4

Comparative analysis of discrete data

'Few methods have been proposed which are framed explicitly in terms of character change along the lineages of a phylogeny' (Maddison 1990, p. 539).

4.1 Introduction

In this chapter we consider methods for the comparative analysis of discrete characters, or characters that can take only a finite number of states. The methods that we review and develop address the analysis of dichotomous characters—that is, characters that are found in just two states. However, this does not impose any limitations because any multi-category discrete variable can always be represented as a set of dichotomous classifications, each one representing the presence or absence of a particular state. In each branch of the phylogenetic tree, a trait that is initially absent may evolve or a trait that is initially present may disappear. These changes along the branches are called character state transitions, or just transitions. The problem, then, is to determine whether a particular pattern of observed transitions between alternative states of two phenotypic characters throughout a phylogenetic tree represents correlated evolutionary change in the two characters, or just a chance pattern of association.

The idea of a variable alternating between one of two character states finds striking parallels in other fields of biology. Island biogeographers have long confronted the problem of species turnover: species initially absent on an island may at a later time be present as a consequence of immigration, and species initially present may be absent as a result of local extinction. In fact, some of the models that we shall apply to characterize character transitions in a phylogeny are in some cases similar to those used by island biogeographers (e.g. MacArthur and Wilson 1967), and in other cases identical (e.g. Diamond and May 1977, note 7). Similar models have also been employed in the genetics literature to examine the consequences of both reversible mutation (e.g. Kempthorne 1957), and various selection processes (e.g. Felsenstein 1979).

Comparative questions involving discrete variables are widespread in biology. Recent studies range over issues from breeding system and mode of dispersal in plants; group living, coloration, and palatability in insects; sexual selection and breeding systems in birds; and sexual competition and social organization in primates (e.g. Givnish 1980; Harvey and Paxton 1981; Ridley 1983a, 1986b; Sillén-Tullberg 1988; Donoghue 1989; Höglund 1989; Maddison 1990). In Chapter 2 we considered the various biological reasons for closely related species having similar phenotypes. Whenever such similarity is inherited by descent from a common ancestor, we are not justified in treating species as independent points for statistical analysis. Instead, we must identify separate or independent evolutionary origins of the character states of interest. Because this chapter deals with statistical methods for analysing discrete character data obtained from a hierarchically nested phylogeny, we start by presenting a statistical complement to the material contained in Chapter 2. How do the biological processes described in Chapter 2 translate into phylogenetic similarity among discrete character states? And how might this similarity be dealt with? Many of these same themes will recur in Chapter 5 in conjunction with methods of analysis for continuous variables.

We then describe the kinds of transitions between character states that will be observed between two dichotomous discrete variables, and show how two methods designed to analyse discrete data from phylogenies, one developed by Ridley (1983a) and the other by Maddison (1990), treat these transitions. This is followed by examples of the application of both methods. Finally, we describe a general statistical method for treating comparative dichotomous data (a method developed by Pagel from one given by Pagel and Harvey 1989b). Here the parallels with models in island biogeography become apparent. This method makes it possible to use Ridley's and Maddison's methods while taking into account times of phylogenetic divergence.

4.2 The problem

4.2.1 Non-independence of data points

Suppose that we have collected together data on whether the larvae of each of many species of butterfly are warningly coloured or cryptic, and whether they are solitary or gregarious. We might ask whether there is evidence that warning coloration and gregariousness are related. Two dichotomous variables can give rise to four possible combinations of character states. It is tempting merely to tally the number of species that are in each of the four possible combinations of the characters and perform a statistical test of association, such as the chi-square. But most statistical tests assume that

the individual data points are independent, and the proposed chi-square test for association would be no exception. Therefore, for this test to give us correct probability values, we must assume that the phylogenetic structure underlying the evolution of these species allows them to be independent. Felsenstein (1985a) gives just such an example for continuous variables, which we shall use again in Chapter 5 and have adopted here for discrete variables (Fig. 4.1).
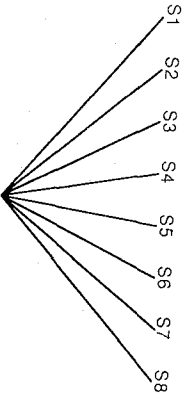
Fig. 4.1. Simultaneous radiation of eight species from a single common ancestor.

The eight species in Fig. 4.1 are all assumed to have radiated at the same time from a single common ancestor. On the assumptions that selective forces are not correlated among lineages and that the underlying genetical structure does not constrain evolution, the changes in any one lineage are independent of the changes in the other lineages, and so species can be treated as independent data points in an analysis. However, consider the more realistic phylogeny shown in Fig. 4.2.
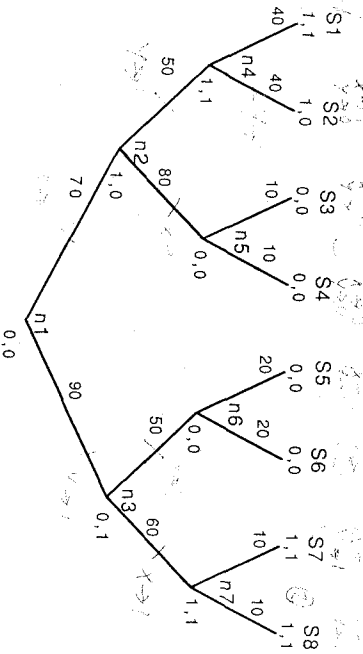
Fig. 4.2. Eight extant species (S1 to S8) which have evolved by a dichotomous branching process. Character states for two discrete variables (X,Y) are given for each species, as well as for ancestral species at each node (n1 to n7). The numbers given at the left of each branch represent times between successive nodes. Branch lengths are not drawn to scale.

We should be reluctant to treat species from this phylogeny as independent. Species S3 and S4, for example, share all of their evolutionary history except for the relatively small amount of independent evolution since they split from their common ancestor: the evolutionary changes leading to their contemporary character states have not been independent. Thus, to count both of them in a statistical test would overestimate the number of times that their particular combination of characters has evolved.

The usual symptom of non-independence is that closely related species tend to be more alike than more distantly related species. This point, of similarity associated with recent shared ancestry, is grasped intuitively, but what evolutionary models are implicit in it?

4.2.2 Two views of how non-independence arises

One view sees similarity associated with phylogeny arising essentially out of a constraint on how fast species can evolve (similar to 'phylogenetic time lags' in Chapter 2). For each of the branches in Fig. 4.2 we can write a variable t_{ij} indicating the length (in units of time or numbers of generations) of the branch connecting node i with node j. If we know the probability of a character changing state per unit of time and we assume that rates of change do not vary, then it is fairly easy to see that unless the probability is high or the branch length is very long, species with very recent common ancestors are more likely to have similar character states than more distantly related species. Thus, for example, consider that the branches leading from nodes n4, n5, n6, and n7 to their respective species are each of length 1.0 instead of the lengths shown on Fig. 4.2. Further, assume that the probability of a particular character changing state per unit time is less than 0.5. Then, more often than not, the species-pairs in the terminal branches will have the same character state, but not because they have each independently evolved it.

There is no doubt an upper bound to the rate of evolution, but whether it is important for the characters typically studied in comparative analyses is less certain. For some characters it probably is, others not. For example, the odds against a marsupial mammal evolving, in a single speciation, into a placental mammal are astronomical. Examples such as this may suffer from adopting the wrong level of analysis. That is, there are very likely intermediate degrees of placentation that can vary among species in, for example, the same genus. But the deeper problem with the 'not enough time' point of view is that it suggests that the only reason for phylogenetic similarity is that species have not spent enough time apart to become very different. This, surely, is not the whole story.

The second point of view, and the one implicit in the adaptationist explanations that we termed 'phylogenetic niche conservatism' and

'different adaptive responses' in Chapter 2, is that phylogenetic similarity reflects a similarity of selective forces. On this view, two species in the same genus are similar because they have shared histories of selective forces up to the point of their splitting; and because they are very likely to share many features in their current environments. The latter point is virtually guaranteed if we believe that similar species tend to be adapted to similar environments. It does not require any special models of speciation.

4.2.3 Implications of non-independence for statistical tests

Either viewpoint on phylogenetic similarity has important implications for testing comparative hypotheses. Species may not be independent with respect to some character because they have not, even in spite of selection to diverge, had enough time to become different. Non-independence will also arise due to shared selective forces and similar phenotypic responses to those forces. Correlating the character states of two variables across species, then, could include many data points that were not independent. As a consequence, the degrees of freedom for the statistical test will be overestimated and we cannot trust the probability values associated with the statistic.

We could, alternatively, restrict ourselves to higher nodes of the tree on the grounds that they are more likely to be independent. But this will be palliative at best, because exactly the same logic that was applied to the species can also be applied to differences among the higher nodes. The use of higher nodes rests implicitly on the assumptions of long enough branches or high enough rates of evolutionary change (or both), or such very different selective forces among the lineages that the higher nodes can be considered independent with respect to the characters under considera-tion.

