Networks of information token recurrences derived from genomic sequences may reveal hidden patterns in epidemic outbreaks: A case study of the 2019-nCoV coronavirus.

Markus Luczak-Roesch[1*]

1 Victoria University of Wellington, School of Information Management

*** markus.luczak-roesch@vuw.ac.nz**

Abstract

Profiling the genetic evolution and dynamic spreading of viruses is a crucial when responding to epidemic outbreaks. We aim to devise novel ways to model, visualise and analyse the temporal dynamics of epidemic outbreaks in order to help researchers and other people involved in crisis response to make well-informed and targeted decisions about from which geographical locations and time periods more genetic samples may be required to fully understand the outbreak. Our approach relies on the application of Transcendental Information Cascades to a set of temporally ordered nucleotide sequences and we apply it to real-world data that was collected during the currently ongoing outbreak of the novel 2019-nCoV coronavirus. We assess information-theoretic and network-theoretic measures that characterise the resulting complex network and suggest touching points and temporal pathways that are of interest for deeper investigation by geneticists and epidemiologists.

Introduction

How far a virus will eventually spread at local, regional, national and international levels is of central concern during an epidemic outbreak due to the severe consequences large-scale epidemics have on human well-being [4], the stability and coherence of social systems [8], and the global economy [21]. A thorough understanding of the genetic characteristics of a virus is crucial to understand the way it is (or may be) transmitted (e.g. human-to-human transmissibility) in order to be able to anticipate the potential reach of an outbreak and develop effective countermeasures. The recent outbreak of a new type of coronavirus – 2019-nCoV – provides plenty of examples of the way researchers seek to quickly approach the aforementioned problems of genetic decomposition of the virus [2, 7] and modelling of its transmission dynamics [12].

Here we present a novel approach to gain insights into the transmission dynamics of an epidemic outbreak. Our method is based on Transcendental Information Cascades [13,14] and combines gene sequence analysis with temporal data mining to uncover potential relationships between a virus' genetic evolution and distinct occurrences of infections. We apply this new approach to publicly available genomic sequence data from the currently ongoing outbreak of the 2019-nCoV coronavirus.

Our results show pathways of similarity between certain clusters of 2019-nCoV virus samples taken in different geographical locations, and indicate coronavirus cases that are candidates for further investigation into the human touch points of these patients in

order to derive a detailed understanding of how the virus may have spread, and how and why it has genetically evolved. Our analysis provides a unique view to the dynamics of the 2019-nCoV coronavirus outbreak that complements knowledge obtained using other state-of-the-art methods.

Materials and methods

Research data and preprocessing

We obtained a total of 61 genomic sequences that were available via the *GISAID EpiFlu Database*™ [1] as of Tuesday, February 4th 2020.

These data were preprocessed using the R programming language. First, we derived a metadata object that allows to store the sequence identifier, the collection location, the collection date, and the raw nucleotide sequence. The metadata object was then filtered to keep only those sequences for which the collection date was not empty and which featured a length that fell within a margin of $1,000$ of the estimated $30,000$ nucleotides that are currently assumed to make up the genomic code of the 2019-nCoV coronavirus. This left us with a total of 56 genomic sequences. We then exported a single CSV file containing only the raw nucleotide sequences in ascending order by the collection date of the sample. This structure is the standard input format for the *genes-CODON-samplesequence* tokeniser featured in the Transcendental Information Cascades R toolchain that is available as free and open scientific software[2].

Construction of Transcendental Information Cascades

Transcendental Information Cascades (TICs) are a method that first transforms any kind of sequential data into a network of recurring *information tokens*, to then exploit this network as a kind of analytical middle ground between views to the source data that would otherwise only be accessible individually through distinct analytical frameworks [13, 14]. The approach falls into the category of contemporary approaches that make use of network theory for the study of dynamical complex systems [3, 16, 19]. In the field of phylogenetic and phylogeographic analysis our approach can be compared to haplotype networks for example [1, 15] but with the benefit that the resulting network preserves access to temporal patterns at both macroscopic (e.g. whole generation time periods) and microscopic (e.g. short time period bursts) resolution.

At an abstract level a Transcendental Information Cascade is constructed from some sequentially ordered source data in the following way: At first the data from each distinct sequence step is processed to extract unique information tokens. These tokens form the *identifier set* for that particular sequence step. Then a directed network is generated with one vertex for each sequence step and edges between any two vertices that have a particular token from their identifier set in common and that does not occur in any identifier set of vertices that represent sequence steps between these two. In other words, identifier sets represent information token co-occurrences and edges represent information token recurrences.

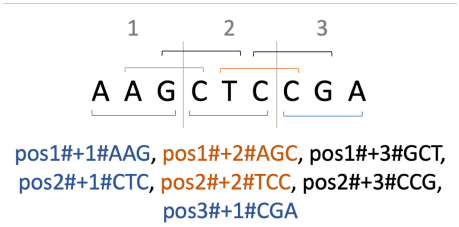

Figure 1. Example of the unique codon tokenisation used in this study. The example sequence comprises three window steps and shows how the +1 (marked in blue), +2 (marked in orange) and +3 (marked in black) reading frames lead to seven unique codon tokens extracted from the nucleotide sequence.

As the information tokens of interest in this study we encode unique codon identifiers as (a) their position when sliding a window of size 3 in steps of size 3 over the nucleotide

[1] https://www.gisaid.org/
[2] https://github.com/vuw-c2lab/transcendental-information-cascades

sequence (e.g. "pos1", "pos2", ...), (b) a flag that indicates the reading frame ("+1", "+2" or "+3") that captures the respective triplet in the current window, and (c) the actual matched nucleotide triplet that constitutes the matched codon. The use of this tokenisation is motivated by previous work on phylogenetic profiling and gene sequencing [5, 10] that suggested that codons suit well for assessing similarities (and dissimilarities) at the gene, chromosome, or genome levels [5, 11, 20].

An example of the particular encoding we use is visualised in Figure 1. It is this tokenisation of the original nucleotide sequences that we use in the TIC approach to construct the information token recurrence network from. This means in our TIC instantiation the nodes represent distinct nucleotide sequences of the virus obtained from patient samples ordered by the date of their collection by researchers (random order in case of same date collections). The edges are unique codon identifier recurrences matched between different nucleotide sequences.

Analysis of Transcendental Information Cascades

The TIC network we derive in the way described before can be analysed in a variety of ways. Information-theoretic and network-theoretic measures have been introduced as the base framework to start the analysis from [13]. In particular we perform (1) an analysis of the information token entropy and evenness over time, (2) an analysis of network clusters of sequences that can be detected in the TIC network, and (3) an analysis of the intra-cluster and inter-cluster nucleotide similarity of the original sequences.

Information token entropy and evenness We perform an analysis of information token entropy and information token evenness in order to understand whether the virus evolution can be considered an open or a closed system. Therefore we assess both measures in an accumulated fashion for each progression step through the ordered nucleotide sequences. In a closed system the information token entropy should strive towards an equilibrium of maximum entropy and not feature any wave-like up and down patterns. If we observe wave-like patterns it means nucleotide sequences at later progression stages feature codon identifiers that have not been observed in earlier nucleotide sequences.

So to measure information token entropy we assess which unique codon identifiers were seen up to a progression step and how often these were seen. This allows to compute the probability for every codon identifier, which then can be used to determine the information token entropy up to that point following the definition of Shannon entropy given as follows:

$$H' = -\sum_{i=1}^{R} p_i \ln p_i \tag{1}$$

Analog to this we can then assess the information token evenness as per Pielou's species evenness index, which is defined as

$$J' = \frac{H'}{H'_{\max}} \tag{2}$$

where H' is the measured Shannon entropy, and $H'_{\max}$ is the maximum entropy that would be measured if all tokens would be equally likely to occur.

Network analysis The Transcendental Information Cascade network we constructed can be mathematically and visually analysed in order to find structures that are of particular significance for a single epidemic outbreak or to compare different outbreaks

structurally. In this study we focus on the former due to the lack of comparative data about other epidemic outbreaks.

We first perform a cluster analysis using the random walk apporach by Pons and Latapy [18] on the weighted network that we can construct from our original TIC network by collapsing all edges between the same vertices to unique edges weighted by the sum of the collapsed edges (see . Afterwards, we analyse visually the network using the open source software Gephi. In Gephi we scale vertices by their degree and edges by their weight. Furthermore, we colour vertices by their cluster membership. We then run the *Yifan Hu* [9] layout algorithm (optimal distance: 10,000; relative strength: 0.2; initial step size: 20; step ratio: 0.95; quadtree max level: 10; theta: 1.2; convergence threshold: $1 * 10^{-4}$; adaptive cooling: enabled) to visualise the network.

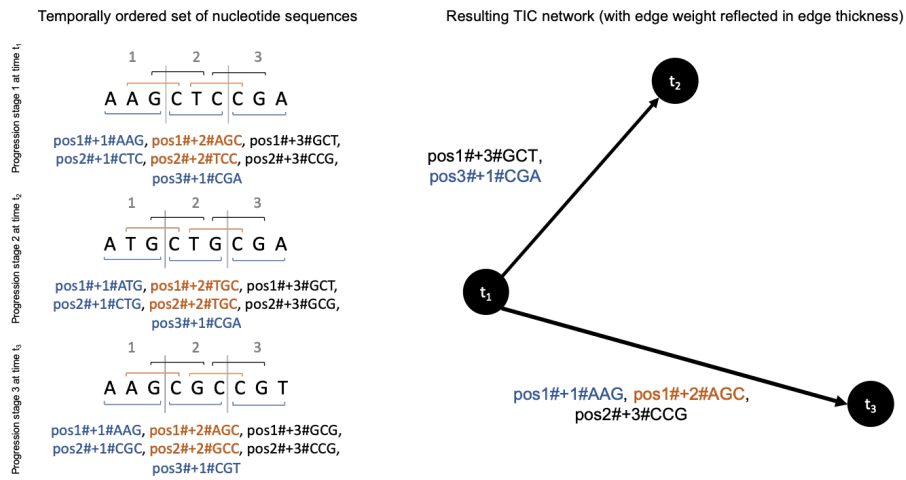

Figure 2. Schematic example of the weighted TIC network constructed from nucleotide sequences with tokenised unique codon identifiers.

Cluster evaluation In order to evaluate the meaningfulness of the TIC network against some baseline, we computer the intra-cluster and inter-cluster similarity of the raw nucleotide sequences. This is done by computing all pairwise comparisons of sequences within individual clusters (intra-cluster similarity) and comparison between all sequences from a cluster with all sequences that are not within that cluster (inter-cluster similarity) using the *compareStrings* function provided by the Biostrings R package [17]. The similarity of two sequences is then simply defined as follows:

$$sim(a,b) = 1 - \frac{number \quad of \quad different \quad nucleotides}{length \quad of \quad compared \quad nucleotide \quad sequence} \tag{3}$$

It is important to note that not all sequences we obtained had the exact same length (mean length of nucleotide sequences: 29861.95; standard deviation: 28.40705). Hence, we performed this comparison only on the sub-sequence from the first nucleotide position to the maximum nucleotide position of the shorter of the two sequences a and b.

Results and discussion

Information-theoretic analysis

Figure 3 shows the graphs for the information token entropy and information token evenness. As one can see, these feature the aforementioned wave-like patterns, which means the system can be regarded *open* in the sense that over time the number of unique codon identifiers grows.

