; OM464776.1; OM015999.1; OM015958.1; OM015597.1; OM016329.1;

347 OL898806.1; OL898861.1; OM016937.1; OM016186.1; OM036549.1; OM051171.1;

348 OM126493.1; OM079115.1; OM099199.1; OM134489.1; OM098796.1; ON618279.1;

349 ON618009.1; OM627701.1; OM356511.1; OM295457.1; ON700063.1; OM033824.1;

350 ON368355.1; OM084700.1; ON208126.1; OM566593.1; OM945690.2; ON030252.1;

351 ON019844.1; OM890075.1; ON020044.1; OM833954.1; ON376082.1; OM084604.1;

352 OP795273.1; ON066609.1; OM352882.1; OM290510.1; OM369978.1; OM199342.1;

353 OM011974.1; OM090274.1; OM043984.1; OM121683.1; OM121624.1; OM175506.1;

354 OM360429.1; OM360221.1; OM358058.1; OM500517.1; OM135027.1; OM742858.1;

355 OM521685.1; OM896558.1; ON694155.1; OM686755.1; OM484260.1; OM332056.1;

356 OM156397.1; OM079447.1; OM134645.1; OM173298.1; OM123082.1; OM116023.1;

357	OM652943.1;	OL994299.1;	OL994920.1;	OM122027.1;	OM121015.1;	OL898817.1;
358	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;
360	OM816604.1;	ON235452.1;	ON334146.1;	OP024162.1;	OP209732.1;	OM354578.1;
361	OM099080.1;	OM297301.1;	OM297438.1;	OM365368.1;	OM449159.1;	OM078863.1;
362	OM096959.1;	OM117155.1;	OM133880.1;	OM077358.1;	OM372005.1;	OM770362.1;
363	OM897488.1;	OM918459.1;	OM918478.1;	OL897115.1;	OL897114.1;	OL986779.1;
364	OL986696.1;	OL987046.1;	ON831866.1;	OM864099.1;	OM863888.1;	OP745925.1;
365	ON831672.1;	OM043643.1;	OM176192.1;	OM226685.1;	OM343689.1;	OM295527.1;
366	OM894975.1;	OM846676.1;	OM822024.1;	OM846844.1;	OM906550.1;	OM015933.1;
367	OM016323.1;	OM016331.1;	OM035685.1;	OM022498.1;	OM156115.1;	OM036875.1;
368	OM099560.1;	OM199246.1;	OM067048.1;	OM079299.1;	OM099911.1;	OM116588.1;
369	OM097010.1;	OM173300.1;	OM805961.1;	OM983266.1;	OM983325.1;	ON618010.1;
370	OM084691.1;	ON021265.1;	ON039239.1;	ON056981.1;	ON144127.1;	OM770527.1;
371	OM156164.1;	OM155119.1;	OM199353.1;	OM084630.1;	OM084605.1;	OM084621.1;
372	OM359369.1;	OM411574.1;	OM584789.1;	OM720486.1;	OM429777.1;	ON047062.1;
373	ON065416.1;	OP415118.1;	OM954373.1;	ON042406.1;	OM335528.1;	OM332335.1;
374	OM353626.1;	OM332813.1;	OM197398.1;	OM226919.1;	OM228399.1;	OM225859.1;
375	OM271353.1;	OM159454.1;	OM224473.1;	OM358278.1;	OM361030.1;	OM412141.1;