4.3 Character transitions and discrete variables

If we often cannot consider species or even higher nodes as independent, then how can we use the information in the phylogeny to test for covariation between the two characters? One answer was hinted at in Fig. 4.1. The subtaxa within a monophyletic group share their evolutionary history up to the point of their origins. But, as a first approximation, the evolutionary change along an individual branch is independent of the change in other branches. Thus, one way to test whether two discrete variables have evolved in a correlated fashion is simply to correlate the changes in X and Y throughout a phylogeny. For example, between nodes n1 and n2 of Fig. 4.2, one variable changed from state 0 to state 1 but the other did not, which would count against the idea of correlated change

between characters. Continuing up the tree we could tally all such changes, ignoring branches in which neither character changes, and find their correlation. This is very close to what the method developed and used by Ridley does (Ridley 1983a, 1986b, 1988). We shall return to Ridley's method later on. But for now it is important only to realize that counting changes in different branches of the tree avoids the problem of non-independence from either point of view expressed above.

With two dichotomous characters there are, in fact, 16 possible combinations of change and absence of change that can take place along the branches of a phylogenetic tree. For example, consider two characters, X and Y, that can each exist in state 0 or 1. Possible states at the beginning and end of a branch for each character are 0→0, 0→1, 1→0, and 1→1. Each pair of states for one character can be associated with any of the four possible pairings for the other character, making a total of 16 combinations which we list in Table 4.1 for future reference. Nine of the combinations are evident along the 14 branches of Fig. 4.2: one combination occurs four times, one three times, and the other seven once each. In the following sections, we describe Ridley's (1983a) and Maddison's (1990) methods in terms of these transitions.

Table 4.1 Characters X and Y can each exist in state 0 or 1. There are 16 possible types of branch in a phylogenetic tree classified according to beginning and end states of the two characters. The numbered classes are those referred to in Table 4.3 and subsequently in the text.

Transitions in character X	Transitions in character Y			
	0→1	0→0	1→0	1→1
0→1	[1]	[2]	[3]	[4]
0→0	5	x6	7	x8
1→0	9	10	11	12
1→1	13	x14	15	x16

4.4 Counting evolutionary events

Ridley (1983a) designed his method for counting evolutionary events to get around the problems of non-independence in phylogenies. The method begins with a phylogeny that has been reconstructed according to one of the methods that was described in Chapter 3. It is assumed that the

reconstructed phylogeny represents our best estimate of the true phylogeny. To assign character states to higher nodes, Ridley suggests using outgroup analysis. However, as described in Chapter 3, rules suggested by Maddison et al. (1984) and by Swofford and Maddison (1987) are more appropriate if parsimony is the criterion to be used for reconstructing ancestral character states.

Once the higher nodes have been assigned, one works through the phylogeny, keeping a tally of the number of transitions in the tree (i.e. branches for which the beginning and end-states differ) that have *ended* in each of the four possible combinations of end-states. A transition is defined as a change of state in either or both characters along a branch. Because only the states of characters at the end of each branch are tallied, there will be only four different combinations. Branches along which *no* change occurs are not included. By not including such branches, the method avoids counting species or higher nodes which share character states with an immediate common ancestor, and which thereby cannot be considered independent data points.

The result of the tallying is a two-by-two contingency table, showing the number of branches along which transitions occurred that ended in each of the four states. The hypothesis that changes in X and Y are independent is tested by means of a chi-square statistic. A significant Chi-square (or Fisher's Exact Test) is evidence that the transitions have not been independent; some combinations of characters are more (or less) common than expected by chance. Table 4.2 shows the counts obtained by applying Ridley's method to the data in Fig. 4.2.

Table 4.2 Analysis of evolutionary events using Ridley's method applied to the data from Fig. 4.2. A tally of end-states in branches showing changes in X or Y produces the 2 × 2 contingency table, for which $\chi^2=0.19, \phi= 0.17$, and $P>0.50$.

	Y=0	Y=1
X=0	2	1
X=1	2	2

The logic of Ridley's method can be understood in terms of the 16 types of phylogenetic tree branch types shown in Table 4.1. Four of the 16 types (no change in either character) are ignored by Ridley, whereas the other 12 are treated together in four groups of three depending on the end-states. Thus,

for example, branches with characters X and Y both in end-state 0 can arise from three types of transition (X changing from state 1, Y changing from state 1, or both characters changing from state 1). Ridley's method treats state 1, or both characters changing as equivalent. In Table 4.3 we show, by reference to Table 4.1, which transitions are classified together by Ridley's method.

table 4.3 The way in which transitions from Table 4.1 contribute to Ridley's method of analysis. Types 6, 8, 14, and 16 are not included because they describe branches along which no transitions occur.

	Y=0	Y=1
X=0	7,10,11	5,9,12
X=1	2,3,15	1,4,13

Ridley's method, then, is correctly seen to be a test of whether end-states, defined as the character-states at the end of a branch, tend to be correlated. This interpretation suggests an imaginary 'phylogeny' consisting of many parallel branches. The beginning states of the branches are unimportant provided that one or the other of the characters has changed along the branch. If changes in the two characters are random, then there should not be an association among the end-states. Correlated evolution-ary change, however, will produce a correlation among the end-states. The method does not utilize the direction of evolutionary change, or which variable changed first. These will be key differences between Ridley's and Maddison's approaches (Section 4.5).

4.4.1 Applying Ridley's method: lekking and dimorphism

An example of the use of Ridley's method comes from Höglund's study of lekking and sexual dimorphism in size and plumage among birds. Höglund (1989) found a significant association between lekking and both size and plumage dimorphism across 114 bird species: lekking was associated with a higher proportion of the species showing size dimorphism and colour dimorphism (Table 4.4).

However, Höglund questioned these results on the grounds that all three variables show strong phylogenetic associations. To determine whether a relationship between lekking and dimorphism has actually arisen through the correlated evolution of the two characters, Höglund constructed phylogenies for his species, determined likely ancestral character states,

Table 4.4 Lekking and sexual dimorphism in size and plumage in birds. Each data point represents a single species. The associations between lekking and size and plumage dimorphism are both significant at $P<0.001$ in this comparison with species counted as independent points. (After Höglund 1989).

Mating system	Sexual size dimorphism		Plumage dimorphism	
	Present	Absent	Present	Absent
Lekking	69	18	55	33
Non-lekking	13	13	8	18

and counted the number of independent instances of the evolution of lekking and of sexual dimorphism in size and plumage in the Tetraonidae (grouse and pheasants). Fig. 4.3 shows the phylogeny for the Tetraonidae. The phylogeny illustrates many of the principles of reconstructing ancestral character states, as well as the steps for conducting Ridley's suggested type of analysis.
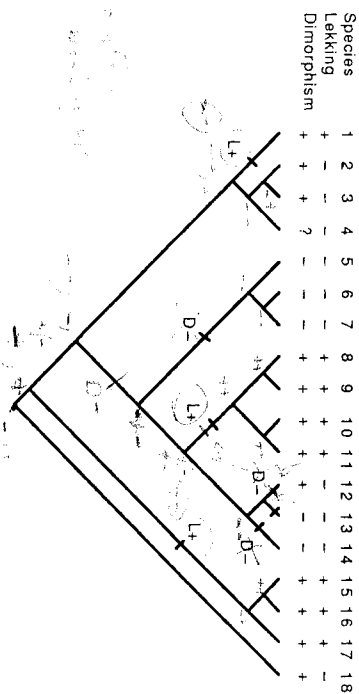

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Lekking	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dimorphism	+	+	+	?	-	-	+	+	+	+	+	+	-	+	+	+	+	-

Fig. 4.3. Lekking and sexual size dimorphism in the Tetraonidae. All species in this family are sexually dimorphic for plumage. Transitions on the phylogenetic tree (L, lekking; D, dimorphism; +, gain; −, loss) are calculated using the rules given by Maddison et al. (1984), described in Chapter 3 (Box 3.1). Species referred to by number are: (1) Centrocercus urophasianus; (2) Dendragapus falcipennis; (3) D. canadensis; (4) D. obscurus; (5) Lagopus leucurus; (6) L. mutus; (7) L. lagopus; (8) Tetrao urogallus; (9) T. parvirostris; (10) T. tetrix; (11) T. mlokosiewiczi; (12) Bonasa sewerzowi; (13) B. bonasia; (14) B. umbellus; (15) Tympanuchus cupido; (16) T. pallidicinctus; (17) T. phasianellus; (18) Phasianinae (outgroup). (After Höglund 1989).

Höglund's concerns about phylogenetic similarity are borne out: all members of the family Tetraonidae are dimorphic for plumage and yet each of the species is counted as a separate value in the test given in Table 4.4. Focusing on the other two variables, Höglund inferred that sexual dimorphism in size was lost at least three times while lekking appeared at least three times during the evolution of the Tetraonidae. Each of the instances of the loss of size dimorphism occurred in a non-lekking species and each instance of the gain of lekking occurred in a size-dimorphic species, a perfect association within this family.

