Article type: Research article

Full-length title:
Genome sequence analysis enabled deciphering the atypical evolution of COVID-19 epidemics in Marseille, France

Short title: Evolution of SARS-CoV-2 genotypes

Author list: Pierre-Edouard FOURNIER\textsuperscript{1,2}, Philippe COLSON\textsuperscript{1,3}, Anthony LEVASSEUR\textsuperscript{1,3}, Philippe GAUTRET\textsuperscript{1,2}, Marielle BEDOTTO\textsuperscript{1,3}, Véronique FILOSA\textsuperscript{1}, Jeremy DELERCE\textsuperscript{1,3}, Ludivine BRECHARD\textsuperscript{1,2}, Jean-Christophe LAGIER\textsuperscript{1,3}, Hervé CHAUDET\textsuperscript{1,4}, Florence FENOLLAR\textsuperscript{1,2}, Didier RAOULT\textsuperscript{1,3}*

Affiliations: \textsuperscript{1}IHU Méditerranée Infection; \textsuperscript{2}Vecteurs - Infections Tropicales et Méditerranéennes (VITROME), Marseille, France; \textsuperscript{3}Aix-Marseille Univ, Microbes Evolution Phylogeny and Infections (MEPHI); \textsuperscript{4}French Armed Forces Center for Epidemiology and Public Health (CESPA).

* 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; mutations; deletions; diversity

Supplementary Figures: 1
Supplementary Figure 1. Distribution along the viral genome and in viral genes of mutations observed for the various viral genotypes. Sequences from complete genomes were analyzed using the Nextstrain web-tool (https://clades.nextstrain.org/) [1]. 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; c Named Nextstrain clade 20A.EU1; d Named Nextstrain clade 20A.EU2.

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.
References