

Determinants of SARS-CoV-2 transmission to guide vaccination strategy in an urban area

Sarah C. Brüningk,^{1,2,†,§} Juliane Klatt,^{1,2,†,*} Madlen Stange,^{2,3,5,†,††} Alfredo Mari,^{2,3,5,††} Myrta Brunner,^{3,4,§§} Tim-Christoph Roloff,^{2,3,5,*} Helena M.B. Seth-Smith,^{2,3,5,†††} Michael Schweitzer,^{3,5} Karoline Leuzinger,^{6,15,†††} Kirstine K. Sogaard,^{3,5,§§§} Diana Albertos Torres,^{3,5} Alexander Gensch,³ Ann-Kathrin Schlotterbeck,³ Christian H. Nickel,^{7,*} Nicole Ritz,⁹ Ulrich Heininger,⁹ Julia Bielicki,^{9,††††} Katharina Rentsch,¹⁰ Simon Fuchs,¹² Roland Bingisser,⁷ Martin Siegemund,¹¹ Hans Pargger,¹¹ Diana Ciardo,¹³ Olivier Dubuis,¹³ Andreas Buser,¹⁴ Sarah Tschudin-Sutter,^{8,††††} Manuel Battegay,^{8,§§§§} Rita Schneider-Sliwa,⁴ Karsten M. Borgwardt,^{1,2,†,*} Hans H. Hirsch,^{6,8,15,†,††††} and Adrian Egli^{3,5,*†,††††}

¹Machine Learning & Computational Biology, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland, ²Swiss Institute for Bioinformatics (SIB), Lausanne, Switzerland, ³Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland, ⁴Human Geography, Department of Environmental Sciences, University of Basel, Basel, Switzerland, ⁵Clinical Bacteriology and Mycology, University Hospital Basel & University of Basel, Basel, Switzerland, ⁶Clinical Virology, University Hospital Basel & University of Basel, Basel, Switzerland, ⁷Emergency Department, University Hospital Basel, Basel, Switzerland, ⁸Infectious Diseases and Hospital Epidemiology, University Hospital Basel & University of Basel, Basel, Switzerland, ⁹Paediatric Infectious Diseases and Vaccinology, University of Basel Children's Hospital Basel and University of Basel, Basel, Switzerland, ¹⁰Laboratory Medicine, University Hospital Basel, Basel, Switzerland, ¹¹Intensive Care Medicine, University Hospital Basel, Basel, Switzerland, ¹²Health Services for the Basel-City, Basel, Switzerland, ¹³Viollier AG, Allschwil, Switzerland, ¹⁴Regional Blood Transfusion Service, Swiss Red Cross, Basel, Switzerland and ¹⁵Transplantation & Clinical Virology, Department Biomedicine, University of Basel, Basel, Switzerland

[†]These authors contributed equally to this work.

[‡]These authors share senior authorship.

[§]<https://orcid.org/0000-0003-3176-1032>

[¶]<https://orcid.org/0000-0003-4096-566X>

^{††}<https://orcid.org/0000-0002-4559-2535>

^{†††}<https://orcid.org/0000-0002-3871-088X>

^{§§}<https://orcid.org/0000-0002-9694-5396>

^{***}<https://orcid.org/0000-0003-3435-4723>

^{††††}<https://orcid.org/0000-0002-6082-5114>

^{†††††}<https://orcid.org/0000-0002-5654-9356>

^{§§§}<https://orcid.org/0000-0003-3624-7999>

^{****}<https://orcid.org/0000-0001-6619-9284>

^{††††††}<https://orcid.org/0000-0002-3902-5489>

^{†††††††}<https://orcid.org/0000-0003-4875-5368>

^{§§§§}<https://orcid.org/0000-0002-6638-3679>

^{*****}<https://orcid.org/0000-0001-7221-2393>

^{††††††††}<https://orcid.org/0000-0003-0883-0423>

^{†††††††††}<https://orcid.org/0000-0002-3564-8603>

*Corresponding author: E-mail: Adrian.Egli@usb.ch

Abstract

Transmission chains within small urban areas (accommodating ~30 per cent of the European population) greatly contribute to case burden and economic impact during the ongoing coronavirus pandemic and should be a focus for preventive measures to achieve containment. Here, at very high spatio-temporal resolution, we analysed determinants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission in a European urban area, Basel-City (Switzerland). We combined detailed epidemiological, intra-city mobility and socio-economic data sets with whole-genome sequencing during the first SARS-CoV-2 wave. For this, we succeeded in sequencing 44 per cent of all reported cases from Basel-City and performed phylogenetic clustering and compartmental modelling based on the dominating viral variant (B.1-C15324T; 60 per cent of cases) to identify drivers and patterns of transmission. Based on these results we simulated vaccination scenarios and corresponding healthcare system burden (intensive care unit (ICU) occupancy). Transmissions were driven by socio-economically weaker and highly mobile population groups with mostly cryptic transmissions which lacked genetic and identifiable epidemiological links. Amongst more senior population transmission was clustered. Simulated vaccination scenarios assuming 60–90 per cent transmission reduction and 70–90 per cent reduction of severe cases showed that prioritising mobile, socio-economically weaker populations for vaccination would effectively reduce case numbers. However, long-term ICU occupation would also be effectively reduced if senior population groups were prioritised, provided there were no changes in testing and prevention strategies. Reducing SARS-CoV-2 transmission through vaccination strongly depends on the efficacy of the deployed vaccine. A combined strategy of protecting risk groups by extensive testing coupled with vaccination

of the drivers of transmission (i.e. highly mobile groups) would be most effective at reducing the spread of SARS-CoV-2 within an urban area.

1. Introduction

Efforts to understand transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been undertaken at different scales including at a global level (Yuan et al. 2020; Nabil et al. 2020; Hadfield et al. 2018), across continents (Europe and North America (Worobey et al. 2020)), within countries (Austria (Popa et al. 2020)), Brazil (Candido et al. 2020), France (Salje et al. 2020), Iceland (Gudbjartsson et al. 2020), South Africa (Post et al. 2020) and Thailand ((Puenpa et al. 2020)) and in large cities (Beijing (Du et al. 2020), Boston (Lemieux et al. 2020)), Houston (Long et al. 2020) and New York City (Maurano et al. 2020; Bushman et al. 2020; Kissler et al. 2020). In Europe, ~30 per cent of the population live in small urban areas (10k – 300k inhabitants)(EUROSTAT 2011), which accordingly play a major role in SARS-CoV-2 transmission and yet have not been studied. Moreover, to date city-based studies of SARS-CoV-2 transmission have very limited resolution in terms of the proportion of sequenced positive cases (incomplete transmission chains), have a paucity of socio-economic or mobility data (incomplete determinants) or fail to combine analysis of transmission clusters with quantitative, descriptive models accounting for population mixing (De Ridder et al. 2021a). Of those studies describing the distribution of cases together with changes in mobility, none to date rigorously study socio-economic differences between city quarters as determinants of transmission (Long et al. 2020; Bushman et al. 2020; Kissler et al. 2020; Chang et al. 2020). An integrated model considering all of these factors (including epidemiological, geographic, mobility, socio-economic and transmission dynamics information) is anticipated to provide profound insights into the determinants underpinning transmission, which can be used to guide the delivery of vaccines. We here present such an integrated analysis for Basel-City, which is part of a metropolitan area, a functional urban area and a European cross-border area as outlined in the supplement. Basel-City is hence representative of other areas in the European Union classified as such.

Local interventions are most effective in cutting transmission chains in families and small community networks (Koo et al. 2020; Mansilla Domínguez et al. 2020) that represent well-defined (phylogenetic or epidemiological) clusters. However, most infections are acquired from unknown sources and transmitted cryptically, making it essential to identify the key determinants and transmission routes at the city level to improve interventions and vaccination campaigns. To address this challenge, we have combined phylogenetic cluster analysis (Ragonnet-Cronin et al. 2013) and compartmental ordinary differential equation (ODE) modelling based on high density (81 per cent of reported cases assessed) and high-resolution (spatial: housing blocks, temporal: day-by-day) epidemiological, mobility, socio-economic and serology (estimate of unreported cases) data sets from the first coronavirus disease (COVID-19) wave (February–April 2020). Whole-genome sequencing (WGS) of all included cases (44 per cent of all cases successful) allowed the analysis be restricted to a single, dominant viral variant (B.1-C15324T, 60 per cent of cases (Stange et al. 2021)). This ensured that our analyses focused on inherently related cases and enabled estimation of effective reproductive numbers for different socio-economic and demographic

population subgroups to provide the basis for vaccination scenario building.

2. Materials and Methods

2.1 Included data

2.1.1 PCR testing and WGS

All analyses were based on polymerase chain reaction (PCR)-positive (750/7,073 tests) cases of residents of Basel-City between 25 February and 22 April 2020, obtained from the University Hospital Basel (UHB). PCR testing was available rapidly, and frequent testing was established and supported by local guidelines by the end of February 2020, before the first case arrived (Leuzinger et al. 2020). SARS-CoV-2 testing was made available (1) via a walk-in test centre in the city centre affiliated to the UHB, which allowed screening of legal-aged patients with mild and severe symptoms, (2) via the University Children's Hospital for minors on recommendation by the paediatricians and (3) via the obligatory screening of any incoming patients to the UHB irrespective of symptoms. Testing in case of symptoms was covered by the Swiss mandatory health insurance scheme, preventing sampling bias from affluent socio-economic population groups. The total number of positive cases for Basel-City including also external testing sources for the same time range was 928, hence the cases registered at the UHB cover 80.8 per cent of the total case burden (Kanton Basel-Stadt Datenportal data.bs.ch 2021). The ratio of negative to positive PCR tests changed during the local epidemic with a median of 10.6 per cent positive PCR tests (see Fig. 2B). All positive samples were subjected to WGS, 53 per cent resulted in high-quality genomes (i.e. 44 per cent of all cases). Of these 247 (247/411, 60 per cent) contained the monophyletic C15324T mutation in the B.1 lineage (B.1-C15324T) characteristic to the virus variant that originated in this tri-national area (Stange et al. 2021). Only B.1-C15324T cases were used for further analysis.