However, the patterns were not as strong in other families. Table 4.5 shows the distribution of evolutionary changes for all the families combined. The result, in contrast to the highly significant across-species analysis, is not significant. Höglund's careful phylogenetic analysis reveals no significant evolutionary association between dimorphism in size and plumage in lekking birds. These results could change as new phylogenetic data become available, but they stand as a caution against drawing evolutionary inferences from analyses performed across species. Höglund also suggests that an analysis of size or plumage dimorphism may overlook other characters that females use in mate choice. For example, in the absence of plumage dimorphism, song dimorphism may have evolved.

Table 4.5 Lekking and sexual dimorphism in size and plumage in birds. The data points refer to independent comparisons calculated using Ridley's method. Using Fisher's Exact Test, neither association is significant. (After Höglund 1989).

Mating system	Sexual size dimorphism		Plumage dimorphism	
	Present	Absent	Present	Absent
Lekking	11	9	11	8
Non-lekking	6	9	6	7

Ridley has applied his method to a variety of cases where correlated evolution between characters might be expected (Ridley 1983a, 1986a, 1988). For example, he found that precopulatory mate guarding by male invertebrates tends to be evolutionarily associated with moulting by females, presumably because sexual receptivity predictably follows moulting (Ridley 1983a). And in a phylogenetically more diverse sample ranging from flies to fishes, Ridley (1983a) found that animals tend to mate with partners that are the same size as themselves among species in which: (1)

larger females lay more eggs; (2) larger males have an advantage in competition for mates; and (3) the duration of mating is long.

4.5 Tests of directional hypotheses with discrete characters

Ridley's method is designed to detect the pattern but not necessarily the direction of evolutionary change throughout a phylogeny. In this section we describe a test developed by Maddison (1990) for that purpose.

Maddison's test is designed to detect whether changes in one character are concentrated in certain regions of a phylogenetic tree. This might arise if the state of one character somehow makes the evolution of another more likely. Thus, Maddison's test explicitly treats one of the variables as the independent or 'causal' variable and the other as the dependent variable. It then searches for evidence that the likelihood of the dependent variable changing is higher in the presence of one category of the independent variable than in the presence of the other.

An example of an explicitly directional hypothesis of the sort for which Maddison designed his test is given by Sillén-Tullberg's (1988) work on the evolution of gregariousness in butterflies with warningly coloured larvae. Sillén-Tullberg found that gregariousness had evolved 23 times in the butterflies, and in 15 to 18 of these cases, the larvae were warningly coloured. Sillén-Tullberg concluded that this many gains of gregariousness in the presence of warning coloration suggest that warning coloration predisposes butterflies to evolve gregariousness.

Maddison's (1990) criticism of Sillén-Tullberg's approach, however, is that it fails to take into account the distribution of warning coloration in the phylogeny. Maddison points out that if, for example, warning coloration is very common in the butterflies then we would expect gregariousness to evolve in its presence more often than not, just by chance. This is easy to see in an extreme case: if all butterflies are warningly coloured then any evolutionary change toward gregariousness will be in the presence of warning coloration. Maddison's test, then, attempts to take into account the phylogenetic distribution of the traits, in addition to the number and pattern of transitions, in deciding whether a particular pattern is evidence for correlated evolution.

To see how Maddison takes into account the phylogenetic distribution of traits, consider the simple phylogeny in Fig. 4.4.

Two characters are coded on the tree. One, the independent variable can take the states W and B. The other takes the values 0 and 1. We take the ancestral condition for this tree to be $W0$, and thus the tree shows two instances of the evolution of B and two instances of the evolution of 1. To keep this example simple, we assume perfect knowledge of the phylogeny,
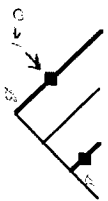
Fig. 4.4. A hypothetical phylogenetic tree. Branches marked in bold refer to lineages in which a character is in state B, with other branches in state W. Diamonds crossing the bold branches refer to evolutionary transitions of a second character from ancestral state 0 to derived state 1.

and that the 1s have evolved in both cases after the Bs. How should we test whether the 1s are more likely to evolve in the presence of B than in the presence of W?

We want to test the likelihood of having two gains of 1 in the presence of B, given the distribution of W and B throughout the tree. To see why, it is important to take into account the distribution of W and B. Consider that there is only one way in Fig. 4.4 that both 1s could have evolved in the presence of B, but many ways that they could have evolved if W and B are ignored. This is the essence of Maddison's test: we must calculate (a) the number of different ways that there are to have the two gains of 1 in the B branches of the tree given that there are two gains total (or more generally, the number of different ways that there are to have x or more gains of 1 in the B branches of the tree given that there are y gains of 1 and z losses of 1 over the whole tree: $x \leq y$); and (b) the number of ways that there are to have two gains of 1 on the tree without regard to whether they occur in W or B branches (or more generally, the number of different ways that there are to have y gains and z losses on the tree without regard to whether they occur in the W or in the B branches). The probability of the particular result under the null hypothesis is a/b. The denominator, then, records the number of ways that the pattern of gains or losses in the dependent variable could have occurred on the tree. The numerator is a subset of this number. It records the number of ways that a pattern of transitions as extreme or more extreme than that observed could have occurred in branches of the tree in which the independent variable takes one particular state (e.g. B).

Maddison's test can be illustrated for the tree in Fig. 4.4. It is straightforward to enumerate all possible outcomes corresponding to the quantities a and b. There turn out to be nine different ways that character state 1 can evolve twice on the tree (Box 4.1). Only one of these corresponds to evolving both 1s along the B branches of the tree. Thus the test yields a probability of 1/9 by chance alone that both instances of the evolution of '1' would occur in the B branches of the tree.

Box 4.1. The nine possible ways that 1 can evolve twice on the phylogenetic tree in Fig. 4.4 (top tree here)
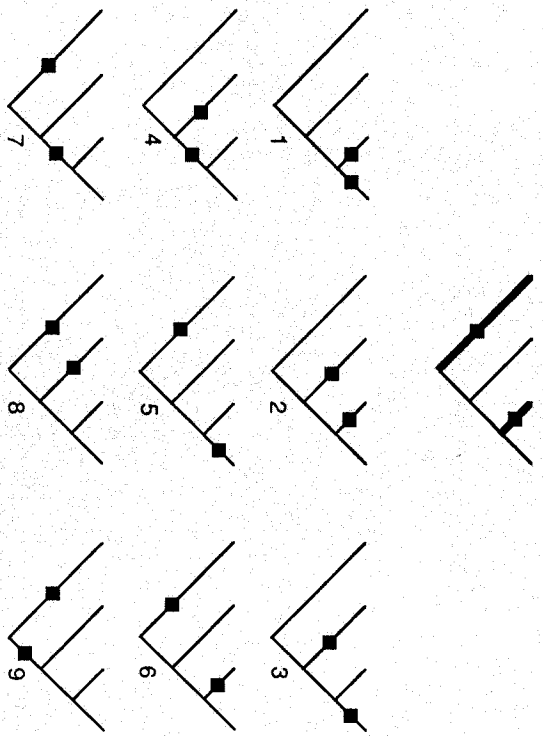
The nine redistributions of character states exhaust the possible ways of evolving the character 1 twice on the tree. Only one of them corresponds to the character evolving both times in the bold branches of the tree. Thus Maddison's test would assign a probability 1/9 to the likelihood that both 0→1 transitions occurred on bold branches by chance alone.

It might be thought that a much simpler test of this hypothesis would be to compare the proportion of B branches in which 1 has evolved with the proportion of W branches in which 1 has evolved. Maddison (1990) correctly points out that such a test ignores the fact that the topology of trees affects the number of ways that events can occur. This is most easily seen with an example. Box 4.2 shows a tree that, like the tree in Fig. 4.4, has two instances of the evolution of 1 in the presence of B, no losses of 1, and four W and two B branches. However, there are 11 possible ways that two instances of 1 could have evolved on the tree in Box 4.2 compared to the nine for the tree in Fig. 4.4 (Box 4.1). Thus, identical patterns of 1s evolving on the B branches of these two trees nevertheless have different probabilities of occurring by chance.

Box 4.2. How tree topology can influence the likelihood of correlated evolutionary events
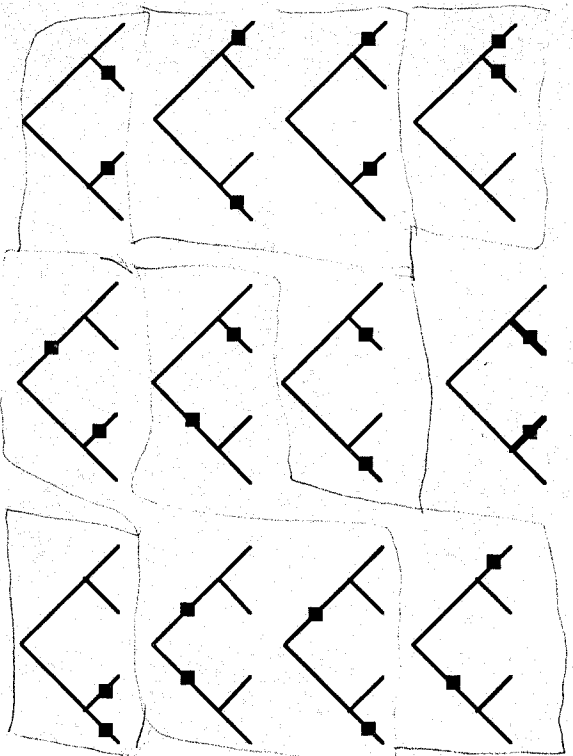
The top centre phylogenetic tree has several similarities with the tree shown in Fig. 4.4: it has the same number of branches in states B (bold) and W, the same number of character state transitions from 0 to 1, and both transitions occur on the B branches. However, the 11 ways that the character transition 0→1 can occur (as enumerated by the trees drawn next to and below the top centre tree) contrast with the nine arrangements for the tree shown in Fig. 4.4 (see Box 4.1).

