Genomic diversity and evolution of coronavirus (SARS-CoV-2) in France from 309 COVID-19-infected patients

Anthony LEVASSEUR[1,2,3#], Jeremy DELERCE[1,2], Aurelia CAPUTO[1,2], Ludivine BRECHARD[1,2], Philippe COLSON[1,2], Jean-Christophe LAGIER[1,2], Pierre-Edouard FOURNIER[2,4], Didier RAOULT[1,2#]

[1] Aix-Marseille Université (AMU), MEPHI (Microbes, Evolution, Phylogeny and Infections), IRD, APHM, Faculté de Médecine, Marseille, France

[2] IHU Méditerranée Infection, Marseille, France

[3] Institut Universitaire de France, Paris, France

[4] Aix-Marseille Université (AMU), IRD, APHM, SSA, VITROME, Marseille, France

Corresponding authors: Didier Raoult: didier.raoult@gmail.com; Anthony Levasseur : anthony.levasseur@univ-amu.fr

Data deposition: genomes were deposited at EMBL-EBI under the BioProject: PRJEB37722

ABSTRACT

The novel coronavirus (SARS-CoV-2) causes pandemic of viral pneumonia. The evolution and mutational events of the SARS-CoV-2 genomes are critical for controlling virulence, transmissibility, infectivity, severity of symptoms and mortality associated to this infectious disease. We collected and investigated 309 SARS-CoV-2 genomes from patients infected in France. Detailed genome cartography of all mutational events (SNPs, indels) was reported and correlated to clinical features of patients. A comparative analysis between our 309 SARS-CoV-2 genomes from French patients and the reference Wuhan coronavirus genome revealed 315 substitution mutations and six deletion events: ten were in 5'/3' UTR, 178 were nonsynonymous, 126 were synonymous and one generated a stop codon. Six different deleted areas were also identified in nine viral variants. In particular, 30 substitution mutations (18 nonsynonymous) and one deletion (Δ21765-21770) concerned the spike S glycoprotein. An average of 7.8 mutational events (+/- 1.7 SD) and a median of 8 (range, 7-9) were reported per viral isolate. Comparative analyses and clustering of specific mutational signatures in 309 genomes disclose several divisions in groups and subgroups combining their geographical and phylogenetic origin. Clinical outcomes of the 309 COVID-19-infected patients were investigated according to the mutational signatures of viral variants. These findings highlight the genome dynamics of the coronavirus 2019-20 and shed light on the mutational landscape and evolution of this virus. Inclusion of the French cohort enabled us to identify 161 novel mutations never reported in SARS-CoV-2 genomes collected worldwide. These results support a global and continuing surveillance of the emerging variants of the coronavirus SARS-CoV-2.

Keywords: French cohort, SARS-CoV-2, coronavirus, evolution, bioinformatics, outbreak, phylogeny, COVID-19 infected patients

INTRODUCTION

The new coronavirus SARS-CoV-2 has now spread to every country in the world. SARS-CoV-2 is classified among the positive-sense single-stranded (ss)RNA viruses as the seventh coronavirus known to infect human[1]. Four genera of coronaviruses have been identified in the subfamily Orthocoronavirinae of the Coronaviridae family (order Nidovirales). Described for the first time in December 2019 in Wuhan, China, the new SARS-CoV-2 coronavirus will naturally accumulate and fix mutations in its RNA genome as the epidemic evolves, partly due to the high immunological pressure in humans. A number of characteristics of coronaviruses are still unknown and one of the key questions is how fast is the SARS-CoV-2 virus evolving? Indeed, mutations could drastically result in difference in virulence, transmissibility, and clinical evolution. Therefore, the frequency and pattern of mutations are essential features to be considered in this ongoing global epidemic.

The genome of SARS-CoV-2 is a positive ssRNA of approximately 29.9 kbases in length, with a 5'cap structure and 3'polyadenine tail[2]. Two large ORFs are translated into polyproteins arising from ribosomal frameshift that are subsequently post-translationally processed to produce 16 proteins conserved between coronaviruses[3]. These 16 proteins are involved in the synthesis of viral RNA and immune evasion[4]. Four structural proteins (spike S, envelope E, membrane M and nucleocapsid N) and at least nine small accessory proteins, including proteins unique to SARS-CoV-2, are produced during viral genome replication[3,5].

By comparing alpha- and beta-coronaviruses, it appears that SARS-CoV-2 exhibits notable genomic features such as an optimized binding to the human receptor ACE2 and a functional polybasic (furin) site in the spike protein with predicted O-linked glycans[6]. The receptor binding domain (RBD) exhibits the most variable fraction of the genome of SARS-CoV-2 (only ~40% amino acid identity with other SARS-CoV coronaviruses). The origin of this virus is debated but two main hypotheses emerged based on either the natural selection in an animal host before zoonotic transfer or natural selection in humans following zoonotic transfer[7]. Nearly identical RBDs found in pangolins coronaviruses and SARS-CoV-2 could indicate an acquisition in SARS-CoV-2 by recombination and

mutations in a parsimonious scenario[6]. Metagenomic sequencing identified coronaviruses close to SARS-CoV-2 from Malayan pangolin and confirmed the natural reservoir of this virus in wild animals and the risk of zoonotic transmission[8].

Being a RNA virus, SARS-CoV-2 is prone to mutate because of the lack of proofreading activity of polymerase. Previous work in SARS-CoV reported an exoribonuclease domain (ExoN) in non-structural protein 14 that could provide proofreading activity protecting the virus from high rate of mutagenesis[9]. Mutations have to be carefully studied and related to the severity of symptoms and clinical evolution associated to patients. In addition, the mutation rate correlates with drug resistance and escape from immune surveillance. The emerging respiratory disease outbreaks caused by the SARS-CoV-2 required an urgent need for comprehensive studies that combine genomic data, epidemiological data, and chart records of the clinical symptoms and clinical evolution of patients. In addition to genomic mutations, transcriptomic studies have revealed several viable deletions in essential proteins, including glycoprotein S reinforcing that several regions are prone to mutate[10].

The 2019–20 coronavirus pandemic was confirmed to have spread to France since January 2020. In this work, 309 genomes of SARS-CoV-2 from French patients monitored in IHU Méditerranée Infection were studied and compared to the isolates from different countries in all continents. Mutational events were evidenced and compared to the reference Wuhan coronavirus isolate. Finally, clinical issues and impacts of these mutations were investigated.

RESULTS

Mutational landscape of the SARS-CoV-2 genomes from the French cohort

The mutational landscape of 309 coronavirus genomes was tracked in terms of SNPs and insertions/deletions (Table 1). As compared to the reference genome SARS-CoV-2 isolated in Wuhan in December 2019, a total of 321 mutational events were reported in the SARS-CoV-2 genomes from 309 patients. A mean of 7.8 mutational nucleotide substitutions (+/- 1.7 SD) and a median of 8 (range, 7-9) were reported per viral isolate. Ten substitutions were retrieved in the non-coding part of the genome. In the coding region, 305 nucleotide substitutions were detected representing 178

nonsynonymous (~58.4%) and 126 synonymous mutations (~41.3%) and one stop codon mutation (~0.3%). Some mutations were commonly distributed in the majority of our isolates whereas 208 substitutions (~68.2%) were unique to one isolate. Regarding missense mutations, we noted drastic changes of aminoacid (AA) in which the physicochemical properties of the AA were severely interfered (Table 2). In particular, the spike gene contained 30 substitutions (18 nonsynonymous, 60%) and included one deletion of six nucleotides (Δ21765-21770). Among mutations retrieved in the spike gene, the missense mutation P330S (22550, CCT>TCT) was embedded at the border of the receptor-binding domain (RBD). A stop mutation was also found in ORF7a at the protein position 38, meaning that the protein was prematurely truncated at the first third of its sequence.

