

Inconsistency of Maximum Parsimony Revisited

SUSANNE SCHULMEISTER

*Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA;
E-mail: schulmei@amnh.org*

Abstract.—Felsenstein (1978, *Syst. Zool.* 27:401–410) showed that the method of maximum parsimony can be inconsistent, i.e., lead to an incorrect result with an infinite amount of data. The situation in which this inconsistency occurs is often called the “Felsenstein zone,” the phenomenon also known as “long-branch attraction.” Felsenstein derived a sufficient inconsistency condition from a model for four taxa with only two different parameters for the probability of change on the five branches connecting the four taxa. In the present paper, his approach is used to derive the inconsistency condition of maximum parsimony from the most general model for four taxa, i.e., with five different parameters for the probabilities of change on the five branches and, for the first time, for characters with k states ($k = 2, 3, 4, 5, 6, \dots$). This is used to determine the factors that can cause the inconsistency of maximum parsimony. It is shown that the probability of change on all five branches and the number of character states play a role in causing inconsistency. [Felsenstein zone; k -state characters; long-branch attraction.]

In statistics, an estimator is said to be consistent if it converges towards the correct value with a growing amount of data, or in other words, if it always gave a correct solution with an infinite amount of data. According to this definition, consistency would be an “all or nothing” statement: a method is either consistent or it is not. Because no method of phylogeny reconstruction is consistent in this sense of the word, i.e., in all situations, it is only relevant to ask which are the situations where a method would give a correct answer with an infinite amount of data. Accordingly, in biological systematics the word “consistent” has been applied in reference to particular situations, conditions, or models. Certain situations in which a method is inconsistent have been referred to as zones or areas of inconsistency.

That maximum parsimony can be inconsistent under specific circumstances was first shown by Felsenstein (1978) and Cavender (1978). This possibility has sometimes been employed as an argument against the use of maximum parsimony. Today, it is known that all methods of phylogeny reconstruction can be inconsistent if their underlying assumptions are violated sufficiently (Steel et al., 1994; inconsistency of maximum likelihood in particular is discussed in Chang, 1996; Gaut and Lewis, 1995; Siddall, 1998; Tuffley and Steel, 1997:601–602; Sullivan and Swofford, 2001; Swofford et al., 2001), which means that they all can give incorrect results if used with real data. Therefore, inconsistency by itself is no argument against the use of a method. (At least not until a method has been found that is consistent under all conditions—which will likely never be the case.) But this does not mean that the topic should be ignored. Determining the conditions that lead to inconsistency can give valuable insights into the mechanisms of a method. And this, in turn, can provide a basis for an improved application of a method.

Inconsistency conditions have been determined for a number of four-taxon cases. Felsenstein (1978) started by determining the inconsistency condition for binary characters and a model in which there are only two different parameters for the probabilities of change on the five branches connecting the four taxa, which I call

the two-parameter model. For the five-parameter model, Penny et al. (1991) gave an inconsistency condition for binary characters and Zharkikh and Li (1992) determined the condition for characters with four states. Still missing are general inconsistency conditions for characters with k states, for the two-parameter as well as the five-parameter model. Both will be given in the present paper. For the two-parameter model and k -state characters, only an approximation has been published previously (Albert et al., 1993).

The inconsistency of maximum parsimony has also been examined for five-taxon trees (Penny et al., 1987; Hendy and Penny, 1989; De Bry, 1992; Zharkikh and Li, 1993), but the present paper will be limited to the four-taxon case.

RESULTS AND DISCUSSION

Inconsistency Conditions for k -State Characters

Felsenstein (1978:406–408) derived a sufficient condition for the (in)consistency of maximum parsimony based on a two-parameter model with four terminal taxa (Fig. 1a). This two-parameter model assumed that a change on the internal branch is as probable as a change on the short terminal branches. But if A, B, C, and D are extant taxa, it is rather improbable that two terminal branches would have even roughly the same length (or probability of change) as an internal branch, much less exactly the same. (It is more likely that the four terminal branches are all relatively long.) Therefore it is necessary to examine a more general model in which each branch has its own probability of change. This five-parameter model is illustrated in Figure 1b.

The inconsistency condition from the five-parameter model is derived here for the first time for k -state characters ($k = 2, 3, 4, 5, 6, \dots$), in order to obtain the most general inconsistency condition for four taxa. We assume that the topology in Figure 1b is the true topology relating the terminal taxa A, B, C, and D. As in Felsenstein's (1978) model, it is assumed that there is a long and a short branch on either side of the internal branch, i.e., $a > b$ and $c > d$ (Fig. 1b). Note that the branch lengths

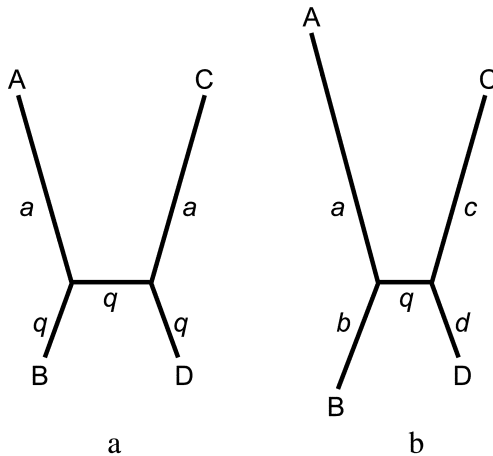


FIGURE 1. The assumed true topologies. (a) For Felsenstein's (1978) two-parameter model. It is identical to his Figure 3, except that P and Q were renamed as a and q. (b) For the five-parameter model.

a , b , c , d , and q represent probabilities, not substitution rate or time.

