T-BAS: A Tool for Real-time Tracking of Biodiversity Across the Tree of Life

Ignazio Carbone

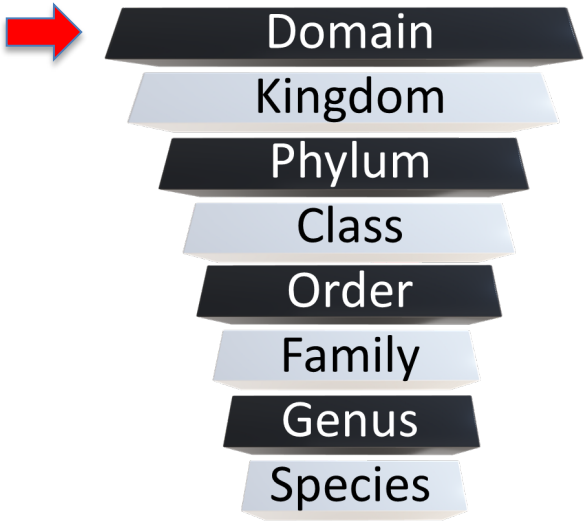
Department of Entomology and Plant Pathology

NC STATE Center for Integrated Fungal Research

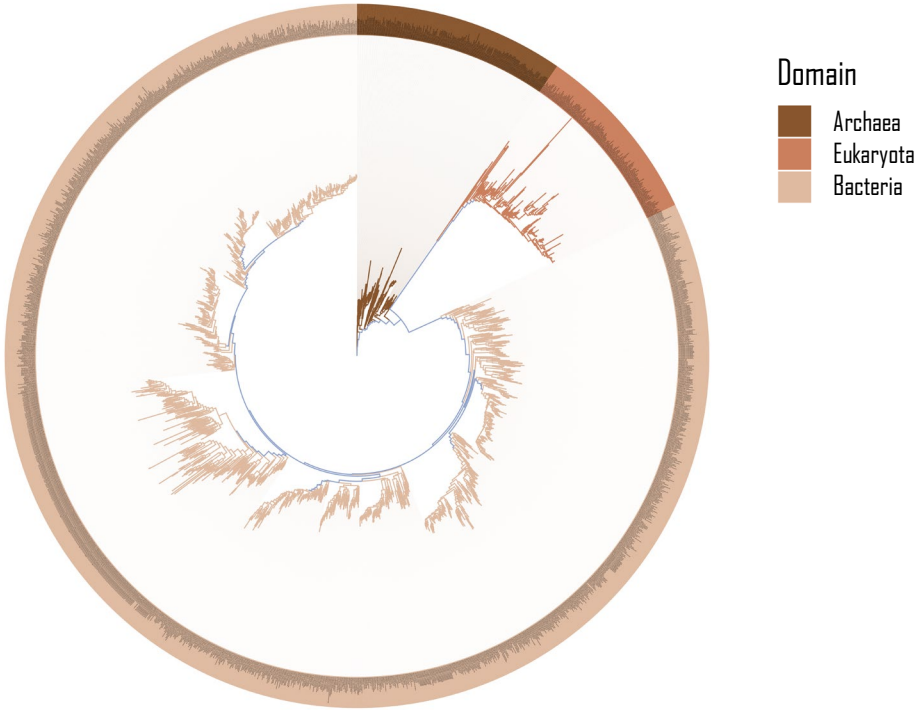
https://tbas.cifr.ncsu.edu/

- Taxonomy should reflect phylogeny

Biological Classification

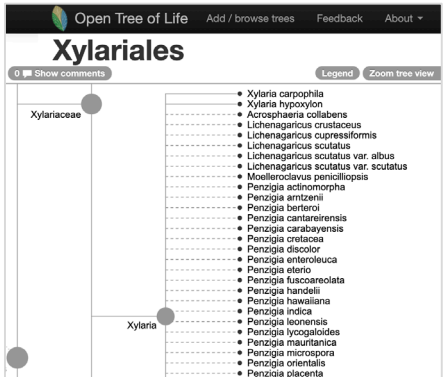


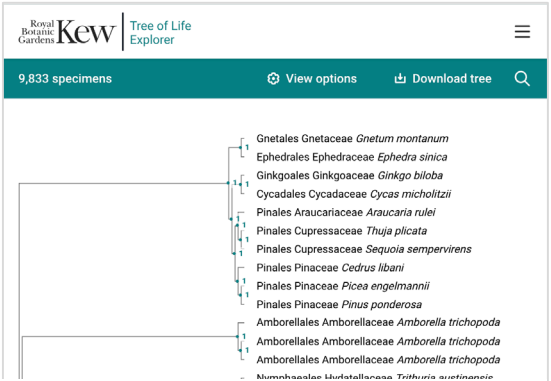
Branches on the Tree of Life

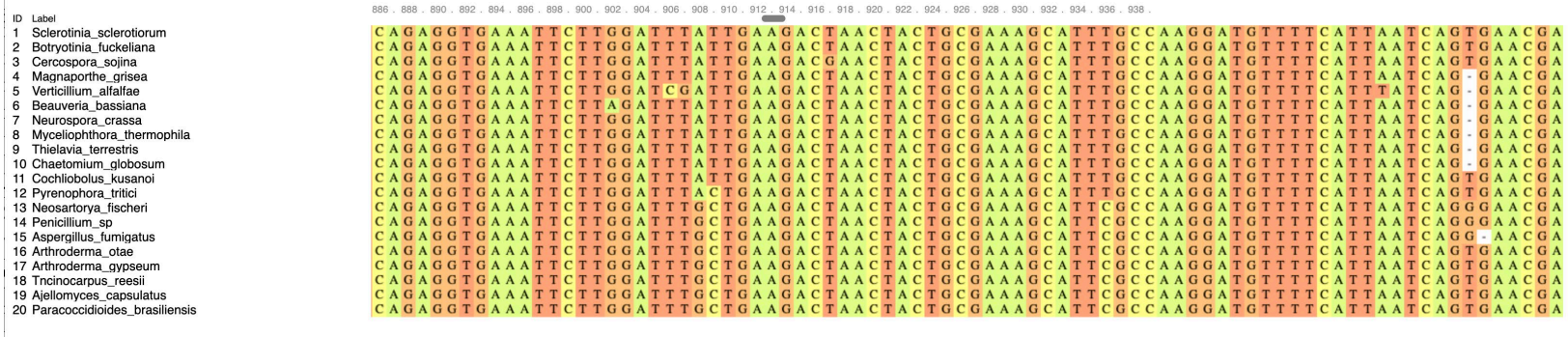


- Current tree of life initiatives synthesize published phylogenetic trees along with taxonomic data

