PNAS Nexus, 2025, 4, pgae561
https://doi.org/10.1093/pnasnexus/pgae561
Advance access publication 17 December 2024
Research Report

Integrating dynamical modeling and phylogeographic inference to characterize global influenza circulation

Francesco Parino[a,1], Emanuele Gustani-Buss[b,1], Trevor Bedford[c,d], Marc A. Suchard[e,f], Nídia S. Trovão[g], Andrew Rambaut[h], Vittoria Colizza[a,i,2], Chiara Poletto[j,*2] and Philippe Lemey[b,*,2]

[a]Sorbonne Université, INSERM, Institut Pierre Louis d'Epidemiologie et de Santé Publique (IPLESP), Paris, France
[b]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven – University of Leuven, Leuven 3000, Belgium
[c]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA
[d]Howard Hughes Medical Institute, Seattle, WA 98109, USA
[e]Departments of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA
[f]Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA 90095, USA
[g]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
[h]Institute of Ecology and Evolution, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom
[i]Department of Biology, Georgetown University, Washington, DC, USA
[j]Department of Molecular Medicine, University of Padova, Padova 35121, Italy
*To whom correspondence should be addressed: Email: chiara.poletto@unipd.it (C.P.); Email: philippe.lemey@kuleuven.be (P.L.)
[1]F.P. and E.G.-B. contributed equally to this work.
[2]V.C., C.P., and P.L. contributed equally to this work.
Edited By Sandro Galea

Abstract

Global seasonal influenza circulation involves a complex interplay between local (seasonality, demography, host immunity) and global factors (international mobility) shaping recurrent epidemic patterns. No studies so far have reconciled the two spatial levels, evaluating the coupling between national epidemics, considering heterogeneous coverage of epidemiological, and virological data, integrating different data sources. We propose a novel-combined approach based on a dynamical model of global influenza spread (GLEAM), integrating high-resolution demographic, and mobility data, and a generalized linear model of phylogeographic diffusion that accounts for time-varying migration rates. Seasonal migration fluxes across countries simulated with GLEAM are tested as phylogeographic predictors to provide model validation and calibration based on genetic data. Seasonal fluxes obtained with a specific transmissibility peak time and recurrent travel outperformed the raw air-transportation predictor, previously considered as optimal indicator of global influenza migration. Influenza A subtypes supported autumn–winter reproductive number as high as 2.25 and an average immunity duration of 2 years. Similar dynamics were preferred by influenza B lineages, with a lower autumn–winter reproductive number. Comparing simulated epidemic profiles against FluNet data offered comparatively limited resolution power. The multiscale approach enables model selection yielding a novel computational framework for describing global influenza dynamics at different scales—local transmission and national epidemics vs. international coupling through mobility and imported cases. Our findings have important implications to improve preparedness against seasonal influenza epidemics. The approach can be generalized to other epidemic contexts, such as emerging disease outbreaks to improve the flexibility and predictive power of modeling.

Keywords: influenza, metapopulation, phylogeography, Bayesian inference

Significance Statement

Despite extensive surveillance efforts to track seasonal influenza epidemics, it still remains unclear how factors acting at different scales (e.g. local epidemics at the community level vs. international coupling through travel) drive the global circulation of the disease. This is partly due to the limitations of current approaches leveraging either genetic or epidemiological data, which, in isolation, are unable to successfully reconcile multiple scales. Here, we introduce an integrative approach based on a multiscale epidemiological model calibrated on worldwide genetic data through phylogeographic inference. Building up from the city level, our approach simulates migration fluxes between epidemics occurring in different countries and identifies model parameterizations that offer better predictions for global influenza circulation than previously attainable.

Competing Interest: The authors declare no competing interests.
Received: March 20, 2024. **Accepted:** November 21, 2024

Introduction

Seasonal influenza viruses cause recurrent epidemics characterized by annual periodicity in temperate countries and by diverse, less regular patterns in the tropics (1). Extensive epidemiological research highlighted the critical role of local sociodemographic and environmental aspects (e.g. weather conditions, school calendar, and increased indoor activity) in the onset and unfolding of influenza waves (2–10). At the same time, international human travel ensures rapid worldwide circulation of influenza (11–13). Phylogeographic studies have reconstructed the global migration patterns of seasonal influenza in extensive detail, revealing limited local persistence of the virus in most regions and highlighting the importance of continual reseeding in determining viral genetic structure, severity, and timing of the epidemics (12, 14–16). The concurrent impact of local and global drivers has also been apparent through analyses of the decline of influenza incidence observed during the coronavirus disease 2019 (COVID-19) pandemic. Both reduction in international travel and social restrictions (due, e.g. to remote working and school closure) were, indeed, found to be associated with the influenza drop (17–19). From a modeling perspective, however, reconciling the two spatial levels represents a major challenge, as the interplay between the local progression of an epidemic in a particular country and the coupling between different epidemics mediated by human mobility remains poorly understood.

The unevenly distributed human host population across countries and seasonal areas and the complex network of human travel acting over both short- and long-range distances (i.e. from city-to-city commuting to international air travel) are key to this challenge (20, 21). In addition, while influenza surveillance is improving worldwide (18), the coverage is still biased, restricting our ability to resolve the spatial dynamics. The quality of epidemiological data is generally higher in temperate areas, but these are the areas that are characterized by a high degree of synchronization of national influenza epidemics in the same hemisphere, thus making spatial effects less identifiable. As a consequence, mathematical models for influenza dynamics at different scales remain difficult to parameterize. Genetic data, however, carry the signature of large-scale circulation dynamics and may, therefore, represent a valuable complementary source to characterize seasonal influenza epidemics, especially when combined with epidemiological data. The recent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has illustrated the importance of combining epidemiological and mobility data, genomic sequences, and metadata to provide insights into viral emergence and spread (22–27). Phylogeographic inference in particular has been widely applied to elucidate SARS-CoV-2 genomic epidemiology, including origins, introductions, routes of dispersal, and drivers associated with variant dissemination, contributing to the effectiveness of systematic genomic surveillance (24–27). However, the development of integrated tools is still relatively limited and opportunities remain to fulfill the full potential of phylodynamic approaches. Different from previous phylogeographic reconstructions, we here propose a novel approach that combines a high-resolution dynamical model for the diffusion of influenza worldwide, informed by extensive demographic, and mobility data, and a generalized linear model of phylogeographic diffusion that allows for inhomogeneous migration rates over time. We use the epidemic model to simulate migration fluxes across countries and evaluate their ability to explain phylogeographic patterns.

Results and discussion

Combining mechanistic epidemic modeling and phylogeographic inference

We combine dynamical modeling and phylogeographic inference by considering the simulated fluxes of infectious cases generated by a data-driven computational model for infectious disease spread at the global scale as predictors for phylogeographic migration rates. For this purpose, we build on GLEAM, the GLobal Epidemic, and Mobility model (28), that integrates high-resolution demographic and mobility data at different spatial scales—air traffic database comprising nearly all commercial air-travels and short-range mobility obtained from national commuting data (28). The global population is distributed among 3362 patches corresponding to large urban areas and traveling of individuals is modeled explicitly based on passenger data. GLEAM has been used to model the short-term outbreak dynamics in the case of the H1N1 influenza pandemic (11, 13), Ebola (29, 30), MERS (31, 32), Zika (33), and COVID-19 (34–37), following prior modeling work considering the global scale (38–40).

To adapt GLEAM to seasonal circulation of influenza, we introduce a more realistic scheme for modeling mobility of individuals that preserves their geographic residence (see Material and Methods for more details). This approach is usually adopted for modeling recurrent travel, such as commuting (28, 41, 42), while a simpler Markovian mobility, model assuming memory-less traveling trajectories, is generally preferred for air travel. Although less realistic, the latter is more parsimonious and has only a limited approximation bias for fast spreading diseases and short-term epidemic dynamics (42). In our analysis of the multiannual influenza propagation, we compare both the Markovian and the recurrent travel approach.

