Molecular Evolution

A Statistical Approach

ZIHENG YANG

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2014
Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2013956540

ISBN 978–0–19–960260–5 (hbk.)
ISBN 978–0–19–960261–2 (pbk.)

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

CHAPTER 3

Phylogeny reconstruction: overview

3.1 Tree concepts

This chapter introduces basic concepts related to phylogenetic trees and discusses general features of tree reconstruction methods. Distance and parsimony methods are described in this chapter as well, while likelihood and Bayesian methods are discussed later in Chapters 4 and 6–8.

3.1.1 *Terminology*

3.1.1.1 *Trees, nodes (vertexes), and branches (edges)*

A phylogeny or phylogenetic tree is a representation of the genealogical relationships among species, among genes, among populations, or even among individuals. Mathematicians define a *graph* as a set of *vertexes* and a set of *edges* connecting the vertexes, and a tree as a connected graph without loops (see, e.g. Tucker 1995, p. 1). Biologists instead use *nodes* for vertexes and *branches* for edges. Here we consider trees for species, but the description also applies to trees of genes or individuals. The *tips*, *leaves*, or *external nodes* represent present-day species, while the *internal nodes* usually represent extinct ancestors for which no sequence data are available. The ancestor of all sequences is the *root* of the tree.

3.1.1.2 *Root of the tree and rooting the tree*

A tree with the root specified is called a *rooted tree* (Figure 3.1a), while a tree with the root unknown or unspecified is called an *unrooted tree* (Figure 3.1b). If the evolutionary rate is constant over time, an assumption known as the *molecular clock*, distance matrix, maximum likelihood (ML) and Bayesian methods can identify the root and produce rooted trees. Such use of the clock assumption to determine the root of the tree is known as *molecular clock rooting*.

Another related rooting method, often used in the analysis of population data, is *midpoint rooting*. First, an unrooted tree is inferred without the clock assumption. Then the most distant pair of sequences is identified, with the distance calculated as the sum of branch lengths connecting the two sequences. The root is then placed at the midpoint between the two sequences. Midpoint rooting also relies on the molecular clock assumption.

When the species are distantly related or the sequences are fairly divergent, the clock assumption is most often violated. Incorrectly assuming the clock then can cause serious

Molecular Evolution: A Statistical Approach. Ziheng Yang. © Ziheng Yang 2014.
Published 2014 by Oxford University Press.

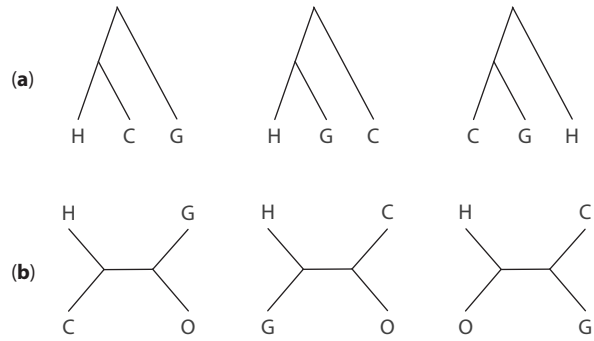


Fig. 3.1 Outgroup rooting. To infer the relationships among human (H), chimpanzee (C), gorilla (G), represented by the three rooted trees in (**a**), we use orangutan (O) as the outgroup. Tree reconstruction methods allow us to estimate an unrooted tree, i.e. one of the trees in (**b**). As the root is along the branch leading to the outgroup, these three unrooted trees for all four species correspond to the three rooted trees for the ingroup species H, C, and G.

errors in phylogeny reconstruction and in rooting. Without the clock and with independent rates for branches on the tree, most tree reconstruction methods are unable to identify the root of the tree and instead produce unrooted trees. Then the commonly used approach to rooting the tree is *outgroup rooting*. Distantly related species, called the *outgroups*, are included in tree reconstruction, while in the reconstructed unrooted tree for all species, the root is placed on the branch leading to the outgroups, so that the subtree for the *ingroups* is rooted. In the example of Figure 3.1, the orangutan is used as the outgroup to root the tree for the ingroup species: human, chimpanzee, and gorilla. In general, outgroups closely related to the ingroup species are better than distantly related outgroups.

In the universal tree of life, no outgroup species exist. Then a strategy is to root the tree using ancient gene duplications that occurred prior to the divergence of all existing life forms (Gogarten et al. 1989; Iwabe et al. 1989). The subunits of ATPase arose through a gene duplication before the divergence of eubacteria, eukaryotes, and archaebacteria. Protein sequences from both paralogues were used to construct a composite unrooted tree, and the root was placed on the branch separating the two duplicates (Gogarten et al. 1989). Elongation factors Tu and G constitute another ancient duplication, and were used in rooting the universal tree of life (Iwabe et al. 1989).

One should note that the output from a tree reconstruction program may look like a rooted tree. The user of the program is expected to know whether the analysis should produce rooted or unrooted trees and to interpret the tree accordingly.

3.1.1.3 *Tree topology, branch lengths, and the parenthesis notation*

The branching pattern of a tree is called the tree structure or *topology*. The length of a branch may represent the amount of sequence divergence or the time period covered by the branch. A tree showing only the topology without the branch length information is

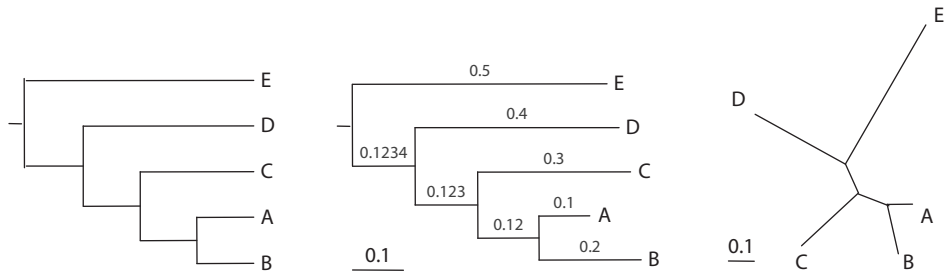


Fig. 3.2 The same tree shown in different styles. (**a**) The cladogram shows the tree topology without branch lengths or with branch lengths ignored. (**b**) In a phylogram, branches are drawn in proportion to their lengths. Here the branch lengths are shown along the branches. (**c**) In an unrooted tree, the location of the root is unknown or ignored.

sometimes called a *cladogram* (Figure 3.2a), while a tree showing both the topology and branch lengths is called a *phylogram* (Figure 3.2b).

For use in computer programs, trees are often represented using the nested parenthesis format or the *Newick format*, named after a lobster restaurant in Dover, New Hampshire, where the format was proposed (Felsenstein 2004, p. 590). For example, the trees in Figure 3.2 may be represented as:

a and **b**: ((((A, B), C), D), E);
b: ((((A: 0.1, B: 0.2): 0.12, C: 0.3): 0.123, D: 0.4): 0.1234, E: 0.5);
c: (((A, B), C), D, E);
c: (((A: 0.1, B: 0.2): 0.12, C: 0.3): 0.123, D: 0.4, E: 0.6234);.

Each internal node is represented by a pair of parentheses, which groups its daughter nodes, while the order of the daughter nodes is arbitrary. The outmost pair of parentheses groups the daughter nodes of the root. Tip nodes are represented by their names. A node can be followed by a semicolon together with a number that is the length of the branch ancestral to the node. Branch lengths here are measured by the expected number of nucleotide substitutions per site, like the sequence distance discussed in Chapter 1.

This format is natural to represent rooted trees. Unrooted trees are represented by placing the root at an arbitrary internal node and by having a trifurcation at the root. The representation is not unique, as the root can be placed anywhere on the tree. For example, the unrooted tree of Figure 3.2c can also be represented as '(A, B, (C, (D, E)));'.

Just as the Newick format does not represent the same tree in a unique way, there is much arbitrariness when a tree is drawn. For example, the root can be on the top, at the bottom, or on the side. To decide whether the different trees are equivalent, think about whether they represent the same evolutionary/genealogical relationships: for example, the three trees shown in Figure 3.3 are identical.

Because of different rates of evolution in different lineages, there may not be direct correspondence between evolutionary *relatedness* and sequence *distance* between two species: two closely related species may not have the smallest sequence distance. The distance is the amount of sequence evolution and is equal to the sum of branch lengths on the paths connecting the two species, while the relatedness is measured by the time of divergence between the two species (see Problem 3.3).

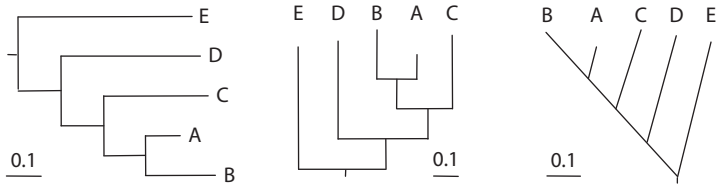


Fig. 3.3 All three trees are identical as they represent the same genealogical relationships among A, B, C, D, and E. This is the same tree of Figure 3.2b: ((((A: 0.1, B: 0.2): 0.12, C: 0.3): 0.123, D: 0.4): 0.1234, E: 0.5).

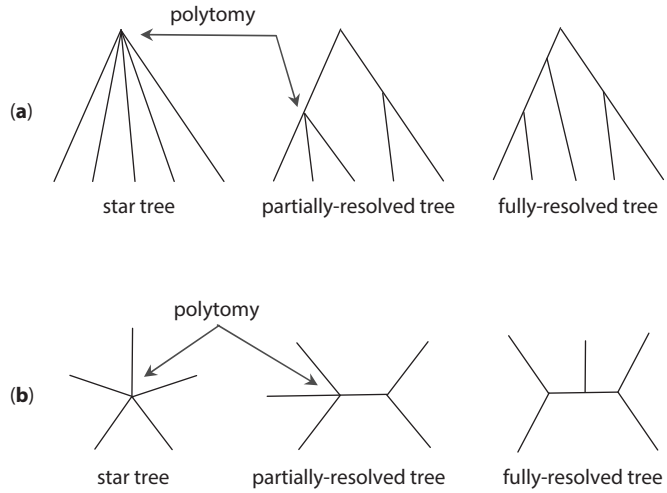


Fig. 3.4 Unresolved and resolved phylogenetic trees. (**a**) Rooted trees. (**b**) Unrooted trees.

3.1.1.4 *Bifurcating and multifurcating trees*

The number of branches connected to a node is called the *degree* of the node. Leaves have a degree of 1. If the root node has a degree greater than 2 or a non-root node has a degree greater than 3, the node represents a *polytomy* or *multifurcation*. A tree with no polytomies is called a *binary tree*, *bifurcating tree*, or *fully resolved tree*. The most extreme unresolved tree is the *star* or *big-bang* tree, in which the root is the only internal node (see Figure 3.4 for example). A polytomy representing truly simultaneous species divergences is sometimes called a *hard polytomy*. It would seem extremely unlikely for one species to diverge into several at exactly the same time, and it may be argued that hard polytomies do not exist. Most often the polytomy represents lack of information in the data to resolve the relationships within a clade (a group of species). Such polytomies are called *soft polytomies*.

3.1.1.5 *The number of trees*

We can work out the total number of unrooted trees by the following *stepwise addition algorithm* (Cavalli-Sforza and Edwards 1967) (Figure 3.5). We start with the single tree for the first three species. This has three branches to which the fourth species can be added. Thus there are three possible trees for the first four species. Each four-species tree has

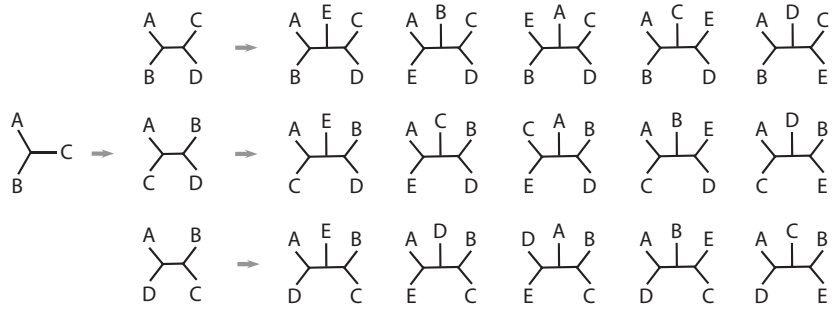


Fig. 3.5 Enumeration of all trees for five taxa A, B, C, D, and E using the stepwise addition algorithm.

five branches, to which the fifth species can be added, resulting in five different five-species trees for each four-species tree. There are thus 3×5 possible trees for five species. In general, a tree of the first $n-1$ species has $(2n-5)$ branches, to which the nth species can be added, so that each of the $(n-1)$-species trees generates $(2n-5)$ distinct n-species trees. Thus the total number of unrooted bifurcating trees for n species is

$$U_n = U_{n-1} \cdot (2n-5) = 3 \cdot 5 \cdot 7 \cdot \cdots \cdot (2n-5) = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot \cdots \cdot (2n-5)}{2 \cdot 4 \cdot \cdots \cdot (2n-6)} = \frac{(2n-5)!}{2^{n-3}(n-3)!}. \tag{3.1}$$

To work out the number of rooted trees for n species, note that each unrooted tree has $(2n-3)$ branches, and the root can be placed on any of those branches, generating $(2n-3)$ rooted trees from each unrooted tree. Thus the number of rooted trees for n species is

$$R_n = U_n \times (2n-3) = U_{n+1} = \frac{(2n-3)!}{2^{n-2}(n-2)!}. \tag{3.2}$$

In certain applications, we also need the concept of *labelled histories*. For example, under the coalescent model, the Yule model of pure birth, or the birth-death process model, all possible labelled histories have the same probability (Aldous 2001). A labelled history is a rooted tree with the internal nodes rank-ordered according to their ages (Edwards 1970). Thus a rooted tree may correspond to several labelled histories. For example, the symmetrical rooted tree for four species $((a, b), (c, d))$ corresponds to two labelled histories, depending on whether or not the common ancestor of a and b is older than the common ancestor of c and d. The asymmetrical rooted tree $(((a, b), c), d)$ corresponds to a single labelled history since there is only one ordering of the internal nodes. The number of possible labelled histories for n sequences is

$$H_n = \frac{n(n-1)}{2} \times \frac{(n-1)(n-2)}{2} \times \cdots \times \frac{2 \cdot 1}{2} = \frac{n!(n-1)!}{2^{n-1}}. \tag{3.3}$$

The counting is done by the so-called coalescent process, which traces the genealogy backwards in time to find common ancestors. Initially there are n lineages, so there are $\binom{n}{2} = \frac{1}{2}n(n-1)$ possible ways of choosing two lineages to join (to coalesce). After the first coalescent event, there will be $n-1$ lineages left so there are $\frac{1}{2}(n-1)(n-2)$ possible ways of choosing two lineages to join, and so on. The last coalescence joins two lineages at the root of the tree. Obviously this coalescent process of joining lineages respects the order of

Table 3.1 The numbers of unrooted trees (U_n), rooted trees (R_n), and labelled histories (H_n) for n species

n	Unrooted trees (U_n)	Rooted trees (R_n)	Labelled histories (H_n)
3	1	3	3
4	3	15	18
5	15	105	180
6	105	945	2,700
7	945	10,395	56,700
8	10,395	135,135	1,587,600
9	135,135	2,027,025	57,153,600
10	2,027,025	34,459,425	2,571,912,000
20	$\sim 2.22 \times 10^{20}$	$\sim 8.20 \times 10^{21}$	$\sim 5.64 \times 10^{29}$
50	$\sim 2.84 \times 10^{74}$	$\sim 2.75 \times 10^{76}$	$\sim 3.29 \times 10^{112}$

coalescent events or the ranking of node ages on the tree; it thus enumerates the labelled histories correctly. The coalescent process is discussed in detail in Chapter 9.

As we can see from Table 3.1 , the number of unrooted trees U_n increases explosively with the number of species n. The number of rooted trees R_n and the number of labelled histories H_n rise even faster.

3.1.1.6 *Distance between trees*

Sometimes we would like to measure how different two trees are. For example, we may be interested in how different the trees estimated from different genes are, or how different the estimated tree is from the true tree in a computer simulation conducted to evaluate a tree reconstruction method. A commonly used measure of topological distance between two trees is the *partition distance* defined by Robinson and Foulds (1981) (see also Penny and Hendy 1985). We give the definition for unrooted trees first. Note that each branch on the tree defines a *bipartition* or *split* of the species; if we chop the branch, the species will fall into two mutually exclusive sets. For example, branch b in tree T_1 of Figure 3.6 partitions the eight species into two sets: (1, 2, 3) and (4, 5, 6, 7, 8). This partition is also present on tree T_2. Partitions defined by terminal branches are in all possible trees and are thus not informative for comparisons between trees. Thus we focus on internal branches only. Partitions defined by branches b, c, d, and e of tree T_1 are the same as partitions defined by branches b', c', d', and e' of tree T_2, respectively. The partition defined by branch a of tree T_1 is not in tree T_2, nor is the partition defined by branch a' of tree T_2 in tree T_1. The partition distance is defined as the total number of bipartitions that exist in one tree but not in the other. Thus T_1 and T_2 have a partition distance of 2. As an unrooted binary tree of n species has $(n-3)$ internal branches, the partition distance ranges from 0 (if the two trees are identical) to $2(n-3)$ (if the two trees do not share any bipartition).

The partition distance can be equivalently defined as the number of contractions and expansions needed to transform one tree into the other. Removing an internal branch by reducing its length to zero is a contraction, while creating an internal branch is an expansion. Trees T_1 and T_2 of Figure 3.6 are separated by a contraction (from T_1 to T_0) and an expansion (from T_0 to T_2), so that their partition distance is 2.

For rooted trees, we use the same definition as for unrooted trees, but imagine the existence of an outgroup species attached to the root. As a rooted binary tree of n species

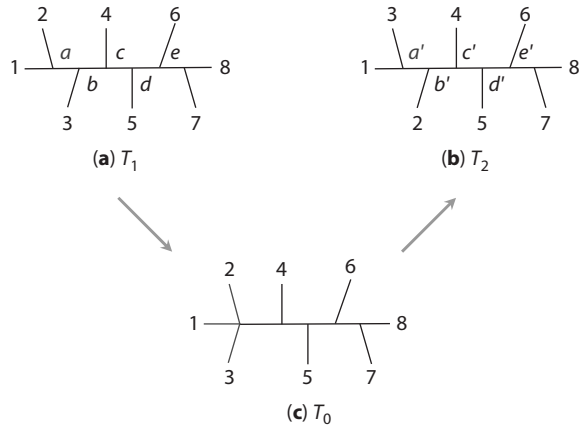


Fig. 3.6 The partition distance between two trees T_1 and T_2 is the total number of bipartitions that are in one tree but not in the other. It is also the number of contractions and expansions needed to change one tree into another. A contraction converts T_1 into T_0 and an expansion converts T_0 into T_2, so the distance between T_0 and T_1 is 1 while the distance between T_1 and T_2 is 2.

has $(n-2)$ internal branches, the partition distance ranges from 0 (if the two trees are identical) to $2(n-2)$ (if the two trees do not share any bipartition).

The partition distance has limitations. First, the distance does not recognize certain similarities between trees. The three trees in Figure 3.7 are identical concerning the relationships among species 2–7 but do not share any bipartitions, so that the partition distance between any two of them is the maximum possible. Indeed, the probability that a random pair of unrooted trees achieve the maximum distance is 70–80% for n = 5–10, and is even greater for larger n. Figure 3.8 shows the distribution of partition distance between two random unrooted trees for the case of n = 10. Second, the partition distance ignores branch lengths in the tree. Intuitively, two trees that are in conflict around short internal branches are less different than two trees that are in conflict around long internal branches. It is unclear how to incorporate branch lengths in a definition of tree distance. One such measure has been suggested by Kuhner and Felsenstein (1994), and is defined as the sum of squared differences between branch lengths in the two trees

$$B_s = \sum\nolimits_i (b_i - b'_i)^2, \tag{3.4}$$

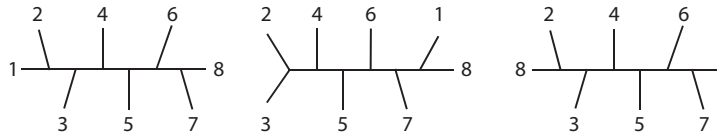


Fig. 3.7 Three trees that do not share any bipartitions and thus achieve the maximum partition distance.

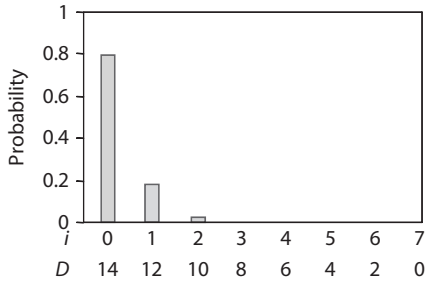


Fig. 3.8 The probability that two random trees from all possible unrooted trees of ten species share i bipartitions or have partition distance D. Note that $D = 2 \times (10 - 3 - i)$.

where b_i and b'_i are branch lengths in the two trees, respectively. If a branch exists in one tree but not in the other, the missing branch has length 0 in the calculation. Third, the partition distance may be misleading if either of the two trees has multifurcations. Suppose we conduct a computer simulation to compare two tree reconstruction methods, using an unrooted binary tree to simulate datasets, and use the partition distance to measure performance: $P = 1 - D/D_{\max}$, where $D_{\max} = 2(n-3)$ is the maximum distance and D is the distance between the true tree and the estimated tree. When both the true tree and the estimated tree are binary, P is the proportion of bipartitions in the true tree that are recovered in the estimated tree. Suppose that with no information in the data, one method returns the star tree as the estimate while the other method returns an arbitrarily resolved binary tree. Now for the first method, $D = (n-3) = 1/2D_{\max}$, so that $P = 50\%$, which may seem very impressive. The second method has a performance of $P = 1/3$ when $n = 4$ or nearly 0 for large n, since a random tree is very unlikely to share any bipartition with the true tree. However, the two methods clearly have the same performance, and the measure based on the partition distance is unreasonable for the first method.

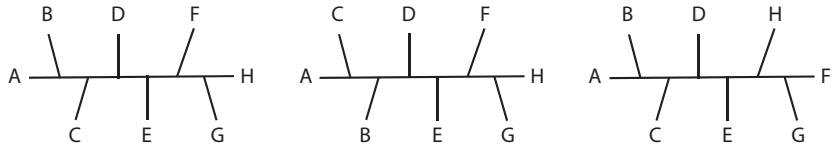
3.1.1.7 *Consensus trees*

While the partition distance measures how different two trees are, a consensus tree summarizes common features among a collection of trees. Many different consensus trees have been defined; see Bryant (2003) for a comprehensive review. Here we introduce two of them.

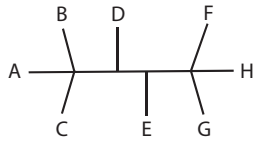
The *strict consensus tree* shows only those branches (partitions or splits) that are shared among all trees in the set, with those not supported by all trees collapsed into polytomies. Consider the three trees in Figure 3.9a. The strict consensus tree is shown in Figure 3.9b. The group (A, B) is in the first and third trees but not in the second, while (A, B, C) is in all three trees. Thus the strict consensus tree shows (A, B, C) as a trichotomy, as well as (F, G, H). The strict consensus tree is a conservative way of summarizing the trees and may not be very useful as it often produces the star tree.

The *majority-rule consensus tree* shows branches or splits that are supported by at least half of the trees in the set. It is also common practice to show the percentage of trees that support every node on the consensus tree (Figure 3.9c). For example, the group (A, B) is in two out of the three trees and is thus shown in the majority-rule consensus tree as resolved, with the percentage of support (2/3) shown next to the node. It is known that all groups that occur in more than half of the trees in the set can be shown on the same consensus tree without generating any conflict.

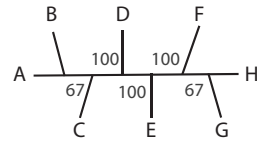
Like the partition distance, the majority-rule consensus tree, as a summary of trees in the set, has limitations. Suppose that there are only three distinct trees in the set, which



(**a**) Three trees for eight species



(**b**) Strict consensus tree



(**c**) Majority-rule consensus tree

Fig. 3.9 Three trees for eight species (**a**) and their strict consensus tree (**b**) and majority-rule consensus tree (**c**).

are the trees of Figure 3.7, each occurring in proportions around 33%. Then the majority-rule consensus tree will be the star tree. In such cases, it appears more informative to report the first few whole trees with the highest support values.

It may be fitting to emphasize here that polytomies in a consensus tree are a heuristic way of summarizing (or visualizing) phylogenetic uncertainties, and do not represent simultaneous speciation events. The consensus tree may thus be unsuitable for use in downstream phylogenetic analyses, such as molecular clock dating, because used in such an analysis, the consensus tree with polytomies is treated as an exact mathematical model of simultaneous speciation. Instead one should use a fully resolved tree inferred from the data, such as the ML tree or the neighbour-joining (NJ) tree.

3.1.1.8 *Monophyly, paraphyly, clade, and clan*

While phylogenetics is concerned with inference of the phylogeny or reconstruction of the evolutionary relationships of the species, classification or taxonomy is the science of describing, naming, and classifying organisms. It is now widely accepted that phylogeny should be the basis of taxonomic classifications. While classification is beyond the scope of this book, some terms are commonly used in molecular phylogenetic analysis, and will be described here.

A monophyletic group includes all the descendants of a common ancestor. Such a group is also called a *clade*. Taxa in a monophyletic group are more *closely related*; i.e. they have more recent common ancestors than those outside the group. We also use the term *sister species* when two species are each other's closest relatives. A group that includes some descendants of a common ancestor but excludes some others is *non-monophyletic*. Some authors distinguish two types of non-monophyly: paraphyly and polyphyly. A *paraphyletic* group does not include all of the descendants of a single common ancestor. For example, 'apes' include chimpanzees, gorillas, orangutans, and gibbons but exclude humans. Apes are thus a paraphyletic group. Similarly 'reptiles' are a paraphyletic group (Figure 3.10). Reptiles include crocodiles, lizards, and turtles but exclude birds (because birds have novel anatomy and behaviour), but crocodiles are more closely related to birds than they are to other reptiles. A *polyphyletic* group includes species that have multiple common ancestors but excludes some other descendants of those ancestors so that the last common ancestor

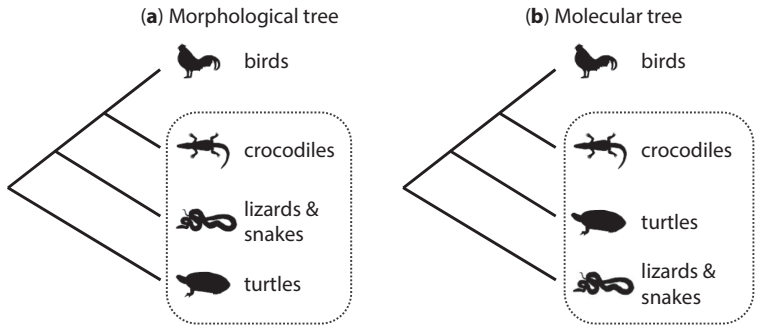


Fig. 3.10 Traditional morphological tree (**a**) and recent molecular tree (**b**) for birds, crocodiles, lizards, and turtles (e.g. Iwabe et al. 2005; Crawford et al. 2012). No matter which of those two trees is correct, 'reptiles' (circled) are a paraphyletic group.

of the group is not a member of the group. A polyphyletic group is often the result of erroneous taxonomic classification based on morphological similarities that are the result of convergent evolution. For example, 'pachyderms' include elephants, rhinoceroses, hippopotamuses, etc. Those mammals all have thick skins but belong to different orders and are a polyphyletic group. 'Vultures' are another polyphyletic group. The Old and New World vultures have striking similarities (such as bald heads) due to convergent evolution. However, the Old World vultures evolved from birds of prey (such as eagles, kites, hawks) while the New World vultures evolved from storks.

Use of terms such as monophyly and clade implies a knowledge of the root of the tree. As most phylogeny reconstruction methods produce unrooted trees, those terms are sometimes applied to unrooted trees as well, with the assumption that the root is in a place such that the use of those terms would be sensible. Wilkinson et al. (2007) recommend the use of the term *clan* (instead of clade) to mean a group of species identified on an unrooted tree. When we cut an internal branch on an unrooted tree, the species will fall into two groups (partitions or splits). These are two clans, and one of them must be a clade.

3.1.2 *Species trees and gene trees*

The phylogeny representing the relationships among a group of species is called the *species tree* or *organismal* tree. The phylogeny for sequences at a particular gene locus from those species is called the *gene tree*. A number of factors may cause the gene tree to differ from the species tree.

