

# Practical Bioinformatics for Biologists (BIOS441/641) Study Guide for the Final Exam

December 9th, 11am - 12:30am, MO444

**15 questions will be selected from below (total 25 points)**

Overview.pptx

1. One sentence to describe what does bioinformatics do (slide 20)

Molecular-biology.pptx

2. What is the three letter code for amino acid letter D, Y, W, P, Q, K, R (slide 6)
3. What are the major protein databases? (slide 20)
4. Describe FASTA format (slide 27)

Ncbi-entrez.pptx

5. If you see an NCBI accession numbers start with NM\_, what does it mean? What about XP\_, NC\_, NW\_, NZ\_? (slide 15)

Ncbi-blast.pptx

6. If you want to find more distant homologs in BLAST search, what scoring matrix you want to use? (slide 8)
7. Why BLAST hits with higher scores tend to have lower E-value? Why E-value is dependent on the database and query size? (slide 9)
8. What is the difference between blastn and tblastn (slide 13)
9. What member protein database does nr include? (slide 16)

Ncbi-geo.pptx

10. If you see NCBI accession numbers GPLxxx, what does it mean? What about GSM, GSE, GDS? (slide 8, 13)
11. What's the difference between GEO dataset and series? Is it correct if I say all datasets are built from series? (slide 7, 13)
12. What can GEO2R do for you? Is it for GSE or GDS? (slide 34)

Ebi-go.pptx

13. What is the EBI database that is equivalent to NCBI GEO? What about nr? (slide 7)
14. What are the two member databases of UniProt? How are they different? (slide 15)
15. Describe what was the GO project designed for? (slide 25)
16. Is GO database a sequence database? Why? (slide 26)

#### Ebi-interpro.pptx

17. After a genome is sequenced, what are the immediate next steps? (slide 4)
18. Briefly describe SCOP's protein classification scheme (slide 10)
19. Can a protein belong to multiple families and why (slide 14)

#### EBI-tools.pptx

20. What is the advantage of tfastx over tblastn? (slide 9)
21. Describe the difference between global and local alignment. (slide 16)
22. What is the most accurate tool for pair-wise local alignment, for global alignment? What is the equivalent tool in the BLAST package? (slide 15, 20)
23. What is the program of the EMBOSS package that allows you to do DNA translation?

#### Phylogeny-1.pptx

24. Describe the applications of a phylogenetic tree? (slide 5)
25. What are the basic steps for phylogeny reconstruction and recommended tools? (slide 9)
26. Draw a graph to explain the concept of ortholog and paralog (slide 13)
27. Give the following newick format: ((A,B),(C,D),E); draw a graph to show the phylogeny (slide 37)

#### Linux-cmd1.pptx

28. Explain the following commands: (slide 16)  
ls  
ls /disk4
29. Define working directory and path (slide 21)
30. Can you name a file project 1.doc in Linux and why? (slide 22)
31. What is the special use of Tab key in using Ubuntu terminal? What about up arrow? (slide 24)
32. Explain the following commands (slide 26):  
cd ~  
cd ..
33. What is the command to list all files including hidden ones under your current folder in long listing format and sort it based on modification time? (slide 28)
34. Suppose you are at /home/, and you want to copy a folder there called bioinfo to your home, what is the command? (slide 33)
35. How to delete an un-empty folder called bioinfo (slide 34)

## Linux-cmd2.pptx

36. If you are at your home and you want to list things in the root folder and save the output as a file called list in your home, what is the command? (slide 10)
37. How do you switch between command mode and edit mode in vi? (slide 11)
38. Explain the following command used in vi: (slide 12)  
:1,\$s/-/negative/g
39. Briefly compare wget and lftp in terms of how they work differently. (slide 25)
40. How do you check the space use of your home? (slide 30)

## Linux-cmd3.pptx

41. Why is the '|' key called pipe in Unix command-line programming? (slide 3)
42. Explain the following command: (slide 5, 6)  
less file | cut -f1,3,6 > file2  
less file | sort -k 3,3nr > file3  
less file | grep -v '>' | head  
less file | sed '/^\$/d' | head  
less file | sed -n '101,200p'  
less file | awk -F "\t" '\$5=="22"'  
less file | awk '{print \$3,\$1}'