Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Research paper

SARS-CoV-2 variants lacking ORF8 occurred in farmed mink and pangolin

Filipe Pereira a, b, *

a Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas, Coimbra 3000-456, Portugal
b IDENTIFICA genetic testing, Rua Simão Bolívar 259 3° Diretor Tras, Maia 4470-214, Portugal

ARTICLE INFO

Keywords:
COVID-19
Coronaviruses
Transmissibility
Nonsense mutations

ABSTRACT

The SARS-CoV-2 Variant of Concern 202012/01 (VOC-202012/01) is rapidly spreading worldwide owing to its substantial transmission advantage. The variant has changes in critical sites of the spike protein with potential biological significance. Moreover, VOC-202012/01 has a mutation that inactivates the ORF8 protein, whose absence can change the clinical features of the infection. Why VOC-202012/01 is more transmissible remains unclear, but spike mutations and ORF8 inactivation stand out by their known phenotypic effects. Here I show that variants combining relevant spike mutations and the absence of ORF8 occurred in SARS-CoV-2 and related viruses circulating in other host species. A truncated ORF8 (Q23stop) occurred in a SARS-CoV-2-related virus from a pangolin seized in China in 2017, also with several mutations in critical spike sites. Strikingly, I found that variants without ORF8 (E19stop) and with the N501T spike mutation circulated in farmed mink and humans from Denmark. Although with differences to VOC-202012/01, the identification of these variants highlights the danger of having reservoirs of SARS-CoV-2 and related viruses where more transmissible variants may occur and spill over to humans.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the cause of coronavirus disease 2019 (COVID-19), has spread rapidly throughout the world following its emergence in Wuhan, China (Zhou et al., 2020). New sequence variants constantly arise during virus replication, some of which increase in frequency locally or worldwide. The spread of new variants over the course of an outbreak most often results from chance events, like when a limited number of individual viruses establish a new population during transmission (Grubaugh et al., 2020). However, the increase in frequency can result from a competitive advantage due to differences in viral infectivity, transmissibility or reactivity with neutralizing antibodies (Duffy et al., 2008).

The SARS-CoV-2 Variant of Concern 202012/01 (VOC-202012/01), lineage B.1.1.7, is the first for which there is compelling evidence that it has a substantial transmission advantage (Chand et al., 2020; Davies et al., 2020; Rambaut et al., 2020). Epidemiological evidence points to an estimated difference in reproduction numbers between VOC and non-VOC ranging between 0.4 and 0.7 (Volk et al., 2021). VOC-202012/01 originated in the UK in late Summer to early Autumn 2020 and rapidly spread worldwide. The variant has several mutations changing important sites in the interaction between the virus spike protein and the host angiotensin I converting enzyme 2 (ACE2). Moreover, VOC-202012/01 has a mutation (C27972T; Q27stop) that truncates the ORF8 protein or renders it inactive (Chand et al., 2020). It has been previously shown that SARS-CoV-2 can persist without a functional ORF8 protein (Gong et al., 2020; Pereira, 2020, 2021; Su et al., 2020; Young et al., 2020). Patients infected with a SARS-CoV-2 variant lacking ORF8 had better clinical outcomes than those infected with wild-type variants, showing a longer duration of symptoms and a possible later onset (Young et al., 2020).

SARS-CoV-2 is thought to have been transmitted to humans from bats, although it is possible that other mammalian species may have acted as “intermediate” hosts in the transmission (Andersen et al., 2020; Zhang and Holmes, 2020). The Malayan pangolins (Manis javanica) are likely candidates for intermediates by the discovery of SARS-CoV-2-related viruses in several animals illegally imported into the Chinese provinces of Guangdong and Guangxi (Han, 2020; Lam et al., 2020). Several species can potentially be infected by SARS-CoV-2 via their ACE2 proteins (Damas et al., 2020). During the COVID-19 pandemic, SARS-CoV-2 has been transmitted from humans to pets and other wild and domestic animals (Oreshkova et al., 2020; Sailleau et al., 2020; Sit...
et al., 2020). At least in the case of farmed mink, viruses had spilled back to humans (Munnink et al., 2021). The circulation of SARS-CoV-2 in other species increases the emergence of new variants, some of which may have significant differences to the ancestral state resulting from the adaptation to the new host. Therefore, the transmission back to humans may introduce variants with higher transmissibility or with the potential to escape the protection resulting from existing vaccines. In this regard, several mutations have been detected in the spike protein during the passage of SARS-CoV-2 through mink populations (Koopmans, 2021; Munnink et al., 2021). It is therefore important to identify variants with modifications of biological relevance, such as in the spike or accessory genes, which may have circulated in domestic or wild animals. Those variants may pose a risk for the ongoing COVID-19 pandemic. Here I show that SARS-CoV-2 variants without the ORF8 occurred in both pangolin and farmed mink, always in the background of mutations in spike sites of biological relevance. These results suggest that mutations as those observed in VOC-202012/01 may occur in other host species, introducing an additional risk to the ongoing pandemic.

