



WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Featured content](#)
[Current events](#)
[Random article](#)
[Donate to Wikipedia](#)
[Wikipedia store](#)

Interaction

[Help](#)
[About Wikipedia](#)
[Community portal](#)
[Recent changes](#)
[Contact page](#)

Tools

[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

Print/export

[Create a book](#)
[Download as PDF](#)
[Printable version](#)

Languages

[Català](#)
[Deutsch](#)
[فارسی](#)
[Français](#)
[日本語](#)
[Português](#)
[中文](#)

 [Edit links](#)

[Create account](#) [Log in](#)

Article [Talk](#)

[Read](#) [Edit](#) [View history](#)

Maximum parsimony (phylogenetics)

From Wikipedia, the free encyclopedia



This article **needs additional citations for verification**. Please help [improve this article](#) by [adding citations to reliable sources](#). Unsourced material may be challenged and removed. *(July 2015)*



This article **may be too long to read and navigate comfortably**. Please consider [splitting](#) content into sub-articles, [condensing](#) it, or adding or removing [subheadings](#). *(July 2015)*

In [phylogenetics](#), **maximum parsimony** is an [optimality criterion](#) under which the [phylogenetic tree](#) that minimizes the total number of character-state changes is to be preferred. Under the maximum-parsimony criterion, the optimal tree will minimize the amount of [homoplasy](#) (i.e., [convergent evolution](#), [parallel evolution](#), and [evolutionary reversals](#)). In other words, under this criterion, the shortest possible tree that explains the data is considered best. The principle is akin to [Occam's razor](#), which states that—all else being equal—the simplest hypothesis that explains the data should be selected. Some of the basic ideas behind maximum parsimony were presented by James S. Farris ^[1] in 1970 and [Walter M. Fitch](#) in 1971.^[2]

Maximum parsimony is an intuitive and simple criterion, and it is popular for this reason. However, although it is easy to *score* a phylogenetic tree (by counting the number of character-state changes), there is no algorithm to quickly *generate* the most-parsimonious tree. Instead, the most-parsimonious tree must be found in "tree space" (i.e., amongst all possible trees). For a small number of taxa (i.e., less than nine) it is possible to do an [exhaustive search](#), in which every possible tree is scored, and the best one is selected. For nine to twenty taxa, it will generally be preferable to use [branch-and-bound](#), which is also guaranteed to return the best tree. For greater numbers of taxa, a [heuristic search](#) must be performed.

Because the most-parsimonious tree is always the shortest possible tree, this means that—in comparison to the "true" tree that actually describes the evolutionary history of the organisms under study—the "best" tree according to the maximum-parsimony criterion will often underestimate the actual evolutionary change that has occurred. In addition, maximum parsimony is not statistically consistent. That is, it is not guaranteed to produce the true tree with high probability, given sufficient data. As demonstrated in 1978 by [Joe Felsenstein](#),^[3] maximum parsimony can be inconsistent under certain conditions, such as [long-branch attraction](#).

Contents [hide]

- [Alternate characterization and rationale](#)
- [In detail](#)
- [Character data](#)
- [Taxon sampling](#)
- [Analysis](#)
- [Problems with maximum parsimony phylogeny estimation](#)
- [Criticism](#)
- [Alternatives](#)
 - [8.1 Minimum evolution](#)
- [References](#)

Alternate characterization and rationale [\[edit\]](#)

The maximization of parsimony (preferring the simpler of two otherwise equally adequate theorizations) has proven useful in many fields. Occam's razor, a principle of theoretical parsimony suggested by [William of Ockham](#) in the 1320s, asserted that it is vain to give an explanation which involves more assumptions than necessary.

Alternatively, phylogenetic parsimony can be characterized as favoring the trees that maximize explanatory power by minimizing the number of observed similarities that cannot be explained by inheritance and common descent.^{[4][5]} Minimization of required evolutionary change on the one hand and maximization of observed similarities that can be explained as homology on the other may result in different preferred trees when some

observed features are not applicable in some groups that are included in the tree, and the latter can be seen as the more general approach.^{[6][7]}

While evolution is not an inherently parsimonious process, centuries of scientific experience lend support to the aforementioned principle of parsimony (*Occam's razor*). Namely, the supposition of a simpler, more parsimonious chain of events is preferable to the supposition of a more complicated, less parsimonious chain of events. Hence, parsimony (*sensu lato*) is typically sought in constructing phylogenetic trees, and in scientific explanation generally.^[8]

In detail ^[edit]

Parsimony is part of a class of character-based tree estimation methods which use a [matrix](#) of discrete phylogenetic characters to infer one or more optimal [phylogenetic trees](#) for a set of [taxa](#), commonly a set of [species](#) or reproductively isolated [populations](#) of a single species. These methods operate by evaluating candidate phylogenetic trees according to an explicit [optimality criterion](#); the tree with the most favorable score is taken as the best estimate of the phylogenetic relationships of the included taxa. Maximum parsimony is used with most kinds of phylogenetic data; until recently, it was the only widely used character-based tree estimation method used for morphological data.

