Smooth Skyride through a Rough Skyline: Bayesian Coalescent-Based Inference of Population Dynamics

Vladimir N. Minin, Erik W. Bloomquist,† and Marc A. Suchard†‡*

*Department of Statistics, University of Washington; †Department of Biostatistics, School of Public Health, University of California, Los Angeles; and ‡Departments of Biomathematics and Human Genetics, David Geffen School of Medicine, University of California, Los Angeles

Kingman's coalescent process opens the door for estimation of population genetics model parameters from molecular sequences. One paramount parameter of interest is the effective population size. Temporal variation of this quantity characterizes the demographic history of a population. Because researchers are rarely able to choose a priori a deterministic model describing effective population size dynamics for data at hand, nonparametric curve-fitting methods based on multiple change-point (MCP) models have been developed. We propose an alternative to change-point modeling that exploits Gaussian Markov random fields to achieve temporal smoothing of the effective population size in a Bayesian framework. The main advantage of our approach is that, in contrast to MCP models, the explicit temporal smoothing does not require strong prior decisions. To approximate the posterior distribution of the population dynamics, we use efficient, fast mixing Markov chain Monte Carlo algorithms designed for highly structured Gaussian models. In a simulation study, we demonstrate that the proposed temporal smoothing method, named Bayesian skyride, successfully recovers "true" population size trajectories in all simulation scenarios and competes well with the MCP approaches without evoking strong prior assumptions. We apply our Bayesian skyride method to 2 real data sets. We analyze sequences of hepatitis C virus contemporaneously sampled in Egypt, reproducing all key known aspects of the viral population dynamics. Next, we estimate the demographic histories of human influenza A hemagglutinin sequences, serially sampled throughout 3 flu seasons.

Introduction

Accurate estimation of population size dynamics has important implications for public health and conservation biology (Pybus et al. 2003; Shapiro et al. 2004; Biek et al. 2006). In this paper, we propose a statistically novel model to infer population size dynamics from molecular sequences. We build our modeling framework upon Kingman's coalescent process, a powerful tool in the population genetics arsenal for studying probabilistic properties of genealogies relating individuals randomly sampled from a population of interest (Kingman 1982). Because genealogical shapes leave their imprints in the genomes of sampled individuals, the coalescent allows for the inference of population genetics parameters, including population size dynamics, directly from the observed genomic sequences (Hein et al. 2005).

Many coalescent-based estimation algorithms rely on simple parametric forms to characterize the evolution of the population size dynamics over time. Advantageously, these deterministic functions contain a relatively small number of parameters to be estimated (Kuhner et al. 1998; Drummond et al. 2002). However, justifying strong parametric assumptions can be difficult and may require laborious and computationally expensive testing of many candidate functional forms to find an appropriate description of the population size trajectory. An extreme alternative to parametric population size models is the classical skyline plot estimation proposed by Pybus et al. (2000). This estimation procedure relies on a piecewise constant population dynamics model. Because the number of free parameters in this model is equal to the number of independently distributed observations, the classical skyline plot approach results in very noisy estimates.

Key words: coalescent, smoothing, effective population size, Gaussian Markov random fields.

E-mail: vminin@u.washington.edu.

Mol. Biol. Evol. 25(7):1459–1471. 2008
doi:10.1093/molbev/msn090
Advance Access publication April 11, 2008

To arrive at a middle ground between overly stringent parametric and noisy classical skyline plot approaches, 3 extensions to the classical skyline plot estimation have been recently proposed. Strimmer and Pybus (2001) develop generalized classical skyline plot estimation. These authors employ a model selection approach, based on the Akaike Information Criterion correction (AIC_c), to reduce the number of free parameters in the classical skyline plot. Drummond et al. (2005) and Opgen-Rhein et al. (2005) use multiple change-point (MCP) models to estimate the population size dynamics in a Bayesian framework. These methods approximate the effective population size trajectory with a step function, defined by estimable change-point locations and step heights. One of the main advantages of MCP models is the ease of incorporating them into a joint Bayesian estimation of genealogies and population genetics parameters as demonstrated by Drummond et al. (2005). Both proposed MCP models share the same weakness as they require fairly strong prior decisions. Drummond et al. (2005) a priori fix the total number of change points, a critical parameter in their model that controls the smoothness of the population size trajectory. Opgen-Rhein et al. (2005) bypass the problem of fixing the number of change points through reversible jump Markov chain Monte Carlo (MCMC) sampling (Green 1995). However, these authors use an informative and very influential prior for the number of change points in their model. Therefore, in both MCP approaches choosing an appropriate level of smoothness of population size dynamics remains problematic.

We propose to smooth population size trajectories explicitly. We choose piecewise constant demographic model of Pybus et al. (2000) as our point of departure. Our goal is to construct a smooth skyride through a rough classical skyline profile. Our construction is accomplished by imposing a Gaussian Markov random field (GMRF) smoothing prior on the parameters of the piecewise constant population size trajectory. We make our smoothing prior "time aware" to penalize effective population size changes between "small"

consecutive (Ghedin et. al 2005) intercoalescent intervals more than changes between intervals of larger size. To achieve this desirable behavior of our smoothing prior, we equip each consecutive pair of intercoalescent intervals with an appropriate smoothing weight. We show through a simulation study that the time-aware prior is very effective in capturing important characteristics of "true" population size trajectories and is superior to a uniform, time-ignorant GMRF prior. The extension of Kingman's coalescent to serially sampled (heterochronous) data by Rodrigo and Felsenstein (1999) opens the door for coalescent-based inference for measurably evolving populations (Drummond et al. 2003). We show that the GMRF smoothing can be easily incorporated into analyses of both isochronous (contemporaneously sampled) and heterochronous data.

Through simulation, we compare performance of the Bayesian skyride with Opgen-Rhein's MCP (ORMCP) model (Opgen-Rhein et al. 2005) and a Bayesian skyline plot, a MCP model implemented in the software package BEAST (Drummond et al. 2005). In 3 simulation scenarios that we consider, we find that the Bayesian skyride performs as well or better than both MCP approaches. Although the small number of our simulations prevent us from a detailed comparison of the methods, we nevertheless can conclude that the Bayesian skyride is a competitive alternative to the MCP models and requires substantially weaker prior assumptions. We demonstrate the utility of the proposed method by applying it to 2 real data sets. We analyze isochronous sequences of hepatitis C virus (HCV) and demonstrate that the Bayesian skyride is able to recover all previously inferred characteristics of the Egyptian HCV population dynamics. We proceed with an investigation of intraseason population dynamics of human influenza virus. Our analysis of heterochronous influenza data from 3 seasons demonstrates that estimation of influenza population dynamics holds promise for predicting peak infection time within a flu season.

Methods

Coalescent Background

We start with a random population sample of n sequences. Coalescent theory provides a stochastic process that produces genealogies relating these sampled sequences. The process starts at sampling time $t = 0$ and proceeds backward in time as t increases, coalescing n individuals one pair at a time until the time to the most recent common ancestor (TMRCA) of the sample is reached (Kingman 1982). In this work, we ignore extensions of the coalescent that allow for modeling of the effects of selection (Krone and Neuhauser 1997), population structure (Notohara 1990), and recombination (Hudson 1983). Instead, we focus on the coalescent with variable effective population size.

The effective population size is an abstract quantity that brings populations with different reproductive models to a "common denominator," namely the Wright–Fisher model (Kingman 1982). One can obtain the census population size by appropriate scaling of the effective population size. Because the dynamics of the effective population size plays a very important role in shaping coalescent-based genealogies, it should be possible to solve the inverse problem and recover the effective population size trajectories from known genealogies.