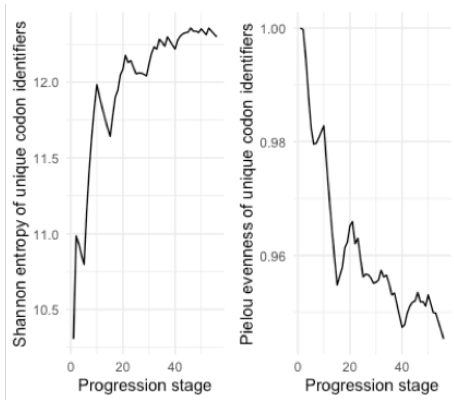

Figure 3. Information-theoretic assessment. Both entropy and evenness graphs feature wave-like patterns, indicating that the set of observed unique codon identifiers grows with increasing progression stages while the distribution of all observed unique codon identifiers remains rather even.

It is important to note that the evenness of the distribution of codon identifiers is relatively high, converging around a value of 0.95. This means that those codon identifiers that have been observed are quite evenly distributed.

These results allow the interpretation that (a) there is evolving variation in the genomic sequences we studied, but (b) there is no indication for the dominance of a particular variance.

Network analysis

The weighted network derived from the original TIC network features a modularity of 0.735 and a diameter of 6, which mainly reflects its small size and its particular structure that results from preserving temporal order in the sense that edges can only ever direct forward in time (i.e. to vertices that represent a later stage). However, with an average degree of 6.571 we find that there is certainly structure that differs from just simple step-by-step forward paths that lead to the consecutively next node. This latter result suggests that some codon paths branch and merge, emphasising that there is an evolving variation in the studied genomic sequences.

From the cluster analysis in combination with a visual inspection of the resulting graph (please refer to Figure 4 at the end of this paper as well as Table 9 in the appendix) we find the following patterns that are noteworthy:

- At the center of the network one cluster emerges (yellow, cluster 3) that has only few strong connections within the cluster but is densely connected to all other clusters. Most genome sequences in this cluster are from China (Wuhan, Shenzhen, Hangzhou, Zhejiang) dominated by 7 samples from the original outbreak region in Wuhan. But this cluster then also features three geographic outliers (France, Germany, Taiwan). This cluster has no temporal focus, it comprises samples that were collected early as well as later during the outbreak.

- The blue cluster in the northern region of the network visualisation (cluster 6) comprises three strongly similar genome sequences that stem consistently from the early stages of the outbreak in Wuhan.

- There is a kind of transition cluster (light blue, cluster 5) that starts at progression stage 41 (sample from Guangdong), then transitions via stage 42 to Guangzhou, followed by three samples from Foshan (stages $43 - 45$), before a final strong similarity has been detected to the nucleotide sequence in stage 48 (from Guangdong again).

- Out of the center emerges one cluster (dark red, cluster 1) that comprises almost exclusively a sequence of samples found outside of China (USA, Singapore, Australia, England, Finland)

We suggest that while there are some obvious pathways of high similarity due to temporal order of the samples taken, there are still very good indications for geographical-dependencies and transition phases that deserve further investigation. We want to draw particular attention in this respect to the transition point from cluster 3 to cluster 1 as well as cluster 3 and cluster 5. It will be particularly important to go back to qualitatively assess the human-to-human connections the respective subjects had from which these samples were taken, and the places they have been to.

Inter- and intra-cluster similarity

From our measurement of the intra-cluster similarity (see Table 1) we find that the genomic similarity at the level of nucleotides is the lowest in the central transition cluster 3, which is also the largest of all clusters. This supports the interpretation we made based on the network analysis that represented strength of codon identifier similarity, which was overall low for this cluster, and that the pathways in and out of this cluster may be promising candidates for deeper investigation.

Table 1. Intra-cluster similarities.

Cluster	Avg. similarity	SD	Min. similarity	Max. similarity	Cluster size
1	0.998	0.0025	0.9912	0.9999	10
2	0.9999	0.0001	0.9998	1	10
3	0.282	0.102	0.2507	1	18
4	0.9997	0.0003	0.9989	1	9
5	0.9999	0	0.9999	1	6
6	0.9999	0.0001	0.9998	0.9999	3

The inter-cluster analysis results show that clusters 2, 3 and 5 are reasonably distinct. However, clusters 1, 4 (light green in the network visualisation) and 6 feature sequences with high similarity at the nucleotide level, which contrasts their distinct clustering in the network. Looking at the samples in these three clusters suggests that clusters 4 and 6 are mainly representative of samples that were taken in the original outbreak region (Wuhan) and that the samples in cluster 1 could be from human subjects that were in direct contact with infected people in that region and basically exported the virus from there. It will again be important to deeper investigate the intersection points between the human subjects in these three clusters (geographically as well as in human contact).

Conclusion

In this paper we presented a novel way to model, visualise and analyse the temporal dynamics of an epidemic outbreak. We applied Transcendental Information Cascades to capture the recurrence and co-occurrence of unique codon identifiers over distinct and temporally ordered nucleotide sequences derived from genetic samples taken from different human subjects. Our case study was focused on the recent outbreak of the novel 2019-nCoV coronavirus. We find clusters and evolution pathways of virus genomes in the constructed Transcendental Information Cascade that should be investigated further for potential human-to-human transmission paths and genetic evolution of the 2019-nCoV coronavirus.

Our study was performed as a snapshot case study on a limited sample of virus data. While this was at the time of this study all data that was available, we will need to re-execute this analysis continuously as more data becomes available to check whether any of our results change. We also suggest that it will be important to perform a wider variety of analysis of the generated TIC data (e.g. testing of different network clustering

Table 2. Inter-cluster similarities.

Inter-cluster pair	Avg. similarity	SD	Min. similarity	Max. similarity
1-2	0.2593	0.0004	0.2581	0.2597
1-3	0.2668	0.0136	0.2507	0.2935
1-4	0.9988	0.0019	0.9932	1.0000
1-5	0.2584	0.0005	0.2568	0.2589
1-6	0.9989	0.0019	0.9940	1.0000
2-3	0.2657	0.0111	0.2521	0.2882
2-4	0.2595	0.0000	0.2594	0.2596
2-5	0.2843	0.0000	0.2843	0.2845
2-6	0.2595	0.0000	0.2594	0.2596
3-4	0.2670	0.0135	0.2526	0.2933
3-5	0.2669	0.0126	0.2521	0.2883
3-6	0.2670	0.0136	0.2526	0.2933
4-5	0.2585	0.0000	0.2585	0.2586
4-6	0.9998	0.0003	0.9990	1.0000
5-6	0.2585	0.0000	0.2585	0.2586

methods, application of dynamical systems analyses techniques) to get more profound insights. Our analysis also does not take measurement biases into account that may results from different techniques used by different laboratories for DNA extraction and analysis [6].

In summary, the approach presented in this paper is exploratory and hypotheses generating in nature. It does not confirm or reject any of the hypotheses we generated about the potential transmission and evolution pathways nor does it provide causal explanation for the patterns found. However, the purpose of our work is to provide an alternative view to the epidemic dynamics of the 2019-nCoV outbreak (and potential other outbreaks in the future) that is hidden to state-of-the-art methods, and that may trigger geneticists and epidemiologists to look at pathways in epidemic outbreaks they would not normally be aware of.

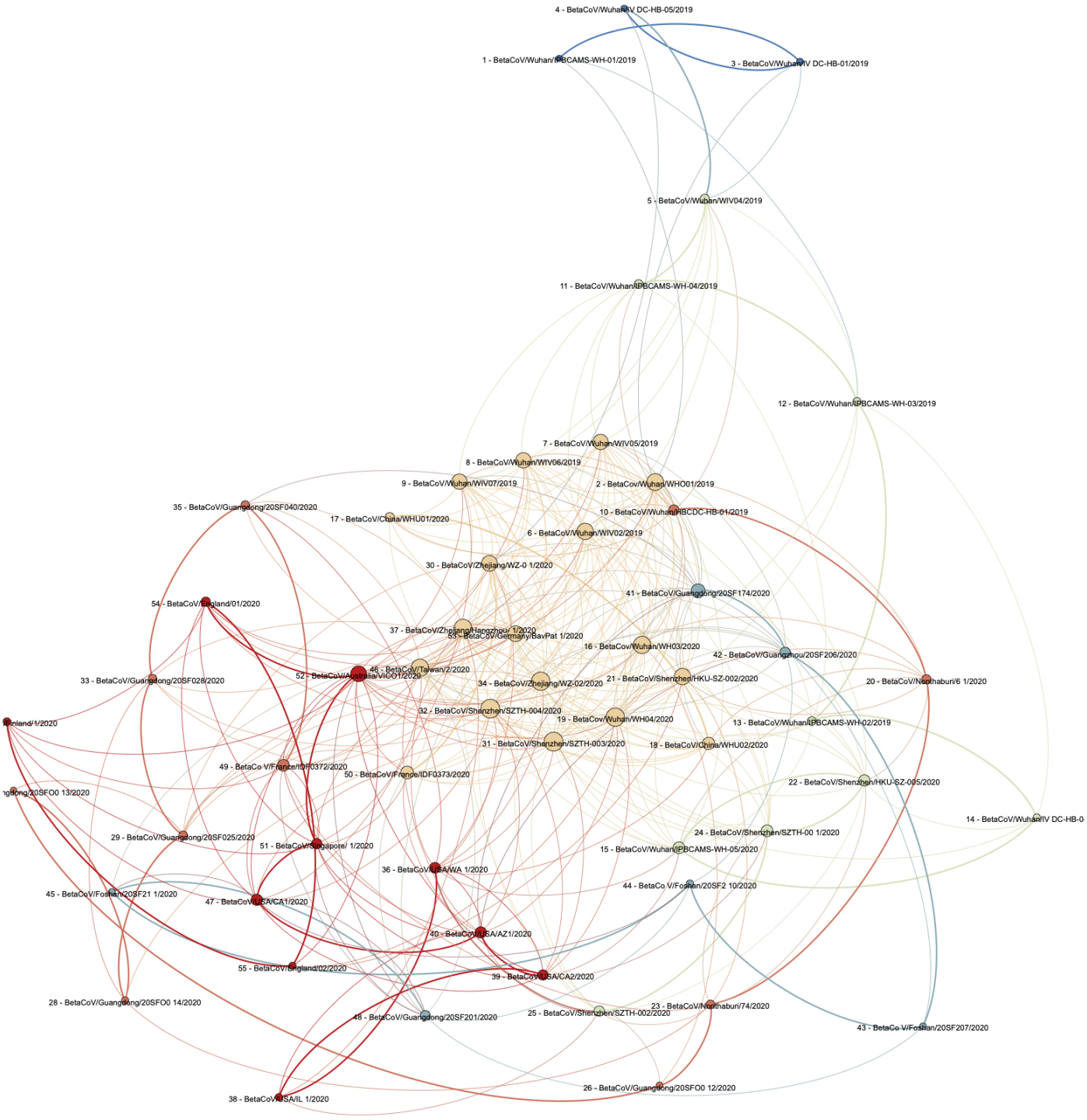

Figure 4. Weighted TIC network. From the original TIC network we generate a weighted network by collapsing all edges between the same vertices to unique edges weighted by the sum of the collapsed edges. The colours represent the six identified network clusters (dark red - cluster 1, orange - cluster 2, yellow - cluster 3, light green - cluster 4, light blue - cluster 5, blue - cluster 6). Node size represents the degree and edge thickness represents edge weight. Numeric indices before the genome identifiers represent the progression stage index of the particular vertex in the sequence of all ordered nucleotide sequences we analysed.