376	OM496298.1;	OM044048.1;	OM121864.1;	OM224477.1;	OM227379.1;	OM228453.1;
377	OM622156.1;	OM906370.1;	OM970683.1;	ON117965.1;	OM198667.1;	OM357800.1;
378	OM357161.1;	OM335230.1;	OM261124.1;	OM077578.1;	OM497172.1;	OM625194.1;
379	OM907131.1;	ON047464.1;	OM911851.1;	OM042846.1;	OM155337.1;	OM097339.1;
380	OM116805.1;	OM134409.1;	OM686782.1;	OM695863.1;	OM724725.1;	OM174366.1;
381	OM822132.1;	OM822106.1;	OM822105.1;	OM822485.1;	OM135143.1;	OM125829.1;
382	OM098855.1;	OM156118.1;	OM155114.1;	OM863926.1;	OP359104.1;	ON209298.1;
383	ON232806.1;	ON421981.1;	ON811217.1;	OM698275.1;	ON052769.1;	ON060006.1;
384	ON060007.1;	ON060009.1;	OM843171.1;	OM843276.1;	OM843550.1;	OM843316.1;
385	OM843340.1;	ON049267.1;	ON450720.1;	ON250163.1;	ON256603.1;	ON480422.1;
386	OM888844.1;	OM890089.1;	ON009425.2;	ON082904.1;	OM901275.1;	OM877094.2;
387	OM877095.2;	OM877096.2;	OM877097.2;	ON378542.1;	ON389858.1;	ON389889.1;
388	ON390359.1;	OM936703.1;	ON352711.1;	ON378000.1;	ON177702.1;	ON205494.1;
389	ON378633.1;	ON617689.1;	ON619375.1;	OM567618.1;	OM659585.1;	OM770913.1;
390	OM781641.1;	OM533441.1;	OM533458.1;	OM570235.1;	OM570252.1;	OM570249.1;
391	OM283361.1;	OM283362.1;	OM283320.1;	OM283343.1;	ON618014.1;	ON618018.1;
392	ON618019.1;	ON618363.1;	ON311615.1;	ON383919.1;	OP579158.1;	OP054411.1;
393	ON633107.1;	ON414693.1;	ON422887.1;	OP364296.1;	OP629673.1;	ON363097.1;
394	OP633561.1;	ON458445.1;	ON592247.1;	ON549687.1;	ON067040.1;	ON321116.1;

395 ON199452.1; ON200331.1; OM861064.1; OM969592.1; ON019120.1; ON221861.1;
396 OM861619.1; ON091288.1; ON151370.1; ON233850.1; ON236456.1; ON296711.1;
397 ON535443.1; ON624524.1; ON377450.1; ON397268.1; ON239032.1; ON373649.1;
398 ON481637.1; ON701163.1; ON312677.1; ON349263.1; ON377487.1; ON377609.1;
399 OM638574.1; OM911616.1; OM988767.1; ON019770.1; OM988769.1; ON468158.1;
400 ON608924.1; ON604965.1; ON535763.1; ON378227.1; ON378238.1; ON728470.1.

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

402 Deduced recombinant spike protein between Omicron variants, BA.1 and BA.2 which showed as the
403 BA.1_S:T19I_L24-_P25-_P26-_A27S_V213G_S371F_T376A_D405N_R408S was used as a query sequence
404 for NCBI protein BLAST search. Then the whole genome sequence of BA.1 and BA.2 recombinant-Omicron
405 isolates bearing some of specific amino acids mutation observed in variant BA.1 and BA.2 in the spike protein
406 region. And then the nucleotide sequences of detected SARS-CoV-2 genome were as follows, GenBank

407 Accession No.: OM360636.1; OM410816.1; OM429902.1; OM497964.1; OM565587.1;
408 OM628132.1; ON549899.1; ON449685.1; ON176765.1; OM628094.1; ON099844.1;
409 OM942313.1; ON395480.1; ON171854.1; ON172005.1; ON076710.1; ON928946.1;
410 OM932113.1; OM942438.1; OM989528.1; OM499181.1; ON414822.1; OM878325.1;
411 ON103067.1; ON103153.1; ON419036.1; ON928719.1; ON337887.1; ON420444.1;
412 ON146520.1; OM469541.1; OM904085.1; ON254531.1; OM881098.1; ON373310.1.

413

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

415 Nucleotide sequences were searched using the keyword PRI PR-UPRRP 2020 (Search details: PRI[All Fields]

416 AND (PR[All Fields] AND UPRRP[All Fields]) AND 2020[All Fields]). And then the search resultants were

417 all SARS-CoV-2 isolate genome sequences. Among these sequences, SARS-CoV-2 Omicron variant related

418 sequences were picked up as follows, GenBank Accession No.: ON928761.1; ON928660.1;

419 ON928794.1; ON928762.1; ON928848.1; ON928741.1; ON928918.1; ON928680.1;

420 ON928975.1; ON928949.1; ON928673.1; ON928865.1; ON928716.1; ON928663.1;

421 ON928779.1; ON928896.1; ON928946.1; ON928912.1; ON928704.1; ON928873.1;

422 ON928813.1; ON928898.1; ON928765.1; ON928912.1; ON928883.1; ON928957.1;

423 ON928880.1; ON928699.1; ON928724.1; ON928941.1.

424

425 The genomes were aligned using SnapGene software. Numbering of nucleotide and amino acids number of

426 each protein are provided by using Wuhan-Hu-1 (NC_045512.2; COVID-19/Wuhan-Hu-1CHN/2019/First

