

TITLE PAGE

Full-length title:

Dramatic increase in the SARS-CoV-2 mutation rate and low mortality rate during the second epidemic in summer in Marseille

Short title: Increase of SARS-CoV-2 mutation rate

Author list: Philippe COLSON^{1,2} [‡], Anthony LEVASSEUR^{1,2} [‡], Jeremy DELERCE^{1,2}, Hervé CHAUDET^{1,3}, Vincent BOSSI¹, Mariem BEN KHEDHER¹, Pierre-Edouard FOURNIER^{1,2}, Jean-Christophe LAGIER^{1,2}, Didier RAOULT^{1,2} *

Affiliations: ¹ IHU Méditerranée Infection; ² Aix-Marseille Univ., Microbes Evolution Phylogeny and Infections (MEPHI); ³ French Armed Forces Center for Epidemiology and Public Health (CESPA), Vecteurs - Infections Tropicales et Méditerranéennes (VITROME), Marseille, France.

[‡] These authors contributed equally to this article

*** Address for correspondence:** Didier Raoult, IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: didier.raoult@gmail.com

Key words: SARS-CoV-2; Covid-19; genome; variant; heterogeneity

Word count: summary, 195; text, 1,449

Figures: 4; **Table:** 1

ABSTRACT

We investigated the evolution of 691 full-length genome sequences obtained from patients diagnosed with SARS-CoV-2 between February and August 2020. We show that the sequences of the past epidemic (February-May) majoritarily disappeared and those of the current epidemic (June-August) belong to new genotypes exhibiting a dramatically higher mutation rate.

TEXT

The SARS CoV-2 outbreak that started in Wuhan, China, in December 2019, has rapidly spread around the world (1) (<https://coronavirus.jhu.edu/map.html>). In Marseille, southeastern France, the first case was diagnosed on 02/27/2020 (2). At IHU Méditerranée Infection, we have been monitoring the daily number of SARS CoV-2 PCR tests and positive cases since then (<https://www.mediterranee-infection.com/covid-19/>). We implemented a systematic testing strategy and carried out (as on 09/03/2020) 212,194 PCRs for 147,813 individuals. Of these, 10,192 were positive (**Figure 1**). This led us to perform culture inoculation for >5,000 samples (3) and next-generation sequencing (NGS) (4) for 1,380 samples, particularly from patients with a low cycle threshold (Ct) value (<20) of the PCR test.

The study

From 29 January 2020, we diagnosed SARS-CoV-2 infections by PCR by testing nasopharyngeal swab fluids sent to the clinical microbiology laboratory of IHU Méditerranée Infection, as previously described (5). Monitoring of the epidemic in Marseille shows that around 20 May, SARS-CoV-2 cases showed an almost total disappearance, and a re-increase was observed during July to reach in early September between 100 and 150 diagnoses per day. The number of patients tested and positive cases were 86,358 and 6,858, respectively, between February and May, and 60,768 and 3,337, respectively, between June and September. These data do not appear to be biased because the proportion of positive tests was in the same order of magnitude in February-April (7.9%) and July-August (5.5%), and we continuously tested voluntary peoples regardless of whether they were symptomatic or asymptomatic since the first case in Marseille (**Figure 1**).

We performed whole genome sequencing for PCR positivity with a cycle threshold

(Ct) value < 30. SARS-CoV-2 genomes were obtained from nasopharyngeal swab fluid by next-generation sequencing using Illumina technology with the Illumina Nextera XT Paired end strategy on a MiSeq instrument (Illumina Inc., San Diego, CA, USA), as previously described (2) (**Appendix**). Genome consensus sequences were generated with the CLC Genomics workbench v.7 by mapping on the SARS-CoV-2 genome GenBank Accession no. MN908947 (Wuhan-Hu-1 isolate) with the following thresholds: 0.8 for coverage and 0.9 for similarity.

Sequences from 691 complete genomes (**Appendix: supplementary File S1**) that were obtained were analyzed using the Nextstrain web-tool (<https://clades.nextstrain.org/>) (6). They were compared to sequences available in the GISAID database (<https://www.gisaid.org/>). Phylogenetic trees were reconstructed by using the nextstrain/ncov tool (<https://github.com/nextstrain/ncov>) and visualized with iTOL (<https://itol.embl.de/>) (**Appendix**). Analysis of full-length SARS-CoV-2 genomes showed that a particular genotype that we named “Marseille 1a” appeared in July, which has not been described in the literature. It combines eight new mutations (**Figure 2**) whose association is unknown and arises as a long branch on the phylogenetic tree built with all genomes obtained in our laboratory. Based on this finding, we accelerated SARS-CoV-2 genome sequencing performed from PCR-positive samples collected in July and August because we believed that a new genotype had set in, and we designed a qPCR specific for the Marseille 1 genotype. However, after analyzing 691 full-length genomes, we were able to demonstrate that there were in fact seven new clades (named “Marseille 1” to “Marseille 7”) (**Figures 3A-B**) that emerged since June, only viruses of 4 genotypes observed between February and May being found since July (**Table 1**). Sequences from the same genotype than three of these clades (2, 5 and 6) were found in the GISAID database (<https://www.gisaid.org/>). The first observed clade (Marseille 1a) seems to have almost disappeared after July. The index case was imported from North

Africa, and the first cases thereafter diagnosed were linked to ships connecting North Africa to Marseille; other cases did not appear to have a specific origin.

