ELSEVIER

Molecular dating when rates vary

John J. Welch and Lindell Bromham

Centre for the Study of Evolution, School of Life Sciences, University of Sussex, Falmer, Brighton, UK, BN1 9QG

Molecular-dating techniques potentially enable us to estimate the time of origin of any biological lineage. Such techniques were originally premised on the assumption of a 'molecular clock'; that is, the assumption that genetic change accumulated steadily over time. However, it is becoming increasingly clear that constant rates of molecular evolution might be the exception rather than the rule. Recently, new methods have appeared that enable the incorporation of variable rates into molecular dating. Direct comparisons between these methods are difficult, because they differ in so many respects. However, the assumptions about rate change on which they rely fall into a few broad categories. Improving our understanding of molecular evolution will be an important next step towards evaluating and improving these methods.

Molecular dating

The advent of molecular-dating techniques, which estimate evolutionary timescales from comparisons between DNA or protein sequences, has transformed many areas of biology [1]. Such techniques offer us an insight into the history of lineages with a poor or non-existent fossil record (e.g. [2]). They have also been used to challenge important evolutionary hypotheses successfully (e.g. suggesting a more recent common ancestry between humans and chimps than had been assumed from fossils alone [3,4]). But molecular dates have also been controversial, particularly when they are at odds with other lines of evidence, such as the fossil record (e.g. [5]).

To appreciate why molecular dating has been controversial, consider in more detail what the process involves. All molecular-dating methods convert measures of the genetic distance between sequences into estimates of the time at which the lineages diverged. The genetic distance estimates require topology (the branching tree structure of the relevant lineages) and branch lengths (estimates of the number of substitutions that have occurred in each lineage). To convert these into measures of time, the methods also require one or more externally derived dates, usually based on fossil or biogeographical evidence (typically, these are treated as point estimates, although increasingly, upper or lower bounds are used [6]). Finally, to extrapolate from the 'known' dates to the rest of the tree, molecular dating has commonly relied on the rate constancy assumption (the assumption that molecular evolution occurred at a steady rate over the whole tree). The central role of this assumption is evident from the fact that the term 'molecular clock', which in its broadest sense

refers to the possibility of inferring age from genetic distance, is also commonly used to refer to rate constancy.

Unfortunately, there is increasing evidence that the assumption of rate constancy is often violated, and that the DNA of even closely related species might evolve at different rates [1,7–9]. This lack of rate constancy has been called 'the single most fundamental obstacle to developing an accurate estimate of times of origination' [5].

Here, we review methods for dealing with rate heterogeneity in molecular dating, and pay particular attention to those methods that incorporate multiple rates directly into the estimation. First, however, consider a response to rate heterogeneity that does not require the use of variable-rate dating methods.

Excluding anomalous sequences

A common and reasonable response to rate heterogeneity has been to try to identify those lineages with 'anomalous' rates, or those genes or sites that are most subject to rate variation, and then exclude these from the analysis. Many dating studies that assume rate constancy will be of this type [10]. To identify rate-variable lineages, various tests of rate constancy can be used [7,11–15]. In practice, there are two possible difficulties with the exclusion approach. First, some of the tests used have low power for the kind of data typically used, and so only dramatic departures from rate constancy are likely to be detected [16–20]. This low power has serious implications, because any rate variation

---

Glossary

Bayesian estimation: estimation of the complete probability distribution of molecular rates and divergence dates, given the molecular sequence data. This 'posterior distribution' is proportional to the likelihood weighted by 'prior' distribution(s) that express our beliefs about the dates and rates in the absence of the molecular sequence data under consideration.

Likelihood: probability of observing the sequence data given that it evolved under a particular model with particular parameter values (e.g., with particular rates and dates). Maximum Likelihood methods find the parameters values that yield the highest likelihood for the data and model under consideration.

Molecular clock: in the broad sense, refers to molecular dating, but also refers to the rate constancy assumption.

Molecular dating: estimating the date of divergence of two or more lineages by comparing their DNA (or protein) sequence data.

Nonidentifiability: parameters in a model are said to be nonidentifiable if different sets of parameter values account for the data with equal likelihood. In molecular dating, rates and dates are nonidentifiable if each branch of a phylogeny is assigned its own rate.

Penalized likelihood estimation: penalized likelihood methods, which are closely related to Bayesian methods, weight the likelihood model of branch lengths by a penalty function applying some data-independent constraints to the parameter values.

Rate-smoothing assumption: assumption that rates of molecular evolution changed gradually over the tree. Introduced by Sanderson [6], but similar assumptions underlie many Bayesian implementations.

Rate testing: methods for identifying lineages, or groups of lineages, whose rate of evolution differs significantly from that of the rest of the tree.

Corresponding author: Welch, J.J. (J.J.Welch@sussex.ac.uk).
Available online 23 February 2005

that remains undetected can result in consistently biased date estimates. Second, the exclusion approach is practical only if rate variation is the exception rather than the rule; otherwise, a large proportion of the sequences have to be excluded.

The variable rate (so-called 'relaxed clock') methods were developed in response to these difficulties. Although there are many such methods, they can be classified into a few broad types, as described in the following sections. Table 1 summarizes the methods and the most important differences between them.

Methods using a small number of rates (local molecular clocks)

The first group of variable rate methods involves a relatively straightforward extension of the conventional constant-rate approach. These methods assume that different parts of the phylogeny are characterized by different rates, or 'local molecular clocks'. As long as the number of rates assigned is small, they can be jointly estimated with the divergence times, as is done with a single fixed rate. If the number of rates is too large, however, then the rates and dates become 'nonidentifiable' (i.e. an infinite

number of rate and date combinations are equally probable [14,21,22]). Because the number of rates is limited, these methods face the fundamental difficulty of how and where to place the rate changes.