427 Isolate) as a reference strain for the definition of mutations.

428

429

430 **References**

431

- 432 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,
433 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
434 associated with human respiratory disease in China. *Nature*, 579(7798), 265-269 (2020) doi:10.1038/s41586-020-2008-3
- 435 2. R. A. Khailany, M. Safdar and M. Ozaslan: Genomic characterization of a novel SARS-CoV-2. *Gene Rep*, 19,
436 100682 (2020) doi:10.1016/j.genrep.2020.100682
- 437 3. Y. M. Bar-On, A. Flamholz, R. Phillips and R. Milo: SARS-CoV-2 (COVID-19) by the numbers. *Elife*, 9 (2020)
438 doi:10.7554/eLife.57309
- 439 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
440 the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215-220 (2020)
441 doi:10.1038/s41586-020-2180-5
- 442 5. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham and J. S. McLellan:
443 Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260-1263 (2020)
444 doi:10.1126/science.abb2507
- 445 6. B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T.
446 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.
447 Sheffield, C. McDanal, L. G. Perez, H. Tang, A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Saphire and D. C.
448 Montefiori: Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus.
449 *Cell*, 182(4), 812-827 e19 (2020) doi:10.1016/j.cell.2020.06.043
- 450 7. A. Carrasco-Montalvo, A. Bruno, D. de Mora, M. Olmedo, J. Garces, M. Paez, M. Regato-Arrata, M. Gonzalez,
451 J. Romero, O. Mestanza, B. Freire-Paspuel, A. Gaviria, S. A. Orlando, M. A. Garcia-Bereguian and L. Patino: First Report
452 of SARS-CoV-2 Lineage B.1.1.7 (Alpha Variant) in Ecuador, January 2021. *Infect Drug Resist*, 14, 5183-5188 (2021)
453 doi:10.2147/IDR.S319439
- 454 8. S. Zarate, B. Taboada, J. E. Munoz-Medina, P. Isa, A. Sanchez-Flores, C. Boukadida, A. Herrera-Estrella, N.
455 Selem Mojica, M. Rosales-Rivera, B. Gomez-Gil, A. G. Salas-Lais, C. E. Santacruz-Tinoco, H. Montoya-Fuentes, J. E.
456 Alvarado-Yaah, G. M. Molina-Salinas, G. E. Espinoza-Ayala, J. A. Enciso-Moreno, R. M. Gutierrez-Rios, A. Loza, J.
457 Moreno-Contreras, R. Garcia-Lopez, X. Rivera-Gutierrez, A. Comas-Garcia, R. M. Wong-Chew, M. E. Jimenez-Corona,
458 R. M. Del Angel, J. A. Vazquez-Perez, M. Matias-Florentino, M. Perez-Garcia, S. Avila-Rios, H. G. Castelan-Sanchez, L.
459 Delaye, L. P. Martinez-Castilla, M. Escalera-Zamudio, S. Lopez and C. F. Arias: The Alpha Variant (B.1.1.7) of SARS-
460 CoV-2 Failed to Become Dominant in Mexico. *Microbiol Spectr*, 10(2), e0224021 (2022) doi:10.1128/spectrum.02240-21
- 461 9. H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E. J.
462 San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A. J. Glass, S. Engelbrecht, G.
463 Van Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, N.-y. Hsiao, S. Korsman, M.-A. Davies, L. Tyers, I.
464 Mudau, D. York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A. Alisoltani-Dehkordi, A. Godzik, C. K.
465 Wibmer, B. T. Sewell, J. Lourenço, L. C. J. Alcantara, S. L. Kosakovsky Pond, S. Weaver, D. Martin, R. J. Lessells, J. N.