Mutations observed in these seven different viral genotypes are located in most SARS-CoV-2 genes including structural and non-structural genes among which *nsp2*, *nsp3* (predicted phosphoesterase), *nsp5* (membrane glycoprotein), *nsp12* (RNA-dependent RNA polymerase), *S* (Spike glycoprotein), *ORF3a*, *E* (membrane glycoprotein), *M* (membrane glycoprotein), *ORF8* and *N* (Nucleocapsid phosphoprotein) (**Figure 2**). Analysis of the heterogeneity of the 204 sequences produced from June to August compared to the 487 sequences produced from February to May shows a large difference (mean $\pm$ standard deviation for genetic distance: $7.6 \times 10^{-4} \pm 3.8 \times 10^{-4}$ vs. $2.3 \times 10^{-4} \pm 1.1 \times 10^{-4}$, respectively; $p < 2.2 \times 10^{-16}$; **Appendix: supplementary Figure S1**), which indicates that the virus mutation rate is accelerating dramatically. The more distant genome relative to that of the Wuhan-1 isolate exhibited 29 mutations and was of clade Marseille 5. Interestingly, but preliminary, the mortality of SARS-CoV-2-positive patients hospitalized since mid-June is lower than that of those hospitalized between February and May [9 deaths/1,958 cases (0.5%) vs. 162 deaths/5,929 cases (2.7%), respectively (<https://www.mediterranee-infection.com/covid-19/>; (5)); $p = 6.6 \times 10^{-11}$], whereas the proportion of positive and ambulatory patients was similar for these two time periods. There is no bias regarding the comparison of the mortality rate between these two periods because our testing strategy and clinical and therapeutic management have remained the same (5). This reduced mortality rate appears to be a general phenomenon in France and Europe with a low fatality rate of this summer outbreak compared to the one that occurred from February to May (<https://covid19-country-overviews.ecdc.europa.eu/>; <https://github.com/CSSEGISandData/COVID-19>). It seems that the current genotypes are at the origin of epidemic bursts, which we do not know if they will relapse and continue, but the first detected became minority in August when we could

distinguish some others (**Figure 3B; Table 1**). In addition to the complete genomes described in this work, the identification of Marseille 1 variant based on specific real-time qPCR (**Appendix**) allowed us to estimate that they were involved in 65 (23%) of 259 infections during the period July–mid August.

Conclusions

Overall, as recently pointed out by Tomaszewski et al. (7) who described for viral genomes available until May 2020 a mutational shift from the spike and replication complex to genes encoding other non-structural proteins that interact with host defense pathways, it appears that the mutation rate of SARS-CoV-2 is accelerating since May, mostly involving C-to-U mutations. The SARS-CoV-2 mutation rate increase generates viral genotypes more distant from the initial Wuhan strain than observed from March to April. This seems to result in epidemics of limited duration, at least for the first new genotype that we identified, and is associated with less severity overall at this stage of the development of this new epidemic.

Acknowledgments

We are grateful to Ludivine Brechard, Olivia Ardizzoni, Madeleine Carrera, Vera Esteves-Vieira, Laurence Thomas, Priscilla Jardot, Raphael Tola, and Audrey Giraud-Gatineau for their technical help.

This manuscript has been edited by Wiley Editing Services (<https://wileyeditingservices.com/en/>): order No DIOUL_2.

Author contributions

Conceived and designed the experiments: DR. Contributed materials/analysis tools: PC, JD,

AL, HC, MB, JCL, and PEF. Analyzed the data: PC, JD, AL, HC, MB, PEF, and DR. Wrote the manuscript: DR and PC. All authors approved the final manuscript version.

Funding

This work was supported by the French Government under the “Investments for the Future” program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-IAHU-03 and was also supported by Région Provence Alpes Côte d’Azur and European funding FEDER PRIMMI (Fonds Européen de Développement Régional-Plateformes de Recherche et d’Innovation Mutualisées Méditerranée Infection), FEDER PA 0000320 PRIMMI.

Conflicts of interest

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

Ethics

The study was approved by the ethical committee of the University Hospital Institute Méditerranée Infection (N°: 2020-016-2). Access to the patients’ biological and registry data issued from the hospital information system was approved by the data protection committee of Assistance Publique-Hôpitaux de Marseille (APHM) and was recorded in the European General Data Protection Regulation registry under number RGPD/APHM 2019-73.

Biographical sketch

Philippe Colson is Professor of Virology at IHU Méditerranée Infection in Marseille, France.

157 His field of interest is the diagnosis and molecular epidemiology of human viral infections
158 including human immunodeficiency virus and hepatitis viruses, and the genomics of giant
159 viruses.

160

161

REFERENCES

1. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;10-2012.
2. Colson P, Lagier JC, Baudoin JP, Bou KJ, La Scola B, Raoult D. Ultrarapid diagnosis, microscope imaging, genome sequencing, and culture isolation of SARS-CoV-2. *Eur J Clin Microbiol Infect Dis*. 2020;39:1601-3.
3. La Scola B, Le BM, Andreani J, Hoang VT, Grimaldier C, Colson P, et al. Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. *Eur J Clin Microbiol Infect Dis*. 2020; 39:1059-1061.
4. Levasseur A, Delerce J, Caputo A, Brechard L, Colson P, Lagier JC, et al. Genomic diversity and evolution of coronavirus (SARS-CoV-2) in France from 309 COVID-19-infected patients. *bioRxiv* 2020; doi: <https://doi.org/10.1101/2020.09.04.282616>.
5. Lagier JC, Million M, Gautret P, Colson P, Cortaredona S, Giraud-Gatineau A, et al. Outcomes of 3,737 COVID-19 patients treated with hydroxychloroquine/azithromycin and other regimens in Marseille, France: A retrospective analysis. *Travel Med Infect Dis*. 2020;36:101791.
6. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34:4121-3.
7. Tomaszewski T, DeVries RS, Dong M, Bhatia G, Norsworthy MD, Zheng X, et al. New Pathways of Mutational Change in SARS-CoV-2 Proteomes Involve Regions of Intrinsic Disorder Important for Virus Replication and Release. *bioRxiv* 2020;doi: <https://doi.org/10.1101/2020.07.31.231472>.

