C246 Homework Assignment 4
Due February 18, 2002 5 PM by email to homework@c246.lbl.gov

1. (10 points) Why do we need to assess the significance of alignments?

2. (15 points) Why are heuristic search algorithms acceptable for database searching?

3. (15 points) How would you compute lambda and u?

4. (10 points) Using the amino acid frequency table found at http://www.tiem.utk.edu/~gross/bioed/webmodules/aminoacid.htm and the Sij's from BLOSUM 62, compute lambda and u given k=0.13 (correct to at least one significant digit).

5. (15 points) You perform a database search in November 1986 against SWISS-PROT release 3 (just SWISS-PROT not TrEMBL) which had 969,641 amino acids in 4,160 sequence entries. You find a matching sequence using BLAST with a reported E-value of $3.6*10^{-7}$. Using FASTA, the same sequence matches with a reported E-value of $8.1*10^{-8}$.What would the E-values of alignments of these two sequences using the same versions of BLAST and FASTA be if the search were performed in June 1992, March 1995, and November 2001?

6. (10 points) Imagine you performed a database search with the sequence ANICETESTPEPTIDE using BLAST and BLOSUM62 and gap penalties of 11 for existence and 1 for extension and parameters gapped-K=0.0410, gapped-lambda=0.267, effective database length=277,038,849. You received the following alignment with a score of 23.5 bits:

```
NICETESTPEPTID
NICENDNLPKPVLD
```

Ignoring penalties and other second order effects, what would the e-value be?

7. (10 points) Describe how you could make a BLAST search more sensitive (at the expense of being slower) by modifying BLAST's parameters. (HINT: look at the command-line options for BLAST).

8. (15 points) Modify your dynamic programming program from assignment 2 to read in a FASTA file and perform a global alignment of the first two sequences found in the FASTA file

Extra Credit: (10 points) What's wrong with question 5?