

# Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters

MARK PAGEL

*School of Mathematical Sciences, Queen Mary and Westfield College, University of London, London E1 4NS, U.K.*

## SUMMARY

I present a new statistical method for analysing the relationship between two discrete characters that are measured across a group of hierarchically evolved species or populations. The method assesses whether a pattern of association across the group is evidence for correlated evolutionary change in the two characters. The method takes into account information on the lengths of the branches of phylogenetic trees, develops estimates of the rates of change of the discrete characters, and tests the hypothesis of correlated evolution without relying upon reconstructions of the ancestral character states. A likelihood ratio test statistic is used to discriminate between two models that are fitted to the data: one allowing only for independent evolution of the two characters, the other allowing for correlated evolution. Tests of specific directional hypotheses can also be made. The method is illustrated with an application to the Hominoidea.

## 1. INTRODUCTION

Comparative biologists frequently search for evidence of the correlated evolution of two or more traits across a group of species. The simplest comparative test correlates across species the values of the two traits. However, to treat species as the units of analysis in a comparative study assumes that the traits under investigation evolved independently in each of them. A parallel set of arguments can be made for populations within a species.

Ridley (1983) and Maddison (1990) have developed tests for discrete characters that avoid the problem of counting species. Ridley's test includes only instances in which one or both characters is inferred to change along the branch of a phylogeny, ignoring branches in which changes do not occur. A  $\chi^2$  statistic is used to assess whether the changes in the two variables are associated. Maddison's (1990) method makes use of all of the branches in the phylogeny. The method assesses whether an inferred pattern of changes in two variables along the branches of the phylogeny is unlikely given all possible ways the set of changes could have evolved on the tree. Harvey & Pagel (1991) provide a fuller account of these methods, including examples of their application to comparative data.

Ridley's and Maddison's methods provide innovative solutions to some of the key problems in the analysis of comparative data. However, these tests have several features that limit their utility for comparative hypothesis testing, especially as well-evaluated phylogenies with information on branch lengths become increasingly available. Neither method, for example, is based upon a stochastic model

of evolution. This means that both methods treat all types of evolutionary change as equally likely, and that neither is able to incorporate information about the lengths of the branches of the phylogenetic tree. Maddison's method is designed to detect the order of changes in two binary characters; however, its power to do this is greatly reduced by branches in which a simultaneous change occurs in both characters.

A characteristic of all existing comparative methods, whether for categorical (e.g. Ridley 1983; Maddison 1990) or continuous data (e.g. Felsenstein 1985; Grafen 1989; Pagel & Harvey 1989; Gittleman & Kot 1990; Harvey & Pagel 1991; Pagel 1992), is that they rely upon a single set of inferred character states at the internal nodes of the phylogeny. Methods for categorical traits typically employ parsimony and out-group methods, whereas methods for continuous traits usually rely upon maximum likelihood inference to assign to each node of a phylogeny a value of each character. The implication for comparative hypothesis testing is that all resulting statistical tests are dependent upon these inferences. The correct representation of a character at each of the internal nodes of a phylogeny is a probability distribution of character states. The probability distribution represents our implicit belief that all states of the character are possible at each node, even if some are more likely or more parsimonious than others.

A new generation of comparative methods that is free of the dependency on a single set of inferred values at the internal nodes is possible. By representing the character states as probability distributions, all possible transitions among states along the branches of the phylogeny can be studied, weighted by their prob-



ability of occurrence. A related point that I discuss informally in a later section is that for all the same reasons that a single set of inferred character states at the internal nodes is undesirable, so is reliance upon a single phylogeny. It is possible, in principle, to develop methods that study simultaneously all possible character state transitions integrated over all possible phylogenies, weighted by their probability of occurrence. This would remove any dependence of the conclusions upon either a single phylogeny or set of character states. However, to simplify presentation, here I describe the first step in this process; a method for the analysis of correlated evolutionary change in categorical characters on a given phylogeny.

## 2. A CONTINUOUS-TIME MARKOV MODEL

The method I shall develop in the following sections uses a continuous-time Markov model to characterize evolutionary changes along each of the branches of a phylogenetic tree. Jukes & Cantor (1969), Kimura (1980), Felsenstein (1981), and Hasegawa *et al.* (1985) and Goldman (1993) have employed similar models to study base substitution in nucleic acid data, and thus to develop estimates of phylogenies of species. The model I develop is not addressed to the problem of building phylogenies, but rather is designed to estimate simultaneous transition rates in pairs of binary characters arrayed on a phylogeny, and to use those rates to test ideas about correlated evolution of the two characters. The method incorporates information on branch lengths, can in principle detect the order of changes even in branches showing simultaneous change, and makes use of all possible character states at each ancestral node. This means that the eventual hypothesis test is independent of any particular method for assigning ancestral values, and of any particular assignment of those values.

The method tests the hypothesis of correlated evolution by comparing the fit of two different models to the observed data set. First, the method fits a model to the data in which the two characters are treated as evolving independently. The goodness of fit of this model to the data is then compared to that of a more complicated model in which the characters evolve in a correlated fashion. The model of correlated change is justified as a representation of the data if it fits the data significantly better than the model of independent change. This result also supports the inference of correlated evolution. Significance is assessed by means of a likelihood ratio statistic. I describe in the following sections the derivation of the models of independent and correlated change, and outline a general procedure for testing the hypothesis of correlated evolution.

