MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 423–434 doi: 10.1111/j.1755-0998.2011.02988.x

INVITED TECHNICAL REVIEW

Skyline-plot methods for estimating demographic history from nucleotide sequences

SIMON Y. W. HO*† and BETH SHAPIRO‡
*Centre for Macroevolution and Macroecology, Research School of Biology, Australian National University, ACT 0200, Australia, †School of Biological Sciences, University of Sydney, NSW 2006, Australia, ‡Department of Biology, The Pennsylvania State University, University Park, PA 16802–5301, USA

Abstract

Estimation of demographic history from nucleotide sequences represents an important component of many studies in molecular ecology. For example, knowledge of a population's history can allow us to test hypotheses about the impact of climatic and anthropogenic factors. In the past, demographic analysis was typically limited to relatively simple population models, such as exponential or logistic growth. More flexible approaches are now available, including skyline-plot methods that are able to reconstruct changes in population sizes through time. This technical review focuses on these skyline-plot methods. We describe some general principles relating to sampling design and data collection. We then provide an outline of the methodological framework, which is based on coalescent theory, before tracing the development of the various skyline-plot methods and describing their key features. The performance and properties of the methods are illustrated using two simulated data sets.

Keywords: Bayesian inference, coalescent, genealogy, phylogenetic analysis, skyline plot

Received 4 October 2010; revision received 14 December 2010; accepted 24 December 2010

Introduction

The demographic history of a population leaves a signature in the genomes of its modern representatives. Reconstructing this history can allow us to gain useful insights into various evolutionary and population-genetic processes, for example by testing correlations between demographic and palaeoclimatic events (Drummond *et al.* 2005; Campos *et al.* 2010), examining the factors driving past population dynamics (Finlay *et al.* 2007; Atkinson *et al.* 2008; Stiller *et al.* 2010), and tracing the transmission and spread of viruses (Kitchen *et al.* 2008; Magiorkinis *et al.* 2009).

A range of methods are available for estimating demographic patterns using nucleotide sequence data (Emerson *et al.* 2001). Most of these assume that the population history is adequately described by a simple parametric model, such as exponential or logistic growth. The demographic history can then be inferred by evaluating support for the candidate models or by estimating the values of the parameters in a flexible population model (Harpending *et al.* 1998; Weiss & von Haeseler 1998). For example, values of the growth parameter in an exponential-growth model can take positive, negative or zero values, suggesting respective trends of increasing, decreasing or constant population size.

In reality, population histories are often more complex than those described by simple parametric models. This concern has motivated the development of nonparametric and semi-parametric methods for inferring demographic history from sequence data or from an estimated genealogy (e.g. Fu 1994; Polanski *et al.* 1998; Pybus *et al.* 2000). The 'skyline plot' framework, introduced by Pybus *et al.* (2000), enables estimation of historical patterns of population size from a genealogy without the need for a priori restrictions on possible demographic models. A number of methodological extensions have subsequently been described, giving rise to a small family of skyline-plot methods (Table 1).

All skyline-plot methods are fundamentally based on coalescent theory, which quantifies the relationship between the genealogy of the sequences and the demographic history of the population (Kingman 1982a,b). In the coalescent framework, lineages are traced backwards from a sample of sequences, with pairs of lineages randomly coalescing until a single common ancestor is reached. The genealogy of any given locus represents a

Correspondence: Simon Ho, Fax: +61 2 91140979;
E-mail: simon.ho@sydney.edu.au

Table 1 Comparison of skyline-plot methods for estimating demographic history from DNA sequence data

Method	Software	Estimation of coalescent error	Estimation of phylogenetic error	Able to analyse heterochronous sequences	Able to analyse multiple loci simultaneously	Reference
Classical skyline	GENIE, APE	No	No	Yes	No	Pybus *et al.* (2000)
Generalized skyline	GENIE, APE	No	No	No	No	Strimmer & Pybus (2001)
Bayesian MCP	APE	Yes	No	No	No	Opgen-Rhein *et al.* (2005)
Bayesian skyline	BEAST	Yes	Yes	Yes	No	Drummond *et al.* (2005)
Bayesian skyride	BEAST	Yes	Yes	Yes	No	Minin *et al.* (2008)
Extended Bayesian skyline	BEAST	Yes	Yes	Yes	Yes	Heled & Drummond (2008)

single realization of this stochastic process, the conditions of which are determined by the history of the population, natural selection and other factors (Donnelly & Tavaré 1995). Reconstruction of this demographic history involves estimating the genealogy and inferring the effective population size at different points along the genealogical timescale. The effective population size reflects the number of individuals that contribute offspring to the descendent generation and is almost always smaller than the census population size. This coalescent framework gives rise to most of the fundamental properties of skyline-plot methods, including their attendant assumptions and limitations.

We begin this technical review by describing some of the issues associated with selection of an appropriate data set, particularly with respect to sampling design. We then outline the basic methodological framework before tracing the development of the various skyline-plot methods and describing their key features. The performance and properties of the methods are illustrated using two simulated data sets.

Compiling a data set

The process of inferring demographic history begins with collection of nucleotide sequence data from the population of interest. As one would expect, the data set needs to be informative for reliable demographic estimation. Given monetary and other limitations, sampling design forms an important aspect of the research project. It is, however, a significant challenge to devise recommendations as to how to compile the ideal data set, particularly because the strength of the demographic signal will vary considerably among populations and among species. We know, for example, that the mitochondrial DNA of woolly mammoths contains relatively few informative sites compared with that of either bison or muskoxen (Shapiro *et al.* 2004; Debruyne *et al.* 2008; Gilbert *et al.* 2008; Campos *et al.* 2010). Presumably, this is because of differences in population histories as well as biological factors such as mutation rate and generation time. This results in significantly less power to infer demographic history given the same alignment size. Below, we outline some general principles relating to sampling design for skyline-plot analysis. However, it should be borne in mind that the importance and relevance of these principles will vary among data sets.

One of the basic principles of data selection is to satisfy the assumptions of the models used in the analysis. The coalescent framework makes several simplifying assumptions about the population from which the samples are drawn (Donnelly & Tavaré 1995). To minimize the violation of these assumptions, sequences should ideally be obtained from individuals that have been randomly sampled from a panmictic population. The sampled sequences are also assumed to be orthologous, nonrecombining and neutrally evolving. Data sets can comprise any number of loci; a single locus might be sufficient for some populations, but demographic estimates are considerably improved by including multiple unlinked loci (Heled & Drummond 2008). Markers should be chosen carefully to optimize information content relative to sequencing effort.