In the metapopulation scheme, GLEAM transmission dynamics occur within patches ruled by a compartmental model specific to seasonal influenza (2, 3, 11, 43, 44) (Material and Methods and Section 1 of the Supplementary Material) and accounting for: (i) a temporary immunity to the virus of average duration D, with values explored between 1 and 8 years; (ii) a geographically dependent seasonal transmission in temperate areas (11, 44, 45) varying sinusoidally in time between a minimum and a maximum basic reproductive number, R_{min} (explored values: 0.5 and 0.75) and $R_{max} \in [1.25,2.5]$, respectively, with November 15, December 15, and January 15 tested as dates of maximum transmission in the northern hemisphere and minimum transmission in the southern hemisphere; (iii) a constant transmission with a basic reproductive number equal to R_{max} in the tropics. Discrete stochastic simulations at the individual level provide numerical trajectories for the global seasonal dynamics with a time resolution of 1 day. Results display autumn–winter waves interspersed by subcritical spring–summer transmission in the temperate hemispheres, and an irregular continuous circulation in the tropics. The output is summarized as fluxes during the April–September and October–March epochs between the countries considered by the phylogeographic approach.

To evaluate different parameterizations of our dynamical model, we test the resulting fluxes in a phylogeographic approach. To this purpose, we adopt a generalized linear model (GLM) extension of discrete phylogeographic diffusion (12) that accommodates time-inhomogeneous migration dynamics. The GLM-diffusion approach allows modeling the intensity of location exchange between discrete states along a phylogeny as a function of a number of potential predictors (12). Using epoch modeling (46), we allow for different location exchange processes, and hence different predictors, across different time intervals in the

evolutionary history. Specifically, we consider the difference in model-based fluxes during alternating April–September and October–March epochs. We employ phylogenetic reconstructions based on sequence data sets that were previously analyzed to reconstruct more than a decade of global seasonal migration dynamics of influenza A H3N2, H1N1 (prior to the H1N1/09 pandemic), B Victoria (VIC), and Yamagata (YAM) between 2000 and 2012 (16). We model the phylogeographic process between the countries of sampling for these different influenza subtypes and lineages. We use BEAST (47) to estimate parameters of the time-inhomogeneous GLM-diffusion approach through Bayesian inference while averaging over a set of time-measured trees.

Model selection and parameter estimation

We evaluated different sets of migration fluxes, including simple passenger fluxes based on air travel data, simulated migration fluxes from the Markovian and the recurrent travel version of GLEAM, either as homogeneously aggregated fluxes over time (annual fluxes) or as two-epoch level fluxes (seasonal fluxes). Small-scale simulations according to a number of different fluxes indicate that the GLM diffusion model performs well in yielding support for true flux predictors (Fig. S2).

To avoid including an excessive amount of migration flux predictors in a single GLM diffusion model analysis, our systematic evaluation relied on a stepwise approach. We first tested Markovian against recurrent travel fluxes for different epidemiological parameters (R_{min}, R_{max} and D) while conditioning on seasonal fluxes with a peak time in January in the northern hemisphere. Next, we tested seasonal against annual fluxes conditioning on recurrent travel fluxes and a peak time in January. Finally, we tested the January peak time against peak times in December and November conditioning on seasonally aggregated recurrent travel fluxes. In each of these analyses, we include simple air passenger fluxes as a baseline predictor, and we perform the analysis with and without a predictor based on the residuals of a regression of sample sizes against population sizes to assess the potential impact of sample sizes. Figure 1 summarizes the marginal posterior inclusion probabilities for the different comparisons, while inclusion probabilities for individual flux predictors are provided in Tables S3–S6.

Our analyses support GLEAM fluxes based on recurrent travel for H3N2, H1N1, and YAM, demonstrating (i) an improvement of dynamical model predictions over simple air travel for most of the influenza variants and (ii) the importance of accounting for memory in the origin of travel trajectories of individuals. For these three variants, seasonally aggregated fluxes also outperform annual fluxes and a peak time in January outperforms earlier peak times (with an inclusion probability of ∼1). For H3N2 specifically, seasonal fluxes strongly outperforms annual fluxes with and without residual predictor (inclusion probability of ∼1) (Table S4) and a similar support was detected for H1N1 (inclusion probability of ∼0.95). The inclusion probability for YAM is marginally lower when employing the residual predictor (0.72). When comparing the magnitude of best-supported predictors, the support for seasonal fluxes is approximately 20 times stronger than annual flux. In the case of VIC, GLEAM fluxes are only supported in the analysis without sample size residual. Standard phylogeographic analyses of these data sets have previously shown that VIC is associated with the highest degree of persistence and the lowest overall migration (16), so offering less information to support GLEAM-based fluxes, in particular when sample heterogeneity can explain a considerable degree of migration variability.

For the seasonal GLEAM fluxes based on recurrent travel and based on a January 15 peak time, we next summarize the support for the different values of R_{max}, R_{min}, and D (Fig. 2). For H3N2, the parameter combination including $R_{max} = 2.25$, $R_{min} = 0.75$, and $D =$ 2 years yields the highest flux inclusion probability (inclusion probability ∼1 and 0.68 with and without sample size residual, respectively). The parameter values in this combination are also clearly the ones that are preferred across all combinations in analyses with and without residual predictor (Fig. 2). In Section 4 of the Supplementary Material, we investigate differences in fluxes generated by different parameterizations and note that the GLM selects fluxes that are distinct from those generated by other parameterizations. The H1N1 analysis with sample size residual also finds this parameter combination to be the best supported, but not as strongly so (0.59 inclusion probability), while the analysis without sample size residual finds marginally lower support (0.33 inclusion probability) compared with a parameter combination including $R_{max} = 1.50$, $R_{min} = 0.75$, and $D = 1$ years (0.52 inclusion probability).

The analysis of YAM without residual predictor recovers the same general preference for D and R_{min} (2 years and 0.75, respectively), but finds the strongest support for a somewhat lower R_{max} of 2.00 (0.77 inclusion probability). The same parameterization is supported by the analysis with residual predictor, but with somewhat lower inclusion probability (0.56). Without residual predictor, the analysis of VIC supports the same parameterization as YAM (0.60 inclusion probability, Fig. 2).

The rapid immune waning we identify, with average duration of immune protection between 1 and 3 years, is in agreement with previous works (48–51), but some studies have estimated longer immunity duration (2, 3). Smaller transmissibility for influenza B compared with A/H3N2 (as estimated for both Yamagata and Victoria) is also consistent with previous works (16, 49). Those studies typically use incidence curves at the country level, whereas our selection is based on influenza migration patterns encoded in the genetic data. The spatial coupling, thus, carries the signature of country seasonal waves. For the period between two consecutive epidemics, early analyses of genetic data in temperate areas suggested that inter-seasonal circulation of influenza was dominated by the importation of cases from other seasonal areas, with negligible local transmission following importation (52–54). Recent improvements in out-of-season surveillance are providing increasing evidence for the sporadic generation of cases during the spring–summer period (55–57), in agreement with spring–summer transmissibility of $R_{min} = 0.75$ supported by our analyses.