The 'quartet method' [17,23] is one of the simplest local clock methods. This method combines two pairs of species, each of which has a known date of divergence. A rate can be estimated for each pair, and this enables the date of the divergence between the pairs to be estimated (Figure 1a). Although the rate change is arbitrarily placed at the midpoint root, rate tests are used to exclude quartets in which members of the pairs have significantly different rates.

Although the quartet method neatly avoids problems of topological uncertainty (because groups with undisputed relationships can be chosen), it is difficult to combine estimates from multiple quartets in a meaningful way unless they are phylogenetically independent [23]. Furthermore, the method does not avoid the difficulties associated with the other rate-testing approaches mentioned above (i.e. the low power of tests, and the necessity of excluding large amounts of data).

A related approach that avoids these difficulties is the local molecular clock method of Yoder and Yang [24],

Table 1. Variable-rate molecular-dating methods

Method	Branch lengths[a]	Node age constraints	Rate change[c]	Topology	Statistics[d]	Software, if available	Refs
Quartet method	Substitution model	Two fixed-point calibrations	Local clocks (two rates per quartet)	Quartets	ML	Qdate (http://evolve.zoo.ox.ac.uk)	[17]
Local molecular clocks I	Substitution model	Multiple fixed point calibrations	Local clocks	Fixed	ML	PAML (http://abacus.gene.ucl.ac.uk)	[24]
Local molecular clocks II	Poisson distribution	Bounds on multiple nodes	Local clocks	Fixed	ML	R8s (http://ginger.ucdavis.edu/r8s)	[27]
Local molecular clocks III	Substitution model	Various node age priors	Local clocks	Variable	Bayesian	BEAST (http://evolve.zoo.ox.ac.uk/beast)	
Nonparametric rate smoothing (NPRS)	Fixed estimates	Bounds on multiple nodes	Rate-smoothing penalty function (Quadratic model)	Fixed	Point estimate	R8s (http://ginger.ucdavis.edu/r8s)	[6]
Penalized likelihood rate smoothing	Poisson distribution	Bounds on multiple nodes	Rate-smoothing penalty function (Quadratic model)	Fixed	MPL	R8s (http://ginger.ucdavis.edu/r8s)	[27,31]
Bayesian estimation I	MVN[b]	Node age prior incorporating bounds on multiple nodes	'Rate-smoothing' prior (Stationary Lognormal model)	Fixed	Bayesian	Multidivtime (http://statgen.ncsu.edu/thorne)	[28,32]
Bayesian estimation II	Substitution model	Node age prior with single fixed point calibration	Various rate change priors	Fixed	Bayesian	PhyBayes (http://aix1.uottawa.ca/~sarisbro)	[35,36]
Bayesian estimation III	Substitution model	Various node age priors	Various rate change priors	Variable	Bayesian	BEAST (http://evolve.zoo.ox.ac.uk/beast)	
Huelsenbeck et al.	Substitution model	Single fixed-point calibration	Compound Poisson process incorporating the Gamma Multiplier model	Fixed	Bayesian (rates not estimated)		[30]
Heuristic rate smoothing (AHRS)	Stage 1: MVN Stage 2: substitution model	Multiple fixed-point calibrations	Stage 1: rate smoothing (Stationary Lognormal model) Stage 2: local clocks	Fixed	ML	PAML (http://abacus.gene.ucl.ac.uk)	[44]
Cutler's method	Gaussian distribution	Bounds on multiple nodes	Implicit in the likelihood model	Fixed	ML (single rate estimated)		[45]

[a]See Box 2.
[b]MVN, multivariate normal approximation of likelihood.
[c]See Box 1.
[d]See Box 3: ML, Maximum Likelihood estimate; MPL, Maximum Penalized Likelihood estimate.
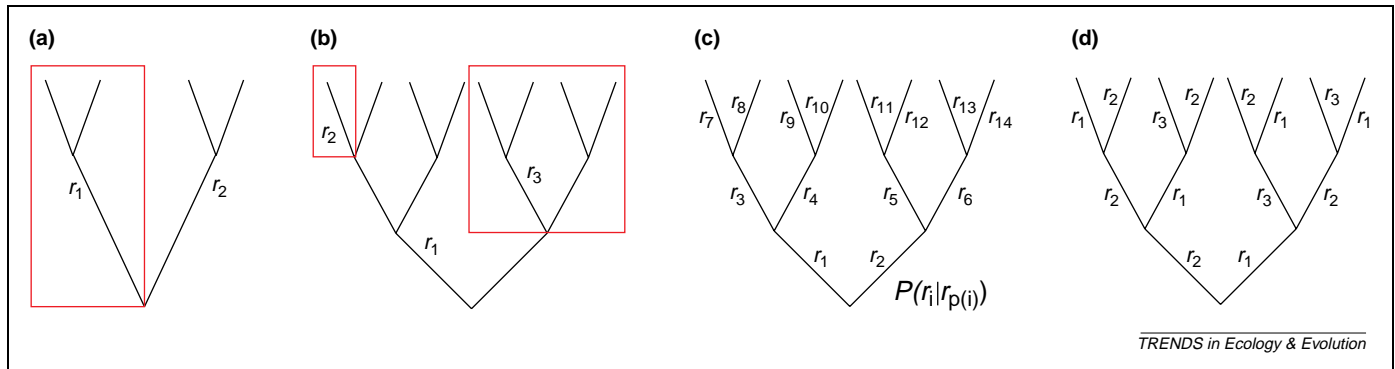
Figure 1. Examples of variable-rate molecular-dating methods. Each of (a–d) shows an example phylogeny to be dated, and indicates where different rates, denoted r_i, have been assigned to different branches. In the quartet method **(a)** [17], trees of four species are used, and both internal nodes must have external calibrations. The date of the basal node is then estimated on the condition that a free-rate model (in which each external branch is assigned its own rate) is not preferred to the two-rate model shown. With the local molecular clock method **(b)** [24], complete phylogenies are used, and placement of the additional rates, here r_2 and r_3, relies on the effective identification of 'anomalous' lineages. Methods allowing many different rates **(c)** [6,31,32,35] must specify prior expectations for the value of each rate, r_i, as a function of the rate of its parental branch $r_{p(i)}$ [in (c), for example, $r_{p(4)} = r_1$]. According to the statistical framework used, these expectations are expressed either as a penalty function, or as a Bayesian prior (Boxes 1 and 3). In Yang's combined method **(d)** [44], the placement of a small number of rate classes (three in the example shown) is determined with the aid of prior expectations, as in (c). The (three) rates are then estimated as in (b).

