

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/320884315>

Multiple Approaches to Phylogenomic Reconstruction of the Fungal Kingdom

Chapter in *Advances in genetics* · November 2017

DOI: 10.1016/bs.adgen.2017.09.006

CITATIONS

5

READS

203

2 authors:



Charley McCarthy

University of Nottingham

65 PUBLICATIONS 50 CITATIONS

SEE PROFILE



David A Fitzpatrick

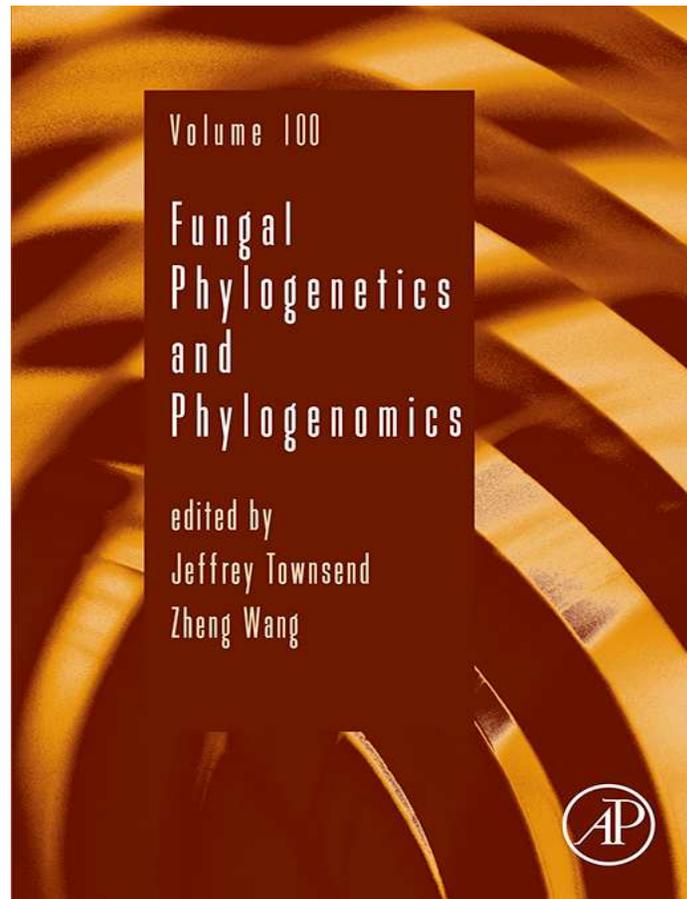
National University of Ireland, Maynooth

187 PUBLICATIONS 2,478 CITATIONS

SEE PROFILE

**Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.**

This chapter was originally published in the book *Advances in Genetics, Vol. 100* published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

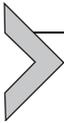
<http://www.elsevier.com/locate/permissionusematerial>

From Charley G.P. McCarthy and David A. Fitzpatrick, Multiple Approaches to Phylogenomic Reconstruction of the Fungal Kingdom. In: Jeffrey P. Townsend and Zheng Wang, editors, *Advances in Genetics, Vol. 100*, Burlington: Academic Press, 2017, pp. 211-266.

ISBN: 978-0-12-813261-6

© Copyright 2017 Elsevier Inc.

Academic Press



Multiple Approaches to Phylogenomic Reconstruction of the Fungal Kingdom

Charley G.P. McCarthy, David A. Fitzpatrick¹

Maynooth University, Maynooth, County Kildare, Ireland

¹Corresponding author: e-mail address: david.fitzpatrick@nuim.ie

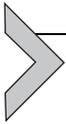
Contents

1. Introduction	212
1.1 The Phylogeny of the Fungal Kingdom	212
1.2 <i>Saccharomyces cerevisiae</i> and the Origin of Modern Fungal Genomics	213
1.3 Fungal Genomics and Phylogenomics Beyond the Yeast Genome	214
1.4 The 1000 Fungal Genomes Project	215
2. Phylogenomic Reconstructions of the Fungal Kingdom	216
2.1 Supermatrix Phylogenomic Analysis of Fungi	225
2.2 Parsimony Supertree Phylogenomic Analysis of Fungi	232
2.3 Bayesian Supertree Phylogenomic Analysis of Fungi	240
2.4 Phylogenomics of Fungi Based on Gene Content	244
2.5 Alignment-Free Phylogenomic Analysis of Fungi	247
3. A Genome-Scale Phylogeny of 84 Fungal Species From Seven Phylogenomic Methods	252
3.1 Higher-Level Genome Phylogeny of the Fungal Kingdom	252
3.2 Multiple Phylogenomic Methods Show Moderate Support for the Modern Designations of Mucoromycota and Zoopagomycota	254
3.3 Pezizomycotina as a Benchmark for Phylogenomic Methodologies	255
3.4 The Use of Phylogenomics Methods in Fungal Systematics	257
4. Concluding Remarks	259
Acknowledgments	260
References	260

Abstract

Fungi are possibly the most diverse eukaryotic kingdom, with over a million member species and an evolutionary history dating back a billion years. Fungi have been at the forefront of eukaryotic genomics, and owing to initiatives like the 1000 Fungal Genomes Project the amount of fungal genomic data has increased considerably over the last 5 years, enabling large-scale comparative genomics of species across the kingdom. In this chapter, we first review fungal evolution and the history of fungal genomics.

We then review in detail seven phylogenomic methods and reconstruct the phylogeny of 84 fungal species from 8 phyla using each method. Six methods have seen extensive use in previous fungal studies, while a Bayesian supertree method is novel to fungal phylogenomics. We find that both established and novel phylogenomic methods can accurately reconstruct the fungal kingdom. Finally, we discuss the accuracy and suitability of each phylogenomic method utilized.



1. INTRODUCTION

1.1 The Phylogeny of the Fungal Kingdom

The fungi are one of the six kingdoms of life sensu Cavalier-Smith, sister to the animal kingdom, and are thought to span approximately 1.5 million species found across a broad range of ecosystems (Baldauf & Palmer, 1993; Berbee & Taylor, 1992; Cavalier-Smith, 1998; Hawksworth, 2001; Nikoh, Hayase, Iwabe, Kuma, & Miyata, 1994). While the overall fossil record of the fungi is poor due to their simple morphology, fungal fossils have been identified dating back to the Ordovician period approximately 400 million years ago (Redecker, 2000) and molecular clock analyses suggest that the fungi originated in the Precambrian eon approximately 0.76–1.06 billion years ago (Berbee & Taylor, 2010). Classic studies into fungal evolution were based on the comparison of morphological or biochemical characteristics; however, the broad range of diversity within the fungal kingdom had limited the efficacy of some of these studies (Berbee & Taylor, 1992; Heath, 1980; Léjohn, 1974; Taylor, 1978). Since the development of phylogenetic approaches within systematics and the incorporation of molecular data into phylogenetic analyses, our understanding of the evolution of fungi has improved substantially (Guarro, Gené, & Stchigel, 1999).

Initial phylogenetic analyses of fungal species had revealed that there were four distinct phyla within the fungal kingdom: the early-diverging Chytridiomycota and Zygomycota, and the Ascomycota and Basidiomycota. The Chytridiomycota grouping was later subject to revision (James et al., 2006), and in their comprehensive classification of the fungal kingdom in 2007 Hibbet et al. formally abandoned the phylum Zygomycota (Hibbett et al., 2007). Instead, Hibbet et al. treated zygomycete species as four *incertae sedis* subphyla (Entomophthoromycotina, Kickellomycotina, Mucoromycotina, and Zoopagomycotina) and subsequently described one subkingdom (the Dikarya) and seven phyla namely Chytridiomycota, Neocallimastigomycota, Blastocladiomycota, Microsporidia, Glomeromycota, Ascomycota, and

Basidiomycota (Hibbett et al., 2007). More recent phylogenetic classification of the zygomycetes has led to the circumscription of the Mucoromycota and Zoopagomycota phyla (Spatafora et al., 2016). Furthermore, recent phylogenetic analyses have shown that *Rozella* species occupy a deep branching position in the fungal kingdom (James et al., 2006; Jones, Forn, et al., 2011), the clade containing these species are now termed the Cryptomycota phylum (Jones, Forn, et al., 2011; Jones, Richards, Hawksworth, & Bass, 2011).

1.2 *Saccharomyces cerevisiae* and the Origin of Modern Fungal Genomics

In terms of genomic data, fungi are by far the highest sampled eukaryotic kingdom, with assembly data available for over 1000 fungal species on the NCBI's GenBank facility as of May 2017. Many of these species also have multiple strains sequenced (the most extreme example being *S. cerevisiae*, which has over 400 strain assemblies available on GenBank). This reflects both the ubiquity of fungi in many areas of biological and medical study and the relative simplicity of sequencing fungal genomes with modern sequencing technology. Fungi have been the exemplar group in eukaryote genetics and genomics, from the first determination of a nucleic acid sequence taken from *S. cerevisiae* by Holley and company in the late 1960s to the sequencing of the first eukaryotic genome in the mid-1990s (Goffeau et al., 1996; Holley et al., 1965). The genome of *S. cerevisiae* was sequenced through a massive international collaboration that grew to involve approximately 600 scientists in 94 laboratories and sequencing centers from across 19 countries between 1989 and 1996 (Engel et al., 2014; Goffeau et al., 1996; Goffeau & Vassarotti, 1991). Throughout the early 1990s, each of the standard 16 nuclear chromosomes of *S. cerevisiae*, sourced from the common laboratory strain 288C and its isogenic derivative strains AB972 and FY1679, was individually sequenced and published by participating researchers (Engel et al., 2014 briefly summarize each of these sequencing projects) with the initial publication of chromosome III involving 35 European laboratories on its own (Oliver et al., 1992). The complete genome sequence of *S. cerevisiae* 288C was finally published in 1996, with 5885 putative protein-coding genes and 275 transfer RNA genes identified across the genome's ~12 million base pairs (Goffeau et al., 1996).

In the intervening years the *S. cerevisiae* 288C reference genome has been constantly updated and refined as individual genes or chromosomes have been reanalyzed or even resequenced, and all of these revisions have been recorded and maintained by the *Saccharomyces* Genome Database (Fisk et al., 2006). It is worth noting, however, that such was the attention paid

to the original sequencing project by its contributors that the most recent major update of the *S. cerevisiae* 288C reference genome, a full resequencing of the derivative AB972 strain using far less labor-intensive modern sequencing and annotation techniques, made only minor alterations to the original genome annotation overall (Engel et al., 2014). Much of our understanding regarding the processes of genome evolution in eukaryotes since 1996 has also been derived from the study of the *S. cerevisiae* 288C genome, including the confirmation that the *S. cerevisiae* genome had undergone a whole-genome duplication (WGD) event (Kellis, Birren, & Lander, 2004; Wolfe & Shields, 1997), the effect of interspecific hybridization on genome complexity (De Barros Lopes, Bellon, Shirley, & Ganter, 2002), evidence that interdomain horizontal gene transfer (HGT) from prokaryotes into eukaryotes has occurred (Hall & Dietrich, 2007), to the ongoing development of an entirely synthetic genome through the Sc2.0 project (Annaluru et al., 2014).

1.3 Fungal Genomics and Phylogenomics Beyond the Yeast Genome

As more model organisms from other eukaryotic kingdoms had their genomes sequenced, *S. cerevisiae* 288C provided a useful comparison as the reference fungal genome, even for more complex eukaryotes like *Drosophila melanogaster*. However, the later sequencing of other model fungal species *Schizosaccharomyces pombe* and *Neurospora crassa* showed the limits of relying solely on *S. cerevisiae* as a reference for the entire fungal kingdom, particularly the latter; *N. crassa* was found to have a far larger genome than either *S. cerevisiae* or *S. pombe* and over 57% of genes predicted in *N. crassa* had no homolog in either of the other two sequenced fungal genomes (Galagan et al., 2003; Galagan, Henn, Ma, Cuomo, & Birren, 2005; Wood et al., 2002). Borne out of a lull in fungal genomic advances and the increasing sophistication of sequencing technology, the Fungal Genome Initiative (FGI) was set up by a number of research organizations in the early 2000s, under the aegis of the Broad Institute (Cuomo & Birren, 2010). Collaborators within the FGI were tasked with the sequencing and annotating the genomes of over 40 species from across the fungal kingdom, with a broad scope of species selected for analysis, medically significant human fungal pathogens like *Candida albicans* and *Aspergillus fumigatus*, commercially important species such as *Penicillium chrysogenum* and *Sclerotinia sclerotiorum*, as well as basal fungal species such as *Phycomyces blakesleeanus* (Cuomo & Birren, 2010). Between 2004 and 2012, in approximately the same amount

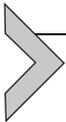
of time it had taken to sequence each individual chromosome of *S. cerevisiae* 288C in the 1990s, over 100 fungal genomes were sequenced and made publicly available on facilities like GenBank and the Joint Genome Institute (JGI)'s Genome Portal website (Benson et al., 2013; Grigoriev, Nordberg, et al., 2011).

The steady increase in genomic data available for fungi from the first decade of this century on, while still sampled mainly from the Ascomycota and Basidiomycota phyla, allowed for a greater range of fungal genomic analyses to be conducted. This included phylogenomic analyses of the fungal kingdom using a variety of different methods (which we will discuss in detail in the following section) and comparative investigations such as analysis of the evolution of pathogenicity in genera like *Candida* or *Aspergillus* (Butler et al., 2009; Galagan, Calvo, Cuomo, et al., 2005; Jackson et al., 2009), the extent of inter-/intra-kingdom HGT both to and from fungal genomes (Fitzpatrick, Logue, & Butler, 2008; Marcet-Houben & Gabaldón, 2010; Richards et al., 2011; Szöllősi, Davín, Tannier, Daubin, & Boussau, 2015), identification of clusters of secondary metabolites (Keller, Turner, & Bennett, 2005; Khaldi et al., 2010), and syntenic relationships across *Saccharomyces* and *Candida* (Byrne & Wolfe, 2005; Fitzpatrick, O'Gaora, Byrne, & Butler, 2010). The wealth of genomic data available for some fungal orders or classes has allowed for easier automation of the sequencing and annotation of novel-related species, through the development of reference transcriptomic or proteomic data for gene prediction software such as AUGUSTUS or quality assessment software for genome assembly such as BUSCO (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015; Stanke, Steinkamp, Waack, & Morgenstern, 2004).

1.4 The 1000 Fungal Genomes Project

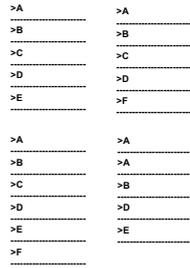
The recent deluge of genomic data available for the fungal kingdom comes as a result of the 1000 Fungal Genomes Project, an initiative headed by the JGI. The project (which can be found at <http://genome.jgi.doe.gov/pages/fungi-1000-projects.jsf>) aims to provide genomic sequence data from at least one species from every circumscribed fungal family, either from projects headed by the JGI, projects which have been incorporated into the MycoCosm database or through community-led nomination and provision of sequencing material. The project has an inbuilt preference for sequencing projects arising from families with no sequenced species to date, or only one other reference genome at the time of nomination. Assembly and

annotation data are then hosted at the JGI's MycoCosm facility as well as other publically available databases (Grigoriev et al., 2014). This community-wide effort has led to a staggering increase in the number of fungal genomes available within the last 5 years; Grigoriev et al. (2014) quoted the number of genomes present in MycoCosm at over 250 at the end of 2013; as of May 2017 there are 772 fungal genomes available to download from the facility, with another 500 species nominated for sequencing. The project has seen a large increase particularly in the amount of data available from fungal phyla outside of the Dikarya, with 58 genomes currently available from the zygomycetes, the Chytridiomycota, Neocallimastigomycota, and Blastocladiomycota. There are many other fungal families with species yet to be nominated for sequencing, including many families from the Pezizomycotina subphylum within Ascomycota and the Chytridiomycota phylum. It is hoped that the wealth of fungal genomic data arising from the 1000 Fungal Genomes Project will help, among countless other scenarios, to fuel the search for novel biosynthetic products and to better understand the ecological effects of different families within the fungal kingdom (Grigoriev, Cullen, et al., 2011). The initiative will also enable the large-scale comparative analysis of hundreds of fungal species from across the fungal kingdom, including kingdom-level phylogenomic reconstructions.



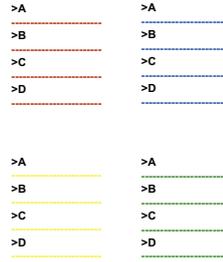
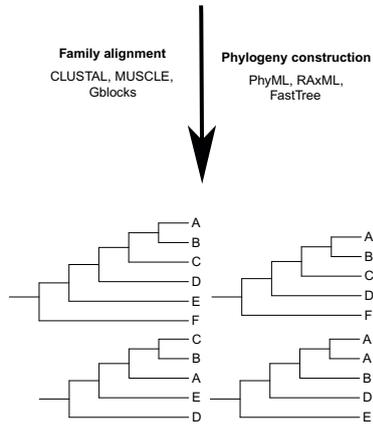
2. PHYLOGENOMIC RECONSTRUCTIONS OF THE FUNGAL KINGDOM

Phylogenetic inference arising from molecular data has, in the past, predominately relied on single genes or small numbers of highly conserved genes or nuclear markers. While usually these markers make for robust individual phylogenies, potential conflicts can occur between individual phylogenies depending on the marker(s) used. The selection of such markers may also overlook other gene families which may be phylogenetically informative, such as gene duplication events or HGT events (Bininda-Emonds, 2004). With the advent of genome sequencing and the increasing sophistication of bioinformatics software and techniques, it has become common practice to reconstruct the evolutionary relationships of species by utilizing large amounts of phylogenetically informative genomic data. Such data can include ubiquitous or conserved genes, individual orthologous and paralogous gene phylogenies, shared genomic content, or compositional signatures of genomes (Fig. 1). Methods of phylogenomic analysis, in other words phylogenetic reconstruction of species using genome-scale data, have



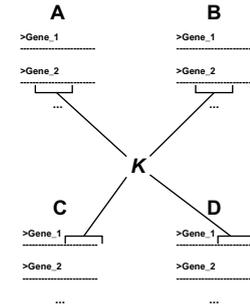
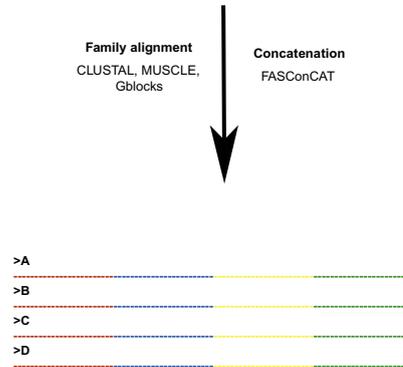
N number of gene families

Depending on size and distribution of taxa, aligned families can be used to generate input data for supertree analysis



Gene families

Gene families that are ubiquitous within taxa (e.g., KOGs) can be concatenated into a "superalignment"



Generating composition vectors

Initial generation of a vector based on overlapping strings of length K per query genome.

