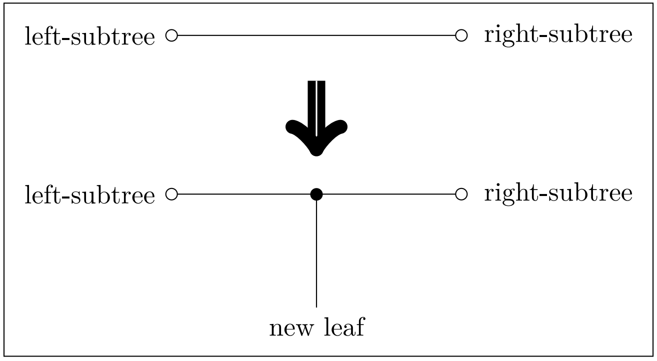
Genome Informatics 2020 Final exam.

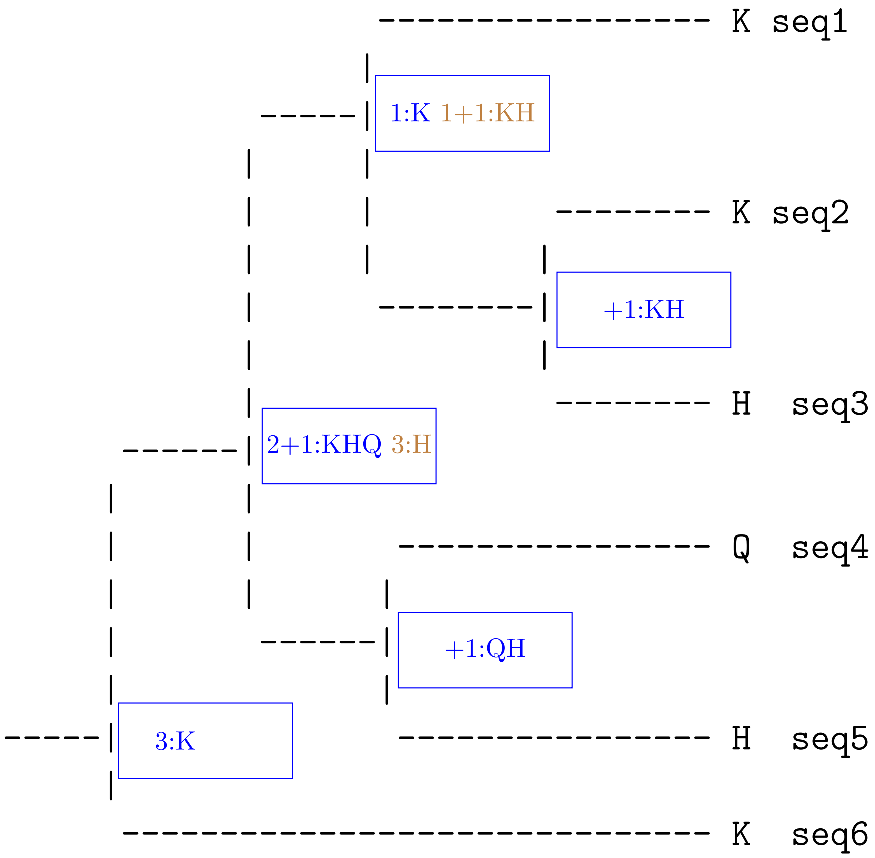
Your Name: ____

Problem 1.
The following problem asked about phylogenetic trees (binary tree graphs) with 6 leaves {a,b,c,d,e,f}. Here by binary tree, we mean complete, but not necessarily balanced trees. As always, explain your answers (okay to use a memorized formula, but justify it in any case).
Problem 1a How many nodes (including leaves) does an unrooted tree with 6 leaves have?
Problem 1b How many nodes (including leaves and the root) does an *rooted* tree with 6 leaves have?
Problem 1c How many distinct unrooted trees exist with 6 leaves.
Problem 1d How many distinct *rooted* trees exist with 6 leaves?

Solution:
Observation 1: in a finite size complete binary tree with internal nodes, at least one of the internal nodes must connect two leaves.
Observation 2: a 2-leaf unrooted tree obviously has just two nodes.
Problem 1a If a 6-leaf unrooted tree has k_6 leaves and a total of n_6 nodes, we can merge two neighbors leaves and the internal node joining them into a single new leaf node; thus producing a 5-leaf unrooted tree with $k_5 = k_6 - 1$ leaves and $n_5 = n_6 - 2$ total nodes. Repeat this thrice more to obtain a 2-leaf unrooted tree with $n_6 - 2 \times 4$ nodes. By Obseration 2 above, $n_6 - 2 \times 4 = 2$, so $n_6 = 10$.
Problem 1b A rooted tree has one more, so a rooted tree with 6 leaves has 11 nodes.
Problem 1c To enumerate tree topologies, it seems easier to use induction from a base case. In complete binary trees, internal nodes have degree 3, so it is easy to see that the only such tree with three leaves has a single internal node connecting the three leaves. Using e and t to denote the number of nodes and topologies, we have $e_3 = 3$, $t_3 = 1$. If we want to add a 4th leaf, we can add a new edge to the middle of any existing edge, to produce a new tree with 4 leaves and two more edges than before; so $t_4 = e_3 = 3$ and $e_4 = e_3 + 2 = 5$. The figure below illustrates adding a leaf to a tree with $i - 1$ leaves, by splitting the tree into a left and right subtree. By definition, there are t_{i-1} trees with $i - 1$ leaves; and each one has e_{i-1} edges to which a new leaf could be attached, so the general recursion formula for the number of trees with one more leaf is $t_i = t_{i-1} e_{i-1}$. Applying this forward: $t_5 = t_4 e_4 = 3 \times 5 = 15$, $e_5 = e_4 + 2 = 7$, $t_6 = t_5 e_5 = 15 \times 7 = 105$.
Problem 1d When converting an unrooted tree to a rooted one, the root may be placed on any edge of the unrooted tree. How many edges does an unrooted tree with 6 leaves have? By the logic in the figure below, it should have 2 more edges than one with 5 leaves, and repeating this logic should have $3 \times 2 = 6$ more edges than the base case of a tree with 3 leaves and 3 edges. Combining information on trees with 6 leaves, we have 105 unrooted tree topologies, each with 9 edges at which the root can be place for a total of $105 \times 9 = 945$ rooted trees.



