

From Sequence to Tree

Hands-on Tools for Phylogenies

Pierrot Van der Aa

[pvderaa @vub.be](mailto:pvderaa@vub.be)

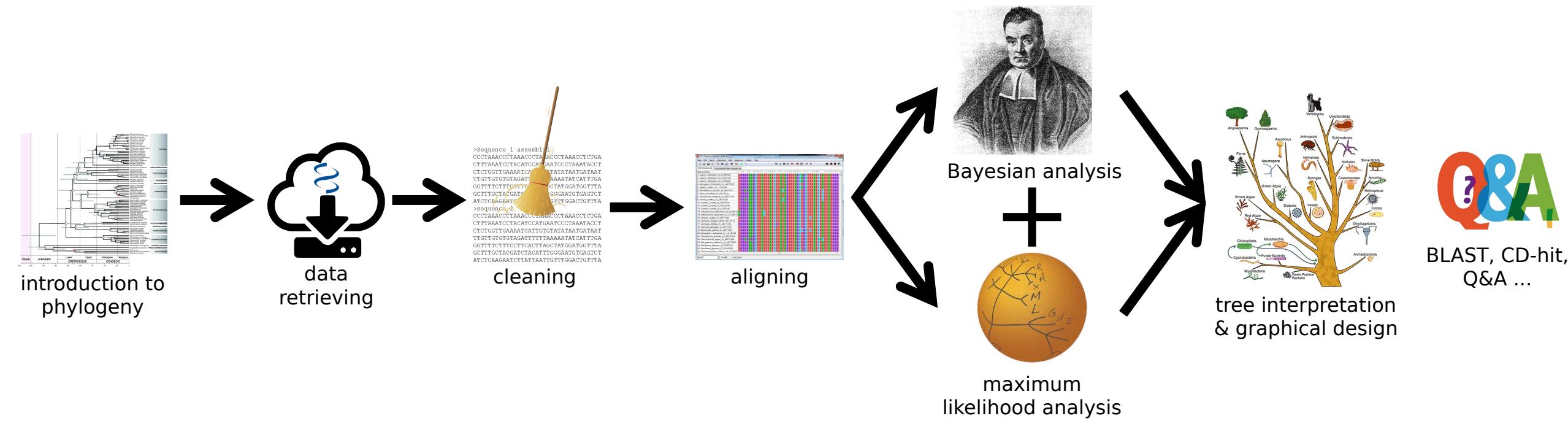
08th-11th of October 2019



UNIVERSITY
OF ABERDEEN



Content of the tutorial

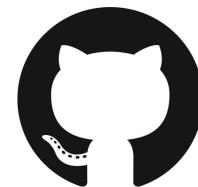


Usefull links

Path to the programs

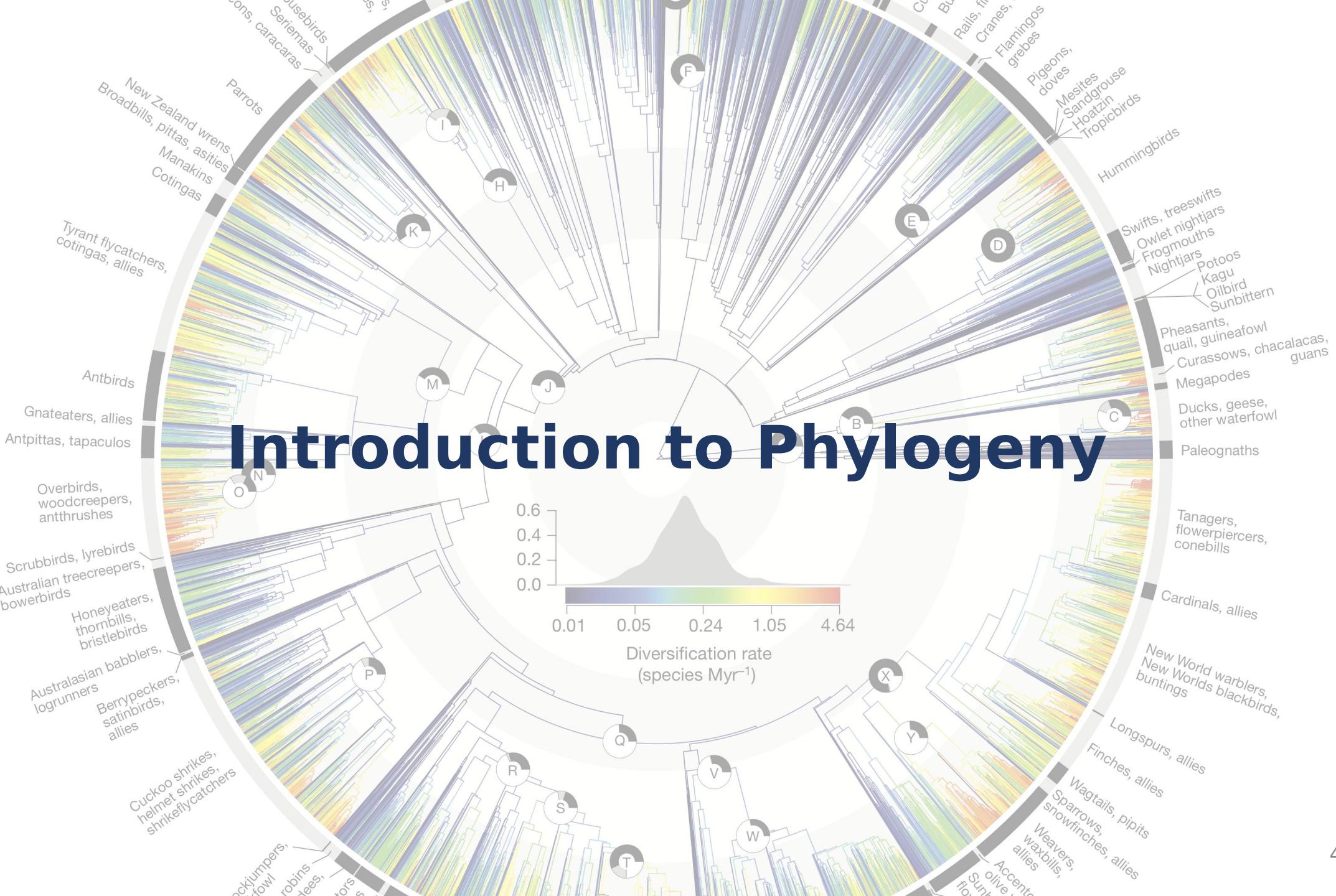
F:\shortcuts\
2019\
Life Science and Medicine\
Biological Sciences\
Plant & Soil Science\
Phylogeny Course

Link to the files

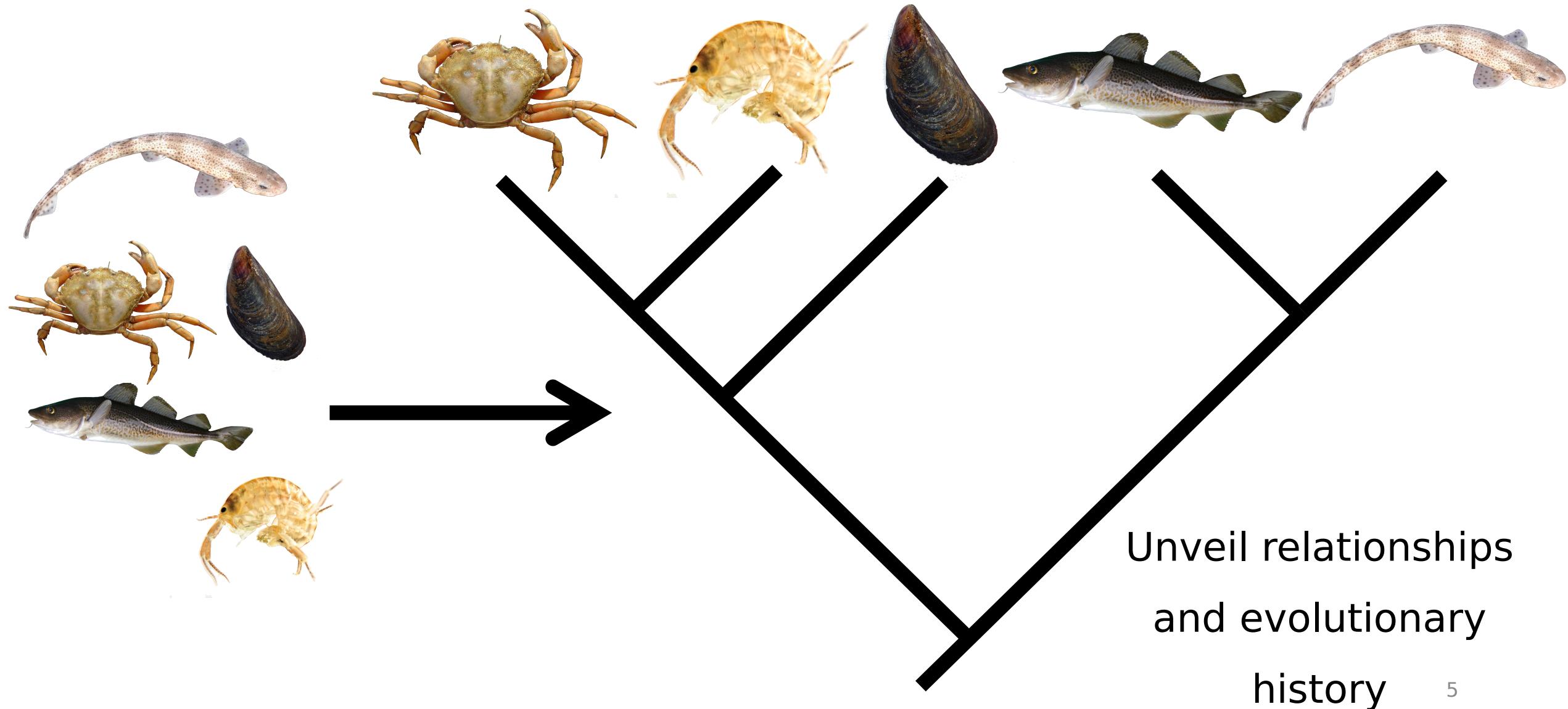


<https://github.com/PierrotVdAa/PhylogenyAberdeen2019>

Introduction to Phylogeny



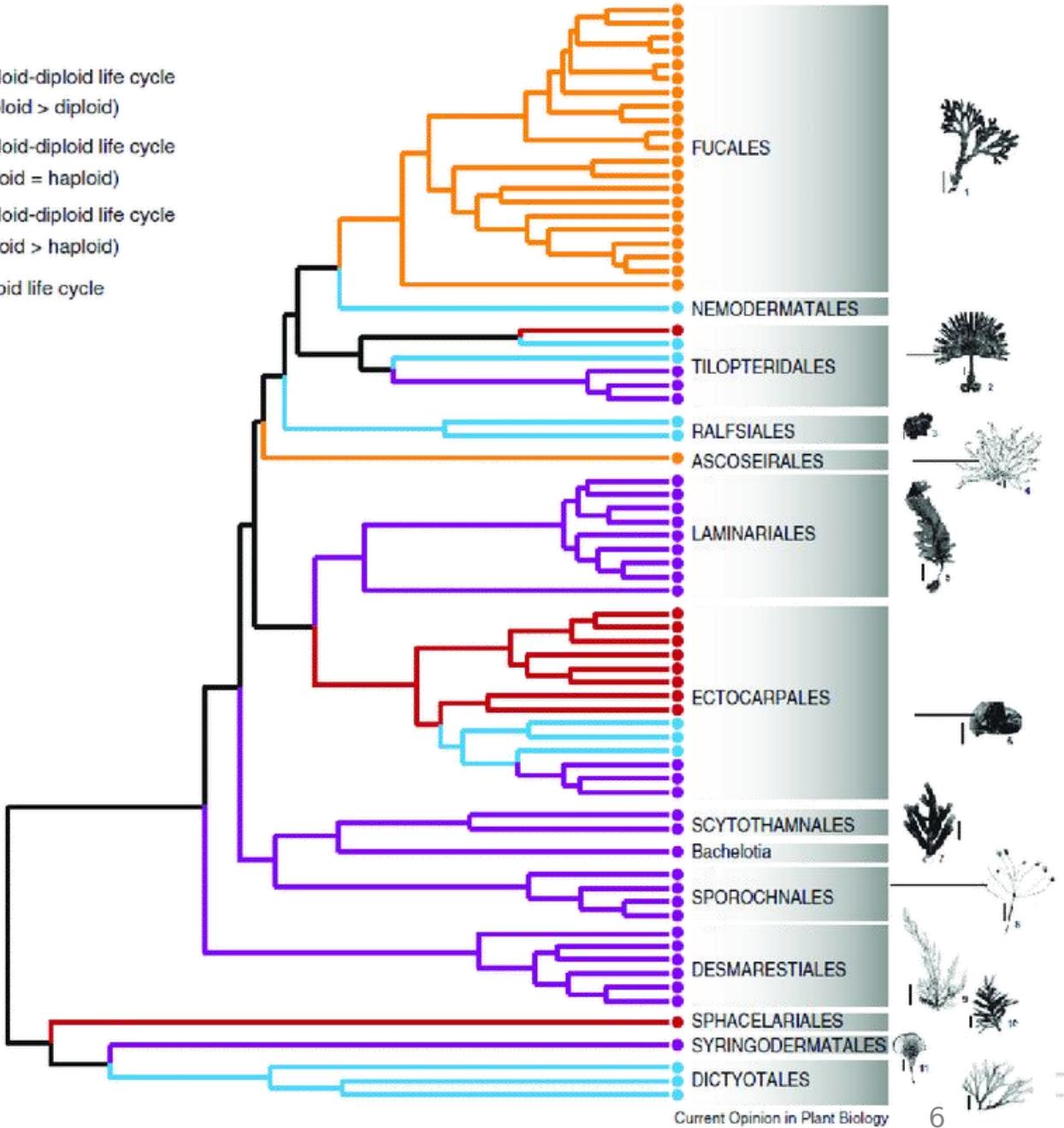
What is phylogeny?



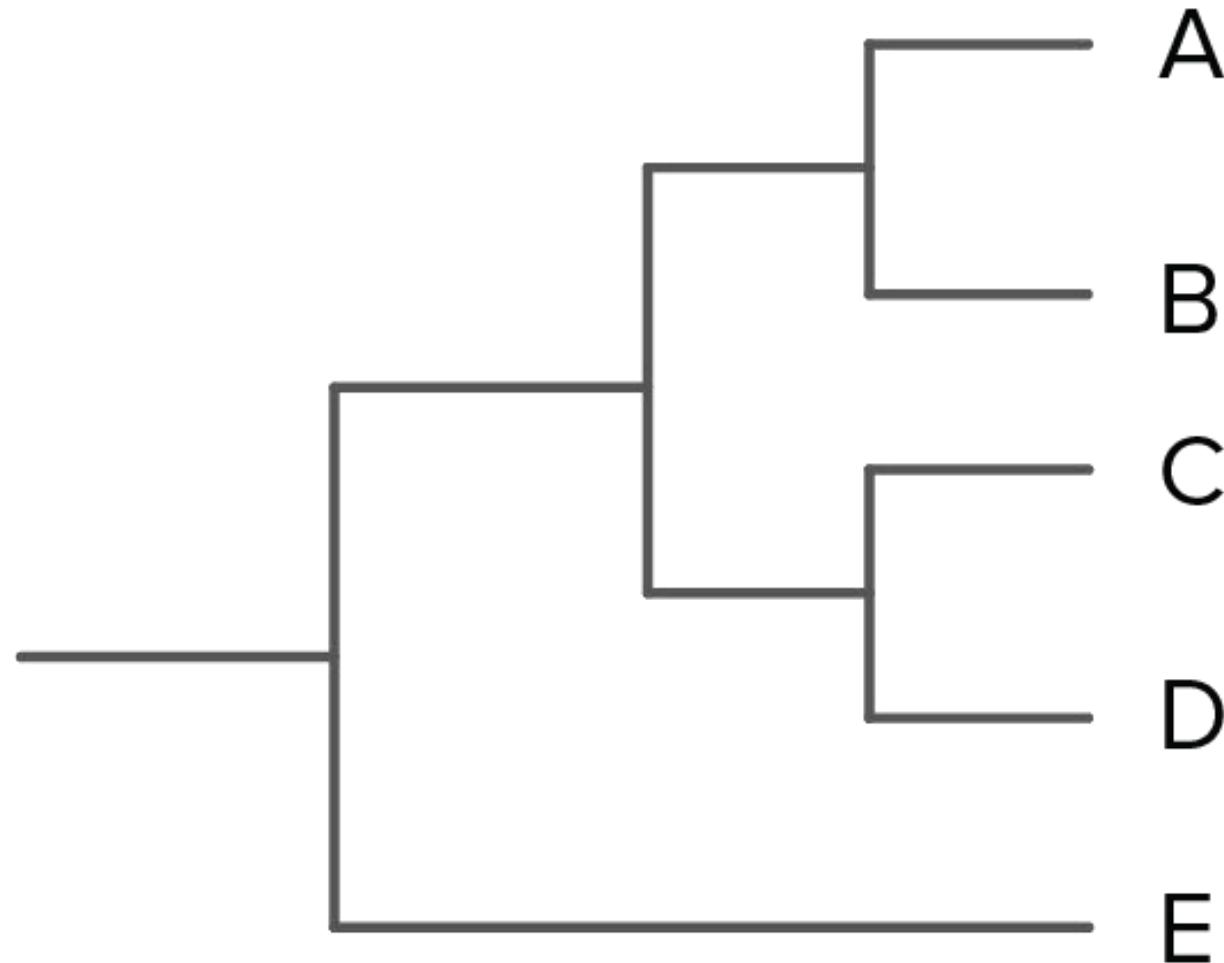
Uses of phylogeny

- Relationships
- Evolutionary history
- Trait analysis
- Time calibration
- Interaction with the environment
- Diversification pattern
- Conservation

- Haplod-diploid life cycle (haploid > diploid)
- Haplod-diploid life cycle (diploid = haploid)
- Haplod-diploid life cycle (diploid > haploid)
- Diploid life cycle

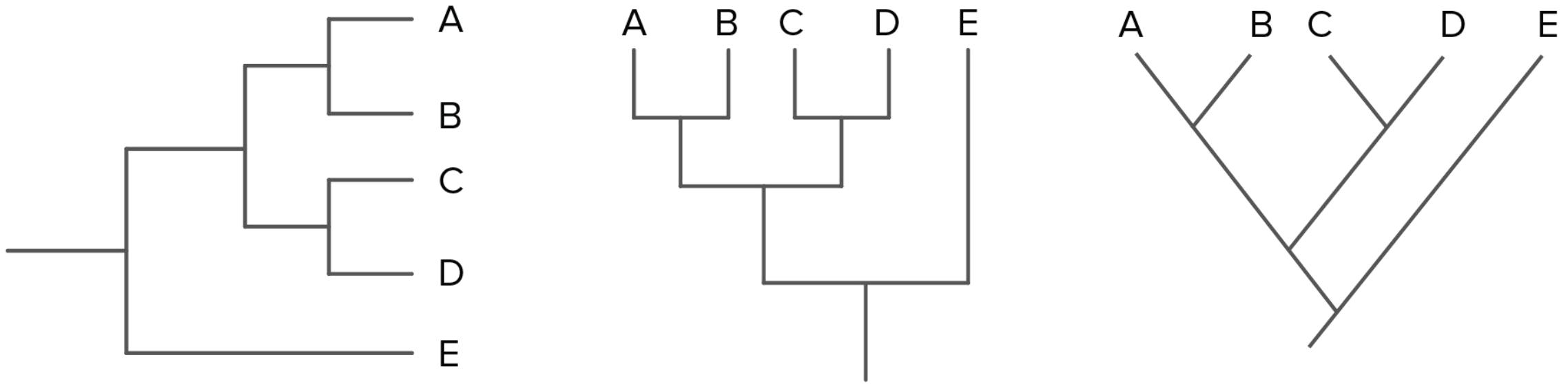


Reading a tree



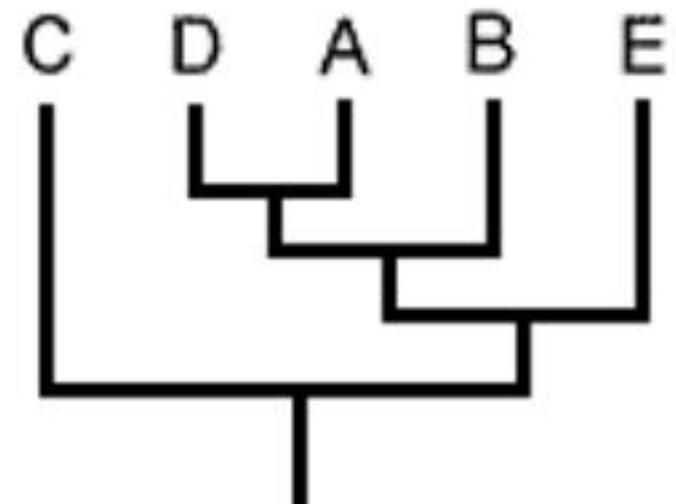
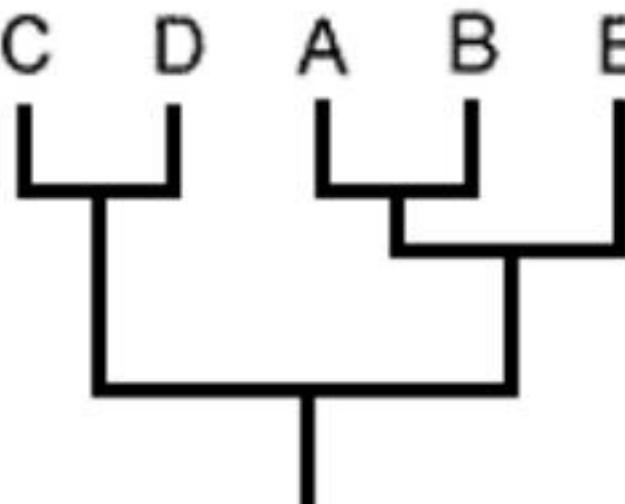
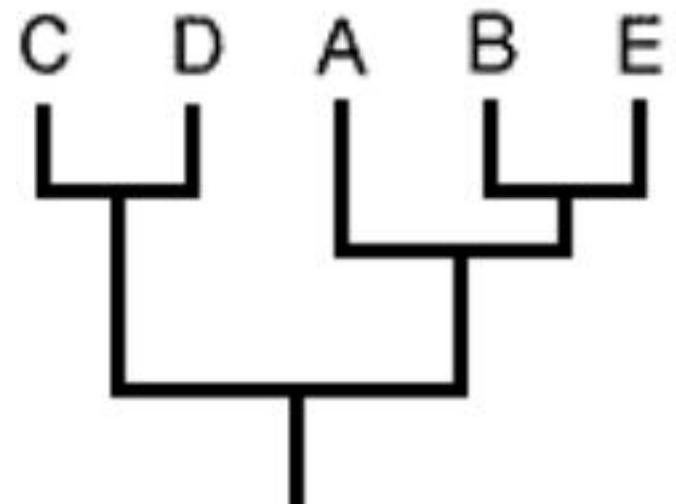
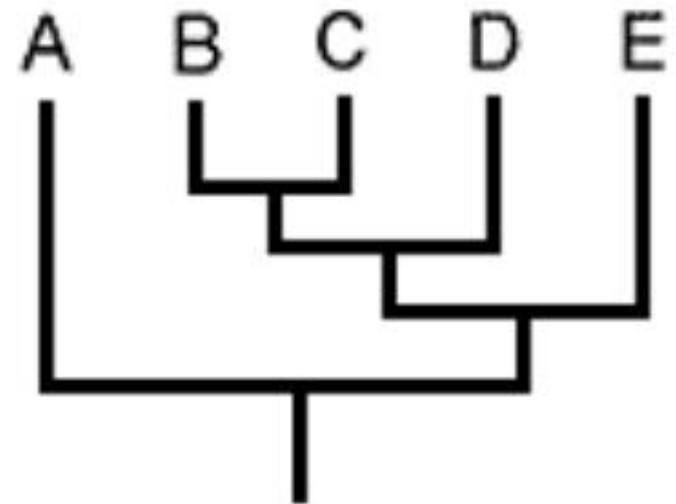
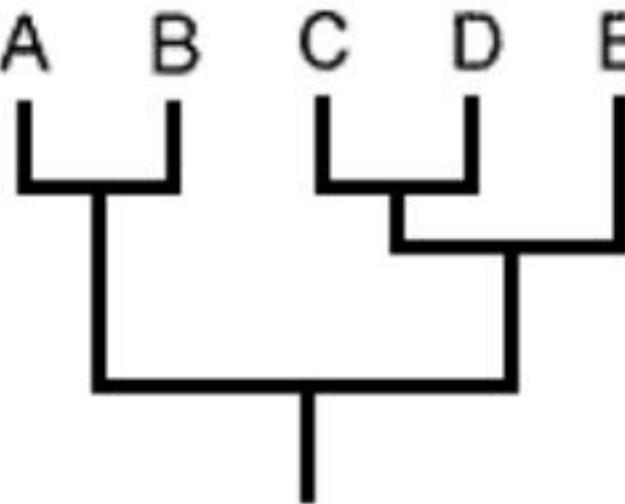
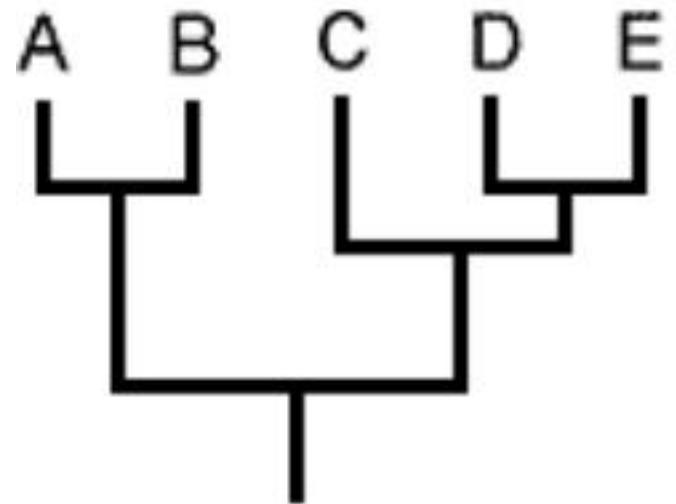
- A is sister of B
- A is more closely related to B than to C

Are those trees different or the same?



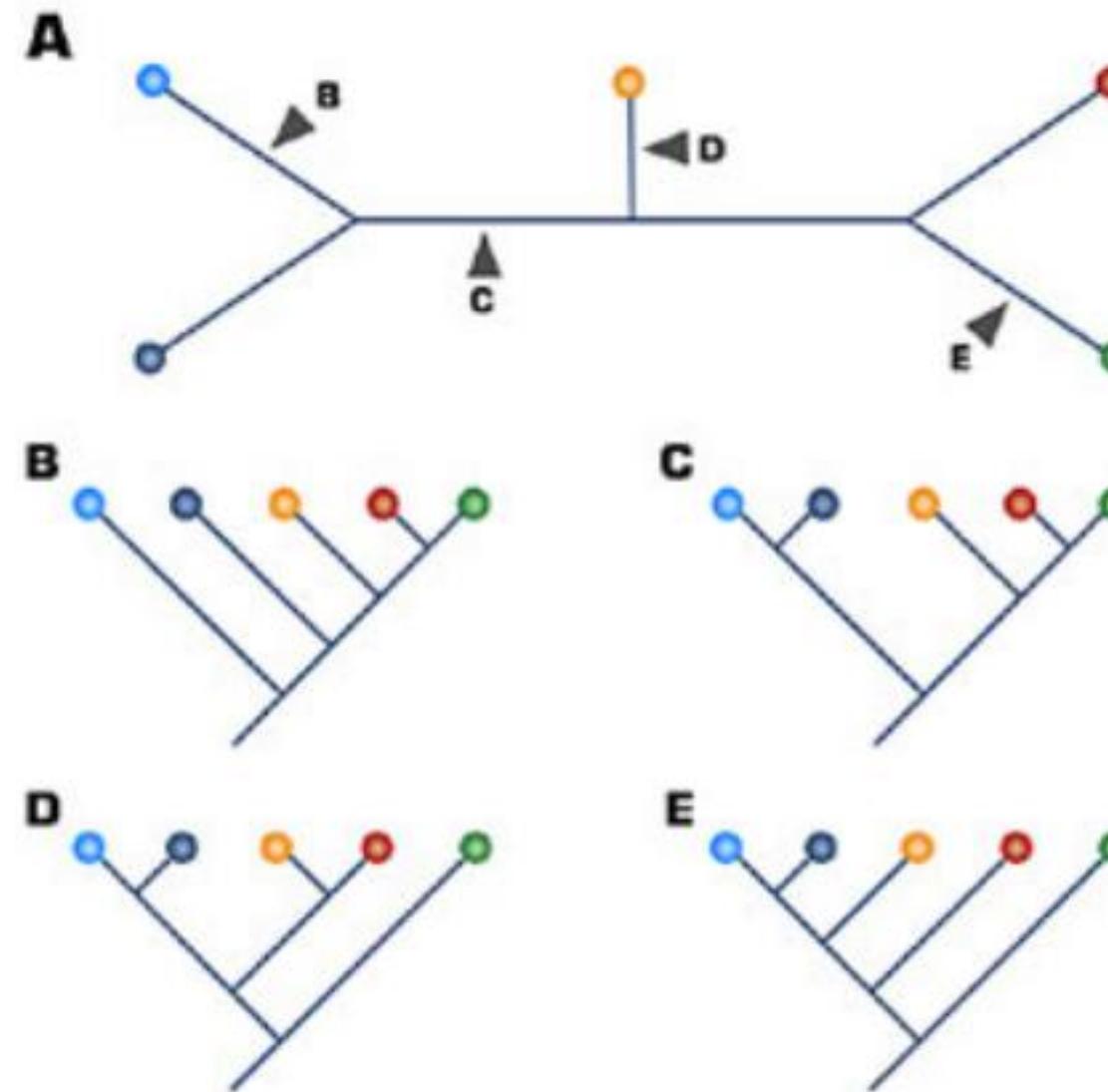
Those figures are different representation of a same tree

Are those trees different or the same?