Predicted influenza dynamics

We compare country-level epidemics simulated by GLEAM with recurrent travel and epidemic profiles reconstructed from FluNet data to evaluate model predictions against available surveillance data. The number of influenza-positive samples stored for each subtype and lineage in FluNet has been extensively used for reconstructing the timing and shape of the epidemic peaks (1, 8, 17, 48, 58). We compute monthly distributions of cases averaged over the period 2004/2005–2014/2015 as an indicator of the typical influenza behavior of the country (Section 5 of the Supplementary Material). Figure 3 compares the H3N2 epidemic profile with the simulated one for a set of countries in each seasonal area. The timing and shape of the epidemic waves are well reproduced by the model in the majority of temperate countries (average correlation 0.83 ± 0.02 for the northern hemisphere and

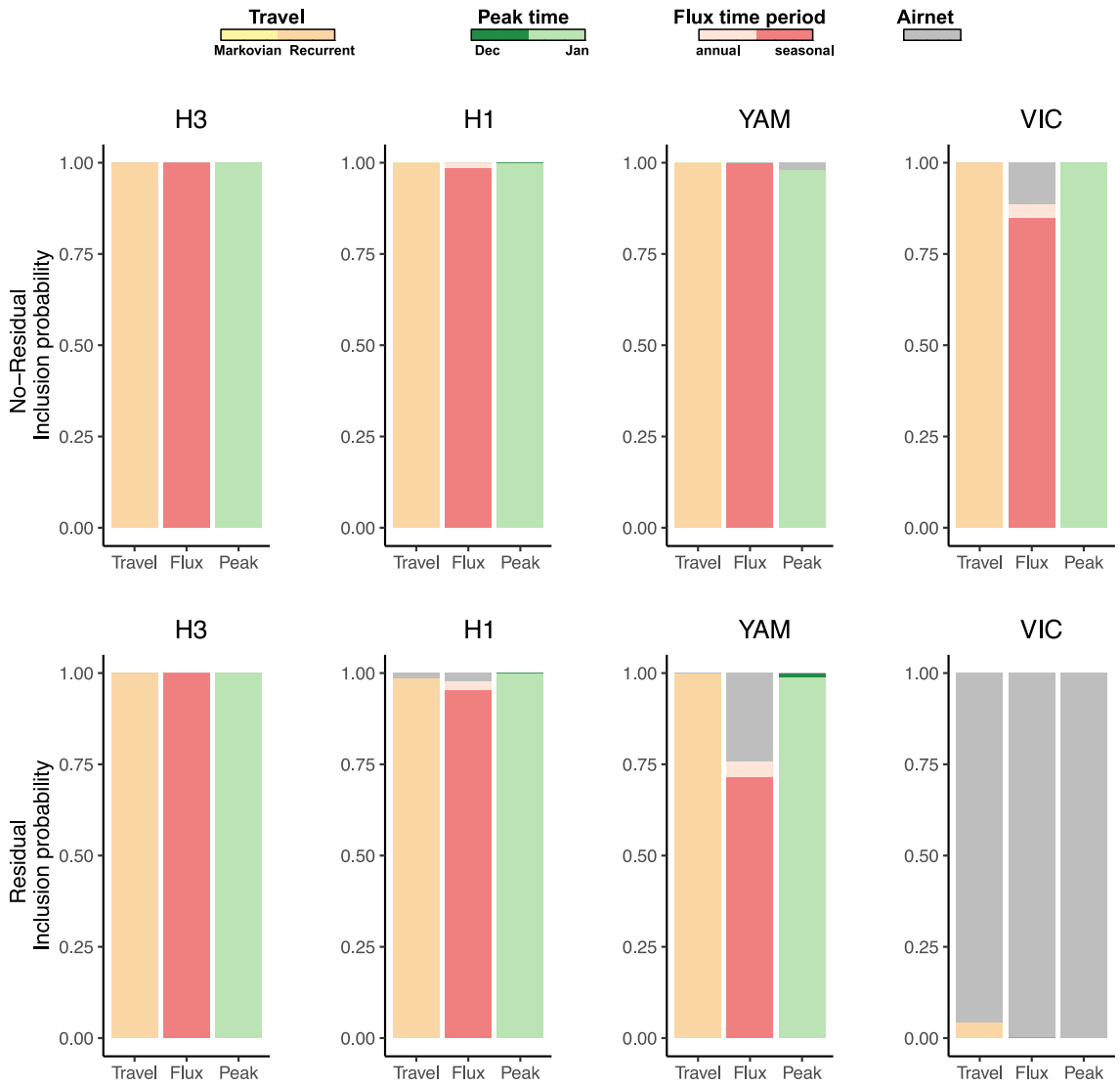

Fig. 1. Marginal posterior inclusion probabilities associated with air travel data or with GLEAM-based fluxes comparing recurrent against Markovian travel, seasonal against annual fluxes and different peak times. For peak times, we performed an analysis comparing November 15 against December 15 and an analysis comparing December 15 to January 15, but we only show the latter for simplicity as December 15 outperformed November 15.

0.66 ± 0.08 for the southern hemisphere), representing also the region with the highest availability of country records. In the tropics, correlations are generally lower (average correlation 0.10 ± 0.06) because of the rather noisy and flat average epidemic profiles (Table S8). Cross-correlation analysis confirms this picture as it reveals that optimal lag for optimizing FluNet correlation in each country falls within −1 to 1 month in the great majority of cases in the temperate areas—further supporting the consistency between simulations and FluNet profiles in the region—while less so for tropical countries (Section 5 of the Supplementary Material). For large countries, strong spatial fragmentation of the population and climatic heterogeneity complicate analyses at the national scale (59), as illustrated for China.

These results were obtained with the best parameterization for H3N2 as selected by the phylogeographic GLM. Other parameter sets performed worse, but a number of different simulated scenarios also showed a similarly high correlation with the data (see Fig. S5 and corresponding paragraph in the Supplementary Material for other influenza subtypes and lineages). This indicates that the degree of information carried by incidence data is limited, likely due to the high level of synchronization between country waves, together with nonuniform surveillance coverage and quality.

Simulated influenza circulation shows a strong coupling between Europe and North America during the October–March period, together with the central role of Southeast Asia and China as influenza sources for the Asian continent and Oceania during April–September and October–March periods. In particular, Southeast Asia is one of the main sources of importation for Australia, Japan, and Korea, India and Europe. South America appears to be disconnected from Asian regions (see Fig. 4 and Tables S9 and S10), but plays a role as seeder of influenza in North America and Europe during April–September.

The pattern shown here is largely consistent with the results of previous phylogeographic reconstructions (12, 14–16). We looked more in depth to importation fluxes across European countries and found a West to East migration pattern compatible with the West-to-East gradient in peak timing observed in the region (58, 60). This highlights that human population distribution, human mobility and seasonal variation in transmission are important drivers of influenza circulation at the global and continental levels.

