

**Discovery of additional ancient genome duplications
in yeasts**

**Kyle T. David¹, Linda Horianopoulos^{2,3}, Carla Gonçalves⁴, Jacob L. Steenwyk⁵, Ana
Pontes⁴, Paula Gonçalves⁴, Chris Todd Hittinger², Matt Pennell⁶, and Antonis Rokas^{1,*}**

¹Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA;
Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235, USA

² Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Center for Genomic
Science Innovation, DOE Great Lakes Bioenergy Research Center, Wisconsin Energy Institute,
University of Wisconsin-Madison, Madison, WI 53726, USA

³Department of Food Science, University of Guelph, Guelph, ON, Canada, N1G 2W1

⁴UCIBIO, i4HB, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa, Caparica 2829-516, Portugal

⁵Howard Hughes Medical Institute and the Department of Molecular and Cell Biology,
University of California, Berkeley, CA 94720, United States

⁶ Department of Computational Biology, Cornell University, Ithaca NY 14853 USA

*Corresponding author: antonis.rokas@vanderbilt.edu

Lead contact: Antonis Rokas

Keywords: Whole Genome Duplication, Yeasts, Polyploidy, Convergent Evolution

Summary

Whole genome duplication (WGD) has had profound macroevolutionary impacts on diverse lineages^{1,2}, preceding adaptive radiations in vertebrates^{3–5}, teleost fish^{6,7}, and angiosperms^{8–10}. In contrast to the many known ancient WGDs in animals^{11,12} and especially plants^{13–15}, we are aware of evidence for only four in fungi^{16,17}. The oldest of these occurred ~100 million years ago (mya) and is shared by ~60 extant Saccharomycetales species^{18,19}, including the baker's yeast *Saccharomyces cerevisiae*. Notably, this is the only known ancient WGD in the yeast subphylum Saccharomycotina. The dearth of ancient WGD events in fungi remains a mystery¹⁶. Some studies have suggested that fungal lineages that experience chromosome²⁰ and genome¹⁶ duplication quickly go extinct, leaving no trace in the genomic record, while others contend that the lack of known WGD is due to an absence of data^{16,17}. Under the second hypothesis, additional sampling and deeper sequencing of fungal genomes should lead to the discovery of more WGD events. Coupling hundreds of recently published genomes from nearly every described Saccharomycotina species with three additional long-read assemblies, we discovered three novel WGD events. While the functions of retained duplicate genes originating from these events are broad, they bear similarities to the well-known WGD that occurred in the Saccharomycetales¹⁸. Our results suggest that WGD may be a more common evolutionary force in fungi than previously believed.

Results and Discussion

Evidence for three whole genome duplication events in the Dipodascales

To detect signatures of ancient WGD we first used the gene/species tree reconciliation algorithm implemented in OrthoFinder v3.0.1b1²¹ to obtain a preliminary estimate of gene duplication

across the Saccharomycotina phylogeny. High rates of gene duplication along specific lineages of the species tree can be indicative of WGD events and have been used to identify ancient WGDs in plants²²⁻²⁴, animals²⁵, and fungi²⁶. An initial analysis using a 400-species backbone phylogeny (Figure S1) successfully recovered the known ancient WGD near the base of Saccharomycetales (labelled WGD1). This lineage possessed the second-highest rate of duplications, over 10x higher than the average internal lineage. However, the lineages with the first and third highest rates of duplication occurred in the Dipodascales, a clade separated from the Saccharomycetales by ~300 million years (my) of evolution, prompting further investigation.

