

Comparative genome analysis of Basidiomycete fungi

Robert Riley¹, Asaf Salamov¹, Bernard Henrissat², Laszlo Nagy³, Daren Brown⁴, Benjamin Held⁵, Scott Baker⁶, Robert Blanchette⁷, Bastien Boussau⁸, Sharon L. Doty⁹, Kirsten Fagnan¹, Dimitris Floudas³, Anthony Levasseur², Gerard Manning¹⁰, Francis Martin¹¹, Emmanuelle Morin¹¹, Robert Otillar¹, Antonio Pisabarro¹², Jonathan Walton¹³, Ken Wolfe¹⁴, David Hibbett³, Igor Grigoriev¹

¹DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598

²Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, CNRS, Marseille cedex 9, France

³Department of Biology, Clark University, 950 Main Street, Worcester, MA 01610

⁴USDA, Peoria, IL, 61604

⁵University of Minnesota 495, Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108;

⁶Pacific Northwest National Laboratory, PO Box 999 Richland, WA 99352

⁷University of Minnesota, 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108;

⁸Department of Integrative Biology, University of California, Berkeley, CA 94720-3140;

⁹School of Environmental and Forest Sciences; College of the Environment, University of Washington, Seattle, WA 98195-2100

¹⁰Razavi-Newman Center for Bioinformatics, Salk Institute for Biological Sciences, San Diego, CA 92186

¹¹Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1136 Université Henri Poincaré, Interactions Arbres/Micro-organismes, 54280 Champenoux, France

¹²Depto. de Producción Agraria, Univ. Pública de Navarra, 31006 Pamplona, Spain

¹³DOE Plant Research Lab, 612 Wilson Road, Room 210, Michigan State University, E. Lansing MI 48824

¹⁴UCD School of Medicine & Medical Science, Conway Institute, University College Dublin, Dublin 4, Ireland

