Evolution, virulence and immunogenicity of relevant SARS-CoV-2 Spike mutants

M. Coscolla[1], P. Ruiz-Rodriguez[1], C. Frances[1], R. Geller[1], A. Chiner-Oms[2], S. Jimenez-Serrano[2], A. Marina[2], L. Martinez-Priego[3], G. D'Auria[1,3], M. Lopez[2], I. Carcino-Muñoz[2], M. Torres-Puente[2], M. Bracho[1,4], P. Ruiz-Hueso[3], F. Gonzalez-Candelas[1,4], I. Comas[2,4], SeqCOVID-SPAIN consortium.

1.I²SysBio, University of Valencia-CSIC, FISABIO Joint Unit - Paterna (Spain)

2.Instituto de Biomedicina de Valencia (IBV-CSIC) - València (Spain)

3.Sequencing and Bioinformatic Service of FISABIO (Valencian Region Foundation for the Promotion of Health and Biomedical Research - València (Spain)

4.CIBER in Epidemiology and Public Health (Spain)

31st ECCMID Online 9 – 12 July 2021

ESCMID EUROPEAN SOCIETY OF CLINICAL MICROBIOLOGY AND INFECTIOUS DISEASES

ABSTRACT

We have detected **two mutations** in the Spike protein of SARS-CoV-2 sequences at amino acid positions **1163** and **1167** which appeared independently in multiple transmission clusters and **different genetic backgrounds**, indicating they may increase viral fitness.

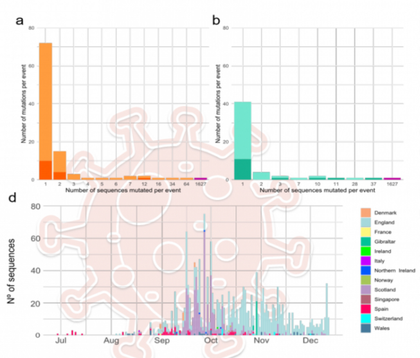
Both mutations appeared together in a cluster within **clade 20E**. This cluster is characterized by 12 additional single nucleotide polymorphisms but no deletions. The available structural information of the S protein in the pre- and post-fusion conformations we predict **that both mutations confer rigidity**, that potentially could decrease viral fitness.

Despite the multiple and successful appearance of two **HR2 mutations** during the first year of SARS-CoV-2 evolution, in **vitro stability, infectivity, or antibody escape does not seem to play a role**

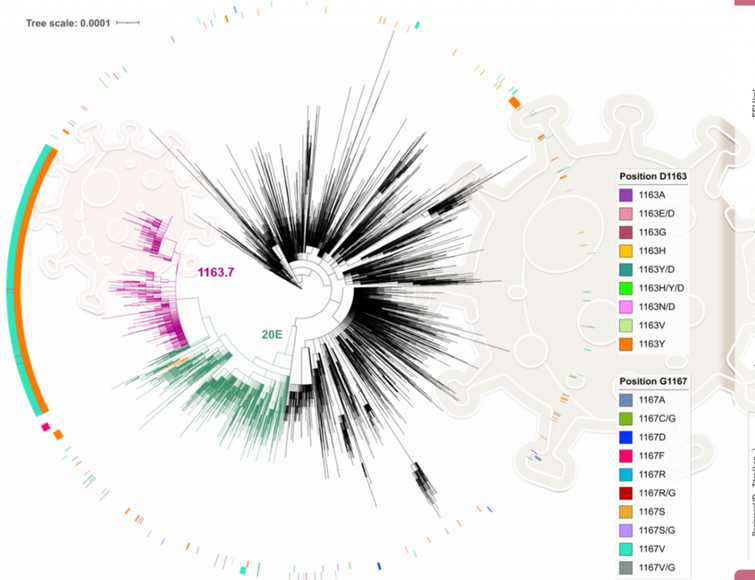
MATERIAL & METHODS

- A total of 5,017 clinical samples were received, sequenced, and analysed by the **SeqCOVID consortium** from all autonomous communities of **Spain**. These samples were confirmed as SARS-CoV-2 positive by RT-PCR carried out by Clinical Microbiology Services from each hospital.
- To build the global alignment, all sequences from **GISAID** including all the pandemic periods since the first known case sequenced until end of 2020 were downloaded . Sequences were aligned against the SARS-CoV-2 reference genome and maximum-likelihood phylogenies were reconstructed from the masked alignment with GTR model and collapsing near-zero branches.
- We used the phylogeny of **10,450** sequences enriched with all sequences mutated in **1163** and **1167** to quantify the minimum number of mutational events involving positions 1163 and 1167 in S protein.
- Pseudotyped VSV with the desired sequence if the SARS-CoV-2 sequence was used for phenotyping. VSV with different S SARS-CoV-2 sequences were used to infect Vero and hACE2–TMPTSS2 cells and estimate **in vitro virulence**. **Antibodies neutralization** was tested using sera from infected patients from first wave, second wave, and Pfizer vaccinated individuals.

