

Program of identificate species dividing line manual

For Everyone

<http://45.76.122.117:8000/home/>
Free webservises

Documentation

Zhaotong, 2018.01

Author:
WangRui

1 Introduction

We coding this program to solve problem that identification of the species dividing line,input a sequence file (.fasta ...) and output a bar chart(.png file type) about frequency distribution of distance,so we can get a frist gap,this gap corssponding a pecentage that we need.Like under bar chart Figure 1,the frist gap corssponding a 2.5 percentage that read is automatic by program,when we get the dividing line(a percentage),so using it to modify previous construted phylogenetic tree that we can get species tree.

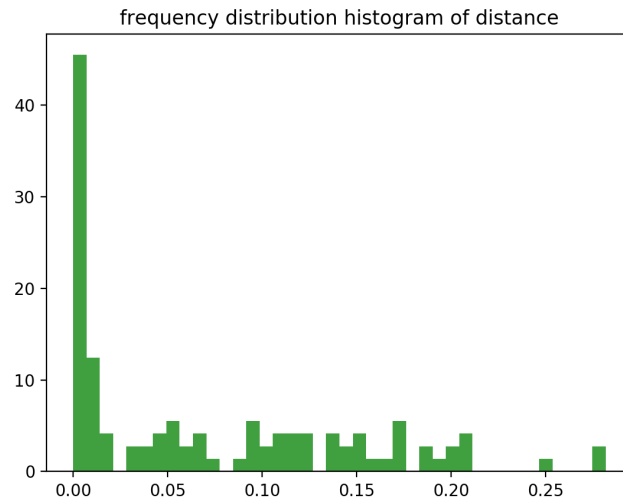
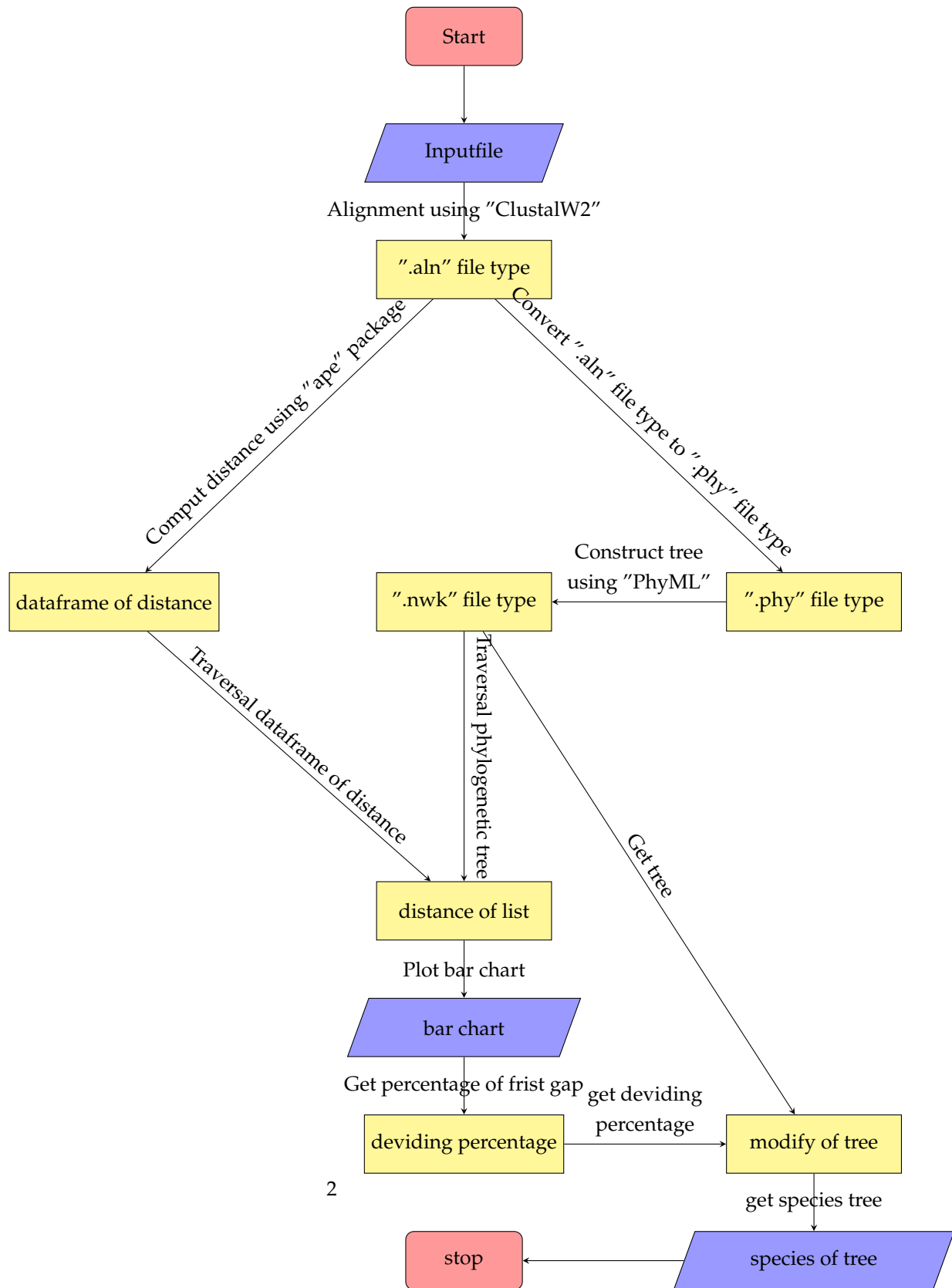


