Investigating the first wave of the COVID-19 pandemic in Ukraine using epidemiological and genomic sequencing data

Yuriy Gankin[a,*], Vladimir Koniukhovskii[b,*], Alina Nemira[c,*], Gerardo Chowell[e], Thomas A. Weppelmann[d], Pavel Skums[c,**], Alexander Kirpich[e,**]

[a] *Quantori, Cambridge, Massachusetts, United States of America*
[b] *EPAM Systems, Saint Petersburg, Russian Federation*
[c] *Department of Computer Science, Georgia State University, Atlanta, Georgia, United States of America*
[d] *Department of Internal Medicine, University of South Florida, Tampa, Florida, United States*
[e] *Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, Georgia, United States of America*

Abstract

The novel coronavirus SARS-CoV-2 emerged in China in December 2019 and has rapidly spread around the globe. The World Health Organization declared COVID-19 a pandemic in March 2020 just three months after the introduction of the virus. Individual nations have implemented and enforced with varying degrees of success a variety of social distancing interventions to slow the virus spread. Investigating the role of non-pharmaceutical interventions on COVID-19 transmission in different settings is an important research. While most transmission modeling studies have focused on the dynamics in China, neighboring Asian counties, Western Europe, and North America, there is a scarcity of studies for Eastern Europe. This study starts to fill this gap by analyzing the characteristics of the *first epidemic wave* in Ukraine using mathematical and statistical models together with epidemiological and genomic sequencing data. Using an agent-based model, the trajectory of the first wave in terms of cases and deaths and explore the im-

*Equal Contribution
**Equal Contribution
Corresponding Author: Alexander Kirpich
Email address: akirpich@gsu.edu (Alexander Kirpich)

pact of quarantine strategies via simulation studies have been characterized. The implemented stochastic model for epidemic counts suggests, that even a small delay of weeks could have increased the number of cases by up to 50%, with the potential to overwhelm hospital systems. The genomic data analysis suggests that there have been multiple introductions of SARS-CoV-2 into Ukraine during the early stages of the epidemic with *eight distinct* transmission clusters identified. The basic reproduction number for the epidemic has been estimated independently both from case counts data and from genomic data. The findings support the hypothesis that, the public health measures did not have a decreasing effect on the existing viral population number at the time of implementation, since strains were detected after the quarantine date. However, the public health measures did help to prevent the appearance of new (and potentially more virulent) SARS-CoV-2 variants in Ukraine.

Keywords: SARS-CoV-2, agent-based, COVID-19, Ukraine, phylogenetics, phylodynamics

---

1. Introduction

The COVID-19 pandemic started in December 2019 in the Chinese city of Wuhan located within the province of Hubei [1][2][3][4][5]. Despite the efforts of Chinese authorities, the virus rapidly spread throughout China, Asia, Europe, and the Americas. The World Health Organization (WHO) declared the spread of the novel coronavirus a pandemic in March, 2020 [6] just three month after the first reported case. The virus would later be named SARS-CoV-2 and the resulting spectrum of disease COVID-19 [7]. Despite having much lower case-fatality rate than the severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), the novel coronavirus has already claimed more lives than both combined [8]. As of March 1, 2021 there were more than 114 million infections with over 2.5 million deaths [9]. Though multiple treatments have been proposed and vaccination has now begun in earnest, during the early pandemic period, social distancing interventions were the only tools at the disposal of public health authorities to prevent and mitigate transmission [10][11][12]. The attempts to implement and enforce such measures by governments around the world were highly variable and included frequent sanitation of public spaces, enforced social distancing, wearing of masks, and closure of schools,

churches, and other mass gatherings [13] [14][15].

Even well after a year since the epidemic started, fundamental questions regarding the effects of isolation and lockdown measures [16][17] and the genomic evolution of SARS-CoV-2 [18][19][20][21] during the first epidemic wave remain. Additionally, recent modeling efforts aimed at shedding light on those questions have mostly focused on China [1][2][3][4], the rest of Asia [22][23][24], Western and Central Europe [25][23][26], and North America [27][28][29][30][5], largely neglecting Africa, the Middle East, and Eastern Europe.

There is a scarcity of modeling studies aimed at elucidating the transmission dynamics and control in Eastern Europe, where the post-socialist economies and the corresponding healthcare systems are inherently different from Western Europe. The available transmission models for Eastern Europe are based on relatively simple SIR or similar compartmental models [31][32][33] where individuals are assigned to groups and all individuals within a given group are expected to have the same characteristics. However, no agent-based modeling studies have been conducted to evaluate the impact of spatial heterogeneities in key transmission drivers such as population density [32][33] [34][35][36][37][38][39]. Furthermore, research studies on the COVID-19 pandemic in Eastern Europe have not incorporated genomic data analysis of circulating viral strains and genomic patterns. In this paper, we sought to fill the knowledge gap for the Ukrainian epidemic [40], which provides a unique setting for studying the COVID-19 epidemic in the ex-USSR healthcare system, but where quarantine policies similar to the rest of Europe have been implemented.

2. Methods

2.1. Agent-based Stochastic Model

The Ukrainian territory can also be geographically covered with the local incidence stochastic modeling framework that has been presented earlier and illustrated in the context of the COVID-19 epidemic in the state of Massachusetts, USA [41], with some model adjustments and modifications for Ukraine. The adjustments are necessary since the country territory, the available incidence and mortality data quality and data reporting mechanisms do vary from country to country.

The local stochastic modeling framework is utilized to investigate the COVID-19 epidemic in Ukraine and assess the course of the epidemic under

different quarantine scenarios. The framework is represented by an agent-based model in which every newly infected individual is represented as a programming language object with multiple characteristics of that individual [41]. This analytic framework and the corresponding model have been adopted for Ukraine, since the available surveillance data and country geography are different. The summary of the framework and the model structure together with the adaptation details are outlined below:

