3

Reconstructing phylogenetic trees and ancestral character states

What might have been is an abstraction
Remaining a perpetual possibility
Only in a world of speculation. (T.S. Eliot, *Burnt Norton*)

3.1 Introduction

We described in the previous chapter how living organisms come to contain information about their evolutionary history, and why this means that comparative analyses must utilize phylogenetic information. This chapter asks what phylogenetic information we need for comparative tests, and how we might obtain it. It will become evident that although defining the information that we need is fairly straightforward, getting it is another matter. In fact the overriding message of this chapter will be that in the absence of a good fossil record we can never be sure of evolutionary history. We often produce quite different pictures of the past by basing our reconstructions on evolutionary models that make different assumptions about the roles of processes such as natural selection and genetic drift. An important task for biologists is to define which models are based on the most biologically realistic assumptions.

A phylogeny, which we treat as synonymous with a phylogenetic tree, is a genealogical history of a group, hypothesizing ancestor-descendant relationships (Levinton 1988, p. 49). Comparative tests seek evidence for correlated evolutionary change between the states of two or more characters. As a consequence, comparative biologists need to know in what lineages and at what times evolutionary changes occurred. This means that we need to know the branching pattern of the phylogenetic tree which links the species in our sample, and the positions on that tree of specified evolutionary events. More often that not, the timings of branching patterns and evolutionary events are given in relative rather than absolute terms.

The structure of the phylogenetic tree is used to tell us when any pair of species last had a common ancestor. Usually, the phylogeny simply tells us

which pairs of species had more recent common ancestors than other pairs. As data and techniques improve, however, inferred phylogenies increasingly include approximate dates of particular branching events so that the absolute rather than relative timing of occurrence of the most recent common ancestor is specified. The positions in the tree or timings of character state changes are usually provided by specifying ancestral character states at each node of the tree. If consecutive nodes have the same character state, it is assumed that no character change occurred along the branch linking the nodes, but if consecutive nodes have different character states we conclude that evolutionary change occurred along the branch. Our aim in this chapter, then, is to describe and assess techniques that are used to reconstruct both phylogenetic trees and ancestral character states.

The reconstruction of phylogenies has been the subject of considerable and often intemperate debate for many years (Felsenstein 1986; Hull 1988) and, more recently, the accumulation of molecular data has added a new level of interest. It is not our intention to review this debate, as such reviews have been undertaken regularly in the past by a variety of authors using different perspectives. Recent examples include Ridley (1986*a*), Hull (1988), and Sober (1989). Instead, we shall focus on the central issues which must be resolved if phylogenies are to be reconstructed with reasonable accuracy, and which bear upon determining hypothetical ancestral character states.

The next section, which considers the reconstruction of phylogenetic trees, asks which procedures for classifying organisms are most likely to result in the production of taxonomies that best describe phylogenetic relationships. Our conclusion is that cladistically defined taxonomies are usually the most suitable, but that some of the procedures commonly used by cladists can be improved upon to provide better estimates of true phylogenies. We go on to examine issues in tree reconstruction that are relevant to comparative studies. In particular, we show how the revolution in molecular biology can provide improved phylogenies, so long as the molecular information used is of the right kind and is from the right taxa. Finally, we discuss methods that can be used to reconstruct ancestral character states at the nodes of trees of known structure.

3.2 Reconstruction of phylogenetic trees

Once the need to use phylogeny in comparative studies has been accepted, the natural procedure has been to refer to standard taxonomies. Unfortunately, it is usually the case that traditional taxonomic relationships are not even *meant* to describe hypothetical phylogenetic relationships. For example, for reasons we shall describe below, common

taxonomic practice classifies crocodiles and lizards together in a taxonomic group from which birds are excluded, even though crocodiles have a more recent ancestor in common with birds than they have with lizards. This is a crucially important point for the comparative biologist, because the use of taxonomies that are known to contradict phylogenetic relationships will often lead to incorrect conclusions from comparative analyses: garbage in, garbage out.

Which taxonomies best represent phylogenetic relationships and which are best avoided? Different schools of taxonomy are easily identified: there are pheneticists, cladists, transformed cladists, evolutionary taxonomists, and others. Most contemporary taxonomies were constructed by evolutionary taxonomists and, as we shall see, their taxonomies explicitly were not designed to describe phylogenies. For entertainingly pithy accounts of the differences among the various schools, see Ridley (1983*b*, 1986*a*).

3.2.1 The schools of taxonomy

Evolutionary taxonomists have been careful to distinguish among the reasons for phenotypic similarity among species. Similarity can result from either convergent evolution, parallel evolution, or identity by descent (see Fig. 1.5). For evolutionary taxonomists, only those characters that are identical by descent should be used to decide upon taxonomic affinity. Convergently evolved characters, that is characters that are not homologous, should not be used to place species in the same taxonomic group. So much for similarity among characters, but what of phenotypic differences among species? Here, evolutionary taxonomists take a different stand: phenotypic difference often takes precedence over phylogenetic relatedness in the production of a taxonomy. One consequence of using phenotypic divergence to construct classifications is that some taxonomic groups are not monophyletic, by which we mean that they do not contain all the descendants of a particular common ancestor.[5] The birds, lizards, and crocodiles mentioned above are a case in point. Birds and crocodiles are more closely related phylogenetically, than either group is to lizards. However, because birds with their wings, feathers and beaks have evolved to look quite different from crocodiles, birds are placed in the Class Aves while crocodiles are placed with the phenotypically similar lizards in the Class Reptilia.

If there is one thing that virtually all comparative biologists are agreed upon, it is that taxonomic groups should be monophyletic, because only

[5] We define a monophyletic group as one containing all the descendants of a particular common ancestor, thus following Hennig's (1966) definition rather than those of Simpson (1961) and Mayr (1969). See Holmes (1980) and Ridley (1986*a*) for useful historical accounts of the concept.

then will a hierarchical taxonomy be isomorphic with a phylogenetic tree. Nevertheless, comparative biologists often use standard taxonomies derived by evolutionary taxonomists as substitutes for phylogenetic relationships. Yet, for example, 'only about half of the classes of the Chordata are thought to be monophyletic . . . and even within the Mammalia orders such as Insectivora and Carnivora are believed not to be monophyletic' (Felsenstein 1985*a*, p. 7). Many comparative studies of mammals continue to use Simpson's (1945) 'evolutionary taxonomy' as their baseline, despite the availability of classifications of the mammalian radiations which more accurately represents phylogeny (e.g. Eisenberg 1981, and examples in Benton 1988*a*, *b*). The classifications of evolutionary taxonomists are not satisfactory for comparative studies.

Pheneticists do not even consider phylogeny as a factor in the construction of their taxonomies. Instead, pheneticists construct taxonomies on the principle of phenotypic similarity. All characters for which information is available, whether evolutionarily convergent or divergent, are included to construct measures of phenotypic similarity. Then statistical methods for detecting clusters are used to produce a hierarchical classification (Sokal and Sneath 1963; Sneath and Sokal 1973). Different cluster statistics produce different taxonomic groupings, and none of them even purport to produce phylogenetic groupings. Furthermore, even if an appropriate cluster statistic could be defined, phenetic methods would still be inappropriate for identifying phylogenetic groupings. One important reason is that, as we shall discuss in the next section, character states that are primitive to a phylogenetic group should not be used for inferring relationships *within* that group (see Fig. 3.1). Pheneticists, however, treat

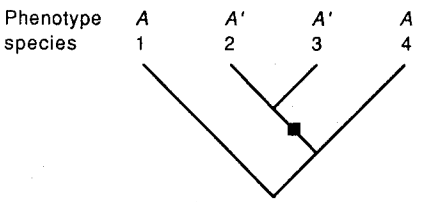


Fig. 3.1. A phylogenetic tree in which ■ represents an evolutionary transition of a character from state *A* to state *A'*. Phenetic similarity on the basis of that character would correctly group species 2 and 3, but incorrectly group species 1 with 4. An analysis using only shared derived character states would correctly group species 2 and 3, but leave other groupings undetermined.

derived and primitive character states equally.[6] The techniques of pheneticists, therefore, are not satisfactory for our purpose.

