

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

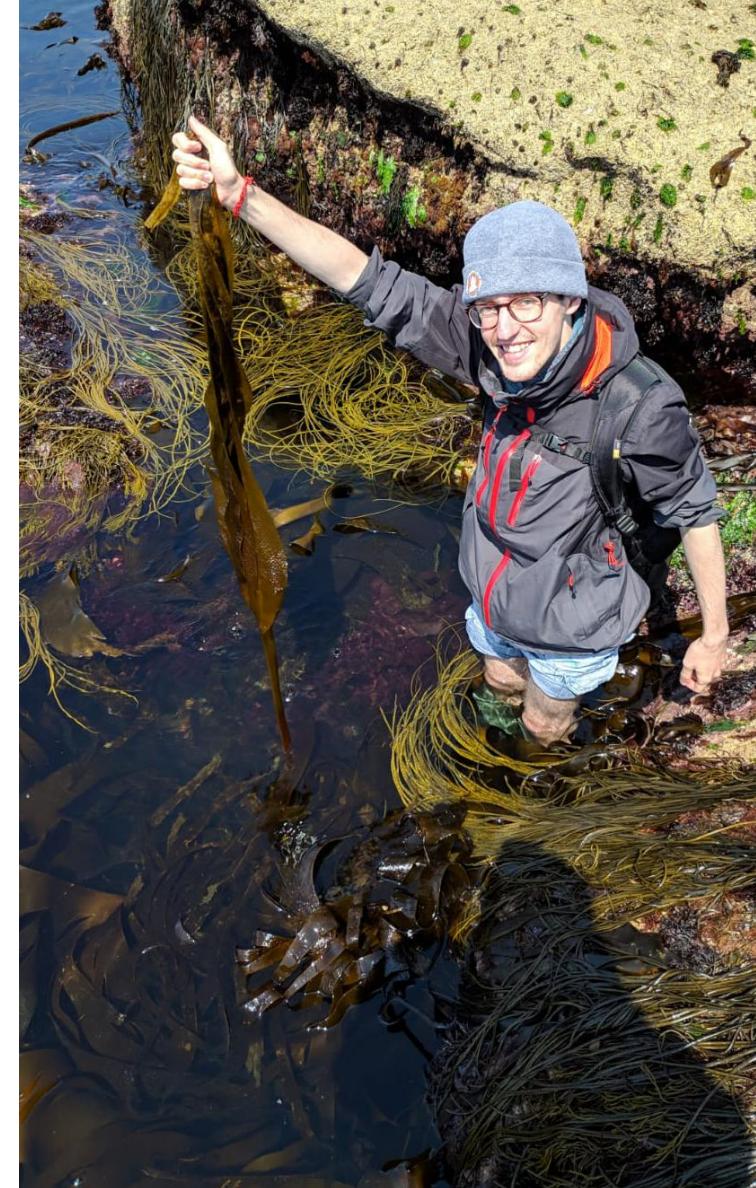


UNIVERSITY
OF ABERDEEN



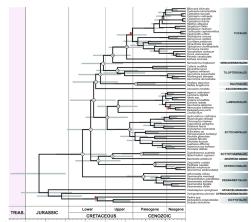
Introducing yourself

- Name
- Background
- Current project
- Dataset
- OS
- Expectations

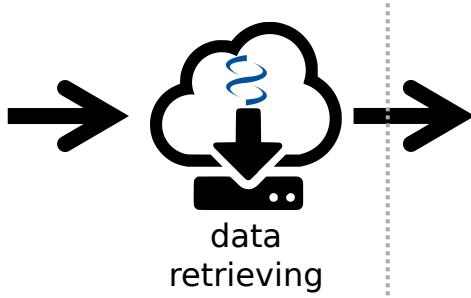


Program of the course

Day 1



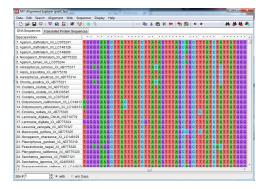
introduction to phylogeny



Day 2

```
>Sequence 1 assembled  
CCCTAAACCTTAACATCCCAAGCCTAACCTCTGA  
CTTAAATCCATACATCCCAAGCCTAACCTCTGA  
CTCTGGTGAAATGATTTATATAATGATAAT  
TTCTGGTGATGATTTATATAATGATAAT  
GGTTTCCTTCCTTGATGATTTATATAATGATAAT  
GCTTTCTTGATGATTTATATAATGATAAT  
ATCGTGTGATGATTTATATAATGATAAT  
<Sequence 2>  
CCCTAAACCTTAACATCCCAAGCCTAACCTCTGA  
CTTAAACCTTAACATCCCAAGCCTAACCTCTGA  
CTCTGGTGAAATCATGGTGTATATAATGATAAT  
TTCTGGTGATGATTTATATAATGATAAT  
GGTTTCCTTCCTTGATGATTTATATAATGATAAT  
GCTTTCTTGATGATTTATATAATGATAAT  
ATCGTGTGATGATTTATATAATGATAAT
```

cleaning



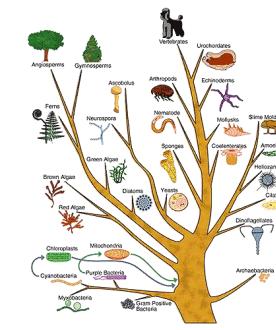
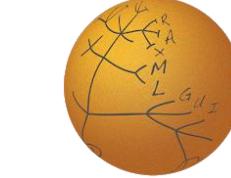
Day 3



Bayesian analysis



maximum likelihood analysis

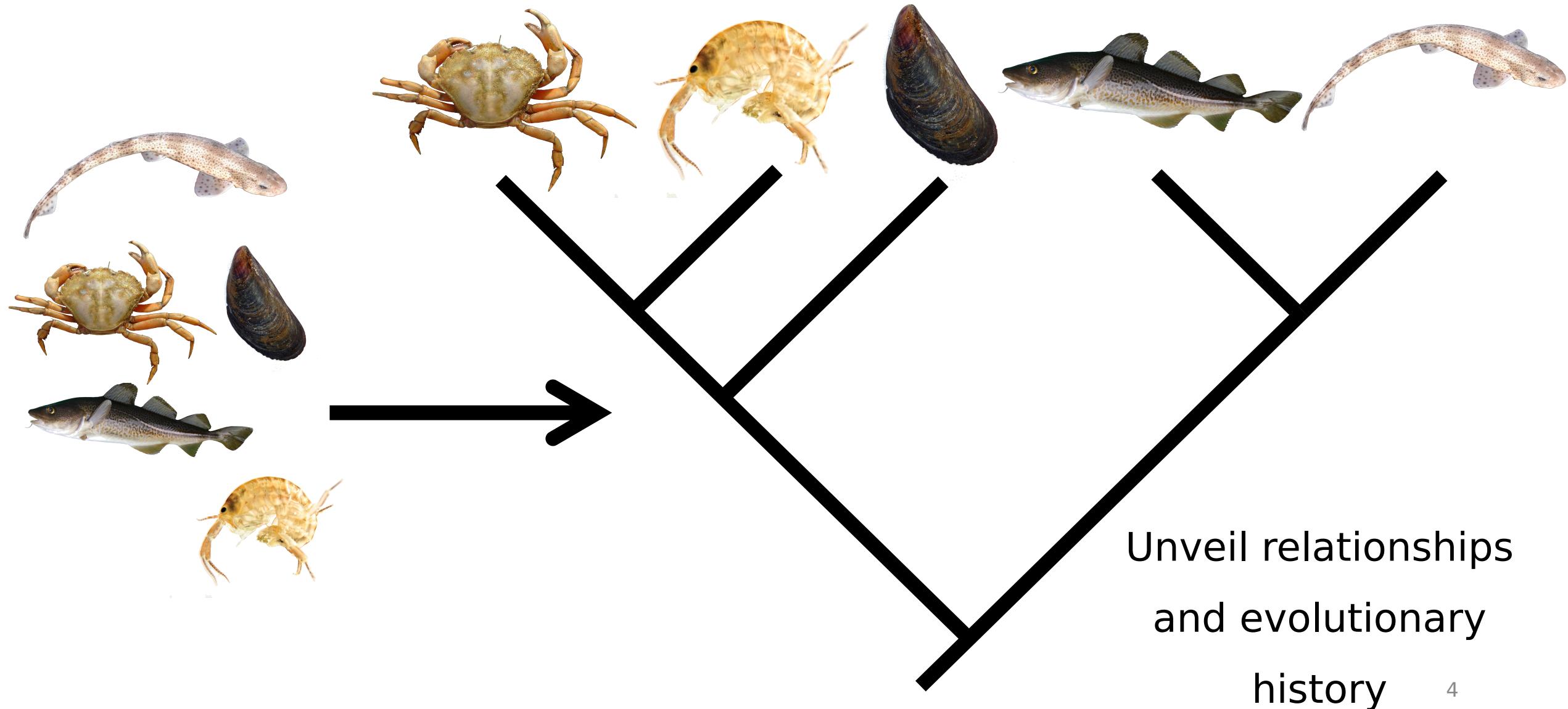


Day 4

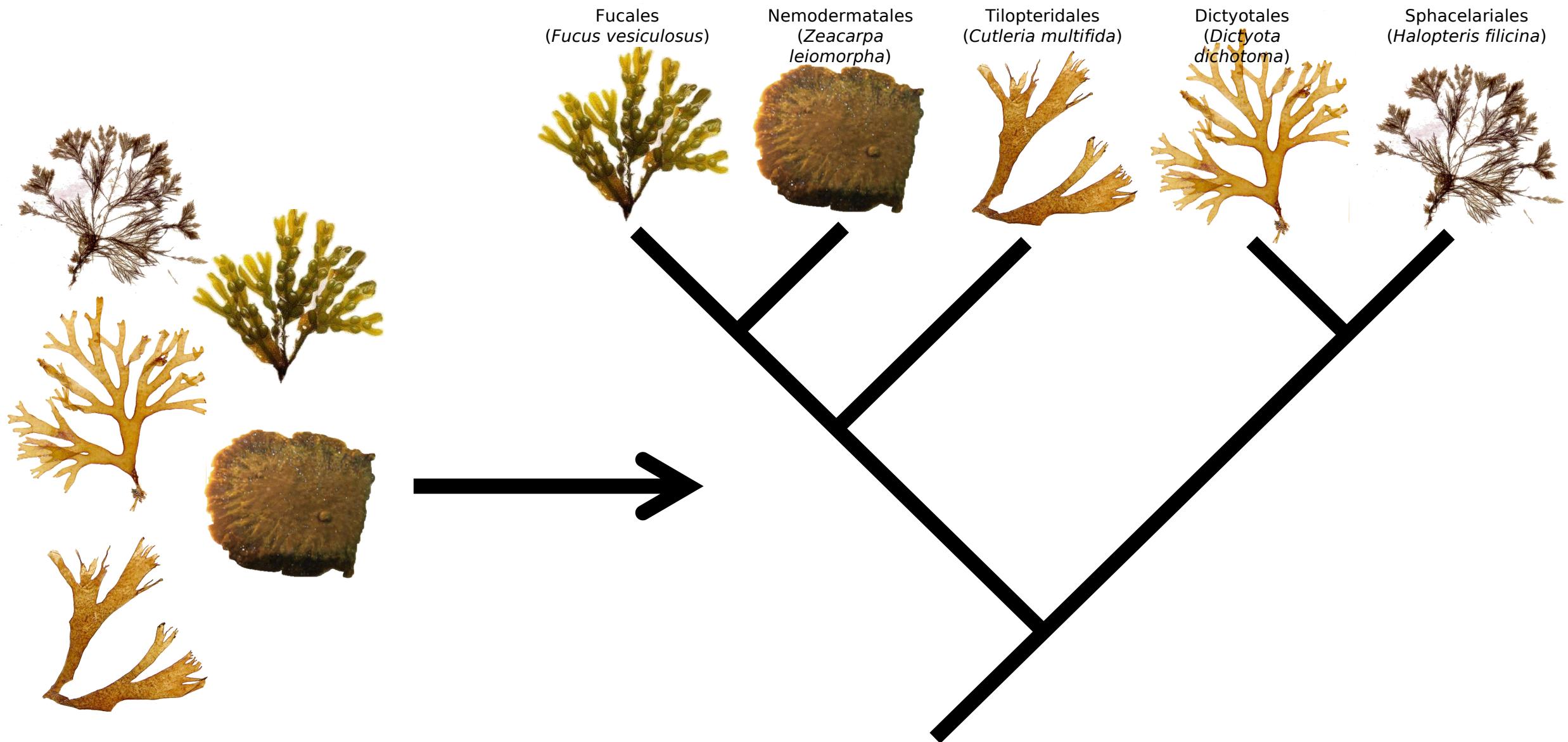


BLAST, CD-hit,
Q&A ...

What is phylogeny?



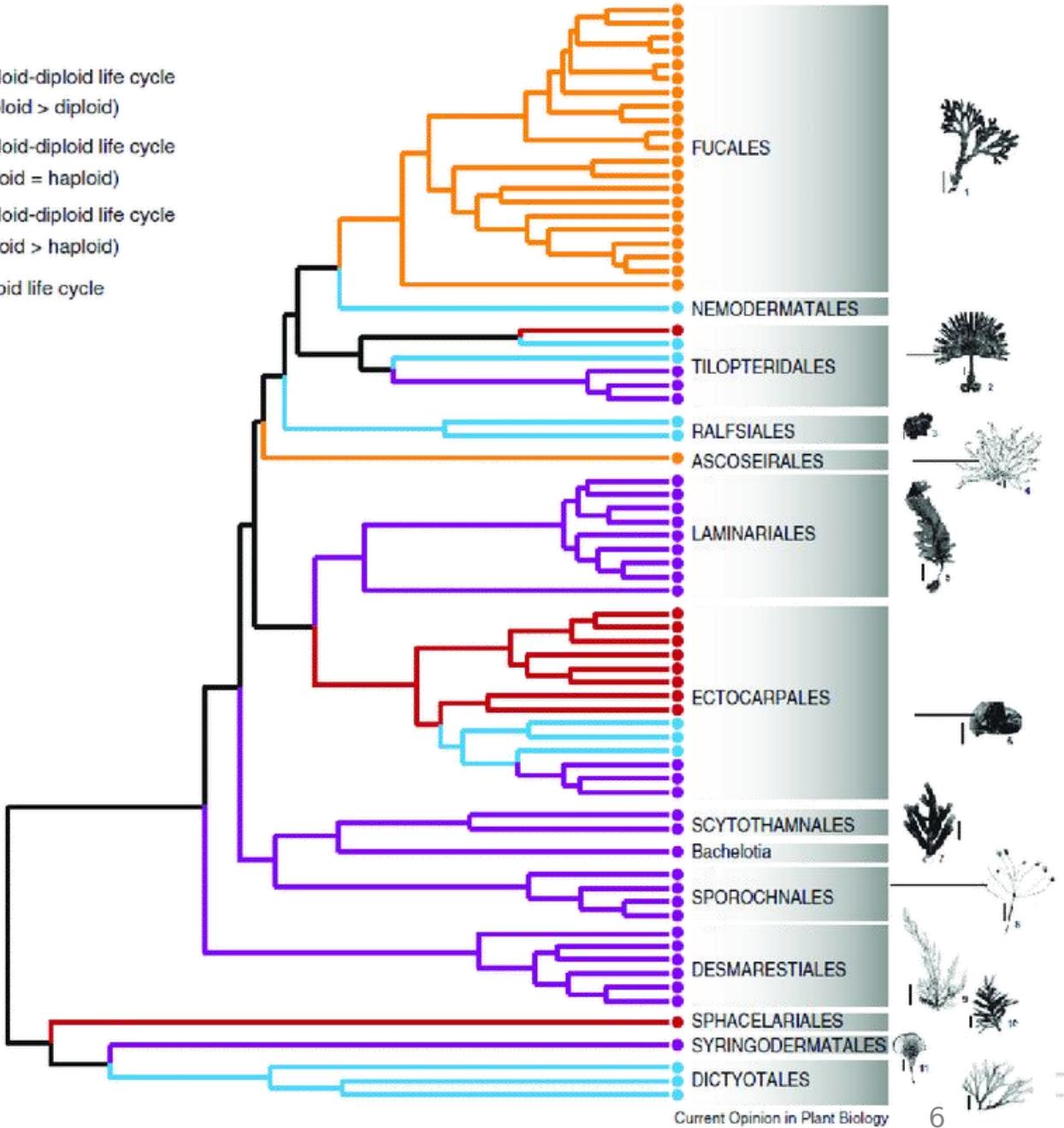
Why apply phylogeny to brown seaweeds?



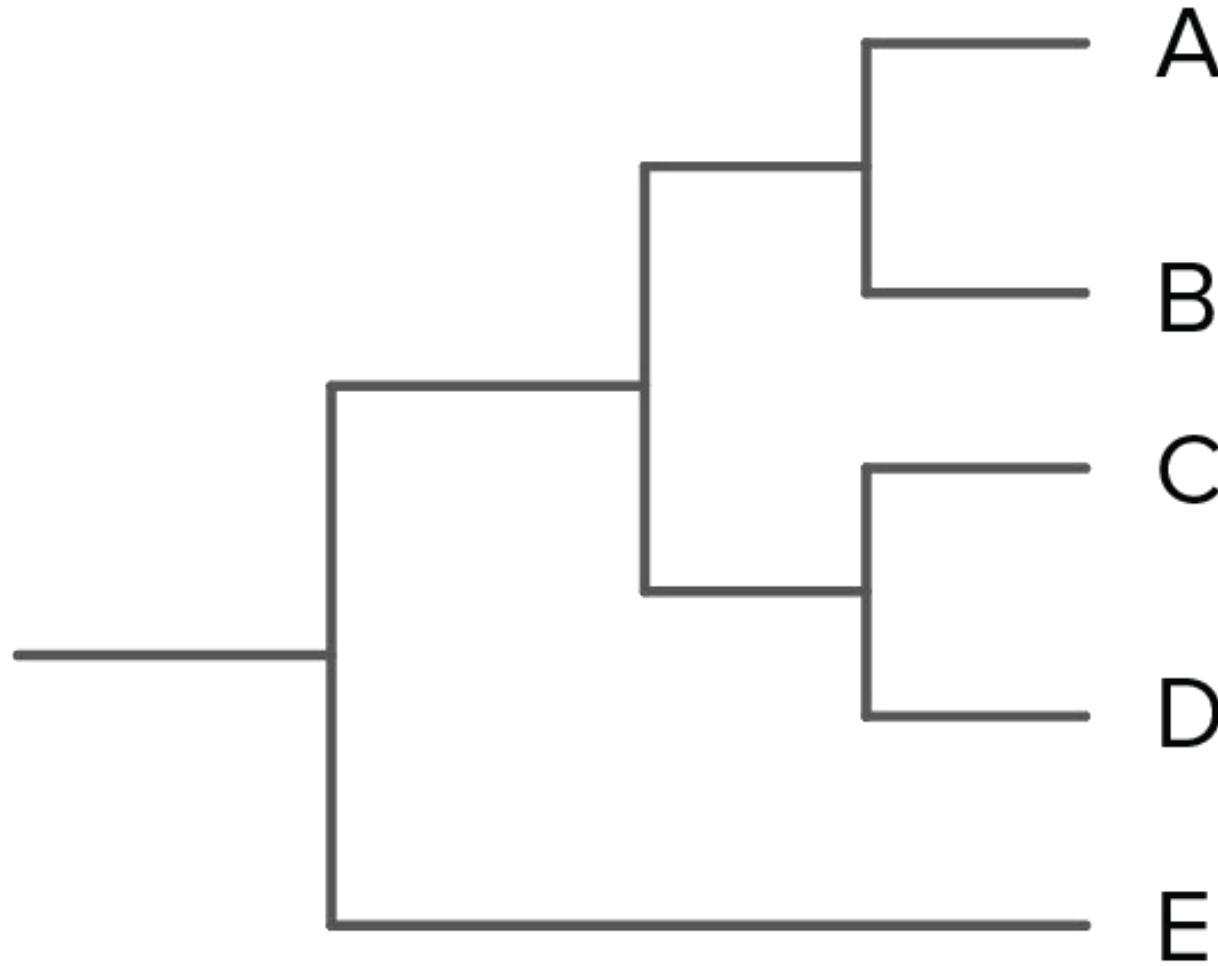
Uses of phylogeny

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

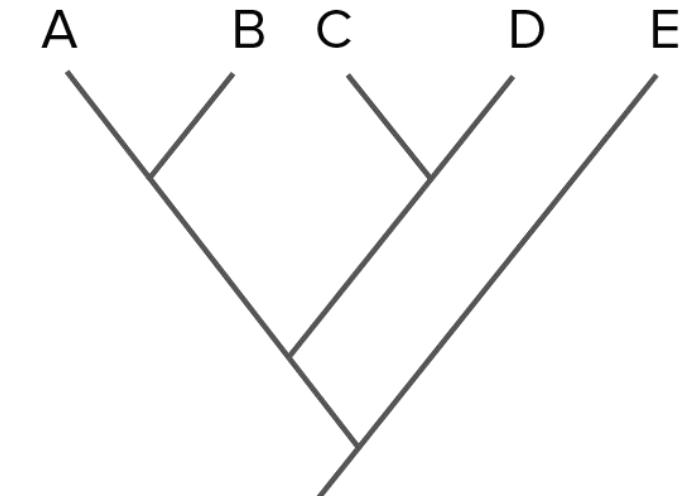
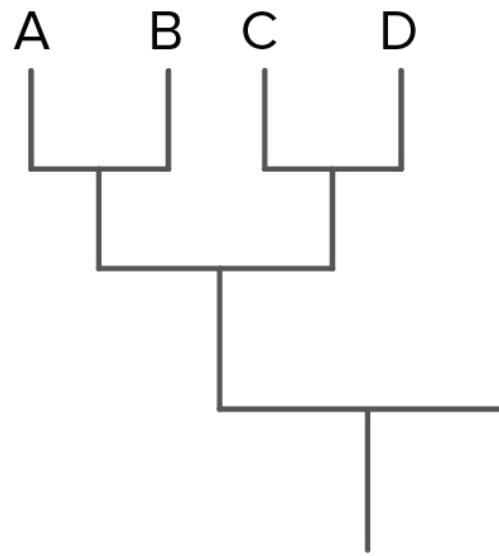
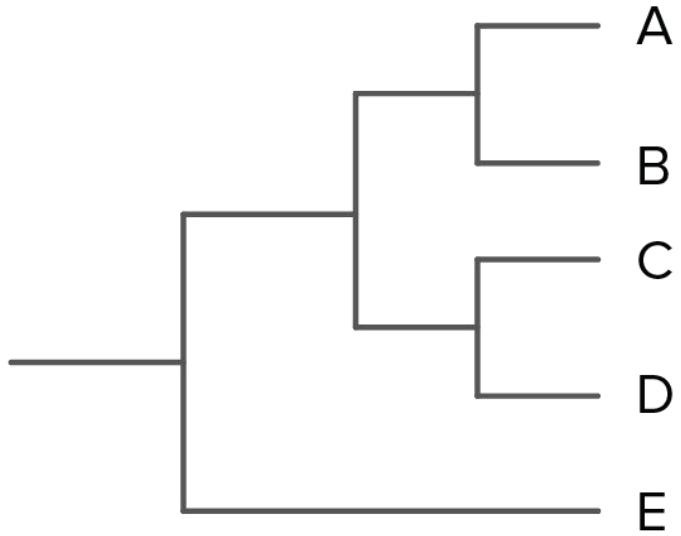
- Haplod-diploid life cycle (haploid > diploid)
- Haplod-diploid life cycle (diploid = haploid)
- Haplod-diploid life cycle (diploid > haploid)
- Diploid life cycle



