Inferring the historical patterns of biological evolution

Mark Pagel

School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, UK

Phylogenetic trees describe the pattern of descent amongst a group of species. With the rapid accumulation of DNA sequence data, more and more phylogenies are being constructed based upon sequence comparisons. The combination of these phylogenies with powerful new statistical approaches for the analysis of biological evolution is challenging widely held beliefs about the history and evolution of life on Earth.

Since 1981, the number of articles reporting phylogenies based on gene-sequence information has been increasing exponentially, with more than half appearing since 1996 (Fig. 1), as new sequence information becomes available at exhausting rates. The phylogenies span taxonomic groups ranging from viruses to bacteria, fungi, plants and animals. The prospect of describing in detail the patterns of descent within many of the major groups of organisms, seen as fanciful just ten years ago, is now realistic.

However, the influence of these new phylogenies extends beyond cataloguing the relatedness of species. All the events of biological evolution are played out somewhere along the branches of phylogenetic trees. As a consequence, these phylogenetic trees preserve traces of the historical evolutionary processes that gave rise to the diversity of contemporary species. This raises the intriguing possibility that the combination of a phylogeny and information on species can be used to infer what the past was like and how the present came about.

I shall not be concerned here with the reconstruction of phylogenies themselves, as this has received much attention (see, for example, ref. 1). Rather, I shall describe the recent advances in statistical modelling of evolution on phylogenetic trees, particularly the use of maximum-likelihood techniques, that are providing researchers with new ways to investigate the evolution of life on Earth. What sets the new statistical techniques apart from conventional non-statistical methods or palaeontological approaches is the range of characteristics that can be investigated and the nuances of historical trends of evolution that can be characterized. The result is that a new era of biological studies is emerging in which statistical approaches applied to phylogenies and information about species form an independent branch of historical enquiry.

Four areas of historical evolutionary enquiry have benefited most from the new statistical approaches: reconstruction of ancestral character states, using a phylogeny in combination with a statistical description of how the traits of organisms evolve, to discover the most probable characteristics of ancestral species; estimation of the timings of historical evolutionary events; assessment of the tempo of evolution, which in turn allows the testing of punctuational or gradual models of evolution; and comparative studies, using correlation and regression to investigate which features of organisms change with which other features or with aspects of their environment, which can provide evidence for the temporal order of changes in two traits, suggesting probable causal pathways.

To illustrate the progress being made in these four areas, I draw on recent studies in which statistical models applied to phylogenies of organisms have been used to infer unexpected features of the common ancestor to life, to challenge conventional views about the Cambrian explosion and the effects of the Cretaceous–Tertiary (K–T) extinction, to reconstruct ancient proteins, to investigate gene–culture coevolution in human societies, and to test widely held theories of mammalian brain-size evolution.

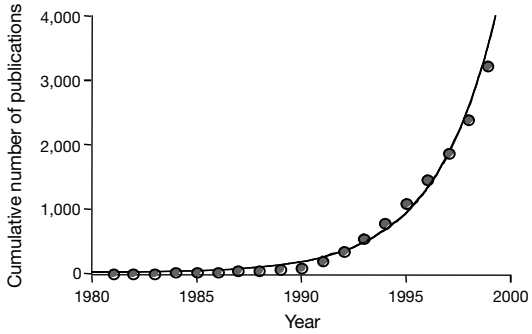

Figure 1 Cumulative number of publications in the Science Citation Index since 1981 that cite the terms 'molecular' and 'phylogeny' in the key words or abstract. The actual number of articles reporting gene-sequence-based phylogenies will be higher. The value for 1999 is extrapolated from the first seven months' data. The fitted curve accounts for more than 97% of the variance and has a period doubling time of just over two years.

Statistical models

The motivation for using statistical models to infer historical patterns of evolution lies in the belief that the diversity of contemporary species reflects the action of various evolutionary processes[2], including the rate and tempo of evolution, timings, correlations and ancestral states. A statistical model specifies by its parameters the way in which a process unfolds over time, or how a past feature leaves traces in the present. This is equivalent to identifying the contemporary signature or imprint of historical events.

The statistical parameters of a model of evolution are typically estimated by maximum-likelihood methods, in which the observed data on species and the model are represented in a common probabilistic framework. Likelihood methods regard the observed data as a fixed observation and seek the values of the statistical parameters that provide the most probable description of those data, given the model of evolution[3]. The likelihood does not describe either the probability that the events under study happened (they did) or that the model is true. Rather, it describes the likelihood that a given process as opposed to some other is responsible for the observed data. These properties make likelihood particularly suited to historical inference problems, in which the observed data arise only once.

If the model of evolution is a hypothesis to explain the data, likelihood chooses the hypothesis that best fits those data[3]. Hypotheses (models) are tested by the likelihood ratio statistic (LR), defined as $LR = -2\log_e[H_1/H_2]$, where, by convention, H_1 is the likelihood associated with the hypothesis that fits the data less well. If the hypotheses are special cases of one another, then LR approximates to a χ^2 statistic with degrees of freedom equal to the difference in the number of free parameters in the two models.

Otherwise, LR greater than 4.0 is conventionally taken as evidence that one of the two hypotheses explains the data significantly better than the other[3].

I shall assume here that the phylogenetic tree is known independently and without error, and shall focus instead on how phylogenies can be used to reveal the past. Nevertheless, it may be useful to define some terminology of phylogenetic trees: the root is the common ancestor to all the species in a phylogenetic tree; branches emanate from the root, tracing the course of evolution to descendants; these descendants reside at nodes of the tree; branches emanate from these nodes, eventually reaching the tips, or the contemporary organisms. Branches have lengths, typically measured in units of time or genetic divergence. A lineage can be loosely defined as a given pathway from the root of the phylogeny along connected branches to some ancestral node or contemporary species. See Box 1 for an explanation of some statistical terms.

Reconstructing ancestral states

Reconstructions of ancestral character states make it possible in principle to describe what the past was like, to discover how traits evolve and to understand function. They are increasingly being used to reconstruct proteins and genes that existed millions of years ago[4,5], and may provide insight into how genes will respond when subjected to forced-evolution methods. New statistical models incorporate explicit models of trait evolution with, amongst other things, the capability to detect directional trends, thereby raising the possibility of reconstructing ancestral features outside the range of features observed in the data.

The Markov-transition process is the statistical model widely used to describe the evolution of traits that adopt only a finite number of states. It is routinely used in phylogeny reconstruction[1] and in comparative methods[2,6]. More recently, it has been applied to reconstructing ancestral character states on phylogenies[7–12], including protein evolution, studies of sexual selection, and habit and diet preferences[8].

The Markov approach estimates the rates at which a discrete character makes transitions among its possible states as it evolves through time. These rates are sufficient to calculate the most probable states at ancestral nodes of the phylogeny. The maximum-likelihood estimate of an ancestral state is determined by independently calculating the likelihood of observing the species data, having successively fixed the particular ancestral node at the possible states of the character; the state with the largest likelihood corresponds to the most probable state at that node[12].