No sequence alignments

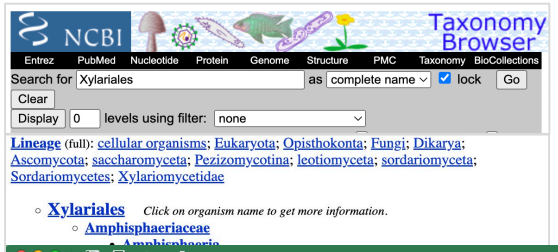


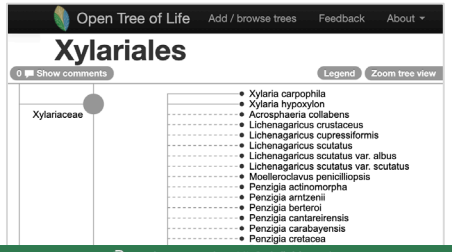


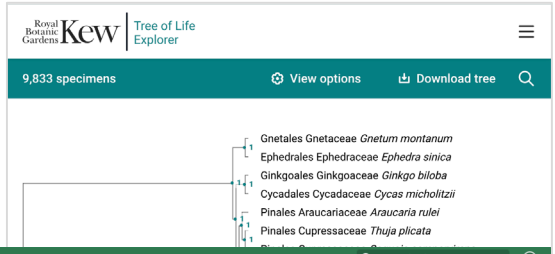


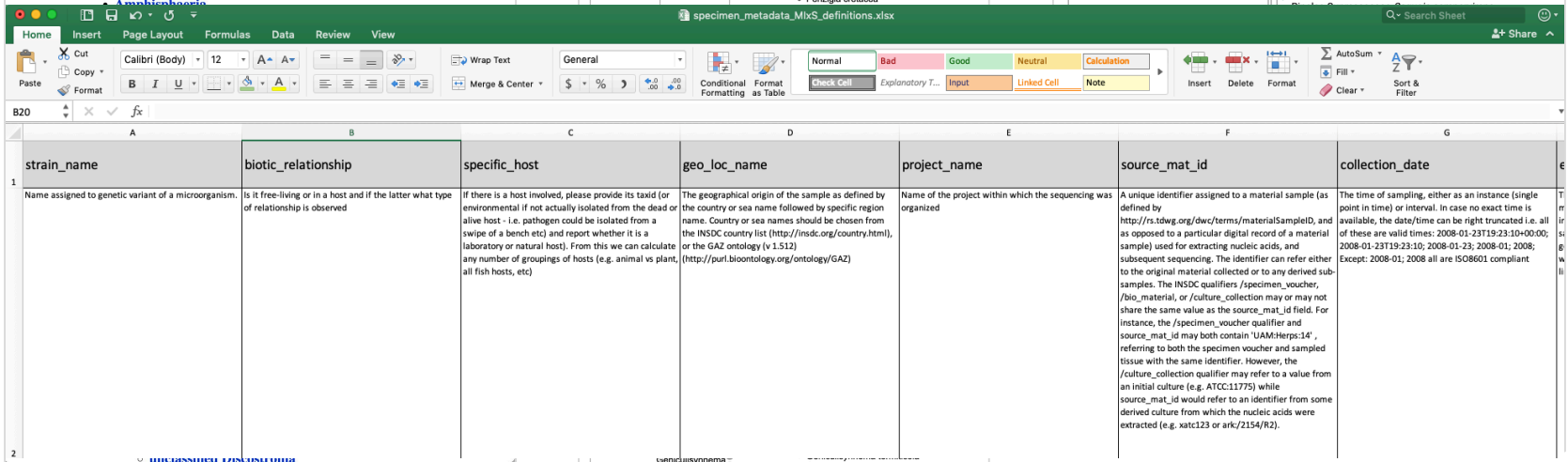
- Current tree of life initiatives synthesize published phylogenetic trees along with taxonomic data

