Quantification of the spread of SARS-CoV-2 variant B.1.1.7 in Switzerland

Chaoran Chen[1,2], Sarah Nadeau[1,2,*], Ivan Topolsky[1,2,*], Marc Manceau[1,2,*], Jana S. Huisman[1,2,3*], Kim Philipp Jablonski[1,2], Lara Fuhrmann[1,2], David Dreifuss[1,2], Katharina Jahn[1,2], Christiane Beckmann[4], Maurice Redondo[4], Olivier Kobel[4], Christoph Noppen[4], Lorenz Risch[5], Martin Risch[5], Nadia Wohlwend[5], Sinem Kas[5], Thomas Bodmer[5] Tim Roloff[2,6,7], Madlen Stange[2,6,7],Adrian Egli[6,7], Isabella Eckerle[8,15], Laurent Kaiser[8,16,17], Rebecca Denes[9], Mirjam Feldkamp[9], Ina Nissen[9], Natascha Santacroce[9], Elodie Burcklen[9], Catharine Aquino[10], Andreia Cabral de Gouvea[10], Maria Domenica Moccia[10], Simon Grüter[10], Timothy Sykes[10], Lennart Opitz[10], Griffin White[10],Laura Neff[10], Doris Popovic[10], Andrea Patrignani[10], Jay Tracy[10], Ralph Schlapbach[10], Emmanouil T. Dermitzakis[11,12], Keith Harshman[11], Ioannis Xenarios[11], Henri Pegeot[11], Lorenzo Cerutti[11], Deborah Penet[11], Anthony Blin[11], Melyssa Elies[11], Christian Althaus[13], Christian Beisel[1,&], Niko Beerenwinkel[1,2,&], Martin Ackermann[3,14,&], Tanja Stadler[1,2,+]

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

[2] Swiss Institute of Bioinformatics, Switzerland

[3] Department of Environmental Systems Science, ETH Zürich, Swiss Federal Institute of Technology, Zurich, Switzerland

[4] Viollier AG, Allschwil, Switzerland

[5] Dr Risch, Labormedizinisches Zentrum, Switzerland

[6] Clinical Bacteriology and Mycology, University Hospital Basel, Basel, Switzerland

[7] Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland

[8] Geneva Center for Emerging Viral Diseases and Laboratory of Virology, Geneva University Hospitals, Geneva, Switzerland

[9] Genomic Facility Basel, Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

[10] Functional Genomics Center Zurich, ETH Zürich and University of Zurich, Zurich, Switzerland

[11] Health 2030 Genome Center, Geneva, Switzerland

[12] University of Geneva Medical School, Geneva, Switzerland

[13] Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

[14] Department of Environmental Microbiology, Eawag, Dubendorf, Switzerland

[15] Department of Microbiology and Molecular Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[16] Division of Infectious Diseases, Geneva University Hospitals, Geneva, Switzerland

[17] Department of Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[*,&] These authors contributed equally to this work

[+] Correpondence: tanja.stadler@bsse.ethz.ch

April 8, 2021

Keywords: Pandemic, SARS-CoV-2, COVID-19, B.1.1.7, transmission advantage

Abstract

Background. In December 2020, the United Kingdom (UK) reported a SARS-CoV-2 Variant of Concern (VoC) which is now coined B.1.1.7. Based on the UK data and later additional data from other countries, a transmission advantage of around 40-80% was estimated for this variant [1, 2, 3].

Aim. The goal of this study is to estimate the transmission fitness advantage and the effective reproductive number of B.1.1.7 through time for data from Switzerland.

Methods. We collected genomic surveillance data from 11.8% of all SARS-CoV-2 confirmed cases across Switzerland between 14.12.2020 and 11.03.2021. It allows us to determine the relative proportion of the B.1.1.7 variant on a daily basis and to quantify the transmission advantage of the B.1.1.7 variant on a national and a regional scale.

Results. We propose a transmission advantage of 43-52% of B.1.1.7 compared to the other circulating variants. Further, we estimate a reproductive number for B.1.1.7 above 1 for Jan. 1, 2021 until now while the reproductive number for the other variants was below 1. In particular, for the time period up to Jan. 17 we obtain a reproductive number of 1.24 [1.07-1.41] and from Jan. 18 until March 1 we obtain 1.18 [1.06-1.30] based on the whole genome sequencing data. For March 10-16, we obtain 1.14 [1.00-1.26] based on all confirmed cases among which B.1.1.7 is dominant at this stage. Switzerland tightened measures on 18.01.2021 and released measures on 01.03.2021.

Conclusion. In summary, the dynamics of increase in the frequency of B.1.1.7 is as expected based on the observations in the UK. B.1.1.7 increased in absolute numbers exponentially with the point estimate for the doubling time being around 2-3.5 weeks. Our plots are available online and are currently regularly updated with new data to closely monitor the spread of B.1.1.7.

1 Introduction

Reports of an increased transmissibility of SARS-CoV-2 variant B.1.1.7 (501Y.V1) were released in mid-December 2020 [4, 5, 6]. This variant carries the N501Y mutation in the spike protein which may increase the ACE2 receptor affinity [7]. Within only a few months, the variant was able to become the dominant lineage in the UK.

Since these first reports, great efforts were made in Switzerland to detect and trace B.1.1.7 [8]. The first cases of B.1.1.7 were confirmed on 24.12.2020 and retrospective analyses identified B.1.1.7 in samples dating back to October [8]. In total, 1370 infections with B.1.1.7 were confirmed up to 05.02.2021 [8].

An increase in prevalence does not necessarily imply a transmission fitness advantage of the virus. For example, a variant termed 20A.EU1 spread rapidly during summer 2020 across Europe. However, data suggests that extended travel and superspreading events rather than a viral transmission advantage caused that spread [9].

However, the variant B.1.1.7 rapidly increased in frequency in many high-prevalence regions across the UK. Such re-occurring patterns are hard to explain without a transmission advantage. Analyses of the B.1.1.7 variant in the UK suggest that the variant has a transmission fitness advantage anywhere between 40 and 80% [1, 2, 3]. Davies et al. [3] further use data from Denmark obtaining again similar estimates. Quantitative analysis of data on the spread of the 501Y mutation in Switzerland also suggests a transmission advantage in roughly that range [10].

In this paper, we collected whole-genome sequencing data based on samples from Viollier AG to determine the proportion of B.1.1.7 through time. Further, we used data from the diagnostic laboratory Dr Risch which screens all their samples for the B.1.1.7 variant. Finally, all patients who tested positive with CT values below 32 at the University Hospital Geneva (HUG) were sequenced starting 23.12.2021 to provide a detailed view on the situation in the Lake Geneva region. We then used this data to quantify the transmission fitness advantage of B.1.1.7 for Switzerland as well as for the seven Swiss economic regions (Grossregionen). We further calculated the reproductive number for B.1.1.7 and show how the B.1.1.7 case numbers developed through time.

The core plots presented here are currently updated regularly on [11] and part of it is currently incorporated into the Swiss National COVID-19 Science Task Force website [12]. The code and data

used for this paper are publically available on Github [13]. All sequences are available on GISAID.

2 Methods

2.1 Data

The main source of data is the whole-genome sequences based on samples from Viollier AG – a large diagnostic lab processing SARS-CoV-2 samples from across Switzerland. Every week, SARS-CoV-2 samples were randomly selected for sequencing among all positively tested samples at the Viollier AG. For each sample, we know the date of the test and the canton in which the test was performed. The sequencing protocols and bioinformatics procedures are described in the supplementary materials (section A.1).

In the time frame of interest – from 14.12.2020 to 11.03.2021, a total of 9772 sequences were generated and checked for the B.1.1.7 variant. Thus, per week we produced around 780 sequences. In the same time period, 184'165 people were confirmed with a SARS-CoV-2 infection, thus we provide sequences for 5.3% of all cases. We define a sequence to be a B.1.1.7 sample if at least 80% of the lineage-defining, non-synonymous nucleotide changes according to [4] are present.

Additionally, we have data from the Dr Risch medical laboratories. In our considered time period, Dr Risch screened 12019 samples for B.1.1.7. Details about the procedure are described in the supplementary materials section A.2.

