

1 **TITLE PAGE**

2

3 **Full-length title:**

4 **Analysis of SARS-CoV-2 variants from 24,181 patients exemplifies the role of**
5 **globalisation and zoonosis in pandemics**

6 **Short title (for the running head):**

7 **Globalisation, zoonosis, and SARS-CoV-2 variants**

8 **Author list: Philippe COLSON^{1,2,3} [¥], Pierre-Edouard FOURNIER^{1,2,3} [¥], Hervé**
9 **CHAUDET^{1,4}, Jérémy DELERCE¹, Audrey GIRAUD-GATINEAU^{1,3,4,5}, Linda**
10 **HOUHAMDI¹, Claudia ANDRIEU¹, Ludivine BRECHARD¹, Marielle BEDOTTO¹,**
11 **Elsa PRUDENT¹, Céline GAZIN¹, Mamadou BEYE¹, Emilie BUREL¹, Pierre**
12 **DUDOUE^{1,2}, Hervé TISSOT-DUPONT^{1,2}, Philippe GAUTRET^{1,3,4}, Jean-Christophe**
13 **LAGIER^{1,2,3}, Matthieu MILLION^{1,2,3}, Philippe BROUQUI^{1,2,3}, Philippe PAROLA^{1,3,4},**
14 **Michel DRANCOURT^{1,2,3}, Bernard LA SCOLA^{1,2,3}, Anthony LEVASSEUR^{1,2,3}, Didier**
15 **RAOULT^{1,2,3} ***

16 **Affiliations:** ¹ IHU Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille,
17 France; ² Microbes Evolution Phylogeny and Infections (MEPHI), Institut de Recherche pour
18 le Développement (IRD), Aix-Marseille Univ., 27 Boulevard Jean Moulin, 13005 Marseille,
19 France; ³ Assistance Publique-Hôpitaux de Marseille (AP-HM), 264 rue Saint-Pierre, 13005
20 Marseille, France; ⁴ Vecteurs–Infections Tropicales et Méditerranéennes (VITROME),
21 Institut de Recherche pour le Développement (IRD), Aix-Marseille Univ., 27 Boulevard Jean
22 Moulin, 13005 Marseille, France; ⁵ French Armed Forces Center for Epidemiology and Public
23 Health (CESPA), Marseille, France

24 [¥] Contributed equally

25 *** Corresponding author:** Didier Raoult, IHU Méditerranée Infection, 19-21 Boulevard Jean

26 Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email:

27 didier.raoult@univ-amu.fr

28 **Key words:** SARS-CoV-2; variant; mutant; classification; travel; introduction; epidemics

29 **Word counts:** abstract, 249; text, 6,290

30 **Figures:** 7; **Tables:** 2; **References:** 56

31 **Supplementary Material: Supplementary Figures:** 3; **Supplementary Tables:** 3

32

33

34

ABSTRACT

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

After the end of the first epidemic episode of SARS-CoV-2 infections, as cases began to rise again during the summer of 2020, we at IHU Méditerranée Infection in Marseille, France, intensified the genomic surveillance of SARS-CoV-2, and described the first viral variants. In this study, we compared the incidence curves of SARS-CoV-2-associated deaths in different countries and reported the classification of SARS-CoV-2 variants detected in our institute, as well as the kinetics and sources of the infections. We used mortality collected from a COVID-19 data repository for 221 countries. Viral variants were defined based on ≥ 5 hallmark mutations shared by ≥ 30 genomes. SARS-CoV-2 genotype was determined for 24,181 patients using next-generation genome and gene sequencing (in 47% and 11% of cases, respectively) or variant-specific qPCR (in 42% of cases). Sixteen variants were identified by analysing viral genomes from 9,788 SARS-CoV-2-diagnosed patients. Our data show that since the first SARS-CoV-2 epidemic episode in Marseille, importation through travel from abroad was documented for seven of the new variants. In addition, for the B.1.160 variant of Pangolin classification (a.k.a. Marseille-4), we suspect transmission from mink farms. In conclusion, we observed that the successive epidemic peaks of SARS-CoV-2 infections are not linked to rebounds of viral genotypes that are already present but to newly-introduced variants. We thus suggest that border control is the best mean of combating this type of introduction, and that intensive control of mink farms is also necessary to prevent the emergence of new variants generated in this animal reservoir.

55

TEXT

56

57 INTRODUCTION

58 Following its emergence in December 2019 in Wuhan, China, SARS-CoV-2 was
59 declared to be a pandemic in March 2020 and its incidence since then has shown different
60 epidemic patterns according to the country (<https://ourworldindata.org/>;
61 <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>) (1). In Asia, a
62 bell-shaped curve typical of seasonal viral respiratory infections has been observed. In
63 Western developed countries, successive phases have occurred that have associated an initial
64 bell-shaped curve and additional waves with one or more peaks. In Marseille, France, after a
65 first epidemic period characterised by a bell-shaped curve of incidence followed by a period
66 of almost zero incidence in May and June 2020, an epidemic burst of SARS-CoV-2 cases
67 occurred in July. As this differed from the single bell-shaped curve typically observed for
68 most respiratory viral infections, we suspected this corresponded to the introduction of new
69 SARS-CoV-2 genotypes. Consequently, in the summer of 2020 we expanded our assessment
70 of viral genetic diversity using genome next-generation sequencing. This allowed us to detect
71 and report as early as 7 September 2020 several SARS-CoV-2 lineages that occurred
72 concurrently or successively, and were associated with different genetic, epidemiological and
73 clinical features (2-4).

74 These observations required defining and naming these different lineages to
75 communicate about them, which is the very principle of taxonomy. Classifying and naming
76 emerging infectious agents is essential (5) but the task is particularly difficult in the case of
77 SARS-CoV-2 for several reasons. First, as for other RNA viruses, SARS-CoV-2 evolves
78 through continuous genetic variation with a high mutation rate (6, 7). Second, SARS-CoV-2
79 has caused a dramatically high number of infections (over 195 million human cases

80 worldwide and 5.9 million cases in France by 31 July 2021
81 (<https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>). In addition,
82 there are animal reservoirs, a major one being mink (8, 9). SARS-CoV-2 is, therefore,
83 currently a rapidly evolving virus with a mutation rate estimated to be 9.8×10^{-4}
84 substitutions/site/year (10) and characterised by a high rate of lineage turnover (11). Several
85 SARS-CoV-2 specific classification and naming systems including those of GISAID
86 (<https://www.gisaid.org/> (12)), Nexstrain (<https://clades.nextstrain.org/> (13)), Pangolin
87 (<https://cov-lineages.org/pangolin.html> (11)), or World Health Organization (WHO)
88 (<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>). But, although these are
89 of great interest and usefulness, other classification and naming systems are also valuable, as
90 previously acknowledged (11), particularly when it comes to better exploring genomic
91 epidemiology on a local scale through real-time next-generation sequencing. Indeed, some
92 classifications have become complicated with the advent of multiple new variants. In
93 addition, local genetic trends may be underrepresented in the tremendous flow of available
94 genomes. Moreover, correlations between genotypic patterns and epidemiological and clinical
95 characteristics can be accurately performed at institutional or local levels through a large
96 network of people who are responsible for patient diagnosis and clinical management, while
97 remote collection and analysis of data may have poorer performance. Thus, we aimed at using
98 our own definition, classification and naming of SARS-CoV-2 lineages and at correlating
99 these lineages with our local epidemiological data in order to better investigate and monitor
100 their genetic features and their sources and geographical origins. Here, we describe our
101 classification and naming system of SARS-CoV-2 lineages and the history and characteristics
102 of the major lineages detected among patients diagnosed at our institute, which sharply
103 highlights the role of travel and globalisation in the current pandemic.
104

105 **RESULTS**

106 **Broad diversity according to countries of the patterns of incidence of SARS-CoV-2-** 107 **associated deaths**

108 Clustering of the incidence curves of SARS-CoV-2-associated deaths for 221
109 countries delineated ten major groups of countries (Fig. 1A). Each group comprised a
110 mean±standard deviation of 12±4 countries (range, 7–21). A broad range of patterns was
111 revealed through these groups, with countries experiencing one or more epidemic waves of
112 varying intensity and duration combined or not with periods of lower and uneven incidence
113 (Fig. 1B). These data indicate that the incidence of SARS-CoV-2-associated deaths did not
114 follow a single bell-shaped curve except in countries from Group 1, which includes China,
115 Australia and a few countries in Central Asia and the Middle East. Instead, the incidence of
116 SARS-CoV-2-associated deaths evolved elsewhere with multiple epidemic patterns according
117 to time and geographic location. The worldwide distribution of these patterns showed that the
118 majority of Western and Northern European countries (including France), the USA and
119 Canada, which are part of Group 10, shared a same pattern (Figs. 1A and 1C). Similarly,
120 several countries in Asia and South America shared the same patterns leading to their
121 classification in Groups 1–4. In contrast, a broad panel of patterns was observed for Africa.
122 Countries that experienced a single peak of SARS-CoV-2-associated deaths (Groups 1, 4 and
123 7) were mostly in Asia, Oceania, the Middle East, South America, Africa, and Eastern
124 Europe; exceptions were Luxembourg, Austria and Switzerland. Among the countries which
125 experienced several peaks of SARS-CoV-2-associated deaths, those in Group 10 were the
126 developed industrialised countries of Western Europe and North America. These countries are
127 particularly involved in globalisation through tourism, business, and immigration (14). In
128 Marseille and its geographical area, the overall pattern of incidence of SARS-CoV-2-
129 associated deaths was similar to that of countries in Group 10, which includes France, with a

130 bell-shaped curve in March–April 2020 then a short period with a very low incidence
131 followed by a longer period with two peaks of incidence and erratic incidence between these
132 peaks (Fig. 2A). This pattern of incidence with several peaks suggests multiple introductions
133 of new variants that evolved abroad and caused distinct epidemics.

134 **Increase of SARS-CoV-2 genetic diversity over time**

135 A total of 54,703 (10.6%) out of 513,805 patients tested in our institute for SARS-
136 CoV-2 infection between 29 January 2020 and 18 August 2021 were positive by real-time
137 reverse transcription-PCR (qPCR) (<https://www.mediterranee-infection.com/covid-19/>) (Fig.
138 2A). We observed a bell-shaped curve of incidence between late February and early May
139 2020. This was followed by a period with almost no diagnoses (between 0 and 8 per day;
140 mean±standard deviation, 2.5 ± 2.2 per day) which lasted from 9 May to 5 July 2020. We
141 performed whole SARS-CoV-2 genome sequencing using next-generation sequencing
142 procedures, since we diagnosed the first infection on 27 February 2020 (15) and primarily
143 obtained viral genomes from 309 patients for the March–April period (2, 16). In July 2020,
144 the incidence of SARS-CoV-2 diagnoses increased again (Fig. 2A). As the occurrence of a
145 second epidemic two months after the end of the first one was unexpected for a respiratory
146 virus infection, we suspected that it was linked to the introduction of new SARS-CoV-2
147 genotypes. This led us to quickly expand the next-generation sequencing of SARS-CoV-2
148 genomes to test this hypothesis. We primarily obtained 382 additional genomes during
149 summer 2020 (2), and ultimately analysed the viral genomes obtained from a total of 9,788
150 patients sampled between 27 February 2020 and 18 July 2021. We observed a significant
151 increase in the mean number of mutations in SARS-CoV-2 genomes between April 2020 and
152 July 2020 (mean (±standard deviation) number of mutations per genome: 8.5 ± 2.0 versus
153 15.0 ± 4.4 , respectively; $p < 10^{-3}$ (ANOVA test)) (Fig. 2B; Table S1). This mean number
154 continued to increase during the summer of 2020, and the increase was significant between

155 July and August (20.0 ± 5.0). Thereafter, significant increases in the mean number of mutations
156 per genome were observed between September and December (25.2 ± 4.3), between November
157 2020 (23.0 ± 4.1) and February 2021 (26.4 ± 5.7), and between February and March 2021
158 (34.1 ± 4.2). Similar trends were observed for the number of amino acid changes in the spike
159 protein (Fig. S1). Overall, when analysing the pairwise genetic distance between genomes
160 obtained over time, two major periods with increased viral genetic diversity compared to
161 previous months were identified, during the summer of 2020 compared to before June 2020,
162 mean \pm standard deviation of pairwise genetic distances between genomes being 8.28×10^{-4}
163 $\pm 3.50 \times 10^{-4}$ versus $2.34 \times 10^{-4} \pm 1.24 \times 10^{-4}$, respectively ($p < 10^{-3}$), then in January–March 2021
164 when the mean \pm standard deviation pairwise genetic distances between genomes was
165 $13.93 \times 10^{-4} \pm 4.48 \times 10^{-4}$ compared to $7.14 \times 10^{-4} \pm 4.44 \times 10^{-4}$ ($p < 10^{-3}$) (Fig. 3).

166 **The Marseille IHU Méditerranée Infection classification system of SARS-CoV-2** 167 **variants**

168 Since the summer of 2020, we intensified our activity of SARS-CoV-2 genotyping
169 using next-generation sequencing and also through performing partial spike gene sequencing
170 and implementing variant-specific qPCR (17, 18). To characterise the increasing SARS-CoV-
171 2 genome diversity observed during the summer of 2020, we promptly implemented a viral
172 genome classification system and a nomenclature of SARS-CoV-2 variants. The term “SARS-
173 CoV-2 variant” had scarcely been used in the scientific literature before September 2020 (Fig.
174 S2) and no definition had been proposed to differentiate variants from mutants. We first
175 reported the emergence of SARS-CoV-2 variants on 7 September, 2020 (2), which included
176 Marseille-2 (later named B.1.177 according to Pangolin classification) and Marseille-4
177 (B.1.160) variants that were then reported in October 2020 to have emerged and, for the case
178 of Marseille-2, to have spread across Europe (19).

179 There are no universally recognised or used strategies and criteria to classify viruses
180 below the species level, and only biological classifications and nomenclatures have been
181 proposed (20, 21) (Box 1). Classification and naming systems have been specifically
182 proposed for SARS-CoV-2 including those of GISAID (<https://www.gisaid.org/>) (12),
183 Nextstrain (<https://clades.nextstrain.org/>) (13), Pangolin ([https://cov-
lineages.org/pangolin.html](https://cov-
184 lineages.org/pangolin.html)) (11), and the WHO ([https://www.who.int/en/activities/tracking-
SARS-CoV-2-variants/](https://www.who.int/en/activities/tracking-
185 SARS-CoV-2-variants/)) (Box 2). However, we chose to implement our own classification
186 system and nomenclature and to use names rather than only numbers. This approach fitted the
187 viral genomic epidemiology in our geographical area, and was more understandable and
188 easier to use, which is the very purpose of classification and nomenclature. Our strategy
189 consisted in detecting all mutations in all genomes obtained from patients diagnosed in our
190 institute and in assigning a dynamic nomenclature to track SARS-CoV-2 genetic patterns.