To improve the resolution of gene family evolution within both clades, we increased our sampling in Saccharomycetales (135 genomes) and Dipodascales (184 genomes) and reperformed the gene/species tree reconciliation analysis. The lineage containing WGD1 was again successfully identified by a spike in duplication rate, experiencing 53.1 gene duplications / my (Figure 1). Two internal spikes were also evident in Dipodascales. The most recent ancestral lineage to the genera *Dipodascus* and *Geotrichum* (spanning from 201-132 mya) had a duplication rate of 21.8 duplications / my, leading us to hypothesize that a WGD may have occurred along this lineage (labelled WGD2). Similarly, the most recent ancestral lineage to the species *Magnusiomyces tetraspermus* and *Saprochaete suaveolens* (spanning from 62-38 mya) had a duplication rate of 23.62 duplications / my (labelled WGD3) (Figure 1). Additionally, the highest species-specific duplication rate by far belonged to *M. magnusii* with 130 gene duplications / my, 2.4x higher than any other Dipodascales species. *M. magnusii* also possesses the largest known Saccharomycotina genome (43.2Mb, 3.4x larger than *S. cerevisiae*), leading us

to hypothesize another WGD (labelled WGD4, spanning from 27-0 mya) specific to this species alone.

While gene/species tree reconciliation approaches are useful in identifying putative WGDs, they are limited in their ability to distinguish true WGD from bouts of single or segmental gene duplication events²⁷. Additionally, reconciliation approaches are generally not sensitive to biological processes, such as gene conversion or unbalanced paralog retention (e.g., via paralog loss in outgroup taxa) that can cause duplication events to appear younger²⁸. Such is likely the case in our dataset, where both WGD2 and WGD3 are directly ancestral to lineages with similarly elevated rates of duplication (Figure 1). High duplication rates are therefore necessary but not sufficient for WGD inference, nor for the confident identification of the timing of WGD on a phylogeny.

A complementary approach to identify WGD involves the detection of colinear segments of paralogs (called multiplicons)^{29,30}. Many multiplicons distributed throughout a genome are difficult to explain without invoking WGD and provide compelling evidence^{1,2}. However, such analysis requires highly contiguous assemblies. Assemblies with fewer than 100 contigs were available for one post-WGD2 (*G. candidum*: 28 contigs) and one post-WGD3 (*Sap. suaveolens*: 12 contigs) species. To generate additional evidence for hypothesized WGDs and their placements on the yeast phylogeny we re-sequenced three additional species: *D. fermentans* for WGD2 (19 contigs), *M. tetraspermus* for WGD3 (33 contigs) and *M. magnusii* for WGD4 (71 contigs). These 5 genomes, along with 81 other highly contiguous Saccharomycotina genome assemblies (Data S1) were then searched for multiplicons (Figure 2A).

90
91 Of the 86 highly contiguous genomes, every species predicted to have experienced a
92 WGD had more colinear paralogs than those that did not (Figure 2A). As expected, cumulative
93 multiplicon size negatively scaled with age. The genome predicted to have experienced the most
94 recent WGD4 had the largest sum of total multiplicons (6,459 paralogs), followed by the two
95 post-WGD3 genomes (3,409 & 3,352 paralogs, Figure 2A). Between the two oldest WGD
96 events, post-WGD2 genomes retained more duplicates on average (865.5 paralogs) than post-
97 WGD1 genomes (737.8 paralogs). By contrast, the average non-WGD genome possessed 46.5
98 colinear paralogs. Interestingly, whereas multiplicons for most post-WGD species occurred in
99 duplicates, in *M. magnusii*, they appear in tetraplicate (Figure 2B), providing further evidence for
100 WGD4 as an event specific to this species. WGD4 is also supported by examination of
101 multiplicon size in *M. magnusii*, which is nearly twice (189.5%) that of sister species *Sap.*
102 *suaveolens*.

103
104 Multiplicons between *D. fermentans* and *G. candidum* had many orthogroups in common,
105 as did *Sap. suaveolens*, *M. tetraspermus*, and *M. magnusii* (Figure 2B). However, multiplicons
106 between post-WGD2 and post-WGD3 species (*D. fermentans* and *Sap. suaveolens*, for example)
107 contained largely distinct orthogroups (Data S2). Of all possible cross-species multiplicon pairs
108 that shared a Jaccard similarity index of at least 0.25 with respect to orthogroup membership,
109 only 1.7% were between species descended from different WGD events, consistent with our
110 inference that WGD2 and WGD3 represent two distinct events (Figure 1).

