Characterising the epidemic spread of influenza A/H3N2 within a city through phylogenetics

Journal Article

Author(s):
Müller, Nicola F.; Wüthrich, Daniel; Goldman, Nina; Sailer, Nadine; Saalfrank, Claudia; Brunner, Myrta; Augustin, Noémi; Seth-Smith, Helena M.B.; Hollenstein, Yvonne; Syedbasha, Mohammedyaseen; Lang, Daniela; Neher, Richard A.; Dubuis, Olivier; Naegele, Michael; Buser, Andreas; Nickel, Christian H.; Ritz, Nicole; Zeller, Andreas; Lang, Brian Matthew (ID); Hadfield, James; Bedford, Trevor; Battegay, Manuel; Schneider-Sliwa, Rita; Egli, Adrian; Stadler, Tanja (ID)

PLOS PATHOGENS

Characterising the epidemic spread of influenza A/H3N2 within a city through phylogenetics

Nicola F. Müller[1,2]*, Daniel Wüthrich[2,3,4], Nina Goldman[5], Nadine Sailer[5], Claudia Saalfrank[5], Myrta Brunner[5], Noémi Augustin[5], Helena MB Seth-Smith[2,3,4], Yvonne Hollenstein[3,4], Mohammedyaseen Syedbasha[3,4], Daniela Lang[3,4], Richard A. Neher[2,6], Olivier Dubuis[7], Michael Naegele[7], Andreas Buser[8], Christian H. Nickel[9], Nicole Ritz[10], Andreas Zeller[11], Brian M. Lang[1,2], James Hadfield[12], Trevor Bedford[12], Manuel Battegay[4,13], Rita Schneider-Sliwa[5], Adrian Egli[3,4◦]*, Tanja Stadler[1,2◦]*

1 ETH Zürich, Department of Biosystems Science and Engineering, 4058 Basel, Switzerland, 2 Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland, 3 Clinical Bacteriology and Mycology, University Hospital Basel, Basel, Switzerland, 4 Department of Biomedicine, University of Basel, Basel, Switzerland, 5 Human Geography, Department of Environmental Sciences, University of Basel, Basel, Switzerland, 6 Biozentrum, University of Basel, Basel, Switzerland, 7 Viollier AG, Allschwil, Switzerland, 8 Regional Blood Transfusion Service, Swiss Red Cross, Basel, Switzerland, 9 Emergency Department, University Hospital Basel, Basel, Switzerland, 10 Pediatric Infectious Diseases and Vaccinology, University Children's Hospital Basel and University of Basel, Basel Switzerland, 11 Institute for Family Medicine, University of Basel, Basel, Switzerland, 12 Fred Hutchinson Cancer Research Center, Seattle, USA, 13 Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel, Switzerland

◦ These authors contributed equally to this work.
* nicola.felix.mueller@gmail.com (NFM); tanja.stadler@bsse.ethz.ch (TS)

Abstract

Infecting large portions of the global population, seasonal influenza is a major burden on societies around the globe. While the global source sink dynamics of the different seasonal influenza viruses have been studied intensively, its local spread remains less clear. In order to improve our understanding of how influenza is transmitted on a city scale, we collected an extremely densely sampled set of influenza sequences alongside patient metadata. To do so, we sequenced influenza viruses isolated from patients of two different hospitals, as well as private practitioners in Basel, Switzerland during the 2016/2017 influenza season. The genetic sequences reveal that repeated introductions into the city drove the influenza season. We then reconstruct how the effective reproduction number changed over the course of the season. While we did not find that transmission dynamics in Basel correlate with humidity or school closures, we did find some evidence that it may positively correlated with temperature. Alongside the genetic sequence data that allows us to see how individual cases are connected, we gathered patient information, such as the age or household status. Zooming into the local transmission outbreaks suggests that the elderly were to a large extent infected within their own transmission network. In the remaining transmission network, our analyses suggest that school-aged children likely play a more central role than pre-school aged children. These patterns will be valuable to plan interventions combating

the spread of respiratory diseases within cities given that similar patterns are observed for other influenza seasons and cities.

Author Summary

As shown with the current SARS-CoV-2 pandemic, respiratory diseases can quickly spread around the globe. While it can be important to understand how diseases spread around the globe, local spread is most often the main driver of novel infections of respiratory diseases such as SARS-CoV-2 or influenza. We here use genetic sequence data alongside patient information to better understand what drives the local spread of influenza by looking at the 2016/2017 influenza season in Basel, Switzerland as an example. The genetic sequence data allows us to reconstruct how the transmission dynamics changed over the course of the season, which we then compare to trends in humidity and temperature and times when schools were open or closed. Additionally, the genetic sequence data allows us to see how individual cases are connected. Using patient information, such as age and household status our analyses suggest that the elderly mainly transmit within their own transmission network. Additionally, they suggest that school aged children, but not necessarily pre-school aged children are important drivers of the local spread of influenza.

Introduction

With a large fraction of the population being infected annually and up to 650,000 deaths per year, seasonal influenza causes a major burden on societies around the globe (https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)). Through rapid evolution, influenza strains evade host immunity, allowing them to reinfect large fractions of a population every year. In order to prevent infections, limited public health resources have to be streamlined as efficiently as possible [1]. The planning of interventions is dependent upon knowledge of the dynamics of epidemic spread of influenza viruses in a city environment, which includes understanding the drivers of the spread of seasonal influenza between individuals. Incidence and prevalence data can be used to some extent to infer such dynamics. However, they lack the information about how individual cases are epidemiologically related.

Phylogenetics allows us to see how individual cases are epidemiologically connected. This is done by reconstructing the evolutionary relationship between temporally spaced samples of genetic sequence data, isolated from different infected individuals. The resulting phylogenetic tree displays how samples are related to each other, and branch lengths in calendar time display the elapsed time. The phylogenetic tree can therefore be interpreted as an approximation of the transmission chain of the sampled cases. Such a view on part of the influenza transmission chain allows to further quantify the epidemiological dynamics which gave rise to the observed phylogenetic tree using phylodynamic methods [2]. Phylogenetics and phylodynamics thus allows us to elucidate past epidemiological dynamics [3, 4] or to infer migration patterns [5, 6].

Several studies have used phylogenetic approaches to study how influenza and its subtypes spread globally [7–11]. On an intermediate scale, college campuses have been studied by using phylogenetics, revealing extensive mixing of influenza strains [12]. On the smallest scale, studies have been performed to investigate person-to-person transmission of influenza in households [13]. There is, however, a gap in studies that seek to describe transmission of influenza

on a city scale. In contrast to college campuses, cities constitute highly heterogeneous societies with various different living arrangements and vastly different social and age groups. This means that lessons learnt about influenza transmission from college campuses are not necessarily transferable to cities. Children, for example, have been repeatedly described to be over proportionally affected by influenza. During the 2009 Influenza A/H1N1 pandemic, school-aged children have been shown to have the highest seroprevalence of all age groups in the USA [14]. A study on the incidence of seasonal influenza A/H3N2 in different age groups found that incidence of influenza A/H3N2 was highest in children, without a strong difference between school and preschool children [15]. A seroconversion study with samples collected between 2009 and 2011 found strong age dependency for H1N1, but not H3N2 [16].

