Sequencing SARS-CoV-2 in Slovakia: An Unofficial Genomic Surveillance Report

Broňa Brejová[1], Viktória Hodorová[2], Kristína Boršová[3,2], Viktória Čabanová[3], Tomáš Szemes[4,5], Matej Mišík[6],
Boris Klempa[3], Jozef Nosek[2], Tomáš Vinař[1]

[1] Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia
[2] Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia
[3] Biomedical Research Center of the Slovak Academy of Sciences, Bratislava, Slovakia
[4] Comenius University Science Park, Bratislava, Slovakia
[5] Geneton Ltd., Ilkovičova 8, Bratislava, Slovakia
[6] Institute of Health Analyses, Ministry of Health, Slovakia

Abstract: We present an unofficial SARS-CoV-2 genomic surveillance report from Slovakia based on approximately 3500 samples sequenced between March 2020 and May 2021. Early samples show multiple independent imports of SARS-CoV-2 from other countries. In Fall 2020, three virus variants (B.1.160, B.1.1.170, B.1.258) dominated as the number of cases increased. In November 2020, B.1.1.7 (alpha) variant was introduced in Slovakia and quickly became the most prevalent variant in the country ($> 75\%$ of new cases by early February 2021 and $> 95\%$ in mid-March).

1 Introduction

Genome sequence of the SARS-CoV-2 virus continually changes over time. The mutations eventually result in the emergence of new variants including those with higher infectivity, the ability to evade the immune system response, and causing milder or more severe clinical manifestations. It is therefore of utmost importance to constantly monitor virus alterations by genome sequencing. Such monitoring provides a means for understanding virus evolution and transmission, identification and characterisation of variants of concern (VoC), improvement of the tools for molecular diagnostics (e.g. RT-qPCR assays), as well as rapid adjustment of the health policy measures.

By the end of June 2021, global sequencing efforts yielded more than 2 millions genome sequences of the SARS-CoV-2 isolates from different geographical regions of the world. These sequences are available in public databases such as the GISAID initiative (`http://www.gisaid.org/`) and the European Nucleotide Archive (ENA, `https://www.ebi.ac.uk/ena/`) which allow rapid data sharing and provide robust resources for genomic epidemiology.

In this report, we summarize the results of the SARS-CoV-2 genomic surveillance in Slovakia during the first and second wave of the COVID-19 pandemic between March 2020 and May 2021.

2 Overview of Genomic Surveillance in Slovakia

The first SARS-CoV-2 samples in Slovakia were sequenced in March of 2020 by Comenius University Science Park using viral RNA amplified in VERO E6 cells and Illumina sequencing platform. Majority of samples in 2020 were sequenced using Oxford Nanopore MinION using the ARTIC PCR-tiling protocol originally developed for sequencing of Ebola and Zika virus samples (Quick et al., 2016, 2017), evaluating a variety of primer pools in the process (Brejova et al., 2021a). In March 2021, a consortium of laboratories formed a genomic surveillance team that started routine sequencing of SARS-CoV-2 from clinical samples. The samples for sequencing are selected by the Public Health Authority of Slovakia and distributed to individual sequencing laboratories. Two laboratories (Public Health Authority and Comenius University Science Park) use Illumina sequencing, and Biomedical Centre of Slovak Academy of Sciences in collaboration with Comenius University in Bratislava is using MinION sequencing. All groups use variations of the ARTIC PCR-tiling protocol. Additional samples were sequenced at the Veterinary University in Zvolen and outside Slovakia (in Austria and Germany). Table 1 summarizes these efforts. Additional data for genomic surveillance have been obtained through differential qPCR testing designed to distinguish between common SARS-CoV-2 variants (Kováčová et al., 2021).

3 Early Cases (March-June 2020)

The first case of SARS-CoV-2 infection in Slovakia has been documented in Kostolište (Malacky district in western Slovakia) on March 6, 2020, in a 52 year old man (The Slovak Spectator, 2020). The next day, two members of his family were tested positive for the infection (SK-BMC1), including his son who returned from Venice, Italy, and presumably got infected while traveling (Úrad verejného zdravotníctva SR, 2020). In the following days, additional cases were confirmed in unrelated persons in

Table 1: **Overview of the sequenced SARS-CoV-2 samples from Slovakia.** The table shows the number of samples submitted to GISAID with collection dates between March 2020 and March 2021, split by quarter and sequencing team. The numbers in parentheses indicate the median response time in days (from collection to submission).

	Q1 2020	Q2 2020	Q3 2020	Q4 2020	Q1 2021
BMC + Comenius University	1 (196)	8 (175)	11 (37)	63 (39)	591 (14)
CeMM Austria	-	-	-	60 (76)	88 (63)
Charite Berlin	2 (162)	-	-	-	-
Comenius University Science Park	4 (20)	-	-	42 (108)	1086 (27)
Public Health Authority	-	-	-	-	193 (79)
Veterinary Institute Zvolen	-	4 (386)	4 (291)	5 (181)	7 (78)

Table 2: Early sequenced cases from Slovakia with collection dates between March and June 2020.