An infinite number of characters, all of which are assumed to have the same probability of change on a given branch, is assumed to evolve on the topology of Figure 1b, starting with the plesiomorphic character state at the left end of the internal branch. (The correct position of the source of the sequence (which is "evolved" on the topology) would be on one of the terminal branches instead of at the end of it (if we assume dichotomous speciation), but this doesn't make a difference with a time-reversible model.) Any character can evolve in a way that produces either an uninformative or an informative pattern if analyzed with Fitch parsimony (Fitch, 1971). In the latter case, the pattern could support any of the three topologies by which the four taxa could be related. Maximum parsimony is inconsistent if the probability of obtaining a pattern supporting the incorrect topology (AC)(BD) is higher than the probability of obtaining a pattern supporting the correct topology (AB)(CD), i.e., if $\text{Prob}[(AC)(BD)] > \text{Prob}[(AB)(CD)]$ (Felsenstein, 1978). (Because $a > b$ and $c > d$, observing the pattern (AD)(BC) is less probable than (AC)(BD), which will be shown below and means that (AD)(BC) can be ignored.) Figures 2 and 3 show the scenarios by which the characters with k states ($k = 2, 3, 4, 5, \dots$) can evolve into the two patterns.

Because there is no defined plesiomorphic condition in an unordered character, the plesiomorphic character state will be defined as some base X in Figures 2 and 3. One change leads to any of the other $k - 1$ bases, let it be Y. If Y changes, this can either be a reversal to the base X, or a transformation to a third base which is called Z. Also, X can change directly to Z. Of course, there could be a further change to a fourth character state, but that would lead to an uninformative character state pattern in our limited four-taxon tree. The characters are assumed to evolve according to the Jukes-Cantor model (Jukes and Cantor, 1969), i.e., all possible changes among the k

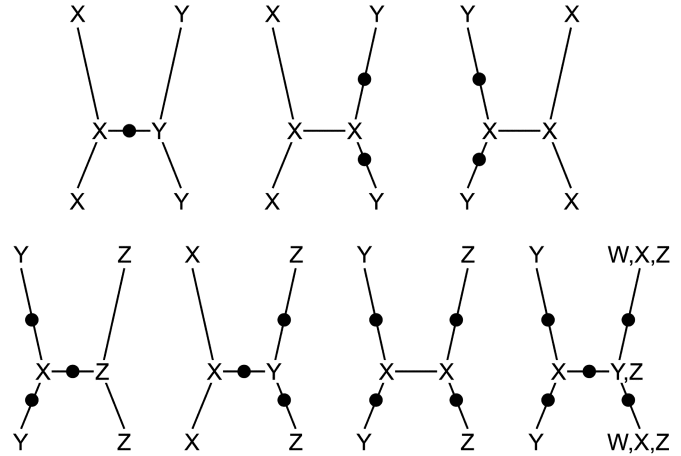


FIGURE 2. All possible scenarios that lead to a character state distribution that actually reconstructs as the topology (AB)(CD), for a character with $k > 2$. The origin is assumed to be at the left internal node.

states are equally likely to occur and the bases all have the same frequency $1/k$. The probability that a change can be observed on a branch is therefore maximally $(k - 1)/k$ (see Li, 1997:59-63).

Figure 2 shows all possible scenarios for a k -state character ($k > 2$) to evolve into a character state distribution that supports the topology (AB)(CD). For a character with two states, only the first three scenarios and, to some extent, the last scenario are possible. However, if $k = 2$ is entered in the equation below, the terms representing the impossible scenarios in the equation below become zero. Therefore, the equation can be used for all $k > 1$. The probability that a character produces this pattern is the sum of the probabilities of all scenarios:

$$\text{Prob}[(AB)(CD)] = (1 - a)(1 - b)q(1 - c)(1 - d)$$

$$+ (1 - a)(1 - b)(1 - q)c\left(\frac{1}{k - 1}\right)d$$

$$+ a\left(\frac{1}{k - 1}\right)b(1 - q)(1 - c)(1 - d)$$

$$+ a\left(\frac{1}{k - 1}\right)b\left(\frac{k - 2}{k - 1}\right)q(1 - c)(1 - d)$$

$$+ (1 - a)(1 - b)q\left(\frac{k - 2}{k - 1}\right)c\left(\frac{1}{k - 1}\right)d$$

$$+ a\left(\frac{1}{k - 1}\right)b(1 - q)\left(\frac{k - 2}{k - 1}\right)c\left(\frac{1}{k - 1}\right)d$$

$$+ a\left(\frac{1}{k - 1}\right)b\left(\frac{1}{k - 1}\right)qc\left(\frac{1}{k - 1}\right)d$$

$$+ a\left(\frac{1}{k - 1}\right)b\left(\frac{k - 2}{k - 1}\right)q\left(\frac{k - 2}{k - 1}\right)c\left(\frac{1}{k - 1}\right)d$$