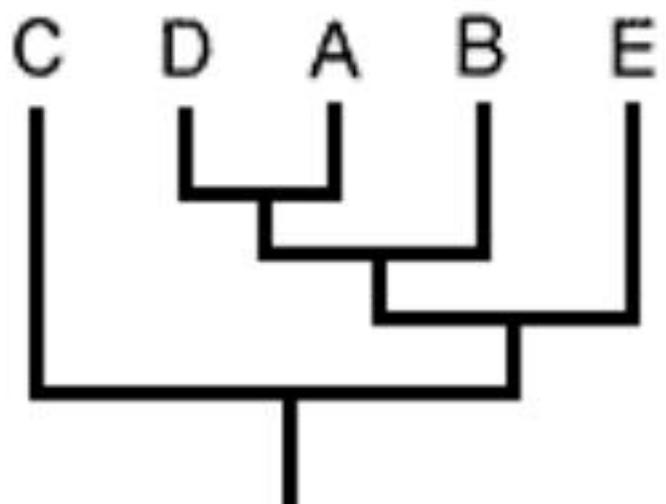
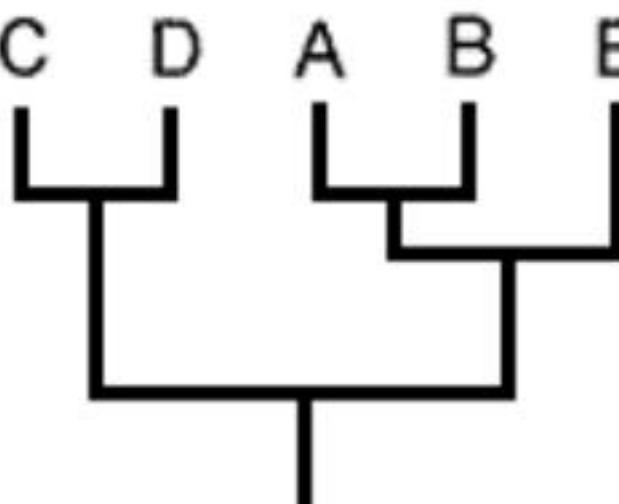
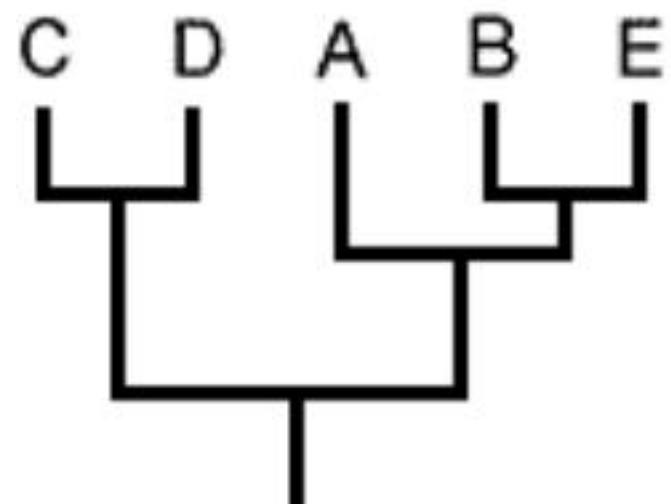
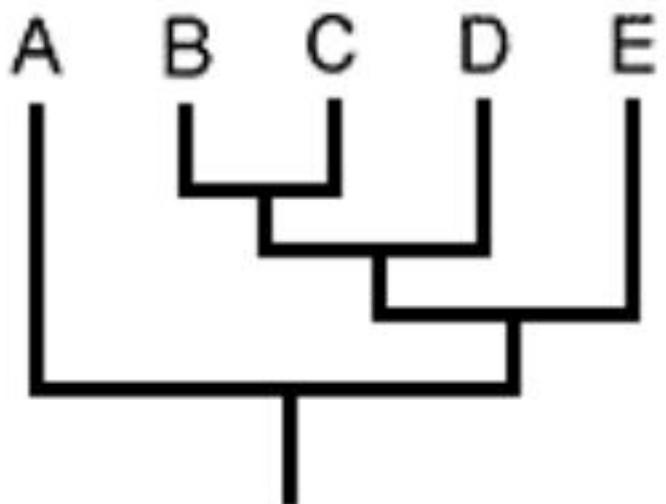
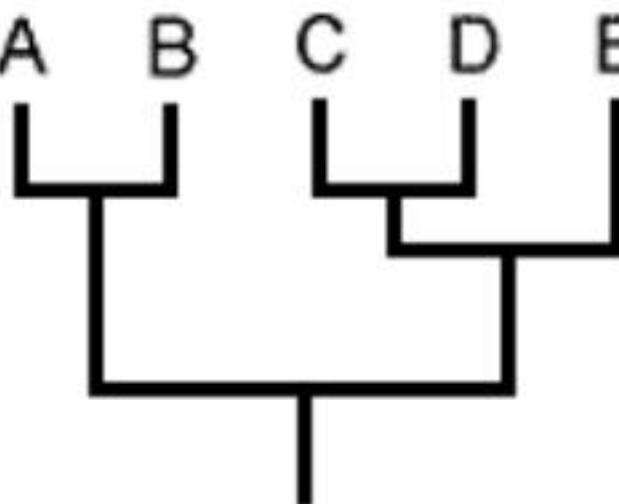
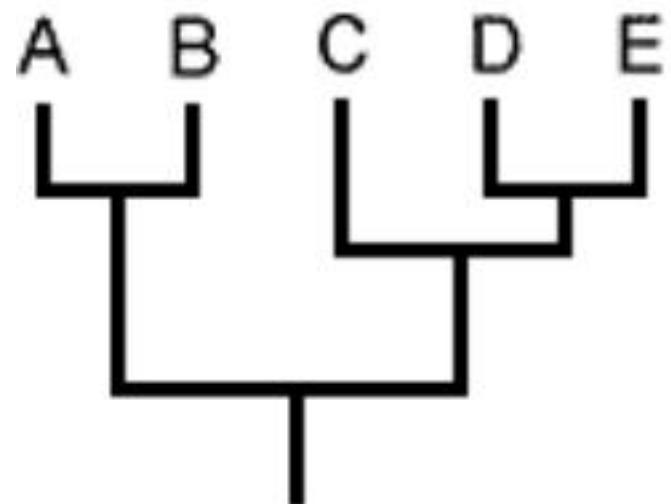
Reading a tree



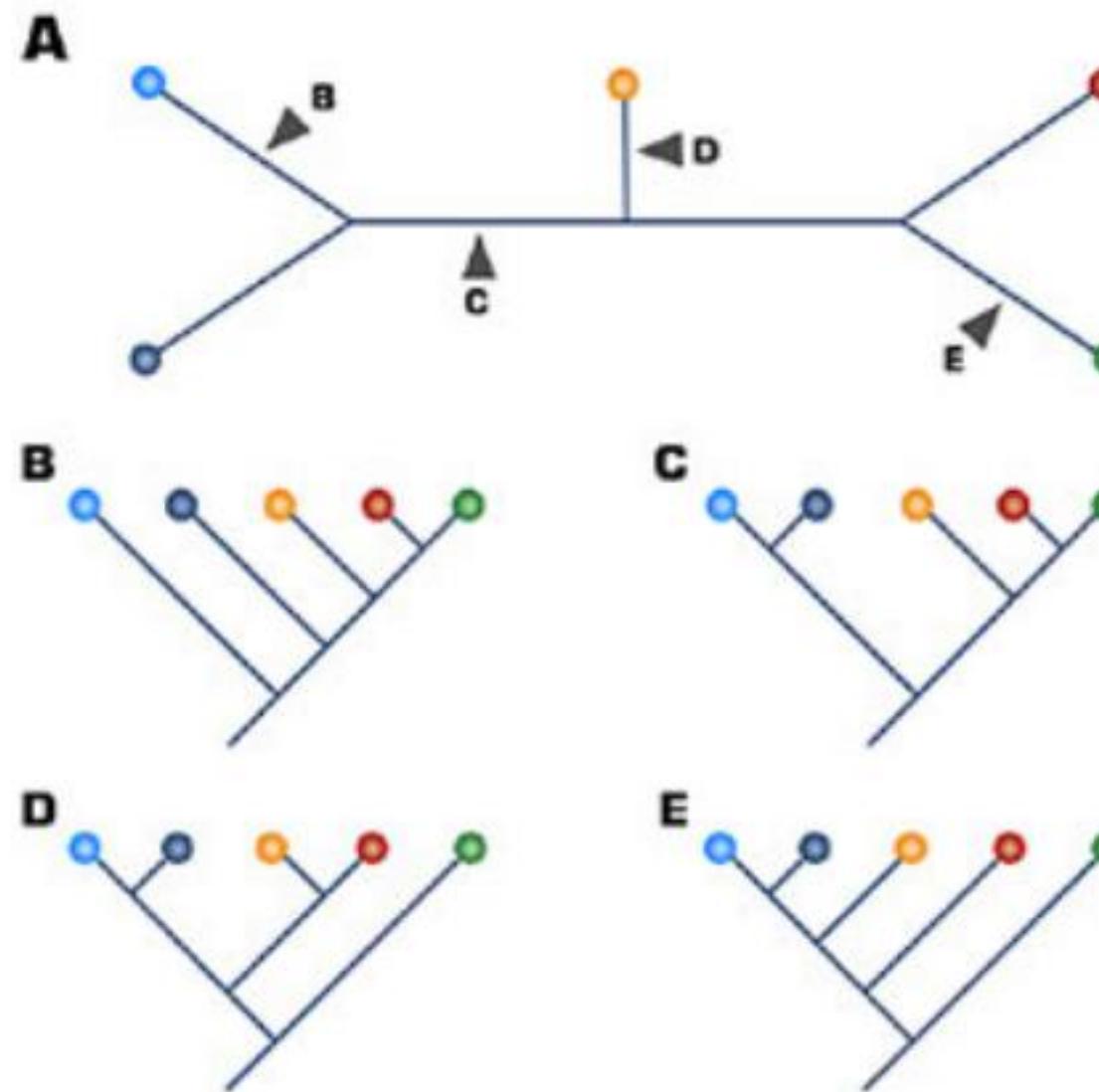
Are those trees different or the same?



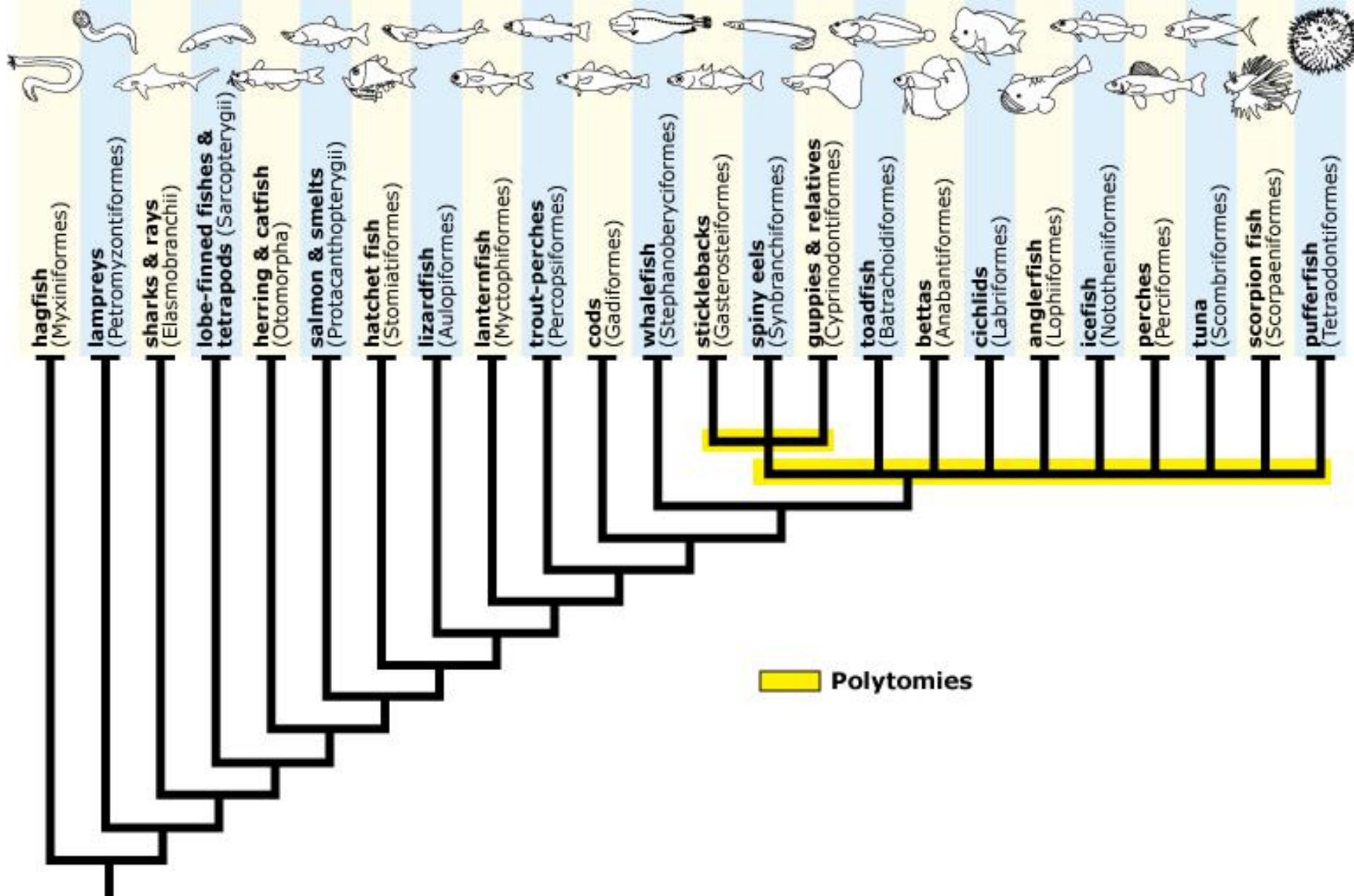
Are those trees different or the same?



