Bayesian molecular clock dating of species divergences in the genomics era

Mario dos Reis[1,2], Philip C. J. Donoghue[3] and Ziheng Yang[1]

Abstract | Five decades have passed since the proposal of the molecular clock hypothesis, which states that the rate of evolution at the molecular level is constant through time and among species. This hypothesis has become a powerful tool in evolutionary biology, making it possible to use molecular sequences to estimate the geological ages of species divergence events. With recent advances in Bayesian clock dating methodology and the explosive accumulation of genetic sequence data, molecular clock dating has found widespread applications, from tracking virus pandemics and studying the macroevolutionary process of speciation and extinction to estimating a timescale for life on Earth.

Molecular clock
The hypothesis that the rate of molecular evolution is constant over time or among species. Thus, mutations accumulate at a uniform rate after species divergence, keeping time like a timepiece.

Tree of Life
The evolutionary tree depicting the relationships among all the living species of organisms, calibrated to the geological time.

[1]*Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK.*
[2]*School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK.*
[3]*School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK.*
Correspondence to M.d.R. and Z.Y.
m.dosreisbarros@qmul.ac.uk; z.yang@ucl.ac.uk

doi:10.1038/nrg.2015.8
Published online 21 Dec 2015

Five decades ago, Zuckerkandl and Pauling published two seminal papers in which they proposed the concept of the molecular evolutionary clock[1,2]; that is, that the rate of evolution at the molecular level is approximately constant through time and among species. The idea arose when the pioneers of molecular evolution compared protein sequences (haemoglobins, cytochrome *c* and fibrinopeptides) from different species of mammals[1,3,4] and observed that the number of amino acid differences between species correlated with their divergence time based on the fossil record. The field of molecular evolution was revolutionized by this hypothesis, albeit not without controversy[5–7,8] (BOX 1), and biologists took on the task of using the molecular clock as a technique for inferring the dates of major species divergence events in the Tree of Life[9].

From the outset, the molecular clock was not perceived as a perfect timepiece but rather as a stochastic clock in which mutations accumulate at random intervals, albeit at approximately the same rate in different species, thus keeping time as a clock does. Initial statistical clock dating methodology that was based on distance and maximum likelihood methods assumed a perfectly constant rate of evolution (the 'strict' clock) and used fossil-age calibrations as point values (even though the fossil record can never provide a precise date estimate for a clade).

Subsequent tests of the molecular clock[10,11] showed that it is often 'violated'; that is, the molecular evolutionary rate is not constant, except in comparisons of closely related species, such as the apes. Multiple factors might influence the varying molecular evolutionary rates among species (such as generation time, population size, basal metabolic rate and so on); however, the exact mechanisms of rate variation and the relative importance of these factors are still a matter of debate[7,12,13]. When the clock is violated, methods for dealing with rate variation include the removal of species that exhibit unusual rates from the analyses[14], as well as the so-called local-clock models, which arbitrarily assign branches to rate classes[15,16].

Sophisticated statistical models that take into account uncertainty in the fossil record as well as variation in evolutionary rate — and thus enable the strict clock assumption to be 'relaxed' — were not developed until the advent of Bayesian methods in the late 1990s and early 2000s. It is now generally acknowledged that the molecular clock cannot be applied globally or to distantly related species. However, for closely related species, or in the analysis of population data, the molecular clock is a good approximation of reality (BOX 2).

Next-generation sequencing technologies and advances in Bayesian phylogenetics over the past decade have led to a dramatic increase in molecular clock dating studies. Examples of recent applications of the molecular clock include the rapid analysis of the 2014 Ebola virus outbreak[17], the characterization of the origin and spread of HIV[18] and influenza[19,20], ancient DNA studies to reconstruct a timeline for the origin and migration patterns of modern humans[21–23], the use of time trees to infer macroevolutionary patterns of speciation and extinction through time[24,25], and the co-evolution of life and the Earth[26,27]. Knowledge of the absolute times of species divergences has proved critically important

Box 1 | **The clock and the neutral theory of molecular evolution**

Zuckerkandl and Pauling provided a justification for the molecular clock by suggesting that amino acid changes that accumulate between species are mostly those with little or no effect on the structure and function of the protein, thus reflecting the background mutational process at the DNA level[1]. This hypothesis was formalized by Kimura[106] and by King and Jukes[107] in the neutral theory of molecular evolution, which asserts that most of the genetic variation that we observe (either polymorphisms within species or divergence between species) is due to chance fixation of selectively neutral mutations, rather than due to fixation of advantageous mutations driven by natural selection[6]. Thus, the molecular clock was soon entwined in the controversy surrounding the neutral theory, which was initially proposed to explain the surprising finding of high levels of polymorphism in natural populations[108,109]. If molecular evolution is dominated by neutral mutations, which have little influence on the survival or reproduction of the individual, then an approximately constant rate of evolution is plausible. Indeed, under this theory, the rate of molecular evolution is equal to the neutral mutation rate, which can be assumed to be similar among species with similar life histories.

Most mutations that arise in a generation in a large population are lost by chance within a small number of generations. This is true not only for neutral and deleterious mutations, but also for advantageous mutations unless the advantage is extremely large. For example, if a mutation offers a 1% selective advantage (which is a very large advantage), there is only about 2% chance that the mutation will eventually spread through the whole population[110]. The minority of mutations that are eventually fixed in the population are known as substitutions. Viewed over a very long timescale, this process of new mutations reaching fixation, replacing previous wild-type alleles, is the process of molecular evolution. Suppose the total mutation rate is μ per generation, and a fraction f_0 of the mutations is neutral. The rest of the mutations are deleterious and are removed by natural selection, and do not contribute to the evolutionary process. There are $2N \times \mu f_0$ neutral mutations per generation for a diploid population of size N. The chance that a neutral mutation will eventually reach fixation is $1/(2N)$, because there are $2N$ alleles in the population and each has the same chance of reaching fixation. The molecular substitution rate per generation r (that is, the number of mutations per generation that reach fixation in the population) is thus equal to the number of new neutral mutations produced in each generation multiplied by the probability that they will eventually reach fixation; that is:

$$r = 2N\mu f_0 \times 1/(2N) = \mu f_0 \quad (1)$$

In other words, the substitution rate is equal to the neutral mutation rate (μf_0)[111]. According to this neutral mutation-random drift theory (or the neutral theory), the rate of molecular evolution reflects the neutral mutation rate independently of the population size. Thus, the molecular clock holds if μ and f_0 are approximately constant through time and similar among closely related species.

Hence, the neutral theory offers an explanation for the molecular clock, and for a time the clock was considered the most important evidence supporting the neutral theory[6]. Proteins with different functional constraints may have different proportions of neutral mutations (f_0), so that they have different rates of neutral mutation and their clocks tick at different rates. Extensive reviews of the clock-neutral theory controversy are given elsewhere[6,7,112].

for the interpretation of newly sequenced genomes[23,28]. Exciting new developments in Bayesian phylogenetics include: relaxed clock models to accommodate the violation of the clock[29–31]; modelling of fossil preservation and discovery to generate prior probability distributions of divergence times to be used as calibrations in molecular clock dating[32]; and the integration of morphological characters from modern and extinct species in a combined analysis with sequencing data[33,34].

In this Review we discuss the history, prospects and challenges of using molecular clock dating to estimate the timescale for the Tree of Life, particularly in the genomics era, and trace the rise of the Bayesian molecular clock dating method as a framework for integrating information from different sources, such as fossils and genomes. We do not discuss non-Bayesian clock dating methods[35–38], which typically do not adequately accommodate different sources of uncertainty in a dating analysis. These methods usually involve less computation and may thus be useful for analysing very large data sets for which the Bayesian method is still computationally prohibitive. A detailed review of non-Bayesian clock dating can be found elsewhere[39].

Likelihood
The probability of the observed data given the model parameters viewed as a function of the parameters with the data fixed. In Bayesian clock dating, likelihood is calculated using the sequence data (and possibly morphological data) under a model of character evolution.