Maximum parsimony is the principal alternative method to likelihood reconstructions of ancestral states of discrete characters, and is still the more widely used. Parsimony reconstructions choose the state of the ancestor to minimize the number of evolutionary events required to explain the character states of the species. Parsimony works well when change is rare or branches are short (see, for example, ref. 13), because, under these circumstances, the probability of a character changing in a branch of the phylogenetic tree is not strongly related to its length. But parsimony methods can perform poorly, especially when rates of character evolution are high and the phylogeny includes some long branches[1,14,15]. This is because parsimony does not take into account the lengths of the branches, so it has a tendency to underestimate the amount of change in long branches. Under some other circumstances, maximum-likelihood and parsimony methods can be shown to be formally equivalent[16].

For traits that naturally vary along a continuous scale, constant-variance random-walk (sometimes called brownian motion) models are the analogue to the Markov-transition model. In the conventional random-walk model, traits evolve in each instant of 'time' dt with a mean change of zero and unknown and constant variance, σ^2. Time may be chronological or some other unit of divergence, such as genetic distance. The evolutionary process is presumed to unfold independently at each instant of time and along each of the branches of the phylogeny. The expected variance of a given species' trait value is then $t\sigma^2$, where t records the total path length (time or distance) from the root to that species. Trait values of species or lineages that have diverged more from the root are expected to have larger variances and so are less reliable observations. Closely related species will tend to have similar trait values, even under a random walk, as they share most of their evolutionary history. A generalized least squares (GLS) approach[2,8,17,18] provides a natural framework in which to represent these features that arise from phylogenetic associations, while simultaneously estimating the variance parameter of trait evolution.

The standard GLS model, by presuming that traits evolve according to an unbiased random walk (neutral drift), cannot detect any directional trends of trait evolution along the branches of the tree. Historical trends such as a phyletic increase in size (for example, Cope's law) will be masked. Consequently, this model always estimates the ancestral state at the root of the phylogeny as falling somewhere within the range of observed values in the species data. Ancestral states obtained from squared-change parsimony or from the popular independent-contrasts approaches for comparative studies are equivalent to those obtained from the constant-variance random-walk model[8]. A general treatment for reconstructing ancestral states of continuously varying characters can be found in ref. 8.

Recent directional GLS models[2,18] for continuous traits can detect historical trends of trait evolution. These models examine the correlation between the species' trait values and the total phylogenetic distance or path length from the root of the tree. If a

Box 1

Statistical terms and models

Generalized least squares (GLS)	A statistical method suitable for analysing data that may not be statistically independent or may not have the same expected variances[17]. It is well suited to phylogenetic applications, owing to the expected similarity amongst species associated with phylogenetic relatedness, and to the different variances arising from differing total path lengths from the root to species. Neutral-drift and directional GLS models both use these features, but the directional model can also detect trends of trait evolution.
Independent contrasts	A method for analysing data from phylogenies that uses the phylogeny to identify a set of mutually independent comparisons between pairs of species, pairs of nodes, or a node and a species (see refs 54, 58).
Likelihood	An amount proportional to the probability of observing the data, given some model[3]. In an evolutionary context, the model often characterizes an aspect of trait evolution.
Likelihood ratio	A way of statistically testing for a difference between two likelihoods (instantiations of models).
Maximum likelihood	A set of techniques for choosing the parameters of a statistical model in such a way as to provide the most probable description of the observed data, given the model.
Markov process	A process in which the probability that a trait takes some value depends only on the value of the trait in the previous unit of time—the Markov no-memory property. For discrete characters, like nucleotides or amino acids, the probability of being in state i at time t depends only on what the state was at time $t-1$.
Maximum parsimony	In a phylogenetic context, a set of techniques for reconstructing phylogenies or ancestral states on phylogenies that prefers solutions requiring the least amount of change.
Random walk	A process in which the state of a variable at time t is given by its starting point (time 0) plus the sum of all the random changes to the variable from time 0 to time t. In the neutral-drift model, changes are randomly sampled from a distribution with a mean of zero and a fixed variance. In a directional drift model, the mean of the distribution of changes is different from zero.
Squared-change parsimony	A method for reconstructing ancestral character states that minimizes the sum of squared changes along the branches of the phylogenetic tree. It produces the same estimates as methods used in independent contrasts.

directional trend exists, species that have diverged more from the root will also tend to have changed more in a given direction, that is, they will be larger or mature earlier, for example. Under these circumstances the directional model will reconstruct ancestral states more accurately and, importantly, it can use the trend to reconstruct the character state at the root of the tree to lie outside of the range of observed values in the data.

Ribonuclease evolution. Maximum-parsimony reconstructions of ancient artiodactyl ribonucleases[19]—an enzyme involved in foregut digestion—assign glycine as the ancestral state at a key amino-acid position, with a transition to aspartic acid about 40 million years ago (Fig. 2). This transition may correspond to the adoption of true ruminant digestion[19]. Maximum-likelihood reanalyses of the same data[7] reveal high rates of character evolution at this site: the estimated 'half-life' for replacement of glycine by aspartic acid is approximately 42 million years (Myr) and for aspartic acid by glycine is 68 Myr (ref. 7), over a tree length spanning 450 Myr. The relatively short half-lives arise because the phylogenetic distribution of traits (Fig. 2) imply that there were at least three historical changes of the amino acid in these artiodactyls. Likelihood reconstructs the ancestral state as aspartic acid, opposite to the inference derived from parsimony (Fig. 2) and implies at least four, as opposed to three, evolutionary transitions on the tree.

Why is there a difference? The likelihood analysis detects that at least two of the four aspartic acid-to-glycine replacements that its reconstructions minimally imply occur in long branches (those leading to camels and cows). These changes are not improbable owing to the high rates of evolution. The parsimony analysis, by ignoring the branch-length information, probably underestimates the amount of change in these longer branches. Parsimony therefore prefers the solution of three events to four, even though these events are reconstructed to occur in comparatively short branches, and require one reversal (an aspartic acid to glycine) transition. Likelihood considers these changes to be relatively improbable. If the branches were all set to the same length, parsimony and likelihood would return the same answers for these data. Maximum likelihood seeks the most probable explanation of the observed data (including the tree and branch lengths), given the model of evolution, not necessarily the solution with the fewest events.

The likelihood analysis raises doubts about the suitability of parsimony for these data. It also raises the fundamental question of whether parsimony is the best general model for trait evolution. The Markov-transition model in a likelihood framework has an advantage in that it takes into account the length of branches in the tree and can adjust to low or high rates of change. The ribonucleases may not be an isolated case: parallel and convergent evolution at the molecular level may be more common than once believed[20–24], and examples of high rates of morphological trait evolution are easy to find[11].

The common ancestor to life. The common ancestor to all cellular life may have arisen more than 3.8 billion years ago[25] when the Earth's environment was hot, unstable and wracked by volcanoes. Phylogenetic trees of life typically reveal that the extant hyperthermophilic bacteria and archaeal species, which inhabit environments of extreme temperatures, have some of the deepest and oldest branches[26–28] (Fig. 3), and it is consequently a widely endorsed textbook view that the common ancestor of life was adapted to hot conditions.

