

Severe acute respiratory syndrome coronavirus 2: virus mutations in specific European populations

F. Coppée¹, J. R. Lechien², A.-E. Declèves¹, L. Tafforeau³ and S. Saussez²

1) Department of Metabolic and Molecular Biochemistry, 2) Department of Human Anatomy and Experimental Oncology and 3) Department of Cell Biology, Research Institute for Health Sciences and Technology, University of Mons, Belgium

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is being intensively studied, particularly its evolution, in the increasingly available sequences between countries/continents with classical phylogenetic tree representation. More recently, certain protein mutations have been correlated with specific functional impacts. Our clinical data from patients suggest that clinical symptoms differ between European countries. Among other factors, SARS-CoV-2 mutations could explain these disparities. Our analyses point to an association of diverse mutations, including co-evolving ones, in a few SARS-CoV-2 proteins within specific countries. We therefore suggest combining clinical information from patients and the determination of the associated SARS-CoV-2 genome to better understand the specific symptoms.

© 2020 The Authors. Published by Elsevier Ltd.

Keywords: Amino acid replacements, co-evolved mutations, coronavirus disease 2019, european countries, G- S- and V-clades

Original Submission: 3 May 2020; **Revised Submission:** 7 May 2020; **Accepted:** 9 May 2020

Article published online: 20 May 2020

Corresponding author: F. Coppée, 6 avenue du champ de Mars, 7000 Mons, Belgium.

E-mail: frederique.coppee@umons.ac.be

Introduction

Our research group and co-investigators analysed the epidemiological and clinical data of 1420 European individuals with mild-to-moderate coronavirus disease 2019 (COVID-19) [1]. The included individuals with similar inclusion criteria came from Spain (30%), Italy (10%) and French-speaking populations from different European countries (Switzerland, Belgium and France; 60%). Interestingly, Bayesian analyses have reported that, depending on the countries, clinical symptoms were clearly identified, which was confirmed in two recent letters [2,3]. Headache accounted for 72.3% and 75.7% of French-speaking and Spanish individuals, respectively, whereas only 40.4% reported headache in the Italian population. Similarly, 67.8% of French-speaking and 72.4% of Spanish individuals presented with nasal obstruction, against only 53.7% of Italian individuals. Loss of smell was significantly more frequent in Spanish (70.5%) and French-speaking (73.3%) populations compared with the Italian population (50.0%). However, the prevalence of cough did not vary with country. Variations in the occurrence of these COVID-19 clinical symptoms among countries might be the result of virus mutations, angiotensin-converting enzyme-2 polymorphisms and other factors [1]. In this paper, we will focus on the potential virus mutation hypothesis through the GISAID database [4] (<https://www.gisaid.org>).

Methods

We analysed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) sequences using CoV_GLUE [5] (<http://cov-glue.cvr.gla.ac.uk/>, last accessed 17 April 2020: 9028 available sequences ('low coverage' excluded), including 4973 European sequences, reporting 2334 non-synonymous mutations).

Results

The two main mutations (S-D614G and nsp12-P323L) that diverge from the SARS-CoV-2 NCBI Reference Sequence (NC_045512) are retrieved in all continents with, as expected, only three cases in Asia. The first mutation, D614G in the S protein (found in 2342 samples), determining the virus clade 'G', frequently co-evolves with the P323L mutation in the nsp12 protein (found in 2318 samples). Indeed, in Spain, France, Italy and Switzerland both mutations were reported, respectively, in 42 sequences out of 145 available, in 173 out of 205 sequences,

