

Maximum parsimony working example of phylogeny (part-1)

Maximum parsimony is a **character-based** method of phylogeny. It means we do not require the distance matrix of the sequences. It is based on Occam's razor problem-solving principle of Monk William of Ockham (1280-1350) which states "*Entities should not be multiplied without necessity*". In other words, "The best hypothesis is the one requiring the smallest number of assumptions". The principle of maximum parsimony (MP) in phylogeny inference involves the identification of a tree topology that requires the smallest number of changes to explain the observed differences. The shortest pathway leading to these is selected as the best tree.

To reduce to complexity of the problem we use the informative site theory to identify the position of the sequences, which have to use for the phylogeny. It is assumed that not all position of the sequences is informative for inferring the phylogeny or finding evolutionary relationship between organisms.

An informative site is a position which fulfill two conditions

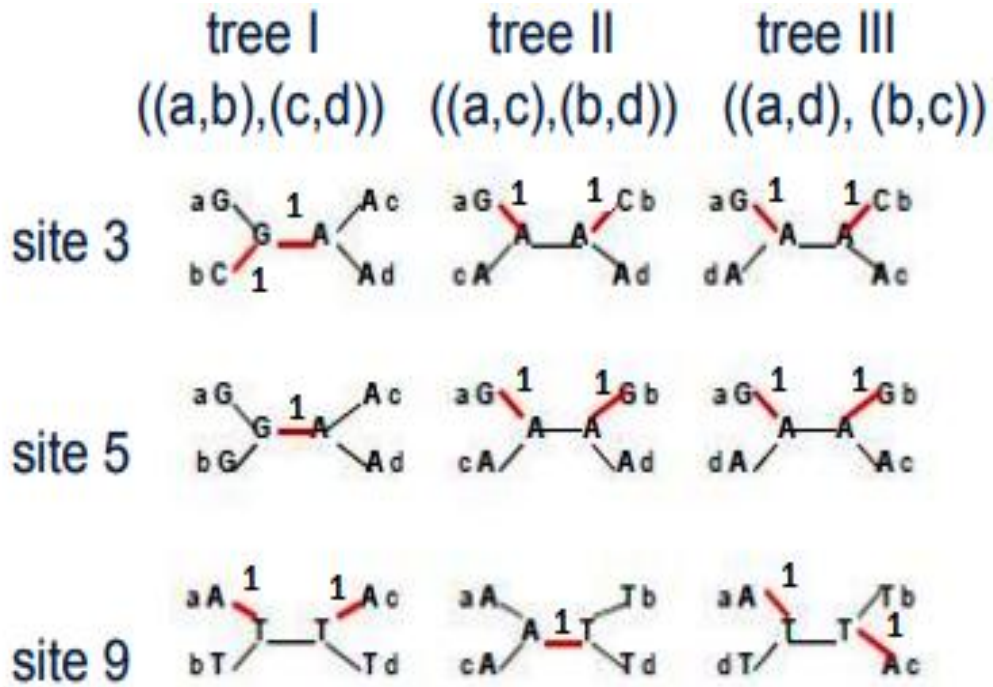
- The sites should occupied by at least two different type nucleotide.
- The two different nucleotides should present at least twice at that position.

Suppose we have four OUT having nine site and we have to make a phylogenetic tree using maximum parsimony approach.

	1	2	3	4	Site 5	6	7	8	9
Sequence	-----								
1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G
					*		*		*

First, we have to find the informative sites in this alignment described above. In this alignment only position no 5, 7 and 9 fulfill the information site criteria. Based on that we can create three possible unrooted tree per site as follows.

We have following observation



We have to choose that tree which have minimum no of mutation to explain the given tree topology.

For site 3 tree, I requires total 2 mutation and tree II and tree III requires only one mutations.

For site 5 tree I requires total 1 mutation and tree II and tree III requires 2 mutations .

For site 9 tree I requires total 2 mutation and tree II and tree III requires only one mutations .

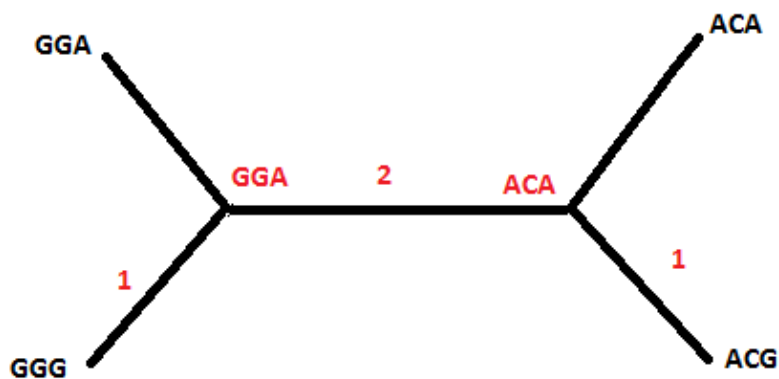
Summing changes:

	site 5	site 7	site 9	Sum
Tree I	1	1	2	4
Tree II	2	2	1	5
Tree III	2	2	2	6

The sum total for all sites is minimum (4) for tree I one so this is most parsimonious tree.

