

Supplementary Materials for

Swiss public health measures associated with reduced SARS-CoV-2 transmission
using genome data

Sarah A. Nadeau, Timothy G. Vaughan, Christiane Beckmann, Ivan Topolsky, Chaoran Chen, Emma Hodcroft, Tobias Schär, Ina Nissen, Natascha Santacroce, Elodie Burcklen, Pedro Ferreira, Kim Philipp Jablonski, Susana Posada-Céspedes, Vincenzo Capece, Sophie Seidel, Noemi Santamaria de Souza, Julia M. Martinez-Gomez, Phil Cheng, Philipp P. Bosshard, Mitchell P. Levesque, Verena Kufner, Stefan Schmutz, Maryam Zaheri, Michael Huber, Alexandra Trkola, Samuel Cordey, Florian Laubscher, Ana Rita Gonçalves, Sébastien Aeby, Trestan Pilonel, Damien Jacot, Claire Bertelli, Gilbert Greub, Karoline Leuzinger, Madlen Stange, Alfredo Mari, Tim Roloff, Helena Seth-Smith, Hans H. Hirsch, Adrian Egli, Maurice Redondo, Olivier Kobel, Christoph Noppen, Niko Beerenwinkel, Richard A. Neher, Christian Beisel, and Tanja Stadler*

*Correspondence to: tanja.stadler@bsse.ethz.ch

This PDF file includes:

Supplementary Text
Figs. S1 to S7
Tables S1 to S3

Supplementary Text

Sensitivity analyses for identifying introductions

Here we describe different sensitivity analyses we performed for the definition of an introduction.

Number of focal Swiss sequences analyzed. To assess whether the number of identified introductions saturates as we add more sequences, we did a sub-sampling analysis. We sub-sampled the Swiss genome sequences used in our main analysis from 5% of confirmed cases down to 1% of confirmed cases. For each new sub-sampling value, we calculated the number of introductions (singletons and transmission chains) that we would have identified given the sub-sampled sequence set. Figure S3A shows that as we add sequences, we do not reach saturation. Therefore, if we were to include even more sequences, we would almost certainly identify more introductions into Switzerland.

Ratio of foreign context to focal Swiss sequences analyzed. Next, we assessed how summary statistics about identified introductions and the runtime of our pipeline changed as we added more and more foreign context sequences. We kept the number of focal Swiss sequences constant at 5% of confirmed cases each week and added foreign context sequences at a 1:1, 2:1, and 3:1 ratio to focal Swiss sequences. Figure S3B shows that as we add foreign context sequences, we identify more numerous, smaller introductions. However, the greatest differences come from the different assumptions about how to resolve polytomies (few vs. many introductions), not the ratio of foreign context to focal Swiss sequences. Therefore, we chose to present results using the 2:1 ratio to balance speed (smaller dataset = faster tree search convergence) and information content (larger dataset = more precise estimates of introduction dates).

Criteria for identifying introductions. After deciding on a ratio of foreign context to focal Swiss sequences, we assessed how summary statistics about identified introductions changed depending on the precise heuristic definition of an introduction. Given phylogenetic trees with polytomies, we identified introductions from these trees as described in the Materials and methods above. Here, we varied (a) the number of export events allowed from each introduction and (b) the maximum number of export events allowed to occur along each single internal branch. Figure S3C shows that increasing (a) yields fewer, larger introductions. Increasing (b) for each level of (a) has a much smaller effect. We note that again, the greatest differences come from the different assumptions about how to resolve polytomies. Therefore, we chose to present results using an introduction definition based on (a) maximum three exported lineages and (b) maximum one consecutive export on each internal branch. This allows for some exports from Swiss introductions but not arbitrarily many.

Sensitivity analyses for phylodynamic modelling

Here we describe a sensitivity analysis and some example intermediate outputs from our phylodynamic analysis.

Sampling proportion prior. We wanted to check that our phylodynamic estimates of a transmission damping factor are robust to our prior on the sampling proportion. Therefore, we repeated our analyses using two different priors on the sampling proportion. The first prior was a uniform distribution between 0 and 1. This broad prior allows the sampling proportion to assume any plausible value. The second prior was a uniform distribution between 0 and an upper bound equal to the number of genome sequences analyzed in a week divided by the number of confirmed

cases that week. This narrower prior assumes that there are at least as many cases each week as confirmed cases. Figure S5A shows that in Switzerland, the estimated sampling proportion in late fall 2020 varies greatly depending on the prior. A drop in SARS-CoV-2 diversity in Switzerland during this period might explain why the inference under the broader sampling prior estimates a proportion corresponding to fewer individuals than we know were infected during this time. Figure S5b shows that the effective reproductive number estimates in fall 2020 for Switzerland more closely match estimates based on confirmed case data alone when the sampling proportion is treated as a fitting parameter, i.e., under the first, broad prior. Therefore, we report results under this prior in the main text. In Figure S6A, we show that the damping factor results are qualitatively similar between the two sampling proportion priors.

