

Gábor Kemenesi^{1,2,3,*}, Safia Zeghib^{1,2,3,d}, Balázs A Somogyi^{1,2,3}, Gábor Endre Tóth^{1,2,3}, Krisztián Bányai⁴, Norbert Solymosi⁵, Peter M Szabo⁶, István Szabó⁷, Ádám Bálint⁷, Péter Urbán⁸, Róbert Herczeg⁸, Attila Gyenesei^{8,9}, Ágnes Nagy¹⁰, Csaba István Pereszlényi¹⁰, Gergely Csaba Babinszky¹⁰, Gábor Dudás¹⁰, Gabriella Terhes¹¹, Viktor Zöldi¹², Róbert Lovas¹³, Szabolcs Tenczer¹³, László Kornya¹⁴, Ferenc Jakab^{1,2,3,*}

¹ Virological Research Group BSL-4 Laboratory, Szentágotthai Research Centre, University of Pécs, Pécs, Hungary

² Institute of Biology, Faculty of Sciences, University of Pécs, Pécs, Hungary

³ National Coronavirus Research Centre, University of Pécs, Pécs, Hungary

⁴ Institute for Veterinary Medical Research, Centre for Agricultural Research, Budapest, Hungary

⁵ University of Veterinary Medicine Budapest, Centre for Bioinformatics, Budapest, Hungary

⁶ Translational Discovery, Stromal Biology, Bristol-Myers Squibb, Princeton, NJ, USA

⁷ National Food Safety Office, Budapest, Hungary

⁸ Szentágotthai Research Centre, Bioinformatics Research Group, Genomics and Bioinformatics Core Facility, University of Pécs, Pécs, Hungary.

⁹ Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland

¹⁰ Hungarian Defense Forces, Military Medical Centre, Budapest, Hungary

¹¹ Institute of Clinical Microbiology, University of Szeged, Faculty of Medicine, Szeged, Hungary

¹² Independent researcher, Vantaa, Finland

¹³ Institute for Computer Science and Control, Hungarian Academy of Sciences

¹⁴ Central Hospital of Southern Pest, Budapest

d These authors contributed equally

***Corresponding author: professor Ferenc Jakab, Ph.D, e-mail: jakab.ferenc@pte.hu; Gabor Kemenesi, Ph.D, e-mail: kemenesi.gabor@pte.hu**

Multiple SARS-CoV-2 introductions shaped the early outbreak in Central Eastern Europe: comparing Hungarian data to a worldwide sequence data-matrix

Severe Acute Respiratory Syndrome Coronavirus 2 is the third highly pathogenic human coronavirus in history. Since the emergence in Hubei province, China, during late 2019 the situation evolved to pandemic level. Following China, Europe was the second epicenter of the pandemic. To better comprehend the detailed founder mechanisms of the epidemic evolution in Central-Eastern Europe, particularly in Hungary, we determined the full-length SARS-CoV-2 genomes from 32 clinical samples collected from laboratory confirmed COVID-19 patients over the first month of disease in Hungary. We applied a haplotype network analysis on all available complete genomic sequences of SARS-CoV-2 from GISAID database as of the 21th of April, 2020. We performed additional phylogenetic and phylogeographic analyses to achieve the recognition of multiple and parallel introductory events into our region. Here we present a publicly available network imaging of the worldwide haplotype relations of SARS-CoV-2 sequences and conclude the founder mechanisms of the outbreak in Central-Eastern Europe.

Introduction

Following the 2002 SARS (Severe Acute Respiratory Syndrome) pandemic and the discovery of MERS (Middle Eastern Respiratory Syndrome) coronavirus in 2012, the third highly pathogenic human coronavirus in history emerged in Hubei province, China, during late 2019. The novel virus was subsequently named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and the acute respiratory disease as coronavirus disease 19 (COVID-19) ¹. Currently, SARS-CoV-2 is responsible for the ongoing coronavirus pandemic spreading on all inhabited continents. As of April 26, 2020, the confirmed case numbers surpassed 3 million worldwide and the disease associated mortality rate exceeded 200,000 ².

At the onset of the second week of March, Europe became the next epicenter of the pandemic, following China, as reported by the World Health Organization ³. By the end of April, more than one million laboratory confirmed cases were reported from all European countries ⁴. The first two Hungarian cases were officially confirmed on March 4th, according to the data of ECDC Communicable Disease Threats Report ⁵. Border closures and universal ban regarding public gatherings was announced on March 17.

To better comprehend the detailed founder mechanisms of the epidemic evolution in Central-Eastern Europe, particularly in Hungary, we determined the full-length SARS-CoV-2 genomes from 32 clinical samples collected from laboratory confirmed COVID-19 patients over the first month of disease in Hungary. Our virus sampling started from this date and spanned the first two weeks of country-wide mitigation regulations (March 17 through April 2, 2020).

