

Sequence analysis tools – practical

The purpose of this practical is to give a basic understanding of common tools for similarity searching (BLAST) and multiple alignment (ClustalW), and some their configurable parameters. The task is to use BLAST to identify a protein fragment, select the full length homologs from several species, align them using ClustalW, and visualize the results using JalView. Please don't hesitate to try different settings and play around with parameters!

1. Go to the BLAST homepage.

<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>

2. Use a few minutes to get familiar with the page, note the different options that are available from the main page.

3. Choose “protein blast”, and on the following page, choose “Swissprot” from the “Database” dropdown list.

4. Open the “Algorithm parameters” section by pressing the link. Make note of the default parameters for “Word size” and “Matrix”.

5. Paste the sequence below (including the header line starting with “>”) into the query box at the top of the page.

>protein_fragment

MGQTGKKSEKGPVCWRKRKVKSEYMRLRQLKRFRRRADEVKSMFSSNRQKILERTEILNQEWKQRRIQPVHI

LTSVSSLRGTRECSVTSDDLDFPTQVIPLKTLNAVASVPIMYSWSPLQQNFMVEDETVLHNIPYMGDEVLD

6. Check the “Show results in a new window” checkbox, and then start the query by pressing the “BLAST” button. It may take some minutes to run the query.

7. The result page should be fairly self explanatory. Look at it and try to understand the different parts of the page. If you have questions, please ask one of the teachers!

8. Based on the information in the result page, what is your conclusion – what protein/species does the fragment come from? On what do you base these conclusions?

9. In another window, go back to the query page and try out different combinations of “Matrix” and “Word size” in the “Algorithm parameters” section. Do the results differ for different settings? How? Why? Do the different results cause you to change your original conclusions? Why/why not?

10. Select the top six hits from your first query by checking the boxes next to the pairwise alignments, then retrieve them using the button “Get selected sequences” above.

11. In the “Display” list box at the bottom of the page, choose FASTA format.

12. In the “Show” list box at the top of the page, choose 10 (this is to make sure that all 6 sequences end up on the same page)

13. In the “Send to” list box, choose “Text”.

14. In a new browser window, go to the EBI ClustalW homepage.

<http://www.ebi.ac.uk/Tools/clustalw2/index.html>

15. Paste the sequences (including headers) into the sequence box, press “Run”.

16. Study the result page and try to understand the basics of it. Don’t hesitate to ask a teacher if you have questions!

17. JalView is a multiple alignment visualization tool that makes it more convenient to study alignments by eye. Start it by using the “Start JalView” at the top of the result page.

18. What are your conclusions?

19. Go back to the query page and play around with different parameter settings. Each parameter has a link with a brief explanation. For example, choose different values for the “GAP OPEN” parameter. How/why do they affect the end result?