SARS-CoV-2 Situation Report - 2020-05-01

Mapping the phylodynamics of SARS-CoV-2 in New Zealand using Transcendental Information Cascades

Markus Luczak-Roesch[1,2]

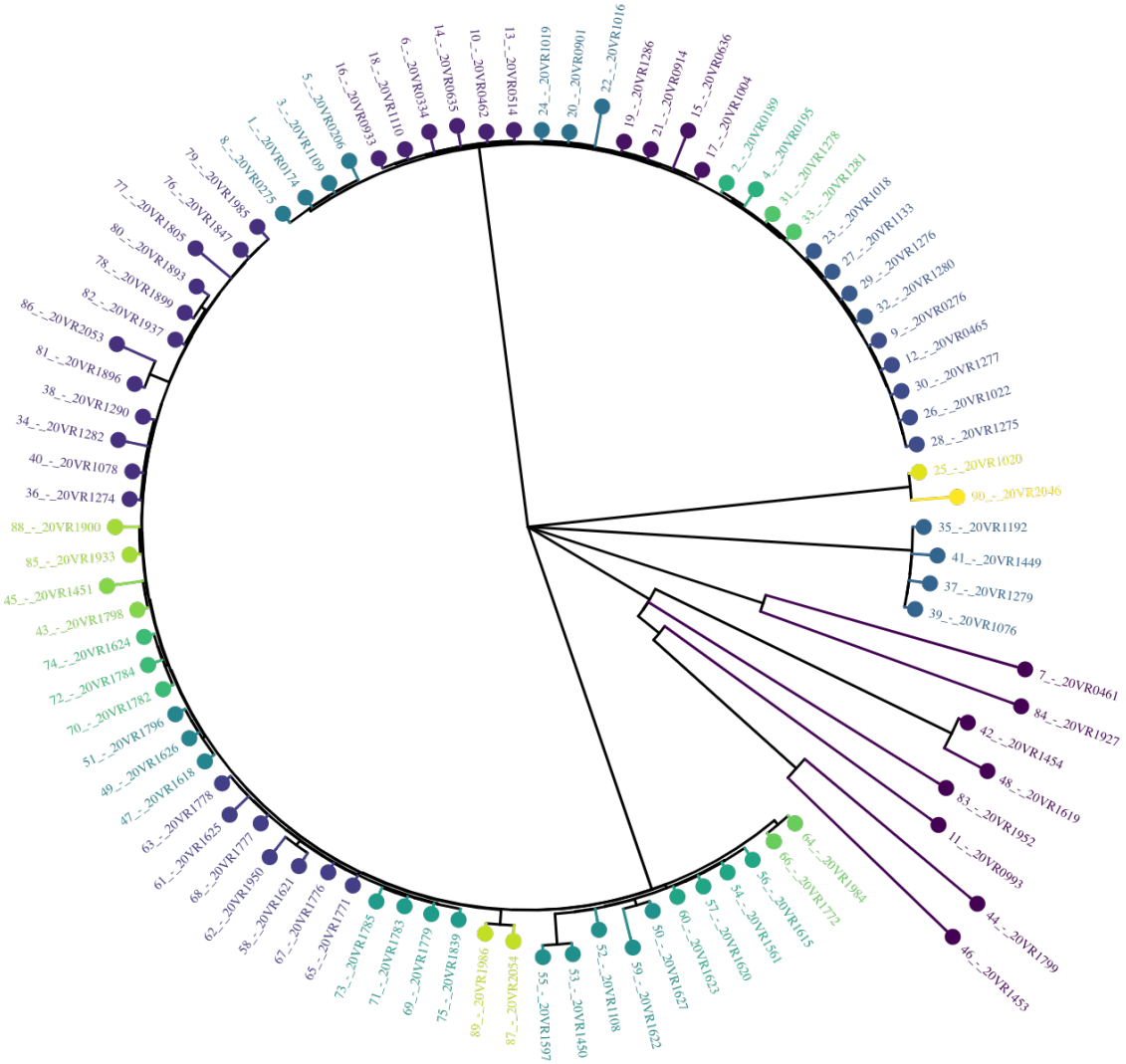
1. Victoria University of Wellington, School of Information Management, New Zealand
2. Te Pūnaha Matatini - the Centre for Complex Systems and Networks, New Zealand