Taking both datasets together, 11.8% of all SARS-CoV-2 cases across Switzerland during our considered time period (14.12.2020 to 11.03.2021) were screened for the B.1.1.7 variant. While Viollier processes samples from all over Switzerland, the intensity varies across regions. The set of sequenced samples inherits this uneven geographical distribution with the result that, relatively to the number of all cases, over eight times more samples were sequenced from the region Nordwestschweiz than from the region Ticino (Table S1). Dr Risch's data has a different geographic distribution than Viollier and, for example, much better coverage of Eastern Switzerland (Table S2). In summary, both datasets differ in their geographic biases.

In order to obtain a detailed view on the Lage Geneva region, we additionally obtained a dataset from the University Hospital Geneva (HUG). All patients who tested positive there were sent for whole genome sequencing (section A.1). Between 23.12.2020 and 04.03.2021, HUG sequenced 2074

samples from Lake Geneva region which cover 7% of all cases in that timeframe.

In what follows, the analyzed data is the number of sequences per day and the number of identified B.1.1.7 variants per day. We present results on the different datasets and compare these results.

2.2 Statistical inference

We fit a logistic model to the data to estimate the logistic growth rate a and the sigmoid's midpoint t_0. From that, we derive an estimate of the transmission fitness advantage under a continuous (f_c) and discrete (f_d) model which we consider as two extremes with the actual dynamics being described by a process in between. Further, we estimate the reproductive number R for the B.1.1.7 and non-B.1.1.7 cases. The mathematical derivations are described in the supplemantary materials in the sections A.3 and A.4.

We finally display the expected number of confirmed cases in the future under the continuous model. We initialize the model on 01.01.2021 with the estimated number of B.1.1.7 and non-B.1.1.7 cases on that day. We assume a reproductive number for the non-B.1.1.7 as estimated on the national level for 01.01.2021-17.01.2021. Further, we assume that the expected generation time is 4.8 days and the fitness advantage is the estimated f_c for the region and dataset of interest (Table 1).

3 Results

We estimate the logistic growth rate a and the sigmoid's midpoint t_0 based on the Viollier and Risch data (Table 1). Taking the estimates of both datasets together, we obtain a growth rate a of 0.07-0.09 per day for Switzerland. For each economic region, the estimated uncertainty interval of a overlaps with the Swiss-wide uncertainty. We have little data for two out of seven regions (Ticino and Central Switzerland; <1100 sequences in total) resulting in very wide uncertainty intervals. From the t_0 estimates, we observe that the Geneva region were about 2 weeks ahead of the rest of Switzerland with respect to the spread of B.1.1.7, which confirms estimates from [10]. Our initial analyses of the data in January projected that B.1.1.7 will become dominant in Switzerland in March 2021. Indeed, our latest data points suggest a frequency of B.1.1.7 in confirmed cases of around 80% for March 11.

In Fig. 1 and 2, we graphically illustrate the logistic growth in frequency of B.1.1.7 and show the daily data together with an estimate of the proportion of B.1.1.7 under the logistic growth model.

As a validation, we additionally analyzed HUG data. The estimates for the Lake Geneva region based on Viollier data agree very well with independent estimates based on HUG data (Table 1 and Figure S2).

Next, we estimate the reproductive number for B.1.1.7 and non-B.1.1.7 on a national scale (Fig. 3). We note that measures were strengthened on January 18, 2021, and then relaxed on March 1, 2021. During 01.01.2021-17.01.2021, the reproductive number for B.1.1.7 was significantly above 1 (Viollier: 1.24 [1.07-1.41], Risch: 1.46 [1.21-1.72]) while the reproductive number of non-B.1.1.7 was below 1 (Viollier: 0.83 [0.65-1.00] , Risch: 0.81 [0.67-0.96]).

Based on the Viollier data, the reproductive number for B.1.1.7 did not change drastically after 18.01.2021 (1.18 [1.06-1.30]) and agrees well with the Risch estimate (1.15 [1.01,1.29]). We note though that the doubling time for a reproductive number of 1.24 is 15 days while it is 20 days for a reproductive number of 1.18 which makes a substantial difference in the overall dynamics. The reproductive number for non-B.1.1.7 did not change much either after 18.01.2021 (Viollier: 0.80 [0.68-0.91], Risch 0.85 [0.76,0.93]).

As expected, the ratio of the reproductive numbers for B.1.1.7 and non-B.1.1.7 for the Viollier data stays roughly constant throughout January and February. For the Risch data, the ratio of the reproductive numbers for B.1.1.7 and non-B.1.1.7 dropped in January. This is not expected under our models. We suspect the reason is a bias in data (such as preferential inclusion of B.1.1.7 in early January) as all estimates but the Risch estimate for B.1.1.7 in early January agree.

Our data does not allow us to estimate the variant-specific reproductive number for March yet. However, for the time period March 10-16, we can estimate the reproductive number based on all confirmed cases [14] (1.14 [1.00-1.26]). Since this estimate is based on confirmed cases from the second part of March when we project around 90-95% of all confirmed cases being B.1.1.7 infections, this estimate is only slightly underestimating the reproductive number of B.1.1.7. In summary, the reproductive number for B.1.1.7 was above 1 since January while the reproductive number for the other variants was below 1.

Next, we use the average reproductive number during 01.01.2021-17.01.2021 to calculate the transmission fitness advantage f_c of B.1.1.7 under our continuous-time model. Further, we calculate the transmission fitness advantage f_d under a discrete-time model. In both cases, we assume a generation time g of 4.8 days (which is the mean generation time assumed in the estimation of the reproductive

number). The fitness values are shown in Table 1. On a national level, we estimate a fitness advantage of 43-52% across methods and datasets. The regional estimates have an overlap with this interval. We note that we use the national reproductive number for the regional f_c estimate. Since Ticino had a lower and the Lake Geneva region a higher reproductive number averaged over all variants [14], the f_c for Ticino may be an underestimate, and the f_c for Lake Geneva an overestimate.

Next, we show the expected dynamics of the epidemic under the continuous model using parameter values based on epidemic conditions in the first half of January (Fig. 4 and 5). We show the development of confirmed case numbers based on this model in the blue (B.1.1.7) and green (non-B.1.1.7) areas. In particular, in January, the model projects a decline in overall case numbers due to a decline in non-B.1.1.7 variants. However, the B.1.1.7 variant is increasing. Under this model, once B.1.1.7 becomes dominant, the total number of cases will be increasing again. It is important to note that this simple model is intended to highlight if dynamics changed compared to early January 2021, and thus it assumes that the transmission dynamics did not change since then. In particular, the model does not include effects of tightening of measures on 18.01.2021, relaxing measures on 01.03.2021, vaccination, immunity after infection, and population heterogeneity.

We investigate if our empirical data follows this trend (lines in Fig. 5). Throughout January, the model and the empirical data follow the same trends across datasets and regions with the exception of Ticino and the Lage Geneva region (see supplementary materials, section A.5). Based on the Viollier data, for the national level and the five regions with a good match in January, we observe that the empirical total case numbers are below the model projections starting early February. This is in line with a slight reduction of the reproductive number after 18.01.2021.

The Risch data follows the model very well until March. Thus, changing the model from the Viollier parameters (R-value 0.83, fitness advantage 0.48) to the Risch parameters (R-value 0.81, fitness advantage 0.45) leads to a very good fit until mid-March highlighting that in a phase of overall exponential growth, slight changes of the parameters can have large consequences on the total number of cases. In other words, a slight reduction of transmission in late January and / or a slight misspecification of the Viollier parameters can explain the recent mismatch of the total number of cases and the Viollier projections.