Selecting individuals

An ideal study sample comprises individuals randomly sampled from across the range of the population of interest. The sample should include a sufficient number of individuals so that the diversity of the population can be captured (Drummond *et al.* 2005).

In highly structured populations, it is advisable to analyse subpopulations separately to satisfy the assumption of panmixia (e.g. Miller *et al.* 2009). In many cases, the degree of population structure is difficult to judge in advance because it is often determined by analysis of genetic markers (Subramanian *et al.* 2009). The impact of population structure on skyline-plot methods has not yet been investigated in depth, but it is likely to lead to biased estimates of model parameters such as population

size and mutation rate (Strimmer & Pybus 2001; Drummond *et al.* 2005; Navascués & Emerson 2009). In turn, this bias can produce patterns in the demographic plot that reflect changes in the degree of structure rather than changes in the overall population size (Pannell 2003).

Some skyline-plot methods are able to accommodate heterochronous sequences, such as those from serially sampled viruses or fossil remains, provided that the ages of the sequences are taken into account (see Table 1). If heterochronous samples are used, the concept of random sampling should ideally be extended to include the temporal as well as geographic range of the population (Navascués *et al.* 2010). However, the researcher is rarely at liberty to choose the individuals for sequencing, as sample availability will depend on extrinsic factors. In studies of viruses, availability will often be governed by historical epidemiological practices, while the design of temporal sampling will depend on the mutation rate of the virus (Duffy *et al.* 2008; Firth *et al.* 2010). In the case of ancient DNA, sample preservation is a key determinant, whereas the monetary cost of radiocarbon analysis can place a constraint on the number of samples that can be dated. It is also difficult to gain a reliable estimate of past population distributions and structuring and to evaluate whether these features have remained constant over time. In analyses of heterochronous sequences, including those combining ancient and modern DNA, it is necessary to assume temporal continuity of the study population (Subramanian *et al.* 2009; Firth *et al.* 2010; Navascués *et al.* 2010).

Selecting loci

The selection of genomic loci is not always straightforward, with a variety of factors needing to be taken into consideration. Because skyline-plot methods are applied at the population level, it is preferable to use loci with a high evolutionary rate so that there is appreciable variation among sampled individuals. Generally, increasing the amount of information in the alignment, either by increasing sequence length or by focusing on variable regions, will improve the precision of phylogenetic estimation of the genealogy (Heled & Drummond 2008).

In skyline-plot analyses of animal populations, mitochondrial DNA has been the locus of choice (e.g. Gompert *et al.* 2008; Naderi *et al.* 2008; Rajabi-Maham *et al.* 2008). This is because the mitochondrial genome evolves quickly, does not usually recombine, and in most species is maternally inherited. However, the nonrecombining nature of the mitochondrial genome means that it must be treated as a single locus, even when complete sequences are available. In addition owing to its maternal inheritance, mitochondrial DNA only tracks one aspect of a population's history, which can be misleading if population processes are heavily sex-biased.

Increasing the number of loci, rather than increasing the sequence length from each locus, represents a more effective way of improving accuracy and reducing estimation error. For example, doubling the number of independent loci reduces error by a factor of $\sqrt{2}$ (Heled & Drummond 2008). Given that available skyline-plot methods are unable to account for partial linkage, the sampled loci should be unlinked so that they have mutually independent genealogical histories. This can be ensured, for example, by sequencing loci from different chromosomes. Increasing the number of loci also improves the power to detect and to see beyond population bottlenecks (Heled & Drummond 2008).

Skyline-plot methods are based on a relatively simple form of the coalescent in which sequences are assumed to be evolving neutrally. However, there is abundant evidence that the mitochondrial genome is under purifying selection over short time-frames (e.g. Stewart *et al.* 2008; Subramanian 2009). The effect of selection is to shift the distribution of mutations in the genealogy. For example, purifying selection leads to an excess of mutations near the tips of the genealogy (Fu & Li 1993; Williamson & Orive 2002). This will lead to biased reconstructions of demographic history when skyline-plot methods are used. Noncoding nuclear sequences, including intergenic regions and some introns, are generally subject to weaker selective constraints.

Perhaps more than data quantity, skyline-plot methods are sensitive to data quality. Heled & Drummond (2008) found that even low levels of sequence error (0.01%) can lead to a doubling of estimation error, while Axelsson *et al.* (2009) found that spurious demographic trends were reconstructed when postmortem damage was artificially induced in ancient sequence data. This has highlighted the need for exceptional data quality in skyline-plot analyses of nucleotide sequences. The manifold sequencing coverage achieved using high-throughput sequencing methods can reduce the risk and presence of errors. For ancient DNA data, estimation biases caused by postmortem damage can be alleviated by using phylogenetic models of DNA decay (Ho *et al.* 2007; Rambaut *et al.* 2009).

Selection of age calibrations

It is usually of interest not only to infer the relative timing of past demographic events, but also to attach a real timescale to them. To achieve this, independent calibrating information is required to separate genetic branch lengths (measured in mutations per site) into their two components: rate and time duration. In some cases, the mutation rate is known from a previous, independent

study, and this can be incorporated as a fixed value or used to inform the prior distribution of the rate in a Bayesian analysis.

When the rate is unknown, it is instead necessary to incorporate information about time. This can be done by fixing or constraining the age of at least one node in the genealogy, based on an estimate from an independent source of data. For example, information from the fossil record can be used to constrain the timing of a specific divergence event in the tree. Unfortunately, fossil calibrations are rarely available for studies of population-level data. More commonly, biogeographic hypotheses are invoked for the purpose of calibration, whereby the timing of a particular population divergence is assumed to coincide with a dated geological event. The uncertainty associated with palaeontological and biogeographic calibrations is rarely trivial, but this can be readily taken into account if the analysis is done in a Bayesian phylogenetic framework (for a recent review, see Ho & Phillips 2009).

Under some circumstances, such as studies of ancient DNA and serially sampled viruses, the sequence data have been obtained from samples of different ages. If the sampling interval of the sequences is substantial relative to the total evolutionary history of the study population, the sequence ages can provide sufficient calibrating information for the analysis (Rambaut 2000; Drummond *et al.* 2003). Sometimes, the ages are known exactly, for example through hospital or museum records; otherwise, they can be estimated radiometrically, stratigraphically or phylogenetically (Shapiro *et al.* 2011). If heterochronous sequence data are included in the analysis, it is advisable to investigate the effect of sampling times on the demographic inference, particularly if there are apparent temporal biases in sampling (Ho *et al.* 2008; Stiller *et al.* 2010).