Estimating phylogenies is not a trivial problem. A huge number of possible phylogenetic trees exist for any reasonably sized set of taxa; for example, a mere ten species gives over two million possible unrooted trees. These possibilities must be searched to find a tree that best fits the data according to the optimality criterion. However, the data themselves do not lead to a simple, arithmetic solution to the problem. Ideally, we would expect the distribution of whatever evolutionary characters (such as [phenotypic traits](#) or [alleles](#)) to directly follow the branching pattern of evolution. Thus we could say that if two organisms possess a shared character, they should be more closely related to each other than to a third organism that lacks this character (provided that character was not present in the last common ancestor of all three, in which case it would be a [symplesiomorphy](#)). We would predict that bats and monkeys are more closely related to each other than either is to an elephant, because male bats and monkeys possess external [testicles](#), which elephants lack. However, we cannot say that bats and monkeys are more closely related to one another than they are to whales, though the two have external testicles absent in whales, because we believe that the males in the last common ancestral species of the three had external testicles.

However, the phenomena of [convergent evolution](#), [parallel evolution](#), and [evolutionary reversals](#) (collectively termed *homoplasy*) add an unpleasant wrinkle to the problem of estimating phylogeny. For a number of reasons, two organisms can possess a trait not present in their last common ancestor: If we naively took the presence of this trait as evidence of a relationship, we would reconstruct an incorrect tree. Real phylogenetic data include substantial homoplasy, with different parts of the data suggesting sometimes very different relationships. Methods used to estimate phylogenetic trees are explicitly intended to resolve the conflict within the data by picking the phylogenetic tree that is the best fit to all the data overall, accepting that some data simply will not fit. It is often mistakenly believed that parsimony assumes that convergence is rare; in fact, even convergently derived characters have some value in maximum-parsimony-based phylogenetic analyses, and the prevalence of convergence does not systematically affect the outcome of parsimony-based methods.^[9]

Data that do not fit a tree perfectly are not simply "noise", they can contain relevant phylogenetic signal in some parts of a tree, even if they conflict with the tree overall. In the whale example given above, the lack of external testicles in whales is homoplastic: It reflects a return to the condition present in ancient ancestors of mammals, whose testicles were internal. This similarity between whales and ancient mammal ancestors is in conflict with the tree we accept, since it implies that the mammals with external testicles should form a group excluding whales. However, among the whales, the reversal to internal testicles actually correctly associates the various types of whales (including dolphins and porpoises) into the group [Cetacea](#). Still, the determination of the best-fitting tree—and thus which data do not fit the tree—is a complex process. Maximum parsimony is one method developed to do this.

Character data ^[edit]

The input data used in a maximum parsimony analysis is in the form of "characters" for a range of taxa. There is no generally agreed-upon definition of a phylogenetic character, but operationally a character can be thought of as an attribute, an axis along which taxa are observed to vary. These attributes can be physical (morphological), molecular, genetic, physiological, or behavioral. The only widespread agreement on characters seems to be that variation used for character analysis should reflect [heritable variation](#). Whether it must be directly heritable, or whether indirect inheritance (e.g., learned behaviors) is acceptable, is not entirely resolved.

Each character is divided into discrete **character states**, into which the variations observed are classified. Character states are often formulated as descriptors, describing the condition of the character substrate. For example, the character "eye color" might have the states "blue" and "brown." Characters can have two or more states (they can have only one, but these characters lend nothing to a maximum parsimony analysis, and are often excluded).

Coding characters for phylogenetic analysis is not an exact science, and there are numerous complicating issues. Typically, taxa are scored with the same state if they are more similar to one another in that particular attribute than each is to taxa scored with a different state. This is not straightforward when character states are not clearly delineated or when they fail to capture all of the possible variation in a character. How would one score the previously mentioned character for a taxon (or individual) with hazel eyes? Or green? As noted above, character coding is generally based on similarity: Hazel and green eyes might be lumped with blue because they are more similar to that color (being light), and the character could be then recoded as "eye color: light; dark." Alternatively, there can be multi-state characters, such as "eye color: brown; hazel, blue; green."

