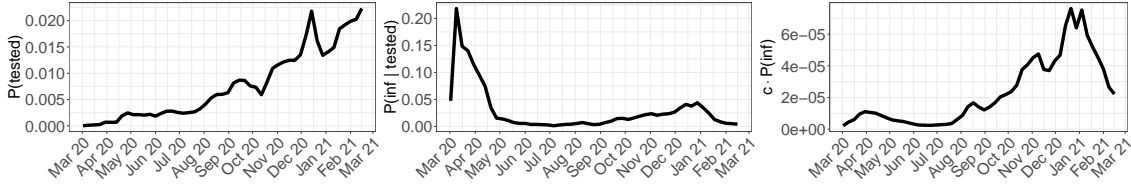
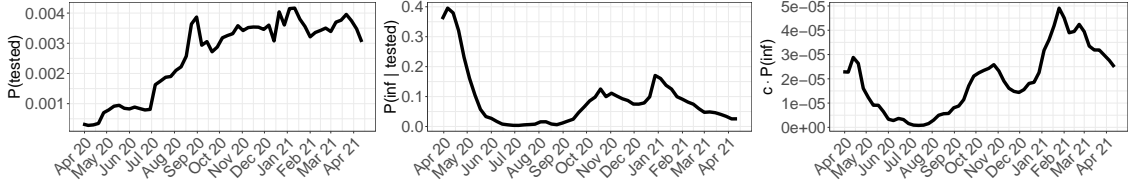


Supplementary Figure 1: Comparison of daily incidence estimates for Scotland using the incidence correlate derived from GInPipe (blue line, median estimates) and EpiInf (green line, median estimates) vs. the reported cases (red line, seven day rolling averages).

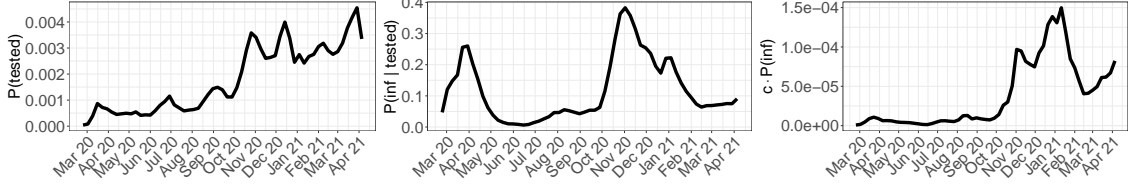
Denmark



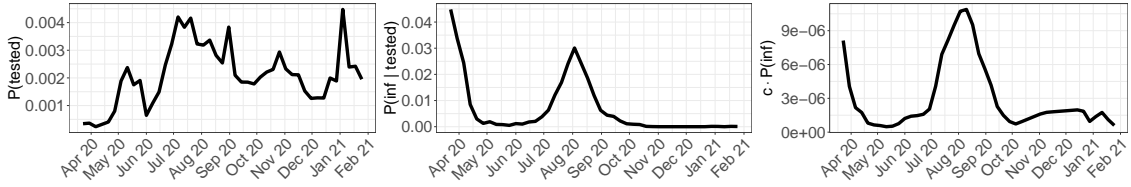
Scotland



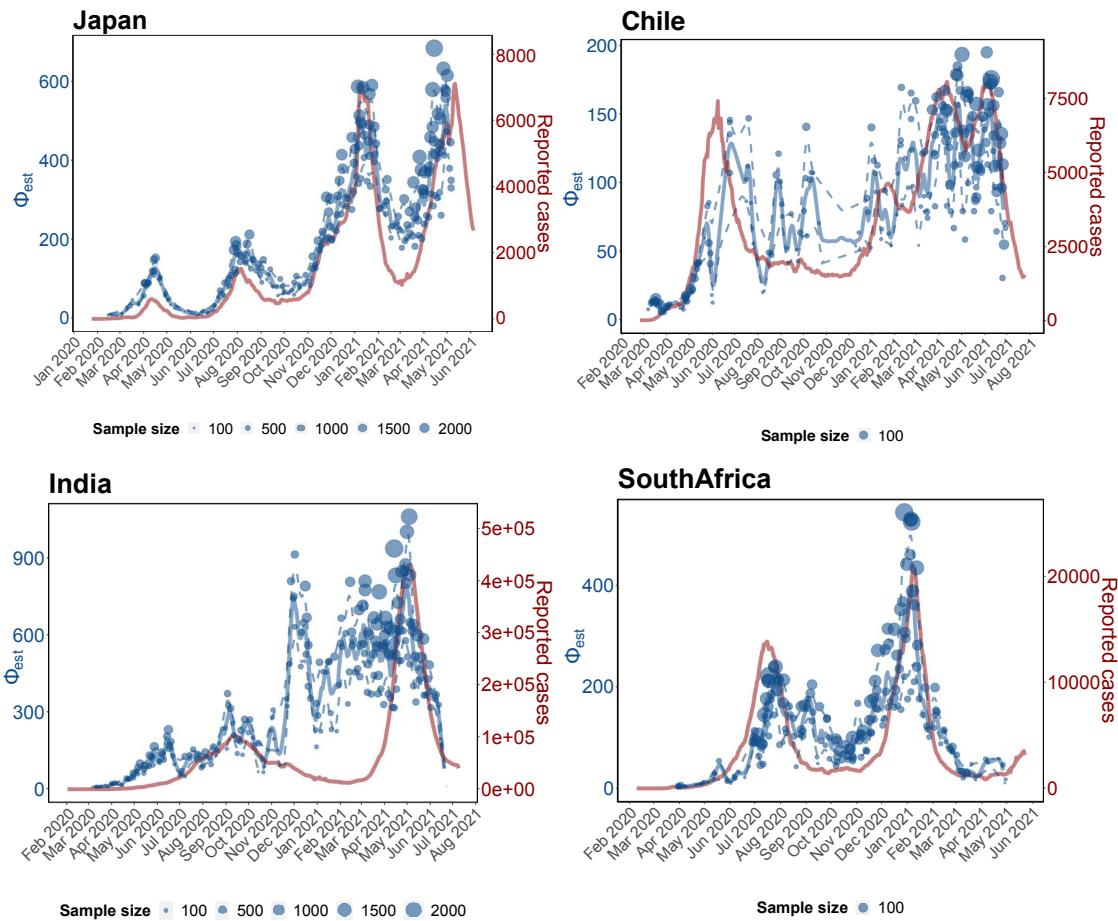
Switzerland



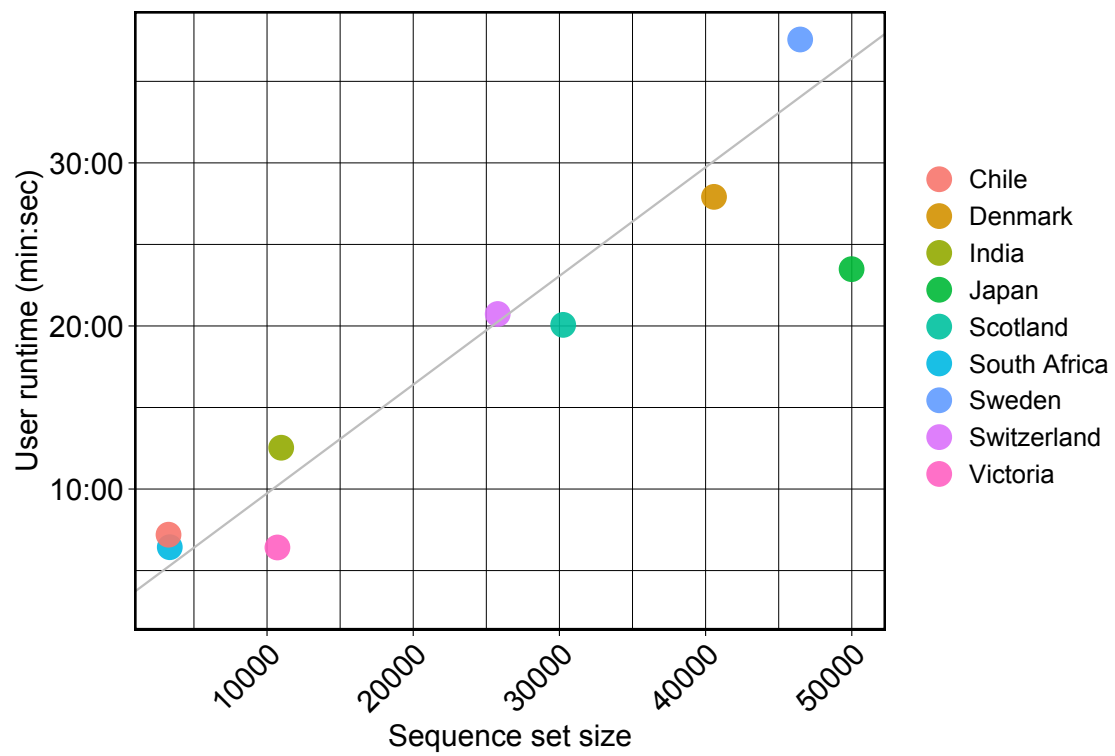
Victoria



Supplementary Figure 2: From left to right: Probability of being tested, conditional probability that a tested individual is infected, and probability of being infected over time. From top to bottom: Denmark, Scotland, Switzerland, and Victoria (Australia).



Supplementary Figure 3: Incidence reconstruction for additional countries. The graphic depicts the genome-based incidence reconstruction (in blue) using the proposed method (left axis) vs. the 7 days rolling average of newly reported cases in red (right axis). Blue dots depict ϕ_b point estimates of the incidence correlate, where the size of the dot is related to the number of sequences used to infer ϕ_b . The solid and dashed blue lines denote the median smoothed trajectories and their 5th and 95th percentiles. **Upper left.** Japan ($n = 49.986$ sequences) **Upper right.** Chile ($n = 3.261$ sequences) **Lower left.** India ($n = 19.979$ sequences) **Lower right.** South Africa ($n = 3.347$ sequences)



Supplementary Figure 4: Runtime of GInPipe scales linear with the number of sequences. The graphic depicts the runtime of executing the entire GInPipe pipeline on a standard laptop (2.3 Ghz, 2 cores) for different data sets.

Supplementary Note 1

The aim of Supplementary Note 1 is to evaluate, based on *in silico* simulated outbreaks, whether GInPipe can infer the number of infected individuals solely from viral genetic data. Moreover, we want to study the robustness of the method to missing data, unbalanced, time-dependent or biased sampling, introduction of unrelated sequence variants, its ability to reconstruct non-smooth dynamics, as well as its sensitivity to changes in the pathogen mutation rate and selective pressure.

SN.1 *In Silico* evaluation of GInPipe

As a proof of concept, we set up a simulation of a virtual outbreak: We formulate a stochastic epidemiological process by which a population of infectious agents progress with a time-dependent replication rate. This generates the number of infected individuals at each time step (‘ground truth’). In our simulations, each infected individual can contribute to the pandemic and is associated with one viral genome. These genomes can be transmitted and randomly mutated in every time step. The set of genomes constitutes the data source for GInPipe. More precisely, we use the *in silico* generated viral genomes to compute the incidence correlate $\phi(t)$ using GInPipe. The derived $\phi(t)$ estimates are then compared with the number of infected individuals $N(t)$ at each time step.

SN.1.1 Summary and Limitations

We thoroughly tested the performance of our method and probed its weaknesses based on the *in silico* simulations. We found that the method’s ϕ estimates linearly correlate with the infectious population size N_{true} . Different binning strategies, which are used to reconstruct the temporal incidence profiles yield complementary results. However, ϕ estimates can become inaccurate when the population of infectious individuals is small. ‘Outliers’ are however eliminated through application of a convolutional filter (smoothing). The ϕ estimates are unbiased in the case of missing data. However, the method becomes less accurate when too little viral sequences are available and when the “evolutionary signal” vanishes. Time-dependent changes in the sampling proportion have no effect on GInPipe. Evolutionary biased sampling (e.g. over representation of closely related sequences) decreases the evolutionary signal. This could, in principle, be problematic, if drastic changes, e.g. from a representative- to a biased sampling occur.

The importation of variants (introductions) that subsequently are community transmitted, do not affect the performance of the method. The same applies to introductions which are not community transmitted. In the extreme case of persistently large frequencies of non-transmitted introductions, however, the slope of the linear correlation function may change. Yet, a linear relationship between the incidence estimate ϕ and the infectious population size N_{true} is retained and the reconstruction of the pandemic trajectory is still possible.

GInPipe is able to reconstruct very steep increases and decreases in the population dynamic. Selective pressure has no effect on GInPipe’s incidence reconstructions. With regards to applying GInPipe to other respiratory viruses, low mutation rates decrease the “evolutionary signal”, but may be compensated by increasing the bin sizes, at the expense of temporal resolution.

SN.1.2 Outbreak simulation

For the evaluation of the method, we simulate the evolutionary dynamics of a viral outbreak. We simulate a time-discrete (generation-wise) population growth, where each simulation starts at time t_0 and ends at $t_{\text{final}} = 120$ (below, we also evaluate the corresponding continuous-time process). At t_0 , the outbreak starts with $N_{\text{init}} = 50$ identical sequences of length $L = 1000$. The nucleotides of the initial sequence are chosen randomly. In each of the t_{final} generations, the current population of size $N(t)$ replicates with replication rate ρ and mutates with mutation rate μ . To simulate non-monotonic outbreak trajectories, we chose a sinusoidal rate, depending on time t :

$$\rho(t) = \frac{\sin(t * 0.11)}{15} + 1.02 \quad (\text{SN.1})$$

The number of sequences in the new generation $t + 1$ is sampled from the Poisson distribution with

$$N(t + 1) \sim \mathcal{P}(N(t) \cdot \rho(t)) \quad (\text{SN.2})$$

The sequences for the next generation are chosen randomly with replacement.

The sampled sequences of the new generation are mutated with mutation rate $\mu = 0.0001$. Analogous to the previous step, the total number of positions to be mutated is sampled from a Poisson distribution:

$$n_{\text{mut}} \sim \mathcal{P}(\mu \cdot N(t + 1) \cdot L) \quad (\text{SN.3})$$

These mutations are randomly chosen across all sequences and positions without replacement (below, we also simulate selection pressure). Each nucleotide is mutated according to the Kimura substitution type model [1] with fixed rates for transitions ($A \leftrightarrow G$, $C \leftrightarrow T$ with $\alpha = 0.6$) and both types of transversions ($A \leftrightarrow C$, $G \leftrightarrow T$ with $\beta = 0.2$, as well as $A \leftrightarrow T$, $C \leftrightarrow G$ with $\gamma = 0.2$). During the course of simulation, each position could in principle be mutated several times.

This procedure results in a sequence set for each generation, which is written into a fasta file. The header of each sequence contains an ID and a date (arbitrary start date + t days) to conform with the format required for the computational pipeline (<https://github.com/KleistLab/GInPipe>). We ran 10 simulations with the settings stated above. The trajectory of the population dynamics are depicted in Figure SN.5.

Moreover, we sub-sampled the the generation-wise data set in different scenarios (section SN.1.7) and created fasta files for the respective scenarios.

SN.1.3 Execution of the method

For each of the generated (sub-)samples we ran the whole pipeline as described in the *Methods* section (main manuscript). ‘Temporal bins’ were spanning 2, 4, 6, 8, and 10 days (here days correspond to generations) and ‘equal size bins’ were created comprising 2, 5, and 7 percent of all sequences. We evaluated the normalised ϕ_b estimates for each bin b according to eq. (1) in the main manuscript.

SN.1.4 Incidence estimate ϕ_b correlates linearly with infectious population size N_{true}

As depicted in Figure SN.6 (left), the bin-wise ϕ_b estimates show a strong linear relationship to the ground truth N_{true} with high significance (Pearson correlation coefficient $r = 0.98$, $p < 10^{-16}$). More specifically, we observe a linear scaling c of the incidence estimate ϕ with N_{true} of the form

$$N_{\text{true}}(t) = c \cdot \phi(t) + \text{error}. \quad (\text{SN.4})$$

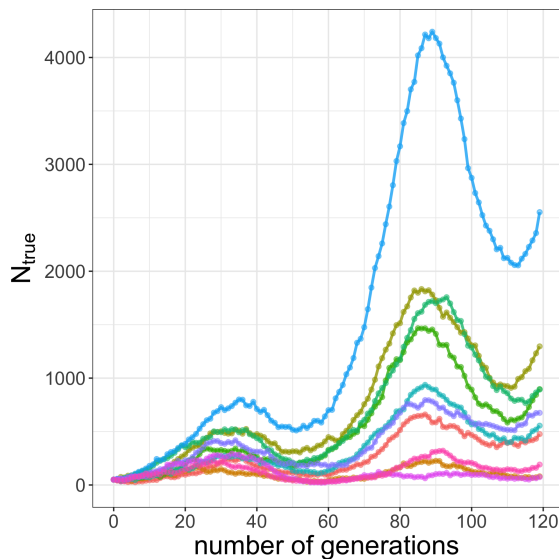


Figure SN.5: Trajectories for 10 simulations of population growth showing $N_{\text{true}}(t)$.

The deviation grows for larger values, because the considered evolutionary model is a Poisson process with error $= \sqrt{\phi(t)} \cdot \epsilon$, where ϵ is a zero-centered, symmetric, additive error.

In Figure SN.6 (right), we compare the percent deviation from the linear regression $\hat{\phi}(t)$:

$$\% \text{deviation} = \frac{\phi_b(t_b) - \hat{\phi}(t_b)}{\hat{\phi}(t_b)} \cdot 100, \quad (\text{SN.5})$$

with t_b denoting the *mean* date of all sequences in bin b . The deviation measure allows to center the approximation errors and henceforth analyse how distinct parameters may affect the method's performance. However, below, we also assess differences in the true vs. GInPipe reconstructed effective population sizes R_e .

We can see that the majority of the point estimates are quite accurately approximated with low percent deviation. Moreover, the estimation seems unbiased (symmetrically dispersed around 0% deviation) and large deviations are rare. Outliers are most prominent when we bin sequences by equal amounts of days, scattering in both directions, hence under- and overestimating. Creating bins of equal size, however, is slightly more prone to underestimating of $c \cdot N_{\text{true}}$.

SN.1.5 Effect of binning strategies

Next, we investigated the sources of outliers. In Figure SN.7, we show violin plots that colour-code the spanned days per bin (left) and the number of sequences per bin (right). Large deviations from the linear fit are particularly occurring for very small infectious population sizes N_{true} , which is indicated by the smaller markers in Fig. SN.7.

Moreover, overestimation for the 'equal size' binning is caused by bins spanning large temporal windows, around 20 days (see red dots in SN.7 (left), third violin plot). The outliers that occur for binning with equal amounts of days, to the contrary, are mainly caused by bins containing only very few sequences (red dots in the central violin plot in Figure SN.7 (left)). In these cases, the true population size is already small (small dots) and hence the amount of sequences used to infer the parameter are also low. Therefore, it may happen that the time until the next mutational event lies outside the particular bin, and not enough information is given to infer the (very low) population size.

We can conclude, that the choice of the binning strategy is important for accurate results. On the one hand, the bin has to be large enough in order to contain mutational information, but on

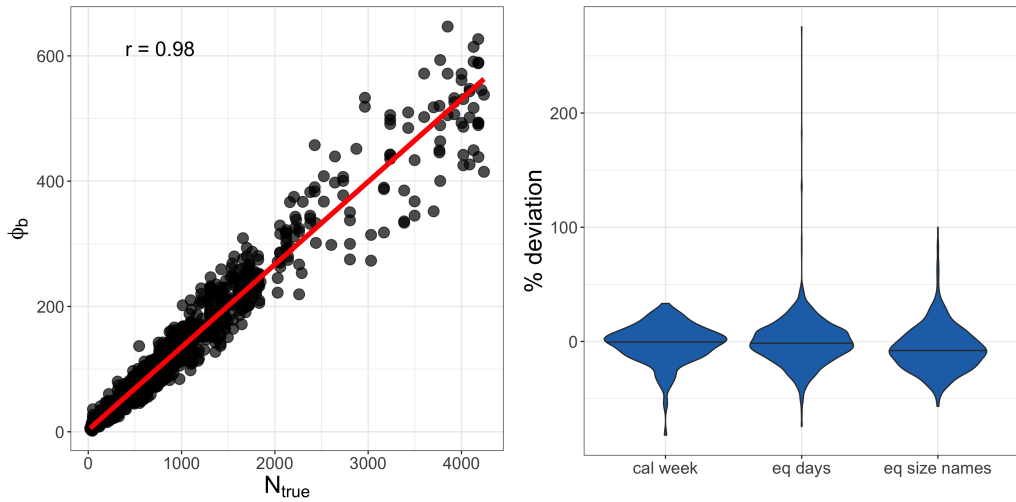


Figure SN.6: Linear relation between N_{true} and ϕ . **Left:** Dotplot showing N_{true} versus the estimated point estimates ϕ_b per bin (full data set). The red line shows the linear fit. **Right:** Violin plot of the percent deviation of ϕ_b from the linear fit for the different binning types (i) calendar week, (ii) equal amount of days and (iii) equal number of sequences per bin.