| ID	Collection date	Pangolin lineage	Location	Mut. count
hCoV-19/Slovakia/ChVir-1996/2020	2020-03	B.1	Bratislava	6
hCoV-19/Slovakia/ChVir-1998/2020	2020-03	B.1	Bratislava	7
hCoV-19/Slovakia/SK-BMC1/2020	2020-03-06	B.1	Kostolište	6
hCoV-19/Slovakia/SK-BMC5/2020	2020-03-06	B.1	Košice	7
hCoV-19/Slovakia/SK-BMC2/2020	2020-03-07	B.1	Bratislava	7
hCoV-19/Slovakia/SK-BMC6/2020	2020-03-08	B.11	Martin	2
hCoV-19/Slovakia/UKBA-212/2020	2020-03-31	B.1.1	Bratislava	7
hCoV-19/Slovakia/UKBA-101/2020	2020-04-03	B.1	Bratislava	11
hCoV-19/Slovakia/UKBA-203/2020	2020-04-06	B.1.1	Žehra	10
hCoV-19/Slovakia/UKBA-204/2020	2020-04-06	B.1.1	Bratislava	11
hCoV-19/Slovakia/UKBA-205/2020	2020-04-13	B.1.1	Žehra	11
hCoV-19/Slovakia/60007-VHU_185/2020	2020-04-15	B.1	Hencovce	9
hCoV-19/Slovakia/60020-VHU_1250/2020	2020-04-24	B.1.1	Martin	8
hCoV-19/Slovakia/60023-VHU1453_2020/2020	2020-04-27	B.40	Petrovany	12
hCoV-19/Slovakia/UKBA-207/2020	2020-04-29	B.1.1	Pezinok	11
hCoV-19/Slovakia/UKBA-208/2020	2020-05-06	B.1.1	Pezinok	8
hCoV-19/Slovakia/60056-VHU_4065/2020	2020-05-15	B.1.1	Martin	8
hCoV-19/Slovakia/UKBA-209/2020	2020-06-30	B.1.1.70	Bratislava	16
hCoV-19/Slovakia/UKBA-210/2020	2020-06-30	B.1.131	Bratislava	12
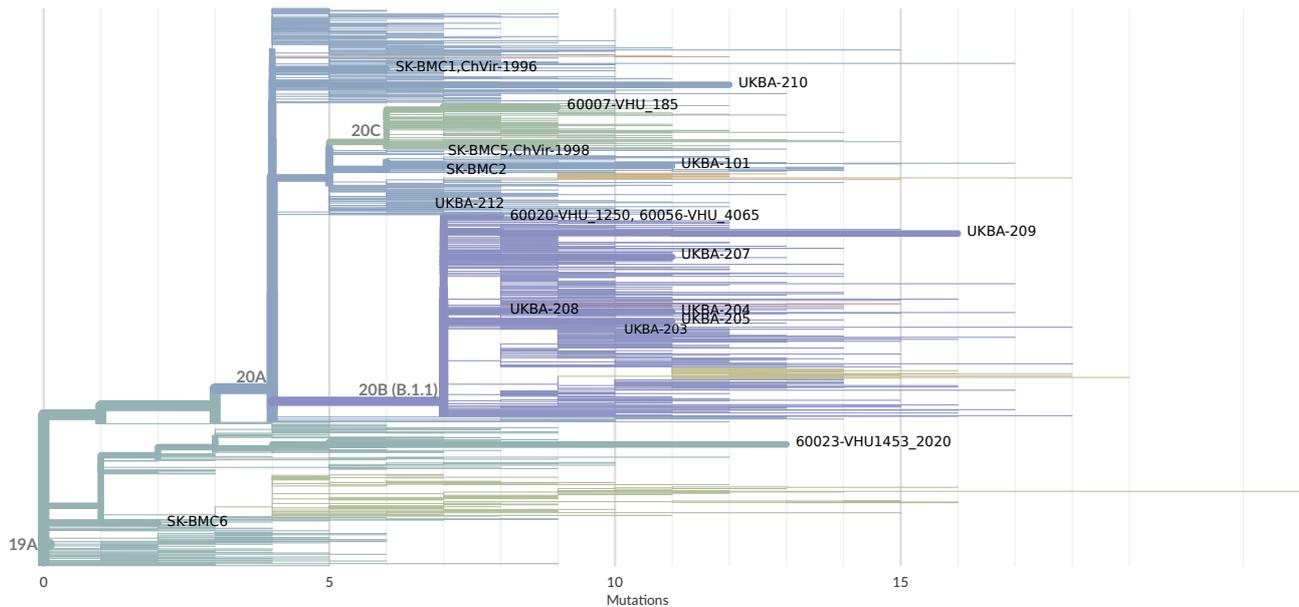
Figure 1: Phylogeny of early cases from Slovakia on the background of randomly chosen samples from GISAID.

Bratislava, Košice, and Martin. Rapid introduction of prevention and control measures including the diagnostics based on real-time quantitative polymerase chain reaction (RT-qPCR), contact tracing, national lockdown, and quarantine for travelers led to substantial reduction of the virus spreading in the country.

The phylogeny of 19 early samples collected between March and June 2020 (Table 2, Figure 1) suggests at least six unrelated import events, likely through routine international travel. The closest matches from GISAID database based on the Jaccard index include samples from France (SK-BMC5), Netherlands (SK-BMC6), Ireland (UKBA-204), Scandinavia (60007-VHU_185), Serbia (UKBA-209), Ukraine (UKBA-210), and the United Kingdom (UKBA-207, UKBA-208). Note that these are locations where particular mutation combinations were common; sampled individuals could have contracted COVID-19 elsewhere. In contrast, sample 60023-VHU1453 shows very little similarity with other GISAID samples.

Three samples (SK-BMC2, SK-BMC5, UKBA-101) were related to the fallout from a documented superspreading event at a medical conference in Boston, MA at the end of February (Lemieux et al., 2021). Sample UKBA-212 is identical to 3757 additional samples from all over the world (including United Kingdom, United States, Portugal, and Italy), which combined an earlier mutation Spike:D614G, which increases the infectivity and stability of virions, leading to higher viral loads and competitive fitness (Plante et al., 2021), with mutations in nucleoprotein N:R203K and N:G204R. The large number of identical samples indicates a very fast spread perhaps following a superspreading event, and the group has become a foundation for evolution of B.1.1 lineage; see also the analysis of Austrian samples (Popa et al., 2020) and the global

analysis of early SARS-CoV-2 lineages (Gómez-Carballa et al., 2020). In Slovakia, nine early samples are classified to the B.1.1 lineage (including sublineage B.1.1.70).

4 Rise in the Fall of 2020

The release of restrictions during the summer 2020 increased the number of imported cases, which has been followed by community transmission. The quick spread of the infection throughout the country with about 5.5 million inhabitants raised the cumulative number of infected people to 267,147 as of December 31, 2020. While in many European countries, B.1.177 (EU1) lineage has become dominant in the Fall of 2020 (Hodcroft et al., 2021), different three lineages appear to have achieved a substantial prevalence in Slovakia between September and November 2020, namely B.1.1.170, B.1.160 (EU2), and B.1.258 (Figure 2).

B.1.1.170 lineage is characterized by mutation P822H (C5184A) in NSP3 peptidase C16 domain required for proteolytic processing of replicase polyprotein. This lineage has been observed in other countries in high numbers, including Denmark, Germany, and United Kingdom; however, in neither of these countries B.1.1.170 represented a significant percentage of cases in the population (Figure 3). Instead, the high absolute number of B.1.1.170 cases reflects the large scale of the sequencing programs in these countries.