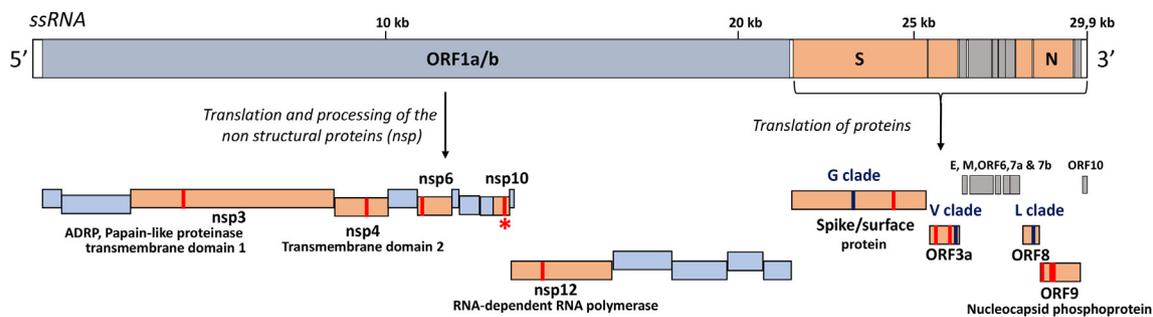


FIG. 1. Schematic of the viral genome and encoded proteins (at scale). (Upper panel) single-strand RNA (ssRNA) genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and position of the open reading frames (ORF) based on the reference sequence (NC_045512). (Bottom panel) All the encoded non-structural (processed after translation of two polyproteins, ORF1a and ORF1b, generated by a -1 ribosome frameshift near the 3' end of ORF1a) and structural proteins. Non-synonymous mutations reported in the text are indicated in red (with an asterisk pointing to a nonsense mutation and a larger band corresponding to close mutations) or in blue (which refers to a specific virus clade). The represented mutated proteins are indicated with their associated function or localization. It is known that in SARS-CoV, the proteins nsp3, nsp4 and nsp6, through their transmembrane domains, are involved in the replicative and transcription complex [13]; and that nsp10 is a critical co-factor for activation of multiple replicative enzymes [14–16].

in 30 out of 44 sequences and in 47 out of 48 sequences. In Belgium, 251 D614G and 231 P323L variants were found out of 342 sequences.

The ORF8-L84S (third most frequent mutation), that determines the virus clade 'S', appeared in 740 sequences but was only reported in 71 European cases, including 44 samples from Spain (collected from 1 to 12 March 2020). The remaining 27 sequences were reported in Iceland (12, including 11 persons who travelled from the USA), the UK (5), the Netherlands (4), France (2), Belgium (1), Portugal (1), Germany (1) and Greece (1). However, this mutation was not reported in Italy and Switzerland, nor in other European countries. In addition, we show that L84S amino acid substitution is co-evolving with three other mutations: nsp4-F308Y, ORF3a-G196V and N-S197L. These mutations are less reported, only 60/61 times. In Spain, 37 sequences showed these four combined mutations. In other European countries, all of these mutations were also found in the above-mentioned samples from: France (2 Grand Est/Strasbourg), Greece (1), Portugal (1), the Netherlands (1) and the UK (2). These four combined mutations were also retrieved in seven Australian patients, in four Chilean patients who travelled from Europe, in one Brazilian patient (returning from Madrid, reported as family cluster), and in one patient from each of the USA, Senegal and Georgia. Moreover, in one of the UK patients and in all seven Australian patients, in addition to the S197L mutation, the N protein presents the P13L replacement.

The fourth frequent mutation (ORF3a-Q57H found in 734 sequences) was reported in 101 samples from France and in 32 from Belgium but was not found in samples from Italy and Spain.

The fifth and sixth mutations (N-R203K and N-G204R) were found in four Spanish, ten Italian, ten Belgian and six French samples.

Both the seventh and eighth mutations (nsp6-L37F and ORF3a-G251V, the latest corresponding to the 'V' clade extensively associated with UK patient sequences) were found in five samples from Spain and two from Italy; and each was, respectively, found in 11 and 18 sequences from Belgian patients and in six and seven from French patient sequences.

At least three specific mutations were only reported in Belgian patients: nsp10-Y126* (51 samples), S-S943P (22 samples) and N-F171C (four samples). In addition, the nsp3-A534V variant was only reported in 11 samples from Belgium and ten from Luxembourg.

All the above-mentioned mutations are positioned on the SARS-CoV-2 genome in Fig. 1.

Discussion

Even though our sequence analysis is not exhaustive, it points to an association between diverse mutations, including co-evolving ones, in specific SARS-CoV-2 proteins and specific countries. All of the above discussed mutations (except the N-R203K) present modifications in their physicochemical properties (hydropathy, volume, chemical, charge, hydrogen acceptor/donor atoms or polarity classes). Particularly, the L84S substitution in the ORF8 protein, which is mainly found in Spain for European countries, is the one showing the most physicochemical changes. These mutations could therefore affect the function of the corresponding proteins: the non-

structural proteins (nsp4, nsp6, nsp12, ORF3a, ORF8) as well as the spike S protein and the nucleocapsid N protein. It was reported that S, ORF8 and ORF3a proteins are significantly different from those of other known SARS-like coronaviruses and could be linked to changes in pathogenicity/transmission [6]. A recent study proposed that these proteins are involved in the inhibition of the haem anabolic pathway and could be linked to a wide range of infections and diseases [7]. Moreover, the S Clade mutation in ORF8 is frequently associated with mutations in ORF3a, nsp4 and the N proteins. A recent sequence analysis in an Italian population also reported co-evolved mutations (the above reported two main mutations) and other synonymous mutations in nsp1 and nsp3 [8].

