Phylogenetic trees from high-throughput sequence data analysis

Adriana Rego

Meta_Microbial Workshop

Metagenomic and bioinformatic insights into microbial communities

@AdrianaRego10

adrianairego@gmail.com

ciimar

U.PORTO

FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Sponsorship

SPECO

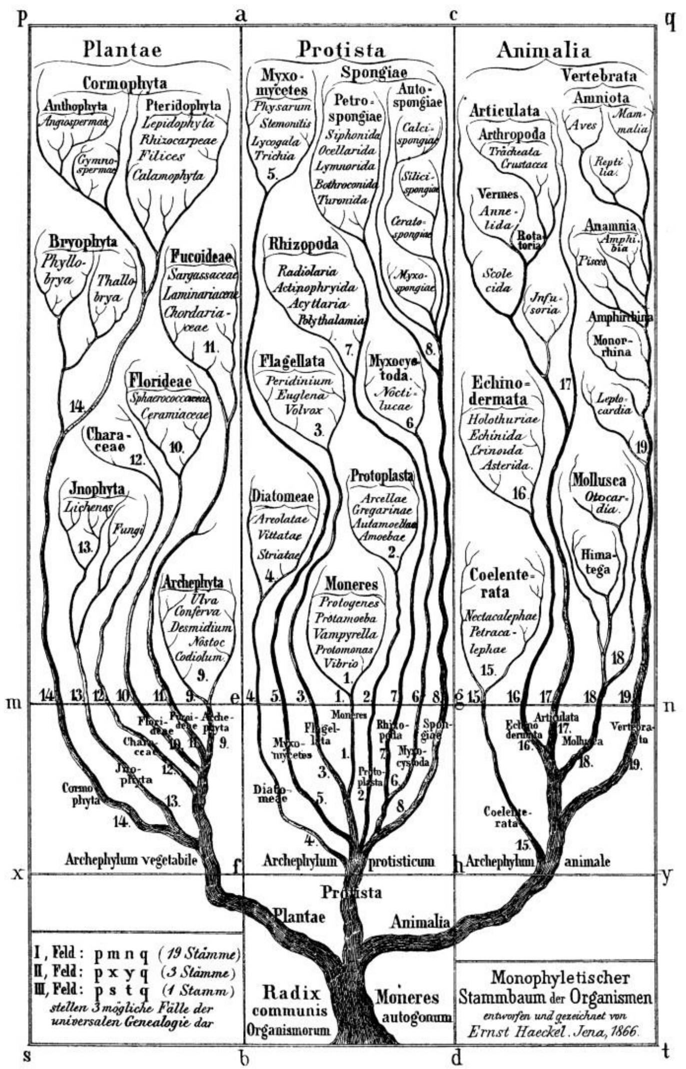
BIOPORTUGAL S.A.
Químico, Farmacêutica

Phylogenetic analysis

A **phylogenetic tree** or **evolutionary tree** is a diagrammatic representation of evolutionary relationships.

Relationships among taxa are inferred based on **homology** (inheritance from a common ancestor, commonly observed as patterns of sequence similarity).

Traditionally, phylogenies were constructed using **morphological data** only, but the advances in DNA sequencing enabled genetic information to be incorporated into phylogenetic analyses.



The earliest Tree of Life, by Ernest Haeckel (1866).

1 Basic concepts

Phylogenetic analysis

Steps in a phylogenetic analysis

1 Planning

- Start with a question
- Select a model (organisms or gene family)
- Choose approppriate molecular markers
- Collect or generate data (sequences)

2 Executing

- Alignment
- Phylogenetic method
- Phyogenetic analysis and tree reconstruction

3 Evaluate

- Tree evaluation
- Tree interpretation and new knowledge

1 Planning

Molecular marker genes

Examples

16S rRNA, 23S rRNA, ITS, rpoB, gyrB, dnaK, dsrAB, amoA, amoB, mip, horA, hitA, recA, ica, frc, oxc.

A **molecular marker** is defined as any DNA sequence which shows polymorphism and can be detected using a molecular technique.

Molecular markers of bacteria should
have several characteristics:

- housekeeping genes present in **all** bacterial species;
- High **polymorphism**, which make them distinguishable in different bacterial species;
- Highly **conserved** in some regions, which are easy to design appropriate primers to amplify by PCR**.**

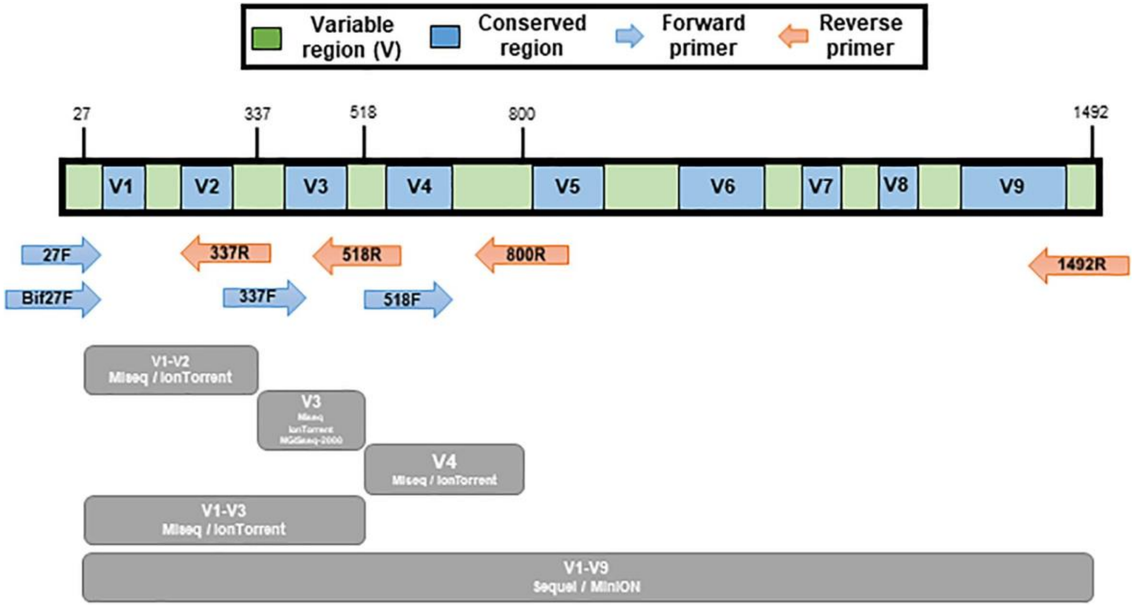
1 Planning

Marker genes in bacteria

16S rRNA gene

The sequence of 16S rRNA gene:

- Is **universal** in bacteria.
- Shows **evolutionary distance** and relationships between organisms and provides statistical and valid measurements for bacterial identification.
- Comprises several **conversed** and **variable** regions useful to design primers.



(Adapted from Park et.al 2021)

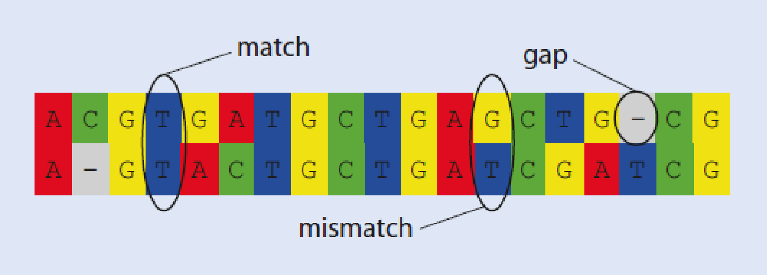
2 Executing

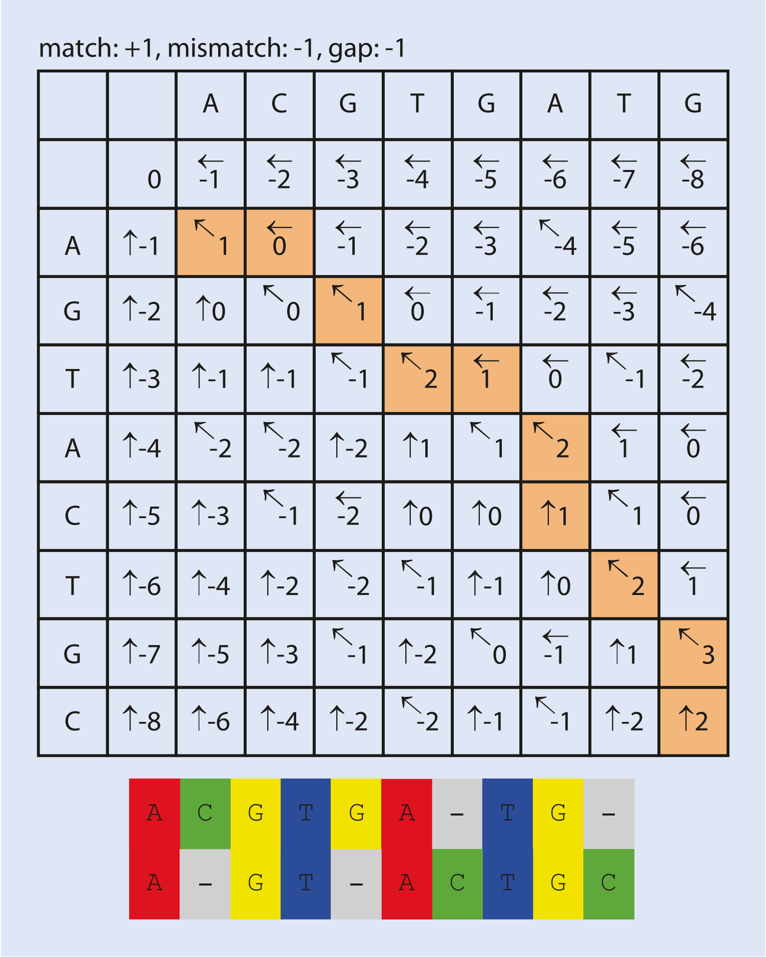
Alignments

For phylogenetic analyses, we should compare homologous genomic/gene positions → DNA sequences need to be aligned.

Alignments are hypotheses of **positional homologies** between nucleotides or amino acids of sequences.

The **Needleman and Wunsch algorithm** finds the optimal pairwise alignments of two sequences, which can contain matches, mismatches and gaps.


match
gap
A C G T G A T G C T G A G C T G – C G
A – G T A C T G C T G A T C G A T C G
mismatch



match: +1, mismatch: -1, gap: -1

		A	C	G	T	G	A	T	G
	0	←-1	←-2	←-3	←-4	←-5	←-6	←-7	←-8
A	↑-1	↖1	←0	←-1	←-2	←-3	↖-4	←-5	←-6
G	↑-2	↑0	↖0	↖1	←0	←-1	←-2	←-3	↖-4
T	↑-3	↑-1	↑-1	↖-1	↖2	←1	←0	↖-1	←-2
A	↑-4	↖-2	↖-2	↑-2	↑1	↖1	↖2	←1	←0
C	↑-5	↑-3	↖-1	←-2	↑0	↑0	↑1	↖1	←0
T	↑-6	↑-4	↑-2	↖-2	↖-1	↑-1	↑0	↖2	←1
G	↑-7	↑-5	↑-3	↖-1	↑-2	↖0	←-1	↑1	↖3
C	↑-8	↑-6	↑-4	↑-2	↖-2	↑-1	↖-1	↑-2	↑2

A	C	G	T	G	A	–	T	G	–
A	–	G	T	–	A	C	T	G	C

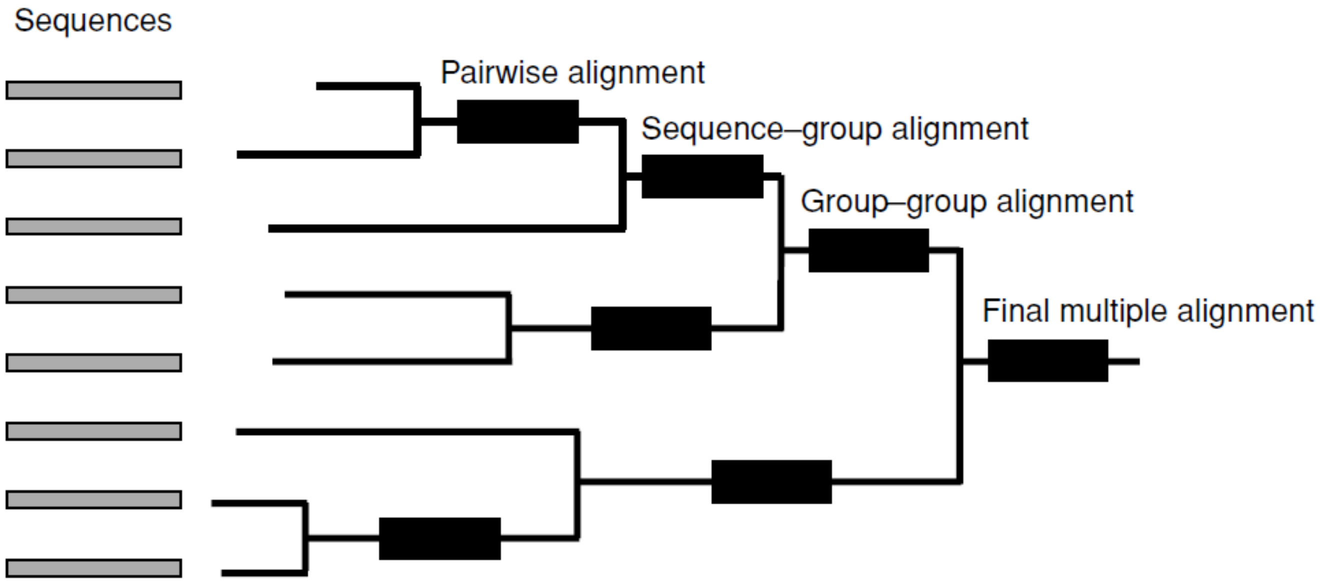
(Adapted from Bleidorn 2017)

2 Executing

Alignments

Multiple sequence alignments (MSA)

Progressive MSA



(Adapted from Des Higgins and Philippe Lemey in *The Phylogenetic Handbook*)

2 Executing

Alignments

For HTS data, are needed **fast** and accurate **multiple sequence alignments** (MSA).