191 A phylogenetic tree was built that included all genomic sequences of SARS-CoV-2
192 obtained in our institute (Fig. 4A). Based on tree topology and to distinguish them from
193 SARS-CoV-2 mutants, we delineated SARS-CoV-2 variants as groups of viral genomes
194 carrying a set of at least five mutations differentiating them from any other viral genomes and
195 obtained from at least 30 (initially 10) different patients. Genomes were, therefore,
196 unambiguously assigned to a given variant based on the presence of particular sets of
197 mutations (Table 1). This differentiated variants from mutants, the genomes of which do not
198 harbour this minimal set of hallmark mutations. We first delineated seven Marseille variants
199 of SARS-CoV-2 as early as 7 September 2020, 2 of which did not eventually reach the
200 threshold number of 30 genomes (2). This was the first description of an emergence and
201 expansion of several SARS-CoV-2 variants responsible for distinct epidemics in a same
202 geographical area. Three additional variants were then defined in 2020 (Fig. 5) and four in
203 2021, making a total of 12 (Table 1; Figs. 4A and 4B). In addition, we delineated 9 other viral

204 lineages comprising genomes harbouring a specific set of at least five mutations but obtained
205 from <30 different patients; some of these lineages may grow and become *bona fide* variants
206 as additional members will be obtained. Besides, we detected four other SARS-CoV-2
207 variants that were defined as variants of concern (VOC; Box 2) and were named Alpha, Beta,
208 Gamma and Delta by the WHO ([https://www.who.int/en/activities/tracking-SARS-CoV-2-](https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/)
209 [variants/](https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/)). Regarding the nomenclature of the variants we defined, we used our city name,
210 Marseille, followed by a number according to the chronology of their description (e.g.:
211 Marseille-1, Marseille-2,...) or one of their hallmark amino acid substitutions (e.g.: Marseille-
212 501). Hereafter, we will use both the Marseille and Pangolin nomenclatures to describe the
213 genotypes of the SARS-CoV-2 genomes obtained in our institute.

214 **Emergence and outcome of the SARS-CoV-2 variants that circulated in our** 215 **geographical area**

216 *Overall prevalence of the Marseille SARS-CoV-2 variants*

217 SARS-CoV-2 genotyping was performed and successful for 24,181 (44%) of the
218 54,703 SARS-CoV-2-positive patients diagnosed in our institute as of 18 August 2021 (Table
219 2). SARS-CoV-2 genotyping was performed by genome sequencing in our institute for at
220 least 12% of positive diagnoses of SARS-CoV-2 infection monthly, and for at least one fifth
221 and one third of positive diagnoses monthly for 14 and 7 of the 18 months from February
222 2020 to July 2021, respectively (Table S2). Genotype was obtained primarily by whole
223 genome next-generation sequencing in 11,387 (47%) cases and by partial genome next-
224 generation sequencing (spike fragment) in 2,621 (11%) cases, and retrospectively or
225 prospectively (since January 2021) by in house variant-specific qPCR assays gradually
226 designed and implemented subsequent to the detection of new variants (17, 18) in 10,173
227 (42%) cases. The most frequently detected variants were, in decreasing order, the Alpha
228 (a.k.a. UK) variant, first detected in late December 2020 (Pangolin lineage B.1.1.7; n= 9,294

229 patients (38.4%)); the Marseille-4 variant, first detected in July 2020 (B.1.160; 6,125
230 (25.3%)); the Delta (a.k.a. Indian) variant, first detected in April 2021 (B.1.617.2; 4,113
231 (17.0%)); the Marseille-2 variant, first detected in August 2020 (B.1.177; 850 (3.5%)); and
232 the Beta (a.k.a. South African) variant, first detected in January 2020 (B.1.351; 575 (2.4%)).
233 Other variants of significant prevalence and/or interest in our geographical area were the
234 Marseille-8 variant (B.1.1.269; 176 (0.7%)); the Marseille-5 variant (B.1.367; 124 (0.5%));
235 the Marseille-1 variant (B.1.416; 123 (0.5%)); the Marseille-484K.V3 variant (B.1.525; 112
236 (0.5%)), the Gamma (a.k.a. Brazilian) variant (P.1; 87 (0.4%)); the Marseille-10 variant
237 (B.1.221; 91 (0.4%)); and the Marseille-501 variant (A.27; 74 (0.3%)).

238 *Dynamics of SARS-CoV-2 variants in the IHU Méditerranée Infection cohort of SARS-CoV-2-*
239 *infected patients*

240 Based on our definition of variants, we were able to identify several since the summer
241 of 2020, which we would currently call variants of concern (VOC) (Box 2). After having
242 defined and named the Marseille variants, we attempted to characterise their dynamic and to
243 analyse their epidemiological features. For some of them, we were able to find their source.
244 During the onset of the SARS-CoV-2 epidemic in late February 2020, we observed two
245 distinct SARS-CoV-2 genotypes that did not make it possible to define variants based on our
246 classification system. They corresponded to Nextstrain clades 19A and 20B that briefly co-
247 circulated in our geographical area during the first period of the SARS-CoV-2 epidemic
248 (Table 2; Figs. 6A, 6B, and 7) (22, 23). The first clade was diagnosed from 14 patients in
249 March 2020. The second, which corresponds to viruses harbouring the D614G substitution in
250 the spike, was initially retrieved from patients who stayed in northwest Italy and at the
251 French/Italian border, and quickly established in the Marseille area as a majority clade
252 together with clades 20A and 20C until May 2020.

253 We named the first variant we defined Marseille-1 (corresponds to Pangolin clade
254 B.1.416), which emerged on week 27 of 2020 and predominated in July 2020 before rapidly
255 disappearing at the end of August 2020 (3) (Table S3; Figs. 6A, 6B, and 7) (22, 23). Its onset
256 coincided with the resumption on 1 July 2020 of maritime connections between Marseille and
257 the Maghreb. As 16 of the first 20 diagnosed cases (80%) were detected in people who had
258 travelled by boat or on passenger boats circulating between the Maghreb and Marseille, we
259 assumed that this variant was brought to Marseille from the Maghreb. Furthermore, genomic
260 analyses showed that it originated from sub-Saharan Africa, mostly Senegal (3).

261 The Marseille-4 variant (B.1.160) was the most prevalent SARS-CoV-2 variant in our
262 geographical area in 2020 and was only outdone by the Alpha (UK) variant in 2021 (4). It
263 replaced the Marseille-1 variant in August 2020 and faded away in April 2021. It was first
264 detected in south-west France and in Marseille. Its emergence coincided with the re-increase
265 in France of SARS-CoV-2 incidence starting with a department in northwest France
266 (Mayenne) close to the Eure-et-Loir department where a mink farm was found to be
267 massively infected with SARS-CoV-2 (4, 24). The only genome from this farm, obtained
268 from a mink sampled in mid-November 2020 but released in March 2021, is a Marseille-4
269 variant (GISAID Accession no. EPI_ISL_1392906). In addition, the co-occurrence of 13
270 hallmark mutations in the Marseille-4 variant without any available genome showing a
271 gradual appearance of these mutations suggested that genetic evolution had been overlooked.
272 These data led us to conclude, based on genomic, geographical and temporal coincidences,
273 that the source of the Marseille-4 epidemic was the mink. The Marseille-4 variant was
274 progressively replaced from January 2021 by the epidemic linked to the Alpha (UK) variant.
275 Meanwhile, it circulated in 76 countries and accounted for 31,299 genomes in the GISAID
276 database as of 30 June 2021, mostly obtained from European countries (in 29,906 cases
277 (96%)) (Fig. S3).

278 During July and August 2020, two additional variants, Marseille-2 and Marseille 5,
279 emerged in our geographical area, which became the second and third most prevalent variants
280 from week 11 of 2020 after the Marseille-4 variant. The Marseille-2 variant (B.1.177) most
281 likely emerged in Spain, where it accounted for two-thirds of released genomes and it spread
282 to our geographical area and through Europe during the summer of 2020 (19). The Marseille-
283 5 variant (B.1.367) disappeared in October 2020 and its source and origin are unknown. The
284 Marseille-8 and Marseille-9 variants were first detected in August 2020. They accounted
285 overall for approximatively 1.2% of the cases of SARS-CoV-2 infections in our geographical
286 area, but nonetheless circulated until March and January 2021, respectively. The source and
287 origin of these variants are unknown.

288 The Marseille-501 variant emerged in Marseille in January 2021 as the first variant of
289 lineage 19B harbouring amino acid substitution N501Y in the spike protein (25). A majority
290 of the genomes primarily available from the GISAID database were linked to Mayotte, which
291 is part of the Comoros archipelago located in the Indian Ocean from where originated some of
292 the patients we diagnosed to be infected with this variant. The Marseille-484K.V3-20A
293 variant emerged in April 2021 in Marseille and was the majority variant apart from Beta
294 (South African) and Gamma (Brazilian) variants to harbour amino acid substitution E484K in
295 the spike protein. This variant was first described in Nigeria on 11 December 2020, but we
296 have not documented its importation from this country.

297 The Alpha variant emerged in December 2020 in France and in Marseille and faded
298 away in June 2021 after having been responsible for the majority of SARS-CoV-2 infections
299 from February 2021. The first cases were part of a family cluster during the Christmas
300 holidays after some family members came from England by train. We also had smaller
301 epidemics with the Beta (South African) and the Gamma (Brazilian) variants from January
302 2021. Regarding the Gamma variant, we first detected it among people returning from the

303 Comoros who had travelled through Ethiopia then Tunisia to Marseille. Between February
304 and April 2021, we also detected three cases of a rare spike E484K-harboured SARS-CoV-2,
305 which we named Marseille-484K.V1 (26). The three genomes were phylogenetically closest
306 to a genome from the Republic of Guinea (a.k.a. Guinea-Conakry), where one of the patients
307 had travelled.

308 Most recently, we have mostly detected infections with the Delta (B.617.2) variant, in
309 4,113 patients. The two first patients diagnosed with this variant in late April 2021 were an
310 Indian sailor re-embarking in Marseille after the resumption of maritime travel (La Scola *et*
311 *al.*, 2021), and a physiotherapist who had travelled to the Republic of the Maldives, in
312 Southeast Asia, and provided care upon his return to France in an accommodation facility for
313 dependent elderly people, being involved in a cluster among residents and health care
314 workers. Finally, in mid-July 2021, we detected a variant classified in the Pangolin lineage
315 B.1.621, that was first identified in Colombia in January 2021. Information was available for
316 one of the six patients infected with this variant, a 10-year-old Colombian girl who developed
317 symptoms and was diagnosed in our institute four days after she arrived in France.

318 Overall, we identified several SARS-CoV-2 variants that were responsible as early as
319 the summer of 2020 for distinct epidemics causing several waves of SARS-CoV-2 incidence
320 (Figs. 6A, 6B, and 7) (22, 23). The epidemics caused by these different variants each
321 exhibited a bell-shaped curve, and they occurred successively or concomitantly, therefore
322 contributing to the total case load. We documented that new SARS-CoV-2 variants were
323 introduced in our geographical area by boat in one case (Marseille-1), by train in one case
324 (Alpha), and by plane in five cases (Alpha, Beta, Marseille-501, Marseille-484K.V1,
325 B.1.617.V2, and B.1.621 variants). In addition, the importation of Marseille-484K.V3 from
326 Nigeria through air travel is suspected. Importation from abroad through travel therefore
327 accounted for a substantial source of variants and, subsequently, of cases. The other source

328 identified for the emergence of a new variants is the mink epidemics in Northern France, with
329 the case of Marseille-4 (4).

330

331 **DISCUSSION**

332 In this study, we wanted to clarify the classification and naming of SARS-CoV-2
333 genotypes we implemented in our institute in order to have a simple identification of SARS-
334 CoV-2 variants and to characterize their mutations, their source, and spread. This is essential
335 to get a better understanding of SARS-CoV-2 epidemiology. Since the SARS-CoV-2
336 emergence in France, we have performed in our institute extensive surveillance, identification
337 and monitoring of SARS-CoV-2 variants. Currently, the viral genotype has been determined
338 for 44% of the patients diagnosed in our institute and the viral genome was obtained in
339 approximately one fifth of patients diagnosed with SARS-CoV-2. As early as the summer of
340 2020, we increased our surveillance of SARS-CoV-2 genotypes which allowed us to decipher
341 several features of the current SARS-CoV-2 pandemic on a local scale. Before the summer of
342 2020, very few studies had pointed out an increase of SARS-CoV-2 genetic and amino acid
343 diversity. Noticeably, Tomaszewski *et al.* had reported new pathways of mutational changes
344 in SARS-CoV-2 proteomes (27).