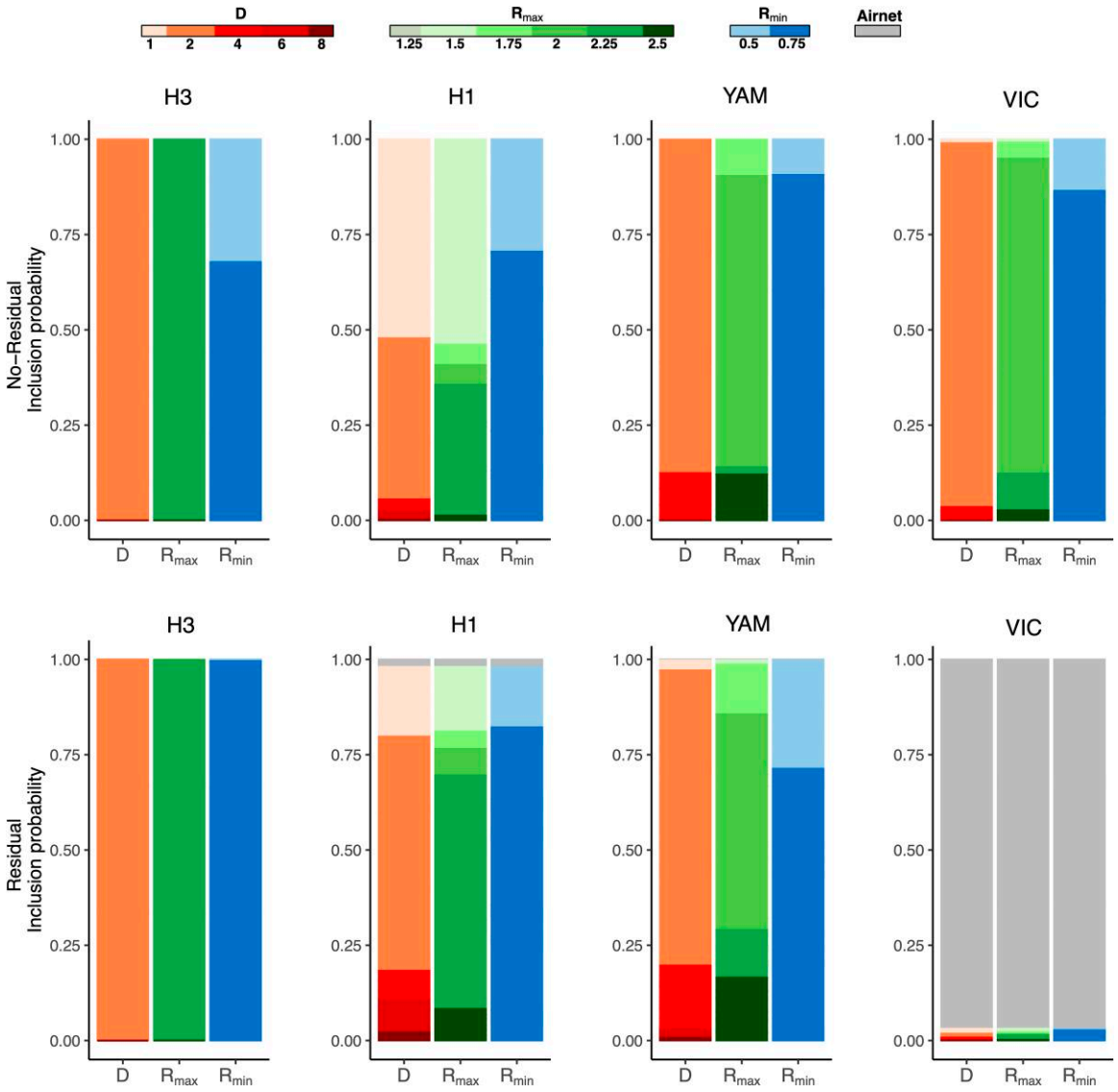

Fig. 2. Marginal posterior inclusion probabilities associated with recurrent travel distribution of GLEAM fluxes for the three parameters R_{max}, R_{min}, and D, without and with residual predictor.

The multiscale nature of influenza dynamics generated by GLEAM with recurrent travel allows simultaneously reconstructing within country spread and global virus circulation, shedding light on the dynamical coupling among countries underlying seasonal epidemic waves. In Fig. 5, we compare local transmission with case importations according to the region of origin. The contribution of importations to local epidemics is important during out-of-season periods as a seeding component that can generate long transmission chains at the beginning of the influenza season. Model predictions on the geographical origin of importations may therefore carry important epidemiological information about the approaching season. Fig. 5 highlights the differences in the behavior between specific countries. A high level of geographical mixing is observed for Australia where importations during summer and at the beginning of the influenza season originate from Southeast Asia, Europe, and North America. For Japan and the United States on the other hand, a geographical pattern emerges in which the large majority of importations originate from a specific region, i.e. Southeast Asia for Japan and South America for United States.

Limitations

Our model captures global circulation patterns that largely explain both incidence and genetic data, with selected parameterizations that are generally consistent for H3N2, H1N1, and YAM. While phylogeographic analysis shows that migration rates are substantially correlated for H3N2, H1N1, and YAM, these are also characterized by different degrees of persistence in specific locations (16). Strain-specific antigenic evolution and its interplay with demography and age structure can affect migration patterns, which is not accounted for in our model. For instance, H1N1 hits more severely the younger population that mix more at the local geographical scale but travel less frequently over long-range distances compared with the adult population (16). This effect has been shown to impact the spatial invasion and the local persistence of an infection (16, 61). Differences in strain-specific patterns may arise also from complex interactions between subtypes that are difficult to capture in a general seasonal model. Further fine-tuning of simulation parameters may also assist in better capturing subtle differences in the strain-specific dynamics. Subdividing countries into three seasonal areas is a standard

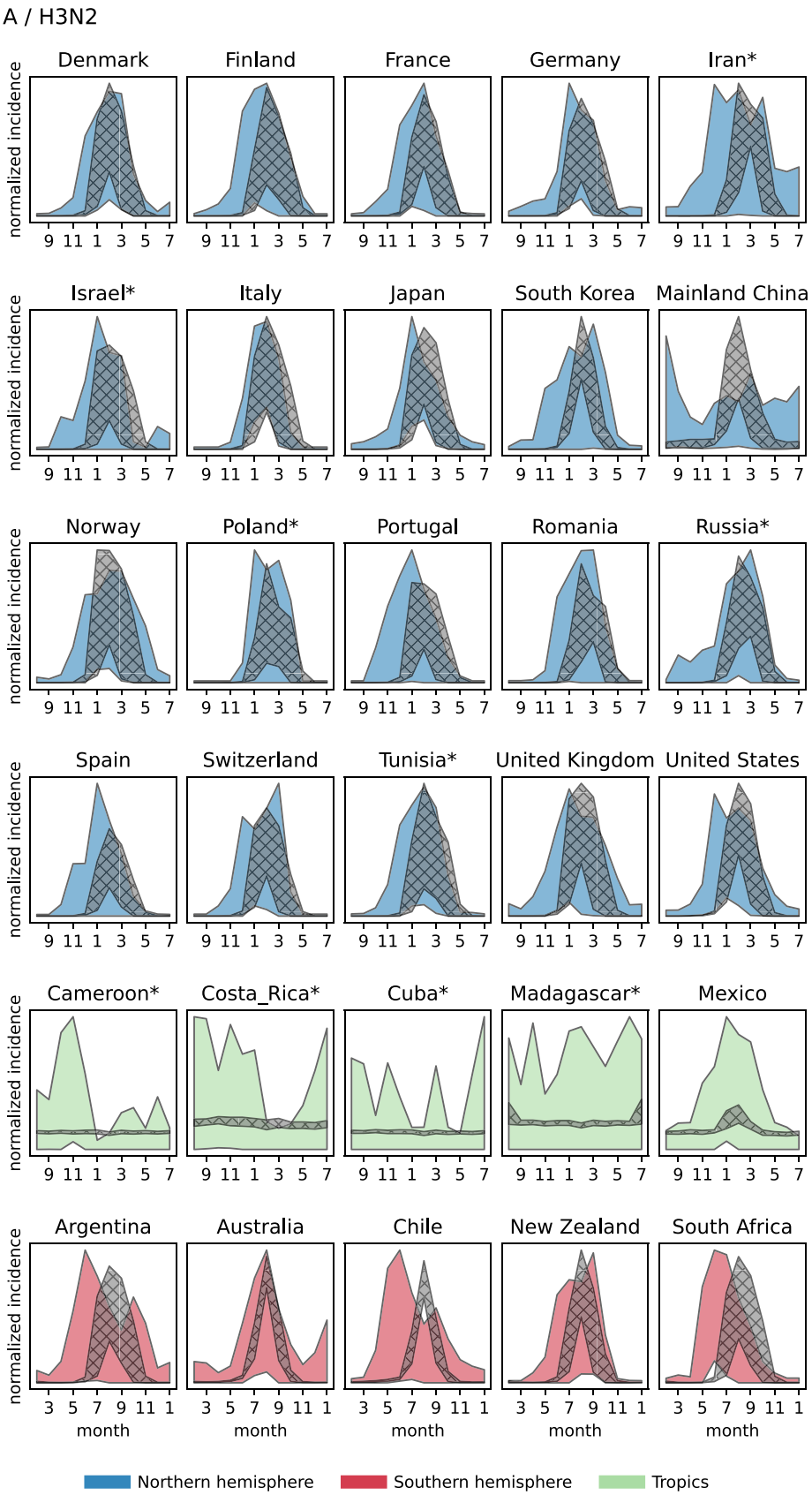

Fig. 3. Annual epidemic profiles for 30 selected countries from FluNet H3N2 samples (colored) and simulations with the best-supported scenario (gray). For each seasonal region, selected countries are representative of the whole set for average correlation and its dispersion. Shaded areas show the 95% CI of the normalized incidence (Section 5 of the Supplementary Material). The * indicates countries for which no genetic data were available for calibration.