B.1.160 (EU2) lineage is characterized by amino acid substitutions M234I and A376T in N, M324I (G9526T) in NSP4, A176S, V767L, K1141R, E1184D in ORF1b, and S477N in S (Fournier et al., 2021). The position 477 in the receptor binding motif of the Spike protein plays a crucial role in the interaction with the human receptor ACE2
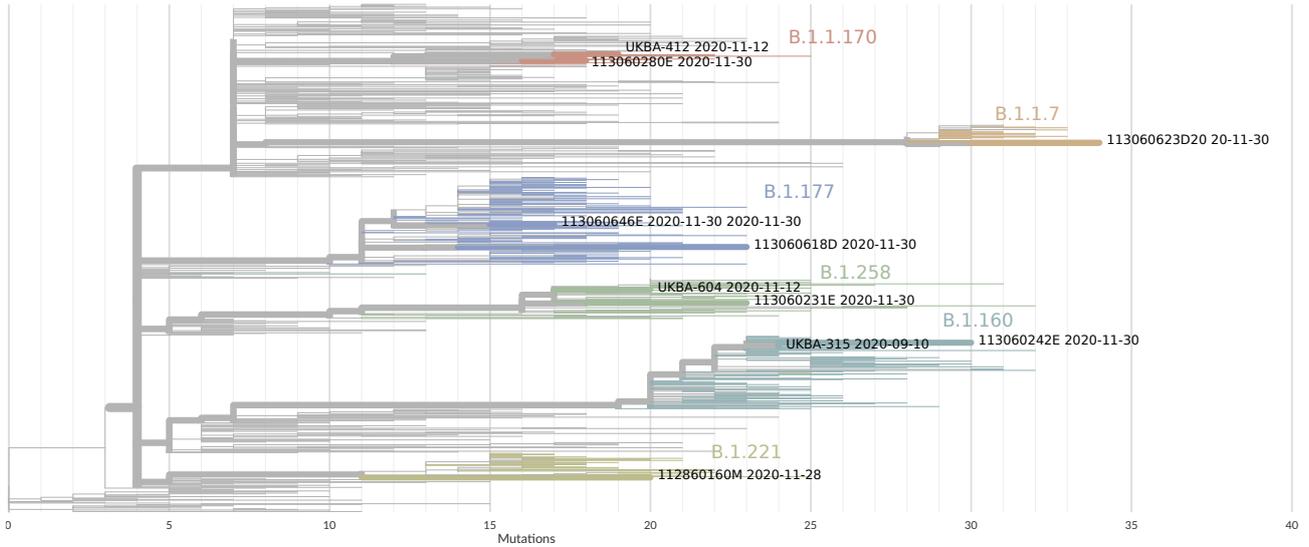
Figure 2: **Selected samples from Slovakia collected between September and December 2020 on the background of randomly chosen samples from GISAID.** Lineages B.1.1.170, B.1.1.7, B.1.258, B.1.160 and B.1.221 are overrepresented among the background samples for better visibility.

Figure 3: **The percentage of SARS-CoV-2 lineages among sequenced samples collected between September and November 2020 in several countries.** The number in parenthesis indicates the total number of sequenced samples deposited in GISAID with collection dates within this time period. The percentage for each studied lineage includes also all its sublineages.

(Singh et al., 2021), and the mutant has been shown to increase the infectivity (Chen et al., 2020b); the same mutation has also emerged in an unrelated outbreak in Australia (Chen et al., 2020a). Besides Slovakia, the B.1.160 was highly prevalent in Hungary, Austria, and Switzerland (Figure 3).

B.1.258 lineage harbours Spike protein receptor binding domain mutation N439K, shown to enhance the binding affinity of the Spike protein to human immune response (Thomson et al., 2021), and the sublineages prevalent in Central Europe combine this mutation with ΔH69/ΔV70 deletion, facilitating escape from immune response (Kemp et al., 2021), which is also one of the characteristic muta-

tions of B.1.1.7 (alpha) lineage emerging later. Lineage B.1.258 likely originated in Switzerland (Brejova et al., 2021b), and spread mainly in Central European countries, including Czech Republic, Slovakia, Poland, Austria, and Germany (Figure 3).

Interestingly, B.1.177 (EU1) lineage, wide-spread across many European countries during Fall of 2020, does not show any evidence of increased transmissibility. It seems to have become dominant simply by repeated introduction to respective countries by summertime travelers, undermining local efforts to keep SARS-CoV-2 cases low (Hodcroft et al., 2021). In Slovakia, this lineage was discovered only at the end of November 2020 and sporadically appeared in sequencing samples since then.

Figure 4 shows how the number of mutations increases over time in selected lineages. Besides lineages shown in Figure 3, we have also included lineages P.1 and B.1.351 which are known for a high number of mutations. Other samples from GISAID were used as a background for the analysis. For each group, we have applied linear regression to estimate the mutation rate, which varied from 0.8 to 1.7 mutations per month with background mutation rate 1.4 mutations per month. Note however that some lineages show low correlation coefficients.

Lineages B.1.177, B.1.1.170 and B.1.221 have a similar number of mutations compared to the background over time, while the remaining lineages exhibit higher mutation counts. Lineages B.1.160 (EU2) and B.1.258 show an interesting evolutionary pattern, where a branch leading to the lineage in the phylogenetic tree is associated with a surge in mutations. After this initial surge, the evolutionary rate stabilizes again at the mutation rate close to the background. Such unusual genetic divergence has also been observed in B.1.1.7 (Rambaut et al., 2020) and
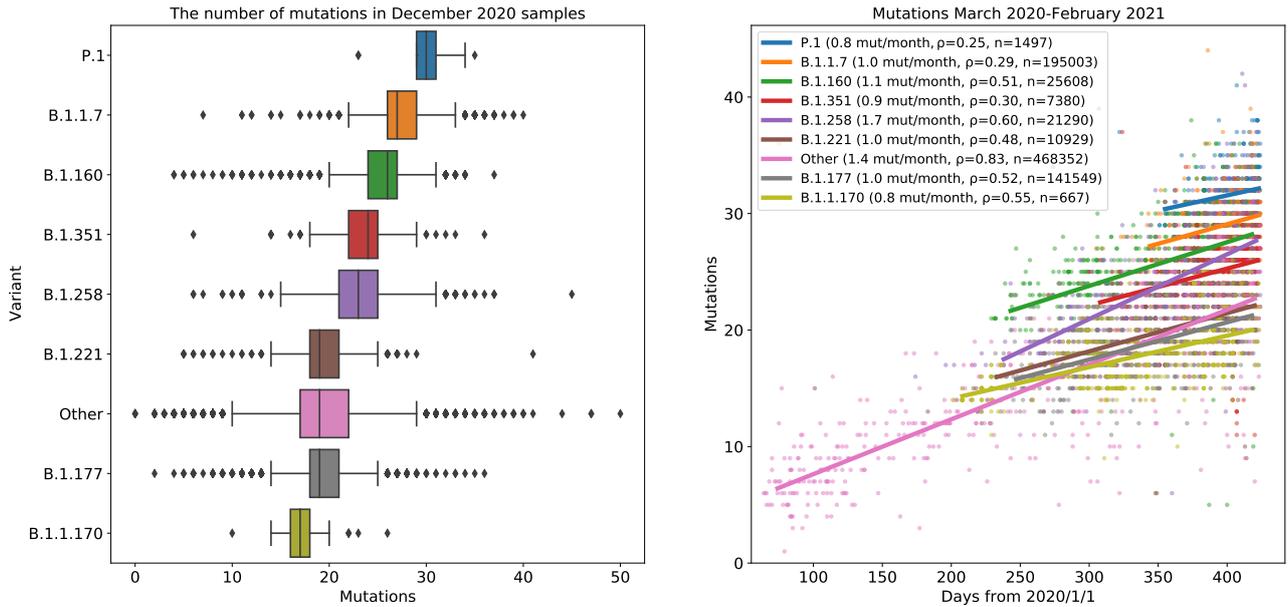
Figure 4: **Evolution of SARS-CoV-2 lineages over time.** (left) The number of mutations compared to the reference virus sequence in samples collected in December 2020. The samples were split according to identified lineage. (right) Accumulation of mutations in individual variants over time. Linear regression shows the base mutation rate of 1.4 substitutions per month (line "other", Pearson correlation 0.83), and mutation rates in selected lineages varying between 0.8 mutations per month in P.1 and 1.7 mutations per month in B.1.258.