Interestingly, we identified six putative deleted regions distributed in nine viral variants among the genes nsp1, nsp2, spike S, ORF3a and ORF7a. These deletions removed entire codons for nsp1 (Δ510-518 and Δ686-694, three AA), nsp2 (Δ1605-1607, one AA), spike protein (Δ21765-21770, two AA), ORF7a (Δ27434-27436, one AA) and consequently, conserved open reading frames. Moreover, the deletion event retrieved in nsp1 (position Δ686-694) was found in three different patients and the deletion in nsp2 (position Δ1605-1607) in two different patients that were not epidemiologically-related, meaning that these two evolutionary events could be a convergent evolution of the virus. In addition, one single nucleotide deletion impacted ORF3a at the position Δ26161, this frameshift mutation lead to a truncated protein missing 19 AA with a stop codon generated at the protein position 259.

The occurrence of specific mutations in the 309 genomes was studied and classified. In order to organize our dataset into meaningful groups and identify coherent patterns, a double hierarchical clustering of the detected mutations was generated (Figure 1). The hierarchical clustering revealed preponderant mutations shared by a majority of the isolates. It divided our 309 patients into five main clusters corresponding to specific mutational signatures. Regarding the clustering of mutational events, we detected the presence and co-occurrence of specific mutations in different clusters (Figure 1). Among the main clusters, we pointed the cluster 1 (44 patients, 14.2%, positions [28881-28882-28883]) with two nonsynonymous mutations in protein N (R203K; G204R). Cluster 2 (39 patients,

12.6%, position 15324) contains a synonymous mutation (C15324T). Cluster 3 (126, 100 and 211 patients, at positions 2416, 8371, 25563, respectively) includes one synonymous mutation (C2416T), and two nonsynonymous mutations (nsp3: Q1884H; ORF3a: Q57H). Cluster 4 (68 patients, 22%, position 1059) contains one nonsynonymous mutation (nsp2: T85I). Finally, cluster 5 (from 297 to 303 patients, 96-98%, positions 241, 3037, 14408, 23403) displays one mutation in 5'UTR (C241T), one synonymous mutation (C3037T) and two nonsynonymous mutations (nsp12b: P314L, S protein D614G).

The genomes of all 309 isolates were integrated in a global phylogenetic reconstruction including all other 200 genomes of SARS-CoV-2 available in France (Figure 2 + supplementary figure 1). We observed different clades corresponding to the previous mutational patterns revealed by the hierarchical clustering. These mutations were positioned in phylogenetic trees. Globally, the distribution of the 309 SARS-CoV-2 genomes from our patients was distributed across all lineages. Noteworthy, we observed a high density of isolates from our cohort in the cluster 3a corresponding to the mutation nsp3: G8371T, Q1884H. Phylogenetic reconstruction was also carried out including all complete genomes available worldwide (5519 genomes on 15th April 2020). All viral isolates of our French cohort were mainly distributed in clades including European isolates (Figure 3). Again, the mutation in the cluster 3a (nsp3: G8371T, Q1884H) regrouped almost exclusively the variants from our French cohort (98/105 genomes). Compared to the Chinese reference, we observed a prevalent distribution of specific mutations at positions 241, 3037, 14808 and 23403 that spread in approximately ~3570 (+/-7) genomes representing ~64.6 % of the isolates worldwide. These mutations have been reported in variants from all continents. Although, we detected some clades depending on the common geographical origin of the patients, a global distribution of variants in all continents was observed, therefore meaning and confirming that SARS-CoV-2 is circulating worldwide.

Clinical outcomes of the 309 COVID-19-infected patients were investigated and integrated in our evolutionary analyses (Figure 1-3). Among these patients, we classified the outcomes in three categories including poor clinical outcome (PClinO, defined by either death or transfer to intensive care unit or hospitalization for 10 days or more), poor virological outcome (PVirO, defined by viral

shedding persistence at day 10) and good clinical outcome (GO) (Supplementary Table 1). Coronavirus genomes isolates from 38 patients'isolates with PVirO were widely distributed across the groups, including diverse mutational events meaning that no correlation between higher viral loads and the genomic specificity of the virus was evidenced as previously reported (supplementary figure 2). For the 10 patients with PClinO, a majority of isolates were also distributed to all groups (supplementary figure 2). An exception concerned two patients (IHUCOVID-0318, IHUCOVID-0333) with one PClinO and one death that were clusterized together in a group of seven different isolates. A common nonsynonymous mutation in the membrane protein M (L129M, CTG>ATG, position 26907) was found in this cluster.

DISCUSSION

Systematic gene level mutational analysis of the genomes allowed us to identify several unique features of the SARS-CoV-2 genome. In this study, we reported 315 substitution mutations including 126 synonymous mutations. Although these later mutations conserved AA, it has been reported that the fitness of RNA viruses with genomes harboring synonymous mutations could be affected[11]. Regarding the 178 nonsynonymous mutations, alteration due to the replacement of specific AA depends on several factors including their physicochemical properties, steric hindrance and the position of the AA in the tridimensional structure of the resulting protein. A key element of coronavirus host range is determined by the binding affinity between the spike S protein and the cellular receptor[12]. Here, thirty substitution mutations (including 18 nonsynonymous) were reported in the spike protein S. The spike protein mediates viral entry into host cells by recognizing angiotensin-converting enzyme 2 (ACE2) as its receptor[13]. We detected a missense mutation P330S (CCT>TCT) located at the border of the RBD[14]. Furthermore, a deletion of six nucleotides Δ21765-21770 removed two AA in the resulting protein. Mutations in the spike surface glycoprotein (A930V (24351C>T)) have already been reported in the Indian SARS-CoV-2[15]. All these mutations in protein S could influence host range and transmissibility of the virus. Consequently, the spike S glycoprotein is a potential target among potential therapeutic strategies. Focusing on the mutations identified in our study enables to drive the strategy to target special positions of the protein. In this line, specific

antiviral therapies on S protein should be conducted *in vitro* (involving RBD-ACE2 blockers, S cleavage inhibitors, neutralizing antibodies, inhibitors and small interfering)[16]. Additionally, truncated proteins were also detected for ORF7a with a stop codon mutation (position 38 on 122 AA protein length) and for ORF3a with a frameshift mutation leading to an aborted protein missing 19 terminal AA. These mutations represent major changes in the protein architecture that have to be experimentally tested in order to assess the impact on protein function.

In evolutionary biology, convergent evolution is a central concept. The importance of a mutation is greatly strengthened if one can show that the fitness has evolved independently multiple times[17]. Two deletion events (Δ686-694; Δ1605-1607) were detected in three and two variants while the COVID-19-infected patients were not epidemiologically-related. These results strongly support a convergent evolution and could suggest convergent adaptation of the virus.