Ambiguities in character state delineation and scoring can be a major source of confusion, dispute, and error in phylogenetic analysis using character data. Note that, in the above example, "eyes: present; absent" is also a possible character, which creates issues because "eye color" is not applicable if eyes are not present. For such situations, a "?" ("unknown") is scored, although sometimes "X" or "-" (the latter usually in **sequence** data) are used to distinguish cases where a character cannot be scored from a case where the state is simply unknown. Current implementations of maximum parsimony generally treat unknown values in the same manner: the reasons the data are unknown have no particular effect on analysis. Effectively, the program treats a ? as if it held the state that would involve the fewest extra steps in the tree (see below), although this is not an explicit step in the algorithm.

Genetic data are particularly amenable to character-based phylogenetic methods such as maximum parsimony because protein and nucleotide sequences are naturally discrete: A particular position in a **nucleotide sequence** can be either **adenine**, **cytosine**, **guanine**, or **thymine** / **uracil**, or a **sequence gap**; a position (**residue**) in a **protein sequence** will be one of the basic **amino acids** or a sequence gap. Thus, character scoring is rarely ambiguous, except in cases where **sequencing** methods fail to produce a definitive assignment for a particular sequence position. Sequence gaps are sometimes treated as characters, although there is no consensus on how they should be coded.

Characters can be treated as unordered or ordered. For a binary (two-state) character, this makes little difference. For a multi-state character, unordered characters can be thought of as having an equal "cost" (in terms of number of "evolutionary events") to change from any one state to any other; complementarily, they do not require passing through intermediate states. Ordered characters have a particular sequence in which the states must occur through evolution, such that going between some states requires passing through an intermediate. This can be thought of complementarily as having different costs to pass between different pairs of states. In the eye-color example above, it is possible to leave it unordered, which imposes the same evolutionary "cost" to go from brown-blue, green-blue, green-hazel, etc. Alternatively, it could be ordered brown-hazel-green-blue; this would normally imply that it would cost two evolutionary events to go from brown-green, three from brown-blue, but only one from brown-hazel. This can also be thought of as requiring eyes to evolve through a "hazel stage" to get from brown to green, and a "green stage" to get from hazel to blue, etc.

There is a lively debate on the utility and appropriateness of character ordering, but no consensus. Some authorities order characters when there is a clear logical, **ontogenetic**, or evolutionary transition among the states (for example, "legs: short; medium; long"). Some accept only some of these criteria. Some run an unordered analysis, and order characters that show a clear order of transition in the resulting tree (which practice might be accused of **circular reasoning**). Some authorities refuse to order characters at all, suggesting that it biases an analysis to require evolutionary transitions to follow a particular path.

It is also possible to apply differential weighting to individual characters. This is usually done relative to a "cost" of 1. Thus, some characters might be seen as more likely to reflect the true evolutionary relationships among taxa, and thus they might be weighted at a value 2 or more; changes in these characters would then count as two evolutionary "steps" rather than one when calculating tree scores (see below). There has been much discussion in the past about character weighting. Most authorities now weight all characters equally, although exceptions are common. For example, **allele frequency** data is sometimes pooled in bins and scored as an ordered character. In these cases, the character itself is often downweighted so that small changes in allele frequencies count less than major changes in other characters. Also, the third **codon** position in a coding **nucleotide sequence** is particularly labile, and is sometimes downweighted, or given a weight of 0, on the assumption that it is more likely to exhibit homoplasy. In some cases, repeated analyses are run, with

characters reweighted in [inverse proportion](#) to the degree of homoplasy discovered in the previous analysis (termed [successive weighting](#)); this is another technique that might be considered [circular reasoning](#).

Character state changes can also be weighted individually. This is often done for [nucleotide sequence](#) data; it has been empirically determined that certain base changes (A-C, A-T, G-C, G-T, and the reverse changes) occur much less often than others. These changes are therefore often weighted more. As shown above in the discussion of character ordering, ordered characters can be thought of as a form of character state weighting.

Some systematists prefer to exclude characters known to be, or suspected to be, highly homoplastic or that have a large number of unknown entries ("?"). As noted below, theoretical and simulation work has demonstrated that this is likely to sacrifice accuracy rather than improve it. This is also the case with characters that are variable in the terminal taxa: theoretical, congruence, and simulation studies have all demonstrated that such polymorphic characters contain significant phylogenetic information.^{[\[citation needed\]](#)}

Taxon sampling ^{[\[edit\]](#)}

The time required for a parsimony analysis (or any phylogenetic analysis) is proportional to the number of [taxa](#) (and characters) included in the analysis. Also, because more taxa require more branches to be estimated, more uncertainty may be expected in large analyses. Because data collection costs in time and money often scale directly with the number of taxa included, most analyses include only a fraction of the taxa that could have been sampled. Indeed, some authors have contended that four taxa (the minimum required to produce a meaningful unrooted tree) are all that is necessary for accurate phylogenetic analysis, and that more characters are more valuable than more taxa in phylogenetics. This has led to a raging controversy about taxon sampling.

