Lost in translation: challenges of transferring information from phylogenies into taxonomy

Romina Gazis*, Jonathan Rees, Rick Ree, Stephen Smith, David Hibbett

IMC10

Build a Tree of Life that is……

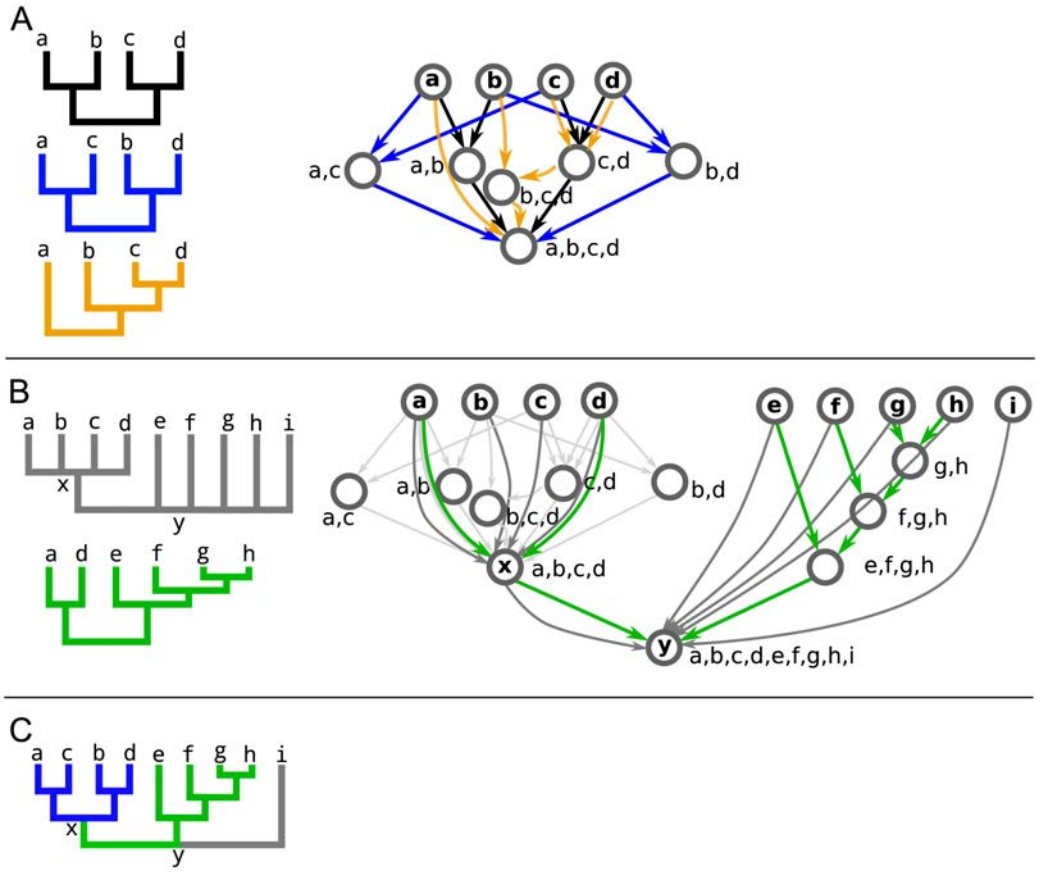
- Complete: contains ALL biodiversity
- Dynamic: continuously updated as more data become available
- Available digitally to browse, search, and download

How?

- Synthesize digitally available phylogenies and taxonomies into a synthetic tree
- Graph building method

Graph building method

- Alternative method to supermatrix and supertree approaches
- Allows the merge of two phylogenies with incomplete overlapping taxa (at least one taxa in common)
- The resolution of conflict can be automatized based on specific criteria (i.e., year, # of genes, # taxa)



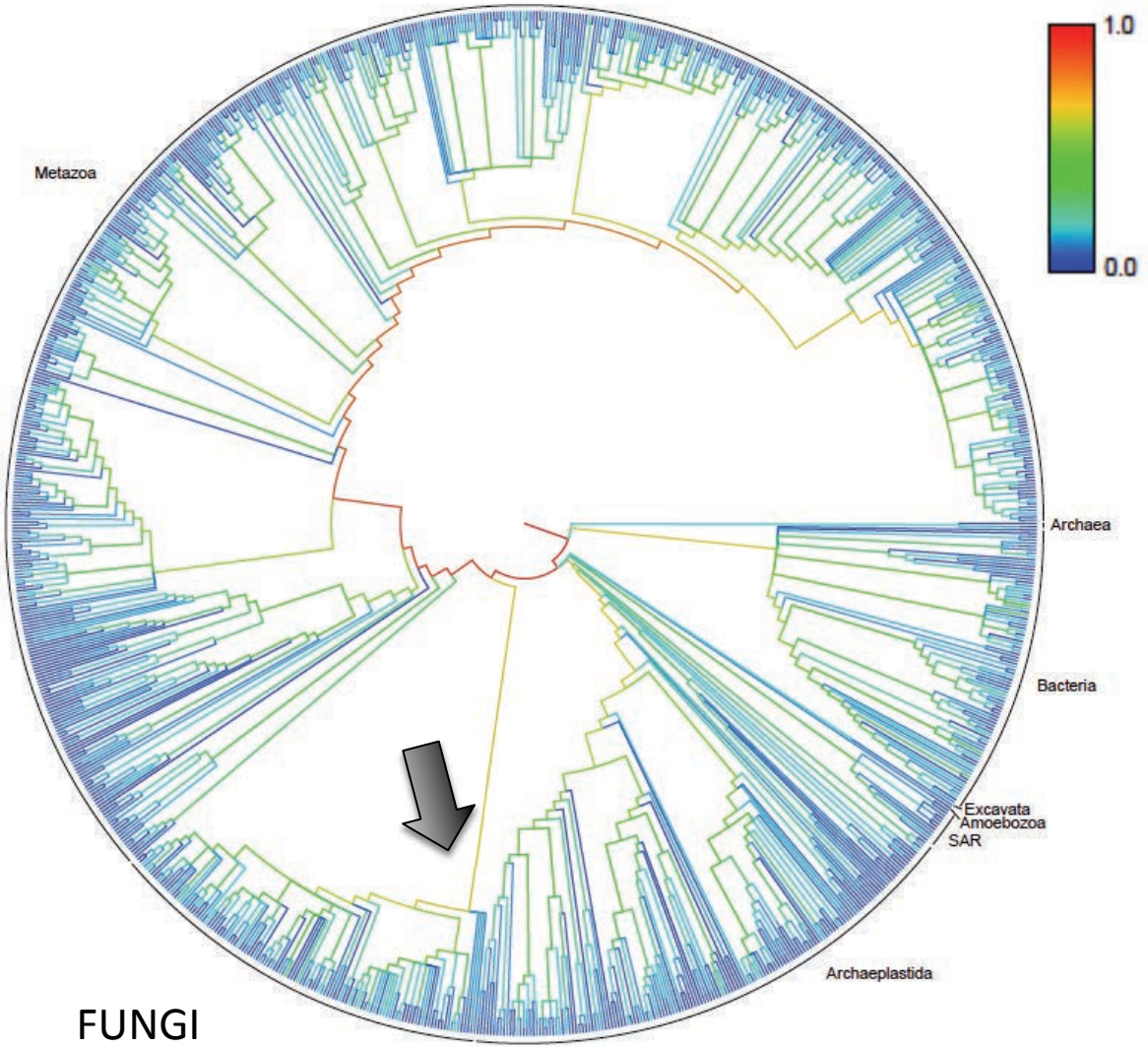
Smith, Brown, Hinchliff. 2013. Analyzing and synthesizing phylogenies using tree alignment graphs. Plos Computational Biology.