Phylogenetic trees were reconstructed by including isolates from France and all isolates worldwide. Isolates were grouped according to continents, but the large circulation of the virus was evidenced. Interestingly, we detected only two isolates belonging to the S-type in our 309 patients' cohort[18]. Noteworthy, an almost exclusive group including SARS-CoV-2 genomes from 98 patients monitored in IHU Mediterranée Infection was phylogenetically close and regrouped in a clade of 105 genomes. Classification based on mutation profiles enables the inclusion of patients in specific groups and correlates this clustering with the clinical features in order to favor a better approach to treatments. We reported widespread variants that extent to different groups and pointed five clustering groups that spread to different lineages. Regarding the clinical evolution of patients, PClinO and PVirO were widely distributed in all groups identified in our clustering analysis as well as in the phylogenetic trees as previously observed[19]. A single exception was reported in a subgroup with a missense mutation in the gene encoding membrane protein M (L129M, CTG>ATG) in which two PClinO (including on death) were observed. This specific mutation in membrane protein M (L129M) has to be carefully investigated in future large cohort as potential determinant of clinical evolution. But at this stage, more data have to be integrated to state any conclusion or consequence about this mutation. Overall, these

results suggested a clinical outcome that is not dependent on viral variants but rather a patient-dependent clinical outcome.

In conclusion, the mutational signatures and evolution of the coronavirus SARS-CoV-2 from 309 patients monitored in IHU Mediterranée Infection were unraveled. Inclusion of the French cohort allowed us to identify 161 novel mutations never reported before in coronavirus isolates worldwide. This mutations list represents a huge library of potential targets to test experimentally. The isolation and the genomic studies of the 309 SARS-CoV-2 genomes are paving the way for new diagnostic, therapeutic and prophylactic approaches.

METHODS

Patients

The study was conducted at Assistance Publique-Hôpitaux de Marseille (AP-HM), Southern France in the Institut Hospitalo-Universitaire (IHU) Méditerranée Infection (https://www.mediterranee-infection.com/). All data from patients were anonymized. The study was conducted on patients included from February 29th to April 4th. Individuals with PCR-documented SARS-CoV-2 RNA from a nasopharyngeal sample[20] were selected for genome sequencing. Viral loads ranged as follows: 93 patients with PCR Ct <16; 69 patients with [16<Ct<17]; 95 patients with [17<Ct<18]; 30 patients with [18<Ct<19]; 17 patients with [19<Ct<20] and 5 patients with Ct >20.

Genome sequencing and assembling

From nasopharyngeal samples, genomic RNA was extracted using the EZ1 biorobot with the EZ1 DNA tissue kit (Qiagen, Hilden, Germany) as previously described [21]. In the reverse transcription (RT) step, cDNA was reverse transcribed from total viral RNA samples using TaqMan Reverse Transcription (Life technologies applied biosystems) following the manufacturer's recommendations. PCR program was set as follows: 25°C, 10 min, 48°C, 30 min, 95°C, 5 min, 4°C. DNA Polymerase I Large Klenow Fragment (BioLabs) was used for generating double stranded cDNA. 20 µL of each sample were added to a mix (10 µL) of: Buffer 2 Neb 10X, DNA polymerase large Klenow Fragment,

dNTPs working solution (10mM) for 1 h at 37°C. cDNA were purified using beads Agencourt AMPure (Beckman Coulter) and then sequenced with the paired end strategy on a MiSeq sequencer (Illumina Inc, San Diego, CA,USA) with the Nextera XT DNA sample prep kit (Illumina). To prepare the paired-end library, the « tagmentation » step fragmented and tagged the DNA. Then, limited cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter Inc, Fullerton, CA, USA), libraries were normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in single 39-hours run in 2x250-bp as previously described [21,22]. After sequencing, reads were mapped on the reference SARS-CoV-2 isolate Wuhan-Hu-1 (MN908947.3) using CLC Genomics workbench v.7 with the following thresholds: 0.8 for coverage and 0.9 for similarity. For each sample, detailed per-base coverage information was extracted and all mutations with a coverage ≥5 were considered.

Hierarchical clustering

Data were ordered and grouped using the distance matrix computation and hierarchical clustering analysis calculated by MultiExperimentViewer[23]. A double clustering was carried out based on Pearson correlation as a distance matrix and complete linkage clustering as a linkage method.

Sequence dataset

All genomes available on April 15 of SARS-CoV-2 were downloaded from GISAID (Global Initiative; https://www.gisaid.org/) with acknowledgment [24,25]. All genomes were deposited at EMBL-EBI under the BioProject: PRJEB37722

ABBREVIATIONS

AA: Amino Acid

GO: good clinical outcome

Indel: Insertion and deletion

ORF: Open Reading Frame

PClinO: poor clinical outcome

PVirO: poor virological outcome

RBD: Receptor Binding Domain

SD: Standard Deviation

SNP: Single Nucleotide Polymorphism

UTR: Untranslated Transcribed Region

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the Ethical Committee of the University Hospital Institute Méditerranée Infection (N°: 2020-016). Informed consent was obtained from all patients. The study was performed in accordance to the good clinical practices recommended by the Declaration of Helsinki.

AVAILABILITY OF DATA AND MATERIALS

All genomes were deposited at EMBL-EBI under the BioProject: PRJEB37722

COMPETING INTERESTS

Authors declare no competing interests

FUNDING

This work was supported by a grant from the French State managed by the National Research Agency under the "Investissements d'avenir (Investments for the Future)" program with the reference ANR-10-IAHU-03 (Méditerranée Infection) and by the Région Provence-Alpes-Côte-d'Azur and the European funding FEDER PRIMI. This research was supported by a grant from the Institut Universitaire de France (IUF, Paris, France) allocated to Prof. Anthony Levasseur.

AUTHOR'S CONTRIBUTIONS

AL and DR supervised and conceived the project. AL designed the evolutionary studies and interpreted the data. AL and JD performed the experiments and analyzed the data. JCL, PEF, PC, AC, LB helped with data acquisition. AL wrote the manuscript. The authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Respectful thanks to all patients and their families enrolled in this study. The authors would like to thank all the engineers, technicians, and clinicians who contributed to this investigation. The authors gratefully acknowledge Olivia Ardizzoni, Vincent Bossi, Madeleine Carrara and Sofiane Regoui for help in sequencing the genomes. The authors acknowledge Yolande Obadia and Audrey Giraud-Gatineau for help in data acquisition. We gratefully thank the authors, the originating and submitting laboratories for their sequence and metadata shared through GISAID (Global Initiative; https://www.gisaid.org/). We acknowledge all the members of CoV-GLUE, Nextstrain (https://nextstrain.org/), and virological.org for sharing their analysis in real-time.

REFERENCES

1. Corman VM, Muth D, Niemeyer D, Drosten, C. 2018. Hosts and Sources of Endemic Human Coronaviruses. Adv. Virus Res. 100:163–188.

2. Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. 2020. Features, Evaluation and Treatment Coronavirus (COVID-19). StatPearls. Treasure Island (FL).

3. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in China. Nature. 579:265-269.

4. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE. 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol. 331:991-1004.

5. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 579:270-273.

6. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. 2020. The proximal origin of SARS-CoV-2. Nat Med. In press

7. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. 2019. Nat. Rev. Microbiol. **17**, 181–192.

