Evolutionary emergence of SARS-CoV-2 and a caution for pharmaceutical development

Swapnil Tichkule[1,2], Aaron Jex[1,3] and Cock van Oosterhout[4]

1. Population Health and Immunity, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia
2. Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, VIC, Australia
3. Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Melbourne, VIC, Australia
4. School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, UK

Email: tichkule.s@wehi.edu.au

Introduction

COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

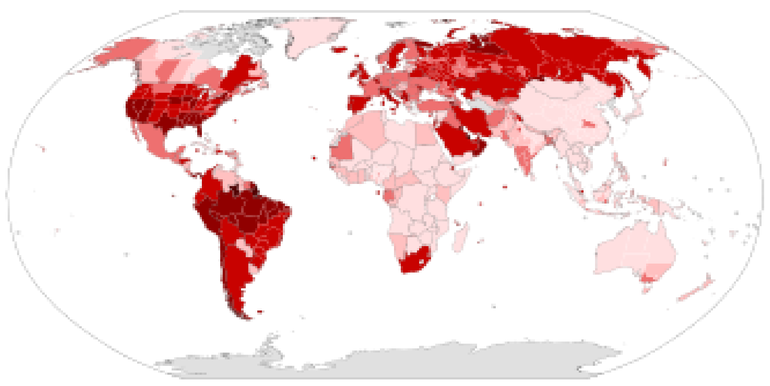
"How did SAR-CoV-2 evolve virulence to the human host?"

Disease burden - As of 23 July

>15.2 million cases worldwide

188 Countries and territories

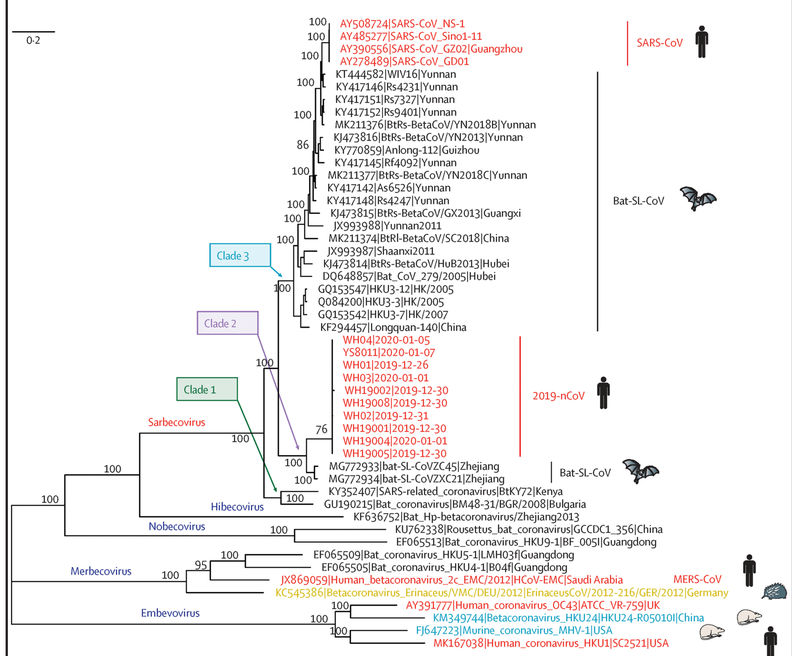
> 623,000 Deaths



SARS-CoV-2

SARS-CoV-2 belongs to subgenus Sarbecovirus, genus Betacoronavirus.

After SARS-CoV (2003) and MERS-CoV (2012), SARS-CoV-12 (2019) is the 3rd pathogenic coronavirus infecting humans.



Phylogenetic tree representing SARS-CoV-2 (2019-nCoV) and representatives of Betacoronaviruses [4].

COVID-19 Research and Motivation

- Scientists across the globe are trying to elucidate the genome characteristics of SARS-CoV-2 to prevent and control the spread of the disease.
- This study contributes towards better understanding of the both evolutionary emergence of the SARS-CoV2, as well as its disease progression.
- First, we determined the evolutionary link between SARS-CoV-2 and the immediate ancestors of SARS-related coronaviruses.
- Secondly, we traced the temporal genomic signature in SARS-CoV2 since its emergence from December-2019

I. Origin of SARS-CoV-2

Identification of infection source is crucial in outbreaks of zoonotic pathogens.

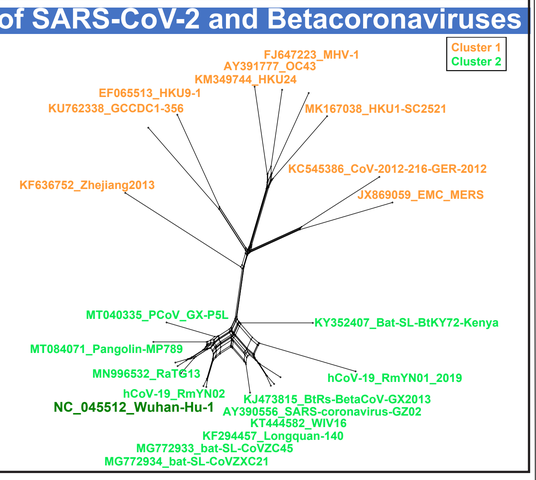
? SARS-CoV-2 evolve through genetic introgression after hybridisation?

