To investigate the behavior of the Bayesian skyline plot model, we analyzed two simulated data sets using our MCMC method. We followed Strimmer and Pybus (2001) and performed the following simulations: (1) Coalescent trees were simulated under two demographic models. We are grateful to Korbinian Strimmer for providing us early access to a stimulating manuscript describing related research. In addition, A.J.D. thanks Chris Holmes and Alexandre Pintore for helpful discussions. A.J.D. and B.S. are funded by the Wellcome Trust. A.R. and O.G.P. are funded by the Royal Society. Considerable progress in the field of population genetic inference has been made during the past decade, following parallel increases in computer processing speed and available gene sequence data. Most current methods are based on coalescent theory, a stochastic process that describes how population genetic processes determine the shape of the genealogy of sampled gene sequences. Coalescent-based inference methods enable population genetic parameters to be estimated directly from gene sequence data under a variety of scenarios, including recombination (Griffiths and Marjoram 1996; Kuhner, Yamato, and Felsenstein 2000; Fearnhead and Donnelly 2001), population subdivision (Bahlô and Griffiths 2000; Beerli and Felsenstein 2001), and review size (Kuhner, Yamato, and Felsenstein 1998; Beaumont 1999; Drummond et al. 2002). It should be noted that if $f\mu(\mu)$ is chosen to be a probability density function (i.e., a proper prior), then this second scheme could also be used for contemporaneous sequences to allow uncertainty in the scaling between mutations per site and time. (the product of the effective population size and the generation length in years). The thick solid line is the median estimate, and the book the 95% HPD limits. The thick dashed customer the mean estimate for the four-parameter model used in Pybus et al. (2003) (see text for details). The plot shows a sharp increase in the effective number of infections in the early 20th century, probably caused by viral contamination of injectable antischistosomiasis treatment that was widely used in Egypt from 1920s (see text for details). The demographic history of these data was analyzed by Pybus et al. © The Author 2005. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oupjournals.org (the product of the effective population size and the generation length in radiocarbon years). The thick solid line is the median estimate, and the dashed lines show the 95% HPD limits. The darker underlay is the estimated median and 95% HPD limits of the four-parameter model used in Shapiro et al. (2004) (see text for details). The four-parameter model captures a large part of the underlying demographic signal; however, significant deviation between the two plots is evident in the most recent $\approx 150,000$ radiocarbon years. The Bayesian skyline plot suggests that the bison species went through a severe bottleneck around $10,000$ radiocarbon years ago, coincident with the time when many North American megafaunal species went extinct. For comparison, the time of first human settlement in Alaska and the cold period around the Last Glacial Maximum are also indicated. The past dynamics of HCV in Egypt are therefore of considerable interest. Represents the effective population size within each grouped interval. Effective population size does not change at the end of sample-ended intervals. Thus, the log likelihood of the piecewise demographic model is given by: In equation (7), the vector $\Theta$ contains the parameters of the substitution model (such as transition/transversion ratio $[\kappa]$, shape parameter for gamma-distributed rate across sites $[\alpha]$, and proportion of invariant sites $[p_{inv}]$), and the parameter $\mu$ is a mutation rate that scales the genealogy from units of mutations per site to units of time. In the case of heterochronous sequences, the posterior distribution sampled in our scheme is: In addition to this model, we also introduce a simple smoothing on $\Theta$ which represents our belief that effective population size is autocorrelated through time. The prior distribution we assume in all subsequent simulations and analyses is that, going back in time, each new population size is drawn from an exponential distribution with a mean equal to the previous population size. During preparation of the manuscript, we became aware of a related method (Opgen-Rhein, Fahrmeir, and Strimmer 2005) that uses rjMCMC to estimate smooth demographic functions directly from a single
reconstructed genealogy. This approach, unlike the Bayesian skyline plot, does not take
phylogenetic error into account and is therefore less appropriate for data sets containing limited
genetic variation. However, the use of rjMCMC by Strimmer and coworkers is preferable to our
method of fixing the amount of smoothing a priori (i.e., fixing the number of groupings, m). The
two methods are therefore complementary especially as they use different information as
starting data. A Bayesian skyline plot (m = 15) derived from a sample (n = 191) of mtDNA control
region sequences from ancient bison (Bison cf. priscus) preserved in permafrost in Beringia. The
x axis is in units of radiocarbon years in the past, and the y axis is equal to Here we extend a
previously published MCMC method (Drummond et al. 2002) to sample the parameters of the
Bayesian skyline plot model described above. Our implementation samples both \( \theta \) and \( \Lambda \), but
for computational reasons, we do not use rjMCMC to sample the hyperparameter m. Instead,
we condition all our runs on a fixed value of m because the resulting posterior demographic
function is highly consistent for a range of a priori values of m (data not shown).