Logged trees. Finally, we logged phylogenetic trees for a few introductions. These trees were sampled by the Markov chains in the phylodynamic analyses. Note that the damping factor results are jointly inferred from all the branching events across introductions in each time period. But since logging all introductions would yield huge log files, we logged just a few individual introductions under each set of model assumptions for visual inspection. For each set of model assumptions and each month, we logged trees for the 50th and 95th percentile largest introductions that were first sampled that month and eventually yielded > 2 samples. Figure S7 shows as an example summary trees for these introductions under the phylodynamic analysis for Switzerland with an un-bounded sampling proportion prior.

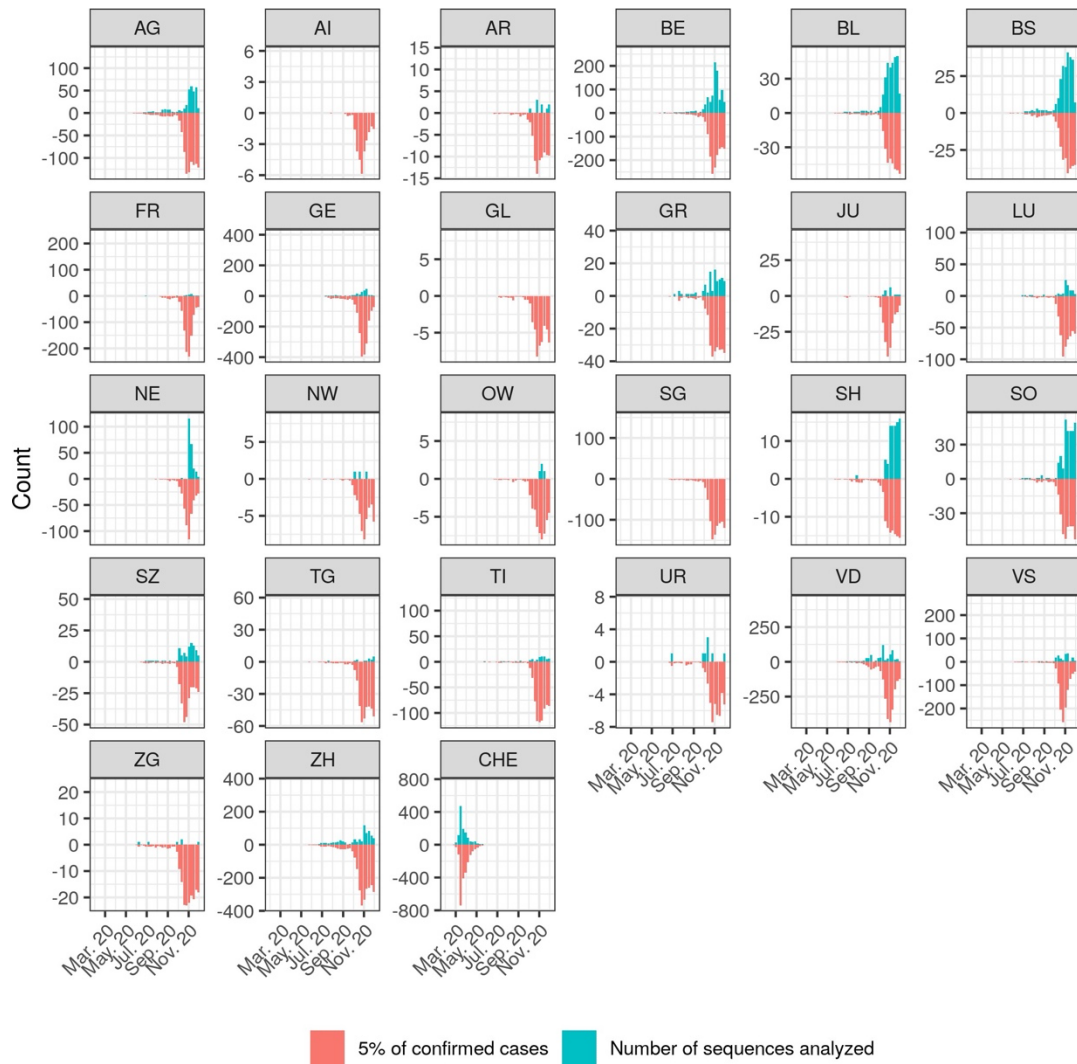


Fig. S1.

Spatiotemporal representativeness of analyzed genome sequences. The mirror y-axis aims to contrast temporal trends in confirmed cases (red) versus analyzed sequences (blue); all values are positive counts. Facet titles are standard abbreviations for Swiss cantons. “CHE” represents Swiss-wide cases and sequences, since case count reporting by canton began mid-May 2020.

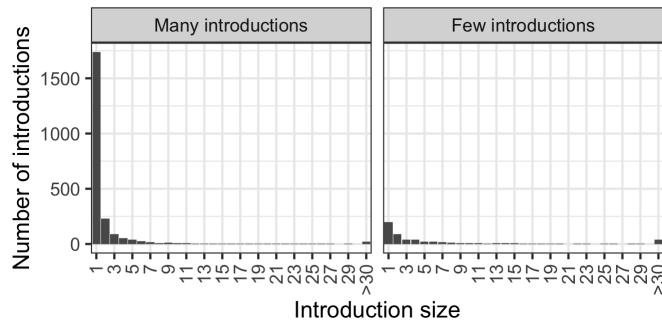


Fig. S2.
Size distribution of estimated introductions.