2. Materials and methods

2.1. Genomic sequences from SARS-CoV-2 and related viruses

Genomic sequences from SARS-CoV-2 and related viruses were obtained from the GISAID Initiative (https://www.gisaid.org/) on January 11, 2021. I searched for ORF8 nonsense mutation is all available sequences from Canis lupus familiaris, Chlorocebus sabaensis, Felis catus, Manis javanica, Manis pentadactyla, Mus musculus, Mustela lutreola, Neovison vison, Panthera leo, Panthera tigris jacksoni, Rhinolophus affinis, Rhinolophus malayanus and Rhinolophus shameli. The sequences from pangolin (M. javanica) and mink (N. vison) with ORF8 nonsense mutations were classified into clades using the Nextclade tool (https://clades.nextstrain.org/) and the Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN) tool running in GISAID (Rambaut et al., 2020a).

2.2. SARS-CoV-2 related virus from pangolin (M. javanica) sample P1E

The raw sequence reads from the pangolin sample P1E were extracted from the SRA database, BioProject accession number PRJNA606875, experiment SRX7732093 (Lam et al., 2020). A blast analysis (Altschul et al., 1990) was performed to identify reads including the ORF8 position 27,960, where a mutation causes the premature stop codon Q23stop. The reads were assembled to the SARS-CoV-2 reference using the Geneious mapper under the ‘Highest sensitivity’ setting, running on Geneious Prime 2019.0.4 (https://www.geneious.com). Reads matching position 27,960 close to their ends were excluded to avoid sequencing artifacts. The final dataset included 152 reads (mean length = 300.1; SD = 0.5) holding the position 27,960.

2.3. SARS-CoV-2 in farmed mink (N. vison) from Denmark

All SARS-CoV-2 complete genomic sequences from Denmark with mutation ORF8 E19stop were downloaded from GISAID. The sequences were aligned using the MAFFT version 7 online service (Katoh et al., 2019), with the light-weight option for MSA of full-length SARS-CoV-2 genomes. The SARS-CoV-2 genome with accession number NC_045512.2 was used as reference. The maximum-likelihood phylogenetic tree was constructed using the PhyML 3.0 (Guindon et al., 2010) running in the ATGC bioinformatics platform (http://www.atgc-montpellier.fr). The GTR + I substitution model was selected using the SMS: Smart Model Selection in PhyML (Lefort et al., 2017), with option AIC (Akaike Information Criterion). The branch support was estimated using a bootstrap of 100 trees. The 3D structure of the spike glycoprotein (PDB: 6aqj, EM 4.2 Angstrom) in complex with host cell receptor ACE2 were obtained using the CoVuserver enabled by GISAID.

3. Results

3.1. A SARS-CoV-2 variant with a truncated ORF8 (Q23stop) occurred in a pangolin (M. javanica) from Guangxi, China

SARS-CoV-2-related viruses have been identified in Malayan pangolins (M. javanica), but is still unknown if they were at the origin of the COVID-19 pandemic (Lam et al., 2020; Dimonaco et al., 2021). I found that a SARS-CoV-2-related virus from a pangolin sample (sample P1E; EPI_ISL_410539) had a truncated ORF8 resulting from a Q23stop mutation (Table 1). This pangolin was seized by the Guangxi Customs, China, during their routine anti-smuggling operations in 2017 (Lam et al., 2020).

The pangolin sample P1E and human SARS-CoV-2 reference genomes had 80.3% of identical sites in the ORF8 gene, in line with previous studies (Mohammad et al., 2020; Pereira, 2020). Both lineages diverge by 16 out of 121 amino acids in the ORF8 protein (86.9% of identity). I aligned the ORF8 gene sequences from all pangolin samples reported in Lam et al. with the SARS-CoV-2 reference sequence (Fig. 1A). After excluding a sequence with a large sequencing gap in ORF8, I found that all pangolin sample had an equal ORF8 gene, with exception of the Q23stop mutation at sample P1E. The similarity among samples from different pangolins suggests that the Q23stop mutation was recent, perhaps even within the sampled individual. If the lineage of sample P1E had acquired the ORF8 Q23stop mutation before, one should expect the accumulation of other mutations in a gene encoding a non-functional protein that was not under selection. The possibility of Q23stop resulting from a sequencing error should be considered in old samples stored in unknown conditions for a few years. However, I found that 151 out of 152 sequencing reads for ORF8 had a T that causes the premature stop codon, with a single read showing the reference C nucleotide. It is therefore unlikely that errors would generate such a clear pattern in obtained sequences.