No specimen metadata



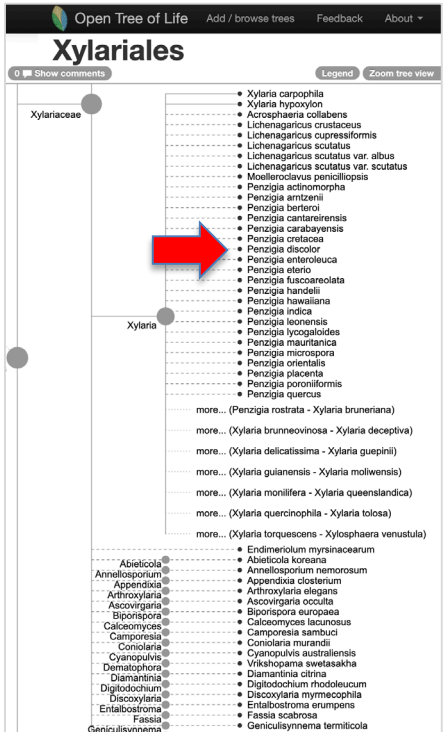


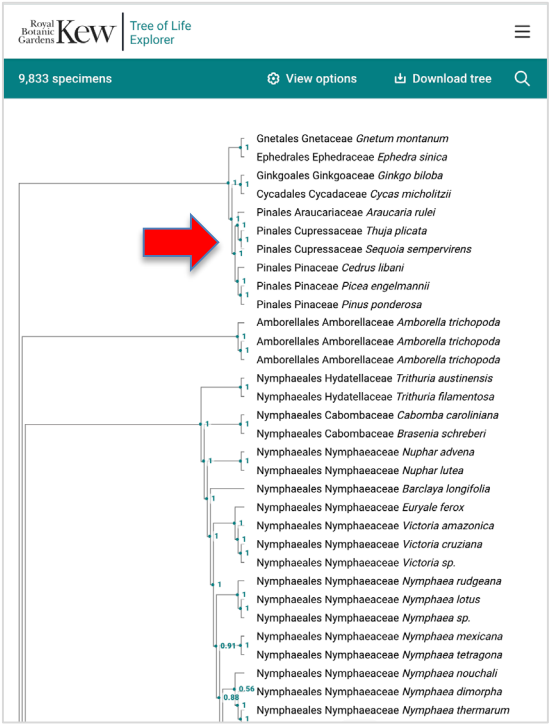




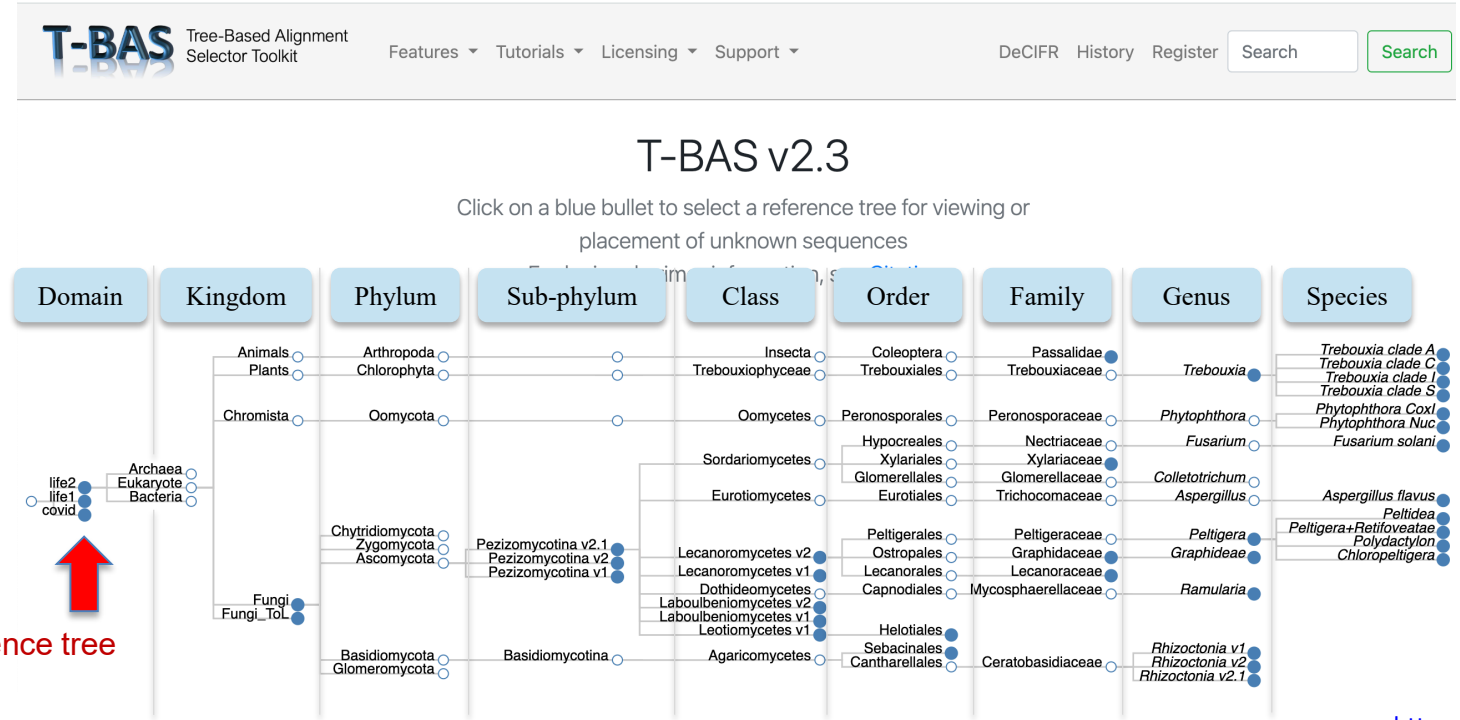
- Current tree of life initiatives synthesize published phylogenetic trees along with taxonomic data

