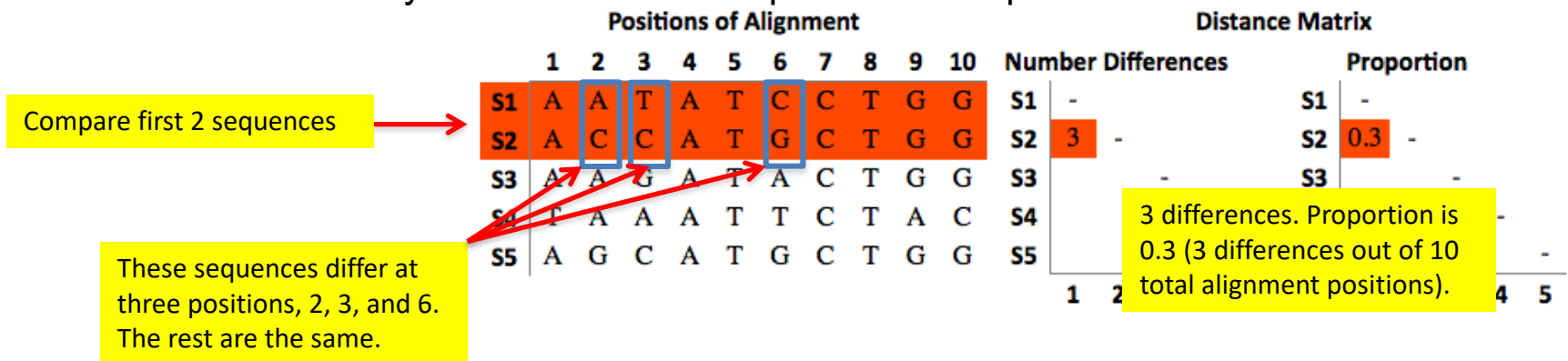# Distance Matrix: Concept Mode

This tutorial explains how to create a Distance Matrix from a multiple sequence alignment. The "distances" calculated are a measure of how different the sequences in the alignment are from one another. A distance of 0 means the sequences are identical, and a difference of 1 means they have no similarity at all. CENTRAL IDEA: The more similar the sequence, the more closely related they are (and the organisms they come from.

Concept Mode

(1) Shows how the initial distance matrix is built from pair-wise comparisons of all the sequences in the alignment.
(2) Shows how the distance matrix is used to build a tree by joining "neighbors".

Step 1: Compare sequences – All sequences are compared to all other sequences and the number of differences are counted. The proportion is the number of differences divided by the total number of positions compared.

Positions of Alignment

	1	2	3	4	5	6	7	8	9	10
S1	A	A	T	A	T	C	C	T	G	G
S2	A	C	C	A	T	G	C	T	G	G
S3	A	A	G	A	T	A	C	T	G	G
S4	T	A	A	A	T	T	C	T	A	C
S5	A	G	C	A	T	G	C	T	G	G

Distance Matrix

Number Differences

	S1	S2	S3	S4	S5
S1	-				
S2	3	-			
S3			-		
S4				-	
S5					-

Proportion

	S1	S2	S3	S4	S5
S1	-				
S2	0.3	-			
S3			-		
S4				-	
S5					-

Compare first 2 sequences

These sequences differ at three positions, 2, 3, and 6. The rest are the same.

3 differences. Proportion is 0.3 (3 differences out of 10 total alignment positions).
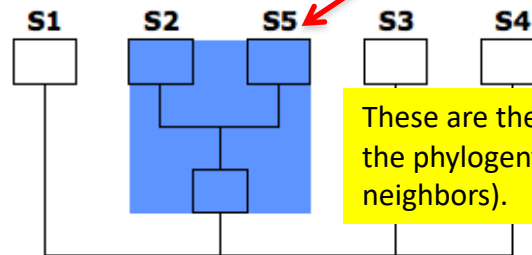
Continue to next pair.

STEP   RESET   QUIZ MODE

Distance Matrix: Concept Mode

Step 2: After completing the Matrix, the algorithm joins the sequences with the fewest differences (the most similar sequences).

In this example: Sequence 2 and 5 are the most similar. They are the first to be joined.

These are the first to be joined in the phylogeny (nearest neighbors).

(S1,(S2,S5),S3,S4)

STEP   RESET   QUIZ MODE

Distance Matrix: Concept Mode

Step 3: After joining the first two tips, the sequences are joined as a single group and the distance matrix is recalculated. Again, the closest pair of sequences (or sequences groups) are joined in the tree.
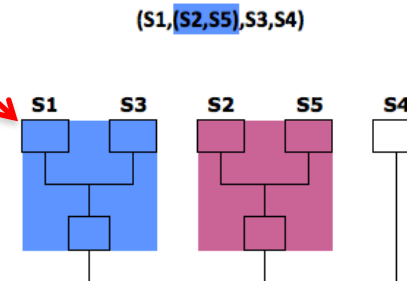
Recalculated Matrix: In this example, the matrix averages the distances of S2 and S5 to all other sequences.

	S1	(S2,S5)	S3	S4
S1	-			
(S2,S5)	0.3	-		
S3	0.2	0.3	-	
S4	0.5	0.6	0.5	-
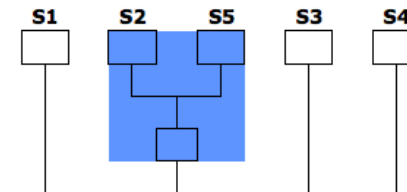
Then the algorithm finds the smallest distance and joins that pair. In this case S1 and S3.

(S1,**(S2,S5)**,S3,S4)

The line below the phylogeny is the text version of the tree. Notice how the parentheses indicate the relationships.

	(S1,S3)	(S2,S5)	S4
(S1,S3)	-		
(S2,S5)	0.3	-	
S4	0.5	0.6	-

The process is repeated until there is only one pair left.

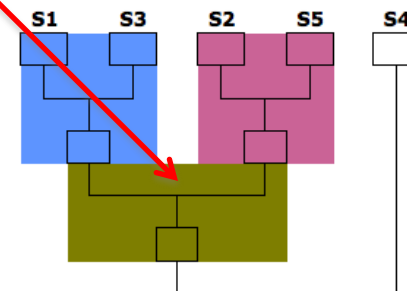(**(S1,S3)**,**(S2,S5)**,S4)

Example Recalculation:
The distance of (S1,S3) to S4 is 0.5
The distance of (S2,S5) to S4 is 0.6

The average of these two is 0.55

	((S1,S3),(S2,S5))	S4
((S1,S3),(S2,S5))	-	
S4	0.55	-

(**((S1,S3),(S2,S5))**,S4)

Distance Matrix: Quiz Mode

In the QUIZ mode, the user calculates the sequence differences and proportion of differences among the sequences, checking the values in the distance matrix tables.