

From Sequence to Tree: Practical Phylogenetic Techniques

Pierrot Van der Aa
pvderaa@vub.be

This week of lecture will be focused on practical aspects and hands on phylogenetic tools. The goal is to follow the path linking the raw file containing the sequences to the phylogenetic tree. In order to achieve this goal, one method will be used and can be split in the steps below. Each of them includes the programs that I would like you to install on beforehand in order to save some time during the practical. It is also important that you already create an account on the Cipres gateway because the account validation can take some time (<https://www.phylo.org/>).

Introduction to phylogeny (Day 1, morning)

We will start by a brief reminder on the importance, use and understanding of phylogenies. This part focuses on the reading and interpretation of existing trees.

Goal: critically analyse the different information provided by different types of trees.

Software: None

Online platform: None

Getting sequences (Day 1, afternoon)

The first step in the construction of phylogenies is to get DNA sequences. One way is to generate the sequence from available samples. The other way is to access open data online (in this case, on the NCBI platform).

Goal: search and retrieve DNA sequences from GenBank (NCBI)

Software: None

Online platform: <https://www.ncbi.nlm.nih.gov/nuccore>

Cleaning the sequences (Day 2, morning)

The downloaded sequences have different structures of their titles which may lead to issues in some scripts or programs, we then need to give them a proper name according to a pattern.

Goal: search and retrieve DNA sequences from GenBank (NCBI)

Software: Notepad++ (Windows) (<https://notepad-plus-plus.org/download/v7.7.1.html>), BBEdit (Mac) (<https://www.barebones.com/products/bbedit/bbedit12.html>) or any other similar text editor

Online platform: None

Align the sequences (Day 2, afternoon)

Now that the sequence file is ready to be used, we can open it in a sequence editor and align the sequences in order to compare them.

Goal: align the sequence using muscle

Software: MEGA (Windows, Mac) (<https://www.megasoftware.net/>, select your os (Windows, Mac, ...), Graphical GUI and MEGA X), Seaview (Windows, Mac) (<http://doua.prabi.fr/software/seaview>)

Online platform: MAFFT (<https://mafft.cbrc.jp/alignment/server/>)

Phylogenetic Analyses (Bayesian analyses) (Day 3, morning)

Using the previously aligned file, we will construct a Bayesian tree. We will discuss the pros and cons of this technique and will mainly focus on the use the Cipres gateway.

Goal: construct a Bayesian tree

Software: MrBayes (Windows, Mac) (<http://nbisweden.github.io/MrBayes/download.html>), (if times allows: Tracer (<https://github.com/beast-dev/tracer/releases>))

Online platform: Cipres (<https://www.phylo.org/>)

Phylogenetic Analyses (Maximum likelihood) (Day 3, afternoon)

Using the previously aligned file, we will construct a maximum likelihood tree. We will discuss the pros and cons of this technique and will mainly focus on the use of the online platforms.

Goal: construct a maximum likelihood tree

Software: RAXML (Windows, Mac) (<https://github.com/stamatak/standard-RAXML>), (if times allows: jmodeltest (<https://github.com/ddarriba/jmodeltest2>)(or googledrive link))

Online platform: IQTREE (<http://iqtree.cibiv.univie.ac.at/>), PhyML (<http://atgc.lirmm.fr/phyml/>)

Tree interpretation and graphical customisation (Day 4, morning)

Now that we have run the analyses, we will open the output files, analyse them and learn how to present a tree.

Goal: Provide useful and correct information on a tree

Software: FigTree (Windows, Mac) (<http://tree.bio.ed.ac.uk/software/figtree/>), Inkscape (optional) (Windows, Mac) (<https://inkscape.org/release/inkscape-0.92.4/>)

Online platform: None

Q&A and Blast (Day 4, afternoon)

The remaining time can be used to explore your own datasets, finish unfinished tasks, answer questions or learn to use the Blast tool provided by NCBI used for identifying sequences from an unknow specimen or gene.

Goal: Feel comfortable with the different steps

Software: None

Online platform: NCBI Blast (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)