# Inferring evolutionary processes from phylogenies

## MARK PAGEL

Pagel, M. 1998. Inferring evolutionary processes from phylogenies.—*Zool. Scr. 26*: 331–348

Evolutionary processes shape the regular trends of evolution and are responsible for the diversity and distribution of contemporary species. They include correlated evolutionary change and trajectories of trait evolution, convergent and parallel evolution, differential rates of evolution, speciation and extinction, the order and direction of change in characters, and the nature of the evolutionary process itself — does change accumulate gradually, episodically, or in punctuational bursts. Phylogenies, in combination with information on species, contain the imprint of these historical evolutionary processes. By applying comparative methods based upon statistical models of evolution to well resolved phylogenies, it is possible to infer the historical evolutionary processes that must have existed in the past, given the patterns of diversity seen in the present. I describe a set of maximum likelihood statistical methods for inferring such processes. The methods estimate parameters of statistical models for inferring correlated evolutionary change in continuously varying characters, for detecting correlated evolution in discrete characters, for estimating rates of evolution, and for investigating the nature of the evolutionary process itself. They also anticipate the wealth of information becoming available to biological scientists from genetic studies that pin down relationships among organisms with unprecedented accuracy. © 1998 The Norwegian Academy of Science and Letters

*Mark Pagel, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.*

## Introduction

The past 10 years have witnessed a revolution in comparative biology. As recently as the middle 1980's comparative studies based upon simple analyses of species data points were common, perhaps in the majority. Now, in a little over 10 years, nearly a complete reversal of practice has occurred: most investigators realise that 'species analyses' are invalid in principle and can often produce spurious results. Methods that account for the influence that phylogenetic relationships among species can exert on the support for evolutionary hypotheses are now almost routinely used in comparative studies.

The context for the growth in the use of comparative methods is an explosive increase in the use of phylogenies in biological studies, especially molecular phylogenies. Figure 1 plots the cumulative number of articles that have been catalogued by the Science Citation Index since 1981 in which the words 'phylogeny' or 'molecular' and 'phylogeny' are mentioned in the title, abstract, or keywords. Following a slow rate of increase throughout the early 1980's both curves enter an exponential phase of growth in the early 1990's that does not yet show any sign of declining. The use of molecular phylogenies is an increasing proportion of the total number of phylogenetic studies as time progresses, and will likely overtake morphological phylogenies early in the next century. These data suggest a pattern of cultural or 'meme' evolution in the biological sciences that can be modelled well by simple models of contact transmission.

Phylogenies bring opportunities to biology, by offering more than a mere description of the genealogical relationships amongst a group of organisms. The main thesis I wish to advance in this article is that, coupled with data on the characteristics, behaviours, and lifestyles of species, phylogenies can be used to infer the historical evolutionary processes that must have existed in the past, given the patterns of diversity seen in the present. This is a simple but important point because it asserts that the combination of phylogeny and contemporary data are sufficient to infer what the past was like and how it got to be like the present. To the extent that the assertion is true, comparative statistical methods can be used to discover — in a manner that might be described as 'statistical palaeontology' — the nature of the evolutionary processes as they advanced along what would become the branches of the phylogenetic tree. Moreover, it can do this in many cases without ever explicitly reconstructing the past.

## Unifying statistical comparative biology

A hint that one can discover the nature of the historical processes of evolution from comparative data emerged from analysis of two different classes of comparative method (Pagel 1993). 'Directional' comparative methods use reconstructions of ancestral character states at each of the nodes of the phylogeny to calculate changes between ancestral and descendant forms along the branches of a phylogenetic tree (e.g., Huey 1987; Huey and Bennett 1987). Pairs of characters so reconstructed are then used to find the correlation between changes along the branches
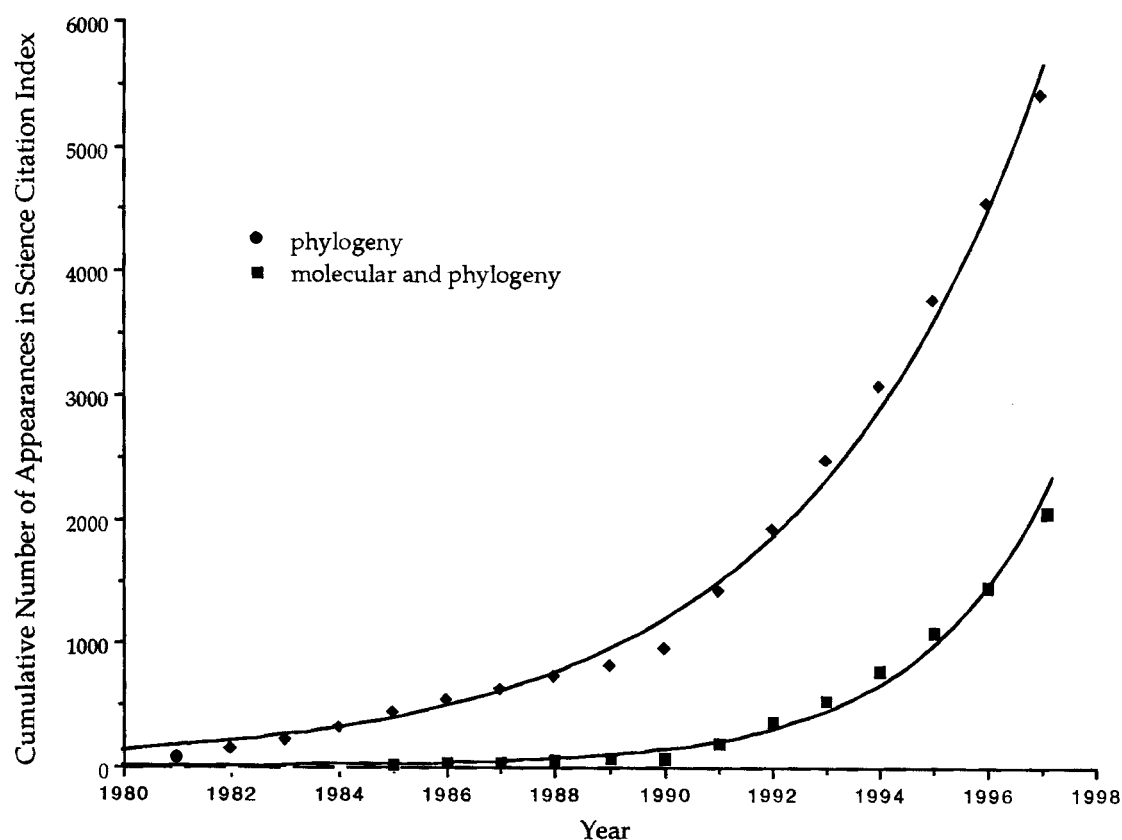
*Fig. 1.* The cumulative number of articles in the Science Citation Index in which the keywords 'phylogeny' (upper curve) or 'molecular' and 'phylogeny' are mentioned in the title, abstract, or keywords, plotted against year since 1981. The citations including 'molecular' have been removed from the upper curve. Both sets of data can be described by exponentially increasing curves.

of the phylogeny. Proponents of directional analyses emphasise the usefulness of these methods for detecting the direction and rate of evolution over evolutionary time.

'Cross-sectional' comparative studies do not estimate changes along the branches of a tree, but rather make comparisons among the contemporary tips of the phylogeny (the species) and among reconstructions of the character states at internal nodes of the phylogeny. The simple species-correlation and the well known 'independent contrasts' techniques (Felsenstein 1985; Pagel and Harvey 1989a; Grafen 1989; Harvey and Pagel 1991; Pagel 1992; Garland *et al.* 1992) that make comparisons between pairs of taxa defined by the phylogeny, are cross-sectional techniques. Cross-sectional techniques are thought to be most useful for relating the contemporary form and function of traits to their environments or to each other. Investigators frequently employ both classes of method in an attempt to reveal different features in the data (examples reviewed in Pagel 1993).

Two discoveries emerged from analysing these two classes of method. One is that the two classes of method are really one. The other is that the methods from the different classes estimate the same statistical parameter. Directional and cross-sectional methods for finding correlations between pairs of characters all estimate the correlation or regression that must have held as the two traits evolved along the branches of the phylogeny. For this reason I called the parameter the methods estimate the 'evolutionary regression coefficient' (Pagel 1993). The analytical result showing their equivalence is demonstrated empirically in Fig. 2. The figure plots the results of direc-

tional tests versus cross-sectional comparative tests on the same data sets, derived from several published studies. If the two classes of technique indeed measure the same thing, we expect that the results will be highly correlated and the slope relating them will be 1.0: this is what the figure reveals. The significance of this result for methods of comparative biology is that a large number of apparently different comparative methods can be collapsed into one. The significance for inferring evolutionary processes is that (Pagel 1993: 202) 'the historical pattern of evolutionary change along the branches of a phylogeny is preserved in the variation among the tips of the phylogeny'.

## Which evolutionary processes and how to estimate them?

The evolutionary process that cross-sectional and directional comparative methods attempt to estimate is that of a correlation between two traits arising and persisting over evolutionary time, yielding a relationship amongst the contemporary species. Many other processes can be identified. One can ask about the *direction* of evolutionary change (small to large?, slow to fast?) or the *order* in which two variables change (do two horned rhinoceroses evolve from one-horned species or *vice versa?*). One can ask about *rates of evolution* and whether rates vary with some ecological, taxonomic, or other variable. Molecular sequence data provide an important source of information for estimating evolutionary processes. Questions about rates of evolution at silent versus coding positions, or at different codon positions can be addressed within a comparative frame-
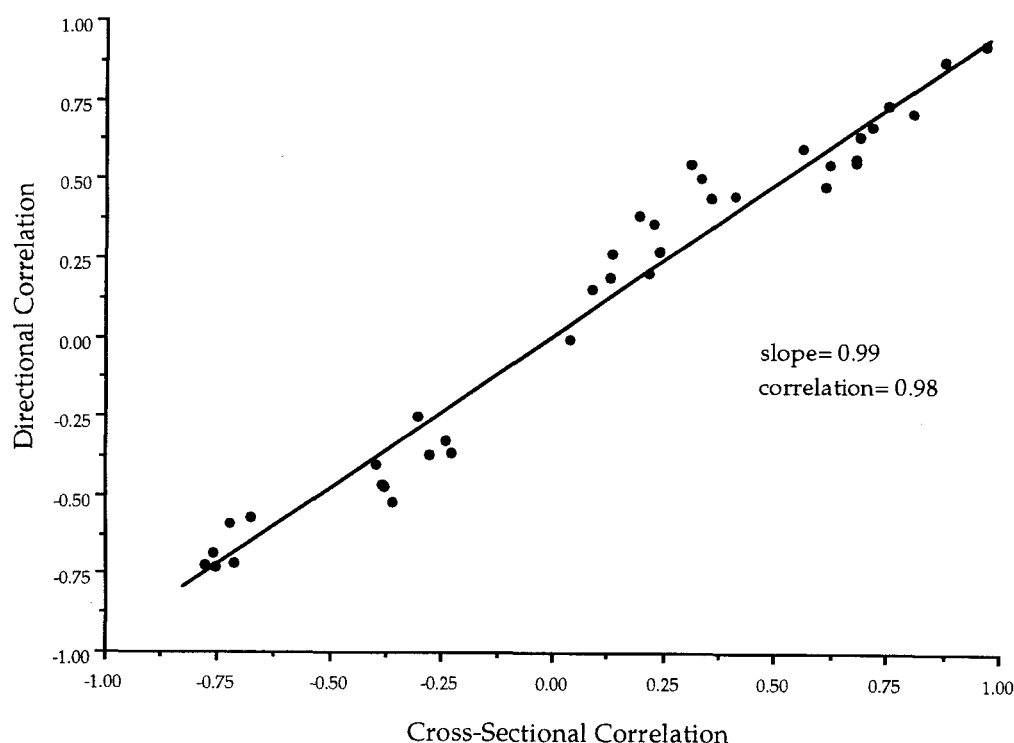
*Fig. 2.* Correlation coefficients derived from 'directional' analyses applied to comparative data sets plotted against 'cross-sectional' correlations derived from the same data sets. Data come from several published studies (see Pagel 1993 for studies).

work. Evolution in different regions of the same gene or in two or more genes, can be investigated as a discrete events, such as the presence or absence of a particular amino acid, or as a continuous event such as the number of substitutions to a region of a gene. *Molecular clocks* and their estimation are a special case of rates of evolution. *Ancestral character states* are not so much a process as a category but they too can be estimated from the phylogeny in a statistical manner.