Cladists attempt to use phylogenetic branching as a basis for classification. Historically, cladists have used a set of rules that Willi Hennig (1966) devised with the express purpose of reconstructing phylogenetic trees as accurately as possible. Hennig's central claim was that a hierarchy of shared derived characters can be used to recognize a hierarchy of monophyletic groups which, itself, is a hierarchy of recency of common ancestry (see Fig. 3.1). *Pattern cladists* (Nelson 1979; Patterson 1980; Nelson and Platnick 1981, 1984) use Hennig's rules without the underlying philosophy of reconstructing phylogenetic trees, so they need not concern us here. Because phylogenetic trees are our goal, Hennig's approach makes a good start.

3.2.2 Shared derived characters as the basis for inferring phylogenetic relationship

Character states that are primitive to a monophyletic group are not useful for inferring evolutionary relationships within that group (see Fig. 3.1). Ridley (1986*a*, p. 54) gives a biological example: 'Suppose we wish to classify a baboon, a crocodile, and a cow relative to each other, and study the states of their limbs. The baboon and the crocodile have five toes, the cow two; but the fact that a baboon and a crocodile both possess a pentadactyl limb—the ancestral condition for tetrapods—is not evidence that these two species are phylogenetically closer to each other than is either to the cow. Shared ancestral characters do not reveal phylogenetic relationship'. Indeed, it is because lizards and crocodiles retain more ancestral character states in common than either group does with birds that evolutionary taxonomists classify lizards together with crocodiles in the Reptilia and birds separately in the Aves.

It was appreciated by some taxonomists early this century that only shared derived characters should be used to establish phylogenetic relationship (see Felsenstein 1982). For example, Mitchell (1901, pp. 181–2) described the practice as 'merely a codification of criteria in common employment among naturalists' and Le Gros Clark and Sonntag (1926, p. 478) declared that the aardvark *Orycteropus afer* 'shares primitive features with most Edentata, but these do not imply relationship'. However, Hennig's (1966) book, which prescribed a taxonomic procedure that was explicitly based on the shared derived character criterion, has been the most influential factor in the general acceptance of this procedure.

[6] It might be argued here that it is not possible to determine which character is derived and which primitive until a phylogeny is known! As we shall see in Section 3.2.5 and 3.3, a variety of methods are available to determine ancestral character states.

Hennig's procedure assumes that ancestral states for each character are known, that evolution of character states is irreversible, and that each character can change only once in a phylogenetic tree. Consider the data set given in Table 3.1 (simplified from Felsenstein 1982). All five characters (1 to 5) recorded in five species (*A* to *E*) can occur in either of two states, 0 or 1, of which 0 is ancestral.

Table 3.1 Each of five characters (numbered 1 to 5) which can occur in one of two states (0 or 1) are scored in each of five species (labelled A to E). A phylogenetic tree based on change in the first four characters but making certain assumptions (see text) about the evolution of character states is given in Fig. 3.2.

Species	Character				
	1	2	3	4	5
A	1	1	0	0	0
B	0	0	1	0	1
C	1	1	0	0	1
D	0	0	1	1	0
E	0	1	0	0	1

Given Hennig's assumptions, each character defines a monophyletic group. For example, character 1 defines (*AC*) and character 2 defines (*ACE*). Taken together, the first three characters define a unique phylogenetic tree, with which the fourth character also agrees (Fig. 3.2).

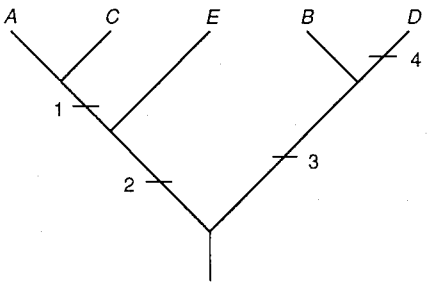


Fig. 3.2. The unique phylogenetic tree determined by the evolution of characters 1 to 4 in Table 3.1, under the assumption that each character can change state only once in the tree. The transition branch for each character is marked.

However, character 5 defines the monophyletic group (*BCE*) which contradicts the phylogenetic tree constructed by characters 1 to 4. There is no phylogenetic tree that can be constructed using all characters with Hennig's criteria. This is what Felsenstein (1982, p. 381) refers to as 'Hennig's dilemma'. How can it be resolved?

3.2.3 Resolving Hennig's dilemma: parsimony, compatibility, and maximum likelihood

Relaxing Hennig's assumptions

The resolution of Hennig's dilemma reduces to a matter of philosophy. If we use a hypothetico-deductive approach to tree-building, Hennig has put forward a series of assumptions which together constitute an hypothesis that the data can falsify. We might then change the hypothesis by relaxing one or more of Hennig's assumptions. We shall discuss a number of such possibilities below. For the purposes of the present discussion, we consider one scenario. What would be the consequence of relaxing Hennig's assumption that each character can evolve the derived state on only one occasion?

For the example discussed above (Table 3.1), if we allow characters to evolve their derived states on more than one occasion, we should be faced with a new problem: *all* potential trees can be assigned character states so that they fit the data! All trees could be chosen so that the state for each character at each node is 0. All evolutionary changes for characters 1 to 5 would then occur in the terminal branches linking each species to its most recent node. There are 11 origins of derived character states in such trees, corresponding to the 11 derived character states labelled 1 in Table 3.1. Some of the trees could be drawn with fewer state changes. For example, the tree in Fig. 3.2 would need three additional changes to deal with the evolution of 0 to 1 in character 5. Those changes would occur along the terminal branches leading to species *B*, *C*, and *E*, making seven character state changes in all. We need a criterion for choosing among all the different trees. One possibility is to choose the tree with the least number of character changes, summed across all characters. This would be a parsimony criterion.

Felsenstein (1982) reviews various relaxations of Hennig's assumptions, with their associated parsimony criteria. First, the same derived characters may be allowed to evolve more than once in the tree ('Camin-Sokal parsimony', after the method used by Camin and Sokal 1965); the most parsimonious tree is the one with the smallest number of separate derivations. Second, derived characters can originate only once in a phylogeny but may be lost many times ('Dollo parsimony', named loosely after Dollo's law which states that complex characters will not have evolved more than once); the most parsimonious tree is the one in which derived characters are lost on the smallest number of occasions. Third, both multiple derivations and multiple losses are allowed with the consequence that the ancestral character state for the common ancestor of all the species in the tree is not automatically specified; the most parsimonious tree is the one which shows the smallest number of character state transitions.[7] A fourth and final method allows an intermediate or polymorphic state (01) in any character; the most parsimonious tree is that with such polymorphisms persisting for the shortest amount of time. It is assumed that the polymorphism between any two character states can arise just once and that the derived character state can be reached only from the polymorphic condition (Inger 1967; Farris 1978; Felsenstein 1979). Other rules of character change are discussed by Gillespie (1986*a*) and by Maddison and Maddison (1989).

Hypothetico-deductive versus statistical methods for deciding among phylogenetic trees

The procedure just outlined is: (1) to decide on a set of rules (specified by Hennig's assumptions in the case discussed above); (2) if necessary, to relax one or more of the rules so that a number of trees are possible; (3) to choose that tree for which, summed across all characters, Hennig's original assumptions are broken the smallest number of times (a parsimony criterion). Alternatively, we might have decided to choose the tree containing the largest number of characters that break none of the rules (a compatibility criterion).