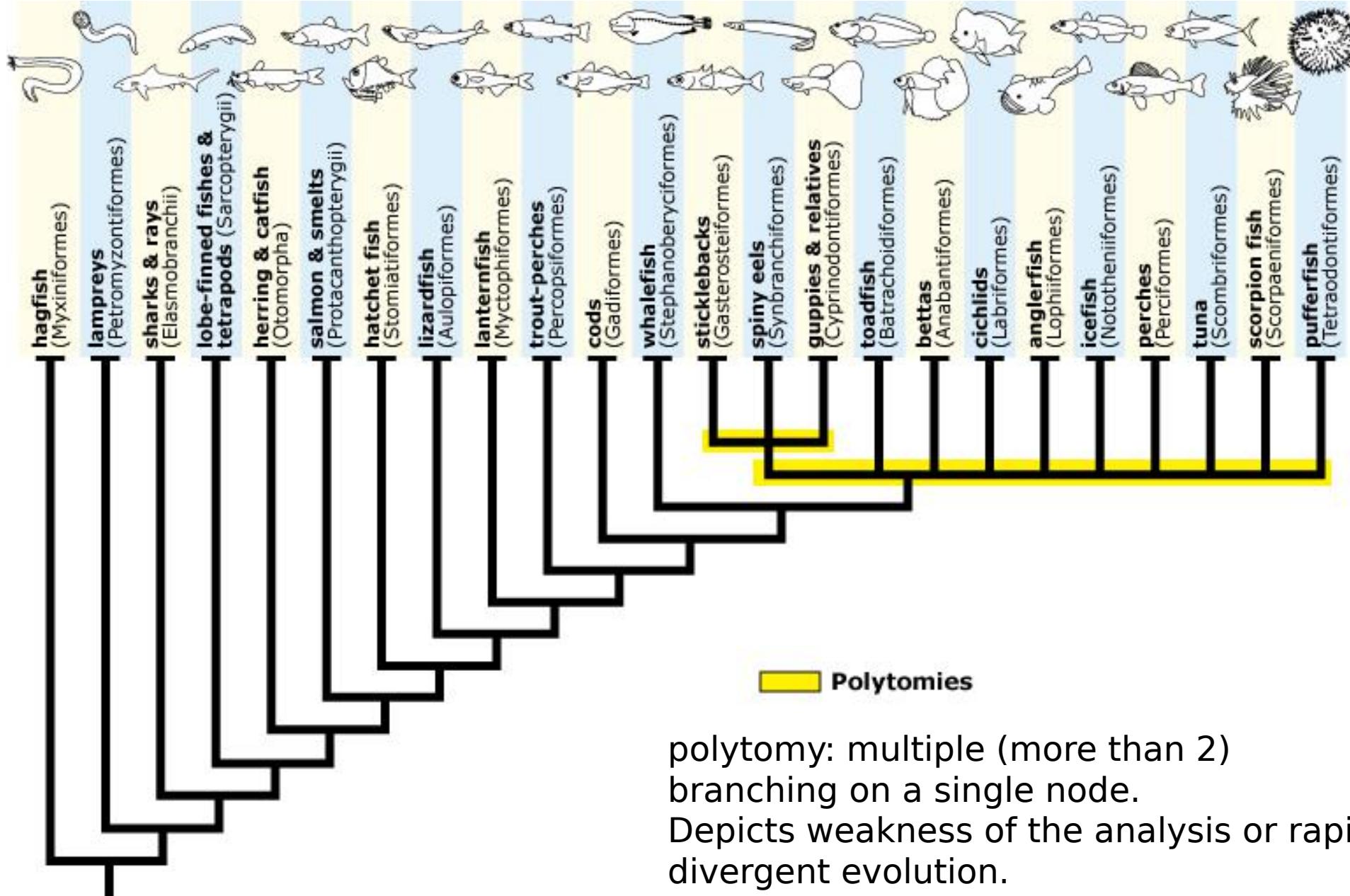
Those figures are different representation of a same tree

Rooted or not

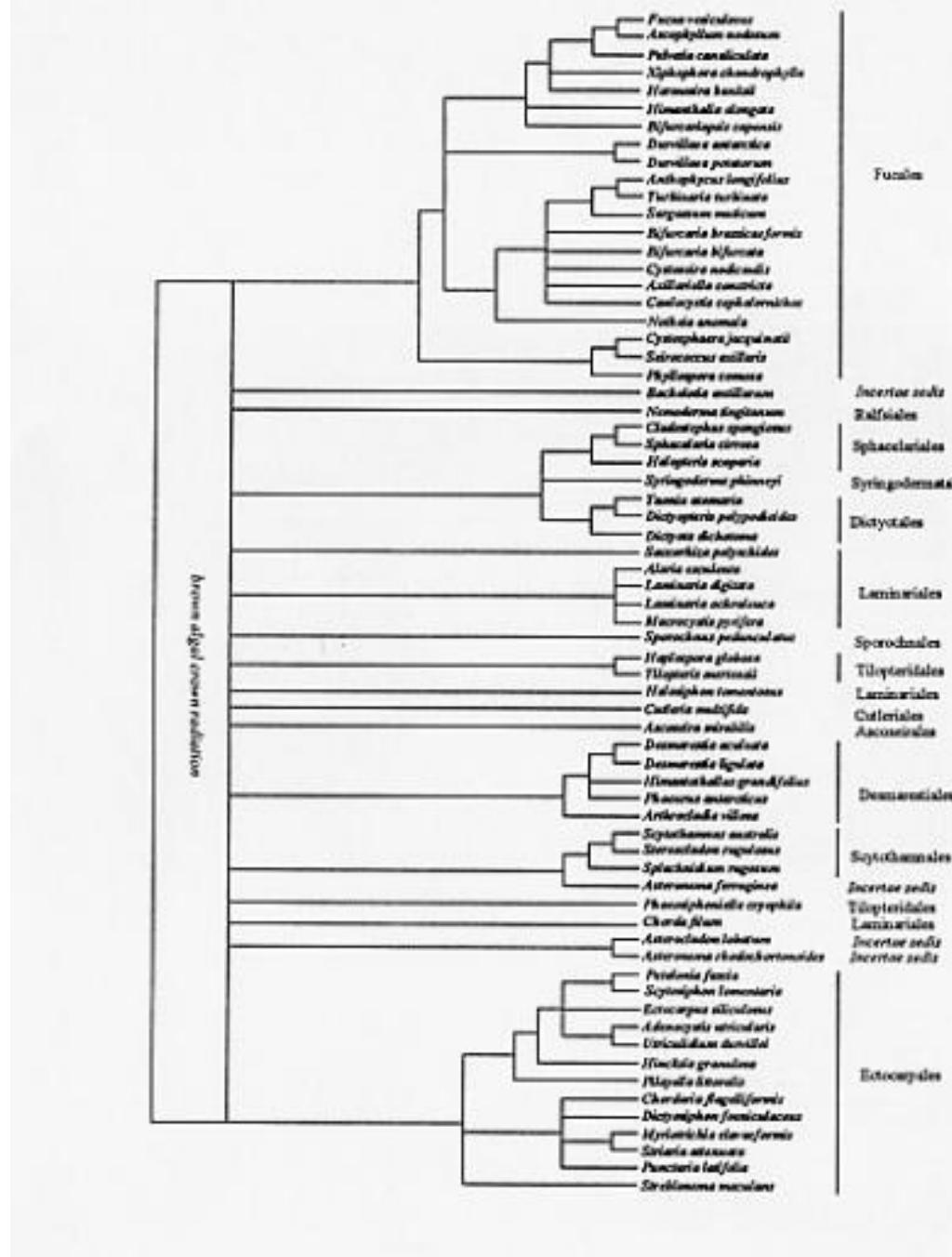


There are multiple rooted trees possible out of a single unrooted tree.
The analyses performed produce unrooted trees.

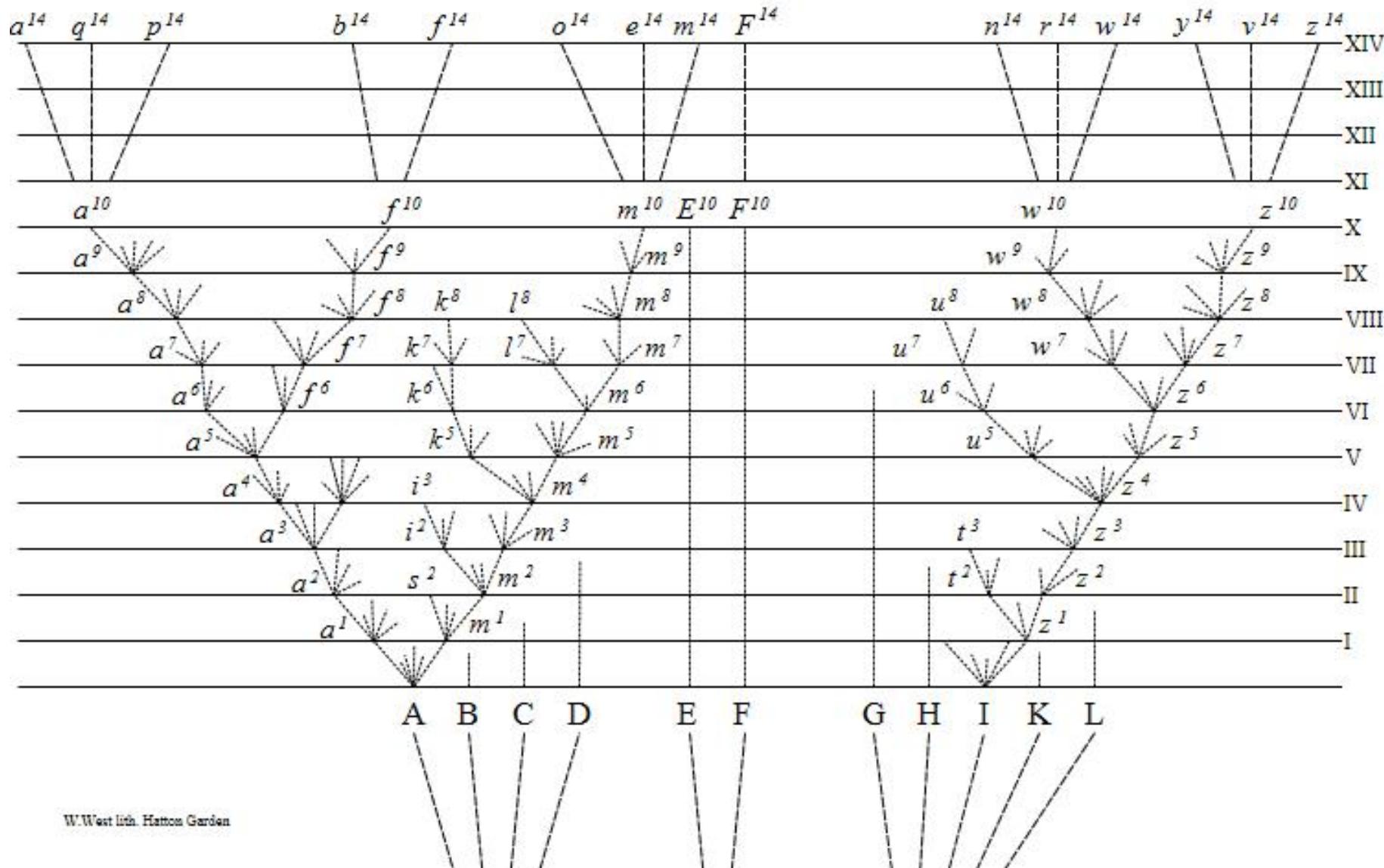
Polytomies



Polytomies

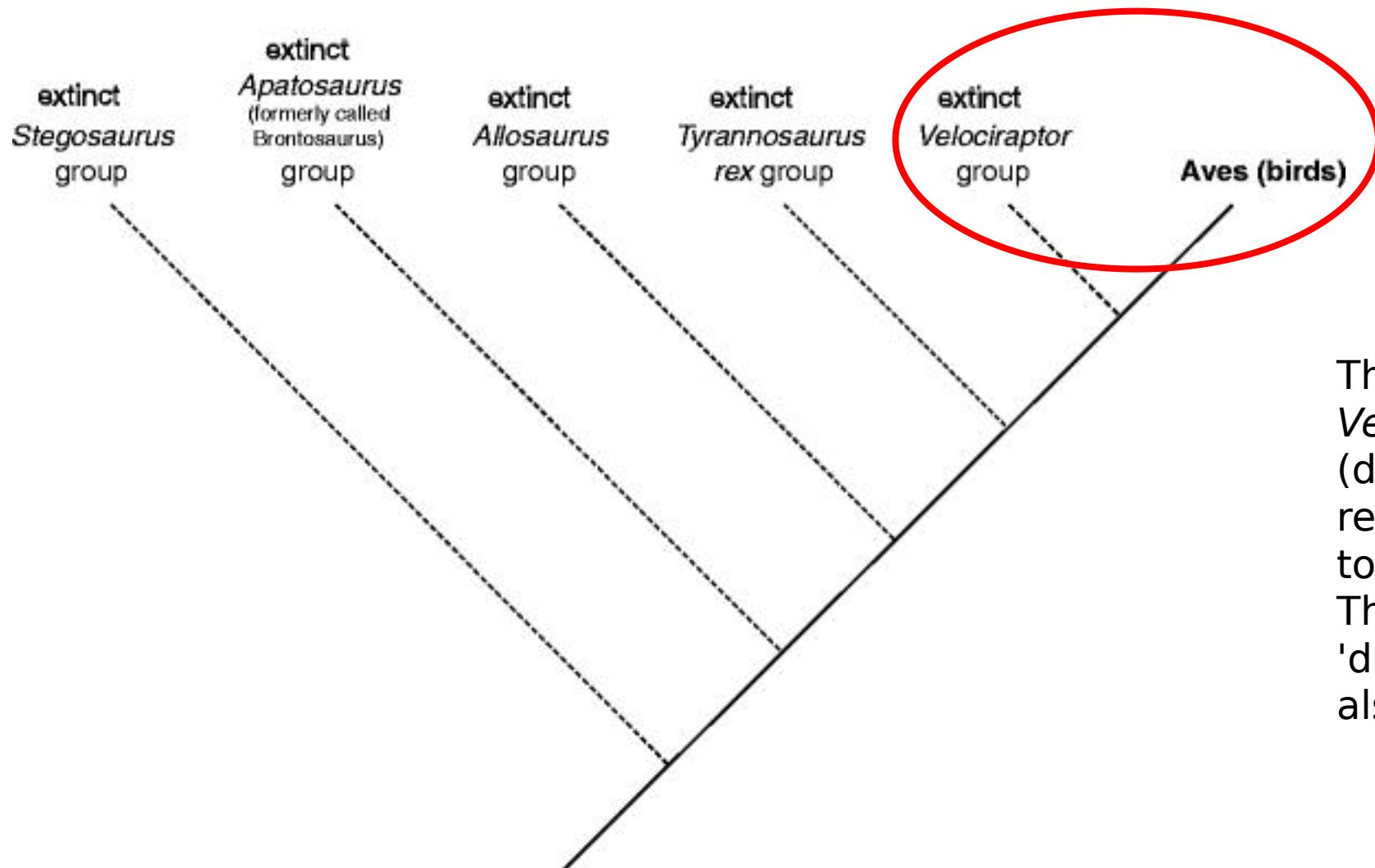


Practical: critical analysis of trees



A simple tree

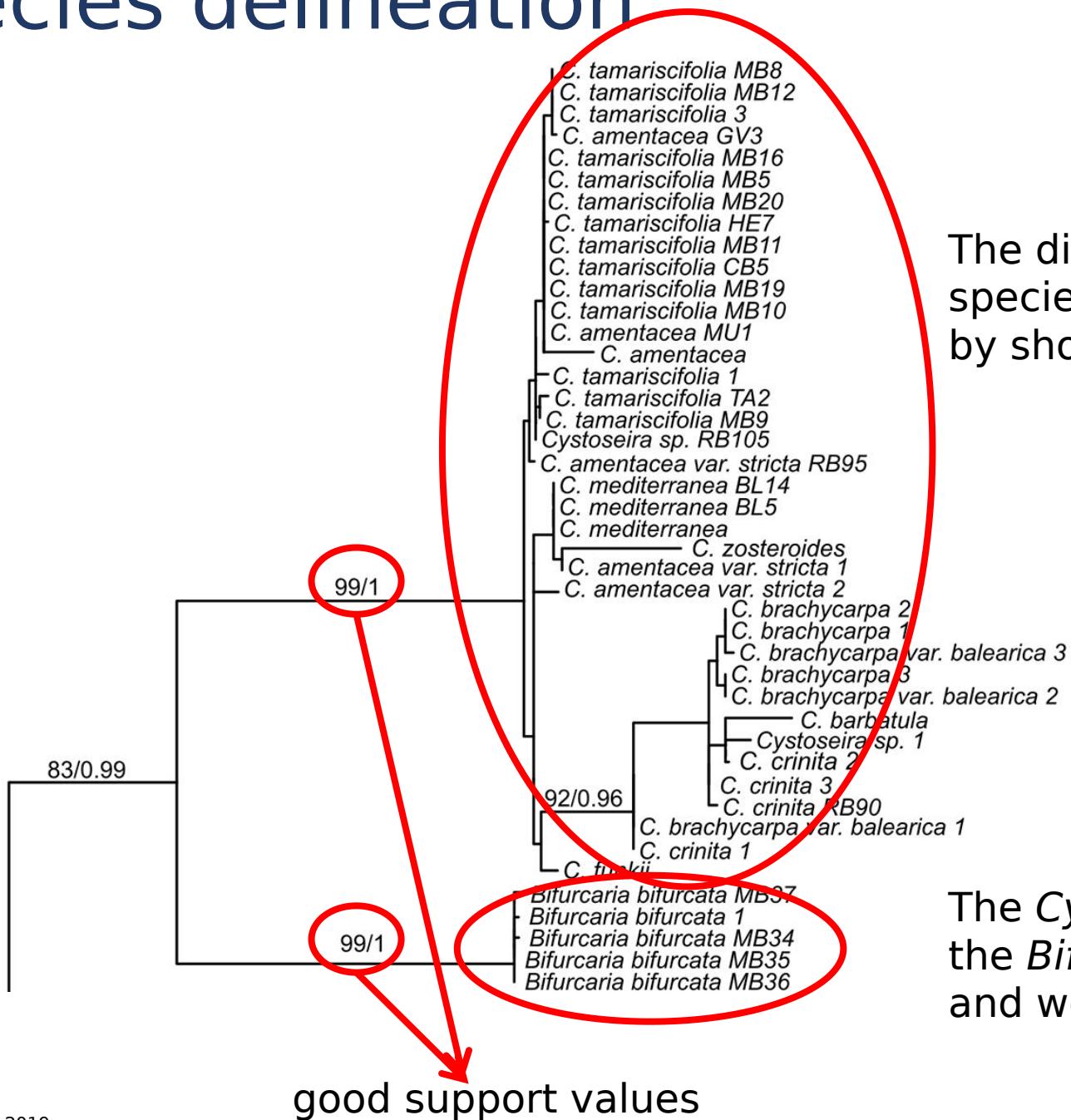
A simplified phylogenetic tree for the dinosaurs



This tree shows that the *Velociraptor* genus (dinosaur) is more closely related to all bird taxa than to the other dinosaur. Therefore, the term 'dinosaur' is only valid if it also includes the birds.

all of the dinosaur lineages are extinct except for Aves (birds)

Species delineation



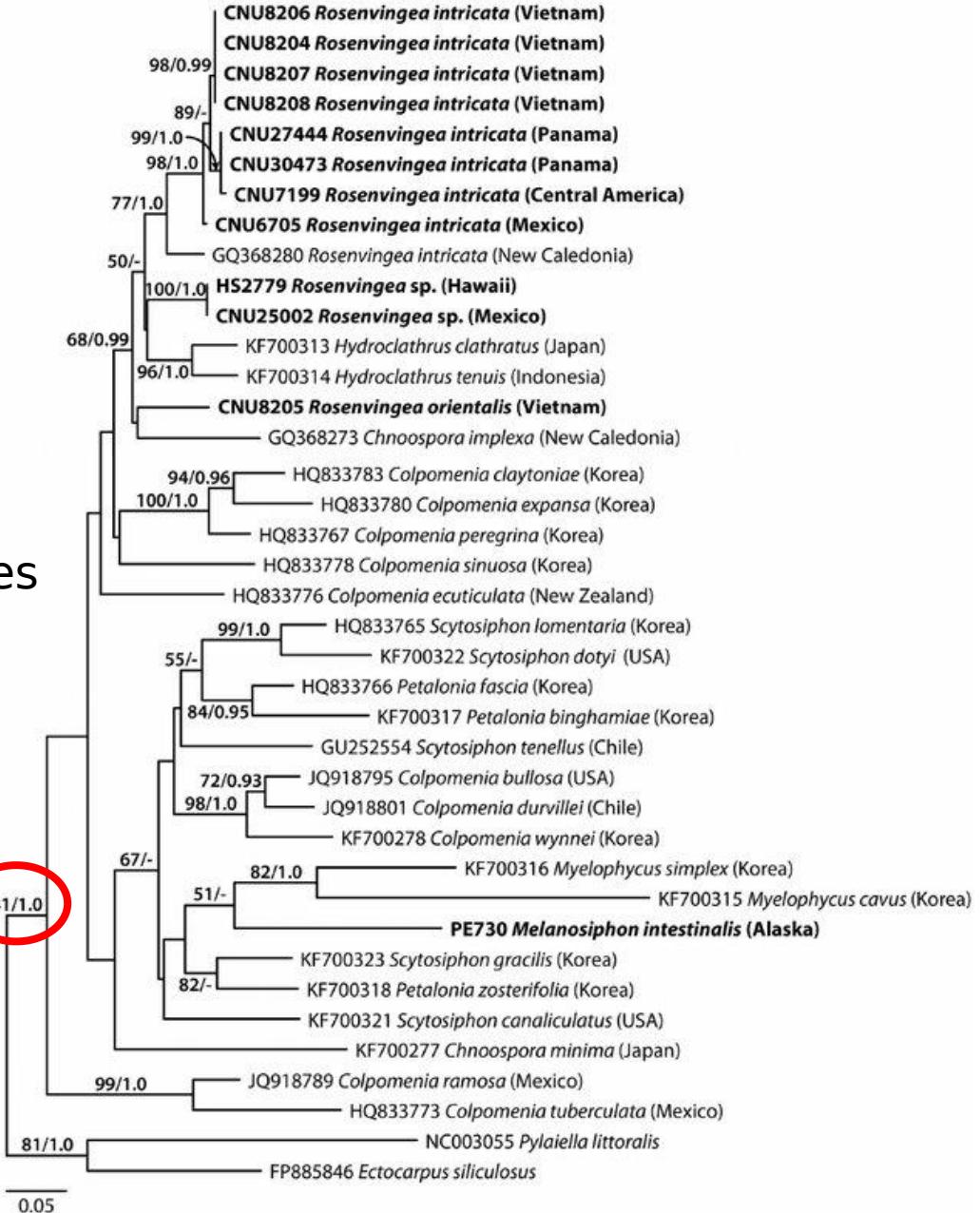
The different *Cystoseira tamariscifolia* species are closely related to each other by short branches.

The *Cystoseira* genus is separated from the *Bifucaria* genus by a long branches and well supported relationship.

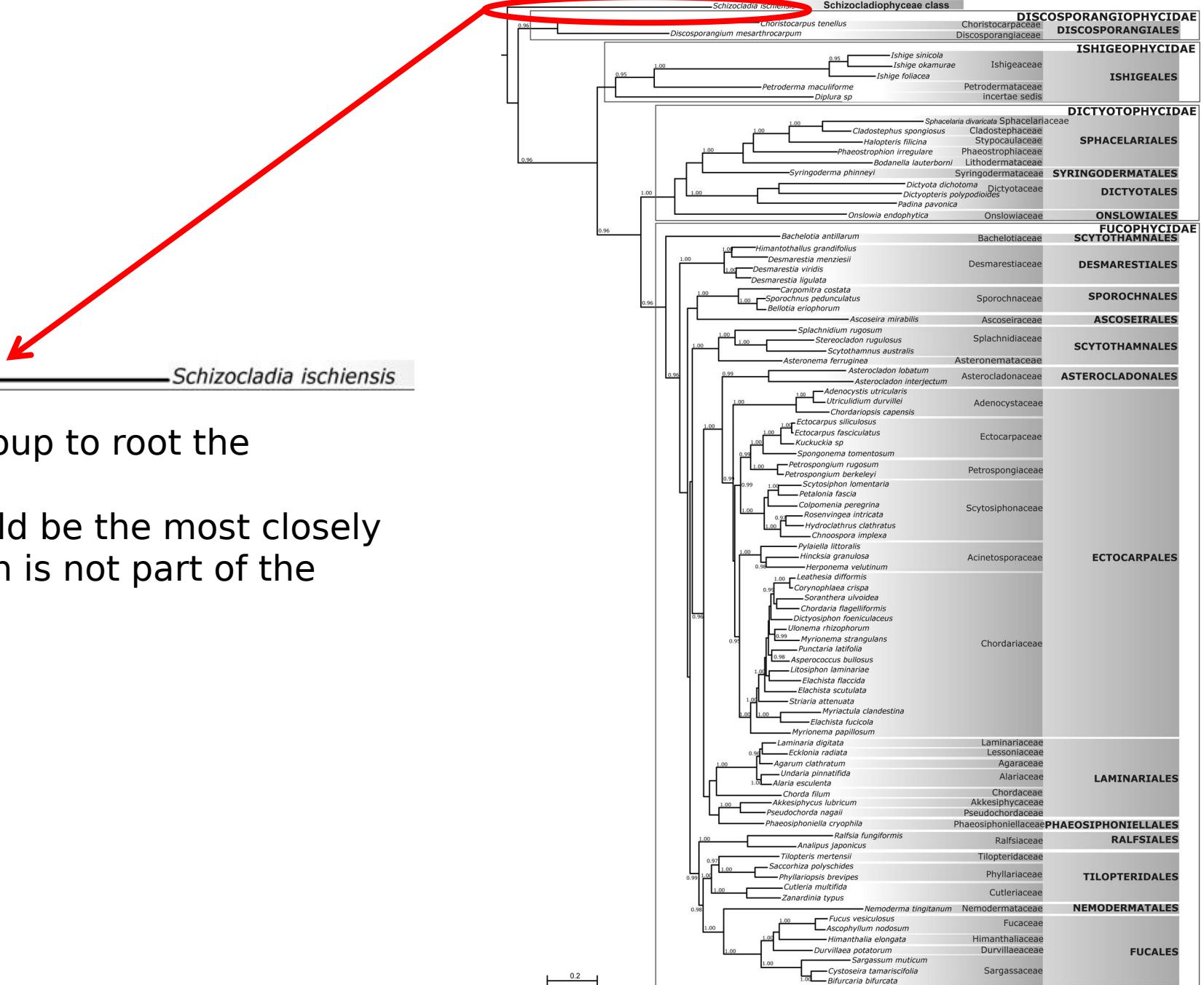
Reliability of the relationships

Maximum-likelihood / Bayesian
bootstrap / posterior probabilities

81/1.0

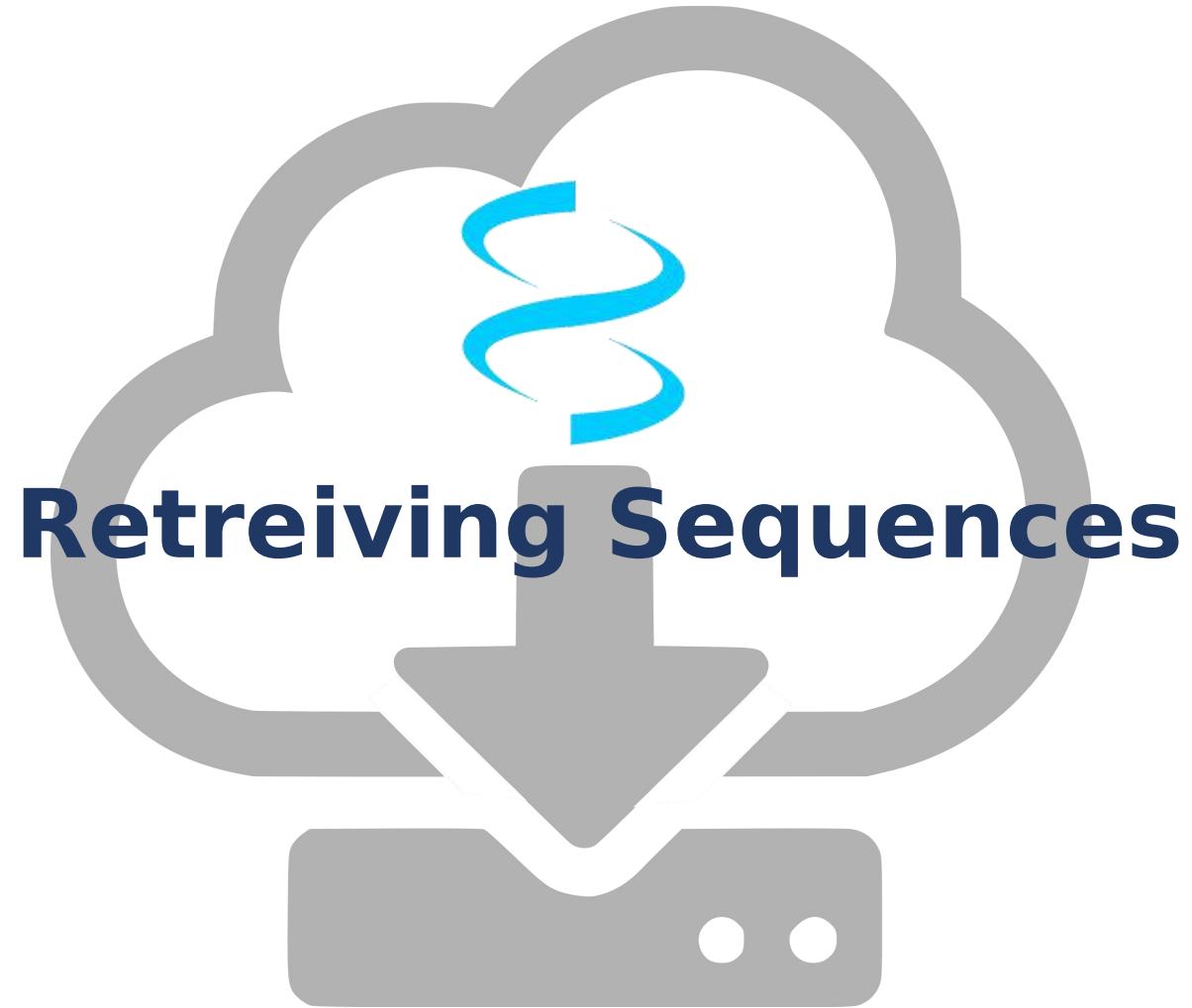


Practise



Choice of an outgroup to root the unrooted tree.

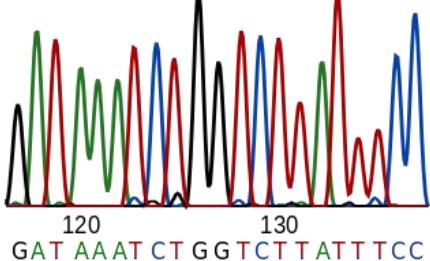
The outgroup should be the most closely related taxon which is not part of the studied group.