First, phylogeny reconstruction errors may cause the estimated gene tree to be different from the species tree even if the true gene tree agrees with the species tree. The estimation errors may be either random, due to the limited amount of sequence data, or systematic, due to deficiencies of the tree reconstruction method or serious violations of its assumptions. One such case is convergent evolution. For example, the lysozyme has apparently undergone convergent evolution in ruminants (e.g. the cow) and the leaf-eating colobine monkeys (e.g. the langur), as it is recruited as a bacteriolytic enzyme in the fermentative foreguts of those animals (Stewart et al. 1987). As a result, the stomach lysozymes of mammals from those two groups share some physico-chemical and catalytical

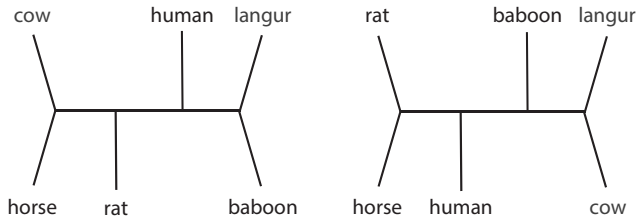


Fig. 3.11 Convergent evolution in the stomach lysozyme of the cow and the langur misleads phylogeny reconstruction methods to infer an incorrect gene tree. The organismal tree for six species of mammals is the one in (**a**), and this should also be the true gene tree for the lysozyme. However, the parsimony (and ML) methods incorrectly infer the gene tree to be the one in (**b**), grouping the cow and the langur together. Drawn following Stewart et al. (1987).

properties as well as certain key amino acids. When the protein sequences are used in tree reconstruction, an incorrect tree is inferred, grouping the cow and the langur together (Figure 3.11).

Second, during the early stages of evolution near the root of the universal tree of life, there appears to have been substantial lateral (horizontal) gene transfer (LGT). As a result, different genes or proteins may have different gene trees, in conflict with the species tree. The LGT appears to be so extensive that some researchers question the concept of a universal tree of life (see, e.g. Doolittle 1998). Third, gene duplications, especially if followed by gene losses, can cause the gene tree to be different from the species tree if paralogous copies of the gene are used for phylogeny reconstruction (Figure 3.12a). Note that

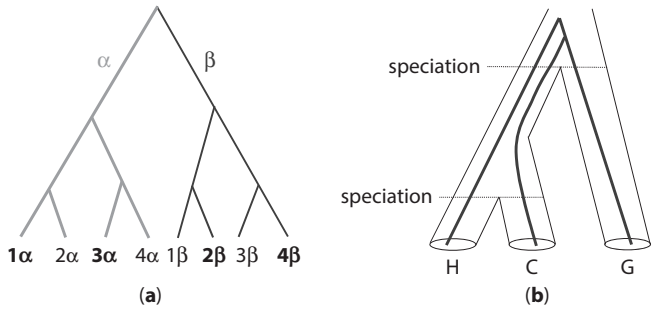


Fig. 3.12 Conflict between species tree and gene tree can be due to gene duplication (**a**) or ancestral polymorphism (**b**). In (**a**), a gene duplicated in the past, creating paralogous copies α and β, followed by divergences of species 1, 2, 3, and 4. If we use gene sequences 1α, 3α, 2β, 4β for phylogeny reconstruction, the true gene tree is $((1\alpha, 3\alpha), (2\beta, 4\beta))$, different from the species tree $((1, 2), (3, 4))$. In (**b**), the species tree is ((human, chimpanzee), gorilla). However, due to ancestral polymorphism and incomplete lineage sorting, the true gene tree is (human, (chimpanzee, gorilla)).

paralogues are genes that originated from gene duplications and may not reflect species relationships, while orthologues are genes that originated from species divergences and thus track speciation events (Fitch 1970). Fourth, *ancestral polymorphism* or polymorphism in ancestral species means that gene sequences in different modern species may be descendants of different ancestral sequences, so that the gene genealogy may fail to track the species phylogeny, a phenomenon called *incomplete lineage sorting*. An example is shown in Figure 3.12b. Here the species tree for human, chimpanzee, and gorilla is ((H, C), G). However, because of sequence variations (polymorphisms) in the extinct ancestral species, the true gene tree is (H, (C, G)). The probability that the gene tree differs from the species tree is greater if the speciation events are closer in time (i.e. if the species tree is almost a star tree) and if the population size of the H-C common ancestor is greater. Such information concerning the gene tree-species tree conflict can be used to estimate the effective population sizes of extinct common ancestors and to infer phylogeographic processes. We will discuss modern computational approaches to such inference using multiple-locus sequence data later, in Chapter 9.

3.1.3 *Classification of tree reconstruction methods*

Here we consider some overall features of phylogeny reconstruction methods. First, some methods are *distance based*. In those methods, distances are calculated from pairwise comparisons of sequences, and the resulting distance matrix is used in subsequent analysis. A cluster algorithm is often used to convert the distance matrix into a phylogenetic tree (Everitt et al. 2001). The most popular methods in this category include UPGMA (Unweighted Pair-Group Method using Arithmetic Averages, Sneath 1962) and NJ (neighbour-joining, Saitou and Nei 1987). Other methods are *character based*, which attempt to fit the characters (nucleotides or amino acids, say) observed in all species at every site to a tree. Maximum parsimony (Fitch 1971b; Hartigan 1973), ML (Felsenstein 1981), and Bayesian methods (Rannala and Yang 1996; Mau and Newton 1997; Li et al. 2000) are all character based. Distance methods are often computationally faster than character-based methods, and can be easily applied to analyse different kinds of data as long as pairwise distances can be calculated.

Tree reconstruction methods can also be classified as being either *algorithmic* (cluster methods) or *optimality* based (search methods). The former include UPGMA and NJ, which use cluster algorithms to arrive at a single tree from the data as the best estimate of the true tree. Optimality-based methods use an optimality criterion (objective function) to measure a tree's fit to data, and the tree with the optimal score is the estimate of the true tree (Table 3.2). In the maximum parsimony method, the tree score is the minimum number of character changes required for the tree, and the *maximum parsimony tree* or *most parsimonious tree* is the tree with the smallest tree score. The ML method uses the log

Table 3.2 Optimality criteria used for phylogeny reconstruction

Method	Criterion (tree score)
Maximum parsimony	Minimum number of changes, minimized over ancestral states
Maximum likelihood	Log likelihood score, optimized over branch lengths and model parameters
Minimum evolution	Tree length (sum of branch lengths, often estimated by least squares)
Bayesian	Posterior probability, calculated by integrating over branch lengths and substitution parameters

likelihood value of the tree to measure the fit of the tree to the data, and the *maximum likelihood tree* is the tree with the highest log likelihood value. In the Bayesian method, the posterior probability of a tree is the probability that the tree is true given the data. The tree with the maximum posterior probability is the estimate of the true tree, known as the maximum *a posteriori* (MAP) *tree*. In theory, methods based on optimality criteria have to solve two problems: (i) calculation of the criterion (tree score) for a given tree and (ii) search in the space of all trees to identify the tree with the best score. The first problem can be expensive if the tree is large, but the second is much worse when the number of sequences is greater than 20 or 50 because of the huge number of possible trees. As a result, heuristic algorithms are used for tree searches. Optimality-based search methods are usually much slower than algorithmic cluster methods.

Some tree reconstruction methods are model based. Distance methods use nucleotide or amino acid substitution models to calculate pairwise distances. Likelihood and Bayesian methods use substitution models to calculate the likelihood function. These methods are clearly model based. Parsimony does not make explicit assumptions about the evolutionary process. Opinions differ as to whether the method makes any implicit assumptions, and, if so, what they are. We will return to this issue in Chapter 5.

3.2 Exhaustive and heuristic tree search

3.2.1 *Exhaustive tree search*

For parsimony and likelihood methods of tree reconstruction, which evaluate trees according to an optimality criterion, one should in theory calculate the score for every possible tree and then identify the tree having the best score. Such a strategy is known as *exhaustive search* and is guaranteed to find the best tree. As mentioned above, the stepwise addition algorithm provides a way of enumerating all possible trees for a fixed number of species (Figure 3.5).

An exhaustive search is, however, computationally unfeasible except for small datasets with, say, fewer than ten taxa. For the parsimony method, a branch-and-bound algorithm has been developed to speed up the exhaustive search (Hendy and Penny 1982). Even so, the computation is feasible for small datasets only. For the likelihood method, such an algorithm is not available. Thus computer programs use heuristic algorithms to search in the tree space, which are not guaranteed to find the optimal tree.

3.2.2 *Heuristic tree search*

Heuristic search algorithms may be grouped into two categories. The first includes hierarchical *cluster algorithms*. These may be subdivided into *agglomerative* methods, which proceed by successive fusions of the n species into groups, and *divisive* methods, which separate the n species successively into finer groups (Everitt et al. 2001). Whether each step involves a fusion or fission, the algorithm involves choosing one out of many alternatives, and the optimality criterion is used to make that choice. The second category of heuristic tree search algorithms includes *tree-rearrangement* or *branch-swapping* algorithms. They propose new trees through local perturbations to the current tree, and the optimality criterion is used to decide whether or not to move to a new tree. The procedure is repeated until no improvement can be made in the tree score. We describe two cluster algorithms in this subsection and a few branch-swapping algorithms in the next.

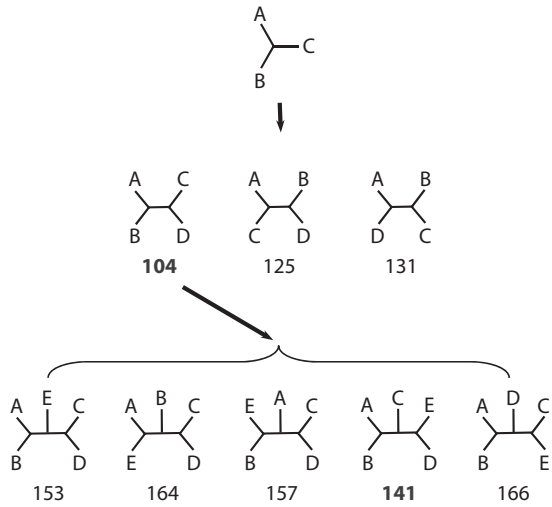


Fig. 3.13 Stepwise addition algorithm under the maximum parsimony criterion. The tree score is the minimum number of changes required by the tree.

Stepwise addition or *sequential addition* is an agglomerative algorithm. It adds sequences one by one, until all sequences are in the tree. When each new sequence is added, all the possible locations are evaluated and the best is chosen using the optimality criterion. Figure 3.13 illustrates the algorithm for the case of five sequences, using the parsimony score as the optimality criterion. Note that this algorithm of heuristic tree search is different from the stepwise addition algorithm for enumerating all possible trees explained in Figure 3.5. In the heuristic search, the locally best subtree is selected at each step, and trees that can be generated from the suboptimal subtrees are ignored. In our example, the ten five-species trees on the second and third rows of Figure 3.5 are never visited in the heuristic search. Thus the algorithm is not guaranteed to find the globally optimal tree. It is less clear whether one should add the most similar sequences or the most divergent sequences first. A common practice is to run the algorithm multiple times, adding sequences in a random order.

Star decomposition is a divisive cluster algorithm. It starts from the star tree of all species, and proceeds to resolve the polytomies by joining two taxa at each step. From the initial star tree of n species, there are $n(n-1)/2$ possible pairs, and the pair that results in the greatest improvement in the tree score is grouped together. The root of the tree then becomes a polytomy with $(n-1)$ taxa. Every step of the algorithm reduces the number of taxa connected to the root by one. The procedure is repeated until the tree is fully resolved. Figure 3.14 shows an example of five sequences, using the log likelihood score for tree selection.

For n species, the stepwise addition algorithm evaluates three trees of four species, five trees of five species, seven trees of six species, with a total of $3+5+7+\cdots+(2n-5) = (n-1)(n-3)$ trees. In contrast, the star decomposition algorithm evaluates $n(n-1)/2 + (n-1)(n-2)/2+\cdots+3 = \frac{1}{6}n(n^2-1)-7$ trees in total, all of which are for n species. Thus for $n > 4$, the star decomposition algorithm evaluates many more and bigger trees than the stepwise addition algorithm and is expected to be much slower. The scores for trees constructed during different stages of the stepwise addition algorithm are not directly

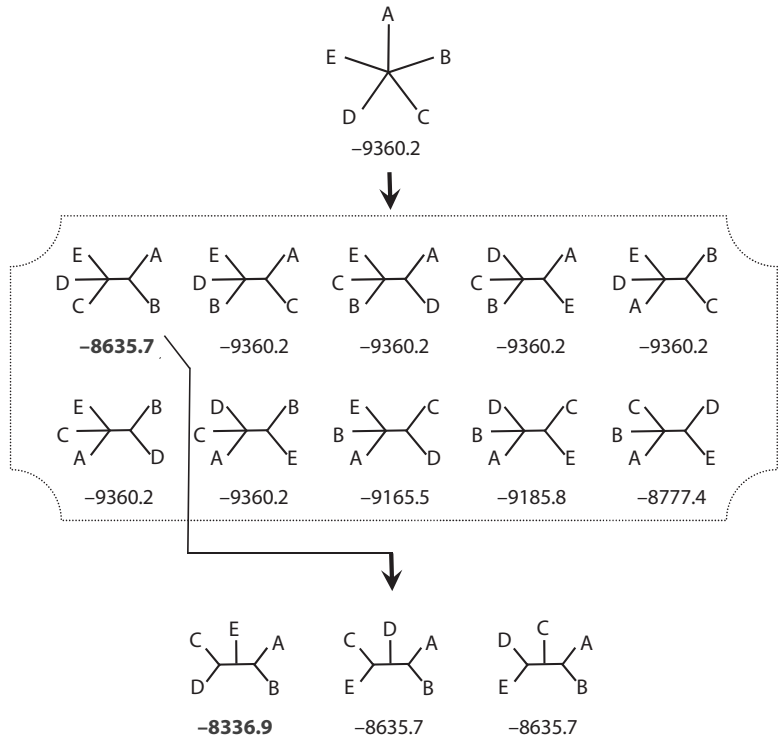


Fig. 3.14 Star decomposition algorithm under the likelihood criterion. The tree score is the log likelihood value calculated by optimizing branch lengths on the tree.

comparable as the trees are of different sizes. Trees evaluated in the star decomposition algorithm are all of the same size and their tree scores are comparable.

Both the stepwise addition and star decomposition algorithms produce resolved trees of all *n* species. If we stop at the end of either algorithm, we have an algorithmic cluster method for tree reconstruction based on the optimality criterion. However, in most programs, trees generated from these algorithms are treated as starting trees and subjected to local rearrangements. Below are a few such algorithms.

3.2.3 *Branch swapping*

Branch swapping or tree rearrangements are heuristic algorithms of hill climbing in the tree space. An initial tree is used to start the process. This can be a random tree, or a tree produced by stepwise addition or star decomposition algorithms, or by other faster tree reconstruction methods such as NJ. The branch-swapping algorithm generates a collection of neighbour trees around the current tree. The optimality criterion is then used to decide which neighbour to move to. The branch-swapping algorithm affects our chance of finding the best tree and the amount of computation it takes to do so. If the algorithm generates too many neighbours, each step will require evaluation of too many candidate trees. If the algorithm generates too few neighbours, we do not have to evaluate many trees at each step, but there may be many local peaks in the tree space (see below), and the search can easily get stuck at a local peak.

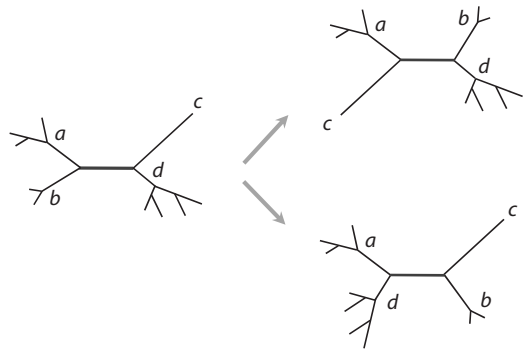


Fig. 3.15 The NNI algorithm. Each internal branch in the tree connects four subtrees or nearest neighbours (*a*, *b*, *c*, *d*). Interchanging a subtree on one side of the branch with another on the other side constitutes an NNI. Two such rearrangements are possible for each internal branch.

Nearest neighbour interchange (NNI). Each internal branch defines a relationship among four subtrees, say, *a*, *b*, *c*, and *d* (Figure 3.15). Suppose the current tree is $((a, b), c, d)$ and the two alternative trees are $((a, c), b, d)$ and $((a, d), b, c)$. The NNI algorithm allows us to move from the current tree to the two alternative trees, by swapping a subtree on one side of the branch with a subtree on the other side. An unrooted tree for n species has $n - 3$ internal branches. The NNI algorithm thus generates $2(n - 3)$ immediate neighbours. The neighbourhood relationships among the 15 trees for five species are illustrated in Figure 3.17.

Two other commonly used algorithms are *subtree pruning and regrafting* (SPR) and *tree bisection and reconnection* (TBR) (Swofford et al. 1996). In the former, a subtree is pruned and then reattached to a different location on the tree (Figure 3.16a). In the latter, the tree

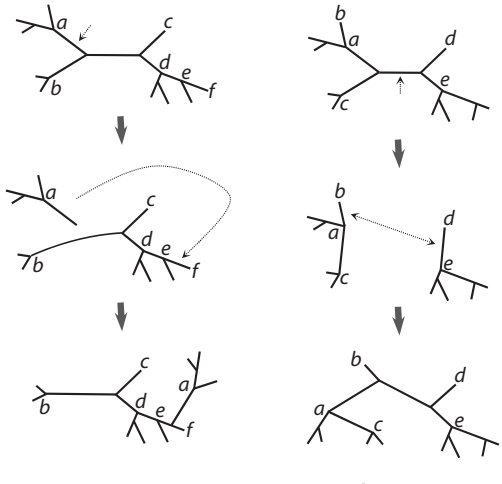


Fig. 3.16 (**a**) Branch swapping by SPR. A subtree (for example, the one represented by node *a*) is pruned, and then reattached to a different location on the tree. (**b**) Branch swapping by TBR. The tree is broken into two subtrees by cutting an internal branch. Two branches, one from each subtree, are then chosen and rejoined to form a new tree.

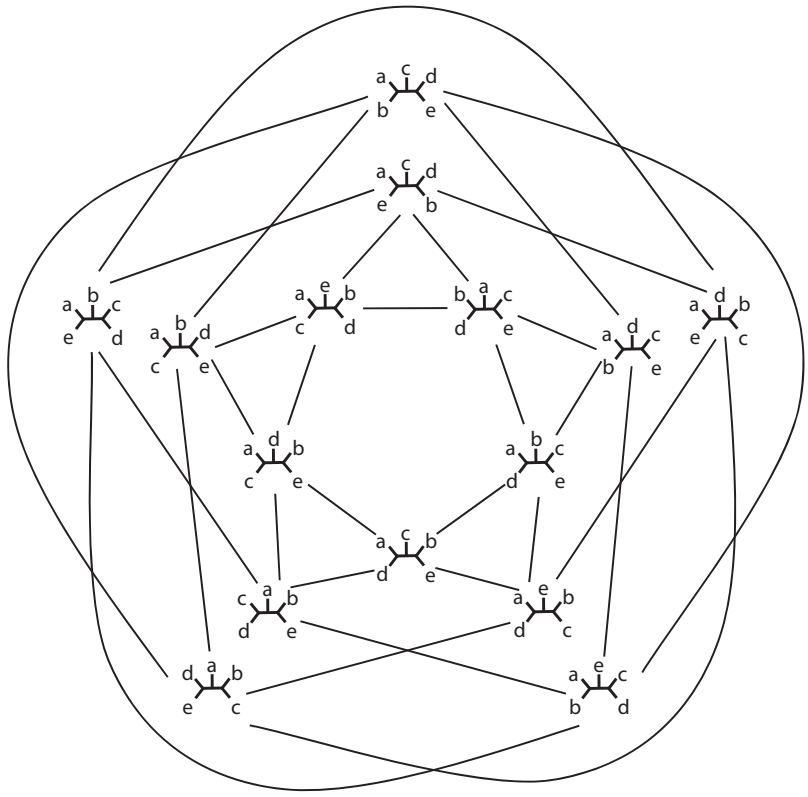


Fig. 3.17 The 15 trees for five species, with neighbourhood relationships defined by the NNI algorithm. Trees that are neighbours under NNI are connected. Note that this visually appealing representation has the drawback that trees close by may not be neighbours. Drawn following Felsenstein (2004).

is cut into two parts by chopping an internal branch and then two branches, one from each subtree, are chosen and rejoined to form a new tree (Figure 3.16b). TBR generates more neighbours than SPR, which in turn generates more neighbours than NNI.

3.2.4 *Local peaks in the tree space*

Maddison (1991) and Charleston (1995) discussed local peaks or tree islands in the tree space. Figure 3.18 shows an example for five species and 15 trees. The neighbourhood relationship is defined using the NNI algorithm (see Figure 3.17). Each tree has four neighbours, while the ten other trees are two NNI steps away. The parsimony tree lengths for the two trees on the top of the graph, T_1 and T_2, are 1366 and 1362. T_1 is the best tree by the likelihood and Bayesian methods, while T_2 is the most parsimonious tree. Other trees are much worse than those two trees by both the likelihood and parsimony criteria. The eight trees that are neighbours of T_1 or T_2 have tree lengths ranging from 1406 to 1438, while the five trees that are two steps away from T_1 and T_2, have tree lengths ranging from 1488 to 1500. Trees T_1 and T_2 are separated from each other by other trees of much poorer scores and are thus local peaks. They are local peaks for the SPR and TBR algorithms as well. Also T_1 and T_2 are local peaks when the data are analysed under

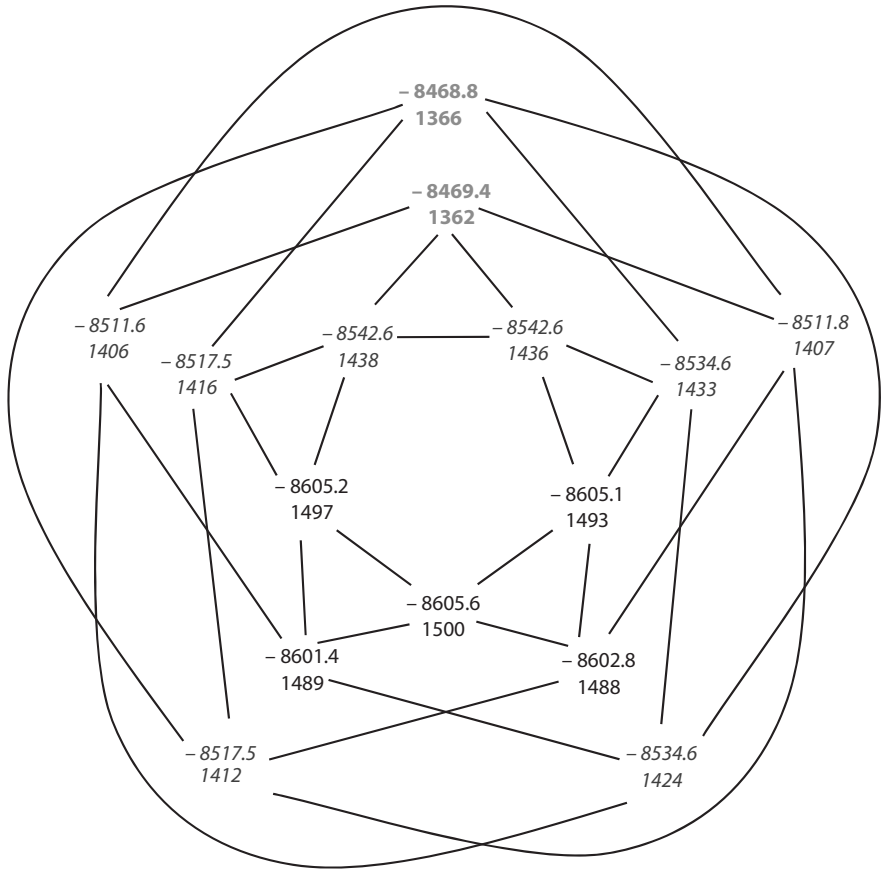


Fig. 3.18 Local peaks in the tree space. The log likelihood values (above) and parsimony scores (below) for the 15 trees of Figure 3.17, shown in the same locations. The dataset was simulated following the construction of Mossel and Vigoda (2005). It consists of 2,000 nucleotide sites simulated under JC69 using the top two trees in Figure 3.17: T_1: ((*a*, *b*), *c*, (*d*, *e*)) and T_2: ((*a*, *e*), *c*, (*d*, *b*)), with 1,000 sites from each tree. All branch lengths are fixed at 0.1. Trees T_1 and T_2 are two local optima under both parsimony and likelihood criteria. The posterior probabilities for T_1 and T_2 are ~0.64 and ~0.36, respectively.

ML. Indeed for this dataset, the rank order of the 15 trees is almost identical under the likelihood and parsimony criteria. Similarly the dataset may pose serious computational problems for Bayesian Markov chain Monte Carlo algorithms, as discussed by Mossel and Vigoda (2005).

One can design a branch-swapping algorithm under which trees T_1 and T_2 are neighbours. However, such an algorithm will define a different neighbourhood relationship among trees, and may have different local peaks or may have local peaks for different datasets. The problem should be more serious for larger trees with more species, as the tree space is much larger. Similarly, in larger sequence datasets with more sites, the peaks tend to be higher and the valleys deeper, making it very difficult to traverse between peaks (Salter 2001).

3.2.5 *Stochastic tree search*

An optimization algorithm that always goes uphill may get stuck at a local peak. Some algorithms attempt to overcome the problem of local peaks by allowing downhill moves. They can work under either parsimony or likelihood criteria.

The first such algorithm is *simulated annealing* (Metropolis et al. 1953; Kirkpatrick et al. 1983). This is inspired by annealing in metallurgy, a technique involving heating and controlled cooling of a metal to reduce defects. The heat causes the atoms to move at random, exploring various configurations, while the slow cooling allows them to find configurations with low internal energy. In a simulated annealing algorithm of optimization, the objective function is modified to have a flattened surface during the early (heating) stage of the search, making it easy for the algorithm to move between peaks. At this stage downhill moves may be accepted nearly as often as uphill moves. The 'temperature' is gradually reduced as the simulation proceeds, according to some 'annealing schedule'. At the final stage of the algorithm, only uphill moves are accepted, as in a greedy algorithm. Simulated annealing algorithms are highly specific to the problem, and their implementation is more art than science. The efficiency of the algorithm is affected by the neighbourhood function (branch-swapping algorithms) and the annealing schedule. Implementations in phylogenetics include Goloboff (1999) and Barker (2004) for parsimony, and Salter and Pearl (2001) for likelihood. Fleissner et al. (2005) used simulated annealing for simultaneous sequence alignment and phylogeny reconstruction.

A second stochastic tree search algorithm is the *genetic algorithm*. A 'population' of trees is kept in every generation; these are allowed to 'breed' to produce trees of the next generation. The algorithm uses operations that are similar to mutation and recombination in genetics to generate new trees from the current ones. The 'survival' of each tree into the next generation depends on its 'fitness', which is the optimality criterion. Lewis (1998), Katoh et al. (2001), and Lemmon and Milinkovitch (2002), among others, have implemented genetic algorithms to search for the ML tree.

A third stochastic tree search algorithm is the Bayesian Markov chain Monte Carlo (MCMC) algorithm. This is a statistical approach and produces not only a point estimate (the tree with the highest likelihood or posterior probability) but also a measure of uncertainty in the point estimate through posterior probabilities estimated during the search. While MCMC algorithms allow downhill as well as uphill moves, high peaks and deep valleys in the search space can cause serious computational problems. In this regard, we note that both simulated annealing and genetic algorithms have been used to design advanced MCMC algorithms for Bayesian computation. We will discuss Bayesian phylogenetic methods in Chapters 7 and 8.

3.3 Distance matrix methods

Distance methods of phylogeny reconstruction involve two steps: (i) calculation of the distance between every pair of species and (ii) reconstruction of a phylogenetic tree from the distance matrix. The first step has been discussed in Chapters 1 (for nucleotide sequence data) and 2 (for amino acid and codon sequence data). Here we discuss the second step. We describe two optimality-based methods (least-squares and minimum evolution) and one cluster algorithm (neighbour-joining). All distance methods treat the matrix of pairwise distances as observed data. Some of them in addition make use of the variances (and even the covariances) of the estimated distances. After the distance matrix is calculated, the original sequence alignment is no longer used.