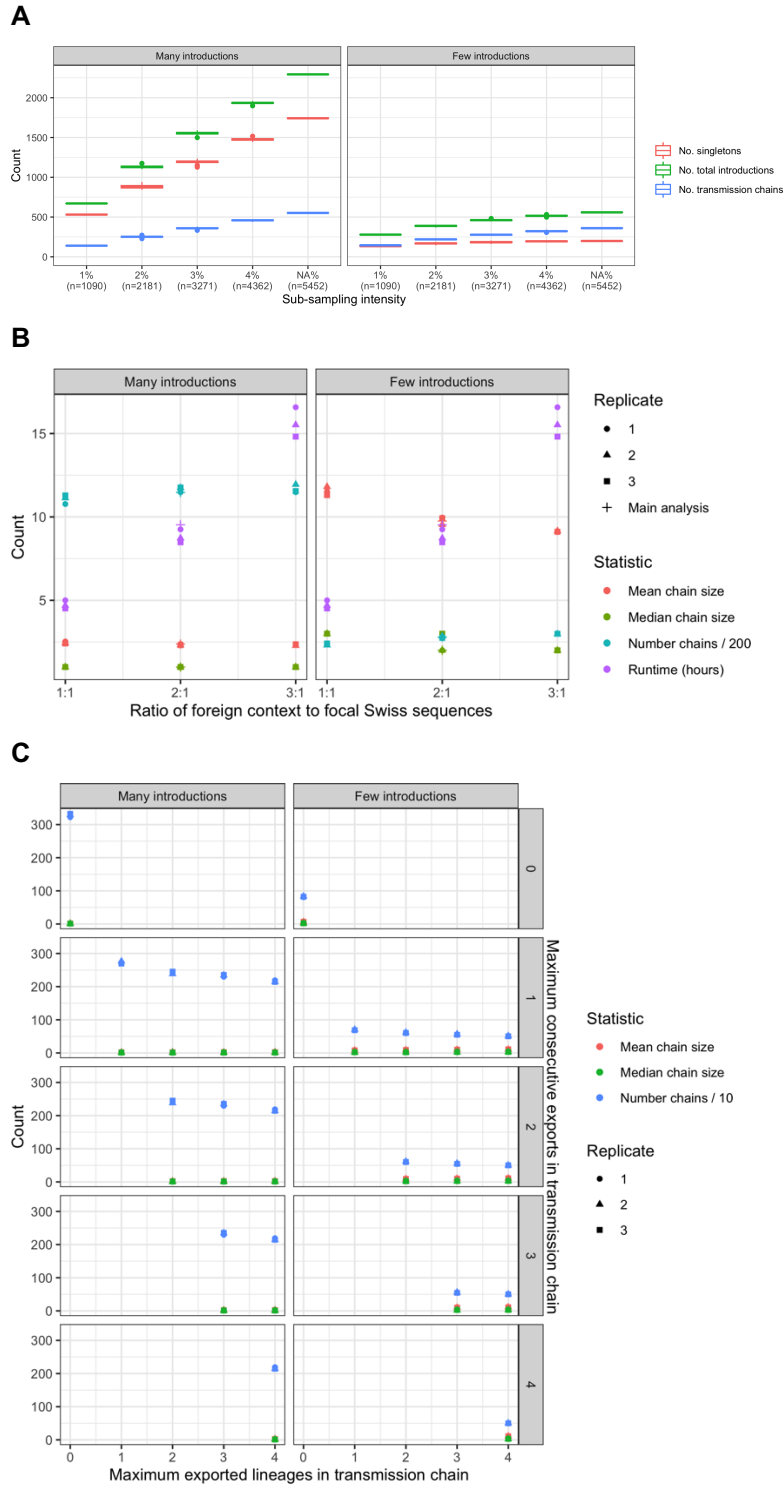


Fig. S3.

Sensitivity analyses for the definition of an introduction, see the supplementary text for more information. (A) shows sensitivity to number of focal sequences analyzed, (B) shows sensitivity to the ratio of foreign context to focal sequences analyzed, and (C) shows sensitivity to the heuristic thresholds used to define an introduction given a tree. All statistics were generated under two different polytomy assumptions giving rise to either many or few introductions.