Results and Discussion

In order to understand the origin of Hungarian-based COVID-19 epidemics and provide baseline data for the evaluation of future epidemic events we applied a network analysis on complete genomic sequence data of SARS-CoV-2 available in GISAID database⁶ current to April 21. The network showed negative exponential degree distribution which is common regarding scale-free networks⁷. This characteristic network is typical for epidemics⁸. However, several nodes represented a higher frequency in the lower part of the plot which is the tendency associated with small-world networks⁹ (Supplementary Figure 1.). Altogether, a total of 147 clusters were identified with a Girvan-Newmann community detection algorithm¹⁰. In consideration of this approach, a total of nine main clusters were described which together serve as the base for the remaining smaller clusters (**Figure 1.**). Although the investigated network contained relatively high number of clusters, its diameter is 25, which infers the farthest distance in the matrix between two sequence is 25 steps, whilst the average path length is 8.91 steps. The high cluster rate was supported by the ratio of these two measures. The proportion of present edges from all possible edges in the network was 0.004 (edge density).

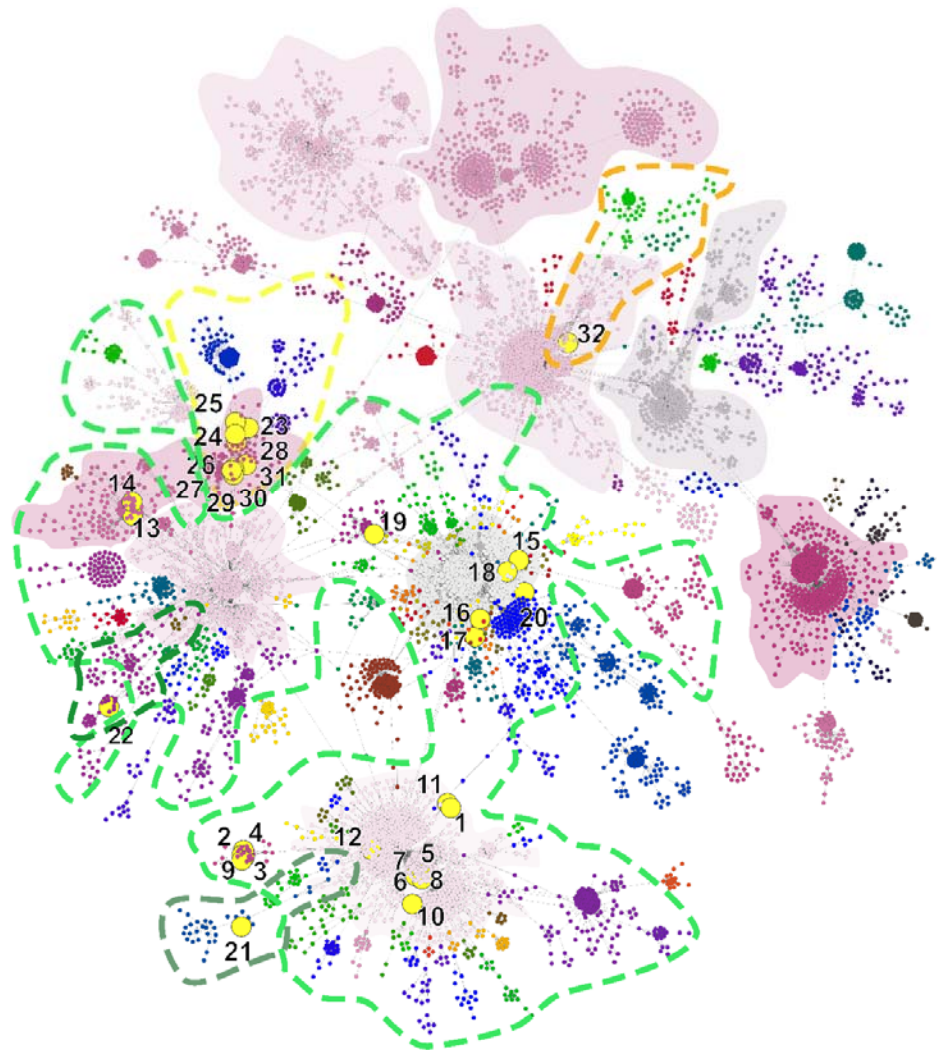


Figure 1.: Genetic network analysis of 7864 SARS-CoV-2 complete genomic sequences. Hungarian strains are indicated with numbered yellow dots – numbers referring to Table 1. The nine major clades are represented by a solid color. Genetic lineages are marked with colored dotted lines, where green lines are bordering B 1, B 1.1 and B 1.11; yellow and orange lines mark B 1.5 and B 3 respectively.