FIGURES

Figure 1. Number of PCR tests, positive diagnoses, and deaths from February to September

A. Number of PCR tests performed at IHU Méditerranée Infection; B: Number of PCR-positive patients performed at IHU Méditerranée Infection; C: Number of deaths among SARS-CoV-2-positive patients in Marseille public hospitals (Assistance Publique-Hôpitaux de Marseille).

Figure 2. Microarray showing the distribution along the viral genome and in viral genes of mutations observed for the various viral genotypes

Sequences from complete genomes that were obtained were analyzed using the Nextstrain web-tool (<https://clades.nextstrain.org/>) (6). They were compared to sequences available in the GISAID database (<https://www.gisaid.org/>).

Representation is adapted from Nextclade sequence analysis web application output (<https://clades.nextstrain.org/results>).

^a In reference to genome GenBank Accession no. NC_045512.2 (Wuhan-Hu-1 isolate);^b green: U; yellow: G; blue: C; red: A.

3CL: 3C-like proteinase; E: Envelope protein; H: NTPase/helicase domain; M: Membrane glycoprotein; N: nucleocapsid phosphoprotein; nsp9: ssRNA-binding protein; nsp14: 3'-to-5' exonuclease; nsp15: EndoRNase; PLpro: Predicted phosphoesterase; RdRp: RNA-dependent RNA polymerase; S: Spike glycoprotein; Syn.: synonymous.

Figure 3. Phylogenetic trees based on SARS-CoV-2 full-length genomes obtained during the two epidemics from February to May and from June to August and monthly

proportions of the newly-identified viral genotypes

A: Phylogenetic trees based on SARS-CoV-2 full-length genomes obtained during the two epidemics from February to May and from June to August. Phylogenetic trees were reconstructed by using the GISAID TreeTool in v2.0 (<https://www.gisaid.org/epiflu-applications/upcoming-features-in-v20/treetool-app/>) that performs an initial approximate maximum likelihood phylogeny reconstruction using FastTree then a refinement by RaXML.

A1: Phylogenetic tree was reconstructed from 691 full-length viral genomes obtained from clinical samples collected from February to August 2020; this tree can be visualized at a greater size in a separate file available in the **Appendix (supplementary Figure S2)**. A2, A3: Phylogenetic trees were reconstructed from 487 and 304 full-length viral genomes obtained from samples collected from February to May 2020 (A2) and from June to August 2020 (A3), respectively. Genomes from February, March, April, May, June, July, and August are labeled with green, orange, light blue, pink, yellow, red, and blue backgrounds, respectively.

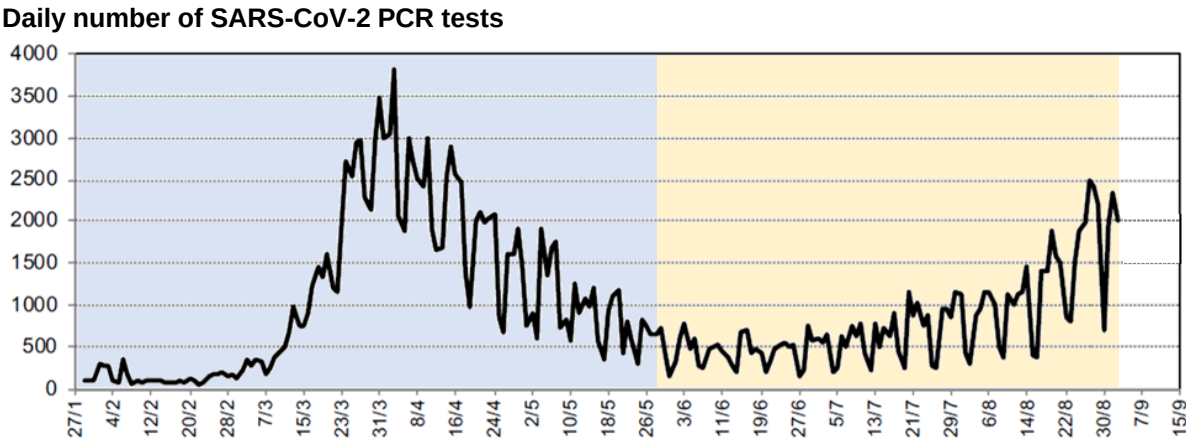
B. Monthly proportions of the newly-identified viral genotypes. The monthly proportions of viral genotypes correspond to that of full-length genomes obtained from samples collected during a given month from June to August 2020 and classified in genotypes 1 to 7. B1: Distribution of the proportions of genomes classified in Marseille clades 1-7 according to the month for months from February to August. B2: A: Distribution of the proportions of genomes classified in the various Marseille clades 1-7 and their subclades according to the month for months from June to August