References

1. H.-J. Bandelt, P. Forster, and A. Röhl. Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, 16(1):37–48, 1999.

2. D. Benvenuto, M. Giovannetti, A. Ciccozzi, S. Spoto, S. Angeletti, and M. Ciccozzi. The 2019-new coronavirus epidemic: evidence for virus evolution. *Journal of Medical Virology*, 2020.

3. S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.

4. O. Cenciarelli, S. Pietropaoli, A. Malizia, M. Carestia, F. D'Amico, A. Sassolini, D. Di Giovanni, S. Rea, V. Gabbarini, A. Tamburrini, et al. Ebola virus disease 2013-2014 outbreak in west africa: an analysis of the epidemic spread and response. *International journal of microbiology*, 2015, 2015.

5. P. Cliften, P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton, J. Majors, R. Waterston, B. A. Cohen, and M. Johnston. Finding functional features in saccharomyces genomes by phylogenetic footprinting. *science*, 301(5629):71–76, 2003.

6. A. S. Devonshire, A. S. Whale, A. Gutteridge, G. Jones, S. Cowen, C. A. Foy, and J. F. Huggett. Towards standardisation of cell-free dna measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Analytical and bioanalytical chemistry*, 406(26):6499–6512, 2014.

7. N. Dong, X. Yang, L. Ye, K. Chen, E. W.-C. Chan, M. Yang, and S. Chen. Genomic and protein structure modelling analysis depicts the origin and infectivity of 2019-ncov, a new coronavirus which caused a pneumonia outbreak in wuhan, china. *bioRxiv*, 2020.

8. J. Faulkner, M. Schaller, J. H. Park, and L. A. Duncan. Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes & Intergroup Relations*, 7(4):333–353, 2004.

9. Y. Hu. Efficient, high-quality force-directed graph drawing. *Mathematica Journal*, 10(1):37–71, 2005.

10. M. Kellis, N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, 423(6937):241, 2003.

11. C. Li, X. Yu, and N. Helal. Similarity analysis of dna sequences based on codon usage. *Chemical physics letters*, 459(1-6):172–174, 2008.

12. T. Liu, J. Hu, M. Kang, L. Lin, H. Zhong, J. Xiao, G. He, T. Song, Q. Huang, Z. Rong, et al. Transmission dynamics of 2019 novel coronavirus (2019-ncov). *bioRxiv*, 2020.

13. M. Luczak-Roesch, K. O'Hara, J. D. Dinneen, and R. Tinati. What an entangled web we weave: An information-centric approach to time-evolving socio-technical systems. *Minds and Machines*, 28(4):709–733, 2018.

14. M. Luczak-Roesch, R. Tinati, M. Van Kleek, and N. Shadbolt. From coincidence to purposeful flow? properties of transcendental information cascades. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 633–638. IEEE, 2015.

15. I. Manolopoulou, L. Legarreta, B. C. Emerson, S. Brooks, and S. Tavaré. A bayesian approach to phylogeographic clustering. *Interface focus*, 1(6):909–921, 2011.

16. N. Masuda and P. Holme. Detecting sequences of system states in temporal networks. *Scientific reports*, 9(1):1–11, 2019.

17. G. R. D. S. Pagès H, Aboyoun P. *Biostrings: Efficient manipulation of biological strings*, 2019. R package version 2.54.0.

18. P. Pons and M. Latapy. Computing communities in large networks using random walks. In *International symposium on computer and information sciences*, pages 284–293. Springer, 2005.

19. M. A. Porter. Nonlinearity+ networks: A 2020 vision. *arXiv preprint arXiv:1911.03805*, 2019.

20. P. M. Sharp and W.-H. Li. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research*, 15(3):1281–1295, 1987.

21. R. D. Smith. Responding to global infectious disease outbreaks: lessons from sars on the role of risk perception, communication and management. *Social science & medicine*, 63(12):3113–3123, 2006.

Appendix

The software developed for this particular analysis is part of the open scientific software "R toolchain to construct and analyse Transcendental Information Cascades" that is available via `https://github.com/vuw-c2lab/transcendental-information-cascades/tree/master/input/2020-02-03-2019nCoV`. For reproduction purposes it is necessary to obtain the raw genomic data used in this study from the GISAID EpiFlu Database™.

For further reference we provide the raw results of some of our analysis. We particularly point to the tables that provide references for relationships between nucleotide sequences and the cluster membership of nucleotide sequences.

Table 3. Edge weight between distinct nucleotide sequences in the Transcendental Information Cascade we constructed.

Source sequence	Target sequence	Edge weight
11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	12 - BetaCoV/Wuhan/IPBCAMS-WH-03/2019	29885
3 - BetaCoV/Wuhan/IV DC-HB-01/2019	4 - BetaCoV/Wuhan/IV DC-HB-05/2019	29883
4 - BetaCoV/Wuhan/IV DC-HB-05/2019	5 - BetaCoV/Wuhan/WIV04/2019	29883
17 - BetaCoV/China/WHU01/2020	18 - BetaCoV/China/WHU02/2020	29879
14 - BetaCoV/Wuhan/IV DC-HB-04/2020	15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	29872
12 - BetaCoV/Wuhan/IPBCAMS-WH-03/2019	13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	29870
13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	14 - BetaCoV/Wuhan/IV DC-HB-04/2020	29867
39 - BetaCoV/USA/CA2/2020	40 - BetaCoV/USA/AZ1/2020	29862
38 - BetaCoV/USA/IL 1/2020	39 - BetaCoV/USA/CA2/2020	29850
26 - BetaCoV/Guangdong/20SFO0 12/2020	27 - BetaCoV/Guangdong/20SFO0 13/2020	29843
27 - BetaCoV/Guangdong/20SFO0 13/2020	28 - BetaCoV/Guangdong/20SFO0 14/2020	29831
28 - BetaCoV/Guangdong/20SFO0 14/2020	29 - BetaCoV/Guangdong/20SF025/2020	29831
54 - BetaCoV/England/01/2020	55 - BetaCoV/England/02/2020	29830
43 - BetaCo V/Foshan/20SF207/2020	44 - BetaCo V/Foshan/20SF2 10/2020	29828
42 - BetaCoV/Guangzhou/20SF206/2020	43 - BetaCo V/Foshan/20SF207/2020	29825
24 - BetaCoV/Shenzhen/SZTH-00 1/2020	25 - BetaCoV/Shenzhen/SZTH-002/2020	29813
44 - BetaCo V/Foshan/20SF2 10/2020	45 - BetaCoV/Foshan/20SF21 1/2020	29801
41 - BetaCoV/Guangdong/20SF174/2020	42 - BetaCoV/Guangzhou/20SF206/2020	29797
51 - BetaCoV/Singapore/ 1/2020	52 - BetaCoV/Australia/VICO1/2020	29736
55 - BetaCoV/England/02/2020	56 - BetaCoV/Finland/1/2020	29643
1 - BetaCoV/Wuhan/IPBCAMS-WH-01/2019	3 - BetaCoV/Wuhan/IV DC-HB-01/2019	29346
23 - BetaCoV/Nonthaburi/74/2020	26 - BetaCoV/Guangdong/20SFO0 12/2020	29290
22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	24 - BetaCoV/Shenzhen/SZTH-00 1/2020	29263
36 - BetaCoV/USA/WA 1/2020	38 - BetaCoV/USA/IL 1/2020	29091
33 - BetaCoV/Guangdong/20SF028/2020	35 - BetaCoV/Guangdong/20SF040/2020	29072
52 - BetaCoV/Australia/VICO1/2020	54 - BetaCoV/England/01/2020	28883
45 - BetaCoV/Foshan/20SF21 1/2020	48 - BetaCoV/Guangdong/20SF201/2020	28741
40 - BetaCoV/USA/AZ1/2020	47 - BetaCoV/USA/CA1/2020	28651
20 - BetaCoV/Nonthaburi/6 1/2020	23 - BetaCoV/Nonthaburi/74/2020	28587
29 - BetaCoV/Guangdong/20SF025/2020	33 - BetaCoV/Guangdong/20SF028/2020	28478
47 - BetaCoV/USA/CA1/2020	51 - BetaCoV/Singapore/ 1/2020	28287
35 - BetaCoV/Guangdong/20SF040/2020	49 - BetaCo V/France/IDF0372/2020	27834
25 - BetaCoV/Shenzhen/SZTH-002/2020	36 - BetaCoV/USA/WA 1/2020	27734
10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	20 - BetaCoV/Nonthaburi/6 1/2020	27513
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	27049
5 - BetaCoV/Wuhan/WIV04/2019	11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	26905
30 - BetaCoV/Zhejiang/WZ-0 1/2020	50 - BetaCoV/France/IDF0373/2020	26605
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	46 - BetaCoV/Taiwan/2/2020	20289
31 - BetaCoV/Shenzhen/SZTH-003/2020	32 - BetaCoV/Shenzhen/SZTH-004/2020	20102
18 - BetaCoV/China/WHU02/2020	46 - BetaCoV/Taiwan/2/2020	3045
5 - BetaCoV/Wuhan/WIV04/2019	6 - BetaCoV/Wuhan/WIV02/2019	848
52 - BetaCoV/Australia/VICO1/2020	53 - BetaCoV/Germany/BavPat 1/2020	839
53 - BetaCoV/Germany/BavPat 1/2020	54 - BetaCoV/England/01/2020	837
10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	16 - BetaCov/Wuhan/WH03/2020	812
16 - BetaCov/Wuhan/WH03/2020	20 - BetaCoV/Nonthaburi/6 1/2020	806
6 - BetaCoV/Wuhan/WIV02/2019	11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	771
33 - BetaCoV/Guangdong/20SF028/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	771
34 - BetaCoV/Zhejiang/WZ-02/2020	35 - BetaCoV/Guangdong/20SF040/2020	771
36 - BetaCoV/USA/WA 1/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	762
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	38 - BetaCoV/USA/IL 1/2020	762
8 - BetaCoV/Wuhan/WIV06/2019	10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	755

Table 4. Edge weight between distinct nucleotide sequences in the Transcendental Information Cascade we constructed (contd.).

