bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota

Xing-Xing Shen[1,*], Jacob L. Steenwyk[2], Abigail L. LaBella[2], Dana A. Opulente[3], Xiaofan Zhou[4], Jacek Kominek[3], Yuanning Li[2], Marizeth Groenewald[5], Chris Todd Hittinger[3], and Antonis Rokas[2,*]

[1]Institute of Insect Sciences, Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China

[2]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA

[3]Laboratory of Genetics, Wisconsin Energy Institute, Center for Genomic Science Innovation, DOE Great Lakes Bioenergy Research Center, J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53706, USA

[4]Guangdong Laboratory for Lingnan Modern Agriculture, Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Centre, South China Agricultural University, Guangzhou 510642, China

[5]Westerdijk Fungal Biodiversity Institute, 3584 CT, The Netherlands

*Correspondence: xingxingshen@zju.edu.cn (XXS) and antonis.rokas@vanderbilt.edu (AR)

Keywords: Fungi, evolutionary rate, GC content, phylogenomics, molecular dating, ancestral state reconstruction, genome evolution, budding yeasts, genome size

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Ascomycota, the largest and best-studied phylum of fungi, contains three subphyla: Saccharomycotina (budding yeasts), Pezizomycotina (filamentous fungi), and Taphrinomycotina (fission yeasts); organisms from all three subphyla have been invaluable as models in diverse fields (e.g., biotechnology, cell biology, genetics, and medicine). Despite its importance, we still lack a comprehensive genome-scale phylogeny or understanding of the similarities and differences in the mode of genome evolution within this phylum. To address these gaps, we examined 1,107 genomes from Saccharomycotina (332), Pezizomycotina (761), and Taphrinomycotina (14) species to infer the Ascomycota phylogeny, estimate its timetree, and examine the evolution of key genomic properties. We inferred a robust genome-wide phylogeny that resolves several contentious relationships and estimated that the Ascomycota last common ancestor likely originated in the Ediacaran (~563 ± 68 million years ago). Comparisons of genomic properties revealed that Saccharomycotina and Pezizomycotina, the two taxon-rich subphyla, differed greatly in their genome properties. Saccharomycotina typically have smaller genomes, lower GC contents, lower numbers of genes, and higher rates of molecular sequence evolution compared to Pezizomycotina. Ancestral state reconstruction showed that the genome properties of the Saccharomycotina and Pezizomycotina last common ancestors were very similar, enabling inference of the direction of evolutionary change. For example, we found that a lineage-specific acceleration led to a 1.6-fold higher evolutionary rate in Saccharomycotina, whereas the 10% difference in GC content between Saccharomycotina and Pezizomycotina genomes stems from a trend toward AT bases within budding yeasts and toward GC bases within filamentous fungi. These

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

44 results provide a robust evolutionary framework for understanding the diversification of the

45 largest fungal phylum.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Main

The fungal phylum Ascomycota is one of most diverse phyla of eukaryotes with ~65,000 known species that represent approximately three quarters of all known species of fungi[1]. The Ascomycota is divided in three subphyla. The Saccharomycotina subphylum is a lineage of more than 1,000 known species and 12 major clades[2]; commonly referred to as budding yeasts. Species in this lineage include the model organism *Saccharomyces cerevisiae*[3] and several notable pathogens, such as the human commensal *Candida albicans*[4] and the multidrug-resistant emerging pathogen *Candida auris*[5]. The Pezizomycotina subphylum contains more than 63,000 described species in 13 classes[6]; commonly referred to as filamentous fungi. This subphylum contains several major plant and animal pathogens belonging to diverse genera, such as *Fusarium*, *Aspergillus*, *Zymoseptoria*, and *Magnaporthe*[7–10]. Finally, the Taphrinomycotina subphylum contains ~140 described species in 5 classes[6]; commonly referred to as fission yeasts. This subphylum contains the model organism *Schizosaccharomyces pombe* and the human pathogen *Pneumocystis jirovecii*[11,12].

To better understand the evolution of species diversity and ecological lifestyles in Ascomycota fungi, a robust framework of phylogenetic relationships and divergence time estimates is essential. In the last two decades, several studies have aimed to infer the Ascomycota phylogeny, either using a handful of gene markers from hundreds of taxa[13–17] or using hundreds of gene markers from tens of taxa[18–21]. To date, the most comprehensive "few-markers-from-many-taxa" phylogeny used a 6-gene, 420-taxon (8 Taphrinomycotina, 16 Saccharomycotina, and 396 Pezizomycotina) data matrix[13], whereas the most

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

68 comprehensive genome-scale phylogeny used an 238-gene, 496-taxon (12 Taphrinomycotina,
69 76 Saccharomycotina, and 408 Pezizomycotina) data matrix[22] but was inferred using
70 FastTree, a program that is faster but typically yields phylogenies that have much lower
71 likelihood scores than those obtained by IQ-TREE and RAxML/RAxML-NG[23]. Key
72 relationships supported by these studies include the monophyly of each subphylum and class
73 and the sister group relationship of subphyla Saccharomycotina and Pezizomycotina. In
74 contrast, relationships among classes are contentious between studies, particularly with
75 respect to relationships between the 13 classes in Pezizomycotina[6]. For example, there is
76 disagreement whether the sister group to the rest of classes in the Pezizomycotina is class
77 Pezizomycetes[14], class Orbiliomycetes[17], or a clade comprised of both[19].
78
79 Previous molecular clock-based estimates of divergence times for Ascomycota have all been
80 based on few-markers-from-many-taxa data matrices[14,15,24–26], resulting in age estimates for
81 key events in Ascomycota evolution that have wide intervals. For example, analysis of a 6-
82 gene, 121-taxon (1 Saccharomycotina, 118 Pezizomycotina, and 2 Taphrinomycotina) data
83 matrix inferred that the origin of the phylum Ascomycota took place 531 million years ago
84 (mya) (95% credibility interval (CI): 671-410 mya) (see their Scenario 4 in Table 3)[15], while
85 analysis of a 4-gene, 145-taxon (12 Saccharomycotina, 129 Pezizomycotina, and 4
86 Taphrinomycotina) data matrix inferred that the phylum originated 588 mya (95% CI: 773-
87 487 mya)[14]. More importantly, the sparser taxon sampling of previous studies has prevented
88 estimation of divergence times of several key divergence events of higher taxonomic ranks[24–

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

89 [26] and stymied our understanding of their evolutionary pace. While these studies have
90 significantly advanced our understanding of Ascomycota evolution, a comprehensive,
91 genome-scale phylogeny and timetree stemming from the sampling of hundreds of genes
92 from thousands of taxa from the phylum are still lacking.
93
94 A robust phylogenomic framework would also facilitate comparisons of genome evolution
95 across the subphyla of Ascomycota. For example, the three subphyla differ in their genome
96 sizes, with the genomes of Pezizomycotina species being notably larger (~42 Mb) than those
97 of Saccharomycotina (~13 Mb) and Taphrinomycotina (~14 Mb)[27]. While several recent
98 studies have analyzed major lineages within the two taxon-rich subphyla,
99 Saccharomycotina[2,20,28] and Pezizomycotina[29–31], comparisons of genome evolution across
100 the two subphyla are lacking. For example, a recent analysis of the tempo and mode of
101 genome evolution in 332 Saccharomycotina found evidence of high evolutionary rates and
102 reductive evolution across this subphylum[2], but whether budding yeasts are faster evolving
103 than filamentous fungi remains unknown. However, a recent analysis of 71 Ascomycota
104 genomes showed that Pezizomycotina have much higher levels of gene order divergence than
105 Saccharomycotina[21]. Similarly, genome-wide examinations of horizontal gene transfer events
106 in dozens to more than a hundred Ascomycota genomes have revealed that Pezizomycotina
107 acquired significantly higher numbers of genes from prokaryotic donors than
108 Saccharomycotina [32,33]. Although these studies have contributed to our understanding of
109 certain evolutionary processes in the phylum, we still know relatively little about the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

evolution of Ascomycota genomes and their properties.

There are currently more than one thousand genomes from Ascomycota species that are publicly available, which span the diversity of Saccharomycotina (332 genomes representing all 12 major clades), Pezizomycotina (761 genomes representing 9 / 13 classes), and Taphrinomycotina (14 genomes representing 4 / 5 classes) (1,107 genomes as of December 14, 2018). These 1,107 genomes represent a much larger and representative source of genomic data across the entire Ascomycota phylum than previously available, providing a unique opportunity to infer a genome-scale phylogeny and timetree for the entire subphylum and compare the mode of genome evolution across its subphyla.

