Published in final edited form as:

Evolution. 2009 December ; 63(12): 3136–3146. doi:10.1111/j.1558-5646.2009.00788.x.

The Tempo and Mode of Evolution of Transposable Elements as Revealed by Molecular Phylogenies Reconstructed from Mosquito Genomes

Claudio J. Struchiner^{1,5}, Eduardo Massad^{2,6}, Zhijian Tu^{3,7}, and José M. C. Ribeiro^{4,8}

¹ENSP/FIOCRUZ and IMS/UERJ, Av. Brasil, 4365, Rio de Janeiro, Brasil 21040 360

²FM/USP, Av. Dr. Arnaldo, 455, São Paulo, Brasil 01246 903

³Department of Biochemistry, Virginia Polytechnic Institute and State University, 309 Engel Hall, Blacksburg, Virginia 24061

⁴Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, 12735 Twinbrook Parkway, Room 2E32D, National Institutes of Health, Rockville, Maryland 20852

Abstract

Although many mathematical models exist predicting the dynamics of transposable elements, there is a lack of available empirical data to validate these models and inherent assumptions. Genomes can provide a snapshot of several transposable-element families in a single organism, and these could have their demographics inferred by coalescent analysis, allowing for the testing of theories on TE amplification dynamics. Using the available genomes of the mosquitoes *Aedes aegypti* and *Anopheles gambiae*, we indicate that such an approach is feasible. Our analysis follows four steps: (i) mining the two mosquito genomes currently available in search of TE families; (ii) fitting, to selected families found in (i), a phylogeny tree under the general time-reversible (GTR) nucleotide substitution model with an uncorrelated lognormal relaxed clock (UCLN) and a non-parametric demographic model; (iii) fitting a non-parametric coalescent model to the tree generated in (ii); (iv) fitting parametric models motivated by ecological theories to the curve generated in (iii).

Keywords

Demography reconstruction; genomes; population dynamics; transposable element

Transposable elements (TEs) are parasitic DNA sequences that can constitute the majority of the genome in eukaryotes. Class I TEs are retroviral-like elements that transpose by means of a reverse transcriptase, while class II TEs transpose by means of a transposase enzyme. Miniature inverted repeat transposable elements (MITE) are nonautonomous elements that parasitize transposases from autonomous (possibly class II) elements. Other nonautonomous elements, possibly parasitizing class I retrotransposases, also exist. They can insert in different places within a host genome (transposition), changing location and/or increasing copy number during successive generations.

Correspondence to: Claudio J. Struchiner.

⁵stru@fiocruz.br

⁶edmassad@dim.fm.usp.br

⁷jaketu@vt.edu

⁸jribeiro@niaid.nih.gov

TE host invasion is accomplished through a mechanism of self replication that increases the element copy number in the gametes, thus 'cheating' Mendellian segregation law. To the extent that this cheating of the Mendellian segregation law is numerically larger than the host fitness cost, the element succeeds in establishing itself into a new genome (Ribeiro and Kidwell 1994; Silva et al. 2004b). Each event of transposition, however, can antagonistically decrease host fitness while increasing the frequency of TE-bearing gametes. Accordingly, successful invasion is predicted to be associated with TE regulation at the organism level, a function depending on TE genomic copy number, and with total invasion of the host species by the TE at the population level (i.e., every member of the species carries the element). Eventually, mutation of most, or all, element copies leads to degeneration, and only inactive remnants of previous invasions are detected in the form of degenerate sequences of reverse transcriptases, transposases, or terminal repeats (Robertson 1993; Lohe et al. 1995; Clark and Kidwell 1997). Different species vary in their total TE genomic content; this may be due to their refractoriness to TE invasion and/or to the speed at which they can remove TE sequences following TE invasion.

Common to the previous modeling exercises of TE invasion is the lack of empirical data against which the putative models could be fitted and checked. This situation could be remedied, at least in part, by exploring the analytic potential of coalescence models fitted to molecular data (Pybus et al. 2001b; Drummond et al. 2002; Brookfield 2005b; Katzourakis et al. 2005; Brookfield and Johnson 2006; Johnson and Brookfield 2006). Although the initial coalescent models proposed have been restricted to very simple demographic scenarios such as constant size and exponential or logistic growth, more recent models can relax those assumptions and allow for more complex population trajectories (Drummond et al. 2006). The demographic component implied by this approach is of great interest because it can help in identifying the molecular and ecologic approximate conditions under which a transposable element spreads through a host population.

In this context, the interaction between mobile DNA sequences and the organisms in which they reside encompasses a more general framework, referred to in the literature as the 'ecology of the genome' (Kidwell and Lisch 2001; Brookfield 2005a), and can be assessed using a modeling approach analogous to the field of population dynamics. While coalescence has been classically used to study genetic variation of single genes in a population (Marjoram and Tavaré 2006), its methodology has also been successfully used to track RNA virus variation, such as HIV evolution, including virus demographics in single individuals (Rodrigo et al. 1999; Seo et al. 2002; Lemey et al. 2003; Robbins et al. 2003; Paraskevis et al. 2004; Hue et al. 2005; Lemey et al. 2005; Carvajal-Rodriguez et al. 2006; Edwards et al. 2006a; Edwards et al. 2006b; Kuhner 2006), and even to estimate the basic reproductive number (R_0) of hepatitis viruses (Pybus et al. 2001a). Coalescent theory, however, has never been used to attempt reconstruction of TE dynamics using TE sequences found in whole genomes.

We here test the use of coalescent models to predict the demographics of TEs found in mosquito genomic sequences. Our analysis follows four steps: *i*) mining two mosquito genomes in search of TE families; *ii*) fitting, to selected families found in (*i*), a phylogeny tree under the GTR nucleotide substitution model with an UCLN relaxed clock and a nonparametric demographic model; *iii*) fitting a nonparametric coalescent model to the tree generated in (*ii*); and *iv*) fitting parametric models motivated by ecological theories to the curve generated in (*iii*). Results indicate that the predicted amplification trajectories of the elements tested conform to the known biology of their families (explosive amplification of autonomous class II elements, long-term stability of class I, and staggered progression in MITEs) (Silva et al. 2004b), suggesting that the power of coalescent models can be applied to the study of the evolution of TEs found in whole genomes.