2.1.2 Geographic mapping and socio-economic stratification

Basel-City is divided into 21 urban quarters and had a population of ~200k in 2020 (Statistisches Amt Basel-Stadt: Bevölkerungsstatistik 2021) (Supplementary Fig. S5). Each PCR test ($N=7,073$), irrespective of the result, was linked to the patient's place of residence anonymised at the scale of statistical (housing) blocks (a city block partitioned by e.g. streets and rivers) in ArcMap 10.7 (by Environmental Systems Research Institute). For each of these housing blocks, where privacy legislation permitted, Basel-City's Cantonal Statistical Office provided socio-economic indicators for the year of 2017 (most recent available). These included (1) living space (per capita in m^2), (2) share of 1-person private households, (3) median income in Swiss Francs and (4) population seniority (percentage of citizens aged over 64 years). According to these indicators, blocks were allocated to one of three socio-economic tertiles (T1: ≤ 33 rd percentile, T2: 33rd to 66th percentile, T3: > 66 th percentile, N/A: no available data or censored) where possible (e.g. Figure 3A). Generally, sparsely populated blocks displayed a maximum of three positive cases and had to be excluded from analysis. The choice of tertiles (over e.g. quartiles) was a pragmatic decision to account for the trade-off between cases per partitioning and socio-economic diversity covered.

All following analyses with respect to socio-economic factors were based on these partitions which were assigned independently for each indicator.

2.1.3 Serology and intensive care unit data

SARS-CoV-2 antibody responses were determined in a total of 2,019 serum samples collected from individuals between 25 February and 22 May, 2020, to account for seroconversion. Serology information was used to estimate the fraction of unreported cases as follows: An estimated 1.88 per cent (38/2,019) of the Basel-City population was infected with SARS-CoV-2. Of these 60 per cent would be attributed to the B.1-C15324T variant, leading to a percentage of 88 per cent of unreported/unsequenced cases to consider (calculated as $1 - \frac{n_{B.1-C15324T}}{p_{infected} \cdot n_{population}}$, with $n_{population}$ being the population of the city, p , $n_{B.1-C15324T}$ the reported case counts and fraction of the B.1-C15324T variant, and $p_{infected}$ the estimated fraction of the total infected population).

Finally, we included data on the number and age distribution of COVID-19 intensive care unit (ICU) patients during the relevant period from UHB, a tertiary hospital with a capacity of 44 ICU beds: 4.5 per cent of reported SARS-CoV-2-positive cases were admitted to ICU and median length of ICU stay was 5.9 days (interquartile range (IQR) 1.5–12.9). Forty per cent of these patients were younger than 64 years.

2.1.4 Phylogenetic inference and cluster analysis

Whole SARS-CoV-2 genomes from Basel-City patients were assembled using our custom analysis pipeline COVID-19 Genome Analysis Pipeline (Stange et al. 2021) (github.com/appliedmicrobiologyresearch/SARS-CoV-2). Global sequences and metadata were downloaded from Global Initiative on Sharing Avian Influenza Data (GISAID) (Elbe and Buckland-Merrett 2017; Shu and McCauley 2017) (as of 22 October 2020; 155,278 consensus sequences). Sequences with more than 10 per cent N's (27,013) and with incomplete dates (43,466) were removed. 84,799 sequences remained which were joined with Basel genomes. Filtering for the period of interest until 22 April retained 39,913 genomes, of which 411 are from Basel-City residents dating from 26 February (first case) to 22 April 2020.

To infer relatedness among the viral genomes and spread of SARS-CoV-2 in Basel-City, a time-calibrated phylogeny that was rooted to the first cases in Wuhan, China, from December 2019 was inferred using a subset of the global genomes. For subsetting, we included 30 genomes per country and month, whereby all genomes from Basel-City were retained, totalling 3,495 genomes, using the nextstrain software v.2.0.0 (nextstrain.org) and augur v.8.0.0 (Hadfield et al. 2018) as described in detail by Stange et al. (Stange et al. 2021).

The resulting global phylogeny was used to infer phylogenetic clusters in Basel-City. First, polytomies, which are caused by identical genomes in the tree were resolved using ETE3 v.3.1.1 (Huerta-Cepas et al. 2016). Cluster Picker v.1.2.3 (Ragonnet-Cronin et al. 2013) was then used to identify clusters in the resolved tree. Bootstrap support thresholds were set to 0 since enforced bifurcation (previous step) resulted in internal nodes without support values (options *initial threshold* = 0, *main support threshold* = 0). The maximum genetic distance defining a cluster and minimal cluster size were set to 0.0004 (*genetic distance threshold* = $4e-4$) and 5 (*large cluster threshold* = 5), respectively. Identified clusters were consolidated manually with epidemiological data (occupation in a health service job, resident of a care home, contact to positive

cases, onset of symptoms and place of infection) to confirm the suitability of the divergence parameter and accuracy of identified clusters. Cluster Matcher v.1.2.4 (Ragonnet-Cronin et al. 2013) was then used to combine ancillary geographic (quarter) and socio-economic or demographic information that were subdivided into tertiles on identified clusters.

To test whether related genomes in Basel-City cluster according to (1) quarter, (2) living space per person, (3) share of one-person households, (4) median income or (5) seniority, with the null hypothesis of random distribution of cases and hence clusters across tertiles, a custom python-script for a random permutation test was performed (Egli et al. 2020)¹. The results for clustering within and among urban quarter and tertiles in socio-economic determinants were visualised using circos v.0.69 (Krzywinski et al. 2009). The circos plots visualise the within-tertile sample size as ticks (each tick a sample/sequence) circularly arranged, the amount of between-tertile associations (via phylogenetic clustering of samples) by connecting lines (the thicker the lines the more samples are associated between tertiles) and P-value for within- and between-tertile accumulation of clusters via a colour code applied to the ring (within-tertile) and the links (between-tertile).

2.2 Mechanistic modelling

2.2.1 Mobility data and transport model

We employed the official traffic model provided by the traffic department of Basel-City (Bau- und Verkehrsdepartement Basel-Stadt Mobilität / Mobilitätsstrategie 2020) consisting of the 2016 average A-to-B traffic on a grid of ~1,400 counting zones for foot, bike, public motorised transport and private motorised transport. We further obtained weekly averages of pass-by traffic for the same count zones over the period of the first wave of the pandemic for the categories of combined foot and bike traffic, as well as private motorised traffic. Additionally, weekly public-transport passenger loads were provided by both the Swiss Federal Rail Company and the local public transport services.

From these data sets, we computed the spatio-temporal variation of mobility within the city: Spatial variation was obtained by aggregating A-to-B traffic between counting zones, first to the statistical block level (by identifying the nearest housing block with respect to a zone's centroid) and second to tertile level via a statistical block's association with a socio-economic indicator tertile. This resulted in a three-by-three mobility matrix M_{jk} representing within-tertile/inter-tertile mobility on/off its diagonal. This matrix was unity normalised since only relative differences were relevant in our model. Temporal variation was obtained as the weighted sum over the public and private transport time series according to the relevant transport mode contribution. This sum was then normalised and smoothed with a univariate spline resulting the final time series for temporal mobility variation (denoted as $\alpha_{mob}(t)$).

2.2.2 Dynamic changes in social interaction

SARS-CoV-2 transmission is contact-based. While the number of contacts potentially taking place within a day and a city is largely influenced by human mobility as estimated above, the

¹ github.com/appliedmicrobiologyresearch/Influenza-2016-2017/blob/master/permutation_test_influenza_cluster.py. This script counts the number connections of samples that are assigned to the same cluster within tertiles and between any two tertiles and then it performs a permutation test (10,000 repeats) that randomly assigns tertiles to the samples and then re-counts connections of shared clusters within and between tertiles to assess significance.

risk of a contact becoming a transmission event is further determined by the precautions taken by the two individuals being in contact (such as washing hands, wearing masks and distance keeping). Both aspects together—mobility and risk-mitigating social behaviour within a (sub-)population—eventually result in an effective, time-dependent, reproductive number characterising the virus's transmission within that (sub-)population. Hence, there are three relevant time series: changes in the overall effective reproductive number, in mobility and in social behaviour. While the computation of the temporal variations in mobility was described above, the overall time-dependent effective reproductive number is obtained by applying a Kalman filter (Kalman et al. 1960; Welch et al. 1995) to the daily case counts of individuals having newly contracted the B.1-C15324T variant of SARS-CoV-2 in all of Basel-City. To this end, we focus on the reduced dynamics as described by the E , P and U compartments with constant times T_{inc} , $T_{inf,P}$ and $T_{inf,U}$ (see Fig. 1) obtained through a grid search. Changes in the number of susceptibles S is slow compared to the other compartments, which allows us to approximate that number as constant and thereby linearise the dynamics—enabling the use of a Kalman filter in the first place. The measurement used to update the filter is $P(t)$ as observed via the positive contribution $p_{sq} \cdot P(t)/T_{inf,P}$ to the daily increment of $U(t)$. Assuming a multiplicative model, the time dependence of residual transmission risk stemming from lack of precaution in social interaction (denoted as $\alpha_{soc}(t)$) is obtained by point-wise division of the time dependence of the effective reproductive number by the mobility time series (depicted in Fig. 3). Thus we are adhering to the logic that in the extreme case of zero mobility, no transmission can take place despite a finite risk of transmission rooted in a lack of precautions, while on the other hand in the case of zero risk of transmission due to perfect precautions, no transmission can take place despite non-zero mobility. Such logic dictates the choice of a multiplicative rather than additive model.

