Comparing the transmission potential from sequence and surveillance data of 2009 North American influenza pandemic waves

Venkata R. Duvvuri [a, b, c, d, e, *], Joseph T. Hicks [d, e], Lambodhar Damodaran [d, e],
Martin Grunnill [a, c], Thomas Braukmann [a], Jianhong Wu [c],
Jonathan B. Gubbay [a, b], Samir N. Patel [a, b], Justin Bahl [d, e, f, **]

[a] Public Health Ontario, Toronto, Ontario, Canada
[b] Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
[c] Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada
[d] Center for the Ecology of Infectious Disease, Department of Infectious Diseases, Institute of Bioinformatics, University of Georgia, Athens, Georgia
[e] Department of Epidemiology and Biostatistics, Institute of Bioinformatics, University of Georgia, Athens, Georgia
[f] Duke-NUS Graduate Medical School, Singapore

ARTICLE INFO

ABSTRACT

Technological advancements in phylodynamic modeling coupled with the accessibility of real-time pathogen genetic data are increasingly important for understanding the infectious disease transmission dynamics. In this study, we compare the transmission potentials of North American influenza A(H1N1)pdm09 derived from sequence data to that derived from surveillance data. The impact of the choice of tree-priors, informative epidemiological priors, and evolutionary parameters on the transmission potential estimation is evaluated. North American Influenza A(H1N1)pdm09 hemagglutinin (HA) gene sequences are analyzed using the coalescent and birth-death tree prior models to estimate the basic reproduction number (R_0). Epidemiological priors gathered from published literature are used to simulate the birth-death skyline models. Path-sampling marginal likelihood estimation is conducted to assess model fit. A bibliographic search to gather surveillance-based R_0 values were consistently lower (mean $\leq$ 1.2) when estimated by coalescent models than by the birth-death models with informative priors on the duration of infectiousness (mean $\geq$ 1.3 to $\leq$2.88 days). The user-defined informative priors for use in the birth-death model shift the directionality of epidemiological and evolutionary parameters compared to non-informative estimates. While there was no certain impact of clock rate and tree height on the R_0 estimation, an opposite relationship was observed between coalescent and birth-death tree priors. There was no significant difference (p = 0.46) between the birth-death model and surveillance R_0 estimates. This study concludes that tree-prior methodological differences may have a substantial impact on the transmission potential estimation as well as the evolutionary parameters. The study also reports a consensus between the sequence-based R_0 estimation and surveillance-based R_0

* Corresponding author. Public Health Ontario, Toronto, Ontario, Canada.
** Corresponding author. Center for the Ecology of Infectious Disease, Department of Infectious Diseases, Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA.
E-mail addresses: venkata.duvvuri@oahpp.ca (V.R. Duvvuri), justin.bahl@uga.edu (J. Bahl).
Peer review under responsibility of KeAi Communications Co., Ltd.

estimates. Altogether, these outcomes shed light on the potential role of phylodynamic modeling to augment existing surveillance and epidemiological activities to better assess and respond to emerging infectious diseases.

1. Introduction

Influenza A viruses (IAVs) are a common cause of human respiratory infections that pose a major challenge to public health and significantly contribute to the economic burden on the healthcare system (Dawood et al., 2012; Iuliano et al., 2018). Through antigenic drift and antigenic shift mechanisms, IAVs produce seasonal epidemics annually and pandemics at irregular intervals. These random introductions of novel IAVs into the human population have led to several pandemic events throughout human history, including the 2009 influenza pandemic (Morens & Taubenberger, 2011). The 2009 H1N1 pandemic emerged in Mexico from a novel reassortment virus in swine with a unique genetic composition [named influenza A(H1N1)pdm09] which resulted in an estimated 284,500 deaths globally in the first 12 months of circulation (Dawood et al., 2012). The repeated but random emergence of novel IAV strains warrants a rapid assessment of transmission potential at the early stages of epidemics to assess pandemic potential and devise the most appropriate public health interventions. The basic reproduction number (R_0) is an epidemiological quantity used to measure the transmission potential of disease by estimating the average number of secondary cases produced by a single index case in the early epidemic. An $R_0 = 0$ indicates no transmission, $0 < R_0 \leq 1$ indicates transmission that is ultimately self-limiting, and $R_0 > 1$ indicates transmission with potential for resulting in an epidemic or endemic spread, if not prevented by intervention or chance extinction. Estimates of R_0 assume a population that is naïve to a newly introduced pathogen and therefore is only comprised of susceptible individuals. Alternatively, the effective reproduction number (R_e), measures the transmission potential of pathogen in a population that may have some level of preexisting immunity (through either vaccination or previous infection) and may be protected under non-pharmaceutical interventions (NPIs) (Anderson & May, 1992.; Diekmann et al., 1990).

Advancements in phylodynamic modeling (Drummond & Rambaut, 2007; Grenfell et al., 2004) coupled with the accessibility of real-time pathogen sequence data (Shu & McCauley, 2017) have created alternative opportunities for the rapid characterization of epidemics. *Phylodynamic modeling* is a predictive framework that unifies mathematical epidemiology and statistical phylogenetics to study evolutionary and epidemiological dynamics (Drummond AJ, 2015). During model inference, phylodynamic modeling software platforms require the user to select various prior assumptions to calculate the likelihood of a parameter set for a given genealogy. A key decision during analysis is the selection of the assumed underlying branching process of the evolutionary tree (henceforth referred to as the tree-prior), which are broadly grouped into two families (coalescent and birth-death) that differ fundamentally in how the model views the direction in time (du Plessis & Stadler, 2015; Stadler et al., 2013). Previously, coalescent models have been used to assess the early 2009 pandemic transmission potential (Fraser et al., 2009), the effects of sample size and sample temporal range on the precision and accuracy of parameter estimation (Hedge et al., 2013), and the transmission potential among different geographic regions (Su et al., 2015). However, these studies on the 2009 pandemic have yet to investigate the impact of selecting coalescent versus birth-death tree priors on the basic reproduction number estimation.

In the current study, we conducted a comparative investigation of coalescent and birth-death tree-prior models using hemagglutinin (HA) sequences from the 2009 H1N1 pandemic waves of the three largest North American countries: Mexico, the United States, and Canada. This study had three main objectives: i) to assess the impact of evolutionary parameters and tree-prior choice on the estimation of transmission potential; ii) to understand the impact of informative priors and non-informative priors (default) on the estimation of transmission potential; and iii) to evaluate the relative consensus of reproduction numbers estimated from sequence data and surveillance data.

2. Materials and methods

2.1. Collection of sequences

Between April to December 2009, Mexico reported three pandemic waves of influenza A(H1N1)pdm09 virus cases (Volz et al., 2009), while the United States, and Canada each reported two pandemic waves (Nelson et al., 2011; Zepeda-Lopez et al., 2010). Publicly available full-length influenza A(H1N1)pdm09 hemagglutinin (HA) gene sequences isolated from patient samples during the wave periods in Mexico, the United States and Canada were retrieved from the NCBI Influenza Virus Database (Bao et al., 2008) and GISAID (Shu & McCauley, 2017). Mexico's second and third wave sequences were combined and considered a single second wave due to the overlap of timelines and the distribution of available sequences from Mexico (Supplemental Fig. S1). Sequences were aligned with MUSCLE (Edgar, 2004). Root-to-tip divergence regression was performed to ensure the presence of a molecular clock signal within the time-stamped sequence data using TempEst v1.5 (Rambaut et al., 2016) based on maximum likelihood (ML) phylogenetic trees inferred with RAxML v8.2.10 (Stamatakis, 2014).

2.2. Datasets

The sequence data were divided into fourteen datasets based on pandemic wave and geographic location to investigate the dynamics of individual pandemic waves at various geographic levels, including at the national and state/province level, depending on data availability. Supplemental Table 3 contains a list of accession numbers of HA sequences used in this study.

