REVIEWS

THE MODERN MOLECULAR CLOCK

Lindell Bromham* and David Penny‡

The discovery of the molecular clock — a relatively constant rate of molecular evolution — provided an insight into the mechanisms of molecular evolution, and created one of the most useful new tools in biology. The unexpected constancy of rate was explained by assuming that most changes to genes are effectively neutral. Theory predicts several sources of variation in the rate of molecular evolution. However, even an approximate clock allows time estimates of events in evolutionary history, which provides a method for testing a wide range of biological hypotheses ranging from the origins of the animal kingdom to the emergence of new viral epidemics.

The molecular revolution has influenced all areas of biology. The human genome project has markedly reduced the time it takes to identify candidate disease genes, molecular markers are now a standard tool in conservation biology and agricultural techniques from pest control to improving crop yields depend on molecular methods. Although it is less in the public eye than these frontline applications, molecular data has also revolutionized the study of evolutionary biology. One way in which DNA evidence is transforming evolutionary biology is by providing a 'molecular clock' to put a new timescale on the history of life. The molecular clock arises from the observation that the amount of difference between the DNA of two species is a function of the time since their evolutionary separation. This provides a universal tool not only for placing past evolutionary events in time (BOX 1), but also for exploring the mechanisms and processes of evolution.

The evolutionary dates measured by molecular clocks have been controversial, particularly if they clash with estimates taken from more traditional sources such as the fossil record[1]. This conflict has been most evident in the debate over the timing of several important evolutionary events. One is the origin of the main types (phyla) of animal, which appear as an 'explosion' of fossil forms in the early Cambrian. Another is the replacement of dinosaurs by modern birds and mammals, traditionally placed in the early Tertiary following evidence of an extraterrestrial impact 65 million years ago (Mya) (BOX 2). Resolving the timing of these events

has clear implications for our understanding of the tempo and mode of important transitions in evolution. But the controversy over the molecular clock is not just the concern of evolutionary theorists: the practical benefits of the molecular clock are immense. For example, molecular clock dates that place the last common ancestor of the main pandemic strain of human immunodeficiency virus (HIV) in the 1930s have been used to counter claims that the virus was spread through the distribution of contaminated polio vaccine in the 1950s (allegedly manufactured using tissue from chimpanzees infected with simian immunodeficiency virus)[2].

How did DNA so rapidly become an irreplaceable tool in evolutionary biology? In this review we discuss the history of the molecular clock and the theory behind this unexpected and useful research tool. We examine how the observation of a molecular clock (a relatively constant rate of molecular change) has had two main effects on evolutionary biology. First, the constancy of rates has provided a window on the evolutionary process, by showing that molecular evolution is consistent with a neutral model rather than primarily driven by the action of natural selection. Second, because molecular change is expected to accumulate steadily over time, it provides a new tool for investigating the evolutionary history of biological lineages (BOX 1). We also discuss the problems that plague the accuracy of molecular date estimates, and new techniques that have been designed to overcome these limitations.

*Centre for the Study of Evolution,
School of Biological Sciences,
University of Sussex, Falmer,
Brighton BN1 9QG, UK.
‡Allan Wilson
Centre for Molecular Ecology and Evolution,
Massey University,
Palmerston North,
New Zealand.
Correspondence to L.B.
e-mail:
l.d.bromham@sussex.ac.uk

Box 1 | **Measuring evolutionary time with a molecular clock**

The basic approach for estimating molecular dates is to take a measure of the genetic distance between species, then use a calibration rate (the number of genetic changes expected per unit time) to convert the genetic distance to time. There are many available methods, ranging from a simple division of genetic distance by a calibration rate to more sophisticated MAXIMUM LIKELIHOOD[26, 66] or BAYESIAN APPROACHES[57, 64], which estimate molecular dates along with other parameters of models of the DNA substitution process. The reliability of all molecular clock methods depends on the accuracy with which genetic distance is estimated, and on the appropriateness of the calibration rate.

The simplest way to estimate genetic distance is to count the number of differences between the DNA (or protein) sequences for a particular gene or genes. However, this approach will fail to count 'multiple hits' (repeated changes), so a model of sequence evolution is needed to estimate the true number of substitutions that have occurred, using the observed number[67]. Selection of the substitution model governs the accuracy of genetic distance estimates and is therefore crucial to the accuracy of molecular dating. The substitution model should account for not only the relative frequency of different types of substitution, but also the variation in substitution rate between sites.

To use genetic distance to estimate divergence time, a calibration rate is needed that states the amount of genetic change expected per unit of evolutionary time. Because rates of change can vary between taxa and between genes, the calibration rate is usually calculated for each data set using a known date of divergence to estimate the rate for the whole phylogeny. The known date is most commonly taken from the fossil record, but other sources include biogeography (for example, the formation of an island[10]) or sample dates for bacteria and viruses[2,62,63]. Calibration dates might be used to interpolate (that is, to estimate divergences that fall between calibration points), but are more commonly applied to extrapolation (in which the divergence to be estimated falls beyond the calibration point). However, extrapolation has few constraints on the variances and biases, so uncertainty increases with the distance between the calibration point and the estimated node. The choice of calibration date is crucial to the accuracy of molecular dates, as shown by the difference in date estimates from studies using the same data and methods but different calibration dates[48]. The variation in rate of molecular evolution between lineages warns against using a calibration rate calculated for one lineage to date another, often distantly related, lineage[1,28,48]. The validity of particular calibration dates is frequently the subject of debate[1]. Sources of variability in fossil calibration points include poor sampling (by region, time period or taxon), uncertainty in the geological date of the fossil and errors in identification or phylogeny.

Discovery of the molecular clock

An early and unexpected finding from the molecular revolution was the discovery that a given protein has a characteristic rate of evolution, but that genes differ in their characteristic rates. Emile Zuckerkandl and Linus Pauling[3] compared protein sequences (particularly haemoglobins) from different species, plotting the number of amino-acid changes between the protein sequences against species age estimated from fossil evidence. They reported a strikingly linear rate of accumulation of amino-acid differences over evolutionary time. Other proteins also showed a constant rate of molecular evolution across species, but with each protein having a different characteristic rate: histones were exceptionally slow, cytochrome *c* was faster (but slower than haemoglobins) and fibrinopeptides were faster still. This relative equality of evolutionary rates for a given protein was unexpected; it had been assumed that, as with morphological evolution, there would be large variation in the rate of change both between species and over evolutionary time.

