University of Strathclyde Science

Evolutionary and Structural Analysis of Pathogen Proteins.

Final year UG project 2024-25
2024-10-21 (Week 5)

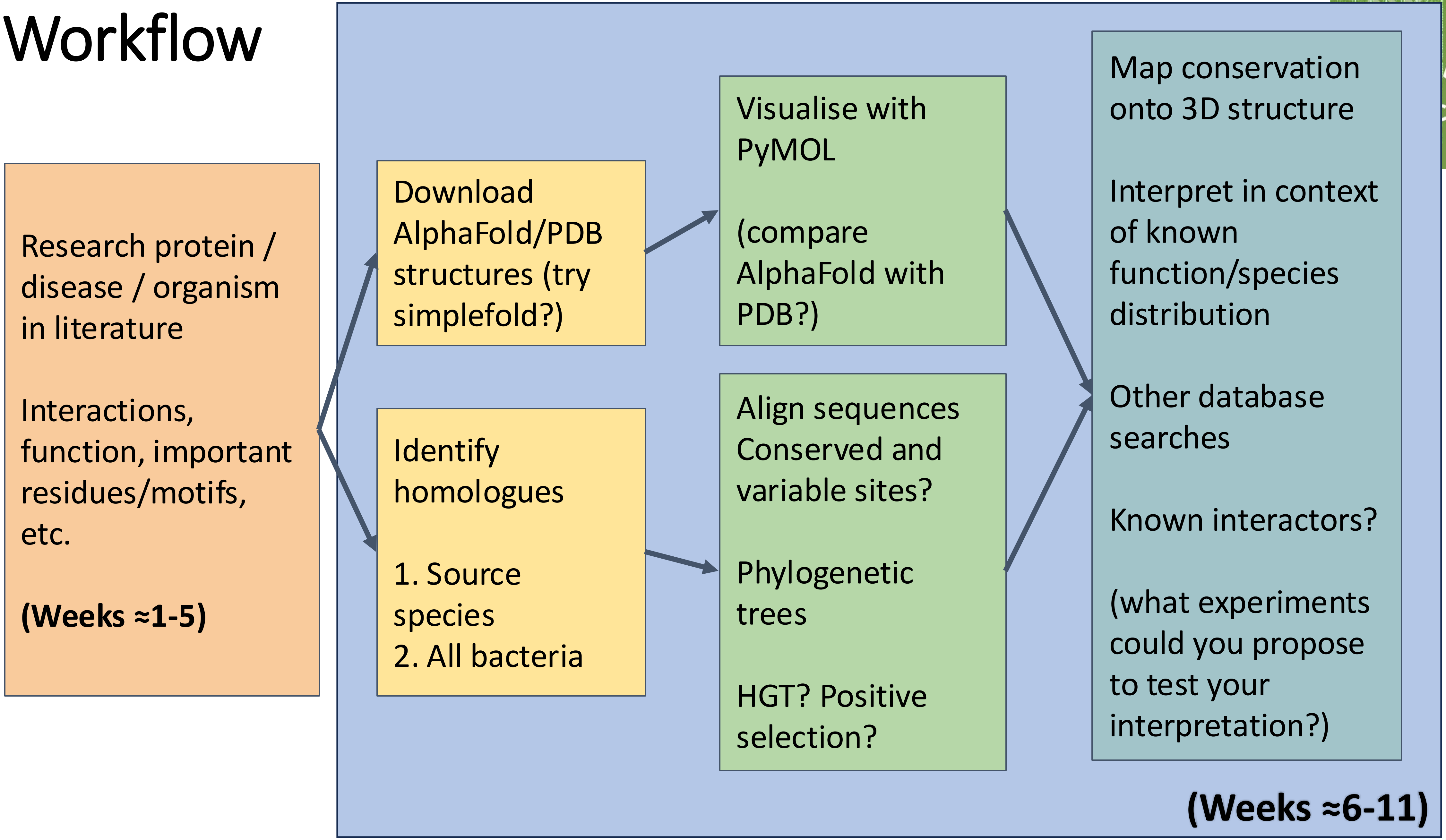
Candidate proteins – start points

Any changes needed?

Organism	Host	Gene/Protein	PHI accession	Student
Escherichia coli	*Homo sapiens*	*espY*	PHI:8647	LB
Shigella flexneri	*Homo sapiens*	*ipaJ*	PHI:9253	LT
Candida albicans	*Mus musculus*	*sap6*	PHI:10193	IM
Pseudomonas aeruginosa	*Homo sapiens*	*tplE*	PHI:6646	AE
Vibrio vulnificus	*Mus musculus*	*vvhA*	PHI:6877	JT

http://www.phi-base.org/

Workflow



Research protein / disease / organism in literature

Interactions, function, important residues/motifs, etc.

(Weeks ≈1-5)

Download AlphaFold/PDB structures (try simplefold?)

Identify homologues

1. Source species
2. All bacteria

Visualise with PyMOL

(compare AlphaFold with PDB?)

Align sequences Conserved and variable sites?

Phylogenetic trees

HGT? Positive selection?

Map conservation onto 3D structure

Interpret in context of known function/species distribution

Other database searches

Known interactors?

(what experiments could you propose to test your interpretation?)

(Weeks ≈6-11)

Your questions/comments

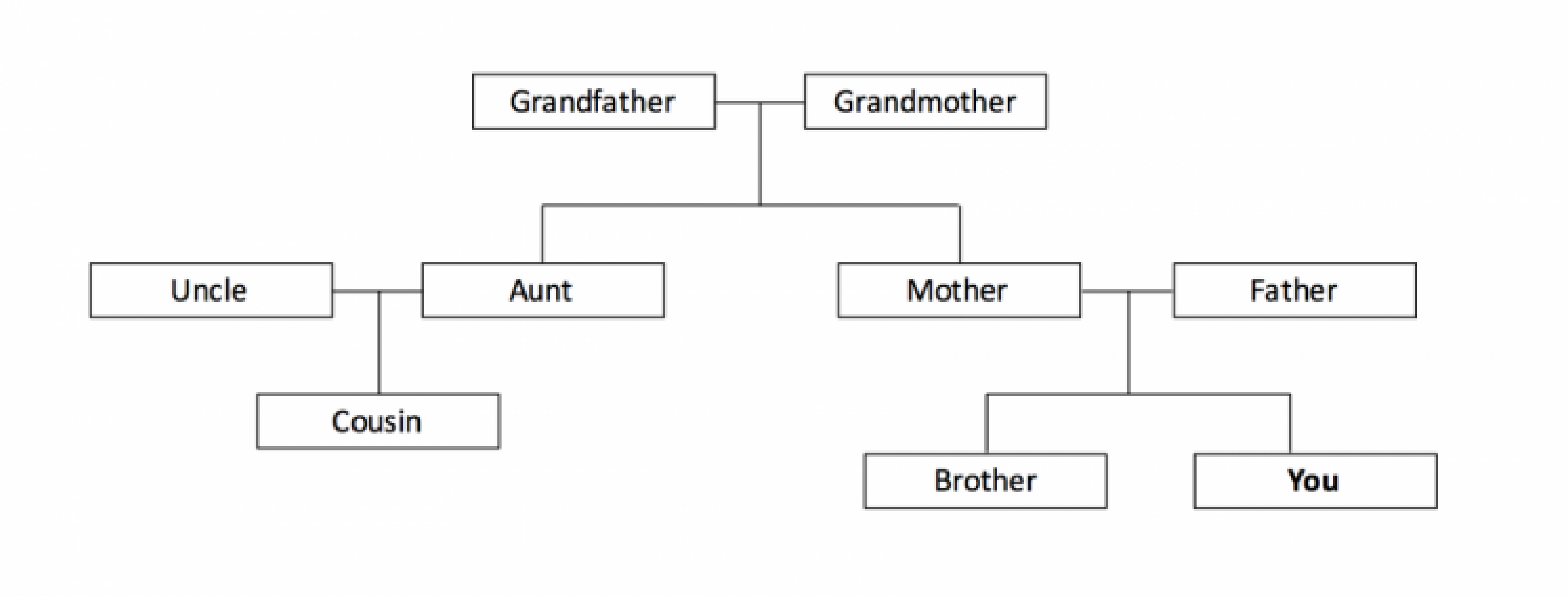
(What would you like to talk about?)

Thesis Introductions

A very quick introduction to building a phylogenetic tree

A Family Tree

- Family trees are not a good model of how bacteria (or species in general) evolve



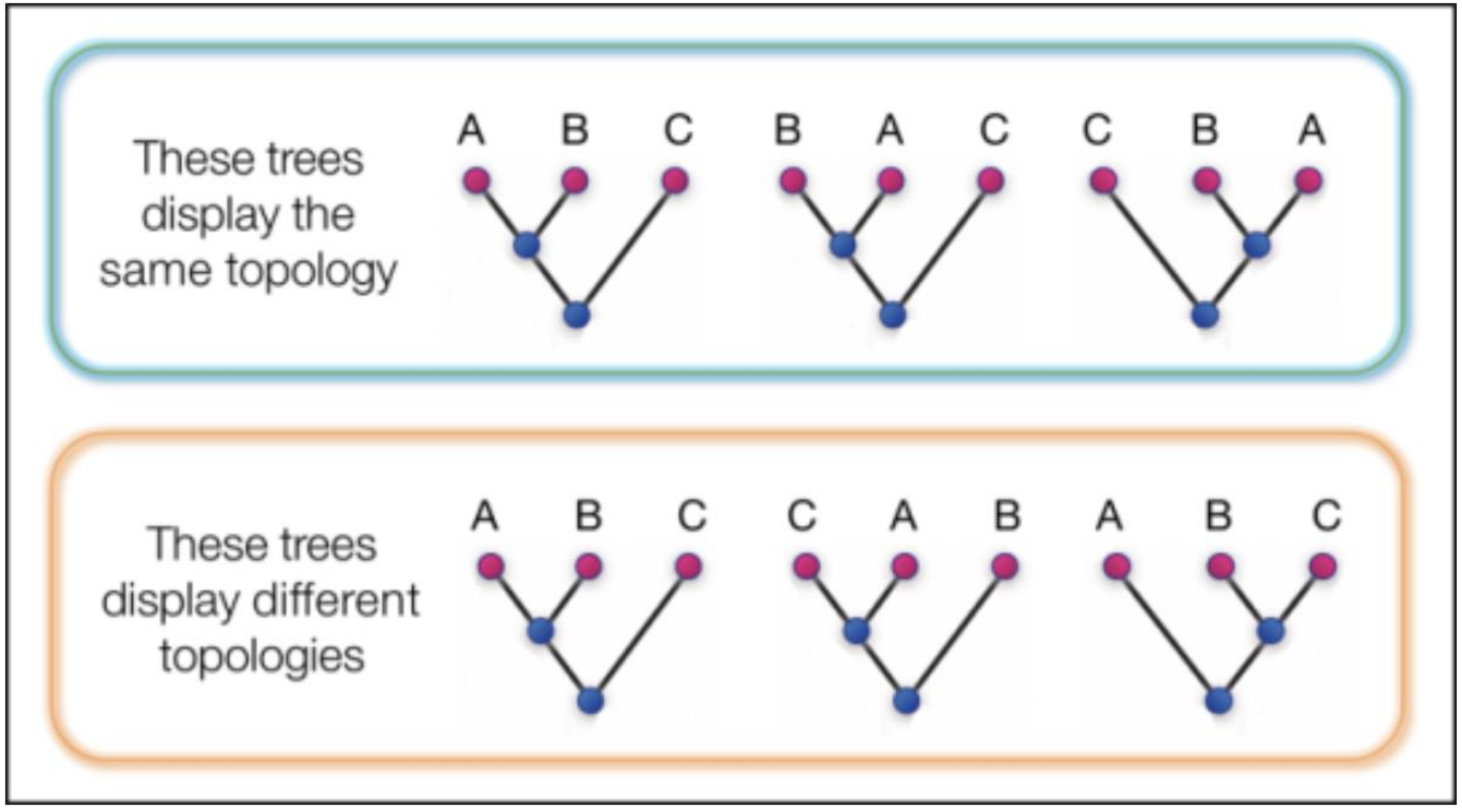
A Brief Introduction to Phylogenetics

- Online introductory course: https://www.ebi.ac.uk/training/online/courses/introduction-to-phylogenetics/
- Conor Meehan’s introductory course: https://conmeehan.github.io/PathogenDataCourse/IntroToPhylogenetics.html