August 2013

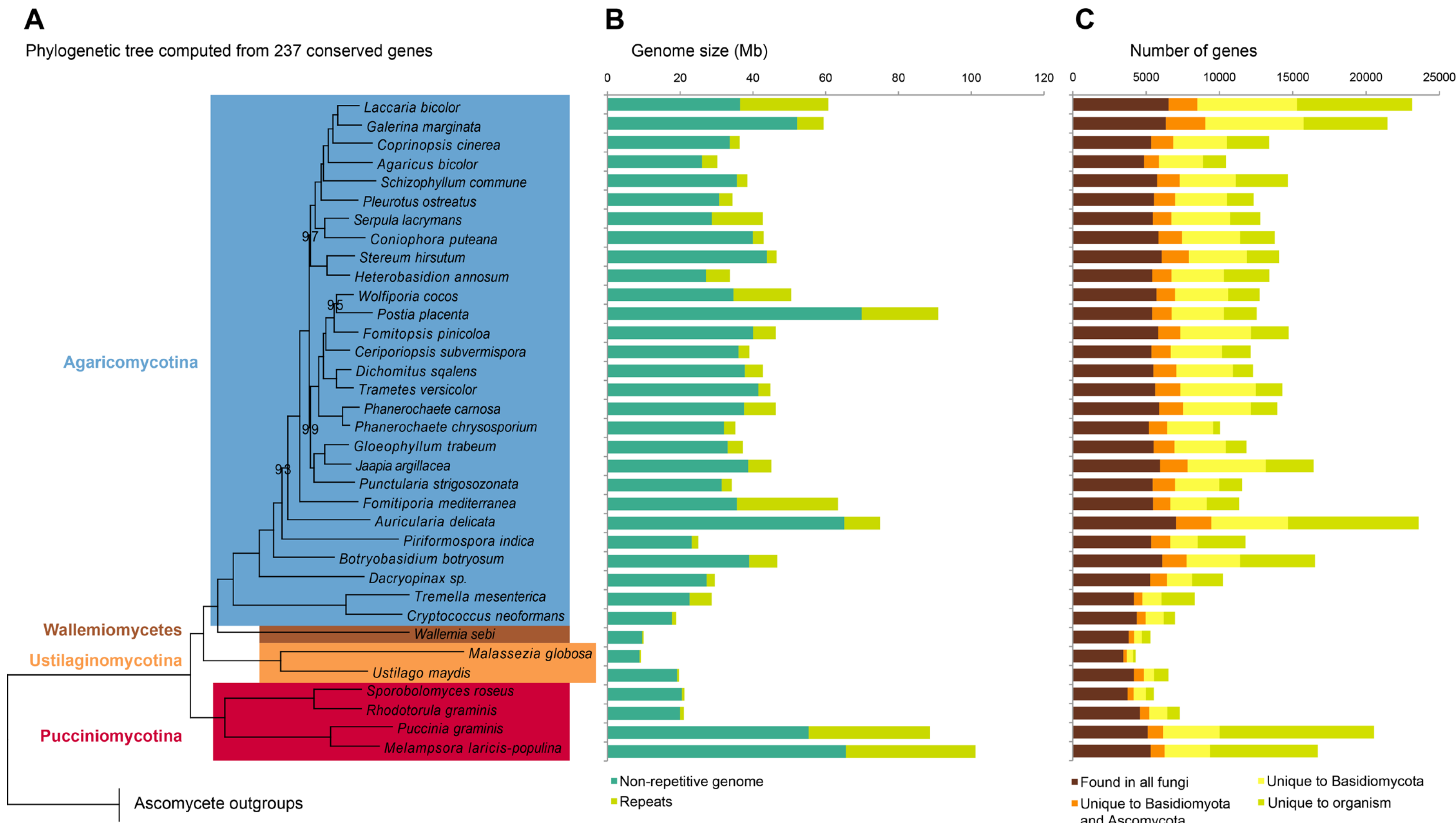
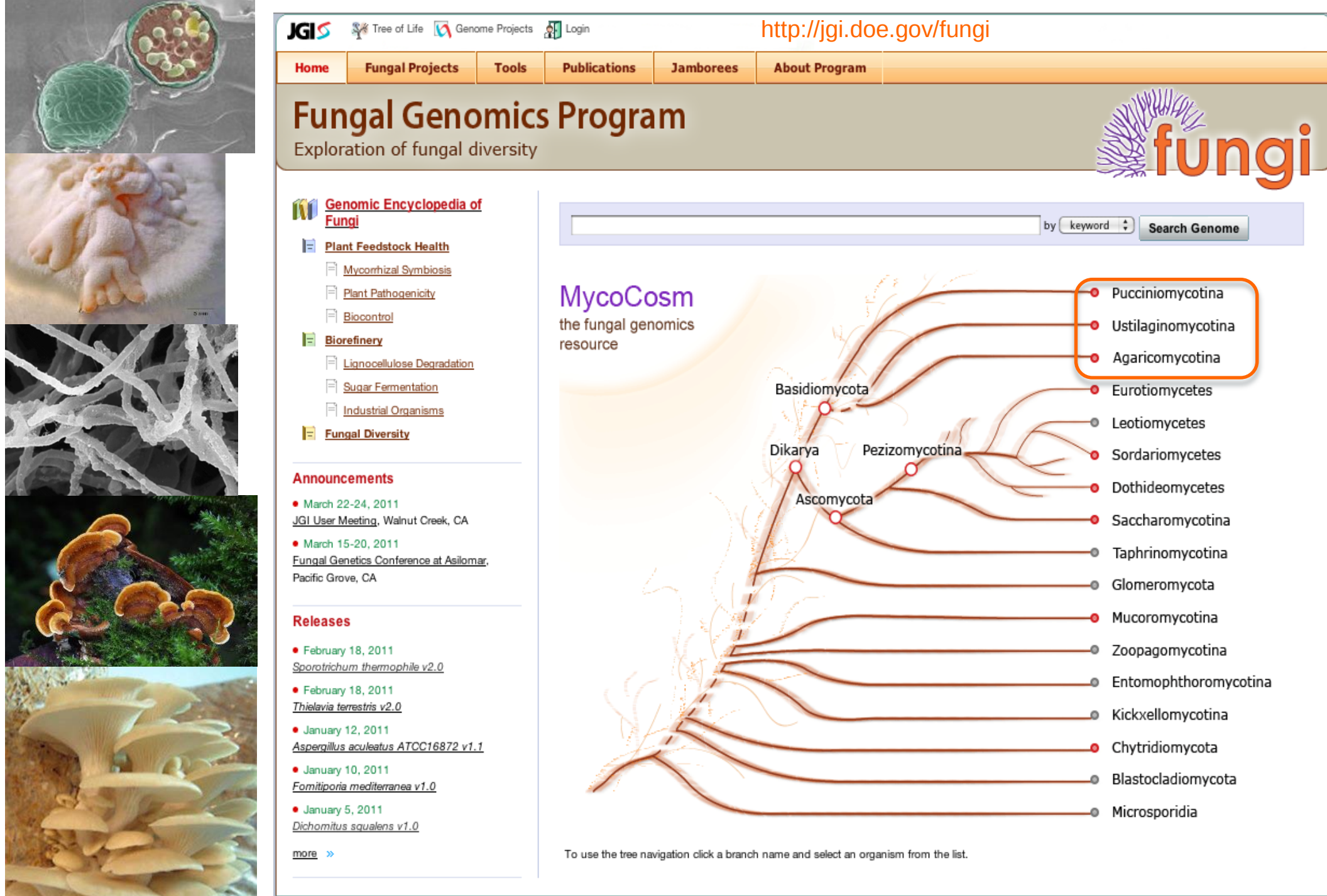
The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