Hungarian genomes are dispersed within four main clusters out of the nine (**Figure 2.**). The genome designated SARS-CoV-2/human/Hungary/620/27_03_2020 is positioned in the main Chinese cluster (incorporating early Chinese sequences and data from Hangzhou, in January) and closely connected to a Taiwanese sequence on March 23rd (**Figure 2.**, Cluster C). Apart from other Hungarian sequences, this is the only indication for the introduction from a non-European source towards Hungary at the examined time-period in consideration of the available sequence data. The remainder of the sequences are dispersed among three other main clusters: A, B and I (**Figure 2.**). A and I clusters are structured by mostly the

Western-European sequences, whilst B is a dominant cluster in the USA. Although sampling bias may largely alter the conclusions for the exact geographic origin of a particular strain, the main patterns as multiple introductions from different sources can be concluded.

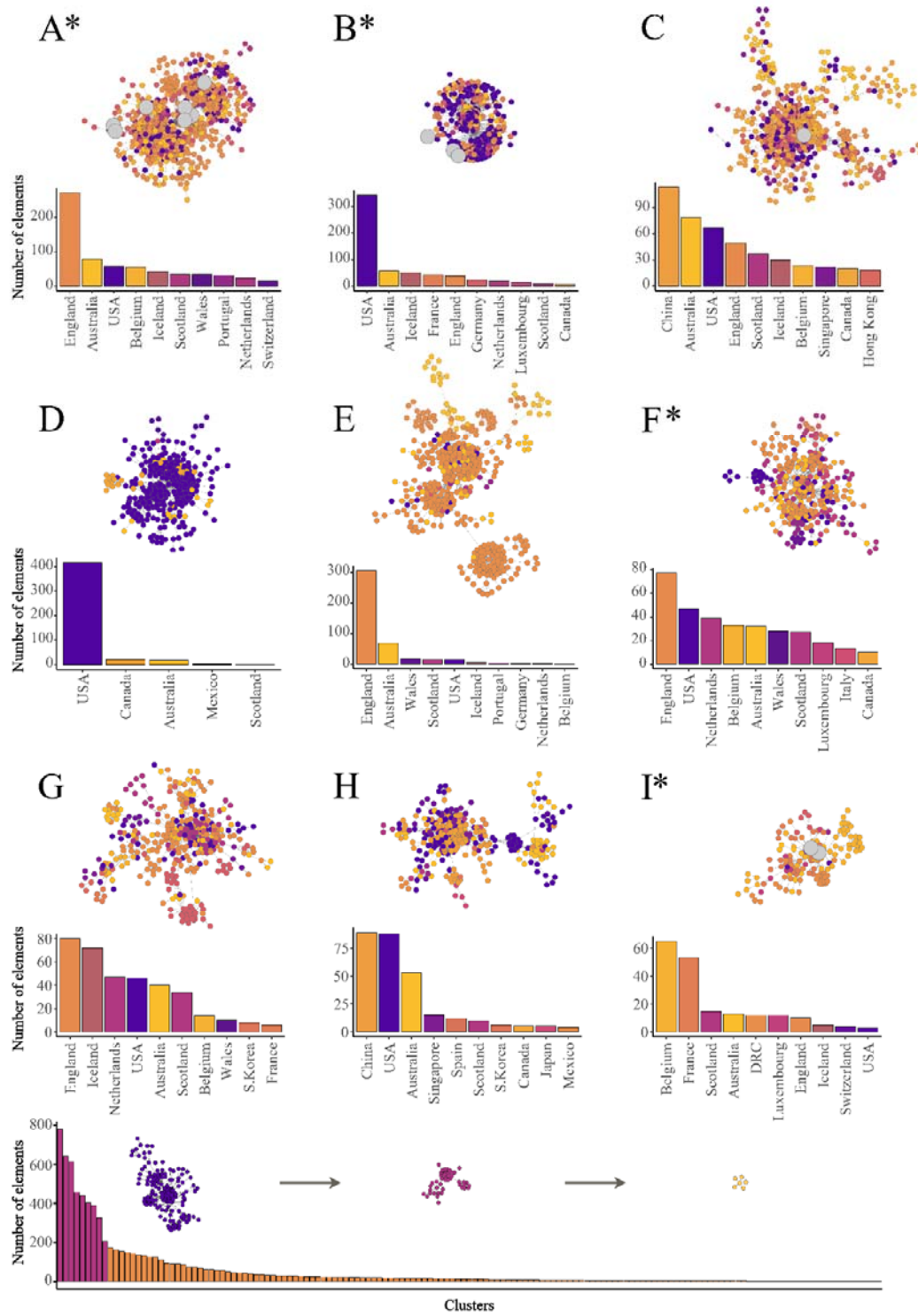


Figure 2: Representation of the nine main clusters which serve as the baseline regarding the worldwide haplotype network of SARS-CoV-2 genomes as of April 21, 2020. The ten most common countries of each cluster are summarized in a column graph and represented using different colors. Hungarian sequences are depicted by enlarged grey dots. Star indicates the participation of a sample from this study indirectly to the specified main cluster (i.e., participation in a sub-cluster relevant to the main cluster). Number of elements within each remaining (n=147) smaller cluster is indicated as a simple column chart at the bottom of the figure.

