Quantifying SARS-CoV-2 spread in Switzerland based on genomic sequencing data

Authors: Sarah Nadeau[1,2], Christiane Beckmann[3], Ivan Topolsky[1,2], Timothy Vaughan[1,2], Emma Hodcroft[2,4], Tobias Schär[1], Ina Nissen[1], Natascha Santacroce[1], Elodie Burcklen[1], Pedro Ferreira[1,2], Kim Philipp Jablonski[1,2], Susana Posada-Céspedes[1,2], Vincenzo Capece[1], Sophie Seidel[1,2], Noemi Santamaria de Souza[5], Julia M. Martinez-Gomez[6], Phil Cheng[6], Philipp Bosshard[6], Mitchell P. Levesque[6], Verena Kufner[7], Stefan Schmutz[7], Maryam Zaheri[7], Michael Huber[7], Alexandra Trkola[7], Samuel Cordey[8], Florian Laubscher[8], Ana Rita Gonçalves[9], Karoline Leuzinger[10,11], Madlen Stange[10], Alfredo Mari[10], Tim Roloff[10], Helena Seth-Smith[10], Hans Hirsch[10], Adrian Egli[10], Maurice Redondo[3], Olivier Kobel[3], Christoph Noppen[3], Niko Beerenwinkel[1,2], Richard A. Neher[2,4], Christian Beisel[1], and Tanja Stadler[1,2,*]

Affiliations:

[1]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland.

[2]SIB Swiss Institute of Bioinformatics, Basel, Switzerland.

[3]Viollier AG, Allschwil, Switzerland.

[4]Biozentrum, University of Basel, Basel, Switzerland.

[5]Insitute of Microbiology, D-BIOL, ETH Zürich, Zürich, Switzerland.

[6]Department of Dermatology, University of Zürich, University of Zürich Hospital, Switzerland.

[7]Institute of Medical Virology, University of Zürich, Zürich, Switzerland.

[8]Laboratory of Virology, Division of Infectious Diseases and Division of Laboratory Medicine, University Hospitals of Geneva & Faculty of Medicine, University of Geneva, Geneva, Switzerland.

[9]Swiss National Reference Centre for Influenza, University Hospitals of Geneva, Geneva, Switzerland.

[10]University Hospital Basel, Basel, Switzerland.

[11]Transplantation & Clinical Virology, Department Biomedicine, University of Basel, Basel, Switzerland.

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

Abstract:

Pathogen genomes provide insights into their evolution and epidemic spread. We sequenced 1,439 SARS-CoV-2 genomes from Switzerland, representing 3-7% of all confirmed cases per week. Using these data, we demonstrate that no one lineage became dominant, pointing against evolution towards general lower virulence. On an epidemiological level, we report no evidence of cryptic transmission before the first confirmed case. We find many early viral introductions

from Germany, France, and Italy and many recent introductions from Germany and France. Over the summer, we quantify the number of non-traceable infections stemming from introductions, quantify the effective reproductive number, and estimate the degree of undersampling. Our framework can be applied to quantify evolution and epidemiology in other locations or for other pathogens based on genomic data.

One Sentence Summary: We quantify SARS-CoV-2 spread in Switzerland based on genome sequences from our nation-wide sequencing effort.

Main Text:

SARS-CoV-2 originated in late 2019 in Wuhan, China and has been spreading rapidly around the globe in 2020. After the initial outbreak in China, Europe soon became the epicenter of the pandemic. As a counter measure, European countries implemented a variety of non-pharmaceutical interventions (NPIs), ranging from border closures to social distancing regulations (*1*). Case numbers declined to a low in early summer but are now increasing again across Europe (*2*). Switzerland is centrally located in Europe and is well-connected with other European countries. Switzerland also implemented some of the same NPIs as other European countries. Thus, we hypothesize that viral introduction and epidemiological dynamics within Switzerland are exemplary for a European country.

We characterize the SARS-CoV-2 epidemic in Switzerland from the first confirmed case on Feb. 24th until Aug. 31st based on SARS-CoV-2 genome sequences. We define an “early epidemic” from Feb. 24th until the low point in daily cases around Jun. 1st and a "summer epidemic” from Jun. 1st until the end of our sampling period on Aug. 31st. To characterize the epidemic over the complete period, we generate 1,439 sequences from our nation-wide sequencing project, representing 3-7% of all laboratory-confirmed cases per week across Switzerland. We couple these data with 675 additional sequences from the early epidemic collected by labs in Basel (*3*) , Geneva and Zurich to represent in total 5% of all laboratory-confirmed cases in Switzerland until Aug. 31st. We combine this Swiss dataset with additional sequences from across the globe to address five main goals:

(i) to assess the change in viral genomic diversity within Switzerland,

(ii) to estimate the timing of introduction events and longevity of transmission chains within Switzerland,

(iii) to estimate the geographic origin of introductions and geographic spread of transmission chains within Switzerland,

(iv) to quantify the relative importance of introductions versus local transmission through time, and

(v) to quantify the effective reproductive number and case underreporting in Switzerland.