Motoo Kimura and Tomoko Ohta[4] explained the constant characteristic rate for each protein by suggesting that most amino-acid changes in a protein were effectively neutral — that is, changing the amino-acid sequence had no influence on the fitness of an organism and, therefore, the rate of change was not affected by natural selection. This differed from previous models that had considered that most amino-acid changes would be either favourable (positively selected) or deleterious

(removed by selection) (FIG. 1a). Kimura and Ohta reasoned that advantageous mutations would be relatively rare, deleterious mutations would be rapidly removed by selection and that a large proportion of possible amino-acid changes would have no practical effect on the functioning of the protein (FIG. 1b). The rate of accumulation of these neutral mutations would be influenced only by the mutation rate, and so would be relatively constant, as long as the base mutation rate remained unchanged (BOX 3). This is a fundamental result; it predicts that the long-term rate of neutral molecular evolution in species is the same as the neutral mutation rate in individuals.

If most amino-acid changes are neutral, or deleterious, then the predicted constant rate of protein change produces a molecular clock. But why do different proteins have different characteristic rates of change? The differences in characteristic rates could be explained by assuming that proteins varied in the proportion of amino-acid positions that were neutral (so that changing the amino acid had no selective effect) or constrained (so that any mutation was probably deleterious). Dickerson[5] compared what was then known about the protein structures of histones, cytochrome *c*, haemoglobins and fibrinopeptides, and concluded that their different rates of change could be explained by the proportion of neutral sites that each protein contained — the greater the proportion of neutral sites, the faster the rate of molecular evolution. According to this model, the rate of gene evolution is governed by the overall mutation rate and the proportion of sites at which changes are neutral[6] (FIG. 1b). In this way, the

MAXIMUM LIKELIHOOD
The maximum-likelihood method takes a model of sequence evolution (essentially a set of parameters that describe the pattern of substitutions) and searches for the combination of parameter values that gives the greatest probability of obtaining the observed sequences.

BAYESIAN APPROACH
A method that selects the tree that has the greatest posterior probability (the probability that the tree is correct), under a specific model of substitution.

neutral theory transformed comparative biochemistry. Previously, the amino acids that changed between species were considered important; now the constant sites are considered the most functionally important. Recent attempts to measure the proportion of amino-acid changes that are driven by selection might further inform this model of molecular evolution[7].

Initially, there was a controversy over whether neutral evolution was 'non-Darwinian'[8]. However, the two phenomena explained by the neutral theory — the rate of fixation of mutations and the high level of polymorphisms in natural populations — were predicted qualitatively by Darwin[9]: "Variations neither useful nor injurious would not be affected by natural selection, and would be left either a fluctuating element, as perhaps we see in certain polymorphic species, or would ultimately become fixed…". This is a concise and prescient statement of the neutral theory: that neutral mutations ("variations neither useful nor injurious") would result in polymorphisms ("fluctuating elements") that could become fixed in the population. The neutral theory also echoes Darwin's emphasis that variation arises constantly, irrespective of any "needs" of the organism, and his demonstration of the unity of the evolutionary processes that underlie changes at the level of species, phyla and kingdoms. So, the neutral theory has contributed significantly to the neo-Darwinian understanding of evolutionary processes.

The molecular clock is a 'sloppy' clock

The neutral theory predicts that, for a given mutation rate and proportion of neutral sites, the rate of molecular evolution should be constant. This constancy of rates is broadly supported by observation: the amount of divergence between genes tends to increase with the time since evolutionary separation, and in many cases the increase seems linear, as in the classic example of globin genes[3]. In a more recent example, the nucleotide distance between sister species on Hawaiian islands, plotted against geological estimates of island age, gave impressively linear relationships for both birds and fruitflies[10] (FIG. 2). However, empirical studies have also shown a great deal of variation in the rate of molecular evolution. The neutral theory allows for two sources of rate variation in the molecular clock: the 'sloppiness' of the 'tick rate' and changes in the mutation rate. These types of rate variation do not necessarily arise from different mechanisms. However, they do give rise to two types of error in molecular date estimates that contribute to 'residual effects' (unevenness of substitution rate in a lineage) and 'lineage effects' (variation in substitution rate between lineages) on the rate of molecular evolution[11].

A variable 'tick rate'

The molecular clock is probabilistic, not deterministic — it ticks at irregular intervals, not regularly as does a metronome. This distribution of intervals between substitution events ('ticks') — commonly described by a POISSON DISTRIBUTION[12] — adds large confidence intervals to date estimates because the time taken to produce the observed substitutions cannot be known precisely. In fact, the molecular clock is generally even sloppier: observations indicate that the pattern of substitution intervals for many genes has a broader distribution, which might be described as an 'overdispersed' Poisson distribution[11,13,14]. So, where does this extra variation come from?

Many processes could affect the number of substitutions per unit time, primarily by changing the balance between the relative influences of selection and drift, either for specific genes or sites in genes, or across whole genomes. The three-dimensional structure assumed by proteins has an important effect on the patterns of substitutions, as modelled by a structurally constrained neutral model[15]: although many sequences can fold to the same three-dimensional structure, different sequences have different numbers of 'neighbours' that they can mutate to and still maintain thermodynamic stability. Neighbouring sequences will tend to have similar proportions of variable sites: for

POISSON DISTRIBUTION
A discrete frequency distribution of the number of independent events per time interval, for which the mean value is equal to the variance.

MACROEVOLUTION
Evolution at, or above, the level of species; the patterns and processes of diversification and extinction of species over evolutionary time.

MICROEVOLUTION
The process of evolution in populations: changing allele frequencies over generations, due to selection or drift.

HOX GENE COMPLEX
A group of linked regulatory genes that are involved in patterning the animal body axis during development.

Box 2 | Controversial clocks for explosive radiations

Molecular clocks have been applied to a wide range of evolutionary questions, but none has attracted the degree of controversy that has surrounded molecular date estimates for several evolutionary radiations — the 'Cambrian explosion' of animal phyla and the radiations of modern birds and mammals in the early Tertiary[1]. These explosive radiations are part of a series of mass extinctions followed by rapid evolutionary radiations inferred from the global fossil record — a pattern that is considered to have fundamentally shaped the history of life. But molecular dates for the origins of animal phyla[27] and orders of mammals and birds[68] are generally twice as old as the dates of origin indicated by the fossil record. DNA evidence has been used to argue that these explosive radiations are an artefact of a disjunct fossil record: the early history of the animal kingdom, and modern birds and mammals, could simply be hidden from view. Conversely, many palaeontologists suggest that the discrepancy between molecular and palaeontological dates shows that molecular clocks can be systematically inaccurate, because some process associated with explosive radiations could speed the molecular clock[69,70] (although no supporting evidence has yet been found for this hypothesis[49]).