2.2.3 SEIR model

We used a compartmental two-arm SEIR model (Chang et al. 2020; Chinazzi et al. 2020; Li et al. 2020) including sequenced and unsequenced/unreported cases that is outlined in Fig. 1 using the following compartments: S (susceptibles), E (exposed, latency period T_{inc}), P (presymptomatic, infectious time $T_{inf,P}$), I (reported infectious, isolated), U_i (unreported infectious, infectious time $T_{inf,U}$), U_r (unreported recovered). The initial number of susceptibles was fixed to the relevant population. All other compartments were initialised as zeros, apart from a seed in E corresponding to the first reported cases. In summary, our model is based only on six free parameters: the tertile-specific reproductive number R_j (three parameters, range $[0, 20]$, this is multiplied by $\alpha_{soc}(t)$ and $\alpha_{mob}(t)$ to include dynamic variation), the initial number of exposed in a single tertile in which the first cases were recorded (range $[0, 20]$), the infectious times $T_{inf,P}$ (range $[2, 12]$ days) and latency period T_{inc} (range $[2, 12]$ days) (He et al. 2020). Since it was not possible to distinguish the fit for $T_{inf,P}$ and $T_{inf,U}$, we assumed a value of two days for the latter infectious time. The ODE system was implemented in python (version 3.8.) using the scipy functions 'odeint' to iteratively solve the system of equations.

2.2.4 Fitting procedure and evaluation

In total 247 cases within the time period from 25 February until 22 April were included. For all data a 7-day moving window average was taken to account for reporting bias on weekends.

Due to the loss of single sequencing plate, missing numbers on 29, 30 and 31 March were imputed by assuming a constant ratio of the B.1-C15324T variant amongst the sequenced samples. Simulations were initialised on 22 February, the estimated date of the occurrence of the initial exposed cases (Stange et al. 2021).

Model fitting was performed on absolute (i.e. not cumulative) infected case numbers (dI) simultaneously for all partitions based on the least squares method using the 'lmfit' library (version 1.0.2 (Newville et al. 2014)) with default parameters. Posterior probability distributions of the fitted parameters were estimated using the Markov Chain Monte Carlo method implemented via the 'emcee' algorithm (Foreman-Mackey et al. 2013) of the 'lmfit' library. We report median values with 95 per cent confidence intervals corresponding to the range of the 2.275th and 97.275th percentiles. We compare effective reproductive numbers corresponding to the normalisation of R by the effective mobility contribution ($\sum_k M_{jk}$):

$$R_{eff,j} = \frac{R_j}{\sum_k M_{jk}}. \quad (1)$$

Significance levels of R_{eff} between tertiles are scored based on a comparison to 99 random partitions of the statistical blocks to a 5 per cent, Bonferroni corrected significance threshold.

2.2.5 Scenario simulation

The impact of mobility relative to social interaction was analysed by recalculating the predicted epidemic trajectory under the constraint of constant intra-urban mobility ($\alpha_{mob}(t) = 1$, scenario M1) or fully restricted ($\alpha_{mob}(t) = 0$, scenario M2) mobility, corresponding to perfect isolation of the affected city areas. These scenarios were compared to the baseline of the actual reduction in mobility (scenario M0).

Vaccination scenarios were simulated as for both 90 per cent and 70 per cent effective vaccines to prevent COVID-19 resembling current vaccine candidate data (AstraZeneca 2020; Mahase 2020), as well as a range of vaccine efficacies to prevent SARS-CoV-2 transmission (60 per cent and 90 per cent). This was achieved by moving the fraction of the vaccinated and not transmitting population from the susceptible to the recovered compartment U_r and calculating the spread of the pandemic with constant effective reproductive number and intra-city mobility. We accounted for a change in social interaction behaviour following vaccination by assigning a mean social interaction score of the vaccinated and not-vaccinated population amongst the initial susceptibles ($\alpha_{soc,vacc}(t) = 0.75$, $\alpha_{soc,novacc}(t) = 0.5$). Mobility was modelled as 100 per cent ($\alpha_{mob}(t) = 1$). Two scenarios were investigated and compared to the no vaccine scenario (V0): (1) vaccination of a fixed population fraction (one or two thirds) randomly throughout the population (scenario V1), (2) vaccination of the corresponding number of individuals from different socio-economic groups (scenario V2 (exclusively from T1 median income), V3 (exclusively from T3 seniority) and V4 (50 per cent from T1 median income and 50 per cent from T3 seniority)). In order to gauge the benefit of a particular vaccination scenario, we calculated the time to reach 50 per cent of ICU capacity. The UHB has a total of 44 ICU beds. During the first wave, 4.5 per cent of reported SARS-CoV-2-positive cases were admitted to ICU, and their median length of ICU stay was 5.9 days (IQR, 1.5–12.9). If considering additional unreported cases (captured by serological testing), the percentage of patients requiring ICU admission was 1 per cent. Of all SARS-CoV-2 patients with ICU stay, 40 per cent were younger than 64 years, resulting in a probability of an under-64-year-old

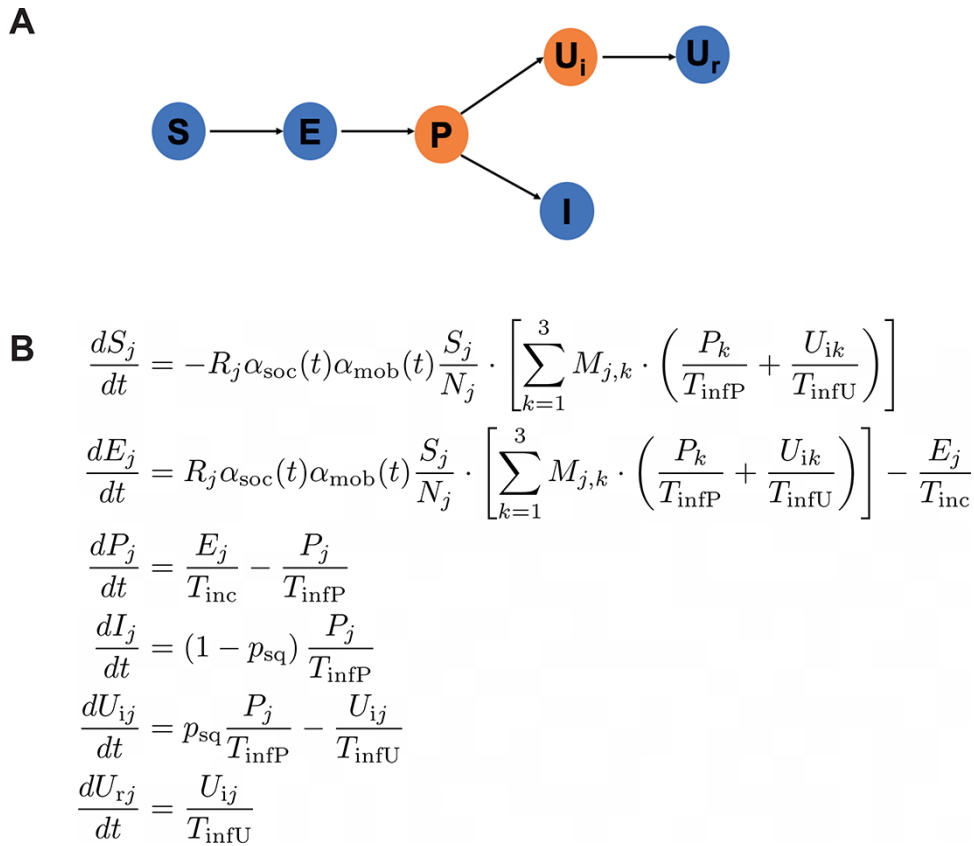


Figure 1. Overview of the susceptible-exposed-infected-recovered (SEIR) model. A) Conceptual overview. We accounted for susceptibles (S), exposed (E, incubation time T_{inc}) and pre-symptomatic yet infectious cases (P). After a presymptomatic time T_{infP} , cases were separated according to the estimated proportion of reported and sequenced cases, p_{sq} , into either reported infectious (I) or unreported infectious (U_i , reproductive number R). Since our data did not include information on recovered patients, a ‘recovered’ compartment was not included following I. It was assumed that reported cases remained isolated. The unreported compartment transitions to recovery (U_r) after an infectious time T_{infU} . B) Relevant model equations to incorporate connectivity and exchange between the defined tertiles (index j). Cross-contamination was included through the mobility matrix M_{jk} and relevant temporal variation of mobility and social interaction (weighting factors $\alpha_{\text{mob}}(t)$ and $\alpha_{\text{soc}}(t)$).

infected case to be admitted to ICU of 0.5 per cent. In case of vaccinated populations not at random, we adjust the relevant fraction of ICU cases based on the represented proportions of *geq* and < population fractions within all susceptibles.

3. Results

3.1 SARS-CoV-2 spread and clustering

We observed 29 viral lineages in Basel-City (Supplementary Fig. S7), with 247 genomes (60.0 per cent) belonging to the B.1-C15324T variant (Stange et al. 2021) (Figs. 2A and B). Applying a genetic divergence threshold and manual verification using available epidemiological data, a total of 128 phylogenetic clusters were determined across all samples, of which 70 belonged to lineage B.1-C15324T (Fig. 2C). Associating phylogenetic clusters (comprising two or more cases) with tertiles of socio-economic and demographic determinants, we found that for median income, T1 contained the most and T3 contained the least clusters (Fig. 2C), which reflects the differences in case counts between tertiles and hence leads to similar numbers of cases per cluster between tertiles. Testing for significant clustering of related cases within and among tertiles of different socio-economic factors showed that all among-tertile transmissions were spread randomly (non-significant). Only some within-tertile transmissions were significant, namely for blocks with either the highest living space per person (T3), lowest share of 1-person households

(T1), highest seniority (T3) or high median income households (T3) (Fig. 2D)). Results are exemplary visualised for median income in Fig. 2E. Conversely, positive cases that belong to lower socio-economic/demographic tertiles (either lower income, less living space or lower share of senior residents) are less likely to be members of the same phylogenetic cluster, indicating cryptic transmission predominates here. Further information regarding the number of clusters and cases within different geographic city quarters are provided in the supplement (Figure S6). Here we find clustering of cases in 4 out of 20 city quarters, 2 of which (Riehen and Bruderholz) belong to the most affluent quarters of Basel (fourth quartile (Q4)), the other two (Am Ring and Iselin) to the middle field (Q3 and Q2, respectively). We also observed within-quarter clustering of cases (Figure S6).

