

Genome Informatics 2021 Fall Final exam.

Exam written by Paul Horton ©2021.

Your Name: \_\_\_\_\_

**Problem 1.**

Assume an organism evolves with constant substitution rates such that after  $t = 1000$  generations the probabilities of the form  $P[a|a, t = 1000]$ ,  $P[a|c, t = 1000]$ , etc. are as shown.

$$S(t = 1000) = \begin{bmatrix} & \text{c} & \text{t} & \text{a} & \text{g} \\ \text{c} & 0.95 & 0.03 & 0.01 & 0.01 \\ \text{t} & 0.03 & 0.95 & 0.01 & 0.01 \\ \text{a} & 0.01 & 0.01 & 0.95 & 0.03 \\ \text{g} & 0.01 & 0.01 & 0.03 & 0.95 \end{bmatrix}$$

**Task:** fill in the numbers for  $S(t = 2000)$

	c	t	a	g
c				
t				
a				
g				

Your Name: \_\_\_\_\_

**Problem 2.**

1. Describe precisely the set of DNA strings accepted by the following context free grammar and the size of that set (in other words, how many distinct strings are accepted by that grammar).
2. Give an example of one of the strings, sketch it as a stem-loop structure and show its derivation from the grammar.

$$\begin{aligned} S &\rightarrow cW_1g \mid gW_1c \\ W_1 &\rightarrow aW_2u \mid uW_2a \\ W_2 &\rightarrow aW_3u \mid uW_3a \\ W_3 &\rightarrow cW_4g \mid gW_4c \\ W_4 &\rightarrow aW_5 \mid cW_5 \mid gW_5 \mid uW_5 \\ W_5 &\rightarrow aW_6 \mid cW_6 \mid gW_6 \mid uW_6 \\ W_6 &\rightarrow aW_7 \mid cW_7 \mid gW_7 \mid uW_7 \\ W_7 &\rightarrow a \mid c \mid g \mid u \end{aligned}$$

Your Name: \_\_\_\_\_

**Problem 3.**

	<b>A</b>		
	<i>w</i>	<i>x</i>	<i>y</i>
<i>x</i>	8		
<i>y</i>	10	6	
<i>z</i>	7	11	13

	<b>B</b>		
	<i>w</i>	<i>x</i>	<i>y</i>
<i>x</i>	8		
<i>y</i>	8	4	
<i>z</i>	6	8	8

	<b>C</b>		
	<i>w</i>	<i>x</i>	<i>y</i>
<i>x</i>	8		
<i>y</i>	13	9	
<i>z</i>	6	9	16

The three matrices above indicate distances between 4 species  $\{w, x, y, z\}$ . Two of the three matrices hold additive distances and one of those is also ultrametric.

**Tasks:**

1. Determine which of the three matrices  $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$  are additive and which one is ultrametric (and show the reason).
2. Use UPGMA to infer a phylogenetic tree (with branch lengths) of the one that is ultrametric. Show your calculation.
3. Use Neighbor-Joining to infer a phylogenetic tree (with branch lengths) of the one which is additive but not ultrametric. Show at least the first iteration of the neighbor-joining algorithm.