One can also investigate the nature of the evolutionary process itself. Some comparative methods assume *constant variance* (sometimes called Brownian motion) models of evolution, in which variance accumulates at a constant rate along each branch of the phylogeny. It is possible to seek the optimal transformation of branch lengths in a phylogeny to produce constant variance on the transformed scale. This then reveals on what scale the characters are evolving at a constant rate. It is also possible to test hypotheses about *punctuational change* versus *gradual change* by asking whether evolutionary change in a branch is independent of the length of the branch. Theories of the nature of evolution can be tested: is the molecular evolution observed in a given set of sequences neutral or does it show evidence of selection?

The question of how to estimate a particular evolutionary process will be the subject of much of the remainder of this essay. It will turn out that phylogenies are not just important to estimating evolutionary processes, they in some sense define how to estimate them. This means that the phylogenetic dog can really bite if ignored.

## Statistical framework

R. A. Fisher invented the statistical framework of maximum likelihood (ML) and it is now widely used in

estimation problems including in applications of phylogeny reconstruction (e.g., Edwards 1972; Felsenstein 1981; Friday 1989; Goldman 1990; Hillis *et al.* 1996). Maximum likelihood is also well suited to estimating evolutionary processes because it provides a framework within which the data, the phylogeny, and the underlying model of evolution (which contains parameters specifying the evolutionary process to be estimated) can all be represented.

A maximum likelihood estimate of a parameter in a phylogenetic context is the value of the parameter that makes the observed data most likely given the statistical model, and the phylogeny. Formally one wishes to maximise

$$L = f(D|Tr,M),$$

where $L$ is the likelihood, $D$ is the data set (values on the character states of a group of species), $Tr$ is the phylogeny, $M$ is the model of evolution, and the expression is read '$L$ is a function $f$ of the data given the phylogeny and the model of evolution'. The model of evolution $M$ is a statistical representation of the evolutionary process. For example, $M$ might be the bivariate normal probability of observing two traits, $x$ and $y$, given the unknown variances and covariances of those traits. More generally, $M$ might include a parameter estimating the most likely ancestral character states at the root of the tree, the rate at which the characters must have evolved, the variance of the evolutionary process (i.e., how variable was it), and the covariation that must have held between the two characters (if any). If $M$ is a model of discrete character evolution, it might additionally identify the rate of transition between character states, and ask whether transitions in one character predispose changes in another. $M$ therefore contains the evolutionary parameters of interest: $M = g(\alpha, \beta, \gamma,...)$,

where $\alpha$, $\beta$, $\gamma$,... can be the ancestral character states, rates of evolution, transitions between character states, correlations, variances, covariances, and so on, to be estimated from the data given the phylogeny.

Maximum likelihood can produce biologically relevant quantitative estimates of all these parameters by taking into account the lengths of the branches of the phylogenetic tree. Branch lengths can represent real time in which rates correspond to changes per unit time, or in units of genetic or other change. In addition, estimates of the uncertainty associated with estimates can be found and incorporated back into the estimates themselves.

An important feature of likelihood approaches in a phylogenetic context is that they can estimate processes on phylogenies without ever explicitly reconstructing the values of ancestral character states. This questionable practice (see Pagel 1994 for discussion) commits the logical error of first inferring ancestral states from species' data, and then treating them in later analyses as fixed and known without error. A likelihood approach can avoid reconstructing ancestral states, and thereby avoid these and other difficulties. The likelihood of a given hypothesis can be found having integrated over all possible values at the ancestral nodes. The different values of the ancestral character are not all equally likely and so this naturally weights their contribution to the overall likelihood. This approach reflects more accurately the view that any character state is possible at any given node, but that some are simply more likely than others. It also avoids the awkwardness that confronts investigators when many equally parsimonious reconstructions of the ancestral character states exist and the hypothesis must be tested on each one.

Hypothesis testing with maximum likelihood proceeds by means of the likelihood ratio statistic, LR. The statistic compares the goodness of fit of a model to the data with that of a simpler model that lacks one or more of the parameters. The LR statistic is then defined as

$$LR = -2\log_e[H_0/H_1],$$

where $H_0$ represents the simpler model and $H_1$ the model containing the parameters representing the evolutionary processes one wishes to estimate. If the simpler model is a subset of the more complicated one, the $LR$ statistic is asymptotically distributed as a chi-squared variate with degrees of freedom equal to the difference in the number of parameters between the two models, i.e., $LR \sim \chi^2(v)$, where $v$ is the number of degrees of freedom.

A criticism that is sometimes levelled at a statistical approach to phylogenetic-based hypothesis testing is that it is meaningless to attach a probability to an event that has already happened. If we observe that wings exist, we know with certainty (probability $= 1.0$) that they evolved. The misunderstanding embodied in this criticism is that the p-value derived from the likelihood approach (or for that matter from classical statistical hypothesis tests) is not a statement about the probability that an event occurred. Rather, it is a statement of the probability that a given process is responsible for the event. Typically, one tests which of two different and unobserved processes best describes the observed data. We may observe that a coin comes up 'heads' ten times in ten flips. We would not attach a probability to the statement that the coin has

been flipped ten times and come up heads on each trial. However, we might very well wish to attach a probability to the statement that the process generating the outcome of ten heads is unbiased. We could then compare that to the probability that the process is biased by some specified amount. These two probabilities can be calculated exactly for this case. Similarly, wings do indeed exist, but we can still ask about the likelihood that they have evolved gradually or in bursts, or whether their evolution correlates to change in other characters. A p-value can be attached to a phylogenetic hypothesis (that is, to a phylogeny) itself. This is not a statement of the probability that the species evolved (we know that they did), but rather one of the probability that they evolved according to the branching pattern and evolutionary processes implied by the phylogeny.

## Molecular clocks and the phylogenetic dog

Phylogenies not only are necessary to doing comparative tests properly, they may even in some cases define how to measure the phenomenon at hand. Fitch *et al.* (1991) presents striking evidence for a regular molecular clock of evolution in two different genes from the influenza virus. Fitch's data also illustrate how ignoring phylogeny when characterising the evolutionary process can produce misleading results. I will develop this example in some detail to show how a combination of theory and the phylogeny inform the search for the signal of the underlying evolutionary process; in this case, the signal of a molecular clock.

Samples of the influenza virus obtained at different times allow one to construct a phylogeny with a measured temporal component. The inferred number of nucleotide substitutions to the gene as measured from the root of the phylogeny can then be plotted against the date the sequence was obtained, yielding the plots in Fig. 3. These show that the rate of evolution is surprisingly regular for each gene — metronome like in fact — although ticking at a different speed in the two genes. Or do the data show this?

Each data point in the plot of Fig. 3 measures on the y-axis the genetic distance from a tip of the phylogeny down to the root, and on the x-axis the year the sample was drawn. Because of the peculiar 'cactus' like structure of these phylogenies, this procedure means that the genetic divergence represented along the stem of the tree is counted repeatedly in the graph. For example, the 1985 sample includes all of the divergence save for the length of the tip leading to the 1987 sample. The plotted points, in effect, are successive sums, each successive point being in large measure the sum of the previous ones, plus a small measure of independent change. The artifical data in Fig. 4 show how quickly successive sums come to produce straight lines with little variation. The data for these plots were generated from either poisson or uniform random number generators with the same mean, and can be thought of as representing rates of evolution in branches of a phylogeny. The poisson process is often used to model rates of nucleotide substitutions (Gillespie 1991). The uniform distribution was chosen as a process with no obvious
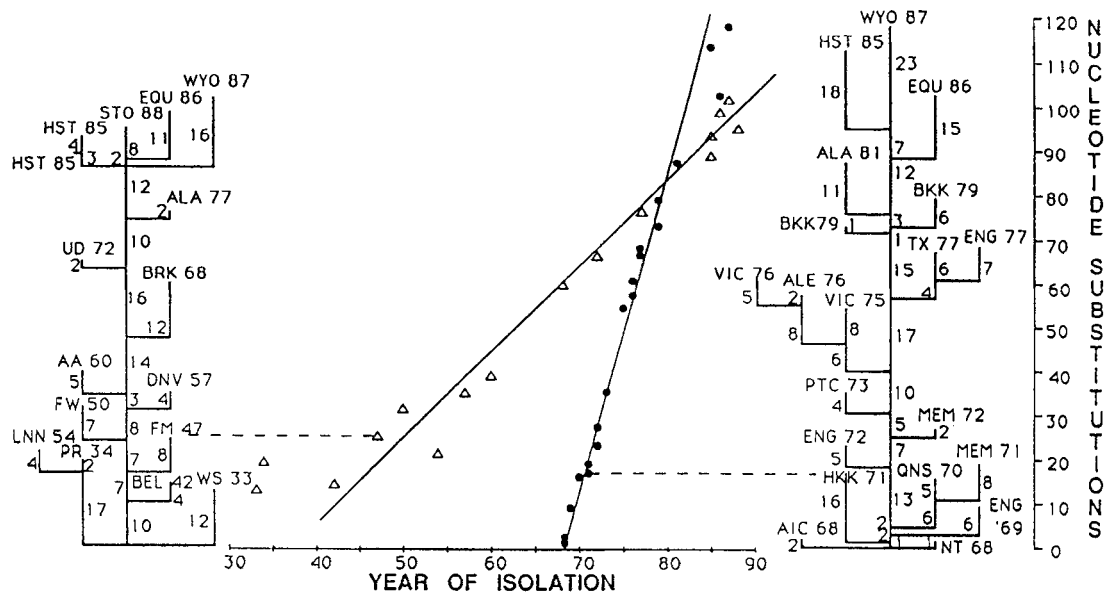
*Fig. 3.* Phylogenies of two genes in the influenza virus. The gene-tree of the non-structural gene is on the left and that of the haemagglutinin gene is on the right. Numbers following the abbreviations at the tips indicate the year that the sample was drawn. The data points plot the number of nucleotide substitutions to each gene when compared to the root, against the year that the sample was drawn, with the non-structural gene plot represented by the open triangles. From Fitch *et al.* (1991) who suggests a regular molecular clock running at different rates in the two genes.
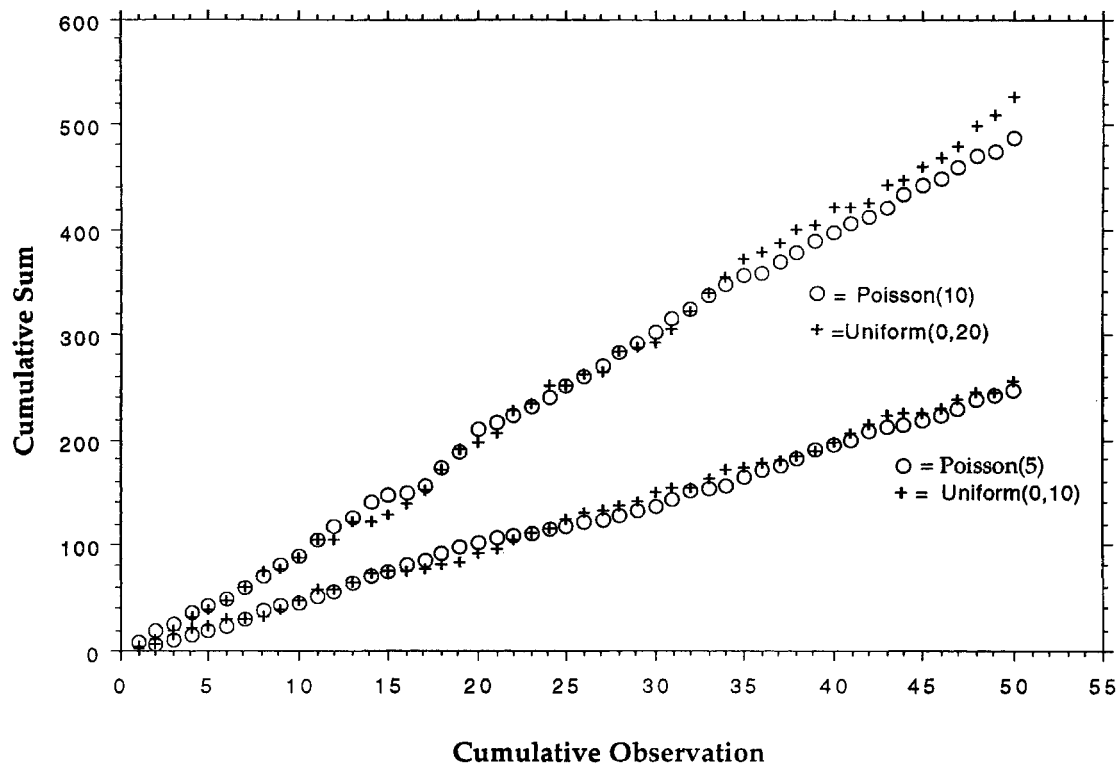


*Fig. 4.* Artificial molecular clocks. The data represent the cumulative sum of a series of random numbers plotted against the number of observations in the sum. The mean of the poisson random numbers is in parentheses. The mean of the uniform random numbers is the mid-point of its range. Poisson and uniform series produce indistinguishable and metronome like 'molecular clocks' despite having very different variances and representing different underlying processes.

biological underpinning. The random numbers were added to get their cumulative sum which is plotted against the number of observations in the sum (1, 2, 3 and so on) in Fig. 4.