In an effort to fill that gap, we studied the local spread of influenza and the factors contributing to it in the city of Basel, Switzerland during the 2016/2017 Influenza season which was dominated by Influenza A/H3N2. To do so, influenza samples together with the age and residential address were collected from around 663 patients from the University Hospital (USB), the Children's Hospital of Basel (UKBB) and patients of private practices from around the city. 85% of these sequences were from the USB and the UKBB and the rest from private practitioners around the city. Since these patients were sick enough to seek medical help, our dataset will represent a sub sample of the overall population that was infected with influenza and experienced more severe symptoms.

Around 200 of all patients also provided additional information through filling out a survey. The survey asked questions about family status, financial status, and demographics. Details on the data collection are provided in [17]. The spatial and survey data were analyzed in a different study [18]. We here assess the importance of introductions of influenza into a city for seeding a seasonal epidemic, the overall dynamics of transmission throughout the season, and explore the impact of different age groups on the epidemic.

Methods and material

Data collection and sequencing

We collected all data in the 2016/17 influenza season as described in [17]. Sequencing was performed as described in [19]. Raw Illumina reads were trimmed with Trimmomatic 0.36 [20]. Alignment of paired-end reads was done by using bowtie 2.2.3 [21], using strain A/New York/18/2014 as a reference. The aligned reads were sorted by using samtools 1.2 [22]. Variants were called and filtered by using lofreq 2.1.2 [23]. Variant calling was done for sites with a coverage of at least 100. Sites with a coverage of less than 100 were assumed to be unknown and were denoted as N, that is any possible nucleotide (Details on the exclusion of sequences are described in S1 Text). Exact input specification can be found at https://github.com/nicfel/FluBaselPhylo/tree/master/Sequences. The consensus sequences from this study were deposited in GenBank (numbers MN299375-MN304713).

Timed phylogenetic tree based on the HA segment

We combined the Basel sequences with all sequences (as of the 17th of Juli 2018) from https://www.gisaid.org sampled between January 1st 2016 and December 31st 2017 for which at least the segments HA, NA and MP were available. We first aligned all consensus sequences using muscle v3.8.3129. We then built an initial phylogeny from the HA segment alone by using RaxML 8.2.1 [24] and obtained branch lengths in calendar time via timetree [25] using the nextstrain pipeline [26]. The inferred time tree is available here https://github.com/nicfel/FluBaselPhylo/tree/master/TimeTree. The tree was plotted using ggtree [27].

Initial clustering based on nucleotide differences

We then calculated the average nucleotide difference between any of the sequences and sequences from Basel. In order to split the dataset into manageable pieces, we first grouped any two sequences from Basel together if they were within an average nucleotide difference of 0.0025 per position. If the full genome for two sequences was available, this would correspond to about 32 different positions on the full genome. For an average clock rate of $2.9*10^{-3}$ per site and year, this would correspond to a pairwise phylogenetic distance of just below 1 year. Sets of sequences from Basel are only split into two groups if the two closest related sequences of each group exceeds this distance. Based on this initial grouping, we added sequences that were not from Basel to each cluster if they were at maximum 0.0025/2 mutations per position away from any of the sequences from Basel. The factor of 2 is only to reduce the number of non-Basel sequences in each of these initial clusters to be computationally tractable.

Phylogenetic trees of initial clusters

We next estimated rates of evolution for each genomic segment using the SRD06 model [28] and a strict clock model from 200 full genome influenza A/H3N2 sequences sampled in California, New York and Europe between 2010 and 2015 in Beast 2.5 [29]. These sequences were downloaded from fludb.org, were not used otherwise and are an independent dataset. We allowed each segment to have its own phylogeny in order to avoid reassortment to bias the estimates of evolutionary rates. Each of the segments, as well as the first two and third codon position was allowed to have its own rate scaler. We ran 10 independent MCMC chains each for 10^8 iterations and then combined them after a burn-in of 10%. These estimated evolutionary rates are long-term rates of influenza A/H3N2. Since the effects of selection over short time periods are smaller compared to longer time periods. The evolutionary rates can be expected to be faster for shorter time windows [30]. We therefore expect the pairwise distances estimated for our data from the 2016/17 outbreak using these rates to be an overestimate of the actual divergence times. The xml and log files for the analysis can be found here https://github.com/nicfel/FluBaselPhylo/tree/master/EvolutionaryRates.

We next reconstructed the phylogenetic trees of all initial clusters by using the full genomes of all samples in the initial clusters. We fixed the evolutionary rates to be equal to the mean evolutionary rates as estimated previously, with the mean evolutionary rate being $2.9*10^{-3}$ per site and year, as well as fixed the rates of the SRD06 site model to the rates estimates using the influenza A/H3N2 datasets samples over many years. As a population prior, we used a constant coalescent model with an effective population size being shared among all initial clusters. We then estimated a distribution of phylogenies for each initial cluster, assuming that all segments share the same phylogeny. If reassortment happened, it would increase the distance between samples. Due to using fixed evolutionary rates as estimated in the previous analysis, reassortment will not bias evolutionary rates. Hence, reassortment events will increase the pairwise distance between isolates separated by reassortment, but will not bias the distance between isolates that are not.

Local cluster identification

To identify sets of sequences from Basel that were likely transmitted locally, we used the phylogenetic tree distributions for each initial cluster and reconstructed the ancestral states using parsimony. We made some modifications to the standard algorithm for ancestral state reconstruction. To reflect our prior belief that Basel is unlikely to act as a relevant source of influenza on a global scale, we classified internal nodes that are not exclusively classified to be in Basel as not in Basel. Since the flu season is only a few months long, we additionally assumed that

lineages are unlikely to persist in Basel for more than 0.1 years without being sampled. To reflect that assumption, we classified internal nodes that are more than 0.1 years from a sample from Basel to be either in a location other than Basel, or to be in an unknown location. We then defined sequences to be in the same local cluster if all their ancestors are inferred to be in Basel. We get these local clusters for each iteration of the MCMC. From the grouping of sequences into local clusters as described above, sequences can be classified into different local clusters over the course of the MCMC. For the estimation of effective reproduction rates we however require each sequence to be in a distinct local clusters. To do so, we randomly picked an iteration of the MCMC and then chose the local clusters present in that iteration. In order to account for uncertainty in the local transmission cluster assignment, we repeated each analysis 10 times with randomly chosen iterations. The exact workflow, including BEAST2 input files can be found at https://github.com/nicfel/FluBaselPhylo/tree/master/LocalClusters. While alternative model based approaches exist to reconstruct locations of internal nodes (e.g. [31]), these approaches themselves make strict assumptions that are violated when studying the spread of diseases on a city scale. Also, it is unknown how well they perform when migration between individual locations is very strong.

