

Alignment Assignment

This assignment can also be found on line on the web site with active links by going through the syllabus or the announcements pages.

Many resources for multiple sequence alignments are now available on line. <http://cnx.org/content/m11026/latest/> has a great deal of information about BLAST as well as a tutorial. <http://pbil.univ-lyon1.fr/alignment.html> has a list of available alignment software.

Several sites actually have the tools available for doing the alignments on line. All that you have to do is upload a FASTA file, and the computer running the web site will do the alignment. These systems use different methods to align these sequences, so many will produce different alignments from the same input file. We want you to try a number of these programs on your own time, and investigate the results.

We will be using a FASTA file of partial calmodulin sequences from ten species of *Conus*. This is a really cool genus of predatory tropical reef snails with beautiful shells and complex cocktails of oligopeptide toxins. Calmodulin is involved in the release of these toxins. These sequences were put on genbank by T.F. Duda Jr., but never published.

Download the file from http://ib.berkeley.edu/courses/ib200a/ceph_sequences.fasta.

Go to the following web sites and generate alignments of the sequences. It should be clear how to do this. It may take a few seconds, when it is done download the alignment files that the program generates. In some cases you will have to copy the alignments from web pages to text files. Additional information about the alignment methods is available on the web sites.

<http://bibiserv.techfak.uni-bielefeld.de/dialign/submission.html>

<http://baboon.math.berkeley.edu/mavid/>

<http://align.genome.jp/> - this web site has three different available methods of alignment, ClustalW, MAFFT, and PRRN. You should generate a separate alignment using each of these methods.

Make the following observations for each alignment:

- 1) Visually inspect the output in Macclade, Clustal, Winclada or Mesquite and describe the “gappiness” and any observed “errors”
- 2) Record the total number of characters in the aligned matrix.
- 3) Record the number of parsimony informative and uninformative characters (use Mop Uninformative Characters in Winclada or a similar command in MacClade. PAUP* and Mesquite can also give you the same information).
- 4) Save the matrix in the appropriate format and run a parsimony analysis in Winclada/Nona, PAUP* or PHYLLIP. This matrix is small enough for you to run an exhaustive search.
- 5) Record the number of minimum length trees found, and the number of steps they have.
- 6) Make a strict consensus tree and note the resolution. Are the consensus trees from the different alignments all compatible? If not, where do they conflict with each other? (Make sure that you root the trees with the same taxon for doing this comparison.)

Now you are an unstoppable alignment maven!!