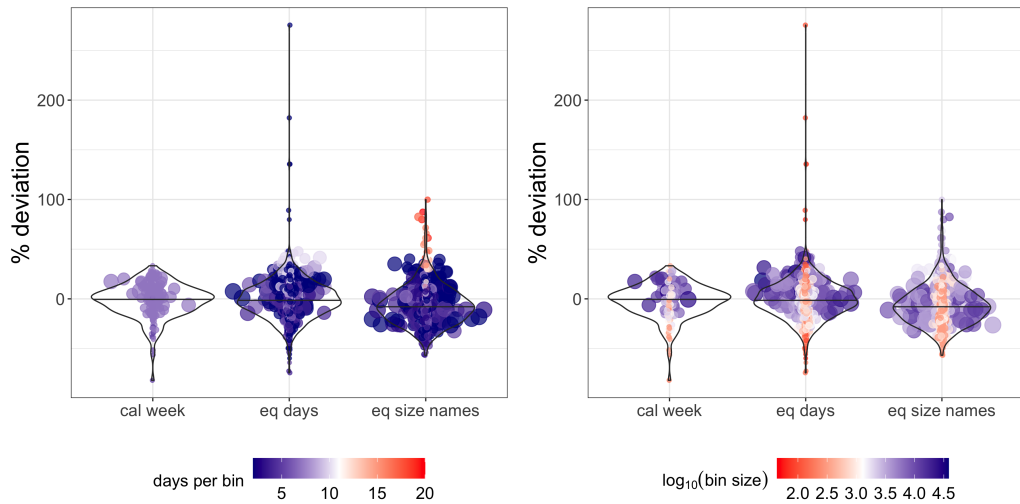


Figure SN.7: Deviation of ϕ from the linear correlation to N_{true} . Incidence correlate ϕ was calculated for the different binning types, i.e. by calendar week, equal days and equal bin size. The size and dispersion of data points indicate the value of N_{true} . **Left:** Colour and transparency indicate the days spanned per bin, with blue and transparent for low values, red and solid for large values. **Right:** Colour and transparency indicate the number of sequences per bin (in \log_{10}), with blue and transparent for large bin sizes, red and solid for small bin sizes.

the other hand not too large such that the population dynamics can still be temporarily resolved. Moreover, prediction accuracy declines with decreasing population size N_{true} .

SN.1.6 Smoothing corrects for outliers

The point estimates are smoothed using a convolution filter with a window size of 7 (days). To infer statistical uncertainty in ϕ profiles, we conducted a sub-sampling procedure of the ϕ point estimates (details in the *Methods* section, main article).

In Figure SN.8 (left), we show the smoothed reconstruction ϕ_{est} of the simulated dynamics of N_{true} , which we saw in Figure SN.5. Figure SN.8 (right) also shows, in a representative simulation, that the incidence correlates ϕ (blue dots and line) yield a proportional estimate of N_{true} (red line) and reflects the overall epidemiological dynamic quite accurately.

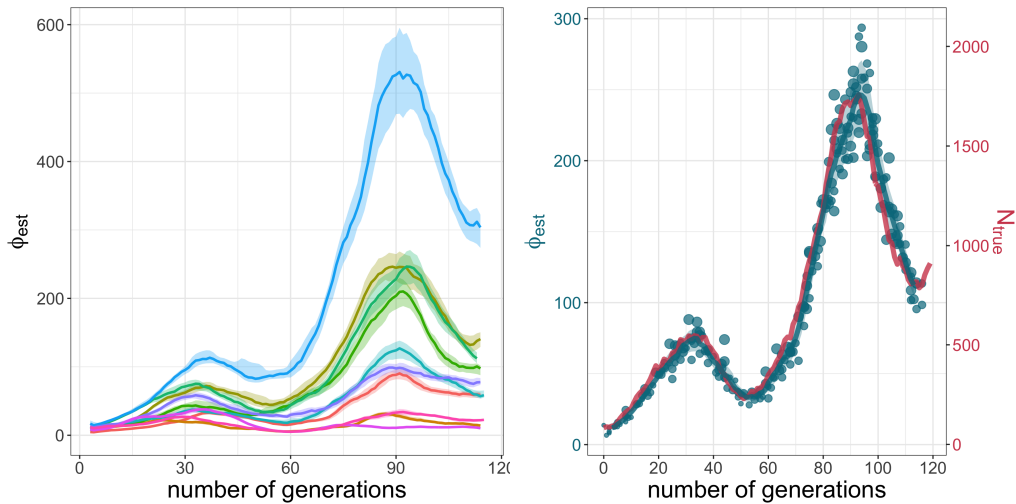


Figure SN.8: Reconstruction of outbreak dynamics. **Left:** The solid lines denote the smoothed median of the ϕ estimates for the 10 stochastic simulations of population growth. The coloured shades present the area between the 5th and 95th percentile derived from sub-sampling over ϕ_b point estimates. **Right:** The blue line shows the smoothed median of the sub-sampled ϕ_b estimates (points) for one of the simulations. The red line shows true incidence per generation N_{true} .

SN.1.7 Missing data does not affect incidence reconstructions

In the evaluation above, we have used all sequences that emerge in the evolutionary process. For a more realistic setting, we evaluated the estimation of ϕ by taking sub-samples of the total sequence set: We sub-sampled a certain fraction of all sequences (WSREL = proportional sub-sampling), and drew the same amount of sequences per generation (WSABS = unbalanced sub-sampling). For the sub-sampling schemes, we tested the following parameters:

- WSREL: fraction of sequences used for inference $r_{\text{sub}} \in \{0.9, 0.7, 0.5, 0.3, 0.1\}$
- WSABS: fixed amount of sequences at each generation (up to) $a_{\text{sub}} \in \{500, 350, 150, 50\}$

The sequences are randomly chosen without replacement at every time point (below we also evaluate other sampling schemes).

In Figure SN.9, we show violin plots of the percent deviation, akin to Eq. (SN.5) for the raw point estimates (yellow) and the smoothed estimates (green) for different sub-sampling schemes. We can generally observe that ϕ is strongly correlated with the population size N_{true} ($r \geq 0.94$), even in scenarios with significant under-sampling. Moreover, ϕ is an unbiased predictor of the incidence (median deviation centered around 0). The deviation from the linear correlation line slightly increases with more severe under-sampling, i.e. the ϕ estimates are still linearly correlated with N_{true} , but the ϕ estimates become less accurate.

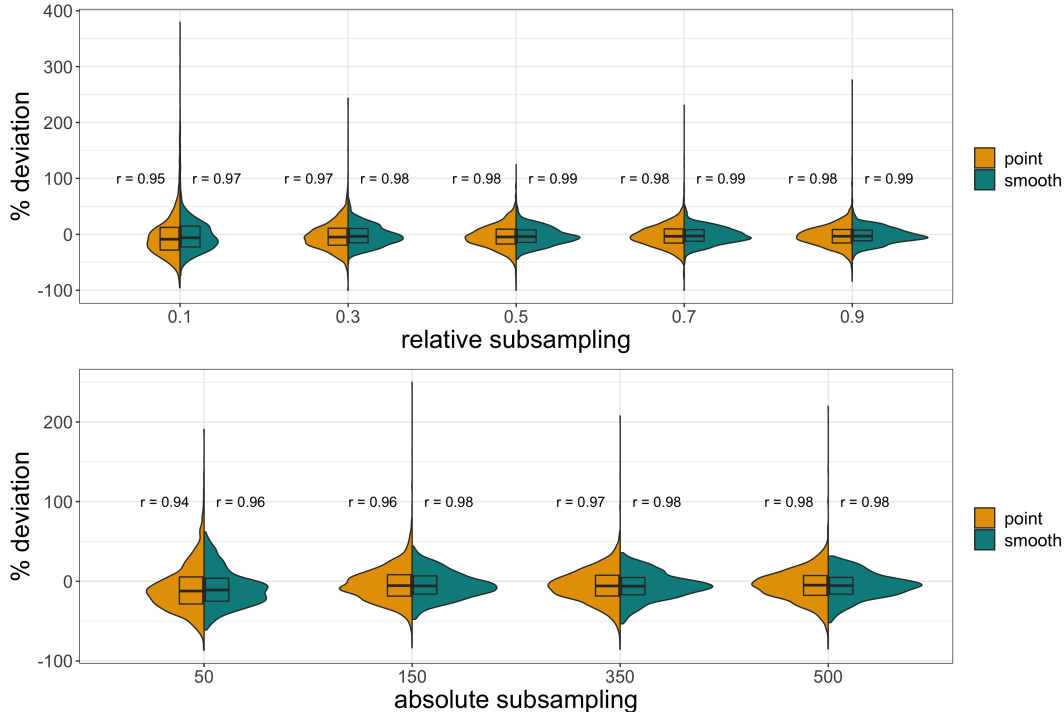


Figure SN.9: Split violin plot showing the percent deviation of the linear fit for ϕ estimates for respective levels of subsampling. The box indicates the interquartile range (IQR) with the median. The left, yellow part shows the deviation for the point estimates, the right, green part depicts the deviation for the smoothed estimates. On each side of the violin plot, the r value indicates the respective level of linear relationship of ϕ and N_{true} . **Top:** Proportional sub-sampling. **Bottom:** Absolute sub-sampling.

SN.1.8 Effects of time-dependent sampling rate on incidence reconstruction

In order to further investigate whether there are any effects of the sampling proportion on incidence reconstruction, we assessed the prediction for a quasi-constant population size with different scenarios of changing subsampling over time. For this, we chose the growth rate $\rho(t) = 1$, i.e. the population of infected individuals will stochastically fluctuate around its initial value, Fig. SN.10, upper left.

First, we inspect the incidence reconstructions without temporal changes in sub-sampling. In Fig. SN.10 (upper right), we show the incidence reconstructions for which 100% of the data is available for GInPipe. In Fig. SN.10 (lower left and right), the median reconstructions are shown for sub-sampled data.

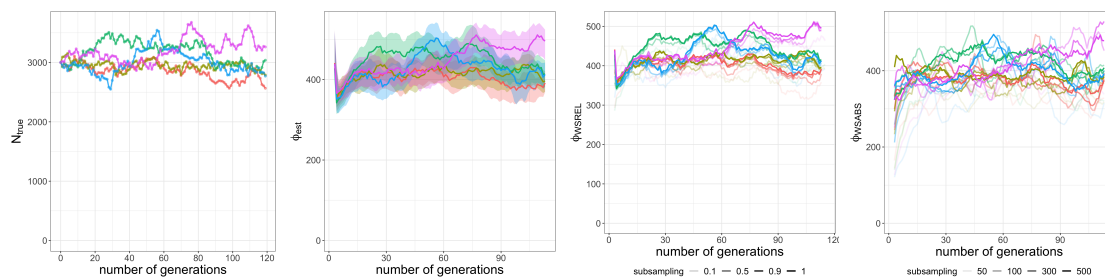


Figure SN.10: Simulations for pandemic dynamics with no apparent trend. **Left:** True population size over time for 5 simulations starting with 3000 sequences and no increase/decrease in the population of infected individuals other than stochastic noise. **Center left:** The solid lines denote the smoothed median of the ϕ estimates for the stochastic simulations. The shaded areas denote the 90% confidence interval. **Center right:** Smoothed median of the ϕ estimates with relative sub-sampling of sequences with proportions indicated below the graphic. **Right:** Smoothed median of the ϕ estimates with absolute sub-sampling with number of sequences as indicated below the graphic.

We observe that the sub-sampling does not affect our estimates drastically. However, the sub-sampling is not time-dependent in these simulations.

Next, we inspect the effect of drastic changes in the sampling rate. For this, we evaluate oscillating changes in the sampling rate, Fig. SN.11 (upper right panels), as well as a sudden increase from a sampling rate of 0.1 to 1.0 Fig. SN.11 (lower left panels) and from 0.5 to 1.0 Fig. SN.11 (lower right panels).

In essence, we observe no systematic effect of temporal changes in the sampling proportion on the incidence reconstructions. This suggests that sudden increases of genomic surveillance of SARS-CoV-2, as observed in many countries after the appearance of the alpha variant (B.1.1.7), do not impact incidence reconstructions using GInPipe. For the extremely low sampling rates in the oscillating case (around generation 45 and 105) we observe inaccuracies and large variance. Therefore, we conclude that temporal changes in the sampling rate do not introduce systematic biases for incidence reconstruction. In line with previous observations (Fig. SN.4), we can however conclude that very low amounts of data increase uncertainty in the incidence reconstructions.

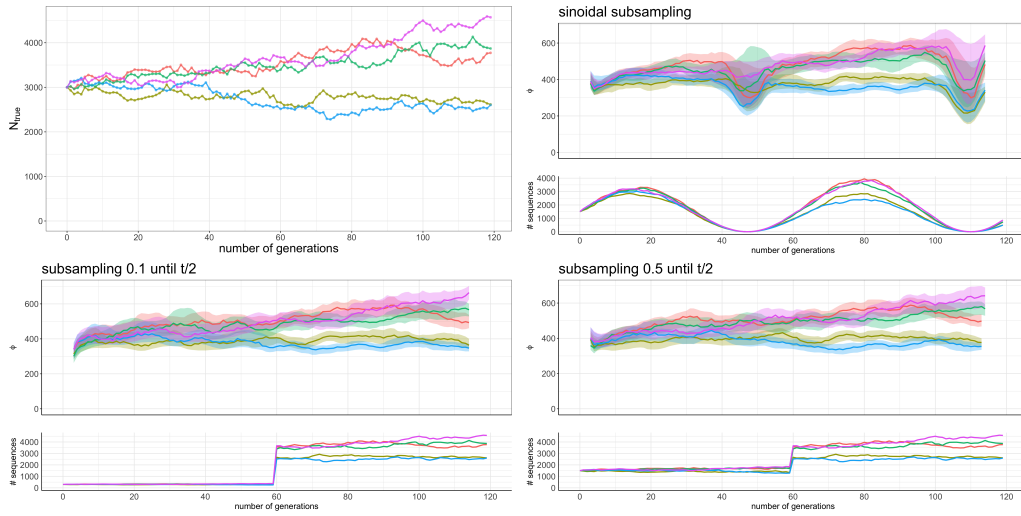


Figure SN.11: Incidence reconstructions with steady population size and time-dependent sub-sampling. Incidence reconstruction shown as smoothed median of the ϕ estimates (solid lines). The coloured shades present the area between the 5th and 95th percentile derived from sub-sampling over the point estimates. The panels below the incidence reconstruction show the number of sampled sequences for the prediction. **Top left:** Simulated population dynamics (5 simulations). **Top Right:** Reconstructed incidence history when the sampling proportion is oscillating. **Bottom left:** Reconstructed incidence dynamics when the sampling proportion abruptly increases from 0.1 to 1.0. **Bottom right:** Incidence reconstructions when the sampling proportion abruptly changed from 0.5 to 1.0.

SN.1.9 Incidence reconstruction when sequences are sampled by similarity

In the previous sections, we assumed that sequence sampling is representative, i.e. that the sub-sampled sequences are randomly selected. In this section, we evaluate if GInPipe is able to reconstruct incidence histories when sequences are sampled with an similarity-bias. This is supposed to emulate extreme events when sequencing is entirely unbalanced, e.g. there is a large under-detection and detection is driven *entirely* by contact tracing. For example, after an infected individual is detected, any *forward* and *backward* contacts are investigated and positive cases are sequenced. At the same time, there would be very little diagnosis without a known contact. The sequenced samples would then be highly similar to the sequence from the first detected case, if they are linked by transmission events.

For the sequence sub-sampling process, instead of choosing the sequences randomly with the same probability, we assign a similarity value $d_t(s)$ for each sequence s to the previous generation, which is used as a probability measure for being sampled in the next generation. The similarity measure is given by the hamming distance (number of differing nucleotides) to the preceding sequence. If the previous generation contains various sequence types, the minimum distance is chosen:

$$d_t(s_i) = \min_{s_j} (d(s_i, s_j)), \forall s_i \in S_t, s_j \in S_{t-1}, \quad (\text{SN.6})$$

with S_t being the sequence population at time t , and S_{t-1} the sequence population in the preceding generation $t - 1$.

We chose to half the chance of being sampled with every nucleotide difference. The probability p_t of each sequence s_i to be chosen at time t is given by

$$p_t(s_i) = \frac{1}{D} \cdot 2^{-d_t(s_i)}, \quad (\text{SN.7})$$

with $D = \sum_{j \in S_t} \frac{1}{2^{d_t(s_j)}}$ being a normalising factor.

Figure SN.12 (top left) shows the simulated population dynamics. Fig. SN.12 (top center) shows the respective incidence reconstructions without sub-sampling (full data set), i.e. the sequences are not sampled by similarity yet. As expected, the incidence estimations with the full data set yields good results ($r = 0.98$ for the smoothed estimates).

When the sequence set is sub-sampled by similarity as described above, we can see a scaling effect, Fig. SN.12 (upper right). I.e., the smaller the sequence set (in terms of the proportion of the entire sequence set), the lower the ϕ values. Interestingly, the profiles are maintained Fig. SN.12 (upper right, Pearson correlations $r \geq 0.95$), while the slope between the populations size and the smoothed incidence estimate decreases with a decreasing proportions of sub-sampled sequences. Therefore, although a strong linear relationship can be maintained with proportional sub-sampling, the scaling factor is affected, Fig. SN.12 (top right and bottom).

This observation is explained by a decrease in the ‘evolutionary signal’ when sequences are sampled by similarity. This effect is most pronounced if the sampled sequence set becomes small.

Sampling non-proportional subsets over time, i.e. taking the same amount of sequences independent of the population size at each generation (50–500 sequences), leads to a irregular re-scaling of the underlying dynamic. The linear relationship of the incidence correlate and the true population size is debilitated (Pearson correlation ranging from 0.85 to 0.95 for the smoothed ϕ estimates). This effect is depicted in Fig. SN.13, where the smoothed ϕ estimates for different levels of similarity-biased sub-sampling for one of the simulations are shown. On the left, the different shades of purple indicate the proportion of sub-sampling. As can be seen, the scaling factor decreases with the amount of sequences, but the dynamic can be retained. However, when non-proportional amounts of sequences are drawn, indicated by the shades of green in the right panel, the sub-sampling-driven scaling effect is uneven over time and the dynamic of the incidence correlate can become distorted. For the case of 500 and 350 sequences per generation, the estimates reach similar values for the first wave (around $t = 30$) compared with the entire data set

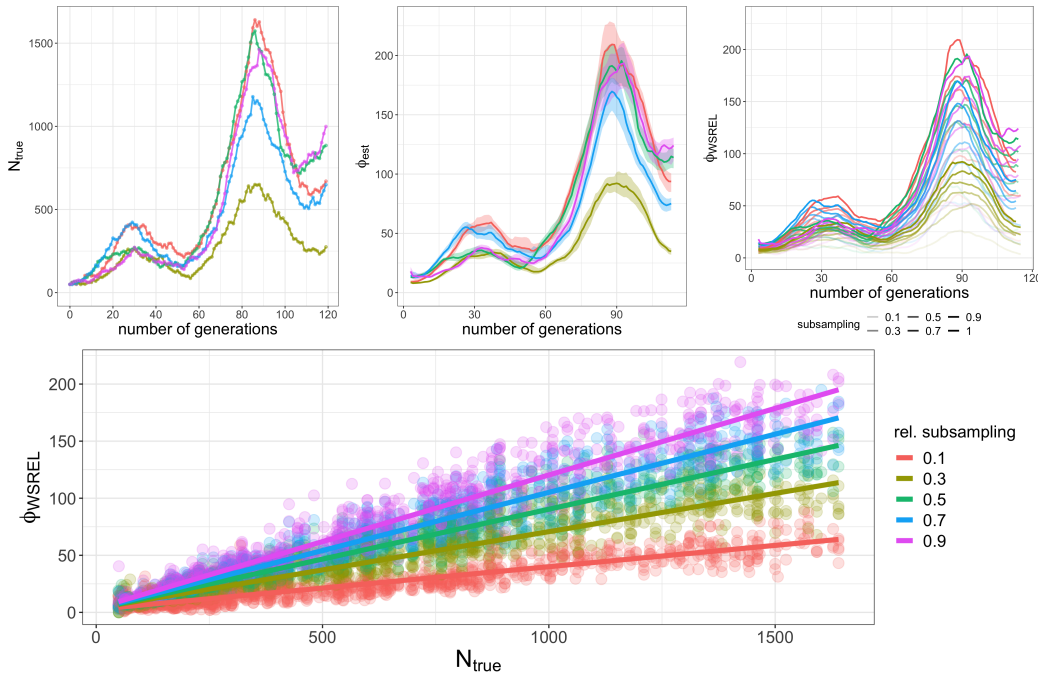


Figure SN.12: Incidence reconstruction when sequences are sampled by similarity. Top left: Simulated population dynamics. **Top center:** The solid lines denote the smoothed median of the incidence reconstructions with GInPipe using all available genetic data (no biased sub-sampling). The coloured shades present the area between the 5th and 95th percentile. **Top right:** Smoothed median of the re-sampled ϕ estimates with relative sub-sampling biased by similarity. **Bottom** Population size N_{true} vs. incidence correlate ϕ for different sub-samplings from the top right panel.

(the black dashed line). However, the second wave (around $t = 90$) is down-scaled and does not reflect the different magnitudes of the two waves.

In conclusion, when sequences are sampled by similarity, the ‘evolutionary signal’ decreases and the scaling between the true population size and the incidence estimate decreases. If this biased sampling is not changing through time, incidence histories can still be reconstructed. If the sampling strategy would change over the time course of the pandemic (e.g. from biased to representative sampling), difficulties with GInPipe-based incidence reconstruction may arise. Most other methods (e.g. phylodynamics) are likely to encounter similar difficulties when the data is systematically perturbed. For GInPipe, and likewise for any other method that utilizes pathogen genome data, a representative sampling of the population is important to guarantee that meaningful results are obtained.

