Pfeiffer, H., Steighner, R., Fisher, R., Mornstad, H., Yoon, C. L. and Holland, M. M. (1998) Mitochondrial DNA extraction and typing from isolated dentin—experimental evaluation in a Korean population. *Int. J. Leg. Med.*, **111**, 309–313.
Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. and Feldman, M. W. (1999) Population growth of human Y chromosomes: a study of Y-chromosome microsatellites. *Molec. Biol. Evoln*, **16**, 1791–1798.
Relethford, J. (1998) Genetics of modern human origins and diversity. *A. Rev. Anthropol.*, **27**, 1–23.
Roewer, L., Kayser, M., de Knijff, P., Anslinger, K. and Betz, A. (2000) A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forens. Sci. Int.*, **114**, 31–43.
Rubin, D. B. and Schenker, N. (1986) Efficiently simulating the coverage properties of interval estimates. *Appl. Statist.*, **35**, 159–167.
Ruiz Linares, A., Nayar, K., Goldstein, D. B., Hebert, J. M., Seielstad, M. T., Underhill, P. A., Lin, A. A., Feldman, M. W. and Cavalli-Sforza, L. L. (1996) Geographic clustering of human Y-chromosome haplotypes. *Ann. Hum. Genet.*, **60**, 401–408.
Schug, M. D., Wetterstrand, K. A., Gaudette, M. S., Lim, R. H., Hutter, C. M. and Aquadro, C. F. (1998) The distribution and frequency of microsatellite loci on Drosophila Melanogaster. *Molec. Evoln*, **7**, 57–70.
Sherry, S. T., Harpending, H. C., Batzer, M. A. and Stoneking, M. (1997) *Alu* evolution in human populations: using the coalescent to estimate effective population size. *Genetics*, **147**, 1977–1982.
Siguroardottir, S., Helgason, A., Gulcher, J. R., Stefansson, K. and Donnelly, P. (2000) The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet.*, **66**, 1599–1609.
Stephens, M. (2001) Inference under the coalescent. In *Handbook of Statistical Genetics* (eds D. J. Balding, C. Cannings and M. Bishop), ch. 8. Chichester: Wiley.
Stephens, M. and Donnelly, P. (2000) Inference in molecular population genetics (with discussion). *J. R. Statist. Soc.* B, **62**, 605–635.
Sykes, B. (1999) The molecular genetics of European ancestry. *Phil. Trans. R. Soc. Lond.* B, **354**, 131–139.
Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. and Feldman, M. W. (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natn. Acad. Sci. USA*, **97**, 7360–7365.
Tremblay, M. and Vézina, H. (2000) New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *Am. J. Hum. Genet.*, **66**, 651–658.
Tully, G., Bär, W., Brinkmann, B., Carracedo, A., Gill, P., Morling, N., Parson, W. and Schneider, P. (2001) Considerations by the European DNA Profiling (EDNAP) Group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forens. Sci. Int.*, **124**, 83–91.
Tully, G., Sullivan, K. M., Nixon, P., Stones, R. E. and Gill, P. (1996) Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics*, **34**, 107–113.
Weber, J. L. and Wong, C. (1993) Mutation of human short tandem repeats. *Hum. Molec. Genet.*, **2**, 1123–1128.
Weir, B. S. and Cockerham, C. C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
Wilkinson-Herbots, H. M. (1998) Genealogy and subpopulation differentiation under various models of population structure. *J. Math. Biol.*, **37**, 535–585.
Wilson, I. J. and Balding, D. J. (1998) Genealogical inference from microsatellite data. *Genetics*, **150**, 499–510.
Wright, S. (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Discussion on the paper by Wilson, Weale and Balding

Ziheng Yang (*University College London*)

It is my great pleasure to propose the vote of thanks. The past two decades have seen a rapid accumulation of genetics data, and the focus of theoretical population genetics has shifted from forward mathematical modelling to inference from real data. The introduction of the coalescent model, 'a time machine which runs evolution backwards' (Edwards, 1970), and the development of computation-intensive statistical methods, such as Markov chain Monte Carlo (MCMC) and importance sampling, have made it possible to implement likelihood-based inference methods under biologically interesting models. In this paper, the authors describe their flexible MCMC algorithm for Bayes inference, which implements the standard coalescent model as well as models of deterministic population size change and population subdivision. The computer program handles all major types of genetics data, such as DNA sequences, short tandem repeats, single-nucleotide polymorphisms and unique event polymorphisms. It will no doubt become a powerful tool for population genetics analysis.

I would like to draw attention to some related work and to make two comments on the authors' algorithm. The population split model of the authors does not allow gene flow (migration) after the split, and it is quite similar to the model for estimating ancestral population sizes on a species phylogeny by using data from multiple loci. A maximum likelihood method was developed by Takahata *et al.* (1995) for two or three species under the infinite sites model, and Yang (2002) and Rannala and Yang (2003)

implemented MCMC algorithms for Bayes inference for general species trees under finite sites models. Two observations that were made in those studies appear relevant here. First, simple methods based on summary statistics (such as the proportion of loci at which the gene tree does not match the species tree) may not use the information in the data efficiently and can be very misleading. Second, to estimate parameters in a complex model incorporating multiple factors, it seems necessary to combine data from multiple loci.

The first of my comments concerns the robustness of the posterior to the prior and to model assumptions. To some extent, one might argue that the authors' models are overparameterized. The simplest (standard coalescent) model has two parameters N and μ, whereas the likelihood depends on their product (θ) only. The model is not identifiable if uniform rather than gamma priors are used for N and μ. Although specifying priors for both parameters is a natural way of accounting for uncertainties in μ when we are estimating N, I wonder to what extent we are simply obtaining what we put in (both through the prior and through the model that is assumed). Overparameterization may produce slow convergence in the MCMC algorithm and strong correlation between parameters in the posterior, in which case the marginal credibility intervals are not an adequate summary of the joint posterior. The problem should be more serious under more sophisticated models of population growth and structure. I would like to see some comments on those issues. A way forward seems to be a combined analysis of data from multiple loci (nuclear, mitochondrial and Y-chromosome) to estimate the shared parameters such as the population growth rate and splitting times, while accommodating differences between loci (in mutation rate, population size, etc.).