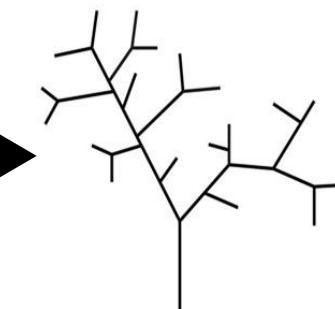
Research workflow for genetic datasets



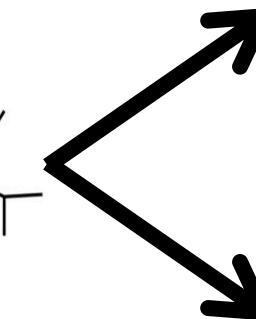
Sampling



Sequencing



Analysing the data



Publishing in a journal



NCBI

EMBL



DDBJ

DNA Data Bank of Japan

Publishing in a database

Genetic databases

<https://www.ncbi.nlm.nih.gov/>

NCBI Resources How To Sign in to NCBI

All Databases Search

NCBI National Center for Biotechnology Information

NCBI Home Resource List (A-Z) All Resources Chemicals & Bioassays Data & Software DNA & RNA Domains & Structures Genes & Expression Genetics & Medicine Genomes & Maps Homology Literature Proteins Sequence Analysis Taxonomy Training & Tutorials Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit Deposit data or manuscripts into NCBI databases 

Download Transfer NCBI data to your computer 

Learn Find help documents, attend a class or watch a tutorial 

Develop Use NCBI APIs and code libraries to build applications 

Analyze Identify an NCBI tool for your data analysis task 

Research Explore NCBI research and collaborative projects 

Popular Resources

PubMed Bookshelf PubMed Central BLAST Nucleotide Genome SNP Gene Protein PubChem

NCBI News & Blog

New search helps you find prokaryotic proteins 30 Sep 2019

The latest improvement in the NCBI search experience is designed to help 30 Sep 2019

Protein BLASTDBs are accession-based 30 Sep 2019

The version 5 BLAST (dbV5) protein databases are now accession-based. You can access these databases and the 26 Sep 2019

Virus hunting in the cloud codeathon, v2 26 Sep 2019

We are pleased to announce the second



Retrieving sequences (1)

<https://www.ncbi.nlm.nih.gov/>

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information

All Databases ▾

- GTR
- HomoloGene
- Identical Protein Groups
- MedGen
- MeSH
- NCBI Web Site
- NLM Catalog
- Nucleotide** 1
- OMIM
- PMC
- PopSet
- Probe
- Protein
- Protein Clusters
- PubChem BioAssay
- PubChem Compound
- PubChem Substance
- PubMed
- SNP
- Sparcle

Search 2

NCBI The National Center for Biotechnology Information advances science and health by providing access to genomic information.

NCBI Mission | Organization | NCBI News & Blog

Submit Transfer manuscripts to NCBI databases

Download Transfer NCBI data to your computer

Learn Find help documents, attend a class or watch a tutorial

Develop Use NCBI APIs and code libraries to build applications

Analyze Identify an NCBI tool for your data analysis task

Research Explore NCBI research and collaborative projects

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

- December 4 Webinar: Human population genetic variation data at NCBI 25 Nov 2019
On Wednesday, December 4, 2019 at 12 PM NCBI staff will present a webinar on
- New release of the Prokaryotic Genome Annotation Pipeline with updated tRNAscan and protein models 20 Nov 2019
A new version of the Prokaryotic Genome

Retreiving sequences (2)

<https://www.ncbi.nlm.nih.gov/nuccore>



The screenshot shows the NCBI Nucleotide search interface. At the top, there is a navigation bar with links for NCBI, Resources, How To, and Sign in to NCBI. Below the navigation bar, the main search interface has "Nucleotide" selected in a dropdown menu. A red circle highlights the "Advanced" link next to a "1" icon, indicating one search result. On the right side of the interface, there is a large DNA sequence visualization and a main content area titled "Nucleotide". The content area contains a brief description of the Nucleotide database and its sources. Below the main content, there are three sections: "Using Nucleotide", "Nucleotide Tools", and "Other Resources", each listing several links.

Using Nucleotide

- [Quick Start Guide](#)
- [FAQ](#)
- [Help](#)
- [GenBank FTP](#)
- [RefSeq FTP](#)

Nucleotide Tools

- [Submit to GenBank](#)
- [LinkOut](#)
- [E-Utilities](#)
- [BLAST](#)
- [Batch Entrez](#)

Other Resources

- [GenBank Home](#)
- [RefSeq Home](#)
- [Gene Home](#)
- [SRA Home](#)
- [INSDC](#)

Retreiving sequences (3)

<https://www.ncbi.nlm.nih.gov/nuccore/advanced>

Nucleotide Advanced Search Builder

((Ralfsiales[Organism]) AND rbcL[Gene Name]) AND Japan

Edit Clear

Builder

1: Organism Show index list
2: Gene Name Show index list
AND
3: All Fields Show index list
AND
 All Fields Show index list

Search or Add to history

History

There is no recent history

1: select the type of information you're looking for.

2: select an operator (AND, OR, NOT) in order to include/exclude the different keywords in the search.

3: click on the 'search' button.

Retreiving sequences (4)

The screenshot shows the NCBI Nucleotide search interface. The search query in the top bar is: ((Raltsiales[Organism]) AND rbcL[Gene Name]) AND Japan. The results page displays 20 items out of 27, with the first item being a 1,409 bp linear DNA sequence from *Neoralfsia expansa* chloroplast, isolate RspG.

Filters (circled 1):

- Species: Protists (27)
- Molecule types: genomic DNA/RNA (27)
- Source databases: INSDC (GenBank) (27)
- Sequence Type: Nucleotide (27)
- Genetic compartments: Chloroplast, Plastid (27)
- Sequence length: Custom range...
- Release date: Custom range...
- Revision date: Custom range...

Results (circled 2):

Items: 1 to 20 of 27

1. *Neoralfsia expansa* chloroplast **rbcL** gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: RspG
1,409 bp linear DNA
Accession: AB250079.1 GI: 189233072
Protein Taxonomy
GenBank FASTA Graphics

2. *Neoralfsia expansa* chloroplast **rbcL** gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: RspE
1,326 bp linear DNA
Accession: AB250077.1 GI: 189233068
Protein Taxonomy
GenBank FASTA Graphics

3. *Raltsia* sp. RspK chloroplast **rbcL** gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds
1,305 bp linear DNA
Accession: AB250083.1 GI: 528313568
Protein Taxonomy
GenBank FASTA Graphics

4. *Raltsia* sp. RspJ chloroplast **rbcL** gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds
1,351 bp linear DNA

Search details:
("Raltsiales"[Organism] AND rbcL[Gene Name]) AND Japan[All Fields] AND (protists[filter] AND biomol_genomic[PROP] AND is_nuccore[filter])

Recent activity: Turn Off Clear

- 1: check the filters (do you have **DNA** sequences from the species group you study (Protists, Plants, ...))
- 2: you can sort your results according to different criteria

Retreiving sequences (5)

The screenshot shows the NCBI Nucleotide search interface. A search query has been entered into the search bar: ((Ralfsiales[Organism]) AND rbcL[Gene Name]) AND Japan. The results pane displays 27 items found, with the first four listed:

1. *Neoralfsia expansa* chloroplast **rbcL** gene for ribulose 1,5-bisphosphate large subunit, partial cds, isolate: RspG
2. *Neoralfsia expansa* chloroplast **rbcL** gene for ribulose 1,5-bisphosphate large subunit, partial cds, isolate: RspE
3. *Ralfsia* sp. RspK chloroplast **rbcL** gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds
4. *Ralfsia* sp. RspJ chloroplast **rbcL** gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds

A modal dialog box is open over the results, titled "Send to:" with "Filters: Manage Filters". It contains the following options:

- Choose Destination:**
 - Complete Record
 - Coding Sequences
 - Gene Features
- Format:** FASTA (selected)
- Sort by:** Default order
- Create File** button (circled with red number 3)

The "Search details" section at the bottom of the dialog box shows the search query: ("Ralfsiales"[Organism] AND rbcL[Gene Name]) AND Japan[All Fields] AND (protists[filter] AND biomol_genomic[PROP] AND is_nuccore[filter]).

1: click on 'send to:'

2: select 'complete record', 'file', 'FASTA' format and the sorting order you want.

3: click on 'create file' and save the file containing the 27 sequences from the search query.

Reading a sequence page (1)

The screenshot shows a search results page for 'Ralfsiales[Organism] AND rbcL[Gene Name] AND Japan'. The results list several entries, with the first one circled in red. The circled entry is: 'Neoralfsia expansa chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: RspG'.

1: click on one of the sequence.

2: the sequence name you've clicked on appears here.

3: unique ID number to find back the sequence (should figure in your article).

4: organism name and its classification (may be outdated).

5: reference of the article in which the author used the sequence.

The screenshot shows the detailed sequence page for the entry circled in the previous screenshot. The page includes the sequence name, GenBank ID (AB250079.1), and links to FASTA and Graphics. Below this, a table provides detailed information about the sequence:

LOCUS	AB250079	1409 bp	DNA	linear	PLN 25-SEP-2013
DEFINITION	Neoralfsia expansa chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: RspG.				
ACCESSION	3 AB250079				
VERSION	AB250079.1				
KEYWORDS					
SOURCE	chloroplast Neoralfsia expansa				
ORGANISM	Neoralfsia expansa				
4	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ralfsiales, Neoralfsiaceae; Neoralfsia.				
REFERENCE	1				
AUTHORS	Lim,P., Sakaguchi,M., Hayunda,T., Kogame,K., Thang,S. and Kawai,H.				
TITLE	Molecular phylogeny of crustose brown seaweeds (Ralfsiales, Phaeophyceae) inferred from rbcL sequences resulting in proposal for Neoralfsiaceae fam. nov				
JOURNAL	Phycologia 46, 456-466 (2007)				
REFERENCE	2	(bases 1 to 1409)			
AUTHORS	Lim,P.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-FEB-2006) Contact:Phaik-Eem Lim Kobe University Research Center for Inland Seas, Department of Science; Rokkodai1-1, Kobe, Hyogo Prefecture 657-8501, Japan				
FEATURES	Location/Qualifiers				
source	1..1409				

Reading a sequence page (2)

NCBI Resources How To

Nucleotide Nucleotide Advanced

GenBank

Neoralfsia expansa chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: RspG

GenBank: AB250079.1

FASTA Graphics

Go to:

LOCUS AB250079 1409 bp DNA linear PLN 25-SEP-2013
DEFINITION Neoralfsia expansa chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: RspG.
ACCESSION AB250079
VERSION AB250079.1

1: click on FASTA.

2: the sequence name you've clicked on appears here.

3: sequence name in fasta format starts with a '>' sign.

Under it, the DNA sequence is displayed.

4: you can download the single sequence from here using 'send to' and proceeding as previously shown.

2

Neoralfsia expansa chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: RspG

GenBank: AB250079.1

GenBank Graphics

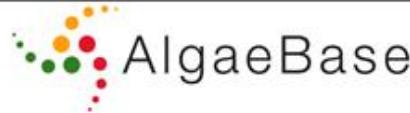
3

>AB250079.1 Neoralfsia expansa chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: RspG

TAATTCCATATGCTAAATCCCATACTGGGATGCTGATTATAATGTTAAAGCTACTCATTTCTACCTCT
TTTCCGCATAACTCCGCAACGGGTGTTGAGGGCGCTGCTGCTGTTGCTGGTGAGTCTTCC
ACTGCAACATGGACTGTAGTTGGACGGATTATAACAGCTGTGACGCTCACCGAGCTAAAGCTTATC
GTGTTGATCCAGTCCAGGCACAAACGATCAATACTTGTCTACATAGCTTACGAATGTGATTATTGA
AGAGGGTTCTTAGCCAACTTAACGGCTTCATTATTGGTAATGTTTTGGTTAAAGCAGTAAAAGCA
TTACGTTAGAAGATATGAGAATTCTCTTGCTTATCTGAAAACGTTCCAAGGTCCTGCTACAGGTGAA
TTGAGAAAAGAGAACGACTAGATAAAATTGGTCGCTTTATTAGGTGCAACTGTAAAACCTAAGTTAGG
ACTTTCTGGAAAAACTATGGCTGTTGTACGAAGGATTATGTTGGTCTGACTTCTAAAGAT
GACGAGAAATTAAACTCACAACCATTATCGCCTGGAAAGAGCGTTCTTATACTGTATGGAAGGGTTA
ACCGTGCCTGAGCTGCACAGGTGAAGTTAAAGGTTCTTATCTAAATGTTACTGCTTCTACAATTGAACA
AATGTATGAGCGTGCTGACTACGCACATGATATTGGTAGTGTAAATTGTAATGATCGACTTAGTTGGT
TATACAGCTATTCAAACATGGCAATTGGCAAGAAAAGCTCAAATGATTTACATTACACCGTCTG
GTAACCTACTTATGCACGTCAAGAAAAACACGGTATTAACTTACAGGTTATCTGAAATGGATCGTAT
GTGTTGGCGTAGATWCATTATGCAGGTACAGTTGAGTAAACTGGAGGGTGTACCTTACATTAAATGGTAAAAA
GGTTTCTATAATACATTACTATTAAACTCAGTTAAAATTAAATTAGCTGAAGGTATATTTTCGATATGG
ACTGGGCATCTTCTAGAAAATGTGTACCGGTAGCTTGGGTTGAATCCATTGTTGGTCAAATGCATCAA
CTCTTATTACTTAGGTGATGTAGTTACAATTGGGGGCGGTACGATTGGTCAACCGAGATGGTATT
CAATCAGGTGCAACTGCTAACCGTGTGCACTAGAAGCAATGGTATTAGCTCGTAACGAAGGTCGTGATT
ACGTGGCAGAAGGACCTGAAATTCTACGTAATGCTGCAAGTACTGTGGACCTCTAAAGCTGTTAGA
TTTATGGAAAGATATCACCTTCGATTATACTTCAACTGATACACCTGACTTCGTTGAAGTTGCAACTGAA
AGCAAATAA

4

Send to:



content about team notulae algarum links contact search

genus · species · literature · journals · images · common names · distribution · glossary · taxonomy browser · higher taxonomy

156,436 species and infraspecific names are in the database, 22,021 images, 60,452 bibliographic items, 447,756 distributional records.

Ascophyllum nodosum (Linnaeus) Le Jolis

Publication details

Ascophyllum nodosum (Linnaeus) Le Jolis 1863: 96

Published in: Le Jolis, A. (1863). Liste des algues marines de Cherbourg. *Mémoires de la Société Impériale des Sciences Naturelles de Cherbourg* 10: 5-168, pls I-IV.

Download PDF

Original description: Download PDF

Type species

The type species (holotype) of the genus *Ascophyllum* is *Ascophyllum laevigata* Stackhouse.

Status of name

This name is of an entity that is currently accepted taxonomically.

Basionym

Fucus nodosus Linnaeus



Ascophyllum nodosum (Linnaeus) Le Jolis

Cuan na Beirtri Bhui, Conamara, Co. na Gaillimhe,

Ireland; yellow bands of intertidal beds

© Michael Guiry (mike.guiry@nuigalway.ie)

Classification:

Empire Eukaryota
Kingdom Chromista
Phylum Ochrophyta
Class Phaeophyceae
Subclass Fucophycidae
Order Fucales
Family Fucaceae
Genus *Ascophyllum*

Uses and Compounds

Taxonomy

References

Submit Feedback

Submit Reference

Links

Genbank

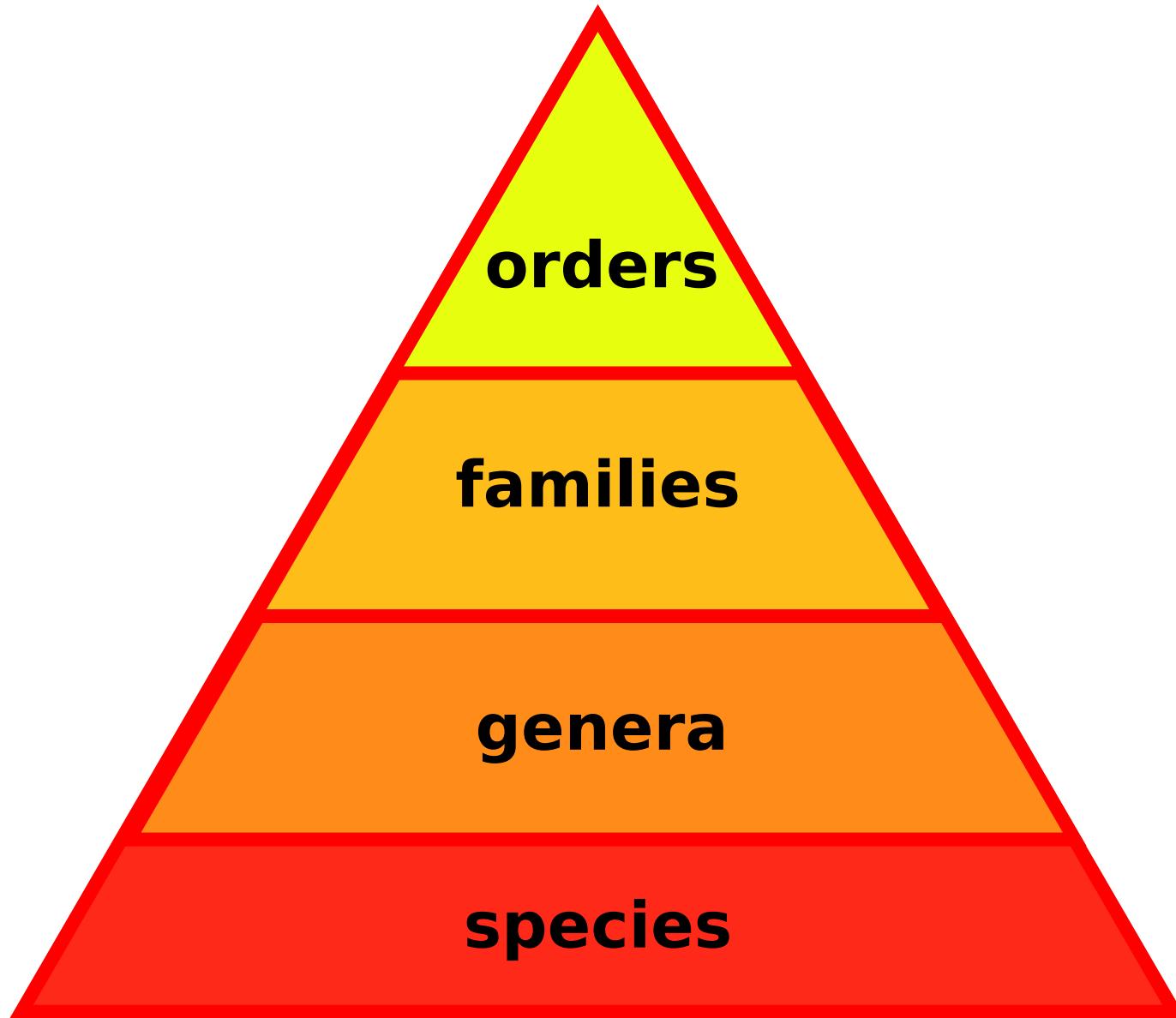
Index Nominum Algarum

Google

Biodiversity Heritage Library

Pictures

Reminder of classification



Ralfsiales



Ralfsiaceae

Ralfsia

Ralfsia verrucosa (Areschoug) Areschoug

Presentation of the dataset from Lim et al. (2007)

Phycologia (2007) Volume 46 (4), 456–466

Published 5 July 2007

Molecular phylogeny of crustose brown algae (Ralfsiales, Phaeophyceae) inferred from *rbcL* sequences resulting in the proposal for Neoralfsiaceae fam. nov.

PHAIK-EEM LIM^{1*}, MOTOHIRO SAKAGUCHI¹, TAKEAKI HANYUDA¹, KAZUHIRO KOGAME², SIEW-MOI PHANG³ AND HIROSHI KAWAI¹

¹Kobe University Research Center for Inland Seas, Rokkodai, Kobe 657-8501, Japan

²Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

³Institute of Biological Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia

P.-E. LIM, M. SAKAGUCHI, T. HANYUDA, K. KOGAME, S.-M. PHANG AND H. KAWAI. 2007. Molecular phylogeny of crustose brown algae (Ralfsiales, Phaeophyceae) inferred from *rbcL* sequences resulting in the proposal for Neoralfsiaceae fam. nov. *Phycologia* 46: 456–466. DOI: 10.2116/06-90.1

The order Ralfsiales was established to accommodate the brown algal taxa having a crustose thallus, an isomorphic life history, discoid early development of the thallus and containing a single, plate-shaped chloroplast without pyrenoids in each cell. However, the validity of the order has been questioned by many researchers because several exceptions to these criteria have been found within the order. Molecular phylogenetic analysis of the taxa assigned to the order, using *rbcL* DNA sequences, reveals that Ralfsiales is not a monophyletic group but is separated into two major groups, excluding Lithodermataceae, which were not included in the present analysis: clade I, comprising the members of Ralfsiaceae, Mesosporaceae, *Analipus japonicus* and *Heteroralfsia saxicola*; and clade II, consisting of *Diplura* species, sister to the Ishigeales clade. On the basis of these results, we propose emendment of the Ralfsiales to contain only species having (1) discoidal early development of the thallus; (2) intercalary plurilocular gametangia with terminal cells and terminal unilocular zoidangia; and (3) a crustose phase in the life history. Furthermore, we propose the establishment of the new family Neoralfsiaceae to accommodate the new genus *Neoralfsia*, on the basis of *Ralfsia expansa*.

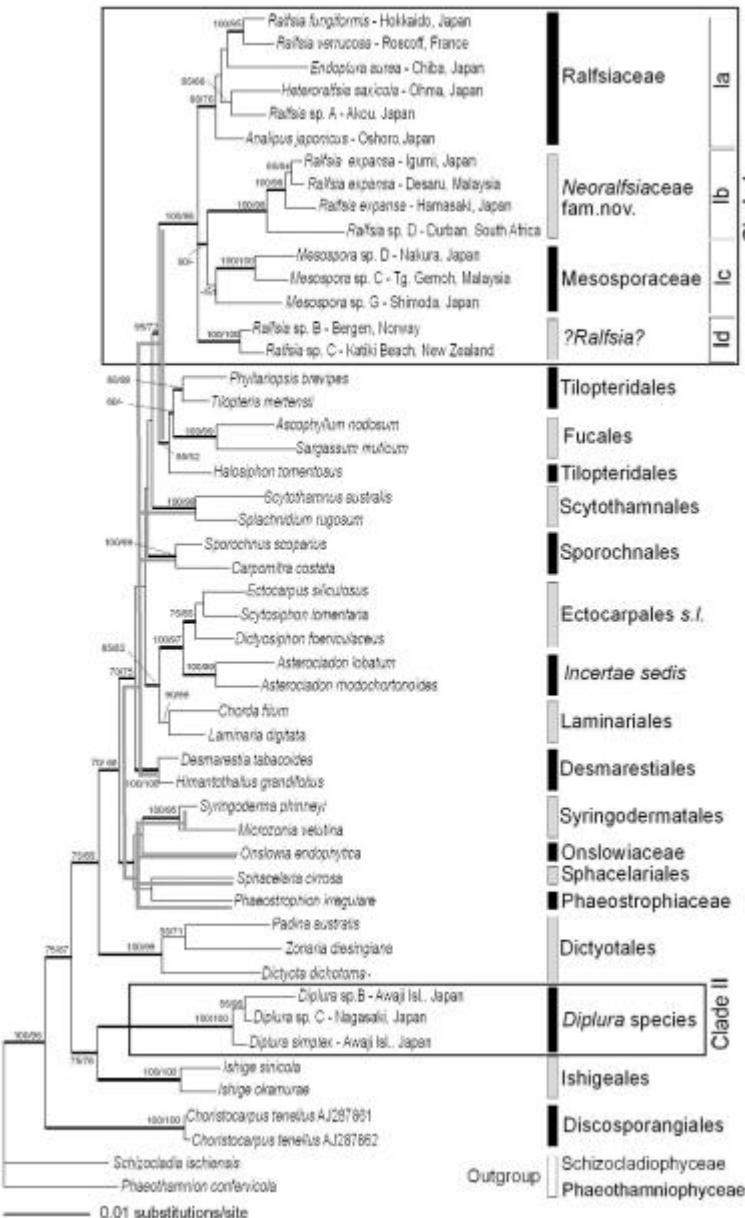
KEY WORDS: Molecular phylogeny, *Neoralfsia*, Neoralfsiaceae, Ralfsiaceae, Ralfsiales, *rbcL*, Taxonomy

Presentation of the dataset from Lim2007

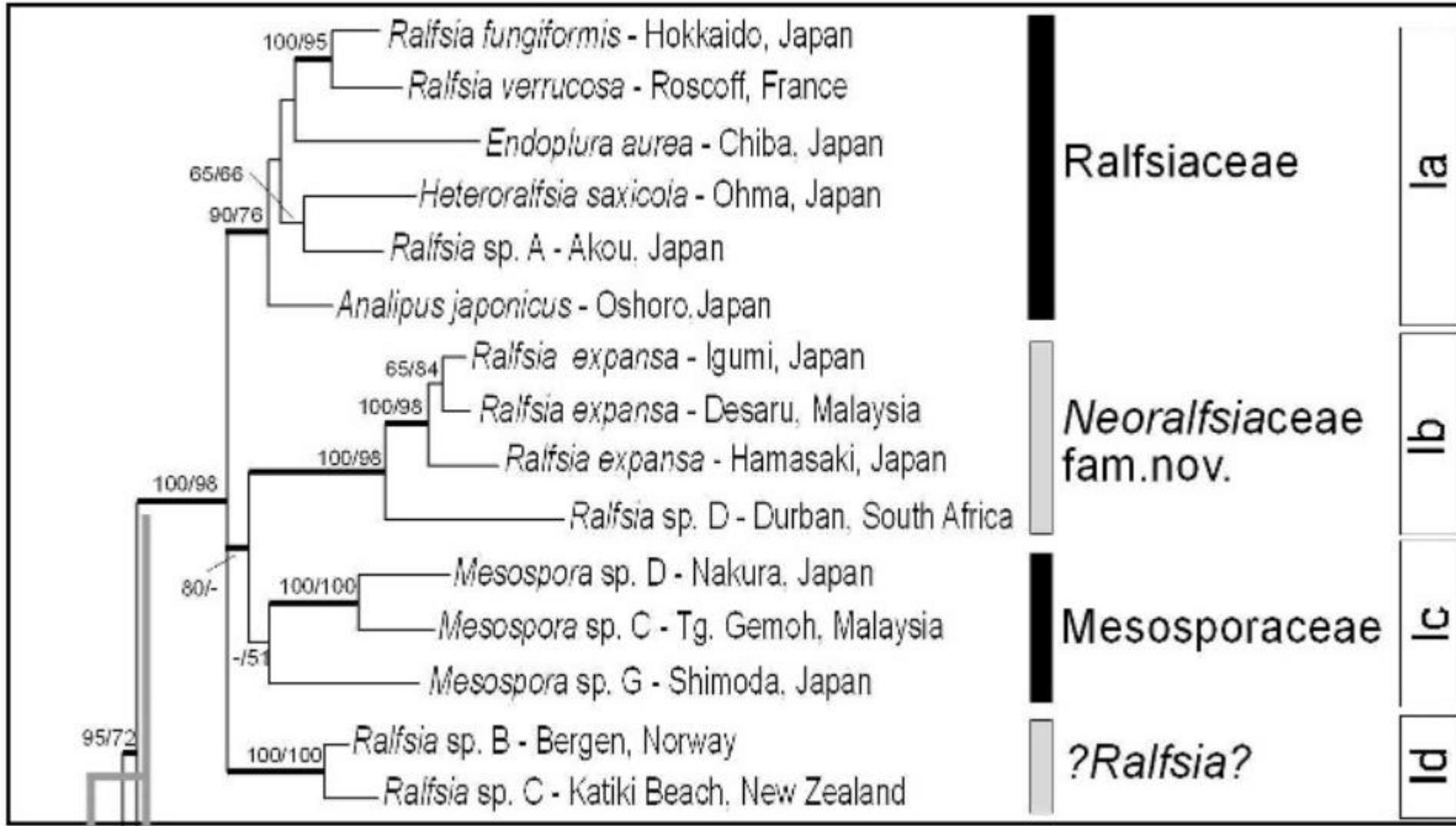
Name of taxa	Origin (published year/collector/specimen no.)	DDBJ accession number for <i>rbcL</i>
<i>Ralfsiaceae</i>		
<i>Analipus japonicus</i> (Harvey) Wynne	Oshoro, Hokkaido Pref., Japan (H. Kawai) (KU-883)	AB264042
<i>Heteroralfsia saxicola</i> (Okamura et Yamada) Kawai	Ohma, Aomori Pref., Japan (H. Kawai) (KU-882)	AB250070
<i>Endophura aurea</i> Hollenberg	Inubouzaki, Chiba Pref., Japan (K. Kogame) (KU-d2273)	AB264039
<i>Ralfsia fungiformis</i> (Gunnerus) Setchell et Gardner	Akkeshi, Hokkaido Pref., Japan (K. Kogame) (KU-d2206)	AB250071
<i>Ralfsia verrucosa</i> (Areschoug) J. Agardh	Roscoff, Brittany, France (H. Kawai) (KU-d2305)	AB250072
<i>Ralfsia</i> sp. A	Akou, Hyogo Pref., Japan (S. Uwai) (KU-d2259)	AB250073
<i>Ralfsia</i> sp. B	Bergen, Norway (H. Kawai) (KU-d2315)	AB250074
<i>Ralfsia</i> sp. C	Katiki Beach, New Zealand (H. Kawai) (KU-d2201)	AB250075
<i>Ralfsia</i> sp. D (<i>Ralfsia expansa</i> related species)	Durban, South Africa (H. Kawai) (KU-2317)	AB250076
<i>Ralfsia expansa</i> (J. Agardh) J. Agardh	Hamasaki, Ishigaki Isl., Okinawa Pref., Japan (P.-E. Lim) (KU-d2132)	AB250077
<i>Ralfsia expansa</i> (J. Agardh) J. Agardh	Desaru, Johor, Malaysia (P.-E. Lim) (KU-d2317)	AB250078
<i>Ralfsia expansa</i> (J. Agardh) J. Agardh	Igumi, Hyogo Pref., Japan (H.Uchida) (KU-d2243)	AB250079
<i>Diplura simplex</i> Tanaka et Chihara	Maruyama, Hyogo Pref., Japan (P.-E. Lim) (KU-d2582)	AB250084
<i>Diplura</i> sp. B	Maruyama, Hyogo Pref., Japan (P.-E. Lim) (KU-d2574)	AB250086
<i>Diplura</i> sp. C	Teguma, Nagasaki Pref., Japan (A. Tanaka) (KU-d2247)	AB250087