3.3.1 *Least-squares method*

The least-squares (LS) method takes the pairwise distances as observed data and estimates branch lengths on any given tree by trying to match those distances as closely as possible, i.e. by minimizing the sum of squared differences between the observed and expected distances. The expected distance between two species is calculated as the sum of branch lengths along the path on the tree connecting the two species. The minimum sum of squared differences achieved on the tree then measures the fit of the tree to the distance data and is used as the tree score. The tree with the best (least) score is the LS tree, which is the estimate of the true tree. This method was proposed by Cavalli-Sforza and Edwards (1967; see also Edwards and Cavalli-Sforza 1963b), who called it the *additive-tree method*.

More formally, let the observed (calculated) distance between species i and j be d_{ij} and the expected distance be δ_{ij}, which is the sum of branch lengths along the path from species i to j on the tree. Their difference is the error $e_{ij} = d_{ij} - \delta_{ij}$. The closer the errors are to zero, the better the tree and branch lengths fit the data. The LS method estimates the branch lengths by minimizing the sum of the squared errors:

$$S = \sum_{i<j} (d_{ij} - \delta_{ij})^2. \tag{3.5}$$

For example, the pairwise distances (d_{ij}) calculated under the K80 model for the mitochondrial data of Brown et al. (1982) are shown in Table 3.3. These are taken as observed data. Now consider the tree ((human, chimpanzee), gorilla, orangutan), with its five branch lengths $t_0, t_1, t_2, t_3,$ and t_4 (Figure 3.19). The expected distances in the tree are thus $\delta_{12} = t_1 + t_2$ between the human and the chimpanzee, $\delta_{13} = t_1 + t_0 + t_3$ between the human and the gorilla, and so on. The sum of squared differences is then

$$S = (d_{12} - \delta_{12})^2 + (d_{13} - \delta_{13})^2 + (d_{14} - \delta_{14})^2 + (d_{23} - \delta_{23})^2 + (d_{24} - \delta_{24})^2 + (d_{34} - \delta_{34})^2. \tag{3.6}$$

In this setup, the distances (d_{ij}) are the observed data and the δs (or more precisely, the five branch lengths $t_0, t_1, t_2, t_3,$ and t_4) are the unknown parameters to be estimated. The

Table 3.3 Pairwise distances for the mitochondrial DNA sequences

1. Human				
2. Chimpanzee	0.0965			
3. Gorilla	0.1140	0.1180		
4. Orangutan	0.1849	0.2009	0.1947	
	1. Human	2. Chimpanzee	3. Gorilla	4. Orangutan

Note: The distance matrix is symmetrical so that only the lower-triangular part is shown. The diagonals are zero.

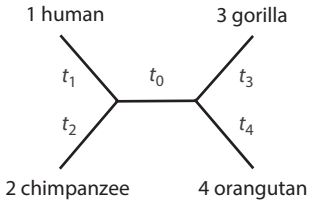


Fig. 3.19 A tree to demonstrate the LS criterion for estimating branch lengths.

Table 3.4 LS branch lengths under K80 for the distance matrix of Table 3.3

Tree	t_0 for internal branch	t_1 for H	t_2 for C	t_3 for G	t_4 for O	S_j
τ_1: ((**H**, **C**), **G**, **O**)	**0.008840**	**0.043266**	**0.053280**	**0.058908**	**0.135795**	**0.000035**
τ_2: ((H, G), C, O)	0.000000	0.046212	0.056227	0.061854	0.138742	0.000140
τ_3: ((C, G), H, O)						
τ_0: (H, C, G, O)						

values of branch lengths that minimize S are the LS estimates. These can be found numerically to be $\hat{t}_0 = 0.008840$, $\hat{t}_1 = 0.043266$, $\hat{t}_2 = 0.053280$, $\hat{t}_3 = 0.058908$, and $\hat{t}_4 = 0.135795$, with the corresponding tree score $S = 0.00003547$. Similar calculations can be done for the other two trees. Indeed, the other two binary trees both converge to the star tree, with the internal branch length estimated to be 0; see Table 3.4. Here we assumed that the branch lengths are estimated under the nonnegative constraint. The tree ((human, chimpanzee), gorilla, orangutan) has the smallest S and is called the LS tree. It is the LS estimate of the true phylogeny. Note that two optimizations are involved: the optimization of branch lengths to calculate the tree score S, and the search in the space of trees for the one with the best tree score, i.e. the LS tree.

Estimation of branch lengths on a fixed tree by the LS criterion uses the same principle as calculating the line of best fit $y = a + bx$ on a scatter plot. If there are no constraints on the branch lengths, the solution is analytical and can be obtained by solving a set of linear equations (Cavalli-Sforza and Edwards 1967). Efficient algorithms that require less computation and less space have also been developed by Rzhetsky and Nei (1993) and Bryant and Waddell (1998). Those algorithms may produce negative branch lengths, which are not meaningful biologically. If the branch lengths are constrained to be nonnegative (as in the above example), the problem becomes one of constrained optimization, which is expensive. However, if we ignore the interpretation of branch lengths, the unconstrained LS is at least consistent: when more and more data are available and the distances approach their true values, the LS tree converges to the true tree. Simulation studies suggest that constraining branch lengths to be nonnegative leads to improved performance in tree reconstruction (e.g. Kuhner and Felsenstein 1994; Gascuel 1997). However, most computer programs implement the LS method without the constraint. It is noted that when the estimated branch lengths are negative, they are most often close to zero.

The LS method described above (i.e. the criterion S of equation (3.5)) uses equal weights for the different pairwise distances and is known as the ordinary least squares (OLS). As in the case of fitting a straight line to a scatter plot, OLS is based on the assumptions that the errors are independent and have equal variance, or equivalently that the (observed) distances are independent and have equal variance. These assumptions are incorrect in the case of pairwise distances. First, larger distances tend to have larger variances. Second, the distances may be correlated because they share branch lengths on the tree. For example, in the tree of Figure 3.19 the distances d_{12} and d_{13} involve the same branch length t_1 so that they both tend to be larger if t_1 is larger: indeed d_{12} and d_{13} have a positive covariance that is equal to the variance of branch length t_1 (Nei and Jin 1989).

The standard approach to dealing with unequal variances is weighted least squares (WLS), which minimizes

$$S = \sum_{i<j} w_{ij}(d_{ij} - \delta_{ij})^2, \tag{3.7}$$

where the weight $w_{ij} = 1/\text{var}(d_{ij})$ (Bulmer 1990). In the method of Fitch and Margoliash (1967), $w_{ij} = 1/d_{ij}^2$ is used. Note that OLS is a special case of WLS with $w_{ij} = 1$. A further extension to WLS that accommodates the correlations (covariances) between the distances, as well as the unequal variances, is the generalized least squares (GLS). While computer simulations suggest that WLS works better than OLS in tree reconstruction, WLS, and especially GLS, involve more computation and are not commonly used.

3.3.2 *Minimum evolution method*

In the LS method discussed above, the LS criterion is used both to estimate the branch lengths on a given tree and to search for the best tree in the tree space. The minimum S for a tree achieved by optimizing its branch lengths is the score for that tree, and at least in theory all possible trees should be compared to find the one with the best score, the LS tree.

In the minimum evolution (ME) method, the LS criterion is usually used to estimate the branch lengths, but tree selection relies on the sum of branch lengths (the tree length). This is based on the plausible but heuristic idea that the true tree is most likely to be the one that involves the minimum amount of evolutionary change. A number of researchers had the same idea at about the same time, including Edwards and Cavalli-Sforza (1963a), Camin and Sokal (1965), and Eck and Dayhoff (1966) (see Edwards 1996, 2009a; Felsenstein 2004). In their analysis of blood group allele frequencies to reconstruct the human population relationships, Edwards and Cavalli-Sforza (1963a) arrived at the *principle of minimum evolution*, which states that 'The most plausible estimate of the evolutionary tree is that which invokes the minimum net amount of evolution'. While the word 'principle' was used, it was intended from the start to be an approximation to the ML method (Edwards 1996). Also it was intended to apply to both continuous and discrete characters. For discrete characters, the amount of evolutionary change should be the minimum number of character changes; so this ME method is now known as parsimony (Camin and Sokal 1965) (see §3.4). For distance data, the amount of evolutionary change is the sum of branch lengths on the tree. Phylogeny reconstruction under the ME criterion based on distances is studied in detail by Kidd and Sgaramella-Zonta (1971) and Rzhetsky and Nei (1993). Gascuel et al. (2001) and Desper and Gascuel (2005) provided excellent reviews of this class of methods.

Variations exist in the practical implementation of the ME principle. First, branch lengths are usually estimated using LS, but as discussed above, variations exist concerning whether the variances and covariances of the observed distances are taken into account and whether the branch lengths are optimized under the nonnegative constraint. Second, several definitions of the tree length exist, differing in their treatment of negative branch lengths. Gascuel et al. (2001) analysed the consistency properties of those variations, and the results are summarized in Table 3.5. A further definition (or estimation method) of tree length is described by Pauplin (2000). This will be described in the next subsection in our discussion of the NJ method.

3.3.3 *Neighbour-joining method*

The simplest distance method is perhaps UPGMA (Sneath 1962). This is a cluster algorithm based on the molecular clock assumption and generates rooted trees. It is thus applicable to population data or closely related species but is not suitable for inferring species phylogenies in general, as the clock is often violated when the sequences are divergent.

Table 3.5 Consistency status of minimum evolution method for phylogeny reconstruction

	Tree length			
Method	All-BL	Positive-BL	Absolute-BL	Nonnegative-BL
Ordinary LS	Consistent	Consistent	Consistent	Unknown
Weighted LS	Inconsistent	Inconsistent	Inconsistent	Inconsistent
Generalized LS	Inconsistent	Inconsistent	Inconsistent	Inconsistent

Note: Branch lengths are estimated using Ordinary LS, Weighted LS, or Generalized LS. They are then summed to give the tree length, which is minimized according to the ME criterion. All-BL means that the tree length is calculated as the sum of all branch lengths (both positive and negative) (Rzhetsky and Nei 1993); Positive-BL means the sum of the positive branch lengths, ignoring the negative ones (Swofford and Olsen 1990); Absolute-BL means the sum of the absolute values of the branch lengths (Kidd and Sgaramella-Zonta 1971); and Nonnegative-BL means the sum of the (nonnegative) branch lengths estimated under the nonnegative constraint. For any of the method for estimating branch lengths and for calculating tree length, the ME method selects the shortest tree as being the estimate of the true phylogeny. From Gascuel et al. (2001).

Here we discuss NJ, which is a divisive cluster algorithm proposed by Saitou and Nei (1987). See §3.2.2 for a discussion of divisive and agglomerative cluster algorithms. NJ does not require the clock assumption and produces unrooted trees. It is widely used because it is computationally fast, produces reasonable trees, and has easy-to-use software implementations (Tamura et al. 2011). It starts with a star tree and then chooses a pair of nodes (neighbours) to join (Figure 3.20). A new node is then created to replace the two joined nodes, reducing the number of nodes connected to the root by one and reducing the dimension of the distance matrix by one. The procedure is repeated until the tree is fully resolved. The branch lengths are updated during every step of the algorithm.

Suppose at the current stage of the algorithm, there are r nodes connected to the root (node o). Out of the $r(r-1)/2$ possible pairs of nodes, the pair that minimizes the following Q criterion is chosen to be neighbours for joining:

$$Q_{ij} = (r-2)d_{ij} - \sum_{k=1}^{r}(d_{ik} + d_{jk}), \text{ for } i < j \le r. \tag{3.8}$$

Suppose nodes i and j are the selected nodes and they are joined to form node u. NJ estimates the length of the branch i–u as

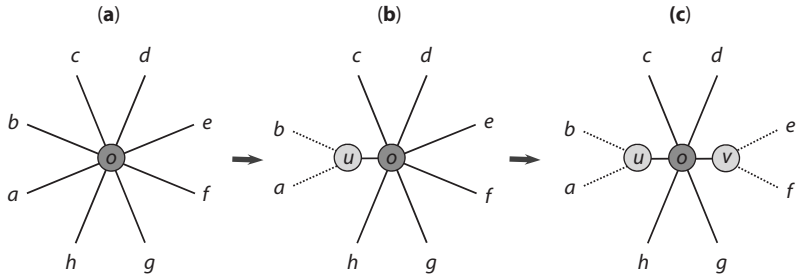


Fig. 3.20 The neighbour-joining method of tree reconstruction is a divisive cluster algorithm. It starts with the star tree (**a**), and chooses a pair of nodes (say, *a* and *b*) to join. The two joined nodes are replaced by a new node (node *u*), reducing the number of nodes connected to the root (node *o*) by one (**b**). The process is repeated until the tree is fully resolved.

$$d_{iu} = \frac{1}{2}d_{ij} + \frac{1}{2(r-2)}\left[\sum_{k=1}^{r} d_{ik} - \sum_{k=1}^{r} d_{jk}\right]. \tag{3.9}$$

The branch length d_{ju} is calculated similarly. Finally, NJ replaces i and j by u in the distance matrix, using the reduction formula

$$d_{uk} = \frac{1}{2}(d_{ik} - d_{iu}) + \frac{1}{2}(d_{jk} - d_{ju}), \tag{3.10}$$

where k is any node connected to the root (o) other than i and j. Equations (3.8)–(3.10) are due to Studier and Keppler (1988). They are equivalent to and computationally more efficient than those given by Saitou and Nei (1987), according to Gascuel (1994). The two versions always construct the same tree, both in terms of topology and branch lengths.

While good performance of NJ was noted early in computer simulations, its assumptions were not well understood until after mathematical analysis several years later. The discussion below draws heavily on Gascuel and Steel (2006).

Saitou and Nei (1987; see also Nei and Kumar 2000) provided a proof of the consistency of NJ. They showed that pair selection using their equivalent of equation (3.8) minimizes the OLS estimate of the tree length. Accordingly, they considered NJ to be an ME method, minimizing the OLS tree length. However, this proof is not strictly correct because it applies only to the first step of the algorithm, when all the nodes connected to the root are tips, but does not apply to the later steps, when some nodes are interior nodes resulting from joining early neighbours (Gascuel and Steel 2006). Also the consistency of the ME method under OLS tree lengths was not established until Rzhetsky and Nei (1993), later than Saitou and Nei (1987).

Nevertheless, NJ is based on an ME criterion, but the tree length is estimated using a different method from OLS (Gascuel and Steel 2006). Pauplin (2000) studied a method for calculating the tree length directly using the distance matrix. Note that for the example tree of four tips of Figure 3.21a, the tree length is given as

$$l = \frac{1}{2}(d_{ac} + d_{cd} + d_{db} + d_{ba}). \tag{3.11}$$

The rule here is to traverse the tree by visiting pairs of tips in the clockwise direction, i.e. in the order a, c, d, and b, as indicated by the dashed lines. This way each branch on the tree is passed twice, hence the factor $\frac{1}{2}$. If the tree is perfectly additive with the distances to be the true values, equation (3.11) will give the true tree length. Otherwise if the distances are estimates, equation (3.11) will give an estimate of the true tree length. However, the same tree can be drawn in different ways, so that this estimate of tree length is not unique. Then it is natural, as suggested by Pauplin (2000), to average over all possible ways of drawing the same tree. In our example, there is a second way of drawing the same tree, shown in Figure 3.21b, and this gives $l = \frac{1}{2}(d_{ad} + d_{dc} + d_{cb} + d_{ba})$. Averaging over the two ways gives

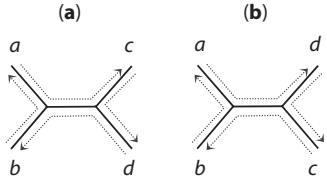


Fig. 3.21 Two different ways of drawing the same tree for four tips to explain Paulin's (2000) direct calculation of tree length from the distance matrix.

$$l = \frac{1}{2}(d_{ab} + d_{cd}) + \frac{1}{4}(d_{ac} + d_{ad} + d_{bc} + d_{bd}). \tag{3.12}$$

In general, Pauplin's (2000) estimate of tree length on a binary tree is

$$l = \sum_{i<j} w_{ij} d_{ij}, \tag{3.13}$$

which averages over all pairwise distances d_{ij}, with the weight w_{ij} to be $\frac{1}{2}$ raised to the power of the number of interior nodes on the path between i and j. This estimate was extended to multifurcating trees by Semple and Steel (2003), in which case the weight w_{ij} is calculated as follows. Consider the directed path from i to j, and for each interior node on the path, count the number of outgoing branches. Multiply those numbers and the reciprocal of the product will be w_{ij}. For example, with the tree of Figure 3.20c, we have $w_{ab} = 1/2$, $w_{cd} = 1/5$, $w_{ae} = 1/(2 \times 5 \times 2)$, and so on. For w_{ae}, note that the path from a to e passes three interior nodes (u, o, v), and the numbers of outgoing branches at those nodes are 2, 5, and 2 respectively. Semple and Steel showed that with this redefinition of w_{ij}, the tree length estimate of equation (3.13) gives exactly the average of estimates of the form of equation (3.11) over all possible ways of drawing the same tree.

Going back to our discussion of NJ, Gascuel and Steel (2006) showed that pair selection in the NJ algorithm by equation (3.8) is equivalent to minimizing the tree length defined in equation (3.13). Of course NJ is a cluster algorithm and does not search for the globally optimal tree under the criterion. One may wonder whether an exhaustive search or a more thorough search than NJ can lead to better performance. The answer to this question is 'Yes'. Indeed this ME method, based on the tree length of equation (3.13), was proposed by Pauplin (2000) and implemented by Desper and Gascuel (2002) as the *balanced ME* method in their FASTME program. Desper and Gascuel's (2002) simulations suggest that FASTME performs better than NJ and other available distance methods (see also Vinh and von Haeseler 2005). Note that equation (3.13) has some flavour of WLS, because a large distance d_{ij} tends to be separated by more interior nodes so that the weight w_{ij} will tend to be smaller (Desper and Gascuel 2004). Desper and Gascuel (2004) also showed that the balanced ME method is consistent.

In summary, NJ is an ME method, but it minimizes the tree length of equation (3.13), not the OLS estimate of tree length. Furthermore the tree length of equation (3.13) is better than the OLS estimate. This explains some counterintuitive results observed in several simulation studies (Gascuel 1997, 2000; Nei et al. 1998). Nei et al. (1998) found that minimizing the OLS tree length leads to poorer performance than NJ. The results prompted the authors to question the optimization principle. The result is unusual, as NJ was justified on the ground that it was based on the ME principle but a more correct implementation (by a more thorough search in the tree space) of the ME principle actually leads to poorer performance than NJ. The optimization principle is justified, but the criterion being optimized is important.

As implied above, a major concern with any distance matrix method is that large distances are poorly estimated, and it is important to take into account their large variances, for example, by using WLS. Besides WLS, Gascuel (1997) modified the formula for updating branch lengths in the NJ algorithm to incorporate approximate variances and covariances of distance estimates. This method, called BIONJ, is close to WLS, and was found to outperform NJ, especially when substitution rates are high and variable among lineages. Another modification is the weighted NJ or WEIGHBOR method of Bruno et al. (2000). This uses an approximate likelihood criterion for joining nodes to accommodate the fact that large distances are poorly estimated. Computer simulations suggest that

WEIGHBOR produces trees similar to ML, and is more robust to the problem of long-branch attraction (see §3.4.5) than NJ (Bruno et al. 2000). Another idea, due to Ranwez and Gascuel (2002), is to improve distance estimates. When calculating the distance between a pair of sequences, those authors used a third sequence to break the long distance into two parts and used ML to estimate three branch lengths on the tree of the three sequences; the pairwise distance is then calculated as the sum of the two branch lengths. Simulations suggest that the improved distance, when combined with the NJ, BIONJ, and WEIGHBOR algorithms, led to improved topological accuracy.

3.4 Maximum parsimony

3.4.1 *Brief history*

Felsenstein (2004) and Edwards (2009a) have published accounts of the early history of phylogeny reconstruction methods. Edwards and Cavalli-Sforza (1963a) suggested the *minimum evolution principle* (later renamed the minimum evolution method) as an approximation to the ML solution. For discrete characters, the amount of evolution should be measured by the minimum number of character changes on the tree. In modern terminology, the method applied to discrete data is known as parsimony, while ME refers to methods minimizing the sum of branch lengths after correcting for multiple hits, as discussed in last section. For discrete morphological characters, Camin and Sokal (1965) suggested the use of the minimum number of changes as a criterion for tree selection, justifying it by arguing that evolution follows the shortest paths, a view sharply criticized by Edwards (1996, 2009a). For molecular data, minimizing changes on the tree to infer ancestral proteins appears most natural and was practised by many pioneers in the field, for example, by Pauling and Zuckerkandl (1963) and Zuckerkandl (1964) as a way of 'restoring' ancestral proteins for 'paleogenetic' studies of their chemical properties, and by Eck and Dayhoff (1966) to construct empirical matrices of amino acid substitution rates. Fitch (1971b) was the first to present a systematic algorithm to enumerate all and only the most parsimonious reconstructions. Fitch's algorithm works on binary trees only. Hartigan (1973) considered multifurcating trees as well and provided a mathematical proof for the algorithm. Since then, much effort has been made to develop fast algorithms for the parsimony analysis of large datasets; see, e.g. Ronquist (1998), Nixon (1999), and Goloboff (1999).

3.4.2 *Counting the minimum number of changes on a tree*

The minimum number of character changes at a site on a given tree is often called the *character length* or *site length*. The sum of character lengths over all sites in the sequence is the minimum number of required changes for the entire sequence and is called the *tree length*, *tree score*, or *parsimony score*. The tree with the smallest tree score is the estimate of the true tree, called the *maximum parsimony tree* or the *most parsimonious tree*. It is common, especially when the sequences are very similar, for multiple trees to be equally best; i.e. they have the same minimum score and are all shortest trees.

Suppose the data for four species at a particular site are AAGG, and consider the minimum number of changes required by the two trees of Figure 3.22. We calculate this number by assigning character states to the extinct ancestral nodes. For the first tree, this is achieved by assigning A and G to the two nodes, and one change (A $\leftrightarrow$ G on the internal branch) is required. For the second tree, we can assign either AA (shown) or

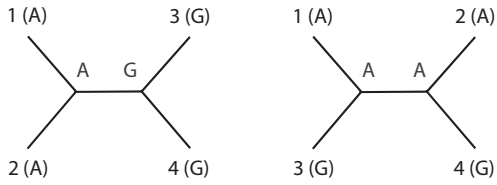


Fig. 3.22 Data AAGG at one site for four species mapped onto two alternative trees ((1, 2), 3, 4) and ((1, 3), 2, 4). The tree on the left requires a minimum of one change while the tree on the right requires two changes to explain the data.

GG (not shown) to the two internal nodes; in either case, a minimum of two changes is required. Note that the set of character states (nucleotides) at a site assigned to ancestral nodes is called an *ancestral reconstruction*. The total number of reconstructions at each site is thus $4^{(n-2)}$ for nucleotides or $20^{(n-2)}$ for amino acids as a binary unrooted tree of n species has $n-2$ interior nodes. The reconstruction that achieves the minimum number of changes is called the *most parsimonious reconstruction*. Thus, for the first tree, there is one single most parsimonious reconstruction, while for the second tree, two reconstructions are equally parsimonious. The algorithm of Fitch (1971b) and Hartigan (1973) calculates the minimum number of changes and enumerates all the most parsimonious reconstructions at a site. We will not describe this algorithm here. Instead we describe in the next subsection a more general algorithm due to Sankoff (1975), which is very similar to the likelihood algorithm to be discussed in Chapter 4.

Some sites do not contribute to the discrimination of trees and are thus noninformative. For example, a constant site, at which the different species have the same nucleotide, requires no change for any tree. Similarly a *singleton* site, at which two states are observed but one is observed only once (e.g. TTTC or AAGA), requires one change for every tree and is thus not informative. Perhaps more strikingly, a site with data AAATAACAAG (for ten species) is not informative either, as a minimum of three changes are required by any tree, which is also achieved by every tree by assigning A to all ancestral nodes. For a site to be a *parsimony-informative* site, at least two characters have to be observed, each at least twice. Note that the concepts of informative and noninformative sites apply to parsimony only. In distance and likelihood methods, all sites including the constant sites affect the calculation and should be included.

We often refer to the observed character states in all species at a site as a *site configuration* or *site pattern*. The above discussion means that for four species, only three *site patterns* are informative: *xxyy*, *xyxy*, and *xyyx*, where x and y are any two distinct states. It is obvious that those three site patterns 'support' the three trees T_1: $((1,2),3,4)$; T_2: $((1,3),2,4)$; and T_3: $((1,4),2,3)$, respectively. Suppose the numbers of sites with those site patterns are $n_1, n_2,$ and n_3, respectively. Then $T_1, T_2,$ or T_3 is the most parsimonious tree if $n_1, n_2,$ or n_3 is the greatest among the three.

3.4.3 *Weighted parsimony and dynamic programming*

The algorithm of Fitch (1971b) and Hartigan (1973) assumes that every change has the same cost. In weighted parsimony, different weights are assigned to different types of character changes. Rare changes are penalized more heavily than frequent changes. For example, transitions are known to occur at a higher rate than transversions and can be assigned a lower cost (weight). Weighted parsimony uses a *step matrix* or *cost matrix* to specify the cost of every type of change. An extreme case is *transversion parsimony*, which gives a penalty of 1 for a transversion but no penalty for a transition. Below we describe Sankoff's (1975) dynamic programming algorithm, which calculates the minimum cost at

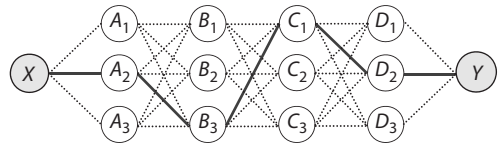


Fig. 3.23 Caravan travelling example used for illustrating the dynamic programming algorithm. It is required to determine the shortest route from *X* to *Y*, through four countries A, B, C, and D. Stops between neighbouring countries are connected, with their distances known.

a site and enumerates the reconstructions that achieve this minimum given any arbitrary cost matrix.

We first illustrate the basic idea of dynamic programming algorithms using a fictitious example of a caravan travelling on the silk route. We start from the source city *X*, Chang-an in central China, to go to the destination *Y*, Baghdad in Iraq (Figure 3.23). The route goes through four countries *A*, *B*, *C*, and *D*, and has to pass one of three caravan stops in every country: $A_1, A_2,$ or A_3 in country *A*; $B_1,$ $B_2,$ or B_3 in country *B*; and so on. We know the distance between any two stops in two neighbouring countries, such as XA_2 and A_1B_2. We seek to determine the shortest distance and the shortest route from *X* to *Y*. An obvious strategy is to evaluate all possible routes, but this can be expensive as the number of routes (3^4 in the example) grows exponentially with the number of countries. A dynamic programming algorithm answers many smaller questions, with the new questions building on answers to the old ones. First we ask for the shortest distances (from *X*) to stops $A_1, A_2,$ and A_3 in country *A*. These are just the given distances. Next we ask for the shortest distances to stops in country *B*, and then the shortest distances to stops in country *C*, and so on. Note that the questions at every stage are easy given the answers to the previous questions. For example, consider the shortest distance to C_1, when the shortest distances to $B_1, B_2,$ and B_3 are already determined. This is just the smallest among the distances of the three routes going through $B_1,$ $B_2,$ or B_3, with the distance through B_j $(j = 1,\ 2,\ 3)$ being the shortest distance (from *X*) to B_j plus the distance between B_j and C_1. After the shortest distances to $D_1, D_2,$ and D_3 are determined, it is easy to determine the shortest distance to *Y* itself. It is important to note that adding another country to the problem will add another stage in the algorithm, so that the amount of computation grows linearly with the number of countries.

We now describe Sankoff's algorithm. We seek to determine the minimum cost for a site on a given tree as well as the ancestral reconstruction that achieves that minimum. We use the tree of Figure 3.24 as an example. The observed nucleotides at the site at the six tips are CCAGAA. Let $c(x, y)$ denote the cost of change from state x to state y, so $c(x, y) = 1$ for a transitional difference and $c(x, y) = 1.5$ for a transversion (Figure 3.24).

