Bioinformatics, computer lab VIII
University of Potsdam WS 2019/2020, S. Hartmann

Evolutionary analysis of DNA sequences

Objectives: at the end of this computer exercise, you should be able to use the online alignment tool Clustal Omega to compute and view a sequence alignment, and to describe how molecular sequences can be used to learn about the evolution of (here: maternal) lineages.

The data

Today you will analyze two different regions in the mitochondrial genome of five extant/modern humans, one Neandertal, and a chimpanzee. The sequences from modern humans come from different geographic regions; this information is given in the sequence IDs:

J1 Russia	R1 Canada	C5 Caucasian
H2 Europe caucasian	L1 sub-Saharan Africa	

The two different mitochondrial regions are stored in the files '*COX_sequences.fa*' and '*HVR_sequences.fa*', and both are available for download from Moodle. 'COX' stands for cytochrome c oxidase subunit I, and this file contains a sequence fragment of 1562 bases. 'HVR' stands for hypervariable region I, this fragment is 385 bases in length.

1 COX: A multiple sequence alignment

The Clustal program can be used again to compute today's alignments, but you can do this either online (using ClustalO) or in terminal (using ClustalW). Choose one of these options:

☞ EITHER compute the COX alignment as you did two weeks ago, using the clustalw program in the terminal.

☞ OR direct your browser to `http://www.ebi.ac.uk/Tools/msa/clustalo/`.

- Paste the sequences from the COX file into the form at the bottom of the page. Select that you have DNA data. Click on the "Submit" button.
- When the alignment has been computed, a page with the results will be displayed. You can briefly view them here (with or without the display adjusted to show colors).
- Click on the tab "Download Alignment File" and save the alignment.

☞ Regardless of whether the alignment was computed online or on your local machine, use the Jalview alignment viewer to view the alignment, just like you did two weeks ago. Adjust the colors for viewing nucleotides.

- What kinds of changes (mutations) can you observe in this alignment? List two or three specific examples.
- In approximately how many of the alignment columns can you observe changes? (20%? 50%? 80%?)
- Among which sequences (or groups of sequences) did the most changes occur? The fewest changes? Can you guess (or 'predict') which of the sequences will be related closely or distantly to each other? Draw a tree diagram that supports your hypothesis.

2 COX: A phylogenetic tree

☞ You can get a sense for how similar or different these sequences are, but we need a computer to accurately calculate their relationships. From within the Jalview window, click on the menu "Calculate", then select "Calculate Tree or PCA". Note that this only works if no sequence of the alignment is selected. For computing a Neighbor Joining tree you can choose PID (percent identity) or DNA (the substitution matrix shown here is used: `http://tinyurl.com/y78q3fhf`). Recall the main steps of the neighbor-joining algorithm as discussed last week. For which step is the computation of identities or substitution scores required?

☞ Select the Neighbor Joining option that you find more appropriate. A new window will open that displays the relationships of the sequences.

☞ Does this tree agree with your impression from the sequence alignment?

3 HVR: Alignment and phylogeny

☞ Repeat the analysis for the other sequence fragment: compute a multiple sequence alignment and a neighbor joining phylogeny from the HVR region. Answer the same questions that were asked for the COX region above.

4 Summary

☞ Which of the two fragments, COX or HVR, is more suitable for studying the evolution of modern humans?

During this exercise, you have only analyzed a small part of the mitochondrial genome of humans, but this type of analysis is exactly what is frequently done in evolutionary bioinformatics. In fact, the HVR fragment was chosen by the Genographic project mentioned in class. This DNA region has been sequenced from thousands of people from around the world to find out about human migration patterns.