- Phylogenetics is the reconstruction of evolutionary history from genetic/genomic data
  - **Input**: Aligned protein sequence data
  - **Output**: A tree estimating evolutionary relationships
- **Phylogenetic reconstruction is a mathematical activity**
- The biology in phylogenetics comes from three places:
  - Aligning the input sequence set correctly (evolutionary equivalence)
  - The model of substitution used (e.g. how likely is residue/base A to be replaced/substituted by residue/base B?)
  - The assumption of a bifurcating tree (this doesn’t apply to some methods, e.g. splitstree, but other assumptions do apply there)

Phylogenetic Trees (Topology)

- We assume that species evolve by a series of branching events
  - e.g. assume that species cannot interbreed so, when one species splits into two, it is an **irrevocable** branching event
- https://sipbs-compbiol.github.io/BM211-Workshop-5/


These trees display the same topology
A B C
B A C
C B A
These trees display different topologies
A B C
C A B
A B C

(A, B) are more closely related to each other than to C

Three different relationships:

(A, B) more closely related to each other than to C

(A, C) more closely related to each other than to B

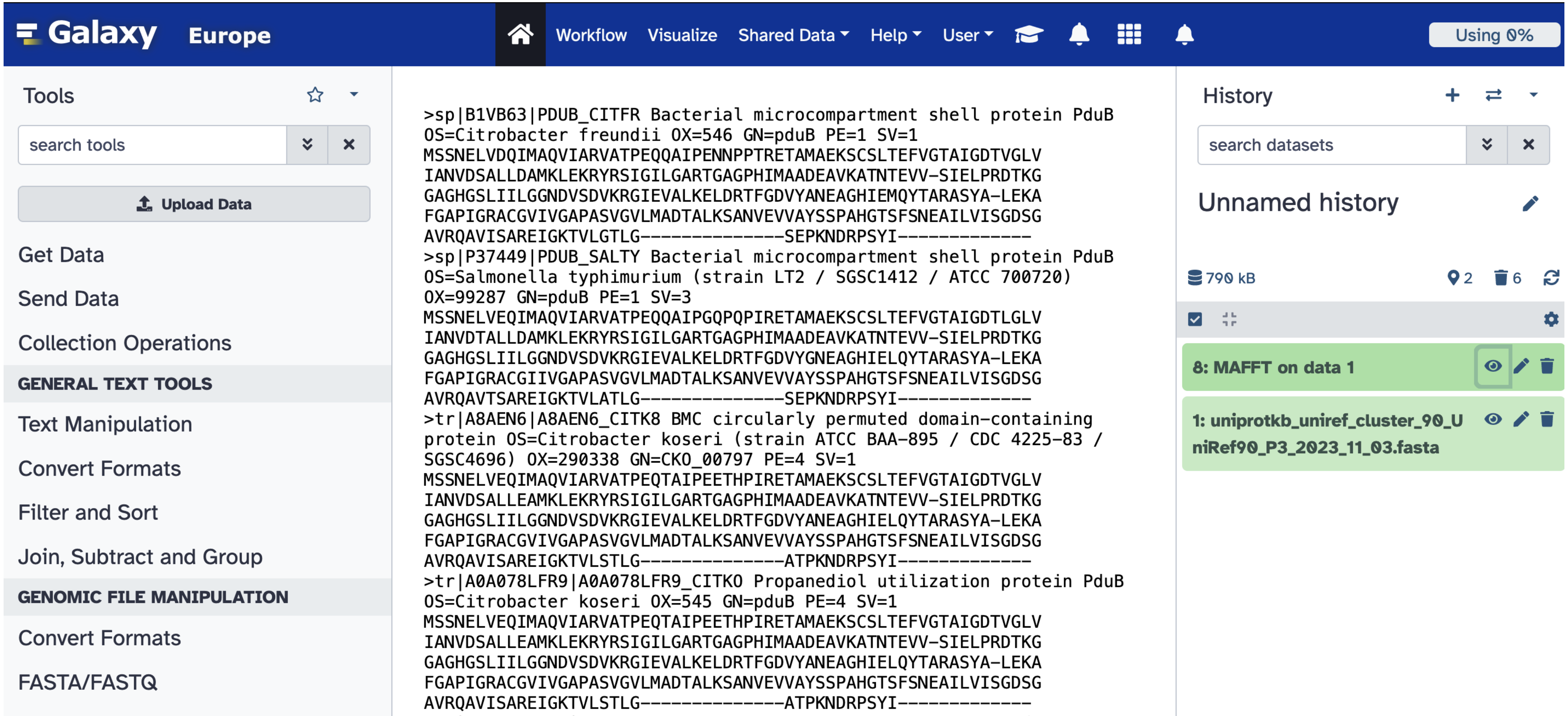
(B, C) more closely related to each other than to A

Input sequence alignment

- The goal is that each column in the alignment represents a single ***evolutionarily equivalent*** position – subject to similar selection pressures
  - We can then make inferences based on what changes are permitted at that position
  - Structural equivalence can (but does not always) imply functional equivalence
- Many gaps in a column mean information is missing and inference is less robust (they bias the tree)
  - Remove "gappy columns", e.g. https://vicfero.github.io/trimal/
- The larger amino acid alphabet means that alignments are more robust than nucleotide sequence alignments
- Codon degeneracy means that amino acid alignments can mask relevant evolutionary change, or mask "saturation"
  - Best practice for low/moderate divergence: align amino acid sequences and backtrace the coding sequence to nucleotide to make the tree
  - For highly divergent sequences, amino acid-based trees may be more robust

How To Make a (Simple) Tree in Galaxy

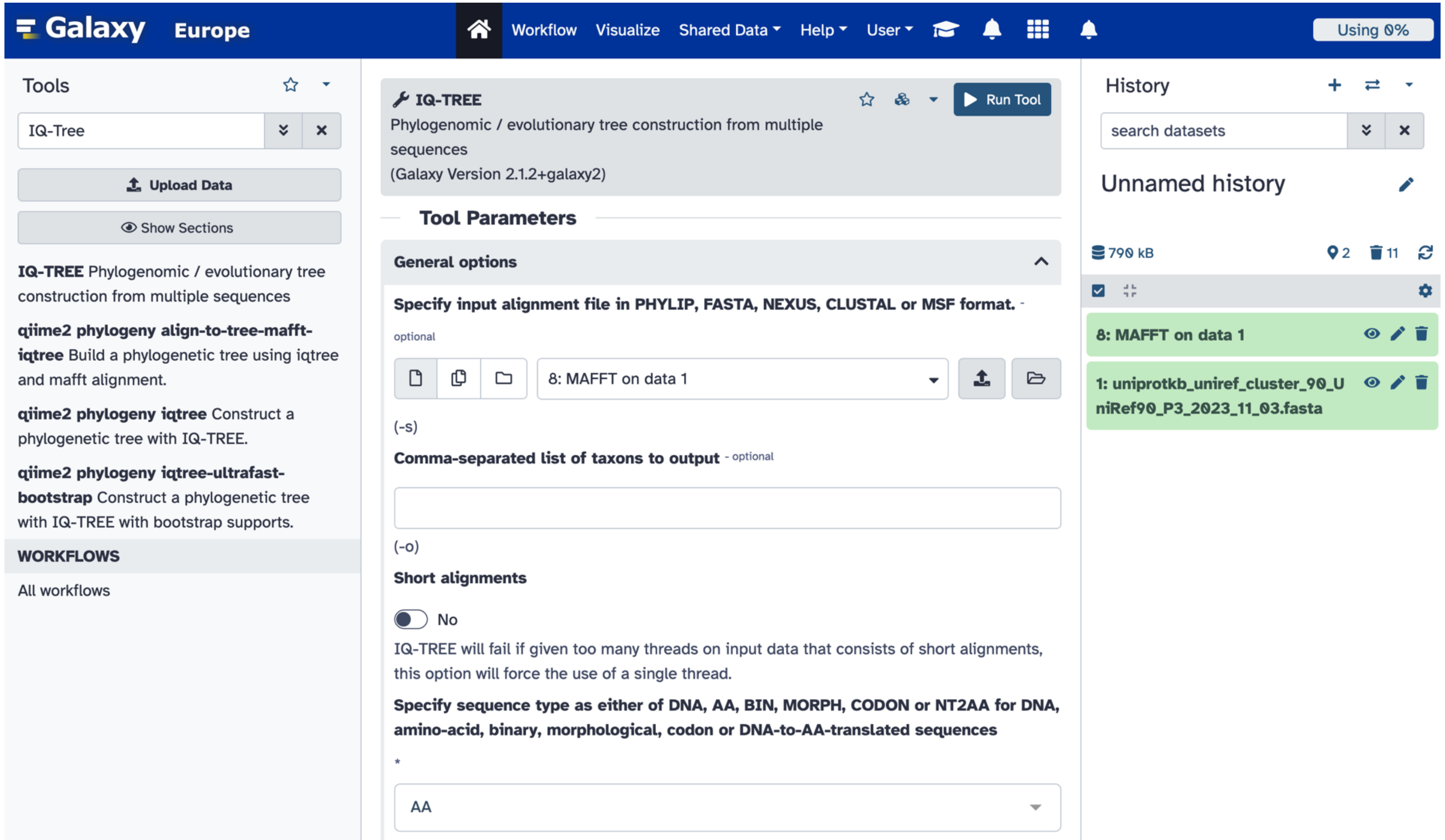
- Start with a FASTA (protein) sequence alignment



```
>sp|B1VB63|PDUB_CITFR Bacterial microcompartment shell protein PduB
OS=Citrobacter freundii OX=546 GN=pduB PE=1 SV=1
MSSNELVDQIMAQVIARVATPEQQAIPENNPPTRETAMAEKSCSLTEFVGTAIGDTVGLV
IANVDSALLDAMKLEKRYRSIGILGARTGAGPHIMAADEAVKATNTEVV-SIELPRDTKG
GAGHGSLIILGGNDVSDVKRGIEVALKELDRTFGDVYANEAGHIEMQYTARASYA-LEKA
FGAPIGRACGVIVGAPASVGVLMADTALKSANVEVVAYSSPAHGTSFSNEAILVISGDSG
AVRQAVISAREIGKTVLGTLG--------------SEPKNDRPSYI-------------
>sp|P37449|PDUB_SALTY Bacterial microcompartment shell protein PduB
OS=Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)
OX=99287 GN=pduB PE=1 SV=3
MSSNELVEQIMAQVIARVATPEQQAIPGQPQPIRETAMAEKSCSLTEFVGTAIGDTLGLV
IANVDTALLDAMKLEKRYRSIGILGARTGAGPHIMAADEAVKATNTEVV-SIELPRDTKG
GAGHGSLIILGGNDVSDVKRGIEVALKELDRTFGDVYGNEAGHIELQYTARASYA-LEKA
FGAPIGRACGIIVGAPASVGVLMADTALKSANVEVVAYSSPAHGTSFSNEAILVISGDSG
AVRQAVTSAREIGKTVLATLG--------------SEPKNDRPSYI-------------
>tr|A8AEN6|A8AEN6_CITK8 BMC circularly permuted domain-containing
protein OS=Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 /
SGSC4696) OX=290338 GN=CKO_00797 PE=4 SV=1
MSSNELVEQIMAQVIARVATPEQTAIPEETHPIRETAMAEKSCSLTEFVGTAIGDTVGLV
IANVDSALLEAMKLEKRYRSIGILGARTGAGPHIMAADEAVKATNTEVV-SIELPRDTKG
GAGHGSLIILGGNDVSDVKRGIEVALKELDRTFGDVYANEAGHIELQYTARASYA-LEKA
FGAPIGRACGVIVGAPASVGVLMADTALKSANVEVVAYSSPAHGTSFSNEAILVISGDSG
AVRQAVISAREIGKTVLSTLG--------------ATPKNDRPSYI-------------
>tr|A0A078LFR9|A0A078LFR9_CITKO Propanediol utilization protein PduB
OS=Citrobacter koseri OX=545 GN=pduB PE=4 SV=1
MSSNELVEQIMAQVIARVATPEQTAIPEETHPIRETAMAEKSCSLTEFVGTAIGDTVGLV
IANVDSALLEAMKLEKRYRSIGILGARTGAGPHIMAADEAVKATNTEVV-SIELPRDTKG
GAGHGSLIILGGNDVSDVKRGIEVALKELDRTFGDVYANEAGHIELQYTARASYA-LEKA
FGAPIGRACGVIVGAPASVGVLMADTALKSANVEVVAYSSPAHGTSFSNEAILVISGDSG
AVRQAVISAREIGKTVLSTLG--------------ATPKNDRPSYI-------------
```