Problem 2.



The graph above represents a phylogenetic tree; each leaf shows the amino acid. The amino acids come from a column of a well aligned multiple alignment of a protein common to all of the leaf species. Internal node represent ancestor species.

Use maximum parsimony to infer minimum sets of substitutions which could explain the leaf amino acids. In the boxes, write

Problem 2a the minimum number of substitutions needed to explain the leaves under that ancestor.

Problem 2b the amino acid(s) of the ancestor which are compatible with that minimum number of substitutions. and write the minimum number of substitutions needed to explain the leaves under that Use maximum parsimony to infer the ancestor amino acid(s) compatible with the minimum number of substitutions which can describe this tree.

Solution: Results of inference by the traditional parsimony inference algorithm are shown in blue in the figure above. New substitutions which must be postulated when combining subtrees are indicated by '+1'. Additional possibilities for minimal substitution reconstruction of the subtree {seq1, seq2, seq3} are shown in brown. The traditional parsimony algorithm would infer only K for this common ancestor, but assuming H also allows reconstructing the tree {seq1, . . ., seq5} with a minimum number of substitutions (total of 3). However in this particular case seq6 is K, so assuming K allows for fewer substitutions when reconstructing the entire tree.

Your Name: ______________________________

Problem 3.

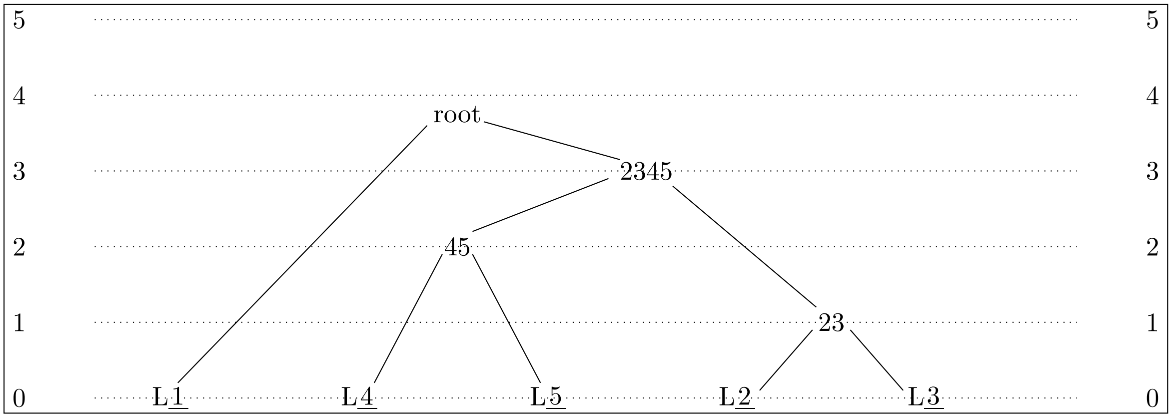
					inferred nodes		
	L2	L3	L4	L5	23	45	2345
L1	9	9	7	5	9	6	7.5_4
	L2	2_1	6	6	X	X	X
		L3	6	6	X	X	X
			L4	4_2	6	X	X
				L5	6	X	X
					23	6_3	X
						45	X
							2345

Problem 3a Fill in all relevant distances computed by UPGMA.
Relevant distances between nodes (leaves or inferred) are shown, with “23” indicating an inferred nodes with leaves 2 and 3. ’X’ indicates distances between leaves and inferred nodes containing those leaves, which are not necessary for UPGMA. Distances which where the minimum at step i of UPGMA are shown in blue with subscript i, i.e. 4_2, indicates that 4 was the minimum distance at step 2 of UPGMa. Note that in the third step two distances were tied at a minimum 6. The solution shown merges 23 with 45 to produce 2345; but merging L1 with 45 to produce 145 would have been an equally good solution, and interestingly the root height would be 3.5 which is different from the solution shown.

Problem 3b Sketch the UPGMA inferred tree on the next page.

Problem 3c Is this UPGMA tree reliable? State your reasons.
Unfortunately, the UPGMA tree is not reliable because the leaf distances given are not ultrametric; e.g. the distances between L1, L4 and L5 are {7,5,4}, none of which are equal to each other.

Sketch the tree inferred by UPGMA, with the leaves at height 0 and inferred ancestor nodes at the appropriate height. Indicate leaves by filling in “L1”, “L2” etc. where there is a “L_”. Show the root node as “root”.



Extra Credit:
Try inferring the tree by a method other than UPGMA. Justify the method used and discuss the reliability of the inferred tree. Show your work above (or on the back of a page) and sketch the tree in the box below.