The progression of the Ukrainian epidemic is modeled in discrete time slots that correspond to days and are indexed by variable t. For each modeled individual k the corresponding personal characteristics include multiple parameters such as Cartesian coordinates denoted as $(x_k; y_k)$, the time of infection onset $tinf_k$ in relation to some baseline (which is April 12th, 2020 in the model) and the detection time det_k which corresponds to the time between the infection onset and the diagnosis (i.e. the identification of infected individual k as such). The personal characteristics also include disease stage vector variable $stg_k = (stg_{1(k)}; stg_{2(k)}; stg_{3(k)})$ that contains the duration of disease stages such that $stg_{1(k)} + stg_{2(k)} + stg_{3(k)} = det_k$. The current infection status of the individual is stored in $st_k(t)$ variable and is different from day to day. This state variable can take values: $st_k(t) = 0$ i.e. the individual k is part of the model input data; $st_k(t) = 1$ i.e. the individual has been infected but has not been detected as such yet; $st_k(t) = 2$ i.e. the individual is infected and has been detected; $st_k(t) = 3$ i.e. the individual has recovered from the virus and is immune, and $st_k(t) = 4$ corresponds to the deceased status. In addition to that the individual k has a severity variable $sever_k$ (i.e. 1 - lethal, 2 - severe, and 3 - mild), a duration variable dur_k that characterizes the duration of the disease and is generated based on the $sever_k$ parameter. The individual reproduction number $R_{0(k)}$ stores the average number of infected individuals by the individual k across multiple runs. The summaries of those individual reproduction numbers $R_{0(k)}$ across individuals are used for the estimation of the population basic reproduction number $\mathcal{R}_0$ which serves as a characteristic of the entire epidemic [42]. The corresponding individual-specific transmission parameters are $cont_k = (\mu_{cont(k)}, \sigma^2_{cont(k)})$ which are used to generate the number of induced infection for each day t. The other characteristics of the individual k include age variable age_k as well as, the infectivity parameters such as radius rad_k up to which the individual can infect the other individuals and probability $p_{cont(k)}$ of infecting other individuals during each day. In summary, the entire vector of characteristics for individual k has the form:

$$\Big((x_k, y_k), tinf_k, det_k, stg_k, age_k, rad_k, p_{cont(k)}, cont_k, R_{0(k)}, sever_k, dur_k, st_k(t)\Big). \quad (1)$$

The model initial cases are generated based on the first subset of reported cases up to a certain reported date. Within the model multiple circular local geographic epicenters $\boldsymbol{E} = \{E_1, E_2, \ldots, E_{\mathcal{I}}\}$ are characterized by the center coordinates and the corresponding radius R_i for $i = 1, 2, \ldots, \mathcal{I}$ where each E_i represents a local epicenter and $\mathcal{I}$ is the total number of epicenters. The epicenters are used to model geographic areas of the infection introduction into the population and the center coordinates that correspond to the centroid regions that are defined from the available epidemiological data. For Ukraine they are defined from administrative and surveillance data that are available for cities and districts. In addition to the local disease introduction epicenters, the model incorporates areas of high population density $\boldsymbol{P} = \{P_1, P_2, \ldots, P_{\mathcal{J}}\}$. Those centers are characterized by the center coordinates and the corresponding radius R_j for $j = 1, 2, \ldots, \mathcal{J}$ where each P_j represents a large city or another densely populated area and $\mathcal{J}$ is the total number of such areas.

The set of reported cases and the corresponding time slots is divided into two subsets based on the reporting dates. The first set of indexes and corresponding cases are used to define the model input and related conditions. The second set of indexes and corresponding cases are used for both the model calibration (i.e. finding the set of parameters when the difference between the model-produced outputs and the reported data is the smallest possible) and for the model validation (i.e. comparing the remaining reported cases that are not used for calibration with the aligned with model-produced outputs for the model justification).

The Ukrainian surveillance data were obtained from the National Security and Defence Council of Ukraine web site [43]. The reports by region were available daily since the beginning of the epidemic in March 2020. The smallest geographic resolution of the data has been an administrative district called "oblast" in Ukrainian. Data for those geographic regions that are not under the control of the Ukrainian government have not been reported by the National Security and Defence Council of Ukraine and have not been used in the subsequent modeling [43]. The first confirmed case in Ukraine was reported on March 3, 2020 and was a Ukrainian who has recently returned from Italy, while the first death was reported on March 13, 2020 [44][40]. The Ukrainian government has started to implement quarantine measure on

March 12, 2020 [40]. Those measures were slightly softened on April 24, 2020 and many services resumed even though the quarantine itself lasted till the end of June 2020 [40]. The initially reported cases from March 12, 2020 until April 12, 2020 represented the first set of cases and were retrospectively stochastically incorporated into the model as the model initial conditions [41]. For the second set of reported cases the model-produced outputs from April 22, 2020 until July 12, 2020 were used for model calibration, while the outputs from July 13, 2020 until August 1, 2020 were used for model validation and comparison with the reported data. The optimization of model parameters has been performed by minimizing the sum of squared differences between the model-produced outputs (across multiple runs) and the reported calibration data by using the Nelder–Mead numerical minimization method [45]. Due to the implemented quarantine by the Ukrainian government the model fit was performed with the assumption that reduced transmission probability parameters were used during the model fitting period. The August 1, 2020 modeling end date has been selected to be in agreement with the available genomic data acquisition times [46] and to be conclusive for the first epidemic wave progression in the immunity-naive population. The population basic reproduction number $\mathcal{R}_0$ [42] which is a characteristic of the entire epidemic at early stages has been estimated from the model-produced distribution quantiles (5%, median, 95%) of the reproduction numbers of individuals within the model and summarized across multiple stochastic runs.

The additional details about the model formulation, parametrization, and calibration methods have been presented earlier [41] while the corresponding model code and the tool for the Ukrainian epidemic have been made public on GitHub [47].

2.2. Genomic Epidemiology Analysis

SARS-CoV-2 genomes together with the corresponding sampling times were extracted from GISAID [46] database on January 14, 2021. Among them in total 60 high-quality Ukrainian sequences suitable for genomic analysis have been identified and used for the analysis. The sampling range for available Ukrainian sequences was from April 24, 2020 to August 7, 2020.

The first analysis part was the identification of virus introductions into the country and within-country transmission clusters. For this part a maximum likelihood phylogeny was constructed using Nextstrain build for SARS-CoV-2 [48] with the default country-specific subsampling settings [49]. The Nextstrain algorithm selected a representative subsample of 6479 GISAID

sequences which included Ukrainian sequences, the sequences with close genomes, as well as a representative subsample of sequences from other geographic regions. Besides the time-labeled phylogeny Nextstrain was used to infer global genomic lineages and regional ancestral states. In the phylogenetic tree constructed by Nextstrain, the Ukrainian transmission clusters were identified as clades with the most recent common ancestors (MRCA) estimated as originating from Ukraine.

The second analysis part was the phylodynamic analysis which was performed using BEAST v1.10.4 [50] for 60 Ukrainian sequences. The strict molecular clock, HKY+Γ nucleotide substitution model and a tree prior with exponential growth coalescent were used for BEAST model settings. Parameters as evolutionary rate (clock.rate), population size, growth rate were adjusted in BEAUti v 1.10.4 with following priors as CTMC Rate Reference, LogNormal[1.0, 2.0], Laplace[0, 100], respectively. Other parameters were used with default prior settings. The model parameters were estimated by using 10,000,000 iterations (steps) of Markov Chain Monte Carlo (MCMC) methods with sampling at each 1000 steps, deleting the initial 10 % as burn-in. MCMC behaviour was satisfactory with reasonable effective sample sizes above 200 calculated by Tracer v1.10.4. [51].