Resolving this controversy is important not only for assessing the reliability of molecular dates, but also for understanding the mechanisms of MACROEVOLUTION. Some researchers suggest that the pace and degree of change in explosive radiations is too great to be explained by Darwinian MICROEVOLUTIONARY processes, and that special mechanisms must be invoked, such as rapid change by developmental macromutations involving the HOX GENE COMPLEX[70,71]. Research into the dynamics of molecular evolution, and the development of more sophisticated molecular dating methods, will contribute to this important debate.

---

example, sequences for which many sites can change and maintain three-dimensional structure will tend to have neighbouring sequences that also have many variable sites, and so neighbours will tend to have similar rates of molecular evolution. This autocorrelation between substitution rates for neighbouring sequences will also show a higher variance in the number of substitutions per unit time than is expected from a simple Poisson process. As the three-dimensional structure of a protein evolves, the position and proportion of variable sites in the sequence is expected to change over time — a process described by the covarion model of protein evolution[6,16].

In addition, the strength of selection on particular sites can change, producing bursts of substitutions as a molecule is adapted to a new role or responds to changes in another part of the genome. For example, ruminants (such as cattle) and leaf-eating langur monkeys have independently evolved herbivory, exploiting a food resource that poses biochemical challenges. They have co-opted several existing enzymes (such as lysozyme and ribonuclease) into the task of leaf digestion and the genes that code for these proteins show a burst of amino-acid changes as the proteins evolve into their new role[17,18].

Not only can particular sites in a gene undergo a change in selection pressure, but also the selection coefficient of an entire gene can change. This is most obvious in pseudogenes — copies of a gene that have been rendered non-functional by mutation. Thereafter, any change to the non-functional sequence has no further effect on fitness, so all substitutions are neutral. Pseudogenes, and other sequences with no effect on fitness, have a fast rate of molecular evolution because most changes to the sequence are allowed. Similarly, the patterns of selection in duplicated copies of a gene might change as they evolve new roles, and this might be reflected in the rates of molecular evolution. Although it is possible that the overall rate of adaptation or morphological change of a species might influence the molecular clock, so far there has been little evidence to support this proposition from either experimental studies[19] or comparative analyses[20,21].

Another source of variability comes from the effect of population size on the rate of fixation of mutations. Tomoko Ohta[22,23] extended the neutral theory by recognizing the critical role of effective population size (FIG. 1c ; BOX 4). Small populations are more severely affected by stochastic fluctuations in allele frequencies, so genetic drift (random sampling error on allele frequencies) can overpower selection for alleles with small selection coefficients. Therefore, the fixation of nearly-neutral alleles of small selective effect is expected to be greatest in small populations[23,24]. So, if a population undergoes a marked reduction in population size due to an environmental catastrophe, this event might be accompanied by a burst of fixation of nearly-neutral alleles. In this way, population fluctuations might add to the sloppiness of the molecular clock.
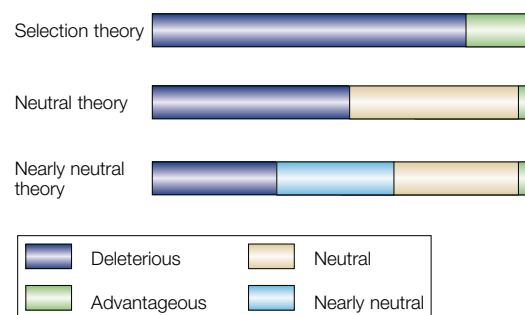
What is the practical effect of the unevenness of

Figure 1 | **Selectionist, neutral and nearly neutral theories. a** | Selectionist theory: early neo-Darwinian theories assumed that all mutations would affect fitness and, therefore, would be advantageous or deleterious, but not neutral. **b** | Neutral theory: the neutral theory considered that, for most proteins, neutral mutations exceeded those that were advantageous, but that differences in the relative proportions of neutral sites would influence the rate of molecular evolution (that is, more neutral sites would produce a faster overall rate of change) (BOX 3). **c** | Nearly neutral theory: the fate of mutations with only slight positive or negative effect on fitness will depend on how population size affects the outcome (BOX 4). Figure modified with permission from REF. 22.

substitution rate? Some researchers assume, as a working model, that fluctuations in rate will be scattered randomly over the phylogeny. If so, then local variations in substitution rate might average out over long timescales, and simply add noise to the molecular clock, making date estimates imprecise. Molecular clock techniques can be designed to express the sloppiness that arises from the variation in substitution intervals by modelling the substitution process as a Poisson process[25,26]. The confidence intervals around the date estimates using this technique can be very large, showing the importance of considering variability when using molecular clocks to estimate dates of divergence[27]. But, in addition to imprecision arising from the variability in the substitution process, the tick rate of the clock can also vary consistently between different species, such that some branches of the phylogeny have a faster rate of molecular evolution than others. In this case, molecular date estimates could be systematically inaccurate[28].

Different rates for different groups

Although it is broadly true that the longer two species are separated, the more protein or DNA differences will accumulate between them, the nearly neutral theory of molecular evolution (BOX 4) leads us to expect that the rate of molecular evolution can vary in three ways: through changes to mutation rate, population size or selective coefficients. Because these factors can differ between species or over time, they give rise to lineage effects (variation in substitution rate between species)[11].

Mutation rate clearly varies between taxa[29] and much of this variation is due to differences in repair equipment (biochemical factors). Mutations arise both through errors during DNA replication, and through

damage that is not repaired. Both of these processes are governed by an array of enzymes that can vary in repair efficiency between species (and might have different effects on different types of mutation[30]). RNA viruses provide an extreme example: they copy their genomes using highly error-prone RNA polymerases or reverse transcriptases that lack proofreading function. This contributes to the high rates of molecular evolution in the retrovirus HIV — around a million times faster than the rate of evolution of mammalian genomes, which use a battery of replication and repair enzymes to reduce the mutation rate.

But DNA-repair efficiency varies even among species with similar genomes. For example, murid rodents (rats, mice and hamsters) have less efficient excision-repair mechanisms than humans[31] and this is consistent with the higher rate of molecular evolution in murid rodents than in hominids[32]. Even within mammalian cells, different replication and repair enzymes vary in mutation rate: mitochondrial DNA is copied by DNA polymerase-γ, which has a higher error rate than other mammalian DNA polymerases, and this contributes to the higher mutation rate of mitochondrial genes over nuclear genes. Furthermore, the error rate of repair enzymes can be altered by mutation, leading to variants with higher or lower rates of mutation, which can increase in frequency due to drift, or could be selected for under some conditions.

