TITLE PAGE

Article type: Research article

Full-length title:

Emergence and outcome of the SARS-CoV-2 "Marseille-4" variant

Short title: Outcome of the Marseille-4 genotype

Author list: Pierre-Edouard FOURNIER[1,2*], Philippe COLSON[1,3], Anthony LEVASSEUR[1,3], Christian DEVAUX[1,3], Philippe GAUTRET[1,2], Marielle BEDOTTO[1,3], Jeremy DELERCE[1,3], Ludivine BRECHARD[1,3], Lucile PINAULT[1,3], Jean-Christophe LAGIER[1,3], Florence FENOLLAR[1,2], Didier RAOULT[1,3]*

Affiliations: [1] IHU Méditerranée Infection; [2] Vecteurs - Infections Tropicales et Méditerranéennes (VITROME), Marseille, France ; [3] Aix-Marseille Univ, Microbes Evolution Phylogeny and Infections (MEPHI), Marseille, France.

*** Contact details for correspondence:**

Pierre-Edouard Fournier, IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: pierre-edouard.fournier@univ-amu.fr

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

Word count: abstract, 303; text, 1,904

Figures: 4; **Table:** 1; **References:** 10

Abstract

Background. In Marseille, France, following a first COVID-19 outbreak between April and May 2020, a second epidemic wave occurred from June, involving ten new successive or concomitant SARS-CoV-2 variants. Of these, the Marseille-4 variant caused an epidemic that started abruptly in August, rapidly replacing other variants, and is still ongoing.

Methods and results. By sequencing the genomes of 1,038 SARS-CoV-2 strains which we deposited in the GISAID database (https://www.gisaid.org/) and compared to the other SARS-CoV-2 sequences available in the same database, we found that the Marseille-4 variant was characterized by a specific combination of 20 mutations compared to the Wuhan-Hu-1 strain. In particular, we found one mutation leading to an amino acid change S477N in the receptor binding domain of the spike protein that is targeted by the current vaccine. Using a specifically-designed qPCR assay, we observed that 12 to 100% of SARS-CoV-2 infections in Marseille were caused by this variant. Altogether, we identified 2106 cases of infection by this variant from September 1st, 2020, to January 20th, 2021. Compared to the 20A variant that predominated during the first epidemic phase, the Marseille-4 variant was associated to a higher frequency of hypoxemia. In addition, eleven patients developed a confirmed Marseille-4 infection months after a first SARS-CoV-2 infection, suggesting either a short term protective immunity or a lack of cross-immunity between the earlier strains and this variant. Its origin remains unknown but its sudden appearance, involving 13 hallmark mutations, points toward an animal reservoir, possibly the mink.

Conclusion. Together with the 20I/501Y.V1 variant in the United Kingdom and the 20H/501Y.V2 variant in South Africa, Marseille-4 should be considered as a major SARS-CoV-2 variant that emerged since summer 2020, became predominant locally and rapidly spread in Europe, Asia, Africa and North America. The protective role of the current vaccines against this variant should be evaluated.

TEXT

Introduction

The SARS CoV-2 epidemic that started in Wuhan, China, in December 2019, has rapidly spread around the world (https://coronavirus.jhu.edu/map.html). From January 2020, at the Mediterranee Infection institute (IHU) in Marseille, we set up the routine diagnosis of SARS-CoV-2 by PCR (*1*, *2*). The first SARS-CoV-2-infected patient was diagnosed at the IHU on 02/27/2020 (*3*, *4*) (https://www.mediterranee-infection.com/covid-19/). Since then, we performed more than 450,000 SARS-CoV-2 PCR tests, 2,000 virus isolations by cell culture, 2,000 whole genome sequencings and took care of 14,000 positive cases. In Europe, the SARS-CoV-2 epidemic was characterized by two major episodes. The first one, herein referred to as episode 1, started in February and almost ended in May (*5*). However, at the end of June a second episode suddenly occurred, exhibiting an atypical epidemic curve which led us to suspect that the two episodes were caused by distinct viral variants. Hence, we performed whole genome sequencing of SARS-CoV-2 strains over time to characterize their genetic diversity. This enabled us to identify 10 distinct genomic patterns that successively or concomitantly spread in the Marseille area (*6*). Of these, two variants were identified at high frequency in the population of individuals diagnosed at the IHU. The Marseille-1 variant caused mild infections in younger patients and predominated from the end of June to the end of July 2020 (*5*). We accumulated evidence indicating that this virus originated in Africa and was brought to Marseille by ferry boat travelers and sailors from North Africa (*5*). In France, the Marseille-1 variant did not spread outside Marseille and vanished rapidly. On July 29th, 2020, a new variant was identified and named Marseille-4 (**Figures 1, 2**).

Identification and circulation of the Marseille-4 variant

The highly transmissible Marseille-4 SARS-CoV-2 variant identified in Marseille at the end of July 2020 rapidly became predominant, reaching 100% of identified viral strains in

the area on November 2nd. Using genome sequences available through the GISAID database (https://www.gisaid.org/), we traced back the outbreaks of this variant in different countries. The first case of infection with the Marseille-4 variant, named 20A.EU2 in the Nexstrain classification (https://clades.nextstrain.org/) (*7*), was detected in a German patient on March 24th. Then, two cases were detected in a Balearic island, Spain, on May 29th and June 18th. Additional cases were detected in Southwestern France from July 9th, then in Denmark, and from August 1st in other European countries and other regions of France (**Figures 1, 2; Supplementary Figure S1**) (*7*). It was detected from September in North America (Canada, then USA), Australia and New Zealand, from October in Asia (Thailand, Hong Kong, Singapore and South Korea) and Africa (Tunisia and Morroco) and from December in Israel. In Marseille, 269 Marseille-4 complete genomes were sequenced from infected patients, and a Marseille-4-specific qPCR (Supplementary material) was designed that enabled rapid identification of an additional 1579 cases. Overall, this variant caused 2106 cases and accounted for about two-thirds of all SARS-CoV-2 viruses tested from September 2020 to January 2021 in our place.

Genomic features

The Marseille-4 variant evolved from clade 20A strains and is characterized by a combination of 20 mutations compared to the Wuhan-Hu-1 strain. Among these mutations, 13 (C4543T, G5629T, G9526T, C11497T, G13993T, G15766T, A16889G, G17019T, G22992A, C25710T, T26876C, G28975C, and G29399A) are hallmarks of this variant (**Supplementary Figures S2**). We provisionally subdivided the Marseille-4 variant into 11 subgroups (Marseille-4-A1 to Marseille-4-J) with a genetic drift ranging from 21 to 24 mutations compared to the Wuhan-Hu-1 strain (**Table 1**). Strikingly, comparative genomics shows that the set of 13 hallmark mutations appeared altogether. They are losses of a G in 7 cases and of a C in three cases, and are scattered along the viral genome. Seven (46%) are nonsynonymous

mutations, including two located in the RNA-dependent RNA polymerase (RdRp) (Nsp14; A176S and V767L), two in the NTPase/helicase (Nsp13; K1141R and E1184D), two in the nucleocapsid (N; M234I and A376T) and one in the spike glycoprotein (S; S477N). The latter substitution lies between mutations observed in viruses infecting humans and others seen in viruses infecting minks (**Figure 4**) (*8*). In addition, it is located in a domain of the spike protein that is a major target for neutralizing antibodies and the current vaccines (*9*). These data could explain the lack of resistance to infection by this Marseille-4 variant among people previously infected with different strains that circulated earlier during the first phase of the 2020 pandemic. Fifteen additional mutations (C222U, C503U, G2600U, A2647G, C8937U, G18105U, C23191U, G25534U, U26442C, G26720U, G27877U, C27942U, G28086U, G29701A, G29511U) have been observed in ≥ 5 viral genomes obtained in our institute. Overall, 283 nucleotide positions are mutated in ≥ 1 Marseille-4 genomes, mostly in the Nsp3 and S genes. They were most frequently C>U (36%), G>U (25%), U>C (8%), G>A (6%), and A>G (5%) mutations, and U>- deletions (6%). Phylogenetically, the Marseille-4 variant fell within a group of isolates from Europe only (**Figure 3; Supplementary Figure S3**).