8. Lam TS, Shum MH, Zhu HC, Tong YG, Ni XB, Liao YS, Wei W, Cheung WY, Li WJ, Li LF, Leung GM, Holmes EC, Hu YL, Guan Y. 2020. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature. In press

9. Smith EC, Blanc H, Vignuzzi M, Denison MR. 2013. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. PLoS Pathog. 9, e1003565

10. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom K, Zambon M, Ellis J, Lewis PA, Hiscox JA, Matthews DA. 2020. Characterisation of the transcriptome and proteome of SARS-CoV-2 using direct RNA sequencing and tandem mass spectrometry reveals evidence for a cell passage induced in-frame deletion in the spike glycoprotein that removes the furin-like cleavage site. bioRxiv (https://doi.org/10.1101/2020.03.22.002204)

11. Cuevas JM, Domingo-Calap P, Sanjuán R. 2012. The fitness effects of synonymous mutations in DNA and RNA viruses. Mol Biol Evol. 29:17-20.

12. Kuo L, Godeke GJ, Raamsman MJ, Masters PS, Rottier PJ. 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. J. Virol. 74**:**1393-1406.

13. Li F, Li W, Farzan M, Harrison SC. 2005. Structure of SARS coronavirus spike receptor binding domain complexed with receptor. Science. 309,1864–1868.

14. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L. 2020. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol. In press

15. Sardar R, Satish D, Birla S, Gupta D. 2020. Comparative analyses of SAR-CoV2 genomes from different geographical locations and other coronavirus family genomes reveals unique features potentially consequential to host-virus interaction and pathogenesis. bioRxiv, doi: https://doi.org/10.1101/2020.03.21.001586

16. Du L, He Y, Zhou Y, Liu S, Zheng B-J, Jiang S. 2009. The spike protein of SARS-CoVda target for vaccine and therapeutic development. Nat Rev Microbiol. 7:226e36.

17. Currie A. 2012. Convergence as evidence. Brit. J. Phil. Sci. 64,763–786.

18. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J, Lu J. 2020. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev. 3:nwaa036

19. Million M, Lagier JC, Gautret P, Colson P, Fournier PE, Amrane S, Hocquart M, Mailhe M, Esteves-Vieira V, Doudier B, Aubry C, Correard F, Giraud-Gatineau A, Roussel Y, Berenger C, Cassir N, Seng P, Zandotti C, Dhiver C, Ravaux I, Tomei C, Eldin C, Tissot-Dupont H, Honoré S, Stein A, Jacquier A, Deharo JC, Chabrière E, Levasseur A, Fenollar F, Rolain JM, Obadia Y, Brouqui P, Drancourt M, La Scola B, Parola P, Raoult D. 2020. Early treatment of 1061 COVID-19 patients with hydroxychloroquine and azithromycin, Marseille, France. Travel Med Infect Dis, In press

20. Amrane S, Tissot-Dupont H, Doudier B, Eldin C, Hocquart M, Mailhe M, DudouetP, Ormières E, Ailhaud L, Parola P, Lagier JC, Brouqui P, Zandotti C, Ninove L, Luciani L, Boschi C, La Scola B, Raoult D, Million M, Colson P, Gautret P. 2020. Rapid viral diagnosis and ambulatory management of suspected COVID-19 cases presenting at the infectious diseases referral hospital in Marseille, France, - January 31st to March 1st, 2020: A respiratory virus snapshot. Travel Med Infect Dis. 101632.

21. Colson P, Lagier JC, Baudoin JP, Bou Khalil J, 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. 1-3. doi:10.1007/s10096-020-03869-w

22. Tall ML, Ndongo S, Ngom II, Delerce J, Khelaifia S, Raoult D, Fournier PE, Levasseur A. *Massilimicrobiota timonensis* gen. nov., sp. nov., a new bacterium isolated from the human gut microbiota New Microbes New Infect. 2019 Sep; 31: 100574.

23. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 34: 374-378.

24. Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 10:33-46.

25. Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data – from vision to reality. EuroSurveillance. 22(13).

FIGURES, TABLES AND ADDITIONAL FILES

Table 1. Mutational events detected in the 309 SARS-CoV-2 genomes as compared to the reference Wuhan coronavirus.

Table 2. Detailed mutations detected in the 309 SARS-CoV-2 genomes from French patients.

Figure 1. Double hierarchical clustering of mutational events in SARS-CoV-2 genomes collected in 309 COVID-19 infected French patients. Top tree: Viral isolates from 309 patients. Left tree: Position of identified mutations according to the reference Wuhan genome. Clusters (1 to 5) were labeled with numbers ranging from 1 to 5. Missing positions were labeled in grey. The mutation type is represented by a colour scale from green (position identical to the reference), red (nonsynonymous mutation), orange (synonymous mutation) and blue (deleted position).

Figure 2. Phylogenetic tree of coronavirus in France. The phylogenetic tree was reconstructed from 509 viral genomes collected in France (309 from this study). Number represents the cluster

identified in the double hierarchical clustering. The reference Wuhan coronavirus isolate was underlined in purple. Red star corresponds to PClinO, blue star corresponds to PVirO, grey star corresponds to the loss of follow-up and green star corresponds to deceased patient. Main mutations were labeled along branches.

Figure 3. Phylogenetic tree of coronavirus worldwide. The phylogenetic tree was reconstructed from 5519 viral genomes collected around the world. Numbers (1 to 5) represent the cluster identified in the double hierarchical clustering. Red star corresponds to PClinO, blue star corresponds to PVirO, grey star corresponds to the loss of follow-up and green star corresponds to deceased patient.

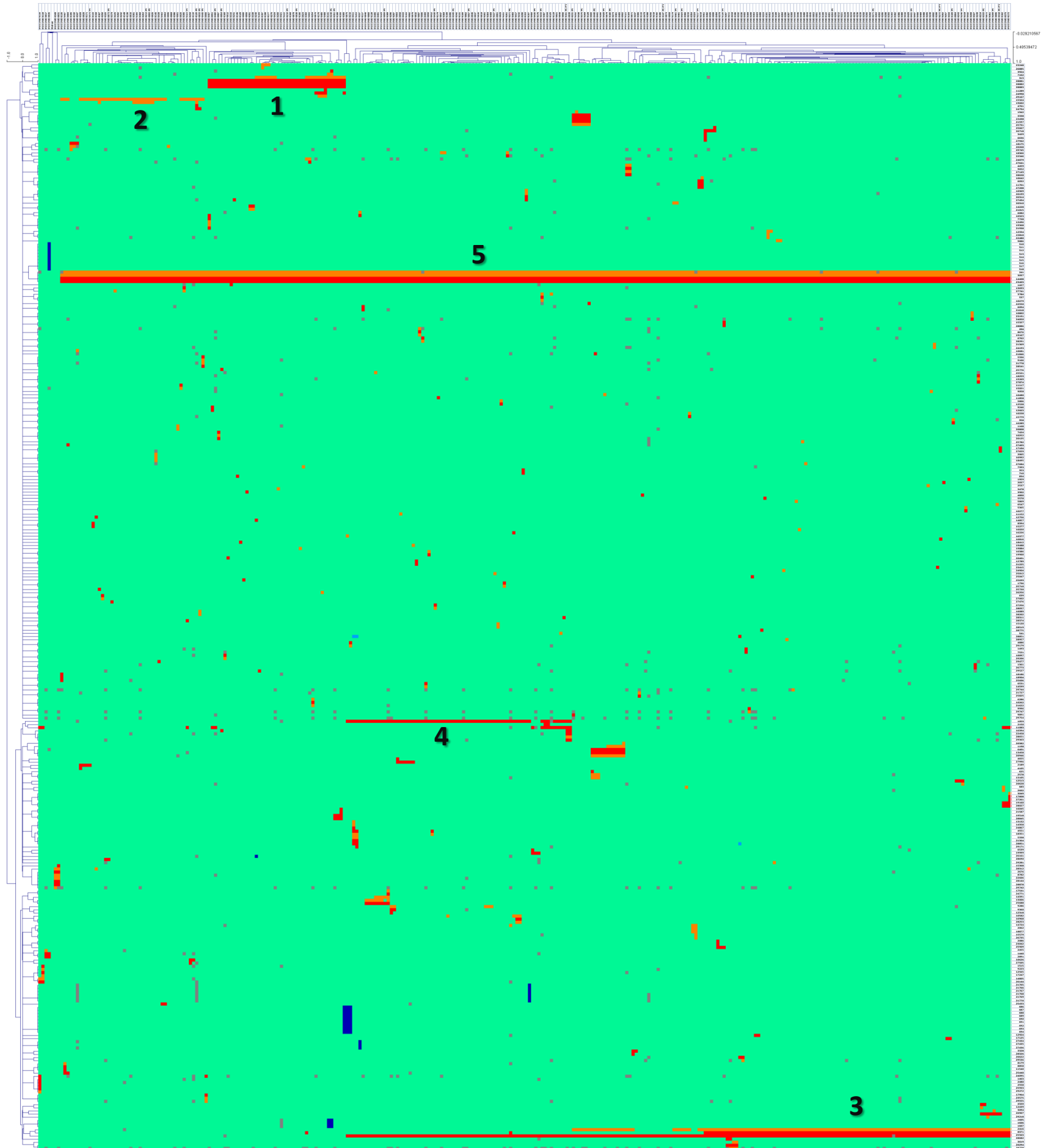
Supplementary Table 1. Baseline characteristics of the 309 COVID-19 infected patients.

