Genetic Databases

Introduction

Online databases have become an essential tool for researchers in Biology. These databases store a huge volume of publicly accessible data that allows society to gain a deeper understanding of molecular biology. In this activity we will be looking at three different online databases.

Activity 1

3.1 S1: Use of a database to determine differences in the base sequence of a gene in two species (Allott & Mindorff, 2014)

One outcome of the Human Genome Project is that the techniques that were developed have enabled the sequencing of other genomes. This allows gene sequences to be compared. The results of this comparison can be used to determine evolutionary relationships. Also, the identification of conserved sequences allows species to be chosen for exploring the function of that sequence.

1. Go to the Gen Bank (http://www.ncbi.nlm.nih.gov/genbank/)

2. Choose 'gene' from the search menu

3. Enter the name of a gene plus the organism, such as cytochrome oxidase 1 (COX1) for pan (Chimpanzee)

4. Under the 'Genomic regions, transcripts and products' click 'go to nucleotide: FASTA'

5. Copy the sequence that appears and paste it into a text file by using the following steps on a mac:
   a. Open textedit
   b. Click "new document"
   c. Under format click "make plain text"
   d. Paste your sequence

6. You can paste multiple sequences in the same text file (note: if you delete the numbers at the top but keep the >gi| followed by the name of the organism whose sequence you have this will make the final steps easier.

7. Repeat for five species of your choice (include at least one plant and one bacteria in your search and save the file with a suitable name)

8. Download the software ClustalX 2.1 from http://www.clustal.org/download/current/

9. Add the application to your applications folder

10. Run the application

11. In the File menu choose 'load sequences'

12. Select your file. Your sequences should appear in the Clustal X window.

13. Under the alignment menu choose 'Do complete alignment'.

14. The output files can be analysed in another program to produce phylogenetic trees (DP Topic 5.4)

Activity 2

3.1 A2: Comparison of the number of genes in humans with other species

3.2 A2: Comparison of genome size in T2 phage, *Escherichia coli, Drosophila melanogaster, Homo sapiens and Paris japonica.*

3.2 A3 Comparison of diploid chromosome numbers of *Homo sapiens, Pan troglodytes, Canis familiaris, Oryza sativa, Parascaris equorum*

This practical has been adapted from Inthinking.co.uk and Allott & Mindorff 2014.

You will use the Ensembl database to complete the table below:

Species	*Number of genes*	*Genome Size/million base pairs*	*Diploid number*
Homo sapiens			46
T2 Phage	███████	0.18	███████
E. coli		5	███████
D. melanogaster		140	
P. japonica	███████	150,000	███████
P. troglodytes			48
C. familiaris			78
O. sativa			24
P. Equorum	███████████████		4

1. Search for Ensembl Genomes (http://ensemblgenomes.org/)
2. Use the tabs in the top right to Navigate between Vertebrates (*H. sapiens, P.Troglodytes, C. familiaris),* Metazoa (*D. melanogaster*), Bacteria (*E.coli*) and Plants (*O. sativa).*
3. In each tab search for the species in question under 'browse genome'.
4. When the species if selected click 'view karyotype'
5. Count the number of chromosomes in the karyotype image to find the diploid number.
6. The number of base pairs will give you the Genome size
7. The number of coding genes will give you the number of genes

Activity 3

3.2 S1: Use of databases to identify the locus of a human gene and its polypeptide product (Allott & Mindorff, 2014)

The locus of a gene is its particular position on the homologous chromosomes. Online databases can be used to find the locus of human genes. There is an example of such a database in the Online Mendelian Inheritance in Man website, maintained by John Hopkins University.

1. *Google the abbreviation OMIM to open the homepage*
2. *Choose search 'Gene Map' in the Advanced Search settings*
3. *Enter the name of a gene into the Search Gene Map box. This will bring up a table with information about the gene including its locus.*
4. *Complete the table below.*

Gene Name	Description of Gene	Chromosome locus
DRD4		
CFTR		
HBB		
F8		
TDF		

5. *Alternatively you can select a chromosome 1-22 or one of the sex chromosomes. A complete sequence of gene loci will be displayed.*