

Genome Informatics 2020 Final exam.

Exam written by Paul Horton ©2020.

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?

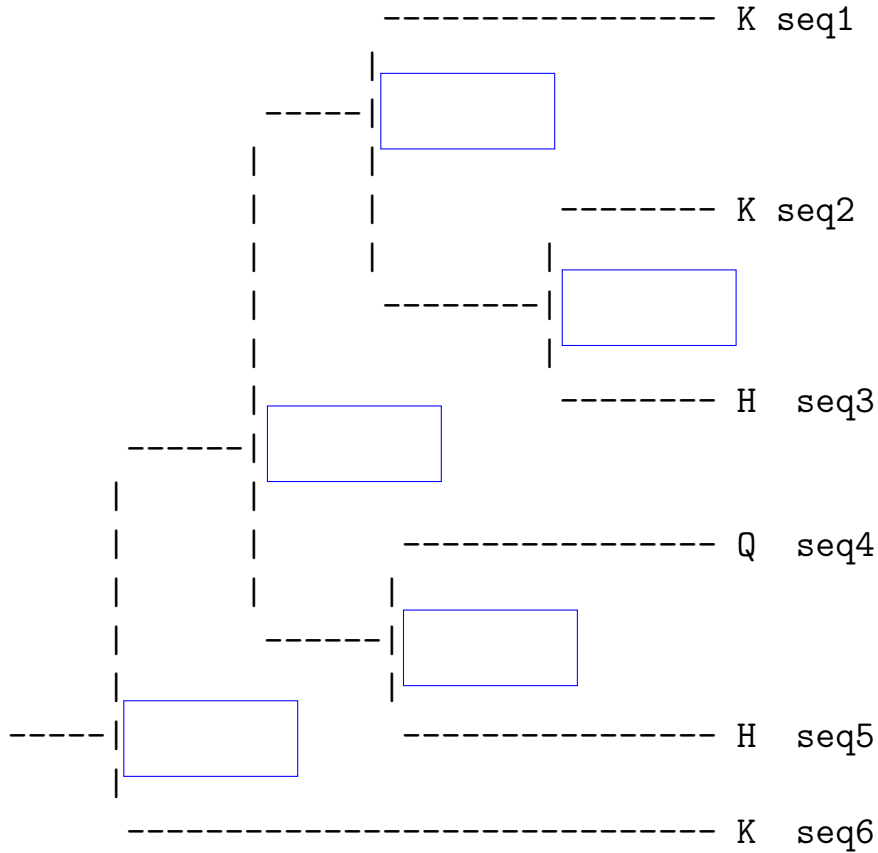
Problem 1b How many nodes (including leaves and the root) does an *rooted* tree with 6 leaves have?

Problem 1c How many distinct unrooted trees exist with 6 leaves.

Problem 1d How many distinct *rooted* trees exist with 6 leaves?

Your Name: _____

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.

Your Name: _____

Problem 3.

	L2	L3	L4	L5	inferred nodes		
L1	9	9	7	5			
	L2	2	6	6			
		L3	6	6			
			L4	4			
				L5			

Problem 3a Fill in all relevant distances computed by UPGMA.

The above matrix shows distances between species L1 to L5. Use UPGMA (Unweighted Pair Group Method Arithmetic averages) to infer a phylogenetic tree. Indicate merged nodes in the blue boxes, so for example write “124” to indicate a node above leaves L1 and L2 and L4.

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

Problem 3c Is this UPGMA tree reliable? State your reasons.

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”.

5		5
4		4
3		3
2		2
1		1
0	<div style="display: flex; justify-content: space-around; padding: 0 10px;"> L_ L_ L_ L_ L_ </div>	0

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.

5		5
4		4
3		3
2		2
1		1
0	<div style="display: flex; justify-content: space-around; padding: 0 10px;"> L_ L_ L_ L_ L_ </div>	0

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.

	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

A

	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

B

	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

C

	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