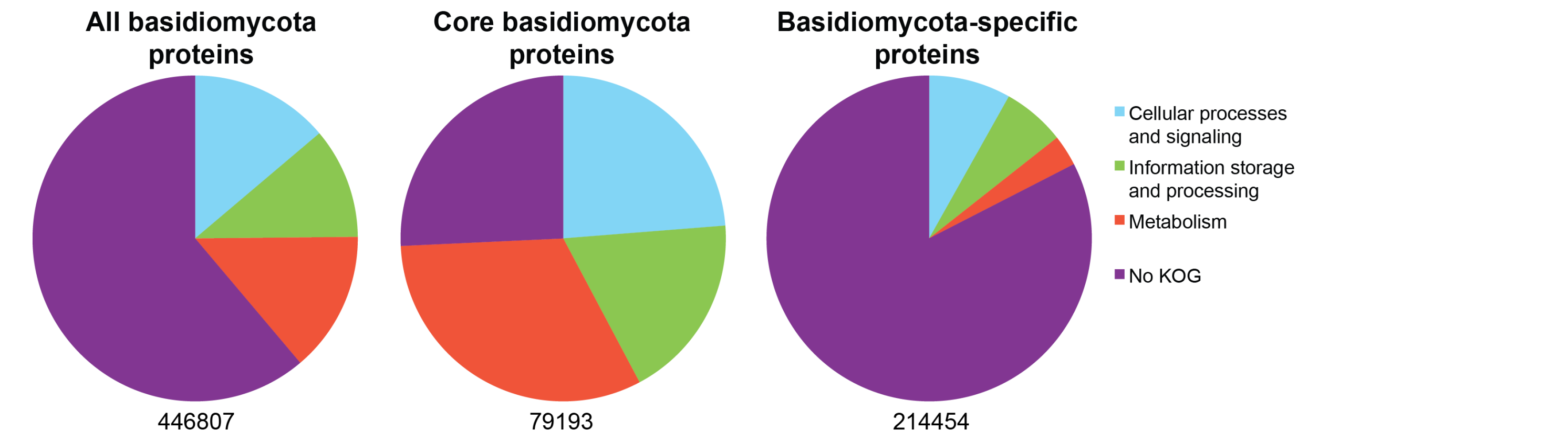
This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any 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.