which builds on the work of Kishino and Hasegawa [25] (see also [26,27]). In this method, a few rate classes are assigned to portions of the whole rooted tree (Figure 1b). The placement of these rate groups relies on the effective identification of anomalous lineages or groups, for example, by rate testing or the use of external (or prior) knowledge [24,25]. Alternatively, rates can be placed after informal examination of branch length estimates obtained without assuming rate constancy. For example, in their investigation of primate speciation dates, Yoder and Yang [24] noticed that their marsupial outgroup taxa had slower rates than did their eutherian ingroup, and so assigned a local clock to each of these two groups. A drawback of this approach, as Yoder and Yang point out, is that using the data to assign rate placements precludes the use of the same data to test formally the adequacy of those placements.

Methods using many rates (rate smoothing)

Whereas local molecular clock methods rely on rate changes being relatively infrequent, other methods have been developed in which the rate can change many times (Figure 1c). In these methods, the placement of (potential) rate changes is fixed in advance, but because so many are assigned, the difficulty of deciding where to place them is largely avoided (there are, however, some relatively minor differences in the placements between implementations [22,28–30]). To avoid problems of nonidentifiability, these methods must rely on strong *a priori* assumptions about the way in which rates change over the tree. Sets of rates that conform to these assumptions are then favoured during the estimation procedure.

Within this broad class of methods, two general approaches can be identified. The first approach was introduced in a pioneering paper by Sanderson [6] where it was termed 'rate smoothing'. Here, the prior assumption about rates is that small changes in rate are more likely than are large changes. This assumption is embodied in a penalty function that is minimized during the estimation (Box 1). In a later implementation [31], the

rate-smoothing function is combined with a likelihood model of branch lengths (Box 2), thus generating a 'penalized likelihood' estimate (Box 3).

The second approach, pioneered by Thorne *et al.* [32], uses Bayesian statistics, a framework in which prior beliefs about parameters are exploited in their estimation [28,32–37]. In Bayesian methods, the prior assumptions about rates are expressed in a formal probability distribution, so patterns of rate change that depart from these assumptions are assigned lower probability values.

These two approaches are based on different schools of statistics, each of which has passionate advocates. However, in terms of their assumptions about rate change, the methods are remarkably similar. Indeed, most Bayesian studies to date have relied on the same basic assumption as rate smoothing, namely that rates change gradually over the tree. This assumption, however, can be embodied in a variety of stochastic models of rate change, some of which are considered in Box 1.

In addition to their similar assumptions, even on purely formal grounds, the Bayesian and penalized likelihood approaches are quite closely related (Box 3). Nonetheless, there are important differences. A benefit of the Bayesian approaches is that standard methods exist for choosing between the different rate change models (although, in practice, these are rarely applied) [36,38,39]. Within the penalized likelihood framework, by contrast, the penalty function tends to be viewed as a mere statistical expediency; as a result, the assumptions about rate change tend to be less closely examined. However, the greater explicitness of the Bayesian approach also brings with it additional problems. In particular, Bayesian statistics requires that prior probabilities be specified for all divergence dates, similar to those specified for the rates (Box 3). In some studies, this 'date prior' seems to have been a major determinant of the date estimates obtained, even though its form cannot be adequately justified in terms of the prior knowledge available [28,39–43]. The influence of the prior diminishes, however, with the amount of sequence data used [33,34,42].

Box 1. Different models of rate change

Dating methods that rely on *a priori* assumptions about rate change use a variety of models. The most common of these are presented in Table I in a common notation. We express the models as penalty functions, as might be used in a penalized likelihood approach [31,44] (the relation to the equivalent Bayesian approach is explained in Box 3). Each of the terms shown is proportional to the penalty applied to a single rate, r, given the rate of its parent branch, r_p. The full penalty function contains a similar term for each rate specified (and, in some cases, normalizing constants that do not depend on r). In addition, the basal rate is dealt with slightly differently by different implementations.

