

TITLE PAGE

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Type of article: Short communication

Full-length title:

First cases of infection with the 21L/BA.2 Omicron variant in Marseille, France

Short title (for the running head):

Emergence of the 21L/BA.2 Omicron variant in Southern France

Author list: Philippe COLSON^{1,2,3} *, Jérémy DELERCE¹, Mamadou BEYE¹, Anthony LEVASSEUR^{1,2}, Céline BOSCHI^{1,2,3}, Linda HOUHAMDI^{1,3}, Hervé TISSOT-DUPONT^{1,2,3}, Nouara YAHY⁴, Matthieu MILLION^{1,2,3}, Bernard LA SCOLA^{1,2,3}, Jacques FANTINI⁴, Didier RAOULT^{1,2}, Pierre-Edouard FOURNIER^{1,3,5} *

Affiliations: ¹ IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; ² Aix-Marseille Univ., Institut de Recherche pour le Développement (IRD), Microbes Evolution Phylogeny and Infections (MEPHI), 27 boulevard Jean Moulin, 13005 Marseille, France; ³ Assistance Publique-Hôpitaux de Marseille (AP-HM), 264 rue Saint-Pierre, 13005 Marseille, France; ⁴ Aix-Marseille Université, INSERM UMR S 1072, 51 boulevard Pierre Dramard, 13015 Marseille, France; ⁵ Aix-Marseille Univ., Institut de Recherche pour le Développement (IRD), Vecteurs – Infections Tropicales et Méditerranéennes (VITROME), 27 boulevard Jean Moulin, 13005 Marseille, France.

*** Corresponding authors:** Philippe Colson, IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: philippe.colson@univ-amu.fr; Pierre-Edouard Fournier, IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

26 402; email: pierre-edouard.fournier@univ-amu.fr.

27 **Keywords:** SARS-CoV-2, variant, Omicron, travel, emergence, southern France

28 **Word counts:** abstract, 196; text, 1,997

29 **Figures:** 3; **Table:** 1 **References:** 20

30

31

ABSTRACT

32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

The SARS-CoV-2 21K/BA.1, 21L/BA.2, and BA.3 Omicron variants have recently emerged worldwide. To date, the 21L/BA.2 Omicron variant has remained very minority globally but became predominant in Denmark instead of the 21K/BA.1 variant. Here we describe the first cases diagnosed with this variant in south-eastern France. We identified thirteen cases using variant-specific qPCR and next-generation sequencing between 28/11/2021 and 31/01/2022, the first two cases being diagnosed in travellers returning from Tanzania. Overall, viral genomes displayed a mean (\pm standard deviation) number of 65.9 ± 2.5 (range, 61-69) nucleotide substitutions and 31.0 ± 8.3 (27-50) nucleotide deletions, resulting in 49.6 ± 2.2 (45-52) amino acid substitutions (including 28 in the spike protein) and 12.4 ± 1.1 (12-15) amino acid deletions. Phylogeny showed the distribution in three different clusters of these genomes, which were most closely related to genomes from England and South Africa, from Singapore and Nepal, or from France and Denmark. Structural predictions pointed out a significant enlargement and flattening of the 21L/BA.2 N-terminal domain surface compared with that of the 21K/BA.2 Omicron variant, which may facilitate initial viral interactions with lipid rafts. Close surveillance is needed at global, country and center scales to monitor the incidence and clinical outcome of the 21L/BA.2 Omicron variant.

52

TEXT

53

54 **Introduction**

55 SARS-CoV-2 variants have been detected since summer 2020^{1,2} and revealed of
56 critical interest regarding viral transmissibility, load, and escape to natural or vaccine
57 immunity.^{3,4} The Omicron variant is currently the predominant variant of concern in many
58 countries worldwide (<https://covariants.org/per-country>).^{5,6} It has been reported to exhibit
59 considerable escape to antibodies elicited by vaccination^{7,8} and to be associated with lower
60 clinical severity including in our center.^{8,9} It was first detected in early November in
61 Botswana and thereafter, in many countries, its incidence has rapidly exceeded that of the
62 Delta variant that had predominated since the summer of 2021 ([https://covariants.org/per-](https://covariants.org/per-country)
63 [country](https://covariants.org/per-country)).^{5,6} As a matter of fact, Omicron is composed of three branches corresponding to
64 three variants named Nextstrain clade^{10,11} 21K (or Pangolin lineage¹² BA.1), 21L (or BA.2),
65 and lineage BA.3 that is part of the 21M Omicron clade. Until very recently, unlike the
66 21K/BA.1 Omicron variant, the 21L/BA.2 Omicron variant has remained a very minority in
67 most countries and globally, including in South Africa from where it seems to originate. Here
68 we describe the emergence in south-eastern France of this variant.

69

70 **Materials and methods**

71 Nasopharyngeal samples were collected from patients in our university hospital
72 institute (Méditerranée Infection; <https://www.mediterranee-infection.com/>) and tested for
73 SARS-CoV-2 infection by real-time reverse transcription PCR (qPCR) as previously
74 described.^{2,13} Then qPCR assays specific of variants were performed according to French
75 recommendation, as previously reported.^{2,13,14} This included detection of spike mutations
76 L452R, K417N, E484K, and/or P681H (Thermo Fisher Scientific, Waltham, USA), combined

77 with testing with the TaqPath COVID-19 kit (Thermo Fisher Scientific) that target viral genes
78 ORF1, N (nucleocapsid) and S (spike).

79 Genomic identification of the 21L/BA.2 Omicron variant was performed through next-
80 generation sequencing with the Oxford Nanopore technology (ONT) on a GridION instrument
81 (Oxford Nanopore Technologies Ltd., Oxford, UK) or with the Illumina COVID-seq protocol
82 on the NovaSeq 6000 instrument (Illumina Inc., San Diego, CA, USA), as previously
83 described.^{2,13,14} Sequence read processing and genome analysis were performed as previously
84 described.^{2,13,14} Fastq files were processed differently according to the sequencing technology.
85 Briefly, for ONT reads, fastq files were processed with the ARTIC field bioinformatics
86 pipeline (<https://github.com/artic-network/fieldbioinformatics>). Sequencing reads were
87 basecalled with Guppy (v.4.0.14) and aligned to the Wuhan-Hu-1 genome GenBank accession
88 no. NC_045512.2 using minimap2 (v2.17-r941) (<https://github.com/lh3/minimap2>). Reads
89 were cleaned with Guppyplex. Mapping was cleaned with ARTIC align_trim. Variant calling
90 was performed using Medaka and Longshot. Consensus genome sequences were built with
91 Bcftools. Illumina NovaSeq reads were basecalled with the Dragen Bcl Convert pipeline
92 (v3.9.3; https://emea.support.illumina.com/sequencing/sequencing_software/bcl-convert.html
93 (Illumina Inc.)), mapping was performed with the bwa-mem2 tool ([https://github.com/bwa-
94 mem2/bwa-mem2](https://github.com/bwa-)) on the Wuhan-Hu-1 genome. Mapping was cleaned with Samtools
95 (<https://www.htslib.org/>). Variant calling was performed with freebayes
96 (<https://github.com/freebayes/freebayes>) and consensus genomes were built with Bcftools
97 (<https://samtools.github.io/bcftools/bcftools.html>).