Some popular **MSA** programs for HTS:

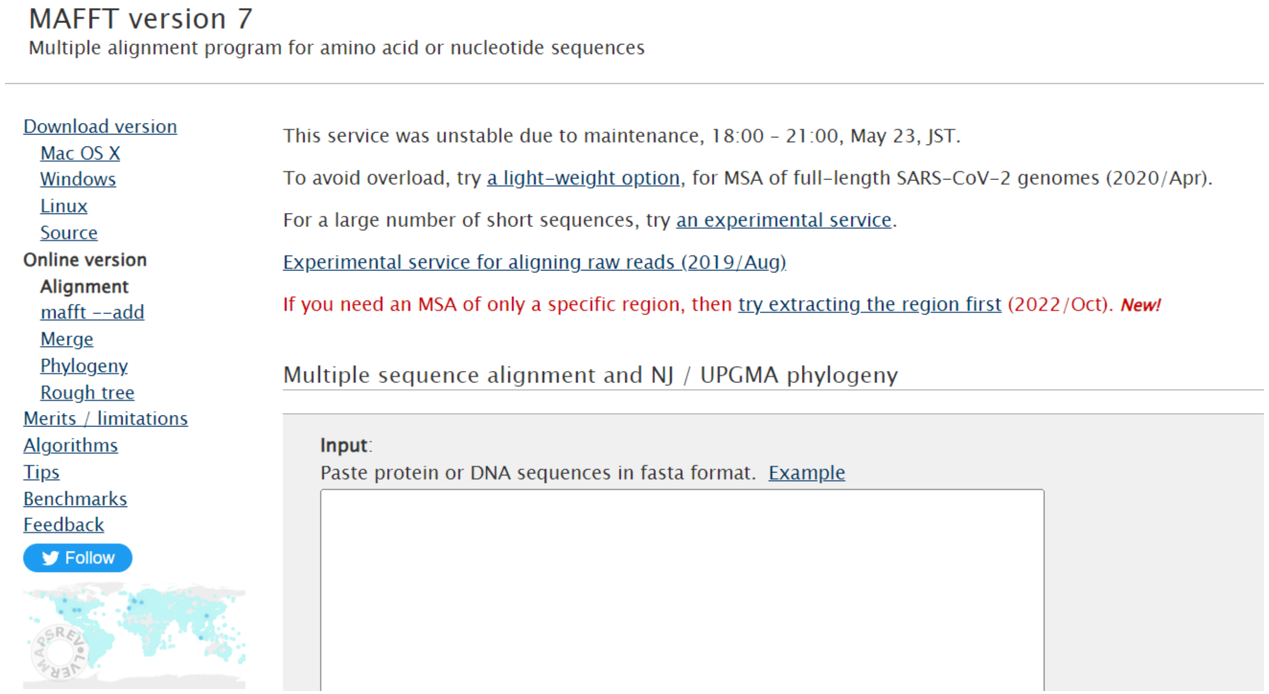
MAFFT - https://mafft.cbrc.jp/alignment/software/
MUSCLE - https://drive5.com/muscle5/
T-COFFEE - https://tcoffee.crg.eu/apps/tcoffee/index.html

2 Executing

Alignments

MAFFT is a MSA program for unix-like operating systems. It offers a range of multiple alignment methods, L-INS-i (accurate; for alignment of <~**200** sequences), FFT-NS-2 (fast; for alignment of <~**30,000** sequences).

-command line and online version (limited to 100 sequences)
- L-INS-i is one of the most accurate MSA methods currently available.
- suitable for SSU rRNA alignments



MAFFT version 7
Multiple alignment program for amino acid or nucleotide sequences

This service was unstable due to maintenance, 18:00 - 21:00, May 23, JST.

To avoid overload, try a light-weight option, for MSA of full-length SARS-CoV-2 genomes (2020/Apr).

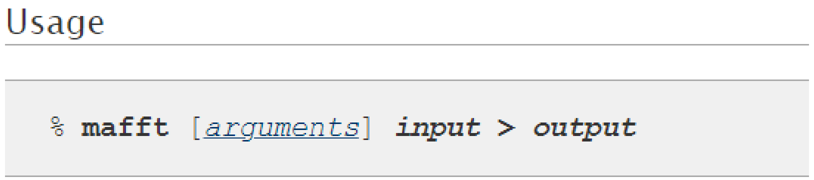
For a large number of short sequences, try an experimental service.

Experimental service for aligning raw reads (2019/Aug)

If you need an MSA of only a specific region, then try extracting the region first (2022/Oct). *New!*

Multiple sequence alignment and NJ / UPGMA phylogeny

Input:
Paste protein or DNA sequences in fasta format. Example



Usage

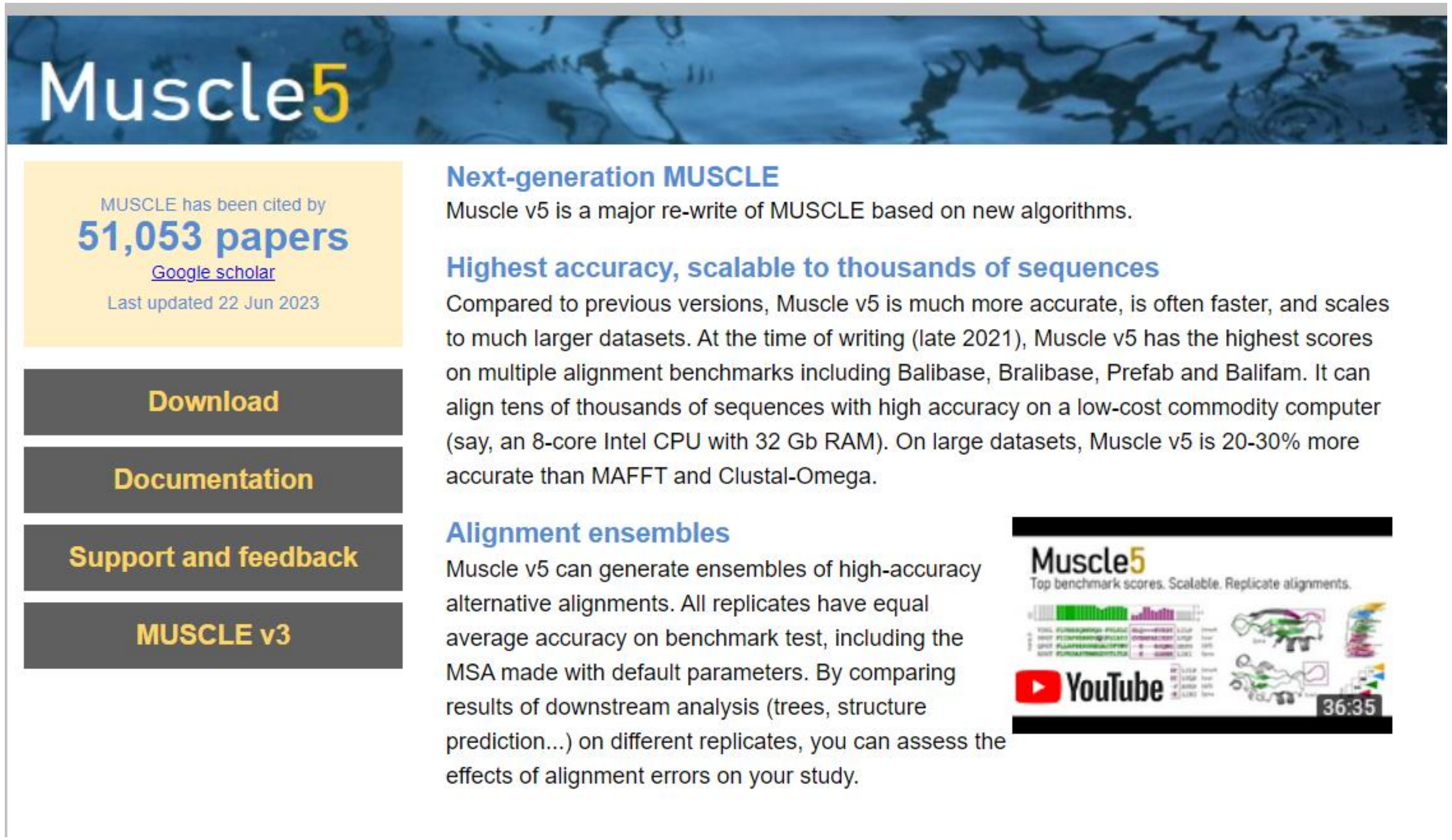
```
% mafft [arguments] input > output
```

2 Executing

Alignments

MUSCLE v5 performs accurate and faster MSAs, capable of escalating to larger datasets.

- on large datasets, Muscle v5 is 20-30% more accurate than MAFFT
- No graphical user interface



MUSCLE has been cited by
51,053 papers
Google scholar
Last updated 22 Jun 2023

Next-generation MUSCLE

Muscle v5 is a major re-write of MUSCLE based on new algorithms.

Highest accuracy, scalable to thousands of sequences

Compared to previous versions, Muscle v5 is much more accurate, is often faster, and scales to much larger datasets. At the time of writing (late 2021), Muscle v5 has the highest scores on multiple alignment benchmarks including Balibase, Bralibase, Prefab and Balifam. It can align tens of thousands of sequences with high accuracy on a low-cost commodity computer (say, an 8-core Intel CPU with 32 Gb RAM). On large datasets, Muscle v5 is 20-30% more accurate than MAFFT and Clustal-Omega.

Alignment ensembles

Muscle v5 can generate ensembles of high-accuracy alternative alignments. All replicates have equal average accuracy on benchmark test, including the MSA made with default parameters. By comparing results of downstream analysis (trees, structure prediction...) on different replicates, you can assess the effects of alignment errors on your study.

Examples

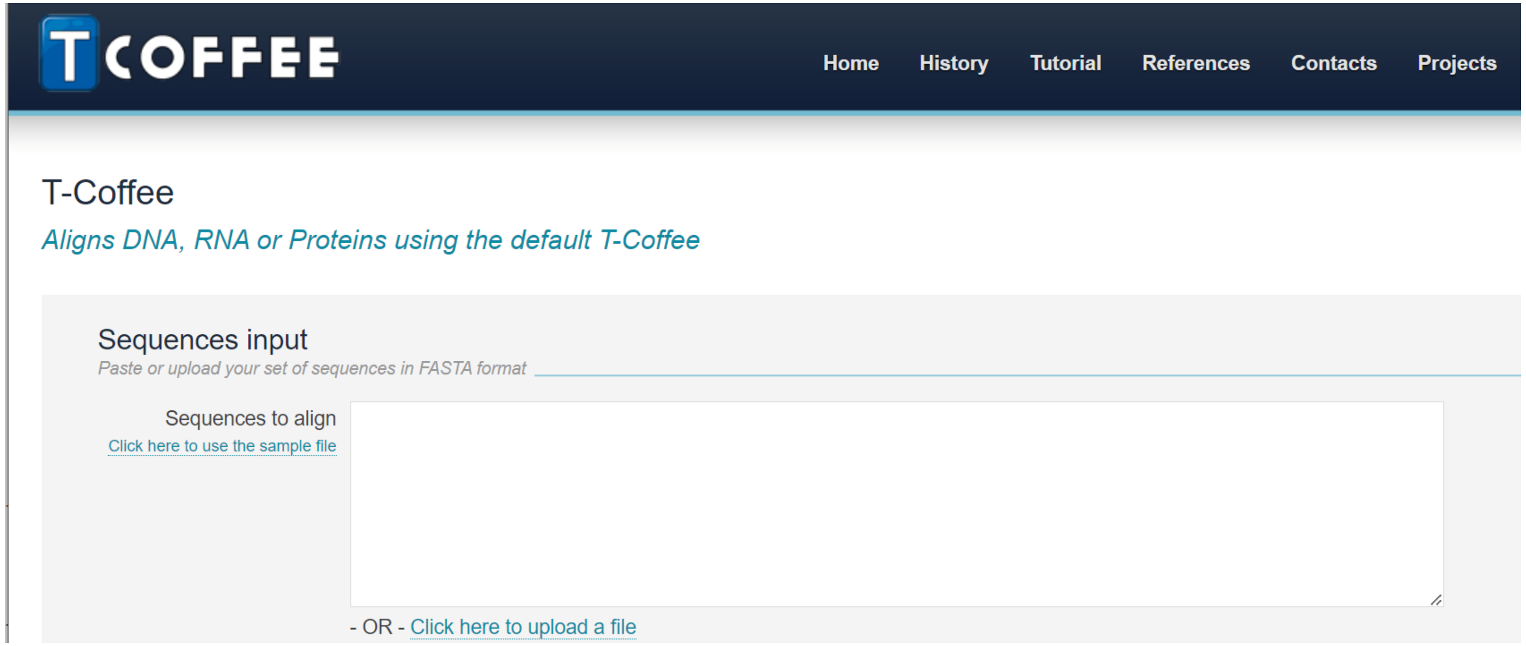
```
muscle -align seqs.fa -output aln.afa
```

2 Executing

Alignments

T-Coffee can be used to align sequences or to combine the output of your favourite alignment methods (Clustal, Mafft, Probcons, Muscle...) into one unique alignment (M-Coffee).