Instead of the minimum cost for the whole tree, we calculate the minimum costs for many subtrees. We refer to a branch on the tree by the node it leads to or by the two nodes it connects. For example, branch 10 is also branch 8–10 in Figure 3.24. We say that each node *i* on the tree defines a subtree, referred to as subtree *i*, which consists of branch *i*, node *i*, and all its descendant nodes. For example, subtree 3 consists of the single tip branch 10–3 while subtree 10 consists of branch 8–10 and nodes 10, 3, and 4. Define $S_i(x)$ as the minimum cost incurred on subtree *i*, given that the mother node of node *i* has

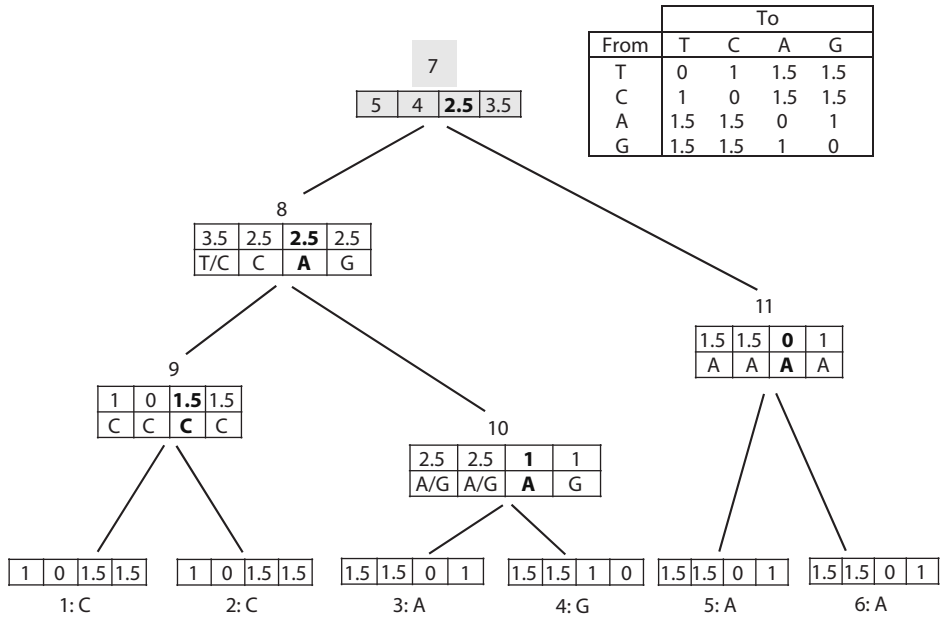


Fig. 3.24 Dynamic programming algorithm for calculating the minimum cost and enumerating the most parsimonious reconstructions using weighted parsimony. The site has observed data CCAGAA. The cost vector at each node gives the minimum cost of the subtree induced by that node (which includes the node itself, its mother branch and all its descendants), given that the mother node has nucleotides T, C, A, or G. The nucleotides at the node that achieved the minimum cost are shown below the cost vector. For example, the minimum cost of the subtree induced by node 3 (including the single branch 10–3) is 1.5, 1.5, 0, or 1, if node 10 has T, C, A, or G, respectively. The minimum cost of the subtree induced by node 10 (including branches 8–10 and nodes 10, 3 and 4) is 2.5, 2.5, 1, or 1, if node 8 has T, C, A, or G, respectively; the said minimum is achieved by node 10 having A/G, A/G, A, or G, respectively. The cost vectors are calculated for every node, starting from the tips and proceeding towards the root. At the root (node 7), the cost vector gives the minimum cost of the whole tree as 5, 4, 2.5, or 3.5, if the root has T, C, A, or G, respectively.

state x. Thus $\{S_i(\mathrm{T}), S_i(\mathrm{C}), S_i(\mathrm{A}), S_i(\mathrm{G})\}$ constitutes a cost vector for subtree i at node i. They are like the shortest distances to stops in a particular country in the caravan example. We calculate the cost vectors for all nodes on the tree, starting with the tips and visiting a node only after we have visited all its descendant nodes. For a tip node i, the subtree is just the tip branch and the cost is simply read from the cost matrix. For example, tip 3 has the cost vector $\{1.5, 1.5, 0, 1\}$, meaning that the (minimum) cost of subtree 3 is 1.5, 1.5, 0, or 1, if mother node 10 has T, C, A, or G, respectively (Figure 3.24). If the nucleotide at the tip is undetermined, the convention is to use the minimum cost among all compatible states (Fitch 1971b). For an interior node i, suppose its two daughter nodes are j and k. Then

$$S_i(x) = \min_y [c(x, y) + S_j(y) + S_k(y)]. \tag{3.14}$$

Note that subtree i consists of branch i plus subtrees j and k. Thus the minimum cost of subtree i is the cost along branch i, $c(x, y)$, plus the minimum costs of subtrees j and k, minimized over the state y at node i. We use $C_i(x)$ to record the state y that achieved the minimum.

Consider node 10 as an example, for which the cost vector is calculated to be $\{S_{10}(\mathrm{T}), S_{10}(\mathrm{C}), S_{10}(\mathrm{A}), S_{10}(\mathrm{G})\} = \{2.5, 2.5, 1, 1\}$. Here the first entry, $S_{10}(\mathrm{T}) = 2.5$, means that the minimum cost of subtree 10, given that mother node 8 has T, is 2.5. To see this, consider the four possible states at node 10: y = T, C, A, or G. The (minimum) cost on subtree 10 is 3 = 0 + 1.5 + 1.5, 4, 2.5, or 2.5, if node 10 has the state y = T, C, A, or G, respectively (and if node 8 has T). Thus the minimum is 2.5, achieved by node 10 having y = A or G; i.e. $S_{10}(\mathrm{T}) = 2.5$ and $C_{10}(\mathrm{T})$ = A or G (Figure 3.24). This is the minimization over y in equation (3.14). Similarly, the second entry in the cost vector at node 10, $S_{10}(\mathrm{C}) = 2.5$, means that the minimum cost for subtree 10, given that node 8 has C, is 2.5. This minimum is achieved by having $C_{10}(\mathrm{C})$ = A/G at node 10.

Similar calculations can be done for nodes 9 and 11. We now consider node 8, which has daughter nodes 9 and 10. The cost vector is calculated to be $\{3.5, 2.5, 2.5, 2.5\}$, meaning that the minimum cost of subtree 8 is 3.5, 2.5, 2.5, or 2.5, if mother node 7 has T, C, A, or G, respectively. Here we derive the third entry $S_8(\mathrm{A}) = 2.5$, with mother node 7 having A. By using the cost vectors for nodes 9 and 10, we calculate the minimum cost on subtree 8 to be 5 = 1.5 + 1 + 2.5, 4, 2.5, or 4.5, if node 8 has T, C, A, or G, respectively (and if mother node 7 has A). Thus $S_8(\mathrm{A}) = 2.5$ is the minimum, achieved by node 8 having $C_8(\mathrm{A}) = \mathrm{A}$.

The algorithm is applied successively to all nodes in the tree, starting from the tips and moving towards the root. This upper pass calculates $S_i(x)$ and $C_i(x)$ for all nodes i except the root. Suppose the root has daughter nodes j and k and note that the whole tree consists of subtrees j and k. The minimum cost of the whole tree, given that the root has y, is $S_j(y) + S_k(y)$. This cost vector is $\{5, 4, 2.5, 3.5\}$, for y = T, C, A, G at the root (Figure 3.24). The minimum is 2.5, achieved by having A at the root. In general, if j and k are the daughter nodes of the root, the minimum cost for the whole tree is

$$S = \min_{y}[S_j(y) + S_k(y)]. \tag{3.15}$$

After calculation of $S_i(x)$ and $C_i(x)$ for all nodes through the upper pass, a down pass reads out the most parsimonious reconstructions. In our example, given A for the root, node 8 achieves the minimum for subtree 8 by having A. Given A at node 8, nodes 9 and 10 should have C and A, respectively. Similarly given A for the root, node 11 should have A. Thus the most parsimonious reconstruction at the site is $y_7y_8y_9y_{10}y_{11}$ = AACAA, with the minimum cost 2.5.

3.4.4 *Probabilities of ancestral states*

Obviously the ancestral states reconstructed by parsimony may not always be the true states. Many authors thus recognized the desirability of calculating the probability that the parsimony reconstructions are the true states (Fitch 1971b; Maddison and Maddison 1982). This can only be achieved by the use of a character evolution model. Unfortunately most of those calculations do not appear to be correct (e.g. Fitch 1971b; Schluter 1995; Pagel 1999) or relevant (e.g. Maddison 1995). We defer the discussion of such calculations to §4.4, where the correct approach is described.

3.4.5 *Long-branch attraction*

Felsenstein (1978b) demonstrated that the parsimony method can be statistically inconsistent under certain combinations of branch lengths on a four-species tree. When the

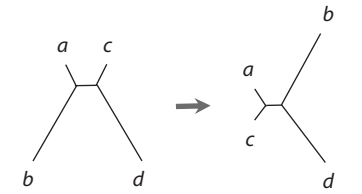


Fig. 3.25 Long-branch attraction. When the correct tree (T_1) has two long branches separated by a short internal branch, parsimony tends to recover a wrong tree (T_2) with the two long branches grouped together.

amount of data (the number of sites) increases to infinity, it becomes more and more certain that the most parsimonious tree is an incorrect tree.

The tree Felsenstein used has the characteristic shape shown in Figure 3.25a, with two long branches separated by a short internal branch. The estimated tree by parsimony, however, tends to group the two long branches together (Figure 3.25b). This phenomenon is now known as 'long-branch attraction'. Using a simple model of character evolution, Felsenstein calculated the probabilities of observing sites with the three site patterns *xxyy*, *xyxy*, *xyyx*, where *x* and *y* are any two distinct characters, and found that Pr(*xyxy*) > Pr(*xxyy*) when the two long branches are much longer than the three short branches. This calculation will be described later in Chapter 4 (see Problem 4.3). Thus with more and more sites in the sequence, it will be increasingly certain that more sites will have pattern *xyxy* than pattern *xxyy*, and that parsimony will recover the wrong tree T_2 instead of the true tree T_1. The phenomenon has been demonstrated in many simulated and real datasets (see, e.g. Huelsenbeck 1998) and is due to the failure of parsimony to correct for parallel changes on the two long branches. Likelihood and distance methods using simplistic and unrealistic evolutionary models show the same behaviour.

3.4.6 *Assumptions of parsimony*

A discussion of the assumptions underlying the parsimony method of phylogeny reconstruction is provided in Chapter 5. Here we comment on a few obvious concerns on the parsimony reconstruction of ancestral states. First, the method ignores branch lengths. Some branches on the tree are longer than others, meaning that they have accumulated more evolutionary changes than other branches. It is thus illogical to assume that a change is as likely to occur on a long branch as on a short one, as parsimony does, when character states are assigned to ancestral nodes on the tree. Second, the simple parsimony criterion ignores different rates of changes between nucleotides. Such rate differences are taken into account by weighted parsimony through the use of a step matrix, although determining the appropriate weights may be nontrivial. In theory, how likely a change is to occur on a particular branch should depend on the length of the branch as well as the relative rate of the change. If one attempts to derive appropriate weights from the observed data, one will naturally be led to the likelihood method, which uses a Markov chain model to describe the nucleotide substitution process, relying on probability theory to accommodate unequal branch lengths, unequal substitution rates between nucleotides, and any other features of the evolutionary process. This is the topic of next chapter.

3.5 Problems

3.1 Draw the tree

(((human: 0.040, chimpanzee: 0.052): 0.016, gorilla: 0.059) : 0.047, orangutan: 0.090, gibbon: 0.125);

The branch lengths are the MLEs under JC69 obtained from the mitochondrial data of Brown et al. (1982). Identify the most distant pair of species and use midpoint rooting to root the tree. Draw the resulting rooted tree.

3.2 Write two equivalent Newick representations of the tree in Figure 3.9b.

3.3 The following rooted tree is shown in Figure 3.26:

(a:0.05, (c: 0.07, ((b:0.015, f:0.12) :0.01, (d:0.01, e:0.4) :0.005) :0.03) :0.025);

Which of the following statements are incorrect?
(a) Species d and e are most closely related.
(b) Sequences b and d are most similar.
(c) Species b is more closely related to d than to e.
(d) Species d is more closely related to c than to f.

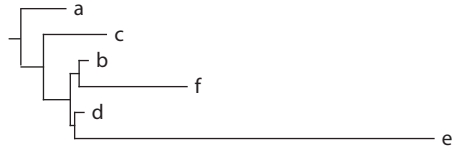


Fig. 3.26 A tree showing branch lengths for Problem 3.3.

3.4 Calculate the partition distance between the two trees of Figure 3.11.

3.5 Use the three trees of Figure 3.27 to construct the majority-rule consensus tree, and show the support values for the nodes on it.

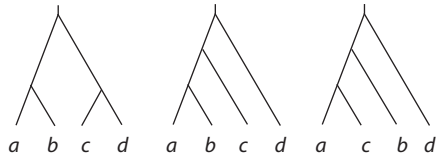


Fig. 3.27 Three rooted trees for constructing the majority-rule consensus tree in Problem 3.5.

References

Ababneh, F., L. S. Jermiin, C. Ma, and J. Robinson. 2006. Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. *Bioinformatics* **22**, 1225–1231.

Abascal, F., D. Posada, and R. Zardoya. 2007. MtArt: a new model of amino acid replacement for Arthropoda. *Mol. Biol. Evol*. **24**, 1–5.

Abramowitz, M., and I. A. Stegun. 1972. *Handbook of Mathematical Functions*. Dover, New York.

Adachi, J., and M. Hasegawa. 1996a. MolPhy Version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Computer Science Monographs* **28**, 1–150.

Adachi, J., and M. Hasegawa. 1996b. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol*. **42**, 459–468.

Adachi, J., P. J. Waddell, W. Martin, and M. Hasegawa. 2000. Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J. Mol. Evol*. **50**, 348–358.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Contr. AC* **19**, 716–723.

Akam, M. 1995. Hox genes and the evolution of diverse body plans. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **349**, 313–319.

Akashi, H. 1994. Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. *Genetics* **136**, 927–935.

Akashi, H. 1995. Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in Drosophila DNA. *Genetics* **139**, 1067–1076.

Akashi, H. 1999a. Within- and between-species DNA sequence variation and the 'footprint' of natural selection. *Gene* **238**, 39–51.

Akashi, H. 1999b. Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. *Genetics* **151**, 221–238.

Akashi, H., P. Goel, and A. John. 2007. Ancestral state inference and the study of codon bias evolution: implications for molecular evolutionary analysis of the *Drosophila melanogaster* subgroup. *PloS One* **2**, e1065.

Akerborg, O., B. Sennblad, and J. Lagergren. 2008. Birth-death prior on phylogeny and speed dating. *BMC Evol. Biol*. **8**, 77.

Alba, R., P. M. Kelmenson, M.-M. Cordonnier-Pratt, and L. H. Pratt. 2000. The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms. *Mol. Biol. Evol*. **17**, 362–373.

Albert, V. A. 2005. *Parsimony, Phylogeny, and Genomics*. Oxford University Press, Oxford, UK.

Aldous, D. J. 2001. Stochastic models and descriptive statistics for phylogenetic trees. *Stat. Sci*. **16**, 23–34.

Alfaro, M. E., and M. T. Holder. 2006. The posterior and the prior in Bayesian phylogenetics. *Ann. Rev. Ecol. Syst*. **37**, 19–42.

Allman, E. S., and J. A. Rhodes. 2006. The identifiability of tree topology for phylogenetic models, including covarion and mixture models. *J. Comput. Biol*. **13**, 1101–1113.

Allman, E. S., C. Ane, and J. A. Rhodes. 2008. Identifiability of a Markovian model of molecular evolution with gamma-distributed rates. *Adv. Appl. Prob*. **40**, 228–249.

Altekar, G., S. Dwarkadas, J. P. Huelsenbeck, and F. Ronquist. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**, 407–415.

Andersen, L. N., T. Mailund, and A. Hobolth. 2014. Efficient computation in the IM model. *J. Math. Biol*. in press.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. *Nature* **437**, 1149–1152.

Ané, C., B. Larget, D. A. Baum et al. 2007. Bayesian estimation of concordance among gene trees. *Mol. Biol. Evol.* **24**, 412–426.

Anisimova, M. 2012. Parametric models of codon substitution. Pp. 12–33 *in* G. Cannarozzi, and A. Schneider, eds. *Codon Evolution: Mechanisms and Models*. Oxford University Press, New York.

Anisimova, M., and O. Gascuel. 2006. Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. *Syst. Biol.* **55**, 539–552.

Anisimova, M., and C. Kosiol. 2009. Investigating protein-coding sequence evolution with probabilistic codon substitution models. *Mol. Biol. Evol.* **26**, 255–271.

Anisimova, M., and C. Kosiol. 2012. Selection on the protein-coding genome. Pp. 113–140 *in* M. Anisimova, ed. *Evolutionary Genomics: Statistical and Computational Methods, Volume 2*. Springer, New York.

Anisimova, M., and D. A. Liberles. 2007. The quest for natural selection in the age of comparative genomics. *Heredity*. **99**: 567–579.

Anisimova, M., and Z. Yang. 2007. Multiple hypothesis testing to tetect adaptive protein evolution affecting individual branches and sites. *Mol. Biol. Evol.* **24**, 1219–1228.

Anisimova, M., J. P. Bielawski, and Z. Yang. 2001. The accuracy and power of likelihood ratio tests to detect positive selection at amino acid sites. *Mol. Biol. Evol.* **18**, 1585–1592.

Anisimova, M., J. P. Bielawski, and Z. Yang. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol. Biol. Evol.* **19**, 950–958.

Anisimova, M., R. Nielsen, and Z. Yang. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**, 1229–1236.

Antoniak, C. E. 1974. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Stat.* **2**, 1152–1174.

Aris-Brosou, S. 2007. Dating phylogenies with hybrid local molecular clocks. *PLOS One* **2**, e879.

Atchadé, Y. F., G. O. Roberts, and J. S. Rosenthal. 2011. Towards optimal scaling of metropolis-coupled Markov chain Monte Carlo. *Stat. Comput.* **21**, 555–568.

Atkinson, A. C. 1970. A method of discriminating between models. *J. R. Stat. Soc. B.* **32**, 323–353.

Avise, J. C. 2000. *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Baele, G., P. Lemey, T. Bedford et al. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167.

Bahlo, M., and R. C. Griffiths. 2000. Inference from gene trees in a subdivided population. *Theor. Popul. Biol.* **57**, 79–95.

Barker, D. 2004. LVB: parsimony and simulated annealing in the search for phylogenetic trees. *Bioinformatics* **20**, 274–275.

Barrier, M., R. H. Robichaux, and M. D. Purugganan. 2001. Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10208–10213.

Barry, D., and J. A. Hartigan. 1987a. Statistical analysis of hominoid molecular evolution. *Stat. Sci.* **2**, 191–210.

Barry, D., and J. A. Hartigan. 1987b. Asynchronous distance between homologous DNA sequences. *Biometrics* **43**, 261–276.

Baudry, E., and F. Depaulis. 2003. Effect of misoriented sites on neutrality tests with outgroup. *Genetics* **165**, 1619–1622.

Bauer, A. M., J. F. Parham, R. M. Brown et al. 2011. Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. *Proc. R. Soc. Lond. B. Biol. Sci.* **278**, 490–492.

Baum, D. A. 1992. Phylogenetic species concepts. *Trends Ecol. Evol.* **7**, 1–2.

Baum, D. A. 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* **56**, 417–426.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pp. 289–303 *in* P. C. Hoch, and A. G. Stephenson, eds. *Molecular and Experimental Approaches to Plant Biosystematics*. Missouri Botanical Garden, St. Louis.

Bayes, T. 1763. An essay towards solving a problem in the doctrine of chance, with an Introduction and an Appendix by Richard Price. *Philos. Trans. R. Soc. Lond.* **53**, 370–418.

Beaumont, M. A. 1999. Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013–2029.

Beerli, P. 2004. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Mol. Ecol.* **13**, 827–836.

Beerli, P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341–345.

Beerli, P., and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**, 763–773.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4563–4568.

Benner, S. A. 2001. Natural progression. *Nature* **409**, 459.

Benner, S. A. 2002. The past as the key to the present: resurrection of ancient proteins from eosinophils. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4760–4761.

Benton, M. J., P. C. J. Donoghue, and R. J. Asher. 2009. Calibrating and constraining molecular clocks. Pp. 35–86 *in* B. S. Hedges, and S. Kumar, eds. *The Timetree of Life*. Oxford University Press, Oxford, UK.

Berg, B. A., and T. Neuhaus. 1991. Multicanonical algorithms for 1st order phase-transitions. *Phys. Lett. B* **267**, 249–253.

Berger, J. O., and J. M. Bernardo. 1992. On the development of reference priors (with Discussion). Pp. 35–60 *in* J. M. Bernardo, J. O. Berger, D. V. Lindley, and A. F. M. Smith, eds. *Bayesian Statistics 4*. Oxford University Press, Oxford, UK.

Berger, J. O., J. M. Bernardo, and D. Sun. 2009. The formal definition of reference priors. *Ann. Stat.* **37**, 905–938.

Bernardo, J. M. 1979. Reference posterior distributions for Bayesian inference. *J. R. Stat. Soc. B* **41**, 113–147.

Bernardo, J. M. 2005. Reference analysis. *Handb. Stat.* **25**, 17–90.

Berry, I. M., R. Ribeiro, M. Kothari et al. 2007. Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J. Virol.* **81**, 10625–10635.

Berry, V., and O. Gascuel. 1996. On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. *Mol. Biol. Evol.* **13**, 999–1011.

Besag, J., and P. J. Green. 1993. Spatial statistics and Bayesian computation. *J. R. Stat. Soc. B* **55**, 25–37.

Bielawski, J. P., and J. R. Gold. 2002. Mutation patterns of mitochondrial H- and L-strand DNA in closely related Cyprinid fishes. *Genetics* **161**, 1589–1597.

Bielawski, J. P., and Z. Yang. 2001. Positive and negative selection in the DAZ gene family. *Mol. Biol. Evol.* **18**, 523–529.

Bielawski, J. P., and Z. Yang. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **59**, 121–132.

Bielawski, J. P., K. Dunn, and Z. Yang. 2000. Rates of nucleotide substitution and mammalian nuclear gene evolution: approximate and maximum-likelihood methods lead to different conclusions. *Genetics* **156**, 1299–1308.

Bielawski, J. P., K. A. Dunn, G. Sabehi, and O. Beja. 2004. Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14824–14829.

Bierne, N., and A. Eyre-Walker. 2003. The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation between the synonymous substitution rate and codon usage bias. *Genetics* **165**, 1587–1597.

Bininda-Emonds, O. R. P. 2004. *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life*. Kluwer Academic, Dordrecht, the Netherlands.

Bishop, M. J., and A. E. Friday. 1985. Evolutionary trees from nucleic acid and protein sequences. *Proc. R. Soc. Lond. B. Biol. Sci.* **226**, 271–302.

Bishop, M. J., and A. E. Friday. 1987. Tetropad relationships: the molecular evidence. Pp. 123–139 *in* C. Patterson, ed. *Molecules and Morphology in Evolution: Conflict or Compromise?* Cambridge University Press, Cambridge, UK.

Bishop, M. J., and E. A. Thompson. 1986. Maximum likelihood alignment of DNA sequences. *J. Mol. Biol*. **190**, 159–165.

Bjorklund, M. 1999. Are third positions really that bad? A test using vertebrate cytochrome *b*. *Cladistics* **15**, 191–197.

Bjorkman, P. J., S. A. Saper, B. Samraoui et al. 1987a. Structure of the class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512.

Bjorkman, P. J., S. A. Saper, B. Samraoui et al. 1987b. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**, 512–518.

Blanquart, S., and N. Lartillot. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol. Biol. Evol*. **23**, 2058–2071.

Blanquart, S., and N. Lartillot. 2008. A site- and time-heterogeneous model of amino acid replacement. *Mol. Biol. Evol*. **25**, 842–858.

Bollback, J. P. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**, 88.

Bouchard-Coté, A., S. Sankararaman, and M. I. Jordan. 2012. Phylogenetic inference via sequential Monte Carlo. *Syst. Biol*. **61**, 579–593.

Bourlat, S. J., T. Juliusdottir, C. J. Lowe et al. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88.

Boussau, B., and M. Gouy. 2006. Efficient likelihood computations with nonreversible models of evolution. *Syst. Biol*. **55**, 756–768.

Box, G. E. P. 1979. Robustness in the strategy of scientific model building. Pp. 202 *in* R. L. Launer, and G. N. Wilkinson, eds. *Robustness in Statistics*. Academic Press, New York.

Box, G. E. P., and M. E. Muller. 1958. A note on the generation of random normal deviates. *Ann. Math. Stat*. **29**, 610–611.

Brandley, M. C., A. D. Leqach, D. L. Warren, and J. A. McGruire. 2006. Are unequal clade priors problematic for Bayesian phylogenetics? *Syst. Biol*. **55**, 158–146.

Braverman, J. M., R. R. Hudson, N. L. Kaplan et al. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783–796.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**, 795–803.

Brenner, S. 1988. The molecular evolution of genes and proteins: a tale of two serines. *Nature* **334**, 528–530.

Brent, R. P. 1973. *Algorithms for Minimization without Derivatives*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Brinkmann, H., M. van der Giezen, Y. Zhou et al. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst. Biol*. **54**, 743–757.

Britten, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**, 1393–1398.

Britton, T. 2005. Estimating divergence times in phylogenetic trees without a molecular clock. *Syst. Biol*. **54**, 500–507.

Bromham, L. 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Phil. Trans. R. Soc. B: Biol. Sci*. **366**, 2503–2513.

Bromham, L., and D. Penny. 2003. The modern molecular clock. *Nat. Rev. Genet*. **4**, 216–224.

Bromham, L., D. Penny, A. Rambaut, and M. D. Hendy. 2000. The power of relative-rates tests depends on the data. *J. Mol. Evol*. **50**, 296–301.

Brooks, S. P., P. Giudici, and G. O. Roberts. 2003. Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *J. R. Stat. Soc. B*. **65**, 3–39.

Brown, J. M., and A. R. Lemmon. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst. Biol*. **56**, 643–655.

Brown, J. M., S. M. Hedtke, A. R. Lemmon, and E. M. Lemmon. 2010. When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. *Syst. Biol*. **59**, 145–161.

Brown, W. M., E. M. Prager, A. Wang, and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.* **18**, 225–239.

Brunet, M., F. Guy, D. Pilbeam et al. 2002. A new hominid from the upper Miocene of Chad, central Africa. *Nature* **418**, 145–151.

Bruno, W. J. 1996. Modeling residue usage in aligned protein sequences via maximum likelihood. *Mol. Biol. Evol.* **13**, 1368–1374.

Bruno, W. J., and A. L. Halpern. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. *Mol. Biol. Evol.* **16**, 564–566.

Bruno, W. J., N. D. Socci, and A. L. Halpern. 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.* **17**, 189–197.

Bryant, D. 2003. A classication of consensus methods for phylogenetics. Pp. 163–184 *in* M. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts, eds. *BioConsensus, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, Providence, Rhode Island.

Bryant, D., and P. J. Waddell. 1998. Rapid evaluation of least-squares and minimum-evolution criteria on phylogenetic trees. *Mol. Biol. Evol.* **15**, 1346–1359.

Bryant, D., R. Bouckaert, J. Felsenstein et al. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* **29**, 1917–1932.

Bulmer, M. G. 1990. Estimating the variability of substitution rates. *Genetics* **123**, 615–619.

Burgess, R., and Z. Yang. 2008. Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol. Biol. Evol.* **25**, 1979–1994.

Burridge, C. P., D. Craw, and J. M. Waters. 2006. River capture, range expansion, and cladogenesis: the genetic signature of freshwater vicariance. *Evolution* **60**, 1038–1049.

Bustamante, C. D., R. Nielsen, and D. L. Hartl. 2003. Maximum likelihood and Bayesian methods for estimating the distribution of selective effects among classes of mutations using DNA polymorphism data. *Theor. Popul. Biol.* **63**, 91–103.

Bustamante, C. D., J. Wakeley, S. Sawyer, and D. L. Hartl. 2001. Directional selection and the site-frequency spectrum. *Genetics* **159**, 1779–1788.

Bustamante, C. D., R. Nielsen, S. A. Sawyer et al. 2002. The cost of inbreeding in Arabidopsis. *Nature* **416**, 531–534.

Butler, G., M. D. Rasmussen, M. F. Lin et al. 2009. Evolution of pathogenicity and sexual reproduction in eight Candida genomes. *Nature* **459**, 657–662.

Calderhead, B., and M. Girolami. 2009. Estimating Bayes factors via thermodynamic integration and population MCMC. *Comput. Stat. Data Analysis* **48**, 4028–4045.