3.2 Spatio-temporal variation of mobility and social interaction patterns

Figure 3A and Figures S2 and S3 show Basel-City’s partition and the corresponding mobility graph. Importantly, the statistical blocks per tertile do not form a single, geographically connected entity. We observe that mobility varies by transport modality and tertile (Fig. 3A inset). For example, for low-median income (T1) the share of private motorised traffic and mobility is more pronounced than in the tertiles of higher median income (T2 and T3).

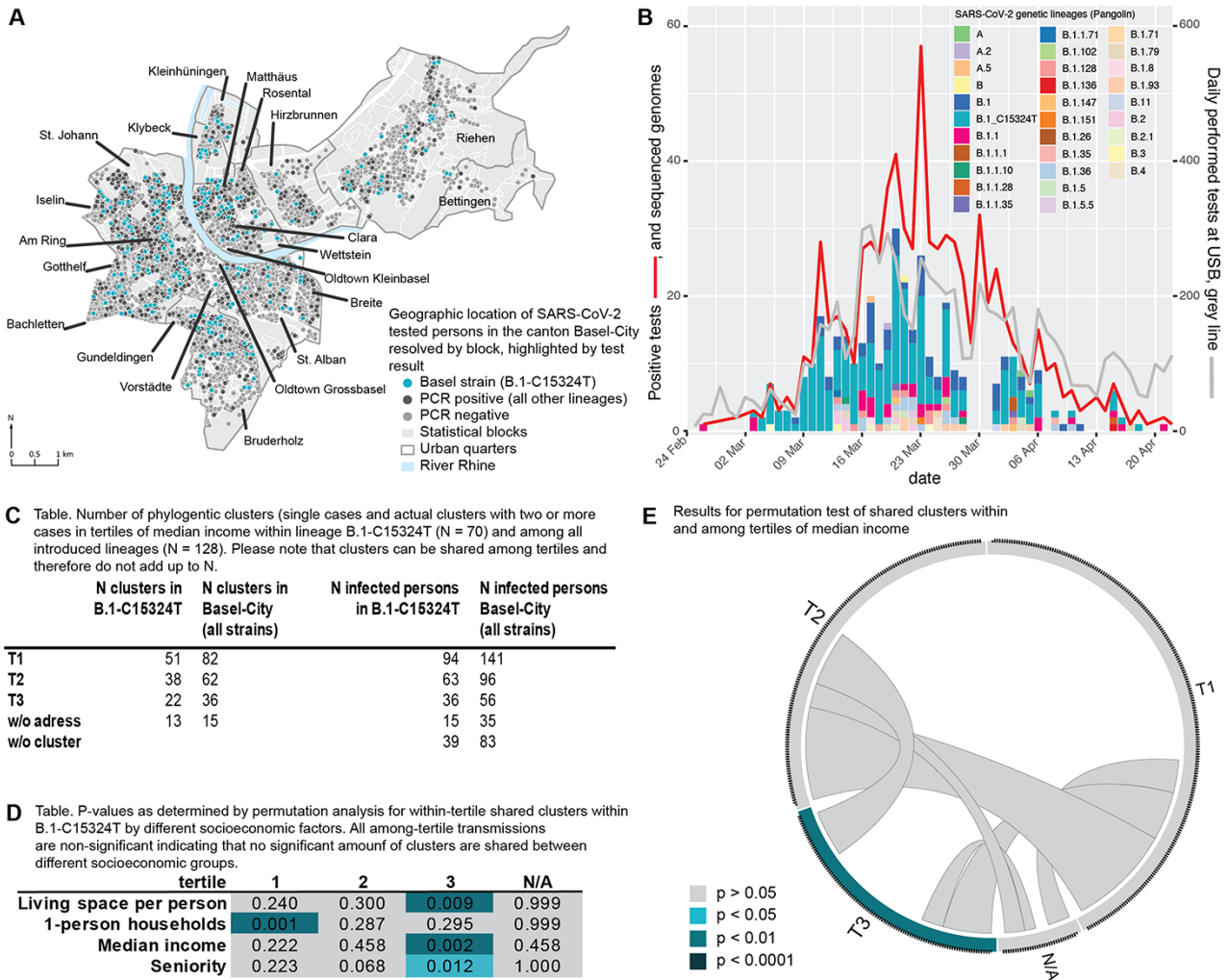


Figure 2. SARS-CoV-2 transmission in and among socio-economic and demographic groups during the first COVID-19 wave in Basel-City. A) Spatial positive/negative case distribution throughout the city with the most dominant SARS-CoV-2 variant (B.1-C15324T), the focus of this study, highlighted in turquoise. B) Epidemiological curve for Basel-City and distribution of phylogenetic lineages (pangolin nomenclature) from 25 February to 22 April 2020. C) Summary for inferred phylogenetic clusters within (1) all lineages and (2) the major variant B.1-C15324T in tertiles of median income. High number of infected people within a tertile with a low number of clusters indicates presence of large transmission clusters whereas large number of clusters and low number of people infected within a tertile indicates random infections and cryptic transmission. D) Results of a significance test for transmission between tertiles of different socio-economic factors. T1: low, T2: intermediate, T3: high, N/A: no available data or censored for privacy reasons. E) Visualisation of a significance test for transmission within (indicated on circle edges) and among (intra-circle connections) tertiles of median income.

Figure 3B shows for each partition the summed edge weights of the mobility graph accounting for the mobility contribution to the final effective reproductive number. We observe that low and median income populations are more mobile than their wealthier counterparts. For living space per person or percentage of senior citizens, mobility was comparable between tertiles with a trend towards higher mobility within the younger population groups.

Dynamic changes in mobility were assessed by agglomerating normalised traffic counts for public and private transport modalities (Fig. 3C). There was a clear drop in mobility for both public and private transport modes around the onset of the national lockdown date (12 March 2020). The decrease was more pronounced for public transport, resulting in a weighted average mobility drop of approximately 50 per cent (Fig. 3D). Figure 3D also shows the dynamic change in social interaction contribution to B.1-C15324T case numbers. Despite noticeable fluctuation, social interaction contribution decreased on average over time. These data also

reflect variation in case reporting which affected the estimated effective reproductive number.

3.3 Spatio-temporal spread of the epidemic and its socio-economic determinants

Unreported cases appeared to be a driving force of the transmission (88 per cent for the sequenced B.1-C15324T variant). Figures 4A–C (and Figure S4) show the SEIR model fit to data for each median income tertile. Independent of the underlying partition, the model provided adequate fits (quantified by a root mean squared error < 6 for cumulative, or < 1 for absolute case numbers for all fits). For the parameters shared between all socio-economic partitions we obtained T_{inf} 2.3 (2.1,2.6) days, T_{inc} 2.4 (2.3,3.4) days and E_0 12.8 (8.1,14.3). The corresponding dynamic change of the effective reproductive number (R_{eff}) is given in Figs. 4D–F. We observed a drop in R_{eff} following the dynamic changes in mobility and social interaction. Importantly, there was a significant difference (achieved significance level below the Bonferroni corrected

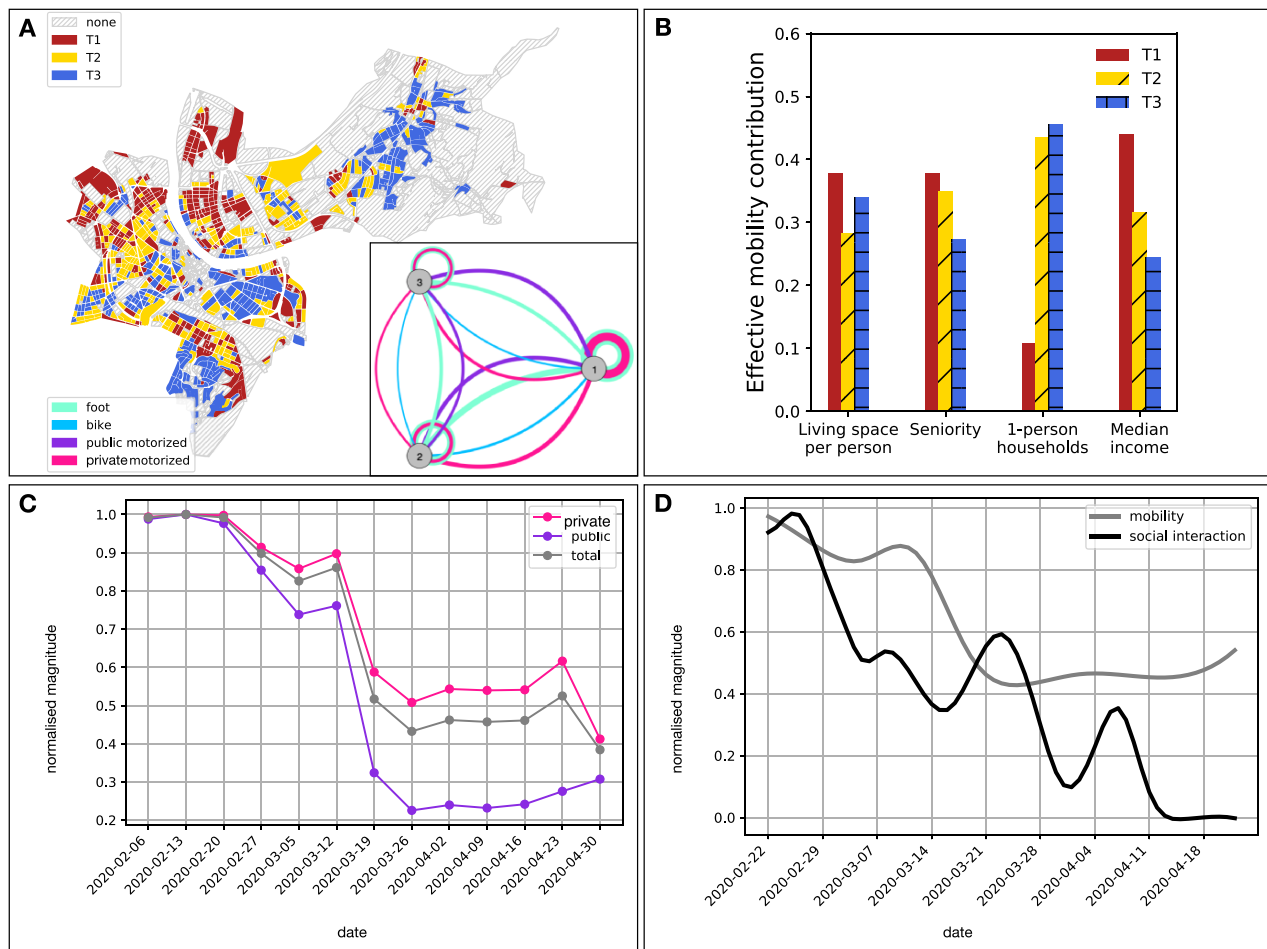


Figure 3. Spatio-temporal variation of mobility patterns within the Canton of Basel-City. A) Basel-City and its delineation with respect to statistical blocks coloured according to the partition into tertiles T1, T2 and T3 of increasing median income. Inset: resulting mobility graph, with nodes representing tertiles and edge widths representing the strength of effective connectedness through mobility by means of various modes of transport, as computed from the traffic model provided by the traffic department of the Canton of Basel-City. B) Relative mean contribution of mobility to a socio-economic tertile's effective reproductive number associated with the major variant B.1-C15324T. C) Normalised temporal development of private and public transport as well as their weighted sum during the first wave of the pandemic in Basel-City. D) Smoothed relative temporal development of social interaction and mobility contribution to the effective reproductive number associated with the major variant B.1-C15324T.