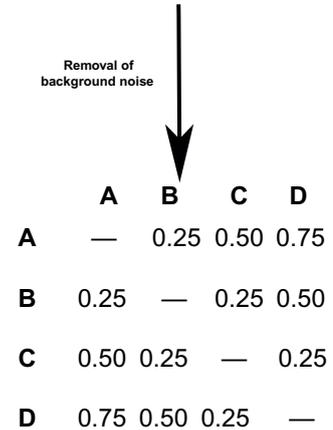


Fig. 1 See figure legend on next page

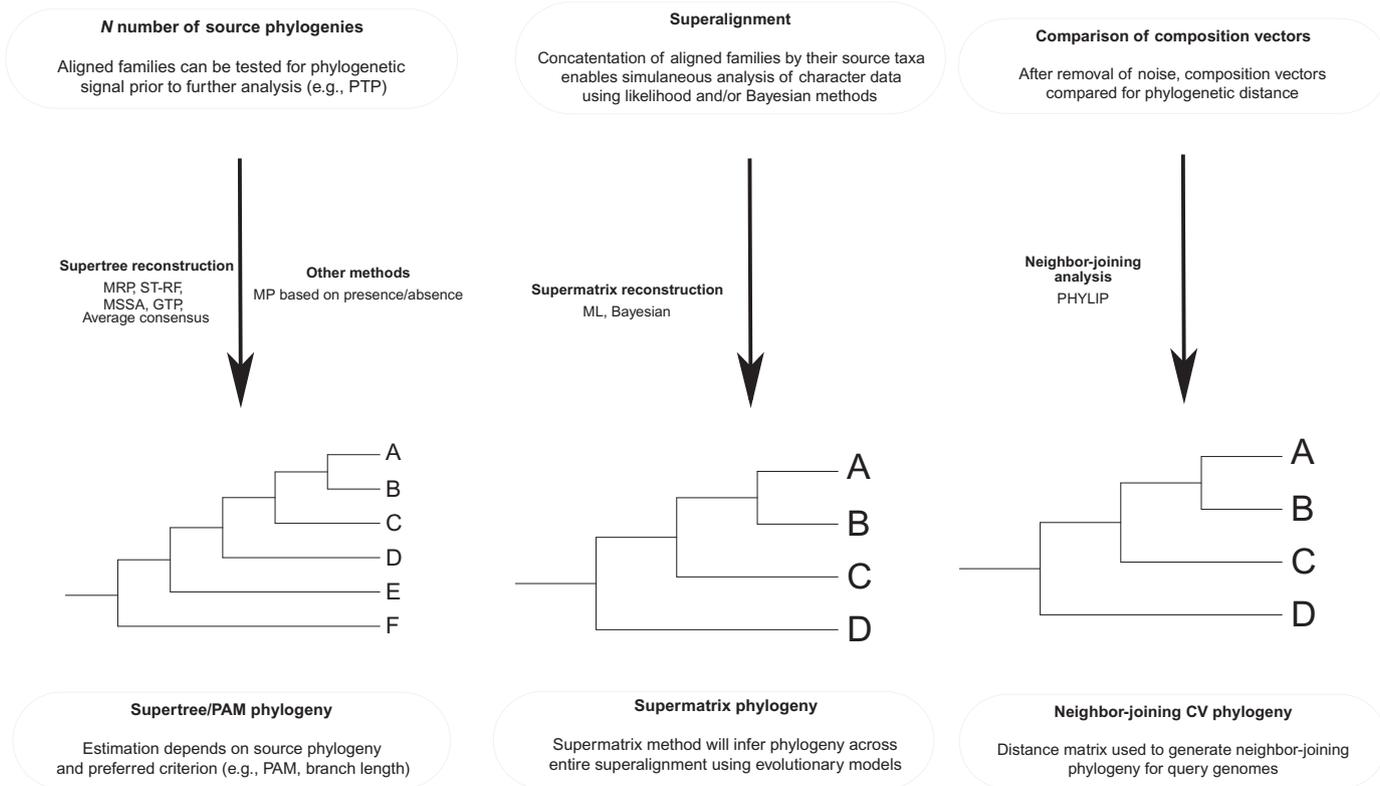


Fig. 1 Illustrative comparison of common phylogenomic methods. *Left*: supertree and presence–absence methods, *middle*: supermatrix methods, and *right*: composition vector methods.

all been developed for each of these types of potential phylogenetic marker and each comes with their advantages and disadvantages. Many phylogenomic analyses of the fungal kingdom have been carried out using these methods.

In this section, we review in turn each established approach to phylogenomic reconstruction from molecular data present in the literature and review each approach's application in previous fungal phylogenomic analyses. To demonstrate both the application and accuracy of all of these approaches to reconstructing phylogeny from genome-scale data, we have conducted our own phylogenomic analyses of the fungal kingdom using each method (Fig. 2). We have carried out such analyses to take advantage of both the greater coverage of the fungal kingdom arising from the 1000 Fungal Genomes Project and the advances in phylogenetic methodologies in the years following many of the analyses that we review below. In total, 84 fungal genomes from across 8 phyla (Table 1) were selected for our

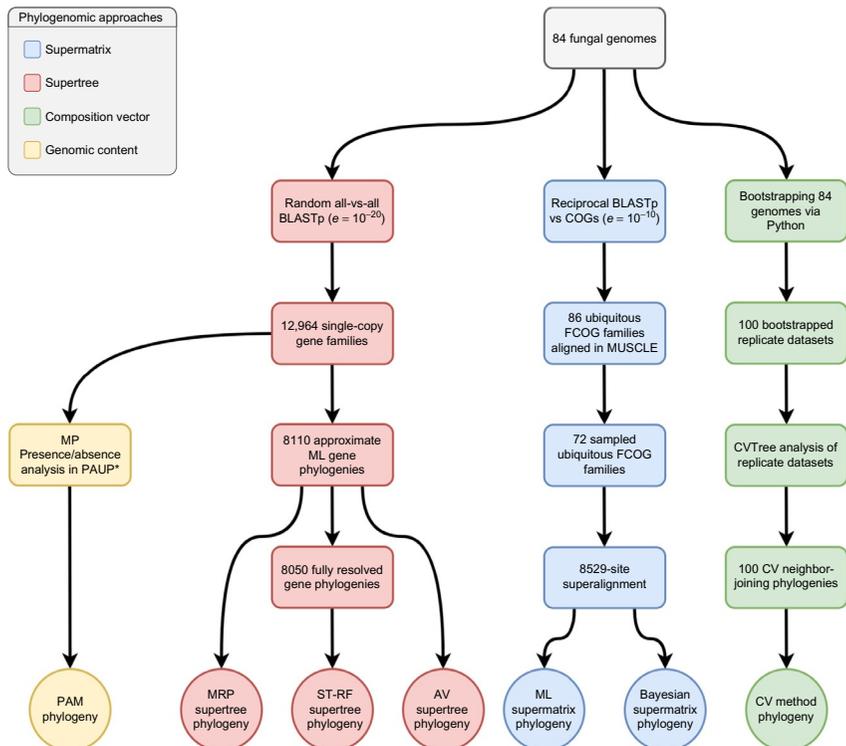


Fig. 2 Summary of the methodology of all 7 phylogenomic analyses of 84 fungal species carried out in this review.

Table 1 List of Species Used in Phylogenomic Analysis

Species	Phylum	Subphylum	Class	MycCosm ID
<i>Bipolaris maydis</i>	Ascomycota	Pezizomycotina	Dothideomycetes	CocheC4_1
<i>Cenococcium geophilum</i>	Ascomycota	Pezizomycotina	Dothideomycetes	Cenge3
<i>Hysterium pulicare</i>	Ascomycota	Pezizomycotina	Dothideomycetes	Hyspu1_1
<i>Zymoseptoria tritici</i>	Ascomycota	Pezizomycotina	Dothideomycetes	Mycgr3
<i>Aspergillus niger</i>	Ascomycota	Pezizomycotina	Eurotiomycetes	Aspni7
<i>Coccidioides immitis</i>	Ascomycota	Pezizomycotina	Eurotiomycetes	Cocim1
<i>Endocarpon pusillum</i>	Ascomycota	Pezizomycotina	Eurotiomycetes	EndpusZ1
<i>Exophiala dermatitidis</i>	Ascomycota	Pezizomycotina	Eurotiomycetes	Exode1
<i>Phaeoconiella chlamydospora</i>	Ascomycota	Pezizomycotina	Eurotiomycetes	Phach1
<i>Blumeria graminis</i>	Ascomycota	Pezizomycotina	Leotiomycetes	Blugr1
<i>Botrytis cinerea</i>	Ascomycota	Pezizomycotina	Leotiomycetes	Botci1
<i>Arthrobotrys oligospora</i>	Ascomycota	Pezizomycotina	Orbiliomycetes	Artol1
<i>Dactylellina haptotyla</i>	Ascomycota	Pezizomycotina	Orbiliomycetes	Monha1
<i>Pyronema omphalodes</i>	Ascomycota	Pezizomycotina	Pezizomycetes	Pyrco1
<i>Tuber melanosporum</i>	Ascomycota	Pezizomycotina	Pezizomycetes	Tubme1
<i>Coniochaeta ligniaria</i>	Ascomycota	Pezizomycotina	Sordariomycetes	Conli1
<i>Hypoxylon</i> sp. EC38	Ascomycota	Pezizomycotina	Sordariomycetes	HypEC38_3

<i>Magnaporthe grisea</i>	Ascomycota	Pezizomycotina	Sordariomycetes	Maggr1
<i>Neurospora crassa</i>	Ascomycota	Pezizomycotina	Sordariomycetes	Neucr_trp3_1
<i>Ophiostoma piceae</i>	Ascomycota	Pezizomycotina	Sordariomycetes	Ophpic1
<i>Phaeoacremonium minimum</i>	Ascomycota	Pezizomycotina	Sordariomycetes	Phaal1
<i>Xylona heveae</i>	Ascomycota	Pezizomycotina	Xylonomycetes	Xylhe1
<i>Candida albicans</i>	Ascomycota	Saccharomycotina	Saccharomycetes	Canalb1
<i>Lipomyces starkeyi</i>	Ascomycota	Saccharomycotina	Saccharomycetes	Lipst1_1
<i>Ogataea polymorpha</i>	Ascomycota	Saccharomycotina	Saccharomycetes	Hanpo2
<i>Saccharomyces cerevisiae</i>	Ascomycota	Saccharomycotina	Saccharomycetes	SacceM3707_1
<i>Saitoella complicata</i>	Ascomycota	Taphrinomycotina	N/A	Saico1
<i>Pneumocystis jirovecii</i>	Ascomycota	Taphrinomycotina	Pneumocystidomycetes	Pneji1
<i>Schizosaccharomyces cryophilus</i>	Ascomycota	Taphrinomycotina	Schizosaccharomycetes	Schcy1
<i>Schizosaccharomyces japonicus</i>	Ascomycota	Taphrinomycotina	Schizosaccharomycetes	Schja1
<i>Schizosaccharomyces octosporus</i>	Ascomycota	Taphrinomycotina	Schizosaccharomycetes	Schoc1
<i>Schizosaccharomyces pombe</i>	Ascomycota	Taphrinomycotina	Schizosaccharomycetes	Schpo1
<i>Protomyces lactucaedebilis</i>	Ascomycota	Taphrinomycotina	Taphrinomycetes	Prola1
<i>Taphrina deformans</i>	Ascomycota	Taphrinomycotina	Taphrinomycetes	Tapde1_1
<i>Agaricus bisporus</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Agabi_varbur_1

Continued

Table 1 List of Species Used in Phylogenomic Analysis—cont'd

Species	Phylum	Subphylum	Class	MycCosm ID
<i>Auricularia subglabra</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Aurde3_1
<i>Botryobasidium botryosum</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Botbo1
<i>Fibulorhizoctonia</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Fibsp1
<i>Gloeophyllum trabeum</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Glotr1_1
<i>Heterobasidion annosum</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Hetan2
<i>Jaapia argillacea</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Jaar1
<i>Punctularia strigosozonata</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Punst1
<i>Serendipita indica</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Pirin1
<i>Serpula lacrymans</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	SerlaS7_3_2
<i>Sistotremastrum suecicum</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Sissu1
<i>Sphaerobolus stellatus</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Sphst1
<i>Wolfiporia cocos</i>	Basidiomycota	Agaricomycotina	Agaricomycetes	Wolco1
<i>Calocera cornea</i>	Basidiomycota	Agaricomycotina	Dacrymycetes	Calco1
<i>Dacryopinax primogenitus</i>	Basidiomycota	Agaricomycotina	Dacrymycetes	Dacsp1
<i>Basidioascus undulatus</i>	Basidiomycota	Agaricomycotina	Geminibasidiomycetes	Basun1
<i>Cryptococcus neoformans</i>	Basidiomycota	Agaricomycotina	Tremellomycetes	Cryne_JEC21_1
<i>Cutaneotrichosporon oleaginosus</i>	Basidiomycota	Agaricomycotina	Tremellomycetes	Triol1

<i>Wallemia sebi</i>	Basidiomycota	Agaricomycotina	Wallemiomycetes	Walse1
<i>Leucosporidium creatinivorum</i>	Basidiomycota	Pucciniomycotina	Microbotryomycetes	Leucr1
<i>Microbotryum lychnidis-dioicae</i>	Basidiomycota	Pucciniomycotina	Microbotryomycetes	Micld1
<i>Rhodotorula graminis</i>	Basidiomycota	Pucciniomycotina	Microbotryomycetes	Rhoba1_1
<i>Mixia osmundae</i>	Basidiomycota	Pucciniomycotina	Mixiomycetes	Mixos1
<i>Puccinia graminis</i>	Basidiomycota	Pucciniomycotina	Pucciniomycetes	Pucgr2
<i>Tilletiaria anomala</i>	Basidiomycota	Ustilaginomycotina	Exobasidiomycetes	Tilan2
<i>Malassezia sympodialis</i>	Basidiomycota	Ustilaginomycotina	Malasseziomycetes	Malsy1_1
<i>Sporisorium reilianum</i>	Basidiomycota	Ustilaginomycotina	Ustilaginomycetes	Spore1
<i>Ustilago maydis</i>	Basidiomycota	Ustilaginomycotina	Ustilaginomycetes	Ustma1
<i>Allomyces macrogynus</i>	Blastocladiomycota	N/A	Blastocladiomycetes	GCA_000151295.1
<i>Catenaria anguillulae</i>	Blastocladiomycota	N/A	Blastocladiomycetes	Catan2
<i>Batrachochytrium dendrobatidis</i>	Chytridiomycota	N/A	Chytridiomycetes	GCA_000149865.1
<i>Rhizoclosmatium globosum</i>	Chytridiomycota	N/A	Chytridiomycetes	Rhihy1
<i>Spizellomyces punctatus</i>	Chytridiomycota	N/A	Chytridiomycetes	Spipu1
<i>Gonapodya prolifera</i>	Chytridiomycota	N/A	Monoblepharidomycetes	Ganpr1
<i>Rozella allomycis</i>	Cryptomycota	N/A	N/A	Rozal1_1
<i>Rhizophagus irregularis</i>	Mucoromycota	Glomeromycotina	Glomeromycetes	Gloin1

Continued

Table 1 List of Species Used in Phylogenomic Analysis—cont'd

Species	Phylum	Subphylum	Class	MycCosm ID
<i>Mortierella elongate</i>	Mucoromycota	Mortierellomycotina	N/A	Morel2
<i>Phycomyces blakesleeanus</i>	Mucoromycota	Mucoromycotina	N/A	Phybl2
<i>Rhizopus oryzae</i>	Mucoromycota	Mucoromycotina	N/A	Rhior3
<i>Umbelopsis ramanniana</i>	Mucoromycota	Mucoromycotina	N/A	Umbra1
<i>Anaeromyces robustus</i>	Neocallimastigomycota	N/A	Neocallimastigomycetes	Anasp1
<i>Neocallimastix californiae</i>	Neocallimastigomycota	N/A	Neocallimastigomycetes	Neosp1
<i>Orpinomyces</i> sp. C1A	Neocallimastigomycota	N/A	Neocallimastigomycetes	Orpsp1_1
<i>Piromyces finnis</i>	Neocallimastigomycota	N/A	Neocallimastigomycetes	Pirfi3
<i>Basidiobolus meristosporus</i>	Zoopagomycota	Entomophthoromycotina	Basidiobolomycetes	Basme2fnSC
<i>Conidiobolus thromboides</i>	Zoopagomycota	Entomophthoromycotina	Entomophthoromycetes	Conth1
<i>Coemansia reversa</i>	Zoopagomycota	Kickxellomycotina	N/A	Coere1
<i>Linderina pennispora</i>	Zoopagomycota	Kickxellomycotina	N/A	Linpe1
<i>Martensiomycetes pterosporus</i>	Zoopagomycota	Kickxellomycotina	N/A	Marpt1
<i>Ramicandelaber brevisporus</i>	Zoopagomycota	Kickxellomycotina	N/A	Rambr1

Genome data from MycoCosm (<http://genome.jgi.doe.gov/programs/fungi/index.jsf>) has previously been published and MycoCosm ID is given in final column. GENBANK accessions given for *Allomyces macrogynus* and *Batrachochytrium dendrobatidis*.

large-scale phylogenomic reconstructions of the fungal kingdom. Where possible, we included at least one published representative genome from each order covered by the 1000 Fungal Genomes Project in our dataset. All genomic data were ultimately obtained from the JGI's MycoCosm facility (Grigoriev et al., 2014). Our analyses include the first phylogenomic reconstruction of fungi carried out using a Bayesian supertree approach, and the first large-scale gene content approach to fungal phylogenomics that has been conducted in at least a decade. We discuss, in brief, the methodology and the results of each reconstruction and their accuracy (or otherwise) in reconstructing the phylogeny of both basal fungal lineages and the Dikarya. In Section 3, we discuss the overall phylogeny of the fungal kingdom arising from our analyses and compare with previous literature.