Early attempts to estimate the time tree of life

Time trees, or phylogenies with absolute divergence times, provide incomparably richer information than a species phylogeny without temporal information, as they make it possible for species divergence events to be calibrated to geological time, from which correlations can be made to events in the Earth's history and, indeed, to other events in biotic evolution (that is, by calibrating independent but potentially interacting lineages to the same timescale), thus allowing for macroevolutionary hypotheses of species divergences and extinctions to be tested.

As the first protein and DNA sequences became available for a diversity of species, biologists started using the molecular clock as a simple but powerful tool to estimate species divergence times. Underlying the notion that molecules can act as a clock is the theory that the genetic distance between two species, which is determined by the number of mutations accumulated in genes or proteins over time, is proportional to the time of species divergence (BOX 1). If the time of divergence between two species is known — from fossil evidence, from a geological event (such as continental break-up or island formation) or from sample dates for bacteria and viruses — the genetic distance between these species can be converted into an estimate of the rate of molecular evolution, which can be applied to all nodes on the species phylogeny to produce estimates of absolute geological times of divergence (BOX 2). One of the first applications of this idea was by Sarich and Wilson[40], who used a molecular clock to infer the immunological distance of albumins. By assuming a divergence time of 30 Ma between the apes and New World monkeys, they calculated the age of the last common ancestor of humans and African apes (chimpanzees and gorillas) as 5 Ma. This work ignited one of the first 'fossils versus molecules' controversies as, at the time, the divergence between human and African apes was thought to be over 14 Ma on the basis of the ages of the fossils *Ramapithecus* and *Sivapithecus*[41]. The controversy was settled once it was recognized that the fossils are more closely related to the orang-utan than to the African apes.

In response to the expanding genetic sequence data sets that resulted from the PCR revolution in the late 1990s, molecular clock dating was applied to a broad range of species. These studies generated considerable controversy because the clock estimates were much older than the dates suggested by the fossil record, sometimes twice as old[42], and many palaeontologists considered the discrepancy to be unacceptably large[43]. Examples include Mesoproterozoic estimates for the timing of the origin and diversification of the animal phyla relative

Fossil-age calibrations
Constraints on the timing of lineage divergence in molecular clock dating. They are established through fossil-based minimum and maximum constraints on clade ages (node calibrations) or through the inclusion of dated fossil species in the analysis (tip calibrations).

Clade
A group of species descended from a common ancestor.

Bayesian methods
Statistical inference methodologies in which statistical distributions are used to represent uncertainties in model parameters. In Bayesian clock dating, priors on times and rates are combined with the likelihood (the probability of the sequence data) to produce the posterior of times and rates.

to their Phanerozoic fossil record[44], a Triassic origin of flowering plants relative to a fossil record beginning in the Cretaceous[45], and a Jurassic or Cretaceous origin of modern birds and placental mammals relative to fossil evidence that is mostly confined to the period after the end-Cretaceous mass extinction[46,47].

The early dating studies suffer from a number of limitations[48,49]. For example, many studies assumed a strict clock even for distantly related species, and most used point fossil calibrations without regard for their uncertainty[25,47]. Sometimes, secondary calibrations — that is, node ages estimated in previous molecular clock dating studies — were used[48]. Despite their limitations, these studies encouraged much discussion about the nature of the fossil record and the molecular clock[49] and inspired the development of more sophisticated methods. These early studies proposed a timescale for life on Earth that has now been revised in the newer genome-scale analyses[24,50,51].

The Bayesian method of clock dating

The Bayesian method was introduced into molecular clock dating around the year 2000 in a series of seminal papers by Jeff Thorne and colleagues[29,52,53]. The method has been developed greatly since then[30,31,54,55], emerging as the dominant approach to divergence time estimation owing to its ability to integrate different sources of information (in particular, fossils and molecules) while accommodating the uncertainties involved.

The Bayesian method is a general statistical methodology for estimating parameters in a model. Its main feature is the use of statistical distributions to characterize uncertainties in all unknowns. One assigns a prior probability distribution on the parameters, which is combined with the information in the data (in the form of the likelihood function) to produce the posterior probability distribution. In molecular clock dating, the parameters are the species divergence times (t) and the evolutionary rates (r). Given the sequence data (D), the posterior of times and rates is given by the Bayes theorem as follows:

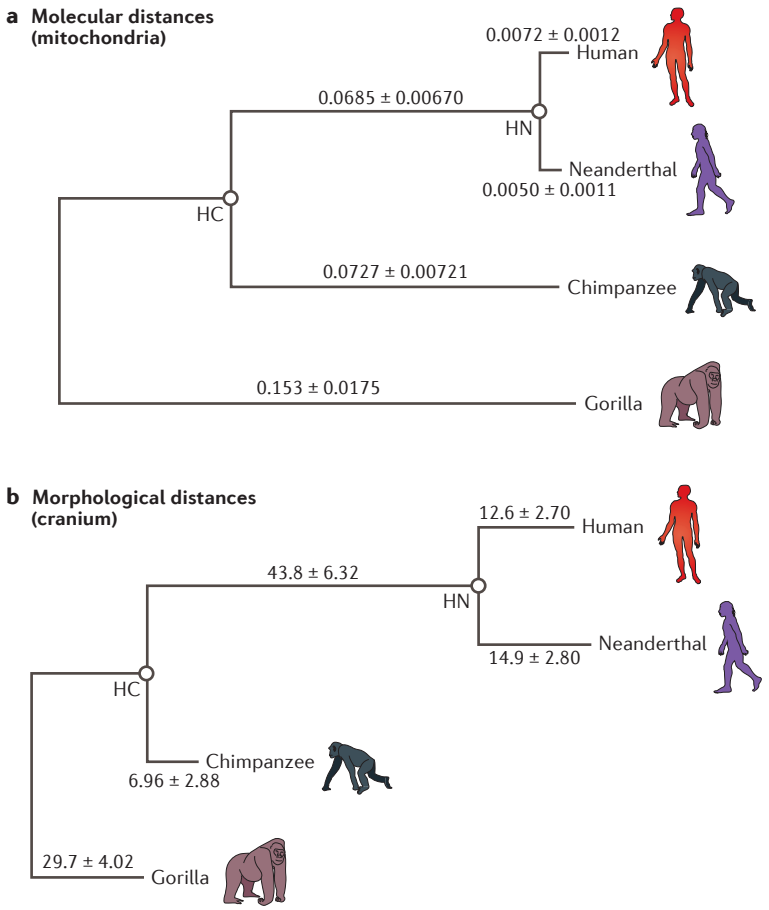
$$f(t, r|D) = \frac{1}{z} f(t)\, f(r|t)\, L(D|t, r) \qquad (2)$$

Here, $f(t)$ is the prior on divergence times, which is often specified using a model of cladogenesis (of speciation and extinction[54,56], and so on) and incorporates the fossil calibration information[52,54]; $f(r|t)$ is the prior on the rates of branches on the tree, which is specified using a model of evolutionary rate drift[29–31]; and $L(D|t, r)$ is the likelihood or the probability of the sequence data, which is calculated using standard algorithms[11]. FIGURE 1

Box 2 | Clock-like molecular evolution versus non-clock-like morphological evolution

Molecular sequences can evolve at a nearly constant rate among close species. An alignment of the human (H), Neanderthal (N), chimpanzee (C) and gorilla (G) mitochondrial genomes (15,889 bp) was analysed by maximum-likelihood under the GTR+Γ_4 model[113,114] to estimate the branch lengths without the assumption of a molecular clock. The molecular distance (see the figure, part **a**) from the common ancestor of human–chimpanzee (HC) to the human (± standard error) is $d_{H\text{-}HC} = 0.0757 \pm 0.00681$, and that from HC to the chimpanzee is $d_{C\text{-}HC} = 0.0727 \pm 0.00721$. These distances are nearly identical, as would be expected under the molecular clock hypothesis. Indeed, the strict clock hypothesis is not rejected by a likelihood-ratio test[11] ($P = 0.60$). The rate constancy of the mitochondrial genome allows us to date the age of the common ancestor of the human and Neanderthal (HN). Under the clock, the times are proportional to the distances, so that $t_{HN}/t_{HC} = 0.0072/0.0757 = 0.0951$. The fossil record suggests that the HC ancestor lived 10–6.5 Ma (REF. 115). Thus, we obtain 0.95–0.62 Ma for the age of the HN ancestor.