significance threshold of 5 per cent, see Table 1) in R_{eff} between statistical blocks of the highest and lowest median income. For all socio-economic partitions the obtained parameter distributions are summarised as histograms with median values and 95 per cent confidence intervals in Figs. 4G–I. Here, we found that blocks with higher living space per person or higher median income had a significantly lower R_{eff} (< 1.7) compared to the maximum R_{eff} observed in the relevant partition. A partitioning based on the share of senior residents did not result in significant differences in R_{eff} . Differences in R_{eff} are due to two factors: the effective mobility contribution (Fig. 3B) and the modelled reproductive number (R , eq.(1)). In particular, the tertile with the highest share of median income (T3) showed less mobility compared to T2 and T1, emphasising differences in R_{eff} . By contrast, mobility in the T1 and T3 tertiles of living space per person were more similar (Fig. 3B), yet differences in R_{eff} were significant, indicating that the transmission was not dominated by mobility alone.

3.4 Impact of mobility changes and modelling of vaccination scenarios

We simulated the developments of the first wave of the epidemic under the assumption of different mobility scenarios and

modelled two future vaccination strategies. Figure 5A displays the results for mobility scenarios as observed with up to 50 per cent mobility reduction (scenario MO) and 100 per cent mobility (scenario M1). Peak case numbers (12 April) would have been far higher in the case of no reduction in mobility (M1 $\sim 150k$ vs 25k cases). Mobility reduction hence played a vital role for the containment of SARS-CoV-2.

Figure 5B shows the results for an outbreak scenario (denoted as V1) in which a random fraction of the population (33 per cent or 66 per cent) received a vaccine that provided either 60 per cent or 90 per cent vaccine transmission reduction, with 90 per cent severe COVID-19 case prevention. As expected, we observe that higher vaccine efficacy or higher population fraction vaccinated reduces the slope and plateau of the epidemic curve. Scenarios for less effective vaccines are shown in Figure S8. Effectiveness to prevent severe COVID-19 solely affects the rate of severe cases and hence ICU occupancy and the time point of reaching 50 per cent ICU occupancy (Fig. 5D vs Figure S8F). It should be noted that vaccination of the population at random is an artificial scenario, applied here only to demonstrate the impact of vaccination efficacy relative to the population fraction vaccinated. This scenario serves as a baseline comparison for two more realistic

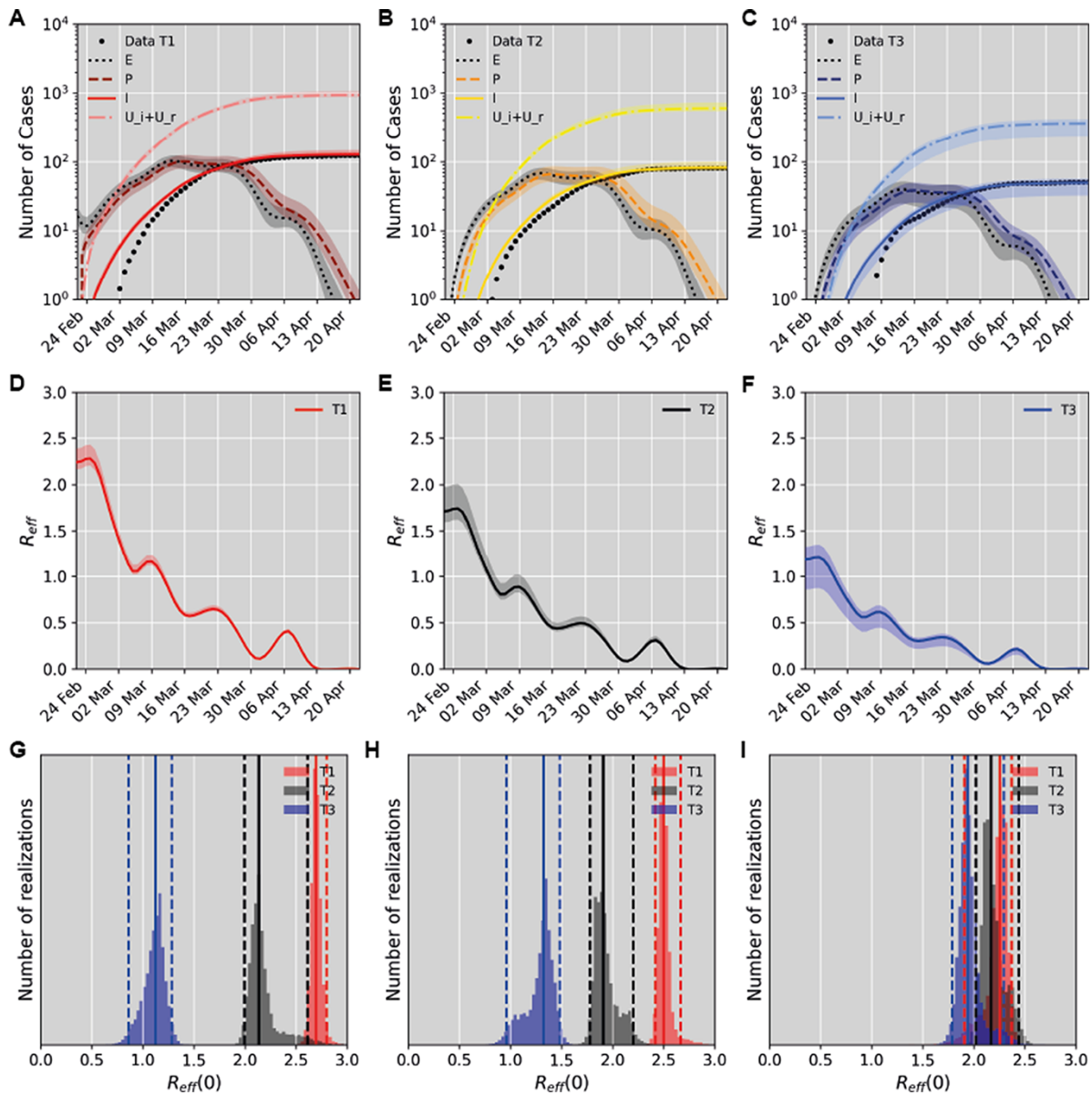


Figure 4. Model fit to the case number time series. A–C) Fit results for a partition based on median income. Data points are shown together with model simulations (median (lines), and 95 per cent confidence bounds (shaded)) for the different tertiles T1 (low, A), T2 (intermediate, B) and T3 (high median income, C). Model compartments E (exposed), P (presymptomatic), I (reported infectious) and $U_i + U_r$ (sum of the unreported infectious and recovered cases) are shown. Fits achieved a root mean squared error of 0.75, 0.49 and 0.34 cases for absolute and 5.3, 3.2 and 2.5 cases for cumulative cases for T1–T3. D–F) The dynamic variation of the effective reproductive number for each of the tertiles shown in A–C. Median values (lines) are shown with 95 per cent confidence bounds (shaded). G–I) Pre-lockdown reproductive number for each socio-economic partition. Parameter distributions obtained using Markov Chain Monte Carlo analysis (histograms) are shown with median values (solid lines) and indicated 95 per cent confidence bounds (dashed lines). Results are shown for partitions based on median income (G), living space per person (H) and share of senior residents (I).

vaccination strategies given in Fig. 5C. Here, as a proof of concept, vaccines that provide 90 per cent transmission reduction and 90 per cent reduction of severe cases are delivered to 23 per cent of the population are presented. For scenario V2, vaccination is prioritised in what we determined as determinants of SARS-CoV-2 transmission—individuals with low income, who have fewer options to socially distance (e.g. by working from home) and hence were more likely to be exposed to and/or transmit the virus (reflected by a higher R_{eff}). With this strategy, the slope of the epidemic curve would be reduced compared to randomly vaccinating the same number of subjects from the whole population. Figure 5D describes the corresponding development of ICU occupancy for the scenarios modelled, revealing that scenario

V2 leads to a delay of approximately 11 days to reach the 50 per cent ICU capacity mark as compared to scenario V1. In scenario V3, resembling the approach by several countries, priority was given to the population group with the highest share of senior residents, which had lower mobility than the rest of the population (Fig. 3B) but constitutes 60 per cent of ICU cases. We observe that scenario V3 resulted in a marginally steeper epidemic curve (Fig. 5C) and would yield 50 per cent ICU capacity at a similar time as a random vaccination strategy (Fig. 5D). However, the total number of cases at this time would be approximately double in scenario V3 compared to V1 (Fig. 5C), whereas the overall peak ICU occupancy would be lowest in V3 (Fig. 5D). These simulations suggest that—in case of vaccines reducing SARS-CoV-2

transmission—vaccination of population groups driving transmission are most effective in reducing the slope of the epidemic curve, whereas vaccination of high-risk groups reduces healthcare system burden in the long term. The presented effects strongly depend on the specific vaccine characteristics and population fraction vaccinated. In Fig. 5E and F we finally show the potential effects of a mixed vaccination scenario (V4) giving equal priority to senior and highly mobile population groups as a representative example. This mixed mode provides a possible compromise with lower case numbers compared to V3, but also delayed and reduced ICU occupancy relative to V1.