2.1 Supermatrix Phylogenomic Analysis of Fungi

The two best-established alignment-based approaches to reconstructing phylogeny on a genomic scale are the “supertree” method, in which a consensus phylogeny is derived from many individual gene phylogenies (discussed in Section 2.2), and the “supermatrix” method which we discuss here. Supermatrix method phylogeny is the simultaneous analysis of a phylogenetic matrix, also referred to as a “superalignment,” constructed from all available character data from a given set of taxa. Generally supermatrices are constructed from concatenating highly conserved markers (e.g., rRNA genes, mitochondrial markers) for small-scale multigene phylogenies, and from homologs of conserved orthologous genes (known as COGs, or sometimes KOGs in eukaryotes) for genome-scale phylogenies (de Queiroz & Gatesy, 2007; Koonin et al., 2004). Supermatrix approaches can also incorporate statistically powerful maximum-likelihood and Bayesian methods of phylogenomic analysis. Described in simple terms, given an alignment of sequences and a suitable evolutionary model, maximum-likelihood phylogenetic analysis examines all possible trees by their possible parameters (e.g., topology, site support, branch length) and returns the most likely phylogenetic tree for the alignment (Page & Holmes, 1998). Similarly, Bayesian analysis incorporates phylogenetic likelihoods to calculate the posterior probability of a phylogeny, which is the probability of that phylogeny given the alignment data (Huelsenbeck, Ronquist, Nielsen, & Bollback, 2001).

One advantage of a supermatrix approach to phylogenomic analysis over a supertree approach is the retention of character evidence in analysis in the former approach; most supertree methods can be considered estimations

using individual trees based on summarized character data, at least two steps removed from any actual sequence data, whereas a supermatrix approach entails direct analysis of combined character data (Creevey & McInerney, 2009; de Queiroz & Gatesy, 2007). Supermatrix methods also have the potential to resolve deep branches and reveal so-called hidden supports within phylogenies that supertree methods may overlook (de Queiroz & Gatesy, 2007). However, supermatrix analysis requires ubiquitous sequences from all taxa being investigated, which restricts the available pool of character data and may overlook miss important phylogenetic information from phylogenies with gene deletion, gene duplication, or horizontal gene transfer events that supertree methods can utilize (Creevey & McInerney, 2009). Compositional biases may also have an effect on supermatrix methods, though phylogenetic models have been developed which can ameliorate errors that these biases may induce during analysis (Lartillot, Brinkmann, & Philippe, 2007; Lartillot & Philippe, 2004). In practice, many phylogenomic analyses utilize both supertree and supermatrix methods in tandem to reconstruct phylogeny in a “total evidence” approach (Kluge, 1989) and will often comment on the taxonomic congruence (or otherwise) of the resulting phylogenies.

2.1.1 Fungal Phylogenomics Using the Supermatrix Approach

Supermatrix analysis has been widely used in fungal phylogenomics. One of the initial comparisons of individual gene phylogenies with genome-scale species phylogenies used a maximum-parsimony analysis among other methods to reconstruct the phylogeny of seven *Saccharomyces* species and *C. albicans*; the authors showed that incongruence among individual gene phylogenies could be resolved with high support using a concatenated alignment (Rokas, Williams, King, & Carroll, 2003). Initial genome-based phylogenies of Ascomycota using 17 genomes and both supertree and supermatrix methods resolved both Pezizomycotina and Saccharomycotina, as well as placing *S. pombe* as an early-diverging branch within Ascomycota (Robbertse, Reeves, Schoch, & Spatafora, 2006). Robbertse et al. (2006) generated a superalignment of 195,664 amino acid characters in length derived from 781 gene families, which produced identical topologies under both neighbor-joining and maximum-likelihood criteria. The first large-scale phylogenomic analysis of fungi used a 67,101-character superalignment derived from 531 eukaryotic COGs found in 21 fungal genomes, all of which were sampled from Ascomycota and Basidiomycota (Kuramae, Robert, Snel, Weiß, & Boekhout, 2006). A more extensive phylogenomic analysis from the same year produced 2 highly congruent genome phylogenies from 42 fungal

genomes using 2 methods: a matrix representation with parsimony (MRP) supertree derived from 4805 single-copy gene families (which we discuss in greater detail in [Section 2.2.1](#)), and a 38,000-character superalignment derived from 153 ubiquitous gene families ([Fitzpatrick, Logue, Stajich, & Butler, 2006](#)).

Most of the relationships resolved in [Fitzpatrick et al. \(2006\)](#) were further supported by a 31,123-character superalignment from 69 proteins conserved in up to 60 fungal genomes generated by [Marcet-Houben, Marceddu, and Gabaldón \(2009\)](#), although they found a large degree of topological conflict within a 21-species Saccharomycotina clade ([Marcet-Houben & Gabaldón, 2009](#); [Marcet-Houben et al., 2009](#)). A later follow-up analysis to [Fitzpatrick et al. \(2006\)](#) by [Medina, Jones, and Fitzpatrick \(2011\)](#) reconstructed the phylogeny of 103 fungal species by performing Bayesian analysis on a 12,267-site superalignment derived from 87 gene families with a phyletic range of over half of their dataset, in addition to supertree analysis ([Medina et al., 2011](#)). A recent phylogenomic analysis of 46 fungal genomes, including 25 zygomycetes species, reconstructed the phylogeny of the early-diverging fungal lineages using a 60,383-character superalignment ([Spatafora et al., 2016](#)). Another recent phylogenomic analysis used a 28,807-site superalignment derived from 136 gene families from 40 eukaryotic genomes to investigate the evolution of sourcing carbon from algal and plant pectin in early-diverging fungi ([Chang et al., 2015](#)). Finally, a comparison of the dynamics of genome evolution between 28 Dikarya species and cyanobacteria used a supermatrix phylogeny of 24,514 amino acid characters from 529 fungal gene families with large phyletic range as a scaffold to infer rates of intrakingdom HGT within Dikarya that were near similar to those within cyanobacteria ([Szöllősi et al., 2015](#)).

To extend the analyses described above, we carried out supermatrix analysis using maximum-likelihood and Bayesian methods on a superalignment constructed from orthologous genes conserved throughout 84 species from 8 phyla within the fungal kingdom.

2.1.2 Phylogenomic Reconstruction of 84 Fungal Species From 72 Ubiquitous Gene Families Using Maximum-Likelihood and Bayesian Supermatrix Analysis

A reciprocal BLASTp search was carried out between all protein sequences from our 84-genome dataset and 458 core orthologous genes (COGs) from *S. cerevisiae* obtained from the CEGMA dataset, with an *e*-value cutoff of 10^{-10} ([Camacho et al., 2009](#); [Parra, Bradnam, & Korf, 2007](#)), from which

456 COG families were retrieved (2 *S. cerevisiae* COGs did not return any homologs). From these, 86 ubiquitous fungal COG families, i.e., families containing a homolog from all 84 input species, were identified. Each ubiquitous fungal COG family was aligned in MUSCLE, and conserved regions of each alignment were sampled in Gblocks using the default parameters (Castresana, 2000; Edgar, 2004). Fourteen alignments did not retain any character data after Gblocks filtering and were removed from further analysis. The remaining 72 sampled alignments were concatenated into a superalignment of 8529 aligned positions using the Perl program FASconCat (Kück & Meusemann, 2010). This superalignment was bootstrapped 100 times using Seqboot (Felsenstein, 1989), and maximum-likelihood phylogenetic trees were generated for each individual replicate using PhyML with an LG + I + G amino acid substitution model as selected by ProtTest (Darriba, Taboada, Doallo, & Posada, 2011; Guindon et al., 2010). A consensus phylogeny was generated from all 100 individual replicate phylogenies using CLANN (Creevey & McInerney, 2005). Markov Chain Monte Carlo (MCMC) Bayesian phylogenetic inference was carried out on the same superalignment using PhyloBayes MPI with the default CAT + GTR amino acid substitution model, running 2 chains for 1000,000 iterations and sampling every 100 iterations (Lartillot & Philippe, 2004; Lartillot, Rodrigue, Stubbs, & Richer, 2013). Both chains were judged to have converged after 100,000 iterations and a consensus Bayesian phylogeny was generated with a burn-in of 1000 trees. Both supermatrix phylogenies were visualized using the Interactive Tree of Life (iTOL) website and annotated according to the NCBI's taxonomy database (Federhen, 2012; Letunic & Bork, 2016). Both supermatrix phylogenies were rooted at *Rozella allomycis*, which is the most basal species in evolutionary terms in our dataset (Jones, Forn, et al., 2011) and is the root for all the phylogenies we present hereafter (Figs. 3 and 4).

2.1.3 Supermatrix Analyses of 84 Fungal Species Accurately Reconstructs the Fungal Kingdom

We reconstructed the phylogeny of the fungal kingdom by generating a superalignment of 72 concatenated ubiquitous gene families and performing ML analysis using PhyML and Bayesian analysis using a parallelized version of PhyloBayes. Both ML and Bayesian analysis reconstruct the phylogeny of our fungal dataset with a high degree of accuracy relative to other kingdom phylogenies in the literature and in most cases recover the eight fungal phyla in our dataset (Figs. 3 and 4). Here, we discuss the results of both our analyses with regard to the basal fungal lineages, and the two Dikarya

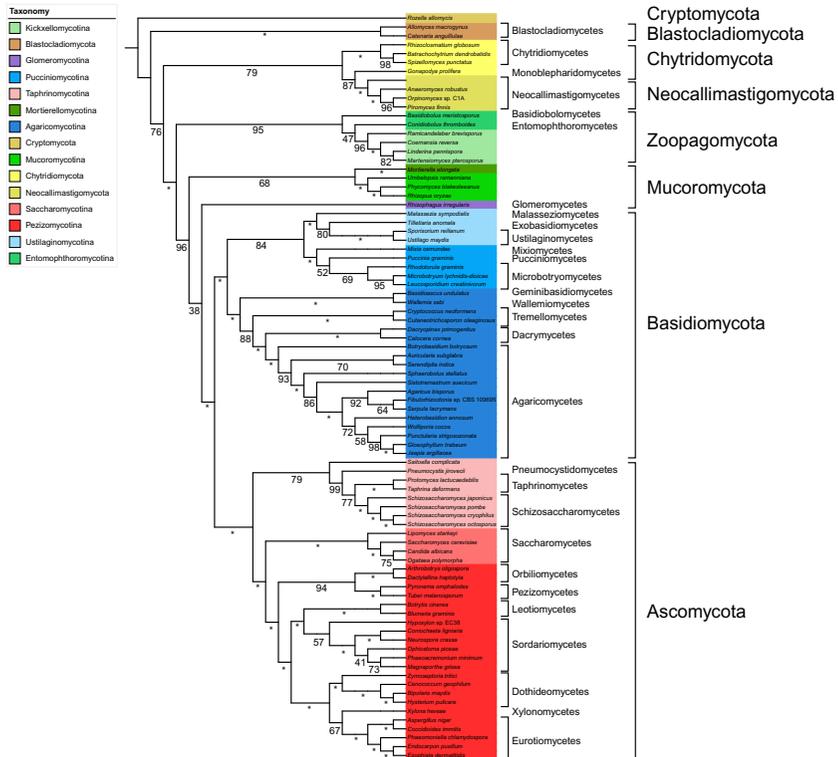


Fig. 3 ML phylogeny of 84 fungal species from a 8529-character superalignment derived from 72 ubiquitous fungal COG families sampled in Gblocks using PhyML with a LG+I+G model. Bootstrap supports shown on branches. Maximum bootstrap support designated with an asterisk (*).

phyla. Further in this chapter, we use these supermatrix analyses as the point of comparison for our other phylogenomic methods.

2.1.3.1 Basal Fungi

In our ML supermatrix phylogeny, Blastocladiomycota emerge as the earliest-diverging fungi with maximum bootstrap support (henceforth abbreviated to BP) after rooting at *R. allomycis* (Fig. 3). Chytridiomycota and Neocallimastigomycota are placed as sister clades with 79% BP, surprisingly the Chytridiomycota species *Gonapodya prolifera* branches as sister to Neocallimastigomycota (87% BP). The Chytridiomycetes class is monophyletic with maximum bootstrap support, as is the Neocallimastigomycetes class (Fig. 3). The former zygomycetes phylum Zoopagomycota is strongly supported as a monophyletic clade with 95% BP (Fig. 3). The other former

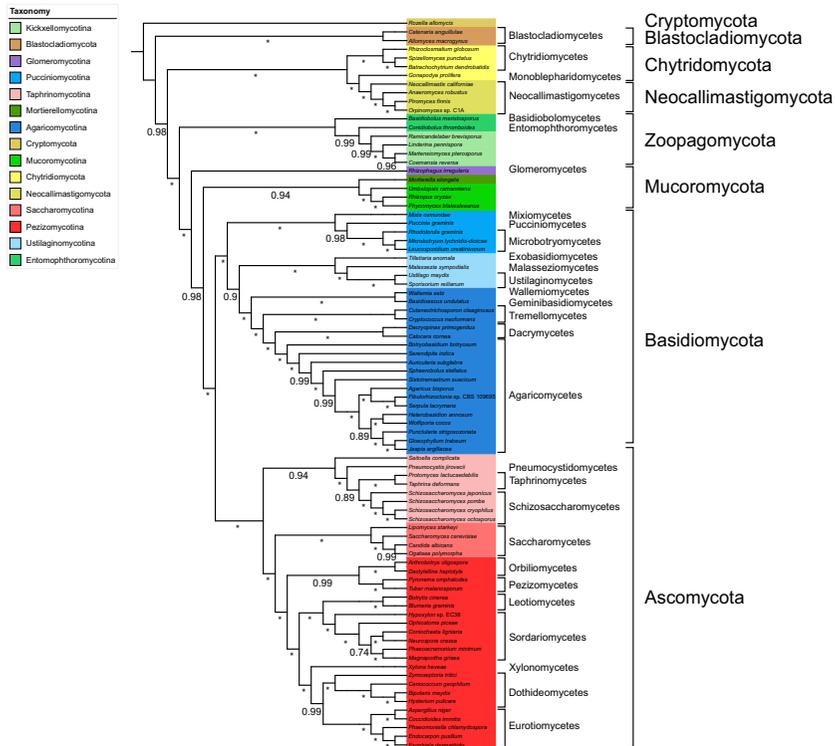


Fig. 4 Bayesian phylogeny of 84 fungal species from an 8529-character superalignment derived from 72 ubiquitous fungal COG families sampled in Gblocks using PhyloBayes MPI with a CAT+GTR model. Posterior probabilities shown on branches with a burn-in of 1000 trees. Maximum posterior probability support designated with an asterisk (*).

zygomycetes phylum Mucoromycota is paraphyletic and split between a clade containing the Mucoromycotina and Mortierellomycotina species *Mortierella elongata* that has 68% BP, and the Glomeromycotina species *Rhizophagus irregularis* branching basal to Dikarya with lower support (38% BP). The placement of Mucoromycota as the closest phyla to Dikarya has near-maximum support (96% BP) which matches other analysis (Spatafora et al., 2016).

The Bayesian supermatrix phylogeny is in near-total agreement with the ML phylogeny in resolving the relationships of the basal fungi in our dataset (Fig. 4). The relationship between Chytridiomycota and Neocallimastigomycota in the Bayesian phylogeny mirrors that seen in the ML phylogeny, with all branches receiving maximum support as monophyletic with a Bayesian posterior probability (henceforth abbreviated to PP) equal to 1 (Fig. 4). The

Zoopagomycota are monophyletic with full support, with a topology matching the ML phylogeny with strong branch support throughout (Fig. 4). There is also a close association between the three Mucoromycota subphyla: Glomeromycota branches earlier in the Bayesian phylogeny than in the ML phylogeny, which receives maximum support in the Bayesian phylogeny, and the sister relationship between Mucoromycotina and *M. elongata* receives strong support (0.94 PP) in the Bayesian phylogeny (Fig. 4). Both the ML and Bayesian place the Mucoromycota as the basal phylum that is most closely related to Dikarya (Fig. 4).

2.1.3.2 Basidiomycota

In the ML phylogeny, the three subphyla within Basidiomycota are fully resolved with maximum BP, with 84% BP for the placement of Ustilagomycotina and Pucciniomycotina as sister clades (Fig. 3). *Basidioascus undulatus* and *Wallemia sebi* branch at the base of Agaricomycotina with maximum BP, while the other classes with the subphyla are all fully supported. There is also high support (88% BP) for the placement of Tremellomycetes as sister to Dacrymycetes and Agaricomycetes (Fig. 3). The Tremellomycetes, including *Cryptococcus neoformans*, are monophyletic. The Dacrymycetes are also monophyletic with maximum BP. The forest saprophyte *Botryobasidium botryosum* is placed at the base of the Agaricomycetes, which has some strong intraclade resolution with weaker branch supports toward the tips of the clade (Fig. 3). *Malassezia sympodialis*, a commensal fungi of humans and animals, is placed at the base of the Ustilagomycotina. The Exobasidiomycetes species *Tilletiaria anomala* branches between *M. sympodialis* and the Ustilagomycetes. The Pucciniomycotina are monophyletic with full support (Fig. 3). The most highly represented Pucciniomycotina class, the Microbotryomycetes, are monophyletic with 69% BP (Fig. 3).