How To Make a (Simple) Tree in Galaxy

- Use IQ-Tree to generate a phylogenetic tree: **SPECIFY AA sequence type!**



IQ-Tree

IQ-TREE Phylogenomic / evolutionary tree construction from multiple sequences

qiime2 phylogeny align-to-tree-mafft-iqtree Build a phylogenetic tree using iqtree and mafft alignment.

qiime2 phylogeny iqtree Construct a phylogenetic tree with IQ-TREE.

qiime2 phylogeny iqtree-ultrafast-bootstrap Construct a phylogenetic tree with IQ-TREE with bootstrap supports.

IQ-TREE
Phylogenomic / evolutionary tree construction from multiple sequences
(Galaxy Version 2.1.2+galaxy2)

Tool Parameters

General options

Specify input alignment file in PHYLIP, FASTA, NEXUS, CLUSTAL or MSF format. - optional

8: MAFFT on data 1

(-s)

Comma-separated list of taxons to output - optional

(-o)

Short alignments

No

IQ-TREE will fail if given too many threads on input data that consists of short alignments, this option will force the use of a single thread.

Specify sequence type as either of DNA, AA, BIN, MORPH, CODON or NT2AA for DNA, amino-acid, binary, morphological, codon or DNA-to-AA-translated sequences

*

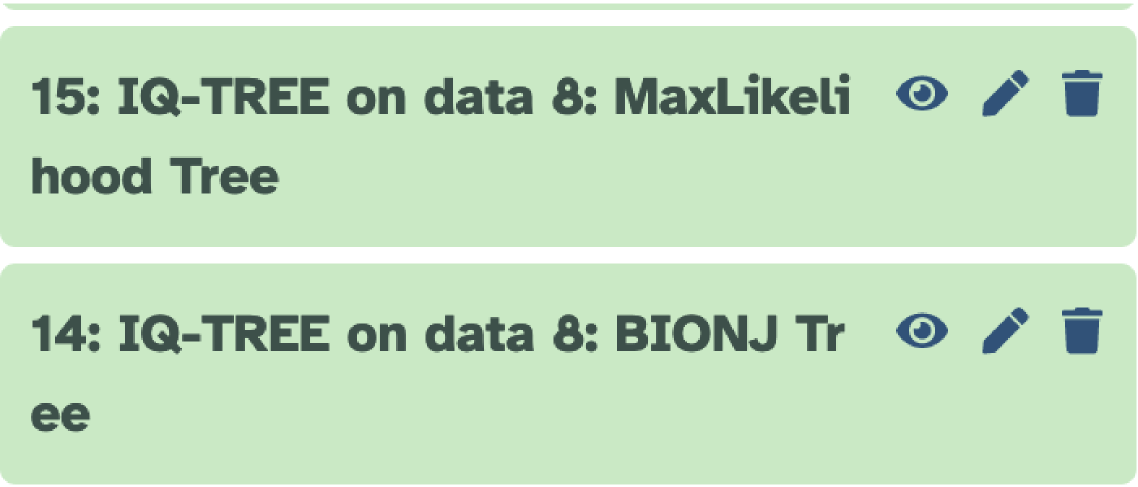
AA

8: MAFFT on data 1

1: uniprotkb_uniref_cluster_90_UniRef90_P3_2023_11_03.fasta

How To Make a (Simple) Tree in Galaxy

- Produces more than one tree
  - BioNJ: Neighbour-Joining (tree-building algorithm)
  - Maximum Likelihood (fitting a tree to the data)
- **Trees produced by different approaches (or with different parameters/inputs) are often different**
  - **This is not bad!** It’s something to note in the discussion – is the tree ***robust***?
  - **Justify choices** (as much as possible) in the thesis
  - There are many parameters/options to choose – it’s fine to use defaults, but state clearly that you did so in your Methods.