345 First, it appears relatively clear that in countries which have been isolated, either
346 geographically (islands that have closed their access) or politically (China and Korea), there
347 has been only one major epidemic episode with a bell-shaped incidence curve. In Europe and
348 the United States, where such a policy has not been implemented or has been made
349 impossible by political circumstances, several epidemics have occurred successively or
350 concurrently with different viral variants. Marseille, which is a very particular city due to its
351 geographical location, has for centuries been at the forefront of imports of epidemics
352 including of plague and cholera (28), and operated in an identical way during the SARS-CoV-

353 2 pandemic. Thus, we observed that the first diagnosed SARS-CoV-2 infections were
354 imported cases involving people who had stayed in northwest Italy and at the French/Italian
355 border. This preceded a first epidemic phase during which international borders were closed.
356 Their reopening in early summer 2020 was followed by a short epidemic. This was linked to a
357 SARS-CoV-2 variant that we named Marseille-1, which was imported from North Africa by
358 boat and originated from sub-Saharan Africa, and was probably not very transmissible, since
359 this epidemic ended two months after its emergence and did not spread beyond the Marseille
360 region. Another variant, which we named Marseille-2 (B.1.177), was introduced to our
361 geographical area during the summer of 2020 from Spain, where it appears to have originated
362 (19). Later, in 2021, the Alpha (B.1.1.7, or UK) variant was first imported in Marseille by
363 French people returning from the UK. In addition, in the Marseille population which includes
364 lots of people of Comorian origin, we observed the onset of small epidemics with the Beta
365 (B.1.351, or South African) variant among people travelling by plane from the Comoros, and
366 with the Marseille-501 (A.27) variant that involved people of Comorian origin including from
367 Mayotte (25). Similarly, the reopening of sea cruises in the spring of 2020 was associated
368 with an importation of the Delta (B.1.617.2, or Indian) variant by an Indian sailor who had
369 travelled from India to Marseille (29), and a second case identified of this variant was in a
370 traveller returning from the Republic of the Maldives located in the Indian Ocean. Two other
371 SARS-CoV-2 variants were imported in 2021 from the Republic of Guinea and from
372 Colombia. We therefore observed that the importation from abroad through travel accounted
373 for a substantial source of variants and subsequently of infections, highlighting the role of
374 globalisation in the spread of SARS-CoV-2 pandemic. Marseille is a city with many
375 immigrant adults and children, and its geographical location and harbour make it very open to
376 travellers. This explains the importation of infectious agents of various origins, which has
377 been recognised in Marseille since antiquity (28). The considerable role of travel in the

378 introduction and subsequent spread of new SARS-CoV-2 variants on the national or
379 continental scales has also been reported in a recent phylogeographical study that analysed
380 genomic, epidemiological, and mobility data (including ours) collected from ten European
381 countries between January and October 2020 (30). This study found that the intensity of
382 international travel predicted the spread of SARS-CoV-2 with the introduction of new viral
383 variants, and it estimated that more than half of the lineages that were spreading in late
384 summer 2020 had been newly introduced to countries since mid-June 2020. Such
385 importations into countries of SARS-CoV-2 variants and their impact on the pattern of
386 incidence of SARS-CoV-2 cases have also been reported in other studies (31). This sharply
387 confirms what we described as early as at the beginning of September 2020 (2) with the
388 emergence of seven new SARS-CoV-2 including one (Marseille-1) for which we had already
389 found a source (3). Thus, we were the first to consider and highlight that the re-increase of
390 SARS-CoV-2 incidence observed during the summer of 2020 in Marseille and in France may
391 have resulted from the importation of new variants rather than from a rebound of previous
392 viral genotypes (2, 3). This aligns with the fact that countries for which cross-border travel
393 was the most limited geographically and/or politically by closing the borders experienced a
394 single epidemic peak, whereas countries with high levels of international travel experienced
395 several epidemic peaks and longer periods with a significant incidence of infections. Hence,
396 these findings indicate that while local control measures such as lockdowns and curfews may
397 limit viral spread, they are inefficient to avoid the introduction of new viral genotypes. The
398 role of local lockdowns in controlling the viral spread has been controversial. This measure
399 has been found to be irrelevant by some researchers (32). In Japan, lockdown was not
400 associated with a reduction of SARS-CoV-2 incidence, whereas travel restrictions were (33).

401 Second, our findings indicate that the role of epizootics in dense animal herds has been
402 overlooked in the generation of variants, their transmission to humans, and their contribution

403 to the SARS-CoV-2 pandemic. Mink was revealed as a major source of SARS-CoV-2
404 diversity and a key element in the expansion of the human pandemic (8, 9). Data have
405 accumulated indicating that the current pandemic, which is primarily a zoonosis, continued
406 being a zoonosis with the massive involvement of minks, while other animals, including
407 domestic animals such as cats and dogs, have also been affected (9). It has been demonstrated
408 that minks were a source of new variants that fuelled the pandemic in Denmark (8), and
409 SARS-CoV-2 spike protein gene variants harbouring amino acid substitutions N501T and
410 G142D were also found to circulate in both humans and minks in the United States (34). In
411 Marseille, we accumulated indications that the Marseille-4 variant originated from mink. This
412 hypothesis took us a long time to confirm due to the very late release, in March 2021, of the
413 viral sequence obtained from a mink farm in Eure-et-Loir in mid-November 2020.
414 Interestingly, the detection of two epidemics, in Marseille and retrospectively in Mayenne
415 (Northwest France), appeared to be unrelated as the absence of SARS-CoV-2 genotyping
416 initially carried out in Mayenne and Eure-et-Loir prevented the demonstration that they were
417 due to the same variant and that virus transmission could have occurred from farm minks
418 located in a neighbouring department. Interestingly, the Marseille-4 variant was responsible
419 for a large epidemic through Europe and was probably considerably overlooked as the global
420 concern about SARS-CoV-2 variants only emerged several months after we described it.

421 These findings show that SARS-CoV-2 genomic surveillance and the definition of
422 variants are valuable for a better assessment of SARS-CoV-2 epidemiology. In addition, they
423 allow correlating viral genotypes with the clinical severity and outcome of infections. Indeed,
424 several studies have reported different clinical pattern of infections according to the SARS-
425 CoV-2 variant (35, 36). We previously compared the clinical symptoms and outcomes of
426 patients infected with different variants we identified. Compared with viruses of the
427 Nextstrain clade 20A that circulated between February and May 2020 in Marseille, patients

428 infected with the Marseille-1 variant were less likely to report dyspnoea and rhinitis and to be
429 hospitalised (35). In addition, patients infected with the Marseille-4 variant were more likely
430 to exhibit fever and to be hospitalised than those infected with the Marseille-1 variant, and
431 they were also more likely to exhibit fever than those infected with a virus of Nextstrain clade
432 20A. Also, patients infected with N501Y-harboring SARS-CoV-2 variants (namely, Alpha
433 (B.1.1.7), Beta (B.1.351) and Gamma (P.1) variants) were less likely to be hospitalised than
434 those infected with the Marseille-1 variant and with viruses of Nextstrain clade 20A that
435 circulated between February and May 2020 in Marseille (35). Furthermore, a key issue from
436 the present study is highlighting the fluctuations in incidence of SARS-CoV-2 infections that
437 we observed resulted from the succession or combination of distinct epidemics caused by
438 different viral variants rather than from the rebound of persisting lineages, which we primarily
439 described in September 2020 and was confirmed at larger scale (30). Since then, these genetic
440 evolutionary and epidemiological patterns continued to be observed and the surveillance of
441 SARS-CoV-2 variants, the role of travel and globalisation in their spread, and the turnover of
442 major VOC has come to the frontline, particularly with the Alpha, Beta, Gamma, and Delta
443 variants (37, 38).

444 Overall, previous findings have revealed that epidemic waves mostly observed in
445 developed Western countries are, in fact, distinct epidemics due to different imported SARS-
446 CoV-2 variants. Therefore, as the major identified sources for these SARS-CoV-2 variants is
447 cross border travel and animal reservoirs, efficient border controls and surveillance of
448 mustelid farms are critical in order to avoid the introduction of new SARS-CoV-2 variants
449 and the occurrence of new epidemics.

450

451 **MATERIALS AND METHODS**

452 **Epidemic curves for SARS-CoV-2-associated deaths in various countries**

453 We used the cumulated daily mortality from the COVID-19 data repository operated
454 by Johns Hopkins University (39) available at [https://github.com/CSSEGISandData/COVID-](https://github.com/CSSEGISandData/COVID-19)
455 19 (accessed on 25 June 2021), which covers 221 countries and locations from 22 April 2020
456 to 24 June 2021. Data mining of time series is a complex scientific field where the analysis
457 process must be tailored to the characteristics of the series and the clustering objectives (40,
458 41). Our objective was to cluster the dynamic of epidemics at country scale as reflected by
459 COVID-19 mortality, taking in account the time location of the different series peaks but not
460 the country mortality incidences. A threshold of 500 deaths was used for countries to analyse
461 their SARS-CoV-2-associated death curves. We transformed the cumulative death counts into
462 smoothed (seven days centred moving average) daily death rates. The resulting series were
463 then clustered using the Pearson distance after normalisation as L2-normed lock-step distance
464 measure (42) and the Ascendant Hierarchical Classification algorithm with Ward aggregation.
465 For determining the resulting clusters, we used a dynamic cluster detection method (43) with
466 a minimum cluster size set at five countries. We then mapped the geographic distribution of
467 cluster members. The data mining process was done using 'R' (<https://www.r-project.org/>)
468 with the packages 'vegan', 'forecast', 'dynamictreecut' and 'rnaturalearth'.

469 **Study period and clinical samples**

470 SARS-CoV-2 genotyping was performed from nasopharyngeal samples tested
471 between 27 February 2020 and 18 August 2021 (18 months) at the IHU Méditerranée
472 Infection institute (<https://www.mediterranee-infection.com/>). Specimens with a cycle
473 threshold value (Ct) lower than 20 were selected as a in priority for whole genome sequencing
474 of SARS-CoV-2 genomes, and those with a Ct between 20 and 30 were included secondarily
475 to ensure a more comprehensive coverage of the study period. Partial sequencing of the
476 SARS-CoV-2 spike gene generating a 1,854 nucleotide-long sequence was performed in 2020
477 and, mainly, between January and April 2021, as previously described (25). For respiratory

478 specimens with Ct values > 30 or those with Ct values < 30 but from which genome
479 sequences were not obtained, we identified those harbouring specific Marseille genotypes
480 using qPCR targeting variant-specific regions, as previously described for the Marseille-1
481 variant (3) and the Marseille-4 variant (4, 18). Additional in house and commercial variant-
482 specific qPCR assays were used for other variants. The study was approved by the ethics
483 committee of the University Hospital Institute Méditerranée Infection under No. 2020-016-3.

484 **Genome sequencing**

485 Viral RNA was extracted from 200 µL of nasopharyngeal swab fluid using the EZ1
486 Virus Mini Kit v2.0 on an EZ1 Advanced XL instrument (Qiagen, Courtaboeuf, France) or
487 the KingFisher Flex system (Thermo Fisher Scientific, Waltham, MA, USA), following the
488 manufacturer's instructions. SARS-CoV-2 genome sequences were obtained by next-
489 generation sequencing with various procedures since February 2020 until August 2021: (i)
490 with the Illumina technology using the Nextera XT paired end strategy on MiSeq instruments
491 (Illumina Inc., San Diego, CA, USA), as previously described (3); (ii) with the Illumina
492 COVIDSeq protocol on a NovaSeq 6000 instrument (Illumina Inc.) since April 2021; or (iii)
493 with Oxford Nanopore technology (ONT) on MinION or GridION instruments (Oxford
494 Nanopore Technologies Ltd., Oxford, UK), as previously described (3). Next-generation
495 sequencing with ONT was performed without or with (since March 2021) synthesized cDNA
496 amplification using a multiplex PCR protocol with ARTIC nCoV-2019 V3 Panel primers
497 purchased from Integrated DNA technologies (IDT, Coralville, IA, USA) according to the
498 ARTIC procedure (<https://artic.network/>). After its extraction, viral RNA was reverse-
499 transcribed using SuperScript IV (ThermoFisher Scientific) prior to cDNA second strand
500 synthesis with Klenow Fragment DNA polymerase (New England Biolabs, Beverly, MA,
501 USA) when performing NGS on the Illumina MiSeq instrument (Illumina Inc.) (3),
502 LunaScript RT SuperMix kit (New England Biolabs) when performing NGS with the ONT, or

503 according to the COVIDSeq protocol (Illumina Inc.) following the manufacturer's
504 recommendations. Generated cDNA was purified using Agencourt AMPure XP beads
505 (Beckman Coulter, Villepinte, France) and quantified using Qubit 2.0 fluorometer
506 (Invitrogen, Carlsbad, CA, USA), and fragment sizes were analysed on an Agilent Fragment
507 analyser 5200 (Agilent Inc., Palo Alto, CA). Next-generation sequencing performed on the
508 NovaSeq 6000 instrument followed the Illumina COVIDSeq protocol (Illumina Inc.), which
509 included first strand cDNA synthesis from extracted viral RNA; cDNA amplification with
510 two COVIDSeq primer pools; fragmentation and tagging of PCR amplicons with adapter
511 sequences; clean up; PCR amplification (7 cycles) of tagged amplicons; pool, clean up,
512 quantification and normalization of libraries; and library sequencing on a NovaSeq 6000
513 sequencing system SP flow cell (Illumina Inc.).

514 **Genome sequence analyses**

515 Genome consensus sequences were generated by mapping on the SARS-CoV-2
516 genome GenBank accession no. NC_045512.2 (Wuhan-Hu-1 isolate) with the CLC Genomics
517 workbench v.7 (with the following thresholds: 0.8 for coverage and 0.9 for similarity)
518 (<https://digitalinsights.qiagen.com/>), as previously described (3), or the Minimap2 software
519 (44). A frequency of the majority nucleotide $\geq 70\%$ and a nucleotide depth ≥ 10 (when
520 sequence reads were generated on the NovaSeq Illumina instrument (Illumina Inc.)) or ≥ 5
521 (when sequence reads were generated on the MiSeq Illumina instrument) were used.
522 Detection of mutations was performed using the freebayes tool
523 (<https://github.com/freebayes/freebayes>) (45) with a mapping quality score of 20. SAMtools
524 was used for soft clipping of ARTIC primers (<https://artic.network/>), and to remove PCR
525 duplicates (46). Sequences described in the present study have been deposited on the GISAID
526 sequence database (<https://www.gisaid.org/>) (12) and can be retrieved online using the
527 GISAID online search tool with “Marseille” as keyword, then selecting sequence names

528 containing “IHU” or “MEPHI”. In addition, they have been deposited on the IHU Marseille
529 Infection website: [https://www.mediterranee-infection.com/sequences-genomiques-sars-cov-](https://www.mediterranee-infection.com/sequences-genomiques-sars-cov-2-completes-partielles-sequences-spike-protein-jusquen-mai-2021/)
530 [2-completes-partielles-sequences-spike-protein-jusquen-mai-2021/](https://www.mediterranee-infection.com/sequences-genomiques-sars-cov-2-completes-partielles-sequences-spike-protein-jusquen-mai-2021/).

531 Numbers of nucleotide changes in the SARS-CoV-2 genomes and of amino acid
532 changes in the SARS-CoV-2 spike protein were obtained using the Nextclade tool
533 (<https://clades.nextstrain.org/results>). They were plotted using the GraphPad software v5.01
534 (<https://www.graphpad.com/>). Pairwise nucleotide distances between SARS-CoV-2 genomes
535 were computed using the MEGA7 software v10.2.5 (<https://www.megasoftware.net/>).

