Phylogeny of birds

Podtytuł prezentacji

Overview of the project

Phylogeny_of_birds.py is a tool prepared for biologists who are not fluent in Python language but they know how to use Bash shell. The program allows to make a phylogeny and amino acid content analysis for chosen species of birds.

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 Igtree (http://www.iqtree.org/)
5. Install Muscle (https://www.drive5.com/muscle/)
6. Install Phyml (http://www.atgc-montpellier.fr/phyml/)

Main 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.

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.

Requirements

- Unix system like Ubuntu or Ubuntu for Windows 10 emulator.
- Mmseq2 installed (optional) (https://github.com/soedinglab/MMseqs2)
- Igtree installed (optional) (http://www.iqtree.org/)
- Muscle installed (optional) (https://www.drive5.com/muscle/)
- Phyml installed (optional) (http://www.atgc-montpellier.fr/phyml/)

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.

Phylogeny analysis

Phylogeny analysis for inexperienced users:
 In file Path.txt write your path to Phyml.
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. Dividision 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.Before clustering type: python3 Phylogeny_of_birds.py -in_file actual_analysis_organisms_names --no_download -merge .
2. Run mmseq2
3. *Trees with paralogs:*
Type: python3 POB.py -in_file actual_analysis_organisms_names --no_download -clusters_no_rep DB_clu_seq.fasta

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

Phylogeny analysis for experienced users

4. Alignment using Muscle
5. Draw trees using PhyML
6. Bootstrap using Iqtree
7. Draw tree using command:
python3 Phylogeny_of_birds.py -in_file actual_analysis_organisms_names
--no_download -mc name.newick

Thank you for your attention

Marcin Klimko

Elżbieta Karpińska

Michał Maruchin

Bibliography

- https://github.com/soedinglab/MMseqs2
- http://www.iqtree.org/
- https://www.drive5.com/muscle/
- http://www.atgc-montpellier.fr/phyml/
- Prum, R., Berv, J., Dornburg, A. et al. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, 569–573 (2015). https://doi.org/10.1038/nature15697