98 Nucleotide and of amino acid changes in viral genomes relatively to the Wuhan-Hu-1
99 isolate genome were obtained using the Nextclade tool (<https://clades.nextstrain.org/>).^{10,11}
100 Nextstrain clades and Pangolin lineages were determined using the Nextclade web application
101 (<https://clades.nextstrain.org/>)^{10,11} and Pangolin web application ([5](https://cov-</p></div><div data-bbox=)

102 lineages.org/pangolin.html)¹², respectively. Genome sequences described here were deposited
103 in the GISAID sequence database (<https://www.gisaid.org/>) (Table 1).¹⁵ Finally, phylogeny
104 was reconstructed with the nextstrain/ncov tool (<https://github.com/nextstrain/ncov>) then
105 visualized with Auspice (<https://docs.nextstrain.org/projects/auspice/en/stable/>). The genomes
106 the closest genetically to those obtained here were selected using Usher
107 (<https://genome.ucsc.edu/cgi-bin/hgPhyloPlace>) and the GISAID BLAST tool
108 (<https://www.epicov.org/epi3/>) then incorporated in phylogeny with all 21L/BA.2 Omicron
109 variant genomes from France available in GISAID.

110 This study was approved by the ethics committee of University Hospital Institute
111 Méditerranée Infection (N°2022-008). Access to patients' biological and registry data issued
112 from the hospital information system was approved by the data protection committee of
113 Assistance Publique-Hôpitaux de Marseille and recorded in the European General Data
114 Protection Regulation registry under number RGPD/APHM 2019-73.

115

116 **Results**

117 Thirteen infections with the 21L/BA.2 Omicron variant were diagnosed in our
118 university hospital institute from patients sampled between 27/12/2021 and 31/01/2022 (Table
119 1). First cases were in two spouses in their 60s diagnosed late 2021 five days after returning
120 from a travel in Zanzibar, Tanzania. They received a third dose of Pfizer-BioNTech COVID-
121 19 vaccine three weeks before diagnosis. The third case had contacts with migrant patients of
122 unknow SARS-CoV-2 status and a SARS-CoV-2-positive case (not tested in our institute)
123 who met students from different countries. This third patient received a third dose of Pfizer-
124 BioNTech COVID-19 vaccine seven weeks before diagnosis. Two other patients originate
125 from the Netherlands and the United Kingdom. No information was available for the other
126 eight patients.

127 All 21L/BA.2 Omicron variant-positive respiratory samples exhibited the same
128 combination of spike mutations as screened by real-time qPCR: negativity for L452R, and,
129 when performed, positivity for K417N and P681H and negativity for E484K and P681R
130 (Table 1). In addition, the TaqPath COVID-19 kit (Thermo Fisher Scientific, Waltham, USA)
131 provided positive signals for all three genes targeted (ORF1, S, and N), except for one sample
132 that showed positivity for the N gene but negativity for both ORF1 and S genes, which was
133 most likely due to a low viral load (qPCR cycle threshold, 32). Thus, 21L/BA.2 Omicron
134 variant-infected patients could be distinguished by qPCR screening from the Delta (L452R-
135 positive) and Omicron 21K (negative for S gene detection by the TaqPath COVID-19 assay)
136 variants that co-circulated in southern France at the time of Omicron 21L/BA.2 emergence.

137 Thirteen 21L/BA.2 Omicron genomes were obtained. Analysis of those larger than
138 28,000 nucleotides showed the presence of a mean (\pm standard deviation) of 65.9 ± 2.5 (range,
139 61-69) nucleotide substitutions and 31.0 ± 8.3 (27-50) nucleotide deletions, which resulted in
140 49.6 ± 2.2 (45-52) amino acid substitutions and 12.4 ± 1.1 (12-15) amino acid deletions. All nine
141 patients' viruses harboured the same set of 28 amino acid substitutions and three contiguous
142 amino acid deletions in their spike protein (Figure 1a). In addition, they shared the same set of
143 amino acid substitutions: these included (i) seven substitutions located in other structural
144 proteins (4, 2 and one in the nucleocapsid, membrane, and envelope proteins, respectively);
145 (ii) 12 substitutions located in non-structural proteins including four in Nsp4, two in Nsp3 (a
146 papain-like protease with phosphatase activity¹⁶, and one each in Nsp1, Nsp5 (a 3C-like
147 proteinase), Nsp12 (RNA-dependent RNA polymerase), Nsp13 (helicase), Nsp14 (3'-5'-
148 exonuclease with proofreading activity), and Nsp15 (an endoribonuclease); and (iii) one
149 substitution located in ORF9b, a regulatory protein. Finally, three contiguous amino acid
150 deletions were located in the nucleocapsid protein and three others were located in ORF9b. Of
151 the 28 amino acid substitutions present in the spike of the 21L/BA.2 Omicron variant, 20 are

152 shared with the 21K/BA.1 as well as the BA.3 Omicron variants

153 (<https://covariants.org/variants/>) (Figure 1a).^{5,6,17}

154 Phylogeny performed with the nextstrain/ncov tool

155 (<https://github.com/nextstrain/ncov>) shows that the nine 21L/BA.2 Omicron variant genomes

156 obtained in our institute were part of three clusters. Two genomes that were retrieved from the

157 two patients who travelled in Tanzania were clustered with genomes obtained in England and

158 South Africa (Figure 2). The genome retrieved from the Dutch patient was clustered with two

159 genomes obtained in Nepal and Singapore. All other six genomes were most closely related to

160 genomes from France and Denmark. As the first two cases we diagnosed were most likely

161 infected with the 21L/BA.2 Omicron variant during their travel in Tanzania, we sought for

162 this variant in GISAID among genomes from this country, but as of 02/02/2022 only three

163 genomes (EPI_ISL_8917336, EPI_ISL_8917337, EPI_ISL_9391124) were available from

164 this country: they were obtained from samples collected in December 2021 and belong to the

165 21K/BA.1 Omicron variant.