The Bayesian phylogeny reflects the ML phylogeny in its resolution of the Basidiomycota as monophyletic with full support (Fig. 4). The phylogeny places Pucciniomycotina at the base of the phylum with maximum support. Resolution of branches within Pucciniomycotina is substantially improved under Bayesian phylogeny (Fig. 4). There is high support (0.9 PP) for a sister relationship between Ustilagomycotina and Agaricomycotina (Fig. 4). The Exobasidiomycetes species *T. anomala* now branches at the base of the Ustilagomycotina, which is resolved with maximum PP. There is maximum support for the placement of *M. sympodialis* as sister to the Ustilagomycetes, which are monophyletic (Fig. 4). As in the ML phylogeny, *B. undulatus* and *W. sebi* branch at the base of Agaricomycotina with maximum support, while

the other classes with the subphyla all have maximum support and have similar topology under Bayesian analysis. There is a large improvement in the support of branches in the Agaricomycotina in the Bayesian phylogeny relative to the ML phylogeny (Fig. 4).

2.1.3.3 Ascomycota

Both the ML and Bayesian supermatrix phylogenies display near-identical topologies for the Ascomycota, and Bayesian analysis shows stronger support for some branches toward the tips of the phylogeny than the ML phylogeny does (Figs. 3 and 4). The three subphyla within Ascomycota are fully resolved, with maximum BP support for Saccharomycotina and Pezizomycotina and 79% BP for the monophyly of Taphrinomycotina in the ML phylogeny (contrast with 0.94 PP for the monophyly of Taphrinomycotina in the Bayesian phylogeny; Figs. 3 and 4). The placement of Taphrinomycotina as an ancestral clade within Ascomycota is fully supported, and within Taphrinomycotina, there is high support (77% BP/0.89 PP) for a sister relationship between Schizosaccharomycetes and Taphrinomycetes. Six of the seven classes within Pezizomycotina in our dataset with two or more representatives (i.e., all bar Xylonomycetes) are monophyletic, most of which receive maximum BP and/or PP support. Many of the relationships between classes are also well supported in both phylogenies, with lower support (67% BP) for a sister relationship between the Xylonomycetes species *Xylona heveae* and the Eurotiomycetes class in the ML phylogeny; in the Bayesian phylogeny *X. heveae* branches sister to a clade containing Dothideomycetes and Eurotiomycetes with maximum PP support (Figs. 3 and 4). The Dothideomycetes are monophyletic in both phylogenies and branch into two clades with high support under both ML and Bayesian reconstruction (Figs. 3 and 4). The Orbiliomycetes and Pezizomycetes are placed as the most basal Pezizomycotina classes, with strong support (94% BP/0.99 BP) for a sister relationship (Figs. 3 and 4). The Leotiomycetes and Sordariomycetes are also placed as a sister clades with maximum support in both phylogenies. The major difference in the resolution of the Sordariomycetes between the supermatrix phylogenies is the stronger branch supports within the order under Bayesian analysis (Figs. 3 and 4).

2.2 Parsimony Supertree Phylogenomic Analysis of Fungi

The most common supertree methods for reconstructing genome phylogenies are grounded in parsimony methods, in which changes to character states (i.e., evolutionary events such as presence of a given taxon in a tree or even a tree branch) are calculated and phylogeny is reconstructed using

as little state changes as possible. The first supertree construction method to see widespread use in large-scale phylogenetic and phylogenomic analysis was the MRP method. MRP, which was developed independently by Baum (1992) and Ragan (1992), enables the use of source phylogenies with overlapping or missing taxa in generating a consensus phylogeny (Baum, 1992; Ragan, 1992). The method generates a matrix (referred to as a Baum–Ragan matrix) where each column represents one internal branch in each given source phylogeny such that the number of columns within the matrix is equal to the number of internal branches across all source phylogenies, and assigns a score of 1 to taxa from a given source phylogeny P which are present in the clade defined by internal branch A , 0 to taxa present in P but not within the clade defined by A , and ? to taxa that are not present in P (Creevey & McInerney, 2009). The Baum–Ragan matrix is then subject to parsimony analysis, with equal weighting given to each source phylogeny, and reconstructs the supertree phylogeny with the minimum of evolutionary changes required which includes all taxa represented across all source phylogenies. Similar parsimony methods, most notably gene tree parsimony (Slowinski & Page, 1999), extend MRP to include source phylogenies containing duplicated taxa; however, we do not cover such methods in this subsection. Parsimony-based supertree methods like MRP are generally quite accurate in reconstructing phylogeny for large datasets, although some issues have been observed (which we discuss in Section 2.3).

2.2.1 Matrix Representation With Parsimony Analysis in Fungal Phylogenomics

Many phylogenomic analyses of fungi have used parsimony methods. The first large-scale phylogenomic analysis of fungi to use MRP in supertree reconstruction was by Fitzpatrick et al. (2006), who carried out a phylogenomic reconstruction of fungi using 42 genomes from Dikarya and the zygomycete *Rhizopus oryzae* using both supertree and supermatrix methods (Fitzpatrick et al., 2006). Using a random BLASTp approach to identify homologous gene families, where randomly selected query sequences are sequentially searched against a full database and then both query sequences and homologs (if any) are sequentially removed from the database, Fitzpatrick et al. (2006) utilized 4805 single-copy gene phylogenies for MRP supertree reconstruction using the software package CLANN (Creevey & McInerney, 2005, 2009). The MRP phylogeny resolved the Pezizomycotina and Saccharomycotina subphyla within Ascomycota and inferred the Sordariomycetes and the Leotiomycetes as sister classes within Pezizomycotina. The MRP phylogeny also resolved two major clades

within the Saccharomycotina: a monophyletic clade of species that translate the codon CTG as serine instead of leucine (the “CTG clade”), and a grouping of species that have undergone whole genome duplication (the “WGD clade”) and their closest relatives. The authors compared the MRP phylogeny with a maximum-likelihood supermatrix phylogeny reconstructed using 38,000 characters from 153 gene families (as detailed in the previous subsection); both were highly congruent with conflict only in the placement of the sole Dothideomycetes species represented, *Stagonospora nodourum*. The authors also complemented their MRP phylogeny with two other supertree methods implemented in CLANN: a most similar supertree analysis (MSSA) method phylogeny which was identical to the MRP supertree (Creevey et al., 2004) and an average consensus (AV) method phylogeny based on branch lengths (Lapointe & Cucumel, 1997), which the authors believed to suffer from long-branch attraction in the erroneous placement of some species within the WGD clade in Saccharomycotina (Fitzpatrick et al., 2006). A follow-up analysis to Fitzpatrick et al. (2006) by Medina et al. (2011) using 103 genomes was extended to include multicopy gene families using the gene tree parsimony (GTP) method and successfully resolved the major groupings within the fungal kingdom (Medina et al., 2011). Using both a random BLASTp and a Markov Clustering Algorithm (MCL)-based approach with varying inflation values to identify orthologous gene families, the authors used as many as 30,012 single and paralogous gene phylogenies as input for supertree reconstruction.

As a follow-up to the supertree reconstructions of the fungal kingdom by Fitzpatrick et al. (2006) and Medina et al. (2011), we ran supertree analysis for 84 fungal species using MRP and AV methods and source phylogenies identified via a random BLASTp approach described later.

2.2.2 Phylogenomic Reconstruction of 84 Fungal Species From 8110 Source Phylogenies Using MRP and AV Supertree Methods

Following Fitzpatrick et al. (2006), families of homologous protein sequences within our 84-genome dataset were identified using BLASTp with an e -value cutoff of 10^{-20} by randomly selecting a query sequences from our database, finding all homologous sequences via BLASTp (Camacho et al., 2009), and removing the entire family from the database before reformatting and repeating. 12,964 single-copy gene families, which contained no more than one homolog from 4 or more taxa, were identified. Each single-copy gene family was aligned in MUSCLE, and conserved regions of each alignment were sampled using Gblocks with the default parameters (Castresana, 2000; Edgar, 2004). Sampled alignments were tested for phylogenetic signal using

the PTP test as implemented in PAUP* with 100 replicates (Faith & Cranston, 1991; Swofford, 2002). 8110 sampled alignments which retained character data after Gblocks filtering and passed the PTP test were retained for phylogenomic reconstruction. 8110 approximately maximum-likelihood gene phylogenies were generated with FastTree, using the default JTT + CAT protein evolutionary model (Price, Dehal, & Arkin, 2010). All 8110 single-copy gene phylogenies were used to generate a matrix representation with parsimony (MRP) supertree using CLANN, with 100 bootstrap replicates (Creevey & McInerney, 2005). To complement the MRP supertree, an average consensus (AV) supertree was generated from the same input dataset in CLANN, with 100 bootstrap replicates. Both supertrees were visualized in iTOL and annotated according to the NCBI's taxonomy database. Both supertrees were rooted at *R. allomyces* (Figs. 5 and 6).

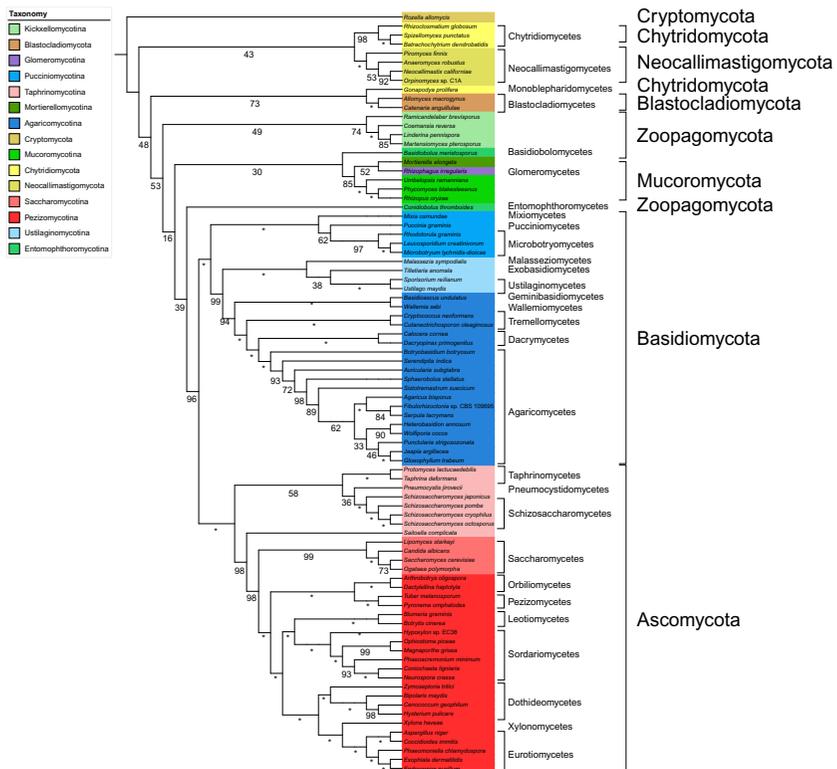


Fig. 5 Matrix representation with parsimony (MRP) phylogeny of 84 fungal species derived from 8110 source phylogenies. Bootstrap supports shown on branches. Maximum bootstrap support designated with an asterisk (*).

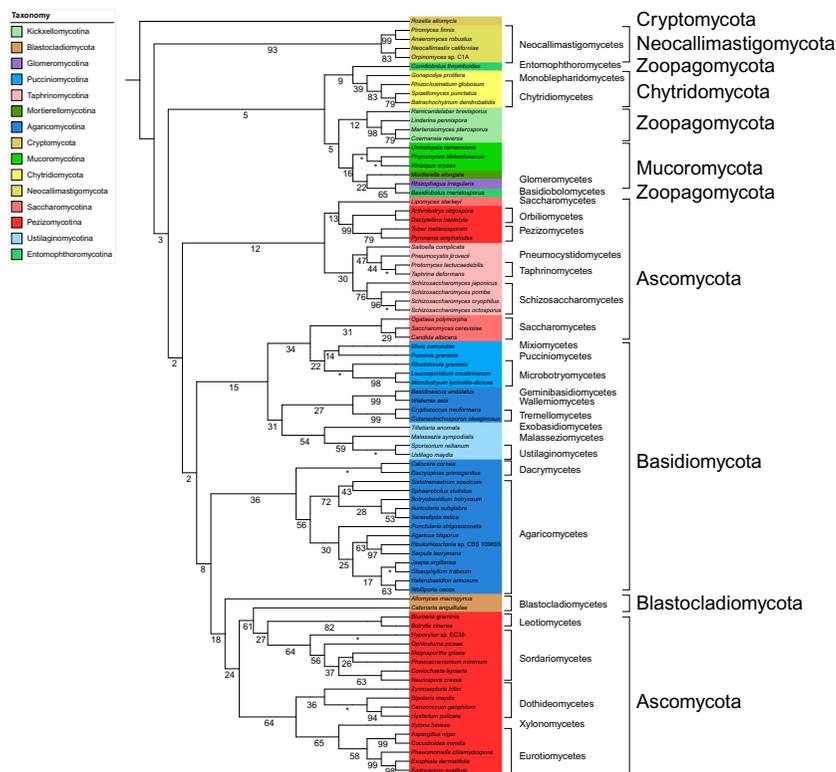


Fig. 6 Average consensus (AV) phylogeny of 84 fungal species derived from 8110 source phylogenies. Bootstrap supports shown on branches. Maximum bootstrap support designated with an asterisk (*).

2.2.3 MRP Phylogenomic Analysis of 84 Fungal Species Is Highly Congruent With Supermatrix Phylogenomic Analyses

We reconstructed the overall phylogeny of 8110 single-copy source phylogenies from our 84-genome dataset using an MRP supertree method analysis as implemented in CLANN (Fig. 5). MRP supertree reconstruction of the fungal kingdom recovers the majority of the eight fungal phyla in our dataset and is effective in resolving the Dikarya. However, there is poorer resolution of some of the basal phyla due to smaller taxon sampling perhaps having a negative influence on the distribution of basal taxa within our source phylogenies (we return to this in Section 3). Overall our MRP analysis is highly congruent with our supermatrix phylogenies detailed earlier, with some variation in the placement and resolution in some branches. We discuss the results of our MRP analysis for the basal fungal lineages and both Dikarya

phyla and note some of the congruences and incongruences where noteworthy with our supermatrix phylogenies (Figs. 3–5).

2.2.3.1 Basal Fungi

After rooting at *R. allomycis*, the Neocallimastigomycota and Chytridiomycota (bar *G. prolifera*) emerge as the earliest-diverging fungal lineages. *G. prolifera* branches basal to the Blastocladiomycota with 73% BP (Fig. 5). This arrangement of the Neocallimastigomycota, Chytridiomycota, and Blastocladiomycota has poor support in general (43% BP for a sister relationship between Neocallimastigomycotina and 4 Chytridiomycota species); however with the exception of the aforementioned placement of *G. prolifera* the individual phyla receive maximum or near-maximum support as monophyletic (Fig. 5). Zoopagomycota is paraphyletic in our MRP phylogeny; a monophyletic Kickxellomycotina clade receives 74% BP support (Fig. 5), while as in the supermatrix phylogenies (Figs. 3 and 4) Entomophthoromycotina is paraphyletic. In our MRP analysis, *Basidiobolus meristosporus* branches at the base of Mucoromycota and *Conidiobolus thromboides* branches at the base of Dikarya, but those relationships are poorly supported (30% and 39% BP, respectively; Fig. 5). The Glomeromycotina species *R. irregularis* branches sister to the Mortierellomycota representative *M. elongata* with weak support (52% BP), but Murocomycota (the placement of Glomeromycotina, Mortierellomycota, and Mucoromycotina) receives higher support (85% BP). The monophyly of Mucoromycotina is also fully supported (Fig. 5). Overall many of the associations between basal phyla we observed in our supermatrix phylogenies are present in our MRP analysis as well; however, the overall placement of the basal fungal lineages varies between supermatrix and MRP analyses, such as the placement of Blastocladiomycota as a later-diverging clade than either Chytridiomycota or Neocallimastigomycota under MRP supertree analysis (Figs. 3–5).

2.2.3.2 Basidiomycota

The Basidiomycota are recovered with maximum support in our MRP phylogeny (Fig. 5). The Pucciniomycotina emerge as the most basal subphylum with maximum support, with *Mixia osmundae* branching at the base of the subphylum and *Puccinia graminis* placed as sister to the Microbotryomycetes (who are monophyletic with 97% BP). This reflects the topology of Pucciniomycotina seen in our supermatrix phylogenies (Figs. 3–5). The Ustilagomycotina and Agaricomycotina branch as sister subphyla with 99% BP and both are monophyletic; the former is fully supported

at the branch level and the latter has 94% BP. *M. sympodialis* is placed at the base of Ustilagomycotina, reflecting the resolution of the Ustilagomycotina under ML supermatrix analysis (Figs. 3 and 5). In the Agaricomycotina, *W. sebi* and *B. undulatus* branch at the base of the subphylum with maximum support. The three larger classes from Agaricomycotina in our dataset (Agaricomycetes, Dacrymycetes, and Tremellomycetes) are all monophyletic and are recovered with maximum support (Fig. 5). The MRP phylogeny of the Basidiomycota is highly congruent overall with the supermatrix phylogenies, with comparable branch support (Figs. 3–5).