Results and Discussion

A genome-scale phylogeny of the fungal phylum Ascomycota

To infer a genome-scale phylogeny of Ascomycota fungi, we employed 1,107 publicly available genomes from species belonging to Ascomycota (Saccharomycotina: 332; Pezizomycotina: 761; Taphrinomycotina: 14) and six outgroups from the sister fungal phylum Basidiomycota. All genomes were retrieved from the NCBI GenBank database, ensuring that only one genome per species was included (Supplementary Tables 1 and 2). Analysis of genome assembly completeness reveals that 1,021/1,113 (~92%) genomes have more than 90% of the 1,315 full-length BUSCO genes[34] (Supplementary Fig. 1).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1,315 BUSCO genes from 1,107 Ascomycota fungi and six outgroups were used to construct a phylogenomic data matrix (see Methods). After constructing the multiple amino acid sequence alignment and trimming ambiguous regions for each of these 1,315 BUSCO genes, we kept only the 815 BUSCO genes that had taxon occupancy of ≥ 50% for each subphylum (i.e., ≥ 7 Taphrinomycotina, ≥ 166 Saccharomycotina, and ≥ 381 Pezizomycotina) and whose amino acid sequence alignments were ≥ 300 sites in length. In the final set of 815 BUSCO genes, alignment lengths range from 300 to 4,585 amino acid sites (average = 690) and numbers of taxa range from 851 to 1,098 (average = 1,051) (Supplementary Table 3). The final data matrix contains 1,107 taxa, 815 genes, and 562,376 amino acid sites.

Inference using concatenation- and coalescent-based approaches yielded a robust, comprehensive phylogeny of the Ascomycota phylum (Fig. 1). The vast majority of internodes in both the concatenation-based (1,103 / 1,110; 99%) and the coalescent-based phylogeny (1,076 / 1,110; 97%) received strong (≥ 95%) support and were congruent between the phylogenies inferred using the two approaches; only 46 / 1,110 (4%) internodes were incongruent between the two phylogenies (Supplementary Figs. 2 and 3).

Our higher-level phylogeny of Ascomycota is generally more congruent with previous genome-scale phylogenies[2,18,19,35] than with few-genes-from-many-taxa phylogenies[13–16], particularly with respect to relationships among the nine classes in the subphylum Pezizomycotina. For example, genome-scale studies, including ours, consistently favor a

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

clade consisting of Pezizomycetes and Orbiliomycetes as the sister group to the rest of the Pezizomycotina [18,19], while studies based on a few genes recovered either Orbiliomycetes[15–17] or Pezizomycetes[14] as the sister class to the rest of the Pezizomycotina (Fig. 2a). Our phylogeny also strongly supported the placement of the class Schizosaccharomycetes, which includes the model organism *Schizosaccharomyces pombe*, as the sister group to the class Pneumocystidomycetes, which contains the human pathogen *Pneumocystis jirovecii* (Fig. 2b). Interestingly, a recent genome-scale study of 84 fungal genomes showed that our result is consistent with the phylogeny inferred using an alignment-free composition vector approach but not with the phylogeny inferred using maximum likelihood, which instead recovered Schizosaccharomycetes as the sister group to Taphrinomycetes[18]. Finally, both concatenation- and coalescent-based approaches supported the placement of the subphylum Saccharomycotina as the sister group to the subphylum Pezizomycotina (Figs. 1 and 2c). This result is consistent with most previous studies that analyze multiple sequence alignment data[16,18,20,36–38], but not with a recent study that analyzed genomic data with an alignment-free method and placed the subphylum Saccharomycotina as the sister group to the subphylum Taphrinomycotina[39].

To evaluate whether our genome-scale data matrix robustly resolved the three historically contentious branches discussed in the previous paragraph, we quantified the distribution of phylogenetic signal for alternative topologies of these three phylogenetic hypotheses at the level of genes and sites using a maximum likelihood framework presented by Shen et al.[40].

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

First, we found that phylogenetic support for each of the three branches stemmed from many genes, i.e., it was not dominated by a small number of genes with highly disproportionate influence (Supplementary Table 4). Second, we found that the topology recovered by both concatenation- and coalescent-based approaches in our study had significantly the highest frequencies of supporting genes and supporting sites (*G*–test), ranging from 0.45 to 0.65, in all three branches examined (Fig. 2a-c, Supplementary Tables 4 and 5). Importantly, none of two alternative conflicting phylogenetic hypotheses for each of the three branches received frequencies of supporting genes and supporting sites that were equal or greater than 1/3 (0.33), the value expected if the relationships among the taxa were represented by a polytomy. The very small fraction of branches where concatenation- and coalescent-based inference conflicted (<5%) and the robust support of individual genes and sites for specific historically contentious branches (Fig. 2a-c) suggest that the coupling of genome-scale amounts of data and comprehensive taxon sampling will provide robust resolution to major lineages of the tree of life[2,41].

A genome-scale evolutionary timetree of the fungal phylum Ascomycota

We next used the robust phylogeny, a relaxed molecular clock approach, and six widely accepted time calibration nodes (see Methods), to infer the timescale of evolution of Ascomycota. We inferred the origin of the phylum to have taken place 563 million years ago (mya) (95% credibility interval (CI): 631–495 mya); the origin of the subphylum Saccharomycotina 438.4 mya (CI: 590–304 mya); the origin of the subphylum

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

194 Pezizomycotina 407.7 mya (CI: 631–405 mya); and the origin of Taphrinomycotina crown
195 group 530.5 mya (CI: 620–417 mya). Notably, the taxonomic placement of all budding yeasts
196 into a single class, Saccharomycetes, whose origin coincides with the origin of the
197 subphylum Saccharomycotina, means that the last common ancestor of this sole class of
198 budding yeasts is much more ancient than those of any of the 9 classes (based on current
199 taxon sampling) in the subphylum Pezizomycotina (Supplementary Fig. 4 and Supplementary
200 Table 6). For example, the most ancient class in Pezizomycotina is Pezizomycetes, whose
201 origin is dated 247.7 mya (CI: 475-193 mya) (Supplementary Fig. 4 and Supplementary Table
202 6). The other outlier, albeit with much larger confidence intervals, is class Neolectomycetes
203 in Taphrinomycotina, which we estimate to have originated 480.4 mya (CI: 607-191 mya)
204 (Supplementary Fig. 4 and Supplementary Table 6).
205
206 Comparison of our inferred dates of divergence to those of a recent study using a 4-gene,
207 145-taxon data matrix[14] shows that our estimates are younger (563 vs 588 mya for
208 Ascomycota and 408 vs 458 mya for Pezizomycotina; sparser taxon sampling in the previous
209 study prevents comparison of dates for Saccharomycotina and Taphrinomycotina). This result
210 is consistent with findings of previous studies[42,43], where inclusion of large numbers of genes
211 was found to also result in younger estimates of divergence times, perhaps because of the
212 influence of larger amounts of data in decreasing the stochastic error involved in date
213 estimation. In summary, generation of a genome-scale timetree for more than 1,000
214 ascomycete species spanning the diversity of the phylum provides a robust temporal

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

framework for understanding and exploring the origin and diversity of Ascomycota lifestyles[44].

Contrasting modes of genome evolution in fungal phylum Ascomycota

To begin understanding the similarities and differences in the modes of genome evolution between subphyla, we focused on examining the evolution of seven different genomic properties between Saccharomycotina (332 taxa) and Pezizomycotina (761 taxa), the two most taxon-rich subphyla in Ascomycota (Fig. 3). Specifically, we found that Saccharomycotina exhibited a 1.6-fold higher evolutionary rate (on average, 1.80 substitutions per site in Saccharomycotina vs. 1.12 substitutions per site in Pezizomycotina), 1.24-fold lower GC content (40% vs. 50%), 3-fold smaller genome size (13 Mb vs. 39 Mb), 1.9-fold lower number of protein-coding genes (5,734 vs. 10,847), 1.3-fold lower number of DNA repair genes (41 vs. 54), 1.2-fold higher number of tRNA genes (179 vs. 146), and 1.3-fold smaller estimates of non-synonymous to synonymous substitution rate ratio (d_N/d_S) (0.053 vs. 0.063), compared to Pezizomycotina (Fig. 3a, Table 1, and Supplementary Table 7).