466 Bhiman, C. Williamson and T. de Oliveira: Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*,
467 592(7854), 438-443 (2021) doi:10.1038/s41586-021-03402-9

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

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

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

479 13. J. Quarleri, V. Galvan and M. V. Delpino: Omicron variant of the SARS-CoV-2: a quest to define the
480 consequences of its high mutational load. *Geroscience*, 44(1), 53-56 (2022) doi:10.1007/s11357-021-00500-4

481 14. A. Gowrisankar, T. M. C. Priyanka and S. Banerjee: Omicron: a mysterious variant of concern. *Eur Phys J Plus*,
482 137(1), 100 (2022) doi:10.1140/epjp/s13360-021-02321-y

483 15. A. Vitiello, F. Ferrara, A. M. Auti, M. Di Domenico and M. Boccellino: Advances in the Omicron variant
484 development. *J Intern Med*, 292(1), 81-90 (2022) doi:10.1111/joim.13478

485 16. Y. Fan, X. Li, L. Zhang, S. Wan, L. Zhang and F. Zhou: SARS-CoV-2 Omicron variant: recent progress and
486 future perspectives. *Signal Transduct Target Ther*, 7(1), 141 (2022) doi:10.1038/s41392-022-00997-x

487 17. J. Fonager, M. Bennedbaek, P. Bager, J. Wohlfahrt, K. M. Ellegaard, A. C. Ingham, S. M. Edslev, M. Stegger,
488 R. N. Sieber, R. Lassauniere, A. Fomsgaard, T. Lillebaek, C. W. Svarrer, F. T. Moller, C. H. Moller, R. Legarth, T. V.
489 Sydenham, K. Steinke, S. J. Paulsen, J. A. S. Castruita, U. V. Schneider, C. H. Schouw, X. C. Nielsen, M. Overvad, R. T.
490 Nielsen, R. L. Marvig, M. S. Pedersen, L. Nielsen, L. L. Nilsson, J. Bybjerg-Grauholm, I. H. Tarpgaard, T. S. Ebsen, J. U.
491 H. Lam, V. Gunalan and M. Rasmussen: Molecular epidemiology of the SARS-CoV-2 variant Omicron BA.2 sub-lineage
492 in Denmark, 29 November 2021 to 2 January 2022. *Euro Surveill*, 27(10) (2022) doi:10.2807/1560-
493 7917.ES.2022.27.10.2200181

494 18. L. B. Shrestha, C. Foster, W. Rawlinson, N. Tedla and R. A. Bull: Evolution of the SARS-CoV-2 omicron
495 variants BA.1 to BA.5: Implications for immune escape and transmission. *Rev Med Virol*, 32(5), e2381 (2022)
496 doi:10.1002/rmv.2381

497 19. 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 and M. J.
498 Ma: Omicron subvariants escape antibodies elicited by vaccination and BA.2.2 infection. *Lancet Infect Dis*, 22(8), 1116-
499 1117 (2022) doi:10.1016/S1473-3099(22)00410-8

500 20. S. Torii, C. Ono, R. Suzuki, Y. Morioka, I. Anzai, Y. Fauzyah, Y. Maeda, W. Kamitani, T. Fukuhara and Y.
501 Matsuura: Establishment of a reverse genetics system for SARS-CoV-2 using circular polymerase extension reaction. *Cell*
502 *Rep*, 35(3), 109014 (2021) doi:10.1016/j.celrep.2021.109014