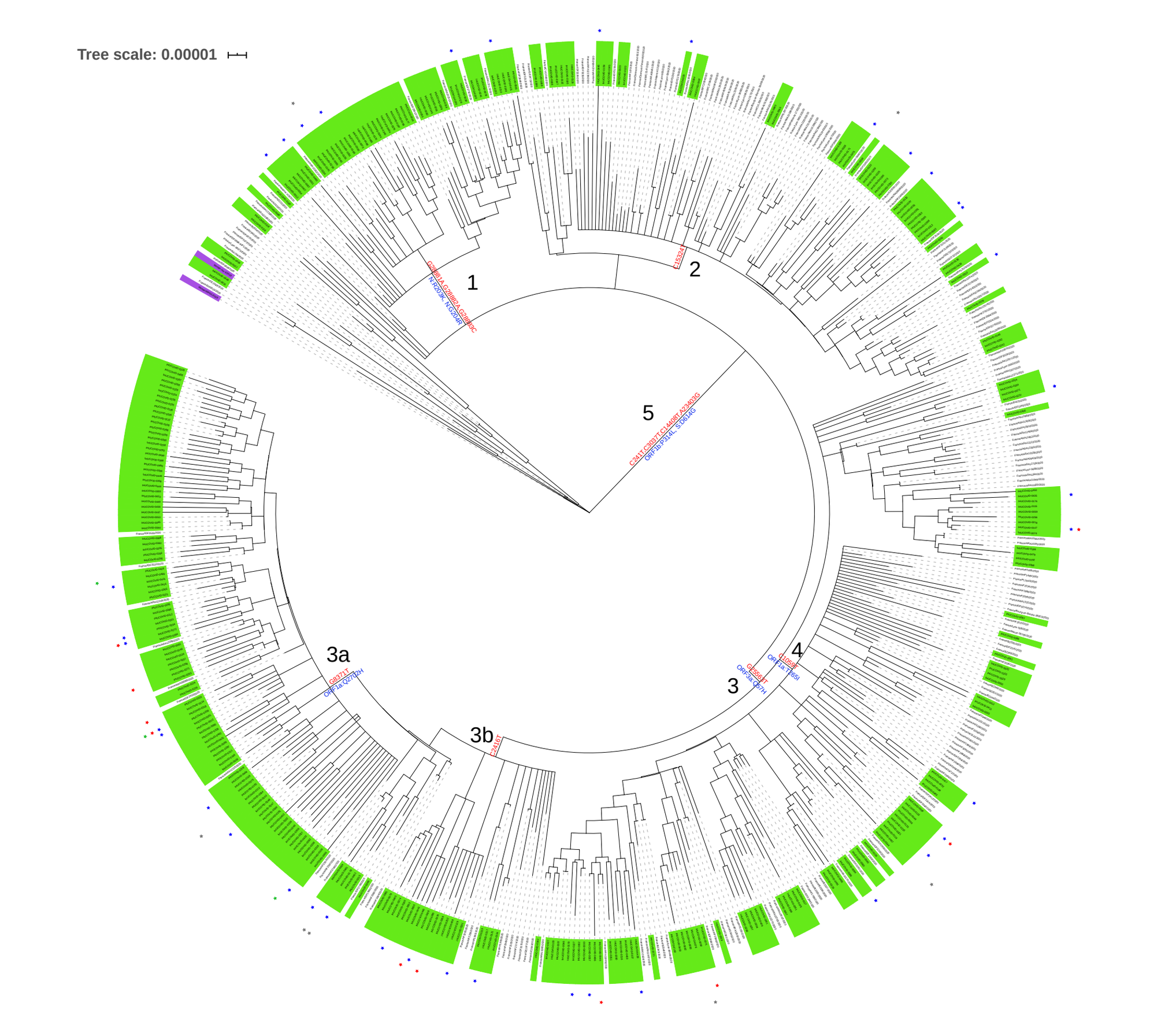
Supplementary Figure 1. Emergence and phylogenetic tree of coronavirus in France.