For the three largest clusters identified by Nextstrain, separate phylogenies were constructed, and an exponential growth rates f of the infected population size had been estimated for each phylogeny using BEAST with the same parameter settings that were used for all 60 sequences described earlier. These three separate estimates $\hat{f}$ were used for the estimation of the basic reproduction numbers $\mathcal{R}_0$, that were alternatives to the one from the transmission model described earlier. The estimates of the basic reproduction numbers $\mathcal{R}_0$ from genomic data were produced with the assumption that SARS-CoV-2 generation intervals (i.e. times between infection onset and onward transmission) were gamma-distributed [52]. The estimates of $\mathcal{R}_0$ from genomic data were derived using the formula [53][54][55][56]:

$$\hat{\mathcal{R}}_0 = \left[1 + \frac{\hat{f}\hat{\sigma}^2}{\hat{\mu}}\right]^{\frac{\hat{\mu}^2}{\hat{\sigma}^2}}, \tag{2}$$

where $\hat{\mu}$ and $\hat{\sigma}$ were the estimates of mean and standard deviation of the aforementioned gamma distribution of SARS-CoV-2 viral population. The values $\hat{\mu} = 5.20$ and $\hat{\sigma} = 1.72$ were taken as estimates from the literature [57]. The other values from the same manuscript [57] but reported for different

region ($\hat{\mu} = 3.95$ and $\hat{\sigma} = 1.51$) were also considered but the estimates of $\mathcal{R}_0$ did not change by much. The results are summarized for those pairs of estimates.

The corresponding 95% uncertainty intervals for $\mathcal{R}_0$ were produced separately for each cluster. Since the formula (2) defines a strictly monotone transformation of f, the application of this transformation to both sides of 95% highest posterior density (HPD) intervals produced by BEAST for f results in desired 95% uncertainty bounds for $\mathcal{R}_0$.

3. Results

3.1. Agent-based Stochastic Model

In addition to the model fit to the reported data two alternative simulation scenarios were considered with later quarantine implementation dates. The first scenario corresponds to output of the model that has been calibrated based on the reported data. The other two scenarios correspond to hypothetical simulation scenarios which could have happened provided the quarantine measures were implemented one (on April 19, 2020) and two (on April 26, 2020) weeks later after the model initialization time (i.e. April 12, 2020). The visual results of first scenario model fit and the corresponding outputs are summarized in Figure 1. Blue curves in Figure 1. correspond to the reported data. They are captured by the model fits which is also indicated by the corresponding median and 90% pointwise model prediction bands across five hundred runs. The calibration interval is highlighted by green background. The model-produced values outside the calibration interval (i.e. after July 12, 2020) are still in a good agreement with the reported data which indicate the predicting abilities of the model [41].

For each of the three considered scenarios the median value across five hundred simulations were computed at each time point and presented together with the corresponding 5-th and 95-th percentiles across five hundred stochastic realizations to form the 90% prediction intervals. The corresponding results are summarized in Tables 1. and 2. for the model-predicted cases and deaths, respectively. The three scenario summaries from Table 1. can be directly compared. For the first (fitted) scenario on June 1, 2020 the model predicts $25,619$ cumulative cases (with the 90% PI $(20,644; 30,694)$), for the second scenario the model predicts $30,321$ cumulative cases (with the 90% PI $(24,345; 35,932)$), and for the third scenario the model predicts

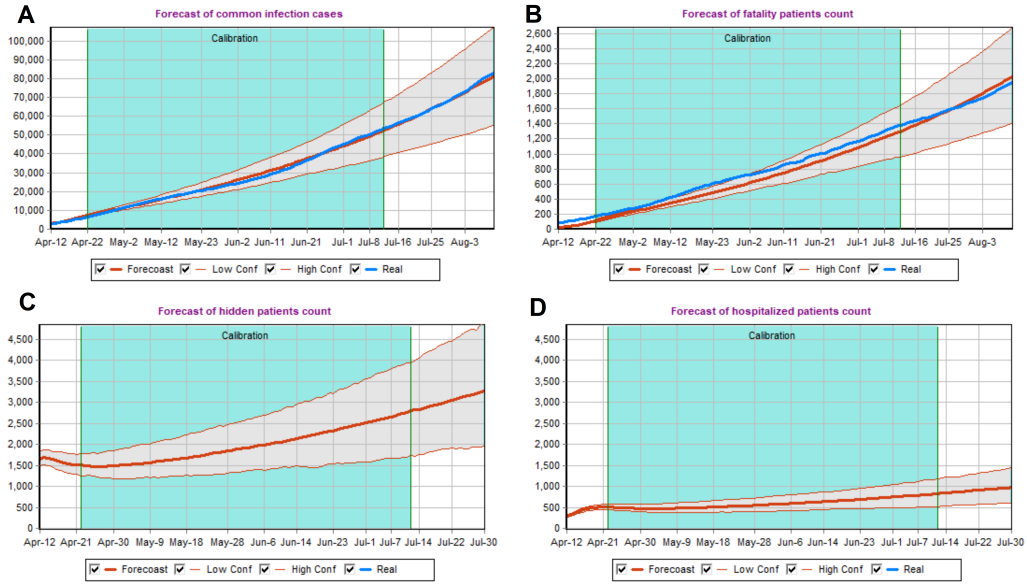

Figure 1: The model predictions together with the reported cases (Panel A) and reported death (Panel B) are presented. The hidden cases (Panel C) and hospitalized cases (Panel D) produced by the model are also presented. The model calibration time interval is highlighted in green. Red lines correspond to the median of the five hundred model-produced runs together with the corresponding 90% prediction bands to quantify the uncertainty. The reported cumulative data for cases and deaths are displayed in blue for visual comparison.

35,396 cumulative cases (with the 90% PI (28,999; 42,343)). For comparison, the actual number of reported cases by June 1, 2020 was 24,012 [43]. On August 1, 2020 for the first (fitted) scenario the model predicts 70,749 cumulative cases (with the corresponding 90% PI (49,131; 92,805)), for the second scenario the model predicts 85,929 cumulative cases (with the corresponding 90% PI (59,992; 109,638)), and for the third scenario the model predicts 104,795 cumulative cases (with the corresponding 90% PI (78,394; 132,849)). For comparison the actual number of reported cases by August 1, 2020 was 71,056 [43] which validates the model fit since August 1, 2020 was outside of the calibration interval. The second and third scenarios produce larger number of cases in comparison to the first fitted scenario. The median estimates can be compared directly. For example the second scenario results in 18% predicted increase in cumulative number of cases on June 1, 2020 and in 21% predicted increase in cumulative number of cases on August 1, 2020 in

comparison to the first scenario. The third scenario results in 38% predicted increase in cumulative number of cases on June 1, 2020 and in 48% predicted increase in cumulative number of cases on August 1, 2020 when compared to the first scenario.