Methods

We construct Transcendental Information Cascades over the genomic sequences ordered by their date of collection as described in [1]. The tokenisation captures all unique codons in a particular position in the genomic sequences (truncated flanks to the consensus range 56 to 29,797 [2]) and in the +1, +2 and +3 reading frame. This results in a directed acyclic graph from which we can derive a phylogenetic tree through community detection.

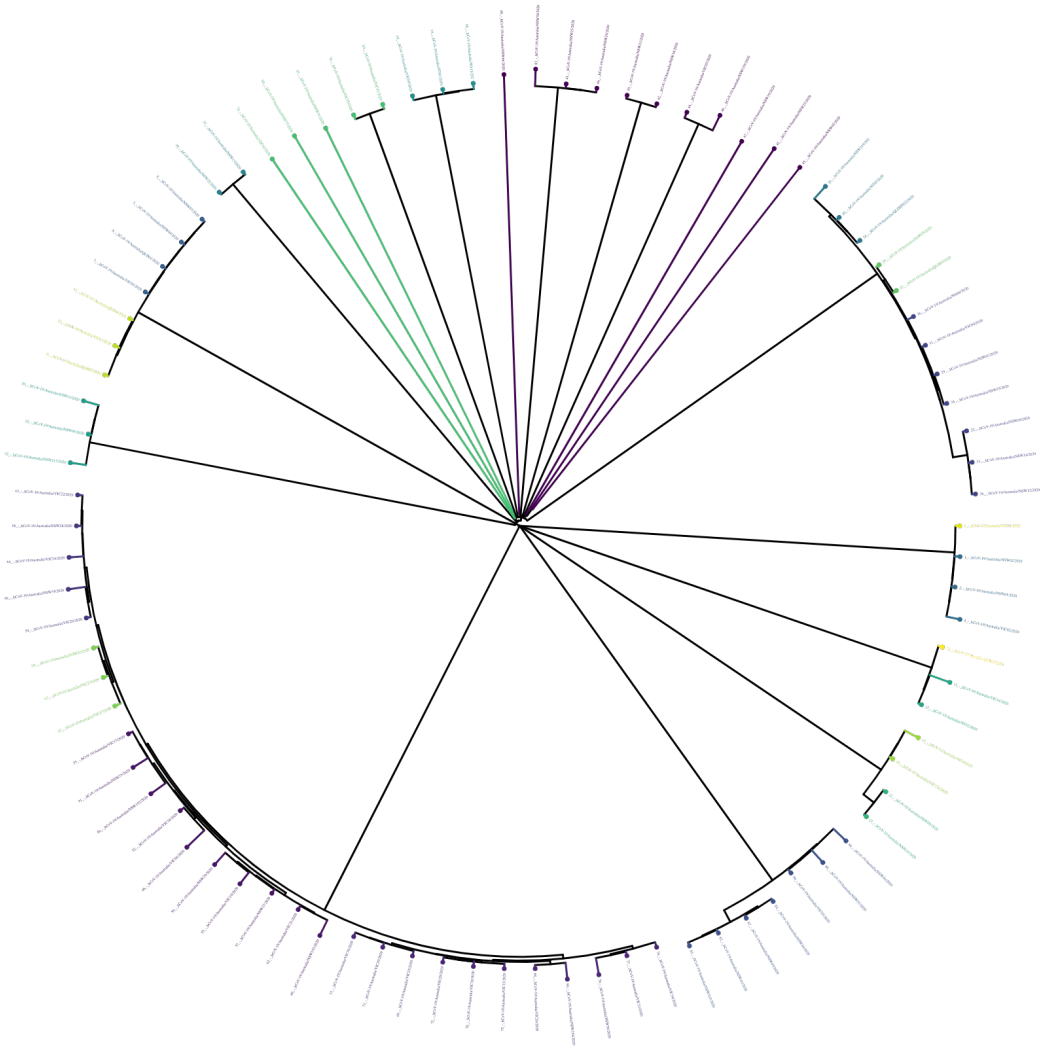
Additionally, we also construct the Transcendental Information Cascade over the currently available case metadata. In particular we tokenise age (in 5-year brackets), gender, hospitalisation status, country linked to overseas travel, way of transmission, status of symptoms and cluster membership. This can lead to just one integrated additional network layer that links cases that share common metadata properties or one layer per metadata property.