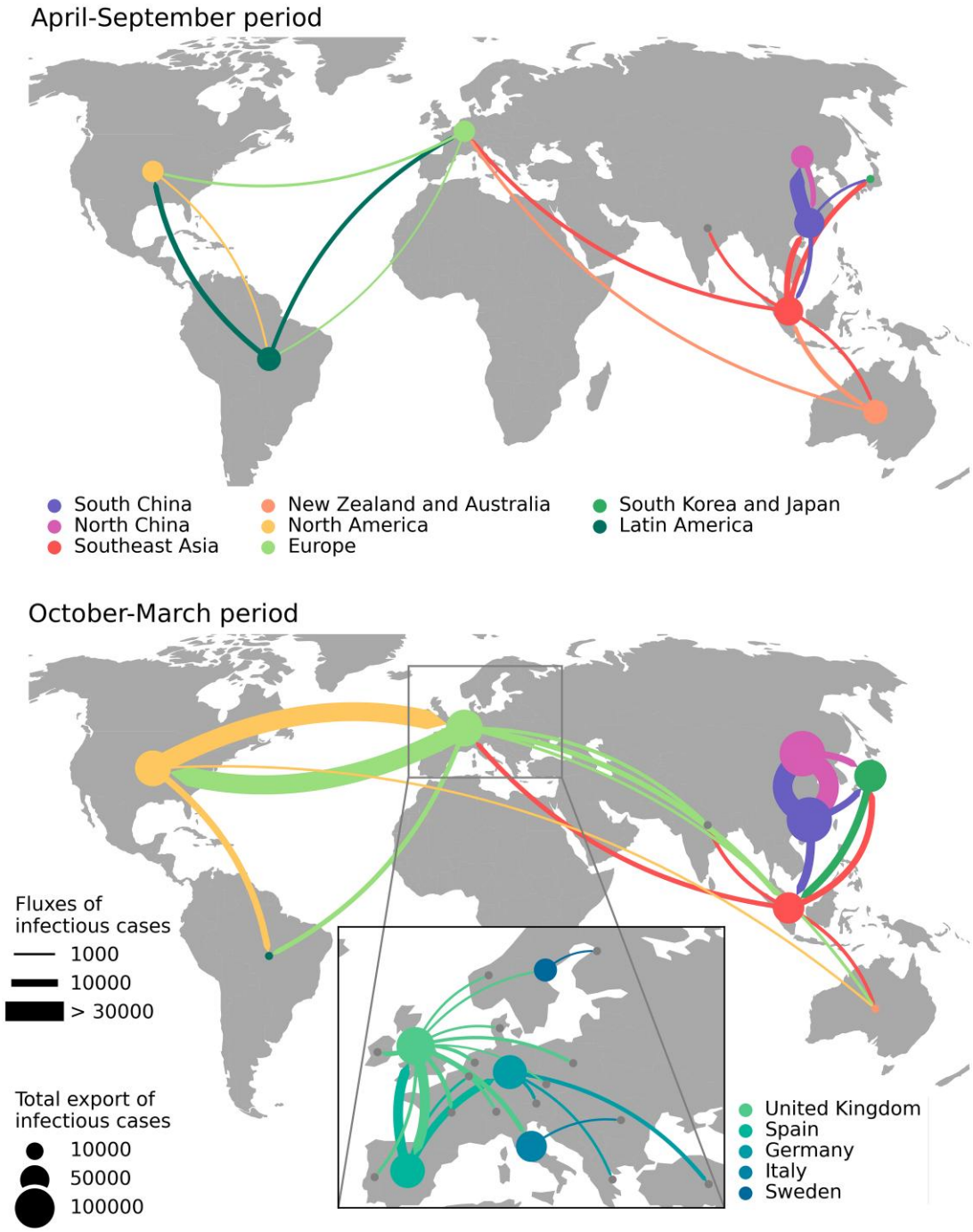

Fig. 4. Dominant fluxes of cases for the two epochs. To enhance clarity, countries have been grouped into geographic areas. We have considered here the same region repartition as in (16). The plots show for each region the fluxes responsible for at least 60% of the importations. In the case of Europe, we show the top 20 countries in terms of importations, each featuring only the most significant importation flux. Fluxes between countries/areas are color-coded according to their country/area of origin.

approach (11, 44, 45). Yet, a uniform model of seasonality in these regions overlooks the environmental and human factors that act at both the country and sub-country levels, which are known to shape epidemic dynamics (6, 8, 58, 62). This could limit our ability to realistically describe influenza patterns, especially in the tropics and southern hemisphere where epidemics are less synchronized compared with the northern hemisphere (6, 58). Heterogeneity in vaccination coverage represents another country-specific aspect that can alter influenza dynamics. The lack of comprehensive information on vaccine coverage by country and the fact that vaccine effectiveness varies from 1 year to another according to the subtype challenge efforts to account for this ingredient in a global, multiannual model of influenza circulation.

Regarding surveillance data, our model requires as input high-quality genetic data collected over several years. However, in many developing countries surveillance is still poor or has improved too recently for a sufficiently long and accurate time series to be available—considering also that influenza migration patterns were highly perturbed during the COVID-19 pandemic years. As many of those countries are located in the tropics or the