Estimation of the effective reproduction number and sampling probability

We then estimated the effective reproduction number through time as well as the sampling proportions and phylogenies from all these local clusters jointly using BDSKY [32]. We assumed the effective reproduction number to be piecewise constant in intervals of 2 days and allowed it to change every 2 days. We then assumed the difference between the log effective reproduction number in interval t (log $R_{eff}(t)$) and in interval t-1 (log $R_{eff}(t-1)$) to be distributed around $N(0,\sigma)$, with σ being estimated in the MCMC [33]. Additionally, we assume the log R_{eff} at the most recent time interval and the one at the very last time interval to be normally distributed in log space around $N(-0.6931,0.1)$. This means that we assume the R_{eff} to change in a continuous way, which can lead to an underestimation of differences in R_{eff}, if the R_{eff} changes abruptly. This adapted version of BDSKY is available on https://github.com/nicfel/bdsky.

We assumed the rate at which an infected individual transitions to being non-infectious to be 0.25 per day. The birth-death model assumes the number of samples over time to be informative about the population dynamics, meaning that the results can be biased if the sampling proportion of individuals would change over time. Since we, however, followed the same procedure for inclusion of patients throughout the epidemic season, our assumption of sampling over time should hold. The BDSKY model additionally conditions on survival [32], meaning that it computes the probability of observing a phylogenetic tree conditional on observing at least one lineage and assumes the host population to be unstructured. We further assume, as is standard, that there is no transmission rate variability between individuals, such as would, for example, be caused by having super spreaders. Additionally, we do not model the process of how lineages are introduced into Basel and how this might change over time.

The weather data used for the correlation analysis was obtained from www.meteoblue.com. This data is based on measurements of weather stations which are then used in simulations to estimate local weather variables (see https://content.meteoblue.com/nl/specifications/data-sources).

Defining connectedness between individuals

We define two individuals to be connected if their pairwise phylogenetic distance is less than 0.1 years. If we assume two individuals to be isolated at the same time, this cutoff would

correspond to a common ancestor that was at most 18.25 days ago. Considering that the evolutionary rates we used to perform these inferences are long term rates and therefore lower than the actual short term rates [30], we expect that the cutoff values are effectively lower in reality. This means that if we use a cutoff of 0.1 years, even individuals that are at an inferred pairwise distance of 0.1 years are very likely more closely related than that. To avoid biases originating from these cutoff values, we repeated all analyses that are based on cutoffs with of 0.05, 0.15, 0.2 and 0.3 years as the cutoff value.

Connectedness across age and family status groups

We estimate the average number of connections members from each of the six categorial age/family status groups have, according to the above definition of connectedness. To do so, we model the number of connections an individual from a group has as a negative binomial distribution. This allows us to model the number of connections an individual from a group has, while taking the variance of the relationship between the group label and the number of connections into account. This is in contrast to, for example, the poisson or geometric models. We assess overall model fit with an ANOVA and then perform Tukey contrasts, comparing all pairs of age groupings [34]. We correct for multiple testing by using Schaffer's method, which is similarly conservative to bonferroni, but takes into account the dependencies enforced in a linear modelling framework [35].

Age mixing patterns

To identify mixing patterns between the six categorial age/family status groups, we again use the definition of connection of two patients.

We use two different approaches to estimate how different groups are connected to each other. First, we use multinomial logistic regression to estimate the probability that a member from one group is connected to a member from another group. As weights, we use the inverse number of samples from each group. This implicitly assumes that individuals from each group have the same probability of being infected. Children however might have higher rates of infection, and we therefore expect this weighting to underestimate the role of children and to overestimate the role of adults.

Second, we use a permutation approach. Between any two groups a and b, we compute the probability of them being associated with one another as follows: For each combination of groups a and b, we count the number of pairs that are associated with one another. We then randomly permute the age labels 10^6 times. For each permutation, we calculate if the number of pairs between these groups is greater or smaller than what we observed. From these values, we then compute the probability that age groups a and b are positively (P_{ab}^+) or negatively (P_{ab}^-) associated with one another as:

$$P_{ab}^+ \quad = \quad \frac{1}{nrperm} \sum_{r=1}^{nrperm} Pairs_{ab}^{observed} > Pairs_{ab}^{random}$$

$$P_{ab}^- \quad = \quad \frac{1}{nrperm} \sum_{r=1}^{nrperm} Pairs_{ab}^{observed} < Pairs_{ab}^{random}$$

With $Pairs_{ab}^{observed}$ the number of pairs we observe in the data and $Pairs_{ab}^{random}$ the number of pairs we observe after we permute the group to patient labels. Because this test does not have an underlying model for how many connections there are between individual groups, we here use a Bonferroni correction instead of the Schaffer's method to correct for multiple testing. In

order to test the sensitivity of these estimates, we repeated this analysis using cut-off values of 0.05, 0.1, 0.15, 0.2 and 0.3 years.

Results

Introduction of new lineages into the city drive the local epidemic

We first assessed how the 663 sampled cases in Basel compare to 11,000 sequences sampled from around the world between January 2016 and December 2017, by inferring a phylogenetic tree using the hemagglutinin (HA) sequences. The Basel sequences span the existing global diversity (see Fig 1a), suggesting strong exchange (most likely importation into Basel) of viruses with other areas around the globe. We, however, did not find isolates in Basel that were part of the same clade as the vaccine strain in that season (i.e. 3c2), which is consistent with very few cases of this clade in that season (see Fig 1A).

The number of sampled sequences in Basel peaked at the end of 2016, with a smaller peak at the beginning of February of 2017 (see Fig 1b). The sequenced cases were dominated by the 3c2 sub-clades A1, A1a and A3, and we observe that a peak in A1a cases mainly contributed to the peak in February of 2017.

Basel sequences cluster into local transmission clusters within the global diversity (see Materials and Methods). We obtained around 240 local clusters (see Fig 1c and https://nextstrain.org/community/jameshadfield/basel-flu/1), suggesting that the sampled sequences were the result of around 240 influenza introductions from areas outside of Basel. In order to investigate if this number is a strict lower bound for the number of introductions, we use random subsets of the 663 sequences to re-estimate the number of introductions. We find that the number of estimated introductions grows approximately linearly with the number of sequences in a subset (see Fig 1c). This suggests that with additional sampling effort, we would have captured more introductions. With its own international airport, the airport of Zürich nearby, and a major rail hub, Basel is well connected to the rest of Europe and the world. As such, people working in Basel often do not live in the city and commute daily from elsewhere in Switzerland, Germany and France. Basel is a tourist destination and often hosts international conferences, attracting people from all over the world. This connectedness likely drives these introductions of influenza into the city.

Quantification of the overall local epidemic following the introduction into the city

After introduction into the city, we next study how influenza is transmitted locally. To get an estimate of how long lineages persist locally, we additionally estimated the tree heights of local transmission clusters that had at least 2 sampled sequences (see Fig 1e). We estimated that the average tree height of a local outbreak clusters with at least 2 sampled sequences was around 30 days. These tree heights provide an estimate of the lower bound to how long lineages spread locally on average and suggest that the average local transmission cluster persists for at least 1 month.