Basic methodological framework

Reconstructing demographic history from a sequence alignment involves two distinguishable and separable steps: (i) estimating the genealogy from the sequence data and (ii) estimating the population history based on the genealogy. Some methods combine these two steps in a single analytical framework, allowing the genealogy and population history to be coestimated from the alignment (Drummond *et al.* 2002, 2005). The two steps are described in detail below.

Estimating the genealogy

The first step is to estimate the genealogy of the sampled individuals from the aligned sequences. The genealogy includes the relationships among the individuals (tree topology) as well as their times of divergence (node times) (Fig. 1a). The node times can be relative or absolute, depending on whether calibrating information is available. Genealogical estimation can be done using standard phylogenetic methods, such as those implemented in a maximum-likelihood or Bayesian framework (e.g. Swofford 2003; Drummond & Rambaut 2007).

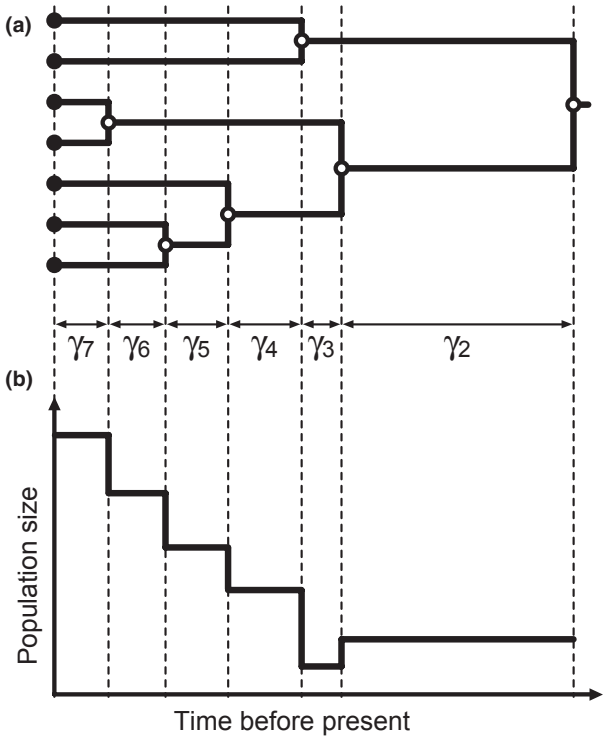

Fig. 1 Estimation of demographic history from a genealogy. (a) An estimated genealogy with branch lengths proportional to time. Filled circles denote terminal nodes (sampled individuals), while empty circles denote internal nodes. Coalescent intervals, denoted by γ_i, are delineated by chronologically successive nodes in the tree. (b) Population size (N) is estimated for each coalescent interval according to the relationship $N_i = \gamma_i i(i-1)/2$, where i denotes the number of lineages in a given coalescent interval.

One condition is that the branch lengths of the tree are proportional to time, with time being measured in mutations, years or generations. If there is significant rate heterogeneity among lineages, an estimate of the genealogy can be obtained in a relaxed-clock framework (Drummond *et al.* 2006). Some relaxed-clock methods require a fixed tree topology to estimate node times (e.g. Thorne *et al.* 1998; Sanderson 2002).

Estimation of the genealogy carries an amount of error, with respect to the topology as well as the branch lengths. The uncertainty in the genealogical estimate, here termed 'phylogenetic error', can be substantial if the genealogy contains short internal branches. This can be the result of either a low mutation rate or short time intervals between nodes. Moreover, given the generally modest amount of genetic variation at the intraspecific level,

especially in slowly evolving organisms such as large vertebrates, branch lengths in the genealogy often have a large amount of stochastic variance. For the purposes of reconstructing the demographic history, however, it is not always crucial that the genealogy is well resolved, particularly when estimates are being averaged across a large number of trees (as in Bayesian methods) (Drummond *et al.* 2005).

Estimating population history from the genealogy

The second step is to estimate the population history from the genealogy. This procedure depends only on the timing of the coalescent events and not on the exact genealogical relationships among the sequences (Pybus *et al.* 2000). For example, coalescent events occurring in rapid succession are normally indicative of small population size. To reconstruct demographic history, skyline-plot methods take advantage of a relatively simple relationship between the population size and the expected length of the coalescent interval. Specifically, the mean population size in each interval can be estimated by the product of the interval size (γ_i) and $i(i-1)/2$, where i is the number of genealogical lineages in the interval (Fig. 1a) (Hudson 1982; Kingman 1982b; Tajima 1983). Thus, this relationship gives an estimate of the population size for each coalescent interval in the estimated genealogy (Fig. 1b), producing a piecewise reconstruction of the demographic history that bears a superficial resemblance to the eponymous skyline of a city (Pybus *et al.* 2000). An extension of the coalescent by Rodrigo & Felsenstein (1999) allows the inclusion of heterochronous data.

Reconstruction of the population history from the genealogy usually involves considerable uncertainty, referred to here as 'coalescent error'. Coalescence is a stochastic process; any single genealogy only represents a single, random realization of this process. In particular, estimation of the population size in each coalescent interval is subject to a large amount of error and is equivalent to estimating the mean of an exponential distribution given only a single sample from the distribution (Minin *et al.* 2008). Coalescent error increases towards the root of the genealogy, where population history is reconstructed from fewer lineages. For example, the population size in the oldest coalescent interval (γ_2 in Fig. 1) is estimated from just two lineages. This is a significant consideration; if the population has remained at a constant size, the oldest coalescent interval represents, on average, half of the entire genealogy and contributes to most of the variance in the overall depth of the genealogy (Hein *et al.* 2005).