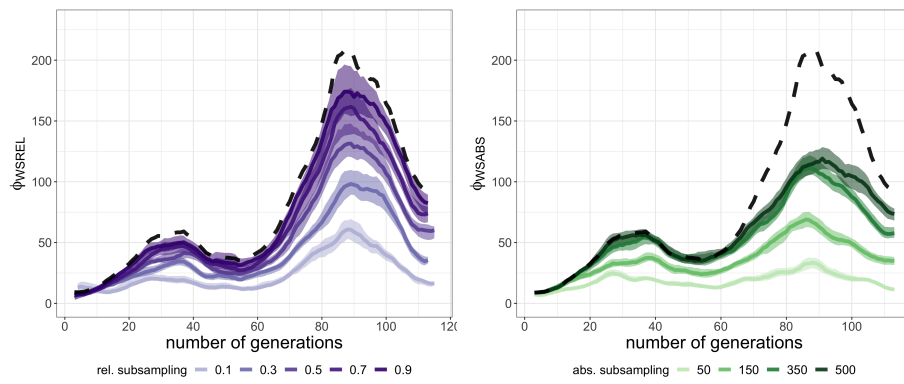
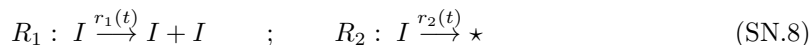


Figure SN.13: Incidence reconstruction with similarity-biased sampling for different levels of sub-sampling for one of the simulations. The shades indicate the level of sub-sampling. The black dashed line shows the estimates with all data. **Left:** Proportional sub-sampling. **Right:** Absolute sub-sampling.

SN.1.10 Incidence reconstructions for a continuous-time population dynamic

In the previous sections, we used a minimalistic approach for simulating the outbreak dynamics. In particular, we performed a time-discretisation, i.e. for each discrete time $t \in (0, 1, \dots, 120)$, the population size was estimated by sampling from a Poisson distribution, $N(t+1) \sim \mathcal{P}(N(t) \cdot \rho(t))$. Here, to verify that the time discretisation has absolutely no effect on the *in silico* assessment of the GInPipe method, we also provide simulations using the time-continuous birth-death process. We introduce the two reactions:



where R_1 denotes the ‘birth’ reaction, where an infected individual infects another individual. Reaction R_2 denotes the ‘death’ reaction, where an infected individual becomes non-infectious. The reaction rates of the model above are given as $r_1(t) = \beta(t) \cdot I(t)$ and $r_2(t) = I(t) \cdot \delta$. Furthermore, we set $\delta = 0.5$ (1/day) and then derive $\beta(t) = \rho(t) \cdot \delta$, where $\rho(t)$ is given in eq. (SN.1). We then simulate the time-continuous model using the Gillespie algorithm, which draws the time Δt to the next event (firing of reaction R_1 or R_2) from an exponential distribution. After each population dynamic event, we update the number of sequences (either cloning or deleting one sequence) and place $n_{\text{mut}} \sim \mathcal{P}(\mu \cdot N(t + \Delta t) \cdot L \cdot \Delta t)$ mutations onto the set of sequences. After simulating a few outbreaks, Fig. SN.14 (left), we use GInPipe to reconstruct the incidence dynamics using the standard setting, as shown in Fig. SN.14 (middle and right). As can be seen, the choice of modelling approach (discrete time vs. continuous time) has no effect of GInPipe’s ability to reconstruct incidence dynamics.

While the population dynamics may also be modelled by more complex S(E)IR-type models, the model above (eq. (SN.8)) can be easily deduced from such models by assuming that a sufficient number of susceptibles is available. However, the choice of structural model does not affect any of the analyses made in Supplementary Note 1.

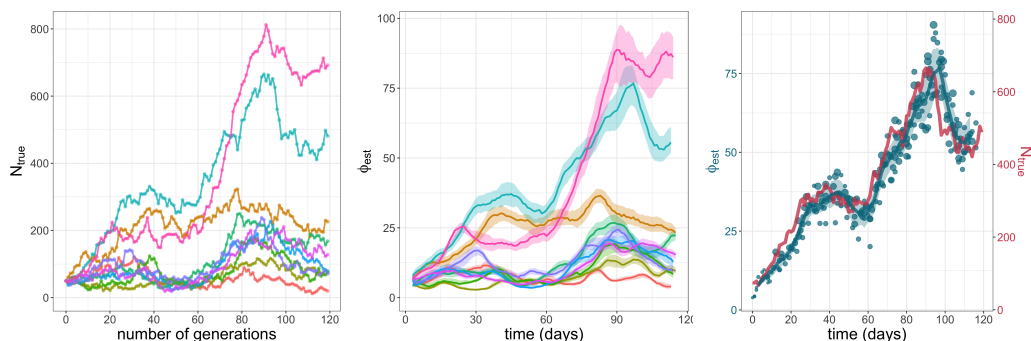


Figure SN.14: Simulations with birth-death-process. Left: True population size. **Middle:** The solid lines denote the smoothed median of the ϕ estimates for the stochastic simulations. The shaded areas denote the 90% confidence interval. **Right:** An example simulation of the epidemic (red curve) vs. the reconstructed incidence history.

SN.1.11 Comparison of R_e estimates

GInPipe allows to infer the effective reproduction number R_e , based on the Wallinga-Teunis method, as described in the *Methods* section in the main manuscript. In the *Results* section, we assess the accuracy of GInPipe by comparing the R_e values inferred from smoothed ϕ estimates with R_e values calculated from the simulated pandemic N_{true} . In Figure SN.15, we see the individual $\log(R_e)$ profiles for the 10 simulations from Section SN.1.2 inferred with N_{true} (bold

cyan lines) and the according smoothed ϕ estimates (red lines). The identity plot in Fig. 1 in the main manuscript shows the $\log(R_e)$ values inferred with N_{true} vs. ϕ , with the respective proportion of qualitatively agreeing or disagreeing predictions in the four quadrants. The accuracy is given by the proportion of qualitatively agreeing R_e values, yielding a value of 0.92 using the complete data set.

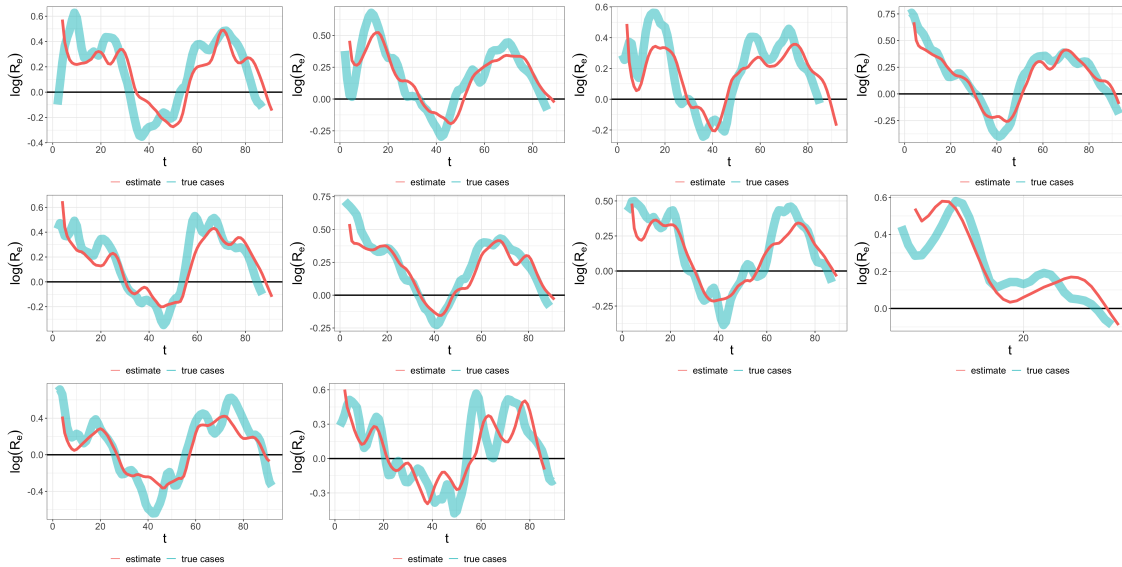


Figure SN.15: Individual R_e profiles. Individual R_e profiles for 10 simulations estimated from the true population sizes (thick cyan lines) and from ϕ estimates inferred with GInPipe (thin red lines).

Complementing the evaluations in the main manuscript, we additionally assess the accuracy for the sub-sampled data (Fig. SN.16). Here, we also observe good results, achieving accuracies between 0.89 and 0.92. In the case of very low sub-sampling proportions (10%), however, the accuracy declines to 78%, due to noisy (“wiggly”) dynamics.

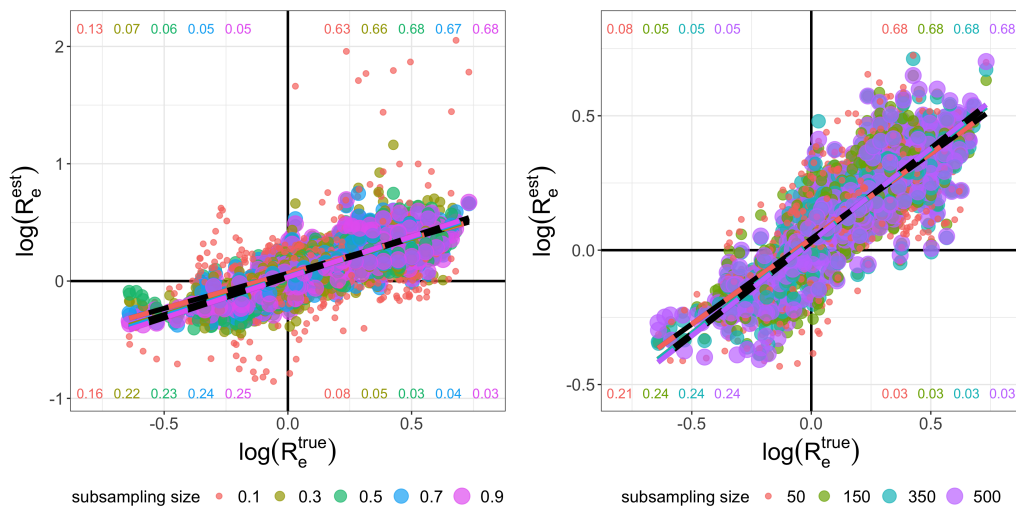


Figure SN.16: Dotplot showing the true $\log(R_e)$ values (x-axis) vs. the R_e values estimated using GInPipe for different levels of sub-sampling. **Left:** Relative sub-sampling (10, 30, 50, 70 or 90% of the data sampled.) **Right:** Absolute sub-sampling (a fixed number of either 50, 150, 350 or 500 sequences are sampled per time point, irrespective of the actual population size).

SN.1.12 Introductions of new sequences do not affect prediction accuracy

Above, we evaluated an isolated outbreak, where the diversity of mutants is emerging solely through evolution, i.e. reproduction and mutation. However, in case of a widespread epidemic, several outbreaks at different locations could lead to populations which evolve in parallel, causing different strains to emerge. Foreign strains may then be introduced into a local epidemic at any point in time.

To test whether the introduction of foreign strains affects the proposed method, we added $x_i \in \{50, 100, 500, 1000\}$ ‘introduction sequences’ at random time points to the initial outbreak. In all simulations, the ‘introduction sequences’ have 3% of the genome randomly mutated with regards to the founder sequence of the local outbreak. We performed 5 simulations that began with 1–10 start sequences. All sequences were considered to evolve with the same replication and mutation rates.

For a representative simulation, Figure SN.17 (left) shows the simulated epidemiological dynamics for different numbers of introductions. On the right, we see the corresponding correlations of the incidence estimate ϕ with the true population size N_{true} . The inferred ϕ estimates show a strong linear correlation with N_{true} across all simulated scenarios ($r = 0.99$, $p < 10^{-16}$).

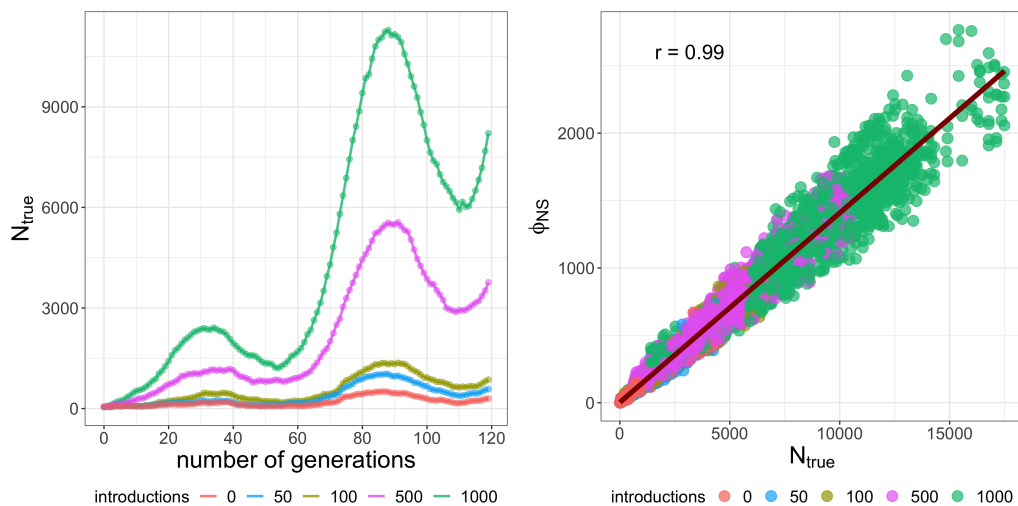


Figure SN.17: Simulations with foreign sequence introductions. **Left:** True population dynamics for the original outbreak (0 introductions) and after introduction of new sequences, each in a representative simulation. The total amount of introduced sequences at random time points is colour-coded. **Right:** Dotplot showing N_{true} versus the point estimates ϕ using 5 simulations per ‘number of introductions’ scenario. The colours indicate the number of introduced foreign sequences to the original outbreak. The black line shows the linear fit.

In Figure SN.18 we evaluate how the relative amount of introductions affects the percent deviation, akin to Eq. (SN.5). We can see that for the majority of the simulations the deviation is around zero with no apparent trend of increasing dispersion (prediction inaccuracy), when large numbers of foreign variants become introduced.

In summary, the predictions are not perturbed by introductions of foreign sequences. Potential biases from single introductions are compensated by ongoing local evolution (community transmission) of the imported sequences.

To investigate the effect of introductions without the compensating effect of community transmissions, we simulated an extreme scenario, where the introduced sequences are only added to the local outbreak and do not replicate or mutate. They simply occur as additional haplotypes in the population. In reality this would be the case, if infected people from a different region get tested,

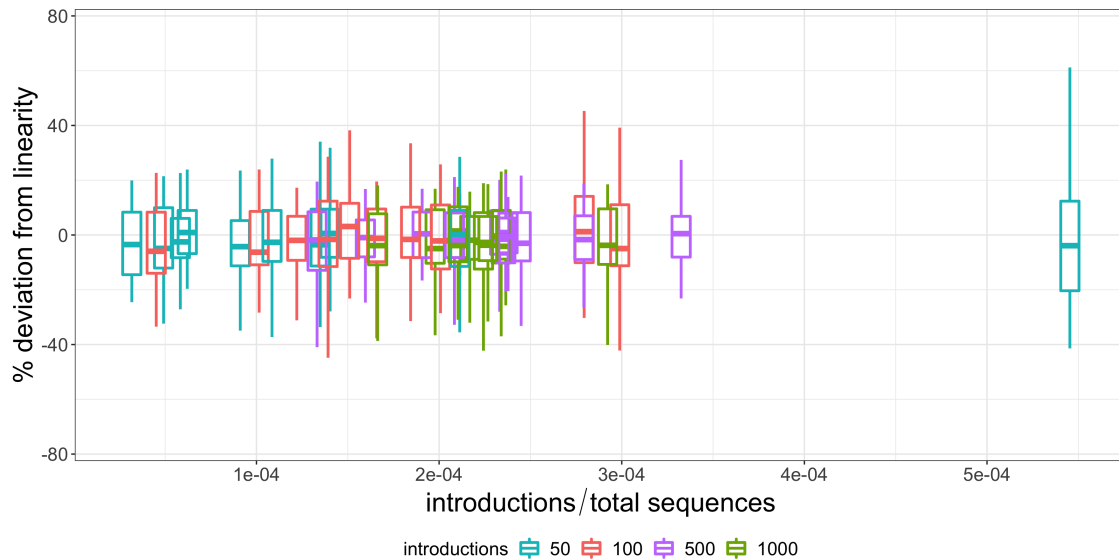


Figure SN.18: Deviation from linear correlation for smoothed ϕ with introduced sequences. Boxplots showing for each simulation the percent deviation of smoothed ϕ from the linear fit to N_{true} in comparison to the ratio of introductions and total amount of emerging sequences in the evolutionary process. The box includes the interquartile range (IQR) with median, together with the lower and upper $1.5 \times \text{IQR}$ whiskers. The colour indicates the number of introductions.

their virus sequenced and subsequent quarantine would completely prevent further transmission of imported variants.

For illustration, Figure SN.19 (top left) shows a simulation where introduced variants are not onwards transmitted. Unlike in Fig. SN.17 (left), where imported variants spread, we can see that the outbreak dynamics are not affected by the non-replicating introductions.

The corresponding ϕ estimates in Figure SN.19 (top right) show the following: In simulations, where 50 or 100 variants were introduced, the corresponding ϕ estimate remain quantitatively unchanged in comparison to a ‘zero-introduction’ simulation. In the extreme case, where 500, and even 1000 variants were introduced, (pink and green lines in Figure SN.19, top right), the dynamical profile of the number of infected individuals is still very closely captured (highlighting the 2.5 waves of infection). However, quantitatively the ϕ estimates are slightly increased in comparison to the ‘zero-introduction’ simulation. Thus, if one would like to compare a setting, where there are many non-replicating introductions, to a setting with replicating- or ‘zero-’ introductions, the quantitative interpretation of the ϕ estimates with regards to the true number of infected individuals N_{true} might differ to some extent. There may be a slightly different slope for the linear correlation function of ϕ vs. N_{true} .

In Figure SN.19 (bottom left), we systematically assess whether large relative amounts of non-replicating introductions decrease prediction accuracy, i.e. increase deviation from the linear regression function for the specific simulation. We do not discover a systematic trend, even if there were more non-replicating introductions than regionally transmitting viruses. We monitor the relative amount of introductions over time for the corresponding smoothed ϕ estimates in the panel above (Fig. SN.19, bottom right). Looking at the ratios of introductions vs. the effective population size per generation for the case of 1000 introductions (green line), we observe that the method is still able to reconstruct the dynamic and attain a linear relationship, despite having consecutively more than 15% introductions during the first wave and partly almost reaching the number of evolving sequences.

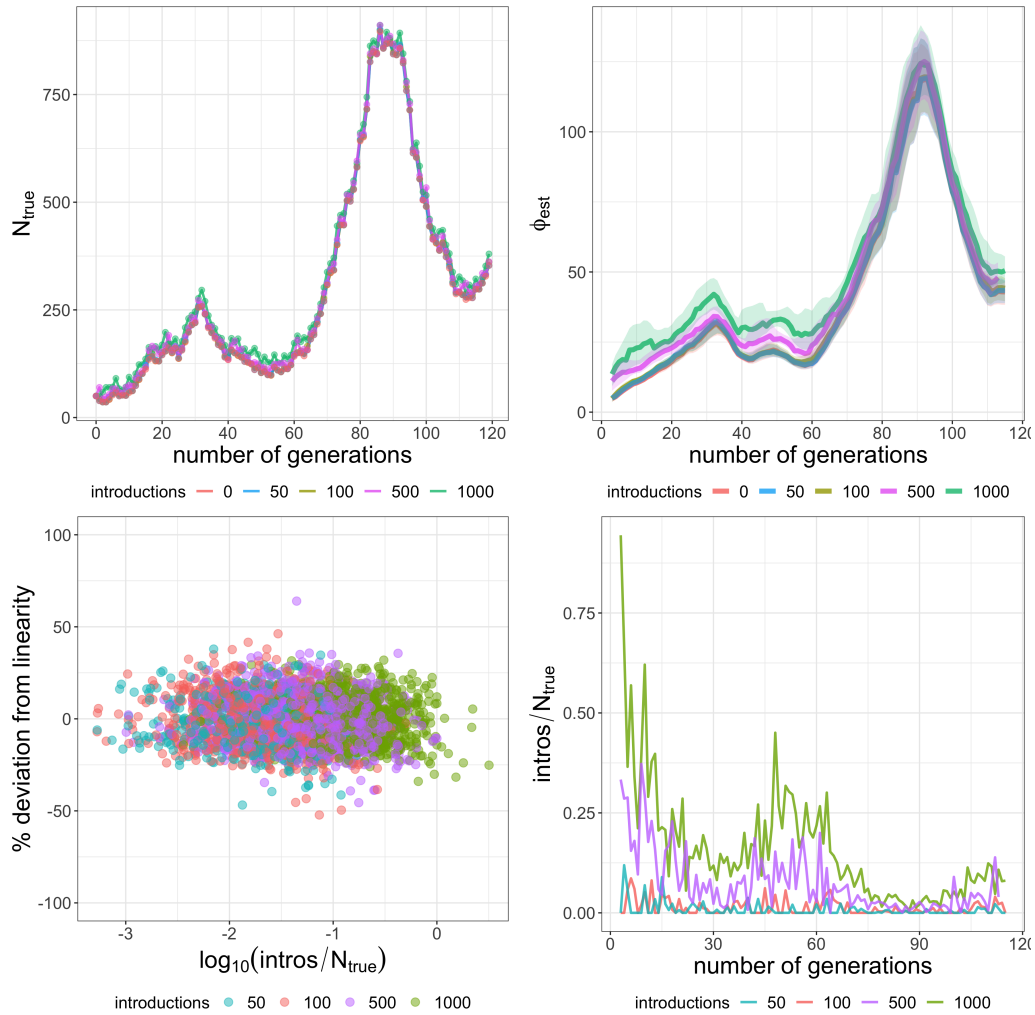


Figure SN.19: Simulations for introduced variants that do not spread in the population. The total amount of introduced sequences at random time points is colour coded. **Top left:** True population size over time for the original outbreak (0 introductions) and after introduction of new sequences for a representative simulation. **Top right:** The solid lines denote the smoothed median of the ϕ estimates for one stochastic simulation with different amounts of introductions. The shaded areas denote the 90% confidence interval. **Bottom left:** Dotplot showing the percent deviation of the smoothed ϕ estimates from the respective linear fit to N_{true} in relation to the relative frequency (\log_{10}) of introductions per time point. **Bottom right:** Relative amount of new introductions over time, corresponding to the simulated trajectories in the top right figure.