Materials and Methods

IDENTIFICATION AND CLASSIFICATION OF TEs

The algorithm used to identify and classify TEs was based on characteristic patterns of local alignments induced by these repeats and is implemented in the software PILER (Edgar and Myers 2005). As input data to PILER, we used the published (release 37, Feb/2006) genomes of the mosquitoes *Anopheles gambiae* (Holt et al. 2002; Mongin et al. 2004) and *Aedes aegypti* (Aedes aegypti Sequencing Project [version 8/1/2005], Broad Institute of Harvard & MIT [www.broad.mit.edu] and The Institute for Genomic Research [www.tigr.org]). The mosquito lines used were inbred, thus minimizing rare variants that could accumulate in terminal branches resembling population growth. The candidate sequences identified by PILER were then annotated by RepeatMasker/Repbase (Jurka et al. 2005) and TEfam (available from <http://tefam.biochem.vt.edu/tefam/>). The sequences were also checked for similarities and relatedness to transposase and other TE-coding sequences deposited in several databases. This material is organized as a spreadsheet database, analogously to AnoXcel (Ribeiro et al. 2004), and available upon request from the authors. In this work, we concentrate our attention on 4 TE families with 19 members or more. The Tc3 (Class II) and Juan (Class I NLTR) families used in this work have been previously described (Shao et al. 2001; Biedler and Tu 2007). We additionally include in our analysis two novel transposons, a novel retrovirus found in *Ae. aegypti* named BossaNova, and a very abundant MITE found in *An. gambiae* (containing 176 members and a palindromic terminal repeat of 18 nucleotides). The online supplementary material makes available the DNA sequences and descriptive statistics of the families that are the object of our attention.

After we applied the above pipeline, we became aware of the work by Smith et al. (Smith et al. 2007). The sequence of steps we took is very similar to theirs, the main differences residing in the way we checked for similarities and relatedness to transposases and other TE coding sequences deposited in the various databases.

SEQUENCE ALIGNMENT

Multiple alignment of the sequences in each family was obtained using the programs MUSCLE (Edgar 2004), BlastAlignP (Belshaw and Katzourakis 2005) and T-Coffee (Notredame et al. 2000). Some of the families, such as Juan, were too numerous to be handled by T-Coffee. Where possible, we compared all three alignments as judged by the ``aln_compare'` option under T-Coffee (data not shown). The final alignments used in this work were obtained with the program MUSCLE with parameters set to yield maximum accuracy.

PHYLOGENY RECONSTRUCTION

The past population dynamics of a TE family, here interpreted as the amplification dynamics, can be inferred from nucleotide sequence data using a population genetic model called the coalescent (Kingman 1982; Griffiths and Tavaré 1994). The coalescent framework requires a demographic model that describes the effective population size through time. Our analysis is based on the reconstruction of a phylogeny tree as a function of a demographic process and a nucleotide substitution model. As an outcome of the analysis, we concurrently obtain estimates of the evolutionary rate, the parameters of the nucleotide substitution model, the phylogeny, and the ancestral population dynamics. We are mostly interested in the latter. The reconstitution of the demographic process followed a nonparametric approach, which can then be used as a first approximation toward identifying an appropriate parametric demographic model (Drummond et al. 2005).

The analyses presented here were performed under a Bayesian framework using a new class of relaxed-clock models able to estimate simultaneously the phylogeny and divergence times.

The method is described in (Drummond et al. 2006) and implemented in the program BEAST. For each alignment considered here, we allowed the application to run for 100,000,000 steps following a discarded burn-in of 20,000,000 steps. The general properties of the chain and convergence of the chain to a stationary distribution were visually inspected using the software Tracer 1.3 (Drummond et al. 2006). Additional options, common to all datasets, include the choice of the GTR model of nucleotide substitution (Rodriguez et al. 1990) with gamma-distributed rate heterogeneity among sites (Yang 1994) and a proportion of invariant sites. We also assumed a Bayesian skyline coalescent prior (Drummond et al. 2005) and fixed the mean substitution rate to 1 in all analyses, resulting in time being in units of substitutions per site.

BAYESIAN NONPARAMETRIC INFERENCE OF DEMOGRAPHIC HISTORY

BEAST outputs a smooth estimate of the demographic function using the Bayesian skyline method (Drummond et al. 2005), which does take the phylogenetic error into account but requires that the amount of smoothing be fixed a priori. A related method developed by (Opgen-Rhein et al. 2005) relaxes this requirement; however, this approach takes the phylogeny tree as fixed and, unlike the Bayesian skyline plot, does not take phylogenetic error into account. The intervals given by this latter approach are expected to underestimate the true interval.

The algorithm described in (Opgen-Rhein et al. 2005) has been implemented as function ``mcmc.popsiz'` as part of package APE (Paradis et al. 2004) and written in the statistical computer language R (R Development Core Team 2007).

PARAMETRIC ECOLOGICAL MODELING

(Katzourakis et al. 2005) have suggested a simple deterministic model with four parameters to describe the dynamics of the human endogenous retrovirus reconstructed from phylogenetics data,

$$\begin{aligned}\frac{dA}{dt} &= bA(1-p) - iA - dA \\ \frac{dI}{dt} &= bAp + iA - dI\end{aligned}\tag{1}$$

with the following notation: A - replication-active elements; I - replication-inactive elements; b - rate of fixation of new elements in the host population; p - proportion of new elements that are replication-inactive as a result of mutations accumulated through time; i - rate of inactivating mutations, per active element, generated by the process of host germline cell division; d - rate of removal from the genome of both active and inactive elements by recombinational deletion. Rates b and d represent, respectively, the fixation of birth and death events in the host population.