2.3. Literature review of A(H1N1)pdm09 basic reproduction numbers

A bibliographic search was conducted using the keywords "reproduction number and influenza," "reproductive number and influenza," "R_0 and influenza," "reproduction rate and influenza," and "reproductive rate and influenza" in the NCBI PubMed to collect the estimated reproduction numbers using incidence data and/or laboratory case confirmed data of 2009 pandemic various waves from Mexico, the United States and Canada (Biggerstaff et al., 2014). Literature related to the reproduction number estimates of seasonal H1N1 that are not the influenza A (H1N1)pdm09 strains, seasonal H3N2, and other subtypes of influenza were excluded.

2.4. Phylodynamic modeling

For each dataset, three different approaches were implemented within a Bayesian framework to compare influenza A(H1N1)pdm09 epidemiological quantities given various prior demographic assumptions. We chose two coalescent-based models and one non-parametric birth-death skyline model. Coalescent model analyses were implemented within BEAST v1.10.1 (Suchard et al., 2018), while BEAST v2.5.1 (Bouckaert et al., 2014) was used for birth-death skyline model estimation. The remaining phylogenetic prior assumptions remained constant across analyses. We used an HKY85 substitution model with four gamma rate categories to estimate the rate of evolution and an uncorrelated relaxed molecular clock model that assumes heterogeneous substitution rates across phylogenetic branches. Three independent runs of 100 million generations were performed for each dataset and model. The results from three independent runs were combined after removing 10% burn-in to achieve an Effective Sample Size (ESS) of >200, as diagnosed in Tracer v1.7.1. The estimated uncertainty is indicated by 95% highest posterior density (HPD) values. Tree heights were subtracted from the date of the most recent tip in each dataset to present the time of the most recent common ancestor (TMRCA).

2.5. Estimation of reproduction number from the coalescent models

The two implemented coalescent models varied in the assumed population growth trajectory. The exponential and logistic coalescent population models use an exponential growth function to capture the early growth phase of a new epidemic and a logistic growth function to model the reaching of a peak virus population, respectively. Such models allow for the estimation of growth rates for influenza A(H1N1)pdm09 viral populations. The population growth rate estimates (r, in years) were then used to calculate the epidemic potential for first wave (all susceptible population due to novel strain) and second wave (large proportion of susceptible due to novel clades of influenza A(H1N1)pdm09 and partial immune populations) using $R_0 \approx R_e = \frac{1}{M(-r)}$. Where, M is a moment-generating function of the generation interval (Wallinga & Lipsitch, 2007). For a gamma approximation of the generation interval this leads to $R_0 \approx R_e = \left(1 + \frac{r}{\beta}\right)^{\alpha}$, where r is growth rate, and both α (shape) and β (rate) are the gamma distribution parameters (McBryde et al., 2009, BEAST documentation). Here the choice of the gamma distribution is more biologically realistic than a normal distribution, as the generation of intervals must be non-negative. The mean (μ) of the gamma distribution is equal to $\frac{\alpha}{\beta}$, with its variance (σ^2) equal to $\frac{\alpha}{\beta^2}$ (Bolker, 2008). Therefore, the gamma distribution parameters $\alpha = \frac{\mu^2}{\sigma^2}$ and $\beta = \frac{\mu}{\sigma^2}$. We used mean generation time distribution of influenza A(H1N1)pdm09: mean (μ) = 2.6 days, and standard deviation (σ) = 1.3 days (Cauchemez et al., 2009).

2.6. Estimation of reproduction numbers from birth-death skyline models

In addition to the coalescent-based models, birth-death skyline (BDSKY v1.4.3) models implemented in the BEAST v2.5.1 (Bouckaert et al., 2014) were also employed to study the A(H1N1)pdm09 viral transmission potential during the pandemic waves across various geographic regions. BDSKY is based on the birth-death (BD) process where each infected individual may transmit infection at a rate of transmission (λ) and eventually becomes non-infectious (δ, rate of becoming non-infectious). Each non-infectious individual moves into the susceptible population with probability 's' (Stadler et al., 2013). As the BD prior accounts for the forward flow of time, BDSKY allows for changes in the effective reproduction number (R_e, for clarity, R_0 is used throughout the rest of this article) throughout the pandemic period as a result of host immune variability and/or effectiveness of preventive and control measures.

We set the BD prior parameters using available epidemiological and evolutionary knowledge of the influenza A(H1N1) pdm09 virus. Since δ is the reciprocal of the duration of infectiousness, we used mean = 52 year^{-1}, SD = 1 (infectious period: ~7 days for seasonal influenza (Leekha et al., 2007) and mean = 77 year^{-1}, SD = 1 (infectious period: 4.69 days for the influenza A(H1N1)pdm09 (Tuite et al., 2010) to explore the impact of δ priors on the R_0 estimation with log-normal prior

distribution, and the upper value of real parameter = 10 based on the influenza dynamics. Based on the value for the duration of infectiousness priors, the BDSKY models will be referred to as BDSKY-52 and BDKSY-77, and the non-informative BDSKY model (using a default set of priors) implementation will be referred to as BDKSY-N (where N standards for non-informative). In the case of R_0 priors, the published community-based reproduction numbers were used to set the upper value of the real parameter: Mexico (mean 1.52, 95% CI 1.47 to 2.0), United States (mean 2.35, 95% CI 1.80 to 2.88) and Canada (mean 1.67, 95% CI 1.44 to 2.0) with dimension 1. The R_0 was estimated using the log-normal distribution with default mean and standard deviation. We used one dimension to capture the point estimate of the reproduction number for each dataset. Other BDSKY parameter prior distributions were selected based on the recommendations in previous literature (Barido-Sottani et al., 2018).

2.7. Model selection

Path-sampling marginal likelihood estimation (MLE) was conducted to assess the relative fit of the demographic models to each dataset. The MLE for path sampling ran with 100 path steps and one million lengths of chains (Baele et al., 2012). The log-marginal likelihood estimates from the MLE were used to compute the log Bayes factor [log (BF)] by comparing model M1 and model M2 using:

$$\log BF(M1, M2) = \Pr(D|M1) / \Pr(D|M2),$$

where $M1$, $M2$ are models, D is data, $\Pr(D|M1)$ and $\Pr(D|M2)$ are MLEs of each model. In the case of BDSKY models, BDSKY with non-informative prior was used as model M1 to assess the fit of informative prior models, BDSKY-52 and BDSKY-77. The following interpretations of the log (BF): 0 to 1.1 (no support to either of the model), 1.1 to 3.0 (positive support), 3.0 to 5.0 (strong support), and >5.0 (overwhelming support) were used to determine the best fit model (Kass RE, 1995).

2.8. Statistical analysis

R v4.2.1 software environment was used in the statistical analysis. The Mann-Whitney Unpaired test was conducted to observe the mean difference between reproduction number (R_0) estimates of coalescent and birth-death models. The Kruskal-Wallis test was used to compare the mean difference of R_0 estimates from sequence and surveillance data. Visualizations were generated using tidyverse v1.3.2, ggplot2 v3.3.6, ggpubr v0.4.0, ggthemes v4.2.4, and gridExtra v2.3 packages implemented in R v4.2.1.

3. Results

To quantify the epidemiological and evolutionary parameters, HA gene sequence data from both pandemic waves of North American influenza A(H1N1)pdm09 (Table 1) were used in the coalescent (CE: exponential growth and CL: logistic growth) and Birth-Death Skyline (BDSKY-52, BDSKY-77 with informative priors and BDSKY-N with non-informative priors) models. Supplemental Fig. S1 represents the sequence distribution across Mexico, Canada, and the United States, overlaying the HA sequences on the weekly magnitude and temporal distribution of 2009 pandemic waves represented by the influenza A(H1N1)pdm09 positive test data reported to the CDC.