Empirical, theoretical, and simulation studies have led to a number of dramatic demonstrations of the importance of adequate taxon sampling. Most of these can be summarized by a simple observation: a phylogenetic data matrix has dimensions of characters *times* taxa. Doubling the number of taxa doubles the amount of information in a matrix just as surely as doubling the number of characters. Each taxon represents a new sample for every character, but, more importantly, it (usually) represents a new *combination* of character states. These character states can not only determine where that taxon is placed on the tree, they can inform the entire analysis, possibly causing different relationships among the remaining taxa to be favored by changing estimates of the pattern of character changes.

The most disturbing weakness of parsimony analysis, that of [long-branch attraction](#) (see below) is particularly pronounced with poor taxon sampling, especially in the four-taxon case. This is a well-understood case in which additional character sampling may not improve the quality of the estimate. As taxa are added, they often break up long branches (especially in the case of fossils), effectively improving the estimation of character state changes along them. Because of the richness of information added by taxon sampling, it is even possible to produce highly accurate estimates of phylogenies with hundreds of taxa using only a few thousand characters.^{[\[citation needed\]](#)}

Although many studies have been performed, there is still much work to be done on taxon sampling strategies. Because of advances in computer performance, and the reduced cost and increased automation of molecular sequencing, sample sizes overall are on the rise, and studies addressing the relationships of hundreds of taxa (or other terminal entities, such as genes) are becoming common. Of course, this is not to say that adding characters is not also useful; the number of characters is increasing as well.

Some systematists prefer to exclude taxa based on the number of unknown character entries ("?") they exhibit, or because they tend to "jump around" the tree in analyses (i.e., they are "wildcards"). As noted below, theoretical and simulation work has demonstrated that this is likely to sacrifice accuracy rather than improve it. Although these taxa may generate more most-parsimonious trees (see below), methods such as agreement subtrees and reduced consensus can still extract information on the relationships of interest.

It has been observed that inclusion of more taxa tends to lower overall support values ([bootstrap](#) percentages or decay indices, see below). The cause of this is clear: as additional taxa are added to a tree, they subdivide the branches to which they attach, and thus dilute the information that supports that branch. While support for individual branches is reduced, support for the overall relationships is actually increased. Consider analysis that produces the following tree: (fish, (lizard, (whale, (cat, monkey)))). Adding a rat and a walrus will probably reduce the support for the (whale, (cat, monkey)) clade, because the rat and the walrus may fall within this clade, or outside of the clade, and since these five animals are all relatively closely related, there should be more uncertainty about their relationships. Within error, it may be impossible to determine any of these animals' relationships relative to one another. However, the rat and the walrus will probably add character data that cements the grouping any two of these mammals exclusive of the fish or the lizard; where the initial analysis

might have been misled, say, by the presence of fins in the fish and the whale, the presence of the walrus, with blubber and fins like a whale but whiskers like a cat and a rat, firmly ties the whale to the mammals.

To cope with this problem, [agreement subtrees](#), [reduced consensus](#), and [double-decay analysis](#) seek to identify supported relationships (in the form of "n-taxon statements," such as the four-taxon statement "(fish, (lizard, (cat, whale)))") rather than whole trees. If the goal of an analysis is a resolved tree, as is the case for [comparative phylogenetics](#), these methods cannot solve the problem. However, if the tree estimate is so poorly supported, the results of any analysis derived from the tree will probably be too suspect to use anyway.

Analysis [\[edit\]](#)

A maximum parsimony analysis runs in a very straightforward fashion. Trees are scored according to the degree to which they imply a parsimonious distribution of the character data. The most parsimonious tree for the dataset represents the preferred hypothesis of relationships among the taxa in the analysis.

Trees are scored (evaluated) by using a simple algorithm to determine how many "steps" (evolutionary transitions) are required to explain the distribution of each character. A step is, in essence, a change from one character state to another, although with ordered characters some transitions require more than one step. Contrary to popular belief, the algorithm does not explicitly assign particular character states to nodes (branch junctions) on a tree: the least number of steps can involve multiple, equally costly assignments and distributions of evolutionary transitions. What is optimized is the total number of changes.

