

MIT 10.555 Bioinformatics: Principles, Methods and Applications
Spring 2003

Problem Set 1

Assigned: February 4, 2003

Due: February 18, 2003

Problem 1: Impressions of Bioinformatics

Describe in 1500 words or less your understanding of and your interest in bioinformatics. Suggest biological problems you are interested in and how bioinformatics may help solve these problems.

Problem 2: Local Alignment: The Forest and the Trees

sequence x: TAAGTACTAT

sequence y: AAGACAT

Find the optimal local alignment of the sequences above using the Smith-Waterman algorithm. Show the scoring matrix of all possible alignments that you used. What bioinformatics problems would be suited to a local alignment? (see Smith, T. F., Waterman, M. S., *J. Mol. Biol.* (1981) 147:195-197)

Problem 3: Bayes Theorem: Two Polymerases and a Genome

3a. State Bayes Theorem. Does it make sense? Why is this theorem useful?

3b. Give a relevant example of Bayesian probability, appropriately describing each term of the theorem. Does it make sense in relation to your example?

3c. The error rate of some common DNA synthesis polymerases, such as human polymerase III, is 1 in every 100,000 nucleotides. However when they err, DNA error correcting polymerases, such as human polymerase I, correct the error with an error rate of 1 in every 100 nucleotides when an error is present. However, human polymerase I is prone to correct a nucleotide despite an absence of error, doing so 1 in every 10,000 instances. Given that polymerase I corrected a nucleotide, what is the probability that this nucleotide was synthesized incorrectly? If there are 6 billion nucleotides (3 billion base pairs) in the human genome how many do you expect to fall into this probability category? What does this say about the human DNA replication machinery?