Using the complete haplotype network dataset as a backbone, we applied additional phylogenetic and phylogeographic analyses (Supplementary Figure S2). It is likely that occupation-related movement within the EU resulted in multiple introductory events from Western-European host countries towards Central-Eastern Europe. This observation is further supported by recent narrative analysis on the Nextstrain online platform focusing on Eastern European processes of SARS-CoV-2 pandemic evolution¹¹. Similarly to Hungary and possibly to the entire region, there were eleven separate introductions to Poland, based on the currently available sequence data¹¹. In order to leverage additional support regarding this phenomenon, we applied a local Nextstrain database workflow in the addition of the sequences from this manuscript (Supplementary Figure S3)¹². As a result of this analysis and considering the observation from Poland, we were able to lend more support for the regular and dispersed introductions into Central-Europe. In addition to regular movement, the border restrictions as outbreak mitigation measures fixed a narrow timescale for individuals returning to Hungary and likely facilitated the parallel introductory events dispersed throughout the country. Based on genetic lineage categorization using PANGOLIN software, 20 out of the total 32 Hungarian sequences fell into the most dominant (i.e., most sequenced) lineage B.1 (**Table 1**). Dominance may largely depend on sampling heterogeneity between geographic regions and countries. However, it substantiates the connection of Hungary regarding SARS-CoV-2 cases to multiple European sources and provides additional support for the network analysis.

Table 1: Summary of the PANGOLIN software analysis. The table indicates the numbers of Figure 1. and offers additional details for each sample. Background data is also noted where it was available. Different background color of SARS-CoV-2 strain names highlight different genetic lineage groups represented with the same colored dotted lines on Figure 1

Across the phylogenetic tree (Supplementary Figure S2A), several of the Hungarian sequences were interspersed and mainly clustered with European sequences (England, France, Iceland and Germany) and supported with high posterior probabilities (>80%) while only one Hungarian sequence clustered with a North-American sequence (PP = 95%). These observations elegantly support the scenario regarding multiple individual introductions. In parallel, local clusters were also observed (PP = 100%) indicating local transmission even within the short timeframe of sampling. Moreover, several of the local clusters had very low PP indicating missing data which is likely to be the consequence of insufficient contact tracing and subsequent missing sequence data.

Within our dataset, the phylogeographic analysis indicated China as the root location (diffusion origin) (Supplementary Figure S2B). Moreover, the virus seemed to spread out to

Hungary mainly from Western European countries, nevertheless local transmissions also contributed to disease spread within the country. The data correspond with the epidemiological history of SARS-2-CoV-2 in Hungary⁴.

As a main result of this study, we present and provide a large-scale haplotype network backbone in reference to the immediate analysis of pandemic evolution of SARS-CoV-2. It is a rapid and useful tool to assess the origin of particular sequences and the acquisition of important data for regarding public health mitigation actions, discovering unidentified infection sources or super-spreading events on a large-scale. In general, it provides the network-based opportunity of rapid, genetic distance-based analysis for all available sequence data, in any context. Herein, we offer this network file available for any researchers to facilitate the understanding of SARS-CoV-2 pandemic evolution. The network file is suitable to visualize any available sequences, available at late April 2020, in its context to all known sequence data.

The importance of early, country-based mitigation measures are thoroughly exemplified on this dataset. We presented the emergence of multiple virus clusters from various sources in Hungary during the early phase of the epidemic. However, the publicly available epidemiologic data indicate a predominance of confirmed cases in and adjacent to the capital city, Budapest. Possibly, this phenomenon is due to effective mitigation by limiting individual movement, application of social distancing and border restrictions¹³. Therefore, we believe a pan-European, coordinated mitigation policy will be beneficial to prevent significant mixture of European clusters during future epidemics.

Our research further highlights the importance of genomic epidemiologic tools for public health decision making. The combination of different methods (i.e., network analysis and phylogenetic approaches) may greatly facilitate the understanding of COVID-19 outbreak evolution.

Methods

Sample collection

Oro-pharyngeal swab samples were obtained from 32 patients during the period from March 17 to April 2. Within the frame of a country-wide collaboration network regarding SARS-CoV-2 research, nucleic-acid samples were received from University Hospitals at Szeged and Budapest and from the Hungarian Defense Forces, Military Medical Center. Ethical approval was obtained from the University of Pécs, Ethics Committee, under the registration number: 8218-PTE2020.