- 503 21. T. Y. Taha, I. P. Chen, J. M. Hayashi, T. Tabata, K. Walcott, G. R. Kimmerly, A. M. Syed, A. Ciling, R. K.
504 Suryawanshi, H. S. Martin, B. H. Bach, C. L. Tsou, M. Montano, M. M. Khalid, B. K. Sreekumar, G. Renuka Kumar, S.
505 Wyman, J. A. Doudna and M. Ott: Rapid assembly of SARS-CoV-2 genomes reveals attenuation of the Omicron BA.1
506 variant through NSP6. *Nat Commun*, 14(1), 2308 (2023) doi:10.1038/s41467-023-37787-0
- 507 22. W. Wang, X. Peng, Y. Jin, J. A. Pan and D. Guo: Reverse genetics systems for SARS-CoV-2. *J Med Virol*, 94(7),
508 3017-3031 (2022) doi:10.1002/jmv.27738
- 509 23. 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
510 substitution in the HA protein changes the replication and pathogenicity of the 2009 pandemic A (H1N1) influenza viruses
511 in vitro and in vivo. *Virol J*, 7, 325 (2010) doi:10.1186/1743-422X-7-325
- 512 24. L. A. Silva, S. Khomandiak, A. W. Ashbrook, R. Weller, M. T. Heise, T. E. Morrison and T. S. Dermody: A
513 single-amino-acid polymorphism in Chikungunya virus E2 glycoprotein influences glycosaminoglycan utilization. *J Virol*,
514 88(5), 2385-97 (2014) doi:10.1128/JVI.03116-13
- 515 25. K. A. Tsetsarkin, D. L. Vanlandingham, C. E. McGee and S. Higgs: A single mutation in chikungunya virus
516 affects vector specificity and epidemic potential. *PLoS Pathog*, 3(12), e201 (2007) doi:10.1371/journal.ppat.0030201
- 517 26. X. C. Tang, S. S. Agnihothram, Y. Jiao, J. Stanhope, R. L. Graham, E. C. Peterson, Y. Avnir, A. S. Tallarico, J.
518 Sheehan, Q. Zhu, R. S. Baric and W. A. Marasco: Identification of human neutralizing antibodies against MERS-CoV and
519 their role in virus adaptive evolution. *Proc Natl Acad Sci U S A*, 111(19), E2018-26 (2014) doi:10.1073/pnas.1402074111
- 520 27. J. Sui, D. R. Aird, A. Tamin, A. Murakami, M. Yan, A. Yammanuru, H. Jing, B. Kan, X. Liu, Q. Zhu, Q. A.
521 Yuan, G. P. Adams, W. J. Bellini, J. Xu, L. J. Anderson and W. A. Marasco: Broadening of neutralization activity to
522 directly block a dominant antibody-driven SARS-coronavirus evolution pathway. *PLoS Pathog*, 4(11), e1000197 (2008)
523 doi:10.1371/journal.ppat.1000197
- 524 28. J. ter Meulen, E. N. van den Brink, L. L. Poon, W. E. Marissen, C. S. Leung, F. Cox, C. Y. Cheung, A. Q.
525 Bakker, J. A. Bogaards, E. van Deventer, W. Preiser, H. W. Doerr, V. T. Chow, J. de Kruif, J. S. Peiris and J. Goudsmit:
526 Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS Med*,
527 3(7), e237 (2006) doi:10.1371/journal.pmed.0030237
- 528 29. L. van Dorp, D. Richard, C. C. S. Tan, L. P. Shaw, M. Acman and F. Balloux: No evidence for increased
529 transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun*, 11(1), 5986 (2020) doi:10.1038/s41467-020-
530 19818-2

531
532
533

534 **Competing interests**

535 The authors declare no competing interests.

536 **Figure legends**

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

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

547

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

549 (A) Different amino acids and synonymous change of nucleotides in S proteins of SARS-CoV-2 Omicron BA.1,
550 BA.1.1 isolates and BA.1-0.1s compared to that of SARS-CoV-2 Wuhan-Hu-H1. Nucleotide deletions and
551 insertions those were "deletion¹" (deletion: nt 21,766-21,771), "deletion²" (deletion: nt 21,987-21,995),
552 "deletion³" (deletion: nt 22,194-22,196), and "insertion⁴" (insertion between 22,206-22,196), introduced the
553 amino acids changes H69_V70-, G142D_V143_Y144_Y145-, N211I_L212- and 215ins.EPE, respectively.
554 (B) Different amino acids and synonymous nucleotide change in S proteins in SARS-CoV-2 Omicron BA.1.1-

555 0.1 isolates. Amino acids different from Wuhan-Hu-H1 found in each variant, Alpha: B.1.1.7, Beta: B.1.351,
556 Gamma: P1, Delta: B.1.617.2.63, Mu: B.1.621, Omicron: BA.1, BA.2 are highlighted with red, orange, green,
557 yellow, lime, deep sky blue, and blue violet, respectively. Amino acid changes common to Omicron:BA.1 and
558 BA.2 are highlighted with purple.

559

560 **Fig. 3. Estimated homologous recombination breakpoints of the SARS-CoV-2 S gene of Omicron BA.1-**
561 **0.1 or BA.1.1-0.1**

562 Sequence alignment of the amino acids and their coding nucleotides (nt.21,746-21,787; nt.22,658-22,702;
563 nt.22,976-23,011 and nt.23,582-23,620) containing the mutation point of SARS-CoV-2 S gene of Omicron
564 BA.1 variant compared to that of SARS-CoV-2 Wuhan-Hu-H1. Changed nucleotides and amino acids of
565 Omicron BA.1 are shown in red letters. Estimated homologous recombination breakpoints of the SARS-CoV-
566 2 S gene of Omicron BA.1-0.1 or BA.1.1-0.1 are shown by asterisks.