The same summaries can be produced for the median number of deaths in Table 2. For the first (fitted) scenario on June 1, 2020 the model predicts 602 cumulative death (with the 90% PI $(501; 716)$) while the actual number of reported cases by June 1, 2020 was 718 [43]. On August 1, 2020 for the first (fitted) scenario the model predicts $1,757$ cumulative death (with the corresponding 90% PI $(1,243; 2,285)$) and the actual number of reported cases by August 1, 2020 was $1,709$ [43]. This also validates the model predicting abilities since August 1, 2020 was outside of the calibration interval. The second (simulated) scenario results in 18% increase in cumulative number of deaths predicted on June 1, 2020 and in 22% increase in cumulative number of deaths predicted on August 1, 2020. The third (simulated) scenario results in 36% increase in cumulative number of deaths predicted on June 1, 2020 and in 49% increase in cumulative number of deaths predicted on August 1, 2020.

Date	**June 1, 2020**	**July 1, 2020**	**August 1, 2020**
April 12, 2020	$25,619$ $(20,644; 30,694)$	$44,209$ $(33,260; 55,592)$	$70,749$ $(49,131; 92,805)$
April 19, 2020	$30,321$ $(24,345; 35,932)$	$53,599$ $(39,549; 66,519)$	$85,929$ $(59,992; 109,638)$
April 26, 2020	$35,396$ $(28,999; 42,343)$	$64,590$ $(50,293; 79,412)$	$104,795$ $(78,394; 132,849)$

Table 1: The predicted number of cumulative cases produced by the model over time for three different quarantine scenarios for three quarantine initiation dates together with the corresponding 90% prediction intervals.

Date	**June 1, 2020**	**July 1, 2020**	**August 1, 2020**
April 12, 2020	602 $(501; 716)$	$1,088$ $(831; 1,361)$	$1,757$ $(1,243; 2,285)$
April 19, 2020	712 $(581; 840)$	$1,313$ $(994; 1,625)$	$2,145$ $(1,529; 2,759)$
April 26, 2020	817 $(678; 969)$	$1,576$ $(1,225; 1,946)$	$2,619$ $(1,951; 3,286)$

Table 2: The predicted number of cumulative death produced by the model over time for three different quarantine scenarios for three dates together with the corresponding 90% prediction intervals.

For both the number of cases and for the number of deaths, the hypothetical scenarios two and three demonstrate the importance of the early

quarantine measures which cause the reduction in transmission probability parameters and, therefore, a reduction in the number of cases, and (more importantly) deaths. At the same time the results for the non-quarantine scenarios should be interpreted as sensitivity analysis since the Ukrainian government has implemented quarantine measures from the beginning of the epidemic and there were no data to estimate non-quarantine transmission probability parameters. This non-quarantine probability parameters have been adopted from the previous analysis [41] so the scenarios two and three correspond to hypothetical scenarios. The population basic reproduction number $\mathcal{R}_0$ estimates from the first scenario were 0.881 (median) and 0.870 (mean) with the corresponding 90% confidence interval from gentiles equal to $(0.21; 1.53)$.

3.2. Genomic epidemiology of SARS-CoV-2

For the research described in the first part of genomic methods the phylogenetic tree of 6479 global SARS-CoV-2 sequences (that include Ukrainian) has been constructed and presented in panels A and B of Figure 2.

The sequences were evaluated using a proposed classification [58]. The corresponding lineages of SARS-CoV-2 had the following distribution: $B1$ - 19 sequences, $B1.1.127$ - 14 sequence, $B1.1.29$ - 10 sequence, $B.1.1.243$ - 5 sequences, $B.1.1.247$ - 3 sequences, $B.1.1.208$ - 2 sequences, and single sequence for $B.1.1.167$, $B.1.1.194$, $B.1.1.220$, $B.1.1.238$, $B.1.1.257$, and $B.1.131$. Based on the proposed classification [58] analyzed lineages from Ukraine belong to lineage B.

At the same time according to the build-in Nextstrain classification [48] the analyzed SARS-CoV-2 variants from Ukraine belong to lineages $19A$, $20A$ and $20B$. The corresponding lineages of SARS-CoV-2 had the following distribution: $19A$ - 1 sequence $(1,67\%)$, 20A - 31 sequence $(51,67\%)$ and $20B$ - 28 sequences $(46,67\%)$. Those lineages are visualized on a global tree in panel B of Figure 2.

Eight intra-Ukrainian transmission clusters with at least two sequences in each of them have been identified on a global tree and summarized in Figures A1-A8 in Appendix. Cluster seven contained three sequences which have been phylogenetically close, but the MRCA from Ukraine was only confirmed on a Ukrainian tree but was not confirmed on a global tree. Therefore, based on available data, there is no way to say for sure whether it has indeed been a cluster.

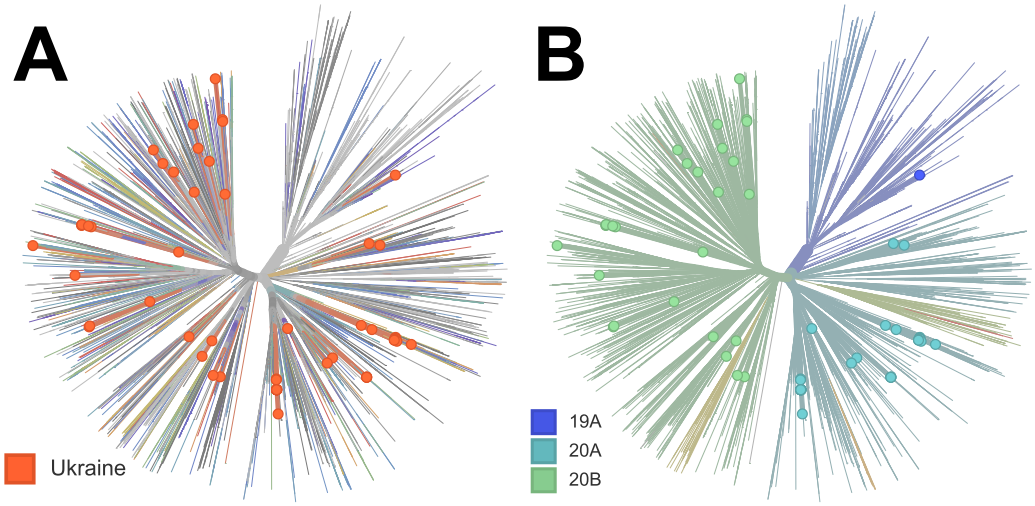

Figure 2: The global phylogenetic tree of SARS-CoV-2 genomes: A) distribution of Ukrainian SARS-CoV-2 genomes inside global SARS-CoV-2 population, B) the same tree with global SARS-CoV-2 lineages highlighted.