Direct sequencing and primary data analysis from patient samples

Nucleic acid samples were extracted directly from oro-pharyngeal swab samples using a Direct-zol™ RNA MiniPrep Plus extraction kit (Zymo Research) and in full compliance to the manufacturers' recommendations. Reverse transcription and multiplex PCR were performed on the basis of information provided by the Artic Network initiative¹⁴. Both the concentration and the quality of the PCR products were measured and checked using the Agilent 4200 TapeStation System and ThermoFisher Scientific Qubit 3 Fluorometer. The 32 sequencing libraries were prepared using 98 overlapping amplicons covering the whole viral

genome. The libraries were then quantitatively checked, barcoded and sequenced on 5 flow cells using Oxford Nanopore MinION Flow Cells (R9.4.1).

During primary data analysis, we used RAMPART to track the sequencing process in “real-time” in order to acquire instant information regarding the quality of samples and the coverage of the amplicons. Sequencing reads of samples with sufficient amplicon coverage were mapped and consensus sequences generated by the bioinformatics pipeline built within the Artic Network protocol.

Genome data analysis

SARS-CoV-2 genomes (n=7864) were downloaded from GISAID database on 21 April, 2020. Only complete (>29000 base-pair length) and high quality (with <1% Ns, <0.05% unique amino acid mutations and no insertion/deletion unless verified by submitter) sequences were used for network construction. To quantify the sequence similarity, percent identity was calculated based on the BLAST¹⁵ alignment for each paired sequence.

First, using the resulted similarity matrix, a fully connected, edge-weighted network was constructed, where each node represented a COVID sequence, while the edges represented their potential connections, and the edge weights (similarity values). Secondly, the edge weights were transformed (100-weight) in the full network to make high values low and low values high. Next, a minimum spanning tree (MST) was identified. In a spanning tree, every node has only one or two connections. If multiple edges have the same minimum weight, the algorithm will randomly pick one and not select all links with the same values. To manage this issue, the graph with additional edges was modified by adding every edge for each node having an equal or higher weight than the edges in the initial MST to the corresponding node. All data analyses were performed using the R 3.6.2 on Linux¹⁶, for network creation, and the Igraph package was applied¹⁷.

In regards to the generation of time-scaled phylogenetic tree, 105 SARS-CoV-2 genomes were retrieved from GISAID⁶ following a manual selection based on the network analysis. The sequences were aligned in MAFFT¹⁸ with default parameters. Subsequently, both best-fitting substitution model and the maximum likelihood phylogenetic tree with ultra-bootstrapping were implemented in IQTREE webserver^{19,20}. The resulting tree was subjugated to a root-to-tip regression analysis in TempEst²¹ to assess the clock-likeness regarding the data. A positive correlation was observed between sampling time and root-to-tip genetic divergence indicating the suitability of the dataset for molecular clock analysis using the Beast v1.10.4 package. The KHY+I substitution model with the uncorrelated lognormal relaxed clock, in addition to the coalescent exponential population growth model, were applied²². The MCMC chains were run for 200 million iterations and sampled every 10,000 cycles, or generations, with 10% discarded as burn in. We explored the effective sample sizes in Tracer (ESS>200)²³. Moreover, to explore the phylogeographic diffusion of SARS-COV-2 in continuous space, the lognormal relaxed random walk diffusion model and a lognormal uncorrelated relaxed clock model were implemented in the same package, were next employed. Thus, the Maximum clade credibility tree was visualized in SpreaD3²⁴.

Lineage assignment of the Hungarian sequences was performed using the PANGOLIN (Phylogenetic Assignment of Named Global Outbreak LINEages) software v1.0, which uses a recently published lineage nomenclature^{25,26}.

The datasets generated during and analysed during the current study are available in the NDEX-The Network Data Exchange repository, [<http://www.ndexbio.org/#/network/2c66e15b-8eeb-11ea-aaef-0ac135e8bacf>].