No placement of unknown sequences



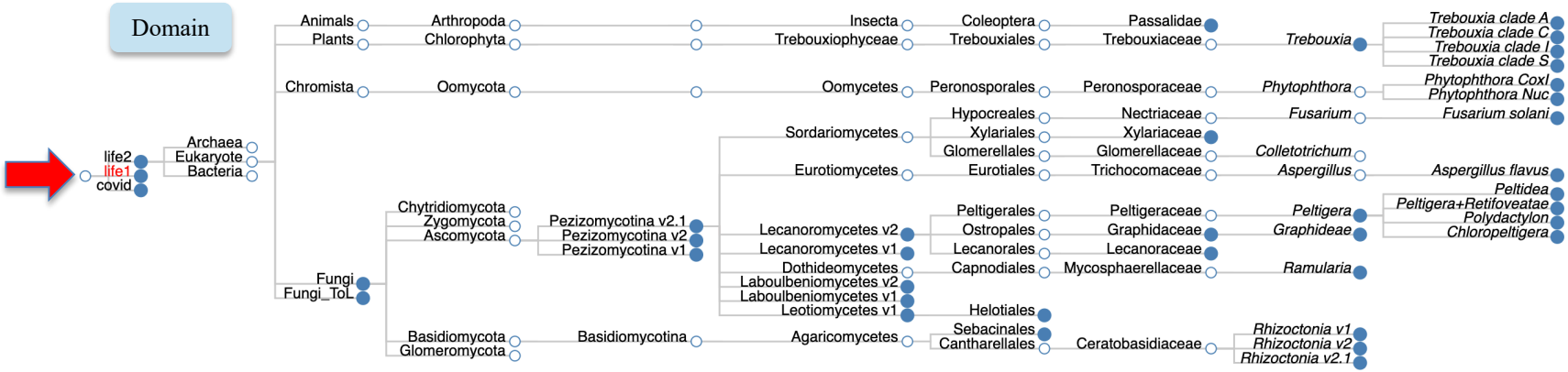


- Phylogenetic integration of unknown sequences with reference trees, alignments and metadata from cultured specimens