166 The earliest 21L/BA.2 Omicron variant genome available from GISAID was obtained

167 in South Africa from a sample collected on 17/11/2021 (EPI_ISL_6795834). As of

168 02/02/2022, most of the 37,521 21L/BA.2 Omicron variant genomes were obtained in

169 Denmark (n= 24,138; 64%) (Figure 3a). Other countries with the greatest number of genomes

170 were United Kingdom (n= 4,637 cases; 12%), India (n= 3,073 cases; 8%), Germany (n= 1,104

171 cases; 2.9%), and Philippines (n= 890 cases; 2.4%). Overall, Europe, Asia, North America,

172 Africa and Oceania accounted for 34,498, 5,071, 398, 377, and 184 genomes, respectively.

173 South Africa, where the 21L/BA.2 Omicron was first described, and Botswana accounted for

174 only 304 and 62 genomes, respectively, amongst 5,550 and 1,449 genomes deposited in

175 GISAID and obtained from samples collected since 01/12/2021, respectively. Finally, only 86

176 genomes (0.2%) were available for France out of 38,350 genomes deposited in GISAID and

177 obtained from samples collected since 01/12/2021, while 18,219 21K/BA.1 Omicron variant
178 genomes (48%) were available for the same period of time.

179 Molecular modeling of Omicron 21L/BA.2 spike protein was performed as previously
180 described¹⁸ by introducing the appropriate mutations and deletions in the framework of a
181 complete 14-1,200 structure of the original 20B SARS-CoV-2 (Wuhan-Hu-1 isolate with
182 D614G substitution)¹⁸ and by incorporating the missing amino acids with the Robetta protein
183 structure prediction tool [<https://rosetta.bakerlab.org/>] before energy minimization with the
184 Polak-Ribière algorithm (Figure 1b).¹⁹ The new 21L/BA.2 Omicron variant displays several
185 common structural features with its close relative, the 21K/BA.1 Omicron variant: many
186 mutations exist that are chiefly distributed in the N-terminal domain (NTD), the receptor
187 binding domain (RBD), and the S1-S2 cleavage site. As for the 21K/BA.1 Omicron variant,
188 the electrostatic surface potential of the RBD is mostly positive, whereas the NTD is
189 constituted by a patchwork of electronegative, electropositive, and neutral regions. A key
190 difference between both 21L/BA.2 and 21K/BA.1 Omicron spike proteins is the significant
191 enlargement and flattening of the 21L/BA.2 Omicron NTD surface compared with that of the
192 21K/BA.1 Omicron variant.¹⁸ This structural change is due to the lack of deletion 143-145 in
193 the 21L/BA.2 Omicron variant. The flat surface of the 21L/BA.2 Omicron NTD may facilitate
194 the initial interaction of the virus with lipid rafts,¹⁹ especially since the surface gain
195 corresponds to an electropositive area (located on the left of the NTD in Figure 1b). Overall,
196 one could hypothesize that the 21L/BA.2 Omicron variant NTD is better adapted to the
197 electronegative surface of lipid rafts than that of the 21K/BA.1 Omicron variant.

198

199 **Discussion**

200 It is currently unknown if this 21L/BA.2 Omicron variant would rise considerably in
201 prevalence and compete the currently predominant Omicron 21K/BA.1, which has spread

202 massively and quickly in countries with a high level of vaccine coverage.⁵ However, the very
203 recent rise of the 21L/BA.2 Omicron variant in Denmark where it became predominant over
204 the 21K/BA.1 Omicron variant that predominated until then suggests that such
205 epidemiological change may occur in other countries worldwide (Figure 3b, 3c).²⁰ In our
206 institute we diagnosed 16,285 SARS-CoV-2 infections between 28/11/2021 (first detection of
207 the Omicron variant) and 02/02/2022, during which 66% of infections were identified as due
208 to the 21K/BA.1 Omicron variant. A first study conducted in Denmark has reported a higher
209 contagiousness with the 21L/BA.2 Omicron variant (n= 2,122 primary household patients)
210 than with the Omicron 21K/BA.1 variant (n= 5,702 primary household patients).²⁰ Secondary
211 attack rates were 39% and 29% among households, respectively, and susceptibility to
212 infection was reported to be significantly increased for unvaccinated (odd ratio (OR), 2.2) as
213 well as full-vaccinated (2.5) and booster-vaccinated (3.0) people. No data to our knowledge is
214 currently available regarding the frequency of asymptomatic and mild and severe clinical
215 forms with this 21L/BA.2 Omicron variant.

216 As for the 21K/BA.1 and 21M/BA.3 Omicron variants, the origin of the 21L/BA.2
217 Omicron variant is currently unclear. The great number of amino acid substitutions in the
218 spike protein and receptor binding domain of these viruses has fuelled several hypotheses that
219 include overlooked virus evolution in people with low access to viral diagnosis and genome
220 sequencing, in an immunocompromised chronically-infected patient, or in animals.⁵ A closest
221 known Omicron's ancestor has been estimated to date back to mid-2020.⁵ Another finding is
222 that despite the tremendous amount of genome sequences available in GISAID (7,790,928 as
223 of 02/02/2022) we are still unable to predict the emergence, and outcome of new variants.
224 This supports the real-time close surveillance of the emergence, spread and vanishing of
225 SARS-CoV-2 variants through molecular and genomic surveillance. It is also worthy of
226 interest to assess phenotypically through inoculation on permissive cells the susceptibility of

227 emerging variants to neutralization by anti-spike antibodies elicited by prior infection or by
228 vaccination, which is on-going in our laboratory for the 21L/BA.2 Omicron variant.

229

230

231 **Acknowledgments**

232 We are grateful to Ludivine Brechard, Claudia Andrieu, Emilie Burel, Elsa Prudent, Céline
233 Gazin, and Marielle Bedotto for their technical help.

234 **Funding**

235 This work was supported by the French Government under the “Investments for the Future”
236 program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-
237 IAHU-03, and was also supported by Région Provence Alpes Côte d’Azur and European
238 funding FEDER PRIMMI (Fonds Européen de Développement Régional-Plateformes de
239 Recherche et d’Innovation Mutualisées Méditerranée Infection), FEDER PA 0000320
240 PRIMMI, by the Emergen French consortium
241 (<https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/consortium-emergen>).

