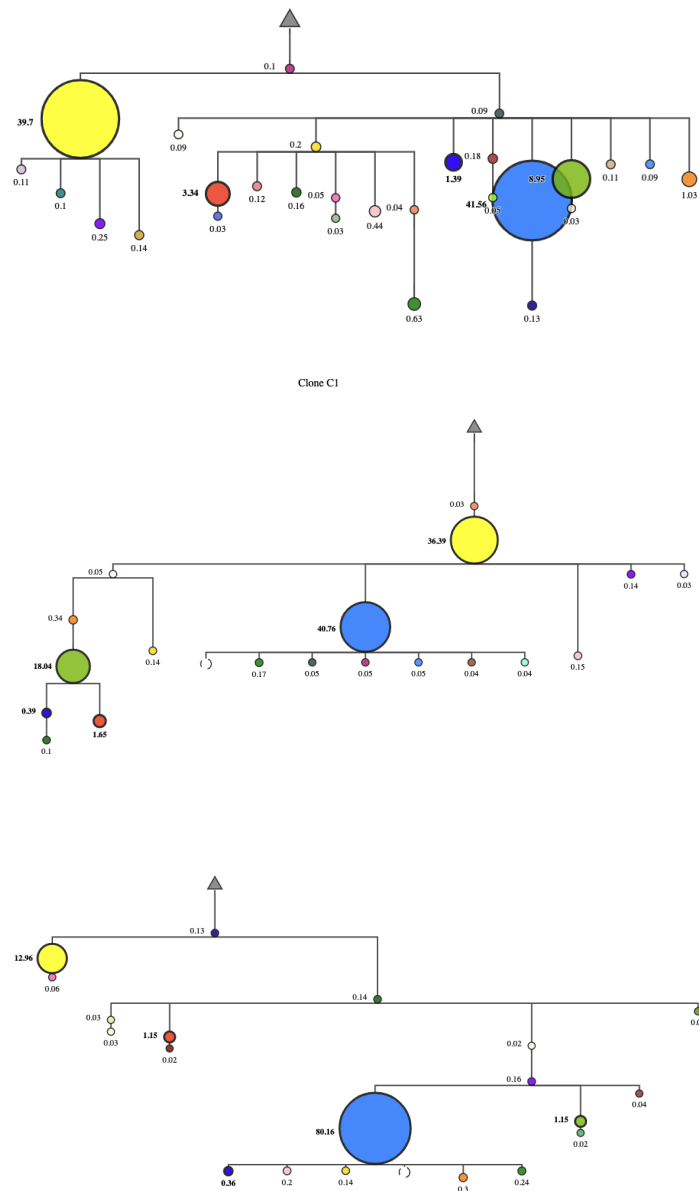# Week Four 21/02/2022

This week's objective: Describe a phylogenetic tree

New week, new challenges!
You have already mastered running Clonaltree, and you can now have multiple highly variable trees.
The next question is how we can interpret these trees' information.

Clone C1

Generally, questions about phylogenetic relationships within or between assemblages[1] tend to ask three types of questions: how much, how different, or how regular? These questions reflect three dimensions of a phylogenetic tree: richness, divergence, and regularity.

1. Richness (how much?)

   Richness metrics sum up the quantity of phylogenetic differences present in an assemblage, and we can further distinguish metrics according to the type of basic units they sum across. The aim is to understand how much evolutionary history is associated with a set of sequences.

2. Divergence (How different?)

   The divergence dimension contains metrics that average the distribution of units extracted from a phylogenetic tree. The aim is to understand how closely related (on average) are sequences within a clone.

3. Regularity (How regular ?)

   The regularity dimension contains metrics that characterize how the phylogenetic tree differs from a star phylogeny (i.e., a phylogeny in which all species are equally unrelated). In other words, these metrics quantify how regularly species are located along the phylogenetic tree and how evenly distant they are from each different species. The aim is to understand how evenly evolutionary history is distributed between sequences within a clone.

For this week, we will work on the Richness and Divergence dimensions of our trees. As you can guess, a vast collection of phylo-diversity metrics are available. We will start with the most simple ones based on branch length. What you will calculate is the Phylogenetic diversity ("PD"). PD is a measure of biodiversity based on phylogeny. Faith (1992) defined the phylogenetic diversity of a set of species as equal to the sum of the lengths of all those branches on the tree that span the members of the set. The branch lengths on the tree are informative because they count the relative number of new features arising along that part of the tree.

Faith and Richards (2012) explored the loss of PD for corals and presented a simple PD example:
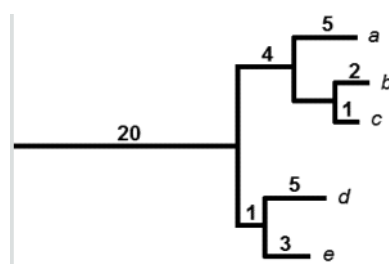
A hypothetical phylogenetic tree for species a through e. Branch lengths are shown above branches. The PD is 41 for this set of species (20 + 5 + 4 + 2 + 1 + 5 + 1 + 3). If species "a" was lost, five units of PD would be lost.

---

[1] Species assemblage, in biology: all the species that exist in a particular habitat. In our case, species are the representative sequences of clonotypes within a clone.

So **for a given tree**, let's calculate :

- **The number of branches**
- **Sum of all branch lengths (PD)**
- **Sum of all branch lengths divided by the number of clonotypes of the lineage (avPD)**

You can write a python script that does these calculations **systematically for each tree generated by Clonaltree.**

In the upcoming weeks, we will work on other metrics, and also we will try to incorporate the sequences abundance in our calculations.

Have fun!

Reference :

Tucker, C.M., Cadotte, M.W., Carvalho, S.B., Davies, T.J., Ferrier, S., Fritz, S.A., Grenyer, R., Helmus, M.R., Jin, L.S., Mooers, A.O., Pavoine, S., Purschke, O., Redding, D.W., Rosauer, D.F., Winter, M. and Mazel, F. (2017), A guide to phylogenetic metrics for conservation, community ecology and macroecology. Biol Rev, 92: 698-715. https://doi.org/10.1111/brv.12252