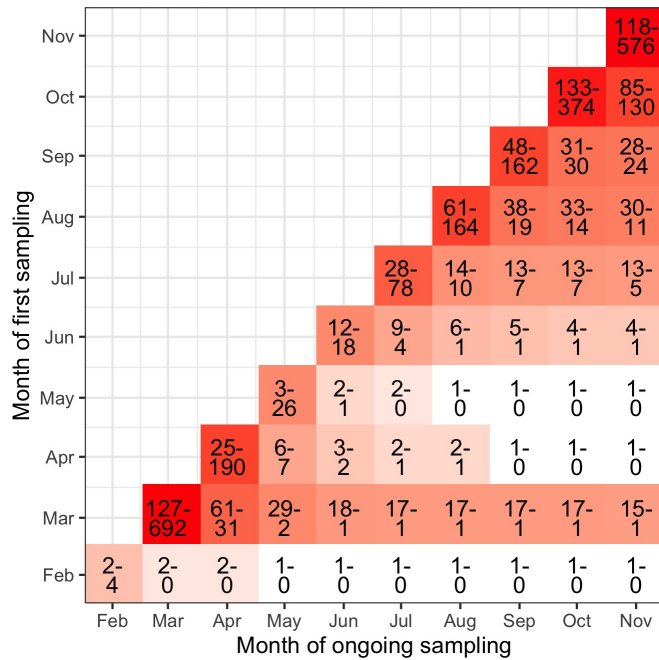


Fig. S4.

Heatmap of the number of newly sampled introductions in Switzerland each month (diagonal entries) and the number continuing to persisting into each following month (off-diagonal entries). Introduction are counted once in the month they are first sampled (“Month of first sampling”) and one every following month (“Month of ongoing sampling”) until the date of the latest sample. The ranges are between two point estimates generated assuming either few or many introductions.

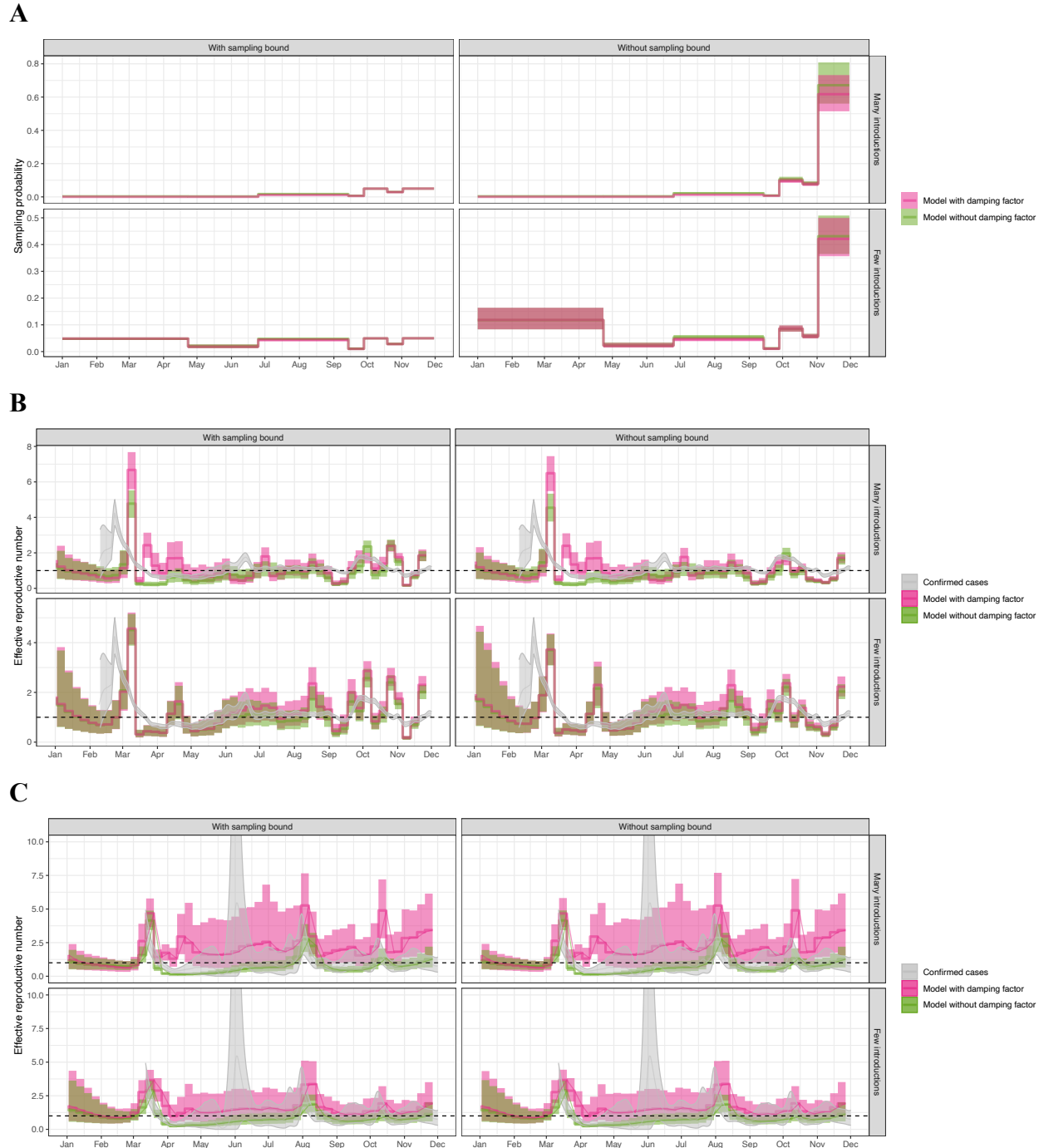


Fig. S5.