To meet these goals, we perform a phylogenetic analysis of the sequence dataset as detailed in the Materials and Methods. Briefly, we pair Swiss SARS-CoV-2 sequences with sequences from abroad that are either genetically similar to Swiss sequences (“similarity dataset”) or that stem from countries with lots of travel connections to Switzerland (“context dataset”). We then

construct a dated maximum-likelihood phylogenetic tree using all three datasets that approximates the transmission history of the considered cases.

We define Swiss transmission chains based on this phylogenetic tree. Essentially, Swiss sequences that are more similar to each other than to sequences from abroad are proposed to belong to the same transmission chain. We apply two different extreme criteria to define transmission chains, namely, a "maximum introductions, minimum local transmission" assumption and a "minimum introductions, maximum local transmission" assumption (see Materials and Methods for details). In this way, we refrain from arbitrarily resolving phylogenetic uncertainty. Instead, we provide upper- and lower-bound estimates for epidemiological parameters of the Swiss epidemic based on these two extreme assumptions. All Swiss sequences that are not linked to a transmission chain are coined singletons. Importantly, each Swiss transmission chain and each singleton is assumed to result from an independent introduction event into Switzerland.

After identifying Swiss introductions, we estimate their geographic origins using a parsimony-based approach. For this, we consider only sequences in the context dataset. Since the context dataset is constructed based on travel connections with Switzerland and not foreign sequencing effort, we aim to reconstruct the true source location of introductions rather than reflect biased sampling.

We extract answers for goals (i-iv) directly from the phylogenetic tree and the corresponding introductions and transmission chains. Finally, we perform additional Bayesian phylodynamic analyses of the identified Swiss transmission chains and singletons to achieve goal (v). Goals (i, iii, iv, v) require a sequence set which is representative for the Swiss epidemic and thus we use only Swiss sequences from our nation-wide SARS-CoV-2 sequencing project. For goal (ii) we use all available Swiss sequences.

Results

(i) Change in viral genomic diversity

First, we assess how the SARS-CoV-2 population changed through time within Switzerland. By sequencing 3-7% of confirmed cases each week (Fig. S1) from across the country (Fig. S2), we argue that we obtain a representative picture of SARS-CoV-2 genetic diversity in Switzerland through time. Here we use the Nextstrain nomenclature standard to describe genetic diversity.

Fig. 1a shows that in Switzerland 20A strains dominated until early summer, then 20B increased in frequency throughout the summer, and finally 20A again became dominant in late August. Since 20B strains increased in frequency while case numbers in Switzerland dropped, and later numbers of new hospitalizations remained low despite rising case numbers (*2*), one may hypothesize that a single 20B strain evolved lower virulence and/or lower pathogenicity before becoming dominant during the summer epidemic. However, on the Swiss phylogenetic tree many lineages cross from the early epidemic into the summer epidemic. Taking June 1 to be the start of the summer epidemic, 114 lineages give rise to 721 sequences from the summer epidemic, with the 8 most fecund lineages producing 60% of sequences from the summer epidemic (Fig. 1b). Similar results are obtained using different date cutoffs between the early and summer epidemic (Fig. S3).

In summary, we show that although the dominant clade in the Swiss epidemic has changed several times, it is implausible that any single strain evolved lower virulence (or any other remarkable phenotype) and then became dominant.

(ii) Timing of introduction events and longevity of transmission chains

Next, we estimated the timing of viral introductions into Switzerland by focusing on the identified introductions. We note that phylogenetic results need to be interpreted carefully due to incomplete sampling and phylogenetic uncertainty. In fact, in Fig. S4 we show that our sampling is not saturated, meaning that if we were to sequence more genomes, we would find more introductions. As mentioned, we account for phylogenetic uncertainty by presenting results robust to two different assumptions about the degree of local transmission vs. introductions.

Our insights on the introduction times of SARS-CoV-2 lineages into Switzerland rely on the fact that a lineage could have been introduced into Switzerland anytime between the date of the most recent common ancestor (MRCA) of a Swiss transmission chain and the date the MRCA attaches to the rest of the tree. For a singleton, the introduction could have occurred anytime between the attachment point and the sample date. Fig. 2 shows that for the early epidemic we cannot pinpoint transmission chain introduction dates very precisely. Fig. S5 shows that the same holds for singletons. As others have noted, there are not enough informative mutations to resolve the ordering of many transmission events (*4*) or to confidently estimate MRCA dates (*5*). However, with more samples, the MRCAs may become older and the attachment times may become younger, reducing this uncertainty.