Table 1
Datasets used in the study.

Dataset	Pandemic Wave	Wave duration (in months)	Number of Sequences
MEX	First	Apr to Aug 2009 (5)	42
MEX	Second	Sep to Dec 2009 (4)	31
CAN-All	First	Apr to Aug 2009 (5)	69
CAN-All	Second	Sep to Dec 2009 (4)	73
CAN-Ontario	First	Apr to Aug 2009 (5)	32
CAN-Ontario	Second	Sep to Dec 2009 (4)	37
CAN-Quebec	Second	Sep to Dec 2009 (4)	38
USA-All	First	Apr to Aug 2009 (5)	342
USA-All	Second	Sep to Dec 2009 (4)	542
USA-NewYork	First	Apr to Aug 2009 (5)	68
USA-NewYork	Second	Sep to Dec 2009 (4)	79
USA-Texas	First	Apr to Aug 2009 (5)	30
USA-Wisconsin	First	Apr to Aug 2009 (5)	32
USA-Wisconsin	Second	Sep to Dec 2009 (4)	86

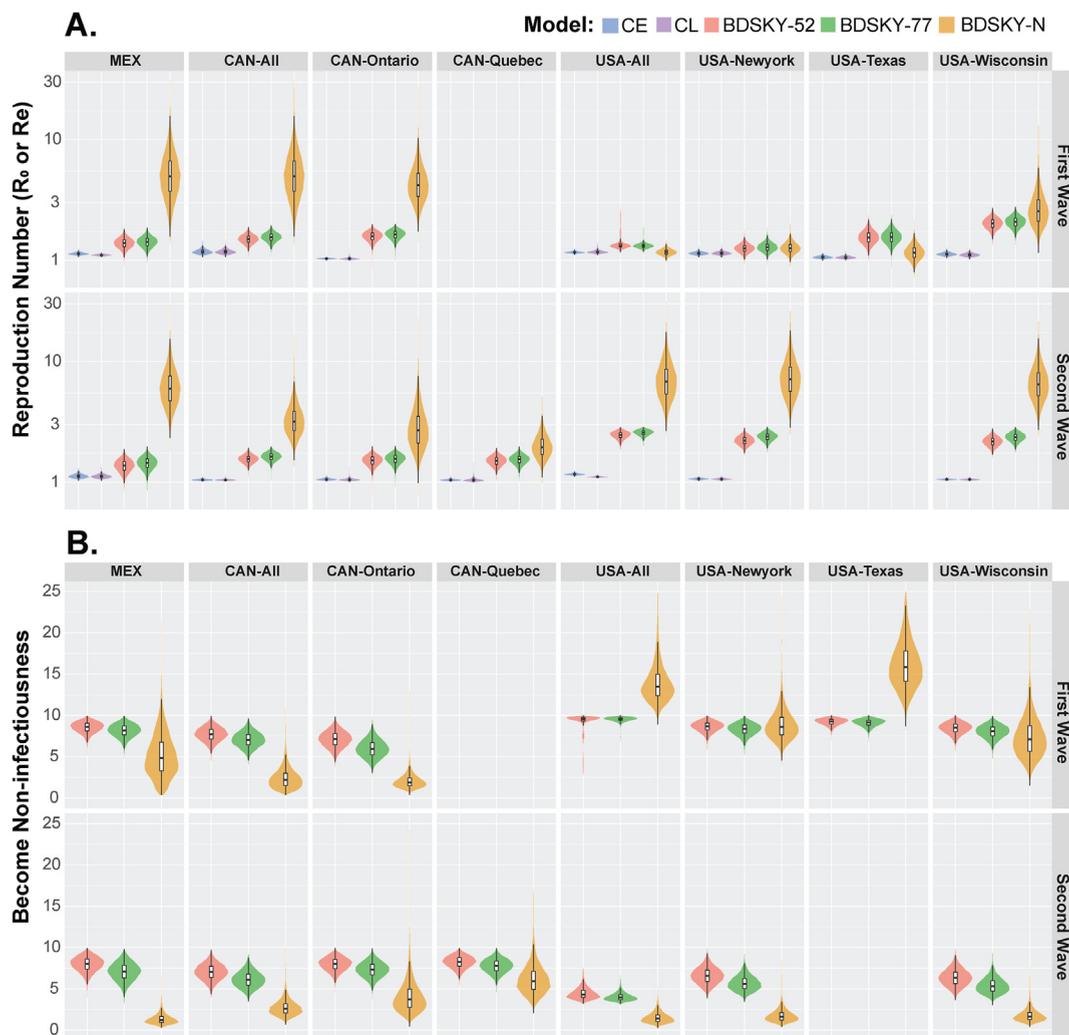Regions: CAN. Canada, MEX. Mexico, and USA. United States of America.

Fig. 1. Epidemiological parameter estimates from coalescent and BDSKY models across 2009 pandemic waves and North American geographic regions. A. Transmission potential (R_0 or R_e) and B. Become non-infectiousness (δ). BDSKY: Birth-Death Skyline, CE & CL: Coalescent exponential growth and coalescent logistic growth models. Empty plot: no data available.

3.1. Tree-priors on the transmission potential estimation

Fig. 1 presents transmission potential observed from coalescent and birth-death tree priors across all geographic regions during both pandemic waves. From both coalescent growth models, the observed mean R_0 estimates at the national level were 1.09–1.19 (95% HPD 1.07 to 1.28) and 1.04 to 1.16 (95% HPD 1.0 to 1.27) during the periods of first and second waves, respectively. The mean R_0 estimates from the state/provincial level of the first wave ranged from 1.10 to 1.14 (95% HPD 0.99 to 1.26), while in the second wave, 1.00 to 1.14 (95% HPD 1.00 to 1.26), respectively. The Mann-Whitney test revealed no significant difference ($p = 0.9489$) between mean R_0 estimates of both coalescent exponential (CE) and coalescent logistic (CL) models.

In contrast to the coalescent models, BDSKY models (BDSKY-52, BDSKY-77) estimated higher mean R_0 values with the same datasets, which are significantly different from the mean R_0 estimates of CE ($p = 0.0021$) and the CL ($p = 0.0006$) based on the Mann-Whitney test (Fig. 1A). The mean R_0 estimates at the national level ranged from 1.31 to 1.55 (95% HPD 0.85 to 1.99) and from 1.37 to 2.56 (95% HPD 0.88 to 2.88) during the first and second wave periods, respectively. The highest mean R_0 estimates, i.e., >2.0, were noticed from the first wave in Wisconsin (2.02–2.08, 95% HPD 1.39–2.81) and the second wave of the United States (2.45–2.56, 95% HPD 1.9 to 2.88), New York (2.21–2.38, 95% HPD 1.73–2.87) and Wisconsin (2.18–2.35, 95% HPD 1.64–2.88).

Next, to examine the role of informative and non-informative priors on the rate of becoming non-infectiousness (δ), a measure of the rate of infectious individuals being removed from the population through immunity or mortality, and reproduction number (R_0 and R_e) on the dynamics of pandemic waves, we fitted two BDSKY models with informative priors (BDSKY-52, and BDSKY-77 with dimensions 10) and one BDSKY model (BDSKY-N) without prior information. The R_0 estimation from BDSKY-N revealed that USA and Mexico showed lower mean R_0 (1.16—2.77, 95% HPD 0.65 to 8.65) during the first wave and higher mean R_0 (6.84—7.65, 95% HPD 1.70 to 16.86) during the second wave. In contrast an opposite trend was found with the Canada and provincial datasets (Fig. 1A). Comparing of BDSKY-N estimates with BDSKY-52 and BDSKY-77 revealed the impact of informative priors in skewing the R_0 estimates either upward or downward (Fig. 1A).

