

Large scale sequencing of Dothideomycetes provides insights into genome evolution and adaptation

Sajeet Haridas¹, Pedro Crous², Manfred Binder², Joseph Spatafora³, Igor Grigoriev¹

¹DOE Joint Genome Institute, Walnut Creek, CA. USA.

²CBS Fungal Biodiversity Centre, Utrecht, The Netherlands

³Oregon State University, Corvallis, OR. USA.

March 2015

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

LBNL- 178328

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Large scale sequencing of Dothideomycetes provides insights into genome evolution and adaptation.

Sajeet Haridas^{1*}, Pedro Crous², Manfred Binder², Joseph Spatafora³, Igor Grigoriev¹

¹DOE Joint Genome Institute, Walnut Creek, CA, USA; ²CBS Fungal Biodiversity Centre, Utrecht, The Netherlands;

³Oregon State University, Corvallis, OR, USA.

*sharidas@lbl.gov



Abstract

Dothideomycetes is the largest and most diverse class of ascomycete fungi with 23 orders 110 families, 1300 genera and over 19,000 known species. We present comparative analysis of 70 Dothideomycete genomes including over 50 that we sequenced and are as yet unpublished. This extensive sampling has almost quadrupled the previous study of 18 species and uncovered a 10 fold range of genome sizes. We were able to clarify the phylogenetic positions of several species whose origins were unclear in previous morphological and sequence comparison studies. We analyzed selected gene families including proteases, transporters and small secreted proteins and show that major differences in gene content is influenced by speciation.

Significance

Fungi in class Dothideomycetes show a high level of ecological diversity including many plant pathogens infecting a broad range of hosts, some of which are involved in the production of food, feed, fiber and biofuel.

We present the genome characteristics of 50 newly sequenced Dothideomycetes along with the genomes of several other previously published Dothideomycetes genomes.

Much of this data was generated as part of the 1000 Fungal Genomes Project - A collaborative project that engages the broad mycological community and aims to sequence unsampled fungal diversity.