This model can be directly fitted to the nonparametric description of the demographic dynamics that was generated as an output of `mcmc.popsiz` (Opgen-Rhein et al. 2005). Model fitting occurs under a Bayesian framework using WinBUGS Markov chain Monte Carlo (MCMC) software (Gilks et al. 1994). This software has the flexibility of a simulation-based approach to estimate parameters of models described as systems of differential equations when used in conjunction with WBDiff (WinBUGS Differential Interface (WBDiff)); available from: www.winbugs-development.org.uk/, an add-on to WinBUGS. Similar to other MCMC approaches, the properties of parameter estimates can be investigated by simulation from a posterior distribution.

The above setup provides a flexible environment in which to investigate alternative models in population ecology. We fitted models that accommodate distinct population growth patterns accounting for population regulation and limiting factors originating from density dependence

mechanisms and interaction between active and inactive elements (Quesneville and Anxolabehere 1998; Le Rouzic and Deceliere 2005; Struchiner et al. 2005; Le Rouzic and Capy 2006). The system of differential equations described next is an example of the kind of mechanisms determining TE dynamics that can be investigated in this setting:

$$\begin{aligned}\frac{dA}{dt} &= bA \left(\frac{1}{(k+I)} \right) - iA - d_A A \\ \frac{dI}{dt} &= iA - d_I I\end{aligned}\tag{2}$$

Model (2) differs from model (1) in three important ways: *i*) we allow for different deletion rates for active (d_A) and inactive (d_I) elements; *ii*) all elements are born active ($p = 0$); and *iii*) the birth rate is inversely proportional to the number of inactive elements (I) and k is the background regulation mechanism of TE replication, i.e., the resultant element birth regulation in the absence of inactive elements. Informal model comparison is obtained via a deviance information criterion (DIC) as described in (Spiegelhalter et al. 2002).

Results

Our goal was to utilize coalescent models to verify whether a statistical evaluation of different mathematical models of TE life histories fitted to the historical demography of particular element families is realistic. Toward this end, we selected four representative samples among the TE families detected using automated de novo repeat identification and classification procedures as implemented in the program PILER (Edgar and Myers 2005). The identification algorithm is based on the signature left by characteristic patterns of local alignments induced by these repeats present on the genome assembly for the African malaria mosquito, *An. gambiae* and for the yellow fever mosquito, *Ae. aegypti*. Our samples contemplate representatives from distinct mechanisms of transposition leading to large copy numbers, including RNA-mediated (Class I) and DNA-mediated (Class II) as well as MITE families (Table 1). Autonomous Class I retrotransposons of the type containing long terminal repeats were not abundantly found in the mosquito genomes searched by our pipeline, having less than 20 copies per genome, and were not further investigated, because coalescent theory predicts that a random sample of size n has a probability to contain the most recent common ancestor (MRCA) according to $(n-1)/(n-1)$, thus a sample of size 19 has 90% and one of 39 a 95% chance to contain the MRCA for that group (Saunders et al. 1984; Nordborg 2001). The actual sequences used in this work as well as a table summarizing the families studied are available from the online supplementary material.

Each of the four datasets was analyzed using the program BEAST with a UCLN relaxed clock (Drummond et al. 2006). For all analyses, the GTR (Rodriguez et al. 1990) model of nucleotide substitution was used with gamma-distributed rate heterogeneity among sites (Yang 1994) and a proportion of invariant sites. A Bayesian skyline coalescent prior (Drummond et al. 2005) was assumed. The final tree topology for each family was selected from the Bayesian posterior distribution using a maximum clade credibility criterion. The target tree, along with the information from a sample of trees produced by BEAST summarizing the posterior probabilities of the nodes in the target tree, the posterior estimates, and 95% highest posterior density (HPD) intervals of the node heights and the rates, are presented as files in nexus format (available online as supplemental material) suitable for visualization with the program FigTree (Rambaut 2006).

Next, we inferred the demographic history from genealogical trees using the method described in (Opgen-Rhein et al. 2005). In this context, inference and model selection is done using reversible jump Markov chain Monte Carlo (rjMCMC). We used as input data to this method the phylogeny reconstructed by BEAST. Figure 1 presents the different profiles under which

the effective population size (Emerson et al. 2001) unfolds in time. Tc3 and BossaNova families (Figs. 1A and D), both representing autonomous elements, exemplify amplification trajectories that rises explosively at first, reaching an equilibrium thereafter. Juan and Mighty families (Figs. 1B and C) have a much slower start up and increase, following a pattern apparently determined by a fluctuating growth rate.

Having collected empirical data on TE dynamics from mosquito genomes and established their demographic history, we can use these data to fit competing copy number growth models, diagnosis model fitting, and estimate model parameters and associated uncertainty in the estimation process. Among several possible models that could be considered, those described (see Materials and Methods) by the systems of differential equations (1) and (2) take into account different mechanisms determining TE dynamics. Both models discriminate between active and inactive elements. Notice that this terminology is distinct from autonomous and nonautonomous elements. For example a nonautonomous element may be active (by means of some transposase) and become inactive when it loses the necessary completeness of its inverted repeats. Under model (1), the growth of TEs is limited solely by the deletion rate common to both inactive and active elements. Under model (2), active and inactive elements carry their own deletion rates. An additional control mechanism dependent on density of inactive elements is also put into action. We fit both models to the demographic histories obtained as above using the WBDiff interface for WINBUGS (see Materials and Methods). Figure 2 shows the observed demographic histories and those predicted by models (1) and (2) for the Tc3 family. Model (2) provides a good fit for the Tc3 trajectory. In particular, model (2) describes well the more recent stable behavior of Tc3 family (Fig. 2B). Model (1) does a good job for most of the Tc3 trajectory but fails to fit its most recent behavior by predicting a downward trend instead of the observed stability. The 95% HPD interval of the posterior parameter estimates for both models fitted to the Tc3 and Mighty families are presented in Table 2.