-web interface and command line option



T-Coffee

Aligns DNA, RNA or Proteins using the default T-Coffee

Sequences input

Paste or upload your set of sequences in FASTA format

Sequences to align

Click here to use the sample file

- OR - Click here to upload a file

mode	**Expresso**
Seq.	All
Accuracy	High

```
t_coffee foo.seq -mode
expresso
```

2 Executing

Alignments

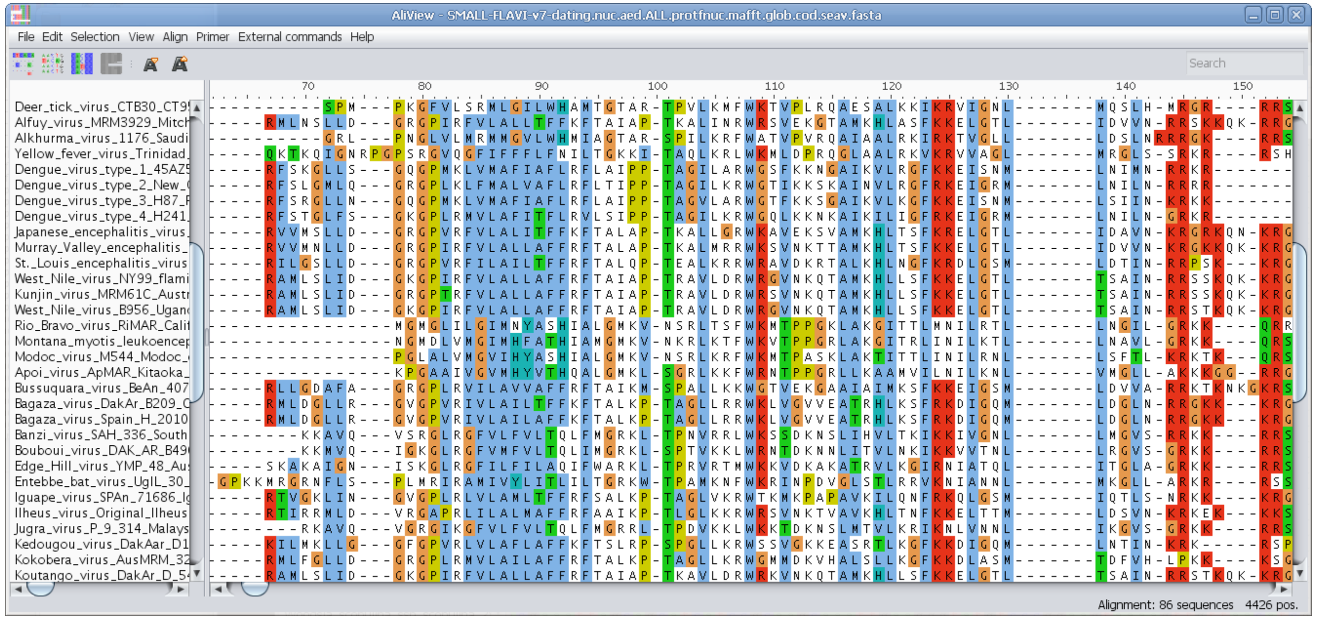
Visualization of MSA from HTS data

AliView - https://ormbunkar.se/aliview/
- one of the fastest alignment viewer and editor
- created to work with large alignments.

Automatic **trimming of MSA**
trimAl - http://trimal.cgenomics.org/trimal

- automated removal of spurious sequences or poorly aligned regions from a MSA
- Command line and webserver available

AliView


AliView - SMALL-FLAVI-v7-dating.nuc.aed.ALL.protfnuc.mafft.glob.cod.seav.fasta
File Edit Selection View Align Primer External commands Help
Search

```
                                   70        80        90        100       110       120       130       140       150
Deer_tick_virus_CTB30_CT9      ----------SPM---PKGFVLSRMLGILWHAMTGTAR-TPVLKMFWKTVPLRQAESALKKIKRVIGNL--------MQSLH-MRGR----RRS
Alfuy_virus_MRM3929_Mitch      -----RMLNSLLD---GRGPIRFVLALLTFFKFTAIAP-TKALINRWRSVEKGTAMKHLASFKKELGTL--------IDVVN-RRSKKQK-RRG
Alkhurma_virus_1176_Saudi      ----------GRL---PNGLVLMRMMGVLWHMIAGTAR-SPILKRFWATVPVRQAIAALRKIRKTVGLL--------LDSLNRRRGK----RRS
Yellow_fever_virus_Trinidad_   -----QKTKQIGNRPGPSRGVQGFIFFFLFNILTGKKI-TAQLKRLWKMLDPRQGLAALRKVKRVVAGL--------MRGLS-SRKR----RSH
Dengue_virus_type_1_45AZ5      -----RFSKGLLS---GQGPMKLVMAFIAFLRFLAIPP-TAGILARWGSFKKNGAIKVLRGFKKEISNM--------LNIMN-RRKR-------
Dengue_virus_type_2_New_(      -----RFSLGMLQ---GRGPLKLFMALVAFLRFLTIPP-TAGILKRWGTIKKSKAINVLRGFRKEIGRM--------LNILN-RRRR-------
Dengue_virus_type_3_H87_F      -----RFSRGLLN---GQGPMKLVMAFIAFLRFLAIPP-TAGVLARWGTFKKSGAIKVLKGFKKEISNM--------LSIIN-KRKK-------
Dengue_virus_type_4_H241_      -----RFSTGLFS---GKGPLRMVLAFITFLRVLSIPP-TAGILKRWGQLKKNKAIKILIGFRKEIGRM--------LNILN-GRKR-------
Japanese_encephalitis_virus_   -----RVVMSLLD---GRGPVRFVLALITFFKFTALAP-TKALLGRWKAVEKSVAMKHLTSFKRELGTL--------IDAVN-KRGRKQN-KRG
Murray_Valley_encephalitis_    -----RVVMNLLD---GRGPIRFVLALLAFFRFTALAP-TKALMRRWKSVNKTTAMKHLTSFKKELGTL--------IDVVN-KRGKKQK-KRG
St._Louis_encephalitis_virus   -----RILGSLLD---GRGPVRFILAILTFFRFTALQP-TEALKRRWRAVDKRTALKHLNGFKRDLGSM--------LDTIN-RRPSK---KRG
West_Nile_virus_NY99_flami     -----RAMLSLID---GKGPIRFVLALLAFFRFTAIAP-TRAVLDRWRGVNKQTAMKHLLSFKKELGTL--------TSAIN-RRSSKQK-KRG
Kunjin_virus_MRM61C_Austr      -----RAMLSLID---GRGPTRFVLALLAFFRFTAIAP-TRAVLDRWRSVNKQTAMKHLLSFKKELGTL--------TSAIN-RRSSKQK-KRG
West_Nile_virus_B956_Ugand     -----RAMLSLID---GKGPIRFVLALLAFFRFTAIAP-TRAVLDRWRGVNKQTAMKHLLSFKKELGTL--------TSAIN-RRSTKQK-KRG
Rio_Bravo_virus_RiMAR_Calif    ----------------MGMGLILGIMNYASHIALGMKV-NSRLTSFWKMTPPGKLAKGITTLMNILRTL--------LNGIL-GRKK----QRR
Montana_myotis_leukoencep      ----------------NGMDLVMGIMHFATHIAMGMKV-NKRLKTFWKVTPPGRLAKGITRLINILKTL--------LNAVL-GRKK----QRS
Modoc_virus_M544_Modoc_(       ----------------PGLALVMGVIHYASHIALGMKV-NSRLKRFWKMTPASKLAKTITTLINILRNL--------LSFTL-KRKTK---QRS
Apoi_virus_ApMAR_Kitaoka_      ----------------KPGAAIVGVMHYVTHQALGMKL-SGRLKKFWRNTPPGRLLKAAMVILNILKNL--------VMGLL-AKKKGG--RRG
Bussuquara_virus_BeAn_407      -----RLLGDAFA---GRGPLRVILAVVAFFRFTAIKM-SPALLKKWGTVEKGAAIAIMKSFKKEIGSM--------LDVVA-RRKTKNKGKRS
Bagaza_virus_DakAr_B209_C      -----RMLDGLLR---GVGPVRIVLAILTFFKFTALKP-TAGLLRRWKLVGVVEATRHLKSFRKDIGQM--------LDGLN-RRGKK---KRG
Bagaza_virus_Spain_H_2010      -----RMLDGLLR---GVGPVRIVLAILAFFKFTALKP-TAGLLRRWKLVGVVEATRHLKSFRKDIGQM--------LDGLN-RRGKK---KRG
Banzi_virus_SAH_336_South      --------KKAVQ---VSRGLRGFVLFVLTQLFMGRKL-TPNVRRLWKSSDKNSLIHVLTKIKKIVGNL--------LMGVS-RRKK----RRS
Bouboui_virus_DAK_AR_B49       --------KKMVQ---IGKGLRGFVMFVLTQLIMGRKL-SPTVKKLWRNTDKNNLITVLNKIKKVVTNL--------LRGVS-GRKK----RRS
Edge_Hill_virus_YMP_48_Au      -----SKAKAIGN---ISKGLRGFILFILAQIFWARKL-TPRVRTMWKKVDKAKATRVLKGIRNIATQL--------ITGLA-GRKK----RRS
Entebbe_bat_virus_UgIL_30_     -GPKKMRGRNFLS---PLMRIRAMIVYLITLILTGRKW-TPAMKNFWKRINPDVGLSTLRRVKNIANNL--------MKGLL-ARKR----RSS
Iguape_virus_SPAn_71686_Ig     -----RTVGKLIN---GVGPLRLVLAMLTFFRFSALKP-TAGLVKRWTKMKPAPAVKILQNFRKQLGSM--------IQTLS-NRKK----KRG
Ilheus_virus_Original_Ilheus   -----RTIRRMLD---VRGAPRLILALMAFFRFAAIKP-TLGLKKRWRSVNKTVAVKHLTNFKKELTTM--------LDSVN-KRKEK---KKS
Jugra_virus_P_9_314_Malays     --------RKAVQ---VGRGIKGFVLFVLTQLFMGRRL-TPDVKKLWKKTDKNSLMTVLKRIKNLVNNL--------IKGVS-GRKK----RRS
Kedougou_virus_DakAar_D1       -----KILMKLLG---GFGPVRLVLAFLAFFKFTSLRP-SPGLLKRWSSVGKKEASRTLKGFKKDIGQM--------LNTIN-KRK-----RSP
Kokobera_virus_AusMRM_32       -----RMLFGLLD---GRGPLRMVLAILAFFRFTALKP-TAGLLKRWGMMDKVHALSLLKGFKKDLASM--------TDFVH-LPKK----KSG
Koutango_virus_DakAr_D_5       -----RAMLSLID---GKGPIRFVLALLAFFRFTAIAP-TKAVLDRWRKVNKQTAMKHLLSFKKELGTL--------TSAIN-RRSTKQK-KRG
```

Alignment: 86 sequences 4426 pos.

trimAl
A tool for automated alignment trimming

2 Executing

Phylogenetic analysis

Most commonly used **phylogenetic methods** for HTS data:

Maximum-parsimony – Character-based method - Which tree has the fewest mutations?
- fast, appropriate for very similar sequences and a small number of sequences

Maximum-Likelihood – Probabilistic method - Which tree has the highest probability given the observed alignment?
- Suitable for very dissimilar sequences

For **HTS** data, popular softwares include:
RaxML - https://github.com/amkozlov/raxml-ng
FastTree - http://www.microbesonline.org/fasttree/
PhyML - http://www.atgc-montpellier.fr/phyml/
IQ-TREE - http://www.iqtree.org/

2 Executing

Phylogenetic analysis

Software for phylogenetic inference

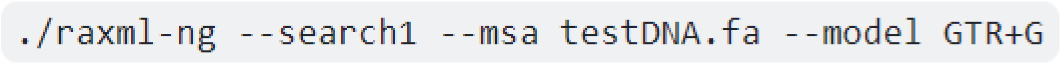
RaxML - https://github.com/amkozlov/raxml-ng

RAxML-NG is a phylogenetic tree inference tool which uses maximum-likelihood (ML) optimality criterion.