By contrast, evolutionary rates of morphological characters may be much more variable (see the figure, part **b**). The 151 cranium landmark measurements from the same four species[116] were aligned and analysed using maximum likelihood under Felsenstein's trait-evolution model[117]. The morphological branch lengths (in units of expected accumulated variance) are shown on the tree. From the branch lengths $b_{H\text{-}HC} = 56.4 \pm 6.87$ and $b_{C\text{-}HC} = 6.96 \pm 2.88$, we see that the human cranium has changed 8.1 times as fast as the chimpanzee since the split of the two species. Driven by natural selection, the human cranium has rapidly become larger and rounder, with a smaller and more protracted face.



Neutral theory
Also termed the neutral mutation-random drift theory; claims that evolution at the molecular level is mainly random fixation of mutations that have little fitness effect.

Neutral mutations
Mutations that do not affect the fitness (survival or reproduction) of the individual.

Advantageous mutations
Mutations that improve the fitness of the carrier and are favoured by natural selection.

Deleterious mutations
Mutations that reduce the fitness of the carrier and are removed from the population by negative selection.

Substitution
Mutations that spread into the population and become fixed, driven either by chance or by natural selection.

Relaxed clock models
Models of evolutionary rate drift over time or across lineages developed to relax the molecular clock hypothesis.

illustrates the Bayesian clock dating of equation (2) in a two-species case.

Direct calculation of the proportionality constant z in equation (2) is not feasible. In practice, a simulation algorithm known as the Markov Chain Monte Carlo algorithm (MCMC algorithm) is used to generate a sample from the posterior distribution. The MCMC algorithm is computationally expensive, and a typical MCMC clock-dating analysis may take from a few minutes to several months for large genome-scale data sets. Methods that approximate the likelihood can substantially speed up the analysis[29,57,58]. For technical reviews on Bayesian and MCMC molecular clock dating see REFS 59,60.

Nearly a dozen computer software packages currently exist for Bayesian dating analysis (TABLE 1), all of which incorporate models of rate variation among lineages (the episodic or relaxed clock models envisioned by Gillespie)[61]. All of these programs can also analyse multiple gene loci and accommodate multiple fossil calibrations in one analysis.

Limits of Bayesian divergence time estimation

Estimating species divergence times on the basis of uncertain calibrations is challenging. The main difficulty is that molecular sequence data provide information about molecular distances (the product of times and rates) but not about times and rates separately. In other words, the time and rate parameters are unidentifiable. Thus, in Bayesian clock dating, the sequence distances are resolved into absolute times and rates through the use of priors. In a conventional Bayesian estimation problem, the prior becomes unimportant and the Bayesian estimates converge to the true parameter values as more and more data are analysed. However, convergence on truth does not occur in divergence time estimation. The use of priors to resolve times and rates has two consequences. First, as more loci or increasingly longer sequences are included in the analysis but the calibration information does not change, the posterior time estimates do not converge to point values and will instead involve uncertainties[31,54,62]. Second, the priors on times and on rates have an important impact on the posterior time estimates even if a huge amount of sequence data is used[62,63]. Errors in the time prior and in the rate prior can lead to very precise but grossly inaccurate time estimates[62,64]. Great care must always be taken in the construction of fossil calibrations and in the specification of priors on times and on rates in a dating analysis[65,66].

As the amount of sequence data approximates genome scale, the molecular distances or branch lengths on the phylogeny are essentially determined without any uncertainty, as are the relative ages of the nodes. However, the absolute ages and absolute rates cannot be known without additional information (in the form of priors). The joint posterior of times and rates is thus one-dimensional. This reasoning has been used to determine the limiting posterior distribution when the amount of sequence data (that is, the number of loci or the length of the sequences) increases without bound[31,54]. An infinite-sites plot can be used to determine whether the amount of sequence data is saturated or whether including more sequence data is likely to improve the time estimates (FIG. 2). The theory has been extended to the analysis of large but finite data sets to partition the uncertainties in the posterior time

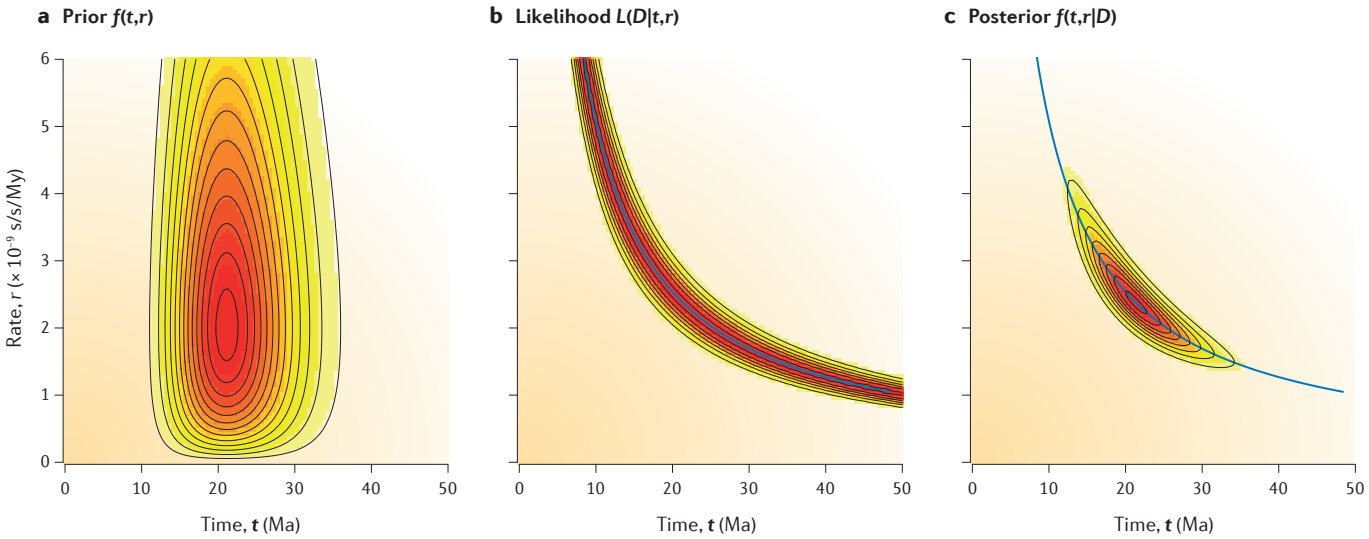


Figure 1 | **Bayesian molecular clock dating.** We estimate the posterior distribution of divergence time (t) and rate (r) in a two-species case to illustrate Bayesian molecular clock dating. The data are an alignment of the 12S RNA gene sequences from humans and orang-utans, with 90 differences at 948 nucleotides sites. The joint prior (part **a**) is composed of two gamma densities (reflecting our prior information on the molecular rate and on the geological divergence time of human–orang-utan), and the likelihood (part **b**) is calculated under the Jukes–Cantor model. The posterior surface (part **c**) is the result of multiplying the prior and the likelihood. The data are informative about the molecular distance, $d = tr$, but not about t and r separately. The posterior is thus very sensitive to the prior. The blue line indicates the maximum likelihood estimate of t and r, and the molecular distance d, with $\hat{t}\hat{r} = \hat{d}$. When the number of sites is infinite, the likelihood collapses onto the blue line, and the posterior becomes one-dimensional[62].