Status of the project

- Synthetic taxonomy with >2.5 million taxa
  → OpenTreeTaxonomy = OTT v. 2.6 (now v. 2.8)
- 1st version of the synthetic tree with all taxonomy taxa plus phylogenetic resolution from > 300 studies (May)
- 2nd version of the synthetic tree with all taxonomy taxa plus phylogenetic resolution from > 500 studies (July)
- Several thousand other studies in database in line for synthesis
- New web application for curation (more user friendly)

Synthetic tree version 1 (v. 2 in progress)

- 336 source phylogenies + many taxonomic sources
- 142,325 clades
- **~ 83% of the clades have only taxonomy as support**



* Lineages with less than 500 taxa are collapsed

Status for Fungi

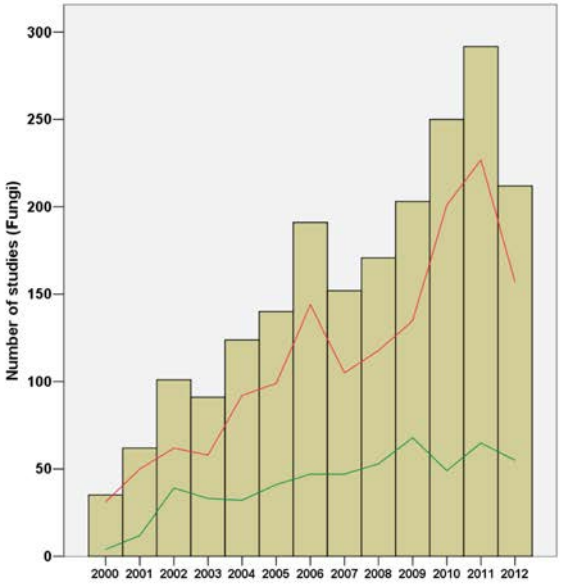
- Synthetic taxonomy
  - OTT v. 2.8 – mainly based on Index Fungorum

  Tips: 346,480 / 65,153 hidden (- incertae sedis, fossils)
- Synthetic tree contains ONLY 12 phylogenetic studies
  - Version 2 = ~60 fungal tree sources
- ~ 1,000 studies waiting to be included in synthesis

Challenges (Fungi):

DATA availability → DATA curation → Real conflict?

Low number of available phylogenies	Rooting of trees Mapping of terminals Determine ingroup	Real conflict because of gene history OR artifact of current taxonomy



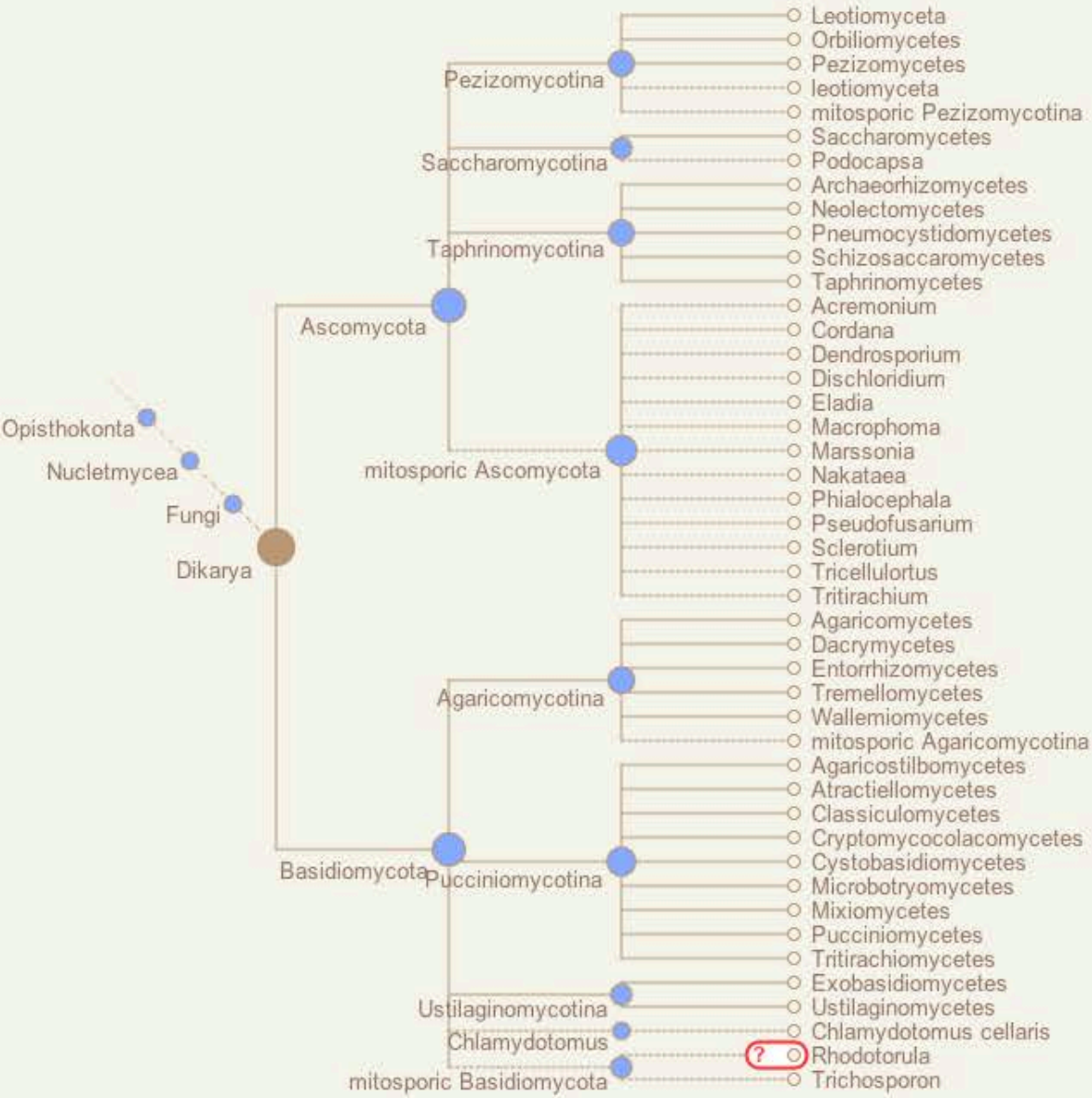
- Most of the studies needed extra curation (re-rooting, label editing and mapping of terminals)
- Conflicts between taxonomic databases (**IF** / GBIF / NCBI)

