

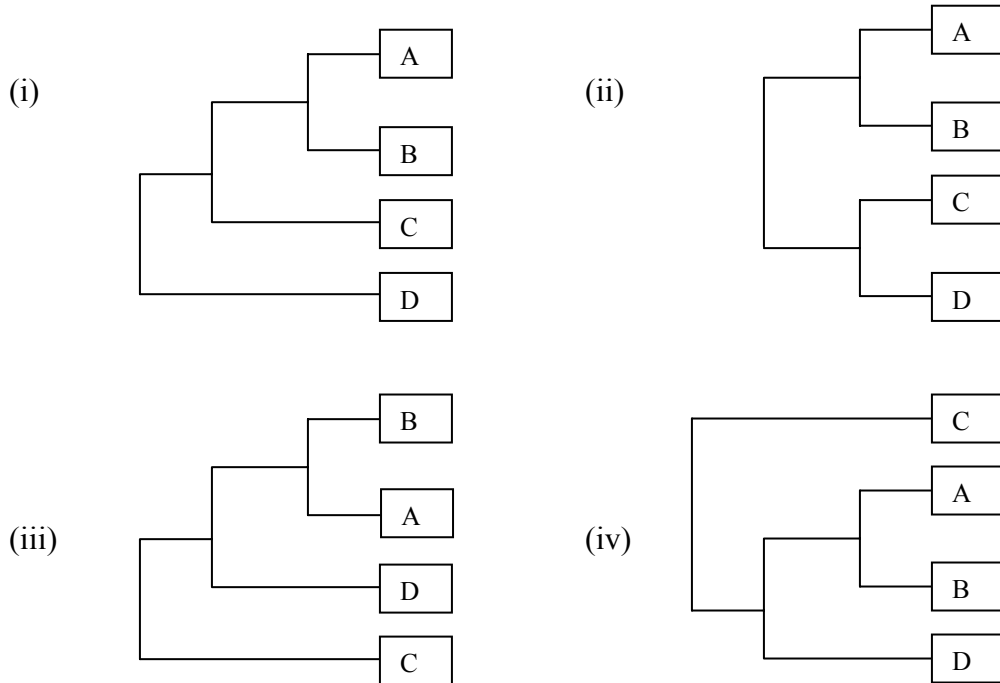
Chapter 8 SELF-TEST Phylogenetic methods

1. The number of possible distinct unrooted trees with 5 species is

- A. 15
- B. 25
- C. 5!
- D. $\frac{5!}{3 \times 2!}$

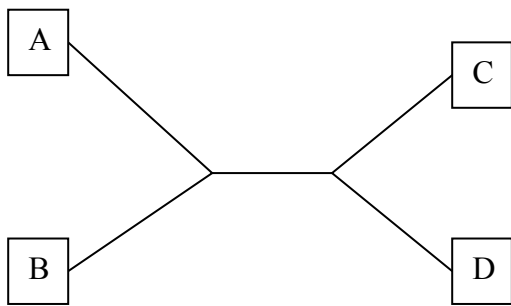
A is correct. Remember it's the product of the odd numbers $5 \times 3 \times 1 = 15$.

2. The following four trees are unrooted and branch lengths are not drawn to scale. Which statement is correct?



- A. All four trees are non-equivalent.
- B. All four trees are equivalent.
- C. Only trees (i) and (iii) are equivalent.
- D. Only trees (iii) and (iv) are equivalent.

B is correct. They are all equivalent because they are all the same as



3. If the same four trees are now taken to be rooted, which statement is correct?

- A. All four trees are non-equivalent.
- B. All four trees are equivalent.
- C. Only trees (i) and (iii) are equivalent.
- D. Only trees (iii) and (iv) are equivalent.

D is correct. Trees (iii) and (iv) are equivalent because C branches first, D second and A and B last.

4. Which of these statements about phylogenetic methods is correct?

- A. Bootstrapping can be used as a measure of confidence of the evolutionary model used in the phylogeny.
- B. Bootstrapping cannot be done if the rate of substitution varies across sites.
- C. Transition substitutions will usually saturate at a smaller divergence time than transversion substitutions.
- D. Long branch attraction arises when the lengths of the sequences in the analysis vary a great deal.

A – wrong – bootstrapping tells you about something statistical uncertainty but not about choice of evolutionary model.

B – wrong – in fact bootstrapping can be done with just about any method.

C – correct – because transitions happen faster

D – wrong - long branch attraction depends the lengths of the branches not the lengths of the sequences!