other VoCs (Figure 4), and is likely indicative of positive selection. One possible mechanism is a selective pressure upon the within-patient virus population in immunodeficient or immunosuppressed chronically infected patients treated with convalescent plasma and antiviral drugs (Choi et al., 2020; Avanzato et al., 2020; Kemp et al., 2021).

While in Spring of 2020, Slovakia was one of the countries with the best response to the pandemic situation (Serhan, 2020), the Fall was characterized by a steep rise in cases. Besides slow and inconsistent response of government institutions to the worsening situation, the genomic surveillance from that period highlights Slovakia's position at the crossroads of Central Europe, which likely led to importation of new variants from neighbouring countries. Combined effect of these variants, some of which share evolutionary characteristics with later identified variants of concern, likely contributed to worsening of the pandemic situation.

5 Introduction of B.1.1.7

Variant B.1.1.7 (alpha) was first observed on September 20, 2020 in Kent, United Kingdom (Rambaut et al., 2020) and quickly spread throughout the United Kingdom and the world. Out of unusually large number of mutations in the spike protein, substitution N501Y in the receptor-binding domain has been identified to increase the binding affinity to human ACE2 receptors, the ΔH69/ΔV70 deletion increases infectivity and mediates cell-cell fusions,

ΔY144 is localized in an antibody supersite epitope, and P681H is adjacent to biologically significant furin cleavage site (Meng et al., 2021; Gupta, 2021).

The first documented case in Slovakia was collected in Bratislava on November 30, 2020. Due to the low volume of sequencing at the time, it is difficult to estimate the prevalence and the spread of the B.1.1.7 variant over time. However, antigen mass testing in the city of Trenčín in Western Slovakia on December 19-20, followed by qPCR re-testing on a voluntary basis, yielded 148 PCR-positive samples collected on December 22 (Mesto Trenčín, 2021), out of which 122 were later re-tested using tests designed to differentiate between B.1.1.7 and B.1.258 (both carrying ΔH69/ΔV70), and variants that do not harbour this deletion (Kováčová et al., 2021). Out of these, 4 samples (or 3.3%) were identified as B.1.1.7, and this was also confirmed by sequencing of selected samples (Brejova et al., 2021b). The fraction of B.1.1.7 cases in the city of Trenčín has quickly risen to 76% ($n = 21$) in as little as 42 days, according to the nation-wide differential qPCR testing on February 3, 2021 (see `https://github.com/Institut-Zdravotnych-Analyz/covid19-data`).

Interestingly, the increase from 3.3% to 76% in 42 days is fast in comparison with other countries. Among the countries with a high number of sequenced samples, Denmark took 48 days to rise from $\approx 3.3\%$ to $\approx 76\%$ (between December 29 and February 15, based on at least $n = 1492$ samples in each 14-day sliding window), in United Kingdom and Switzerland, a similar rise in the fraction of cases took 61 days (October 31-December 31, $n = 8307$)

and 62 days (December 19-February 19, $n = 597$) respectively, and in Germany it took even longer (77 days between December 17 and March 4, $n = 304$). While it is difficult to speculate on why the B.1.1.7 was able to achieve domination in Slovakia so quickly, it is worthwhile to point out that during this period, Slovakia was under various forms of nation-wide lockdown. It has been demonstrated that effectiveness of lockdown measures differs between old variants and B.1.1.7 (Vöhringer et al., 2020). Also, a fatigue from following the rules and inconsistencies in the government imposed interventions likely caused people to selectively choose to follow certain rules while rejecting others, based mostly on their own experience. Such an approach may have increased the lineage-based differences in effectiveness of mitigation.

One possible explanation for fast spread of B.1.1.7 in Slovakia are repeated imports by workers and students visiting home during the Christmas holidays. In fact, one of the first B.1.1.7 outbreaks detected in Slovakia was in a marginalized community in the Eastern Slovakia (Pavlovce nad Uhom), where the link to travel from the United Kingdom was clearly established. However, we show below that this may not be the main factor.

Interestingly, 74% of B.1.1.7 sequenced cases collected in Slovakia between November 2020 and May 2021, form a separate clade in the phylogenetic tree (called B.1.1.7ce for the purpose of this paper), characterized by mutations C5944T and G28884C (Figures 5 and 6). While the first of these is silent, the second causes R204P substitution in the nucleoprotein IDR2 region. Note that this site was mutated from G to R at the base of lineage B.1.1. The G28884C mutation also extends an existing span of three consecutive mutations at positions 28881-28883 compared to the reference, this region being characterized as a mutational hotspot of the N protein (Azad, 2021). Additional mutations A28095T and T15096C likely happened after the emergence of B.1.1.7, but prior to the characteristic mutations C5944T and G28884C. The former mutation introduces a stop codon K68* in gene ORF8; mutations in ORF8 being potentially linked to immune evasion (Zhang et al., 2021). The earliest case in GISAID belonging to B.1.1.7ce sublineage has been collected in Switzerland on November 9, 2020; the early cases from Bratislava (November 30, 2020) and Trenčín (December 22, 2020) also belong to the ce sublineage. Besides Slovakia, sublineage B.1.1.7ce represented 91% of sequenced B.1.1.7 samples from the Czech Republic, 69% in Hungary, and 57% in Austria. (Surprisingly Colombia and French Guiana also showed over 40% cases of B.1.1.7 belonging to the B.1.1.7ce sublineage, but the number of sequenced genomes are quite small, and thus the sample may not be representative). In contrast, B.1.1.7ce constituted only 0.3% of B.1.1.7 samples in the United Kingdom.

