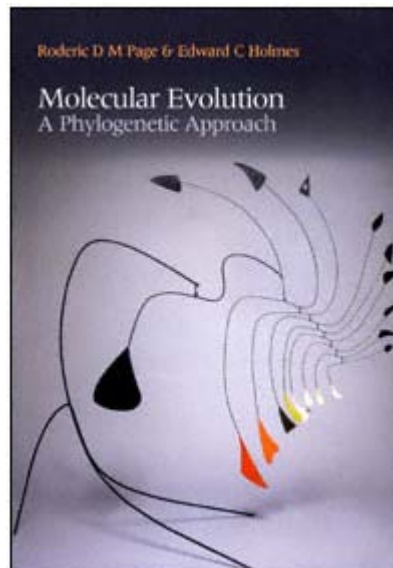




title: Molecular Evolution : A Phylogenetic Approach
author: Page, Roderic D. M.; Holmes, Edward C.
publisher: Blackwell Publishing Ltd.
isbn10 | asin:
print isbn13: 9780865428898
ebook isbn13: 9780632061488
language: English
subject: Molecular evolution, Evolutionary genetics.
publication date: 1998
lcc: QH390.P34 1998eb
ddc: 572.8/38
subject: Molecular evolution, Evolutionary genetics.

~~much the same way, when estimating branch lengths using least squares we can obtain tree distances that are less than the distances observed between the sequences (e.g. Table 6.1). Again, we cannot have less evolutionary change than we actually observe in the data.~~

~~6.3~~

~~Discrete Methods~~

~~In contrast to distance methods, discrete methods operate directly on the sequences, or on functions derived from the sequences, rather than on pairwise distances. Hence they endeavour to avoid the loss of information that occurs when sequences are converted into distances. The two major discrete methods are maximum parsimony (MP) and maximum likelihood (ML). Maximum parsimony chooses the tree (or trees) that require the fewest evolutionary changes. Maximum likelihood chooses the tree (or trees) that of all trees is the one that is most likely to have produced the observed data. We also consider two methods that treat the data not as individual sites but as 'splits' (see Chapter 2): spectral analysis and split decomposition.~~

6.4

Maximum Parsimony

The data for maximum parsimony comprise individual nucleotide sites. For each site the goal is to reconstruct the evolution of that site on a tree subject to the constraint of invoking the fewest possible evolutionary changes. Consider the following four sequences:

1 ATATT

2 ATCGT

3 GCAGT

4 GCCGT

Figure 6.11 shows the unrooted tree ((1,2),(3,4)) and two possible reconstructions of the evolution of the first site on that tree. In each reconstruction we have postulated which nucleotide the two internal nodes (ancestral sequences)

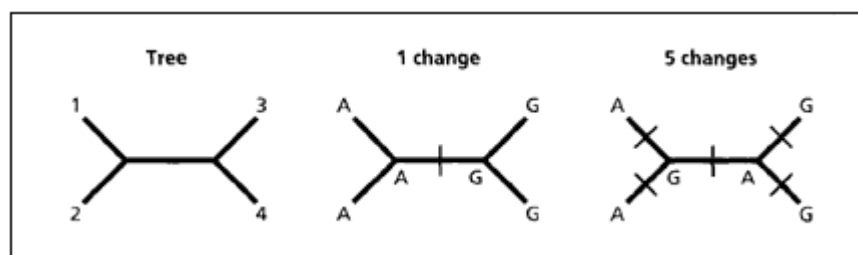


Fig. 6.11

An unrooted tree for four sequences 14 and two possible reconstructions of the evolution of the same site on that tree. One reconstruction requires one change, the other requires five. The former is more parsimonious.

possess. On branches where the nucleotide at each end differs we postulate a substitution. Note that our reconstruction does not specify where along that branch the change took place, merely that it did. Under the principle of parsimony the reconstruction that invokes a single substitution is preferred over the less parsimonious reconstruction that requires five substitutions. We can now find the most parsimonious reconstructions for each of the remaining sites (Fig. 6.12). Note that for the third site there are two equally parsimonious reconstructions, both requiring two steps.

The total number of evolutionary changes on a tree (often referred to as the tree's length) is simply the sum of the number of changes at each site. Hence, if we have k sites, each with a length of l , then the length L of the tree is given by

$$L = \sum_{i=1}^k l_i \quad (6.4)$$

Hence the tree in Fig. 6.11 has a length of $1 + 1 + 2 + 1 + 0 = 5$ steps. The other two trees have lengths of six and seven steps (Table 6.3), so the tree $((1,2),(3,4))$ is the most parsimonious tree.