In order to quantify the amount of local transmission, we estimated the effective reproduction number (R_{eff}) to be between 1 and 1.5 for most of the season, which agrees with previous estimates of the effective reproduction number for seasonal influenza [36]. The R_{eff} peaks in December and in February (see Fig 1f) with the 95% credible interval excluding and R_{eff} of 1 in December. In January, we inferred a drop in the effective reproduction number with the estimated median being below 1 (see Fig 1f).
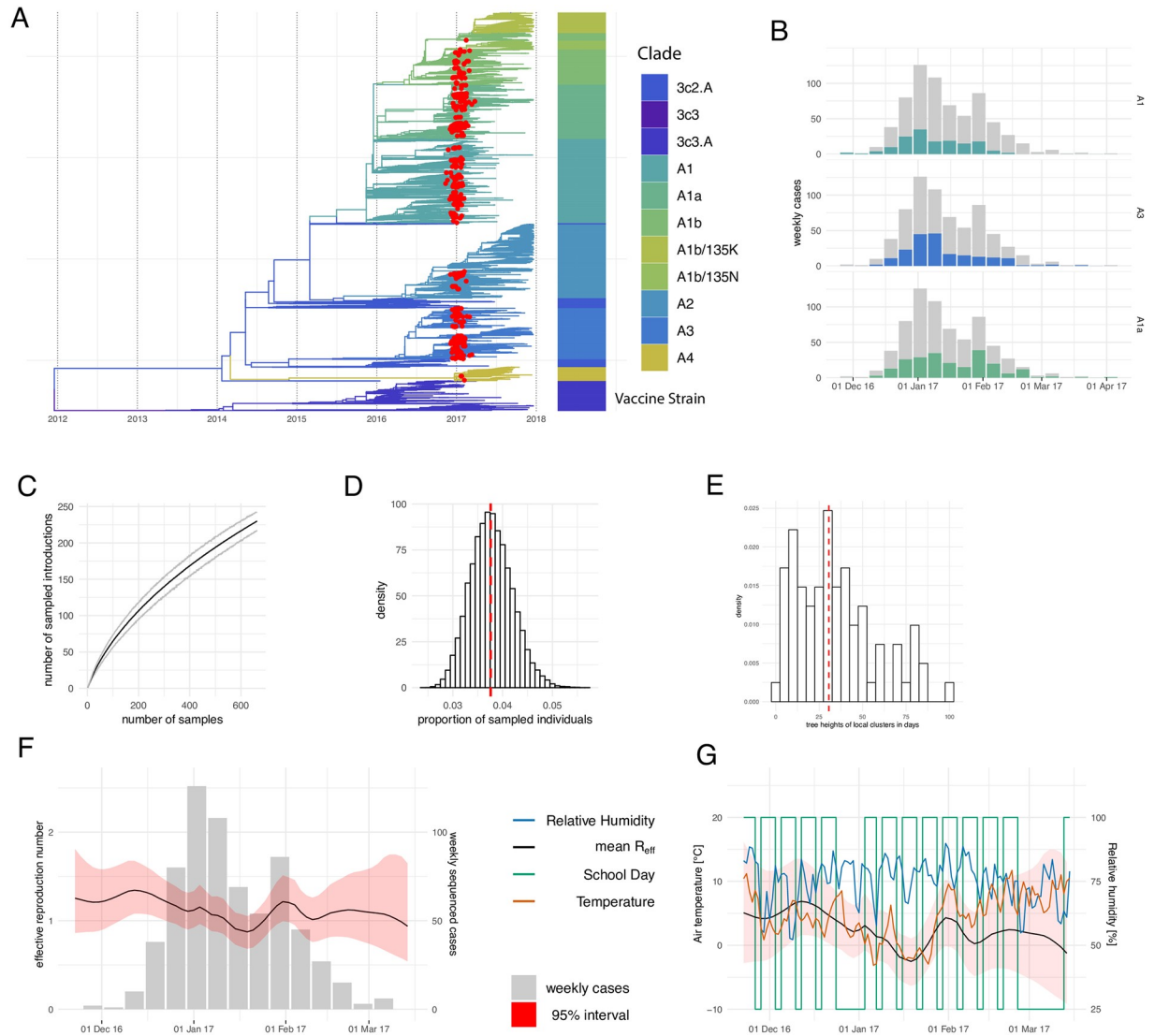
Fig 1. The local spread of influenza H3N2 in Basel in the 2016/2017 season. A Time tree of HA segment reconstructed from sequences from this study and sequences from around the world. The HA sequences sampled in this study (red) are dispersed across almost all clades present between 2016–2018. This indicates that the diversity of samples in Basel is similar to the diversity of samples around the globe. **B** Number of sequenced weekly cases for the three most abundant HA clades. The number of cases for the corresponding clade is shown in color and the overall number of cases is shown in grey. **C** Number of introductions depending on how many random sequences from Basel are used. **D** Estimated proportion of sampled individuals averaged over ten different BDSKY runs using different classifications of local sequences into local clusters. Estimates for individual classifications are shown in S1 Fig. **E** Estimates of tree heights of local clusters, which can be used as a lower bound to how long lineages persisted in the city. **F** Estimates of the effective reproduction number through time inferred from all local cluster jointly by using BDSKY. The black line is the mean estimate and the red area denotes the 95% interval. The grey bars denote the number of cases per week that were sequenced and used in the BDSKY analysis. **G** Comparison of the inferred effective reproduction number with temperature and relative humidity and if a day is a school day or not. The effective reproduction number curves are averaged over the 10 different classifications of sequences into local clusters. Comparisons for the individual random classifications are shown in S2 Fig.

The trend in the overall number of cases is similar to the trends in other places in Switzerland (http://meldesysteme.bagapps.ch/sentinella/publikationen/2017%20Saisonbericht%20Grippe%202016_2017_d.pdf), where doctoral consultations for influenza like illnesses also peaked in early January. Further, these estimates are comparable to the overall trend of influenza cases during the 2016/2017 season in Europe (https://ecdc.europa.eu/en/publications-data/summary-influenza-2016-2017-season-europe).

We next investigated potential factors determining the changes in R_{eff}. The number of influenza cases over the years show a strong seasonality, with the majority of cases occurring in the winter months in both the northern and southern hemisphere [7]. Relative humidity and temperature have been described to drive influenza transmission [37]. Additionally, the effect of school closures on the spread of pandemic influenza has been discussed [38, 39]. Thus we investigated potential correlations of R_{eff} with temperature, relative humidity and school days (i.e. days when children go to school). As we only studied one season, these correlations have to be interpreted with caution, and analyses of other seasons are needed to confirm the potential correlations. Neither humidity nor school days showed a significant correlation: relative humidity stayed fairly constant over the season, and both low and high R_{eff} are found during times when schools were open (Fig 1g and S4, S5 and S6 Figs). To account for autocorrelation, we performed the correlation analysis, averaging the temperature, relative humidity and mean number of school days over 4, 6, 8 and 10 days instead of just 2 days. We find the mean temperature to be significantly correlated with the mean R_{eff} in both scenarios (see S3 and S4 Figs).