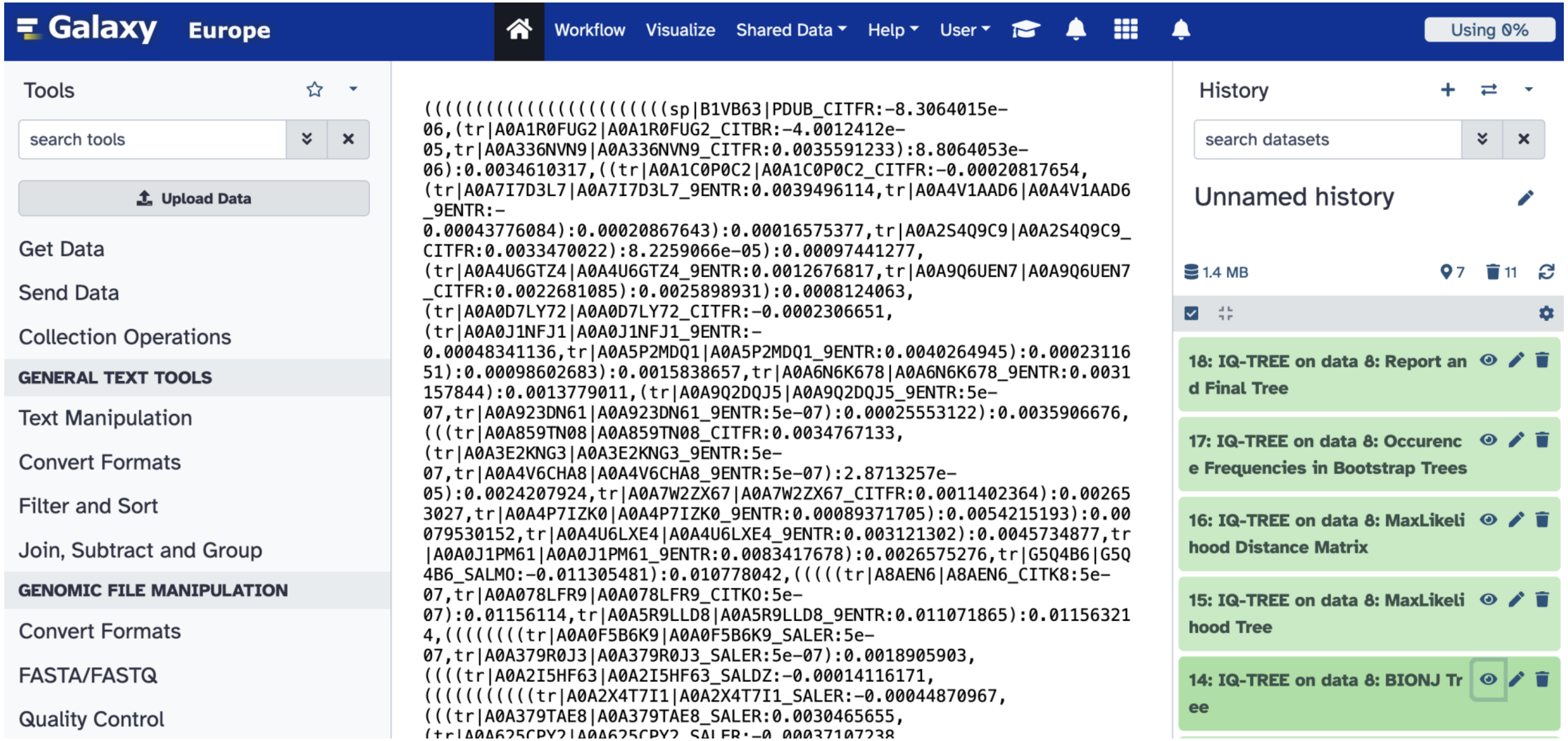


15: IQ-TREE on data 8: MaxLikelihood Tree

14: IQ-TREE on data 8: BIONJ Tree

How To Make a (Simple) Tree in Galaxy

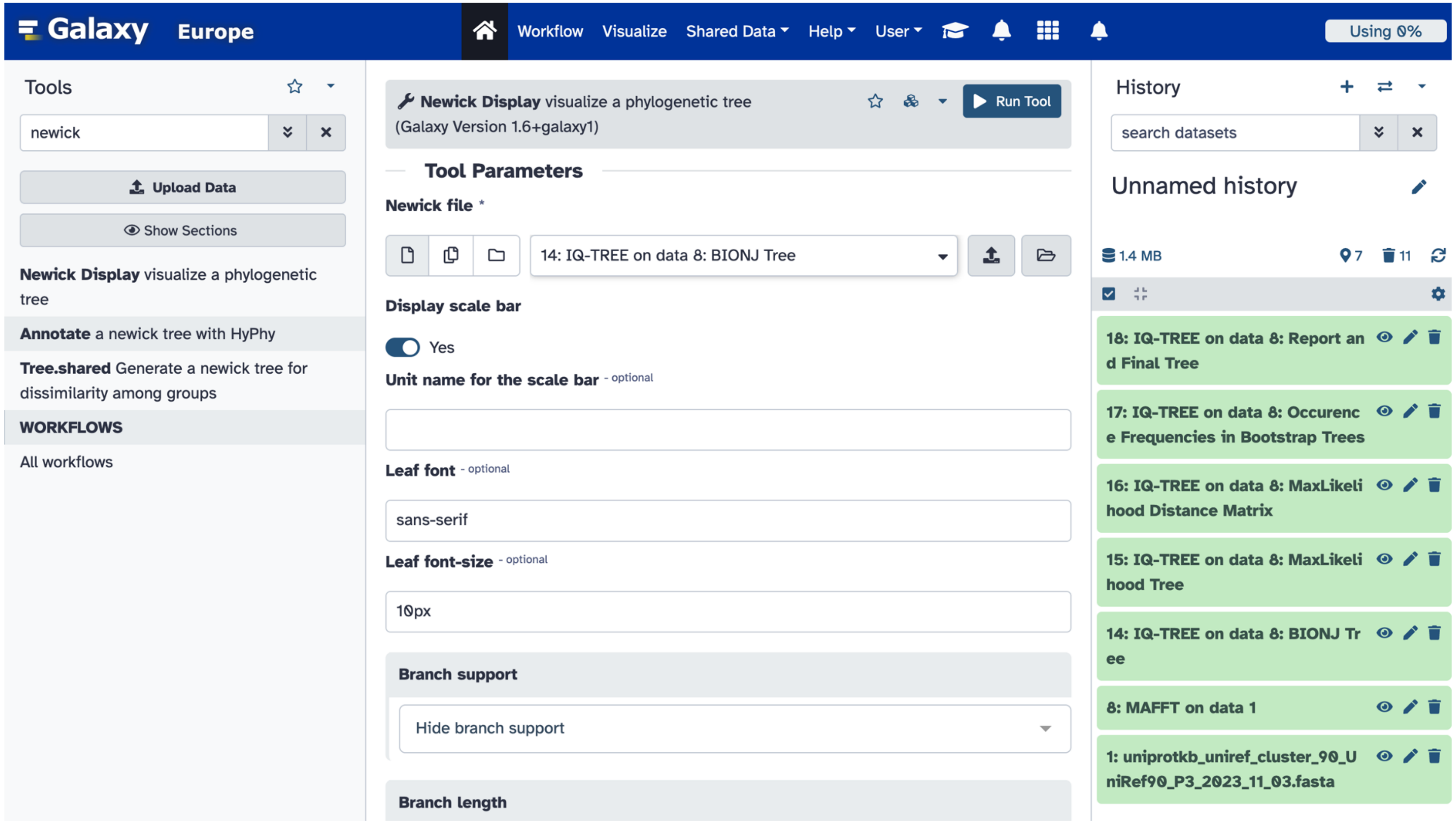
- Raw tree data (Newick format) looks cryptic – you don’t have to read this data yourself: it’s for computers



```
(((((((((((((((((((((((sp|B1VB63|PDUB_CITFR:-8.3064015e-
06,(tr|A0A1R0FUG2|A0A1R0FUG2_CITBR:-4.0012412e-
05,tr|A0A336NVN9|A0A336NVN9_CITFR:0.0035591233):8.8064053e-
06):0.0034610317,((tr|A0A1C0P0C2|A0A1C0P0C2_CITFR:-0.00020817654,
(tr|A0A7I7D3L7|A0A7I7D3L7_9ENTR:0.0039496114,tr|A0A4V1AAD6|A0A4V1AAD6
_9ENTR:-
0.00043776084):0.00020867643):0.00016575377,tr|A0A2S4Q9C9|A0A2S4Q9C9_
CITFR:0.0033470022):8.2259066e-05):0.00097441277,
(tr|A0A4U6GTZ4|A0A4U6GTZ4_9ENTR:0.0012676817,tr|A0A9Q6UEN7|A0A9Q6UEN7
_CITFR:0.0022681085):0.0025898931):0.0008124063,
(tr|A0A0D7LY72|A0A0D7LY72_CITFR:-0.0002306651,
(tr|A0A0J1NFJ1|A0A0J1NFJ1_9ENTR:-
0.00048341136,tr|A0A5P2MDQ1|A0A5P2MDQ1_9ENTR:0.0040264945):0.00023116
51):0.00098602683):0.0015838657,tr|A0A6N6K678|A0A6N6K678_9ENTR:0.0031
157844):0.0013779011,(tr|A0A9Q2DQJ5|A0A9Q2DQJ5_9ENTR:5e-
07,tr|A0A923DN61|A0A923DN61_9ENTR:5e-07):0.00025553122):0.0035906676,
(((tr|A0A859TN08|A0A859TN08_CITFR:0.0034767133,
(tr|A0A3E2KNG3|A0A3E2KNG3_9ENTR:5e-
07,tr|A0A4V6CHA8|A0A4V6CHA8_9ENTR:5e-07):2.8713257e-
05):0.0024207924,tr|A0A7W2ZX67|A0A7W2ZX67_CITFR:0.0011402364):0.00265
3027,tr|A0A4P7IZK0|A0A4P7IZK0_9ENTR:0.00089371705):0.0054215193):0.00
079530152,tr|A0A4U6LXE4|A0A4U6LXE4_9ENTR:0.003121302):0.0045734877,tr
|A0A0J1PM61|A0A0J1PM61_9ENTR:0.0083417678):0.0026575276,tr|G5Q4B6|G5Q
4B6_SALMO:-0.011305481):0.010778042,(((((tr|A8AEN6|A8AEN6_CITK8:5e-
07,tr|A0A078LFR9|A0A078LFR9_CITKO:5e-
07):0.01156114,tr|A0A5R9LLD8|A0A5R9LLD8_9ENTR:0.011071865):0.01156321
4,((((((((tr|A0A0F5B6K9|A0A0F5B6K9_SALER:5e-
07,tr|A0A379R0J3|A0A379R0J3_SALER:5e-07):0.0018905903,
((((tr|A0A2I5HF63|A0A2I5HF63_SALDZ:-0.00014116171,
(((((((((((tr|A0A2X4T7I1|A0A2X4T7I1_SALER:-0.00044870967,
(((tr|A0A379TAE8|A0A379TAE8_SALER:0.0030465655,
(tr|A0A625CPY2|A0A625CPY2_SALER:-0.00037107238
```

How To Make a (Simple) Tree in Galaxy

- You can visualize your tree(s) directly in Galaxy (e.g. with Newick Display)



Tools search: newick

Newick Display visualize a phylogenetic tree

Annotate a newick tree with HyPhy

Tree.shared Generate a newick tree for dissimilarity among groups

Newick Display visualize a phylogenetic tree (Galaxy Version 1.6+galaxy1)