-developed for handling large datasets with low memory consumption and advanced search algorithms

-command line version

Usage example



```
./raxml-ng --search1 --msa testDNA.fa --model GTR+G
```

2 Executing

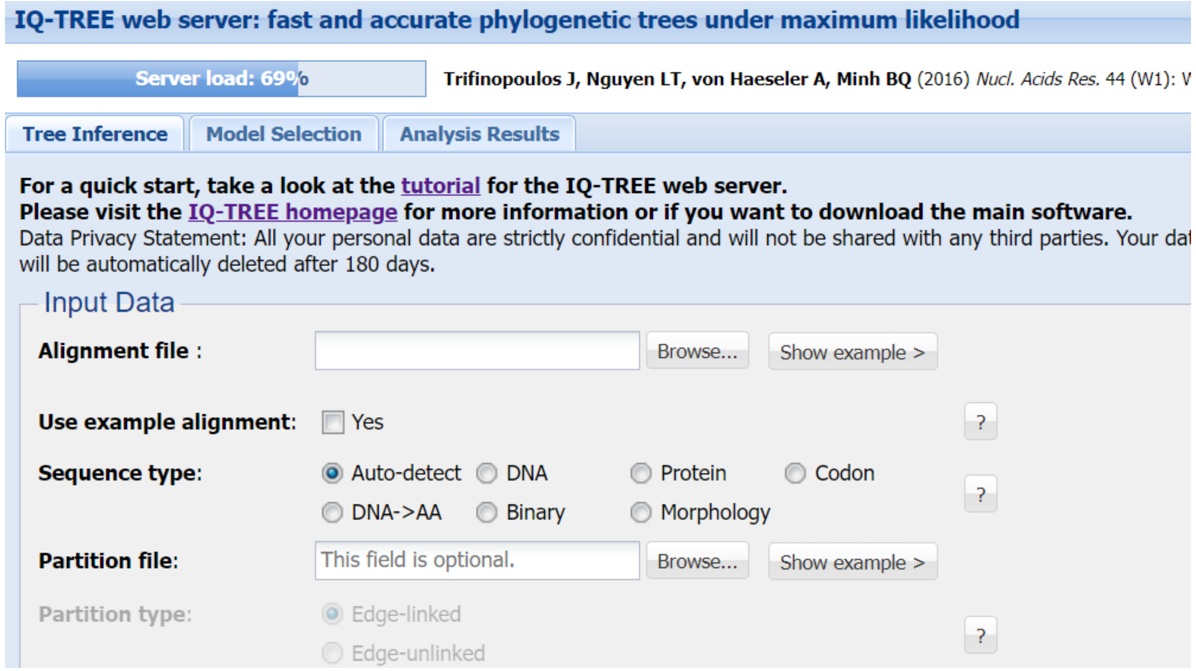
Phylogenetic analysis

Software for phylogenetic inference

IQ-TREE - http://www.iqtree.org/

A fast and effective algorithm to infer phylogenetic trees by ML.
- IQ-TREE compares favorably to RAxML and PhyML in terms of likelihoods with similar computing time.

- supports datasets with thousands of sequences or millions of alignment sites
- -command line and webserver available (http://iqtree.cibiv.univie.ac.at/)



IQ-TREE web server: fast and accurate phylogenetic trees under maximum likelihood

Server load: 69%

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) *Nucl. Acids Res.* 44 (W1): V

For a quick start, take a look at the tutorial for the IQ-TREE web server.
Please visit the IQ-TREE homepage for more information or if you want to download the main software.
Data Privacy Statement: All your personal data are strictly confidential and will not be shared with any third parties. Your dat will be automatically deleted after 180 days.

Input Data

Alignment file : Browse... Show example >

Use example alignment: □ Yes ?

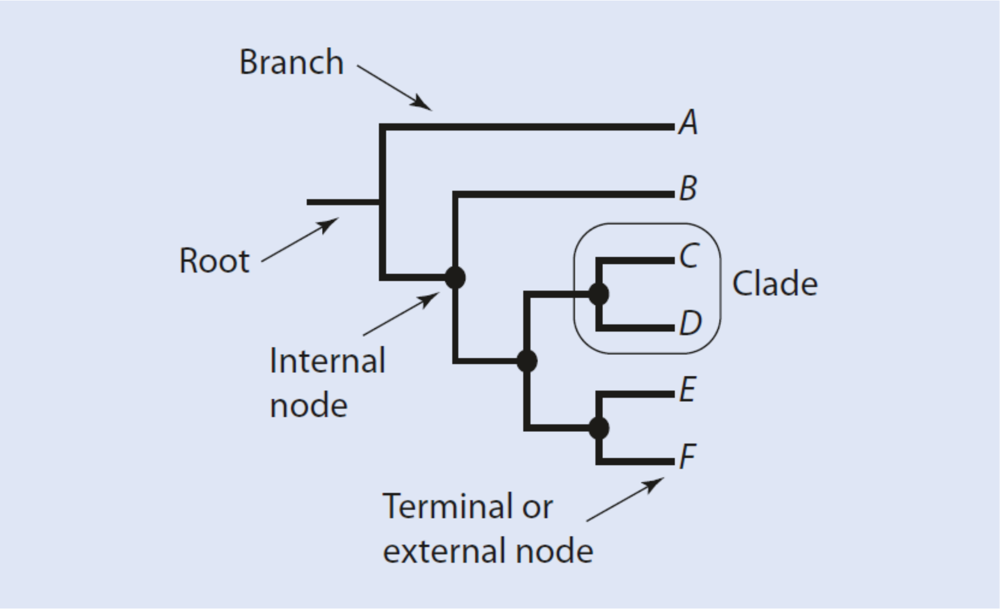
Sequence type: ● Auto-detect ○ DNA ○ Protein ○ Codon ○ DNA->AA ○ Binary ○ Morphology ?

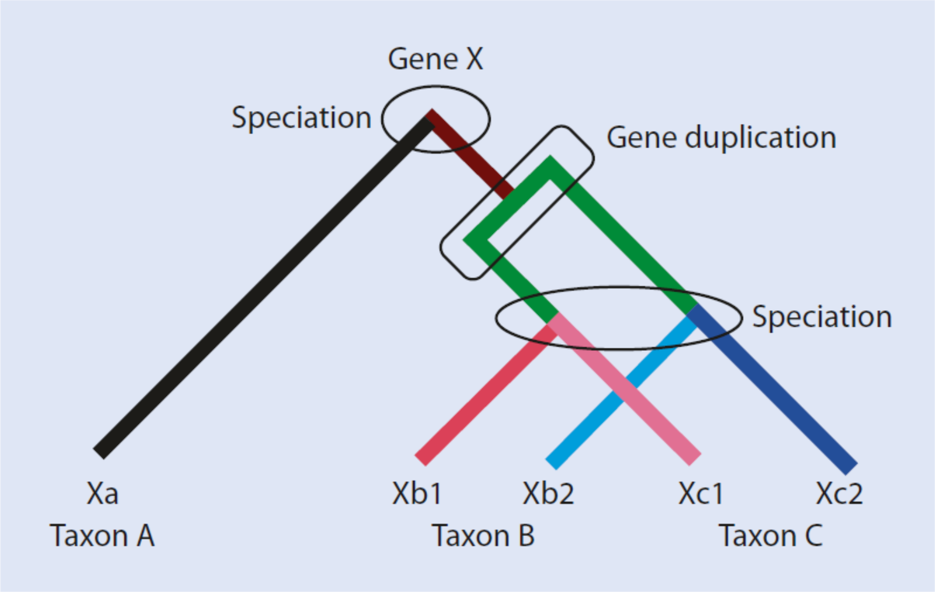
Partition file: This field is optional. Browse... Show example >

Partition type: ● Edge-linked ○ Edge-unlinked ?

3 Evaluate

How to interpret a phylogenetic tree?





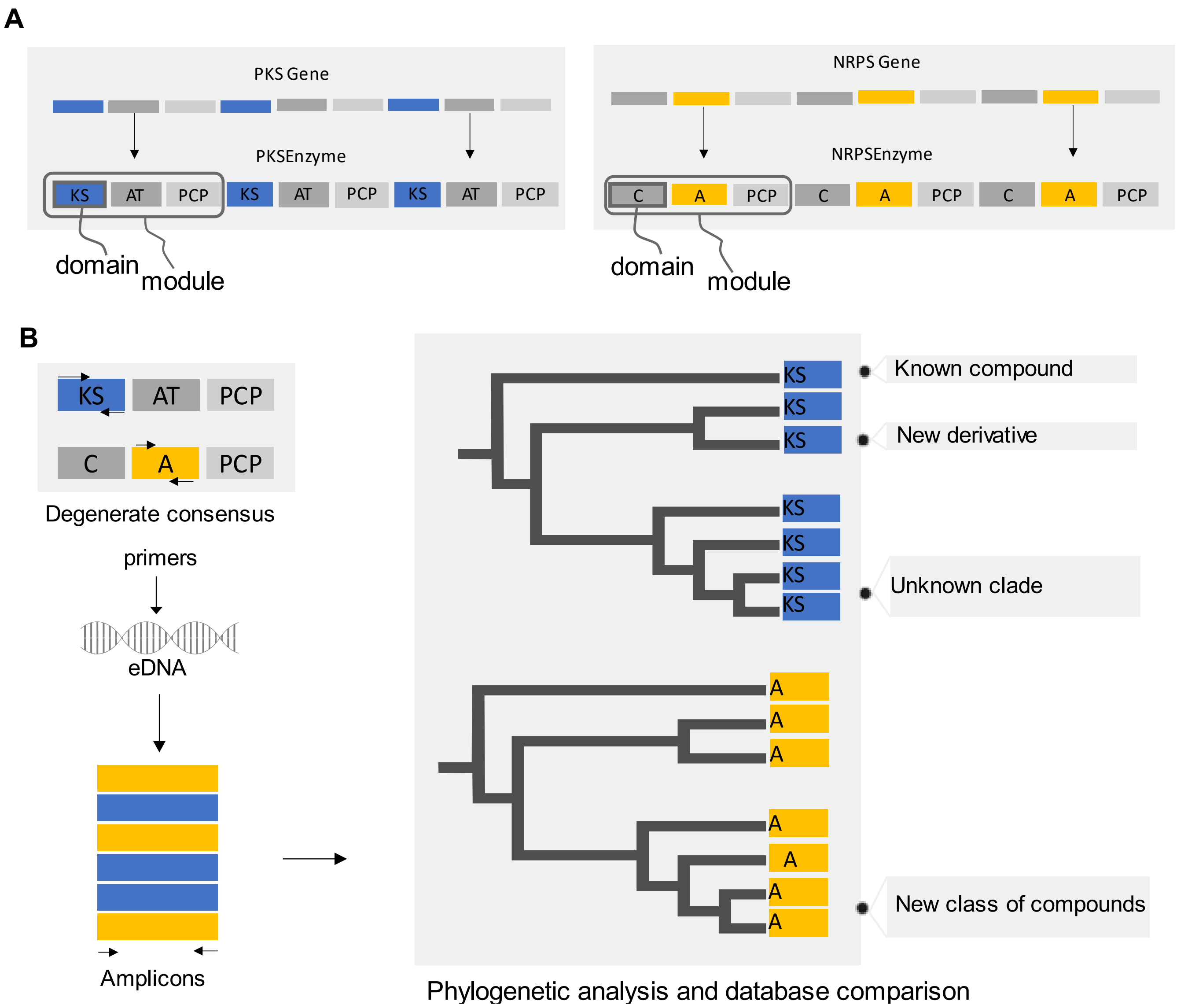
(Adapted from Bleidorn 2017)

5 Examples

Functional genes

Biosynthetic domains (KS and A) are highly conserved and have proven to be very informative in a phylogenetic context.

Biosynthetic gene clusters (BGCS) encoding for structurally related metabolites are predicted to share **common ancestry** and exhibit high sequence **identity** among these conserved domains.