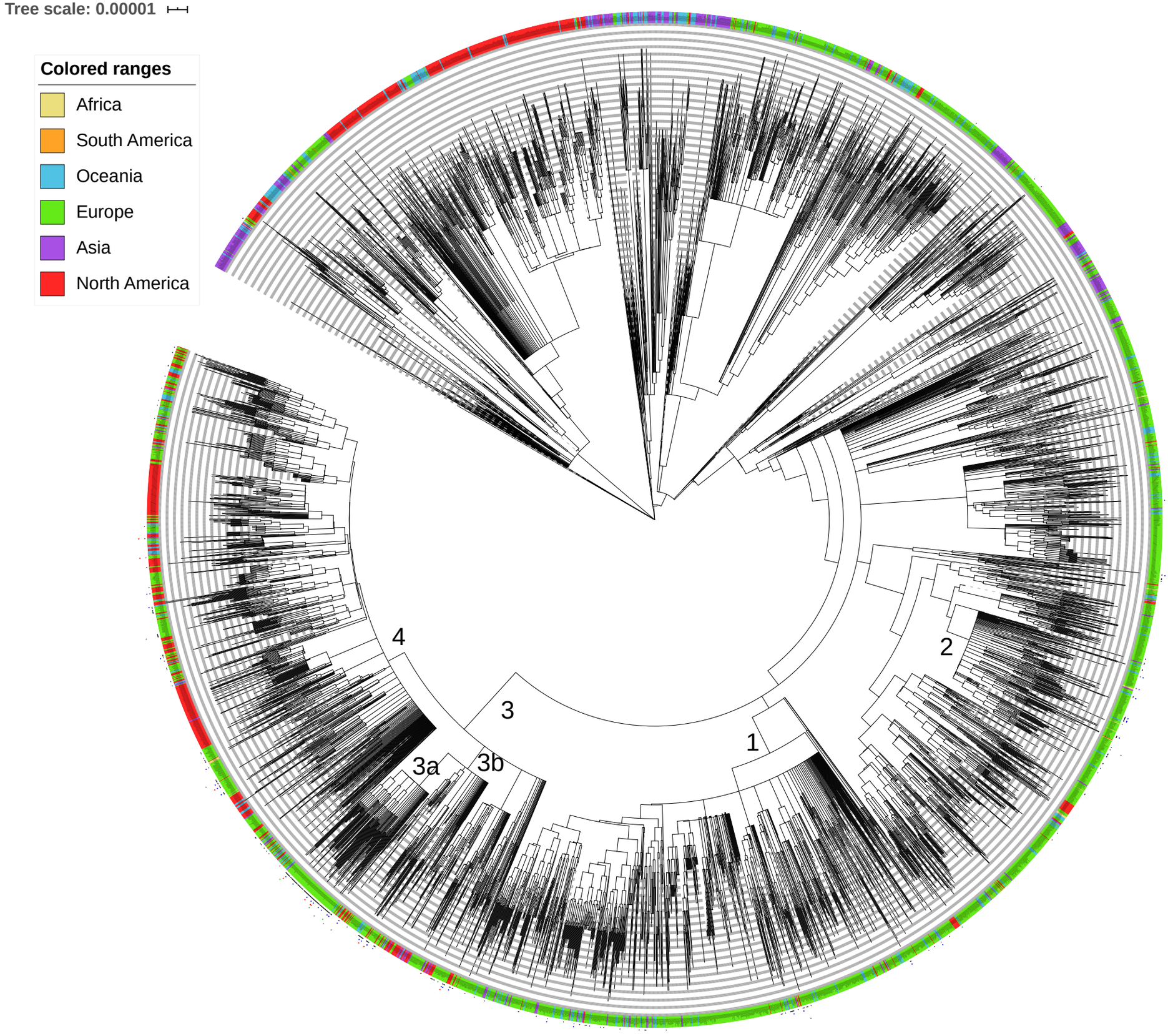
Supplementary Figure 2. Phylogenetic tree of coronavirus from the cohort of 309 COVID-19 infected patients. The phylogenetic tree was reconstructed from 309 viral genomes in this study. Numbers (1 to 5) represent the cluster identified in the double hierarchical clustering. The reference Wuhan coronavirus isolate was underlined in purple. Red star corresponds to PClinO, blue star corresponds to PVirO, grey star corresponds to the loss of follow-up and green star corresponds to deceased patient. Mutations were labeled along branches.

AUTHOR INFORMATION

Correspondence to Didier Raoult didier.raoult@gmail.com or Anthony Levasseur anthony.levasseur@univ-amu.fr


1
2
5
4
3


Tree scale: 0.00001
1
2
3
3a
3b
4
5
C21532T
G28881A,G28882A,G28883C
N:R203K, N:G204R
C241T,C3037T,C14408T,A23403G
ORF1b:P314L, S:D614G
G6371T
ORF1a:Q2718H
C24116T
G25563T
ORF3a:Q57H
C10545T
ORF1a:T265I


Tree scale: 0.00001
Colored ranges
Africa
South America
Oceania
Europe
Asia
North America
1
2
3
3a
3b
4

	Nb S	*Nb NS*	*Deletion event*
SARS-CoV-2 genome	126	178	6
nsp1	3	5	2
nsp2	11	17	1
nsp3	17	27	0
nsp4	6	7	0
nsp5	9	5	0
nsp6	1	5	0
nsp7	0	1	0
nsp8	3	5	0
nsp9	1	3	0
nsp10	2	0	0
nsp12a	0	1	0
nsp12b	12	11	0
nsp13	5	8	0
nsp14	9	7	0
nsp15	6	5	0
nsp16	4	2	0
S (spike protein)	12	18	1
ORF3a	4	13	1
E (envelope)	1	1	0
M (membrane)	4	4	0
ORF6	0	3	0
ORF7a	2	5	1
ORF7b	0	1	0
ORF8	3	4	0
N (nucleocapsid)	11	20	0

S: synonymous mutation; NS : nonsynonymous mutation

Table 1. Mutational events detected in the 309 SARS-CoV-2 genomes as compared to the reference Wuhan coronavirus.

Position genome	Region/ gene	Mutation nt	Mutation aa	Nb isolates
204	5UTR	G>T		1
241	5UTR	C>T		297
313	nsp1	CTC>CTT	L	20
353	nsp1	GGC>AGC	G30S	1
510	nsp1	GGT>X		1
511	nsp1	GGT>X		1
512	nsp1	CAT>X		1
513	nsp1	CAT>X		1
514	nsp1	CAT>X		1
515	nsp1	GTT>X		1
516	nsp1	GTT>X		1
517	nsp1	GTT>X		1
518	nsp1	ATG>X		1
537	nsp1	GAA>GGA	E91G	1
541	nsp1	CTC>CTT	L	1
625	nsp1	AAG>AAT	K120N	1
639	nsp1	AAG>AGG	K125R	1
683	nsp1	CTA>TTA	L	2
686	nsp1	AAG>X		3
687	nsp1	AAG>X		3
688	nsp1	AAG>X		3
689	nsp1	TCA>X		3
690	nsp1	TCA>X		3
691	nsp1	TCA>X		3
692	nsp1	TTT>X		3
693	nsp1	TTT>X		3
694	nsp1	TTT>X		3
710	nsp1	CTT>ATT	L149I	1
834	nsp2	TTC>TGC	F10C	1
958	nsp2	TGC>TGT	C	1
1059	nsp2	ACC>ATC	T85I	68
1168	nsp2	CGA>CGG	R	1
1198	nsp2	GAA>GAG	E	6
1361	nsp2	GTT>TTT	V186F	1
1427	nsp2	CAT>TAT	H208Y	1
1440	nsp2	GGC>GAC	G212D	2
1463	nsp2	GGT>AGT	G220S	1
1493	nsp2	GTG>ATG	V230M	1
1515	nsp2	CAT>CGT	H237R	1
1605	nsp2	AAT>X		2
1606	nsp2	AAT>X		2
1607	nsp2	GAC>X		2
1736	nsp2	GTG>ATG	V311M	1
1942	nsp2	GTG>GTT	V	1
1959	nsp2	GCC>GTC	A385V	1
2037	nsp2	GCT>GTT	A411V	1
2062	nsp2	GCC>GCT	A	2
2065	nsp2	TAC>TAT	Y	1
2086	nsp2	CAG>CAT	Q427H	1
2150	nsp2	GAT>AAT	D449N	2
2189	nsp2	CTT>TTT	L462F	4
2416	nsp2	TAC>TAT	Y	126
2455	nsp2	CTC>CTT	L	1
2480	nsp2	ATT>GTT	I559V	1
2537	nsp2	TTG>CTG	L	1
2558	nsp2	CCA>TCA	P585S	1
2578	nsp2	AGT>AGC	S	3
2602	nsp2	GTT>GTG	V	1
2676	nsp2	CCT>CTT	P624L	2
2891	nsp3	GCA>ACA	A58T	2
3037	nsp3	TTC>TTT	F	303
3049	nsp3	GAT>GAC	D	1
3094	nsp3	CCA>CCG	P	1
3328	nsp3	CAG>CAT	Q203H	2
3338	nsp3	GTG>TTG	V207L	6
3429	nsp3	ACA>ATA	T237I	1
3478	nsp3	GTT>GTC	V	1
3619	nsp3	GTC>GTT	V	2
3784	nsp3	GTC>GTT	V	2
3902	nsp3	CCA>TCA	P395S	1
3924	nsp3	CCT>CTT	P402L	1
4105	nsp3	AAG>AAT	K462N	1
4320	nsp3	GCC>GTC	A534V	2
4459	nsp3	ATA>ATC	I	2
4551	nsp3	ACA>ATA	T611I	3
4655	nsp3	CGG>TGG	R646W	1
4761	nsp3	GAA>GGA	E681G	1
4800	nsp3	AAA>AGA	K694R	1
4886	nsp3	CCT>TCT	P723S	1
5336	nsp3	AAT>GAT	N873D	1
5338	nsp3	AAT>AAC	N	2
5365	nsp3	TAC>TAT	Y	1
5570	nsp3	ATG>GTG	M951V	1
5806	nsp3	TGC>TGT	C	1
5869	nsp3	TAC>TAT	Y	1
6032	nsp3	GCA>ACA	A1105T	2
6082	nsp3	GAT>GAC	D	1
6094	nsp3	CAG>CAT	Q1125H	1
6286	nsp3	ACC>ACT	T	1
6401	nsp3	CCA>TCA	P1228S	11