SN.1.13 Reconstruction of extreme (un-smooth) dynamics

To further test the limits of GInPipe, we set up two experiments with population dynamics that resemble step-functions. Obviously, these dynamics are quite unlikely to happen in reality and are therefore considered as extreme cases to push the pipeline to the limit.

In the first experiment, the population of infected individuals stochastically fluctuates around its initial value of ≈ 2000 for 60 generations, after which the population suddenly decreases by a factor 10 ($N(t = 61) = \mathcal{P}(N(t = 60) \cdot 0.1)$, Fig. SN.20) and henceforth remains around a value of ≈ 200 for the remaining time points. In the second experiment, the population of infected individuals stochastically fluctuates around its initial value of ≈ 100 for 60 generations, after which it increases by a factor five ($N(t = 61) = \mathcal{P}(N(t = 60) \cdot 5)$, Fig. SN.21).

Fig. SN.20 (upper left) shows 10 simulations where the population sizes suddenly decrease by a factor 10. Fig. SN.20 (upper right) shows the corresponding smoothed median ϕ estimates and Fig. SN.20 (lower panels) shows two individual simulations and the corresponding incidence reconstructions.

Notably, while GInPipe generally reconstructs the population dynamics quite well, we do observe a “smearing out” when the population size changes drastically at generation 60, Fig. SN.20 (upper left vs. upper right). This “smearing out” is likely due to the binning of temporally adjacent sequences, and may be overcome by defining bins that do not span the region of abrupt change (if the latter was known).

Likewise, for the case of the step-function that increases the population size, Fig. SN.21 (upper left), we observe reasonable good incidence reconstructions, Fig. SN.21 (upper right). However, while many reconstructed incidence histories map the population dynamics reliably (example in Fig. SN.21, lower left), we can also observe more drastic delays in some simulations Fig. SN.21, lower right). This effect appears to be stronger for the step function increasing the populations size (Fig. SN.21), than for the step function decreasing the population size (Fig. SN.20). We believe that this effect is mainly due to a sampling bias with regards to the number of new mutations (i.e. in some simulations the new mutations appear later than the population size increases). In fact, such extreme dynamics could in theory appear if a single super-spreading event is responsible for the vast majority of new infections. In view of the spatial scales we analysed (countries or federal states) it is however unlikely that a single event dominates the pandemic. Our assessments of GInPipe for non-smooth population dynamics (step-wise changes of the number of infected individuals) revealed that the pipeline, with default setting may “smear out” abrupt changes in the number of infected individuals. For sudden changes that increase the number of infected individuals, temporal delays may be encountered. However, we have to note that this technical test may have limited relevance to a real-world application where the population of infected individuals changes rather smoothly over time.

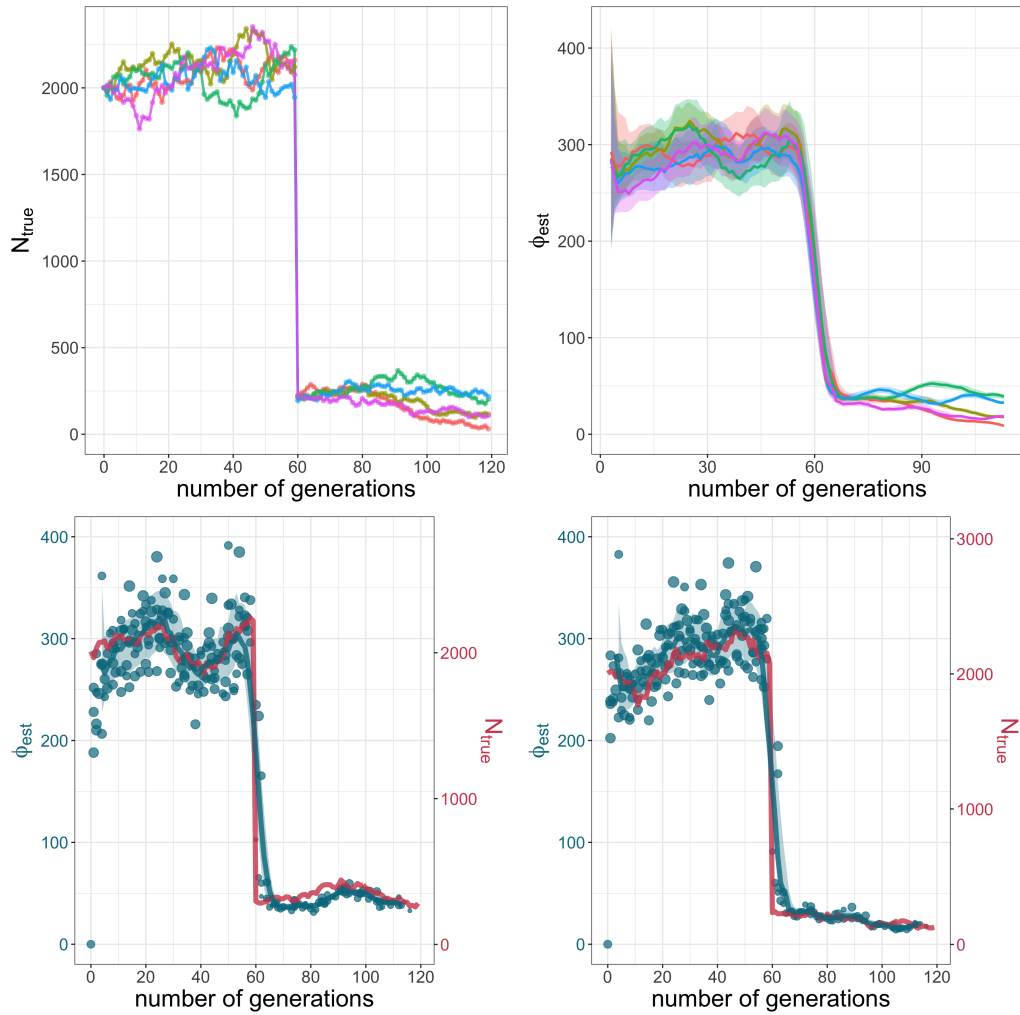


Figure SN.20: Simulations for a drastic decline of infected individuals. Top left: Simulated pandemic trajectories (5 simulations) starting with 3000 infected individuals that drop by a factor of ≈ 10 after 60 generations. **Top right:** The solid lines denote the smoothed median of the ϕ estimates for the respective stochastic simulations. The shaded areas denote the 90% confidence interval. **Bottom left and right:** Two individual simulations (red lines) and the corresponding incidence reconstructions (blue dots, lines (median) and shaded areas (90% confidence interval))

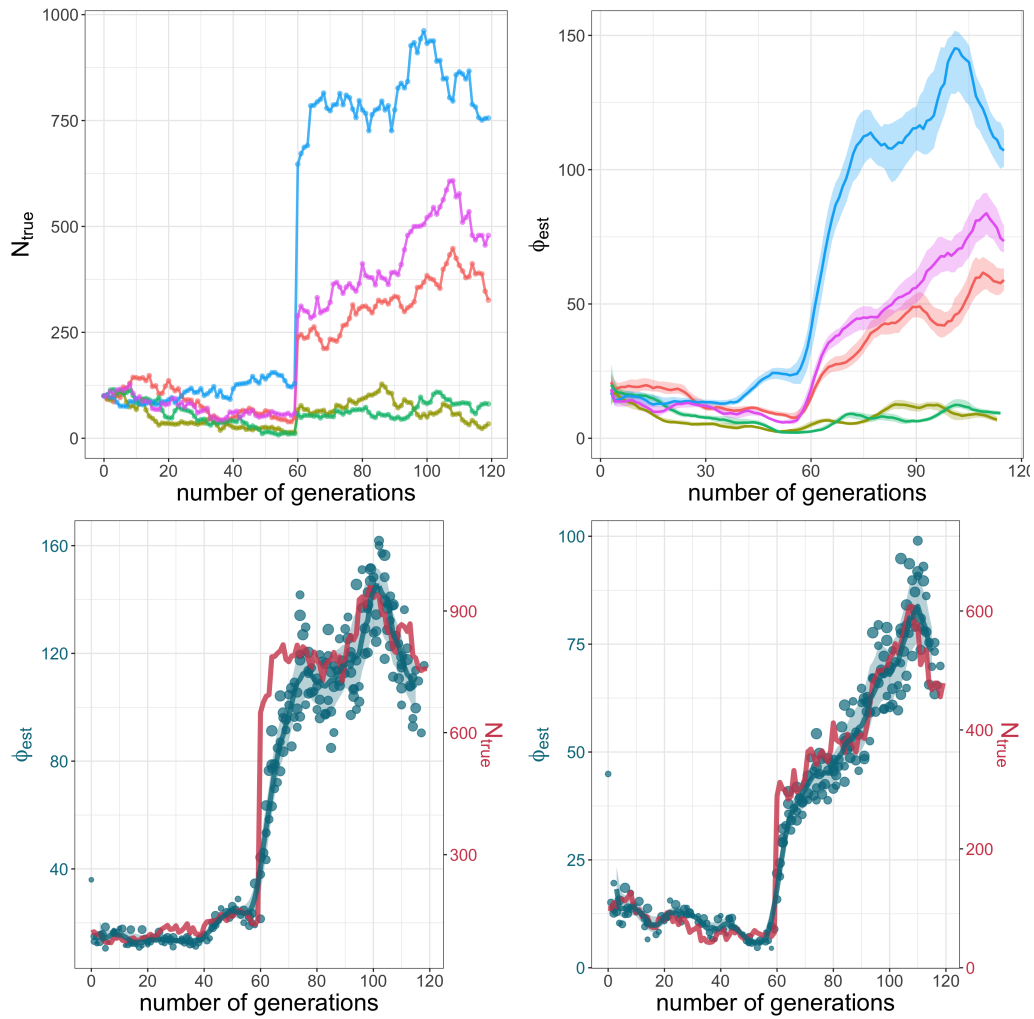


Figure SN.21: Simulations for a drastic increase of infected individuals. Top left: Simulated pandemic trajectories (5 simulations) starting with 100 infected individuals that increase by a factor of ≈ 5 after 60 generations. **Top right:** The solid lines denote the smoothed median of the ϕ estimates for the respective stochastic simulations. The shaded areas denote the 90% confidence interval. **Bottom left and right:** Two individual simulations (red lines) and the corresponding incidence reconstructions (blue dots, lines (median) and shaded areas (90% confidence interval))

SN.1.14 Effect of selective pressure on incidence reconstruction

In next scenario, we wanted to analyse if selection pressure affects GInPipe’s ability to reconstruct the incidence history. To this end, we assign a fitness value $f(i, m_i)$ to each mutation m_i at position i . We then compute the fitness of a sequence f_s as the product of the fitness of all mutations contained in that sequence. Sequences s are then chosen to replicate based on their fitness, i.e. with probability $f_s / \sum_r f_r$. The mutations can either be beneficial (fitter than the wild type) or deleterious (lower survival chance than the wild type).

Individual fitness values were drawn from a log-normal distribution. To assess different levels of stochasticity and yet derive comparable simulations, we considered $\ell \in N_f \subset L$ positions to affect fitness (with $f(\ell, m_\ell) \neq 1$). The remaining positions $k \in L \setminus N_f$ were considered to be neutral ($f(k, m_k) = 1 =$ wild type fitness). We chose the single fitness values according to

$$f(\ell, m_\ell) = e^X, \text{ with } X \sim \mathcal{N}(0, \sigma^2) \tag{SN.9}$$

The fitness of a sequence s is given by the product the position-wise fitness values $f(i, m_i)$ in the respective nucleotide state m_i .

$$f_s = \prod_{i=1}^L f_s(i, m_i) \tag{SN.10}$$

$$= \prod_{i=1}^L \exp(X_i) \tag{SN.11}$$

$$= \exp\left(\sum_{i=1}^L X_i\right) \tag{SN.12}$$

Our total variance, which denotes the convolution of all mutation-wise variances, was set to $\sigma_{\text{tot}}^2 = 0.2$. This setting was chosen, such that the overall change in fitness remains within a realistic range (e.g. the fitness advantage of the SARS-CoV-2 Delta variant over the wild type is assumed to be ≈ 2). Consequently, the variances of the individual fitness values $f(\ell, m_\ell)$ were set to $\sigma^2 = \frac{\sigma_{\text{tot}}^2}{N_f}$. For

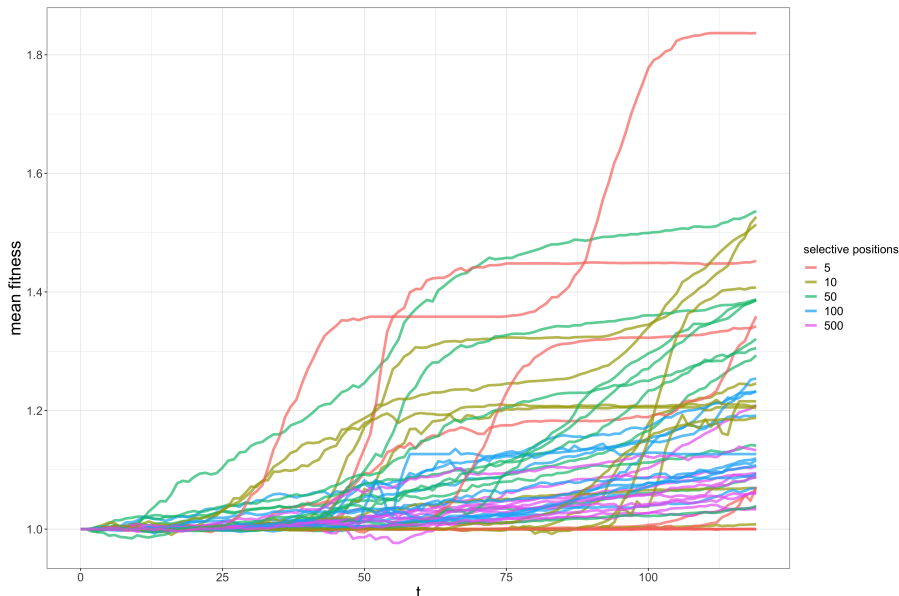


Figure SN.22: Population-average fitness for 10 simulations respectively with $N_f \in \{5, 10, 50, 100, 500\}$ selective sites.

illustration purposes, Figure SN.22 shows the *population average* fitness for 10 simulations each, using the evolutionary setting as above, with $N_f = \{5, 10, 50, 100, 500\}$ non-neutral sites. With the choice of a normalised variance, smaller numbers of non-neutral sites correspond to larger fitness changes per site, making the simulations with e.g. $N_f = 5$ more stochastic. On the contrary, with $N_f = 500$, each non-neutral mutation confers very small changes with regards to fitness, making the evolutionary adaptation more transient. The plot shows that the composition of the sequence set is driven by selective pressure: I.e. the sequence composition is changed such that the average fitness increases over time.

Fig. SN.23(top left) shows the simulated population dynamics using the evolutionary setting with $N_f = 50$ non-neutral sites. Fig. SN.23(top right) shows the GInPipe reconstructed incidence histories and Fig. SN.23(lower panels) shows the reconstructed incidence histories with different sub-sampling schemes (relative vs. absolute sub-sampling). As can be seen in Fig. SN.23, GInPipe can reconstruct incidence histories very well when particular positions are under selective pressure.

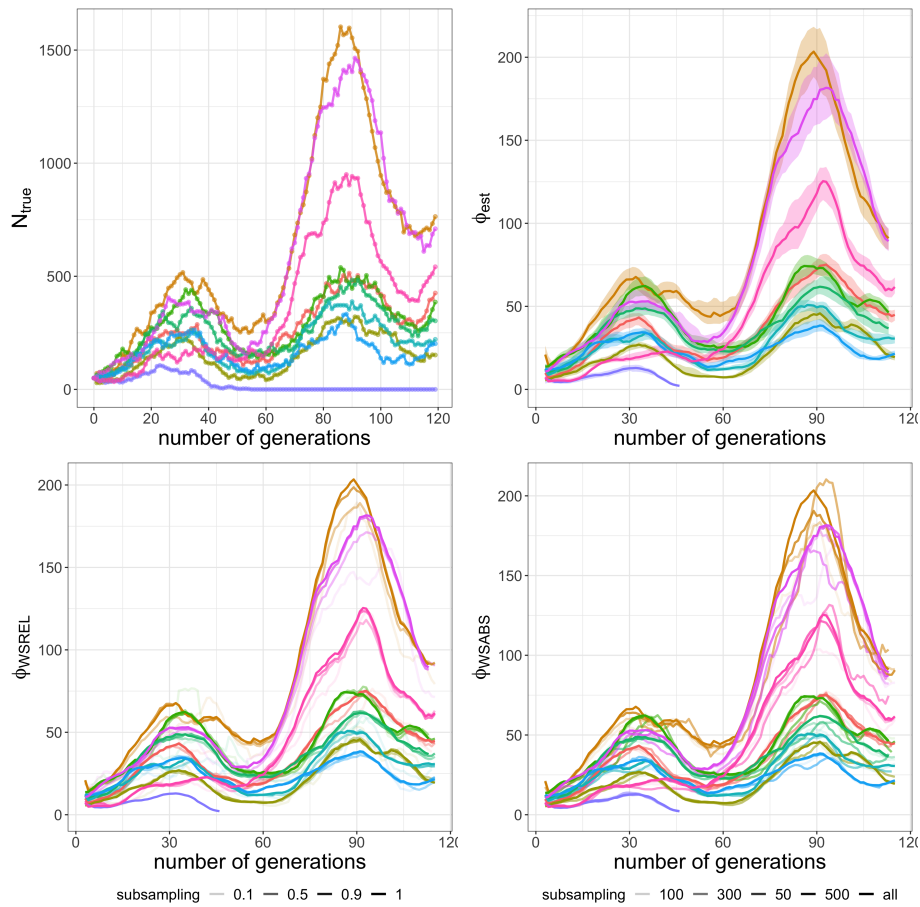


Figure SN.23: Simulations with selective pressure. Top left: Population dynamics with selection pressure. **Top right:** The solid lines denote the smoothed ϕ estimates using all sequence data. The shaded areas denote the 90% confidence interval. **Bottom left:** Smoothed ϕ estimates with relative sub-sampling. **Bottom right:** Smoothed ϕ estimates with absolute sub-sampling.

SN.1.15 Effect of mutation rate on incidence reconstruction

In this section we assess how changes in the mutation rate affect the incidence reconstruction with GInPipe. Note that a severely lowered mutation rate would decrease the evolutionary information contained in any sequence set and increase stochasticity, by turning evolutionary events into rare events. We performed simulations with the same setting as in section SN.1.3, but this time increased or decreased the mutation rate 10-fold.

10-fold lowered mutation rate. Simulations using a low mutation rate of $\mu = 0.00001$ are depicted in Fig. SN.24 (upper row, left most panel). Reconstructed incidence dynamics using GInPipe with default settings are shown in Fig. SN.24 using the full data set (upper row, center left panel), as well as sub-sampled data sets (upper row, the two panels on the right). For the full data set, GInPipe can still reconstruct the dynamics (Pearson correlation of $r = 0.9$), but for the sub-sampled data sets the reconstructions worsen (Pearson correlation coefficients of $r = 0.7-0.9$).

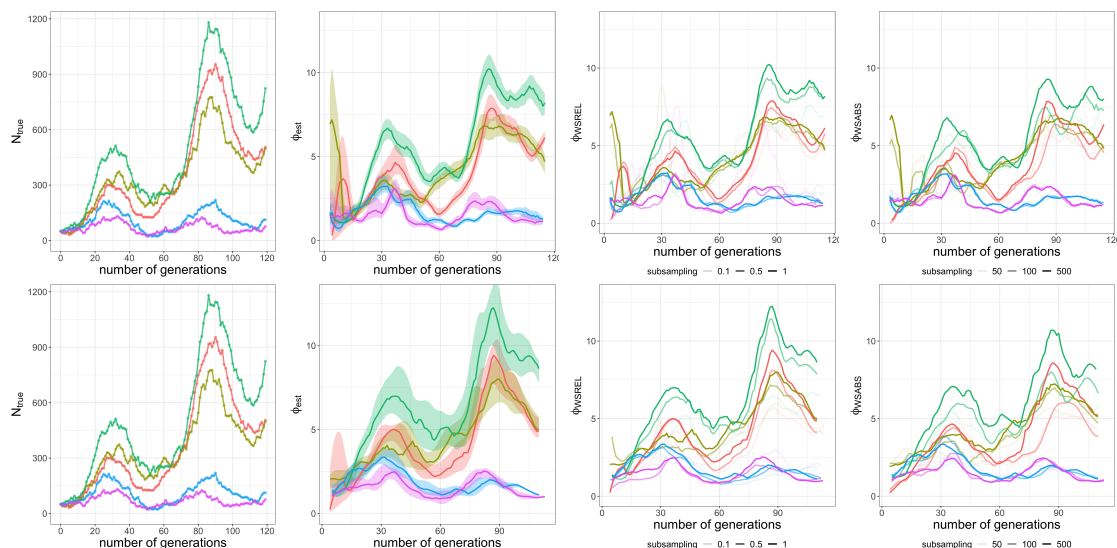


Figure SN.24: Incidence reconstructions with a 10-fold lowered mutation rate ($\mu = 0.00001$). **Top row:** Original binning strategy. **Bottom row:** Adjusted binning strategy. **Left:** Simulated pandemic trajectories (5 simulations). **Center left:** The solid lines denote the smoothed median of the ϕ estimates for the respective stochastic simulations using all sequences. The shaded areas denote the 90% confidence interval. **Center right and rightmost panel:** Smoothed median of the ϕ estimates for the respective stochastic simulations using subsets of the sequences (center right: relative subsampling, far right: uneven subsampling). The transparency of the curve indicates the level of subsampling.