https://tbas.cifr.ncsu.edu/

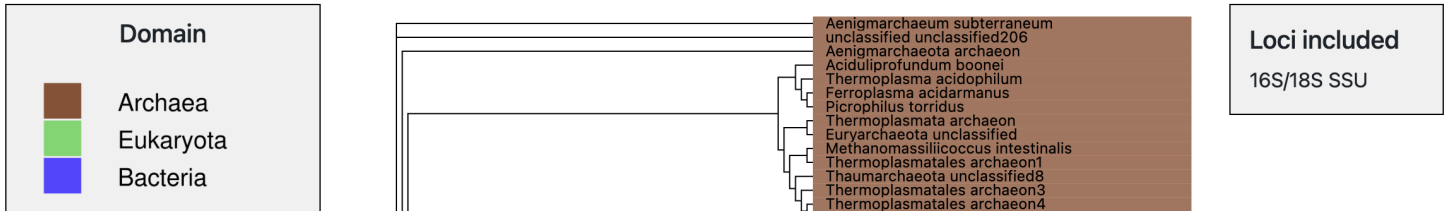
Single-locus phylogeny-based placement of 16S/18S rRNA



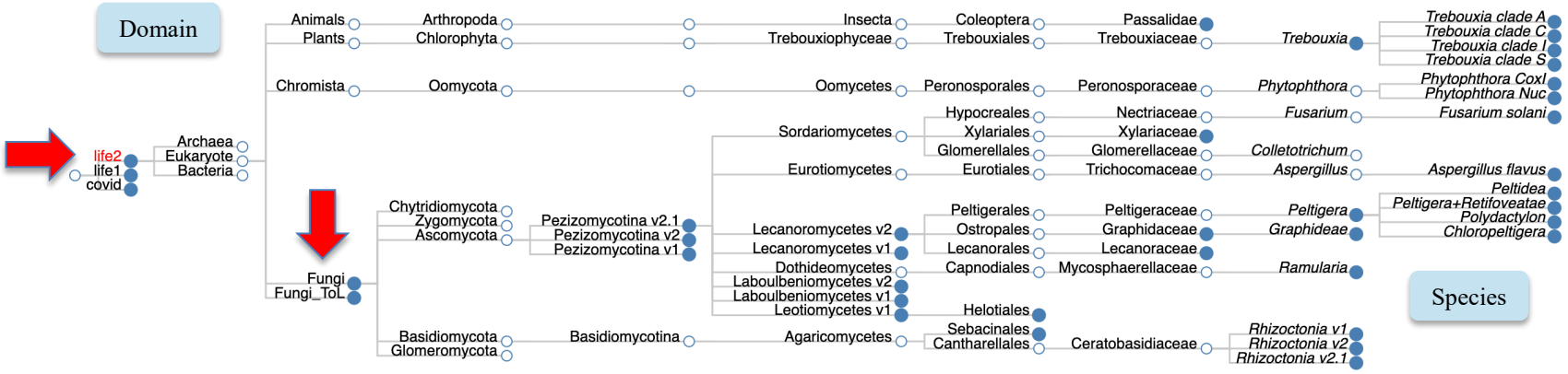
- Targeted Amplicon Sequencing
- Genomes
- Metagenomes