Abstract

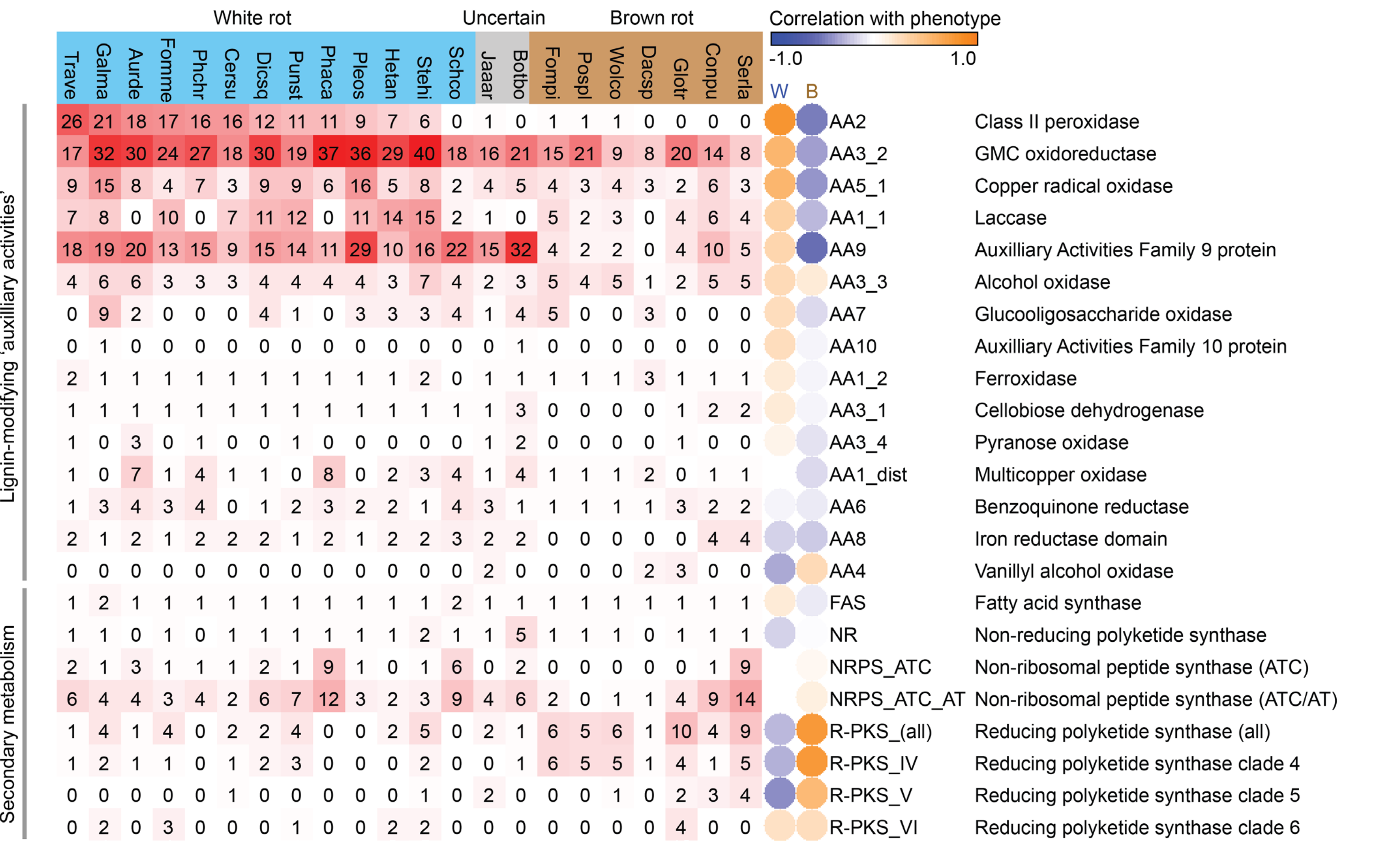
Fungi of the phylum Basidiomycota (basidiomycetes), make up some 37% of the described fungi, and are important in forestry, agriculture, medicine, and bioenergy. This diverse phylum includes symbionts, pathogens, and saprotrophs including the majority of wood decaying and ectomycorrhizal species. To better understand the genetic diversity of this phylum we compared the genomes of 35 basidiomycetes including 6 newly sequenced genomes. These genomes span extremes of genome size, gene number, and repeat content. Analysis of core genes reveals that some 48% of basidiomycete proteins are unique to the phylum with nearly half of those (22%) found in only one organism. Correlations between lifestyle and certain gene families are evident. Phylogenetic patterns of plant biomass-degrading genes in Agaricomycotina suggest a continuum rather than a dichotomy between the white rot and brown rot modes of wood decay. Based on phylogenetically-informed PCA analysis of wood decay genes, we predict that that *Botryobasidium botryosum* and *Jaapia argillacea* have properties similar to white rot species, although neither has typical ligninolytic class II fungal peroxidases (PODs). This prediction is supported by growth assays in which both fungi exhibit wood decay with white rot-like characteristics. Based on this, we suggest that the white/brown rot dichotomy may be inadequate to describe the full range of wood decaying fungi. Analysis of the rate of discovery of proteins with no or few homologs suggests the value of continued sequencing of basidiomycete fungi.