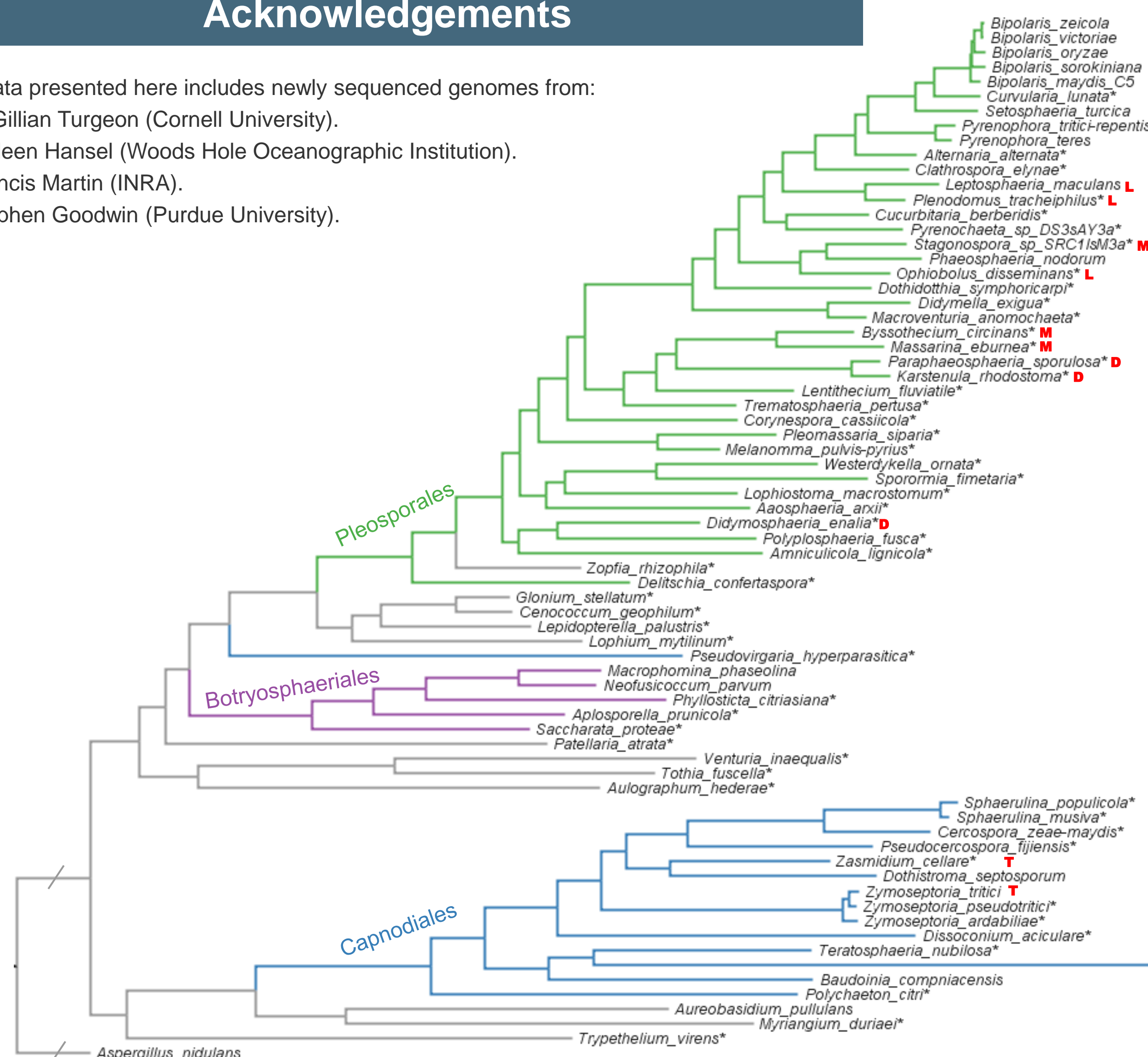
Conclusions

- Repeat content varies widely and is unrelated to lifestyle or phylogenetic position.
- The newly proposed phylogeny is supported by genome synteny.
- An expanded species specific gene set is an indicator of insufficient sampling within the clade.
- Most differences are organism specific rather than clade dependent suggesting niche specific speciation.
- These 70 genomes and 400 more are available on MycoCosm (jgi.doe.gov/fungi).