Skyline-plot methods

Simulations

To allow the properties of different skyline-plot methods to be compared, two data sets were generated via simulation. Each data set comprised 150 sequences of 2000 nucleotides, simulated with a mutation rate of 10^{-7} mutations/site/year. In the first scenario, sequences were generated according to a demographic model in which the population grew at an exponential rate of 10^{-4} until 50 000 years ago then remained at a constant size of 10^6 from 50 000 years ago until the present (Fig. 2a). In the

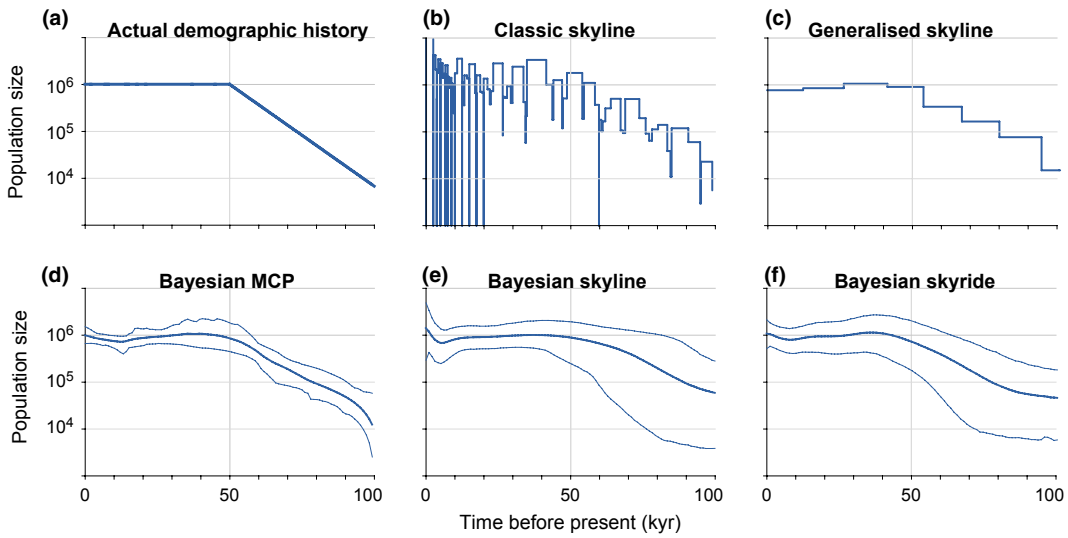

Fig. 2 Performance of different skyline-plot methods for a simulated data set. The actual demographic history used for simulation is shown in panel A, with the faint vertical line at 50 kyr indicating a change-point in the demographic function. The remaining panels show the reconstructions of demographic history by five skyline-plot methods. Note the logarithmic scale on the *y*-axis. Further details of the simulation model are given in the Appendix.

second scenario, the population remained at a constant size of 10^5 until 50 000 years ago then remained at a constant size of 10^6 from 50 000 years ago to the present (Fig. 3a). Further details about the simulated data sets are given in the Appendix.

Classic skyline

The first skyline-plot method was developed by Pybus *et al.* (2000) and is now referred to as the classic skyline plot. In this method, the genealogy of the sequences needs to be obtained independently and is assumed to be known without error (i.e. phylogenetic error is assumed to be negligible). A separate population size is estimated for each coalescent interval in the genealogy, following the relationship described above (Fig. 1). Owing to the number of free parameters, this method tends to produce noisy reconstructions of demographic history. This is particularly the case when the genealogy contains a large number of short, possibly zero-length, branches (Pybus *et al.* 2000).

Classic skyline plots estimated from the two simulated data sets are shown in Figs 2b and 3b. In both cases, some semblance of the underlying trend is discernable, but the plots are very noisy. In particular, there appears to be considerable noise towards the present, which is a consequence of the large number of short coalescent intervals. If data from multiple loci are available, a classic skyline plot can be produced from each genealogy, and the plot can be averaged across these. Assuming that the different loci reflect the same demographic history, which might not be true when they have different modes of inheritance, this should remove some of the noise present in individual skyline plots (Pybus *et al.* 2000).

The classic skyline plot has been implemented in the software GENIE 3.0 (Pybus & Rambaut 2002) and in the *Analyses of Phylogenetics and Evolution* (*APE*) package (Paradis *et al.* 2004) (Table 1). The only form of input that is required is an estimate of the genealogy, or even just a list of the sizes of the coalescent intervals (γ_i in Fig. 1). The classic skyline plot is able to accommodate heterochronous sequences.

Generalized skyline

The presence of short coalescent intervals can lead to a large amount of noise in the demographic reconstructions obtained using the classic skyline plot. To address this problem, Strimmer & Pybus (2001) introduced the generalized skyline plot, which removes short intervals by grouping them with their neighbours if they are below a certain length (ε). Choosing the optimal value of ε, which can be done objectively using the corrected Akaike information criterion (Akaike 1974), represents a trade-off between the removal of noise from the skyline plot and the preservation of the underlying demographic signal.

Generalized skyline plots estimated from the two simulated data sets are shown in Figs 2c and 3c. For each plot, the optimal value of ε was chosen using the corrected Akaike information criterion. The plots retain key features of the demographic history that were evident

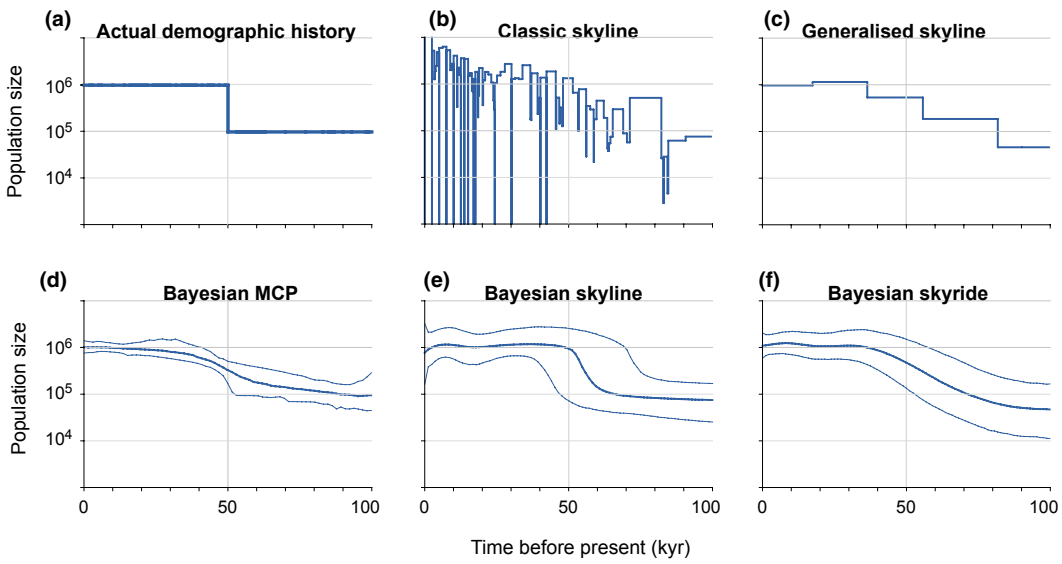

Fig. 3 Performance of different skyline-plot methods for a simulated data set. The actual demographic history used for simulation is shown in panel A, with the faint vertical line at 50 kyr indicating a change-point in the demographic function. The remaining panels show the reconstructions of demographic history by five skyline-plot methods. Note the logarithmic scale on the *y*-axis. Further details of the simulation model are given in the Appendix.

in the classic skyline plot (Figs 2b and 3b), but the noise from short coalescent intervals has been removed. Note that the population increase in the first scenario (Fig. 2c) is reconstructed as a series of large, instantaneous changes in population size, whereas the instantaneous population increase in the second scenario (Fig. 3c) is not recovered.