536 **Real-time reverse transcription-PCR specific of Marseille variants**

537 For specimens with Ct values > 30 or those with Ct values < 30 but from which genome
538 sequences were not obtained, we identified those harbouring specific Marseille genotypes
539 using qPCR targeting variant-specific regions. We previously described qPCR specific of the
540 Marseille-1 variant (3) and the Marseille-4 variant (4). We used other in house variant-
541 specific qPCR systems for which the sequences of primers and probes are as follows,
542 provided in 5'-3' orientation. For the Beta (B.1.351) variant, qPCR targets the envelope gene
543 and uses forward primer C_SA_3_MBF: TGAATTGCAGACACCTTTTGA, reverse primer
544 C_SA_3_MBR: CAACCCTTGGTTGAATAGTCTTG, and probe C_SA_3_MBP:
545 TGACATCTTCAATGGGGAATGT. For the Gamma (P.1) variant, qPCR targets the
546 nucleocapsid gene and uses forward primer C_Bra_2_MBF:
547 GTCAAGCCTCTTCTCGTTCCT, reverse primer C_Bra_2_MBR:
548 AAAGCAAGAGCAGCATCACC, and probe C_Bra_2_MBP:
549 GCAGCTCTAAACGAACTTCTCCTG. For the Delta (B.1.617.2) variant, two qPCR were
550 used that target the envelope gene: the first one uses forward primer C_IND_1_MBF:
551 AATCTTGATTCTAAGGTTGGTGGT, reverse primer C_IND_1_MBR:
552 TGCTACCGCCTGATAGATT, and probe C_IND_1_MBP:

553 TTACCGGTATAGATTGTTTAGGAAGTCT; the second one uses forward primer
554 P681_MBF: TGACATACCCATTGGTGCAG, reverse primer P681_MBR:
555 GGCAATGATGGATTGACTAGC, and probe P681_MBP: ATTCTCATCGGCGGGCA.
556 Finally, for the Marseille-2 (B.1.177) variant, qPCR targets the nsp16 gene and uses forward
557 primer C2_2_MBF: CATGGTGGACAGCCTTTGTT, reverse primer C2_2_MBR:
558 CATCTATTTGTTTCGCGTGGTT, and probe C2_2_MBP:
559 AATGCCTCATCATCTGAAGCAT. PCR conditions were the same than those described in
560 (25), (3) and (4). Finally, we used the Applied Biosystems TaqPath COVID-19 kit (Thermo
561 Fisher Scientific, Waltham, USA) then the TaqMan SARS-CoV-2 mutation assay H69-V70
562 (Thermo Fisher Scientific) for Alpha (B.1.1.7) variant detection, and TaqMan SARS-CoV-2
563 mutation assays L452R and P681R (Thermo Fisher Scientific) to detect the Delta (B.1.617.2)
564 variant.

565 **SARS-CoV-2 classification and naming system**

566 Phylogeny reconstructions based on the SARS-CoV-2 genomes obtained in our
567 laboratory were performed using the nextstrain/ncov tool (<https://github.com/nextstrain/ncov>)
568 then visualised with Auspice (<https://docs.nextstrain.org/projects/auspice/en/stable/>) or
569 FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>); the limitation of sampling per date
570 of collection of respiratory specimens was deactivated. Genome sequences were kept for
571 analyses when larger than 24,000 nucleotides. Within reconstructed trees, we identified and
572 counted all mutations between nodes and between nodes and leaves. This allowed us to define
573 the set of mutations for the most recent ancestor of the clusters corresponding to each of the
574 variants.

575 **Statistical analyses**

576 Statistical tests were carried out using R 4.0.2 (<https://cran.r-project.org/>) and
577 GraphPad software v5.01 (<https://www.graphpad.com/>). A $p < 0.05$ was considered
578 statistically significant.

579

580

581 **Acknowledgments**

582 We are thankful to the IHU Méditerranée Infection COVID-19 Task Force that
583 notably includes Sophie Amrane, Camille Aubry, Sofiane Bakour, Karim Bendamardji, Cyril
584 Berenger, Nadim Cassir, Anne Darmon, Claire Decoster, Catherine Dhiver, Barbara Doudier,
585 Sophie Edouard, Véronique Filosa, Séverine Guitton, Marie Hocquart, Morgane Mailhe,
586 Florine Pelisson, Coralie Porcheto, Isabelle Ravaux, Piseth Seng, Laurence Thomas, Raphael
587 Tola, Christelle Tomei, Catherine Triquet. This manuscript has been edited for English
588 language by a native English speaker.

589

590 **Author contributions**

591 Conceptualization, D.R., P.C., and P.-E.F.; methodology, D.R., P.C., P.E.F., A.L.; data
592 and investigation, P.C., P.-E.F., H.C., J.D., A.G.-G., L.H., C.A., L.B., M.B., E.P., C.G.,
593 M.Be., E.B., P.D., H.T.-D., P.G., J.-C.L., M.M., P.B., P.P., F.F., M.D., B.L.S., A.L.; original
594 draft preparation, D.R., P.C., and P.-E.F.; review and editing of the manuscript, all authors;
595 supervision, D.R.;. All authors have read and agreed to the published version of the
596 manuscript.

597

598 **Funding**

599 This work was supported by the French Government under the “Investments for the
600 Future” programme managed by the National Agency for Research (ANR), Méditerranée-

601 Infection 10-IAHU-03 and was also supported by Région Provence Alpes Côte d'Azur and
602 European funding PRIMMI (Plateformes de Recherche et d'Innovation Mutualisées
603 Méditerranée Infection), FEDER PA 0000320 PRIMMI.

604

605 **Data Availability Statement**

606 Genome sequences have been submitted to the GISAID sequence database
607 (<https://www.gisaid.org/>).

608

609 **Conflicts of interest**

610 The authors have no conflicts of interest to declare. Funding sources had no role in the
611 design and conduct of the study; collection, management, analysis, and interpretation of the
612 data; and preparation, review, or approval of the manuscript.

613

614 **Ethics**

615 All data were generated as part of the routine work at Assistance Publique-Hôpitaux
616 de Marseille (Marseille university hospitals), and this study results from routine standard
617 clinical management. The study was approved by the ethics committee of the University
618 Hospital Institute Méditerranée Infection under No. 2020-016-3.

619

620

621

REFERENCES

- 622 1. D. Cucinotta, M. Vanelli. WHO Declares COVID-19 a Pandemic. *Acta Biomed.* **91**,
623 157-160 (2020).
- 624 2. P. Colson *et al.* Dramatic increase in the SARS-CoV-2 mutation rate and low mortality
625 rate during the second epidemic in summer in Marseille. *IHU. pre-prints* (2020).
626 <https://doi.org/10.35088/68c3-ew82>.
- 627 3. P. Colson *et al.* Introduction into the Marseille geographical area of a mild SARS-CoV-
628 2 variant originating from sub-Saharan Africa: An investigational study. *Travel. Med.*
629 *Infect Dis* **40**, 101980 (2021).
- 630 4. P. E. Fournier *et al.* Emergence and outcomes of the SARS-CoV-2 'Marseille-4' variant.
631 *Int J Infect Dis* **106**, 228-236 (2021).
- 632 5. R. R. Sokal. Classification: purposes, principles, progress, prospects. *Science.* **185**,
633 1115-1123 (1974).
- 634 6. E. Domingo, C. Perales. Viral quasispecies. *PLoS. Genet.* **15**, e1008271 (2019).
- 635 7. M. Eigen. On the nature of virus quasispecies. *Trends Microbiol.* **4**, 216-218 (1996).
- 636 8. B. B. Oude Munnink *et al.* Transmission of SARS-CoV-2 on mink farms between
637 humans and mink and back to humans. *Science* **371**, 172-177 (2021).
- 638 9. F. Fenollar *et al.* Mink, SARS-CoV-2, and the Human-Animal Interface. *Front*
639 *Microbiol.* **12**, 663815 (2021).
- 640 10. D. L. van Dorp *et al.* No evidence for increased transmissibility from recurrent
641 mutations in SARS-CoV-2. *Nat. Commun.* **11**, 5986-19818 (2020).
- 642 11. A. Rambaut *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist
643 genomic epidemiology. *Nat. Microbiol.* **5**, 1403-1407 (2020).
- 644 12. E. Alm *et al.* Geographical and temporal distribution of SARS-CoV-2 clades in the
645 WHO European Region, January to June 2020. *Euro. Surveill* **25**, 2001410 (2020).

- 646 13. J. Hadfield *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*.
647 **34**, 4121-4123 (2018).
- 648 14. M. Chinazzi *et al.* The effect of travel restrictions on the spread of the 2019 novel
649 coronavirus (COVID-19) outbreak. *Science*. **368**, 395-400 (2020).
- 650 15. P. Colson *et al.* Ultrarapid diagnosis, microscope imaging, genome sequencing, and
651 culture isolation of SARS-CoV-2. *Eur. J Clin Microbiol Infect Dis*. **39**, 1601-1603
652 (2020).
- 653 16. A. Levasseur *et al.* Genomic diversity and evolution of coronavirus (SARS-CoV-2) in
654 France from 309 COVID-19-infected patients. *bioRxiv* (2020). doi:
655 <https://doi.org/10.1101/2020.09.04.282616>.
- 656 17. M. Bedotto, P. E. Fournier, L. Houhamdi, P. Colson, D. Raoult. Implementation of an
657 in-house real-time reverse transcription-PCR assay to detect the emerging SARS-CoV-2
658 N501Y variants. *J Clin. Virol* **140**, 104868 (2021).
- 659 18. M. Bedotto *et al.* Implementation of an in-house real-time reverse transcription-PCR
660 assay for the rapid detection of the SARS-CoV-2 Marseille-4 variant. *J Clin. Virol* **139**,
661 104814 (2021).
- 662 19. E. B. Hodcroft *et al.* Spread of a SARS-CoV-2 variant through Europe in the summer of
663 2020. *Nature*. **595**, 707-712 (2021).
- 664 20. International Committee on Taxonomy of Viruses (ICTV). Virus Taxonomy: The ICTV
665 Report on Virus Classification and Taxon Nomenclature. The Online (10th) Report of
666 the International Committee on Taxonomy of Viruses. (2020).
667 https://talk.ictvonline.org/ictv-reports/ictv_online_report/. 21. C. M. Fauquet *et al.*
668 Geminivirus strain demarcation and nomenclature. *Arch. Virol* **153**, 783-821 (2008).

- 669 22. Evolution of the weekly incidence of SARS-CoV-2 mutants and variants extrapolated to
670 the total number of cases, based on their proportions of genotyped cases . Audiovisual
671 Material (2021). <https://doi.org/10.35081/nz2g-f980>.
- 672 23. Introduction of SARS-CoV-2 mutants and variants in the Marseille geographical area
673 through travel from abroad. Audiovisual Material (2021). [https://doi.org/10.35081/ffd4-](https://doi.org/10.35081/ffd4-1y77)
674 [1y77](https://doi.org/10.35081/ffd4-1y77).
- 675 24. Santé Publique France. Incidence rate of SARS-CoV-2 infections by French department
676 per sliding week. (2021). <https://geodes.santepubliquefrance.fr/>.
- 677 25. P. Colson *et al.* Spreading of a new SARS-CoV-2 N501Y spike variant in a new
678 lineage. *Clin. Microbiol. Infect* 27, 1352.e1-1352.e5. Online ahead of print. (2021).
- 679 26. P. Colson *et al.* Limited spread of a rare spike E484K-harboring SARS-CoV-2 in
680 Marseille, France. *IHU. pre-prints* (2021). doi: <https://doi.org/10.35088/2ngw-sa32>.
- 681 27. T. Tomaszewski *et al.* New Pathways of Mutational Change in SARS-CoV-2 Proteomes
682 Involve Regions of Intrinsic Disorder Important for Virus Replication and Release.
683 *Evol. Bioinform. Online.* **16**, 1176934320965149 (2020).
- 684 28. R. Barbieri, P. Colson, D. Raoult, M. Drancourt. Two-millennia fighting against port-
685 imported epidemics, Marseille. I *IHU. pre-prints* (2021). doi:
686 <https://doi.org/10.35088/84a4-me41>.
- 687 29. B. La Scola *et al.* SARS-CoV-2 variant from India to Marseille: the still active role of
688 ports in the introduction of epidemics. *Travel. Med Infect Dis* **42**, 102085 Epub 2021
689 May 21 (2021). doi: 10.1016/j.tmaid.2021.102085.
- 690 30. P. Lemey *et al.* Untangling introductions and persistence in COVID-19 resurgence in
691 Europe. *Nature.* **595**, 713-717 (2021).
- 692 31. A. McLaughlin *et al.* Early and ongoing importations of SARS-CoV-2 in Canada.
693 medRxiv (2021). doi: <https://doi.org/10.1101/2021.04.09.21255131>.

- 694 32. E. R. Melnick, J. P. A. Ioannidis. Should governments continue lockdown to slow the
695 spread of covid-19? *BMJ*.369 m1924 (2020). doi: 10.1136/bmj.m1924.
- 696 33. T. Hideki, M. Skidmore. A Cross-Country Analysis of the Determinants of Covid-19
697 Fatalities. *CESifo Working Papers* **9028** (2021).
698 [https://www.cesifo.org/en/publikationen/2021/working-paper/cross-country-analysis-](https://www.cesifo.org/en/publikationen/2021/working-paper/cross-country-analysis-determinants-covid-19-fatalities)
699 [determinants-covid-19-fatalities](https://www.cesifo.org/en/publikationen/2021/working-paper/cross-country-analysis-determinants-covid-19-fatalities).
- 700 34. H. Y. Cai, A. Cai. SARS-CoV2 spike protein gene variants with N501T and G142D
701 mutation-dominated infections in mink in the United States. *J. Vet. Diagn. Invest.* **33**,
702 939-942 (2021).
- 703 35. T. L. Dao *et al.* Clinical outcomes in COVID-19 patients infected with different SARS-
704 CoV-2 variants in Marseille, France. *Clin. Microbiol. Infect.* 10 (2021). S1198-
705 743X(21)00270-6 Online ahead of print (2021). doi: 10.1016/j.cmi.2021.05.029.
- 706 36. P. Domingo *et al.* Not all COVID-19 pandemic waves are alike. *Clin. Microbiol. Infect.*
707 **27**, 1040-1040 (2021).
- 708 37. W. T. Harvey *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat.*
709 *Rev. Microbiol.* 1-16 (2021).
- 710 38. R. C. Del, P. N. Malani, S. B. Omer. Confronting the Delta Variant of SARS-CoV-2,
711 Summer 2021. *JAMA*. 10 (2021).
- 712 39. E. Dong, H. Du, L. Gardner. An interactive web-based dashboard to track COVID-19 in
713 real time. *Lancet Infect. Dis.* **20**, 533-534 (2020).
- 714 40. T. W. Liao. Clustering of time series data—a survey. *Pattern Recognition* **38**, 1857-
715 1874 (2005).
- 716 41. T. Fu. A review on time series data mining. *Eng. Appl. Artif. Intell.* **24**, 164-181 (2011).