Because the mutation rate is affected by the biochemical properties of replication and repair enzymes, it can be altered by mutations in these enzymes that make them more, or less, error prone. This variation in mutation rate between individuals can be selected for under certain conditions. Laboratory experiments on bacteria show that a 'mutator' allele, which raises the mutation rate, can be indirectly selected for because it increases the chance of an advantageous mutation arising[33]. But high mutation rates carry the cost of an increased number of deleterious mutations, leading to a compromise mutation rate that balances the costs and benefits of mutation and repair. The mutation rates of complex eukaryotes also show evidence of being shaped by selection, which balances the costs of improved accuracy and repair against the costs of deleterious mutations, as DNA-repair rates are neither at the maximum achievable level nor at the minimum toler-

able level, and it is possible that different species evolve different optimum mutation rates.

The mutation rate can also be influenced by the LIFE HISTORY of a species and, potentially, by environmental variables. If by-products of the metabolic process — such as oxygen radicals — have a mutagenic effect on DNA[34], then species with higher metabolic rates might generate higher concentrations of mutagens and incur more DNA damage[35]. This is consistent with the observations that ECTOTHERMS have lower substitution rates than ENDOTHERMS[36], that smaller-bodied species of vertebrates with faster metabolic rates have higher substitution rates than larger-bodied species[35,37–39] and that animal mitochondria, which are at the site of oxidative phosporylation, have a higher rate of molecular evolution than nuclear rates[34] (it would be interesting to extend these studies to plants, in which the rate of mitochondrial and chloroplast evolution seems not to be faster than nuclear evolution). However, there is, as yet, no evidence that environmental temperature influences the rate of molecular evolution[40,41].

Even species with the same basic DNA replication efficiency per cell division could vary in the rate of accumulation of DNA copy errors. For a given rate of copy error, a species that copies its DNA more frequently per unit time (for example, mice, which can produce offspring within three months of birth) will accumulate more copy-error mutations per unit time than a species with a longer generation time (for example, humans, who take more than a decade to reach maturity). In this sense, the mutation rate might be best measured as a per-generation rate, particularly in animals that have separate germline and somatic cells. The predicted correlation of generation time with mutation rate[24] has been supported by studies that compare the rates of molecular evolution with generation time (or its correlate, body size) across a range of genes for several vertebrate species[35,37–39]. The generation time effect is complicated by the relative contribution of DNA damage to the mutation rate — DNA damage that accumulates between cell divisions could be independent of division rate, even though it might only be repaired during cell proliferation[42].

The nearly neutral theory also predicts that any consistent effect on effective population size could influence substitution rate (BOX 4). These effects include ecological factors (such as a reduced population size on

LIFE HISTORY
The reproductive strategy of an organism.

ECTOTHERM
A 'cold-blooded' organism, such as a reptile, for which body temperature is dependent on the environment.

ENDOTHERM
A 'warm-blooded' organism, such as a mammal or bird, for which body temperature is maintained independently of the environment.

EFFECTIVE POPULATION SIZE
(N_e). The equivalent number of breeding adults in a population after adjusting for complicating factors, such as reproductive dynamics. It is usually less that the actual number of living or reproducing individuals (the census population size N).

FIXATION
An increase in allele frequency to the point at which all individuals in a population are homozygous.

Box 3 | The neutral theory and the molecular clock

According to the neutral theory of Kimura[4,72] (FIG. 1b), the overall mutation rate (μ) is the sum of deleterious (μ^-), neutral (μ^o), and positive (μ^+) mutations, $\mu = \mu^- + \mu^o + \mu^+$. The theory focused on neutral mutations because advantages mutations (μ^+) were considered relatively rare, and in large populations deleterious mutations (μ^-) would be eliminated by negative selection. The mutation rate is expressed per individual per unit time — which might be per generation, per year or per cell replication. The other parameter required is the EFFECTIVE POPULATION SIZE (N_e). For a haploid taxon, the number of mutations per time period $= N_e \cdot \mu^o$, the probability of fixation of a neutral mutation $= 1/N_e$ and, therefore, the number of neutral mutations fixed per time period $= N_e \cdot \mu^o/N_e = \mu^o$.

In other words, although a larger population produces more mutations, the probability of a specific mutation being FIXED into the population declines proportionally with population size. So, according to a neutral model, population size cancels out to leave the molecular evolution rate determined by the mutation rate (μ^o). For diploids, the mutations double ($2N_e$), but the probability of fixation halves ($1/2N_e$) and so population size still cancels out.

---

islands[43]), correlates of life-history evolution (such as inbreeding or EUSOCIALITY[44]) and aspects of niche or lifestyle (such as endoparasitism[45]). It is possible that the population size effect could also account for the intriguing association between the diversification rate of flowering plant lineages (speciation minus extinction) and rates of molecular evolution[46]. Given the lack of a relationship between phenotypic evolution and molecular rates[19–21], it seems unlikely that the observed faster rate of molecular evolution in more rapidly diversifying lineages is a result of species adapting to new niches. However, it could be related to population size effects if rapidly speciating lineages undergo frequent population subdivision[40].

How reliable is the molecular clock?

The molecular clock allows a valuable insight into the biochemical, selective and population processes that underlie genetic evolution. It also provides a remarkable tool for investigating evolutionary history: if the rate of molecular evolution is relatively constant, then the amount of genetic difference between two species gives a measure of the time since their evolutionary separation (BOX 1). This molecular timescale can provide insights into the history of all organisms from which we can obtain genetic sequences. This is valuable in the case of organisms with no fossil record — such as viruses — or for which the sampling of the fossil record is patchy in time or space. Because of the universality of DNA (or RNA), molecular clocks can reach all timescales of evolution, from population divergences to the last common ancestor of the five kingdoms of life.

But this apparently simple technique has resulted in some controversial molecular dates. Some seem to contradict other lines of evidence, such as the study that produced a molecular date estimate for the split between kingdoms that is markedly younger than the earliest fossils[47]. Other molecular dates have been used to challenge the reliability of the fossil record, and engendered debate over the tempo and mode of macroevolution, such as the discrepancy between molecular and fossil dates for the origin of animal phyla (BOX 2). In some cases, molecular dates produced by different methods

contradict each other[1,48,49]. These controversial cases should act as a warning when interpreting molecular dates. This is problematic when molecular clocks are applied to groups for which no other historical information is available for comparison, such as the origins of new emerging diseases such as HIV (REF. 2).