Presentation of the dataset from Lim2007 (tree)

462 *Phycologia*, Vol. 46 (4), 2007



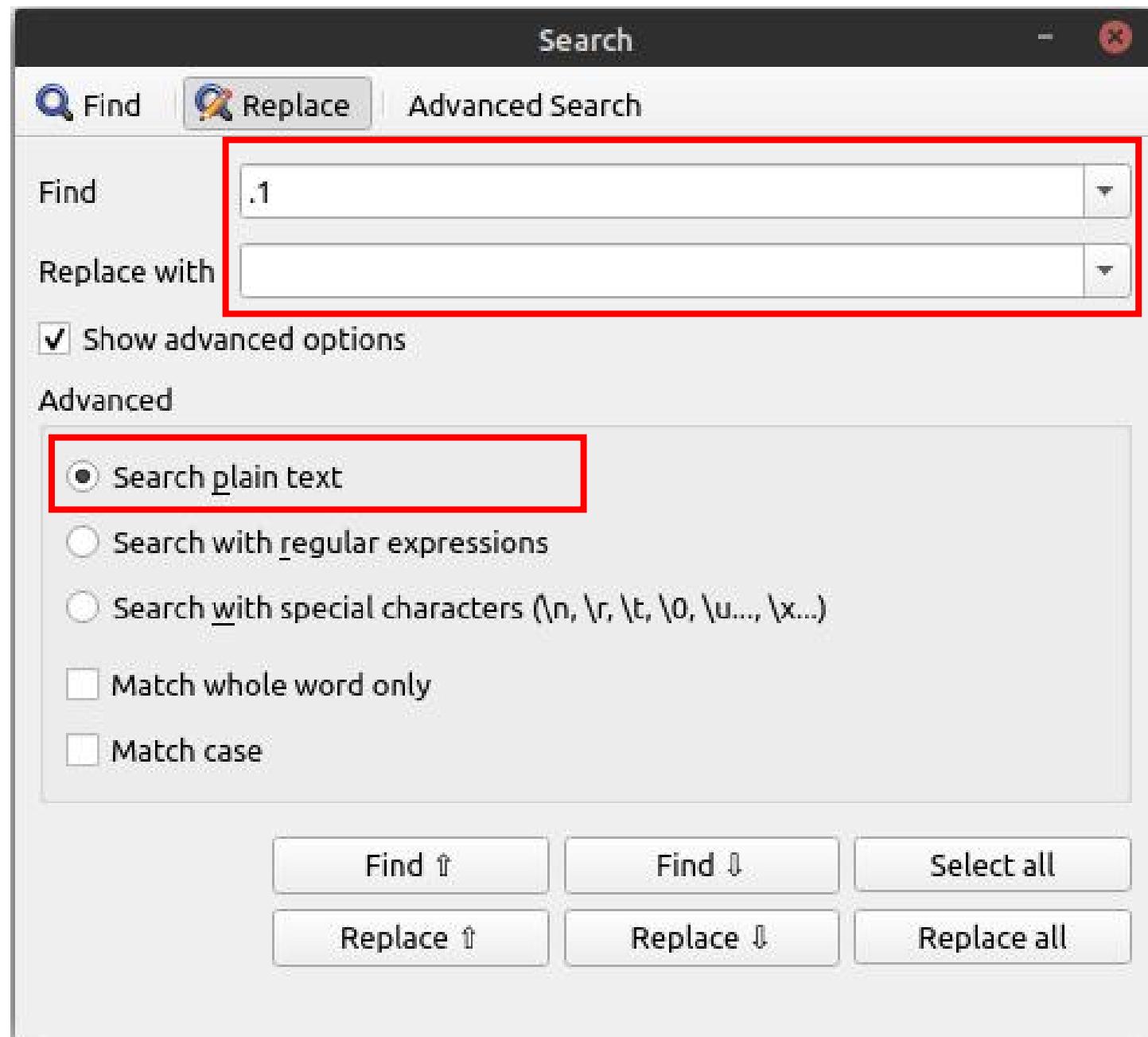
Presentation of the tree we will do



Cleaning Sequences

```
>Sequence_1 assembly1  
CCCTAAACCCCTAAACCCCTAAACCCCTAAACCTCTGA  
CTTAAATCCTACATCCATGAATCCCTAAATACCT.  
CTCTGGTTGAAAATCATTGTGTATAATGATAAT  
TTGTTGTGTAGATTTTAATAATCATTTGA  
GGTTTCCTTCCTTCACTTAAGGATGGTTA  
GCTTGCTACGATCTACATTGATAATGTGAGTCT  
ATCTCAAGAACATTATAATTGGACTGTTA  
CCCTAAACCCCTAAACCCCTAAACCCCTAAACCTCTGA  
CTTAAATCCTACATCCATGAATCCCTAAATACCT.  
CTCTGGTTGAAAATCATTGTGTATAATGATAAT  
TTGTTGTGTAGATTTTAAAAATATCATTTGA  
GGTTTCCTTCCTTCACTTAGCTATGGATGGTTA  
GCTTGCTACGATCTACATTGGAAATGTGAGTCT  
ATCTCAAGAACATTATAATTGGACTGTTA
```

Cleaning the sequence file (1)



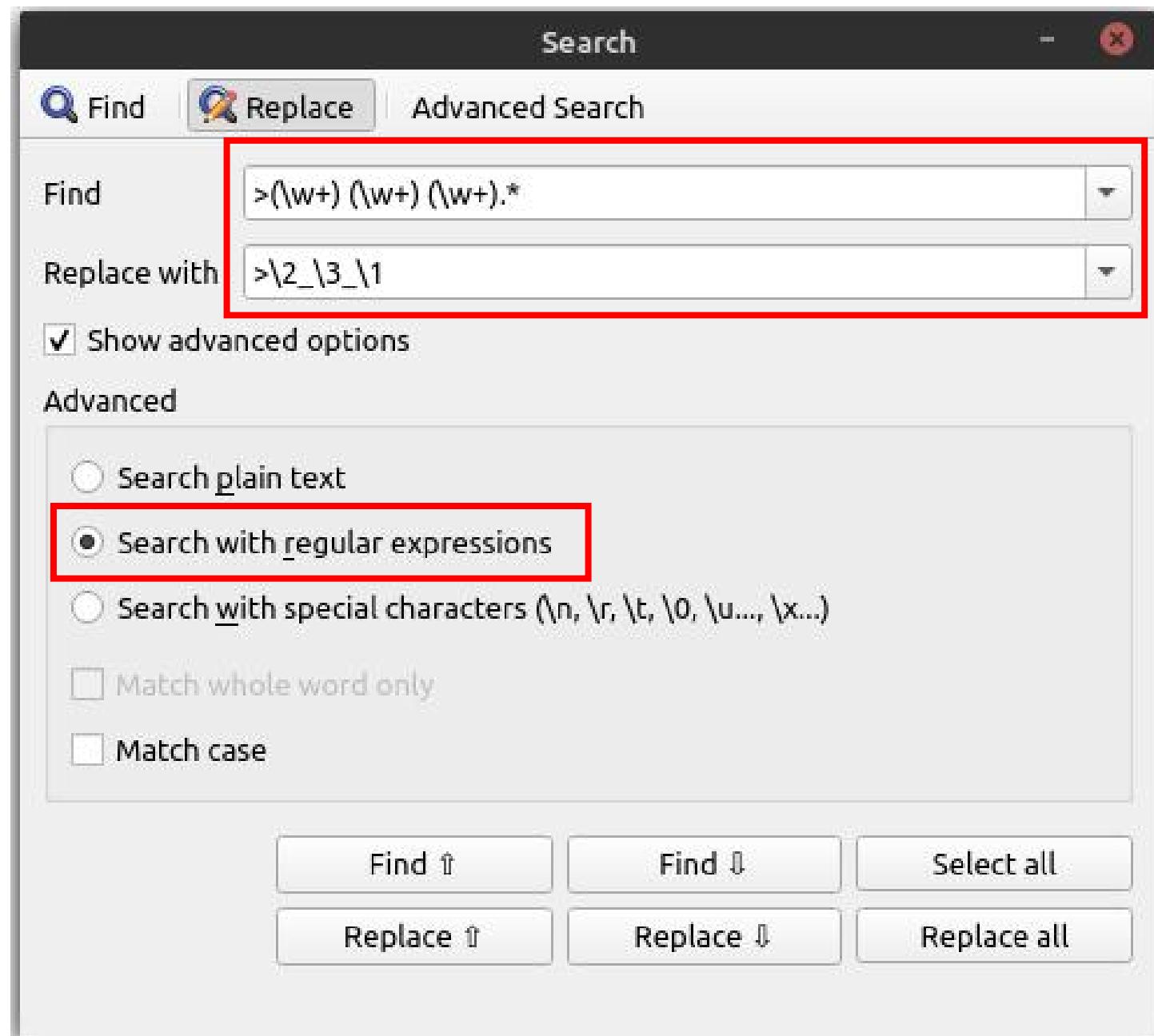
Open the sequence.fasta file previously downloaded with notepad++, sublimetext, ...

Click on **CTRL + H** to open the replace window

Fill in the fields as shown on the screenshot and click on replace all.

This takes off all the .1 at the end of the accession numbers (version of the sequence, not relevant)

Cleaning the sequence file (2)



Click on **CTRL + H** to open the replace window

Fill in the fields as shown on the screenshot and click on replace all.

Takes only the first three 'words' (\w+) and put them in a different order separated by '_'

Normalise the names to this format:

>Genus_species_GenBankAccessionNumber

Sort file for unique haplotypes

<http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi?cmd=cd-hit>

CD-HIT Suite: Biological Sequence Clustering and Comparison

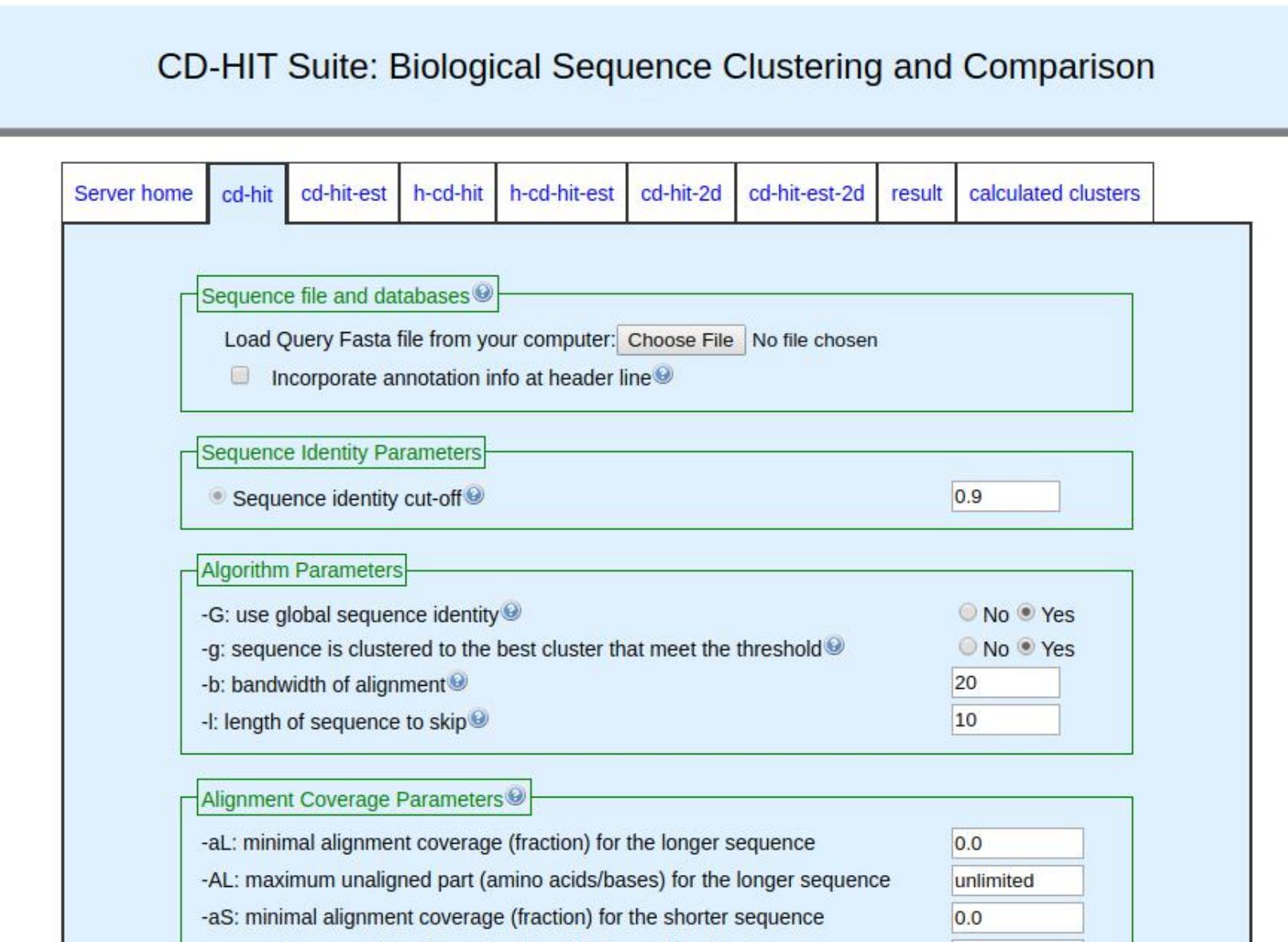
Server home cd-hit cd-hit-est h-cd-hit h-cd-hit-est cd-hit-2d cd-hit-est-2d result calculated clusters

Sequence file and databases
Load Query Fasta file from your computer: Choose File No file chosen
 Incorporate annotation info at header line

Sequence Identity Parameters
 Sequence identity cut-off 0.9

Algorithm Parameters
-G: use global sequence identity
-g: sequence is clustered to the best cluster that meet the threshold
-b: bandwidth of alignment
-l: length of sequence to skip
No Yes
No Yes
20
10

Alignment Coverage Parameters
-aL: minimal alignment coverage (fraction) for the longer sequence 0.0
-AL: maximum unaligned part (amino acids/bases) for the longer sequence unlimited
-aS: minimal alignment coverage (fraction) for the shorter sequence 0.0



CD-HIT can be used for sorting datafiles with redundant sequences.

If you set the Sequence identity cut-off to 1, the resulting file will only contain unique sequences. When two sequences are 100% identical and that they differ in length, the algorithm keeps the longest sequence.



Blast

Basic Local Alignment Search Tool

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

The screenshot shows the NCBI BLAST homepage. At the top, there's a navigation bar with the NIH logo, U.S. National Library of Medicine, NCBI National Center for Biotechnology Information, and a 'Sign in to NCBI' link. Below the navigation bar, the 'BLAST®' logo is on the left, and 'Home', 'Recent Results', 'Saved Strategies', and 'Help' links are on the right. A teal vertical bar on the left contains the word 'NEWS'. The main content area features a section titled 'Basic Local Alignment Search Tool' with a brief description of what BLAST does. To the right, there's a news box with a teal header containing the text 'End of updates for BLAST+ version 4 databases (dbV4)', followed by 'Start moving to the new version 5 databases!', the date 'Fri, 27 Sep 2019 16:00:00 EST', and a 'More BLAST news...' link. Below this, there are three large buttons for 'Nucleotide BLAST', 'blastx', and 'tblastn', each with a corresponding diagram. There's also a 'Protein BLAST' button. At the bottom, there's a 'BLAST Genomes' link.

U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

Sign in to NCBI

BLAST®

Home Recent Results Saved Strategies Help

NEWS

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

End of updates for BLAST+ version 4 databases (dbV4)

Start moving to the new version 5 databases!

Fri, 27 Sep 2019 16:00:00 EST

[More BLAST news...](#)

Web BLAST

Nucleotide BLAST

nucleotide ► nucleotide

blastx

translated nucleotide ► protein

tblastn

protein ► translated nucleotide

Protein BLAST

protein ► protein

BLAST Genomes

Blastn (1) = Blast of nucleotide sequences on nucleotide sequences

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

The screenshot shows the NCBI BLAST homepage. At the top, there's a navigation bar with the NIH logo, U.S. National Library of Medicine, NCBI National Center for Biotechnology Information, and a 'Sign in to NCBI' link. Below the navigation is the BLAST logo and links for Home, Recent Results, Saved Strategies, and Help. A teal vertical bar on the left says 'NEWS'. In the center, there's a section titled 'Basic Local Alignment Search Tool' with a description of what BLAST does and a 'Learn more' link. To the right, there's a news box about the end of updates for BLAST+ version 4 databases (dbV4), with a date (Fri, 27 Sep 2019 16:00:00 EST) and a 'More BLAST news...' link. Below this are three main search tool sections: 'Nucleotide BLAST' (highlighted with a red border), 'blastx' (translated nucleotide ▶ protein), and 'tblastn' (protein ▶ translated nucleotide). To the right is a 'Protein BLAST' section (protein ▶ protein). At the bottom, there's a 'BLAST Genomes' link.

U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

Sign in to NCBI

BLAST®

Home Recent Results Saved Strategies Help

NEWS

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

End of updates for BLAST+ version 4 databases (dbV4)

Start moving to the new version 5 databases!

Fri, 27 Sep 2019 16:00:00 EST

[More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide

Protein BLAST
protein ▶ protein

BLAST Genomes

Blastn (2)

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

The screenshot shows the BLAST Standard Nucleotide BLAST search interface. Key fields highlighted with red circles and numbers are:

- 1: Enter Query Sequence (text input field for sequence entry).
- 2: Database (radio button selection for Standard databases (nr etc.)).
- 3: Organism (optional text input field for organism name or ID).
- 4: BLAST (button to initiate the search).

Other visible fields include: Enter accession number(s), gi(s), or FASTA sequence(s); Or, upload file; Job Title; Align two or more sequences; Choose Search Set; Entrez Query; Program Selection; and Algorithm parameters.

- 1: Enter your sequence (copy-paste or import your sequence file) can be multiple at the same time
- 2: Check those parameters (default)
- 3: Can be used to refine/exclude results to select a 'good' outgroup
- 4: Usually better to show the results in a new window for an easier repeatability/correction of the analysis

Blastn (3)

The screenshot shows the BLAST search results page. At the top, there's a header with the NIH logo, U.S. National Library of Medicine, and a 'Log in' button. Below the header, the URL is 'BLAST® > blastn suite > results for RID-0JAEZEWX014'. The main content area has sections for 'Edit Search', 'Save Search', and 'Search Summary'. On the left, there's a table of search parameters: Job Title (Unknown sequence), RID (0JAEZEWX014, search expires on 12-31 05:34 am, Download All), Program (BLASTN, Citation), Database (nt, See details), Query ID (lcl|Query_33195), Description (None), Molecule type (dna), Query Length (1425), and Other reports (Distance tree of results, MSA viewer). A red dashed circle highlights the 'Distance tree of results' link. Below these, a red oval highlights the navigation tabs: Descriptions (selected), Graphic Summary, Alignments, and Taxonomy.

The result of the Blast provides information on the search query and allow different filters.

The section below displays results in different formats (see further).

NB: BLAST provides a 'distance tree' which gives you a first idea of the possible related species to yours. This **cannot be** the tree you publish.

Blastn (4)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100 ?

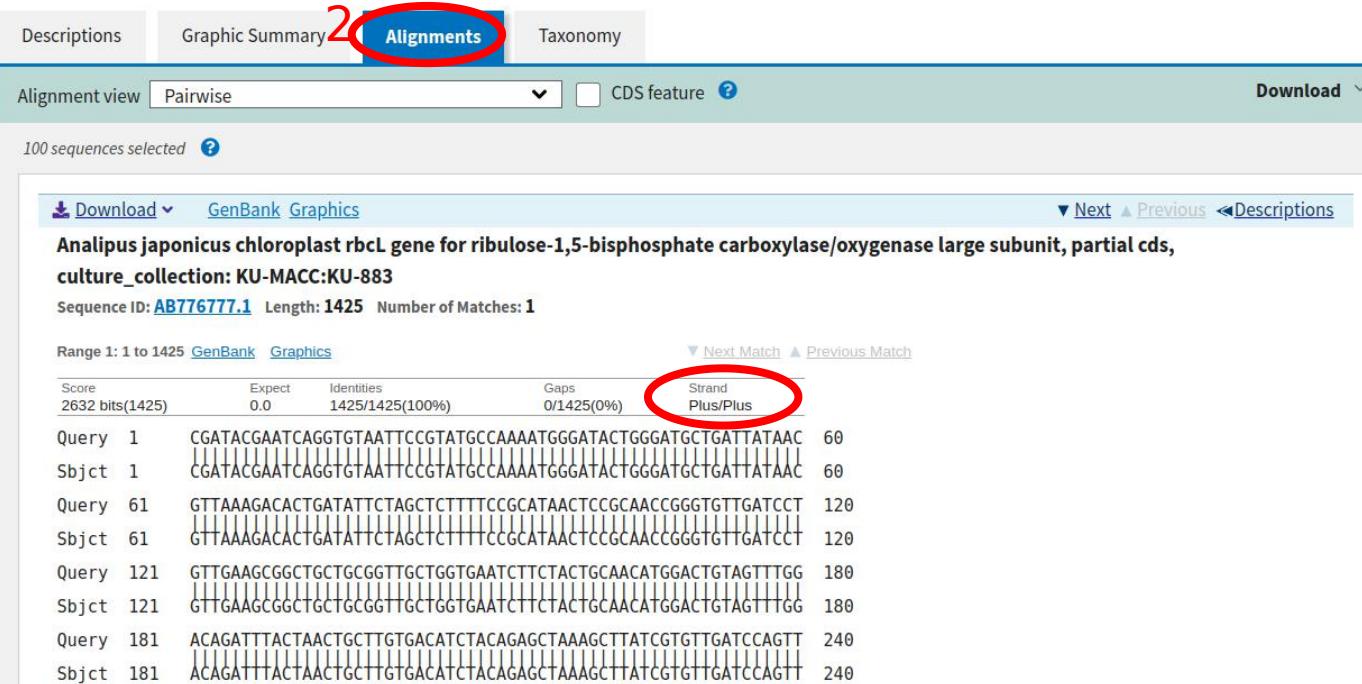
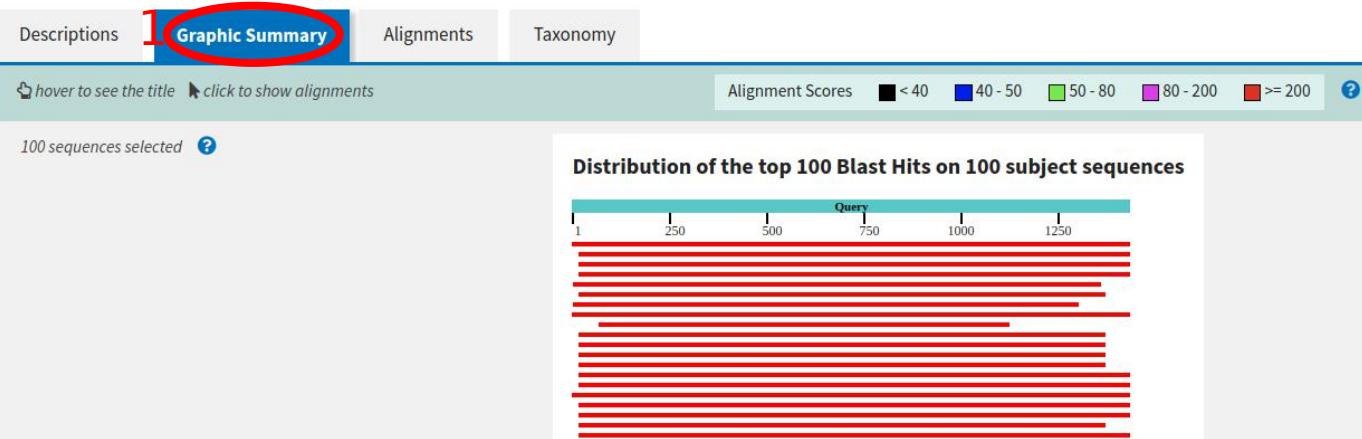
select all 100 sequences selected GenBank Graphics Distance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Analipus japonicus chloroplast rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, culture collection: K	2632	2632	100%	0.0	100.00%	AB776777.1
<input checked="" type="checkbox"/>	Analipus japonicus chloroplast rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxygenase, partial cds	2569	2569	98%	0.0	99.57%	AB264042.1
<input checked="" type="checkbox"/>	Ralfsia fungiformis chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	2093	2093	98%	0.0	92.97%	AB250071.1
<input checked="" type="checkbox"/>	Heteroralfsia saxicola chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	2043	2043	98%	0.0	92.84%	AB250070.1
<input checked="" type="checkbox"/>	Ralfsia tenebris voucher GWS019661 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL).gene, partial cds; chloroplast	2034	2034	94%	0.0	93.85%	MH593212.1
<input checked="" type="checkbox"/>	Ralfsia verrucosa chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	1991	1991	94%	0.0	93.38%	AB250072.1
<input checked="" type="checkbox"/>	Ralfsia unimaculata voucher GWS021144 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL).gene, partial cds; chloroplast	1960	1960	90%	0.0	94.04%	MH593196.1
<input checked="" type="checkbox"/>	Platysiphon glacialis chloroplast rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	1923	1923	100%	0.0	91.03%	AB776772.1
<input checked="" type="checkbox"/>	Analipus japonicus voucher GWS035897 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL).gene, partial cds; chloroplast	1921	1921	73%	0.0	99.71%	MH593219.1
<input checked="" type="checkbox"/>	Ralfsiaceae sp. MLN-2017a voucher FCME-PTM8982 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL).gene, partial cd	1919	1919	94%	0.0	92.42%	KX831608.1
<input checked="" type="checkbox"/>	Ralfsiaceae sp. MLN-2017a voucher FCME-PTM8973 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL).gene, partial cd	1908	1908	94%	0.0	92.27%	KX831607.1
<input checked="" type="checkbox"/>	Ralfsiaceae sp. MLN-2017a voucher FCME-PTM8979 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL).gene, partial cd	1908	1908	94%	0.0	92.27%	KX831606.1
<input checked="" type="checkbox"/>	Ralfsiaceae sp. MLN-2017a voucher FCME-PTM8970 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL).gene, partial cd	1903	1903	94%	0.0	92.19%	KX831605.1
<input checked="" type="checkbox"/>	Phyllariopsis purpurascens chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, strain: Spain	1899	1899	98%	0.0	91.00%	AB045251.1
<input checked="" type="checkbox"/>	Phyllariopsis purpurascens chloroplast rbcL gene for ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, strain: Spain	1893	1893	98%	0.0	90.93%	AB045249.1
<input checked="" type="checkbox"/>	Saccorhiza polyschides chloroplast rbcL, rbcS genes for ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, ribulose-1,5-bispho	1890	1890	100%	0.0	90.61%	AB545978.1

The first tab (Description) gives you a link to the closest match for your sequence. The best sequence is the one with:

- the highest percentage of identity
 - the best query cover
- (- a good total score)

Blastn (5)



1: The 'Graphic summary' part provides similar results as 'Descriptions' but more visual and highlights the aligned segments and their score.

2: 'Alignments' is mainly used for proteins but can be informative to inspect the amount of gaps and the orientation of the strands in the alignment

Presentation of TreeBase

<https://treebase.org>

The screenshot shows the TreeBase homepage. At the top, there is a blue header bar with the TreeBASE logo and the text "A Database of Phylogenetic Knowledge". Below the header is a navigation menu on the left side with the following items: Search data, Upload data, Documentation, Technology overview, Upload Tutorial, Web service APIs, About, Background, People, Partnerships, References, Data Management, Journals, and Contact. On the right side, there is a main content area titled "Welcome to TreeBASE". The content describes TreeBASE as a repository of phylogenetic information, accepting user-submitted phylogenetic trees and data. It mentions that data can be submitted to TreeBASE and will be available to the public once peer-reviewed. It also states that such data are only available to publication editors or reviewers using a special access URL. The content is produced and governed by "The Phyloinformatics Research Foundation, Inc.". Below this, it provides statistics: as of April 2014, TreeBASE contains data for 4,076 publications written by 8,777 different authors. These studies analyzed 8,233 matrices and resulted in 12,817 trees with 761,460 taxon labels that mapped to 104,593 distinct taxa. There is also a section for "Some recent additions" which lists new features including richer annotation of metadata, improved normalization of names, visualization and editing of trees using Phylowidget, and the ability to search or browse tree topology.

TreeBASE
A Database of Phylogenetic Knowledge

Search data
Upload data
Documentation
Technology overview
Upload Tutorial
Web service APIs
About
Background
People
Partnerships
References
Data Management
Journals
Contact

Welcome to TreeBASE

TreeBASE is a repository of phylogenetic information, specifically user-submitted phylogenetic trees and the data used to generate them. TreeBASE accepts all kinds of phylogenetic data (e.g., trees of species, trees of populations, trees of genes) representing all biotic taxa. Data in TreeBASE are exposed to the public if they are used in a publication that is in press or published in a peer-reviewed scientific journal, book, conference proceedings, or thesis. Data used in publications that are in preparation or in review can be submitted to TreeBASE but will not be available to the public until they have passed peer review. Aside from the submitter, such data are only available to the publication editors or reviewers using a special access URL. TreeBASE is produced and governed by [The Phyloinformatics Research Foundation, Inc.](#).

As of April 2014, TreeBASE contains data for 4,076 publications written by 8,777 different authors. These studies analyzed 8,233 matrices and resulted in 12,817 trees with 761,460 taxon labels that mapped to 104,593 distinct taxa.

Some recent additions:

The current release includes a host of new features and improvements over the previous TreeBASE prototype. New features include:

- Richer annotation of metadata (journal DOIs, specimen georeferences, Genbank accession numbers, etc)
- A mapping between taxon labels and taxonomic names in uBio and NCBI for improved normalization of names
- The ability to visualize and edit trees using Phylowidget
- The ability to search or browse tree topology

LinkedIn icon
Twitter icon

Online database where **some** researchers upload their dataset (aligned sequences).

Should only be a complement to your own analysis.