Given that transmission chain MRCAs are the earliest evidence of local transmission, we do not find evidence for cryptic spread in Switzerland prior to detection of the first case on Feb. 24th. The oldest estimated MRCA is Feb. 22nd (90th percentile confidence interval Feb. 17th – Mar. 7th). In fact, the majority of lineages which generated Swiss cases during the early epidemic were still circulating outside of Switzerland in the weeks Feb. 26th – Mar. 10th (Fig. S6a). In comparison, only a few were still circulating outside of Switzerland after the lockdown and border closures were implemented in Switzerland on Mar. 17th (Fig. S6a).

About a third of all sampled Swiss transmission chains and more than half of the transmission chains of the early epidemic (here defined as starting spreading prior to June 1) began spreading locally around Mar. 4 – 24th (Fig. S6b). Three of these transmission chains persisted from March until August and due to incomplete sampling, we may underestimate the longevity of some other transmission chains. Finally, transmission chains in the summer epidemic largely started spreading locally after Jun. 15th, the date Switzerland lifted its partial border closure.

These results highlight that there is no evidence of cryptic transmission prior to the first confirmed case in Switzerland, that many introduction events occurred just before Swiss border closures, and that new introduction events occurred just after borders were re-opened.

(iii) Geographic origin and geographic spread of transmission chains

We asked whether the source of Swiss introductions changed over the course of the epidemic thus far. To do this, we estimated the geographic location of attachment points of Swiss introductions. We used only our context dataset of foreign sequences to generate these estimates, which should reduce the effect of uneven foreign sampling efforts. However, we note that we may underestimate the contributions of Italy, France, and Germany during some summer months as we lack sequences from some of these countries (Fig. S7).

In Fig. 2 we show that while initially the origin of introduced strains was from our neighboring countries France, Italy, and Germany, the source location shifted more towards Belgium and the Netherlands in May – July and then, despite under-sampling, to Germany and France more recently. These results were roughly consistent across three different random samplings of context sequences (Fig. S8 – S10). Some of our source location estimates can be confirmed with independent information on case travel history. Where this information is available, our estimates for transmission chain origins largely agree (Fig. 2a).

To investigate epidemic spread between Swiss cantons, we tallied how often samples from the same transmission chain were found in different Swiss cantons (Fig. S11). We observe significantly more mixing than expected by chance between the neighboring cantons of Basel-Stadt and Basel-Land, with 16 transmission chains containing samples from both cantons. None of the cantons mixed less than expected based on the available data.

We conclude that in general, SARS-CoV-2 introduction dynamics into Switzerland shifted from neighboring countries early in the epidemic towards more non-neighboring countries as border restrictions eased and then finally back to neighboring countries near the end of the summer holidays. Within Switzerland, there is evidence that two neighboring cantons share a single intermixed epidemic.

(iv) Relative importance of introductions versus local transmission

Next, we assessed the relative importance of introductions compared to local transmissions over time. Importantly, the chance of sampling a local transmission compared to an introduction increases with increased sampling. Thus, local transmission may be even more significant than we report here.

Fig. 3a shows that we sample many new introductions in March (up to 28-161 in the most extreme week) compared to almost none from mid-May to mid-June. These trends coincide with closed borders in Switzerland from Mar. 17th – Jun. 15th. We then had a stable period with 1-28 introductions per week until the end of July. In August, towards the end of the summer holidays, we see a steady increase in introductions. The probability for a new case to arise from a local transmission (as opposed to an introduction) rose throughout the epidemic to above 80% in July. In August, this probability dropped to around 40%-90% as introductions again increased (Fig. 3a).

To quantify the importance of these trends, we estimate the number of samples that are un-preventable vs. preventable via perfect contact tracing. To do this, we rely on the insight that cases generated between an introduction index case's infection date and case confirmation (we assume a 10-day period (*6*)) cannot be prevented because contact tracers are not yet aware of the transmission chain. We estimate that there were >100 such un-preventable secondary cases in the week prior to the lockdown. This number decreased to <10 per week in late spring, and more recently increased to around 50 per week. On the flipside, Fig. 3b shows that over the summer, 25% or more of the new infections each week could have been prevented under the optimistic assumption that contact tracing could have prevented all cases confirmed more than 10 days after the index case.

These results highlight the fact that transmission by introductions is harder to contain than local transmission as we cannot identify them quickly through contact tracing. As local transmission increases, contact tracing becomes a very effective strategy to prevent onward transmission.

(v) Effective reproductive number

Finally, we quantify the effective reproductive number (R_e) and under-sampling in Switzerland by fitting a phylodynamic model to the identified Swiss transmission chains and singletons. We only present estimates after May 1st because prior to this time, low viral diversity contributes to high uncertainty in the size and number of transmission chains (i.e. Fig. 3a shows that the estimated number of introductions varies widely depending on the assumptions used).