Check Additivity
We observe that the distances given are not ultrametric. Thus we might check if the distances are at least additive. Consider 4 leaves with interleaf distances $\{d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34}\}$. These distances are additive iff the three sums $(d_{12} + d_{34}), (d_{13} + d_{24}), (d_{14} + d_{23})$ are equal, or two of them are and the third is shorter. A set of interleaf distances a tree with $L \geq 4$ leaves is additive iff all size four subsets of its leaves are. Let's check if our distances are additive.

$$
\begin{aligned}
&(d_{12} + d_{34}), (d_{13} + d_{24}), (d_{14} + d_{23}) = (9+6), (9+6), (7+2) = 15, 15, 9 \checkmark \\
&(d_{12} + d_{35}), (d_{13} + d_{25}), (d_{15} + d_{23}) = (9+6), (9+6), (5+2) = 15, 15, 7 \checkmark \\
&(d_{12} + d_{45}), (d_{15} + d_{24}), (d_{14} + d_{25}) = (9+4), (5+6), (7+6) = 13, 11, 13 \checkmark \\
&(d_{15} + d_{34}), (d_{13} + d_{54}), (d_{14} + d_{35}) = (5+6), (9+4), (7+6) = 11, 13, 13 \checkmark \\
&(d_{25} + d_{34}), (d_{35} + d_{24}), (d_{45} + d_{23}) = (6+6), (6+6), (4+2) = 12, 12, 6 \checkmark
\end{aligned}
$$

The distances are additive. Thus we could use use the Saitou & Nei neighbor joining algorithm to compute an unrooted tree.

Normalized Distance Neighbor Joining
The first step in neighbor joining is computing the normalization factors r_i.

$$
\begin{aligned}
r_i &= \frac{1}{|L| - 2} \sum_{k \in L} d_{ik} & r_1 &= \tfrac{1}{3}(9+9+7+5) = 10 \\
r_2 &= \tfrac{1}{3}(9+2+6+6) = 7\tfrac{2}{3} & r_3 &= \tfrac{1}{3}(9+2+6+6) = 7\tfrac{2}{3} \\
r_4 &= \tfrac{1}{3}(7+6+6+4) = 7\tfrac{2}{3} & r_5 &= \tfrac{1}{3}(5+6+6+4) = 7
\end{aligned}
$$

Using the formula:

$$D_{ij} = d_{ij} - (r_i + r_j) \tag{1}$$

The normalized distances would become:

	L2	L3	L4	L5
L1:	9 - 10 - $7\frac{2}{3}$ = -$8\frac{2}{3}$	9 - 10 - $7\frac{2}{3}$ = -$8\frac{2}{3}$	7 - 10 - $7\frac{2}{3}$ = -$10\frac{2}{3}$	5 - 10 - 7 = -12
	L2:	2 - $7\frac{2}{3}$ - $7\frac{2}{3}$ = -$13\frac{1}{3}$	6 - $7\frac{2}{3}$ - $7\frac{2}{3}$ = -$9\frac{1}{3}$	6 - $7\frac{2}{3}$ - 7 = -$8\frac{2}{3}$
		L3:	6 - $7\frac{2}{3}$ - $7\frac{2}{3}$ = -$9\frac{1}{3}$	6 - $7\frac{2}{3}$ - 7 = -$8\frac{2}{3}$
			L4:	4 - $7\frac{2}{3}$ - 7 = -$10\frac{2}{3}$

The distance between L2 and L3 is the smallest (most negative), so we add a node N23 placed at distances:

$$
\begin{aligned}
d_{1,23} &= \tfrac{1}{2}(d_{1,2} + d_{1,3} - d_{2,3}) = \tfrac{1}{2}(9+9-2) = 8 \\
d_{23,4} &= \tfrac{1}{2}(d_{2,4} + d_{3,4} - d_{2,3}) = \tfrac{1}{2}(6+6-2) = 5 \\
d_{23,5} &= \tfrac{1}{2}(d_{2,5} + d_{3,5} - d_{2,3}) = \tfrac{1}{2}(6+6-2) = 5 \\
d(N23, L2) &= \tfrac{1}{2}(d_{2,3} + r_2 - r_3) = \tfrac{1}{2}(d_{23} + 7\tfrac{2}{3} - 7\tfrac{2}{3}) = \tfrac{1}{2}d_{23} = 1 \\
d(N23, L3) &= \tfrac{1}{2}(d_{2,3} + r_3 - r_2) = \tfrac{1}{2}(d_{23} + 7\tfrac{2}{3} - 7\tfrac{2}{3}) = \tfrac{1}{2}d_{23} = 1
\end{aligned}
$$

After merging L2 and L3, the new R's become:

$$\begin{aligned}
r_1 &= \tfrac{1}{2}(d_{1,23} + d_{1,4} + d_{1,5}) &&= \tfrac{1}{2}(8+7+5) = 10\\
r_{23} &= \tfrac{1}{2}(d_{1,23} + d_{23,4} + d_{23,5}) &&= \tfrac{1}{2}(8+5+5) = 9\\
r_4 &= \tfrac{1}{2}(d_{1,4} + d_{23,4} + d_{4,5}) &&= \tfrac{1}{2}(7+5+4) = 8\\
r_5 &= \tfrac{1}{2}(d_{1,5} + d_{23,5} + d_{4,5}) &&= \tfrac{1}{2}(5+5+4) = 7
\end{aligned}$$

And the distances after merging L2 and L3 are:

	N23	L4	L5
L1:	8 - 10 - 9 = -11	7 - 10 - 8 = -11	5 - 10 - 7 = -12
	N23:	5 - 9 - 8 = -12	5 - 9 - 7 = -11
		L4:	4 - 8 - 7 = -11