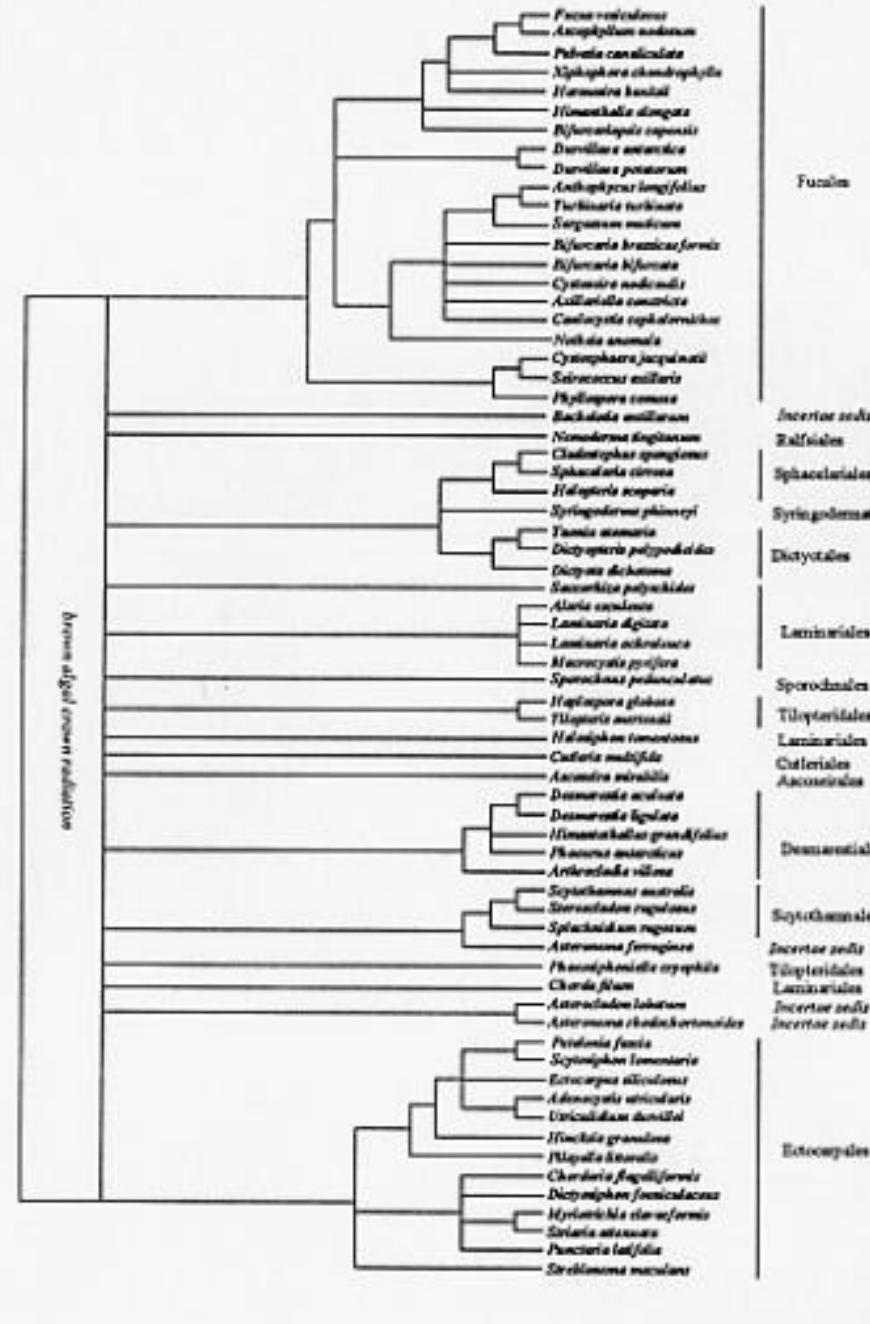
Rooted or not



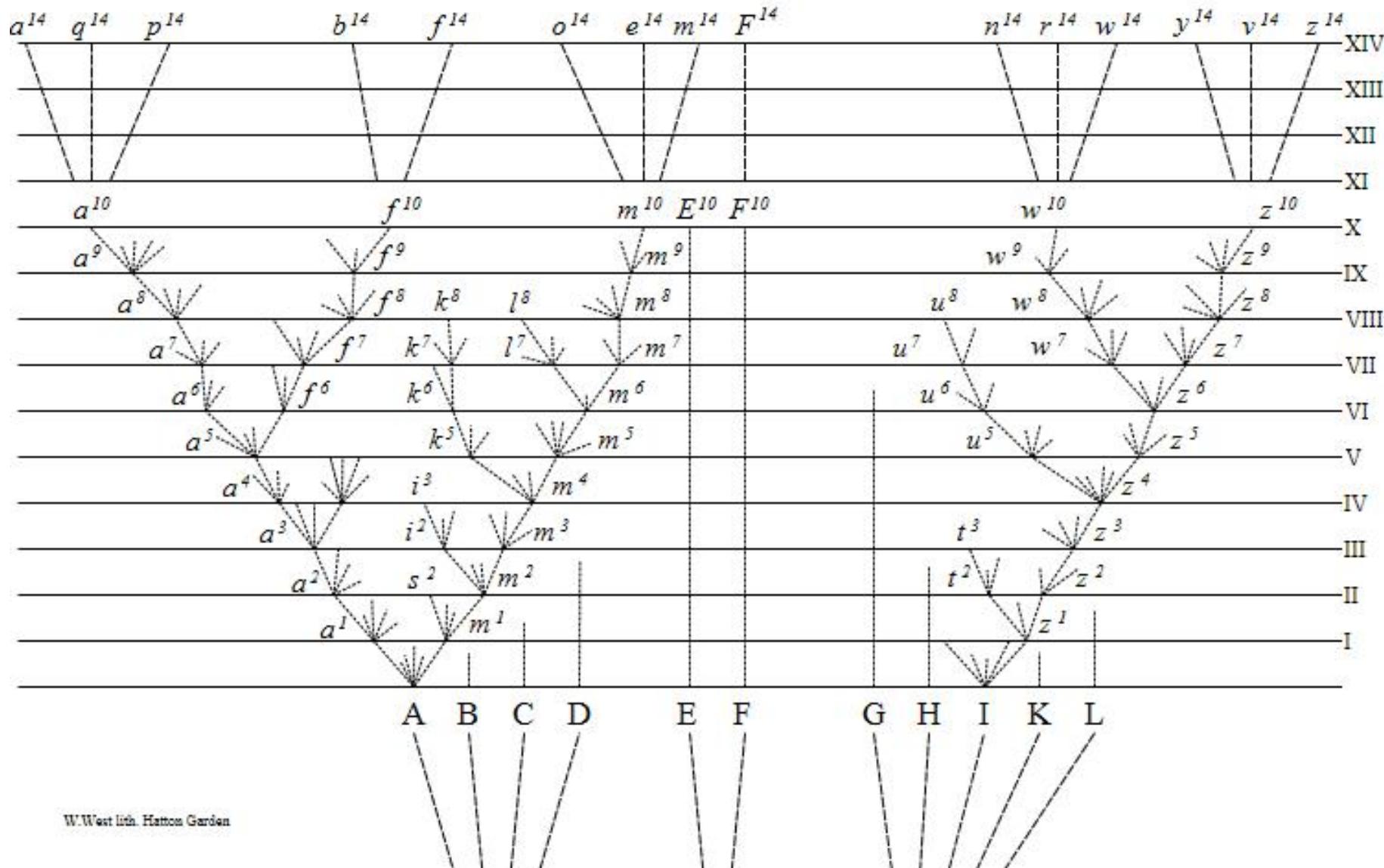
Polytomies



Polytomies

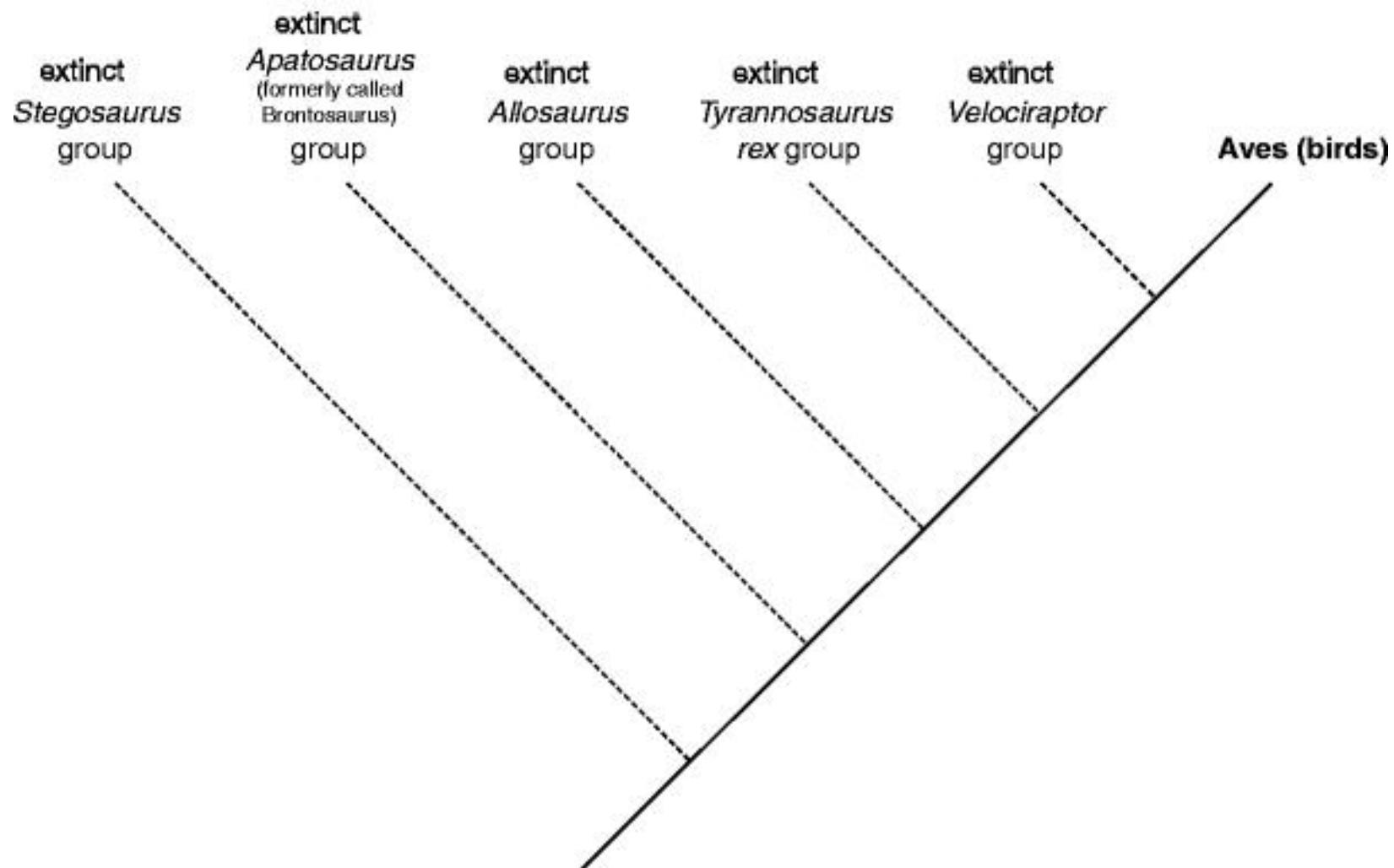


Practical: critical analysis of trees



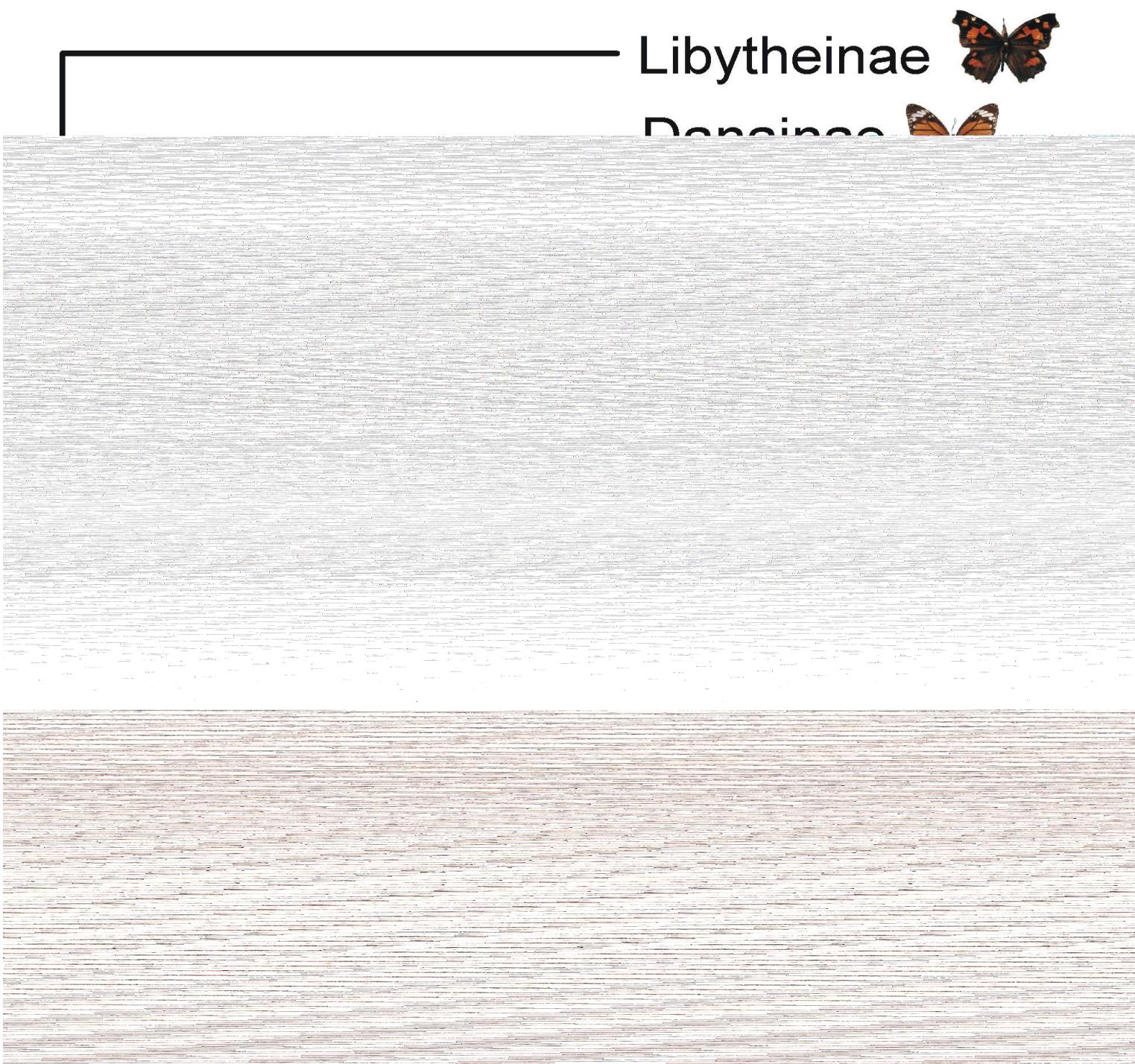
A simple tree

A simplified phylogenetic tree for the dinosaurs



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

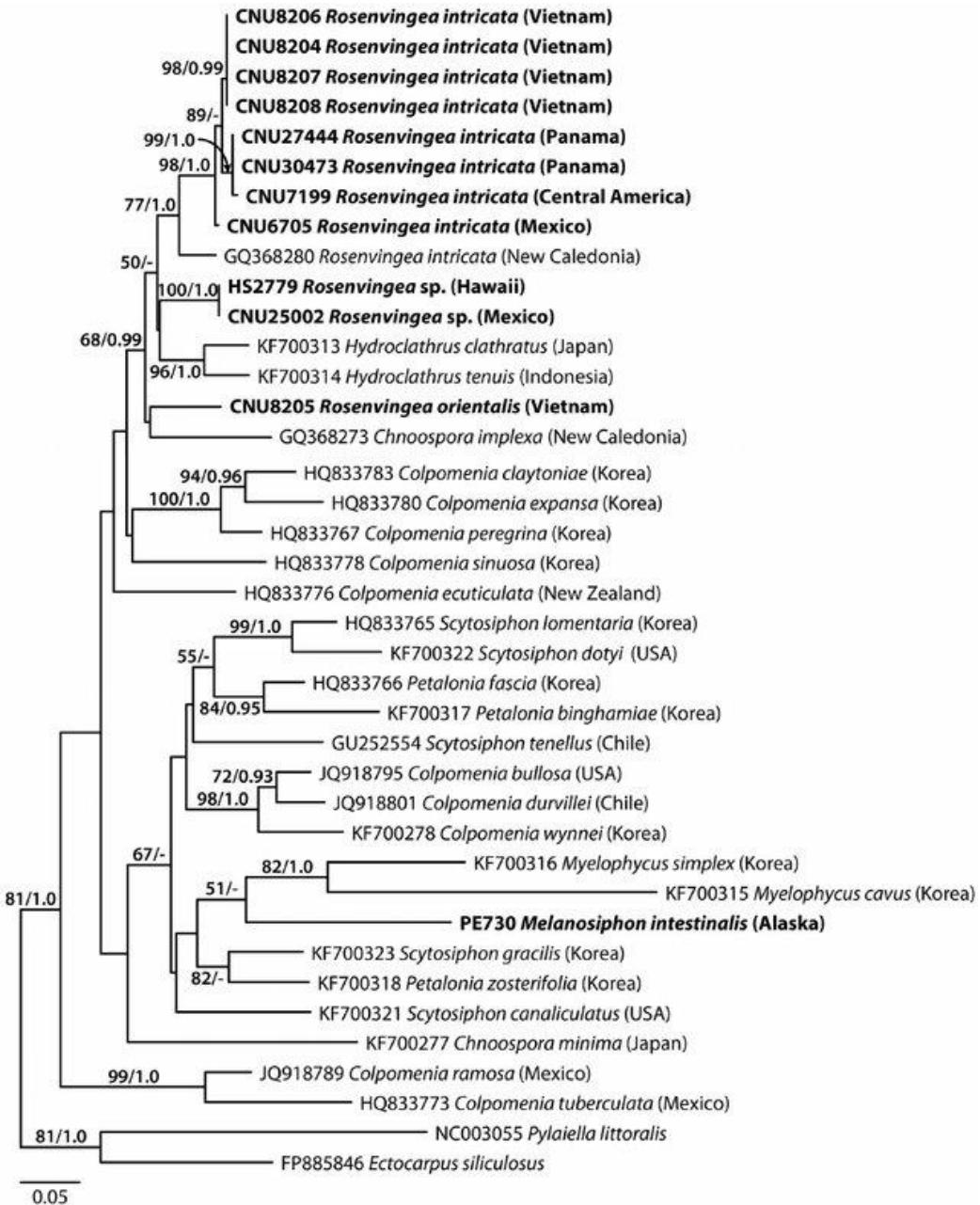
A simple tree



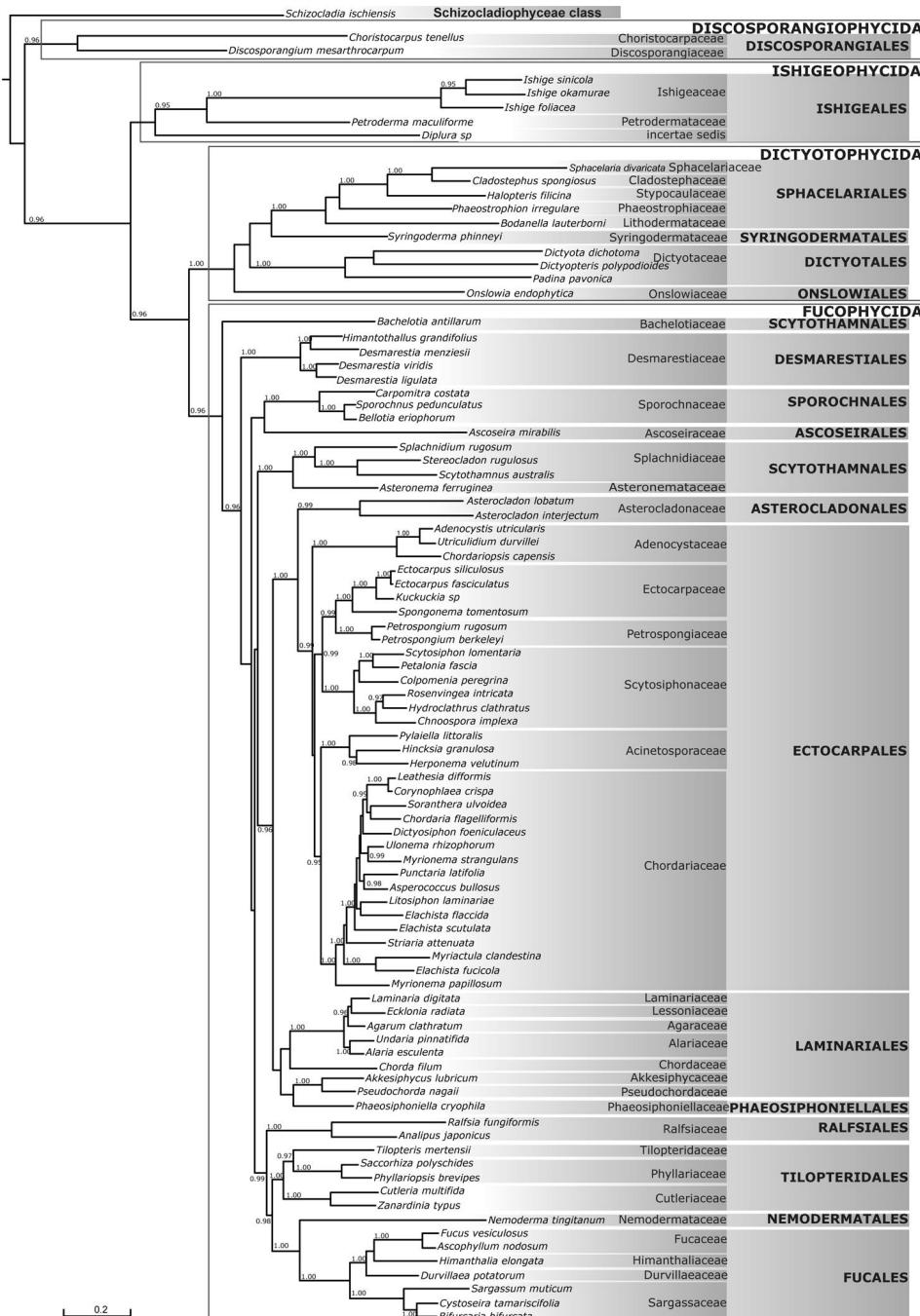
Species delineation



Reliability of the relationships



Practise



Time calibrated

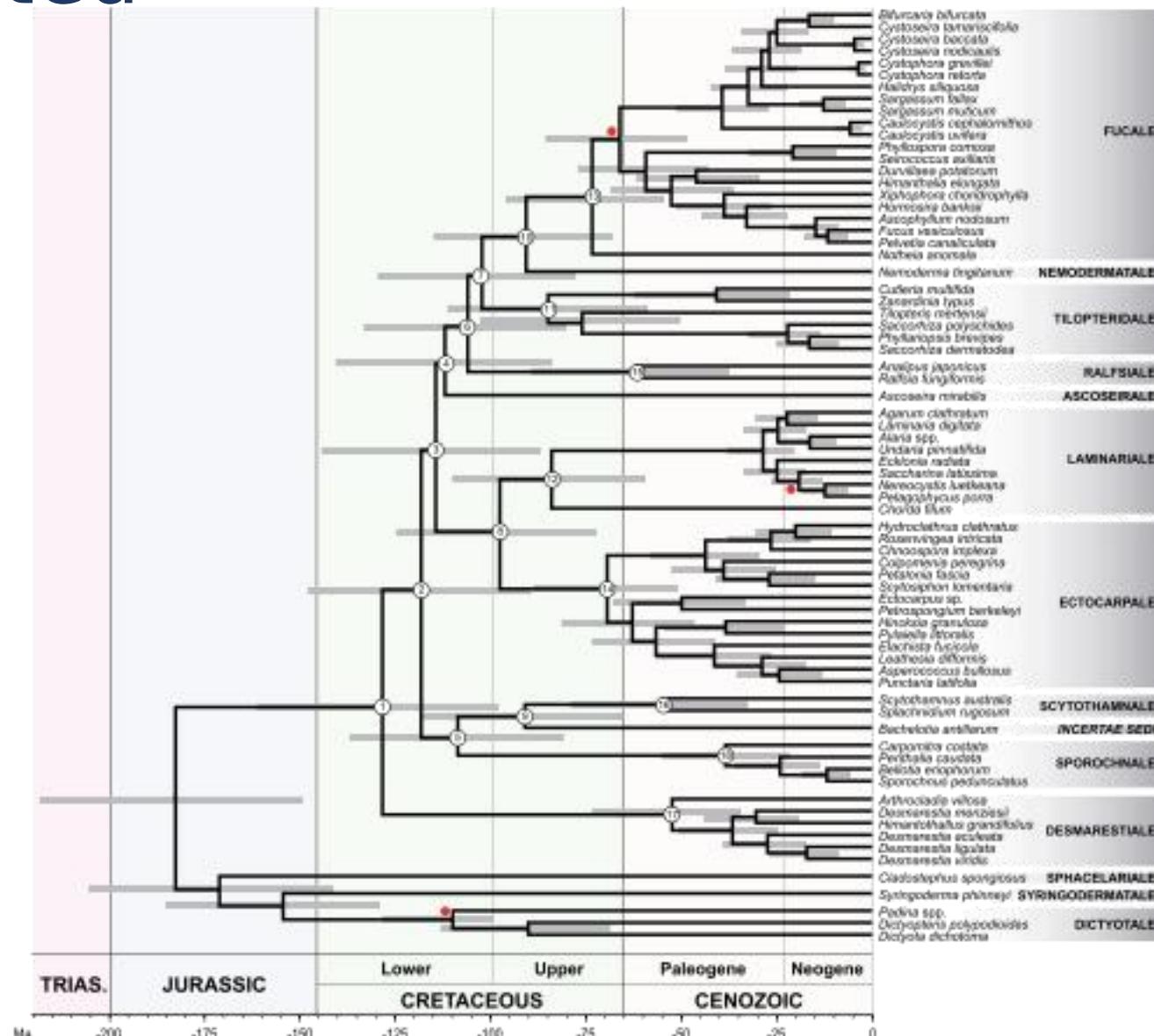
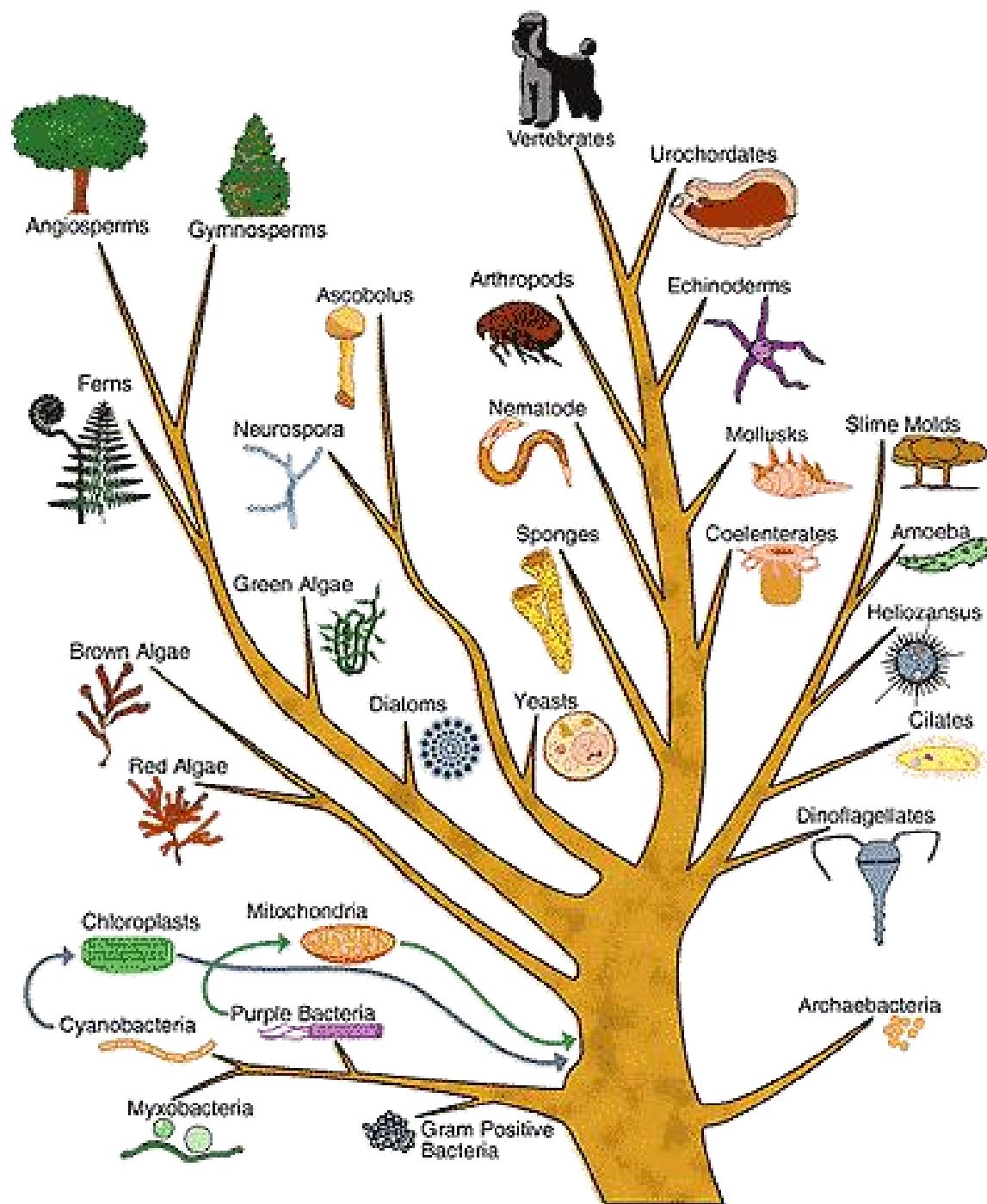


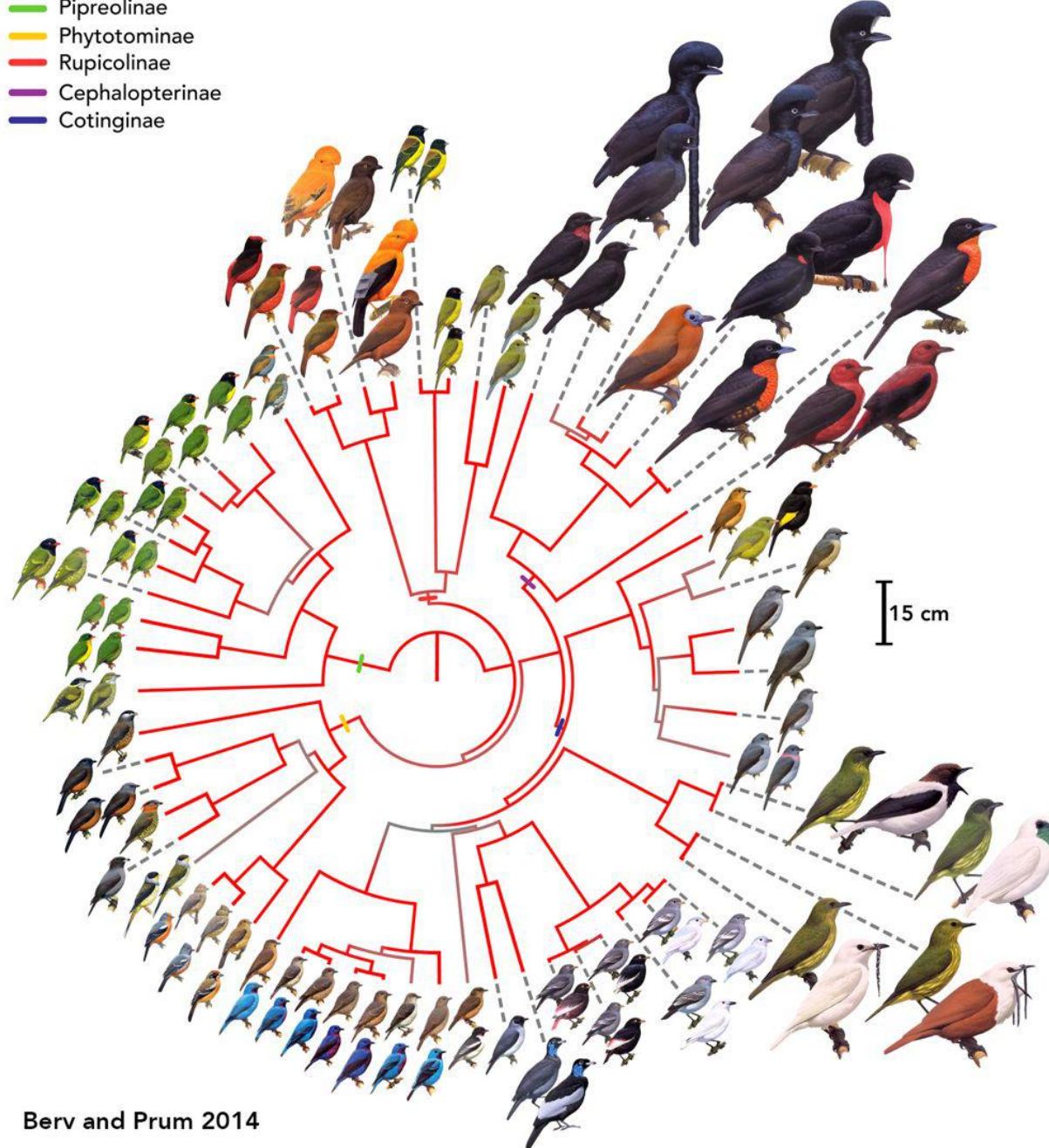
Fig. 2. Chronogram resulting from the Bayesian relaxed molecular clock analysis performed with BEAST (Drummond and Rambaut, 2007). The grey bars display the 95% HPD (highest probability density) interval of node ages. Details for nodes labelled with numbers are provided in Table 3. The red circles mark the three nodes that were time-constrained with fossils as described in text: the Sargassaceae and the (*Nereocystis*–*Pelagophycus*) lineage were constrained under a uniform prior with 13 Ma as lower boundary for their respective stem nodes, and the Dictyotales were constrained under a uniform prior with 13 Ma as lower boundary for their crown node. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Cartoon one



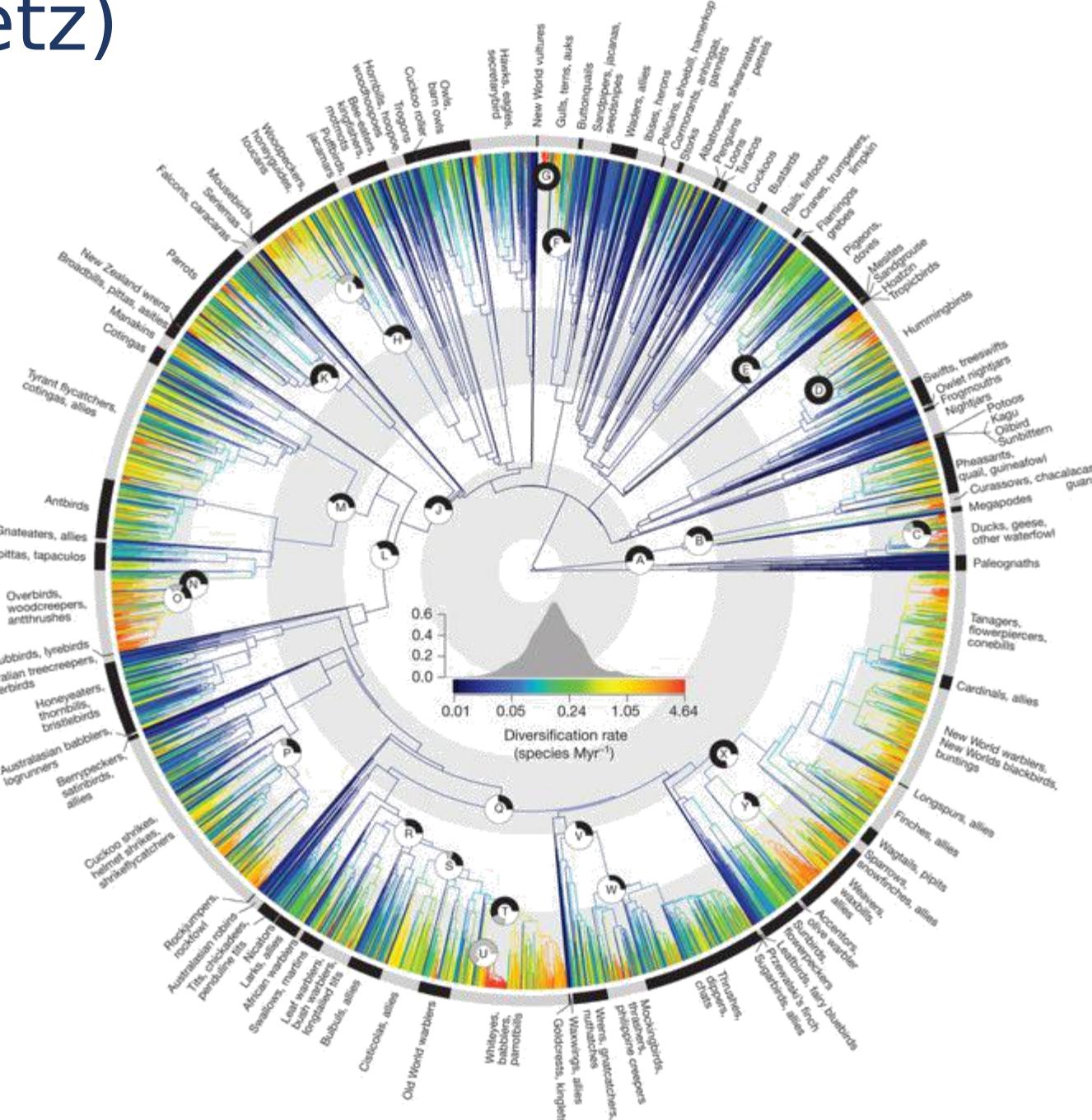
round tree

- Pipreolinae
- Phytotominae
- Rupicolinae
- Cephalopterinae
- Cotinginae

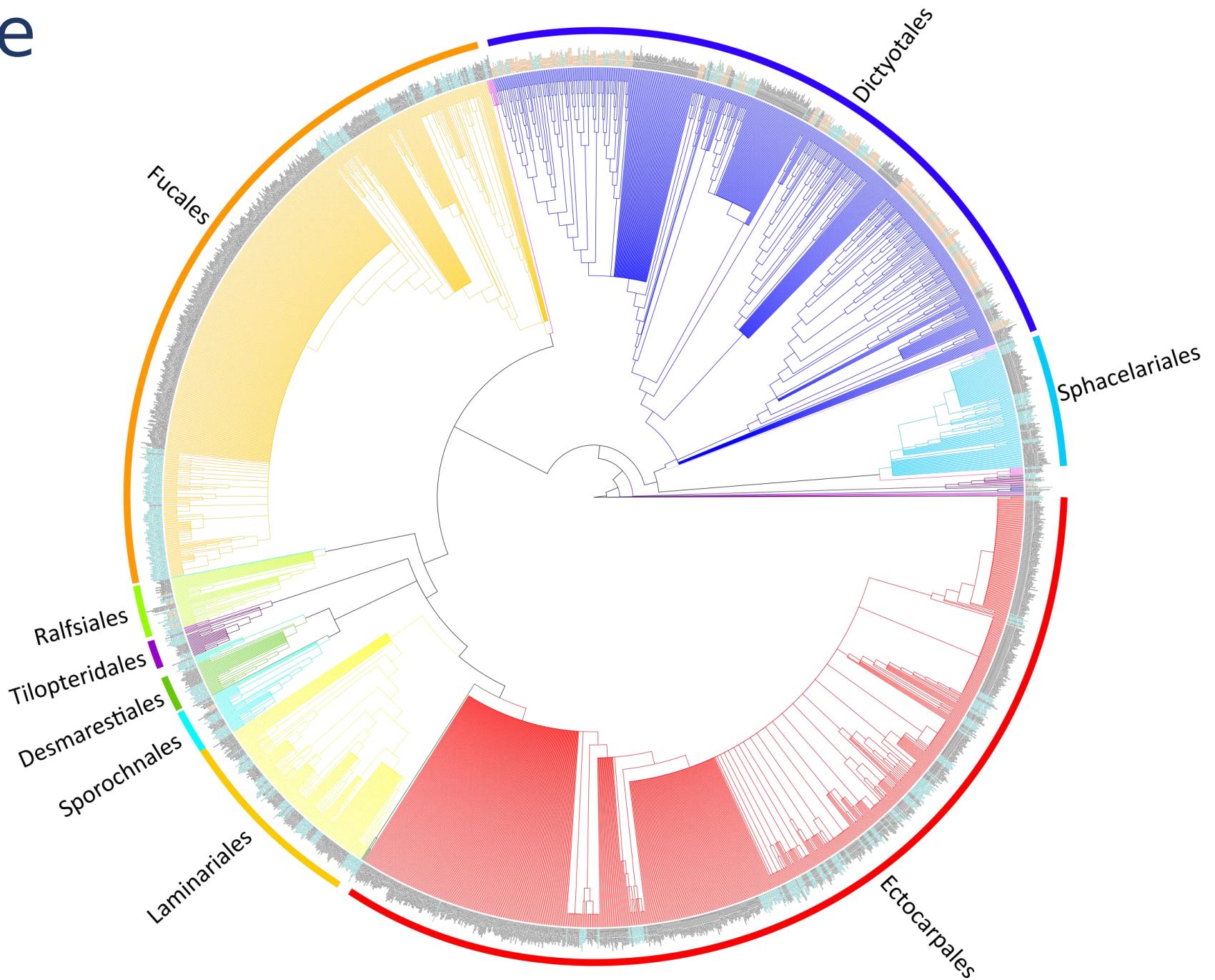


Berv and Prum 2014

Supertree (Jetz)



Megatree



Explore internet

Google

phylogeny



All

Images

Videos

News

Books

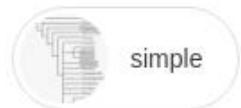
More

Settings

Tools

Collections

SafeSearch ▾



simple



human



evolution



tree



bacteria



animal



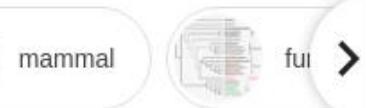
plant



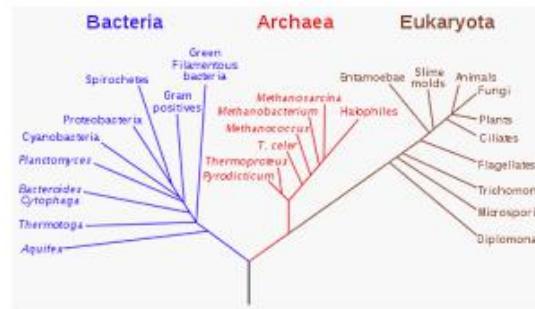
fish



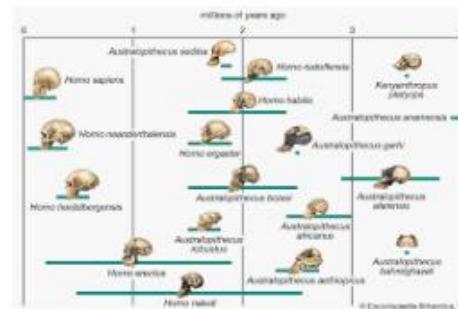
mammal



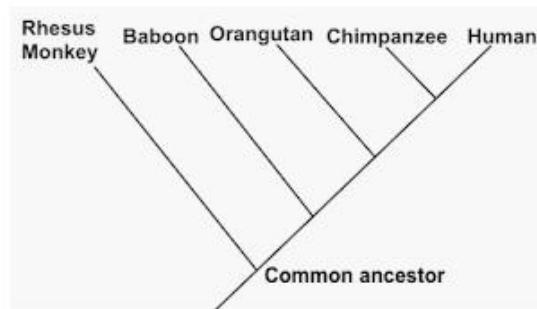
full



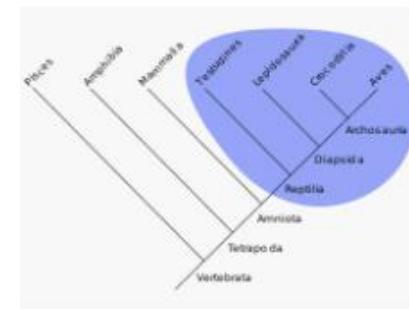
Phylogenetic tree - Wikipedia
en.wikipedia.org



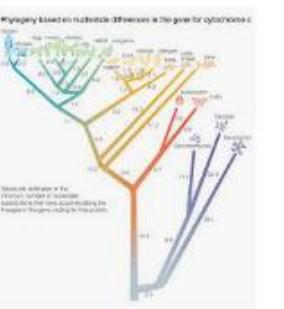
Phylogeny | biology | Britannica.com
britannica.com



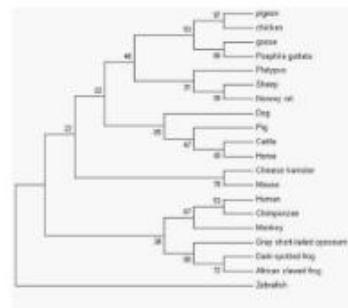
Phylogeny (practice) | Evolution | Khan ...
khanacademy.org