Indeed, the phylogenetic tree of choice could be decided upon using both parsimony and compatibility criteria. For example, a parsimony criterion, such as the minimum number of character state changes along the branches of the tree, might be used for choosing the ancestral state at each node for each character. A second criterion, such as that tree which is the most parsimonious for the greatest number of characters, could be used to decide upon the final phylogenetic tree. This would be a type of compatibility criterion. The most compatible tree would not necessarily have been chosen if the second criterion had been to select the most parsimonious tree, defined as that tree requiring the minimum total number of changes summed across all characters. For example, Table 3.2 gives the results of two alternative tree structures when a number of taxa were classified according to each of three characters. *A* is the chosen tree

[7] With multiple states, this model is either 'Wagner parsimony' (defined as the 'Wagner method' by Kluge and Farris 1969) when character states are in an ordered sequence (0, 1, 2, 3 . . .) with only single-step transitions allowed, or 'Fitch parsimony' when any state-to-state transition is possible.

when our compatibility criterion is used in the second stage of the process, but *B* is the chosen tree when parsimony is used. An important point to be made here is that, under our compatibility criterion, the number of characters for which the chosen tree is the most parsimonious is a subset of all the characters used, and the number of changes in the tree required by characters not belonging to that subset may be considerable.

Table 3.2 Several species are classified according to each of three characters. The minimum number of character state changes required to produce the same two hypothetical phylogenetic trees differs for each character. Summed across all characters, tree B is the most parsimonious tree, requiring a total of 13 character state changes. However, using a simple criterion of compatability, tree A which is the most parsimonious tree for both characters 1 and 2 would be chosen.

Character	1		2		3	
Tree	A	B	A	B	A	B
Character state changes	3	4	3	4	8	5

However, there are problems associated with applying either compatibility or parsimony within a hypothetico-deductive framework because the hypothesis of choice is falsified by the chosen tree if any rule specifying that hypothesis is ever broken. This can lead to what may seem to be an awkward situation: 'that phylogenetic hypothesis which has been rejected the least number of times is to be preferred over its alternates' (Wiley 1975, p. 243). Nevertheless, as we shall see below, parsimony and compatibility methods do, in fact, have an important part to play in phylogenetic tree reconstruction and in the reconstruction of ancestral character states. The important issue for our discussion will be to find the conditions under which they can be expected to yield the true phylogeny.

A different approach (Edwards and Cavalli-Sforza 1964, and championed in particular by Felsenstein 1973*a*, *b*, 1985*b*, *c*), is to view phylogenetic tree reconstruction as a statistical problem. One statistical procedure is to seek the maximum likelihood tree, which is that phylogenetic tree with the highest probability of having produced the observed data under a certain set of probabilities of character change. This statistical approach fits nicely with biological intuition. After all, provided that all transition probabilities from one state to another are greater than zero, no tree is impossible, it's just that some trees are more likely than

others. The problem is put into perspective by considering gene frequency data. Felsenstein (1985*b*, pp. 300–301) points out that, given appropriate levels of mutation, selection and drift: 'Any given pattern of gene frequencies could arise on any given phylogeny, although with lower probabilities on some phylogenies than others. The notion of falsification is called into question', because no tree could be falsified by any pattern of gene frequencies.

Maximum likelihood is a statistical estimation procedure which, in the case of phylogeny reconstruction, must be based on a model of evolutionary change. Maximum likelihood estimators have the property of consistency over a wide range of evolutionary rates and tree topologies. This property ensures that the estimator is increasingly likely to yield the correct tree as more data are collected. However, maximum likelihood solutions can be technically and computationally difficult to achieve even under some of the simplest (and most unrealistic) models of change. For example, Edwards and Cavalli-Sforza (1964) developed their method for constructing evolutionary trees based on the parsimony criterion of minimum evolution because they were unable to produce a suitable maximum likelihood solution for evolution under genetic drift. As a result, it has become common practice to use compatibility or parsimony procedures because they are tractable even though evolution does not necessarily proceed that way. If we are to use compatibility and parsimony procedures instead of maximum likelihood, we need to know the circumstances under which they yield maximum likelihood solutions (or are acceptable under some other statistical criterion), and then determine if those circumstances are biologically realistic.

When compatibility and parsimony procedures converge on the correct tree

Compatibility finds that tree which is compatible with the greatest number of characters given the rules of character change used (Estabrook 1972, 1980; Le Quesne 1969). The tree in Fig. 3.2, for example, is compatible with 4 of 5 characters under Hennig's assumptions for the ways characters can change. Character 5 would have needed to change three times to be compatible with the tree, but Hennig's rules say that it cannot have done this. When many characters are used in the construction of a tree, very few characters may be compatible with the chosen tree so that, in fact, the majority of characters do not help define the tree structure. Methods to incorporate information from non-compatible characters were pioneered by Estabrook *et al.* (1977) and by Estabrook and Anderson (1978). It might seem that compatibility methods return us to a hypothetico-deductive approach, but Felsenstein (1979, 1981*a*) has discussed some conditions under which compatibility methods provide maximum likelihood trees.

The latter are known in these circumstances to converge on the true trees as more data are collected. One circumstance involves the case where some characters change sufficiently rarely to produce only a single origin of a derived state while others produce many origins of the derived states. The characters with multiple origins contain little useful information and can, indeed, mislead taxonomists; they may be excluded by compatibility methods. A useful discussion of the role of compatibility methods in the reconstruction of phylogenies, including references to examples from buttercups to birds, is given by Meacham and Estabrook (1985).

Another possible resolution of Hennig's dilemma involves using parsimony which, in the case discussed in Section 3.2.2, means relaxing one of two of Hennig's assumptions: either a character state can arise more than once or character change is reversible. The most parsimonious tree is that which minimizes the number of additional evolutionary changes in the tree. Interestingly, Hennig (1950, 1965, 1966) prescribed neither compatibility nor parsimony. Instead, he recommended re-checking the character states in the original data. The hypothesis should never be refuted!

But is the most parsimonious tree most likely to be the correct tree? Not always (Cavender 1978; Felsenstein 1978; Hendy and Penny 1989). As an example, Felsenstein (1978, 1983) produced a phylogenetic tree containing four living species connected by branches along which rates of character state change had been different. The tree is reproduced in Fig. 3.3 and, for the sake of discussion, is assumed to be the correct tree. The probability of

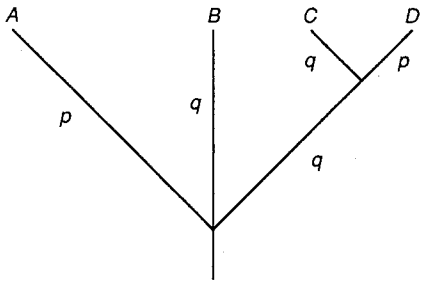


Fig. 3.3. A phylogenetic tree that is likely to produce data that will cause parsimony or compatibility methods to give misleading results. The correct tree is shown, with p and q being probabilities of character state change along the different branches of the tree. The wrong tree is likely to be determined when p is much larger than q because then B and C are likely to be more similar than are C and D. Details are given in the text. (After Felsenstein 1978, 1983).

character state change in any single branch is either p or q, but p is much larger than q. Transitions from character state 0 to state 1 are as likely as transitions from 1 to 0. When Wagner parsimony is used (forward and backward transitions allowed) on a single dichotomous character to reconstruct the tree from the character states of the extant species, the reconstructed tree is more likely to be incorrect than correct.