The closest pair is L1, L5; so create node N15 placed at distances:

$$\begin{aligned}
d_{15,23} &= \tfrac{1}{2}(d_{1,23} + d_{23,5} - d_{1,5}) = \tfrac{1}{2}(8+5-5) = 4\\
d_{15,4} &= \tfrac{1}{2}(d_{1,4} + d_{4,5} - d_{1,5}) = \tfrac{1}{2}(7+4-5) = 3\\
d(L1, N15) &= \tfrac{1}{2}(d_{1,5} + r_1 - r_5) = \tfrac{1}{2}(5+10-7) = 4\\
d(L5, N15) &= \tfrac{1}{2}(d_{1,5} + r_5 - r_1) = \tfrac{1}{2}(5+7-10) = 1
\end{aligned}$$

After merging L1, L5, the new R's become:

$$\begin{aligned}
r_{15} &= d_{15,23} + d_{15,4} = 4+3 = 7\\
r_{23} &= d_{15,23} + d_{23,4} = 4+5 = 9\\
r_4 &= d_{15,4} + d_{23,4} = 3+5 = 8
\end{aligned}$$

And the distances after merging L1 and L5 are:

	N23	L4
N15:	$d_{15,23} - r_{15} - r_{23} = 4 - 7 - 9 = -12$	$d_{15,4} - r_{15} - r_4 = 3 - 7 - 8 = -12$
	N23:	$d_{23,4} - r_{23} - r_4 = 5 - 9 - 8 = -12$

All normalized distances are now equal: $D_{15,23} = D_{15,4} = D_{23,4} = -12$.
I arbitrarily choose to merge N23 with L4 creating node N234 with distances:

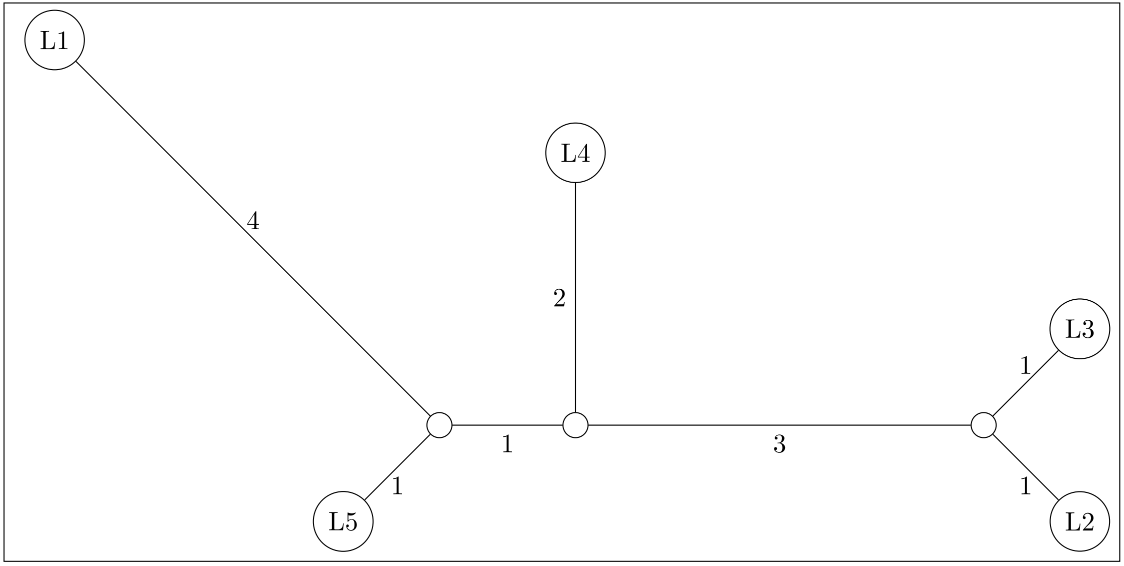
$$\begin{aligned}
d_{15,234} &= \tfrac{1}{2}(d_{15,23} + d_{15,4} - d_{23,4}) = \tfrac{1}{2}(4+3-5) = 1\\
d(N23, N234) &= \tfrac{1}{2}(d_{23,4} + r_{23} - r_4) = \tfrac{1}{2}(5+9-8) = 3\\
d(L4, N234) &= \tfrac{1}{2}(d_{23,4} + r_4 - r_{23}) = \tfrac{1}{2}(5+8-9) = 2
\end{aligned}$$

There are only 2 nodes now, which should be connected by distance $d_{15,234} = 1$.

For convenience let’s gather the edge length information computed on the previous pages:

edge	len	edge	len
L2,N23	1	L3,N23	1
L1,N15	4	L5,N15	1
N23,N234	3	L4,N234	2
N15,N234	1		

Which can be plotted in this way:



Unrooted tree inferred by the Neighbor-Joining method of Saitou & Nei

UPGMA after distance “correction” In this section I inferred a rooted tree in an *ad hoc* manor, inspired by the distance normalization technique to get a hint for an *ad hoc* way to correct the distances so that UPGMA can be used.