We assume for the moment that $\mathbf{g}$ is a known genealogy relating the n sampled sequences. Suppose that function $N_e(t)$ describes the time evolution of the effective population size as we move into the past. Given $N_e(t)$, we need to compute the probability of observing $\mathbf{g}$ under the coalescent with variable population size. To achieve this, it suffices to construct a probability density function over the intercoalescent times $\mathbf{u}=(u_2,\ldots,u_n)$ induced by $\mathbf{g}$, where $u_k = t_k - t_{k-1}$, t_k is the time of the $(n-k)$th coalescent event for $k=2,\ldots,n$ and $t_n = 0$ is the time at which sequences are sampled (Felsenstein 1992; Pybus et al. 2000). Griffiths and Tavaré (1994) show that the joint density of intercoalescent times can be obtained by multiplying conditional densities

$$\Pr\left(u_k \middle| t_k\right) = \frac{k(k-1)}{2N_e(u_k + t_k)} \exp\left[- \int_{t_k}^{u_k + t_k} \frac{k(k-1)}{2N_e(t)} \mathrm{d}t \right], \quad (1)$$

where time is measured in units of generations. Given a simple parametric form for $N_e(t)$, it is straightforward to estimate the parameters characterizing $N_e(t)$ in a likelihood-based framework, possibly integrating over genealogies using MCMC sampling (Kuhner et al. 1998; Drummond et al. 2002). However, lack of prior knowledge about the appropriate parametric form for $N_e(t)$ stimulates the current development of nonparametric and semiparametric methods of estimating $N_e(t)$ (Pybus et al. 2000; Drummond et al. 2005; Opgen-Rhein et al. 2005).

Piecewise Demographic Model for Isochronous Data

Let us assume that all sampled sequences were collected effectively at the same time, meaning that differences between sampling times are negligible compared with the TMRCA of the sample. We start with a critical assumption that $N_e(t)$ can change its value only at coalescent times, $N_e(t) = \theta_k$ for some $\theta_k > 0$ and $t_k < t \leq t_{k-1}$, $k=2,\ldots,n$. Informally, we can plug the piecewise constant $N_e(t)$ into equation (1) and arrive at

$$\Pr\left(u_k \middle| \theta_k\right) = \frac{k(k-1)}{2\theta_k} \exp\left[- \frac{k(k-1)u_k}{2\theta_k} \right]. \quad (2)$$

See Pybus et al. (2000) and Strimmer and Pybus (2001) for more details on the likelihood of the piecewise demographic model. It is important to notice that because the right-hand side of equation (2) does not depend on t_k, intercoalescent intervals $\mathbf{u}$ are independent of each other under the piecewise constant demographic model. Therefore, estimating interval-specific population sizes θ_k is equivalent to estimating the rate of an exponential distribution after drawing only a single realization from this distribution. Maximizing likelihood function (2) with respect to θ_k yields

$$\hat{\theta}_k = \frac{k(k-1)u_k}{2}. \quad (3)$$

Such estimators of interval-specific population sizes are called classical skyline plots (Pybus et al. 2000). Strimmer and Pybus (2001) quickly recognize that estimators (3) have substantial variance resulting from overfitting and introduce a generalized skyline plot. In their new method, the authors restrict the number of different population sizes across intervals. Intervals that do not exceed a predefined threshold ϵ borrow their population sizes from the neighboring intervals. The authors choose ϵ by maximizing a second-order extension of the AIC_c. The Bayesian skyline plot of Drummond et al. (2005) groups intervals using a MCP model providing a Bayesian extension of the generalized skyline plot.

Throughout this section, we have assumed that time is measured in units of generations. When analyzing isochronous data, branches of genealogies are often estimated in units of average number of substitutions per site. Even if we are able to estimate branches in units of clock time, the generation time may be unknown. However, this inability to identify time in units of generations does not limit estimation of the dynamics of demographic histories. If estimated intercoalescent intervals are not measured in units of generations and are rescaled as $u_k = cu_k$, then plugging u_k^*s into the likelihood function (2), we can recover the rescaled effective population size trajectory $N_e^*(t) = cN_e(t)$.

Piecewise Demographic Model for Heterochronous Data

We now turn to the piecewise demographic model for sequences sampled at sufficiently different time points. As before, we assume that genealogy $\mathbf{g}$ relating the sampled sequences is known and fixed. Moreover, branch lengths of $\mathbf{g}$ satisfy constraints imposed by sampling times $\mathbf{s}$. The sampling times divide each intercoalescent interval k into subintervals $\mathbf{w}_k = (w_{k0}, \ldots, w_{kj_k})$, where $j_k \in \{0, \ldots, n-1\}$ is the number of "distinct" sampling times occurring during interval k, $\sum_{j=0}^{j_k} w_{kj} = u_k$, and the interval that ends with the $(n-k)$th coalescent event is always indexed by $k0$. To each subinterval kj, we attach the number of lineages n_{kj} present in the genealogy at the beginning of this interval. See figure 1 for an example genealogy with labeled intercoalescent intervals, subintervals, and numbers of lineages. For heterochronous data, we still assume that $N_e(t) = \theta_k$ for some $\theta_k > 0$ and $t_k < t \leq t_{k-1}$ for $k = 2, \ldots, n$.

Rodrigo and Felsenstein (1999) extend the coalescent likelihood to incorporate heterochronous data. The authors distinguish between coalescent and sampling events. In our notation, subintervals labeled as $k0$ end with a coalescent event. Each such subinterval contributes an exponential density to the coalescent likelihood, where the exponential rate depends on the number of lineages present and the effective population size in the interval. Subintervals ending with a sampling event contribute to the likelihood the probability of no coalescence or equivalently the probability that an exponentially distributed coalescence time is greater than the subinterval length. Because in our notation only subintervals with indices $k0$ end with a coalescence event, the likelihood of observing subintervals $\mathbf{w}_k$ comprising intercoalescent interval k is

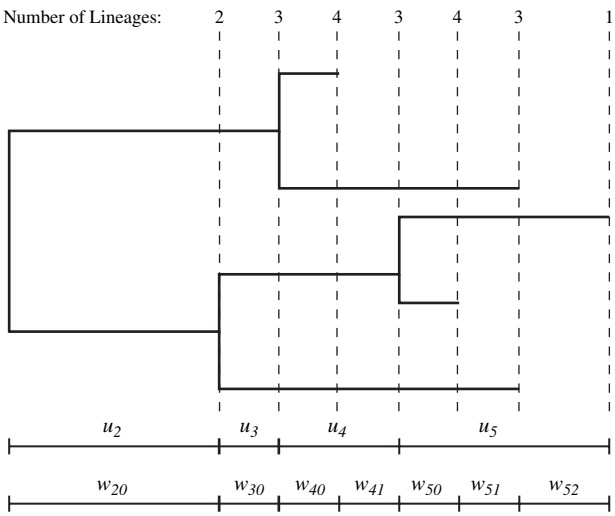

FIG. 1.—Example of a genealogy with intercoalescent interval notation. Times of coalescence and sampling events are depicted as vertical dashed lines with numbers of lineages present at these times shown above the lines. Below the genealogy, we mark the boundaries of intercoalescent intervals together with their lengths $(u_2, \ldots, u_5)$. We show how sampling events interrupt the intercoalescent intervals and produce subintervals with lengths $(w_{20}, \ldots, w_{52})$ at the bottom of the figure.

$$\Pr(\mathbf{w}_k|\theta_k) = \frac{n_{k0}(n_{k0}-1)}{2\theta_k} \exp\left[-\frac{\sum_{j=0}^{j_k} n_{kj}\left(n_{kj}-1\right)w_{kj}}{2\theta_k}\right]. \quad (4)$$

The maximum likelihood estimate of the piecewise constant effective population size is

$$\hat{\theta}_k = \frac{\sum_{j=0}^{j_k} n_{kj}\left(n_{kj}-1\right)w_{kj}}{2}. \quad (5)$$

By analogy with the likelihood for isochronous data, Pybus and Rambaut (2002) call estimates (5) a classical skyline plot. The authors group intercoalescent intervals using the same AIC-based algorithm as in the isochronous case to arrive at a generalized skyline plot.

Working with heterochronous data has 2 major advantages. First, such data allow one to estimate branch lengths of genealogies in units of time and simultaneously estimate the mutation rate (Drummond et al. 2002, 2003). Secondly, the sampling times provide additional information for the effective population size estimation. Although the sampling subintervals do not allow one to observe more coalescence events, these subintervals serve as censored time-to-event data. Unfortunately, because there are at most n distinct sampling times, the improvement in the information content may not be dramatic. Therefore, estimation of intercoalescent interval-specific effective population sizes remains problematic without further modifications.

Temporally Smoothed Piecewise Demographic Model

Because the heterochronous likelihood (4) reduces to the isochronous likelihood (2) when $j_k = 0$ for all

$k=2,\ldots,n$, from now on we assume that that observed data come in the form of intercoalescent subintervals $\mathbf{w}=(\mathbf{w}_2,\ldots,\mathbf{w}_n)$. We first transform the intercoalescent interval-specific effective population sizes onto the whole real line via

$$\gamma_k = \log\theta_k, k = 2,\ldots,n, \tag{6}$$

and following equation (4), consider the likelihood

$$\Pr(\mathbf{w}|\gamma) = \prod_{k=2}^{n} \Pr(\mathbf{w}_k|\gamma_k), \tag{7}$$

where $\gamma=(\gamma_2,\ldots,\gamma_n)$.

