

Installation:

Basic installation:

1. Copy the Phylogeny_of_birds.py to the choosen directory.

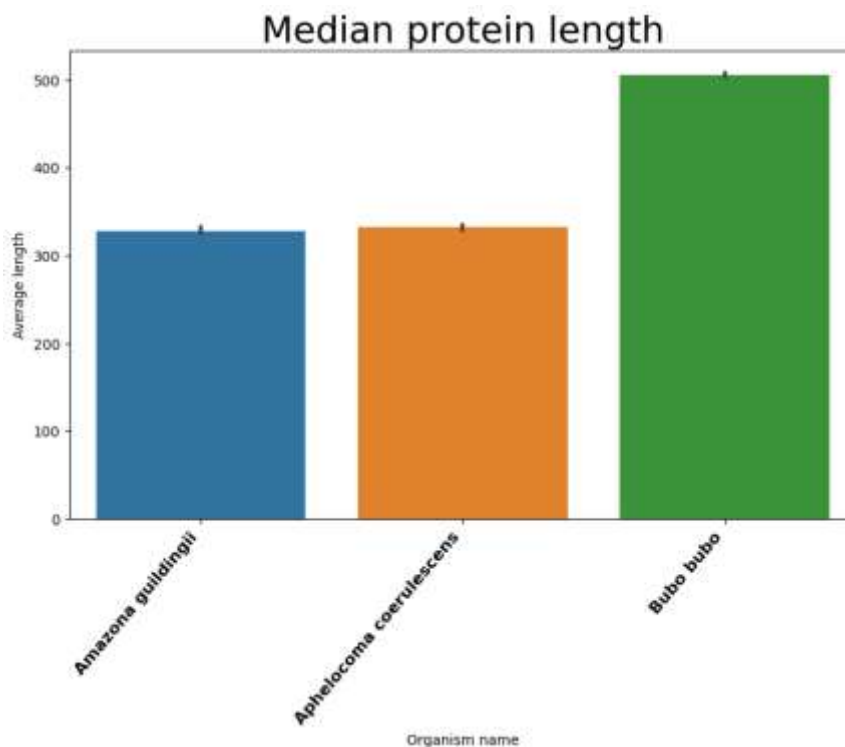
Advanced installation:

1. Copy the Phylogeny_of_birds.py to the choosen directory.
2. Copy Path.txt to the same directory as Phylogeny_of_birds.py.
3. Install Mmseq2 (<https://github.com/soedinglab/MMseqs2>)
4. Install Iqtree (<http://www.iqtree.org/>)
5. Install Muscle (<https://www.drive5.com/muscle/>)
6. Install PhymI (<http://www.atgc-montpellier.fr/phymI/>)

Functions:

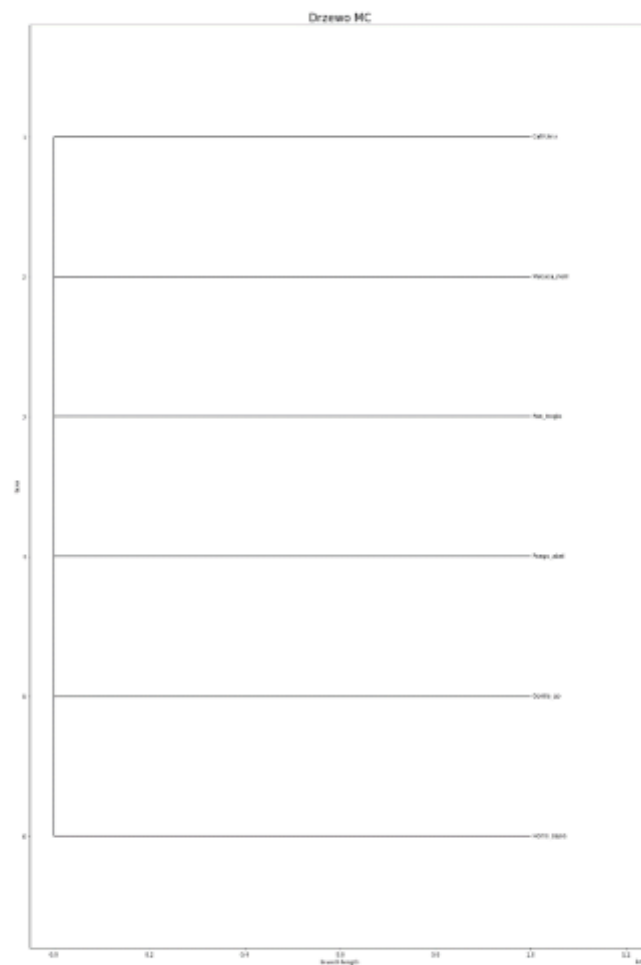
If you used the basic installation, you only have access to some program functions, such as: median protein length, average protein length, amino acid content table, amino acid content plot.