The Marseille-4 variant harbors the S477N substitution within the RBD of the spike glycoprotein that attaches the virion to the cell membrane by binding to the viral receptor ACE2, and mediates viral entry (*10*), and is a major target of neutralizing antibodies (*11*) and the current vaccines (*9*) (**Figure 4**). S477N adds to the D614G substitution that was reported to increase the stability of spike trimers and confers greater affinity for ACE2 (*12*). It is worthy to note that the first genome available in the GISAID database that originates from Germany on March 24th 2020, did not harbor this S477N substitution, which may explain that it did not apparently spread further. Other critical mutations may be substitution Q57H in ORF3a, a viroporin that forms ion channels and was reported as required for viral replication, virulence and release, and is also predicted to be a pro-apoptotic protein (*13*, *14*), and

substitutions A176S in the RdRp and K1141R and E1184D in the NTPase/helicase.

In search for the origin of the Marseille-4 variant

The origin of the Marseille-4 variant is currently unknown. It emerged abruptly with its block of specific mutations, with no known intermediate form, while the SARS-CoV-2 epidemic had almost ended in France and Europe (**Figure 3**). This apparently discontinuous evolution of SARS-CoV-2 genomes is abnormal, particularly if we consider that after its first detection this variant had shown a subsequent mutation rate similar to that of other lineages (e.g., mutation in the RdRp did not alter the polymerase fidelity). Although we cannot exclude that the missing intermediate exists but has not been sequenced so far from COVID-19 patients, this could also suggest that there is an overlooked reservoir in which the virus was submitted to a selection pressure that favored a particular increase in mutation accumulation. Interestingly, among the 10,516 sequences from the Marseille-4 variant in the GISAID database, the 272 genomes from our laboratory had close relatives with those originating from Northern Europe, mostly Denmark (3,366), the UK (2,652) and Switzerland (1,147) (**Supplementary Figure S1**). A phylogenetic tree was constructed that included genomes from mink and human SARS-CoV-2 strains. Mink isolates were divided into five and six main groups, for the samples from the Netherlands and Denmark, respectively (**Supplementary Figure S4**). We observed a common phylogenetic node between mink isolates, the Marseille-4, Marseille-5, Marseille-6 variants and the 20H/501Y.V2 variant from England. This node pointed to the above-described common mutation, Q57H in ORF3a. The rapid emergence of the Marseille-4 variant during summer 2020, after the end of the first epidemic phase, may point toward an animal reservoir. Mink farms were identified as reservoirs and sources of SARS-CoV-2 mutants in the Netherlands in April (*15*), and in Denmark in June 2020 (*16*). In France, one of the four mink farms was infected and animals were culled. SARS-CoV-2 is an epizootic agent that caused an outbreak in humans before

being transferred to mink in which it spread rapidly through densely caged animal and subsequently became a source for human infection. To date, more than 800 human infections from minks have been reported (*15*). One hypothesis that may account for the discontinuous mutation rate could be that a human SARS-CoV-2 from infected caregivers infected mink, then the frequency of viral mutations changed in the mink due to a different host selection pressure, and this mink-adapted virus (with multiple mutations) became a new source of virus to infect humans. A genome obtained from a German patient sampled on March 24th (EPI_ISL_7079562020-03-24) is atypical as it is devoid of the S477N substitution, one of the Marseille-4 hallmark mutations, but harbors more mutations (31) than the other Marseille-4 strains, including in Nsp2, Nsp3, S and N proteins, and in ORF1b, particularly the Nsp14 exonuclease, which has a proofreading activity (*8*). The evolutionary relationships of this genome with other Marseille-4 genomes warrant a further investigation with the availability of other genomes obtained from samples collected during the same period.

Clinical findings: Marseille-4 variant may escape immunity conferred by a first SARS-CoV-2 infection

Compared to the 20A variant that predominated during episode 1, the Marseille-4 variant was associated with a lower frequency of cough, rhinitis and olfactory and gustatory disorders (**Table 2**). By contrast, hypoxemia was more frequent in patients infected with Marseille-4 variant. It was reported that differences observed in COVID-19 severity may in part be associated with dysfunction of cellular immune responses to SARS-CoV-2 and/or weakness of neutralizing humoral response (*17*). We diagnosed two successive COVID-19 infections, separated by more than 4 months, in 11 patients. The first infection was diagnosed when only the Wuhan-Hu-1 genotype was circulating in Marseille (*18*), and we obtained genomic or PCR (1 and 10 patients, respectively) confirmation that the second episode was caused by the Marseille-4 variant. This suggests either a short protective immunity (only a

few weeks or months) as previously observed with seasonal coronaviruses (*19*), or a lack of cross-immunity between different SARS-CoV-2 variants, allowing Marseille-4 to evade immune protection elicited by another earlier variant. This may be related to the S477N mutation which could change the affinity of RBD for ACE2 and decrease the sensitivity of the variant virus to anti-RBD-specific neutralizing antibodies (*20*).

Conclusion

The evolution of the SARS-CoV-2 virus may reflect the generation of new variants in different ecosystems that may spread with globalization and replace the original variant issued from Wuhan. The ecosystems allowing this selection may consist of human groups isolated for a while, or animal reservoirs such as the huge mink farms. Large concentrations of farmed minks were infected by human SARS-CoV-2. Under these conditions, sub-speciation may occur (*21*). In the present case, the re-connection of isolated ecosystems (either countries and/or farmed animals) where different variants had developed generated new outbreaks in countries that were exposed to incoming populations such as travelers. This was in particular the case for Mediterranean countries and, in France, for Marseille that received an elevated number of tourists in summer 2020. We believe that the segregation of viral strains in isolated geographical areas and in animal reservoirs, may indeed contribute to explain the differences observed among epidemic curves around the world. This would help to understand the mechanism of the second episode that developed in Marseille, initially caused by an African variant that disappeared, and then by emerging new variants linked to different areas of Europe, including those hosting huge mink farms. Finally, the role of the treatment of COVID-19 by remdesivir (*22*) or hyperimmune plasma (*19*) in generating and selecting variants may also have contributed to the new outbreaks observed in the most developed countries.

Acknowledgments

We are grateful to Olivia Ardizzoni, Vincent Bossi, Madeleine Carrera, Vera Esteves-Vieira, Laurence Thomas, Priscilla Jardot and Raphael Tola for their technical help, and to Audrey Giraud-Gatineau and Léa Luciani for their help with data analysis.

Ethical approval

The study was approved by the ethical committee of the Mediterranee Infection institute under reference 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.

Funding

This work was supported by the French Government under the "Investments for the Future" programme managed by the National Agency for Research (ANR), Mediterranee-Infection 10-IAHU-03 and was also supported by Region Provence Alpes Cote d'Azur and European funding FEDER PRIMMI (Fonds Européen de Developpement Regional-Plateformes de Recherche et d'Innovation Mutualisées Mediterranée Infection), FEDER PA 0000320 PRIMMI.

Transparency declaration

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.

REFERENCES

1. J. C. Lagier *et al.*, *Travel. Med. Infect. Dis.* **36**, 101791 (2020).
2. P. Colson *et al.*, *Int. J. Infect. Dis.* **101**, 121 (2020).
3. P. Colson *et al.*, *Eur. J. Clin. Microbiol. Infect. Dis.* **39**, 1601 (2020).
4. B. D. Brahim *et al.*, *Front Microbiol.* **11**, 2014 (2020).
5. P. Colson *et al.*, *Travel. Med. Infect. Dis.* **in press**, (2021).
6. P. Colson et al. *IHU preprints*. doi: https://doi.org/10.35088/68c3-ew82 (2020).
7. E. B. Hodcroft *et al.*, *medRxiv* https://doi.org/10.1101/2020.10.25.20219063, (2020).
8. R. F. Garry, *Virological. org* **PrePrint**, (2021).
9. L. Dai, G. F. Gao, *Nat. Rev. Immunol.* (2020).
10. J. Lan *et al.*, *Nature* **581**, 215 (2020).
11. C. O. Barnes *et al.*, *Cell* **182**, 828 (2020).
12. B. Korber *et al.*, *Cell* **182**, 812 (2020).
13. M. Bianchi, A. Borsetti, M. Ciccozzi, S. Pascarella, *Int. J Biol. Macromol.* **170**, 820 (2021).
14. P. T. W. Law *et al.*, *J. Gen. Virol.* **86**, 1921 (2005).
15. B. B. Oude Munnink *et al.*, *Science* **371**, 172 (2021).
16. A. S. Hammer *et al.*, *Emerg. Infect. Dis.* **27**, (2020).
17. C. R. Moderbacher *et al.*, *Cell* **183**, 996 (2020).
18. .P. Colson *et al. J. Infect.* S0163-4453(20)30706-4. doi: 10.1016/j.jinf.2020.11.011. Online ahead of print. (2020).
19. A. W. D. Edridge *et al.*, *Nat. Med.* **26**, 1691 (2020).
20. E. Andreano *et al.*, *bioRxiv* preprint, doi: https://doi.org/10.1101/2020.12.28.424451 (2020).
21. C. Darwin, *On the origin of species* (John Murray, London, 1859).