Camin, J. H., and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. *Evolution* **19**, 311–326.

Cannarozzi, G., and A. Schneider. 2012. *Codon Evolution: Mechanisms and Models*. Oxford University Press, New York.

Cao, Y., K. S. Kim, J. H. Ha, and M. Hasegawa. 1999. Model dependence of the phylogenetic inference: relationship among Carnivores, Perissodactyls and Cetartiodactyls as inferred from mitochondrial genome sequences. *Genes Genet. Syst.* **74**, 211–217.

Cao, Y., J. Adachi, A. Janke et al. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *J. Mol. Evol.* **39**, 519–527.

Cao, Y., A. Janke, P. J. Waddell et al. 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *J. Mol. Evol.* **47**, 307–322.

Carlin, B. P., and S. Chib. 1995. Bayesian model choice through Markov chain Monte Carlo. *J. R. Stat. Soc. B* **57**, 473–483.

Carlin, B. P., and T. A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, London.

Carroll, S. B. 1995. Homeotic genes and the evolution of the arthropods and chordates. *Nature* **376**, 479–485.

Carroll, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36.

Carstens, B. C., and T. A. Dewey. 2010. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American Myotis bats. *Syst. Biol*. **59**, 400–414.

Cartwright, R. A. 2005. DNA assembly with gaps (Dawg): simulating sequence evolution. *Bioinformatics* **21**, iii31–38.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1966. Estimation procedures for evolutionary branching processes. *Bull. Int. Stat. Inst*. **21**, 803–808.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* **21**, 550–570.

Cavender, J. A. 1978. Taxonomy with confidence. *Math. Biosci.* **40**, 271–280.

Chang, B. S., and M. J. Donoghue. 2000. Recreating ancestral proteins. *Trends Ecol. Evol*. **15**, 109–114.

Chang, J. T. 1996a. Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. *Math. Biosci*. **137**, 51–73.

Chang, J. T. 1996b. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci*. **134**, 189–215.

Charleston, M. A. 1995. Toward a characterization of landscapes of combinatorial optimization problems, with special attention to the phylogeny problem. *J. Comput. Biol*. **2**, 439–450.

Chen, F.-C., and W.-H. Li. 2001. Genomic divergences between humans and other Hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet*. **68**, 444–456.

Chen, M.-H., and Q.-M. Shao. 1999. Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Comput. Graph. Stat*. **8**, 69–92.

Chen, M.-H., L. Kuo, and P. Lewis. 2014. *Bayesian Phylogenetics: Methods, Algorithms, and Applications*. Chapman & Hall/CRC, London. in press.

Cheon, S., and F. Liang. 2009. Bayesian phylogeny analysis via stochastic approximation Monte Carlo. *Mol. Phylogenet. Evol*. **53**, 394–403.

Chernoff, H. 1954. On the distribution of the likelihood ratio. *Ann. Math. Stat*. **25**, 573–578.

Chor, B., and S. Snir. 2004. Molecular clock fork phylogenies: closed form analytic maximum likelihood solutions. *Syst. Biol*. **53**, 963–967.

Chor, B., B. R. Holland, D. Penny, and M. D. Hendy. 2000. Multiple maxima of likelihood in phylogenetic trees: an analytic approach. *Mol. Biol. Evol*. **17**, 1529–1541.

Chung, Y., and C. Ané. 2011. Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. *Syst. Biol*. **60**, 261–275.

Clark, B. 1970. Selective constraints on amino-acid substitutions during the evolution of proteins. *Nature* **228**, 159–160.

Clark, N. L., J. E. Aagaard, and W. J. Swanson. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* **131**, 11–22.

Collins, T. M., P. H. Wimberger, and G. J. P. Naylor. 1994. Compositional bias, character-state bias, and character-state reconstruction using parsimony. *Syst. Biol*. **43**, 482–496.

Comeron, J. M. 1995. A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. *J. Mol. Evol*. **41**, 1152–1159.

Cooper, A., and R. Fortey. 1998. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol*. **13**, 151–156.

Cox, D. R. 1961. Tests of separate families of hypotheses. *Proc. 4th Berkeley Symp. Math. Stat. Prob*. **1**, 105–123.

Cox, D. R. 1962. Further results on tests of separate families of hypotheses. *J. R. Stat. Soc. B*. **24**, 406–424.

Cox, D. R., and D. V. Hinkley. 1974. *Theoretical Statistics*. Chapman and Hall, London.

Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sinauer Assoc., Sunderland, Massachusetts.

Cranston, K. A., B. Hurwitz, D. Ware et al. 2009. Species trees from highly incongruent gene trees in rice. *Syst. Biol*. **58**, 489–500.

Crawford, N. G., B. C. Faircloth, J. E. McCormack et al. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol. Lett*. **8**, 783–786.

Cummings, M. P., S. P. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* **12**, 814–822.

Cummings, M. P., S. A. Handley, D. S. Myers et al. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* **52**, 477–487.

Cutler, D. J. 2000. Understanding the overdispersed molecular clock. *Genetics* **154**, 1403–1417.

Dagan, T., Y. Talmor, and D. Graur. 2002. Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. *Mol. Biol. Evol.* **19**, 1022–1025.

Dalquen, D. A., M. Anisimova, G. H. Gonnet, and C. Dessimoz. 2012. ALF: a simulation framework for genome evolution. *Mol. Biol. Evol.* **29**, 1115–1123.

Dasmahapatra, K. K., G. Iamas, F. Simpson, and J. Mallet. 2010. The anatomy of a 'suture zone' in Amazonian butterflies: a coalescent-based test for vicariant geographic divergence and speciation. *Mol. Ecol.* **19**, 4283–4301.

Datta, G. S., and M. Ghosh. 1996. On the invariance of noninformative priors. *Ann. Stat.* **24**, 141–159.

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap Methods and their Application*. Cambridge University Press, , Cambridge, UK.

Dawid, A. P. 1992. Prequential analysis, stochastic complexity and Bayesian inference (with Discussion). Pp. 109–125 *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds. *Bayesian Statistics*. Clarendon Press, Oxford, UK.

Dawkins, R., and J. R. Krebs. 1979. Arms races between and within species. *Proc. R. Soc. Lond. B. Biol. Sci.* **205**, 489–511.

Dayhoff, M. O., R. V. Eck, and C. M. Park. 1972. Evolution of a complex system: the immunoglobulins. Pp. 31–40. *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Maryland.

Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt. 1978. A model of evolutionary change in proteins. Pp. 345–352. *Atlas of Protein Sequence and Structure, Volume 5, Suppl. 3*. National Biomedical Research Foundation, Washington DC.

De Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* **56**, 879–886.

Dean, A. M., and J. W. Thornton. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.* **8**, 675–688.

DeBry, R. 2001. Improving interpretation of the decay index for DNA sequences. *Syst. Biol.* **50**, 742–752.

DeBry, R. W. 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. *Mol. Biol. Evol.* **9**, 537–551.

Deely, J. J., and D. V. Lindley. 1981. Bayes empirical Bayes. *J. Am. Stat. Assoc.* **76**, 833–841.

Degnan, J. H., and N. A. Rosenberg. 2006. Discordance of species trees with their most likely gene trees. *PLoS Genet.* **2**, e68.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**, 332–340.

Degnan, J. H., and L. A. Salter. 2005. Gene tree distributions under the coalescent process. *Evolution* **59**, 24–37.

Degnan, J. H., N. A. Rosenberg, and T. Stadler. 2012. The probability distribution of ranked gene trees on a species tree. *Math. Biosci.* **235**, 45–55.

DeGroot, M. H., and M. J. Schervish. 2002. *Probability and Statistics*. Addison-Wesley, Boston, Massachusetts.

Delson, E., I. Tattersall, J. A. Van Couvering, and A. S. Brooks. 2000. Pp. 166–171 *in* E. Delson, I. Tattersall, J. A. Van Couvering, and A. S. Brooks, eds. *Encyclopedia of Human Evolution and Prehistory*. Garland, New York.

Desper, R., and O. Gascuel. 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J. Comput. Biol.* **9**, 687–705.

Desper, R., and O. Gascuel. 2004. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol. Biol. Evol.* **21**, 587–598.

Desper, R., and O. Gascuel. 2005. The minimum-evolution distance-based approach to phylogenetic inference. Pp. 1–32 *in* O. Gascuel, ed. *Mathematics of Evolution and Phylogeny*. Oxford University Press, Oxford, UK.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. *Nature* **400**, 354–357.
Diggle, P. J. 1990. *Time Series: A Biostatistical Introduction*. Oxford University Press, Oxford, UK.
Dimmic, M. W., J. S. Rest, D. P. Mindell, and R. A. Goldstein. 2002. rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *J. Mol. Evol.* **55**, 65–73.
Dobzhansky, T. G. 1937. *Genetics and the Origin of Species*. Columbia University, New York.
Donnelly, P., and S. Tavaré. 1997. *Progress in Population Genetics and Human Evolution*. Springer-Verlag, New York.
Donnelly, P., and S. Tavaré. 2005. Coalescents and genealogical structure under neutrality. *Ann. Rev. Genet*. **29**, 401–421.
Doolittle, F. W. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet*. **14**, 307–311.
Doolittle, R. F., and B. Blomback. 1964. Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. *Nature* **202**, 147–152.
Dornburg, A., F. Santini, and M. E. Alfaro. 2008. The influence of model averaging on clade posteriors: an example using the triggerfishes (Family Balistidae). *Syst. Biol.* **57**, 905–919.
Doron-Faigenboim, A., and T. Pupko. 2007. A combined empirical and mechanistic codon model. *Mol. Biol. Evol.* **24**, 388–397.
dos Reis, M., and Z. Yang. 2011. Approximate likelihood calculation for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172.
dos Reis, M., and Z. Yang. 2013a. The unbearable uncertainty of Bayesian divergence time estimation. *J. Syst. Evol.* **51**, 30–43.
dos Reis, M., and Z. Yang. 2013b. Why do more divergent sequences produce smaller nonsynonymous/synonymous rate ratios in pairwise sequence comparisons? *Genetics* **195**, 195–204.
dos Reis, M., J. Inoue, M. Hasegawa et al. 2012. Phylogenomic data sets provide both precision and accuracy in estimating the timescale of placental mammal evolution. *Proc. R. Soc. Lond. B. Biol. Sci.* **279**, 3491–3500.
Douady, C. J., F. Delsuc, Y. Boucher et al. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* **20**, 248–254.
Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214.
Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88.
Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307–1320.
Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192.
Duda, T. F., and S. R. Palumbi. 2000. Evolutionary diversification of multigene families: allelic selection of toxins in predatory cone snails. *Mol. Biol. Evol.* **17**, 1286–1293.
Duret, L. 2002. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* **12**, 640–649.
Duret, L., M. Semon, G. Piganeau et al. 2002. Vanishing GC-rich isochores in mammalian genomes. *Genetics* **162**, 1837–1847.
Dutheil, J., T. Pupko, A. Jean-Marie, and N. Galtier. 2005. A model-based approach for detecting coevolving positions in a molecule. *Mol. Biol. Evol.* **22**, 1919–1928.
Eck, R. V., and M. O. Dayhoff. 1966. Inference from protein sequence comparisons. *in* M. O. Dayhoff, ed. *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Maryland.
Edgeworth, F. Y. 1885. Observations and statistics. *Trans. Cam. Phil. Soc.* **14**, 138–169.
Edwards, A. W. F. 1970. Estimation of the branch points of a branching diffusion process (with discussion). *J. R. Stat. Soc. B.* **32**, 155–174.
Edwards, A. W. F. 1974. A problem in the doctrine of chances. Pp. 43–60 *in* O. Barndorff-Nielsen, P. Balaesild, and G. Schou, eds. *Proceedings of the Conference on Foundational Questions in Statistical Inference*. Institute of Mathematics, University of Aarhus, Denmark.
Edwards, A. W. F. 1992. *Likelihood*. John Hopkins University Press, London.

Edwards, A. W. F. 1996. The origin and early development of the method of minimum evolution for the reconstruction of phylogenetic trees. *Syst. Biol.* **45**, 79–91.
Edwards, A. W. F. 2009a. Statistical methods for evolutionary trees. *Genetics* **183**, 5–12.
Edwards, A. W. F., and L. L. Cavalli-Sforza. 1963a. A method for cluster analysis (Abstract). *The 5th International Biometrics Conference*, Cambridge, UK.
Edwards, A. W. F., and L. L. Cavalli-Sforza. 1963b. The reconstruction of evolution (Abstract). *Ann. Hum. Genet*. **27**, 105.
Edwards, A. W. F., and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. *Phenetic. Phylogenet. Classificat. Syst. Assoc. Publ.* **6**, 67–76.
Edwards, S. V. 2009b. Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1–19.
Edwards, S. V., and P. Beerli. 2000. Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839–1854.
Edwards, S. V., W. B. Jennings, and A. M. Shedlock. 2005. Phylogenetics of modern birds in the era of genomics. *Proc. R. Soc. B.* **272**, 979–992.
Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Ann. Stat*. **7**, 1–26.
Efron, B. 1986. Why isn't everyone a Bayesian? (with discussion). *Am. J. Stat. Assoc*. **40**, 1–11.
Efron, B., and D. V. Hinkley. 1978. Assessing the accuracy of the maximum likelihood estimator: observed and expected information. *Biometrika* **65**, 457–487.
Efron, B., and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, London.
Efron, B., and R. J. Tibshirani. 1998. The problem of regions. *Ann. Stat*. **26**, 1687–1718.
Efron, B., E. Halloran, and S. Holmes. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. U.S.A*. **93**, 13429–13434 [corrected and republished article originally printed in Proc. Natl. Acad. Sci. U.S.A. 1996, **93**, 7085–7090].
Ence, D. D., and B. C. Carstens. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. *Mol. Ecol. Resour*. **11**, 473–480.
Erixon, P., B. Svennblad, T. Britton, and B. Oxelman. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* **52**, 665–673.
Everitt, B. S., S. Landau, and M. Leese. 2001. *Cluster Analysis*. Arnold, London.
Ewens, W. J. 1990. Population genetics theory – the past and the future. Pp. 177–227 *in* S. Lessard, ed. *Mathematical and Statistical Developments of Evolutionary Theory*. Kluwer Academic, Amsterdam.
Excoffier, L., and Z. Yang. 1999. Substitution rate variation among sites in the mitochondrial hypervariable region I of humans and chimpanzees. *Mol. Biol. Evol.* **16**, 1357–1368.
Eyre-Walker, A. 1998. Problems with parsimony in sequences of biased base composition. *J. Mol. Evol*. **47**, 686–690.
Farris, J. S. 1969. A successive approximation approach to character weighting. *Syst. Zool*. **18**, 374–385.
Farris, J. S. 1973. A probability model for inferring evolutionary trees. *Syst. Zool*. **22**, 250–256.
Farris, J. S. 1977. Phylogenetic analysis under Dollo's law. *Syst. Zool*. **26**, 77–88.
Farris, J. S. 1983. The logical basis of phylogenetic analysis. Pp. 7–26 *in* N. Platnick, and V. Funk, eds. *Advances in Cladistics*. Columbia University Press, New York.
Farris, J. S. 1989. The retention index and the rescaled consistency index. *Cladistics* **5**, 417–419.
Fay, J. C., and C.-I. Wu. 2001. The neutral theory in the genomic era. *Curr. Opinion Genet. Dev*. **11**, 642–646.
Fay, J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413.
Fay, J. C., and C. I. Wu. 2003. Sequence divergence, functional constraint, and selection in protein evolution. *Ann. Rev. Genomics Hum. Genet*. **4**, 213–235.
Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. *Trends Genet*. **28**, 342–350.
Felsenstein, J. 1973a. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet*. **25**, 471–492.
Felsenstein, J. 1973b. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool*. **22**, 240–249.

Felsenstein, J. 1978a. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool*. **27**, 401–410.
Felsenstein, J. 1978b. The number of evolutionary trees. *Syst. Zool*. **27**, 27–33.
Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol*. **17**, 368–376.
Felsenstein, J. 1983. Statistical inference of phylogenies. *J. R. Stat. Soc. A*. **146**, 246–272.
Felsenstein, J. 1985a. Phylogenies and the comparative method. *Am. Nat*. **125**, 1–15.
Felsenstein, J. 1985b. Confidence limits on phylogenies with a molecular clock. *Evolution* **34**, 152–161.
Felsenstein, J. 1985c. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Ann. Rev. Genet*. **22**, 521–565.
Felsenstein, J. 1992. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet. Res*. **59**, 139–147.
Felsenstein, J. 2001a. Taking variation of evolutionary rates between sites into account in inferring phylogenies. *J. Mol. Evol*. **53**, 447–455.
Felsenstein, J. 2001b. The troubled growth of statistical phylogenetics. *Syst. Biol*. **50**, 465–467.
Felsenstein, J. 2004. *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts.
Felsenstein, J., and G. A. Churchill. 1996. A hidden Markov model approach to variation among sites in rate of evolution. *Mol. Biol. Evol*. **13**, 93–104.
Felsenstein, J., and H. Kishino. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Syst. Biol*. **42**, 193–200.
Felsenstein, J., and E. Sober. 1986. Parsimony and likelihood: an exchange. *Syst. Zool*. **35**, 617–626.
Ferguson, T. 1973. Bayesian analysis of some nonparametric problems. *Ann. Stat*. **1**, 209–230.
Ferreira, M. A. R., and M. A. Suchard. 2008. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat*. **36**, 355–368.
Filip, L. C., and N. I. Mundy. 2004. Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates. *Mol. Biol. Evol*. **21**, 1504–1511.
Fisher, R. 1930a. The distribution of gene ratios for rare mutations. *Proc. R. Soc. Edin*. **50**, 205–220.
Fisher, R. 1930b. *The Genetic Theory of Natural Selection*. Clarendon Press, Oxford, UK.
Fisher, R. A. 1970. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
Fitch, W. M. 1970. Distinguishing homologous from analogous proteins. *Syst. Zool*. **19**, 99–113.
Fitch, W. M. 1971a. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool*. **20**, 406–416.
Fitch, W. M. 1971b. Rate of change of concomitantly variable codons. *J. Mol. Evol*. **1**, 84–96.
Fitch, W. M. 1976. Molecular evolutionary clocks. Pp. 160–178 *in* F. J. Ayala, ed. *Molecular Evolution*. Sinauer Associates, Sunderland, Massachusetts.
Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. *Science* **155**, 279–284.
Fitch, W. M., R. M. Bush, C. A. Bender, and N. J. Cox. 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. U.S.A*. **94**, 7712–7718.
Fleissner, R., D. Metzler, and A. von Haeseler. 2005. Simultaneous statistical multiple alignment and phylogeny reconstruction. *Syst. Biol*. **54**, 548–561.
Fletcher, R. 1987. *Practical Methods of Optimization*. Wiley, New York.
Fletcher, W., and Z. Yang. 2009. INDELible: a flexible simulator of biological sequence evolution. *Mol. Biol. Evol*. **26**, 1879–1888.
Fletcher, W., and Z. Yang. 2010. The effect of insertions, deletions and alignment errors on the branch-site test of positive selection. *Mol. Biol. Evol*. **27**, 2257–2267.
Foote, M., J. P. Hunter, C. M. Janis, and J. J. Sepkoski. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* **283**, 1310–1314.
Forsberg, R., and F. B. Christiansen. 2003. A codon-based model of host-specific selection in parasites, with an application to the influenza A virus. *Mol. Biol. Evol*. **20**, 1252–1259.
Foster, P. G. 2004. Modeling compositional heterogeneity. *Syst. Biol*. **53**, 485–495.
Freeland, S. J., and L. D. Hurst. 1998. The genetic code is one in a million. *J. Mol. Evol*. **47**, 238–248.

Friel, N., and A. N. Pettitt. 2008. Marginal likelihood estimation via power posteriors. *J. Roy. Stat. Soc. B* **70**, 589–607.

Frigessi, A., C. R. Hwang, and L. Younes. 1992. Optimal spectral structure of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields. *Ann. Appl. Prob.* **2**, 610–628.

Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and backgroud selection. *Genetics* **147**, 915–925.

Fu, Y. 1994. Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. *Genetics* **138**, 1375–1386.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Fujisawa, T., and T. G. Barraclough. 2013. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* **62**, 707–724.

Fujita, M. K., and A. D. Leaché. 2011. A coalescent perspective on delimiting and naming species: a reply to Bauer et al. *Proc. R. Soc. Lond. B. Biol. Sci.* **278**, 493–495.

Fujita, M. K., A. D. Leaché, F. T. Burbrink et al. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol. Evol.* **27**, 480–488.

Fukami-Kobayashi, K., and Y. Tateno. 1991. Robustness of maximum likelihood tree estimation against different patterns of base substitutions. *J. Mol. Evol.* **32**, 79–91.

Fukami, K., and Y. Tateno. 1989. On the maximum likelihood method for estimating molecular trees: uniqueness of the likelihood point. *J. Mol. Evol.* **28**, 460–464.

Gadagkar, S. R., and S. Kumar. 2005. Maximum likelihood outperforms maximum parsimony even when evolutionary rates are heterotachous. *Mol. Biol. Evol.* **22**, 2139–2141.

Galtier, N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol. Biol. Evol.* **18**, 866–873.

Galtier, N., and M. Gouy. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* **15**, 871–879.

Galtier, N., N. Tourasse, and M. Gouy. 1999. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**, 220–221.

Gascuel, O. 1994. A note on Sattath and Tversky's, Saitou and Nei's, and Studier and Keppler's algorithms for inferring phylogenies from evolutionary distances. *Mol. Biol. Evol.* **11**, 961–963.

Gascuel, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**, 685–695.

Gascuel, O. 2000. On the optimization principle in phylogenetic analysis and the minimum-evolution criterion. *Mol. Biol. Evol.* **17**, 401–405.

Gascuel, O., and M. Steel. 2006. Neighbor-joining revealed. *Mol. Biol. Evol.* **23**, 1997–2000.

Gascuel, O., D. Bryant, and F. Denis. 2001. Strengths and limitations of the minimum evolution principle. *Syst. Biol.* **50**, 621–627.

Gaucher, E. A., and M. M. Miyamoto. 2005. A call for likelihood phylogenetics even when the process of sequence evolution is heterogeneous. *Mol. Phylogenet. Evol.* **37**, 928–931.

Gaucher, E. A., S. Govindarajan, and O. K. Ganesh. 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–707.

Gaut, B. S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. *Evol. Biol.* **30**, 93–120.

Gaut, B. S., and P. O. Lewis. 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. *Mol. Biol. Evol.* **12**, 152–162.

Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**, 398–409.

Gelman, A., and X. L. Meng. 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Stat. Sci.* **13**, 163–185.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* **7**, 457–511.

Gelman, A., G. O. Roberts, and W. R. Gilks. 1996. Efficient Metropolis jumping rules. Pp. 599–607 *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds. *Bayesian Statistics 5*. Oxford University Press, Oxford, UK.

Gelman, S., and G. D. Gelman. 1984. Stochastic relaxation, Gibbs distributions and the Bayes restoration of images. *IEEE Trans. Pattn. Anal. Mach. Intel.* **6**, 721–741.
Georgelis, N., J. R. Shaw, and L. C. Hannah. 2009. Phylogenetic analysis of ADP-glucose pyrophosphorylase subunits reveals a role of subunit interfaces in the allosteric properties of the enzyme. *Plant Physiol.* **151**, 67–77.
Geyer, C. J. 1991. Markov chain Monte Carlo maximum likelihood. Pp. 156–163 *in* E. M. Keramidas, ed. *Computing Science and Statistics: Proc. 23rd Symp. Interface*. Interface Foundation, Fairfax Station.
Geyer, C. J. 1992. Practical Markov chain Monte Carlo. *Stat. Sci.* **7**, 473–511.
Gilks, W. R., S. Richardson, and D. J. Spielgelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
Gill, P. E., W. Murray, and M. H. Wright. 1981. *Practical Optimization*. Academic Press, London.
Gillespie, J. H. 1984. The molecular clock may be an episodic clock. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 8009–8013.
Gillespie, J. H. 1986a. Natural selection and the molecular clock. *Mol. Biol. Evol.* **3**, 138–155.
Gillespie, J. H. 1986b. Rates of molecular evolution. *Ann. Rev. Ecol. Syst.* **17**, 637–665.
Gillespie, J. H. 1991. *The Causes of Molecular Evolution*. Oxford University Press, Oxford, UK.
Gillespie, J. H. 1998. *Population Genetics: a Concise Guide*. John Hopkins University Press, Baltimore, Maryland.
Godsill, S. J. 2001. On the relationship between Markov chain Monte Carlo methods for model uncertainty. *J. Comput. Graph. Stat.* **10**, 230–248.
Gogarten, J. P., H. Kibak, P. Dittrich et al. 1989. Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6661–6665.
Gojobori, T. 1983. Codon substitution in evolution and the 'saturation' of synonymous changes. *Genetics* **105**, 1011–1027.
Gojobori, T., W. H. Li, and D. Graur. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. *J. Mol. Evol.* **18**, 360–369.
Golding, G. B. 1983. Estimates of DNA and protein sequence divergence: an examination of some assumptions. *Mol. Biol. Evol.* **1**, 125–142.
Golding, G. B., and A. M. Dean. 1998. The structural basis of molecular adaptation. *Mol. Biol. Evol.* **15**, 355–369.
Goldman, N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analysis. *Syst. Zool.* **39**, 345–361.
Goldman, N. 1993a. Simple diagnostic statistical tests of models for DNA substitution. *J. Mol. Evol.* **37**, 650–661.
Goldman, N. 1993b. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**, 182–198.
Goldman, N. 1994. Variance to mean ratio, R(t), for Poisson processes on phylogenetic trees. *Mol. Phylogenet. Evol.* **3**, 230–239.
Goldman, N. 1998. Phylogenetic information and experimental design in molecular systematics. *Proc. R. Soc. Lond. B Biol. Sci.* **265**, 1779–1786.
Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736.
Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**, 652–670.
Goldman, N., J. L. Thorne, and D. T. Jones. 1998. Assessing the impact of secondary structure and solvent accessibility on protein evolution. *Genetics* **149**, 445–458.
Goldstein, D. B., and D. D. Pollock. 1994. Least squares estimation of molecular distance–noise abatement in phylogenetic reconstruction. *Theor. Popul. Biol.* **45**, 219–226.
Goldstein, R. A., and D. D. Pollock. 2006. Observations of amino acid gain and loss during protein evolution are explained by statistical bias. *Mol. Biol. Evol.* **23**, 1444–1449.
Goloboff, P. A. 1999. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* **15**, 415–428.
Goloboff, P. A., and D. Pol. 2005. Parsimony and Bayesian phylogenetics. Pp. 148–159 *in* V. A. Albert, ed. *Parsimony, Phylogeny, and Genomics*. Oxford University Press, Oxford, UK.
Golub, G. H., and C. F. Van Loan. 1996. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland.

Gonnet, G. H., M. A. Cohen, and S. A. Benner. 1992. Exhaustive matching of the entire protein sequence database. *Science* **256**, 1443–1445.

Goswami, A., and P. Upchurch. 2010. The dating game: a reply to Heads (2010). *Zool. Scr.* **39**, 406–409.

Götesson, A., J. S. Marshall, D. A. Jones, and A. R. Hardham. 2002. Characterization and evolutionary analysis of a large polygalacturonase gene family in the oomycete pathogen Phytophthora cinnamomi. *Mol. Plant Microbe Interact.* **15**, 907–921.

Grantham, R. 1974. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864.

Graur, D., and W.-H. Li. 2000. *Fundamentals of Molecular Evolution*. Sinauer Associates, Massachusetts.

Graur, D., and W. Martin. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* **20**, 80–86.

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Green, P. J. 2003. Trans-dimensional Markov chain Monte Carlo. Pp. 179–196 *in* P. J. Green, N. L. Hjort, and S. Richardson, eds. *Highly Structured Stochastic Systems*. Oxford University Press, Oxford, UK.

Green, P. J., and X. L. Han. 1992. Metropolis methods, Gaussian proposals and antithetic variables. Pp. 142–164 *in* P. Barone, A. Frigessi, and M. Piccioni, eds. *Stochastic Models, Statistical Methods & Algorithms in Image Analysis*. Springer, New York.

Green, P. J., and S. Richardson. 2001. Modelling heterogeneity with and without the Dirichlet process. *Scand. J. Stat.* **28**, 355–375.

Griffiths, R. C., and S. Tavaré. 1994. Sampling theory for neutral alleles in a varying environment. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **344**, 403–410.

Grimmett, G. R., and D. R. Stirzaker. 1992. *Probability and Random Processes*. Clarendon Press, Oxford.