242 **Competing interests**

243 All authors have no conflicts of interest to declare. Didier Raoult has been a consultant for
244 Hitachi High-Technologies Corporation, Tokyo, Japan from 2018 to 2020. He is a scientific
245 board member of Eurofins company and a founder of a microbial culture company (Culture
246 Top). Funding sources had no role in the design and conduct of the study; collection,
247 management, analysis, and interpretation of the data; and preparation, review, or approval of
248 the manuscript.

249 **Author contributions**

250 Study conception and design: Philippe Colson, Didier Raoult, Jacques Fantini, and Pierre-
251 Edouard Fournier. Materials, data and analysis tools: Philippe Colson, Jeremy Delerce,

252 Mamadou Beye, Anthony Levasseur, Céline Boschi, Linda Houhamdi, Hervé Tissot-Dupont,
253 Nouara Yahi, Matthieu Million, Jacques Fantini. Data analyses: Philippe Colson, Pierre-
254 Edouard Fournier, Bernard La Scola, Didier Raoult, Jeremy Delerce, Mamadou Beye,
255 Anthony Levasseur, Jacques Fantini, and Nouara Yahi. Writing of the first draft of the
256 manuscript: Philippe Colson, Jacques Fantini, and Pierre-Edouard Fournier. All authors read,
257 commented on, and approved the final manuscript.

258 **Data availability**

259 The dataset generated and analyzed during the current study are available in the GISAID
260 database (<https://www.gisaid.org/>).

261 **Ethics approval**

262 This study has been approved by the ethics committee of University Hospital Institute (IHU)
263 Méditerranée Infection (N°2022-008). Access to the patients' biological and registry data
264 issued from the hospital information system was approved by the data protection committee
265 of Assistance Publique-Hôpitaux de Marseille (APHM) and was recorded in the European
266 General Data Protection Regulation registry under number RGPD/APHM 2019-73.

267

268

269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292

REFERENCES

1. Lemey P, Ruktanonchai N, Hong SL, et al. Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature* 2021;595:713-7.
2. Colson P, Fournier PE, Chaudet H, et al. Analysis of SARS-CoV-2 variants from 24,181 patients exemplifies the role of globalisation and zoonosis in pandemics. *Front Microbiol* 2022; in press.
3. Li J, Lai S, Gao GF, Shi W. 2021. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* 600:408-418.
4. Harvey WT, Carabelli AM, Jackson B, et al. COVID-19 Genomics UK (COG-UK) Consortium, Peacock SJ, Robertson DL. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021;19:409-424.
5. Mallapaty S. Where did Omicron come from? Three key theories. *Nature* 2022;602:26-28.
6. Hodcroft E. 2012. CoVariants: SARS-CoV-2 mutations and variants of interest. Available from: <https://covariants.org/>.
7. Cao Y, Wang J, Jian F, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* 2021 Dec 23. doi: 10.1038/s41586-021-04385-3. Epub ahead of print.
8. Espenhain L, Funk T, Overvad M, et al. Epidemiological characterisation of the first 785 SARS-CoV-2 Omicron variant cases in Denmark, December 2021. *Euro Surveill.* 2021 Dec;26(50). doi: 10.2807/1560-7917.ES.2021.26.50.2101146.
9. Houhamdi L, Gautret P, Hoang VT, Fournier PE, Colson P, Raoult D. Characteristics of the first 1119 SARS-CoV-2 Omicron variant cases, in Marseille, France, November-

- 293 December 2021. *J Med Virol* 2022 Jan 20. doi: 10.1002/jmv.27613. Epub ahead of
294 print.
- 295 10. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen
296 evolution. *Bioinformatics* 2018;34:4121-4123.
- 297 11. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment,
298 mutation calling and quality control for viral genomes. *Zenodo* 2021.
299 <https://doi.org/10.5281/zenodo.5607694>.
- 300 12. Rambaut A, Holmes EC, O'Toole A, et al. A dynamic nomenclature proposal for SARS-
301 CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;5:1403-1407.
- 302 13. Colson P, Delerce J, Burel E, et al. Emergence in Southern France of a new SARS-
303 CoV-2 variant of probably Cameroonian origin harbouring both substitutions N501Y
304 and E484K in the spike protein. *medRxiv* 2021.12.24.21268174; doi:
305 <https://doi.org/10.1101/2021.12.24.21268174>.
- 306 14. Colson P, Gautret P, Delerce J, et al. The emergence, spread and vanishing of a French
307 SARS-CoV-2 variant exemplifies the fate of RNA virus epidemics and obeys the Black
308 Queen rule. *medRxiv* 2022.01.04.22268715; doi:
309 <https://doi.org/10.1101/2022.01.04.22268715>
- 310 15. Alm E, Broberg EK, Connor T, et al. WHO European Region sequencing laboratories
311 and GISAID EpiCoV group (2020) Geographical and temporal distribution of SARS-
312 CoV-2 clades in the WHO European Region, January to June 2020. *Euro. Surveill*
313 25:2001410.
- 314 16. Prates ET, Garvin MR, Pavicic M, et al. Potential Pathogenicity Determinants Identified
315 from Structural Proteomics of SARS-CoV and SARS-CoV-2. *Mol Biol Evol*
316 2021;38:702-715.

- 317 17. Desingu PA, Nagarajan K, Dhama K. Emergence of Omicron third lineage BA.3 and its
318 importance. *J Med Virol*. 2022 Jan 18. doi: 10.1002/jmv.27601. Epub ahead of print.
- 319 18. Fantini J, Yahi N, Colson P, Chahinian H, La Scola B, Raoult D. The puzzling
320 mutational landscape of the SARS-2-variant Omicron. *J Med Virol* 2022 Jan 8. doi:
321 10.1002/jmv.27577. Epub ahead of print.
- 322 19. Fantini J, Yahi N, Azzaz F, Chahinian H. Structural dynamics of SARS-CoV-2 variants:
323 A health monitoring strategy for anticipating Covid-19 outbreaks. *J Infect* 2021;83:197-
324 206. doi: 10.1016/j.jinf.2021.06.001. Epub 2021 Jun 3.
- 325 20. Plesner Lyngse F, Thure Kirkeby C, Denwood M, et al. Transmission of SARS-CoV-2
326 Omicron VOC subvariants BA.1 and BA.2: Evidence from Danish Households.
327 *medRxiv* 2022.01.28.22270044; doi: <https://doi.org/10.1101/2022.01.28.22270044>.
- 328
- 329

330

FIGURE LEGENDS

331

332 **Figure 1. Map of the Omicron 21L/BA.2 spike protein with signature amino acid**
333 **substitutions and deletions (a) and structural features of 21L/BA.2 Omicron variant**
334 **spike protein (b)**

335 a: Amino acid substitutions and deletions shared with the 21K/BA.1 Omicron variant are
336 indicated by a red font.