This is because a derived character found in only one species could have arisen in any terminal branch, thus providing no information concerning relationships among species. However, of the informative combinations of character states, the most likely for *ABCD* are 0110 or 1001 simply because B and C are likely to share the same character state if q is low enough: B would be classified with C, or A with D. Felsenstein demonstrates that if a tree was reconstructed using Wagner parsimony, so long as $p^2 > q(1\text{-}q)$, the most parsimonious tree probably would not show C being classified with D. Indeed, as more characters are included, each with the same rates of change, it becomes less and less likely that the most parsimonious or the most compatible tree will be the correct tree (Felsenstein 1979).

Most comparative studies have been analysed using statistical methods that assume equal branch lengths and equal rates of transition along each branch (as we shall see in Chapters 4 and 5). The dangers of making these assumptions without supporting evidence are clear from Felsenstein's example. In practice, rates of change for many characters may both covary with each other and vary among lineages for a variety of reasons. For example, neoteny or other changes in developmental timing may result in the seemingly co-ordinated evolution of many characters. Similarly, changes in rates of hormone production can affect the development of many characters. And, as we saw in the previous chapter, key innovations (Lauder 1981; Liem and Wake 1985) may herald predictable sequences of change in many characters.

Felsenstein (1983, p. 324) concluded that 'parsimony methods are well justified if the rates of change are sufficiently small or if they are sufficiently equal in different lineages'. In the example given above, rates of change differ among lineages. Sober (1989, p. 198), argues that Felsenstein's two conditions may be over-restrictive: because such conditions may be sufficient to make parsimony produce a maximum likelihood tree, this does not mean that parsimony methods would fail to produce approximations to maximum likelihood trees under other modes of evolution with higher rates of character change. Sober's distinction is between sufficient and necessary conditions. Hendy and Penny (1989), however, report that parsimony can converge on the wrong tree even when rates of change are equal, or when rates are unequal but 'low', for trees with more than four species. Hendy and Penny report that this is most likely to occur when the root of the tree divides into a single species on one

side, with the remaining species on the other side (as in Fig. 3.1). Parsimony may converge to the correct tree if one or more outgroups to the single species can be added to the tree.

In summary, we can never be sure that we have constructed the correct tree, and so we must rely on methods that will at least converge on the correct tree as more information is collected. This is the property of consistency. Maximum likelihood techniques for inferring trees are well justified in recognizing that any phylogenetic tree is possible (given that the probability of character change is not zero), and give consistent estimates under a very wide range of conditions. Unfortunately, maximum likelihood can quickly become computationally impractical for large numbers of species. Parsimony and compatibility methods can often be counted on to give consistent estimates of the true phylogeny when the probability of character change is small and roughly equal throughout all lineages (but there problems even here, as pointed out by Hendy and Penny 1989). Where there is evidence for parallel evolution or a character shows a large number of independent transitions, then these assumptions are questionable, and the resulting phylogeny should be treated with caution. Characters to be used for tree reconstruction are best chosen bearing those principles in mind. As we shall see in the next section, appropriate choice of DNA or amino acid sequence data may fulfil the twin criteria of low and equal rates of change.

3.2.4 Molecular data

Different DNA sequences diverge at different rates

A frequent assumption made when analysing data from macromolecules to produce phylogenetic trees is the constancy of the molecular clock: divergence at the molecular level occurs at a reasonably constant rate (in terms of absolute time or number of generations). A critical evaluation of the clock concept for the evolution of a variety of macromolecules is provided by Thorpe (1982). Fortunately, the clock runs at different rates for different types of macromolecule and for different regions of the same molecule, so an appropriate molecule may be picked for the question being asked.

When using comparisons among sequence data for reconstructing phylogenetic trees, sequences should be chosen which have diverged sufficiently so that the taxa in the sample can be discriminated from each other, while still bearing varying degrees of resemblance through common ancestry. Ribosomal RNA sequence analysis is useful for discriminating distant levels of relationship because of its relatively conservative structure, and has been used to throw new light on the origin of molluscs (Ghiselin 1988). Solution hybridization of total single-copy nuclear DNA

(DNA hybridization) provides useful data for closer relatives; it cannot be used to provide evidence about relationships between different phyla but, among birds, it works at all taxonomic levels, at least above the tribe (Sibley and Ahlquist 1983). Some regions of the mitochondrial DNA genome can be used to distinguish even closer phylogenetic relatives, particularly among species and populations (Wilson *et al.* 1985). Finally, the use of appropriate probes to either detect DNA single-locus polymorphisms (Quinn and White 1987) or distinguish between hypervariable DNA sequences, such as DNA fingerprinting (Jeffreys *et al.* 1985; Jeffreys 1987), are rapidly replacing analyses of electrophoretically detected enzyme polymorphisms to determine familial relationships within local single-species populations (see Burke 1989).

One particularly comprehensive application to date of molecular techniques for phylogenetic inference is Sibley and Ahlquist and Monroe's (Sibley and Ahlquist 1983, 1984, 1985; Sibley, Ahlquist and Monroe 1988) studies of birds and hominoid primates using nuclear DNA hybridization. Radioactively labelled DNA strands of about 500 nucleotides in length from a focal species were hybridized with similar strands from a number of phylogenetic relatives. The rate of dissociation of the duplexes provided a measure of similarity between the sequences. These molecular phylogenies agree, for the most part, with those derived from morphological and biogeographical data (Diamond 1983), but there have been some surprises. For example, Australian passerines seem to be more closely related to each other than they are to passerines from other continents, thus suggesting that previous classifications based on morphological comparisons had failed to detect considerable convergent morphological evolution. While DNA hybridization studies offer considerable promise for reconstructing phylogenies by estimating the amount of divergence among single-copy fragments of vertebrate genomes, we should be wary about treating many published studies as providing more than suggestive evidence (see Sarich *et al.* 1989). In particular, some measures of divergence that have been used may not be reliable and it is not possible to calculate alternative measures in the absence of published raw data.

For several reasons, mitochondrial DNA is proving a particularly powerful tool in phylogenetic tree reconstruction (Avise *et al.* 1987). For example, it is easy to isolate and assay, it has a simple genetic structure (lacking repetitive DNA, transposable elements, pseudogenes, and introns), and it is inherited without recombination. The order of genes (rather than the sequence of bases within them) is stable among three mammalian genera and an amphibian but differs between them and *Drosophila*, thus providing a possible route for phylogenetic reconstruction among very distantly related taxa (Harrison 1989). Nucleotide sequences can evolve about ten times faster than in most single-copy

nuclear DNA (Vawter and Brown 1986), although this relative rate varies among taxa, in large part because of varying rates of nuclear DNA evolution (Britten 1986; Moritz *et al.* 1987). Avise *et al.* (1987) describe mitochondrial DNA phylogenetic trees as 'self pruning' because mitochondria are usually maternally inherited (for exceptions see Kondo *et al.* 1990). For example, if females produce an average of one and a variance of five daughters according to a binomial distribution, then all mitochondria are likely to be inherited from the same foundress after $2n$ generations, where n is the number of females in the population (Avise *et al.* 1984).

Mitochondrial DNA has already established its credentials for phylogenetic inference among lower level taxa, for it has succeeded in several instances where other techniques have failed (see Moritz *et al.* 1987). In the past, restriction site maps or restriction fragment mobilities have been used for most studies, but in the future we can expect base sequence data to become more widely available. The reason for this optimism is the newfound ability to produce multiple copies of particular DNA fragments using the polymerase chain reaction. (In the past it was necessary to 'clone' the DNA using recombinant DNA technology, and to propagate it using *Escherichia coli*. With PCR, these steps are bypassed.) For example, using the method, Golenberg *et al.* (1990) have sequenced an 820 base pair DNA fragment from a 17 to 20 million-year-old magnolia (*Magnolia latahensis*) chloroplast gene. When they compared the sequence cladistically with sequences from homologous regions in other species, they found that it never grouped outside those of the other Magnoliidae that were examined. The polymerase chain reaction was introduced in 1985 (Saiki *et al.* 1985) but has now been both simplified and automated to the extent that large quantities of DNA can be obtained from a single molecule (Li *et al.* 1988), and sequenced directly (McMahon *et al.* 1987).