2.2.3.3 Ascomycota

Our MRP phylogeny supports the Ascomycota as a monophyletic group with maximum BP (Fig. 5). There is greater support along many deeper branches in the Ascomycota in our MRP phylogeny than in our ML supermatrix phylogeny and support is comparable with our Bayesian phylogeny; we ascribe this to a larger abundance of smaller source phylogenies containing closely related Ascomycotina species in our dataset (Figs. 3–5). Taphrinomycotina emerges as the earliest-diverging lineage but is paraphyletic; *Saitoella complicata* branches as an intermediate between Taphrinomycotina and a Saccharomycotina–Pezizomycotina clade with 98% BP, while the remaining members are monophyletic with weak support (58% BP). *Pneumocystis jirovecii* is placed as a sister taxon to Schizosaccharomycetes in our MRP analysis with weak support (36% BP); in the supermatrix phylogenies it was sister to Taphrinomycetes. The Taphrinomycetes and Schizosaccharomycetes themselves are monophyletic with maximum BP (Fig. 5). The Saccharomycotina are monophyletic with 99% BP (Fig. 5). The six larger classes (i.e., all bar Xylonomycetes) in our dataset from Pezizomycotina are all supported as monophyletic and receive maximum BP, with Pezizomycetes and Orbiliomycetes branching as the basal sister clades (Fig. 5). The MRP phylogeny mirrors Bayesian supermatrix reconstruction in placing a single origin for three classes (Xylonomycetes, Eurotiomycetes, and Dothideomycetes) with maximum support (Figs. 4 and 5). As in both supermatrix phylogenies, Dothideomycetes are split into two clades with high or maximum support. In the Sordariomycetes, MRP analysis reflects the ML supermatrix phylogeny in placing *Hypoxylon* sp. EC58 at the base of the class (Figs. 3 and 5). The MRP phylogeny of the Ascomycota is highly congruent with both of our supermatrix phylogenies with comparable branch supports, which is aided by the broad range of genomic data available for the phylum (Figs. 3–5).

2.2.4 Average Consensus Phylogenomic Reconstruction of 84 Fungal Species Is Affected by Long-Branch Attraction Artifacts

To complement our MRP phylogeny, we generated an average consensus (AV) method supertree phylogeny (Fig. 6) using the same set of input phylogenies as implemented in CLANN following Fitzpatrick et al. (2006). AV phylogeny infers phylogeny based on the branch lengths of source phylogenies, by computing the average value of the path-length matrices associated with said source phylogenies, and then using a least-squares method to find the source matrix closest to this average value (Lapointe & Cucumel, 1997). The tree that is associated with this source matrix is the average consensus phylogeny for the total set of source phylogenies, and the method is thought to work best with a set of source phylogenies of similar size (Lapointe & Cucumel, 1997). Our AV phylogeny was rooted at *R. allomycis* (Fig. 6). Given the results we obtained from our AV phylogeny, we believe that the method is susceptible to long-branch attraction (Felsenstein, 1978), as reported by Fitzpatrick et al. (2006). Long-branch attraction occurs when two very divergent taxa or clades with long branch lengths (i.e., many character changes occurring over time) are inferred as each other's closest relative due to convergent evolution of a given character (e.g., amino acid substitution), and is a common problem in parsimony and distance-based methods (Bergsten, 2005; Felsenstein, 1978). In the AV phylogeny, we recovered the two Blastocladiomycota species in our dataset within a large paraphyletic Pezizomycotina clade (Fig. 6). Additionally, the Ascomycota are paraphyletic: one clade containing two Pezizomycotina classes (Pezizomycetes and Orbiliomycetes), the Taphrinomycotina and the Saccharomycotina species *Lipomyces starkeyi* places at the base of Dikarya, while three Saccharomycotina species (including *S. cerevisiae*) appear as a sister clade to Pucciniomycotina (Fig. 6). The Agaricomycotina are also paraphyletic; Tremellomycetes and two basal Basidiomycota species (*B. undulatus* and *W. sebi*) appear closer to Ustilagomycota (Fig. 6). Many of the supports throughout the tree are extremely poor (almost all of the incongruences we highlighted all have <40% BP), which seems to be another effect of long-branch attraction (Fig. 6). Due to the breadth of fungal taxa, we have sampled for our multiple analyses, and the timescale of the evolution of the fungal kingdom being approximately 1 billion years old, it is unsurprising that a method using branch lengths to infer a close relationship between actually distantly related species that both have long branches, a classic example of the "Felsenstein Zone" (Bergsten, 2005; Huelsenbeck & Hillis, 1993). Ultimately, our AV phylogeny (Fig. 6) seems to confirm one of the concerns of Fitzpatrick et al. (2006) in a much more stark fashion that the AV method is not appropriate

for large-scale phylogenomic reconstructions containing taxa sampled from across many phyla without prior predictive analysis of the potential for long branch attraction in such datasets (Su & Townsend, 2015).

2.3 Bayesian Supertree Phylogenomic Analysis of Fungi

While parsimony-based supertree reconstructions are generally reliable, concerns have been raised in the past as to some of the underlying methodology of MRP reconstruction and the effects that factors like input tree sizes (Pisani & Wilkinson, 2002; Wilkinson, Thorley, Pisani, Lapointe, & McInerney, 2004). There has long been the desire for a supertree method that infers phylogeny from source trees with more statistical rigor like Bayesian and maximum-likelihood inference methods. While Bayesian and ML analyses are the standard for supermatrix reconstruction, such methods have been difficult to implement in the past for supertree analysis due to computational limitations, most of which is down to the necessity of tree searching for the best supertree (i.e., calculating likelihoods for all possible supertrees given a set of source phylogenies).

It is only in recent years that phylogenomic inference based on ML and Bayesian methods has been implemented for supertree analysis; one such model for supertree likelihood estimation was first described by Steel and Rodrigo (2008) and then refined the following year (Bryant & Steel, 2009; Steel & Rodrigo, 2008). The Steel and Rodrigo method of likelihood estimation (henceforth referred to as ST-RF) is based on modeling the incongruences between input gene phylogenies and a corresponding unknown or provided supertree phylogeny. Two recent implementations of ST-RF ML analysis have been reported: the first a heuristic method of estimating approximate ML supertrees based on subtree pruning and regrafting implemented in the Python software L.U.St. by Akanni, Creevey, Wilkinson, and Pisani (2014), and the second a heuristic Bayesian MCMC criterion by Akanni, Wilkinson, Creevey, Foster, and Pisani (2015) implemented in the Python software package p4 (Akanni et al., 2014, 2015; Foster, 2004). Akanni et al. (2015) tested the Bayesian MCMC implementation on both a large kingdom-wide metazoan dataset and a smaller Carnivora dataset, notably the analysis produced a Bayesian supertree in full agreement with both the literature on metazoan relationships and a previous MRP supertree analysis on the same dataset (Holton & Pisani, 2010).

No parametric supertree reconstruction has been carried out for the fungal kingdom to date, and with that in mind we reconstructed the phylogeny

of our 84-genome dataset with the MCMC Bayesian criterion developed by Akanni et al. (2015) using a slightly amended gene phylogeny dataset from our MRP and AV supertree phylogenies.

2.3.1 Heuristic MCMC Bayesian Supertree Reconstruction of 84 Fungal Genomes From 8050 Source Phylogenies

MCMC Bayesian supertree analysis was carried out on the single-copy phylogeny dataset using the ST-RF model as implemented in p4 (Akanni et al., 2015; Foster, 2004; Steel & Rodrigo, 2008). As ST-RF analysis is currently only implemented in p4 for fully bifurcating phylogenies, 60 phylogenies were removed from the total single-copy phylogeny dataset, for an input dataset of 8050 gene phylogenies. Two separate MCMC analyses with 4 chains each were ran for 30,000 generations with $\beta=1$, sampling every 20 generations. The analyses converged after 30,000 generations, and a consensus phylogeny based on posterior probability of splits was generated from 150 supertrees sampled after convergence following Akanni et al. (2015). This consensus phylogeny was visualized in iTOL and annotated according to the NCBI's taxonomy database, and rooted at *R. allomycis* (Fig. 7).

2.3.2 Supertree Reconstruction With a Heuristic MCMC Bayesian Method Highly Congruent With MRP and Supermatrix Phylogenies

Using 8050 of the 8110 individual gene phylogenies which we identified in our MRP supertree analysis, we have reconstructed the first parametric supertree of the fungal kingdom (Fig. 7). We selected the ST-RF MCMC Bayesian supertree reconstruction method implemented in p4 for reconstruction over the heuristic method implemented in L.U.St. due to tractability issues regarding large datasets in the latter method (Akanni et al., 2014, 2015). Two ST-RF analyses were carried out for 30,000 generations, and the analyses were adjudged to have converged after 20,000 generations. To construct a phylogeny from our MCMC analysis, we sampled 150 trees generated after convergence and built a consensus tree in p4, where branch support values are the estimated posterior probabilities of a given split (i.e., bipartition) within a phylogeny (Fig. 7). Our ST-RF MCMC analysis is highly congruent with both our MRP supertree phylogeny and supermatrix phylogenies and supports the monophyly of the majority of the eight fungal phyla in our dataset (Fig. 7). Below, we detail the resolution of the basal and Dikarya lineages under ST-RF analysis.

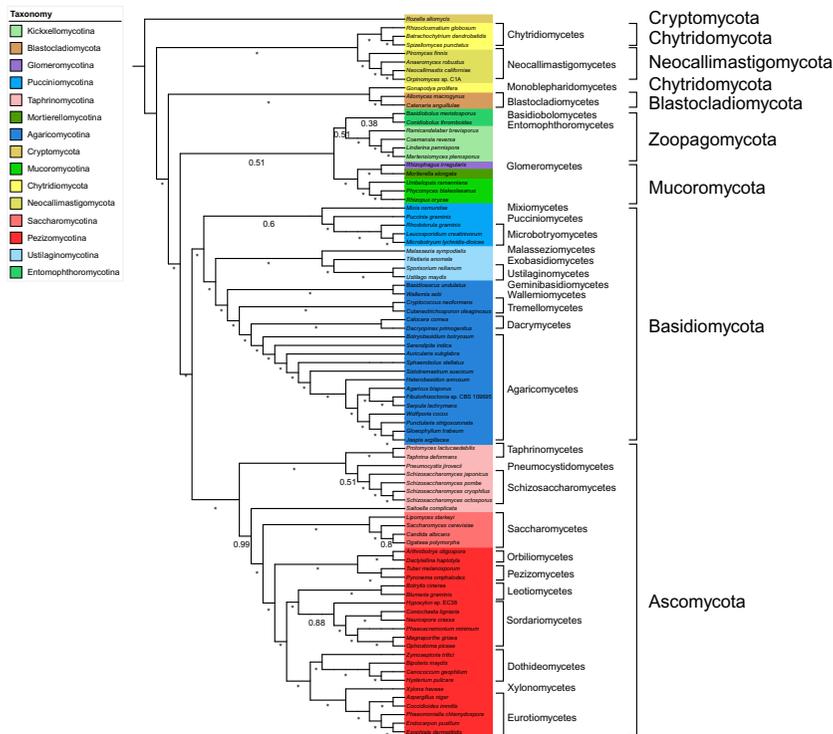


Fig. 7 MCMC Bayesian supertree phylogeny of 84 fungal species derived from 8050 fully bifurcating source phylogenies. Phylogeny generated in p4 using ST-RF model of maximum-likelihood supertree estimation running for 30,000 generations with $\beta=1$. Posterior probabilities of bipartition(s) within 150 trees sampled after convergence shown on branches. Maximum posterior probability support designated with an asterisk (*).

2.3.2.1 Basal Fungi

After rooting at *R. allomycis*, the Neocallimastigomycota and Chytridiomycota (except *G. prolifera*) form a sister group relationship with maximum PP (Fig. 7). The Blastocladiomycota emerge after this branch, and the Chytridiomycota species *G. prolifera* branches as sister to the phylum with maximum PP (Fig. 7). There is weak support (0.51 PP) for a monophyletic clade containing both former zygomycetes phyla Zoopagomycota and Mucoromycota as sister clades (Fig. 7). Notably, unlike MRP and supermatrix analysis, ST-RF phylogeny places the Entomophthoromycotina as monophyletic but with very weak support (0.38 PP). There is also weak support for the placement the Entomophthoromycotina as basal within Zoopagomycota. Kickxellomycotina are monophyletic with maximum

support. The monophyly of Mucoromycota is fully supported, with *R. irregularis* (Glomeromycotina) and *M. elongata* (Mortierellomycotina) branching as sister taxa (Fig. 7).

2.3.2.2 Basidiomycota

The Basidiomycota are supported as a monophyletic group with maximum PP (Fig. 7). There is weak support for the monophyly of Pucciniomycotina (0.6 PP); however, the deeper branches within the subphyla are all fully supported and their topology reflects both the MRP supertree and ML supermatrix phylogenies discussed earlier (Figs. 3, 5, and 7). There is full support for a sister relationship between Ustilaginomycotina and Agaricomycotina, and both these subphyla are fully supported. In Ustilaginomycotina, *M. sympodialis* is the basal species with maximum support (Fig. 7), as in our supermatrix and MRP supertree phylogenies. The topology of the Agaricomycotina is nearly identical on the class level to both the MRP and supermatrix phylogenies, with *B. undulatus* and *W. sebi* branching as basal species, the Tremellomycetes forming a monophyletic intermediate clade, and a fully supported sister relationship between the Dacrymycetes and the Agaricomycetes (Fig. 7).

2.3.2.3 Ascomycota

The monophyly of the Ascomycota is supported with maximum PP, as is the monophyly of two of the three subphyla in Ascomycota (Fig. 7). Taphrinomycotina is paraphyletic as in the MRP phylogeny, with *S. complicata* branching sister to Saccharomycota with near-maximum support (0.99 PP) and the remaining Taphrinomycotina species are placed as a monophyletic clade with maximum PP (Figs. 5 and 7). The Taphrinomycetes branch at the base of the Taphrinomycotina clade, and there is weak support (0.51 PP) for the placement of *P. jirovecii* as sister to the Schizosaccharomycotina (Fig. 7). The Saccharomycotina are fully supported as monophyletic (1.0 PP) with *L. starkeyi* placed at the base of the subphyla. The monophyly of the Pezizomycotina is also fully supported and there is maximum support for the monophyly of the six larger represented classes within the subphylum (Fig. 7). Additionally, the relationships between the individual classes within Pezizomycotina are identical to the topology seen in both the MRP supertree phylogeny and the ML supermatrix phylogeny (Figs. 3, 5, and 7). The Orbiliomycetes and Pezizomycetes branch as the earliest-diverging clades within Pezizomycotina with maximum PP, the Sordariomycetes and Leotiomycetes are sister classes with maximum PP and a monophyletic

Dothideiomycetes–Xylonomycetes–Eurotiomycetes clade receives maximum PP (Fig. 7).

2.4 Phylogenomics of Fungi Based on Gene Content

A common alternative to phylogenomic reconstruction using gene phylogenies is to take a “gene content” approach in which evolutionary relationships between species are derived from shared genomic content, such as the presence or absence of conserved orthologous genes (COGs) or the overall proportion of shared genes between two species, working under the assumption that species that share more of their genome are closely related (Snel, Bork, & Huynen, 1999; Snel, Huynen, & Dutilh, 2005). In the case of presence–absence analyses, a matrix can be constructed for the species under investigation, which can then have their phylogeny reconstructed via parsimony methods. Analyses based on proportions of shared genes can entail the construction of distance matrices for all input species, with values equal to the inverse ratio of shared genes (i.e., if two species share 75% of their genes, their distance is 0.25), which is then used to construct a neighbor-joining phylogeny. The advantages of such approaches are the relative tractability of parsimony or distance-based gene content methods, and their potential to use more information from genomes rather than the sourcing of data from smaller sets of gene families required by supertree or supermatrix approaches (Creevey & McInerney, 2009). However, the gene content approach is by its very nature a “broad strokes” approach and can ignore potentially important phylogenetic information from individual gene phylogenies such as HGT events, and assumes the same evolutionary history for missing orthologs or genomic content among species (Page & Holmes, 1998).

2.4.1 Gene Content Approaches to Phylogenomics in Fungi

Gene content approaches to phylogenomic reconstruction have seen application in a number of phylogenomics studies, although its greatest use predated many of the now common supertree and supermatrix methods. One of the earliest phylogenomic studies used a distance-based approach based on shared gene content to reconstruct the phylogeny of 13 unicellular species, including *S. cerevisiae* (Snel et al., 1999). Another study used a weighted distance matrix approach to reconstruct the phylogeny of 23 prokaryote and eukaryote species, including *S. cerevisiae* and partial genomic data from *S. pombe* (Tekaiia, Lazcano, & Dujon, 1999). The most extensive gene content-based phylogenomic reconstruction of fungi was an analysis of 21 fungal genomes and 4 other eukaryote genomes in 2006 (Kuramae et al., 2006). In their

analysis, the authors generated a presence–absence matrix (PAM) of 4852 COGs in fungal genomes as a complement to a supermatrix phylogeny using 531 concatenated proteins which was reconstructed using four different methods (MP, ML, neighbor-joining, and Bayesian inference). The authors reconstructed the phylogeny of all 25 genomes using this PAM and found that the PAM phylogeny differ most in the placement of *S. pombe* within Saccharomycetes as opposed to its basal position in Ascomycetes as seen in their supermatrix reconstructions (Kuramae et al., 2006).

To test the accuracy of inferring the phylogeny of a large genomic dataset using simple parsimony methods based on shared genomic content, we carried out a simple parsimony-based PAM phylogenomic reconstruction of 84 fungal species based on the presence of orthologs from single-copy gene families.

2.4.2 Phylogenomic Reconstruction of 84 Fungal Species Based on COG PAM

A simple PAM was generated for 84 fungal genomes based on their representation across 12,964 single-copy gene families identified via the random BLASTp approach detailed in Section 2.2. Parsimony analysis of this matrix was carried out using PAUP* with 100 bootstrap replicates. The resultant consensus phylogeny generated by PAUP* was visualized using iTOL and annotated according to the NCBI's taxonomy database. The phylogeny was rooted at *R. allomyces* (Fig. 8).