337 b: Structural model of the Omicron 21L/BA.2 spike protein with mutations highlighted in red
338 atomic spheres (left panel) or in electrostatic surface rendering (right panel). Note the flat
339 surface of the N-terminal domain that faces lipid rafts of the host cell membrane. The S1-S2
340 cleavage site is indicated by an arrow. The color scale for the electrostatic surface potential
341 (negative in red, positive in blue, neutral in white) is indicated.

342 NTD, N-terminal domain; RBD, Receptor binding domain.

343

344 **Figure 2. Phylogeny reconstruction based on genomes of the 21L/BA.2 Omicron variant**
345 **obtained in the present study**

346 Figure 2a incorporated genome sequences of 21K/BA.1, 21L/BA.2 and BA.3 Omicron
347 variants. Figure 2b is a zoom of the 21L/BA.2 Omicron cluster of Figure 2a.

348 Phylogenetic tree was built using the nextstrain/ncov tool (<https://github.com/nextstrain/ncov>)
349 then visualized with Auspice (<https://docs.nextstrain.org/projects/auspice/en/stable/>). X-axis
350 shows time. The 21L/BA.2 Omicron genomes the closest genetically to those obtained in our
351 institute were selected using the Usher tool (<https://genome.ucsc.edu/cgi-bin/hgPhyloPlace>)
352 and the GISAID BLAST tool (<https://www.epicov.org/epi3/>) and they were incorporated in
353 the phylogenetic analysis in addition to all 21L/BA.2 Omicron variant genomes from France
354 available in GISAID as of 02/02/2022. Sequences obtained in our laboratory (IHU

355 Méditerranée Infection, Marseille, France) are indicated by a dark blue arrow and their
356 GISAID identifier is indicated. Countries are indicated when they are not France.

357

358 **Figure 3. Number of genomes of the SARS-COV-2 21L/BA.2 Omicron variant available**
359 **in GISAID and chronology of collections of respiratory samples from where they were**
360 **obtained**

361 a: Number of genomes of the SARS-COV-2 21L/BA.2 Omicron variant available in the
362 GISAID sequence database (<https://www.gisaid.org/>)¹⁵ as of 02/02/2022.

363 b: Chronology of SARS-CoV-2 diagnoses with the 21L/BA.2 Omicron variant for genomes
364 deposited in the GISAID sequence database and obtained worldwide.

365 c: Chronology of SARS-CoV-2 diagnoses with the 21L/BA.2 Omicron variant for genomes
366 deposited in the GISAID sequence database and obtained in France or in our university
367 hospital institute.

368 The number of genomes was analyzed until 02/02/2022. Total number of genomes analyzed
369 was 36,428. A total of 1,093 genomes were excluded as the date of sample collection was
370 uncomplete (days or months were lacking).

371

372

373

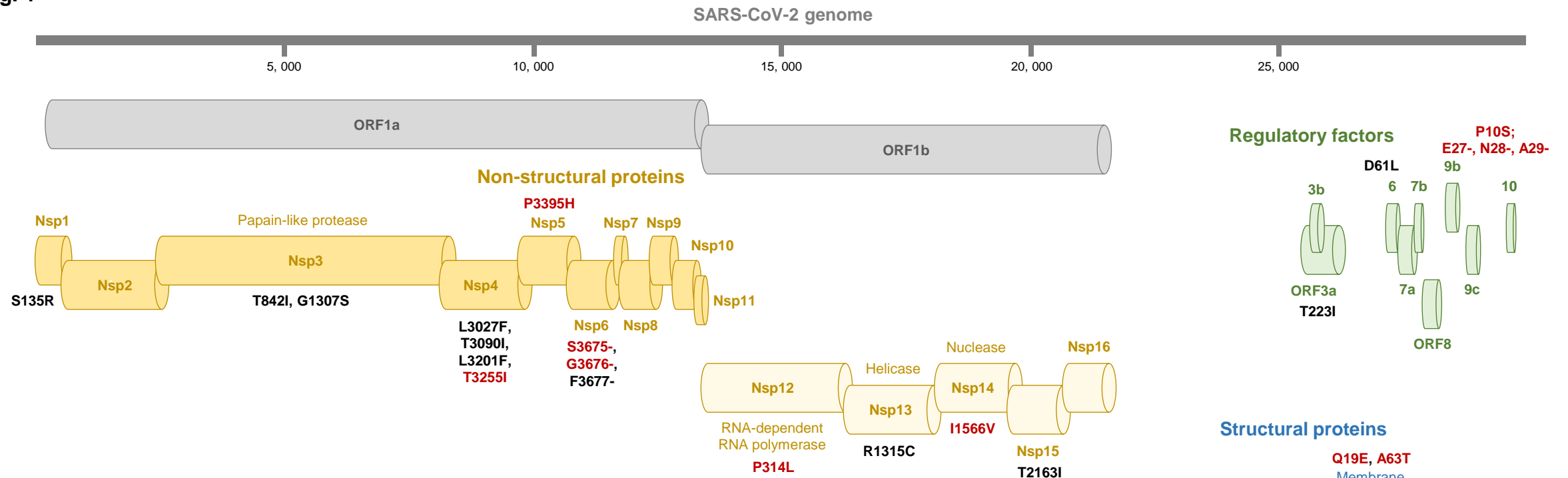
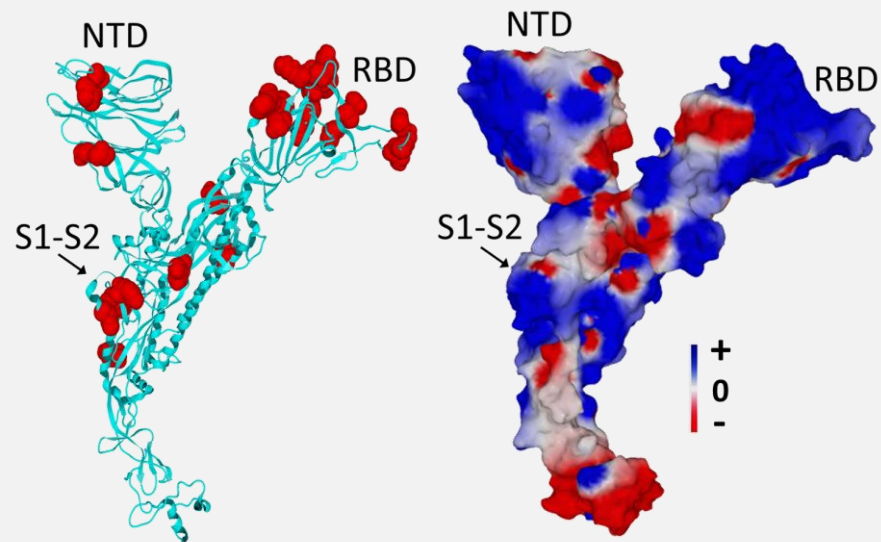
TABLES