Sequence Alignment

Before alignment

AliView - *Woodsia_chloroplast_min4_20131109_v2.excluded.nexus

File Edit Selection View Align Primer External commands Help

Search

310 320 330 340 350 360 370 380 390

Woodsia_alpina_4x_F4_F135
Woodsia_andersonii_8x_F56
Woodsia_andersonii_8x_F75
Woodsia_canescens_2x4x_F
Woodsia_cochisensis_4x_F4
Woodsia_cycloloba_8x_F72
Woodsia_elongata_2x4x_F1
Woodsia_fragilis_4x_F19_F1
Woodsia_glabella_2x_F86
Woodsia_glabella_2x_F88_F1
Woodsia_gracilima_2x_F55
Woodsia_ilvensis_2x_F3_F21
Woodsia_indusiosa_4x_F91
Woodsia_intermedia_4x_F71
Woodsia_lanosa_8x_F119
Woodsia_macrochlaena_2x_I
Woodsia_manchuriensis_2x_F
Woodsia_mollis_3x4x_F18_F
Woodsia_burgessiana_syn_n
Woodsia_montevidensis_ecu
Woodsia_neomexicana_4x_F
Woodsia_obtusa_2x4x_F70
Woodsia_okamotoi_4x8x_F1
Woodsia_oregana_ssp_orega
Woodsia_aff_phillipsii_4x_F6
Woodsia_plummerae_4x_F31
Woodsia_polystichoides_2x_F
Woodsia_rosthorniana_4x_F1
Woodsia_scopulina_ssp_sco
Woodsia_phillipsii_like_mexio
Woodsia_sp_cyclolobalike1

Selected: Woodsia_cochisensis_4x_F48_F152_F195 | pos: 353 | pos (ungaped): 346 | Selected sequences: 1 | columns: 1 | total selected characters: 1

Alignment: 35 sequences 6604 pos.

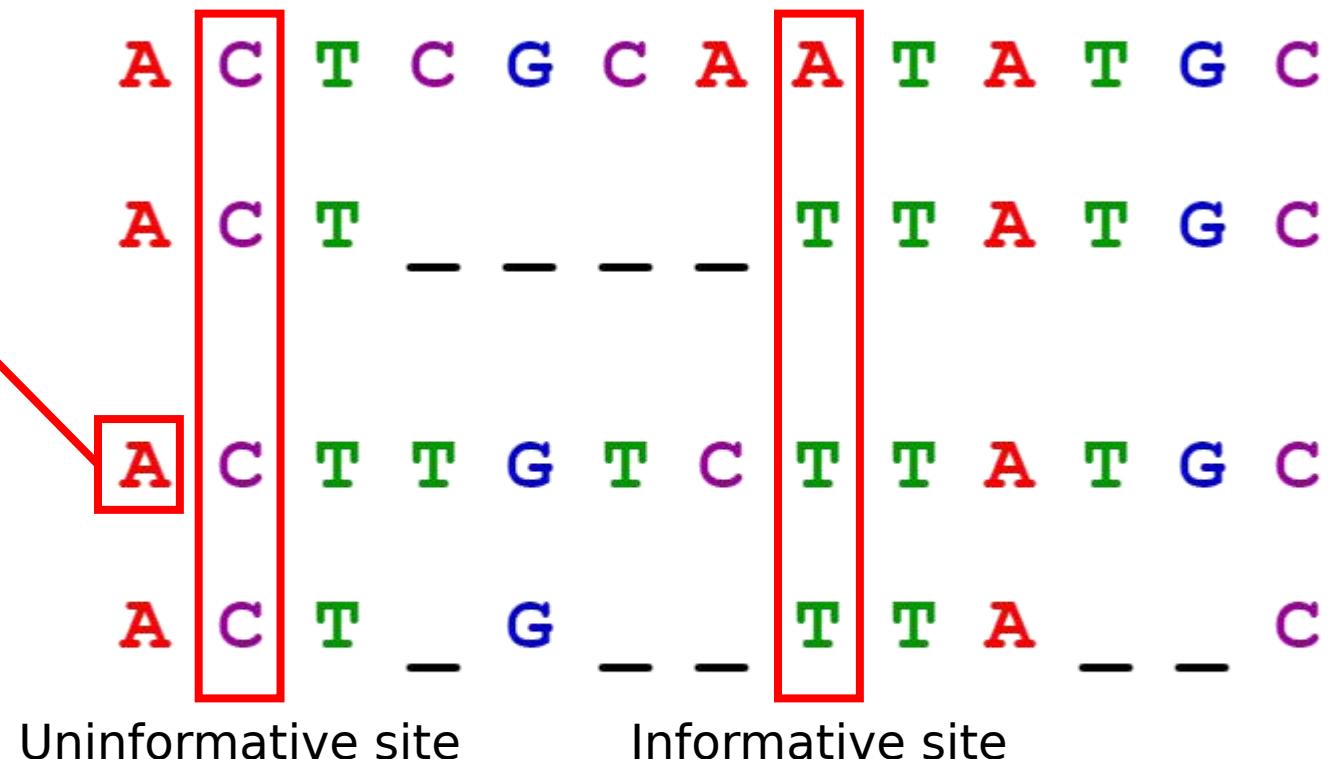
4 /

After alignment



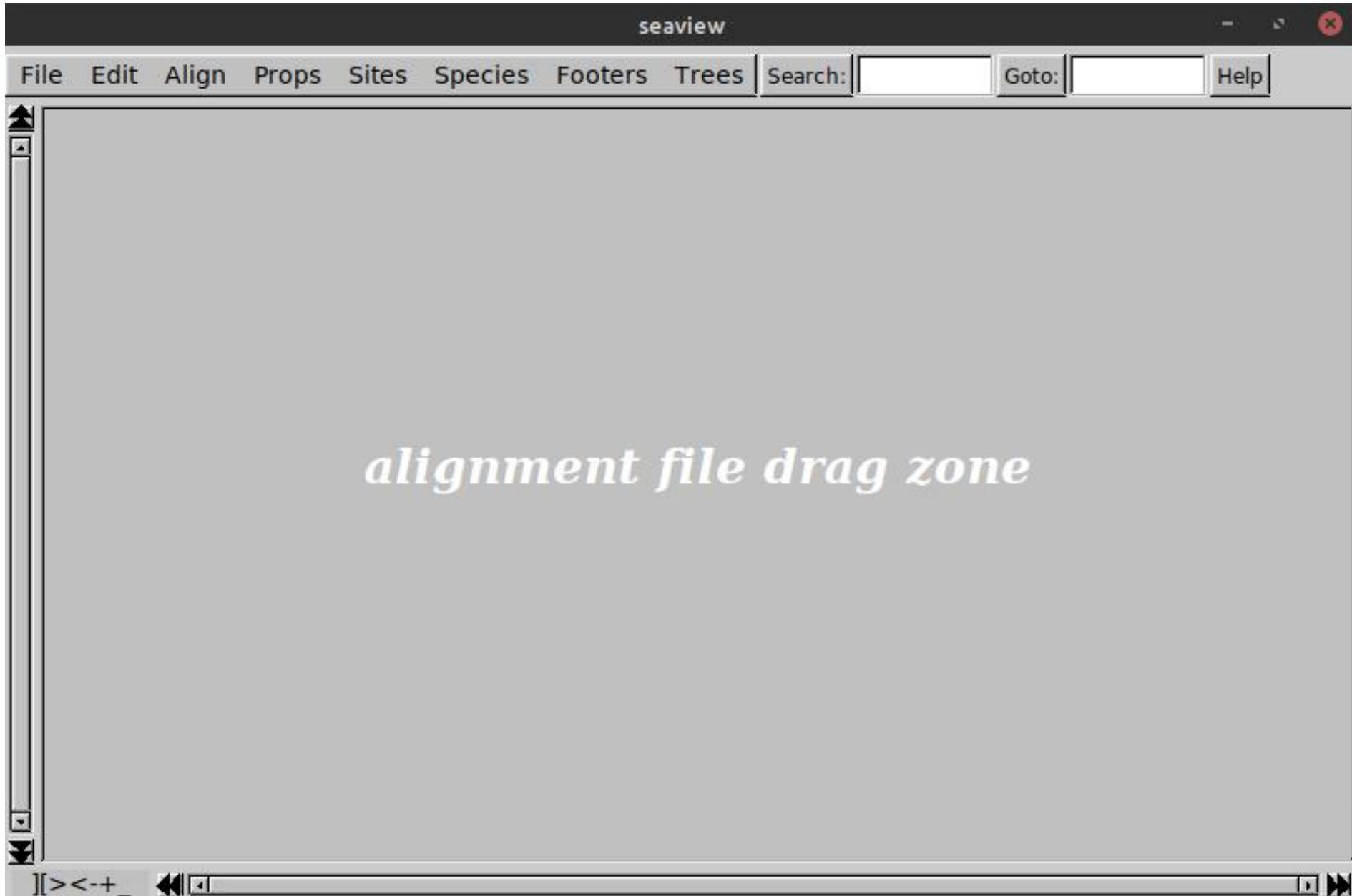
Sequence analysis

A is called the
character state
of the first codon position
(=character)



Seaview (1)

S
E
A
-



Might seem rather basic but is less resource demanding than MEGA (see further) so better for big dataset that you wish to analyse with the same algorithm than the one of MEGA.

Start by dropping your .fasta file in the seaview window

Seaview (2) S E A -

for sequence alignment

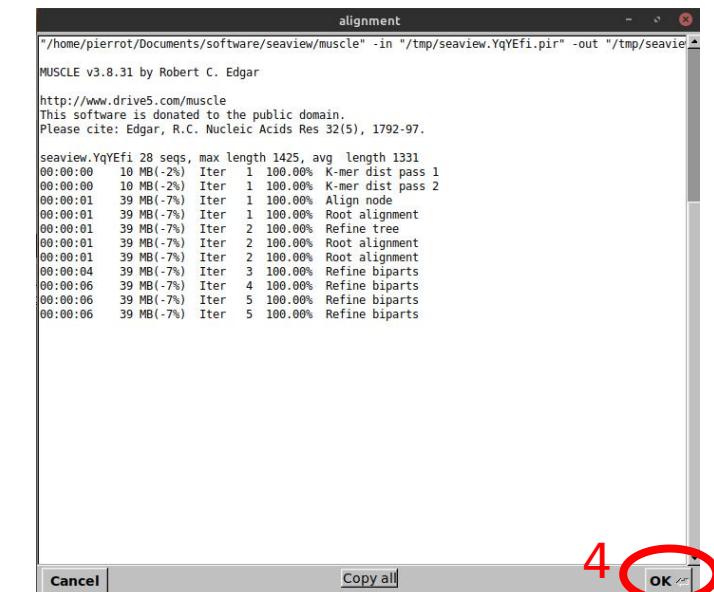
The screenshot shows the Seaview interface with the title "Ralfsiales_Japan_bis.fasta". The menu bar includes File, Edit, Align (circled in red), Props, Sites, Species, Footers, Trees, Search, Goto, and Help. The "Align" option is circled in red with a number 1. The "Align all" option in the dropdown menu is also circled in red with a number 3. The "Alignment options" option in the dropdown menu is circled in red with a number 2. The main window displays a sequence alignment of multiple DNA or RNA sequences. The sequences are color-coded by nucleotide (A, T, C, G). The top sequence is labeled "q:1 Pos:1|1 [Analipus japonicus AB776777] 72". The left sidebar lists various species names and their corresponding sequence IDs.

1: Select the menu 'Align'

2: Check that 'Alignment options' is on 'muscle'

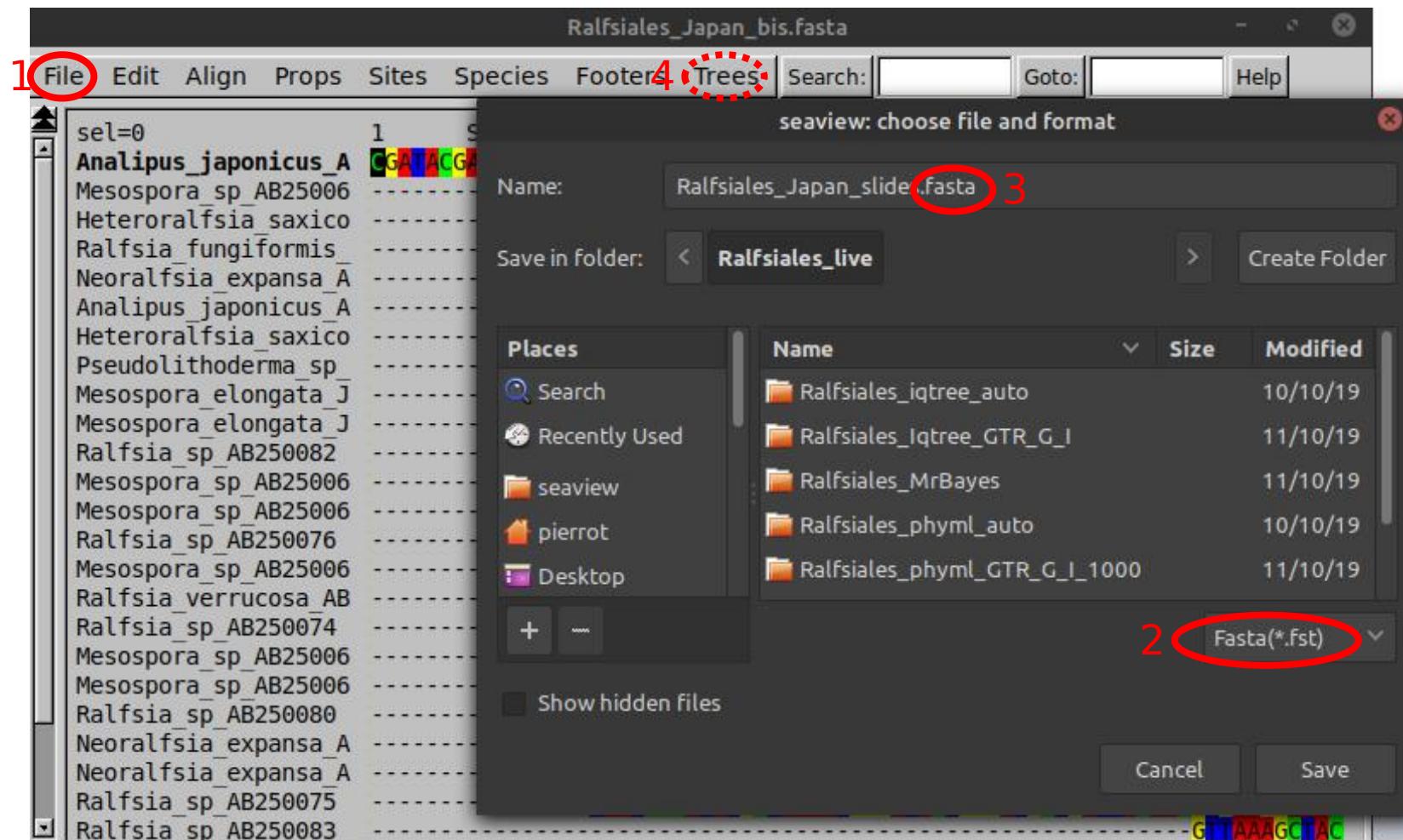
3: Click on 'Align all'

4: When the alignment is finished, click OK



Seaview (4)

for file conversion



After aligning or just to convert files, you can click on 'File' (1), 'Save as' and then choose your format in the bottom right corner (2).

Don't forget to add the extension (not automatic on Seaview) (3):

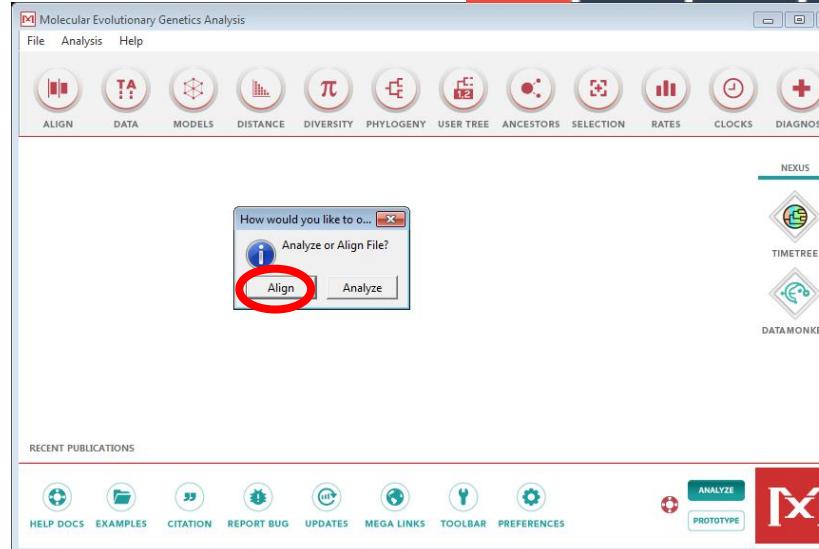
- .fasta (alignment file)
- .nex (Bayesian analyses)
- .phy (maximum likelihood analyses)

NB: you can visualize trees in Seaview (4)

MEGA (1) M | E | G | A

Molecular Evolutionary
Genetics Analysis

for sequence alignment



1

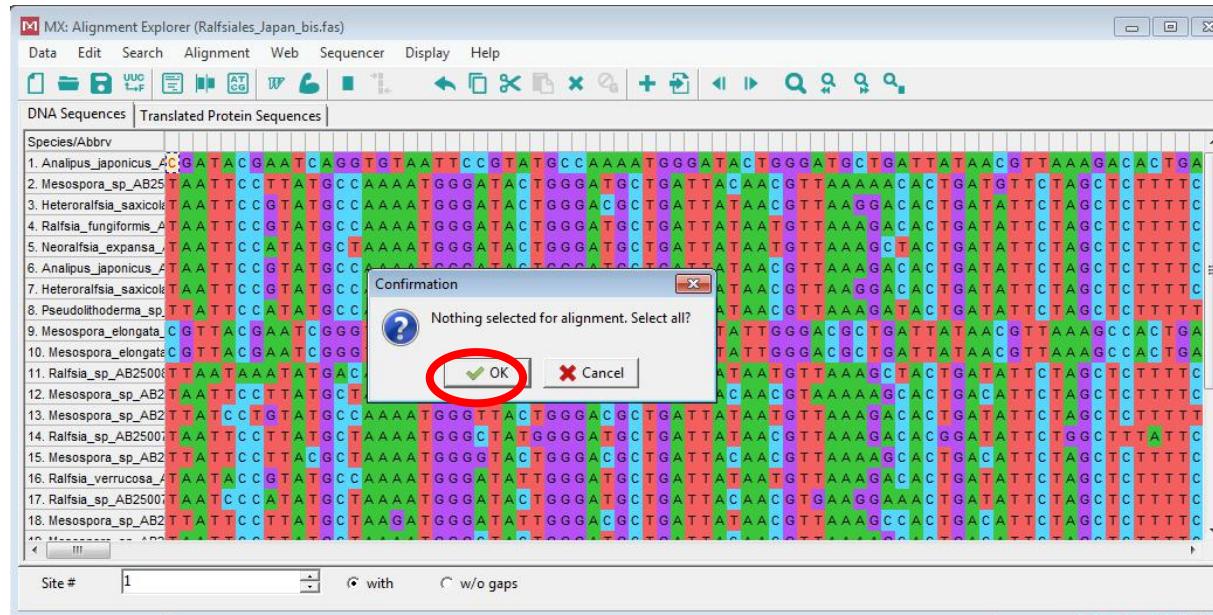
1: drag and drop the sequence file in the MEGA interface and select click on 'Align'.

The screenshot shows the MEGA Alignment Explorer window titled 'MX: Alignment Explorer (Ralfsiales_Japan_bis.fas)'. The window has a toolbar with various icons. The 'Align DNA' button in the toolbar is circled in red. The main area displays a grid of DNA sequences for 18 different species, each represented by a unique color. The sequences are aligned horizontally, showing homologous positions. The bottom of the window has a 'Site #' input field and some other controls.

2

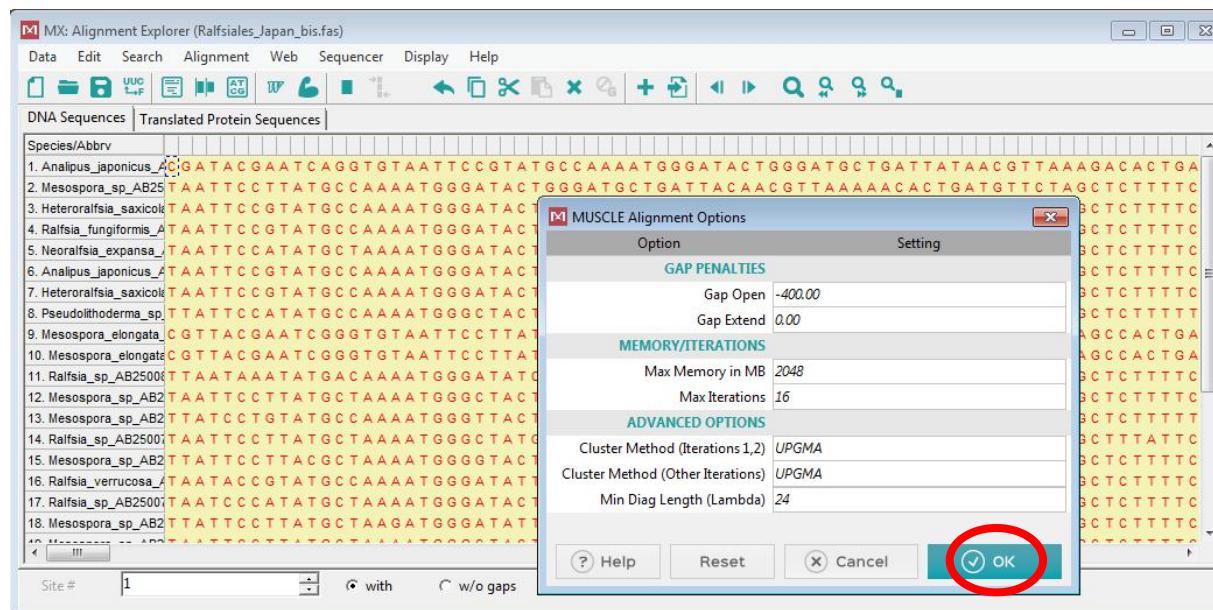
2: Sequences can be aligned using muscle by clicking on the arm symbol and then 'Align DNA'.

1

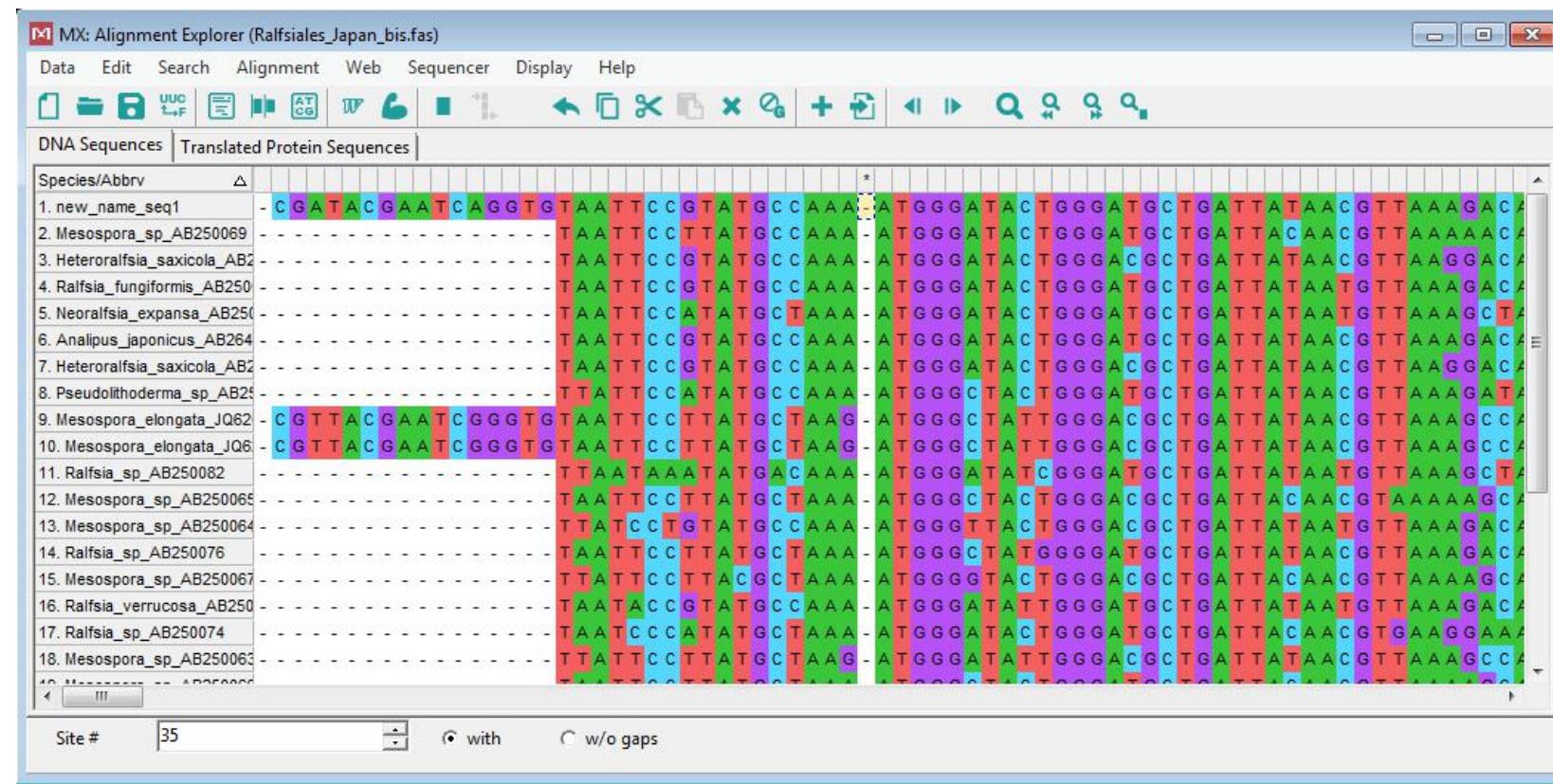


1: You might get an error message if you haven't selected sequences to align. By clicking 'OK' you select all sequences from the file.

2



2: The pop-up box of parameters can be left as default.



MEGA allows an easy overview on the sequences and a visual inspection of the data. Moreover, you can add, delete parts of your sequences and change the sequence names easily. Another practical feature is the possibility to sort the sequence by name. (in this example, a gap at the 35th position of all sequences has been inserted and the name of the first sequence has been changed)

MEGA Web Browser: Nucleotide BLAST: Search nucleotide databases using a nucleotide query

File Edit View Navigate Help

<https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&LAYOUT=OneWindows&AUTC> C + Add To Alignment

Nucleotide BLAST: Search nucleotide databases using a nucleotide query

NIH U.S. National Library of Medicine NCBI Sign in to NCBI

BLAST® » blastn suite Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blast blastp blastx tblastn tblastx

Enter Query Sequence BLASTN programs search nucleotide databases using a nucleotide query. more... Reset page Bookmark

Enter accession number(s), gi(s), or FASTA sequence(s)
ATGGGCTATTGGGACGCTGATTATACTGTTAAAGCCACTGACATTCTAGCTCTTCCGCATAACTCCGCAAC
CGGGTGTGATCCTGTTGAGGCCGCTGCGGTTGCTGGCAATCTTCACTGCAACATGGACTGTAGTTG
GACAGATTACTAAGCTCTTGACATTACAGAGCTAAAGCTTACCGTGTGATCCGGTTCAGGAACAAAC
GATCAGTATTTGCATACATAGCATATGAATGTGATTATTCAAGAAGGGTCTTAGCTAACTTAACAGCCT

Clear Query subrange From
To

Or, upload file Choose File No file chosen

Job Title Enter a descriptive title for your BLAST search

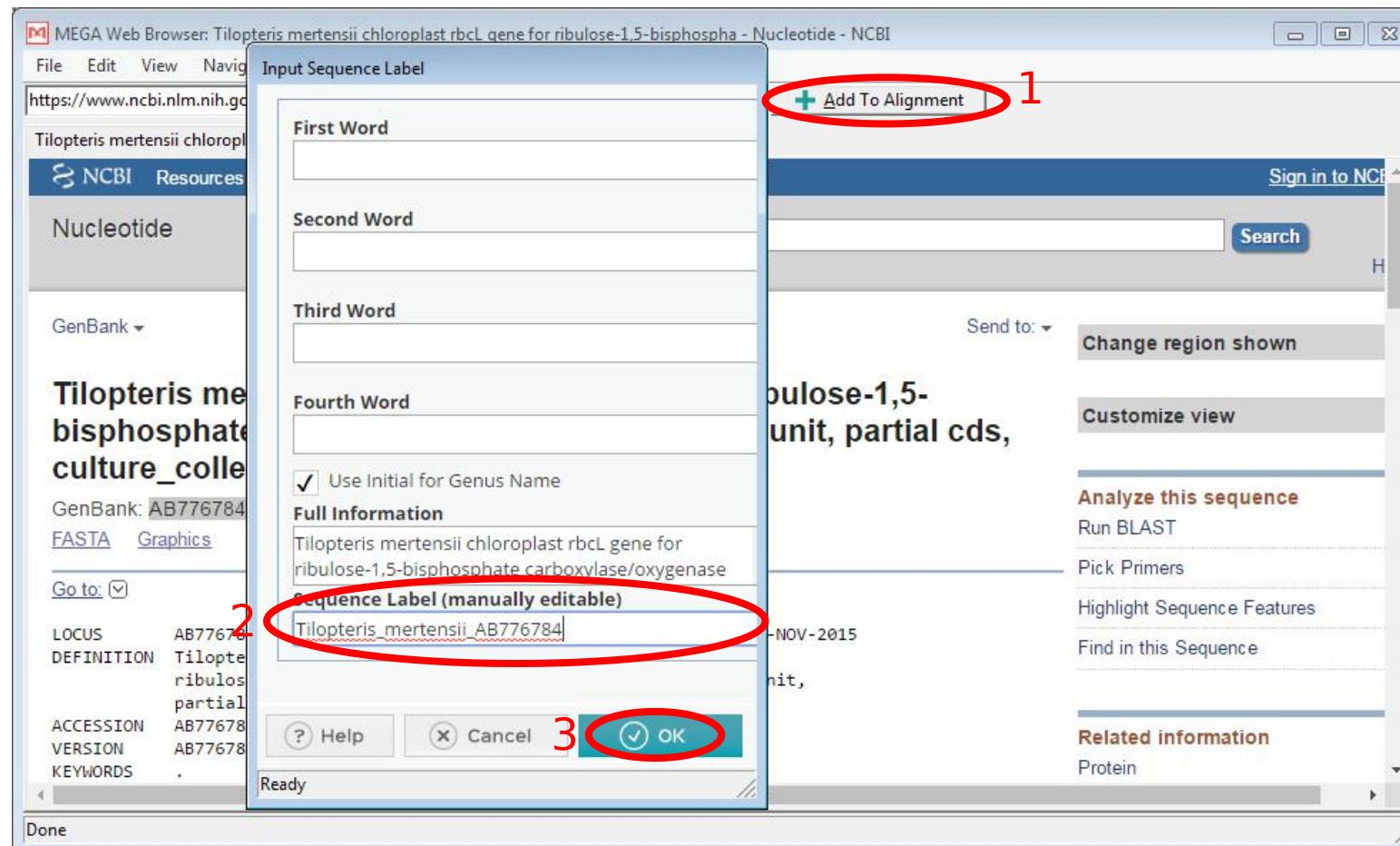
Align two or more sequences

BLAST results will be displayed in a new format by default [New](#)

https://www.ncbi.nlm.nih.gov/

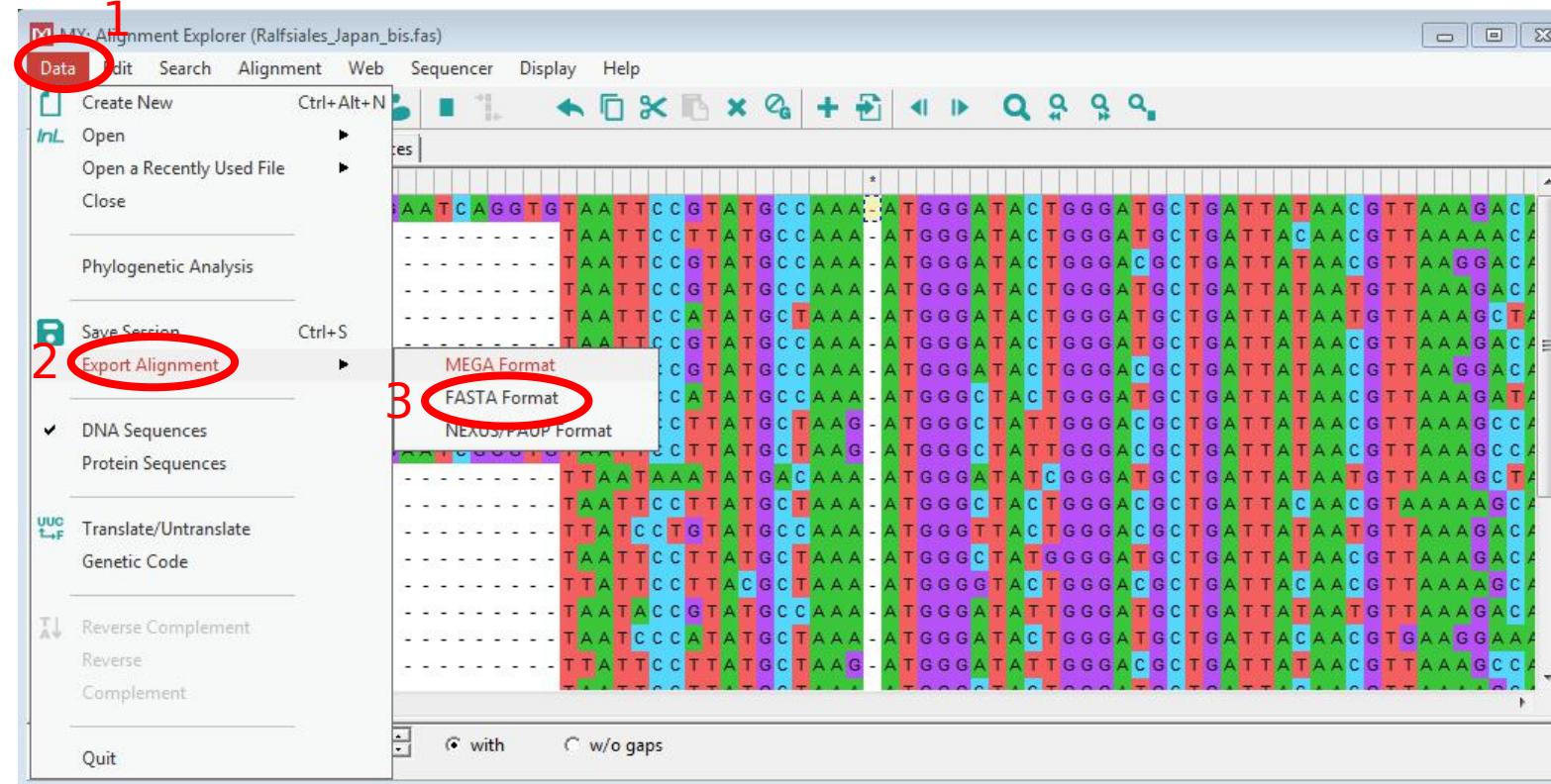
If you have a strange or unknown sequence, you can select it and then click on the 

This opens a BLAST interface as already seen previously.