Molecular evolutionary theory leads us to expect two types of error in molecular clock estimates. First, the sloppy nature of the substitution rate results in large variance around the amount of genetic difference expected for any given time period, adding a large degree of imprecision to molecular date estimates[27]. Second, the nearly neutral theory predicts that the rate of molecular evolution is influenced by mutation rate, population size and the relative proportions of sites with different selective coefficients; these factors might differ between genes, between species and over time, potentially resulting in consistent bias in date estimates[28]. Molecular dating techniques must take into account both the imprecision and potential bias of date estimates.

Problems with the accuracy and precision of molecular clocks can be demonstrated by comparing studies that use different genes, calibrations or methods, and produce different date estimates for the same divergence[48]. For example, published molecular dates for the split between rodents and primates vary by as much as 100% (REF. 48). Part of the difficulty of assessing the reliability of molecular clocks is that molecular dates are frequently presented as point estimates, or with sample errors that represent the difference between estimates in the same study. Few studies present molecular date estimates with confidence intervals that accurately portray the variance in the clock due to the sloppiness of the tick rate, or the lineage variation in rates. Such confidence intervals allow molecular dates to be used to test evolutionary hypotheses within the bounds of the accuracy and precision of molecular clocks, by asking whether the range of possible date estimates is consistent with a specific evolutionary hypotheses[26] (for example, that the animal phyla arose in the early Cambrian[27]) (BOX 2). They would also allow the identification of cases for which molecular date estimates from differ-

EUSOCIALITY
A life-history strategy in which only a subset of members of a group produce their own offspring, and others act as non-reproductive helpers, as in honeybees or naked molerats.

GENETIC DRIFT
The random fluctuation that occurs in allele frequencies as genes are transmitted from one generation to the next. This is because allele frequencies in any sample of gametes that perpetuate the population might not represent those of the adults in the previous generation.

VARIANCE
A measure of the variation around the central class of a distribution (the average squared deviation of the observations from their mean value).

Box 4 | The nearly neutral theory: population size and the molecular clock

Tomoko Ohta[22–24] provided an important extension of the neutral theory by recognizing the crucial role of effective population size in the influence of selection. Slightly deleterious mutations will tend to be removed by selection in very large populations; however, they can be fixed by chance in smaller populations in which selection can be overpowered by random sampling events (GENETIC DRIFT). Ohta showed that the fixation of these nearly neutral mutations with small selection coefficients (s), whether positive or negative, would be governed by chance events in small populations, just as if they were neutral. So, a mutation would be effectively neutral if $|s| < 1/4N_e$, where N_e is the effective population size. In other words, whether a mutation behaves according to the neutral expectation is determined not simply by the selection properties of the mutation (s), but also the size of the population in which it arises (N_e). Ohta then considered the fate of a range of slightly deleterious or slightly advantageous mutations with a VARIANCE on their selection coefficients (σ_s) — some will be slightly deleterious and some slightly advantageous. Mutations are divided into three categories: mutations for which selection is the predominant force ($4N_e\sigma_s > 3$); nearly neutral mutations, which are governed by both selection and drift ($3 \geq 4N_e\sigma_s \leq 0.2$); and effectively neutral mutations, the fate of which is determined only by drift ($4N_e\sigma_s < 0.2$). So, the nearly neutral theory describes how the rate of molecular evolution can vary not only with changes in the mutation rate, but also through the changing balance between selection and drift.
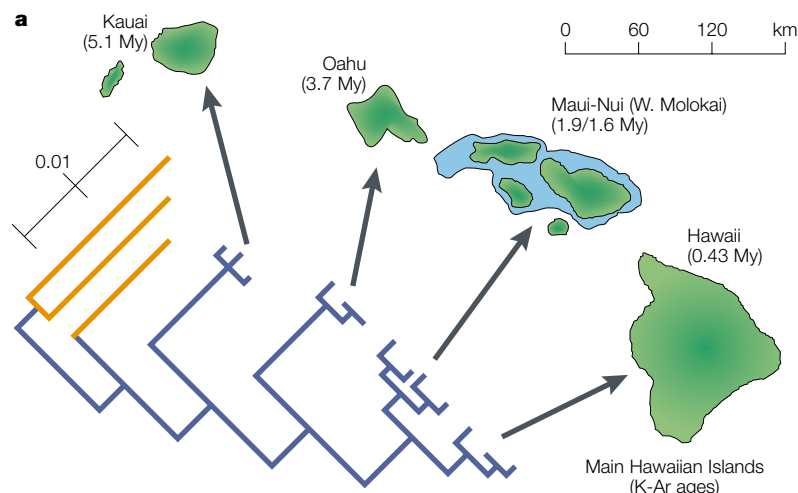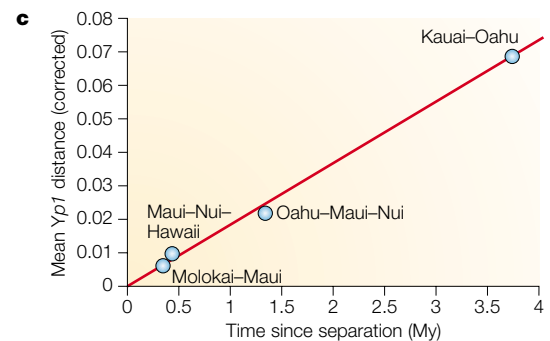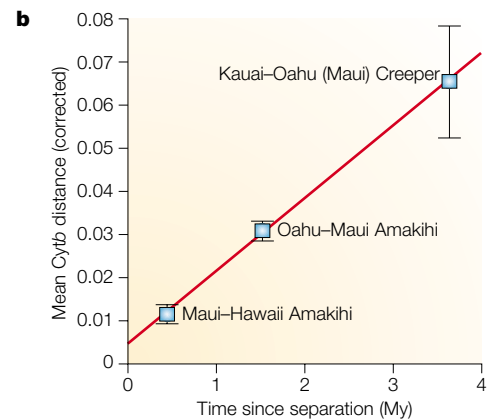

Figure 2 | **A molecular clock for the Hawaiian islands. a** | The volcanic origin of the Hawaiian islands has produced a chain of islands of increasing geological age. The phylogenetic relationships of island endemic birds (for example, the drepananine (honeycreeper) species such as the amakihi, *Hemignathus virens* and the akiapolaau *Hemignathus wilsoni*, shown in the tree) and fruitflies (*Drosophila* spp.) reflect this volcanic 'conveyer belt', with the species of the oldest islands forming the deepest branch of the tree, and the younger islands on the tips of the tree. Orange lines represent the outgroups. **b,c** | Molecular dates for *Hemignathus* (panel **b**) and *Drosophila* (panel **c**) confirm this order of colonization, and produce a remarkably linear relationship between genetic divergence and time when DNA distance is plotted against island age. My, million years. Figures reproduced with permission from REF. 10 © (1998) Blackwell Publishing.

ent studies conflicted significantly, rather than simply varied within the expected bounds of error.