My second comment concerns the proposal algorithm. We must make many quite arbitrary decisions, and it is often unclear which make an efficient algorithm. I am particularly interested in how the authors' data augmentation approach, which averages over ancestral states (at ancestral nodes in the genealogy) during the MCMC run, compares with the alternative approach of calculating that average directly by using the peeling or pruning algorithm (Felsenstein, 1981). The latter computation is linear with the sample size n even though the space of the ancestral states grows exponentially with n. A further saving is achieved if the proposal alters only parts of the genealogical tree, as duplicated computation in the unchanged parts is avoided. The posterior distribution of the ancestral states, if needed, can easily be recovered as well, at least for the root. Let the data be x, the ancestral states be y and the parameters in the model be θ. $p(y|x, \theta)$ is easily calculated from the pruning algorithm (Yang *et al.*, 1995), and

$$p(y \mid x) = \int p(y, \theta \mid x)\, \mathrm{d}\theta = \int f(y \mid x, \theta)\ f(\theta \mid x)\, \mathrm{d}\theta$$

can be calculated by averaging over the MCMC sample. I can see that for microsatellite data (short tandem repeats) the authors' use of ancestral states to generate proposals may increase the acceptance rate and the efficiency of the algorithm. However, this is less straightforward to implement for sequence data, and tying the proposal step (to change trees) with the mutation model or data type might complicate the algorithm. I would be delighted to hear any comments on the relative efficiency of those two strategies.

To conclude, the authors have produced a powerful and versatile program package that can accommodate different types of genetics data under several important population genetic models. I congratulate the authors for this achievement and have great pleasure in proposing the vote of thanks.

David Stephens (*Imperial College London*)

This is a very interesting paper that draws together aspects of many of the previously published methods of analysis and describes the implementation of the authors' own software in the analysis of some common types of DNA sequence data. The authors use extensions of the standard coalescent model and use a Markov chain Monte Carlo (MCMC) algorithm to analyse a variety of human-derived DNA data sets, thus extending the ground breaking paper of Wilson and Balding (1998). The modelling extensions in particular are very important, and the authors demonstrate that their algorithms can cope with the more complex models, albeit at some increased computational burden. The authors are to be congratulated on the major achievement of the development of a robust, accessible and efficient computational package.

I think that several important issues are raised by the paper.

The models

Models for the different data sets, in increasing order of complexity, the genealogical models used, are the standard coalescent (fixed population size, with population growth) or splitting–coalescent (with and without populations growth), the coalescent with population splitting and the coalescent with population

splitting and growth. Thus we have a rich and comprehensive treatment set of (nested) models, within which some key parameters are interpretable across all models. The prior specification for parameters in the models appears, however, to be quite problematic: for example, it is not clear that the prior specifications under the different growth models are sufficiently comparable to allow a straightforward posterior interpretation; contrast, for example, the prior and posterior for the TMRCA value for each model. Is it possible routinely to calibrate the priors for ease of comparison?

The algorithm

The algorithm that is used is an MCMC algorithm implemented through the BATWING package, using standard approaches to such MCMC problems. The authors clearly prefer the augmented likelihood approach, where the MCMC algorithm traverses the joint parameter–coalescent tree space, and it is clear that this offers a broader range of modelling possibilities. Do the authors have an opinion on the inferential and algorithmic performance of their method compared with non-MCMC methods?

Model selection and validation

The results that are presented for the various data sets raise the question of model selection and validation. For example, for Table 2, and given the results presented, what should the geneticist infer about TMRCA for the C96 data? We have the results for a range of models, but which set of results is most appropriate? Can population growth models be verified independently, with some form of time-stamped data?

Overall, for all the data sets, there is little or no discussion of model validity or comparison for the various models proposed; there is now the facility to fit sophisticated population genetics models to such sequence data but no real guide on how to compare the adequacy or otherwise of the utilized models *a posteriori*. When presented with, for example, the wealth of results in Tables 2, 4 or 6, what inference should the practitioner draw? The subjective Bayesian approach is, of course, completely coherent and to be recommended, but this does not remove the need for a thorough investigation of the effect of a range of prior specifications, or a *post hoc* model validation or assessment exercise. Although prior parameters are carefully chosen in each example, often on the basis of genuine prior opinion or historical data, the sensitivity of inference

(a) to the prior specification and
(b) any individual datum sequence

is not really discussed. In addition, the posterior behaviour as the sample size n changes may be of some interest, especially when n is small. Could the authors comment on, for example, the utility of bootstrap resampling—which is common in phylogenetics—or leave-one-out validation, population subsampling etc. to examine the stability of the posterior?

At the moment I am left with the feeling that I have no real idea about which of the models proposed (population structures and prior specifications) best represents the data. Much attention is given to understanding the convergence properties of the MCMC algorithm; I would regard the validity of the inference to be equally important. Current simulation-based Bayesian inference provides several different methods for assessing and comparing the fits of different models. First (approximations to) marginal likelihood quantities, the calculation of which for coalescent models are discussed and described in Stephens and Donnelly (2000) for example, can be used, and the model with the highest marginal likelihood preferred. Secondly, variable dimension MCMC methods can be used to compute posterior model probabilities. Neither method would impose a tremendous computational burden (with a slightly amended algorithm), but, I suspect, would detect any serious inconsistencies in prior or model specification.

The issue of identifiability—which parameters are inferable from the data—is not discussed at any length in the paper, but it is widely acknowledged that some parameters will always be estimated poorly (as they are only technically and not practically identifiable from the data alone). Apart from two parameters that are well known to be aliased (N and μ), are there any other parameters that display a similar strong dependence? This would be detected by inspection of joint posterior sample plots; none are included in the paper, but I assume that the authors have used such plots—if, for example, the TMRCA parameter is strongly correlated with other parameters in the posterior, reporting marginal results is questionable.

The simulation study is an attempt to verify consistency of the posterior, i.e. whether or not Bayesian posterior analysis regularly derives the correct result. The answer seems, generally, to be yes. What may be informative here would be to study performance for varying sample size; presumably (one hopes), in these simulations (taking, for example, $n = 30$ or $n = 120$ for the sample size), the principal reason that the posterior interval does not include the true value of the parameter is that the sample size is quite small. In the absence of technical results describing the asymptotic behaviour of the posterior it may be useful to

see what happens when the sample size is gradually increased beyond the practically realistic values that are selected in the paper.

In summary, I congratulate the authors on their significant contribution; the paper draws together several previously proposed methods of MCMC analysis for DNA sequence data and makes modelling and some algorithmic advances. Geneticists now have an array of models and computational algorithms with which to analyse their data. At this stage, there is little guidance on how to assess whether their inferences about the unobserved parameters of interest are credible in the light of the observed data, and conditional on all aspects of their model specification. Nevertheless, the work will be of considerable use and interest to geneticists and statisticians; I gladly second the vote of thanks to the authors.

The vote of thanks was passed by acclamation.

Kevin J. Dawson (*Rothamsted Research, Harpenden*)

The authors have extended the earlier approach of Wilson and Balding (1998) and Beaumont (1999) to allow for data sets where individuals have been sampled from separate populations and have incorporated the 'phylogeny' of these subpopulations, the 'supertree', as a parameter of the model. The assignment of individuals to contemporary subpopulations is also treated as a parameter of the model, about which we are uncertain. Each branch of the supertree is also associated with an *effective population size*. Migration between branches of the supertree (i.e. between ancestral subpopulations) is not allowed under the present model. I hope that this aspect of reality will be incorporated in the model in the near future.