Region	a	t_0	$f_d = exp(ag) - 1$	$f_c = ag/R_c$
Switzerland (Viollier)	0.08 [0.08; 0.09]	65 [64; 66]	0.49 [0.47; 0.52]	0.48 [0.46; 0.50]
Switzerland (Risch)	0.08 [0.07; 0.08]	69 [68; 70]	0.45 [0.43; 0.48]	0.45 [0.43; 0.47]
Central Switzerland	0.09 [0.07; 0.11]	73 [67; 79]	0.51 [0.37; 0.66]	0.50 [0.38; 0.62]
Espace Mittelland	0.07 [0.07; 0.08]	64 [62; 66]	0.41 [0.37; 0.45]	0.41 [0.38; 0.44]
Grossregion Nordwestschweiz	0.09 [0.08; 0.10]	67 [65; 68]	0.53 [0.48; 0.59]	0.52 [0.47; 0.56]
Grossregion Tessin	0.07 [0.04; 0.11]	77 [69; 86]	0.42 [0.20; 0.64]	0.43 [0.24; 0.61]
Grossregion Zurich	0.09 [0.08; 0.10]	68 [65; 70]	0.55 [0.47; 0.62]	0.53 [0.47; 0.58]
Lake Geneva region	0.10 [0.09; 0.11]	54 [52; 56]	0.61 [0.51; 0.71]	0.57 [0.50; 0.65]
Lake Geneva region (HUG)	0.09 [0.09; 0.10]	52 [51; 53]	0.56 [0.50; 0.62]	0.54 [0.49; 0.59]
Ostschweiz	0.10 [0.08; 0.11]	65 [62; 68]	0.58 [0.46; 0.69]	0.55 [0.46; 0.64]

Table 1: Estimates of the growth rate a and the sigmoid's midpoint t_0 (measured in days after Dec. 14)as well as the transmission fitness advantages f_d and f_c are reported. In the f_c calculation, the Swiss-wide estimate of the reproductive number for the time period 01.01.2021-17.01.2021 is assumed for the R_c. Data stem from the Viollier dataset unless stated otherwise.

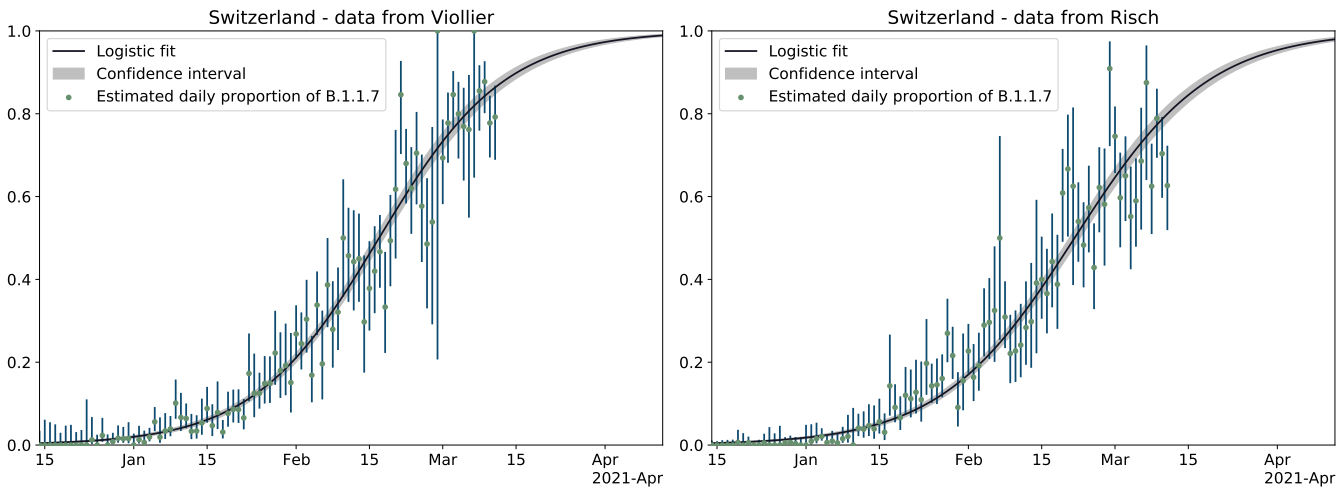

Figure 1: Logistic growth of frequency of B.1.1.7 in Switzerland. Green points are the empirical proportions of B.1.1.7 for each day (i.e. number of B.1.1.7 samples divided by the total number of samples). Blue vertical lines are the estimated 95 % uncertainty of this proportion for each day, assuming a simple binomial sampling and Wilson uncertainty intervals. A logistic growth function fit to the data from all of Switzerland is shown in black with the 95 % uncertainty interval of the proportions in gray (i.e. $p(t)$ from Eqn. 2 and 5 in the supplementary material section A.3).

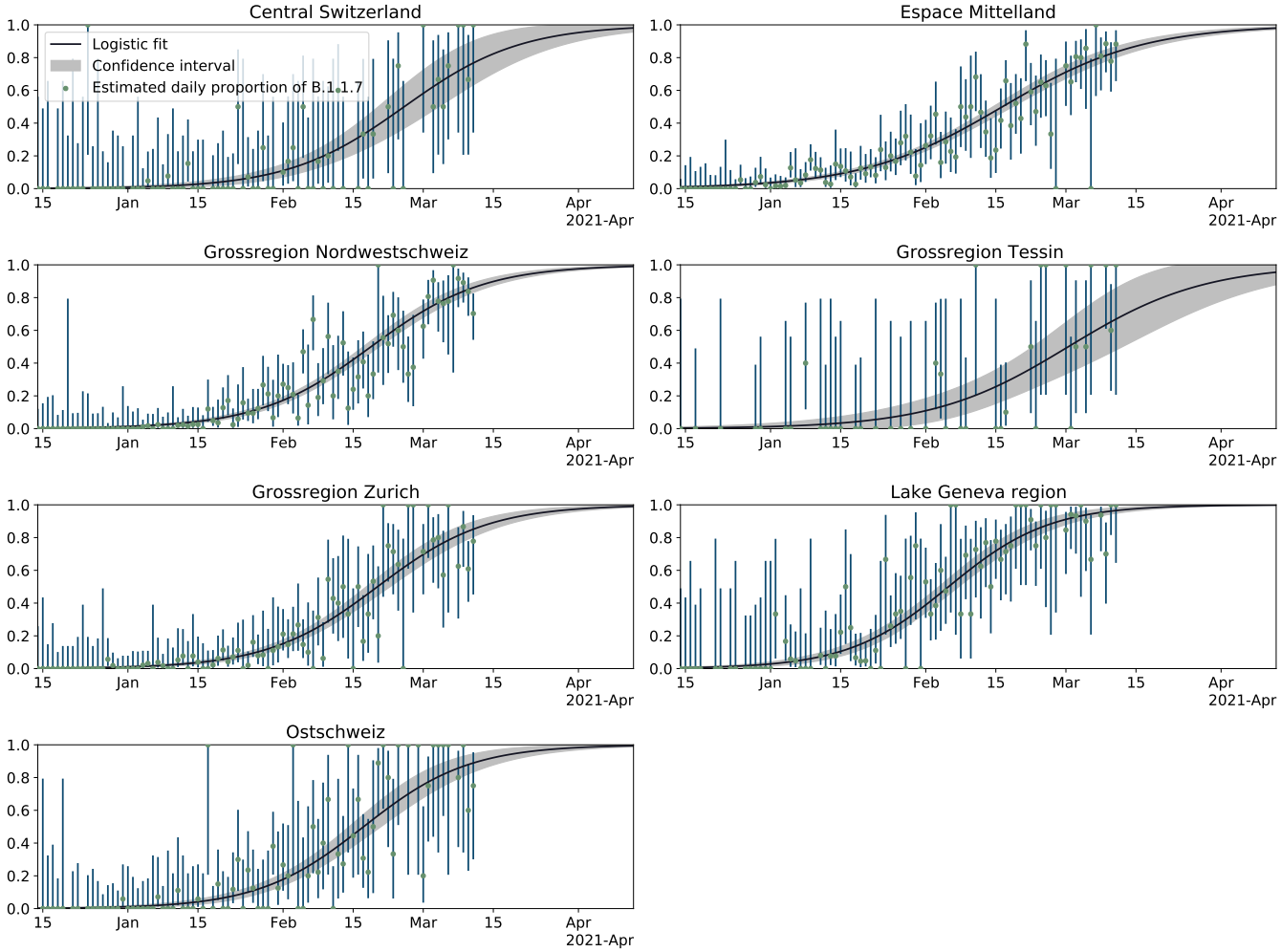

Figure 2: Logistic growth of frequency of B.1.1.7 in the seven economic regions of Switzerland. For details see legend of Fig. 1.