For simple trees, Maddison gives an algorithm for computing the two values required for his test. However, this algorithm quickly becomes impractical to use for complex trees and so Maddison has developed a simulation technique for such cases. Maddison re-analysed Sillén-Tullberg's data set with this method and found that 15 to 18 gains of gregariousness in the presence of warning coloration could easily have occurred by chance, given that warning coloration is so widespread in the butterflies (Table 4.6 and Fig. 4.5).

4.5.1 Transitions among character states

Maddison's method uses the 16 transitions in Table 4.1 differently from Ridley's method. Returning to the example of Höglund's (1989) study of

Table 4.6 Data from Sillén-Tullberg's (1988) study of warning coloration (aposematism) and gregariousness in butterflies. Each record is of an evolutionary event. There are 15 to 18 instances of the evolution of gregariousness following the evolution of warning coloration. Using Sillén-Tullberg's estimates (see text): probability of 15 or more out of 20 (two-tailed test) = 0.04, probability of 15 or more out of 23 = 0.21, and probability of 18 or more out of 23 = 0.01. Families included are the Papilionidae, Pieridae, and the Nymphalidae. Re-analysis by Maddison's (1990) method, based on 23 gains and six losses of gregariousness and taking into account the phylogenetic distribution of warning coloration, gives P> 0.05 for both 15 or more and 18 or more gains in the warningly coloured branches. (See Fig. 4.5).

State of larvae	Solitary	Gregarious
Evolution of aposematism	9	0
State of larvae	Aposematic	Cryptic
Evolution of gregariousness	15	5
Evolution of aposematism and gregariousness inseparable	3	

Fig. 4.5. Frequency distribution of the expected number of changes to gregariousness in the warningly coloured branches of Sillén-Tullberg's (1988) phylogenetic tree of butterfly caterpillars, assuming the observed case of 23 gains and six losses in the tree as a whole. The distribution is based on 10 411 computer simulations. More than 20 gains would be judged significant; Sillén-Tullberg observed 15 to 18 gains. (After Maddison 1990).

lekking and sexual dimorphism, we showed how Ridley's method counted any transition in the phylogeny that resulted in either lekking and dimorphism or non-lekking and monomorphism as evidence for the relationship. But, for example, there are three ways that an end-state of lekking and dimorphism can arise: (1) a monomorphic lekking taxon may become dimorphic; (2) a dimorphic non-lekking taxon may evolve lekking; or (3) a monomorphic non-lekking taxon may show a double transition to dimorphism and lekking. Transitions with qualitatively different causalities are treated equivalently.

Maddison, on the other hand, would use each of these transitions differently. The first would be used as evidence for the directional hypothesis that the evolution of lekking precedes the evolution of dimorphism. The second outcome would be taken as evidence against the hypothesis. To see why, recall that Maddison's test involves counting the number of different ways that, in the present example, the observed
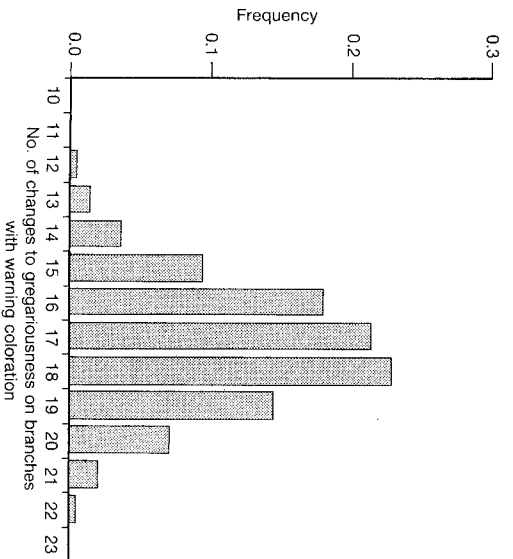
number of transitions to dimorphism in the lekking branches of the tree plus any other gain of dimorphism (not to mention losses) could have evolved on the tree. The more times that dimorphism has evolved in non-lekking branches, the greater the quantity a in Maddison's test will be, other things equal (e.g. compare the number of different combinations of five things taken five at a time versus five things taken three at a time). Because this number is the numerator of Maddison's test, it will increase the overall probability. The third outcome might be ignored since the order of the changes in a double transition is ambiguous. Alternatively, branches in which both characters change together could be assigned to first one then the other directional category to determine what effect, if any, their inclusion has on the results.

Maddison's method, then, explicitly tests whether the probability of the dependent variable (Y) changing from, say, 0 to 1 depends upon the state of the independent variable (X). This is equivalent to testing whether the probability associated with the cell numbered 5 in Table 4.1 differs from that of cell 13, taking into account back transitions and lack of change.

4.5.2 Maddison's method applied: the evolution of dioecy

Donoghue (1989) used Maddison's method to investigate the evolution of dioecy in flowering plants. Dioecious plants have separate male and female flowers, with only one sex per plant. Monoecious species have bisexual flowers. Givnish (1980) showed a strong correlation across species between dioecy and fleshy propagules that are dispersed by animals, and monoecy and propagules dispersed by wind: 339 of 384 monoecious species had wind-dispersed propagules, while 402 of 420 dioecious species had fleshy propagules ($\chi_1^2 = 570$, $P < 0.00001$). Givnish argued on the basis of this highly significant association that dioecy would be favoured in plants with fleshy propagules.

Donoghue (1989), however, was concerned that strong phylogenetic associations in these traits had led to gross inflation of the strength of the association that Givnish found. Further, Donoghue was interested in Givnish's explicitly directional suggestion that dioecy evolves after the evolution of fleshy propagules, and so used Maddison's approach. The cladogram in Fig. 4.6 shows the evolution of dioecy/monoecy and fleshy/non-fleshy propagules in the Gymnosperms. Ancestral states were constructed according to parsimony rules (Swofford and Maddison 1987). The mode of dispersal was uncertain in some taxa and so Donoghue tested his ideas in two ways using the cladogram in Fig. 4.6: one in which all equivocal taxa were taken to be animal dispersed and one in which they were treated as wind dispersed. Further, some branches showed simultaneous change in both characters. The results did not depend on which way the equivocal taxa were counted and so we confine our remarks only to the effect of the branches in which both characters changed.

When the branches with changes in both characters were treated as evidence for the relationship, Maddison's test gave a significant result: there were five or more gains of dioecy in branches with fleshy propagules, out of seven gains and two losses overall (Fig. 4.6, Table 4.7, $P = 0.02$). However, when the equivocal branches were not counted in favour of the hypothesis, the result became non-significant ($P = 0.13$). Donoghue cautioned that the uncertainties about phylogenies mean that his results cannot be taken as an unambiguous test of Givnish's idea, but rather as a beginning that will encourage further phylogenetic study. Leaving aside such difficulties, this example and Höglund's (1989) lekking example demonstrate how badly we can be misled when species are treated as independent data points.

4.6 A statistical model of evolutionary change

Ridley's and Maddison's methods both account for non-independence of species and provide ways of counting only independent instances of
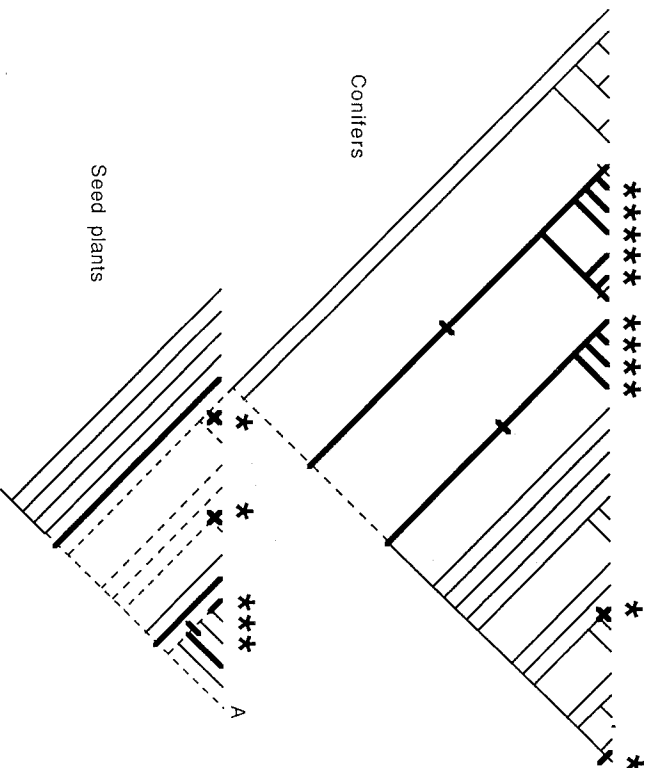
Fig. 4.6. A cladogram of seed plants showing parsimonious reconstructions of the evolution of fleshy propagules (animal dispersal) and dioecy. The phylogeny of angiosperms (A) is omitted from this figure. Branches in bold indicate fleshy propagules, non-bold branches dry propagules, and dashed branches are equivocal. Taxa known to be dioecious are marked by asterisks. Bold crossbars indicate the origins of dioecy and non-bold crossbars reversal to monoecy. (After Donoghue 1989).