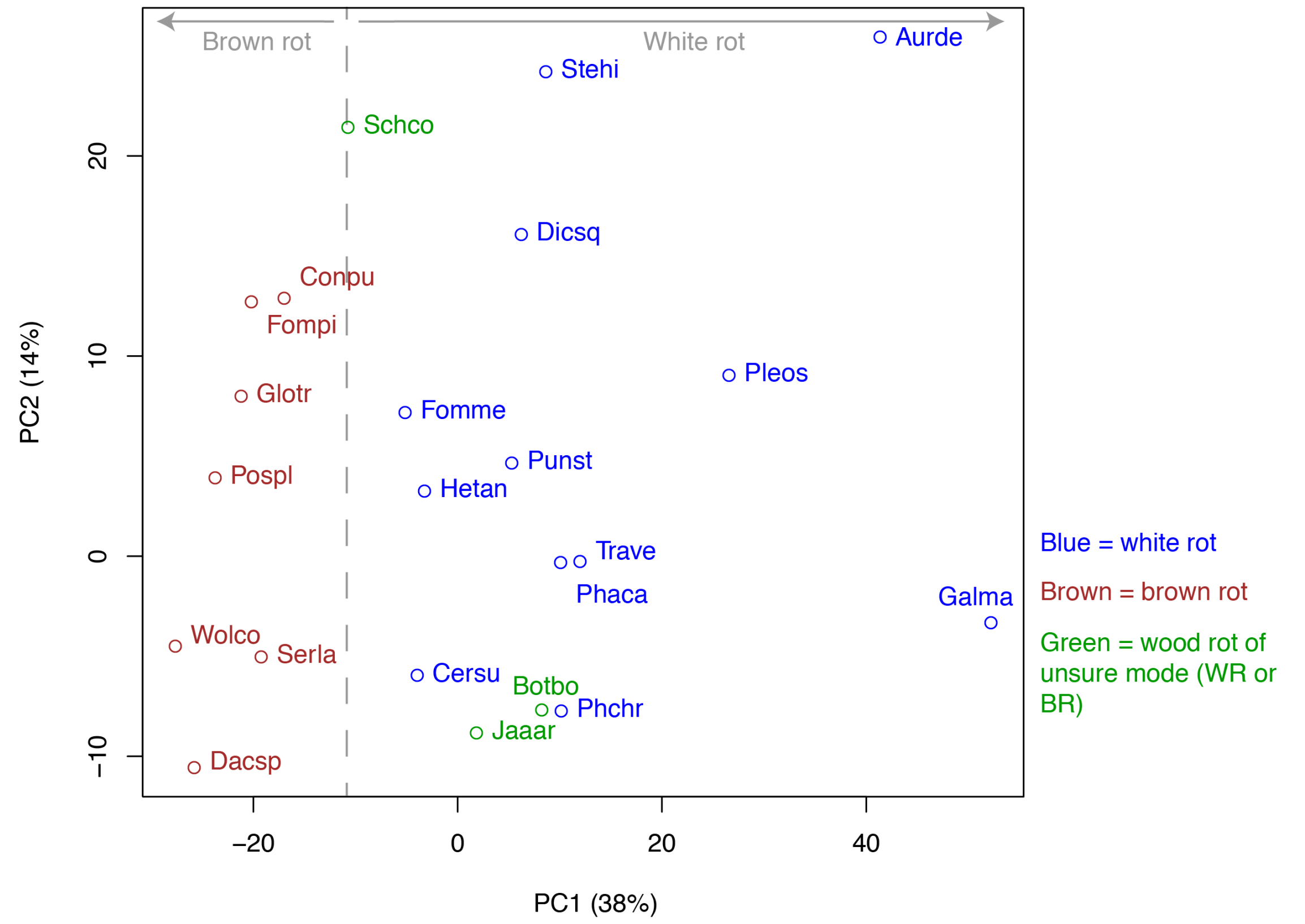
Phylogeny, genome size and repeats, gene conservation. **A.** Maximum likelihood tree of 35 basidiomycetes based on concatenated alignments of 237 widely conserved genes. Ascomycete outgroups are omitted from figure. Bootstrap values of branches are 100% except where indicated. **B.** Repeat content in basidiomycetes is highly variant, ranging from 1% to 44%. **C.** About 42% of basidiomycete genes are conserved in all published fungi. Some 10% are unique to Basidiomycota and Ascomycota; 26% are unique to basidiomycetes, and another 22% are organism-specific within basidiomycetes, which, while biased by taxon sampling, highlights the novel functions yet to be uncovered in the phylum.