Most of the models in Table I can reasonably be described as rate smoothing, because each penalizes large changes in rate. Some of the differences are due solely to different interpretations of this principle (e.g. the Quadratic model applies least penalty when $r = r_p$, whereas the Lognormal model, expressed as a probability distribution, ensures that $E[\log(r)] = \log(r_p)$ and the Stationary Lognormal model, that $E[r] = r_p$). The models that depart most from the rate-smoothing assumption contain strong directional trends (i.e. they penalize rate increases more or less severely than they penalize rate decreases). For example, under the Exponential model, larger rates will always be more harshly penalized, whereas, under the Ornstein-Uhlenbeck Process, the preferred rate decreases towards zero over time. Neither of these properties has any biological justification [30,39].

The other important difference between the models is that some penalize more harshly rate changes that occur over shorter time periods (i.e. the penalty decreases with t). For example, under the Lognormal model, log rates are assumed to change in a quasi-continuous manner analogous to Brownian motion, whereas under the Quadratic model, the time separating the rate changes is ignored. In general, the time-dependent models seem more plausible; however, time-independent penalties might be reasonable if rate changes were thought to occur primarily at speciation events, and taxon sampling was reasonably complete. (Questions of time dependence apply only when rate assignments are fixed in advance of the estimation [30].)

Finally, although we know of no such method available, a rate-smoothing procedure similar to Cutler's method [45] could penalize deviations from the overall mean rate, rather than from the parental rate.

Table I. Models of rate change

Model name	Function[a]	Refs
Quadratic	$(r - r_p)^2$	[6,31]
Lognormal	$\dfrac{1}{2\beta t}\left[\log(r/r_p)\right]^2 + \log\left(r\sqrt{2\beta t}\right)$	[32,35]
Stationary Lognormal	$\dfrac{1}{2\beta t}\left[\log(r/r_p) + \beta t/2\right]^2 + \log\left(r\sqrt{2\beta t}\right)$	[28,35,44]
Exponential	r/r_p	[35,36]
Ornstein-Uhlenbeck Process	$(r - r_p e^{-\beta t})^2 / \left[t(1 - e^{-2\beta t})/\beta\right]$	[35,36]
Gamma Multiplier	$r/r_p - \beta\log(r/r_p)$	[30]

[a]Notation: r, rate of the branch under consideration; r_p, rate of its parental branch; t, time separating current and parental branch (defined slightly differently in different implementations); and β, an additional parameter.

Finally, the implementations of the methods discussed in this section also differ substantially in their treatment of the branch length estimates. These differences, which are unrelated to the assumptions about rate change (Box 1), and only tangentially related to the statistical methods used (Box 3), are discussed in Box 2.

Combined methods

Rate testing and methods relying on the rate-smoothing assumption are in many ways complementary (one

Box 2. Branch length estimation in molecular dating

Molecular-dating methods differ in their treatment of branch length estimation. Each approach represents a compromise between the competing demands of model accuracy, simplicity and computational tractability.

Some methods implement a full likelihood substitution model [24,35,36]. These models take full account of the uncertainty in estimates, resulting from multiple substitutions at a single site (which can be important even at low divergence [38]) and from the stochastic nature of molecular evolution. In addition, they can distinguish between different kinds of substitution (e.g. transitions and transversions) and incorporate rate variation between sites. Model-testing methods can be used to find the best model for the data [38,62,63]. However, this approach often introduces several additional parameters (a particular problem for Maximum Likelihood methods) and can be computationally intensive.

To avoid these problems, other dating methods do not use the molecular data directly in the estimation, but instead use previously obtained branch length estimates as 'pseudo-data'. These branch length estimates can be obtained via Maximum Likelihood, retaining many of the advantages of the approach described above; however, there remains the question of how to deal with their associated uncertainty.

At one extreme, Sanderson's nonparametric rate smoothing (NPRS; [6]), assumes that branch lengths are known with complete certainty, and attributes all of the differences between sister branches to variation in the rate of evolution. (In this respect, NPRS is the exact reverse of many molecular-dating studies, which assume complete rate constancy, but accept that this can result in many different branch lengths.) By ignoring uncertainty in the branch length estimates, NPRS avoids the thorny problem, faced by other rate-smoothing approaches, of weighting the two sources of uncertainty during the estimation (Box 3 [31,32]).

Other methods of Sanderson [27,31] follow Langley and Fitch [11,15], and use a greatly simplified likelihood function, assuming that the substitution number is a Poisson-distributed random variable, centred on the branch-length estimate. Cutler's method [45] extends this approach by using a Gaussian (normal) distribution instead of a Poisson distribution. (The Gaussian distribution has an arbitrarily large variance, whereas the variance of a Poisson distribution is equal to its mean.) The relation of these simplified models to the full substitution models is not well understood, but it is possible that some of the inflation of variance detected by Cutler [45] and Langley and Fitch [11,15] is due to characteristics of molecular evolution that are neglected by their likelihood models, rather than to rate variation in any stronger sense [31,54,55,64].

Finally, Thorne *et al.* [32,44] introduced a useful compromise between the approaches discussed above. These authors approximate their likelihood substitution model with the use of a multivariate normal distribution, whose means and covariance matrix are estimated from the molecular data.

dealing with rare, large changes in rate, the other with small, frequent changes). As such, a promising approach might be to combine them in some way; for example, by identifying anomalous lineages with rate testing, and then applying rate smoothing to the remaining taxa. The methods of Huelsenbeck *et al.* [30] and Yang [44] can also be viewed as hybrid approaches, combining elements of both rate smoothing and local molecular clocks, and to some extent bridging the gap between them. Both methods present original solutions to the problem of placing rate changes on the tree.