As lineages become separated for longer times, repeated substitutions at the same sites result in lower rates of sequence divergence (see Fig. 3.4 after Hixson and Brown 1986; Brown *et al.* 1982). Under such circumstances, a shift of attention to analysing base substitutions only from areas of the mitochondrial genome that change less rapidly (either RNA-coding sequences or second codon positions of protein genes) can allow reasonable resolution at higher taxonomic levels (Moritz *et al.* 1987). However, there is increasing evidence that rates of mitochondrial DNA divergence differ among lineages (e.g. Templeton 1987; Moritz *et al.* 1987). We have already seen how differing rates of evolution pose problems for phylogeny reconstruction based on parsimony analyses (see above and Fig. 3.3).

Comparisons between the structure of protein molecules have also been used for phylogenetic inference, sometimes as direct amino acid sequence data (e.g. Fitch and Margoliash 1970) and sometimes more indirectly as

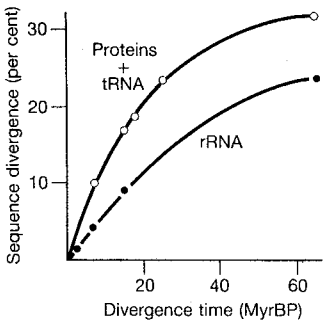


Fig. 3.4. The dynamics of mitochondrial DNA sequence divergence for primates, cow, and mouse (after Moritz *et al.* 1987). As lineages become separated for longer periods of time, repeated substitutions at the same site causes the apparent substitution rate to drop dramatically. Different parts of the mitochondrial DNA genome change at different rates, with regions coding for ribosomal RNA (rRNA) changing more slowly than those coding for transfer RNA (tRNA) and proteins.

immunological distances (e.g. Sarich 1977; Wilson *et al.* 1977; Wyles and Gorman 1980).

Models of molecular evolution

While, as we have seen, the molecular revolution is beginning to supply us with new data that will allow the increased resolution of phylogenetic trees, this does not mean that the data will analyse themselves. Indeed, there are formidable statistical problems associated with using those data correctly, and the development of better models based on an understanding of the forces involved in molecular evolution has become imperative (Gillespie 1986*a*, *b*, 1987; Friday 1989).

Gillespie (1986*a*) pointed out that many models for molecular evolution, whether constructed to deal with nucleotide, codon, or amino acid substitutions, assume that substitutions at different sites are independent events (in fact, independent stationary Markov processes). If that assumption is correct, queuing theory predicts that substitutions along lineages of set lengths will conform closely to a Poisson process. The ratio of the variance in the number of substitutions to the mean (R) should be one under a Poisson process. However, for protein sequence data at both nuclear and mitochondrial loci, R is usually much greater than one, with a modal value between 2.5 and 3.5. In a study of DNA sequences, the range is appreciably greater, extending to 35.0 for amino acid substitutions and to 19.0 for silent substitutions (Gillespie 1986*b*). This means that the variance in

the rate of evolution is at least an order of magnitude greater than expected from the neutral theory for some sequences (Gillespie 1987). Gillespie (1986*a*) points to four possible explanations for the deviations from a Poisson process. First, evolution within lineages could occur in bursts. Second, rates of evolutionary change could vary among lineages. Third, single mutations might affect more than one substitutional site. Fourth, correlated or compensatory substitutions occur. Distinguishing among these explanations and the suggestion that Gillespie's 'set lengths' were incorrect (Esteal 1990) is an important task.

Many interpretations of molecular evolution, as well as our understanding of the timing of branching points, depend on the notion of a molecular clock. As Gillespie (1986*a*, p. 659) points out, the evidence that R is greater than one 'does not argue against the molecular clock. Rather, it argues against using a Poisson process as a model of the clock'. Of the four explanations given above for the deviation of molecular evolution from a Poisson process, only the one that suggests rates of evolutionary change differing among lineages is at odds with the usual working definition of the molecular clock (see Thorpe 1982).

Many of the patterns already described, as well as others, argue strongly against purely neutral theories for molecular evolution. One alternative among the many possible theories which might account for some of the patterns in the data by allowing a role for selection was developed by Gillespie (see Gillespie 1987). The central concept is that of a 'molecular landscape'. At any point in time, the fixed allele at a locus is the most fit in the current environment among those alleles one mutational step away. When the environment changes, a previously deleterious allele may become favoured, and become fixed in the population. Now, a new set of mutants can be produced which were two mutational steps away from the previously fixed allele. One of these new mutants may be fitter than the currently fixed allele, and become fixed in the population. The process will continue until all the mutants that are one mutational step away from the currently fixed allele are less fit than it is. It can be seen that, following a change in the environment, this simple model can produce bursts of evolution, thus explaining deviations from the Poisson process outlined previously.

Almost all models of molecular evolution to date have been based on the neutral allele theory. If we are to meet the challenge of specifying models of molecular evolution, the demonstrated inadequacies of the strict neutral theory should prompt the development of realistic alternatives. Despite the 'extraordinary mathematical difficulties that these models present' (Gillespie 1987, p. 34), we hope that this section has demonstrated the critical role that they are destined to play in the phylogenetic and comparative analyses of the future. Phylogenetic trees and ancestral character states must be reconstructed on some theoretical foundation.

DNA–DNA hybridization data and the principles of cladism

The sequence of base pairs along the genome can be treated like any other character. If we knew the sequence along homologously derived sections of the genome for different species, we could use the agreement or disagreement at each of the sites as a character for cladistic purposes, according to Hennig's shared-derived criterion (e.g. Wölters and Erdmann 1986; Ghiselin 1988).

However, one technique for assessing the similarity of genomes from different species, DNA–DNA hybridization, has been criticized on the grounds that it cannot distinguish primitive from derived characters. Thus, the strength of the bond between between two strands of DNA will be a function of the number of sites at which they agree, regardless of whether the similarity is primitive or derived. The result, then, critics assert, is that DNA–DNA hybridization data have all the problems of phenetic approaches to classification. Not surprisingly, this charge has been denied by Sibley and Ahlquist (1987).

The debate over whether DNA–DNA hybridization identifies shared-derived characters revolves around the idea of how well the technique assesses total evolutionary change, which Springer and Krajewski (1989), whose arguments we summarize here, denote T. Assume that all genomes to be compared are equivalent in the sense that all sites in the genome of one species have homologous sites arranged in the same sequence in the genome of the other species. We wish to measure the pairwise differences between nucleotide sequences of a pair of species. The measure T is defined as a tally of all the point mutations that have occurred since the pair of species last had a common ancestor. T would include convergence events and is, by definition, a measure of total evolutionary change between two species. T is also an inverse measure of shared-derived characters: higher values of T, given our assumptions about genomic equivalence, mean fewer shared-derived characters.

Now, define a measure D, which is a simple tally of the number of homologous sites which differ in sequence between the two species. D will not count convergent events, and so will underestimate T. D, then, is a non-additive measure of the distance between two species. DNA–DNA hybridization can be thought of as an analogue measure of D. Thus, if we know the function linking D to melting points, we can convert temperature to genetic distance. This is what Sibley and Ahlquist attempt to do.