If so, what were the "parental" sequences?

How long ago did this hybridisation happen?

Evolutionary relationship of SARS-CoV-2 and Betacoronaviruses

Figure 1. SplitsTree network analysis of 22 species representing Betacoronavirus. Cluster2 with SARS-CoV-2 shows loops which illustrates phylogentic incongruencies and potential recombination.



Recombination analysis and determining ancestor

Recombination analysis determined genetic introgression and hybridized regions in SARS-CoV-2 (Fig 2A).

Identified major (recombinant-free region) and minor (recombinant events) parents of SARS-CoV-2 with network analysis (Fig 2B-C).

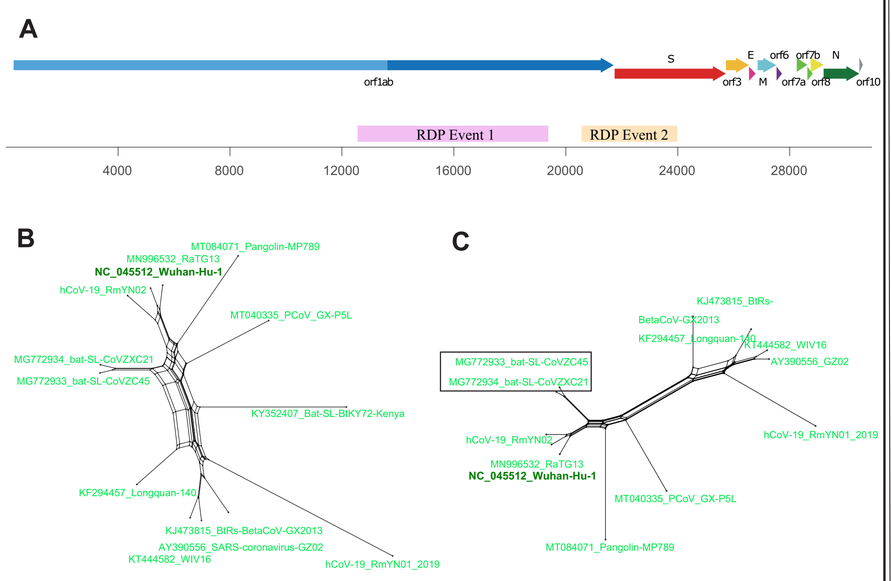


Figure 2. A. Genomic position of two recombinant breakpoints on SARS-Cov-2 genome. **B.** Network of cluster2 showing larger loops. **C.** Network of major parent (recombinant-free) showing bat-SARS-like-CoV (box) are closely associated with SARS-CoV-2 (Wuhan) apart from RmYN02, RaTG13, highly similar species; without any loops.

Similarity and divergence between SARS-Cov-2 and bat-SL-CoV

HYBRIDCHECK program was used to determine similarity and hybridzation in SARS-CoV-2 by considering multiple triplet represention (hybrid=SARS-CoV-2, major= bat-SL-CoVZC45, minor=other).

The divergence of SARS-Cov-2 from bat-SL-CoV is estimated to be 31-40 and 48-63 years ago according to substution rate of 1.24E-3 and 8E-4 resp.

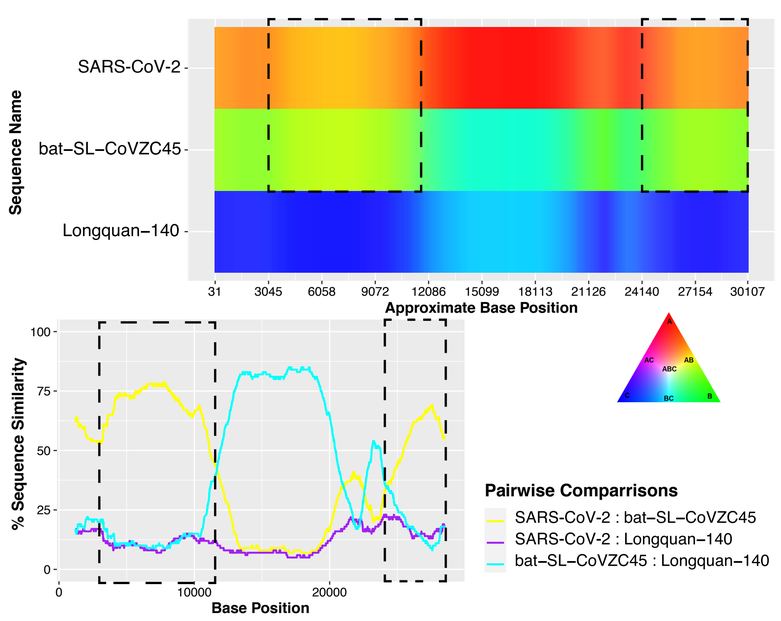


Figure 3. Yellow colour in both the plots represents similarity between the SARS-CoV-2 and bat-SL-CoV. Lower % of similarity represents hybridiasation in the SARS-CoV-2 genome.