We invoke a common assumption stating that the effective population size changes continuously through time. To infuse this minimal prior knowledge into our Bayesian model, we devise a GMRF prior distribution for the vector γ. This prior penalizes the differences between components of γ as

$$\Pr(\gamma|\tau) \propto \tau^{(n-2)/2} \exp\left[-\frac{\tau}{2}\sum_{k=2}^{n-1} \frac{(\gamma_{k+1}-\gamma_k)^2}{\delta_k} \right], \tag{8}$$

where τ is the overall precision of the GMRF and δ_k is the distance between sites $k+1$ and k on the 1-dimensional lattice $\{2,\ldots,n\}$. A uniform GMRF smoothing with equal distances $\delta_2=\cdots=\delta_{n-1}=1$ is often assumed for temporal and spatial smoothing. However, in the piecewise demographic model, the penalty for the dissimilarity between adjacent intercoalescent, interval-specific, effective population sizes should depend on the interval sizes. Therefore, we construct a time-aware GMRF prior based on midpoint distances between intercoalescent intervals,

$$\delta_k = \frac{u_k + u_{k+1}}{2}. \tag{9}$$

Midpoint distances have been successfully used before in molecular evolution by Thorne et al. (1998) to construct an autocorrelated prior for evolutionary rates on genealogy branches. The midpoint distances' 2D analogs, the distances between centroids of 2-dimensional areas, are widely used in spatial statistics (Elliott et al. 2000).

We conclude our prior specification by assigning a gamma prior to the GMRF precision parameter τ,

$$\Pr(\tau) \propto \tau^{a-1} e^{-\beta\tau}. \tag{10}$$

Bernardinelli et al. (1995) highlight the importance of priors for the GMRF precision parameter and advise against diffuse priors. However, a researcher normally has no knowledge about the smoothness of the effective population size trajectory a priori. Therefore, in our examples, we choose $\alpha = \beta = 0.001$ making prior (10) relatively uninformative, with expectation 1 and variance 1,000.

We estimate γ and τ by MCMC sampling from the posterior distribution of these parameters. Moreover, we implement our Bayesian skyride method in the software package BEAST to estimate effective population size trajectories and genealogies simultaneously. The details of our MCMC algorithm can be found in Appendix A. Appendix B contains instructions on using the Bayesian skyride in BEAST.

Testing Significance of the Effective Population Size Changes

It is common to assess significance of effective population size changes by visually inspecting quantiles of the marginal posterior distributions of effective population sizes. Such an informal approach can be deceptive. In using it, one attempts to draw conclusions about the difference of 2 possibly highly correlated random variables based on their marginal distributions. Moreover, the correlation between values of the effective population size cannot be ignored if we indeed believe that $N_e(t)$ changes continuously through time. As an alternative, we propose to use Bayes factors to formally test the significance of effective population size changes when $\mathbf{g}$ is assumed to be known.

We start with a 1-sided hypothesis test. Suppose that we a priori fix 2 intercoalescent intervals $i < j$ that correspond to some historical events of interest. Then, we may be interested in testing hypotheses $H_0{:}\gamma_i < \gamma_j$ versus $H_1{:}\gamma_i \geq \gamma_j$. The Bayes factor

$$B_{01} = \frac{\Pr(\mathbf{w}|H_0)}{\Pr(\mathbf{w}|H_1)} = \frac{\Pr(H_0|\mathbf{w})/\Pr(H_0)}{\Pr(H_1|\mathbf{w})/\Pr(H_1)}. \tag{11}$$

allows one to quantify the evidence in favor of the null hypothesis H_0 and against the alternative H_1 (Kass and Raftery 1995). Through the ergodic theorem, we approximate the posterior probability $\Pr(H_0|\mathbf{w})$ with a fraction of MCMC samples satisfying $\gamma_i < \gamma_j$. The GMRF prior (8) implies that $\gamma_i - \gamma_j \,\big|\, \tau \sim N\left(0, \frac{\sum_{k=i}^{j-1}\delta_k}{\tau}\right)$. Therefore, $\Pr(H_0) = \Pr(H_1) = 0.5$ regardless of the prior choice for τ as long as the prior is proper.

Clearly, this 1-sided test is impractical if the direction of the effective population size change is irrelevant or if one is interested in testing effective population size differences among multiple intercoalescent intervals. Therefore, we also consider the hypothesis $H_0 : \gamma_{i_1}=\cdots=\gamma_{i_j}$, where $\{i_1,\ldots,i_j\}$ is a subset of $\{2,\ldots,n\}$. The alternative hypothesis H_1 states that not all $\gamma_{i_1},\ldots,\gamma_{i_j}$ are equal. To estimate the Bayes factor in equation (11), we calculate 2 marginal likelihoods, $\Pr(\mathbf{w}|H_0)$ and $\Pr(\mathbf{w}|H_1)$, from the MCMC output using the harmonic mean estimator (Newton and Raftery 1994). This Bayes factor estimation procedure requires sampling from the posterior distribution of (τ,γ) under $j-1$ linear constraints on γ imposed by hypothesis H_0. Fortunately, sampling from GMRFs under linear constraints adds very little computational cost to the unconstrained sampling algorithm (Rue and Held 2005). Because the harmonic mean estimator of the marginal likelihood is not the most efficient, one could also use the method of Chib and Jeliazkov (2001). Alternatively, it may be possible to adapt the generalized Savage–Dickey ratio of Verdinelli and Wasserman (1995) to test the sharp hypothesis H_0.

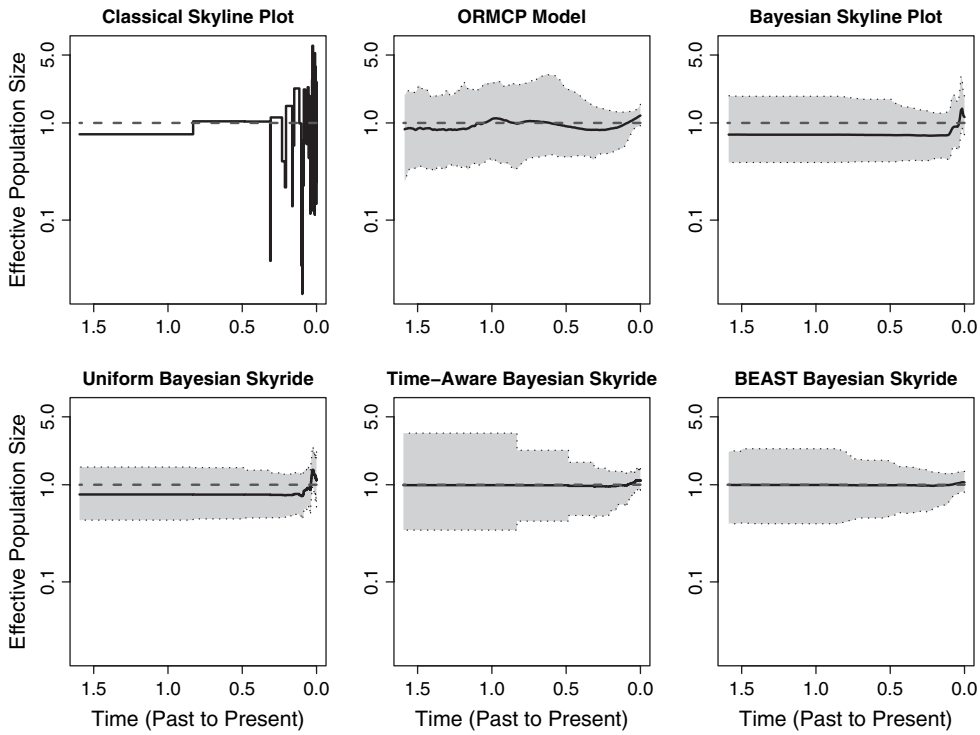

FIG. 2.—Constant population size simulation. We present a classical skyline plot (solid black line) in the top left part of the figure. The other 5 plots show posterior median (solid black line) and 95% BCIs (gray shading) of the effective population size under the ORMCP model, Bayesian skyline plot, time-aware and uniform Bayesian skyrides with a fixed genealogy, and BEAST Bayesian skyride method. In all 6 plots, the dashed lines represent the true population size trajectory that was used for simulations. Here and in all subsequent plots of effective population sizes, we use the log transformation of the population size axis.

