

Name:

Gene:

Task 5: Creation of edited and annotated multiple sequence alignment (MSA)

In this Milestone, you will perform a multiple sequence alignment (MSA) of the 10-12 sequences of similar genes from diverse organisms which you found in Milestone 2.

To perform the multiple sequence alignment, you will be using the CLUSTALW program available from the EBI (European Bioinformatics Institute):

<http://www.ebi.ac.uk/clustalw/>

After inspecting the multiple sequence alignment produced by ClustalW, if you find portions of your sequence that do not align well, you may want to consider removing these non-homologous regions, particularly if they are interfering with optimal alignment within the protein family. This may be useful if the sequences differ radically in length, especially at their ends.

In the process of performing multiple sequence alignments, as an intermediary step, pairwise alignments are determined. Is the ranking of pairwise alignments between *your yeast sequence* and other organism sequences consistent with what you observed in Milestone 2?

Which pairwise alignments (out of all pairwise alignments, not necessarily including your yeast gene) have the highest scores?

Is the phylogram tree (based on your multiple sequence alignment) consistent with what you expected?

In studying your alignment (the JalView option may be useful here), are there conserved regions and non-conserved regions within your protein family?

How well does your yeast gene seem to fit your protein family?

When submitting the Milestone, please include your actual multiple sequence alignment (such as the output provided by JalView). The alignment should be colour-coded to indicate structure/function relationships (i.e., identity and similarity columns). Also, please label each of your sequences with the organism's real name (e.g., Arabidopsis, fish, etc.).

As a last step, you should re-run your multiple sequence alignment, and in the "output" options, you should select "gcg MSF" as the output format. You should then save the "alignment" file (the file ending with .aln) which is generated by ClustalW. The multiple sequence alignment saved in this file will be used in the next Milestone to construct an evolutionary tree (phylogeny) using the software package PAUP (Phylogenetic Analysis Under Parsimony).

Due date: March 16th 2005

Please e-mail your completed assignments to:

Btjaden@wellesley.edu

Dwebb@wellesley.edu