Results



Phylogenetic tree of NZ SARS-CoV-2 samples (90 with metadata as of today) inferred from the TIC network through random walk community detection (colour indicates cluster membership within branch/sub-tree, number of clusters: 23). We filtered out edges with weight larger than 300 and consider only the +1 reading frame. Additionally we filtered out codons that are overrepresented according to RSCU index over all samples [3,4]. Branch length reflects codon similarity in the +1 reading frame between genomes.

Phylogenetic tree of the first 90 SARS-CoV-2 samples from Belgium. Same filters applied as for NZ tree. We highlight that it is obvious that the tree shows a lower number of clusters (17) but a higher diversity of these clusters, indicating that there is a higher degree of within-country transmission while the New Zealand tree suggests imported cases that are not substantially passed on to other members of the community except immediate close contacts.



Phylogenetic tree of the first 90 SARS-CoV-2 samples from Australia. Same filters applied as for NZ tree above. This tree shows higher structural similarity to the phylodynamics we captured for New Zealand data (AUS: 21 clusters). In the light of the current understanding of the containment success of New Zealand and Australia compared to Belgium, this is reasonable and suggestive that there is indeed no evidence for wide-ranging uncontrolled community transmission within Australia and New Zealand.

Metadata TIC for location of person (TA) and cluster membership as per EPI database.

Suggestions for further case investigation

- 20VR0993 - 20VR1952 - 20VR1927 - 20VR0461
- 20VR1625 - 20VR1777 - 20VR1778
- 20VR1771 - 20VR1776 - 20VR1621 - 20VR1950

Limitations

This situation report is based on the limited amount of genomic data that is available as of today and a preliminary analysis of the codon recurrence tracing approach using Transcendental Information Cascades only. The network and the inferred phylogenetic tree have not been validated against state-of-the art methods such as phylogenetic trees constructed using a maximum likelihood or Bayesian approach. It is important to note that the tree only reflects codon similarity as per network clustering and is not reflective of nucleotide mutations as in traditional phylogenetic trees. The analysis is meant to complement traditional phylogenomic analysis with an alternative macroscopic angle to the phylodynamics of the epidemic outbreak.

Acknowledgements

We want to thank Matthew Tansley for his work on visualising the case data on a geographical map. This work has also been greatly supported by Michael Thingnes' work on

the pyhton scripts to construct Transcendental Information Cascades from DNA sequences. We also want to thank Tanja Karl for her invaluable feedback throughout this analysis process.

References

[1.] Luczak-Roesch, M. (2020). Networks of information token recurrences derived from genomic sequences may reveal hidden patterns in epidemic outbreaks: A case study of the 2019-nCoV coronavirus. medRxiv.

[2.] Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., ... & Yuan, M. L. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265-269.

[3.] Plotkin, J. B., & Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nature Reviews Genetics, 12(1), 32-42.

[4.] Dilucca, M., & Pavlopoulou, A. (2020). Analysis of codon usage and evolutionary rates of the 2019-nCoV genes. bioRxiv.