## 3. INDEPENDENT CHANGES IN TWO VARIABLES

Consider a variable that can take two states, 0 and 1. We want to write equations to characterize the probability that in a branch of a phylogeny beginning in state 0, the variable will either remain the same or change to state 1, over over amount of time  $t$  with

probabilities  $P_{00}(t)$  and  $P_{01}(t)$ . We shall also write the equations for  $P_{11}(t)$  and  $P_{10}(t)$ , corresponding, respectively, to a branch beginning and ending in state 1, or changing from state 1 to state 0 over time  $t$ . A parallel set of equations will be used to characterize changes in a second variable. If the two variables change independently of each other, then their joint probabilities of change in any given branch of the tree are given by the product of their separate probabilities of change.

Assume that a variable evolves according to a Markov process such that the probability of change is independent in each branch of the phylogenetic tree, and that the probability of changing from one state to the next depends only upon the state at the beginning of the branch, and not on events previous to that. The probability that the variable changes from state  $i$  to state  $j$  over some time interval  $t+dt$  is given by

$$P_{ij}(t+dt) = P_{ii}(t) q_{ij} dt + P_{ij}(t) (1 - q_{ji}) dt, \quad (1)$$

where  $t$  is an arbitrary time period,  $dt$  is a short interval of time, and the  $q_{ij}$  are transition rate parameters denoting, respectively, the rate of change from state  $i$  to state  $j$  over an infinitesimally short time interval. The  $1 - q_{ji}$  term represents the probability of staying in state  $j$ ; that is, the probability of not changing from state  $j$  to state  $i$ . Equation (1), then, merely states that a character can change from state  $i$  to state  $j$  either by remaining unchanged for a period  $t$  followed by a transition to state  $j$ , or by changing to state  $j$  during time period  $t$ , then remaining in state  $j$  over period  $dt$ . The set of equations describing the four possible probabilities of interest can be represented in matrix form as the familiar Chapman-Kolmogorov equation

$$\mathbf{P}(t+dt) = \mathbf{P}(t) (\mathbf{I} + \mathbf{Q} dt), \quad (2)$$

$$\begin{bmatrix} 1 - P_{01}(t+dt) & P_{01}(t+dt) \\ P_{10}(t+dt) & 1 - P_{10}(t+dt) \end{bmatrix} = \begin{bmatrix} P_{00}(t) & P_{01}(t) \\ P_{10}(t) & P_{11}(t) \end{bmatrix} \begin{bmatrix} (1 - q_{01}) dt & q_{01} dt \\ q_{10} dt & (1 - q_{10}) dt \end{bmatrix},$$

where  $\mathbf{P}(t)$  is a  $2 \times 2$  matrix of  $P_{ii}$ 's  $= 1 - P_{ij}$ 's, and  $P_{ij}$ 's (where  $i$  and  $j$  refer to the two possible character states),  $\mathbf{I}$  is the identity matrix, and the matrix  $\mathbf{Q}$  is a  $2 \times 2$  matrix of the transition rates.

Equation (2) is the mathematical representation of the Markov process. To solve for the  $\mathbf{P}(t)$  in terms of the rate parameters in  $\mathbf{Q}$  first find

$$\frac{d\mathbf{P}(t)}{dt} = \frac{\mathbf{P}(t+dt) - \mathbf{P}(t)}{dt} = \mathbf{P}(t) \mathbf{Q}, \quad (3)$$

from which

$$\mathbf{P}(t) = \exp[\mathbf{Q}t] + c, \quad (4)$$

where  $c$  is the constant of integration. Because  $\mathbf{P}(0) = \mathbf{I}$ ,  $c = 0$  and therefore  $\mathbf{P}(t) = \exp[\mathbf{Q}t]$ . To complete the matrix calculation  $\exp[\mathbf{Q}t]$  it is necessary to take advantage of the result that  $\exp[\mathbf{Q}t]$  is equal to  $\mathbf{C} \exp[\mathbf{D}t] \mathbf{C}^{-1}$  where  $\mathbf{C}$  contains the eigenvectors of  $\mathbf{Q}$  and  $\mathbf{D}$  is a diagonal matrix containing the eigenvalues of  $\mathbf{Q}$ . The value of  $t$  is a constant for any given  $\mathbf{P}(t)$  and



is taken from the particular branch to which that  $P(t)$  applies. Thus,

$$P(t) = C \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix} C^{-1},$$

where the  $\lambda_i$  are the eigenvalues of  $Q$ . The Chapman–Kolmogorov equations therefore allow for a solution of the  $P(t)$  solely in terms of the rate parameters in  $Q$  and the length of the branch,  $t$ . For example,  $P_{01}(t)$  will be a function of the rate of transition from 0 to 1, balanced over time  $t$  by the rate of transition from 1 to 0: the larger  $q_{01}$  is relative to  $q_{10}$ , the greater the probability that at the end of time  $t$  a character beginning in state 0 will be in state 1.

#### 4. FINDING THE LIKELIHOOD FOR TWO VARIABLES

Given the assumption of independence between  $X$  and  $Y$  the likelihood of the data given the model and the phylogenetic tree will be a product of the likelihoods for each character. To find the likelihood of each character, first consider one particular realization of character states for the  $X$  variable at the tips of the tree and of assignments of ancestral states at the interior nodes. The likelihood of this realization is given by the product over all of the branches of the tree of the appropriate probability for each branch as derived from the Markov process. This product, in turn, is summed over all possible assignments of character states at the interior nodes, including the root, although the values observed at the tips are held constant. Repeat this procedure for the  $Y$  variable, then find the product of the two likelihoods.