Table 1 | **Sample of Bayesian programs that use the molecular clock to estimate divergence times***

Program	Method	Brief description	Refs
Beast	Bayesian	Comprehensive suite of models. Particularly strong for the analysis of serially sampled DNA sequences. Includes models of morphological traits	132
DPPDiv	Bayesian	Dirichlet relaxed clock model[71]. Fossilized birth–death process prior to calibrate time trees[56]	133
MCMCTree	Bayesian	Comprehensive suite of models of rate variation. Fast approximate likelihood method that allows the estimation of time trees using genome alignments[57]	134
MrBayes	Bayesian	Large suite of models for morphological and molecular evolutionary analysis. Comprehensive suite of models of rate variation	135
Multidivtime	Bayesian	The first Bayesian clock dating program. Introduced the geometric Brownian model and the approximate likelihood method	29,53
PhyloBayes	Bayesian	Broad suite of models. Uses data augmentation to speed up likelihood calculation and can be efficiently used in parallel computing environments (MPI enabled)	136, 137
r8s	Penalized likelihood	Very fast (uses Poisson densities on inferred mutations to approximate the likelihood). Suitable for the analysis of large phylogenies. Suitable for estimating relative ages (by fixing the age of the root to 1). Does not deal with fossil and branch length uncertainty correctly[138]	139
TreePL	Penalized likelihood	Similar to r8s	140

*The Bayesian programs listed were chosen for their ability to accommodate multiple calibrations with uncertainties (bounds or other probability densities), multiple loci of sequence data and relaxed clock models. Penalized likelihood programs are listed as they are related to the Bayesian method[138].

Prior probability distributions
Distributions assigned to parameters before the analysis of the data. In Bayesian clock dating, the prior on divergence times is specified using a branching model, possibly incorporating fossil calibration information, and the prior on evolutionary rates is specified using a model of rate drift (a relaxed-clock model).

Morphological characters
Discrete features or continuous measurements of different species that are informative about phylogenetic relationships.

Phylogeny
A tree structure representing the evolutionary relationship of the species.

Posterior probability distribution
The distribution of the parameters (or models) after analysis of the observed data. It combines the information in the prior and in the data (likelihood).

Likelihood-ratio test
A general hypothesis-testing method that uses the likelihood to compare two nested hypotheses, often using the χ^2.

Markov chain Monte Carlo algorithm
(MCMC algorithm). A Monte Carlo simulation algorithm that generates a sample from a target distribution (often a Bayesian posterior distribution).

Jukes–Cantor model
A model of nucleotide substitution in which the rate of substitution between any two nucleotides is the same.

estimates according to different sources: uncertain fossil calibrations and finite amounts of sequence data[62,63]. Application of the theory to the analysis of a few real data sets (including genome-scale data) has indicated that most of the uncertainty in the posterior time estimates is due to uncertain calibrations rather than to limited sequence data[24,66].

Relaxed clock models — the prior on rates

Unsurprisingly, divergence time estimation under the strict molecular clock is highly unreliable when the clock is seriously violated. In early studies it was common to remove genes and/or lineages that violated the clock from the analysis[14], but this method does not make efficient use of the data and is impractical when the clock is violated by too many genes or species. Relaxed clock models have been developed to allow the molecular rate to vary among species. The first methods were developed under the penalized-likelihood and maximum-likelihood frameworks[67,68]. In Bayesian clock dating, such models are integrated into the analysis as the prior on rates.

Several types of relaxed clock models have been implemented, using either continuous or discrete rates. In the geometric Brownian motion model[29,31,52] (also known as the autocorrelated-rates model) the logarithm of the rate drifts over time as a Brownian motion process (FIG. 3a). Let $y_0 = \log(r_0)$ and $y_t = \log(r_t)$, where r_0 is the ancestral rate at time 0 while r_t is the rate time t later. Then:

$$y_t \mid y_0 \sim N(y_0, tv) \quad (3)$$

That is, given y_0 (or the ancestral rate r_0), y_t has a normal distribution with mean y_0 and variance tv (or r_t has a log-normal distribution). Thus, rates on descendent branches are similar to the rate of the ancestral branch, especially if the branches cover short timescales; furthermore, the variance of the rate increases with the passage of time. An unappealing property of Brownian motion is that it does not have a stationary distribution. Over a very long timescale, the log-rate can drift to very negative or very positive values with the rate becoming near zero or very large, and the variance of the rate tends to approach infinity with time. This does not seem to be realistic. A model that does not have this property is the (geometric) Ornstein–Uhlenbeck model (FIG. 3b). The logarithm of the rate follows Brownian motion with a dampening force, leading to a stationary distribution. This model (and the related Cox–Ingersoll–Ross model)[55] looks promising and merits further research. Notably, an early implementation of the Ornstein–Uhlenbeck model[69] to clock dating inadvertently assumed that evolutionary rates drift to zero with time[70]. Another type of relaxed clock model assumes a small number of distinct rates on the tree and assigns branches to the rate classes through a random process[71–73]. It is also possible to assume that the rates for branches on the tree do not correlate and are random draws from the same common distribution such as the log-normal[30,31] (FIG. 3c).

Fossil calibrations — the prior on times

Molecular clock analyses are most commonly calibrated using evidence from the fossil record[74,75]. Geological events such as the closure of the Isthmus of Panama or continental break-ups can also be used as calibrations, although such calibrations may also involve many uncertainties owing to assumptions about vicariance, species dispersal potential, and so on[76]. In Bayesian clock dating,

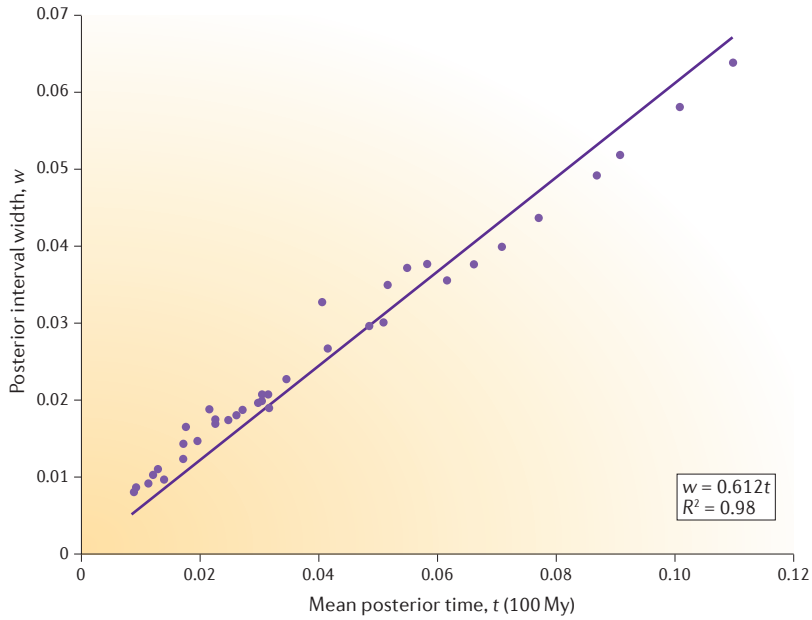


Figure 2 | **Infinite-sites plot for Bayesian clock dating of divergences among 38 cat species.** There are 37 nodes on the tree and 37 points in the scatter plot. The *x* axis is the posterior mean of the node ages and the y axis is the 95% posterior credibility interval (CI) width of the node ages. Here the slope (0.612) indicates that every million years of species divergence adds 0.612 million years of uncertainty in the posterior CI. When the amount of sequence data is infinite the points will fall onto a straight line. Here, the high correlation ($R^2 = 0.98$) indicates that the amount of sequence data is very high, and the large uncertainties in the posterior time estimates are mostly due to uncertainties in the fossil calibrations; including more sequence data is unlikely to improve the posterior time estimates. Reproduced from Inoue, J., Donoghue, P. C. J. & Yang, Z. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* **59**(1), 74–89 (2010), by permission of the Society of Systematic Biologists.

calibration information is incorporated in the analysis through the prior on times.

It has long been recognized that the fossil record is incomplete — temporally, spatially and taxonomically — and long time gaps may exist between the oldest known fossils and the last common ancestor of a group. The first known appearance of a fossil member of a group cannot be interpreted as the time and place of origination of the taxonomic group[77]. For example, during the 1980s the oldest known members of the human lineage were the Australopithecines, dating to around 4 Ma (REF. 41), providing a minimum age for the divergence time between humans and chimpanzees. However, since 2000, several fossils belonging to the human lineage have been discovered in quick succession, including *Ardipithecus* (4.4 Ma), *Orrorin* (6 Ma) and *Sahelanthropus* (7 Ma), which pushed the age of the human–chimpanzee ancestor to over 7 Ma (REF. 78). Some groups have no known fossil record, such as the Malagasy lemurs for which only a few hundred-year-old sub-fossils are known[79]. The oldest fossil in their sister lineage (the galagos and lorises) dates to 38 Ma, indicating a minimum 38 My gap in the fossil record of lemurs[80]. Clearly, fossil ages provide good minimum-age bounds on clade ages, but assuming that clade ages are the same as that of their oldest fossil is unwarranted and incorrect[81,82].

