

Bioinformatics: Life Sciences on Your Computer

Week 2 Peer Assignment

1. Find the protein sequence with gi number 165893770.
 - a. From what organism is this sequence?
 - b. Perform a BLAST search against the nr protein database, with no organism limits. List the total score and E-value of the match from the PDB database (3OJ7_A).
 - c. Redo the same BLAST search as in part B, but this time, limit your search to the pdb database and limit your output to archamoebae (taxid:555406). What is the total score and E-value of 3OJ7_A in this BLAST search?
 - d. Explain any discrepancy in E-value/score between the two BLAST searches (for 3OJ7_A).
2. Align the following protein sequences using BLAST pairwise alignment (default values):
>gi|45549037|ref|NP_476761.2| abnormal wing discs CG2210-PA [Drosoph...]
>gi|35068|emb|CAA35621.1| Nm23 protein [Homo sapiens]
 - a. What is the score? The Expect (E) value? Are these two proteins homologous?
 - b. What is the % identity? The % positives? Briefly explain the difference.
 - c. Go to the human protein record and find the reference by Rosengard et al. What does awd (the name of the fly protein) stand for?
 - d. On the surface, it might seem surprising that two such proteins would align. According to the abstract, describe how tumor metastases could be related to abnormal wing disc development in flies?

3. Run BLAST and PSI-BLAST using the sequence below.

```
>gi|169791755|pdb|2VL0|B Chain B, X-Ray Structure Of A Pentameric  
Ligand Gated Ion Channel From Erwinia Chrysanthemi (Elic)  
APADNAADARPVDVSVSIFINKIYGVNTLEQTYKVDGYIVAQWTGKPRKTPGDKPLIVENTQIRWINNG  
LWVPALEFINVVGSPDTGNKRLMLFPDGRVIYNARFLGSFSNDMDFRLFPFDRQQFVLELEPFSYNNQQL  
RFSDIQVYTENIDNEEIDEWIRKASTHISDIRYDHLSSVQPNQNEFSRITVRIDAVRNPSYYLWSFILP  
LGLIIAASWSVFWLESFSERLQTSFTLMLTVVAYAFYTSNILPRLPYTTVIDQMIIAGYGSIFAAILLII  
FAHHRQANGVEDLLIQRCLAFPLGFLAIGCVLVIRGITL
```

Search for homologs in Archaea in the nr database. Report the BLAST results and the new hits from PSI-BLAST iteration 2 (accession numbers are fine).