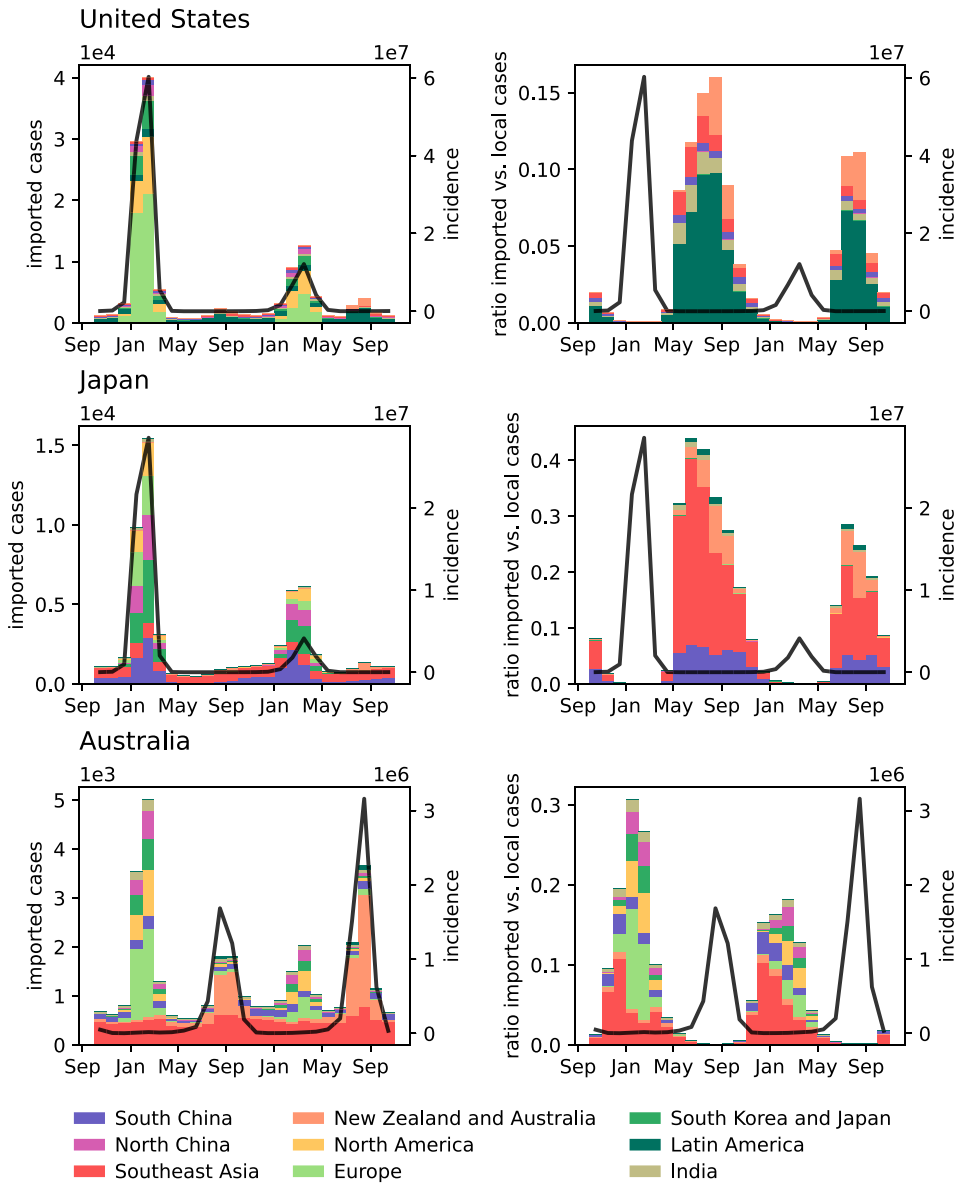

Fig. 5. Left: Number of imported cases (bars, left axis) for two consecutive years from a single stochastic realization—for the best parameterization obtained for H3N2. Right: ratio of imported vs. local cases (bars, left axis). In both plots, simulated incidence is reported with solid black line (right axis) for reference. On both left and right bars are assigned colors based on their respective area of origin (same repartition as in Fig. 4).

southern hemisphere, this further limits our ability to describe epidemics in these regions. Eventually, the global evaluation of the model performance against incidence data requires comparing and averaging metrics computed on countries with different geographical extent, population size and quality of surveillance systems. In the absence of a well-established weighting criterion, we have opted for homogeneous weights to compute average correlations recognizing that different countries have independent surveillance systems.

Conclusions

By designing a combined approach, we are able to employ genetic data to validate and calibrate a dynamical model for the multiscale spread of influenza. The model simulates within-country epidemics, spatial coupling mediated by human mobility, and thus the resulting global circulation of the virus. The information encoded in the genetic data allows for unambiguous identification of the essential epidemiological parameters, whereas incidence data offer lower resolution power. We were able to show that population distribution, local mobility, and international travel, as well as seasonality are fundamental ingredients to accurately model influenza migration patterns.

We have studied a decade of influenza dynamics before the COVID-19 pandemic. Following SARS-CoV-2 emergence in 2020, the global influenza circulation has been substantially altered with potential long-term consequences (17, 63), as illustrated by the probable extinction of the B Yamagata variant (64). In such a situation, the long-term and global-scale description of influenza dynamics is more important than ever to identify viral evolutionary pathways for the prediction of vaccine composition, to inform projections on the approaching influenza season in a given region, and, on a more fundamental level, to disentangle the interplay between endogenous and exogenous factors in shaping regional epidemic waves. Phylodynamic approaches may become an invaluable tool to achieve these goals. Our study provides the starting point of a

new methodological approach that can be further extended with additional ingredients and data layers to better capture geographical variations in transmission, strain-specific features, and their interactions in the post-COVID-19 pandemic era. This will make it possible to improve the description of the source–sink dynamics. Going beyond robust multiannual patterns, a more refined model will address year-by-year variations opening the door to probabilistic forecasting and scenario analysis. This will require performance evaluation metrics that are more sophisticated than simple correlation analyses to account for uncertainty in the predictions (65). Advances in global influenza surveillance (18, 62), along with the increased availability of large-scale datasets (8) will undoubtedly instigate further model developments. For instance, incidence data could be directly employed for calibration, e.g. to resolve situations where genetic data are unable to unambiguously select the best parameterization. In addition, the flexible multistep structure of our approach makes it adaptable to a variety of epidemic models, infectious diseases, and epidemic scenarios.

Material and methods

GLEAM

The GLEAM mobility layer integrates the global flight network with the daily commuting patterns between adjacent patches (66) (see Section 1 of the Supplementary Material). The short-range commuting is accounted for by defining effective patch mixing, based on a time-scale separation approach (11, 28). Air travel mobility is modeled explicitly as a discrete-time multinomial process (67). In Markovian GLEAM, the probability that an individual travels from patch i to patch j is $p_{ij} = w_{ij}\Delta t/N_i$, with N_i being the population size of i and w_{ij} being the daily flux of passengers from i to j in the air transportation data. Traveling probability does not account for the location of residence of individuals. In the recurrent travel GLEAM, the daily flux of travelers w_{ij} is subdivided into individuals resident in i and departing for j, and individuals visiting i and returning to the residence location j (41, 68–70). Leaving and returning home are modeled as distinct processes with average trip duration assumed to be 15 days (28) (i.e. a return rate $1/15\,\text{days}^{-1}$) and departing rate derived from w_{ij} (Section 1.2.3 of the Supplementary Material). Influenza transmission dynamics is modeled within each patch through a compartmental model where individuals are divided in susceptible, latent, symptomatic infectious (that may or may not travel dependent on the severity of symptoms), asymptomatic infectious and recovered, i.e. immune to the virus. The average duration of the exposed and infection period are set to 1.1 and 2.5 days, respectively (11). Given the stochastic nature of the model, each parameterization generates a collection of possible time evolutions for the observables, such as prevalence, peak of infection, number of imported cases, etc., at the spatial resolution of a single patch and time resolution of a day, that can be aggregated at the desired level in time and space.

Phylogeographic analysis

We combine a GLM parameterization of discrete phylogeographic diffusion (12) with epoch modeling (46) in a Bayesian full probabilistic framework (see Section 2 of the Supplementary Material). Both approaches represent extensions of continuous-time Markov chain (CTMC) processes implemented in a Bayesian phylogenetic framework (71). The GLM diffusion model parameterizes the CTMC transition rates as a log linear function of a number of potential predictors and allows estimating both the size of the contribution and the inclusion probability of each predictor. Previous applications have demonstrated how this model averaging approach can identify the predictor or set of predictors that adequately explain the dynamics among locations (12, 72) or among host transitioning (73). Here, we adapt this approach to compare the fit of individual model-based fluxes as predictors of phylogeographic diffusion. In order to model heterogeneity in migration rates through time, and hence to allow for different fluxes predicting these time-variable rates, we adopt an epoch modeling approach (46). The epoch approach partitions evolutionary history into an arbitrary number of time intervals or epochs, separated by transition times, and allows specifying a potentially different CTMC parameterization for each epoch. Here, we set up transition times every six months separating each time the period from March 21 to September 20 from the period from September 21 to March 20 (for brevity, the two periods are named throughout the text April–September and October–March, respectively). We apply two different alternating GLM parameterizations to this April–September and October–March epoch setup.

We perform inference under the GLM and epoch model using Markov chain Monte Carlo (MCMC) integration using BEAST (47).