22. B. Choi *et al.*, *N. Engl. J. Med.* **383**, 2291 (2020).

FIGURES

Figure 1. Schematic of the evolution of the SARS-CoV-2 Marseille-4 variant in Europe.

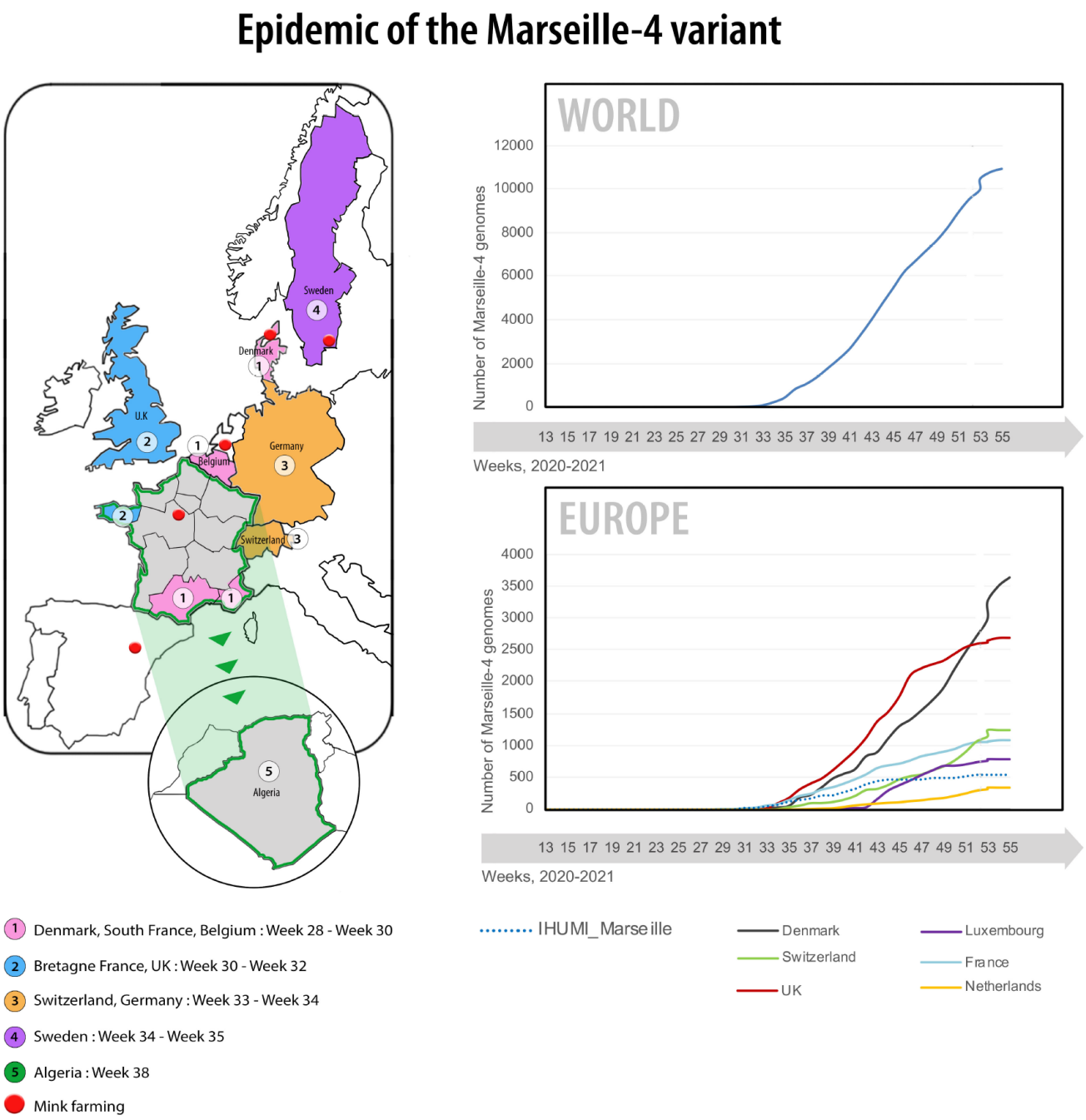
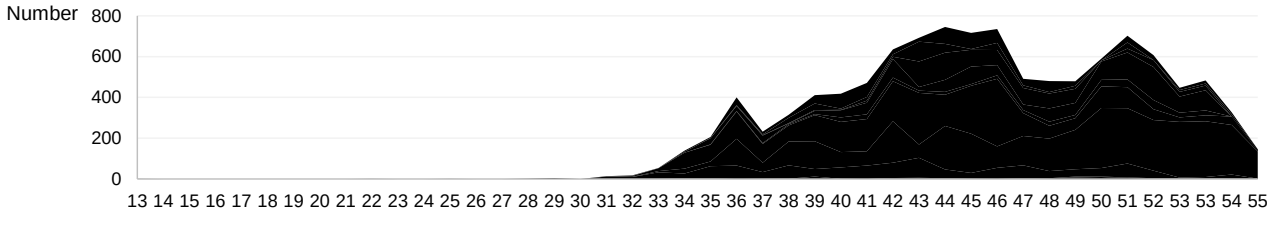


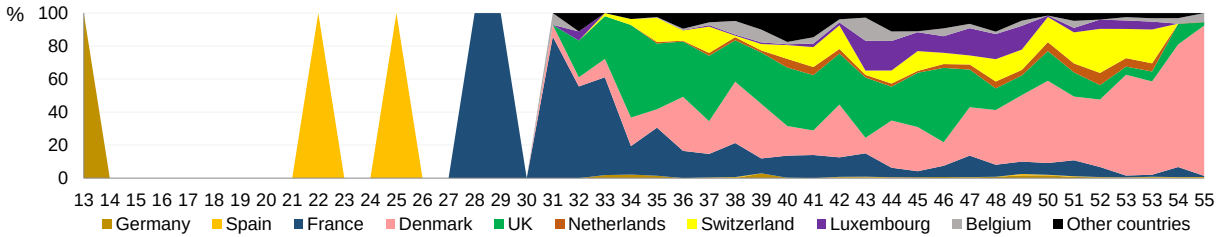
Figure 2. Evolution of the Marseille-4 variant over time.

a. World

a1. Weekly number of genomes of the Marseille-4 variant

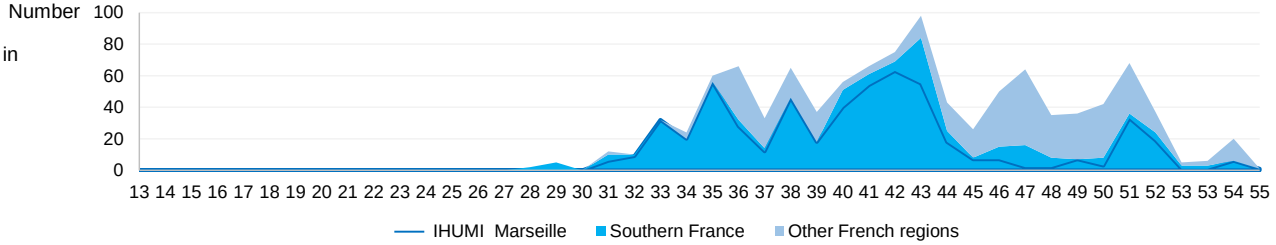


a2. Weekly frequency normalized to 100% of the countries where genomes of the Marseille-4 variant were obtained