Another common hallmark of ancient WGD is large increases in the number of chromosomes across lineages³¹. However, fungi possess small, loosely-packed chromosomes, which make karyotyping difficult³². Furthermore, Saccharomycotina yeasts have extremely diverse telomeres^{33,34} and centromeres^{35,36}, which make chromosome number estimation challenging even for highly-contiguous assemblies. Despite these obstacles, chromosome count estimates from several *Dipodascales* species exist, including one representative from each proposed WGD. For species not inferred to have undergone WGD, count estimates range from four³⁷ to five^{38,39} chromosomes. This estimate increases to seven chromosomes for the post-WGD3 species *Sap. suaveolens*⁴⁰ and eight⁴¹ or nine⁴² for the post-WGD2 species *G. candidum*. Finally, 13 chromosomes have been inferred from pulsed field gel electrophoresis for *M. magnusii*⁴³, which is predicted to have undergone two rounds of WGD (both WGD3 and WGD4). Therefore, while limited chromosome count estimates in *Dipodascales* preclude a formal analysis, we interpret the available data to be consistent with our conclusions (Table 1).

WGD can occur both within lineages (autopolyploidization) or between lineages via hybridization^{26,44} (allopolyploidization). WGD1 was originally hypothesized to be an autopolyploidization event due to the highly conserved gene order within descendent genomes¹⁹. However, a gene/species tree reconciliation analysis later recovered a lineage pre-dating WGD1 with a higher rate of gene duplications than the WGD1 lineage itself²⁶. The authors posit this unusual finding could be explained by gene tree discordance produced through hybridization²⁶, which suggests allopolyploidization. As our reconciliation analysis recovered no such peak in gene duplication rates preceding any of the novel WGD events (Figure 1), we hypothesize that

these events occurred either through autopolyploidization or through allopolyploidization between closely related parent species.

In contrast to ancient WGD, which was presumed to be extremely rare in Saccharomycotina, recent allopolyploidization events are quite common. Polyploidy is a frequent outcome of hybridization between extant taxa, including in several industrial strains^{45–50}. To prevent these events from being mistaken for ancient WGD, no known hybrids were included in the dataset. However, as a relatively young event, WGD4 bears many similarities to recent allopolyploidization events. WGD4 is specific to just one species, and sequence identity between ohnologs is much greater (82%) than older WGDs in the Dipodascales (WGD2: 46%, WGD3: 69%) (Data S3). Therefore, further investigation is warranted to better elucidate the timing and mechanism of WGD4 to determine whether it represents a recent allopolyploidization of a cryptic hybrid or a true “ancient” WGD.

Functional consequences of ancient genome duplication in yeasts

WGD is often referred to as an engine of evolutionary innovation^{2,4}, preceding jumps in phenotypic complexity in plants⁵¹ and animals⁵. In fungi, it has been hypothesized that WGD1 facilitated aerobic glycolysis in several yeast species^{52,53}. Also known as the “Crabtree/Warburg Effect”, this biochemical process empowers baking, brewing, and winemaking. However, others have noted that some species that have not undergone WGD1 are still capable of aerobic glycolysis to a lesser extent^{54,55}. As this example illustrates, identifying cause-effect relationships from rare evolutionary events, such as WGD, is exceedingly difficult^{56,57}. The discovery of

additional WGD events in Saccharomycotina affords new opportunities for understanding how WGD events contribute to evolutionary innovation.

To test whether certain functional gene classes were more likely to be retained in duplicate following WGD, and if these classes were convergently shared across events, we ran enrichment analysis on retained ohnologs for each event. Notably, WGD4 ohnologs were significantly enriched (adjusted $p < 0.05$) for only five InterPro Gene Ontology⁵⁸ categories (Data S4) and for none of the Kyoto Encyclopedia of Genes and Genomes⁵⁹ (KEGG) pathways (Data S5). As 76.6% of all *M. magnusii* genes are ohnologs, we hypothesize that not enough time has passed since WGD4 for genes to become lost and patterns within retained copies to emerge. As discussed above, it is not yet known whether WGD4 represents a truly ancient WGD or a recent allopolyploidization. Therefore, we turn our attention to the three older WGD1, WGD2, and WGD3 (WGD1,2,3 for short) for the remainder of this discussion.

