

Advanced tutorial for using OrthoFinder3

OrthoFinder3 provides a new workflow to assign new genes from new species to an already inferred set of orthogroups for a smaller, core group of species. If you are analysing >100 species, we recommend that you use this scalable implementation.

To do this effectively, we need to pick a good set of core species. If we are running OrthoFinder on some bees, some moths, and some flies, we want to build our core orthogroups on a phylogenetically diverse set of those species. If instead we chose randomly and ended up using all moths as our core species, we might end up with less accurate orthogroups, and a longer runtime.

If you have a phylogenetic tree, you can use tools like Phylogenetic Diversity Analyzer to pick a representative set (<http://www.cibiv.at/software/pda/>). Or, you could just pick a diverse set yourself. We recommend that you pick 64 species for your core set.

First, run OrthoFinder on the subset of 64 species

```
orthofinder [options] -f <core>
```

Then, add the additional species to the results of the core run

```
orthofinder [options] --assign <additional> --core <Results_Dir>
```

This workflow also makes it really quick and easy to add new species to previous OrthoFinder runs. For example, if your research group works on various species of Angiosperms you might collectively share a core OrthoFinder results folder with a phylogenetically diverse set of species, which individual researchers could easily add any new species to.

An important note is that this workflow requires multiple sequence alignments, so unfortunately you cannot use it to add to OrthoFinder2 results that were run with the default

-M dendroblast option.

Using Outgroups

You can make orthogroup inference more accurate by including outgroup species. You just need to make sure that you use the correct N*.tsv file in Phylogenetic_Hierarchical_Orthogroups to look at the orthogroups (use species tree to discover which one)