evolutionary changes throughout a phylogeny. However, missing from both approaches is any way of taking into account the differing amounts of evolutionary change that might be expected to take place in branches of different lengths. This is an important point, because, for example, we may expect by chance that a variable will be more likely to change in a longer branch of a tree than in a shorter branch. Correlated change in a longer branch may be more likely by the same reasoning. We need a way of correcting for these differences. This section describes a method sketched out by Pagel and Harvey (1989b) and formally developed by Pagel (unpublished manuscript) for the evolution of discrete characters that takes into account both evolutionary branch lengths and the probability of character transitions. Ridley's and Maddison's methods can be derived

Table 4.7 Counts of the evolution of dioecy in gymnosperm lineages with fleshy propagules versus wind-dispersed propagules (data from Donoghue 1989). Analysis of this data set using Maddison's method which, despite the small number of changes, by taking into account the number of branches with fleshy and wind-dispersed propagules, yields P=0.02 if equivocal data are counted in favour of the hypothesis, but P=0.13 when equivocal data are not counted.

Dependent variable	Independent variable		
	Fleshy propagules	Wind-dispersed propagules	Equivocal propagules
Gains of dioecy	5	1	1
Losses of dioecy	2	0	0

from this model by making special assumptions about branch lengths and transition probabilities (see Section 4.9).

4.6.1 Evolutionary change in discrete characters

Consider again the hypothetical phylogeny in Fig. 4.2. We shall assume that it is the true phylogeny, and that we know the branch lengths and the character states at higher nodes. Each character can take one of two states, 0 or 1. A branch beginning in 0 may change to 1 in a given branch, or remain in state 0. The same is true for branches beginning in state 1. Now, given a null hypothesis of evolutionary change which treats the two characters under consideration as undergoing independent evolution, we can derive the expected values of the characters and of the variances of character changes in each branch of the phylogeny. The observed states can then be transformed to take into account these expectations. Then, we can correlate the transformed variables to test for covariation in the two characters.

We assume that the evolution of a dichotomous character can be modelled by a process that allows the character to change from one state to the other with specified probability per unit of time. The probability of a change is assumed to be the same in all branches of the phylogeny. For probabilities greater than zero, the model assumes that, over long enough time periods, the character is very likely to change state. Over very short time periods, the character can change but it is less likely to do so. An analogy with base substitutions can be used to illustrate what we mean by a probability of a character change. Assume that a base changes state at

Assumptions

some constant rate, r, when measured over long enough time periods. Presumably all of the forces required to cause a base change are present in the base or its environment, but nevertheless the base does not change every $1/r$ units of time, rather there will be a distribution of events (most likely Poisson), with a mean of $1/r$. This distribution says something about the probability of all the forces required for a base change coming together, and the variance of that process. Similarly, for more complex characters we assume that each has the potential to change, but when on average it actually does change depends upon the likelihood of all of the forces for change occurring together or perhaps in the right sequence. Transition probabilities, then, encapsulate our knowledge of all the forces acting on a variable in a particular branch.

More formally, assume that we have a dichotomous variable, X, which can take the values 0 and 1. Define αdt as the probability that X changes from state 0 to state 1 over some small unit of time dt:

$$P_{01}(dt) = \alpha dt. \tag{4.1}$$

Similarly, we can define βdt as the probability of X changing from 1 to 0 over a small unit of time dt:

$$P_{10}(dt) = \beta dt. \tag{4.2}$$

Note that α and β do not have to add to 1.0. For simplicity in the present discussion, we assume that α and β are constants. That is, the probability of a transition does not change in different branches of the tree. However, to make the equations more general one could simply attach subscripts i and j to any transition probability to indicate the two nodes that define the branch in the tree it is described. Similar transition probabilities between character states could be defined for a second character Y, and for any number of other variables (the equations describing a second and additional variables will be identical to those for α and β).

The probability of a character X ending in state 1 in a branch of arbitrary length t that began in state 0 can be modelled by a process that allows X to change forward and backward an indefinite number of times. The probability is a function of the relative values of α and β and the length of the branch:

$$P_{01}(t) = \frac{\alpha}{\alpha + \beta}(1 - \exp[-(\alpha + \beta)t]). \tag{4.3}$$

Similarly, if the character begins in state 1, then the probability that after time t it is in state 0 is given by:

$$P_{10}(t) = \frac{\beta}{\alpha + \beta}(1 - \exp[-(\alpha + \beta)t]).$$ (4.4)

Diamond and May (1977, note 7) reported eqns (4.3) and (4.4) in their study of species' turnover rates on islands as a function of immigration and local extinction rates. By the usual methods, we can derive expected values for the mean and variance of the characters. The expected value can be thought of as follows. If evolution was randomly re-run many times, some of the times a branch beginning in state 0 will end in state 1, the remainder of the times it would end in state 0. The average of these many 'trials' is the expected value. The variance is just the variance of these outcomes. The expected character state over an arbitrary amount of time t is given by[9]:

$$E[X_{01}(t)] = \frac{\alpha}{\alpha + \beta}(1 - \exp[-(\alpha + \beta)t]) = \mu_{01}.$$ (4.5)

The expected variance of the character after time t is given by:

$$E[VAR(X_{01})] = \mu_{01}(1 - \mu_{01}).$$ (4.6)

Similar expressions can be derived for the probability of a transition from 1 to 0. Thus, the expected character state after time t for a branch beginning in state 1 is:

$$E[X_{10}(t)] = 1 - \frac{\beta}{\alpha + \beta}(1 - \exp[-(\alpha + \beta)t]) = \mu_{10}.$$ (4.7)

The expected variance is given by:

$$E[VAR(X_{10})] = \mu_{10}(1 - \mu_{10}).$$ (4.8)

The expected values in eqns (4.5) to (4.8) can be used to transform all observations in the phylogeny to have the same mean and variance. Thus, for any given end-state (0 or 1) along a branch that began in 0, a standard

R error?

---

9 Equations (4.5) to (4.8), depend upon the arbitrary assignment of 0 and 1 to the two character states. For different values, the equations will change appropriately and the same ultimate answers will be obtained.

score can be formed by subtracting its mean (eqn 4.5) and dividing by the square root of its variance (eqn 4.6). Similarly, subtracting eqn (4.7) and dividing by the square root of eqn (4.8) will convert to a standard score an end-state that began in state 1. Doing this for each branch will scale all observations to a mean of zero and a standard deviation of one. The size of the standard score (either positive or negative) gives an indication of the degree of change relative to what would have been expected by chance. All the equations for transforming X and a second character, Y, to standard scores are shown in Box 4.3, along with an example transformation.

The formulae in eqns (4.5) to (4.8) yield positive standardized scores for transitions either from 1 to 0 or from 0 to 1, and negative standardized scores for a lack of change, be it either a branch beginning and ending in 0

Box 4.3. Equations for finding the mean and variance of change along a branch of a phylogenetic tree, and the computation of standard scores

I. Transition probabilities
$$P_{01}(dt) = \alpha dt$$
$$P_{10}(dt) = \beta dt$$

II. Expected character state after time t
$$\mu_{01} = \frac{\alpha}{\alpha + \beta}(1 - \exp[-(\alpha + \beta)t])$$
$$\mu_{10} = 1 - \frac{\beta}{\alpha + \beta}(1 - \exp[-(\alpha + \beta)t])$$

III. Expected variance after time t
$$\sigma_{01}^2 = \mu_{01}(1 - \mu_{01})$$
$$\sigma_{10}^2 = \mu_{10}(1 - \mu_{10})$$

IV. Computation of a standard score

Assume that character X has changed from 0 to 1 over a branch of length 50 and that $\alpha = 0.01$ and $\beta = 0.01$

$$\mu_X = \frac{0.01}{0.01 + 0.01}(1 - \exp[-(0.01 + 0.01)50]) = 0.32$$
$$\sigma_X^2 = 0.32(1 - 0.32) = 0.22$$
$$\text{standardized score} = \frac{1 - 0.32}{\sqrt{0.22}}.$$

or beginning and ending in 1. By convention, however, we assign (arbitrarily) a positive sign to any branch ending in 1 and a negative sign to any branch ending in 0. This convention says, in effect, that a transition from 0 to 1 or a branch remaining in 1 are the same kind of evidence. For example, Y and X both changing from 0 to 1, or Y and X both remaining in state 1, are evidence that Y and X evolve together.