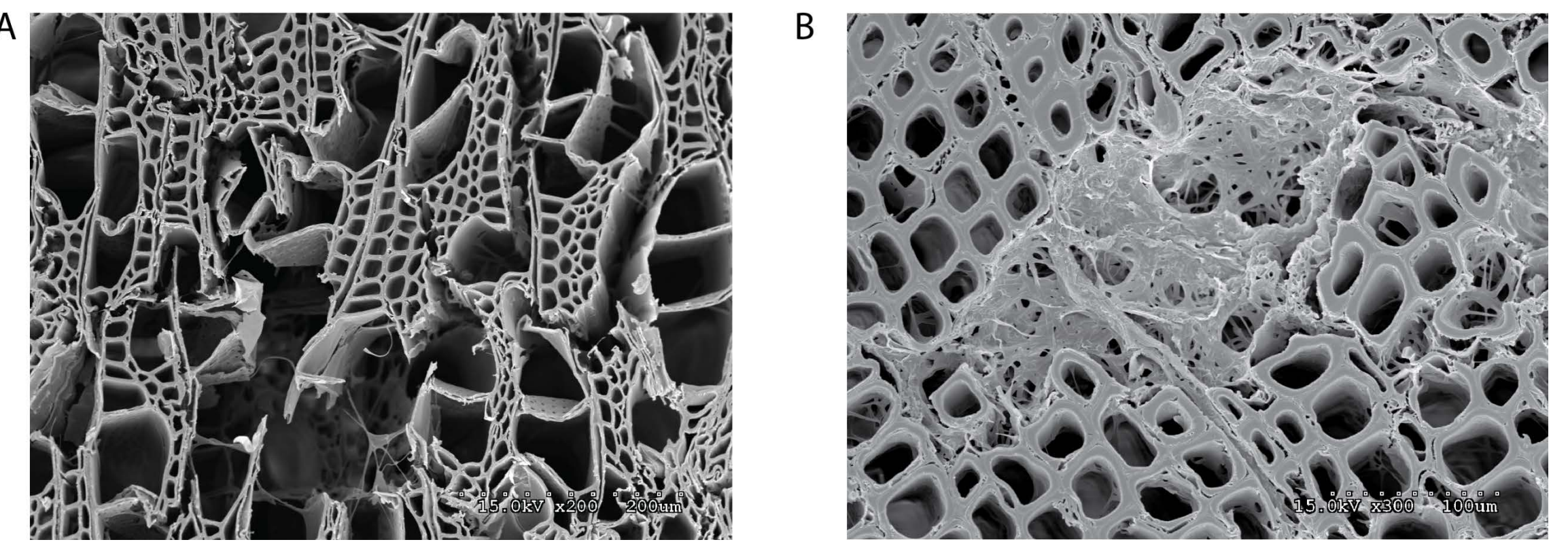
Core genes of Basidiomycota. Notice that half of basidiomycete proteins have no KOG annotation (function unknown); 92% of core basidiomycete proteins have a KOG annotation (putative function predicted) and that 78% of non-core basidiomycete proteins have no KOG annotation (function unknown)



