COMPUTING LABORATORY 1 - ONLINE BIOINFOMATIC TOOLS

There are a number of useful molecular biology online tools which one can take advantage of. There are too numerous to list here and give examples of. Consequently, I will provide two examples of these for two different situations.

In the first example, the use of BLAST – a homology search tool is used to determine the closet homologs of a query sequence. In the second example, the same query sequence is analysed in greater details and a phylogenetic tree with its nearest neighbors is constructed using a regimented approach.

1. Go to the URL http://trishul.sci.gu.edu.au/courses/ss13bmm/ and click on
   Module 2 "Molecular Biology DataBases"

2. Go to F Exercise 2.
   (i)   Copy the sequence from "Locus to the end of file (EOF) indicated by //
   (ii)   Now click on ReadSeq and paste the sequence in the Top Box. Format from Any to GenBank
   (iii)   Change the "OutPut Format Box" to Pearson/FASTA/fa and use View in browser button rather than download to file button.
   (iv)   Click Submit; A new box will appear with the sequence, which has been converted, from GenBank to FASTA. Copy the sequence which will now be used for two types of analysis, namely BLAST (GenBank database) and PHYLOGENY (Ribosomal DataBase Project, RDP DataBase)

FOR BLAST:

BLAST is the most widely used molecular biology tool. It matches your query sequence against a database and assists in finding the most homologous sequence to your query sequence. I will explain BLAST in simple terms but there is a lot that you can learn provided you are interested and want to spend time on it. If you are interested and need further help do feel free to contact me for advice.

There are 5 different BLAST tools – you will be using a nucleotide to nucleotide search ie you have a DNA sequence and you will match it against the DNA database.

1. Go back to the homepage of 3010BBS at the URL http://trishul.sci.gu.edu.au/courses/ss13bmm/
   -- Click on "Similarity Searches in Molecular DataBases" (Module 2: Bioinformatics)
   -- Now click on BLAST which will take you to the NCBI pages

2. Click on Nucleotide-nucleotide BLAST [blastn] under Nucleotide BLAST section (the first section). Note all the other flavors of BLAST.

3. Copy your sequence and hit submit. In response a new page will be shown, hit return again. Now wait and respond to any notices which may appear. If everything goes well you will be able to see your query report in around 3 to 5 minutes.

4. Peruse the report and I will take you through some of the main features of the report.

FOR PHYLOGENY:

Phylogeny is a stepwise approach to determine a "theoretical" relationship of your query sequence to other sequences. There are a few steps involved in phylogenetic analysis and each step uses a special algorithm. The algorithms have been developed using various hypothesis and you will see these as you pass through the various steps :

(i) Similarity search: one must search for the most homologus sequences against your query sequence in a database . You have used BLAST search tool and NCBIs GenBank database for this.

(ii) Multiple pairwise sequence alignment: You will then need to download some of these homogs and together with your sequence align these using bioinformatics software. Automatic alignment software can be used - Clustal is widely used for this purpose. A manual alignment editor is then used to improve the alignments. In case of 16S rRNA, secondary constraints are used to guide alignments. So, it is not a straightforward task as is the case with protein sequence alignments.

(iii) The next task is to ensure that only "like is compared with like" that is any columns containing alignment gaps are removed from further analysis – a "masking" process is used for this purpose.

(iv) The next task is to determine the similarities between the sequences. A scoring matrix is produced from this exercise.

(v) Now the most similar sequences are clustered together using a clustering algorithm

 (vi) Finally you will need to view the product – phylogenetic tree) that you have made. You may also need to make changes to the tree.

Constructing phylogeny at http://rdp.cme.msu.edu/html/, the Ribosomal DataBase Project (RDP II):
Go to the site and choose the Phylip Interface. Phylip is the most widely used suite (collection) of programs used in phylogeny. Phylip requires pre-formatted alignments in the Phylip format. None of the programs allow you to align and / or edit sequences.

Note that the server requires that a query sequence file be stored in your computer which it can access (up loads the sequence).

You will be instructed in the use of the various simple steps.