234 Table 1. Number of genomes per genotype and month from February to May and from June to August

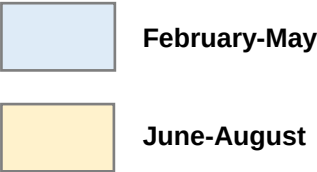
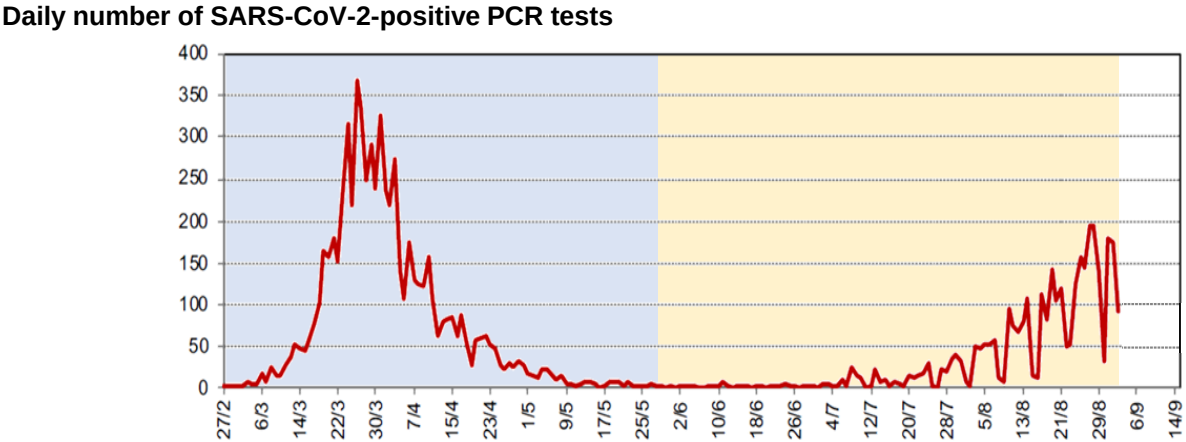
Clade nomenclature	Marseille clade	Number of mutations relative to the clade original	Number of mutations relative to the Wuhan-Hu-1 genome	Months								All months	Hallmark mutations
				February	March	April	May	June	July	August			
Total number of genomes:				1	259	204	23	5	40	159	691		
Genotypes that were present both between February-May and June-August													
20A/15324U		1	5	0	25	17	0	0	0	5	47	--C15324U	
20A/15324U-3		1	6	0	0	1	8	1	0	0	10	---G9479U	
20A/20268G		1	5	1	0	1	0	0	1	0	3	---A20268G	
20B		3	7	0	23	13	0	0	8	10	54	---G28881A,G28882A,G28883C	
20C-2		1	7	0	2	2	0	1	0	2	7	----C10582U	
Genotypes that disappeared between February-May and June-August													
19A / 19B	/	/	/	0	7	0	0	0	0	0	7		
20A		4	4	0	3	4	0	0	0	0	7	-C241U,C3037U,C14408U,A23403G	
20A/15324U-1		1	6	0	3	2	2	0	0	0	7	---C2189U	
20A/15324U-1a		1	7	0	0	15	0	0	0	0	15	----G4105U	
20A/15324U-4		1	6	0	8	0	0	0	0	0	8	---C19602U	
20A/18877U		1	6	0	1	1	0	0	0	0	2	---C18877U	
20A/20268G-2		1	6	0	4	6	0	0	0	0	10	---C18175U	
20A/25563U		1	5	0	10	6	1	0	0	0	17	--G25563U	
20A/25563U-1		1	6	0	8	2	0	0	0	0	10	---C2416U	
20A/25563U-1a		2	8	0	0	1	0	0	0	0	1	----C13458U,C20946U	
20A/25563U-1a1		1	9	0	1	5	0	0	0	0	6	-----C6401U	
20A/25563U-1a1a		1	10	0	4	5	0	0	0	0	9	-----A1198G	
20A/25563U-1b		1	7	0	54	65	2	0	0	0	121	----G8371U	
20A/25563U-1b1		1	8	0	7	1	0	0	0	0	8	-----C26907A	
20A/25563U-1b2		2	9	0	1	5	0	0	0	0	6	-----C9996U,G29747U	
20A/25563U-1b3		1	8	0	1	4	0	0	0	0	5	-----G26718U	
20A/25563U-1b4		1	8	0	2	3	1	0	0	0	6	-----G25909U	
20A/25563U-1b5		1	8	0	0	0	5	0	0	0	5	-----G25767U	
20A/25563U-1b6		2	9	0	3	3	0	0	0	0	6	-----A20294G,G29779C	
20A/25563U-1b7		1	8	0	0	1	0	0	0	0	1	-----A12334U	
20A/25563U-1b7a		1	9	0	1	4	0	0	0	0	5	-----C15810U	
20A/25563U-1b8		1	8	0	0	3	2	0	0	0	5	-----C2910U	
20A/25563U-1b9		3	10	0	3	2	0	0	0	0	5	-----G11083U,C19160U,G28027U	
20A/25563U-1c		4	10	0	5	1	0	0	0	0	6	-----G3338U,G15438U,C21597U,C25731U	
20B-1		1	8	0	11	1	0	0	0	0	12	---C313U	
20B-1a		2	10	0	3	2	0	0	0	0	5	-----G19518U,G28845U	
20B-1b		1	9	0	3	2	0	0	0	0	5	-----G24998U	
20C		1	6	0	37	13	1	0	0	0	51	---C1059U	
20C-1		1	7	0	4	1	0	0	0	0	5	-----C9286U	
20C-3		2	8	0	5	0	0	0	0	0	5	-----U13006C,C25688U	
20C-3a		1	9	0	1	5	0	0	0	0	6	-----C14391U	
20C-4		1	7	0	5	1	0	0	0	0	6	-----G27996U	
20C-5		1	7	0	14	6	1	0	0	0	21	-----G11083U	
Genotypes that appeared between February-May and June-August													
20A/15324U-2	Marseille 3	7	12	0	0	0	0	0	1	17	18	---C1912U,G5210A,C17470U,C21191U,A23148G,U27125C,C28854U	
20A/18877U-1		1	7	0	0	0	0	0	1	0	1	-----C26735U	
20A/18877U-1a	Marseille 4	13	20	0	0	0	0	0	1	55	56	-----C4543U,G5629U,G9526U,C11497U,G13993U,G15766U,A16889G,G17019U,G22992A,C25710U,U26876C,G28975C,G29399A	
20A/18877U-1a1	"	1	21	0	0	0	0	0	0	5	5	-----G28086U	
20A/20268G-1a	Marseille 1	6	12	0	0	0	0	1	19	1	21	-----G1181U,C1625U,G22894A,C25886U,G28198U,G28851U	
20A/20268G-1a1	"	1	13	0	0	0	0	0	5	5	10	-----C22088U	
20A/20268G-1a1a	"	1	14	0	0	0	0	0	0	5	5	-----G5378A	
20A/20268G-1b	Marseille 7	3	9	0	0	0	0	0	0	5	5	-----C2706U,C25731U,G27463C	
20A/25563U-1b10	Marseille 6	10	17	0	0	0	0	0	1	5	6	-----C9430U,A15477U,C18395U,A20622U,G20623U,A20624U,C23730U,A26319G,C28854U,G29044A	
20C-2a		2	9	0	0	0	0	2	2	0	4	-----C27804U,C28830A	
20C-2a1	Marseille 5	17	26	0	0	0	0	0	1	20	21	-----C3099U,G4960U,C4965U,C6070U,C7303U,C7564U,C9246U,C10279U,C10301A,C10525U,C10688U,G11851U,C14230A,G21800U,G27632U,G29402U,G29779U	
20C-2a1a	"	3	29	0	0	0	0	0	0	6	6	-----U8743C,G16377U,G16647A	
20C-2a1b	"	1	27	0	0	0	0	0	0	5	5	-----C9142U	
20D	Marseille 2	7	11	0	0	0	0	0	0	6	6	--U445C,C6286U,G21255C,C22227U,C26801G,C28932U,G29645U	
20D1	"	5	16	0	0	0	0	0	0	7	7	---G4006U,C5170U,G5504A,G11132U,U26609C	