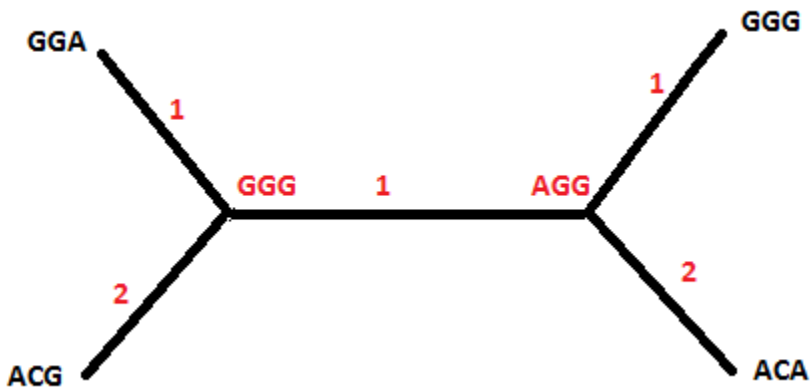
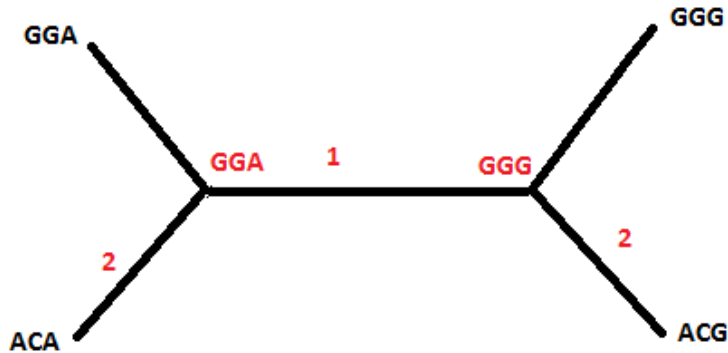
We can solve this example by taking care of all these three information sites together as well

- 1 G G A
- 2 G G G
- 3 A C A
- 4 A C G



Tree I total sum of changes = 1 + 2 + 1 = 4

Tree II total sum of changes = $2 + 1 + 2 = 5$



Tree III total sum of changes = $1 + 2 + 1 + 1 + 2 = 6$

Out of three tree configuration total changes required to explain tree configuration is minimum for tree I so this is the most parsimonious.

In the case of 4 OTUs an informative site can favor only one of the three possible alternative trees. For example, site 5 favors tree I over trees II and III, and is thus said to support tree I. The tree supported by the largest number of informative sites is the most parsimonious tree. In the cases where more than 4 OTUs are involved, an informative site may favor more than one tree and the maximum parsimony tree may not necessarily be the one supported by the largest number of informative sites.

FITCH'S PARSIMONY

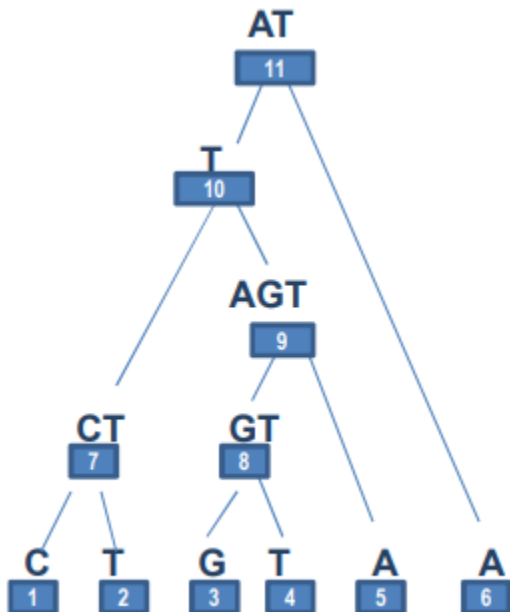
- The set at an interior node is the intersection of its two immediately descendant sets if the intersection is not empty.
- Otherwise, it is the union of the descendant sets.
- For every occasion that a union is required to form the nodal set, a nucleotide substitution at this position must have occurred at some point during the evolution for this position. Thus, counting the number of unions gives the minimum number of substitutions required to account for descendant nucleotides from a common ancestor, given the phylogeny assumed at the outset.

Example:

Let us consider the case of six OTUs, and one particular site, at which the nucleotides are as follows.

 <i>site</i>
OTU 1	C
OTU 2	T
OTU 3	G
OTU 4	T
OTU 5	A
OTU 6	A

Assume that there are five ancestors 7, 8, 9, 10, & 11 for these 6 OTUs



➤ One possible tree topology for the example site. The nucleotide at nodes 7, 8 and 9 cannot be determined uniquely under the parsimony rule. At node 10 T is chosen as it is shared by the sets at the two descendant nodes, 7 and 9. The nucleotide at node 11 cannot be determined uniquely. Parsimony requires it to be either A or T.

➤ At nodes 7, 8 and 10 nucleotide A could be included as a possible ancestral nucleotide because A is a possible common ancestral nucleotide (node 11) of all the six OTUs

➤ The nucleotide at node 10 is the intersection of the sets at nodes 7 and 9. The set at node 9 is the union of the sets at nodes 8 and 5.

➤ Counting the number of unions gives the minimum number of substitutions required to account for descendant nucleotides from a common ancestor, given the phylogeny assumed at the outset. In the example, this number is 4.

- There are many other alternative trees, each of which requires 3 substitutions. Thus, unlike the case of four OTUs, an informative site may favor many alternative trees.

References:

1. Molecular Evolution, by Wen-Hsiung Li, Sinauer Associates Inc, ISBN: 978-0878934638