374 **Table 1. Main epidemiological and virological features of cases identified with infection with the SARS-CoV-2 21L/BA.2 Omicron**
 375 **variant**

Case no.	Adult age *	Epidemiological data	Clinical data	Diagnostic qPCR Ct	Results of qPCR used to screen for the presence of SARS-CoV-2 spike substitutions	Results of the TaqPath COVID-19 qPCR assay (Targets: ORF1, S, and N genes)	Genome GISAID Id.
1	Yes	Back to travel from Zanzibar (Tanzania)	Three doses of vaccine	21	L452R-Neg.; P681H-Neg.; E484K-Neg.	Pos. for all three genes	EPI_ISL_9161702
2	Yes	Back to travel from Zanzibar (Tanzania)	Three doses of vaccine	12	L452R-Neg.; P681H-Neg.; E484K-Neg.	Pos. for all three genes	EPI_ISL_9184187
3	Yes	No travel abroad	Three doses of vaccine	18	L452R-Neg.; P681H-Neg.; E484K-Neg.	Pos. for all three genes	EPI_ISL_9161106
4	Yes	No data	No data	16	L452R-Neg.; P681H-Neg.; E484K-Neg.	Pos. for all three genes	EPI_ISL_9184306
5	Yes	Dutch nationality	No data	20	L452R-Neg.; P681H: N.t.; E484K: N.t.	Pos. for all three genes	EPI_ISL_8709900
6	Yes	No travel abroad	Not vaccinated	27	L452R-Neg.; P681H-Neg.; E484K-Neg.	Pos. for all three genes	EPI_ISL_9184305
7	Yes	No data	No data	31	L452R-Neg.; P681H-Neg.; E484K-Neg.	Pos. for all three genes	EPI_ISL_9186024
8	Yes	No data	No data	18	L452R-Neg.; P681H: N.t.; E484K-Neg.	Pos. for all three genes	EPI_ISL_9486836
9	Yes	No data	No data	22	L452R-Neg.; K417N-Pos.; E484K: N.t.	Pos. for all three genes	EPI_ISL_9479322
10	Yes	No data	No data	23	L452R-Neg.; K417N-Pos.; E484K: N.t.	Pos. for all three genes	EPI_ISL_9479323
11	Yes	UK nationality	No data	32	L452R-Neg.; K417N-Pos.; E484K: N.t.	N gene-pos.; ORF1 and S genes-neg.	EPI_ISL_9517119
12	Yes	No data	No data	30	L452R-Neg.; K417N-Pos.; E484K: N.t.	Pos. for all three genes	EPI_ISL_9468068
13	Yes	No data	No data	16	L452R-Neg.; K417N-Pos.; E484K: N.t.	Pos. for all three genes	EPI_ISL_9479324

376 Ct, cycle threshold value; E, glutamic acid; H, histidine; Id., identifier; K, lysin; L, leucin; P, proline; R, arginine; N.d., no data; S, spike; N, nucleocapsid; N.d., no
 377 data; Neg., negative; ORF, open reading frame; Pos., positive. Some samples not tested for variant-specific qPCR assays were tested directly by next-generation
 378 sequencing. * Age greater than 18 years
 379

