

Name:

## Bioinformatics Take Home Exam #2 Due 9/30/16

All questions worth 1pt

1. Assuming equal frequency of the different building blocks, two random protein sequences are on average \_\_\_\_\_ and nucleotide sequences are on average \_\_\_\_\_?  
A. 25% identical and 5% identical.  
B. 5% identical and 25% identical.  
C. 25% identical and 40% identical.  
D. 95% identical and 75% identical.  
E. None of the above.
2. \_\_\_\_\_ sequences reach saturation before \_\_\_\_\_ sequences reach saturation, so \_\_\_\_\_ sequences can be used to look further back in time.  
A. Nucleotide, protein, nucleotide  
B. Protein, nucleotide, nucleotide  
C. Nucleotide, protein, protein  
D. Protein, nucleotide, protein  
E. None of the above
3. The universe and the earth are approximately how old, respectively?  
A. 20 billion years old and 500 million years old  
B. 14 billion years old and 4.5 billion years old  
C. 16 billion years old and 6 billion years old  
D. 2 billion years old and 450 million years old  
E. 1 million years old and 4500 thousand years old
4. **True/False** Life could have inhabited the Earth prior to 3.5 billion years ago and LUCA sometime before 3.5 billion years ago.
5. Who drew the first phylogenetic trees?  
A. Lamarck  
B. Darwin  
C. Mayr  
D. Henning  
E. Peter Simon Pallas
6. **True/False** BLINK, from NCBI, stands for Boolean-link and links to pre-computed Boolean searches.
7. **True/False**- Social Darwinism is correct in that charity has stopped evolution in our species.
8. **True/False**- When inteins first begin to decay they lose the DNA-binding domain first, while the protein-binding domain must stay functional or it will destroy the function of the host proteins.
9. **True/False**- The finding that the ribosomal protein alone is responsible for the catalysis of translation is an argument against the RNA world hypothesis.
10. **True/False** Among Site Rate Variation (ASRV) means that some sites will undergo multiple substitutions while other sites do not undergo any substitutions. Due to ASRV, protein and nucleotide sequences take longer to become saturated with substitutions than without ASRV.
11. What Boolean operators can be used in NCBI/Entrez searches?

12. Match the terms on the left with the definitions on the right- 6 pts

mRNA	1. The process of making a protein from an RNA template
tRNA	2. A molecular parasite that splices itself out at the RNA level
rRNA	3. A molecular parasite that splices itself out at the DNA level
transcription	4. The process of making RNA from DNA
replication	5. RNA that binds an A.A. & matches it with mRNA triplet
translation	6. A molecular parasite that splices itself out at the protein level
intein	7. An RNA copy of a gene, used in the process of making proteins
intron	8. Part of a host gene's transcript left after RNA parasite is spliced out
exon	9. The host protein, which is spliced back together
extein	10. RNA that makes up the ribosome and catalyzes protein synthesis
Does not exist	11. Process of creating a new DNA molecule, from DNA stran

13. Inteins are composed of which of the following domains? Choose 2.

- A. Self-splicing domain
- B. Walker motif
- C. Nucleotide binding domain (GRASP)
- D. Hydrolase domain
- E. Helix-turn-Helix DNA binding domain
- F. Homing endonuclease domain

14. Which of the following are databases available through the NCBI aka Entrez? Circle all that apply-

- A. [BioProject \(formerly Genome Project\)](#)
- B. [Bookshelf](#)
- C. [Database of Genome Survey Sequences \(dbGSS\)](#)
- D. [GenBank](#)
- E. [Genome Reference Consortium \(GRC\)](#)
- F. [NCBI Help Manual](#)
- G. [Nucleotide Database](#)
- H. [Protein Database](#)
- I. [PubMed Central \(PMC\)](#)
- J. [Taxonomy](#)
- K. All of the above and many many more.

15. Sequences that do not show significant similarity-

- A. are not homologous
- B. might never-the-less be homologous
- C. are homologous

16. If the following searches were conducted in PubMed for articles, what would the searches return? Please draw Venn diagrams to illustrate your answers (i.e. depict each of the individual searches as a circle). 2pts.

- A. Gogarten J NOT Gogarten JP

B. Gogarten JP AND Doolittle WF

C. Gogarten J OR ATPsynthase

D. Inteins AND (Gogarten JP OR Swithers K)

17) If multiple search terms are connected by Boolean operators without parenthesis, the NCB interface will start evaluation

- A) from the left
- B) from the right
- C) Using a priority of AND over OR

18. What does the abbreviation NCBI stand for?

19. What is the Black queen hypothesis?

- A. Life is like an arms race, where all lifeforms have to run faster and faster just to stay in place.
- B. Gaia favors cooperation, which is the driving force behind multi-species communities being interdependent.
- C. Leaky goods can be taken up by all members in a community, so selection for streamlined genomes will result in all members of a community producing only a subset of the required leaky goods.
- D. DNA based organisms took over from the RNA world, after DNA was created by a virus in an act of genome warfare.
- E. None of the above.

20. **True/False-** The Modern Synthesis does not give any weight to the effects of mutations themselves.

21. How might mutual aid be selected for?

- A. Trick question: it cannot be selected for, because even if a stingy species is going extinct, it cannot decide to stop being stingy.
- B. When cooperation results in more offspring for both, the entire community thrives and spreads.
- C. When cooperation is between close relatives, like siblings, helping each other survive leads to some of your own genes being passed on.
- D. When one bird helps defend another bird's nest, that second bird will remember and return the favor when the first bird is under attack.
- E. B-D.

22. Why did Darwin consider the term "Coral of Life"?

- A) Because an herbarium specimen of a red algae that was wrongly labeled as a coral looked very similar to a phylogenetic tree.
- B) Because he recognized that fusion of lineages is an important process in evolution of species, and fan corals often have strands that fuse, in contrast, tree branches only branch and never fuse.
- C) Because a tree has living cells in the root, stem and leaves, whereas a phylogenetic tree has living representatives only at the tips.

**Extra credit question2**

1. Two random nucleotide sequences with equal frequencies of A, G, T, and Cs without alignment have an average percent identity of 25%. How would the average percent identity change, if the frequencies for the nucleotides are not equal. Use composition with 20%G 20%C and 30%A,30%T as an example.

2. Which of the following features of life is inescapable and will surely be found in all alien life discovered?

A. DNA

B. RNA

C. The central dogma (DNA →RNA →Proteins)

D. Ester linked lipid bilayers

E. Parasites

F. All of the above