Figure 2C and D show the trajectories of active and inactive elements for the Tc3 family predicted by each model using the parameter estimates presented in Table 2. The predicted trajectories of Tc3 active and inactive elements (Fig. 2D) are biologically plausible under model (2). Active elements follow a trajectory similar to an epidemic outbreak and are postdated by the inactive elements. The density-dependent mechanism, with inactive elements controlling the growth of active elements, allows for a dynamic equilibrium with both forms coexisting in recent times. Under model (1), the trajectories of active and inactive elements seem to favor the master gene hypothesis (Deininger et al. 1992), where a small number of active elements maintain the family prevalent among the host population. Model (1) also implies a higher turnover (higher birth, inactivation, and deletion rates) of elements than the dynamics implied by model (2); however, these predictions are not compatible with the observed stability in more recent times (Silva et al. 2004b).

Both models fit poorly the observed trajectory for the Mighty MITE family, which lacks transposase (Fig. 3). As expected, this family seems to follow neither the master gene hypothesis nor the simple inactive element-based density-dependent regulatory mechanism implied by model (2). Mighty observed trajectory seems to be the composite outcome of complex mechanisms, either host originated or, more probably, TE originated (Quesneville and Anxolabehere 1998), as the product of TE re-invasions or reactivation and the interaction with transposase acting in trans produced by other families present in the same mosquito host. We have not tested for alternative mechanisms other than those presented here. Also, BossaNova follows a dynamics similar to Tc3 and Juan experiences an amplification pattern similar to Mighty except for the “bumpy” behavior displayed by Mighty. Since we only intended to provide a “proof of concept” that the parametric modeling could be informative, we have not pursued the model fitting exercise to these two additional families.

Discussion

Here, we describe the amplification dynamics of four specific TE families, found in the draft genomes of the dengue and malaria mosquito vectors, using coalescent theory, a population genetic model that describes how the demographic history of a population determines the ancestral relationships of individuals sampled from it. Our approach should be regarded as the best possible approximation given the lack of empirical data generated by experiments explicitly targeting this objective. Are the results presented in this paper sound and can they stand up to closer scrutiny?

The following points should be carefully considered in putting our contribution into perspective: (1) The probability sampling mechanism leading to the data used in our analyses cannot be easily expressed explicitly. The assembled genomes are the product of the DNA sequencing steps applied to a group of inbred but nonclonal individual mosquitoes reared in laboratory cages. Indeed the number of different haplotypes made it difficult to assemble the genome of *An. gambiae* (Holt et al. 2002), and the draft genome of *Ae. aegypti* consists of over one thousand scaffolds; only 35% of the total nucleotide sequence has been assigned to particular chromosome arms (Nene et al. 2007). Many of these scaffolds consists of relatively small pieces of DNA sequences that could not be assembled in great part due to the presence of repetitive DNA, in many cases the result of transposable elements that vary from individual to individual in their chromosomal insertion sites. In the case of the *An. gambiae* genome, these unmapped sequences are collectively joined in the so called 'unknown' chromosome, where many TE sequences are found. Accordingly, the final DNA sequence posted to the database and subject to the analysis may not reflect precisely the actual genomic material of any of the contributing individual mosquitoes. Nevertheless, the available set represents at worst a sample snapshot of the current population of TEs in an organism. As mentioned above, coalescent theory predicts that a random sample of size n has a probability to contain the MRCA according to $(n-1)/(n+1)$ (Saunders et al. 1984; Nordborg 2001), allowing proper use of coalescent methods to investigate TE phylogeny and demographics from these draft genomes, when families of sufficient size are used. (2) The phylogenies reconstructed here require a multiple sequence alignment as input. By specifying the alignment, we are implying that all the nucleotides in a given column are homologous to one another. We make use of automated methods of alignment in two instances in our work: *i*) as the first step in the automated de novo repeat identification and classification procedures, and *ii*) the multiple alignments of the sequences comprising each family and used as input data for the reconstruction of the tree topology. The results of phylogeny programs are sensitive to changes in the alignment, and this source of uncertainty is not accounted for in the analysis. (3) TEs follow a complex 'life cycle' (Silva et al. 2004a). With the passage of time, all TE sequences degenerate, and after a hundred million years or so, they become unrecognizable. The method of TE detection utilized is capable of identifying remnants of previous invasions in the form of degenerate sequences representing intermediate steps of the whole process. Truncation of our analysis to the first steps of TE life cycle is an imposition of the methodology used. (4) Most phylogenetic methods assume that the evolutionary dynamics of the nucleotide sites are independently and identically distributed and that they evolved under the same stationary, reversible, and homogeneous conditions. These assumptions arise from the need to render phylogenetic methods tractable and easy to use, and they are unlikely to be realistic. The model of nucleotide substitution used here accounts for the observation that different sites in a nucleotide sequence may evolve at different rates by modeling rate heterogeneity across sites using a Γ distribution. We used this approach for the whole alignment but did not consider that different parts of the alignment may require different Γ distributions. For example, mutations caused by reverse transcriptase occur at a rate orders of magnitude higher than 'normal' mutations (Lemey et al. 2007), but these should occur in a relatively short period of the element's invasion, thus sharpening the explosive amplification trajectory type. Nor did we consider that variation among some sites is not

independent and that the distribution of variable sites may vary across sequences. (5) Recombination has been proposed as a confounding factor in the reconstruction of phylogenetic trees (Posada and Crandall 2002; Carvajal-Rodriguez et al. 2006). Recombination can occur in adjacent copies of very similar sequences, and recombinant hybrid P elements have indeed been described in *Drosophila* (Svoboda et al. 1995). A recent work analyzing recombination in *Drosophila*'s P elements indicated the recombination rates to be small, leading mostly to deletions or inversions, but also of fusion of truncated elements (Liang and Sved 2009). A logical extension to the work would be to partition the alignment and estimate the evolutionary rates for different TE structures separately. Violation of the assumed stationary, reversible, and homogeneous conditions or lack of additional mechanisms inducing genome heterogeneity, such as recombination, may lead to compositional differences in the aligned sequences and hence to errors in phylogenetic estimates.

