Multiple alignment and phylogeny exercise: The phylogeny of HIV

Description of the data

The protein gp120 is the crucial envelope protein of HIV facilitating binding to and fusion with the target cell (human CD4 lymphocytes).

Take a look at the HIV sequences data in fasta format: The file `gp120.fasta` contains 27 gp120 sequences (envelope protein) from HIV-1, HIV-2 and SIV. As you can see, they do not have the same lengths.

Regarding the alignment, it is of interest that gp120 contains 9 conserved disulfide bridges. Also relevant is the so-called V3-loop. This is a surface-exposed highly immunogenic antibody-binding and hypervariable (immunological escape) region of gp120, which has been extensively sequenced. The location of the V3 loop in the gp120 molecule can be seen in this schematic visualization.

Multiple alignment

1. Go to the ClustalW server at The Baylor College of Medicine Search Launcher in Houston, Texas.

2. Copy the entire contents of the `gp120.fasta` file and paste it into the sequence input window.

3. Click "Perform Search" and wait for the results to appear.

4. Look at the alignment.   Check the alignment length, and that all cysteines (18) are correctly aligned.

If you want, you can try to repeat the alignment with different settings of the options. You can read about the options in the HELP file.

Computing the phylogeny

In this part of the exercise we will use a number of programs from the PHYLIP package, implemented in a server at The Institut Pasteur, Paris.

First, we use the program `protdist` to compute a distance matrix from the alignment we made, and then we use the program `neighbor` to calculate Neighbor-joining and UPGMA trees based on this matrix.

1. Copy the aligned sequences in FASTA format from the ClustalW output page (that is the part of the page highlighted with a green background).

2. Go to the `protdist` input form and paste the sequences into the alignment window.

3. Write your own e-mail address in the top field and click "Run protdist." **Note:** all programs at this server require you to give your e-mail address; if a program takes more than a few minutes to run,

     you will be notified by e-mail when it is finished, and then you can access the results by clicking on a link.

4. From the results page, take a look at "outfile."

5. Select "neighbor" from the menu and click "Run the selected program on outfile".

6. Now you are at the `neighbor` input form with your distance matrix file already loaded. Click "Run neighbor."

7. Take a look at "outfile." Remember: this is an *unrooted* tree!

8. If your computer has a postscript viewer installed, you can get a graphical version of the tree:

    1. Select "drawtree" from the menu and click "Run the selected program on treefile"
    2. Using the "Apple Laserwriter (with Postscript)" format setting, click "Run drawtree"
    3. View the postscript file "plotfile.ps".
    4. To change the appearance of the tree, go back to the `drawtree` input form, click "Advanced drawtree form," and change some of the "Drawtree options."

9. Now, go back to the `neighbor` input form and click "Advanced neighbor form".

10. Change "Distance method" from "Neighbor-joining" to "UPGMA" before you click "Run neighbor." This will produce a *rooted* tree.

11. If your computer has a postscript viewer installed, you can get a graphical version of the tree as described above. However, to plot a *rooted* tree, you should select the program "drawgram" instead of "drawtree".

Other resources

This exercise might also be carried out using the [ClustalW server](#) or the [ClustalW and Jalview server](#) at EBI (European Bioinformatics Institute). Jalview is a java-based multiple alignment editor, which shows your alignment in cool colours and makes it possible to do manual post-processing of the alignment. However, this service is slower than the others, and Jalview itself tends to run painfully slow on some computers.

[PAUP](#) is a widely used package of programs for inferring phylogenies, and exists in versions for most computer platforms.

[TreeView](#) is a simple program for displaying phylogenies on Apple Macintosh and Windows PCs.

The [BCM Search Launcher](#) has a [multiple alignment page](#) offering several programs different from ClustalW. Note: the other algorithms tend to be slower and have stricter size limitations.

For more links concerning multiple alignments, see the [Multiple Alignment Resource Page](#) from the [GNA-VSNS Biocomputing course](#).

For more links concerning phylogenetics servers at the WWW, see [this list](#) from the PHYLIP WWW site.