The effect of selection on the genealogical process at loci which are tightly or loosely linked, or even unlinked, to the targets of selection (Barton, 1998) means that parameters of the genealogical process are no longer strictly determined by demography and should be treated as locus-specific parameters, or at least as parameters specific to tightly linked regions of the genome. It makes sense to combine information across marker loci on the Y-chromosome, as the authors have done, since these completely linked loci share the same gene genealogy. The assumption of complete linkage probably also applies to the mitochondrial genome. In contrast, the gene genealogies at unlinked or loosely linked autosomal (and X-chromosome) genes can be assumed to be statistically independent (unless the sample is drawn from a very small or otherwise closely inbred population). Here we should be much more cautious about assuming common values for effective population sizes across loci. The discrepancy between the inferences based on the Y-chromosome loci and the (autosomal) β-globin locus illustrate this point. Frequentist methods have been developed for identifying loci which have *outlying* genealogical histories (Bowcock *et al.*, 1991; Beaumont and Nichols, 1996; Vitalis *et al.*, 2001). It would be preferable to make these decisions within a Bayesian framework to make full use of the information provided by the data.

I presume that we should take the TMRCA of 29000 years BP for the Y-chromosome, based on the most general model, as being the more reliable estimate. The problem of reconciling this with the much earlier data for the colonization of Australia (between 40000 and 30000 years BP) is intriguing. The solution offered by the authors appears to be a selective sweep at the Y-chromosome, which could have extended worldwide, reaching Australia some time after 29000 years BP. This seems plausible. Have the authors considered how long it might have taken for such a selective sweep on the Y-chromosome to spread worldwide, or whether several geographically more restricted selective sweeps could have been responsible? How committed are they to such a recent TMRCA for the Y-chromosome?

Alexei Drummond (*University of Oxford*) **and Geoff Nicholls** (*University of Auckland*)

In elaborating a Bayesian Metropolis–Hastings Markov chain Monte Carlo (MCMC) framework for coalescent-based inference the authors provide an attractive alternative to both importance sampling (Griffiths and Tavare, 1994) and maximum likelihood MCMC methods (Kuhner *et al.*, 1995). Our comments arise from insight gained from our own published work on coalescent-based Bayesian MCMC kernels (Drummond *et al.*, 2002). The authors used data augmentation of sequences at internal nodes rather than the standard analytical peeling algorithm (Felsenstein, 1981). Data augmentation allows more complicated models of mutation and likelihood calculations are simplified. They studied small data sets, with only 13 variable sites in the H97 data set, leading to a small state space of ancestral sequences to sample. However, for larger more variable data sets, peeling will certainly be preferable, especially if ancestral sequences are nuisance parameters. Roughly speaking, MCMC sampling is slowed by diffuse distributions. When mutation rates are low, the distribution over ancestral sequences on a fixed tree is concentrated on a small set. At high mutation rates the MCMC algorithm must explore a relatively large set of ancestral sequences on each tree. In contrast the work in peeling is fixed. In our studies on temporally spaced leaf

data, we made comparable MCMC programs with and without peeling. Peeling had a clear advantage on large data sets. For sequences on a tree, it is straightforward to implement peeling so that local MCMC tree operations generate likelihood calculations that have $O\{\log(n)\}$ time complexity rather than $O(n)$ (where n is the number of leaves). Therefore, peeling is not as slow per update as might be expected.

A second consideration involves the separation of Θ into effective population size N_e and mutation rate μ. For contemporaneous sequences with no external calibrations of time, N_e and μ are confounded. This is true in the authors' work, and all information about N_e and μ beyond their product derives from the priors. Hence, their posterior of μ is almost identical to their prior (Tables 3, 4 and 7).

Finally, there is often doubt about what our state of knowledge actually is (so which prior we should use) and doubt about how to represent our knowledge mathematically (for high dimensional priors it is easy to write down a prior which, when sampled, produces typical realizations that are dramatically unrepresentative of our prior knowledge). Because of the conflation of N_e and μ, the primary achievement of this study is to give a method for converting prior knowledge about N_e and μ into knowledge about the timing of human origins, with an explicit quantification of uncertainty.

R. C. Griffiths (*Oxford University*)

The Melanesian data set that is considered by the authors is interesting in that it conforms to the infinitely many sites model of mutation with no recombination. The data are then equivalent to an essentially unique gene tree, constructed as a perfect phylogeny, whose vertices are labelled by mutations. Labelling of vertices is unique up to permutations of mutation labels along single edges. There is mathematical detail about the tree nature of this data set and ages of mutations in Griffiths and Tavaré (1999). Questions relating to the stochastic nature of the ancestral gene tree back in time can then be asked and answered by simulating trees back in time conditional on the topology of the gene tree by using computationally intensive methods with a combination of sequential importance sampling, Markov chain Monte Carlo or Bayesian methods. Fig. 6 shows an average gene tree from the Melanesian data using sequential importance sampling on coalescent histories with a proposal distribution of Stephens and Donnelly. The gene trees are drawn to scale with mean ages of mutations and TMRCA calculated as weighted means with likelihood weights on each simulation run. In this method each simulation run is independent. Assuming a 25-year generation time and an effective population size of 20000, the TMRCA estimates in the two trees are 1.08 million and 3.01 million years. Of interest are the mean ages of clades underneath mutations, such as the mutation at site 1358. Fig. 6(a) is constructed with $\theta = 2.55$ and Fig. 6(b) is constructed by assuming that the data are single-nucleotide polymorphism data as an illustration to see the effect. In Fig. 6(b) there is no mutation

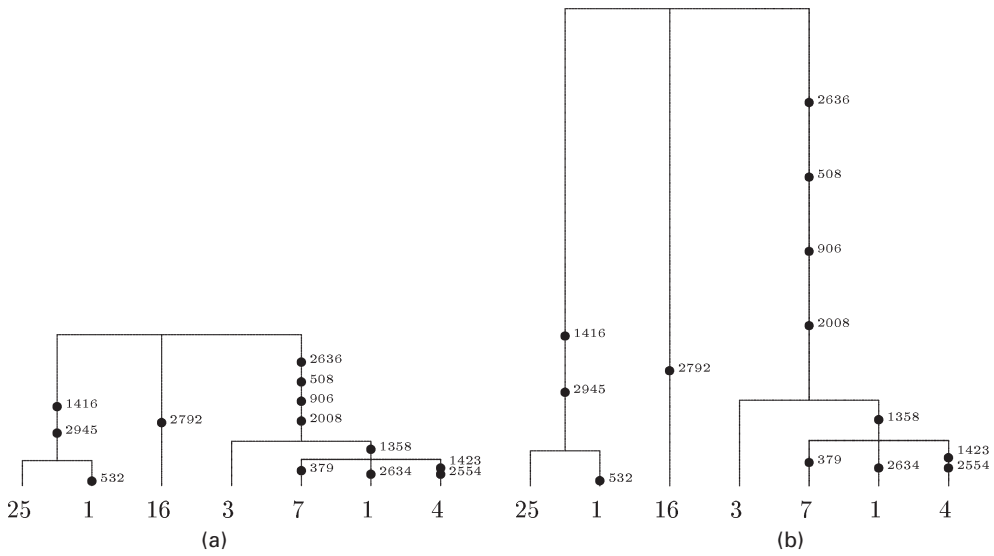

Fig. 6. Melanesian gene trees drawn to scale with expected TMRCA, ages of mutations and ages of clades: in (a) the mutation rate is $\theta = 2.55$ and in (b) the data are assumed to be single-nucleotide polymorphism data and the tree times are computed conditional on the mutant sites segregating, with no assumption about the mutation rate