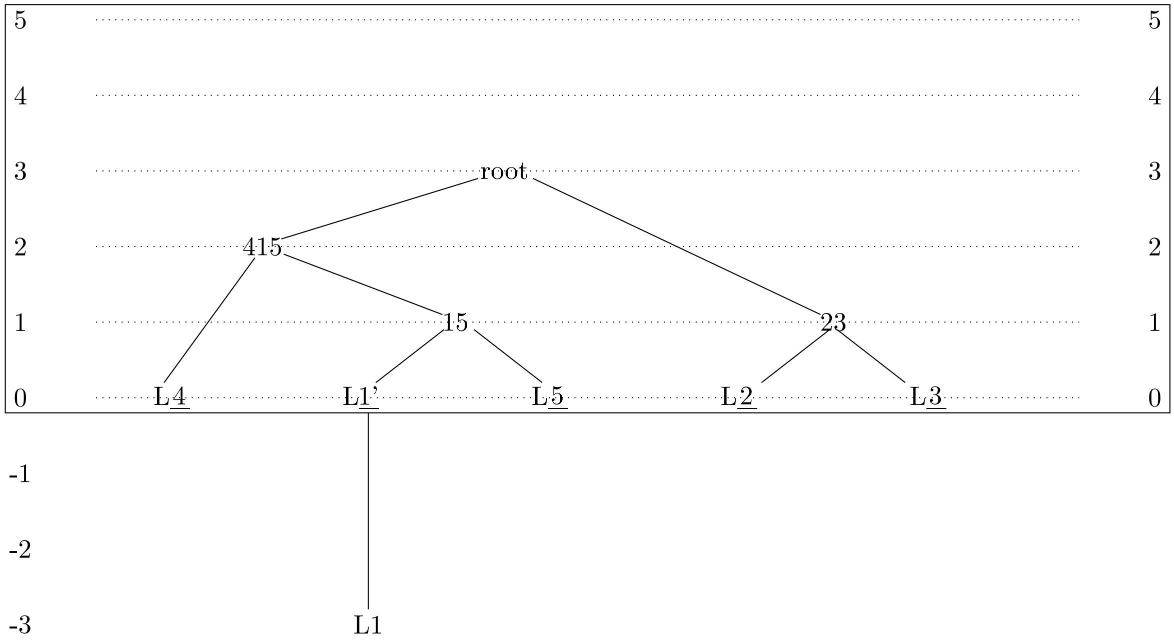
Notice that since $r_2 = r_3$ they cancel out when computing edge lengths and the result is exactly the same as the edge height which UPGMA would give. The reason for this is that the original distances are *almost* ultrametric. In fact if one ignores, leaf L1, the remaining distances are ultrametric (all 6 except for $d_{23} = 2$ and $d_{45} = 4$, see the light green squares in the distance matrix in the figure on the previous page). Noting that r_1 is much bigger than the other r_i’s, one might suspect that L1 has experienced accelerated evolution and its distance to other nodes can perhaps be adjusted by subtracting a constant k. In other words look for a subtrahend k which can make the distances additive.

$$\begin{array}{rl}
 & k = 3 \quad \text{Ultrametric?} \\
(d_{12} - k, d_{13} - k, d_{23}) = (9 - k, 9 - k, 2) \rightarrow & (6, 6, 2) \checkmark \\
(d_{12} - k, d_{14} - k, d_{23}) = (9 - k, 7 - k, 6) \rightarrow & (6, 4, 6) \checkmark \\
(d_{12} - k, d_{15} - k, d_{25}) = (9 - k, 5 - k, 6) \rightarrow & (6, 2, 6) \checkmark \\
(d_{13} - k, d_{14} - k, d_{34}) = (9 - k, 7 - k, 6) \rightarrow & (6, 4, 6) \checkmark \\
(d_{13} - k, d_{15} - k, d_{35}) = (9 - k, 5 - k, 6) \rightarrow & (6, 2, 6) \checkmark
\end{array}$$

So an idea is to perform UPGMA on the distances after substracting 3 from all distances involving L1.

	L2	L3	L4	L5	23	15	415
L1':	6	6	4	2	6	X	X
L2:		2	6	6	X	X	X
L3:			6	6	X	X	X
L4:				4	6	4	X
L5:					6	X	X
23:						6	6
15:							X

Results of UPGMA after "correcting" distances from L1 is shown below. L1 is placed at -3 to indicate its actual distance from other leaves.



Your Name: ____

Problem 4.
The following page has shows 6 amino acid scoring matrices (look on the back as well). One of them is the PAM70 matrix. Select it and explain why.

The answer is **C**.
Why? Considering that PAM70 reflects an evolutionary distance of 70 accepted point mutations per 100 amino acid positions, mismatches between highly similar amino acids should be rewarded or at least not penalized heavily. Thus we can eliminate matrices {**B**,**D**,**E**} since all of their mismatch scores are negative (**B** is in fact the PAM10 matrix and the others are shuffled versions of PAM10). To choose among {A,C,F} we need some background knowledge about which amino acids are similar to each other. From examples we worked on in class you should know that similar amino acids pairs include {(K,R), (D,E), (I,L), (S,T)}. Although matrix **F** happens to give (K,R) a good score, matrix **C** is the only one giving good scores to all those pairs of similar amino acids.