Fig. 1

A.



B.



C.

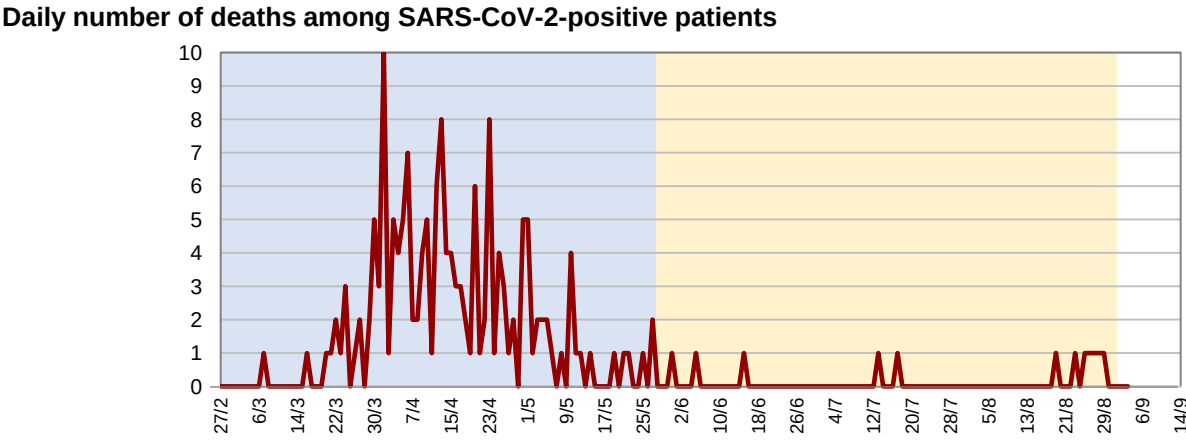


Fig. 2

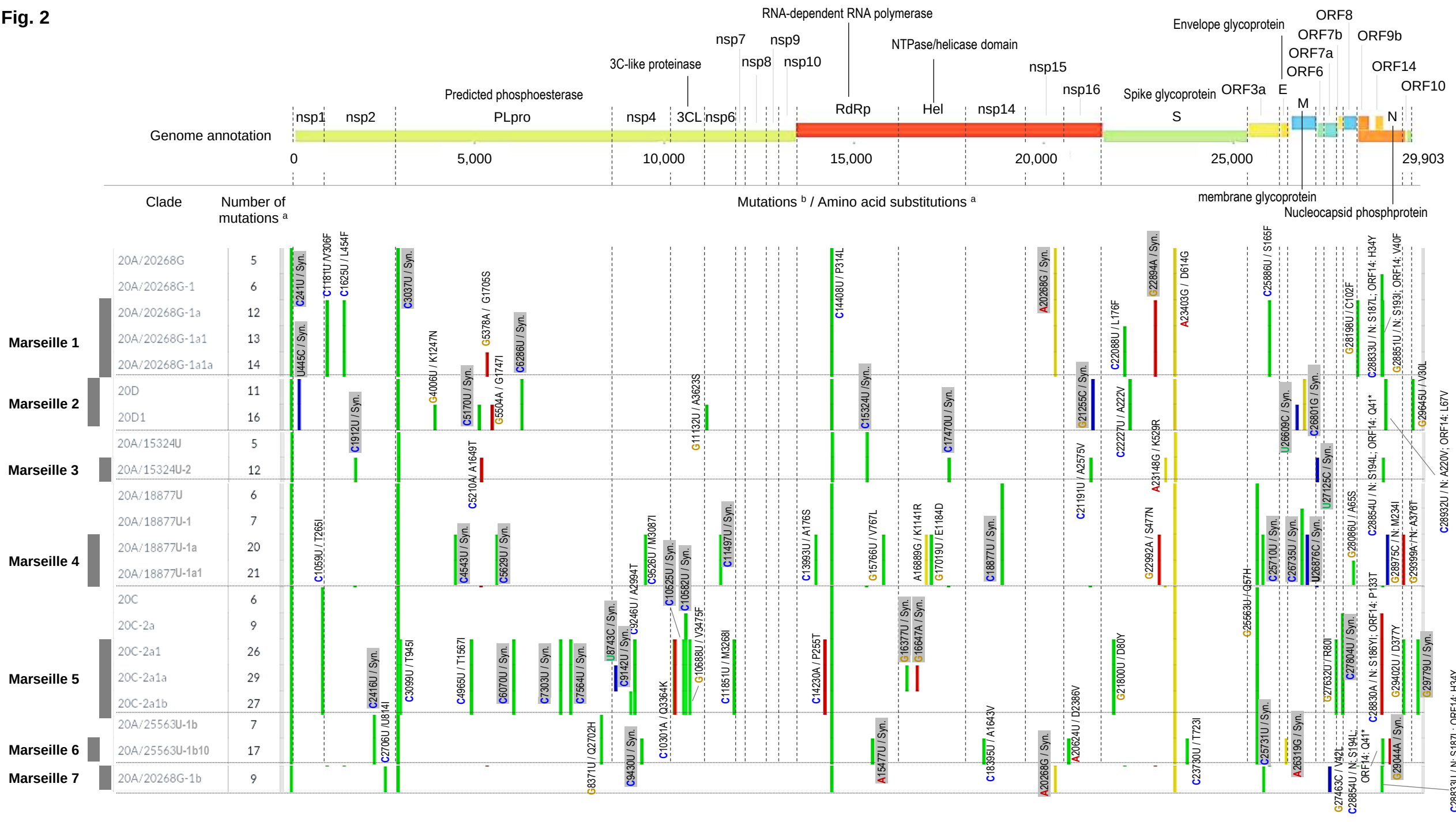