4. Discussion

This analysis evaluates complementary aspects of the spread of SARS-CoV-2 within a medium-sized European urban area, including local transmission analysed by phylogenetic tree inference and clustering, and the overall spread described by a compartmental SEIR model enabling simulation of vaccination strategies. The main strength of this study lies in the high degree of diverse and detailed data included as well as the complementary models and analyses.

Patterns of SARS-CoV-2 transmission have previously been discussed from different angles either via network and transmission

Table 1. Achieved level of significance (ALS) of maximum differences in R_{eff} associated with a partition of housing blocks according to various socio-economic indicators. ALSs have been obtained by comparing these differences in R_{eff} with those obtained from 99 bootstrapping random partitions.

Socio-economic indicator	ALS
Living space per person	1 per cent
Median income	2 per cent
Fraction of 1-person households	2 per cent
Fraction of residents aged above 64	45 per cent

modelling (Jay et al. 2020), by statistical evaluation (De Ridder et al. 2021b) or by phylogenetic clustering based on genomic sequencing data (Bluhm et al. 2020). Independent of model choice, the importance of socio-economic factors has been suggested previously (Jay et al. 2020; De Ridder et al. 2021a). However, analyses focused on metropolitan areas only may be biased towards their underlying socio-economic and demographic characteristics making it important to also quantify SARS-CoV-2 transmission in other urban areas, such as in a city context as well as in countries across all continents and stages of economic potential. Modelling studies provide the foundation of scenario predictions and estimation of effective reproductive numbers but require

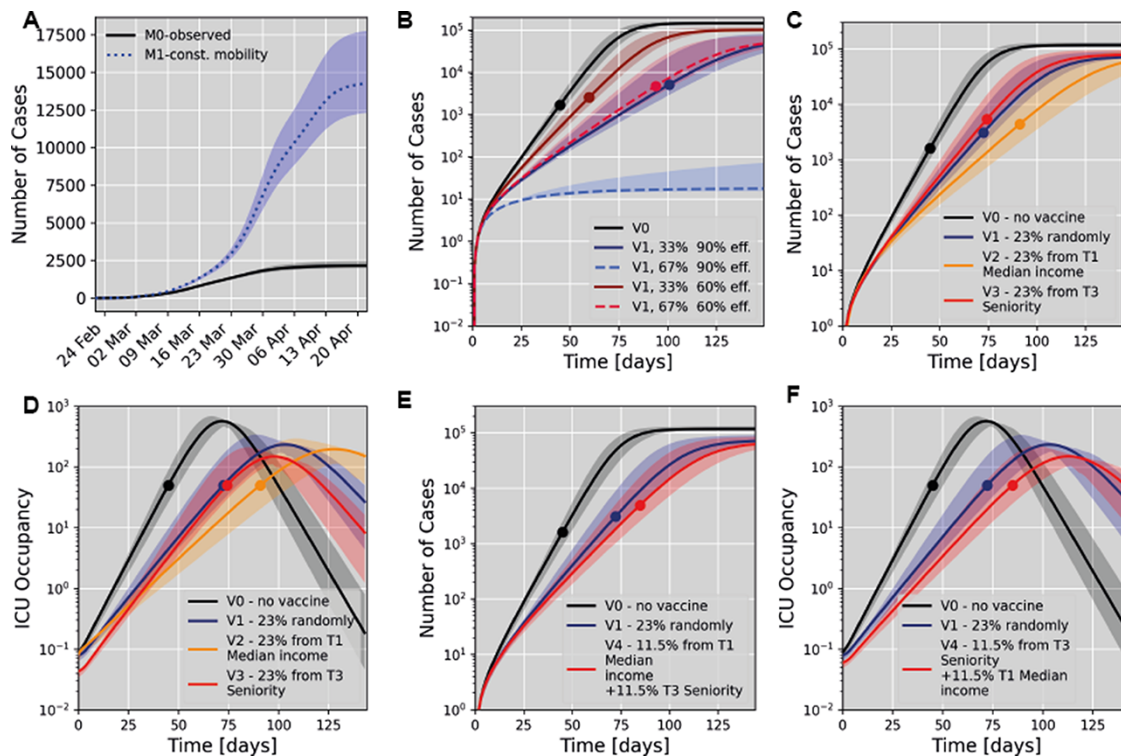


Figure 5. Scenario simulations for a partition based on median income. All subsets A,B,C and E show the total number of infected cases (i.e. the sum of compartments I , U_i , and U_r). A) Influence of the mobility pattern on the total number of infected cases during the first wave (sum of reported and unreported cases) modelling either no change in mobility (no lockdown scenario, M1) or the observed scenario (MO) is shown. B) Simulation of simulated vaccination effects if a specific percentage of all citizens was randomly selected for vaccination at given efficacy (V1). We compare this to the scenario of no vaccine (V0). * Here vaccination of different population fractions at either 60 per cent or 90 per cent efficacy to prevent SARS-CoV-2 transmission were modelled. Median values (lines) and 95 per cent confidence bounds (shaded) are shown. Dots indicate the time of reaching a 50 per cent ICU occupancy. C) Simulation of future vaccination effects based on a partition according to median income. Scenario V2 models vaccination of 23 per cent of all citizens selected from the tertile with the lowest median income (T1). Scenario V3 models vaccination of 23 per cent of all citizens selected from the tertile with the highest share of senior residents (T3). In C–E we model 90 per cent vaccine efficacy (transmission and severe COVID-19) and compare with scenarios V0 and V1. D) Temporal evolution of ICU occupancy for the scenarios modelled in C. E) Simulation of a mixed vaccination strategy giving equal priority to senior citizens and mobile population groups. F) Temporal evolution of ICU occupancy for the scenarios modelled in E).

balancing the trade-off between detail described and the number of data points available. Accordingly, many published models rely on publicly available case numbers without being able to directly relate socio-economic parameters and specific geographic locations per case and/or are only performed for large metropolitan areas (Chang et al. 2020). This has led to biases in evaluations and the neglect of a considerable population share living outside major cities (53 per cent/41 per cent in Europe/worldwide) (United Nations, Department of Economic and Social Affairs, Population Division 2019; European Commission 2021). In this study we chose Basel-City as a case study of a European urban area. Despite the strong economic status of Switzerland and an obligatory health insurance, implying potentially less extreme socio-economic gradients than other countries, we demonstrate that socio-economic background impacts the probability to acquire and transmit SARS-CoV-2. It would be expected that in cities containing more pronounced socio-economic gradients, these disparities would emphasise transmission patterns of SARS-CoV-2 between and within socio-economic groups.

The success of our analyses is based on the optimised choice of evaluation and model, high data-density and -quality, rather than large absolute case numbers. It is difficult to distinguish the spread of competing viral variants within the same population and to account for new introductions in a model since classical ODE or agent-based models are relying on the assumption of uninterrupted transmission chains. Sequencing information is essential to reliably inform on such transmission patterns. Yet given the cost of such analyses, WGS covering entire epidemic waves is often unfeasible. We included 81 per cent of all reported cases in the study time frame and geographic area, which allowed us to restrict our analysis to a subset of 247 phylogenetically related cases consisting of a single SARS-CoV-2 variant. Given the ~200k inhabitants of Basel-City, this implies one of the largest per-capita sequencing densities of reported studies to date. We deliberately choose a simple way to incorporate socio-economic, demographic and mobility information into our compartmental model since more complex network approaches would be unfeasible for limited case numbers. The use of mobility and socio-economic data in our models is unique since we include regularly collected data analysed by the statistical office of Basel-City, providing a high spatial and temporal resolution network of the inner city mobility patterns. In contrast to mobile phone data (Jay et al. 2020; Rader et al. 2020; Kissler et al. 2020; Chang et al. 2020), our data are not subjected to privacy legislation and are hence expected to be more readily available for other medium-sized urban areas around the world, making our analysis transferable. We do not hold information on the duration and specific location of individuals, but a continuum estimate of population mixing that aligns well with the concept of a compartmental ODE model. Mobility and the reduction thereof have been suggested as a proxy to evaluate the reduction of the spread of SARS-CoV-2 (Douglas et al. 2020; Rader et al. 2020; Candido et al. 2020); however, there has also been a change in hygiene practices and social interaction behaviour. In our SEIR model we separate these two contributions, allowing for an easier translation of our model for scenario building. We further addressed unreported cases which were the driving force of infection outside the observed clusters. Our ~73 per cent estimate of unreported cases overall (not limited to B.1-C15324T, i.e. $1 - n_{total} / (P_{infected} * n_{population})$) falls within the range of previous reports within Europe (Gudbjartsson et al. 2020; Burki 2020).

The SEIR model evaluates general trends of transmission, such as effective reproductive numbers, and vaccine scenario building. To complement this, we employ phylogenetic analysis

to identify transmission clusters. This comprehensive evaluation showed that socio-economic brackets characterised by low median income and smaller living space per person were associated with significantly larger effective reproductive numbers. In line with previous results (Jay et al. 2020), we suggest that population groups from a weaker socio-economic background are more mobile and at higher risk for SARS-CoV-2 infection/transmission originating from multiple sources via cryptic transmission. This aligns with the possibility that low socio-economic status may relate to jobs requiring higher personal contact and unavoidable mobility (Reeves and Rothwell 2020), which has been shown to increase the risk of infection by 76 per cent (Rodríguez-Barranco et al. 2021). By contrast, phylogenetic clusters were predominantly discovered within higher socio-economic, or more senior groups, implying a spread within the same social network. It is likely that those individuals are retired, or have had the ability to work from home, a pattern that has been observed also in other cities (Wilson 2020). Effective contact tracing and testing strategies could be most efficient for these groups, which were not driving SARS-CoV-2 transmission.

