

## **LAB 2 Pairwise Alignment and BLAST**

### **1) Pairwise alignment (You need to have two nucleotide or protein sequences)**

You can search “tomato aldehyde oxidase” – a family of genes I cloned for your lab exercises. TAO1 accession number AAG22605; TAO2: AAG22606; TAO3: AAG22607

There are two types of pairwise alignment – global alignment or local alignment. Read the section on “Global and local alignments”

[https://en.wikipedia.org/wiki/Sequence\\_alignment](https://en.wikipedia.org/wiki/Sequence_alignment)

Also please visit: <https://www.ebi.ac.uk/Tools/psa/>

For a global alignment: [http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](http://www.ebi.ac.uk/Tools/psa/emboss_needle/)

For a local alignment: [https://www.ebi.ac.uk/Tools/psa/emboss\\_water/](https://www.ebi.ac.uk/Tools/psa/emboss_water/)

NCBI “blast2sequence” program is a type of local alignment – after you choose a program, then click “Align two or more sequences”

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

**Note: if two sequences have a low similarity, they may not be aligned properly when using a local alignment tool, thus, you need to use the global alignment tool.**

### **2) BLAST**

Read Chapter 16 in the NCBI handbook (can do it at home)

<http://www.ncbi.nlm.nih.gov/books/NBK21097/>

If you need any help, click help at

[https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastDocs](https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs)

**A guide:**

[http://proteomics.vsu.edu/courses/BIOL4800\\_6900/ExtraReadings/HowTo\\_BLAST\\_Guide.pdf](http://proteomics.vsu.edu/courses/BIOL4800_6900/ExtraReadings/HowTo_BLAST_Guide.pdf)

**A video:** <https://www.youtube.com/watch?v=mvjHYMgJDTQ>

**Do blast search with your protein sequence of interest -**

Find homologous protein sequences and then retrieve these sequences and their corresponding cds (coding DNA sequence) using links in the blast output and GenBank record. You may explore different blast programs and change parameters for testing purpose. For example, change e-value, limit the number of output, change scoring matrix, filter on/off, etc.

**Manually retrieve 10 – 30 homologous sequences in FASTA format (including proteins and protein-coding DNA sequences) from at least 10 different species.**

Make sure to save DNA and protein sequences in different files (text file), make sure to include species information in the definition line.