Series 5 - BLAST exercise solutions

Genomics and bioinformatics - Week 5

October 18, 2011

WORK IN PROGRESS.....

1 BLAST

1. Do you get any matches to `fragment_007`? Which parameters did you use? Record the alignment statistics for the top hits.

   Using the Nucleotide Collection (nr/nt) Database and optimizing for "more dissimilar sequences" (discontiguous megablast)

2. Extract the sequence of the hit with the highest query coverage (this may not necessarily be the top hit) and perform another nucleotide BLAST, using the same parameters. Record the alignment statistics for the top hits.

3. What differences do you observe between the two BLAST results?

4. To which parameter could you attribute the changes in the E-values?

5. What is the default threshold for the E-value on NCBI BLAST?

6. Do you have any significant hits suggesting a possible function for `fragment_007`?

1.1 Protein BLAST

Using the python function from the previous exercise, obtain the amino acid sequences for `fragment_007.fasta` in three reading frames. Choose the appropriate reading frame and save the corresponding protein sequence in `aa_007.fasta` format.

Perform a protein BLAST with `aa_007.fasta`

1. Are any well-known protein domains found?

2. Do you get any significant hits? Record the alignment statistics for the top hits.

3. What is the possible function of the protein encoded by `fragment_007`?

4. Which species is most predominant in your BLAST output?

5. Could you have obtained the same results using another BLAST program, without having to translate the nucleotide sequence of `fragment_007`?

6. How do results from Protein BLAST compare with the results from Nucleotide BLAST?

2 Finding orthologs

Putative ortholog of *Saccharomyces cerevisiae* Pho2p in *Candida glabrata* is `CAGL0L07436g`

In a paper comparing the phosphate signal transduction in *S. cerevisiae* and *C. gabrata* (Genetics 182:471-9,2009), the authors identified the ortholog of Pho2p in *C. gabrata*.

Are the findings of the paper consistent with your observations? Yes