Analysis of standard Pearson's correlations among the seven genomic properties revealed that two pairs exhibited statistically significant contrasting patterns between Saccharomycotina and Pezizomycotina. Specifically, evolutionary rate shows negative correlation with GC content in Saccharomycotina but positive correlation in Pezizomycotina and GC content

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

236 shows negative correlation with number of DNA repair genes in Saccharomycotina but
237 positive correlation in Pezizomycotina (Fig. 3b). These correlations are largely consistent
238 before (i.e., standard Pearson's correlations) and after (i.e., phylogenetically independent
239 contrasts) accounting for correlations due to phylogeny (Supplementary Table 8).
240
241 For each of the seven properties, we used our genome-scale phylogeny (Fig. 1) to infer the
242 ancestral character states and reconstruct their evolution in the Saccharomycotina ancestor
243 and the Pezizomycotina ancestor. Comparison of ancestral states along branches on the
244 Saccharomycotina part of the phylogeny to those on the Pezizomycotina part of the
245 phylogeny shown that all genomic properties, except the number of tRNA genes, exhibited
246 different modes of evolution (Fig. 4 and Table 1). For example, most Saccharomycotina
247 branches exhibit evolutionary rates of at least 1.0 amino acid substitutions / site, whereas
248 those of Pezizomycotina exhibit evolutionary rates between 0.7 and 1.4 substitutions / site
249 (Fig. 4a). However, the inferred values for these properties in the Saccharomycotina last
250 common ancestor and in the Pezizomycotina last common ancestor nodes are quite similar.
251 For example, the inferred state values for the Saccharomycotina last common ancestor and
252 the Pezizomycotina last common ancestor are 1.1 and 0.9 substitutions / site for evolutionary
253 rate and 43% and 47% for GC content (Table 1), respectively. Interestingly, the same trends
254 are also observed across lineages, such as Lipomycetaceae, which is the sister group to the
255 rest of the Saccharomycotina, and the clade consisting of Pezizomycetes and Orbiliomycetes,
256 which is the sister group to the rest of the Pezizomycotina (Fig. 4a and b).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Comparison of the trait values for the seven genome properties between extant Saccharomycotina and Pezizomycotina branches to those of the Saccharomycotina and Pezizomycotina last common ancestors showed that evolutionary rate, GC content, genome size, and number of protein-coding genes were the properties with the highest amounts of evolutionary change (Figs. 3 and 4, Table 1). Ancestral state reconstruction also enabled inference of the direction of evolutionary change for each of the evolutionary properties. For example, the Saccharomycotina and Pezizomycotina last common ancestors, as well as branches in Lipomycetaceae and branches across Pezizomycotina, exhibit similar evolutionary rates, whereas the rest of the nodes and branches in the Saccharomycotina part of the phylogeny exhibit much higher evolutionary rates. This pattern suggests that the higher levels of genomic diversity in Saccharomycotina stem from an acceleration of evolutionary rate that occurred within the subphylum, after the divergence of Lipomycetaceae from the rest of the Saccharomycotina (Fig. 4a and b).

Why do Saccharomycotina exhibit higher evolutionary rates compared to Pezizomycotina? Studies in other lineages, such as vertebrate[45] and invertebrate[46] animals, have previously shown that evolutionary rate is positively associated with generation time. Assuming that mutation rates are equal, species with shorter generation times will replicate their genomes more frequently, accruing more mutations per unit time. While the generation times of most fungi in our phylogeny are unknown, the generation times of model organisms in

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

278 Saccharomycotina are thought to be shorter than those in Pezizomycotina. For example, the
279 doubling time of the budding yeasts *S. cerevisiae* and *C. albicans* under optimal conditions is
280 90 min[47,48], while that of the filamentous fungi *Aspergillus nidulans* and *Neurospora crassa* is
281 between 2-3 hours[49,50]. An alternative but not mutually exclusive explanation may be that
282 Saccharomycotina have, on average, 13 fewer DNA repair genes (41) than Pezizomycotina
283 (54) (Fig. 3 and Table 1), since it is well established that absence or loss of DNA repair genes
284 increase mutation rates[51–53]. The lower numbers of DNA repair genes in budding yeasts, but
285 not their higher evolutionary rate, was also recently reported in a recent analysis of 328
286 ascomycete proteomes by Milo et al.[54]. Finally, other life-history traits (e.g., smaller cell size,
287 faster metabolism, and larger population size) that have been associated with variation in the
288 rate of molecular evolution[55] might also contribute to higher evolutionary rates of the
289 Saccharomycotina.
290
291 Variation in genomic GC content has historically been of broad interest in biology[56]. Average
292 GC content values of different genomic regions (e.g., intergenic regions, protein-coding
293 regions) in Saccharomycotina are consistently lower than those in Pezizomycotina
294 (Supplementary Fig. 5). Similarly, gene-wise average estimates of GC content showed that all
295 815 BUSCO genes in Saccharomycotina have lower GC content values than those in
296 Pezizomycotina (Supplementary Fig. 6). Moreover, we found that the frequencies of amino
297 acids encoded by GC-rich codons in Saccharomycotina are much lower than those of amino
298 acids encoded by GC-rich codons in Pezizomycotina (Supplementary Fig. 7). Ancestral state

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

299 reconstruction of genomic GC content along branches on the phylogeny shows that the
300 Saccharomycotina and Pezizomycotina last common ancestors, as well as branches in
301 Lipomycetaceae and branches in classes Pezizomycetes and Orbiliomycetes, exhibit
302 intermediate GC content around 45%. In contrast, GC content of most branches within the
303 rest of Saccharomycotina (i.e., all major clades of Saccharomycotina, including extant taxa,
304 except Lipomycetaceae) evolved toward 40%, while GC content within the rest of
305 Pezizomycotina (i.e., all classes, including extant taxa, except Pezizomycetes and
306 Orbiliomycetes) evolved toward 50%. This pattern suggests that the evolution of lower levels
307 of GC content in Saccharomycotina occurred after the divergence of Lipomycetaceae from
308 the rest of Saccharomycotina and that the evolution of higher levels of GC content in
309 Pezizomycotina occurred after the divergence of the clade consisting of Pezizomycetes and
310 Orbiliomycetes from the rest of Pezizomycotina (Fig. 4a and b).
311
312 Why are Pezizomycotina genomes more GC-rich compared to Saccharomycotina genomes?
313 There are two possible explanations. The first one is that mutational biases have skewed the
314 composition of Saccharomycotina genomes toward AT content[57]. For example, Steenwyk et
315 al. showed that *Hanseniaspora* budding yeasts with higher AT content lost a greater number
316 of DNA repair genes than those with lower AT content[53], suggesting that the loss of DNA
317 repair genes is associated with AT richness. Consistent with these results, we found that
318 Pezizomycotina genomes contain a higher number of DNA repair genes than
319 Saccharomycotina (Fig. 3 and Table 1). The second potential, not necessarily mutually

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

exclusive, explanation is that mutational biases have skewed Pezizomycotina genomes toward GC richness. It was recently shown that increasing GC-biased gene conversion (gBGC), a process associated with recombination that favors the transmission of GC alleles over AT alleles[58], can result in a systematic underestimate of d_N/d_S in birds[59]. If this is true for Ascomycota, due to the higher GC content of Pezizomycotina genomes, we would expect that their d_N/d_S would be underestimated due to the higher levels of gBGC compared to Saccharomycotina. Consistent with this expectation, by calculating differences in d_N/d_S before and after accounting for gBGC across 815 codon-based BUSCO genes, we found that the underestimate of d_N/d_S in Pezizomycotina is 2-fold higher than that in Saccharomycotina (Pezizomycotina: average of differences in d_N/d_S = 0.004; Saccharomycotina: average of differences in d_N/d_S = 0.002) (Supplementary Fig. 8).