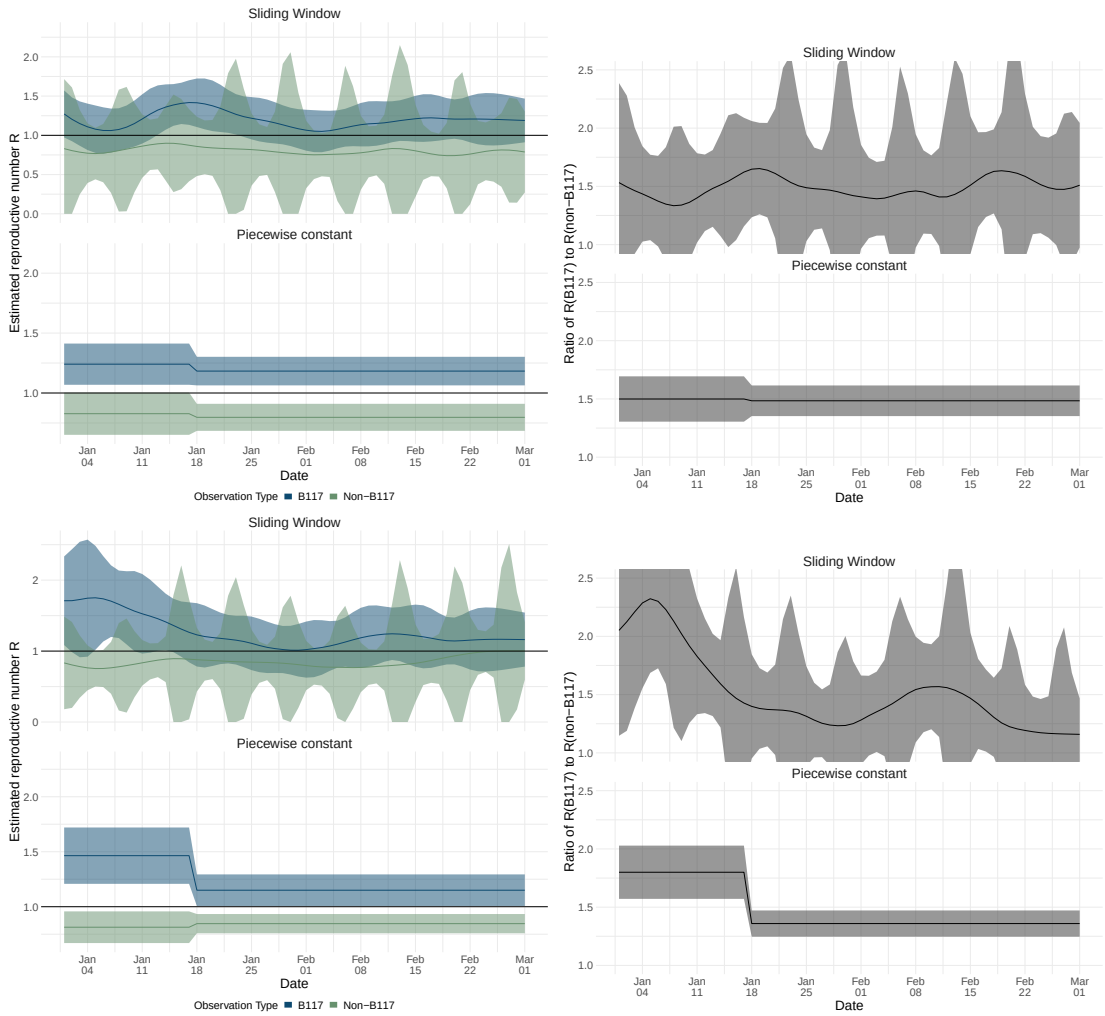

Figure 3: Estimates of the effective reproductive number R of the B.1.1.7 variant and non-B.1.1.7 variants. Results in the top row are based on Viollier data, and in the bottom row based on Risch data. Within each panel, the top row shows the results of the continuously varying R estimation, and the bottom of the piecewise constant R estimation. The left column shows the R estimates, whereas the right shows the ratio between R estimated for B.1.1.7 and R estimated for all non-B.1.1.7 variants. The confidence intervals for the R of non-B117 variants show a 7-day periodicity due to lower case reporting on weekends.

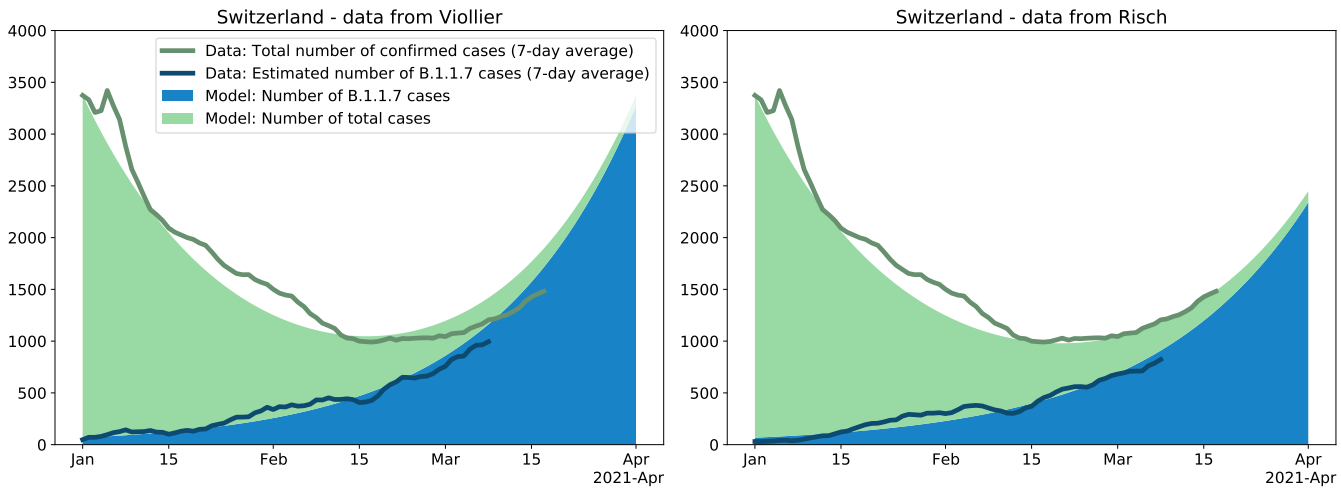

Figure 4: Change in the number of B.1.1.7 variants and in the number of all cases through time for Switzerland. Based on the average reproductive number R_c for Switzerland estimated for the time period 01.01.2021-17.01.2021 (i.e. prior to the tightening of measures on 18.01.2021) and the transmission fitness advantage f_c for the same time period, we plot the expected number of B.1.1.7 variants (blue) and the expected number of non-B.1.1.7 variants (green) under the continuous model. The model is initialized on Jan. 1 with the total number of cases and the estimated number of B.1.1.7 cases on that day. This model is compared to data: The dark green line is the total number of confirmed cases (7-day average). The dark blue line is the estimated number of confirmed B.1.1.7 cases (7-day average); this number is the product of the total number of confirmed cases for a day by the proportion of the B.1.1.7 variant for that day. If the empirical data develops as the model, the dark blue line follows the upper end of the blue area and the dark green line follows the upper end of the green area.

