1   **Article type: Research article**

2

3   **Full-length title:**

4   **Genome sequence analysis enabled deciphering the atypical evolution of COVID-19 epidemics in**

5   **Marseille, France**

6   **Short title: Evolution of SARS-CoV-2 genotypes**

7

8   **Author list: Pierre-Edouard FOURNIER[1,2], Philippe COLSON[1,3], Anthony LEVASSEUR[1,3],**

9   **Philippe GAUTRET[1,2], Marielle BEDOTTO[1,3], Véronique FILOSA[1], Jeremy DELERCE[1,3],**

10   **Ludivine BRECHARD[1,2], Jean-Christophe LAGIER[1,3], Hervé CHAUDET[1,4], Florence**

11   **FENOLLAR[1,2], Didier RAOULT[1,3]***

12

13   **Affiliations:** [1] IHU Méditerranée Infection; [2] Vecteurs - Infections Tropicales et Méditerranéennes

14   (VITROME), Marseille, France ; [3] Aix-Marseille Univ, Microbes Evolution Phylogeny and Infections

15   (MEPHI); [4] French Armed Forces Center for Epidemiology and Public Health (CESPA).

16   [¥] These authors contributed equally to this article

17   *** Address for correspondence:** Didier Raoult, IHU - Méditerranée Infection, 19-21 boulevard Jean

18   Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email:

19   didier.raoult@gmail.com

20

21   **Key words:** SARS-CoV-2; Covid-19; genome; variant; mutations; deletions; diversity

22   **Supplementary Figures:** 1

23

24 **Supplementary Figure 1. Distribution along the viral genome and in viral genes of mutations observed for the various viral genotypes.**

25 Sequences from complete genomes were analyzed using the Nextstrain web-tool (https://clades.nextstrain.org/) [1]. They were compared to

26 sequences available in the GISAID database (https://www.gisaid.org/). Representation is adapted from Nextclade sequence analysis web

27 application output (https://clades.nextstrain.org/results).

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

29 Nextstrain clade 20A.EU1; [d] Named Nextstrain clade 20A.EU2.

30 3CL: 3C-like proteinase; E: Envelope protein; H: NTPase/helicase domain; M: Membrane glycoprotein; N: nucleocapsid phosphoprotein; nsp9:

31 ssRNA-binding protein; nsp14: 3'-to-5' exonuclease; nsp15: EndoRNase; PLpro: Predicted phosphoesterase; RdRp: RNA-dependent RNA

32 polymerase; S: Spike glycoprotein; Syn.: synonymous.

33

34
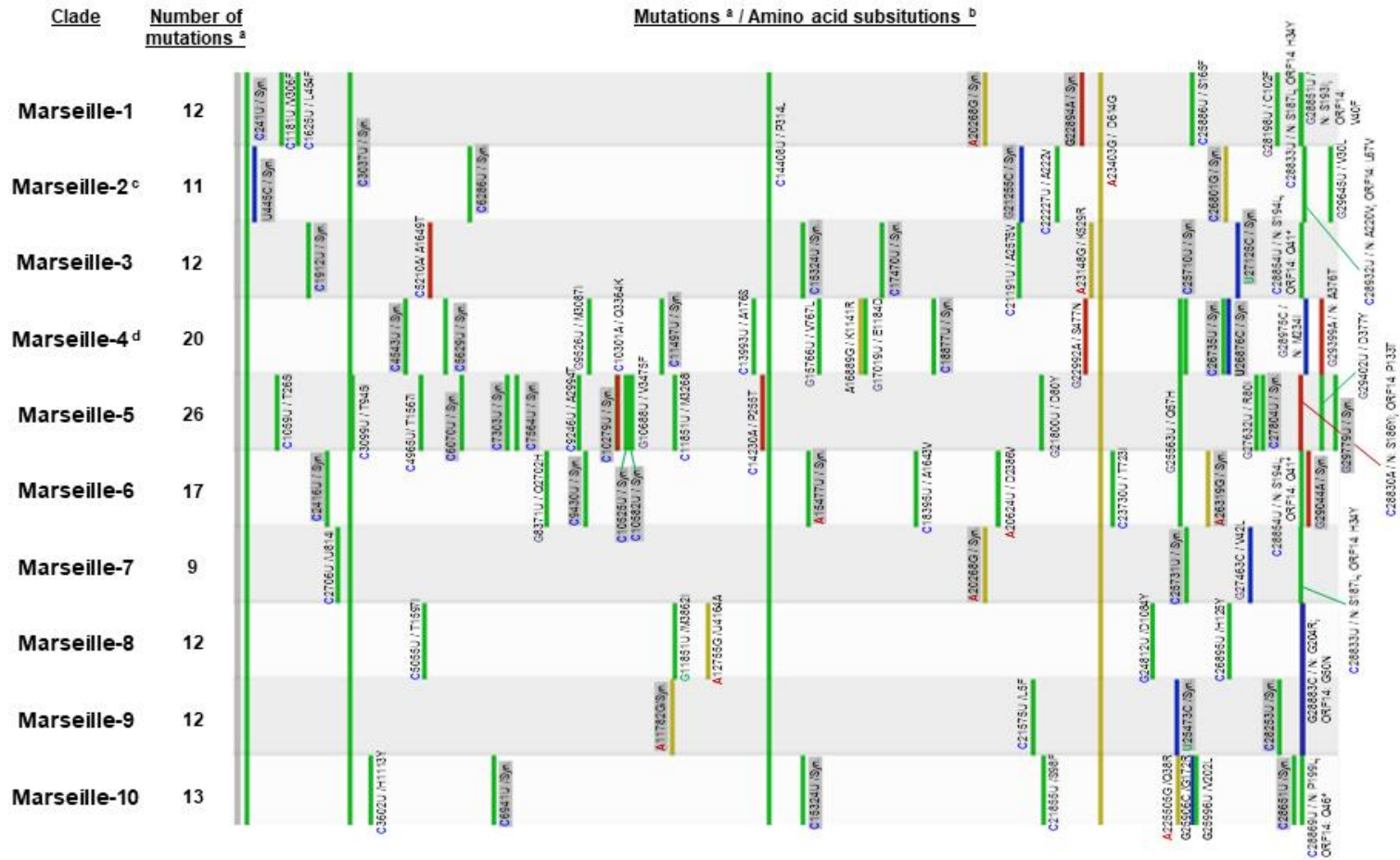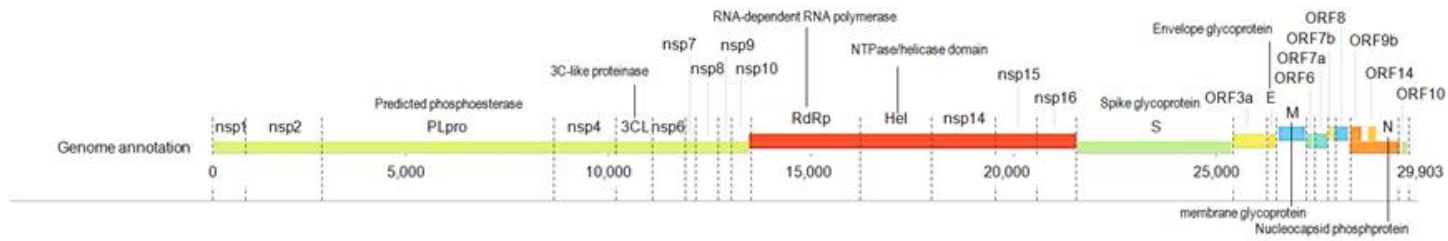
References

(1)   Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics **2018**; 34:4121-3.