References

1. Gorbalenya, A. E. *et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* vol. 5 536–544 (2020).
2. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* (2020) doi:10.1016/S1473-3099(20)30120-1.
3. WHO Director-General’s opening remarks at the media briefing on COVID-19 - 13 March 2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-mission-briefing-on-covid-19---13-march-2020>.
4. Situation update for the EU/EEA and the UK, as of 30 April 2020. <https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea>.
5. Communicable disease threats report, 1-7 March 2020, week 10. <https://www.ecdc.europa.eu/en/publications-data/communicable-disease-threats-report-1-7-march-2020-week-10>.
6. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, 30494 (2017).
7. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* (80-.). **286**, 509–512 (1999).
8. Keeling, M. J. & Eames, K. T. . Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307 (2005).
9. Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
10. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7821–7826 (2002).
11. Nextstrain / narratives / ncov / sit-rep / 2020-05-01. https://nextstrain.org/narratives/ncov/sit-rep/2020-05-01?f_region=Europe.
12. Nextstrain. <https://nextstrain.org/>.
13. Remuzzi, A. & Remuzzi, G. COVID-19 and Italy: what next? *The Lancet* vol. 395 1225–1228 (2020).
14. Quick, J. Forked from Ebola virus sequencing protocol. (2020) doi:10.17504/protocols.io.bbmuk6w.
15. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
16. R: The R Project for Statistical Computing. <https://www.r-project.org/>.
17. Csárdi, G. & Nepusz, T. *The igraph software package for complex network research*.
18. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
19. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).

20. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2017).
21. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, (2016).
22. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, (2018).
23. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
24. Bielejec, F. *et al.* Spred3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Mol. Biol. Evol.* **33**, 2167–2169 (2016).
25. GitHub - hCoV-2019/pangolin: Software package for assigning SARS-CoV-2 genome sequences to global lineages. <https://github.com/hCoV-2019/pangolin>.
26. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* 2020.04.17.046086 (2020) doi:10.1101/2020.04.17.046086.

Acknowledgements

On behalf of the project, “Genomic Epidemiology of SARS-CoV-2 in Hungary” we are grateful for the usage of MTA Cloud (<https://cloud.mta.hu/>) which significantly aided in achieving the results published in this paper. Gabor Kemenesi was supported by the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences. Balazs A Somogyi was supported by the ÚNKP-19-3 New National Excellence Program of the Ministry For Innovation and Technology. The research was performed in collaboration with the Genomics and Bioinformatics Core Facility at the Szentágotthai Research Center of the University of Pécs. Bioinformatics infrastructure was supported by ELIXIR Hungary (<http://elixir-hungary.org/>).

Authors Contribution

Conceptualization: GK, SZ, BAS, KB; Sample processing: GET, PU, ÁN, CIP, GCP, GD, GT; Sequencing: GET, PU; Bioinformatic support: RH, AGY; Sequence manipulation and analysis: SZ, BAS, KB, NS, PMS, IS, AB, RL, ST; Writing and editing: GK, BAS, SZ, FJ; Medical and epidemiological revision: VZ, JK; Supervising: FJ, GK. All authors reviewed the manuscript.

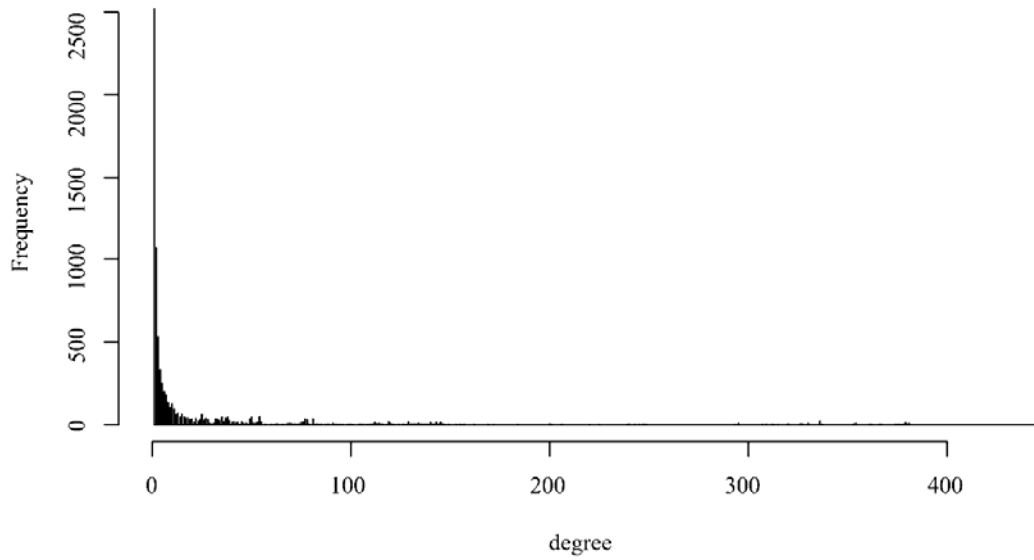
Competing interests

The authors declare no competing interests.