Fig. 1B presents the observed rate estimates of δ (in days) with the informative BDSKY models (BDSKY-52 and BDSKY-77) and non-informative BDSKY model (BDSKY-N). Overall, with BDSKY-N models, we observed the highest mean δ estimates with wider intervals from first wave datasets than from the second wave datasets. The higher mean δ estimates (7.34—16.20, 95% HPD 0.98—26.35) were observed from first wave datasets of USA (13.95, 95% HPD 9.35—20.53), Texas (16.20, 95% HPD 7.7—26.35), New York (8.83, 95% HPD 3.9—14.76) and Wisconsin (7.34, 95% HPD 0.98—15.06) while lower mean δ estimates were observed from the first wave of Canada (mean 2.38, 95% HPD 0.15—6.47) and Mexico (mean 5.16, 95% HPD 0.17—13.03). In contrast to the first wave, the lowest mean δ rates with close-fitting intervals were noticed from the second wave of USA (1.51, 95% HPD 0.17—3.70), New York (1.76, 95% HPD 0.19—4.19) and Wisconsin (1.78, 95% HPD 0.23—4.19) and Mexico (1.36, 95% HPD 0.16—3.39). From the Canadian provincial datasets, we observed 2-3-fold higher δ estimates (4.05—6.14, 95% HPD 0.55—11.86) than from the Canadian dataset (2.71, 95% HPD 0.48—5.76). BDSKY models with prior information narrowed wide intervals or increased narrow intervals estimated with BDSKY-N, where mean δ estimates were 5.91—9.52 (95% HPD 2.91—10 days) and 4.45 to 8.23 (95% HPD 3.04—9.99 days) during the first wave and second wave, respectively. Altogether, these results suggest that there was an impact of coalescent and birth-death tree priors on the R_0 estimation and that informative priors used in the birth-death model can influence R_0 estimation.

3.2. Clock rate and tree height on transmission potential estimation

Molecular clock rates and tree heights estimated from analyses assuming coalescent and birth-death tree-priors are presented in Supplemental Fig. S2. Except for the second wave in the USA, the mean clock rates (substitutions/site/year) estimated from the coalescent and BDSKY models with a relaxed molecular clock revealed that coalescent estimates were found to be lower compared with the informative BDSKY models, BDSKY-52, and BDSKY-77.

The mean of the time of the most recent common ancestor (TMRCA) of both coalescent models ranged from November 08, 2008 to March 25, 2009 (95% HPD March 09, 2008 to May 11, 2009) across first wave data sets and overlapped with the informative BDSKY mean TMRCA estimates (February 02, 2009 to April 28, 2009, 95% HPD, November 28, 2008 to May 18, 2009). The estimates from the non-informative BDSKY model (BDSKY-N) were shown to be wider with Mexico and Canada datasets. In contrast narrower HPD estimates were observed with USA datasets nationally and at the state level. For the second wave, despite an overlap, the coalescent TMRCA estimates (December 07, 2008 to April 28, 2009; 95% HPD, January 01, 2008 to August 22, 2009) were found to be variable and higher than the BDSKY-52 and BDSKY-77 estimates (February 27, 2008 to July 31, 2009; 95% HPD, October 20, 2007 to August 28, 2009) across all regions except USA. These second wave TMRCA estimates of the USA (March 02, 2008, 95% HPD November 17, 2007 to June 20, 2008) from informative prior BDSKY models were far from the TMRCA estimates of Mexico (May 09, 2009, HPD 95% March 01, 2009 to June 11, 2009) and Canada (May 21, 2009, HPD 95% March 17, 2009 to July 20, 2009).

We noticed an inconsistency in the clock rate estimates in both waves and the tree height estimation with the second wave. This observation led us to explore the relationship between clock rate and tree height with transmission potential (R_0). A positive correlation was observed between the coalescent clock rate and R_0, while with the birth-death tree-prior, an inverse relationship was noticed (Figs. 2A and 3A). The tree height estimates from the coalescent models were shown to be negatively correlated with R_0, whereas the opposite relationship was true in the case of the birth-death models (Figs. 2B and 3B).

4. Model fit

The log Bayes Factor is a ratio of two marginal likelihood estimations calculated from the path sampling analysis of two models (Table 2) to indicate statistical support for one model over other. Among coalescent models, the log(BF) estimates supported both coalescent exponential growth and logistic growth fit in some regions and pandemic waves. The lack of both coalescent models' support in other regions was also observed. Log(BF) estimates overwhelmingly supported BDSKY models with prior information (BDSKY-52 and BDSKY-77) compared with the non-informative prior model (BDSKY-N). Though BDSKY-52 and BDSKY-77 were supported in all datasets, we observed stronger support from one prior than the other in some cases.

4.1. Comparing R_0 estimates from sequence and surveillance data

A total of seventeen published articles (Supplemental Table S2) were identified from a bibliographic search to gather the surveillance-based R_0 estimates related to first and second waves of influenza A(H1N1)pdm09 from Mexico (Balcan et al.,
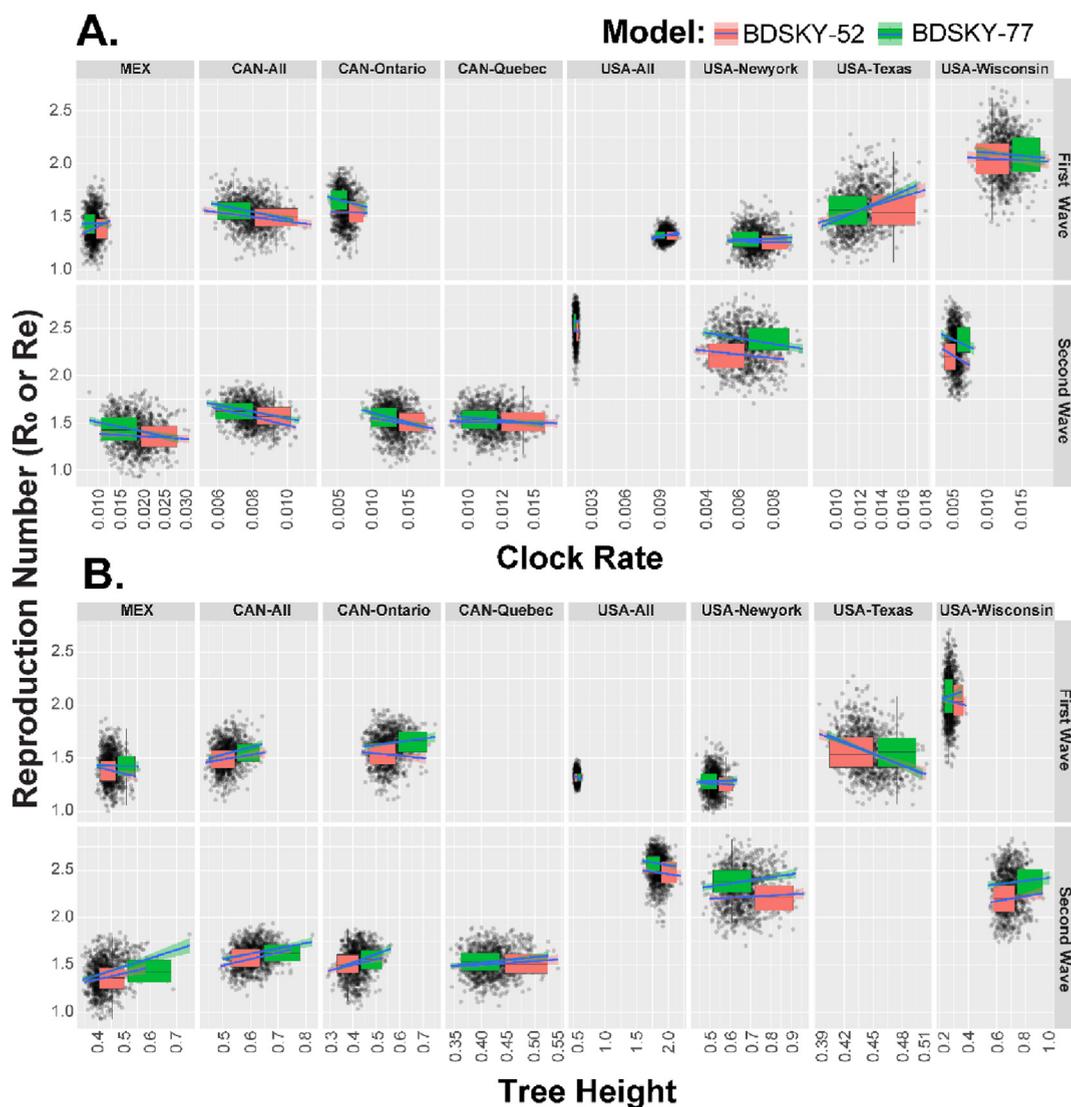
Fig. 2. Relationship of coalescent tree prior's clock rate and tree height on the estimation of transmission potential (R_0). A. Clock rate, and B. Tree height. BDSKY: Birth-Death Skyline, CE & CL: Coalescent exponential and coalescent logistic; and Empty plot: no data available. Data points are posterior results from each model.