2.4.3 COG PAM Approach Displays Erroneous Placement of Branches Within Dikarya

We generated a simple PAM phylogeny for the 84 fungal genomes in our dataset by checking for the presence or absence of all 84 species across the 12,964 single-copy phylogenies we generated during our supertree analyses via random BLASTp searches and using the PAM as input for parsimony analysis (Fig. 8). The simple PAM phylogeny shows some level of congruence with the other phylogenomic analyses described here along certain branches (Fig. 8). The monophyly of Neocallimastigomycota, Chytridiomycota, and Blastocladiomycota all displays maximum or near-maximum BP, and there is 72% BP for a sister relationship between Chytridiomycota and Neocallimastigomycota (Fig. 8). The Zoopagomycota and Mucoromycota are placed in one monophyletic clade with 82% BP, with the two Entomophthoromycotina species in our dataset branching as closely related to the Mucoromycota (Fig. 8). However, some glaring conflicts with the

being the Sordariomycetes and Eurotiomycetes (Fig. 8). In short, our PAM phylogeny is able to retrieve relationships with some level of accuracy within the fungal kingdom, but the method lacks the ability to resolve some of the more divergent relationships within fungi to the degree that some of our supermatrix or supertree phylogenies have illustrated.

2.5 Alignment-Free Phylogenomic Analysis of Fungi

Another alternative to the alignment-based methods of phylogenomic reconstruction we have detailed earlier is the use of a string-based comparison of genomes to infer phylogeny, based on the assumption that under such comparisons each species should have a characteristic genomic signature that can act as a phylogenetic marker (Delsuc, Brinkmann, & Philippe, 2005). Some analyses have thus used signatures such as distribution of protein folds or frequency of oligonucleotides from genetic and genomic data to infer phylogeny (Campbell, Mrázek, & Karlin, 1999; Lin & Gerstein, 2000; Pride, Meinersmann, Wassenaar, & Blaser, 2003). The most widely used alignment-free phylogenomic method, the composition vector (CV) approach, was first implemented by Qi, Luo, and Hao (2004) and by Qi, Wang, and Hao (2004), who used the approach to reconstruct the phylogeny of 87 prokaryote species from 11 bacterial and 2 archaeal phyla (Qi, Wang, et al., 2004). In their analysis, the authors detail the CV method for reconstructing phylogeny using genome-scale data, which we recount as follows:

1. Given a nucleic acid or amino acid sequence of length L in a genome, count the appearances of overlapping strings (i.e., oligonucleotides or oligopeptides) of a length K and construct a frequency vector of length 4^K for nucleic acid sequences and 20^K for amino acid sequences.
2. Subtract background noise, to account for random mutation at the molecular level, from each frequency vector to generate an overall composition vector for a given genome.
3. Calculate a distance matrix for the set of composition vectors corresponding to the set of input genomes.
4. Generate a neighbor-joining phylogeny from the distance matrix using software such as Neighbor or PAUP*.

The main advantages of the composition vector approach over traditional alignment-based methods of inferring phylogeny are the removal of artificial selection of phylogenetic markers from the process of reconstruction (the only variable in the method is K , the length of overlapping oligopeptides), and the relative speed with which the approach can infer phylogeny for large

datasets over alignment-based supertree or supermatrix methods. Hence, it may be useful for quick phylogenomic identification of newly sequenced genomes against published data and as an independent verification step of previous alignment-based phylogenetic or phylogenomic analysis (Wang, Xu, Gao, & Hao, 2009). On that point however, interpreting the accuracy or otherwise of CV phylogenomic reconstructions is generally dependent on prior knowledge of the phylogeny of given taxa derived from alignment-based phylogenetic or phylogenomic analyses. An approach to inferring phylogeny based on nucleotide or amino acid composition may also be susceptible to compositional biases, and there has not been to the best of our knowledge a rigorous analysis of the potential effect these may have on accuracy of phylogenomic inference, as there have been for the supertree or supermatrix methods referred to earlier.

2.5.1 Composition Vector Method Phylogenomics of Fungi

Many of the phylogenomic analyses using the CV method have analyzed large prokaryotic datasets or broad global datasets sampled from many phyla or kingdoms across the three domains of life, whose phylogenies were recovered with quality comparative to alignment-based phylogenomic analyses. The most extensive application of the composition vector approach in fungal phylogenomics was an 85-genome analysis by Wang et al. (2009) using a CV implementation in the software program CVTree (Qi, Luo, et al., 2004; Wang et al., 2009). For their analysis, Wang et al. (2009) reconstructed the phylogeny of the fungal kingdom using 81 genomes from 4 fungal phyla (Basidiomycota, Ascomycota, Chytridiomycota, and Mucoromycota) as well as the microsporidian *Encephalitozoon cuniculi* and 3 eukaryotic outgroup taxa. The authors described the resolution of both the Basidiomycota and Ascomycota in detail in their analysis; the three sub-phyla within Basidiomycota were recovered but with poor bootstrap support due to issues with taxon sampling (only 12 Basidiomycota species had genomic data at the time of the analysis), while the main focus of the authors analysis was on the resolution of 65 Ascomycota species. Within the Ascomycota, the Taphrinomycota (represented by three *Schizosaccharomyces* species) were fully resolved and in the Saccharomycotina the two clades described by Fitzpatrick et al. (2006), the CTG clade and the WGD clade, were also recovered. CV reconstruction recovered four classes within Pezizomycotina; the Dothideomycetes and Eurotiomycetes were placed as sister taxa with maximum support, as were the Sordariomycetes and Leotiomycetes.

To complement our phylogenomic analyses based on source gene phylogenies or identification of shared orthologs, we carried out alignment-free analysis of 84 fungal species using the composition vector method as implemented in CVTree.

2.5.2 Phylogenomic Reconstruction of 84 Fungal Species Using the CV Approach

Composition vector analysis was carried out on 84 genomes using CVTree with $K=5$ (Qi, Luo, et al., 2004). We selected $K=5$ as the best compromise of both computational requirements and resolution power. As the CV method does not generate bootstrapped phylogenies, we generated 100 bootstrap replicates of our 84-genome representative dataset using bespoke Python scripting and ran composition vector analysis on each replicate dataset (Zuo, Xu, Yu, & Hao, 2010). 100 replicate neighbor-joining phylogenies were calculated from their corresponding CVTree output distance matrices using Neighbor (Felsenstein, 1989). The majority-rule consensus phylogeny for all 100 composition vector replicate trees was generated using Consense (Felsenstein, 1989) and was visualized in iTOL, and annotated according to the NCBI's taxonomy database. The phylogeny was rooted at *R. allomycis* (Fig. 9).

2.5.3 Composition Vector Phylogenomic Reconstruction of 84 Fungal Species Is Congruent With Alignment-Based Methods

We carried out composition vector method phylogenomic reconstruction of our 84-genome dataset to complement the alignment-based and genomic content methods we detailed earlier (Fig. 9). Our composition vector analysis displays adequate levels of taxonomic congruence with our supermatrix and supertree analyses detailed in previous sections, supporting all the monophyly of each major fungal phyla and many of the subphyla within (Fig. 9). There are however some variations in topology and support between the basal lineages and within the Pezizomycotina subphylum in our CV phylogeny compared to our supermatrix and supertree phylogenies.

2.5.3.1 Basal Fungi

After rooting at *R. allomycis*, the Neocallimastigomycota emerge as the earliest-diverging fungal lineage (Fig. 9). The monophyly of Neocallimastigomycetes is also fully supported. Monophyletic Blastocladiomycota and Chytridiomycota clades branch as sister phyla with 62% BP. The monophyly of

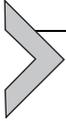
seen under PAM phylogenomic analysis (Figs. 8–9). Like the MRP phylogeny (Fig. 5), *R. irregularis* is within a paraphyletic Mucoromycota clade instead of at the base of the Dikarya as seen in the supermatrix phylogenies (Figs. 3, 4, and 9).

2.5.3.2 Basidiomycota

Pucciniomycotina is placed as the earliest-diverging subphylum within Basidiomycota with 52% BP, and the Ustilagomycotina and Agaricomycotina subphyla are sister clades with 95% BP (Fig. 9). The most-represented class within the Pucciniomycotina, the Microbotryomycetes, is monophyletic with 65% BP (Fig. 9), while unlike the rest of our phylogenies discussed earlier *P. graminis* is placed as the most basal species within Pucciniomycotina. Within the Ustilaginomycotina, *M. sympodialis* are placed as the basal lineage sister to the Exobasidiomycetes representative *T. anomala* similar to its position under ML supermatrix reconstruction and MRP reconstruction (Figs. 3, 5, and 9). The Agaricomycetes are monophyletic with 84% BP, with varying support for relationships within the class but a topology identical to both supermatrix phylogenies and MRP phylogeny with the exception of the placement of Tremellomycetes within a monophyletic ancestral branch adjacent to *B. undulatus* and *W. sebi* (Figs. 3–5 and 9).

2.5.3.3 Ascomycota

Within the Ascomycota, all three subphyla are resolved as monophyletic clades (Fig. 9). Taphrinomycotina is placed as the most basal subphylum within Ascomycota with maximum support, while the Pezizomycotina and Saccharomycotina are sister subphyla with 80% BP (Fig. 9). The Taphrinomycotina are monophyletic with 80% BP, and CV phylogeny displays maximum support for a sister relationship between *P. jirovecii* and the Schizosaccharomycetes and near-maximum (96% BP) support for a similar relationship between *S. complicata* and the two Taphrinomycetes representatives in our dataset (Fig. 9). The Saccharomycotina are monophyletic with 74% support (Fig. 9). All six larger classes from the Pezizomycotina represented in our dataset are resolved as monophyletic. The Orbiliomycetes and Pezizomycetes are placed as both sister subphyla and the earliest-diverging Pezizomycotina clades, both with maximum BP. The Leotiomycetes and Sordariomycetes are also sister clades with 95% BP. As our MRP phylogeny, the Eurotiomycetes are placed as sister to the Xylonomycetes species *X. heveae* with 97% BP (Figs. 5 and 9).



3. A GENOME-SCALE PHYLOGENY OF 84 FUNGAL SPECIES FROM SEVEN PHYLOGENOMIC METHODS

There is a large degree of congruence in the resolution of the fungal kingdom in most of the phylogenomic analyses we described in [Section 2](#), which speaks to the quality of the genomic data we obtained from MycoCosm and the relative accuracy of the majority of the phylogenomic methods we utilized. In constructing a dataset for our analyses, we selected one representative from as many fungal orders as had been sequenced to date; this was to generate a phylogeny that was representative on the order level (though we do not focus on order phylogeny in this review) and to avoid overrepresentation of highly sampled taxa such as Eurotiomycetes or Saccharomycotina. Many of the best-known phylogenetic relationships within the fungal kingdom were recovered in our analyses, such as the monophyly of Dikarya as a whole ([Hibbett et al., 2007](#)). However, our analyses also supports more recent studies that have attempted to resolve outstanding branches of the fungal tree of life ([Spatafora et al., 2016](#)). In this section, we briefly describe the main trends seen across our seven phylogenomic reconstructions of the fungal kingdom and their congruence with previous studies and comment on the reconstructions of both the well-studied and highly represented Pezizomycotina subphylum and some of the newly circumscribed basal phyla. Finally, we discuss the suitability of the phylogenomic methods we have described and applied in this review for future fungal systematics studies.

3.1 Higher-Level Genome Phylogeny of the Fungal Kingdom

Despite variations in the resolution of some branches, there is a trend across the majority of phylogenies conducted of support or partial support for the eight phyla described in our dataset. [Fig. 10](#) shows the congruence on the phylum level within the fungal kingdom in five of our seven phylogenetic reconstructions. We will refer to [Fig. 10](#) and the subfigures ([Figs. 10A–D](#)) in [Fig. 10](#) when comparing the different reconstructions on the phylum level and to the corresponding full phylogenies themselves for comparisons at lower levels here and elsewhere (average consensus and gene content phylogenies are omitted from [Fig. 10](#) on the basis of erroneous placement of taxa). Beginning with the Cryptomycota species *R. allomycis*, the next-earliest-diverging clade within the fungal kingdom is the Blastocladiomycota under both supermatrix analyses followed by Neocallimastigomycota and Chytridiomycota

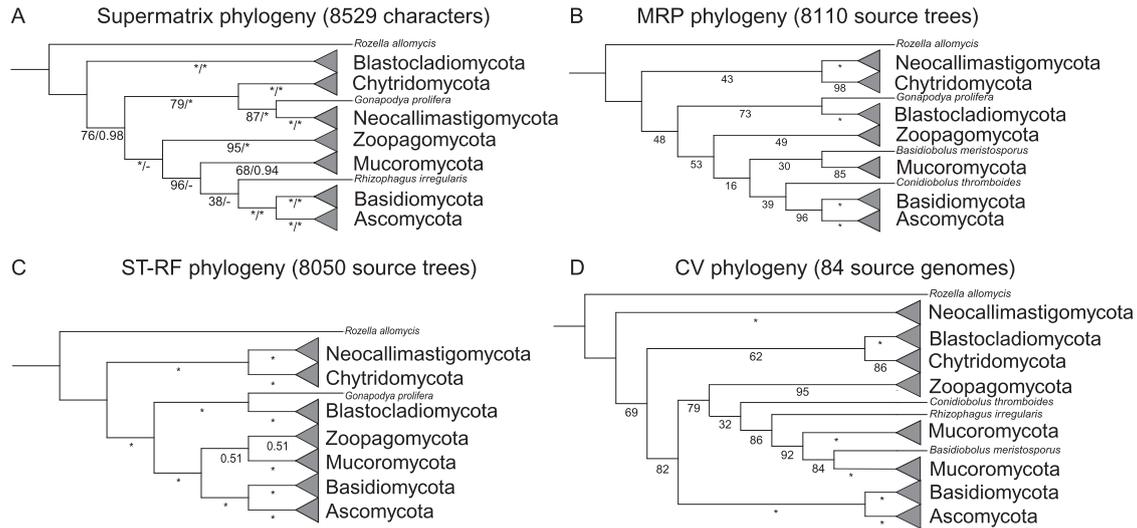
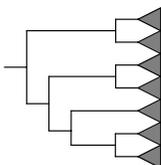


Fig. 10 Congruence of eight fungal phyla under five phylogenomic reconstructions. All clades bar Cryptomycota (represented *Rozella allomycis*) collapsed by phylum, paraphyletic species displayed as individual leaves. *Gonapodya prolifera* = Chytridiomycota, *Rhizophagus irregularis* = Mucoromycota, all other species except *R. allomycis* = Zoopagomycota. Refer to Figs. 3, 4, 5, 7, and 9, respectively, for original phylogenies. (A) ML and Bayesian supermatrix phylogenies. Branch supports given as ML bootstrap supports and, where topology is identical, Bayesian posterior probabilities. Maximum bootstrap or posterior probability support designated with an asterisk (*). (B) MRP supertree phylogeny. Branch supports given as bootstrap supports. Maximum bootstrap support designated with an asterisk (*). (C) MCMC Bayesian supertree phylogeny using ST-RF ML method. Branch supports given as posterior probabilities of bipartition(s). Maximum posterior probability support designated with an asterisk (*). (D) CV phylogeny. Branch supports given as bootstrap supports. Maximum bootstrap support designated with an asterisk (*).



Pezizomycotina	Supermatrix		Supertree			PAM	CV
	ML	PB	MRP	AV	ST-RF		
Orbiliomycetes	✓	✓	✓	✗	✓	✓	✓
Pezizomycetes	✓	✓	✓	✗	✓	✓	✓
Leotiomycetes	✓	✓	✓	✗	✓	✓	✓
Sordariomycetes	✓	✓	✓	✗	✓	✓	✓
Dothideomycetes	✓	Xy	✓	✗	✓	Eu	✓
Xylonomycetes	✓	Do	✓	✗	✓	✓	✓
Eurotiomycetes	✓	✓	✓	✗	✓	Do	✓

Fig. 11 Congruence of Pezizomycotina under seven phylogenomic methods. Placement of classes identical to topology on the left (see text) indicated with a *tick*, varying placement of classes indicated by the first two letters of a class. Average consensus (AV) phylogeny produced paraphyletic Pezizomycotina and so entire column labeled with *crosses*. Refer to text for discussion of topology of Pezizomycotina under AV phylogeny. Refer to Figs. 3–9 for original phylogenies.

(Fig. 10A). Other analyses place Neocallimastigomycota and Chytridiomycota (except *G. prolifera*) as closest to *R. allomyis* (Fig. 10B–D).

We describe the resolution of the former zygomycetes in greater detail later, but in the five phylogenies in Fig. 10 all support at least a sister relationship between the two zygomycetes phyla Zoopagomycota and Mucoromycota. The placement of the Glomeromycotina species *R. irregularis* varies, but Mucoromycota is generally placed as sister to the Dikarya (Fig. 10). The Basidiomycota are fully supported as monophyletic in each of the five phylogenies represented in Fig. 10, and all bar ML supermatrix reconstruction is in exact agreement with the two most extensive fungal genome phylogenies containing all three Basidiomycota subphyla (Medina et al., 2011; Wang et al., 2009). The Ascomycota are also fully supported as monophyletic in each of the five phylogenies represented in Fig. 10, with the only major variation being the placement of *S. complicata* within (or paraphyletic to) Taphrinomycotina (Fig. 10). The Saccharomycotina are monophyletic in all five phylogenies (Fig. 10). We discuss the class-level phylogeny within Pezizomycotina in greater detail in Section 3.3 and Fig. 11, but to briefly summarize here we see strong-to-maximum support for all six of the larger classes that were present in our dataset, and support for the two unofficial “Sordariomyceta” and “Dothideomyceta” groupings within Pezizomycotina (Schoch et al., 2009).