The high variance that arises from residual effects makes date estimates imprecise, but lineage effects, in which different species have different characteristic rates, might bias date estimates — that is, estimates based on the assumption of a constant rate might produce consistently over- or under-estimated dates. For example, the substitution rate in murid rodents is, on average, two to three times faster than the rate in hominid primates (apes and humans)[32]. This difference is probably due to a combination of factors, including differences in DNA-repair mechanisms, generation time, metabolic rate and effective population size. If the genetic distance between rodents and primates is used to estimate their date of divergence by assuming a constant rate of molecular evolution, this rate difference can result in molecular dates that are consistently too old[28].

Compounding these residual and lineage errors in the clock are operational errors in measuring the rate of molecular evolution, which include uncertainty in the relationships between taxa (the evolutionary tree), errors in calibration points and uncertainties in the mechanism of evolution[1,49]. All molecular clock methods rely on accurate estimates of the genetic distance between sequences, and, in some cases, this measurement might be biased. However, the molecular clock is an irreplaceable source of information in evolutionary biology and it would be foolish to abandon it altogether. The challenge is to develop methods that enable this valuable source of historical information to be

exploited, while allowing for the known sources of variation and recognizing the limits of precision.

Estimating dates when rates vary
Several approaches have been taken to using the molecular clock when rates differ between species. The first, and most common, approach is to search for genes that have a uniform rate across all taxa under consideration, by using a 'clock test' (such as the RELATIVE RATES TEST[50], the TAJIMA TEST[51] and the LIKELIHOOD RATIO TEST[25]) to detect and reject sequences that show rate variation between species[27,52–54]. However, these clock tests lack power for shorter sequences (or genes with low rates of change) and will detect only a relatively low proportion of cases of rate variation for the types of sequence that are typically used in molecular clock studies[28]. Failure to detect rate variation can cause systematic error in date estimates, because undetected rate variation can lead to consistently over-estimated dates of divergence[28]. The move towards using much longer sequences might improve the usefulness of this approach in future.

An alternative approach to seeking rate-constant genes is to recognize the degree of rate variation in the data and to use this information to put limits on the possible divergence dates[55]. Estimating maximum and minimum possible rates for a given data set requires several calibration dates, which are not always available, and gives conservative assessments of molecular dates that are suitable only for rejecting particular hypotheses as being inconsistent with the data[55]. A more promising approach is to explicitly model rate variation over the

phylogeny[56,57]. Bayesian methods[57], which allow correlated rate changes between the nodes of a tree according to a specified model, are now generating much excitement in the field. For example, a recent study used Bayesian inference to propose a coordinated increase in substitution rate across all early animal lineages[58], potentially explaining discrepancies between molecular and palaeontological dates for the origin of animal phyla (BOX 2) — although no mechanisms that could produce such a marked concerted change were offered. However, the method relies on the validity of many assumptions, some of which might be inappropriate for the data being tested (such as the assumption that a constant rate of lineage origination and extinction, counter to the proposed 'explosion') (BOX 2). In addition, modelling too many parameters can lead to a loss of power to distinguish between variants of a model[59]. The next few years will undoubtedly see important advances in the development of molecular clock methods that allow variation in rate between taxa.

Ideally, the application of rate-variable molecular clocks to estimating divergence dates would be made more reliable by understanding the processes that generate differences in substitution rate. Molecular evolutionary theory has led to the prediction that species can differ in their rate of molecular evolution through changes to traits that influence mutation rate, population size or selective coefficients, but the measurable effect of these predicted influences has only recently begun to be investigated for a wide range of taxa. Growth in genetic databases will increasingly provide the power to uncover the effect of determinants such as life history[35–39], behaviour[44] and environment[40,41] on the molecular clock, which can then be used to inform methods for estimating evolutionary divergence dates from DNA sequences.

Molecular clocks for emerging diseases

One of the most practical applications of molecular clocks has been in the study of emerging viral diseases. Estimating dates for the origin of viral lineages can be the key to uncovering the cause of disease emergence[2,60], yet viruses leave no fossil record and their origins are often obscure. Molecular data allow the reconstruction of the history of epidemics from extant virus samples. Viral molecular clocks are most commonly calibrated using past samples, such as sequences from stored tissue samples[2,63]. In the case of endogenous retroviruses (embedded in the genome of their host), dates of origin can be estimated without a calibration date by comparing the two long terminal repeats (LTRs) that flank the viral genome — a method that has helped uncover the history of viral DNA in the human genome[60,61]. If many calibration dates are available — for example, if a series of samples from a single infected patient[62] or a collection of archived samples[63] is available — it becomes possible to test the constancy of rates across the viral phylogeny. Several dates can be used to fit a linear relationship between genetic distance and time, which might subsequently be used to estimate

the rate of molecular evolution[62] or the time of origin of the last common ancestor of all included samples[2]. This method runs into statistical problems if non-independent comparisons are used in the regression; however, this problem can be overcome in some cases by using a phylogenetic comparative method[64].

Viral molecular clocks also show how the degree of clock like sequence evolution varies between different taxa or genes. For some viruses — such as Dengue fever[63] or influenza[65] — observable constancy of rates provides a convincing molecular clock. But for other viruses, complex processes of molecular evolution, such as recombination and selection, cast doubt on the information value of genetic distances as a measure of time[2]. Several rate-variable molecular clock methods have been developed and applied to viral sequences with serial sample dates — for example, using maximum likelihood to optimize the substitution rate over a phylogeny of sequences with known isolation dates to estimate the unknown internal nodes[66]. A likelihood ratio test is used to compare the fit of a single substitution rate with a multiple-rate model. Similar methods have been developed that use Bayesian statistics to select the most likely parameters of molecular evolution for a set of viral sequences, including the substitution rates and timing of ancestral nodes[64].