Work-around: Increasing bin sizes. As mentioned above, for very low mutation rates, the binning strategy may produce sequence sets that contain no evolutionary information. Consequently, GInPipe may not robustly infer incidence estimates on these data sets. However, increasing the bin sizes may revert this situation. To demonstrate this presumption, we changed the binning strategy (compare to *Methods* section in the main manuscript) such that the bin sizes increase. In particular, we set: (i) the days per bin to $\{10, 15, 20, 25, 30\}$, (ii) the sequences per bin to $\{5, 10, 15, 20\}$ percent of the entire sequence set and adjusted the (iii) constraints, such that each bin must span at least 5 days and not more than 30 days and contain at least 20 sequences.

As can be seen in Fig. SN.24, this helped to recover the original dynamics (Pearson correlation of $r = 0.95$ (full data set), and $0.87-0.94$ for the sub-sampled data sets), however at the cost of temporal resolution (e.g. the final increase in incidence cannot be recovered).

A comparison of the original binning strategy with the adjusted binning strategy (larger bins) is shown in Fig. SN.25 for one simulation. As can be seen therein (top left panel vs. bottom left panel), the original binning strategy has a much finer time resolution (number of blue dots).

As mentioned before, lowering the mutation rate increases stochasticity, and consequently decreases the signal-to-noise ratio. When sub-sampling the data using the original binning strategy Fig. SN.25 (top center and right panel), this can lead to scenarios where certain bins produce outlier dynamics (large shaded areas). When increasing the bin sizes Fig. SN.25 (bottom center and right panel), these large fluctuations due to extreme outliers could be eliminated.

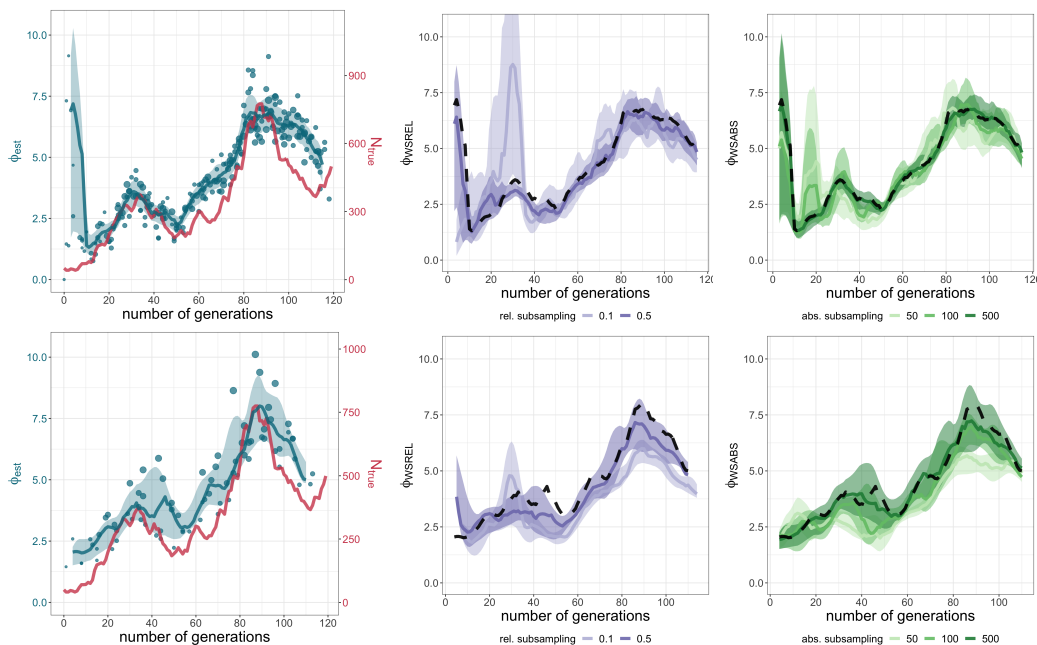


Figure SN.25: Example of incidence reconstructions for one simulation with a 10-fold decreased mutation rate ($\mu = 0.00001$) using different binning strategies. Top row: Original binning strategy as stated in Section SN.1.3. Bottom row: Adapted binning strategy (larger bins). Left: Simulated population dynamics (red line) and GInPipe reconstructed incidence dynamics using all available data (blue dots: ϕ estimates, blue line: smoothed estimate). Center: Incidence reconstructions using relative sub-sampling. The transparency of the lines indicate the level of sub-sampling. Right: Incidence reconstructions using absolute sub-sampling. The transparency of the lines indicate the level of sub-sampling.

10-fold increased mutation rate. Simulations with a 10-fold increased mutation rate of $\mu = 0.001$ are shown in Fig. SN.26 (left) together with GInPipe’s incidence reconstructions using the full data set (center left), as well as sub-sampled data sets (the two panels on the right). For a large mutation rate GInPipe can reconstruct the incidence very well (Pearson correlation of $r = 0.996$). For relative sub-sampling the incidence reconstructions are similarly good ($r > 0.95$), Fig. SN.26 (center right). For absolute sub-sampling Fig. SN.26 (rightmost panel) reconstructions are still very good ($r > 0.92$), except for the lowest absolute sub-sampling (50 sequences), where we observed $r = 0.81$.

Observations and Outlook

In summary, our simulations show that, particularly for low mutation rates, the ‘evolutionary signal’ decreases and that the emergence of novel haplotypes becomes more stochastic. In essence, this means that the signal-to-noise ratio decreases. The loss in evolutionary information can be circumvented by increasing the bin sizes, at the cost of temporal resolution. While the binning strategy presented in the *Methods* is well adapted to SARS-CoV-2, application of GInPipe to other respiratory infections would likely require to adapt the binning strategy to the data. As an outlook, we plan to build filters that automatically adapt binning strategies to the amount of

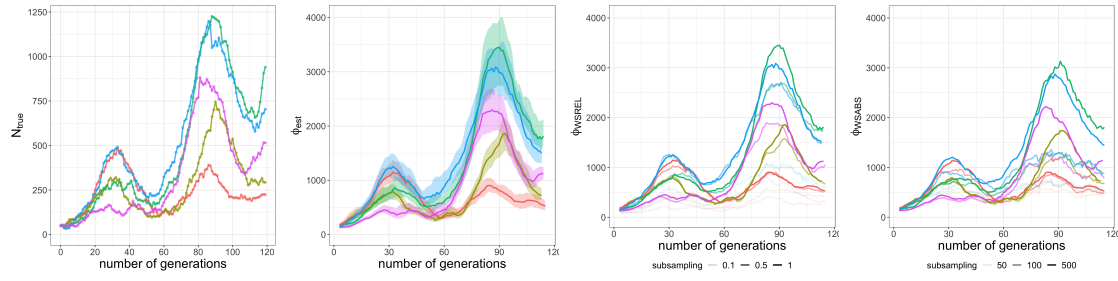


Figure SN.26: Incidence reconstructions with a 10-fold increased mutation rate ($\mu = 0.001$). **Leftmost panel:** Simulated pandemic trajectories (5 simulations). **Center left:** The solid lines denote the smoothed median incidence reconstructions using all sequences. The shaded areas denote the 90% confidence interval. **Center right and rightmost panel:** Smoothed median incidence reconstructions using subsets of the the data (center right: relative sub-sampling; right most: absolute sub-sampling). The transparency of the curve indicates the level of sub-sampling.

evolutionary information in a given data set, and which mark temporal periods where insufficient information is available.

Supplementary Note 2

In this Supplementary Note all parameters used in the phylodynamic inference are stated, such as (i) change dates where major non-pharmaceutical measures were introduced or lifted (these intervals are used to calculate the piece-wise constant $R_e(\tau)$ values); (ii) All final parameter estimates, as well as (iii) ‘lineage-through-time’ plots for phylodynamic reconstruction.

Birth-death-sampling rate change dates

Date	Non-pharmaceutical interventions
28 April 2020	Start of easing of restrictions: Opening of primary schools, later public places
15 June 2020	Further easing of restrictions: bigger private and mass gatherings allowed, home office partially recommended
22 August 2020	Protective masks: Face masks mandatory in closed public spaces
26 October 2020	Reimplementation of light restrictions: private gatherings
26 December 2020	Start of lockdown: closure of schools, restaurants and entertainment venues

Table SN.1: Denmark: Intervention measures used to set the birth-death-sampling rate change dates

Date	Non-pharmaceutical interventions
24 March 2020	Start of national lockdown: Closure of shops, restaurants etc, stay-at-home order
29 May 2020	First lifting of restrictions: Phase 1 of route map out of lockdown
15 July 2020	Phase 3 of route map out of lockdown: Opening restaurants, hairdressers etc.
23 September 2020	Reimplementation of restrictive measures: Restrictions on household visits and national curfew
17 November 2020	Further restrictions: Update on COVID-19 protection levels and travel regulations
05 January 2021	Start of second lockdown for mainland Scotland: stay-at-home order

Table SN.2: Scotland: Intervention measures used to set the birth-death-sampling rate change dates

Date	Non-pharmaceutical interventions
16 March 2020	Start of national lockdown: Closure of shops, museums, restaurants, other gathering places and borders
11 May 2020	Lifting of stricter lockdown measures: Opening of i.a. schools, shops and museums
15 June 2020	Second step in reopening the borders
20 August 2020	Unspecified
19 October 2020	Reimplementation of restrictions: home office recommendation, restrictions on private gatherings and for restaurants

Table SN.3: Switzerland: Intervention measures used to set the birth-death-sampling rate change dates

Date	Non-pharmaceutical interventions
16 March 2020	State of Emergency and first restrictions
11 May 2020	First easing of restrictions
19 July 2020	Reimplementation of restrictions, especially around Melbourne
13 September 2020	First step in roadmap for reopening
26 October 2020	Zero new cases recorded

Table SN.4: Victoria, Australia: Intervention measures used to set the birth-death-sampling rate change dates

Phylogenetic and epidemiological parameter estimates

	median	95% HPDI
become non-infectious rate δ_1	2.5625e+01	[2.3626e+01, 2.7618e+01]
become non-infectious rate δ_2	4.8634e+01	[4.6664e+01, 5.0612e+01]
become non-infectious rate δ_3	4.9078e+01	[4.7126e+01, 5.1035e+01]
become non-infectious rate δ_4	5.0413e+01	[4.8636e+01, 5.2326e+01]
become non-infectious rate δ_5	5.7918e+01	[5.6189e+01, 5.9674e+01]
become non-infectious rate δ_6	5.6963e+01	[5.5201e+01, 5.8822e+01]
reproductive number $R_{e,1}$	1.9810e+00	[1.7938e+00, 2.1807e+00]
reproductive number $R_{e,2}$	7.8879e-01	[6.4151e-01, 9.2349e-01]
reproductive number $R_{e,3}$	9.8603e-01	[9.2555e-01, 1.0466e+00]
reproductive number $R_{e,4}$	1.0571e+00	[1.0197e+00, 1.0928e+00]
reproductive number $R_{e,5}$	1.1692e+00	[1.1397e+00, 1.2008e+00]
reproductive number $R_{e,6}$	1.0611e+00	[1.0354e+00, 1.0884e+00]
sampling proportion s_1	0.0000e+00	[0.0000e+00, 0.0000e+00]
sampling proportion s_2	2.6598e-03	[1.0540e-03, 4.8091e-03]
sampling proportion s_3	3.2960e-04	[1.4211e-04, 5.7886e-04]
sampling proportion s_4	2.0136e-03	[1.3163e-03, 2.8224e-03]
sampling proportion s_5	8.8009e-03	[8.2494e-03, 9.0000e-03]
sampling proportion s_6	7.9828e-03	[7.9259e-03, 8.0000e-03]
sampling proportion s_7	1.2953e-02	[1.2798e-02, 1.3000e-02]
frequency parameter π_A	2.9857e-01	[2.9799e-01, 2.9912e-01]
frequency parameter π_C	1.8344e-01	[1.8295e-01, 1.8389e-01]
frequency parameter π_G	1.9603e-01	[1.9552e-01, 1.9649e-01]
frequency parameter π_T	3.2197e-01	[3.2140e-01, 3.2255e-01]
transition/transversion bias	6.0220e+00	[5.5656e+00, 6.5410e+00]
gamma shape parameter	6.5450e-01	[1.3687e-03, 3.0386e+00]
clock rate	8.0000e-04	[8.0000e-04, 8.0000e-04]

Table SN.5: Denmark: Median and 95% highest posterior density interval (HPDI) of the posterior sample of the model-specific phylogenetic and epidemiological parameters. Subscript of the latter indicates the interval for which it is inferred. Intervals are labelled numerically from the start of the process to the date of the last sample with boundaries as specified in Table SN.1

	median	95% HPDI
become non-infectious rate δ_1	2.6001e+01	[2.4104e+01, 2.8018e+01]
become non-infectious rate δ_2	4.9892e+01	[4.8056e+01, 5.1859e+01]
become non-infectious rate δ_3	4.8036e+01	[4.6171e+01, 5.0075e+01]
become non-infectious rate δ_4	5.0577e+01	[4.8539e+01, 5.2410e+01]
become non-infectious rate δ_5	5.7387e+01	[5.5567e+01, 5.9139e+01]
become non-infectious rate δ_6	5.4086e+01	[5.2269e+01, 5.6012e+01]
become non-infectious rate δ_7	4.9587e+01	[4.7647e+01, 5.1547e+01]
reproductive number $R_{e,1}$	2.3968e+00	[2.1690e+00, 2.6398e+00]
reproductive number $R_{e,2}$	9.6825e-01	[9.3264e-01, 1.0074e+00]
reproductive number $R_{e,3}$	8.8451e-01	[8.0327e-01, 9.6796e-01]
reproductive number $R_{e,4}$	1.4277e+00	[1.3730e+00, 1.4844e+00]
reproductive number $R_{e,5}$	9.5832e-01	[9.2419e-01, 9.9162e-01]
reproductive number $R_{e,6}$	1.2506e+00	[1.2013e+00, 1.2968e+00]
reproductive number $R_{e,7}$	1.2581e+00	[1.1723e+00, 1.3468e+00]
sampling proportion s_1	0.0000e+00	[0.0000e+00, 0.0000e+00]
sampling proportion s_2	4.3825e-02	[3.1846e-02, 5.7429e-02]
sampling proportion s_3	4.2829e-02	[3.6254e-02, 5.0000e-02]
sampling proportion s_4	7.2367e-03	[5.6379e-03, 7.9998e-03]
sampling proportion s_5	3.9707e-02	[3.3169e-02, 4.6317e-02]
sampling proportion s_6	9.9777e-03	[9.9019e-03, 1.0000e-02]
sampling proportion s_7	5.4729e-03	[5.3869e-03, 5.5000e-03]
sampling proportion s_8	7.3337e-03	[6.5973e-03, 7.6000e-03]
frequency parameter π_A	2.9858e-01	[2.9802e-01, 2.9907e-01]
frequency parameter π_C	1.8353e-01	[1.8309e-01, 1.8396e-01]
frequency parameter π_G	1.9605e-01	[1.9564e-01, 1.9659e-01]
frequency parameter π_T	3.2184e-01	[3.2130e-01, 3.2241e-01]
transition/transversion bias	6.0886e+00	[5.6013e+00, 6.6120e+00]
gamma shape parameter	7.1777e-01	[1.5681e-03, 3.0007e+00]
clock rate	8.0000e-04	[8.0000e-04, 8.0000e-04]

Table SN.6: Scotland: Median and 95% highest posterior density interval (HPDI) of the posterior sample of the model-specific phylogenetic and epidemiological parameters. Subscript of the latter indicates the interval for which it is inferred. Intervals are labelled numerically from the start of the process to the date of the last sample with boundaries as specified in Table SN.2

	median	95% HPDI
become non-infectious rate δ_1	2.6673e+01	[2.4635e+01, 2.8638e+01]
become non-infectious rate δ_2	4.9550e+01	[4.7621e+01, 5.1421e+01]
become non-infectious rate δ_3	4.8491e+01	[4.6607e+01, 5.0447e+01]
become non-infectious rate δ_4	4.9093e+01	[4.7222e+01, 5.1156e+01]
become non-infectious rate δ_5	5.4159e+01	[5.2311e+01, 5.5945e+01]
become non-infectious rate δ_6	5.3518e+01	[5.1617e+01, 5.5328e+01]
reproductive number $R_{e,1}$	3.7731e+00	[3.4827e+00, 4.0733e+00]
reproductive number $R_{e,2}$	6.0013e-01	[5.4912e-01, 6.5069e-01]
reproductive number $R_{e,3}$	1.0066e+00	[9.1301e-01, 1.0976e+00]
reproductive number $R_{e,4}$	1.0697e+00	[1.0268e+00, 1.1102e+00]
reproductive number $R_{e,5}$	1.2218e+00	[1.1859e+00, 1.2583e+00]
reproductive number $R_{e,6}$	1.2443e+00	[1.2194e+00, 1.2701e+00]
sampling proportion s_1	0.0000e+00	[0.0000e+00, 0.0000e+00]
sampling proportion s_2	2.8901e-02	[2.2041e-02, 3.6251e-02]
sampling proportion s_3	1.2186e-02	[1.0287e-02, 1.3000e-02]
sampling proportion s_4	5.6926e-03	[3.3451e-03, 8.4517e-03]
sampling proportion s_5	2.1886e-02	[1.7928e-02, 2.5921e-02]
sampling proportion s_6	1.0948e-02	[1.0777e-02, 1.1000e-02]
sampling proportion s_7	1.9905e-03	[1.9590e-03, 2.0000e-03]
frequency parameter π_A	2.9853e-01	[2.9787e-01, 2.9914e-01]
frequency parameter π_C	1.8344e-01	[1.8286e-01, 1.8392e-01]
frequency parameter π_G	1.9598e-01	[1.9550e-01, 1.9656e-01]
frequency parameter π_T	3.2206e-01	[3.2144e-01, 3.2270e-01]
transition/transversion bias	5.1855e+00	[4.8250e+00, 5.5949e+00]
gamma shape parameter	7.7207e-01	[1.0269e-03, 3.1043e+00]
clock rate	8.0000e-04	[8.0000e-04, 8.0000e-04]

Table SN.7: Switzerland: Median and 95% highest posterior density interval (HPDI) of the posterior sample of the model-specific phylogenetic and epidemiological parameters. Subscript of the latter indicates the interval for which it is inferred. Intervals are labelled numerically from the start of the process to the date of the last sample with boundaries as specified in Table SN.3

	median	95% HPDI
become non-infectious rate δ_1	2.6169e+01	[2.4199e+01, 2.8134e+01]
become non-infectious rate δ_2	4.9617e+01	[4.7653e+01, 5.1587e+01]
become non-infectious rate δ_3	5.1130e+01	[4.9280e+01, 5.3008e+01]
become non-infectious rate δ_4	5.3928e+01	[5.2086e+01, 5.5703e+01]
become non-infectious rate δ_5	4.8734e+01	[4.6809e+01, 5.0700e+01]
become non-infectious rate δ_6	4.8756e+01	[4.6749e+01, 5.0668e+01]
reproductive number $R_{e,1}$	1.4985e+00	[1.3534e+00, 1.6480e+00]
reproductive number $R_{e,2}$	8.5330e-01	[7.5130e-01, 9.2257e-01]
reproductive number $R_{e,3}$	1.5831e+00	[1.5003e+00, 1.6653e+00]
reproductive number $R_{e,4}$	8.2566e-01	[7.7671e-01, 8.7320e-01]
reproductive number $R_{e,5}$	2.3464e-01	[3.3521e-07, 5.6699e-01]
reproductive number $R_{e,6}$	2.0824e+00	[1.3328e+00, 2.9249e+00]
sampling proportion $s_{g,1}$	0.0000e+00	[0.0000e+00, 0.0000e+00]
sampling proportion $s_{g,2}$	2.9634e-02	[1.9242e-02, 4.1241e-02]
sampling proportion $s_{g,3}$	1.7328e-01	[1.3147e-01, 2.1603e-01]
sampling proportion $s_{g,4}$	6.5030e-02	[4.3376e-02, 9.1276e-02]
sampling proportion $s_{g,5}$	2.2637e-01	[8.7877e-02, 4.1711e-01]
sampling proportion $s_{g,6}$	1.6848e-01	[8.4786e-03, 4.3823e-01]
sampling proportion $s_{g,7}$	1.7360e-02	[2.1027e-08, 1.0043e-01]
sampling proportion $s_{D,2,1}$	0.0000e+00	[0.0000e+00, 0.0000e+00]
sampling proportion $s_{D,2,2}$	5.0018e-02	[4.0377e-03, 9.8919e-02]
sampling proportion $s_{D,2,3}$	2.5750e-02	[1.6234e-07, 8.5113e-02]
sampling proportion $s_{D,2,4}$	9.9060e-02	[9.6127e-02, 1.0000e-01]
sampling proportion $s_{D,2,5}$	9.9453e-02	[9.7659e-02, 1.0000e-01]
sampling proportion $s_{D,2,6}$	8.5009e-02	[5.6538e-02, 9.9999e-02]
sampling proportion $s_{D,2,7}$	4.9863e-02	[1.6285e-06, 9.4899e-02]
frequency parameter π_A	2.9860e-01	[2.9781e-01, 2.9947e-01]
frequency parameter π_C	1.8357e-01	[1.8284e-01, 1.8423e-01]
frequency parameter π_G	1.9609e-01	[1.9536e-01, 1.9677e-01]
frequency parameter π_T	3.2173e-01	[3.2082e-01, 3.2255e-01]
transition/transversion bias	6.3017e+00	[5.4814e+00, 7.2111e+00]
gamma shape parameter	7.1495e-01	[1.0263e-03, 3.0786e+00]
clock rate	8.0000e-04	[8.0000e-04, 8.0000e-04]