The artificial molecular clocks qualitatively resemble those Fitch *et al.* (1991) reports. The poisson and uniform cumulative sums are virtually indistinguishable, and the collections of points have slopes equal to their mean 'rates'. These results give a clear hint that the regularity of the

clocks in Fig. 3 may be misleading. Successive sums tell virtually nothing about the underlying process, apart from the mean. 'Molecular clocks' whose rates are distributed uniformly on some interval are indistinguishable from those that conform to the poisson process. The nearly straight lines of data points, far from saying that the underlying process is regular and constant, hide, in the case of the uniform random numbers, the fact that the next 'rate' to be observed can fall with equal probability anywhere

on the sampling interval. Were one to use the molecular clocks of Fig. 3 to predict the time to a common ancestor given two or more samples, the confidence intervals on the estimate would be far too narrow.

To account for the redundancy in the data Fitch plots, one can use the phylogeny to define a collection of rates which represent a non-overlapping sample from the underlying process. Figure 5 shows what happens to Fitch's data if all of the phylogenetic overlap in the influenza data points is removed. Each point in these figures measures the amount of evolution and the time along a single branch of the phylogeny. These points now show much more scatter, more in line with what one expects of a stochastic molecular clock. Ignoring phylogeny gives a misleading idea as to the nature of the underlying evolutionary process.

*Characterising the molecular clock process*

But what is the evolutionary process that Fitch *et al.* (1991) reports? Each point in Fig. 5 represents an amount of evolution per unit time, potentially an observation from a poisson process. Using the likelihood procedure defined in the previous section, one can define M in terms of the poisson distribution, and ask what poisson rate maximises the likelihood of observing these data. Formally, this requires maximising

$$L(u) = \prod_{i=1}^{n} POI_i(r_i|u = \lambda t)$$

where

$$POI_i(r_i|u = \lambda t) = \frac{\lambda t_i^{r_i} Exp\,(-\lambda t_i)}{r_i!}$$

designates the probability of observing *r* events (nucleotide substitutions) in time *t* given a poisson process with instantaneous rate $\lambda$, where $\lambda$ is the mean rate of substitution per locus per unit time, and is the evolutionary process to be estimated. The separate branches of Fitch's phylogeny each provide values of *t* and *r*, and from these the value of $\lambda$ that makes the set of observations most likely can be estimated (more advanced models allowing for multiple hits and other eventualities — such as gamma rate variation models — could be used, if these have not already been employed to infer the number of substitutions per branch of the phylogenetic tree itself).

The likelihood $L(u)$ is the product of the poisson probabilities of observing the *n* different rates (one for each branch of the tree) defined by the phylogeny. The histograms in Fig. 6 show the individual rates (obtained from Fig. 5 by dividing the number of substitutions by the time axis). These rates are the raw units of the evolutionary process and can be shown to fit poisson distributions quite well, with mean rates of $u=2.6$ and $u=6.8$, respectively. These are quite close to the slopes of the lines from Fitch's plots, but this clock is now revealed to be a quite variable one. Confidence intervals constructed around these rates will be much wider than those inferred from Fitch's plots. Consequently, estimates of the time to a most recent common ancestor of these viral sequences will also have much wider confidence intervals that estimates constructed from Fitch's analyses.

Fitch's data are unusual in that the samples have known ages. More typically, the rate of the molecular clock is obtained from a fossil or other dated sample and then this rate is used to infer the ages of samples for which only the inferred number of nucleotide changes is known. In such cases the variability in estimating both the molecular clock itself as well as in estimating the inferred rate of nucleotide changes per unit branch length (as above) must be considered. A recent paper (Vilá *et al.* 1997) reported a most likely age for the origin of domesticated dogs of approximately 135,000 years, based upon a molecular clock estimate applied to variation amongst dog breeds in a 261 base-pair sequence of the mitochondrial DNA control region. The estimated age is interesting in the light of archaeological evidence suggesting a much more recent origin of domestic dogs of roughly 14,000 years. Taking into account the variation in the gene sequence data and in the molecular clock (assuming a poisson substitution model), the 95% confidence intervals around the 135,000
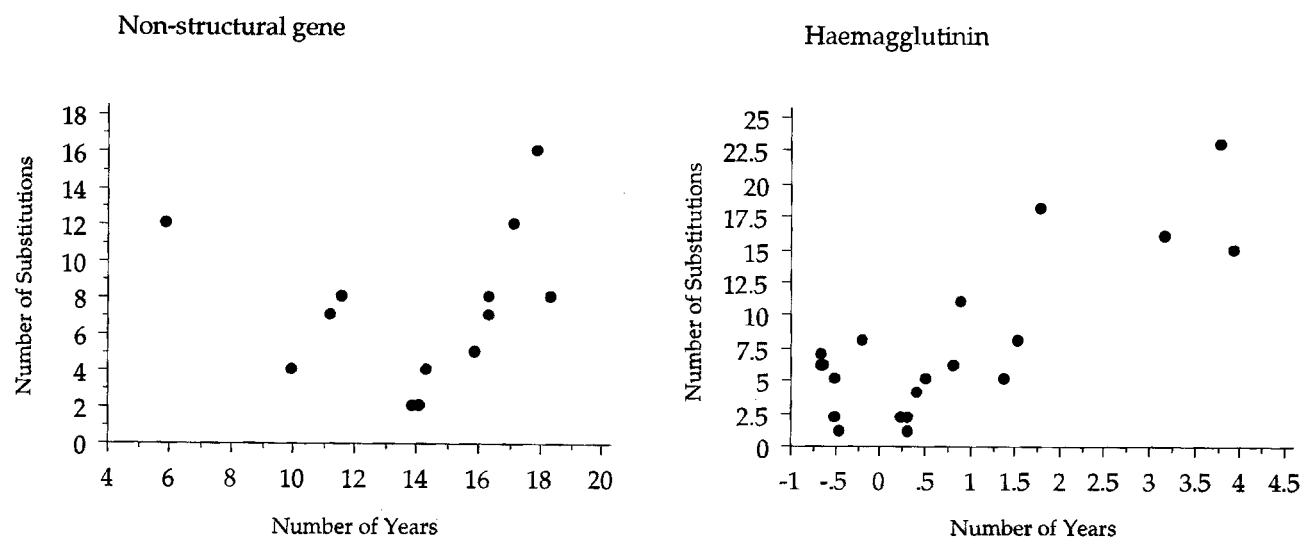


*Fig. 5.* Number of substitutions per branch for the two influenza genes of Fig. 3, plotted against the length of the branch. These data remove the overlap that is present in the data of Fig. 3 and reveal a much less regular clock of evolution.
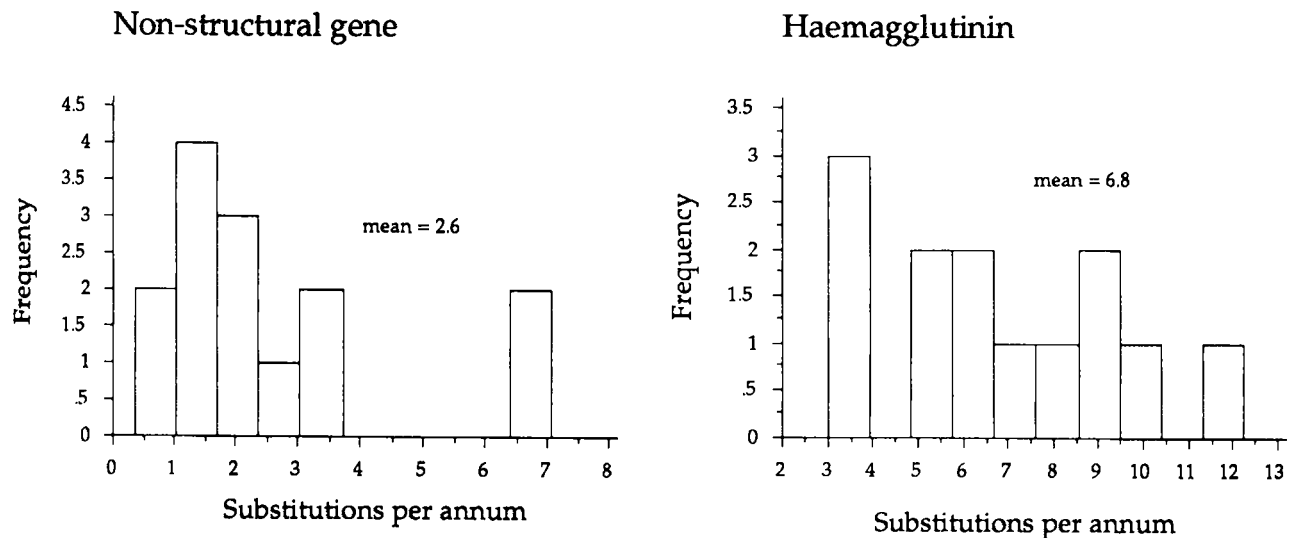
## Non-structural gene

## Haemagglutinin



*Fig. 6.* Histograms of the rates derived from the data of Fig. 5. Under the assumption of a poisson molecular clock, these data should conform to poisson curves with means given by the mean rate (see text). These data reveal the variability in the molecular clock that is hidden by the plots of Fig. 3.

year ago figure include 14,000 (Pagel, unpublished). The 135,000 date remains the most likely, but by conventional criteria, the more commonly accepted 14,000 figure cannot be ruled out under the poisson model.

### Evolutionary change in continuous characters: the generalised least squares model

Molecular clocks arise from the covariation of a process of substitution and time. In the foregoing example I decomposed this evolutionary process into an observed rate in each branch of the phylogeny. This allowed examination of the statistical nature of the process itself. More generally one might be interested in how some quantitative trait evolves with time or with genetic distance. Alternatively, researchers frequently will be interested in whether two or more traits covary among a group of species. Either of these general phenomena can be investigated solely from data on species and a phylogeny, by means of a standard statistical model known as generalised least squares, or GLS (e.g., Johnston 1963; Weisberg 1985).

*Investigating the evolution of a single character*

The GLS approach can be applied to characterise how a variable evolves in 'operational time' (real time, genetic distance, or other measures of divergence). In this context, one is not asking whether two traits of the species covary, but rather, whether a trait covaries with evolutionary divergence itself. Comparative methods for the analysis of continuous traits assume that traits accumulate variance at each tick of the evolutionary clock according to some constant variance (sometimes called 'Brownian motion') process. The GLS model treats each species' value of the trait as an observation to be predicted from the regression of the trait on total path length from the root to the species. Denoting the trait observed in the ith species $y_i$, then $y_i = \alpha + \beta \Sigma t_i + e_i$, where $\alpha$ is the y-axis intercept of the $y$ values regressed on total path length, $\beta$ is the slope of the line relating $y$ to operational time (here measured in units

of genetic or other divergence as time from the root to the tips will be the same for all species), the summation is over all of the branches leading to the $i^{th}$ species, and $e_i$ is random error. Effectively, one estimates the correlation between the value of a trait and total divergence. As will be discussed below, the ability to estimate $\alpha$ is an important difference between the GLS and 'independent contrast' approaches. Apart from estimating $\alpha$, the GLS model does not explicitly calculate values at the internal nodes of the phylogeny. The value of the trait in each species is presumed to arise from a process in which the trait evolves along each branch of the tree at some rate $\beta$ per unit branch length. Variation in trait values amongst species arises from differing path lengths from the root to the tips. The parameter $\beta$, then, defines the evolutionary process to be estimated, given the phylogeny.