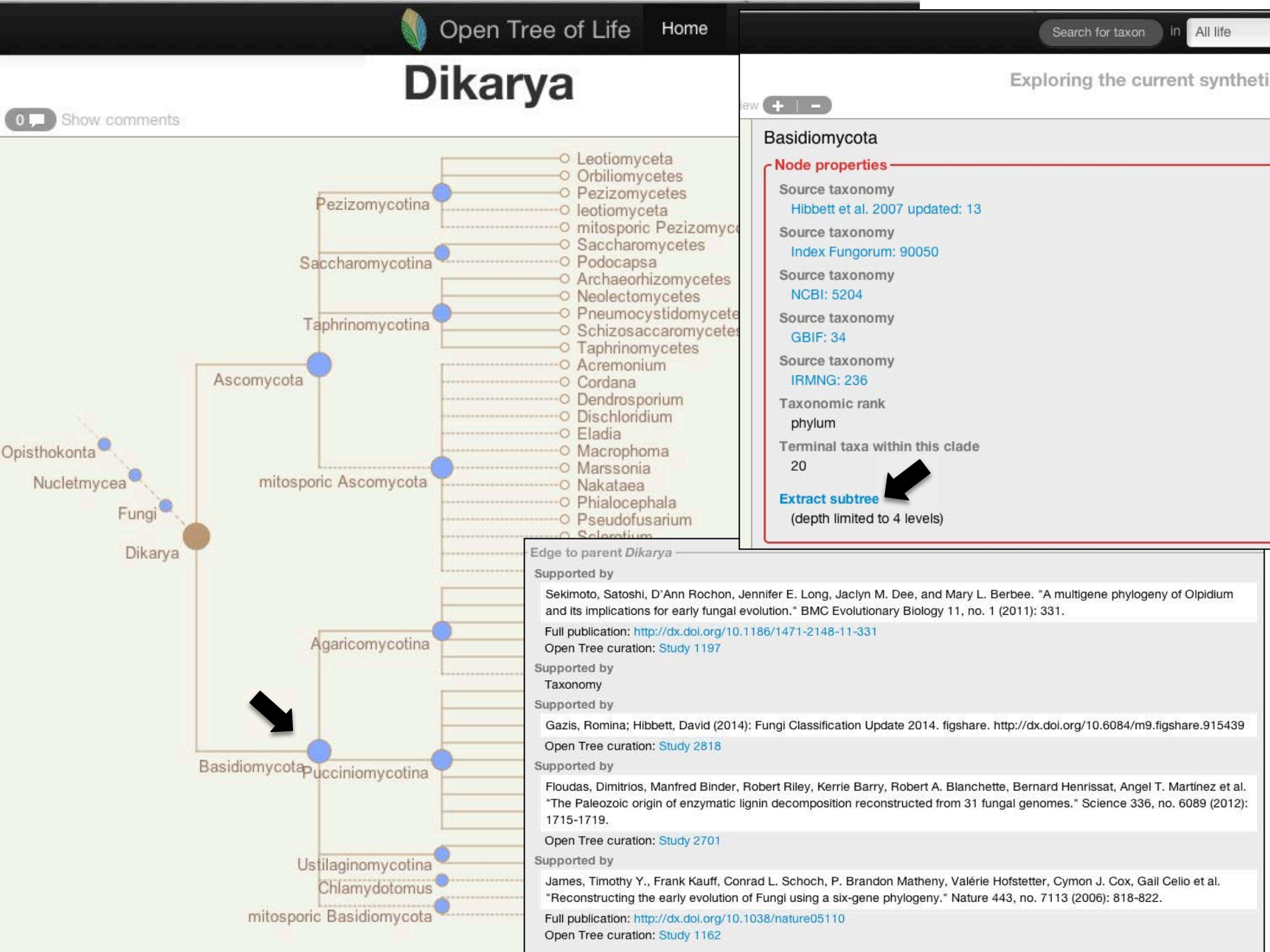
- If there is too much conflict with the OTT taxonomy the trees do not pass to synthesis