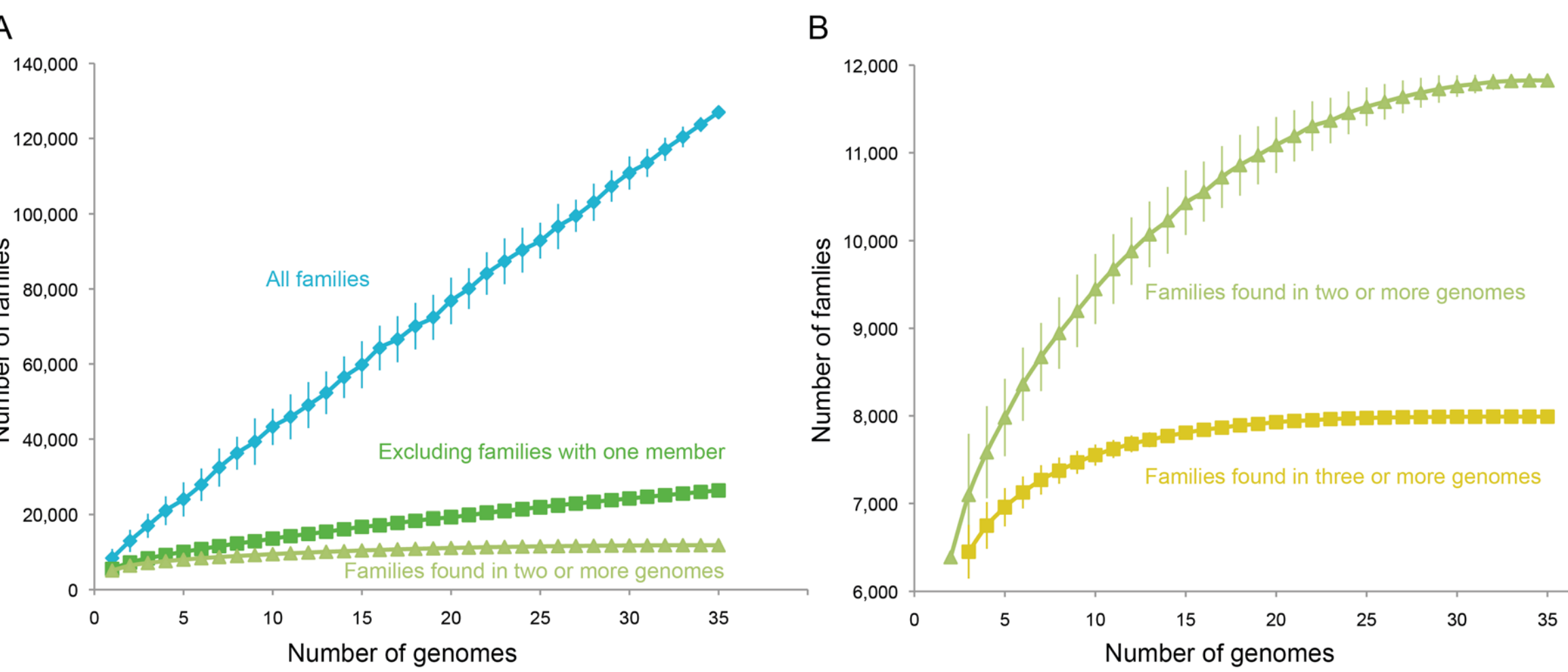
Lignin-degrading and secondary metabolism in wood-decaying fungi. Gene copy number is shaded red/white and independent contrasts correlation(36) of enzyme with rot type shaded orange/blue. Notice the white/brown rot dichotomy with respect to the lignin peroxidase-containing category AA2 (containing the PODs), and blurred line between white/brown rots with the other enzymes. *B. botryosum* and *J. argillacea* lack PODs but possess other enzymes potentially involved in lignin breakdown. The distribution of secondary metabolism enzymes suggests that the reducing polyketide synthase (R-PKS) enzymes are expanded in the brown rot fungi.



Wood-decaying basidiomycetes plotted on the first two principal components from phylogenetic PCA of CAZymes (including lignin-related auxiliary activities) of the organisms. Note the separation of brown rot fungi (brown), white rot fungi (blue) on PC1, which explains 38% of the variability. *B. botryosum* and *J. argillacea* group with *C. subvermispora* and *P. chrysosporium*, suggesting possible similarities to these fungi known to produce white rot. *S. commune*, previously characterized as able to produce white rot, is midway between white and brown rot. Though the brown rot fungi are all together on PC1, the polypores *F. pinicola*, *P. placenta*, and *W. cocos*, are spread out on PC2 (14% of variability), suggesting the power of phylogenetic PCA to deconvolute phylogenetic similarities of organisms.



Wood decay experiments indicating white rot-like mode of decay in *Botryobasidium botryosum* and *Jaapia argillacea*. **A** Micrograph of *B. botryosum* on aspen wood with vessel, fiber and parenchyma cell walls degraded. All cell wall layers are degraded resulting in small voids in the wood. Mycelia are visible growing through voids. **B** Micrograph of *J. argillacea* on pine showing an area within the wood where the fungus has caused a localized simultaneous decay of the cells. Residual cell wall material and hyphae fill the degraded zone.



Rate of protein family discovery from newly sequenced basidiomycete genomes. MCL clustering of protein sequences was used to identify protein families. For each number of genomes, samples of genomes of a given number were randomly generated, and the protein families were counted. Panel A illustrates that many new protein families are still being discovered from each new basidiomycete genome sequenced, and that the gains are more modest, but still increasing, for families with more than one member. Panel B indicates that most of the families conserved in multiple genomes have been discovered by sequencing efforts.