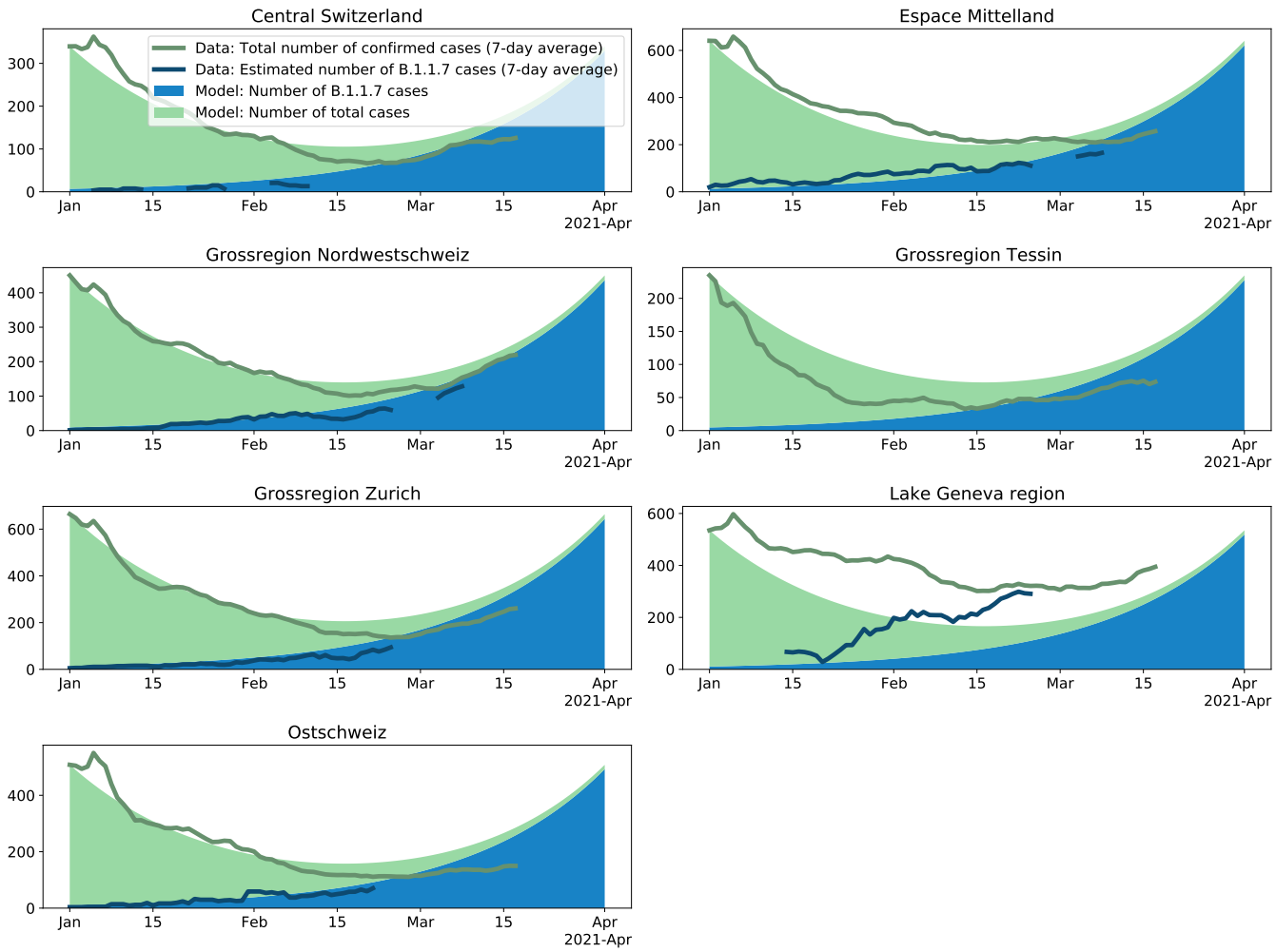

Figure 5: Change in the number of B.1.1.7 variants and in the number of all cases through time for the seven Swiss economic regions. For details see legend of Fig. 4. We use the reproductive number estimated for the whole of Switzerland for the continuous-time model such that we can compare to what extend regions differ from the national dynamic. The regional transmission fitness advantage is taken from Table 1.

4 Discussion

We have quantified the transmission fitness advantage and the effective reproductive number of B.1.1.7 for two Swiss-wide datasets. One dataset also allows us to obtain estimates for the seven Swiss economic regions. Swiss-wide estimates point towards a transmission fitness advantage of 43-52%. Based on our early sequencing data, we already projected in January that B.1.1.7 will become dominant in March. The same conclusion was reached by [10] in January based on Swiss data tracking 501Y mutations. On 11.03.2021 indeed around 80% of analyzed confirmed cases carry the B.1.1.7 variant. Since cases are confirmed 8-11 days after the time of infection in Switzerland, actually almost all new infections may already be caused by B.1.1.7 as of today, March 26. Thus, the Swiss epidemic is now a B.1.1.7 epidemic.

The increase in the frequency of B.1.1.7 in Switzerland occurred as expected given the large transmission advantage. Unless measures such as contact tracing target in particular B.1.1.7, this increase in relative frequency will occur given a transmission advantage. However, variables such as implemented measures, the adherence to measures, or levels of immunity in the population determine how fast the variants increase in absolute numbers.

We show that in the first half of January, the absolute numbers of B.1.1.7 increased (R-value 1.24 [1.07-1.41] for 01.01.2021-17.01.2021) based on our Swiss-wide sequence dataset from Viollier) while the absolute numbers of all other cases decreased (R-value 0.83 [0.65-1.00]). For the second half of January, the reproductive number for B.1.1.7 and non-B.1.1.7 only decreased slightly according to the Viollier data. The Risch estimates agree very well with this pattern with the exception of B.1.1.7 in early January where a higher reproductive number is estimated. We speculate that the large drop in the reproductive number for B.1.1.7 based on the Risch data is due to a bias in the B.1.1.7 data, possibly early data contained samples preferentially stemming from B.1.1.7 infections.

The reproductive number averaged over all confirmed cases in Switzerland (among which B.1.1.7 is projected to cause 90-95% of all infections) for March 10-16 is estimated to be 1.14 [1.00-1.26]. Taken together, this highlights that B.1.1.7 was spreading exponentially since early January with the point estimate for the doubling time to be around 2-3.5 weeks.

When looking at the Viollier data, we observe that the total number of confirmed cases in February is lower compared to our model scenario using parameter estimates from the first half of January.

This may reflect an overall small slow-down in transmission as also suggested in the reproductive number. However, as observed from the reproductive number calculation, the slowdown was not large and uncertainty intervals are wide. The Risch data agrees very well with the model. We suggest that either a slight reduction of transmission in late January and / or a slight misspecification of the Viollier parameters in the model can explain the recent mismatch of the total number of cases and the Viollier projections (assumed parameters are: reproductive number of non-B.1.1.7 for Viollier is 0.83 and for Risch is 0.81; transmission advantage for Viollier is 0.48 and for Risch is 0.45.) This highlights that the 11.5% of characterized confirmed cases in our dataset is not sufficient to quantify the change in dynamics precisely enough to evaluate the effect of the strengthening of measures on the reproductive number.

Our data reveal that different regions from Switzerland were sampled with different intensities. Thus, we also performed analyses for the seven economic regions in Switzerland. Here, we expect that the homogeneous sampling intensity is met better compared to the national level. We may observe cases though which are imports from other economic regions compared to local transmission (the same problem of course exists on a national level, though the smaller the scale, the more imports we expect). However, the economic regions represent well-defined regions where we expect a lot of mixing within and less mixing across regions. The estimates of the transmission fitness advantage for the economic regions are largely in line with the national estimates, though of course with larger uncertainty. Two of the regions (Ticino, Central Switzerland) have too little data to make precise statements. Geneva is estimated to be around 2 weeks ahead in the B.1.1.7 dynamics compared to Switzerland as a whole. One explanation may be the large number of UK travelers in ski resorts in the Valais in December 2020.

Overall we see a consistent signal for a large transmission advantage of B.1.1.7 in Switzerland. This confirms estimates from studies from the UK [1, 2, 3], Denmark [3], and Switzerland ([10]; looking at 501Y mutations). The strengthening of measures did not result in stopping the spread of B.1.1.7. This resulted in overall exponentially growing case numbers once B.1.1.7 became dominant in March. Our core plots are available on [11] and [12] for real-time monitoring. Further, we will add new Variants of Concern to our webpage once they spread in Switzerland.

Contribution

CC: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing - original draft; SN: Data curation, Writing - review & editing; IT: Data Curation, Resources, Software, Validation; MM: Formal Analysis, Writing - review & editing; JSH: Formal Analysis, Writing - review & editing; KPJ: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - review & editing; LF: Software; DD: Data curation, Software, Validation; KJ: Formal analysis, Validation; CA: Methodology, Writing - review & editing; NB: Conceptualization, Funding acquisition, Project administration, Supervision, Writing - review & editing; MA: Conceptualization, Investigation, Writing - review & editing; TS: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing - original draft; Everyone else: Resources, Writing - review & editing.

Acknowledgement

We thank Richard Neher for useful discussions on the maximum likelihood inference. TS acknowledges funding from the Swiss National Science foundation (Special Call on Coronaviruses; 31CA30_196267 and 31CA30_196348). CA received funding from the European Union Horizon 2020 research and innovation programme - project EpiPose (No 101003688).

References

[1] Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from Linking Epidemiological and Genetic Data. medRxiv. 2021 Jan:2020.12.30.20249034.

[2] Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Eurosurveillance. 2021;26(1):2002106.