The Bayesian method of Huelsenbeck *et al.* [30] contains a model of rate change that, like most similar methods, penalizes large changes. Uniquely, however, the model enables the number of rate changes to vary during the

---

Box 3. Bayesian statistics and penalized likelihood

The penalized likelihood approach of Sanderson [31] finds the set of divergence dates and molecular rates that maximize a function in the form of Eqn I,

$$\text{Log}(L) - \lambda P, \qquad\qquad \text{[Eqn I]}$$

where L denotes the likelihood of obtaining the sequence data given the proposed rates and dates (Box 2), and P denotes a penalty function (Box 1). In Sanderson's implementation (the Quadratic model of Table I, Box 1), this function embodies the rate-smoothing assumption and so increases as the rates become more heterogeneous. The parameter λ determines the relative importance of the rate smoothing and the uncertainty attached to the branch length estimates. In addition, the dates of some nodes are constrained to lie within a given range (usually determined by fossil evidence). We can include these external constraints explicitly by taking the antilog (or exponential) of Eqn I and writing the result as Eqn II.

$$L \times e^{-\lambda P} \times C, \qquad\qquad \text{[Eqn II]}$$

where $C = 1$ if the proposed dates are within their allowable bounds, and $= 0$ if they are not. When constraints are included, the same rates and dates will maximize Eqn I and II.

Depending on the choice of penalty function and the constraints, we could write an expression such as Eqn II a formal probability distribution. The resulting distribution would describe the probabilities of the rate and date values given the sequence data; in Bayesian statistics, this distribution is called the posterior distribution of rates and dates. Estimating the entire posterior distribution is the goal of a Bayesian analysis. By contrast, the equivalent penalized likelihood approach produces only the highest point, or mode, of the posterior distribution (which is equivalent to a maximum *a posteriori* or MAP estimate). As such, a Bayesian approach provides all the information in a penalized likelihood estimation, and more. In particular, confidence intervals for the date estimates are easy to obtain from the width of the posterior distribution.

Crucially, however, a Bayesian approach is only possible if Eqn II can, in fact, be expressed as a proper probability distribution (i.e. a distribution that can be normalized to unity). To guarantee this, $e^{-\lambda P}$ and C must both be expressible as probability distributions (the likelihood, by definition, is always such a distribution). In this case, they are referred to as the prior distributions of molecular rates ($e^{-\lambda P}$) and divergence dates (C), and express our beliefs about these parameters in the absence of the molecular data. Specifying a prior distribution of divergence dates has proved especially difficult for some Bayesian implementations [28,36,39,43]. For example, Sanderson's constraints, as described above, cannot be expressed in this way. (This is because some nodes can be legitimately placed at an indefinite age, and so no normalizing constant can be found.)

In addition to those mentioned above, there are other differences between Bayesian and likelihood methods. Most important are the fundamentally different approaches to 'nuisance parameters' (such as the parameter λ), and the different numerical methods required; both of these are well explained elsewhere [31,33,34,44].

---

estimation, with departures from the expected number of changes penalized. In addition, and again uniquely, these changes can occur anywhere on the tree, not just at a set of predetermined points (such as the internal nodes) [29]. Because there are two distinct kinds of constraint on the rates, with penalties applied to the number of rate changes and to their magnitude, one or other can be given greater prominence. To test the sensitivity of the date estimates to the different weightings, Huelsenbeck *et al.* [30] applied their method to a single data set (complete mitochondrial sequences from 23 mammalian species), while making different prior assumptions about the number of rate changes. They showed that the same data could be interpreted as containing many changes of small effect (the

assumption governing the rate-smoothing approaches) or as containing a few changes of large effect (the assumption behind the local molecular clocks approaches). Crucially, they showed that the date estimates for most nodes varied little in the two cases, although, for a few nodes, estimates differed by as much as 15%.

Although Huelsenbeck *et al.*'s method does assign different rates to different parts of the tree, these rates cannot be directly estimated along with the dates. This is because the number of rate classes is allowed to vary, so each date estimate is associated with many different rate-change scenarios. For this reason, Huelsenbeck *et al.*'s [30] method provides estimates of the 'hyperparameters' that specify the model of rate change, rather than estimating the rates themselves.

Yang's heuristic rate-smoothing method [44] consists of two quite separate stages. The first stage is '*ad hoc* rate smoothing' (AHRS), a penalized likelihood estimation using Kishino *et al.*'s Stationary Lognormal model [28] (Box 1). This stage yields a distinct rate estimate for each branch of the tree. Using these estimates, each branch is assigned to one of a small number of rate groups (branches with similar rates being placed in the same group), after which the rate and date estimates from the first stage are discarded. The second stage is a conventional local molecular clock estimation [24] using the rate group placements determined by the first stage (Figure 1d).

Cutler's method

A quite distinctive approach to incorporating rate variation was developed by Cutler [45]. As this author points out, the observation that sister branches have widely variable numbers of substitutions could have two distinct explanations: (i) that different rates of molecular evolution characterize the lineages; or (ii) that the process of molecular evolution is identical in all lineages, but simply has a high variance (owing, perhaps, to substitutions tending to cluster in time); if this were the case, random sampling alone might have produced the observed variation. Furthermore, possibilities (i) and (ii) might be fundamentally indistinguishable in some cases. Unlike all the methods discussed so far, which are premised on assumption (i), Cutler's method is premised on assumption (ii), so that all lineages are assigned the same basic evolutionary rate, but the process can be highly variable such that rapid bursts of substitutions might occur on some lineages (Box 2). In this way, rate variation is implicit in the way likelihood values are assigned to the branch lengths.