6531	nsp3	GAG>GTG	E1271V	1
6539	nsp3	CAC>TAC	H1274Y	1
6547	nsp3	GAT>GAC	D	1
6636	nsp3	ACT>ATT	T1306I	1
6762	nsp3	ACT>ATT	T1348I	1
7162	nsp3	GAC>GAT	D	2
7393	nsp3	CCG>CCT	P	1
7424	nsp3	TTT>GTT	F1569V	1
7521	nsp3	ACA>ATA	T1601I	1
7728	nsp3	TCT>TTT	S1670F	1
8179	nsp3	CGG>CGT	R	1
8371	nsp3	CAG>CAT	Q1884H	100
8394	nsp3	GCT>GTT	A1892V	1
8676	nsp4	AGT>ATT	S41I	1
8782	nsp4	AGC>AGT	S	2
8858	nsp4	GTC>ATC	V102I	1
9058	nsp4	CCA>CCG	P	1
9166	nsp4	ACC>ACT	T	1
9223	nsp4	CAC>CAT	H	1
9246	nsp4	GCT>GTT	A231V	1
9286	nsp4	AAC>AAT	N	5
9360	nsp4	ACA>ATA	T269I	2
9652	nsp4	ATG>ATT	M366I	2
9886	nsp4	TAC>TAT	Y	2
9891	nsp4	GCT>GTT	A446V	1
9996	nsp4	TCA>TTA	S481L	1
10097	nsp5	GGT>AGT	G15S	1
10228	nsp5	CTC>CTT	L	1
10252	nsp5	TTC>TTT	F	1
10262	nsp5	GCT>TCT	A70S	1
10277	nsp5	CTC>TTC	L75F	1
10279	nsp5	CTC>CTT	L	1
10323	nsp5	AAG>AGG	K90R	2
10369	nsp5	CGC>CGT	R	1
10480	nsp5	AAT>AAC	N	1
10582	nsp5	GAC>GAT	D	4
10626	nsp5	GCA>GTA	A191V	2
10642	nsp5	ACG>ACT	T	1
10681	nsp5	TAC>TAT	Y	1
10802	nsp5	CTA>TTA	L	1
11083	nsp6	TTG>TTT	L37F	19
11117	nsp6	ATT>GTT	I49V	1
11152	nsp6	GTC>GTT	V	1
11289	nsp6	TCT>TTT	S106F	1
11540	nsp6	GTT>TTT	V190F	1
11761	nsp6	AAG>AAT	K263N	2
11858	nsp7	GTA>ATA	V6I	1
12318	nsp8	TCT>TAT	S76Y	1
12334	nsp8	GCA>GCT	A	2
12439	nsp8	CCC>CCT	P	1
12496	nsp8	TAT>TAC	Y	1
12513	nsp8	ACG>ATG	T141M	3
12528	nsp8	ACA>ATA	T146I	1
12565	nsp8	CAG>CAT	Q158H	1
12663	nsp8	GCC>GTC	A191V	1
12756	nsp9	ACT>ATT	T24I	1
12789	nsp9	ACA>ATA	T35I	1
12924	nsp9	CCT>CTT	P80L	2
13006	nsp9	GCT>GCC	A	7
13105	nsp10	TAC>TAT	Y	3
13132	nsp10	CAA>CAG	Q	1
13458	nsp12a	TCG>TTG	S6L	11
13693	nsp12b	ACA>TCA	T76S	1
14228	nsp12b	AAG>AGG	K254R	2
14358	nsp12b	TGC>TGT	C	1
14391	nsp12b	TTC>TTT	F	5
14408	nsp12b	CCT>CTT	P314L	303
14599	nsp12b	CTA>TTA	L	1
14724	nsp12b	TTC>TTT	F	3
14805	nsp12b	TAC>TAT	Y	2
14857	nsp12b	GTT>TTT	V464F	1
15021	nsp12b	CTT>CTC	L	1
15120	nsp12b	GTC>GTT	V	1
15277	nsp12b	CAC>TAC	H604Y	1
15324	nsp12b	AAC>AAT	N	39
15327	nsp12b	ATG>ATT	M620I	1
15368	nsp12b	ACA>ATA	T634I	1
15438	nsp12b	ATG>ATT	M657I	6
15579	nsp12b	AAC>AAT	N	1
15779	nsp12b	AAG>AGG	K771R	1
15810	nsp12b	AAC>AAT	N	1
16020	nsp12b	GTG>GTA	V	1
16045	nsp12b	CTT>TTT	L860F	1
16089	nsp12b	TTG>TTA	L	1
16226	nsp12b	ACA>ATA	T920I	1
16289	nsp13	GCT>GTT	A18V	1
16377	nsp13	CCG>CCT	P	1
16386	nsp13	TGC>TGT	C	1
16394	nsp13	CCA>CTA	P53L	3
16816	nsp13	CAA>AAA	Q194K	1
16968	nsp13	GAG>GAT	E244D	2
16992	nsp13	TTA>TTG	L	1
17104	nsp13	CAT>TAT	H290Y	1