Beginning May 1st, we estimate that R_e was significantly below 1 (Fig. 4a). Coincident with the opening of shops and restaurants on May 11th, R_e increased and our estimates include the threshold of 1 throughout most of the summer. The estimates of R_e based on confirmed case data are contained within our estimated uncertainty intervals for most of the time.

Our phylodynamic analysis co-infers the sampling proportion, which is the proportion of total cases included in the analysis, over time. Fig. 4b shows that the inferred sampling proportion closely matches the fraction of total confirmed cases included in the analysis. Therefore, we do not find evidence of significant undersampling during the summer epidemic in Switzerland.

Overall, genome-based estimates of R_e match independent estimates based on confirmed case data. Pending confirmation with seroprevalence information, genome-based estimates of the sampling proportion suggest most infections over the summer were successfully identified.

Discussion

By analyzing a time-homogenous dataset of Swiss SARS-CoV-2 genome sequences, we generated evolutionary and epidemiological insights on SARS-CoV-2 spread for a centrally located and well-connected European country. The insights that we obtained regarding our goals (i-v) can be directly relevant for public health policies.

First, knowing that no one lineage evolved lower virulence and then became dominant means that the current low numbers of hospitalizations and deaths relative to case numbers in Switzerland (*2*) cannot be explained by viral evolution. In fact, the shift in case age distribution towards younger people (*7*) may well explain the low numbers.

On an epidemiological level, we observe no cryptic spread prior to the first confirmed case in Switzerland, suggesting that the Swiss early surveillance was well-functioning. As recent studies have highlighted (*8*, *9*), early identification is crucial in order to buy time for preparing a pandemic response plan. With currently increasing case numbers across Europe, the usefulness of maintaining and expanding quarantine rules is becoming a widely discussed topic. In the summer epidemic in Switzerland, we identify first Belgium and the Netherlands as important sources for introductions and then more recently, France and Germany. Of these countries, only Belgium was put on the Swiss quarantine list (from Aug. 20) shortly before the end of our study period on Aug. 31. Since then, quarantine has been made mandatory for travelers returning from Belgium, the Netherlands, and many regions of France and Germany. Given our baseline estimates for introductions in the absence of mandatory quarantine, it will be interesting to assess the impact of quarantine rules going forward.

Finally, we estimate that, in comparison to the early epidemic where a 5-10 time underreporting was observed based on seroprevalence studies (*10*), confirmed case counts over the summer closely match genome-based estimates of the total number of infected individuals. This suggests that the SARS-CoV-2 epidemic in Switzerland can in fact be well-monitored. However, this estimate should be assessed for plausibility through additional seroprevalence studies.

We will continue to update our genome-based results on SARS-CoV-2 evolution and epidemiology as the epidemic continues and plan to publish them on the Swiss National COVID-19 Scientific Task Force website and a Swiss Nextstrain site (*11*). Going forward, we call for linking epidemiological metadata, contact tracing information, clinical data, and genomic data in order to extract the maximum amount of information about epidemiological dynamics. Such information is crucial in order to understand where infections happen and thus to design specific and efficient public health interventions.

References and Notes

1. ECDC, Download data on country response measures to COVID-19, (available at https://www.ecdc.europa.eu/en/publications-data/download-data-response-measures-covid-19).

2. COVID-19 Re, (available at https://ibz-shiny.ethz.ch/covid-19-re/).

3. M. Stange, A. Mari, T. Roloff, H. M. Seth-Smith, M. Schweitzer, M. Brunner, K. Leuzinger, K. K. Søgaard, A. Gensch, S. Tschudin-Sutter, S. Fuchs, J. Bielicki, H. Pargger, M. Siegemund, C. H. Nickel, R. Bingisser, M. Osthoff, S. Bassetti, R. Schneider-Sliwa, M. Battegay, H. H. Hirsch, A. Egli, SARS-CoV-2 phylogeny during the early outbreak in the Basel area, Switzerland: import and spread dominated by a 1 single B.1 lineage variant (C15324T) 2 3, doi:10.1101/2020.09.01.20186155.

4. M. Worobey, J. Pekar, B. B. Larsen, M. I. Nelson, V. Hill, J. B. Joy, A. Rambaut, M. A. Suchard, J. O. Wertheim, P. Lemey, The emergence of SARS-CoV-2 in Europe and the US, 1–26 (2020).

5. S. Duchene, L. Featherstone, M. Haritopoulou-Sinanidou, A. Rambaut, P. Lemey, G. Baele, *bioRxiv*, in press, doi:10.1101/2020.05.04.077735.