Phylogenetic tree reconstruction from a randomly selected subset of B.1.1.7 cases from Slovakia, Czech Republic, and Austria (Figure 6) shows a large radiation at the base of B.1.1.7ce clade, which suggests a rapid spread of this clade before further mutations had a chance to accumulate. High percentage of these samples in Central European countries contradicts the theory of repeated introduction by independent travelers from the United Kingdom (with only 0.3% of B.1.1.7ce cases out of all B.1.1.7 cases), and instead suggests a fast community spread directly within Central Europe. This theory is further supported by the data from four screenings by differential qPCR tests performed between February 3 and March 17, 2021, initially showing high percentage of B.1.1.7 cases in Western Slovakia and spreading over time to the eastern parts of the country (Figure 7; Kováčová et al. (2021)).

Lineage B.1.1.7 constitutes 96.7% of 3184 GISAID samples collected in Slovakia between February and May 2021. Other lineages previously present in Slovakia, such as B.1.258, B.1.160, B.1.1.170, B.1.177, constitute 54 samples in total (1.7%). Other lineages occured sporadically and in many cases have been linked directly to international travel with only a limited community spread. Lineage B.1.351 (VoC Beta) was found in 27 samples, and 25 samples belong to 12 additional lineages, including ECDC variants of interest B.1.617.1 (Kappa) and B.1.621 (European Centre for Disease Prevention and Control, 2021a).

6 Discussion and Conclusions

Throughout 2020, genomic surveillance of COVID-19 pandemic in Slovakia consisted almost exclusively from uncoordinated activities of individual researchers, which has resulted in highly uneven sequencing coverage, both in time and regionally. Nevertheless, the information collected has provided a unique insight into a progression of COVID-19 pandemic in Slovakia and allowed us to identify both common and unique trends compared to the neighbouring countries. The situation changed in March 2021 with the establishment of coordinated efforts involving the Public Health Authority, Comenius University, and Biomedical Center of Slovak Academy of Sciences. Since then, the Public Health Authority has been selecting and distributing positively tested PCR samples to individual labs for sequencing and the number of sequenced samples typically exceeded 500 samples per week in June 2021.

Yet, there is space for improvements. A major problem currently lies with the logistics, where samples are delivered to the sequencing labs two weeks or longer after their collection. Depending on the laboratory and sequencing technology used, the time from receiving samples to sequencing results can be as short as two days or as long as one week. The information obtained through sequencing is thus much delayed and has only a limited value for treatment, epidemiological response, or as the basis for rapid public policy decisions. Examples from Denmark, United Kingdom, Netherlands, and other countries (see e.g. (Oude Munnink et al., 2020)) show that these logistic issues can be solved and the response time can be decreased dramatically. In fact, governments in these countries routinely use
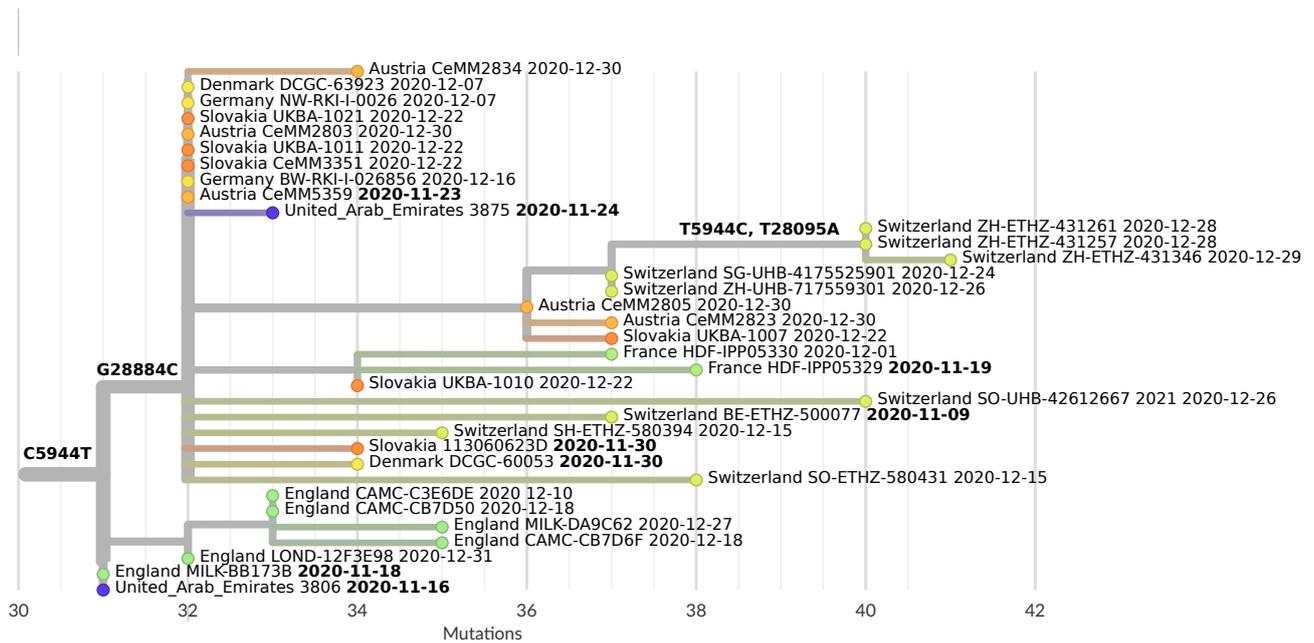
Figure 5: **Early cases of B.1.1.7ce clade.** Collection date is shown for November samples. Note that this particular reconstruction includes apparent back mutations at positions 5944 and 28095. However, these may be caused by data processing artefacts as these positions are within the ARTIC v3 primer binding sites, which may render them invisible to some computing pipelines typically used for processing SARS-CoV-2 sequencing data.
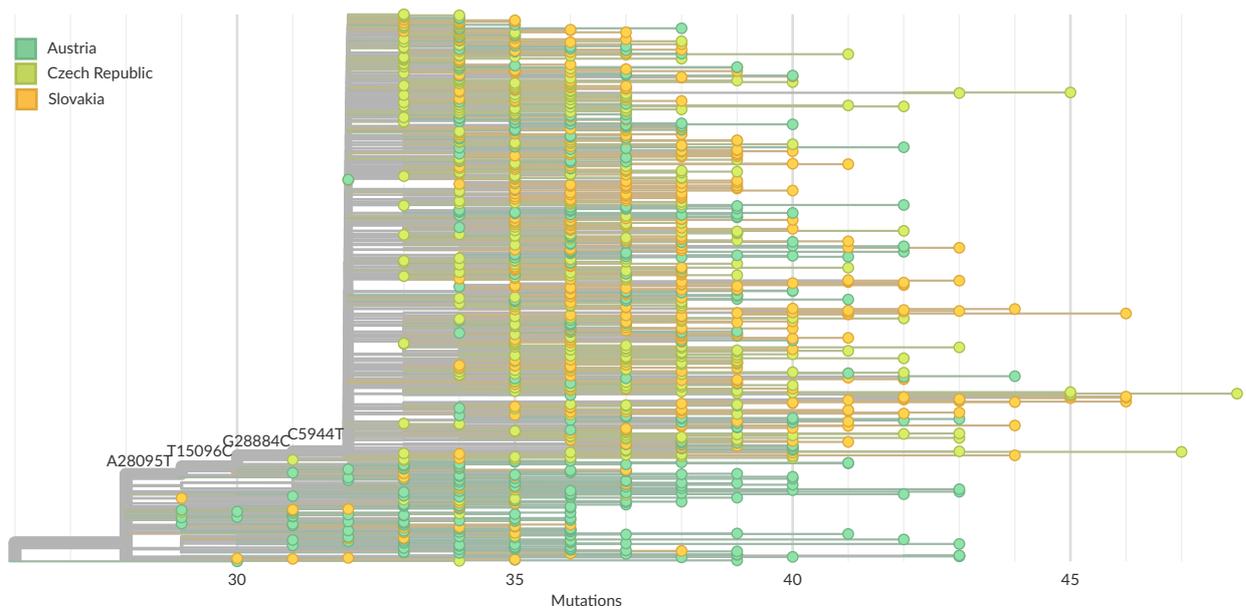