To illustrate these points, consider the hypothetical phylogeny in figure 1. Following Felsenstein (1983), let  $P(u, v, t)$  represent the probability that a branch beginning in state  $v$  ends in state  $u$ , after time period  $t$ . For a given variable,  $P(u, v, t)$  will take only two forms for branches leading to the tips of the tree because  $u$  will be a constant. For all other branches both  $u$  and  $v$  can vary from 0 to 1 so four possibilities arise in each variable. Then the likelihood for this tree will be given by,

$$L(I) = \prod_{s9=0}^{X \text{ and } Y} \sum_{s8=0}^1 \sum_{s7=0}^1 \sum_{s6=0}^1 P(s8, s9, t8) \times P(s3, s8, t3) P(s7, s8, t7) P(s5, s7, t5) \times P(s4, s7, t4) P(s6, s9, t6) \times P(s1, s6, t1) P(s2, s6, t2), \quad (5)$$

where the summations are from 0 to 1 at each node. This likelihood is a function of four parameters, two for the  $X$  variable and two for the  $Y$  variable. One then searches the likelihood surface to find the maximum likelihood solution to  $L(I)$ . This is the set of four parameters that makes  $L(I)$  largest.

Although the likelihood in equation (5) is found over all possible assignments of the ancestral character states, this is not strictly necessary. One may have prior information or beliefs about the ancestral values at the root of the tree. In such instances, the root of the tree

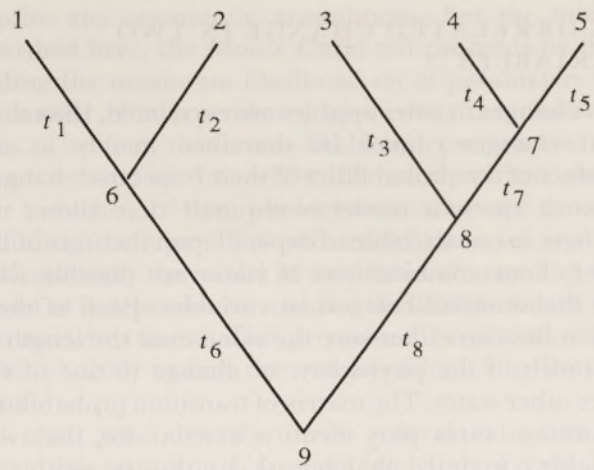


Figure 1. A phylogeny of five species. The  $t_i$  refer to the lengths of the branches.

can either be fixed (e.g. fix  $s9$  in the example of figure 1), or the likelihoods corresponding to the four different possible states at the root can be weighted by one's priors. However, either of these actions is a hypothesis and one should be prepared to defend the values attached to the priors, or the value of a fixed root, against the charge that the eventual conclusions may depend upon them.

A difficulty that arises in connection with computing the likelihood is that the number of different possible assignments of character states to the interior nodes of the trees is proportional to  $4^{n-1}$  where  $n$  is the number of tips of the tree and four is all possible combinations of two dichotomous variables. This produces a very large number of sets of characters states over which to find the likelihood if equation (5) is followed. For a tree with only 11 tips there are already over one million possible distinct sets. Fortunately a substantial savings can be obtained by recognizing that some of the terms in the likelihood products do not change from one summation to the next.

Felsenstein (1993) encountered this problem in connection with estimating tree topologies and developed his 'pruning' algorithm (1983; pp. 251–252) to reduce the computational effort. Felsenstein's pruning algorithm can be applied in this instance (I refer the reader to his article for further details) by recognizing that the four nested summations in (5) can be distributed along the product terms:

$$L(I) = \prod_{s9=0}^{X \text{ and } Y} \sum_{s8=0}^1 \left[ \left( \sum_{s7=0}^1 P(s8, s9, t8) P(s3, s8, t3) \times \sum_{s6=0}^1 P(s7, s8, t7) P(s5, s7, t5) P(s4, s7, t4) \right) \times \sum_{s6=0}^1 P(s6, s9, t6) P(s1, s6, t1) \times P(s2, s6, t2) \right]. \quad (6)$$

No expression of the form  $P(u, v, t)$  can appear to the left of a summation over  $u$ . Applying the pruning algorithm makes the computing task linear rather than exponential in  $n$ : that is, proportional to  $4(n-1)$ .



## 5. CORRELATED CHANGE IN TWO VARIABLES

If changes in two variables are correlated, then their joint changes cannot be described merely as the product of the probabilities of their respective changes. In such cases, a model is required that allows the changes in one variable to depend upon the state of the other. Four combinations of states are possible with two dichotomous categorical variables. Each of these four states can either stay the same over the length of a branch of the phylogeny, or change to one of the three other states. The matrix of transition probabilities (transition rates plus identity matrix) for the two-variable correlated change case can thus be written as

$$I + Q_D = \begin{matrix} X, Y \\ \begin{matrix} 0, 0 \\ 0, 1 \\ 1, 0 \\ 1, 1 \end{matrix} \end{matrix} \begin{bmatrix} 0, 0 & 0, 1 & 1, 0 & 1, 1 \\ 1 - & q_{12} & q_{13} & 0 \\ (q_{12} + q_{13}) & & & \\ q_{21} & 1 - & 0 & q_{24} \\ & (q_{21} + q_{24}) & & \\ q_{31} & 0 & 1 - & q_{34} \\ & & (q_{31} + q_{34}) & \\ 0 & q_{42} & q_{43} & 1 - \\ & & & (q_{42} + q_{43}) \end{bmatrix},$$

where the subscript 'D' on the matrix  $Q_D$  denotes that this matrix contains the transition rate parameters for the case of dependent or correlated change in two characters. The subscripts on the rate parameters refer to the beginning and ending character states for each particular transition, where 1 = 0, 0; 2 = 0, 1; 3 = 1, 0; 4 = 1, 1. Thus, for example,  $q_{12}$  denotes a transition from state  $X = 0, Y = 0$  to state  $X = 0, Y = 1$ .