Cutler's method resembles rate smoothing in that departures from rate constancy are (in effect) penalized during the estimation. However, unlike all of the models in Box 1, Cutler's method does not assume that bursts of substitutions are most likely to occur on closely related lineages or, alternatively, that rapidly evolving lineages are most likely to give rise to other rapidly evolving lineages. Rather, the method penalizes departures from the overall mean rate of the tree, regardless of the smoothness with which the changes take place.

Comparing the methods

As is well known, in some cases, different molecular-dating studies have produced wildly different date estimates

for the same evolutionary event (e.g. [5,43,46]). Perhaps the most notorious example has been the origin of the major metazoan phyla [5,47], where molecular date estimates have ranged from ∼575 million years ago (Mya) [36,48] to >1200 Mya [49]. But it is not obvious to what extent these differences are due to the different methods or to the different data used (the studies use different species, sequences and fossil calibrations) [42,46,50]. To disentangle these factors, an increasing number of studies have compared date estimates obtained from the same data but with the use of different dating methods [26,28,30,35,36,41,51].

Most such studies contain variable rate and constant rate estimates and, in almost every case, conclude that relaxing the assumption of rate constancy has a dramatic influence on the results [28,35,36,51]. Although this conclusion is unequivocal and important, it does not address the reliability of most published constant-rate studies. This is because, unlike the methodological comparisons, most constant-rate studies will rate test their data and exclude certain lineages [10,48].

No such caveat applies to comparisons between different variable rate methods. Here, results are more mixed, with some methods yielding similar estimates, and some not [17,26,35,41]. Explaining the differences that remain is complicated, because the dating methods differ not only in the way in which they deal with rate variation, but also in their treatment of the other aspects of molecular dating, such as branch length estimation, the use of external fossil constraints and the statistics reported. A few of these differences are necessary corollaries of the assumptions made about evolutionary rates, but most are more-or-less arbitrarily associated with different implementations (Table 1).

Prior knowledge of rate change: beyond rate smoothing?

Although empirical comparisons between the methods can be difficult to interpret, we can still ask about the *a priori* reasonableness of the rate change assumptions on which the methods rely. This task is simplified by realizing that these assumptions fall into two broad categories. Several methods rely on rate changes being uncommon, but potentially large (so that lineages departing from rate constancy can be detected and excluded, or assigned their own local molecular clock). Other methods assume that rate changes are frequent, but small. Which of these is the most likely to reflect biological reality? The answer depends to some extent on the kind and cause of the rate variation being considered.

Some rate variation might be due to the 'sloppiness' of the molecular clock resulting from widespread noise in substitution rates (owing perhaps to correlated changes in paired nucleotides, positive selection or demographic fluctuations) [1]. Rate-smoothing approaches (in the broad sense), which model rate change as a random walk, might be best placed to deal with variation of this kind. By contrast, local molecular clocks might be better suited to situations where lineages have substantial and persistent differences in rates of molecular evolution [1,52–55].

Unfortunately, to move beyond these fairly innocuous observations, we need to increase our understanding of molecular evolution. For example, many studies have demonstrated that murid rodents have consistently faster substitution rates than do hominids [7]. But extrapolating from such results to other lineages is complicated because so many factors might be responsible for the difference [56,57]. Explanations that have been suggested include differences in DNA repair efficiency (rats and mice lack some of the excision repair pathways that are found in primates), generation time (the mouse germ line might be copied 100 times more often than is the human genome per unit of evolutionary time, and so accumulate more DNA copy errors), metabolic rate (a cell in a mouse burns more energy than does a human cell and thus might accumulate more DNA-damaging metabolites), effective population size (following the predictions of the nearly neutral theory [58]), or other factors. Comparative studies have identified species characteristics that are correlated with variation in rate of evolution, such as body size, generation time, metabolic rate and fecundity [56,57,59]. Substitution rate has been also been found to be positively related to diversification rate, suggesting that tree shape itself might influence rates [60,61].

Although this long (and probably incomplete) list of factors that can influence rates might lead to pessimism about the possibility of accurate molecular dating, if correlations between traits and the rate of molecular evolution are found to be common and reasonably robust, then they might be used, as prior information, to constrain date estimates. Using this prior information would then involve specifying how the relevant traits changed over the tree. However, we currently lack information that could be used in this way with sufficient confidence. The problems will be even more acute in a Bayesian framework in which our prior knowledge (messy as it is) must be corralled into formal probability distributions (Box 3).

Conclusion

Molecular-dating methods are a boon to evolutionary biology because they can be applied to all species, areas and time periods, and can give a glimpse into the past of lineages for which crucial historical information is missing. But molecular dating is greatly complicated by variation in the rate of molecular evolution between lineages. Nevertheless, a variety of methods are available for molecular dating in the presence of rate variation; many of these have been implemented in freely available software packages and are becoming widely used. Comparing the methods is difficult, because of the many differences between implementations. If future studies confirm initial findings that different methods can lead to very different date estimates, it will be crucial to discriminate between their different assumptions about rate change, and if necessary, to develop new, and more realistic models. An important part of this process will be investigating the causes and dynamics molecular rate variation.