Phylodynamic estimates for (A) the sampling probability in Switzerland and the time-varying effective reproductive number R_e in Switzerland (B) and New Zealand (C). R_e estimates are overlaid on estimates generated from case count data (18) in grey. Additionally, R_e estimates from the models with a damping factor (pink) are the “baseline” R_e before introduction-specific damping (i.e. before application of a damping factor once introductions are older than 2-days post sampling).

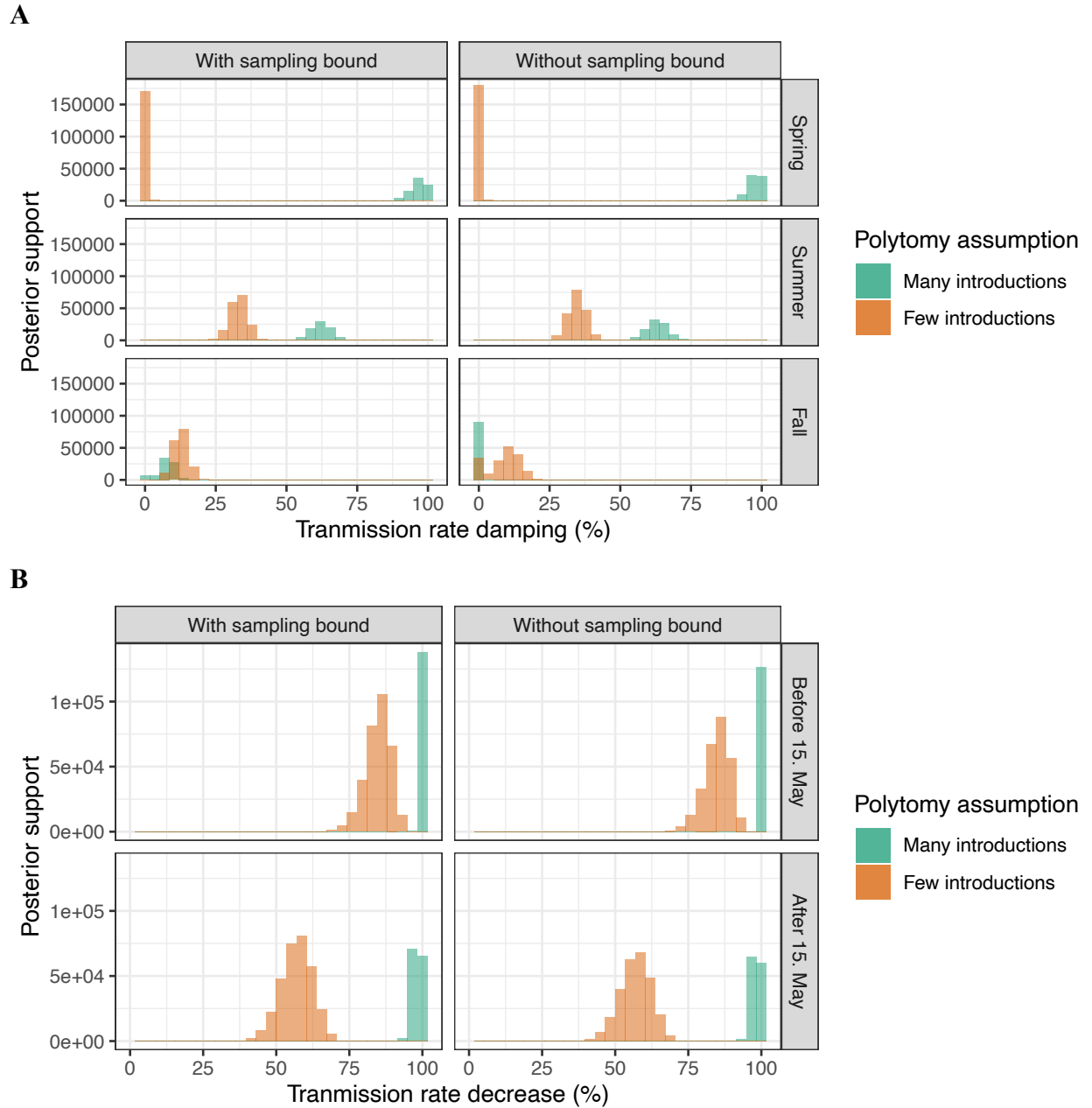
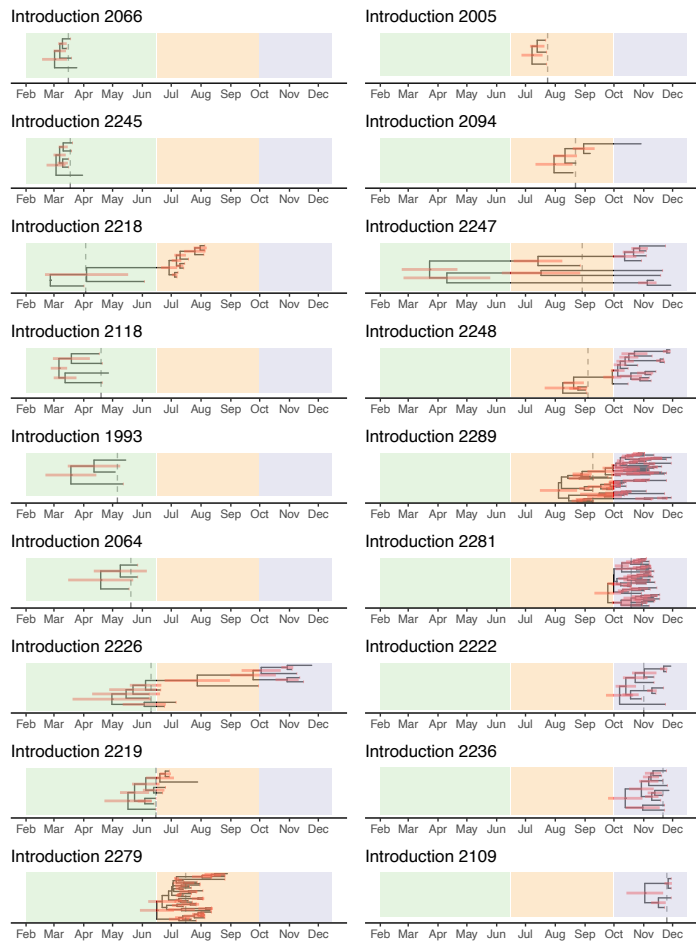
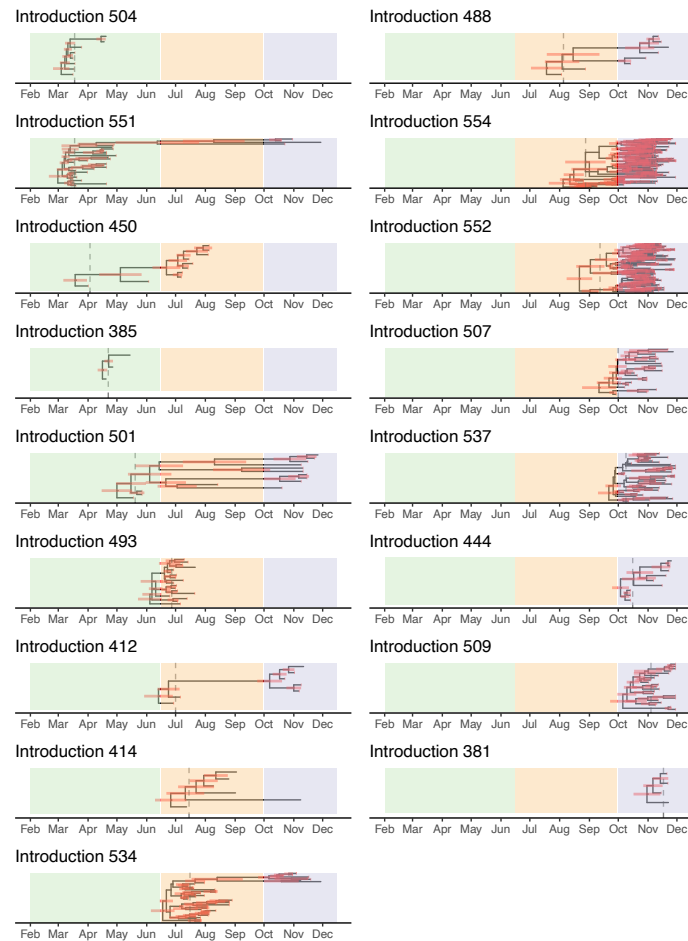