(Equation 1)

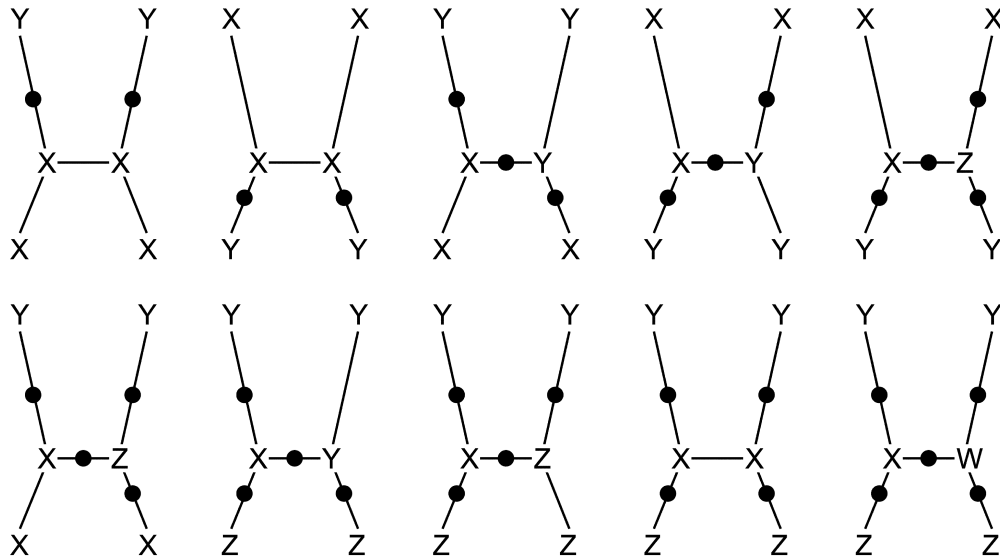


FIGURE 3. All possible scenarios that lead to a character state distribution that actually reconstructs as the topology (AC) (BD), for a character with $k > 3$. Again, the origin is assumed to be at the left internal node.

Figure 3 shows all possible scenarios for a k -state character ($k > 3$) to evolve into a pattern supporting the topology (AC)(BD). For characters with two or three states, only some of the scenarios are possible. However, if $k = 2$ or $k = 3$ is entered in the equation below, the terms for these impossible scenarios become zero. Therefore, the equation can be used for all $k > 1$.

$$\begin{aligned}
 \text{Prob}[(\overline{AC})(\overline{BD})] &= a(1-b)(1-q)\left(\frac{1}{k-1}\right)c(1-d) \\
 &+ (1-a)b(1-q)(1-c)\left(\frac{1}{k-1}\right)d \\
 &+ a(1-b)\left(\frac{1}{k-1}\right)q(1-c)\left(\frac{1}{k-1}\right)d \\
 &+ (1-a)b\left(\frac{1}{k-1}\right)q\left(\frac{1}{k-1}\right)c(1-d) \\
 &+ (1-a)b\left(\frac{k-2}{k-1}\right)q\left(\frac{1}{k-1}\right)c\left(\frac{1}{k-1}\right)d \\
 &+ a(1-b)\left(\frac{k-2}{k-1}\right)q\left(\frac{1}{k-1}\right)c\left(\frac{1}{k-1}\right)d \\
 &+ a\left(\frac{k-2}{k-1}\right)b\left(\frac{1}{k-1}\right)q(1-c)\left(\frac{1}{k-1}\right)d \\
 &+ a\left(\frac{k-2}{k-1}\right)b\left(\frac{1}{k-1}\right)q\left(\frac{1}{k-1}\right)c(1-d) \\
 &+ a\left(\frac{k-2}{k-1}\right)b(1-q)\left(\frac{1}{k-1}\right)c\left(\frac{1}{k-1}\right)d \\
 &+ a\left(\frac{k-2}{k-1}\right)b\left(\frac{k-3}{k-1}\right)q\left(\frac{1}{k-1}\right)c\left(\frac{1}{k-1}\right)d
 \end{aligned}$$

(Equation 2)

Maximum parsimony is inconsistent if and only if $\text{Prob}[(\overline{AC})(\overline{BD})] > \text{Prob}[(\overline{AB})(\overline{CD})]$, as long as $\text{Prob}[(\overline{AC})(\overline{BD})] > \text{Prob}[(\overline{AD})(\overline{BC})]$, which is always the case for $k > 1$. Hence, the actual (i.e., sufficient and necessary) inconsistency condition for unweighted Fitch parsimony (Fitch, 1971) and four terminal taxa connected by five branches with five probability parameters (as in Fig. 1b) and for characters with k states ($k = 2, 3, 4, 5, \dots$) is

$$\left(\frac{1}{k-1}\right)(a-d)(c-b) > q \left\{ \begin{aligned} &1 - a - b - c - d + \left(\frac{k}{k-1}\right)(ac + bd) \\ &+ \left(\frac{k^2 - 2k}{(k-1)^2}\right)(ab + ad + bc + cd) \\ &- \left(\frac{k^3 - 2k^2}{(k-1)^3}\right)(abc + abd + acd + bcd) \\ &+ \left(\frac{k^4 - 2k^3}{(k-1)^4}\right)abcd \end{aligned} \right\}$$

(Inequality 1)

Conclusions from this inconsistency condition will be drawn in another section.