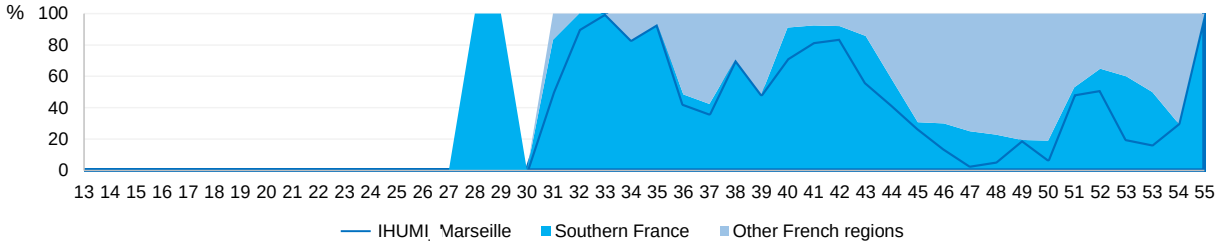


b. France

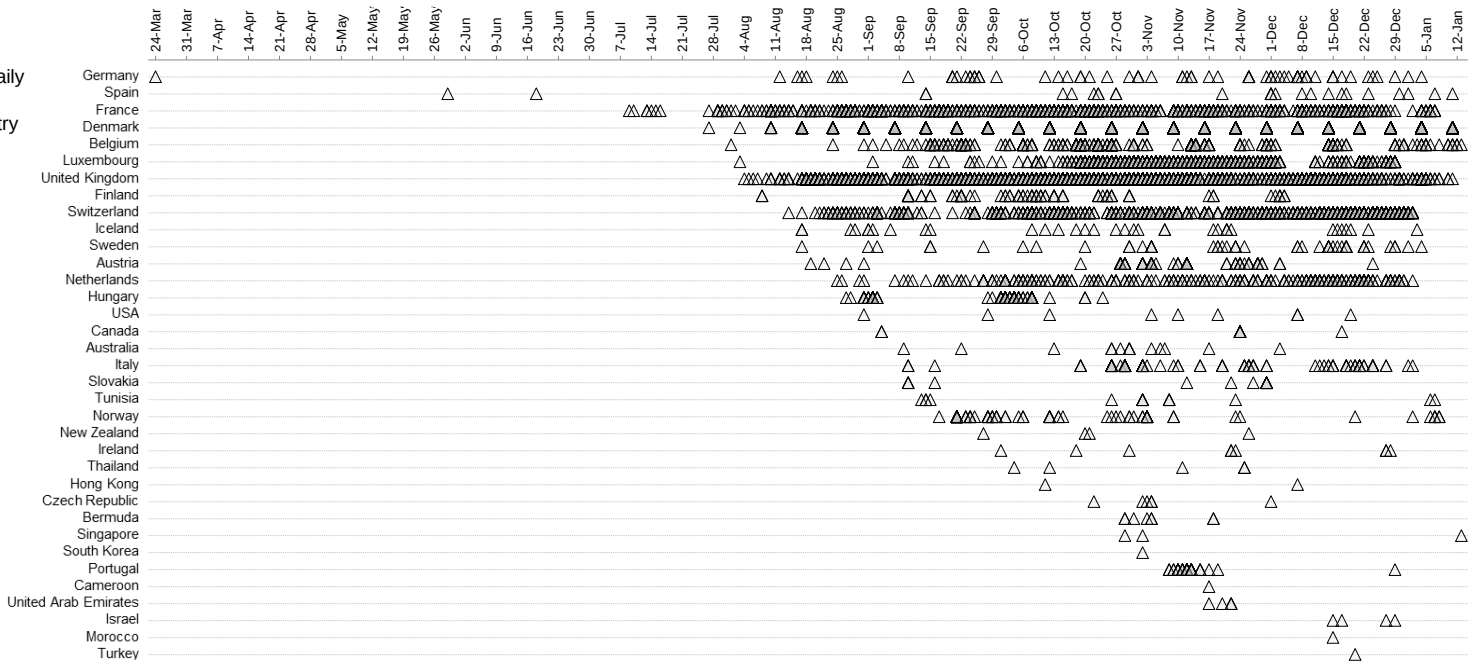
b1. Weekly number of genomes of the Marseille-4 variant in French regions



b2. Weekly frequency normalized to 100% of the French regions where genomes of the Marseille-4 variant were obtained



c. Time distribution of the daily number of genomes of the Marseille-4 variant per country



a. 1: Weekly number of genomes of the Marseille-4 variant worldwide; **2:** Weekly frequency normalized to 100% of the countries where genomes of the Marseille-4 variant were obtained.

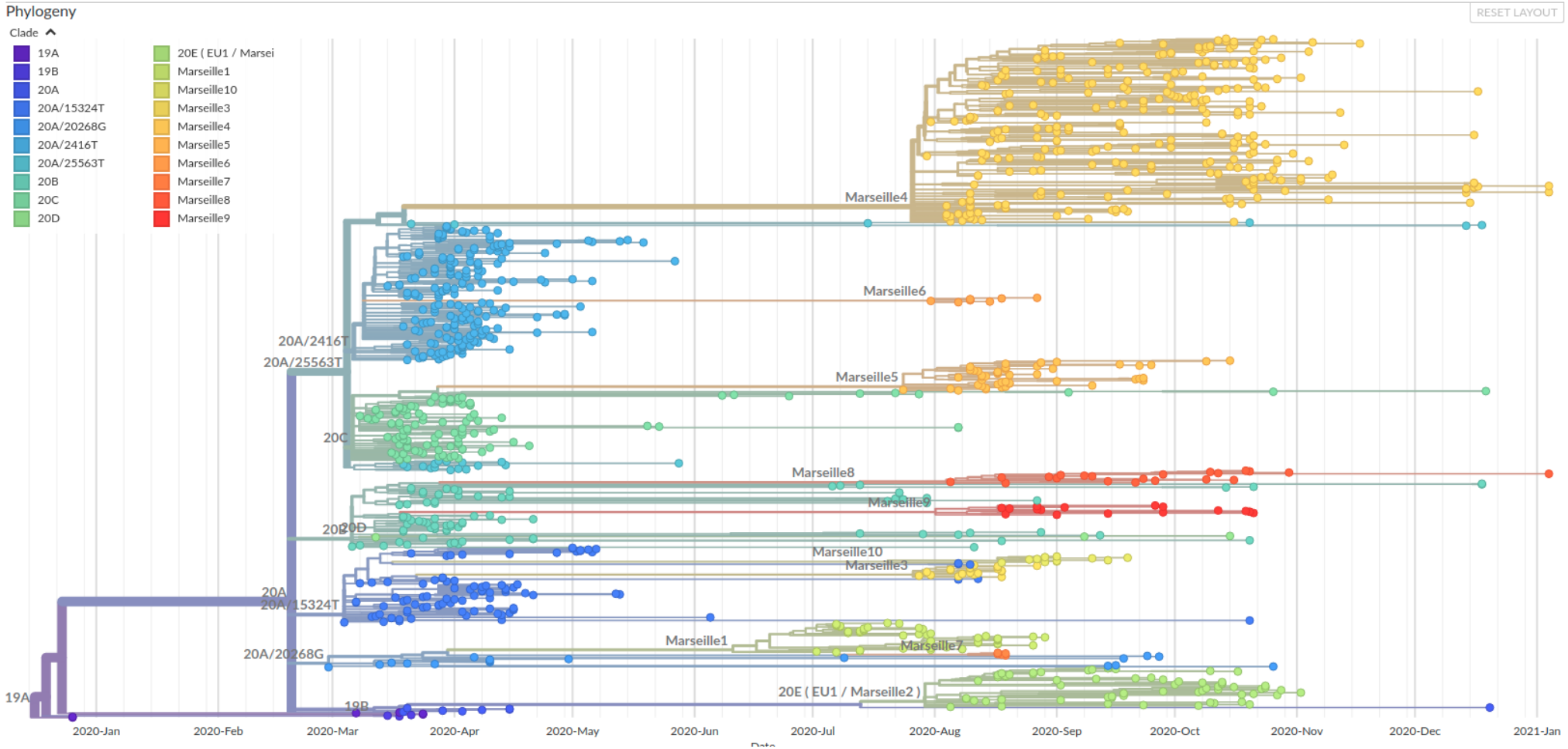
b. 1: Weekly number of genomes of the Marseille-4 variant in French regions; **2:** Weekly frequency normalized to 100% of the French regions where genomes of the Marseille-4 variant were obtained

c. Time distribution of the daily number of genomes of the Marseille-4 variant per country.

Figure 3. Genome sequence-based phylogenetic trees showing the evolution of SARS-CoV-2 Marseille-4 variant strains.