3.2 Multiple Phylogenomic Methods Show Moderate Support for the Modern Designations of Mucoromycota and Zoopagomycota

There is moderate support for the recent designation of the zygomycetes phyla Zoopagomycota and Mucoromycota by Spatafora et al. (2016) across

most of our phylogenomic methods (Fig. 10). Previously the species within these two phyla were classified within Zygomycota, a phylum-level classification that had dated back to the 1950s until it was formally disputed by Hibbett et al. (2007). Six *incertae sedis* zygomycetes subphyla were later circumscribed (Hoffmann, Voigt, & Kirk, 2011), and subsequent phylogenetic analyses informally classified the zygomycetes subphyla into two groups, which were later established as Mucoromycota and Zoopagomycota (Chang et al., 2015; Spatafora et al., 2016).

Our phylogenomic analyses included 11 species from the 2 zygomycetes phyla, with the best resolution found in the ST-RF phylogeny where Zoopagomycota and Mucoromycota are placed as sister phyla with 0.51 PP and branch sister to Dikarya (Fig. 10C). Notably, our ST-RF phylogeny is the only phylogeny that resolves Entomophthoromycotina as a monophyletic clade (Fig. 7), albeit with extremely weak posterior probability support (0.38 PP). Within Zoopagomycota in our ST-RF phylogeny, Entomophthoromycotina branch as the basal clade with 0.51 PP, sister to Kickxellomycotina (Fig. 7). Our ST-RF phylogeny also places *R. irregularis* (Glomeromycotina) adjacent to *M. elongata* (Mortierellomycotina) within the Mucoromycota (Fig. 7). Within Mucoromycota, Mortierellomycotina and Mucoromycotina are supported as sister subphyla throughout the majority of our phylogenies (e.g., Bayesian supermatrix analysis, Fig. 4), with high to maximum support. Both of these phylum-level topologies are in agreement with Spatafora et al. (2016), though their phylogeny does not support a distinctive monophyletic branch containing both Zoopagomycota and Mucoromycota (Fig. 10C). The majority of our remaining phylogenomic analysis all shows some degree of support for both Zoopagomycota and Mucoromycota in relative agreement with Spatafora et al. (2016); however, in each of these phylogenies there is some conflict in either subphylum-level topology or lower BP/PP support due to issues of taxon sampling or low gene tree coverage in our dataset (of our 8110 source phylogenies for MRP analysis over 3500 contain 7 taxa or less; Fig. 10). With greater sampling of species from these lineages, we hope to see more consistent support of both the Zoopagomycota and Mucoromycota in future genome phylogenies using these methods, in line with what appears to be moderate-to-strong support for the new classification in our analyses based on total evidence (Kluge, 1989).

3.3 Pezizomycotina as a Benchmark for Phylogenomic Methodologies

The Pezizomycotina are by far the most sampled subphylum within the fungal kingdom in terms of genome sequencing (375 Pezizomycotina species

have genomic data available from MycoCosm as of May 2017). Reflecting this, 22 Pezizomycotina species representing 7 classes are present in our 84-genome dataset (>25% of our final dataset). As a well-represented clade within our dataset at both the subphylum and individual class level, we are able to see how multiple phylogenomic analyses conducted in a total evidence approach (Kluge, 1989) are able to resolve a single clade of closely related classes containing some important ecological and pathogenic fungi. In every phylogenomic reconstruction, we attempted bar average consensus (AV) phylogeny, Pezizomycotina were monophyletic with maximum bootstrap or posterior probability branch support, and every class within Pezizomycotina is monophyletic with high or maximum BP or PP support (Figs. 3–5 and 7–9). There is a consistent trend within each of these phylogenies in the resolution of relationships between Pezizomycetes classes:

1. The Orbiliomycetes and Pezizomycetes always branch as the basal classes within Pezizomycotina and are always sister taxa (Figs. 3–5 and 7–9).
2. The relationship between Sordariomycetes and Leotiomycetes (within “Sordariomyceta” sensu Schoch et al., 2009) is always present and is fully supported in each phylogeny (Figs. 3–5 and 7–9).
3. The relationship between Dothideomycetes, Xylonomycetes, and Eurotiomycetes (within “Dothideomyceta” sensu Schoch et al., 2009) is always present and is fully supported in each phylogeny (Figs. 3–5 and 7–9).

Fig. 11 displays on the left the topology of the Pezizomycotina classes supported under ML supermatrix reconstruction, MRP supertree reconstruction, and ST-RF supertree reconstruction (Figs. 3, 5, and 7) and indicates the congruence (or otherwise) of Pezizomycotina under every phylogenomic analysis we attempted (Figs. 3–9). All methods bar AV are highly congruent in their resolution of the Pezizomycotina subphylum, with placement of the Xylonomycetes class the most notable variation. Even within the highly aberrant AV phylogeny, sister relationships such as those between Orbiliomycetes and Pezizomycetes or the association of classes within Sordariomyceta or Dothideomyceta can still be observed, though with lower resolution and support (Fig. 6). There is a high degree of congruence between our genome phylogenies of Pezizomycotina (Fig. 11) and the most extensive molecular phylogenies of Pezizomycotina that we could find in the literature derived from either small concatenated sets or whole genomes (Medina et al., 2011; Spatafora et al., 2006; Wang et al., 2009). The relative consistency of our analyses both with each other and with previous literature suggests that the resolution of Pezizomycotina could be considered a good benchmark for the accuracy of novel or existing

phylogenomic methods (e.g., ST-RF analysis) when incorporated into a total evidence analysis, as the subphylum is large and diverse (the 10th edition of Ainsworth & Bisby's Dictionary of the Fungi estimates close to 70,000 Pezizomycetes species) but also densely sampled in genomic terms and containing a number of genomes of reference quality (Kirk, Cannon, Minter, & Stalpers, 2008).

3.4 The Use of Phylogenomics Methods in Fungal Systematics

Phylogenomic analyses with larger datasets across a wider spectrum of taxa are becoming more and more computationally tractable as methods of identifying potential phylogenetic markers on a genome-wide scale (e.g., identification and reconstruction of orthologous gene phylogenies in supertree analysis) and genome-scale reconstruction improve. In as much as the majority of our multiple analyses strongly support the major phyla of the fungal kingdom, we can also treat our analyses as measures of the accuracy of each of these phylogenomic methods in the reconstruction of large datasets. Supermatrix, MRP and ST-RF supertree, and CV method reconstructions all appear to arrive at relatively congruent results and may be useful for approximating a total evidence style approach for phylogenomic analyses of fungi. Simplified parsimony methods like our PAM phylogeny or branch length-based methods like our average consensus phylogeny may be useful for the reconstruction of smaller but well-represented datasets (for example, our PAM phylogeny does reconstruct the Pezizomycotina with support and topology close to supertree and supermatrix phylogenies) but for phylum or kingdom-wide analyses issues such as long-branch attraction begin to emerge (Bergsten, 2005). Long-branch attraction is thought to be an issue with MRP reconstruction as well, and while it is likely a factor in the weaker supports in some of the ancestral branches in our MRP phylogeny (for example, the weak supports in some of the internal branches grouping the basal phyla together), the MRP phylogeny seems to have been relatively immune to the topological effects of long-branch attraction that are very apparent in our branch length-dependent average consensus method phylogeny (Pisani & Wilkinson, 2002).

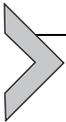
For our supertree analyses, we identified groups of orthologous proteins using a sequential random BLASTp approach as implemented by Fitzpatrick et al. (2006), where a random sequence from a given database is searched against that entire database, and then the sequence and its homologs (if any) are removed and the database reformatted (Fitzpatrick et al., 2006). Overall, this

ad hoc approach to identifying orthology within our dataset seems to have been sufficient as a first step to generating source gene phylogenies; however, it may have had an impact downstream on resolution of internal branches within our MRP analysis. It is possible that a random BLASTp approach is too conservative, in that the orthologous families it identifies are missing members or that two “separate” orthologous families may in fact be one large orthologous family. Other established methods of identifying orthologous families, such as the OrthoMCL pipeline, have been used in phylogenomic analyses and can be tuned for granularity (i.e., orthologous cluster size) which may produce broader source phylogenies (Li, Stoeckert, & Roos, 2003). However, the large SQL-dependent computational overhead required for the current implementation of OrthoMCL was not considered suitable for an analysis of this scale.

Most of the phylogenomic methods we attempted are relatively tractable even for a dataset as large as ours. Depending on computational resources and available data, some of the methods we have discussed may be more appropriate for future fungal phylogenomic analyses than others. The most common techniques like MRP analysis and both ML and Bayesian supermatrix analysis were both tractable and produced phylogenies with largely congruent topologies and supports on most branches (although we should note that we utilized the parallelized version of PhyloBayes for our Bayesian analysis). The heuristic MCMC Bayesian supertree reconstruction we attempted using the ST-RF model as implemented in p4 was also relatively tractable despite not being parallelized, and Akanni et al. (2015) note that the method is far more efficient than the approximate ML reconstruction implemented in L.U.St. (Akanni et al., 2015). However, ST-RF analysis using either p4 or L.U.St. is currently only able to use fully resolved input phylogenies. While in our case this meant only 60 single-copy phylogenies (<1% of our total dataset) had to be removed before carrying out analysis, this may cause issues for more polytomous datasets. Bayesian and ML supertree reconstruction is certainly a promising development for phylogenomics, and hopefully methods like ST-RF should see more widespread use in future phylogenomic analysis as they mature.

Phylogenomic reconstruction using average consensus as implemented in CLANN was extremely inefficient time-wise and returned a severely erroneous phylogeny, so while it is certainly desirable for branch lengths to be incorporated in supertree reconstruction, a branch length-based method like AV is not appropriate for this kind of large-scale analysis. While PAM method reconstruction was straightforward to carry out, as we state earlier there were issues with erroneous placement of taxa and as such we

do not recommend the method for large-scale datasets. Finally, composition vector method analysis produced a phylogeny relatively congruent to our alignment-based methods at $K=5$. Other CV method analyses have recommended K -values between 5 and 7 for most datasets (Zuo, Li, & Hao, 2014), however with the size of our dataset and the increase in computational resources required for generating distance matrices for eukaryotic genomes at $K>5$ in CVTree we felt that $K=5$ was the best compromise between accuracy and computational tractability. We would recommend however as in Section 2.5 that CV analysis should be used in conjunction with alignment-based methods for eukaryotic datasets, as interpretation of CV analysis requires a priori knowledge of the phylogeny of a given dataset.



4. CONCLUDING REMARKS

Fungi make up one of the major eukaryotic kingdoms, with an estimated 1.5 million member species inhabiting a diverse variety of ecological niches and an evolutionary history dating back over a billion years. It is imperative that evolutionary relationships within the fungal kingdom are well understood by analysis of as much quality phylogenetic data as is available with the most accurate methodologies possible. In this chapter, we discussed the evolutionary diversity of the fungal kingdom and the important role that fungi have had in the area of genomic and phylogenomics. We have reviewed previous phylogenomic analyses of the fungal kingdom over the last decade, and using seven phylogenomic methods, we have reconstructed the phylogeny of 84 fungal species across 8 fungal phyla. We found that established supermatrix and supertree methods produced relatively congruent phylogenies that were in large agreement with the literature. We also conducted the first analysis of the fungal kingdom using a heuristic MCMC Bayesian approach to supertree reconstruction previously used in Metazoa and found that this novel supertree approach resolves the fungal kingdom with a high degree of accuracy. The majority of our analyses overall show moderate-to-strong support of the newly assigned zygomycete phyla Mucoromycota and Zoopagomycota and strongly support the monophyly of Dikarya, while within the highly sampled Pezizomycotina subphylum there is a large amount of congruence between different phylogenomic methods as to the resolution of class relationships within the subphylum. We also conclude that supermatrix and supertree analyses remain the exemplar methods of phylogenomic reconstruction for fungi, based on their accuracy and

computational tractability. We believe through both our discussion of the ecological diversity of the fungal kingdom and the history of its study on the genomic level we have demonstrated the need for a robust fungal tree of life with a broad representation, and that through our multiple phylogenomic analysis we have generated an important backbone for future comparative genomic analysis of fungi, particularly with the constantly increasing amount of quality genomic data arising from the 1000 Fungal Genomes Project and its certain use in future studies.

ACKNOWLEDGMENTS

We wish to acknowledge the JGI and all individual contributors to the 1000 Fungal Genomes Project for both the sheer scope of their undertaking and the quantity and quality of genomic data that they have made publically available and that we were able to use in this chapter. We also wish to acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. C.G.P.M. is funded by an Irish Research Council Government of Ireland Postgraduate Scholarship (Grant No. GOIPG/2015/2242).

REFERENCES

- Akanni, W. A., Creevey, C. J., Wilkinson, M., & Pisani, D. (2014). L.U.St: A tool for approximated maximum likelihood supertree reconstruction. *BMC Bioinformatics*, *15*(1), 183. <https://doi.org/10.1186/1471-2105-15-183>.
- Akanni, W. A., Wilkinson, M., Creevey, C. J., Foster, P. G., & Pisani, D. (2015). Implementing and testing Bayesian and maximum-likelihood supertree methods in phylogenetics. *Royal Society Open Science*, *2*(8), 140436. <https://doi.org/10.1098/rsos.140436>.
- Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquandano, G., Richardson, S. M., et al. (2014). Total synthesis of a functional designer eukaryotic chromosome. *Science*, *344*(6179), 55–58. <https://doi.org/10.1126/science.1249252>.
- Baldauf, S. L., & Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(24), 11558–11562. <https://doi.org/10.1073/pnas.90.24.11558>.
- Baum, B. R. (1992). Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon*, *41*(1), 3–10. <https://doi.org/10.2307/1222480>.
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Research*, *41*(D1), D36–42. <https://doi.org/10.1093/nar/gks1195>.
- Berbee, M. L., & Taylor, J. W. (1992). Detecting morphological convergence in true fungi, using 18S ribosomal RNA gene sequence data. *Biosystems*, *28*(1–3), 117–125.
- Berbee, M. L., & Taylor, J. W. (2010). Dating the molecular clock in fungi—How close are we? *Fungal Biology Reviews*, *24*(1–2), 1–16. <https://doi.org/10.1016/j.fbr.2010.03.001>.
- Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, *21*(2), 163–193. <https://doi.org/10.1111/j.1096-0031.2005.00059.x>.
- Bininda-Emonds, O. R. P. (2004). The evolution of supertrees. *Trends in Ecology and Evolution*, *19*(6), 315–322. <https://doi.org/10.1016/j.tree.2004.03.015>.

- Bryant, D., & Steel, M. (2009). Computing the distribution of a tree metric. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 6(3), 420–426. <https://doi.org/10.1109/TCBB.2009.32>.
- Butler, G., Rasmussen, M. D., Lin, M. F., Santos, M. C. M. A. S., Sakthikumar, S., Munro, C. A., et al. (2009). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature*, 459(7247), 657–662. <https://doi.org/10.1038/nature08064>.
- Byrne, K. P., & Wolfe, K. H. (2005). The yeast gene order browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, 15(10), 1456–1461. <https://doi.org/10.1101/gr.3672305>.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 421. <https://doi.org/10.1186/1471-2105-10-421>.
- Campbell, A., Mrázek, J., & Karlin, S. (1999). Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), 9184–9189. <https://doi.org/10.1073/pnas.96.16.9184>.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4), 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>.
- Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biological Reviews of the Cambridge Philosophical Society*, 73(3), 203–266.
- Chang, Y., Wang, S., Sekimoto, S., Aerts, A. L., Choi, C., Clum, A., et al. (2015). Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biology and Evolution*, 7(6), 1590–1601. <https://doi.org/10.1093/gbe/evv090>.
- Creevey, C. J., Fitzpatrick, D. A., Philip, G. K., Kinsella, R. J., O'Connell, M. J., Pentony, M. M., et al. (2004). Does a tree-like phylogeny only exist at the tips in the prokaryotes? *Proceedings Biological Sciences/The Royal Society*, 271(1557), 2551–2558. <https://doi.org/10.1098/rspb.2004.2864>.
- Creevey, C. J., & McInerney, J. O. (2005). Clann: Investigating phylogenetic information through supertree analyses. *Bioinformatics*, 21(3), 390–392. <https://doi.org/10.1093/bioinformatics/bti020>.
- Creevey, C. J., & McInerney, J. O. (2009). Trees from trees: Construction of phylogenetic supertrees using CLANN. *Methods in Molecular Biology*, 537, 139–161. https://doi.org/10.1007/978-1-59745-251-9_7.
- Cuomo, C. A., & Birren, B. W. (2010). The fungal genome initiative and lessons learned from genome sequencing. *Methods in Enzymology*, 470(C), 833–855. [https://doi.org/10.1016/S0076-6879\(10\)70034-3](https://doi.org/10.1016/S0076-6879(10)70034-3).
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*, 27(8), 1164–1165. <https://doi.org/10.1093/bioinformatics/btr088>.
- De Barros Lopes, M., Bellon, J. R., Shirley, N. J., & Ganter, P. F. (2002). Evidence for multiple interspecific hybridization in *Saccharomyces sensu stricto* species. *FEMS Yeast Research*, 1(4), 323–331. [https://doi.org/10.1016/S1567-1356\(01\)00051-4](https://doi.org/10.1016/S1567-1356(01)00051-4).
- Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5), 361–375. <https://doi.org/10.1038/nrg1603>.
- de Queiroz, A., & Gatesy, J. (2007). The supermatrix approach to systematics. *Trends in Ecology and Evolution*, 22(1), 34–41. <https://doi.org/10.1016/j.tree.2006.10.002>.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>.