These results should be accounted for during vaccine prioritisation depending on the relevant vaccine characteristics and ICU capacity available. Our simulation framework provides flexibility to model various scenarios and vaccine efficacies but did not account for fatalities due to limited case numbers during the studied period in Basel-City. In case of a combined effect of vaccines to both prevent SARS-CoV-2 transmission and protect against COVID-19 (Swan et al. 2020; Moghadas et al. 2020; Anderson et al. 2020), vaccination of individuals driving transmission in addition to the protection of high-risk groups would be ideal to arrive at a combined concept of protection from and containment of SARS-CoV-2. Vaccinating high-risk groups reduces the number of hospitalised and ICU patients in the short term, the spread of the pandemic would, however, be more effectively contained by vaccinating the transmission drivers. This finding is in agreement with previous results obtained across different countries (Bubar et al. 2021). By restricting vaccination to only risk groups, a larger fraction of the general population will be exposed to SARS-CoV-2, implying that contact and travel restrictions would remain necessary to contain transmission. Such measures come at great economic cost. Based on our results it would be recommended to follow a combined strategy to employ extensive testing where transmission chains are traceable, e.g. among less mobile population groups, and to combine this with a vaccination strategy aiming to prevent cryptic transmission. It should however, be noted that vaccination prioritisation is a highly political and ethical topic. Identifying mobile population groups may not be based on the mobility behaviour of the individual and similarly discrimination purely based on income has to be prevented, too. It would potentially instead be an option to prioritize based on specific work environment, accounting for the option to work from home and how much interaction with e.g. customers is required.

It is important to clearly state additional assumptions and limitations of our approach to put it in context with previous studies. All data used in this analysis provide the highest level of detail achievable in the setting of an urban area, yet it they are far from individual level data or a large scale population level analysis. The estimate of the fraction of unreported cases and the assumption of it to be constant both over time and between population groups is a simplification that was inevitable in light of the available data. It is assumed that testing rates may be biased towards socio-economic levels. In contrast to other countries, COVID-19 testing costs were covered by the obligatory health insurance in

Switzerland which may reduce, yet not fully prevent testing bias. It would be expected that the reported difference in transmission between socio-economic groups may indeed be stronger than reported in this analysis. Importantly, we observed no large differences in the ratios of performed tests over the population within tertiles of different median incomes: 0.046, 0.04 and 0.035 for T1, T2 and T3, respectively. This implies only a small systematic sampling bias. Moreover, we did not account for bias in the number of unreported cases with respect to socio-economic indicators. This includes possible variation in the fraction of unreported cases in general, as well as variation in the sequencing success rate which is dictated by the patient's viral load. To address this, we confirmed that for a partitioning based on median income the sequencing success rates were comparable among tertiles (T1: 0.76, T2: 0.67, T3: 0.77).

The choice of a continuum model excluded the possibility to account for stochastic superspreading events since cases are modelled in a successive fashion that only ends with isolation (reported cases) or recovery (unreported cases). Our WGS analysis showed that, despite initial effects of mass gatherings (Stange et al. 2021), nucleotide diversity evolved over time indicating successive transmission chains. The relevant infectious and incubation times, as well as the reproductive numbers represent an average across all of these events. We further also did not account for continuous imports of cases. However, the restriction to a single variant predominantly reported in Basel-city 'limited' the impact of case im-/exports. Realistically, cases of the B.1-C15324T variant could have been re-introduced from commuters entering Basel. At the same time, the SEIR model assumes that Basel citizens would have only infected others within the city and it could be suggested that the aforementioned im- and export effects cancel each other out to some extent. We did not perform our analysis based on all reported cases (independent of the respective variant) since during the first wave, returning travellers were quickly isolated and several small transmission chains were quickly interrupted. Only the B.1-C15324T variant dominated throughout the relevant observation period. We conclude that the arguments in favour of using inherently related cases outweigh those for increasing the number of included cases to improve statistical power.

Finally, we assume constant mobility and social interaction behaviour changes across all tertiles. It was confirmed that the drop in population mobility was comparable within different quarters of Basel-City; however, we were unable to estimate if social interaction varied. As such, this remains a fundamental limitation of our approach. Analysis of e.g. mobile phone data (Chang et al. 2020) could help to evaluate this aspect to some extent.

Despite these limitations, we were able to obtain comparable results in terms of the impact of mobility and socio-economic status as previously reported, motivating the application of our approach for meaningful scenario building. Alternative modelling approaches, e.g. phylodynamic birth–death skyline models (Lai et al. 2020), could provide an elegant additional angle to combine phylogenetic analysis with the estimation of effective reproductive numbers. Such methods have previously been applied by several groups to model viral transmission (Stadler et al. 2013; Ratmann et al. 2019; Seemann et al. 2020) and should be considered for future evaluation of this rich data set.

In conclusion, high-resolution city-level epidemiological studies are essential for understanding factors affecting pandemic transmission chains and thereby supporting tailored public health information campaigns and vaccination distribution strategies at

the municipal level. We here provided an example of such an analysis within a representative medium-sized European city at the core of the Greater Basel area and part of the Upper Rhine Region Metropolitan Economy, which suggests that the findings and modelling approaches presented may be readily translated to other such areas.

Data Availability

The SEIR model code used for this submission will be available on <https://github.com/BorgwardtLab/BaselEpi.git>. Code that was used for phylogenetic inference and calculation of significance of clusters in specified groups is available at <https://github.com/appliedmicrobiologyresearch>. SARS-CoV-2 whole genomes from Basel-City are available at gisaid.com and at European Nucleotide Archive under accession number PRJEB39887.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Acknowledgement

We greatly appreciate the input and data received from Construction and Traffic department Canton Basel-City, Baselland Transport AG, Basler Verkehrs-Betriebe, Autobus AG Liestal, SBB Federal Railways and Statistical Office of the Canton of Basel-City and want to specifically thank Björn Lietzke, Lukas Mohler and Madeleine Imhof (all Statistical Office of the Canton of Basel-City) from Construction and Traffic Department of the Canton of Basel-City Michael Redle and Kathrin Grotrian, Matthias Hofmann (Basler Verkehrs-Betrieb), Roman Stingelin (Autobus AG), Nadine Ruch (SBB AG) and Stefan Burtschi (Baselland Transport AG) for their support. We thank Christine Kiessling, Magdalena Schneider, Elisabeth Schultheiss, Clarisse Straub and Rosa-Maria Vesco (University Hospital Basel) for excellent technical assistance with next-generation sequencing. Computations were performed at sciCORE (<http://scicore.unibas.ch>) scientific computing facility at the University of Basel. Data exchange was organised via the BioMedIT node between the University of Basel and ETH Zurich, Department of Biosystem Science and Engineering. We thank all authors, who have shared their genomic data on GISAID, especially the Stadler Lab from ETH Zurich for sharing Swiss sequences. A full table (csv) outlining the originating and submitting labs is included as a supplementary file. We finally thank Dr. A. Jermy (Geminate Science Consulting) for his critical review of the manuscript.

Funding

This study was supported by the Alfried Krupp Prize for Young University Teachers of the Alfried Krupp von Bohlen und Halbach-Stiftung (K.B.). SCB was supported by the Swiss National Science Foundation (Spark 190647).

Competing Interests. The authors declare that they have no competing financial interests.

Correspondence

Correspondence and requests for materials should be addressed to Adrian Egli, adrian.egli@usb.ch, Clinical Bacteriology and Mycology, University Hospital Basel, Petersgraben, CH-4031 Basel, Switzerland, Phone: +41 61 556 57 49.

Ethical statement

Ethical approval was given by the local ethical committee Ethik Kommission Nordwest und Zentralschweiz (EKNZ No. 2020-00769, to be found at <https://ongoingprojects.swissethics.ch>) and the project was registered at clinicaltrials.gov under NCT04351503.

Author Contributions

A.E., H.H.H. and K.B. devised the project. S.C.B. and J.K. developed and performed the mathematical modelling. M.St. performed and interpreted phylogenetic analyses. S.C.B., J.K. and M.St. led the writing and revising of the report. A.M. provided the genome assembly pipeline. T.C.R. and H.S.S. prepared viral RNA for sequencing. My.B. and R.S.S. provided geographical expertise. R.S.S. provided detailed background on the representative status of Basel-City. K.K.S. collected clinical data. K.R., D.A.T., A.G., A.K.S., M.Sch. analysed serology samples. D.C. and O.D. provided serology samples from Viollier AG. K.L. provided virological expertise. A.B. provided serology samples from the blood transfusion service. J.B., S.T.S. and S.F. provided public health and epidemiological expertise. H.P., M.Si., C.H.N., R.B. and M.B. provided clinical expertise and valuable discussion on the results. N.R., U.H. and J.B. provided epidemiological expertise. All author reviewed and edited the manuscript.