Figure 1: bar chart

2 algorithm

Here we describe the details of program,this program is mainly coded by python language(python3.6 distribustion),using PhyML[GG03] to construt phylogenetic tree,using rpy2 module to concatenate python and ape package[PCS04](a R language package),cluswtal2[LBB⁺07] be used for building phylogenetic tree,the software flow chart is as follows:



2.1 what python packages we used?

We used Biopython[EBPB12](but we need import Bio,not Biopython in python program),rpy2[rpy],itertools,pandas,numpy,maltplotlib,copy and pydot;You can look for them if you aren't familiar with these python package.

2.2 How to use it?

We provide a website(...) for you, you only need a browser and open the url to upload your file(must be ".fasta, ..." type file),email and so on,we will send the results to you by email,the normal results include a picture of bar chart and a species tree file that is ".nwk" type file.

References

- [EBPB12] ERIC, Talevich ; BRANDON, M. I. ; PETER, J.A. C. ; BRAD, A. C.: Bio.Phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. In: *BMC Bioinformatics* 209 (2012). <http://dx.doi.org/10.1186/1471-2105-13-209>. – DOI 10.1186/1471-2105-13-209
- [GG03] GUINDON, S. ; GASCUEL, O.: PhyML : "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.". In: *Systematic Biology* 52 (2003), Nr. 5, S. 696–704
- [LBB⁺07] LARKIN, MA ; BLACKSHIELDS, G ; BROWN, NP ; CHENNA, R ; MCGETTIGAN, PA ; MCWILLIAM, H ; VALENTIN, F ; WALLACE, IM ; WILM, A ; LOPEZ, R ; THOMPSON, JD ; GIBSON, TJ ; HIGGINS, DG: In: *Bioinformatics* 23 (2007), Nr. 21, 2947–2948 S. <http://dx.doi.org/10.1093/bioinformatics/btm404>. – DOI 10.1093/bioinformatics/btm404
- [PCS04] PARADIS, E. ; CLAUDE, J. ; STRIMMER, K.: APE: analyses of phylogenetics and evolution in R language. In: *Bioinformatics* 20 (2004), S. 289–290
- [rpy] *rpy2*. <https://rpy2.bitbucket.io/>,