We fit both time-homogeneous and epoch GLM models with the flux predictors to phylogenetic histories for influenza A subtypes H3N2 and H1N1 and influenza B lineages Yamagata and Victoria, previously analyzed by Bedford et al. (16). The phylogenetic reconstructions are based on hemagglutinin (HA) gene sequence data sets covering a time interval from 2000 to 2012 and representing roughly equitable spatiotemporal distributions across global regions. The original data sets comprise 4,006, 2,144, 1,455, and 1,999 HA sequences for H3N2, H1N1, YAM, and VIC, respectively. We fit our models to the same empirical tree distributions as used in the original work. We run sufficiently long MCMC chains to ensure adequate mixing as assessed by effective sample size estimates. We perform discrete diffusion phylogenetic simulations using πBUSS (74) to explore the performance of the approach (see Section 3 of the Supplementary Material).

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

P.L., V.C., C.P., E.C.G.B. and F.P. acknowledge funding from EU Horizon 2020 grants MOOD (H2020-874850, publication cataloged as MOOD 102). V.C. acknowledges support from Horizon Europe grant ESCAPE (101095619) and Agence Nationale de la Recherche projects DATAREDUX (ANR-19-CE46-0008-03). P.L., M.A.S., and A.R. acknowledge funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement no. 725422-ReservoirDOCS) and from the Wellcome Trust Collaborative Award, 206298/Z/17/Z. M.A.S. acknowledges support from US National Institutes of Health grants U19 AI135995, R01 AI153044, and R01 AI162611. P.L. acknowledges support by the Special Research Fund, KU Leuven ("Bijzonder Onderzoeksfonds," KU Leuven, OT/14/115), and the Research Foundation—Flanders ("Fonds voor Wetenschappelijk Onderzoek—Vlaanderen," G066215N, G0D5117N, and G0B9317N). C.P. acknowledges funding by the Cariparo Foundation through the program Starting Package and the Department of Molecular Medicine of the University of Padova

through the program SID from BIRD funding. T.B. is a Howard Hughes Medical Institute Investigator and is supported by NIH NIGMS R35 GM119774. The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Author Contributions

T.B., V.C., C.P., P.L.: conceptualization; F.P., E.G.-B., T.B., M.A.S., N.S.T., A.R., V.C., C.P., P.L.: data curation; F.P., E.G.-B., T.B., M.A.S., N.S.T., A.R., V.C., C.P., P.L.: formal analysis; F.P., E.G.-B., T.B., M.A.S., N.S.T., A.R., V.C., C.P., P.L.: investigation; F.P., E.G.-B., T.B., V.C., C.P., P.L.: methodology; F.P., E.G.-B., T.B., M.A.S., N.S.T., A.R., V.C., C.P., P.L.: software; V.C., C.P., P.L.: supervision; F.P., E.G.-B., T.B., M.A.S., N.S.T., A.R., V.C., C.P., P.L.: validation; F.P., E.G.-B., T.B., V.C., C.P., P.L.: visualization; F.P., E.G.-B., V.C., C.P., P.L.: writing—original draft; F.P., E.G.-B., T.B., M.A.S., N.S.T., A.R., V.C., C.P., P.L.: writing—review & editing.

Preprints

This manuscript was posted on a preprint: https://doi.org/10.1101/2024.03.14.24303719.

Data Availability

The data sets consisting of influenza HA gene sequences covering the time interval from 2000 to 2012 are available at https://github.com/blab/global-migration. The global influenza model developed in the study was based on GLEAM, which is publicly available at http://www.gleamviz.org/. The BEAST code is openly accessible at https://github.com/beast-dev/beast-mcmc. Model simulation outputs that were used as BEAST input files are available at https://github.com/phylogeography/GLEAM-phylogeography.
The global flight network data are commercially available from the International Air Transport Association (IATA, https://www.iata.org/en/contact-support). For the present study, we used data of 2013.

References

1 Finkelman BS, *et al.* 2007. Global patterns in seasonal activity of influenza A/H3n2, A/H1n1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PLoS One*. 2(12):e1296.
2 Truscott J, *et al.* 2012. Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. *J R Soc Interface*. 9(67):304–312.
3 Axelsen JB, Yaari R, Grenfell BT, Stone L. 2014. Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proc Natl Acad Sci U S A*. 111(26):9538–9542.
4 Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M. 2010. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol*. 8(2):e1000316.
5 Charu V, *et al.* 2017. Human mobility and the spatial transmission of influenza in the United States. *PLoS Comput Biol*. 13(2):e1005382.
6 Tamerius JD, *et al.* 2013. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathog*. 9(3):e1003194.
7 Tamerius J, Viboud C, Shaman J, Chowell G. 2015. Impact of school cycles and environmental forcing on the timing of pandemic influenza activity in Mexican states, May-December 2009. *PLoS Comput Biol*. 11(8):e1004337.
8 Deyle ER, Maher MC, Hernandez RD, Basu S, Sugihara G. 2016. Global environmental drivers of influenza. *Proc Natl Acad Sci U S A*. 113(46):13081–13086.
9 Ewing A, Lee EC, Viboud C, Bansal S. 2017. Contact, travel, and transmission: the impact of winter holidays on influenza dynamics in the United States. *J Infect Dis*. 215(5):732–739.
10 De Luca GD, *et al.* 2018. The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium. *BMC Infect Dis*. 18(1):29.
11 Balcan D, *et al.* 2009. Seasonal transmission potential and activity peaks of the new influenza A(H1n1): a Monte Carlo likelihood analysis based on human mobility. *BMC Med*. 7(1):45.
12 Lemey P, *et al.* 2014. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3n2. *PLoS Pathog*. 10(2):e1003932.
13 Tizzoni M, *et al.* 2012. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Med*. 10(1):1–31.
14 Russell CA, *et al.* 2008. The global circulation of seasonal influenza A (H3n2) viruses. *Science*. 320(5874):340–346.
15 Rambaut A, *et al.* 2008. The genomic and epidemiological dynamics of human influenza A virus. *Nature*. 453(7195):615–619.
16 Bedford T, *et al.* 2015. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*. 523(7559):217–220.
17 Bonacina F, *et al.* 2023. Global patterns and drivers of influenza decline during the COVID-19 pandemic. *Int J Infect Dis*. 128:132–139.
18 Dhanasekaran V, *et al.* 2022. Human seasonal influenza under COVID-19 and the potential consequences of influenza lineage elimination. *Nat Commun*. 13(1):1721.
19 Davis WW, Mott JA, Olsen SJ. 2022. The role of non-pharmaceutical interventions on influenza circulation during the COVID-19 pandemic in nine tropical Asian countries. *Influenza Other Respir Viruses*. 16(3):568–576.
20 Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. 2004. The architecture of complex weighted networks. *Proc Natl Acad Sci U S A*. 101(11):3747–3752.
21 Colizza V, Barrat A, Barthelemy M, Valleron A-J, Vespignani A. 2007. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med*. 4(1):e13.
22 Faucher B, *et al.* 2024. Drivers and impact of the early silent invasion of SARS-CoV-2 Alpha. *Nat Commun*. 15(1):2152.
23 Lemey P, *et al.* 2020. Accommodating individual travel history and unsampled diversity in bayesian phylogeographic inference of SARS-CoV-2. *Nat Commun*. 11(1):5110.
24 Lemey P, *et al.* 2021. Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature*. 595(7869):713–717.
25 Hodcroft EB, *et al.* 2021. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature*. 595:707–712.
26 Dudas G, *et al.* 2021. Emergence and spread of SARS-CoV-2 lineage b.1.620 with variant of concern-like mutations and deletions. *Nat Commun*. 12(1):5769.
27 L-H Tsui J, *et al.* 2023. Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA. 1. *Science*. 381(6655):336–343.
28 Balcan D, *et al.* 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci U S A*. 106(51):21484–21489.
29 Gomes MFC, *et al.* 2014. Assessing the international spreading risk associated with the 2014 west African Ebola outbreak. *PLoS*