Concluding Remarks

In this study, we took advantage of the recent availability of the genome sequences of 1,107 Ascomycota species from Saccharomycotina (332), Pezizomycotina (761), and Taphrinomycotina (14) to infer a genome-scale phylogeny and timetree for the entire phylum and compare the mode of genome evolution across its subphyla. Leveraging genome-scale amounts of data from the most comprehensive taxon set to date enabled us to test the robustness of our inference for several contentious branches, potentially resolving controversies surrounding key higher-level relationships within the Ascomycota phylum. For example, our study robustly supported Saccharomycotina as the sister group to

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Pezizomycotina and a clade comprised of classes Pezizomycetes and Orbiliomycetes as the sister group to the rest of the Pezizomycotina. Our first genome-scale timetree suggests the last common ancestor of Ascomycota likely originated in the Ediacaran period. Examination of mode of genome evolution revealed that Saccharomycotina, which contains the single currently described class Saccharomycetes, and Pezizomycotina, which contains 13 classes, exhibited greatly contrasting evolutionary processes for seven genomic properties, in particular for evolutionary rate, GC content, and genome size. Our results provide a robust evolutionary framework for understanding the diversification of the largest fungal phylum.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Methods

Data collection

To collect the greatest possible set of genome representatives of the phylum Ascomycota as of 14 December, 2018, we first retrieved the 332 publicly available Saccharomycotina yeast genomes (https://doi.org/10.6084/m9.figshare.5854692) from a recent comprehensive genomic study of the Saccharomycotina yeasts[2]. We then used "Pezizomycotina" and "Taphrinomycotina" as search terms in NCBI's Genome Browser (https://www.ncbi.nlm.nih.gov/genome/browse#!/eukaryotes/Ascomycota) to obtain the basic information of strain name, assembly accession number, assembly release date, assembly level (e.g., contig, scaffold, etc.), and GenBank FTP access number for draft genomes from the subphyla Pezizomycotina and Taphrinomycotina, respectively. For species with multiple isolates sequenced, we only included the genome of the isolate with the highest assembly level and the latest release date. We next downloaded genome assemblies from GenBank data via FTP access number (ftp://ftp.ncbi.nlm.nih.gov/genomes/). Collectively, we included 332 species representing all 12 major clades of the subphylum Saccharomycotina[2], 761 species representing 9 / 13 classes of the subphylum Pezizomycotina[1,6], and 14 species representing 4 / 5 classes of the subphylum Taphrinomycotina[1,6]. Finally, we used the genomes of six representatives of the phylum Basidiomycota as outgroups. Detailed information of taxonomy and source of the 1,113 genomes in our study is provided in Supplementary Tables 1 and 2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Assessment of genome assemblies and phylogenomic data matrix construction

To assess the quality of each of the 1,113 genome assemblies, we used the Benchmarking Universal Single-Copy Orthologs (BUSCO), version 3.0.2[34]. Each assembly's completeness was assessed based on the presence / absence of a set of 1,315 predefined orthologs (referred to as BUSCO genes) from 75 genomes in the OrthoDB Version 9 database[60] from the Ascomycota database, as described previously[28,61]. In brief, for each BUSCO gene, its consensus orthologous protein sequence among the 75 reference genomes was used as query in a tBLASTn search against each genome to identify up to three putative genomic regions, and the gene structure of each putative genomic region was predicted by AUGUSTUS v 3.2.2[62]. Next, the sequences of these predicted genes were aligned to the HMM-profile of the BUSCO gene. BUSCO genes in a given genome assembly were considered as single-copy, "full-length" if there was only one complete predicted gene present in the genome, duplicated, "full-length" if there were two or more complete predicted genes present in the genome, "fragmented" if the predicted gene was shorter than 95% of the aligned sequence lengths from the 75 reference species, and "missing" if there was no predicted gene present in the genome.

To construct the phylogenomic data matrix, we started with the set of 1,315 single-copy, full-length BUSCO genes from 1,107 representatives of the phylum Ascomycota and six outgroups. For each BUSCO gene, we first translated nucleotide sequences into amino acid sequences, taking into account the different usage of the CUG codon in Saccharomycotina[2,63].

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Next, we aligned the amino acid sequences using MAFFT v7.299b[64] with the options "--thread 4 --auto --maxiterate 1000" and trimmed amino acid alignments using the trimAl v1.4.rev15[65] with the options "-gappyout -colnumbering". We mapped the nucleotide sequences on the trimmed amino acid alignment based on the column numbers in the original alignment and to generate the trimmed codon-based nucleotide alignment. Finally, we removed BUSCO gene alignments whose taxon occupancy (i.e., percentage of taxa whose sequences were present in the trimmed amino acid alignment) was < 50% for each subphylum (i.e., < 7 Taphrinomycotina, < 166 Saccharomycotina, and < 381 Pezizomycotina) or whose trimmed alignment length was < 300 amino acid sites. These filters resulted in the retention of 815 BUSCO gene alignments, each of which had ≥ 50% taxon occupancy for each subphylum and alignment length ≥ 300 amino acid sites.

Phylogenetic analysis

For each of 815 BUSCO genes, we first inferred its best-fitting amino acid substitution model using IQ-TREE multi-thread version 1.6.8 [66] with options "-m TEST -mrate G4" with the Bayesian information criterion (BIC). We then inferred best-scoring maximum likelihood (ML) gene tree under 10 independent tree searches using IQ-TREE. The detailed parameters for running each gene were kept in log files (see the Figshare repository). We inferred the concatenation-based ML tree using IQ-TREE on a single node with 32 logical cores under a single "LG +G4" model with the options "-seed 668688 -nt 32 -mem 220G -m LG+G4 -bb 1000", as 404 out of 815 genes favored "LG +G4"[67,68] as best-fitting model (see Supplementary Table 3). We also inferred the coalescent-based species phylogeny with

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ASTRAL-III version 4.10.2[69,70] using the set of 815 individual ML gene trees. The reliability of each internal branch was evaluated using 1,000 ultrafast bootstrap replicates[71] and local posterior probability[72], in the concatenation- and coalescence-based species trees, respectively. We visualized phylogenetic trees using the R package *ggtree* v1.10.5[73].

We used the non-Bayesian RelTime method, as implemented in the command line version of MEGA7[74] to estimate divergence times. The very large size of our data matrix, both in terms of genes as well as in terms of taxa, prohibited the use of computationally much more demanding methods, such as the Bayesian MCMCTree method[75,76]. The concatenation-based ML tree with branch lengths was used as the input tree. Six time calibration nodes, which were retrieved from the TimeTree database [77], were used for molecular dating analyses: the *Saccharomyces cerevisiae* – *Saccharomyces uvarum* split (14.3 mya – 17.94 mya), the *Saccharomyces cerevisiae* - *Kluyveromyces lactis* split (103 mya – 126 mya), the *Saccharomyces cerevisiae* - *Candida albicans* split (161 mya – 447 mya), the origin of the subphylum Saccharomycotina (304 mya – 590 mya), the *Saccharomyces cerevisiae* – *Saitoella complicata* split (444 mya – 631 mya), and the origin of the subphylum Pezizomycotina (at least 400 mya) based on the *Paleopyrenomycites devonicus* fossil[78].

Examination of seven genome properties

As the subphylum Taphrinomycotina (No. species = 14) has a much smaller number of species than the subphylum Saccharomycotina (No. species = 332) and the subphylum Pezizomycotina (No. species = 761) in our dataset, we focused our analyses on the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

436 comparisons of seven genome properties (evolutionary rate, GC content, genome size,
437 number of genes, number of DNA repair genes, number of tRNA genes, and d_N/d_S) between
438 Saccharomycotina and Pezizomycotina. Specifically, for a given taxon, 1) evolutionary rate is
439 a sum of path distances from the most common ancestor of the subphyla Saccharomycotina
440 and Pezizomycotina to its tip on the concatenation-based ML tree (Fig. 1); 2) GC content is
441 the percentage of guanine-cytosine nucleotides in genome; 3) genome size is the total number
442 of base pairs in genome in megabases (Mb); 4) number of genes is the number of protein-
443 coding genes in genome. The gene structure was predicted with AUGUSTUS v3.3.1 [79] on
444 *Aspergillus fumigatus* and *Saccharomyces cerevisiae* S288C trained models for
445 Pezizomycotina and Saccharomycotina, respectively; 5) number of DNA repair genes was
446 estimated by counting the number of unique protein-coding genes with GO terms related to
447 DNA repair using InterProscan version 5 [80]; 6) number of tRNA genes is the number of tRNA
448 genes inferred to be present using the tRNAscan-SE 2.0 program[81]; and 7) d_N/d_S was
449 estimated by calculating the average of the ratio of the expected numbers of non-synonymous
450 (d_N) and synonymous substitutions (d_S) across 815 trimmed codon-based BUSCO gene
451 alignments under the YN98 (F3X4)[82] codon model and the free ratio model using bppml and
452 MapNH in the bio++ libraries[83], following the study by Bolívar et al.[59].
453
454 **Statistical analyses**
455 All statistical analyses were performed in R v. 3.4.2 (R core team 2017). Pearson's correlation
456 coefficient was used to test for correlations among seven variables. To account for phylogenetic
457 relationships of species in correlation analysis, we used the R package ape v5.1[84] in order to

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

458 compute phylogenetically independent contrasts following the method described by
459 Felsenstein[85].

