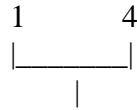


Problem Set 6

1. The pairwise distance matrix is:

		Taxon		
		2	3	4
Taxon				
	1	5	7	2
	2	5	3	
	3	5		

The pair of taxa with the smallest distance are Taxon 1 and Taxon 4 (distance=2). Join these two taxa together first:

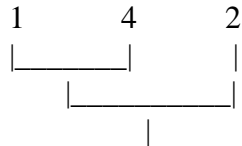


Next, recalculate the distance matrix, combining Taxa 1 and 4:

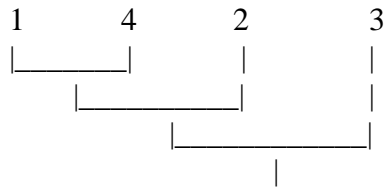
		Taxon	
		2	3
Taxon			
	1,4	4	6
	2		5

The distance between (1,4) and 2 is the average of the distance between 1 and 2 (5 from first table) and the distance between 4 and 2 (3 from first table), i.e. $\frac{1}{2}(5 + 3) = 4$. Similarly, the distance between (1,4) and 3 is $\frac{1}{2}(7 + 4) = 6$.

From this new table, the pair of taxa with the shortest distance is (1,4) and 2, so they are joined together next:



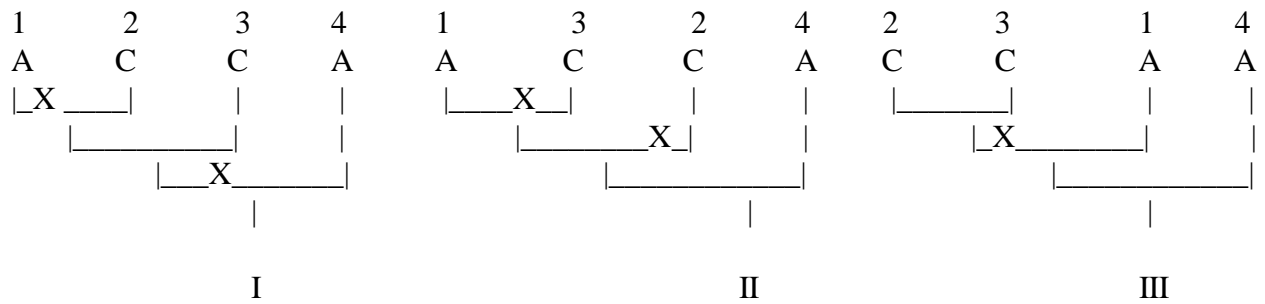
This leaves only one unconnected taxon left: taxon 3. This is then joined to the taxon ((1,4),2):



This is the final phylogeny.

2. The three possible phylogenies involving taxa 1, 2 and 3 (with 4 as the outgroup) are shown below. Also shown are the nucleotide identities of each taxon at the first nucleotide site in the sequences. The minimum transitions are shown for each phylogeny (note that phylogenies I and II are topologically equivalent. Each requires two transitions; different possibilities are shown on each tree).

Taxon	Nucleotide Sequence
1	A C T A G C A A A G
2	C C A C A C T A G C
3	C C A T A C T C G C
4	A C T A G C T A A C



Thus, to explain the data at the first nucleotide site, Model I and II require 2 transitions, while Model III requires only one. Repeating this process for each nucleotide site yields the following table of minimum number of transitions required:

Nucleotide Site	Model I	Model II	Model III
1	2	2	1
2	0	0	0
3	2	2	2
4	2	2	2
5	2	2	1
6	0	0	0
7	1	1	1
8	1	1	1
9	2	2	1
10	1	1	1
TOTAL	13	13	10

The total transitions required for each model for ALL of the 10 nucleotide sites is obtained by summing down the columns. Model III requires a total of only 10 transitions, while Models I and II require a minimum of 13. By the criterion of parsimony, then, Model III is favored.

Note that nucleotide sites 2, 3, 4, 6, 7, 8, and 10 are uninformative--they require the same minimum number of transitions for each Model. Only sites 1, 5 and 9 are informative. In this case, they all point to Model III being preferred. In a real dataset, not all informative sites will point to the same model. The model preferred is the one with that requires the smallest TOTAL number of transitions.

3. First consider Character A. The phylogeny, with the associated states of Character A are shown in the following phylogeny. First determine the character states of the nodes a, b and c by using the downpass. For node a, its descendent character states are 1 and 0, so in the downpass, we assign a value of 1, 0 to this node. This comes from Rule 2: If the two descendent nodes do not have a character state in common, assign the union of the two sets of character states. Next, for node b, the two descendent character states are 1,0 (node a) and 1 (taxon 2). Using Rule 1 (if two descendent nodes have a character state in common, assign that character state to the node), we assign a character state of 1 to node b. Finally, using Rule 2 again, we assign 1, 0 to node c, since it's descendents have states 1 (node b) and 0 (taxon 3).

Next, we use an uppass to resolve ambiguities generated in the downpass. The root node (node c) can not be resolved because there is no character state below it. Node b, according to Rule 2b of step 2, is 1,0. Finally, node a, using Rule 1 of step 2, is also 1,0.

Taxon	1	4	2	3	Node	Downpass	Uppass
Character State	1	0	1	0			
					a	1,0	1,0
					b	1	1,0
					c	1,0	1,0

The node assignments are shown on the following diagram. From these nodes, we can determine where character state transitions occurred (indicated by + on the tree). If the root (node c) is state 0, then nodes a and b are also state 0. One transition occurred from 0 to 1 between nodes c and b, while another occurred between node a and taxon 1. Alternatively, if the root is state 1, then nodes b and c are also 1. There was one transition from 1 to 0 between node c and taxon 3, and another between node a and taxon 4. In both cases, the most reconstructions require a minimum of 2 transitions.

Taxon	1	4	2	3	Node	Downpass	Uppass
Character State	1	0	1	0			
					a	1,0	1,0
					b	1	1,0
					c	1,0	1,0

For Character B, we have the following Downpass-Uppass analysis:

Taxon	1	4	2	3	Node	Downpass	Uppass
Character State	1	0	0	0			
	a				a	1,0	0
	b				b	0	0
	c				c	0	-0

|

This then gives the following node reconstructions and transitions:

Taxon	1	4	2	3	Downpass	Uppass
Character State	1	0	0	0		
	+0				1,0	0
	0				0	0
	0				0	0

There is one transition from 0 to 1 between node a and taxon 1.

For Character C, we have the following Downpass-Uppass analysis:

Taxon	1	4	2	3	Node	Downpass	Uppass
Character State	1	0	0	1			
	a				a	1,0	0
	b				b	0	0
	c				c	1,0	1,0

|

This then gives the following node reconstructions and transitions:

Taxon	1	4	2	3	Downpass	Uppass
Character State	1	0	0	1		
	1,0				1,0	1,0
	1,0				0	1,0
	1,0				1,0	1,0

If the root is 0, then nodes a and b are 0 and there are transitions from 0 to 1 between node c and taxon 3 and between node a and taxon 1. If the root is 1, then nodes a and b are 1 and there are transitions from 1 to 0 between node b and taxon 2 and between node a and taxon 4; or, nodes a and b are 0 and there is a transition from 1 to 0 between nodes c and b, and a transition from 0 to 1 between node a and taxon 1.