Results

Simulated Genealogies

We examine the ability of the Bayesian skyride to recover effective population size dynamics in a simulation study. Because the number of individuals in a population at time $t = 0$ affects only the time measurement units, we always start with $N(0) = 1.0$ in our simulations. First, we simulate a genealogy assuming the constant population size. Next, we use the molecular sequence evolution simulator of Rambaut and Grassly (1997) to generate sequence data on the tips of the simulated genealogy. We assume a molecular clock and use the HKY model (Hasegawa et al. 1985) with a transition/transversion ratio fixed to 2.5. In this and subsequent simulations, we choose mutation rates such that the root heights of simulated genealogies, measured in expected number of substitutions per site, vary between 0.15 and 0.4. Such mutation rates produce sequences with realistic levels of divergence seen in genomic data of rapidly evolving pathogens. We choose this parameter regime because both of our real data examples concern viral evolution. We summarize our posterior inference results in figure 2. In the top left plot of this figure, we show the classical skyline plot based on the simulated genealogy. The rest of the plots in figure 2 demonstrate results of estimating the effective population size trajectory using the ORMCP and Bayesian skyline plot models, fixed-tree time-aware and uniform Bayesian skyrides, and BEAST Bayesian skyride with the time-aware weighting scheme. The posterior medians (solid black lines) obtained using the fixed-tree time-aware and the BEAST Bayesian skyrides nearly perfectly match the true effective population size (dashed line in all 6 plots). The 95% Bayesian credible intervals (BCIs), shown as gray shaded areas, become wider near the TMRCA. This behavior is natural because under the constant population size model, the relatively long time to coalescence of the last 2 lineages leaves very little information for effective population size estimation near the root of the genealogy. Posterior medians of $N_e(t)$, obtained under the ORMCP and Bayesian skyline plot models, as well as under the uniform Bayesian skyride, underestimate the effective population size. It is clear that small intercoalescent intervals, relative to the TMRCA, near the root of the genealogy mislead these methods. We investigate the effect of this bias on the frequentist coverage properties of the methods' 95% BCIs via a simulation study. After simulating 1,000 genealogies under the constant population size coalescent model, we analyze these genealogies using the ORMCP model, fixed-tree uniform, and time-aware Bayesian skyrides. We then calculate the percentage of time each method's BCIs fully cover the true population size trajectory. The ORMCP model, uniform, and time-aware Bayesian skyride coverages amount to 61.8%, 89.9%, and 94.1%, respectively. These numbers clearly illustrate the frequentist coverage advantages of the time-aware Bayesian skyride.

Next, we simulate a genealogy assuming that $N_e(t) = e^{-500t}$. Because we proceed in time from present to past in our simulations, the negative growth constant implies an

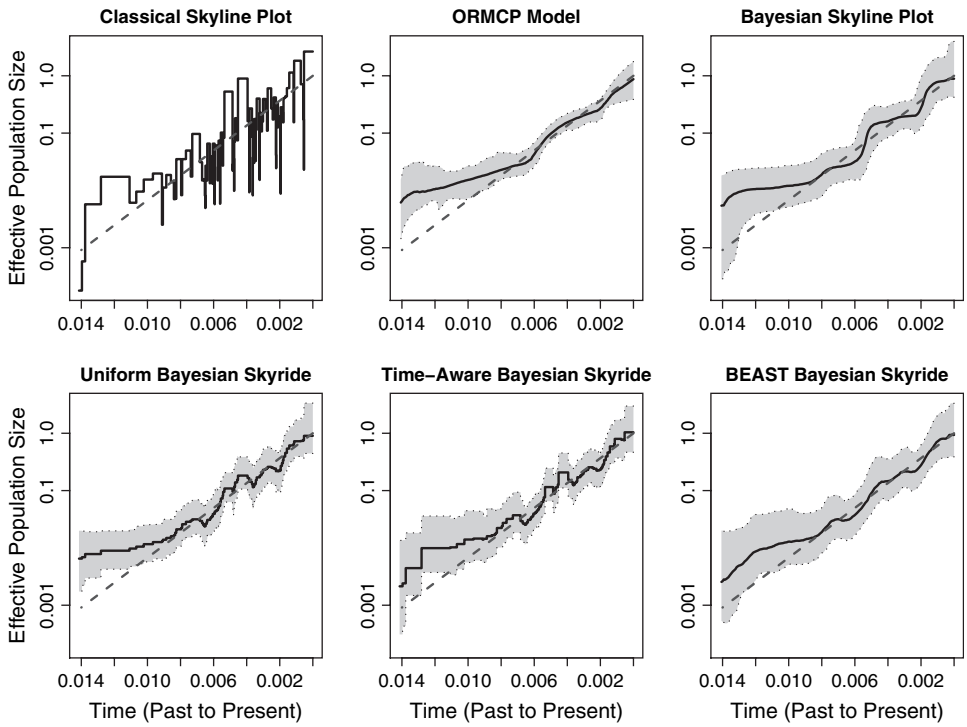

FIG. 3.—Exponential growth simulation. See figure 2 for the legend explanation.

exponentially growing population. This exponential growth is shown on the log scale as a straight dashed line in all plots of figure 3. All models perform reasonably well in this simulation. Here, as in the constant population size case, intercoalescent intervals near the root of the genealogy cause estimation problems. One "unusually" large intercoalescent interval in this region provokes the uniform Bayesian skyride, the ORMCP model, and the Bayesian skyline plot model into overestimating the effective population size in the proximity of this interval. However, the fixed-tree and BEAST time-aware Bayesian skyride methods are less prone to such overestimation. We believe that this desirable behavior results from our weighting scheme that prohibits rapid changes of the effective population size during short time periods.

Finally, we generate a genealogy from a coalescent process assuming that a population experiences a bottleneck during the population's evolutionary history. More specifically, we set

$$N_e(t) = \begin{cases} e^{-10t} & 0 \le t < 0.04, \\ e^{25t} & 0.04 \le t < 0.1, \text{ and} \\ e^{-50t} & t \ge 0.1. \end{cases} \qquad (12)$$

According to the piecewise exponential function (12), exponential growth of a population is followed by exponential decay and then subsequent exponential regrowth. We illustrate this demographic history on the log scale with a dashed piecewise linear curve in all 4 plots of figure 4. Looking at the classical skyline plot in the top-left corner of figure 4, it is clear that the simulated intercoalescent intervals do not permit an accurate and detailed reconstruction of the population dynamics. However, both the fixed-tree and BEAST time-aware Bayesian skyrides capture all important features of the true trajectory, some better than the others. For example, the initial exponential growth is estimated very well by this model. The time-aware Bayesian skyride detects the subsequent decay in the effective population size but does not fully recover the exponential nature of this decline. The second exponential growth phase is not recovered well. However, the posterior distribution of the effective population size under the time-aware Bayesian skyride suggests that the population bottleneck is indeed followed by an increase of the effective population size. The uniform Bayesian skyride performs slightly worse than its time-aware analog. The Bayesian skyline plot model also performs well. However, this method misses the decay phase of the population dynamics and instead predicts a constant population size for this time period. The ORMCP model clearly oversmoothes the effective population size trajectory. This model suggests that the population has been increasing during its entire history and completely misses the bottleneck component of the demographic history. It is possible to improve the performance of the ORMCP model by substantially increasing the prior mean number of change points in the model (data not shown). However, it is not clear how to choose an appropriate value of this prior parameter a priori and post hoc adjustments invalidate inference by employing the data more than once.