The value of $\beta$ could be estimated by a simple linear regression were the species independent. However, owing to the phylogenetic structure of the data, the species will share some portion of the path lengths leading from the root to the tips. For closely related species this may be most of the path length. As a consequence, unless evolutionary change per branch is large and unrelated to phylogeny, closely related species will tend to have similar values of the trait. The estimate of $\beta$ must be adjusted for this non-independence. The phylogeny defines the expected covariance (lack of independence) between any pair of species as proportional to their shared path length from the root. A species' expected variance is proportional to its total path length. The standard GLS maximum likelihood estimate of $\beta$ is obtained in this context from solving

$$\beta(T'V^{-1}T)(T'V^{-1}Y),$$

where $T$ is a vector of length $n$ containing the total path lengths from the root to the tips and contains one value for each of the $n$ species, $V$ is a $n \times n$ variance-covariance matrix of the shared path lengths among all species (Fig. 7), and $Y$ is a $n$-element vector of the observed $y$ values, one for each species. Common regression models assume that all of the off-diagonal elements of $V$ are zero (species are independent). The value of $\beta$ defined by this formula
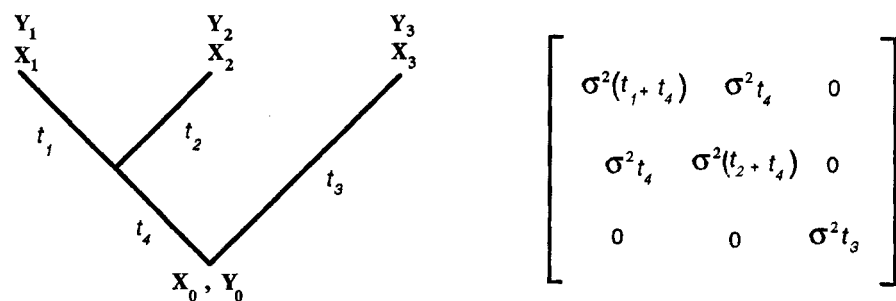
**Variance-Covariance matrix**



*Fig. 7.* An example of a phylogeny with branch lengths and the variance–covariance matrix it implies. Under a constant variance model of evolution, the expected variability in a character is directly proportional to the time it has been evolving. The expected covariance between any two species is proportional to their shared branch lengths. The variance–covariance matrix shows how the phylogeny specifies these amounts. The proportionality constant is the variance of the evolutionary process.

can be shown to maximise the likelihood of observing the set of $y$ values across species given the phylogeny, and the evolutionary process defined by $\beta$. Formally $\beta$ maximises

$$L(\beta) = \prod_{i=1}^{n} P_i(y_i | Tr, \beta).$$

To estimate $\alpha$ one merely replaces the T vector above by a matrix in which the first column is all ones, and the second column is the total path lengths as before. Then $\beta$ is a vector containing two elements, $\alpha$ and the estimate of the slope $\beta$ discussed before.

The significance of $\beta$ is tested by means of the LR statistic. The goodness of fit of a model in which the regression coefficient (or covariance or correlation coefficient-all equivalent hypothesis tests) is allowed to take its maximum likelihood value, is compared to a model in which the regression is set to zero or to some constant, c:

$$LR = -2\log_e[L(\beta = c)/L(\beta = \hat{\beta}],$$

The two models differ by one parameter if only a single regression is estimated.

### An application to gc content: testing neutral theory

Kimura (1968) suggested that the fate of most mutations is governed more by the stochastic effects of drift than by selection. The $G + C$ nucleotide content of a gene or genome can be used to test the hypothesis that variation in GC content amongst species arises from neutral drift, against the hypothesis that variation in GC content arises from selection. Sueoka as early as 1962 offered the idea of 'GC pressure' to explain the large variation observed in GC content across mammals. Sueoka (Sueoka 1962, Sueoka 1988) suggests various forces bias nucleotide substitutions upwards or downwards from the fifty percent $G + C$ content expected given neutrality. Figure 8 shows a phylogeny of 40 mammal species derived from an array of sources (Krakauer and Pagel 1998). We used this phylogeny in combination with data on the prion gene to test a number of hypotheses about the evolution of the prion gene in mammals (see also Krakauer *et al.* 1996, and below), including ideas about its GC content.

If some systematic and directional 'pressure' influences the GC content of the prion gene, then total GC content of the gene should increase or decrease with evolutionary distance of the species from the root of the tree. Using the measure $\mu_D$ of Jermiin *et al.* (1994), the total synonymous (i.e., silent replacements) GC content for each species can be calculated and then correlated with the total genetic distance of the species from the root of the tree. Silent sites should be relatively free of the effects of selection and therefore provide a useful test of Sueoka's ideas against those of neutral theory. Figure 9 shows that there is a slight negative correlation ($r = -0.15$, $p \approx 0.15$) suggesting a tendency for GC content to decline as evolutionary divergence increases. Neutral theory expects no correlation (although apparent correlations could arise from a random walk). The data suggest that with respect to synonymous GC substitutions in mammalian prion genes, Kimura may be wrong.

This result, however, ignores phylogeny. I have applied the GLS approach to the data of Figs 8 and 9. The GLS estimate of the correlation (equivalent to estimating $\beta$ but of more immediate interest here) is $r = -0.024$, substantially lower than the value obtained across species. The LR value is 0.01, which is far less than the critical value of chi-squared with one degree of freedom of 3.85. To the extent that this analysis is representative it suggests that GC content does not have any tendency to vary with genetic divergence, unlike the result perhaps implied from the analysis conducted across species.

Why the difference in the two results? The GLS method estimates the slope or correlation that must have held along each branch of the tree. By doing this it will be less influenced by large differences in GC content between taxonomic groups, differences that may have evolved only once, but then been retained in each of the species in that group. Krakauer and Pagel (1998) report that taxonomic groups show large differences in total GC content.

### Characterising the nature of the evolutionary process in a single trait

The GC example illustrates how one can analyse the evolution of a single character and study its relationship to operational time in the phylogeny. It is also possible to ask whether variance in some variable, such as GC content, accumulates as a constant function of branch length, as is assumed by the statistic or whether it evolves at a constant rate in some transformation of the branch lengths. The hypothesis of constant variance can be investigated by repeating the previous analysis while simultaneously trans-
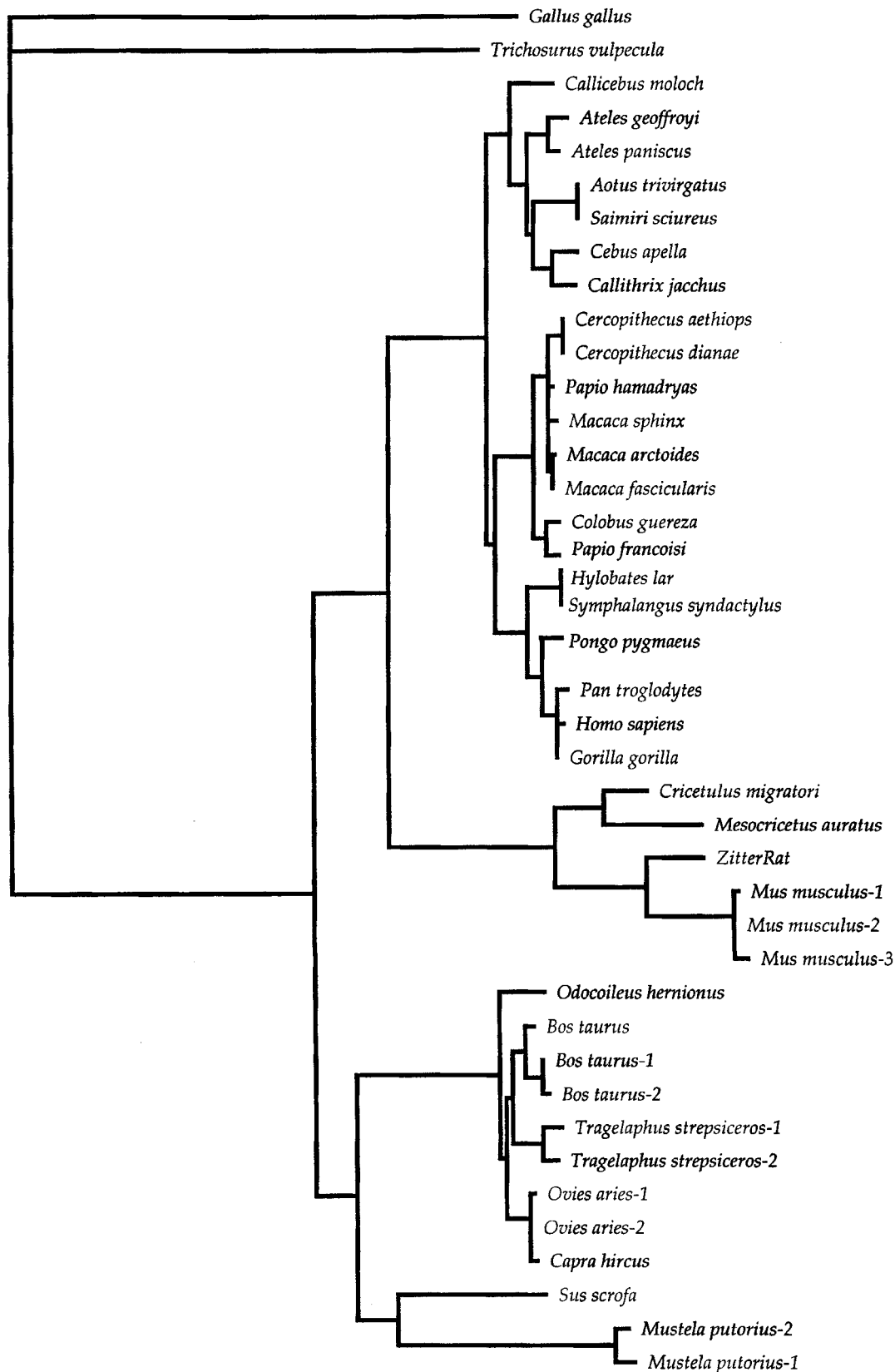
*Fig. 8.* Phylogeny of the prion protein for 41 mammal species, and an avian outgroup. The phylogeny is a best-guess composite species-tree derived from a number of literature sources (Krakauer *et al.* 1998).