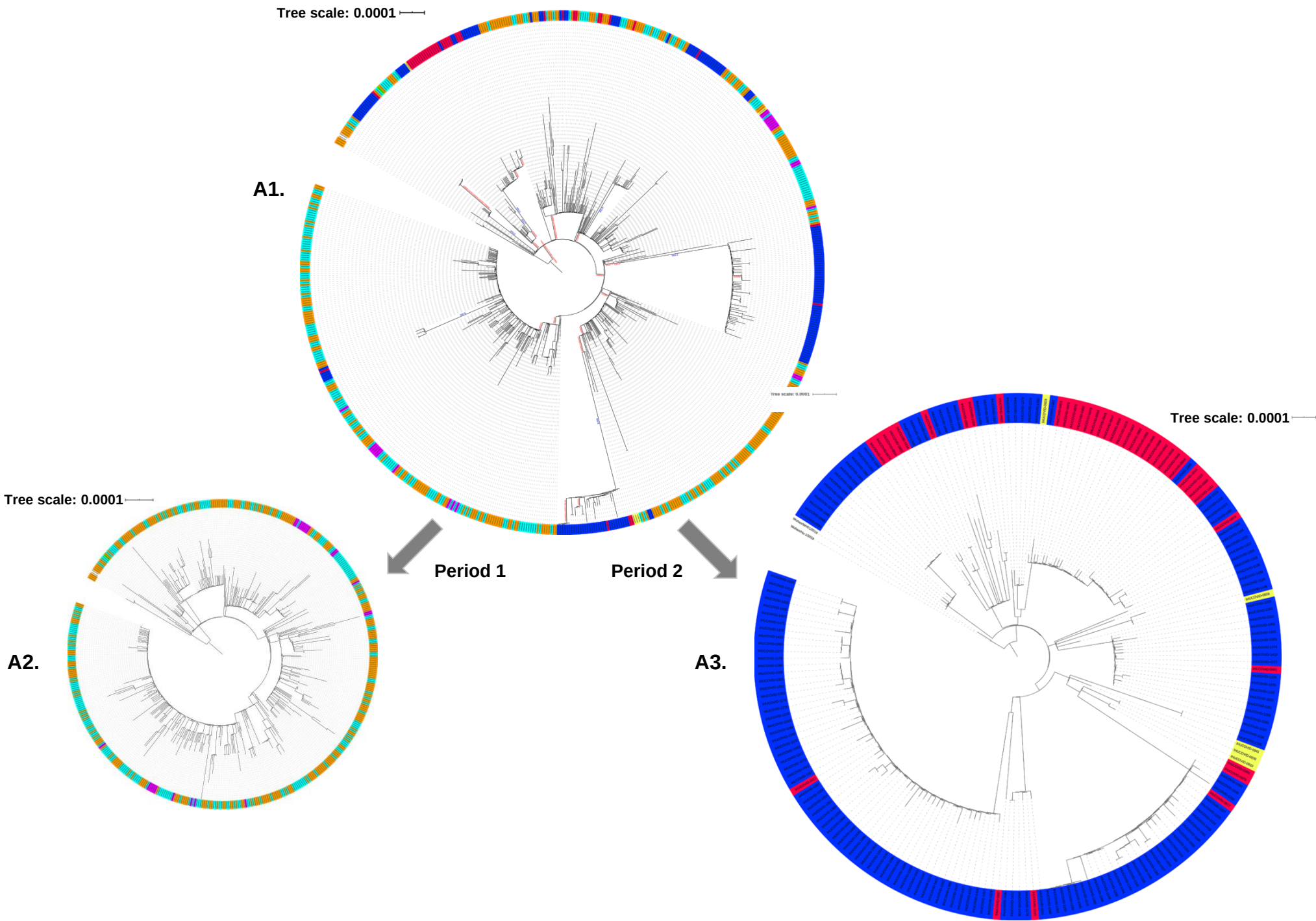
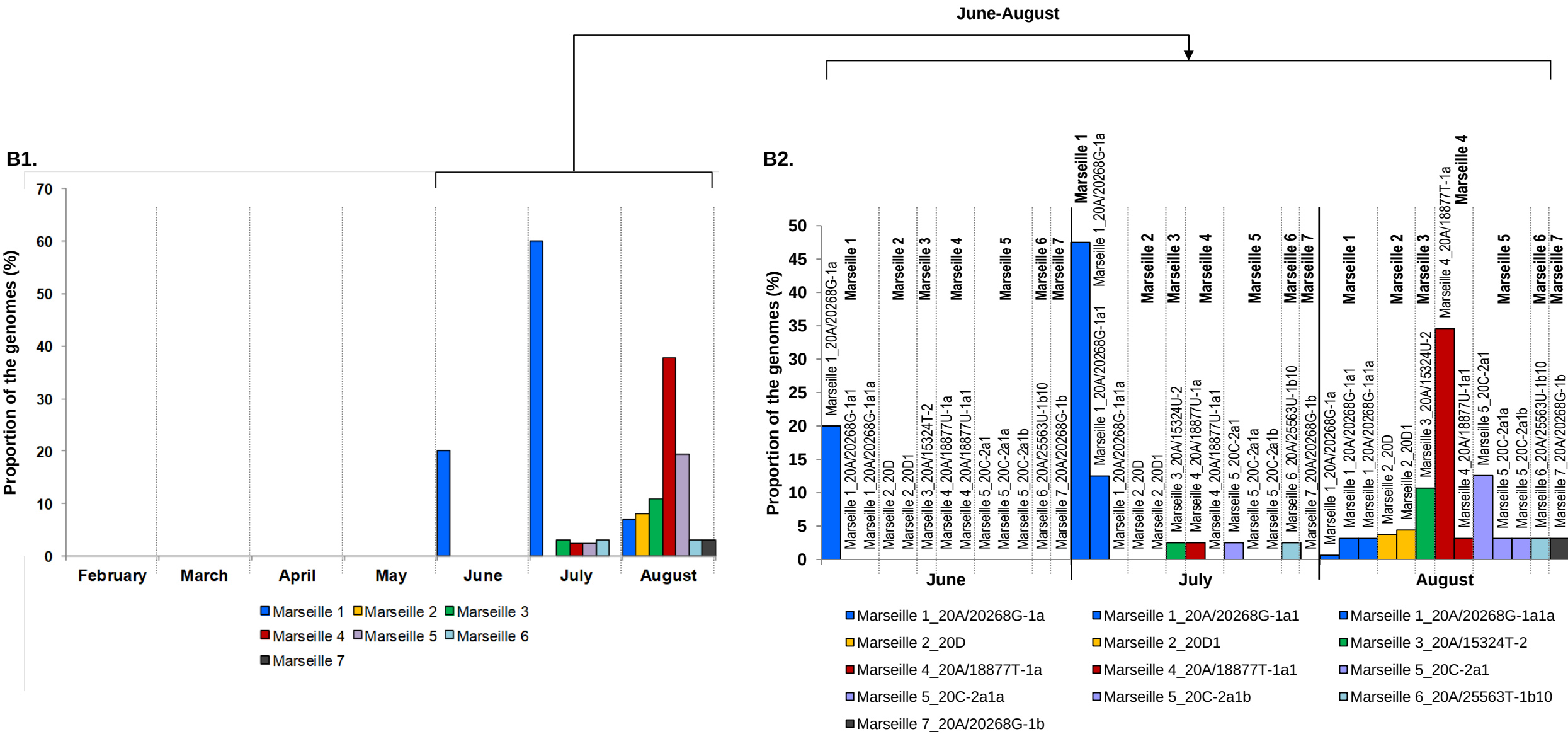