Acknowledgements

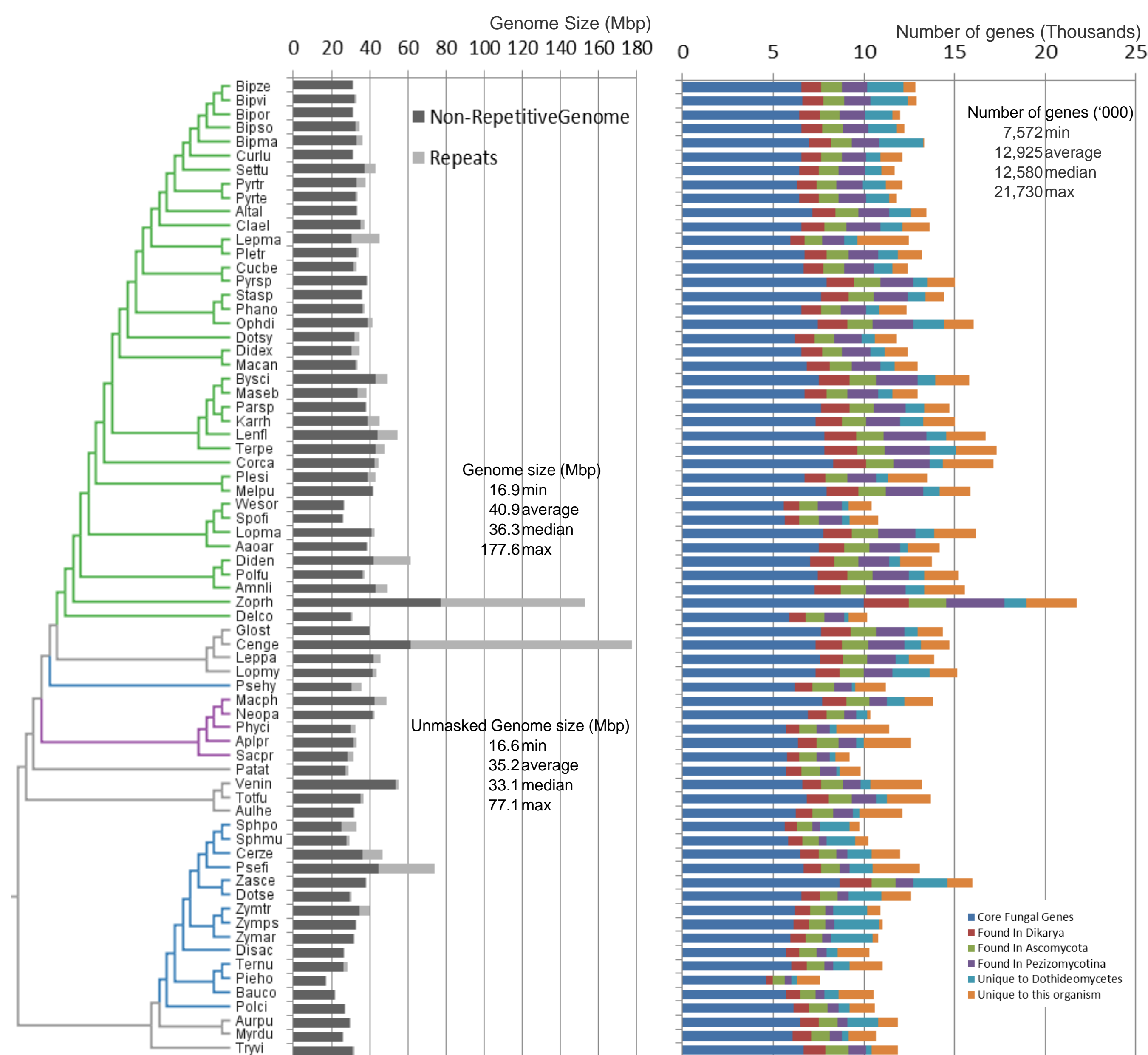
The data presented here includes newly sequenced genomes from:

- B. Gillian Turgeon (Cornell University).
- Colleen Hansel (Woods Hole Oceanographic Institution).
- Francis Martin (INRA).
- Stephen Goodwin (Purdue University).