The functional effects of WGD appeared widespread. Ohnologs were significantly enriched (adjusted $p < 0.01$) in at least six of the seven main KEGG pathway categories across WGD1,2,3 (Data S5). Despite the diversity of affected pathways, commonalities remain. Four related metabolic pathways were significantly enriched (adjusted $p < 0.01$) in one of the three WGD events (Data S5). For example, WGD2 ohnologs were enriched in the propanoate metabolic pathway. Propanoate esters and other related compounds are responsible for the characteristic fruity scent of *G. candidum*⁶⁰, a popular yeast in cheesemaking⁶¹ as well as one of the major natural flavorants of cocoa^{62,63}. Many other *Geotrichum/Dipodascus* species are similarly known for their flavor-enhancing properties⁶⁴.

179

180 As mentioned above, the metabolic effects of WGD1 have been studied previously^{52–54,65}.
181 Enzymes and hexose transporters involved in glycolysis have been preferentially retained in
182 post-WGD1 species, in particular those genes whose dosage most greatly increases glycolytic
183 flux⁶⁵. We find all ten of these genes have been retained in duplicate from at least one of the
184 three newly reported WGD events, even those not retained from WGD1. Three gene families,
185 *HXT*, *ENO*, and *PYK* were retained across all four events. The enzyme products of *ENO* and
186 *PYK* catalyze the last two steps of glycolysis, whose overexpression greatly increases glycolytic
187 rates^{55,66}. *HXT* is also notable, as dosage of these hexose transporter-encoding genes has the
188 single biggest positive impact on glycolytic flux by a wide margin⁶⁵.

189

190 It has also been hypothesized that WGD1 coincided with the rise of angiosperms and the
191 increased availability of simple sugars found in fruit and nectar¹⁸. Our discovery of multiple
192 WGDs reveals that these events are not tied to specific time periods. However, we do find that
193 WGD events are associated with metabolic specialism in yeasts. Post-WGD species metabolize
194 significantly fewer substrates than species that have not undergone WGD (phylogenetic ANOVA
195 $p < 2.2 \times 10^{-16}$), a result that persists even if WGD1 taxa are excluded (phylogenetic ANOVA $p = 1 \times 10^{-3}$).
196 We hypothesized that additional gene copies in preexisting metabolic pathways produced by
197 WGD events facilitate higher-throughput metabolic output on glucose, reducing evolutionary
198 pressure to metabolize alternative substrates. However, we found that post-WGD species have
199 significantly *reduced* growth rates on the simple sugars mannose (phylogenetic ANOVA $p = 4.6 \times 10^{-2}$)
200 and fructose (phylogenetic ANOVA $p = 1.6 \times 10^{-3}$), while differences on glucose were
201 insignificant (phylogenetic ANOVA $p = 0.9$). Clearly, more work is required to disentangle the

effects of WGD on metabolic rate, growth rate, and niche breadth, an effort we hope will be aided by the discovery of these additional events.

While metabolic pathways were differentially enriched across WGD events, other pathways exhibited stronger signatures of convergence. In particular, several signaling pathways were consistent beneficiaries of WGD. The mitogen-activated protein kinase (MAPK), glucagon, and insulin signaling pathways were all significantly enriched (adjusted $p < 0.01$) across WGD1,2,3 (Data S5, Figure 3, S2). MAPK is an ancient family of kinases that drive phosphorylation cascades, which trigger a variety of cellular mechanisms of *S. cerevisiae* in response to stressful conditions, such as high osmolarity⁶⁷, damage to the cell wall⁶⁸, or starvation⁶⁹. While fungi do not naturally produce glucagon or insulin, these general pathways are deeply conserved across eukaryotes^{70,71}. These pathways contain protein kinases AMPK and PKA which serve important nutrient sensing functions and trigger cellular responses accordingly^{72,73}. Both AMPK and PKA contain duplicate genes from all four WGD events (Figure S2).