For characters with $k = 2$ states (binary characters), the actual inconsistency condition is

$$(a-d)(c-b) > q\{1 - a - b - c - d + 2(ac + bd)\}$$

(Inequality 2)

For characters with $k = 4$ states, e.g., DNA sequences without gaps, the inconsistency condition is:

$$\frac{1}{3}(a-d)(c-b) > q \left\{ \begin{array}{l} 1 - a - b - c - d + \frac{4}{3}(ac + bd) \\ + \frac{8}{9}(ab + ad + bc + cd) \\ - \frac{32}{27}(abc + abd + acd + bcd) + \frac{128}{81}abcd \end{array} \right\} \quad (\text{Inequality 3})$$

A different form of this inequality (i.e., for $k = 4$) was published by Zharkikh and Li (1992: formula 6), albeit containing a slight error. In their formulas 6 and 7, the plus sign in front of Q_5 should be replaced with a minus sign. The corrected version of their Inequality 6 would be

$$0 > \frac{1}{3}(Q_4 - Q_1)(Q_3 - Q_2) - Q_5 \left\{ \begin{array}{l} \frac{2}{9}(Q_1 - Q_2)(Q_4 - Q_3) + \frac{1}{2}\left(1 - \frac{4}{3}Q_1\right) \\ \times \left(1 - \frac{4}{3}Q_2\right)\left(1 - \frac{4}{3}Q_3\right)\left(1 - \frac{4}{3}Q_4\right) \\ + \frac{1}{8}\left[\left(1 - \frac{4}{3}Q_1\right) + \left(1 - \frac{4}{3}Q_2\right)\right] \\ \times \left[\left(1 - \frac{4}{3}Q_3\right) + \left(1 - \frac{4}{3}Q_4\right)\right] \end{array} \right\} \quad (\text{Inequality 4})$$

(Because both their Inequalities 6 and 7 have this error, they equal each other out in inequality 8, so that inequality 8 is correct.)

As mentioned above, Felsenstein (1978) showed that the probability of getting a pattern which supports the topology (AD)(BC) is lower than Prob [(AC)(BD)], given $a > q$. This still has to be proven for the five-parameter model with k states.

$$\begin{aligned} & \text{Prob} [(AD)((BC)) \text{ (for } k > 1 \text{ and } a, b, c, d, \text{ and } q)] \\ &= a(1-b)(1-q)(1-c) \left(\frac{1}{k-1}\right) d \\ &+ (1-a)b(1-q) \left(\frac{1}{k-1}\right) c(1-d) \\ &+ a(1-b) \left(\frac{1}{k-1}\right) q \left(\frac{1}{k-1}\right) c(1-d) \\ &+ (1-a)b \left(\frac{1}{k-1}\right) q(1-c) \left(\frac{1}{k-1}\right) d \\ &+ (1-a)b \left(\frac{k-2}{k-1}\right) q \left(\frac{1}{k-1}\right) c \left(\frac{1}{k-1}\right) d \end{aligned}$$

$$\begin{aligned} &+ a(1-b) \left(\frac{k-2}{k-1}\right) q \left(\frac{1}{k-1}\right) c \left(\frac{1}{k-1}\right) d \\ &+ a \left(\frac{k-2}{k-1}\right) b \left(\frac{1}{k-1}\right) q \left(\frac{1}{k-1}\right) c(1-d) \\ &+ a \left(\frac{k-2}{k-1}\right) b \left(\frac{1}{k-1}\right) q(1-c) \left(\frac{1}{k-1}\right) d \\ &+ \left(\frac{k-2}{k-1}\right) ab(1-q) \left(\frac{1}{k-1}\right) c \left(\frac{1}{k-1}\right) d \\ &+ \left(\frac{k-2}{k-1}\right) ab \left(\frac{k-3}{k-1}\right) q \left(\frac{1}{k-1}\right) c \left(\frac{1}{k-1}\right) d \end{aligned} \quad (\text{Equation 3})$$

This leads to

$$\begin{aligned} & \text{Prob} [(AC)(BD)] - \text{Prob} [(AD)(BC)] \\ &= \left(\frac{1}{k-1}\right)(a-b)(c-d) \left(1 - \left(\frac{k}{k-1}\right)q\right) \end{aligned} \quad (\text{Equation 4})$$

It can be seen that this expression clearly is ≥ 0 as long as $a > b$, $c > d$, and $q < \frac{k-1}{k}$. Because $a > b$ and $c > d$ by definition, and because the probability q (as well as a , b , c , and d) cannot be larger than $\frac{k-1}{k}$ if the model is to be consistent with a continuous-time symmetric Markov process of base substitution (after Felsenstein, 1983; see also Li, 1997:59–63), the assumption Prob [(AC)(BD)] > Prob [(AD)(BC)] is correct.