parameter θ, with the tree being calculated conditional on the segregating sites, supposing that these are the only sites sampled. The tree is approximately three times higher than the tree in Fig. 6(a) because there is no assumption that sites other than those sampled are non-segregating. Lower parts of the tree may be reasonable but an estimate of 3.01 million years seems too high for TMRCA in a biological sense. Although this type of computation can take substantial time, this is a small gene tree and the computing time on a modest personal computer was 260 s with 3 million runs for Fig. 6(a) and 864 s with 10 million runs for Fig. 6(b).

Hilde M. Wilkinson-Herbots (*University College London*)

The interpretation of mitochondrial DNA evidence in court has long been a problem of considerable practical interest, and the method described by Wilson, Weale and Balding is a substantial step forward. However, the mitochondrial DNA 'minisequences' to which their method is applied are mainly useful to exclude suspects quickly and cheaply. Before taking a defendant to court on the basis of mitochondrial DNA evidence, a match for a much longer mitochondrial DNA sequence would normally have been established. For example, the forensic mitochondrial DNA database published by Piercy *et al.* (1993) consists of mitochondrial DNA sequences of approximately 800 nucleotide sites in length. If the authors' method can be readily applied to such 'long' sequences, then their work is of significant practical interest indeed. If however, the analysis of a substantial set of such long sequences proves computationally too demanding at present, then it may be possible to reduce the computational complexity of the problem by taking account of information about the genealogical structure of the mitochondrial DNA gene pool obtained by other methods. Various researchers have studied the major haplogroups that are present in the UK white Caucasian population or the European population as a whole (see for example Wilkinson-Herbots *et al.* (1996) and references therein). Each haplogroup corresponds to a major branch of the genealogical tree and its frequency can be estimated directly from relevant mitochondrial DNA databases. I plan to investigate whether it is possible to use a modified version of the authors' method to estimate the match probability of any individual mitochondrial DNA haplotype, given the frequency of the haplogroup to which it belongs, and focusing primarily on the reduced data set for the haplogroup concerned (although correlations between haplogroups may cause additional difficulties).

Another point where it would be useful to take into account findings from other studies concerns the relative mutation rates of the different nucleotide sites that are included in the minisequence. Whereas part of the polymorphism at the mitochondrial DNA minisequence is due to a few stable, ancient mutations (three of the single-nucleotide polymorphisms characterize major branches of the genealogical tree for the UK white Caucasian mitochondrial gene pool), six of the single-nucleotide polymorphisms included in the minisequence are known to have very high mutation rates (see Wilkinson-Herbots *et al.* (1996) and references therein for evidence at some of these sites). If the authors' method is to be used to evaluate mitochondrial DNA evidence in court cases, then it is important to take this known mutation rate heterogeneity into account, as it may affect the estimates of the match probabilities of uncommon haplotypes.

Mark A. Beaumont (*University of Reading*)

This study provides a significant advance on the original ground breaking paper by Wilson and Balding (1998) and is currently the only approach to allow Bayesian inference of parameters in a model with both population structure and population growth. Methods developed by Wakeley (1999) and Wakeley *et al.* (2001) allow for likelihood-based inference with population structure and growth, and highlight the need to model both aspects jointly. In addition to the effects of population structure there are other phenomena that have the potential to vitiate conclusions drawn about historical changes in population size. These include the effects of ascertainment (where polymorphic loci are deliberately chosen) (Beaumont, 1999; Wakeley *et al.*, 2001), selection at linked sites and the effects of initial population contractions followed by growth (Calmet, 2003). The potential effect of these on inferences from Y-chromosome data are reviewed in Beaumont (2003).

In genealogical modelling there is a general problem of non-identifiability of parameters in the likelihood, which has traditionally been avoided through the use of scaled parameters such as θ. An important innovation in Tavaré *et al.* (1997) and Wilson and Balding (1998) was the use of background information on mutation rates and population sizes to allow for inference on all the parameters of interest. However, it seems to me that only on mutation rates are there grounds to use strongly informative priors. The current and ancestral population sizes and growth rates in the model are unlikely to bear any relation to any estimate of current of historical population size because of their sensitivity to historical metapopulation structure (Wakeley, 2001), the intricate details of which we cannot hope to include directly in our models. In models of population growth, even with proper priors for the mutation rate, improper priors on other

parameters lead to improper posterior distributions (Beaumont, 1999). Therefore, even though we know little about N in the demographic models, it is necessary to impose some limitation on population sizes, and there is a temptation for this to be motivated by the need to obtain good convergence of the Markov chain Monte Carlo algorithm. Yet inferences about changes in population size are sensitive to the priors for N. For example, with the model described in Storz and Beaumont (2002) and Storz *et al.* (2002), if the prior assumes current and ancestral population sizes that are too high the posterior distribution tends to support a model of population growth. This sensitivity could be straightforwardly examined by using several different priors. However, given the difficulties in obtaining convergence, this aspect is probably the most weakly developed part of current Bayesian approaches in population genetics. In conclusion, it seems to me that a period of consolidation is needed in which the sensitivity to model assumptions and specification of the priors is evaluated.

The following contributions were received in writing after the meeting.

Stuart J. E. Baird (*University of California, Berkeley*)

Wilson, Weale and Balding develop a flexible class of Metropolis–Hastings algorithms for drawing inferences about population histories and mutation rates from DNA sequence data. Population structure is generalized to allow splitting of an ancestral population into any number of separate panmictic units. The class of models of population structure that are applicable for inferring human population history is a tiny subset of those that are interesting for making inferences about evolution. Motivated by an interest in broader scale evolutionary inference, inspired by earlier work by two of the authors (Wilson and Balding, 1998), and in consultation with Ian Wilson, a complementary set of algorithms has recently been developed (Baird, 2003) which allows generalization over a wider class of models of population structure. The approach achieves this generality with a trade-off against computation time. A Markov chain Monte Carlo simulation is created with a state consisting of a tree of paths through discrete space and time (Fig. 7). Movement is on a two-dimensional stepping-stone lattice. Between discrete opportunities for movement demes are undisturbed by migration events, and so coalescent probabilities can be described following standard coalescent theory. Proposed transition on the chain state can most succinctly be described as a series of dance steps allowing nodes and paths on the tree to be moved in space and time. The transitions are designed such that change in the tree state is localized, bounded by the nodes connected to the part of the tree being moved and consistent with the stepping-stone paradigm. The process of the Markov chain Monte Carlo simulation can be visualized by iterating the chain and sampling the positions of the lineage paths that make up the tree. Animating the resulting snapshots of the state suggests a label for this approach: the dancing trees algorithm.