The proportion of all nucleotides that are either guanine or cytosine (the G+C content) of ribosomal RNA is a reliable indicator of the environmental temperature of an organism[29], so an estimate of the G+C content of the root of the tree of life provides evidence for the environmental conditions that prevailed when the common ancestor to life arose. The technical difficulty in estimating the G+C content of the common ancestor to life lies in a limitation of the standard Markov model of gene-sequence evolution as applied to estimating phylogenetic trees. These models are classified as homogeneous and stationary, that is, they presume for computational purposes that all lineages have the same frequencies of nucleotides (often taken to be the relative frequencies of A, G, C, and T or U in the data), and that all lineages are in equilibrium in the sense that these relative frequencies have not changed over time. Observed differences in base composition among lineages are treated as chance variation arising from neutral drift. A consequence of these assumptions is that homogeneous and stationary Markov models are constrained always to assign the ancestral nucleotide frequencies within the range observed in the data, typically as the relative frequencies observed in the extant species.

In an important development, Galtier *et al.* (ref. 29) suggest a non-homogeneous, non-equilibrium Markov model of gene-sequence evolution. The model allows lineages to evolve lineage-specific relative nucleotide frequencies, thereby exploiting the observation that lineages separated for billions of years can have

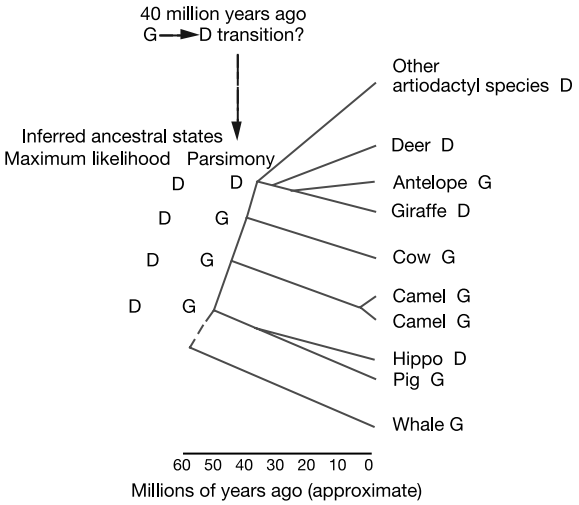

Figure 2 A phylogeny of the artiodactyls with approximate branch lengths, adapted from ref. 19. More recent phylogenies place the hippoptamus as the sister group to whales[77], but this difference does not qualitatively affect the results reported here. The G (glycine) and D (aspartic acid) indicate the species values, and, separately, the ancestral states derived from parsimony[19] and maximum likelihood[7]. Jermann *et al.*[19] postulate a transition from G to D at position 38 of the ribonuclease gene, corresponding to the adoption of true ruminant digestion in artiodactyls. Maximum-likelihood methods favour D as the ancestral state (see text). Whales are included to show a putative outgroup. The line to the whales is dotted and the root values are not recorded because Jermann *et al.*[19] and Schluter[7] considered only the artiodactyls.

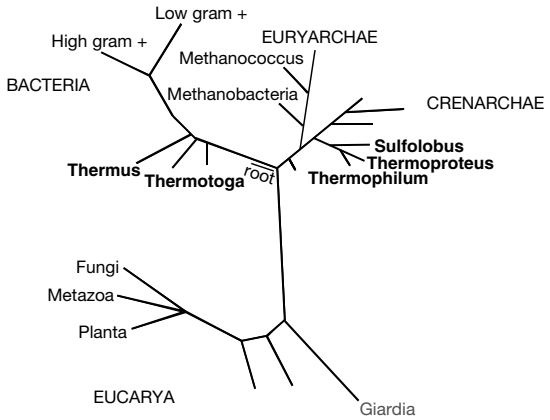

Figure 3 A tree of life showing the approximate relationships among the bacteria, archaea (eurarchaea and crenarchaea) and eukaryotes, adapted from ref. 27. Extant hyperthermophilic organisms are in bold and are close to the root of the tree of life.

Table 1 G+C nucleotide content for the four main groups in the tree of life

Group	G+C nucleotide content		
	All species (LSU)	Extreme sequences	
		SSU	LSU
Eucarya	50.98%	71.0, 70.9	65.7, 62.8
Crenarchae	59.63%	65.4, 65.2	60.4, 59.7
Euryarchae	55.08%	65.1, 65.0	58.7, 58.4
Bacteria	52.78%	62.4, 63.2	58.3, 57.4
Estimated ancestral state (common ancestor to life)	54.0%	57.3%	55.5%
Typical hyper-thermophile		≥60%	

Extreme sequences represent the two species with the highest G+C content within each group (from ref. 29). Sequences are from short-subunit (SSU) and long-subunit (LSU) ribosomal RNA.

widely divergent G+C contents. If the nucleotide frequencies diverge between lineages systematically with underlying genetic divergence, then a directional trend of evolution is suggested, rather than neutral drift. The model can detect such directional trends, and so can potentially estimate ancestral states that lie outside the range of features observed in the species data.

Galtier *et al.*'s model estimates the ancestral G+C content of both short- and long-subunit ribosomal RNA at an intermediate value of 52–54%, with narrow confidence limits, when applied to sequences from the euryarchae, eucarya, crenarchaea and bacteria, including thermophilic and non-thermophilic species (Table 1). Analysis of only those species with the highest G+C content from each of the four principal groups shows that these results are not simply some average of the observed data: extraordinarily, the estimated ancestral sequences for this subset have lower G+C content than any of the extreme sequences. Computer analyses confirm the reliability of the estimates and the capability of the model to reconstruct the ancestral sequence correctly in simulated data[29].

Differences in the G+C contents of ribosomal sequences in lineages that diverged perhaps more than three billion years ago preserve the trace of an ancient and long-term directional trend of evolution. The non-equilibrium statistical model recovers these traces, enabling it, when evolution is traced backwards in time, to estimate the ancestral G+C content at a value lower than suggested from previous analyses. The significance of Galtier's *et al.*'s results[29] is that the G+C content in the ranges they estimate is not compatible with the high G+C content (typically 60 of more) that would indicate a thermophilic common ancestor to life.

Extant hyper-thermophilic species, rather than being ancestral, may be derived descendants of a common ancestor to life that arose in environments of moderate temperature: their deep-rooting phylogenetic position is not a reliable indicator of the ancestral G+C content. Reminiscent of the ribonuclease genes discussed above, hyper-thermophilia may have evolved independently several times. The reconstructed ancestral states do not unequivocally rule out a thermophilic origin of life, but no evidence for it can be claimed from the RNA sequences most likely to harbour that evidence.

Influenza evolution. Directional models can prove crucial for correctly estimating the ancestral state of a continuous trait. Figure 4 shows the cactus-like structure of phylogenetic trees of the haemagglutinin and non-structural 1 (NS1) influenza genes, based on samples drawn at known times over a number of years[30]. Branches that end before the present represent extinct lineages of the influenza virus. The plots record the number of nucleotide substitutions to each gene, as measured from the root of the tree, against the year in which the sample was obtained, and reveal striking clock-like regularity.