The inconsistency condition for the five-parameter model for k -state characters ($k > 1$), which was derived above, can easily be transformed into an inconsistency condition for a two-parameter model with k states by replacing c by a and replacing b and d by q , which leads to

$$\begin{aligned} & \frac{1}{(k-1)}a^2 \\ & > q \left\{ \begin{array}{l} 1 - \left(\frac{2k-4}{k-1}\right)a - \left(\frac{2k-1}{k-1}\right)q + \left(\frac{k}{k-1}\right)(a^2 + q^2) \\ + \left(\frac{4k^2-8k}{(k-1)^2}\right)aq - \left(\frac{2k^3-4k^2}{(k-1)^3}\right)(a^2q + aq^2) \\ + \left(\frac{k^4-2k^3}{(k-1)^4}\right)a^2q^2 \end{array} \right\} \end{aligned} \quad (\text{Inequality 5})$$

The Causes of Inconsistency

Albert et al. (1993:759) mentioned that the inconsistency condition for a character with k states and two probability parameters is approximated by $\left(\frac{1}{k-1}\right)a^2 > q$, based on a personal communication with Felsenstein.

They did not give the exact inconsistency condition for k states and they did not state how the approximation was derived.

In the approximation of the inconsistency condition $a^2 > q$ for binary characters (Felsenstein, 1978) and the approximation for k -state characters given by Albert et al. (1993), only **the length of the long branches** in relation to the length of the internal branch seem to be decisive for inconsistency. In Felsenstein's (1978: 407) words: "For small values of P and Q, the condition for inconsistency amounts to requiring that simultaneous changes on two long segments of the tree be more probable a priori than one change on a short segment." The short terminal branches seem to be unimportant for inconsistency, as judged by these conditions and Felsenstein's statement.

But the probability of having $c = a$ and $b = d = q$ is zero for real data, which means that the two-parameter model is an unrealistic simplification of the four-taxon case and that a model with a different parameter for each of the branches must be used. Zharkikh and Li (1992) developed the consistency condition for the five-parameter model for DNA characters ($k = 4$) (see above), but the purely analytical part of their consistency research ended at this point; they did not derive any conclusions directly from their inconsistency condition. They set up a hypothetical tree like that in Figure 4 that evolved with a constant substitution rate overall. They then allowed one branch at a time to depart from this standard substitution rate (the molecular clock) and determined how much the substitution rate in a certain branch must differ from the standard rate to cause maximum parsimony to become

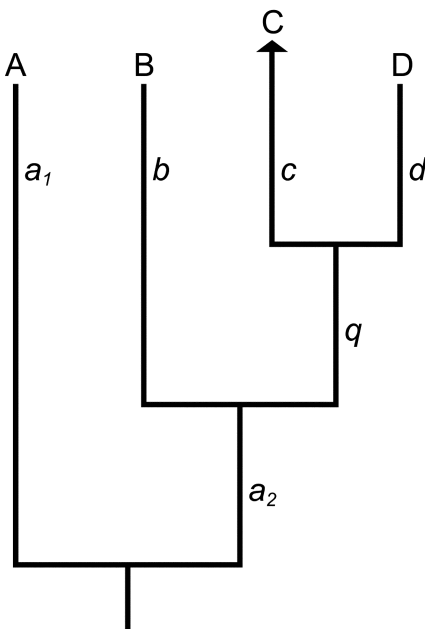


FIGURE 4. This figure illustrates the factors that influence inconsistency if taxa A, B, and D evolve with the same constant substitution rate and C evolves faster than the others. An explanation is given in the text.

inconsistent. When they varied the length of the outgroup branch, they found that parsimony remains consistent no matter how short or long the outgroup branch was. But they did not notice that this is only true for their example. This illustrates that their approach (which was mainly aimed at examining the bootstrap method) was not able to determine all the factors that cause inconsistency in maximum parsimony.

However, if the five-parameter inconsistency conditions for binary characters (Inequality 2) and for k -state characters (Inequality 1) are directly examined, it can be seen that it is rather **the difference in length between the (opposing) long and short terminal branches** in relation to the length of the internal branch that is the most important factor causing the inconsistency of maximum parsimony. The shorter the internal branch, the smaller the difference between the opposing long and short terminal branches needs to be to cause inconsistency.

Approximations

For certain cases, it can be helpful to simplify the inconsistency conditions through approximation. Equivalent to the form of the inconsistency condition of Zharkikh and Li (1992) (see above), the inconsistency condition for all $k > 1$ and five probability parameters can also be written as

$$\left(\frac{1}{k-1}\right)(a-d)(c-b) > q \left\{ \begin{aligned} &\left(\frac{k}{2(k-1)^2}\right)(a-b)(c-d) \\ &+ \left(\frac{k-2}{k}\right)\left(1-\left(\frac{k}{k-1}\right)a\right)\left(1-\left(\frac{k}{k-1}\right)b\right) \\ &\times \left(1-\left(\frac{k}{k-1}\right)c\right)\left(1-\left(\frac{k}{k-1}\right)d\right) \\ &+ \frac{1}{2k}\left[\left(1-\left(\frac{k}{k-1}\right)a\right)+\left(1-\left(\frac{k}{k-1}\right)b\right)\right] \\ &\times \left[\left(1-\left(\frac{k}{k-1}\right)c\right)+\left(1-\left(\frac{k}{k-1}\right)d\right)\right] \end{aligned} \right. \quad \text{(Inequality 6)}$$