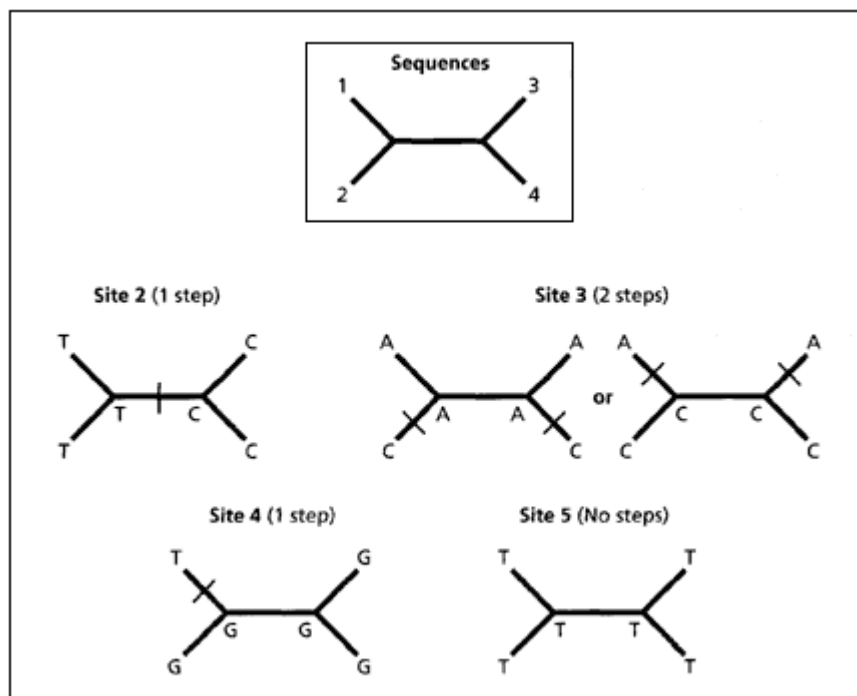


Fig. 6.12

Most parsimonious reconstructions for sites 25. There are two equally parsimonious reconstructions for site 3.

Table 6.3 Changes required for each site to fit the three possible trees for four sequences. Sites

Tree	Sites	1	2	3	4	5	Total
((1,2),(3,4))	1	1	1	2	1	0	5
((1,3),(2,4))	2	2	2	1	1	0	6
((1,4),(2,3))	2	2	2	2	1	1	7

Notice that the number of steps at sites 4 and 5 are the same for all three trees, hence these sites do not discriminate between the three alternative trees. Such sites are termed phylogenetically uninformative; sites that are invariant (each sequence has the same nucleotide at the same position) or sites where only one sequence has a different nucleotide are examples of such sites.

6.4.1

~~Generalised Parsimony~~

~~When counting changes in the above example, each substitution was accorded the same 'cost', so that the transition A → G counts as one step, as does the transversion A → C. Another way of representing this is by means of a step matrix (Fig. 6.13).~~

~~In the first model, which we have implicitly used above, each substitution has equal cost, so each cell in the corresponding step matrix has the entry '1'. Note that the cost of not changing is zero; if the nucleotide at each end of a branch is the same then we infer that no change has occurred.~~

~~The second model assigns a higher cost to transversions, in this case twice the cost of a transition. This is equivalent to saying the transversions are rarer.~~

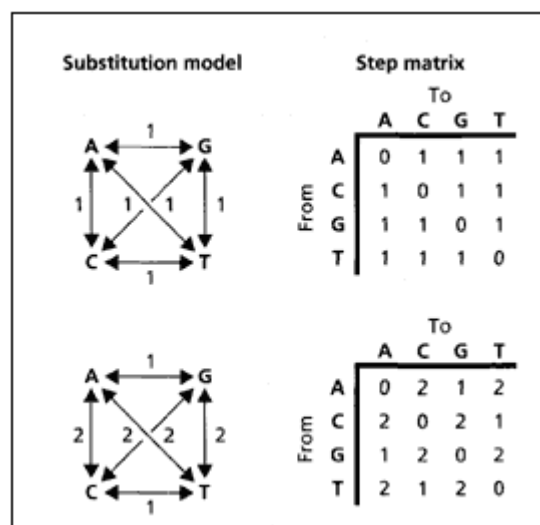


Fig. 6.13

Two different substitution models and the corresponding step matrices. The value in each cell in a step matrix is the 'cost' of the corresponding substitution.