The TE copy number trajectories presented here can be interpreted as the composition of four main regulatory mechanisms: *i*) host population regulation, *ii*) host-TE interaction through the impact that TEs might have on host fitness, *iii*) host-originated mechanisms of regulation of TE dynamics, and *iv*) TE self regulation (Quesneville and Anxolabehere 1998; Struchiner et al. 2005). The intricate dynamics between TEs and their host genomes are further complicated by the ability of some TEs to reinvade their original host genome (Silva et al. 2004a) and the demographic dynamics of the vector population. Coalescence exponential models fitted to mitochondrial sequence data in three closely related species of mosquitoes, *Anopheles dirus* species A, C, and D, from Southeast Asia indicate a large population expansion (Walton et al. 2000). Separating the confounding effects of the above mechanisms on the long-term population history of TEs and their hosts can be difficult. Models (1) and (2) used to fit the data are clearly inappropriate to capture all the complexity determined by the various regulatory mechanisms involved and should be regarded as 'null models', i.e., models that function as the starting point against which more elaborated hypotheses can be tested.

We used the most flexible methods currently available to estimate the effective population size from the reconstructed phylogeny trees. These methods do not explicitly allow for the mechanisms of TE activation and inactivation, and the amount of imprecision implied by using coalescent models to estimate the numbers of transposable elements of various types and activities in the past is not known. The approach adopted here serves the purpose to collect the available empirical evidences, albeit imperfect as they are, and inform future simulation exercises on the evolutionary trajectory of those elements. The next step in this research agenda will be paved by revised models that explicitly account for the TE life cycle in the coalescent models. Therefore, estimates of demographic patterns can be distorted by bouts of TE replication and quiescence and caution should be exercised in the analyses before jumping into the conclusions. However, the simulation studies done by Katzourakis et al. (Katzourakis et al. 2005) indicate that the tree patterns can discriminate among different levels of activity implying that general demographic patterns, such as explosive invasions or multiple periods of growth, can still be recovered.

Because reconstructed phylogenies represent time in units of nucleotide substitutions per site, some parameters are estimated as functions of the substitution rate. These parameters can be transformed back into their natural units using the substitution rate of the TE family concerned. We made no attempt to guess TE mutation rate and express parameters and graphs in natural units of time; however, if we assume that the mutation rate in mosquitoes is similar to *Drosophila melanogaster* (Haag-Liautard et al. 2007), near $1e-8$ per generation and assuming 10 generations per year, a rough value for mosquito mutation rate of $1e-7$ per year may be used to translate to time the events shown in Figure 1. The explosive amplifications of TC3 and BossaNova thus might have occurred at 3.3 and 2 million years ago, respectively. To the extent these amplifications occurred immediately following TE invasion of their host genomes, the

figure for the *An. gambiae* TC3 element indicates the invasion occurred much after the separation of the Anopheline and Culicine subfamilies, about 150 million years ago, and thus this invasion should not be reflected in the *Aedes* genome. Similarly, BossaNova invasion of the *Aedes* genome should not have affected *Anopheles*; however, the sibling species of *An. gambiae* should show TC3 invasions, as they might have originated much later at only 6,000 years ago, coincident with human agricultural settlement (Ayala and Coluzzi 2005). The proposed genome sequencing of the sibling species of *An. gambiae* should reveal Tc3 in all subspecies.

The demographic dynamics of a population reconstructed via coalescent theory has proven useful in different contexts in which the premises of this analytic approach cannot be completely fulfilled. Similarly, using the same approach, we inferred, for all TE families in our analysis, biologically plausible dynamics compatible with both common knowledge in the literature and the results of theoretical simulation models. Our results suggest that coalescent theory can also yield approximate TE demography and dynamics using genome sequences, thus allowing estimates of population parameters and supplying the empirical data necessary for TE population model hypothesis testing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank NIAID intramural editor Brenda Rae Marshall for assistance and acknowledge support from the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health.

Because J.M.C.R. is a government employee and this is a government work, the work is in the public domain in the United States. Notwithstanding any other agreements, the NIH reserves the right to provide the work to PubMedCentral for display and use by the public, and PubMedCentral may tag or modify the work consistent with its customary practices. You can establish rights outside of the U.S. subject to a government use license.

Funded in part by a grant to the Regents of the University of California from the Foundation for the National Institutes of Health through the Grand Challenges in Global Health initiative, and by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases. CJS and EM were partially supported by the Brazilian Research Council (CNPq), FAPESP and FAPERJ.

LITERATURE CITED

- Ayala FJ, Coluzzi M. Chromosome speciation: Humans, *Drosophila*, and mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:6535–6542. [PubMed: 15851677]
- Belshaw R, Katzourakis A. BlastAlign: a program that uses blast to align problematic nucleotide sequences. *Bioinformatics (Oxford, England)* 2005;21:122–123.
- Biedler JK, Tu Z. The Juan non-LTR retrotransposon in mosquitoes: genomic impact, vertical transmission and indications of recent and widespread activity. *BMC evolutionary biology* 2007;7:112. [PubMed: 17620143]
- Brookfield JFY. The ecology of the genome - Mobile DNA elements and their hosts. *Nature Reviews Genetics* 2005a;6:128–136.
- Brookfield JFY. Evolutionary forces generating sequence homogeneity and heterogeneity within retrotransposon families. *Cytogenetic and Genome Research* 2005b;110:383–391. [PubMed: 16093690]
- Brookfield JFY, Johnson LJ. The evolution of mobile DNAs: When will transposons create Phylogenies that look as if there is a master gene? *Genetics* 2006;173:1115–1123. [PubMed: 16790583]