II. Progression of COVID19

Spike (S) protein is responsible for host cell entry.

It is highly polymorphic region of SARS-CoV-2 genome.

Multiple recombination events were detected in this protein.

? Variablity of S protein play role in disease progression ?

Diversifying selection in Spike protein

Dataset: 1. December 2019 - January 2020 (n=92)

2. April 2020 - 4th May 2020 (n=148)

7 polymorphic sites in S protein of total 240 sequences.

Identified: codon 614 under diversifying selection (Fig 4A)

codon 614, only non-synonymous mutation (Fig 4B)

D614G mutation diversifies between early (Dec-Jan) and late (April-May) sampled sequences. Supported by phylogenetic analysis (Fig 4).

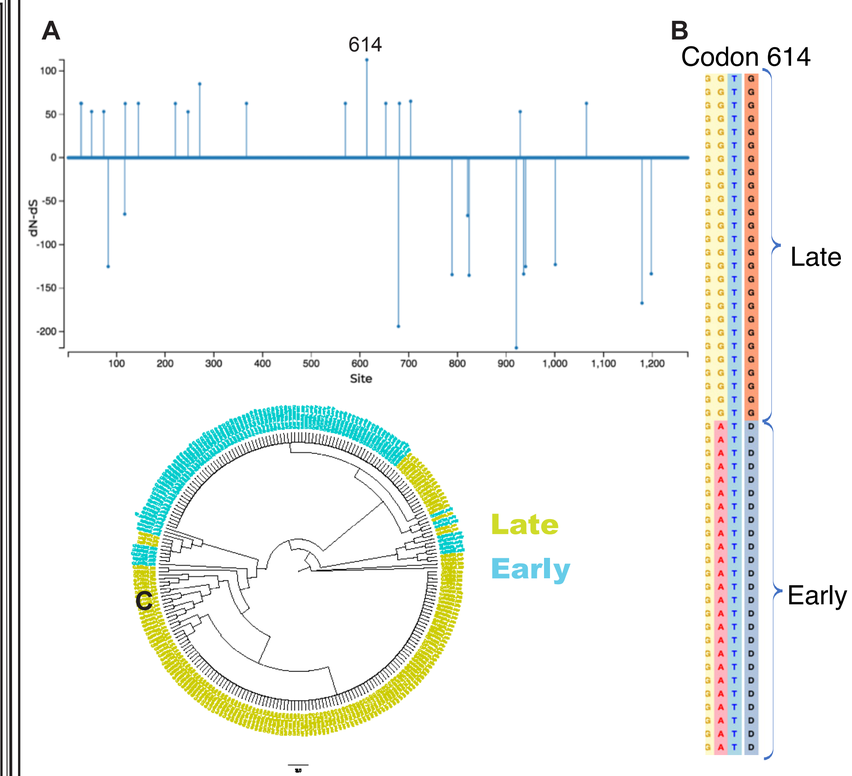


Figure 4. A. dn-ds graph showing diversifying selection on codon 614**. B.** Non-synonymous mutation at codon 614 and difference between early and late sampled sequences. **C.** Phylogenetic analysis showing cseggregation between the same.

Conclusion

- Introgression between SARS-CoV-2 and bat-SARS-related coronaviruses.
- SARS-CoV-2 diverged <70 years ago from bat-SARS related coronaviruses.
- Demonstrated potential origins of SARS-CoV-2 from SARS-related coronaviruses associated with the spill-over in humans.
- Different time-scales identified mutation (D614G) with a diversifying selection in S protein.
- D614G changes a peptide that is thought to interact with the human cell
- D614G is evolving in recent outbreaks of COVID-19 across the world.
- S protein-otential target for therapeutics, pharmaceuticals will

Reference

1. Zhou, Hong, et al. "A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein." Current Biology (2020).

2. Korber, Bette, et al. "Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2." bioRxiv (2020).

3. Boni, Maciej F., et al. "Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic." bioRxiv (2020).

4. Kumar, Dharmender, Lalit Batra, and Mohammad Tariq Malik. "Insights of Novel Coronavirus (SARS-CoV-2) disease outbreak, management and treatment." AIMS Microbiology 6.3 (2020): 183.

5. Li, Xingguang, et al. "Evolutionary history, potential intermediate animal host, and cross□ species analyses of SARS-CoV-2." Journal of medical virology 92.6 (2020): 602-611.

6. MacLean, Oscar A., et al. "No evidence for distinct types in the evolution