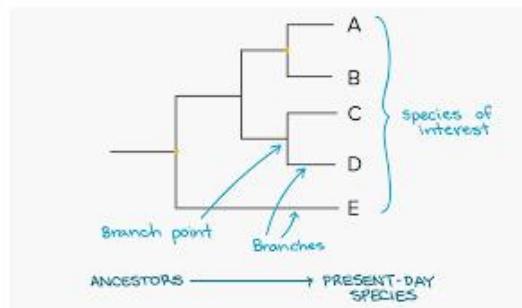
Phylogeny - Simple English Wikipedia...
simple.wikipedia.org



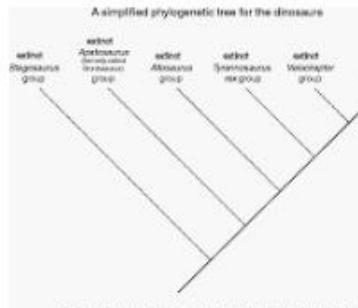
Phylogeny | biology | B...
britannica.com



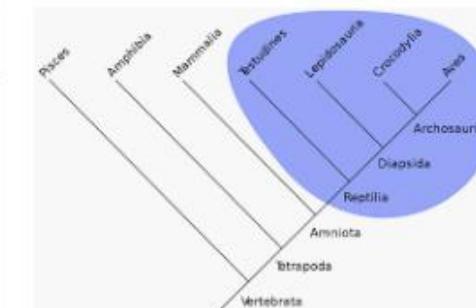
Reed Bio 342
reed.edu



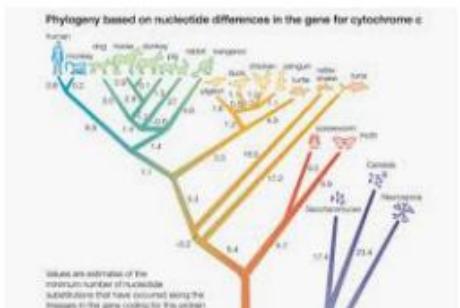
Phylogeny review (article) | Evolution ...
khanacademy.org



How to Read a Phylogenetic Tree ...
link.springer.com



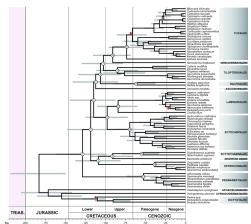
Phylogeny - Simple English Wikipedia ...
simple.wikipedia.org



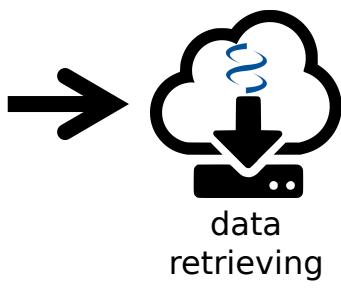
What Is Phylogeny? Definition and Exampl...
thoughtco.com

Program of the course

Day 1



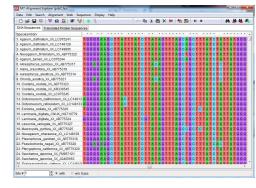
introduction to phylogeny



```
>Sequence 1 assembled  
CCCTAAACCTTAACATCCCAAGATCCCTAACCTCTGA  
CTTAAATCCTACATCCCAAGATCCCTAACCTCTGA  
CTCTGGTGAAATGATTTATATAATGATAAT  
TTCTGTGTGTAGATTTATATAATGATAAT  
GGTTTCCTTCCTTGATTTATGATGGTTTA  
GCTTTTAA/GATTTATGATGGATGTGGCT  
ATCGTGGTGTGGCT  
<Sequence 2>  
CCCTAAACCTTAACATCCCAAGATCCCTAACCTCTGA  
CTTAAACCTACATCCCAAGATCCCTAACCTCTGA  
CTCTGGTGAAATCATGTGTATATAATGATAAT  
TTCTGTGTGTAGATTTATATAATGATAAT  
GGTTTCCTTCCTTGATTTATGATGGTTTA  
GCTTTTAA/GATTTATGATGGATGTGGCT  
ATCGTGGTGTGGCT
```

cleaning

Day 2



aligning

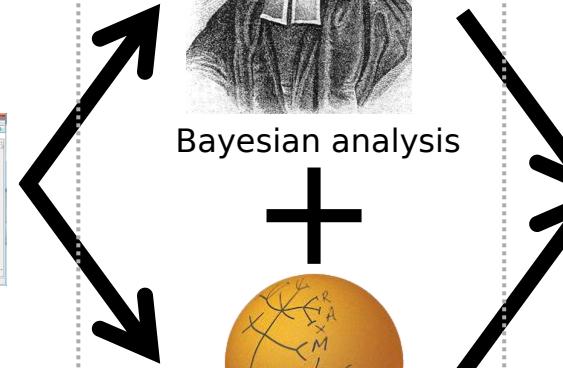
Day 3



Bayesian analysis



maximum likelihood analysis



tree interpretation & graphical design



BLAST, CD-hit,
Q&A ...

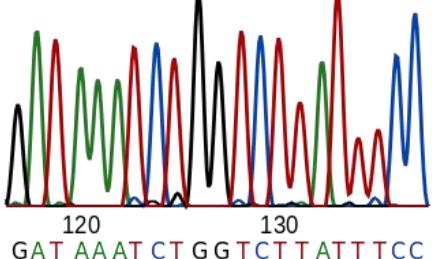


<https://www.ncbi.nlm.nih.gov/nuccore>

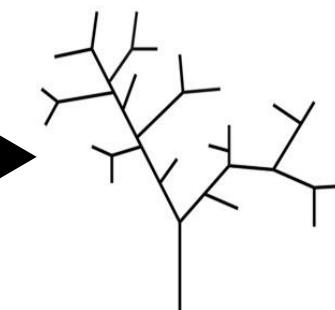
Research workflow for genetic datasets



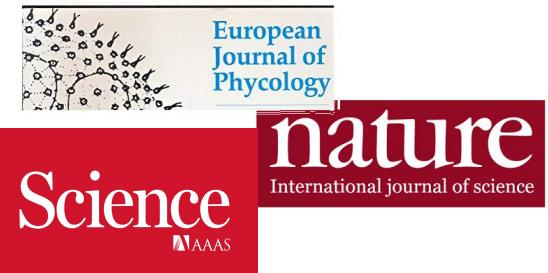
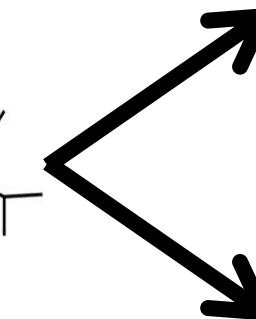
Sampling



Sequencing



Analysing the data



Publishing in a journal



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

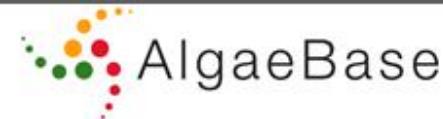
Protein BLASTDBs are accession-based 30 Sep 2019

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

Virus hunting in the cloud codeathon, v2 26 Sep 2019

We are pleased to announce the second





[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 *Ascophylla 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](#)

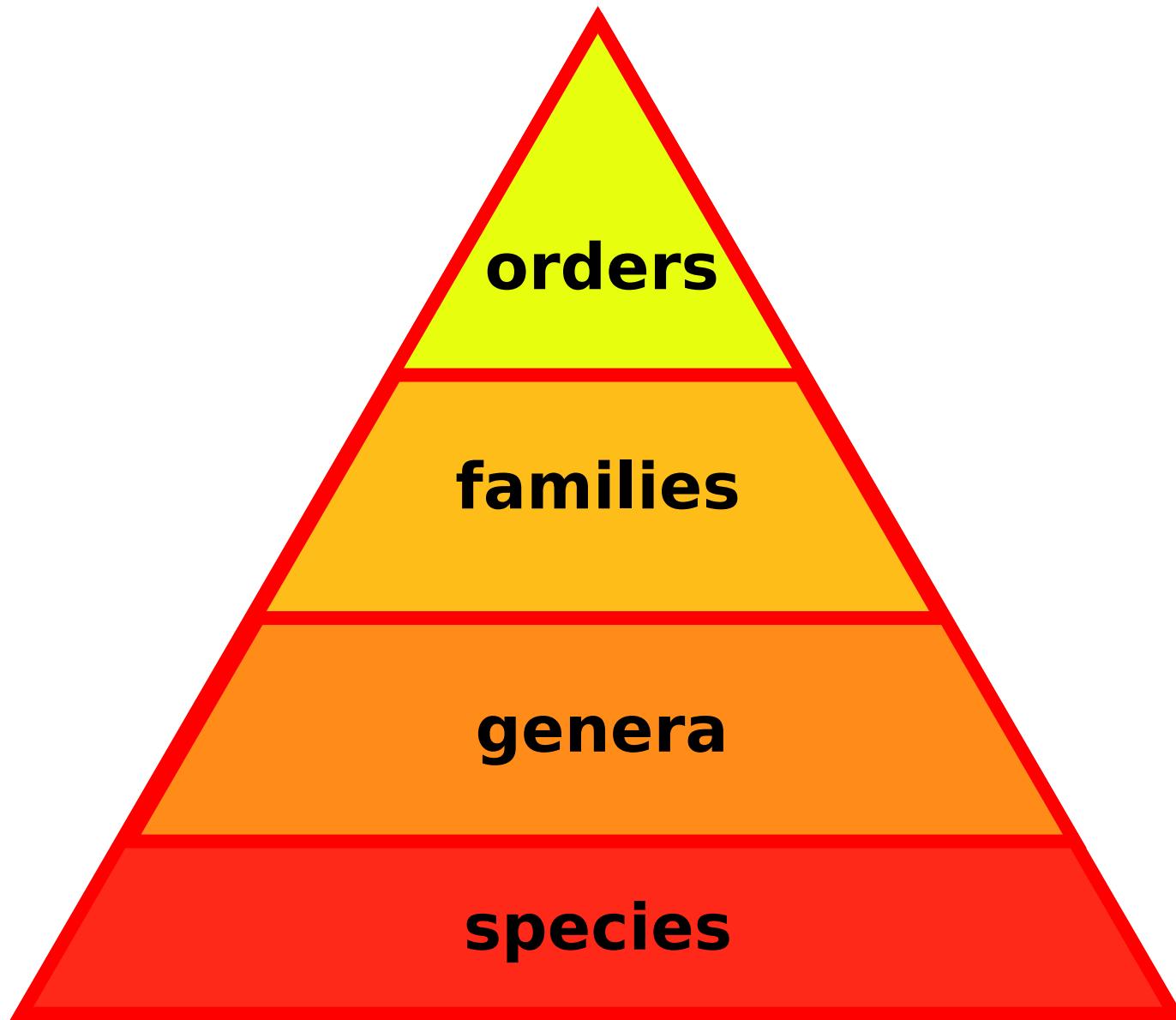
[Index Nominum Algarum](#)

[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*.

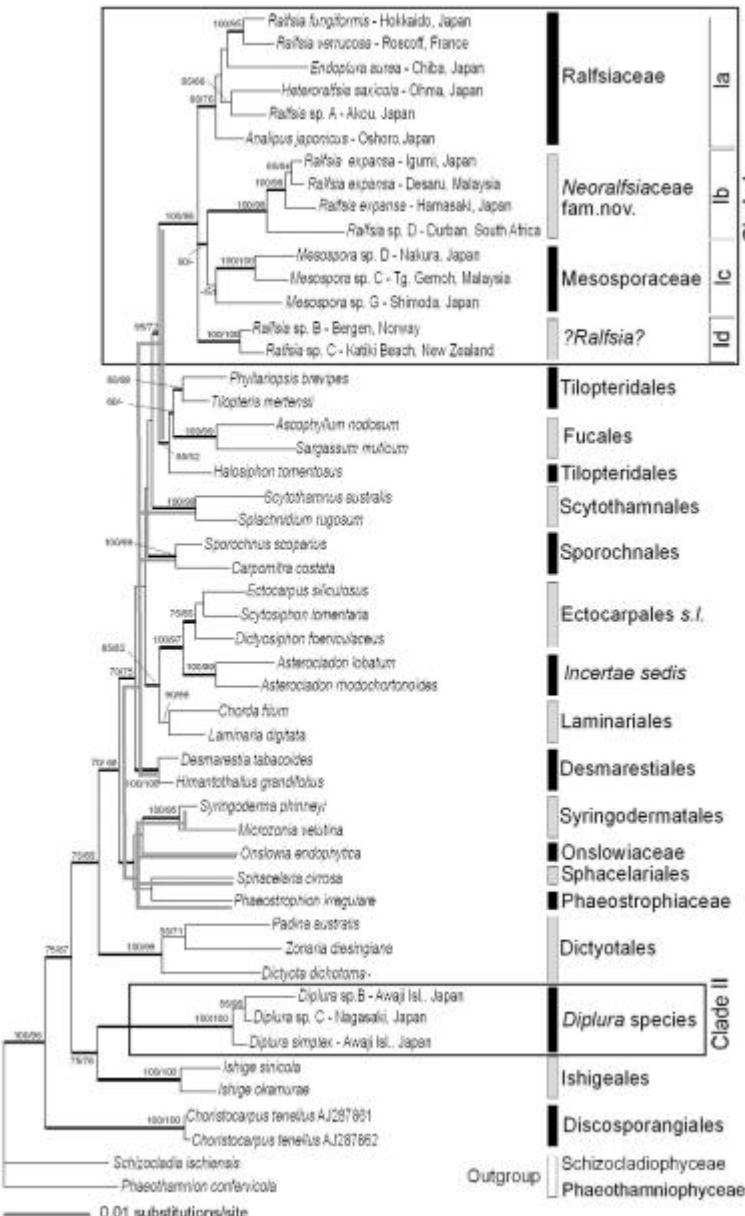
KEY WORDS: Molecular phylogeny, *Neoralfsia*, Neoralfsiaceae, Ralfsiaceae, Ralfsiales, *rbcL*, Taxonomy

Presentation of the dataset from Lim2007

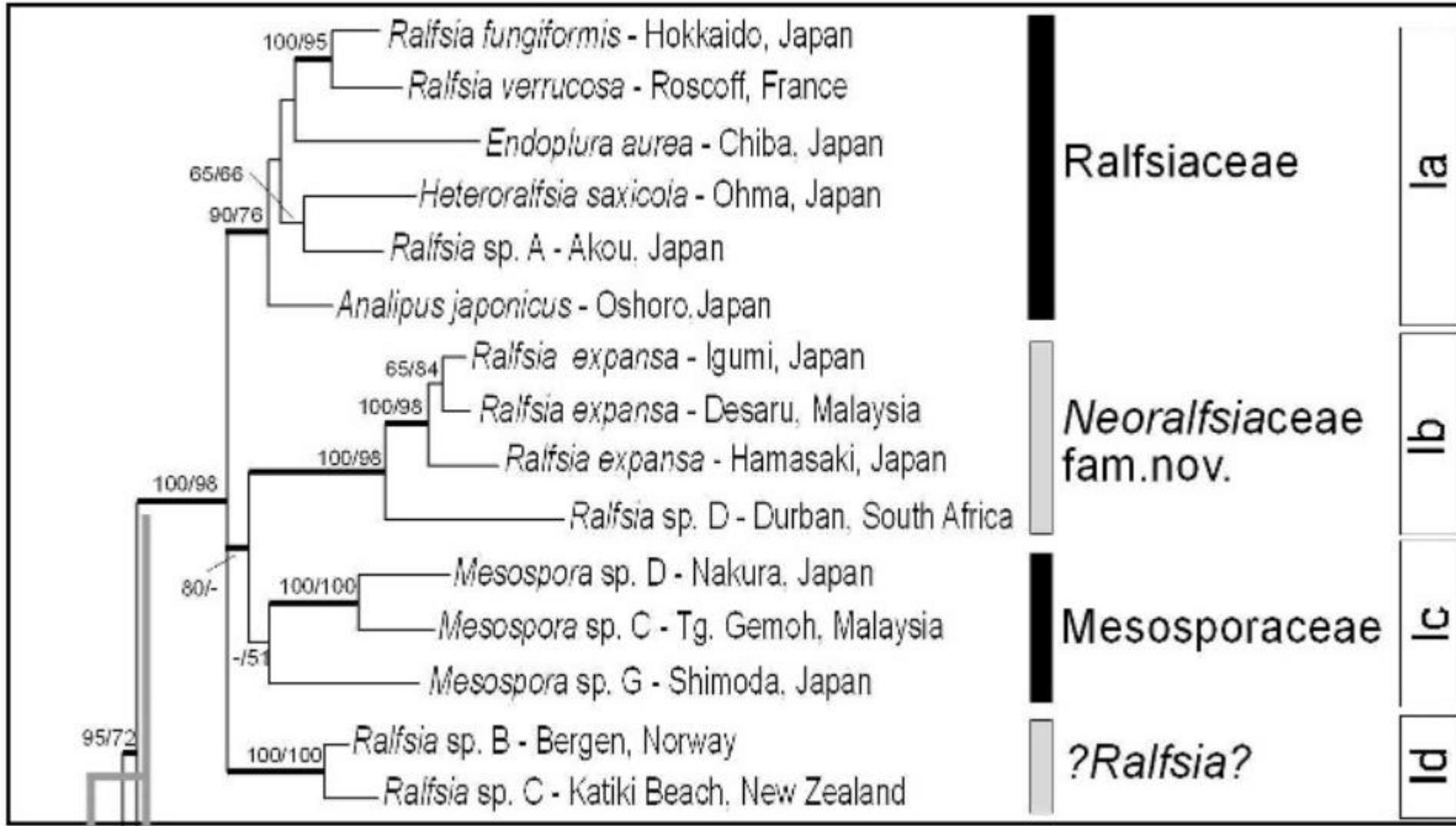
Name of taxa	Origin (published year/collector/specimen no.)	DDBJ accession number for <i>rbcL</i>
<i>Ralfsiaceae</i>		
<i>Analipus japonicus</i> (Harvey) Wynne	Oshoro, Hokkaido Pref., Japan (H. Kawai) (KU-883)	AB264042
<i>Heteroralfsia saxicola</i> (Okamura et Yamada) Kawai	Ohma, Aomori Pref., Japan (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