If you want to add sequences directly to your alignment (an outgroup for example), click on 

This opens a NCBI interface as already seen previously. Once you've found your sequence of interest, you can click on 'Add to Alignment' (1) and edit the name that you want to give to that sequence (2)



To save your alignment:

1: Data

2: Export Alignment

3: Fasta Format

MAFFT MAFFT

<https://mafft.cbrc.jp/alignment/server/>

MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences



Download version

[Mac OS X](#)

[Windows](#)

[Linux](#)

[Source](#)

Online version

[Alignment](#)

[mafft --add](#)

[Merge](#)

[Phylogeny](#)

[Rough tree](#)

[Merits / limitations](#)

[Algorithms](#)

[Tips](#)

[Benchmarks](#)

[Feedback](#)

[Follow](#)

For a large number of short sequences, try [an experimental service](#).

[Experimental service for aligning raw reads \(2019/Aug\)](#) **New!**

Memory usage is more strictly restricted than before, in order to make the service more stable (2019/Aug/29).

If your job is unexpectedly terminated with an error message, "Cannot allocate ..." or "Allocation error ...", then [let us know](#).

Multiple sequence alignment and NJ / UPGMA phylogeny

Input:

Paste protein or DNA sequences in fasta format. [Example](#)

or upload a **plain text** file: No file chosen

Use [DASH](#) to add homologous structures (protein only) **New! 2018/Dec/23**

Output original plus DASH sequences Output original sequences only

Give structural alignment(s) externally prepared

Allow unusual symbols (Selenocysteine "U", Inosine "i", non-alphabetical characters, etc.) [Help](#)

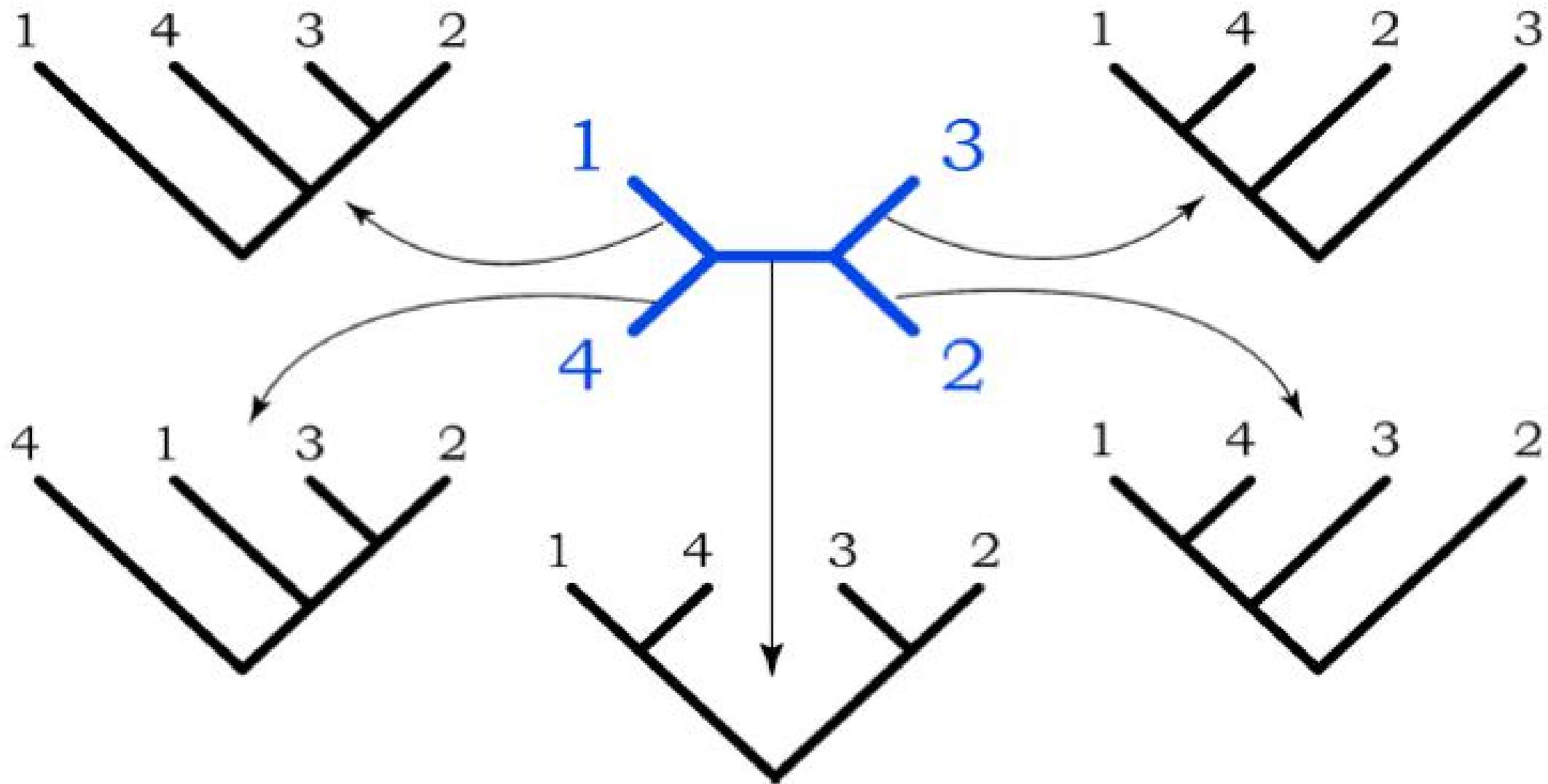
Input your sequences and leave the other parameters as defaults.

MAFFT is normally meant for aligning protein sequences but can be used for big datasets of nucleotide sequences. The algorithm is different than muscle, keep that in mind when you try to compare to results from a muscle alignment (MEGA, seaview, etc.)

Maximum Likelihood Analyses



Number of trees possible



Number of trees possible

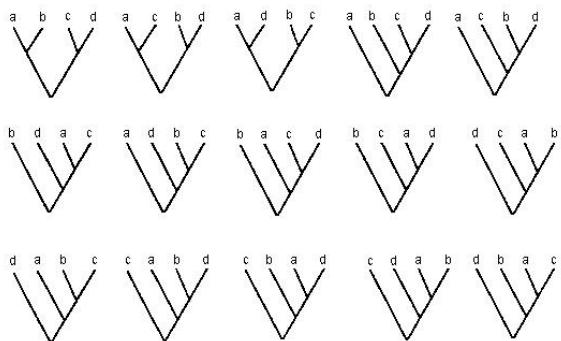
Industries	Unrooted trees	Rooted trees
n	$\prod_{i=3}^n (2i - 5)$	$\prod_{i=2}^n (2i - 3)$
3	1	3
4	3	15
5	15	105
10	2×10^6	3×10^7
15	8×10^{12}	2×10^{14}
20	2×10^{20}	8×10^{21}
25	3×10^{28}	1×10^{30}
50	3×10^{74}	3×10^{76}
100	2×10^{182}	3×10^{184}
500	1×10^{1277}	1×10^{1280}

Maximum likelihood analysis

Phylip file (.phy)

Likelihood: Probability to obtain the alignment
given a certain evolutionary model

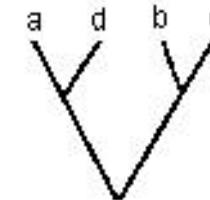
Workflow



Creation of trees



Likelihood calculation of each tree



Tree with highest likelihood

Drawback

Fixed evolutionary model

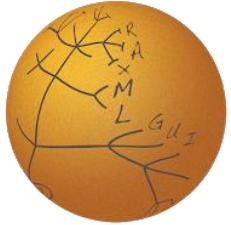
Manual approach

jModelTest



jModelTest

1: Define the evolutionary model suitable for the dataset with jModelTest.



2: Run the analysis with RAxML

This option is more and more left behind because it requires to run the analysis on your own computer. This tutorial will present two online alternatives.

PhyML (1) ATGC

<http://www.atgc-montpellier.fr/phym/>

PhyML online execution

1 

Input Data

Sequences (PHYLIP format)

Data Type

File Example (DNA file) (from Phylogenetic Handbook)

DNA Amino-Acids

Substitution Model

Automatic model selection by SMS

AIC (Akaike Information Criterion)

BIC (Bayesian Information Criterion)

Selection criterion

If you use SMS, please cite:
"SMS: Smart Model Selection in PhyML."
Vincent Lefort, Jean-Emmanuel Longueville, Olivier Gascuel.
Molecular Biology and Evolution, 34(9):2422-2424, 2017.

Set by user

Substitution model: HKY85

Equilibrium frequencies: optimized empirical

Transition / transversion ratio (DNA models): fixed estimated

Proportion of invariable sites: fixed estimated

Number of substitution rate categories: fixed estimated

Gamma shape parameter: fixed estimated

Tree Searching

Starting tree(s): Choose file No file chosen

Type of tree improvement: NNI

File BIONJ

Number of random starting tree: yes no

Branch Support

Fast likelihood-based method: aLRT SH-like

Use aLRT or aBayes to save computing time

yes no

Perform bootstrap: yes no

Name of your analysis

Your email 

Please confirm your email

Execute & email results

1: Upload your file.phy

2: The other parameters can be left as default or you can change 'Perform bootstrap' to 1000 or 1000

3: It is recommended, given the time that an analysis can take, to provide and email address

- *model_choice.csv* → can be downloaded in order to know which model was selected by the program if left to automatic (the model should be precised in your article)
- *task_name.nwk* → Consensus tree to analyse
- *task_name.pdf* **OR** *task_name.svg* → Consensus tree in a vectorial format for graphical customisation (the topology cannot be edited anymore)

IQTREEE (1) IQ-TREE

http://iqtree.cibiv.univie.ac.at/

Input Data

Alignment file : 1 C:\fakepath\Ralfsiales_Japan_bit

Use example alignment: Yes

Sequence type: 2 DNA Auto-detect DNA->AA Binary

Partition file: This field is optional.

Partition type: Edge-linked Edge-unlinked

Substitution Model Options

Substitution model: 3 Auto

FreeRate heterogeneity: Yes [+R]

Rate heterogeneity: Gamma [+G] Invar. sites [+I]

#rate categories: 4

State frequency: Empirical (from data) AA model (from matrix) ML-optimized

Codon F1x4 Codon F3x4

Ascertainment bias correction: Yes [+ASC]

1: Upload your file.phy

2: You can (optional) precise that you uploaded DNA sequences

3: Let IQTREE select the model itself but thick the FreeRate heterogeneity box

IQTREEE (2) IQ-TREE

<http://iqtree.cibiv.univie.ac.at/>

Branch Support Analysis

Bootstrap analysis: None Ultrafast Standard

Number of bootstrap alignments: 1 **10000**

Create .ufboot file: Yes (write bootstrap trees to .ufboot file)

Maximum iterations: 10000

Minimum correlation coefficient: 0.99

Single branch tests:

SH-aLRT branch test: No Yes #replicates: 1000

Approximate Bayes test: Yes

IQ-TREE Search Parameters

Perturbation strength: 0.5

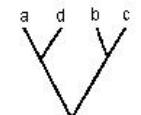
IQ-TREE stopping rule: 100

Email (optional, to retrieve results): **2**

SUBMIT JOB

1: For publications, set the bootstrap value to 10 000 (for quick overview, you can leave this parameter to default values

2: it's again advised to provide an email address to retrieve the results more easily



- *name_of_your_input_file.contree* → Consensus tree to analyse
- *name_of_your_input_file.iqtree* → Output of the program (model chosen, tree, etc.)
-  *name_of_your_input_file.log* → log file containing the different steps of the program (can be used to trace back execution errors)
- *name_of_your_input_file.treefile* → Consensus and best tree found by the analysis (allows a comparison of the results)

A black and white portrait of Thomas Bayes, an 18th-century English statistician and Presbyterian minister. He is shown from the chest up, wearing a dark coat over a white cravat and a patterned waistcoat. His hair is powdered and powdered white.

Bayesian Analyses

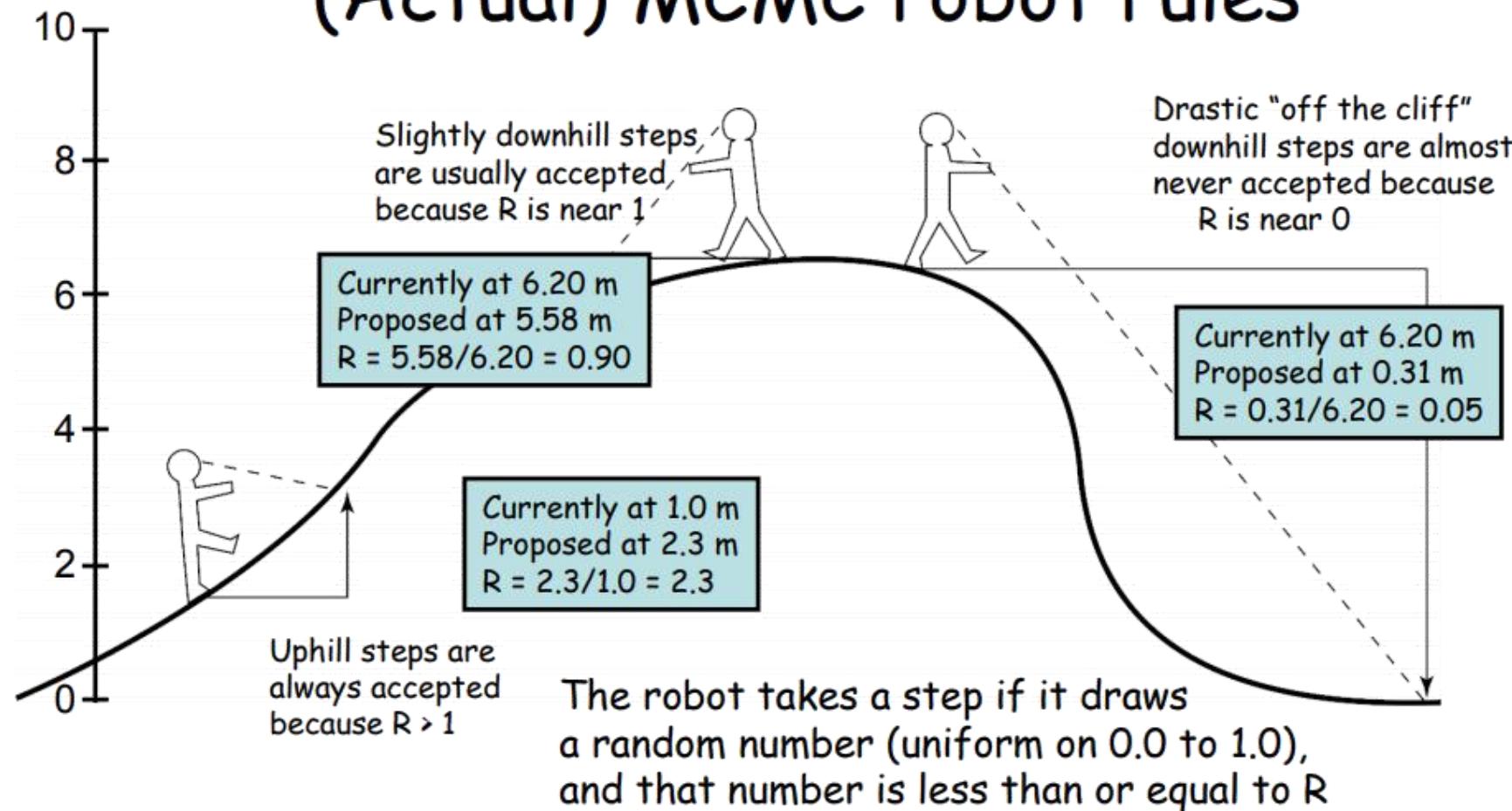
Bayesian analysis

Nexus file (.nex)

Bayesian analysis is based on comparison of prior and posterior probabilities

The evolutionary model is **not** fixed

(Actual) MCMC robot rules



MrBayes (1)



<https://www.phylo.org/portal2/home.action>

The screenshot shows the 'All Data' section of the CIPRES portal. On the left, there's a sidebar with 'Folders' and 'Total Storage: 7.96 GB'. Below that is a list with 'Aberdeen2019' (highlighted with a red box), 'Data (2)' (circled in red), and 'Tasks (2)'. A red number '1' is placed next to 'Aberdeen2019'. A red number '2' is placed next to the 'Upload Data' button, which is highlighted with a red circle. The main area shows a message: 'There are currently 2 data items in this folder. (Items 1 - 2 are shown here.)'. Below this is a table titled 'User Data' with columns: Select all, User Data ID, Label, Size, Data Format, Date Created, Parse Status. Two items are listed: 2241737 (Ralfsiales_Japan.fasta, 36.95 KB, Unknown, 10/9/19, 07:08, Unknown) and 2242645 (Ralfsiales_Japan_bis_aligned.nex, 41.56 KB, Unknown, 10/10/19, 05:48, Unknown). At the bottom are buttons for 'Move selected to' (set to '5,8S test') and 'Delete Selected'.

Select all	User Data ID	Label	Size	Data Format	Date Created	Parse Status
<input type="checkbox"/>	2241737	Ralfsiales_Japan.fasta	36.95 KB	Unknown	10/9/19, 07:08	Unknown
<input type="checkbox"/>	2242645	Ralfsiales_Japan_bis_aligned.nex	41.56 KB	Unknown	10/10/19, 05:48	Unknown

In order to run MrBayes, we will use the online platform of Cipres. After logging in on the platform:

1: Go to 'Data'

2: Click on 'Upload Data'

MrBayes (2)



<https://www.phylo.org/portal2/home.action>

Upload File

Upload your files **1** Choose Files No file chosen

You can select multiple files.

MSIE 9 and below support single uploads only.

You can also enter your data manually below

Label (required)

Label

Data:

Enter your data

2 Save

Cancel

1: Select your sequence file to upload (.nex for Bayesian analyses)

2: Click on 'Save'

MrBayes (3)



<https://www.phylo.org/portal2/home.action>

Folders

Total Storage: 7.96 GB

Aberdeen2019
Data (2)
Tasks (2)

1

Current CPU Hr Usage: 28 Explain this?
Running Tasks: 0
There are currently 2 tasks in this tab. (Items 1 - 2 are shown here.)

2 Create New Task

Show 20 records on each page

Page 1 of 1

Select All	Label	Tool	Input	Parameters	Date Created	Action
<input type="checkbox"/>	Raftsiales_Japan_Mrbayes_500000	Mrbayes on XSEDE	View (1)	View (75)	10/10/19, 05:49	View Output
<input type="checkbox"/>	Raftsiales_Japan_alignment_muscle	Muscle	View (1)	View (28)	10/9/19, 07:09	View Output

Move selected to 5.BS test GO

Kill Selected Delete Selected

1: Go to 'Tasks'

2: Click on 'Create New Task'

MrBayes (4)



<https://www.phylo.org/portal2/home.action>

Create new task

Task Summary Select Data Select Tool Set Parameters

You may edit your task using the tabs above.
Current CPU Hr Usage: 28 [Explain this?](#)

Description

Input 1 [Select Input Data](#)

Tool 2 [Select Tool](#)

Input Parameters 3 [Set Parameters](#)

[Save Task](#) [Save and Run Task](#) [Discard Task](#)

Saved tasks can be run later from the task list
XSEDE tasks are limited to 168 hours. Non-XSEDE tasks are limited to 72 hours.

1: Click on 'Select Input Data' and Cipres will automatically guide you through the three different steps (1, 2 and 3).

When those steps are completed, give a clear name that you will understand to your task (not 'tree').

MrBayes (5)



<https://www.phylo.org/portal2/home.action>

Create new task

Task Summary

Select Data

Select Tool

Set Parameters

You can choose the following data.

Select One	Label	Size	Data Format	Date Created	Parse Status
<input type="radio"/>	Ralfsiales_Japan.fasta	36.95 KB	Unknown	10/9/19, 07:08	Unknown
1 <input checked="" type="radio"/>	Ralfsiales_Japan_bis_aligned.nex	41.56 KB	Unknown	10/10/19, 05:48	Unknown

2

Select Data

Cancel

1: Select the nexus file containing your data

2: Click on the green box 'Select Data'

MrBayes (6)



<https://www.phylo.org/portal2/home.action>

Create new task

Task Summary Select Data **Select Tool** Set Parameters

If there is a tool or a feature you need, please [let us know](#).

[BALi-Phy on XSEDE \(3.2\)](#) ⓘ - BALi-Phy estimates multiple sequence alignments and evolutionary trees.

[BEAST2 on XSEDE \(2.1 - 2.6.1\)](#) ⓘ - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

[BEAST on XSEDE \(1.8.0 - 1.10.4\)](#) ⓘ - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

[BlastN \(2.2.1\)](#) ⓘ - Search DBs for Nucleotide Sequence similarity

[Clearcut \(1.0.9\)](#) ⓘ - Fast Implementation of Relaxed Neighbor Joining

[ClustalW \(2.1\)](#) ⓘ - Create Multiple Alignments from Sequences

[Consense \(Phylip 3.66\)](#) ⓘ - Find A Consensus Tree

[DPPDIV on XSEDE \(1.0\)](#) ⓘ - Estimating species divergence times and lineage-specific substitution rates on a fixed topology run on XSEDE

[ExaBayes on XSEDE \(1.5\)](#) ⓘ - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

[FastML on XSEDE \(3.1\)](#) ⓘ - Fast (Approximate) Maximum Likelihood tree construction - run on XSEDE

[FastTreeMP on XSEDE \(2.1.10\)](#) ⓘ - Fast (Approximate) Maximum Likelihood tree construction - run on XSEDE

[GARLI 2.01 on XSEDE \(2.01\)](#) ⓘ - Genetic Algorithm for Rapid Likelihood Inference - run on XSEDE.

[GARLI.conf Creator \(2.0\)](#) ⓘ - Creates a Garli.conf file for up to five partitions

[G-PhoCS on XSEDE \(1.3\)](#) ⓘ - A Generalized Phylogenetic Coalescent Sampler

[Guidance2 on XSEDE \(2.02\)](#) ⓘ - Accurate detection of unreliable alignment regions - run on XSEDE.

[IMa3 on XSEDE \(1.11\)](#) ⓘ - IMa3p - Parallel MCMC and inference of ancient demography under the Isolation with Migration (IM) model

[IQ-Tree on XSEDE \(1.6.10\)](#) ⓘ - Efficient phylogenomic software by maximum likelihood, run on XSEDE

[jModelTest2 on XSEDE \(2.1.6\)](#) ⓘ - Statistical selection of best-fit models of nucleotide substitution, run on XSEDE

[LogCombiner on XSEDE \(1.8.4-2.6.0\)](#) ⓘ - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

[MAFFT on XSEDE \(7.402\)](#) ⓘ - Multiple alignment program for amino acid or nucleotide sequences; parallel version

[Migrate-N on XSEDE \(3.6.11; 4.2.14\)](#) ⓘ - Estimation of Population Sizes and Gene Flow using the Coalescent

[ModelTest-NG on XSEDE \(0.1.5\)](#) ⓘ - Statistical selection of best-fit models of nucleotide and protein substitution, run on XSEDE

[MrBayes Restart on XSEDE \(3.2.x\)](#) ⓘ - Tree Inference Using Bayesian Analysis - run on XSEDE

1 [MrBayes on XSEDE \(3.2.7a\)](#) ⓘ - Tree Inference Using Bayesian Analysis - run on XSEDE

[Muscle \(3.7\)](#) ⓘ - Create Multiple Alignments from Sequences or Profiles

The list of programs offered by the Cipres Science Gateway appears.

1: Select 'MrBayes on XSEDE'

NB: feel free to use this platform to perform other kind of analyses (a muscle alignment for example).

MrBayes (7)



<https://www.phylo.org/portal2/home.action>

Task Summary Select Data Select Tool Set Parameters

MrBayes on XSEDE: Tree Inference Using Bayesian Analysis - run on XSEDE (John P. Huelsenbeck and Fred Ronquist)

Simple Parameters

Choose the MB version you wish to run * MrBayes 3.2.6 MrBayes 3.2.7a

My Data Contains a MrBayes Data Block (CHECK THIS OR MrBayes BLOCK ENTRIES WILL BE OVERWRITTEN!!!) *

My MrBayes Block species nruns= 2

My MrBayes Block species nchains= 4

Maximum Hours to Run (click here for help setting this correctly) 20

My Data Type Is (only one data type can be used through the web form, see help below) *

Set the Seed Number (set seed=)

Set the Swapseed (set swapseed=)

I need more memory

Use scientific notation

How many decimals should we print?

Run BEAGLE

Specify (only) one outgroup

We will mainly use default parameters but if you intend to further work on phylogeny, it's a section worth trying to further investigate.

1: You can change the amount of hours you want the analysis to run but it's not compulsory

2: Click on 'Advanced Parameters'



<https://www.phylo.org/portal2/home.action>

Parameters for MCMC

Number of Generations (Ngen=) (circled)

Number of Runs (nruns=)

Number of Chains to Run (nchains=)

Temperature parameter (temp=)

How often should swap of states be attempted

How many swaps should be tried per generation

How often should the Markov chain be sampled?

Write acceptance ratios of moves and swaps to file? + Yes No

Minimum frequency for a partition to be included (minpartfreq)

Record acceptance ratios for all chains? + Yes No

Discard a proportion of the sampled values as burnin when calculating the convergence diagnostic? + Yes No

Specify the fraction of the sampled values discarded as burnin

Specify the number of sampled values discarded as burnin

Stop early if the convergence diagnostic falls below the stop value? + Yes No

Please enter the stop value

Number of random perturbations to apply to user starting tree.

Save branch length information? + Yes No

Should taxa be ordered before trees are printed to file? + Yes No

Scroll down until the 'Parameters for MCM' section and set the number of generations to a number exceeding one million (50M more or less for a publication depending on the kind of dataset).

MrBayes (9)



<https://www.phylo.org/portal2/home.action>

Set Sumt parameters

1 **Sumt Burnin Value** 5000000

Discard a specified proportion of samples instead of a specific number(Relburnin=Yes) [Not Mandatory] Yes No

Specify the fraction of samples to be discarded (Burninfrac=) 0.25

How many .t files should be summarized (Sumt nrungs=) 2

How many trees should be in the Sumt model 1

Minimum probability of partitions to display in Sumt (0.05 = 95%) 0.05

Type of consensus tree + 50% Majority Rule All Compatible Groups

Choose the output format for your consensus tree (Conformat=) [Not Mandatory] Figtree Simple

Show Tree Probabilities + Yes No

Set Sump parameters

2 **Sump Burnin Value** 5000000

Discard a specified proportion of samples instead of a specific number(Relburnin=Yes) [Not Mandatory] Yes No

Specify the fraction of samples to be discarded (Burninfrac=) 0.25

How many '.p' files from independent analyses will be summarized (sump Nrungs=) 2

3 **Save Parameters** Reset Cancel

Scroll down to the last two parameters sections ('Set Sumt parameters' and 'Set Sump parameters').