Figure 6: A randomly selected subset of B.1.1.7 samples from Slovakia, Czech Republic and Austria.
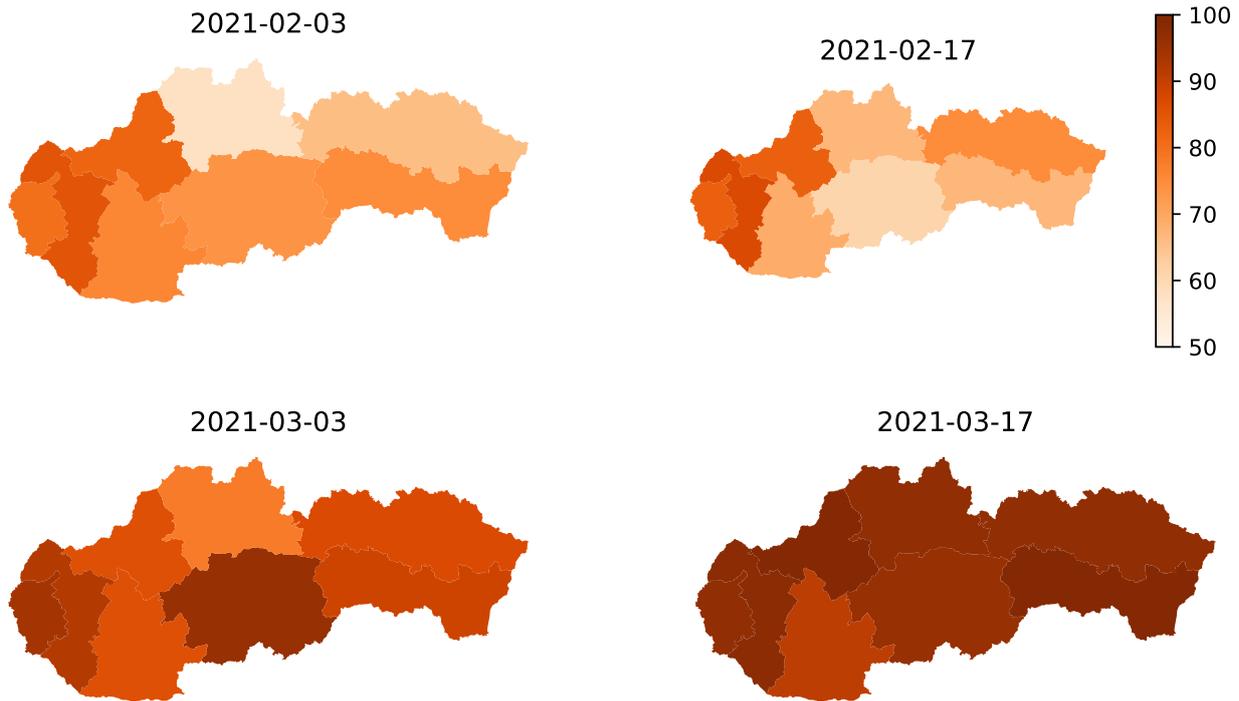
Figure 7: **Percentage of B.1.1.7 among samples positively tested by qPCR in different regions of Slovakia.** All samples positively tested on these dates were re-tested with differential qPCR tests designed to distinguish B.1.1.7, B.1.258, and variants that do not harbour ΔH69/ΔV70 deletion (data: Institute of Health Analyses).

the information obtained through sequencing to fine-tune the pandemic mitigation measures.

According to the recommendations from the ECDC (European Centre for Disease Prevention and Control, 2021b), in choosing samples for sequencing, the priority should be given to the representative sampling for the purpose of surveillance of emerging variants (even those that are not yet characterized as VoCs). This can be combined with targeted monitoring of outbreaks, vaccine escape and reinfection, long-term persistent infections, monitoring of travel, etc. For the purpose of data analysis, it is essential that the reasons for choosing a particular sample for sequencing is known to the researchers analyzing the data; yet this information is not provided by the Public Health Authority, and data analysts have no influence in developing the sampling strategy. Based on recently improving epidemiological situation, there is currently an ambition to sequence all samples with sufficient viral load. Nevertheless, this issue is likely to reappear once the situation worsens and the selection of samples is again necessary.

While some of the analyses requiring integration of epidemiology and genomics data are conceptually straightforward (such as monitoring the prevalence of lineages over time), other tasks, such as recognizing mutations spreading due to a selective advantage rather than a random drift, are much more involved and require complex bioinformatics and modeling expertise (see, e.g. (Vöhringer et al., 2020)). At present, we are not aware of any plans of

establishing a team that would have such an expertise, enough redundancy to perform such analyses regularly, unobstructed access to all necessary data, and regular communication with experts elsewhere on these matters.

7   Methods

SARS-CoV-2 genomic sequences and their metadata (including date of collection and submission, country, and Pangolin lineage) were downloaded from GISAID (Shu and McCauley, 2017) on June 18, 2021. The database contained 2,012,564 sequences. Out of these, we have used 1,950,347 sequences with fewer than 5kbp of missing sequence.