567

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

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

574 Amino acid changes common to Omicron: BA.1 and BA.2 are shown in purple. Synonymous nucleotide
575 changes, c2470u observed in Omicron:BA.1.1 mainly shown with blue. Synonymous and non-synonymous
576 changes, u10135c of nsp5, L106F in ORF3, and D343G in N protein that subset observed ~40% of Omicron:
577 BA.1-0.1 are highlighted with emerald-green. Undetermined nucleotide(s) or amino acid(s) are shown as UD
578 or X, respectively.

579

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

582 **(A)** Different amino acids and synonymous change of nucleotides in S proteins of SARS-CoV-2 Omicron BA.1-
583 BA.2 recombinant isolates compared to that of SARS-CoV-2 Wuhan-Hu-H1. Nucleotide deletions, "deletion⁵"
584 (deletion: nt 21,633-21,641), introduced the amino acids changes L24-_P25-_P26-_A27S. **(B)** Different amino
585 acids and synonymous change of nucleotides in S proteins of SARS-CoV-2 Omicron BA.1.1, Omicron BA.1-
586 BA.2 recombinant isolate, which is highlighted with magenta (GenBank: ON928946.1), Omicron BA.2 and
587 Omicron 2-0.1(K440N) those detected in Puerto Rico in 2020. Amino acids different from Wuhan-Hu-H1 found
588 in each variant, Alpha: B.1.1.7, Beta: B.1.351, Gamma: P1, Delta: B.1.617.2.63, Mu: B.1.621, Omicron: BA.1,
589 BA.2 are highlighted with red, orange, green, yellow, lime, deep sky blue, and blue-violet, respectively. Amino
590 acid changes common to Omicron: BA.1 and BA.2 are highlighted with purple.

591 **Supplemental Figure 1**

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

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

598

599 **Supplemental Figure 2**

600 **Different amino acids and synonymous change of nucleotides in S proteins of SARS-CoV-2 Omicron**

601 **BA.2 isolates and BA.2-0.1s those compared to that of SARS-CoV-2 Wuhan-Hu-H1.**

602 Nucleotide deletions, "deletion⁵" (deletion: nt 21,633-21,641), introduced the amino acids changes L24-
603 _P25-
604 _P26-
605 _A27S. Amino acids different from Wuhan-Hu-H1 found in each variant, Alpha: B.1.1.7, Beta: B.1.351,
606 Gamma: P1, Delta: B.1.617.2.63, Mu: B.1.621, Omicron: BA.1, BA.2 are highlighted with red, orange, green,
607 yellow, lime, deep sky blue, and blue-violet, respectively. Amino acid changes common to Omicron: BA.1
608 and BA.2 are highlighted with purple.

607

608 **Supplemental Figure 3 Estimated homologous recombination breakpoints of the SARS-CoV-2 S gene**

609 **of Omicron BA.2-0.1 or BA.1-BA.2 recombinant.**

610 (A) Sequence alignment of the amino acids and its coding nucleotides (nt. 22,658-22,702) containing the
611 mutation point of SARS-CoV-2 S gene of Omicron BA.2 variants compared to that of SARS-CoV-2 Wuhan-
612 Hu-H1. (B) Sequence alignment of the amino acids and its coding nucleotides (nt. 22,178-22,222) containing
613 the mutation point of SARS-CoV-2 S gene of Omicron BA,1, BA.2 variant and BA.1-BA.2 recombinant
614 isolate compared to that of SARS-CoV-2 Wuhan-Hu-H1. Changed nucleotides and amino acids of Omicron
615 variants BA.1, BA.2, and BA.1-BA.2 recombinant isolate compared to that of SARS-CoV-2 Wuhan-Hu-H1
616 sequences are shown in red letters. Asterisks showed an estimated homologous recombination breakpoint of
617 the SARS-CoV-2 S gene of Omicron BA.2-0.1.
618

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
 UCUGUCCUAUAUAAUCCGCAUCAUUUCCACUUUUAAGUGUUAU
 UCUGUCCUAUAUAAUCUCGCACCAUUUUCACUUUUAAGUGUUAU

 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
 AUCUAUCAGGCCGGUAACAACCUUGUAAUGGUGUU
 **
 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
 UAUCAGACUCAGACUAGUCUCAUCGGCGGGCACGUAGU

 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