2009; Boelle et al., 2009; Cauchemez et al., 2009; Chowell et al., 2011; Fraser et al., 2009; Pourbohloul et al., 2009; White & Pagano, 2010; Yang et al., 2009), the United States(Lessler et al., 2009; Sugimoto et al., 2011; White et al., 2009) and Canada (He et al., 2013; Hsieh et al., 2010; Mostaco-Guidolin et al., 2011, 2012; Sharomi et al., 2011; Tuite et al., 2010). The mean R_0 estimates for the first wave were in the range of 1.22–2.3 (95% CI 1.05 to 2.5) from Mexico, 1.7 to 3.3 (95% CI 1.4 to 3.6) from the United States, and 1.3 to 2.45 (95% CI 1.12 to 3.75) from Canada. The mean R_0 estimates for the second wave were 1.62 (95% CI 1.61 to 1.63) for Mexico and 1.22 to 1.91 (95% CI 1.43 to 2.03) for Canada. There were no reports on the R_0 estimates from USA second wave.

Fig. 4 depicts the sequence-based R_0 values estimated by coalescent and birth-death tree-priors and surveillance-based R_0 values collected from the published literature. Highly significant mean R_0 differences were observed between the coalescent and BDSKY (p = 1.4E-10) estimates, as well as between coalescent and surveillance (p = 1.4E-12). No significant mean difference (p = 0.46) was observed between the BDSKY and the surveillance R_0 estimates. In summary, comparing means using the Kruskal-Wallis test revealed that the BDSKY R_0 estimates are in consensus with the surveillance-based R_0 estimates.

Fig. 3. Relationship of birth-death tree-prior's clock rate and tree height on the estimation of transmission potential (R_0). A. Clock rate, and B. Tree height. BDSKY: Birth-Death Skyline, CE & CL: Coalescent exponential and coalescent logistic; and Empty plot: no data available. Data points are posterior results from each model.

5. Discussion

Our comprehensive analysis, demonstrated the impact of tree-priors, informative and non-informative epidemiological priors, and evolutionary parameters (clock rate and tree height) in assessing viral transmission potential (R_0) during successive pandemic waves. Agreement was found in R_0 estimates derived from pathogen sequence and surveillance data, thereby illustrating the reliability of sequence data in inferring epidemiological parameters.

Model fit analysis revealed that the birth-death model with informative priors had better support for inferring transmission potential than the simple coalescent growth models. This is most likely a reflection of differences in the methodology of the birth-death model compared to coalescent model. Namely, projecting time forward with respect to transmission (births) and recovery (death), when sampling parameters to estimate the transmission potential (Drummond AJ, 2015; Stadler et al., 2013). Little or no difference was observed between the R_0 estimates of coalescent exponential and coalescent logistic models, consistent with previous studies (Fraser et al., 2009; Hedge et al., 2013). Compared to the informative birth-death tree-prior and surveillance data estimates, our coalescent models appeared to underestimate R_0 across all geographic regions and different sample sizes during both pandemic waves; however, these coalescent estimates overlapped with previously reported coalescent R_0 estimates both in North America and other parts of the world, including Europe, Australia, and Singapore (Fraser et al., 2009; Hedge et al., 2013; Su et al., 2015). These lower R_0 estimates suggest that the coalescent

Table 2
Bayes factors to determine the best fitting model.

Dataset	Pandemic Wave	log(BF) based on log-MLEs of two comparative models			Model Fit based on log(BF) ≥ 1.1	
		CE vs. CL	BDSKY-52 vs. BDSKY-N	BDSKY-77 vs. BDSKY-N	Coalescent growth models	Informative prior BDSKY models
MEX	First	3.91	19.42	19.73	CE	both
MEX	Second	0.03	31.34	31.45	none	both
CAN-All	First	0.43	16.71	19.30	none	both
CAN-All	Second	2.17	249.94	15.65	CL	BDSKY-52
CAN-Ontario	First	0.54	14.02	13.73	none	both
CAN-Ontario	Second	0.51	16.73	16.25	none	both
CAN-Quebec	Second	0.51	21.30	22.26	none	both
USA-All	First	3.37	4231.78	75.74	CL	BDSKY-52
USA-All	Second	0.61	9096.50	9850.93	none	BDSKY-77
USA-NewYork	First	3.9	24.46	26.45	CE	both
USA-NewYork	Second	1.98	4.11	4.87	CE	both
USA-Texas	First	1.3	17.44	16.58	CE	both
USA-Wisconsin	First	0.98	15.72	16.59	none	both
USA-Wisconsin	Second	3	19.69	1188.17	CL	BDSKY-77

CAN. Canada, MEX. Mexico, and USA. United States of America; Italicized numbers: supported by Bayes Factor. Italicized & bold-faced numbers: showed strong support between BDSKY-52 and BDSKY-77. none: no support from both coalescent models; both: BDSKY-52 and BDSKY-77 supported. A log Bayes Factor (logBF) represents better fit over a model where log(BF) = 0 to 1.1 (no support to either of the model), 1.1 to 3.0 (positive support), 3.0 to 5.0 (strong support), and >5.0 (overwhelming support) (Kass RE, 1995).

growth model's dependence on a single parameter (i.e., deterministic population size) limits accuracy in capturing transmissibility strength (Stadler et al., 2013; Drummond AJ, 2015). Hence, these lower coalescent R_0 estimates need to be verified by alternative birth-death skyline models, which allow capturing changes in the epidemiological parameters through time based on pathogen sequence data (Stadler et al., 2013).

The highest mean R_0 estimates via BDSKY modeling with informative priors (mean R_0 = 2.18 to 2.56), from the United States, are in alignment with previous phylodynamic studies (Duchene et al., 2019; Featherstone et al., 2021). Duchene et al. (2019) reported the highest mean R_0 estimate was 3.01 (95% CI 2.5 to 3.7) using a birth-death-sampling SIR (BDSIR) model with 328 whole genome sequences (WGS) collected from April to December 2009 and an informative δ prior (μ = 85, σ = 15; based on the reciprocal of the infectious period ~4.45 days). Featherstone et al. (2021) produced an R_0 estimate of 2.10 (95% HPD 1.60–2.54), with a BDSKY model using 639 WGS (sampling time: early April to October 2009) with an informative δ prior (μ = 91, σ = 1 based on the reciprocal of the infectious period ~4.0 days). The higher R_0 estimates for the fall wave in New York (2.21–2.38, 95% HPD 1.73 to 2.87) and Wisconsin (2.18–2.35, 95% HPD 1.64 to 2.88) are likely associated with combinatory role of epidemiological drivers: circulation of a single dominant viral lineage, i.e., clade 7 (Nelson et al., 2011), fall school openings (Chao et al., 2010; Earn et al., 2012; Gog et al., 2014; Lessler et al., 2009), low absolute humidity (Gog et al., 2014; Shaman et al., 2011), vulnerable remote communities (Mostaco-Guidolin et al., 2011, 2012) and spatio-temporality (He et al., 2015).