6. J. Scire, S. Nadeau, T. Vaughan, G. Brupbacher, S. Fuchs, J. Sommer, K. N. Koch, R. Misteli, L. Mundorff, T. Götz, T. Eichenberger, C. Quinto, M. Savic, A. Meienberg, T. Burkard, M. Mayr, C. A. Meier, A. Widmer, R. Kuehl, A. Egli, H. H. Hirsch, S. Bassetti, C. H. Nickel, K. S. Rentsch, W. Kübler, R. Bingisser, M. Battegay, S. Tschudin-Sutter, T. Stadler, Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft. *Swiss Med. Wkly.* **150**, w20271 (2020).

7. BAG, Situation in Switzerland, (available at https://www.bag.admin.ch/bag/en/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/situation-schweiz-und-international.html#-640157857).

8. M. Worobey, J. Pekar, B. B. Larsen, M. I. Nelson, V. Hill, J. B. Joy, A. Rambaut, M. A. Suchard, J. O. Wertheim, P. Lemey, The emergence of SARS-CoV-2 in Europe and North America. *Science (80-.).* (2020), doi:10.1126/SCIENCE.ABC8169.

9. T. Bedford, A. L. Greninger, P. Roychoudhury, L. M. Starita, M. Famulare, M.-L. Huang, A. Nalla, G. Pepper, A. Reinhardt, H. Xie, L. Shrestha, T. N. Nguyen, A. Adler, E. Brandstetter, S. Cho, D. Giroux, P. D. Han, K. Fay, C. D. Frazar, M. Ilcisin, K. Lacombe, J. Lee, A. Kiavand, M. Richardson, T. R. Sibley, M. Truong, C. R. Wolf, D. A. Nickerson, M. J. Rieder, J. A. Englund, J. Hadfield, E. B. Hodcroft, J. Huddleston, L. H. Moncla, N.

F. Müller, R. A. Neher, X. Deng, W. Gu, S. Federman, C. Chiu, J. S. Duchin, R. Gautom, G. Melly, B. Hiatt, P. Dykema, S. Lindquist, K. Queen, Y. Tao, A. Uehara, S. Tong, D. MacCannell, G. L. Armstrong, G. S. Baird, H. Y. Chu, J. Shendure, K. R. Jerome, Cryptic transmission of SARS-CoV-2 in Washington state. *Science (80-.).*, eabc0523 (2020).

10. S. Stringhini, A. Wisniak, G. Piumatti, A. S. Azman, H. Baysson, D. De Ridder, D. Petrovic, S. Schrempft, S. Yerly, I. Arm Vernez, O. Keiser, S. Hurst, D. Trono, D. Pittet, L. Gétaz, F. Chappuis, I. Eckerle, N. Vuilleumier, B. Meyer, A. Flahault, I. Guessous, Repeated seroprevalence of anti-SARS-CoV-2 IgG antibodies in a population-based sample from Geneva, Switzerland. *medRxiv* (2020), doi:10.1101/2020.05.02.20088898.
11. Nextstrain / groups / swiss / ncov / switzerland, (available at https://nextstrain.org/groups/swiss/ncov/switzerland).
12. J. Quick, N. D. Grubaugh, S. T. Pullan, I. M. Claro, A. D. Smith, K. Gangavarapu, G. Oliveira, R. Robles-Sikisaka, T. F. Rogers, N. A. Beutler, D. R. Burton, L. L. Lewis-Ximenez, J. G. De Jesus, M. Giovanetti, S. C. Hill, A. Black, T. Bedford, M. W. Carroll, M. Nunes, L. C. Alcantara, E. C. Sabino, S. A. Baylis, N. R. Faria, M. Loose, J. T. Simpson, O. G. Pybus, K. G. Andersen, N. J. Loman, Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1266 (2017).
13. S. Posada-Céspedes, D. Seifert, I. Topolsky, K. J. Metzner, N. Beerenwinkel, *bioRxiv*, in press, doi:10.1101/2020.06.09.142919.
14. nextstrain/ncov, (available at https://github.com/nextstrain/ncov).
15. N. De Maio, C. Walker, R. Borges, L. Weilguny, G. Slodkowicz, N. Goldman, Issues with SARS-CoV-2 sequencing data (2020), (available at https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473).
16. W-L/ProblematicSites_SARS-CoV2, (available at https://github.com/W-L/ProblematicSites_SARS-CoV2).
17. FSO, Tourist accommodation in the first half of 2020, (available at https://www.bfs.admin.ch/bfs/en/home/statistics/tourism.gnpdetail.2020-0305.html).
18. FSO, Cross-border Commuters Statistics in 2nd quarter 2020, (available at https://www.bfs.admin.ch/bfs/en/home/statistics/work-income.gnpdetail.2020-0507.html).
19. ECDC, Download today's data on the geographic distribution of COVID-19 cases worldwide, (available at https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide).
20. Nextclade, (available at https://clades.nextstrain.org/).
21. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, E. Teeling, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
22. T. H. To, M. Jung, S. Lycett, O. Gascuel, Fast Dating Using Least-Squares Criteria and Algorithms. *Syst. Biol.* **65**, 82–97 (2016).
23. Nextstrain, Nextstrain / ncov / global, (available at https://nextstrain.org/ncov/global?c=region).
24. C. J. Villabona-Arenas, W. P. Hanage, D. C. Tully, Phylogenetic interpretation during outbreaks requires caution. *Nat. Microbiol.* **5** (2020), pp. 876–877.
25. N. F. Müller, D. Wüthrich, N. Goldman, N. Sailer, C. Saalfrank, M. Brunner, N. Augustin, H. M. B. Seth-Smith, Y. Hollenstein, M. Syedbasha, D. Lang, R. A. Neher, O. Dubuis, M. Naegele, A. Buser, C. H. Nickel, N. Ritz, A. Zeller, B. M. Lang, J. Hadfield, T. Bedford,