16S/18S SSU reference tree

View Tree Data | Place Unknowns



Multi-locus phylogeny-based placement of 16 ribosomal proteins

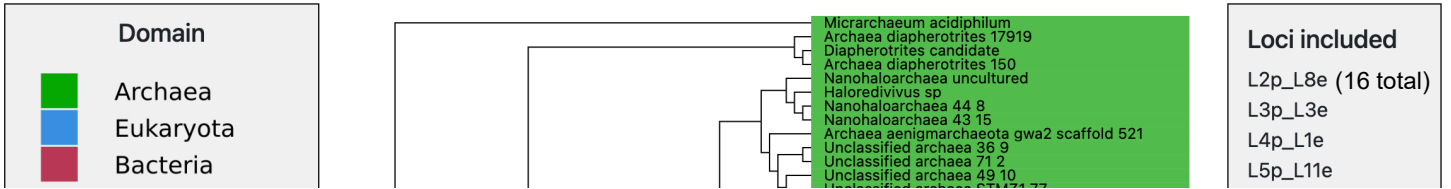


Increasing taxon sampling

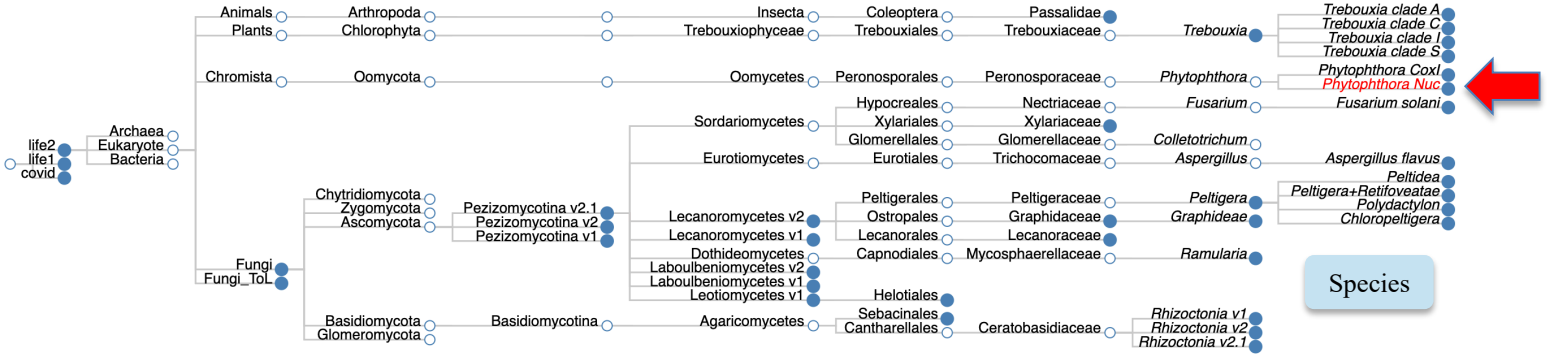
- Multi-locus phylogeny
- Genome-scale phylogeny

multilocus ribosomal protein reference tree

View Tree Data | Place Unknowns



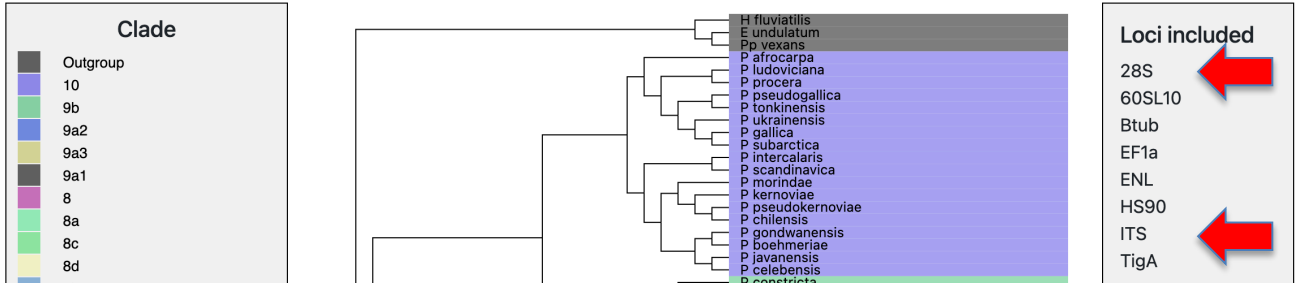
Multi-locus phylogeny-based placement for species delimitation



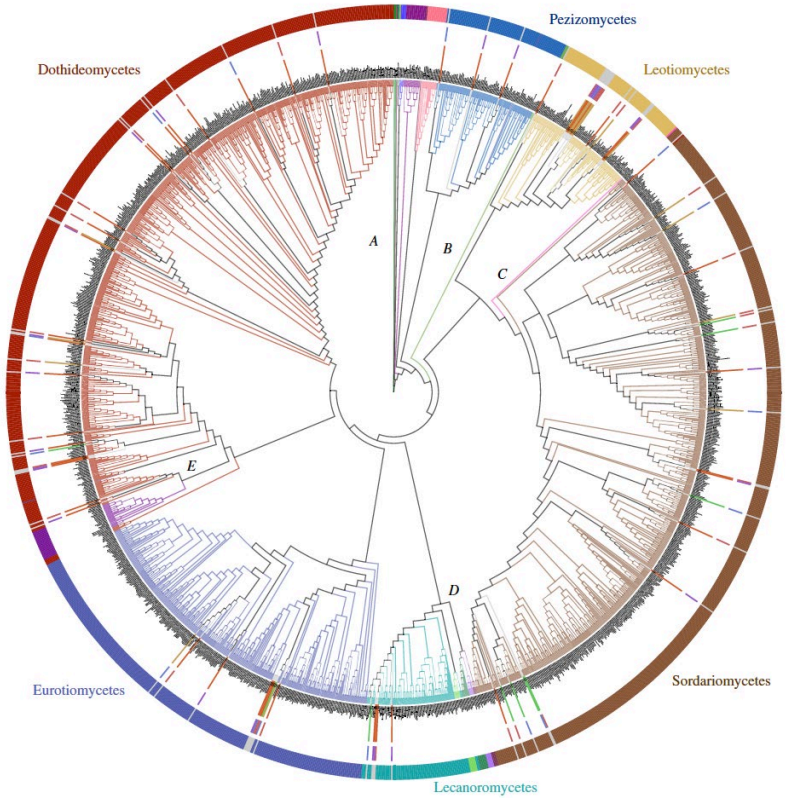
- Multi-level placement
- Population genetics/genomics