The set of scaled observations on X and Y obtained by applying eqns (4.5) to (4.8) to each of the branches of a phylogeny can be tested using a Pearson correlation coefficient, under the assumption that change or lack of change in each branch is an independent event, and that the underlying distribution of standardized Y and X values throughout a phylogeny is bivariate normal. However, because the model does not assume normality a non-parametric test of correlation might be used instead. Either test assumes that the characters are evolving independently *in each branch* of the tree, and that ancestral conditions are known independently of the species values. In practice, neither of these assumptions is likely to be met. Ancestral conditions will be reconstructed from the species' values. This introduces some dependency between the higher nodes and the tips. As a result, treating each branch of the tree as an independent data point probably overestimates the true number of degrees of freedom. For example, two sister species with the same character state will cause the reconstructed ancestral condition to be that state, whether it was in fact or not: the reconstructed ancestral condition, then, is not independent of the two species.

The extent of non-independence between the tips and the higher nodes depends upon the distribution of character states among extant species. An extreme example, that of all species having the same character state, fully determines all higher nodes. Further, the reason for a lack of character change in some branches may be shared among them. Without simulation studies of a variety of realistic situations, it is difficult to know how much bias will be introduced if all branches of the tree are counted in the test. A bifurcating phylogeny of n species will always have 2n − 2 branches, but probably closer to n − 1 true degrees of freedom if the higher nodes are reconstructed from the tips. But consider a phylogeny with a large number of branches in which only a few show a character change. If the model developed here is used, and all branches are counted, a significant result is likely to emerge. One suggestion, then, is to use the data from all the branches but to assign only n − 1 degrees of freedom to the statistic. A more conservative approach is to use all of the data to estimate likelihoods but only use in the statistical test the data from branches in which one or both variables change.

4.7 Estimating transition probabilities by maximum likelihood

It is necessary to have information on branch lengths, and on the probabilities of transitions in order to apply the model we have described. Branch length information is becoming increasingly available (see, for example, references in Springer and Krajewski 1989), and we assume for the moment that the limiting information will be on transition probabilities. It is possible to estimate transition probabilities, if one is willing to assume that the likelihood of a transition in a particular direction (e.g. 0 to 1) for a particular character is constant throughout the tree. We describe here a maximum likelihood approach for estimating the transition probabilities from real data.

We assume that a phylogeny is available along with branch lengths. The problem is to estimate the transition probabilities in each character in such a way as to maximize the likelihood of the particular distribution of characters given the phylogeny. We describe here how to find the transition probabilities for the character X. However, all of the equations for a second or other characters are the same. The overall likelihood of the distribution of character states in X for a particular tree and values of α and β is given by:

$$L(D) = \Pi(P_{01}(t_i))\,\overset{0\rightarrow 1}{\Pi(1-P_{01}(t_j))}\,\overset{0\rightarrow 0}{\Pi(P_{10}(t_k))}\,\overset{1\rightarrow 0}{\Pi(1-P_{10}(t_l))}\,.\overset{1\rightarrow 1}{} \qquad (4.9)$$

where the first product is found over the branches showing a transition from 0 to 1, the second product is over the branches showing 'a transition' from 0 to 0, and the third and fourth products are for transitions from 1 to 0 and 1 to 1, respectively. We seek the values of α and β that maximize $L(D)$ where "D" signifies that the likelihood is found for the data given the tree. Equation (4.9) can be expressed as a log-likelihood:

$$\log_e[L(D)] = \Sigma\log_e(P_{01}(t_i)) + \Sigma\log_e(1-P_{01}(t_j)) + \Sigma\log_e(P_{10}(t_k))$$
$$+ \Sigma\log_e(1-P_{10}(t_l)). \qquad (4.10)$$

Analytical solutions to these equations will not be possible in most instances, so an iterative search procedure must be used to find the values of α and β that maximize equation (4.10).

In Box 4.4, we analyse the correlation between X and Y for the phylogeny in Figure 4.2 using maximum likelihood estimates obtained from an iterative solution of the log-likelihood equation (4.10).

Box 4.4. Standard score analysis of the data from Fig. 4.2 using maximum likelihood estimates of the transition probabilities

I. By maximum likelihood estimation

$$X: \alpha = 0.00857, \beta = 0.00585$$
$$Y: \alpha = 0.010913, \beta = 0.014635$$

II. Standard score analysis

Branch	X	Y	t	μ_X	μ_Y	S_X	S_Y	$\frac{X-\mu_X}{S_X}$	$\frac{Y-\mu_Y}{S_Y}$
n4→S1 *	1→1	1→1	40	0.18	0.37	0.38	0.48	+0.47	+0.76
n4→S2	1→1	1→0	40	0.18	0.37	0.38	0.48	+0.47	-1.31
n5→S3 *	0→0	0→0	10	0.08	0.10	0.27	0.30	-0.29	-0.33
n5→S4 *	0→0	0→0	10	0.08	0.10	0.27	0.30	-0.29	-0.33
n6→S5 *	0→0	0→0	20	0.15	0.17	0.36	0.38	-0.42	-0.45
n6→S6 *	0→0	0→0	20	0.15	0.17	0.36	0.38	-0.42	-0.45
n7→S7 *	1→1	1→1	10	0.05	0.13	0.23	0.34	+0.24	+0.39
n7→S8 *	1→1	1→1	10	0.05	0.13	0.23	0.34	+0.24	+0.39
n2→n4	1→1	0→1	50	0.21	0.31	0.41	0.46	+0.51	1.50
n2→n5	1→0	0→0	80	0.28	0.37	0.45	0.48	-1.61	-0.77
n3→n6	0→0	1→0	50	0.31	0.41	0.46	0.49	-0.66	-1.19
n3→n7	0→1	1→1	60	0.34	0.45	0.48	0.50	1.38	+0.90
n1→n2	0→1	0→0	70	0.38	0.36	0.48	0.48	1.28	-0.75
n1→n3	0→0	0→1	90	0.43	0.38	0.50	0.49	-0.87	1.27

* Branches along which neither character changes.

Note: signs of standardized scores are assigned according to the convention adopted in Section 4.6.1.

III. Spearman rank correlation test for correlated evolution of characters

For the 7 branches in which one or both variables change:

$$r_s = 0.25, P = 0.54.$$

For all 14 branches:

$$r_s = 0.31, P = 0.26.$$

4.7.1 What if branch lengths of the phylogeny are not known?

Even when branch lengths are not known it should still be possible to improve upon merely ignoring the issue of transition probabilities. We can ask, for a given tree with unknown branch lengths, what values of α and β maximize eqn (4.10)? The transition probabilities and the branch lengths are inseparable as all branch lengths are assumed to be equal. We assume that their product is a constant and estimate it by maximum likelihood just as we did previously. Under these conditions, the maximum likelihood estimates of α and β will, at least, take into account the pattern of transitions in the phylogeny, if not the branch lengths along which they occur. Moreover, α and β take values in this case that yield reassuring results. When all branch lengths are assumed to be equal, the values of α and β will be such that the predicted character state for any branch beginning with 0 is simply the proportion of such branches that showed a change from 0 to 1, and the predicted character state for any branch beginning with 1 is just the proportion of such branches showing a transition from 1 to 0: for character X in Fig. 4.2, then the value 0.25 is predicted for all branches beginning in 0 and the value 0.167 is predicted for branches beginning in 1, assuming that all branches are of length 1.0.

4.8 Applying the statistical model to a real data set

Most species of the dog family Canidae can be classified as either carnivorous or omnivorous and also as either showing limited biparental or more extended communal care of the young; data are given in Gittleman (1983). Is there a relationship between diet and the evolution of communal care in canids as suggested by Gittleman (1985)? We can make use of a phylogenetic tree of the canids recently compiled by Wayne and O'Brien (1987) and Wayne et al. (1989) based on isozyme genetic distance data and supported in part by DNA hybridization (Fig. 4.7).