460

461 **Ancestral state reconstruction**

462 To reconstruct ancestral character states for each of seven continuous properties, we used the
463 R package phytools v0.6.44 function *contMap* [86] to infer ancestral character states across
464 internal nodes using the maximum likelihood method with the function *fastAnc* and to
465 interpolate the states along each edge using equation [2] of Felsenstein[85]. The input tree was
466 derived from the concatenation-based ML with branch lengths, which was then pruned to
467 keep the 1,093 taxa from the subphyla Pezizomycotina and Saccharomycotina.

468

469 ## Data availability

470 All genome assemblies and proteomes are publicly available in the Zenodo repository:
471 https://doi.org/10.5281/zenodo.3783970. Multiple sequence alignments, phylogenetic trees,
472 trait ancestral character state reconstructions, log files, R codes, and custom Perl scripts are
473 available on the figshare repository (https://doi.org/10.6084/m9.figshare.12196149;
474 https://figshare.com/articles/
475 Phylogenomics_and_contrasting_modes_of_genome_evolution_in_Ascomycota/12196149 –
476 please note that this link will become active upon publication).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Heitman, J. *et al. The Fungal Kingdom*. (ASM Press, 2017).

2. Shen, X.-X. *et al.* Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* **175**, 1533–1545 (2018).

3. Peter, J. *et al.* Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–344 (2018).

4. Odds, F. C., Brown, A. J. P. & Gow, N. A. R. *Candida albicans* genome sequence: a platform for genomics in the absence of genetics. *Genome Biol.* **5**, 230 (2004).

5. Sharma, C., Kumar, N., Pandey, R., Meis, J. F. & Chowdhary, A. Whole genome sequencing of emerging multidrug resistant *Candida auris* isolates in India demonstrates low genetic variation. *New Microbes New Infect.* **13**, 77–82 (2016).

6. Spatafora, J. W. *et al.* The Fungal Tree of Life: from Molecular Systematics to Genome-Scale Phylogenies. *Microbiol. Spectr.* **5**, (2017).

7. Ma, L.-J. *et al.* Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367–373 (2010).

8. Plissonneau, C., Hartmann, F. E. & Croll, D. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* **16**, 5 (2018).

9. Dean, R. A. *et al.* The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986 (2005).

10. Rokas, A., Mead, M. E., Steenwyk, J. L., Oberlies, N. H. & Goldman, G. H. Evolving moldy murderers: *Aspergillus* section *Fumigati* as a model for studying the repeated

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

499 evolution of fungal pathogenicity. *PLOS Pathog.* **16**, e1008315 (2020).

500 11. Wood, V. *et al.* The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**,
501 871–880 (2002).

502 12. Ma, L. *et al.* Genome analysis of three *Pneumocystis* species reveals adaptation
503 mechanisms to life exclusively in mammalian hosts. *Nat. Commun.* **7**, 10740 (2016).

504 13. Schoch, C. L. *et al.* The Ascomycota Tree of Life: A Phylum-wide Phylogeny
505 Clarifies the Origin and Evolution of Fundamental Reproductive and Ecological Traits.
506 *Syst. Biol.* **58**, 224–239 (2009).

507 14. Beimforde, C. *et al.* Estimating the Phanerozoic history of the Ascomycota lineages:
508 Combining fossil and molecular data. *Mol. Phylogenet. Evol.* **77**, 307–319 (2014).

509 15. Prieto, M. & Wedin, M. Dating the Diversification of the Major Lineages of
510 Ascomycota (Fungi). *PLoS One* **8**, e65576 (2013).

511 16. James, T. Y. *et al.* Reconstructing the early evolution of Fungi using a six-gene
512 phylogeny. *Nature* **443**, 818–822 (2006).

513 17. Spatafora, J. W. *et al.* A five-gene phylogeny of Pezizomycotina. *Mycologia* **98**, 1018–
514 1028 (2006).

515 18. McCarthy, C. G. P. & Fitzpatrick, D. A. Multiple Approaches to Phylogenomic
516 Reconstruction of the Fungal Kingdom. in *Advances in Genetics* 211–266 (Elsevier
517 Inc., 2017).

518 19. Nagy, L. G. & Szöllősi, G. Fungal Phylogeny in the Age of Genomics: Insights Into
519 Phylogenetic Inference From Genome-Scale Datasets. in *Advances in Genetics* 49–72

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(2017).

20. Riley, R. *et al.* Comparative genomics of biotechnologically important yeasts. *Proc. Natl. Acad. Sci. USA* **113**, 9882–9887 (2016).

21. Rajeh, A., Lv, J. & Lin, Z. Heterogeneous rates of genome rearrangement contributed to the disparity of species richness in Ascomycota. *BMC Genomics* **19**, 282 (2018).

22. Higgins, S. A., Schadt, C. W., Matheny, P. B. & Löffler, F. E. Phylogenomics Reveal the Dynamic Evolution of Fungal Nitric Oxide Reductases and Their Relationship to Secondary Metabolism. *Genome Biol. Evol.* **10**, 2474–2489 (2018).

23. Zhou, X., Shen, X.-X., Hittinger, C. T. & Rokas, A. Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets. *Mol. Biol. Evol.* **35**, 486–503 (2018).

24. Taylor, J. W. & Berbee, M. L. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* **98**, 838–849 (2006).

25. Padovan, A. C. B., Sanson, G. F. O., Brunstein, A. & Briones, M. R. S. Fungi Evolution Revisited: Application of the Penalized Likelihood Method to a Bayesian Fungal Phylogeny Provides a New Perspective on Phylogenetic Relationships and Divergence Dates of Ascomycota Groups. *J. Mol. Evol.* **60**, 726–735 (2005).

26. Gueidan, C., Ruibal, C., de Hoog, G. S. & Schneider, H. Rock-inhabiting fungi originated during periods of dry climate in the late Devonian and middle Triassic. *Fungal Biol.* **115**, 987–996 (2011).

27. Stajich, J. E. Fungal Genomes and Insights into the Evolution of the Kingdom.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Microbiol. Spectr. **5**, (2017).

28. Shen, X.-X. *et al.* Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. *G3 Genes|Genomes|Genetics* **6**, 3927–3939 (2016).

29. Kjærbølling, I. *et al.* A comparative genomics study of 23 *Aspergillus* species from section Flavi. *Nat. Commun.* **11**, 1106 (2020).

30. Haridas, S. *et al.* 101 *Dothideomycetes* genomes: A test case for predicting lifestyles and emergence of pathogens. *Stud. Mycol.* **96**, 141–153 (2020).

31. Okagaki, L. H., Sailsbery, J. K., Eyre, A. W. & Dean, R. A. Comparative genome analysis and genome evolution of members of the magnaporthaceae family of fungi. *BMC Genomics* **17**, 135 (2016).

32. Marcet-Houben, M. & Gabaldón, T. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet.* **26**, 5–8 (2010).

33. Wisecaver, J. H., Slot, J. C. & Rokas, A. The Evolution of Fungal Metabolic Pathways. *PLoS Genet.* **10**, e1004816 (2014).

34. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).

35. Wang, H., Xu, Z., Gao, L. & Hao, B. A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evol. Biol.* **9**, 195 (2009).