Soft bounds
Minimum or maximum constraints on a node age with small error probabilities (such as 1% or 5%) used as bounds in clock dating.

However, minimum-age bounds alone are insufficient for calibrating a molecular tree. Recent developments in Bayesian dating methodology have enabled soft bounds and arbitrary probability curves to be used as calibrations[30,54,83]. Soft bounds assign small probabilities (such as 5% or 10%) for the violation of the bounds[54]. These developments have motivated palaeontologists to formulate probabilistic densities for the true clade ages, rather than focusing on the minimum age. A programme has been launched in palaeontology to reinterpret the fossil record to provide both sharp minimum bounds and soft maximum bounds on clade ages[84,85].

We envisage several strategies for generating fossil calibrations, each of which may be appropriate depending on the available data. First, one may use the absence of evidence (the lack of available fossil species in the rock record) as weak evidence of absence and thus construct soft maximum age bounds[81,82]. Together with hard or sharp minimum-age bounds, they can be used as calibrations. This procedure may involve some subjectivity. Second, fossil occurrences in the rock layers can be analysed using probabilistic models of fossil preservation and discovery to generate posterior distributions of node ages, which can be used in subsequent molecular dating studies[32,56,86–88]. Third, if morphological characters are scored for both modern and fossil species then they can be analysed using models of morphological character evolution to estimate node ages, which serve as calibrations in molecular clock dating. It is advisable to fix the phylogeny for modern species while allowing the placement of the fossil species to be determined by the data. Fossil remains are typically incomplete and their phylogenetic placement most often involves uncertainties[89]. It is also possible to analyse the fossil or morphological data and the molecular data in one joint analysis, as discussed below (known as total evidence dating)[34].

Joint analysis of molecular and morphological data

Morphological characters from both fossil species (which have been dated) and modern species may be analysed jointly with molecular data under models of morphological character evolution to estimate divergence times[33,34]. The analysis is statistically similar to the analysis of serially sampled sequences in molecular dating of viral or ancient DNA and proteins (BOX 3). A perceived advantage of this 'tip-calibration' approach is that it is unnecessary to use constraints on node ages (so-called node calibration). The approach also facilitates the co-estimation of time and topology. Recent applications of this strategy to insects[34], arachnids[90,91], fish[92,93] and mammals[94–96] have produced surprisingly ancient divergence times[97].

Although tip calibration offers a coherent framework for integrating information from molecules and fossils in one combined analysis, its current implementation involves a number of limitations, which may underlie these old date estimates. First, current models of morphological character evolution are simplistic and may not accommodate important features of the data well[98]. For example, morphological characters tend to be strongly correlated, but almost all current models assume

Parsimony-informative characters
A discrete character is informative to the parsimony method of phylogenetic reconstruction if at least two states are observed among the species, each state at least twice.

independence. Furthermore, all recent tip-dating studies have analysed discrete morphological characters, but morphologists usually score only variable characters or parsimony-informative characters. Such ascertainment bias, even if correctly accommodated in the model[98], greatly reduces information about branch lengths and divergence times in the data. Whereas the removal of constant characters can be easily accommodated[98], the removal of parsimony-uninformative characters would require too much computation and is not properly accommodated by any current dating software. Second, a tip-calibrated analysis does not place any constraints on the ages of internal nodes on the tree and may thus be very sensitive to the prior of divergence times or the branching process used to generate that prior compared with dating using node calibrations. In a sense, although node dating uses node calibrations that may be subjective, it allows the palaeontologist's common sense to be injected into the Bayesian analysis. By contrast, tip calibration may be unduly influenced by arbitrary choices of priors implemented in the computer program. Third, it is generally the case that there is far more molecular data than morphological characters, and that morphological characters may undergo convergent evolution in distant species and may evolve at much more variable rates than molecules[6]. BOX 2 presents the case of cranial evolution within the hominoids, in which the rate in the human is about eight times as high as the rate in the chimpanzee. Such drastic changes in morphological evolutionary rate contrast sharply with the near-perfect clock-like evolution of the mitochondrial genome from the same species. Characters with drastically variable evolutionary rates, even if the rate variation is adequately accommodated in the model, will not provide much useful time information