Viral shedding of viruses has been shown previously to be increased at lower temperatures in animal models [37] and higher absolute humidity has been shown to favor transmission on a population level [40]. We here observe lower effective reproduction numbers at lower temperatures. The correlation of the effective reproduction number with the temperature however is not necessarily causal, as it, for example, could be due to social behavior being different at lower temperatures. Also, the computed p-values could be inflated due to unaccounted autocorrelation, artificially inflating the number of independent data points that are actually in the datasets.

Along with the R_{eff}, we co-estimated the sampling probability, that is the probability of an infected individual being sampled. Since we followed the same procedure for inclusion of patients throughout the epidemic season [17], we assumed that this probability is constant throughout the influenza season. We estimated the sampling probability to be between 3% and 5% (see Fig 1d and S2 Fig). In contrast to the R_{eff} estimates, this value is more sensitive to the procedure of clustering of sequences into sets of locally transmitted sequences (see S2 Fig). Additionally, the prior probability on the effective reproduction number, as well as the assumed becoming un-infectious rate can influence this estimate [41]. With 663 samples from different patients included in this analysis, this would suggest that between 13260 and 22100 people in Basel or between 8% and 13% of its population of about 171000 were infected with influenza H3N2 during the 2016-17 season. The city limits of Basel are however in reality not fixed and the metropolitan area around the city is substantially larger than the city itself. Furthermore we have sampled patients who went to a doctor or hospital in the city of Basel, but live in the surrounding areas or other parts of the world. We therefore expect the estimate of between 8% and 13% to be an estimate for the upper bound of the number of infected people, rather than the true percentage of infected individuals. These estimates of the overall attack rate of seasonal influenza in Basel are broadly consistent (though on the lower end) with estimates of the attack rate derived from un-vaccinated individuals [42], which range from approximately 20% in children to 10% in adults and assuming a vaccination rate of 12% in the 2016/2017 season in Switzerland [43].

Overall, our analyses suggest that transmission occurred with an effective reproduction number varying between 1-1.5 throughout the season, overall infecting at most 8–13% of the population.

Importance of age groups and family status in the local epidemic spread

After having determined the importance of introductions of influenza into the city and the overall rate at which influenza is spread in the city, we next studied the effect that age and family status has on the overall spread. Fig 2a and 2b and show the patient age distribution within
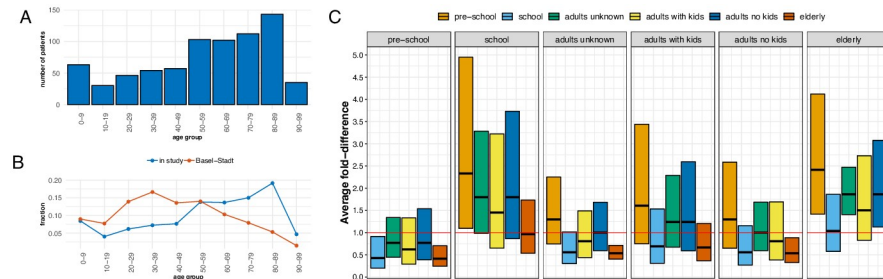
Fig 2. Age distribution and differences in the number of connections to other patients that members of the different categorical patient groups have. A Distribution of patient ages in this study in age group intervals of 10 years. **B** Proportion of different age groups in this study compared to in the city of Basel. **C** Model-based confidence intervals for the difference in average number of connections between each pair of age groups from a negative binomial model. Upper and lower bounds represent 95% confidence intervals for the average fold-difference in connection number between two groups corrected for multiple hypothesis testing using the Tukey method. This means that confidence intervals that do not include 1 are statistically significant. We see that the average connection number for elderly patients is twice the average connection number for preschoolers, adults without children, and adults with unknown status. These values are estimated for a cutoff values of 0.1 years. Estimates for different cutoff values are shown in S7 Fig. Also, the estimates of the mean number of connections in each group are shown in S8 Fig.

https://doi.org/10.1371/journal.ppat.1008984.g002

our samples. In order to study the role of age and family status in spreading influenza, we next subdivided our Basel patients into four different age groups, preschoolers (<7 years old), school-aged children (7 to 17 years), adults (18 to 65 years) and the elderly (>65 years old). We further categorized adults into three subgroups corresponding to family status: adults for whom we know that they live in the same household as children, adults for whom we know that they do not and adults for whom we do not have this information. We thus have overall six different categories of patient groups.

For each individual infected with influenza in each of these categorical groups, we determined the number of patients with influenza viruses isolated below a certain phylogenetic distance. This number, we then define as the number of connections a patient has. A connection exists if the pairwise phylogenetic distance between viruses isolated from two patients is at most 0.1 years. We then evaluate the mean number of connections of a negative binomial distribution for all individuals from each of the six groups. We later repeated the analyses using different cutoff values. This way of defining two viral isolates to be connected is therefore done independently of the above used definition of local transmission clusters. Using pairwise distances allows us to use distance in the transmission chain. Defining two individuals as connected if they are from the same cluster, on the other hand, only says if two sequences originated from the same introduction. Additionally, we confirm with a simulation that the average number of connections we observe empirically are similar to the number of connections we would expect when simulating under a simple SIR model with a 4% sampling proportion (see S9 Fig).

We find that school-aged children are on average connected to more individuals, than preschoolers (see Fig 2c). This difference is statistically significant after multiple hypothesis testing at a cutoff value of 0.1 years, but not other cutoff values. We further, but not significantly, find that adults that reported to live in the same household as children are on average connected to more patients than those that do not live in the same household as children. For the elderly, we find that they have significantly more connections compared to adults with unknown household status, adults without children, and preschoolers. They do not have more connections on average compared to adults living in the same household as children and school aged children. In summary, there is signal for school children and elderly having more connections to other

individuals compared to the three groups unknown household status, adults without children, and preschoolers. The adult living in the same household as children group show tendencies to be connected more in average than the latter three groups, though the data is not informative enough, respectively we do not have enough data, to provide strong evidence for that.

That school aged children and elderly are connected to more individuals than the other groups can have different explanations. The most obvious one is that individuals of these groups are more likely to participate in transmission events, either as a donor or recipient; alternatively, strong mixing within a group and a higher probability of visiting a doctor upon infection and therefore a higher sampling probability could act as an explanation (if members from any group are equally likely to transmit or receive influenza to and from members of any other group, higher sampling probability would increase to number of connections of every group and not just one). Indeed, sampling cases from the Hospital and private practitioners will lead to more severe cases being more likely included in the analysis. In paticular, the elderly and pre-school aged children are overrepresented in our dataset compared to their relative proportion in the population of Basel (see Fig 2A).