Gronau, I., M. J. Hubisz, B. Gulko et al. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* **43**, 1031–1034.

Gu, X. 2001. Maximum-likelihood approach for gene family evolution under functional divergence. *Mol. Biol. Evol.* **18**, 453–464.

Gu, X., and W.-H. Li. 1996. A general additive distance with time-reversibility and rate variation among nucleotide sites. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4671–4676.

Gu, X., Y. X. Fu, and W. H. Li. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* **12**, 546–557.

Guindon, S. 2013. From trajectories to averages: an improved description of the heterogeneity of substitution rates along lineages. *Syst. Biol.* **62**, 22–34.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.

Guindon, S., M. Black, and A. Rodrigo. 2006. Control of the false discovery rate applied to the detection of positively selected amino acid sites. *Mol. Biol. Evol.* **23**, 919–926.

Guindon, S., A. G. Rodrigo, K. A. Dyer, and J. P. Huelsenbeck. 2004. Modeling the site-specific variation of selection patterns along lineages. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12957–12962.

Hacking, I. 1965. *Logic of Scientific Inference*. Cambridge University Press, Cambridge, UK.

Haldane, J. B. S. 1931. A note on inverse probability. *Proc. Cam. Phil. Soc.* **28**, 55–61.

Haldane, J. B. S. 1932. *The Causes of Evolution*. Longmans Green & Co., London.

Halpern, A. L., and W. J. Bruno. 1998. Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol. Biol. Evol.* **15**, 910–917.

Hanson-Smith, V., B. Kolaczkowski, and J. W. Thornton. 2010. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol. Biol. Evol.* **27**, 1988–1999.

Harris, H. 1966. Enzyme polymorphism in man. *Proc. R. Soc. Lond. B. Biol. Sci.* **164**, 298–310.

Hartigan, J. A. 1973. Minimum evolution fits to a given tree. *Biometrics* **29**, 53–65.

Hartl, D. L., and A. G. Clark. 1997. *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts.

Hartl, D. L., E. N. Moriyama, and S. A. Sawyer. 1994. Selection intensity for codon bias. *Genetics* **138**, 227–234.

Harvey, P. H., and M. Pagel. 1991. *The Comparative Method in Evlutionary Biology*. Oxford University Press, Oxford, UK.

Harvey, P. H., and A. Purvis. 1991. Comparative methods for explaining adaptations. *Nature* **351**, 619–624.

Hasegawa, M., and M. Fujiwara. 1993. Relative efficiencies of the maximum likelihood, maximum parsimony, and neihbor joining methods for estimating protein phylogeny. *Mol. Phylogenet. Evol*. **2**, 1–5.

Hasegawa, M., and H. Kishino. 1989. Confidence limits on the maximum-likelihood estimate of the Hominoid tree from mitochondrial DNA sequences. *Evolution* **43**, 672–677.

Hasegawa, M., and H. Kishino. 1994. Accuracies of the simple methods for estimating the bootstrap probability of a maximum likelihood tree. *Mol. Biol. Evol*. **11**, 142–145.

Hasegawa, M., J. Adachi, and M. C. Milinkovitch. 1997. Novel phylogeny of whales supported by total molecular evidence. *J. Mol. Evol*. **44**, S117–S120.

Hasegawa, M., Y. Cao, and Z. Yang. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA: replacement/synonymous rate ratio is much higher within species than between species. *Mol. Biol. Evol.* **15**, 1499–1505.

Hasegawa, M., H. Kishino, and N. Saitou. 1991. On the maximum likelihood method in molecular phylogenetics. *J. Mol. Evol*. **32**, 443–445.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol*. **22**, 160–174.

Hasegawa, M., J. L. Thorne, and H. Kishino. 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst*. **78**, 267–283.

Hasegawa, M., T. Yano, and H. Kishino. 1984. A new molecular clock of mitochondrial DNA and the evolution of Hominoids. *Proc. Japan Acad. B*. **60**, 95–98.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their application. *Biometrika* **57**, 97–109.

Haydon, D. T., A. D. Bastos, N. J. Knowles, and A. R. Samuel. 2001. Evidence for positive selection in foot-and-mouth-disease virus capsid genes from field isolates. *Genetics* **157**, 7–15.

Heads, M. 2005. Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics* **21**, 62–78.

Heath, T. A., M. T. Holder, and J. P. Huelsenbeck. 2012. A Dirichlet process prior for estimating lineage-specific substitution rates. *Mol. Biol. Evol*. **29**, 939–955.

Hebert, P. D., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004. Identification of birds through DNA barcodes. *PLoS Biol*. **2**, 1657–1663.

Hedges, S. B., and S. Kumar. 2004. Precision of molecular time estimates. *Trends Genet*. **20**, 242–247.

Hedges, S. B., P. H. Parker, C. G. Sibley, and S. Kumar. 1996. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229.

Hein, J., J. L. Jensen, and C. N. Pedersen. 2003. Recursions for statistical multiple alignment. *Proc. Natl. Acad. Sci. U.S.A*. **100**, 14960–14965.

Hein, J., M. H. Schieriup, and C. Wiuf. 2005. *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press, Oxford, UK.

Hein, J., C. Wiuf, B. Knudsen et al. 2000. Statistical alignment: computational properties, homology testing and goodness-of-fit. *J. Mol. Biol*. **302**, 265–279.

Heled, J., and A. J. Drummond. 2008. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol*. **8**, 289.

Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol*. **27**, 570–580.

Heled, J., and A. J. Drummond. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst. Biol*. **61**, 138–149.

Heled, J., D. Bryant, and A. J. Drummond. 2013. Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC Evol. Biol*. **13**, 44.

Hellberg, M. E., and V. D. Vacquier. 2000. Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. *Mol. Biol. Evol*. **17**, 458–466.

Hendy, M. D. 2005. Hadamard conjugation: an analytical tool for phylogenetics. Pp. 143–177 *in* O. Gascuel, ed. *Mathematics of Evolution and Phylogeny*. Oxford University Press, Oxford, UK.

Hendy, M. D., and D. Penny. 1982. Branch and bound algorithms ro determine minimum-evolution trees. *Math. Biosci*. **60**, 133–142.

Hendy, M. D., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool*. **38**, 297–309.

Henikoff, S., and J. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A*. **89**, 10915–10919.

Hernandez, R. D., S. H. Williamson, and C. D. Bustamante. 2007. Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol. Biol. Evol*. **24**, 1792–1800.

Hey, J. 2010. Isolation with migration models for more than two populations. *Mol. Biol. Evol*. **27**, 905–920.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747–760.

Hickerson, M. J., C. P. Meyer, and C. Moritz. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst. Biol*. **55**, 729–739.

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol*. **42**, 182–192.

Hillis, D. M., J. J. Bull, M. E. White et al. 1992. Experimental phylogenetics: generation of a known phylogeny. *Science* **255**, 589–592.

Ho, S. Y. W., and M. J. Phillips. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol*. **58**, 367–380.

Hobolth, A., L. N. Andersen, and T. Mailund. 2011. On computing the coalescence time density in an isolation-with-migration model with few samples. *Genetics* **187**, 1241–1243.

Hodgkinson, A., and A. Eyre-Walker. 2011. Variation in the mutation rate across mammalian genomes. *Nat. Rev. Genet*. **12**, 756–766.

Hoekstra, H. E., and J. A. Coyne. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016.

Höhna, S., and A. J. Drummond. 2012. Guided tree topology proposals for Bayesian phylogenetic inference. *Syst. Biol*. **61**, 1–11.

Höhna, S., M. Defoin-Platel, and A. J. Drummond. 2008. Clock-constrained tree proposal operators in Bayesian phylogenetic inference. *8th IEEE International Conference on BioInformatics and BioEngineering. Athens (Greece): BIBE*:7.

Holder, M., and P. O. Lewis. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet*. **4**, 275–284.

Holder, M. T., P. O. Lewis, D. L. Swofford, and B. Larget. 2005. Hastings ratio of the LOCAL proposal used in Bayesian phylogenetics. *Syst. Biol*. **54**, 961–965.

Holmes, I. 2005. Using evolutionary expectation maximization to estimate indel rates. *Bioinformatics* **21**, 2294–2300.

Holmes, S. 2003. Bootstrapping phylogenetic trees: theory and methods. *Stat. Sci*. **18**, 241–255.

Horai, S., K. Hayasaka, R. Kondo et al. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. U.S.A*. **92**, 532–536.

Huang, H., and L. L. Knowles. 2009. What is the danger of the anomaly zone for empirical phylogenetics? *Syst. Biol*. **58**, 527–536.

Hudson, R. R. 1983a. Testing the constant-rate neutral alele model with protein sequence data. *Evolution* **37**, 203–217.

Hudson, R. R. 1983b. Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol*. **23**, 183–201.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Pp. 1–44 *in* D. J. Futuyma, and J. D. Antonovics, eds. *Oxford Surveys in Evolutionary Biology*. Oxford University Press, New York.

Hudson, R. R. 2001. Two-locus sampling distributions and their application. *Genetics* **159**, 1805–1817.

Hudson, R. R., and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. *Evolution* **56**, 1557–1565.

Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Huelsenbeck, J. P. 1995a. The performance of phylogenetic methods in simulation. *Syst. Biol.* **44**, 17–48.
Huelsenbeck, J. P. 1995b. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol. Biol. Evol.* **12**, 843–849.
Huelsenbeck, J. P. 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst. Biol.* **47**, 519–537.
Huelsenbeck, J. P. 2002. Testing a covariotide model of DNA substitution. *Mol. Biol. Evol.* **19**, 698–707.
Huelsenbeck, J. P., and P. Andolfatto. 2007. Inference of population structure under a Dirichlet process model. *Genetics* **175**, 1787–1802.
Huelsenbeck, J. P., and J. P. Bollback. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. *Syst. Biol.* **50**, 351–366.
Huelsenbeck, J. P., and K. A. Dyer. 2004. Bayesian estimation of positively selected sites. *J. Mol. Evol.* **58**, 661–672.
Huelsenbeck, J. P., and K. M. Lander. 2003. Frequent inconsistency of parsimony under a simple model of cladogenesis. *Syst. Biol.* **52**, 641–648.
Huelsenbeck, J. P., and R. Nielsen. 1999. Variation in the pattern of nucleotide substitution across sites. *J. Mol. Evol.* **48**, 86–93.
Huelsenbeck, J. P., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.* **53**, 904–913.
Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.
Huelsenbeck, J. P., and M. A. Suchard. 2007. A nonparametric method for accommodating and testing across-site rate variation. *Syst. Biol.* **56**, 975–987.
Huelsenbeck, J. P., M. E. Alfaro, and M. A. Suchard. 2011. Biologically inspired phylogenetic models strongly outperform the no common mechanism model. *Syst. Biol.* **60**, 225–232.
Huelsenbeck, J. P., J. J. Bull, and C. W. Cunningham. 1996. Combining data in phylogenetic analysis. *Trends Ecol. Evol.* **11**, 152–158.
Huelsenbeck, J. P., B. Larget, and M. E. Alfaro. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* **21**, 1123–1133.
Huelsenbeck, J. P., B. Larget, and D. Swofford. 2000a. A compound Poisson process for relaxing the molecular clock. *Genetics* **154**, 1879–1892.
Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback. 2003. Stochastic mapping of morphological characters. *Syst. Biol.* **52**, 131–158.
Huelsenbeck, J. P., B. Rannala, and B. Larget. 2000b. A Bayesian framework for the analysis of cospeciation. *Evolution* **54**, 352–364.
Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310–2314.
Hughes, A. L. 1999. *Adaptive Evolution of Genes and Genomes*. Oxford University Press, Oxford, UK.
Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170.
Hughes, A. L., T. Ota, and M. Nei. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol. Biol. Evol.* **7**, 515–524.
Hurley, I. A., R. L. Mueller, K. A. Dunn et al. 2007. A new time-scale for ray-finned fish evolution. *Proc. R. Soc. Lond. B. Biol. Sci.* **274**, 489–498.
Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.
Ina, Y. 1995. New methods for estimating the numbers of synonymous and nonsynonymous substitutions. *J. Mol. Evol.* **40**, 190–226.
Inoue, J., P. C. H. Donoghue, and Z. Yang. 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* **59**, 74–89.
Issac, N. J. B., J. Mallet, and G. M. Mace. 2004. Taxonomic inflation: its influence on macroecology and conservation. *Trends Ecol. Evol.* **19**, 464–469.

Ivarsson, Y., A. J. Mackey, M. Edalat et al. 2002. Identification of residues in glutathione transferase capable of driving functional diversification in evolution: a novel approach to protein design. *J. Biol. Chem.* **278**, 8733–8738.

Iwabe, N., K. Kuma, M. Hasegawa et al. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9355–9359.

Iwabe, N., Y. Hara, Y. Kumazawa et al. 2005. Sister group relationship of turtles to the bird-crocodilian clade revealed by nuclear DNA-coded proteins 10.1093/molbev/msi075. *Mol. Biol. Evol.* **22**, 810–813.

Jasra, A., C. C. Holmes, and D. A. Stephens. 2005. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Stat. Sci.* **1**, 50–67.

Jeffreys, H. 1935. Some tests of significance, treated by the theory of probability. *Proc. Cam. Phil. Soc.* **31**, 203–222.

Jeffreys, H. 1961. *Theory of Probability*. Oxford University Press, Oxford, UK.

Jennings, W. B., and S. V. Edwards. 2005. Speciational history of Australian grass finches (Poephila) inferred from thirty gene trees. *Evolution* **59**, 2033–2047.

Jensen, J. D., A. Wong, and C. F. Aquadro. 2007. Approaches for identifying targets of positive selection. *Trends Genet.* **23**, 568–577.

Jermann, T. M., J. G. Opitz, J. Stackhouse, and S. A. Benner. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374**, 57–59.

Jermiin, L. S., V. Jayaswal, F. Ababneh, and J. Robinson. 2008. Phylogenetic model evaluation (Chapter 16) *in* J. M. Keith, ed. *Bioinformatics, Volume I: Data, Sequence Analysis, and Evolution*. Humana Press (Springer), Totowa, New Jersey.

Jin, L., and M. Nei. 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis [Erratum in Mol. Biol. Evol. 1990 7, 201]. *Mol. Biol. Evol.* **7**, 82–102.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *CABIOS* **8**, 275–282.

Jordan, G., and N. Goldman. 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol. Biol. Evol.* **29**, 1125–1139.

Jordan, I. K., F. A. Kondrashov, I. A. Adzhubei et al. 2005. A universal trend of amino acid gain and loss in protein evolution. *Nature* **433**, 633–638.

Jukes, T. H. 1987. Transitions, transversions, and the molecular evolutionary clock. *J. Mol. Evol.* **26**, 87–98.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21–123 *in* H. N. Munro, ed. *Mammalian Protein Metabolism*. Academic Press, New York.

Jukes, T. H., and J. L. King. 1979. Evolutionary nucleotide replacements in DNA. *Nature* **281**, 605–606.

Kafatos, F. C., A. Efstratiadis, B. G. Forget, and S. M. Weissman. 1977. Molecular evolution of human and rabbit ß-globin mRNAs. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5618–5622.

Kalbfleisch, J. G. 1985. *Probability and Statistical Inference, Volume 2, Statistical Inference*. Springer-Verlag, New York.

Kalbfleisch, J. G., and D. A. Sprott. 1970. Application of likelihood methods to models involving large numbers of parameters (with discussions). *J. R. Stat. Soc. B.* **32**, 175–208.

Kao, E. P. C. 1997. *An Introduction to Stochastic Processes*. ITP, Belmont, California.

Karlin, S., and H. M. Taylor. 1975. *A First Course in Stochastic Processes*. Academic Press, San Diego, California.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795.

Katoh, K., K. Kuma, and T. Miyata. 2001. Genetic algorithm-based maximum-likelihood analysis for molecular phylogeny. *J. Mol. Evol.* **53**, 477–484.

Keilson, J. 1979. *Markov Chain Models: Rarity and Exponentiality*. Springer-Verlag, New York.

Keller, A., F. Förster, T. Müller et al. 2010. Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. *Biol. Direct.* **5**, 4.

Kelly, C., and J. Rice. 1996. Modeling nucleotide evolution: a heterogeneous rate analysis. *Math. Biosci.* **133**, 85–109.

Kemeny, J. G., and J. L. Snell. 1960. *Finite Markov Chains*. Van Nostrand, Princeton, New Jersey.

Kendall, D. G. 1948. On the generalized birth-and-death process. *Ann. Math. Stat.* **19**, 1–15.

Kidd, K. K., and L. A. Sgaramella-Zonta. 1971. Phylogenetic analysis: concepts and methods. *Am. J. Hum. Genet*. **23**, 235–252.

Kim, J. 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Syst. Biol*. **45**, 363–374.

Kimura, M. 1957. Some problems of stochastic processes in genetics. *Ann. Math. Stat*. **28**, 882–901.

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

Kimura, M. 1977. Prepondence of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol*. **16**, 111–120.

Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. U.S.A*. **78**, 454–458.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kimura, M. 1987. Molecular evolutionary clock and the neutral theory. *J. Mol. Evol*. **26**, 24–33.

Kimura, M., and T. Ohta. 1971a. *Theoretical Topics in Population Genetics*. Princeton University Press, Princeton, New Jersey.

Kimura, M., and T. Ohta. 1971b. Protein polymorphism as a phase of molecular evolution. *Nature* **229**, 467–469.

Kimura, M., and T. Ohta. 1972. On the stochastic model for estimation of mutational distance between homologous proteins. *J. Mol. Evol*. **2**, 87–90.

King, C. E., and T. H. Jukes. 1969. Non-Darwinian evolution. *Science* **164**, 788–798.

Kingman, J. F. C. 1982a. On the genealogy of large populations. *J. Appl. Prob*. **19A**:27–43.

Kingman, J. F. C. 1982b. The coalescent. *Stochastic Process Appl*. **13**, 235–248.

Kingman, J. F. C. 2000. Origins of the coalescent: 1974–1982. *Genetics* **156**, 1461–1463.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science* **220**, 671–680.

Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol*. **29**, 170–179.

Kishino, H., and M. Hasegawa. 1990. Converting distance to time: application to human evolution. *Methods Enzymol*. **183**, 550–570.

Kishino, H., T. Miyata, and M. Hasegawa. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol*. **31**, 151–160.

Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol*. **18**, 352–361.

Kluge, A. G., and J. S. Farris. 1969. Quantitateive phyletics and the evolution of anurans. *Syst. Zool*. **18**, 1–32.

Knoll, A. H., and S. B. Carroll. 1999. Early animal evolution: emerging views from comparative biology and geology. *Science* **284**, 2129–2137.

Knowles, L. L. 2009. Statistical phylogeography. *Ann. Rev. Ecol. Syst*. **40**, 593–612.

Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. *Syst. Biol*. **56**, 887–895.

Knudsen, B., and M. M. Miyamoto. 2001. A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. *Proc. Natl. Acad. Sci. U.S.A*. **98**, 14512–14517.

Knuth, D. E. 1997. *The Art of Computer Programming: Fundamental Algorithms*. Addison-Wesley, Reading, Massachusetts.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet*. **5**, 288–298.

Kolaczkowski, B., and J. W. Thornton. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* **431**, 980–984.

Kolaczkowski, B., and J. W. Thornton. 2008. A mixed branch length model of heterotachy improves phylogenetic accuracy. *Mol. Biol. Evol*. **25**, 1054–1066.

Kondrashov, A. S., and F. A. Kondrashov. 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**, 351–354.

Kong, A., M. L. Frigge, G. Masson et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475.

Kosakovsky Pond, S. L., and S. D. W. Frost. 2005a. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208–1222.

Kosakovsky Pond, S. L., and S. D. W. Frost. 2005b. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol. Biol. Evol.* **22**, 478–485.

Kosakovsky Pond, S. L., and S. V. Muse. 2004. Column sorting: rapid calculation of the phylogenetic likelihood function. *Syst. Biol.* **53**, 685–692.

Kosakovsky Pond, S. L., and S. V. Muse. 2005. Site-to-site variation of synonymous substitution rates. *Mol. Biol. Evol.* **22**, 2375–2385.

Koshi, J. M., and R. A. Goldstein. 1996a. Probabilistic reconstruction of ancestral protein sequences. *J. Mol. Evol.* **42**, 313–320.

Koshi, J. M., and R. A. Goldstein. 1996b. Correlating structure-dependent mutation matrices with physical- chemical properties. *Pac. Symp. Biocomput.*:488–499.

Koshi, J. M., D. P. Mindell, and R. A. Goldstein. 1999. Using physical-chemistry-based substitution models in phylogenetic analyses of HIV-1 subtypes. *Mol. Biol. Evol.* **16**, 173–179.

Kosiol, C., and N. Goldman. 2005. Different versions of the Dayhoff rate matrix. *Mol. Biol. Evol.* **22**, 193–199.

Kosiol, C., I. Holmes, and N. Goldman. 2007. An empirical codon model for protein sequence evolution. *Mol. Biol. Evol.* **24**, 1464–1479.

Kou, S. C., Q. Zhou, and W. H. Wong. 2006. Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Stat.* **34**, 1581–1619.

Kreitman, M. 2000. Methods to detect selection in populations with applications to the human. *Ann. Rev. Genomics Hum. Genet.* **1**, 539–559.

Kreitman, M., and H. Akashi. 1995. Molecular evidence for natural selection. *Ann. Rev. Ecol. Syst.* **26**, 403–422.

Kronmal, R. A., and A. V. Peterson. 1979. On the alias method for generating random variables from a discrete distribution. *Am. Stat.* **33**, 214–218.

Kryazhimskiy, S., and J. B. Plotkin. 2008. The population genetics of dN/dS. *PLoS Genet.* **4**, e1000304.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* **56**, 17–24.

Kubatko, L. S., B. C. Carstens, and L. L. Knowles. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* **25**, 971–973.

Kuhner, M. K., and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates (Erratum in Mol. Biol. Evol. 1995; **12**, 525). *Mol. Biol. Evol.* **11**, 459–468.

Kuhner, M. K., J. Yamato, and J. Felsenstein. 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* **140**, 1421–1430.

Kumar, S. 2005. Molecular clocks: four decades of evolution. *Nat. Rev. Genet.* **6**, 654–662.

Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920.

Kumar, S., and S. Subramanian. 2002. Mutation rate in mammalian genomes. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 803–808.

Kumar, S., K. Tamura, and M. Nei. 2005a. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* **5**, 150–163.

Kumar, S., A. Filipski, V. Swarna et al. 2005b. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18842–18847.

Laird, C. D., B. L. McConaughy, and B. J. McCarthy. 1969. Rate of fixation of nucleotide substitutions in evolution. *Nature* **224**, 149–154.

Lake, J. A. 1994. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1455–1459.

Lakner, C., P. van der Mark, J. P. Huelsenbeck et al. 2008. Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. *Syst. Biol.* **57**, 86–103.

Lang, S. 1987. *Linear Algebra*. Springer-Verlag, New York.

Langley, C. H., and W. M. Fitch. 1974. An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3**, 161–177.
Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* **16**, 750–759.
Lartillot, N. 2006. Conjugate Gibbs sampling for Bayesian phylogenetic models. *J. Comput. Biol.* **13**, 1701–1722.
Lartillot, N., and H. Philippe. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109.
Lartillot, N., and H. Philippe. 2006. Computing Bayes factors using thermodynamic integration. *Syst. Biol.* **55**, 195–207.
Lartillot, N., and H. Philippe. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 1463–1472.
Lartillot, N., H. Brinkmann, and H. Philippe. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* **7 Suppl 1**, S4.
Lartillot, N., T. Lepage, and S. Blanquart. 2009. PhyloBayes 3, a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288.
Le, S. Q., and O. Gascuel. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320.
Le, S. Q., N. Lartillot, and O. Gascuel. 2008. Phylogenetic mixture models for proteins. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3965–3976.
Leaché, A. D. 2009. Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (sceloporus). *Syst. Biol.* **58**, 547–559.
Leaché, A. D., and M. K. Fujita. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proc. R. Soc. Lond. B. Biol. Sci.* **277**, 3071–3077.
Leaché, A. D., and D. G. Mulcahy. 2007. Phylogeny, divergence times and species limits of spiny lizards (Sceloporus magister species group) in western North American deserts and Baja California. *Mol. Ecol.* **16**, 5216–5233.
Leaché, A. D., and B. Rannala. 2011. The accuracy of species tree estimation under simulation: a comparison of methods. *Syst. Biol.* **60**, 126–137.
Leaché, A. D., and T. W. Reeder. 2002. Molecular systematics of the Eastern Fence Lizard (Sceloporus undulatus): a comparison of Parsimony, Likelihood, and Bayesian approaches. *Syst. Biol.* **51**, 44–68.
Leaché, A. D., R. B. Harris, B. Rannala, and Z. Yang. 2014. The influence of gene flow on Bayesian species tree estimation: a simulation study. *Syst. Biol.* **63**, 17–30.
Leaché, A. D., M. S. Koo, C. L. Spencer et al. 2009. Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proc. Natl .Acad. Sci. U.S.A.* **106**, 12418–12423.
LeCam, L. 1953. On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calf. Publ. Stat.* **1**, 277–330.
Lee, M. S. Y. 2000. Tree robustness and clade significance. *Syst. Biol.* **49**, 829–836.
Lee, Y., and J. A. Nelder. 1996. Hierarchical generalized linear models. *J. R. Stat. Soc. B.* **58**, 619–678.
Lehmann, P. 2002. Structure and evolution of plant disease resistance genes. *J. Appl. Genet.* **43**, 403–414.
Leigh, J. W., E. Susko, M. Baumgartner, and A. J. Roger. 2008. Testing congruence in phylogenomic analysis. *Syst. Biol.* **57**, 104–115.
Lemmon, A. R., and M. C. Milinkovitch. 2002. The metapopulation genetic algorithm: an efficient solution for the problem of large phylogeny estimation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10516–10521.
Lemmon, A. R., and E. C. Moriarty. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst. Biol.* **53**, 265–277.
Leonard, T., and J. S. J. Hsu. 1999. *Bayesian Methods*. Cambridge University Press, Cambridge, UK
Lepage, T., D. Bryant, H. Philippe, and N. Lartillot. 2007. A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* **24**, 2669–2680.
Letsch, H., and K. Kjer. 2011. Potential pitfalls of modelling ribosomal RNA data in phylogenetic tree reconstruction: evidence from case studies in the Metazoa. *BMC Evol. Biol.* **11**, 146.

Lewis, P. O. 1998. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Mol. Biol. Evol.* **15**, 277–283.
Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925.
Lewis, P. O., M. T. Holder, and K. E. Holsinger. 2005. Polytomies and Bayesian phylogenetic inference. *Syst. Biol.* **54**, 241–253.
Lewontin, R. 1989. Inferring the number of evolutionary events from DNA coding sequence differences. *Mol. Biol. Evol.* **6**, 15–32.
Lewontin, R. C., and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics* **54**, 595–609.
Li, S., D. Pearl, and H. Doss. 2000. Phylogenetic tree reconstruction using Markov chain Monte Carlo. *J. Am. Stat. Assoc.* **95**, 493–508.
Li, W.-H. 1976. Distribution of nucleotide differences betwen two randomly chosen cistrons in a subdivided population: the finite island model. *Theor. Popul. Biol.* **10**, 303–308.
Li, W.-H. 1986. Evolutionary change of restriction cleavage sites and phylogenetic inference. *Genetics* **113**, 187–213.
Li, W.-H. 1989. A statistical test of phylogenies estimated from sequence data. *Mol. Biol. Evol.* **6**, 424–435.
Li, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**, 96–99.
Li, W.-H. 1997. *Molecular Evolution*. Sinauer Associates, Massachusetts.
Li, W.-H., and M. Gouy. 1991. Statistical methods for testing molecular phylogenies. Pp. 249–277 *in* M. Miyamoto, and J. Cracraft, eds. *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, Oxford, UK.
Li, W.-H., and M. Tanimura. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* **326**, 93–96.
Li, W.-H., M. Tanimura, and P. M. Sharp. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* **25**, 330–342.
Li, W.-H., C.-I. Wu, and C.-C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**, 150–174.
Li, W. L. S., and A. J. Drummond. 2012. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Mol. Biol. Evol.* **29**, 751–761.
Liang, F. 2005. Generalized Wang–Landau algorithm for Monte Carlo computation. *J. Am. Stat. Assoc.* **100**, 1311–1327.
Liang, F., and W. H. Wong. 2001. Real parameter evolutionary Monte Carlo with applications in Bayesian mixture models. *J. Am. Stat. Assoc.* **96**, 653–666.
Liang, F., C. Liu, and R. J. Carroll. 2007. Stochastic approximation in Monte Carlo computation. *J. Am. Stat. Assoc.* **102**, 305–320.
Liang, F., C. Liu, and R. J. Carroll. 2010. *Advanced Markov chain Monte Carlo: Learning from Past Samples*. Wiley, New York.
Liberles, D. A. 2009. *Ancestral Sequence Reconstruction*. Oxford University Press, New York.
Libertini, G., and A. Di Donato. 1994. Reconstruction of ancestral sequences by the inferential method, a tool for protein engineering studies. *J. Mol. Evol.* **39**, 219–229.
Lindley, D. V. 1957. A statistical paradox. *Biometrika* **44**, 187–192.
Lindley, D. V. 1962. Discussion on 'Confidence sets for the mean of a multivariate normal distribution' by C. Stein. *J. R. Stat. Soc. B.* **24**, 265–296.
Lindley, D. V., and L. D. Phillips. 1976. Inference for a Bernoulli process (a Bayesian view). *Am. Stat.* **30**, 112–119.
Lindsey, J. K. 1974a. Construction and comparison of statistical models. *J. R. Stat. Soc. B.* **36**, 418–425.
Lindsey, J. K. 1974b. Comparison of probability distributions. *J. R. Stat. Soc. B.* **36**, 38–47.
Linhart, H. 1988. A test whether two AIC's differ significantly. *S. Afr. Stat. J.* **22**, 153–161.
Little, R. A. J., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. Wiley, New York.
Liu, J. S. 2001. *Monte Carlo Strategies in Scientific Computing*. Springer, New York.

Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* **24**, 2542–2543.

Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst. Biol.* **56**, 504–514.

Liu, L., L. Yu, and S. V. Edwards. 2010a. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302.

Liu, L., L. Yu, and D. K. Pearl. 2010b. Maximum tree: a consistent estimator of the species tree. *J. Math. Biol.* **60**, 95–106.

Liu, L., L. Yu, D. K. Pearl, and S. V. Edwards. 2009a. Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* **58**, 468–477.

Liu, L., L. Yu, L. Kubatko et al. 2009b. Coalescent methods for estimating phylogenetic trees. *Mol. Phylogenet. Evol.* **53**, 320–328.

Lockhart, P., P. Novis, B. G. Milligan et al. 2006. Heterotachy and tree building: a case study with plastids and Eubacteria. *Mol. Biol. Evol.* **23**, 40–45.

Lockhart, P. J., M. A. Steel, M. D. Hendy, and D. Penny. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**, 605–612.

Lohse, K., and N. H. Barton. 2011. A general method for calculating likelihoods under the coalescent process. *Genetics.* **189**, 977–987.

Löytynoja, A., and N. Goldman. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10557–10562.

Löytynoja, A., and N. Goldman. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635.

Lunter, G., I. Miklos, A. Drummond et al. 2005. Bayesian coestimation of phylogeny and sequence alignment. *BMC Bioinformatics* **6**, 83.

Lynch, M. 2010. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 961–968.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155.

Maddison, D. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Syst. Zool.* **33**, 83–103.

Maddison, D. R., and W. P. Maddison. 2000. *MacClade 4, Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Inc., Sunderland, Massachusettes.

Maddison, W. P. 1995. Calculating the probability distributions of ancestral states reconstructed by parsimony on phylogenetic trees. *Syst. Biol.* **44**, 474–481.

Maddison, W. P. 1997. Gene trees in species trees. *Syst. Biol.* **46**, 523–536.

Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* **55**, 21–30.

Maddison, W. P., and D. R. Maddison. 1982. *MacClade: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Inc., Sunderland, Massachusettes.

Magallon, S., K. W. Hilu, and D. Quandt. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am. J. Bot.* **100**, 556–573.

Makova, K. D., M. Ramsay, T. Jenkins, and W. H. Li. 2001. Human DNA sequence variation in a 6.6-kb region containing the melanocortin 1 receptor promoter. *Genetics* **158**, 1253–1268.

Malcolm, B. A., K. P. Wilson, B. W. Matthews et al. 1990. Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature* **345**, 86–89.

Mallet, J. 2013. Concepts of species. Pp. 679–691 *in* S. A. Levin, ed. *Encyclopedia of Biodiversity*. Academic Press, Massachusetts.

Mallick, S., S. Gnerre, P. Muller, and D. Reich. 2010. The difficulty of avoiding false positives in genome scans for natural selection. *Genome Res.* **19**, 922–933.

Margoliash, E. 1963. Primary structure and evolution of cytochrome c. *Proc. Natl. Acad. Sci. U.S.A.* **50**, 672–679.

Marinari, E., and G. Parisi. 1992. Simulated tempering: a new Monte Carlo scheme. *Europhys. Lett.* **19**, 451–458.

Maritz, J. S., and T. Lwin. 1989. *Empirical Bayes Methods*. Chapman and Hall, London.

Marko, P. B. 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Mol. Biol. Evol.* **19**, 2005–2021.

Marsaglia, G., and W. W. Tsang. 2000. A simple method for generating gamma variables. *ACM Trans. Math. Soft.* **26**, 363–372.

Marshall, D. C. 2010. Cryptic failure of partitioned Bayesian phylogenetic analyses: lost in the land of long trees. *Syst. Biol.* **59**, 108–117.

Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4087–4091.

Massingham, T., and N. Goldman. 2005. Detecting amino acid sites under positive selection and purifying selection. *Genetics* **169**, 1753–1762.

Massingham, T., and N. Goldman. 2007. Statistics of the log-det estimator. *Mol. Biol. Evol.* **24**, 2277–2285.

Maston, G. A., and M. Ruvolo. 2002. Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol. Biol. Evol.* **19**, 320–335.

Mateiu, L. M., and B. Rannala. 2006. Inferring complex DNA substitution processes on phylogenies using uniformization and data augmentation. *Syst. Biol.* **55**, 259–269.

Mau, B., and M. A. Newton. 1997. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. *J. Comput. Graph. Stat.* **6**, 122–131.

Mau, B., M. A. Newton, and B. Larget. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* **55**, 1–12.

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. *Genet. Res. (Camb.)* **23**, 23–35.

Mayr, E. 1942. *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Columbia University Press, New York.

Mayrose, I., N. Friedman, and T. Pupko. 2005. A gamma mixture model better accounts for among site rate heterogeneity. *Bioinformatics* **21**, 151–158.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in Drosophila. *Nature* **351**, 652–654.

McGuire, G., M. C. Denham, and D. J. Balding. 2001. Models of sequence evolution for DNA sequences containing gaps. *Mol. Biol. Evol.* **18**, 481–490.

McVean, G. A., and B. Charlesworth. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929–944.

McVean, M., P. Awadalla, and P. Fearnhead. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241.

Mengersen, K. L., and R. L. Tweedie. 1996. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Stat.* **24**, 101–121.

Meredith, R. W., J. E. Janecka, J. Gatesy et al. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524.

Messier, W., and C.-B. Stewart. 1997. Episodic adaptive evolution of primate lysozymes. *Nature* **385**, 151–154.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth et al. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.

Meyer, A., T. D. Kocher, P. Basasibwaki, and A. C. Wilson. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347**, 550–553.

Minin, V. N., and M. A. Suchard. 2008. Fast, accurate and simulation-free stochastic mapping. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3985–3995.

Minin, V. N., E. W. Bloomquist, and M. A. Suchard. 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25**, 1459–1471.

Mira, A. 2001. Ordering and improving the performance of Monte Carlo Markov chains. *Stat. Sci.* **16**, 340–350.

Miyata, T., and T. Yasunaga. 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its applications. *J. Mol. Evol.* **16**, 23–36.

Miyata, T., S. Miyazawa, and T. Yasunaga. 1979. Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.* **12**, 219–236.

Moler, C., and C. F. Van Loan. 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* **45**, 3–49.

Mooers, A. Ø., and D. Schluter. 1999. Reconstructing ancestor states with maximum likelihood: support for one- and two-rate models. *Syst. Biol*. **48**, 623–633.

Morgan, G. J. 1998. Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock. *J. Hist. Biol*. **31**, 155–178.

Moriyama, E. N., and J. R. Powell. 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *J. Mol. Evol*. **45**, 378–391.

Mossel, E., and S. Roch. 2010. Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. *IEEE/ACM Trans. Comput. Biol. Bioinform*. **7**, 166–171.

Mossel, E., and E. Vigoda. 2005. Phylogenetic MCMC algorithms are misleading on mixtures of trees. *Science* **309**, 2207–2209.

Mulcahy, D. G., B. P. Noonan, T. Moss et al. 2012. Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Mol. Phylogenet. Evol*. **65**, 974–991.

Müller, T., and M. Vingron. 2000. Modeling amino acid replacement. *J. Comput. Biol*. **7**, 761–776.

Muse, S. V. 1996. Estimating synonymous and nonsynonymous substitution rates. *Mol. Biol. Evol*. **13**, 105–114.

Muse, S. V., and B. S. Gaut. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol*. **11**, 715–724.

Muse, S. V., and B. S. Gaut. 1997. Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. *Genetics* **146**, 393–399.

Muse, S. V., and B. S. Weir. 1992. Testing for equality of evolutionary rates. *Genetics* **132**, 269–276.

Nachman, M. W., S. Boyer, and C. F. Aquadro. 1996. Non-neutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. *Proc. Natl. Acad. Sci. U.S.A*. **91**, 6364–6368.

Nakamura, K., T. Oshima, T. Morimoto et al. 2011. Sequence-specific error profile of Illumina sequencers. *Nucl. Acids Res*. **39**, e90.

Nath, H. B., and R. C. Griffiths. 1993. The coalescent in two colonies with symmetric migration. *J. Math. Biol*. **31**, 841–852.

Neal, R. M. 2001. Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Stat*. **9**, 249–265.

Needleman, S. G., and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol*. **48**, 443–453.

Nei, M. 1977. Standard error of immunological dating of evolutionary time. *J. Mol. Evol*. **9**, 203–211.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M. 1996. Phylogenetic analysis in molecular evolutionary genetics. *Ann. Rev. Genet*. **30**, 371–403.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol*. **3**, 418–426.

Nei, M., and L. Jin. 1989. Variances of the average numbers of nucleotide substitutions within and between populations. *Mol. Biol. Evol*. **6**, 290–300.

Nei, M., and S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford, UK.

Nei, M., S. Kumar, and K. Takahashi. 1998. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small [In Process Citation]. *Proc. Natl. Acad. Sci. U.S.A*. **95**, 12390–12397.

Nei, M., J. C. Stephens, and N. Saitou. 1985. Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. *Mol. Biol. Evol*. **2**, 66–85.

Newton, M. A. 1996. Bootstrapping phylogenies: large deviations and dispersion effects. *Biometrika* **83**, 315–328.

Newton, M. A., and A. E. Raftery. 1994. Approximating Bayesian inference with the weigthed likelihood bootstrap. *J. R. Stat. Soc. B* **56**, 3–48.

Neyman, J. 1971. Molecular studies of evolution: a source of novel statistical problems. Pp. 1–27 *in* S. S. Gupta, and J. Yackeleds., *Statistical Decision Theory and Related Topics*. Academic Press, New York.

Nichols, R. 2001. Gene trees and species trees are not the same. *Trends Ecol. Evol.* **16**, 358–364.

Nielsen, R. 1997. Site-by-site estimation of the rate of substitution and the correlation of rates in mitochondrial DNA. *Syst. Biol.* **46**, 346–353.

Nielsen, R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**, 641–647.

Nielsen, R. 2002. Mapping mutations on phylogenies. *Syst. Biol.* **51**, 729–739.

Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**, 885–896.

Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936.

Nixon, K. C. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15**, 407–414.

Nordborg, M. 2007. Coalescent theory. Pp. 843–877 *in* D. Balding, M. Bishop, and C. Cannings, eds. *Handbook of Statistical Genetics*. Wiley, New York..

Norris, J. R. 1997. *Markov Chains*. Cambridge University Press, Cambridge, UK.

Notohara, M. 1990. The coalescent and the genealogical process in geographically structured populations. *J. Math. Biol.* **29**, 59–75.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* **53**, 47–67.

O'Hagan, A., and J. Forster. 2004. *Kendall's Advanced Theory of Statistics: Bayesian Inference*. Arnold, London.

O'Meara, B. C. 2010. New heuristic methods for joint species delimitation and species tree inference. *Syst. Biol.* **59**, 59–73.

Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.

Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98.

Ohta, T. 1992. Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. *Genetics* **130**, 917–923.

Ohta, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **40**, 56–63.

Ohta, T. 2002. Near-neutrality in evolution of genes and gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16134–16137.

Ohta, T., and J. H. Gillespie. 1996. Development of neutral and nearly neutral theories. *Theor. Popul. Biol.* **49**, 128–142.

Ohta, T., and M. Kimura. 1971. On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* **1**, 18–25.

Ohta, T., and H. Tachida. 1990. Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* **126**, 219–229.

Oliver, J. C. 2013. Microevolutionary processes generate phylogenomic discordance at ancient divergences. *Evolution* **67**, 1823–1830.

Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek. 1994. fastDNAML: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* **10**, 41–48.

Opgen-Rhein, R., L. Fahrmeir, and K. Strimmer. 2005. Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo. *BMC Evol. Biol.* **5**, 6.

Osawa, S., and T. H. Jukes. 1989. Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* **28**, 271–278.

Page, R. D., and M. A. Charleston. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Mol. Phylogenet. Evol.* **7**, 231–240.

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

Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* **48**, 612–622.

Pagel, M., and A. Meade. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* **53**, 571–581.
Pagel, M., and A. Meade. 2008. Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3955–3964.
Pagel, M., A. Meade, and D. Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684.
Palumbi, S. R. 1994. Genetic divergence, reproductive isolation and marine speciation. *Ann. Rev. Ecol. Syst.* **25**, 547–572.
Pamilo, P., and N. O. Bianchi. 1993. Evolution of the *Zfx* and *Zfy* genes – rates and interdependence between the genes. *Mol. Biol. Evol.* **10**, 271–281.
Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**, 568–583.
Papadopoulou, A., I. Anastasiou, and A. P. Vogler. 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol. Biol. Evol.* **27**, 1659–1672.
Parham, J., P. Donoghue, C. Bell et al. 2012. Best practices for applying paleontological data to molecular divergence dating analyses. *Syst. Biol.* **61**, 346–359.
Paterson, S., T. Vogwill, A. Buckling et al. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature* **464**, 275–278.
Pauling, L., and E. Zuckerkandl. 1963. Chemical paleogenetics: molecular "restoration studies" of extinct forms of life. *Acta Chem. Scand.* **17**, S9–S16.
Pauplin, Y. 2000. Direct calculation of a tree length using a distance matrix. *J. Mol. Evol.* **51**, 41–47.
Penny, D., and M. D. Hendy. 1985. The use of tree comparison metrics. *Syst. Zool.* **34**, 75–82.
Perler, F., A. Efstratiadis, P. Lomedica et al. 1980. The evolution of genes: the chicken preproinsulin gene. *Cell* **20**, 555–566.
Perna, N. T., and T. D. Kocher. 1995. Unequal base frequencies and the estimation of substitution rates. *Mol. Biol. Evol.* **12**, 359–361.
Peskun, P. H. 1973. Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60**, 607–612.
Philippe, H., Y. Zhou, H. Brinkmann et al. 2005. Heterotachy and long-branch attraction in phylogenetics. *BMC Evol. Biol.* **5**, 50.
Pickett, K. M., and C. P. Randle. 2005. Strange Bayes indeed: uniform topological priors imply non-uniform clade priors. *Mol. Phylogenet. Evol.* **34**, 203–211.
Polley, S. D., and D. J. Conway. 2001. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics* **158**, 1505–1512.
Pons, J., T. G. Barraclough, J. Gomez-Zurita et al. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* **55**, 595–609.
Posada, D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256.
Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Informtaion Criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53**, 793–808.
Posada, D., and K. Crandall. 2001. Simple (wrong) models for complex trees: a case from retroviridae. *Mol. Biol. Evol.* **18**, 271–275.
Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.
Prince, V. E., and F. B. Pickett. 2002. Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* **3**, 827–837.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
Pupko, T., I. Pe'er, R. Shamir, and D. Graur. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* **17**, 890–896.
Pupko, T., D. Huchon, Y. Cao et al. 2002a. Combining multiple data sets in a likelihood analysis: which models are the best? *Mol. Biol. Evol.* **19**, 2294–2307.
Pupko, T., I. Pe'er, M. Hasegawa et al. 2002b. A branch-and-bound algorithm for the inference of ancestral amino-acid sequences when the replacement rate varies among sites: application to the evolution of five gene families. *Bioinformatics* **18**, 1116–1123.

Pybus, O. G., A. Rambaut, and P. H. Harvey. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* **155**, 1429–1437.

Raaum, R. L., K. N. Sterner, C. M. Noviello et al. 2005. Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J. Hum. Evol.* **48**, 237–257.

Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-comptemporaneous sequences into maximum likelihood phylogenetics. *Bioinformatics* **16**, 395–399.

Rambaut, A., and L. Bromham. 1998. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**, 442–448.

Rambaut, A., and N. C. Grassly. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *CABIOS* **13**, 235–238.

Rand, D., M. Dorfsman, and L. Kann. 1994. Neutral and nonneutral evolution of *Drosophila* mitochondrial DNA. *Genetics* **138**, 741–756.

Rand, D. M., D. M. Weinreich, and B. O. Cezairliyan. 2000. Neutrality tests of conservative-radical amino acid changes in nuclear- and mitochondrially-encoded proteins. *Gene* **261**, 115–125.

Rannala, B. 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* **51**, 754–760.

Rannala, B., and Z. Yang. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* **43**, 304–311.

Rannala, B., and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656.

Rannala, B., and Z. Yang. 2007. Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **56**, 453–466.

Rannala, B., and Z. Yang. 2013. Improved reversible jump algorithms for Bayesian species delimitation. *Genetics* **194**, 245–253.

Rannala, B., T. Zhu, and Z. Yang. 2012. Tail paradox, partial identifiability and influential priors in Bayesian branch length inference. *Mol. Biol. Evol.* **29**, 325–335.

Ranwez, V., and O. Gascuel. 2002. Improvement of distance-based phylogenetic methods by a local maximum likelihood approach using triplets. *Mol. Biol. Evol.* **19**, 1952–1963.

Redelings, B. D., and M. A. Suchard. 2005. Joint Bayesian estimation of alignment and phylogeny. *Syst. Biol.* **54**, 401–418.

Reeves, J. H. 1992. Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. *J. Mol. Evol.* **35**, 17–31.

Ren, F., H. Tanaka, and Z. Yang. 2005. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. *Syst. Biol.* **54**, 808–818.

Ren, F., H. Tanaka, and Z. Yang. 2009. A likelihood look at the supermatrix-supertree controversy. *Gene* **441**, 119–125.

Ripley, B. 1987. *Stochastic Simulation*. Wiley, New York.

Roach, J. C., G. Glusman, A. F. A. Smit et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639.

Robbins, H. 1955. An empirical Bayes approach to statistics. *Proc. 3rd Berkeley Symp. Math. Stat. Prob.* **1**, 157–164.

Robbins, H. 1983. Some thoughts on empirical Bayes estimation. *Ann. Stat.* **1**, 713–723.

Robert, C. P., and G. Casella. 2004. *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

Roberts, G. O., and R. L. Tweedie. 1996. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95–110.

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147.

Rocha, E. P., J. M. Smith, L. D. Hurst et al. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J. Theor. Biol.* **239**, 226–235.

Rodrigo, A. G., and J. Felsenstein. 1999. Coalescent approaches to HIV population genetics. pp. 233–271 *in* K. Crandall, ed. *Molecular Evolution of HIV*. Johns Hopkins University Press, Baltimore, Maryland.

Rodrigue, N., H. Philippe, and N. Lartillot. 2008. Uniformization for sampling realizations of Markov processes: applications to Bayesian implementations of codon substitution models. *Bioinformatics* **24**, 56–62.

Rodrigue, N., H. Philippe, and N. Lartillot. 2010. Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4629–4634.

Rodriguez-Trelles, F., R. Tarrio, and F. J. Ayala. 2003. Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13413–13417.

Rodriguez, F., J. F. Oliver, A. Marin, and J. R. Medina. 1990. The general stochastic model of nucleotide substitutions. *J. Theor. Biol.* **142**, 485–501.

Rogers, J. S. 1997. On the consistency of maximum likelihood estimation of phylogenetic trees from nucleotide sequences. *Syst. Biol.* **46**, 354–357.

Rogers, J. S., and D. L. Swofford. 1998. A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. *Syst. Biol.* **47**, 77–89.

Rogers, J. S., and D. L. Swofford. 1999. Multiple local maxima for likelihoods of phylogenetic trees: a simulation study. *Mol. Biol. Evol.* **16**, 1079–1085.

Rokas, A., D. Kruger, and S. B. Carroll. 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science* **310**, 1933–1938.

Ronquist, F. 1998. Fast Fitch-parsimony algorithms for large data sets. *Cladistics* **14**, 387–400.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3, Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Ronquist, F., S. Klopfstein, L. Vilhelmsen et al. 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst. Biol.* **61**, 973–999.

Ronquist, F., M. Teslenko, P. van der Mark et al. 2012b. MrBayes 3.2, efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542.

Rosenberg, N. A., and M. Nordborg. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet*. **3**, 380–390.

Ross, R. 1997. *Simulation*. Academic Press, London.

Ross, S. 1996. *Stochastic Processes*. Springer-Verlag, New York.

Rota-Stabelli, O., Z. Yang, and M. Telford. 2009. MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. *Mol. Phylogenet. Evol.* **52**, 268–272.

Roth, C., M. J. Betts, P. Steffansson et al. 2005. The Adaptive Evolution Database (TAED): a phylogeny based tool for comparative genomics. *Nucl. Acids Res.* **33**, D495–D497.

Rubin, D. B., and N. Schenker. 1986. Efficiently simulating the coverage properties of interval estimates. *Appl. Stat*. **35**, 159–167.

Rubinstein, N. D., I. Mayrose, A. Doron-Faigenboim, and T. Pupko. 2011. Evolutionary models accounting for layers of selection in protein coding genes and their impact on the inference of positive selection. *Mol. Biol. Evol.* **28**, 3297–3308.

Russo, C. A., N. Takezaki, and M. Nei. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* **13**, 525–536.

Rzhetsky, A. 1995. Estimating substitution rates in ribosomal RNA genes. *Genetics* **141**, 771–783.

Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. *Mol. Biol. Evol.* **9**, 945–967.

Rzhetsky, A., and M. Nei. 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol. Biol. Evol.* **10**, 1073–1095.

Rzhetsky, A., and M. Nei. 1994. Unbiased estimates of the number of nucleotide substitutions when substitution rate varies among different sites. *J. Mol. Evol.* **38**, 295–299.

Rzhetsky, A., and M. Nei. 1995. Tests of applicability of several substitution models for DNA sequence data. *Mol. Biol. Evol.* **12**, 131–151.

Rzhetsky, A., and T. Sitnikova. 1996. When is it safe to use an oversimplified substitution model in tree-making? *Mol. Biol. Evol.* **13**, 1255–1265.

Saitou, N. 1988. Property and efficiency of the maximum likelihood method for molecular phylogeny. *J. Mol. Evol.* **27**, 261–273.

Saitou, N., and M. Nei. 1986. The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. *J. Mol. Evol.* **24**, 189–204.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Salter, L. A. 2001. Complexity of the likelihood surface for a large DNA dataset. *Syst. Biol*. **50**, 970–978.

Salter, L. A., and D. K. Pearl. 2001. Stochastic search strategy for estimation of maximum likelihood phylogenetic trees. *Syst. Biol*. **50**, 7–17.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol*. **14**, 1218–1232.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol*. **19**, 101–109.

Sanderson, M. J., and J. Kim. 2000. Parametric phylogenetics? *Syst. Biol*. **49**, 817–829.

Sankoff, D. 1975. Minimal mutation trees of sequences. *SIAM J. Appl. Math*. **28**, 35–42.

Sarich, V. M., and A. C. Wilson. 1967. Rates of albumin evolution in primates. *Proc. Natl. Acad. Sci. U.S.A*. **58**, 142–148.

Sarich, V. M., and A. C. Wilson. 1973. Generation time and genomic evolution in primates. *Science* **179**, 1144–1147.

Satta, Y., M. Hickerson, H. Watanabe et al. 2004. Ancestral population sizes and species divergence times in the primate lineage on the basis of intron and BAC end sequences. *J. Mol. Evol*. **59**, 478–487.

Saunders, I. W., S. Tavaré, and G. A. Watterson. 1984. On the genealogy of nested subsamples from a haploid population. *Adv. Appl. Prob*. **16**, 471–491.

Savage, L. J. 1962. *The Foundations of Statistical Inference*. Metheun & Co., London.

Savill, N. J., D. C. Hoyle, and P. G. Higgs. 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics* **157**, 399–411.

Sawyer, K. R. 1984. Multiple hypothesis testing. *J. R. Stat. Soc. B*. **46**, 419–424.

Sawyer, S. A., and D. L. Hartl. 1992. Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176.

Sawyer, S. L., L. I. Wu, M. Emerman, and H. S. Malik. 2005. Positive selection of primate TRIM5a identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. U.S.A*. **102**, 2832–2837.

Scally, A., J. Y. Dutheil, L. W. Hillier et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169–175.

Scheffler, K., and C. Seoighe. 2005. A Bayesian model comparison approach to inferring positive selection. *Mol. Biol. Evol*. **22**, 2531–2540.

Schluter, D. 1995. Uncertainty in ancient phylogenies. *Nature* **377**, 108–110.

Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, UK.

Schmidt-Lebuhn, A. N., J. M. de Vos, B. Keller, and E. Conti. 2012. Phylogenetic analysis of Primula section Primula reveals rampant non-monophyly among morphologically distinct species. *Mol. Phylogenet. Evol*. **65**, 23–34.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504.

Schneider, A., G. M. Cannarozzi, and G. Gonnet. 2005. Empirical codon substitution matrix. *BMC Bioinformatics* **6**, 134.

Schneider, A., A. Souvorov, N. Sabath et al. 2009. Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biol. Evol*. **2009**, 114–118.

Schoeniger, M., and A. von Haeseler. 1994. A stochastic model for the evolution of autocorrelated DNA sequences. *Mol. Phylogenet. Evol*. **3**, 240–247.

Schott, J. R. 1997. *Matrix Analysis for Statistics*. Wiley, New York.

Schultz, T. R., and G. A. Churchill. 1999. The role of subjectivity in reconstructing ancestral character states: a Bayesian approach to unknown rates, states, and transformation asymmetries. *Syst. Biol*. **48**, 651–664.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat*. **6**, 461–464.

Self, S. G., and K.-Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc*. **82**, 605–610.

Semple, C., and M. Steel. 2003. *Phylogenetics*. Oxford University Press, New York.

Seo, T. K., H. Kishino, and J. L. Thorne. 2004. Estimating absolute rates of synonymous and non-synonymous nucleotide substitution in order to characterize natural selection and date species divergences. *Mol. Biol. Evol.* **21**, 1201–1213.

Shackelton, L. A., C. R. Parrish, U. Truyen, and E. C. Holmes. 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 379–384.

Shan, H., L. Zahn, S. Guindon et al. 2009. Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. *Mol. Biol. Evol.* **26**, 2229–2244.

Shapiro, B., A. Rambaut, and A. J. Drummond. 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23**, 7–9.

Sharp, P. M. 1997. In search of molecular Darwinism. *Nature* **385**, 111–112.

Shaw, K. L. 1998. Species and the diversity of natural groups. pp. 44–56 *in* D. J. Howard, and S. J. Berlocher, eds. *Endless Forms: Species and Speciation*. Oxford University Press, Oxford, U.K.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114–1116.

Shimodaira, H., and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247.

Shindyalov, I. N., N. A. Kolchanov, and C. Sander. 1994. Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Eng.* **7**, 349–358.

Shriner, D., D. C. Nickle, M. A. Jensen, and J. I. Mullins. 2003. Potential impact of recombination on sitewise approaches for detecting positive natural selection. *Genet. Res.* **81**, 115–121.

Siddall, M. E. 1998. Success of parsimony in the four-taxon case: long branch repulsion by likelihood in the Farris zone. *Cladistics* **14**, 209–220.

Siepel, A., and D. Haussler. 2004. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* **21**, 468–488.

Sievers, F., A. Wilm, D. Dineen et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Simes, R. J. 1986. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751–754.

Simmons, M. P., K. M. Pickett, and M. Miya. 2004. How meaningful are Bayesian support values? *Mol. Biol. Evol.* **21**, 188–199.