As with the classic skyline plot, the generalized skyline plot assumes that the genealogy is known without error. Accordingly, estimates of the coalescent times in the genealogy need to be reasonably reliable. Compared with the classic skyline plot, however, the presence of short coalescent intervals will have a less marked effect on the demographic history inferred using the generalized skyline plot (Strimmer & Pybus 2001).

The generalized skyline plot has been implemented in the software GENIE 3.0 (Pybus & Rambaut 2002) and in the *APE* package (Paradis *et al.* 2004). Unlike the classic skyline plot, the generalized skyline plot is unable to accommodate heterochronous sequences (Table 1).

Bayesian multiple-change-point

The Bayesian multiple-change-point method was developed by Opgen-Rhein *et al.* (2005) to smooth the stepped demographic function produced by the generalized skyline plot and to provide an estimate of coalescent error. In implementing this approach, the authors assumed that population size tends to be autocorrelated over time, such that it is unlikely to experience drastic, rapid changes.

In practice, the authors modelled demographic history using a spline, which is a piecewise function comprising a number of polynomial curves demarcated by supporting nodes. First-order splines are used in the multiple-change-point method. This is implemented in a Bayesian framework, allowing reversible-jump Markov chain Monte Carlo (rjMCMC) to be used to sample the number of internal supporting nodes in the spline. Unlike the classic and generalized skyline plots, these change-points do not need to coincide with coalescent events in the genealogy. A posterior sample of splines is obtained from the rjMCMC. The demographic history can then be reconstructed by plotting the mean posterior population size, averaged over the sampled splines, for each time point across a chosen time-frame. Coalescent error can be estimated by plotting the 95% credibility interval for the population size at each time point. Because the population size is averaged over a large number of samples from the posterior, a relatively smooth plot tends to be produced using this approach.

Owing to its implementation in a Bayesian framework, a prior distribution needs to be specified for each parameter. In a strict Bayesian approach, an arbitrary prior can be selected for the population history, for example allowing the population size to be constant or to follow a gamma distribution. Alternatively, Opgen-Rhein *et al.* (2005) suggested that an empirical Bayes approach can be taken, whereby the skyline plot can be used to provide a prior mean estimate of the population history.

Bayesian multiple-change-point plots estimated from the two simulated data sets are shown in Figs 2d and 3d. The first demographic scenario is reconstructed reasonably well. Note that the plot presents a smooth change in the population size, in contrast with the stepped nature of the classic (Fig. 2b) and generalized (Fig. 2c) skyline plots. The Bayesian multiple-change-point method does not recover the stepped change in the second simulation. This result is caused by the assumption of population autocorrelation, whereby drastic population-size changes are treated as being very unlikely. However, this penalty can be revised by altering the prior distributions of relevant parameters in the model.

The Bayesian multiple-change-point method is available in the *APE* package (Paradis *et al.* 2004), implemented in the R statistical environment (Table 1). An estimate of the genealogy is the only input that is required. As with the classic and generalized skyline plots, the Bayesian multiple-change-point method assumes that the genealogy and coalescent times are known. In principle, however, the Bayesian multiple-change-point method can be extended to MCMC sampling that takes phylogenetic uncertainty into account (Opgen-Rhein *et al.* 2005).

Bayesian skyline

In most cases, it is impractical to ignore phylogenetic error in the inferred genealogy. Uncertainty in estimates of node times can be substantial, especially given the intraspecific nature of the sequence data used for skyline-plot analyses. To address this problem, Drummond *et al.* (2005) presented the Bayesian skyline plot, implemented in a framework in which the genealogy, demographic history and substitution-model parameters are coestimated in a single analysis. The resulting plot of population history, which is averaged across the posterior sample of population sizes over time (as in the Bayesian multiple-change-point method), includes credibility intervals that represent the combined phylogenetic and coalescent uncertainty.

The Bayesian skyline plot has its roots in the generalized skyline plot, with which it shares several defining characteristics. Both methods employ a piecewise-constant model in which the population size is constant in each interval and changes instantaneously between successive intervals (Fig. 1). Both methods allow multiple coalescent intervals to be grouped, thereby reducing the

noise associated with short coalescent intervals. Unlike the generalized skyline plot, however, the Bayesian skyline plot requires that the number of groups be chosen a priori. In the absence of rigorous guidelines for choosing the number of groups, this represents a somewhat subjective step in the analysis. It is possible that the estimate of demographic history is relatively robust to the chosen number of groups across a range of values (Drummond *et al.* 2005). Nevertheless, choosing an excessive number of groups can increase estimation error and can be problematic in analyses of uninformative data sets (Heled & Drummond 2008).

As with the Bayesian multiple-change-point method, population sizes in successive coalescent intervals are assumed to be correlated. In each interval, the population size is sampled from an exponential distribution centred on the population size in the preceding interval. This reduces the probability of large population changes between intervals.

Bayesian skyline plots estimated from the two simulated data sets are shown in Figs 2e and 3e. For both scenarios, the 95% credibility intervals of the plot are wider than those obtained using the Bayesian multiple-change-point method, owing to the fact that they include the phylogenetic uncertainty associated with estimating the genealogy. In contrast to the Bayesian multiple-change-point method, the Bayesian skyline plot produces a recognizable reconstruction of the population jump in the second scenario. This is because of the piecewise-constant nature of the latter population model, which permits instantaneous changes in population size. However, the 95% credibility intervals are very wide around the population-change event. For an example of the application of the Bayesian skyline plot to real data, see Box 1.

The Bayesian skyline plot is implemented in the phylogenetic software BEAST (Drummond & Rambaut 2007). The Bayesian framework allows the genealogy and demographic history to be coestimated (Table 1). Consequently, estimates of model parameters are integrated over phylogenetic and coalescent uncertainty. For this reason, the Bayesian skyline plot can be a useful method even when a specific population parameter is of primary interest and the demographic history is a 'nuisance' parameter to be integrated out. A relaxed molecular clock can be employed to accommodate rate variation among lineages (Drummond *et al.* 2006).