Curr. 6. https://doi.org/10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5.
30 Poletto C, *et al.* 2014. Assessing the impact of travel restrictions on international spread of the 2014 west African Ebola epidemic. *Euro Surveill*. 19(42):20936. https://doi.org/10.2807/1560-7917.es2014.19.42.20936.
31 Poletto C, *et al.* 2014. Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach. *Euro Surveill*. 19(23):20824. https://doi.org/10.2807/1560-7917.es2014.19.23.20824.
32 Poletto C, Colizza V, Boëlle P-Y. 2016. Quantifying spatiotemporal heterogeneity of MERS-CoV transmission in the Middle East region: a combined modelling approach. *Epidemics*. 15:1–9.
33 Zhang Q, *et al.* 2017. Spread of Zika virus in the Americas. *Proc Natl Acad Sci U S A*. 114(22):E4334–E4343.
34 Chinazzi M, *et al.* 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. 368(6489):395–400.
35 Davis JT, *et al.* 2021. Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. *Nature*. 600(7887):127–132.
36 Pullano G, *et al.* 2020. Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. *Euro Surveill*. 25(4):2000057.
37 Gilbert M, *et al.* 2020. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *Lancet*. 395(10227):871–877.
38 Rvachev LA, Longini IM. 1985. A mathematical model for the global spread of influenza. *Math Biosci*. 75(1):3–22.
39 Grais RF, Ellis JH, Glass GE. 2003. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *Eur J Epidemiol*. 18(11):1065–1072.
40 Hufnagel L, Brockmann D, Geisel T. 2004. Forecast and control of epidemics in a globalized world. *Proc Natl Acad Sci U S A*. 101(42):15124–15129.
41 Balcan D, Vespignani A. 2011. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nat Phys*. 7(7):581–586.
42 Keeling MJ, Danon L, Vernon MC, House TA. 2010. Individual identity and movement networks for disease metapopulations. *Proc Natl Acad Sci U S A*. 107(19):8866–8870.
43 Longini IM, Halloran ME, Nizam A, Yang Y. 2004. Containing pandemic influenza with antiviral agents. *Am J Epidemiol*. 159(7):623–633.
44 Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ. 2006. Delaying the international spread of pandemic influenza. *PLoS Med*. 3(6):e212.
45 Wen F, Bedford T, Cobey S. 2016. Explaining the geographical origins of seasonal influenza A (H3N2). *Proc Biol Sci*. 283(1838):20161312. https://doi.org/10.1098/rspb.2016.1312.
46 Bielejec F, Lemey P, Baele G, Rambaut A, Suchard MA. 2014. Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. *Syst Biol*. 63(4):493–504.
47 Suchard MA, *et al.* 2018. Bayesian phylogenetic and phylodynamic data integration using beast 1.10. *Virus Evol*. 4(1):vey016.
48 He D, *et al.* 2015. Global spatio-temporal patterns of influenza in the post-pandemic Era. *Sci Rep*. 5(1):11013.
49 Yang W, Lau EcHY, Cowling BJ. 2020. Dynamic interactions of influenza viruses in Hong Kong during 1998–2018. *PLoS Comput Biol*. 16(6):e1007989.
50 Yuan H, Kramer SC, Lau EcHY, Cowling BJ, Yang W. 2021. Modeling influenza seasonality in the tropics and subtropics. *PLoS Comput Biol*. 17(6):e1009050.
51 Dorigatti I, Cauchemez S, Ferguson NM. 2013. Increased transmissibility explains the third wave of infection by the 2009 H1N1 pandemic virus in England. *Proc Natl Acad Sci U S A*. 110(33):13422–13427.
52 Viboud C, Nelson MI, Tan Y, Holmes EC. 2013. Contrasting the epidemiological and evolutionary dynamics of influenza spatial transmission. *Philos Trans R Soc Lond B Biol Sci*. 368(1614):20120199.
53 Nelson MI, Simonsen L, Viboud C, Miller MA, Holmes EC. 2007. Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS Pathog*. 3(9):e131.
54 Nelson MI, *et al.* 2006. Stochastic processes are key determinants of short-term evolution in influenza A virus. *PLoS Pathog*. 2(12):e125.
55 Ghedin E, *et al.* 2010. Unseasonal transmission of H3n2 influenza A virus during the swine-origin H1n1 pandemic. *J Virol*. 84(11):5715–5718.
56 Ross ZP, *et al.* 2015. Inter-seasonal influenza is characterized by extended virus transmission and persistence. *PLoS Pathog*. 11(6):e1004991.
57 Kelly HA, Grant KA, Tay EL, Franklin L, Hurt AC. 2013. The significance of increased influenza notifications during spring and summer of 2010-11 in Australia. *Influenza Other Respir Viruses*. 7(6):1136–1141.
58 Alonso WJ, *et al.* 2015. A global map of hemispheric influenza vaccine recommendations based on local patterns of viral circulation. *Sci Rep*. 5(1):17214.
59 Yu H, *et al.* 2013. Characterization of regional influenza seasonality patterns in China and implications for vaccination strategies: spatio-temporal modeling of surveillance data. *PLoS Med*. 10(11):e1001552.
60 Caini S, Alonso WJ, Séblain CE-G, Schellevis F, Paget J. 2017. The spatiotemporal characteristics of influenza A and B in the who European region: can one define influenza transmission zones in Europe? *Euro Surveill*. 22(35):30606.
61 Apolloni A, Poletto C, Colizza V. 2013. Age-specific contacts and travel patterns in the spatial spread of 2009 H1n1 influenza pandemic. *BMC Infect Dis*. 13(1):176.
62 Alonso WJ, *et al.* 2007. Seasonality of influenza in Brazil: a traveling wave from the Amazon to the subtropics. *Am J Epidemiol*. 165(12):1434–1442.
63 Chen Z, *et al.* 2024. Covid-19 pandemic interventions reshaped the global dispersal of seasonal influenza viruses. *Science*. 386(6722):eadq3003.
64 Caini S, *et al.* 2024. Probable extinction of influenza B/Yamagata and its public health implications: a systematic literature review and assessment of global surveillance databases. *Lancet Microbe*. 5(8):100851.
65 Bracher J, Ray EL, Gneiting T, Reich NG. 2021. Evaluating epidemic forecasts in an interval format. *PLoS Comput Biol*. 17(2):e1008618.
66 International Air Transport Association (IATA), 2013. https://www.iata.org/.
67 Halloran ME, Longini IM, Struchiner CJ. 2010. Binomial and stochastic transmission models. In: *Design and analysis of vaccine studies. Statistics for biology and health*. New York (NY): Springer. p. 63–84.
68 Sattenspiel L, Dietz K. 1995. A structured epidemic model incorporating geographic mobility among regions. *Math Biosci*. 128(1-2):71–91.
69 Keeling MJ, Rohani P. 2002. Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecol Lett*. 5(1):20–29.

70 Poletto C, Tizzoni M, Colizza V. 2012. Heterogeneous length of stay of hosts' movements and spatial epidemic spread. *Sci Rep*. 2(1):476.
71 Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol*. 5(9):e1000520.
72 Nelson MI, *et al*. 2015. Global migration of influenza a viruses in swine. *Nat Commun*. 6(1):6696.
73 Faria NR, Suchard MA, Rambaut A, Streicker DG, Lemey P. 2013. Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Philos Trans R Soc Lond B Biol Sci*. 368(1614):20120196.
74 Bielejec F, *et al*. 2014. πBUSS: a parallel BEAST/BEAGLE utility for sequence simulation under complex evolutionary scenarios. *BMC Bioinformatics*. 15(1):133.