In terms of the spike gene, sample P1E and human SARS-CoV-2 had
Fig. 1. SARS-CoV-2-related viruses in Malayan pangolins (*Manis javanica*). A) Phylogeny of SARS-CoV-2 and related viruses sampled in bats and pangolins. SARS-CoV-2 from the COVID-19 epidemic are coloured in red, while related SARS-like coronaviruses are coloured in blue. The sample P1E (EPI_ISL_410539) is highlighted by a green box. Tree obtained from the Nextstrain website. B) and C) Identity plot for the alignment of ORF8 (B) and spike (C) genes from SARS-CoV-2-related viruses sequenced from five pangolin samples collected in Guangxi, China. Highlighted in B) is the ORF8 gene region with the mutation C27960T causing the premature stop codon Q23stop observed in sample P1E. The alignments include the reference SARS-CoV-2 genome (NC_045512.2). The most conserved positions are indicated by green bars.
Fig. 2. SARS-CoV-2 variants with truncated ORF8 observed in farmed mink (Neovison vison). A) Maximum-likelihood phylogenetic tree with all SARS-CoV-2 complete genome sequences with the ORF8 E19stop mutation detected in Denmark. Sequences obtained from mink are shown in red. The cluster of sequences with the spike N501T mutation is highlighted by an orange box. Bootstrap support values are shown in main branches. B) Predicted 3D structure of the spike glycoprotein (grey ribbon) in complex with host cell receptor ACE2 (green ribbon), with amino acids 501 (yellow circles) and 614 (blue circles) highlighted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
4. Discussion

The ORF8 accessory gene is believed to interfere with immune responses and many putative cellular interactors have been identified, but its precise function remains unknown (Gordon et al., 2020; Flower et al., 2021). Expression of ORF8 is not essential for the virus replication, as suggested by the several variants lacking this gene identified worldwide (Pereira, 2020, 2021; Zinzula, 2021; Alkhansa et al., 2021). This fact was first noticed in a lineage with an ORF8 382-nucleotide deletion detected in Singapore and Taiwan (Gong et al., 2020; Su et al., 2020). Curiously, middle and late phases of the 2002/2003 SARS epidemic were characterized by the spread of SARS-CoV with either partial or complete deletions of the ORF8 gene (Guan et al., 2003; C.S.M.E. Consortium, 2004). It remains unclear if the truncated ORF8 may have changed the SARS-CoV replication capacity or virulence in a way that favoured its adaptation to humans and/or its spread during the SARS epidemic (Guan et al., 2003; C.S.M.E. Consortium, 2004; Lau et al., 2005; Chen et al., 2007). Whatever the case, deletions at ORF8 were the most remarkable changes observed during the SARS epidemic. The same may be happening in SARS-CoV-2, as more transmissible lineages become dominant as the pandemic progresses.

Patients infected with variants lacking ORF8 had less severe symptoms and a possible prolonged infection period that may increase the opportunities for the virus spread (Young et al., 2020). Interestingly, the VOC-202012/01 lacks ORF8 in addition to several spike mutations. It remains to be determined if the absence of ORF8 contributes to the increased transmissibility. The longer duration of symptoms resulting from the absence of ORF8 may increase the opportunity for transmission, and the milder symptoms may increase undiagnosed cases and the number of transmissions.

The identification of a pangolin harbouring a SARS-CoV-2-related virus without a functional ORF8 suggests that lossing this accessory gene may be a common event. In addition to SARS-CoV, deletion of accessory genes have been observed in the Middle East respiratory syndrome related coronavirus (MERS-CoV) (Lamers et al., 2016), feline coronavirus (FCoV) (Kenny et al., 2001), porcine respiratory coronavirus (PRCV) (Chen et al., 2019) and mouse hepatitis virus (MHV) (de Haan et al., 2002), often leading to an attenuation of virulence. It is clear that accessory genes play a dynamic role in such viruses, often varying along the progression of the pandemics.

The pangolin sample P1E has 92 spike amino acid substitutions in relation to the reference SARS-CoV-2, including V483Q, F490Y, N501T and P681M. The spike protein consists of 2 subunits (S1 and S2) that recognize and binding ACE2 (Gu et al., 2020). Mutation N501T gives rise to a hydrogen bond to lysine 353 (van Dorp et al., 2020) and was shown to enhance ACE2 affinity (Starr et al., 2020). According to GISAID, spike mutation N501T occurred 47 times (0.02% of all samples with Spike sequence) in 9 countries.