References

1 Bromham, L. and Penny, D. (2003) The modern molecular clock. *Nat. Rev. Genet.* 4, 216–224
2 Holmes, E.C. (2003) Molecular clocks and the puzzle of RNA virus origins. *J. Virol.* 77, 3893–3897
3 Sarich, V.M. and Wilson, A.C. (1967) Immunological time scale for hominid evolution. *Science* 158, 1200–1203
4 Goodman, M. *et al.* (1998) Towards a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* 9, 585–598
5 Smith, A.B. and Peterson, K.J. (2002) Dating the time of origin of major clades: molecular clocks and the fossil record. *Annu. Rev. Earth Planet. Sci.* 30, 65–88
6 Sanderson, M.J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14, 1218–1231
7 Wu, C.I. and Li, W.H. (1985) Evidence for higher rates of nucleotide substitutions in rodents than in man. *Proc. Natl. Acad. Sci. U. S. A.* 82, 1741–1745
8 Woolfit, M. and Bromham, L. (2003) Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol. Biol. Evol.* 20, 1545–1555
9 Davies, T.J. *et al.* (2004) Environmental energy and evolutionary rates in flowering plants. *Proc. R. Soc. Lond. B Biol. Sci.* 271, 2195–2200
10 Hedges, S.B. and Kumar, S. (2003) Genomic clocks and evolutionary timescales. *Trends Genet.* 19, 200–206
11 Langley, C.H. and Fitch, W.M. (1973) The constancy of evolution: a statistical analysis of *a* and *b* haemoglobins, cytochrome *c*, and fibrinopeptide *A*. In *Genetic Structure of Populations* (Morton, N.E., ed.), pp. 246–262, University of Hawaii Press
12 Sarich, V.M. and Wilson, A.C. (1973) Generation time and genomic evolution in primates. *Science* 179, 1144–1147
13 Takezaki, N. *et al.* (1995) Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* 12, 823–833
14 Felsenstein, J. (2004) *Inferring Phylogenies*, Sinauer Associates
15 Langley, C.H. and Fitch, W.M. (1974) An estimation of the constancy of the rate of molecular evolution. *J. Mol. Evol.* 3, 161–177
16 Tajima, F. (1993) Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135, 599–607
17 Rambaut, A. and Bromham, L. (1998) Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* 15, 442–448
18 Bromham, L. *et al.* (2000) The power of relative rates tests depends on the data. *J. Mol. Evol.* 50, 296–301
19 Robinson, M. *et al.* (1998) Sensitivity of the relative-rate test to taxonomic sampling. *Mol. Biol. Evol.* 15, 1091–1098
20 Scherer, S. (1989) The relative-rate test of the molecular clock hypothesis: a note of caution. *Mol. Biol. Evol.* 6, 436–441
21 Chang, J.T. (1996) Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. *Math. Biosci.* 137, 51–73
22 Rannala, B. (2002) Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* 51, 754–760
23 Bromham, L. *et al.* (1998) Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12386–12389
24 Yoder, A.D. and Yang, Z. (2000) Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* 17, 1081–1090
25 Kishino, H. and Hasegawa, M. (1990) Converting distance to time: application to human evolution. *Methods Enzymol.* 183, 550–570
26 Yang, Z. and Yoder, A.D. (2003) Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Syst. Biol.* 52, 705–716
27 Sanderson, M.J. (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19, 301–302
28 Kishino, H. *et al.* (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* 18, 352–361
29 Penny, D. *et al.* (1998) Estimating times of divergence with a change of rate: the orang-utan/African ape divergence. *Mol. Biol. Evol.* 15, 608–610
30 Huelsenbeck, J.P. *et al.* (2000) A compound Poisson process for relaxing the molecular clock. *Genetics* 154, 1879–1892

31 Sanderson, M.J. (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109
32 Thorne, J.L. *et al.* (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15, 1647–1657
33 Holder, M. and Lewis, P.O. (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* 4, 275–284
34 Huelsenbeck, J.P. *et al.* (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51, 673–688
35 Aris-Brosou, S. and Yang, Z. (2002) Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. *Syst. Biol.* 51, 703–714
36 Aris-Brosou, S. and Yang, Z. (2003) Bayesian models of episodic evolution support a late Precambrian explosive diversification of the metazoa. *Mol. Biol. Evol.* 20, 1947–1954
37 Thorne, J.L. and Kishino, H. (2002) Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51, 689–702
38 Arbogast, B.S. *et al.* (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu. Rev. Ecol. Syst.* 33, 707–740
39 Welch, J.J. *et al.* Molecular dates for the "Cambrian explosion": the influence of prior assumptions. *Syst. Biol.* (in press)
40 Yoder, A.D. and Yang, Z. (2004) Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol. Ecol.* 13, 757–773
41 Pérez-Losada, M. *et al.* (2004) Unraveling the evolutionary radiation of the thoracican barnacles using molecular and morphological evidence: a comparison of several divergence time estimation approaches. *Syst. Biol.* 53, 244–264
42 Douzery, E.J.P. *et al.* (2004) The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. U. S. A.* 101, 15386–15391
43 Sanderson, M.J. *et al.* (2004) Molecular evidence on plant divergence times. *Am. J. Bot.* 91, 1656–1665
44 Yang, Z. (2004) A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zool. Sin.* 50, 645–656
45 Cutler, D.J. (2000) Estimating divergence times in the presence of an overdispersed molecular clock. *Mol. Biol. Evol.* 17, 1647–1660
46 Sanderson, M.J. and Doyle, J.A. (2001) Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. *Am. J. Bot.* 88, 1499–1516
47 Fortey, R.A. *et al.* (2003) Phylogenetic fuses and evolutionary 'explosions': conflicting evidence and critical tests. In *Telling the Evolutionary Time: Molecular Clocks and the Fossil Record* (Donoghue, P.C.J. and Smith, M.P., eds), pp. 41–65, CRC Press
48 Peterson, K.J. *et al.* (2004) Estimating metazoan divergence times with a molecular clock. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6536–6541
49 Wray, G.A. *et al.* (1996) Molecular evidence for deep Precambrian divergences among metazoan phyla. *Science* 274, 568–573
50 Sanderson, M.J. (2003) Molecular data from 27 proteins do not support a Precambrian origin of land plants. *Am. J. Bot.* 90, 954–956
51 Ho, S.Y.W. *et al.* Accuracy of rate estimation using relaxed clock models, with a critical focus on the early metazoan radiation. *Mol. Biol. Evol.* (in press)
52 Gillespie, J.H. (1989) Lineage effects and the index of dispersion of molecular evolution. *Mol. Biol. Evol.* 6, 636–647
53 Smith, N.G.C. and Eyre-Walker, A. (2003) Partitioning the variation in mammalian substitution rates. *Mol. Biol. Evol.* 20, 10–17
54 Cutler, D.J. (2000) Understanding the overdispersed molecular clock. *Genetics* 154, 1403–1417
55 Gillespie, J.H. (1991) *The Causes of Molecular Evolution*, Oxford University Press
56 Bromham, L. *et al.* (1996) Determinants of rate variation in mammalian DNA sequence evolution. *J. Mol. Evol.* 43, 610–621
57 Martin, A.P. and Palumbi, S.R. (1993) Body size, metabolic rate, generation time and the molecular clock. *Proc. Natl. Acad. Sci. U. S. A.* 90, 4087–4091
58 Ohta, T. (1972) Population size and rate of evolution. *J. Mol. Evol.* 1, 305–314