Fourteen carnivore species are included in the sample, providing a phylogenetic tree with 22 branches along which path lengths can be measured. The Spearman rank correlation between standardized scores for the two variables obtained by applying the procedures in Sections 4.6 and 4.7 is highly significant ($r_s = 0.55$, $P = 0.01$). However, perhaps more appropriately, using only the five branches in which one or both characters changes (Section 4.6), the correlation becomes non-significant ($r_s = 0.50$, $P = 0.32$).
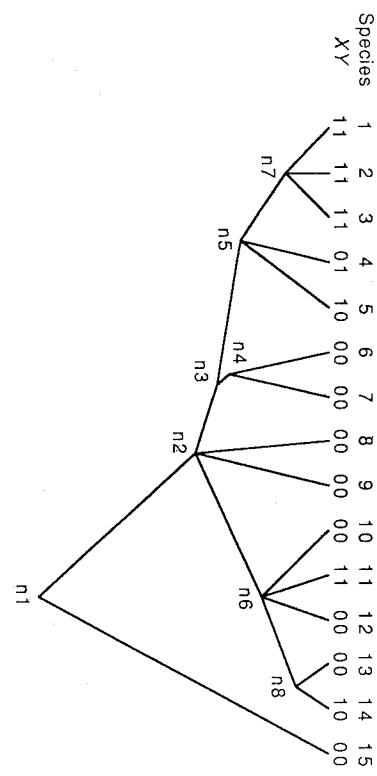
Species	XY
1	11
2	11
3	01
4	10
5	00
6	00
7	00
8	00
9	00
10	11
11	00
12	10
13	00
14	00
15	00

(tree nodes: n7, n5, n4, n3, n2, n6, n8, n1)

Fig. 4.7. Phylogenetic tree for fourteen canid species, with the black bear *Ursus americanus* used as an outgroup. Species 1 to 4 are: (1) grey wolf *Canis lupus*; (2) coyote *Canis latrans*; (3) cape hunting dog *Lycaon pictus*; (4) black-backed jackal *Canis mesomelas*; (5) bush dog *Speothos vanaticus*; (6) maned wolf *Chrysocyon brachyurus*; (7) crab-eating fox *Cerdocyon thous*; (8) grey fox *Urocyon cinereoargenteus*; (9) raccoon dog *Nyctereutes procyonoides*; (10) cape fox *Vulpes chama*; (11) red fox *Vulpes vulpes*; (12) fennec fox *Fennecus zerda*; (13) kit fox *Vulpes velox*; (14) arctic fox *Alopex lagopus*. Ancestral character states were calculated using Maddison *et al.*'s (1984) method (see Chapter 3, Box 3.1). Character X state 0 is omnivory, and state 1 is carnivory. Character Y state 0 is non-communal care, and state 1 is communal care. Approximate times before present (my) and ancestral character states (in parentheses) estimated according to the parsimony procedure described in Box 3.1, for each node are: n1 16.75 (00); n2 9.0 (00); n3 7.0 (00); n4 6.5 (00); n5 6.0 (11); n6 5.25 (00); n7 3.0 (11); n8 2.75 (00). (Phylogeny after Wayne *et al.* 1989).

4.9 Relationship of the statistical model to previous methods

4.9.1 The method of counting evolutionary events

Ridley's method for counting evolutionary events succeeds in deriving a set of independent data points, at least within the limitations of the methods used to reconstruct ancestral character states. However, each of the evolutionary events is given equal weight in the final chi-square or Fisher's exact test. But the model in the previous section shows how the likelihood of a character transition depends upon the length of the branch in which it occurs. What assumptions are implicit in assigning all branches equal weight?

Ridley's test can be understood in terms of the model presented in the previous section by noting that the chi-square statistic for a 2 × 2 contingency table is related to the correlation coefficient by the formula chi-square = $n r^2$ where n is the number of pairs of observations and r is the correlation coefficient between two dichotomous discrete variables (also known as the phi-coefficient). Thus, Ridley's test is simply a correlation of changes along the interior branches of a phylogeny, where all changes are assigned a 1 or a 0, and branches in which no change occurs are ignored.

The assumptions implicit in assigning equal weight to all transitions are made more clear by considering the example in Box 4.5. There we have calculated standardized scores after making all branch lengths in Fig. 4.2 equal to 10 and setting all four transition probabilities (α and β for X and Y) to a single value such that the standardized scores all take the values + 1.0 or −1.0. Doing this yields a phi-coefficient (or equivalently a chi-square) for the overall test that is identical to that obtained by Ridley's test. This result will be true in general, not just for the example of Fig. 4.2. Ridley's test, then, in weighting all changes equally assumes implicitly that all branch lengths are the same, and that α = β for both characters (however α and β for X need not be equal to α and β for Y).

A further assumption of Ridley's model is that branches in which no change occurs are irrelevant to the hypothesis; that is, Ridley's method only counts branches in which one or both of the characters change. By ignoring branches in which no change occurs, Ridley's method is, in terms of the statistical model, assigning them a path length of zero (that is, the path does not exist).

4.9.2 The method of directional change

How does Maddison's test relate to the statistical model? The main feature of Maddison's test is that it considers all possible combinations of the ways that a certain pattern of transitions could evolve. As our worked example showed, this is equivalent to finding all possible redistributions on the tree of one of the characters, namely, the one taken as the dependent variable, while holding constant the placement on the tree of the independent variable transitions. Each of these redistributions is then counted once in arriving at the values for the numerator and denominator of Maddison's test.

But these different ways of assigning the transitions to the tree are not equally likely if branch lengths differ. We can estimate the likelihood for any possible redistribution of the characters in a tree in terms of the model in eqns (4.1) to (4.4). This statement would be equivalent to that given in eqn (4.9) and would be calculated for the transitions in the dependent variable. In terms of Maddison's model, a different probability statement can be calculated from eqn (4.9) for each of the possible redistributions of

Box 4.5. Comparison of Ridley's method of counting evolutionary events with the general statistical model, assuming equal branch lengths and transition probabilities

The data from Fig. 4.2 are used, but path lengths are assumed to be equal.

I. Method of counting evolutionary events

A tally of end-states in branches showing changes in X or Y produces the 2 x 2 contingency table:

		X	
		0	1
Y	0		2
	1	1	2

$\chi_1^2 = 0.19$ $\phi = \sqrt{\dfrac{0.19}{7}} = 0.17$ $P = 0.66$

II. Statistical model

Assume all branch lengths = 10

Set $\alpha = \beta = 0.50$

From eqns (4.5) and (4.7) or Box 4.3:

$\mu_X = \dfrac{0.5}{(0.5 + 0.5)} \left(1 - \exp[-(0.5 + 0.5)10]\right) = 0.50, \; \mu_Y = 0.5.$

From eqns (4.6) and (4.8):

$\sigma_X = \sigma_Y = \sqrt{0.5^2} = 0.5.$

This general result is true for *any* common path length. For example, if all branch lengths are 50 and the transition probabilities are 0.10, the means and standard deviations are still 0.50. In this example, the standardized scores are arbitrarily given the same sign if the end-states match: +1 for end-states = 1, -1 for end-states = 0 (see end of Section 4.6.1). This is equivalent to Ridley's method.

Branch	End-State		Standardized score	
	X	Y	X	Y
n4→1 *	1	1	1	1
n4→2	1	0	1	-1
n5→4 *	0	0	-1	-1
n5→3 *	0	0	-1	-1
n2→n4	1	0	1	-1
n3→n6	0	0	-1	-1
n7→7 *	1	0	1	-1
n7→8 *	1	0	1	-1
n6→5 *	0	0	-1	-1
n1→n2	1	0	1	-1
n1→n3	0	1	-1	1
n3→n7	1	1	1	1
n2→n5	0	0	-1	-1
n6→6 *	0	0	-1	-1

When branches with no change (marked *) are ignored, a 2 x 2 contingency table comparing the number of branches for which the standardized Xs and Ys take the same or different values (+1 or -1) produces the same result as the previous method:

		Standardized X	
		-1	1
Y	-1		2
	1	1	2

Association statistics are, of course, also identical.

The analysis derived from Ridley's model is seen to be equivalent to the general model if all branch lengths are assumed to be equal, and $\alpha = \beta$ for each character. These circumstances will yield standardized scores of ±1. Ridley's method does not count branches along which there is no change of character state.

the character transitions on the tree. These different likelihood statements will not all take the same value, because, depending upon branch lengths, a transition may be more or less likely in one part of the tree than in another. The sum of the probabilities of all possible redistributions would conform to Maddison's quantity *b*. But now this quantity would not assign each redistribution an equal weight, but rather a weight based on the probability of that particular redistribution. A subset of the possible redistributions of the character on the tree will conform to the quantity *a* in Maddison's test. The test statistic would then be a weighted version of Maddison's, where each of the possible redistributions pertaining to Maddison's quantities *a* and *b* would be weighted by its probability of occurrence:

$$P(Y|X) = \frac{\Sigma(y_a)}{\Sigma(y_b)} \qquad (4.11)$$

where $P(Y|X)$ is the probability of the character transitions in *Y* (the dependent variable) arising by chance given the distribution of the *X* variable transitions, y_a are the weighted redistributions according to Maddison's quantity *a*, and y_b are all of the possible weighted redistributions.

We have calculated the different likelihoods for the nine possible redistributions of character states for the phylogeny of Fig. 4.4, according to the branch lengths shown in Box 4.6, and using the transition probability derived from the maximum likelihood solution. In this instance, only α is estimated since there are no back transitions in the tree, and thus β = 0.0. Not surprisingly, some of the redistributions are substantially less likely than others. For example, the redistribution corresponding to the two transitions occurring in the shortest branches of the tree is about 1/60 as likely as the tree with the two transitions in the longest branches. The ratio test for the observed tree (Fig. 4.4) is 0.076. This compares with 1/9 or 0.111 for Maddison's test. The statistical model gives a slightly smaller probability in this instance by recognizing that one of the transitions occurred in a very short path segment.