- Carvajal-Rodriguez A, Crandall KA, Posada D. Recombination estimation under complex evolutionary models with the coalescent composite-likelihood method. *Molecular biology and evolution* 2006;23:817–827. [PubMed: 16452117]
- Clark JB, Kidwell MG. A phylogenetic perspective on P transposable element evolution in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 1997;94:11428–11433. [PubMed: 9326626]
- Deininger PL, Batzer MA, Hutchison CA, Edgell MH. Master Genes in Mammalian Repetitive DNA Amplification. *Trends in Genetics* 1992;8:307–311. [PubMed: 1365396]
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *Plos Biology* 2006;4:699–710.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 2002;161:1307–1320. [PubMed: 12136032]
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* 2005;22:1185–1192. [PubMed: 15703244]
- Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinformatics* 2004;5:1–19. [PubMed: 14706121]
- Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. *Bioinformatics (Oxford, England)* 2005;21(Suppl 1):i152–158.
- Edwards CT, Holmes EC, Pybus OG, Wilson DJ, Viscidi RP, Abrams EJ, Phillips RE, Drummond AJ. Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. *Genetics* 2006a;174:1441–1453. [PubMed: 16951087]
- Edwards CT, Holmes EC, Wilson DJ, Viscidi RP, Abrams EJ, Phillips RE, Drummond AJ. Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. *BMC evolutionary biology* 2006b;6:28. [PubMed: 16556318]
- Emerson BC, Paradis E, Thebaud C. Revealing the demographic histories of species using DNA sequences. *Trends in Ecology & Evolution* 2001;16:707–716.
- Gilks WR, Thomas A, Spiegelhalter DJ. A Language and Program for Complex Bayesian Modeling. *Statistician* 1994;43:169–177.
- Griffiths RC, Tavaré S. Sampling Theory for Neutral Alleles in a Varying Environment. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 1994;344:403–410.
- Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Charlesworth B, Keightley PD. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 2007;445:82–85. [PubMed: 17203060]
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanagan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 2002;298:129–149. [PubMed: 12364791]
- Hue S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:4425–4429. [PubMed: 15767575]

- Johnson LJ, Brookfield JFY. A test of the master gene hypothesis for interspersed repetitive DNA sequences. *Molecular Biology and Evolution* 2006;23:235–239. [PubMed: 16221895]
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 2005;110:462–467. [PubMed: 16093699]
- Katzourakis A, Rambaut A, Pybus OG. The evolutionary dynamics of endogenous retroviruses. *Trends in Microbiology* 2005;13:463–468. [PubMed: 16109487]
- Kidwell MG, Lisch DR. Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution* 2001;55:1–24. [PubMed: 11263730]
- Kingman JFC. On the Genealogy of Large Populations. *Journal of Applied Probability* 1982;19A:27–43.
- Kirkpatrick M, Slatkin M. Searching for Evolutionary Patterns in the Shape of a Phylogenetic Tree. *Evolution* 1993;47:1171–1181.
- Kuhner MK. Robustness of coalescent estimators to between-lineage mutation rate variation. *Molecular biology and evolution* 2006;23:2355–2360. [PubMed: 16956973]
- Le Rouzic A, Capi P. Population genetics models of competition between transposable element subfamilies. *Genetics* 2006;174:785–793. [PubMed: 16888345]
- Le Rouzic A, Deceliere G. Models of the population genetics of transposable elements. *Genetical Research* 2005;85:171–181. [PubMed: 16174335]
- Lemey P, Pond SLK, Drummond AJ, Pybus OG, Shapiro B, Barroso H, Taveira N, Rambaut A. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *Plos Computational Biology* 2007;3:282–292.
- Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme AM. Tracing the origin and history of the HIV-2 epidemic. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100:6588–6592. [PubMed: 12743376]
- Lemey P, Van Dooren S, Vandamme AM. Evolutionary dynamics of human retroviruses investigated through full-genome scanning. *Molecular biology and evolution* 2005;22:942–951. [PubMed: 15635055]
- Liang X, Sved JA. Properties of re-arranged P elements in *Drosophila melanogaster*. *Heredity* 2009;102:342–348. [PubMed: 19174840]
- Lohe AR, Moriyama EN, Lidholm DA, Hartl DL. Horizontal Transmission, Vertical Inactivation, and Stochastic Loss of Mariner-Like Transposable Elements. *Molecular Biology and Evolution* 1995;12:62–72. [PubMed: 7877497]
- Marjoram P, Tavaré S. Modern computational approaches for analysing molecular genetic variation data. *Nature reviews* 2006;7:759–770.
- Mongin E, Louis C, Holt RA, Birney E, Collins FH. The *Anopheles gambiae* genome: an update. *Trends Parasitol* 2004;20:49–52. [PubMed: 14747013]
- Nene, V.; Wortman, JR.; Lawson, D.; Haas, B.; Kodira, C.; Tu, Z.; Loftus, B.; Xi, Z.; Megy, K.; Grabherr, M.; Ren, Q.; Zdobnov, EM.; Lobo, NF.; Campbell, KS.; Brown, SE.; Bonaldo, MF.; Zhu, J.; Sinkins, SP.; Hogenkamp, DG.; Amedeo, P.; Arensburger, P.; Atkinson, PW.; Bidwell, S.; Biedler, J.; Birney, E.; Bruggner, RV.; Costas, J.; Coy, MR.; Crabtree, J.; Crawford, M.; deBruyn, B.; DeCaprio, D.; Eglmeier, K.; Eisenstadt, E.; El-Dorri, H.; Gelbart, WM.; Gomes, SL.; Hammond, M.; Hannick, LI.; Hogan, JR.; Holmes, MH.; Jaffe, D.; Johnston, JS.; Kennedy, RC.; Koo, H.; Kravitz, S.; Kriventseva, EV.; Kulp, D.; LaButti, K.; Lee, E.; Li, S.; Lovin, DD.; Mao, C.; Mauceli, E.; Menck, CFM.; Miller, JR.; Montgomery, P.; Mori, A.; Nascimento, AL.; Naveira, HF.; Nusbaum, C.; O'Leary, S.; Orvis, J.; Pertea, M.; Quesneville, H.; Reidenbach, KR.; Rogers, Y-H.; Roth, CW.; Schneider, JR.; Schatz, M.; Shumway, M.; Stanke, M.; Stinson, EO.; Tubio, JMC.; VanZee, JP.; Verjovski-Almeida, S.; Werner, D.; White, O.; Wyder, S.; Zeng, Q.; Zhao, Q.; Zhao, Y.; Hill, CA.; Raikhel, AS.; Soares, MB.; Knudson, DL.; Lee, NH.; Galagan, J.; Salzberg, SL.; Paulsen, IT.; Dimopoulos, G.; Collins, FH.; Birren, B.; Fraser-Liggett, CM.; Severson, DW. Genome Sequence of *Aedes aegypti*, a Major Arbovirus Vector. 2007. p. 1718–1723.
- Nordborg, M. Coalescent Theory. In: Balding, DJ.; Bishop, M.; Cannings, C., editors. *Handbook of statistical genetics*. Wiley, Chichester; New York: 2001.
- Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 2000;302:205–217. [PubMed: 10964570]