Box 1: Comparing the demographic histories of two bear species

Various studies of ancient DNA have been able to shed light on the relative impacts of climatic and anthropogenic factors on megafaunal populations (Greenwood 2009; Hoelzel 2010). Stiller *et al.* (2010) used a Bayesian skyline approach to investigate the population dynamics of the cave bear (*Ursus spelaeus*) in the late Pleistocene, with a view to understanding the causes of its extinction about 24 000 years ago.

Separate alignments were assembled from the mitochondrial D-loop of 25 ancient and 15 modern brown bears (*Ursus arctos*; 177 nucleotides) and 59 ancient cave bears (251 nucleotides). The sampling times of the sequences, most of which were estimated by radiocarbon dating, were able to provide temporal calibrating information. Each data set was analysed in BEAST (Drummond & Rambaut 2007) using a Bayesian skyline plot with 10 groups (Drummond *et al.* 2005).

Stiller *et al.* (2010) found that the population of brown bears remained relatively constant throughout the past 100 000 years (Fig. 4). In contrast, cave bears experienced a striking decline during the 25 000 years prior to their extinction. Support for this demographic estimate was confirmed using a Bayes-factor comparison with a simpler constant-size model. The authors conducted a range of further analyses to investigate the possible effects of biased sampling and other confounding factors, but were unable to find any challenges to the validity of their demographic estimate.

The results suggest that the terminal population decline in cave bears was unlikely to have been driven entirely by the climatic changes of the Last Glacial Maximum, which began about 30 000 years ago. Instead, it is possible that modern humans competed with cave bears for suitable cave sites. This study demonstrates the effective combination of heterochronous sequences and the Bayesian skyline plot. Stiller *et al.* (2010) were able to recover a clear demographic pattern from a relatively short mitochondrial alignment, even when phylogenetic and coalescent error was taken into account.

Bayesian skyride

Minin *et al.* (2008) developed the Bayesian skyride method with the aim of providing an alternative model of demographic change. Like the Bayesian skyline and multiple-change-point methods, the Bayesian skyride assumes that there is some degree of autocorrelation in the population size. In this method, differences between population sizes in successive coalescent intervals are penalized, with the penalty either being dependent on the lengths of the coalescent intervals ('time-aware') or being independent of time. This is done using a smoothing prior based on a Gaussian Markov random

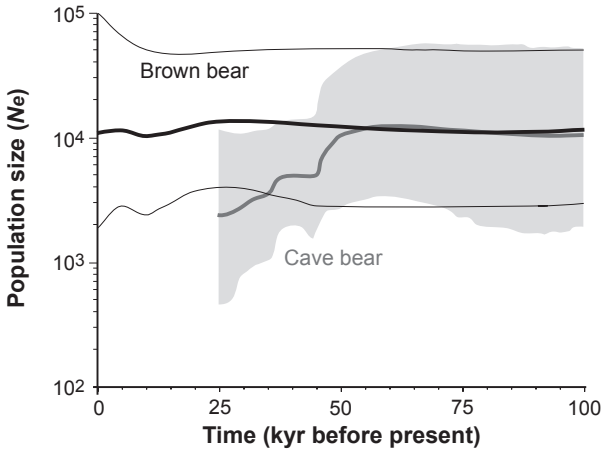

Fig. 4 Demographic histories of brown bear (*Ursus arctos*) and cave bear (*Ursus spelaeus*) estimated using the Bayesian skyline plot from mitochondrial D-loop sequences. The extinction of the cave bear is indicated by the termination of the skyline plot around 24 000 years before present. The figure is modified from Stiller *et al.* (2010).

field. The degree of smoothing is governed by a single parameter, the value of which is informed by the data.

The time-aware penalty on population-size changes stands in contrast with the time-independent approach taken in the Bayesian skyline plot. In the latter, the penalty depends only on the magnitude of the change, irrespective of the length of the coalescent interval. In the time-aware approach implemented in the Bayesian skyride method, the penalty is effectively reduced for longer coalescent intervals. This is equivalent to assuming that population size changes gradually over time.

Bayesian skyride plots estimated from the two simulated data sets are shown in Figs 2f and 3f. The plots are smoother than those obtained using the Bayesian skyline plot. The Bayesian skyride does not recover the sudden population change in the second scenario, a result reminiscent of that obtained using the Bayesian multiple-change-point method.

The Bayesian skyride method has been implemented in the Bayesian phylogenetic software BEAST (Drummond & Rambaut 2007) (Table 1). As with the Bayesian skyline plot, which is available in the same programme, it is possible to coestimate the genealogy and demographic history.

Extended Bayesian skyline

Previous skyline-plot methods were based on individual genealogies, precluding the simultaneous analysis of multiple loci. This does not represent an ideal approach, however, because there is considerable coalescent error associated with estimates made using a single locus. This is because an individual genealogy only represents a single realization of a stochastic process (the coalescent). Heled & Drummond (2008) presented the extended Bayesian skyline plot, which permits the analysis of multiple unlinked loci. Increasing the number of independent loci allows the uncertainty in the coalescent to be assessed, leading to an improvement in the reliability of demographic inference and a substantial reduction in estimation error.

In the extended Bayesian skyline plot, demographic reconstruction is conditioned on the estimated genealogies of the loci in the data set. The different loci are assigned population-size factors to account for any differences in the mode of inheritance. For example, the effective population size of biparentally transmitted, diploid autosomal DNA is about four times that of maternally transmitted, haploid mitochondrial DNA.

One of the advantages in analysing multiple loci is that there is a considerable improvement in the ability to detect past population bottlenecks. Extreme bottlenecks can cause the demographic signal to be erased at a given locus, preventing recovery of the population history prior to the bottleneck event. Without the availability of sequences antedating the event (i.e. ancient DNA), the only possibility of recovering the demographic history comes from the analysis of multiple loci. In this case, there is a chance that some of the loci have preserved the prebottleneck demographic signal.

The extended Bayesian skyline plot permits the use of a piecewise-linear model to describe demographic history, allowing the population size to change continuously along each interval. Previous skyline methods employed a piecewise-constant model in which the population size remained constant in each coalescent interval and changed instantaneously between successive intervals, which is perhaps less biologically realistic.

Unlike the standard Bayesian skyline plot, the extended Bayesian skyline plot allows the number of groups of coalescent intervals to be determined using stochastic search variable selection (Kuo & Mallick 1998), rather than needing to be chosen a priori. If there is only a single group, the demographic trend collapses to that of a constant population size. By default, the prior on the number of groups is given a mean of ln(2), which places a probability of 0.5 on a constant population size.