In order to assess the potential explanations, we investigated how strongly or weakly different age groups are connected with each other. In particular, instead of just looking at how many patients from any age groups an individual is connected to, we now assess how patients from age groups are connected between each other. We did so using two different approaches. First, we estimated the probability that an individual from a group is connected to a member from the same or a different group by using multinomial logistic regression with the inverse number of samples from each group as weights. Second, we use permutation testing to estimate the probability that the number of connections between different groups is significantly higher or lower than expected if all groups would be equally connected. To do so, we again use the definition of a connection between pairs of patients from the last section.

For both approaches, we find that mostly school-aged children are associated with other school-aged children, and the elderly are associated with other elderly people. We additionally find some indication of higher association between children of any age and adults living in the same household as children than children of any age and adults without children (see Fig 3). Adults living in the same household as children, on the other hand, are estimated to have low association to other adults without children and much higher association to children, whereas adults without children are mostly associated to other adults without children (see Fig 3a).

Increased sampling of the elderly relative to the other age groups is likely to occur, since the elderly are more likely to visit a doctor when in case of infections with influenza [44]. The elderly are indeed over-represented in this study (see Fig 2b). Thus, strong mixing within group and high sampling, might explain the increased connectivity of the elderly.

The second group that we found to have many connections to other patients where school-aged children (see Fig 2c). When looking to which groups these were connected, we found them to be associated with other school-aged children. However, they are unlikely to suffer from more symptoms than preschoolers [45] and therefore should not be overrepresented in our dataset compared to pre-schoolers. Also, based on Fig 2b, we do not see evidence for over-sampling of this group compared to preschoolers. We therefore interpret our results as school-aged children being involved in more transmission events compared to the other patient groups, including preschoolers. Furthermore, adults living in the same household as children might get mainly infected by the children and not by other adults, which does however not mean that adults do not play a crucial role in introducing novel lineages into the city. Indeed, children have been previously reported to be a strong driver of influenza transmission [14, 15, 39].
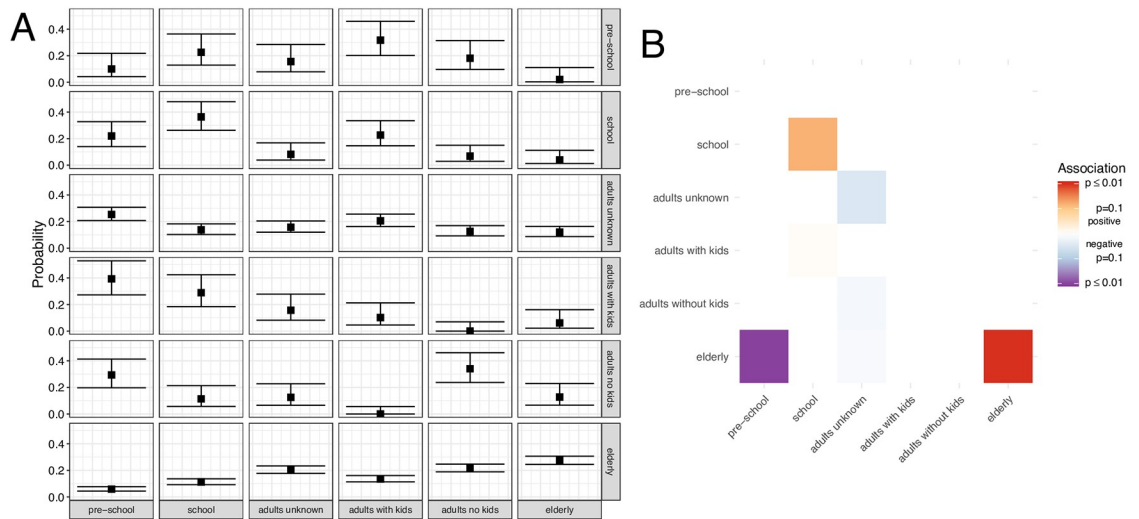
Fig 3. Mixing of different age groups. Here we show the mixing patterns between the different categorial patient groups. In contrast to Fig 2, we here ask between which groups connections exists and not just if individuals within these groups have more or less connections than individuals within other groups. We define pairs of patients to be connected if their pairwise phylogenetic distance was below 0.1 years. Results for other thresholds are shown in S10 and S11 Figs. **A** Probability that an individual from the group in each row is connected to a random individual from the group in a column. These probabilities were calculated by using the inverse number of samples from each group as weights. Upper and lower bounds correspond to 95% confidence intervals around the estimated probability. **B** The color of each tile in the heatmap corresponds to the p-value for either positive (red) or negative (blue) associations. These p-values are bonferroni corrected for the number of comparisons (42). We estimate these p-values by randomly permuting the group to patient labels and then comparing the number of pairs of interactions we observe in the data vs. when randomly permuting.

These interpretations however are based on the analysis of influenza isolates from one season and city and will therefore need to be repeated in different seasons and cities to get a more complete understanding of the transmission patterns of influenza across age groups.

Discussion

In absence of deep knowledge of the important drivers of the local spread of SARS-CoV-2, governments around the world resorted to closing down societies to reduce the burden of COVID-19. Better understanding of how a disease is spread can help optimizing non-pharmaceutical interventions in order to reduce the burden on societies, while still effectively reducing transmission.

One of the diseases that are major reoccurring burdens on societies are seasonal influenza viruses. Seasonal influenza annually infects a large portion of the global population and while its global spread has been studied extensively, its local spread remains largely unstudied. Our results are based on one of the most densely sampled genetic datasets of influenza sequences to date. Additionally, we connected the genetic information to patient information such as age for all and more personal information for a subset of the patients, providing unparalleled resolution to study how influenza spreads locally. The 2016/17 season for which we collected data was dominated by influenza A/H3N2. Based on this data, we observe that hundreds of introductions initiate the seasonal influenza epidemic in the studied city of Basel, that the overall spread varies throughout the season, and that school aged children seem to play a more important role in local outbreaks than preschoolers, while elderly have their own transmission chains.

It will be interesting to see how these results transfer to other cities and seasons with potentially other social structure or other geographical location. In particular, the subtypes that circulate and their ability to escape host immunity and seasons dominated by different influenza types may influence mixing patterns. For the future, it will be particularly interesting to see if seasons where other subtypes such as influenza A/H1N1 or influenza B dominate show the same or differing patterns that we observed. While such studies on a population level requires great effort in recruiting patients as well as in sequencing viruses, they can greatly improve our understanding of how influenza spreads locally. This will hopefully allow us to streamline public health interventions in the most efficient way possible, and thus, help to reduce the great burden on societies caused by the seasonal flu.