380

Fig. 1**a.****b.**

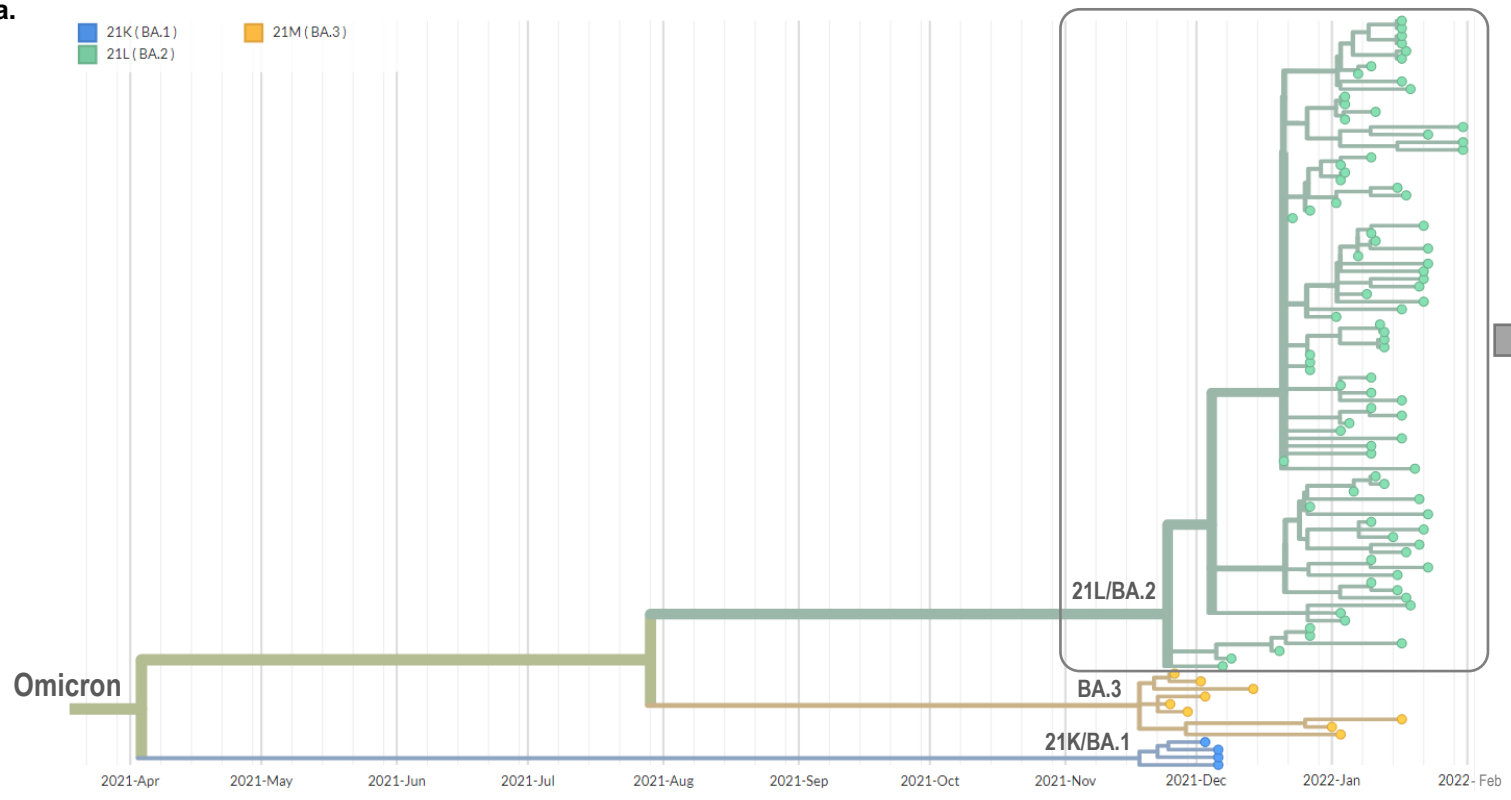
Spike

S

T19I, A27S, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K; L24-, P25-, P26-

Fig. 2

a.



b.

