Bioinformatics for Biologists

Spring 2021

New questions are added at

the end.

Study guide for quiz 1-2

1. Know the full name, 3-letter and 1-letter abbreviation of the 20 amino acids.
2. Know the physicochemical properties (hydrophobicity, size, and charge) of the amino acids.
3. Know how different amino acids approximately relate to each other based on hydrophobicity, size, and charge
4. Be familiar with the amino acid Venn diagram
5. Why are some amino found classified in opposing categories, such as both hydrophobic and charged?

6. How do genomes evolve?
7. What are homologs?
8. What are orthologs?
9. What are paralogs?
10. What are common scenarios for different gene copies after gene duplication?
11. What is an isoform and what is a homolog? How can you tell them apart?
12. What is alternative splicing?

13. What does FASTA format look like?

14. What information does a substitution matrix hold?
15. Give an example of a substitution matrix
16. What is a local alignment?
17. What is a global alignment?
18. Which matrix does BLAST use as default?

19. Why is BLAST faster than Smith-Waterman?
20. What does it mean for an algorithm to be heuristic?
21. What does it mean for an algorithm to be exhaustive?
22. How do gap penalties influence your BLAST results?
23. Know the different databases and query data types for the 5 blast algorithms

    blastn: Search **nucleotide** database using a **nucleotide** query
    blastp: Search **protein** database using a **protein** query
    blastx: Search **protein** database using a **translated nucleotide** query
    tblastn: Search **translated nucleotide** database using a **protein** query
    tblastx: Search **translated nucleotide** database using a **translated nucleotide** query

24. What do you need to evaluate which of the following alignments is better?

PARIS		PARIS
PA--S	or	PAS--

25. What is the score of the alignment, given the following information

    Gap opening -10, gap extension -2.

    A A W A D T F F F A R
    A C W A D S - - - A K

(A)

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

Study guide for quiz 3

(all questions from previous study guides are also included)

26. What is a progressive alignment?
27. State 2 differences between ClustalW to Muscle.
28. What assumption is made for amino acid residues in a column in the multiple sequence alignment?
29. Describe how ClustalW works
30. How does Muscle work?
31. Why do we build multiple sequence alignments?
32. Be familiar with fasta, clustal, and phylip format (will work with these in class on Monday Jan 25)
33. Know how to convert a simple tree format with 4 taxa to a tree and back.
34. What is a species tree?
35. Where can you extract a species tree?
36. What is the purpose of rooting a phylogenetic tree?
37. How can a tree be rooted?

38. What is the procedure for building a phylogenetic tree using a character-based/discretemethod?
39. What are the main components of the model of evolution? What do they describe? When testing for a model of evolution for a MSA, you find that the best model of evolution is LG+G+F+I. What does this tell you about your MSA?
40. How are distance and discrete trees fundamentally different in how they are constructed?
41. Where do the starting and the final trees for the likelihood trees come from?
42. What are two types of support values that can be used to evaluate how supported different clades in the tree?