forming the total path length non-linearly via a parameter I call $\kappa$ (Pagel 1994). This parameter is estimated by raising branch lengths to the power $\kappa$, and finding that value of $\kappa$ that maximises the overall likelihood. If the maximum likelihood estimate of $\kappa$ is not significantly different from 1.0, then the constant variance model provides a reason-

able fit to the untransformed data. The maximum likelihood estimate of $\kappa$ for these data was 0.87, which is not significantly different from 1.0 by a likelihood ratio statistic (Table I): a process of Brownian motion cannot be rejected as descriptive of the synonymous GC content in the prion gene.
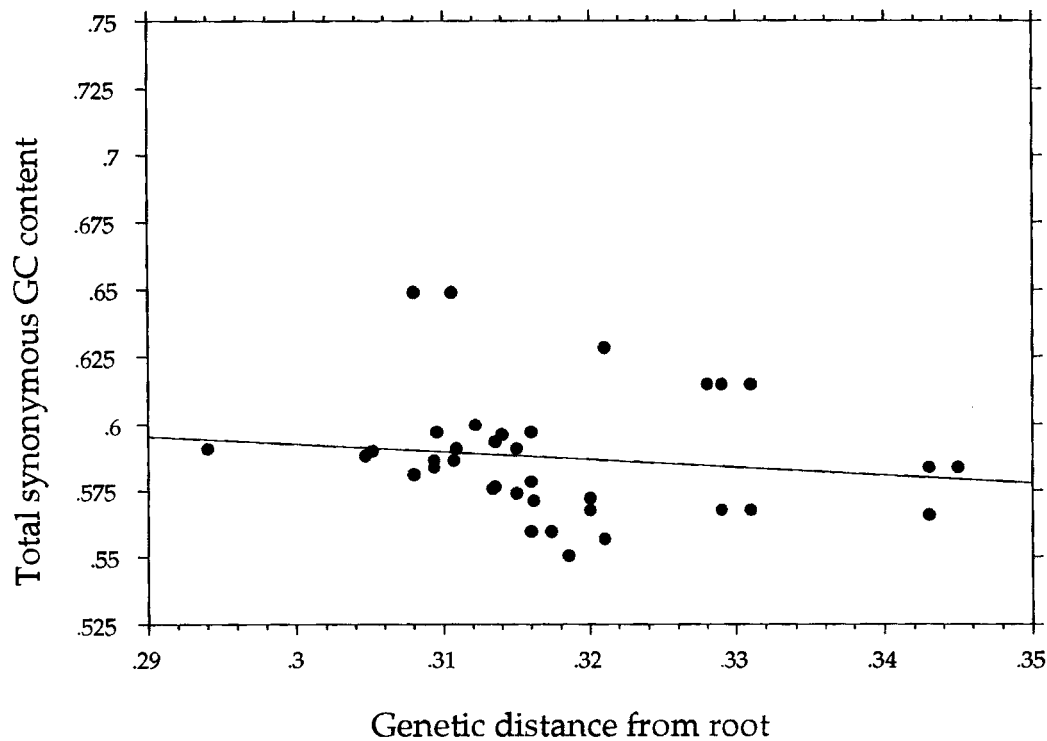
*Fig. 9.* The synonymous GC content of the prion protein gene plotted against the total genetic path length from the root to each species. The correlation between the two measures is $r = -0.15$, $p \approx 0.15$. Taking into account phylogenetic associations in the GLS model (text) reduces the correlation to $r = -0.024$.

Table I. *Transformation of the total path length from the root to the species to produce the best fit to the constant variance model, applied to prion protein data from Fig. 8. Total path length is raised to the power $\kappa$. Various values of $k$ yield better or worse likelihoods. The best fitting $\kappa$ (0.87, likelihood ratio closest to zero) is not significantly different from a $\kappa$ of 1.0.*

| Value of $\kappa$ | log-likelihood | Likelihood Ratio versus $\kappa = 1.0$ |
|---|---|---|
| 0.50 | −298.8 | 8.8, p < 0.01 |
| 0.75 | −294.4 | 0.0, ns |
| 0.87* | −294.1 | 0.6, ns |
| 1.00 | −294.4 | |
| 1.25 | −296.5 | 4.2, p < 0.05 |
| 1.5 | −301.5 | 14.2, p < 0.001 |

* 0.87 is the maximum likelihood estimate for these data

Estimating $\kappa$ can provide insights into the nature of the evolution of a trait. Assume for sake of illustration that all total path lengths are greater than one. Values of $\kappa$ also greater than 1.0 for such data lengthens long path lengths disproportionately to shorter ones. This might suggest that variance in the trait accumulates faster than overall divergence. The reverse holds for values of $\kappa$ less than 1.0: longer branch lengths are disproportionately shortened, suggesting that variance in the trait reaches an upper (or lower) limit while overall divergence is still progressing. A value of $\kappa$ not different from 0.0 suggests that evolutionary change is independent of branch length, as might be expected from a punctuational view of evolution.

*The association between two characters evolving along the branches of the phylogeny*

The more common use of the GLS is to investigate whether two or more characters covary on the phylogeny, that is, to estimate the 'evolutionary regression coefficient' (Pagel

1993). The implicit model of evolution for this problem is $y_i = \alpha + \beta x_i + e_i$, signifying that some variable $y$ is a linear function of another variable $x$, plus the $y$-axis intercept and some error. The parameter $\beta$ now is the regression of the $y$ variable on the $x$ variable rather than the regression of $y$ on total path length. Once again the model allows estimation of $\alpha$, which is lost in the independent contrast approaches. The variables $y$ and $x$ are traits observed in a number of species. Specifically, $\beta$ estimates the amount of change in $y$ for a unit change in $x$ along the branches of the phylogenetic tree, that is, how $y$ changed as $x$ evolved over time. This approach can easily be extended to any number of predictor variables. The likelihood function is formalised as a statement of the likelihood of observing the sets of $y$ and $x$ variables in the species, given the phylogeny $(Tr)$, and the regression coefficient, $\beta$. This likelihood can be written as

$$L(\beta) = \prod_{i=1}^{n} P_i(x_i, y_i | Tr, \beta),$$

where now one estimates the value of the parameter $\beta$ that makes the pairs of data points $(x,y)$ most likely given the phylogeny and the model of evolution that includes $\beta$. This value can be found analytically by least squares according to the well known formula

$$\beta = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

where $X$ is a $n \times m$ matrix of the $X$ or predictor variables, n is the number of species and m is the number of predictor variables, $V$ is, as before, an $n \times n$ variance-covariance matrix of the shared path lengths in the phylogeny, $Y$ is the vector of observations on the variable to be explained by the $X$ variables, and $\beta$ is the maximum likelihood estimator. If there is more than one $X$ variable, then $\beta$ is a

vector containing a regression coefficient corresponding to each variable. With only one $X$ variable, only a single regression coefficient is estimated. The regression coefficient is tested for significance by a likelihood ratio test, in which the likelihood of a model in which $\beta$ is forced to be zero or some constant is compared to a model in which $\beta$ is allowed to take its maximum likelihood value.

As with the analyses described for a single variable it is possible with two or more variables to estimate a scaling parameter $\kappa$ that transforms the branch lengths non-linearly (Pagel 1994). By using $\kappa$ one can find the branch length transformation that maximises the association between two traits, thereby revealing in what scale they can best be described as evolving linearly with respect to each other.

### Differences between GLS and 'Independent Contrasts' approaches

The GLS estimator of the regression coefficient accounts for the lack of independence among species without forming 'independent contrasts'. The independent contrasts techniques (Felsenstein 1985; Grafen 1989; Pagel and Harvey 1989a; Harvey and Pagel 1991; Pagel 1992; Garland *et al.* 1992) calculate differences or 'contrasts' between pairs or groupings of species or other nodes of the phylogeny. They can all be seen as attempts to use the phylogeny to define data points that are statistically independent. With unresolved phylogenies these contrast techniques are still useful. However, with a resolved phylogeny, independent contrasts are not required. The GLS approach solves the problem of non-independence, as before, by specifying the nature of the expected correlation among species via the variance–covariance matrix, $V$, which is given by the phylogeny (Fig. 7).

Independent contrast approaches remove information about the y-axis intercept, and cannot estimate $\alpha$ as defined in the equations above. The GLS approach can estimate $\alpha$, and can even find values of $\alpha$ that fall outside of the range of observed values among species. In practice all ancestral character state reconstruction methods estimate ancestral values that fall within the range of observed variation. Thus, for example, the estimated body size at the root of the mammalian tree will fall somewhere between the largest and smallest mammals included in the sample. The GLS approach is not restricted this way. If the relationship between the value of a character and total path length is strong, the estimate of $\alpha$ may fall well outside the range of observed scores. On the other hand, $\alpha$ may be estimated poorly (have a large standard error) if total path length from the root to the tip is not predictive of the value of the trait in a species. A method of analysis known as restricted maximum likelihood (REML) estimation (e.g., Rao and Kleffe 1988) can be applied to the regression problem to get estimates of $\beta$ that do not depend upon estimating $\alpha$. The REML solution is roughly equivalent to finding $\beta$ while integrating over all possible values of $\alpha$.

GLS estimates of the evolutionary regression coefficient are expected in some instances to be freer of phylogenetic effects than independent contrasts approaches. Accordingly, the GLS may give different answers to previously well studied problems (see also Hansen and Martins (1996)

for an application of the GLS approach incorporating a stabilizing selection model). I have used the GLS approach to estimate the evolutionary regression coefficient relating brain size to body size in Old World primates, using Purvis' (Purvis 1995) primate phylogeny and branch lengths. A large number of studies of brain and body size evolution suggests a value of 0.67 to 0.75 for this important parameter and much theory depends upon its value (see Pagel and Harvey 1989b). The GLS model returns a correlation of 0.94 but more importantly a regression coefficient of 0.48 (Table II). The GLS estimate of 0.48 may be much closer to the true value. Pagel and Harvey (1989b) found a regression of approximately 0.75 within each of the three primate sub-orders. However, the average of 18 regression coefficients of brain on body size fitted separately within primate genera — estimates free in principle of phylogenetic effects — was 0.47. If the GLS estimates holds up in additional analyses, theory about the evolutionary relationship of brain size to body size may be called into question.

### Analysis of discrete characters: the continuous-time Markov model

The evolution of discrete characters on a phylogeny provides investigators with the opportunity to test ideas about likely ancestral character states, about which characters evolve first and which follow, and about correlations among characters. The approach I have described elsewhere (Pagel 1994) and called 'Discrete' adopts a markov-transition rate model of trait evolution within a maximum likelihood framework. The method fits by maximum likelihood a continuous-time markov model to the data in which the two binary characters are allowed to evolve on the tree independently. Each character can adopt two states, 0 or 1. To calculate transitions from one state to the other requires two parameters per character, or four total for the model of independence (Fig. 10). Then, the method fits a model in which the two characters evolve in a correlated fashion. This is achieved by considering the four possible states that two binary characters can adopt (0,0; 0,1; 1,0; 1,1) and allowing one of the variables to change state in any branch of the tree, yielding eight possible transitions. Transitions of both characters simultaneously are estimated as combinations of single-state transitions. This model therefore requires eight parameters (see also Pagel 1994).

In the $L = f(D|Tr,M)$ structure, $M$ contains the markov-model specifying the transition rates among the character states for the model of independence and for the model

Table II. *Estimating the evolutionary regression coefficient for brain weight on body weight in Old World Primates. Ordinary across-species regressions on log-log plots typically yield slopes of between 0.67 and 0.75. A set of within-genera analyses (see text) yielded a slope of 0.47. The GLS model (text) agrees with this lower figure.*

| Analysis | Regression Coefficient |
| --- | --- |
| Across-species | 0.67–0.75 |
| Within-genera | 0.47 |
| Generalised Least Squares | 0.48 |

## Independent transitions in two binary traits

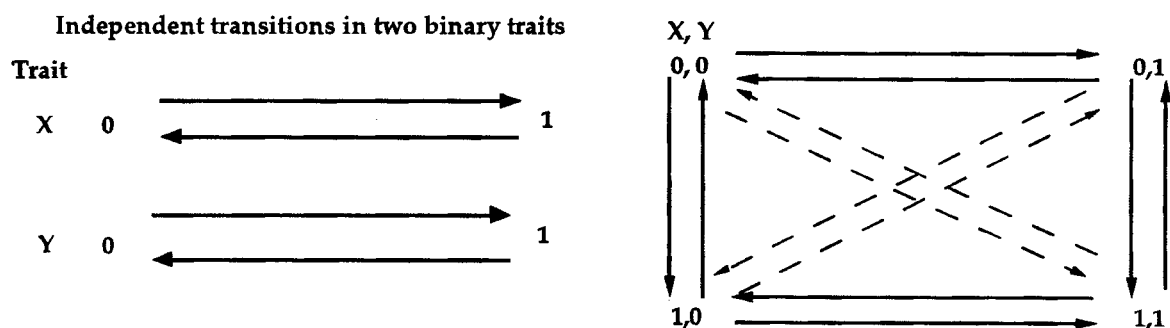## Linked transitions between two binary traits



*Fig. 10.* Diagrams of the model of independent evolution and the model of correlated evolution (Pagel 1994, and text). To characterise evolutionary transitions in two binary traits assumed to be evolving independently requires estimating the rate of transitions from state '0' to state '1' and back again for each trait, or four traits total. When the transition of one character depends upon the background state of the other character, eight parameters are required corresponding to the states shown in the right hand diagram. For technical reasons (Pagel 1994, and text), the dashed arrows representing 'dual' transitions are forced to be zero. This enables one to test whether the transition from, say, state '0,0' to state '1,1' proceeds via '0,1' or via '1,0', thereby yielding a test of the direction of evolutionary change.

of correlated evolution. The parameters estimated in the markov-model describe instantaneous rates of evolution along branches of the tree, and are chosen by maximum likelihood to make the observed data among species most likely given the underlying phylogeny. The models are fitted allowing the character states at each node to take all possible values. This removes any dependency of the results on the particular set of ancestral values that might be assigned by parsimony or other reconstructions of the ancestral nodes, and means that the problem of ambiguity in ancestral character state reconstructions is avoided.