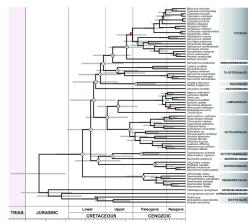


Presentation of the tree we will do

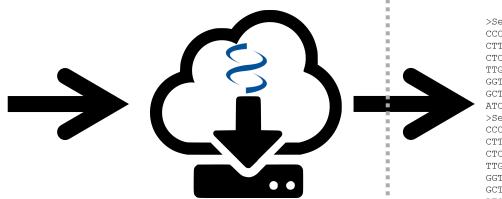


Program of the course

Day 1



introduction to phylogeny



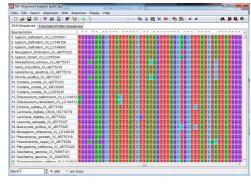
data retrieving



<https://notepad-plus-plus.org/downloads/>

Day 2

>Sequence 1 assembled
CCCTAAACCTTAACCTTAAACCTCTGA
CTTAAATCCATCCATCCATGATCCCTAAATACCT
CTCTGGTGAAATGATATATAATGATAAT
TTCTTGTTGTAGATGATATGATATGATATGATAAT
GGTTTCCTTCCTGATATGATATGATATGATATGATAAT
GCTTTTCA/GATATGATATGATATGATATGATATGATAAT
ATGATATGATATGATATGATATGATATGATATGATAAT
<Sequence 2>
CCCTAAACCTTAACCTTAAACCTCTGA
CTTAAACCTTACATCCATGATCCCTAAATACCT
CTCTGGTGAAATCATGATATGATATGATATGATAAT
TTCTTGTTGTAGATGATATGATATGATATGATAAT
GGTTTCCTTCCTGATATGATATGATATGATATGATAAT
GCTTTCTCGATCTGATATGATATGATATGATAAT
ATGATATGATATGATATGATATGATATGATATGATAAT



aligning

Day 3



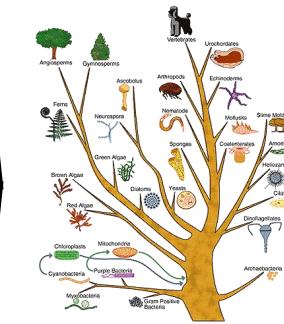
Bayesian analysis



maximum likelihood analysis

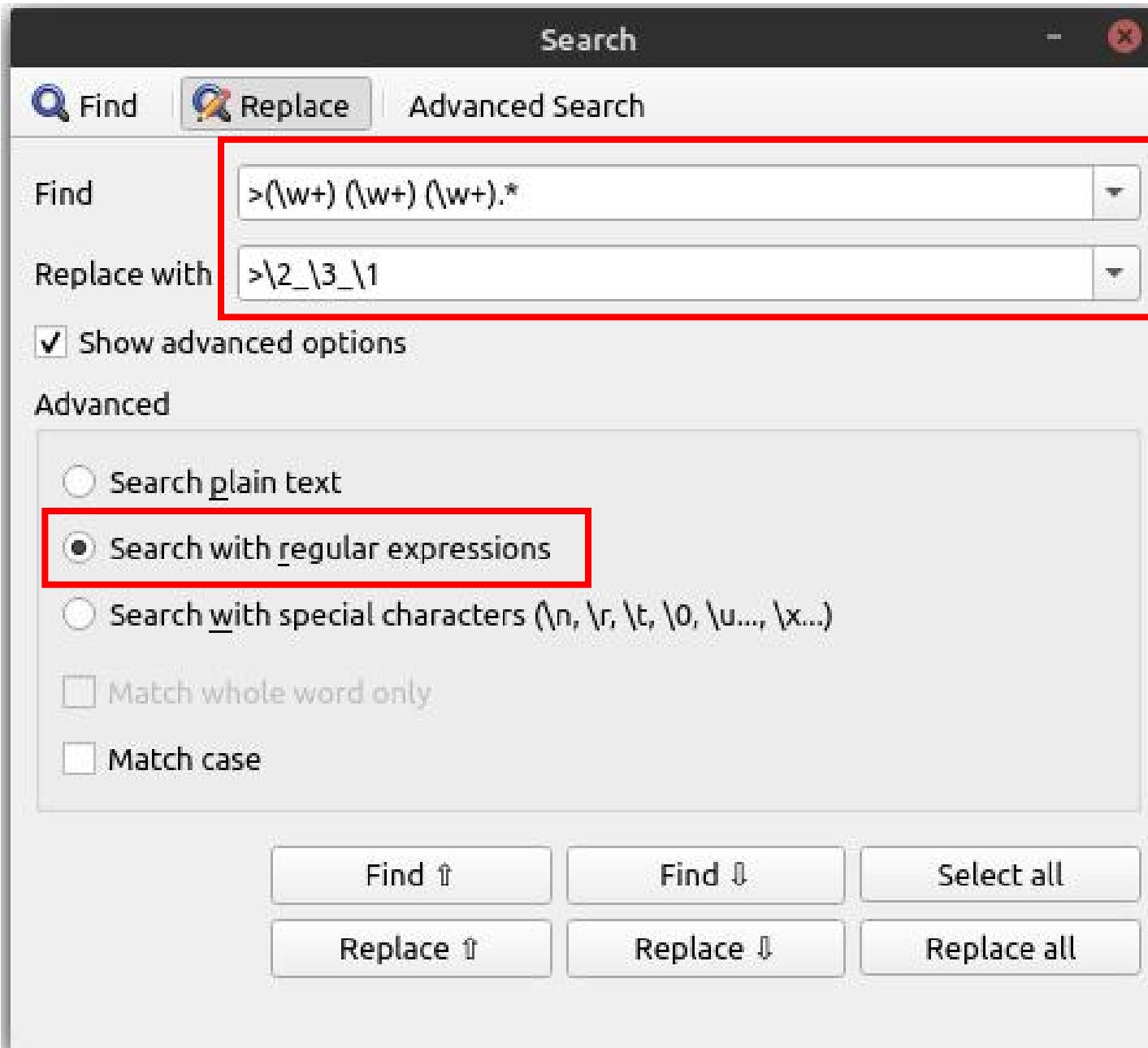


BLAST, CD-hit,
Q&A ...



tree interpretation & graphical design

Regular expression used



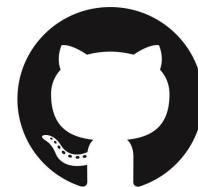
Takes only the first three
'words' (\w+) and
put them in a different order
separated by '_'

Usefull links

Path to the programs

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

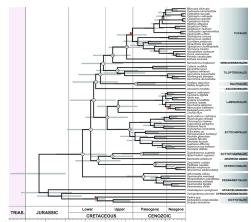
Link to the files



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

Program of the course

Day 1



introduction to phylogeny



data retrieving

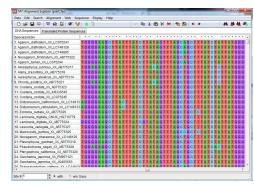
```
>Sequence 1 assembled  
CCCTAAACCTTAACCTTAAACCTCTGA  
CTTAAATCCTACATCCATGATCCCTAAATACCT  
CTCTGGTGAAATGATATATAATGATAAT  
TTCTGGTGATGATGATATATGATGATAAT  
GGTTTCCTTCCTGATGATGATGATGATGATGAT  
GCTTTCCTGATGATGATGATGATGATGATGAT  
ATGATGATGATGATGATGATGATGATGATGAT  
<Sequence 2>  
CCCTAAACCTTAACCTTAAACCTCTGA  
CTTAAATCCTACATCCATGATCCCTAAATACCT  
CTCTGGTGAAATGATATATAATGATAAT  
TTCTGGTGATGATGATATATGATGATAAT  
GGTTTCCTTCCTGATGATGATGATGATGATGAT  
GCTTTCCTGATGATGATGATGATGATGATGAT  
ATGATGATGATGATGATGATGATGATGATGAT
```

cleaning



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

Day 2

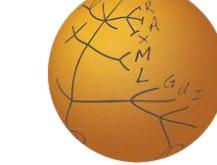


aligning

Day 3



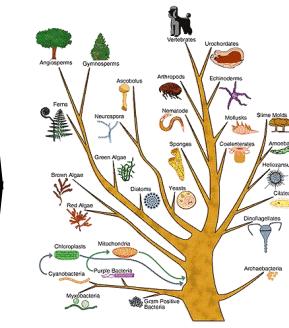
Bayesian analysis



maximum likelihood analysis



BLAST, CD-hit,
Q&A ...



tree interpretation & graphical design

Presentation of Blast

Basic Local Alignment Search Tool

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

The screenshot shows the NCBI BLAST homepage with the following elements:

- Header:** NIH U.S. National Library of Medicine, NCBI National Center for Biotechnology Information, Sign in to NCBI.
- Title:** BLAST®
- Breadcrumbs:** Home > Basic Local Alignment Search Tool
- Section: 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](#)
- News:** 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 Tools:**
 - Nucleotide BLAST**: nucleotide ▶ nucleotide
 - blastx**: translated nucleotide ▶ protein
 - tblastn**: protein ▶ translated nucleotide
 - Protein BLAST**: protein ▶ protein
- BLAST Genomes**

Presentation of TreeBase

<https://treebase.org>



A Database of Phylogenetic Knowledge

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 [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:

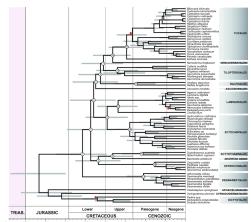
The current release includes a host of new features and improvements over the previous TreeBASE prototype. New features include:

- 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 on tree topology



Program of the course

Day 1



introduction to phylogeny



data retrieving



<https://www.megasoftware.net>



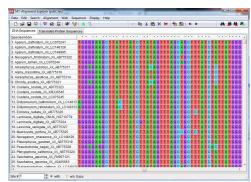
<http://doua.prabi.fr/software/seaview>

MAFFT

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

Day 2

>Sequence_1 assembled
CCCTAAACCTTAAACCTTAAACCTCTGA
CTTAAATCCTACATCCATGATCCCTAAATACCT
CTCTGGTGAAATGATATATAATGATAAT
TTCTGGTGATGATGATGATGATGATGATGATGAT
GGTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT
GCTTTTGTGATGATGATGATGATGATGATGATGAT
ATGATGATGATGATGATGATGATGATGATGATGAT
>Sequence_2
CCCTAAACCTTAAACCTTAAACCTCTGA
CTTAAACCTTACATCCATGATCCCTAAATACCT
CTCTGGTGAAATCATGATGATATAATGATAAT
TTCTGGTGATGATGATGATGATGATGATGATGAT
GGTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT
GCTTTGCTGATGATGATGATGATGATGATGATGAT
ATGATGATGATGATGATGATGATGATGATGATGAT

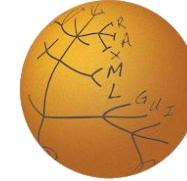


aligning

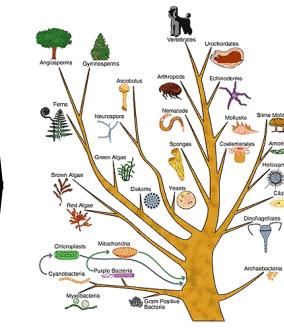
Day 3



Bayesian analysis



maximum likelihood analysis



tree interpretation & graphical design



BLAST, CD-hit,
Q&A ...

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.

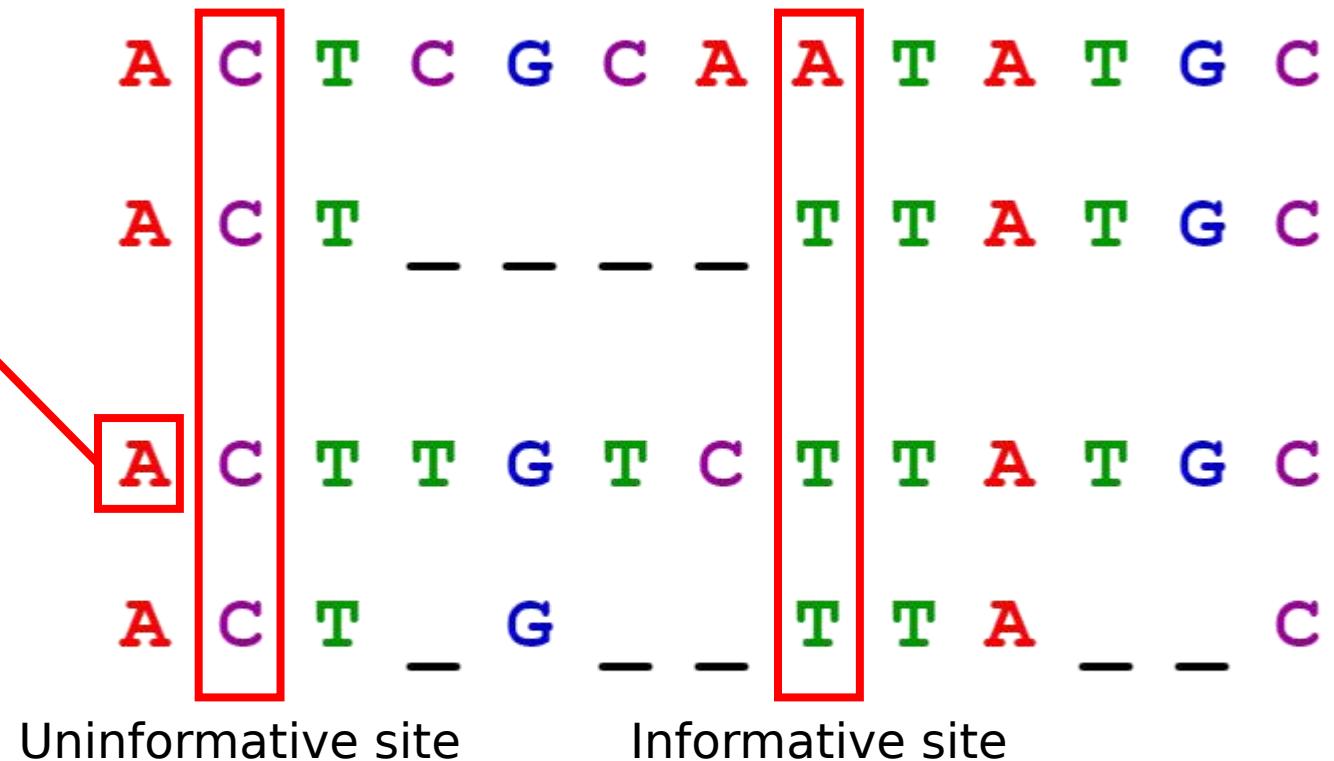
41

After alignment

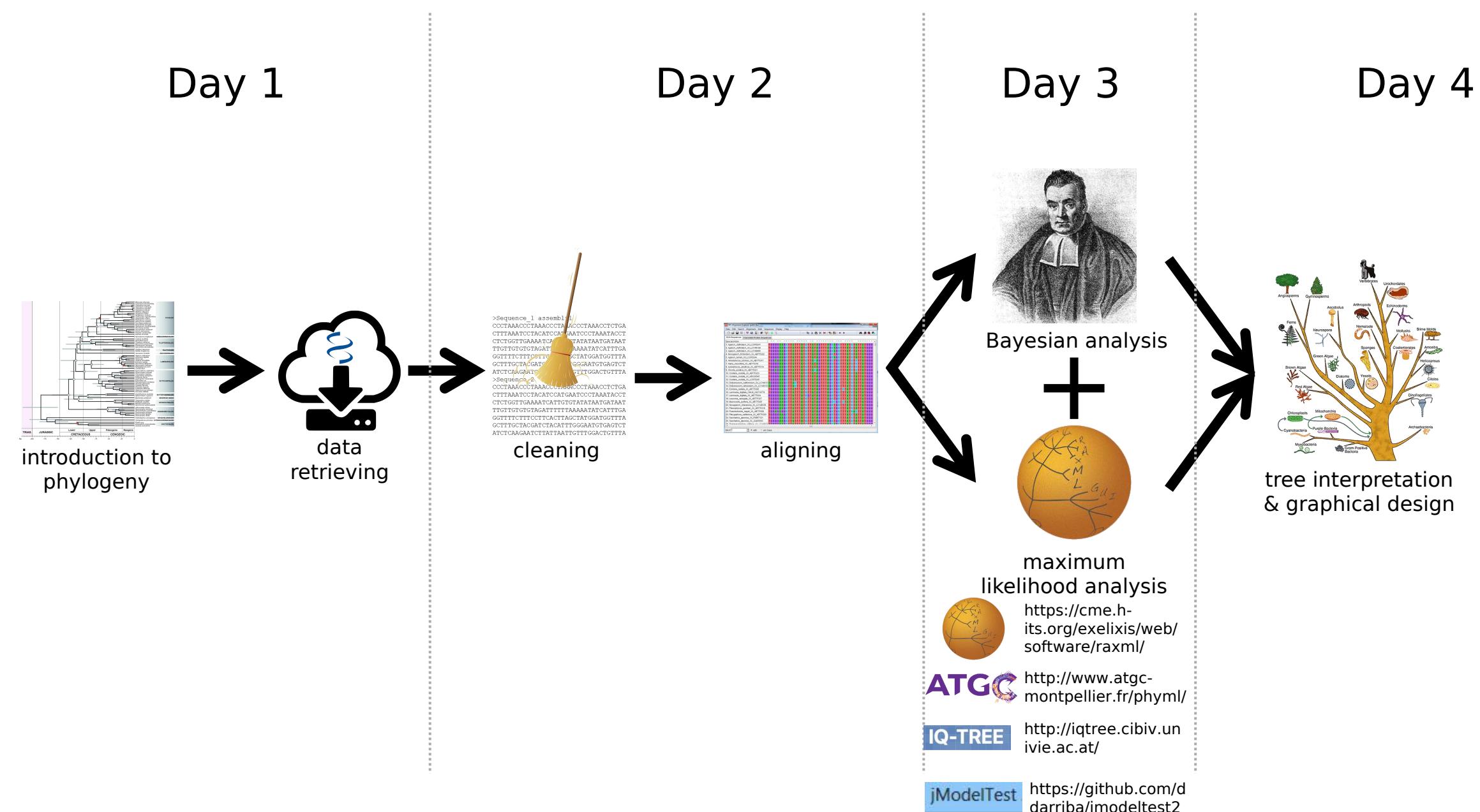


Sequence analysis

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



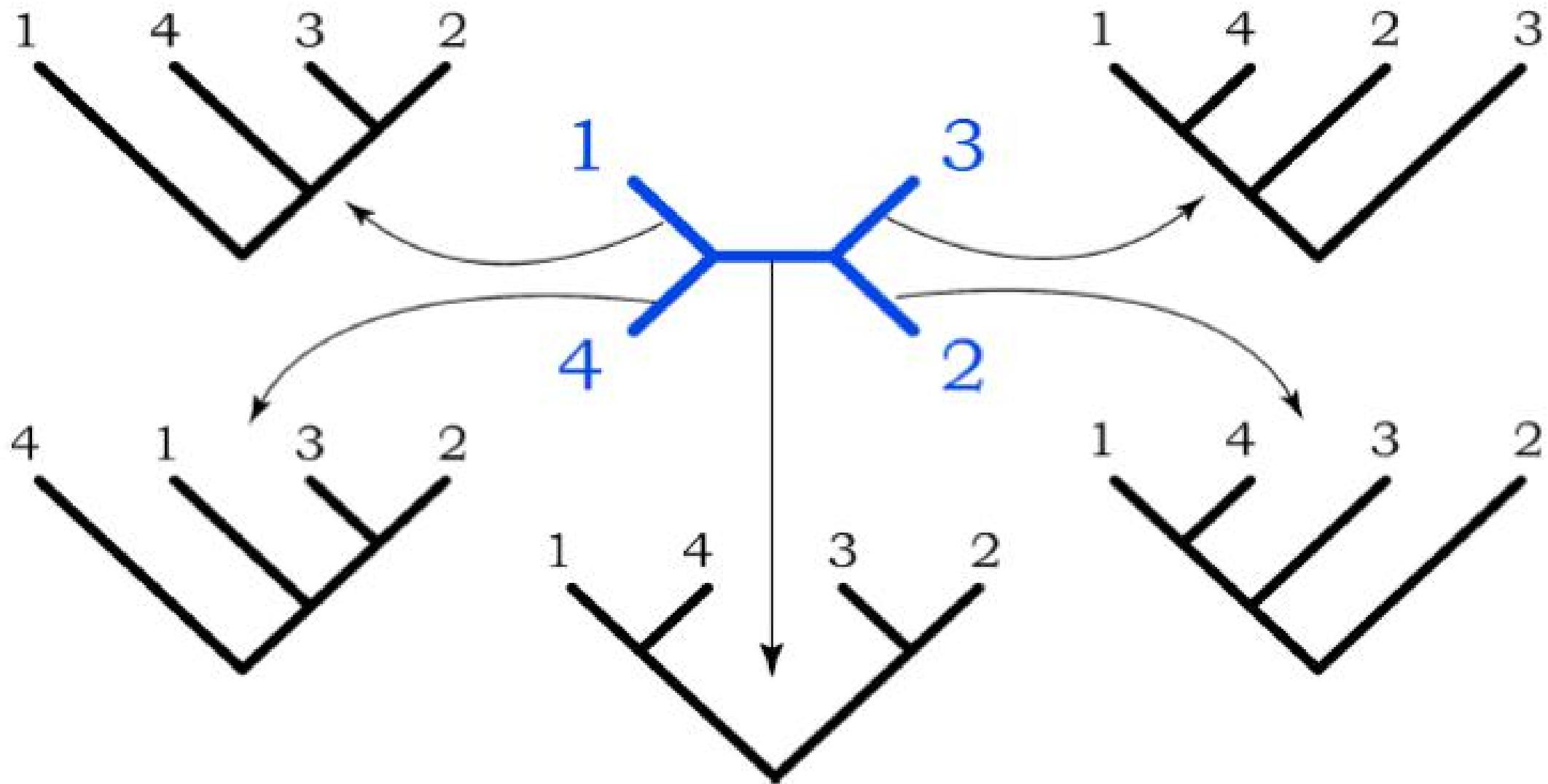
Program of the course



Q&A

BLAST, CD-hit,
Q&A ...