Source sequence	Target sequence	Edge weight
19 - BetaCov/Wuhan/WH04/2020	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	748
32 - BetaCoV/Shenzhen/SZTH-004/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	747
48 - BetaCoV/Guangdong/20SF201/2020	49 - BetaCo V/France/IDF0372/2020	742
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	19 - BetaCov/Wuhan/WH04/2020	731
20 - BetaCoV/Nonthaburi/6 1/2020	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	727
6 - BetaCoV/Wuhan/WIV02/2019	7 - BetaCoV/Wuhan/WIV05/2019	725
35 - BetaCoV/Guangdong/20SF040/2020	41 - BetaCoV/Guangdong/20SF174/2020	717
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	23 - BetaCoV/Nonthaburi/74/2020	715
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	41 - BetaCoV/Guangdong/20SF174/2020	715
9 - BetaCoV/Wuhan/WIV07/2019	17 - BetaCoV/China/WHU01/2020	713
5 - BetaCoV/Wuhan/WIV04/2019	7 - BetaCoV/Wuhan/WIV05/2019	711
34 - BetaCoV/Zhejiang/WZ-02/2020	41 - BetaCoV/Guangdong/20SF174/2020	709
7 - BetaCoV/Wuhan/WIV05/2019	19 - BetaCov/Wuhan/WH04/2020	703
7 - BetaCoV/Wuhan/WIV05/2019	11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	698
2 - BetaCov/Wuhan/WHO01/2019	9 - BetaCoV/Wuhan/WIV07/2019	686
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	685
46 - BetaCoV/Taiwan/2/2020	47 - BetaCoV/USA/CA1/2020	685
46 - BetaCoV/Taiwan/2/2020	53 - BetaCoV/Germany/BavPat 1/2020	678
40 - BetaCoV/USA/AZ1/2020	46 - BetaCoV/Taiwan/2/2020	676
16 - BetaCov/Wuhan/WH03/2020	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	674
8 - BetaCoV/Wuhan/WIV06/2019	16 - BetaCov/Wuhan/WH03/2020	664
29 - BetaCoV/Guangdong/20SF025/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	656
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	654
7 - BetaCoV/Wuhan/WIV05/2019	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	653
9 - BetaCoV/Wuhan/WIV07/2019	30 - BetaCoV/Zhejiang/WZ-0 1/2020	651
19 - BetaCov/Wuhan/WH04/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	644
18 - BetaCoV/China/WHU02/2020	30 - BetaCoV/Zhejiang/WZ-0 1/2020	641
8 - BetaCoV/Wuhan/WIV06/2019	41 - BetaCoV/Guangdong/20SF174/2020	629
32 - BetaCoV/Shenzhen/SZTH-004/2020	33 - BetaCoV/Guangdong/20SF028/2020	628
32 - BetaCoV/Shenzhen/SZTH-004/2020	41 - BetaCoV/Guangdong/20SF174/2020	617
2 - BetaCov/Wuhan/WHO01/2019	17 - BetaCoV/China/WHU01/2020	613
16 - BetaCov/Wuhan/WH03/2020	41 - BetaCoV/Guangdong/20SF174/2020	607
19 - BetaCov/Wuhan/WH04/2020	53 - BetaCoV/Germany/BavPat 1/2020	605
16 - BetaCov/Wuhan/WH03/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	603
7 - BetaCoV/Wuhan/WIV05/2019	53 - BetaCoV/Germany/BavPat 1/2020	600
16 - BetaCov/Wuhan/WH03/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	596
8 - BetaCoV/Wuhan/WIV06/2019	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	594
6 - BetaCoV/Wuhan/WIV02/2019	19 - BetaCov/Wuhan/WH04/2020	587
8 - BetaCoV/Wuhan/WIV06/2019	34 - BetaCoV/Zhejiang/WZ-02/2020	587
2 - BetaCov/Wuhan/WHO01/2019	30 - BetaCoV/Zhejiang/WZ-0 1/2020	585
8 - BetaCoV/Wuhan/WIV06/2019	31 - BetaCoV/Shenzhen/SZTH-003/2020	584
46 - BetaCoV/Taiwan/2/2020	49 - BetaCo V/France/IDF0372/2020	584
8 - BetaCoV/Wuhan/WIV06/2019	9 - BetaCoV/Wuhan/WIV07/2019	579
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	17 - BetaCoV/China/WHU01/2020	571
19 - BetaCov/Wuhan/WH04/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	550
50 - BetaCoV/France/IDF0373/2020	53 - BetaCoV/Germany/BavPat 1/2020	547
29 - BetaCoV/Guangdong/20SF025/2020	30 - BetaCoV/Zhejiang/WZ-0 1/2020	546
49 - BetaCo V/France/IDF0372/2020	50 - BetaCoV/France/IDF0373/2020	546
22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	23 - BetaCoV/Nonthaburi/74/2020	545
10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	544
23 - BetaCoV/Nonthaburi/74/2020	24 - BetaCoV/Shenzhen/SZTH-00 1/2020	544
25 - BetaCoV/Shenzhen/SZTH-002/2020	26 - BetaCoV/Guangdong/20SFO0 12/2020	544
35 - BetaCoV/Guangdong/20SF040/2020	36 - BetaCoV/USA/WA 1/2020	544
19 - BetaCov/Wuhan/WH04/2020	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	539
6 - BetaCoV/Wuhan/WIV02/2019	10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	538
35 - BetaCoV/Guangdong/20SF040/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	536
2 - BetaCov/Wuhan/WHO01/2019	3 - BetaCoV/Wuhan/IV DC-HB-01/2019	535
45 - BetaCoV/Foshan/20SF21 1/2020	46 - BetaCoV/Taiwan/2/2020	535
1 - BetaCoV/Wuhan/IPBCAMS-WH-01/2019	2 - BetaCov/Wuhan/WHO01/2019	534
18 - BetaCoV/China/WHU02/2020	20 - BetaCoV/Nonthaburi/6 1/2020	534
40 - BetaCoV/USA/AZ1/2020	41 - BetaCoV/Guangdong/20SF174/2020	533
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	16 - BetaCov/Wuhan/WH03/2020	532
20 - BetaCoV/Nonthaburi/6 1/2020	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	532
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	532
47 - BetaCoV/USA/CA1/2020	48 - BetaCoV/Guangdong/20SF201/2020	532
2 - BetaCov/Wuhan/WHO01/2019	6 - BetaCoV/Wuhan/WIV02/2019	530
47 - BetaCoV/USA/CA1/2020	49 - BetaCo V/France/IDF0372/2020	530
6 - BetaCoV/Wuhan/WIV02/2019	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	529
49 - BetaCo V/France/IDF0372/2020	51 - BetaCoV/Singapore/ 1/2020	529
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	41 - BetaCoV/Guangdong/20SF174/2020	527
46 - BetaCoV/Taiwan/2/2020	48 - BetaCoV/Guangdong/20SF201/2020	526
18 - BetaCoV/China/WHU02/2020	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	525
18 - BetaCoV/China/WHU02/2020	19 - BetaCov/Wuhan/WH04/2020	524
45 - BetaCoV/Foshan/20SF21 1/2020	47 - BetaCoV/USA/CA1/2020	522
49 - BetaCo V/France/IDF0372/2020	53 - BetaCoV/Germany/BavPat 1/2020	521
30 - BetaCoV/Zhejiang/WZ-0 1/2020	33 - BetaCoV/Guangdong/20SF028/2020	518
2 - BetaCov/Wuhan/WHO01/2019	8 - BetaCoV/Wuhan/WIV06/2019	516
10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	17 - BetaCoV/China/WHU01/2020	511
48 - BetaCoV/Guangdong/20SF201/2020	51 - BetaCoV/Singapore/ 1/2020	510
18 - BetaCoV/China/WHU02/2020	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	503
34 - BetaCoV/Zhejiang/WZ-02/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	502
16 - BetaCov/Wuhan/WH03/2020	30 - BetaCoV/Zhejiang/WZ-0 1/2020	501

Table 5. Edge weight between distinct nucleotide sequences in the Transcendental Information Cascade we constructed (contd.).

