

## Bioinformatics: Life Sciences on Your Computer

### Week 2 Practice Problems - Part 3

*These are practice problems for PSI-BLAST, and a related algorithm Delta-BLAST.*

1. Use the human RAD51 protein (accession: CAG38796) as a query in a PSI-BLAST search of the RefSeq database, limited to enterobacteria sequences. Change the PSI-BLAST threshold to 0.0001.
  - a. How many hits can definitively be called homologous?
  - b. What is the E value for the below threshold sequence WP\_014543474.1?
  - c. Run the second iteration of PSI-BLAST. Include all checked proteins with E values less than 0.0001. How many new sequences are added (answer “very few,” “some” or “a lot”)?
  - d. What enzyme name appears in most descriptions?
  - e. What is the second iteration E value for WP\_014543474.1?
  - f. Do a third iteration. Now how many new hits?
  - g. Can you conclude that all PSI-BLAST hits are homologs? Why or why not?
2. Run a Delta-BLAST with the silkworm insulin protein (P26726). Limit to human proteins in the RefSeq database.
  - a. How many human homologs appear to have the insulin domain?
  - b. Run a second iteration. How many new sequences are added?
3. Run PSI-BLAST with the following sequence. Use the RefSeq database, limit to bacteria and change the PSI-BLAST threshold to 0.0001.

```
> PFC_10595 [Pyrococcus furiosus COM1]
MDLERALNAFHSMKVNQLTPPCSQMPIVEEDSPIIDVLKLLRTRHHVWVVNNRKMMLVGIIRYLDIIDI
FLPPKGARLGAVSSVFKSILSGAEKAGDIMERNFLTINEDATVLEALEKMKRYKIQILALIDEEGKLGGE
ISLRILINEFLRLIRVGGG
```

- a. Record the approximate number of new hits in the first and second iteration.
- b. Are there still sequences being added on the fifth iteration?
- c. Are most of those descriptions “hypothetical?”

### **Answers**

- 1.a. 5
  - b. 2e-04
  - c. a lot
  - d. recombinase
  - e. 4e-75 (answer can vary with PSI-BLAST)
  - f. some (~30)
  - g. No, PSI\_BLAST finds putative homologs and results should be investigated. However, the fact that most from iteration 2 are recombinases suggest strongly that those new hits are indeed related.
- 
- 2.a. 6
  - b. none

3. a. first: 4; second: it maxed out the 500 shown;  
b. no  
c. yes