Tool Parameters

Newick file *

14: IQ-TREE on data 8: BIONJ Tree

Display scale bar

Yes

Unit name for the scale bar - optional

Leaf font - optional

sans-serif

Leaf font-size - optional

10px

Branch support

Hide branch support

Branch length

History: search datasets

Unnamed history

18: IQ-TREE on data 8: Report and Final Tree

17: IQ-TREE on data 8: Occurence Frequencies in Bootstrap Trees

16: IQ-TREE on data 8: MaxLikelihood Distance Matrix

15: IQ-TREE on data 8: MaxLikelihood Tree

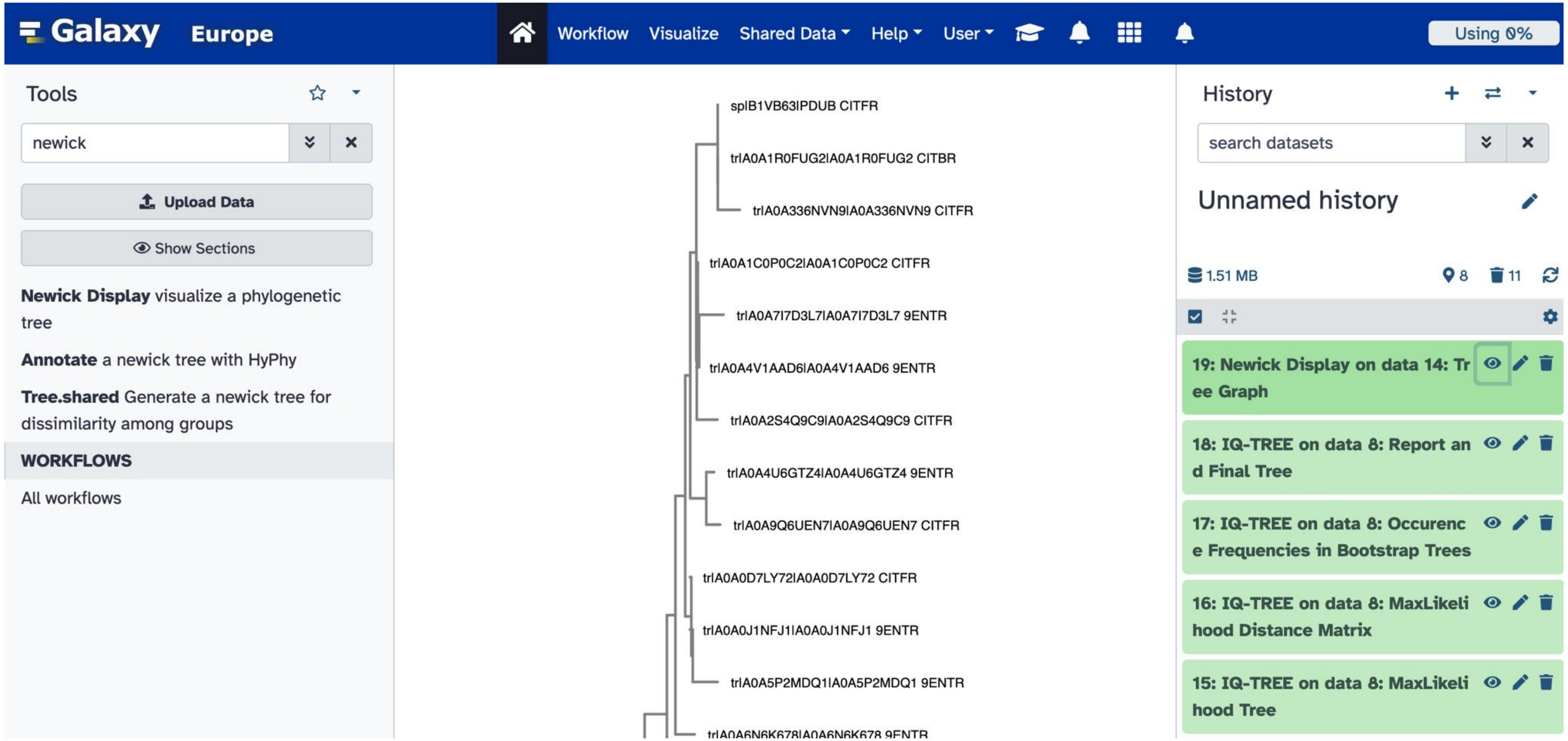
14: IQ-TREE on data 8: BIONJ Tree

8: MAFFT on data 1

1: uniprotkb_uniref_cluster_90_UniRef90_P3_2023_11_03.fasta

How To Make a (Simple) Tree in Galaxy

- You can visualize your tree(s) directly in Galaxy (e.g. with Newick Display)
  - TBH it doesn't do a great job


Galaxy Europe
Workflow
Visualize
Shared Data
Help
User
Using 0%
Tools
newick
Upload Data
Show Sections
WORKFLOWS
All workflows
History
search datasets
Unnamed history
1.51 MB
8
11
splB1VB63lPDUB CITFR
trlA0A1R0FUG2lA0A1R0FUG2 CITBR
trlA0A336NVN9lA0A336NVN9 CITFR
trlA0A1C0P0C2lA0A1C0P0C2 CITFR
trlA0A7I7D3L7lA0A7I7D3L7 9ENTR
trlA0A4V1AAD6lA0A4V1AAD6 9ENTR
trlA0A2S4Q9C9lA0A2S4Q9C9 CITFR
trlA0A4U6GTZ4lA0A4U6GTZ4 9ENTR
trlA0A9Q6UEN7lA0A9Q6UEN7 CITFR
trlA0A0D7LY72lA0A0D7LY72 CITFR
trlA0A0J1NFJ1lA0A0J1NFJ1 9ENTR
trlA0A5P2MDQ1lA0A5P2MDQ1 9ENTR
trlA0A6N6K678lA0A6N6K678 9ENTR

Newick Display visualize a phylogenetic tree

Annotate a newick tree with HyPhy

Tree.shared Generate a newick tree for dissimilarity among groups

19: Newick Display on data 14: Tree Graph

18: IQ-TREE on data 8: Report and Final Tree

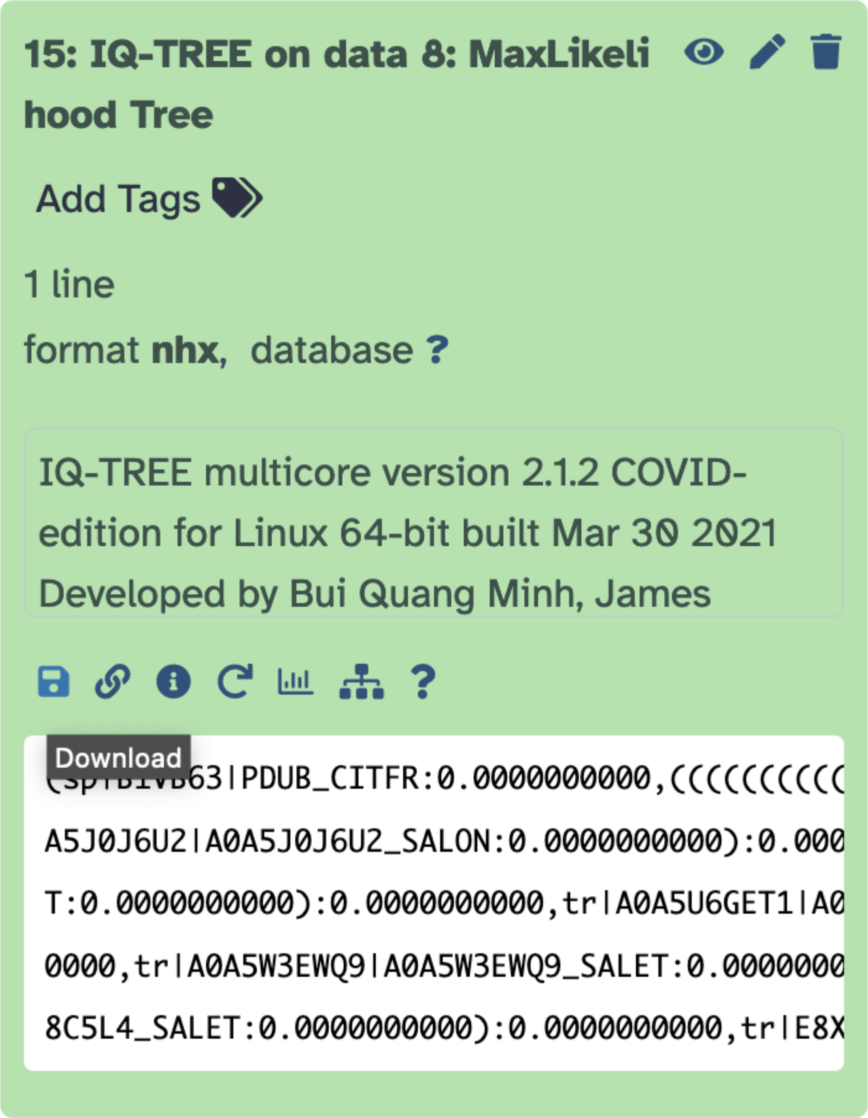
17: IQ-TREE on data 8: Occurence Frequencies in Bootstrap Trees

16: IQ-TREE on data 8: MaxLikelihood Distance Matrix

15: IQ-TREE on data 8: MaxLikelihood Tree

How To Make a (Simple) Tree in Galaxy

- Download your (Newick) tree file
  - This will let you use better visualization tools
- FigTree
  - http://tree.bio.ed.ac.uk/software/Figtree/
- DendroScope
  - https://uni-tuebingen.de/en/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/algorithms-in-bioinformatics/software/dendroscope/
- iTOL (Interactive Tree of Life)
  - https://itol.embl.de/



15: IQ-TREE on data 8: MaxLikelihood Tree

Add Tags

1 line

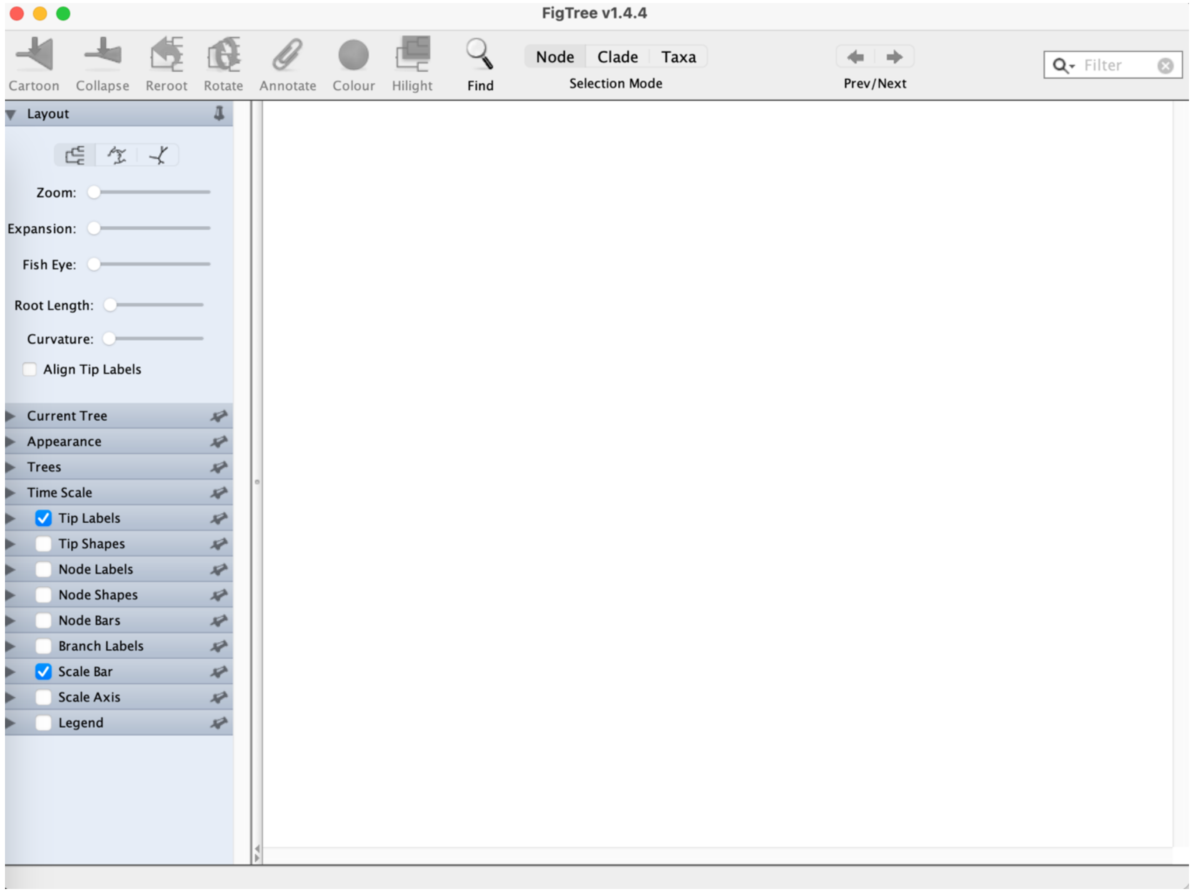
format **nhx**, database **?**

IQ-TREE multicore version 2.1.2 COVID-edition for Linux 64-bit built Mar 30 2021
Developed by Bui Quang Minh, James