59 Mooers, A.Ø. and Harvey, P.H. (1994) Metabolic rate, generation time and the rate of molecular evolution in birds. *Mol. Phylogenet. Evol.* 3, 344–350

60 Barraclough, T.G. and Savolainen, V. (2001) Evolutionary rates and species diversity in flowering plants. *Evolution* 55, 677–683

61 Webster, A.J. *et al*. (2003) Molecular phylogenies link rates of evolution and speciation. *Science* 301, 478

62 Posada, D. and Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818

63 Bollback, J.P. (2002) Bayesian model adequacy and choice in phylogenetics. *Mol. Biol. Evol.* 19, 1171–1180

64 Zheng, Q. (2001) On the dispersion index of a Markovian molecular clock. *Math. Biosci.* 172, 115–128

Forthcoming Conferences

Are you organizing a conference, workshop or meeting that would be of interest to *TREE* readers? If so, please e-mail the details to us at TREE@elsevier.com and we will feature it in our Forthcoming Conference filler.

1–5 July 2005
International Society for Molecular Biology and Evolution Annual Meeting, Auckland, New Zealand
http://www.mbe05.com/

12–13 July 2005
BES/IEEM Conference on Ecological Impact Assessments: Science and Best Practice, Bath, UK
http://www.britishecologicalsociety.org/articles/groups/conservation/bes_ieem_conf/

15–19 July 2005
The 19th Annual Meeting of the Society for Conservation Biology, Brasília, Brazil
http://www.unb.br/ib/zoo/scb2005/

19–21 July 2005
ASAB Summer Meeting, Behavioural Interactions: Visions for the Future, University of Lancaster, UK
http://www.societies.ncl.ac.uk/asab/meetings.html

24–28 July 2005
Association for Tropical Biology and Conservation Annual Meeting, Uberlândia, Brazil
http://www.atbio.org/meetings.html

31 July–5 August 2005
Gordon Research Conference: Evolutionary and Ecological Functional Genomics, Oxford, UK
http://www.grc.uri.edu/05sched.htm#GRC

6–10 August 2005
The Animal Behavior Society 2005 Annual Meeting, Snowbird, UT, USA
http://www.animalbehavior.org/ABS/Program/

7–12 August 2005
90th Annual Meeting of the Ecological Society of America, Montréal, Canada
http://www.esa.org/montreal/

15–20 August 2005
10th Congress of the European Society for Evolutionary Biology, Krakow, Poland
http://www.eseb.org/index.html?http://www.eseb.org/congress.htm

2–4 September 2005
Special BES Symposium, Body Size and the Organisation and Function of Aquatic Ecosystems, University of Hertfordshire, UK
http://www.britishecologicalsociety.org/articles/meetings/current/specialsymposium2005/

13–17 September 2005
World Conference on Ecological Restoration. Ecological Restoration: A Global Challenge. Zaragoza, Spain
http://www.ecologicalrestoration.net

21–23 September 2005
9th Evolutionary Biology Meeting at Marseilles, Marseilles, France
http://www.up.univ-mrs.fr/evol/congres/

25–29 September 2005
International Symposium on Biological control of Aphids and Coccids, University of Yamagata, Japan
http://www.bf.jcu.cz/tix/strita/aphidophaga/tsurprog.html

26–27 September 2005
Royal Society Discussion Meeting: Major Steps in Cell Evolution: Evidence, Timing and Global Impact, London, UK
http://www.royalsoc.ac.uk/events/

17–18 October 2005
Royal Society Discussion Meeting: Evolution of the Antarctic Ice Sheet: New Understandings and Challenges, London, UK
http://www.royalsoc.ac.uk/events/

19–22 October 2005
The Society of Vertebrate Paleontology, Mesa, Arizona, USA
http://www.vertpaleo.org/meetings/future_meetings.html

4–9 December 2005
Combined Conferences of Invertebrate Biodiversity and Conservation, Australian Entomological Society and the Society of Australian Systematic Biologists, Canberra, Australia
http://www.invertebrates2005.com

13–19 August 2006
24th International Ornithological Congress, Hamburg, Germany
http://www.i-o-c.org