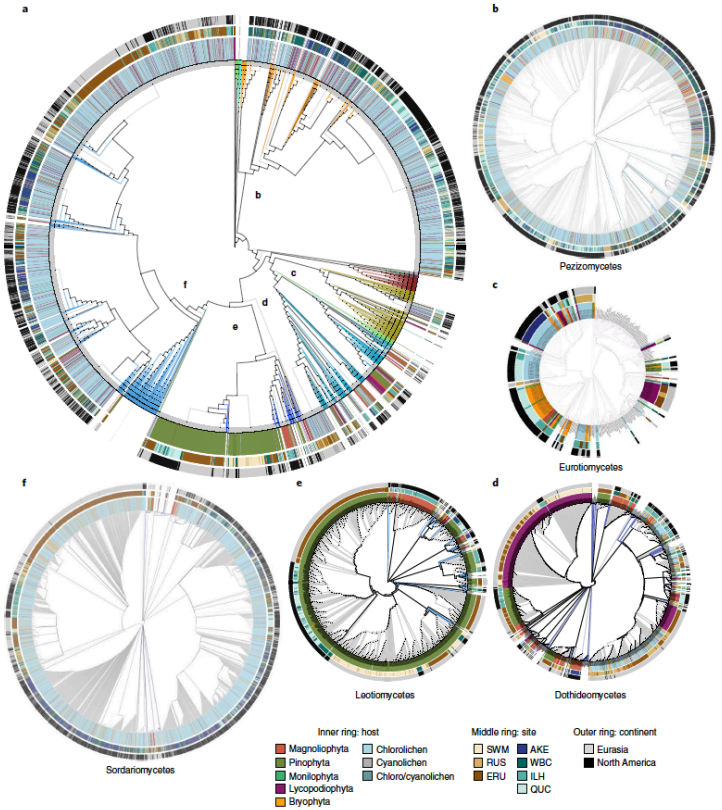
Phytophthora Nuclear

Coomber, A., Saville, A., Carbone, I. and J. B. Ristaino. 2023. An open-access T-BAS phylogeny for emerging Phytophthora species. PLoS One In Press



Integration of taxonomic information, alignments, and collection metadata



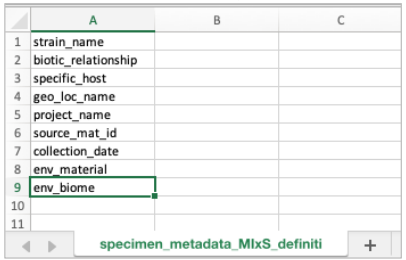


Daru *et al.* (2018). Philos Trans R Soc Lond B Biol Sci 374.

U'Ren *et al.* (2019). *Nat Ecol Evol* **3**, 1430-1437.

T-BAS tracks metadata in phylogenies using MEP data standard

T-BAS Tree-Based Alignment Selector Toolkit — Features ▾ Tutorials ▾ Licensing ▾ Support ▾ DeCIFR History Register



	A	B	C
1	strain_name		
2	biotic_relationship		
3	specific_host		
4	geo_loc_name		
5	project_name		
6	source_mat_id		
7	collection_date		
8	env_material		
9	env_biome		
10			
11			

specimen_metadata_MIxS_definiti +

Metadata Enhanced PhyloXML (MEP)

