

09 | Bioinformatics: family tree



Equipment list

- Computer for:
 - aligning multiple sequences
 - creating phylogenetic trees
- Instructions for multiple sequence alignment and phylogenetic tree creation and display

Instructions | Multiple sequence alignments and phylogenetic trees

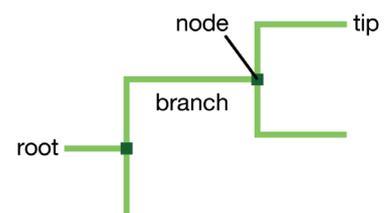
A **phylogenetic tree** is a diagram that represents evolutionary relationships among organisms.

Comparison of complete genomes gives the greatest information, but takes a lot of time and computing power. So the shorter DNA barcode sequence can be used to look at the relatedness of organisms more rapidly and using less computing power.

The relatedness of different organisms is inferred from the similarity of the DNA barcode sequence. The more similar the DNA barcode sequence, the more closely related the organisms are, and the more likely they are to have evolved from a recent common ancestor. The less similar the DNA barcode sequence, the less closely related the organisms are, and the longer ago they are likely to have evolved from a common ancestor.

These evolutionary relationships can be represented on a diagram that shows:

- a common ancestor for the whole tree, the **root**
- the relative time since which species diverged, **branch length**
- where speciation occurred from a common ancestor, a **node**
- the sequences from different species used to construct the phylogenetic tree, the **tips**



In this task you will align multiple DNA sequences, use the number of differences to make a phylogenetic tree, then consider different ways of presenting phylogenetic trees.

From this point onwards if you would prefer to use online instructions follow this link:

app.tango.us/app/workflow/09-Multiple-sequence-alignments-and-phylogenetic-trees-2b2d2a4a33314372be730b878e090abb

□ 1. Find the European Bioinformatics Institute (EBI) Clustal Omega website, with free software for aligning multiple DNA sequences

- Type **EBI Clustal Omega** into an internet search engine.
- Click on the link for **Clustal Omega < EMBL-EBI**.
- Select **Input sequence**, **Sequence type** to be **DNA**.

□ 2. Align the barcode sequences from the invertebrates

- This example uses the text file **09_R_Known-barcodes**, but you should create a phylogenetic tree of the invertebrates identified by your group.
- To use your invertebrate DNA sequences, create a text file, with all DNA sequences in FASTA format.

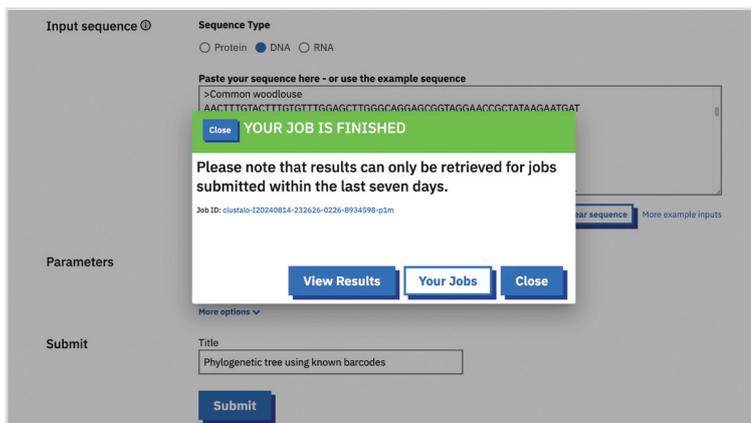
*Remember this is >name on the first line, then the DNA sequence on the next line. See the text file **09_R_Known-barcodes** if you are not sure what this should look like.*

- Copy all names and sequences from the text file with all DNA sequences in FASTA format, and paste them into the **Paste your sequence here** box.
- In the **Parameters, Output format**, choose **ClustalW with character counts**.
- Type a title into the box.
For example, 'Phylogenetic tree using known barcodes'

- Click the blue **Submit** button. The software will use a machine learning approach to align the 5 sequences. Depending on how many searches are submitted at the same time as yours this may take a few minutes.