Viewed in this light, the debate about hybridization measures comes down to whether or not we believe that convergent evolution of nucleotide sequences is a common thing. If it is, and if it differs among lineages, hybridization measures will be distorted. If, however, convergence of nucleotide sequences is not common, hybridization measures will provide a

good indication of the number of shared-derived characters between two lineages, even though the measure does not count individual characters. Springer and Krajewski (1989, p. 314) conclude that if hybridization measures are additive the 'essence of cladistic intent (i.e. that net amounts of synapomorphy [shared-derived characters] are evidence of phylogenetic relationships) is not compromised, even though precise identification of character states is lacking'.

Cases of both parallel and convergent evolution at the molecular level have, in fact, been identified (e.g. Romero-Herrera *et al.* 1978; Liao *et al.* 1986; Stewart *et al.* 1987). In their study of myoglobin amino acid sequence differences among mammals, Romero-Herrera *et al.* (1978, p. 63) discussed two reasons for non-divergent evolutionary change. First, only a particular subset of all possible amino acid substitutions is consistent with myoglobin retaining its function as both a short- and long-term oxygen store. Second, not all of those substitutions can be accomplished by a single base change. However, once one base change has occurred, this may open new avenues for subsequent change. As a consequence 'constraints demanded by the functional morphology of the molecule itself and the constraints of the genetic code . . . contribute to parallel change in different lineages'. Furthermore, Romero-Herrera *et al.* found evidence to suggest that particular amino acid sequence changes in the myoglobin molecule may be favoured by a species' lifestyle. For example, the aquatic mammals (Cetaceans and Pinnipeds) seemed independently to have evolved an excess of arginine residues, any adaptive significance of which remains unknown.

Another example of apparent evolutionary convergence at the molecular level comes from Stewart *et al.*'s (1987) comparative study of lysozyme sequence data. Lysozyme is used to fight invading bacteria and many mammals have moderate to high levels in tears, saliva, white blood cells, and tissue macrophages. However, foregut fermenters such as langurs (*Presbytis entellus*) and cows also use lysozyme to digest bacteria that pass from the fermentative foregut into the true stomach. The *c* class lysozyme of foregut fermenters must work at low *p*H and be unusually resistant to breakdown by pepsin. The true phylogenetic tree, which puts cows closer to horses and langurs to baboons, is not the most parsimonious tree accounting for amino acid sequence divergence in *c* class lysozyme among the four species. Instead, langurs are more similar to cows than they are to baboons. Two alternatives to evolutionary convergence—horizontal transfer of genetic material between the ancestor of a cow and that of a langur, and gene duplication with homologous copies of the lysozyme coding gene being activated in the langur and the cow—were effectively discounted by Stewart *et al.* (1987).

Even though there is no simple way to estimate the degree of convergence at the molecular level using DNA–DNA hybridization, there is little doubt that most sequence evolution is predominantly divergent (Stewart *et al. 1987*). Sibley and Ahlquist (1987) do not believe that convergence is a serious problem. However, Felsenstein (1984) suggests that additivity of hybridization distances must be supported empirically, while Wilson *et al.* (1977) discuss some early evidence for additivity in hybridization data.

3.2.5 Rooting the tree

Our discussion up to this point has assumed that we know the origin or root of the phylogenetic tree. However, if characters can change state reversibly during evolution, as they do under the assumptions of Wagner parsimony, roots of phylogenetic trees are not automatically specified as they are under Hennig's original scheme, where the ancestor is a species with all characters in the ancestral state. Unrooted trees obtained by Wagner parsimony were termed 'Wagner Networks' by Farris (1970). When methods are available to identify ancestral forms and thereby root the tree, 'Wagner Networks' become 'Wagner Trees'. Unrooted trees may be adequate for many non-directional comparative studies, but for directional comparisons we must be able to distinguish ancestral from derived nodes (for a discussion of the difference between directional and non-directional comparisons, see Section 1.4.2).

Stevens (1980) and others have considered a variety of techniques that can help to root trees, of which outgroup comparison (including gene duplication events (Iwabe *et al.* 1989)), ontogeny, and paleontology are probably the most useful. The goal in each case is to identify the ancestral condition by some independent means.

Outgroup comparison involves expanding a monophyletic group by one step to include a new taxon which is likely to possess the ancestral state. A nice example of outgroup comparison being used to root a phylogeny is provided by Carson and colleagues' study of 103 species of picture-winged Hawaiian *Drosophila*. (Carson *et al.* 1967; Carson and Kaneshiro 1976; Carson 1983). The unrooted tree was derived from a comparison of chromosome inversions in the different species. *Drosophila primaeva* and *D. attigua*, a pair of sibling species, were treated as the outgroup and used to root the tree—the ancestral node with the most similar inversion patterns to those of *D. primaeva* and *D. attigua* was chosen as the root of the tree. There were two reasons for deciding upon them as the outgroup. First, they possess the chromosome pattern of closely related continental forms which are thought to represent the stock from which the Hawaiian *Drosophila* originated. Second, they are restricted to the oldest island of the Hawaiian chain.

The second method for rooting trees is based on ontogenetic comparisons among species. If evolution of some known characters proceeds by

what Gould (1977) calls 'terminal addition', so that derived characters develop later during development than primitive characters, then primitive character states may sometimes be identified using ontogenetic criteria. The value of ontogenetic comparisons to identify primitive states therefore depends on the extent to which a contemporary version of von Baer's (1828) first law holds true: the primitive features of a taxon appear earlier in the embryo than derived features. (von Baer, who was not an evolutionist, would have called them 'general' and 'special' features (Gould 1977)). Such character state changes are, by definition, not reversible and therefore provide information that can help root trees. However, evolution by terminal addition is not a universal phenomenon (see Gould 1977; Kluge and Strauss 1985; Brooks *et al.* 1985*a*, *b*). For example, many cases of neoteny (paedomorphosis) are well documented, which is equivalent to terminal deletion of many characters, and therefore to character state reversal. It might be argued that once complex characters have been lost they will not evolve again (Dollo's Law) but, not surprisingly perhaps, nature even provides exceptions to that generalization. Frogs lost their teeth in the Jurassic, which might be considered the straightforward loss of a fairly complex character. However, the potential to develop teeth seems to have been retained. Indeed, one South American frog *Amphignathodon* has actually re-evolved true teeth in its lower jaw (see Futuyma 1986). Nevertheless, such examples really are exceptions to a useful general rule, and several studies point to the value of comparative ontogenetic sequences in the reconstruction of ancestral character states (e.g. Miyazaki and Mickevich 1982; Alberch and Gale 1983). Although the evidence will often be tentative, if von Baer's first law is appropriately re-phrased and cautiously applied, it is probably 'true enough to be usable' (Ridley 1986*a*, p. 67) in many cases (Fink 1982).

The final technique for helping root trees is to locate appropriate fossil material as corroborative evidence, which may even be used to support the ontogenetic evidence (Miyazaki and Mickevich 1982). However, fossil material can be misleading. For example, taxa with the ancestral character state may be first detected in the fossil record at a later date than those with the derived character state, thus leading to the wrong character state being identified as primitive. That does not, of course, imply that anyone will find a Pre-Cambrian rabbit. Because of the many notorious gaps in the fossil record, there remains considerable debate about its general value as a reliable indicator of the direction of character change (Schaeffer *et al.* 1972; Ghiselin 1972; Paul 1982; Ridley 1986*a*).

3.2.6 On the incompleteness of phylogenetic trees

As we have emphasized, inferring the structure of phylogenetic trees is tricky—we are never sure that we have the correct tree. Comparative biologists should be aware of the fact that they may well be working with the wrong tree! A further problem is that most phylogenies are incomplete in two important ways which mean that comparative methods have to be designed to accommodate the inadequacies (see Section 5.2.1). First, many nodes have more than two daughter branches, and second, many branches which may in fact have been quite different are given the same lengths.