The largest three clusters are presented in Figure 3. The presence of such distinct clusters suggests multiple introductions of SARS-CoV-2 into Ukraine. This agrees well with the virus introduction patterns observed in other countries [59], and emphasizes the contribution of global movement of people to the rapid spread of SARS-CoV-2.

For the research described in the second part of genomic methods the phylodynamic analysis of Ukrainian sequences has been performed. This analysis allowed the estimation of timelines of the eight largest cluster introductions to the country and the comparison of those timelines with the timeline of the viral control measures applied in Ukraine. The inferred phylogeny is depicted in Figure 4 with MRCAs of the eight clusters highlighted as green circles. The vertical red bar time separator in Figure 4 corresponds to the formal quarantine date introduction in Ukraine i.e. to March 12, 2020 [40] [60]. The formal quarantine date followed by the closure of the country's land borders and a ban on foreign citizens' entry on March 16, 2020 [61] and the closure of international air, rail and bus services on March 17, 2020 [62].

Times to the most recent common ancestor (TMRCA) for the identified eight clusters are in the range from January 28, 2020 to June 4, 2020 which is visually summarized in Figure 5. TMRCAs of four clusters (50%) fall before the announced quarantine date, and the sequences from these clusters

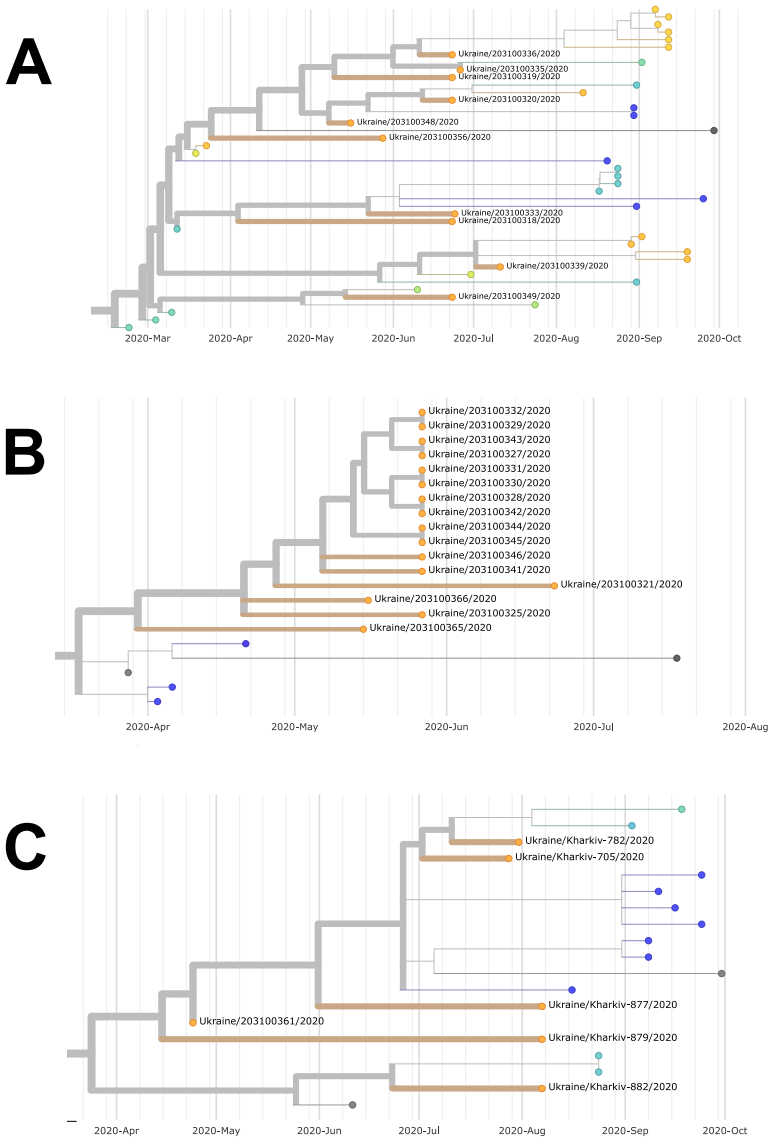

Figure 3: The three largest Ukrainian clusters of SARS-CoV-2 sequences that have been identified in the global tree of 6479 sequences. The first cluster is presented in panel A, the second in panel B, and the third in panel C.

form 80% (36 sequences) of the total number of clustered sampled Ukrainian sequences. This is visualized in the bar chart in Figure 5. Clusters with

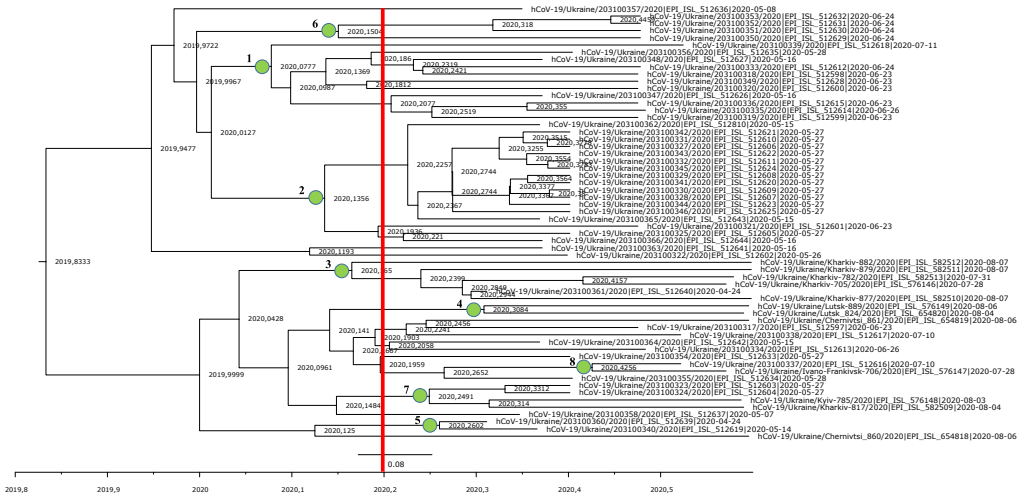

Figure 4: The SARS-CoV-2 clusters are presented in the Ukrainian phylogenetic tree. Clusters marked by light green circles and numbered from one to eight. The quarantine start date in Ukraine has been March 12, 2020, which is indicated by a vertical red bar time separator.