Calculating the critical immunization threshold (Fine et al., 2011) using these BDSKY R_0 values suggested that 53%–67% of fully immune people are required to stop the spread of the 2009 H1N1 infection. In contrast, lower vaccine coverage was indicated in the United States from October 2009 to May 2010 based on the lower median 2009 H1N1 monovalent vaccine coverage of 16.7%–40.2% in the US (Centers for Disease Control and Prevention (CDC), 2011). Such low vaccine coverage may have led to the large Fall wave associated with the higher mortality and hospitalizations in high-risk population (Centers for Disease Control and Prevention (CDC), 2010; Nelson et al., 2011; Truelove et al., 2011). Estimating R_0 using early pandemic or epidemic pathogen genome data aids in understanding epidemic behavior. Thereby, allowing public health officials to respond with suitable interventions aimed at breaking infectious disease transmission chains. The accessibility of pathogen genetic data, along with the availability of advanced phylodynamic methods, has enabled the exploration of alternative approaches to characterizing an epidemic early in its development to inform decisions on appropriate mitigation strategies (Armstrong et al., 2019; Attwood et al., 2022; Dudas et al., 2017; Duvvuri et al., 2015; Fourment et al., 2018; Fraser et al., 2009; Pybus et al., 2001; Stadler et al., 2012, 2014; Volz et al., 2017).

Exploration with genomic data has led to the validation of surveillance-based reproduction numbers estimated by conventional mathematical models with the sequence-based estimates by Bayesian phylogenetic modeling (Biek et al., 2007; Duvvuri et al., 2015; Fraser et al., 2009; Fraser et al., 2009, 2009; Pybus et al., 2001; Stadler et al., 2012, 2014). In this study, we observed an agreement between reproduction number estimates of informative BDSKY models (BDSKY-52 and BDSKY-77) and surveillance data during both waves and across all study regions. Similarly, Stadler et al. (2014) reported consistency in recovering epidemiological estimates using 2014 Ebola outbreak sequences in Sierra Leone from a wide range of phylodynamic methods compared with WHO surveillance estimates. In summary, this accumulated evidence from different pathogens supports that R_0 estimated from viral sequence data using birth-death models can be reliable in quantifying the
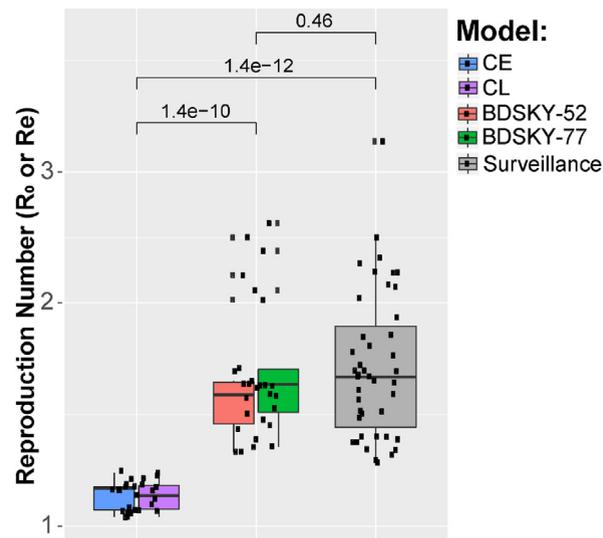
Fig. 4. Correlation of R_0 estimated from sequence and surveillance data. Both CE (coalescent exponential growth) and CL (coalescent logistic growth) are coalescent models. BDSKY: birth-death skyline model; coalescent and BDSKY models were used to estimate R_0 from the HA sequences; surveillance represents published R_0 estimates from the 2009 pandemic H1N1 incidence data.

transmission potential of an epidemic. However, care must be taken while choosing the prior distributions for disease-specific parameters.

In agreement with published studies (Boskova et al., 2018; Moller et al., 2018), the choice of the tree prior had a substantial influence on the clock rate and tree height estimation. We also found that birth-death models with informative priors can inflate clock rate estimates compared to non-informative ones. From previous studies, it has been reported that clock rate can also be influenced by multiple factors, including purifying selection (Ho et al., 2005, 2011), sequence sampling time (Moller et al., 2018), calibration errors, model misspecification, and sequencing errors (Ho et al., 2011). While it is evident that tree priors can impact evolutionary parameters and R_0 estimation, we did not observe any distinct trends between clock rate and R_0 or tree height and R_0. One explanation is that in the birth-death model, the birth rate, death rate, and sampling proportion has a stronger effect on R_0 than the evolutionary parameters.

Our study has important limitations to consider, including sampling frequency for sequence data from a given region and strong informative priors for birth-death models. This study space describes the initial stages of a pandemic during a time when sequencing capabilities were not robust, which can lead to sampling biases. Newer and more extensive methodologies, such as the right-size sampling approach proposed by the CDC and the Association of Public Health Laboratories (APHL), are a solution to these regional sampling biases and allow for better geographic characterization of influenza transmission (Rosenthal et al., 2017). Birth-death models require strong informative priors to accurately estimate the parameters of the birth-death process; this can lead to heavy bias and variability in parameter estimates (Volz & Frost, 2014). This limitation can be addressed by estimating informative priors from implementations of birth-death models for case data and by allowing Bayesian phylogenetic reconstructions to have a longer run-time with lower sampling probabilities (Boskova et al., 2014). The generation interval is a key component in the gamma approximation. Hence, care must be taken when implementing gamma approximation for chronic diseases such as HIV and Tuberculosis (Park et al., 2019).

In conclusion, this comprehensive analysis of the 2009 H1N1 pandemic emphasizes the significant role phylodynamics can play in inferring the transmission potential of an emerging disease. Additionally, we described epidemiological and evolutionary parameters that affect R_0 estimation and provided evidence of the robustness and comparability of R_0 estimation with surveillance-based estimates. The accumulated evidence from the phylodynamic modeling in characterizing infectious disease outbreaks reinforces the importance of pathogen sequence data in public health surveillance and epidemiology (Armstrong et al., 2019; Attwood et al., 2022; World Health Organization, 2021). Hence, this methodology should be implemented in routine genomic epidemiological surveillance to enhance monitoring, assessment, and response capacities.

Funding

Author contributions

Conceptualization: V.R.D. and J.B., methodology: V.R.D., J.T.H. and J.B, formal analysis: V.R.D, funding acquisition: J.B., investigation: V.R.D., supervision: J.B., validation: J.T.H., J.B., visualization: V.R.D., writing - original draft preparation: V.R.D., writing − review & editing: V.R.D., J.T.H., L.D., M.G., T.B., J.W., J.B.G., S.N.P., and J.B.

Declaration of competing interest

All authors declared that they have no competing interests.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2023.02.003.

References

Anderson, R. M., & May, R. M. (Eds.). (1992). *Infectious diseases of humans dynamics and control*. Oxford, UK: Oxford University.

Armstrong, G. L., MacCannell, D. R., Taylor, J., Carleton, H. A., Neuhaus, E. B., Bradbury, R. S., et al. (2019). Pathogen genomics in public health. *New England Journal of Medicine, 381*(26), 2569–2580.