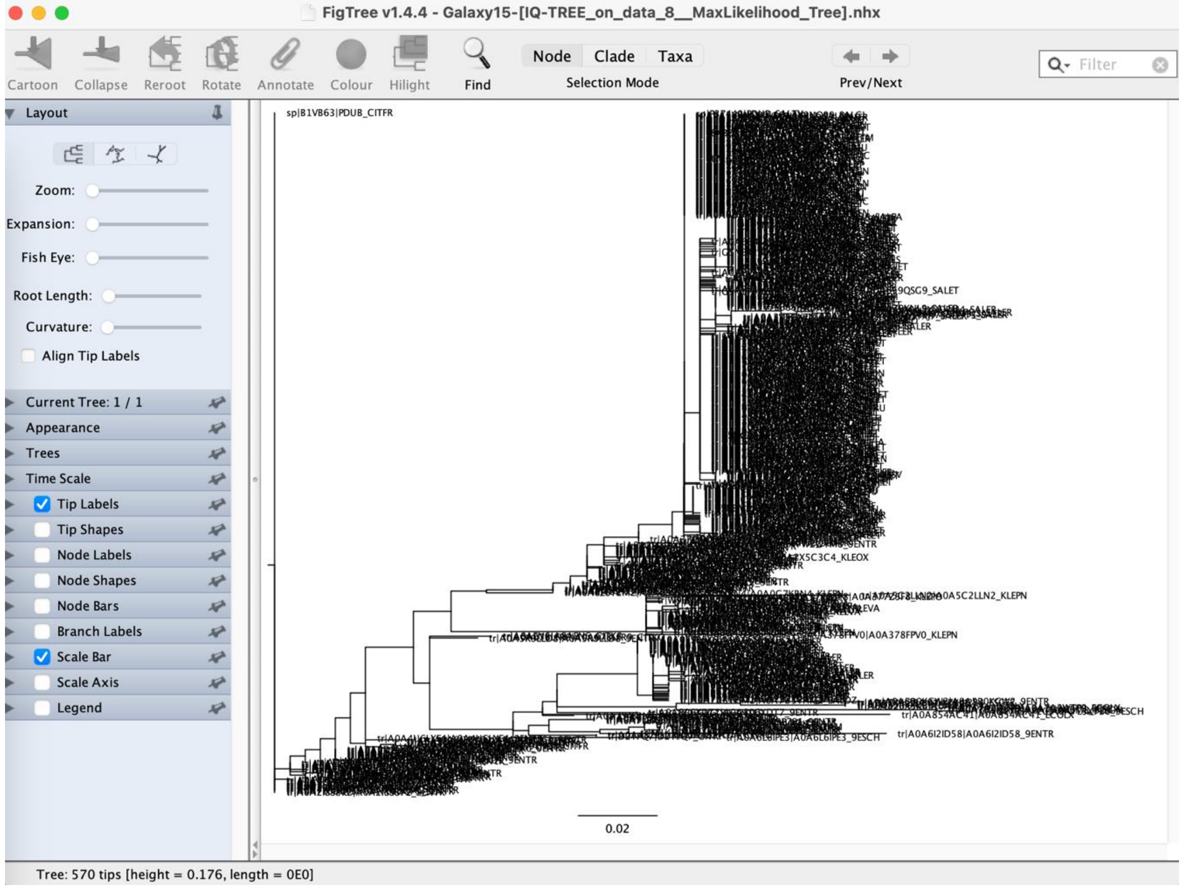
```
(sp|B1VB63|PDUB_CITFR:0.0000000000,(((((((((
A5J0J6U2|A0A5J0J6U2_SALON:0.0000000000):0.000
T:0.0000000000):0.0000000000,tr|A0A5U6GET1|A0
0000,tr|A0A5W3EWQ9|A0A5W3EWQ9_SALET:0.0000000
8C5L4_SALET:0.0000000000):0.0000000000,tr|E8X
```

FigTree

- Open the downloaded Newick (.nhx) file



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

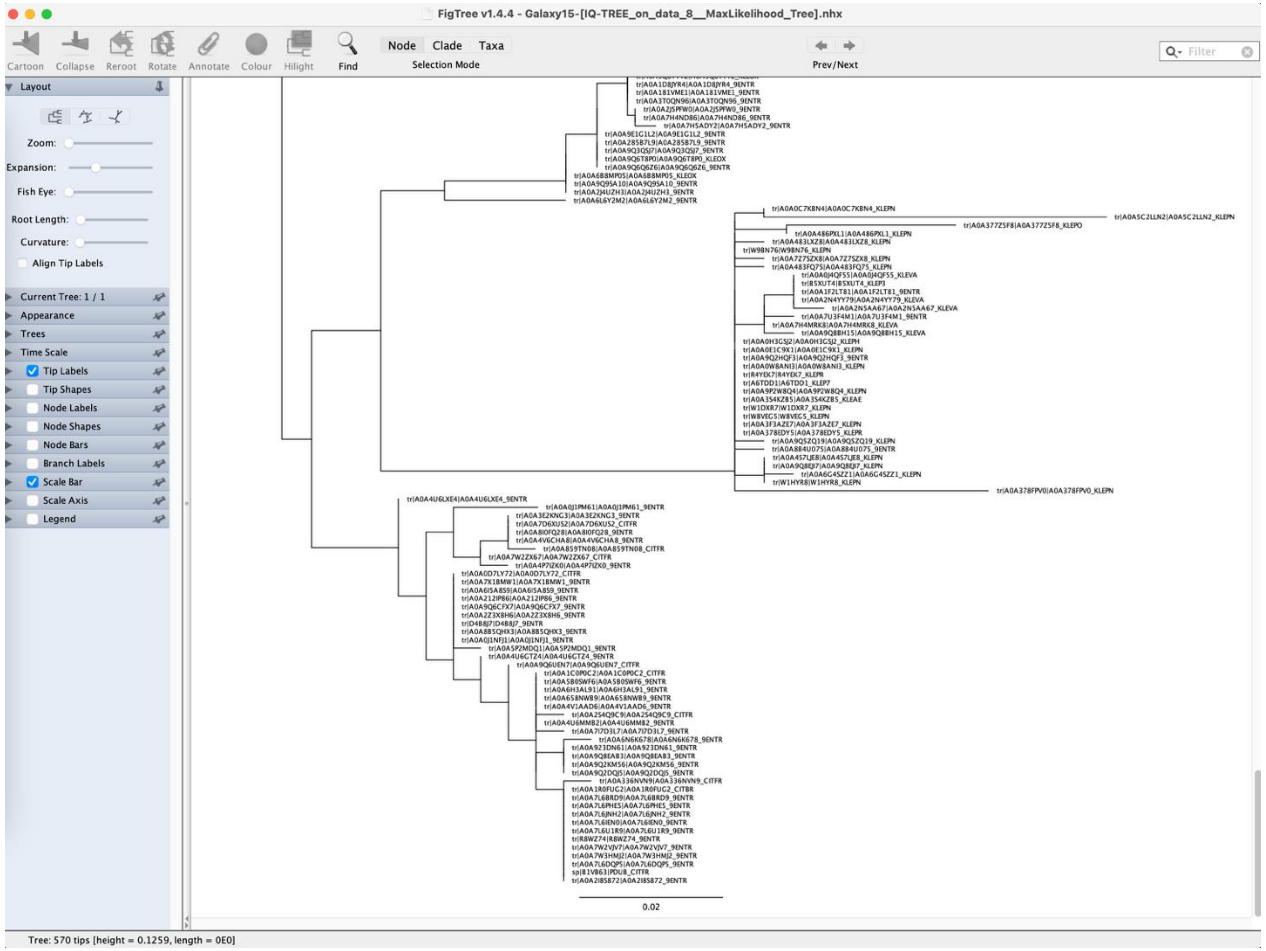


☐ Align Tip Labels
☑ Tip Labels
☐ Tip Shapes
☐ Node Labels
☐ Node Shapes
☐ Node Bars
☐ Branch Labels
☑ Scale Bar
☐ Scale Axis
☐ Legend

Tree: 570 tips [height = 0.176, length = 0E0]

FigTree

- Use the sliders and tree rooting options to make the tree more legible



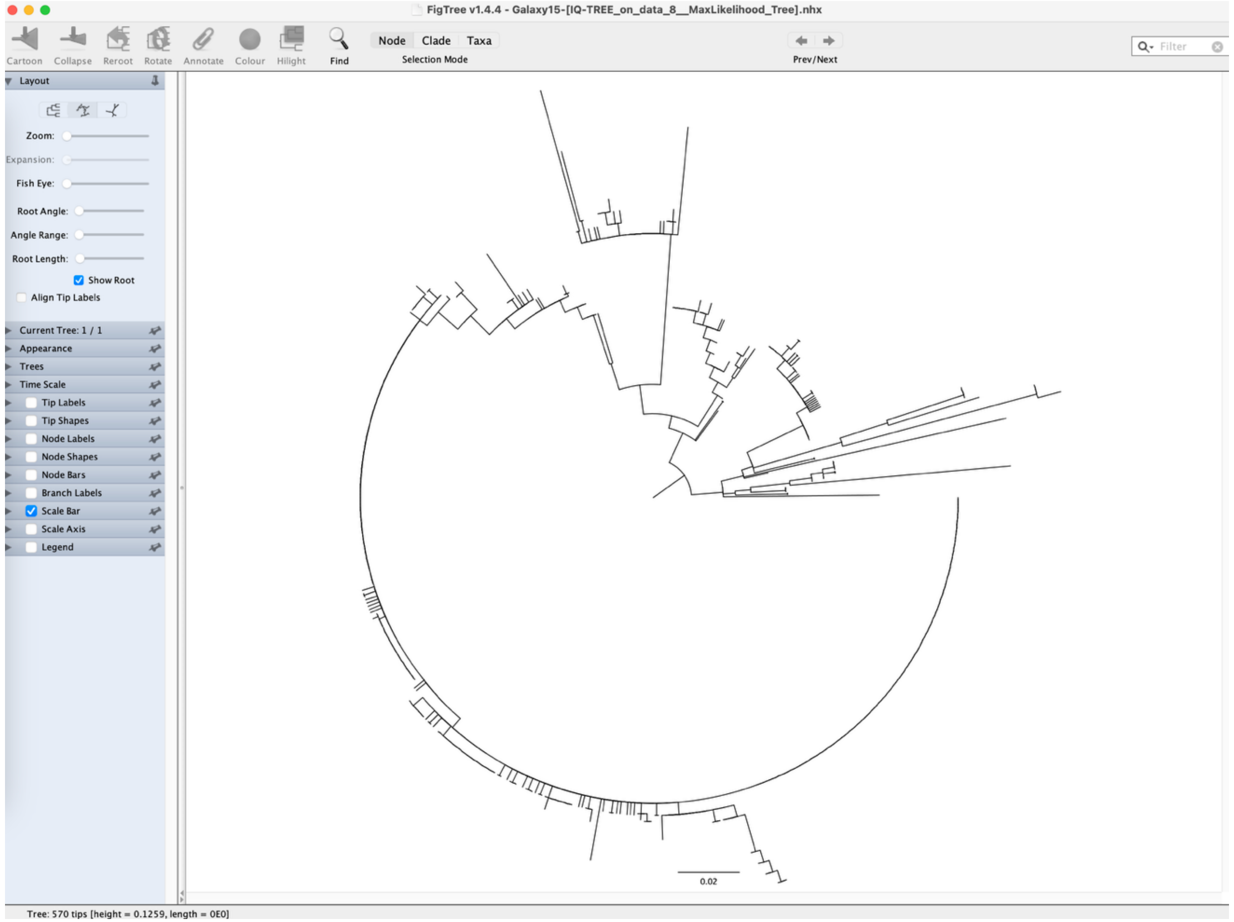
FigTree v1.4.4 - Galaxy15-[IQ-TREE_on_data_8__MaxLikelihood_Tree].nhx