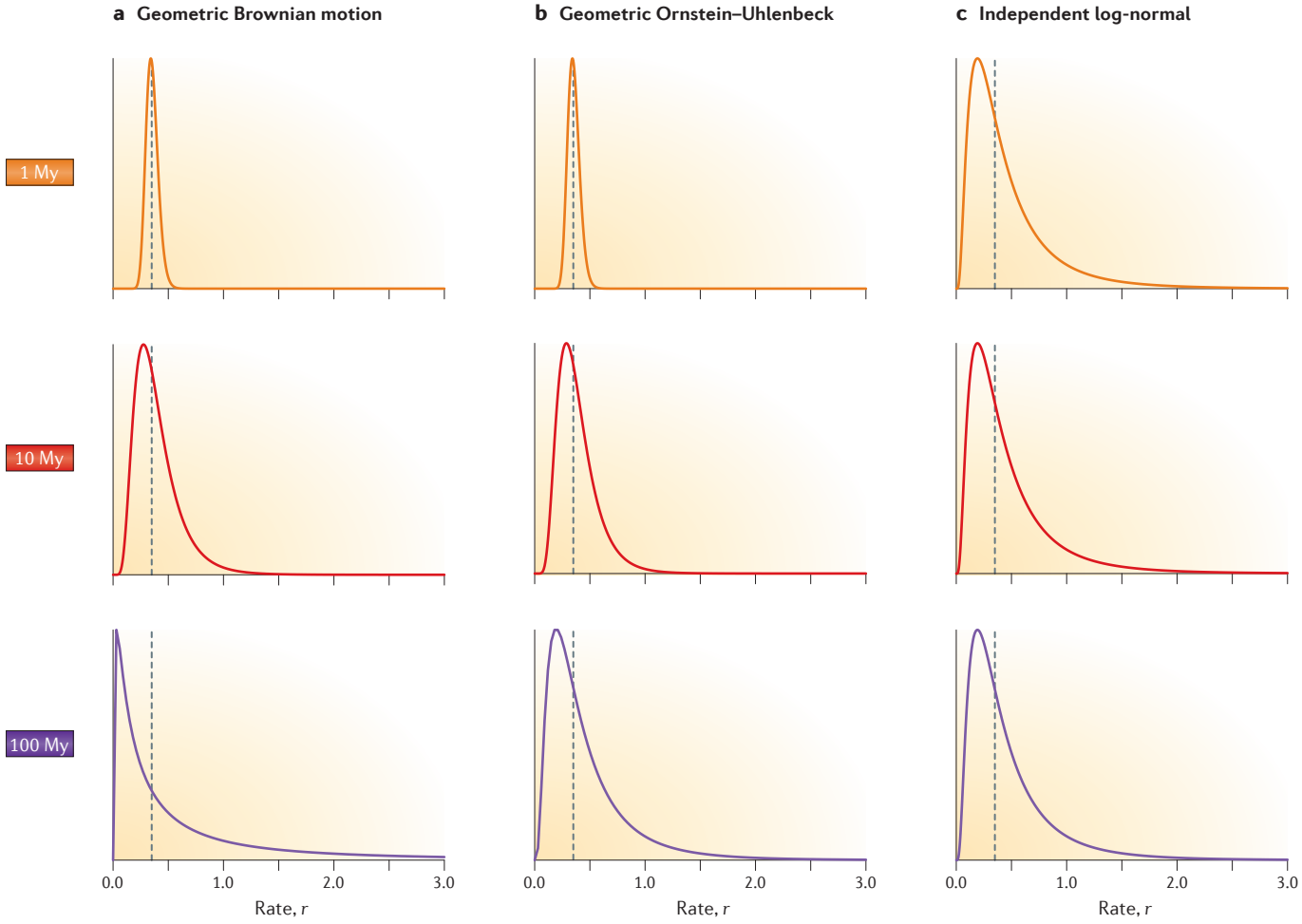


Figure 3 | **Three relaxed clock models of rate drift.** The rate of molecular evolution among lineages (species) is described by a time-dependent probability distribution (plotted here for three time points: 1 My, 10 My and 100 My) since the lineages diverged from a common ancestral rate ($r_0 = 0.35$ substitutions per site per 100 My (represented by the dashed line)). **a** | The geometric Brownian process[29,31,52] (here with drift parameter $v = 2.4$ per 100 My). This model has the undesirable property that the variance increases with time and without bound, and at large times the mode of the distribution is pushed towards zero. **b** | The geometric Ornstein–Uhlenbeck process (here with $v = 2.4$ per 100 My and dampening force $f = 2$ per 100 My) converges to a stationary distribution with constant variance when time is large. **c** | The independent log-normal distribution[30,31] is a stationary process, and the variance of rate among lineages remains constant through time (here with log-variance $\sigma^2 = 0.6$, the same as the long-term log-variance of the Ornstein–Ulhenbeck process above). The branch length (the amount of evolution along the branch) under the rate-drift models of parts **a** and **b** is usually approximated in Bayesian dating software[31,52]; methods for exact calculation have recently been developed[55].

Box 3 | **Dating divergences using serially sampled sequences**

For viral sequences that evolve very quickly, it is possible to observe mutations at the different sampling times of the viral sequences. The different sampling times in combination with the different amounts of evolution that are reflected in the genetic distances can be used to date the divergence events[118–121]. For example, the genome of the 1918 pandemic influenza virus has been sequenced from samples obtained from individuals who died in 1918 and were buried in the Alaskan permafrost[122]. Analysis of the genomic sequences has allowed the estimation of divergence times for the ancestors of the virus[19,20] and has led to proposed scenarios for the origin of the pandemic — for example, a possible swine origin of the virus[123]. Similar approaches have also been used to study the HIV pandemic in humans, tracing its origins from West Africa, its spread in African cities during the mid-twentieth century, and its later spread to the Americas, Europe and the rest of the world[18,124,125].

The strategy of using sequences with sampling dates also applies to studies of ancient DNA (or proteins). Ancient sequence data are informative about times and rates separately, and divergence times can be estimated with high precision if the events to be dated are not much older than the sampling times covered by the data. Analysis of ancient DNA offers exciting prospects for elucidating evolutionary timelines. For example, analysis of several hundred ancient DNA samples from Bison, dating up to 60 Ka, allowed estimation of the timeline of evolution of bison populations, charting the rise and subsequent fall of bison populations in the northern hemisphere throughout the late Pleistocene and Holocene epochs[126]. Other examples of ancient clock studies include dating the origins of horses[127], camels[128] and humans[129]. The approach is limited by our ability to sequence ancient, highly degraded material[130]. The oldest molecular material to be sequenced dates to 0.78–0.56 Ma for DNA[127] and to 80 Ma (controversially) for proteins[131].

for the dating analysis. The small amount of morphological data and the low information content (owing to variable rates) mean that the priors on times and rates will remain important to the dating analysis. Finally, we note that most tip-calibrated studies have not integrated any of the uncertainty associated with fossil dating[97].

Resolving the timeline of the Tree of Life

The molecular clock is now serving as a framework for the integration of genomic and palaeontological data to estimate time trees. Advances in Bayesian clock dating methodology, increased computational power and the accumulation of genome-scale sequence data have provided us with an unprecedented opportunity to achieve this objective. However, considerable challenges remain. Although next-generation sequencing technologies[99] now enable the cheap and rapid accumulation of genome data for many species[100], much work still remains to be carried out to obtain a balanced sampling of biodiversity: some estimates place the fraction of living eukaryotic species that have been described at approximately 14%[101], and sequence data are available for a much smaller and skewed fraction. More seriously, fossils are unavailable for most branches of the Tree of Life, and other sources of information (such as geological events[76] or experimentally measured mutation rates[23]) are only rarely available[102]. The amount of information in fossil morphological characters may never match the information about sequence distances in the genomic data, placing limits on the degree of precision achievable in the estimation of ancient divergence times, because fossil information is essential for resolving sequence distances into absolute times and rates. This problem seems particularly severe in dating ancient divergences, such as the origins of animal phyla[103], because at deeper divergences the quality of fossil data tends to be poor, and the evolutionary rates for both morphological characters and sequence data are highly variable among distantly related species.

Challenges also remain in the development of the statistical machinery necessary for molecular clock dating. Current models of morphological evolution are simplistic and should be improved to accommodate different types of data and to account for the correlation between characters. In the analysis of genomic-scale data sets under relaxed clock models, data partitioning is an important but poorly studied area. The rationale for partitioning the sequence data is that sites in the same partition are expected to share the same trajectory of evolutionary rate drift but those in different partitions do not, so that the different partitions constitute independent realizations of the rate-drift process (for example, geometric Brownian motion). Theoretical analysis suggests that the precision of posterior time estimates is mostly determined by the number of partitions rather than by the number of sites in each partition[63]. However, the different strategies for partitioning large data sets for molecular clock dating analysis are poorly explored. Furthermore, the prior model of rate drift for data of multiple partitions seems to be very important to Bayesian divergence time estimation[53], but currently implemented rate models are highly unrealistic. All current dating programs assume independent rates among partitions, failing to accommodate the lineage effect — the fact that some evolutionary lineages or species tend to be associated with high (or low) rates for almost all genes in the genome[13]. Developing more realistic relaxed clock models for multi-partition data and evaluating their effects on posterior time estimation will be a major research topic for the next few years. Another issue that has been underappreciated in clock dating studies is the fact that speciation events are more recent than gene divergences[104] (a result of the coalescent process of gene copies in ancestral populations), and ignoring this may cause important errors when estimating divergence times[105].

Despite the multitude of challenges, the prospect for a broadly reliable timescale for life on Earth is currently looking more likely than ever before. Genome-scale sequence data are now being applied to resolve iconic controversies between fossils and molecules. For example, Bayesian clock dating using genome-scale data has demonstrated that modern mammals and birds diversified after the K-Pg boundary[24,50] in contrast to non-Bayesian estimates based on limited sequence data that had suggested pre-K-Pg diversification[25,47]. Similarly, Bayesian clock dating analysis of insect genomes has been used to elucidate the time of insect origination in the Early Ordovician[51]. We predict that the explosive increase in completely sequenced genomes, together with the development of efficient Bayesian strategies to analyse morphological and molecular data from both modern and fossil species, will eventually allow biologists to resolve the timescale for the Tree of Life. It seems that in reaching its half-century, the molecular clock has finally come of age.

Coalescent
The process of lineage joining when one traces the genealogical relationships of a sample backwards in time.

K-Pg boundary
The boundary between Cretaceous and Paleogene at 66 Ma. It coincides with a mass extinction, including that of the dinosaurs and many more species.