The presence of individual substitutions was ascertained by mapping individual genomes to the reference hCoV-19/Wuhan/Hu-1/2019 by minimap2 (Li, 2018) and then formatting the result into a multiple alignment in reference sequence coordinates by gofasta tool (Jackson, 2020). To count mutations for Figure 4, each continuous stretch of mutated bases was counted as a single mutation to mitigate impact of occasional local misalignments. In this figure, only sequences with fully specified date, with at most 1kb of missing sequence and at most 50 mutations were used. Sequences with more than 50 mutations were rare in the displayed period. All sublineages of each displayed lineage were included within this lineage. All samples were

used to estimate linear regression, but at most 500 samples (randomly selected) are displayed in the plot per lineage.

Phylogenetic trees were created by Augur and visualized by Auspice (Hadfield et al. 2018; Sagulenko, Puller, and Neher 2018). The tree in Figure 1 contains all samples collected in Slovakia before July 2020. As a background, we have selected 10% of samples from other countries sequenced before March 2020 and 1% of samples selected between March and June 2020 (inclusive). We have removed all samples with more than 25 mutations compared to the reference as likely metadata errors (samples from early 2021 mistakenly marked as 2020). Finally the set of 1830 background samples was reduced to 1199 by removing samples that differed by the presence or absence of at most one mutation from some older sample. The tree in Figure 2 highlights manually selected representative Slovak samples from lineages described in the text. The background sequences were randomly selected with probably 0.2% from samples sequenced before the end of November 2020 with at most 40 mutations compared to the reference. To give a better context for the selected samples, some lineages were overrepresented in the background set. Namely, 10% of samples from B.1.1.170 were added, as well as 1% of samples from B.1.160, B.1.1.7, B.1.221, B.1.258; in both cases using only samples from September to November 2020. Again, the background set was reduced by the removal of samples differing by at most one mutation. The tree in Figure 5 is a clade selected from a bigger tree, which contained three outgroup sequences (reference hCoV-19/Wuhan/Hu-1/2019, an early B.1.1 sample hCoV-19/Slovakia/UKBA-212/2020, and an early B.1.1.7 sample hCoV-19/England/MILK-9E05B3/2020), 968 B.1.1.7 samples collected in 2020 and containing the A28095T mutation (out of all 2990 samples satisfying these criteria, we have again filtered out nearly identical sequences). We have also added all B.1.1.7 sequences from 2020 that contain at least one of the mutations G28884C and C5944T characteristic for B.1.1.7ce clade. We have excluded 8 sequences that appeared as outliers markedly different from other sequences, possibly due to recombination or technical errors. Finally, the tree in Figure 6 contains a selection of B.1.1.7 sequences from Austria, Czech Republic and Slovakia collected between September 2020 and May 2021, excluding samples marked as environmental as well as sequences from Slovakia lacking G28882A mutation. Many samples without this mutation were creating a spurious clade; we believe that the lack of this mutation is due to technical problems with calling the four successive variants 28881-28884 in B.1.1.7ce using very short Illumina reads. From the remaining samples, we have taken 25% of sequences from Slovakia and Austria and 14% from the Czech Republic. After again filtering out nearly identical sequences in each country separately, we were left with 361 Austrian, 382 Czech and 371 Slovak sequences. We added the reference hCoV-19/Wuhan/Hu-1/2019 as an outgroup and removed three outliers.

References

Avanzato, V. A., Matson, M. J., Seifert, S. N., Pryce, R., Williamson, B. N., Anzick, S. L., Barbian, K., Judson, S. D., Fischer, E. R., Martens, C., Bowden, T. A., Wit, E. d., Riedo, F. X., and Munster, V. J. (2020). Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell*, 183(7):1901–1912.e9.

Azad, G. K. (2021). The molecular assessment of SARS-CoV-2 Nucleocapsid Phosphoprotein variants among Indian isolates. *Heliyon*, 7(2).

Brejova, B., Borsova, K., Hodorova, V., Cabanova, V., Gafurov, A., Fricova, D., Nebohacova, M., Vinar, T., Klempa, B., and Nosek, J. (2021a). Nanopore Sequencing of SARS-CoV-2: Comparison of Short and Long PCR-tiling Amplicon Protocols. doi:10.1101/2021.05.12.21256693. medRxiv.

Brejova, B., Hodorová, V., Boršová, K., Čabanová, V., Reizigová, L., Paul, E. D., Čekan, P., Klempa, B., Nosek, J., and Vinař, T. (2021b). B.1.258Δ, a SARS-CoV-2 variant with ΔH69/ΔV70 in the Spike protein circulating in the Czech Republic and Slovakia. *arXiv:2102.04689 [q-bio]*. arXiv: 2102.04689.

Chen, A. T., Altschuler, K., Zhan, S. H., Chan, Y. A., and Deverman, B. E. (2020a). COVID-19 CG: Tracking SARS-CoV-2 mutations by locations and dates of interest. doi:10.1101/2020.09.23.310565. bioRxiv.

Chen, J., Wang, R., Wang, M., and Wei, G.-W. (2020b). Mutations Strengthened SARS-CoV-2 Infectivity. *Journal of Molecular Biology*, 432(19):5212–5226.

Choi, B., Choudhary, M. C., Regan, J., Sparks, J. A., Padera, R. F., Qiu, X., Solomon, I. H., Kuo, H.-H., Boucau, J., Bowman, K., Adhikari, U. D., Winkler, M. L., Mueller, A. A., Hsu, T. Y.-T., Desjardins, M., Baden, L. R., Chan, B. T., Walker, B. D., Lichterfeld, M., Brigl, M., Kwon, D. S., Kanjilal, S., Richardson, E. T., Jonsson, A. H., Alter, G., Barczak, A. K., Hanage, W. P., Yu, X. G., Gaiha, G. D., Seaman, M. S., Cernadas, M., and Li, J. Z. (2020). Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *New England Journal of Medicine*, 383(23):2291–2293.

European Centre for Disease Prevention and Control (2021a). SARS-CoV-2 variants of concern as of 24 June 2021. `https://www.ecdc.europa.eu/en/covid-19/variants-concern`.

European Centre for Disease Prevention and Control (2021b). Sequencing of SARS-CoV-2 - first update. `https://www.ecdc.europa.eu/en/publications-data/sequencing-sars-cov-2`.

Fournier, P.-E., Colson, P., Levasseur, A., Devaux, C. A., Gautret, P., Bedotto, M., Delerce, J., Brechard, L., Pinault, L., Lagier, J.-C., Fenollar, F., and Raoult, D. (2021). Emergence and outcomes of the SARS-CoV-2 'Marseille-4' variant. *International Journal of Infectious Diseases*, 106:228–236.