Source sequence	Target sequence	Edge weight
32 - BetaCoV/Shenzhen/SZTH-004/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	499
5 - BetaCoV/Wuhan/WIV04/2019	8 - BetaCoV/Wuhan/WIV06/2019	497
16 - BetaCov/Wuhan/WH03/2020	17 - BetaCoV/China/WHU01/2020	497
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	20 - BetaCoV/Nonthaburi/6 1/2020	497
25 - BetaCoV/Shenzhen/SZTH-002/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	496
50 - BetaCoV/France/IDF0373/2020	51 - BetaCoV/Singapore/ 1/2020	496
32 - BetaCoV/Shenzhen/SZTH-004/2020	36 - BetaCoV/USA/WA 1/2020	494
19 - BetaCov/Wuhan/WH04/2020	20 - BetaCoV/Nonthaburi/6 1/2020	493
6 - BetaCoV/Wuhan/WIV02/2019	9 - BetaCoV/Wuhan/WIV07/2019	491
8 - BetaCoV/Wuhan/WIV06/2019	11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	488
30 - BetaCoV/Zhejiang/WZ-0 1/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	488
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	485
2 - BetaCov/Wuhan/WHO01/2019	10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	484
34 - BetaCoV/Zhejiang/WZ-02/2020	36 - BetaCoV/USA/WA 1/2020	484
6 - BetaCoV/Wuhan/WIV02/2019	53 - BetaCoV/Germany/BavPat 1/2020	484
7 - BetaCoV/Wuhan/WIV05/2019	16 - BetaCov/Wuhan/WH03/2020	483
5 - BetaCoV/Wuhan/WIV04/2019	10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	482
9 - BetaCoV/Wuhan/WIV07/2019	11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	482
7 - BetaCoV/Wuhan/WIV05/2019	17 - BetaCoV/China/WHU01/2020	482
25 - BetaCoV/Shenzhen/SZTH-002/2020	30 - BetaCoV/Zhejiang/WZ-0 1/2020	482
16 - BetaCov/Wuhan/WH03/2020	19 - BetaCov/Wuhan/WH04/2020	481
18 - BetaCoV/China/WHU02/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	481
19 - BetaCov/Wuhan/WH04/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	481
48 - BetaCoV/Guangdong/20SF201/2020	53 - BetaCoV/Germany/BavPat 1/2020	481
6 - BetaCoV/Wuhan/WIV02/2019	8 - BetaCoV/Wuhan/WIV06/2019	480
32 - BetaCoV/Shenzhen/SZTH-004/2020	50 - BetaCoV/France/IDF0373/2020	479
48 - BetaCoV/Guangdong/20SF201/2020	50 - BetaCoV/France/IDF0373/2020	479
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	53 - BetaCoV/Germany/BavPat 1/2020	479
16 - BetaCov/Wuhan/WH03/2020	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	478
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	30 - BetaCoV/Zhejiang/WZ-0 1/2020	477
18 - BetaCoV/China/WHU02/2020	41 - BetaCoV/Guangdong/20SF174/2020	477
7 - BetaCoV/Wuhan/WIV05/2019	8 - BetaCoV/Wuhan/WIV06/2019	472
9 - BetaCoV/Wuhan/WIV07/2019	10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	472
6 - BetaCoV/Wuhan/WIV02/2019	17 - BetaCoV/China/WHU01/2020	472
47 - BetaCoV/USA/CA1/2020	50 - BetaCoV/France/IDF0373/2020	472
2 - BetaCov/Wuhan/WHO01/2019	7 - BetaCoV/Wuhan/WIV05/2019	469
9 - BetaCoV/Wuhan/WIV07/2019	16 - BetaCov/Wuhan/WH03/2020	468
19 - BetaCov/Wuhan/WH04/2020	30 - BetaCoV/Zhejiang/WZ-0 1/2020	467
6 - BetaCoV/Wuhan/WIV02/2019	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	466
30 - BetaCoV/Zhejiang/WZ-0 1/2020	36 - BetaCoV/USA/WA 1/2020	466
25 - BetaCoV/Shenzhen/SZTH-002/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	465
10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	19 - BetaCov/Wuhan/WH04/2020	463
2 - BetaCov/Wuhan/WHO01/2019	19 - BetaCov/Wuhan/WH04/2020	462
46 - BetaCoV/Taiwan/2/2020	50 - BetaCoV/France/IDF0373/2020	461
7 - BetaCoV/Wuhan/WIV05/2019	9 - BetaCoV/Wuhan/WIV07/2019	460
30 - BetaCoV/Zhejiang/WZ-0 1/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	460
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	459
7 - BetaCoV/Wuhan/WIV05/2019	10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	458
30 - BetaCoV/Zhejiang/WZ-0 1/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	453
19 - BetaCov/Wuhan/WH04/2020	41 - BetaCoV/Guangdong/20SF174/2020	453
34 - BetaCoV/Zhejiang/WZ-02/2020	50 - BetaCoV/France/IDF0373/2020	452
9 - BetaCoV/Wuhan/WIV07/2019	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	448
9 - BetaCoV/Wuhan/WIV07/2019	31 - BetaCoV/Shenzhen/SZTH-003/2020	446
5 - BetaCoV/Wuhan/WIV04/2019	9 - BetaCoV/Wuhan/WIV07/2019	445
6 - BetaCoV/Wuhan/WIV02/2019	31 - BetaCoV/Shenzhen/SZTH-003/2020	445
8 - BetaCoV/Wuhan/WIV06/2019	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	445
32 - BetaCoV/Shenzhen/SZTH-004/2020	53 - BetaCoV/Germany/BavPat 1/2020	445
7 - BetaCoV/Wuhan/WIV05/2019	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	444
8 - BetaCoV/Wuhan/WIV06/2019	17 - BetaCoV/China/WHU01/2020	443
18 - BetaCoV/China/WHU02/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	443
16 - BetaCov/Wuhan/WH03/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	441
30 - BetaCoV/Zhejiang/WZ-0 1/2020	41 - BetaCoV/Guangdong/20SF174/2020	441
34 - BetaCoV/Zhejiang/WZ-02/2020	53 - BetaCoV/Germany/BavPat 1/2020	438
2 - BetaCov/Wuhan/WHO01/2019	16 - BetaCov/Wuhan/WH03/2020	434
9 - BetaCoV/Wuhan/WIV07/2019	53 - BetaCoV/Germany/BavPat 1/2020	434
9 - BetaCoV/Wuhan/WIV07/2019	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	430
6 - BetaCoV/Wuhan/WIV02/2019	16 - BetaCov/Wuhan/WH03/2020	428
2 - BetaCov/Wuhan/WHO01/2019	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	428
9 - BetaCoV/Wuhan/WIV07/2019	41 - BetaCoV/Guangdong/20SF174/2020	428
2 - BetaCov/Wuhan/WHO01/2019	31 - BetaCoV/Shenzhen/SZTH-003/2020	425
9 - BetaCoV/Wuhan/WIV07/2019	19 - BetaCov/Wuhan/WH04/2020	424
8 - BetaCoV/Wuhan/WIV06/2019	19 - BetaCov/Wuhan/WH04/2020	419
2 - BetaCov/Wuhan/WHO01/2019	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	418
6 - BetaCoV/Wuhan/WIV02/2019	30 - BetaCoV/Zhejiang/WZ-0 1/2020	412
18 - BetaCoV/China/WHU02/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	412
16 - BetaCov/Wuhan/WH03/2020	53 - BetaCoV/Germany/BavPat 1/2020	410
7 - BetaCoV/Wuhan/WIV05/2019	34 - BetaCoV/Zhejiang/WZ-02/2020	409
8 - BetaCoV/Wuhan/WIV06/2019	53 - BetaCoV/Germany/BavPat 1/2020	409
8 - BetaCoV/Wuhan/WIV06/2019	30 - BetaCoV/Zhejiang/WZ-0 1/2020	408
7 - BetaCoV/Wuhan/WIV05/2019	30 - BetaCoV/Zhejiang/WZ-0 1/2020	407
18 - BetaCoV/China/WHU02/2020	53 - BetaCoV/Germany/BavPat 1/2020	403
7 - BetaCoV/Wuhan/WIV05/2019	31 - BetaCoV/Shenzhen/SZTH-003/2020	400
7 - BetaCoV/Wuhan/WIV05/2019	41 - BetaCoV/Guangdong/20SF174/2020	400
9 - BetaCoV/Wuhan/WIV07/2019	34 - BetaCoV/Zhejiang/WZ-02/2020	399
2 - BetaCov/Wuhan/WHO01/2019	34 - BetaCoV/Zhejiang/WZ-02/2020	395
6 - BetaCoV/Wuhan/WIV02/2019	34 - BetaCoV/Zhejiang/WZ-02/2020	395
6 - BetaCoV/Wuhan/WIV02/2019	41 - BetaCoV/Guangdong/20SF174/2020	395
2 - BetaCov/Wuhan/WHO01/2019	53 - BetaCoV/Germany/BavPat 1/2020	338
2 - BetaCov/Wuhan/WHO01/2019	41 - BetaCoV/Guangdong/20SF174/2020	336
31 - BetaCoV/Shenzhen/SZTH-003/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	256

Table 6. Edge weight between distinct nucleotide sequences in the Transcendental Information Cascade we constructed (contd.).

Source sequence	Target sequence	Edge weight
18 - BetaCoV/China/WHU02/2020	32 - BetaCoV/Shenzhen/SZTH-004/2020	239
30 - BetaCoV/Zhejiang/WZ-0 1/2020	32 - BetaCoV/Shenzhen/SZTH-004/2020	220
31 - BetaCoV/Shenzhen/SZTH-003/2020	33 - BetaCoV/Guangdong/20SF028/2020	207
35 - BetaCoV/Guangdong/20SF040/2020	46 - BetaCoV/Taiwan/2/2020	204
2 - BetaCov/Wuhan/WHO01/2019	32 - BetaCoV/Shenzhen/SZTH-004/2020	201
9 - BetaCoV/Wuhan/WIV07/2019	32 - BetaCoV/Shenzhen/SZTH-004/2020	196
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	53 - BetaCoV/Germany/BavPat 1/2020	189
31 - BetaCoV/Shenzhen/SZTH-003/2020	46 - BetaCoV/Taiwan/2/2020	188
31 - BetaCoV/Shenzhen/SZTH-003/2020	41 - BetaCoV/Guangdong/20SF174/2020	185
34 - BetaCoV/Zhejiang/WZ-02/2020	46 - BetaCoV/Taiwan/2/2020	182
7 - BetaCoV/Wuhan/WIV05/2019	32 - BetaCoV/Shenzhen/SZTH-004/2020	169
25 - BetaCoV/Shenzhen/SZTH-002/2020	32 - BetaCoV/Shenzhen/SZTH-004/2020	162
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	32 - BetaCoV/Shenzhen/SZTH-004/2020	158
31 - BetaCoV/Shenzhen/SZTH-003/2020	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	158
16 - BetaCov/Wuhan/WH03/2020	32 - BetaCoV/Shenzhen/SZTH-004/2020	157
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	50 - BetaCoV/France/IDF0373/2020	157
31 - BetaCoV/Shenzhen/SZTH-003/2020	53 - BetaCoV/Germany/BavPat 1/2020	156
8 - BetaCoV/Wuhan/WIV06/2019	32 - BetaCoV/Shenzhen/SZTH-004/2020	155
32 - BetaCoV/Shenzhen/SZTH-004/2020	46 - BetaCoV/Taiwan/2/2020	154
31 - BetaCoV/Shenzhen/SZTH-003/2020	36 - BetaCoV/USA/WA 1/2020	152
29 - BetaCoV/Guangdong/20SF025/2020	32 - BetaCoV/Shenzhen/SZTH-004/2020	151
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	46 - BetaCoV/Taiwan/2/2020	149
30 - BetaCoV/Zhejiang/WZ-0 1/2020	46 - BetaCoV/Taiwan/2/2020	147
19 - BetaCov/Wuhan/WH04/2020	32 - BetaCoV/Shenzhen/SZTH-004/2020	146
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	49 - BetaCo V/France/IDF0372/2020	145
31 - BetaCoV/Shenzhen/SZTH-003/2020	50 - BetaCoV/France/IDF0373/2020	145
6 - BetaCoV/Wuhan/WIV02/2019	32 - BetaCoV/Shenzhen/SZTH-004/2020	129
19 - BetaCov/Wuhan/WH04/2020	46 - BetaCoV/Taiwan/2/2020	129
51 - BetaCoV/Singapore/ 1/2020	54 - BetaCoV/England/01/2020	101
8 - BetaCoV/Wuhan/WIV06/2019	46 - BetaCoV/Taiwan/2/2020	97
16 - BetaCov/Wuhan/WH03/2020	46 - BetaCoV/Taiwan/2/2020	85
22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	25 - BetaCoV/Shenzhen/SZTH-002/2020	75
6 - BetaCoV/Wuhan/WIV02/2019	46 - BetaCoV/Taiwan/2/2020	63
2 - BetaCov/Wuhan/WHO01/2019	46 - BetaCoV/Taiwan/2/2020	61
9 - BetaCoV/Wuhan/WIV07/2019	46 - BetaCoV/Taiwan/2/2020	61
7 - BetaCoV/Wuhan/WIV05/2019	46 - BetaCoV/Taiwan/2/2020	51
44 - BetaCo V/Foshan/20SF2 10/2020	48 - BetaCoV/Guangdong/20SF201/2020	35
12 - BetaCoV/Wuhan/IPBCAMS-WH-03/2019	14 - BetaCoV/Wuhan/IV DC-HB-04/2020	21
36 - BetaCoV/USA/WA 1/2020	39 - BetaCoV/USA/CA2/2020	18
47 - BetaCoV/USA/CA1/2020	56 - BetaCoV/Finland/1/2020	13
27 - BetaCoV/Guangdong/20SFO0 13/2020	29 - BetaCoV/Guangdong/20SF025/2020	12
40 - BetaCoV/USA/AZ1/2020	51 - BetaCoV/Singapore/ 1/2020	12
47 - BetaCoV/USA/CA1/2020	52 - BetaCoV/Australia/VICO1/2020	12
52 - BetaCoV/Australia/VICO1/2020	56 - BetaCoV/Finland/1/2020	12
14 - BetaCoV/Wuhan/IV DC-HB-04/2020	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	11
32 - BetaCoV/Shenzhen/SZTH-004/2020	52 - BetaCoV/Australia/VICO1/2020	10
23 - BetaCoV/Nonthaburi/74/2020	28 - BetaCoV/Guangdong/20SFO0 14/2020	9
28 - BetaCoV/Guangdong/20SFO0 14/2020	33 - BetaCoV/Guangdong/20SF028/2020	9
19 - BetaCov/Wuhan/WH04/2020	52 - BetaCoV/Australia/VICO1/2020	9
1 - BetaCoV/Wuhan/IPBCAMS-WH-01/2019	12 - BetaCoV/Wuhan/IPBCAMS-WH-03/2019	8
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	24 - BetaCoV/Shenzhen/SZTH-00 1/2020	8
14 - BetaCoV/Wuhan/IV DC-HB-04/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	7
23 - BetaCoV/Nonthaburi/74/2020	49 - BetaCo V/France/IDF0372/2020	7
3 - BetaCoV/Wuhan/IV DC-HB-01/2019	5 - BetaCoV/Wuhan/WIV04/2019	6
13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	6
36 - BetaCoV/USA/WA 1/2020	40 - BetaCoV/USA/AZ1/2020	6
38 - BetaCoV/USA/IL 1/2020	40 - BetaCoV/USA/AZ1/2020	6
41 - BetaCoV/Guangdong/20SF174/2020	43 - BetaCo V/Foshan/20SF207/2020	6
42 - BetaCoV/Guangzhou/20SF206/2020	44 - BetaCo V/Foshan/20SF2 10/2020	6
39 - BetaCoV/USA/CA2/2020	47 - BetaCoV/USA/CA1/2020	6
39 - BetaCoV/USA/CA2/2020	51 - BetaCoV/Singapore/ 1/2020	6
47 - BetaCoV/USA/CA1/2020	54 - BetaCoV/England/01/2020	6

Table 7. Edge weight between distinct nucleotide sequences in the Transcendental Information Cascade we constructed (contd.).