Number of trees possible



Number of trees possible

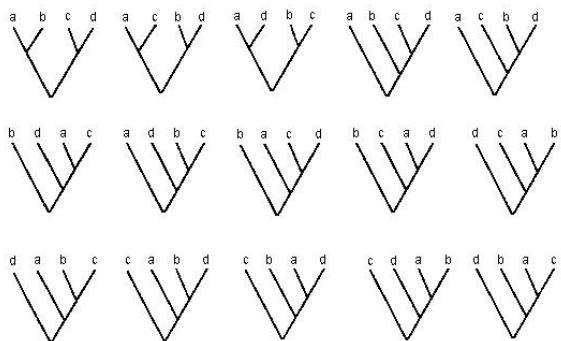
Industries	Unrooted trees	Rooted trees
n	$\prod_{i=3}^n (2i - 5)$	$\prod_{i=2}^n (2i - 3)$
3	1	3
4	3	15
5	15	105
10	2×10^6	3×10^7
15	8×10^{12}	2×10^{14}
20	2×10^{20}	8×10^{21}
25	3×10^{28}	1×10^{30}
50	3×10^{74}	3×10^{76}
100	2×10^{182}	3×10^{184}
500	1×10^{1277}	1×10^{1280}

Maximum likelihood analysis

Phylip file (.phy)

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

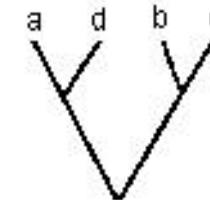
Workflow



Creation of trees



Likelihood calculation of each tree

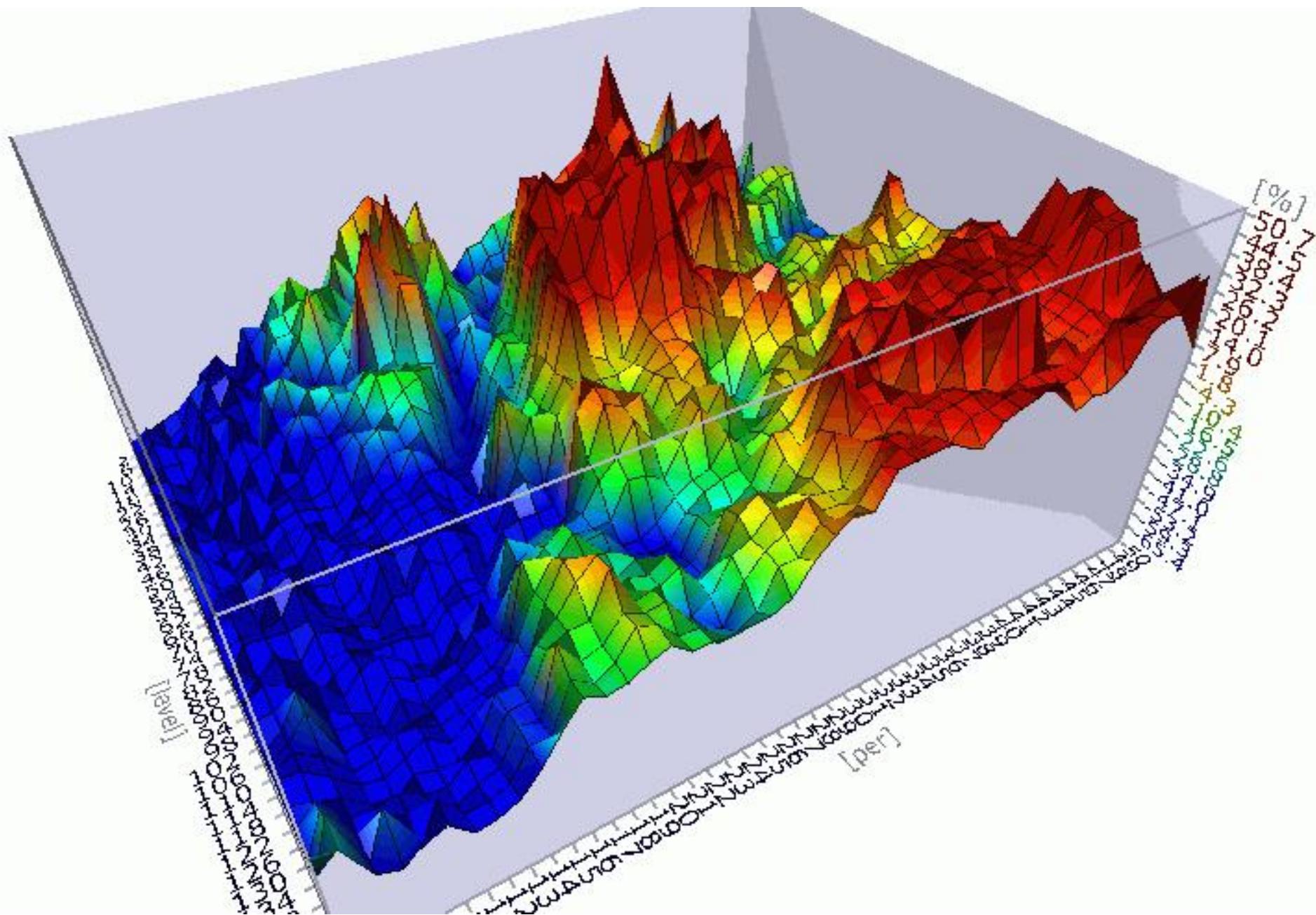


Tree with highest likelihood

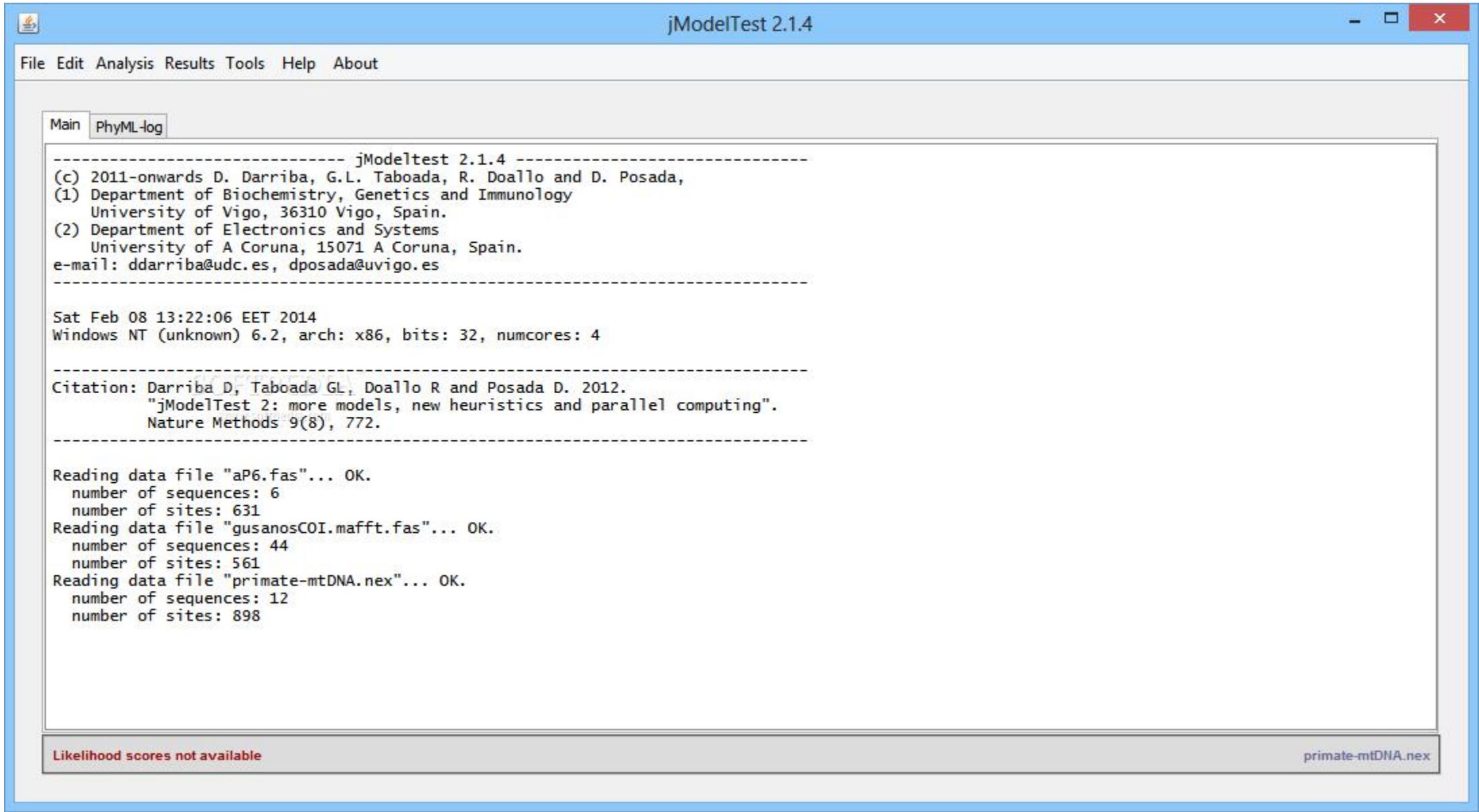
Drawback

Fixed evolutionary model

Tree space exploration



Model selection (jmodeltest)



jModelTest 2.1.4

File Edit Analysis Results Tools Help About

Main PhyML-log

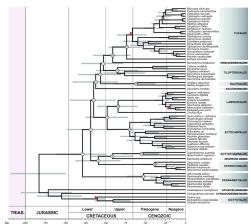
```
----- jModeltest 2.1.4 -----
(c) 2011-onwards D. Darriba, G.L. Taboada, R. Doallo and D. Posada,
(1) Department of Biochemistry, Genetics and Immunology
    University of Vigo, 36310 Vigo, Spain.
(2) Department of Electronics and Systems
    University of A Coruna, 15071 A Coruna, Spain.
e-mail: ddarriba@udc.es, dposada@uvigo.es
-----
Sat Feb 08 13:22:06 EET 2014
Windows NT (unknown) 6.2, arch: x86, bits: 32, numcores: 4

-----
Citation: Darriba D, Taboada GL, Doallo R and Posada D. 2012.
"jModelTest 2: more models, new heuristics and parallel computing".
Nature Methods 9(8), 772.
-----
Reading data file "aP6.fas"... OK.
    number of sequences: 6
    number of sites: 631
Reading data file "gusanosCOI.mafft.fas"... OK.
    number of sequences: 44
    number of sites: 561
Reading data file "primate-mtDNA.nex"... OK.
    number of sequences: 12
    number of sites: 898

Likelihood scores not available
primate-mtDNA.nex
```

Program of the course

Day 1



introduction to phylogeny

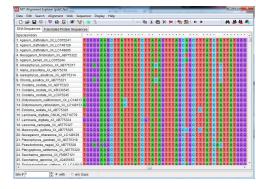


data retrieving

Day 2

```
>Sequence 1 assembled  
CCCTAAACCTTAACCCCTTAAACCTTACCTCTGA  
CTTTAAATCCTACATCCAAATATCCCTAAATACCT  
CTCTGGTGAAATGATATATAATGATAATGATAAT  
TTCTTGTGTGATAGATATGATAATGATAATGATAAT  
GGTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT  
GCTTTTCTTCAATGATATGATAATGATAATGATAAT  
ATGATAATGATAATGATAATGATAATGATAATGATAAT  
<Sequence 2>  
CCCTAAACCTTAACCCCTTAAACCTTACCTCTGA  
CTTTAAACCTTACATCCAAATATCCCTAAATACCT  
CTCTGGTGAAATCATGATGATAATGATAATGATAAT  
TTCTTGTGTGATATGATAATGATAATGATAATGATAAT  
GGTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT  
GCTTTCTTCAATGATATGATAATGATAATGATAAT  
ATGATAATGATAATGATAATGATAATGATAATGATAAT
```

cleaning

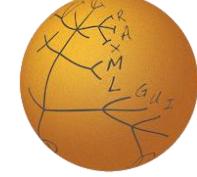


aligning

Day 3



Bayesian analysis



maximum likelihood analysis

 <http://nbisweden.github.io/MrBayes/download.html>

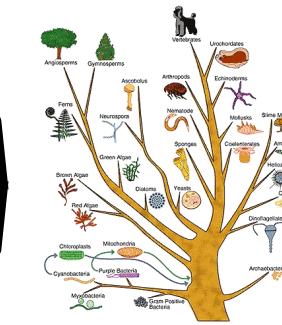


<https://www.phylo.org/portal2/home.action>



<https://github.com/beast-dev/tracer/releases>

Day 4



tree interpretation & graphical design



BLAST, CD-hit,
Q&A ...

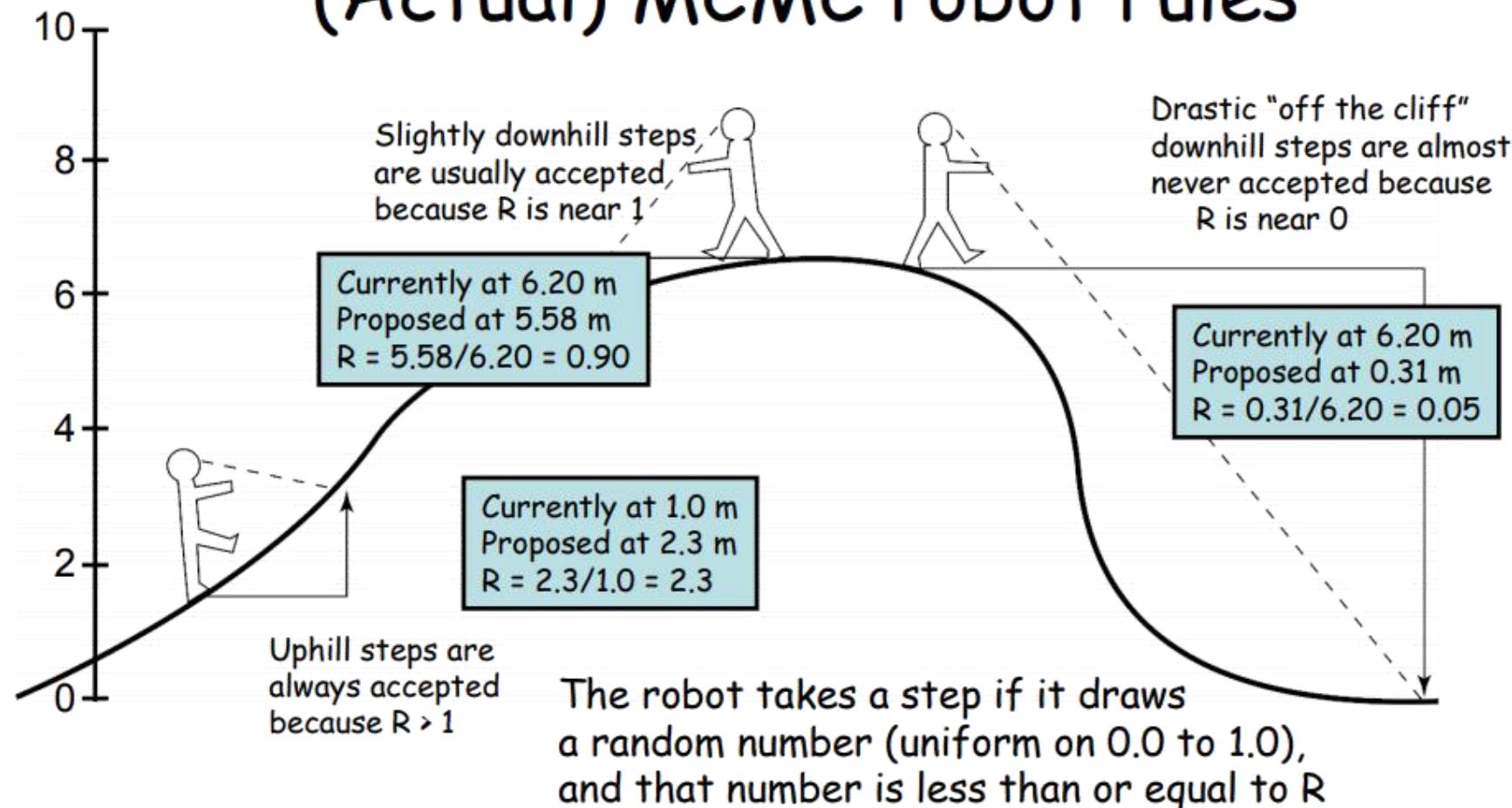
Bayesian analysis

Nexus file (.nex)

Bayesian analysis is based on comparison of prior and posterior probabilities

The evolutionary model is **not** fixed

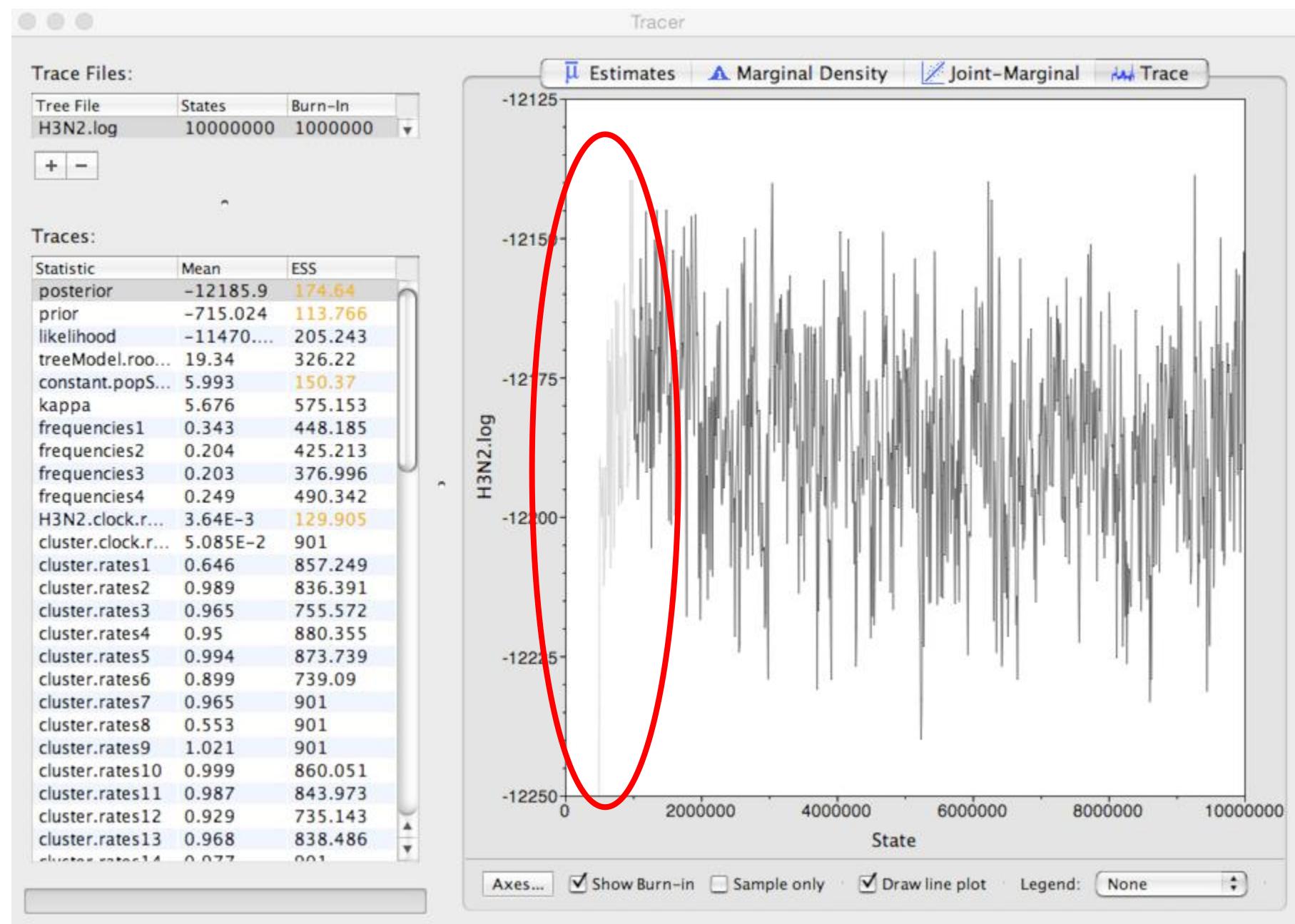
(Actual) MCMC robot rules



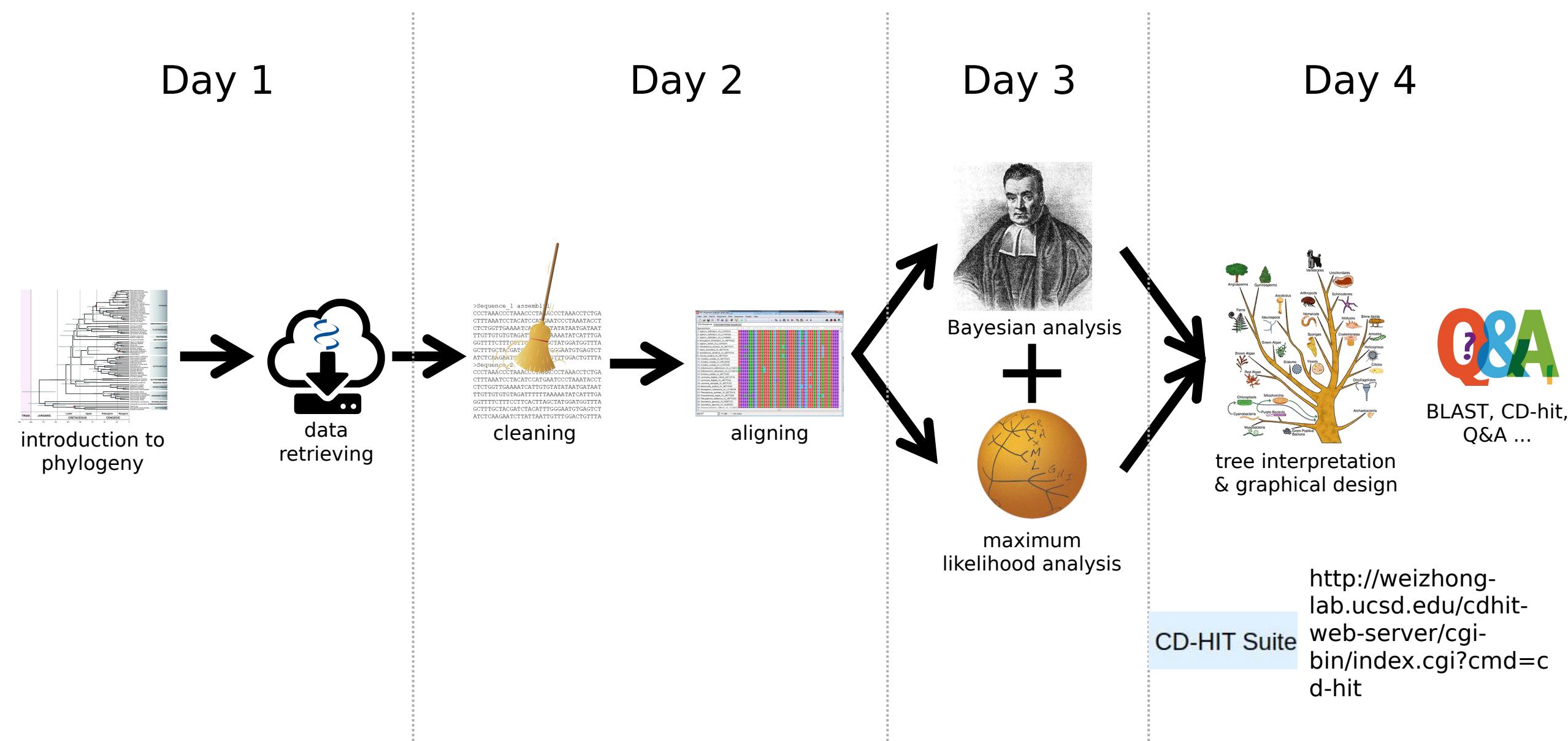
Tracer



goal: set the burn-in value



Program of the course



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: No file chosen
 Incorporate annotation info at header line