A

```
    A   R   N   D   C   Q   E   G   H   I   L   K   M   F   P   S   T   W   Y   V
A   8  -5  -1  -2  -4  -5   0  -3  -4  -6   2   1  -4  -2  -4  -3  -1  -4  -6  -6
R  -5   9  -2  -5  -8 -11  -5  -1  -4  -9  -9  -7 -10  -9  -4  -9  -9  -5  -6  -4
N  -1  -2   9  -9   4  -3  -7  -5  -5  -7  -8  -3  -4  -6  -5  -7  -7  -4  -9  -4
D  -2  -5  -9   7  -7  -9  -2   0   0  -5  -1  -3  -5  -3  -3  -4  -4  -2  -3  -5
C  -4  -8   4  -7   8  -2  -7  -4  -6  -2  -9  -6  -1  -9  -5  -9 -10  -6  -7   0
Q  -5 -11  -3  -9  -2  13   0  -3  -9  -8  -8  -6  -4 -11 -10  -7 -10  -8 -10  -9
E   0  -5  -7  -2  -7   0   8  -1  -4  -2   0  -3  -6  -5  -5   2  -6  -4  -6  -3
G  -3  -1  -5   0  -4  -3  -1   5   1  -3  -3   1  -6  -2  -3  -2  -1   2   0  -4
H  -4  -4  -5   0  -6  -9  -4   1   5  -3  -2  -2  -4  -1  -1  -4  -1   1   0  -2
I  -6  -9  -7  -5  -2  -8  -2  -3  -3  10  -2  -5   2  -4   0   0  -7  -2  -6   1
L   2  -9  -8  -1  -9  -8   0  -3  -2  -2   7  -1  -3   2  -4  -1   0  -3  -4  -5
K   1  -7  -3  -3  -6  -6  -3   1  -2  -5  -1   6  -5   0  -5   0   3   0  -1  -3
M  -4 -10  -4  -5  -1  -4  -6  -6  -4   2  -3  -5   6  -6   0  -5  -8  -4  -7   1
F  -2  -9  -6  -3  -9 -11  -5  -2  -1  -4   2   0  -6   6  -4  -2   3  -3  -2  -4
P  -4  -4  -5  -3  -5 -10  -5  -3  -1   0  -4  -5   0  -4   6  -6  -5  -1  -3   3
S  -3  -9  -7  -4  -9  -7   2  -2  -4   0  -1   0  -5  -2  -6   6  -2  -1  -5  -4
T  -1  -9  -7  -4 -10 -10  -6  -1  -1  -7   0   3  -8   3  -5  -2   6  -2  -1  -5
W  -4  -5  -4  -2  -6  -8  -4   2   1  -2  -3   0  -4  -3  -1  -1  -2   6  -3  -1
Y  -6  -6  -9  -3  -7 -10  -6   0   0  -6  -4  -1  -7  -2  -3  -5  -1  -3   6  -6
V  -6  -4  -4  -5   0  -9  -3  -4  -2   1  -5  -3   1  -4   3  -4  -5  -1  -6   7
```

B

```
    A   R   N   D   C   Q   E   G   H   I   L   K   M   F   P   S   T   W   Y   V
A   7 -10  -7  -6 -10  -7  -5  -4 -11  -8  -9 -10  -8 -12  -4  -3  -3 -20 -11  -5
R -10   9  -9 -17 -11  -4 -15 -13  -4  -8 -12  -2  -7 -12  -7  -6 -10  -5 -14 -11
N  -7  -9   9  -1 -17  -7  -5  -6  -2  -8 -10  -4 -15 -12  -9  -2  -5 -11  -7 -12
D  -6 -17  -1   8 -21  -6   0  -6  -7 -11 -19  -8 -17 -21 -12  -7  -8 -21 -17 -11
C -10 -11 -17 -21  10 -20 -20 -13 -10  -9 -21 -20 -20 -19 -11  -6 -11 -22  -7  -9
Q  -7  -4  -7  -6 -20   9  -1 -10  -2 -11  -8  -6  -7 -19  -6  -8  -9 -19 -18 -10
E  -5 -15  -5   0 -20  -1   8  -7  -9  -8 -13  -7 -10 -20  -9  -7  -9 -23 -11 -10
G  -4 -13  -6  -6 -13 -10  -7   7 -13 -17 -14 -10 -12 -12 -10  -4 -10 -21 -20  -9
H -11  -4  -2  -7 -10  -2  -9 -13  10 -13  -9 -10 -17  -9  -7  -9 -11 -10  -6  -9
I  -8  -8  -8 -11  -9 -11  -8 -17 -13   9  -4  -9  -3  -5 -12 -10  -5 -20  -9  -1
L  -9 -12 -10 -19 -21  -8 -13 -14  -9  -4   7 -11  -2  -5 -10 -12 -10  -9 -10  -5
K -10  -2  -4  -8 -20  -6  -7 -10 -10  -9 -11   7  -4 -20 -10  -7  -6 -18 -12 -13
M  -8  -7 -15 -17 -20  -7 -10 -12 -17  -3  -2  -4  12  -7 -11  -8  -7 -19 -17  -4
F -12 -12 -12 -21 -19 -19 -20 -12  -9  -5  -5 -20  -7   9 -13  -9 -12  -7  -1 -12
P  -4  -7  -9 -12 -11  -6  -9 -10  -7 -12 -10 -10 -11 -13   8  -4  -7 -20 -20  -9
S  -3  -6  -2  -7  -6  -8  -7  -4  -9 -10 -12  -7  -8  -9  -4   7  -2  -8 -10 -10
T  -3 -10  -5  -8 -11  -9  -9 -10 -11  -5 -10  -6  -7 -12  -7  -2   8 -19  -9  -6
W -20  -5 -11 -21 -22 -19 -23 -21 -10 -20  -9 -18 -19  -7 -20  -8 -19  13  -8 -22
Y -11 -14  -7 -17  -7 -18 -11 -20  -6  -9 -10 -12 -17  -1 -20 -10  -9  -8  10 -10
V  -5 -11 -12 -11  -9 -10 -10  -9  -9  -1  -5 -13  -4 -12  -9 -10  -6 -22 -10   8
```