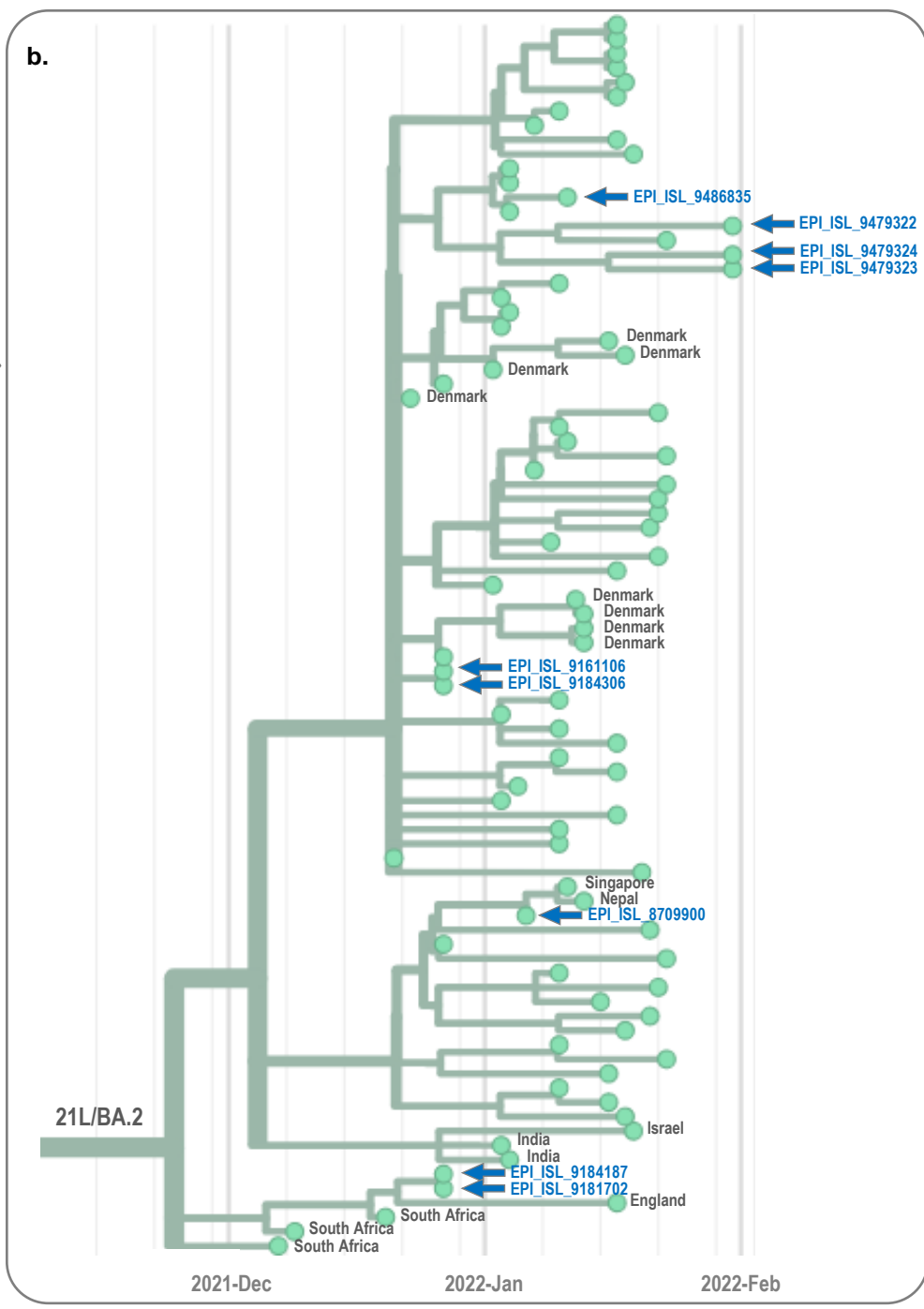
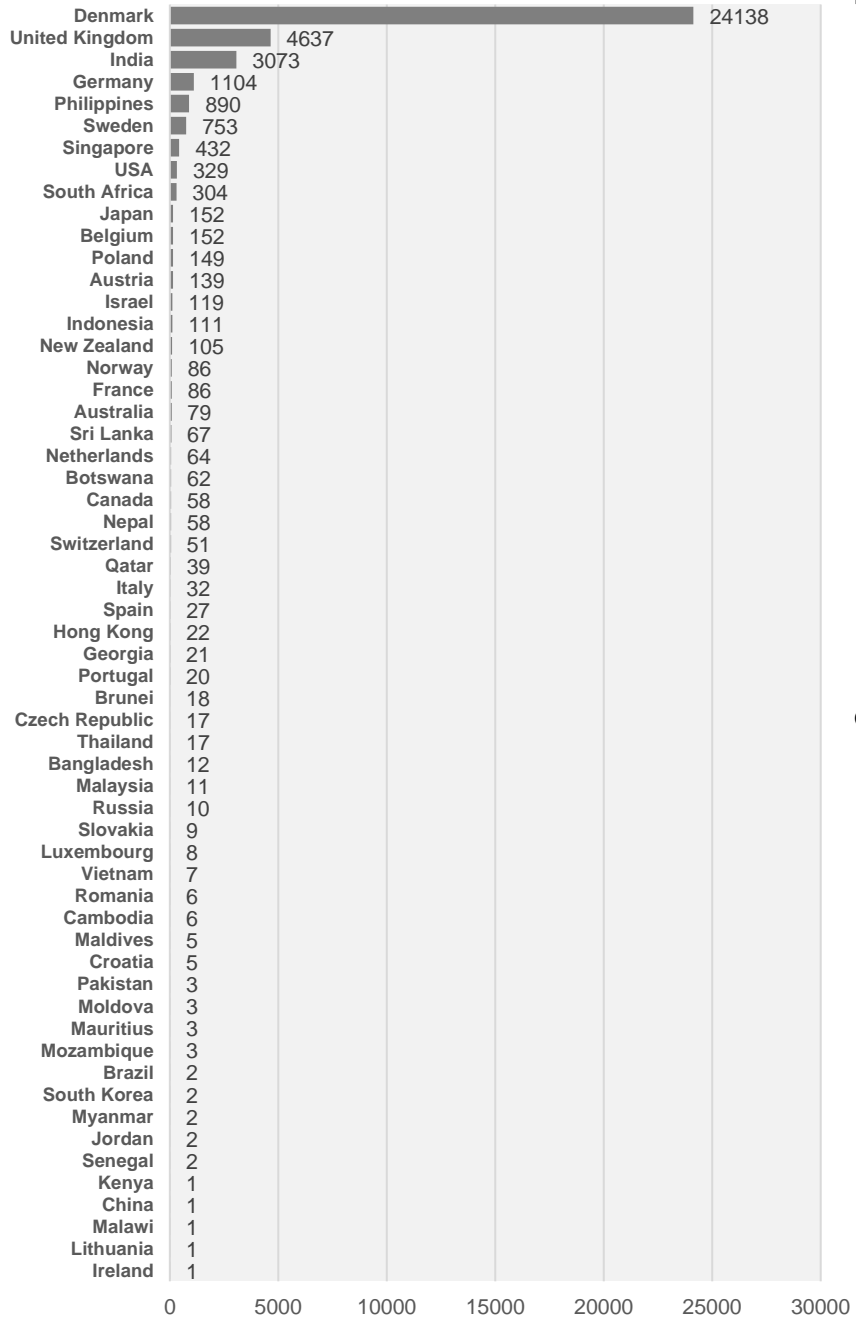
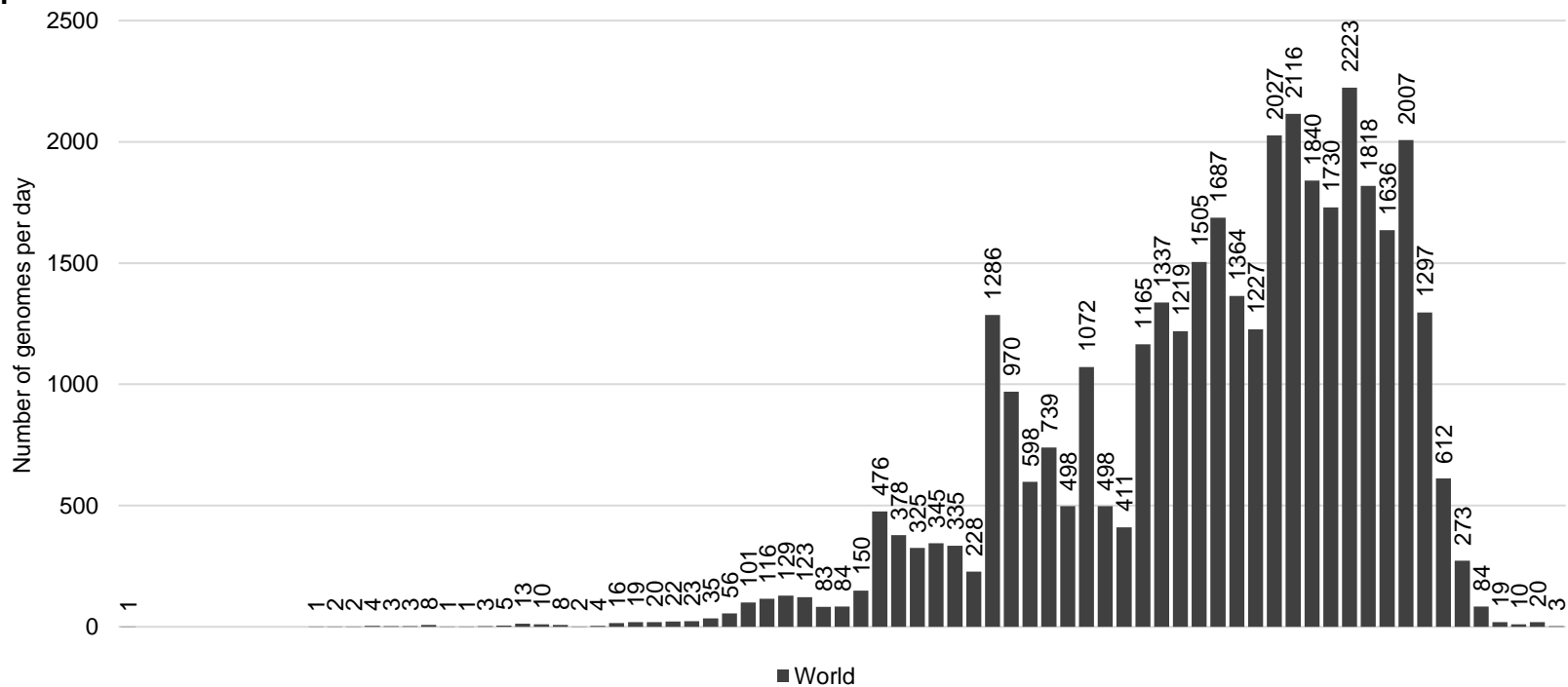


Fig. 3

a.



b.



c.