- Engel, S. R., Dietrich, F. S., Fisk, D. G., Binkley, G., Balakrishnan, R., Costanzo, M. C., et al. (2014). The reference genome sequence of *Saccharomyces cerevisiae*: Then and now. *G3 (Bethesda)*, 4(3), 389–398. <https://doi.org/10.1534/g3.113.008995>.
- Faith, D. P., & Cranston, P. S. (1991). Could a cladogram this short have arisen by chance alone? On permutation tests for cladistic structure. *Cladistics*, 7(1), 1–28. <https://doi.org/10.1111/j.1096-0031.1991.tb00020.x>.
- Federhen, S. (2012). The NCBI taxonomy database. *Nucleic Acids Research*, 40(D1), D136–D143. <https://doi.org/10.1093/nar/gkr1178>.
- Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, 27(4), 401. <https://doi.org/10.2307/2412923>.
- Felsenstein, J. (1989). PHYLIP—Phylogeny inference package—v3.2. *Cladistics*, 5(2), 164–166. <https://doi.org/10.1111/j.1096-0031.1989.tb00562.x>.
- Fisk, D. G., Ball, C. A., Dolinski, K., Engel, S. R., Hong, E. L., Issel-Tarver, L., et al. (2006). *Saccharomyces cerevisiae* S288C genome annotation: A working hypothesis. *Yeast*, 23(12), 857–865. <https://doi.org/10.1002/yea.1400>.
- Fitzpatrick, D. A., Logue, M. E., & Butler, G. (2008). Evidence of recent interkingdom horizontal gene transfer between bacteria and *Candida parapsilosis*. *BMC Evolutionary Biology*, 8(1), 181. <https://doi.org/10.1186/1471-2148-8-181>.
- Fitzpatrick, D. A., Logue, M. E., Stajich, J. E., & Butler, G. (2006). A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evolutionary Biology*, 6(1), 99. <https://doi.org/10.1186/1471-2148-6-99>.
- Fitzpatrick, D. A., O'Gaora, P., Byrne, K. P., & Butler, G. (2010). Analysis of gene evolution and metabolic pathways using the *Candida* Gene Order Browser. *BMC Genomics*, 11(1), 290. <https://doi.org/10.1186/1471-2164-11-290>.
- Foster, P. G. (2004). Modeling compositional heterogeneity. *Systematic Biology*, 53(3), 485–495. <https://doi.org/10.1080/10635150490445779>.
- Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*, 422(6934), 859–868. <https://doi.org/10.1038/nature01554>.
- Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L.-J., Wortman, J. R., Batzoglou, S., et al. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, 438(7071), 1105–1115. <https://doi.org/10.1038/nature04341>.
- Galagan, J. E., Henn, M. R., Ma, L. J., Cuomo, C. A., & Birren, B. (2005). Genomics of the fungal kingdom: Insights into eukaryotic biology. *Genome Research*, 15(12), 1620–1631. <https://doi.org/10.1101/gr.3767105>.
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. *Science*, 274(5287), 546–567. <https://doi.org/10.1126/science.274.5287.546>.
- Goffeau, A., & Vassarotti, A. (1991). The European project for sequencing the yeast genome. *Research in Microbiology*, 142(7–8), 901–903. [https://doi.org/10.1016/0923-2508\(91\)90071-H](https://doi.org/10.1016/0923-2508(91)90071-H).
- Grigoriev, I. V., Cullen, D., Goodwin, S. B., Hibbett, D., Jeffries, T. W., Kubicek, C. P., et al. (2011). Fueling the future with fungal genomics. *Mycology*, 2(3), 192–209. <https://doi.org/10.1080/21501203.2011.584577>.
- Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otilar, R., et al. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, 42(D1), D699–704. <https://doi.org/10.1093/nar/gkt1183>.
- Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., et al. (2011). The genome portal of the department of energy joint genome institute. *Nucleic Acids Research*, 40(D1), 1–7. <https://doi.org/10.1093/nar/gkr947>.
- Guarro, J., Gené, J., & Stchigel, A. M. (1999). Developments in fungal taxonomy. *Clinical Microbiology Reviews*, 12(3), 454–500. [https://doi.org/0893-8512/99/\\$04.00?0](https://doi.org/0893-8512/99/$04.00?0).

- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>.
- Hall, C., & Dietrich, F. S. (2007). The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics*, *177*(4), 2293–2307. <https://doi.org/10.1534/genetics.107.074963>.
- Hawksworth, D. L. (2001). The magnitude of fungal diversity: The 1.5 million species estimate revisited. *Mycological Research*, *105*(12), 1422–1432. <https://doi.org/10.1017/S0953756201004725>.
- Heath, I. B. (1980). Variant mitoses in lower eukaryotes: Indicators of the evolution of mitosis? *International Review of Cytology*, *64*(C), 1–80. [https://doi.org/10.1016/S0074-7696\(08\)60235-1](https://doi.org/10.1016/S0074-7696(08)60235-1).
- Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., et al. (2007). A higher-level phylogenetic classification of the fungi. *Mycological Research*, *111*(5), 509–547. <https://doi.org/10.1016/j.mycres.2007.03.004>.
- Hoffmann, K., Voigt, K., & Kirk, P. M. (2011). *Mortierellomycotina* subphyl. nov., based on multi-gene genealogies. *Mycotaxon*, *115*(1), 353–363. <https://doi.org/10.5248/115.353>.
- Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., et al. (1965). Structure of a ribonucleic acid. *Science (New York, N.Y.)*, *147*(3664), 1462–1465. <https://doi.org/10.1126/science.147.3664.1462>.
- Holton, T. A., & Pisani, D. (2010). Deep genomic-scale analyses of the metazoa reject coelomata: Evidence from single- and multigene families analyzed under a supertree and supermatrix paradigm. *Genome Biology and Evolution*, *2*(1), 310–324. <https://doi.org/10.1093/gbe/evq016>.
- Huelsenbeck, J. P., & Hillis, D. M. (1993). Success of phylogenetic methods in the four taxon case. *Systematic Biology*, *42*(3), 247–264. <https://doi.org/10.1093/sysbio/42.3.247>.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, *294*(5550), 2310–2314. <https://doi.org/10.1126/science.1065889>.
- Jackson, A. P., Gamble, J. A., Yeomans, T., Moran, G. P., Saunders, D., Harris, D., et al. (2009). Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Research*, *19*(12), 2231–2244. <https://doi.org/10.1101/gr.097501.109>.
- James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., et al. (2006). Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature*, *443*(7113), 818–822. <https://doi.org/10.1038/nature05110>.
- Jones, M. D. M., Forn, I., Gadelha, C., Egan, M. J., Bass, D., Massana, R., et al. (2011). Discovery of novel intermediate forms redefines the fungal tree of life. *Nature*, *474*(7350), 200–203. <https://doi.org/10.1038/nature09984>.
- Jones, M. D. M., Richards, T. A., Hawksworth, D. L., & Bass, D. (2011). Validation and justification of the phylum name Cryptomycota phyl. nov. *IMA Fungus*, *2*(2), 173–175. <https://doi.org/10.5598/ima fungus.2011.02.02.08>.
- Keller, N. P., Turner, G., & Bennett, J. W. (2005). Fungal secondary metabolism—From biochemistry to genomics. *Nature Reviews Microbiology*, *3*(12), 937–947. <https://doi.org/10.1038/nrmicro1286>.
- Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, *428*(VN-(6983)), 617–624. <https://doi.org/10.1038/nature02424>.
- Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., et al. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology*, *47*(9), 736–741. <https://doi.org/10.1016/j.fgb.2010.06.003>.

- Kirk, P. M., Cannon, P. F., Minter, D. W., & Stalpers, J. A. (2008). *Ainsworth & Bisby's dictionary of the fungi* (10th ed.). Wallingford, UK: CABI.
- Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among epicrates (Boidae, serpentes). *Systematic Biology*, 38(1), 7–25. <https://doi.org/10.1093/sysbio/38.1.7>.
- Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biology*, 5(2), R7. <https://doi.org/10.1186/gb-2004-5-2-r7>.
- Kück, P., & Meusemann, K. (2010). FASconCAT: Convenient handling of data matrices. *Molecular Phylogenetics and Evolution*, 56(3), 1115–1118. <https://doi.org/10.1016/j.ympev.2010.04.024>.
- Kuramae, E. E., Robert, V., Snel, B., Weiß, M., & Boekhout, T. (2006). Phylogenomics reveal a robust fungal tree of life. *FEMS Yeast Research*, 6(8), 1213–1220. <https://doi.org/10.1111/j.1567-1364.2006.00119.x>.
- Lapointe, F.-J., & Cucumel, G. (1997). The average consensus procedure: Combination of weighted trees containing identical or overlapping sets of taxa. *Systematic Biology*, 46(2), 306–312. <https://doi.org/10.1093/sysbio/46.2.306>.
- Lartillot, N., Brinkmann, H., & Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evolutionary Biology*, 7(Suppl. 1), S4. <https://doi.org/10.1186/1471-2148-7-S1-S4>.
- Lartillot, N., & Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, 21(6), 1095–1109. <https://doi.org/10.1093/molbev/msh112>.
- Lartillot, N., Rodrigue, N., Stubbs, D., & Richer, J. (2013). PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology*, 62(4), 611–615. <https://doi.org/10.1093/sysbio/syt022>.
- Léjohn, H. B. (1974). *Biochemical parameters of fungal phylogenetics*. In T. Dobzhansky, M. K. -Hecht, & W. C. Steere (Eds.), *Evolutionary biology* (pp. 79–125). Boston, MA: Springer. https://doi.org/10.1007/978-1-4615-6944-2_3.
- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1), W242–W245. <https://doi.org/10.1093/nar/gkw290>.
- Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13(9), 2178–2189. <https://doi.org/10.1101/gr.1224503>.
- Lin, J., & Gerstein, M. (2000). Whole-genome trees based on the occurrence of folds and orthologs: Implications for comparing genomes on different levels. *Genome Research*, 10(6), 808–818. <https://doi.org/10.1101/gr.10.6.808>.
- Marcet-Houben, M., & Gabaldón, T. (2009). The tree versus the forest: The fungal tree of life and the topological diversity within the yeast phylome. *PLoS One*, 4(2). e4357 <https://doi.org/10.1371/journal.pone.0004357>.
- Marcet-Houben, M., & Gabaldón, T. (2010). Acquisition of prokaryotic genes by fungal genomes. *Trends in Genetics: TIG*, 26(1), 5–8. <https://doi.org/10.1016/j.tig.2009.11.007>.
- Marcet-Houben, M., Marceddu, G., & Gabaldón, T. (2009). Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. *BMC Evolutionary Biology*, 9(1), 295. <https://doi.org/10.1186/1471-2148-9-295>.
- Medina, E. M., Jones, G. W., & Fitzpatrick, D. A. (2011). Reconstructing the fungal tree of life using phylogenomics and a preliminary investigation of the distribution of yeast

- prion-like proteins in the fungal kingdom. *Journal of Molecular Evolution*, 73(3–4), 116–133. <https://doi.org/10.1007/s00239-011-9461-4>.
- Nikoh, N., Hayase, N., Iwabe, N., Kuma, K., & Miyata, T. (1994). Phylogenetic relationship of the kingdoms Animalia, Plantae, and Fungi, inferred from 23 different protein species. *Molecular Biology and Evolution*, 11(5), 762–768. [https://doi.org/0737-4038/94f1IOS-0005\\$02](https://doi.org/0737-4038/94f1IOS-0005$02).
- Oliver, S. G., Van Der Aart, Q. J., Agostoni-Carbone, M. L., Aigle, M., Alberghina, L., Alexandraki, D., et al. (1992). The complete DNA sequence of yeast chromosome III. *Nature*, 357(6373), 38–46. <https://doi.org/10.1038/357038a0>.
- Page, R. D. M., & Holmes, E. C. (1998). *Molecular evolution: A phylogenetic approach*. Oxford, UK: Blackwell Science.
- Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23(9), 1061–1067. <https://doi.org/10.1093/bioinformatics/btm071>.
- Pisani, D., & Wilkinson, M. (2002). Matrix representation with parsimony, taxonomic congruence, and total evidence. *Systematic Biology*, 51(1), 151–155. <https://doi.org/10.1080/106351502753475925>.
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS One*, 5(3). e9490 <https://doi.org/10.1371/journal.pone.0009490>.
- Pride, D. T., Meinersmann, R. J., Wassenaar, T. M., & Blaser, M. J. (2003). Evolutionary implications of microbial genome tetranucleotide frequency biases. *Genome Research*, 13(2), 145–156. <https://doi.org/10.1101/gr.335003>.
- Qi, J., Luo, H., & Hao, B. (2004). CVTree: A phylogenetic tree reconstruction tool based on whole genomes. *Nucleic Acids Research*, 32(Suppl. 2), W45–7. <https://doi.org/10.1093/nar/gkh362>.
- Qi, J., Wang, B., & Hao, B. I. (2004). Whole proteome prokaryote phylogeny without sequence alignment: A K-string composition approach. *Journal of Molecular Evolution*, 58(1), 1–11. <https://doi.org/10.1007/s00239-003-2493-7>.
- Ragan, M. A. (1992). Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetics and Evolution*, 1(1), 53–58. [https://doi.org/10.1016/1055-7903\(92\)90035-F](https://doi.org/10.1016/1055-7903(92)90035-F).
- Redecker, D. (2000). Glomalean fungi from the ordovician. *Science*, 289(5486), 1920–1921. <https://doi.org/10.1126/science.289.5486.1920>.
- Richards, T. A., Soanes, D. M., Jones, M. D. M., Vasieva, O., Leonard, G., Paszkiewicz, K., et al. (2011). Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(37), 15258–15263. <https://doi.org/10.1073/pnas.1105100108>.
- Robbertse, B., Reeves, J. B., Schoch, C. L., & Spatafora, J. W. (2006). A phylogenomic analysis of the Ascomycota. *Fungal Genetics and Biology*, 43(10), 715–725. <https://doi.org/10.1016/j.fgb.2006.05.001>.
- Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, 425(6960), 798–804. <https://doi.org/10.1038/nature02053>.
- Schoch, C. L., Sung, G. H., López-Giráldez, F., Townsend, J. P., Miadlikowska, J., Hofstetter, V., et al. (2009). The ascomycota tree of life: A phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. *Systematic Biology*, 58(2), 224–239. <https://doi.org/10.1093/sysbio/syp020>.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.

- Slowinski, J. B., & Page, R. D. (1999). How should species phylogenies be inferred from sequence data? *Systematic Biology*, 48(4), 814–825. <https://doi.org/10.1080/106351599260030>.
- Snel, B., Bork, P., & Huynen, M. a. (1999). Genome phylogeny based on gene content. *Nature Genetics*, 21(1), 108–110. <https://doi.org/10.1038/5052>.
- Snel, B., Huynen, M. A., & Dutilh, B. E. (2005). Genome trees and the nature of genome evolution. *Annual Review of Microbiology*, 59(1), 191–209. <https://doi.org/10.1146/annurev.micro.59.030804.121233>.
- Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., et al. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia*, 108(5), 1028–1046. <https://doi.org/10.3852/16-042>.
- Spatafora, J., Sung, G., Johnson, D., Hesse, C., O'Rourke, B., Serdani, M., et al. (2006). A five-gene phylogeny of Pezizomycotina. *Mycologia*, 98(6), 1018–1028. <https://doi.org/10.3852/mycologia.98.6.1018>.
- Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*, 32(Suppl. 2) W309–12 <https://doi.org/10.1093/nar/gkh379>.
- Steel, M., & Rodrigo, A. (2008). Maximum likelihood supertrees. *Systematic Biology*, 57(2), 243–250. <https://doi.org/10.1080/10635150802033014>.
- Su, Z., & Townsend, J. P. (2015). Utility of characters evolving at diverse rates of evolution to resolve quartet trees with unequal branch lengths: Analytical predictions of long-branch effects. *BMC Evolutionary Biology*, 15(86), 86. <https://doi.org/10.1186/s12862-015-0364-7>.
- Swofford, L. D. (2002). *PAUP*: Phylogenetic analysis using parsimony (* and other methods). Version 4.0 beta*. Sunderland, MA: Sinauer.
- Szöllösi, G. J., Davin, A. A., Tannier, E., Daubin, V., & Boussau, B. (2015). Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 370(1678), 20140335. <https://doi.org/10.1098/rstb.2014.0335>.
- Taylor, F. J. R. (1978). Problems in the development of an explicit hypothetical phylogeny of the lower eukaryotes. *Biosystems*, 10(1–2), 67–89. [https://doi.org/10.1016/0303-2647\(78\)90031-X](https://doi.org/10.1016/0303-2647(78)90031-X).
- Tekaia, F., Lazcano, A., & Dujon, B. (1999). The genomic tree as revealed from whole proteome comparisons. *Genome Research*, 9(6), 550–557. <https://doi.org/10.1101/gr.9.6.550>.
- Wang, H., Xu, Z., Gao, L., & Hao, B. (2009). A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evolutionary Biology*, 9(1), 195. <https://doi.org/10.1186/1471-2148-9-195>.
- Wilkinson, M., Thorley, J. L., Pisani, D. E., Lapointe, F.-J., & McInerney, J. O. (2004). *Some desiderata for liberal supertrees*. In O. R. P. Bininda-Emonds (Ed.), Vol. 3. *Phylogenetic supertrees: Combining information to reveal the Tree of Life* (pp. 227–246). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-1-4020-2330-9_11.
- Wolfé, K. H., & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387(6634), 708–713. <https://doi.org/10.1038/42711>.
- Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature*, 415(6874), 871–880. <https://doi.org/10.1038/nature724>.
- Zuo, G., Li, Q., & Hao, B. (2014). On K-peptide length in composition vector phylogeny of prokaryotes. *Computational Biology and Chemistry*, 53(Part A), 166–173. <https://doi.org/10.1016/j.compbiolchem.2014.08.021>.
- Zuo, G., Xu, Z., Yu, H., & Hao, B. (2010). Jackknife and bootstrap tests of the composition vector trees. *Genomics, Proteomics and Bioinformatics*, 8(4), 262–267. [https://doi.org/10.1016/S1672-0229\(10\)60028-9](https://doi.org/10.1016/S1672-0229(10)60028-9).