Because the probabilities $a, b, c,$ and d are maximally $(k-1)/k$, the bracket next to q is maximally

$$\begin{aligned} &\left(\frac{k}{2(k-1)^2}\right)\left(\frac{k-1}{k}\right)^2 + \left(\frac{k-2}{k}\right) + \frac{4}{2k} \\ &= \frac{1}{2k} + \frac{2k-4}{2k} + \frac{4}{2k} = \frac{2k+1}{2k} \end{aligned}$$

(The terms " $-b$ " and " $-d$ " have been ignored and for the terms $1 - (\frac{k}{k-1})a$, etc., the upper bound 1 has been used in order to get the maximum for the content of the bracket; however, this is clearly overestimating the

possible maximum.) Hence, if

$$\frac{1}{k-1}(a-d)(c-b) > q \left(\frac{2k+1}{2k} \right) \quad (\text{Inequality 7})$$

the inconsistency condition is fulfilled. For $k = 2$, $\frac{2k+1}{2k}is \frac{5}{4} = 1.25$; for k growing to infinity this value converges towards 1. Therefore, a sufficient inconsistency condition is

$$\left(\frac{1}{k-1} \right) (a-d)(c-b) > \frac{5}{4}q \quad (\text{Inequality 8})$$

Similarly, the inconsistency condition from the two-parameter model for k -state characters can be written as

$$\left(\frac{1}{k-1} \right) a^2 > q \left[\begin{aligned} & \left(\frac{2}{k-1} \right) a - \left(\frac{1}{k-1} \right) q + \left(\frac{k}{2(k-1)^2} \right) (a-q)^2 \\ & + \left(\frac{k-2}{k} \right) \left(1 - \left(\frac{k}{k-1} \right) a \right)^2 \left(1 - \left(\frac{k}{k-1} \right) q \right)^2 \\ & + \frac{1}{2k} \left(1 - \left(\frac{k}{k-1} \right) a \right) + \left(1 - \left(\frac{k}{k-1} \right) q \right)^2 \end{aligned} \right] \quad (\text{Inequality 9})$$

The bracket next to q is maximally

$$\begin{aligned} & \left(\frac{2}{k-1} \right) a + \left(\frac{k}{2(k-1)^2} \right) a^2 + \left(\frac{k-2}{k} \right) + \frac{4}{2k} \\ & = \frac{4}{2k} + \frac{1}{2k} + \left(\frac{2k-4}{2k} \right) + \frac{4}{2k} \\ & = \frac{(4+1+2k-4+4)}{2k} = \frac{2k+5}{2k} \end{aligned}$$

Hence, a sufficient inconsistency condition for the five-parameter model is

$$\left(\frac{1}{k-1} \right) a^2 > \left(\frac{2k+5}{2k} \right) q \quad (\text{Inequality 10})$$

For $k = 2$, the bracket next to q is 2.25; for $k = 10$ it is 1.25; for k growing to infinity it converges towards 1. Therefore, a sufficient inconsistency condition for the two-parameter model is

$$\left(\frac{1}{k-1} \right) a^2 > \frac{9}{4}q \quad (\text{Inequality 11})$$

Deductions from the General Inconsistency Condition

What can be deduced from the approximation of the general inconsistency condition derived from the five-

parameter model for k -state characters (Inequalities 7 and 8)? As mentioned above, parsimony becomes inconsistent for four terminal taxa if the difference between the opposing long and short terminal branches is relatively large and if the internal branch is relatively short. The shorter the internal branch, the smaller the difference between the terminal branches can be in order to cause inconsistency. The longer the internal branch, the more extreme the difference between the terminal branches must be to cause inconsistency.

It should be kept in mind that the length of the branches in this context stands for the probabilities to observe a change on the respective branches, which is determined by both time and substitution rate. (Or, more precisely, the probability of observing a change in a character with $k = 4$ states on a certain branch is $p = \frac{3}{4} - \frac{3}{4}e^{-4\alpha t}$. See, for example, Li, 1997: 62.)

With the "duration" of the branches staying equal (as is the case for a given set of taxa), and the substitution rate going up (equally in all the branches, i.e., by the same factor), the term $\left(\frac{1}{k-1} \right) (a-d)(c-b)$ will increase much faster (provided that none of the probabilities has reached the maximum value of 3/4) than the term on the right side, which can even decrease. This means that for a given set of four taxa, increasing the overall substitution rate can lead from consistency to inconsistency. In other words, for a given tree with two relatively long terminal branches and two relatively short terminal branches, in which the two long terminal branches are at the two ends of a short internal branch, there is a certain substitution rate above which the consistency condition is not fulfilled (with an infinite data set and stochastically evolving characters, that is). This confirms an observation of Zharkikh and Li (1992).

We can also look at it the other way round: with increasing substitution rate, the differences between the opposing terminal branches can be smaller and smaller in order to cause inconsistency.

Another parameter that influences the consistency is k , the number of character states. The higher the number of character states, the smaller the risk of inconsistency.

