

Name:

Are you a graduate or undergraduate student? Please circle one.

## Bioinformatics Take Home Test #2

(This is an open book exam based on the honors system -- you can use notes, lecture notes, online manuals, and text books.

*Teamwork is not allowed on the exams*, write down your own answers, do not cut and paste from webpages.

If your answer uses a citation, give the source of the quoted text.)

Notes on Formatting Quizzes: Please make sure each answer is only on one page, by using page breaks. Splitting an answer onto two pages tend to lead to grading errors.

Please do not write or type in font smaller than 12 point or write in cursive.

If you submit your quiz via email, please remove the instructions and extras (blank lines, alternative answers for multiple choice questions) from your document, so that only your answers, a minimal amount white space, and optionally the questions, are left.

1. **True/False** Assuming equal frequency of the different building blocks, two random protein sequences are on average 5% identical and nucleotide sequences are on average 25% identical. **1pt**
2. **True/False** Protein sequences reach saturation before nucleotide sequences reach saturation, so nucleotide sequences can be used to look further back in time.
3. **True/False** The universe is (approximately) 13 billion years old and the earth is approximately 6 billion years old. **1 pt**
4. **True/False** Life certainly has inhabited the Earth since no more than 3.5 billion years ago and LUCA lived sometime after that. **1 pt**
5. **True/False** Life could have arisen long before the late heavy bombardment and that the only two lineages survived it, giving rise to the Bacteria and the Archaea/Eukaryotes, and is responsible for the rapid radiation we see in the Bacteria, but this is (unproven). **1 pt**
6. **True/False** BLINK, from NCBI, stands for Boolean-link and is a link to a GUI (graphics user interface) that helps you point and click your way to more advanced searches. **1 pt**
7. **True/False** The late heavy bombardment is the impact that created the moon. **1pt**
8. **True/False**- Inteins are molecular parasites that splice themselves out at the protein level. **1pt**
9. **True/False**- Entrez is so effective because it is highly cross-linked and uses pre-computed searches. **1 pt**
10. **True/False**- Entrez is not effective because it tries to do too much and covers too many databases, making it run slowly. **1 pt**
11. **True/False**- Nucleotide derived co-factors are very common in protein catalyzed reactions, supporting the RNA world hypothesis. **1 pt**
12. **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. **1pt**
13. **True/False**- The finding that the ribosomal RNA alone can perform translation is an argument in favor of the RNA world hypothesis **1 pt**

14. **True/False-** Having the same function is proof of homology and without a shared function, homology is firmly disproven.
15. **True/False** Among Site Rate Variation (ASRV) means that some sites will undergo multiple substitutions while other sites do not undergo any substitutions. This means that due to ASRV protein and nucleotide sequences take longer to become saturated with substitutions than without ASRV. **1 pt**

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

- |   |  |
|---|--|
| a. mRNA                                     | 1. RNA that makes up the ribosome and catalyzes protein synthesis    |
| b. tRNA                                     | 2. The process of making RNA from DNA                                |
| c. rRNA                                     | 3. A molecular parasite that splices itself out at the DNA level     |
| d. transcription                            | 4. The process of making a protein from an RNA template              |
| e. replication                              | 5. A molecular parasite that splices itself out at the protein level |
| f. translation                              | 6. A molecular parasite that splices itself out at the RNA level     |
| g. intein                                   | 7. The host protein, which is spliced back together                  |
| h. intron                                   | 8. Process of creating a new DNA molecule, from DNA strand           |
| i. exon                                     | 9. An RNA copy of a gene, used in the process of making proteins     |
| j. extein                                   | 10. Host DNA left after molecular parasite splices itself out        |
| k. Non-existent thing put in to trip you up | 11. RNA that binds an A.A. & matches it with mRNA triplet            |
| l. Another non-existent thing               | 12. Part of host gene left after RNA parasite is spliced out         |

17. Inteins are composed of which of the following domains? Choose 2. **1pts**

- Self-splicing domain
- Nucleotide binding domain (GRASP)
- Hydrolase domain
- Homing endonuclease domain
- $\beta$ -barrel
- Walker motif

18. What Boolean operations can be used in NCBI/Entrez searches? **1pt**

19. What are the functions of the two domains that are present in full inteins? **1pts**
- Creates a channel through the lipid bilayer, to allow molecules to pass through
  - Cleave uninfected DNA, so that the molecular parasite will spread
  - Cleaves carbohydrates
  - Binds ATP
  - Binds a nucleotide
  - Splices itself out of the host protein, putting the host protein back together
20. Which of the following are databases available through the NCBI aka Entrez? Circle all that apply- **1pt**
- [BioProject \(formerly Genome Project\)](#)
  - [Bookshelf](#)
  - [Database of Genome Survey Sequences \(dbGSS\)](#)
  - [GenBank](#)
  - [Genome Reference Consortium \(GRC\)](#)
  - [NCBI Help Manual](#)
  - [Nucleotide Database](#)
  - [Protein Database](#)
  - [PubMed Central \(PMC\)](#)
  - [Taxonomy](#)
  - All of the above and many many more.
21. Sequences that do not show significant similarity- **1pt**
- are homologous
  - are not homologous
  - might never-the-less be homologous
22. If the following searches were conducted in PubMed for articles, what would the searches return? Please draw ven diagrams to illustrate your answers (i.e. depict each of the individual searches as a circle). **2pts.**
- Gogarten J NOT Gogarten JP
  - Gogarten JP AND Doolittle WF
  - Gogarten J OR ATPsynthase
  - (Gogarten JP OR Swithers K) AND Inteins

23. What does the abbreviation NCBI stand for and why is this site important in the field of bioinformatics? Limit your answer to 30 words or less. **1pt**

24. What is the definition of homology? **1pt**

- How similar two sequences are and degrades over time with sequence evolution
- Similarity due to shared ancestry, i.e. both got it from a common ancestor
- Similarity due to convergent evolution
- The percent of time the A.A. residues of two sequences are similar
- The common ancestor of all cellular life
- Has the same function.

25. Find a primary research article that uses “top scoring BLAST hits” to detect for “Horizontal Gene Transfer.” Record the citation here. **1pt**

Graduate questions- Short essays. Grad students should add more space to record answers.

26. **3pts** Look up a paper that argues that viruses are older and have been around longer than cellular life.

- Cite the paper you found and copy the abstract:
- Does this article take the opinion that viruses or virus-like life-forms are the precursors to cells or that they were parasitizing early pre-cellular life? Circle one or fill in whatever alternative view the authors take:
- What do you think about this paper? List a key idea from this paper. Did this paper change your opinion on viruses? If so, how?

27. **2 pts** If protein space is so big, how were complex functional molecules assembled? What does the multi-dimensional protein landscape look like in terms of functionality?

**Extra credit question for all --**

1. How would the result for a nucleotide sequence change, if the frequencies for the two nucleotides are not equal. Use composition with 40%G 40%C and 10%A, 10%T as an example. **1 pt**

2. Pushing the boundaries of life **2 pt**

a. Imagine we found cells somewhere else in the universe that could make more of themselves and were able to replicate so perfectly that they were incapable of evolution. The conditions of their planet never changed, but there was no mechanism for adaptation, should conditions change. Furthermore, they use a ready-made energy source made from the environment (an ATP equivalent) and have no metabolism of their own. Would these cells be alive? Could something like this even come into existence in the first place?

b. If we found a virus that brought its own ribosome with it into the host cell and used only its own protein manufacturing equipment, would that virus be alive?