TMRCA after the quarantine date with numbers 4,5,7, and 8 constitute 20% (9 sequences) of observed clustered Ukrainian coronavirus population which is visualized in a pie chart in Figure 5. The remaining 15 sequences (not shown) *did no cluster* with any other Ukrainian sequences on the global tree (not shown due to tree complexity). There is not enough information in the available samples to make strong statements about the roles of those sequences in transmission. The analysis findings support the hypothesis that the public health measures taken to control the virus spread in Ukraine did not have a decreasing effect on the viral population numbers. The conclusion came from the fact that four MRCAs were estimated to form before the actual quarantine date and later gave a rise to 60% of analyzed Ukrainian SARS-CoV-2 sequences.

The estimates of the basic reproduction number $\mathcal{R}_0$ are summarized in Table 3. The estimates $\hat{\mathcal{R}}_0$ are presented together with the corresponding 95% uncertainty intervals for two combinations of parameters of the gamma distribution for generation interval [57]. The relevant estimates of the exponential growth rate f are presented in Appendix Table A1.

The estimates $\hat{\mathcal{R}}_0$ from Table 3 are close to each other, and most of the

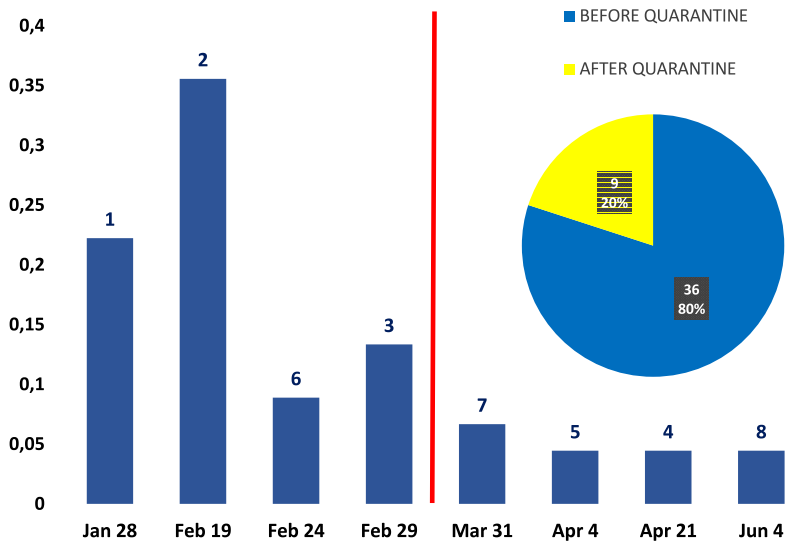

Figure 5: The dates of origin for the most recent common ancestors (MRCAs) and the corresponding cluster proportions for clustered SARS-CoV-2 sequences from Ukraine. The quarantine start date in Ukraine has been March 12, 2020, which is indicated by a vertical red bar time separator. The pie chart displays the counts and the proportions for sequences that 1) their MRCAs arose before the quarantine date, 2) their MRCAs arose after the quarantine date, 3) have not been identified as members of any other Ukrainian cluster.

Cluster	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mathcal{R}}_0$ & 95% CI	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mathcal{R}}_0$ & 95% CI
First	5.20	1.72	1.66 (1.17; 2.29)	3.95	1.51	1.47 (1.13; 1.87)
Second	5.20	1.72	1.26 (0.84; 3.42)	3.95	1.51	1.19 (0.88; 2.54)
Third	5.20	1.72	1.32 (1.02; 1.72)	3.95	1.51	1.24 (1.01; 1.51)

Table 3: The estimates of the basic reproduction number $\mathcal{R}_0$ for the three largest clusters (one, two and threee) together with the corresponding 95% confidence intervals (CI-s) produced from genetic data analysis. The results are summarized for two pairs of gamma distribution parameter estimates ($\hat{\mu} = 5.20$ with $\hat{\sigma} = 1.72$ and $\hat{\mu} = 3.95$ with $\hat{\sigma} = 1.51$) for the exponential growth rate.

estimates are larger than one, and the corresponding confidence intervals cover values above one. Those findings support the hypothesis of ongoing local transmission of SARS-CoV-2 during the first wave in Ukraine.

4. Discussion

In this study, the authors have detailed the epidemic characteristics of the first wave of the COVID-19 pandemic in Ukraine, evaluated effects of timely interventions, and characterized the genomic diversity of circulating viral strains. In contrast to previous coronavirus epidemics, the decreased case fatality rate and presence of asymptomatic transmission has resulted in unprecedented spread of SARS-CoV-2 and a renewed public interest in infectious disease epidemiology [63][64]. It has also provided mathematical modelers an opportunity to explore the effects of varying levels of human connectivity, population density, and different types and intensity of non-pharmacologic interventions [64][65]. With such heterogeneity, we argue that is crucial to rely on stochastic modeling frameworks, such as the one presented in this paper, to evaluate the impact of critical public health decisions in mitigating the effects of this pandemic [66].

To the authors knowledge this is the first paper to characterize the population structure of SARS-CoV-2 viruses circulating in Ukraine. Aside from human factors, the population structure was largely dependent on the number of viral introductions and founder genetics along with differences in the transmissibility of novel variants [67] [68]. For Ukraine, the genomic data analysis results suggest that there have been at least eight introductions of SARS-CoV-2 into Ukraine during the first wave despite the quarantine; four of which occurred after the implementation of quarantine. Unlike quantifying the reduction in incidence, the effect of non-pharmacologic interventions on the genetic diversity is less straightforward [69]. Intuitively, that the earliest possible quarantine date has the potential to reduce the number of new introductions, however on a population level, they contribute only a small fraction to the overall incidence [70]. Just as planning non-pharmacologic interventions, the ability to infer transmission patterns via genomic sequencing and phylogenetic analyses will prove critical to ongoing development of vaccines as the virus continues to evolve [71][72].

The findings from this paper are of course not without limitations. As one of the poorest countries in Europe (based on GDP per capita) much of the health care infrastructure in Ukraine lacks the resources of its neighbors [73]. Both availability of screening tests and the reporting of incidence data during the initial epidemic likely underestimated the burden of disease. Likewise, as the pandemic evolved, the scarcity of genetic sequencing limited the number of sequence comparison in the phylogenetic analysis. As such, the number of

actual viral clusters remains unknown and should be interpreted as *at least eight* clusters. The findings however, do provide a clear indication of a local transmission within Ukraine and were useful to establish a timeline for viral introduction.

In summary, this study was among the first to explore the characteristics of the initial pandemic as it spread to Ukraine and provided additional genomic analyses not previously published. Not only does the use of the local stochastic model support the efficacy of non-pharmacologic measures such as quarantine to reduce the spread of illness, it demonstrated that even small delays in implementation can have large effects on the number of future cases.

5. Acknowledgements

The authors would like to acknowledge the authors, originating and submitting laboratories of the sequences from GISAID. The list of laboratories is provided in Appendix. The authors would also like to acknowledge Sergey Brutsky for his help with Ukraine incidence data collection and acquisition.

6. Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, et al., A novel coronavirus from patients with pneumonia in china, 2019, New England Journal of Medicine (2020).

[2] H. Lu, C. W. Stratton, Y.-W. Tang, Outbreak of pneumonia of unknown etiology in wuhan china: the mystery and the miracle, Journal of Medical Virology (2020).

[3] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, et al., Clinical features of patients infected with 2019 novel coronavirus in wuhan, china, The lancet 395 (2020) 497–506.

[4] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study, The Lancet 395 (2020) 507–513.

[5] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, S. Bernardini, The covid-19 pandemic, Critical Reviews in Clinical Laboratory Sciences 57 (2020) 365–388.

[6] WHO, `https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19---11-march-2020`, 2020. Accessed: 2021-02-20.

[7] WHO, `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it`, 2020. Accessed: 2021-02-20.

[8] E. Mahase, Coronavirus: covid-19 has killed more people than sars and mers combined, despite lower case fatality rate, BMJ (2020).

[9] Johns Hopkins University - Coronavirus Resource Center, `https://coronavirus.jhu.edu/map.html`, 2021. Accessed: 2021-03-01.

[10] W.-H. Chen, U. Strych, P. J. Hotez, M. E. Bottazzi, The SARS-CoV-2 vaccine pipeline: an overview, Current tropical medicine reports (2020) 1–4.

[11] S. Caddy, Russian SARS-CoV-2 vaccine, BMJ: British Medical Journal (Online) 370 (2020).

[12] A. Mullard, How COVID vaccines are being divvied up around the world, Nature (2020). URL: https://doi.org/10.1038/d41586-020-03370-6. doi:10.1038/d41586-020-03370-6.

[13] N. Islam, S. J. Sharp, G. Chowell, S. Shabnam, I. Kawachi, B. Lacey, J. M. Massaro, R. B. D'Agostino, M. White, Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries, bmj 370 (2020).

[14] WHO, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public, 2020. Accessed: 2021-02-20.

[15] CDC, https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html, 2020. Accessed: 2021-02-20.

[16] F. Di Gennaro, D. Pizzol, C. Marotta, M. Antunes, V. Racalbuto, N. Veronese, L. Smith, Coronavirus diseases (covid-19) current status and future perspectives: a narrative review, International journal of environmental research and public health 17 (2020) 2690.

[17] K. Iyengar, A. Mabrouk, V. K. Jain, A. Venkatesan, R. Vaishya, Learning opportunities from covid-19 and future effects on health care system, Diabetes & Metabolic Syndrome: Clinical Research & Reviews 14 (2020) 943–946.

[18] K. Tizaoui, I. Zidi, K. H. Lee, R. Abou Ghayda, S. H. Hong, H. Li, L. Smith, A. Koyanagi, L. Jacob, A. Kronbichler, et al., Update of the current knowledge on genetics, evolution, immunopathogenesis, and transmission for coronavirus disease 19 (covid-19), International Journal of Biological Sciences 16 (2020) 2906.

[19] H. Yi, 2019 novel coronavirus is undergoing active recombination, Clinical Infectious Diseases (2020).

[20] C. Li, Y. Yang, L. Ren, Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species, Infection, Genetics and Evolution (2020) 104285.

[21] D. Benvenuto, M. Giovanetti, A. Ciccozzi, S. Spoto, S. Angeletti, M. Ciccozzi, The 2019-new coronavirus epidemic: evidence for virus evolution, Journal of medical virology 92 (2020) 455–459.

[22] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, V. J. Lee, Presymptomatic transmission of SARS-CoV-2—singapore, january 23–march 16, 2020, Morbidity and Mortality Weekly Report 69 (2020) 411.

[23] N. Yamamoto, G. Bauer, Apparent difference in fatalities between central europe and east asia due to SARS-CoV-2 and covid-19: Four hypotheses for possible explanation, Medical Hypotheses (2020) 110160.

[24] D. Roy, S. Tripathy, S. K. Kar, N. Sharma, S. K. Verma, V. Kaushal, Study of knowledge, attitude, anxiety & perceived mental healthcare need in indian population during covid-19 pandemic, Asian Journal of Psychiatry (2020) 102083.

[25] H. Salje, C. T. Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, et al., Estimating the burden of SARS-CoV-2 in france, Science (2020).

[26] A. Fokas, N. Dikaios, G. Kastis, Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2, Journal of the Royal Society Interface 17 (2020) 20200494.

[27] I. Ghinai, T. D. McPherson, J. C. Hunter, H. L. Kirking, D. Christiansen, K. Joshi, R. Rubin, S. Morales-Estrada, S. R. Black, M. Pacilli, et al., First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the usa, The Lancet (2020).

[28] S. Ellington, P. Strid, V. T. Tong, K. Woodworth, R. R. Galang, L. D. Zambrano, J. Nahabedian, K. Anderson, S. M. Gilboa, Characteristics of women of reproductive age with laboratory-confirmed SARS-CoV-2 infection by pregnancy status—united states, january 22–june 7, 2020, Morbidity and Mortality Weekly Report 69 (2020) 769.

[29] J. D. Silverman, N. Hupert, A. D. Washburne, Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the united states, Science translational medicine 12 (2020).

[30] J. R. Fauver, M. E. Petrone, E. B. Hodcroft, K. Shioda, H. Y. Ehrlich, A. G. Watts, C. B. Vogels, A. F. Brito, T. Alpert, A. Muyombwe, et al., Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the united states, Cell (2020).

[31] A. Issanov, Y. Amanbek, A. Abbay, S. Adambekov, M. Aljofan, A. Kashkynbayev, A. Gaipov, Covid-19 outbreak in post-soviet states: modeling the best and worst possible scenarios (2020).

[32] I. Nesteruk, Long-term predictions for covid-19 pandemic dynamics in ukraine, austria and italy, Europe PMC (2020).

[33] Y. N. Kyrychko, K. B. Blyuss, I. Brovchenko, Mathematical modelling of the dynamics and containment of covid-19 in ukraine, Scientific reports 10 (2020) 1–11.

[34] S. Venkatramanan, B. Lewis, J. Chen, D. Higdon, A. Vullikanti, M. Marathe, Using data-driven agent-based models for forecasting emerging infectious diseases, Epidemics 22 (2018) 43–49.

[35] D. L. Chao, M. E. Halloran, V. J. Obenchain, I. M. Longini Jr, Flute, a publicly available stochastic influenza epidemic simulation model, PLoS computational biology 6 (2010).

[36] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, N. Wang, Modelling disease outbreaks in realistic urban social networks, Nature 429 (2004) 180–184.

[37] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun, et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak, Science 368 (2020) 395–400.

[38] D. L. Chao, A. P. Oron, D. Srikrishna, M. Famulare, Modeling layered non-pharmaceutical interventions against SARS-CoV-2 in the united states with corvid, medRxiv (2020).