Supporting information

S1 Text. Quality criteria for sequences used in this study.
(PDF)

S1 Fig. Estimates of the sampling proportion for different classifications of sequences into local clusters. Each histogram shows an inferred sampling proportion based on a classification of sequences into local clusters. Since these classifications are dependent on which iteration of the MCMC was used for the classification into local clusters, we repeated the analysis using 10 different random iterations. Each subplot shows the estimated sampling proportion when using one of these classifications. The dotted red line shows the median estimate of the sampling proportion over all iterations.
(TIF)

S2 Fig. Estimates of the effective reproduction number over time using different classifications of sequences into local clusters. Each subplot shows the inferred effective reproduction number when using a different iteration of the MCMC for the assignment of Basel sequences into local clusters.
(TIF)

S3 Fig. Estimated p-values for correlation between the effective reproduction number and temperature, relative humidity and school days. Here we show the estimated p-values for the correlation between the effective reproduction number and temperature, relative humidity and school days estimated when the data was averaged over different number of days (x-axis). The estimate p-values are shown for two different time intervals (in different colors). For the orange line, estimates for 1 November 2016 until 1 March 2017 were used and for the blue line, estimated from December until February were used. These plots were generated using the effective reproduction number averaged over 10 different classifications of sequences into local clusters. The equivalent plots for generated using the effective reproduction number estimates of each individual subset is shown in S5 Fig.
(TIF)

S4 Fig. Estimated correlation coefficients between the effective reproduction number and temperature, relative humidity and school days. Here we show the estimated correlation coefficients for the correlation between the effective reproduction number and temperature, relative humidity and school days estimated when the data was averaged over different number of days (x-axis). The estimate p-values are shown for two different time intervals (in different colors). For the orange line, estimates for 1 November 2016 until 1 March 2017 were used and for the blue line, estimated from December until February were used. These plots were generated using the effective reproduction number averaged over 10 different classifications of

sequences into local clusters. The equivalent plots for generated using the effective reproduction number estimates of each individual subset is shown in S6 Fig.
(TIF)

S5 Fig. Estimated p-values for correlation between the effective reproduction number and temperature, relative humidity and school days. Here we show the estimated p-values for the correlation between the effective reproduction number and temperature, relative humidity and school days estimated when the data was averaged over different number of days (x-axis). The estimate p-values are shown for two different time intervals (in different colors). For the orange line, estimates for 1 November 2016 until 1 March 2017 were used and for the blue line, estimated from December until February were used. The different lines of the same color show the p-values estimates using the effective reproduction number estimates of individual classifications of sequences into local clusters.
(TIF)

S6 Fig. Estimated correlation coefficients between the effective reproduction number and temperature, relative humidity and school days. Here we show the estimated correlation coefficients for the correlation between the effective reproduction number and temperature, relative humidity and school days estimated when the data was averaged over different number of days (x-axis). The estimate p-values are shown for two different time intervals (in different colors). For the orange line, estimates for 1 November 2016 until 1 March 2017 were used and for the blue line, estimated from December until February were used. The different lines of the same color show the p-values estimates using the effective reproduction number estimates of individual classifications of sequences into local clusters.
(TIF)

S7 Fig. Fold difference between the average number of connection of individuals from the different groups. Plots are analogue to Fig 2c, but for different thresholds: 0.05 years in plot **A**, 0.1 years in plot **B** (analogue to Fig 2c), 0.15 years in plot **C**, 0.2 years in plot **D** and 0.3 years in plot **E**.
(TIF)

S8 Fig. Average number of patients and individual from a group is connected to. Each subplot shows the average number of connected individuals a patient from the group shown by the color is connected to. Upper and lower bounds correspond to 95% confidence intervals around the average. We consider two patients to be connected if the pairwise phylogenetic distance between the influenza viruses sequenced from them is below a certain threshold. These thresholds are 0.05 years in plot **A**, 0.1 years in plot **B**, 0.15 years in plot **C**, 0.2 years in plot **D** and 0.3 years in plot **E**.
(TIF)

S9 Fig. Average number of patients and individual from a group is connected to for different thresholds. Here, we compare the average number of connections (y-axis) of an individual patient to other patients for different threshold (x-axis) between what we observe empirically and what we observe in simulations. We ran the simulations using either a reproduction number of 1 or 1.25, a becoming uninfectious rate of 0.25 per day and a sampling proportion of 4%.
(TIF)

S10 Fig. Probability that an individual from the group in each row is connected to an individual from the group in a column. Plots are analogue to Fig 3a, but for different thresholds:

0.05 years in plot **A**, 0.1 years in plot **B**, 0.15 years in plot **C**, 0.2 years in plot **D** and 0.3 years in plot **E**.
(TIF)

S11 Fig. Mixing of different age groups when using different cutoff values. Plots are analogue to Fig 3b, but for different thresholds. The thresholds in years are given on top of each subfigure. Unless very large thresholds are used, the elderly are estimated to be positively associated with other members from the same group. School aged children are associated with other school aged children except for very low thresholds where the number of pairs is very low. Elderly being negatively associated with pre-school aged children shows up only at lower thresholds.
(TIF)

Author Contributions

Conceptualization: Manuel Battegay, Rita Schneider-Sliwa, Adrian Egli, Tanja Stadler.

Data curation: Daniel Wüthrich, Nina Goldman, Nadine Sailer, Claudia Saalfrank, Myrta Brunner, Noémi Augustin, Helena MB Seth-Smith, Yvonne Hollenstein, Mohammedyaseen Syedbasha, Daniela Lang, Olivier Dubuis, Michael Naegele, Andreas Buser, Christian H. Nickel, Nicole Ritz, Andreas Zeller, Manuel Battegay, Adrian Egli.

Formal analysis: Nicola F. Müller, Brian M. Lang, James Hadfield.

Funding acquisition: Manuel Battegay, Rita Schneider-Sliwa, Adrian Egli, Tanja Stadler.

Investigation: Nicola F. Müller, Adrian Egli, Tanja Stadler.

Methodology: Nicola F. Müller, Richard A. Neher, Brian M. Lang, Trevor Bedford, Manuel Battegay, Rita Schneider-Sliwa, Adrian Egli, Tanja Stadler.

Project administration: Adrian Egli, Tanja Stadler.

Resources: Rita Schneider-Sliwa, Adrian Egli, Tanja Stadler.

Software: Nicola F. Müller, Brian M. Lang, James Hadfield.

References

1. Fitzner KA, Shortridge K, McGhee S, Hedley A. Cost-effectiveness study on influenza prevention in Hong Kong. Health Policy. 2001; 56(3):215–234. https://doi.org/10.1016/S0168-8510(00)00140-8

2. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. science. 2004; 303(5656):327–332. https://doi.org/10.1126/science.1090727 PMID: 14726583

3. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. Science. 2001; 292(5525):2323–2325. https://doi.org/10.1126/science.1058321

4. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular biology and evolution. 2005; 22(5):1185–1192. https://doi.org/10.1093/molbev/msi103

5. Kühnert D, Stadler T, Vaughan TG, Drummond AJ. Phylodynamics with migration: a computational framework to quantify population structure from genomic data. Molecular biology and evolution. 2016; 33(8):2102–2116. https://doi.org/10.1093/molbev/msw064

6. Müller NF, Rasmussen DA, Stadler T. The structured coalescent and its approximations. Molecular biology and evolution. 2017; 34(11):2970–2981. https://doi.org/10.1093/molbev/msx186

7. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, et al. The global circulation of seasonal influenza A (H3N2) viruses. Science. 2008; 320(5874):340–346. https://doi.org/10.1126/science.1154137 PMID: 18420927