Sequence Identity Parameters

Sequence identity cut-off 0.9

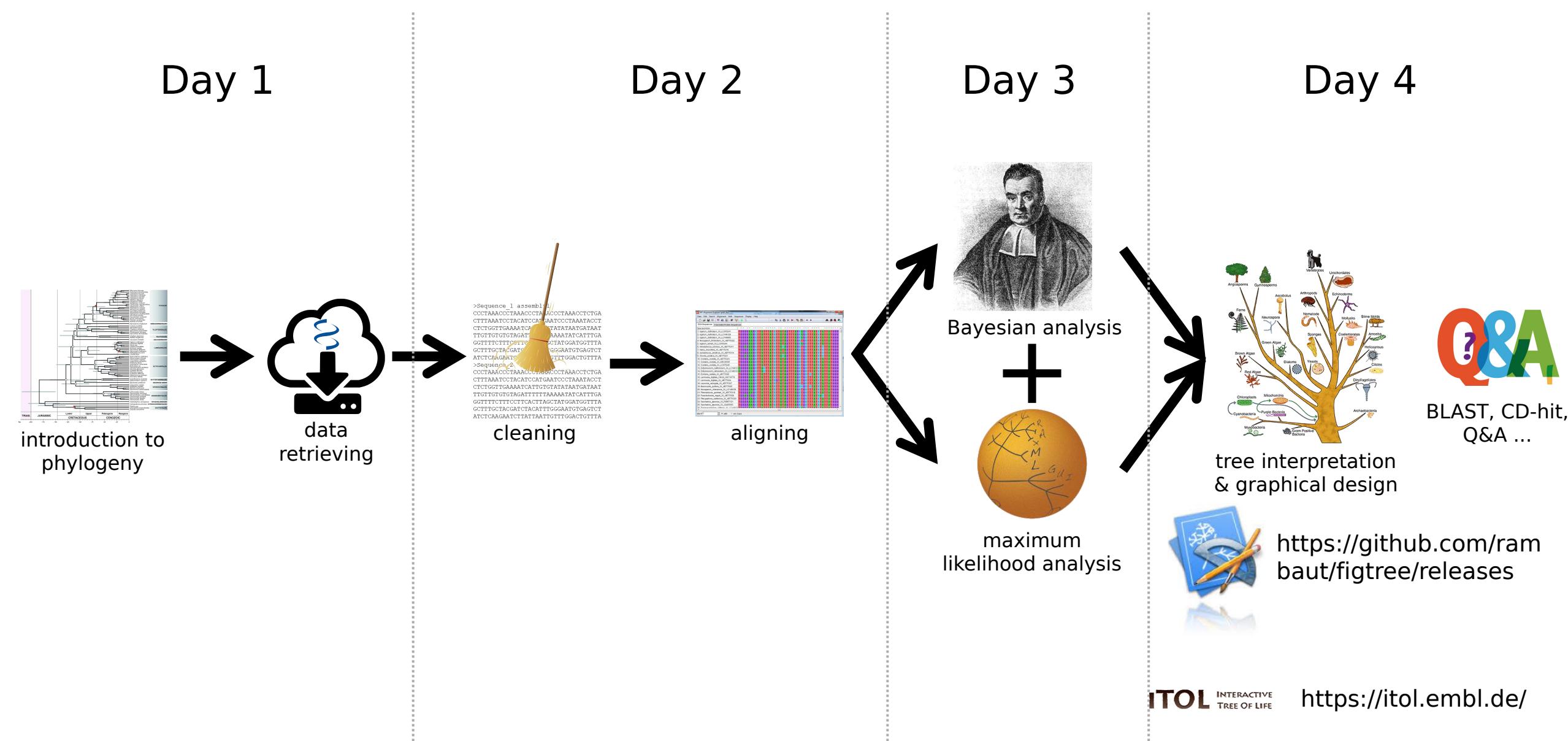
Algorithm Parameters

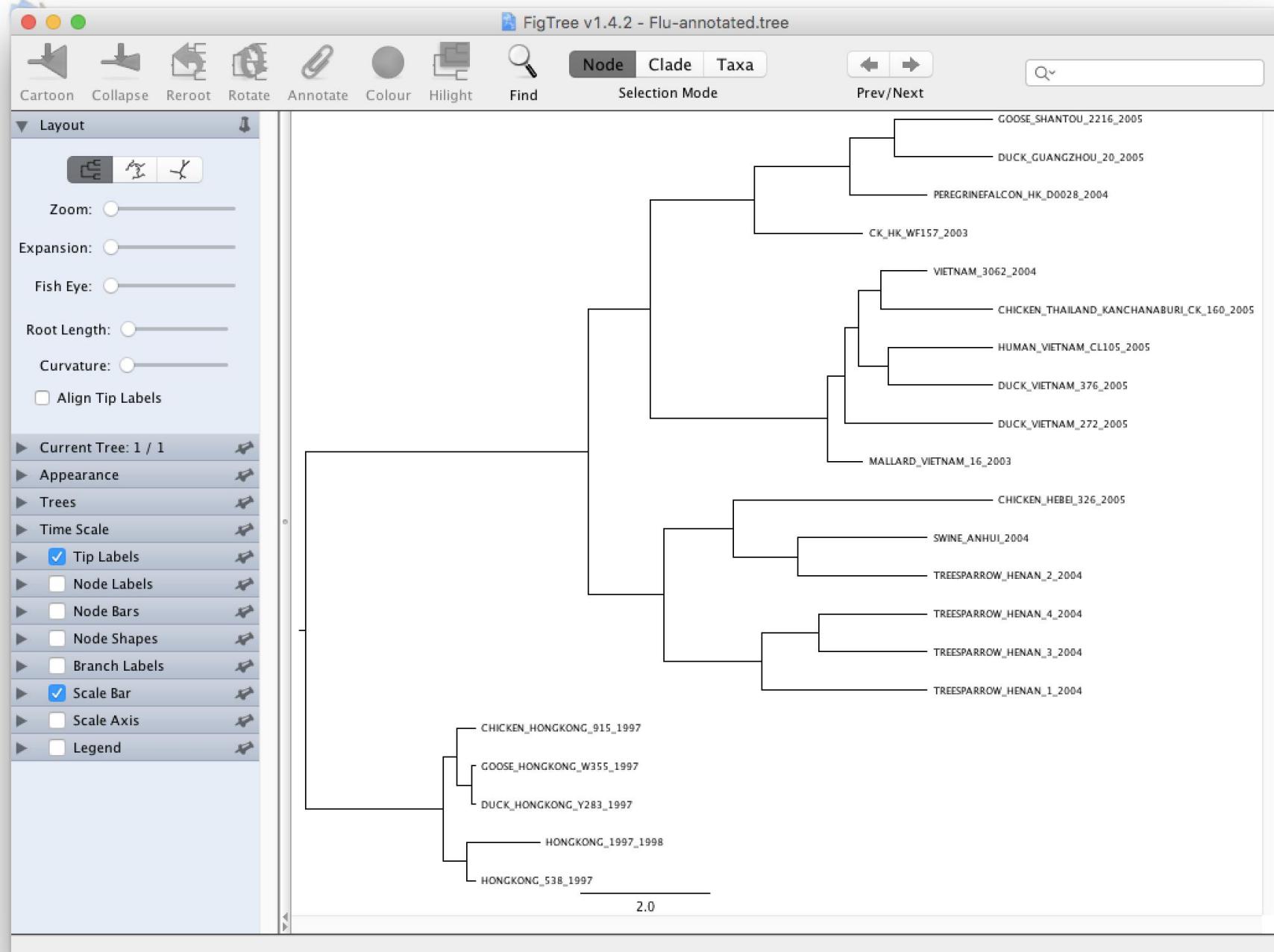
-G: use global sequence identity No Yes
-g: sequence is clustered to the best cluster that meet the threshold No Yes
-b: bandwidth of alignment 20
-l: length of sequence to skip 10

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

Program of the course





Figtree



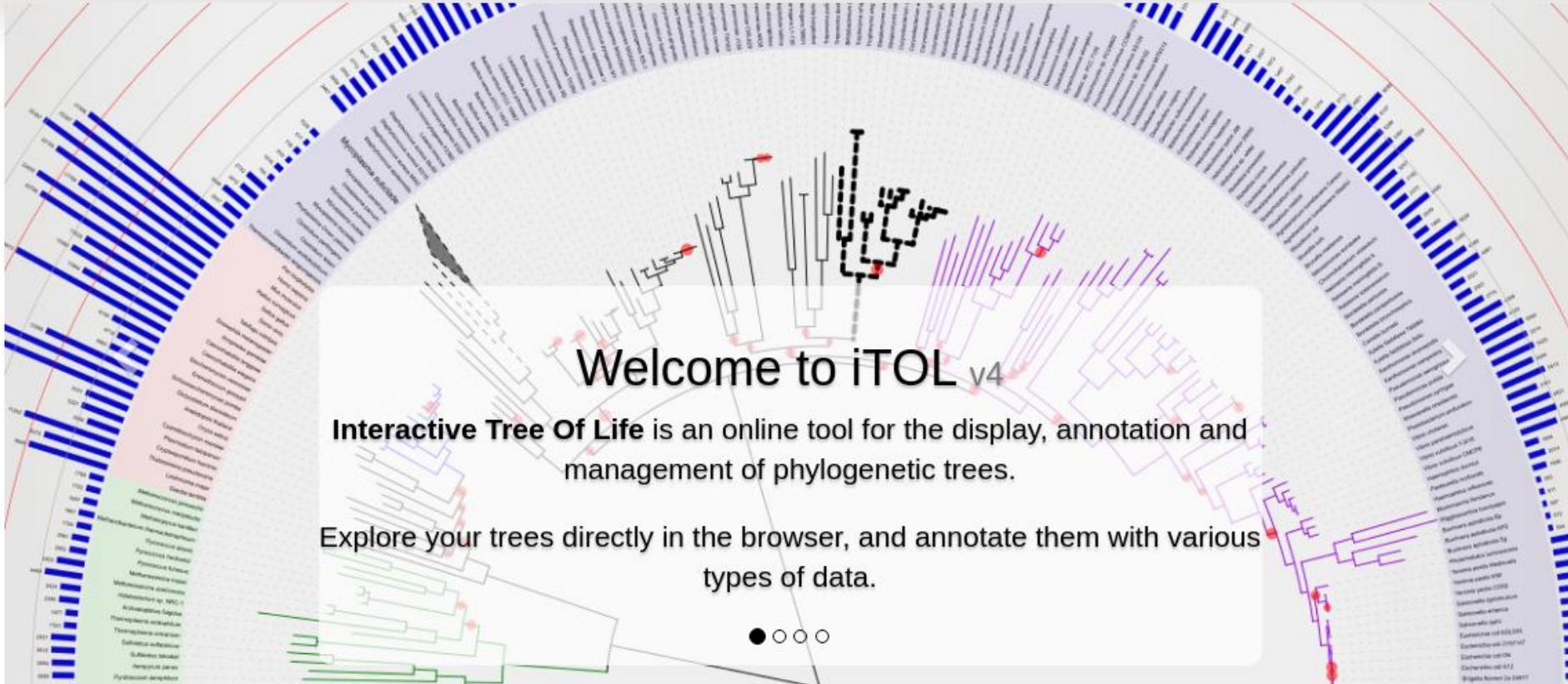
newick file (.nwk or .tre)

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

ITOL INTERACTIVE TREE OF LIFE

Tree of Life Upload Sharing data Help ▾

Login Register



Welcome to iTOL v4

Interactive Tree Of Life is an online tool for the display, annotation and management of phylogenetic trees.

Explore your trees directly in the browser, and annotate them with various types of data.

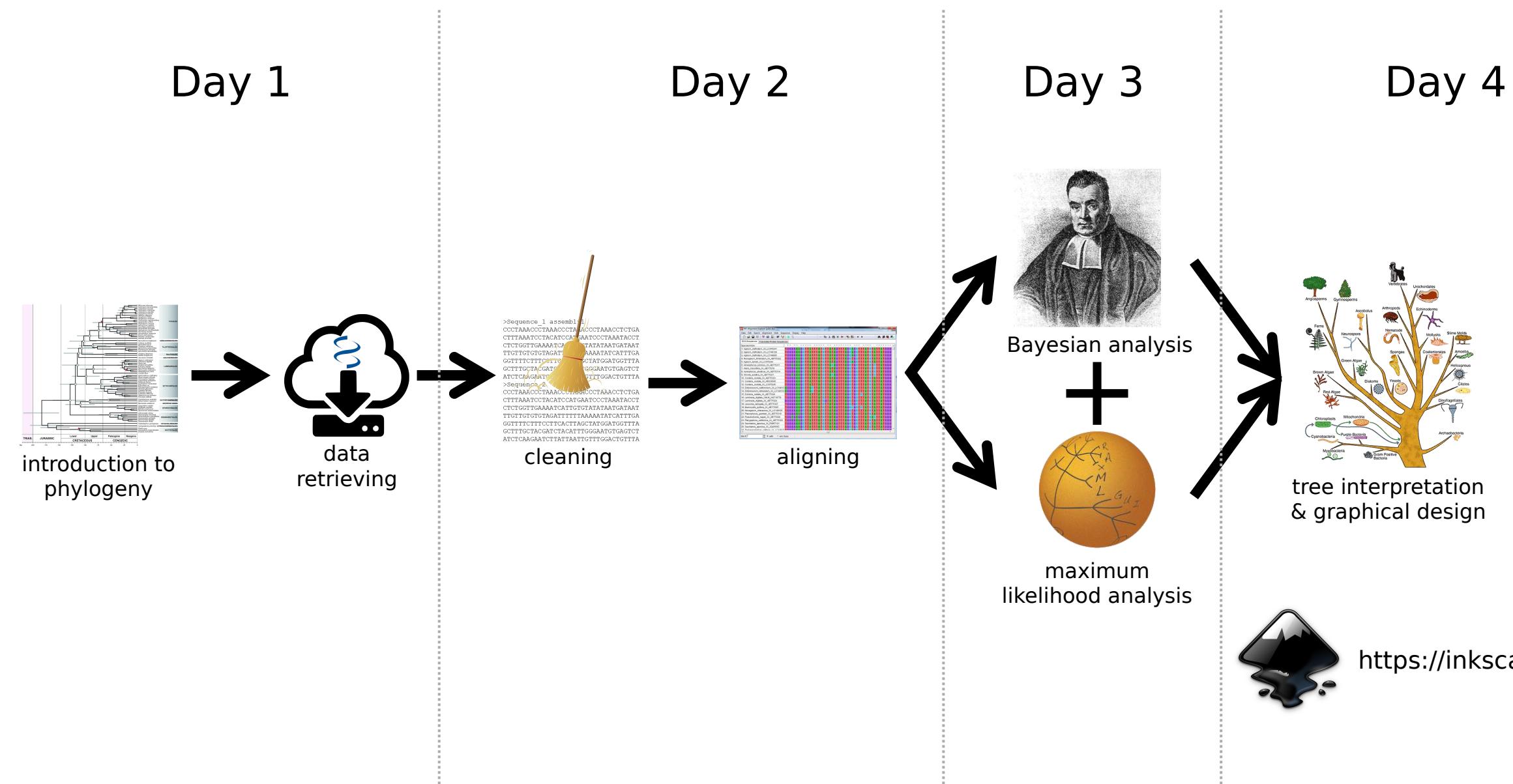
Current changelog: [version 4.4.2](#)

Citation: Letunic and Bork (2019) Nucleic Acids Res doi: 10.1093/nar/gkz239 | Privacy Policy

design & development: biobyte solutions

57

Program of the course



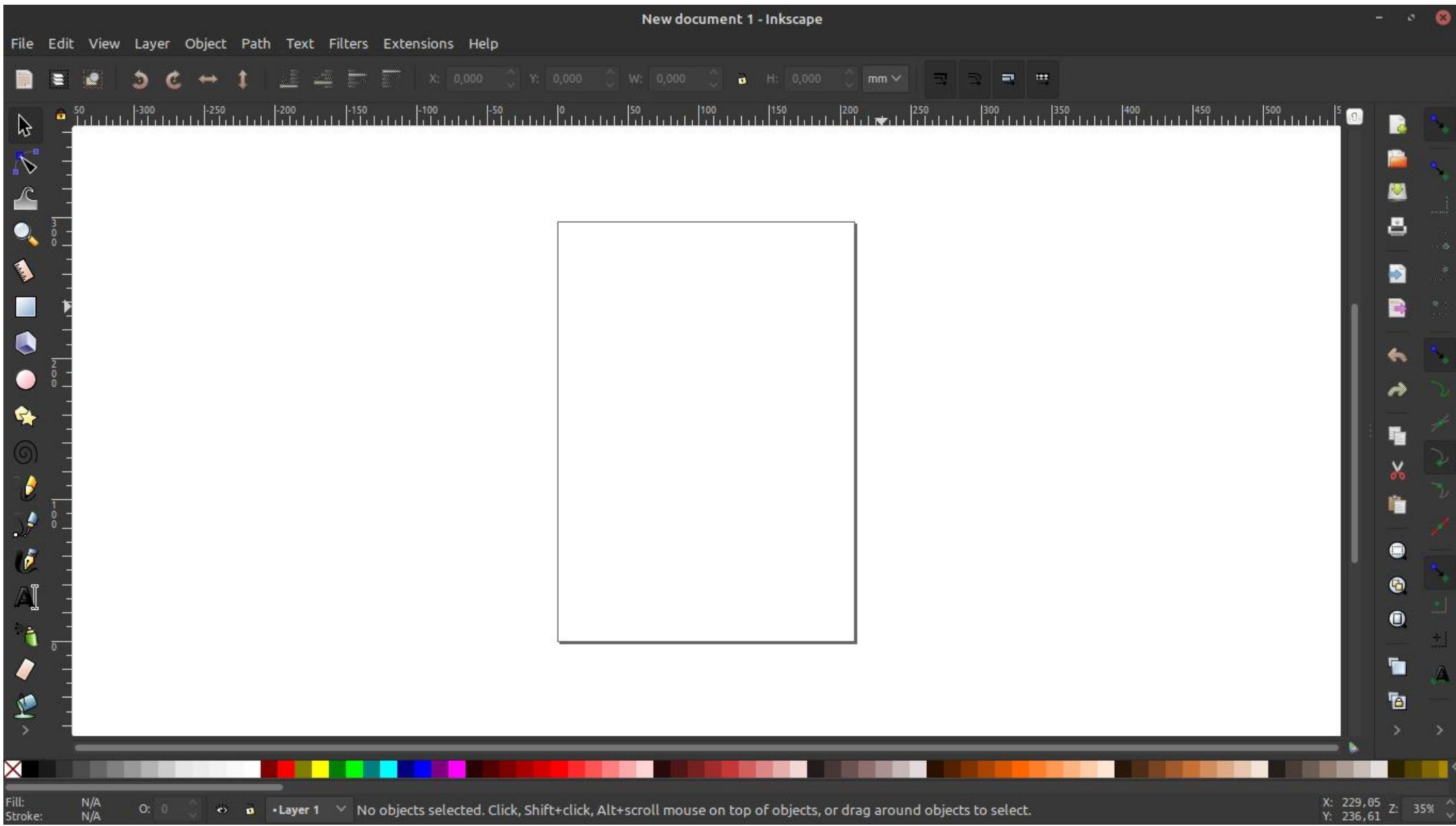
Q&A

BLAST, CD-hit, Q&A ...



<https://inkscape.org/>

Vectorial drawings with Inkscape





Any questions?

How to reference the different tools used

see 'References_to_programs.txt' on the github:

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