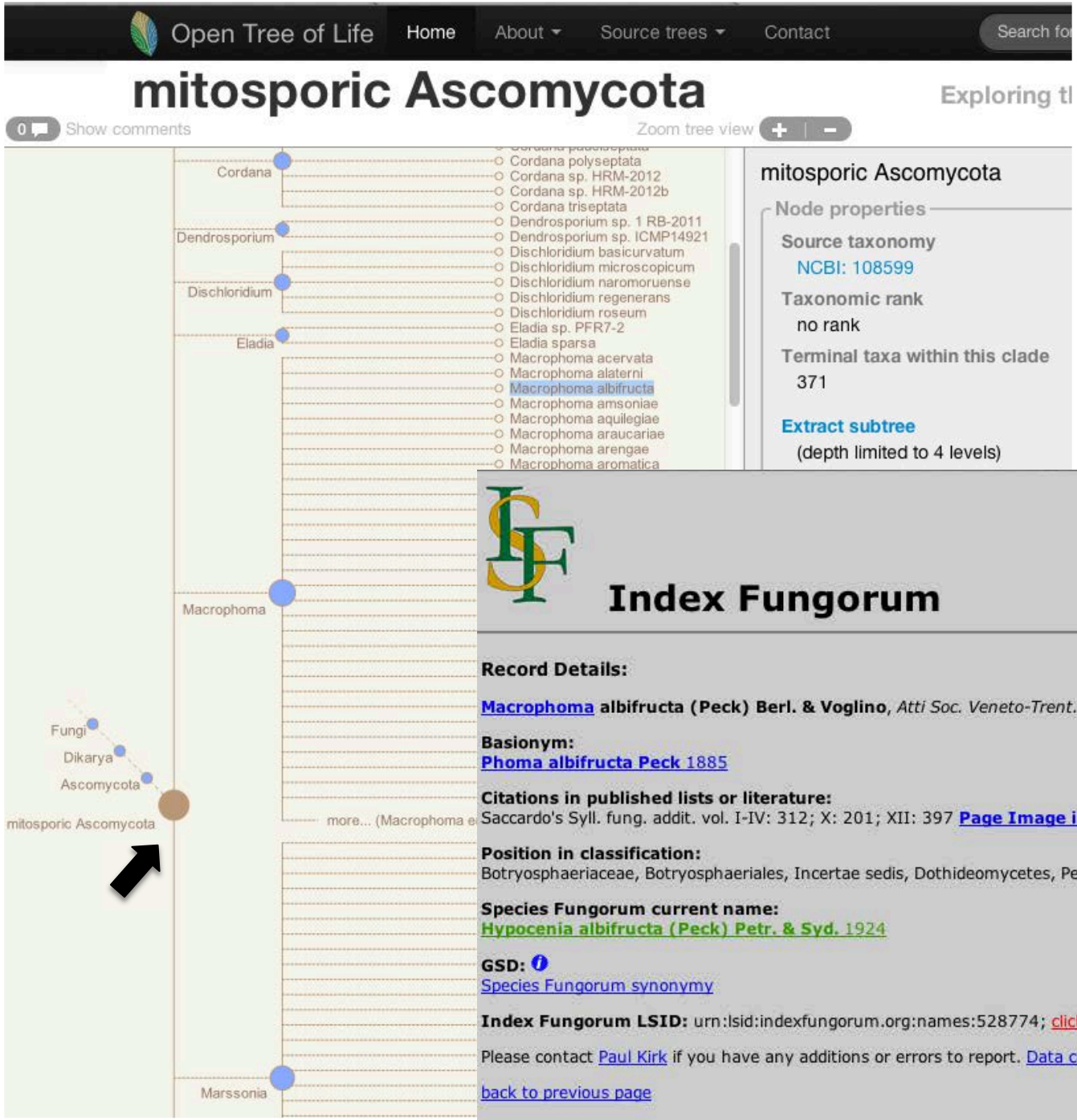
Lost Branches in the Tree of Life. 2013. Plos Biology

