Final exam, Genome Informatics 20221228 *Write your <u>name</u> on each sheet.*

	B	C	D
A	9	6	4
B	x	13	9
C	x	x	8

Problem 1.

Use UPGMA to infer a possible tree for leaves **A**,**B**,**C**,**D** (topology & edge lengths) from the distance matrix shown. Discuss if the inferred tree is reasonable.

|	B	C	D
A	9	6	4
B	x	13	9
C	x	x	8

Problem 2.
Use Saitou Nei Neighbor Joining (NJ) to infer a possible tree for leaves **A**,**B**,**C**,**D** (topology & edge lengths) from the distance matrix shown.

Problem 3.
Use Maximum Parsimony to reconstruct possible sequences of the ancestor nodes for the tree on the following page. Use notation like [ac] to represent sets. For example "t [ac]" denotes t followed by either a or c.

tgagccc

tcagagc

tgtgact

ttgtgcc

caacacc

ctgtacc

ttggtcc
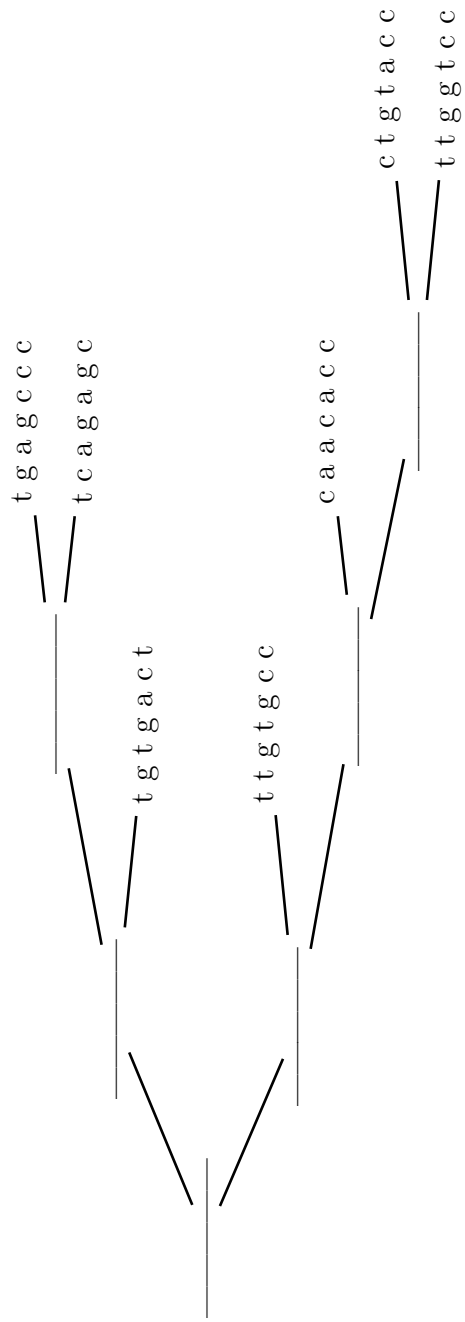
Problem 4.
Assuming the following symmetric substitution costs:

```
     c   g   t
a    3   2   3
c    x   3   2
g    x   x   3
```

Use weighted weighted parsimony to compute the minimum cost of each possible nucleotide for each blank in the tree on the following page.

For notation use vectors in the order [a c g t], for example [6 3 2 5] denotes a minimum cost of 6, 3, 2 or 5; conditioned on the base in that sequence being a,c,g, or t, respectively.

a g

a c a

t g

g