The extended Bayesian skyline plot has been implemented in the Bayesian phylogenetic software BEAST (Drummond & Rambaut 2007). As with the Bayesian skyline and skyride methods, the genealogy, demographic history and other model parameters can be coestimated (Table 1). Different partitions of the data set can be given independent substitution models, enabling the rate and pattern of the evolutionary process to vary among loci.

Table 2 Units of axes in skyline plots

Calibrating information		Units of x-axis in skyline plot and of branch lengths in the genealogy	Quantity measured on y-axis in skyline plot
Type	Units		
None	n/a	Mutations/site	Population size (N_e) × mutation rate (μ)
Mutation rate	Mutations/site/year	Years	Population size (N_e) × generation time* (τ)
	Mutations/site/generation	Generations	Population size (N_e)
Nodal age(s)	Years	Years	Population size (N_e) × generation time* (τ)
	Generations	Generations	Population size (N_e)

*Generation time measured in years.

Using skyline-plot methods

Interpretation of skyline plots

Skyline plots show the relationship between the population size (y-axis) and the amount of change (x-axis). The units on the axes of the plot will depend on the form of the calibrating information. Three situations can arise: (i) no calibrating information, (ii) calibration using a known rate and (iii) calibration using known nodal age(s). These situations are summarized in Table 2.

Choosing a method

The characteristics of the available data will usually determine which of the skyline-plot methods is most appropriate (see Table 1). If the data set comprises sequences that have evolved rapidly over a long timeframe, phylogenetic error might be relatively modest and the generalized skyline plot could be employed. Generally, however, it is prudent to account for coalescent error (Opgen-Rhein *et al.* 2005), and a Bayesian approach is preferable so that the genealogy and demographic history can be coestimated (Drummond *et al.* 2005). The extended Bayesian skyline method should be used if sequences from multiple independent loci are available (Heled & Drummond 2008).

An important issue in coalescent analysis is that of demographic model selection. Visual inspection of skyline plots can be used to determine whether a simple parametric model might be sufficient for describing population history (Pybus & Rambaut 2002). For example, a completely flat skyline plot might indicate that the population has remained at a constant size over time. In this case, the 95% credibility interval of the demographic estimate should be taken into account to determine whether any apparent population-size changes are supported by the data. This provides a relatively informal tool for demographic model selection. However, it is usually desirable to evaluate population trends in a more statistically rigorous manner.

In a Bayesian setting, population models can be compared using Bayes factors. The Bayes factor of model A over model B is given as Pr(data | model A) ÷ Pr(data | model B). Values over 10 indicate strong support for model A over model B (Jeffreys 1961). In principle, this approach can also be used to assess the level of support for population-size changes in skyline plots. Bayes factors can be calculated using the harmonic-mean estimator (Suchard *et al.* 2001), although alternative methods such as thermodynamic integration have superior statistical properties and should be employed if possible (Lartillot & Philippe 2006).

There are several methods for comparing skyline plots with the simplest demographic model (constant size). In the standard Bayesian skyline plot, this can be done by calculating the Bayes factor of an n-group model against a 1-group model. In the extended Bayesian skyline plot, the number of groups of coalescent intervals is chosen using stochastic search variable selection. This presents a straightforward method for testing a putative complex demographic scenario against the hypothesis of constant population size (1-group model).

Concluding remarks

Skyline-plot methods present a powerful set of techniques for inferring past population-size changes from sequence data. Although they are subject to a number of significant limitations, their application to real data has shown the potential of these methods for elucidating complex patterns of demographic history. With the growing availability of multi-locus data sets, the extended Bayesian skyline plot will enable detailed patterns of population history to be reconstructed. Detailed estimates of demographic history will play an increasingly important role in debates over megafaunal

extinction in the late Pleistocene, the response of organisms to climatic change, and viral phylodynamics.

Acknowledgements

S.Y.W.H. is supported by the Australian Research Council and by a start-up fund from the University of Sydney. BS is supported by NSF ARC–0909456 and the NASA Astrobiology Institute.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Atkinson QD, Gray RD, Drummond AJ (2008) mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Molecular Biology and Evolution*, **25**, 468–474.

Axelsson E, Willerslev E, Gilbert MTP *et al.* (2009) The effect of ancient DNA damage on inferences of demographic histories. *Molecular Biology and Evolution*, **25**, 2181–2187.

Campos PF, Willerslev E, Sher A *et al.* (2010) Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5675–5680.

Debruyne R, Chu G, King CE *et al.* (2008) Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Current Biology*, **18**, 1320–1326.

Donnelly P, Tavaré S (1995) Coalescents and genealogical structure under neutrality. *Annual Review in Genetics*, **29**, 401–421.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Nicholls GK, Rodrigo AG *et al.* (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, **161**, 1307–1320.

Drummond AJ, Pybus OG, Rambaut A *et al.* (2003) Measurably evolving populations. *Trends in Ecology and Evolution*, **18**, 481–488.

Drummond AJ, Rambaut A, Shapiro B *et al.* (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology & Evolution*, **22**, 1185–1192.

Drummond AJ, Ho SYW, Phillips MJ *et al.* (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**, e88.

Duffy S, Shackelton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics*, **9**, 267–276.

Emerson BC, Paradis E, Thebaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology and Evolution*, **16**, 707–716.

Finlay EK, Gaillard C, Vahidi SM *et al.* (2007) Bayesian inference of population expansions in domestic bovines. *Biology Letters*, **3**, 449–452.

Firth C, Kitchen A, Shapiro B *et al.* (2010) Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Molecular Biology and Evolution*, **27**, 2038–2051.

Fu Y-X (1994) A phylogenetic estimator of effective population size or mutation rate. *Genetics*, **136**, 685–692.

Fu YX, Li WH (1993) Statistical test of neutrality of mutations. *Genetics*, **133**, 693–709.

Gilbert MTP, Drautz DI, Lesk AM *et al.* (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8327–8332.

Gompert Z, Fordyce JA, Forister ML *et al.* (2008) Recent colonization and radiation of North American Lycaeides (*Plebejus*) inferred from mtDNA. *Molecular Phylogenetics and Evolution*, **48**, 481–490.

Greenwood AD (2009) Ancient DNA and the genetic consequences of late pleistocene extinctions. In: *American Megafaunal Extinctions at the End of the Pleistocene* (ed. Haynes G), pp. 107–123. Springer Science and Business Media BV, Dordrecht, The Netherlands.

Harpending HC, Batzer MA, Gurven M *et al.* (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 1961–1967.