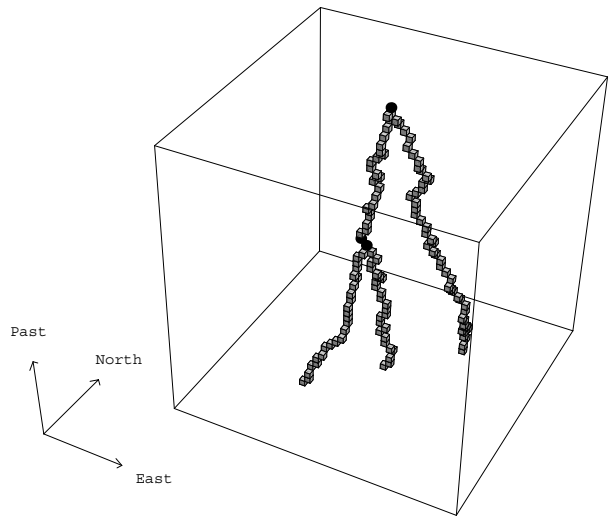

Fig. 7. Example of the explicit state of a genealogy in discrete space–time: cubes represent demes occupied by one lineage; circles represent demes occupied by two lineages

A comparison of inference involving the authors' and dancing trees approaches will be mutually informative. The dancing trees algorithm can be used to define better the set of lineage trees in nature whose history is well approximated by the population splitting model. Conversely this set and others involving islands of panmixis are intrinsically computationally intensive for the dancing trees algorithm. In summary the current work has wide implications: the authors' approach and its complements pave the way towards a sounder understanding of population structure and the evolutionary process.

Martin Lascoux (*Uppsala University*)
In the paper the authors state that 'the present paper is addressed *in part* to statisticians' and, as the paper is being published in the *Journal of the Royal Statistical Society*, this is undoubtedly true. However, I hope that the 'in part' will turn out to be correct as I feel that evolutionary biologists would perhaps benefit most from reading this excellent paper, for which I would like to congratulate the authors. The evolutionary biologists whom I have in mind here are primarily those working with phylogeography, as the coalescent has so far had only a limited influence on this area (at least when humans are not considered). Interestingly, the limited effect that it has had so far can be, at least in part, attributed to precisely the computer program described in the present paper, BATWING, and its predecessor, MICSAT. This is hardly surprising as both programs were tailored for the type of data that are generally produced by phylogeographers, namely variation in non-recombining DNA (chloroplast DNA, mitochondrial DNA, the non-recombining-part of the Y-chromosome). This might also turn out to be one of the main limitations of these programs as only the coalescent analysis of the variation at large numbers of independent loci can lead to more precise inferences for the demographic parameters.

I have two questions for the authors. First, in the light of what has just been said, could their data augmentation approach be extended to include recombination? Second, as shown recently by Ptak and Przeworski (2002), sampling can have an important effect on inferences on the demographic history of species. Interestingly, in his discussion of the paper by Stephens and Donnelly (2000), Ian Wilson has already pointed out that the design and analysis of surveys of different populations have been somewhat neglected by statisticians and geneticists working on the inference of past demographics from molecular variation. Also, in a paper written almost 10 years ago by the first author with N. Barton (Barton and Wilson, 1995) the stage was set for further studies on modelling the coalescent in continuous environments (isolation-by-distance models for instance), i.e. on how to consider jointly the geographical location of individuals and their genotype. How would the authors proceed to find the best sampling strategy to increase our chances of obtaining good estimates of past demographic parameters, given that the sampling strategy depends precisely on knowing something about these parameters? Does this not imply that non-genetics data should explicitly be included in our models?

Raphaël Leblois and Arnaud Estoup (*Centre de Biologie et de Gestion des Populations, Montferrier–Lez*)
We congratulate the authors on their excellent paper. Their methodology represents an important advance towards the goal of fully likelihood-based methods for analysing complex evolutionary scenarios. The treatment of increasingly complex models raises the problem of the validation of methods and programs. Analytical results for the likelihood of a sample of two genes for various population and mutational models can be obtained to check the accuracy of such complicated algorithms (e.g. Nagylaki (1982) and Rousset (1996)). Another important issue is the robustness of algorithms to violations of both the mutation and the demographic assumptions of the model. Simple generation-by-generation coalescence algorithms allow the simulation without approximation of molecular data under virtually any demographic and mutational model and hence can be used to test the robustness and precision of any inferential method.

Because increasingly more models can now be considered, it is crucial to develop criteria for comparing models rather than relying on inferences, from a given model, that fit our beliefs. Did the authors compute the relative likelihood of their four evolutionary models?

The surprisingly low values obtained here for the time since the most recent common ancestor, TMRCA, raise several questions. To what extent could low TMRCA values reflect inappropriate prior assumptions for the mutation rate of microsatellites? More importantly, the possibility of migration between populations is expected to reduce TMRCA substantially as well as time split estimations in a model with no migration. Moreover, since the possibility of homogenizing selection acting on the Y-chromosome may also explain low TMRCA values, it would be worth performing similar analyses on independent and presumably neutral microsatellite loci on autosomal chromosomes.

Finally, in agreement with Fu and Li (1999), our analysis of particularly complex evolutionary histories indicates that, in such cases, inferential methods that are not fully likelihood based still appear to be the

best option available (e.g. Pritchard *et al.* (1999), Estoup *et al.* (2001) and Beaumont *et al.* (2002)). These methods combine the computational convenience of summary statistics with the advantages of the Bayesian paradigm and can handle complex models provided that the simulation of data under the model is feasible. Simulation results have shown that the computational and statistical efficiency of such methods compares favourably with those of the Markov chain Monte Carlo method described here (Beaumont *et al.*, 2002). However, the Markov chain Monte Carlo based method still appears consistently superior to the summary-statistic-based methods, highlighting that it is well worth making the effort to obtain full data inferences if possible.