Synthetic tree – Fungi

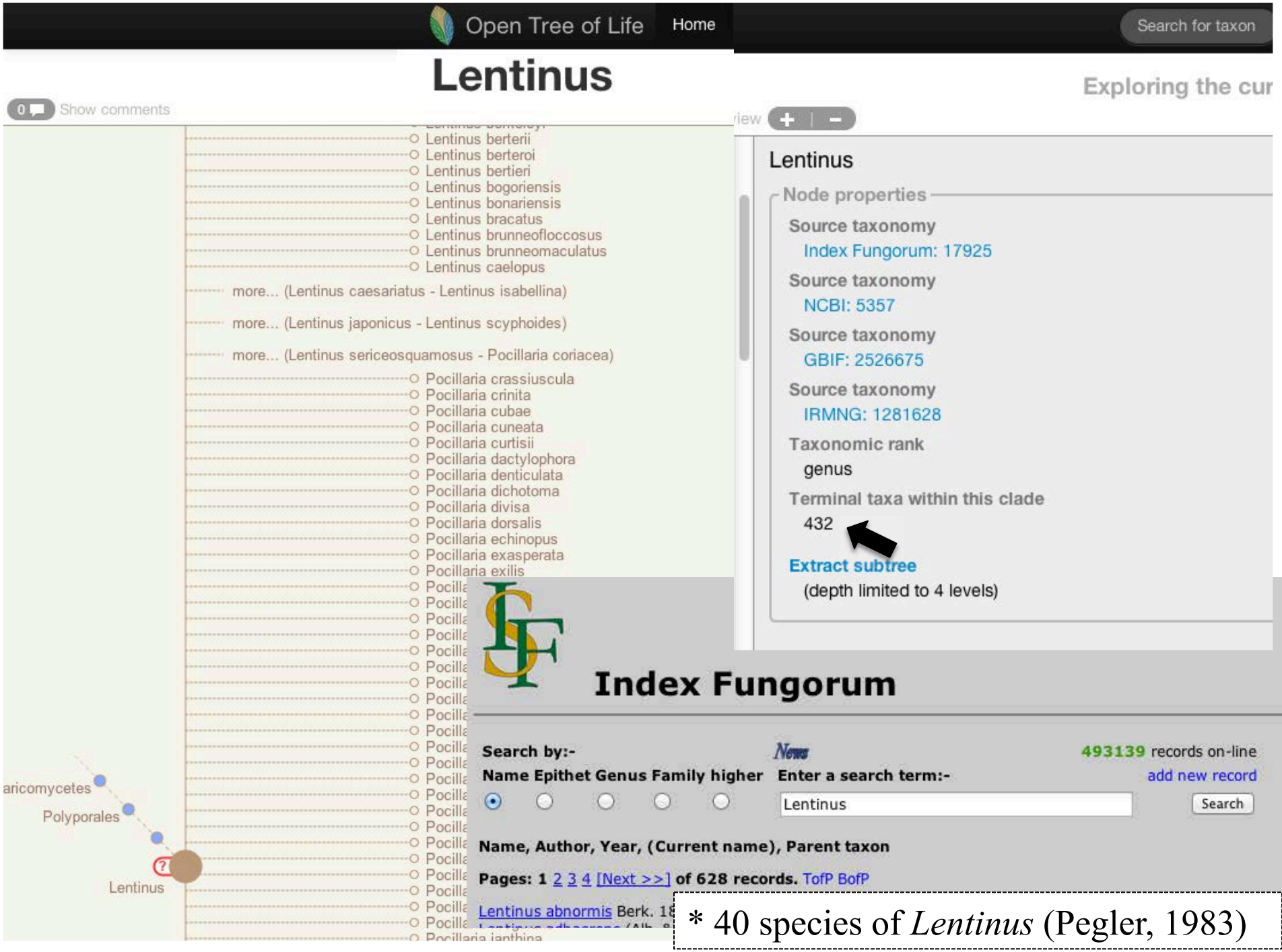
Dikarya




Open Tree of Life
Home
Search for taxon
in
All life
Dikarya
0
Show comments
Exploring the current syntheti
Basidiomycota
Node properties
Source taxonomy
Hibbett et al. 2007 updated: 13
Source taxonomy
Index Fungorum: 90050
Source taxonomy
NCBI: 5204
Source taxonomy
GBIF: 34
Source taxonomy
IRMNG: 236
Taxonomic rank
phylum
Terminal taxa within this clade
20
Extract subtree
(depth limited to 4 levels)
Leotiomyceta
Orbiliomycetes
Pezizomycetes
leotiomyceta
Pezizomycotina
Saccharomycetes
Podocapsa
Saccharomycotina
Archaeorhizomycetes
Neolectomycetes
Pneumocystidomycete
Schizosaccaromycetes
Taphrinomycotina
Taphrinomycetes
Acremonium
Cordana
Ascomycota
Dendrosporium
Dischloridium
Eladia
Macrophoma
Marssonia
Nakataea
Phialocephala
Pseudofusarium
mitosporic Ascomycota
Opisthokonta
Nucletmycea
Fungi
Dikarya
Agaricomycotina
Basidiomycota
Pucciniomycotina
Ustilaginomycotina
Chlamydotomus
mitosporic Basidiomycota
Edge to parent Dikarya
Supported by
Sekimoto, Satoshi, D'Ann Rochon, Jennifer E. Long, Jaclyn M. Dee, and Mary L. Berbee. "A multigene phylogeny of Olpidium and its implications for early fungal evolution." BMC Evolutionary Biology 11, no. 1 (2011): 331.
Full publication: http://dx.doi.org/10.1186/1471-2148-11-331
Open Tree curation: Study 1197
Supported by
Taxonomy
Supported by
Gazis, Romina; Hibbett, David (2014): Fungi Classification Update 2014. figshare. http://dx.doi.org/10.6084/m9.figshare.915439
Open Tree curation: Study 2818
Supported by
Floudas, Dimitrios, Manfred Binder, Robert Riley, Kerrie Barry, Robert A. Blanchette, Bernard Henrissat, Angel T. Martínez et al. "The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes." Science 336, no. 6089 (2012): 1715-1719.
Open Tree curation: Study 2701
Supported by
James, Timothy Y., Frank Kauff, Conrad L. Schoch, P. Brandon Matheny, Valérie Hofstetter, Cymon J. Cox, Gail Celio et al. "Reconstructing the early evolution of Fungi using a six-gene phylogeny." Nature 443, no. 7113 (2006): 818-822.
Full publication: http://dx.doi.org/10.1038/nature05110
Open Tree curation: Study 1162