(Rego 2023)

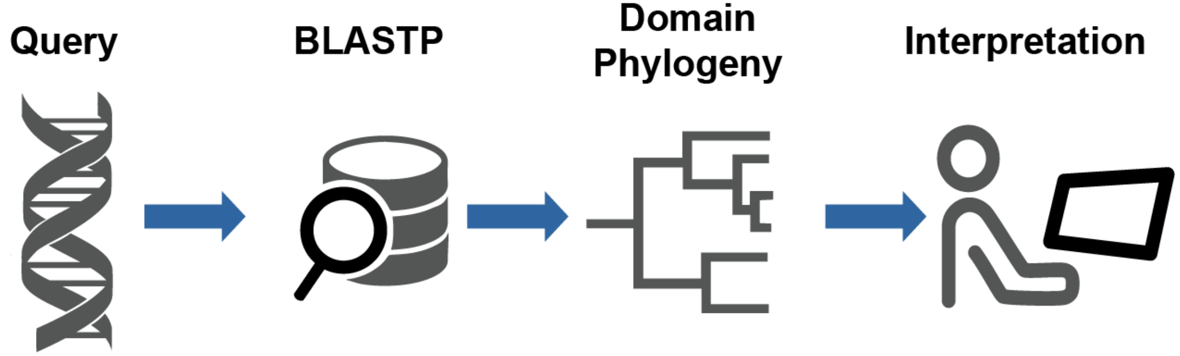
5 Examples

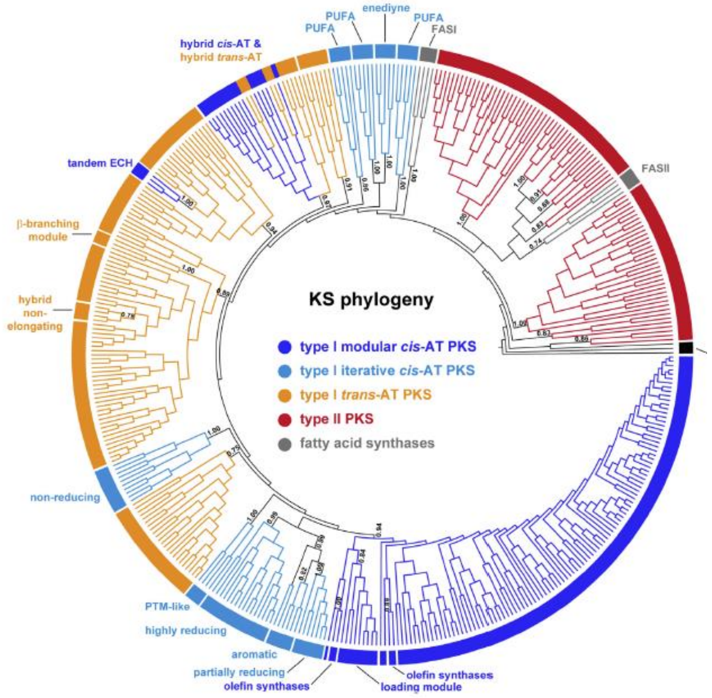
Functional genes

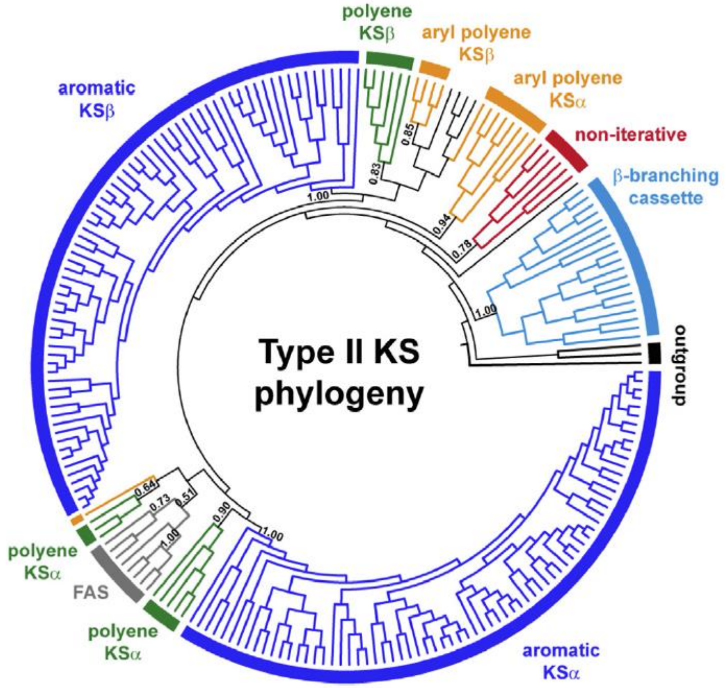
NAPDOS2.0

NaPDoS2, the second-generation Natural Product Domain Seeker, rapidly detects and classifies **ketosynthase (KS)** and **condensation (C)** domains from genomic, metagenomic, or PCR amplicon sequence data.

https://npdomainseeker.sdsc.edu/napdos2/







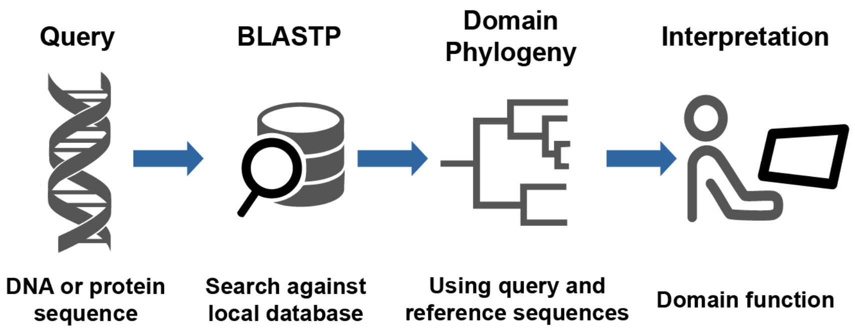
Functional genes

NAPDOS2.0

Main applications of NAPDOS approach

Poorly assembled (fragmented) genomes and metagenomes
Amplicons
Fast estimate of biosynthetic potential
New clades = new functionality

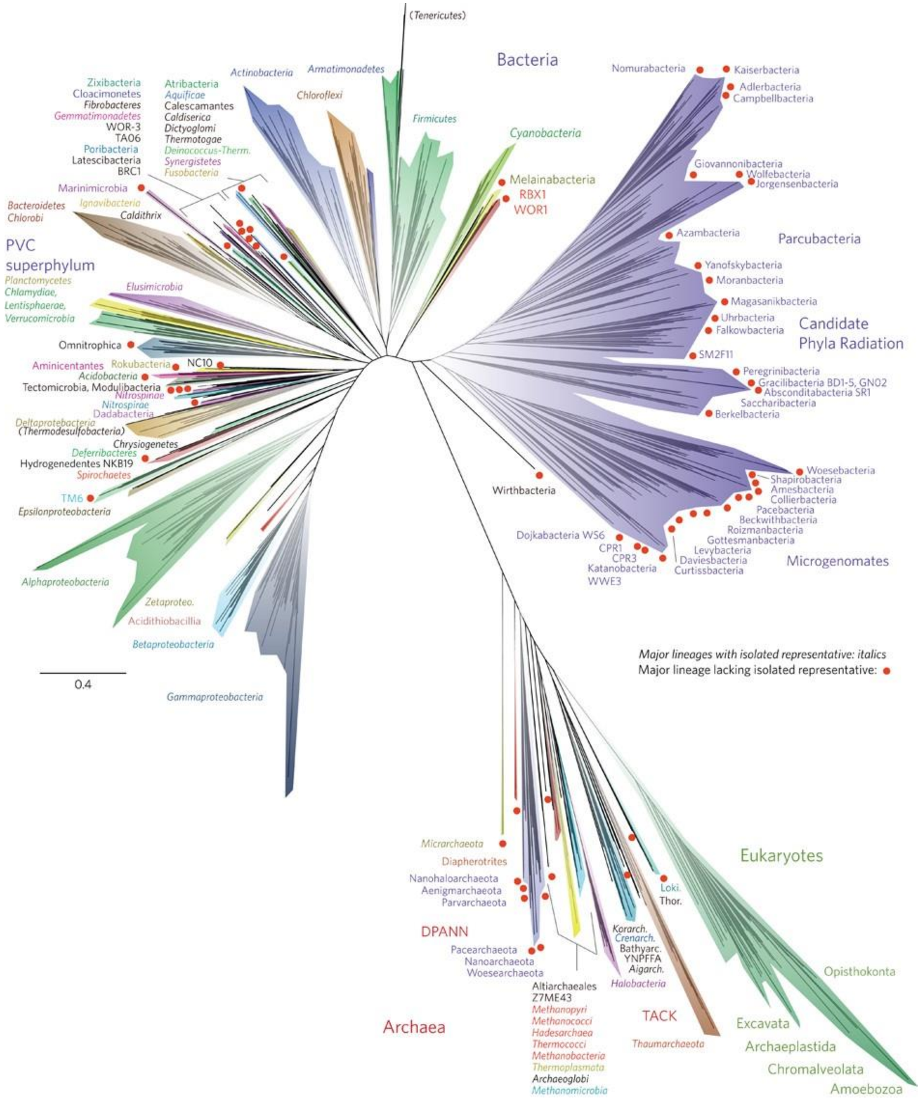
Limitations: comparison to database, new diversity might not be detected



Phylogenomics

Phylogenomics involves the reconstruction of evolutionary relationships by comparing sequences of **whole genomes** or **portions** of genomes.

Genome-scale phylogenetic datasets yield an increase in the **statistical confidence** of inferred relationships, often yielding maximally **supported** species trees.

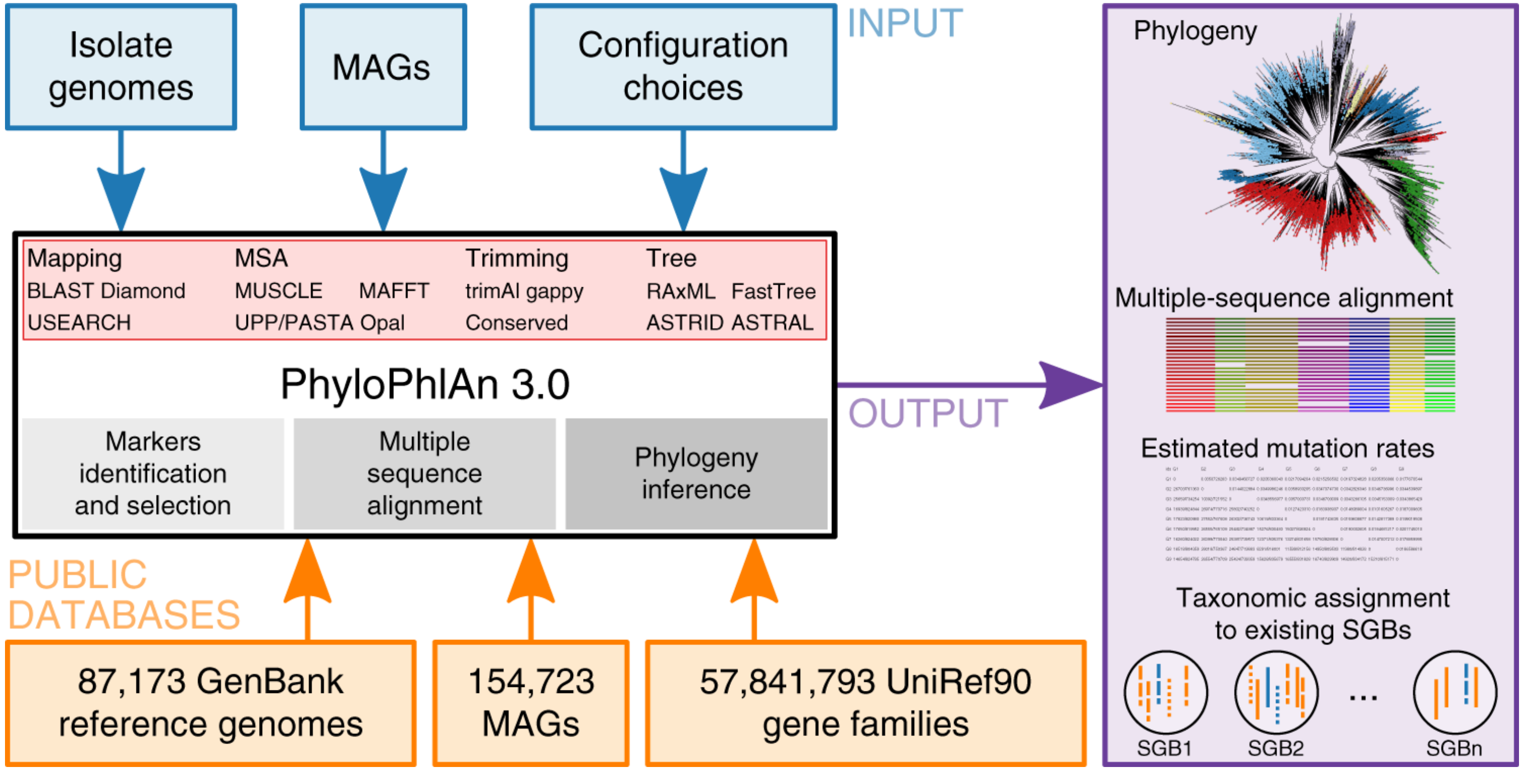


A current view of the tree of life, encompassing the total diversity represented by **sequenced genomes** (Hug et al. 2016).

6 Phylogenomics

PhyloPhlan

PhyloPhlAn can reconstruct strain-level phylogenies using clade-specific informative **phylogenetic markers**, and can also scale to very-large phylogenies comprising **>17,000** microbial species.