The first problem arises because several consecutive real dichotomies may be compounded into a single multiple branching event. For example, three or more species are often included together in the next higher taxonomic grouping—a single genus. More complete information would enable us to resolve many multiple branch points into a series of dichotomies. However, if speciation involves the simultaneous splitting of one species into three or more daughter species, multiple branch points would be accurate representations within the phylogeny.

The second way in which most phylogenies are incomplete concerns branch lengths and the timing of branch points. Not only are chronological timings usually missing from the tree, but relative timings are artificially classified. For example, when species are classified into genera, families, and orders by some systematists, nodes defining the origins of different families are, by implication, given the same times of occurrence. In theory at least, there is a time earlier than which a node denominates family status, and later a genus. As a consequence, nodes that occurred at very nearly the same time can be assigned a different taxonomic status. Similarly, nodes that occurred at very different times can be placed in the same category. The implicit assumption that equivalent taxa originated at the same time (e.g. the most recent common ancestor of all species in one family existed at the same time as the most recent ancestor of any other family) is made by several comparative methods described in Chapter 5. Of course, finer-grained taxonomic distinctions are often used, and the finer-grained the better as far as comparative biology is concerned. As we shall see in Chapters 4, 5, and 6, the better comparative methods can use any number of taxonomic levels.

3.3 Finding ancestral character states

The problem to be discussed in this section—the second task of this chapter—is how to determine hypothetical ancestral character states given a particular tree structure. As we shall see in Chapters 4 and 5, several types of analysis involve reconstructing hypothetical ancestral character states for the characters being analysed. However, as we discussed in the previous section, we have already considered most of the techniques involved because ancestral character state reconstruction often plays a major part in procedures used for reconstructing phylogenies.

Outgroup comparison has frequently been advocated as a method for reconstructing ancestral character states for dichotomously varying characters under the assumption of parsimony. Take the example of attempting to determine the character state of a common ancestor of a monophyletic group consisting of two species with different character states. One possibility is to expand the monophyletic group by one step to include the minimum number of new species with a less recent common ancestor. If all the additional species (the outgroup) demonstrate a single character state, then parsimony suggests that it is the ancestral condition for the original group (the ingroup). If the common ancestor did not contain the character state of the outgroup, at least one additional evolutionary transition would be required to account for the observed species diversity. If there were more than two species in the original ingroup, one possibility would be to decide on the most common character state among those species as being ancestral. Unfortunately, ingroup comparisons frequently do not provide such parsimonious trees as do outgroup comparisons, partly because it is entirely possible for the most common character in a group to be the derived state (Crisci and Stuessy 1980; Ridley 1983*a*).

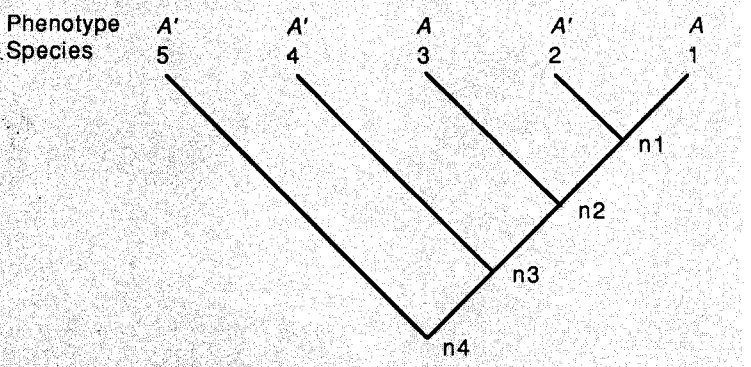
How, in practice, is outgroup comparison used? Futuyma (1986) gives a straightforward example. The anterior legs of species belonging to two butterfly families (Nymphalidae and Danaidae) are very reduced in size, whereas the anterior legs of species belonging to two other families (Papilionidae and Pieridae) are functional. Moths and other insect orders, taken as outgroups, have functional anterior legs, which would thereby be accepted as the hypothesized ancestral state. The families with reduced legs would, under outgroup comparison, be considered to share a more recent common ancestor than either does with members of the two families containing functional legs only.

However, outgroup comparison has a serious shortcoming for allocating ancestral character states to nodes: it may provide the most parsimonious *local* number of character state transitions, but not the single most parsimonious *global* number of state transitions (an improved method is provided by Maddison *et al.* 1984, see Box 3.1). If phylogenetic trees are known, algorithms, simple rules, and computer programs have been developed to determine hypothetical ancestral character states based on the parsimony assumption of the fewest number of character state transitions[8].

When characters are continuously varying with no restraints on direction of change, two methods are available to estimate the hypothetical

[8] Farris 1970; Fitch 1971; Hartigan 1973; Moore *et al.* 1973; Sankoff and Rousseau 1975; Sankoff and Cedergren 1983; Maddison *et al.* 1984; Swofford and Berlocher 1987; Swofford and Maddison 1987.

Box 3.1. Outgroup comparison, parsimony, and the reconstruction of ancestral character states



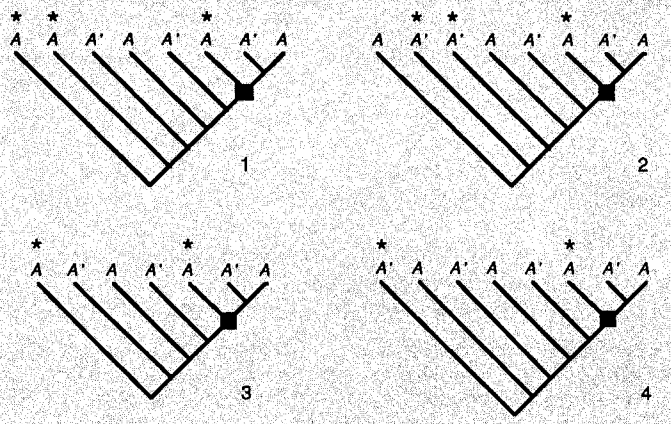
The problem is to allocate ancestral character states to the nodes n1 through n4. The first node, n1, is ambiguous but reference to its outgroup, species 3, suggests that the most parsimonious solution is for nodes n1 and n2 to be labelled *A*. A single transition from $A \rightarrow A'$ is then required from the node linking n1 with species 2. However, such a solution is only locally parsimonious because, if species 4 is considered, then two equally parsimonious allocations of ancestral character states are possible, one with n3, n2 and n1 labelled A' and the other with the same nodes labelled *A*. Consideration of species 5 labels both n3 and n4 as A'.

Maddison *et al.* (1984) provide an algorithm for determining the maximally parsimonious ancestral character states from comparison with local outgroups, under the assumption that all branches are of equal length and the probability of the transition from $A \rightarrow A'$ is the same as $A' \rightarrow A$. The algorithm gives rise to two straightforward rules if outgroups consist of single species or monomorphic taxa. The rules also provide satisfactory resolutions when multi-state characters are involved.

Rule One states that, as we move from the unresolved ingroup (species 1 and 2) out, if the first outgroup and the first doublet (consecutive outgroups with matching characters) have the same character state, then that state is the most parsimonious assignment for n1. (The first outgroup may constitute one of the members of the doublet.) When the first outgroup and the first doublet have different character states, no single most parsimonious allocation of ancestral character states exists. Rule Two, which is applied to cases where there are no doublets, states that if the first and last outgroups have the same character state, that state is the most parsimonious assignment for n1, but if the first and last outgroups have different

character states, no single most parsimonious allocation of ancestral character states exists.