C

```
    A   R   N   D   C   Q   E   G   H   I   L   K   M   F   P   S   T   W   Y   V
A   5  -4  -2  -1  -4  -2  -1   0  -4  -2  -4  -4  -3  -6   0   1   1  -9  -5  -1
R  -4   8  -3  -6  -5   0  -5  -6   0  -3  -6   2  -2  -7  -2  -1  -4   0  -7  -5
N  -2  -3   6   3  -7  -1   0  -1   1  -3  -5   0  -5  -6  -3   1   0  -6  -3  -5
D  -1  -6   3   6  -9   0   3  -1  -1  -5  -8  -2  -7 -10  -4  -1  -2 -10  -7  -5
C  -4  -5  -7  -9   9  -9  -9  -6  -5  -4 -10  -9  -9  -8  -5  -1  -5 -11  -2  -4
Q  -2   0  -1   0  -9   7   2  -4   2  -5  -3  -1  -2  -9  -1  -3  -3  -8  -8  -4
E  -1  -5   0   3  -9   2   6  -2  -2  -4  -6  -2  -4  -9  -3  -2  -3 -11  -6  -4
G   0  -6  -1  -1  -6  -4  -2   6  -6  -6  -7  -5  -6  -7  -3   0  -3 -10  -9  -3
H  -4   0   1  -1  -5   2  -2  -6   8  -6  -4  -3  -6  -4  -2  -3  -4  -5  -1  -4
I  -2  -3  -3  -5  -4  -5  -4  -6  -6   7   1  -4   1   0  -5  -4  -1  -9  -4   3
L  -4  -6  -5  -8 -10  -3  -6  -7  -4   1   6  -5   2  -1  -5  -6  -4  -4  -4   0
K  -4   2   0  -2  -9  -1  -2  -5  -3  -4  -5   6   0  -9  -4  -2  -1  -7  -7  -6
M  -3  -2  -5  -7  -9  -2  -4  -6  -6   1   2   0  10  -2  -5  -3  -2  -8  -7   0
F  -6  -7  -6 -10  -8  -9  -9  -7  -4   0  -1  -9  -2   8  -7  -4  -6  -2   4  -5
P   0  -2  -3  -4  -5  -1  -3  -3  -2  -5  -5  -4  -5  -7   7   0  -2  -9  -9  -3
S   1  -1   1  -1  -1  -3  -2   0  -3  -4  -6  -2  -3  -4   0   5   2  -3  -5  -3
T   1  -4   0  -2  -5  -3  -3  -3  -4  -1  -4  -1  -2  -6  -2   2   6  -8  -4  -1
W  -9   0  -6 -10 -11  -8 -11 -10  -5  -9  -4  -7  -8  -2  -9  -3  -8  13  -3 -10
Y  -5  -7  -3  -7  -2  -8  -6  -9  -1  -4  -4  -7  -7   4  -9  -5  -4  -3   9  -5
V  -1  -5  -5  -5  -4  -4  -4  -3  -4   3   0  -6   0  -5  -3  -3  -1 -10  -5   6
```

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	9	-7	-7	-6	-7	-2	-6	-6	-11	-18	-19	-10	-20	-8	-1	-9	-8	-19	-10	-4
R	-7	12	-8	-17	-15	-17	-11	-4	-3	-17	-19	-12	-20	-2	-10	-7	-8	-7	-4	-7
N	-7	-8	7	-6	-7	-11	-4	-10	-8	-11	-20	-4	-10	-9	-5	-3	-3	-12	-5	-10
D	-6	-17	-6	8	-1	-7	-12	-8	-11	-17	-21	-6	-21	-19	0	-8	-7	-21	-11	-17
C	-7	-15	-7	-1	9	-2	-9	-4	-8	-7	-11	-6	-17	-10	-5	-5	-2	-12	-12	-9
Q	-2	-17	-11	-7	-2	10	-7	-10	-13	-6	-10	-13	-10	-9	-9	-11	-9	-9	-9	-4
E	-6	-11	-4	-12	-9	-7	8	-10	-12	-20	-20	-10	-11	-10	-9	-7	-4	-13	-9	-7
G	-6	-4	-10	-8	-4	-10	-10	7	-9	-12	-18	-10	-20	-11	-7	-6	-7	-20	-13	-2
H	-11	-3	-8	-11	-8	-13	-12	-9	9	-9	-20	-17	-9	-4	-8	-5	-10	-5	-1	-8
I	-18	-17	-11	-17	-7	-6	-20	-12	-9	10	-8	-20	-7	-10	-11	-9	-10	-1	-10	-14
L	-19	-19	-20	-21	-11	-10	-20	-18	-20	-8	13	-21	-22	-9	-23	-19	-8	-7	-22	-5
K	-10	-12	-4	-6	-6	-13	-10	-10	-17	-20	-21	7	-13	-14	-7	-10	-4	-12	-9	-13
M	-20	-20	-10	-21	-17	-10	-11	-20	-9	-7	-22	-13	10	-21	-20	-11	-6	-19	-9	-11
F	-8	-2	-9	-19	-10	-9	-10	-11	-4	-10	-9	-14	-21	7	-13	-10	-12	-5	-5	-12
P	-1	-10	-5	0	-5	-9	-9	-7	-8	-11	-23	-7	-20	-13	8	-9	-7	-20	-10	-15
S	-9	-7	-3	-8	-5	-11	-7	-6	-5	-9	-19	-10	-11	-10	-9	8	-2	-12	-6	-10
T	-8	-8	-3	-7	-2	-9	-4	-7	-10	-10	-8	-4	-6	-12	-7	-2	7	-9	-10	-6
W	-19	-7	-12	-21	-12	-9	-13	-20	-5	-1	-7	-12	-19	-5	-20	-12	-9	9	-12	-12
Y	-10	-4	-5	-11	-12	-9	-9	-13	-1	-10	-22	-9	-9	-5	-10	-6	-10	-12	8	-11
V	-4	-7	-10	-17	-9	-4	-7	-2	-8	-14	-5	-13	-11	-12	-15	-10	-6	-12	-11	9