Fig. S6.

Phylodynamic estimates for the damping factor in (A) Switzerland and (B) New Zealand in different time periods, conditioned on introductions defined using different polytomy assumptions and using different priors on the sampling probability.

A**B****Fig. S7.**

Summary trees from the phylodynamic analysis for Switzerland using different polytomy assumptions (**A**) many introductions and (**B**) few introductions. The 50th and 95th percentile largest introductions among those first sampled each month are shown. The three color regions represent the spring (green), summer (orange) and fall (blue) periods. The vertical dashed line shows the date at which the transmission rate can slow for each introduction - two days after the first sample date. The red bars show the 95% highest posterior density uncertainty in node dates.

Table S1.

Summary of Pango lineages analyzed. If more than 50% of the samples from a lineage in the full, quality-filtered dataset were Swiss, we aggregated them into the parent lineage. The percentage of Swiss samples in the final, aggregated lineage sets are given in column “% lineage Swiss”. Lineage aliases were also aggregated with their extended-form names. A separate phylogeny was constructed for each lineage analyzed.

Lineage analyzed	No. Swiss samples analyzed	Lineages aggregated	% lineage Swiss
B.1.160	1347	B.1.160, B.1.160.10, B.1.160.11, B.1.160.12, B.1.160.14, B.1.160.15, B.1.160.16, AB, B.1.160.19, B.1.160.20, B.1.160.22, B.1.160.26, B.1.160.29, B.1.160.30, B.1.160.31, B.1.160.32, B.1.160.9, B.1.160.16.1, AB.1	19
B.1.177	1260	B.1.177, B.1.177.23, B.1.177.28, B.1.177.43, B.1.177.44, B.1.177.71	4.8
B.1	930	B.1, B.1.214.2	2.2
B.1.1	655	B.1.1, B.1.1.144, B.1.1.327, B.1.1.39, AQ, B.1.1.524	3.1
B.1.221	176	B.1.221	8.1
B.1.1.70	108	B.1.1.70, AP	15
B.1.416.1	105	B.1.416.1	45
B.1.258	101	B.1.258	4.4
B.1.367	60	B.1.367	10
B.1.236	59	B.1.236	33
B.1.1.1.35	53	B.1.1.1.35, C.35	13
B.1.36.1	47	B.1.36.1	35
B.1.128	31	B.1.128	3.3
B.1.93	31	B.1.93	3.3
B.1.1.277	27	B.1.1.277, K	5.8
B.1.1.47	24	B.1.1.47	32
B.1.1.269	19	B.1.1.269	5
B.1.1.1.36	16	B.1.1.1.36, C.36, B.1.1.1.36.2, C.36.2	9.3
B.1.1.10	16	B.1.1.10, L	2.7
B.1.1.7	16	B.1.1.7, Q	0.85
B.1.1.1	15	B.1.1.1, C, B.1.1.1.5, C.5	0.86
B.1.1.189	15	B.1.1.189	12
B.1.146	15	B.1.146	30
B.1.1.232	14	B.1.1.232, AK	3.1
B	11	B	0.44
B.1.1.153	11	B.1.1.153	6.5
B.1.1.305	11	B.1.1.305, AF, B.1.1.305.1, AF.1	8.4

B.1.1.372	11	B.1.1.372	0.95
B.1.177.75	11	B.1.177.75	12
B.1.177.77	11	B.1.177.77	6.1
B.1.1.200.1	10	B.1.1.200.1, AN.1	33
B.1.147	10	B.1.147	0.84
B.1.177.81	10	B.1.177.81	1.8
B.1.1.37	9	B.1.1.37	0.42
B.1.177.33	8	B.1.177.33	4.5
B.1.36	8	B.1.36	0.8
B.1.509	8	B.1.509	2.3
B.1.1.433	7	B.1.1.433	7.8
B.1.1.521	7	B.1.1.521	19
B.1.36.17	7	B.1.36.17	1.2
B.1.8	7	B.1.8	1.8
B.1.91	7	B.1.91	1.6
B.1.177.51	6	B.1.177.51	20
B.1.258.17	6	B.1.258.17	1.4
B.1.467	6	B.1.467	33
B.1.1.242	5	B.1.1.242	35
B.1.1.58	5	B.1.1.58	14
B.1.177.83	5	B.1.177.83	7.8
B.1.177.85	5	B.1.177.85	11
B.1.535	5	B.1.535	0.59
B.40	5	B.40	0.23
B.1.1.218	4	B.1.1.218	5.2
B.1.1.241	4	B.1.1.241, AH	4.2
B.1.1.428	4	B.1.1.428	50
B.1.1.464	4	B.1.1.464, AW	1.2
B.1.258.14	4	B.1.258.14	11
B.1.356	4	B.1.356	0.82
A	3	A	0.15
B.1.1.170	3	B.1.1.170	2.6
B.1.1.231.1	3	B.1.1.231.1, AL.1	0.34
B.1.1.297	3	B.1.1.297, AG	1.9
B.1.1.317	3	B.1.1.317, AS	2.1
B.1.1.371	3	B.1.1.371	6.2
B.1.177.52	3	B.1.177.52, Y	2.8
B.1.177.53	3	B.1.177.53, W	3.6
B.1.389	3	B.1.389	1.5

B.1.474	3	B.1.474	14
B.1.480	3	B.1.480	4.3
B.1.9.5	3	B.1.9.5	11
B.11	3	B.11	1.8
B.3	3	B.3	0.37
A.2	2	A.2	0.22
B.1.1.219	2	B.1.1.219	1.6
B.1.1.243	2	B.1.1.243	4.2
B.1.1.33	2	B.1.1.33, N	0.11
B.1.1.44	2	B.1.1.44	0.58
B.1.1.50	2	B.1.1.50	1.2
B.1.160.28	2	B.1.160.28	1.4
B.1.177.15	2	B.1.177.15, AA	0.21
B.1.177.32	2	B.1.177.32	1.1
B.1.177.53.1	2	B.1.177.53.1, W.1	7.7
B.1.177.55	2	B.1.177.55	0.87
B.1.177.60	2	B.1.177.60, U	2.5
B.1.177.62	2	B.1.177.62	6.2
B.1.177.80	2	B.1.177.80	17
B.1.177.82	2	B.1.177.82	0.63
B.1.177.86	2	B.1.177.86	2.1
B.1.218	2	B.1.218	6.5
B.1.408	2	B.1.408	3.5
B.1.416	2	B.1.416	0.94
B.1.523	2	B.1.523	0.88
B.1.9.4	2	B.1.9.4	12
B.28	2	B.28	0.57
B.4	2	B.4	0.54
B.58	2	B.58	2.2
B.59	2	B.59	1.3
A.5	1	A.5	0.21
B.1.1.1.30	1	B.1.1.1.30, C.30	0.19
B.1.1.142	1	B.1.1.142	6
B.1.1.145	1	B.1.1.145	4.5
B.1.1.198	1	B.1.1.198	0.18
B.1.1.221	1	B.1.1.221	1.2
B.1.1.266	1	B.1.1.266	4.9
B.1.1.28	1	B.1.1.28, P	0.066
B.1.1.294	1	B.1.1.294, M	0.28

