First Midterm Exam

BioE131

October 12, 2011

Email your numbered answers to `ihh@berkeley.edu` as a **single email** with the exact string `MidtermAnswers` in the subject line (no space between the words).

The maximum length of your email message body should be about 1,500 words. This averages around 250 words per question. (Or, according to Claude Shannon, about 4,718 bits per question [Shannon, 1948].)

Answers must be received by 9am the morning after the exam is distributed.

1 Language

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Consider how you think the above text might have been constructed. Outline the design and operation of a program that would generate pieces of text that are different, but have "similar" statistical properties (in your reply, clarify how you have interpreted the word "similar").

2 Signals

Outline several (from four to six) distinct sequence signals/motifs regulating the central or structural dogmas of molecular biology. Mention the lengths of these signals, the biological roles they play, the biophysical principles involved in their recognition, and the approximate number of times you'd expect to find the motif in a million "perfectly random" characters of the appropriate (DNA, RNA or protein) sequence. "Perfectly random" here means uniform, independent, identically distributed characters. **Points awarded for biological diversity.**

3 Perl

Discuss elements of the Perl programming language, and optionally the historical influences that informed them, relating to the following areas:

1. functional programming,
2. imperative programming,
3. UNIX system programming.

Illustrate your answers with three different Perl code snippets, each of which uses one of the above programming styles to find the number of lines in a file.

4 Viruses

Describe a synthetic virus that does something cool. It can be all your own speculative invention, or a design that you read about and really liked. Include some indication of the constraints imposed by using the virus as a platform.

5 Mumps

This section relates to the mumps virus genome.

1. Answer the following questions about mumps evolution:

   (a) What proteins are encoded by the mumps genome?
   (b) Which of these proteins would you expect to evolve the fastest, and why?
   (c) How could you empirically measure the evolutionary rates, and the nature of the selection pressures, acting on mumps proteins?
   (d) What effects (if any) would recombination have on your answers?

2. For each of the following items, give the name of the most appropriate database in which to look for the item, and an accession number for one example of the item:
   (a) a 3D structure of a mumps fusion glycoprotein;
   (b) an entire mumps genome;
   (c) the sequence of a mumps nucleocapsid protein.

3. For each of the following, discuss whether you think it is a practical idea, and why. If it is practical, outline the contributions that computational biology could make:
   (a) a live attenuated mumps vaccine;
   (b) a variant of mumps engineered for gene therapy.

6 Annotation

Answer the following questions:

1. What role does protein kinase C play in signal transduction?

2. Give a database and accession number suitable for investigating whether a given protein sequence contains
   (a) the terminal domain of protein kinase C;
   (b) a protein kinase C phosphorylation site.

3. Give a Perl subroutine that takes one argument (a protein sequence) and returns a list of all the positions that might be a protein kinase C phosphorylation site.

References

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423.