D

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	9	-12	-9	-8	-17	-3	-20	-8	-5	-8	-4	-9	-13	-1	-5	-10	-11	-9	-11	-8
R	-12	8	-20	-9	-10	-11	-20	-7	-7	-9	-10	-10	-7	-9	-13	-4	-6	-11	-12	-4
N	-9	-20	10	-11	-20	-17	-8	-14	-9	-7	-10	-12	-6	-10	-1	-10	-18	-7	-17	-11
D	-8	-9	-11	8	-7	-10	-23	-15	-9	-5	-13	-7	-9	-10	-20	-7	-1	-20	0	-5
C	-17	-10	-20	-7	7	-12	-21	-13	-10	-6	-14	-10	-13	-9	-12	-4	-10	-13	-6	-4
Q	-3	-11	-17	-10	-12	12	-19	-7	-7	-15	-2	-4	-17	-4	-7	-8	-7	-20	-17	-8
E	-20	-20	-8	-23	-21	-19	13	-5	-19	-11	-9	-18	-10	-22	-7	-8	-19	-22	-21	-20
G	-8	-7	-14	-15	-13	-7	-5	9	-10	-9	-12	-2	-4	-11	-12	-6	-4	-11	-17	-10
H	-5	-7	-9	-9	-10	-7	-19	-10	8	-5	-10	-6	-11	-6	-12	-2	-9	-11	-8	-3
I	-8	-9	-7	-5	-6	-15	-11	-9	-5	9	-10	-4	-2	-12	-12	-2	-7	-17	-1	-7
L	-4	-10	-10	-13	-14	-2	-9	-12	-10	-10	7	-11	-9	-5	-5	-12	-8	-21	-19	-9
K	-9	-10	-12	-7	-10	-4	-18	-2	-6	-4	-11	7	-10	-13	-20	-7	-6	-20	-8	-10
M	-13	-7	-6	-9	-13	-17	-10	-4	-11	-2	-9	-10	10	-9	-9	-9	-2	-10	-7	-11
F	-1	-9	-10	-10	-9	-4	-22	-11	-6	-12	-5	-13	-9	8	-12	-10	-10	-9	-11	-5
P	-5	-13	-1	-20	-12	-7	-7	-12	-12	-12	-5	-20	-9	-12	9	-9	-19	-19	-21	-12
S	-10	-4	-10	-7	-4	-8	-8	-6	-2	-2	-12	-7	-9	-10	-9	7	-8	-6	-7	-3
T	-11	-6	-18	-1	-10	-7	-19	-4	-9	-7	-8	-6	-2	-10	-19	-8	9	-20	-6	-7
W	-9	-11	-7	-20	-13	-20	-22	-11	-11	-17	-21	-20	-10	-9	-19	-6	-20	10	-21	-10
Y	-11	-12	-17	0	-6	-17	-21	-17	-8	-1	-19	-8	-7	-11	-21	-7	-6	-21	8	-6
V	-8	-4	-11	-5	-4	-8	-20	-10	-3	-7	-9	-10	-11	-5	-12	-3	-7	-10	-6	7

E

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	8	-5	0	2	-5	-1	-2	-7	-4	-6	-6	-6	-2	-4	0	-7	-3	-5	0	-3
R	-5	6	-11	-2	-4	-2	-4	-9	-3	-6	-2	3	-3	-1	2	-6	0	-9	-2	-4
N	0	-11	13	-7	-10	-3	-8	-2	-8	-4	-10	-10	-9	-9	-8	-3	-6	-11	-5	-9
D	2	-2	-7	6	-6	-2	0	-9	-1	-5	-5	-2	-4	-4	-1	-7	0	-9	-3	-4
C	-5	-4	-10	-6	6	-3	0	-5	-1	0	-3	-5	-3	-1	-4	-5	-5	-4	-4	3
Q	-1	-2	-3	-2	-3	5	-3	-4	2	-6	0	-1	0	1	-3	-5	1	-1	-3	-4
E	-2	-4	-8	0	0	-3	10	-2	-2	2	-6	-7	-5	-3	-2	-7	-5	-9	-6	1
G	-7	-9	-2	-9	-5	-4	-2	8	-6	-1	-7	-10	-7	-6	-9	4	-6	-8	-4	0
H	-4	-3	-8	-1	-1	2	-2	-6	6	-4	-3	-2	-2	1	-3	-4	0	-5	-4	-1
I	-6	-6	-4	-5	0	-6	2	-1	-4	6	-7	-8	-5	-4	-3	-4	-5	-10	-4	1
L	-6	-2	-10	-5	-3	0	-6	-7	-3	-7	6	-1	-3	0	-4	-9	-1	-6	-6	-6
K	-6	3	-10	-2	-5	-1	-7	-10	-2	-8	-1	6	-4	-1	0	-7	3	-9	-1	-5
M	-2	-3	-9	-4	-3	0	-5	-7	-2	-5	-3	-4	7	0	-1	-9	-3	-5	-2	-5
F	-4	-1	-9	-4	-1	1	-3	-6	1	-4	0	-1	0	5	-2	-5	-2	-4	-4	-2
P	0	2	-8	-1	-4	-3	-2	-9	-3	-3	-4	0	-1	-2	7	-8	-1	-9	2	-5
S	-7	-6	-3	-7	-5	-5	-7	4	-4	-4	-9	-7	-9	-5	-8	9	-3	-2	-1	-4
T	-3	0	-6	0	-5	1	-5	-6	0	-5	-1	3	-3	-2	-1	-3	6	-7	1	-3
W	-5	-9	-11	-9	-4	-1	-9	-8	-5	-10	-6	-9	-5	-4	-9	-2	-7	9	-5	-4
Y	0	-2	-5	-3	-4	-3	-6	-4	-4	-4	-6	-1	-2	-4	2	-1	1	-5	8	-6
V	-3	-4	-9	-4	3	-4	1	0	-1	1	-6	-5	-5	-2	-5	-4	-3	-4	-6	7

F