The dates predicted to fall at the roots of these two trees correspond to the occurrence of the common ancestors to the haemagglutinin and NS1 genes, that is, to the genes that would have been ancestral to all of the lineages in the respective trees. The predicted date of the root must be earlier than any of the observations but, for both genes, the drift GLS model predicts that the common ancestor arose after at least one of the known sampling times in the data set (haemagglutinin, 1968.2 ± 0.4; NSI, 1939.1 ± 5.7; asterisks on the time axis of Fig. 4).

What has gone wrong? The directional GLS model (dashed regression lines) identifies that time and molecular divergence strongly covary in these data. By rewinding time down the branches leading to the root, the directional model correctly places the time of the common ancestor outside the range of observed values (the point where the dashed regression lines cross the time axis: haemagglutinin, 1968.8 ± 0.1; NSI, 1931.20 ± 2.3; Fig. 4). The slopes of the directional regression indicate that there were about 6.5 and 1.7 substitutions each year, respectively, for haemagglutinin and the NS1 gene.

The solid regression lines (Fig. 4) are derived from a simple regression across the observed data. Although here they are similar to the GLS regressions, they are incorrect as they fail to take account of the similarities (non-independence) among lineages that arise from phylogenetic associations, and by virtue of giving equal weight to all observations despite large differences in their expected variances. The NS1 regression line, when extrapolated, crosses the time axis at a point after at least two of the observations in the data set.

Directional trends, such as those identified for influenza and in the analysis of G+C content, can reveal fundamental trends in the history of trait evolution that must be accounted for if the past is to be reconstructed accurately. This is true whether the traits under investigation are time or some morphological, life-history, behavioural, genetic or other feature. Little is known about such trends outside careful palaeontological research (see refs 31, 32 for exceptions), but appropriate statistical models can identify them and hence return provocative new answers to old questions.

Timings of evolutionary events

If d is the amount of gene-sequence divergence between two species, and r is the rate of molecular evolution, then their time of divergence is given by $T_{past} = d/2r$, where the 2 accounts for the fact that the genetic divergence has accumulated independently in two branches. This is the widely used molecular clock. The rate of evolution r is obtained by calibrating a given amount of divergence from the date of a fossil or geological event. However, r may vary from lineage to lineage and from time to time. This rate heterogeneity greatly limits the accuracy and credibility of molecular-clock studies[33].

New statistical models are limiting the damage of this critique. A maximum-likelihood approach based on the 'quartet' method can accommodate rate heterogeneity from two independently calibrated clocks to estimate times of divergence[34,35]. The method is applied to two pairs of species (Fig. 5) for which fossil ancestors are known. A calibrated molecular clock is derived for each pair by using the known dates and genetic divergences between the pairs. The two clocks are then used simultaneously to estimate the most likely date of the common ancestor to the four species. Computer simulations show that the method returns good estimates of times of divergence[34,35].

By combining the information from two independently calibrated molecular clocks, the quartet method can accommodate rate heterogeneity between the two pairs of species. However, the method assumes that a constant molecular clock operates within each pair of species in the quartet. This may not be true for any given data set, and so rate constancy within the pairs is tested by fitting a standard Markov model of sequence evolution to the unrooted quartet. The quartet is unrooted at this step because the date at the root is unknown, so the two oldest branches (Fig. 5) are joined. Rates of evolution are allowed to vary freely in each of the five branches of the unrooted quartet. Rate constancy is rejected if this five-rate model fits the data better than does the constrained (two-

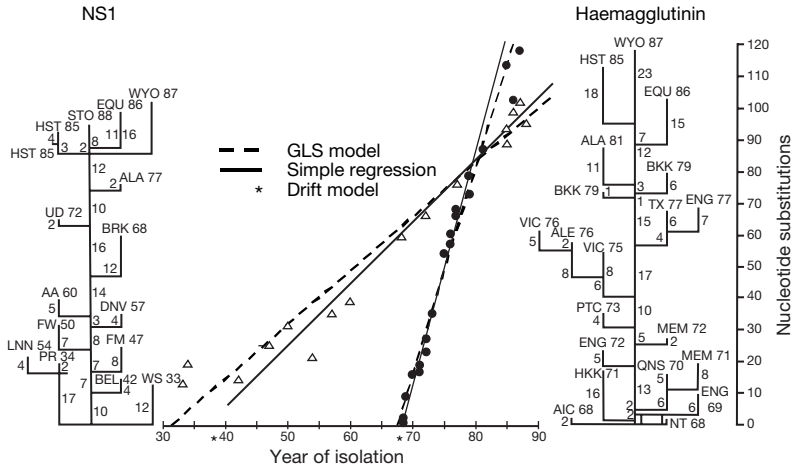

Figure 4 Peculiar cactus-like shape of two gene trees from data on the influenza virus (modified from ref. 30). Left, non-structural 1 (NS1) gene. Right, haemagglutinin gene. Branches that terminate before the present (top) are extinct lineages; sampling times are indicated in the lineage name (for example, ALA77 was obtained in 1977). The length of the vertical lines in the gene trees is proportional to the number of nucleotide substitutions (indicated by the number adjacent to the vertical line). The total number of substitutions inferred to have occurred since the root of the tree is shown in the scale on the right. The graph records the number of substitutions from the root to the tip against the year of sampling; triangles, NS1 gene; circles, haemagglutinin. Three statistical analyses are shown. Asterisks along the time axis indicate inferred root times, as derived from the neutral-drift GLS model; this model predicts the root but ignores the information linking time to genetic divergence. The dashed regression line is derived from the directional GLS model; this model predicts the root and therefore extends to the time axis, signifying the time of the point of zero genetic divergence. Solid regression lines are derived from a simple across-species analysis; the simple regression does not extend to the time axis (and thereby predict the value at the root) because this method fits the line only to the observed data; extrapolating beyond these data is not valid. The directional-GLS and simple regressions differ because the latter weights all observations equally, whereas the GLS models give more weight to observations closer to the root as they are expected to have smaller variances (see text). The simple regression also fails to account for the non-independence among successive data points that derives from counting nucleotide substitutions up the backbone of the tree repeatedly.

rate) model, in which case the gene sequences are excluded as unreliable clocks. Only when rate constancy is supported are sequence data used to estimate molecular clocks, which in turn provide an estimate of divergence time.

The Cambrian explosion? The quartet method returns extraordinary results that challenge conventional wisdom about the dates of the major adaptive radiations of animal phyla associated with the Cambrian explosion. According to the Cambrian-explosion hypothesis, the major phyla and even classes of the animal kingdom emerged suddenly in a rapid evolutionary radiation beginning about 565 Myr ago near the start of the Cambrian era[36]. Representatives of nearly all animal phyla that fossilize are found in the Cambrian rocks. The earliest undisputed metazoan fossils are dated to about 600 Myr[37], reinforcing the view of the Cambrian as a sort of 'Big Bang' of animal evolution[38].