There are many more possible [phylogenetic trees](#) than can be searched exhaustively for more than eight taxa or so. A number of algorithms are therefore used to search among the possible trees. Many of these involve taking an initial tree (usually the favored tree from the last iteration of the algorithm), and perturbing it to see if the change produces a higher score.

The trees resulting from parsimony search are unrooted: They show all the possible relationships of the included taxa, but they lack any statement on relative times of divergence. A particular branch is chosen to root the tree by the user. This branch is then taken to be outside all the other branches of the tree, which together form a [monophyletic](#) group. This imparts a sense of relative time to the tree. Incorrect choice of a root can result in incorrect relationships on the tree, even if the tree is itself correct in its unrooted form.

Parsimony analysis often returns a number of equally most-parsimonious trees (MPTs). A large number of MPTs is often seen as an analytical failure, and is widely believed to be related to the number of missing entries ("?") in the dataset, characters showing too much homoplasy, or the presence of topologically labile "wildcard" taxa (which may have many missing entries). Numerous methods have been proposed to reduce the number of MPTs, including removing characters or taxa with large amounts of missing data before analysis, removing or downweighting highly homoplastic characters ([successive weighting](#)) or removing wildcard taxa (the [phylogenetic trunk](#) method) [a posteriori](#) and then reanalyzing the data.

Numerous theoretical and simulation studies have demonstrated that highly homoplastic characters, characters and taxa with abundant missing data, and "wildcard" taxa contribute to the analysis. Although excluding characters or taxa may appear to improve resolution, the resulting tree is based on less data, and is therefore a less reliable estimate of the phylogeny (unless the characters or taxa are non informative, see [safe taxonomic reduction](#)). Today's general consensus is that having multiple MPTs is a valid analytical result; it simply indicates that there is insufficient data to resolve the tree completely. In many cases, there is substantial common structure in the MPTs, and differences are slight and involve uncertainty in the placement of a few taxa. There are a number of methods for summarizing the relationships within this set, including [consensus trees](#), which show common relationships among all the taxa, and [pruned agreement subtrees](#), which show common structure by temporarily pruning "wildcard" taxa from every tree until they all agree. [Reduced consensus](#) takes this one step further, by showing all subtrees (and therefore all relationships) supported by the input trees.

Even if multiple MPTs are returned, parsimony analysis still basically produces a point-estimate, lacking [confidence intervals](#) of any sort. This has often been levelled as a criticism, since there is certainly error in estimating the most-parsimonious tree, and the method does not inherently include any means of establishing how sensitive its conclusions are to this error. Several methods have been used to assess support.

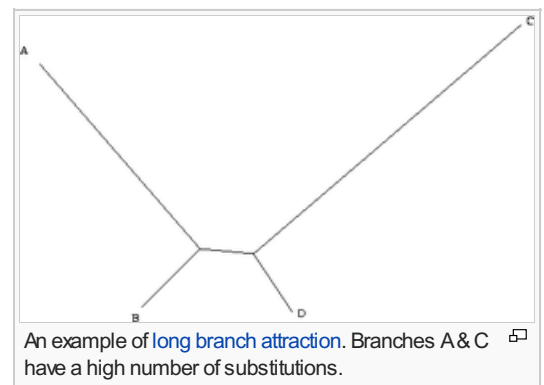
[Jackknifing](#) and [bootstrapping](#), well-known [statistical resampling](#) procedures, have been employed with parsimony analysis. The jackknife, which involves resampling without replacement ("leave-one-out") can be employed on characters or taxa; interpretation may become complicated in the latter case, because the variable of interest is the tree, and comparison of trees with different taxa is not straightforward. The bootstrap, resampling with replacement (sample x items randomly out of a sample of size x, but items can be picked multiple times), is only used on characters, because adding duplicate taxa does not change the result of a

parsimony analysis. The bootstrap is much more commonly employed in phylogenetics (as elsewhere); both methods involve an arbitrary but large number of repeated iterations involving perturbation of the original data followed by analysis. The resulting MPTs from each analysis are pooled, and the results are usually presented on a 50% **Majority Rule Consensus** tree, with individual branches (or nodes) labelled with the percentage of bootstrap MPTs in which they appear. This "bootstrap percentage" (which is not a **P-value**, as is sometimes claimed) is used as a measure of support. Technically, it is supposed to be a measure of repeatability, the probability that that branch (node, clade) would be recovered if the taxa were sampled again. Experimental tests with viral phylogenies suggest that the bootstrap percentage is not a good estimator of repeatability for phylogenetics, but it is a reasonable estimator of accuracy.^[citation needed] In fact, it has been shown that the bootstrap percentage, as an estimator of accuracy, is biased, and that this bias results on average in an underestimate of confidence (such that as little as 70% support might really indicate up to 95% confidence). However, the direction of bias cannot be ascertained in individual cases, so assuming that high values bootstrap support indicate even higher confidence is unwarranted.