Full-length genome sequences obtained in our study were compared to those available in the GISAID database (https://www.gisaid.org/). Phylogenetic trees were reconstructed and visualized by using the Nextclade and iTOL (https://itol.embl.de/) web tools, respectively. **A**. Time-scale phylogenetic tree. **B**. Phylogenetic tree based on mutational events. We used the Nextstrain pipeline for our phylogenetic analyses (https://github.com/nextstrain/ncov/).

A


Phylogeny
RESET LAYOUT
Clade
19A
19B
20A
20A/15324T
20A/20268G
20A/2416T
20A/25563T
20B
20C
20D
20E (EU1 / Marsei
Marseille1
Marseille10
Marseille3
Marseille4
Marseille5
Marseille6
Marseille7
Marseille8
Marseille9
Marseille4
Marseille6
20A/2416T
20A/25563T
Marseille5
20C
Marseille8
Marseille9
20D
Marseille10
Marseille3
20A
20A/15324T
Marseille1
Marseille7
20A/20268G
20E (EU1 / Marseille2)
19A
19B
2020-Jan
2020-Feb
2020-Mar
2020-Apr
2020-May
2020-Jun
2020-Jul
2020-Aug
2020-Sep
2020-Oct
2020-Nov
2020-Dec
2021-Jan

B

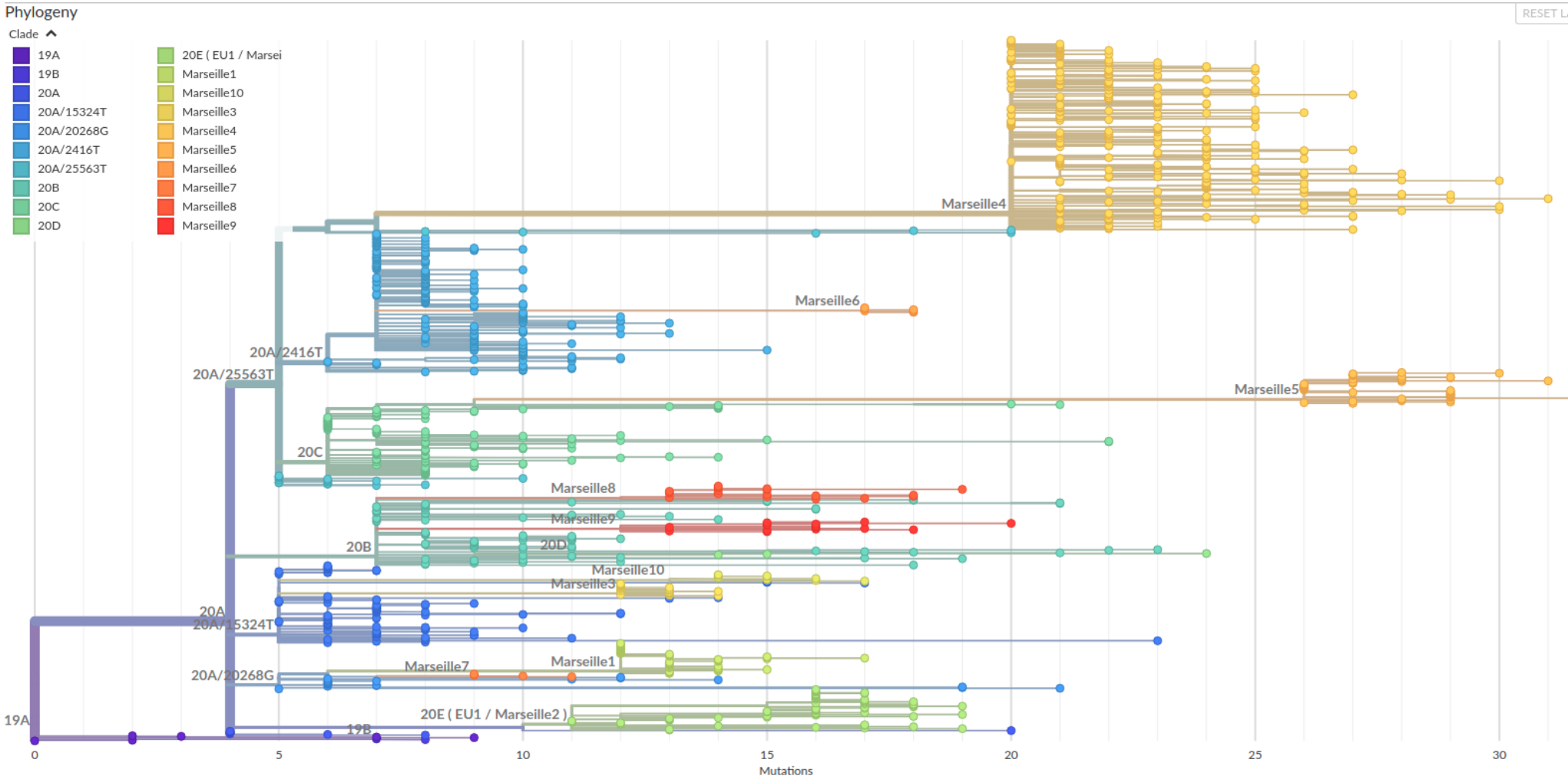
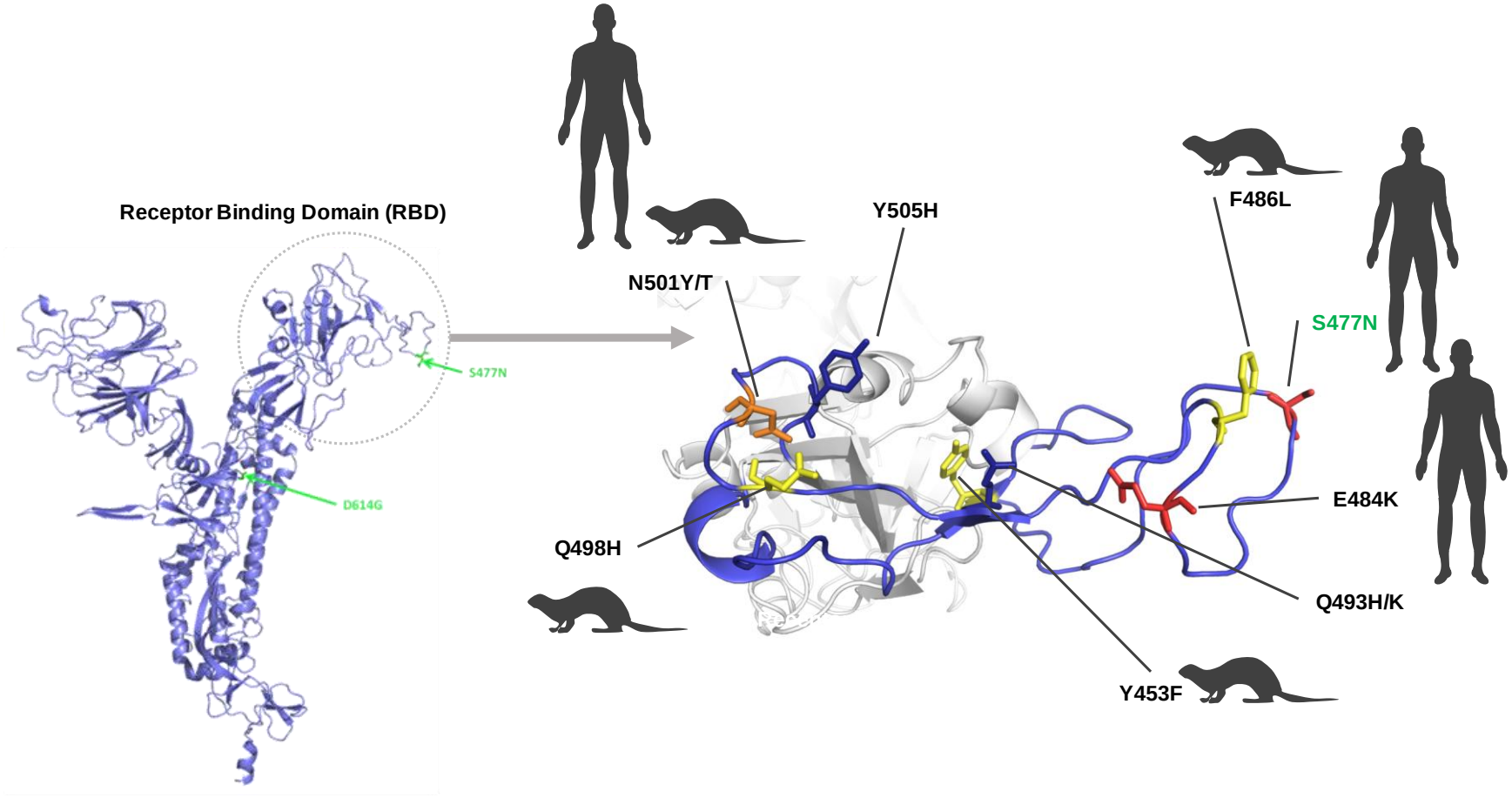