1. Zuckerkandl, E. & Pauling, L. in *Evolving Genes and Proteins* (eds Bryson, V. & Vogel, H. J.) 97–166 (Academic Press, 1965).
The seminal paper proposing the concept of a molecular evolutionary clock. Provides a justification for the clock based on the idea that most amino acid changes may not change the structure and function of the protein.
2. Zuckerkandl, E. & Pauling, L. in *Horizons in Biochemistry* (eds Kasha, M. & Pullman, B.) 189–225 (Academic Press, 1962).
The earliest clock dating paper. Used the idea of approximate rate constancy to calculate the age of the alpha and beta globin duplication event.
3. Margoliash, E. Primary structure and evolution of cytochrome *c. Proc. Natl Acad. Sci. USA* **50**, 672–679 (1963).
4. Doolittle, R. F. & Blomback, B. Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. *Nature* **202**, 147–152 (1964).
5. Morgan, G. J. Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock. *J. Hist. Biol.* **31**, 155–178 (1998).
6. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, 1983).
Authoritative book outlining the neutral theory. Chapter 4 has an extensive discussion of morphological versus molecular rates of evolution.
7. Bromham, L. & Penny, D. The modern molecular clock. *Nat. Rev. Genet.* **4**, 216–224 (2003).
8. Kumar, S. Molecular clocks: four decades of evolution. *Nat. Rev. Genet.* **6**, 654–662 (2005).
9. Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G. & Little, E. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**, 470–477 (1996).
10. Langley, C. H. & Fitch, W. M. An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3**, 161–177 (1974).
11. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376 (1981).
Seminal paper describing how to calculate the likelihood for a molecular sequence alignment and describing a likelihood-ratio test of the clock.
12. Drummond, D. A., Raval, A. & Wilke, C. O. A single determinant dominates the rate of yeast protein evolution. *Mol. Biol. Evol.* **23**, 327–337 (2006).
13. Ho, S. Y. The changing face of the molecular evolutionary clock. *Trends Ecol. Evol.* **29**, 496–503 (2014).
14. Takezaki, N., Rzhetsky, A. & Nei, M. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**, 823–833 (1995).
15. Rambaut, A. & Bromham, L. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**, 442–448 (1998).
16. Yoder, A. D. & Yang, Z. Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* **17**, 1081–1090 (2000).
17. Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
18. Faria, N. R. *et al.* HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61 (2014).
19. Smith, G. J. *et al.* Dating the emergence of pandemic influenza viruses. *Proc. Natl Acad. Sci. USA* **106**, 11709–11712 (2009).
20. dos Reis, M., Hay, A. J. & Goldstein, R. A. Using non-homogeneous models of nucleotide substitution to identify host shift events: application to the origin of the 1918 'Spanish' influenza pandemic virus. *J. Mol. Evol.* **69**, 333–345 (2009).
21. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
22. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
23. Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13**, 745–753 (2012).
24. dos Reis, M. *et al.* Phylogenomic data sets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B. Biol. Sci.* **279**, 3491–3500 (2012).
An example of using the molecular clock with genome-scale data sets to infer the timeline of diversification of modern mammals relative to the end-Cretaceous mass extinction.
25. Bininda-Emonds, O. R. *et al.* The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
26. Hoorn, C. *et al.* Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**, 927–931 (2010).
27. Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).
28. Carbone, L. *et al.* Gibbon genome and the fast karyotype evolution of small apes. *Nature* **513**, 195–201 (2014).
29. Thorne, J. L., Kishino, H. & Painter, I. S. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647–1657 (1998).
Describes the first Bayesian molecular clock dating method. Introduces the geometric Brownian motion model of rate variation among species.
30. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
31. Rannala, B. & Yang, Z. Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **56**, 453–466 (2007).
32. Wilkinson, R. D. *et al.* Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst. Biol.* **60**, 16–31 (2011).
Develops a model of species origination, extinction and fossil preservation and discovery to construct time priors based on data of fossil occurrences.
33. Pyron, R. A. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Syst. Biol.* **60**, 466–481 (2011).
34. Ronquist, F. *et al.* A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst. Biol.* **61**, 973–999 (2012).
Develops a Bayesian 'total-evidence' dating method for the joint analysis of morphological and molecular data.
35. Xia, X. & Yang, Q. A distance-based least-square method for dating speciation events. *Mol. Phylogenet. Evol.* **59**, 342–353 (2011).
36. Tamura, K. *et al.* Estimating divergence times in large molecular phylogenies. *Proc. Natl Acad. Sci. USA* **109**, 19333–19338 (2012).
37. Paradis, E. Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. *Mol. Phylogenet. Evol.* **67**, 436–444 (2013).
38. Fourment, M. & Holmes, E. C. Novel non-parametric models to estimate evolutionary rates and divergence times from heterochronous sequence data. *BMC Evol. Biol.* **14**, 163 (2014).
39. Ho, S. Y. & Duchene, S. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol. Ecol.* **23**, 5947–5965 (2014).
40. Sarich, V. M. & Wilson, A. C. Immunological time scale for Hominoid evolution. *Science* **158**, 1200–1203 (1967).
41. Simons, E. Man's immediate forerunners. *Phil. Trans. R. Soc.* **292**, 21–41 (1981).
42. Cooper, A. & Fortey, R. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol.* **13**, 151–156 (1998).
43. Benton, M. J. & Ayala, F. J. Dating the tree of life. *Science* **300**, 1698–1700 (2003).
44. Wray, G. A., Levinton, J. S. & Shapiro, L. H. Molecular evidence for deep Precambrian divergences. *Science* **274**, 568–573 (1996).
45. Heckman, D. S. *et al.* Molecular evidence for the early colonization of land by fungi and plants. *Science* **293**, 1129–1133 (2001).
46. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229 (1996).
47. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
48. Graur, D. & Martin, W. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* **20**, 80–86 (2004).
49. Hedges, S. B. & Kumar, S. Precision of molecular time estimates. *Trends Genet.* **20**, 242–247 (2004).
50. Jarvis, E. D. *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
51. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
52. Kishino, H., Thorne, J. L. & Bruno, W. J. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* **18**, 352–361 (2001).
53. Thorne, J. L. & Kishino, H. Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* **51**, 689–702 (2002).
54. Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226 (2006).
Develops a method to integrate the birth–death process to construct the time prior jointly with fossil calibrations with soft bounds. Introduces the limiting theory of uncertainty in divergence time estimates.
55. Lepage, T., Bryant, D., Philippe, H. & Lartillot, N. A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* **24**, 2669–2680 (2007).
56. Heath, T. A., Huelsenbeck, J. P. & Stadler, T. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proc. Natl Acad. Sci. USA* **111**, E2957–E2966 (2014).
57. dos Reis, M. & Yang, Z. Approximate likelihood calculation for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172 (2011).
58. Guindon, S. Bayesian estimation of divergence times from large sequence alignments. *Mol. Biol. Evol.* **27**, 1768–1781 (2010).
59. Yang, Z. *Molecular Evolution: A Statistical Approach* (Oxford Univ. Press, 2014).
60. Heath, T. A. & Moore, B. R. in *Bayesian Phylogenetics: Methods, Algorithms, and Applications* (eds Chen, M.-H, Kuo, L. & Lewis, P. O.) 277–318 (Chapman and Hall, 2014).
61. Gillespie, J. H. The molecular clock may be an episodic clock. *Proc. Natl Acad. Sci. USA* **81**, 8009–8013 (1984).
Proposes the idea of an episodic clock, modelling rate evolution through time and among lineages as a stochastic process.
62. dos Reis, M. & Yang, Z. The unbearable uncertainty of Bayesian divergence time estimation. *J. Syst. Evol.* **51**, 30–43 (2013).
63. Zhu, T., Dos Reis, M. & Yang, Z. Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. *Syst. Biol.* **64**, 267–280 (2015).
64. dos Reis, M., Zhu, T. & Yang, Z. The impact of the rate prior on Bayesian estimation of divergence times with multiple Loci. *Syst. Biol.* **63**, 555–565 (2014).
65. Warnock, R. C., Parham, J. F., Joyce, W. G., Lyson, T. R. & Donoghue, P. C. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. Biol. Sci.* **282**, 20141013 (2015).
66. Inoue, J., Donoghue, P. C. J. & Yang, Z. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* **59**, 74–89 (2010).
67. Sanderson, M. J. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**, 1218–1232 (1997).
68. Yang, Z. & Yoder, A. D. 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 (2003).
69. Aris-Brosou, S. & Yang, Z. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Mol. Biol. Evol.* **20**, 1947–1954 (2003).
70. Welch, J. J., Fontanillas, E. & Bromham, L. Molecular dates for the "cambrian explosion": the influence of prior assumptions. *Syst. Biol.* **54**, 672–678 (2005).
71. Heath, T. A., Holder, M. T. & Huelsenbeck, J. P. A. Dirichlet process prior for estimating lineage-specific substitution rates. *Mol. Bio. Evol.* **29**, 939–955 (2012).
72. Drummond, A. J. & Suchard, M. A. Bayesian random local clocks, or one rate to rule them all. *BMC Biol.* **8**, 114 (2010).
73. Huelsenbeck, J. P., Larget, B. & Swofford, D. A compound Poisson process for relaxing the molecular clock. *Genetics* **154**, 1879–1892 (2000).
74. Donoghue, P. C. & Benton, M. J. Rocks and clocks: calibrating the tree of life using fossils and molecules. *Trends Ecol. Evol.* **22**, 424–431 (2007).
75. Ho, S. Y. & Phillips, M. J. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* **58**, 367–380 (2009).
76. Goswami, A. & Upchurch, P. The dating game: a reply to Heads. *Zool. Scripta* **39**, 406–409 (2010).