- 717 42. U. Mori, A. Mendiburu, J. A. Lozano. Similarity Measure Selection for Clustering Time
718 Series Databases. *IEEE Transactions on Knowledge and Data Engineering* **28**, 181-194
719 (2016).
- 720 43. P. Langfelder, B. Zhang, S. Horvath. Defining clusters from a hierarchical cluster tree:
721 the Dynamic Tree Cut package for R. *Bioinformatics*. **24**, 719-720 (2008).
- 722 44. H. Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. **34**,
723 3094-3100 (2018).
- 724 45. E. Garrison, G. Marth. Haplotype-based variant detection from short-read sequencing.
725 *arXiv* (2012). <https://arxiv.org/abs/1207.3907>.
- 726 46. H. Li *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**,
727 2078-2079 (2009).
- 728 47. M. J. Adams, E. J. Lefkowitz, A. M. King, E. B. Carstens. Recently agreed changes to
729 the International Code of Virus Classification and Nomenclature. *Arch. Virol.* **158**,
730 2633-2639 (2013).
- 731 48. C. M. Fauquet, J. Stanley. Revising the way we conceive and name viruses below the
732 species level: a review of geminivirus taxonomy calls for new standardized isolate
733 descriptors. *Arch. Virol* **150**, 2151-2179 (2005).
- 734 49. M. H. Van Regenmortel. Virus species and virus identification: past and current
735 controversies. *Infect Genet. Evol.* **7**, 133-144 (2007).
- 736 50. E. Domingo, R. A. Flavell, C. Weissmann. In vitro site-directed mutagenesis:
737 generation and properties of an infectious extracistronic mutant of bacteriophage Qbeta.
738 *Gene* **1**, 3-25 (1976).
- 739 51. M. Eigen. Molecular self-organization and the early stages of evolution. *Q. Rev.*
740 *Biophys.* **4**, 149-212 (1971).
- 741 52. R. Andino, E. Domingo. Viral quasispecies. *Virology* **479-480**, 46-51 (2015).

- 742 53. M. Eigen. Viral quasispecies. *Sci. Am.* **269**, 42-49 (1993).
- 743 54. E. V. Koonin. Darwinian evolution in the light of genomics. *Nucleic Acids Res.* **37**,
744 1011-1034 (2009).
- 745 55. A. Rambaut, D. Posada, K. A. Crandall, E. C. Holmes. The causes and consequences of
746 HIV evolution. *Nat. Rev. Genet.* **5**, 52-61 (2004).
- 747 56. J. H. Kuhn *et al.* Virus nomenclature below the species level: a standardized
748 nomenclature for laboratory animal-adapted strains and variants of viruses assigned to
749 the family Filoviridae. *Arch. Virol.* **158**, 1425-1432 (2013).
- 750
- 751

752

BOX

753 **Box 1. Definition of virus groups at and below species level**

754 **Viral species:** Defined according to a polythetic classification as monophyletic groups of
755 viruses whose properties can be distinguished from those of other species by multiple criteria
756 including increasingly genetic criteria (<https://talk.ictvonline.org/taxonomy/w/ictv-taxonomy>)
757 (47); (48).

758 Below the level of species, there are currently no universally recognised or used strategies and
759 criteria to classify viruses (20, 21).

760 **Viral clade:** Monophyletic group comprised by a common ancestor and all its descent. They
761 are branches with an independent root that obey a Darwinian evolutionary process.

762 **Viral isolate:** Viral population obtained by culture, generally linked to a particular virus
763 sample (48).

764 **Viral strain:** Virus of a given species with stable and heritable biological, serological, and/or
765 molecular characteristics, and/or links to particular hosts, vectors, or symptoms (48, 49).

766 **Viral quasi-species:** The mutant clouds (or swarms) with closely related genomes generated
767 according to a Darwinian evolutionary process. It includes the accumulation of mutations
768 during the replication of RNA viruses through continuous genetic variation, with possible
769 genetic rearrangement by genetic recombination or genome segment reassortment, followed
770 by competition among the mutants generated and selected in a given environment (7, 50-52).

771 **Viral mutant:** The result of nucleotide substitutions (non-synonymous or synonymous),
772 deletions or insertions. These nucleotide changes occur randomly, gradually, and
773 spontaneously, according to the Darwinian model (6, 53, 54) and possibly at sites critical for
774 viral protein conformation, antigenicity, susceptibility to immune responses, and
775 functionality, including ability to interact with other proteins and enzymatic properties (37,
776 55).

777 **Viral variant:** An isolate that differs slightly by its transmission mode or the symptoms it
778 causes from a reference virus (21, 48) or whose consensus genome sequence differs from that
779 of a reference virus (21, 48, 49). Notwithstanding, no definition of the term “variant” is
780 universal or broadly accepted (21, 56).

781 **Box 2. Major SARS-CoV-2 classification systems**

782 **World Health Organization (WHO) classification ([https://www.who.int/en/activities/tracking-](https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/)** 783 **SARS-CoV-2-variants/**):

784 Defines SARS-CoV-2 ‘Variants of Interest’ (VOI) and ‘Variants of Concern’ (VOC). **VOI:** viral
785 genome has mutations with established or suspected phenotypic implications and either has been
786 identified to cause community transmission/multiple COVID-19 cases/clusters or has been detected in
787 multiple countries. **VOC:** a VOI demonstrated through comparative assessment to be associated with
788 (i) increase transmissibility or detrimental change in COVID-19 epidemiology; (ii) increase in
789 virulence or change in clinical disease presentation; or (iii) decrease in effectiveness of public health
790 and social measures or available diagnostics, vaccines, and/or therapeutics. Nomenclature uses letters
791 of the Greek alphabet (currently from Alpha to Theta as of 1 August 2021).

792
793 **GISAID classification ([https://www.gisaid.org/references/statements-clarifications/clade-and-](https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/)**
794 **lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/**) (*I2*): Defines
795 high-level phylogenetic clades informed by the statistical distribution of genome distances in
796 phylogenetic clusters followed by merging of smaller lineages into major clades (eight currently)
797 based on shared marker mutations. Nomenclature uses actual letters of marker mutations (S, L, V, G,
798 GH, GR, GV, GRY).

799
800 **Nextstrain classification ([https://nextstrain.org/blog/2021-01-06-updated-SARS-CoV-2-clade-](https://nextstrain.org/blog/2021-01-06-updated-SARS-CoV-2-clade-naming)**
801 **naming**) (*I3*): Defines major clades that reach $\geq 20\%$ global frequency for \geq two months, $\geq 30\%$
802 regional frequency for \geq two months, or correspond to a ‘variant of concern’ (VOC). Nomenclature
803 uses a year-letter naming system.

804
805 **Pangolin (for Phylogenetic Assignment of Named Global Outbreak LINEages) classification**
806 (<https://cov-lineages.org/pangolin.html>) (*I1*): Based on a maximum likelihood phylogenetic
807 reconstruction with the following criteria: (i) Phylogenetic evidence of emergence from an ancestral
808 lineage into another geographically distinct population; (ii) \geq one shared nucleotide difference(s) from
809 the ancestral lineage; (iii) \geq five genomes; (iv) \geq one shared nucleotide change(s) among genomes from
810 the lineage; and (v) a bootstrap value $>70\%$ for the lineage node. Regarding the nomenclature: a letter
811 (A for the Wuhan-Hu-1 isolate) is given first, followed by numbers for a maximum of three sublevels
812 after which new descendant lineages are given a new letter.

813
814 **Marseille classification:** Defines groups of viral genomes carrying specific sets of \geq five mutations
815 differentiating them from any other viral genomes and obtained from ≥ 30 different patients.
816 Nomenclature uses our city name, “Marseille”, followed by a number according to the chronology of
817 their description (e.g.: Marseille-1, Marseille-2,...) or to one of their hallmark amino acid substitutions
818 (e.g.: Marseille-501).

819 **FIGURE LEGENDS**

820

821 **Fig. 1. Patterns of incidence curves of SARS-CoV-2-associated deaths according to countries.**

822 (A) Hierarchical clustering of incidence curves of SARS-CoV-2-associated deaths per country

823 (B) Major patterns of incidence curves of SARS-CoV-2-associated deaths defined based on

824 hierarchical clustering, and the example of one country for each pattern.

825 Patterns of incidence curves of SARS-CoV-2-associated deaths are shown in 10 panels numbered

826 from 1 to 10; these numbers correspond to those from the cladogram of Figure 1A. The example of

827 one country is shown in a panel at the right of each of the ten patterns.

828 (C) Worldwide distribution of the ten major patterns of incidence curves of SARS-CoV-2-associated

829 deaths defined based on hierarchical clustering.

830 Colours are as used for Figure 1A.

831

832 **Fig. 2. Chronological distribution of SARS-CoV-2 diagnoses by qPCR at IHU Méditerranée**

833 **Infection institute (A) and of mean (\pm standard deviation) numbers of mutations in SARS-CoV-**

834 **2 genomes obtained per month from patients SARS-CoV-2-diagnosed at IHU Méditerranée**

835 **Infection, Marseille (B).**

836 (A) Chronological distribution of SARS-CoV-2 diagnoses is the mean number of diagnoses by

837 qPCR per sliding windows of 7 days and steps of one day.

838 (B) Numbers of mutations in SARS-CoV-2 genomes are calculated in reference to the genome of the

839 Wuhan-Hu-1 isolate (GenBank Accession no. NC_045512.2). Whiskers indicate 10-90 percentiles.

840 ***: $<10^{-3}$; *: $p<10^{-1}$.

841

842 **Fig. 3. Mean (\pm standard deviation) pairwise genetic distance between SARS-CoV-2 genomes**

843 **obtained from patients SARS-CoV-2-diagnosed at IHU Méditerranée Infection, Marseille,**

844 **according to time periods.**

845

846 **Fig. 4. Phylogenetic trees of 10,773 genomic sequences of SARS-CoV-2 obtained from patients**
847 **SARS-CoV-2-diagnosed at IHU Méditerranée Infection, Marseille.**

848 Phylogeny reconstruction was performed using the nextstrain/ncov tool
849 (<https://github.com/nextstrain/ncov>) then visualised with Auspice
850 (<https://docs.nextstrain.org/projects/auspice/en/stable/>). The genome of the original Wuhan-Hu-1
851 coronavirus isolate (GenBank accession no. NC_045512.2) was added as outgroup. Major (most
852 prevalent) variants are labelled. Marseille variants or WHO clades, and Pangolin, or Nextstrain
853 clades are indicated.

854 (A) X-axis shows the number of mutations compared to the genome of the Wuhan-Hu-1 isolate
855 (GenBank accession no. NC_045512.2).

856 (B) Circular representation of the phylogenetic tree based on 10,773 genomic sequences of SARS-
857 CoV-2 obtained from patients SARS-CoV-2-diagnosed at IHU Méditerranée Infection, Marseille.

858

859 **Fig. 5. Microarray showing the distribution along the viral genome and in viral genes of**
860 **nucleotide and amino acid substitutions observed in comparison with the genome of Wuhan-**
861 **Hu-1 isolate for the various viral variants detected in 2020 in respiratory samples from patients**
862 **diagnosed with SARS-CoV-2 infection at IHU Méditerranée Infection.**

863 Sequences from complete genomes that were obtained were analyzed using the Nextstrain web-tool
864 (<https://clades.nextstrain.org/>) (13). Representation is adapted from Nextclade sequence analysis web
865 application output (<https://clades.nextstrain.org/>).

866 ^a In reference to genome GenBank Accession no. NC_045512.2 (Wuhan-Hu-1 isolate); b Color code
867 for nucleotide mutations: Green: U; yellow: G; blue: C; red: A. nsp9: ssRNA-binding protein; nsp14:
868 3'-to-5'; exonuclease; nsp15: EndoRNase; PLproSyn.: synonymous.

869

870 **Fig. 6. Weekly incidence of each SARS-CoV-2 variants among patients SARS-CoV-2-diagnosed**
871 **at IHU Méditerranée Infection institute.**

872 (A) Three-dimensional plot of weekly proportions accounted by each SARS-CoV-2 mutants and
873 variants among patients SARS-CoV-2-diagnosed.

874 (B) Weekly incidence of each SARS-CoV-2 mutants and variants extrapolated to the total number of
875 cases, based on their proportions of genotyped cases.

876

877 **Fig. 7. Demonstrated or likely sources and origins of SARS-CoV-2 mutants or variants detected**
878 **among patients SARS-CoV-2-diagnosed in our institute.**

879

880

881
882

TABLES

Table 1. Nomenclature and description of Marseille (IHU Méditerranée Infection) variants of SARS-CoV-2

Marseille variant (IHU Méditerranée Infection)	Number of genomes in the IHU Méditerranée Infection dataset	Nextstrain clade	Pangolin lineage	WHO classification	Mutation compared to the genome of the Wuhan-Hu-1 isolate (GenBank accession no. NC_045512.2)
Defined					
Marseille-1	88	20A	B.1.416	-	C28833T, G1181T, C1625T, C25886T, G28198T, G28851T
Marseille-2	617	20E (EU1)	B.1.177	-	T445C, C6286T, C22227T, C26801G, C28932T, G29645T, G21255C
Marseille-3	43	20A	B.1	-	C1912T, G5210A, C17470T, C21191T, A23148G, T27125C, C28854T
Marseille-4	2,442	20A.EU2	B.1.160	-	C18877T, C26735T, G5629T, G9526T, C11497T, G13993T, G15766T, G22992A, C25710T, T26876C, G28975C, G29399A, C4543T, A16889G, G17019T
Marseille-5	135	20C	B.1.367	-	C10582T, C28830A, C3099T, G4960T, C4965T, C6070T, C7303T, C7564T, C9246T, C10279T, C10301A, C10525T, G10688T, G11851T, C14230A, G21800T, G27632T, C27804T, G29402T, G29779T
Marseille-8	188	20B	B.1.1.269	-	C5055T, G11851T, A12755G, G24812T, C26895T
Marseille-9	114	20B	B.1.1.241	-	A11782G, T21570G, C21575T, T25473C, C28253T
Marseille-10	86	20A	B.1.221	-	C21855T, A25505G, G25906C, G25996T, C28651T, C28869T, C3602T, C6941T
Marseille-12	55	20A	-	-	G29734C, C17104T, T7767C, C8047T, C22879A
Marseille-501	47	19B	A.27	-	G28878A, G29742A, A361G, C1122T, C2509T, A9204G, A11217G, C16466T, A18366G, A20262G, T22917G, A23063T, C23520T, C23525T, G23948T, G25218T, T25541C, TTAATCCAG26159T, C27247T, TA27385T, AGATTTTC28247A, A28273T
Marseille-484K.V3	84	21D	B.1.525	Eta	C1498T, A1807G, G2659A, C6285T, T8593C, GTCTGGTTTT11287G, C14407T, C18171T, A20724G, A21717G, C21762T, ATACATGT21764AT, TTTATTA21990TTTA, G23012A, G23593C, T24224C, C24748T, C26305T, T26767C, GTTT27204G, TCTG28277T, C28308G, A28699G, C28887T, G29543T, TTTCACCGAGGCCACGCGGAGTACGATCGAGTGTACAGTGAACAATGCTAGG29727TGG
Marseille-452R-19B	51	19B	A.21	-	G28878A, G29742A, T1885A, C5907T, C10138T, G11417T, C11824T, C18129T, G18181A, T21982C, G22132T, T22917G, G23402A, G25687T, G26062T, C28253T, C29686T
Potential					
Marseille-6	11	20A	B.1	-	C9430T, A15477T, C18395T, A20622T, G20623T, A20624T, C23730T, A26319G, C28854T, G29044A
Marseille-7	17	20A	B.1.416.1	-	C28833T, C2706T, C25731T, G27463C
Marseille-11	19	20A	B.1.221	-	C21855T, A25505G, G25906C, G25996T, C28651T, C28869T, T445C, G13514T, G25785T, C25810T, G29511T, C29708T
Marseille-484K.V1-20B	1	20B	R.1	-	C14340T, G17551A, C18877T, A19167G, C19274A, T19839C, G22017T, G23012A, G23868T, T26604C, C27213T, TAAAA28270TAAA, C28833T, G29527T
Marseille-484K.V2-20B	9	20B	B.1.1.318	-	C3961T, C9072T, C9891T, C10116T, GTCTGGTTTT11287G, G20578A, C21846T, TTTATTA21990TTTA, G23012A, T23287C, C23604A, G23948C, C24382T, C25276A, T26767C, AAACGAACATGAAATTT27886AT, G28209T, A28271G, GCTA28895G, C29769T
Marseille-681H-19B	6	19B	A.23	-	G28878A, G29742A, G11230T, G22661T, G23401T, G28167A, G569A, C5178T, C7389T, C10029T, A15498T, C21646T, T23030C, C23604A, G25323T, C25603G, G25785T, T26767C
Marseille-681R-19B	7	19B	A.23.1	-	C9430T, A15477T, C18395T, A20622T, G20623T, A20624T, C23730T, A26319G, C28854T, G29044A
Marseille-4B	7	-	-	-	C18877T, C26735T, G5629T, G9526T, C11497T, G13993T, G15766T, G22992A, C25710T, T26876C, G28975C, G29399A, C1059T, C17634T, T19584C, G19684T, T25081C, G28378A, G29553T
Marseille-20E.2	14	-	-	-	C3646T, C6636T, C2862T, G5629T, A10652G, C13275T, C20844T, C21575T, G22104T, C24866T