Tree: 570 tips [height = 0.1259, length = 0E0]

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
☑ Tip Labels
□ Tip Shapes
□ Node Labels
□ Node Shapes
□ Node Bars
□ Branch Labels
☑ Scale Bar
□ Scale Axis
□ Legend
tr|A0A1D8JYR4|A0A1D8JYR4_9ENTR
tr|A0A181VME1|A0A181VME1_9ENTR
tr|A0A3T0QN96|A0A3T0QN96_9ENTR
tr|A0A2J5PFW0|A0A2J5PFW0_9ENTR
tr|A0A7H4ND86|A0A7H4ND86_9ENTR
tr|A0A7H5ADY2|A0A7H5ADY2_9ENTR
tr|A0A9E1G1L2|A0A9E1G1L2_9ENTR
tr|A0A285B7L9|A0A285B7L9_9ENTR
tr|A0A9Q3QSJ7|A0A9Q3QSJ7_9ENTR
tr|A0A9Q6T8P0|A0A9Q6T8P0_KLEOX
tr|A0A9Q6Q6Z6|A0A9Q6Q6Z6_9ENTR
tr|A0A6B8MP05|A0A6B8MP05_KLEOX
tr|A0A9Q9SA10|A0A9Q9SA10_9ENTR
tr|A0A2J4UZH3|A0A2J4UZH3_9ENTR
tr|A0A6L6Y2M2|A0A6L6Y2M2_9ENTR
tr|A0A0C7KBN4|A0A0C7KBN4_KLEPN
tr|A0A5C2LLN2|A0A5C2LLN2_KLEPN
tr|A0A377Z5F8|A0A377Z5F8_KLEPO
tr|A0A486PXL1|A0A486PXL1_KLEPN
tr|A0A483LXZ8|A0A483LXZ8_KLEPN
tr|W9BN76|W9BN76_KLEPN
tr|A0A7Z7SZX8|A0A7Z7SZX8_KLEPN
tr|A0A483FQ75|A0A483FQ75_KLEPN
tr|A0A0J4QF55|A0A0J4QF55_KLEVA
tr|B5XUT4|B5XUT4_KLEP3
tr|A0A1F2LT81|A0A1F2LT81_9ENTR
tr|A0A2N4YY79|A0A2N4YY79_KLEVA
tr|A0A2N5AA67|A0A2N5AA67_KLEVA
tr|A0A7U3F4M1|A0A7U3F4M1_9ENTR
tr|A0A7H4MRK8|A0A7H4MRK8_KLEVA
tr|A0A9Q8BH15|A0A9Q8BH15_KLEVA
tr|A0A0H3GSJ2|A0A0H3GSJ2_KLEPH
tr|A0A0E1C9X1|A0A0E1C9X1_KLEPN
tr|A0A9Q2HQF3|A0A9Q2HQF3_9ENTR
tr|A0A0W8ANI3|A0A0W8ANI3_KLEPN
tr|R4YEK7|R4YEK7_KLEPR
tr|A6TDD1|A6TDD1_KLEP7
tr|A0A9P2W8Q4|A0A9P2W8Q4_KLEPN
tr|A0A3S4KZB5|A0A3S4KZB5_KLEAE
tr|W1DXR7|W1DXR7_KLEPN
tr|W8VEG5|W8VEG5_KLEPN
tr|A0A3F3AZE7|A0A3F3AZE7_KLEPN
tr|A0A378EDY5|A0A378EDY5_KLEPR
tr|A0A9Q5ZQ19|A0A9Q5ZQ19_KLEPN
tr|A0A8B4U075|A0A8B4U075_9ENTR
tr|A0A4S7LJE8|A0A4S7LJE8_KLEPN
tr|A0A9Q8EJI7|A0A9Q8EJI7_KLEPN
tr|A0A6G4SZZ1|A0A6G4SZZ1_KLEPN
tr|W1HYR8|W1HYR8_KLEPN
tr|A0A378FPV0|A0A378FPV0_KLEPN
tr|A0A4U6LXE4|A0A4U6LXE4_9ENTR
tr|A0A0J1PM61|A0A0J1PM61_9ENTR
tr|A0A3E2KNG3|A0A3E2KNG3_9ENTR
tr|A0A7D6XUS2|A0A7D6XUS2_CITFR
tr|A0A8I0FQ28|A0A8I0FQ28_9ENTR
tr|A0A4V6CHA8|A0A4V6CHA8_9ENTR
tr|A0A859TN08|A0A859TN08_CITFR
tr|A0A7W2ZX67|A0A7W2ZX67_CITFR
tr|A0A4P7IZK0|A0A4P7IZK0_9ENTR
tr|A0A0D7LY72|A0A0D7LY72_CITFR
tr|A0A7X1BMW1|A0A7X1BMW1_9ENTR
tr|A0A6I5A8S9|A0A6I5A8S9_9ENTR
tr|A0A212IP86|A0A212IP86_9ENTR
tr|A0A9Q6CFX7|A0A9Q6CFX7_9ENTR
tr|A0A2Z3X8H6|A0A2Z3X8H6_9ENTR
tr|D4B8J7|D4B8J7_9ENTR
tr|A0A8B5QHX3|A0A8B5QHX3_9ENTR
tr|A0A0J1NFJ1|A0A0J1NFJ1_9ENTR
tr|A0A5P2MDQ1|A0A5P2MDQ1_9ENTR
tr|A0A4U6GTZ4|A0A4U6GTZ4_9ENTR
tr|A0A9Q6UEN7|A0A9Q6UEN7_CITFR
tr|A0A1C0P0C2|A0A1C0P0C2_CITFR
tr|A0A5B0SWF6|A0A5B0SWF6_9ENTR
tr|A0A6H3AL91|A0A6H3AL91_9ENTR
tr|A0A658NWB9|A0A658NWB9_9ENTR
tr|A0A4V1AAD6|A0A4V1AAD6_9ENTR
tr|A0A2S4Q9C9|A0A2S4Q9C9_CITFR
tr|A0A4U6MMB2|A0A4U6MMB2_9ENTR
tr|A0A7I7D3L7|A0A7I7D3L7_9ENTR
tr|A0A6N6K678|A0A6N6K678_9ENTR
tr|A0A923DN61|A0A923DN61_9ENTR
tr|A0A9Q8EAB3|A0A9Q8EAB3_9ENTR
tr|A0A9Q2KM56|A0A9Q2KM56_9ENTR
tr|A0A9Q2DQJ5|A0A9Q2DQJ5_9ENTR
tr|A0A336NVN9|A0A336NVN9_CITFR
tr|A0A1R0FUG2|A0A1R0FUG2_CITBR
tr|A0A7L6BRD9|A0A7L6BRD9_9ENTR
tr|A0A7L6PHE5|A0A7L6PHE5_9ENTR
tr|A0A7L6JNH2|A0A7L6JNH2_9ENTR
tr|A0A7L6IEN0|A0A7L6IEN0_9ENTR
tr|A0A7L6U1R9|A0A7L6U1R9_9ENTR
tr|R8WZ74|R8WZ74_9ENTR
tr|A0A7W2VJV7|A0A7W2VJV7_9ENTR
tr|A0A7W3HMJ2|A0A7W3HMJ2_9ENTR
tr|A0A7L6DQP5|A0A7L6DQP5_9ENTR
sp|B1VB63|PDUB_CITFR
tr|A0A2I8S872|A0A2I8S872_9ENTR
0.02

FigTree

- Use tree layout options to make the whole tree easier to see/interpret



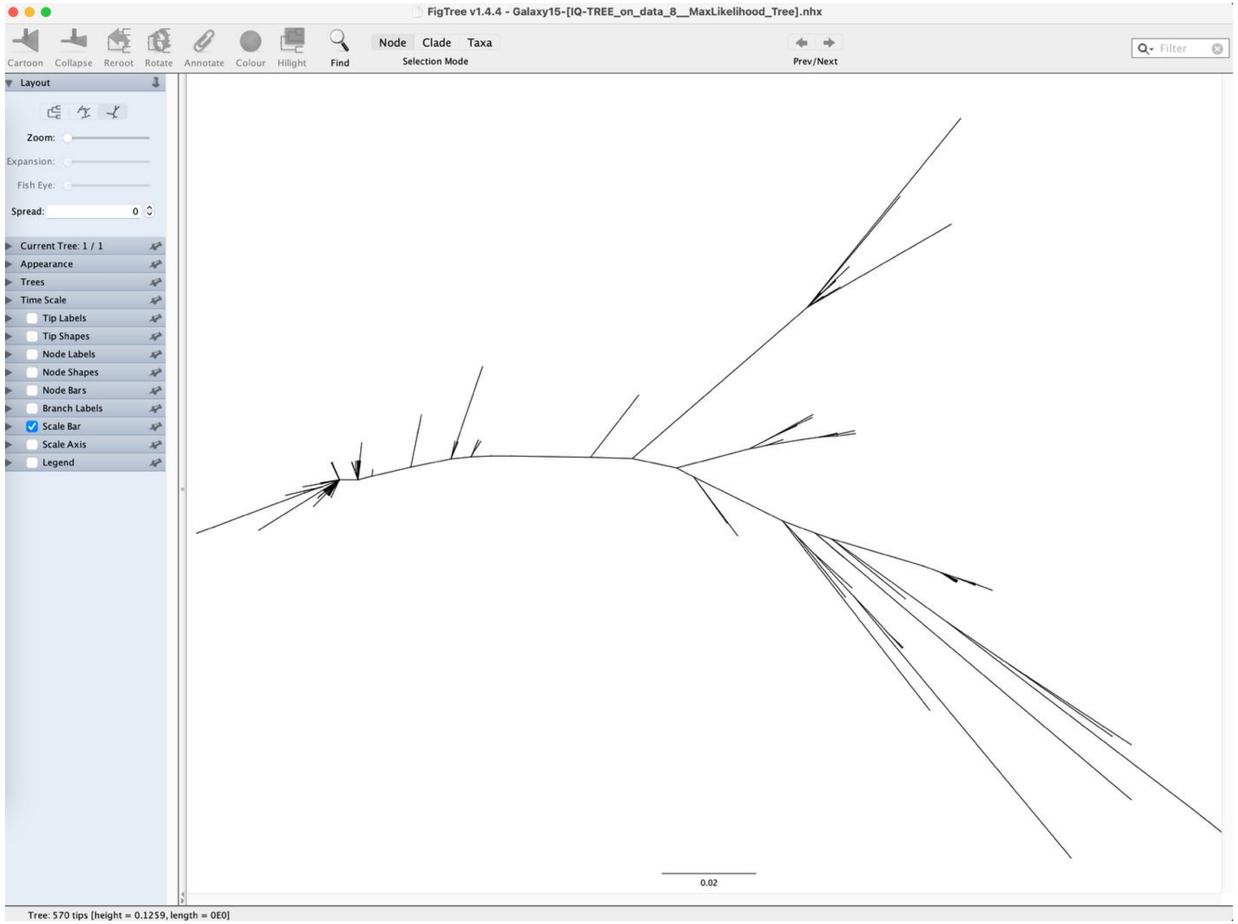
FigTree v1.4.4 - Galaxy15-[IQ-TREE_on_data_8__MaxLikelihood_Tree].nhx