5. The alignment below is thought to be part of a protein-coding gene. Which of the four conclusions would you draw?

Wombat	:	GT	TAA	TG	AG	TG	GT	TAT	CC	AGA	AG	TG	AG	ATA	:	30
Opossum	:	GT	TAA	TG	AG	TG	GT	TAT	TC	AGA	AG	TG	AG	ATA	:	30
Rhinoceros	:	GT	TAA	TG	AG	TG	TTT	TCC	AGA	AG	TG	AA	ATA	:	30	
Pig	:	GT	TAA	TG	AG	TG	TTT	TCT	AGA	AG	CG	AA	ATA	:	30	
Hedgehog	:	GT	GAA	TG	AA	TG	CTT	TCC	AGA	AG	TG	AA	CTG	:	30	
Human	:	GT	TAA	TG	AG	TG	TTT	TCC	AGA	AG	TG	AA	CTG	:	30	
Hare	:	GT	TAA	CG	AG	TG	TTT	TCT	AGA	AG	TG	AA	ATG	:	30	

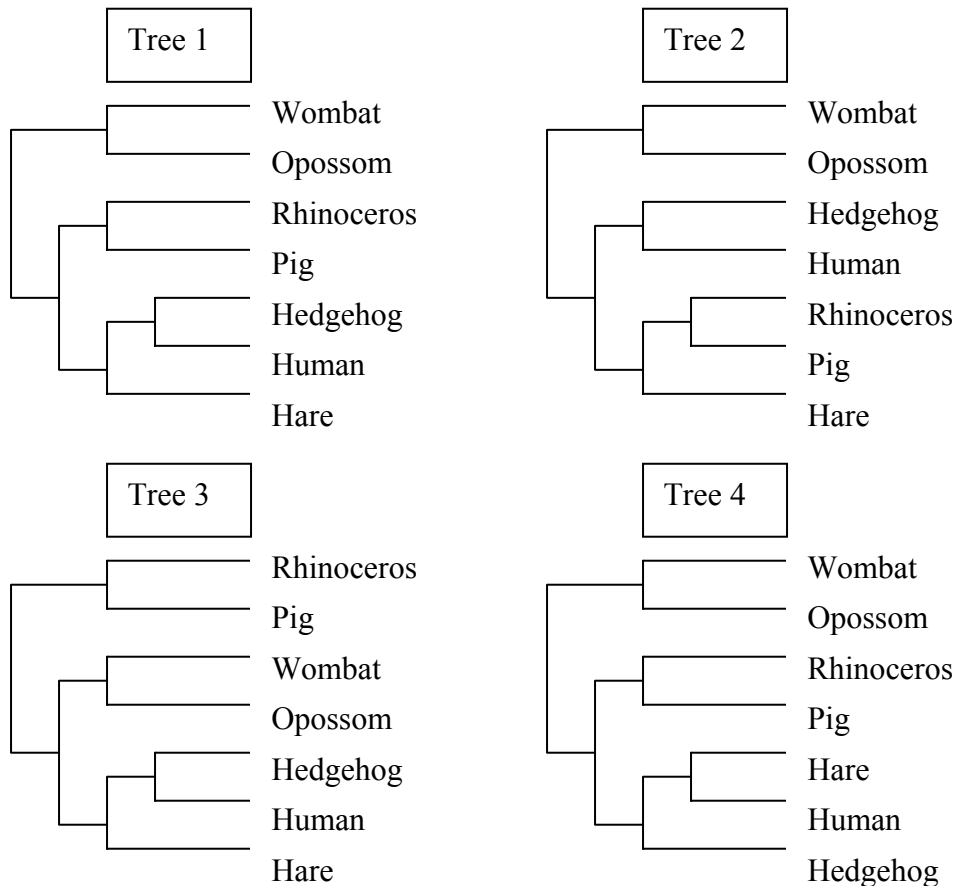
- A . This is probably part of a pseudogene.
- B . This is more likely to be part of an RNA-coding gene than a protein.
- C . There is evidence that transversions occur more frequently than transitions.
- D . There is evidence that synonymous substitutions occur more frequently than non-synonymous substitutions.

A and B – wrong - You can see the codon structure in this gene, so it must be a protein coding gene, not a pseudogene or an RNA gene.

C – wrong – There are 4 sites with TC transitions, 3 sites with AG transitions, 2 sites with a transversion, and 1 site with both a transition and a transversion. The number of transitions seems to be larger (as is usually the case).

D – correct – You could guess this by looking at the pattern of substitution in the third codon positions (sites 3, 6, 9 etc.), and you could confirm it by checking the genetic code table.

6. The following four phylogenetic trees have been proposed for the species in the previous sequence alignment. The standard parsimony criterion (minimization of the required number of mutations) is used to distinguish between the four trees using the above sequence data. Which conclusion would you draw?



According to the parsimony criterion:

- A. Tree 1 is preferable to all other trees.
- B. Trees 1 and 2 are preferable to Trees 3 and 4.
- C. Trees 1 and 3 are preferable to Trees 2 and 4.
- D. Trees 1 and 4 are preferable to Trees 2 and 3.

First you need to work out where the informative sites are in the alignment.

Sites 15 and 27 put Wombat and Opossum together.

Site 28 puts Human and Hedgehog together.