883

884 **Table 2.** Number of the main SARS-CoV-2 variants and mutants diagnosed among the IHU

885 Méditerranée Infection cohort of SARS-CoV-2-infected patients as of 18 August 2021

Variant/mutant name	Genotyping approach			Total	% of all genotypes
	Next-generation genome sequencing	Next-generation spike gene fragment sequencing	qPCR		
Alpha (B.1.1.7, a.k.a. UK)	3,506	2,220	3,568	9,294	38.44
Marseille-4 (B.1.160)	2,564	45	3,516	6,125	25.33
Delta (B.1.617.2, a.k.a. Indian)	1,406	0	2,707	4,113	17.01
Marseille-2 (B.1.177)	639	6	205	850	3.52
Beta (B.1.351, a.k.a. South African)	204	210	161	575	2.38
20A/8371T (1 st period)	488	0	0	488	2.02
20A (1 st period)	323	44	0	367	1.52
20C (1 st period)	310	0	0	310	1.28
20B (1 st period)	290	6	0	296	1.22
20A/15324T (1 st period)	176	4	0	180	0.74
Marseille-8 (B.1.1.269)	174	2	0	176	0.73
Marseille-5 (B.1.367)	124	0	0	124	0.51
Marseille-1 (B.1.416)	123	0	0	123	0.51
20A/25563T (1 st period)	123	0	0	123	0.51
Marseille-9 (B.1.1.241)	117	0	0	117	0.48
Marseille-484K.V3 (B.1.525)	112	0	0	112	0.46
20A/20268G (1 st period)	95	0	0	95	0.39
Marseille-10 (B.1.221)	89	2	0	91	0.38
20A/2416T (1 st period)	89	0	0	89	0.37
Gamma (P.1, a.k.a. Brazilian)	70	1	16	87	0.36
Marseille-501 (A.27)	26	48	0	74	0.31
20E (B.1.177)	56	5	0	61	0.25
Marseille-452R-19B (A.21)	43	2	0	45	0.19
Marseille-3 (B.1)	39	0	0	39	0.16
19B (1 st period)	13	23	0	36	0.15
Marseille-12	36	0	0	36	0.15
20D (B.1.1.1)	31	2	0	33	0.14
Marseille-484K.V2 (B.1.1.318)	20	1	0	21	0.09
Marseille-11 (B.1.221)	19	0	0	19	0.08
19A (1 st period)	15	0	0	15	0.06
B.1.214 (a.k.a. Belgian)	15	0	0	15	0.06
Marseille-7 (B.1416.1)	15	0	0	15	0.06
Marseille-6 (B.1)	9	0	0	9	0.04
Marseille-681R-19B (A.23.1)	7	0	0	7	0.03
Marseille-681H-19B (A.23)	6	0	0	6	0.02
Epsilon (B.1.427/B.1.429)	6	0	0	6	0.02
B.1.621	6	0	0	6	0.02
Marseille-484.V1 (A.23)	3	0	0	3	0.01
Total	11,387	2,621	10,173	24,181	100.0