The entries in this matrix describe with a different parameter each of the possible combinations of beginning and ending states of a branch. If the changes in the two variables are independent, these rate parameters will not differ from the product of the two corresponding rate parameters (one for  $X$  and one for  $Y$ ) from the model of independence (but see §6). If, however, the two variables do change in a correlated fashion, then the rate parameters for the model of correlated change will differ from those predicted from a model of independence.

The values of the rate parameters describing dual transitions – a change in both  $X$  and  $Y$  – are set to zero in  $Q_D$ . This is because the model is designed to detect evolutionary relationships between two variables in which the state of one variable affects the probability of a change in the other. Setting the elements that describe dual transitions over time  $dt$  to zero does not, however, rule out dual transitions over longer time period  $t$ . A dual transition from 0, 0 to 1, 1 observed over time  $t$  can proceed first via a transition from 0, 0 to 1, 0, then from 1, 0 to 1, 1. Alternatively, it could go first from 0, 0 to 0, 1, then from 0, 1 to 1, 1. The model is designed to be able to distinguish these two alternatives (see §6).

Nothing in the mathematics of Markov models requires that the dual transition rate parameters be set to zero. However, were the parameters describing dual

transitions over time  $dt$  not set to zero, the model would treat branches in which a dual transition occurred as a qualitatively different class of phenomena from the others. A dual transition over time  $dt$  means that the two different variables have changed simultaneously, that is, at the exact same instant of time  $dt$ . Such a view is incompatible with the belief that the state of one variable affects the probability of change in the other. Moreover, were the rate parameters representing dual transitions estimated, the model would (most likely mistakenly) attribute to those parameters all of the instances in the phylogeny in which over longer time  $t$  both variables change.

The constraints that each row sums to zero, and that dual transitions have a probability of zero over time  $dt$ , means that the matrix  $Q_D$  is specified completely by eight parameters. None the less, the rate matrix contains all of the parameters of interest for testing specific comparative hypotheses. The idea that, for example, the variable  $Y$  is more likely to change from 0 to 1 when the background variable  $X$  is in state 1 compared to when it is in state 0, is a statement that parameters  $q_{12}$  and  $q_{34}$  are different. The belief that warning colours in insects are more likely to evolve in gregariously living than in solitary insects (e.g. Sillén-Tullberg 1988) is an example of such a hypothesis.

We seek equations to describe the probability of each kind of change represented in  $Q_D$ . There will be sixteen different  $P(t)$ 's, four describing stasis, and twelve describing change in one or both of the variables; one corresponding to each of the conditions represented by the transition rates in  $Q_D$ . As before the  $P(t)$ s are found according to equations (2), (3) and (4), except now the matrices  $P(t+dt)$ ,  $P(t)$ , and  $Q_D$  are  $4 \times 4$  rather than  $2 \times 2$ , and follow the ordering of  $Q_D$ .

Given an equation for each kind of change in  $X$  and  $Y$ , the likelihood of a particular realization of ancestral states on the tree can be found as the product over all branches of the relevant probabilities, including the final branches leading to the tips (the values of which, as before, are fixed). This likelihood is then summed over all possible reconstructions of the ancestral nodes. Call this likelihood  $L(D)$  to denote that it represents the model of dependent or correlated change. As with  $L(I)$  the root of the tree can be fixed or weighted if one has prior beliefs about it.

The set of the eight transition rate parameters that makes  $L(D)$  largest is found by maximum likelihood. Thus,  $L(D)$  takes a similar form to  $L(I)$  for figure 1 except now the likelihood expression does not contain a term for the product over  $X$  and  $Y$ , and the  $P(u, v, t)$  refer to the joint changes in  $X$  and  $Y$  along each branch. Accordingly, each summation is now over the four possible states (numbered as in  $Q_D$ ) that can occur at the beginning or ending of each branch.

$$L(D) = \sum_{s9=1}^4 \sum_{s8=1}^4 \sum_{s7=1}^4 \sum_{s6=1}^4 P(s8, s9, t8) \\ \times P(s3, s8, t3) P(s7, s8, t7) P(s5, s7, t5) \\ \times P(s4, s7, t4) P(s6, s9, t6) \\ \times P(s1, s6, t1) P(s2, s6, t2). \quad (7)$$



As with the expression  $L(I)$ , the summations can be distributed throughout this expression for  $L(D)$  to achieve substantial reductions in the computational effort:

$$L(D) = \sum_{s9=1}^4 \left[ \left( \sum_{s8=1}^4 P(s8, s9, t8) P(s3, s8, t3) \right. \right. \\ \times \sum_{s7=1}^4 P(s7, s8, t7) P(s5, s7, t5) \\ \times P(s4, s7, t4) \left. \right) \\ \times \sum_{s6=1}^4 P(s6, s9, t6) P(s1, s6, t1) \\ \times P(s2, s6, t2) \left. \right]. \quad (8)$$

## 6. STATISTICAL TESTING

The likelihood ratio (LR) statistic, defined here as

$$LR = -2 \log_e [L(I)/L(D)], \quad (9)$$

provides a general procedure for comparing the two likelihoods. If when the null hypothesis is true,  $L(D)$  can be expressed as a particular instance or special case of  $L(I)$ , the LR statistic will be asymptotically distributed as a  $\chi^2$ . The degrees of freedom (d.f.) for the asymptotic  $\chi^2$  test are given by the difference in the number of parameters required to estimate each likelihood, or more precisely, the number of restrictions that must be placed on the parameters of  $L(D)$  to derive  $L(I)$ . Precisely what constitutes sufficient data to ensure that the LR statistic is distributed as a  $\chi^2$  will depend upon the values of the rate parameters. The smaller are the  $q_{ij}$  the more data will be needed.