[3] Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday J, et al. Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. MedRxiv. 2021:2020–12.

[4] Andrew Rambaut, Nick Loman, Oliver Pybus, Wendy Barclay, Jeff Barrett, Alesandro Carabelli, et al. Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations; 2020.

[5] NERVTAG meeting on SARS-CoV-2 variant under investigation VUI-202012/01;. Available at https://app.box.com/s/3lkcbxepqixkg4mv640dpvvg978ixjtf/file/756963730457. Available from: https://app.box.com/s/3lkcbxepqixkg4mv640dpvvg978ixjtf/file/756963730457.

[6] Public Health England - Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01;. Available at https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201. Available from: https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201.

[7] Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell. 2020 Sep;182(5):1295–1310.e20.

[8] Cabecinhas ARG, Roloff T, Stange M, Bertelli C, Huber M, Ramette A, et al. SARS-CoV-2 N501Y Introductions and Transmissions in Switzerland from Beginning of October 2020 to February 2021 - Implementation of Swiss-Wide Diagnostic Screening and Whole Genome Sequencing. medRxiv. 2021 Feb:2021.02.11.21251589.

[9] Hodcroft EB, Zuber M, Nadeau S, Comas I, Candelas FG, Consortium SS, et al. Emergence and Spread of a SARS-CoV-2 Variant through Europe in the Summer of 2020. medRxiv. 2020 Oct:2020.10.25.20219063.

[10] Transmission of SARS-CoV-2 variants in Switzerland;. Available at https://ispmbern.github.io/covid-19/variants/. Available from: https://ispmbern.github.io/covid-19/variants/.

[11] SARS-CoV-2 Variants of Concern in Switzerland;. Available at https://cevo-public.github.io/Quantification-of-the-spread-of-a-SARS-CoV-2-variant/. Available from: https://cevo-public.github.io/Quantification-of-the-spread-of-a-SARS-CoV-2-variant/.

[12] SARS-CoV-2 Variants of Concern in Switzerland;. Available at https://sciencetaskforce.ch/nextstrain-phylogentische-analysen/. Available from: https://sciencetaskforce.ch/nextstrain-phylogentische-analysen/.

[13] github repository for "Quantification of the spread of SARS-CoV-2 variant B.1.1.7 in Switzerland"; Chen et al. 2021;. Https://github.com/cevo-public/Quantification-of-the-spread-of-a-SARS-CoV-2-variant. Available from: https://github.com/cevo-public/Quantification-of-the-spread-of-a-SARS-CoV-2-variant.

[14] Real-time estimates of the reproductive number for SARS-CoV-2;. Available at https://ibz-shiny.ethz.ch/covid-19-re-international/. Available from: https://ibz-shiny.ethz.ch/covid-19-re-international/.

[15] Nadeau SA, Vaughan TG, Sciré J, Huisman JS, Stadler T. The origin and early spread of SARS-CoV-2 in Europe. Proceedings of the National Academy of Sciences. 2021;118(9).

[16] Illumina COVIDSeq Test;. Available at https://emea.illumina.com/products/by-type/ivd-products/covidseq.html. Available from: https://emea.illumina.com/products/by-type/ivd-products/covidseq.html.

[17] ARTIC v3 multiplex PCR amplicon protocol;. Available at https://artic.network/. Available from: https://artic.network/.

[18] Health 2030 Genome Center github;. Available at https://github.com/health2030genomecenter. Available from: https://github.com/health2030genomecenter.

[19] Posada-Céspedes S, Seifert D, Topolsky I, Jablonski KP, Metzner KJ, Beerenwinkel N. V-pipe: a computational pipeline for assessing viral genetic diversity from high-throughput data. Bioinformatics. 2021 01. Btab015. Available from: https://doi.org/10.1093/bioinformatics/btab015.

[20] Stange M, Mari A, Roloff T, Seth-Smith HM, Schweitzer M, Brunner M, et al. SARS-CoV-2 outbreak in a tri-national urban area is dominated by a B. 1 lineage variant linked to mass gathering events. medRxiv. 2020.

[21] Chevin LM. On measuring selection in experimental evolution. Biology letters. 2011;7(2):210–213.

[22] Seabold S, Perktold J. statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference; 2010. .

[23] Huisman JS, Scire J, Angst DC, Neher RA, Bonhoeffer S, Stadler T. Estimation and Worldwide Monitoring of the Effective Reproductive Number of SARS-CoV-2. medRxiv. 2020 Nov:2020.11.26.20239368.

[24] Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. American Journal of Epidemiology. 2013 Nov;178(9):1505–1512.

[25] Coronavirus: Federal Council extends and tightens measures;. Available at `https://www.admin.ch/gov/en/start/documentation/media-releases/media-releases-federal-council.msg-id-81967.html`. Available from: `https://www.admin.ch/gov/en/start/documentation/media-releases/media-releases-federal-council.msg-id-81967.html`.

A Supplementary Material

A.1 Sequencing Protocols for Viollier samples and HUG samples

We perform whole-genome sequencing of the Viollier samples in three facilities. The HUG samples are processed in one of them, the Health 2030 Genome Center. The sequencing protocol for the samples sequenced at the Genomics Facility Basel and the Functional Genomics Center Zurich is described in [15]. Samples processed at the Health 2030 Genome Center used the Illumina COVIDSeq library preparation reagents following the protocol provided by the supplier [16]. These reagents are based on the ARTIC v3 multiplex PCR amplicon protocol [17]. When sufficient volume was available, 8.5ul of RNA extracted from patient nasopharyngeal samples were used in the cDNA synthesis step; if

8.5ul were not available, the maximum volume possible was used. Pooled libraries were sequenced on the Illumina NovaSeq 6000 using a 50-nucleotide pair-end run configuration. Post-sequencing library read de-multiplexing was done using an in-house developed processing pipeline [18]. The downstream bioinformatics procedure to obtain consensus sequences is described in [15, 19].

A.2 Screening procedure at Dr Risch

Dr Risch medical laboratories used the Taqpath assay from Thermofisher for their diagnostic and recorded the S gene target failure (SGTF). SGTF samples are potential B.1.1.7 variants, as the B.1.1.7 variant causes a SGTF due to its deletion at positions 69-70 in the spike protein. Further, Dr Risch medical laboratories screen their samples for the 501Y mutation by a variant-specific PCR test. If a sample is identified as a potential VoC by these procedures, it was initially sent for whole-genome sequencing to the University Hospital Basel in order to confirm the B.1.1.7 variant. The sequencing protocol is described in [20]. For recent samples, the confirmation may still be outstanding or not conducted due to B.1.1.7 now being dominant. However, since typically a SGTF plus a 501Y mutation corresponds indeed to a B.1.1.7 variant, we consider these samples as B.1.1.7 variants even when whole genome sequencing confirmation is lacking.

A.3 Estimating the transmission fitness advantage of a new variant

In what follows, we define two models describing the dynamics with which a new variant with a transmission fitness advantage spreads in a population. The first model is based on the assumption of discrete-time, while the second model is based on the assumption of continuous-time. Both models have been considered extensively in the literature to estimate fitness advantage (see e.g. [21, 3]). While in an epidemic, a generation does not end after a fixed time span (discrete-time model), generations are typically less variable than modeled under an exponential distribution (continuous-time model). Thus we view the two models as two extremes with the actual dynamics being described by a process in between. We provide estimates based on both models and suggest that the true parameter may be anywhere within the ranges spanned by the two models. In the next sections, we provide details of how we estimate the transmission fitness advantage of B.1.1.7 based on daily data of the total number of samples and B.1.1.7 samples under these two models.

A.3.1 Discrete time model

We call X the common (non-B.1.1.7) variants and Y the B.1.1.7 variant. The process starts in generation 0 with x_0 cases caused by variant X and y_0 cases caused by variant Y. Let the number of cases in generation n be $x(n)$ and $y(n)$ for variants X and Y.

Let the reproductive number R_d of variant X in generation n be $R_d(n)$. Let the transmission advantage of variant Y be f_d. Then the reproductive number of Y in generation n is $(1+f_d)R_d(n)$. Thus, we assume a multiplicative fitness advantage.