8. Bedford T, Cobey S, Beerli P, Pascual M. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). PLoS pathogens. 2010; 6(5):e1000918. https://doi.org/10.1371/journal.ppat.1000918

9. Bahl J, Nelson MI, Chan KH, Chen R, Vijaykrishna D, Halpin RA, et al. Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. Proceedings of the National Academy of Sciences. 2011; 108(48):19359–19364. https://doi.org/10.1073/pnas.1109314108 PMID: 22084096

10. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLoS pathogens. 2014; 10(2):e1003932. https://doi.org/10.1371/journal.ppat.1003932 PMID: 24586153

11. Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, et al. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature. 2015; 523(7559):217. https://doi.org/10.1038/nature14460 PMID: 26053121

12. Holmes EC, Ghedin E, Halpin RA, Stockwell TB, Zhang XQ, Fleming R, et al. Extensive geographical mixing of 2009 human H1N1 influenza A virus in a single university community. Journal of virology. 2011; 85(14):6923–6929. https://doi.org/10.1128/JVI.00438-11 PMID: 21593168

13. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes constrain the within and between host evolution of influenza virus. Elife. 2018; 7:e35962. https://doi.org/10.7554/eLife.35962

14. Reed C, Katz JM, Hancock K, Balish A, Fry AM. Prevalence of seropositivity to pandemic influenza A/H1N1 virus in the United States following the 2009 pandemic. PLoS one. 2012; 7(10):e48187. https://doi.org/10.1371/journal.pone.0048187

15. Turbelin C, Souty C, Pelat C, Hanslik T, Sarazin M, Blanchon T, et al. Age distribution of influenza like illness cases during post-pandemic A (H3N2): comparison with the twelve previous seasons, in France. PLoS One. 2013; 8(6):e65919. https://doi.org/10.1371/journal.pone.0065919 PMID: 23755294

16. Kwok KO, Riley S, Perera RA, Wei VW, Wu P, Wei L, et al. Relative incidence and individual-level severity of seasonal influenza A H3N2 compared with 2009 pandemic H1N1. BMC infectious diseases. 2017; 17(1):337. https://doi.org/10.1186/s12879-017-2432-7 PMID: 28494805

17. Egli A, Saalfrank C, Goldman N, Brunner M, Hollenstein Y, Vogel T, et al. Identification of influenza urban transmission patterns by geographical, epidemiological and whole genome sequencing data: protocol for an observational study. BMJ open. 2019; 9(8):e030913. https://doi.org/10.1136/bmjopen-2019-030913 PMID: 31434783

18. Egli A, Goldman N, Müller NF, Brunner M, Wuethrich D, Tschudin-Sutter S, et al. High-resolution influenza mapping of a city reveals socioeconomic determinants of transmission within and between urban quarters. bioRxiv. 2020;.

19. Wüthrich D, Lang D, Müller NF, Neher RA, Stadler T, Egli A. Evaluation of two workflows for whole genome sequencing-based typing of influenza A viruses. Journal of virological methods. 2019; 266:30–33. https://doi.org/10.1016/j.jviromet.2019.01.009

20. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9 (4):357. https://doi.org/10.1038/nmeth.1923

22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

23. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic acids research. 2012; 40(22):11189–11201. https://doi.org/10.1093/nar/gks918 PMID: 23066108

24. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–1313. https://doi.org/10.1093/bioinformatics/btu033

25. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus evolution. 2018; 4(1):vex042.

26. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018; 34(23):4121–4123. https://doi.org/10.1093/bioinformatics/bty407 PMID: 29790939

27. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017; 8(1):28–36. https://doi.org/10.1111/2041-210X.12628

28. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Molecular biology and evolution. 2005; 23(1):7–9.

29. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS computational biology. 2014; 10(4):e1003537. https://doi.org/10.1371/journal.pcbi.1003537 PMID: 24722319

30. Holmes EC, Dudas G, Rambaut A, Andersen KG. The evolution of Ebola virus: Insights from the 2013–2016 epidemic. Nature. 2016; 538(7624):193. https://doi.org/10.1038/nature19790

31. Vaughan TG, Kühnert D, Popinga A, Welch D, Drummond AJ. Efficient Bayesian inference under the structured coalescent. Bioinformatics. 2014; 30(16):2272–2279. https://doi.org/10.1093/bioinformatics/btu201

32. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Sciences. 2013; 110(1):228–233. https://doi.org/10.1073/pnas.1207965110

33. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Molecular biology and evolution. 2012; 30(3):713–724.

34. Montgomery DC. Design and analysis of experiments. John wiley & sons; 2017.

35. Donoghue JR, et al. Implementing Shaffer's multiple comparison procedure for a large number of groups. In: Recent Developments in Multiple Comparison Procedures. Institute of Mathematical Statistics; 2004. p. 1–23.

36. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC infectious diseases. 2014; 14(1):480. https://doi.org/10.1186/1471-2334-14-480

37. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. PLoS pathogens. 2007; 3(10):e151. https://doi.org/10.1371/journal.ppat.0030151

38. Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006; 442(7101):448. https://doi.org/10.1038/nature04795

39. Cauchemez S, Valleron AJ, Boelle PY, Flahault A, Ferguson NM. Estimating the impact of school closure on influenza transmission from Sentinel data. Nature. 2008; 452(7188):750. https://doi.org/10.1038/nature06732

40. Deyle ER, Maher MC, Hernandez RD, Basu S, Sugihara G. Global environmental drivers of influenza. Proceedings of the National Academy of Sciences. 2016; 113(46):13081–13086. https://doi.org/10.1073/pnas.1607747113

41. Stadler T, Kouyos R, von Wyl V, Yerly S, Böni J, Bürgisser P, et al. Estimating the basic reproductive number from viral sequence data. Molecular biology and evolution. 2011; 29(1):347–357. https://doi.org/10.1093/molbev/msr217

42. Somes MP, Turner RM, Dwyer LJ, Newall AT. Estimating the annual attack rate of seasonal influenza among unvaccinated individuals: a systematic review and meta-analysis. Vaccine. 2018; 36(23):3199–3207. https://doi.org/10.1016/j.vaccine.2018.04.063

43. Brunner I, Schmedders K, Wolfensberger A, Schreiber PW, Kuster SP. The Economic and Public Health Impact of Influenza Vaccinations in Swiss Pharmacies in the 2016/17 and 2017/18 Influenza Seasons. Swiss medical weekly. 2019;(149):w20161.

44. Thompson WW, Shay DK, Weintraub E, Brammer L, Bridges CB, Cox NJ, et al. Influenza-associated hospitalizations in the United States. Jama. 2004; 292(11):1333–1340. https://doi.org/10.1001/jama.292.11.1333 PMID: 15367555

45. Troeger CE, Blacker BF, Khalil IA, Zimsen SR, Albertson SB, Abate D, et al. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017. The Lancet Respiratory Medicine. 2019; 7(1):69–89. https://doi.org/10.1016/S2213-2600(18)30496-X