Attwood, S. W., Hill, S. C., Aanensen, D. M., Connor, T. R., & Pybus, O. G. (2022). Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nature Reviews Genetics, 23*(9), 547–562.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., & Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution, 29*(9), 2157–2167.

Balcan, D., Hu, H., Goncalves, B., Bajardi, P., Poletto, C., Ramasco, J. J., 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 Medicine, 7*, 45-7015-7-45.

Bao, Y., Bol, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., et al. (2008). The influenza virus resource at the national center for biotechnology information. *Journal of Virology, 82*(2), 596–601.

Barido-Sottani, J., Boskova, V., Plessis, L. D., Kuhnert, D., Magnus, C., Mitov, V., et al. (2018). Taming the BEAST-A community teaching material resource for BEAST 2. *Systematic Biology, 67*(1), 170–174.

BEAST documentation. Estimating R0 from coalescent growth rates | BEAST Documentation. https://beast.community/estimating_R0. (Accessed 4 February 2023).

Biek, R., Henderson, J. C., Waller, L. A., Rupprecht, C. E., & Real, L. A. (2007). A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of the National Academy of Sciences of the United States of America, 104*(19), 7993–7998.

Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., & Finelli, L. (2014). Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: A systematic review of the literature. *BMC Infectious Diseases, 14*, 480-2334-14-480.

Boelle, P. Y., Bernillon, P., & Desenclos, J. C. (2009). A preliminary estimation of the reproduction ratio for new influenza A(H1N1) from the outbreak in Mexico, march-april 2009. *Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = Eur. Commun. Dis. Bull., 14*(19), Article 19205. https://doi.org/10.2807/ese.14.19.19205-en

Bolker, B. M. (2008). 4.5.2.3. Gamma. In *Ecological models and data in R*. Princeton University Press, ISBN 9781400840908.

Boskova, V., Bonhoeffer, S., & Stadler, T. (2014). Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS Computational Biology, 10*(11), Article e1003913.

Boskova, V., Stadler, T., & Magnus, C. (2018). The influence of phylodynamic model specifications on parameter estimates of the zika virus epidemic. *Virus Evol., 4*(1), vex044.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C., Xie, D., et al. (2014). Beast 2: A software platform for bayesian evolutionary analysis. *PLoS Computational Biology, 10*(4), Article e1003537.

Cauchemez, S., Donnelly, C. A., Reed, C., Ghani, A. C., Fraser, C., Kent, C. K., et al. (2009). Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *New England Journal of Medicine, 361*(27), 2619–2627.

Centers for Disease Control and Prevention (CDC). (2010). Update: Influenza activity - United States, 2009-10 season. *MMWR.Morbidity. Mortality Week. Rep., 59*(29), 901–908.

Centers for Disease Control and Prevention (CDC). (2011). *Final estimates for 2009–10 seasonal influenza and influenza A (H1N1) 2009 monovalent vaccination coverage − United States, august 2009 through may, 2010*. Retrieved 11/28, 2022, from https://www.cdc.gov/flu/fluvaxview/coverage_0910estimates.htm.

Chao, D. L., Halloran, M. E., & Longini, I. M., Jr. (2010). School opening dates predict pandemic influenza A(H1N1) outbreaks in the United States. *The Journal of Infectious Diseases, 202*(6), 877–880.

Chowell, G., Echevarria-Zuno, S., Viboud, C., Simonsen, L., Tamerius, J., Miller, M. A., et al. (2011). Characterizing the epidemiology of the 2009 influenza A/H1N1 pandemic in Mexico. *PLoS Medicine, 8*(5), Article e1000436.

Dawood, F. S., Iuliano, A. D., Reed, C., Meltzer, M. I., Shay, D. K., Cheng, P., et al. (2012). Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: A modelling study. *The Lancet Infectious Diseases, 12*(9), 687–695.

Diekmann, O., Heesterbeek, J. A., & Metz, J. A. (1990). On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology, 28*(4), 365–382.

Drummond Aj, B. R. (2015). *Bayesian evolutionary analysis with BEAST*. Cambridge: Cambridge University Press, ISBN 9781107019652.

Drummond, A. J., & Rambaut, A. (2007). Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology, 7*, 214-2148-7-214.

Duchene, S., Bouckaert, R., Duchene, D. A., Stadler, T., & Drummond, A. J. (2019). Phylodynamic model adequacy using posterior predictive simulations. *Systematic Biology, 68*(2), 358–364.

Dudas, G., Carvalho, L. M., Bedford, T., Tatem, A. J., Baele, G., Faria, N. R., et al. (2017). Virus genomes reveal factors that spread and sustained the ebola epidemic. *Nature, 544*(7650), 309–315.

Duvvuri, V. R., Granados, A., Rosenfeld, P., Bahl, J., Eshaghi, A., & Gubbay, J. B. (2015). Genetic diversity and evolutionary insights of respiratory syncytial virus A ON1 genotype: Global and local transmission dynamics. *Scientific Reports, 5*, Article 14268.

Earn, D. J., He, D., Loeb, M. B., Fonseca, K., Lee, B. E., & Dushoff, J. (2012). Effects of school closure on incidence of pandemic influenza in Alberta, Canada. *Annals of Internal Medicine, 156*(3), 173—181.

Edgar, R. C. (2004). Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792—1797.

Featherstone, L. A., Di Giallonardo, F., Holmes, E. C., Vaughan, T. G., & Duchêne, S. (2021). Infectious disease phylodynamics with occurrence data. *Methods in Ecology and Evolution, 12*(8), 1498—1507.

Fine, P., Eames, K., & Heymann, D. L. (2011). Herd immunity". A rough guide. *Clinical Infectious Diseases : Off. Publ. Infect. Dis. Soc. Am., 52*(7), 911—916.

Fourment, M., Claywell, B. C., Dinh, V., McCoy, C., Matsen Iv, F. A., & Darling, A. E. (2018). Effective online bayesian phylogenetics via sequential Monte Carlo with guided proposals. *Systematic Biology, 67*(3), 490—502.

Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., et al. (2009). Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science (New York, N.Y.), 324*(5934), 1557—1561.

Gog, J. R., Ballesteros, S., Viboud, C., Simonsen, L., Bjornstad, O. N., Shaman, J., Chao, D. L., Khan, F., & Grenfell, B. T. (2014). Spatial transmission of 2009 pandemic influenza in the US. *PLoS Computational Biology, 10*(6), Article e1003635.

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., et al. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science (New York, N.Y.), 303*(5656), 327—332.

Hedge, J., Lycett, S. J., & Rambaut, A. (2013). Real-time characterization of the molecular epidemiology of an influenza pandemic. *Biology Letters, 9*(5), Article 20130331.

He, D., Dushoff, J., Eftimie, R., & Earn, D. J. D. (2013). Patterns of spread of influenza A in Canada. *Proceed.Biol. Sci., 280*(1770), Article 20131174.

He, D., Lui, R., Wang, L., Tse, C. K., Yang, L., & Stone, L. (2015). Global spatio-temporal patterns of influenza in the post-pandemic era. *Scientific Reports, 5*, Article 11013.

Ho, S. Y. W., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., et al. (2011). Time-dependent rates of molecular evolution. *Molecular Ecology, 20*(15), 3087—3101.

Ho, S. Y. W., Phillips, M. J., Cooper, A., & Drummond, A. J. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution, 22*(7), 1561—1568.

Hsieh, Y., Fisman, D. N., & Wu, J. (2010). On epidemic modeling in real time: An application to the 2009 novel A (H1N1) influenza outbreak in Canada. *BMC Research Notes, 3*, 283-0500-3-283.