INPUT

Isolate genomes

MAGs

Configuration choices

Mapping
BLAST Diamond
USEARCH

MSA
MUSCLE MAFFT
UPP/PASTA Opal

Trimming
trimAl gappy
Conserved

Tree
RAxML FastTree
ASTRID ASTRAL

PhyloPhlAn 3.0

Markers identification and selection

Multiple sequence alignment

Phylogeny inference

OUTPUT

Phylogeny

Multiple-sequence alignment

Estimated mutation rates

Taxonomic assignment to existing SGBs

SGB1 SGB2 ... SGBn

PUBLIC DATABASES

87,173 GenBank reference genomes

154,723 MAGs

57,841,793 UniRef90 gene families

https://huttenhower.sph.harvard.edu/phylophlan/

Visualization of phylogenetic/phylogenomic trees

Web Portals

Interactive Tree of Life viewer (iTOL) – https://itol.embl.de/
Evolview - https://www.evolgenius.info/help/

Software/libraries

FigTree - http://tree.bio.ed.ac.uk/software/figtree/
GraphLan - https://huttenhower.sph.harvard.edu/graphlan/
ggtree: an R package for visualization of phylogenetic trees with their annotation data- https://github.com/YuLab-SMU/ggtree

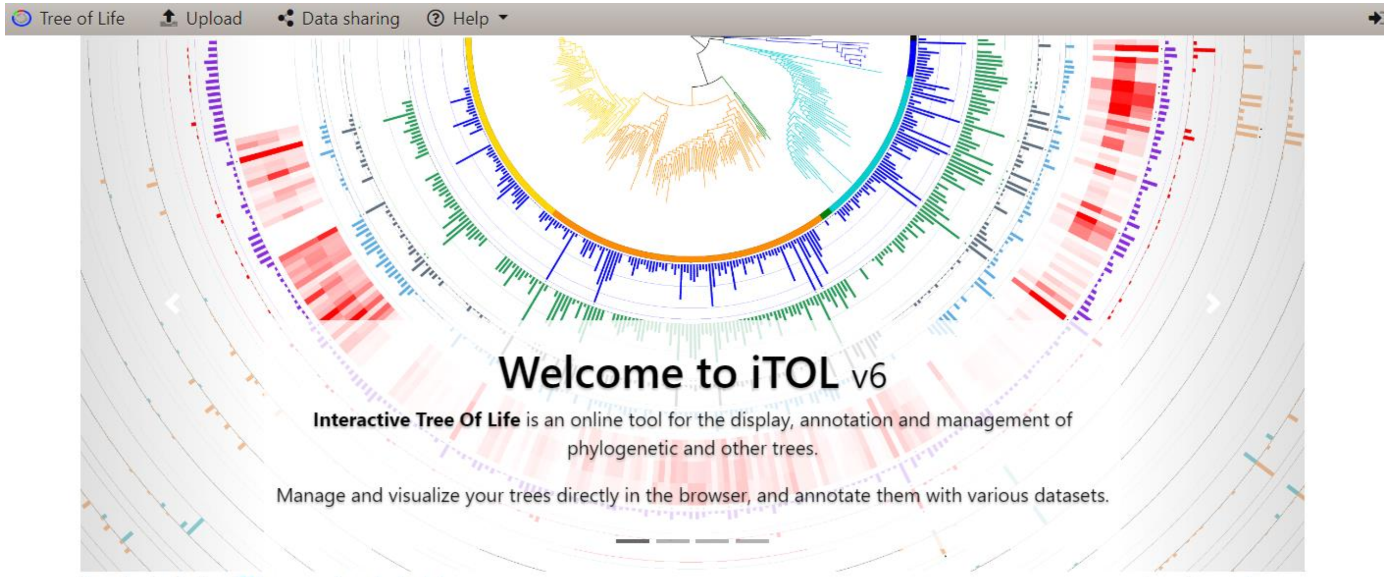
https://en.wikipedia.org/wiki/List_of_phylogenetic_tree_visualization_software

Visualization of phylogenetic/phylogenomic trees

Web Portals

Interactive Tree of Life viewer (**iTOL**) – https://itol.embl.de/

- Web-based
- User friendly (interactive interface)
- Highly costumizable



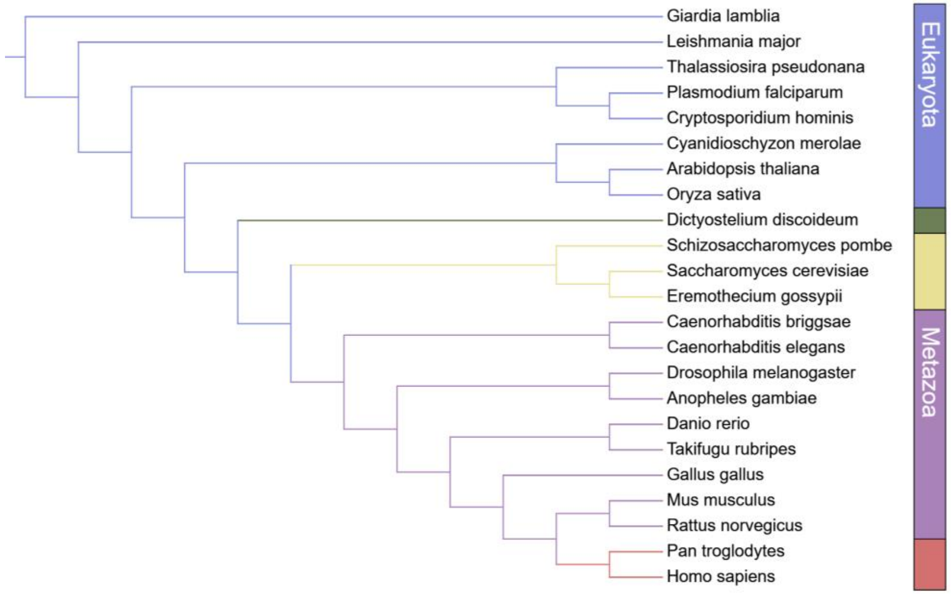
Welcome to iTOL v6

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

Manage and visualize your trees directly in the browser, and annotate them with various datasets.

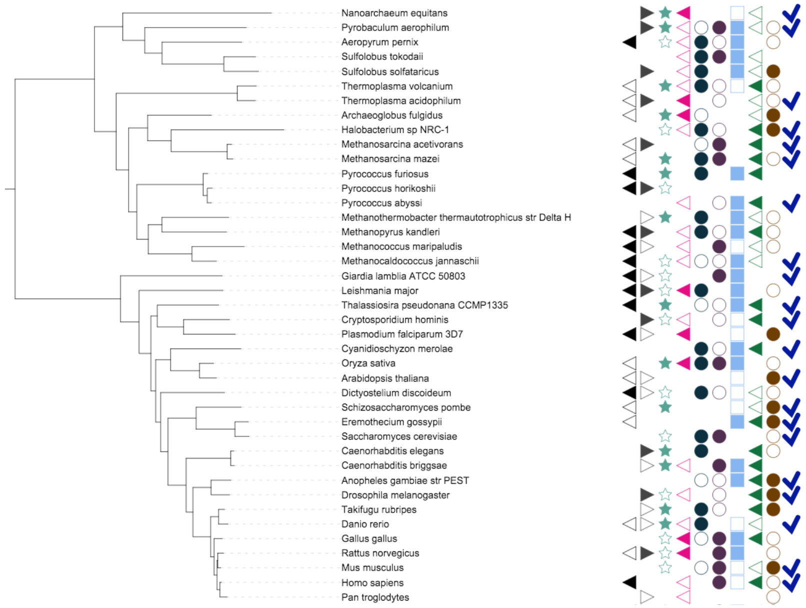
6 Visualization

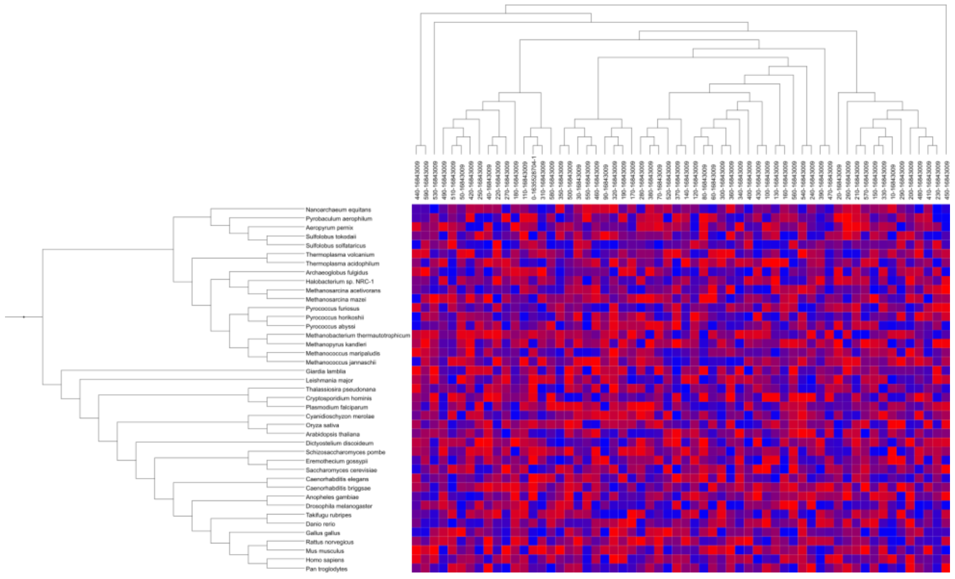
Visualization of phylogenetic/phylogenomic trees



Web Portals

Interactive Tree of Life viewer (**iTOL**) – https://itol.embl.de/





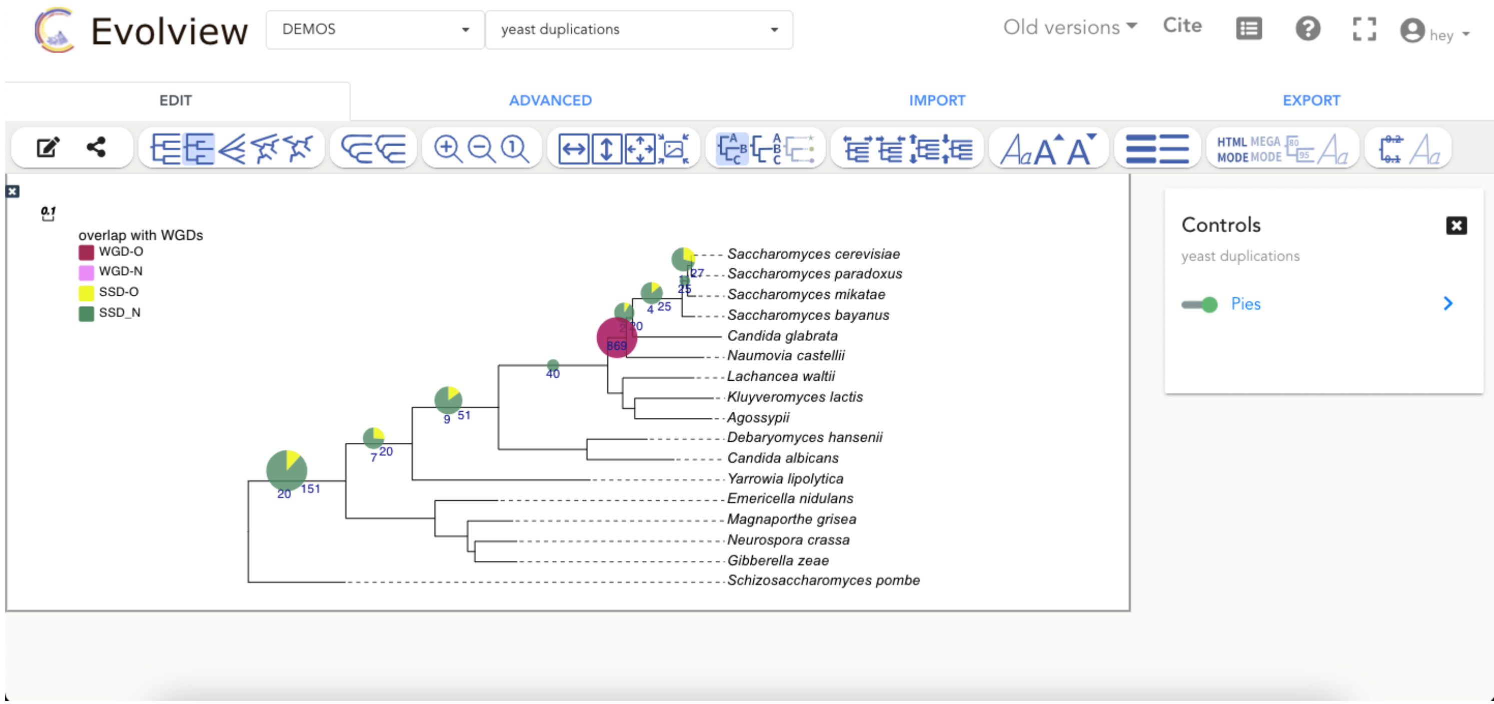
6 Visualization

Visualization of phylogenetic/phylogenomic trees

Web Portals

Evolview - https://www.evolgenius.info/help/

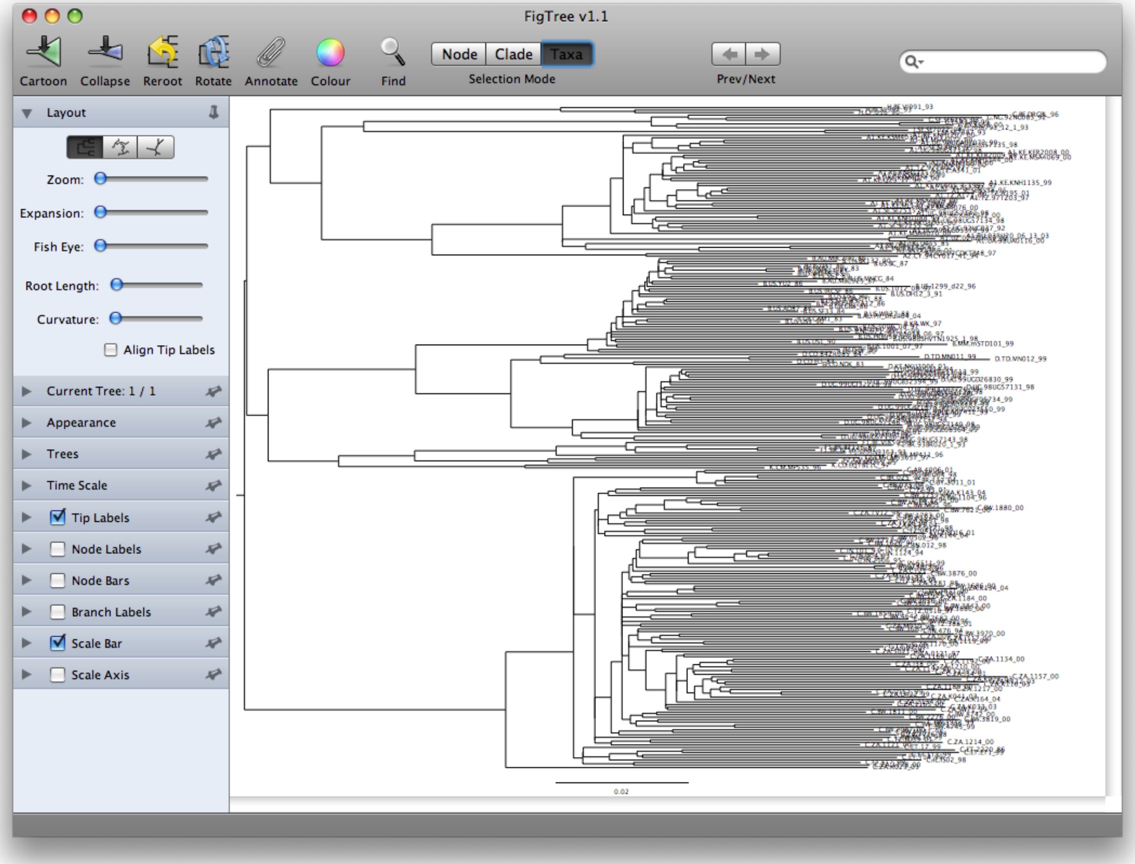
- Web-based
- User friendly (interactive interface)