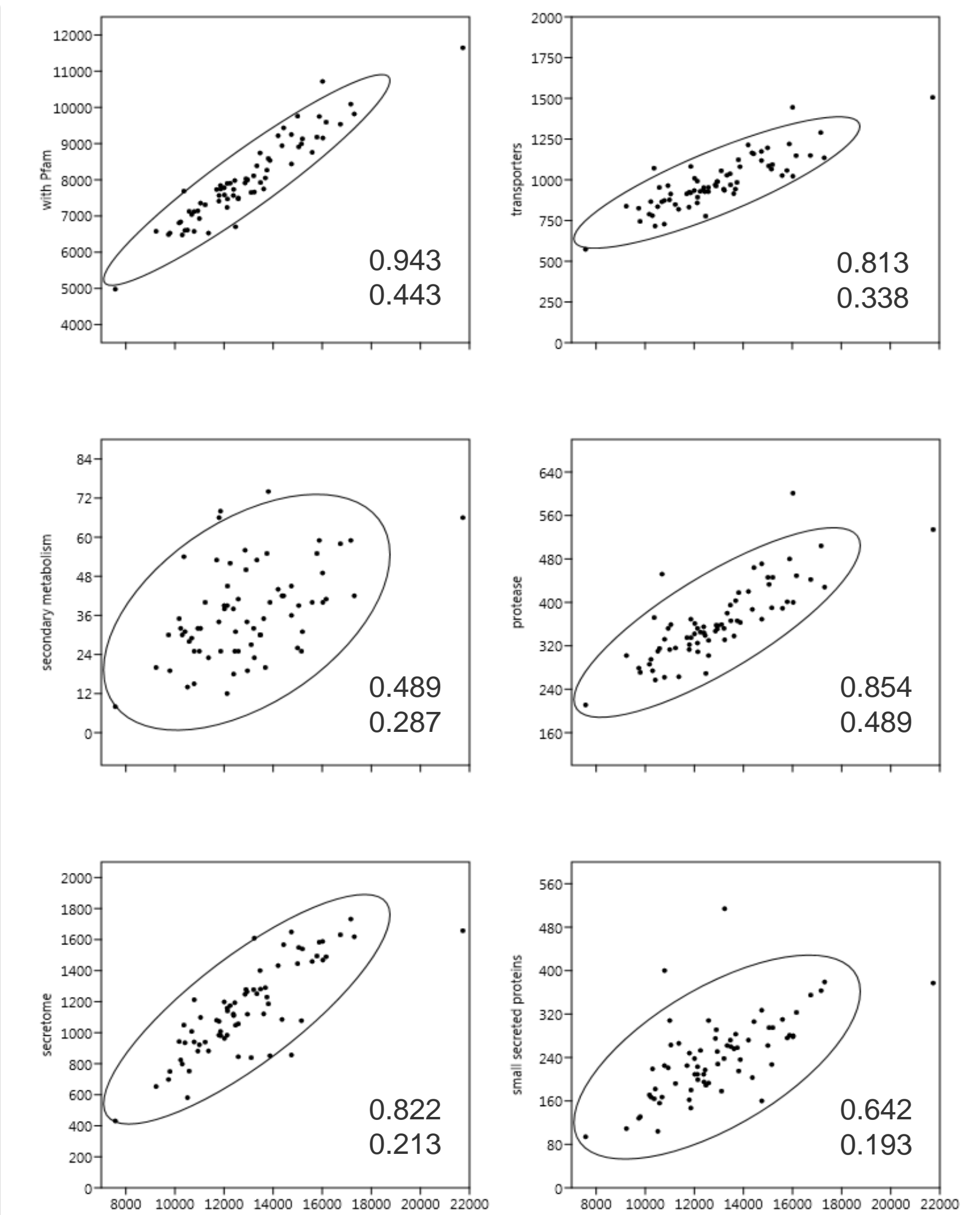
A maximum likely inference of the phylogenetic position of 70 Dothideomycetes genomes with *Aspergillus nidulans* as the outgroup. The tree was constructed using a concatenated list of 300 orthologous proteins using RAxML. While the orders are mostly well separated, there are several families with members in different clades shown with red letters representing **L**eptosphaeriaceae, **M**assariniaceae, **D**idymosphaeriaceae and **T**eratosphaeriaceae.

*represents newly sequenced (previously unpublished) genomes.