1 and 2: Set the Sumt (and Sump) values to 10% of the value set previously for the number of generations.

3: Click on 'Save parameters'

NB: Cipres will give you warning messages, you can ignore them and click on ok

MrBayes (10)



<https://www.phylo.org/portal2/home.action>

Create new task

Task Summary Select Data Select Tool Set Parameters

You may edit your task using the tabs above.
Current CPU Hr Usage: 28 [Explain this?](#)

Description **2** Ralfsiales_Japan_MrBayes_50000000

Input **1** Inputs Set

Tool **1** MrBayes on XSEDE [Click for more info](#)

Input Parameters **1** 76 Parameters Set

3 Save Task Save and Run Task Discard Task

Saved tasks can be run later from the task list
XSEDE tasks are limited to 168 hours. Non-XSEDE tasks are limited to 72 hours.

1: Check that the different parameters are set correctly.

2: Give a meaningful title to your job. Here, the name contains two terms to describe the dataset, one for the type of analysis and one for the main parameter set.

3: Click on 'Save Task' (you might have to click multiple times)

MrBayes (11)



<https://www.phylo.org/portal2/home.action>

Tasks

Refresh Tasks

Current CPU Hr Usage: **28** [Explain this?](#)

Running Tasks: **0**

There are currently 3 tasks in this tab. (Items 1 - 3 are shown here.)

[Create New Task](#)

Show **20** records on each page

« < Page 1 of 1 > »

<input type="checkbox"/> Select All	Label	Tool	Input	Parameters	Date Created	Action
<input checked="" type="checkbox"/> 2	Ralfsiales_Japan_MrBayes_50000000	MrBayes on XSEDE	View (1)	View (75)	1/1/20, 12:28	1

Your job appears in the list of tasks and you can start running it by clicking on 'Run Task'

NB: if you want to run a similar task (i.e. same parameters), you can use the 'Clone' button to get back the same parameters.

MrBayes (12)



output files

<https://www.phylo.org/portal2/home.action>

<input type="checkbox"/> Select all	Tool Output	File Name	File Size (Bytes)		
<input type="checkbox"/>	ALL_FILES	infile.nex.run2.t	262445	View	Download
<input type="checkbox"/>		stderr.txt	0	View	Download
<input type="checkbox"/>		paramfile.txt	1106	View	Download
<input type="checkbox"/>		infile.nex.vstat	8923	View	Download
<input type="checkbox"/>		infile.nex.trprobs	8330	View	Download
<input type="checkbox"/>		infile.nex.parts	2034	View	Download
<input type="checkbox"/>		sumpoutput.out.pstat	986	View	Download
<input type="checkbox"/>		scheduler.conf	65	View	Download
<input type="checkbox"/>		sumpoutput.out.lstat	223	View	Download
<input type="checkbox"/>		stdoutput.txt	57402	View	Download
<input type="checkbox"/>		start.txt	40	View	Download
<input type="checkbox"/>		infile.nex.run1.t	262445	View	Download
<input type="checkbox"/>		infile.nex.tstat	3201	View	Download
<input type="checkbox"/>		infile.nex.ckp	14765	View	Download
<input type="checkbox"/>		infile.nex.mcmc	15922	View	Download
<input type="checkbox"/>		done.txt	49	View	Download
<input type="checkbox"/>		infile.nex.con.tre	15995	View	Download
<input type="checkbox"/>		infile.nex.run1.p	30886	View	Download
<input type="checkbox"/>		_scheduler_stderr.txt	634	View	Download
<input type="checkbox"/>		infile.nex.run2.p	30886	View	Download
<input type="checkbox"/>		term.txt	317	View	Download
<input type="checkbox"/>		infile.nex.ckp~	14765	View	Download
<input type="checkbox"/>		_JOBINFO.TXT	360	View	Download
<input type="checkbox"/>		infile.nex	42560	View	Download
<input type="checkbox"/>	PROCESS_OUTPUT	STDOUT	57402	View	Download
<input type="checkbox"/>		STDERR	0	View	Download

At the end of the analysis, the 'Run Task' button changes into 'View Output' and you will receive an email. You will only need the files circled in red and detailed on the next slide (you only need to download the full circled ones but you can view the dashed on directly on Cipres).

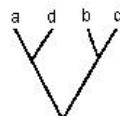
MrBayes (12)



output files



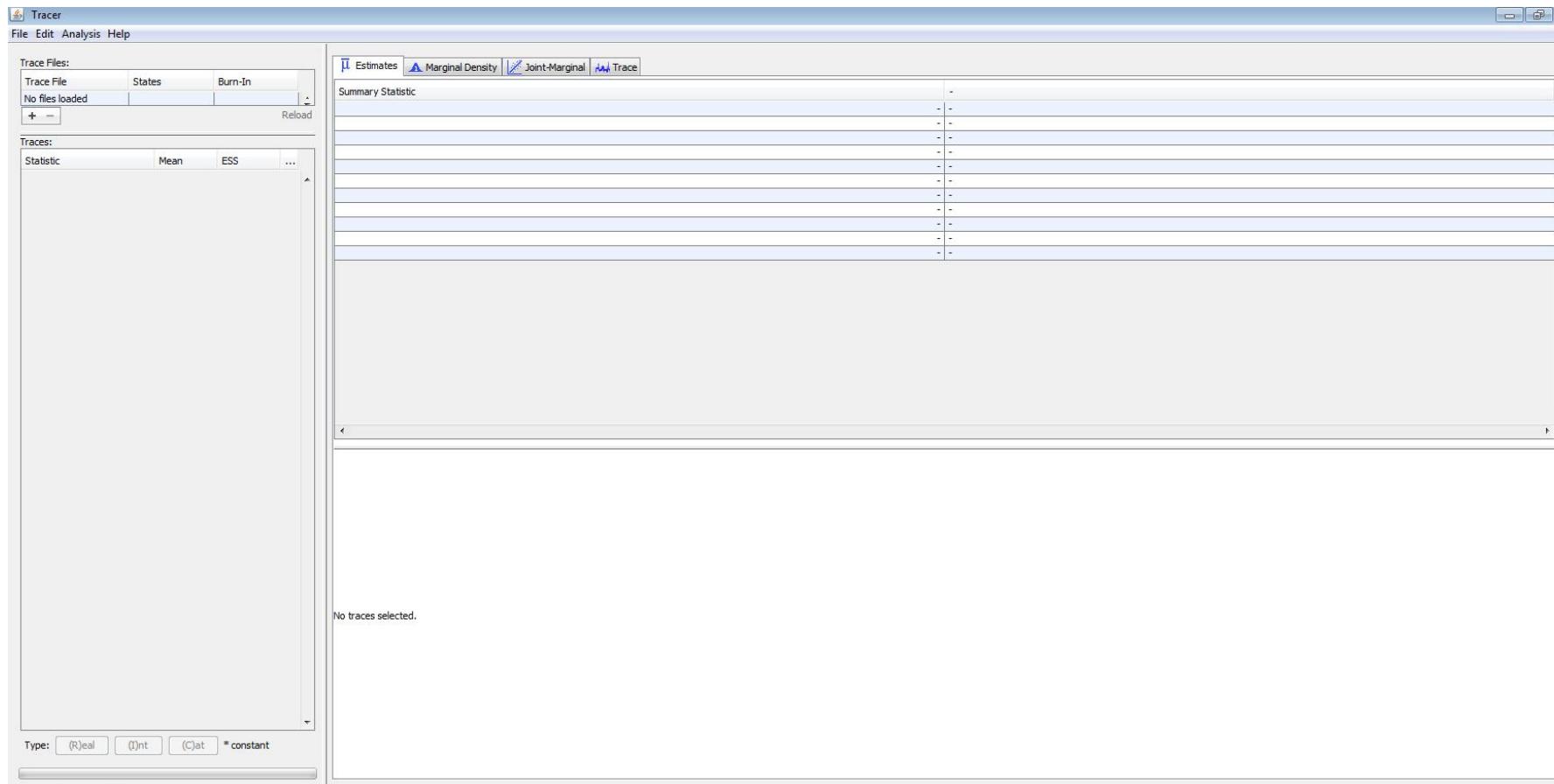
- stdout.txt → log file containing the different steps of the program
(can be used to trace back execution errors)



- infile.nex.con.tre → Consensus tree to analyse

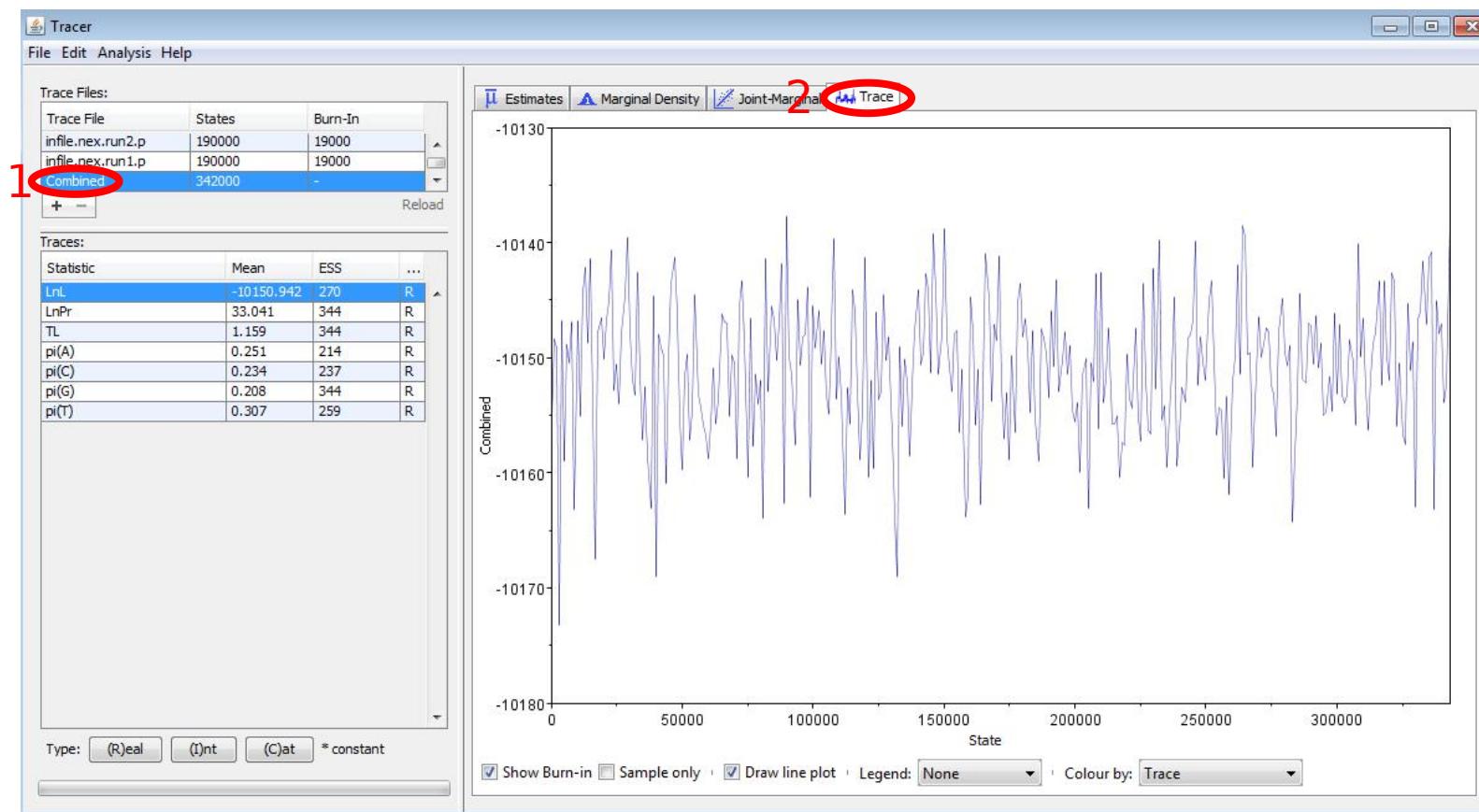
- infile.nex.run1.p & infile.nex.run2.p → Statistics of the two runs performed (see next slides)

Tracer (1)



Open Tracer, drag and drop
your two infile.nex.runX.p in
the window of the program

Tracer (2)



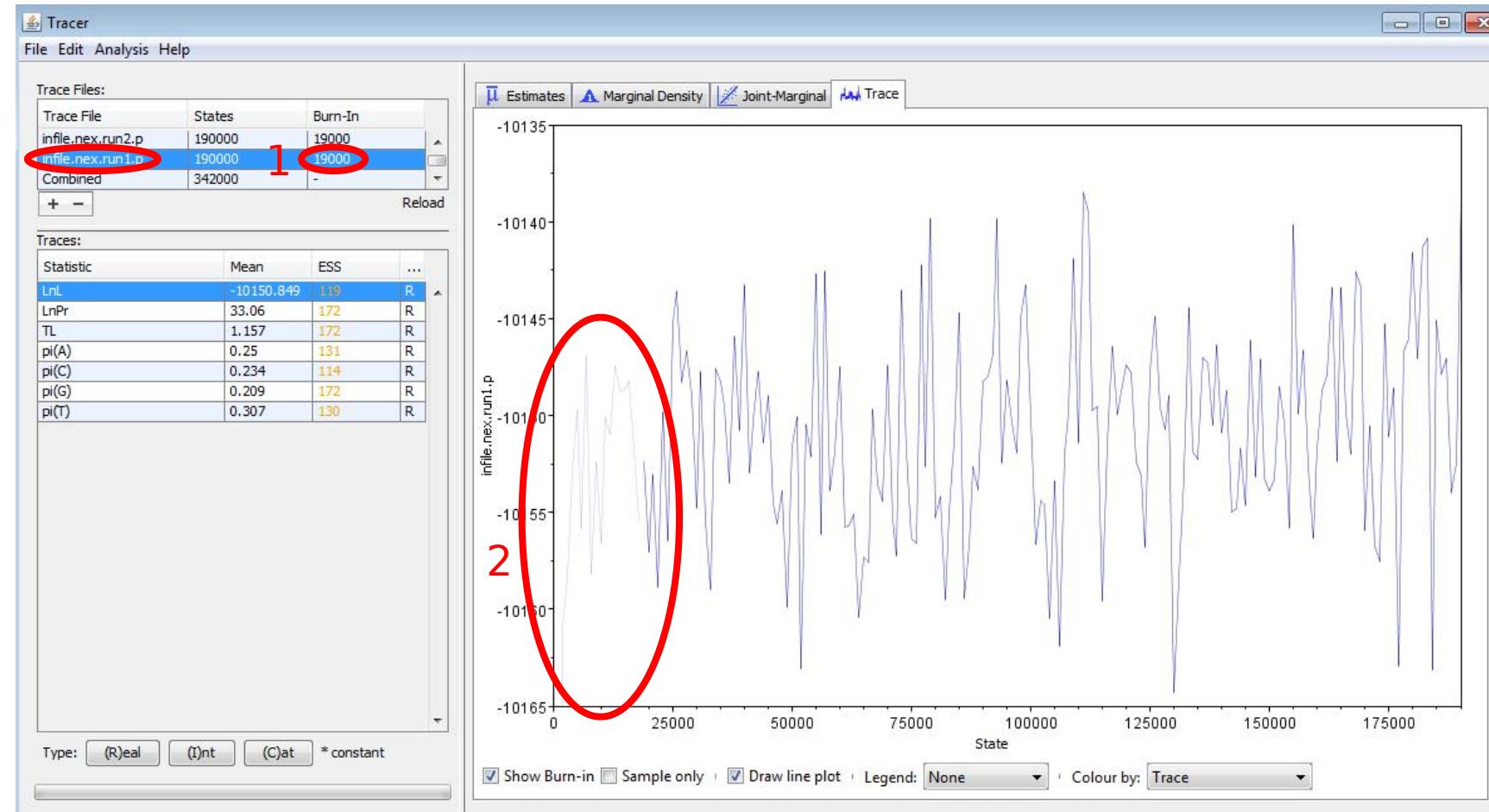
Tracer will help you refine the value you set for the burn-in in the Bayesian analysis (Sump, Sumt).

For that purpose, you want to observe a 'caterpillar' shape which is a plateau in the search of the tree space (the case on the picture on the left here)

1: Select 'Combined' to visualise both run at the same time

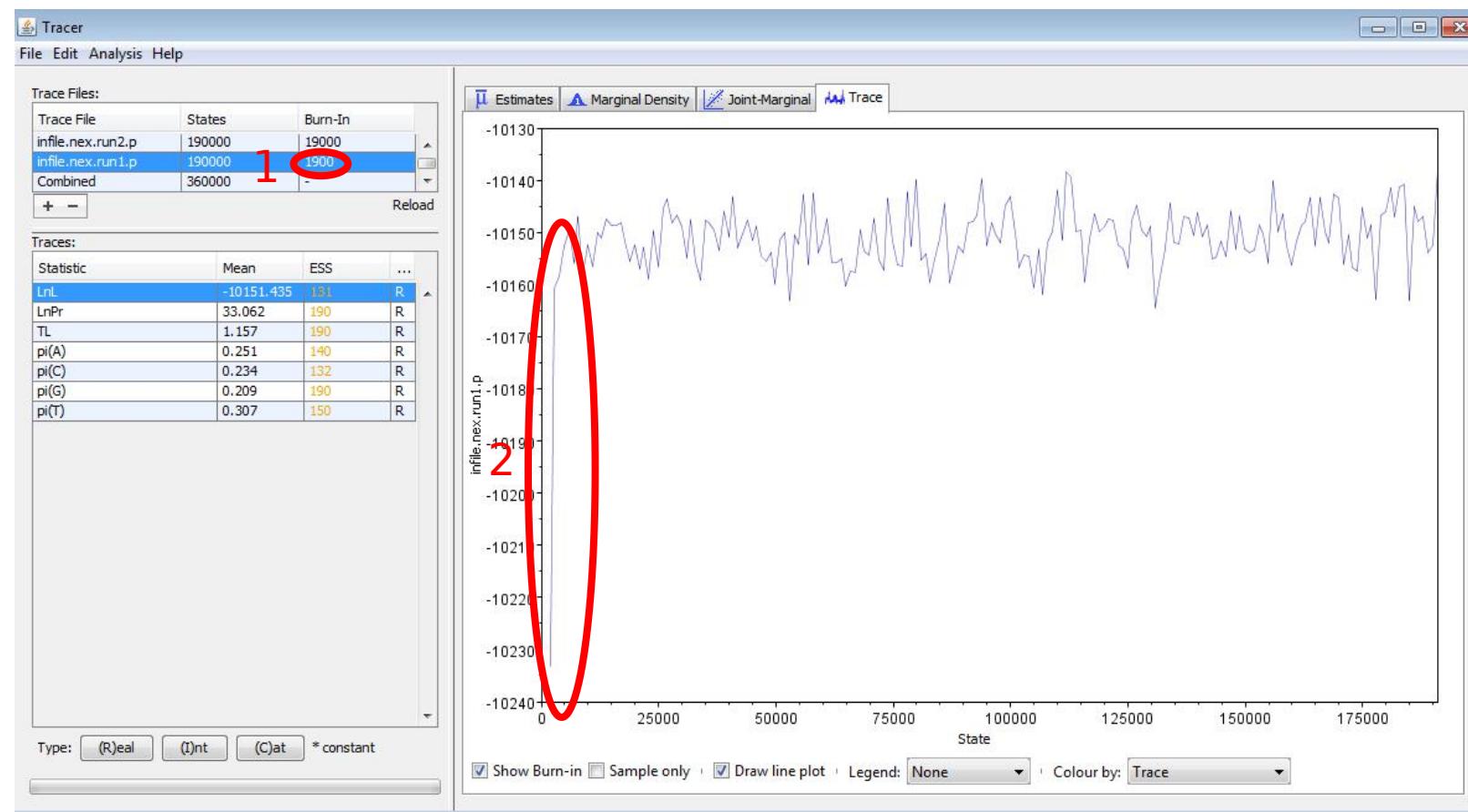
2: Select the tab 'Trace'

Tracer (3)



Analyse the trace of each two runs separately and play with the Burn-in value (1) to see the effect on the trace (2). Here, the 10% burn-in value discard correctly the first trees of the analysis which are too distant from the locally optimal tree.

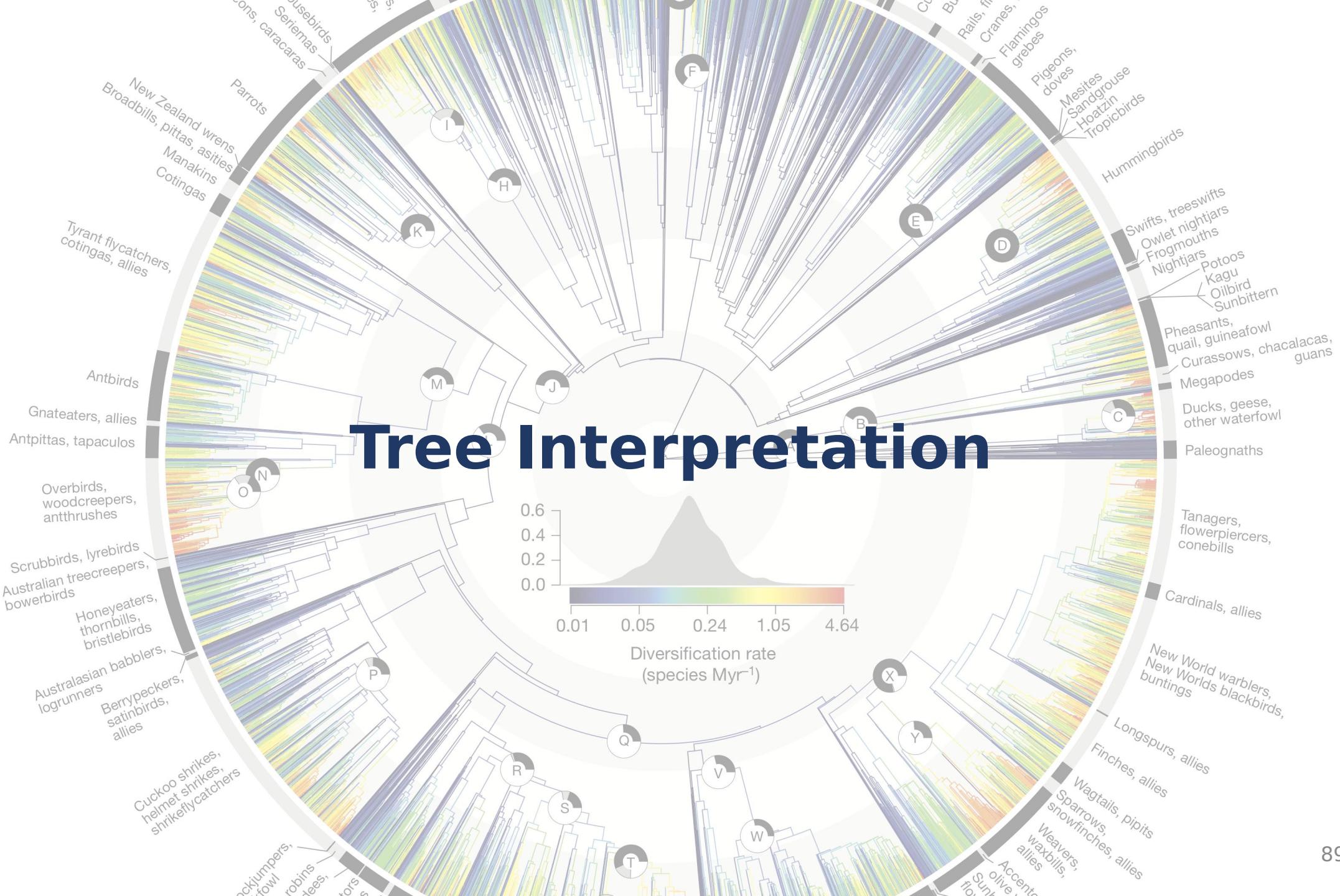
Tracer (4)



For the example, we reduced here the burn-in value to 1% (1) to highlight a bad result for the trace (2).

NB: a too high value for the burn-in is not a big problem but checking the trace allows to detect too low values which might affect the confidence values of the tree and even the topology.

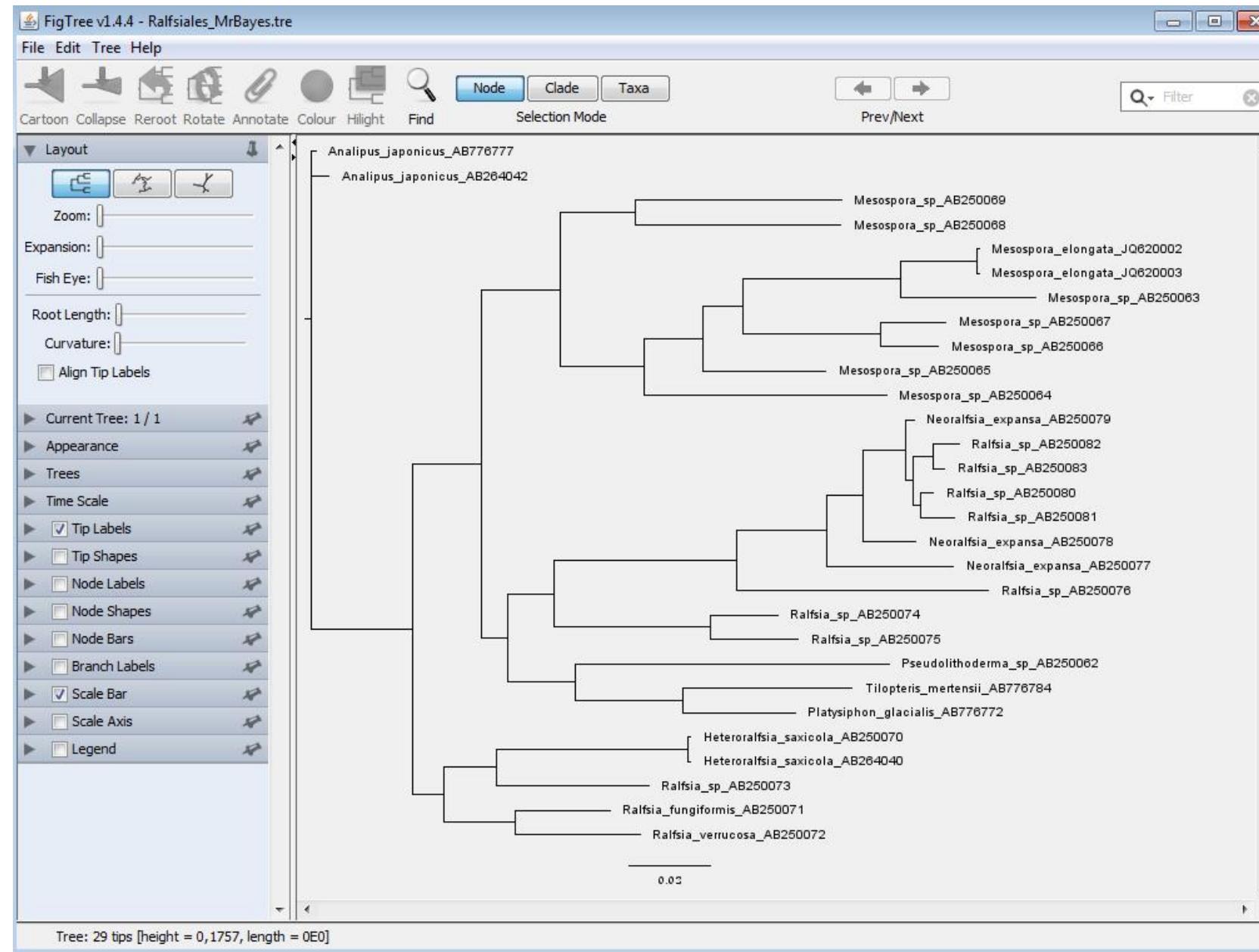
Tree Interpretation



Figtree (1)



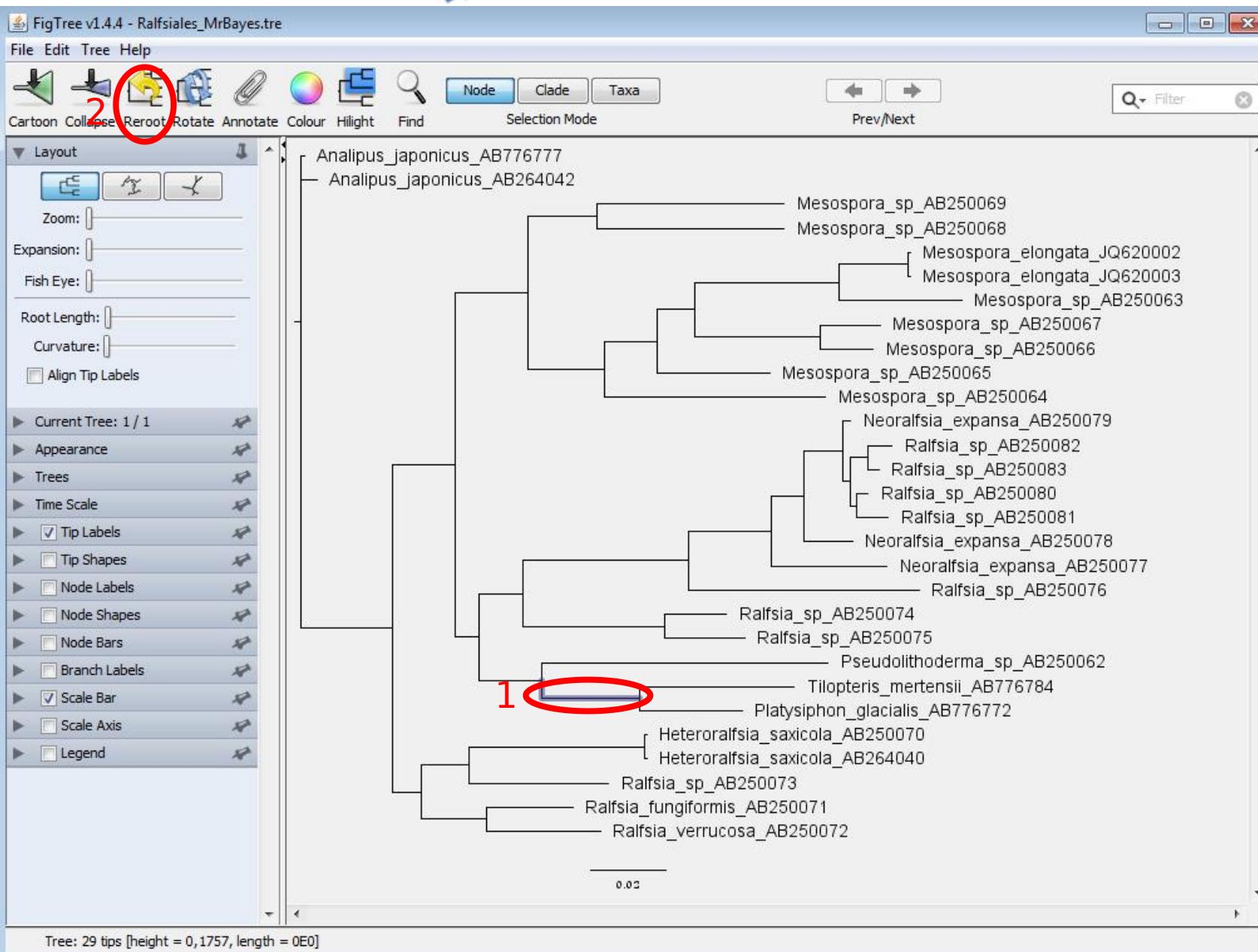
newick file (.nwk or .tre)



Figtree is a tool to visualise phylogenetic trees but does not add any information to the trees produced. The following steps are then only visualisation of the data and presentation for their publication.