Hein J, Schierup MH, Wiuf C (2005) *Gene Genealogies, Variation and Evolution*. Oxford University Press, Oxford.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.

Ho SYW, Phillips MJ (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, **58**, 367–380.

Ho SYW, Heupink TH, Rambaut A *et al.* (2007) Bayesian estimation of sequence damage in ancient DNA. *Molecular Biology & Evolution*, **24**, 1416–1422.

Ho SYW, Saarma U, Barnett R *et al.* (2008) The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS ONE*, **3**, e1615.

Hoelzel AR (2010) Looking backwards to look forwards: conservation genetics in a changing world. *Conservation Genetics*, **11**, 655–660.

Hudson RR (1982) Testing the constant-rate neutral allele model with protein sequence data. *Evolution*, **37**, 203–217.

Jeffreys H (1961) *The Theory of Probability*, 3rd edn. Oxford University Press, Oxford.

Kingman JFC (1982a) The coalescent. *Stochastic Processes and their Applications*, **13**, 235–248.

Kingman JFC (1982b) On the genealogy of large populations. *Journal of Applied Probability*, **19A**, 27–43.

Kitchen A, Miyamoto MM, Mulligan CJ (2008) Utility of DNA viruses for studying human host history: case study of JC virus. *Molecular Phylogenetics and Evolution*, **46**, 673–682.

Kuo L, Mallick B (1998) Variable selection for regression models. *Sankhya B*, **60**, 65–81.

Lartillot N, Philippe H (2006) Computing Bayes factors using thermodynamic integration. *Systematic Biology*, **55**, 195–207.

Magiorkinis G, Magiorkinis E, Paraskevis D *et al.* (2009) The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. *PLoS Medcine*, **6**, e1000198.

Miller HC, Moore JA, Allendorf FW *et al.* (2009) The evolutionary rate of tuatara revisited. *Trends in Genetics*, **25**, 13–15.

Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, **25**, 1459–1471.

Naderi S, Rezaei HR, Pompanon F *et al.* (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 17659–17664.

Navascués M, Emerson BC (2009) Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. *Molecular Ecology*, **18**, 4390–4397.

Navascués M, Depaulis F, Emerson BC (2010) Combining contemporary and ancient DNA in population genetic and phylogeographical studies. *Molecular Ecology Resources*, **10**, 760–772.

Opgen-Rhein R, Fahrmeir L, Strimmer K (2005) Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo. *BMC Evolutionary Biology*, **5**, 6.

Pannell JR (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution*, **57**, 949–961.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Polanski A, Kimmel M, Chakraborty R (1998) Application of a time-dependent coalescence process for inferring the history of population size changes from DNA sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 5456–5461.

Pybus OG, Rambaut A (2002) GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics*, **18**, 1404–1405.

Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, **155**, 1429–1437.

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of Mus musculus domesticus inferred from mitochondrial DNA coalescent, from Iran to Europe. *Molecular Ecology*, **17**, 627–641.

Rambaut A (2000) Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics*, **16**, 395–399.

Rambaut A, Grassly NC (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences*, **13**, 235–238.

Rambaut A, Ho SYW, Drummond AJ *et al.* (2009) Accommodating the effect of ancient DNA damage on inferences of demographic histories. *Molecular Biology and Evolution*, **26**, 245–248.

Rodrigo AG, Felsenstein J (1999) Coalescent approaches to HIV population genetics. In: *Molecular Evolution of HIV* (ed Crandall K), pp. 233–272. Johns Hopkins University Press, Baltimore.

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology & Evolution*, **19**, 101–109.

Shapiro B, Drummond AJ, Rambaut A *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science*, **306**, 1561–1565.

Shapiro B, Ho SYW, Drummond AJ *et al.* (2011) A Bayesian method to estimate unknown sequence ages in a phylogenetic context. *Molecular Biology and Evolution*, **28**, 879–887.

Stewart JB, Freyer C, Elson JL *et al.* (2008) Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biology*, **6**, e10.

Stiller M, Baryshnikov G, Bocherens H *et al.* (2010) Withering away – 25 000 years of genetic decline preceded cave bear extinction. *Molecular Biology and Evolution*, **27**, 975–978.

Strimmer K, Pybus OG (2001) Exploring the demographic history of DNA sequences using the generalized skyline plot. *Molecular Biology and Evolution*, **18**, 2298–2305.

Subramanian S (2009) Temporal trails of natural selection in human mitogenomes. *Molecular Biology and Evolution*, **26**, 715–717.

Subramanian S, Hay JM, Mohandesan E *et al.* (2009) Molecular and morphological evolution in tuatara are decoupled. *Trends in Genetics*, **25**, 16–18.

Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology & Evolution*, **18**, 1001–1013.

Swofford DL 2003 *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology & Evolution*, **15**, 1647–1657.

Weiss G, von Haeseler A (1998) Inference of population history using a likelihood approach. *Genetics*, **149**, 1539–1546.

Williamson S, Orive ME (2002) The genealogy of a sequence subject to purifying selection at multiple sites. *Molecular Biology & Evolution*, **19**, 1376–1384.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data S1 Input files for the simulation analyses carried out in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix

Genealogies with 150 tips were generated under two different demographic scenarios using GENIE 3.0 (Pybus & Rambaut 2002). The two population models are shown graphically in Figs 2a and 3a. In both cases, a generation time of 1 year was assumed so that the *y*-axis of resulting plots is expressed in terms of effective population size.

The demographic function in the first scenario was:

$$N_e(t)\begin{cases} N_0 & \text{if } t<x \\ N_0 e^{-r(t-x)} & \text{otherwise} \end{cases}$$

where $N_0 = 10^6$, $r = 10^{-4}$, and $x = 50\ 000$.

The demographic function for the second scenario was:

$$N_e(t)\begin{cases} N_0 & \text{if } t<x \\ aN_0 & \text{otherwise} \end{cases}$$

where $N_0 = 10^6$, $a = 0.1$, and $x = 50\ 000$.

Sequence evolution was simulated on the two genealogies using Seq-Gen (Rambaut & Grassly 1997). On each genealogy, 150 sequences of 2000 nucleotides were generated, using a mutation rate of 10^{-7} mutations/site/year. The ratio of transitions to transversions was fixed at 5.0, while rates among sites followed a discrete gamma distribution with 6 categories and with a shape parameter of 1.0. For the single-tree skyline-plot methods (classic and generalized skyline plots), ultrametric trees were estimated using the UPGMA method in PAUP* (Swofford 2003). Input files are available as Data S1 (Supporting information.).