Authors

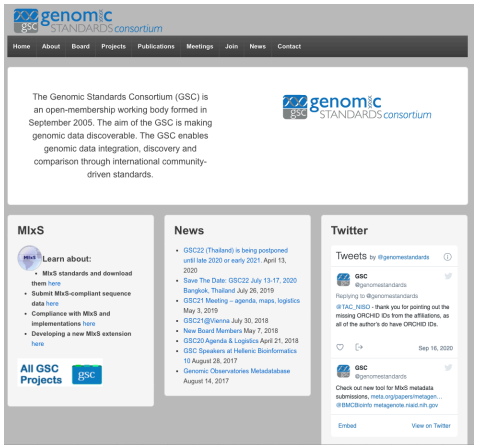
- Jim White
- Ignazio Carbone

Description

In T-BAS, DNA sequences and associated specimen metadata are phylogenetically placed on curated multilocus reference trees and the placement results are saved as Metadata Enhanced PhyloXML (MEP) files. The MEP format allows placements and associated specimen attributes (e.g. host, locality, environmental traits) to be readily viewed, archived and importantly analyzed within a phylogenetic context. MEP files are structured to adhere to the minimum information about any (x) sequence (MIxS) family of standards defined by the Genomic Standards Consortium. A template is provided for users to fill in and submit when performing a phylogeny-based placement in T-BAS. Additional categories of metadata information can be added. MIxS headers and metadata are saved in MEP files as defined in the XML schemas below. The use of MEP files ensures interoperability and retrieval of relevant sequences and metadata for downstream applications. MEP is based on XML, a widely used markup language for representing and sharing information, and PhyloXML, an extension of XML with custom tags for describing evolutionary trees or networks.

The standard pre-defined XML schema for phyloXML is used as a starting point for validating MEP files. PhyloXML includes a phylogeny element that saves the tree information and associated alignments. MEP extends this by adding (1) a tag to each clade that is a leaf in the tree and saving the metadata for that leaf, (2) a gene tag that saves the locus name, the number of sequence characters, and the positions of the excluded unaligned character set (i.e. exset) for each alignment, and (3) an OTUs tag that saves the taxonomic assignments, associated query metadata and sequences for each OTU.

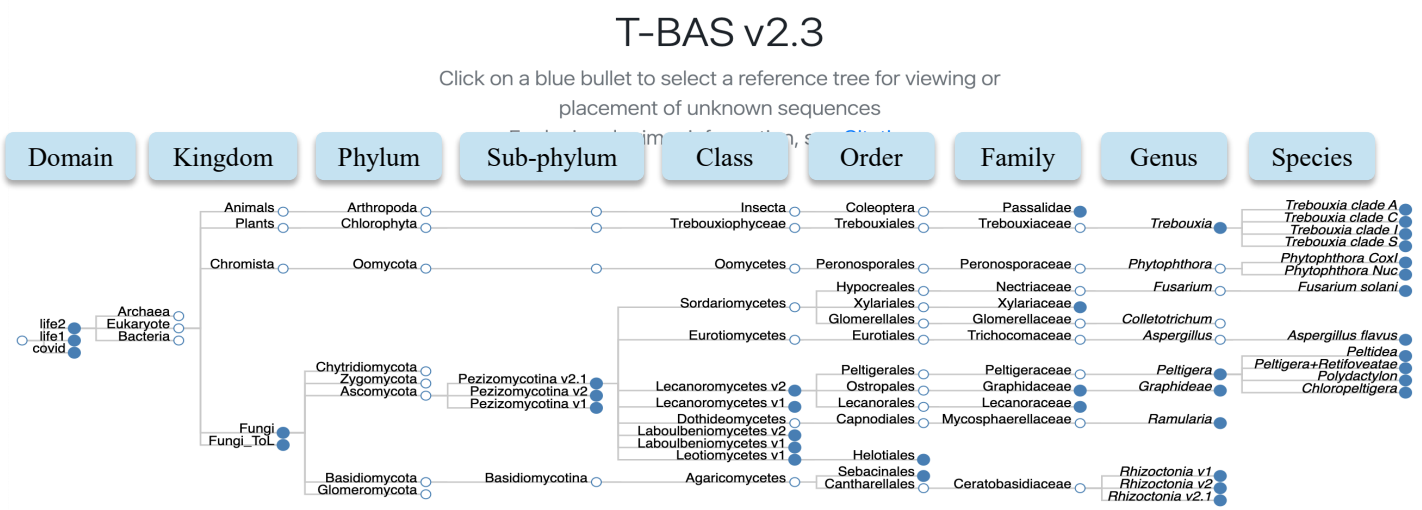
MEP uses two associated schema definitions:



https://decifr.cifr.ncsu.edu/schema.php

T-BAS is a dynamic and open-access tree of life

- Features a phylogenomic database for taxonomy and metadata validation
- Real-time tracking of biodiversity across taxonomic ranks
- Species delimitation and population genetic analysis based on multi-locus genomic data
- Can be used to track and catalog the uncultivated microbial diversity at all taxonomic ranks



T-BAS version 2.3: phylogenetic-data repository and tool for phylogeny-based placement

Funding

- NSF Dimensions of Biodiversity
- NSF Genealogy of Life
- NSF Ecology and Evolution of Infectious Diseases
- Novo Nordisk Foundation Collaborative Crop Resilience Program
- NSF Predictive Intelligence for Pandemic Prevention

Collaborators

- Cyberinfrastructure for Phylogenetic Research (CIPRES)

novo nordisk fonden

ACCESS Advancing Innovation

CIPRES