In summary, the risk of inconsistency of parsimony is increased by

- a smaller number of character states,
- a shorter internal branch (i.e., a low probability of change between the two internal nodes),
- a larger difference between the probability of change on the paired terminal branches, which is promoted by a and/or c being relatively high and b and/or d being relatively small (as is, for example, the case if the outgroup taxon is far away from the ingroup), and
- a higher substitution rate.

Exploration of the Inconsistency Condition

Above, it has been shown under which circumstances maximum parsimony becomes inconsistent. But that treatment was rather abstract. What would the results

mean to four particular terminal taxa? How high is the risk of inconsistency? To examine these questions, the concept of the molecular clock (i.e., the assumption of a constant substitution rate) has been used by previous authors (e.g., Zharkikh and Li, 1992). Whether or not a molecular clock exists in reality has no bearing on these theoretical considerations; they are only intended to serve as an example which is less abstract than the general conclusions above.

For four terminal taxa that behaved according to the molecular clock (i.e., the probability of change in a certain branch is linearly correlated with time), maximum parsimony would be consistent because the long branches would be on one side of the internal branch and the short branches on the other (Fig. 4) (Hendy and Penny, 1989). The interesting question now is how much and in which way the evolution of four particular taxa can deviate from the molecular clock assumption without making parsimony inconsistent.

Under the molecular clock assumption, the outgroup taxon (A) is always at the end of the longest branch, the branch leading to B is intermediate, and the branches leading to C and D are the shortest branches, which are equally long (Fig. 4). This scenario is the starting point. How much would it have to change in order to make parsimony inconsistent? The first requirement is that the branches leading to A and C (or D) (respectively to B and C or D) would have to be the two longest branches. This is, for example, the case if taxon C evolves faster than the others, so that the branch leading to C becomes longer than that leading to B, i.e., $c > b$. The second requirement is that the branch leading to C becomes so long that the inconsistency condition is fulfilled. The question is: by which factor "y" does c have to be higher than d in order to cause inconsistency?

Zharkikh and Li (1992) set up an hypothetical tree like that in Figure 4 in which characters with four states evolved with a constant substitution rate of $\mu = 1 \times 10^{-8}$. In that tree, C and D diverged 45 Myr ago, B and (C+D) parted 50 Myr ago, and outgroup A and ingroup (B+C+D) separated 100 Myr ago. Accordingly, the internal branch is 5 Myr "long." Zharkikh and Li (1992) found that when the substitution rate in the branch leading to C (μ_c) was varied, maximum parsimony becomes inconsistent if $\mu_c > 1.32 \mu$. This is illustrated in Figure 5a. I further explored their example to illustrate how different factors influence the risk of inconsistency. It can be seen that if a_2 was decreased by half, i.e., taxon A was 75 Myr old (instead of 100 Myr), maximum parsimony would become inconsistent if $\mu_c > 1.48 \mu$ (instead of 1.32μ). This is illustrated in Figure 5b. If, instead, the substitution rate was decreased by half, inconsistency would be reached with $\mu_c > 1.55 \mu$ (Fig. 5c). If the internal branch was twice as long, μ_c would have to be larger than 1.64μ to cause inconsistency (Fig. 5d). In summary: reducing a_2 by half by choosing a different outgroup or using characters with half the substitution rate has approximately the same effect (in this example) as if the internal branch was twice as long as it is. If both a closer outgroup and slower characters were used, C would have