M. Battegay, R. Schneider-Sliwa, A. Egli, T. Stadler, *bioRxiv*, in press, doi:10.1101/2020.04.27.052225.
26. T. Stadler, D. Kühnert, S. Bonhoeffer, A. J. Drummond, Birth--death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci.* **110**, 228–233 (2013).
27. R. Bouckaert, T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, others, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
28. M. Hasegawa, H. Kishino, T. aki Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
29. Z. Yang, Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
30. V. N. Minin, E. W. Bloomquist, M. A. Suchard, Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25**, 1459–1471 (2008).
31. D. Probst, Data from BAG Tweets made useful., (available at https://github.com/daenuprobst/covid19-cases-switzerland).

Figures

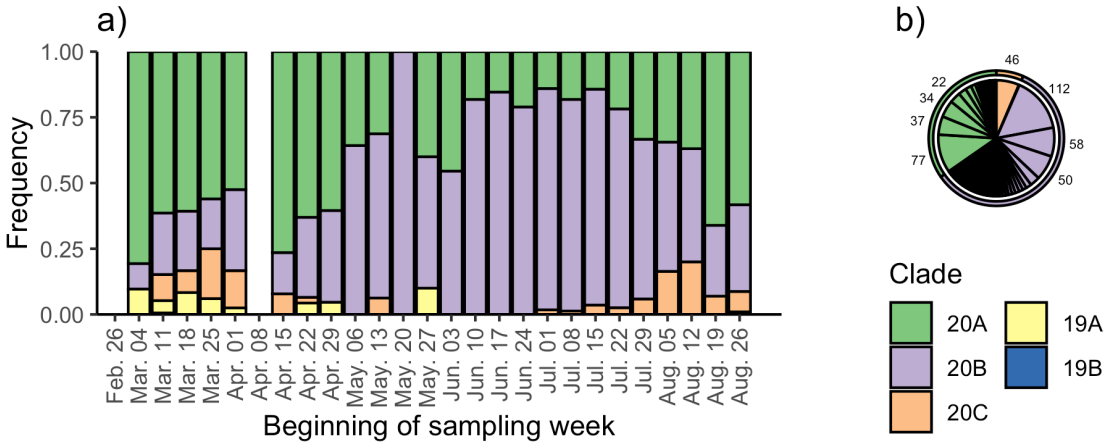

Fig. 1. Genomic diversity of the Swiss epidemic over time based on sequences from our nation-wide sequencing project. **(a)** shows the frequency of different clades over time. **(b)** shows the number of samples from the summer epidemic generated by different lineages from the early epidemic. Each slice represents a unique lineage present on Jun. 1 2020, where the width of the slice is proportional to the number of sampled descendants after Jun. 1. In total, 114 lineages present on Jun. 1 gave rise to 721 samples from the summer epidemic. Combined, the 8 largest lineages (sample numbers shown) gave rise to 60% of the samples from the summer epidemic.