Rasmus Nielsen (*Cornell University, Ithaca*) **and Jody Hey** (*Rutgers University, Piscataway*)

Several likelihood-based methods for analysing data from multiple populations have been developed in recent years. The methods differ with respect to population genetic assumptions and with respect to the type of data that they are applicable to. Wilson, Weale and Balding have chosen a demographic model of population splitting with no migration between populations. Other likelihood-based methods for estimating parameters in demographic models with population splitting have been proposed by Nielsen *et al.* (1998), Nielsen (1998) and Nielsen and Slatkin (2000). In contrast, the likelihood methods of Beerli and Felsenstein (1999, 2001) and Bahlo and Griffiths (2000) assume infinite divergence times among populations but allow for arbitrary levels of migration between populations. The method of Nielsen and Wakeley (2001) allows for both finite divergence and migration but is only applicable to pairs of populations. The method of Wilson, Weale and Balding also differs from these methods by incorporating population growth. Although all the models naturally are simplifications of the true model, the question arises which of these models is most appropriate. There are undoubtedly organisms in which the authors' model is adequate; however, in the human genetics community there appears to be growing concern that human evolution cannot be described by using models that ignore migration. Evidence of gene flow between human populations has been found at local scales (see for example Papiha *et al.* (1997), Lum *et al.* (2002) and Fix (1999)), across continents (Bandelt *et al.*, 2001; Sokal *et al.*, 1991), as well as between continents (Hammer *et al.*, 1998).

To illustrate the effect of migration, we reanalysed the β-globin data set of Harding *et al.* (1997) for 46 European and 24 Asian individuals, using the method in Nielsen and Wakeley (2001) which incorporates both population splitting and migration. In Fig. 8 we present the marginal posterior distribution of the scaled migration parameter M, assuming uniform priors for all parameters. Note that very little of the probability mass is located around $M = 0$. The marginal posterior distribution for the splitting time T is a strictly increasing function of T (not shown). The data appear to be compatible with a model of equilibrium migration as in Beerli and Felsenstein (1999, 2001) and Bahlo and Griffiths (2000), but not with a model of population splitting without migration. Although we consider the authors' method a great improvement on previous methods for estimating population splitting times in the absence of migration, methods that incorporate migration may be more applicable to the analysis of human genetics data.

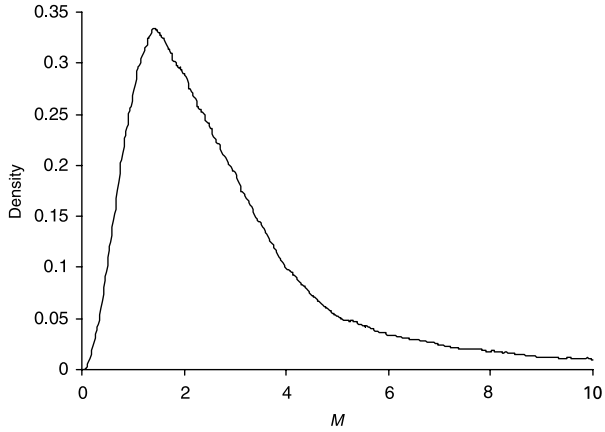

Fig. 8. Marginal posterior distribution for the scaled migration rate M for the β-globin data

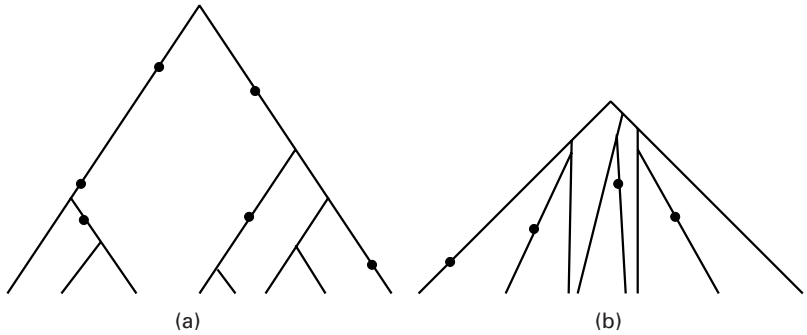

Fig. 9. Genealogy of a sample from a population, assuming (a) constant population size or (b) population growth (note that evolution under constant population size is expected to result in more haplotypes at intermediate frequencies than under a growing population scenario): •, mutations

Michael P. H. Stumpf and Hilde M. Wilkinson-Herbots (*University College London*)
In Section 5.4.2, the authors comment that no appreciable differences in the match probabilities of the mitochondrial DNA minisequences were found when they used the coalescent model with population growth (for which they do not show their results) rather than the standard coalescent model with constant population size. This is surprising, as these two demographic scenarios are expected to lead to different haplotype frequency distributions. Genealogies of populations that have experienced growth tend to be star like. If the mutation rate is not too high compared with the timescale that is involved, this is expected to lead to a very common haplotype and some rare haplotypes, as is illustrated in Fig. 9(b).

In a constant-sized population, by contrast, we expect to obtain more haplotypes at moderate frequencies, giving a different haplotype frequency distribution (see Fig. 9(a)). If a coalescent model with constant population size is used as a prior when estimating match probabilities of haplotypes that actually evolved in an expanding population, we would therefore expect the match probability of the most common haplotype to be underestimated; conversely match probabilities of rare haplotypes may be overestimated. This may at least in part explain the low estimates obtained by the authors (Table 11) for the match probability of the most common haplotype in each of the three ethnic groups considered, compared with the 'naïve' estimate (which is the observed relative frequency of the haplotype)—for a common haplotype we would expect the latter estimate to be reasonably accurate. An inappropriate use of the coalescent model with constant population size might also explain the relatively high estimates that the authors obtained for the match probabilities of the 'similar' and the 'dissimilar' haplotypes that are listed in Table 11 for the Caucasian population, where population growth is believed to have been particularly strong (see also Section 5.2.1.3).

We have verified the above-described effect of population growth on the expected haplotype frequencies by a large number of coalescent simulations under constant population size *versus* population growth (our results are not shown).

The **authors** replied later, in writing, as follows.

We are gratified by the positive and constructive contributions, and we thank all the discussants for their comments. There is considerable overlap of the questions and comments, and we focus on issues raised by several discussants.

More complex demographic models
Geographical structuring of, and migration between, populations is thought to underlie many observed patterns in human DNA data. Lascoux mentions the importance of sampling strategies: our allowance of different population sizes gives some flexibility to incorporate these effects. Nielsen and Hey report evidence of migration in a superset of the H97 data, Leblois and Estoup suggest that ignoring migration may have reduced our TMRCA estimates, and several other contributors mention the desirability of modelling migration within BATWING. We agree. However, the problem of the number of migration parameters rising quadratically with the number of subpopulations (Section 3.1.3) would have been substantial for the 13-subpopulation data set, and an assumption of a common migration parameter would have been suspect. Our splitting model for population structure is unrealistic in some respects but captures some principal aspects of structured data while being computationally relatively unburdensome.