Phylogeny
Clade
19A
19B
20A
20A/15324T
20A/20268G
20A/2416T
20A/25563T
20B
20C
20D
20E (EU1 / Marsei
Marseille1
Marseille10
Marseille3
Marseille4
Marseille5
Marseille6
Marseille7
Marseille8
Marseille9
Marseille4
Marseille6
20A/2416T
20A/25563T
Marseille5
20C
Marseille8
Marseille9
20B
20D
Marseille10
Marseille3
20A
20A/15324T
Marseille7
Marseille1
20A/20268G
20E (EU1 / Marseille2)
19A
19B
0
5
10
15
20
25
30
Mutations

Figure 4. 3D structure of the spike protein showing the amino acid substitutions in the receptor-binding motif of the Marseille-4 variant and of other variants detected in humans and/or minks

The structure was predicted using the Phyre2 web portal (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) and visualized using the Pymol tool v.1.8 (https://pymol.org/2/). Amino acids where a substitution was observed in humans are colored in red, where a substitution was observed in minks are colored in yellow, and those where a substitution was observed in humans and minks are colored in orange.



TABLES

Table 1. Nucleotide mutations and amino acid substitutions in the genomes of SARS-CoV-2 Marseille-4 variants

								Wuhan	20A	20A/25563T	20A/18877T	20A/26735T	Marseille4	Marseille4-A	Marseille4-A1	Marseille4-B	Marseille4-C	Marseille4-D	Marseille4-E	Marseille4-F	Marseille4-G	Marseille4-H	Marseille4-I	Marseille4-J
Nucleotide position	**Gene**	**WT nt**	**Mutated nt**	**Codon change**	**Codon number**	**Amino acid substitution**																		
									0	4	5	6	7	20	21	22	24	21	21	22	21	21	21	21	22
222	5UTR	C	T				17	1	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	N	C
241	5UTR	C	T				1016	11	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	N	T
503	nsp1	C	T	CCT>TCT	80	P80S	17	1	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C
2600	nsp2	G	T	GTT>TTT	599	V599F	9	1	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G
2647	nsp2	A	G	AAA>AAG	614	K	7	1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G
3037	nsp3	C	T	TTC>TTT	106	F	1030	12	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
4543	nsp3	C	T	ACC>ACT	608	T	268	12	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
5629	nsp3	G	T	ACG>ACT	970	T	269	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T
6539	nsp3	C	T	CAC>TAC	1274	H1274Y	6	1	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C
8937	nsp4	C	T	GCA>GTA	128	A128V	4	1	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C
9526	nsp4	G	T	ATG>ATT	324	M324I	269	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T
11497	nsp6	C	T	TAC>TAT	175	Y	268	12	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
13993	nsp12b	G	T	GCT>TCT	176	A176S	269	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T
14408	nsp12b	C	T	CCT>CTT	314	P314L	1028	12	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
15766	nsp12b	G	T	GTG>TTG	767	V767L	268	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T
16889	nsp13	A	G	AAA>AGA	218	K218R	269	12	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G
17019	nsp13	G	T	GAG>GAT	261	E261D	269	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T
18105	nsp14	G	T	CAG>CAT	22	Q22H	6	1	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G
18877	nsp14	C	T	CTA>TTA	280	L	272	12	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T
22992	S	G	A	AGC>AAC	477	S477N	269	12	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A
23191	S	C	T	TTC>TTT	543	F	13	1	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C
23403	S	A	G	GAT>GGT	614	D614G	1028	12	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
25534	ORF3a	G	T	GTT>TTT	48	V48F	7	1	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T
25563	ORF3a	G	T	CAG>CAT	57	Q57H	657	12	G	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
25710	ORF3a	C	T	CTC>CTT	106	L	272	12	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
26442	E	T	C	AAT>AAC	66	N	6	1	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T
26720	M	G	T	GTG>GTT	66	V	16	1	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G
26735	M	C	T	TAC>TAT	71	Y	272	12	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T
26876	M	T	C	ATT>ATC	118	I	269	12	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C
27877	ORF7b	G	T	TGT>TTT	41	C41F	6	1	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G
27942	ORF8	C	T	CAC>TAC	17	H17Y	12	1	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C
28086	ORF8	G	T	GCT>TCT	65	A65S	30	2	G	G	G	G	G	G	T	T	G	G	G	G	G	G	G	G	G
28975	N	G	C	ATG>ATC	234	M234I	268	12	G	G	G	G	G	C	C	C	C	C	C	C	C	C	C	C	C
29399	N	G	A	GCT>ACT	376	A376T	263	12	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A
29511	N	G	T	AGT>ATT	413	S413I	6	1	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G
29701	3UTR	G	A				12	1	G	G	G	G	G	G	G	G	G	G	N	G	G	A	G	G	G

Table 2. Demographics, outcomes and clinical symptoms in patients infected with different SARS-CoV-2 variants

Demographics and outcomes (N=759)	20A (N=339)		Marseille-4 (N=420)		p-value*
	n	%	n	%	
Male gender	151	44.5	216	51.4	0.059
Age (mean ± SD)	50.2 ± 22.3		48.9 ± 23.1		0.41
Hospitalization	53	15.6	68	16.2	0.835
Transfer to intensive care unit	5	1.5	10	2.4	0.44
Death	10	2.9	16	3.8	0.52
Symptoms (N=444)	**20A (N=254)**		**Marseille-4 (N=190)**		**p-value***
	n	%	n	%	
Cough	123	48.4	73	38.4	**0.036**
Rhinitis	106	41.7	37	19.5	**<0.0001**
Anosmia	76	29.9	35	18.5	**0.006**
Ageusia	71	27.9	34	18.0	**0.015**
Dyspnea	72	28.3	42	22.1	0.136
SpO2 <96%	37	14.6	42	22.1	**0.04**

* Chi2 or Fisher exact test for qualitative variables. Student test for quantitative variables

SUPPLEMENTARY MATERIAL

Supplementary Figures: 4; **Supplementary Table:** 1

SUPPLEMENTARY METHODS

Phylogeny

A total of 1038 SARS-CoV-2 genomes were integrated in the phylogenetic analysis. All genomes were aligned using MAFFT version 7 (*1*). A phylogenetic tree was reconstructed using IQ-TREE with the GTR Model with ultra fast bootstrap of 1000 repetitions (*2*). Phylogenetic tree was visualized with iTOL (Interactive Tree Of Life, (https:// itol.embl.de/)).

PCR detection of SARS-CoV-2 Marseille-4 variant (a.k.a. Nextstrain clade 20A.EU2)

A qPCR system targeting nucleotide positions 9,460-9,543 in reference to genome GenBank accession number NC_045512.2 (Wuhan-Hu-1 isolate) within the nsp4 gene, was designed. The primers and probe are described in Supplementary Table S1.