Simonsen, K. L., G. A. Churchill, and C. F. Aquadro. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429.

Singh, N. D., P. F. Arndt, A. G. Clark, and C. F. Aquadro. 2009. Strong evidence for lineage and sequence specificity of substitution rates and patterns in Drosophila. *Mol. Biol. Evol.* **26**, 1591–1605.

Sites, J. W., and J. C. Marshall. 2004. Delimiting species: a renaissance issue in systematic biology. *Trends Ecol. Evol.* **18**, 462–470.

Sitnikova, T., A. Rzhetsky, and M. Nei. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Mol. Biol. Evol.* **12**, 319–333.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genet. Res.* **58**, 167–175.

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562.

Sneath, P. H. A. 1962. The construction of taxonomic groups. Pp. 289–332 *in* G. C. Ainsworth, and P. H. A. Sneath, eds. *Microbial Classification*. Cambridge University Press, Cambridge, UK.

Sober, E. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference*. MIT Press, Cambridge, Massachusetts.

Sober, E. 2004. The contest between parsimony and likelihood. *Syst. Biol.* **53**, 644–653.

Sokal, A. D. 1989. *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*. Lecture Notes for the Cours de Troisieme Cycle de la Physique en Suisse Romande, Lausanne, Switzerland (June 1989).

Sourdis, J., and M. Nei. 1988. Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Mol. Biol. Evol*. **5**, 298–311.
Spencer, M., E. Susko, and A. J. Roger. 2005. Likelihood, parsimony, and heterogeneous evolution. *Mol. Biol. Evol*. **22**, 1161–1164.
Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci. U.S.A*. **100**, 1056–1061.
Stackhouse, J., S. R. Presnell, G. M. McGeehan et al. 1990. The ribonuclease from an ancient bovid ruminant. *FEBS Lett*. **262**, 104–106.
Stadler, T. 2010. Sampling-through-time in birth-death trees. *J. Theor. Biol*. **267**, 396–404.
Stadler, T. 2013. How can we improve accuracy of macroevolutionary rate estimates. *Syst. Biol*. **62**, 321–329.
Stadler, T., and Z. Yang. 2013. Dating phylogenies with sequentially sampled tips. *Syst. Biol*. **62**, 674–688.
Stadler, T., R. Kouyos, V. von Wyl et al. 2012. Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol*. **29**, 347–357.
Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.
Stamatakis, A., A. J. Aberer, C. Goll et al. 2012. RAxML-Light: a tool for computing terabyte phylogenies. *Bioinformatics* **28**, 2064–2066.
Steel, M. 2011. Can we avoid 'SIN' in the house of 'no common mechanism'? *Mol. Biol. Evol*. **60**, 96–109.
Steel, M., and K. M. Pickett. 2006. On the impossibility of uniform priors on clades. *Mol. Phylogenet. Evol*. **39**, 585–586.
Steel, M. A. 1994a. The maximum likelihood point for a phylogenetic tree is not unique. *Syst. Biol*. **43**, 560–564.
Steel, M. A. 1994b. Recovering a tree from the leaf colourations it generates under a Markov model. *Appl. Math. Lett*. **7**, 19–24.
Steel, M. A. 2005. Should phylogenetic models be trying to 'fit an elephant'? *Trends Genet*. **21**, 307–309.
Steel, M. A., and D. Penny. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol. Biol. Evol*. **17**, 839–850.
Stein, C. 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. 3rd Berkeley Symp. Math. Stat. Prob*. **1**, 197–206.
Stein, C. 1964. Inadmissibility of the usual estimator for the variance of a multivariate normal distribution with unknown mean. *Ann. Inst. Math*. **16**, 155–160.
Steiper, M. E., N. M. Young, and T. Y. Sukarna. 2004. Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid-cercopithecoid divergence. *Proc. Natl. Acad. Sci. U.S.A*. **101**, 17021–17026.
Stephens, M. 2000. Dealing with label switching in mixture models. *J. R. Stat. Soc. B*. **62**, 795–809.
Stephens, M., and P. Donnelly. 2000. Inference in molecular population genetics (with discussions). *J. R. Stat. Soc. B*. **62**, 605–655.
Stewart, C.-B., J. W. Schilling, and A. C. Wilson. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* **330**, 401–404.
Stigler, S. M. 1982. Thomas Bayes's Bayesian inference. *J. R. Stat. Soc. A*. **145**, 250–258.
Strimmer, K., and O. G. Pybus. 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol. Biol. Evol*. **18**, 2298–2305.
Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol*. **13**, 964–969.
Strobeck, K. 1987. Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* **117**, 149–153.
Stuart, A., K. Ord, and S. Arnold. 1999. *Kendall's Advanced Theory of Statistics*. Arnold, London.
Studier, J. A., and K. J. Keppler. 1988. A note on the neighbor-joining algorithm of Saitou and Nei. *Mol. Biol. Evol*. **5**, 729–731.
Suchard, M., and A. Rambaut. 2009. Many-core algorithms for statistical phylogenetics. *Bioinformatics* **25**, 1370–1376.

Suchard, M. A., R. E. Weiss, and J. S. Sinsheimer. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol. Biol. Evol.* **18**, 1001–1013.

Suchard, M. A., C. M. Kitchen, J. S. Sinsheimer, and R. E. Weiss. 2003. Hierarchical phylogenetic models for analyzing multipartite sequence data. *Syst. Biol.* **52**, 649–664.

Sueoka, N. 1995. Intrastrand parity rules of DNA base composition of cyprinid fishes in subgenus Notropis inferred from nucleotide and usage biases of synonymous codons. *J. Mol. Evol.* **40**, 318–325.

Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Stat. A – Theory Methods* **7**, 13–26.

Sullivan, J., and D. L. Swofford. 2001. Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Syst. Biol.* **50**, 723–729.

Sullivan, J., K. E. Holsinger, and C. Simon. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA in sigmodontine rodents. *Mol. Biol. Evol.* **12**, 988–1001.

Sullivan, J., D. L. Swofford, and G. J. P. Naylor. 1999. The effect of taxon-sampling on estimating rate heterogeneity parameters on maximum-likelihood models. *Mol. Biol. Evol.* **16**, 1347–1356.

Susko, E. 2008. On the distributions of bootstrap support and posterior distributions for a star tree. *Syst. Biol.* **57**, 602–612.

Susko, E. 2009. Bootstrap support is not first-order correct. *Syst. Biol.* **58**, 211–223.

Susko, E. 2010. First-order correct bootstrap support adjustments for splits that allow hypothesis testing when using maximum likelihood estimation. *Mol. Biol. Evol.* **27**, 1621–1629.

Suzuki, Y. 2004. New methods for detecting positive selection at single amino acid sites. *J. Mol. Evol.* **59**, 11–19.

Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **16**, 1315–1328.

Suzuki, Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16138–16143.

Swanson, W. J., and V. D. Vacquier. 2002a. Reproductive protein evolution. *Ann. Rev. Ecol. Syst.* **33**, 161–179.

Swanson, W. J., and V. D. Vacquier. 2002b. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3**, 137–144.

Swanson, W. J., Z. Yang, M. F. Wolfner, and C. F. Aquadro. 2001a. Positive Darwinian selection in the evolution of mammalian female reproductive proteins. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2509–2514.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail et al. 2001b. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7375–7379.

Swofford, D. L. 2000. *PAUP*: Phylogenetic Analysis by Parsimony*, Version 4. Sinauer Associates, Sanderland, Massachusetts.

Swofford, D. L., and G. J. Olsen. 1990. Phylogeny reconstruction. Pp. 411–501 *in* D. M. Hillis, and C. Moritz, eds. *Molecular Systematics*. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogeny inference. Pp. 407–514 *in* D. M. Hillis, C. Moritz, and B. K. Mable, eds. *Molecular Systematics*. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L., P. J. Waddell, J. P. Huelsenbeck et al. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* **50**, 525–539.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607.

Tajima, F., and M. Nei. 1982. Biases of the estimates of DNA divergence obtained by the restriction enzyme technique. *J. Mol. Evol.* **18**, 115–120.

Tajima, F., and N. Takezaki. 1994. Estimation of evolutionary distance for reconstructing molecular phylogenetic trees. *Mol. Biol. Evol.* **11**, 278–286.

Takahata, N. 1983. Gene identity and genetic differentiation of populations in the finite island model. *Genetics* **104**, 497–512.

Takahata, N. 1988. The coalescent in two partially isolated diffusion populations. *Genet. Res. (Camb.)* **52**, 213–222.

Takahata, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. *Genetics* **122**, 957–966.

Takahata, N., and M. Nei. 1985. Gene genealogy and variance of interpopulational nucleotide differences. *Genetics* **110**, 325–344.

Takahata, N., Y. Satta, and J. Klein. 1995. Divergence time and population size in the lineage leading to modern humans. *Theor. Popul. Biol.* **48**, 198–221.

Takezaki, N., and T. Gojobori. 1999. Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences. *Mol. Biol. Evol.* **16**, 590–601.

Takezaki, N., and M. Nei. 1994. Inconsistency of the maximum parsimony method when the rate of nucleotide substitution is constant. *J. Mol. Evol.* **39**, 210–218.

Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**, 823–833.

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. *Mol. Biol. Evol.* **9**, 678–687.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526.

Tamura, K., D. Peterson, N. Peterson et al. 2011. MEGA5, molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Tanner, M. A., and W. H. Wong. 2000. From EM to data augmentation: the emergence of MCMC Bayesian computation in the 1980s. *Stat. Sci.* **25**, 506–516.

Tateno, Y., N. Takezaki, and M. Nei. 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Mol. Biol. Evol.* **11**, 261–277.

Tautz, D., P. Arctander, A. Minelli et al. 2003. A plea for DNA taxonomy. *Trends Ecol. Evol.* **18**, 70–74.

Tavaré, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect. Math. Life Sci.* **17**, 57–86.

Tavaré, S., C. R. Marshall, O. Will et al. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* **416**, 726–729.

Telford, M. J., M. J. Wise, and V. Gowri-Shankar. 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria. *Mol. Biol. Evol.* **22**, 1129–1136.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of man and the apes. *Evolution* **37**, 221–224.

Thompson, E. A. 1975. *Human Evolutionary Trees*. Cambridge University Press, Cambridge, UK.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.

Thorne, J. L., and H. Kishino. 1992. Freeing phylogenies from artifacts of alignment. *Mol. Biol. Evol.* **9**, 1148–1162.

Thorne, J. L., and H. Kishino. 2005. Estimation of divergence times from molecular sequence data. Pp. 233–256 *in* R. Nielsen, ed. *Statistical Methods in Molecular Evolution*. Springer-Verlag, New York.

Thorne, J. L., N. Goldman, and D. T. Jones. 1996. Combining protein evolution and secondary structure. *Mol. Biol. Evol.* **13**, 666–673.

Thorne, J. L., H. Kishino, and J. Felsenstein. 1991. An evolutionary model for maximum likelihood alignment of DNA sequences [Erratum in *J. Mol. Evol.* 1992, 34, 91]. *J. Mol. Evol.* **33**, 114–124.

Thorne, J. L., H. Kishino, and J. Felsenstein. 1992. Inching toward reality: an improved likelihood model of sequence evolution. *J. Mol. Evol.* **34**, 3–16.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647–1657.

Thornton, J. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet*. **5**, 366–375.
Thornton, J. W., E. Need, and D. Crews. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**, 1714–1717.
Tillier, E. R., and R. A. Collins. 1998. High apparent rate of simultaneous compensatory base-pair substitutions in ribosomal RNA. *Genetics* **148**, 1993–2002.
Tillier, E. R. M. 1994. Maximum likelihood with multiparameter models of substitution. *J. Mol. Evol*. **39**, 409–417.
Tsaur, S. C., and C.-I. Wu. 1997. Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. *Mol. Biol. Evol*. **14**, 544–549.
Tucker, A. 1995. *Applied Combinatorics*. Wiley, New York.
Tuff, P., and P. Darlu. 2000. Exploring a phylogenetic approach for the detection of correlated substitutions in proteins. *Mol. Biol. Evol*. **17**, 1753–1759.
Tuffley, C., and M. Steel. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull. Math. Biol.* **59**, 581–607.
Tuffley, C., and M. Steel. 1998. Modeling the covarion hypothesis of nucleotide substitution. *Math. Biosci*. **147**, 63–91.
Tzeng, Y. H., R. Pan, and W. H. Li. 2004. Comparison of three methods for estimating rates of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol*. **21**, 2290–2298.
Ugalde, J. A., B. S. W. Chang, and M. V. Matz. 2004. Evolution of coral pigments recreated. *Science* **305**, 1433.
Vallender, E. J., and B. T. Lahn. 2004. Positive selection on the human genome. *Hum. Mol. Genet*. **13**, R245–R254.
Vinh, L. S., and A. von Haeseler. 2005. Shortest triplet clustering: reconstructing large phylogenies using representative sets. *BMC Bioinformatics* **6**, 92.
Vuong, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333.
Waddell, P. J., and M. A. Steel. 1997. General time-reversible distances with unequal rates across sites: mixing gamma and inverse Gaussian distributions with invariant sites. *Mol. Phylogenet. Evol*. **8**, 398–414.
Waddell, P. J., D. Penny, and T. Moore. 1997. Hadamard conjugations and modeling sequence evolution with unequal rates across sites. *Mol. Phylogenet. Evol*. **8**, 33–50. [Erratum in *Mol. Phylogenet. Evol*. 1997, **8**, 446]
Wakeley, J. 1994. Substitution-rate variation among sites and the estimation of transition bias. *Mol. Biol. Evol*. **11**, 436–442.
Wakeley, J. 2009. *Coalescent Theory: An Introduction*. Roberts & Co. Publishers, Greenwood Village, Colorado
Wakeley, J., and N. Aliacar. 2001. Gene genealogies in a metapopulation. *Genetics* **159**, 893–905.
Wald, A. 1949. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat*. **20**, 595–601.
Walker, A. J. 1974. New fast method for generating discrete random numbers with arbitrary frequency distributions. *Electron. Let*. **10**, 127–128.
Walsh, J. B. 1995. How often do duplicated genes evolve new functions? *Genetics* **139**, 421–428.
Wang, F., and D. P. Landau. 2001. Efficient, multiple-range random-walk algorithm to calculate the density of states. *Phys. Rev. Lett*. **86**, 2050–2053.
Wang, Y., and J. Hey. 2010. Estimating divergence parameters with small samples from a large number of loci. *Genetics* **184**, 363–379.
Waterston, R. H. Lindblad-TohK. E. Birney et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562.
Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276.
Weadick, C. J., and B. S. Chang. 2012. An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol. Biol. Evol*. **29**, 1297–1300.
Whelan, S., and N. Goldman. 2000. Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. *Mol. Biol. Evol*. **17**, 975–978.

Whelan, S., and N. Goldman. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. *Mol. Biol. Evol.* **18**, 691–699.
Whelan, S., P. Liò, and N. Goldman. 2001. Molecular phylogenetics: state of the art methods for looking into the past. *Trends Genet.* **17**, 262–272.
Wiens, J. J. 2007. Species delimitation: new approaches for discovering diversity. *Syst. Biol.* **56**, 875–878.
Wiley, E. O. 1981. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. Wiley, New York.
Wilkinson-Herbots, H. M. 1998. Genealogy and subpopulation differentiation under various models of population structure. *J. Math. Biol.* **37**, 535–585.
Wilkinson-Herbots, H. M. 2008. The distribution of the coalescence time and the number of pairwise nucleotide differences in the 'isolation with migration' model. *Theor. Popul. Biol.* **73**, 277–288.
Wilkinson-Herbots, H. M. 2012. The distribution of the coalescence time and the number of pairwise nucleotide differences in a model of population divergence or speciation with an initial period of gene flow. *Theor. Popul. Biol.* **82**, 92–108.
Wilkinson, M., F.-J. Lapointe, and D. J. Gower. 2003. Branch lengths and support. *Syst. Biol.* **52**, 127–130.
Wilkinson, M., D. Pisani, J. A. Cotton, and I. Corfe. 2005. Measuring support and finding unsupported relationships in supertrees. *Syst. Biol.* **54**, 823–831.
Wilkinson, M., J. O. McInerney, R. P. Hirt et al. 2007. Of clades and clans: terms for phylogenetic relationships in unrooted trees. *Trends Ecol. Evol.* **22**, 114–115.
Wilkinson, R. D., M. E. Steiper, C. Soligo et al. 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst. Biol.* **60**, 16–31.
Williams, P. D., D. D. Pollock, B. P. Blackburne, and R. A. Goldstein. 2006. Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput. Biol.* **2**, e69.
Williamson, S., and M. E. Orive. 2002. The genealogy of a sequence subject to purifying selection at multiple sites. *Mol. Biol. Evol.* **19**, 1376–1384.
Williamson, S. H., M. J. Hubisz, A. G. Clark et al. 2007. Localizing recent adaptive evolution in the human genome. *PLoS Genet.* **3**, e90.
Wilson, A. C., S. S. Carlson, and T. J. White. 1977. Biochemical evolution. *Ann. Rev. Biochem.* **46**, 573–639.
Wilson, D. J., and G. McVean. 2006. Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172**, 1411–1425.
Wilson, I. J., and D. J. Balding. 1998. Genealogical inference from microsatellite data. *Genetics* **150**, 499–510.
Wilson, I. J., M. E. Weal, and D. J. Balding. 2003. Inference from DNA data: population histories, evolutionary processes and forensic match probabilities. *J. R. Stat. Soc. A* **166**, 155–201.
Wong, W. H., and F. Liang. 1997. Dynamic weighting in Monte Carlo and optimization. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14220–14224.
Wong, W. S., and R. Nielsen. 2004. Detecting selection in noncoding regions of nucleotide sequences. *Genetics* **167**, 949–958.
Wong, W. S. W., Z. Yang, N. Goldman, and R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**, 1041–1051.
Wray, G. A., J. S. Levinton, and L. H. Shapiro. 1996. Molecular evidence for deep Precambrian divergences. *Science* **274**, 568–573.
Wright, F. 1990. The 'effective number of codons' used in a gene. *Gene* **87**, 23–29.
Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**, 97–159.
Wright, S. 1943. Isolation by distance. *Genetics* **28**, 114–138.
Wu, C.-I., and W.-H. Li. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1741–1745.
Wu, C. I., and C. T. Ting. 2004. Genes and speciation. *Nat. Rev. Genet.* **5**, 114–122.
Wyckoff, G. J., W. Wang, and C.-I. Wu. 2000. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, 304–309.
Xia, X. 1998. How optimized is the translational machinery in Escherichia coli, Salmonella typhimurium and Saccharomyces cerevisiae? *Genetics* **149**, 37–44.

Xie, W., P. O. Lewis, Y. Fan et al. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol*. **60**, 150–160.
Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol*. **10**, 1396–1401.
Yang, Z. 1994a. Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. *Syst. Biol*. **43**, 329–342.
Yang, Z. 1994b. Estimating the pattern of nucleotide substitution. *J. Mol. Evol*. **39**, 105–111.
Yang, Z. 1994c. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol*. **39**, 306–314.
Yang, Z. 1995a. Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. *J. Mol. Evol*. **40**, 689–697.
Yang, Z. 1995b. A space-time process model for the evolution of DNA sequences. *Genetics* **139**, 993–1005.
Yang, Z. 1995c. On the general reversible Markov-process model of nucleotide substitution: a reply to Saccone et al. *J. Mol. Evol*. **41**, 254–255.
Yang, Z. 1996a. Maximum-likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol*. **42**, 587–596.
Yang, Z. 1996b. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol*. **11**, 367–372.
Yang, Z. 1996c. Phylogenetic analysis using parsimony and likelihood methods. *J. Mol. Evol*. **42**, 294–307.
Yang, Z. 1997a. How often do wrong models produce better phylogenies? *Mol. Biol. Evol*. **14**, 105–108.
Yang, Z. 1997b. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci*. **13**, 555–556.
Yang, Z. 1998a. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol*. **15**, 568–573.
Yang, Z. 1998b. On the best evolutionary rate for phylogenetic analysis. *Syst. Biol*. **47**, 125–133.
Yang, Z. 2000a. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J. Mol. Evol*. **51**, 423–432.
Yang, Z. 2000b. Complexity of the simplest phylogenetic estimation problem. *Proc. R. Soc. B: Biol. Sci*. **267**, 109–116.
Yang, Z. 2002a. Likelihood and Bayes estimation of ancestral population sizes in Hominoids using data from multiple loci. *Genetics* **162**, 1811–1823.
Yang, Z. 2002b. Inference of selection from multiple species alignments. *Curr. Opinion Genet. Devel*. **12**, 688–694.
Yang, Z. 2004. A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zool. Sinica* **50**, 645–656.
Yang, Z. 2005. Bayesian inference in molecular phylogenetics. Pp. 63–90 *in* O. Gascuel, ed. *Mathematics of Evolution and Phylogeny*. Oxford University Press, Oxford, UK.
Yang, Z. 2006. *Computational Molecular Evolution*. Oxford University Press, Oxford, UK.
Yang, Z. 2007a. PAML 4, Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol*. **24**, 1586–1591.
Yang, Z. 2007b. Fair-balance paradox, star-tree paradox and Bayesian phylogenetics. *Mol. Biol. Evol*. **24**, 1639–1655.
Yang, Z. 2008. Empirical evaluation of a prior for Bayesian phylogenetic inference. *Phil. Trans. R. Soc. Lond. B*. **363**, 4031–4039.
Yang, Z. 2010. A likelihood ratio test of speciation with gene flow using genomic sequence data. *Genome Biol. Evol*. **2**, 200–211.
Yang, Z., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol*. **15**, 496–503.
Yang, Z., and M. dos Reis. 2011. Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol*. **28**, 1217–1228.
Yang, Z., and S. Kumar. 1996. Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. *Mol. Biol. Evol*. **13**, 650–659.

Yang, Z., and R. Nielsen. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**, 409–418.
Yang, Z., and R. Nielsen. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43.
Yang, Z., and R. Nielsen. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**, 908–917.
Yang, Z., and R. Nielsen. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol. Biol. Evol.* **25**, 568–579.
Yang, Z., and B. Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo Method. *Mol. Biol. Evol.* **14**, 717–724.
Yang, Z., and B. Rannala. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst. Biol.* **54**, 455–470.
Yang, Z., and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226.
Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9264–9269.
Yang, Z., and B. Rannala. 2012. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* **13**, 303–314.
Yang, Z., and D. Roberts. 1995. On the use of nucleic acid sequences to infer early branchings in the tree of life. *Mol. Biol. Evol.* **12**, 451–458.
Yang, Z., and C. E. Rodríguez. 2013. Searching for efficient Markov chain Monte Carlo proposal kernels. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19307–19312.
Yang, Z., and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* **19**, 49–57.
Yang, Z., and T. Wang. 1995. Mixed model analysis of DNA sequence evolution. *Biometrics* **51**, 552–561.
Yang, Z., and A. D. Yoder. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Syst. Biol.* **52**, 705–716.
Yang, Z., N. Goldman, and A. Friday. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol. Biol. Evol.* **11**, 316–324.
Yang, Z., N. Goldman, and A. E. Friday. 1995a. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Syst. Biol.* **44**, 384–399.
Yang, Z., S. Kumar, and M. Nei. 1995b. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641–1650.
Yang, Z., I. J. Lauder, and H. J. Lin. 1995c. Molecular evolution of the hepatitis B virus genome. *J. Mol. Evol.* **41**, 587–596.
Yang, Z., R. Nielsen, and M. Hasegawa. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.* **15**, 1600–1611.
Yang, Z., W. S. W. Wong, and R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118.
Yang, Z., R. Nielsen, N. Goldman, and A.-M. K. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.
Yap, V. B., H. Lindsay, S. Easteal, and G. Huttley. 2010. Estimates of the effect of natural selection on protein-coding content. *Mol. Biol. Evol.* **27**, 726–734.
Yoder, A. D., and Z. Yang. 2000. Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* **17**, 1081–1090.
Yokoyama, S. 2002. Molecular evolution of color vision in vertebrates. *Gene* **300**, 69–78.
Yoshida, I., W. Sugiura, J. Shibata et al. 2011. Change of positive selection pressure on HIV-1 envelope gene inferred by early and recent samples. *PLOS One* **6**, e18630.
Yu, N., Z. Zhao, Y. X. Fu et al. 2001. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Mol. Biol. Evol.* **18**, 214–222.
Zang, L.-L., X.-H. Zou, F.-M. Zhang et al. 2011. Phylogeny and species delimitation of the C-genome diploid species in Oryza. *J. Syst. Evol.* **49**, 386–395.
Zardoya, R., and A. Meyer. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Biol. Evol.* **13**, 933–942.

Zhai, W., R. Nielsen, and M. Slatkin. 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Mol. Biol. Evol*. **26**, 273–283.

Zhang, C., B. Rannala, and Z. Yang. 2012. Robustness of compound Dirichlet priors for Bayesian inference of branch lengths. *Syst. Biol*. **61**, 779–784.

Zhang, C., D.-X. Zhang, T. Zhu, and Z. Yang. 2011. Evaluation of a Bayesian coalescent method of species delimitation. *Syst. Biol*. **60**, 747–761.

Zhang, J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J. Mol. Evol*. **50**, 56–68.

Zhang, J. 2003. Evolution of the human ASPM gene, a major determinant of brain size. *Genetics* **165**, 2063–2070.

Zhang, J. 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. *Mol. Biol. Evol*. **21**, 1332–1339.

Zhang, J., and M. Nei. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol*. **44**, S139–146.

Zhang, J., S. Kumar, and M. Nei. 1997. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Mol. Biol. Evol*. **14**, 1335–1338.

Zhang, J., R. Nielsen, and Z. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol*. **22**, 2472–2479.

Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. U.S.A*. **95**, 3708–3713.

Zhang, J., Y. P. Zhang, and H. F. Rosenberg. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet*. **30**, 411–415.

Zhang, L., B. S. Gaut, and T. J. Vision. 2001. Gene duplication and evolution. *Science* **293**, 1551.

Zhang, R., Z. Peng, G. Li et al. 2013. Ongoing speciation in the tibetan plateau gymnocypris species complex. *PLoS One* **8**, e71331.

Zhao, Z., L. Jin, Y. X. Fu et al. 2000. Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. *Proc. Natl. Acad. Sci. U.S.A*. **97**, 11354–11358.

Zharkikh, A. 1994. Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol*. **39**, 315–329.

Zharkikh, A., and W.-H. Li. 1993. Inconsistency of the maximum parsimony method: the case of five taxa with a molecular clock. *Syst. Biol*. **42**, 113–125.

Zharkikh, A., and W.-H. Li. 1995. Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique. *Mol. Phylogenet. Evol*. **4**, 44–63.

Zhou, Y., H. Brinkmann, N. Rodrigue et al. 2010. A Dirichlet process covarion mixture model and its assessments using posterior predictive discrepancy tests. *Mol. Biol. Evol*. **27**, 371–384.

Zhu, L., and C. D. Bustamante. 2005. A composite likelihood approach for detecting directional selection from DNA sequence data. *Genetics* **170**, 1411–1421.

Zhu, T., and Z. Yang. 2012. Maximum likelihood implementation of an isolation-with-migration model with three species for testing speciation with gene flow. *Mol. Biol. Evol*. **29**, 3131–3142.

Zierke, S., and J. Bakos. 2010. FPGA acceleration of the phylogenetic likelihood function for Bayesian MCMC inference methods. *BMC Bioinformatics* **11**, 184.

Zoller, S., and A. Schneider. 2010. Empirical analysis of the most relevant parameters of codon substitution models. *J. Mol. Evol*. **70**, 605–612.

Zuckerkandl, E. 1964. Further principles of chemical paleogenetics as applied to the evolution of hemoglobin. Pp. 102–109 *in* P. H., ed. *Peptides of the Biological Fluids*. Elsevier, Amsterdam.

Zuckerkandl, E., and L. Pauling. 1962. Molecular disease, evolution, and genetic heterogeneity. Pp. 189–225 *in* M. Kasha, and B. Pullman, eds. *Horizons in Biochemistry*. Academic Press, New York.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pp. 97–166 *in* V. Bryson, and H. J. Vogel, eds. *Evolving Genes and Proteins*. Academic Press, New York.

Zwickl, D. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. *Ph.D. Thesis*: University of Texas at Austin.

Zwickl, D. J., and M. T. Holder. 2004. Model parameterization, prior distributions, and the general time-reversible model in Bayesian phylogenetics. *Syst. Biol*. **53**, 877–888.