886

Fig. 1A

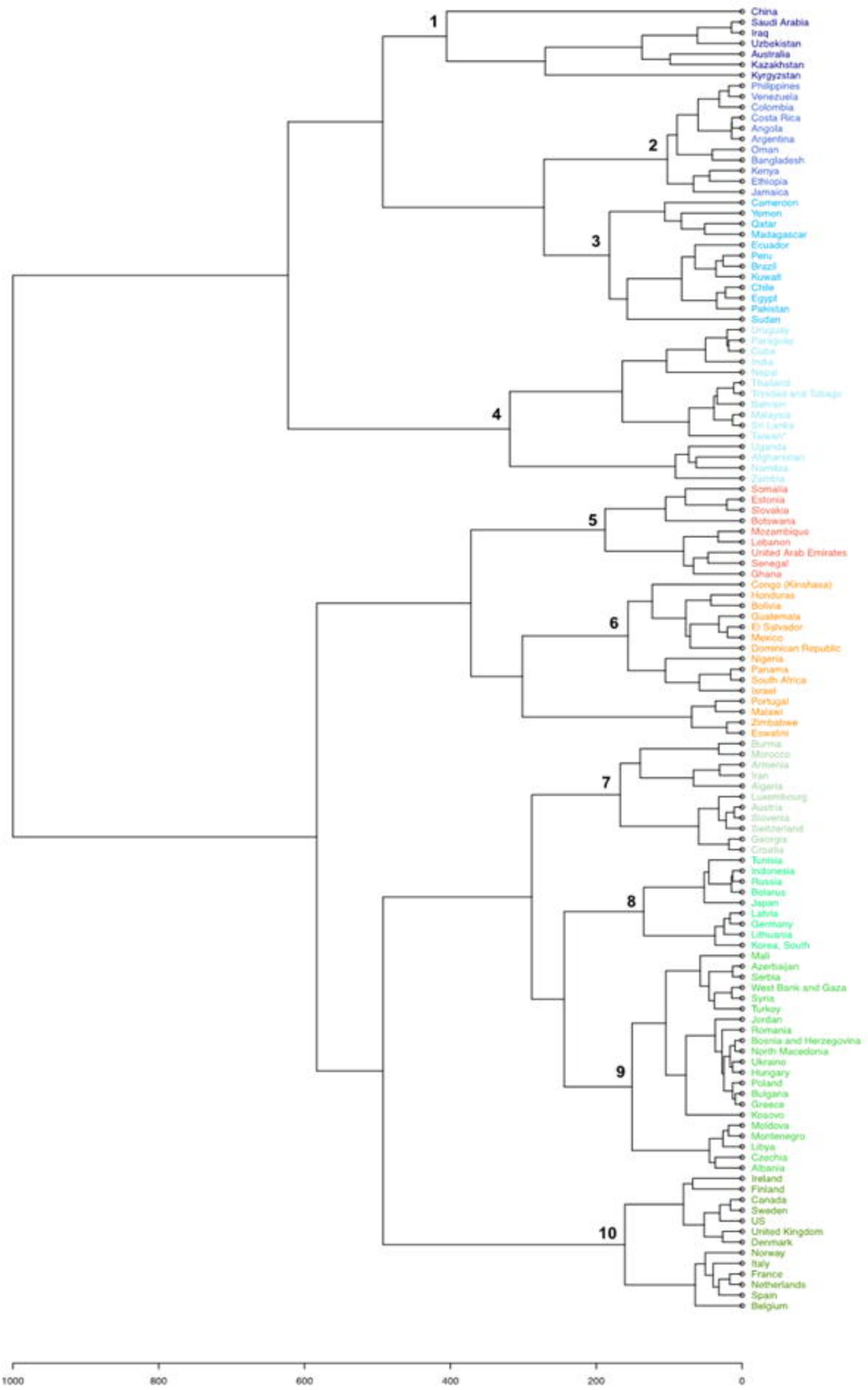


Fig. 1B

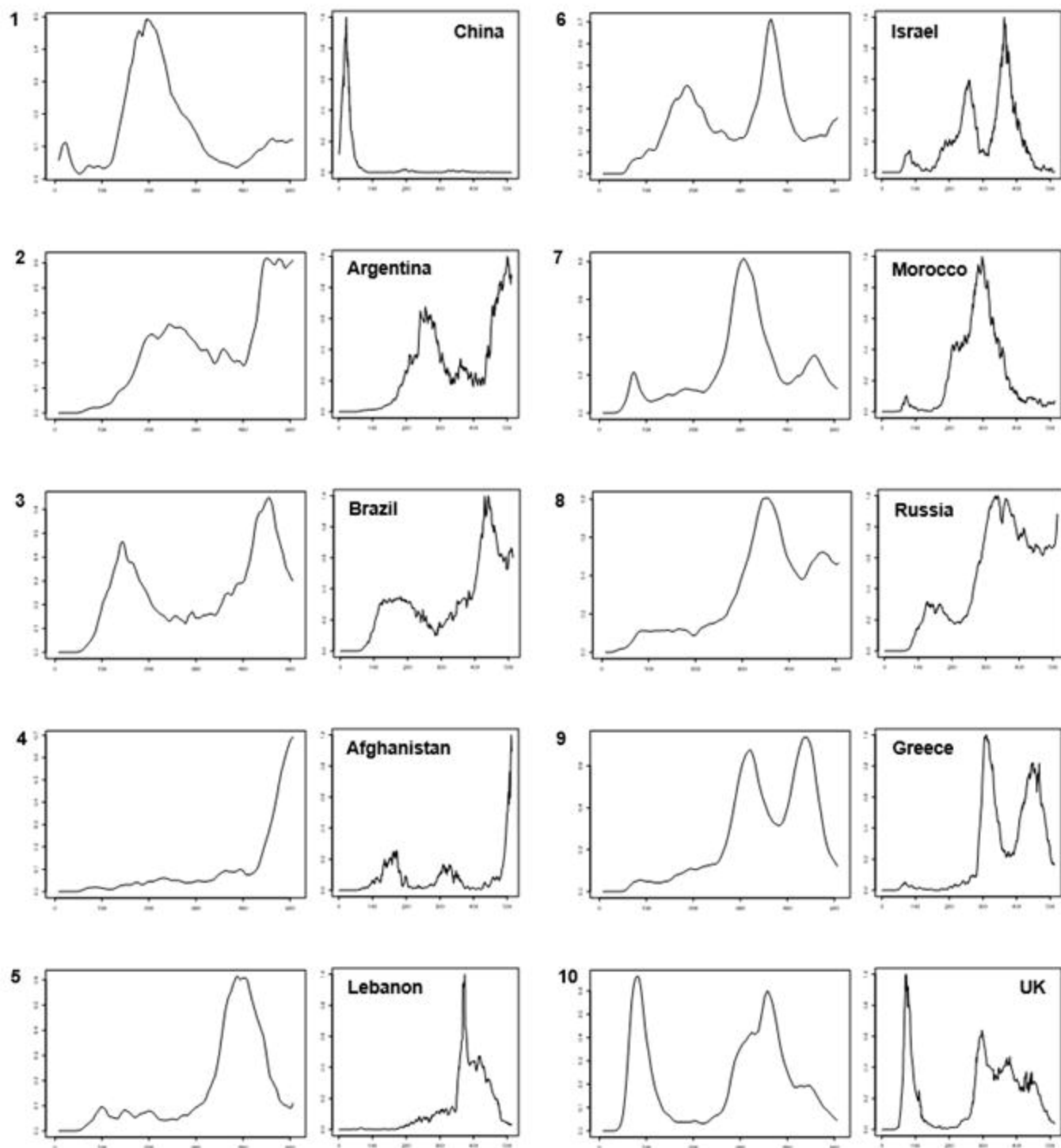


Fig. 1C

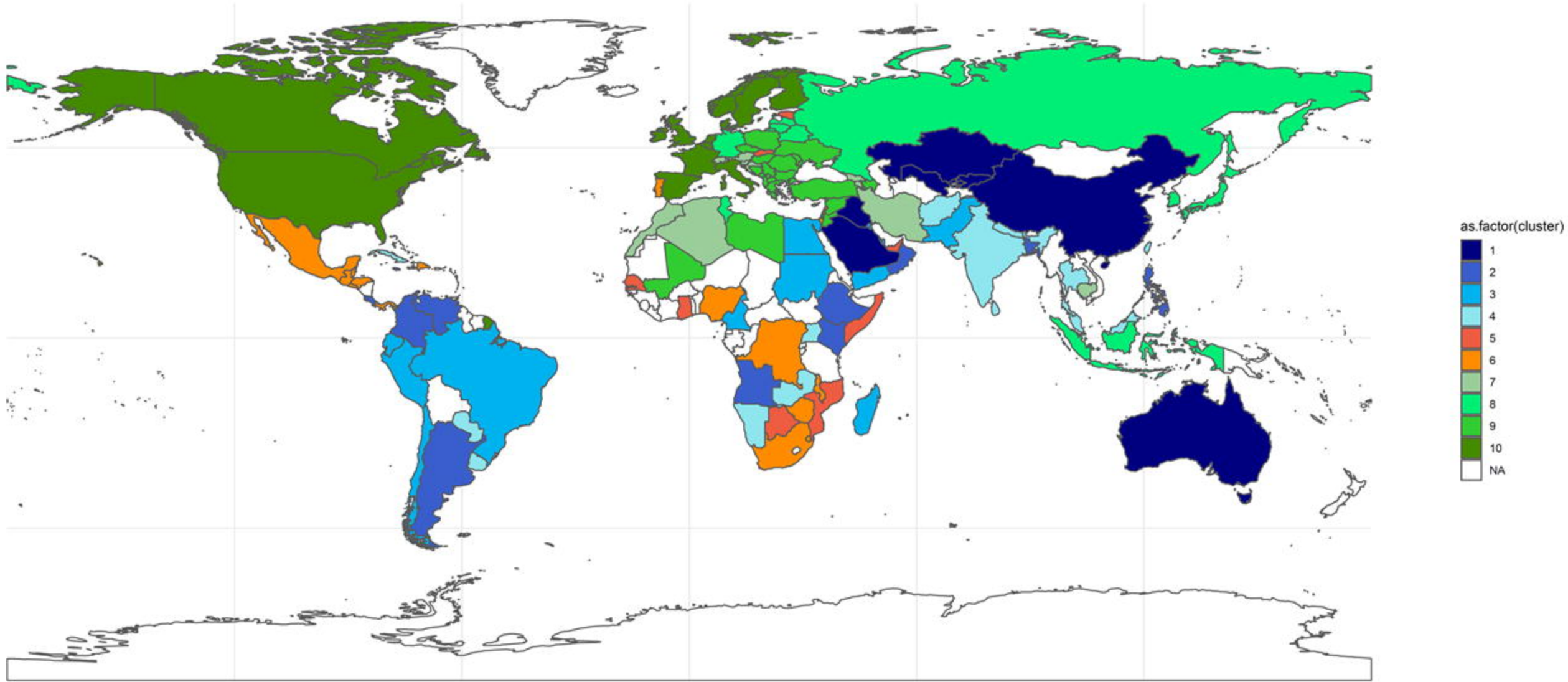
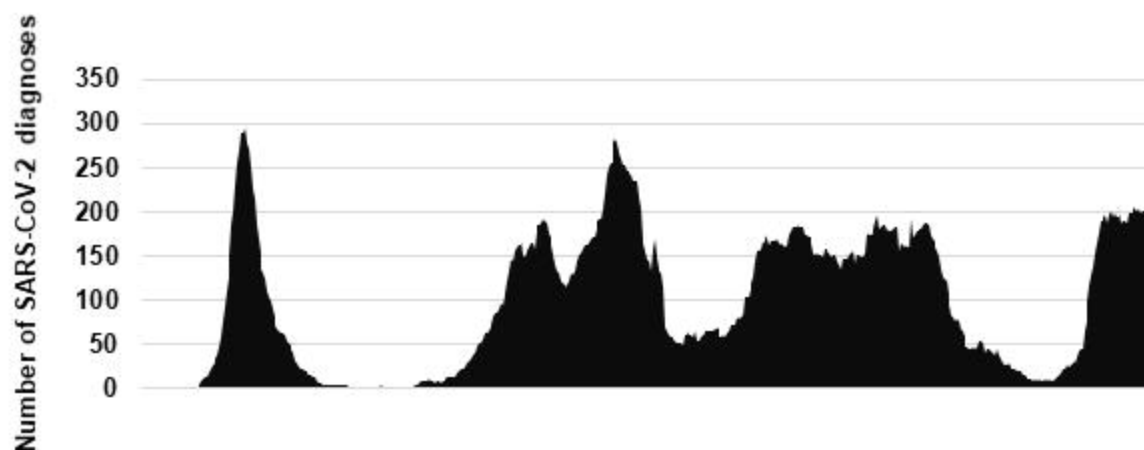