Although computer-intensive methods such as Markov chain Monte Carlo (MCMC) methods permit the analysis of complex models, often generality of model structure is acquired at the expense of efficiency of the algorithm. Inevitably our choice of models was restricted by the computer power that was available at the time: migration would have considerably increased the computational burden. With further increases in computer power, this approach may now be feasible and we shall make efforts to incorporate migration in a future version of BATWING.

Lascoux and Baird want to draw inferences about phylogeography—using the joint information about location and genotype to draw inferences about the processes of evolution. These models are very much more complex. Baird gives an account of his methodology—the coalescent on a two-dimensional grid. He employs auxiliary variables liberally, to try to reduce the computations required for each change in the tree. We commend the approach but note that many auxiliary variables can lead to more problems with mixing.

Yang mentions the similarity of our model to a model for the estimation of ancestral population sizes in closely related species. Our population supertree model is well suited to this application, and it would be interesting to compare results.

Parameterization

Several contributors mention the confounding of effective population size N_e with rate parameters such as μ. This reflects the fact that, if we know only that k events have occurred in an unknown time period, we cannot distinguish a low rate–large time scenario from high rate–small time. For this reason population geneticists have traditionally been limited to working only with $\theta = 2N_e\mu$, a severe limitation since the timescale parameter N_e is required to convert coalescent time units into practically useful units such as generations or years.

Drummond *et al.* (2002) overcame the confounding problem by using time-stamped data, a possibility that is also mentioned by Stephens. Time-stamped data are rarely available for humans, although archaeological evidence can give some help.

BATWING allows users to work either with θ or with N_e and μ separately. As noted by Beaumont, this constitutes a major advance, but it carries the inevitable consequence of sensitivity to the prior: the data are informative about θ, but the 'allocation' of this information between N and μ depends entirely on the prior. Although we agree that there are difficulties with interpreting the available information, there is nevertheless substantial background information about both N_e and μ. Our approach has been to use this information as best we can, making explicit our choice of priors and the evidence on which they are based.

Beaumont and Drummond and Nicholls seem happy with our choice of prior for μ but mention possible problems with priors for N. Our gamma prior is reasonably diffuse, but we agree that it could have some influence on growth rate estimates. We distribute a program with BATWING that allows users to simulate from their prior to explore some of its implications, e.g. about TMRCA.

BATWING works with identifiable parameters internally, so non-identifiable parameters do not cause problems with mixing. Stephens raises the related problem of 'weakly identifiable' parameters: a subset of the parameters such that changes in some members of the subset can be largely compensated (so that the likelihood is almost unchanged) by changes in other members. Weakly identifiable parameters are both sensitive to prior assumptions and potentially problematic for mixing. We highlight one case in the paper: the growth rate and time since the start of growth. Since we work with complex models, involving typically hundreds of parameters, sets of weakly identifiable parameters are practically inevitable, and it is infeasible to diagnose them all. Our approach has been to formulate priors as carefully as possible, and to check mixing as much as we can.

Combining information across loci may help with some cases of weak identifiability. However, Dawson makes the point that selection can affect N_e, and that we should be careful about using the same N_e-values for different loci, even allowing for the difference in the number of chromosomes for Y and nuclear DNA.

Efficiency issues

Stephens asks how our inferential methods compare with other strategies. The importance sampling method of Stephens and Donnelly (2000) works well with one or two linked short tandem repeats but performs less well when there are many linked short tandem repeat loci. We agree with Leblois and Estoup that, because fully likelihood-based methods remain in a phase of development, non-likelihood methods may still be the best option in many settings. The rejection sampling methods that they highlight, briefly described in Section 6, have advantages and there is scope for improving these methods (Beaumont *et al.*, 2002). However, there are no general principles for finding the good summary statistics that are needed, or for assessing the resulting approximation. We also concur with Leblois and Estoup that likelihood-based methods are preferred when available.

Our experience does not accord with Drummond and Nicholls's view that peeling is preferable to the use of auxiliary variables for highly variable data sets. Early versions of BATWING used a peeling algorithm, but the auxiliary variable approach was found to be much better for highly variable short tandem repeat data. We agree that peeling may be preferable for sequence data with low mutation rates, since mixing of auxiliary variables may then be poor. Standard auxiliary variable approaches are unlikely to be efficient for monomorphic sites, but these can be treated separately. We have found that our auxiliary variable approach works well at least for the H97 sequence data set, and computation time is linear in sequence length for both approaches.

Stephens is concerned about model checking, which was not a focus of our paper. We expect that our most general model (splitting with growth) is substantially superior to simpler models that are in widespread use, yet it is still inadequate to capture all important features of the data. We have reported informal model validation via comparison across our models and with the results of other researchers. Nevertheless we agree that quantitative model comparison and assessment are a priority for the future. Under the standard coalescent, increasing the sample size often has little effect of inferences; thus there may be little loss in reserving some data to be used for model testing only, not model fitting.

Mitochondrial DNA match probabilities

Stumpf and Wilkinson-Herbots suggest that a high growth rate should influence the match probability in forensic inference on mitochondrial DNA data. We concur that evidence for growth has been reported from mitochondrial DNA data sets (Excoffier, 2002), and this should indeed have some effect on match probabilities. However, the effect of growth on inference from an observed data set may be much less than its effect on simulations that are not constrained by data.

Wilkinson-Herbots notes that much longer mitochondrial DNA sequences would nowadays be routinely typed, and also that mutation rate heterogeneity is found in such sequences. We explored the effect of a variable mutation rate in a simple way by allowing for different rates at each single-nucleotide polymorphism and also by splitting single-nucleotide polymorphisms into 'high' and 'low' rate categories according to published evidence, but we admit that a more sophisticated rate heterogeneity model would be preferable. It will be worth investigating the proposal of Wilkinson-Herbots to perform BATWING within subclades only (we suggest that fitting an exponential growth model will compensate in part for the different genealogical structure of subclades relative to the whole tree). Because of the relatively high mutation rate of the mitochondrial DNA control region, this would provide an interesting test of the relative merits of peeling algorithms and auxiliary variable approaches.

We do not propose our algorithm for the routine calculation of mitochondrial DNA match probabilities, because of the remaining questions about model validity and because of the computation time that is required for each calculation. Instead, our goals in undertaking the mitochondrial DNA match probability were

(a) to indicate some of the possibilities that genealogical modelling opens and
(b) to check the validity of the naïve estimator against a more sophisticated approach in at least some settings.

We found that, although not providing a bound, the naïve estimator is likely to be adequate in practice for the scenarios that we considered.

References in the discussion

Bahlo, M. and Griffiths, R. C. (2000) Inference from gene trees in a subdivided population. *Theoret. Popln Biol.*, **57**, 79–95.

Baird, S. J. E. (2003) A novel approximation to the spatial coalescent: the dancing trees algorithm . To be published.