Quartet-model estimates derived independently from a large number of genes indicate that the ancient echinoderm–vertebrate and protostome–deuterostome lineages of multicellular animals diverged long before the Cambrian[35]. Excluding genes showing rate heterogeneity within pairs, most divergences are estimated at 1,000 Myr or more, and none occurs more recently than 680 Myr; 95% confidence intervals exclude both the Cambrian origin and the earlier fossil dates of around 600 Myr.

Other recent molecular estimates of the divergence of animal phyla[39], derived from genes tested for rate constancy, broadly support an age of 1,000 Myr. Suggested dates of 704 Myr and 600–670 Myr from the reanalysis[40] of data[39] may be based on underestimates of the rate of the molecular clock[35]. The pre-Cambrian has been described as more of a 'phylogenetic fuse'[35] than a period leading up to a Big Bang. New palaeontological findings lend support to the earlier estimates derived from the molecular data. Controversial fossil evidence of worms 1,000 Myr old has been discounted, but some consensus is emerging for a figure of 680 Myr[41], pre-dating the Cambrian by 120 Myr—a long time in the life of an 'explosion'.

Effects of the K–T extinction. The quartet method returns similarly starting results when applied to the effects of the K–T extinction event. The K–T boundary describes a time approximately 65 Myr ago when the dinosaurs became extinct, possibly as a result of a cataclysmic collision between Earth and an asteroid. Geological and fossil evidence indicate that up to 70% of mammalian and avian taxa were lost[42], and the classical view based on fossil evidence is that the major avian and mammalian orders diverged rapidly in adaptive radiations thereafter. The present diversity of mammalian and avian life on Earth is suggested to be largely an accidental by-product of those lineages lucky enough to survive the K–T extinction[42].

Estimates derived from the quartet method tell a different story, however. At least 22 diverse avian lineages probably survived the K–T extinction[43]. Molecular-clock estimates of the timings of the ordinal diversification of birds and mammals, based on genes tested for rate constancy, substantially precede the K–T event, falling 90 Myr to 113 Myr[44]. Intriguingly, these dates for bird and mammal diversification coincide instead with the break-up of large continental land masses[44].

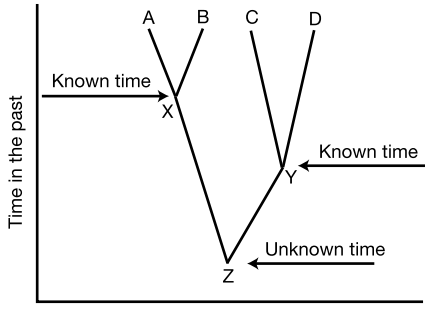

Figure 5 The quartet method simultaneously considers two pairs of species with known times of divergence based on fossils. The genetic distance between each pair is calculated and the molecular clock calibrated from the divergence time, yielding rate estimates μ_X and μ_Y. The unknown branch lengths XZ and YZ can then be found from $XZ = \mu_X(t_Z - t_X)$ and $YZ = \mu_Y(t_Z - t_Y)$, where t_Z is chosen in a standard Markov model of gene-sequence evolution to maximize the likelihood of observing the known times and rates (after Fig. 1 in refs. 34, 35).

Re-evaluating animal diversification. So divergent and provocative are these new findings pertaining to the Cambrian and K–T extinction that they force a re-evaluation of both molecular and fossil data in an effort to understand the evolution of animal diversity millions of years ago. Is it possible that the molecular clocks are wrong? If rates of gene-sequence evolution were higher during the Cambrian period, clocks calibrated from more recent (slower) periods would overestimate the age of the divergence of ancestral lineages. However, Wray *et al.* (ref. 39) provide a detailed discussion of why it is unlikely that undetected rate acceleration during the Cambrian would alter the conclusions of the molecular-clock studies. The earlier dates derived from the new statistical models do not deny the existence of a rich diversity of forms in the Cambrian, but they do suggest that the basic animal body plans emerged long before the Cambrian. Similarly, the K–T events were momentous but the diversity of avian and mammal forms may owe more to adaptation to new environments arising from gradual tectonic, rather than sudden cataclysmic, climatic events.

Molecular-clock studies are providing a rich and independent set of results for anthropologists, biologists and palaeontologists to debate. The date of the domestication of dogs has recently been set back to approximately 135,000 years ago[45], far earlier than the 14,000 years ago favoured by archaeologists, and may have occurred several times independently in different regions of the world. The long-awaited amplification of ancient DNA from the Neanderthal type specimen indicates that the lineages leading to Neanderthals and modern humans diverged 500,000 years ago[46], weighing against speculation and some fictional accounts of gene flow between these groups.

The tempo of evolution

Do traits evolve in fits and starts, or smoothly and gradually? Darwin's preference for gradualism is evident in his remark that "as natural selection acts solely by accumulating slight, successive, favourable variations, it can produce no great or sudden modifications"[47], although some interpret him differently[48]. Proponents of punctuated equilibra[49] contend that the fossil record suggests a different view: that most evolution occurs during relatively short periods of rapid change that are interspersed within far longer periods of stasis, during which time traits are in some form of equilibrium.

The challenge of punctuated equilibria to gradualists is that it begins to unhook the evolution of traits from the regular tug of natural selection. Theorists have responded imaginatively[50]. The sometimes intemperate debate between the two camps has continued for over 25 years.

Critics charge that the fossil record can never provide sufficient resolution to answer questions of the tempo of trait evolution convincingly (see, for example, ref. 51), whereas others emphasize the completeness of the record for some groups[52]. It is little appreciated that a continuum of trait evolution from gradual to punctuational change can be detected on phylogenies, given appropriate scaling of branch lengths. Key questions of gradual versus punctuational evolution of traits can be tested, and over a far wider range of both taxa and traits than is available in the fossil record.

Maximum-likelihood methods in conjunction with the models for discrete and continuously varying traits can estimate scaling parameters to characterize the tempo of trait evolution. The parameter called κ defines for discrete traits[2,6] the relationship between the lengths of individual branches and the probability that a character changes state. For continuously varying characters, the parameter δ scales both the shared phylogenetic path lengths between related species and the total path lengths from the root of the phylogeny to the tips (defined as continuous-variables κ in ref. 6).

These parameters discriminate amongst several modes of trait evolution. Direct gradualism describes traits that change linearly with branch length (discrete traits, $\kappa = 1.0$), or with shared and total path length (continuous traits, $\delta = 1.0$). Scaled gradualism arises when the relationship of trait evolution to branch or path length is nonlinear; κ or $\delta > 1.0$ implies that traits change proportionately more in longer branches, or that longer paths contribute more to trait evolution, as would be true if later (that is, more recent) evolution has contributed more than earlier events; κ or $\delta < 1.0$ indicates traits changing rapidly at first then remaining stable, as might occur in adaptive radiations. Decoupled trait evolution refers to traits evolving by amounts independent of the lengths of the branches or path lengths in the tree (κ or $\delta = 0.0$). Scaled gradualism with κ or $\delta \ll 1.0$ and decoupled trait evolution may be consistent with punctuated equilibria. An alternative to these scaled models allows the rate of trait evolution to vary independently in every branch of the tree[53].

