CSE182 lecture 2 questions

Vineet Bafna

September 26, 2006

The questions are open ended, but should help you understand lectures better. Do these questions make sense? Are they helpful in following the lecture? Constructive feedback is appreciated.

1. Locate NCBI on the web and run Blast with a query. Can you make sense of the output? What are the different variants of Blast?

2. The Blast page has a number of seaecrh parameters. What do they mean? We will return to this in subsequent lectures.

3. What is an alignment?

4. What does percent identity, and percent similarity mean in the Blast output? What are the + signs in the alignment?

5. Consider the DNA strings AGATGGCCCCATCG and CGGTCCCCCGATGG. Compute the best global alignment of the two strings assuming match score of 1, and mismatch/indel penalty of -1. Compute the best local alignment, and the best global alignment. Are the two different?

6. Compute the best global alignment of ACACAACGG and AAAACG using affinie gap costs. Use $+1$ for match, -1 for mismatch, and gap extension, -3 for gap-open. Compare your results for the case when gap open penalty is 0.

7. How much space do the alignment algorithms need? Suppose you want to find all regions of the human genome that matched to some other location on the human genome. Describe in words, and approach to computing this. Present a rough calculation on how much memory your computer needs to have to allow this computation.

8. When you are computing sequence alignments for assembling DNA, local and global alignments are not quite what you need. Instead you need to align a prefix of a string to a suffix of another string. Design an algorithm to compute optimal prefix suffix alignments between two strings.

9. What is the genetic code? How does Blast align a DNA string to a protein string?