To provide a comparative summary of the performance of the Bayesian skyride and MCP models, we report the percent error (PE) for all simulation scenarios. We define this PE as

$$\text{PE} = \int_0^{\text{TMRCA}} \frac{|\hat{N}_e(t) - N_e(t)|}{N_e(t)} dt \times 100, \qquad (13)$$

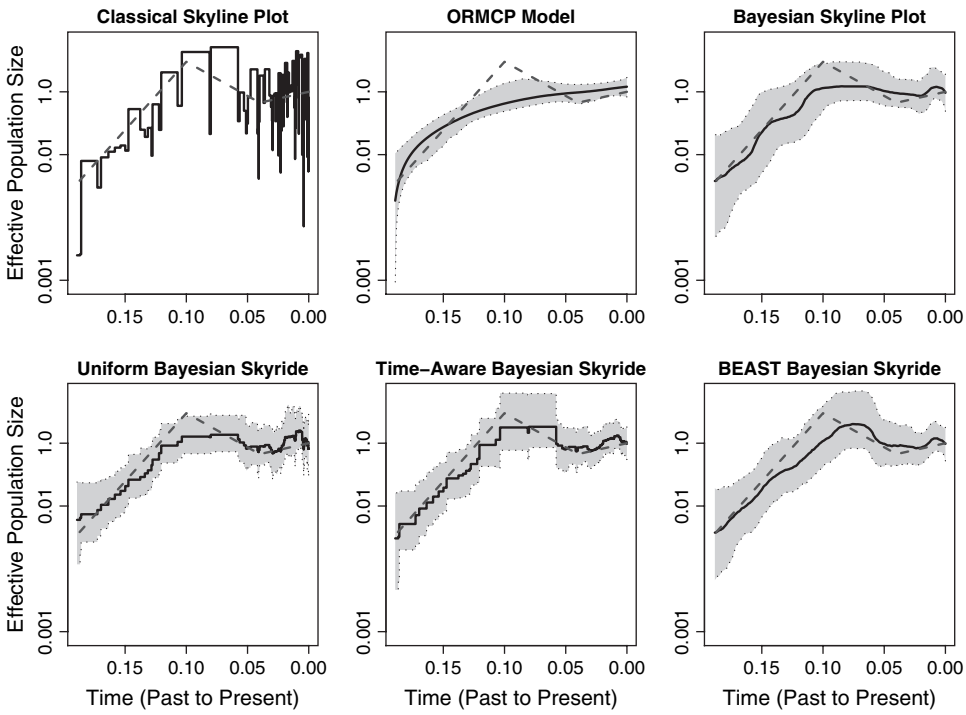

FIG. 4.—Simulated bottleneck. See figure 2 for the legend explanation.

where $\hat{N}_e(t)$ is the estimated posterior median of a population size trajectory and $N_e(t)$ is a true population size trajectory. Table 1 reports PEs for the 6 methods and 3 simulation scenarios. It is only fair to compare the fixed-tree and BEAST methods with each other but not across these 2 method groups. Among the fixed-tree methods, the time-aware Bayesian skyride is a clear leader according to table 1. The BEAST Bayesian skyride performs better than the Bayesian skyline plot for the constant and exponential growth models. Despite a mild oversmoothing effect, produced by the Bayesian skyline plot model in the bottleneck analysis, this method produces PE smaller than the BEAST Bayesian skyride method in the last simulation scenario.

Because running BEAST is time consuming even under a simple parametric demographic model, we use our simulation study to investigate the computational cost of incorporating the Bayesian skyride into BEAST. On a dual-processor Pentium 3.4 Ghz with 4 GB of RAM, the Bayesian skyride analysis took 2.75, 2.30, and 2.21 h for the constant, exponential growth, and bottleneck simulated data sets, respectively. The corresponding running times for the Bayesian skyline plot with 10 change points are 1.70, 1.81, and 1.58 h, indicating that the Bayesian skyride method is only slightly slower than the Bayesian skyline plot. We expect to further optimize our BEAST implementation and improve computational efficiency of the Bayesian skyride.

Table 1
Percent Error in Simulations

Model	Constant	Exponential	Bottleneck
ORMCP	14.0	1.7	7.4
Uniform Bayesian skyride	32.8	1.5	5.9
Time-aware Bayesian skyride	2.8	1.2	4.8
Bayesian skyline plot	38.2	1.6	5.2
BEAST Bayesian skyride	1.7	1.0	5.4

NOTE.—We compare PEs, defined in equation (13), for the ORMCP model, uniform and time-aware fixed-tree Bayesian skyrides, Bayesian skyline plot, and BEAST Bayesian skyride.

Population Dynamics of Egyptian HCV

We analyze 63 HCV sequences, sampled in 1993 in Egypt. Pybus et al. (2003) use this data set to study the population dynamics of Egyptian HCV. The analyzed sequences are derived from the HCV E1 genomic region. Pybus et al. (2003) argue that the approximately random sequence sampling, no sign of population substructure, and other properties of these HCV sequences make them very suitable for the coalescent analysis.

We perform a phylogenetic analysis of the HCV sequences using the BEAST software package (Drummond and Rambaut 2007). Following Pybus et al. (2003), we use a strict molecular clock and the HKY substitution model (Hasegawa et al. 1985). We put a coalescent prior with constant population size on genealogies. However, the wide uniform prior over interval [0, 1000] (measured in years) on the TMRCA together with the abundant phylogenetic information in the HCV sequences should limit the coalescent prior influence on estimated genealogies. To estimate branches in units of years, we use a previously estimated mutation rate in the HCV E1 genomic region, 7.9×10^{-4} substitutions/site/year (Pybus et al. 2001). For estimation of the population dynamics with the fixed-tree

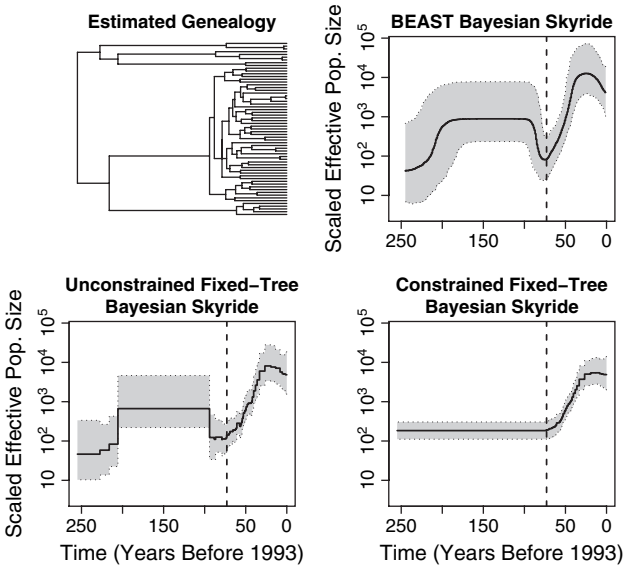

FIG. 5.—Egyptian HCV. In the top left corner, we show the estimated genealogy of the HCV sequences. The rest of the plots demonstrate posterior medians (solid lines) and 95% BCIs (shaded gray areas) of the scaled effective population size trajectories under the BEAST Bayesian skyride (top right), unconstrained (bottom left), and constrained (bottom right) fixed-tree time-aware Bayesian skyride analyses. In the constrained model, the effective population size is forced to be constant prior to the 1920s.

Bayesian skyride, we use the summary, majority clade support genealogy with median node heights depicted in the top-left plot of figure 5. We also use the BEAST Bayesian skyride to estimate the effective population size trajectory and the HCV genealogy simultaneously.

The top-right and bottom-left plots of figure 5 show the posterior medians and 95% BCIs of $N_e(t)$ under the time-aware BEAST and fixed-tree Bayesian skyride models. The similarity between the BEAST and fixed-tree Bayesian skyride results indicates that tree uncertainty does not play a significant role in the estimation of the Egyptian HCV population dynamics. Because time is measured in years, we estimate the effective population size scaled by the generation length per year. These scaled estimates are often interpreted as effective numbers of infections (Pybus et al. 2001). Welch et al. (2005) provide a theoretical justification of such an interpretation.

The exponential growth of HCV infections in the 20th century is the most remarkable aspect of the HCV evolution in Egypt. Pybus et al. (2003) argue that the exponential growth of HCV infections is a result of intravenously administered parenteral antischistosomal therapy (PAT), practiced in Egypt from the 1920s to the 1980s. Our temporal smoothing procedure successfully recovers this exponential growth. However, the effective population size reconstruction is noisier in the time periods preceding the exponential growth phase due to lack of coalescent events. Pybus et al. (2003) hypothesize that the effective number of HCV infections was constant before the start of the exponential growth phase. We test this hypothesis using the fixed-tree Bayesian skyride by constraining intercoalescent effective population sizes to be equal from the TMRCA to the year 1920. In terms of our model parameters, this hypothesis translates to H_0: $\gamma_2 = \cdots = \gamma_3$. The Bayes factor of 12,880 in favor of H_0 decisively supports constant population size hypothesis of Pybus et al..