□ 3. Understanding the results

- When the pop up says that your job is finished, click on the blue **View results** button.

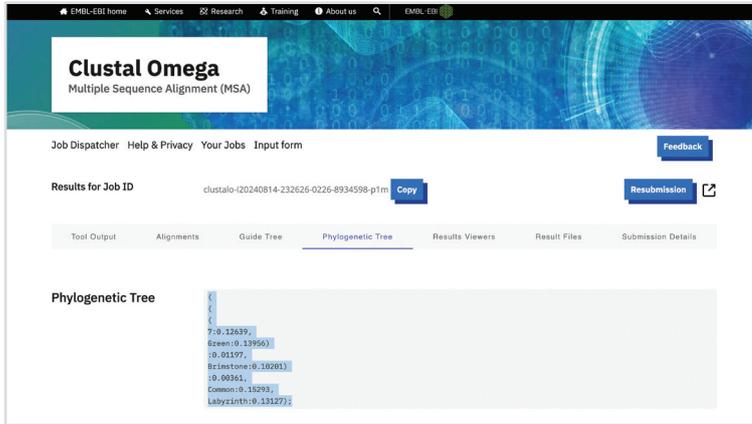


- Under the **Tool output** tab there is an alignment of the sequences. Sequence names are shown on the left hand side. The numbers on the right hand side give the number of base pairs (60 per line). Dashes indicate gaps in the sequence. Asterisks (*) underneath the sequence alignment indicates the same base is present in all of the sequences.

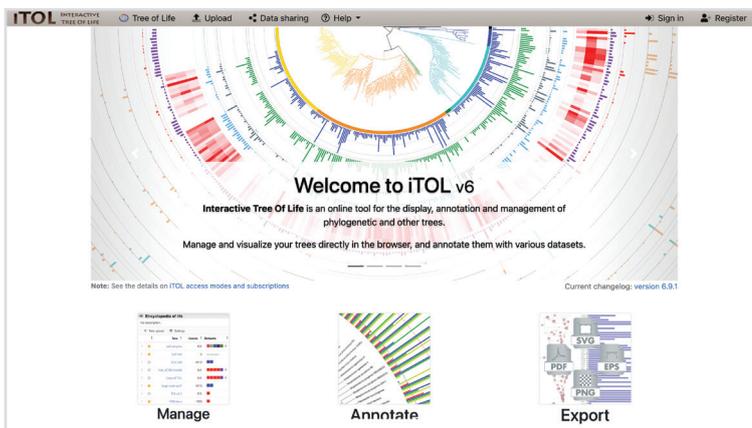
- Scroll to the bottom of the sequence alignment and click on the **Alignment with colours, Show** button.
- To present the results of your sequence alignment of the unknown invertebrate barcodes, scroll back to the top and click on the **Tool output, Download** button. This will produce an **.aln-clustal_num** file. You should open this using Microsoft Word and save it as a plain text document.
- Make sure that you save the alignment in a location and use a file name that you will be able to find again. If you select all text and change the font to **Courier New size 9**, then you will be able to view the alignments arranged on individual lines.

4. Displaying information as a phylogenetic tree using the interactive Tree of Life website

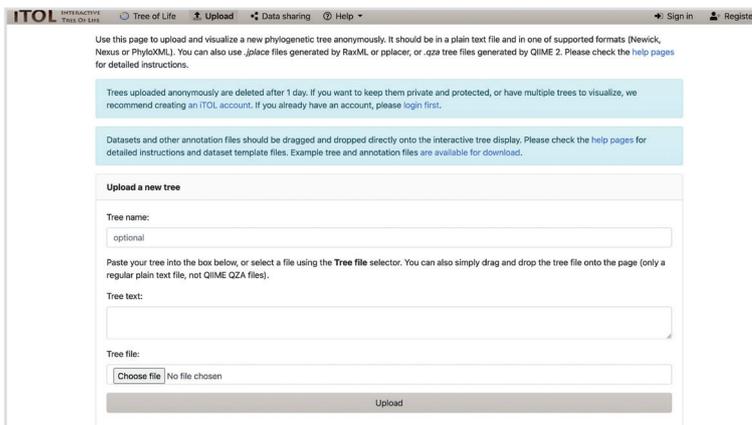
- Click on the **Phylogenetic tree** tab for your results.