Source sequence	Target sequence	Edge weight
32 - BetaCoV/Shenzhen/SZTH-004/2020	42 - BetaCoV/Guangzhou/20SF206/2020	5
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	52 - BetaCoV/Australia/VICO1/2020	4
39 - BetaCoV/USA/CA2/2020	52 - BetaCoV/Australia/VICO1/2020	4
49 - BetaCo V/France/IDF0372/2020	52 - BetaCoV/Australia/VICO1/2020	4
11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	3
19 - BetaCov/Wuhan/WH04/2020	24 - BetaCoV/Shenzhen/SZTH-00 1/2020	3
24 - BetaCoV/Shenzhen/SZTH-00 1/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	3
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	36 - BetaCoV/USA/WA 1/2020	3
25 - BetaCoV/Shenzhen/SZTH-002/2020	38 - BetaCoV/USA/IL 1/2020	3
24 - BetaCoV/Shenzhen/SZTH-00 1/2020	39 - BetaCoV/USA/CA2/2020	3
32 - BetaCoV/Shenzhen/SZTH-004/2020	39 - BetaCoV/USA/CA2/2020	3
25 - BetaCoV/Shenzhen/SZTH-002/2020	40 - BetaCoV/USA/AZ1/2020	3
43 - BetaCo V/Foshan/20SF207/2020	48 - BetaCoV/Guangdong/20SF201/2020	3
29 - BetaCoV/Guangdong/20SF025/2020	49 - BetaCo V/France/IDF0372/2020	3
2 - BetaCov/Wuhan/WHO01/2019	52 - BetaCoV/Australia/VICO1/2020	3
7 - BetaCoV/Wuhan/WIV05/2019	52 - BetaCoV/Australia/VICO1/2020	3
9 - BetaCoV/Wuhan/WIV07/2019	52 - BetaCoV/Australia/VICO1/2020	3
31 - BetaCoV/Shenzhen/SZTH-003/2020	52 - BetaCoV/Australia/VICO1/2020	3
48 - BetaCoV/Guangdong/20SF201/2020	52 - BetaCoV/Australia/VICO1/2020	3
52 - BetaCoV/Australia/VICO1/2020	55 - BetaCoV/England/02/2020	3
54 - BetaCoV/England/01/2020	56 - BetaCoV/Finland/1/2020	3
2 - BetaCov/Wuhan/WHO01/2019	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	2
16 - BetaCov/Wuhan/WH03/2020	24 - BetaCoV/Shenzhen/SZTH-00 1/2020	2
12 - BetaCoV/Wuhan/IPBCAMS-WH-03/2019	31 - BetaCoV/Shenzhen/SZTH-003/2020	2
22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	2
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	39 - BetaCoV/USA/CA2/2020	2
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	42 - BetaCoV/Guangzhou/20SF206/2020	2
2 - BetaCov/Wuhan/WHO01/2019	50 - BetaCoV/France/IDF0373/2020	2
40 - BetaCoV/USA/AZ1/2020	50 - BetaCoV/France/IDF0373/2020	2
8 - BetaCoV/Wuhan/WIV06/2019	52 - BetaCoV/Australia/VICO1/2020	2
34 - BetaCoV/Zhejiang/WZ-02/2020	52 - BetaCoV/Australia/VICO1/2020	2
46 - BetaCoV/Taiwan/2/2020	52 - BetaCoV/Australia/VICO1/2020	2
50 - BetaCoV/France/IDF0373/2020	52 - BetaCoV/Australia/VICO1/2020	2
51 - BetaCoV/Singapore/ 1/2020	56 - BetaCoV/Finland/1/2020	2
5 - BetaCoV/Wuhan/WIV04/2019	12 - BetaCoV/Wuhan/IPBCAMS-WH-03/2019	1
8 - BetaCoV/Wuhan/WIV06/2019	13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	1
2 - BetaCov/Wuhan/WHO01/2019	15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	1
2 - BetaCov/Wuhan/WHO01/2019	20 - BetaCoV/Nonthaburi/6 1/2020	1
13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	1
6 - BetaCoV/Wuhan/WIV02/2019	22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	1
6 - BetaCoV/Wuhan/WIV02/2019	24 - BetaCoV/Shenzhen/SZTH-00 1/2020	1
18 - BetaCoV/China/WHU02/2020	24 - BetaCoV/Shenzhen/SZTH-00 1/2020	1
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	24 - BetaCoV/Shenzhen/SZTH-00 1/2020	1
23 - BetaCoV/Nonthaburi/74/2020	25 - BetaCoV/Shenzhen/SZTH-002/2020	1
16 - BetaCov/Wuhan/WH03/2020	26 - BetaCoV/Guangdong/20SFO0 12/2020	1
4 - BetaCoV/Wuhan/IV DC-HB-05/2019	30 - BetaCoV/Zhejiang/WZ-0 1/2020	1
24 - BetaCoV/Shenzhen/SZTH-00 1/2020	30 - BetaCoV/Zhejiang/WZ-0 1/2020	1
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	1
28 - BetaCoV/Guangdong/20SFO0 14/2020	31 - BetaCoV/Shenzhen/SZTH-003/2020	1
12 - BetaCoV/Wuhan/IPBCAMS-WH-03/2019	32 - BetaCoV/Shenzhen/SZTH-004/2020	1
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	1
24 - BetaCoV/Shenzhen/SZTH-00 1/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	1
29 - BetaCoV/Guangdong/20SF025/2020	34 - BetaCoV/Zhejiang/WZ-02/2020	1
18 - BetaCoV/China/WHU02/2020	36 - BetaCoV/USA/WA 1/2020	1
13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	1
19 - BetaCov/Wuhan/WH04/2020	39 - BetaCoV/USA/CA2/2020	1
31 - BetaCoV/Shenzhen/SZTH-003/2020	39 - BetaCoV/USA/CA2/2020	1
19 - BetaCov/Wuhan/WH04/2020	40 - BetaCoV/USA/AZ1/2020	1
32 - BetaCoV/Shenzhen/SZTH-004/2020	40 - BetaCoV/USA/AZ1/2020	1
7 - BetaCoV/Wuhan/WIV05/2019	42 - BetaCoV/Guangzhou/20SF206/2020	1
16 - BetaCov/Wuhan/WH03/2020	42 - BetaCoV/Guangzhou/20SF206/2020	1
19 - BetaCov/Wuhan/WH04/2020	42 - BetaCoV/Guangzhou/20SF206/2020	1
34 - BetaCoV/Zhejiang/WZ-02/2020	42 - BetaCoV/Guangzhou/20SF206/2020	1
40 - BetaCoV/USA/AZ1/2020	42 - BetaCoV/Guangzhou/20SF206/2020	1
42 - BetaCoV/Guangzhou/20SF206/2020	46 - BetaCoV/Taiwan/2/2020	1
32 - BetaCoV/Shenzhen/SZTH-004/2020	47 - BetaCoV/USA/CA1/2020	1
44 - BetaCo V/Foshan/20SF2 10/2020	47 - BetaCoV/USA/CA1/2020	1
6 - BetaCoV/Wuhan/WIV02/2019	49 - BetaCo V/France/IDF0372/2020	1
40 - BetaCoV/USA/AZ1/2020	49 - BetaCo V/France/IDF0372/2020	1
16 - BetaCov/Wuhan/WH03/2020	50 - BetaCoV/France/IDF0373/2020	1
19 - BetaCov/Wuhan/WH04/2020	50 - BetaCoV/France/IDF0373/2020	1
9 - BetaCoV/Wuhan/WIV07/2019	51 - BetaCoV/Singapore/ 1/2020	1
6 - BetaCoV/Wuhan/WIV02/2019	52 - BetaCoV/Australia/VICO1/2020	1
16 - BetaCov/Wuhan/WH03/2020	52 - BetaCoV/Australia/VICO1/2020	1
13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	53 - BetaCoV/Germany/BavPat 1/2020	1
24 - BetaCoV/Shenzhen/SZTH-00 1/2020	53 - BetaCoV/Germany/BavPat 1/2020	1
51 - BetaCoV/Singapore/ 1/2020	53 - BetaCoV/Germany/BavPat 1/2020	1
32 - BetaCoV/Shenzhen/SZTH-004/2020	54 - BetaCoV/England/01/2020	1
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	54 - BetaCoV/England/01/2020	1
46 - BetaCoV/Taiwan/2/2020	54 - BetaCoV/England/01/2020	1
19 - BetaCov/Wuhan/WH04/2020	55 - BetaCoV/England/02/2020	1
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	55 - BetaCoV/England/02/2020	1
49 - BetaCo V/France/IDF0372/2020	56 - BetaCoV/Finland/1/2020	1

Table 8. Frequency of edge distances.

Edge distance	Frequency
1	637395
2	156944
3	94028
4	63558
7	59381
6	31713
11	29915
10	29726
14	29606
20	27337
9	23259
5	4549
28	4473
12	2659
13	2627
15	2355
8	1885
16	1849
25	1769
18	1731
23	1665
21	1544
22	1464
19	1416
32	1302
35	1227
34	1007
24	970
30	939
29	871
26	725
33	638
46	601
27	538
31	530
44	497
47	484
37	471
17	470
45	412
39	387
51	338
38	97
40	64
43	4
50	3
36	2
48	2
42	1

Table 9. Cluster memberships of the nucleotide sequences.