Table SN.8: Victoria: Median and 95% highest posterior density interval (HPDI) of the posterior sample of the model-specific phylogenetic and epidemiological parameters. Subscript of the latter indicates the interval for which it is inferred. Intervals are labelled numerically from the start of the process to the date of the last sample with boundaries as specified in Table SN.4

Identified Pango lineages and tree reconstructions

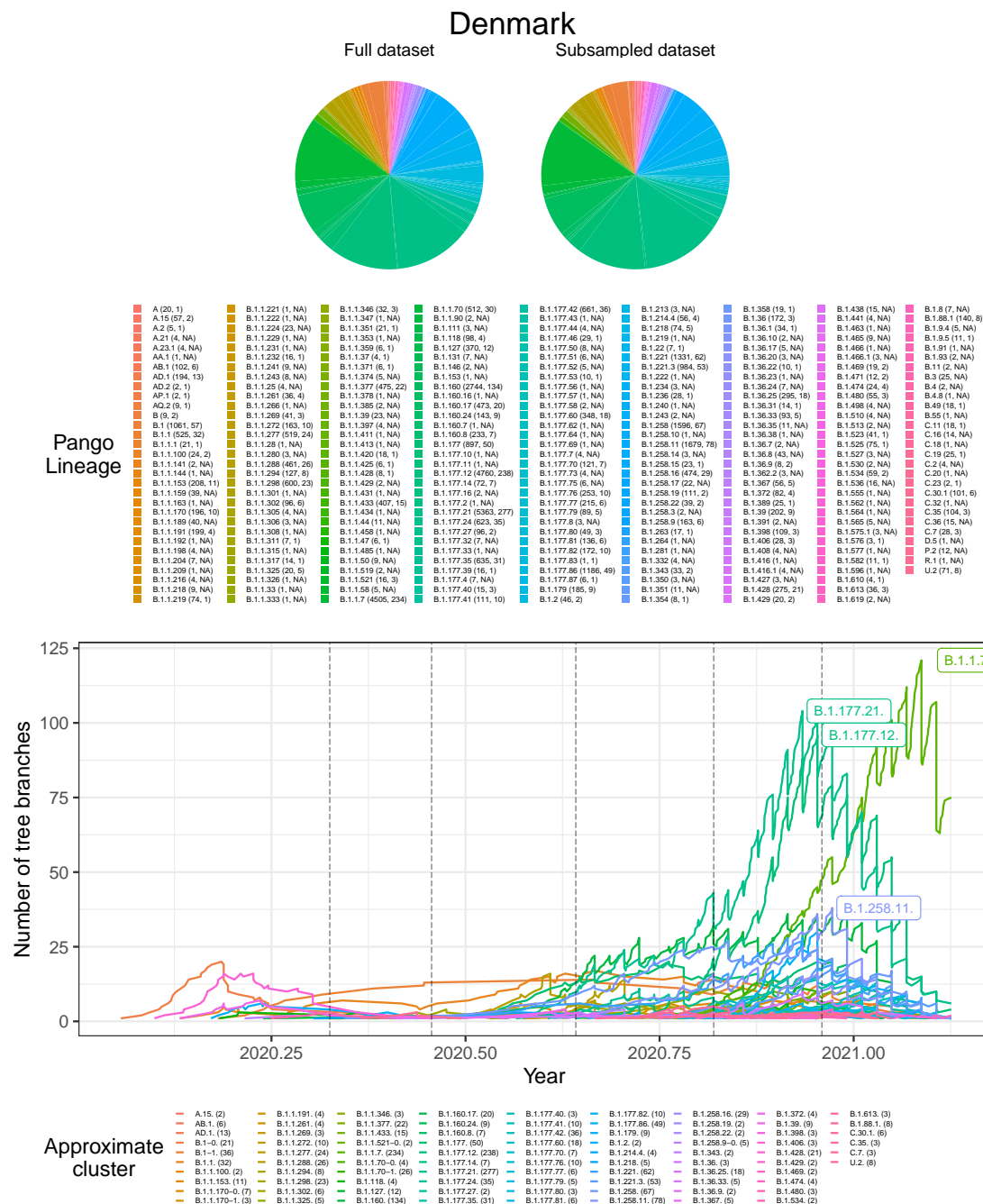


Figure SN.27: Denmark: Top: Pango lineages identified in the full (left) and subsampled (right) data set with the coloured area being proportional to the size of the specific lineage. Values in brackets next to the lineage name in the legend represent the number of sequences identified in the full (left) and subsampled (right) data set. Bottom: Lineage-through-time plot, counting the number of phylogenetic tree branches over time, of maximum clade credibility tree for all identified approximate clusters. Pango lineages that were split due to a longer period without any sample are denoted numerically using a dash separator, e.g. B.1-0 and B.1-1. Vertical dotted lines represent the dates at which birth, death and sampling rate were allowed to change. The number of sequences included for each approximate cluster in the analysis are given in brackets next to the cluster name in the legend.

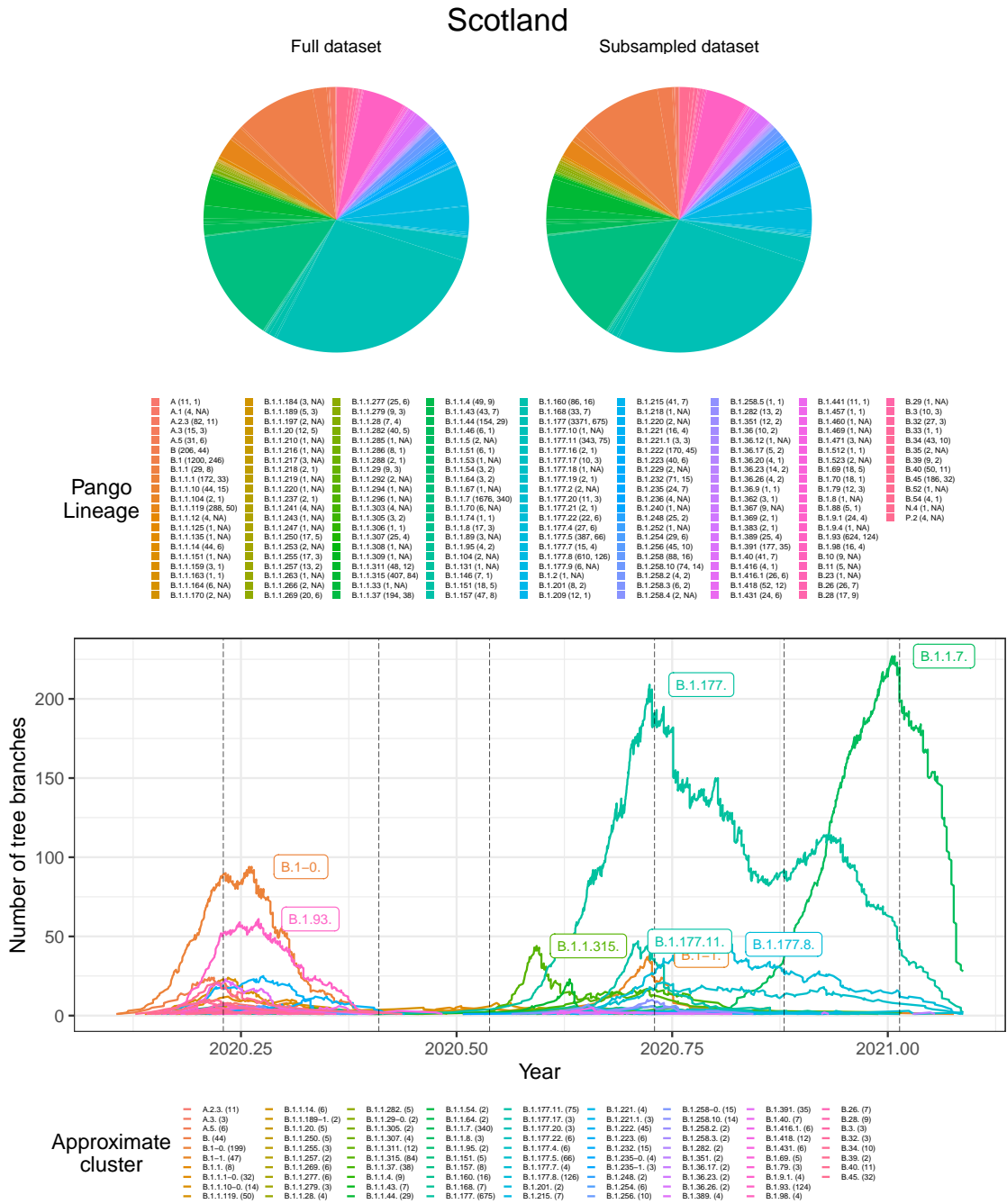


Figure SN.28: Scotland: Top: Pango lineages identified in the full (left) and subsampled (right) data set with the coloured area being proportional to the size of the specific lineage. Values in brackets next to the lineage name in the legend represent the number of sequences identified in the full (left) and subsampled (right) data set. Bottom: Lineage-through-time plot, counting the number of phylogenetic tree branches over time, of maximum clade credibility tree for all identified approximate clusters. Pango lineages that were split due to a longer period without any sample are denoted numerically using a dash separator, e.g. B.1-0 and B.1-1. Vertical dotted lines represent the dates at which birth, death and sampling rate were allowed to change. The number of sequences included for each approximate cluster in the analysis are given in brackets next to the cluster name in the legend.

Switzerland

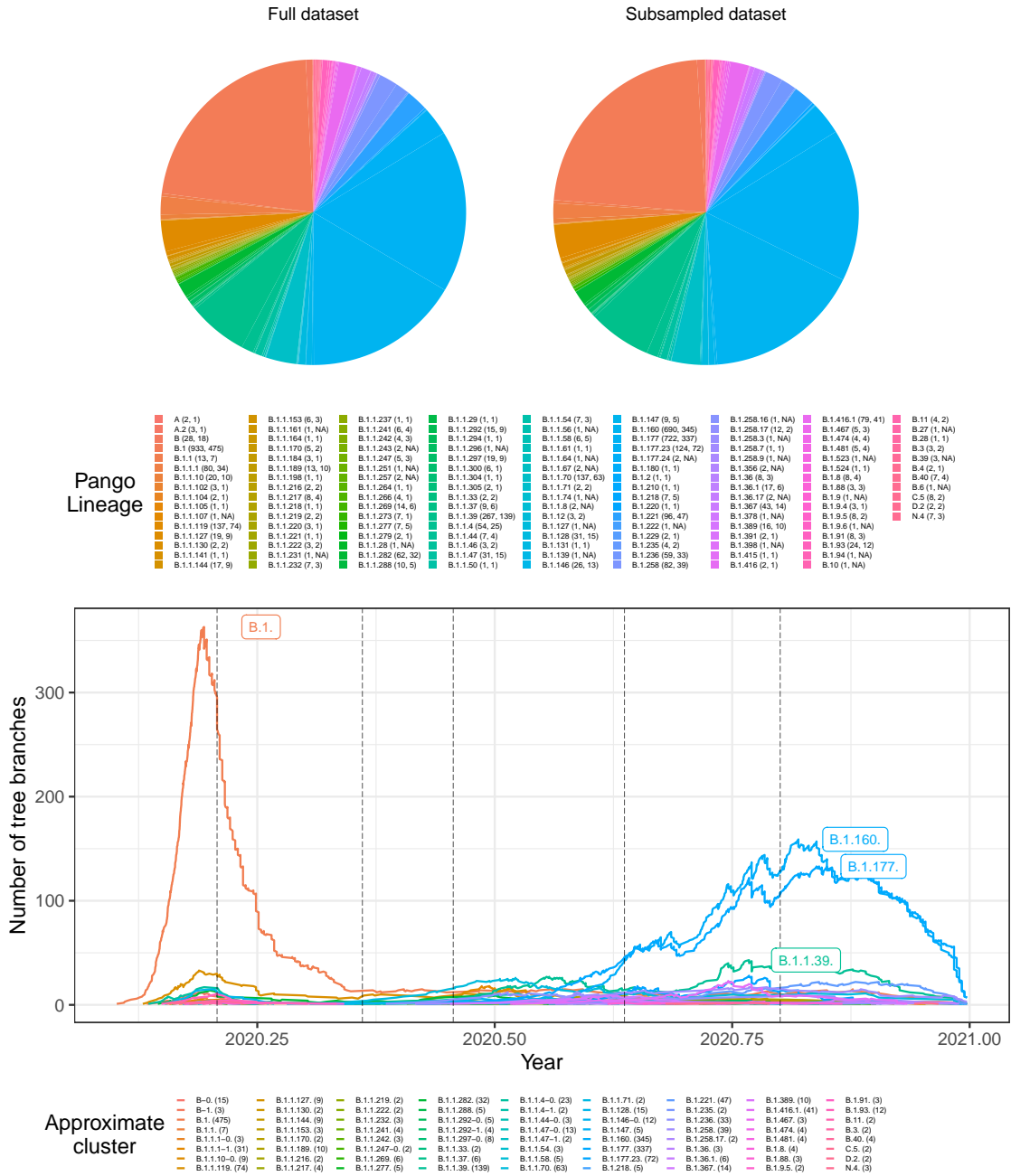


Figure SN.29: Switzerland: Top: Pango lineages identified in the full (left) and subsampled (right) data set with the coloured area being proportional to the size of the specific lineage. Values in brackets next to the lineage name in the legend represent the number of sequences identified in the full (left) and subsampled (right) data set. Bottom: Lineage-through-time plot, counting the number of phylogenetic tree branches over time, of maximum clade credibility tree for all identified approximate clusters. Pango lineages that were split due to a longer period without any sample are denoted numerically using a dash separator, e.g. B-0 and B-1. Vertical dotted lines represent the dates at which birth, death and sampling rate were allowed to change. The number of sequences included for each approximate cluster in the analysis are given in brackets next to the cluster name in the legend.

Victoria



Figure SN.30: Victoria, Australia: Top: Pango lineages identified in the full (left) and subsampled (right) data set with the coloured area being proportional to the size of the specific lineage. Values in brackets next to the lineage name in the legend represent the number of sequences identified in the full (left) and subsampled (right) data set. Due to the lineage-specific subsampling procedure, the relative lineage composition is different between the full and the subsampled dataset. Bottom: Lineage-through-time plot, counting the number of phylogenetic tree branches over time, of maximum clade credibility tree for all identified approximate clusters. Vertical dotted lines represent the dates at which birth, death and sampling rate were allowed to change. The number of sequences included for each approximate cluster in the analysis are given in brackets next to the cluster name in the legend.

Supplementary Note 3

This Supplementary Note contains all weblinks that point to the primary data depicted or discussed in the main manuscript. For example, the data related to (i) changes in non-pharmaceutical interventions, which are used in BEAST2; (ii) case reporting data which is depicted in Fig. 3 (main article); (iii) the number of tests performed and the proportion of positive tests, which are used to calculate the proportion of detected cases; and (iv) changes in testing strategies, which are depicted in Fig. 4 (main article).

Supplementary Links

Changes in non-pharmaceutical interventions

The data sources stated herein are used to determine change dates for phylodynamic reconstruction. The latter is used to compute R_e values that are depicted in Fig. 2 (main manuscript).

Denmark and Switzerland (Phylodynamics). WHO Coronavirus (COVID-19) Dashboard. Link: <https://covid19.who.int/table>, (accessed: 2021-03-10)

Victoria (Phylodynamics). Victorian coronavirus (COVID-19) data. Link: <https://www.dhhs.vic.gov.au/victorian-coronavirus-covid-19-data>, accessed: 2021-03-10)

Scotland (Phylodynamics). Coronavirus (COVID-19): trends in daily data. Link: <https://www.gov.scot/publications/coronavirus-covid-19-trends-in-daily-data>, (accessed: 2021-03-10)

Case reports

Reported SARS-CoV-2 infections are depicted in Fig. 3 (main manuscript) alongside the reconstructed incidence histories using GInPipe.

Denmark and Switzerland. Daily new reported cases. Link: <https://github.com/owid/covid-19-data/tree/master/public/data>, (accessed: 2021-04-27)

Scotland. New reported cases by date. Link: <https://coronavirus.data.gov.uk/details/download>, (accessed: 2021-04-27)

Victoria (AU). Daily new reported cases by state. Link: https://public.tableau.com/profile/vicdhhs#!/vizhome/Cases_15982342702770/DashCasesGSG, (accessed: 2021-04-27)

Number of tests performed and the test positivity

This data is used in Fig. 4 (main manuscript) to compute the case detection rate as outlined in the *Methods* section.

Denmark. A cross-country database of COVID-19 testing. Link: <https://covid.ourworldindata.org/data/owid-covid-data.csv>, (accessed: 2021-03-11)

Scotland. Coronavirus (COVID-19): trends in daily data. Link: <https://www.gov.scot/publications/coronavirus-covid-19-trends-in-daily-data>, (accessed: 2021-03-11)

Switzerland. A cross-country database of COVID-19 testing. Link: <https://covid.ourworldindata.org/data/owid-covid-data.csv>, (accessed: 2021-03-11)

Victoria. Testing for COVID-19. Link: <https://www.covid19data.com.au/testing>, (accessed: 2021-03-19)

Testing Strategy

Data on major changes in testing strategies are depicted in Fig. 4 as vertical dashed lines.

Denmark. Policy responses for Denmark. Link: <https://www.covid19healthsystem.org/countries/denmark/livinghit.aspx?Section=1.5%20Testing&Type=Section>, (accessed: 2021-03-11)

Scotland. Timeline of Coronavirus (COVID-19) in Scotland. Link: <https://spice-spotlight.scot/2021/03/12/timeline-of-coronavirus-covid-19-in-scotland>, (accessed: 2021-03-11)

Switzerland. Policy responses for Switzerland (Covid-19 Health System Response Monitor). Link: <https://www.covid19healthsystem.org/countries/switzerland/livinghit.aspx?Section=1.5%20Testing&Type=Section>, (accessed: 2021-03-18)

Victoria. Updates about the outbreak of the coronavirus disease (COVID-19). Link: <https://www.dhhs.vic.gov.au/coronavirus/updates>, (accessed: 2021-03-11)

Switzerland

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.

All Submitters of data may be contacted directly via www.gisaid.org

Authors are sorted alphabetically.