Fig. 3

A.



B.



Dramatic increase in the SARS COV-2 mutation rate and low mortality rate during the second epidemic outbreak in summer in Marseille

APPENDIX

METHODS

Genome sequencing

SARS-CoV-2 genomes were obtained from nasopharyngeal swab fluid by next-generation sequencing as previously described (1). Briefly, viral RNA was extracted from 200 µL of nasopharyngeal swab fluid using the EZ1 Virus Mini Kit v2.0, and it was reverse transcribed using SuperScript IV (ThermoFisher Scientific, Waltham, MA, USA) prior to cDNA second strand synthesis with Klenow Fragment DNA polymerase (New England Biolabs, Beverly, MA, USA). The generated DNA was purified using Agencourt AMPure XP beads (Beckman Coulter, Villepinte, France) and sequenced using Illumina technology with the Illumina Nextera XT Paired end strategy on a MiSeq instrument (Illumina Inc., San Diego, CA, USA). Genome consensus sequences were generated with the CLC Genomics workbench v.7 by mapping on the SARS-CoV-2 genome GenBank Accession no. MN908947 (Wuhan-Hu-1 isolate) with the following thresholds: 0.8 for coverage and 0.9 for similarity.

Marseille 1 variant specific qPCR

The fragment of the gene encoding the nucleocapside and harboring two mutations separated by 17 nucleotides concurrently present in the Marseille 1 genotype was used as a target for the design of a real-time qPCR assay with a hydrolysis probe specific to the Marseille 1 genotype.

The sequences of primers and probes are as follows: Pri_28833-51_F5 (forward):

CAAGCCTCTTCTCGTTCCTT; Pri_28833-51_R5 (reverse):

GCCAGCCATTCTAGCAGGA; Probe_28833-51_4: 5'FAM-

ACGTAGTCGCAACATTTCAAGAAA-3'TAMRA.