Load your tree in Figtree (here I loaded the Bayesian tree produced during the practical)

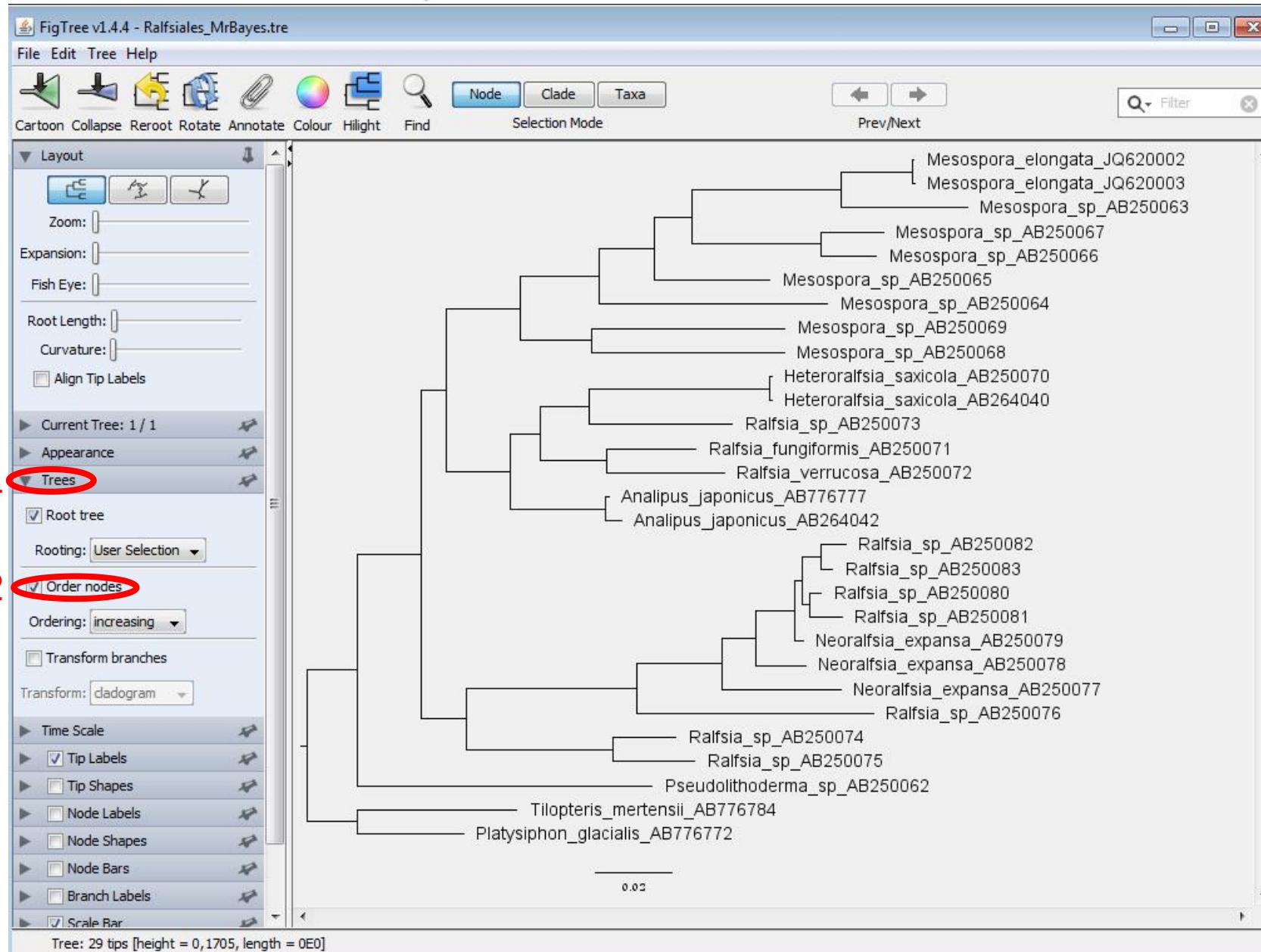
Figtree (2)



1: Select the branch leading to your taxa used as outgroups

2: Click on Reroot

Figtree (3)



1: Click on the 'Trees' menu

2: Thick the 'Order nodes' box
to have a ladderised tree
(easier to read)

Figtree (4)



FigTree v1.4.4 - Ralfsiales_MrBayes.tre

File Edit Tree Help

Cartoon Collapse Reroot Rotate Annotate Colour Hilight Find Node Clade Taxa Selection Mode Prev/Next Filter

Layout

Zoom: Expansion: Fish Eye: Root Length: Curvature: Align Tip Labels

Current Tree: 1 / 1

Appearance

Trees

Time Scale

1 Tip Labels

Display: Names

Colour by: User selection

Font Size: 14

Setup: Colour Font

Format: Decimal

Sig. Digits: 4

Tip Shapes

Node Labels

Node Shapes

Node Bars

Branch Labels

Scale Bar

Scale Axis

Tree: 29 tips [height = 0,1705, length = 0E0]

0.05

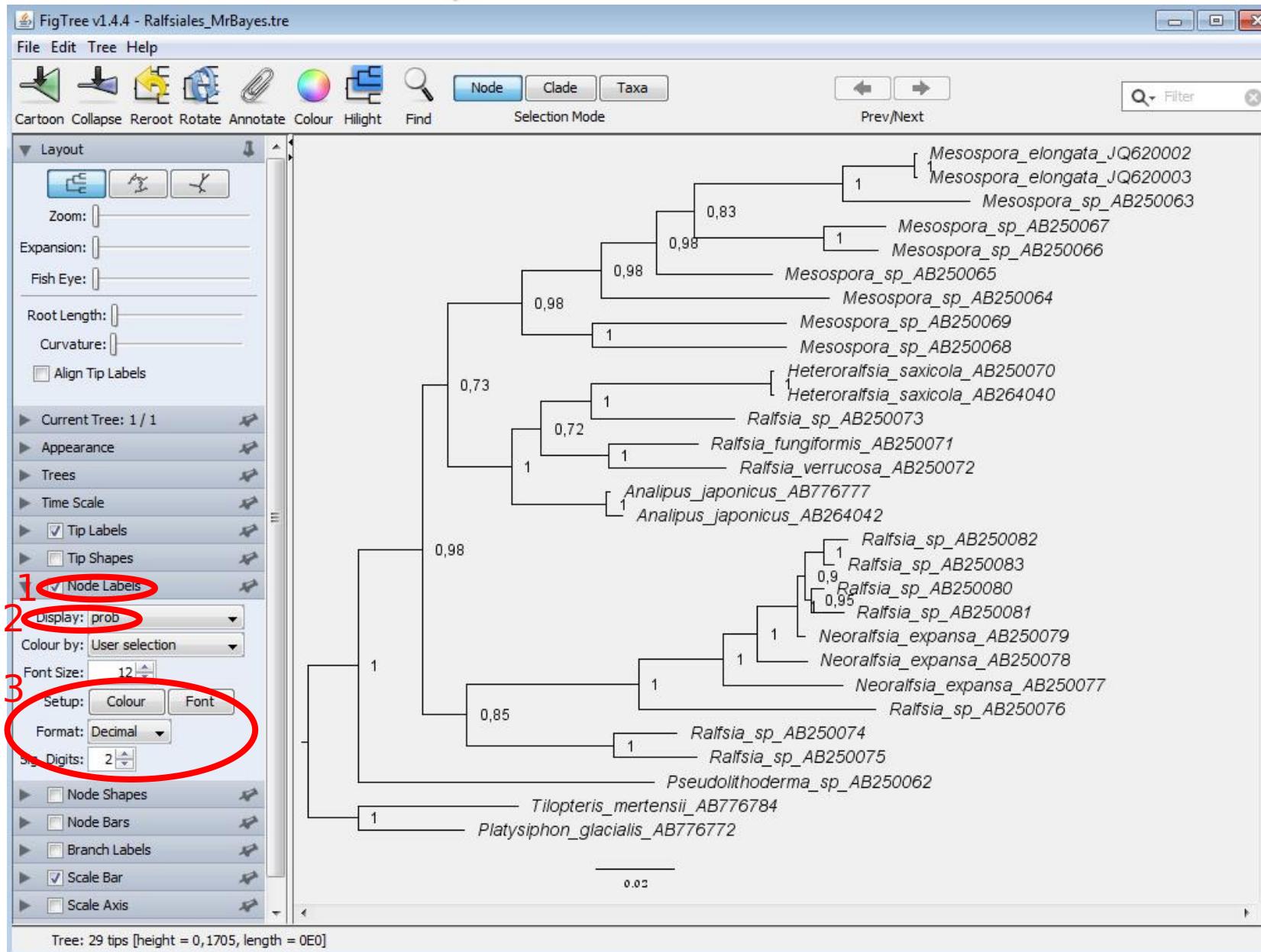
Mesospora_elongata_JQ620002
Mesospora_elongata_JQ620003
Mesospora_sp_AB250063
Mesospora_sp_AB250067
Mesospora_sp_AB250066
Mesospora_sp_AB250065
Mesospora_sp_AB250064
Mesospora_sp_AB250069
Mesospora_sp_AB250068
Heteroralfsia_saxicola_AB250070
Heteroralfsia_saxicola_AB264040
sp_AB250073
formis_AB250071
errucosa_AB250072
76777
3264042
Ralfsia_sp_AB250082
Ralfsia_sp_AB250083
Ralfsia_sp_AB250080
Ralfsia_sp_AB250081
Neoralfsia_expansa_AB250079
Neoralfsia_expansa_AB250078
Ralfsia_sp_AB250077
Ralfsia_sp_AB250074
Ralfsia_sp_AB250075
Pseudolithoderma_sp_AB250062
Tilopteris_mertensii_AB776784
Platysiphon_glacialis_AB776772

You can change the font of the taxon names by:

1: Going in the 'Tip Labels' menu

2: Changing the 'Font'

Figtree (5)



In order to display the confidence value of each nodes of your tree:

1: Go to the menu 'Node Labels' and thick its box

2: In 'Display', select 'prob'

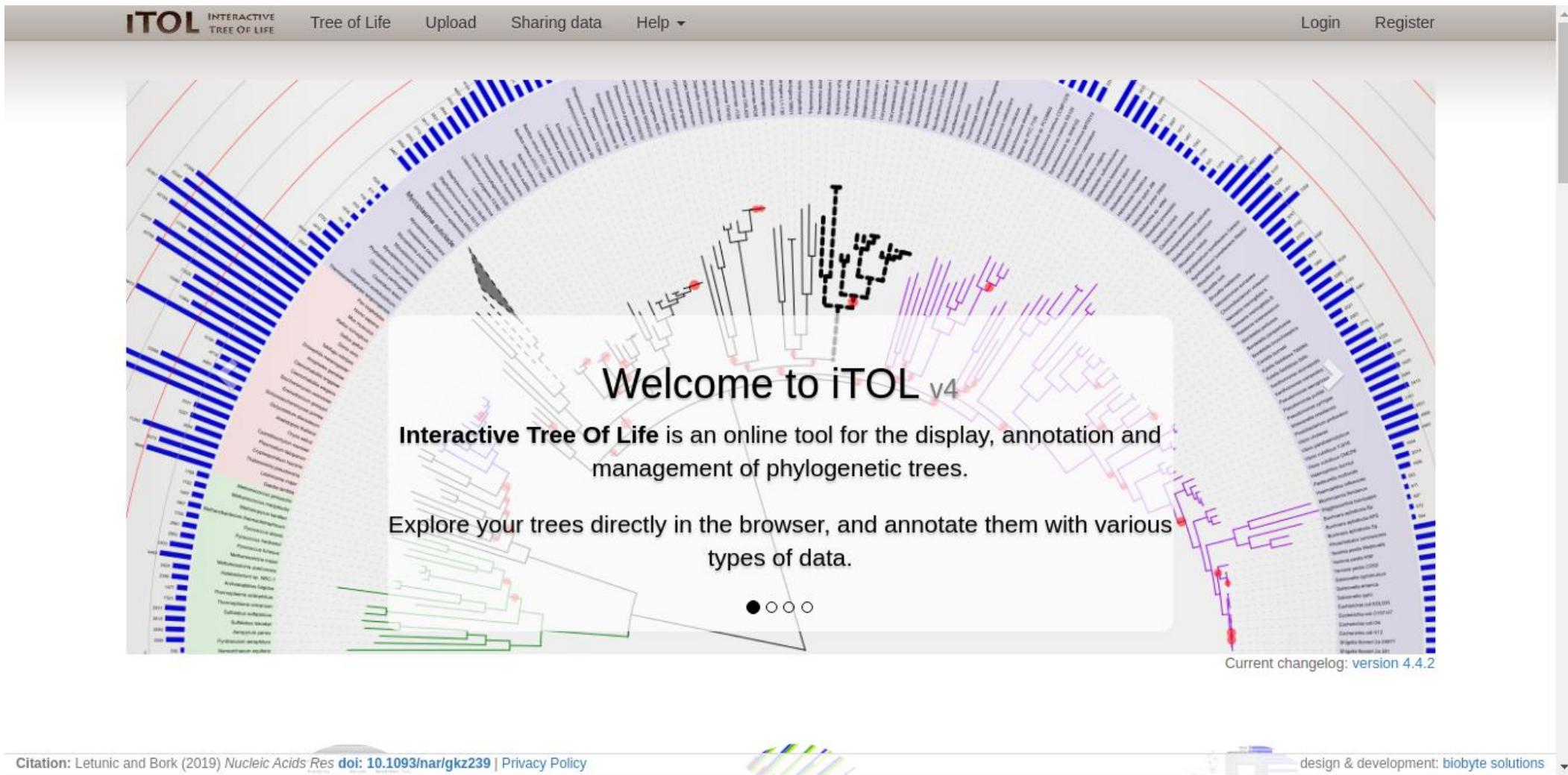
3: Change the font and number of digits to make it more readable.

NB: don't hesitate to explore more features of Figtree

Interactive Tree of Life

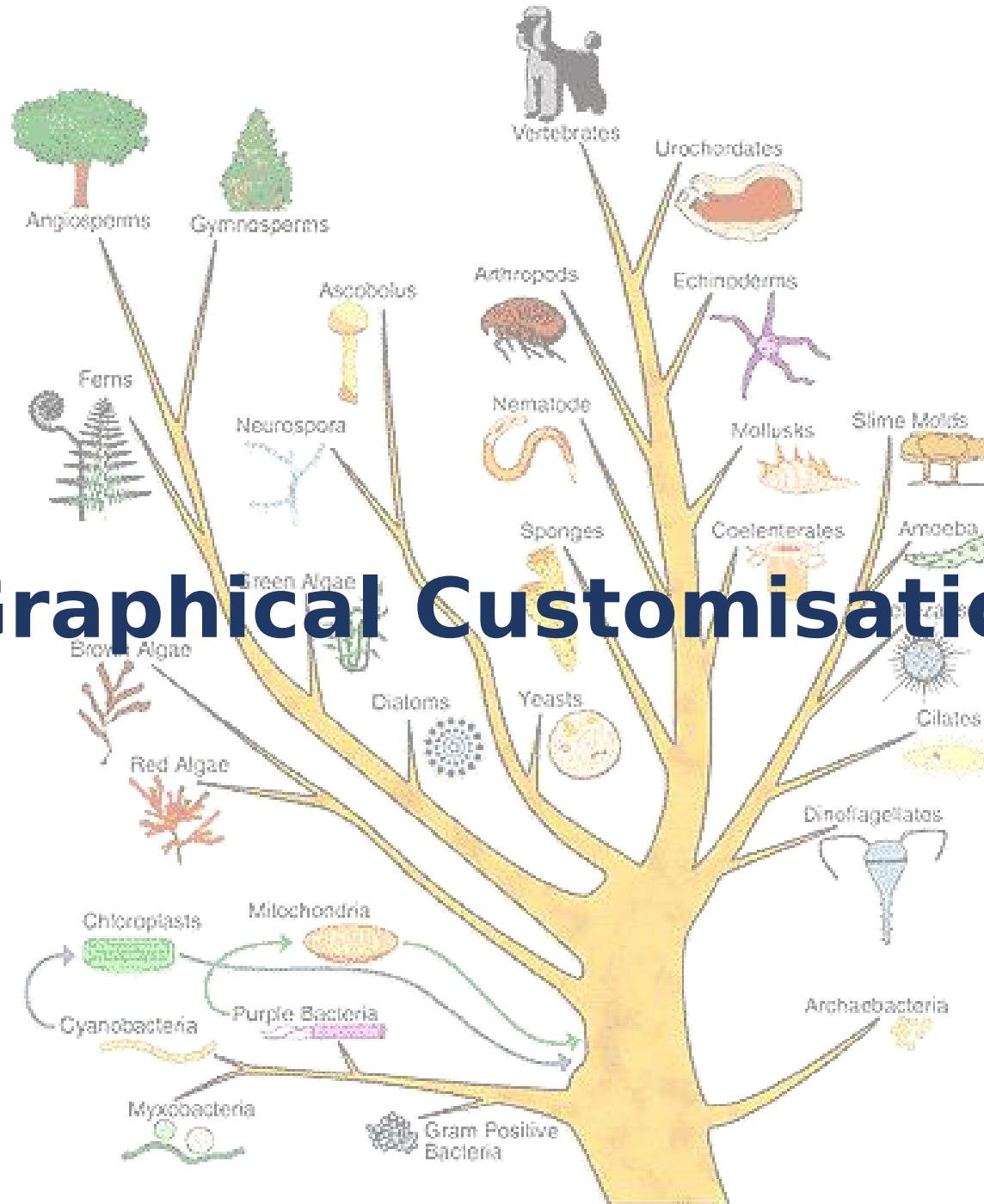
newick file (.nwk or .tre)

<https://itol.embl.de/>

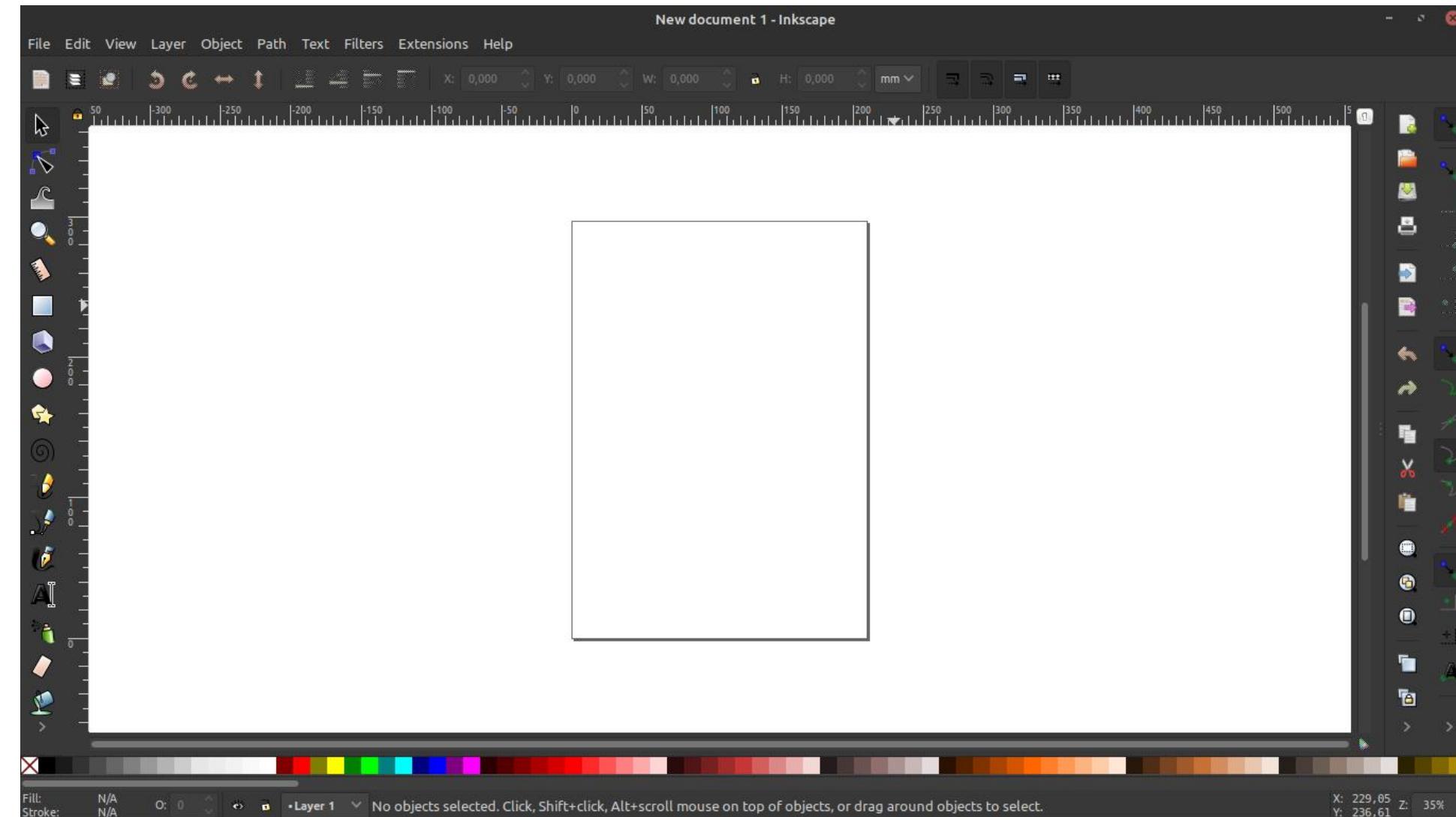


Good online alternative to Figtree

Graphical Customisation

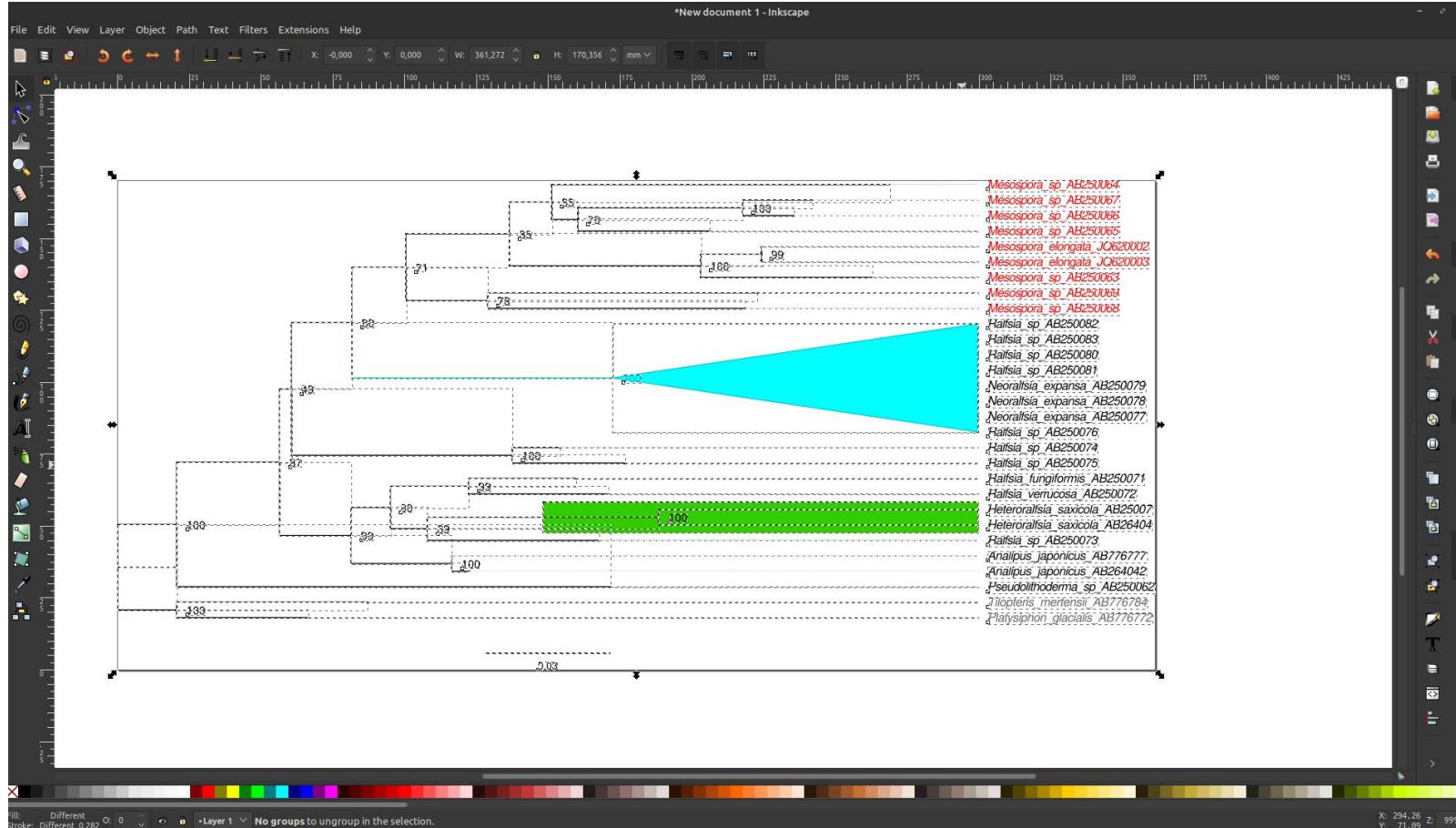


Vectorial drawings with Inkscape



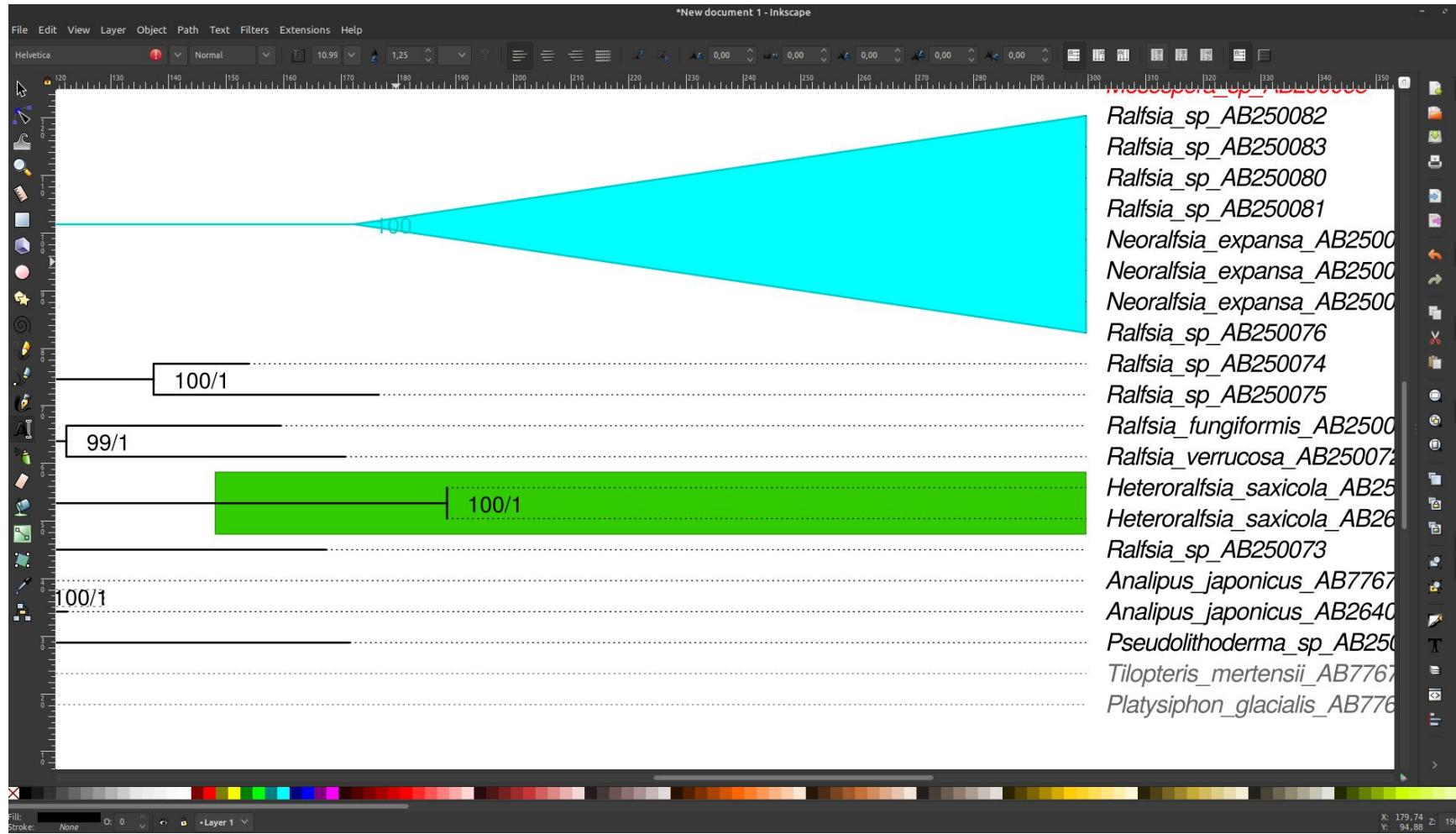
Inkscape is one of the possible drawing tools (gimp, photoshop, etc.) that you can use to further improve the design of your tree after using Figtree. It is recommended to use vectorial files (pdf, svg) to preserve the editability and resolution of your tree.

Vectorial drawings with Inkscape



This tutorial doesn't mean to be exhaustive on vectorial drawing but the following hint will be useful in the framework of phylogenies: when you load your tree (svg or pdf) in inkscape, right click on it and ungroup it (or **CTRL+shift+G**) (may need to be repeated several times) to access the different elements of the tree.

Vectorial drawings with Inkscape



Inkscape was presented during the course to highlight the necessity to add the bootstrap values next to the posterior probabilities on the tree (the choice of the topology is free but has to be mentioned in your publication).

How to reference the different tools used

see 'References_to_programs.txt' on the github:

<https://github.com/PierrotVdAa/PhylogenyAberdeen2019>