Previous studies suggest that the sequence diversity in SARS-CoV-2 proteins would be associated with the virus pathogenicity/transmission [9–12]. In Europe, we observe that several countries are associated with specific or several virus clades or mutations and that the frequency in clinical symptoms of individuals with COVID-19 also varies between these countries.

To conclude, our study suggests that patient clinical information (sequence polymorphisms and symptoms) and the sequence determination of the associated infectious genome should be combined to give better understanding of SARS-CoV-2 pathogenicity and help in the development of adapted treatments.

Acknowledgements

The sequences from the GISAID database were made available by the laboratories reported in the Supplementary material (Table S1). We used the public CoV_Glue resource, developed and maintained by the MRC-University of Glasgow Centre for Virus Research, and the Eishor1A COVID-19 library developed by KUL (Belgium) for some contextual research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nmni.2020.100696>.

Author contributions

FC and JL contributed to conceptualization, investigation, formal analysis and to the writing and original draft preparation.

A-ED and LT contributed to the writing, review and editing; and SS contributed to conceptualization, writing, review and editing.

Declaration of interest

The authors declare no conflicts of interest.

Funding

No funding was used to conduct this research.

References

- [1] Lechien JR, Chiesa-Estomba CM, Place S, et al. Clinical and epidemiological characteristics of 1,420 European patients with mild-to-moderate coronavirus disease 2019. *J Int Med* 2020. <https://doi.org/10.1111/JOIM.13089>.
- [2] Vaira LA, Salzano G, Deiana G, De Riu G. Anosmia and ageusia: common findings in COVID-19 patients. *Laryngoscope* 2020. <https://doi.org/10.1002/lary.28692>.
- [3] Villalba NL, Maouche Y, Ortiz MBA, Sosa ZC, Chahbazian JB, Syrovatkova A, et al. Anosmia and dysgeusia in the absence of other respiratory diseases: should COVID-19 infection be considered? *Eur J Case Rep Intern Med* 2020;7:001641.
- [4] Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. *EuroSurveillance* 2017;22(13). <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.
- [5] Singer JB, Thomson EC, McLauchlan J, Hughes J, Gifford RJ. GLUE: a flexible software system for virus sequence data. *BMC Bioinform* 2018;19(1):532.
- [6] Chan JF, Kok KH, Zhu Z, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* 2020;9:221–36.
- [7] Liu W, Li H. COVID-19: attacks the I- β chain of hemoglobin and captures the porphyrin to inhibit human heme metabolism. *ChemRxiv* 2020. <https://doi.org/10.26434/chemrxiv.11938173.v5>. xiv.
- [8] Lorusso A, Calistri P, Mercante MT, et al. A “One-Health” approach for diagnosis and molecular characterization of SARS-CoV-2 in Italy. *One Health* 2020. <https://doi.org/10.1016/j.onehlt.2020.100135>. preproof.
- [9] Benvenuto D, Giovanetti M, Ciccozzi A, et al. The 2019-new coronavirus epidemic: evidence for virus evolution. *J Med Virol* 2020;92(4).
- [10] Benvenuto D, Angeletti S, Giovanetti M, et al. Evolutionary analysis of SARS-CoV-2: how mutation of non-structural protein 6 (NSP6) could affect viral autophagy. *J Infect* 2020. pii: S0163-4453(20)30186-30189.
- [11] Brufsky A. Distinct viral clades of SARS-CoV-2: implications for modeling of viral spread. *J Med Virol* 2020. <https://doi.org/10.1002/jmv.25902>.
- [12] Rubino S, Kelvin N, Bermejo-Martin JF, Kelvin D. As COVID-19 cases, deaths and fatality rates surge in Italy, underlying causes require investigation. *J Infect Dev Ctries* 2020;14:265–7.

- [13] Knoops K, Kikkert M, van den Worm SHE, et al. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol* 2008;6:e226.
- [14] Bouvet M, Imbert I, Subissi L, Gluais L, Canard B, Decroly E. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc Natl Acad Sci USA* 2012;109:9372–7.
- [15] Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, et al. Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J Biol Chem* 2014;289(37):25783–96.
- [16] Sevajol M, Subissi L, Decroly E, Canard B, Imbert I. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. *Virus Res* 2014;194:90–9.