Bandelt, H. J., Alves-Silva, J., Guimaraes, P. E., Santos, M. S., Brehm, A., Pereira, L., Coppa, A., Larruga, J. M., Rengo, C., Scozzari, R., Torroni, A., Prata, M. J., Amorim, A., Prado, V. F. and Pena, S. D. (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann. Hum. Genet.*, **65**, 549–563.

Barton, N. H. (1998) The effect of hitch-hiking on neutral genealogies. *Genet. Res.*, **72**, 123–133.

Barton, N. H. and Wilson, I. (1995) Genealogies and geography. *Proc. R. Soc. Lond.* B, **349**, 49–59.

Beaumont, M. A. (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.

Beaumont, M. A. (2003) Recent developments in genetic data analysis: what can they tell us about human demographic history? *Heredity*, to be published.

Beaumont, M. and Nichols, R. A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Genetics*, **153**, 2013–2029.
Beaumont, M. A., Zhang, W. and Balding, D. J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
Beerli, P. and Felsenstein, J. (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.
Beerli, P. and Felsenstein, J. (2001) Maximum-likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc. Natn. Acad. Sci. USA*, **98**, 4563–4568.
Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. K. and Cavalli-Sforza, L. L. (1991) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Genetics*, **88**, 839–843.
Calmet, C. (2003) Inférences sur l'histoire des populations à partir de leur diversité genetique: étude de séquences démographiques de type fondation-explosion. *PhD Thesis*. University of Paris, Paris.
Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. and Solomon, W. (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, **161**, 1307–1320.
Edwards, A. W. F. (1970) Estimation of the branch points of a branching diffusion process (with discussion). *J. R. Statist. Soc.* B, **32**, 155–174.
Estoup, A., Wilson, I. J., Sullivan, C., Cornuet, J.-M. and Moritz, C. (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics*, **159**, 1671–1687.
Excoffier, L. (2002) Human demographic history: refining the recent African origin model. *Curr. Opin. Genet. Devlpmnt*, **12**, 675–682.
Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Molec. Evoln*, **17**, 368–376.
Fix, A. G. (1999) *Migration and Colonization in Human Microevolution*. Cambridge: Cambridge University Press.
Fu, Y. X. and Li, W. H. (1999) Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theoret. Popln Biol.*, **56**, 1–10.
Griffiths, R. C. and Tavaré, S. (1994) Sampling theory for neutral alleles in a varying environment. *Phil. Trans. R. Soc. Lond.* B, **344**, 403–410.
Griffiths, R. C. and Tavaré, S. (1999) The age of mutations in gene trees. *Ann. Appl. Probab.*, **9**, 567–590.
Hammer, M. F., Karafet, T., Rasanayagam, A., Wood, E. T., Altheide, T. K., Jenkins, T., Griffiths, R. C., Templeton, A. R. and Zegura, S. L. (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Molec. Biol. Evoln*, **15**, 427–441.
Harding, R. M., Fullerton, S. M., Griffiths, R. C. and Clegg, J. B. (1997) A gene tree for β-globin sequences from Melanesia. *J. Molec. Evoln*, **44**, suppl. 1, S133–S138.
Kuhner, M. K., Yamato, J. and Felsenstein, J. (1995) Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics*, **140**, 1421–1430.
Lum, J. K., Jorde, L. B. and Schiefenhovel, W. (2002) Affinities among Melanesians, Micronesians, and Polynesians: a neutral biparental genetic perspective. *Hum. Biol.*, **74**, 413–430.
Nagylaki, T. (1982) Geographical invariance in population genetics. *Theoret. Popln Biol.*, **99**, 159–172.
Nielsen, R. (1998) Maximum likelihood estimation of population divergence times and population phylogenies under the infinite sites model. *Theoret. Popln Biol.*, **53**, 143–151.
Nielsen, R., Mountain, J. L., Huelsenbeck, J. P. and Slatkin, M. (1998) Maximum-likelihood estimation of population divergence times and population phylogeny in models without mutation. *Evolution*, **52**, 669–677.
Nielsen, R. and Slatkin, M. (2000) Analysis of population subdivision using di-allelic models. *Evolution*, **54**, 44–50.
Nielsen, R. and Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.
Papiha, S. S., Singh, B. N., Lanchbury, J. S., Mastana, S. S. and Rao, Y. S. (1997) Genetic study of the tribal populations of Andhra Pradesh, south India. *Hum. Biol.*, **69**, 171–199.
Piercy, R., Sullivan, K. M., Benson, N. and Gill, P. (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int. J. Leg. Med.*, **106**, 85–90.
Pritchard, J. K., Seielstad, M. T., Prez-Lezaum, A. and Feldman, M. W. (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molec. Biol. Evoln*, **16**, 1791–1798.
Ptak, S. E. and Przeworski, M. (2002) Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet.*, **18**, 559–563.
Rannala, B. and Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using multi-locus DNA sequences. Submitted to *Genetics*.
Rousset, F. (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.
Sokal, R. R., Oden, N. L. and Wilson, C. (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature*, **351**, 143–145.

Stephens, M. and Donnelly, P. (2000) Inference in molecular population genetics (with discussion). *J. R. Statist. Soc.* B, **62**, 605–655.
Storz, J. F. and Beaumont, M. A. (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution*, **56**, 154–166.
Storz, J. F., Beaumont, M. A. and Alberts, S. C. (2002) Genetic evidence for long-term population decline in a savannah-dwelling primate: inferences from a hierarchical Bayesian model. *Molec. Biol. Evoln*, **19**, 1981–1990.
Takahata, N., Satta, Y. and Klein, J. (1995) Divergence time and population size in the lineage leading to modern humans. *Theoret. Popln Biol.*, **48**, 198–221.
Tavaré, S., Balding, D. J., Griffiths, R. C. and Donnelly, P. (1997) Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505–518.
Vitalis, R., Dawson, K. J. and Boursot, P. (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811–1823.
Wakeley, J. (1999) Nonequilibrium migration in human history. *Genetics*, **153**, 1863–1871.
Wakeley, J. (2001) The coalescent in an island model of population subdivision with variation among demes. *Theoret. Popln Biol.*, **59**, 133–144.
Wakeley, J., Nielsen, R., Liu-Cordero, S. N. and Ardlie, K. (2001) The discovery of single-nucleotide polymorphisms—and inferences about human demographic history. *Am. J. Hum. Genet.*, **69**, 1332–1347.
Wilkinson-Herbots, H. M., Richards, M. B., Forster, P. and Sykes, B. C. (1996) Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations. *Ann. Hum. Genet.*, **60**, 499–508.
Wilson, I. J. and Balding, D. J. (1998) Genealogical inference from microsatellite data. *Genetics*, **150**, 499–510.
Yang, Z. (2002) Likelihood and Bayes estimation of ancestral population sizes in Hominoids using data from multiple loci. *Genetics*, **162**, 1811–1823.
Yang, Z., Kumar, S. and Nei, M. (1995) A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*, **141**, 1641–1650.