Gupta, R. K. (2021). Will SARS-CoV-2 variants of concern affect the promise of vaccines? *Nature Reviews Immunology*, 21(6):340–341.

Gómez-Carballa, A., Bello, X., Pardo-Seco, J., Martinón-Torres, F., and Salas, A. (2020). Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. *Genome Research*, 30(10):1434–1448.

Hodcroft, E. B., Zuber, M., Nadeau, S., Vaughan, T. G., Crawford, K. H. D., Althaus, C. L., Reichmuth, M. L., Bowen, J. E., Walls, A. C., Corti, D., Bloom, J. D., Veesler, D., Mateo, D., Hernando, A., Comas, I., Candelas, F. G., Stadler, T., and Neher, R. A. (2021). Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature*, pages 1–9.

Jackson, B. (2020). cov-ert/gofasta. `https://github.com/cov-ert/gofasta`.

Kemp, S. A., Collier, D. A., Datir, R. P., Ferreira, I. A. T. M., Gayed, S., Jahun, A., Hosmillo, M., Rees-Spear, C., Mlcochova, P., Lumb, I. U., Roberts, D. J., Chandra, A., Temperton, N., Sharrocks, K., Blane, E., Modis, Y., Leigh, K. E., Briggs, J. A. G., Gils, M. J. v., Smith, K. G. C., Bradley, J. R., Smith, C., Doffinger, R., Ceron-Gutierrez, L., Barcenas-Morales, G., Pollock, D. D., Goldstein, R. A., Smielewska, A., Skittrall, J. P., Gouliouris, T., Goodfellow, I. G., Gkrania-Klotsas, E., Illingworth, C. J. R., McCoy, L. E., and Gupta, R. K.

(2021). SARS-CoV-2 evolution during treatment of chronic infection. *Nature*, 592(7853):277–282.

Kováčová, V., Boršová, K., Paul, E. D., Radvánszka, M., Hajdu, R., Čabanová, V., Sláviková, M., Ličková, M., Lukáčiková, L., Belák, A., Roussier, L., Kostičová, M., Líšková, A., Maďarová, L., Štefkovičová, M., Reizigová, L., Nováková, E., Sabaka, P., Koščálová, A., Brejová, B., Staroňová, E., Mišík, M., Vinař, T., Nosek, J., Čekan, P., and Klempa, B. (2021). Surveillance of SARS-CoV-2 lineage B.1.1.7 in Slovakia using a novel, multiplexed RT-qPCR assay. doi:10.1101/2021.02.09.21251168. medRxiv.

Lemieux, J. E., Siddle, K. J., Shaw, B. M., Loreth, C., Schaffner, S. F., et al. (2021). Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of super-spreading events. *Science*, 371(6529).

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100.

Meng, B., Kemp, S. A., Papa, G., Datir, R., Ferreira, I. A. T. M., et al. (2021). Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Reports*, 35(13).

Mesto Trenčín (2021). Plošné testovanie v Trenčíne odhalilo vyše 500 pozitívnych. `https://trencin.sk/aktuality/plosne-testovanie-v-trencine-odhalilo-vyse-500-pozitivnych/`.

Oude Munnink, B. B., Nieuwenhuijse, D. F., Stein, M., O'Toole, A., Haverkate, M., Mollers, M., Kamga, S. K., Schapendonk, C., Pronk, M., Lexmond, P., van der Linden, A., Bestebroer, T., Chestakova, I., Overmars, R. J., van Nieuwkoop, S., Molenkamp, R., van der Eijk, A. A., GeurtsvanKessel, C., Vennema, H., Meijer, A., Rambaut, A., van Dissel, J., Sikkema, R. S., Timen, A., and Koopmans, M. (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med*, 26(9):1405–1410.

Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfias, C. R., Mirchandani, D., Scharton, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A. N., Menachery, V. D., Xie, X., Plante, K. S., Weaver, S. C., and Shi, P.-Y. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 592(7852):116–121.

Popa, A., Genger, J.-W., Nicholson, M. D., Penz, T., Schmid, D., Aberle, S. W., Agerer, B., Lercher, A., Endler, L., Colaço, H., Smyth, M., Schuster, M., Grau, M. L., Martínez-Jiménez, F., Pich, O., Borena, W., Pawelka, E., Keszei, Z., Senekowitsch, M., Laine, J., Aberle, J. H., Redlberger-Fritz, M., Karolyi, M., Zoufaly, A., Maritschnik, S., Borkovec, M., Hufnagl, P.,

Nairz, M., Weiss, G., Wolfinger, M. T., Laer, D. v., Superti-Furga, G., Lopez-Bigas, N., Puchhammer-Stöckl, E., Allerberger, F., Michor, F., Bock, C., and Bergthaler, A. (2020). Genomic epidemiology of super-spreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Science Translational Medicine*, 12(573).

Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols*, 12(6):1261–1276.

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589):228–232.

Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D. L., Volz, E., and COVID-19 Genomics Consortium UK (2020). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. virological.org.

Serhan, Y. (2020). Lessons From Slovakia—Where Leaders Wear Masks. `https://www.theatlantic.com/international/archive/2020/05/slovakia-mask-coronavirus-pandemic-success/611545/`. The Atlantic.

Shu, Y. and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22(13):30494.

Singh, A., Steinkellner, G., Köchl, K., Gruber, K., and Gruber, C. C. (2021). Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2. *Scientific Reports*, 11(1):4320.

The Slovak Spectator (2020). Coronavirus confirmed in Slovakia. `https://spectator.sme.sk/c/22351796/coronavirus-confirmed-in-slovakia.html`.

Thomson, E. C., Rosen, L. E., Shepherd, J. G., Spreafico, R., da Silva Filipe, A., et al. (2021). Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell*, 184(5):1171–1187.e20.

Vöhringer, H., Sinnott, M., Amato, R., Martincorena, I., Kwiatkowski, D., Barrett, J. C., and Gerstung, M. (2020). Lineage-specific growth of SARS-CoV-2 B.1.1.7 during the English national lockdown. virological.org.

Zhang, Y., Chen, Y., Li, Y., Huang, F., Luo, B., Yuan, Y., Xia, B., Ma, X., Yang, T., Yu, F., Liu, J., Liu, B., Song, Z., Chen, J., Yan, S., Wu, L., Pan, T., Zhang, X., Li, R., Huang, W., He, X., Xiao, F., Zhang, J., and Zhang, H. (2021). The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. *Proceedings of the National Academy of Sciences*, 118(23):e2024202118.

Úrad verejného zdravotníctva SR (2020). COVID-19: 389 vzoriek je negatívnych, 3 potvrdené prípady. `https://www.uvzsr.sk/index.php?id=4065`.