Dating with confidence

The molecular clock — a relatively constant rate of accumulation of molecular differences between species — was an unexpected discovery that has provided a window on the mechanisms that drive molecular evolution. It has also provided a universal tool for investigating the evolutionary past and processes. However, both theory and observation show that the molecular clock is much more complex than was initially supposed. We cannot expect a universal linear relationship between distance and time because of the many factors that can influence the rate of molecular evolution. This complexity throws up a challenge to develop new methods to allow historical information to be extracted from molecular data, allowing for both the sloppiness of the tick rate of the molecular clock and for the variation in molecular rate between species. There are three approaches to allowing for the variation in the rate of molecular evolution. First, molecular date estimates should be presented with confidence intervals that accurately portray the variance in the rate of molecular evolution, both in and between lineages. Second, new molecular clock methods that incorporate variation in the rate of molecular evolution should be developed. Third, an understanding of the mechanisms that generate the rate variation will inform judgement of the reliability of molecular date estimates (including the identification of cases to which molecular clocks cannot be reasonably applied). The next decade will see a growth in increasingly sophisticated methods that will allow the molecular clock to be applied to an ever-widening set of biological questions.

1. Smith, A. B. & Peterson, K. J. Dating the time of origin of major clades: molecular clocks and the fossil record. *Annu. Rev. Earth Planet. Sci.* **30**, 65–88 (2002).
   A review of the controversy surrounding dates for the Cambrian explosion of animal phyla and the early Tertiary radiation of modern mammals and birds. Written by a palaeontologist and a molecular geneticist, this review takes a critical look at the reliability of both fossil and molecular dates.
2. Korber, B. *et al.* Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789–1796 (2000).
3. Zuckerkandl, E. & Pauling, L. in *Horizons in Biochemistry* (eds Kasha, M. & Pullman, B.) 189–225 (Academic Press, New York, 1962).
4. Kimura, M. & Ohta, T. On the rate of molecular evolution. *J. Mol. Evol.* **1**, 1–17 (1971).
5. Dickerson, R. E. The structure of cytochrome c and rates of molecular evolution. *J. Mol. Evol.* **1**, 26–45 (1971).
6. Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *J. Mol. Evol.* **53**, 711–723 (2001).
7. Smith, N. H. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024 (2002).
8. King, J. L. & Jukes, T. H. Non-Darwinian evolution. *Science* **164**, 788–798 (1969).
9. Darwin, C. *The Origin of Species by Means of Natural Selection* 6th edn Ch. 4 **64** (John Murray, London, 1872).
   Remarkably prescient exposition of the processes of evolution, including a pre-genetic description of the neutral theory, pre-emptively rebutting rumours that neutral theory is 'non–Darwinian'.
10. Fleischer, R. C., McIntosh, C. E. & Tarr, C. L. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-Ar based ages of the Hawaiian islands to estimate molecular evolutionary rates. *Mol. Ecol.* **7**, 533–545 (1998).
11. Gillespie, J. H. *The Causes of Molecular Evolution* (Oxford University Press, Oxford, UK, 1991).
12. Zheng, Q. On the dispersion index of a Markovian molecular clock. *Math. Biosci.* **172**, 115–128 (2001).
    This gives a statistical view of the expected variability in rates that occur when the simple probabilistic models of molecular evolution are allowed to increase in complexity.
13. Bickel, D. R. Implications of fluctuations in substitution rates: impact on the uncertainty of branch lengths and on relative-rate tests. *J. Mol. Evol.* **50**, 381–390 (2000).
14. Cutler, D. J. Estimating divergence times in the presence of an overdispersed molecular clock. *Mol. Biol. Evol.* **17**, 1647–1660 (2000).
15. Bastolla, U., Porto, M., Roman, H. E. & Vendruscolo, M. Lack of self-averaging in neutral evolution of proteins. *Phys. Rev. Lett.* **89**, article no. 208101 (2002).
    This original paper follows the evolution of protein sequences that are restricted in their predicted tertiary structure. It shows, using basic biochemical principles, that the variability in rates of a molecular clock is expected to be higher than for a simple Poisson process.
16. Fitch, W. M. Rate of change of concomitantly variable codons. *J. Mol. Evol.* **1**, 84–96 (1971).
17. Swanson, K. W., Irwin, D. M. & Wilson, A. C. Stomach lysozyme gene of the langur monkey: tests for convergence and positive selection. *J. Mol. Evol.* **33**, 418–425 (1991).
18. Zhang, J. Z., Zhang, Y. P. & Rosenberg, H. F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genet.* **30**, 411–415 (2002).
19. Papadopoulos, D. *et al.* Genomic evolution during a 10,000-generation experiment with bacteria. *Proc. Natl Acad. Sci. USA* **96**, 3807–3812 (1999).
    A laboratory experiment comparing rates of morphological and molecular evolution in bacterial populations. Although adaptive phenotypic evolution was fastest at the beginning, DNA substitutions accumulated steadily through the experiment, indicating that the molecular clock is decoupled from the pace of adaptive evolution.
20. Bromham, L., Woolfit, M., Lee, M. S. Y. & Rambaut, A. Testing the relationship between morphological and molecular rates of change along phylogenies. *Evolution* **56**, 1921–1930 (2002).
21. Wyles, J. S., Kunkel, J. G. & Wilson, A. C. Birds, behavior, and anatomical evolution. *Proc. Natl Acad. Sci. USA* **80**, 4394–4397 (1983).
22. Ohta, T. & Kimura, M. On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* **1**, 18–25 (1971).
23. Ohta, T. Very slightly deleterious mutations and the molecular clock. *J. Mol. Evol.* **26**, 1–6 (1987).
24. Ohta, T. Near-neutrality in evolution of genes and gene regulation. *Proc. Natl Acad. Sci. USA* **99**, 16134–16137 (2002).
    The most recent exposition of the nearly-neutral model, in which the effects of weak selection depend both on the selection coefficient of the mutation and the size of the population in which the mutant occurs.
25. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376 (1981).
26. Rambaut, A. & Bromham, L. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**, 442–448 (1998).
27. Bromham, L., Rambaut, A., Fortey, R., Cooper, A. & Penny, D. Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proc. Natl Acad. Sci. USA* **95**, 12386–12389 (1998).
28. Bromham, L. D., Rambaut, A., Hendy, M. D. & Penny, D. The power of relative rates tests depends on the data. *J. Mol. Evol.* **50**, 296–301 (2000).
29. Drake, J., Charlesworth, B., Charlesworth, D. & Crow, J. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