Table S1. Primers and probe

Name	Sequence (5'-3')	Positions *
Primers:		
Pri_IHU_ C4_5_MBF	GAGGTTTAGAAGAGCTTTTGGTGA	9,460-9,483
Pri_IHU_ C4_5_MBR	CCAGGTAAGAATGAGTAAACTGGTG	9,549-9,573
Probe (6FAM-labelled):		
Pro_IHU_ C4_5_MBP	CCTTATTTCATTCACTGTACTCTG	9,520-9,543

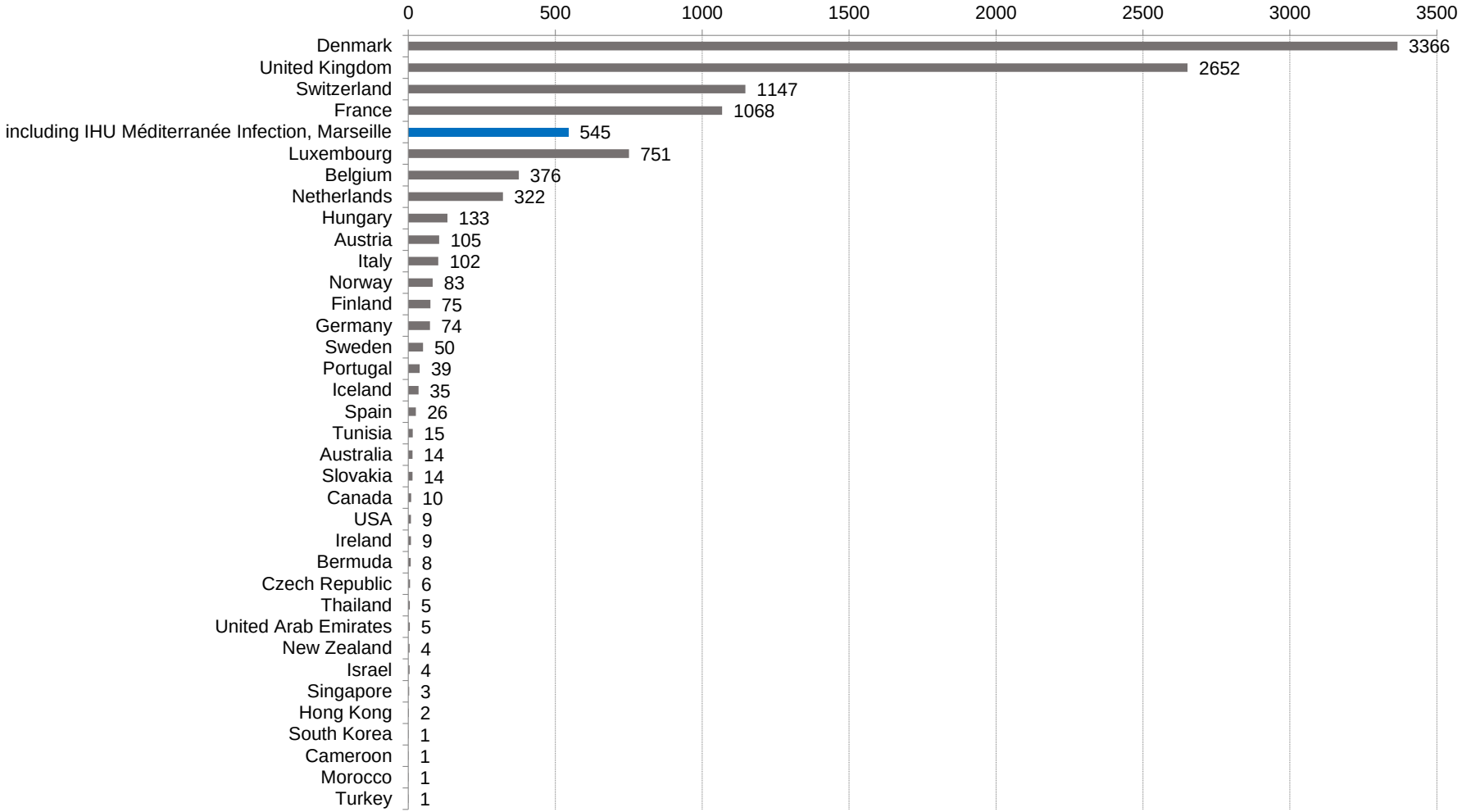
* in reference to genome GenBank accession number NC_045512.2 (Wuhan-Hu-1 isolate). The nucleotide carrying the mutation specific of the Marseille-4 variant is covered by the probe and underlined.

The qPCR was performed by adding 5 µL of extracted viral RNA to 15 µL of reaction mixture containing 5 µL of 4X TaqMan Fast Virus 1-Step Master Mix (Thermo Fisher Scientific, Grand Island, NY, USA), 0.5 µL of forward primer (10 pmol/µL), 0.5 µL of reverse primer (10 pmol/µL), 0.4 µL of probe (10 pmol/µL), and 8.6 µL of water. PCR conditions were as follows: reverse transcription at 50°C for 10 min, then a hold at 95°C for

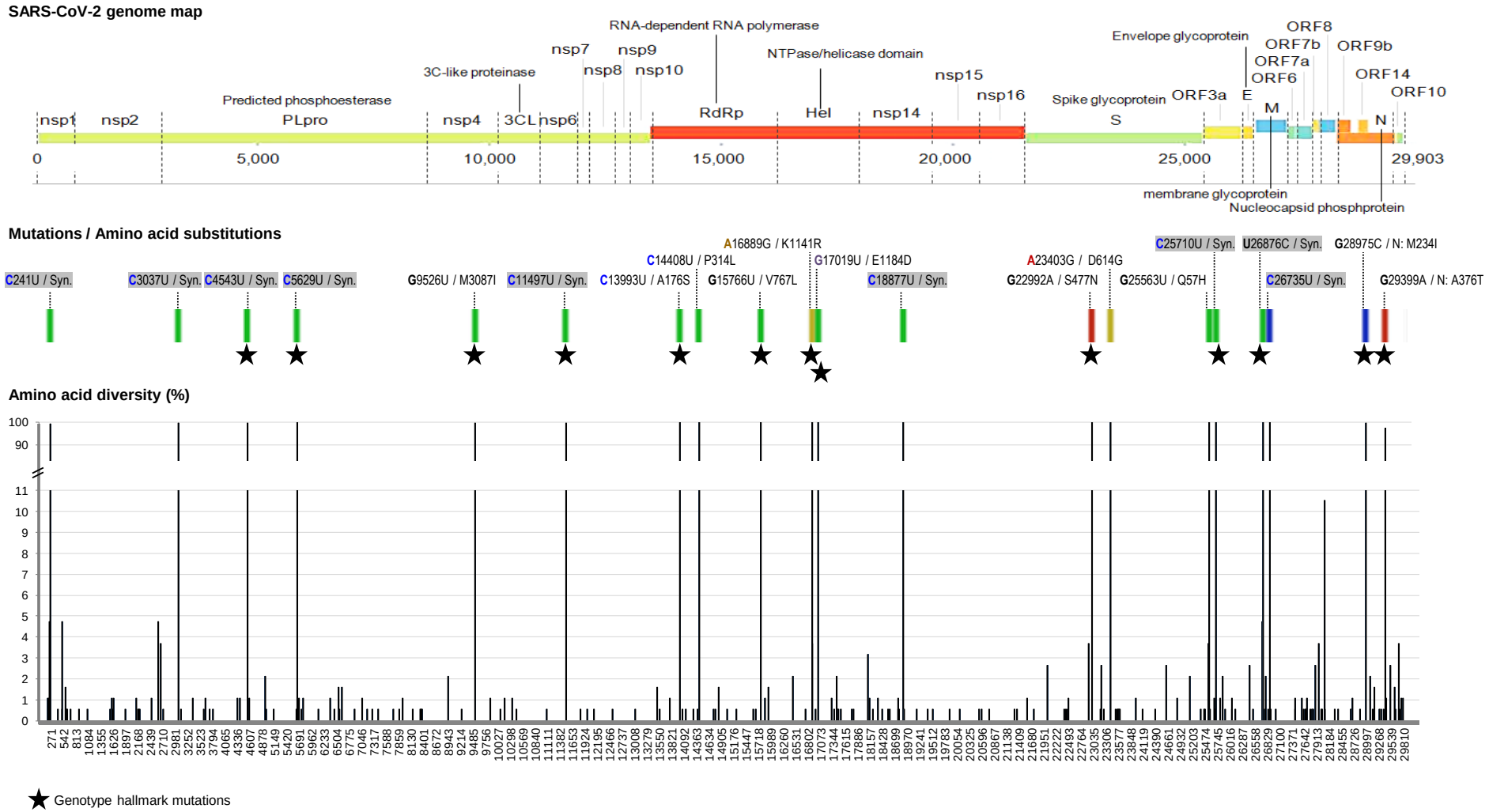
20 sec followed by 40 cycles comprising a denaturation step at 95°C for 15 sec and a hybridization-elongation step at 60°C for 60 sec. This qPCR was run on a LC480 thermocycler (Roche Diagnostics, Mannheim, Germany).