For cases in which  $L(D)$  cannot be expressed as a special case of  $L(I)$  the likelihood ratio may not be distributed as a  $\chi^2$  and a Monte Carlo testing procedure is required. Goldman (1993) provides a thorough review of the literature on hypothesis testing with likelihood ratio tests, and outlines a general Monte Carlo procedure based on Cox (1962), Lindsay (1974*a, b*) and Hall & Wilson (1991). The Monte Carlo procedure can be used to derive the null hypothesis distribution for any likelihood ratio test, and does not

require any asymptotic assumptions. For the model described here, the Monte Carlo test proceeds by first finding the maximum likelihood set of parameters for the model of independence applied to the observed data. The four maximum likelihood parameters are then fixed and used to evolve the two characters along the branches of the phylogenetic tree up to and including the tips. This null model data set is then analysed with the model of correlated change and the model of independence, and the likelihood ratio of the two models is found. This procedure is repeated many times ( $\sim 100$ ) to find the null hypothesis distribution of LR.

I shall now describe three general categories of hypothesis test using  $L(I)$  and  $L(D)$ . Two of these, the omnibus test, and tests of alternative models, require Monte Carlo tests of the sort Goldman (1993) describes. Asymptotic  $\chi^2$  tests can be used to assess contingent and directional hypotheses, although the Monte Carlo procedures can also be used. The tests are summarized in table 1.

The omnibus test compares whether the eight-parameter model describes the observed data significantly better than the four-parameter model, and is accordingly a test for correlated evolution of  $Y$  and  $X$ . The omnibus likelihood  $L(D)$  cannot in general be expressed as a special case of  $L(I)$ . Under the null hypothesis, the separate rate parameters in  $Q_D$  will not necessarily be equal to the product of the appropriate rate parameters for  $Y$  and  $X$  in the model of independence. This is because the rate parameters in  $Q_D$  representing dual transitions have been set to zero. Thus, the likelihood ratio for the omnibus test will not necessarily be asymptotically distributed as  $\chi^2$  with d.f. = 8–4. However, the significance level of the omnibus LR statistic can be ascertained to any desired level of precision by Monte Carlo simulation.

One may wish to test specific hypotheses about the nature of the correlated changes in  $X$  and  $Y$ . These can include tests of whether changes in one variable are contingent upon (more or less likely given) the state of the other, and tests of the temporal ordering and directions of change. In the warning colours example given above, one is interested in whether warning

Table 1. Categories of hypothesis test using  $L(I)$  and  $L(D)$

(The tests shown in the table do not exhaust the range of possible tests but represents ones likely to be of interest. Tests with more than 1 d.f. are also possible. The contingent change tests, temporal order/relative rate tests, and branch length transformation tests will have asymptotic  $\chi^2$  distributions but may also be tested by Monte Carlo simulation. The omnibus test, and tests of alternative models will not in general be distributed as  $\chi^2$ . The test of branch length transformation can be applied to any likelihood, not just to the full model. Subscripts to  $L(D)$  denote the number of parameters.)

test	hypothesis	description	statistic
omnibus $\chi^2$ (4 d.f.)	$L(I) < L(D)$	correlated evolution	$L(I)$ versus $L(D)$
contingent change (1 d.f.)	$q_{12} \neq q_{34}$	change in $Y$ depends upon state of $X$	$L(D_7)$ versus $L(D_8)$
	$q_{21} \neq q_{43}$	change in $Y$ depends upon state of $X$	$L(D_7)$ versus $L(D_8)$
	$q_{13} \neq q_{24}$	change in $X$ depends upon state of $Y$	$L(D_7)$ versus $L(D_8)$
	$q_{31} \neq q_{42}$	change in $X$ depends upon state of $Y$	$L(D_7)$ versus $L(D_8)$
	$q_{12} \neq q_{13}$	order of acquisition of $X$ versus $Y$	$L(D_7)$ versus $L(D_8)$
temporal order or relative rate (1 d.f.)	$q_{42} \neq q_{43}$	order of loss of $X$ versus $Y$	$L(D_7)$ versus $L(D_8)$
alternative models	$q_{ij} = 0$	$q_{ij}$ transitions excluded	$L(D_7)$ versus $L(D_8)$
branch transformation	$\kappa \neq 1$	transformation of branch lengths	$L(D_n)$ versus $L(D_{n+1})$



colours are more likely to evolve when individuals are gregarious than solitary. Taking warning colours as the dependent or  $Y$  variable, this is a hypothesis that two parameters,  $q_{12}$  and  $q_{34}$ , differ, against the null hypothesis that they are the same. If  $q_{12}$  and  $q_{34}$  differ, then this is evidence that changes in  $Y$  are more likely given a particular state of  $X$ . The hypothesis is tested by forcing the relevant parameters in  $\mathbf{Q}_D$  to take the same value: here by setting  $q_{12} = q_{34}$ . This creates a seven parameter model that is a special case of the full eight-parameter model:  $q_{34} = kq_{12}$  where  $k = 1$  under the null. The seven-parameter model is then fitted to the data by maximum likelihood, and the two models are then compared by means of a likelihood ratio test. This test will be asymptotically distributed as a  $\chi^2$  with 1 d.f.