The two influenza genes (Fig. 4), although apparently clock-like in their rates of evolution, yield different trends when δ is used to scale sampling time against number of substitutions. The maximum-likelihood estimate of δ for haemagglutinin is not different from 1.0 ($\hat{\delta} = 1.06$, 95% confidence intervals = 0.89–1.24), indicating that this gene is indeed evolving chronometrically. In contrast, the clock-like behaviour of the NS1 gene hides a more complex tempo. The maximum-likelihood estimate of δ is 0.57 (95% confidence interval = 0.22–0.93), indicating that the gene evolves rapidly in shorter paths (temporally earlier lineages in this example), then slows in the later surviving lineages. In other words, more recent viruses have changed less year to year than did earlier ones, indicating that the gene may have reached some sort of fitness plateau. Substitutions to the ribonuclease gene (Fig. 2) do not depart from direct gradualism ($\hat{\kappa} = 0.98$; 95% confidence intervals = 0.52–1.25), providing further evidence that ancestral reconstructions that take account of branch lengths are preferable for these data. As Fig. 4 shows, trends such as those reported here may not be apparent from an inspection of data, but emerge when the data are subjected to the right statistical analysis.

Correlated evolution

The comparative method is one of evolutionary biology's most enduring traditions for testing hypotheses of adaptation[54]. Comparative methods seek evidence for adaptation in the patterns of correlated trait evolution across contemporary species, that is, in how characteristics of organisms (such as size, shape, life history and behaviour) evolve together. The correlations derived from a group of contemporary organisms can also be shown to have an historical interpretation: under some models of evolution, they reflect the correlation that would have held in the ancestral species as the same traits evolved along what would become the branches of phylogenetic trees[55].

Most investigators now realize that statistical analyses of species must take into account the fact that closely related species tend to be more similar than distantly related ones, and thus that species cannot be considered as statistically independent units of observation[54]. Conventional comparative methods for discrete variables use the phylogeny to reconstruct by parsimony the probable ancestral character states, and then calculate statistics conditional upon those reconstructions[56,57]. The 'independent contrasts' approach[54,55,58–62] has come to dominate comparative analyses of continuously varying characters. These methods use the phylogeny in combination with a neutral-drift model of trait evolution to identify a set of statistically independent comparisons among species.

New maximum-likelihood models of correlated trait evolution for discrete[6] and continuously varying characters[2,18] offer hypothesis tests not available to the earlier methods, and avoid some of the difficulties of those approaches. Correlated evolution of discrete binary characters is modelled in a maximum-likelihood framework by the same Markov-process model as described for reconstructing ancestral character states of discrete characters. Evidence for corre-

lated change in two characters is gathered by comparing the likelihood of two binary characters, each allowed to evolve independently on the tree, to the likelihood obtained when the same characters are modelled as if they are evolving in a correlated fashion[2,6].

Statistical tests for the dominant direction of change (for example, do 'forward' transitions predominate over 'backward' ones) and the temporal order of changes (for example, does character y change before or after character x) are possible by comparing appropriate likelihoods[6]. The temporal-order test detects traces of the probable sequence of evolution from the ancestral states of two characters to the derived states of both (see, for example, ref. 63).

Maximum-likelihood models for investigating correlations amongst continuously varying characters use the same GLS framework as described for modelling trait evolution in the influenza virus[2]. The models return estimates of the correlations and regressions among two or more variables while controlling for phylogenetic associations. Trait evolution along the branches of the phylogenetic tree can be modelled in the GLS framework by the neutral drift or other models, such as the model of directional change.

Under the neutral-drift model, the GLS and independent-contrasts techniques[58,59] return the same correlation. This 'drift correlation' includes variance and covariance in the traits that arise from at least three sources: neutral drift, any directional or other non-random trends, and independent correlated trait evolution. By comparison, in the 'directional correlation', the variance arising from neutral-drift and directional trends is removed. The directional-model correlation therefore estimates solely the independent correlated trait evolution, and will differ from the drift correlation whenever there are directional trends of trait evolution in the data.

A key advantage of the GLS approach, however, lies in its ability to scale phylogenetic path lengths in response to patterns in the data[2]. Several recent studies show that independent-contrasts methods perform worse than simple non-phylogenetic analyses in some circumstances (refs 64–66; P. H. Harvey & A. Rambaut, unpublished data). In one of these[65], traits are modelled to evolve by a larger amount per speciation event early on than later in the phylogeny, and trait values are not necessarily very similar between pairs of closely related species. In short, the variance of evolutionary change is not constant, so the neutral-drift model that the independent-contrast method presumes (see the section Reconstructing ancestral states) does not accurately reflect trait evolution.

Using the GLS approach, the δ scaling parameter, described in conjunction with characterizing the tempo of evolution, would detect early and rapid trait evolution as $\delta < 1.0$ and rescale the path lengths accordingly. A second scaling parameter, denoted λ, detects whether the shared evolutionary histories as specified by the phylogeny produce the patterns of similarity observed in the data. Values of $\lambda < 1.0$ correspond to traits being less similar amongst species than expected from their phylogenetic relationships; $\lambda > 1.0$ suggests the reverse. We might expect $\lambda < 1.0$ for the model of ref. 65. Signatures of trait evolution relevant to the suitability of a comparative method reside in the data. The GLS model, by detecting these signatures, is expected to perform reasonably well, even under circumstances averse to other methods.

The evolution of lactose tolerance. The enzyme lactase confers an ability to digest milk. Human infants can digest milk, but most adults cannot (the widespread tolerance to lactose among some European groups is an exception). The dominant view is that adult lactose tolerance in humans is an adaptation to reduced exposure to the Sun[67]. Both the Sun and the enzyme lactase promote calcium absorption. This hypothesis accounts for the lactose tolerance of some northerly dwelling cultural groups, such as the reindeer-herding Lapps, and the prevalence of lactose tolerance in some European groups.

Table 2 Slope of the regressions of brain volume and basal metabolic rate against body size in mammals

Variable	Species	Correlation	GLS regression slope (95% confidence interval)	Theoretical expectation
Brain volume	23	0.96	0.59 (0.52–0.67)	0.67–0.75
Basal metabolic rate	15	0.95	0.72 (0.60–0.84)	0.75

An alternative to the latitude theory is that adult tolerance to lactose is advantageous in cultures that keep animals for milk. If milk forms a significant portion of the diet, selection pressures on adults to develop the ability to digest it could be strong.