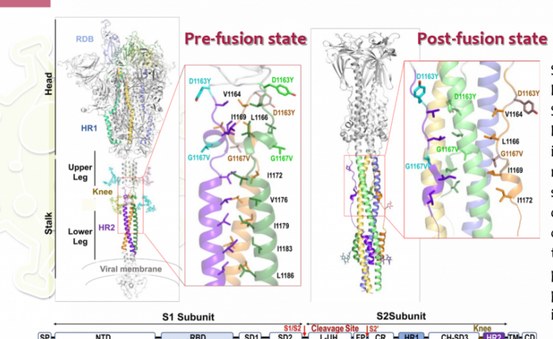
RESULTS 1



Positions **1163** and **1167** in the S protein have mutated independently multiple times in SARS-CoV-2. Most of mutated sequences (94.43%) were found in transmission clusters. Only one change at each position appeared in most clusters: **D1163Y** in 83.33% and **G1167V** in 69.23% clusters. **D1163Y** appeared in 22 transmission clusters and **G1167V** in 8 clusters Interestingly, the biggest cluster included both the **D1163Y** and **G1167V** mutations together and was detected initially in 65 sequences from Spain until December 2020, representing 1.17% of the Spanish sequences. The 1,627 sequences form a monophyletic cluster which within **lineage 20E**, and we denote it as **cluster 1163.7**.

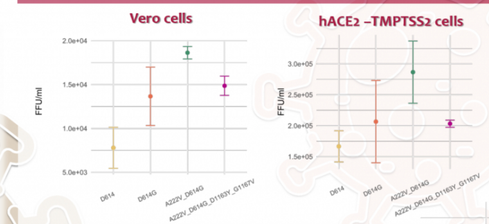


*Maximum-likelihood phylogeny of **10,450** SARS-CoV-2 genomes. The inner circle of the rings represents sequences with amino acid changes in position D1163 of the S protein. The external circle represents sequences with amino acid changes in position G1167 of the S protein. Branches are coloured in magenta for 1163.7, green for clade 20E, and orange for cluster 1163.654.*



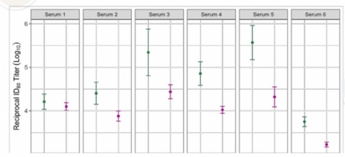
S protein positions **1163** and **1167** are both located within the **HR2 domain.** Specifically, 1167 is present at the beginning of the **HR2 motif** and **1163** in its upstream linker region. **G1167V** mutation is predicted to confer significant rigidity. First, the introduction of a side chain strongly reduces the conformational freedom provided by the **glycine residue**. Second, the presence of the new aliphatic side chain provided by the **valine residue** strongly increases hydrophobicity.

RESULTS 2



20E S genotype enhanced infectivity relative to the **D614 S genotype** by 70% in both Vero and A549-hACE2-TMPRSS2 cells (p-value < 0.05 by unpaired t-test). The **20E S genotype** also showed a trend towards increased infectivity versus the **D614G** mutation alone (35% increase in both cell lines), not statistically significant (p-value > 0.05). **1163.7 S genotype does not increase infectivity *in vitro*.**

We tested if individuals infected with **1163.7** have different viral loads with the **cycle threshold** (Ct). Higher viral loads were observed in individuals infected with **1163.7** and **20E** (**D614G** and **A222V**) compared to the **D614G** alone (**D614G** Ct mean = 25.32, **20E** Ct mean = 21.14, **1163.7** Ct mean = 20.63, p-value < 0.01 for both comparisons by unpaired Wilcoxon test). **This data suggests that, *in vivo viral load is similar* to other 20E genotypes .**



1163.7 S genotype conferred a modest but statistically significant reduction in sensitivity to neutralization by six serum samples tested from the first wave of the pandemic, (ID_{80}; mean = 6.75, range: 1.30-17.68; p-value = 0.008 by paired t-test). In contrast, both **20E** and **1163.7** S genotypes were equally susceptible to sera from patients infected during the second wave.

No significant differences in susceptibility to antibody neutralization from vaccinated donors were observed between the two genotypes.

CLUSTER UPDATE

*In June 2021, we do not detect transmission of **cluster 1163.7**. However, sequences harbouring amino acid changes **D1163Y** or **G1167V** are still appearing in a different PANGO lineages.*

CONCLUSION

1-We have identified **two mutations** in the S protein that are likely to be beneficial for the virus. The largest cluster, and therefore the most **successful** in terms of transmission, includes both mutations together. Additionally, both positions have been reported as positively selected **multiple** times throughout the SARS-CoV-2 phylogeny **indicating a possible fitness advantage.**

2-Structural data suggests that **G1167V** might alter the **flexibility of the S protein stalk.** The stalk flexibility has been suggested to increase avidity for the host receptors.

3-We find the **1163.7 S genotype** to have reduced infectivity compared to the **20E S genotype** in both Vero and A549-hACE2-TMPRSS2 cells. However, we do not see this reduction in viral load in vivo

4-We found a modest but statistically significant reduction in susceptibility to neutralization of the **1163.7 S genotype** compared to the **20E S genotype** in sera from individual infected in the first wave. No difference in neutralization was observed between the two variants in sera from patients infected during the second wave and vaccinated patients.

This work was funded by the Instituto de Salud Carlos III project COV20/00140 and COV20/00437, Spanish National Research Council project CSIC-COV19-021 and CSIC-COVID19-082, and the Generalitat Valenciana (SEJI/2019/011 and Covid_19-SCI). Action co-financed by the European Union through the Operational Program of the European Regional Development Fund (ERDF) of the Valencian Community 2014-2020. MC is supported by Ramón y Cajal program from Ministerio de Ciencia, grant RTI2018-094399-A-I00.