Fig. 2

A



B

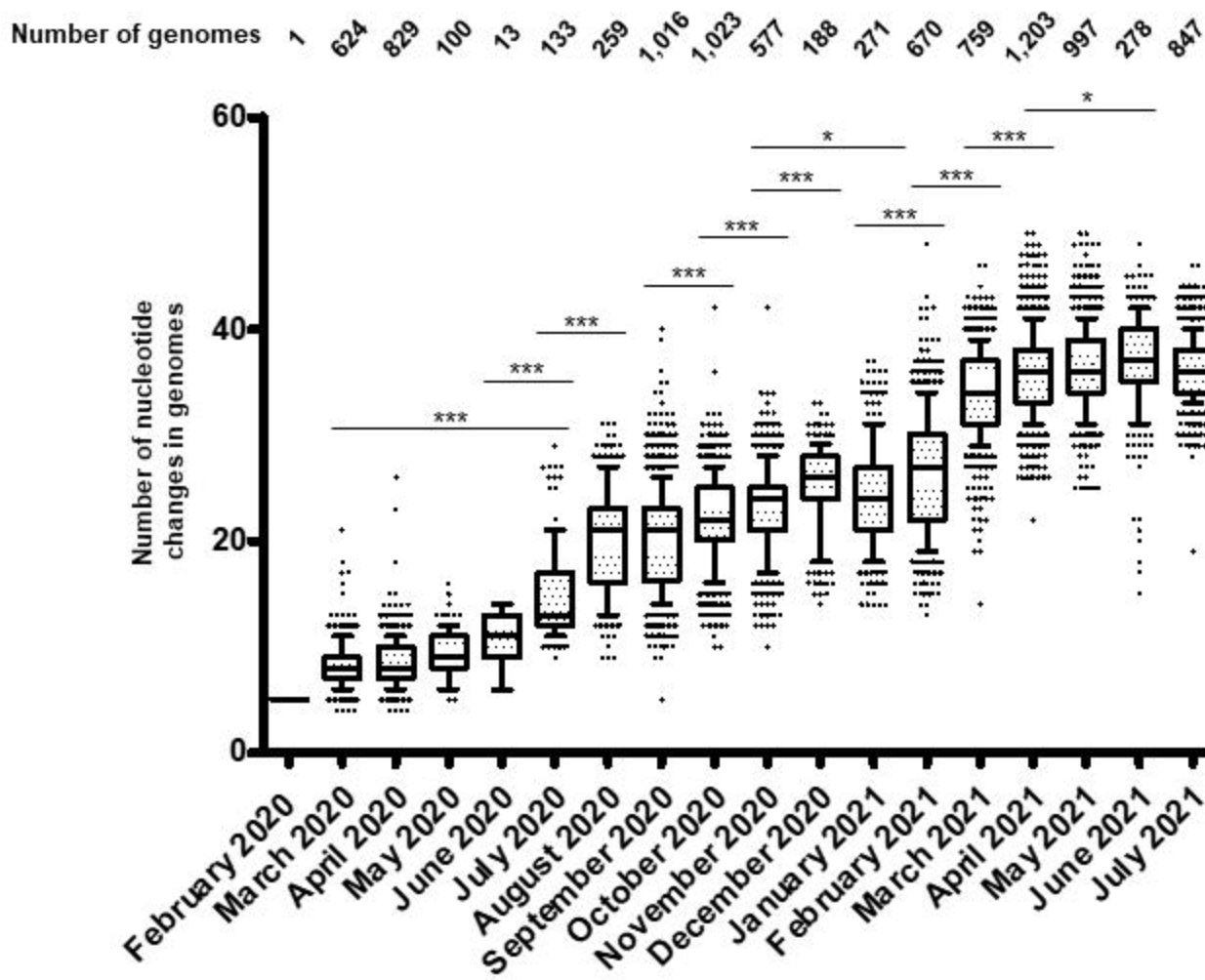


Fig. 3

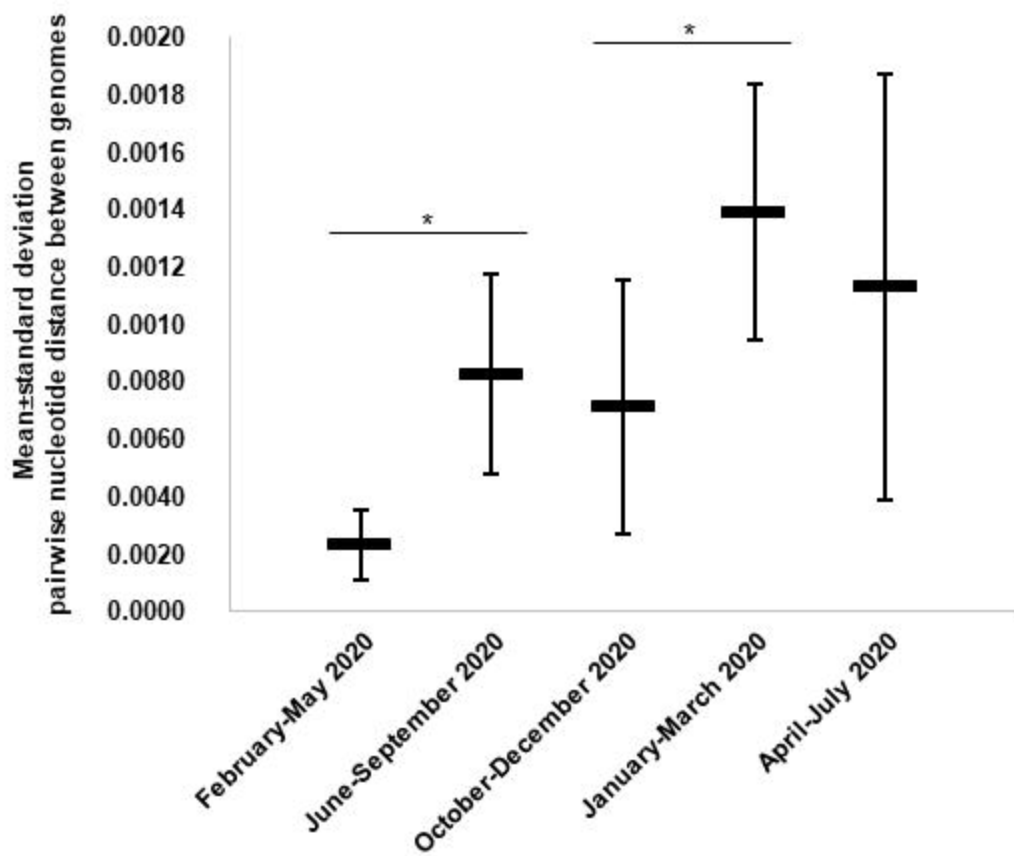


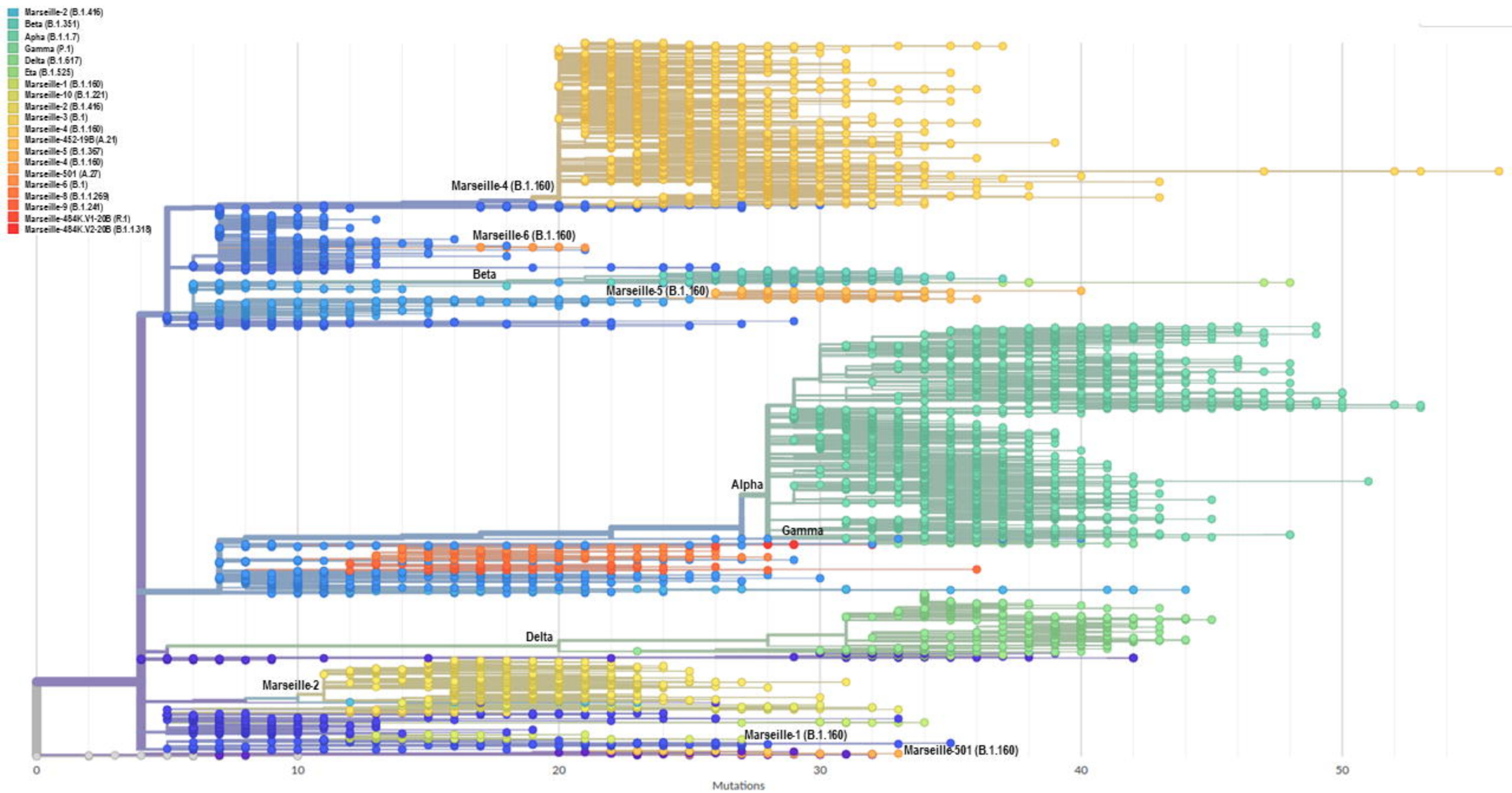
Fig. 4A

Fig. 4B

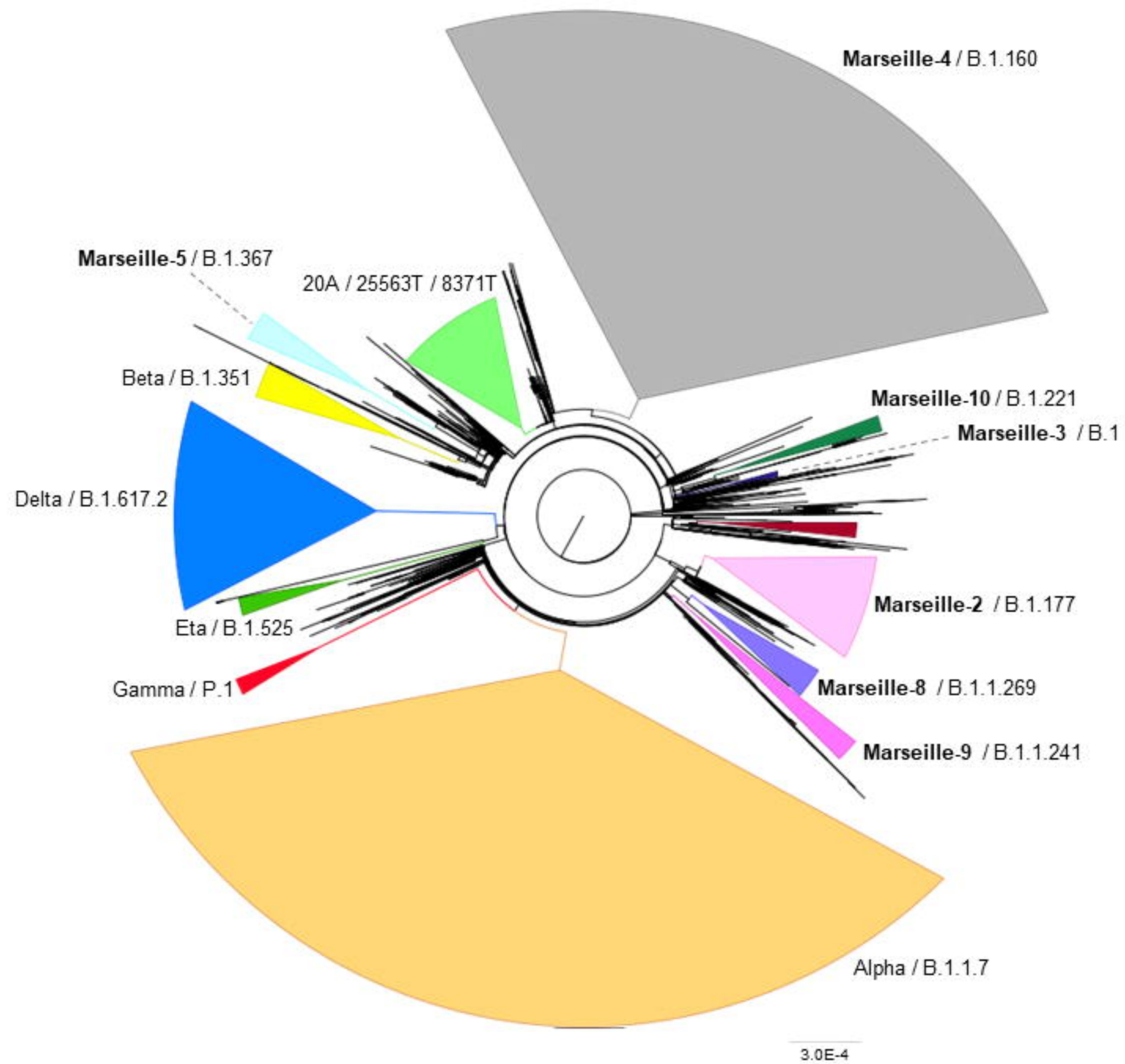


Fig. 5

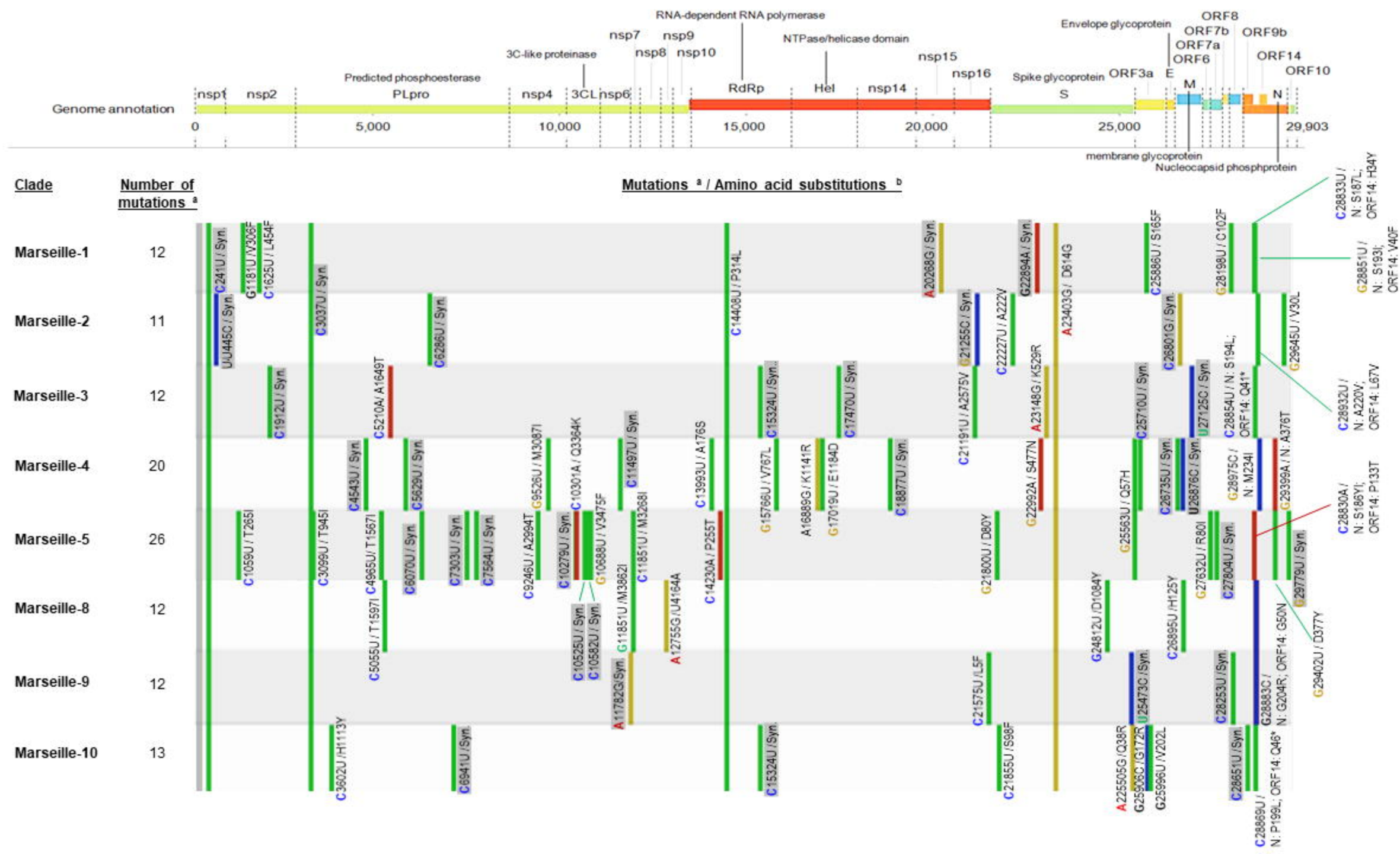


Fig. 6A

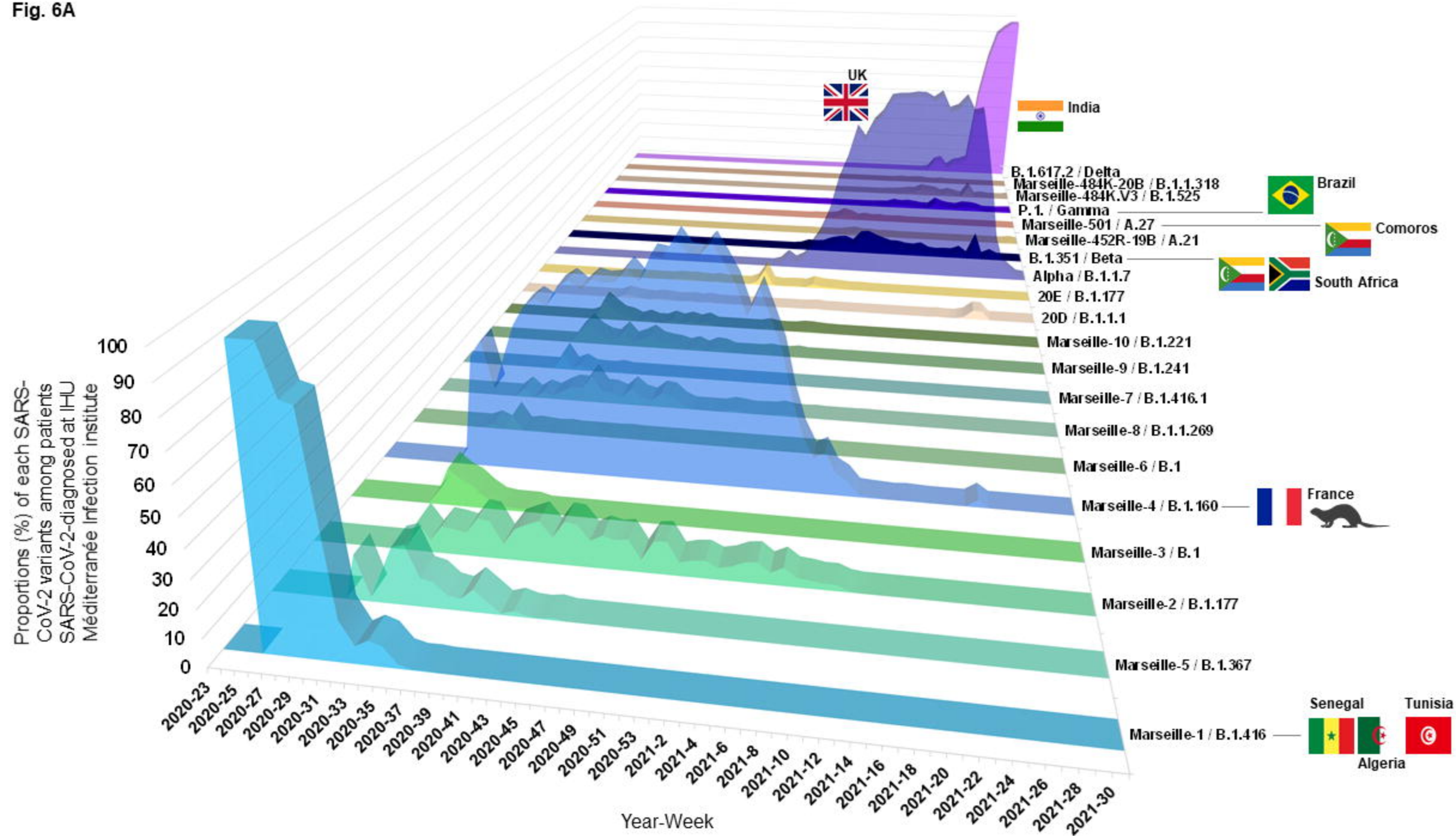


Fig. 6B

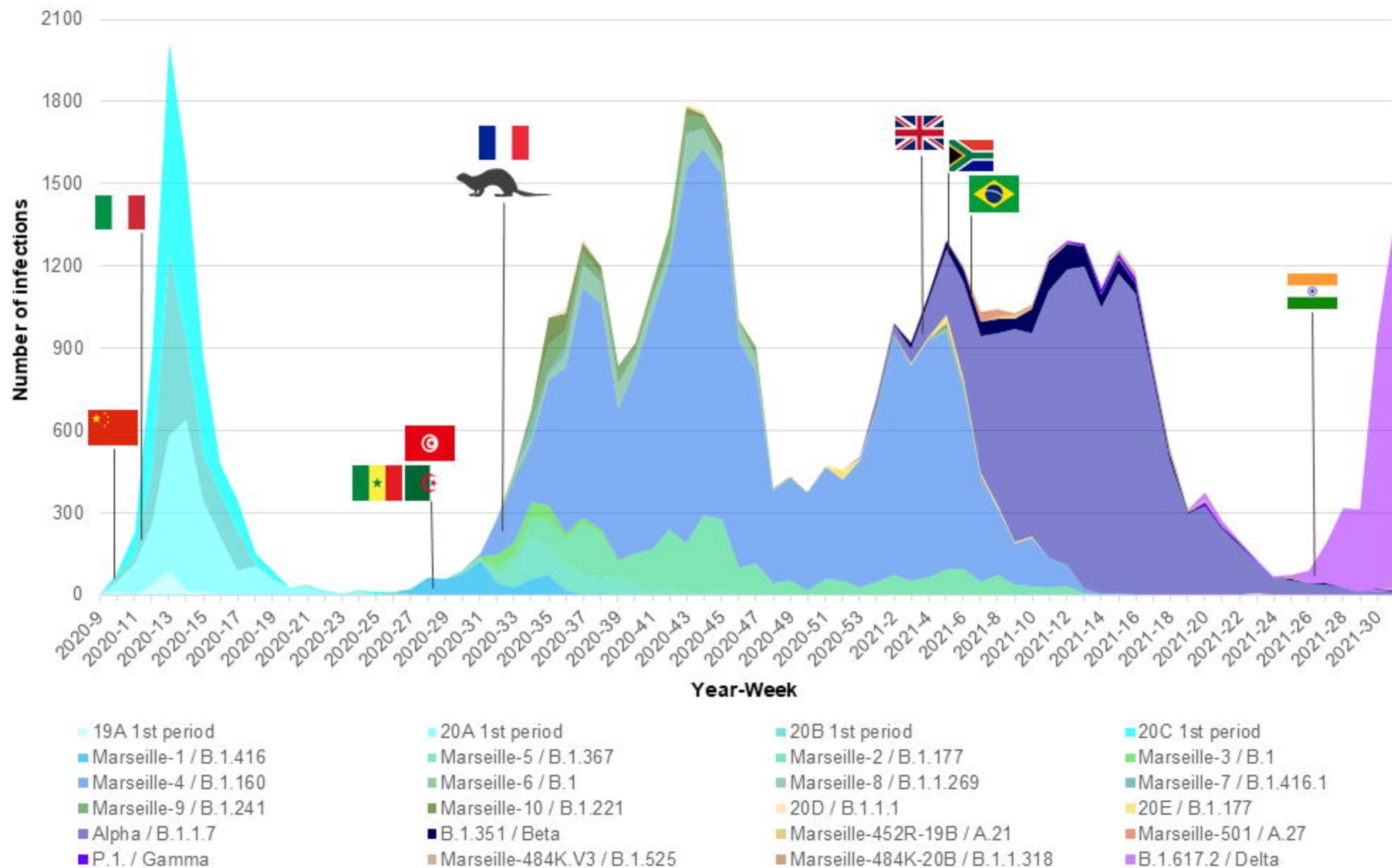


Fig. 7