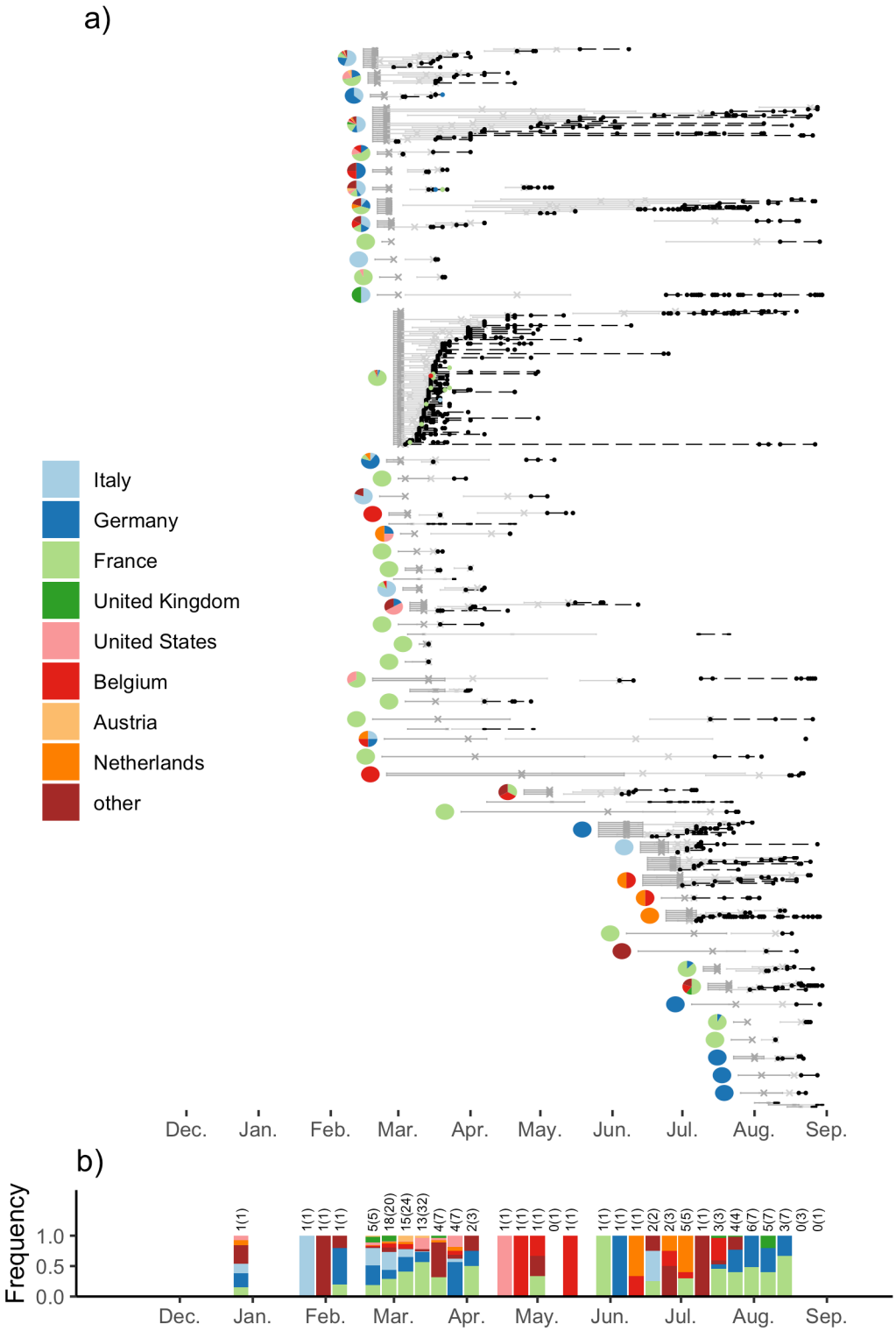

Fig. 2. (a) Swiss transmission chains based on all publicly available Swiss sequences. Dashed lines connect samples from the same transmission chain. Samples with known travel information are colored accordingly. The dark grey intervals are estimates for the attachment time of each transmission chain to the rest of the tree and the light grey intervals are estimates for each transmission chain's MRCA, where crosses are the least-square estimates for the respective date. The pie charts show estimates for the location of each attachment point. We do not infer geographic

origins for transmission chains that cluster with only sequences from the "similarity" dataset. Fig. S4 shows the corresponding plot for singletons. **(b)** The estimated contribution of each country towards introductions (transmission chains and singletons) into Switzerland over time. The bars summarize how frequent each location is across the least-square attachment times (dark grey crosses) dated during each week. The numbers above the bars give the number of attachment points with location information falling in that week; values in parentheses are the total number of attachment points. We note that the bar in late December corresponds to a polytomy with Swiss singleton descendants (Fig. S5). The start of local spread within Switzerland is very likely to have been much later (Fig. S6).