The posterior summary of the effective population size trajectory under the constrained Bayesian skyride is shown in the bottom-right plot of figure 5. Enforcing a constant population size prior to the 1920s generates long-range effects on the estimation of more recent population sizes due to the increase in the GMRF precision. Under the unconstrained Bayesian skyride, the effective population size trajectory shows a slight decrease in the period 1970–1993. However, the constrained model predicts a constant population size during these years. A gradual transition from the intravenous to oral administration of the PAT started in the 1970s (Frank et al. 2000). We would like to test whether this transition caused a significant decrease in the effective number of HCV infections, seen in the estimates under the unconstrained Bayesian skyride. Our null hypothesis H_0:$\gamma_{62} < \gamma_{46}$, where index 46 corresponds to time period 1960–1977 and index 62 corresponds to time period 1992–1993. The Bayes factor of only 3 in favor of H_0 suggests that the decay of the HCV effective population size in the 1970–1993 period is not statistically significant.

Intraseason Population Dynamics of Human Influenza

To study intraseason population dynamics of human influenza, we compile 3 data sets from sequences reported by Ghedin et al. (2005) that correspond to the 3 flu seasons: 1999–2000, 2001–2002, and 2003–2004. The data sets consist of 48, 59, and 72 hemagglutinin sequences, respectively. All sequences derive from H3N2 isolates. Sequence sampling was restricted to New York State. This should diminish the effects of geographical population structure on estimated genealogies. Recombination also should not play a part in shaping intragenic influenza genealogies because homologous intragenic recombination is very rare in influenza viruses (Steinhauer and Skehel 2002). Because sampling dates for all analyzed sequences are available, all 3 flu season data sets are heterochronous.

Similarly to the Egyptian HCV analysis, we obtain a posterior sample of genealogies using the BEAST constant population size model and Bayesian skyride. In the latter analysis, we estimate influenza effective population size trajectory simultaneously with the viral genealogy. Our model specification of nucleotide evolution is nearly identical to the HCV analysis. We test the molecular clock assumption using Bayes factors (Suchard et al. 2003). After analyzing the intraseason data sets with a lognormal relaxed clock model, proposed by Drummond et al. (2006), we compute the Bayes factors in favor of the molecular clock hypothesis using a harmonic mean estimator. The estimated Bayes factors (season 1999–2000: 111,000, 2001–2002: 19, and 2003–2004: 16) strongly support the molecular clock hypothesis in all 3 data sets. Although the molecular clock together with the heterochronous nature of the data sets in principle allow us to estimate the mutation rates simultaneously with other model parameters, we find that these intrahost influenza data have little information about

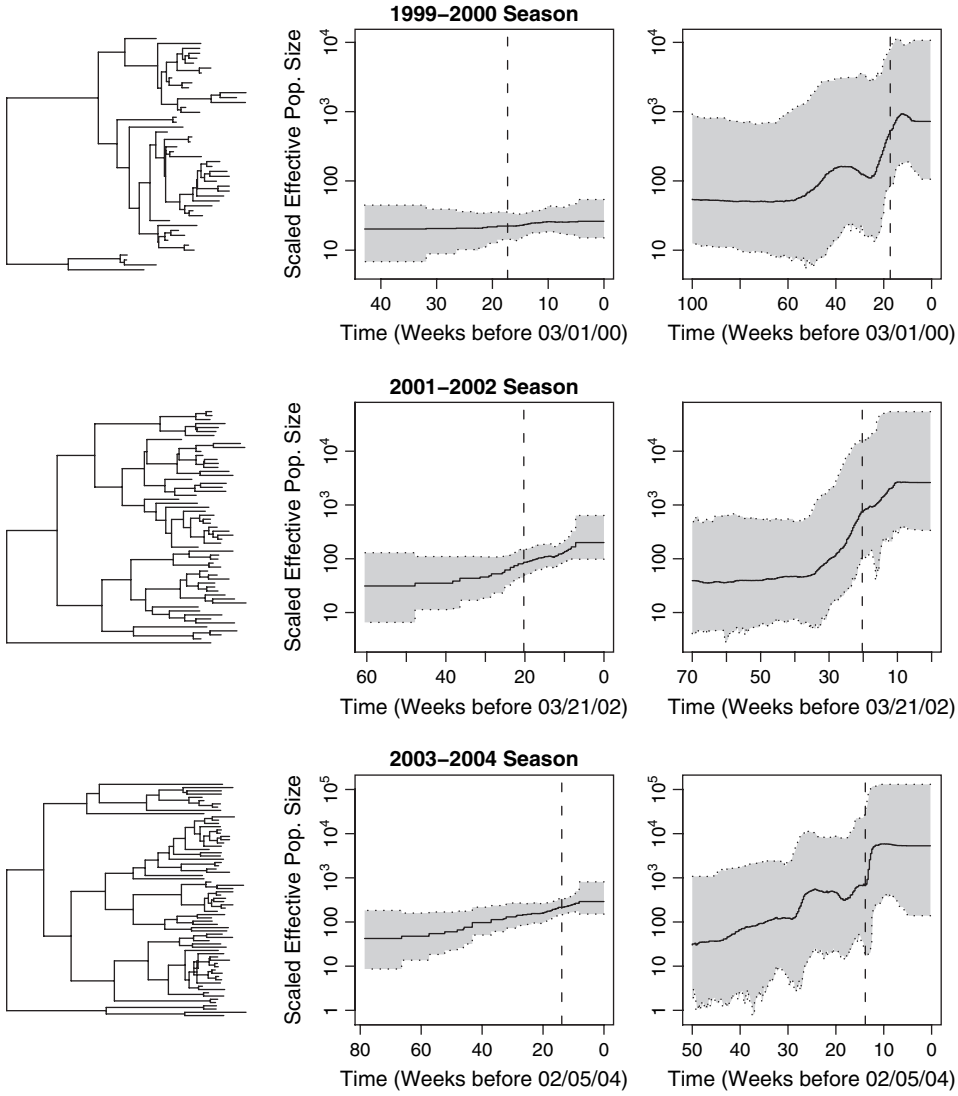

FIG. 6.—Intraseason dynamics of human influenza. For each season, we plot the estimated genealogy (left), the fixed-tree time-aware (middle) and BEAST (right) Bayesian skyride estimates. The vertical dashed lines in the middle and right columns mark 1 October in all 3 seasons.

mutation rates. Therefore, for the mutation rate, we use informative lognormal prior, commensurate with previously obtained estimates of mutation rate in influenza hemagglutinin genes (Fitch et al. 1997; Yang et al. 2007).

From posterior samples of all model parameters in the intraseason data sets, we obtain maximum clade support genealogies with median node heights. We feed these genealogies, shown in the left column of figure 6, into our fixed-tree Bayesian skyride procedure. Branch lengths are measured in units of weeks because these units of time are commonly used for intraseason surveillance of flu epidemics. The middle column of figure 6 shows the fixed-tree Bayesian skyride estimates of $N_e(t)$. Results of the BEAST Bayesian skyride, depicted in the right column of figure 6, significantly differ from the fixed-tree inference. Such discrepancy is a clear indication that genealogical uncertainty cannot be ignored during inference of influenza intraseason population dynamics. Interestingly, applying the fixed-tree Bayesian skyride to a random subset of genealogies, sampled under the constant population size model, does not reveal significant variation of the effective population size trajectory estimates (results not shown). This observation suggests that joint inference of genealogies and $N_e(t)$ should be preferred even when the effect of ignoring genealogical uncertainty is not apparent.