Identifier	Cluster
36 - BetaCoV/USA/WA 1/2020	1
38 - BetaCoV/USA/IL 1/2020	1
39 - BetaCoV/USA/CA2/2020	1
40 - BetaCoV/USA/AZ1/2020	1
47 - BetaCoV/USA/CA1/2020	1
51 - BetaCoV/Singapore/ 1/2020	1
52 - BetaCoV/Australia/VICO1/2020	1
54 - BetaCoV/England/01/2020	1
55 - BetaCoV/England/02/2020	1
56 - BetaCoV/Finland/1/2020	1
10 - BetaCoV/Wuhan/HBCDC-HB-01/2019	2
20 - BetaCoV/Nonthaburi/6 1/2020	2
23 - BetaCoV/Nonthaburi/74/2020	2
26 - BetaCoV/Guangdong/20SFO0 12/2020	2
27 - BetaCoV/Guangdong/20SFO0 13/2020	2
28 - BetaCoV/Guangdong/20SFO0 14/2020	2
29 - BetaCoV/Guangdong/20SF025/2020	2
33 - BetaCoV/Guangdong/20SF028/2020	2
35 - BetaCoV/Guangdong/20SF040/2020	2
49 - BetaCo V/France/IDF0372/2020	2
2 - BetaCov/Wuhan/WHO01/2019	3
6 - BetaCoV/Wuhan/WIV02/2019	3
7 - BetaCoV/Wuhan/WIV05/2019	3
8 - BetaCoV/Wuhan/WIV06/2019	3
9 - BetaCoV/Wuhan/WIV07/2019	3
16 - BetaCov/Wuhan/WH03/2020	3
17 - BetaCoV/China/WHU01/2020	3
18 - BetaCoV/China/WHU02/2020	3
19 - BetaCov/Wuhan/WH04/2020	3
21 - BetaCoV/Shenzhen/HKU-SZ-002/2020	3
30 - BetaCoV/Zhejiang/WZ-0 1/2020	3
31 - BetaCoV/Shenzhen/SZTH-003/2020	3
32 - BetaCoV/Shenzhen/SZTH-004/2020	3
34 - BetaCoV/Zhejiang/WZ-02/2020	3
37 - BetaCoV/Zhejiang/Hangzhou- 1/2020	3
46 - BetaCoV/Taiwan/2/2020	3
50 - BetaCoV/France/IDF0373/2020	3
53 - BetaCoV/Germany/BavPat 1/2020	3
5 - BetaCoV/Wuhan/WIV04/2019	4
11 - BetaCoV/Wuhan/IPBCAMS-WH-04/2019	4
12 - BetaCoV/Wuhan/IPBCAMS-WH-03/2019	4
13 - BetaCoV/Wuhan/IPBCAMS-WH-02/2019	4
14 - BetaCoV/Wuhan/IV DC-HB-04/2020	4
15 - BetaCoV/Wuhan/IPBCAMS-WH-05/2020	4
22 - BetaCoV/Shenzhen/HKU-SZ-005/2020	4
24 - BetaCoV/Shenzhen/SZTH-00 1/2020	4
25 - BetaCoV/Shenzhen/SZTH-002/2020	4
41 - BetaCoV/Guangdong/20SF174/2020	5
42 - BetaCoV/Guangzhou/20SF206/2020	5
43 - BetaCo V/Foshan/20SF207/2020	5
44 - BetaCo V/Foshan/20SF2 10/2020	5
45 - BetaCoV/Foshan/20SF21 1/2020	5
48 - BetaCoV/Guangdong/20SF201/2020	5
1 - BetaCoV/Wuhan/IPBCAMS-WH-01/2019	6
3 - BetaCoV/Wuhan/IV DC-HB-01/2019	6
4 - BetaCoV/Wuhan/IV DC-HB-05/2019	6

Acknowledgements

We want to thank the laboratories and researchers who contributed the genomic research data used in this study and obtained via the *GISAID EpiFlu Database*™ [3] on Tuesday, February 4th 2020. The following tables contain information about the origination of the research data used in this study and give credit to the collecting and submitting researchers and laboratories.

Table 10. Nucleotide sequences used in this study that were obtained from the GISAID EpiFlu Database™.

Sequence identifier	Collecting lab	Address	Submitting lab	Authors
BetaCoV/Wuhan/IPBCAMS-WH-01/2019	Institute of Pathogen Biology	No. 9 Dong Dan San Tiao, Dong Cheng District, Beijing, P.R. China	Institute of Pathogen Biology, Chinese Academy of Medical Sciences	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhigiang Wu, Chao Wu, Yiwei Liu
BetaCov/Wuhan/WHO01/2019	General Hospital of Central Theater Command of People	NO.627 Wuluo Road, Wuchang District, Wuhan, China	BGI	Weijun Chen, Yuhai Bi, Weifeng Shi and Zhenhong Hu
BetaCoV/Wuhan/IV DC-HB-01/2019	National Institute for Viral Disease Control and Prevention	155 Changbai Road, Changping District, Beijing 102206 China	National Institute for Viral Disease Control and Prevention, China CDC	Wenjie TanXiang ZhaoWenling Wang Xuejun MaYongzhong JiangRoujian Lu, Ji Wang, Weimin ZhouPeihua
BetaCoV/Wuhan/IV DC-HB-05/2019	National Institute for Viral Disease Control and Prevention	155 Changbai Road, Changping District, Beijing 102206 China	National Institute for Viral Disease Control and Prevention, China CDC	Wenjie TanXuejun MaXiang ZhaoWenling Wang Yongzhong JiangRoujian LuJi WangPeihua Niu, Weimin Zhou,
BetaCoV/Wuhan/WIV04/2019	Wuhan Jinyintan Hospital	Ist Yintan Road, Dongxihu District, Wuhan	Wuhan Institute of Virology, Chinese Academy of Sciences	Peng Zhou, Xing-Lou Yang, Ding-Yu Zhang, Lei Zhang, Yan Zhu, Hao-Rui Si, Zhengli Shi
BetaCoV/Wuhan/WIV02/2019	Wuhan Jinyintan Hospital	Ist Yintan road, Dongxihu district, Wuhan	Wuhan Institute of Virology, Chinese Academy of Sciences	Peng Zhou, Xing-Lou Yang, Ding-Yu Zhang, Lei Zhang, Yan Zhu, Hao-Rui Si, Zhengli Shi
BetaCoV/Wuhan/WIV05/2019	Wuhan Jinyintan Hospital	Ist Yintan road, Dongxihu district, Wuhan	Wuhan Institute of Virology, Chinese Academy of Sciences	Peng Zhou, Xing-Lou Yang, Ding-Yu Zhang, Lei Zhang, Yan Zhu, Hao-Rui Si, Zhengli Shi
BetaCoV/Wuhan/WIV06/2019	Wuhan Jinyintan Hospital	Ist Yintan road, Dongxihu district, Wuhan	Wuhan Institute of Virology, Chinese Academy of Sciences	Peng Zhou, Xing-Lou Yang, Ding-Yu Zhang, Lei Zhang, Yan Zhu, Hao-Rui Si, Zhengli Shi
BetaCoV/Wuhan/WIV07/2019	Wuhan Jinyintan Hospital	Ist Yintan road, Dongxihu district, Wuhan	Wuhan Institute of Virology, Chinese Academy of Sciences	Peng Zhou, Xing-Lou Yang, Ding-Yu Zhang, Lei Zhang, Yan Zhu, Hao-Rui Si, Zhengli Shi
BetaCoV/Wuhan/HBCDC-HB-01/2019	Wuhan Jinyintan Hospital	1 Yintan Road, Dongxihu District, Wuhan City, Hubei Province, China, 430011	Hubei Provincial Center for Disease Control and Prevention	Bin Fang, Xiang Li, Xiao Yu, Linlin Liu, Bo Yang, Faxian Zhan, Guojun Ye, Xixiang Huo, Jungiang Xu, Bo Yu,
BetaCoV/Wuhan/IPBCAMS-WH-04/2019	Institute of Pathogen Biology	No. 9 Dong Dan San Tiao, Dong Cheng District, Beijing, P.R. China	Institute of Pathogen Biology, Chinese Academy of Medical Sciences	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhigiang Wu, Chao Wu, Yiwei Liu
BetaCoV/Wuhan/IPBCAMS-WH-03/2019	Institute of Pathogen Biology	No. 9 Dong Dan San Tiao, Dong Cheng District, Beijing, P.R. China	Institute of Pathogen Biology, Chinese Academy of Medical Sciences	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhigiang Wu, Chao Wu, Yiwei Liu
BetaCoV/Wuhan/IPBCAMS-WH-02/2019	Institute of Pathogen Biology	No. 9 Dong Dan San Tiao, Dong Cheng District, Beijing, P.R. China	Institute of Pathogen Biology, Chinese Academy of Medical Sciences	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhigiang Wu, Chao Wu, Yiwei Liu
BetaCoV/Wuhan/IV DC-HB-04/2020	National Institute for Viral Disease Control and Prevention	155 Changbai Road, Changping District, Beijing 102206 China	National Institute for Viral Disease Control and Prevention, China CDC	Wenjie TanXiang ZhaoWenling WangXuejyun MaYongzhong JiangRoujian LuJi WangWeimin ZhouPeihua
BetaCoV/Wuhan/IPBCAMS-WH-05/2020	Institute of Pathogen Biology	No. 9 Dong Dan San Tiao, Dong Cheng District, Beijing, P.R. China	Institute of Pathogen Biology, Chinese Academy of Medical Sciences	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhigiang Wu, Chao Wu, Yiwei Liu
BetaCov/Wuhan/WH03/2020	General Hospital of Central Theater Command of People	NO.627 Wuluo Road, Wuchang District, Wuhan, China	BGI	Weijun Chen, Yuhai Bi, Weifeng Shi and Zhenhong Hu

[3] https://www.gisaid.org/

Table 11. Nucleotide sequences used in this study that were obtained from the GISAID EpiFlu Database™(contd.).

Sequence identifier	Collecting lab	Address	Submitting lab	Authors
BetaCoV/China/WHU01/2020	unknown	unknown	State Key Laboratory of Virology, Wuhan University	Chen,L., Liu,W., Zhang,Q., Xu,K., Ye,G., Wu,W., Sun,Z., Liu,F., Wu,K., Mei, Y., Zhang,W., Chen, Y., Li,Y.,
BetaCoV/China/WHU02/2020	unknown	unknown	State Key Laboratory of Virology, Wuhan University	Chen,L., Liu,W., Zhang,Q., Xu,K., Ye,G., Wu,W., Sun,Z., Liu,F., Wu,K., Mei, Y., Zhang,W., Chen, Y., Li,Y.,
BetaCov/Wuhan/WH04/2020	General Hospital of Central Theater Command of People	NO.627 Wuluo Road, Wuchang District, Wuhan, China	BGI	Weijun Chen, Yuhai Bi, Weifeng Shi and Zhenhong Hu
BetaCoV/Nonthaburi/6 1/2020	Bamrasnaradura Hospital	38/4 Tiwanon Rd, Tambon Talat Kwan, Mueang, Nonthaburi 11000, Thailand	1. Department of Medical Sciences, Ministry of Public Health, Thailand	Pilailuk,Okada
BetaCoV/Shenzhen/HKU-SZ-002/2020	The University of Hong Kong - Shenzhen Hospital	1 Haiyuan Ist Road, Futian District, Shenzhen 518053, Guangdong, China	Li Ka Shing Faculty of Medicine, The University of Hong Kong	Chan,J.F.-W., Yuan,S., Kok,K.H., To,K.K.-W., Chu,H., Yang,J., Xing,F., Liu,J., Yip,C.C.-Y., Poon,R.W–S.,
BetaCoV/Shenzhen/HKU-SZ-005/2020	The University of Hong Kong - Shenzhen Hospital	1 Haiyuan Ist Road, Futian District, Shenzhen 518053, Guangdong, China	Li Ka Shing Faculty of Medicine, The University of Hong Kong	Chan,J.F.-W., Yuan,S., Kok,K.H., To,K.K.-W., Chu,H., Yang,J., Xing,F., Liu,J., Yip,C.C.-Y., Poon,R.W–S.,
BetaCoV/Nonthaburi/74/2020	Bamrasnaradura Hospital	38/4 Tiwanon Rd, Tambon Talat Kwan, Mueang, Nonthaburi 11000, Thailand	1. Department of Medical Sciences, Ministry of Public Health, Thailand	Pilailuk,Okada
BetaCoV/Shenzhen/SZTH-00 1/2020	Shenzhen Third People	NO. 29 Bulan road, Longgang district, Shenzhen, China	Shenzhen Key Laboratory of Pathogen and Immunity, National Clinical Research Center for Infectious	Yang Yang, Chenguang Shen, Li Xing, Zhixiang Xu, Haixia Zheng, Yingxia Liu
BetaCoV/Shenzhen/SZTH-002/2020	Shenzhen Key Laboratory of Pathogen and Immunity	NO. 29 Bulan road, Longgang district, Shenzhen, China	Shenzhen Key Laboratory of Pathogen and Immunity, National Clinical Research Center for Infectious Disease,	Yang Yang, Chenguang Shen, Li Xing, Zhixiang Xu, Haixia Zheng, Yingxia Liu
BetaCoV/Guangdong/20SFO0 12/2020	Guangdong Provincial Center for Diseases Control and Prevention	NO.160, Qunxian Road, Panyu District, Guangzhou, China	Department of Microbiology, Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCoV/Guangdong/20SFO0 13/2020	Guangdong Provincial Center for Diseases Control and Prevention	NO.160, Qunxian Road, Panyu District, Guangzhou, China	Department of Microbiology, Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCoV/Guangdong/20SFO0 14/2020	Guangdong Provincial Center for Diseases Control and Prevention	NO.160, Qunxian Road, Panyu District, Guangzhou, China	Department of Microbiology, Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCoV/Guangdong/ 20SF025/2020	Guangdong Provincial Center for Diseases Control and Prevention	NO.160, Qunxian Road, Panyu District, Guangzhou, China	Department of Microbiology, Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie