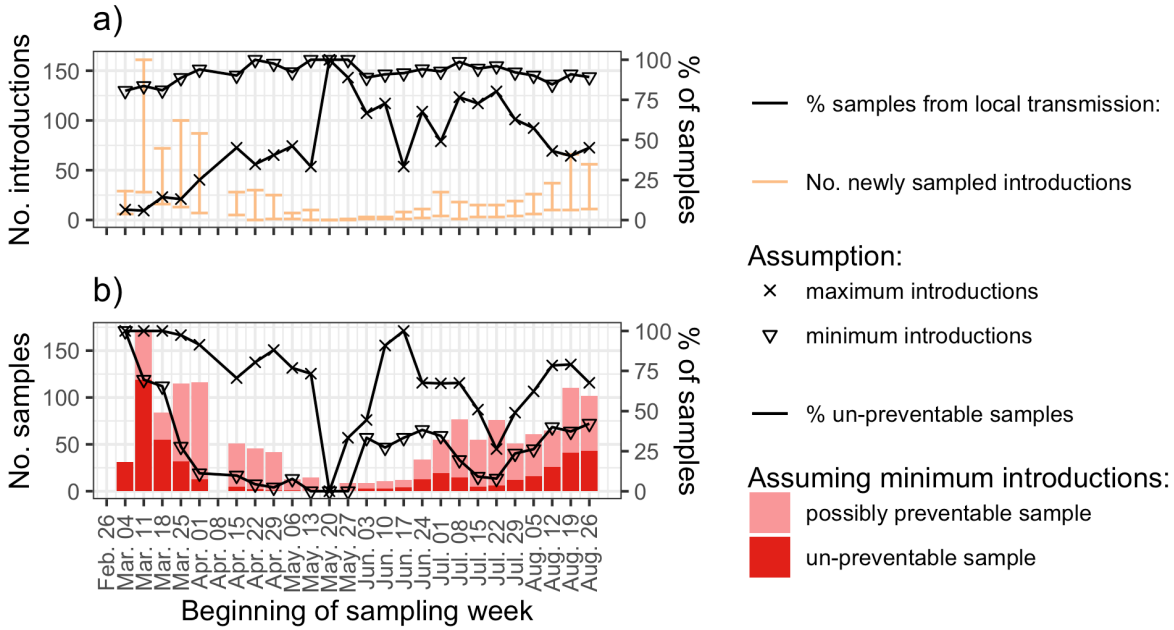

Fig. 3. Importance of local transmission versus introductions through time based on sequences from our nation-wide sequencing project. Here we label the first sample from each introduction the index sample. We then count the number of index samples (i.e. newly sampled introductions) and non-index samples (i.e., samples from local transmission) each week. **(a)** shows the number of newly sampled introductions (transmission chains and singletons) each week in orange and the percentage of samples from local transmission each week in black. **(b)** shows the lower bound number (dark red) and percentage (black) of samples that are un-avoidable even given a perfect test, trace, quarantine, and isolate strategy (see also Supplementary Text). Importantly, un-preventable cases result from transmission chains caused by introductions prior to these introduced transmission chains being traceable.

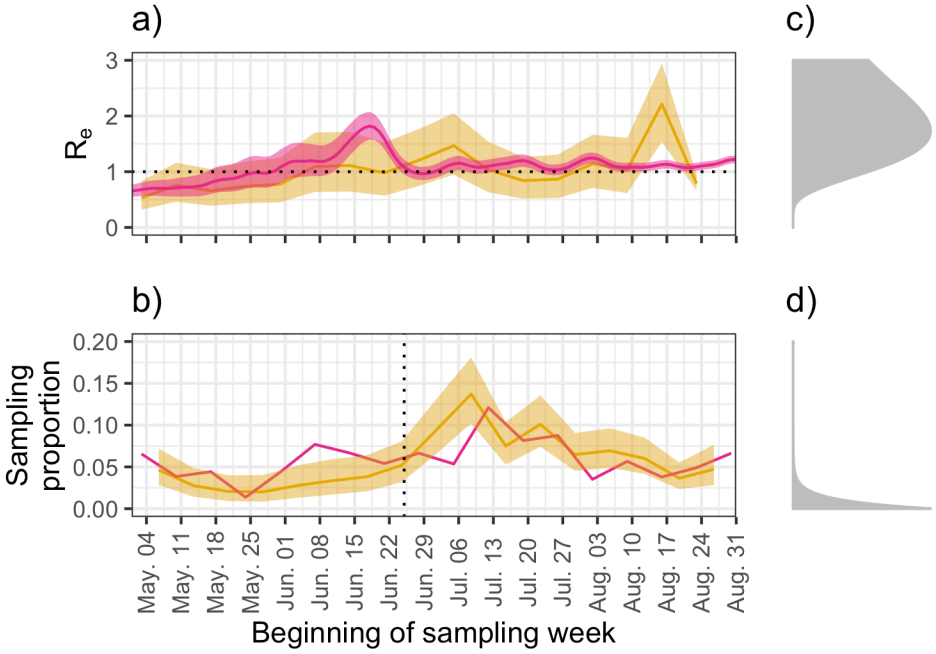

Figure 4. (a) Estimates of the effective reproductive number (R_e) and **(b)** the sampling proportion in Switzerland through time based on genome sequences from our nation-wide sequencing project (yellow) compared with estimates from line-line data (pink, (*2,31*)). **(c, d)** The prior distributions applied to R_e and the sampling proportion. The week-to-week smoothing prior applied to both parameters (see Supplementary Text) is not shown.