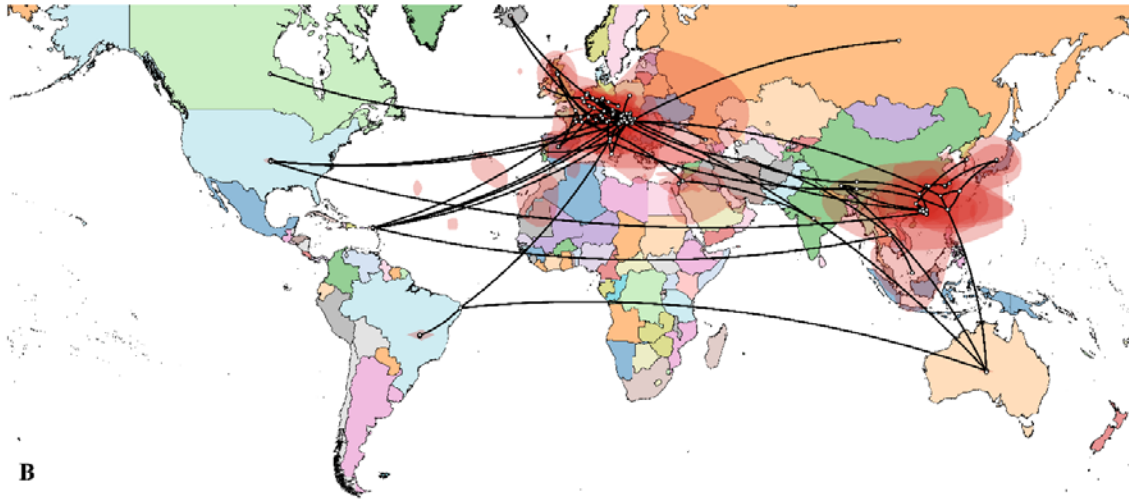
No. on Figure 2.	Taxon name	lineage	SH-ahrt	UFbootstrap	Note
1	SARS-CoV-2/human/Hungary/49/20_03_2020	B.1	100	100	
2	SARS-CoV-2/human/Hungary/55/20_03_2020	B.1	100	100	
3	SARS-CoV-2/human/Hungary/126/22_03_2020	B.1	100	100	
4	SARS-CoV-2/human/Hungary/186/23_03_2020	B.1	100	100	
5	SARS-CoV-2/human/Hungary/105w/21_03_2020	B.1	100	100	
6	SARS-CoV-2/human/Hungary/278w/25_03_2020	B.1	100	100	
7	SARS-CoV-2/human/Hungary/2801w/25_03_2020	B.1	100	100	
8	SARS-CoV-2/human/Hungary/3670w/29_03_2020	B.1	100	100	
9	SARS-CoV-2/human/Hungary/541/27_03_2020	B.1	100	100	
10	SARS-CoV-2/human/Hungary/777/30_03_2020	B.1	100	100	
11	SARS-CoV-2/human/Hungary/175/23_03_2020	B.1	100	100	
12	SARS-CoV-2/human/Hungary/417/25_03_2020	B.1	100	100	
13	SARS-CoV-2/human/Hungary/3597w/28_03_2020	B.1	100	100	
14	SARS-CoV-2/human/Hungary/MBL-3/25_03_2020	B.1	100	100	Travel-related: France to Hungary
15	SARS-CoV-2/human/Hungary/67/20_03_2020	B.1	100	100	
16	SARS-CoV-2/human/Hungary/183/23_03_2020	B.1	100	100	
17	SARS-CoV-2/human/Hungary/419/26_03_2020	B.1	100	100	
18	SARS-CoV-2/human/Hungary/827/30_03_2020	B.1	100	100	
19	SARS-CoV-2/human/Hungary/836/30_03_2020	B.1	100	100	
20	SARS-CoV-2/human/Hungary/792/30_03_2020	B.1	100	100	
21	SARS-CoV-2/human/Hungary/817/30_03_2020	B.1.1	100	93	
22	SARS-CoV-2/human/Hungary/572w/29_03_2020	B.1.11	100	99	
23	SARS-CoV-2/human/Hungary/2/17_03_2020	B.1.5	100	85	
24	SARS-CoV-2/human/Hungary/MBL-2/23_03_2020	B.1.5	100	74	Household infection
25	SARS-CoV-2/human/Hungary/MBL-1/17_03_2020	B.1.5	100	79	Household infection
26	SARS-CoV-2/human/Hungary/66/20_03_2020	B.1.5	100	94	Travel-related: Spain to Hungary
27	SARS-CoV-2/human/Hungary/1788lc/19_03_2020	B.1.5	100	93	
28	SARS-CoV-2/human/Hungary/MBL-464/27_03_2020	B.1.5	100	87	Hospital cluster
29	SARS-CoV-2/human/Hungary/MBL-465/27_03_2020	B.1.5	100	93	Hospital cluster
30	SARS-CoV-2/human/Hungary/MBL-469/27_03_2020	B.1.5	100	93	Hospital cluster
31	SARS-CoV-2/human/Hungary/1136/02_04_2020	B.1.5	85	76	
32	SARS-CoV-2/human/Hungary/620/27_03_2020	B.3	100	87	