36. Naranjo-Ortiz, M. A. & Gabaldón, T. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biol. Rev.* **94**, 2101–2137 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

562 37. Floudas, D. *et al.* The Paleozoic Origin of Enzymatic Lignin Decomposition
563 Reconstructed from 31 Fungal Genomes. *Science* **336**, 1715–1719 (2012).
564 38. Robbertse, B., Reeves, J. B., Schoch, C. L. & Spatafora, J. W. A phylogenomic
565 analysis of the Ascomycota. *Fungal Genet. Biol.* **43**, 715–725 (2006).
566 39. Choi, J. & Kim, S.-H. A genome Tree of Life for the Fungi kingdom. *Proc. Natl.*
567 *Acad. Sci. USA* 201711939 (2017).
568 40. Shen, X.-X., Hittinger, C. T. & Rokas, A. Contentious relationships in phylogenomic
569 studies can be driven by a handful of genes. *Nat. Ecol. Evol.* **1**, 0126 (2017).
570 41. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and
571 the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
572 42. Shen, X.-X. *et al.* Enlarged Multilocus Data set Provides Surprisingly Younger Time
573 of Origin for the Plethodontidae, the Largest Family of Salamanders. *Syst. Biol.* **65**,
574 66–81 (2016).
575 43. Liu, L. *et al.* Genomic evidence reveals a radiation of placental mammals
576 uninterrupted by the KPg boundary. *Proc. Natl. Acad. Sci. USA* **114**, E7282–E7290
577 (2017).
578 44. Hedges, S. B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of life reveals
579 clock-like speciation and diversification. *Mol. Biol. Evol.* **32**, 835–845 (2015).
580 45. Welch, J. J., Bininda-Emonds, O. R. & Bromham, L. Correlates of substitution rate
581 variation in mammalian protein-coding sequences. *BMC Evol. Biol.* **8**, 53 (2008).
582 46. Thomas, J. A., Welch, J. J., Lanfear, R. & Bromham, L. A Generation Time Effect on

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

583 the Rate of Molecular Evolution in Invertebrates. *Mol. Biol. Evol.* **27**, 1173–1180
584 (2010).
585 47. Salari, R. & Salari, R. Investigation of the Best *Saccharomyces cerevisiae* Growth
586 Condition. *Electron. physician* **9**, 3592–3597 (2017).
587 48. London, R. *et al.* An Automated System for Rapid Non-Destructive Enumeration of
588 Growing Microbes. *PLoS One* **5**, e8609 (2010).
589 49. TRINCI, A. P. J. A Kinetic Study of the Growth of *Aspergillus nidulans* and Other
590 Fungi. *J. Gen. Microbiol.* **57**, 11–24 (1969).
591 50. Gillie, O. J. Growth of *Neurospora crassa* in Unstirred Liquid Cultures. *J. Gen.*
592 *Microbiol.* **51**, 179–184 (1968).
593 51. Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic
594 Instability in Cancer. *Cell* **168**, 644–656 (2017).
595 52. Lang, G. I., Parsons, L. & Gammie, A. E. Mutation Rates, Spectra, and Genome-Wide
596 Distribution of Spontaneous Mutations in Mismatch Repair Deficient Yeast. *G3*
597 *Genes|Genomes|Genetics* **3**, 1453–1465 (2013).
598 53. Steenwyk, J. L. *et al.* Extensive loss of cell-cycle and DNA repair genes in an ancient
599 lineage of bipolar budding yeasts. *PLOS Biol.* **17**, e3000255 (2019).
600 54. Milo, S., Misgav, R. H., Hazkani-Covo, E. & Covo, S. Limited DNA repair gene
601 repertoire in Ascomycete yeast revealed by comparative genomics. *Genome Biol. Evol.*
602 **11**, 3409–3423 (2019).
603 55. Bromham, L. The genome as a life-history character: why rate of molecular evolution

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

varies between mammal species. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2503–2513 (2011).

56. Šmarda, P. *et al.* Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc. Natl. Acad. Sci. USA* **111**, E4096–E4102 (2014).

57. Zhu, Y. O., Siegal, M. L., Hall, D. W. & Petrov, D. A. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. USA* **111**, E2310–E2318 (2014).

58. Liu, H. *et al.* Tetrad analysis in plants and fungi finds large differences in gene conversion rates but no GC bias. *Nat. Ecol. Evol.* **2**, 164–173 (2018).

59. Bolívar, P., Guéguen, L., Duret, L., Ellegren, H. & Mugal, C. F. GC-biased gene conversion conceals the prediction of the nearly neutral theory in avian genomes. *Genome Biol.* **20**, 5 (2019).

60. Zdobnov, E. M. *et al.* OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749 (2017).

61. Steenwyk, J. L., Shen, X.-X., Lind, A. L., Goldman, G. H. & Rokas, A. A Robust Phylogenomic Time Tree for Biotechnologically and Medically Important Fungi in the Genera *Aspergillus* and *Penicillium*. *MBio* **10**, 1–25 (2019).

62. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**, ii215–ii225 (2003).

63. Krassowski, T. *et al.* Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. *Nat. Commun.* **9**, 1887 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

64. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

65. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

66. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

67. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).

68. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).

69. Mirarab, S. *et al.* ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).

70. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).

71. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).

72. Sayyari, E. & Mirarab, S. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Mol. Biol. Evol.* **33**, 1654–1668 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

646 73. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an r package for
647 visualization and annotation of phylogenetic trees with their covariates and other
648 associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
649 74. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics
650 Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
651 75. Mello, B., Tao, Q., Tamura, K. & Kumar, S. Fast and Accurate Estimates of
652 Divergence Times from Big Data. *Mol. Biol. Evol.* **34**, 45–50 (2017).
653 76. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**,
654 1586–1591 (2007).
655 77. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for
656 Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
657 78. Taylor, T. N., Hass, H., Kerp, H., Krings, M. & Hanlin, R. T. Perithecial ascomycetes
658 from the 400 million year old Rhynie chert: an example of ancestral polymorphism.
659 *Mycologia* **97**, 269–285 (2005).
660 79. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically
661 mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–
662 644 (2008).
663 80. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
664 *Bioinformatics* **30**, 1236–1240 (2014).
665 81. Lowe, T. M. & Chan, P. P. tRNAscan-SE On-line: integrating search and context for
666 analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54-7 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

667 82. Yang, Z. & Nielsen, R. Synonymous and nonsynonymous rate variation in nuclear
668 genes of mammals. *J. Mol. Evol.* **46**, 409–418 (1998).
669 83. Guéguen, L. *et al.* Bio++: Efficient Extensible Libraries and Tools for Computational
670 Molecular Evolution. *Mol. Biol. Evol.* **30**, 1745–1750 (2013).
671 84. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and
672 evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
673 85. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
674 86. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other
675 things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
676 87. Steenwyk, J. L. & Rokas, A. Treehouse: a user-friendly application to obtain subtrees
677 from large phylogenies. *BMC Res. Notes* **12**, 541 (2019).
678

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Acknowledgments

We thank members of the Rokas and Hittinger labs, especially members of the Y1000+ Project for constructive feedback.

Funding Statement

This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University. X.X.S. was supported by the start-up grant from the "Hundred Talents Program" at Zhejiang University and the Fundamental Research Funds for the Central Universities (No. is appending). X.Z. was supported by the National Key Project for Basic Research of China (973 Program, No. 2015CB150600) and the open fund from Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, College of Crop Science, Fujian Agriculture and Forestry University (GBMUC-2018-005). M.G. was supported by the Royal Netherlands Academy of Arts and Sciences. C.T.H. was supported by the National Science Foundation (DEB-1442148), the USDA National Institute of Food and Agriculture (Hatch Project No. 1020204), in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science No. DE-SC0018409), the Pew Charitable Trusts (Pew Scholar in the Biomedical Sciences), and the Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (H. I. Romnes Faculty Fellow). J.L.S. and A.R. were supported by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program. A.R. was supported by the National Science Foundation (DEB-1442113), the Guggenheim Foundation, and the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Burroughs Wellcome Fund.

Author contributions

Study conception and design: X.X.S., C.T.H., A.R..; Acquisition of data: X.X.S.; Analysis and interpretation of data: X.X.S., J.L.S., A.L.L., D.A.O., X.Z., J.K., Y.L., M.G., C.T.H., A.R..; Drafting of manuscript: X.X.S., A.R.; Critical revision: all authors.