The method tests an 'omnibus' hypothesis of correlated evolution by comparing the likelihoods of the two different models fitted to the observed data set: L(I) represents the likelihood associated with the model of independent evolution, and L(C) denotes the model of correlated evolution. The overall model of correlated evolution is justified as a representation of the data if its likelihood is significantly larger than the model of independent change. Significance is assessed by means of the likelihood ratio statistic $LR = -2\log_e[L(I)/L(C)]$. These two models differ by four parameters (see Pagel 1994), and the LR statistic is in the limit distributed as a chi-squared variate with four degrees of freedom. Analysis of single parameters or of combinations of the parameters can be used to test hypotheses about the order and direction of change in the discrete characters (Lutzoni and Pagel 1997), and to test ideas about the rates of change *per se*. Ward and Seely (1996) present an example of the use of Discrete in their investigations of physiological adaptations in beetles, Cotgreave and Pagel (1997) use it to investigate conservation status in birds, and Mace and Pagel (1997) and Holden and Mace (1997) apply it to phylogenies of human cultures. Price and Birch (1996) and Schluter (1995) provide other examples.

Some commentators have criticised the continuous-time markov approach for finding the likelihood over all branches of the phylogeny (Read and Nee 1995; Grafen and Ridley 1996), thereby failing to distinguish independent evolutionary change from similarity by descent, and leading to inflated Type I error rates. This is a common misunderstanding of the maximum likelihood approach both as applied to estimating evolutionary processes and as applied to estimating phylogenies. The misunderstanding arises for both cases from the belief that the size of the

likelihood ratio statistic is a function of the number of branches in the tree. The likelihood ratio statistic is in fact independent of the size of the phylogeny, but is an increasing function of the difference in the number of parameters in the 'null' versus 'alternative' model. The more parameters that the 'alternative' model contains compared to the 'null' model, the larger the likelihood ratio will be by chance alone. Counting each branch or even dividing each branch into an infinite number of smaller epochs, as the mathematical structure implicitly does, has no influence on the likelihood ratio statistic.

Computer simulation studies of Discrete show that it has Type I error properties close to theoretical expectations Table III). The simulations reported in Table III were produced by independently evolving two characters along the branches of a phylogeny. Phylogenies were taken from the literature or in some cases were generated randomly. The markov model of independence was used to simulate the data, with transition rate parameter values derived either from real data, or chosen to produce differing amounts of variation in the simulated data. The simulated data were then analysed with the model of independence and the model of correlated change, to find the likelihood ratio. This process is repeated to produce a null hypothesis distribution of the LR statistic. If the LR statistic is asymptotically distributed as a chi-squared variate then the mean of the null hypothesis distribution of LR values rounded to the nearest integer should specify the number of degrees of fredom of the chi-squared distribution to which the simulated distribution will most closely correspond.

This is what is observed. The distribution of the likelihood ratio statistic when the null hypothesis is true follows quite closely the theoretical chi-squared distribution with, in the limit, four degrees of freedom (Table III). Generally, phylogenies with more species yield LR statistics with more degrees of freedom as the statistic becomes asymptotic, although the extent of this effect is governed by the amount of variation in the data. Smaller phylogenies or phylogenies with only small amounts of variation at the tips (i.e., either few character changes or character change at equilibrium) tend to produce LR values that are distributed with fewer degrees of fedom. Table III shows that the method conforms closely to chi-squared distributions of 3, 2, or 1 degree of freedom for such cases.

Table III. *Type I error rates for Discrete. Independent character evolution was simulated along the branches of the sixteen phylogenies in the table according to the model of Pagel (1994). The phylogenies come from a variety of literature sources or were generated by computer, and represent a range of types of topologies and branch lengths. Names are arbitrary. No. of tips is the size of the phylogeny, No. of replicates refers to the number of simulation runs on which the statistics are calculated, the approximate $\chi^2$ df is the degrees of freedom of the theoretical chi-squared distribution against which the simulations were assessed, the mean LR and variance are the mean and variance of the simulated likelihood ratios, the $\alpha$ levels refer to the upper tails of the chi-squared distribution. The simulated distributions are expected to be chi-squared with df equal to the mean LR (i.e., non-integer), but show quite close fits to the nearest integer degrees of freedom. The expected variance of a chi-squared is $2*df$.*

| Phylogeny | No. of tips | No. of replicates | Approximate $\chi^2$ df | Mean LR (Variance) | $\alpha = 0.10$ | $\alpha = 0.05$ | $\alpha = 0.01$ |
|---|---|---|---|---|---|---|---|
| 120.1 | 120 | 176 | 4 | 3.94 (7.07) | 0.074 | 0.034 | 0.005 |
| Exp109 | 109 | 500 | 4 | 4.34 (8.49) | 0.102 | 0.060 | 0.010 |
| Primates | 67 | 100 | 4 | 4.02 (7.06) | 0.060 | 0.010 | 0.000 |
| Cecile | 301 | 108 | 4 | 4.13 (7.26) | 0.102 | 0.046 | 0.019 |
| threatened | 80 | 100 | 3 | 2.84 (4.37) | 0.110 | 0.030 | 0.000 |
| species.sims | 36 | 397 | 3 | 3.02 (5.15) | 0.115 | 0.047 | 0.003 |
| 5.8S | 30 | 100 | 3 | 3.11 (5.54) | 0.090 | 0.060 | 0.020 |
| 10.3 | 10 | 190 | 3 | 3.00 (4.89) | 0.100 | 0.026 | 0.000 |
| gene.sims | 36 | 467 | 2 | 2.06 (2.89) | 0.089 | 0.034 | 0.004 |
| Lekking | 18 | 84 | 2 | 1.50 (2.51) | 0.080 | 0.010 | 0.010 |
| rtrees 10.3 | 10 | 471 | 2 | 1.74 (3.33) | 0.063 | 0.027 | 0.006 |
| disc.test | 9 | 91 | 2 | 1.59 (3.43) | 0.055 | 0.011 | 0.011 |
| nrDNA25S | 30 | 98 | 1 | 1.24 (1.86) | 0.170 | 0.041 | 0.010 |
| ihd.monog | 30 | 93 | 1 | 1.25 (2.59) | 0.097 | 0.053 | 0.021 |
| 120.3 | 120 | 490 | 1 | 1.06 (2.09) | 0.110 | 0.043 | 0.010 |
| Ihd.monog.1 | 30 | 82 | 1 | 0.59 (0.63) | 0.024 | 0.012 | 0.000 |
| | | | | grand weighted means | **0.095** | **0.040** | **0.007** |

These results mean that the observed likelihood ratio statistic for any result when using Discrete can be compared to the chi-squared 4 df distribution. If the observed LR exceeds the critical value of the 4 df chi-squared statistic, then the null hypothesis can be rejected with confidence given by $1-\alpha$ or better. If the null hypothesis is not rejected against a chi-squared 4 df distribution, it may still be significant. In these cases it is recommended that the null hypothesis distribution be simulated (see Goldman 1990; Pagel 1994 describes the procedure for Discrete) to get exact p-values.

*Applying discrete to prion protein evolution*

The prion is a protein of approximately 270 amino acids that is normally expressed in the brain of adult vertebrate species. The protein can adopt at least two different three dimensional conformations. The normal constitutively expressed configuration can undergo modification to an altered pathogenic isoform. This form is resistant to normal proteases and therefore accumulates in the brain where it is likely involved in the formation of disease causing plaques. The pathogeneic prion protein is associated with the spongiform encephalopathies, including bovine spongiform encephalopathy (BSE), Creutzfeld-Jacob disease and fatal familial insomnia in humans, and feline spongiform encephalopathy, amongst others. It is hypothesised that the altered form can cause normal prion proteins in the brain spontaneously to adopt the altered or disease causing configuration, further promoting disease. Studies suggest that consumption of infected tissue can effect transmission of prions.

Much interest has focused on whether the prion responsible for BSE can infect humans. We investigated prion protein evolution in mammals by acquiring gene-sequence data on the 33 mammal species for which data were available at that time, yielding a phylogeny similar to that in Fig. 8 (Krakauer *et al.* 1996). This tree revealed an unusual case of convergence at the molecular level in two sites in

the prion gene of cattle and humans. In both cattle (*Bos taurus*) and the lineage including the great apes and humans (not including Orang-utans), we observed a substitution of asparagine to serine at site 143 of the gene and a substitution of tyrosine to histidine at site 155. We wished to know whether this pair of convergent amino acid substitutions was unlikely to have occurred by chance, and might therefore be evidence that for some reason there had been a selective convergence in these two lineages.

To investigate this question we studied these two amino acid substitutions as a pair of discrete changes on the phylogeny. Our observation, then, was of two variables changing from state '0' to state '1' and doing so twice on the tree simultaneously. In addition, it is possible that the two derived states were lost simultaneously in the Orangutan, and changed back to the ancestral amino acids. Our interest was to establish evidence for or against correlated evolutionary transitions from one amino acid to another. Evidence for an association indicates an underlying evolutionary process in which changes to one variable make it more or less likely that another will change. From an observed likelihood ratio of 10.9, and simulations of the null hypothesis distribution suggesting a chi-squared 2 df criterion, we estimated that the probability of a pair of events happening together on the tree by chance alone was about 0.005. From this we were able to estimate the expected total number of pairs of identical derived amino acids expected by chance on the tree. That is, given the overall variation observed in the prion protein, we calculated how many pairs of changes such as we observed could reasonably be expected if we included all amino acid positions of the protein that showed any variation whatsoever across mammals. The expectation was for substantially fewer than one. The events we observed are unlikely to be the result of chance convergence.

Since we produced these results, additional evidence shows that one of the two amino acids also evolves in two other lineages. Nevertheless, the fit of the data to the model of dependent change is still significantly better than to the

model of independent change (Krakauer and Pagel 1996). It may seem surprising that so few correlated changes can give rise to a significant result. Here one must bear in mind the hypothesis being tested. Under the null hypothesis, the assumption is that the two traits can evolve anywhere on the tree. The probability of a change (amino acid substitution) is small for any given branch. Given this, it is unlikely by chance alone that a pair of amino acid substitutions will happen in each of two branches rather than being scattered throughout the phylogeny. A very crude analogy is to imagine throwing darts at a target on which the phylogeny has been drawn. The probability of landing two pairs of darts right on top of each other is very low even if one is trying to do so: it is even lower if the throws are directed at random points on the target (the null hypothesis).

The convergence between cattle and humans at the two sites of the prion gene is interesting in view of the fact that BSE almost certainly transmits to humans (Collinge *et al.* 1996), and yet there is no evidence that scrapie, a similar disease in sheep, does. Sheep lack the two substitutions that cows and humans share. Sites 143 and 155 are also known to be involved in determining the 3-D structure of the prion protein (Riek *et al.* 1996).

### Ancestral character states and the evolutionary process

The likelihood approach to estimating the ancestral character states is to calculate a likelihood for each possible value the ancestral character could have taken. For a binary character there will be two likelihoods, one corresponding to fixing the state of the character at "0", the other to fixing it at "1". These 'partial likelihoods' for a node, correspond to the likelihood associated with that node and all of the branches 'above' it (closer to the tips) in the tree. The partial likelihoods at the root include all of the tree, those closer to the tips include less and less of it. They are useful for identifying how much more 'support' there is for one assignment of the ancestral state at a node compared to the other. Here, 'support' is used in a technical sense (Edwards 1972) to mean the difference between the two partial log-likelihoods, $\log[L_i/L_j]$, where the subscript "i" refers to the larger likelihood and "j" to the smaller.