Supplementary Data

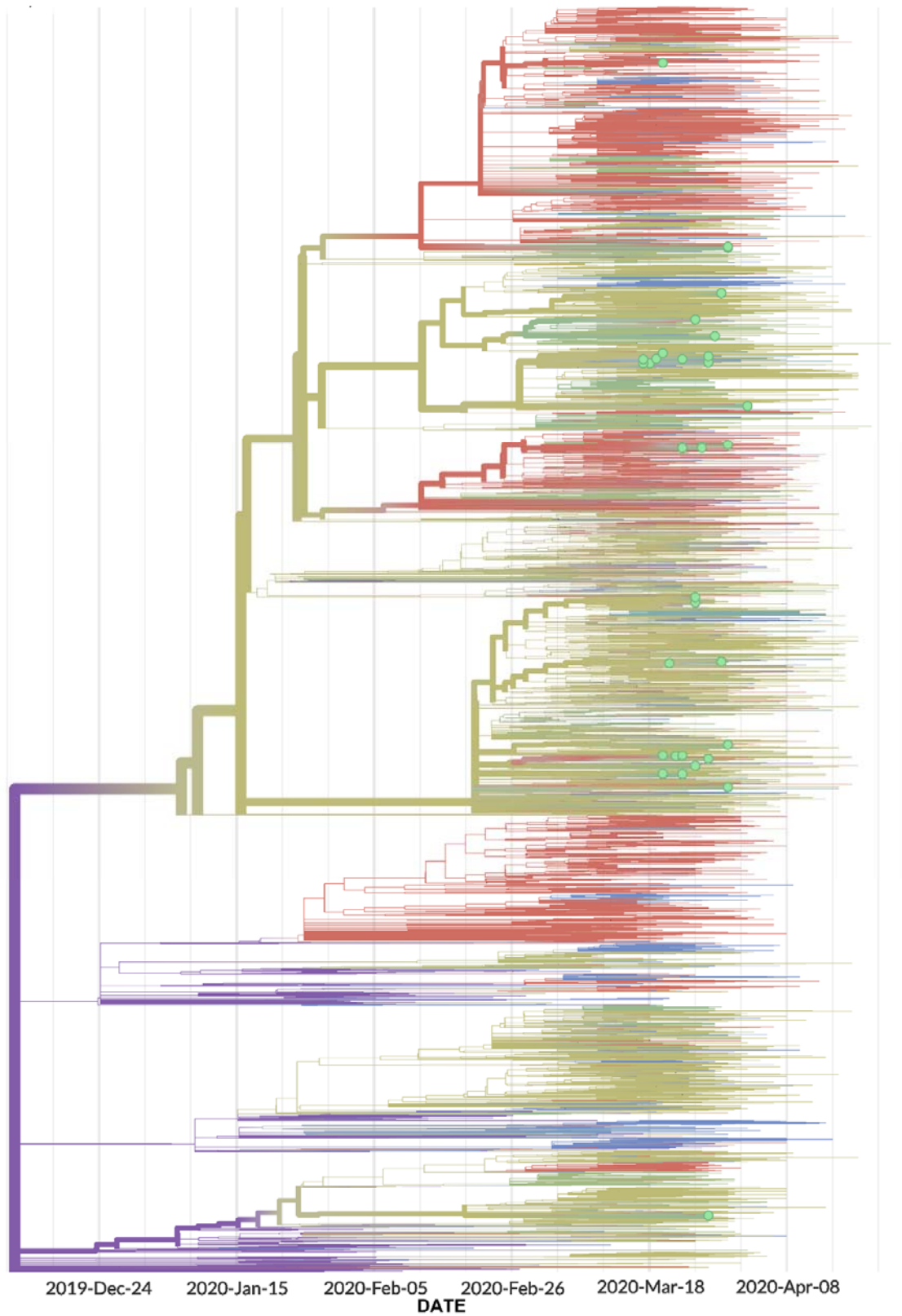
Degree distribution



Supplementary Figure 1. | Degree distribution representing the haplotype network analysis



Supplementary Fig. 2 | A: Time calibrated phylogenetic and phylogeographic visualization of 105 complete SARS-CoV-2 genomes compared to the 32 Hungarian strains. Sequences of this study are highlighted in bold face, colored clades are representing Nextstrain analysis clustering. **B:** The map is a visualization of the sequences presented at the phylogenetic tree.



Supplementary Fig. 3 | Visualization of Hungarian sequence dataset with Nextstrain local workflow. Showing 35 (32 from this manuscript and three additional from GISAID database)

Hungarian sequences compared to 10,869 genomes sampled between Mar 2020 and Apr 2020.