Competing interests

The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure Legends

Fig. 1 Maximum likelihood (ML) phylogeny of 1,107 taxa in the fungal phylum Ascomycota. The concatenation-based ML phylogeny (*lnL*= -269043834.145) was inferred from a set of 815 BUSCO amino acid genes (total 56, 2376 sites) under a single LG + G4 substitution model using IQ-TREE multicore version 1.5.1. The number of species sampled in each subphylum is given in parentheses. Internal branch labels are acronyms for 12 major clades in the subphylum Saccharomycotina and 9 classes in the subphylum Pezizomycotina. The bar next to each species indicates the guanine-cytosine (GC) content. On average, lineages in the subphylum Saccharomycotina have significantly lower GC content (49.6% vs. 40.6%; Wilcoxon rank-sum test; *P*-value = 3.07 x 10^{-103}) but higher evolutionary rate (1.80 substitutions per site vs. 1.12 substitutions per site; Wilcoxon rank-sum test; *P*-value = 6.57 x 10^{-126}) compared to lineages in the subphylum Pezizomycotina. The complete phylogenetic relationships of 1,107 taxa are given in Supplementary Fig. 2 and in the Figshare repository. For easy determination of the relationships among any subset of taxa, the phylogeny is also available through Treehouse[87].

Fig. 2 Distribution of phylogenetic signal for three historically contentious relationships within Ascomycota. For each relationship / internal branch (**a**: which class(es) is the sister group to the rest of the Pezizomycotina?; **b**: what is the relationship among three classes Schizosaccharomycetes, Pneumocystidomycetes, and Taphrinomycetes in the subphylum Taphrinomycotina?; **c**: what is the relationship among three subphyla Pezizomycotina, Saccharomycotina, and Taphrinomycotina in the phylum Ascomycota?), we applied the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

732 framework presented by Shen et al.[40] to examine proportions of genes (left panel) and sites
733 (right panel) supporting each of three competing hypotheses (topology 1 or T1 in red,
734 topology 2 or T2 in green, and topology 3 or T3 in yellow). Note that both concatenation- and
735 coalescent-based approaches supported T1 in our study. Dashed horizontal lines on 1/3 y-axis
736 value denote expectation of proportion of genes / sites under a polytomy scenario. The *G*-test
737 was used to test if the sets of three values are significantly different (***: *P*-value ≤ 0.001).
738 All values are given in Supplementary Tables 4 and 5. Input and output files associated with
739 phylogenetic signal estimation are also deposited in the Figshare repository.
740
741 **Fig. 3 Contrasting patterns for seven genomic properties between Pezizomycotina and**
742 **Saccharomycotina. a**, For each species in Pezizomycotina (colored in red, n=761) and
743 Saccharomycotina (colored in green, n=332), we calculated evolutionary rate, GC content,
744 genome size, number of protein-coding genes, number of DNA repair genes, number of tRNA
745 genes, and d_N/d_S (see the Methods section for details). The Wilcoxon rank-sum test was used
746 to test if the sets of values in two subphyla are significantly different. **b**, Pairwise standard
747 Pearson's correlation coefficient among pairs of the seven genomic properties were
748 conducted using R 3.4.2 for Pezizomycotina (lower diagonal) and Saccharomycotina (upper
749 diagonal), respectively. For each cell, the top value corresponds to *P*-value (NS: *P*-
750 value >0.05; *: *P*-value ≤0.05; **: *P*-value ≤0.01； ***: *P*-value ≤ 0.001), whereas the
751 bottom value corresponds to Pearson's coefficient value. Orange cells denote instances where
752 correlation trends in Pezizomycotina and Saccharomycotina are in opposite directions,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

753 whereas blue cells denote instances where the trends are in the same direction. The detailed
754 values of all seven properties in Pezizomycotina and Saccharomycotina are given in
755 Supplementary Table 7. The correlations among these seven properties are largely consistent
756 before (i.e., standard Pearson's correlations) and after (i.e., phylogenetically independent
757 contrasts) accounting for correlations due to phylogeny (see Supplementary Table 8).
758
759
760 **Fig. 4 Contrasting modes of genome evolution in Pezizomycotina and**
761 **Saccharomycotina. a**, For each of the seven genomic properties examined (see the Methods
762 section for details) , we reconstructed them as continuous traits on the species phylogeny
763 (Fig. 1) and visualized their ancestral states with the R package phytools v0.6.44 [86]. Heatmap
764 bars denote ancestral state values from small (blue) to large (red). Three ancestral state values
765 next to three red dots are shown for the ancestor of the subphyla Pezizomycotina and
766 Saccharomycotina, the ancestor of the subphylum Pezizomycotina, and the ancestor of the
767 subphylum Saccharomycotina, respectively. **b**, Phylogeny key showing the placement of the
768 21 nodes representing the last common ancestors of the 12 major clades in the subphylum
769 Saccharomycotina and of the 9 classes in the subphylum Pezizomycotina; the 21 nodes are
770 indicated by the red dots. The orders of branches in **a** are identical to those in **b**.
771

Table 1. Summary of values for seven genomic properties in extant Saccharomycotina and Pezizomycotina and in the last common ancestors of Saccharomycotina and Pezizomycotina.

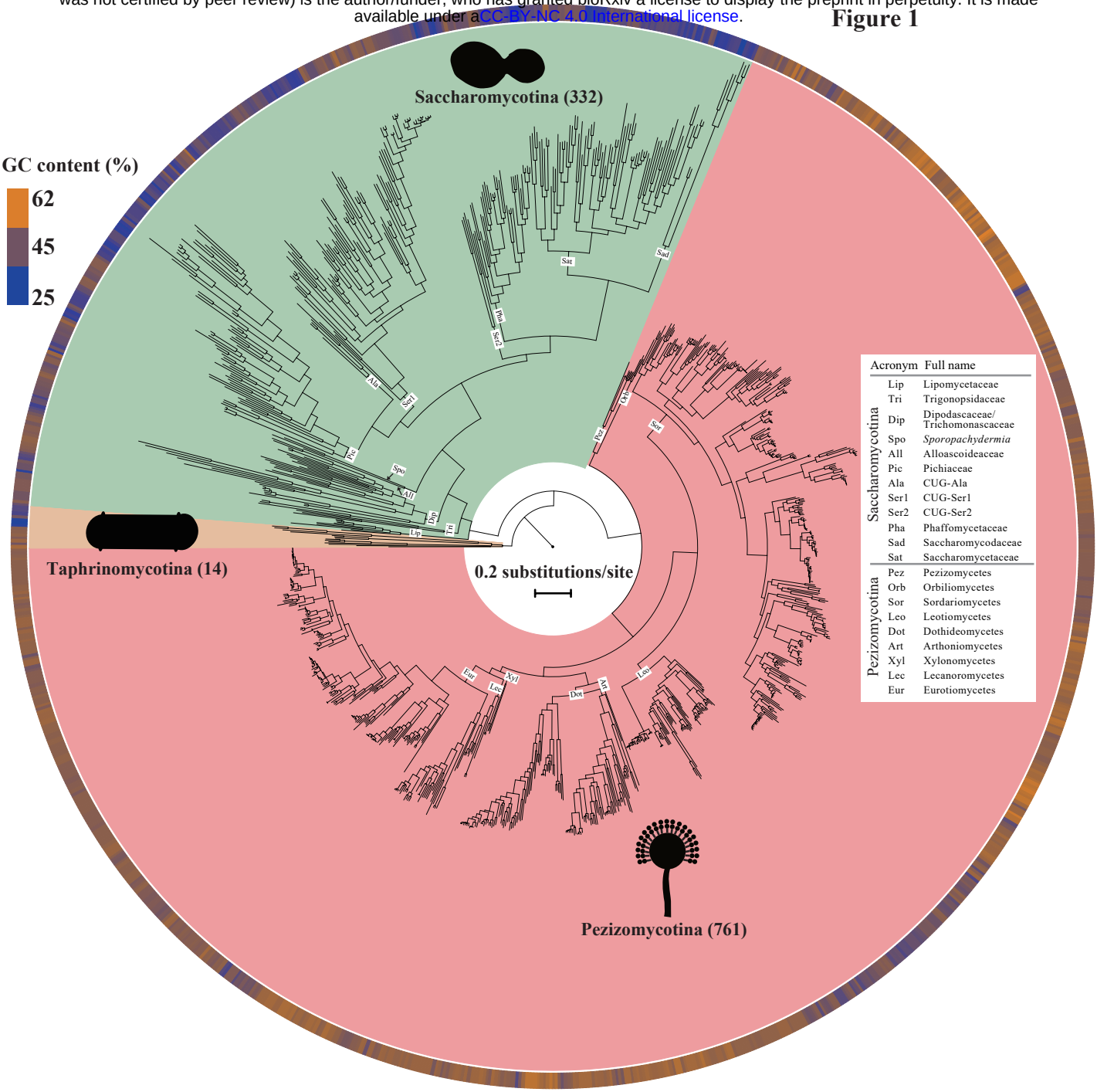
Property	Extant Saccharomycotina* (n=332)	Extant Pezizomycotina* (n=761)	Saccharomycotina ancestor	Pezizomycotina ancestor	Difference between two extant lineages	Difference between two ancestors
Evolutionary rate (amino acid substitutions/ site)	1.80	1.12	1.1	0.9	0.68	0.2
GC content (%)	40	50	43	47	10	4
Genome size (Mb)	13	39	23	42	26	19
No. of genes	5,734	10,847	7,000	9,400	5,113	2,400
No. of DNA repair genes	41	54	44	52	13	8
No. of tRNA genes	179	146	160	170	33	10
d_N/d_S	0.053	0.063	0.052	0.058	0.01	0.006

