

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

3

4 **Anthony LEVASSEUR<sup>1,2,3#</sup>, Jeremy DELERCE<sup>1,2</sup>, Aurelia CAPUTO<sup>1,2</sup>, Ludivine**  
5 **BRECHARD<sup>1,2</sup>, Philippe COLSON<sup>1,2</sup>, Jean-Christophe LAGIER<sup>1,2</sup>, Pierre-Edouard**  
6 **FOURNIER<sup>2,4</sup>, Didier RAOULT<sup>1,2#</sup>**

7

8 <sup>1</sup> Aix-Marseille Université (AMU), MEPHI (Microbes, Evolution, Phylogeny and Infections), IRD,  
9 APHM, Faculté de Médecine, Marseille, France

10 <sup>2</sup> IHU Méditerranée Infection, Marseille, France

11 <sup>3</sup> Institut Universitaire de France, Paris, France

12 <sup>4</sup> Aix-Marseille Université (AMU), IRD, APHM, SSA, VITROME, Marseille, France

13 # Corresponding authors: Didier Raoult: [didier.raoult@gmail.com](mailto:didier.raoult@gmail.com); Anthony Levasseur :  
14 [anthony.levasseur@univ-amu.fr](mailto:anthony.levasseur@univ-amu.fr)

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

16 doi: <https://doi.org/10.1101/2020.09.04.282616>

17

18

19

20

21

22

23 **ABSTRACT**

24 The novel coronavirus (SARS-CoV-2) causes pandemic of viral pneumonia. The evolution and  
25 mutational events of the SARS-CoV-2 genomes are critical for controlling virulence, transmissibility,  
26 infectivity, severity of symptoms and mortality associated to this infectious disease. We collected and  
27 investigated 309 SARS-CoV-2 genomes from patients infected in France. Detailed genome  
28 cartography of all mutational events (SNPs, indels) was reported and correlated to clinical features of  
29 patients. A comparative analysis between our 309 SARS-CoV-2 genomes from French patients and  
30 the reference Wuhan coronavirus genome revealed 315 substitution mutations and six deletion events:  
31 ten were in 5'/3' UTR, 178 were nonsynonymous, 126 were synonymous and one generated a stop  
32 codon. Six different deleted areas were also identified in nine viral variants. In particular, 30  
33 substitution mutations (18 nonsynonymous) and one deletion ( $\Delta$ 21765-21770) concerned the spike S  
34 glycoprotein. An average of 7.8 mutational events (+/- 1.7 SD) and a median of 8 (range, 7-9) were  
35 reported per viral isolate. Comparative analyses and clustering of specific mutational signatures in 309  
36 genomes disclose several divisions in groups and subgroups combining their geographical and  
37 phylogenetic origin. Clinical outcomes of the 309 COVID-19-infected patients were investigated  
38 according to the mutational signatures of viral variants. These findings highlight the genome dynamics  
39 of the coronavirus 2019-20 and shed light on the mutational landscape and evolution of this virus.  
40 Inclusion of the French cohort enabled us to identify 161 novel mutations never reported in SARS-  
41 CoV-2 genomes collected worldwide. These results support a global and continuing surveillance of the  
42 emerging variants of the coronavirus SARS-CoV-2.

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

45

46

47