We have $x(n) = x(0) \times R_d(0)R_d(1)\ldots R_d(n-1)$ and $y(n) = y(0) \times R_d(0)R_d(1)\ldots R_d(n-1)(1+f_d)^n$. If R_d is constant through time, we have

$$x(n) = x(0)R_d^n \text{ and } y(n) = y(0)((1+f_d)R_d)^n.$$

Let the proportion of variant Y at generation n be $p(n)$. We have,

$$\begin{aligned} p(n) &= \frac{y(n)}{x(n)+y(n)} = \frac{y(0)(1+f)^n}{x(0)+y(0)(1+f)^n} \\ &= \frac{p(0)(1+f_d)^n}{1-p(0)+p(0)(1+f_d)^n} \\ &= \frac{1}{1+(1+f_d)^{-n}\left(p(0)^{-1}-1\right)}. \end{aligned}$$

Thus, $p(n)$ is the logistic function. It does not depend on R_d.

If we write time in days t rather than generations n and assume a generation time of g days, we get

$$p(t) = \frac{1}{1+(1+f_d)^{-t/g}\left(p(0)^{-1}-1\right)}. \tag{1}$$

We now switch our parameterization to the more common

$$p(t) = \frac{1}{1+e^{-a(t-t_0)}} \tag{2}$$

for parameter estimation from daily data. Parameter a is the logistic growth rate and parameter t_0 the sigmoid's midpoint.

The two free parameters, a and t_0, are related to the two free parameters in Equation 1, f_d and $p(0)$, as follows:

$$a = \frac{\ln(1+f_d)}{g}, \quad t_0 = g\,\frac{\ln(p(0)^{-1}-1)}{\ln(1+f_d)}.$$

In particular, we get $f_d = e^{ag} - 1$.

A.3.2 Continuous time model

In continuous-time, instead of R_d and generation time g, we define the transmission rate β and the recovery rate μ. Under this model, the reproductive number is $R_c = \beta/\mu$. Further, since an individual in the discrete model recovers after a generation of duration g (during which they left R_d offspring), we note that g is related to the expected time to recovery $1/\mu$ in the continuous model, and in fact assume $g = 1/\mu$ in what follows. Again our initial numbers of the variants X and Y are $x(0)$ and $y(0)$. Calendar time is denoted by a continuous parameter t. We then have in expectation,

$$x(t) = x(0)e^{(\beta-\mu)t}. \tag{3}$$

We note that $\beta - \mu$ is coined the Malthusian growth parameter [3].

Further, we again assume that variant Y has a transmission fitness advantage of f_c, with transmission rate $\beta(1+f_c)$ and recovery rate μ. The population size of the variant at time t is thus

$$y(t) = y(0)e^{(\beta(1+f_c)-\mu)t}. \tag{4}$$

The proportion of the variant in the population at time t is

$$\begin{aligned} p(t) &= \frac{y(0)e^{(\beta(1+f_c)-\mu)t}}{x(0)e^{(\beta-\mu)t} + y(0)e^{(\beta(1+f_c)-\mu)t}} \\ &= \frac{p(0)e^{\beta f_c t}}{1 - p(0) + p(0)e^{\beta f_c t}} \\ &= \frac{1}{1 + e^{-\beta f_c t}(p(0)^{-1} - 1)} \end{aligned}$$

where we again recognize the logistic function. We turn again to the more common parameterization,

$$p(t) = \frac{1}{1 + e^{-a(t-t_0)}} \tag{5}$$

where we thus have $f_c = \frac{a}{\beta}$.

The reproductive number is $R_c = \beta/\mu$ and the mean time to recovery, $1/\mu$, is equaled to g. Then, $\beta = R_c/g$. Thus,

$$a = \frac{f_c R_c}{g}, \quad t_0 = g\frac{\ln(p(0)^{-1} - 1)}{f_c R_c}.$$

In particular, we have $f_c = \frac{ag}{R_c}$. Note that the estimated fitness advantage under this model depends on the reproductive number R_c and is thus changing if R_c is changing through time.

A.3.3 Connection between discrete and continuous time

The discrete and continuous models are very similar. Both have the intitial conditions $x(0)$ and $y(0)$. For the dynamics, the discrete model has parameters R_d and g while the continuous model has parameters β and μ. We have $R_c = \beta/\mu$ and we further assumed that $g = 1/\mu$ ($1/\mu$ is the expected time until recovery in the continuous setting while g is the time to recovery in the discrete setting). The different parameterizations of fitness advantage are coined f_c and f_d. We now determine how f_c and f_d are related.

To compare the two models, we now assume that their overall dynamics for the variant X are the same. After n generations of duration g, we have for variant X,

$$\begin{aligned} & x(0)e^{(\beta-\mu)ng} = x(0)R_d^n \\ \iff & e^{(\beta-\mu)g} = R_d \\ \iff & R_d = e^{R_c - 1} \end{aligned}$$

Using a Taylor expansion for β/μ close to 1, we obtain that indeed $R_d = R_c$.

For the two models to produce the same growth also for variant Y, we require,

$$
\begin{aligned}
& y_0 e^{(\beta(1+f_c)-\mu)ng} = y_0((1+f_d)R_d)^n \\
\Longleftrightarrow & (\beta(1+f_c)-\mu)g = \ln((1+f_d)R_d) \\
\Longleftrightarrow & (\beta-\mu)g + f_c\beta g = \ln(R_d) + \ln(1+f_d) \\
\Longleftrightarrow & f_c\beta g = \ln(1+f_d)) \\
\Longleftrightarrow & f_c = \frac{\ln(1+f_d)}{\beta g} \\
\Longleftrightarrow & f_c = \frac{\ln(1+f_d)}{R_c}.
\end{aligned}
$$

In the last step, we make use of $R_c = \beta/\mu$ and $g = 1/\mu$.

This is equivalent to $f_d = e^{R_c f_c} - 1$. Using a Taylor expansion we get $f_d = 1 + R_c f_c + O((R_c f_c)^2) - 1$ and thus $f_d = R_c f_c$ for small $R_c f_c$.

A.3.4 Maximum likelihood parameter estimation

Next we explain how we estimate a and t_0 of the logistic functions (Eqn. 2 and 5) from our data using maximum likelihood. We consider that we have data at times $t_1, ..., t_d$. At time t_i, we obtained n_i samples, where n_i is fixed, non-random.

We assume that the true number of B.1.1.7 variants at time t_i is a random variable, K_i which is binomially distributed with parameter $p(t_i)$, i.e.

$$
K_i \sim \mathcal{B}\left(n_i, p(t_i)\right), \qquad \text{where } p(t_i) = \frac{e^{a(t_i - t_0)}}{1 + e^{a(t_i - t_0)}} \quad .
$$

In particular, we assume here a deterministic logistic growth model for the increase in the proportion of variant Y (Eqn. 2 and 5), on top of which only the drawing process is random. This model simplifies naturally to a very popular logistic regression. This is an instance of a Generalized Linear Model, where the natural parameter of the binomial distribution is a linear function of predictors, the only predictor considered here being the time t.

We use the *Python* library *statsmodel* [22] to recover maximum likelihood estimates (MLEs) and confidence intervals based on an asymptotic Gaussian distribution for the parameters of the logistic regression based on our data, i.e. the fixed values $t_1, \ldots, t_d$, $n_1, \ldots, n_d$ as well as the numbers of

samples at each time point being the variant B.1.1.7, $k_1, \ldots, k_d$. Parameters a, t_0, f_d, f_c as well as the proportions of variant B.1.1.7 $p(t)$ through time are simple transformations of the parameters of the logistic regression. Their MLEs are the same transformations applied to the MLEs of the logistic regression parameters. The difference between the MLEs and the true parameters are again Gaussian, with a covariance matrix found by applying the delta method. This is used to construct confidence intervals for all these quantities.