Iuliano, A. D., Roguski, K. M., Chang, H. H., Muscatello, D. J., Palekar, R., Tempia, S., et al. (2018). Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *Lancet (London, England), 391*(10127), 1285—1300.

Kass Re, R. A. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773—795.

Leekha, S., Zitterkopf, N. L., Espy, M. J., Smith, T. F., Thompson, R. L., & Sampathkumar, P. (2007). Duration of influenza A virus shedding in hospitalized patients and implications for infection control. *Infection Control and Hospital Epidemiology, 28*(9), 1071—1076.

Lessler, J., Reich, N. G., Cummings, D. A. T., New York City Department of Health and Mental Hygiene Swine Influenza Investigation Team, Nair, H. P., Jordan, H. T., et al. (2009). Outbreak of 2009 pandemic influenza A (H1N1) at a New York city school. *New England Journal of Medicine, 361*(27), 2628—2636.

McBryde, E., Bergeri, I., van Gemert, C., Rotty, J., Headley, E., Simpson, K., Lester, R., Hellard, M., & Fielding, J. (2009). Early transmission characteristics of influenza A(H1N1)v in Australia: Victorian state, 16 May - 3 June 2009. *Euro Surveillance, 14*(42), Article 19363.

Moller, S., du Plessis, L., & Stadler, T. (2018). Impact of the tree prior on estimating clock rates during epidemic outbreaks. *Proceedings of the National Academy of Sciences of the United States of America, 115*(16), 4200—4205.

Morens, D. M., & Taubenberger, J. K. (2011). Pandemic influenza: Certain uncertainties. *Reviews in Medical Virology, 21*(5), 262—284.

Mostaco-Guidolin, L. C., Bowman, C. S., Greer, A. L., Fisman, D. N., & Moghadas, S. M. (2012). Transmissibility of the 2009 H1N1 pandemic in remote and isolated canadian communities: A modelling study. *BMJ Open, 2*(5), Article e001614.

Mostaco-Guidolin, L. C., Greer, A., Sander, B., Wu, J., & Moghadas, S. M. (2011). Variability in transmissibility of the 2009 H1N1 pandemic in canadian communities. *BMC Research Notes, 4*, 537-530500-4-537.

Nelson, M. I., Tan, Y., Ghedin, E., Wentworth, D. E., St George, K., Edelman, L., et al. (2011). Phylogeography of the spring and fall waves of the H1N1/09 pandemic influenza virus in the United States. *Journal of Virology, 85*(2), 828—834.

Park, S. W., Champredon, D., Weitz, J. S., & Dushoff, J. (2019). A practical generation-interval-based approach to inferring the strength of epidemics from their speed. *Epidemics, 27*, 12—18.

du Plessis, L., & Stadler, T. (2015). Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends in Microbiology, 23*(7), 383—386.

Pourbohloul, B., Ahued, A., Davoudi, B., Meza, R., Meyers, L. A., Skowronski, D. M., et al. (2009). Initial human transmission dynamics of the pandemic (H1N1) 2009 virus in north America. *Influenza. Respiratory. Virus., 3*(5), 215—222.

Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C., & Harvey, P. H. (2001). The epidemic behavior of the hepatitis C virus. *Science (New York, N.Y.), 292*(5525), 2323—2325.

Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-gen). *Virus Evol., 2*(1), Article vew007.

Rosenthal, M., Anderson, K., Tengelsen, L., Carter, K., Hahn, C., & Ball, C. (2017). Evaluation of sampling recommendations from the influenza virologic surveillance right size roadmap for Idaho. *JMIR Publ. Health Surveill., 3*(3), Article e57.

Shaman, J., Goldstein, E., & Lipsitch, M. (2011). Absolute humidity and pandemic versus epidemic influenza. *American Journal of Epidemiology, 173*(2), 127—135.

Sharomi, O., Podder, C. N., Gumel, A. B., Mahmud, S. M., & Rubinstein, E. (2011). Modelling the transmission dynamics and control of the novel 2009 swine influenza (H1N1) pandemic. *Bulletin of Mathematical Biology, 73*(3), 515—548.

Shu, Y., & McCauley, J. (2017). Gisaid: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = Eur. Commun. Dis. Bull., 22*(13), Article 30494. https://doi.org/10.2807/1560

Stadler, T., Kouyos, R., von Wyl, V., Yerly, S., Boni, J., Burgisser, P., et al. (2012). Estimating the basic reproductive number from viral sequence data. *Molecular Biology and Evolution, 29*(1), 347—357.

Stadler, T., Kuhnert, D., Bonhoeffer, S., & Drummond, A. J. (2013). Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences of the United States of America, 110*(1), 228—233.

Stadler, T., Kuhnert, D., Rasmussen, D. A., & du Plessis, L. (2014). Insights into the early epidemic spread of ebola in Sierra Leone provided by viral sequence data. *PLoS Currents, 6*, 10, 1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312—1313.

Su, Y. C. F., Bahl, J., Joseph, U., Butt, K. M., Peck, H. A., Koay, E. S. C., et al. (2015). Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. *Nature Communications, 6*, 7952.

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol., 4*(1), Article vey016.

Sugimoto, J. D., Borse, N. N., Ta, M. L., Stockman, L. J., Fischer, G. E., Yang, Y., et al. (2011). The effect of age on transmission of 2009 pandemic influenza A (H1N1) in a camp and associated households. *Epidemiology, 22*(2), 180—187.

Truelove, S. A., Chitnis, A. S., Heffernan, R. T., Karon, A. E., Haupt, T. E., & Davis, J. P. (2011). Comparison of patients hospitalized with pandemic 2009 influenza A (H1N1) virus infection during the first two pandemic waves in Wisconsin. *The Journal of Infectious Diseases, 203*(6), 828—837.

Tuite, A. R., Greer, A. L., Whelan, M., Winter, A., Lee, B., Yan, P., et al. (2010). Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. *Canadian Medical Association Journal : Canadian Med. Assoc. J = Journal De L'Association Medicale Canadienne, 182*(2), 131—136.

Volz, E. M., & Frost, S. D. (2014). Sampling through time and phylodynamic inference with coalescent and birth-death models. *Journal of The Royal Society Interface, 11*(101), Article 20140945.

Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J., & Frost, S. D. W. (2009). Phylodynamics of infectious disease epidemics. *Genetics, 183*(4), 1421—1430.

Volz, E. M., Romero-Severson, E., & Leitner, T. (2017). Phylodynamic inference across epidemic scales. *Molecular Biology and Evolution, 34*(5), 1276—1288.

Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences, 274*(1609), 599—604.

White, L. F., & Pagano, M. (2010). Reporting errors in infectious disease outbreaks, with an application to pandemic influenza A/H1N1. *Epidemiologic Perspectives & Innovations : EP+I, 7*, 12-5573-7-12.

White, L. F., Wallinga, J., Finelli, L., Reed, C., Riley, S., Lipsitch, M., et al. (2009). Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. *Influenza. Respiratory. Virus., 3*(6), 267—276.

World Health Organization. (2021). *Genomic sequencing of SARS-CoV-2: A guide to implementation for maximum impact on public health (who.int).* Retrieved 11/28, 2022, from https://www.who.int/publications/i/item/9789240018440.

Yang, Y., Sugimoto, J. D., Halloran, M. E., Basta, N. E., Chao, D. L., Matrajt, L., et al. (2009). The transmissibility and control of pandemic influenza A (H1N1) virus. *Science (New York, N.Y.), 326*(5953), 729—733.

Zepeda-Lopez, H. M., Perea-Araujo, L., Miliar-Garcia, A., Dominguez-Lopez, A., Xoconostle-Cazarez, B., Lara-Padilla, E., et al. (2010). Inside the outbreak of the 2009 influenza A (H1N1)v virus in Mexico. *PLoS One, 5*(10), Article e13256.