Accession ID	Originating Laboratory	Submitting Laboratory	Authors
EPI_ISL_1002357	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Catharine Aquino, Ivan Topolicky, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Andreia Cabral de Gouveia, Maria Domènica Moccia, Simon Gröler, Timothy Sykes, Lennart Opitz, Griffin White, Laura Neff, Doris Popovic, Andrea Pattignani, Jay Tracy, Ralph Schlapbach, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1004814, EPI_ISL_1004815, EPI_ISL_1004816, EPI_ISL_1004817, EPI_ISL_1004818, EPI_ISL_1004819, EPI_ISL_1004820, EPI_ISL_1004821, EPI_ISL_1004822, EPI_ISL_1004823, EPI_ISL_1004824, EPI_ISL_1004825, EPI_ISL_1004826, EPI_ISL_1004827, EPI_ISL_1004828, EPI_ISL_1004829, EPI_ISL_1004830, EPI_ISL_1004831, EPI_ISL_1004832, EPI_ISL_1004833, EPI_ISL_1004834, EPI_ISL_1004835, EPI_ISL_1004836, EPI_ISL_1004837, EPI_ISL_1004838, EPI_ISL_1004839, EPI_ISL_1004840, EPI_ISL_1004841, EPI_ISL_1004842, EPI_ISL_1004843, EPI_ISL_1004844, EPI_ISL_1004845, EPI_ISL_1004846, EPI_ISL_1004847, EPI_ISL_1004848, EPI_ISL_1004849, EPI_ISL_1004850, EPI_ISL_1004851, EPI_ISL_1004852, EPI_ISL_1004853, EPI_ISL_1004854, EPI_ISL_1004855, EPI_ISL_1004856, EPI_ISL_1004857, EPI_ISL_1004858, EPI_ISL_1004859, EPI_ISL_1004860, EPI_ISL_1004861, EPI_ISL_1004862, EPI_ISL_1004863, EPI_ISL_1004864, EPI_ISL_1004865, EPI_ISL_1004866, EPI_ISL_1004867, EPI_ISL_1004868, EPI_ISL_1004869, EPI_ISL_1004870, EPI_ISL_1004871, EPI_ISL_1004872, EPI_ISL_1004873, EPI_ISL_1004874, EPI_ISL_1004875, EPI_ISL_1004876, EPI_ISL_1004877, EPI_ISL_1004878, EPI_ISL_1004879, EPI_ISL_1004880, EPI_ISL_1004881, EPI_ISL_1004882, EPI_ISL_1004883, EPI_ISL_1004884, EPI_ISL_1004885, EPI_ISL_1004886, EPI_ISL_1004887, EPI_ISL_1004888, EPI_ISL_1004889, EPI_ISL_1004890, EPI_ISL_1004891, EPI_ISL_1004892, EPI_ISL_1004893, EPI_ISL_1004894, EPI_ISL_1004895, EPI_ISL_1004896, EPI_ISL_1004897, EPI_ISL_1004898, EPI_ISL_1004899, EPI_ISL_1004900, EPI_ISL_1004901, EPI_ISL_1004902, EPI_ISL_1004903, EPI_ISL_1004904, EPI_ISL_1004905, EPI_ISL_1004906, EPI_ISL_1004907, EPI_ISL_1004908, EPI_ISL_1004909, EPI_ISL_1004910, EPI_ISL_1004911, EPI_ISL_1004912, EPI_ISL_1004913, EPI_ISL_1004914, EPI_ISL_1004915, EPI_ISL_1004916, EPI_ISL_1004917, EPI_ISL_1004918, EPI_ISL_1004919, EPI_ISL_1004920, EPI_ISL_1004921, EPI_ISL_1004922, EPI_ISL_1004923, EPI_ISL_1004924, EPI_ISL_1004925, EPI_ISL_1004926, EPI_ISL_1004927, EPI_ISL_1004928, EPI_ISL_1004929, EPI_ISL_1004930, EPI_ISL_1004931, EPI_ISL_1004932, EPI_ISL_1004933, EPI_ISL_1004934, EPI_ISL_1004935, EPI_ISL_1004936, EPI_ISL_1004937, EPI_ISL_1004938, EPI_ISL_1004939, EPI_ISL_1004940, EPI_ISL_1004941, EPI_ISL_1004942, EPI_ISL_1004943, EPI_ISL_1004944, EPI_ISL_1004945, EPI_ISL_1004946, EPI_ISL_1004947, EPI_ISL_1004948, EPI_ISL_1004949, EPI_ISL_1004950, EPI_ISL_1004951, EPI_ISL_1004952, EPI_ISL_1004953, EPI_ISL_1004954, EPI_ISL_1004955, EPI_ISL_1004956, EPI_ISL_1004957, EPI_ISL_1004958, EPI_ISL_1004959, EPI_ISL_1004960, EPI_ISL_1004961, EPI_ISL_1004962, EPI_ISL_1004963, EPI_ISL_1004964, EPI_ISL_1004965, EPI_ISL_1004966, EPI_ISL_1004967, EPI_ISL_1004968, EPI_ISL_1004969, EPI_ISL_1004970, EPI_ISL_1004971, EPI_ISL_1004972, EPI_ISL_1004973, EPI_ISL_1004974, EPI_ISL_1004975	Vollter AG	Chaoran Chen, Sarah Nadeau, Ivan Topolicky, Emmanuel Demitziakis, Keith Harshman, Ioannis Xenarios, Henri Peggel, Lorenzo Cerutti, Deborah Penat, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler	
EPI_ISL_1005007, EPI_ISL_1005008, EPI_ISL_1005009, EPI_ISL_1005010, EPI_ISL_1005011, EPI_ISL_1005038, EPI_ISL_1005049, EPI_ISL_1005050, EPI_ISL_1005051, EPI_ISL_1005052, EPI_ISL_1005053, EPI_ISL_1005071, EPI_ISL_1005072, EPI_ISL_1005073, EPI_ISL_1005074, EPI_ISL_1005075, EPI_ISL_1005087, EPI_ISL_1005106, EPI_ISL_1005107, EPI_ISL_1005108, EPI_ISL_1005109, EPI_ISL_1005110, EPI_ISL_1005167, EPI_ISL_1005179, EPI_ISL_1005180, EPI_ISL_1005181, EPI_ISL_1005182, EPI_ISL_1005183, EPI_ISL_1005196, EPI_ISL_1005197, EPI_ISL_1005198, EPI_ISL_1005202, EPI_ISL_1005213, EPI_ISL_1005214	Vollter AG	Chaoran Chen, Sarah Nadeau, Catharine Aquino, Ivan Topolicky, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Andreia Cabral de Gouveia, Maria Domènica Moccia, Simon Gröler, Timothy Sykes, Lennart Opitz, Griffin White, Laura Neff, Doris Popovic, Andrea Pattignani, Jay Tracy, Ralph Schlapbach, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler	
EPI_ISL_1007614, EPI_ISL_1007615	CHUV	Laboratory of genomics and metagenomics, Institute of Microbiology, University Hospital Centre and University of Lausanne, Switzerland	Trestan Pillonel, Damien Jacot, Sébastien Aebly, Gilbert Greut, Claire Benelli
EPI_ISL_1036105, EPI_ISL_1036106, EPI_ISL_1036107, EPI_ISL_1036108, EPI_ISL_1036109, EPI_ISL_1036110, EPI_ISL_1036111, EPI_ISL_1036112, EPI_ISL_1036114, EPI_ISL_1036115, EPI_ISL_1036117, EPI_ISL_1036118, EPI_ISL_1036119, EPI_ISL_1036120, EPI_ISL_1036121, EPI_ISL_1036122, EPI_ISL_1036123, EPI_ISL_1036124, EPI_ISL_1036125, EPI_ISL_1036126, EPI_ISL_1036127, EPI_ISL_1036128, EPI_ISL_1036129, EPI_ISL_1036130, EPI_ISL_1036131, EPI_ISL_1036132, EPI_ISL_1036133, EPI_ISL_1036134, EPI_ISL_1036136, EPI_ISL_1036138, EPI_ISL_1036139	Labormedizinisches Zentrum Dr Risch	University Hospital Basel, Clinical Bacteriology	Tim Rolfert, Madlen Stange, Helena MB Seth-Smith, Alfredo Mari, Karoline Leuzinger, Julia Belkici, Nadia Wohlwend, Martin Risch, Lorenz Risch, Manuel Balthagy, Hans Hirsch, Adrian Egli
EPI_ISL_1119024, EPI_ISL_1119025, EPI_ISL_1119026, EPI_ISL_1119027, EPI_ISL_1119028, EPI_ISL_1119029	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Catharine Aquino, Ivan Topolicky, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Andreia Cabral de Gouveia, Maria Domènica Moccia, Simon Gröler, Timothy Sykes, Lennart Opitz, Griffin White, Laura Neff, Doris Popovic, Andrea Pattignani, Jay Tracy, Ralph Schlapbach, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1119041, EPI_ISL_1119042, EPI_ISL_1119043, EPI_ISL_1119044, EPI_ISL_1119045, EPI_ISL_1119046, EPI_ISL_1119047, EPI_ISL_1119048, EPI_ISL_1119049, EPI_ISL_1119050, EPI_ISL_1119052, EPI_ISL_1119022	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Ivan Topolicky, Emmanuel Demitziakis, Keith Harshman, Ioannis Xenarios, Henri Peggel, Lorenzo Cerutti, Deborah Penat, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1119234, EPI_ISL_1119235, EPI_ISL_1119237	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Catharine Aquino, Ivan Topolicky, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Andreia Cabral de Gouveia, Maria Domènica Moccia, Simon Gröler, Timothy Sykes, Lennart Opitz, Griffin White, Laura Neff, Doris Popovic, Andrea Pattignani, Jay Tracy, Ralph Schlapbach, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1119238	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Ivan Topolicky, Emmanuel Demitziakis, Keith Harshman, Ioannis Xenarios, Henri Peggel, Lorenzo Cerutti, Deborah Penat, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1119241	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Catharine Aquino, Ivan Topolicky, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Andreia Cabral de Gouveia, Maria Domènica Moccia, Simon Gröler, Timothy Sykes, Lennart Opitz, Griffin White, Laura Neff, Doris Popovic, Andrea Pattignani, Jay Tracy, Ralph Schlapbach, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1119244	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Ivan Topolicky, Emmanuel Demitziakis, Keith Harshman, Ioannis Xenarios, Henri Peggel, Lorenzo Cerutti, Deborah Penat, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1119340, EPI_ISL_1119341, EPI_ISL_1119342, EPI_ISL_1119343, EPI_ISL_1119346, EPI_ISL_1119348, EPI_ISL_1119350, EPI_ISL_1119351	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Catharine Aquino, Ivan Topolicky, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Andreia Cabral de Gouveia, Maria Domènica Moccia, Simon Gröler, Timothy Sykes, Lennart Opitz, Griffin White, Laura Neff, Doris Popovic, Andrea Pattignani, Jay Tracy, Ralph Schlapbach, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1119367, EPI_ISL_1119369	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Catharine Aquino, Ivan Topolicky, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Andreia Cabral de Gouveia, Maria Domènica Moccia, Simon Gröler, Timothy Sykes, Lennart Opitz, Griffin White, Laura Neff, Doris Popovic, Andrea Pattignani, Jay Tracy, Ralph Schlapbach, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler
EPI_ISL_1119370	Vollter AG	Department of Biosystems Science and Engineering, ETH Zürich	Chaoran Chen, Sarah Nadeau, Ivan Topolicky, Emmanuel Demitziakis, Keith Harshman, Ioannis Xenarios, Henri Peggel, Lorenzo Cerutti, Deborah Penat, Philipp Jablonksi, Lara Fuhrmann, David Drefuss, Katharina Jahn, Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen, Sophie Seidel, Noemie Santaromia de Souza, Nico Beerenwinkler, Targa Stadler

Japan

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.

All Submitters of data may be contacted directly via www.gisaid.org

Authors are sorted alphabetically.

Accession ID	Originating Laboratory	Submitting Laboratory	Authors
EPI_ISL_408662, EPI_ISL_408666, EPI_ISL_408667, EPI_ISL_408669, EPI_ISL_410531, EPI_ISL_410532	Dept. of Virology III, National Institute of Infectious Diseases	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Shutoku Matsuyama, Naganori Nao, Kazuya Shirato, Makoto Takada, Makoto Kuroda
EPI_ISL_412968, EPI_ISL_412969	Dept. of Pathology, National Institute of Infectious Diseases	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Harutaka Katano, Shutoku Matsuyama, Naganori Nao, Kazuya Shirato, Moto Suzuki, Hideo Hasegawa, Takayuki Wakita, Makoto Takada, Tadaki Suzuki, Makoto Kuroda
EPI_ISL_413459	Takayuki Hishiki Kanagawa Prefectural Institute of Public Health	Takayuki Hishiki Kanagawa Prefectural Institute of Public Health	Hishiki T., Suzuki R., Sakurai J., Usui K., Tanaka Y., Kawajiri J., Kogo Y., Matsuki Y., An T., Hayashizaki Y. and Takasaki T.
EPI_ISL_414511	Department of Pathology, Toshiro Hospital	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Kentaro Itokawa, Takuya Adachi, Masahiro Sano, Jun Yamazaki, Ipppei Miyamoto, Hanuka Nishio, Ja-Mun Chong, Noriko Nakajima, Yuko Sato, Minoru Tobume, Harutaka Katano, Tadaki Suzuki, Makoto Kuroda
EPI_ISL_414524	Department of Microbiology, Ryuta Kumagai Tokyo Metropolitan Institute of Public Health	Tokyo Metropolitan Institute of Public Health	Kumagai R., Yoshida J., Nagashima M., Chiba T. and Sadamasu K.
EPI_ISL_416025	Saitama Medical University Hospital	Saitama Medical University	Kazuo Inai
EPI_ISL_416565, EPI_ISL_416566, EPI_ISL_416567, EPI_ISL_416569, EPI_ISL_416570, EPI_ISL_416571, EPI_ISL_416572, EPI_ISL_416573, EPI_ISL_416574, EPI_ISL_416575, EPI_ISL_416576, EPI_ISL_416577, EPI_ISL_416578, EPI_ISL_416579, EPI_ISL_416580, EPI_ISL_416581, EPI_ISL_416582, EPI_ISL_416583, EPI_ISL_416584, EPI_ISL_416585, EPI_ISL_416586, EPI_ISL_416587, EPI_ISL_416588, EPI_ISL_416589, EPI_ISL_416590, EPI_ISL_416591, EPI_ISL_416592, EPI_ISL_416593, EPI_ISL_416594, EPI_ISL_416595, EPI_ISL_416596, EPI_ISL_416597, EPI_ISL_416598, EPI_ISL_416599, EPI_ISL_416600, EPI_ISL_416601, EPI_ISL_416602, EPI_ISL_416603, EPI_ISL_416604, EPI_ISL_416605, EPI_ISL_416606, EPI_ISL_416607, EPI_ISL_416608, EPI_ISL_416609, EPI_ISL_416610, EPI_ISL_416611, EPI_ISL_416612, EPI_ISL_416613, EPI_ISL_416614, EPI_ISL_416615, EPI_ISL_416616, EPI_ISL_416617, EPI_ISL_416618, EPI_ISL_416619, EPI_ISL_416620, EPI_ISL_416621, EPI_ISL_416622, EPI_ISL_416624, EPI_ISL_416625, EPI_ISL_416626, EPI_ISL_416627, EPI_ISL_416628, EPI_ISL_416629, EPI_ISL_416630, EPI_ISL_416631, EPI_ISL_416633, EPI_ISL_416634	Saitama Medical University	Tsuyoshi Sekizuka, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Tsutomu Kageyama, Shiroi Saito, Bujo Takayama, Hideo Hasegawa, Takuro Takahashi, Hajime Kamiya, Takuya Yamagishi, Moto Suzuki, Takayuki Wakita, Masato Kuroda, Katarina Braun, Gage Moreno, Peter Hoffmann, et al.	
see above	Japanese Quarantine Stations	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Tsutomu Kageyama, Shiroi Saito, Bujo Takayama, Hideo Hasegawa, Takuro Takahashi, Hajime Kamiya, Takuya Yamagishi, Moto Suzuki, Takayuki Wakita, Masato Kuroda
EPI_ISL_418800	University of Wisconsin - Madison: Influenza Research Institute	University of Wisconsin - Madison: AIDS Vaccine Research Laboratories	Katarina Braun, Gage Moreno, Peter Hoffmann, et al.
EPI_ISL_419296	Kochi Prefectural Institute of Public Health	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Akhiko Tokaji, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Hajime Kamiya, Moto Suzuki, Makoto Kuroda
EPI_ISL_419297, EPI_ISL_419298	Chiba Prefectural Institute of Public Health	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Masakatsu Taira, Yushi Hachisu, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Hajime Kamiya, Moto Suzuki, Makoto Kuroda
EPI_ISL_419299, EPI_ISL_419300	Ishikawa Prefectural Institute of Public Health and Environmental Science	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Sanae Kuramoto, Eri Nariai, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Hajime Kamiya, Moto Suzuki, Makoto Kuroda
EPI_ISL_419301, EPI_ISL_419302, EPI_ISL_419303, EPI_ISL_419304, EPI_ISL_419305, EPI_ISL_419306, EPI_ISL_419307, EPI_ISL_419308	Saitama Prefectural Institute of Public Health	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Michio Shirohara, Tsuyoshi Kishimoto, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Hajime Kamiya, Moto Suzuki, Makoto Kuroda
EPI_ISL_419309, EPI_ISL_419310, EPI_ISL_419311	Chiba Prefectural Institute of Public Health	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Masakatsu Taira, Yushi Hachisu, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Hajime Kamiya, Moto Suzuki, Makoto Kuroda
EPI_ISL_420889, EPI_ISL_420890	Takayuki Hishiki Kanagawa Prefectural Institute of Public Health	Takayuki Hishiki Kanagawa Prefectural Institute of Public Health	Hishiki T., Suzuki R., Sakurai J., Usui K., Tanaka Y., Kawajiri J., Kogo Y., Matsuki Y., An T., Hayashizaki Y. and Takasaki T.
EPI_ISL_434561, EPI_ISL_434562	Department of Microbiology, Ryuta Kumagai Tokyo Metropolitan Institute of Public Health	Department of Microbiology, Ryuta Kumagai Tokyo Metropolitan Institute of Public Health	Kumagai R., Yoshida J., Asakura H., Nagashima M., Chiba T. and Sadamasu K.
EPI_ISL_437601	Keio University School of Medicine	Keio University School of Medicine	Kenjiro Kosaki, Yuka Iwasaki, Toshiaki Takenouchi, Hanahiko Sioni,
EPI_ISL_438947, EPI_ISL_438948, EPI_ISL_438949, EPI_ISL_438951, EPI_ISL_438952, EPI_ISL_438953, EPI_ISL_438954, EPI_ISL_438955, EPI_ISL_438956, EPI_ISL_438958, EPI_ISL_438959, EPI_ISL_438961, EPI_ISL_438965, EPI_ISL_438966, EPI_ISL_438967, EPI_ISL_438969, EPI_ISL_438971, EPI_ISL_438972	Keio University School of Medicine	Keio University School of Medicine	Kenjiro Kosaki
see above	Takayuki Hishiki Kanagawa Prefectural Institute of Public Health	Takayuki Hishiki Kanagawa Prefectural Institute of Public Health	Hishiki T., Suzuki R., Sakurai J., Usui K., Tanaka Y., Kawajiri J., Kogo Y., Matsuki Y., An T., Hayashizaki Y. and Takasaki T.
EPI_ISL_454749	Japanese Quarantine Stations	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Tsutomu Kageyama, Shiroi Saito, Bujo Takayama, Hideo Hasegawa, Takuro Takahashi, Hajime Kamiya, Takuya Yamagishi, Moto Suzuki, Takayuki Wakita, Makoto Kuroda
EPI_ISL_462439, EPI_ISL_462440, EPI_ISL_462441, EPI_ISL_462442, EPI_ISL_462443, EPI_ISL_462444, EPI_ISL_462445	unknown	Ryuta Kumagai Tokyo Metropolitan Institute of Public Health	Asakura H., Kumagai R., Yoshida J., Nagashima M., Chiba T., Sadamasu K.
EPI_ISL_462446	unknown	Ryuta Kumagai Tokyo Metropolitan Institute of Public Health	Kumagai R., Yoshida J., Asakura H., Nagashima M., Chiba T., Sadamasu K.
EPI_ISL_468724, EPI_ISL_468725	unknown	Contact: Ryuta Kumagai Tokyo Metropolitan Institute of Public Health	Kumagai R., Yoshida J., Asakura H., Nagashima M., Chiba T., Sadamasu K.
EPI_ISL_469287, EPI_ISL_469288, EPI_ISL_469289, EPI_ISL_469290, EPI_ISL_469291, EPI_ISL_469292, EPI_ISL_469293, EPI_ISL_469294, EPI_ISL_469295, EPI_ISL_469296	Keio University Hospital	Keio University Hospital	Kenjiro Kosaki
EPI_ISL_479676, EPI_ISL_479677, EPI_ISL_479678, EPI_ISL_479679, EPI_ISL_479680, EPI_ISL_479681, EPI_ISL_479682, EPI_ISL_479683, EPI_ISL_479684, EPI_ISL_479685, EPI_ISL_479686, EPI_ISL_479687, EPI_ISL_479688, EPI_ISL_479689, EPI_ISL_479690, EPI_ISL_479691	unknown	Contact: Hiroyuki Asakura Tokyo Metropolitan Institute of Public Health, Department of Microbiology	Asakura H., Yoshida J., Kumagai R., Nagashima M., Chiba T., Sadamasu K.
EPI_ISL_479702, EPI_ISL_479703	Heikokado Institute of Public Health	Pathogen Genomics Center, National Institute of Infectious Diseases	Tsuyoshi Sekizuka, Rika Komagome, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Hajime Kamiya, Moto Suzuki, Makoto Kuroda

Brazil

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.

All Submitters of data may be contacted directly via www.gisaid.org

Authors are sorted alphabetically.