Holden and Mace[68] applied the Markov-process model[6] to a phylogeny of human cultural groups to show that adult lactose tolerance has arisen independently—possibly by selection for rare alleles—up to three times in cultures that keep animals for milk, but not in non-dairying cultures. The latitudinal effect disappeared when the phylogeny was used to take account of non-independence amongst cultural groups. Likelihood-ratio tests of the temporal order of gene versus cultural change reveal the probable course of the evolution of lactose tolerance. The ancestral condition of non-dairying and no adult lactose tolerance was first replaced by dairying perhaps as early as 6,000 to 8,000 years ago, which then favoured tolerance to lactose. The data do not support the alternative scenario that human groups with adult lactose tolerance were more likely to adopt dairying.

The phylogenetic statistical approach is able to untangle the latitudinal and dairying effects in a way that previous analyses could not, and could detect evidence to confirm the important causal order of effects: cultural practices of herding societies appear to have selected for a genetic adaptation.

Mammalian brain-size evolution. Theories of the evolution of mammalian brain size link brain volume to body mass, either through body surface area or through basal metabolic rate[69,70]. The surface-area hypothesis requires a slope of 2/3 between brain volume and body mass; the metabolic-rate hypothesis predicts that the slope of brain volume against body mass will be 3/4, owing to the relationship of basal metabolic rate to body size. Empirically derived slopes generally lie somewhere in the 0.67–0.75 range[71]. New molecular phylogenies of the mammals derived from whole mitochondrial DNA (mtDNA) sequences[72] afford tests of the surface-area and metabolic-rate hypotheses using the features of the GLS models.

Before calculating a correlation, the maximum-likelihood values of the path-length scaling parameter δ and the phylogeny scaling factor λ were estimated: both overlap 1.0 ($\hat{\delta} = 0.96$ (0.078–2.79) and $\hat{\lambda} = 1.0$ (0.94–1.0)), indicating that brain-size evolution is proportional to the branch lengths of this mtDNA phylogeny, and justifying the use of the unscaled phylogeny to correct for non-independence. The GLS model returns a slope relating brain volume to body mass that is lower than the expected 2/3 to 3/4, despite a very strong correlation between the two traits (Table 2). Even in these comparatively small samples, the 95% confidence intervals exclude 3/4 and nearly exclude 2/3. The same GLS applied to a composite primate phylogeny[73] of 59 species yields a slope of 0.48 (95% confidence intervals = 0.39–0.57) between brain and body size[2].

Are the regression slopes peculiar to the model or phylogeny? Applying the same model to estimate the slope of basal metabolic rate against body size, again using the mtDNA phylogeny, gives an empirical result (Table 2) that is very close to the expected value of 3/4 (ref. 70). In the light of these results, the surface-area and metabolic-rate theories for mammalian brain size would seem increasingly difficult to support. Linking brain adaptations to specific environmental demands[71,74] may offer an alternative way of thinking about brain evolution.

Conclusions

The development of statistical modelling techniques for analysing trait evolution on phylogenies is still in its early years but, as the studies reported here show, they are already returning results that question long-standing views of the history of life and the patterns of adaptation. They also raise questions about how we should think about evolution. Is it smooth and gradual, or punctuational? Is parsimony, so long the unquestioned null model of trait evolution, really appropriate, or should it be replaced by models that explicitly consider the lengths of the branches of phylogenetic trees? Future developments[75,76] may make it feasible to estimate the historical processes of evolution simultaneously with estimation of the phylogenetic tree.

The growth of statistical inference techniques brings a new and independent point of view to debates about the history of life. There are bound to be many factional and territorial disputes as their use grows, however. As predominantly visual and tactile primates, we may find it far harder to accept statistical inferences than that which we can see and touch. But as in any historical discipline, new ideas will be judged by how well they explain the existing facts. Statistical techniques for inferring the history and pattern of evolution will prove invaluable for looking into the past in ways, and for kinds of traits, that are out of reach to other approaches. □

Received 21 May; accepted 7 September 1999.

