

Alignments with Clustal

Emily Josephs & Nancy Chen

March 16, 2016

This is simply a brief introduction to multiple sequence alignment with ClustalW.

Typeset conventions: All code is written in **constant-width font**. Keyboard commands are in **bold**.

Input file formats

Gene sequences are stored in the data folder (/data) in fasta format. To get a better feel for fasta format, copy the two fasta files corresponding to your gene to your own directory and use tools like top, less, and vim to look at the files.

What are the two different parts of the fasta file? What comes after the “>” character?

To align two sequences, we need to put these sequences in the same file. We can do that using “cat”

- **cat dumptruck*.fasta > alldumptruck.fasta**

Now, it is time to align these sequences. Why do we want to do this again?

Using Clustal to align sequences

Clustal is a program for aligning two sequences. You can open it with the following command:

- **clustalw**

Using clustal is a bit like using a point-and-click program, except you interact with it in the command line by entering options when prompted. If at any point you get stuck, you can access Clustal’s help. We will walk through it together.

Congrats, you’ve aligned two gene sequences!!