Zoom:
Expansion:
Fish Eye:
Root Angle:
Angle Range:
Root Length:
☑ Show Root
☐ Align Tip Labels

Current Tree: 1 / 1

☐ Tip Labels
☐ Tip Shapes
☐ Node Labels
☐ Node Shapes
☐ Node Bars
☐ Branch Labels
☑ Scale Bar
☐ Scale Axis
☐ Legend

Tree: 570 tips [height = 0.1259, length = 0E0]



FigTree v1.4.4 - Galaxy15-[IQ-TREE_on_data_8__MaxLikelihood_Tree].nhx

Zoom:
Expansion:
Fish Eye:
Spread: 0

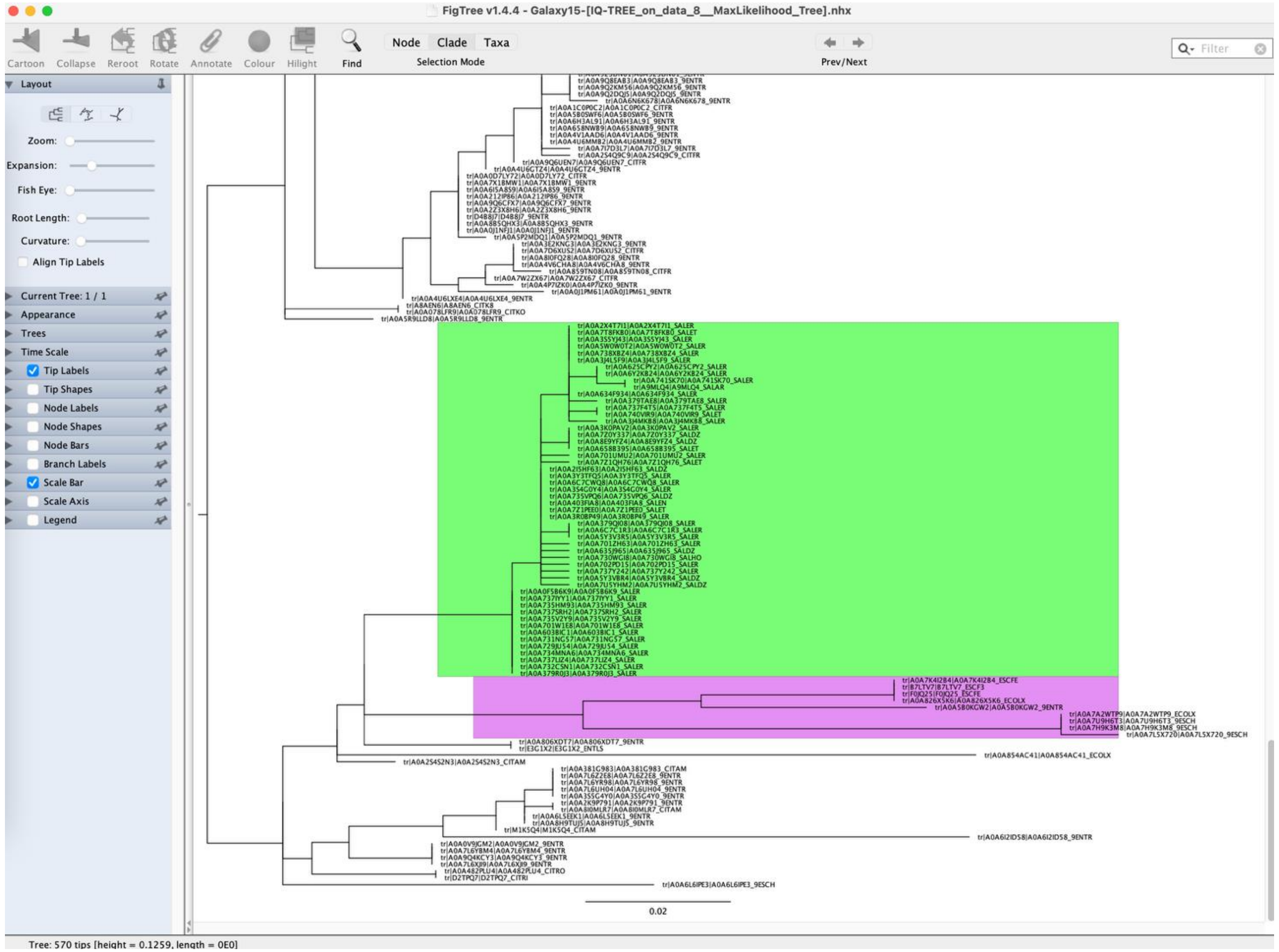
Current Tree: 1 / 1

☐ Tip Labels
☐ Tip Shapes
☐ Node Labels
☐ Node Shapes
☐ Node Bars
☐ Branch Labels
☑ Scale Bar
☐ Scale Axis
☐ Legend

Tree: 570 tips [height = 0.1259, length = 0E0]

FigTree

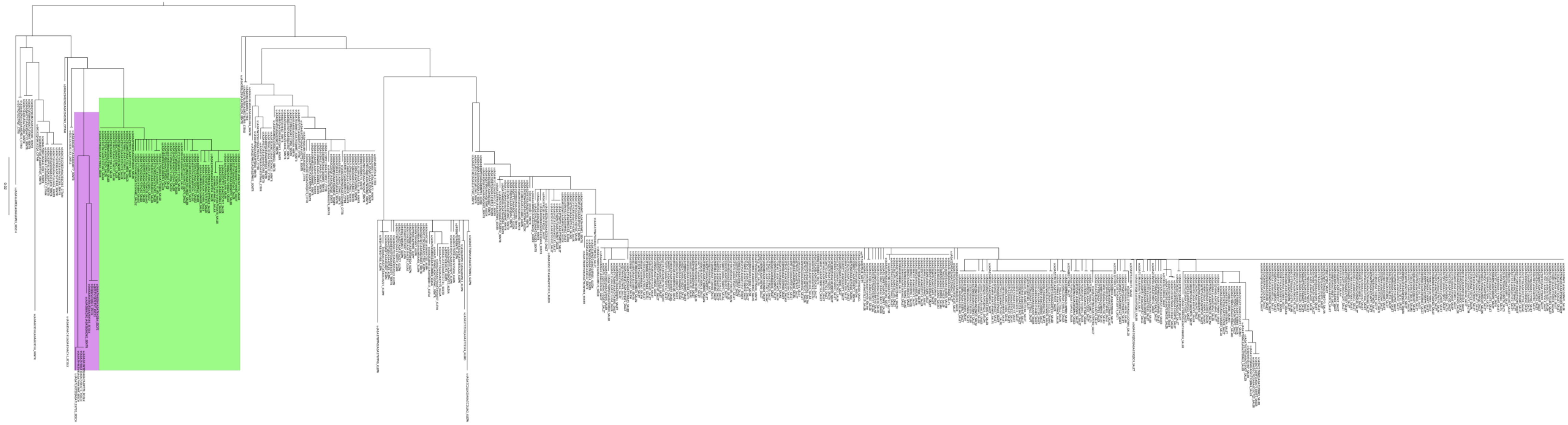
- Use selection and colour options to highlight groups of sequences for the reader



Tree: 570 tips [height = 0.1259, length = 0E0]

FigTree

- Export .png, .jpg, .pdf, .svg files for inserting into your thesis (or presentation)



Going further

University of Strathclyde Science

- Using nucleotide alignments can be more informative and accurate, but backtracing can be tricky – not all proteins have a known coding sequence
  - https://ncfp.readthedocs.io/en/stable/
- Identify the most appropriate substitution model before building the tree
  - (built-in for IQ-tree and RAxML)
- Maximum likelihood methods are the baseline standard
  - IQ-tree, RAxML, etc.
- Bayesian methods are statistically more robust, but are computationally very intensive
  - RevBayes, BEAST
- Bootstrapping gives an estimate of the robustness of your tree to changes in the input data

Next Week’s Group Meetings

Tuesday 20th October 13:30 HW324

Thursday 23rd October 10:30 HW324

Topics to Discuss at Next Meeting

- What would you like to cover?

Useful Links

Useful tools (many others are available)

GalaxyEU: https://usegalaxy.eu/

- Sequence alignment (e.g. MAFFT), phylogenetics (e.g. RaxML), positive selection (e.g. codeML)

iTOL: https://itol.embl.de/

- Visualisation/annotation of phylogenetic trees

PyMOL: https://pymol.org/2/ and/or ChimeraX: https://www.cgl.ucsf.edu/chimerax/

- Protein structure visualisation/annotation

Jalview: http://www.jalview.org/

- Visualisation of multiple sequence alignments

Useful sites/databases

PHI-base: http://www.phi-base.org/

- Proteins involved in host-pathogen interactions, with linked evidence

EMBL AlphaFold: https://www.alphafold.ebi.ac.uk/

- AlphaFold predictions for proteins from model organisms

UniProt: https://www.uniprot.org/

- Protein sequence (including homologous sequences) and functional information with evidence

RCSB/PDB: https://www.rcsb.org/

- Repository of record for protein structures

SIPBS CompBiol Sites

- BM432 Project Pages
  - https://sipbs-compbiol.github.io/bm432-project/

- An incomplete little book of bioinformatics
  - https://sipbs-compbiol.github.io/little-bioinformatics-book/

Project Management Tools

You may want tools to…

- Manage your time
  - E.g. Pomodoro technique (e.g. BeFocused, Pomofocus, Forest)
- Schedule work
  - Reminders (macOS, MS Office)
  - Calendar (macOS, MS Office), with email alerts
  - Trello, Asana, etc.
- Manage your project data and information effectively
  - How to name files
  - Project management guidelines (BM432, 2022-23 session; me and Dr Feeney)
  - How to keep a lab notebook
  - Keeping a computational biology lab notebook: https://doi.org/10.1371/journal.pcbi.1004385
  - Organising a lab book