Another means of assessing support is **Bremer support**, or the **decay index** (which is technically not an index). This is simply the difference in number of steps between the score of the MPT(s), and the score of the most parsimonious tree that does *not* contain a particular clade (node, branch). It can be thought of as the number of steps you have to add to lose that clade; implicitly, it is meant to suggest how great the error in the estimate of the score of the MPT must be for the clade to no longer be supported by the analysis, although this is not necessarily what it does. Decay index values are often fairly low (one or two steps being typical), but they often appear to be proportional to bootstrap percentages. However, interpretation of decay values is not straightforward, and they seem to be preferred by authors with philosophical objections to the bootstrap (although many morphological systematists, especially paleontologists, report both). **Double-decay analysis** is a decay counterpart to **reduced consensus** that evaluates the decay index for all possible subtree relationships (n-taxon statements) within a tree.

Problems with maximum parsimony phylogeny estimation ^[edit]

Maximum parsimony is a very simple approach, and is popular for this reason. However, it is not **statistically consistent**. That is, it is not guaranteed to produce the true tree with high probability, given sufficient data. Consistency, here meaning the monotonic convergence on the correct answer with the addition of more data, is a desirable property of any **statistical method**. As demonstrated in 1978 by **Joe Felsenstein**,^[3] maximum parsimony can be inconsistent under certain conditions. The category of situations in which this is known to occur is called **long branch attraction**, and occurs, for example, where there are long branches (a high level of substitutions) for two characters (A & C), but short branches for another two (B & D). A and B diverged from a common ancestor, as did C and D.



Assume for simplicity that we are considering a single binary character (it can either be + or -). Because the distance from B to D is small, in the vast majority of all cases, B and D will be the same. Here, we will assume that they are both + (+ and - are assigned arbitrarily and swapping them is only a matter of definition). If this is the case, there are four remaining possibilities. A and C can both be +, in which case all taxa are the same and all the trees have the same length. A can be + and C can be -, in which case only one character is different, and we cannot learn anything, as all trees have the same length. Similarly, A can be - and C can be +. The only remaining possibility is that A and C are both -. In this case, however, we group A and C together, and B and D together. As a consequence, when we have a tree of this type, the more data we collect (i.e. the more characters we study), the more we tend towards the wrong tree.

Another complication with maximum parsimony is that finding the most parsimonious tree is an **NP-Hard** problem.^[10] The only currently available, efficient way of obtaining a solution, given an arbitrarily large set of taxa, is by using heuristic methods which do not guarantee that the most parsimonious tree will be recovered. These methods employ **hill-climbing algorithms** to progressively approach the best tree. However, it has been shown that there can be "tree islands" of suboptimal solutions, and the analysis can become trapped in these **local optima**. Thus, complex, flexible heuristics are required to ensure that tree space has been adequately explored. Several heuristics are available, including **nearest neighbor interchange** (NNI), **tree bisection reconnection** (TBR), and the **parsimony ratchet**. This problem is certainly not unique to MP; any method that uses an optimality criterion faces the same problem, and none offer easy solutions.

Criticism [\[edit\]](#)

It has been asserted that a major problem, especially for [paleontology](#), is that maximum parsimony assumes that the only way two species can share the same nucleotide at the same position is if they are genetically related. This asserts that phylogenetic applications of parsimony assume that all similarity is [homologous](#) (other interpretations, such as the assertion that two organisms might *not* be related at all, are nonsensical). This is emphatically not the case: as with any form of character-based phylogeny estimation, parsimony is used to test the homologous nature of similarities by finding the phylogenetic tree which best accounts for all of the similarities.

For example, birds and bats have wings, while crocodiles and humans do not. If these were the only data available, maximum parsimony would tend to group crocodiles with humans, and birds with bats (as would any other method of [phylogenetic inference](#)). We believe that humans are actually more closely related to bats than to crocodiles or birds. Our belief is founded on additional data that were not considered in the one-character example (using wings). If even a tiny fraction of these additional data, including information on skeletal structure, [soft-tissue](#) morphology, integument, behaviour, genetics, etc., were included in the analysis, the faint phylogenetic signal produced by the presence of wings in birds and bats would be overwhelmed by the preponderance of data supporting the (human, bat)(bird, crocodile) tree.