Table 12. Nucleotide sequences used in this study that were obtained from the GISAID EpiFlu Database™(contd.).

Sequence identifier	Collecting lab	Address	Submitting lab	Authors
BetaCoV/Zhejiang/WZ-0 1/2020	Zhejiang Provincial Center for Disease Control and Prevention	3399, Binsheng Road, Binjiang district, Hangzhou, 310051, China	Department of Microbiology, Zhejiang Provincial Center for Disease Control and Prevention	Yin Chen, Yanjun Zhang, Haiyan Mao, Junhang Pan, Xiuyu Lou, Yiyu Lu, Juying Yan, Hanping Zhu, Jian Gao,
BetaCoV/Shenzhen/SZTH-003/2020	Shenzhen Key Laboratory of Pathogen and Immunity	NO. 29 Bulan road, Longgang District, Shenzhen, China	Shenzhen Key Laboratory of Pathogen and Immunity, National Clinical Research Center for Infectious Disease,	Yang Yang, Chenguang Shen, Li Xing, Zhixiang Xu, Haixia Zheng, Yingxia Liu
BetaCoV/Shenzhen/SZTH-004/2020	Shenzhen Key Laboratory of Pathogen and Immunity	NO. 29 Bulan road, Longgang district, Shenzhen, China	Shenzhen Key Laboratory of Pathogen and Immunity, National Clinical Research Center for Infectious Disease,	Yang Yang, Chenguang Shen, Li Xing, Zhixiang Xu, Haixia Zheng, Yingxia Liu
BetaCoV/Guangdong/ 20SF028/2020	Guangdong Provincial Center for Diseases Control and Prevention	NO.160, Qunxian Road, Panyu District, Guangzhou, China	Department of Microbiology, Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCoV/Zhejiang/WZ-02/2020	Zhejiang Provincial Center for Disease Control and Prevention	3399, Binsheng Road, Binjiang district, Hangzhou, 310051, China	Department of Microbiology, Zhejiang Provincial Center for Disease Control and Prevention	Yanjun Zhang, Yin Chen, Haiyan Mao, Junhang Pan, Xiuyu Lou, Yiyu Lu, Juying Yan, Hanping Zhu, Jian Gao,
BetaCoV/Guangdong/ 20SF040/2020	Guangdong Provincial Center for Diseases Control and Prevention	NO.160, Qunxian Road, Panyu District, Guangzhou, China	Department of Microbiology, Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCoV/USA/WA 1/2020	Providence Regional Medical Center	1700 13th St, Everett, WA 98201	Division of Viral Diseases, Centers for Disease Control and Prevention	Queen,K., Tao, Y., Li,Y., Paden,C.R., Lu,X., Zhang,J., Gerber,S.I., Lindstrom,S., Tong,S.
BetaCoV/Zhejiang/Hangzhou- 1/2020	Hangzhou Center for Disease and Control Microbiology Lab	Mingshi Road 568, Hanghzhou, Zhejiang, China	Hangzhou Center for Disease and Control Microbiology Lab	Yu Hua, Wang Haogiu, Li Jun, Yu Xinfeng
BetaCoV/USA/IL 1/2020	IL Department of Public Health Chicago Laboratory	2121 West Taylor, Chicago IL 60612	Pathogen Discovery, Respiratory Viruses Branch, Division of Viral Diseases, Centers for Dieases Control and	Ying Tao, Krista Queen, Clinton R. Paden, Jing Zhang, Yan Li, Anna Uehara, Xiaoyan Lu, Brian Lynch, Senthil
BetaCoV/USA/CA2/2020	California Department of Public Health	850 Marina Bay Parkway Richmond CA 94804	Pathogen Discovery, Respiratory Viruses Branch, Division of Viral Diseases, Centers for Dieases Control and	Anna Uehara, Krista Queen, Ying Tao, Yan Li, Clinton R. Paden, Jing Zhang, Xiaoyan Lu, Brian Lynch, Senthil
BetaCoV/USA/AZ1/2020	Arizona Department of Health Services	250 N. 17th Avenue Phoenix AZ 85007	Pathogen Discovery, Respiratory Viruses Branch, Division of Viral Diseases, Centers for Disease Control and	Ying Tao, Clinton R. Paden, Krista Queen, Anna Uehara, Yan Li, Jing Zhang, Xiaoyan Lu, Brian Lynch, Senthil
BetaCoV/Guangdong/ 20SF174/2020	Guangdong Provincial Center for Diseases Contorl and Prevention	No.160, Qunxian Road, Panyu district	Guangdong Provincial Center for Disease Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie

Table 13. Nucleotide sequences used in this study that were obtained from the GISAID EpiFlu Database™(contd.).

Sequence identifier	Collecting lab	Address	Submitting lab	Authors
BetaCoV/Guangzhou/ 20SF206/2020	Guangdong Provincial Center for Diseases Contorl and Prevention	NO.160, Qunxian Road, Panyu District, Guangzhou, China	Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCo V/Foshan/20SF207/2020	Guangdong Provincial Center for Diseases Control and Prevention	No.160, Qunxian Road, Panyu district	Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCo V/Foshan/20SF2 10/2020	Guangdong Provincial Center for Diseases Control and Prevention	No.160, Qunxian Road, Panyu district	Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCoV/Foshan/20SF21 1/2020	Guangdong Provincial Center for Diseases Control and Prevention	No.160, Qunxian Road, Panyu District, Guangzhou, China	Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCoV/Taiwan/2/2020	Centers for Disease Control	161, Kunyang Street, 11561 Taipei, Taiwan	Centers for Disease Control, R.O.C.	Ji-Rong Yang, Yu-Chi Lin, Jung-Jung Mu, Ming-Tsan Liu, Shu-Ying Li
BetaCoV/USA/CA1/2020	California Department of Public Health	850 Marina Bay Parkway Richmond CA 94804	Pathogen Discovery, Respiratory Viruses Branch, Division of Viral Diseases, Centers for Dieases Control and	Anna Uehara, Krista Queen, Ying Tao, Yan Li, Clinton R. Paden, Jing Zhang, Xiaoyan Lu, Brian Lynch, Senthil
BetaCoV/Guangdong/ 20SF201/2020	Guangdong Provincial Center for Diseases Control and Prevention	NO.160, Qunxian Road, Panyu District, Guangzhou, China	Guangdong Provincial Center for Diseases Control and Prevention	Min Kang, Jie Wu, Jing Lu, Tao Liu, Baisheng Li, Shujiang Mei, Feng Ruan, Lifeng Lin, Changwen Ke, Haojie
BetaCo V/France/IDF0372/2020	Department of Infectious and Tropical Diseases	46, rue Henri Huchard	National Reference Center for Viruses of Respiratory Infections, Institut Pasteur, Paris	Mélanie Albert, Marion Barbet, Sylvie Behillil, Méline Bizard, Angela Brisebarre, Flora Donati, Vincent Enouf, Maud Vanpeene, Sylvie van der Werf, Yazdan Yazdanpanah, Xavier Lescure.
BetaCoV/France/IDF0373/2020	Department of Infectious and Tropical Diseases	46, rue Henri Huchard	National Reference Center for Viruses of Respiratory Infections, Institut Pasteur, Paris	Mélanie Albert, Marion Barbet, Sylvie Behillil, Méline Bizard, Angela Brisebarre, Flora Donati, Vincent Enouf, Maud Vanpeene, Sylvie van der Werf, Yazdan Yazdanpanah, Xavier Lescure.
BetaCoV/Singapore/ 1/2020	Singapore General Hospital	Molecular Laboratory, Division of Pathology	National Public Health Laboratory	Mak, TM
BetaCoV/Australia/VICO1/2020	Monash Medical Centre	246 Clayton Road, Clayton, Melbourne, VIC, Australia	Collaboration between the University of Melbourne at The Peter Doherty Institute for Infection and Immunity, and	Caly,L., Seemann,T., Schultz,M., Druce,J. and Taiaroa,G
BetaCoV/Germany/BavPat 1/2020	Charité Universitätsmedizin Berlin, Institute of Virology; Institut für Mikrobiologie der Bundeswehr, Munich	Institute of Virology, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany; Institut für Mikrobiologie der Bundeswehr, Neuherbergstraße 11, 80937 Munich, Germany	Charité Universitätsmedizin Berlin, Institute of Virology	Victor M Corman, Julia Schneider, Talitha Veith, Barbara Mühlemann, Markus Antwerpen, Christian Drosten, Roman Wölfel
BetaCoV/England/01/2020	Respiratory Virus Unit	61 Colindale Avenue, London, NW9 5EQ, United Kingdom	Respiratory Virus Unit, Microbiology Services Colindale, Public Health England	Monica Galiano, Shahjahan Miah, Richard Myers, Angie Lackenby, Omolola Akinbami, Tiina Talts, Leena Bhaw, Kirstin Edwards, Jonathan Hubb, Joanna Ellis, Maria Zambon.
BetaCoV/England/02/2020	Respiratory Virus Unit	61 Colindale Avenue, London, NW9 5EQ, United Kingdom	Respiratory Virus Unit, Microbiology Services Colindale, Public Health England	Monica Galiano, Shahjahan Miah, Richard Myers, Angie Lackenby, Omolola Akinbami, Tiina Talts, Leena Bhaw, Kirstin Edwards, Jonathan Hubb, Joanna Ellis, Maria Zambon.
BetaCoV/Finland/1/2020	Lapland Central Hospital	Lapland Hospital District, P.O. Box 8041, 96101 Finland	Department of Virology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland	Teemu Smura, Suvi Kuivanen, Hannimari Kallio-Kokko, Olli Vapalahti