One may not be able to establish with either Ridley's or Maddison's tests which of the two variables changes first, when evidence is found for correlated evolution. This is because the inferred values of the ancestral character states will often yield branches of the tree in which a transition in both  $Y$  and  $X$  occurs. The Markov model approach, however, in principle can test for the order of changes even in such instances. To test, for example, whether  $Y$  evolves to state 1 before  $X$  or vice versa requires a comparison of  $q_{12}$  versus  $q_{13}$ . If the rate estimated for  $q_{12}$  exceeds that of  $q_{13}$ , then this is evidence that changes in  $Y$  tend to precede changes in  $X$ . This hypothesis can be tested formally by forcing  $q_{12} = q_{13}$  and comparing the likelihood of the resulting seven parameters model to the full eight parameter model in a likelihood ratio test. Again, this test will be asymptotically distributed as a  $\chi^2$  with 1 d.f.

One may wish to test models of correlated evolution in which certain types of transitions are excluded *a priori*. This is a test of an alternative model. Alternative models are tested by setting to zero the relevant parameter(s) in the matrix  $\mathbf{Q}_D$  prior to finding the maximum likelihood. The likelihood of such a reduced model will have 1 fewer degree of freedom than the full model for each parameter that is set to zero. Alternative models of this sort can be tested in one of two ways. A test with asymptotic  $\chi^2$  properties is to compare the reduced likelihood to the full model likelihood. If the parameter set to zero in the reduced model is significantly different from zero in the full model, the full model will explain the data significantly better than the reduced model. However, tests of reduced models will usually be made when one suspects that a parameter would not contribute a significant amount to the likelihood. If the parameter is not different from zero in the full model, then the likelihood ratio of the reduced and full models will approach 1 ( $LR = 0.0$ ) and is evidence that the reduced model is an adequate representation of the data. However, such a test requires the logically weak position of 'accepting' the hypothesis of no difference between the two models as true.

A better justified test is to compare the reduced likelihood to  $L(I)$ . One can successively remove parameters from the full model  $L(D)$  to create a minimum reduced model that still explains the data significantly better than  $L(I)$ . Unfortunately, for the

same reasons that the omnibus test could not be tested by asymptotic  $\chi^2$ , it will generally be true in tests of reduced models that the reduced  $L(D)$  cannot be expressed as a special case of  $L(I)$ , and Monte Carlo tests of the sort Goldman (1993) reviews are required.

## 7. PHYLOGENIES: RESOLUTION, FOSSILS AND BRANCH LENGTHS

I have assumed throughout that investigators will have to hand a phylogeny of the species to be studied. This is merely a convenience. The solutions for  $L(I)$  and  $L(D)$  and the phylogeny can be estimated simultaneously by maximum likelihood. An even more general solution is possible that removes the dependence of the conclusions upon any one phylogeny. In principle, one can find the maximum likelihood solutions for  $L(I)$  and  $L(D)$  integrated over all possible reconstructions of the phylogeny. This will, of course, be computationally onerous in most cases. However, if the different possible phylogenies are discrete (such as is generally true of 'cladistic' phylogenies) the task may become more feasible. In either case, the set of transition parameters that maximizes the likelihood over all phylogenies and all assignments of the ancestral states in each phylogeny would be sought.

I have also assumed that the phylogenies to be used will be fully resolved or bifurcating phylogenies. It may sometimes be the case, however, that the best phylogeny available is not fully resolved, showing instead nodes from which emanate three or more branches. These nodes can be used with the model described here if they represent a true but not fully resolved picture of the true tree. Including unresolved nodes will very likely reduce the power of the test to discriminate between the models of independence and dependence.

What is the role of fossils in the methods described here? Fossils may provide evidence of a relevant character state in some instances. They can be included at the appropriate spot in the phylogeny and treated as an additional observation, just as one would treat the data at the tips. That is, one would not calculate the likelihood over all possible assignments of the ancestral character states at that node, but rather fix it at the value given by the fossil.

Markov transition models specify the rate parameters in units of time,  $t$ , and the transition rate applies smoothly over the entire time period. However, the relevant times to attach to each branch of the phylogeny may not be simple clock time. If rates of evolution are a function of the number of generations along a branch, then branch lengths could be expressed in numbers of generations. Genetic distances uncorrected for molecular clock assumptions may be the best estimate of the 'opportunity for change' in the sense of indicating the total amount of evolution between pairs of species. Genetic distances can be used directly in place of estimates of  $t$  in units of time. Time, then, in this model of the Markov process is correctly thought of as being measured in units of evolutionary change.

How should investigators proceed when there is no information about the lengths of branches of the



phylogeny? One possibility is simply to make all branches the same length, an implicitly punctuational view of evolution. This is also implicit to any method that ignores branch lengths, such as Ridley's (1983) or Maddison's (1990) method. However, investigators should be cautious about interpreting significant results when equal branch lengths are used because their use may lead to increased Type I error rates (Martins & Garland 1991; Gittleman & Luh 1992; Purvis *et al.* 1994). Another possibility is to assign branch lengths according to an algorithm based upon the branching pattern of nodes in the phylogeny. Assign to each node of the tree a number corresponding to the number of species that can trace their ancestry to it. Then, the branch length between any two nodes (or between a tip and a node) is the difference between their two numbers (Grafen 1989). Although this method for assigning branch lengths seems to have a rough 'face-validity', recent computer simulations (Purvis *et al.* 1994) indicate that it leads to elevated Type I error rates.