It is often stated that parsimony is not relevant to phylogenetic inference because "evolution is not parsimonious." In most cases, there is no explicit alternative proposed; if no alternative is available, any statistical method is preferable to none at all. Additionally, it is not clear what would be meant if the statement "evolution is parsimonious" were in fact true. This could be taken to mean that more character changes may have occurred historically than are predicted using the parsimony criterion. Because parsimony phylogeny estimation reconstructs the minimum number of changes necessary to explain a tree, this is quite possible. However, it has been shown through simulation studies, testing with known [in vitro](#) viral phylogenies, and congruence with other methods, that the accuracy of parsimony is in most cases not compromised by this. Parsimony analysis uses the number of character changes on trees to choose the best tree, but it does not require that exactly that many changes, and no more, produced the tree. As long as the changes that have not been accounted for are randomly distributed over the tree (a reasonable null expectation), the result should not be biased. In practice, the technique is robust: maximum parsimony exhibits minimal bias as a result of choosing the tree with the fewest changes.

An analogy can be drawn with choosing among contractors based on their initial (nonbinding) estimate of the cost of a job. The actual finished cost is very likely to be higher than the estimate. Despite this, choosing the contractor who furnished the lowest estimate should theoretically result in the lowest final project cost. This is because, in the absence of other data, we would assume that all of the relevant contractors have the same risk of cost overruns. In practice, of course, unscrupulous business practices may bias this result; in phylogenetics, too, some particular phylogenetic problems (for example, [long branch attraction](#), described above) may potentially bias results. In both cases, however, there is no way to tell if the result is going to be biased, or the degree to which it will be biased, based on the estimate itself. With parsimony too, there is no way to tell that the data are positively misleading, without comparison to other evidence.

Along the same lines, parsimony is often characterized as implicitly adopting the philosophical position that evolutionary change is rare, or that homoplasy (convergence and reversal) is minimal in evolution. This is not entirely true: parsimony minimizes the number of convergences and reversals that are assumed by the preferred tree, but this may result in a relatively large number of such homoplastic events. It would be more appropriate to say that parsimony assumes only the minimum amount of change implied by the data. As above, this does not require that these were the only changes that occurred; it simply does not infer changes for which there is no evidence. The shorthand for describing this is that "parsimony minimizes assumed homoplasies, it does not assume that homoplasy is minimal."

Parsimony is also sometimes associated with the notion that "the simplest possible explanation is the best," a generalisation of [Occam's Razor](#). Parsimony does prefer the solution that requires the least number of unsubstantiated assumptions and unsupportable conclusions, the solution that goes the least theoretical distance beyond the data. This is a very common approach to science, especially when dealing with systems that are so complex as to defy simple models. Parsimony does not by any means necessarily produce a "simple" assumption. Indeed, as a general rule, most character datasets are so "noisy" that no truly "simple" solution is possible.

Alternatives [\[edit\]](#)

Main article: [Computational phylogenetics](#)

There are several other methods for inferring phylogenies based on discrete character data, including [maximum likelihood](#) and [Bayesian inference](#). Each offers potential advantages and disadvantages. In practice, these methods tend to favor trees that are very similar to the most parsimonious tree(s) for the same dataset; however, they allow for complex modelling of evolutionary processes, are [statistically consistent](#) and are not susceptible to [long-branch attraction](#). Note, however, that the performance of likelihood and Bayesian methods are dependent on the quality of the [model of evolution](#) employed; an incorrect model can produce a biased result. In addition, they are still quite slow relative to parsimony methods, sometimes requiring weeks to run large datasets. Most of these methods have particularly avid proponents and detractors; parsimony especially has been advocated as philosophically superior (most notably by ardent [cladists](#)).^{[[citation needed](#)]} One area where parsimony still holds much sway is in the analysis of morphological data, because—until recently—stochastic models of character change were not available for non-molecular data, and they are still not widely implemented. Parsimony has also recently been shown to be more likely to recover the true tree in the face of profound changes in evolutionary ("model") parameters (e.g., the rate of evolutionary change) within a tree ([Kolaczkowski & Thornton 2004](#)).