Accession ID	Originating Laboratory	Submitting Laboratory	Authors
EPI_ISL_100670, EPI_ISL_100671, EPI_ISL_100672, EPI_ISL_100677, EPI_ISL_100677	Instituto de Biociencia - UNESP-Botucatu-SP	Instituto de Biociencia - UNESP-Botucatu-SP	Lella Sabrina Ullmann; Fábio Sossai Possobon, Camila Dantas Matossi, Paula Rahal, Paulo Inacio da Costa, João Pessoa Araújo Jr.
EPI_ISL_1023841	Centro de Desenvolvimento Tecnológico em Saúde - CDTS	Centro de Desenvolvimento Tecnológico em Saúde - CDTS	Souza T.M., Frittelman-Rodrigues N., De Paula A.D., Saraiva F.B., Ferreira M.A. and Sacramento C.O.
EPI_ISL_1034304, EPI_ISL_1034305, EPI_ISL_1034306	Laboratório de Ecologia de Doenças Transmissíveis na Amazônia, Instituto Leonidas e Maria Deane - Foz de Iguaçu, Amazônia	Laboratório de Ecologia de Doenças Transmissíveis na Amazônia, Instituto Leonidas e Maria Deane - Foz de Iguaçu, Amazônia	Valdineia Nascimento, Victor Souza, André Casado, Fernando Nascimento, George Silva, Ângela Costa, Débora Duarte, Karina Pozzato, Matheus Magalhães, Luciana Gonçalves, Maria Júlia Brandão, Michele Jesus, Felipe Naveson on behalf of the Foz de Iguaçu COVID-19 Genomic Surveillance Network
EPI_ISL_1039691, EPI_ISL_1039692, EPI_ISL_1039693, EPI_ISL_1039694, EPI_ISL_1039695	LACEN do Estado de Goiás	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039696	Instituto Adolfo Lutz - Regional de Presidente Prudente	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039697	Instituto Adolfo Lutz Central	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039698	Lab Loc - Rapexicon da Serra	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039699	Instituto Adolfo Lutz - Regional de Taubaté	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039700, EPI_ISL_1039701	Instituto Adolfo Lutz Central	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039702	Instituto Adolfo Lutz - Regional de Aracatuba	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039703	Instituto Adolfo Lutz - Regional de Taubaté	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039704	Lab Loc - Rapexicon da Serra	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1039705, EPI_ISL_1039706, EPI_ISL_1039707, EPI_ISL_1039708, EPI_ISL_1039709, EPI_ISL_1039710	Instituto Adolfo Lutz Central	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1040623	Secretaria Municipal de Saúde de Piracicaba	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1040624, EPI_ISL_1040625, EPI_ISL_1040626, EPI_ISL_1040627, EPI_ISL_1040628, EPI_ISL_1040629, EPI_ISL_1040630, EPI_ISL_1040631, EPI_ISL_1040632, EPI_ISL_1040633, EPI_ISL_1040634, EPI_ISL_1040635, EPI_ISL_1040636, EPI_ISL_1040637, EPI_ISL_1040638, EPI_ISL_1040639, EPI_ISL_1040640, EPI_ISL_1040641, EPI_ISL_1040642, EPI_ISL_1040643, EPI_ISL_1040644, EPI_ISL_1040645, EPI_ISL_1040646, EPI_ISL_1040647, EPI_ISL_1040648, EPI_ISL_1040649, EPI_ISL_1040650, EPI_ISL_1040651	LACEN do Mato Grosso do Sul	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
see above	LACEN do Estado de Goiás	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1041509	LACEN do Estado de Goiás	Instituto Adolfo Lutz, Interdisciplinary Procedures Center, Strategic Laboratory	Claudio Tavares Sacchi, Claudia Regina Gonçalves, Erica Valeska Ramos Gomes, Karoline Rodrigues Campos
EPI_ISL_1061031, EPI_ISL_1061032	CDL Laboratório Santos e Vidal LTDA	Instituto de Medicina Tropical de São Paulo	Brazil-UK Centre for Arbovirus Discovery, Diagnosis, Genomics and Epidemiology (CADDE) Genomic Network - Instituto de Medicina Tropical
EPI_ISL_1063789	Evandro Chagas Instituto	Evandro Chagas Institute Virology	Santos, M.C.; Silva, A.M.; Junior, W.D.C.; Barbagelata, L.S.; Ferreira, J.A.; Sousa, E.M.A.; da Silva, P.S.; Pinheiro, K.C.L.C.; Sousa Junior, E.C.
EPI_ISL_1067728	Center for Biotechnology and Cell Therapy, São Rafael Hospital, Salvador, Brazil	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Stephane Tosta, Luciana Oliveira, Vanessa Nardy, Patricia Cajado, Marcela Gómez, Breno Dominguez, Jaqueline Gomes, Vagner Fonseca, Marta Giovannetti, Luz Alcántara, Felicidade Pereira, Arabela Leal
EPI_ISL_1067729, EPI_ISL_1067730, EPI_ISL_1067731	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Stephane Tosta, Luciana Oliveira, Vanessa Nardy, Patricia Cajado, Marcela Gómez, Breno Dominguez, Jaqueline Gomes, Vagner Fonseca, Marta Giovannetti, Luz Alcántara, Felicidade Pereira, Arabela Leal
EPI_ISL_1067732	Center for Biotechnology and Cell Therapy, São Rafael Hospital, Salvador, Brazil	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Stephane Tosta, Luciana Oliveira, Vanessa Nardy, Patricia Cajado, Marcela Gómez, Breno Dominguez, Jaqueline Gomes, Vagner Fonseca, Marta Giovannetti, Luz Alcántara, Felicidade Pereira, Arabela Leal
EPI_ISL_1067733, EPI_ISL_1067734, EPI_ISL_1067735	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Stephane Tosta, Luciana Oliveira, Vanessa Nardy, Patricia Cajado, Marcela Gómez, Breno Dominguez, Jaqueline Gomes, Vagner Fonseca, Marta Giovannetti, Luz Alcántara, Felicidade Pereira, Arabela Leal
EPI_ISL_1067736	Center for Biotechnology and Cell Therapy, São Rafael Hospital, Salvador, Brazil	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Stephane Tosta, Luciana Oliveira, Vanessa Nardy, Patricia Cajado, Marcela Gómez, Breno Dominguez, Jaqueline Gomes, Vagner Fonseca, Marta Giovannetti, Luz Alcántara, Felicidade Pereira, Arabela Leal
EPI_ISL_1067737, EPI_ISL_1067738	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Stephane Tosta, Luciana Oliveira, Vanessa Nardy, Patricia Cajado, Marcela Gómez, Breno Dominguez, Jaqueline Gomes, Vagner Fonseca, Marta Giovannetti, Luz Alcántara, Felicidade Pereira, Arabela Leal
EPI_ISL_1068076, EPI_ISL_1068079, EPI_ISL_1068080, EPI_ISL_1068081, EPI_ISL_1068082, EPI_ISL_1068083, EPI_ISL_1068084, EPI_ISL_1068085, EPI_ISL_1068086, EPI_ISL_1068087, EPI_ISL_1068088, EPI_ISL_1068089, EPI_ISL_1068090, EPI_ISL_1068091, EPI_ISL_1068092, EPI_ISL_1068093, EPI_ISL_1068094, EPI_ISL_1068095, EPI_ISL_1068096, EPI_ISL_1068097, EPI_ISL_1068098, EPI_ISL_1068099, EPI_ISL_1068100, EPI_ISL_1068101, EPI_ISL_1068102, EPI_ISL_1068103, EPI_ISL_1068104, EPI_ISL_1068105, EPI_ISL_1068106, EPI_ISL_1068107, EPI_ISL_1068108, EPI_ISL_1068109, EPI_ISL_1068110, EPI_ISL_1068111, EPI_ISL_1068112, EPI_ISL_1068113, EPI_ISL_1068114, EPI_ISL_1068115, EPI_ISL_1068116, EPI_ISL_1068117, EPI_ISL_1068118, EPI_ISL_1068119, EPI_ISL_1068120, EPI_ISL_1068121, EPI_ISL_1068122, EPI_ISL_1068123, EPI_ISL_1068124, EPI_ISL_1068125, EPI_ISL_1068126, EPI_ISL_1068127, EPI_ISL_1068128, EPI_ISL_1068129, EPI_ISL_1068130, EPI_ISL_1068131, EPI_ISL_1068132, EPI_ISL_1068133, EPI_ISL_1068134, EPI_ISL_1068135, EPI_ISL_1068136, EPI_ISL_1068137, EPI_ISL_1068138, EPI_ISL_1068139, EPI_ISL_1068140, EPI_ISL_1068141, EPI_ISL_1068142, EPI_ISL_1068143, EPI_ISL_1068144, EPI_ISL_1068145, EPI_ISL_1068146, EPI_ISL_1068147, EPI_ISL_1068148, EPI_ISL_1068149, EPI_ISL_1068150, EPI_ISL_1068151, EPI_ISL_1068152, EPI_ISL_1068153, EPI_ISL_1068154, EPI_ISL_1068155, EPI_ISL_1068156, EPI_ISL_1068157, EPI_ISL_1068158, EPI_ISL_1068159, EPI_ISL_1068160, EPI_ISL_1068161, EPI_ISL_1068162, EPI_ISL_1068163, EPI_ISL_1068164, EPI_ISL_1068165, EPI_ISL_1068166, EPI_ISL_1068167, EPI_ISL_1068168, EPI_ISL_1068169, EPI_ISL_1068170, EPI_ISL_1068171, EPI_ISL_1068172, EPI_ISL_1068173, EPI_ISL_1068174, EPI_ISL_1068175, EPI_ISL_1068176, EPI_ISL_1068177, EPI_ISL_1068178, EPI_ISL_1068179, EPI_ISL_1068180, EPI_ISL_1068181, EPI_ISL_1068182, EPI_ISL_1068183, EPI_ISL_1068184, EPI_ISL_1068185, EPI_ISL_1068186, EPI_ISL_1068187, EPI_ISL_1068188, EPI_ISL_1068189, EPI_ISL_1068190, EPI_ISL_1068191, EPI_ISL_1068192, EPI_ISL_1068193, EPI_ISL_1068194, EPI_ISL_1068195, EPI_ISL_1068196, EPI_ISL_1068197, EPI_ISL_1068198, EPI_ISL_1068199, EPI_ISL_1068200, EPI_ISL_1068201, EPI_ISL_1068202, EPI_ISL_1068203, EPI_ISL_1068204, EPI_ISL_1068205, EPI_ISL_1068206, EPI_ISL_1068207, EPI_ISL_1068208, EPI_ISL_1068209, EPI_ISL_1068210, EPI_ISL_1068211, EPI_ISL_1068212, EPI_ISL_1068213, EPI_ISL_1068214, EPI_ISL_1068215, EPI_ISL_1068216, EPI_ISL_1068217, EPI_ISL_1068218, EPI_ISL_1068219, EPI_ISL_1068220, EPI_ISL_1068221, EPI_ISL_1068222, EPI_ISL_1068223, EPI_ISL_1068224, EPI_ISL_1068225, EPI_ISL_1068226, EPI_ISL_1068227, EPI_ISL_1068228, EPI_ISL_1068229, EPI_ISL_1068230, EPI_ISL_1068231, EPI_ISL_1068232, EPI_ISL_1068233, EPI_ISL_1068234, EPI_ISL_1068235, EPI_ISL_1068236, EPI_ISL_1068237, EPI_ISL_1068238, EPI_ISL_1068239, EPI_ISL_1068240, EPI_ISL_1068241, EPI_ISL_1068242, EPI_ISL_1068243, EPI_ISL_1068244, EPI_ISL_1068245, EPI_ISL_1068246, EPI_ISL_1068247, EPI_ISL_1068248, EPI_ISL_1068249, EPI_ISL_1068250, EPI_ISL_1068251, EPI_ISL_1068252	Central Public Health Laboratory - LACEN - Bahia, Salvador, Brazil	Stephane Tosta, Luciana Oliveira, Vanessa Nardy, Patricia Cajado, Marcela Gómez, Breno Dominguez, Jaqueline Gomes, Vagner Fonseca, Marta Giovannetti, Luz Alcántara, Felicidade Pereira, Arabela Leal	

India

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.

All Submitters of data may be contacted directly via www.gisaid.org

Authors are sorted alphabetically.

Accession ID	Originating Laboratory	Submitting Laboratory	Authors
EPI_ISL_1034165, EPI_ISL_1034167, EPI_ISL_1034168, EPI_ISL_1034169, EPI_ISL_1034170, EPI_ISL_1034171, EPI_ISL_1034172, EPI_ISL_1034173, EPI_ISL_1034174, EPI_ISL_1034175, EPI_ISL_1034176, EPI_ISL_1034177, EPI_ISL_1034178, EPI_ISL_1034179, EPI_ISL_1034180, EPI_ISL_1034181, EPI_ISL_1034182, EPI_ISL_1034183, EPI_ISL_1034184, EPI_ISL_1034185, EPI_ISL_1034186, EPI_ISL_1034188, EPI_ISL_1034189	see above	see above	see above
EPI_ISL_1039249, EPI_ISL_1055383	NV Influenza	Indian Council of Medical Research-National Institute of Virology, Microbial Containment Complex	Potter V
EPI_ISL_1055384, EPI_ISL_1055796, EPI_ISL_1055770, EPI_ISL_1055771, EPI_ISL_1055786, EPI_ISL_1055788, EPI_ISL_1055789	Virus Research and Diagnostic Laboratory, Kasturba Hospital for Infectious Diseases	National Institute of Virology-Microbial Containment Complex, Indian Council of Medical Research	Pragya D. Yadav, Jayanti Shastri
EPI_ISL_1061417, EPI_ISL_1061420, EPI_ISL_1061421, EPI_ISL_1061431, EPI_ISL_1061434	Indian Council of Medical Research-National Institute of Virology, Microbial Containment Complex	Indian Council of Medical Research-National Institute of Virology, Microbial Containment Complex	Pragya D. Yadav
EPI_ISL_1063400, EPI_ISL_1063401	Virus Research and Diagnostic Laboratory, Medical College Level, Kasturba Hospital for Infectious Diseases	National Institute of Virology-Microbial Containment Complex, Indian Council of Medical Research	Pragya D. Yadav, Jayanti Shastri
EPI_ISL_1063403	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Rupinder Bakshi
EPI_ISL_1063494, EPI_ISL_1063526, EPI_ISL_1063527	Virus Research and Diagnostic Laboratory, Medical College Level, Kasturba Hospital for Infectious Diseases	National Institute of Virology-Microbial Containment Complex, Indian Council of Medical Research	Pragya D. Yadav, Jayanti Shastri
EPI_ISL_1063528	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Rupinder Bakshi
EPI_ISL_1063529	Virus Research and Diagnostic Laboratory, Medical College Level, Kasturba Hospital for Infectious Diseases	National Institute of Virology-Microbial Containment Complex, Indian Council of Medical Research	Pragya D. Yadav, Jayanti Shastri
EPI_ISL_1063530	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Rupinder Bakshi
EPI_ISL_1063531	Virus Research and Diagnostic Laboratory, Medical College Level, Kasturba Hospital for Infectious Diseases	National Institute of Virology-Microbial Containment Complex, Indian Council of Medical Research	Pragya D. Yadav, Jayanti Shastri
EPI_ISL_1063532, EPI_ISL_1063534, EPI_ISL_1063535, EPI_ISL_1063536, EPI_ISL_1063537	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Rupinder Bakshi
EPI_ISL_1063774, EPI_ISL_1063775	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Meil Singh
EPI_ISL_1063776	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Rupinder Bakshi
EPI_ISL_1063777	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Meil Singh
EPI_ISL_1063778, EPI_ISL_1063779	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Rupinder Bakshi
EPI_ISL_1063780, EPI_ISL_1063781, EPI_ISL_1063782, EPI_ISL_1063783	PGIMER, Chandigarh	ICMR-NATIONAL INSTITUTE OF VIROLOGY, MICROBIAL CONTAINMENT COMPLEX	Pragya D. Yadav, Meil Singh
EPI_ISL_1063806	Molecular Laboratory, Vikash Multiplicity Hospital, Bargah	Immunogenetics lab, Institute of Life Sciences, Bhubaneswar	Sunil K. Raghav, Safal Walla, Anup Ghosh, Atmuka Jha, Anil M. Kanampalliar, Shru Aggarwal, Kirtal Hansdah, Rupesh Dash, Rajesh Swain, Punit Prasad, INSACOG Consortium, Ajay Parida
EPI_ISL_1063807	REGIONAL VRDL, ICMR-RMRC BBSR	Immunogenetics lab, Institute of Life Sciences, Bhubaneswar	Sunil K. Raghav, Safal Walla, Anup Ghosh, Atmuka Jha, Anil M. Kanampalliar, Shru Aggarwal, Kirtal Hansdah, Rupesh Dash, Rajesh Swain, Punit Prasad, INSACOG Consortium, Ajay Parida
EPI_ISL_1063808	Immunogenetics lab, Institute of Life Sciences, Bhubaneswar	Immunogenetics lab, Institute of Life Sciences, Bhubaneswar	Sunil K. Raghav, Safal Walla, Anup Ghosh, Atmuka Jha, Anil M. Kanampalliar, Shru Aggarwal, Kirtal Hansdah, Rupesh Dash, Rajesh Swain, Punit Prasad, INSACOG Consortium, Ajay Parida
EPI_ISL_1063809	REGIONAL VRDL, ICMR-RMRC BBSR	Immunogenetics lab, Institute of Life Sciences, Bhubaneswar	Sunil K. Raghav, Safal Walla, Anup Ghosh, Atmuka Jha, Anil M. Kanampalliar, Shru Aggarwal, Kirtal Hansdah, Rupesh Dash, Rajesh Swain, Punit Prasad, INSACOG Consortium, Ajay Parida
EPI_ISL_1063810, EPI_ISL_1063811, EPI_ISL_1063812, EPI_ISL_1063813, EPI_ISL_1063814, EPI_ISL_1063815, EPI_ISL_1063816, EPI_ISL_1063817, EPI_ISL_1063818, EPI_ISL_1063819, EPI_ISL_1063820, EPI_ISL_1063821, EPI_ISL_1063822, EPI_ISL_1063823, EPI_ISL_1063824, EPI_ISL_1063825, EPI_ISL_1063826, EPI_ISL_1063827, EPI_ISL_1063828, EPI_ISL_1063829, EPI_ISL_1063830, EPI_ISL_1063831, EPI_ISL_1063832, EPI_ISL_1063833, EPI_ISL_1063834, EPI_ISL_1063835, EPI_ISL_1063836, EPI_ISL_1063837, EPI_ISL_1063838, EPI_ISL_1063839	see above	see above	see above
EPI_ISL_1063840, EPI_ISL_1063841, EPI_ISL_1063842, EPI_ISL_1063843, EPI_ISL_1063844, EPI_ISL_1063845, EPI_ISL_1063846, EPI_ISL_1063847, EPI_ISL_1063848, EPI_ISL_1063849, EPI_ISL_1063850, EPI_ISL_1063851	see above	see above	see above
EPI_ISL_1063852, EPI_ISL_1063853, EPI_ISL_1063854	INOVA Life Sciences Private Limited, Bhubaneswar	Immunogenetics lab, Institute of Life Sciences, Bhubaneswar	Sunil K. Raghav, Safal Walla, Anup Ghosh, Atmuka Jha, Anil M. Kanampalliar, Shru Aggarwal, Kirtal Hansdah, Rupesh Dash, Rajesh Swain, Punit Prasad, INSACOG Consortium, Ajay Parida
EPI_ISL_1063855, EPI_ISL_1063856, EPI_ISL_1063857, EPI_ISL_1063858, EPI_ISL_1063859, EPI_ISL_1063860, EPI_ISL_1063861	Molecular Laboratory, Vikash Multiplicity Hospital, Bargah	Immunogenetics lab, Institute of Life Sciences, Bhubaneswar	Sunil K. Raghav, Safal Walla, Anup Ghosh, Atmuka Jha, Anil M. Kanampalliar, Shru Aggarwal, Kirtal Hansdah, Rupesh Dash, Rajesh Swain, Punit Prasad, INSACOG Consortium, Ajay Parida
EPI_ISL_1063862, EPI_ISL_1063863, EPI_ISL_1063864, EPI_ISL_1063865, EPI_ISL_1063866, EPI_ISL_1063867, EPI_ISL_1063868, EPI_ISL_1063869, EPI_ISL_1063870, EPI_ISL_1063871, EPI_ISL_1063872, EPI_ISL_1063873, EPI_ISL_1063874, EPI_ISL_1063875, EPI_ISL_1063876, EPI_ISL_1063877, EPI_ISL_1063878			

South Africa

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.

All Submitters of data may be contacted directly via www.gisaid.org

Authors are sorted alphabetically.

Accession ID	Originating Laboratory	Submitting Laboratory	Authors
EPI_ISL_1040548, EPI_ISL_1040550, EPI_ISL_1040551, EPI_ISL_1040552, EPI_ISL_1040553, EPI_ISL_1040554, EPI_ISL_1040555	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040556, EPI_ISL_1040557, EPI_ISL_1040558, EPI_ISL_1040559, EPI_ISL_1040560, EPI_ISL_1040561	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040562	Heideveld Emergency Centre	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040563	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040562	False Bay Hospital w/ FBH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040563, EPI_ISL_1040565, EPI_ISL_1040566, EPI_ISL_1040567, EPI_ISL_1040568	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040569, EPI_ISL_1040590, EPI_ISL_1040591, EPI_ISL_1040592, EPI_ISL_1040593	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040594, EPI_ISL_1040595	Harcover Park CHC w/ HPH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040596, EPI_ISL_1040597, EPI_ISL_1040598, EPI_ISL_1040599, EPI_ISL_1040700	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040701	Mowbray Maternity Hospital w/ MMH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040702	Lenteguur Hospital w/ LGH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040703	Neuweltpark Clinic w/ NYP	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040704, EPI_ISL_1040705, EPI_ISL_1040706, EPI_ISL_1040707, EPI_ISL_1040708	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040709	Lenteguur Hospital w/ LGH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040710	Gugulethu CHC w/ GDH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040711	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040712	Mowbray Maternity Hospital w/ MMH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040713	Gugulethu CHC w/ GDH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040714	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040715	Gugulethu CHC w/ GDH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040716	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040717	Harcover Park CHC w/ HPH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040718	Gugulethu CHC w/ GDH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040719	Heideveld CDC w/ HVP	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040720	Groote Schuur Hospital w/ GSH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson
EPI_ISL_1040721	Gugulethu CHC w/ GDH	NHLSUCT	Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Bruna Galvao, Innocent Mudau, Marvin Hsiao, Kruger Marais, Diana Hardie, Stephen Korsman, Carolyn Williamson

Bibliography

- [1] Motoo Kimura. “Estimation of evolutionary distances between homologous nucleotide sequences”. In: *Proceedings of the National Academy of Sciences* 78.1 (1981), pp. 454–458.