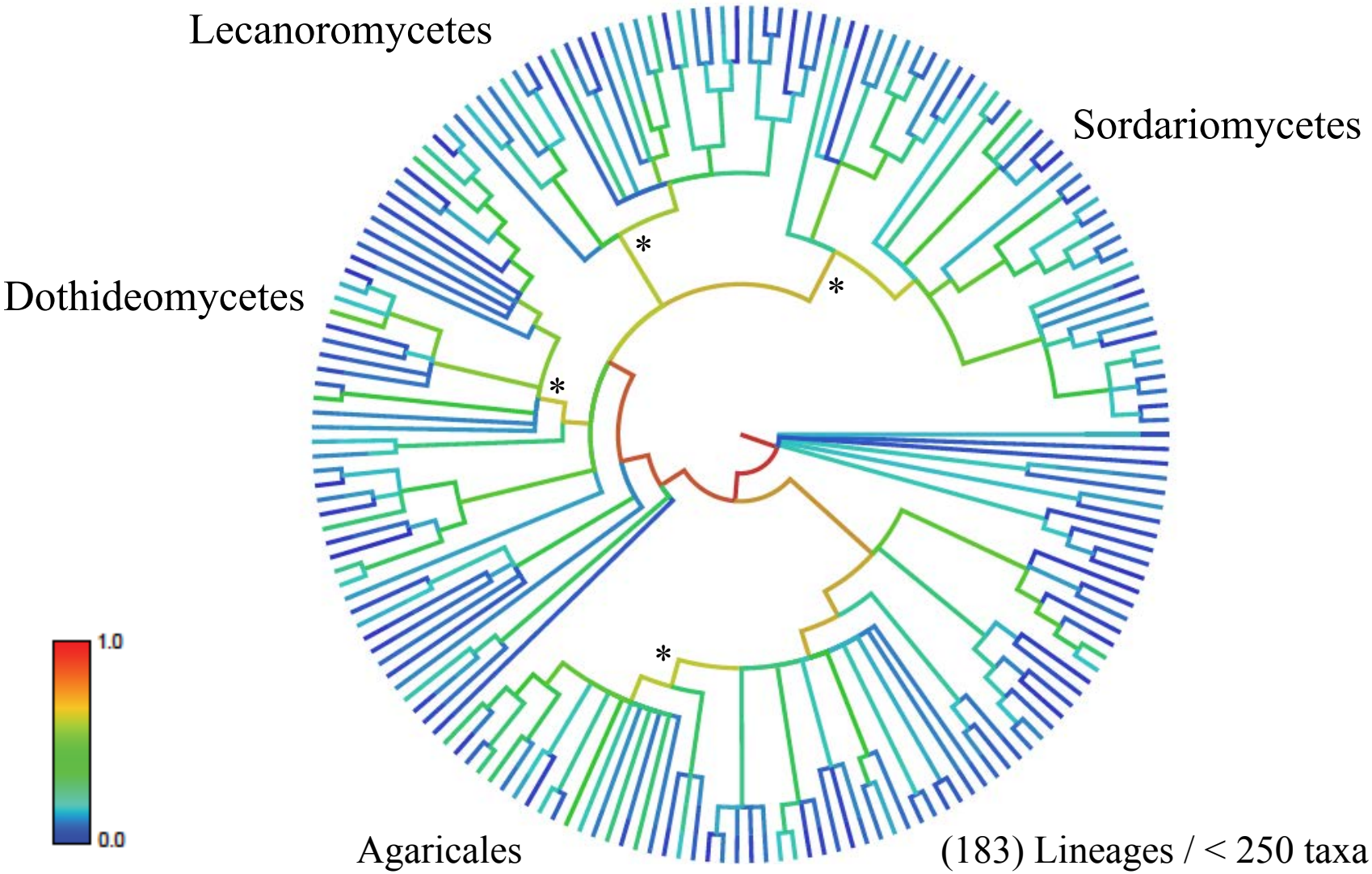


Inconsistencies among public taxonomic databases i.e., NCBI / IF



* 40 species of *Lentinus* (Pegler, 1983)

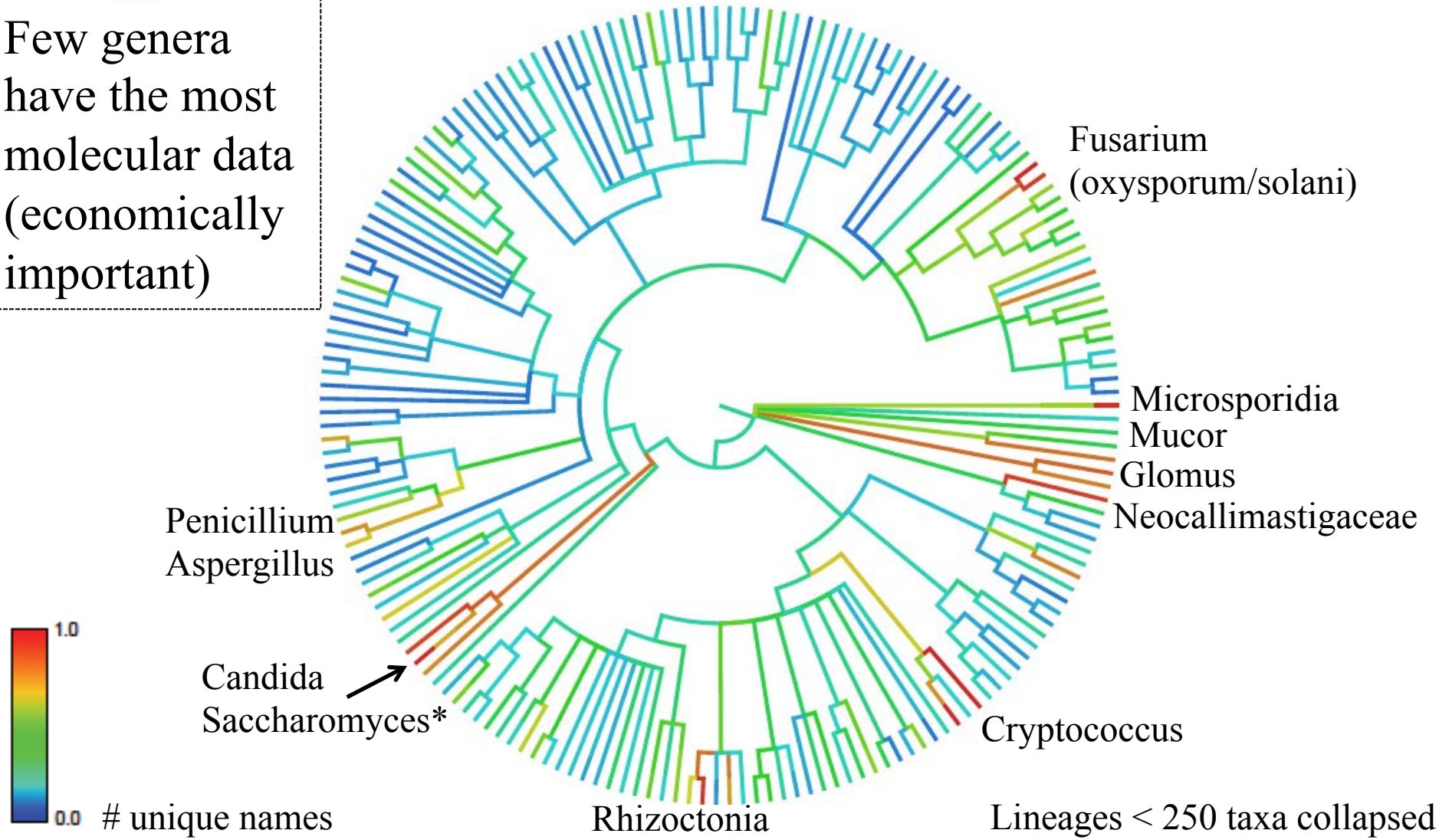
Some metrics for Fungi:
Diversity in the tree (number of taxa)