1. Hillis, D. M., Moritz, C. & Mable, B. K. *Molecular Systematics*, 2nd edn (Sinauer, Sunderland, Massachusetts, 1996).
2. Pagel, M. Inferring evolutionary processes from phylogenies. *Zool. Scripta* **26,** 331–348 (1997).
3. Edwards, A. W. F. *Likelihood* (Cambridge Univ. Press, Cambridge, 1972).
4. Golding, G. B. & Dean, A. M. The structural basis of molecular adaptation. *Mol. Biol. Evol.* **15,** 355–369 (1998).
5. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvak, Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91,** 501–510 (1997).
6. Pagel, M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B* **255,** 37–45 (1994).
7. Schluter, D. Uncertainty in ancient phylogenies. *Nature* **377,** 108–109 (1995).
8. Schluter, D., Price, T., Mooers, A. Ø. & Ludwig, D. Likelihood of ancestor states in adaptive radiation. *Evolution* **51,** 1699–1711 (1997).
9. Yang, Z., Kumar, S. & Nei, M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141,** 1641–1650 (1995).
10. Koshi, J. M. & Goldstein, R. A. Probabilistic reconstruction of ancestral protein sequences. *J. Mol. Evol.* **42,** 313–320 (1996).
11. Mooers, A. Ø. & Schluter, D. Support for one and two rate models of discrete trait evolution. *Syst. Biol.* **48,** 623–633 (1999).
12. Pagel, M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* **48,** 612–622 (1999).
13. Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R. & Molineux, I. J. Experimental phylogenetics: generation of a known phylogeny. *Science* **255,** 589–592 (1992).
14. Collins, T. M., Wimberger, P. H. & Naylor, G. J. P. Compositional bias, character-state bias, and character-state reconstruction using parsimony. *Syst. Biol.* **43,** 482–496 (1994).
15. Maddison, D. R. Phylogenetic methods for inferring the evolutionary history and process of change in discreetly valued characters. *Annu. Rev. Entomol.* **39,** 267–292 (1994).
16. Tuffley, C. & Steel, M. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull. Math. Biol.* **59,** 581–607 (1997).
17. Johnston, J. *Econometric Methods* (McGraw Hill, New York, 1963).
18. Martins, E. P. & Hansen, T. F. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149,** 646–667 (1997).
19. Jermann, T. M., Opitz, J. G., Stackhouse, J. & Benner, S. A. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374,** 57–59 (1995).
20. Li, W.-H. *Molecular Evolution* (Sinauer, Sunderland, Massachusetts, 1997).
21. Yeager, M. & Hughes, A. L. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol. Rev.* **167,** 45–58 (1999).
22. Bull, J. J. *et al.* Exceptional convergent evolution in a virus. *Genetics* **147,** 1497–1507 (1997).
23. Roux, K. H. *et al.* Structural analysis of the nurse shark (new) antigen receptor (NAR): Molecular convergence of NAR and unusual mammalian immunoglobulins. *Proc. Natl Acad. Sci. USA* **95,** 11804–11809 (1998).
24. Crandall, K. A., Kelsey, C. R., Imamichi, H., Lane, H. C. & Salzman, N. P. Parallel evolution of drug resistance in HIV: Failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol. Biol. Evol.* **16,** 372–382 (1999).
25. Mojzsis, S. J. *et al.* Evidence for life on earth before 3,800 million years ago. *Nature* **384,** 55–59 (1996).
26. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51,** 221–271 (1987).
27. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276,** 734–740 (1997).
28. Brown, J. R. & Doolittle, W. F. Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61,** 456–504 (1997).
29. Galtier, N., Tourasse, N. & Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283,** 220–221 (1999).
30. Fitch, W. M., Leiter, J. M. E., Li, X. & Palese, P. Positive Darwinian evolution in human influenza A viruses. *Proc. Natl Acad. Sci. USA* **88,** 4270–4274 (1991).
31. Omland, K. Character congruence between a molecular and a morphological phylogeny for dabbling ducks (*Anas*). *Syst. Biol.* **43,** 369–386 (1994).
32. Omland, K. Correlated ratest of molecular and morphological evolution. *Evolution* **51,** 1381–1393 (1997).
33. Strauss, E. Can mitochondrial clocks keep time? *Science* **283,** 1435–1438 (1999).
34. Rambaut, A. & Bromham, L. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15,** 442–448 (1998).
35. 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).
36. Fortey, R. *Life: An Unauthorised Biography* (HarperCollins, London, 1997).
37. Valentine, J. W., Erwin, D. H. & Jablonski, D. Developmental evolution of metazoan bodyplans: The fossil evidence. *Dev. Biol.* **173,** 373–381 (1996).
38. Bowring, S. A. *et al.* Calibrating rates of early Cambrian evolution. *Science* **261,** 1293–1298 (1993).
39. Wray, G., Levinton, J. S. & Shapiro, L. H. Molecular evidence for deep precambrian divergences among the metazoan phyla. *Science* **274,** 568–573 (1996).
40. Ayala, F. J., Rzhetsky, A. & Ayala, F. J. Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proc. Natl Acad. Sci. USA* **95,** 606–611 (1998).
41. Kerr, R. A. Earliest animals growing younger? *Science* **284,** 411–412 (1999).
42. Raup, D. M. *Mass Extinctions: Bad Genes or Bad Luck?* (Oxford Univ. Press, Oxford, 1991).
43. Cooper, A. & Penny, D. Mass survival of birds across the Cretaceous–Tertiary boundary: molecular evidence. *Science* **275,** 1109–1113 (1997).
44. 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).
45. Vilá, C. *et al.* Multiple and ancient origins of the domestic dog. *Science* **276,** 1687–1689 (1997).
46. Krings, M. *et al.* Neandertal DNA sequences and the origin of modern humans. *Cell* **90,** 19–30 (1997).
47. Darwin, C. *The Origin of Species by Means of Natural Selection*, 6th edn (with corrections and additions to 1872) (John Murray, London, 1888).
48. Bowler, P. J. *Charles Darwin: The Man and His Influence* (Blackwell, Oxford, 1991).
49. Gould, S. J. & Eldredge, N. Punctuated equilibria comes of age. *Nature* **366,** 223–227 (1993).
50. Williams, G. C. *Natural Selection* (Oxford Univ. Press, Oxford, 1992).
51. Hecht, M. K. & Hoffman, A. Why no neo-Darwinism: a critique of paleobiological challenges. *Oxford Surv. Evol. Biol.* **3,** 1–47 (1986).
52. Foote, M. & Sepkoski, J. J. Jr Absolute measures of the completeness of the fossil record. *Nature* **398,** 415–417 (1999).
53. Mooers, A. Ø., Vamosi, S. M. & Schluter, D. Using phylogenies to test macroevolutionary hypotheses of trait evolution. *Am. Nat.* **154,** 249–259 (1999).
54. Harvey, P. H. & Pagel, M. *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991).
55. Pagel, M. Seeking the evolutionary regression coefficient: an analysis of what comparative methods measure. *J. Theor. Biol.* **164,** 191–205 (1993).
56. Ridley, M. *The Explanation of Organic Diversity* (Oxford Univ. Press, Oxford, 1981).
57. Maddison, W. P. A method for testing the correlated evolutionary of two binary characters: are gains and losses concentrated on certain branches of a phylogenetic tree. *Evolution* **44,** 539–557 (1990).
58. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125,** 1–15 (1985).
59. Pagel, M. A method for the analysis of comparative data. *J. Theor. Biol.* **156,** 431–442 (1992).
60. Garland, T. Jr, Harvey, P. H. & Ives, A. R. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* **41,** 18–32 (1992).
61. Martins, E. P. & Garland, T. H. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* **45,** 534–557 (1991).
62. Purvis, A., Gittleman, J. L. & Luh, H. K. Truth or consequences—effects of phylogenetic accuracy on two comparative methods. *J. Theor. Biol.* **167,** 293–300 (1994).
63. Rolland, C., Danchin, E. & de Fraipont, M. The evolution of coloniality in birds in relation to food, habitat, predation, and life-history traits: a comparative analysis. *Am. Nat.* **151,** 514–529 (1998).
64. Ricklefs, R. E. & Starck, J. M. Applications of phylogenetically independent contrasts: A mixed progress report. *Oikos* **77,** 167–172 (1996).
65. Price, T. Correlated evolution and independent contrasts. *Phil. Trans. R. Soc. Lond. B* **352,** 519–529 (1997).
66. Harvey, P. H. & Rambaut, A. Phylogenetic extinction rates and comparative methodology. *Proc. R. Soc. Lond. B* **265,** 1691–1696 (1998).
67. Durham, W. *Coevolution: Genes, Culture, and Human Diversity* (Stanford Univ. Press, Stanford, 1991).
68. Holden, C. & Mace, R. A phylogenetic analysis of the evolution of lactose digestion. *Hum. Biol.* **69,** 605–628 (1997).
69. Jerison, H. J. *Evolution of the Brain and Intelligence* (Academic, New York, 1973).
70. Martin, R. D. Relative brain size and metabolic rate in terrestrial vertebrates. *Nature* **293,** 57–60 (1981).
71. Pagel, M. & Harvey, P. H. Taxonomic differences in the scaling of brain on body weight in mammals. *Science* **244,** 1589–1593 (1989).
72. Arnason, U., Gullberg, A. & Janke, A. Molecular timing of primate divergences as estimated by two nonprimate calibration points. *J. Mol. Evol.* **47,** 718–727 (1998).
73. Purvis, A. A composite estimate of primate phylogeny. *Phil. Trans. R. Soc. Lond. B* **348,** 405–421 (1995).
74. Healy, S. D. & Krebs, J. R. Food storing and the hippocampus in Paridae. *Brain Behav. Evol.* **47,** 195–199 (1996).
75. Wilson, I. & Balding, D. Geneaological inference from microsatellite data. *Genetics* **150,** 499–510 (1998).
76. Larget, B. & Simon, D. L. Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* **16,** 750–759 (1999).
77. Ursing, B. M. & Arnason, U. Analyses of mitochondrial genomes strongly support a hippopotamus–whale clade. *Proc. R. Soc. Lond. B* **265,** 2252–2255 (1998).

Acknowledgements

I thank D. Cox, N. Galtier, G. Laden and A. Rambaut for discussion and access to data. The Leverhulme Trust provided financial support.