77. Darwin, C. *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life* (John Murray, 1859).
78. Brunet, M. *et al.* A new hominid from the upper Miocene of Chad, central Africa. *Nature* **418**, 145–151 (2002).
79. Kistler, L. *et al.* Comparative and population mitogenomic analyses of Madagascar's extinct, giant 'subfossil' lemurs. *J. Hum. Evol.* **79**, 45–54 (2015).
80. Yoder, A. D. & Yang, Z. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol. Ecol.* **13**, 757–773 (2004).
81. Reisz, R. R. & Muller, J. Molecular timescales and the fossil record: a paleontological perspective. *Trends Genet.* **20**, 237–241 (2004).
82. Benton, M. J. & Donoghue, P. C. J. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53 (2007).
83. Warnock, R. C. M., Yang, Z. & Donoghue, P. C. J. Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* **8**, 156–159 (2012).
84. Parham, J. *et al.* Best practices for applying paleontological data to molecular divergence dating analyses. *Syst. Biol.* **61**, 346–359 (2012). **Sets out the criteria required for the establishment of fossil calibrations.**
85. Ksepka, D. T. *et al.* The fossil calibration database – a new resource for divergence dating. *Syst. Biol.* **64**, 853–859 (2015).
86. Marshall, C. R. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* **23**, 165–173 (1997).
87. Tavaré, S., Marshall, C. R., Will, O., Soligos, C. & Martin, R. D. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* **416**, 726–729 (2002).
88. Bracken-Grissom, H. D. *et al.* The emergence of lobsters: phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (decapoda: achelata, astacidea, glypheidea, polychelida). *Syst. Biol.* **63**, 457–479 (2014).
89. Sansom, R. S. & Wills, M. A. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Sci. Rep.* **3**, 2545 (2013).
90. Wood, H. M., Matzke, N. J., Gillespie, R. G. & Griswold, C. E. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. *Syst. Biol.* **62**, 264–284 (2013).
91. Sharma, P. P. & Giribet, G. A revised dated phylogeny of the arachnid order Opiliones. *Front. Genet.* **5**, 255 (2014).
92. Arcila, D. *et al.* An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Mol. Phyl. Evol.* **82**, 131–145 (2015).
93. Alexandrou, M. A., Swartz, B. A., Matzke, N. J. & Oakley, T. H. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Mol. Phyl. Evol.* **69**, 514–523 (2013).
94. Schrago, C. G., Mello, B. & Soares, A. E. Combining fossil and molecular data to date the diversification of New World Primates. *J. Evol. Biol.* **26**, 2438–2446 (2013).
95. Slater, G. J. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous–Palaeogene boundary. *Meth. Ecol. Evol.* **4**, 734–744 (2013).
96. Tseng, Z. J. *et al.* Himalayan fossils of the oldest known pantherine establish ancient origin of big cats. *Proc. Biol. Sci.* **281**, 20132686 (2014).
97. O'Reilly, J. E., Dos Reis, M. & Donoghue, P. C. Dating tips for divergence–time estimation. *Trends Genet.* **31**, 637–650 (2015).
98. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).
99. Metzker, M. L. Sequencing technologies – the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
100. Check Hayden, E. 10,000 genomes to come. *Nature* **462**, 21 (2009).
101. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. & Worm, B. How many species are there on Earth and in the ocean? *PLoS Biol.* **9**, e1001127 (2011).
102. Hipsley, C. A. & Muller, J. Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. *Front. Genet.* **5**, 138 (2014).
103. dos Reis, M. *et al.* Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* **25**, 2939–2950 (2015).
104. Gillespie, J. H. & Langley, C. H. Are evolutionary rates really variable? *J. Mol. Evol.* **13**, 27–34 (1979).
105. Angelis, K. & dos Reis, M. The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. *Curr. Zool.* **61**, 874–885 (2015).
106. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
107. King, C. E. & Jukes, T. H. Non-Darwinian evolution. *Science* **164**, 788–798 (1969).
108. Harris, H. Enzyme polymorphism in man. *Proc. R. Soc. B. Biol. Sci.* **164**, 298–310 (1966).
109. Lewontin, R. C. & Hubby, J. L. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595–609 (1966).
110. Haldane, J. B. S. in *Mathematical Proceedings of the Cambridge Philosophical Society* 838–844 (Cambridge Univ Press, 1927).
111. Kimura, M. Prepondence of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276 (1977).
112. Gillespie, J. H. *The Causes of Molecular Evolution* (Oxford Univ. Press, 1991).
113. Yang, Z. Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* **39**, 105–111 (1994).
114. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
115. Benton, M. J. *et al.* Constraints on the timescale of animal evolutionary history. *Palaeo. Electronica* 18.1.1FC (2015).
116. Gonzalez-Jose, R., Escapa, I., Neves, W. A., Cuneo, R. & Pucciarelli, H. M. Cladistic analysis of continuous modularized traits provides phylogenetic signals in Homo evolution. *Nature* **453**, 775–778 (2008).
117. Felsenstein, J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25**, 471–492 (1973).
118. Rambaut, A. Estimating the rate of molecular evolution: incorporating non-comptemporaneous sequences into maximum likelihood phylogenetics. *Bioinformatics* **16**, 395–399 (2000).
119. Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R. & Rodrigo, A. G. Measurably evolving populations. *Trends Ecol. Evol.* **18**, 481–488 (2003).
120. Stadler, T. & Yang, Z. Dating phylogenies with sequentially sampled tips. *Syst. Biol.* **62**, 674–688 (2013).
121. To, T. H., Jung, M., Lycett, S. & Gascuel, O. Fast dating using least-squares criteria and algorithms. *Syst. Biol.* syv068 (2015).
122. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
123. dos Reis, M., Tamuri, A. U., Hay, A. J. & Goldstein, R. A. Charting the host adaptation of influenza viruses. *Mol. Biol. Evol.* **28**, 1755–1767 (2011).
124. Korber, B. *et al.* Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789–1796 (2000).
125. Worobey, M. *et al.* Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**, 661–664 (2008).
126. Shapiro, B. *et al.* Rise and fall of the Beringian steppe bison. *Science* **306**, 1561–1565 (2004).
127. Orlando, L. *et al.* Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
128. Rybczynski, N. *et al.* Mid-Pliocene warm-period deposits in the High Arctic yield insight into camel evolution. *Nat. Commun.* **4**, 1550 (2013).
129. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
130. Orlando, L., Gilbert, M. T. & Willerslev, E. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* **16**, 395–408 (2015).
131. Schweitzer, M. H. *et al.* Biomolecular characterization and protein sequences of the Campanian hadrosaur B. canadensis. *Science* **324**, 626–631 (2009).
132. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comp. Biol.* **10**, e1003537 (2014).
133. Heath, T. A. A hierarchical Bayesian model for calibrating estimates of species divergence times. *Syst. Biol.* **61**, 793–809 (2012).
134. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
135. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
136. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
137. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
138. Thorne, J. L. & Kishino, H. in *Statistical Methods in Molecular Evolution* (ed. Nielsen, R.) 233–256 (Springer-Verlag, 2005).
139. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
140. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).

Acknowledgements
This work was supported by Biotechnology and Biosciences Research Council (UK) grant BB/J009709/1. M.d.R. wishes to thank the National Evolutionary Synthesis Center, USA, National Science Foundation #EF-0905606, for its support during his research on morphological evolution.

Competing interests statement
The authors declare no competing interests.