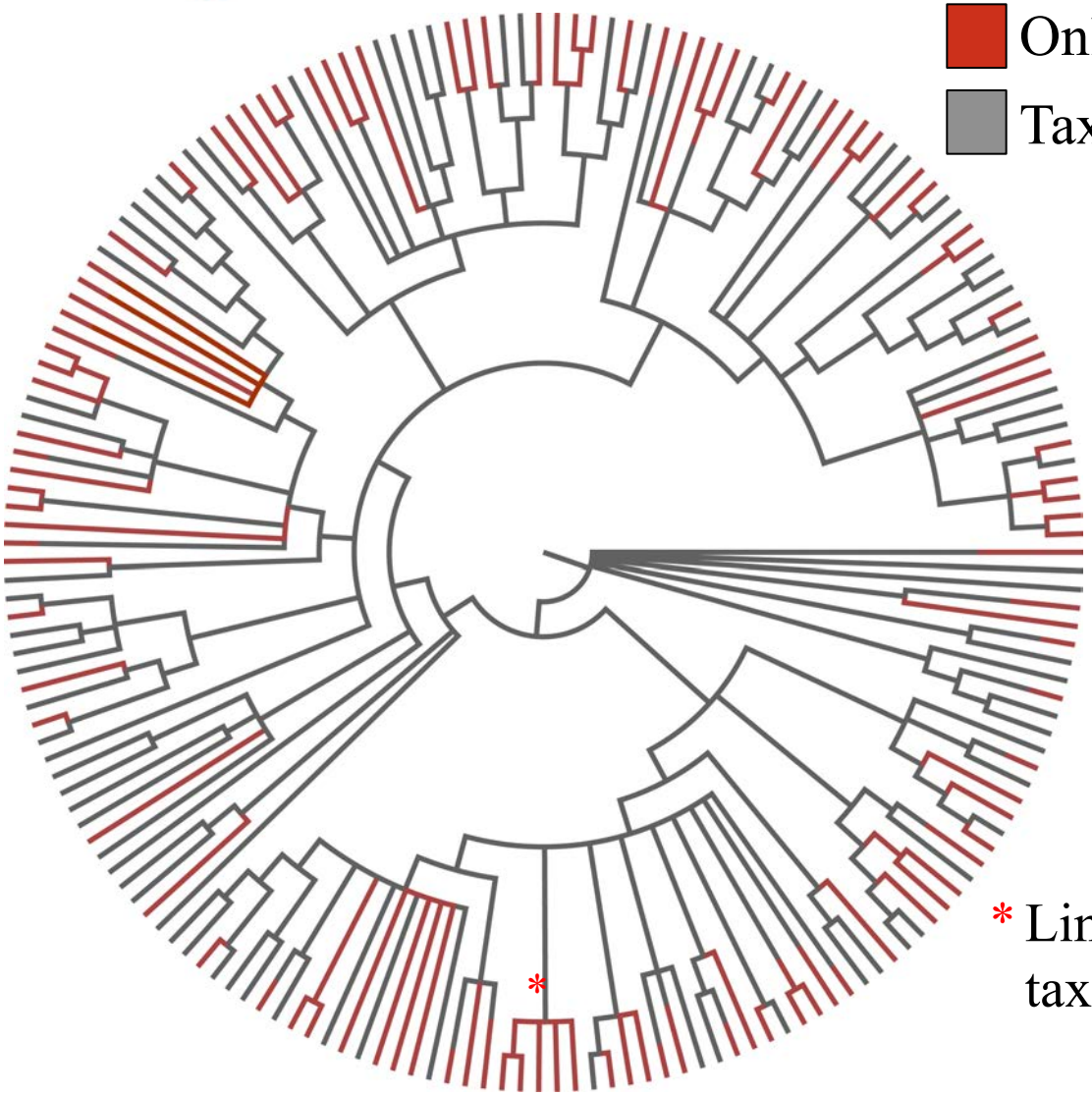
(183) Lineages / < 250 taxa collapsed

Some metrics for Fungi:
Diversity in NCBI (number of taxa represented)

Few genera have the most molecular data (economically important)



Some metrics for Fungi:
Need to add more phylogenies….



Only taxonomy

Taxonomy + @least one source tree

In v. 1 only 12 source trees in synthesis
In v.2 ~60 source trees in synthesis
Many more will be added to order/family/genera regions of the tree