    Observable mutation rates, when measured per genome per generation, are remarkably similar across widely divergent organisms, indicating that natural selection might shape optimum mutation rates.
30. Ota, R. & Penny, D. Estimating changes in mutational mechanisms of evolution. *J. Mol. Evol.* (in the press).
31. Hart, R. W. & Setlow, R. B. Correlation between deoxyribonucleic acid excision-repair and life-span in a number of mammal species. *Proc. Natl Acad. Sci. USA* **71**, 2169–2173 (1974).
32. Li, W.-H., Ellesworth, D. L., Krushkal, J., Chang, B. H.-J. & Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol. Phylogenet. Evol.* **5**, 182–187 (1996).
33. Chao, L. & Cox, E. C. Competition between high and low mutating strains of *Escherichia coli*. *Evolution* **37**, 125–134 (1983).
34. Rand, D. M. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends Ecol. Evol.* **9**, 125–131 (1994).
35. Martin, A. P. & Palumbi, S. R. Body size, metabolic rate, generation time and the molecular clock. *Proc. Natl Acad. Sci. USA* **90**, 4087–4091 (1993).
    Showed a relationship between body size and the rate of molecular evolution for vertebrates using estimates of absolute substitution rates. This paper showed that the life history of a species must influence the rate of molecular evolution.
36. Martin, A. P. Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. *Mol. Biol. Evol.* **12**, 1124–1131 (1995).
37. Bromham, L., Rambaut, A. & Harvey, P. H. Determinants of rate variation in mammalian DNA sequence evolution. *J. Mol. Evol.* **43**, 610–621 (1996).
38. Bromham, L. Molecular clocks in reptiles: life history influences rate of molecular evolution. *Mol. Biol. Evol.* **19**, 302–309. (2002).
39. Mooers, A. Ø. & Harvey, P. H. Metabolic rate, generation time and the rate of molecular evolution in birds. *Mol. Phylogenet. Evol.* **3**, 344–350 (1994).
40. Bromham, L. & Cardillo, M. Testing the link between the latitudinal gradient in species richness and rates of molecular evolution. *J. Evol. Biol.* **16**, 200–207 (2003).
41. Held, C. No evidence for slow-down of molecular substitution rates at subzero temperatures in Antarctic serolid isopods (Crustacea, Isopoda, Serolidae). *Polar Biol.* **24**, 497–501 (2001).
42. Bielas, J. H. & Heddle, J. A. Proliferation is necessary for both repair and mutation in transgenic mouse cells. *Proc. Natl Acad. Sci. USA* **97**, 11391–11396 (2000).
43. Johnson, K. P. & Seger, J. Elevated rates of nonsynonymous substitution in island birds. *Mol. Biol. Evol.* **18**, 874–881 (2001).
44. Schmitz, J. & Moritz, R. F. A. Sociality and the rate of rDNA sequence evolution in wasps (Vespidae) and honeybees (*Apis*). *J. Mol. Evol.* **47**, 606–612 (1998).
45. Moran, N. A. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA* **93**, 2873–2878 (1996).
46. Barraclough, T. G. & Savolainen, V. Evolutionary rates and species diversity in flowering plants. *Evolution* **55**, 677–683 (2001).
47. 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).
48. Bromham, L. D., Phillips, M. J. & Penny, D. Growing up with dinosaurs: molecular dates and the mammalian radiation. *Trends Ecol. Evol.* **14**, 113–118 (1999).
49. Bromham, L. Molecular clocks and explosive radiations. *J. Mol. Evol.* (in the press) .
50. Wu, C.-I. & Li, W.-H. Evidence for higher rates of nucleotide substitutions in rodents than in man. *Proc. Natl Acad. Sci. USA* **82**, 1741–1745 (1985).
51. Tajima, F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607 (1993).
52. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
53. Nei, M. & Glazko, G. V. Estimation of divergence times for a few mammalian and several primate species. *J. Hered.* **93**, 157–164 (2002).
54. Takezaki, N., Rzhetsky, A. & Nei, M. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**, 823–833 (1995).
55. Bromham, L. D. & Hendy, M. D. Can fast early rates reconcile molecular dates to the Cambrian explosion? *Proc. R. Soc. Lond. B* **267**, 1041–1047 (2000).
56. Sanderson, M. J. A nonparametric approach to estimating divergence times in the absence of rate constancy. *J. Mol. Evol.* **14**, 1218–1231 (1997).
57. 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).
    This paper outlined new Bayesian methods for estimating dates of divergence if rates of molecular evolution vary between lineages, by allowing the mutation rate to vary with time, and averages its estimates over a range of alternatives.
58. Aris-Brosou, S. & Yang, Z. 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 (2002).
59. Rannala, B. Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* **51**, 754–760 (2002).
60. Bromham, L. The human zoo: endogenous retroviruses in the human genome. *Trends Ecol. Evol.* **17**, 91–97 (2002).
61. Tristem, M. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the Human Genome Mapping Project database. *J. Virol.* **74**, 3715–3730 (2000).
62. Shankarappa, R. *et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* **73**, 10489–10502 (1999).
63. Twiddy, S. S., Holmes, E. C. & Rambaut, A. Inferring the rate and time-scale of dengue virus evolution. *Mol. Biol. Evol.* **20**, 122–129 (2003).
64. Drummond, A., Pybus, O. G. & Rambaut, A. Inference of viral evolutionary rates from molecular sequences. *Adv. Parasitol.* (in the press).
    A review of the methods used to estimate substitution rates in viruses, including estimating molecular dates when rates vary.
65. Fitch, W. M., Leiter, J. M., Li, X. Q. & Palese, P. Positive Darwinian evolution in human influenza A viruses. *Proc. Natl Acad. Sci. USA* **88**, 4270–4274 (1991).
66. Rambaut, A. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**, 395–399 (2000).
67. Page, R. D. M. & Holmes, E. C. *Molecular Evolution: a Phylogenetic Approach* (Blackwell Science, Oxford, UK, 1998).
68. Madsen, O. *et al.* Parallel adaptive radiations in two major clades of placental mammals. *Nature* **409**, 610–614 (2001).
    Used the quartet method — which uses several calibration dates to allow for differences in substitution rate between lineages — to support the hypothesis that modern mammals arose long before the final extinction of the dinosaurs.
69. Conway Morris, S. Early metazoan evolution: reconciling paleontology and molecular biology. *Am. Zool.* **38**, 867–877 (1998).
70. Valentine, J., Jablonski, D. & Erwin, D. Fossils, molecules and embryos: new perspectives on the Cambrian explosion. *Development* **126**, 851–859 (1999).
71. Carroll, R. C. Towards a new evolutionary synthesis. *Trends Ecol. Evol.* **15**, 27–32 (2000).
72. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, UK, 1983).

🌐 Online links

FURTHER INFORMATION
University of Sussex Centre for the Study of Evolution:
http://www.biols.susx.ac.uk/CSE
Allan Wilson Centre for Molecular Ecology and Evolution:
http://awcmee.massey.ac.nz
Access to this interactive links box is free online.