Exponentially growing influenza populations suggest that the viral diversity was increasing before the start of all 3 flu seasons. The posterior medians of the influenza effective population size trajectories exhibit piecewise exponential shapes. However, the wide BCIs of $N_e(t)$ prevent us from studying local features of these curves. These wide BCIs of the effective population size trajectories appropriately reflect the lack of information about influenza genealogies in the sampled sequences. We only point out that the BEAST Bayesian skyride results suggest that the influenza effective population size was growing slower in 2001–2002 than in 1999–2000 and 2003–2004. This is consistent with the Centers for Disease Control and Prevention surveillance

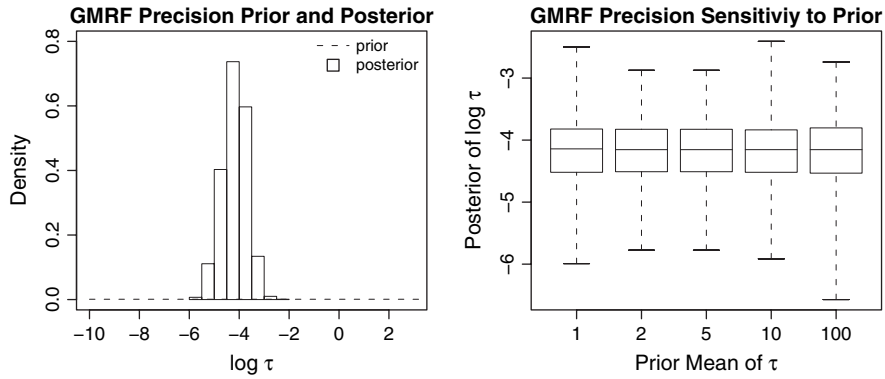

FIG. 7.—Prior sensitivity. In the left plot, we depict the prior density (dashed line) and posterior histogram (vertical bars) of the log-transformed GMRF precision τ. The right plot demonstrates 5 boxplots of the posterior distributions of log τ corresponding to 5 different prior means of τ. In both plots, we use the Egyptian HCV data.

data (http://www.cdc.gov/flu/) reporting a significant delay of flu activity during the 2001–2002 season.

Prior Sensitivity

In Bayesian modeling, it is important to investigate sensitivity of results to the prior assumptions of the model. We place a GMRF smoothing prior on log effective population sizes γ. This prior informs our model only about the smoothness of the population size trajectory, leaving the task of determining an overall level of γ to the likelihood. The GMRF precision parameter τ regulates this degree of smoothness; unfortunately, expert knowledge of τ is rarely known a priori. Using the Egyptian HCV data, we illustrate the prior and posterior distributions of log τ in the left plot of figure 7. We use the log transformation to mitigate boundary effects near 0, facilitating interpretability of distributional summaries. The dramatic difference between the flat and diffuse prior density and highly peaked posterior histogram of log τ suggests that our data alone contain more than sufficient information to estimate τ.

Recall that in all our analyses we set $\alpha = \beta = 0.001$ in τ's prior density (10). To investigate the sensitivity of our results to these hyperprior parameter choices, we reanalyze the Egyptian HCV data using 5 different values of α: 0.001, 0.002, 0.005, 0.01, and 0.1, leaving β unchanged. For these values of α, the prior mean of τ grows from 1, 2, 4, 10, to 100 respectively. We summarize posterior distributions of log τ in the 5 boxplots on the right side of figure 7. These boxplots demonstrate that the posterior distribution of log τ hardly changes when we alter the hyperprior parameter α. We conjecture that this remarkable robustness indicates that our Bayesian temporal smoothing model is very well suited for estimating population size dynamics.

Discussion

We present the Bayesian skyride, a novel coalescent-based, statistical approach for estimating effective population size dynamics. In contrast to previously proposed methods, we explicitly incorporate smoothing into our Bayesian model via a GMRF prior. This strategy allows us to regularize the noisy skyline plot estimates of piecewise constant effective population size trajectories. We make our Bayesian skyride time aware using a weighting scheme based on the sizes of intercoalescent intervals. As with many other smoothing techniques, our GMRF prior has a precision parameter τ that controls the strength of smoothness. Estimating this parameter from data alone can be challenging and often requires injection of prior knowledge (Bernardinelli et al. 1995). However, we find that in all our examples, the intercoalescent intervals contain sufficient information about the GMRF precision τ, eliminating the need of informative priors. Therefore, our method, in contrast to competitors, does not require subjective decisions from the user. The independence of intercoalescent intervals permits us to use very efficient, GMRF-tailored algorithms for sampling from the posterior distribution of the model parameters. This computational efficiency becomes even more critical for integration of the Bayesian skyride into a joint Bayesian estimation of genealogies and population genetics parameters.

We demonstrate that the Bayesian skyride can successfully reconstruct effective population size trajectories under the 3 simulated demographic scenarios. During our simulations, we find that our time-aware Bayesian skyride is superior to the uniform GMRF prior. Therefore, we use the former during our analyses of real data sets. The temporally smoothed scaled effective population size trajectory of the Egyptian HCV demographic history agrees with previous estimates of $N_e(t)$ remarkably well (Pybus et al. 2003; Drummond et al. 2005). Using this example, we illustrate how to formally test one- and two-sided hypotheses in our Bayesian skyride framework. Next, we analyze the intraseason population dynamics of human influenza. We find that fixing a genealogy, estimated under the constant population size demographic model, leads to inadequate estimation of $N_e(t)$. The striking difference between our fixed-tree and BEAST Bayesian skyride methods highlights the importance of joint estimation of the effective population size and the genealogy of sampled sequences.

Other building blocks of the coalescent model may also be needed to accurately determine influenza intraseason population dynamics. Although all influenza sequences were sampled in the same geographical location, the periodic migrational patterns of the virus may have a significant

effect on shaping genealogies relating the sampled sequences. Omitting selection in our coalescent model can also lead to inaccurate estimation of population dynamics. A more detailed coalescent analysis of intrahost influenza evolution should resolve these difficulties.

The influenza example motivates the need for new statistical tools for quantifying commonalities in multiply observed demographic histories. Similar repeated evolutionary patterns occur during intrahost HIV evolution and have important medical implications (Shankarappa et al. 1999). We envision analyzing such repeated patterns using a Bayesian hierarchical framework (Kitchen et al. 2004). Such an approach will require an accurate alignment of observed time intervals and inclusion of external factors that may effect demographic dynamics. The Bayesian skyride is perfectly suitable for the inclusion of such external information as covariates in a generalized linear model framework (MacNab 2003). This proposed methodology will enable statistical testing of environmental effects on demographic histories of populations.

Acknowledgments

We would like to thank John O'Brien for helpful discussions and his advice on the analysis of influenza evolution. We are grateful to Oliver Pybus and another anonymous reviewer for their constructive comments that greatly improved our manuscript. V.N.M. was supported by a Dissertation Year Fellowship from the University of California, Los Angeles Graduate Division. E.W.B. is supported by National Institutes of Health grant AI07370. M.A.S. is an Alfred P. Sloan Research Fellow.

Appendix A: MCMC Sampling Scheme

Fixed Genealogy

We first describe a sampling scheme for the fixed genealogy case. Here, we represent genealogy $\mathbf{g}$ through its vector of intercoalescent and sampling intervals $\mathbf{w}$. To approximate the posterior

$$\Pr(\gamma, \tau|\mathbf{w}) \propto \Pr(\mathbf{w}|\gamma)\Pr(\gamma|\tau)\Pr(\tau), \tag{A1}$$

we parallel the block-updating MCMC scheme of Knorr-Held and Rue (2002). Given current parameter values (τ, γ), we first generate a candidate value for the GMRF precision, $\tau^* = \tau f$, where f is drawn from a symmetric proposal distribution with density $\Pr(f) \propto f + 1/f$ defined on the interval $[1/F, F]$. The tuning constant F controls the distance between the proposed and current values of the GMRF precision. Conditional on τ^*, we propose a new state γ^* for the vector of log-effective population sizes using a Gaussian approximation to the full conditional density

$$\Pr(\gamma|\tau^*, \mathbf{w}) \propto \Pr(\mathbf{w}|\gamma)\Pr(\gamma|\tau^*). \tag{A2}$$

Density (A2) is called a hidden Markov random field because conditional on the Markov field γ, the observed $\mathbf{w}_k$s are distributed independently of each other (Rue et al. 2004). Such a special form of density (A2) allows one to generate samples from its Gaussian approximation using computationally efficient algorithms for sparse matrix computations (Rue 2001). After obtaining a new candidate state (τ^*, γ^*), we accept or reject it in a Metropolis–Hastings step (Metropolis et al. 1953; Hastings 1970).