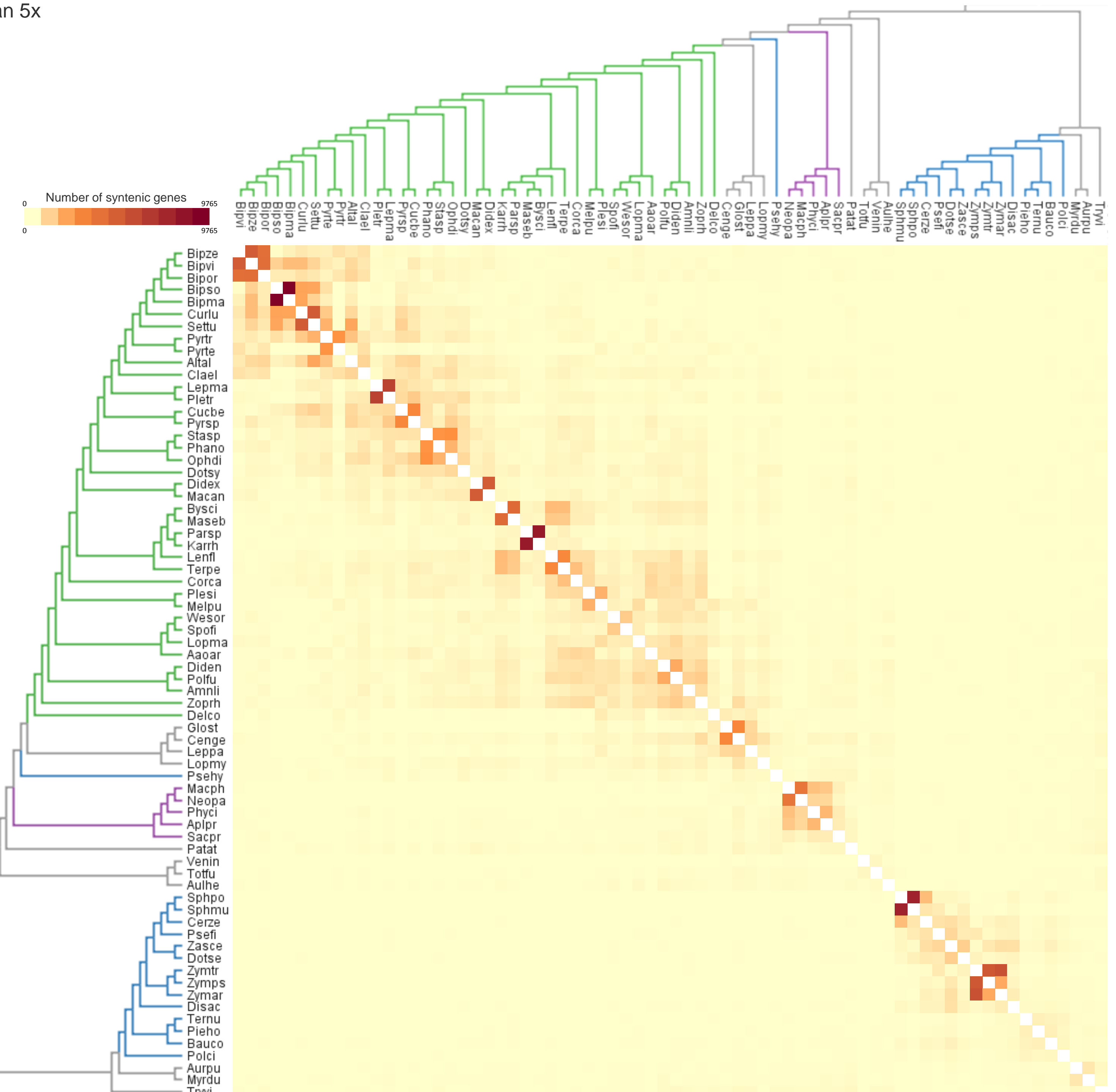


Genome Sizes and Repeat Content. While the assembled genomes show a 10x range of sizes, the un-masked genomes show less than 5x range.

Gene content: Gene clusters representing potential orthologous genes from 450 species across the fungal tree were used to identify genes belonging to the major taxonomic clades.



Annotation characteristics of Dothideomycetes genomes: The numbers represent the Linear r (Pearson) correlation with the number of genes and genome size respectively. Ellipses represent the 95% confidence interval.



A count of syntenic genes supports the proposed phylogeny. We used OrthoMCL to identify orthologous genes and iADHoRe to identify syntenic clusters (minimum cluster size = 5). The number of genes in syntenic clusters between two genomes is an indirect measure of relatedness with closely related genomes showing a higher number of syntenic genes. This measure shows that the abnormally placed genomes such as *Pseudovirgaria* and *Piedraia* indeed belong as shown on the phylogenetic tree.

The genome short names are made of the first three letters of the genus and the first two of the corresponding species names.