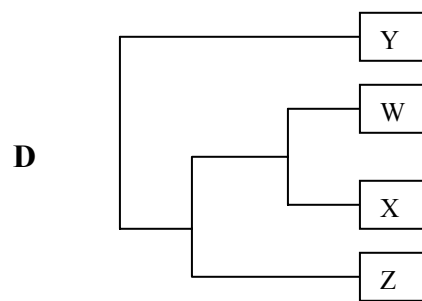
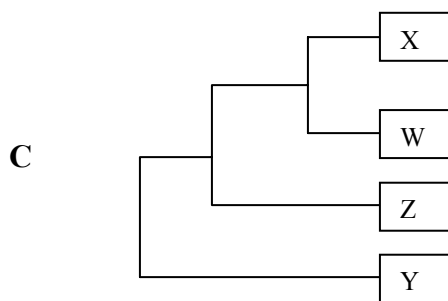
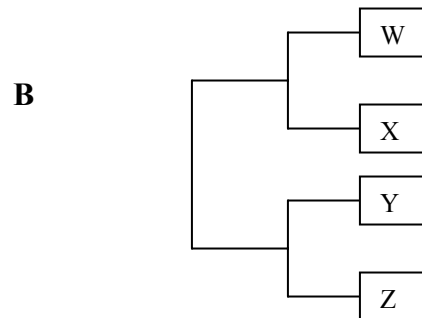
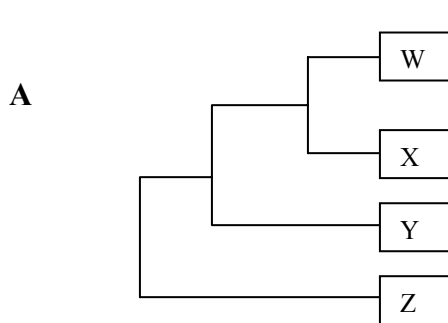
Site 30 puts Hare, Human and Hedgehog together.

All the other sites are non-informative, and can be ignored.

C is correct because all these groupings are contained in both trees 1 and 3. Trees 1 and 3 have the same parsimony score (which means A is wrong, as well as B and D).

7. Which of the four trees A, B, C or D corresponds to the result of the UPGMA algorithm applied to the distance matrix below? (The branch lengths are not drawn to scale).

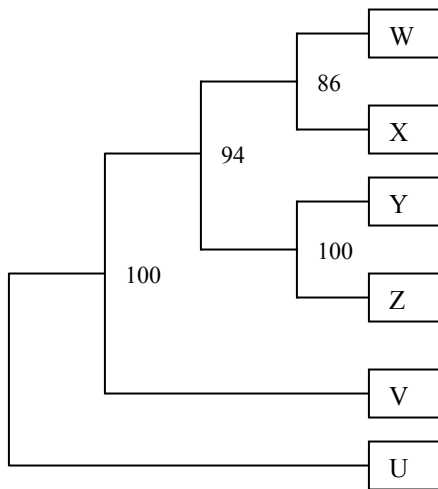
	<i>W</i>	<i>X</i>	<i>Y</i>	<i>Z</i>
<i>W</i>	–	0.3	0.4	0.5
<i>X</i>	0.3	–	0.5	0.5
<i>Y</i>	0.4	0.5	–	0.4
<i>Z</i>	0.5	0.5	0.4	–



Follow the UPGMA algorithm. First join W and X because this is the smallest distance. Then the distance from WX to Y is $(0.4+0.5)/2 = 0.45$, and the distance from WX to Z is $(0.5+0.5)/2 = 0.5$. Both

these distances are larger than the distance from Y to Z. Therefore the next step is to joining Y to Z. This means that tree B is correct.

8. 100 bootstrap data-sets are created for a set of sequences. The Neighbour-Joining method is applied to these data-sets to give 100 trees. The consensus tree of this set of 100 trees is given below with bootstrap percentages indicated. It has been rooted by assuming that U is the outgroup. Which of the four conclusions can be drawn?



- A. The clade X+Y+Z never occurs in the set of 100 trees.
- B. The clade W+X+V never occurs in the set of 100 trees.
- C. The clade W+X+V could occur in up to 6 of the 100 trees.
- D. The clade W+V could occur in up to 14 of the 100 trees.

A – wrong – The numbers on the consensus tree do not rule out the possibility that X is sometimes a sister group to Y+Z a small percentage of the time. We would have to look in the full set of trees to see whether this actually happens.

B – wrong – V could occasionally be the sister group to W+X.

C – correct – this is what the 94 tells us.

D – wrong – if W+V occurred 14 times, this would be consistent with the 86 on the W+X group, but inconsistent with the 94 on the W+X+Y+Z group.

9. Which of the following statements about phylogenetic methods is correct?

- A. The Maximum Likelihood method determines the tree for which the likelihood of the tree given the data is largest.
- B. The Maximum Likelihood tree may sometimes contain branches of zero length.
- C. If the trees from Maximum Likelihood, Parsimony and Neighbour-Joining methods all have the same topology, this must be the correct topology.
- D. In Bayesian phylogenetic methods, if the prior probabilities of two trees are equal, then the posterior probabilities must also be equal.

A – wrong – its the likelihood of the data given the tree. Check the discussion about how likelihoods are normalized and compare this with the discussion about Bayesian posterior probabilities.

B – correct – these look like multifurcations in the tree

C – wrong – unfortunately, it is quite possible for all three methods to give the same wrong topology.

D – wrong – because the posterior probabilities depend on the likelihood of the data on the two trees, which will usually not be the same.