Investigators should consider that a choice of branch lengths is a hypothesis about the way evolution is assumed to have proceeded. A hint that the results in any particular investigation are sensitive to the exact branch lengths used can be obtained from observing the effects on the results of using different rules for assigning the branch lengths. It is also possible with the Markov approach to test statistically whether a transformation of the chosen branch lengths can improve the fit of the model to the data. Let

$$P(t) = \exp[Q_D t^\kappa], \quad (10)$$

be the model of dependence described above but with an additional parameter  $\kappa$  fitted by maximum likelihood. If  $\kappa > 1$  then all branches are effectively increased in length, but longer branches are increased more than shorter ones. Conversely, a value of  $\kappa < 1$  reduces the length of longer branches more than that of shorter ones. A very small value of  $\kappa$  is evidence that the branches are all of comparable length with respect to estimating the rate parameters, effectively a punctuational view of evolution.

The model fitted this way has 9 d.f. and its likelihood can be compared to the likelihood of the untransformed model with 8 d.f. The likelihood ratio statistic will be a  $\chi^2$  with one degree of freedom. The parameter  $\kappa$  in this model is functionally equivalent to the parameter  $\rho$  in Grafen's (1989) method and the parameter  $\alpha$  in Gittleman & Kot's (1990) method, both of which are for use with quantitative traits. One should not expect miracles from  $\kappa$  (nor from  $\alpha$  or  $\rho$ ) because it can only perform monotonic transformations of the branch lengths; it cannot, for example, alter a shorter branch to be longer than another that was originally specified to be longer. Accordingly,  $\kappa$  (and  $\alpha$  and  $\rho$ ) should always be tested for significance, and not just used routinely.

A useful test is to compare the likelihood of a model with the optimal  $\kappa$  with one in which  $\kappa$  is set to zero. This provides a test of whether the optimal transformed branch lengths differ significantly from one in which the branches all have unit length: a direct test of

whether punctuational branch lengths provide a reasonable description of the data.

The model assumes that the rates of change are uniform across all branches. If rates of change vary among clades, for example, the rate parameters will be a composite of these varying rates. All comparative tests must confront this issue. Given very large data sets, it would be possible to estimate rate parameters separately in different groups. The probability of change may be zero for some types of species studied. Were a sufficient number of these species added to a smaller group in which changes did occur, the model could find a significant effect. This raises the possibility that the probability of finding a significant result may increase with the number of such clades that are included. This is a problem of defining the selection criteria for a comparative study. Pagel (1994) discusses the logic of how to decide what species to include (or not) in a comparative test.

## 8. A WORKED EXAMPLE

Clutton-Brock & Harvey (1976) observed that in many Old World monkey species in which the females regularly mate with more than one male during their oestrus, the females develop a prominent reddening and swelling of the perineum at or near the time of ovulation. Swellings are not common in species with single male or monogamous breeding systems. Thus, each species can be scored for their mating system (multi-male or not, and whether they have swellings at oestrus). Here I illustrate how the Markov model outlined in this paper can be applied to questions such as this.

I take as my study group the Hominoidea which includes the gibbons, orang-utan, gorillas, chimpanzees, and humans, plus an outgroup of the black and white colobus (*Colobus guereza*). Figure 2 displays a phylogeny of these species derived from the CO II mitochondrial gene (Horai *et al.* 1992; Disotell 1992; Ruvolo *et al.* 1991). The branch lengths are the minimum number of changes required to draw the tree as derived from PAUP (T. Disotell, personal communication). Branch lengths have not been normalized to create a uniform length from the root to all tips. I use this phylogeny rather than estimate it simultaneously with the rate parameters (see §7) to simplify the illustration of the method. Because this is such a small group, I do not take this as a general test of Clutton-Brock & Harvey's (1976) observation.

The scores for mating system and oestrous advertisement are presented for each species in figure 2 (caption) and are taken from the literature. There is a near perfect correspondence between mating system (M) and oestrous advertisement (+), the only exception being Pongo. It is not necessary to make any assumptions about the character states at any of the internal nodes of the tree. The first step is to find the maximum likelihood values of the transition parameters under the model of independent evolution of mating system and oestrous advertisement, and to calculate  $L(I)$ . The likelihood surface for so few species is nearly flat, making difficult the maximum likelihood search.



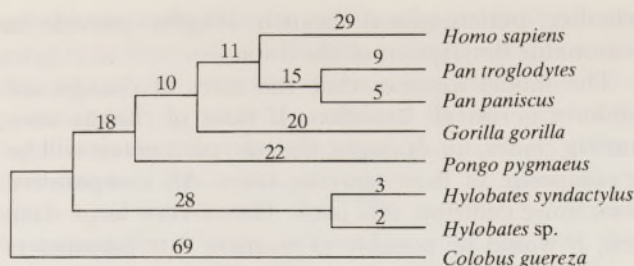


Figure 2. A phylogeny of the Hominoidea including one Old World Monkey. Two characters, social system and oestrous advertisement are scored for each species, where s = single male social system, m = multi-male, and '+' = the presence of oestrous swellings, '-' = absence. *Homo sapiens* (s, -); *Pan* (m, +); *Gorilla* (s, -); *Pongo* (m, -); *Hylobates* (s, -); *Colobus guereza* (s, -).