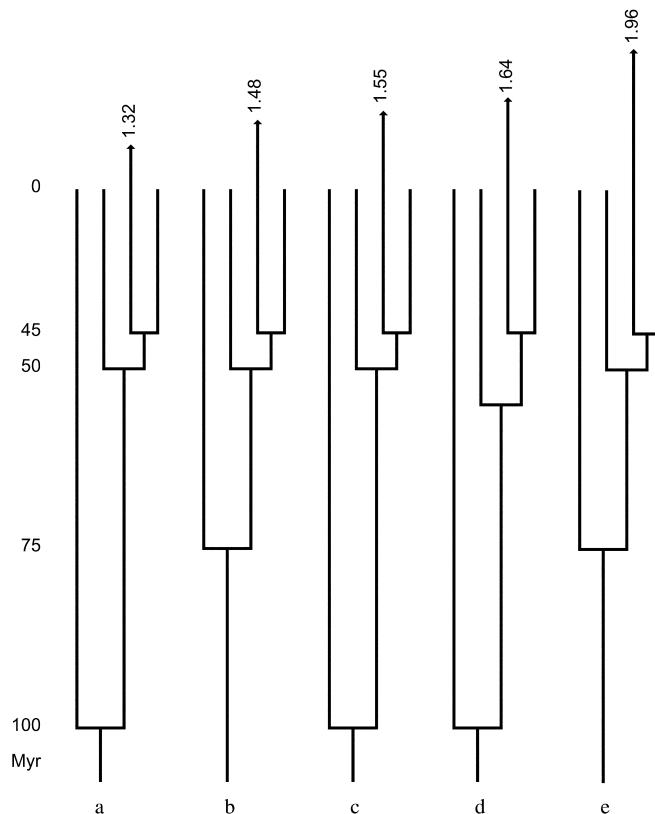


FIGURE 5. Deviations from the molecular clock causing inconsistency in maximum parsimony. The probabilities a , b , and d correspond to the age of the taxa in the example of Zharkikh and Li (1992). (a) The substitution rate in the branches leading to A, B, and D is $m = 1 \times 10^{-8}$. Maximum parsimony is inconsistent in Zharkikh and Li's model if $mc > 1.32 m$. (b) If the outgroup is closer to the ingroup taxa, i.e., the last common ancestor of all four taxa is 75 Myr old instead of 100 Myr, then parsimony is inconsistent if $mc > 1.48 m$. (c) If the substitution rate is half as much as in (a), i.e., $m = 0.5 \times 10^{-8}$, parsimony is inconsistent if $mc > 1.55 m$. (d) If the internal branch is twice as long as in (a) and a and b were changed accordingly, then parsimony is inconsistent if $mc > 1.64 m$. (e) If the substitution rate is half as much as in (a), i.e., the same as in (c), and the outgroup and ingroup diverged 75 Myr ago, i.e., as in (b), parsimony is inconsistent if $mc > 1.96 m$.

to evolve almost twice as fast as D to cause inconsistency (Fig. 5e).

CONCLUSIONS

With this paper, I hope to provide further insights into the "mechanism" of maximum parsimony by examining the conditions that lead to inconsistency in maximum parsimony. If a more realistic model is used, which allows different probabilities for all five branches of the four-taxon tree, it can be seen that all of these five probability parameters as well as the number of character states play a role in causing maximum parsimony to be inconsistent.

The approach used in this paper does not assume that characters change with a constant substitution rate: there is no particular relation between the branches required as in a molecular clock model; it includes the possibility that some taxa evolve faster than others. Therefore

the approach does not necessarily apply only for neutral characters (which are assumed to evolve with a constant substitution rate). But it does assume that the relative branch lengths in terms of probability of change are the same for all characters. This means that if the substitution rate in a branch is increased relative to other branches, this is equally true for all the characters. This assumption can at most be true for nonadaptive characters.

As was pointed out in the introduction, inconsistency is no argument against the use of a particular method of phylogenetic reconstruction, because all methods can give incorrect results in cases for which they are actually used. But this does not mean that the conditions that lead to incorrect results need not be evaluated. On the contrary, to achieve optimal results with a method, one must know of its limitations and potential pitfalls. Only if we are aware of the circumstances that lead to incorrect results can we attempt to avoid them.

Therefore, we first need to find out which circumstances cause a method to become inconsistent and how we could avoid these circumstances. Once this is determined, we need to ask how robust a method is against violations of its assumptions, whether it is possible with real data to avoid inconsistency in a certain method with a satisfying certainty, and whether the effort to achieve this stays within practical limits. Only these arguments can be used to judge different methods. More research on the consistency and reliability of maximum parsimony as well as other methods is required.

ACKNOWLEDGMENTS

I want to thank Mike Steel and Joe Felsenstein, whose comments and suggestions greatly improved the paper. I would also like to thank James Keating for his help.

REFERENCES

- Albert, V. A., M. W. Chase, and B. D. Mishler. 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences. *Ann. Miss. Bot. Gard.* 80:752–766.
- Cavender, J. A. 1978. Taxonomy with confidence. *Math. Biosci.* 40:271–280.
- Chang, J. T. 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* 134:189–215.
- DeBry, R. W. 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. *Mol. Biol. Evol.* 9:537–551.
- Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27:401–410.
- Felsenstein, J. 1983. Inferring evolutionary trees from DNA sequences. Pages 133–150 in *Statistical analysis of DNA sequence data* (B. S. Weir, ed.). Dekker, New York.
- Fitch, W. M. 1971. Toward defining the course of evolution: Minimal change for a specific tree topology. *Syst. Zool.* 20:406–416.
- 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.
- Hendy, M. D., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38:297–309.
- Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pages 21–123 in *Mammalian protein metabolism* (H. N. Munro, ed.). Academic Press, New York.
- Li, W.-H. 1997. *Molecular evolution*. Sinauer Associates, Sunderland, MA.
- Penny, D., M. D. Hendy, and I. M. Henderson. 1987. Reliability of evolutionary trees. *Cold Spring Harbor Symp. Quant. Biol.* 52:857–862.
- Penny, D., M. D. Hendy, and M. A. Steel. 1991. Testing the theory of descent. Pages 155–183 in *Phylogenetic analysis of DNA sequences* (M. M. Miyamoto and J. Cracraft, eds.). Oxford University Press, Oxford, UK.
- Siddall, M. 1998. Success of parsimony in the four-taxon case: Long-branch repulsion by likelihood in the Farris zone. *Cladistics* 14:209–220.
- Steel, M. A., L. A. Székely, and M. D. Hendy. 1994. Reconstructing trees when sequence sites evolve at variable rates. *J. Comput. Biol.* 1:153–163.
- 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.
- Swofford, D. L., P. J. Waddell, J. P. Huelsenbeck, P. G. Foster, P. O. Lewis, and J. S. Rodgers. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* 50:525–539.
- 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.
- Zharkikh, A., and W.-H. Li. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. II. Four taxa without a molecular clock. *J. Mol. Evol.* 35:356–366.
- 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.

First submitted 25 July 2001; reviews returned 8 December 2001;

final acceptance 14 December 2003

Associate Editor: Mike Steel