- Opgen-Rhein R, Fahrmeir L, Strimmer K. Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo. *Bmc Evolutionary Biology* 2005;5. [PubMed: 15651995]
- Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* (Oxford, England) 2004;20:289–290.
- Paraskevis D, Magiorkinis E, Magiorkinis G, Kiosses VG, Lemey P, Vandamme AM, Rambaut A, Hatzakis A. Phylogenetic reconstruction of a known HIV-1 CRF04_cpx transmission network using maximum likelihood and Bayesian methods. *Journal of molecular evolution* 2004;59:709–717. [PubMed: 15693627]
- Posada D, Crandall KA. The effect of recombination on the accuracy of phylogeny estimation. *Journal of molecular evolution* 2002;54:396–402. [PubMed: 11847565]
- Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. *Science* (New York, N.Y) 2001a;292:2323–2325.
- Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. *Science* 2001b;292:2323–2325. [PubMed: 11423661]
- Pybus OG, Harvey PH. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society of London Series B-Biological Sciences* 2000;267:2267–2272.
- Quesneville H, Anxolabehere D. Dynamics of transposable elements in metapopulations: A model of P element invasion in *Drosophila*. *Theoretical Population Biology* 1998;54:175–193. [PubMed: 9733658]
- R Development Core Team. R Foundation for Statistical Computing. Vienna, Austria: 2007. R: A language and environment for statistical computing. [<http://www.R-project.org>]. [ISBN 3-900051-07-0]
- Rambaut, A. FigTree. 2006. FigTree v1.0, Available from <http://beast.bio.ed.ac.uk/FigTree>
- Rambaut, A.; Drummond, A. TreeStat. 2006. TreeStat v1.0, Available from <http://evolve.zoo.ox.ac.uk/treestat/>
- Ribeiro JM, Topalis P, Louis C. Anoxcel: an *Anopheles gambiae* protein database. *Insect Mol Biol* 2004;13:449–457. [PubMed: 15373803]
- Ribeiro JMC, Kidwell MG. Transposable Elements as Population Drive Mechanisms - Specification of Critical Parameter Values. *J Med Entomol* 1994;31:10–16. [PubMed: 8158612]
- Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairoj AS, Brown TM, Salemi M, Vandamme AM, Kalish ML. U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *Journal of virology* 2003;77:6359–6366. [PubMed: 12743293]
- Robertson HM. The Mariner Transposable Element Is Widespread in Insects. *Nature* 1993;362:241–245. [PubMed: 8384700]
- Rodrigo AG, Shpaer EG, Delwart EL, Iversen AK, Gallo MV, Brojatsch J, Hirsch MS, Walker BD, Mullins JI. Coalescent estimates of HIV-1 generation time in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 1999;96:2187–2191. [PubMed: 10051616]
- Rodriguez F, Oliver JL, Marin A, Medina JR. The General Stochastic-Model of Nucleotide Substitution. *Journal of Theoretical Biology* 1990;142:485–501. [PubMed: 2338834]
- Saunders IW, Tavaré S, Watterson GA. On the Genealogy of Nested Subsamples from a Haploid Population. *Advances in Applied Probability* 1984;16:471–491.
- Seo TK, Thorne JL, Hasegawa M, Kishino H. Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach. *Genetics* 2002;160:1283–1293. [PubMed: 11973287]
- Shao H, Qi Y, Tu Z. MsqTc3, a Tc3-like transposon in the yellow fever mosquito *Aedes aegypti*. *Insect molecular biology* 2001;10:421–425. [PubMed: 11881806]
- Silva JC, Loreto EL, Clark JB. Factors that affect the horizontal transfer of transposable elements. *Current Issues in Molecular Biology* 2004a;6:57–71. [PubMed: 14632259]
- Silva JC, Loreto EL, Clark JB. Factors that affect the horizontal transfer of transposable elements. *Current issues in molecular biology* 2004b;6:57–71. [PubMed: 14632259]
- Smith CD, Edgar RC, Yandell MD, Smith DR, Celniker SE, Myers EW, Karpen GH. Improved repeat identification and masking in Dipterans. *Gene* 2007;389:1–9. [PubMed: 17137733]

- Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 2002;64:583–616.
- Struchiner CJ, Kidwell MG, Ribeiro JMC. Population dynamics of transposable elements: Copy number regulation and species invasion requirements. *Journal of Biological Systems* 2005;13:455–475.
- Svoboda YH, Robson MK, Sved JA. P-element-induced male recombination can be produced in *Drosophila melanogaster* by combining end-deficient elements in trans. *Genetics* 1995;139:1601–1610. [PubMed: 7789763]
- Walton C, Handley JM, Tun-Lin W, Collins FH, Harbach RE, Baimai V, Butlin RK. Population structure and population history of *Anopheles dirus* mosquitoes in southeast Asia. *Molecular Biology and Evolution* 2000;17:962–974. [PubMed: 10833203]
- WinBUGS Differential Interface (WBDiff). WBDiff. Available from:
<http://www.winbugs-development.org.uk/>
- Yang ZH. Maximum-Likelihood Phylogenetic Estimation from DNA-Sequences with Variable Rates over Sites - Approximate Methods. *Journal of Molecular Evolution* 1994;39:306–314. [PubMed: 7932792]

Figure 1.

Nonparametric demographic history inferred from the phylogeny tree reconstructed by BEAST for each of the TE families studied (Table 1). Thick lines indicate central estimates (median) and were used to fit the models given by models (1) and (2). Thin lines indicate 95% limits sampled from the posterior distribution by the rjMCMC procedure. The classic skyline plots are depicted by dashed lines. Logarithmic scale is used for the y-axis.

Figure 2.

Parametric trajectories described by models (1) (thin line) and (2) (thick line) fitted to the nonparametric demographic history (dashed lines) reconstructed from Tc3 phylogenies. Logarithmic scale is used for the y-axes in (A), and linear scale is used for the y-axes in (B). (C) and (D), Individual parametric trajectories of active (dot only), inactive (dot and dash) and total (continuous) elements under models (1) and (2) for Tc3 family. Continuous line describing total elements was omitted from (C) for clarity.

Figure 3.

Parametric trajectories described by models (1) (thin line) and (2) (thick line) fitted to the nonparametric demographic history (dashed lines) reconstructed from Mighty phylogenies. Logarithmic scale is used for the y-axes in (A), and linear scale is used for the y-axes in (B). (C) and (D), Individual parametric trajectories of active (dot only), inactive (dot and dash) and total (continuous) elements under models (1) and (2) for Mighty family. Continuous line describing total elements was omitted from (D) for clarity.

Table 1 Descriptive data characterizing the TE sequences analyzed¹

Element	BossaNova	Mighty	Tc3	Juan
Species	<i>Ae. aegypti</i>	<i>An. gambiae</i>	<i>Ae. aegypti</i>	<i>An. aegypti</i>
N	19	162	50	180
Length	933	623	1989	5138
Class	Novel Class II	Novel Class II	Class II	Class I-nonLTR
B ₁ ²		82.60 (77.89 – 87.13)	26.28 (23.80 – 28.62)	89.57 (84.95 – 94.24)
γ ³		-1730.73 (-1950.34 – -1524.25)	-659.62 (-815.61 – -505.58)	-1619.87 (-1934.46 – -1280.44)
Depth ⁴		0.24 (0.21 – 0.28)	0.33 (0.33 – 0.34)	0.22 (0.154 – 0.338)
σ_r ⁵		0.15 (0.04 – 0.23)	3.26 (3.05 – 3.48)	0.37 (0.33 – 0.41)

¹ Statistics were obtained using the program TreeStat (Rambaut and Drummond 2006) from the maximum clade credibility trees inferred under a Bayesian Markov Chain Monte Carlo (MCMC) approach, implemented in the computer program BEAST (Drummond et al. 2006). 95% HPD intervals are shown in parentheses.

² The B₁ statistic measures tree imbalance (Kirkpatrick and Slatkin 1993) by comparing the number of tips falling on either side of each branching point. It depends on tree size and can be compared directly only if the number of elements is identical.

³ The γ statistic measures tree starlikeness (Pybus and Harvey 2000). Starlike trees have short internal branches and long terminal branches.

⁴ Phylogeny depth (genetic distance from root to tips) is given in units of expected nucleotide substitutions per site.

⁵ The coefficient of variation (σ_r) measures clocklikeness. Lower values of σ_r indicate that sequences are evolving in a more clocklike manner (Drummond et al. 2006). Sequence length includes gaps introduced by the alignment process.

Table 2 Posterior summary statistics^{1,2}

	Tc3(M1)³	Tc3(M2)³	Mighty(M1)³	Mighty(M2)³
<i>b</i>	1932.0 (1835.0 – 2041.0)	212.8 (194.7 – 232.9)	2384.0 (2273.0 – 2590.0)	172.0 (143.7 – 223.1)
<i>i</i>	460.9 (420.5 – 518.5)	25.71 (24.26 – 27.29)	212.4 (188.8 – 227.8)	0.27 (0.20 – 0.38)
<i>d_A</i>	7.18 (6.72 – 7.65)	0.13 (0.02 – 0.52)	132.2 (125.2 – 146.6)	5.90 (0.06 – 9.18)
<i>d_I</i>	7.18 (6.72 – 7.65)	15.81 (15.18 – 16.43)	132.2 (125.2 – 146.6)	3.05 (0.66 – 9.94)
<i>P</i>	0.78 (0.75 – 0.80)	-	0.85 (0.84 – 0.85)	-
<i>k</i>	-	0.21 (0.18 – 0.24)	-3.38	(2.92 – 4.20)
DIC	937.3	758.8	2052.1 1719.0	

¹ Obtained from 100,000 MCMC simulations after a burn-in period of 10,000 simulations, for parameter estimates under models described by models 1 (M1) and 2 (M2).

² Estimates obtained using the programs WinBUGS (Gilks et al. 1994) and (WinBUGS Differential Interface (WBDiff)).

³ Numbers in parenthesis represent 95% HPD intervals.

⁴ DIC (deviance information criterion) statistics described in (Spiegelhalter et al. 2002).