SUPPLEMENTARY FIGURES

Supplementary Figure S1. Numbers of SARS-CoV-2 Marseille-4 variant genomes deposited in the GISAID database and available in our sequence database according to the country of origin

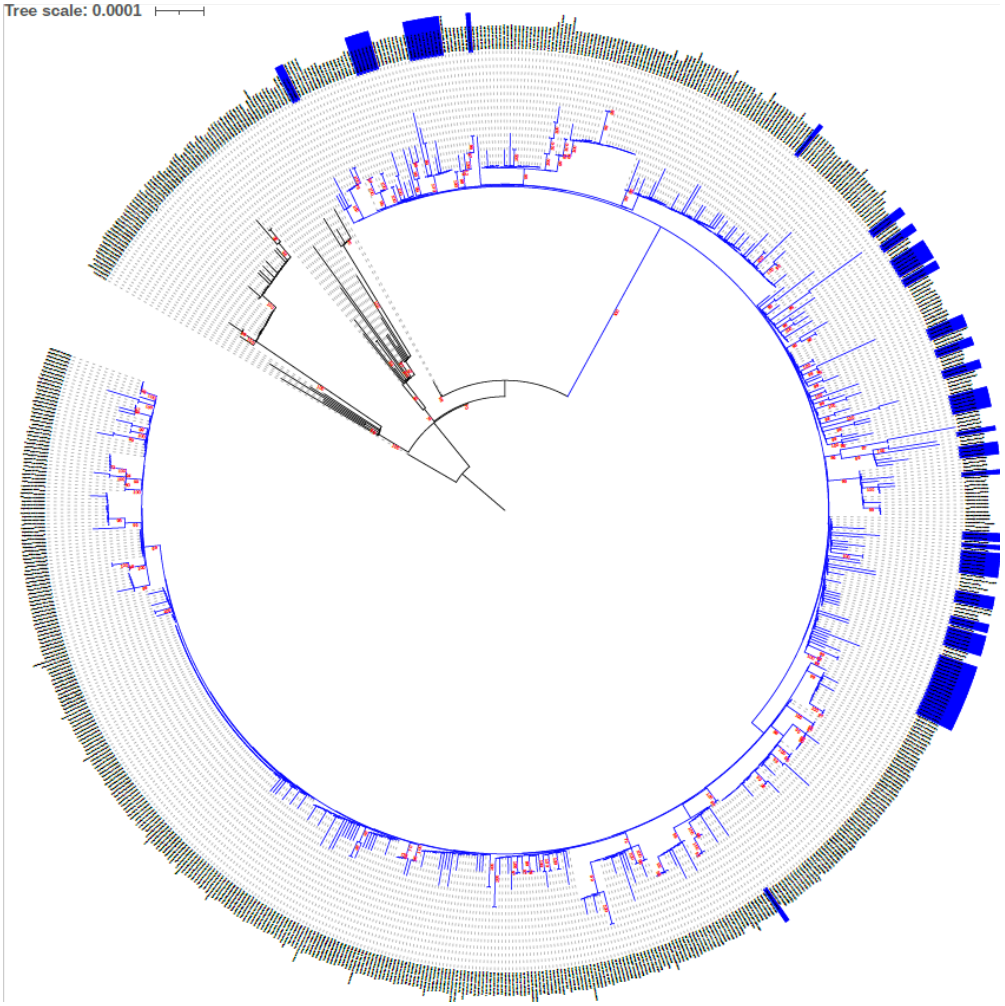


Supplementary Figure S2. Mutations, amino acid substitutions and diversity in SARS-CoV-2 Marseille-4 strains.



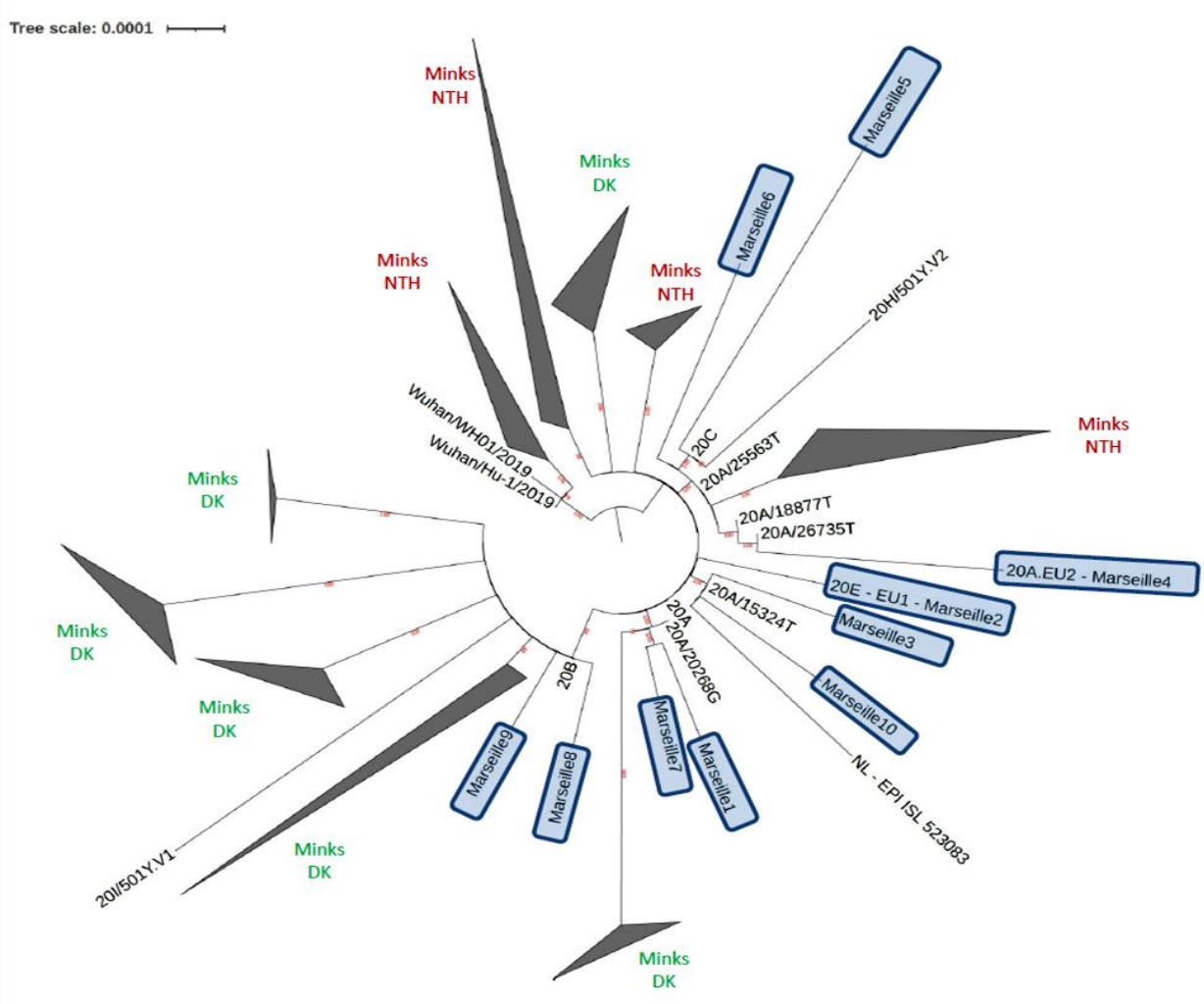
Supplementary Figure S3. Phylogenetic tree based on SARS-CoV-2 full-length genomes.

Phylogenetic tree reconstructed from full-length viral genomes obtained from clinical samples. Phylogenetic trees were reconstructed by using the GISAID TreeTool in v2.0 that performs an initial approximate maximum likelihood phylogeny reconstruction using FastTree then a refinement by RaXML.



Supplementary Figure S4. Phylogenetic tree based on SARS-CoV-2 full-length genomes. A total of 744 genomes of SARS-CoV2 were integrated in a phylogenetic analysis. All genomes were aligned by usinf MAFFT version 7 (*1*). Phylogenetic tree was reconstructed by using IQ-TREE with the GTR model with ultra fast bootstrap of 1000 repetitions (*2*), and visualized with iTOL (Interactive Tree Of Life, (https:// itol.embl.de/)).

DK, Denmark; NTH, The Netherlands.



REFERENCES

1. K. Katoh, D. M. Standley, *Mol. Biol. Evol.* **30**, 772 (2013).

2. B. Q. Minh *et al.*, *Mol. Biol. Evol.* **37**, 1530 (2020).