Based on a number of different searches with random starting points, the value of  $\log [L(I)]$  for this group is -11.91. Next the model of dependence is fitted. The value of  $\log [L(D)]$  is -8.43, representing a slightly better fit to the data than  $L(I)$ . The value of LR is 6.96. Not surprisingly for so few species this turns out to be a non-significant result when tested by Monte Carlo simulation ( $p > 0.12$ ).

The question of whether oestrous advertisement is more likely to evolve in the presence of a multi-male mating system is tested by comparing  $q_{12}$  with  $q_{34}$  from the  $Q_D$  matrix (taking oestrous advertisement as the  $Y$  variable). Their values are respectively 0.29 and 3.45. The larger value of  $q_{34}$  indicates that indeed there is a tendency for swellings to evolve more rapidly in the presence of a multi-male mating system. The question of which changes first in a branch, mating system followed by oestrous advertisement or vice versa can be investigated by comparing the pairs of rate parameters  $q_{12}$  and  $q_{13}$ ; their values are respectively 0.29 and 1.87 indicating that mating system is more likely to change first in a branch. However, none of these differences is significant with  $\chi^2$  tests.

One can also ask whether a monotonic transformation of the branch lengths significantly improves the likelihood. Fitting  $\kappa$  by maximum likelihood reduces the value of  $L(D)$  to -8.09. The value of  $\kappa$  is 1.92 suggesting that the branch lengths did not change substantially. Accordingly, the slight improvement in the likelihood is also not significant by a  $\chi^2$  test, and there is no evidence that  $\kappa$  differs from 1.0 for these data and branch lengths. Programmes to conduct the maximum likelihood searches and to conduct the Monte Carlo simulations are available from the author.

## 9. DISCUSSION

The model presented here provides a very general method for detecting correlated evolution in two categorical characters along the branches of a phylogeny. The weaknesses of the method are that it is computationally intensive, and for some hypothesis tests will require that the investigator generate the null hypothesis distribution. However, the strengths of the method are that it can make use of information on branch lengths, it does not rely upon any one

reconstruction of the ancestral character states, and it can be used to test highly specific hypotheses about the temporal order and direction of changes in two variables. Simulation studies are needed to give guidelines about the minimum number of different species that can be analysed under the model of correlated change.

This work was supported by SERC grant No. GR/F 98727. I thank David Balding, Mark Beaumont, and Diana Pallant for comments on the ideas presented in this paper. Simon Heath wrote the programmes and conducted all of the computing for the worked example. Brook Milligan has independently derived a model similar to the one I report here.

## REFERENCES

- Clutton-Brock, T. H. & Harvey, P. H. 1976 Evolutionary rules and primate societies. In *Growing points in ethology* (ed. P. P. G. Bateson & R. A. Hinde), pp. 195-237. Cambridge University Press.
- Cox, D. R. 1962 Further results on tests of separate families of hypotheses. *Jl R. statist. Soc. B* **24**, 406-424.
- Disotell, T. R. 1992 Molecular evolution of the Papionini (Primates: Cercopithecinae). Ph.D. thesis, Harvard University.
- Felsenstein, J. 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. molec. Evol.* **17**, 368-376.
- Felsenstein, J. 1983 Statistical inference of phylogenies. *Jl R. statist. Soc. A* **146**, 246-272.
- Felsenstein, J. 1985 Phylogenies and the comparative method. *Am. Nat.* **125**, 1-15.
- Gittleman, J. L. & Kot, M. 1990 Adaptation: statistics and a null model for estimating phylogenetic effects. *System. Zool.* **39**, 227-241.
- Gittleman, J. L. & Luh, H.-K. 1992 On comparing comparative methods. *A. Rev. Ecol. Syst.* **23**, 383-404.
- Goldman, N. 1993 Statistical tests of models of DNA substitution. *J. Molec. Evol.* **36**, 182-198.
- Grafen, A. 1989 The phylogenetic regression. *Phil. Trans. R. Soc. Lond. B* **326**, 119-156.
- Hall, P. & Wilson, S. R. 1991 Two guidelines for bootstrap hypothesis testing. *Biometrics* **47**, 757-762.
- Harvey, P. H. & Pagel, M. D. 1991 *The comparative method in evolutionary biology*. Oxford University Press.
- Hasegawa, M., Kishino, H. & Yano, T. 1985 Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *J. molec. Evol.* **22**, 160-174.
- Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S. & Takahata, N. 1992 Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. molec. Evol.* **35**, 32-43.
- Jukes, T. H. & Cantor, C. R. 1969 Evolution of protein molecules. In *Mammalian protein metabolism* (ed. H. N. Munro), pp. 21-132. New York: John Wiley.
- Kimura, M. 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. molec. Evol.* **16**, 111-120.
- Lindsay, J. K. 1974a Comparison of probability distributions. *Jl R. statist. Soc. B* **36**, 38-44.
- Lindsay, J. K. 1974b Construction and comparison of statistical models. *Jl R. statist. Soc. B* **36**, 418-425.
- Maddison, W. P. 1990 A method for testing the correlated evolution of two binary characters: are gains and losses



consequences: effects of phylogenetic accuracy on two comparative methods. *J. theor. Biol.* (In the press.)

- Received 9 September 1993; accepted 20 October 1993*