B.1.1.294.2	1	B.1.1.294.2, M.2	50
B.1.1.315	1	B.1.1.315, AD	1.4
B.1.1.331	1	B.1.1.331	2.4
B.1.1.336	1	B.1.1.336	7.1
B.1.1.355	1	B.1.1.355	2.5
B.1.1.369	1	B.1.1.369	0.046
B.1.1.406	1	B.1.1.406	3.1
B.1.1.409	1	B.1.1.409	0.83
B.1.1.519	1	B.1.1.519	2.6
B.1.1.71	1	B.1.1.71	1.3
B.1.12	1	B.1.12	0.89
B.1.127	1	B.1.127	0.53
B.1.149	1	B.1.149	2.7
B.1.177.31	1	B.1.177.31	50
B.1.177.50.1	1	B.1.177.50.1, Z.1	0.5
B.1.177.53.3	1	B.1.177.53.3, W.3	0.65
B.1.177.6	1	B.1.177.6	0.19
B.1.177.7	1	B.1.177.7	0.03
B.1.177.72	1	B.1.177.72	1.8
B.1.2	1	B.1.2	0.0053
B.1.213	1	B.1.213	4.2
B.1.220	1	B.1.220	1.2
B.1.221.1	1	B.1.221.1	0.34
B.1.229	1	B.1.229	1.1
B.1.258.4	1	B.1.258.4	0.46
B.1.258.7	1	B.1.258.7	0.27
B.1.258.9	1	B.1.258.9	0.65
B.1.36.22	1	B.1.36.22	0.24
B.1.36.24	1	B.1.36.24	4.5
B.1.36.35	1	B.1.36.35	2.1
B.1.397	1	B.1.397	0.86
B.1.398	1	B.1.398	1.4
B.1.400	1	B.1.400	0.099
B.1.406	1	B.1.406	1.9
B.1.415	1	B.1.415	1.4
B.1.513	1	B.1.513	1.4
B.1.520	1	B.1.520	0.1
B.1.540	1	B.1.540	1.6
B.1.88.1	1	B.1.88.1	0.6

B.39	1	B.39	0.26
B.55	1	B.55	0.55
B.6	1	B.6	0.14
None	1	None	1

Table S2.

Sampling proportion change-points for the phylodynamic analysis on Swiss data. The sampling proportion was modelled as a piece-wise-constant function in time, with the following change-points motivated by major shifts in the testing regime or genome sequencing intensity in Switzerland.

Start date	Description
23 April 2020	All symptomatic individuals can get tested
25 June 2020	Government pays for tests for symptomatic individuals
14 September 2020	Genome sampling \ll 5% of confirmed cases
28 September 2020	Number of tests conducted and % positivity dramatically increase, genome sampling also increases
19 October 2020	Genome sampling \ll 5% of confirmed cases again
11 November 2020	Genome sampling increases again

Table S3.

Top 20 largest SARS-CoV-2 sequencing data contributors to GISAID in 2020 by submitting lab.

Submitting lab	Countries represented (ISO codes)	Number of sequences
Wellcome Sanger Institute for the COVID-19 Genomics UK (COG-UK) Consortium	GBR	96441
COVID-19 Genomics UK (COG-UK) Consortium	GBR	71371
Albertsen Lab, Department of Chemistry and Bioscience, Aalborg University, Denmark	DNK	27936
Houston Methodist Hospital	USA	27409
Pathogen Genomics Center, National Institute of Infectious Diseases	JPN; MMR	19708
Department of Biosystems Science and Engineering, ETH Zürich	CHE	11357
MDU-PHL	AUS; TLS	10459
TGen North	USA	9491
Wyoming Public Health Laboratory	USA	9172
Aalborg University	DNK	8439
SeqCOVID-SPAIN consortium/IBV(CSIC)	ESP	8279
Chan-Zuckerberg Biohub	USA	7803
BCCDC Public Health Laboratory	CAN	7646
Laboratoire de santé publique du Québec	CAN	6914
Andersen lab at Scripps Research	JOR; MEX; USA	6258
Utah Public Health Laboratory	USA	5925
MEPHI, Aix Marseille University	FRA	5617
Respiratory Virus Unit, Microbiology Services Colindale, Public Health England	GBR; UKR	5142
deCODE genetics	ISL	5005
Erasmus Medical Center	BEL; BHR; LUX; NLD; SUR	4594