[Distance matrices](#) can also be used to generate phylogenetic trees. [Non-parametric](#) distance methods were originally applied to [phenetic](#) data using a matrix of pairwise distances and reconciled to produce a [tree](#). The [distance matrix](#) can come from a number of different sources, including [immunological distance](#), morphometric analysis, and [genetic distances](#). For phylogenetic character data, raw distance values can be calculated by simply counting the number of pairwise differences in character states ([Manhattan distance](#)) or by applying a model of evolution. Notably, distance methods also allow use of data that may not be easily converted to character data, such as [DNA-DNA hybridization](#) assays. Today, distance-based methods are often frowned upon because phylogenetically-informative data can be lost when converting characters to distances. There are a number of distance-matrix methods and optimality criteria, of which the **minimum evolution** criterion is most closely related to maximum parsimony.

Minimum evolution ^{[[edit](#)]}

The minimum-evolution tree-optimality criterion is similar to the maximum-parsimony criterion in that the tree that has the shortest total branch lengths is said to be optimal. The difference between these two criteria is that minimum evolution is calculated from a distance matrix, whereas maximum parsimony is calculated directly using the character matrix. Like the maximum-parsimony tree, the minimum-evolution tree must be sought in "tree space", typically using a [heuristic search](#) method. The [neighbor-joining](#) algorithm is very fast and often produces a tree that is quite similar to the minimum-evolution tree.

References ^{[[edit](#)]}

- [^] ^a ^b Farris, J. S. (1970). Methods for computing Wagner trees. *Systematic Zoology* 19, 83-92.
- [^] Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specified tree topology. *Systematic Zoology* 20 (4), 406-416
- [^] ^a ^b J. Felsenstein (1978). "Cases in which parsimony and compatibility methods will be positively misleading". *Systematic Zoology* 27 (4): 401–410. doi:10.1093/sysbio/27.4.401 [[][↗](#)[]].
- [^] Farris, J. S. (1983). The logical basis of phylogenetic analysis. In *Advances in Cladistics Vol. 2* (eds. N. I. Platnick, and V. A. Funk), pp. 7-36. Columbia University Press, New York, New York.
- [^] Farris, J. S. (2008). Parsimony and explanatory power. *Cladistics* 24, 1-23.
- [^] De Laet J. (2005). Parsimony and the problem of inapplicables in sequence data. Pp. 81-116 in Albert, V.A. (ed.) *Parsimony, phylogeny and genomics*. Oxford University Press, ISBN 0-19-856493-7
- [^] De Laet, Jan (2014). "Parsimony analysis of unaligned sequence data: maximization of homology and minimization of homoplasy, not Minimization of operationally defined total cost or minimization of equally weighted transformations". *Cladistics*. doi:10.1111/cla.12098 [[][↗](#)[]].
- [^] Jaynes, E.T. (2003) in Bretthorst, G.L. (Ed.), *Probability theory: the logic of science*. Cambridge, UK: Cambridge University Press.
- [^] Sober, E. (1983). "Parsimony in Systematics: Philosophical Issues". *Annual Review of Ecology and Systematics* 14: 335–357. doi:10.1146/annurev.es.14.110183.002003 [[][↗](#)[]].
- [^] William H. E. Day, [Computational complexity of inferring phylogenies from dissimilarity matrices](#) [[][↗](#)[]]. Bulletin of Mathematical Biology, volume 49, number 4, pages 461-467.
- Bremer, Kåre (1994). "Branch support and tree stability". *Cladistics* 10 (3): 295–304. doi:10.1111/j.1096-0031.1994.tb00179.x [[][↗](#)[]].
- Kolaczkowski, B.; Thornton, J. W. (2004). "Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous". *Nature* 431 (7011): 980–984. doi:10.1038/nature02917 [[][↗](#)[]]. PMID 15496922 [[][↗](#)[]].

v · t · e	Topics in phylogenetics [hide]
Relevant fields	Computational phylogenetics · Molecular phylogenetics · Cladistics · Evolutionary taxonomy
Basic concepts	Phylogenetic tree · Phylogenetic network · Long branch attraction · Clade vs Grade · Ghost lineage
Inference methods	Maximum parsimony · Probabilistic methods (Maximum likelihood · Bayesian inference) · Distance-matrix methods (Neighbor-joining · UPGMA · Least squares) · Three-taxon analysis
Current topics	PhyloCode · DNA barcoding · Molecular phylogenetics · Phylogenetic comparative methods · Phylogenetic network · Phylogenetic niche conservatism · Phylogenetics software · Phylogenomics · Phylogeography
Group traits	Symplesiomorphy · Apomorphy · Synapomorphy · Autapomorphy
Group types	Monophyly · Paraphyly · Polyphyly
<div>  List of evolutionary biology topics ·  Evolutionary biology </div>	

Categories: [Phylogenetics](#) | [Computational phylogenetics](#)