- Copy the **Tree data** (all text from the phylogenetic tree box).
- Type **EMBL Interactive Tree of Life** into an internet search engine.
- Click on the link for **iTOL: Interactive Tree Of Life**.



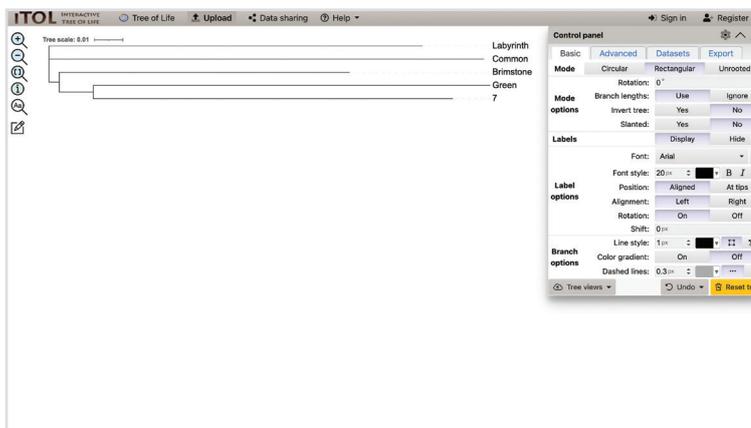
- Click on the **Upload** command from the ribbon of options at the top.



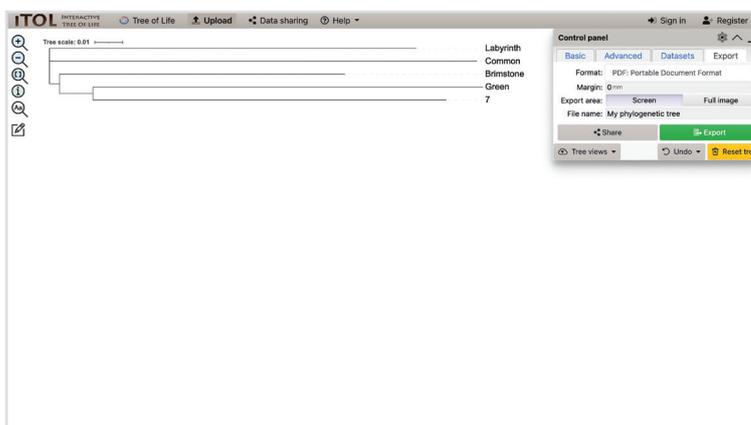
- Type in a tree name.
For example, 'Phylogenetic tree using known barcodes'
- Paste the **Tree data** (copied from the multiple sequence alignment site), into the box labelled **Tree text**.
- Click on the **Upload** button.

□ 5. Choosing the presentation style for the phylogenetic tree

- A rectangular tree should be visible immediately.



- Experiment with different options from the menu. Try adjusting the **mode** (circular / unrooted) and altering the label options and branch options to alter the appearance too.
- Consider the ways in which you could display the data and decide the way in which you wish to show your phylogenetic tree.
- When the phylogenetic tree is shown in a style that satisfies you, then export a copy for use in presenting your results. To do this go to the **Export** tab on the control panel. Choose your format from the drop down menu (PNG or PDF are good choices), add a **file name** and click the big, green **Export** button. Make sure that you save the file using a name and in a location that you will be able to find again.



Questions

- A. Which barcodes (and therefore organisms) are most closely related from this phylogenetic tree?
- B. Look at the organisms most closely related – does this make sense with your knowledge of the invertebrates sampled?
- C. How could you make the phylogenetic tree more accurate?