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 a **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

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

PARIS PARIS

PA--S or PAS--

1. 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 - - - AK



Study guide for quiz 3

(all questions from previous study guides are also included)

1. What is a progressive alignment?
2. State 2 differences between ClustalW to Muscle.
3. What assumption is made for amino acid residues in a column in the multiple sequence alignment?
4. Describe how ClustalW works
5. How does Muscle work?
6. Why do we build multiple sequence alignments?
7. Be familiar with fasta, clustal, and phylip format (will work with these in class on Monday Jan 25)
8. Know how to convert a simple tree format with 4 taxa to a tree and back.
9. What is a species tree?
10. Where can you extract a species tree?
11. What is the purpose of rooting a phylogenetic tree?
12. How can a tree be rooted?
13. What is the procedure for building a phylogenetic tree using a character-based/discretemethod?
14. 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?
15. How are distance and discrete trees fundamentally different in how they are constructed?
16. Where do the starting and the final trees for the likelihood trees come from?
17. What are two types of support values that can be used to evaluate how supported different clades in the tree?

Study guide for quiz 5

All questions from previous study guides and but also questions and applied steps from activities are also included. The quiz will have applied and theoretical questions.

1. What is PDB an abbreviation of?
2. What information can you find in a PDB file?
3. In which file format can you find the x, y, and z coordinates for a protein’s structure?
4. Why can a domain be referred to as an evolutionary unit?
5. What is a fold?
6. How does CATH classify protein structure on the class, architecture, topology (fold), and

Homologous superfamily level?

How many folds are currently known according to CATH?

1. What does the fold distribution look like? Are all folds equally common or are some more common than others? Why may that be?
2. Evaluate if the following statements are true or false:

Different folds evolve with different amino acid substitution rates.

The rate of amino acid substitution depends on the proteins fold and function.

Protein structure and function determine how proteins evolve.

Some protein folds are very common, some protein folds are very rare.

51. What is the difference between a smooth and a rugged protein folding funnel (energy landscape)?

52. How can different signals such as ligands, pH, oxygen, and temperature alter the shape of an energy landscape and consequently, which the main conformations are?

53. What is Cryo-EM (watch the video)?

54. Which are the other 2 primary experimental methods have been used to generate the PDB

data and what are their shortcomings (from the Cryo-EM Nobel video)?

55. What are the current trends for X-ray, NMR, and Cryo-EM protein structure determination?

56. What does the resolution of a protein structure tell you about the protein structure (see

<http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/resolution>)?

57.What can be seen in protein structures at 1Å, 2Å, and 3Å resolution, respectively?

**For THEORY TEST 1, ALL questions from the study guides for previous quizzes and the additional questions below will be included.**

Added 2/16/21 (not on quiz 6, but on theory test)

*If you find that you need to Google to find answers, please contact me (**jliberle@fiu.edu**) by email at any time. Yes, 11 PM on Saturday night is fine, although I may not answer until the next morning…*

1. What type of information is found in the PFAM database?
2. How does PFAM differentiate between a domain and a family?
3. What does the HMM abbreviation mean?
4. What are HMMs used for?
5. What are domain architectures?
6. What are PFAM envelope coordinates?
7. What are PFAM HMM coordinates and how do the show a partial domain?
8. What is a PFAM seed alignment?
9. What is a PFAM full alignment?
10. How is a PFAM HMM made?
11. What is KEGG?
12. Why is homology modeling helpful for studying protein structure?
13. What assumption do we make when we use homology modeling to study the structural features of a protein?
14. What is needed to make a good homology model?
15. Briefly describe how homology modeling works
16. Why is the sequence alignment between the target and template important for the quality of a homology model?
17. What does the RMSD tell us?
18. What is QMEAN, how can it help identify regions of poor quality in a structure or model? What is the difference between Local and Global Quality Estimation?
19. What is a statistical potential?
20. What is Psipred?

(For a prediction and experimental data, know how to define and count:)

1. What is a True Positive?
2. What is a False Positive?
3. What is a True Negative?
4. What is a False Negative?
5. Know how to calculate % accuracy
6. Know how to calculate and how to interpret specificity
7. Know how to calculate and how to interpret sensitivity
8. What is bioinformatics?
9. **When a blast search is performed, the results are by default sorted by E-value. What does the E-value reflect?  What does the E-value depend upon?**