Module 1: Datasets

What species are available for my analysis and how are they related to each other?

Assumption: All species are related to each other, and this relationship can be represented by a tree

Analysis

Task 1: What species are available for my analysis?

1. Go to NCBI Datasets
2. Search for all available eukaryotic genomes
3. Filter for genomes “with RefSeq Annotation”
4. Add column “taxid”
5. Download table in CSV format

Task 2: How are they related to each other?

1. Extract the information from the “taxid” column and save it in a `.txt` file

   Download file with taxids from here

2. Go to NCBI CommonTree and upload your `.txt` file (→ menu: **Add from file**)
3. Download the tree in Phylip format (→ menu: **save as → phylip tree**)
4. Open the iTOL web page
5. Upload the tree into iTOL and explore:
    1. How many animals, how many fungi, how many plants are there? (Tip: the nodes in the tree will be named according to entries in NCBI Taxonomy)
    2. Highlight these three clades with different colors
    3. Compare the circular vs rectangular representation of the tree

Summary and discussion

- Note down your observations and questions
- Discuss with the group

Back to main

From:
https://applbio.biologie.uni-frankfurt.de/teaching/wiki/ - **Teaching**

Permanent link:
https://applbio.biologie.uni-frankfurt.de/teaching/wiki/doku.php?id=compgenomics:modul1

Last update: **2025/09/23 13:09**