Genome sequence analyses

Sequences from complete genomes that were obtained were analyzed using the Nextstrain web-tool (<https://clades.nextstrain.org/>) (2). They were compared to sequences available in the GISAID database (<https://www.gisaid.org/>). Phylogenetic trees were reconstructed by using the nextstrain/ncov tool (<https://github.com/nextstrain/ncov>) and visualized with iTOL (<https://itol.embl.de/>). Distance estimation was performed using MEGA 6 (version 6.06) using the maximum composite likelihood method with uniform rates among sites and partial deletion set at 5%. Genome nucleotide heterogeneity was represented in a box-and-whisker plot of the distributions of nucleotide distances that compared genomes collected from February to May and from June to August using Welch Two Sample t-test. Statistical tests were done using R 4.0.2 (3).

SUPPLEMENTARY FILE

Supplementary File S1: contains the 691 complete genomes that were obtained at IHU Méditerranée Infection.

SUPPLEMENTARY FIGURE LEGEND

Supplementary Figure S1. Box-and-whisker plot of the distance distributions of genomes collected from February to May and from June to August

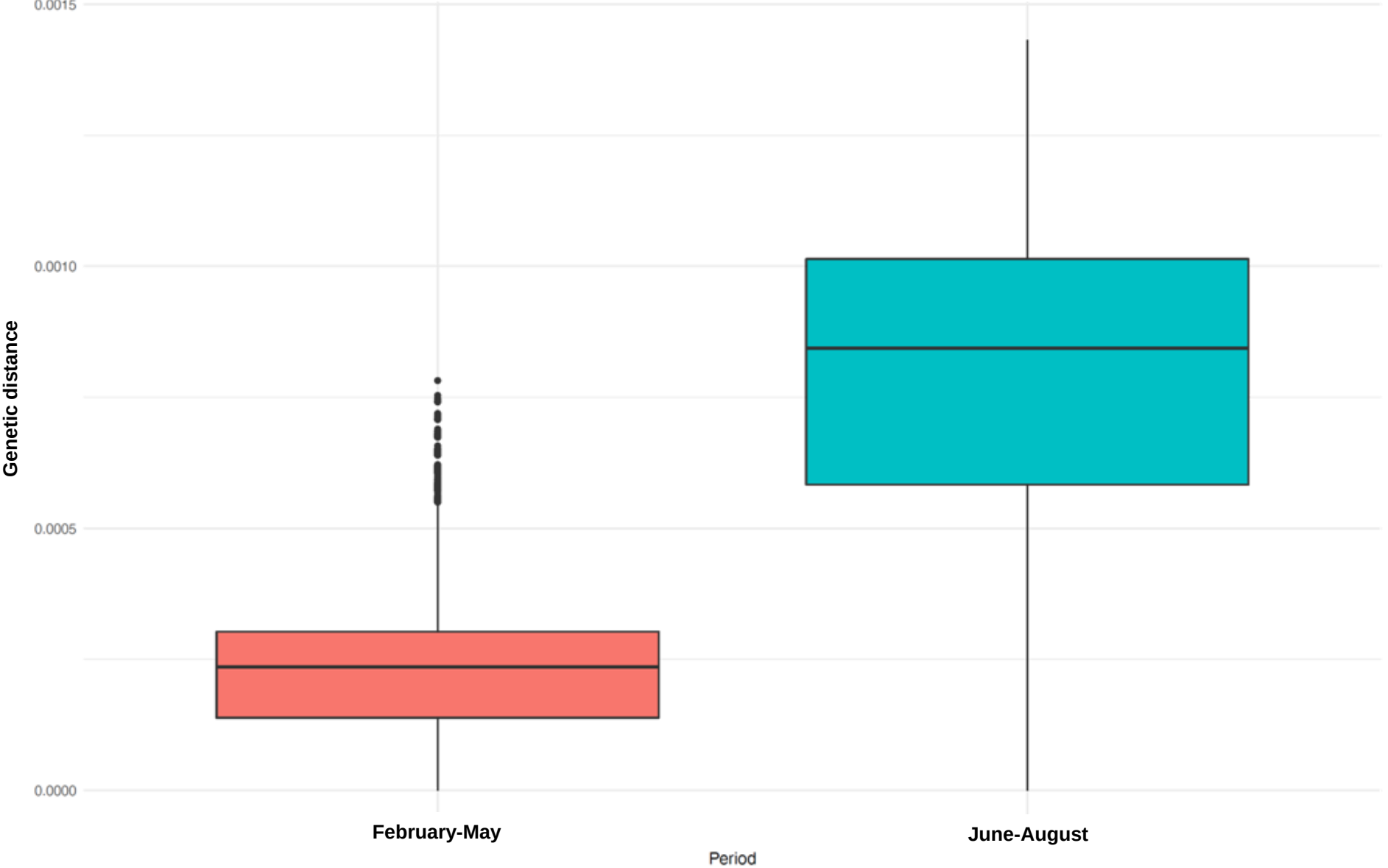
Distance estimation was performed using MEGA 6 (version 6.06) using the maximum composite likelihood method with uniform rates among sites and partial deletion set at 5%. Genome nucleotide heterogeneity was represented in a box-and-whisker plot of the distributions of nucleotide distances that compared genomes collected from February to May and from June to August using Welch Two Sample t-test. Statistical tests were done using R 4.0.2 (3).

Supplementary Figure S2. Phylogenetic tree was reconstructed from 691 full-length viral genomes obtained from clinical samples collected from February to August 2020.

References

1. Colson P, Lagier JC, Baudoin JP, Bou KJ, La Scola B, Raoult D. Ultrarapid diagnosis, microscope imaging, genome sequencing, and culture isolation of SARS-CoV-2. *Eur J Clin Microbiol Infect Dis*. 2020;39:1601-3.
2. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34:4121-3.
3. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria 2020.

Suppl. Fig. S1



Tree scale: 0.0001