[39] N. Hoertel, M. Blachier, C. Blanco, M. Olfson, M. Massetti, M. S. Rico, F. Limosin, H. Leleu, A stochastic agent-based model of the SARS-CoV-2 epidemic in france, Nature Medicine (2020) 1–5.

[40] A. Åslund, Responses to the covid-19 crisis in russia, ukraine, and belarus, Eurasian Geography and Economics (2020) 1–14.

[41] A. Kirpich, V. Koniukhovskii, V. Shvartc, P. Skums, T. A. Weppelmann, E. Imyanitov, S. Semyonov, K. Barsukov, Y. Gankin, Development of an interactive, agent-based local stochastic model of COVID-19 transmission and evaluation of mitigation strategies illustrated for the state of Massachusetts, USA, PLoS ONE 16 (2021) e0247182.

[42] G. N. Milligan, A. D. Barrett, An Essential Guide, Wiley Online Library, 2015.

[43] National Security And Defence Council Of Ukraine - , https://covid19.rnbo.gov.ua/, 2021. Accessed: 2021-02-20.

[44] Kyiv Post, https://www.kyivpost.com/ukraine-politics/developing-first-coronavirus-case-identified-in-ukraine.html, 2020. Accessed: 2021-02-20.

[45] J. A. Nelder, R. Mead, A simplex method for function minimization, The computer journal 7 (1965) 308–313.

[46] GISAID, A global initiative on sharing avian flu data., https://www.gisaid.org/, 2020. Accessed: 2021-02-20.

[47] github.com, https://github.com/quantori/COVID19-Ukraine-Transmission, 2021. Accessed: 2020-11-01.

[48] Nextstrain: Real-time tracking of pathogen evolution., https://www.nextstrain.org/, 2021. Accessed: 2021-02-20.

[49] github.com, https://github.com/nextstrain/ncov, 2021. Accessed: 2020-11-01.

[50] M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, A. Rambaut, Bayesian phylogenetic and phylodynamic data integration using beast 1.10, Virus evolution 4 (2018) vey016.

[51] virological.org - BEAST Settings, https://virological.org/t/phylodynamic-analysis-of-sars-cov-2-update-2020-03-06/420, 2021. Accessed: 2021-02-20.

[52] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al., Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia, New England Journal of Medicine (2020).

[53] BEAST Software, https://beast.community/estimating_R0.html, 2020. Accessed: 2021-02-20.

[54] BEAST Software, https://beast.community/phylodynamics_of_epidemic_influenza, 2020. Accessed: 2021-02-20.

[55] J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers, Proceedings of the Royal Society B: Biological Sciences 274 (2007) 599–604.

[56] N. C. Grassly, C. Fraser, Mathematical models of infectious disease transmission, Nature Reviews Microbiology 6 (2008) 477–487.

[57] T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, N. Hens, Estimating the generation interval for coronavirus disease (covid-19) based on symptom onset data, march 2020, Eurosurveillance 25 (2020) 2000257.

[58] A. Rambaut, E. C. Holmes, Á. O'Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, O. G. Pybus, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology, Nature microbiology 5 (2020) 1403–1407.

[59] J. L. Geoghegan, X. Ren, M. Storey, J. Hadfield, L. Jelley, S. Jefferies, J. Sherwood, S. Paine, S. Huang, J. Douglas, et al., Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in aotearoa new zealand, Nature communications 11 (2020) 1–7.

[60] www.kmu.gov.ua - Quarantine Announcement from March 12,2020, hhttps://www.kmu.gov.ua/news/uryad-prijnyav-nizku-rishen-shcho-mayut-ubezpechiti-ukrayinciv-vid-covid-19-11-03-20?fbclid=IwAR31alZLy55xNAlW82qQAhLH1fKbFxzPtG-YtR5XDvuNj-1CJIagF-_Q6Bs, 2021. Accessed: 2021-02-20.

[61] baltsi.mfa.gov.ua - No Entry for Foreigners - May 26, 2020 , https://baltsi.mfa.gov.ua/ru/news/timchasova-zaborona-do-

24-kvitnya-2020-r-na-peretinannya-derzhavnogo-kordonu-na-vyizd-v-ukrayinu-dlya-inozemciv-ta-osib-bez-gromadyanstva, 2021. Accessed: 2021-02-20.

[62] Closed borders, https://www.kmu.gov.ua/news/uryad-timchasovo-zaboroniv-vyizd-v-ukrayinu-inozemciv-ta-zakriv-mizhnarodne-pasazhirske-aviaspoluchennya, 2021. Accessed: 2021-02-20.

[63] C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, Factors that make an infectious disease outbreak controllable, Proceedings of the National Academy of Sciences 101 (2004) 6146–6151.

[64] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, T. D. Hollingsworth, How will country-based mitigation measures influence the course of the covid-19 epidemic?, The lancet 395 (2020) 931–934.

[65] R. Chaudhry, G. Dranitsaris, T. Mubashir, J. Bartoszko, S. Riazi, A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on covid-19 mortality and related health outcomes, EClinicalMedicine 25 (2020) 100464.

[66] C. M. Peak, L. M. Childs, Y. H. Grad, C. O. Buckee, Comparing non-pharmaceutical interventions for containing emerging epidemics, Proceedings of the National Academy of Sciences 114 (2017) 4023–4028.

[67] J. W. Tang, P. A. Tambyah, D. S. Hui, Emergence of a new SARS-CoV-2 variant in the uk, The Journal of Infection (2020).

[68] R. S. Baric, Emergence of a highly fit sars-cov-2 variant, New England Journal of Medicine (2020).

[69] B. Nussbaumer-Streit, V. Mayr, A. I. Dobrescu, A. Chapman, E. Persad, I. Klerings, G. Wagner, U. Siebert, D. Ledinger, C. Zachariah, et al., Quarantine alone or in combination with other public health measures to control covid-19: a rapid review, Cochrane Database of Systematic Reviews (2020).

[70] T. W. Russell, J. T. Wu, S. Clifford, W. J. Edmunds, A. J. Kucharski, M. Jit, et al., Effect of internationally imported cases on internal spread

of covid-19: a mathematical modelling study, The Lancet Public Health 6 (2021) e12–e20.

[71] T. C. Williams, W. A. Burgers, Sars-cov-2 evolution and vaccines: cause for concern?, The Lancet Respiratory Medicine (2021).

[72] A. Fontanet, B. Autran, B. Lina, M. P. Kieny, S. S. A. Karim, D. Sridhar, Sars-cov-2 variants and ending the covid-19 pandemic, The Lancet (2021).

[73] emerging-europe.com, https://emerging-europe.com/news/imf-ukraine-is-emerging-europes-poorest-country/, 2021. Accessed: 2021-03-01.