* denote average values.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 1



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 2

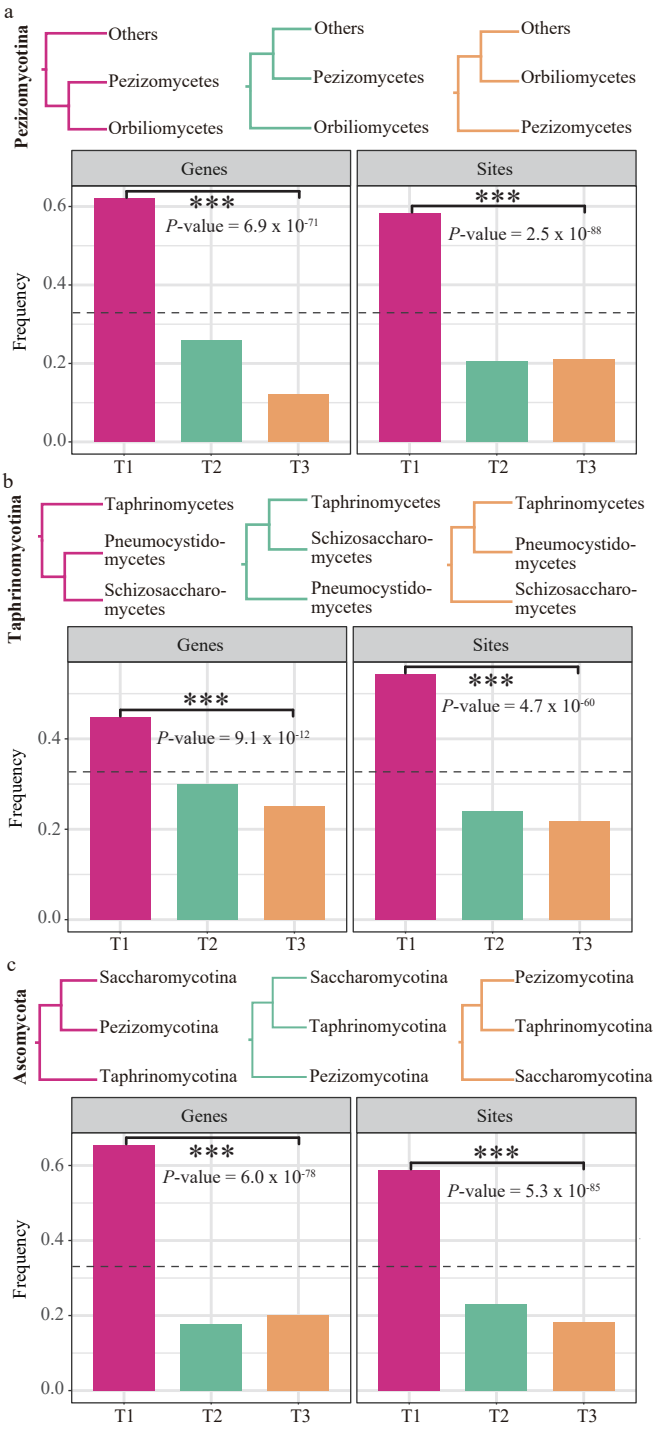


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

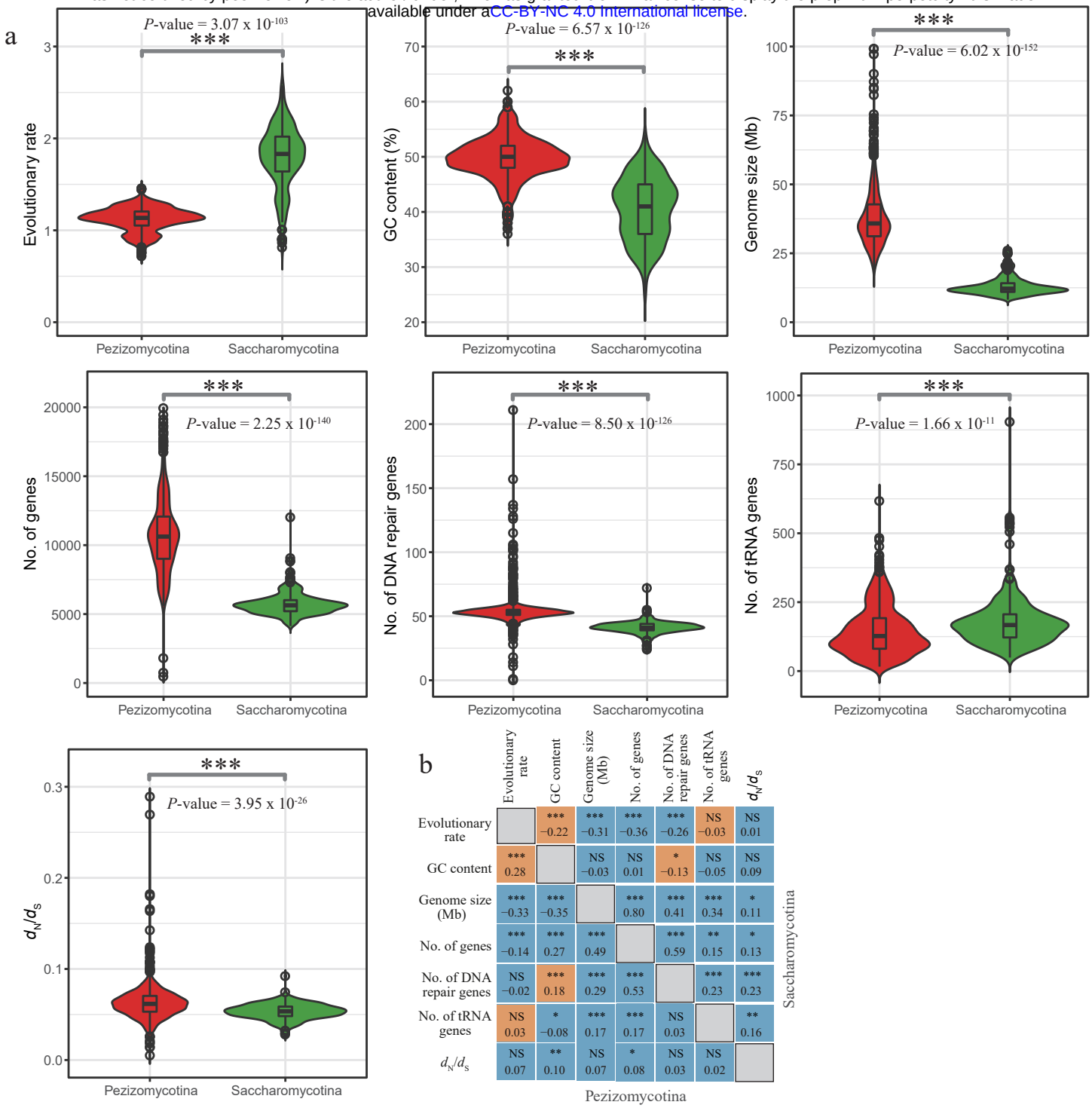


Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.11.088658; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.