A.4 Estimation of the effective reproductive number

We use the number of confirmed cases per day from the Federal Office of Public Health, Switzerland, for 14.12.2020 to 11.03.2021. Then, for each day, we estimate the number of B.1.1.7 variants by multiplying the total number of confirmed cases by the proportion of B.1.1.7 in our dataset (Viollier or Risch). We then estimate an effective reproductive number of the B.1.1.7 variant and of the non-B.1.1.7 variants using these data. For this estimation, we use the method developed in [23]. This method consists of two main parts: first, the observed case data is related to the corresponding time series of infections. We smooth the observations using LOESS smoothing to remove weekend effects. Then, we deconvolve with the delay of infection to symptom onset (gamma distributed with mean 5.3 and sd 3.2) and the delay from symptom onset to case confirmation (gamma distributed with mean 5.5 and sd 3.8). Second, we estimate the effective reproductive number from the time series of infection incidence using EpiEstim [24]. The reported point estimate is the estimate on the original case data. To account for uncertainty in the observation process, the observed daily case incidences are additionally bootstrapped 1000 times, resulting in an ensemble of alternative case incidence time series and corresponding estimated effective reproductive numbers. These are used to construct the 95% confidence interval around the effective reproductive number, and to calculate the standard deviation of the ratios of effective reproductive number estimates (see below).

We perform the estimation of the reproductive number in two different ways. First, we estimate smooth changes in the reproductive number, by estimating it across the entire time series using a 3-day sliding window. Second, we assume the reproductive number was constant during time intervals in which the non-pharmaceutical interventions did not change. Since 18.01.2021, Switzerland has implemented a set of tighter measures (in particular, shops are closed and the size of gatherings is restricted to five people [25]). Thus we fix the reproductive number to be constant between 01.01.2021

and 17.01.2021. Then the reproductive number is allowed to change and again fixed to be constant from 18.01.2021 onwards.

To compare the effective reproductive number R of the B.1.1.7 variant (Y) to that of non-B.1.1.7 variants (X), we take the ratio $\rho = \frac{R_Y}{R_X}$ at every time point. The standard deviation of this ratio σ_ρ was found through Gaussian error propagation of the standard deviation of the individual R estimates (σ_X, σ_Y):

$$\sigma_\rho = \sqrt{\frac{1}{(R_X)^2}\sigma_Y^2 + \frac{(R_Y)^2}{(R_X)^4}\sigma_X^2}.$$

A.5 Discrepancy for Ticino and Lake Geneva

The discrepancy for Ticino and Lake Geneva is not surprising: they had a reproductive number which was different from the national reproductive number in the first half of January. For Ticino, the empirical case numbers drop faster than the model, which is in line with a lower reported reproductive number compared to the national level[14]. For the Lake Geneva region, the empirical case numbers drop slower than the model, which is in line with a higher reported reproductive number compared to the national level[14]. For all regions but Ticino, we have enough data to estimate a reproductive number for the non-B.1.1.7 variants for 01.01.2021-17.01.2021. While for Switzerland, we obtained a point estimate of 0.83, the point estimates for all regions but Lake Geneva are between 0.81-0.83. Thus using the point estimate for all of Switzerland for the regional plots - with the exception of Lake Geneva and Ticino - in Fig. 5 is justified. For Geneva, we obtain a point estimate of 0.88. We use this point estimate in a Lake Geneva specific model (Fig. S1). Again we observe that the total number of confirmed cases dropped recently faster compared to the model. For a discussion on the discrepancy see main text.

A.6 Additional Tables

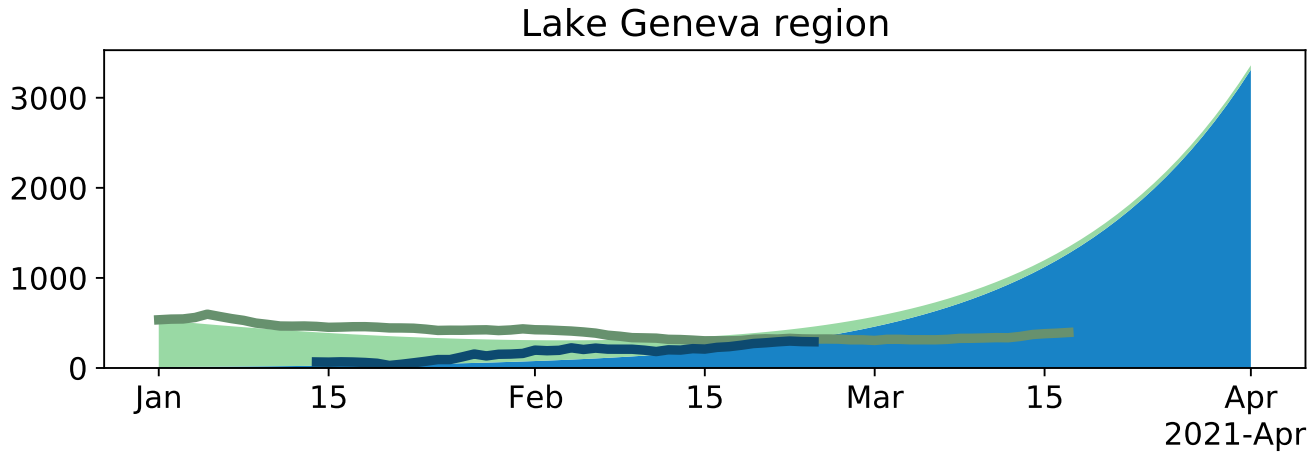

Figure S1: Change in the number of B.1.1.7 variants and in the number of all cases through time for the Lake Geneva region. For details see legend of Fig. 4. Compared to 4, we here use the average reproductive number estimated for non-B.1.1.7 in Geneva for the time period 01.01.2021-17.01.2021. The transmission fitness advantage is calculated based on this reproductive number and the estimate of the growth rate a for the Lake Geneva region.

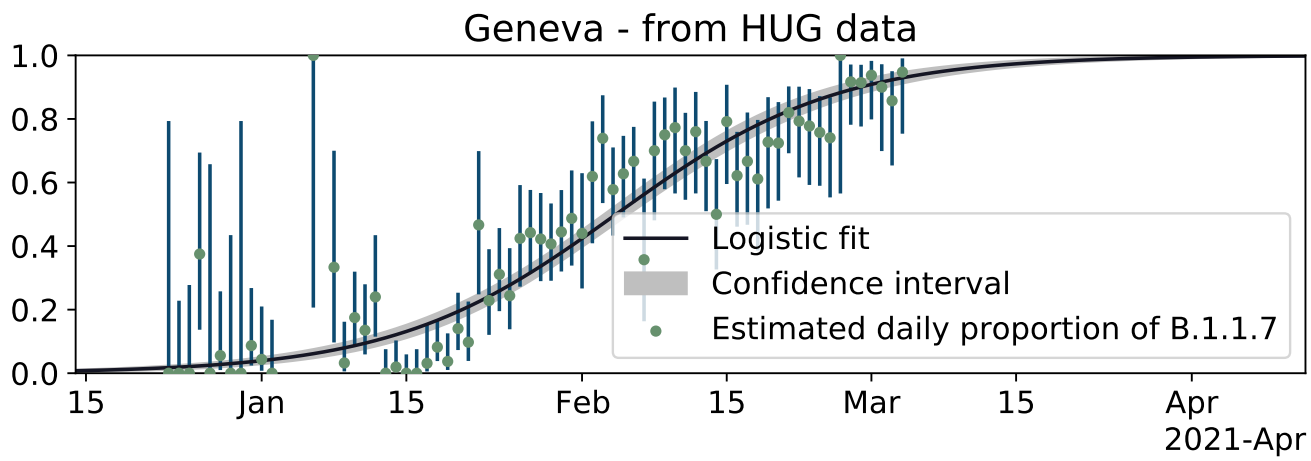

Figure S2: Logistic growth of frequency of B.1.1.7 in the Lake Geneva region based on HUG data.

Grossregion	Total confirmed cases	Sequenced	Proportion
Central Switzerland	17663	503	2.85%
Espace Mittelland	35933	2948	8.20%
Nordwestschweiz	24275	2716	11.19%
Tessin	9442	119	1.26%
Zurich	33615	1820	5.41%
Lake Geneva	36515	807	2.21%
Ostschweiz	26722	859	3.21%
Total	184165	9772	5.31%

Table S1: The proportion of sequenced cases out of all cases for the Viollier dataset.

Grossregion	Total confirmed cases	Sequenced	Proportion
Central Switzerland	17663	321	1.82%
Espace Mittelland	35933	3893	10.83%
Nordwestschweiz	24275	1135	4.68%
Tessin	9442	113	1.20%
Zurich	33615	1090	3.24%
Lake Geneva	36515	405	1.11%
Ostschweiz	26722	4594	17.19%
Total	184165	11551	6.27%

Table S2: The proportion of characterized cases out of all cases for the Risch dataset.