Visualization of phylogenetic/phylogenomic trees

Software

FigTree
- graphical viewer of phylogenetic trees and a program for producing publication-ready figures.
-available for Mac, Windows and Linux
-interactive interface



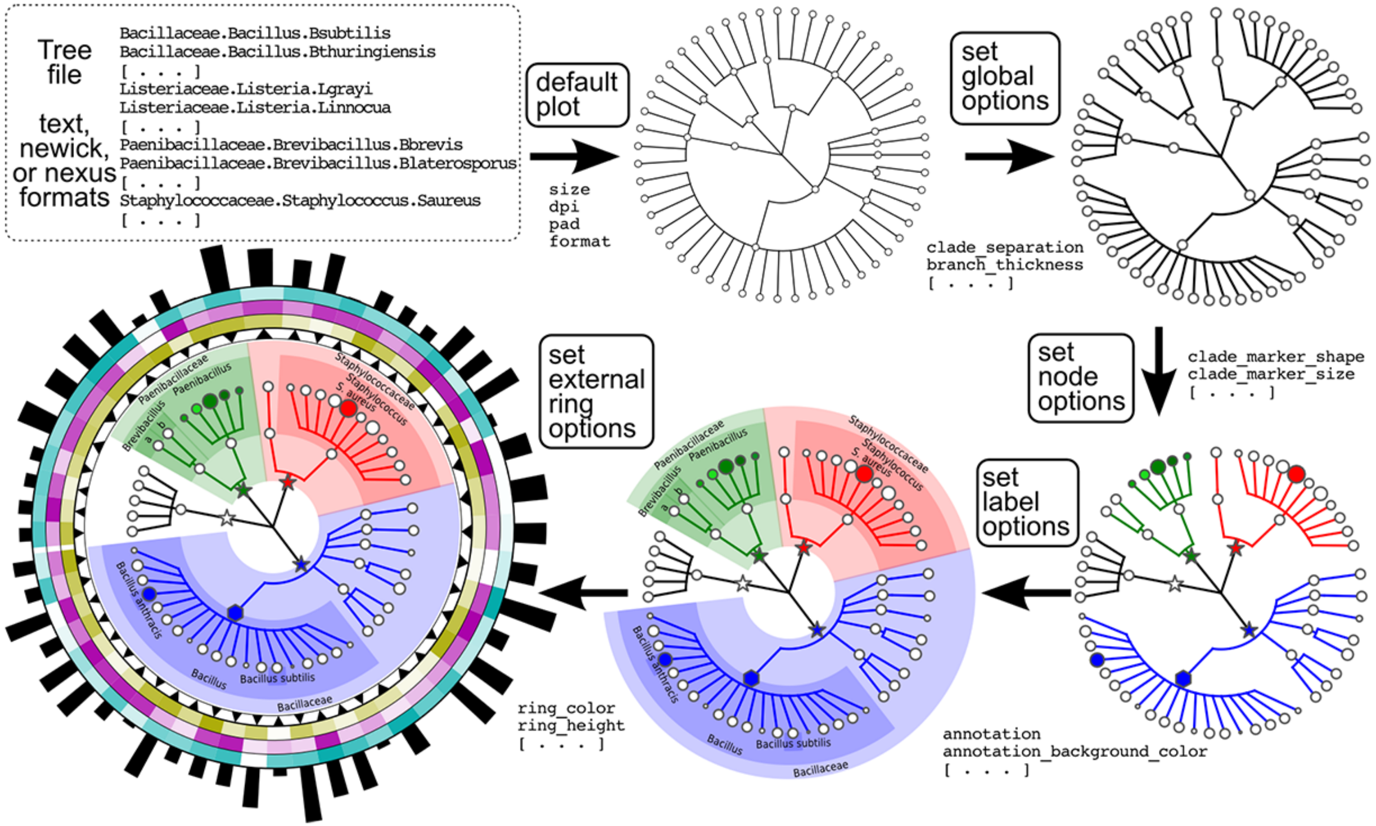
☐ Align Tip Labels
☑ Tip Labels
☐ Node Labels
☐ Node Bars
☐ Branch Labels
☑ Scale Bar
☐ Scale Axis

Visualization of phylogenetic/phylogenomic trees

Software

GraPhlAn
https://github.com/biobakery/graphlan
- is a software tool for producing high-quality circular representations of taxonomic and phylogenetic trees.

- publication-ready representations
- command line

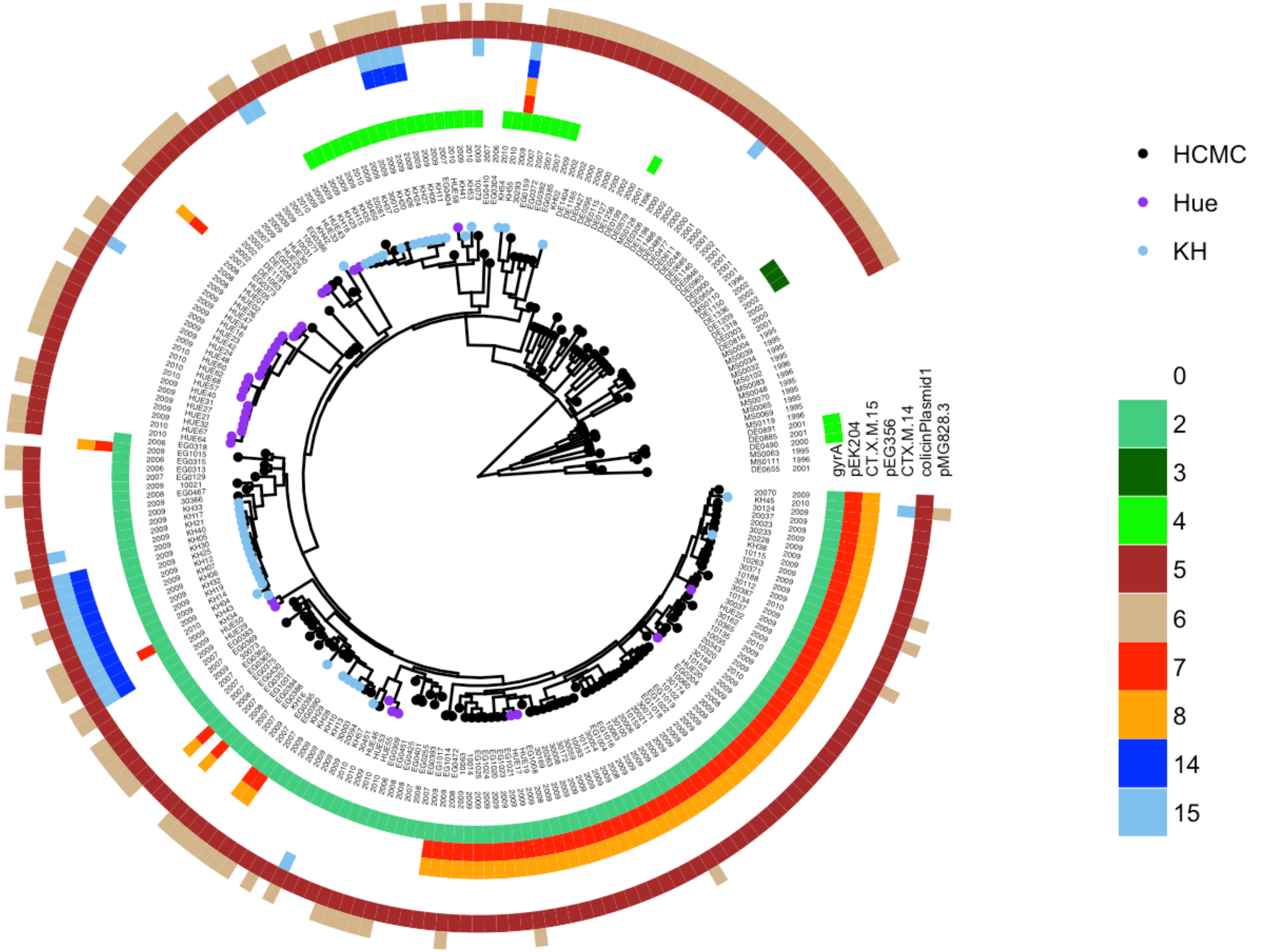


6 Visualization

Visualization of phylogenetic/phylogenomic trees

Software

R package **ggtree** - https://github.com/YuLab-SMU/ggtree
- an extension of the ‘ggplot2’ plotting system
- visualization and annotation of phylogenetic trees and other tree-like structures with their annotation data.

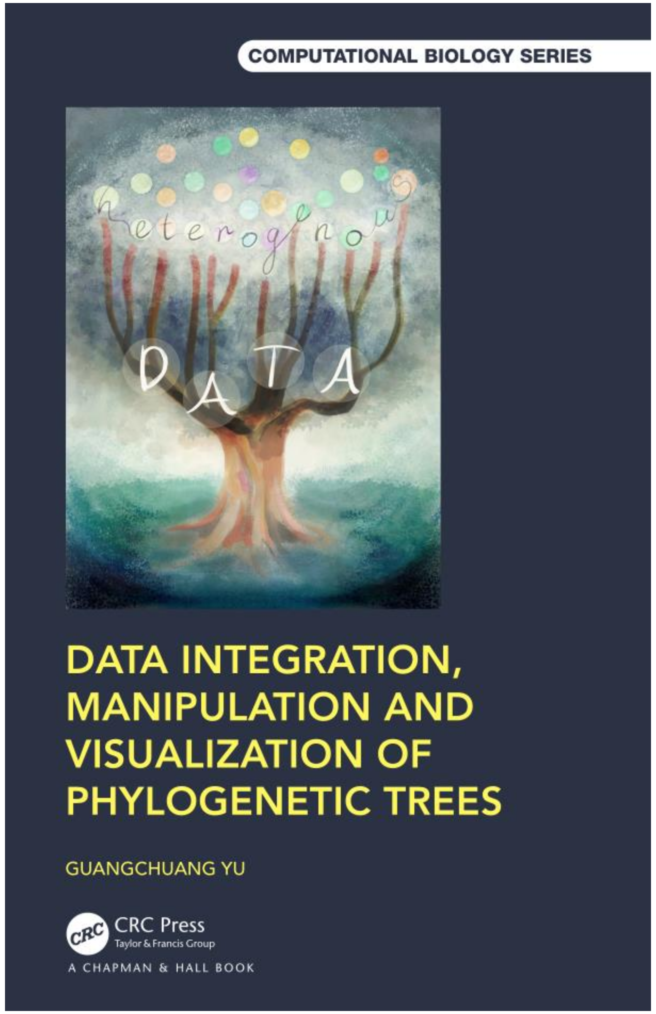


Visualization of phylogenetic/phylogenomic trees

Resources for visualization and annotation of phylogenetic trees using R

This book (https://yulab-smu.top/treedata-book/index.html) is meant as a guide for data integration, manipulation and visualization of phylogenetic trees using a suite of R packages, tidytree, treeio, ggtree and ggtreeExtra.

ggtree book – https://guangchuangyu.github.io/ggtree-book/short-introduction-to-r.html
Practical examples of annotation and visualization of phylogenetic trees using ggtree.