The partial likelihood of a character state that is assigned unequivocally by parsimony may be only slightly larger than the less preferred character. In other cases the two approaches can differ in the value they assign. Schluter (1995) employed Discrete to re-analyse data on an artiodactyl ribonuclease. Schluter estimated the most likely ancestral state of the amino acid at position 38 of the ribonuclease, a critical determinant of enzyme catalytic activity. The best maximum likelihood estimate differed from parsimony-based estimates for several species. Schluter found considerable uncertainty associated with some ancestral positions as well as different evolutionary patterns from those derived from parsimony. Such information cannot be derived from the parsimony approach, and as Schluter argued only within the maximum likelihood framework is it possible to derive testable quantitative models to help choose among the alternatives.

Just as with the GLS model for continuous characters, it is possible to transform branch lengths within the discrete variables model to investigate the nature of the evolutionary process itself. The branch length transformation parameter $\kappa$ discussed previously is used to find that transformation that makes the data best fit the continuous-time markov model. A value of $\kappa = 0$ for discrete data suggests that transitions between character states are independent of branch length. This is consistent with a punctuational view of evolution, and also makes the model mathematically similar to a discrete-time markov approach.

### The order and direction of evolutionary change

The diagram of Fig. 10 highlights eight parameters corresponding to character state transitions in one character while the background character remains unchanged. Combinations of these parameters can be used to test the order of evolutionary change in the two traits (i.e., which changes first) and, in combination with information on the likely ancestral states, the direction of change.

Figure 10 shows two paths by which a pair of variables could evolve from '0,0' to '1,1'. If for example the transition '0,0' to '0,1' is not significant while that leading to '1,0' is, the implication is that the path to '1,1' proceeds via the latter route. One can then construct a 'flow diagram' describing the likely evolutionary pathways (see for example, Mace and Pagel 1997; and Pagel 1994 for further discussion of hypothesis testing with this model).

To draw inferences about the direction of evolutionary change requires first that one develops an explicit inference about the value at the root of the tree. This is easily accomplished within a likelihood framework by simply calculating the likelihood of one or the other states at the root. Then, it is a simple matter to test whether the transition parameter corresponding to changing away from the root value is statistically significant.

### Analysing rates of evolution and speciation: the likelihood approach

In this section I move away from questions of correlated evolution to a discussion of how the maximum likelihood approach can be applied to questions of estimating rates of evolution and speciation. Of interest is whether phylogenies can be used to test ideas about whether rates of evolution differ according to the value of some background character. This background 'character' could simply denote two different clades, or more interestingly denote a character that varies within clades. For example, does the adoption of a mutualistic lifestyle in fungi (Lutzoni and Pagel 1997) lead to increased rates of speciation and molecular evolution?

Several new methods use phylogenies to reconstruct speciation rates through time or ask whether two different clades provide evidence of differing speciation rates (for recent reviews, see Nee *et al.* 1996, see also Garland 1992). However, one may wish to ask questions about whether rates of evolution vary according to some background character that can change state within a given group over

time, such as shown in the diagram below (Fig. 11). Existing methods are not applicable to this problem. Models for speciation rates treat the whole phylogeny as one entity, and the models of differential cladogenesis compare one entire clade to another. Neither of these is useful when the character thought to influence the rate of speciation varies within a given clade.

The simple solution to the problem posed by the diagram is to write down the length of the branches associated with the character in state '1' (black boxes) and compare their mean to that of the branches corresponding to the character in state '0' (white boxes). Alternatively, one could ask whether sub-clades in state '0' are more or less speciose than sister-sub-clades in state '1'. Clades 'A' and 'B' provide one such comparison. These methods are limited, however, by the need in the first method to reconstruct the ancestral values of the background character: the ancestral states are ambiguous in this instance, as they often will be. The second method, though robust, can in this example make use of only a small amount of the information in the tree, and it does not provide estimates of evolutionary rates.

## A likelihood approach to estimating rates of evolution

A statistical approach to estimating rates of evolution can provide maximum likelihood estimates of evolutionary rates and rates of speciation. Consider again the tree in Fig. 11, and define two parameters $\lambda_1$ and $\lambda_2$ which describe the instantaneous rate of branching for branches in the phylogeny with the background character in state 1 versus in state 0. The reciprocals of $\lambda_1$ and $\lambda_2$ estimate the time (or branch length) per speciation event.

Each branch of the tree provides information about the rate of speciation or branching per unit length. If we adopt a model in which we assume that there is a constant probability of speciating per unit branch length (where 'spe-



*Fig. 11.* Arbitrary phylogeny showing species with one of two values of some binary trait (unfilled versus filled boxes). Ancestral character states for the two binary states are ambiguous thereby preventing a simple calculation of the average branch length associated with each state. A maximum likelihood procedure allowing the background states to vary can solve this problem and provide estimates of the rates of evolution associated with each state (see text).

ciation' is evidenced by a branching event in the phylogeny — an implicitly cladistic species concept), then, for inter-nodal branches the values of $\lambda$ can be estimated from the exponential model according to

$$P_{sp}(t_i) = (1 - Exp[-\lambda t_i]),$$

Here $P(t)$ is used to denote the probability of a speciation event in the time t, and the subscript '*sp*' denotes speciation. In this formula, $\lambda$ is the instantaneous branching rate. In branches leading to species or tips, no speciation or branching event has occurred and accordingly one wishes to estimate $\lambda$ from

$$\overline{P_{sp}}(t_i) = Exp[-\lambda t_i],$$

where the 'bar' over the $P(t)$ indicates that speciation has not occurred.

The exponential model fits an explicit model of speciation to the data. An alternative approach is merely to characterise the mean branch lengths corresponding to the two different background characters. The gamma distribution is useful for this purpose, describing a smoothly varying distribution with a lower limit of zero, and with a shape that can range from exponential to symmetric and normal-like. Let

$$Gam(\alpha, \beta) = \frac{t^{\alpha-1} Exp[-t/\beta]}{\alpha^{\beta} \Gamma(\alpha)}$$

designate the probability of observing a branch of length $t$ given that the branch lengths follow a gamma distribution with parameters $\alpha$ and $\beta$. Separate pairs of the parameters $\alpha$ and $\beta$ can then be estimated for the two different background character states. The mean of the gamma distribution is $\alpha\beta$.

The gamma model implicitly treats inter-nodal branches and branches leading to tips as equivalent samples of the rate of branching. The 'length' of a branch leading to a tip may underestimate the rate of branching given that the species at the tip is still evolving (i.e., it has not 'speciated'). Thus, if the rate of evolution is slow and only a short time has elapsed in the branch, few events will have occurred, yielding a short estimated branch length. This can distort the estimates of the difference between the two rate parameters if the tips tend to be associated predominantly with one or the other of the background states. The exponential model on the other hand in principle accounts for this bias. The maximum likelihood estimate of $\lambda$ (instantaneous rate of branching) given only a single branch leading to a tip is infinity reflecting the fact that no branching (speciation) has occurred. How best in practice to account for these potential biases requires much more work.

By applying either the exponential or gamma formula to all of the branches and substituting $\lambda_1$ or $\lambda_2$ (or $\alpha_1, \beta_1$, $\alpha_2, \beta_2$) depending upon the state of the background character, two different rates can be estimated and tested to determine whether they differ. The rates are chosen to maximise the likelihood of observing the particular set of branch lengths given the phylogeny and the background character states.

We wish to estimate the rates in a way that does not depend upon the reconstructions of the background
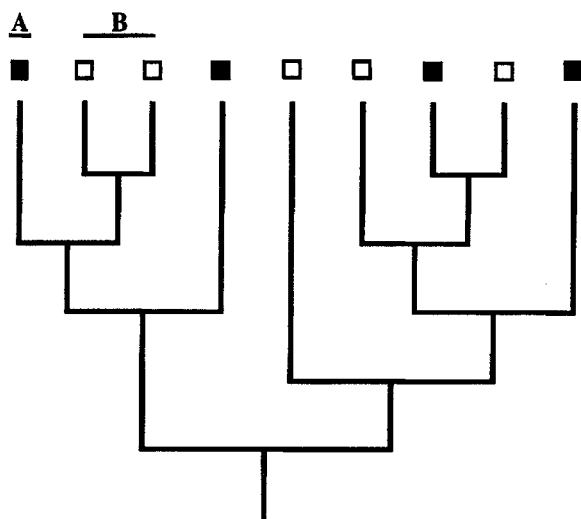
character, which are either not known or only known with error. To avoid this problem, define, following Pagel (1994), a markov-transition model characterising at each node of the phylogeny the likelihood that the background character is in state 0 or 1. Also define the rate of evolution model and now simultaneously find by maximum likelihood the values of $\lambda_1$ and $\lambda_2$ or $\{\alpha_1,\beta_1,\ \alpha_2,\beta_2\}$ and the parameters of the markov model that maximise the likelihood of observing the distribution of branch lengths on the phylogeny, across the background probabilities of the character being in state 1 versus state 0 at each node. Thus, we maximise $L = f(D|Tr, M)$, where $M$ is now some function of the markov-transition model and of the model of rates of evolution. The result of this exercise is estimates of the two rates obtained over all possible values at the ancestral nodes, weighted by the likelihood that a branch was in a particular state.

The hypothesis that speciation rates (or rates of evolution) differ is tested by comparing in the likelihood ratio test the likelihood of the two parameter rates model to a model in which only a single rate parameter is fitted to the data,

$$LR = -2\log_e\left[\frac{L(\lambda_1 = \lambda_2)}{L(\lambda \neq \lambda_2)}\right],$$

where here the formula is illustrated for the exponential model. If LR is significant there is evidence for differential rates of evolution.

## Mutualism in fungi

We have applied this model called Rates to investigate rates of 25S ribosomal DNA evolution in the Omphalina group of fungi (Lutzoni and Pagel 1997). The Omphalina evidence possibly four independent transitions to a mutualist lifestyle (species in italics). Based upon the phylogeny in Fig. 12, the mutualists show apparently increased rates of evolution (longer branches) compared to non-mutualistic sister species (mutualistic species in italics). To test statistically whether the rate of evolution is increased, we applied the Rates model to these data. The raw data consisted of the lifestyle trait for each species, and the branch lengths of the phylogeny. Rates then fits the one and two
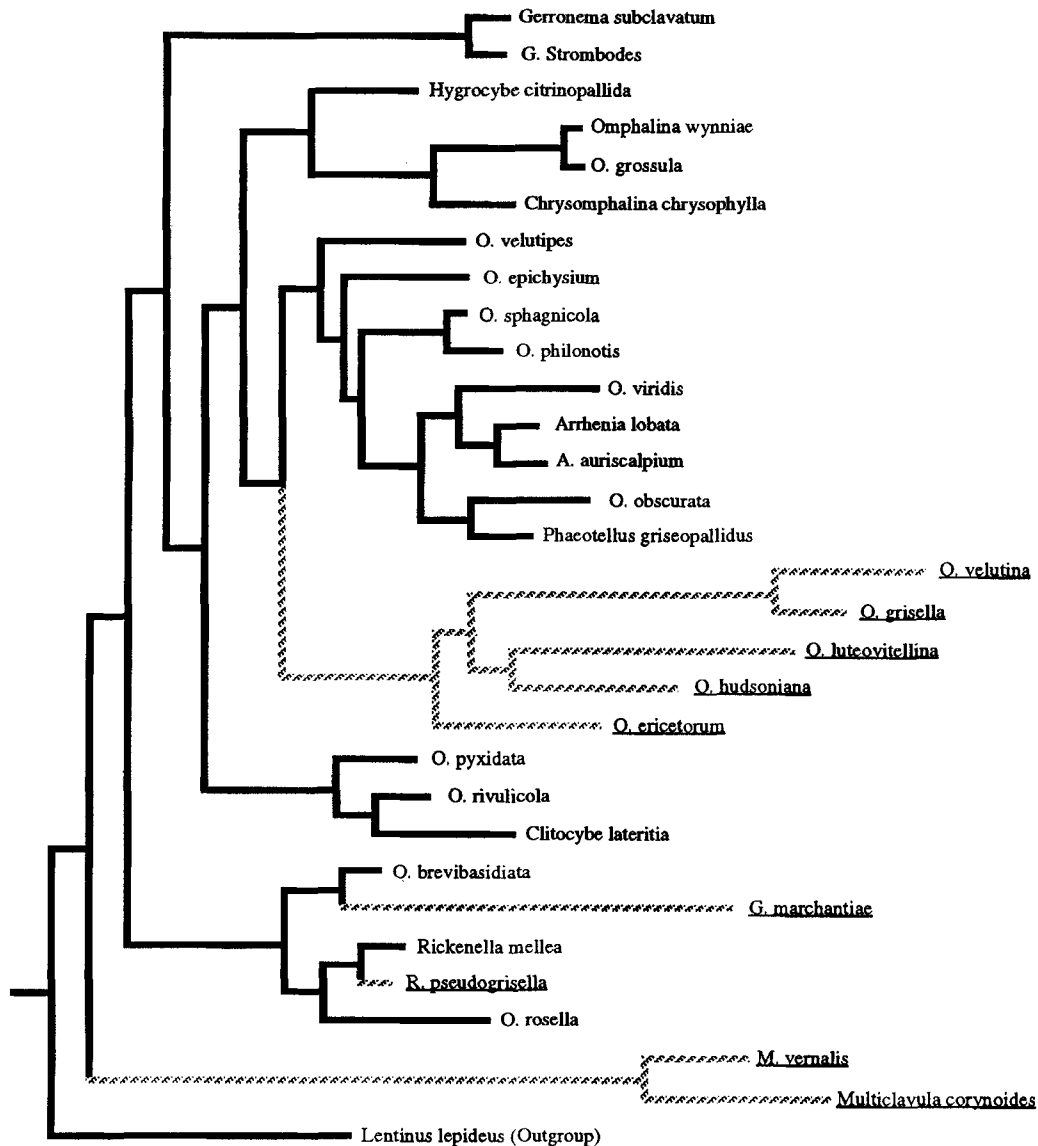