The trees numbered 1 to 4 below, adapted from Maddison *et al.* (1984), demonstrate the use of the rules in practice. The extreme right hand node in each tree may be *A* or *A'*. We need Rules One and Two to determine the ancestral states at the nodes marked ■.



For tree 1, Rule One is applied because there is a doublet. The first outgroup and the first doublet (*) have the matching character state *A*, which is therefore the single most parsimonious ancestral state. Working down through the tree, the ancestral state at each node is *A*.

For tree 2, Rule One is also applied because there is a doublet. The first outgroup and the first doublet (*) do not have matching character states, therefore alternative ancestral character states are equally parsimonious. In this example, using the single outgroup method would designate the node as A whereas an equally parsimonious series of ancestral character states can be constructed with the node designated A'.

For tree 3, Rule Two is applied because there is no doublet. The first and last outgroups (*) have matching character state *A*, which is therefore the single most parsimonious character state.

For tree 4, Rule Two is also applied because there is no doublet. The first and last outgroups (*) do not have matching character states, therefore alternative ancestral character states are equally parsimonious.

A real example where these rules were used to estimate ancestral states for lekking and sexual size dimorphism for the phylogeny of grouse and pheasants (Tetraonidae) is summarized in Chapter 4 (Section 4.4.1 and Fig. 4.3).

The important point for the comparative biologist using parsimony to construct ancestral character states is that alternative equally parsimonious solutions are available for trees 2 and 4. In such cases, it is reasonable to examine the consequences of alternative allocations of ancestral character states. For example, Donoghue (1989) examined the evidence for correlated evolution between dioecy and fleshy propagules versus monoecy and wind-dispersed propagules. As we shall see (Section 4.5.2), he compared his results when ancestral character states at equivocal nodes were counted for the null hypothesis of no correlated change with those when equivocal nodes were counted against the hypothesis.

Outgroup comparison as well as the parsimony algorithm discussed above assume that character change in one direction is as likely as change in the other direction. In fact, once the ancestral character states have been designated, it may be obvious from the resulting allocation of ancestral character states that this assumption is not valid. For example, branches with ancestral nodes *A* may be less likely to change to *A'* than vice versa. If $A \rightarrow A'$ is less likely than $A' \rightarrow A$, the algorithm may have underestimated the number of instances of $A' \rightarrow A$. This is an important but largely untackled problem for comparative studies because reconstructing the most likely character states and designating appropriate probabilities of character change are central to testing many evolutionary hypotheses. We shall return to these issues in Chapter 4.

character states at nodes of the most parsimonious tree. The first is the 'median rule algorithm' (Kluge and Farris 1969) which minimizes the summed absolute changes of a character along the branches of a tree by selecting the nodes of the tree from the median value of the three adjacent nodes; an example of its use is given by Larson (1984). The second method, 'the averaging rule algorithm' developed by Felsenstein (in Huey and Bennett 1987; Huey 1987), iterates for each node the average of the three nearest nodes, and minimizes the sum of squared changes along the branches of the tree. The averaging rule method is a modification of Cavalli-Sforza and Edwards' (1964) minimum evolution method which, when applied to blood group data, provided so nice a fit to hypothetical human migration patterns (Cavalli-Sforza *et al.* 1964; Fig. 3.5).

Cavalli-Sforza and Edwards' minimum evolution model applied one of many genetic distance measures to gene frequency data. As we have seen, the minimum evolution criterion is not based directly on any realistic model for evolutionary change. Indeed, Cavalli-Sforza and Edwards (1967, p. 240) themselves emphasized, 'It certainly cannot be justified on the grounds that evolution *proceeds* according to some minimum principle'

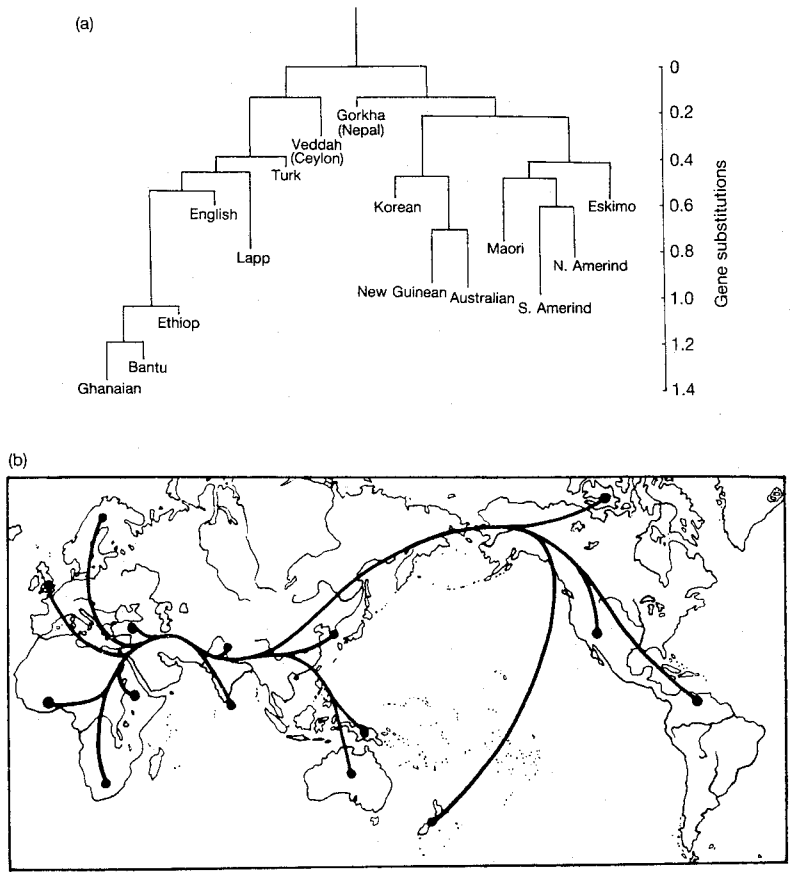


Fig. 3.5. Minimum evolution method used on blood group data from different human populations to produce a gene substitution tree (a), which was then superimposed on a map of the world (b) to show its correspondence with probable historical migration patterns. (After Cavalli-Sforza *et al.* 1964). Subsequent studies have improved on these data.

(their italics). Felsenstein (1985*b*) used a model of evolution for small populations under drift to evaluate several genetic distance measures, and found faults with each of them when initial gene frequencies are extreme. The important point about this exercise is not that Felsenstein's particular model is likely to be realistic, but simply that we need to know how measures of different genetic distance behave under specified conditions. Other comparisons of genetic distance measures under different regimes of evolutionary change (e.g. Latter 1973*a*, *b*; Nei 1976; Reynolds *et al.* 1983), including a neutral mutation model, help towards that end (Felsenstein 1985*b*). But all of this begs the central question: what is the most appropriate model of evolution? And that, perhaps, as the main message of this chapter, provides an appropriate note on which to finish this discussion.

A useful review of available computer programs or packages of programs that can use a variety of specified criteria to reconstruct both phylogenetic trees (including PHYLIP, PAUP and Hennig86) and ancestral character states for characters with specified rules of change (MacClade) can be found in Maddison and Maddison (1989).

3.4 Summary

Comparative methods need to utilize information on phylogenetic tree structure and ancestral character states. Maximum likelihood procedures based on appropriate models of evolution provide one suitable statistical technique for providing that information. Cladistic approaches using parsimony and compatibility criteria can produce approximations to maximum likelihood solutions under some models of evolutionary change. Standard taxonomies are often unsuitable for comparative studies because they do not accurately represent phylogenetic relationships. DNA sequence data from parts of the genome with appropriate amounts of divergence for the taxa being compared can provide particularly useful material for phylogenetic tree reconstruction.