Example:



If you have used an advanced installation, you have access to the full capabilities of the program, such as: making majority phylogeny trees with paralogs and without paralogs and optionally provide bootstrapping to enhance the quality of majority consensus tree.

Example:



Usage:

To start the program type:

```
python3 Phylogeny_of_birds.py -h
```

or

```
python3 Phylogeny_of_birds.py --help
```

and you will see all options.

Type:

```
python3 Phylogeny_of_birds.py --update
```

and you will get available organism list. The you can use these organism name to your analysis.

Example:

```
python3 Phylogeny_of_birds.py -in_file actual_analysis_organisms_names --no_download -median  
median_plot
```

actual_analysis_organisms_names is a file that contains available organism list. You can create your own and choose species from actual_analysis_organisms_names list.

Phylogeny analysis for inexperienced users:

In file Path.txt write your path to PhymL.

Exapmle:

```
/home/user/PHYML/PhyML-3.1_linux64
```

Type:

```
python3 Phylogeny_of_birds.py -in_file actual_analysis_organisms_names --no_download -auto .
```

and program will create a majority consensus tree from choosen species.

NOTE: The auto procedure is very simple, so for clustering there is used only Linclust algorithm which is the fastest, and the trees are made from clusters with paralogs.

The precodure may take a while, it depends on how many species you choose.

Procedure steps:

1. Division of all proteins into clusters using MMseqs2 (Linclust).
2. Division of each cluster into specific file with paralogs.
3. Alignment of all sequences in each cluster using Muscle.
4. Making ML tree of each cluster using PHyml.
5. Making majority consensus tree.
6. Result is in a .png file and newick file.

Phylogeny analysis for experienced users:

1. Clustering with Mmseq2 (it is only example script below):

Before clustering type: `python3 Phylogeny_of_birds.py -in_file actual_analysis_organisms_names --no_download -merge .`

This function will merge the sequence files into one file readable for MMseqs2.

and then:

for example:

```
mmseqs createdb DB.txt DB
```

```
mmseqs linclust DB DB_clu tmp
```

```
mmseqs createsubdb DB_clu DB DB_clu_rep
```

```
mmseqs convert2fasta DB_clu_rep DB_clu_rep.fasta
```

```
mmseqs createtsv DB DB DB_clu DB_clu.tsv
```

```
mmseqs createseqfiledb DB DB_clu DB_clu_seq
```

```
mmseqs result2flat DB DB DB_clu_seq DB_clu_seq.fasta
```

but you can choose what function you want. MMseqs2 is very advanced tool. You can make more accurate clustering than linclust.

2. Modifications of clustered file:

If you want trees with paralogs:

```
Type: python3 POB.py -in_file actual_analysis_organisms_names --no_download -clusters_no_rep DB_clu_seq.fasta
```

if you want trees without paralogs:

NOTE: The program will ask you about minimum number of species per each cluster.

```
Type: python3 Phylogeny_of_birds.py -in_file actual_analysis_organisms_names --no_download -clusters_rep DB_clu_seq.fasta
```

3. Alignment using Muscle (it is only example script below):

Make sh script with content below and run in:

```
#!/bin/bash
```

```
for f in ./*.cluster; do
```

```
    muscle -in $f -phyiout ./$f.phypob
```

```
done
```

4. Draw trees using PhyML:

Make sh script with content below and run it (it is only example script below):

```
#!/bin/bash
```

```
for f in ./*.phypob; do
```

```
    /home/user/PHYML/PhyML-3.1_linux64 -i $f -d aa
```

```
done
```

5. Bootstrap using Iqtree

Make sh script with content below and run it (it is only example script below):

NOTE: You will be asked about bootstrap support value. Typically it is 70.

```
#!/bin/bash
```

```
for f in */*.phy; do
```

```
    iqtree -s $f -m JTT+G -bb 1000 -alrt 1000 -pre $f.tree
```

```
done
```

6. Draw tree:

Type `cat *.phy > name.newick`

and then type:

NOTE: You will be asked about majority consensus threshold. You can choose from 0 to 1. The most popular value is 0.5. If you choose 1 then you will have strict consensus tree.

```
python3 Phylogeny_of_birds.py -in_file actual_analysis_organisms_names --no_download -mc  
name.newick
```

References

- <https://github.com/soedinglab/MMseqs2>
- <http://www.iqtree.org/>
- <https://www.drive5.com/muscle/>
- <http://www.atgc-montpellier.fr/phyml/>