

THE CASE OF THE FLORIDA DENTIST

This practical will introduce you to some of the some of the online bioinformatic tools that we use in our lab. If you get stuck just ask for help. Feel free to explore other parts of these websites and if you finish, have a go at some of the other activities in the resources pages.

bold = things to press or do

> **CAPTIAL** = use the menu bar

→ just information

underlineditalic = has a key word explanation given on page 4

Background

This practical is based on a true story. The case involves a group of patients who claim they contracted HIV from their dentist. If this is true, then we would expect the HIV virus isolated from the dentist to be more related to the patients virus than to other 'control' sequences from the population. To test this hypothesis we are going to compare gene sequences from all these HIV strains by building a phylogenetic tree. We will use the *env* gene which codes for the outer coat of the virus. The *env* sequences for this case have been placed into the GenBank database. For more information see Palca, 1992.

Overview of practical

- Retrieve dentist, control and patient *env* gene sequences
- Arrange these sequences using Clustal into a multiple sequence alignment
- Build a phylogenetic tree of these sequences

Getting your sequences

First we need to get the *env* sequences. I have got most of them for you from the NCBI website (step 1.) but you still need to get one more in step 2.

1. **Open WordPad** and then **open** the text file called **HIV.txt** which is in the following folder: K:\SCIENCE\Dr Green\The florida dentist
Save a copy of this file to your desktop: call it HIV.txt

→ if you look at this file you can see the DNA sequences of the *env* gene from the patients A-H, dentist and the local controls are all in FASTA format. Each sequence has a > followed by the title, then the sequence on a new line. I got the sequences from the NCBI website to save time, now you are going to get the final one.

2. Go to NCBI web page <http://www.ncbi.nlm.nih.gov/>
3. Use the Search bar on the left to search **NUCLEOTIDE**
4. Now **enter** the GI number **326847** into the box.
5. Press **SEARCH**, this will take you to the GenBank page for this gene

→ have a look at the information on this page as it contains information on the organism, sequences (nucleotide and protein) and any relevant references.

6. To retrieve the nucleotide sequence for this gene **scroll up** to the top of this page and **Click** on the **FASTA** button
Select and **copy** this sequence and **paste** it into the HIV.txt file on your desktop that you created in 2.
7. **Rename** this sequence by adding the word **DENTIST** just after the > like I have done below and **SAVE** changes

```
>gi|326847|gb|M90848.1|HIVFLD1 Human immunodeficiency virus type 1, viral sample FLD1, V3 region
```

```
>DENTISTgi|326847|gb|M90848.1|HIVFLD1 Human immunodeficiency virus type 1, viral sample FLD1, V3 region
```

Aligning your sequences

Now you have all your sequences in one file (in FASTA format) we need to produce a *multiple sequence alignment* using a program called Clustal. First we will import all the sequences from your HIV.txt file.

8. Visit the website <http://www.ebi.ac.uk/Tools/msa/clustalw2/>
9. **Copy your sequences and paste into the box**

Now we are going to get Clustal to align the sequences so that each column (hopefully!) contains homologous positions (ones that have evolved from a common position in the ancestor)

10. Set sequences as DNA and press submit
11. Wait for your job to finish
12. Look at the alignment using the scroll bars along the bottom
13. How many positions are identical?

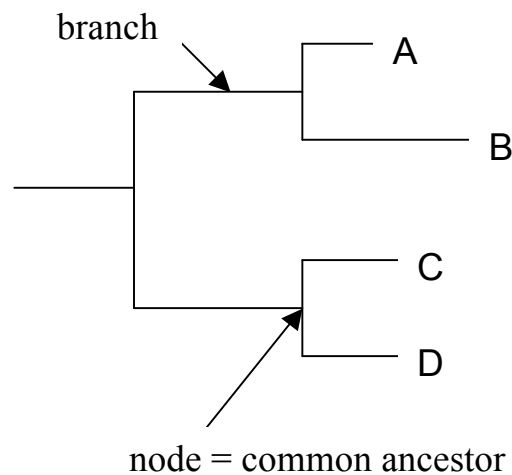
→ * = identical residues, - in the alignment is a gap

Building a tree

1. Stay on the same page as the alignment, scroll up to the top
2. Press the tab that say phylogenetic tree, scroll to the bottom to see this

→ When looking at evolutionary trees the following tips may help:

1. Every node on the tree represents a hypothetical ancestor.
2. Relatedness is to do with common ancestry e.g. C is more closely related to D than it is to either A or B. This is because C shares a common ancestor with D more recently than it does with either A or B.



Look at your tree and look at the positions of the dentist, control and patient sequences. Was the dentist responsible for giving HIV to all, some or none of his patients?