COMPUTATIONAL BIOLOGY SERIES
heterogenous
DATA
DATA INTEGRATION, MANIPULATION AND VISUALIZATION OF PHYLOGENETIC TREES
GUANGCHUANG YU
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

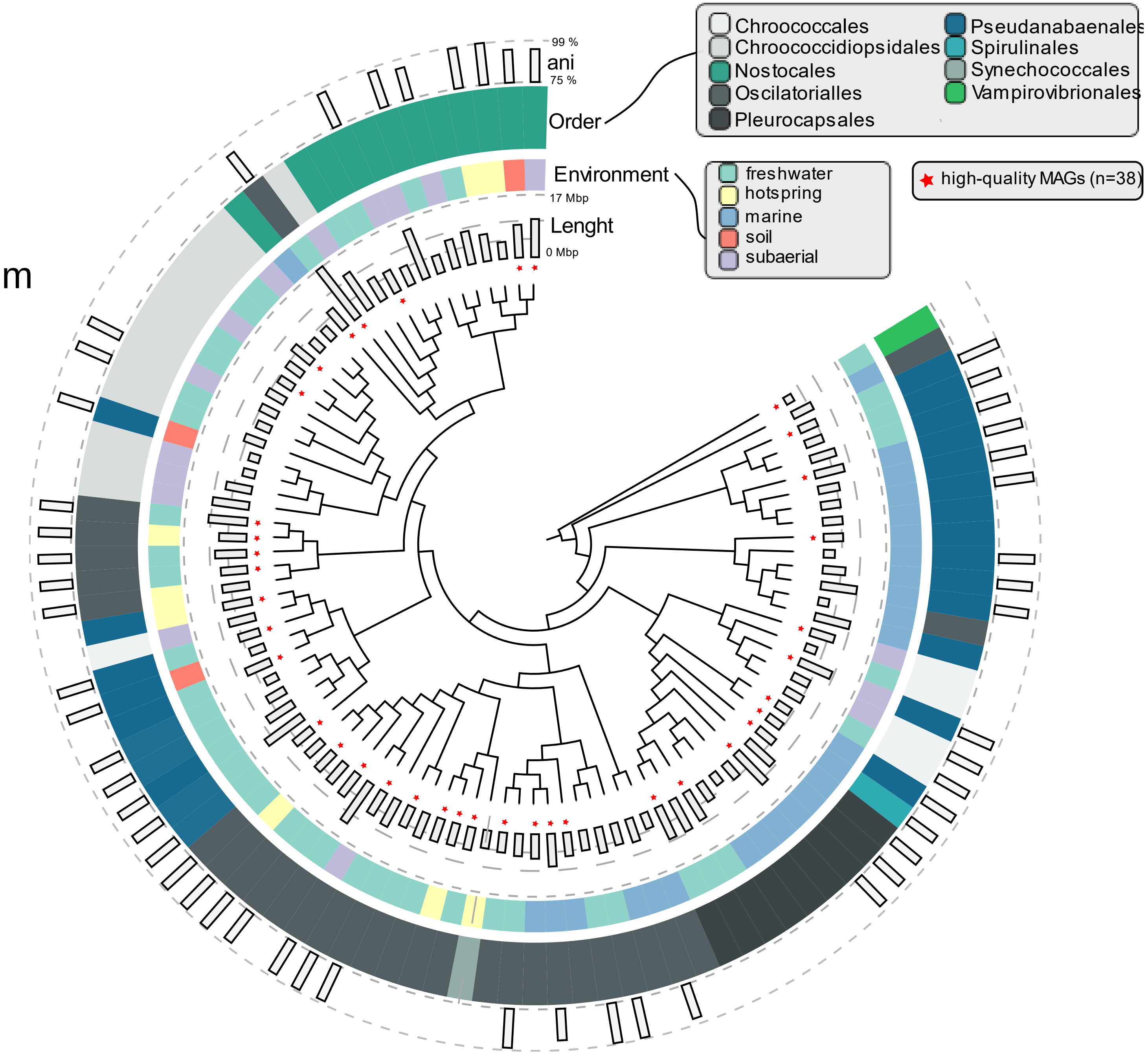
Chapter 4 Visualization and annotation of phylogenetic trees: *ggtree*

7 **Visualization**

Pratical example

Cyanobacteria MAGs - Samples collected from the environment

- Which Cyanonacteria orders are present?
- How novel these MAGs are?
- How are they distributed by environment type?
- What about genome lenght and completeness and contamination?



Pratical example

Workflow followed:

Tree reconstruction - **Phylophlan**

Tree Edition – **iTOL** - include previously obtained metadata (MAG/genome size, completeness/contamination of MAG, ANI to reference genomes) as datasets for iTOL

Import datasets to iTOL

Annotation template files – help page

https://itol.embl.de/help.cgi

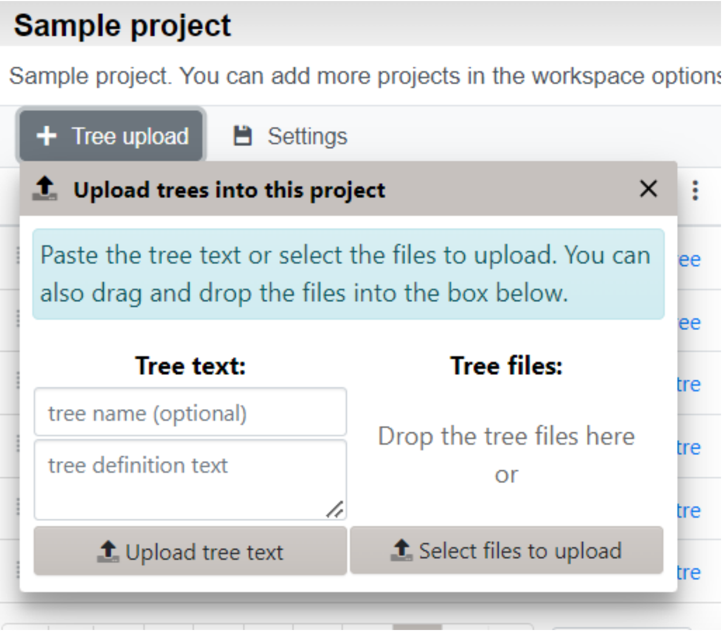
itol.toolkit R package

https://tongzhou2017.github.io/itol.toolkit/index.html

7 Visualization

Pratical example

Import tree into iTOL



Sample project. You can add more projects in the workspace options

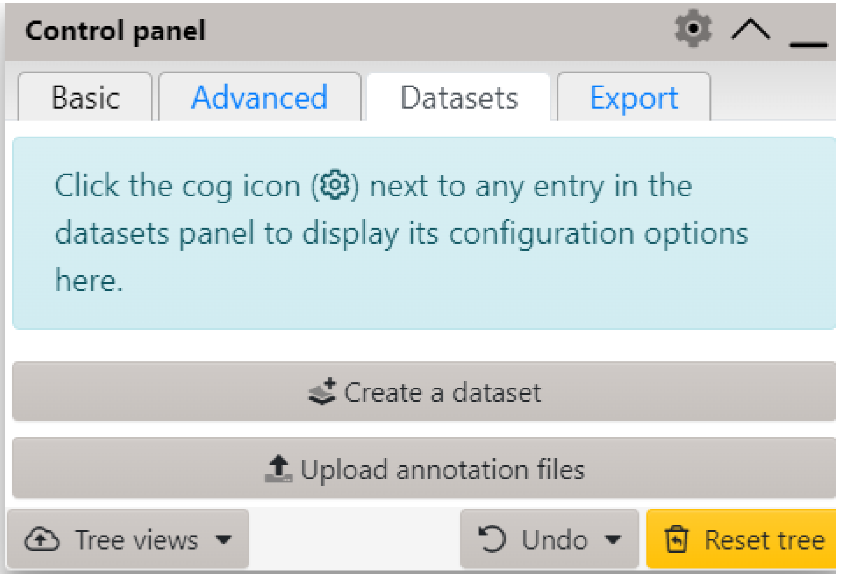
Paste the tree text or select the files to upload. You can also drag and drop the files into the box below.

tree name (optional)

tree definition text

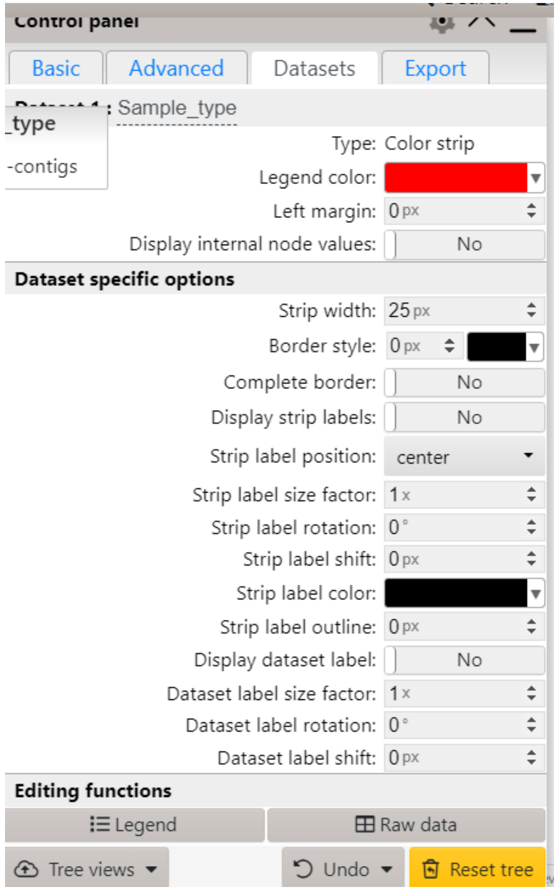
Drop the tree files here or

Import datasets into iTOL



Click the cog icon (⚙) next to any entry in the datasets panel to display its configuration options here.

Control panel for datasets



Dataset 1 : Sample_type

Type: Color strip

Legend color: 🟥

Left margin: 0 px

Display internal node values: No

Dataset specific options

Strip width: 25 px

Border style: 0 px ■

Complete border: No

Display strip labels: No

Strip label position: center

Strip label size factor: 1 x

Strip label rotation: 0 °

Strip label shift: 0 px

Strip label color: ■

Strip label outline: 0 px

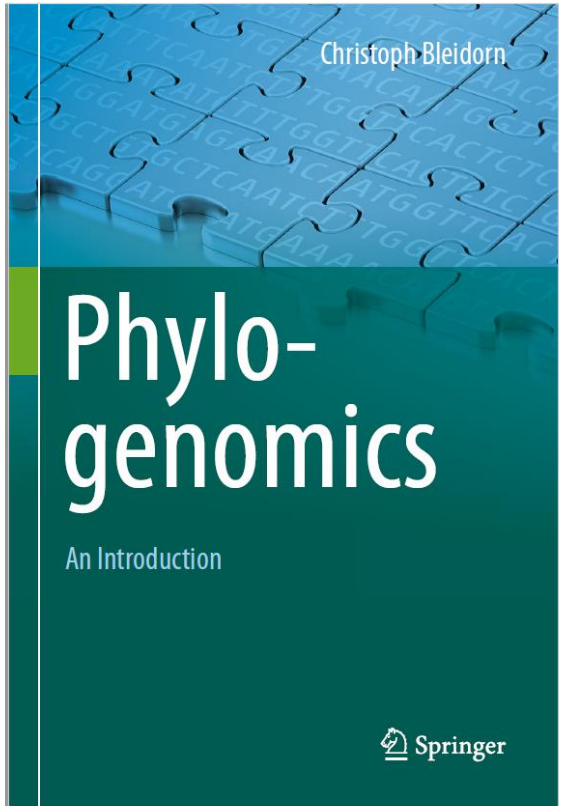
Display dataset label: No

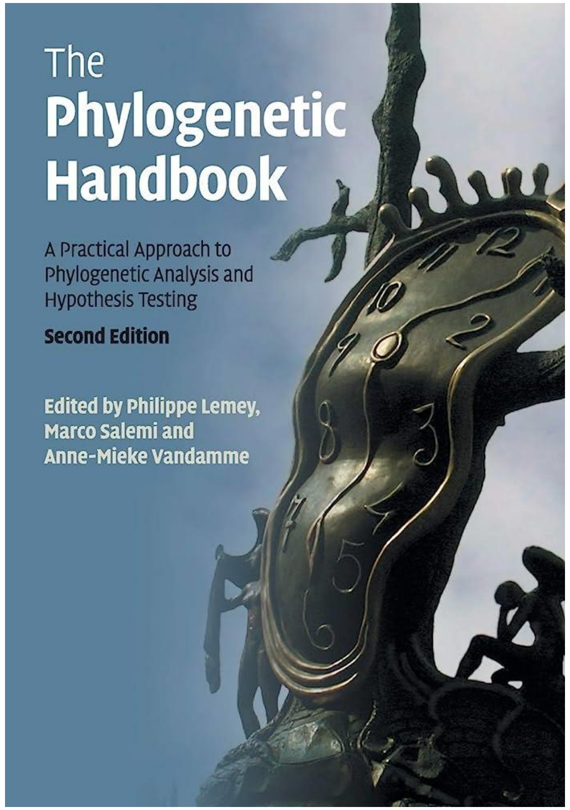
Dataset label size factor: 1 x

Dataset label rotation: 0 °

Dataset label shift: 0 px

Bibliography





Glossary

Clade Monophyletic group in a phylogenetic tree, thereby representing at least two terminals (which share a common ancestor).

Maximum likelihood (ML) Likelihood-based optimality criterion to find the best tree in a phylogenetic analysis through the computation of probabilities of character evolution given an explicit evolutionary model.

Maximum parsimony (MP) Optimality criterion which selects the phylogenetic tree(s) minimizing the total number of character state changes.

Monophyletic group Group containing a (hypothetical) ancestor and all of its descendants.

Orthology Pairs of homologous genes which have emerged through a speciation event are called orthologs.

Root Point of a topology where it is hypothetically connected to the remaining tree of life. Rooted trees are used to polarize character evolution.

Questions?

Thank you!

Any questions?

✉ adrianairego@gmail.com