17125	nsp13	CTC>TTC	L297F	2
17247	nsp13	CGT>CGC	R	1
17541	nsp13	GAC>GAT	D	1
17808	nsp13	AAG>AAT	K524N	1
17964	nsp13	ATG>ATA	M576I	1
18175	nsp14	CCT>TCT	P46S	3
18186	nsp14	ATG>ATT	M49I	1
18259	nsp14	ATC>GTC	I74V	1
18351	nsp14	AAT>AAC	N	3
18366	nsp14	CTA>CTG	L	3
18412	nsp14	GTT>TTT	V125F	1
18495	nsp14	CTT>CTC	L	1
18877	nsp14	CTA>TTA	L	2
18904	nsp14	CGT>TGT	R289C	1
19160	nsp14	TCT>TTT	S374F	3
19167	nsp14	AAA>AAG	K	1
19269	nsp14	AAC>AAT	N	1
19368	nsp14	CCA>CCG	P	1
19488	nsp14	GTC>GTT	V	1
19518	nsp14	TTG>TTT	L493F	3
19602	nsp14	AAC>AAT	N	9
19884	nsp15	TAC>TAT	Y	1
19900	nsp15	GCA>ACA	A94T	1
19999	nsp15	GTT>TTT	V127F	3
20125	nsp15	GGA>CGA	G169R	1
20199	nsp15	TAC>TAT	Y	1
20268	nsp15	TTA>TTG	L	3
20294	nsp15	TAT>TGT	Y225C	4
20310	nsp15	TAT>TAC	Y	2
20382	nsp15	CTA>CTT	L	1
20401	nsp15	TCA>CCA	S261P	1
20628	nsp15	GGC>GGT	G	1
20808	nsp16	CTG>CTA	L	1
20946	nsp16	GTC>GTT	V	11
21015	nsp16	CAT>CAC	H	2
21225	nsp16	TGG>TGT	W189C	1
21364	nsp16	CCA>TCA	P236S	2
21369	nsp16	ATT>ATC	I	1
21587	S	CCA>TCA	P9S	1
21597	S	TCT>TTT	S12F	6
21618	S	ACA>ATA	T19I	1
21622	S	ACC>ACT	T	1
21727	S	TTC>TTT	F	1
21765	S	ATA>X		1
21766	S	ATA>X		1
21767	S	CAT>X		1
21768	S	CAT>X		1
21769	S	CAT>X		1
21770	S	GTC>X		1
21778	S	GGG>GGA	G	1
21846	S	ACT>ATT	T95I	1
21938	S	GTT>TTT	V126F	1
22346	S	GCT>TCT	A262S	1
22458	S	ACA>ATA	T299I	2
22480	S	TTC>TTT	F	1
22550	S	CCT>TCT	P330S	1
22606	S	GCA>GCT	A	2
23042	S	TCA>GCA	S494A	1
23191	S	TTC>TTT	F	1
23248	S	TTC>TTT	F	3
23403	S	GAT>GGT	D614G	303
23415	S	ACA>ATA	T618I	1
23575	S	TGC>TGT	C	1
24050	S	GAT>CAT	D830H	1
24079	S	GAT>GAC	D	2
24095	S	GCT>TCT	A845S	2
24193	S	CTG>CTA	L	1
24771	S	GCA>GTA	A1070V	1
24794	S	GCT>TCT	A1078S	2
24867	S	TGG>TTG	W1102L	1
24904	S	ATC>ATT	I	1
24998	S	GAC>TAC	D1146Y	5
25012	S	GAG>GAA	E	1
25047	S	CCA>CTA	P1162L	1
25433	ORF3a	ACT>ATT	T14I	2
25440	ORF3a	AAG>AAT	K16N	2
25521	ORF3a	TTC>TTT	F	1
25563	ORF3a	CAG>CAT	Q57H	211
25606	ORF3a	GCA>TCA	A72S	1
25667	ORF3a	TCA>TTA	S92L	1
25688	ORF3a	GCT>GTT	A99V	8
25699	ORF3a	GCC>TCC	A103S	1
25710	ORF3a	CTC>CTT	L	1
25720	ORF3a	GCT>TCT	A110S	1
25731	ORF3a	TAC>TAT	Y	6
25776	ORF3a	TGG>TGT	W128C	1
25782	ORF3a	TGC>TGT	C	1
25825	ORF3a	TAT>CAT	Y145H	1
25909	ORF3a	GAT>TAT	D173Y	3
25933	ORF3a	GAA>AAA	E181K	1
26144	ORF3a	GGT>GTT	G251V	2
26161	ORF3a	AAT>X		1
26256	E	TTC>TTT	F	1

26314	E	GTG>TTG	V24L	1
26622	M	CTT>TTT	L34F	2
26718	M	GTG>TTG	V66L	4
26735	M	TAC>TAT	Y	1
26773	M	ATG>ACG	M84T	1
26885	M	AAC>AAT	N	1
26907	M	CTG>ATG	L129M	7
27002	M	GAC>GAT	D	1
27143	M	AAC>AAT	N	2
27208	ORF6	CAT>TAT	H3Y	2
27256	ORF6	ATG>TTG	M19L	1
27261	ORF6	AGG>AGT	R20S	1
27403	ORF7a	ATT>GTT	I4V	1
27434	ORF7a	ACT>X		1
27435	ORF7a	ACT>X		1
27436	ORF7a	TGT>X		1
27476	ORF7a	ACA>ATA	T28I	1
27494	ORF7a	CCT>CTT	P34L	2
27505	ORF7a	GGA>TGA	G38*	1
27641	ORF7a	TCA>TTA	S83L	1
27659	ORF7a	AGA>ATA	R89I	1
27684	ORF7a	TAC>TAT	Y	1
27741	ORF7a	CTC>CTA	L	2
27874	ORF7b	ACT>ATT	T40I	1
27942	ORF8	CAC>TAC	H17Y	1
27996	ORF8	GAC>TAC	D35Y	6
28027	ORF8	TGG>TTG	W45L	3
28037	ORF8	AGA>AGG	R	1
28144	ORF8	TTA>TCA	L84S	2
28202	ORF8	TCG>TCA	S	1
28253	ORF8	TTC>TTT	F	2
28291	N	CCC>CCT	P	1
28311	N	CCC>CTC	P13L	1
28312	N	CCC>CCT	P	4
28326	N	GGT>GTT	G18V	1
28374	N	GGG>GAG	G34E	1
28519	N	GAC>GAT	D	1
28541	N	GCT>ACT	A90T	1
28628	N	GCT>TCT	A119S	2
28651	N	AAC>AAT	N	2
28682	N	GGA>AGA	G137R	2
28690	N	TTG>TTT	L139F	2
28775	N	CCA>TCA	P168S	1
28845	N	CGC>CTC	R191L	3
28851	N	AGT>AAT	S193N	2
28851	N	AGT>ATT	S193I	1
28878	N	AGT>AAT	S202N	2
28881	N	AGG>AAA	R203K	44
28882	N	AGG>AAA	R203K	44
28883	N	GGA>CGA	G204R	44
28886	N	ACT>GCT	T205A	1
28957	N	AAC>AAT	N	1
29167	N	TAC>TAT	Y	3
29171	N	CAT>TAT	H300Y	1
29179	N	CCG>CCT	P	1
29206	N	GCT>GCA	A	1
29218	N	TTC>TTT	F	1
29272	N	TAC>TAT	Y	1
29281	N	GCC>GCT	A	1
29363	N	CCA>TCA	P364S	2
29477	N	GAT>TAT	D402Y	1
29527	N	CAG>CAT	Q418H	1
29536		T>C		1
29555		C>T		1
29742	3UTR	G>A		1
29744	3UTR	G>T		1
29745	3UTR	G>T		1
29747	3UTR	G>T		1
29754	3UTR	C>T		1
29779	3UTR	G>C		3

Table 2. Detailed mutations detected in the 309 SARS-CoV-2 genomes from French patients