

PAL2NAL

PAL2NAL is a web-based tool for converting protein alignments into their respective nucleotide (coding sequence; CDS) alignments. It can also calculate the ratio of the nonsynonymous-to-synonymous substitution rates (dN/dS; also known as Ka/Ks) between two coding sequences.

You can find PAL2NAL at:

<http://www.bork.embl.de/pal2nal>

Example input sequences can be found on the course website in the text file "Alignments.txt": These can be cut-and-pasted into PAL2NAL (do not try to upload the file). The input format is FASTA. The major feature of this format is that the first line begins with the ">" symbol and contains a description of the gene, which can be any text that you want. The following lines contain the sequence using standard nucleotide/amino acid codes or the gap symbol "-".

1. Use PAL2NAL to compare the protein and CDS sequences of human and rat Histone H1.
 - a) Use the default settings for options 1-4 and for option 5 choose "Codon with amino acid". This will show you a combined alignment of the protein and DNA sequences.
 - b) Repeat the above, but change option 5 to "FASTA". This will give you the aligned DNA sequences in FASTA format. In the case of Histone H1, there are no gaps, so the result should look like your input sequences.
 - c) Repeat the above, but change both parts of option 2 to "yes". For option 5 choose "CLUSTAL". This will produce an alignment of the CDS and calculate dN/dS. How many synonymous sites are there? How many nonsynonymous sites? What are the values for dS, dN, and dN/dS? What does the dN/dS value indicate about the evolution of this gene?

2. Repeat all of step 1 using the provided sequences for the *Acp26Aa* gene of *Drosophila melanogaster* (mel) and *D. sechellia* (sec). This is a male-specific "accessory protein" (Acp) that is secreted by the male and transferred to the female along with sperm during mating. Note that only the protein sequences need to be aligned in the input (that is, have gaps in the appropriate places). The CDS sequences do not need to be aligned. The CDS alignment is done by the program.

3. Use FlyBase (www.flybase.org) to download the protein and CDS sequences of the *ocnus* (*ocn*) gene from *D. melanogaster* (Dmel) and *D. sechellia* (Dsec). Then prepare them as input for PAL2NAL and repeat the above steps. Hint: it is helpful to use the "Get FastA" option next to the gene map in the FlyBase entry for *ocn* and choose "translations" or "CDS". You can use the "Orthologs -> Orthologs (via OrthoDB) -> Orthologs in Drosophila species" section of the *D. melanogaster* entry to link to sequences from other *Drosophila* species. Or you can use the "include non-Dmel species" option when searching from the FlyBase home page.