* Lineages within Cantharellales only taxonomy

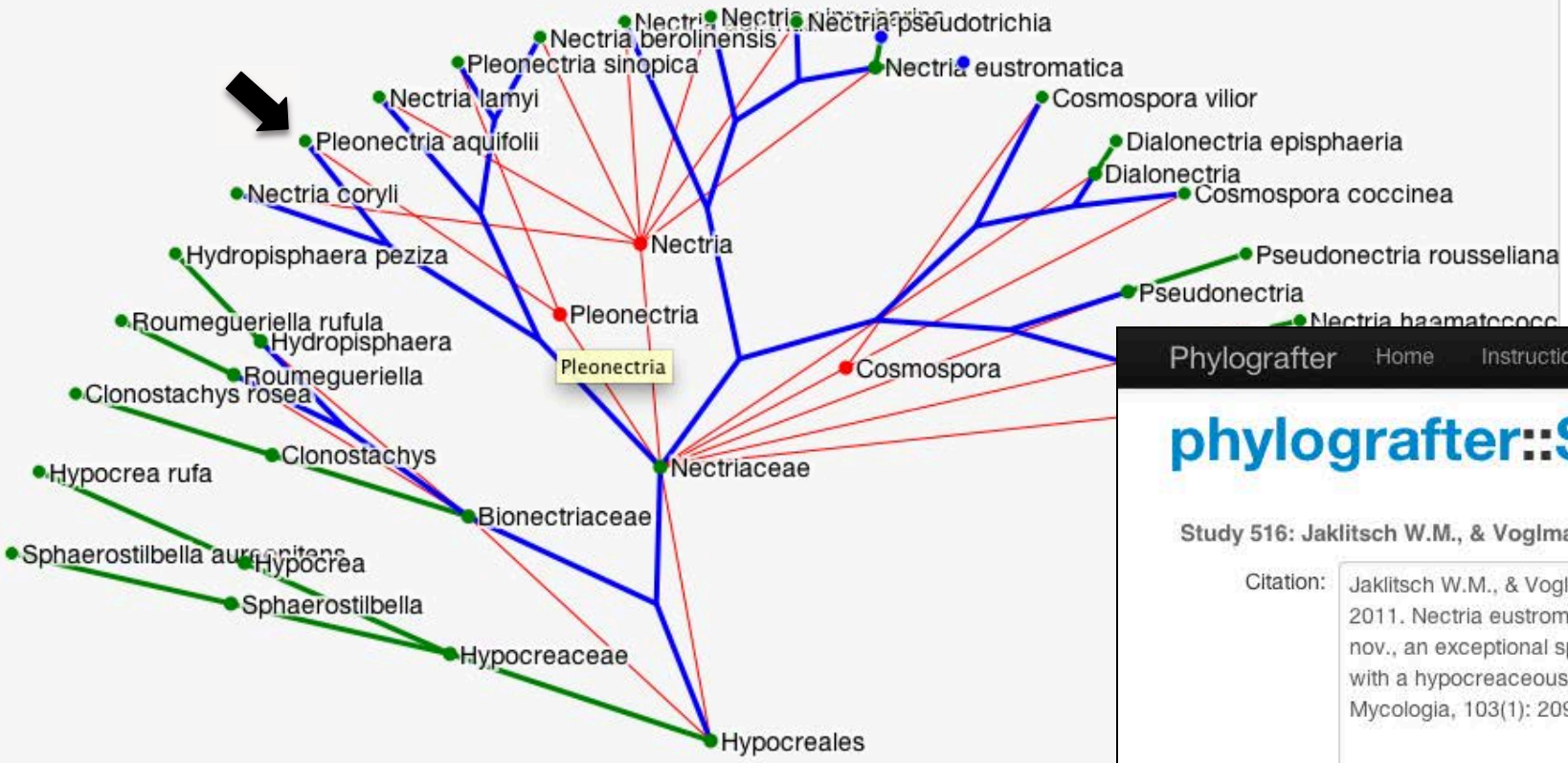
Lineages < 250 taxa collapsed

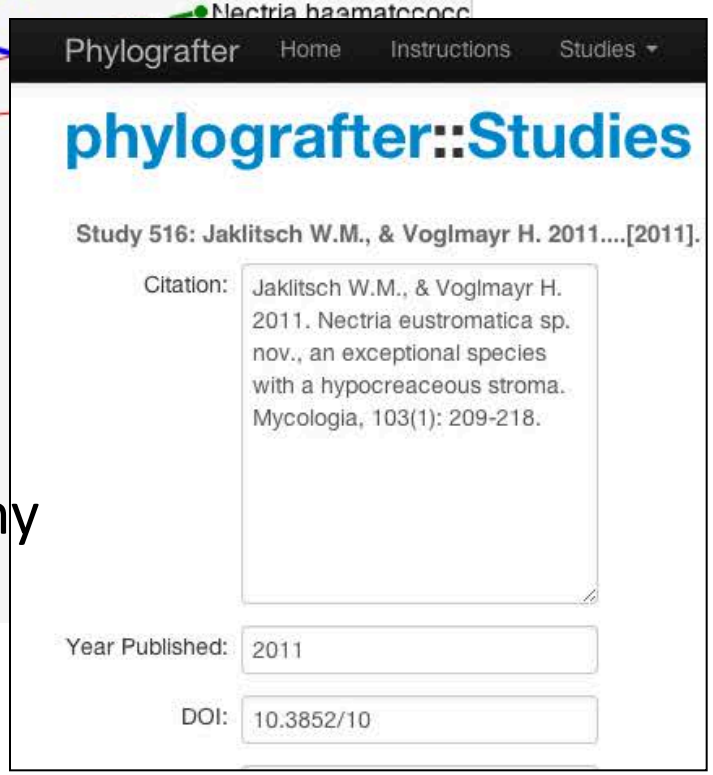
Coming soon… year 3

- Metrics and notification for curators
  (subscribe to your favorite clade!)
  → promote community curation and collaboration among working groups
- Visualize conflict and overlap among/between sources tree and taxonomy
  → help in taxonomic revisions

phylografter::OTT TreeGraph of source tree 728

Taxonomy should reflect phylogeny

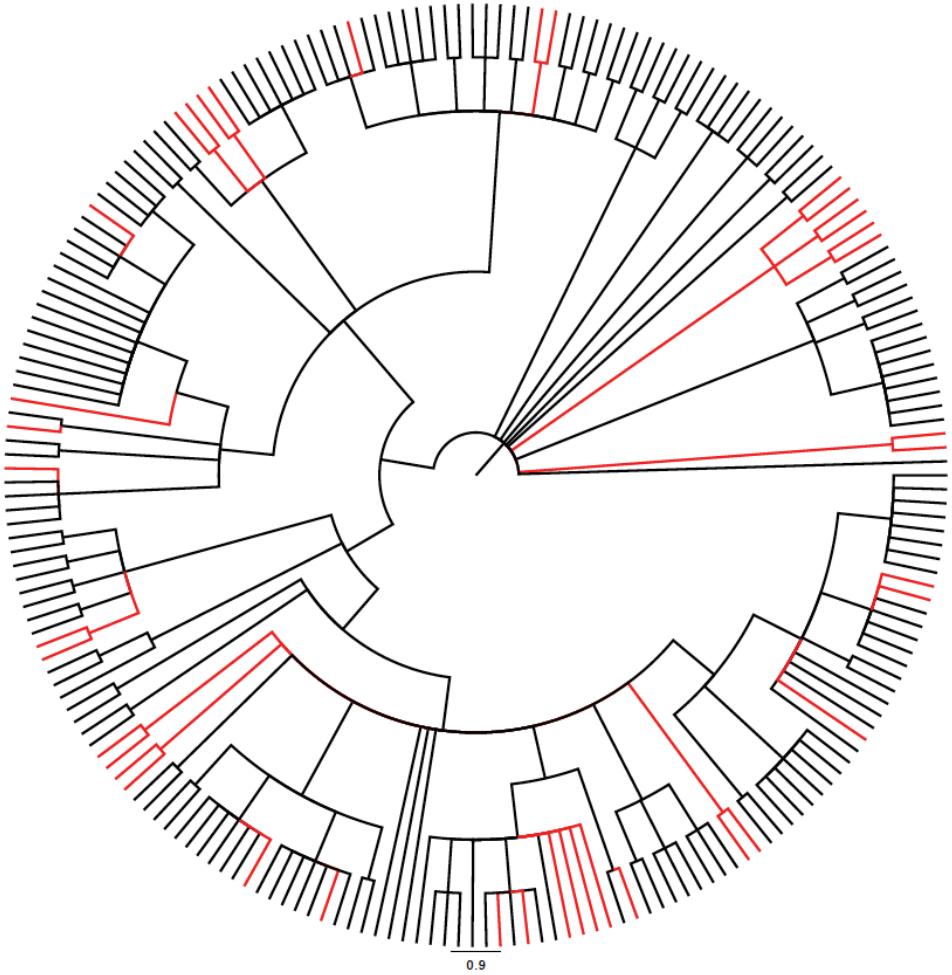




- Relationships not supported by taxonomy
- Relationships supported by taxonomy
- Taxonomy OTT v. 2.6
- Non-monophyletic taxa

Fungal Classification Update

- 180 emails to authors of studies that have produced comprehensive phylogenies/ described new high-rank taxa
- Response: 40% positive – will contribute with tree-files. THANKS!
- Fungal classification update (~2015)



Additions* to the Hibbett et al. 2007 (~33 orders, 10 classes, 2 phylum)

*include taxonomic novelties and rearrangements

THANKS!

*Special thanks to Paul Kirk (Index Fungorum)

OPEN Tree of Life

NSF