*Fig. 12.* A phylogeny of the Omphalina mushrooms derived from 25S nrDNA sequences (after Lutzoni and Pagel 1997). Underlined species form mutualistic associations with an algal species, the remaining species do not. Lutzoni and Pagel (1997) test the hypothesis and find that rates of evolution in the mutualists are higher than those in non-mutualists. See text.

Table IV. *Rates of Evolution in Mutualist and Non-mutualist fungi. The parameter* λ *here is the rate of branching derived from the exponential or constant probability of branching model. The overall* λ *obtained by ignoring the background state of mutualist/non-mutualist is denoted by* λ. *Separate values of* λ *for the two states are as indicated. The higher value for* λ₁, *corresponding to non-mutualists, indicates a higher rate of branching in these species. This translates into less evolutionary change per branch.*

| Model | $\lambda$ | $\lambda_1$ | $\lambda_2$ | LR | p-value |
|-------|------|------|------|-----|---------|
| Rate  | 0.73 | 1.27 | 0.12 | 5.9 | < 0.02  |

parameter models to the data, allowing the lifestyle character to take all possible values at each node (i.e., the model does not depend upon the parsimony reconstructions shown in the diagram).

For these data the LR statistic is significant, indicating substantially higher rates of evolution among mutualists (Table IV). The model fitted to these data characterises the instantaneous rate of speciation per unit branch length. Longer branches by definition 'speciate' at a lower rate because they take longer (or evolve more) prior to reaching a node, or a species. The phylogeny of Fig. 12 has branch lengths in units of expected genetic divergence. Longer branch lengths indicate that the mutualists evolve more per speciation event than non-mutualists. On the assumption that the time from the root to the tips is the same for all species, this result translates into more evolution per unit time for the mutualists.

## Discussion

Biological scientists working in research areas as disparate as genetics, development, conservation, ecology, evolution, behaviour, and disease transmission are united by the fact that organisms evolve according to a hierarchical branching process in which ancestors give rise to descendants, which in turn give rise to yet more descendants. New and more accurate estimates of these hierarchies — phylogenies — are rapidly appearing owing to breakthroughs in molecular genetic techniques. Their use can give insights into some of the most fundamental questions of biology spanning issues from conservation policy, to how basic animal body plans are formed, to evolutionary theory: for example, how do genes reveal the bio-geographic history of a population? What are the genetic innovations that preceded gross changes in body plan and developmental systems such as those differentiating invertebrate and vertebrate life forms? Do the same genes evolve at different rates in different organisms or lineages?

I have described a maximum likelihood approach to estimating evolutionary processes that can be applied to these new phylogenies. The approach allows estimates to be obtained of real biological parameters of interest, such as rates of evolution or speciation, trajectories of trait evolution against time or genetic divergence and against other variables, rates of transition between character states, and the order of change in two variables. The parameters are estimated within a hypothesis testing framework that allows tests of whether parameter estimates differ significantly from zero or from theoretically important constants. In addition, the likelihood framework

allows one to investigate the evolutionary process itself. By searching for branch length transformations, it is possible to ask questions about whether evolutionary change is smooth and gradual or punctuational.

The approach that I have described has assumed that a phylogeny was to hand and known without error. Inaccuracies in phylogenies will influence the reliability of any estimates obtained from a comparative method. One way around this problem would be to estimate the phylogeny and the comparative relations of interest simultaneously in a combined likelihood framework. Until such solutions become available investigators must be careful always to present their phylogenies and their comparative results as hypotheses that depend upon each other.

The molecular revolution is well underway in phylogenetic studies. New phylogenies will continue to arrive at increasing rates. They will frequently overturn older descriptions, and provide information about the amount of divergence in the branches of the tree. Exploiting these new data to test important questions of evolutionary biology and molecular evolution is the key to turning the revolution into more than just an exercise in classification.

## Acknowledgements

## References

Collinge, J., Sidle, K. C. L., Meads, J., Ironside, J. & Hill, A. F. 1996. Molecular analysis of prion strain variation and the etiology of New Variant CJD.—*Nature 383(6602)*, 685–690.

Cotgreave, P. & Pagel, M. 1997. Predicting and understanding rarity: the comparative approach. In *The Biology of Rarity: the causes and consequences of rare-common differences* (eds. W. Kunin and K. Gaston): Chapman and Hall, 237–261.

Edwards, A. W. F. 1972. Likelihood. The John Hopkins University Press.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach.—*J. Mol. Evol. 17*, 368–376.

Felsenstein, J. 1985. Phylogenies and the comparative method.—*Am. Nat. 125*, 1–15.

Fitch, W. M., Leiter, J. M. E., Li, X. & Palese, P. 1991. Positive Darwinian evolution in human influenza A viruses.—*Proc. Natl. Acad. Sci. USA 88*, 4270–4274.

Friday, A. 1989. Quantitative aspects of the estimation of evolutionary trees.—*Folia Primat. 53*, 221–234.

Garland, T. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts.—*Am. Nat. 140*, 509–519.

Garland, T., Harvey, P. H. & Ives, A. R. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts.—*Syst. Biol. 41*, 18–32.

Gillespie, J. H. 1991. *The causes of molecular evolution*. Oxford: Oxford University Press.

Goldman, N. 1990. Maximum likelihood inference of phylogenetic trees.—*Syst. Zool. 39*, 345–361.

Grafen, A. 1989. The phylogenetic regression.—*Phil. Tran. R. Soc. Lond. B 326*, 119–157.

Grafen, A. & Ridley, M. 1996. Statistical tests for discrete cross-species data.—*J. Theor. Biol. 183*, 255–267.

Hansen, T. & Martins, E. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data.—*Evolution 50*, 1404–1417.

Harvey P. H. & Pagel, M. 1991. *The comparative method in evolutionary biology*. Oxford University Press.

Hillis, D., Moritz, M. & Mable, B. K. 1996. *Molecular Systematics* (2nd Edition). Sunderland, Mass: Sinauer.

Holden, C. & Mace, R. 1997. A phylogenetic analysis of the evolution of lactose digestion in adults.—*Hum. Biol. 69*, 605–628.

Huey, R. B. 1987. Phylogeny, history and the comparative method. In *New directions in ecological physiology* (eds M. E. Feder, A. F. Bennett, W. Burgren, & R. B. Huey). Cambridge University Press.

Huey, R. B. & Bennett, A. F. 1987. Phylogenetic studies of co-adaptation: preferred temperatures versus optimal performance temperatures of lizards.—*Evolution 41*, 1098–1115.

Jermiin, L. S., Graur, D., Lowe, R. M. & Crozier, R. 1994. Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome *b* genes.—*J. Mol. Evol. 39*, 160–173.

Johnston, J. 1963. *Econometric Methods*. New York: McGraw Hill.

Kimura, M. 1968. Evolutionary rate at the molecular level.—*Nature 217*, 624–626.

Krakauer, D. C., Pagel, M., Southwood, T. R. E. & Zanotto, P. M. de A. 1996. Phylogenesis of prion protein.—*Nature 380*, 675.

Krakauer, D. C. & Pagel, M. 1996. The genealogy of a toxic protein and the case of BSE.—*Soc. Biol. Hum. Aff. 61*, 45–54.

Krakauer, D. C., Zanotto, P. de A. & Pagel, M. 1998. Prion's progress: patterns and rates of molecular evolution in relation to spongiform disease. *J. Mol. Evol.*, in press.

Lutzoni, F. & Pagel, M. 1997. Accelerated molecular evolution as a consequence of transitions to mutualism.—*Proc. Natl. Acad. Sci. USA 94*, 11422–11427.

Mace, R., & Pagel, M. 1997. Tips, branches, and nodes: seeking adaptation through comparative studies. In *Human Nature: a critical reader* (ed. L. Betzig): 297–310. New York: Oxford University Press.

Nee, S., Holmes, E. C., Rambaut, A. & Harvey, P. H. 1996. Inferring population history from molecular phylogenies. In *New uses for new phylogenies* (eds P. H. Harvey, A. Leigh-Brown, J. Maynard Smith & S. Nee). Oxford University Press.

Pagel, M. 1992. A method for the analysis of comparative data.—*J. Theor. Biol. 156*, 431–442.

Pagel, M. 1993. Seeking the evolutionary regression coefficient: an analysis of what comparative methods measure.—*J. Theor. Biol. 164*, 191–205.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters.—*Proc. R. Soc. Lond. B 255*, 37–45.

Pagel, M. & Harvey, P. H. 1989a. Comparative methods for investigating adaptation depend upon evolutionary models.—*Folia Primat. 53*, 203–220.

Pagel, M. & Harvey, P. H. 1989b. Taxonomic differences in the scaling of brain on body weight among mammals.—*Science 244*, 1589–1593.

Price, T. & Birch, G. L. 1996. Repeated evolution of sexual color dimorphism in passerine birds.—*Auk 113*, 842–848.

Purvis, A. 1995 A composite estimate of primate phylogeny.—*Phil. Trans. R. Soc. Lond. B 348*, 405–421.

Rao, C. R. & Kleffe, J. 1988. *Estimation of variance components and applications*. North Holland: Amsterdam.

Read, A. & Nee, S. 1995. Inference from binary comparative data.—*J. Theor. Biol. 173*: 99–108.

Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R. & Wüthrich, K. 1996. NMR structure of the mouse prion protein domain PrP(121–231).—*Nature 382*, 180–182.

Schluter, D. 1995. Uncertainty in ancient phylogenies.—*Nature 377*, 108–109.

Sueoka, N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition.—*Proc. Natl. Acad. Sci. USA 48*, 582–592.

Sueoka, N. 1988. Directional mutation pressure and neutral molecular evolution.—*Proc. Natl. Acad. Sci. USA 85*, 2653–2657.

Vilá, C. *et al.* 1997. Multiple and ancient origins of the domestic dog.—*Science 276*, 1687.

Ward, D. & Seely, M. K. 1996. Adaptation and constraint in the evolution of the physiology and behaviour of the Namib Desert Tenebrionid beetle genus *Onymacris*.—*Evolution 50*, 1231–1240.

Weisberg, S. 1985. *Applied linear regression*. 2nd edition. Wiley.