Incorporating Genealogical Uncertainty

So far, we have assumed that the genealogy $\mathbf{g}$ is known and fixed. However, we do not observe genealogies relating individuals randomly sampled from a population. Instead, we observe molecular sequence data for each individual on the tips of an unknown genealogy. Sequence data and genealogies are connected through the standard assumption that sequence characters are generated by a mutational process that acts along a hidden genealogy. Therefore, the complete likelihood of observing sequence data $\mathbf{D}$ is $\Pr(\mathbf{D}|\mathbf{g}, \mathbf{Q})$, where $\mathbf{Q}$ is a vector of mutational process model parameters. A priori, we assume that $\mathbf{Q}$ and $\mathbf{g}$ are independent. Probability distribution $\Pr(\mathbf{Q})$ depends on the parameterization of the mutational process model. We use the coalescent as a prior for $\mathbf{g}$ so that

$$\Pr(\mathbf{g}|\gamma) \propto \Pr(\mathbf{w}|\gamma), \tag{A3}$$

where $\Pr(\mathbf{w}|\gamma)$ is defined by equation (7). The posterior distribution of all model parameters becomes

$$\Pr(\mathbf{g}, \mathbf{Q}, \gamma|\mathbf{D}) \propto \Pr(\mathbf{D}|\mathbf{g}, \mathbf{Q})\Pr(\mathbf{Q})\Pr(\mathbf{g}|\gamma)\Pr(\gamma). \tag{A4}$$

To approximate this posterior distribution, we equip the software package BEAST (Drummond and Rambaut 2007) with our GMRF MCMC updating scheme. We then merge our fixed-tree analysis with BEAST MCMC kernels for updating $\mathbf{Q}$ and $\mathbf{g}$ to jointly estimate genealogies and population size trajectories.

Appendix B: BEAST Implementation

We have implemented the time-aware Bayesian skyride in the BEAST software package (Drummond and Rambaut 2007). Users may employ the Bayesian skyride both while assuming a fixed evolutionary tree and while integrating over all possible trees given molecular sequence data and a mutational model. Example BEAST XML input blocks that allow users to place the GMRF prior on effective population size dynamics and update the field parameters are provided below:

```
<!— Define field —>
<gmrfSkyrideLikelihood id="coalescent">
    <populationTree>
        <treeModel idref="treeModel"/>
     </populationTree>
    <populationSizes>
        <parameter id="skyride.logPopSize" dimension="62"/>
    </populationSizes>
    <precisionParameter>
        <parameter id="skyride.precision" dimension="1"/>
    </precisionParameter>
</gmrfSkyrideLikelihood>

<!— Update field parameters —>
<gmrfBlockUpdateOperator weight="3" scaleFactor="10.0"
          autoOptimize="true" maxIterations="200"
          stopValue="0.01">
    <gmrfSkyrideLikelihood idref="coalescent"/>
</gmrfBlockUpdateOperator>
```

Literature Cited

Bernardinelli L, Clayton D, Montomoli C. 1995. Bayesian estimates of disease maps: how important are priors? Stat Med. 14:2411–2431.

Biek R, Drummond A, Poss M. 2006. A virus reveals population structure and recent demographic history of its carnivore host. Science. 311:538–541.

Chib S, Jeliazkov I. 2001. Marginal likelihood from the Metropolis-Hastings output. J Am Stat Assoc. 96:270–281.

Drummond A, Ho S, Phillips M, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology. 4:e88.

Drummond A, Nicholls G, Rodrigo A, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 161:1307–1320.

Drummond A, Pybus O, Rambaut A, Forsberg R, Rodrigo A. 2003. Measurably evolving populations. Trends Ecol Evol. 18:481–488.

Drummond A, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.

Drummond A, Rambaut A, Shapiro B, Pybus O. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 22:1185–1192.

Elliott P, Wakefield J, Best N, Briggs D, editors. 2000. Spatial epidemiology: methods and applications. Oxford (UK): Oxford University Press.

Felsenstein J. 1992. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genet Res. 59:139–147.

Fitch W, Bush R, Bender C, Cox N. 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. Proc Natl Acad Sci USA. 94:7712–7718.

Frank C, Mohamed M, Strickland G, et al. (11 co-authors). 2000. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. Lancet. 355:887–891.

Ghedin E, Sengamalay N, Shumway M, et al. (19 co-authors). 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature. 437: 1162–1166.

Green P. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 82:711–732.

Griffiths R, Tavaré S. 1994. Sampling theory for neutral alleles in a varying environment. Philos Trans R Soc Lond B Biol Sci. 344:403–410.

Hasegawa M, Kishino H, Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 22:160–174.

Hastings W. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 57:97–109.

Hein J, Schierup M, Wiuf C. 2005. Gene genealogies, variation and evolution. New York: Oxford University Press.

Hudson R. 1983. Properties of a neutral allele model with intragenic recombination. Theor Popul Biol. 23:183–201.

Kass R, Raftery A. 1995. Bayes factors. J Am Stat Assoc. 90:773–795.

Kingman J. 1982. On the genealogy of large populations. J Appl Probab. 19:27–43.

Kitchen C, Philpott S, Burger H, Weiser B, Anastos K, Suchard M. 2004. Evolution of human immunodeficiency virus type 1 coreceptor usage during antiretroviral therapy: a Bayesian approach. J Virol. 78:11296–11302.

Knorr-Held L, Rue H. 2002. On block updating in Markov random field models for disease mapping. Scand J Stat. 29:597–614.

Krone S, Neuhauser C. 1997. Ancestral processes with selection. Theor Popul Biol. 51:210–237.

Kuhner M, Yamato J, Felsenstein J. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. Genetics. 149:429–434.

MacNab Y. 2003. Hierarchical Bayesian modeling of spatially correlated health service outcome and utilization rates. Biometrics. 59:305–316.

Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E. 1953. Equation of state calculation by fast computing machines. J Chem Phys. 21:1087–1092.

Newton M, Raftery A. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. J R Stat Soc Ser B. 56:3–48.

Notohara M. 1990. The coalescent and the genealogical process in geographically structured population. J Math Biol. 29: 59–75.

Opgen-Rhein R, Fahrmeir L, Strimmer K. 2005. Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo. BMC Evol Biol. 5:6.

Pybus O, Charleston M, Gupta S, Rambaut A, Holmes E, Harvey P. 2001. The epidemic behavior of the hepatitis C virus. Science. 292:2323–2325.

Pybus O, Drummond A, Nakano T, Robertson B, Rambaut A. 2003. The epidemiology and iatrogenic transmission of hepatitis c virus in Egypt: a Bayesian coalescent approach. Mol Biol Evol. 20:381–387.

Pybus O, Rambaut A. 2002. GENIE: estimating demographic history from molecular phylogenies. Bioinformatics. 18:1404–1405.

Pybus O, Rambaut A, Harvey P. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. Genetics. 155:1429–1437.

Rambaut A, Grassly N. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput Appl Biosci. 13:235–238.

Rodrigo A, Felsenstein J. 1999. The evolution of HIV, chapter Coalescent approaches to HIV population genetics. Baltimore (MD): Johns Hopkins University Press.

Rue H. 2001. Fast sampling of Gaussian Markov random fields. J R Stat Soc Ser B. 63:325–338.

Rue H, Held L. 2005. Gaussian Markov random fields: theory and applications. London: Chapman & Hall. (Monographs on Statistics and Applied Probability; vol. 104)

Rue H, Steinsland I, Erland S. 2004. Approximating hidden Gaussian Markov random fields. J R Stat Soc Ser B. 66: 877–892.

Shankarappa R, Margolick J, Gange S, et al. 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol. 73:10489–10502.

Shapiro B, Drummond A, Rambaut A, et al. (27 co-authors). 2004. Rise and fall of the Beringian steppe bison. Science. 306: 1561–1565.

Steinhauer D, Skehel J. 2002. Genetics of influenza viruses. Annu Rev Genet. 36:305–332.

Strimmer K, Pybus O. 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. Mol Biol Evol. 18:2298–2305.

Suchard M, Weiss R, Sinsheimer J. 2003. Testing a molecular clock without an outgroup: derivations of induced priors on branch-length restrictions in a Bayesian framework. Syst Biol. 52:48–54.

Thorne J, Kishino H, Painter I. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol. 15:1647–1657.

Verdinelli I, Wasserman L. 1995. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. J Am Stat Assoc. 90:614–618.

Welch D, Nicholls G, Rodrigo A, Solomon W. 2005. Integrating genealogy and epidemiology: the ancestral infection and selection graph as a model for reconstructing host virus histories. Theor Popul Biol. 68:65–75.

Yang Z, O'Brien J, Zheng X, Zhu H, She Z. 2007. Tree and rate estimation by local evaluation of heterochronous nucleotide data. Bioinformatics. 23:169–176.

Jody Hey, Associate Editor

Accepted April 7, 2008