Like Ridley's method, the implicit assumptions of Maddison's test in assigning all possible redistributions of characters equal weight are that all branch lengths are equal and both transition probabilities (Maddison's test only studies transitions in the 'dependent variable') are the same. To see why, consider that each of the possible redistributions must be given the same probability by eqn (4.9). That is, all of the *y*s in eqn (4.11) must take the same value. Making all the branch lengths equal in the likelihood statement of eqn (4.9) and choosing α and β appropriately makes all the transitions have a probability of 0.5. Then, each redistribution has the

Box 4.6. The ratio test for the phylogeny shown in Figure 4.4

I. Assume unequal branch lengths as shown in the Figure below:
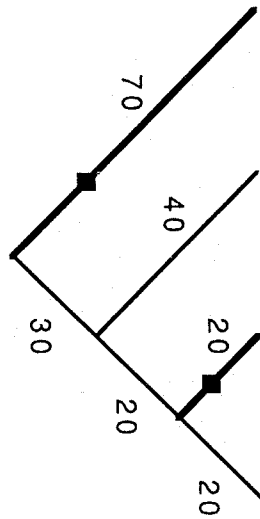
The maximum likelihood estimate for α in this tree is 0.0134. Using this value, we can find the likelihood for each of the nine possible arrangements in Box 4.1.

Arrangement (from Box 4.1)	Likelihood $y = \Pi(1 - e^{-\alpha t})\Pi(e^{-\alpha t})$
1	0.39 * 0.67 * 0.59 * 0.77 * 0.24 * 0.24 = 0.0065
2	0.39 * 0.67 * 0.42 * 0.77 * 0.24 * 0.77 = 0.0149
3	0.39 * 0.67 * 0.42 * 0.77 * 0.24 * 0.77 = 0.0149
4	0.39 * 0.67 * 0.42 * 0.24 * 0.77 * 0.77 = 0.0149
5	0.61 * 0.67 * 0.59 * 0.77 * 0.77 * 0.24 = 0.0328
6	0.61 * 0.67 * 0.59 * 0.77 * 0.24 * 0.77 = 0.0328
7	0.61 * 0.67 * 0.59 * 0.24 * 1.00 * 0.77 = 0.0560
8	0.61 * 0.67 * 0.42 * 0.77 * 0.77 * 0.77 = 0.0757
9	0.61 * 0.33 * 1.00 * 1.00 * 1.00 * 1.00 = 0.2016

$$P(Y|X) = \frac{\Sigma(y_a)}{\Sigma(y_b)}$$

For example, $P(Y|X)$ for arrangement 6 = $\frac{0.0328}{0.4501}$ = 0.073

II. Assume, instead, that all branch lengths are equal. If all branch lengths = 70 and $\alpha = 0.0099$, it follows that the probability of each path, with or without a transition, is 0.50. Then:

$$y = (0.50)^6 \text{ for each redistribution}$$

$$P(Y|X) = \frac{\Sigma Y(a)}{\Sigma Y(b)} = \frac{(0.50)^6}{9(0.50)^6} = \frac{1}{9} \text{ for any redistribution}$$

Thus, Maddison's test makes all arrangements equally likely by setting path lengths equal and choosing transition probabilities to be the same.

same probability of occurrence, and the overall likelihood of any particular outcome is just the sum of the (now all equal) probabilities that make up the quantity a divided by the sum of the (all equal) probabilities that make up the quantity b, or a/b.

4.10 Discussion

This chapter has described Ridley's method for assessing correlated evolutionary change in dichotomous variables, and Maddison's method for detecting the direction of evolutionary change. We showed how both of these methods could also be thought of as special cases of a more general statistical model that takes into account information on branch lengths. The general model can be applied to any data set by using the maximum likelihood procedures for estimating the model's parameters.

In the hypothetical examples that we used to introduce the statistical model, we assumed that the phylogeny and the ancestral states were known without error. This luxury of course will seldom be obtainable with real data sets, and so it is important to recognize some of the limitations of the tests that we have discussed. One difficulty concerns the assumption that each of the branches of the tree can be used as an independent data point in testing for the correlation of changes between Y and X. In a bifurcating tree with n tips, there will always be $2n-1$ apparent degrees of freedom for analyses. However, for two reasons there may be fewer actual degrees of freedom. The first is that if, for example, an ancestral node is in state 0, then each daughter node that radiates from it is constrained either to change to 1 or not at all: they cannot change from 1 to 0. Second, in most instances the ancestral nodes will not be known and will have to be estimated from the species values. This introduces a dependence between the ancestral and descendant states that is not easily quantified, except in

extreme cases. For example, if all species have the same value of the character, then all higher nodes will be reconstructed to have that character (see also discussion at end of Section 4.6.1).

Another difficulty is the assumption that the transition probabilities αdt and βdt are constant throughout the tree. If these two parameters vary considerably, then the scalings of X and Y will be incorrect. There is good reason to expect that αdt and βdt will vary (see Diamond and May 1977 for an example from island biogeography). However, this is not a limitation of the formal models (indeed the problem also plagues Ridley's and Maddison's methods), but of our understanding of which taxa should and should not be included in the same analysis. A related problem, perhaps, is that ancestral character states frequently will have been found according to a parsimony rule, but the statistical model assumes a Poisson process underlying the changes of character states. The effect this might have on the performance of the method has not been investigated.

Computer simulation studies, such as we report in Chapter 5, are needed to determine how seriously these issues may affect the statistical tests. Preliminary evidence concerning a problem of non-independence of branches with a method for continuous variables suggest that the effects may not be serious (see Section 5.10). Lacking evidence from simulation studies, there seem to be two courses of action, one more conservative than the other. The first involves reducing the number of degrees of freedom for statistical tests. Felsenstein (1985a) shows that $n-1$ independent comparisons between species and higher nodes can be derived from a bifurcating phylogeny with n species. Using this number would approximately reduce by half the number of degrees of freedom compared to counting all branches as independent points. A more conservative approach is to use in the final analyses only those branches in which one or the other character changes.

Another class of solutions would avoid conditioning the comparative test on any one particular set of higher nodes.[10] The idea is to estimate the four transition probabilities (X and Y changing from 0 to 1 and 1 to 0) independently of each other from the counts of the numbers of each kind of transition throughout the phylogeny. For any set of transition probabilities it is possible to calculate the likelihood of all possible reconstructions of ancestral character states. This likelihood can be compared with that derived from the set of transition probabilities that maximize the likelihood of finding the extant character states. Much work remains to be done in this area.

The examples that we have used to illustrate the model have all used bifurcating phylogenies. Even if some branches of the tree are not resolved

---

10 We thank Joe Felsenstein for this idea.

to this level, however, it is still possible to calculate standard scores. A multifurcating node can be thought of as representing two or more bifurcating nodes joined (implicitly) by a path or paths of zero length. The issue, then, is whether there have been any transitions along these paths. Probably the safest way to deal with a multifurcating node is to calculate only one standard score for each end-state represented among the sub-nodes (i.e. up to a maximum of two; see also Maddison 1989).

We have chosen a particular evolutionary model to describe the evolution of two dichotomous characters. More realistic alternatives may be possible. For example, rather than conceptualizing transitions between characters states as discrete jumps, there may be an underlying quantitative dimension whose phenotypic expression is two character states. Whether an individual is in one or the other state would depend upon the effects of many genes. Beyond a certain threshold along the dimension, an individual has one of the states, below the threshold it has the other (see discussion in Felsenstein 1988). A feature of the threshold model would be that not all individuals in the same state have the same probability of changing to the other: those closer to the threshold are more likely to change. Although the mathematics of this method are not yet worked out, it seems a promising approach.

Statistical and mathematical models of evolution are often criticized as biologically implausible or unrealistic. However, even techniques that apparently are not based on an evolutionary model may in fact just be special cases of *implicit* evolutionary models. These special cases may be even less realistic than the models. Thus, even if models are biologically unrealistic, they serve important functions. One is to make explicit the consequences of the assumptions made implicitly when using existing methods. For example, different redistributions of character changes on a phylogenetic tree varied 12-fold in their likelihood of occurrence according to the statistical model, and yet all were given the same weight by Maddison's method. An equally important function of models is that they force us to think clearly about the sorts of processes that are thought to give rise to the phenomena we are attempting to understand. This will, no doubt, lead to more complicated models that attempt to incorporate a greater number of the factors thought to be giving rise to the observed phenomena. Nevertheless, it is important to bear in mind that, in practice, results will depend partly on the phenomena under study, and partly on the particular assumptions of the methods used. Where the conclusions depend on the model used, this should be acknowledged explicitly, and some justification should be given for choosing one model over another.

4.11 Summary

This chapter develops a general model for the comparative analysis of discrete data. The model is designed to be used in conjunction with a phylogeny for which branch lengths, and the probabilities of character change, are known. A maximum likelihood estimation procedure is described for estimating the latter, even when branch lengths are not known. Thus, the model can be applied to existing data sets. Existing methods for the analysis of categorical data can be derived as special cases of this model, by making particular assumptions about branch lengths and probabilities of character change.