References

- Anderson, R. M., Vegvari, C., Truscott, J. and Collyer, B. S. (2020), 'Challenges in creating herd immunity to SARS-CoV-2 infection by mass vaccination', *The Lancet* 396: 1614–1616.
- AstraZeneca (2020), 'AZD1222 vaccine met primary efficacy endpoint in preventing COVID-19'.
- Bau- und Verkehrsdepartement Basel-Stadt Mobilität / Mobilitätsstrategie (2020), 'Gesamtverkehrsmodell der Region Basel, Basismodell: "Ist-Zustand 2016"'.
- Bluhm, A., Christandl, M., Gesmundo, F., Ravn Klausen, F., Mančinska, L., Steffan, V., Stilck França, D. and Werner, A. H. (2020), 'SARS-CoV-2 transmission routes from genetic data: A Danish case study', *PLOS ONE* 15: e0241405.
- Bubar, K. M., Reinholt, K., Kissler, S. M., Lipsitch, M., Cobey, S., Grad, Y. H. and Larremore, D. B. (2021), 'Model-informed COVID-19 vaccine prioritization strategies by age and serostatus', *Science* 371: 916–921.
- Burki, T. (2020), 'Mass testing for COVID-19', *The Lancet Microbe* 1: e317.
- Bushman, D. et al. (2020), 'Detection and Genetic Characterization of Community-Based SARS-CoV-2 Infections – New York City, March 2020', *MMWR. Morbidity and mortality weekly report* 69: 918–922.
- Candido, D. S., et al. (2020), 'Evolution and epidemic spread of SARS-CoV-2 in Brazil', *Science (New York, N.Y.)* 369: 1255–1260.
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D. and Leskovec, J. (2020), 'Mobility network models of COVID-19 explain inequities and inform reopening', *Nature*.
- Chinazzi, M. et al. (2020), 'The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak', *Science* 368: 395–400.
- De Ridder, D., Sandoval, J., Vuilleumier, N., Azman, A. S., Stringhini, S., Kaiser, L., Joost, S. and Guessous, I. (2021a), 'Socioeconomically disadvantaged neighborhoods face increased persistence of sars-cov-2 clusters', *Frontiers in Public Health* 8: 1091.
- De Ridder, D., Sandoval, J., Vuilleumier, N., Azman, A. S., Stringhini, S., Kaiser, L., Joost, S. and Guessous, I. (2021b), 'Socioeconomically Disadvantaged Neighborhoods Face Increased Persistence of SARS-CoV-2 Clusters', *Frontiers in Public Health* 8.
- Douglas, J. et al. (2020), 'Phylogenetics reveals the role of human travel and contact tracing in controlling COVID-19 in four island nations', *medRxiv*.
- Du, P. et al. (2020), 'Genomic surveillance of COVID-19 cases in Beijing', *Nature communications* 11: 5503.
- Egli, A. et al. (2020), 'High-resolution influenza mapping of a city reveals socioeconomic determinants of transmission within and between urban quarters', *bioRxiv*.
- Elbe, S. and Buckland-Merrett, G. (2017), 'Data, disease and diplomacy: GISAID's innovative contribution to global health', *Global Challenges* 1: 33–46.
- European Commission (2021), 'GDP per capita, consumption per capita and price level indices'. https://ec.europa.eu/eurostat/statistics-explained/index.php/GDP_per_capita,_consumption_per_capita_and_price_level_indices.
- EUROSTAT (2011), 'Population Data Collection for European Local Administrative Units from 1960 onward'. <https://ec.europa.eu/eurostat/web/nuts/local-administrative-units>, accessed 18 February 2021.
- Foreman-Mackey, D., Hogg, D. W., Lang, D. and Goodman, J. (2013), 'emcee : The MCMC Hammer', *Publications of the Astronomical Society of the Pacific* 125: 306–312.
- Gudbjartsson, D. F., et al. (2020), 'Spread of SARS-CoV-2 in the Icelandic Population', *New England Journal of Medicine* 382: 2302–2315.
- Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T. and Neher, R. A., Kelso, J. (2018), 'NextStrain: Real-time tracking of pathogen evolution', *Bioinformatics* 34: 4121–4123.
- He, X. et al. (2020), 'Temporal dynamics in viral shedding and transmissibility of COVID-19', *Nature Medicine* 26: 672–675.
- Huerta-Cepas, J., Serra, F. and Bork, P. (2016), 'ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data', *Molecular Biology and Evolution* 33: 1635–1638.
- Jay, J., Bor, J., Nsoesie, E. O., Lipson, S. K., Jones, D. K., Galea, S. and Raifman, J. (2020), 'Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States.', *Nature human behaviour*. 4 1294–1302.
- Kalman, R. E. et al. (1960), 'Contributions to the theory of optimal control', *Bol. soc. mat. mexicana* 5: 102–119.
- Kanton Basel-Stadt Datenportal (data.bs.ch 2021) (2021), 'Coronavirus (COVID-19): Fallzahlen Basel-Stadt'. <https://data.bs.ch/explore/dataset/100073>, accessed Feb 2021.
- Kissler, S. M. et al. (2020), 'Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City', *Nature communications* 11: 4674.
- Koo, J. R., Cook, A. R., Park, M., Sun, Y., Sun, H., Lim, J. T., Tam, C. and Dickens, B. L. (2020), 'Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study', *The Lancet Infectious Diseases* 20: 678–688.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. and Marra, M. A. (2009), 'Circos: an information aesthetic for comparative genomics', *Genome research* 19: 1639–45.
- Lai, A., Bergna, A., Acciarri, C., Galli, M. and Zehender, G. (2020), 'Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2', *Journal of Medical Virology* 92: 675–679.
- Lemieux, J. et al. (2020), 'Phylogenetic analysis of SARS-CoV-2 in the Boston area highlights the role of recurrent importation and

- superspreading events', *medRxiv : the preprint server for health sciences* 12: 15.
- Leuzinger, K. et al. 'Epidemiology of Severe Acute Respiratory Syndrome Coronavirus 2 Emergence Amidst Community-Acquired Respiratory Viruses.', *The Journal of infectious diseases*.
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W. and Shaman, J. (2020), 'Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)', *Science* 368: 489–493.
- Long, S. W. et al. (2020), 'Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area.', *medRxiv*.
- Mahase, E. (2020), 'Covid-19: Moderna vaccine is nearly 95% effective, trial involving high risk and elderly people shows', *BMJ: British Medical Journal (Online)* **371**.
- Mansilla Domínguez, J. M., Font Jiménez, I., Belzunegui Eraso, A., Peña Otero, D., Díaz Pérez, D. and Recio Vivas, A. M. (2020), 'Risk Perception of COVID-19 Community Transmission among the Spanish Population.', *International journal of environmental research and public health* 17.
- Maurano, M. T. et al. (2020), 'Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City region.', *Genome research* 30: 1781–1788.
- Moghadas, S. M., Vilches, T. N., Zhang, K., Wells, C. R., Shoukat, A., Singer, B. H., Meyers, L. A., Neuzil, K. M., Langley, J. M., Fitzpatrick, M. C. and Galvani, A. P. (2020), 'The impact of vaccination on COVID-19 outbreaks in the United States.', *medRxiv : the preprint server for health sciences*.
- Nabil, B., Sabrina, B. and Abdelhakim, B. (2020), 'Transmission route and introduction of pandemic SARS-CoV-2 between China, Italy, and Spain.', *Journal of medical virology*.
- Newville, M., Stensitzki, T., Allen, D. B. and Ingargiola, A. (2014), 'LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python'.
- Popa, A. et al. (2020), 'Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2.', *Science translational medicine* 12.
- Post, L. A. et al. (2020), 'A SARS-CoV-2 Surveillance System in Sub-Saharan Africa: Modeling Study for Persistence and Transmission to Inform Policy', *Journal of medical Internet research* 22: e24248.
- Puenpa, J., Suwannakarn, K., Chansaenroj, J., Nilyanimit, P., Yorsaeng, R., Auphimai, C., Kitphati, R., Mungaomklang, A., Kongklieng, A., Chirathaworn, C., Wanlapakorn, N. and Poovorawan, Y. (2020), 'Molecular epidemiology of the first wave of severe acute respiratory syndrome coronavirus 2 infection in Thailand in 2020', *Scientific Reports* 10.
- Rader, B. et al. (2020), 'Crowding and the shape of COVID-19 epidemics', *Nature Medicine* 26: 1829–1834.
- Ragonnet-Cronin, M., Hodcroft, E., Hué, S., Fearnhill, E., Delpech, V., Brown, A. J. and Lycett, S. (2013), 'Automated analysis of phylogenetic clusters', *BMC Bioinformatics* 14: 317.
- Ratmann, O. et al. (2019), 'Inferring HIV-1 transmission networks and sources of epidemic spread in Africa with deep-sequence phylogenetic analysis', *Nature Communications* 10: 1411.
- Reeves, R. and Rothwell, J. (2020), 'Class and covid: How the less affluent face double risks. brookings (March 27)'.
- Rodríguez-Barranco, M., Rivas-García, L., Quiles, J. L., Redondo-Sánchez, D., Aranda-Ramírez, P., Llopis-González, J., Sánchez Pérez, M. J. and Sánchez-González, C. (2021), 'The spread of SARS-CoV-2 in Spain: Hygiene habits, sociodemographic profile, mobility patterns and comorbidities', *Environmental Research* 192: 110223.
- Salje, H. et al. (2020), 'Estimating the burden of SARS-CoV-2 in France.', *Science (New York, N.Y.)* 369: 208–211.
- Seemann, T. et al. (2020), 'Tracking the COVID-19 pandemic in Australia using genomics', *Nature Communications* **11**: 4376.
- Shu, Y. and McCauley, J. (2017), 'GISAID: Global initiative on sharing all influenza data - from vision to reality', *Eurosurveillance* 22: 2–4.
- Stadler, T., Kuhnert, D., Bonhoeffer, S. and Drummond, A. J. (2013), 'Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV)', *Proceedings of the National Academy of Sciences* 110: 228–233.
- Stange, M. et al. (2021), 'SARS-CoV-2 outbreak in a tri-national urban area is dominated by a B.1 lineage variant linked to a mass gathering event', *medRxiv; accepted for publication in PLOS Pathogens* p.2020.09.01.20186155.
- Statistisches Amt Basel-Stadt: Bevölkerungsstatistik (2021), 'Statistisches Jahrbuch des Kantons Basel-Stadt: Wohnbevölkerung nach Geburtsjahr, Heimat und Geschlecht [t01.1.14]'.
- Swan, D. A. et al. (2020), 'Vaccines that prevent SARS-CoV-2 transmission may prevent or dampen a spring wave of COVID-19 cases and deaths in 2021', *medRxiv*.
- United Nations, Department of Economic and Social Affairs, Population Division (2019), *World Urbanization Prospects: The 2018 Revision*, United Nations, New York.
- Welch, G., Bishop, G. et al. (1995), 'An introduction to the kalman filter'.
- Wilson, C. (2020), 'These graphs show how covid-19 is ravaging New York City's low-income neighborhoods', *Time*.
- Worobey, M., Pekar, J., Larsen, B. B., Nelson, M. I., Hill, V., Joy, J. B., Rambaut, A., Suchard, M. A., Wertheim, J. O. and Lemey, P. (2020), 'The emergence of SARS-CoV-2 in Europe and North America', *Science (New York, N.Y.)* 370: 564–570.
- Yuan, F., Wang, L., Fang, Y. and Wang, L. (2020), 'Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity', *Transboundary and emerging diseases*.