4. Discussion

The ORF8 accessory gene is believed to interfere with immune responses and many putative cellular interactors have been identified, but its precise function remains unknown (Gordon et al., 2020; Flower et al., 2021). Expression of ORF8 is not essential for the virus replication, as suggested by the several variants lacking this gene identified worldwide (Pereira, 2020, 2021; Zinzula, 2021; Alkhansa et al., 2021). This fact was first noticed in a lineage with an ORF8 382-nucleotide deletion detected in Singapore and Taiwan (Gong et al., 2020; Su et al., 2020). Curiously, middle and late phases of the 2002/2003 SARS epidemic were characterized by the spread of SARS-CoV with either partial or complete deletions of the ORF8 gene (Guan et al., 2003; C.S.M.E. Consortium, 2004). It remains unclear if the truncated ORF8 may have changed the SARS-CoV replication capacity or virulence in a way that favoured its adaptation to humans and/or its spread during the SARS epidemic (Guan et al., 2003; C.S.M.E. Consortium, 2004; Lau et al., 2005; Chen et al., 2007). Whatever the case, deletions at ORF8 were the most remarkable changes observed during the SARS epidemic. The same may be happening in SARS-CoV-2, as more transmissible lineages become dominant as the pandemic progresses.

Patients infected with variants lacking ORF8 had less severe symptoms and a possible prolonged infection period that may increase the opportunities for the virus spread (Young et al., 2020). Interestingly, the VOC-202012/01 lacks ORF8 in addition to several spike mutations. It remains to be determined if the absence of ORF8 contributes to the increased transmissibility. The longer duration of symptoms resulting from the absence of ORF8 may increase the opportunity for transmission, and the milder symptoms may increase undiagnosed cases and the number of transmissions.

The identification of a pangolin harbouring a SARS-CoV-2-related virus without a functional ORF8 suggests that lossing this accessory gene may be a common event. In addition to SARS-CoV, deletion of accessory genes have been observed in the Middle East respiratory syndrome related coronavirus (MERS-CoV) (Lamers et al., 2016), feline coronavirus (FCoV) (Kenny et al., 2001), porcine respiratory coronavirus (PRCV) (Chen et al., 2019) and mouse hepatitis virus (MHV) (de Haan et al., 2002), often leading to an attenuation of virulence. It is clear that accessory genes play a dynamic role in such viruses, often varying along the progression of the pandemics.
Declaration of Competing Interest

The author declare that he has no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgments

I would like to thank all the researchers who have kindly shared genes on public databases.

Funding

This work was supported by the Fundação para a Ciência e a Tecnologia [RESEARCH 4 COVID-19 project No. 029] and the EOSCsecretariat.eu COVID-19 Fast Track Funding. EOSCsecretariat.eu has received funding from the European Union’s Horizon Programme call H2020-INFRAESC-05-2018-2019, grant Agreement number 831644.

References

Alkhanza, A., Lakkis, G., El Zein, L. 2021. Mutational analysis of SARS-CoV-2 ORF8 during six months of COVID-19 Pandemic. Gene Reports 23, 101024. https://doi.org/10.1016/j.genrep.2021.101024.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403-410.

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. The proximal origin of SARS-CoV-2. Nat. Med. 26 (4), 450-452.

Chand, M., Hopkins, S., Baber, G., Ackison, C., Barclay, W., Ferguson, N., Volf, E., Loman, N., Rambaut, A., Barrett, J. 2020. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01. Public Health England.

Chen, C.-Y., Ping, Y.-H., Lee, H.-C., Chen, K.-H., Lee, Y.-M., Chan, Y.-J., Lien, T.-C., Jap, T.-S., Lin, C.-H., Kuo, I.-S., 2007. Open reading frame 8a of the human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induces apoptosis. J. Infect. Dis. 196 (3), 405-415.

Chen, F., Knutson, T.P., Rossow, S., Saif, L.J., Marthaler, D.G., 2019. Decline of canine coronavirus (CCoV) infections in Middle East. Emerg. Microbes. Infect. 9 (1), 1457-1459. https://doi.org/10.1080/22133123.2020.1866055.

F. Pereira Gene 784 (2021) 145596

Volz, E., Mistry, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinley, W.R., Laydon, D.J., Baber, G. and O’Toole, A. 2021. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. medRxiv. 2021.12.20.21250039.

Yu, Y.C., Anderson, Y., Young, B.E., Fong, S.-W., Chan, Y.-H., Mak, T.-M., Ang, L.W., Anderson, D.E., Lee, C.-P., Amnum, S.N., Lee, B., Goh, Y.S., 2020. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet. 396 (10251), 603-611.

Zhang, Y.Z., Holmes, E.C., 2020. A new perspective on the origin and emergence of SARS-CoV-2. Cell 181 (2), 223-228.

Zhao, X., Yang, X.L., Wang, G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, P., Huang, C.-L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 583 (7813), 285-289.

Zuluina, L. 2021. Lost in deletion: The enigmatic ORF8 protein of SARS-CoV-2. Biochem. Biophys. Res. 538, 116-124.