Conclusion

We report the presence of three previously unknown whole genome duplication events, denoted WGD2, WGD3, and WGD4 in the Dipodascales clade. Despite ~300 my of evolution separating them, WGD2 and WGD3 share similar outcomes with the previously known WGD1 that occurred in Saccharomycetales, as well as other WGDs across eukaryotes. In particular, ohnologs involved in signal transduction were significantly retained across all confirmed ancient WGDs. Genes with many protein-protein interactions, such as those embedded in dense

signaling networks, are expected to be more sensitive to dosage effects, and therefore more likely to be retained following WGD^{7,74}. Indeed, genes coding for signal transducers have both been preferentially retained following WGD in plants⁷⁵ and animals⁷⁶. Genes that are highly sensitive to dosage are also more likely to be associated with genetic disorders⁷⁷ and ohnologs in humans⁷⁸ and other vertebrates⁷⁹ are known to be more closely associated with disease. Our findings support this connection since many pathways significantly enriched (adjusted $p < 0.01$) with ohnologs from WGD1,2,3 are related to human disease. Proteoglycans in cancer, insulin resistance, and COVID-19 were each enriched by all three events (Data S5). The observed functional similarities between ohnologs from events separated across kingdoms suggest that the effects of WGD and other major evolutionary events may be predictable⁸⁰.

Another feature common to WGD-enriched pathways is their role in adaptation under diverse environments. Previous studies in yeasts have shown how metabolic⁶⁵ and signaling^{69,71} pathways provide heterogeneous responses to hostile conditions and to the quantity and quality of available nutrients. The enrichment of these gene families following WGD may explain why modern polyploid *S. cerevisiae* strains are more fit in challenging environments, such as non-optimal carbon sources⁸¹, human hosts⁸², or brewing vats⁸³. This result may further help to address the more widely observed trend of polyploid plant⁸⁴ and animal⁸⁵ species occurring in extreme, rapidly-changing environments.

The nonrandom distribution of retained ohnologs in environmental information and resource processing networks indicates WGD as a potential driver of evolutionary innovation. Gene duplication is often associated with the gain of new function (neofunctionalization)^{4,86,87}.

WGD provides perhaps the only opportunity for trans-regulated networks to neofunctionalize wholesale by making redundant copies of the entire network. However, more work is required to identify whether various modes of functional divergence are involved, such as neofunctionalization or escape from adaptive conflict⁸⁸, or if simply increasing dosage of conserved copies is sufficient to precipitate major evolutionary change⁶⁵. Previously thought to be largely absent from fungi¹⁶, these results underscore the importance of WGD in all eukaryotic kingdoms. We anticipate more fungal WGDs will be discovered as sampling and sequencing continue to improve, and that these events will yield a fuller portrait of eukaryotic genome evolution.

Resource availability

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Antonis Rokas (antonis.rokas@vanderbilt.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- These data have been deposited in a figshare repository and are publicly available as of the date of publication at doi.org/10.6084/m9.figshare.29852876. The newly sequenced genomes are also available at NCBI under the BioProject ID PRJNA1304572.
- All original code has been deposited in a figshare repository and is publicly available as of the date of publication at <https://doi.org/10.5281/zenodo.17940646>.

- Any additional information required to reanalyze the data reported in this paper is available from the Lead Contact upon request.

Acknowledgements

This work was performed using resources contained within the Advanced Computing Center for Research and Education at Vanderbilt University in Nashville, TN. This work was supported by the NSF (grants DBI-2305612 to K.T.D., DEB-2110403 to C.T.H., and DEB-2110404 to A.R.) and the NIH (R35GM151348 to M.P.). This work was partially supported by FCT—Fundação para a Ciência e a Tecnologia, I.P. (FCT/MCTES; <https://www.fct.pt/>) in the scope of projects UIDP/04378/2020, LA/P/0140/2020 and grant PTDC/BIA-EVL/0604/2021 (to C.G.). Research in the Hittinger Lab is also supported by the United States Department of Agriculture National Institute of Food and Agriculture (Hatch Project 7005101), in part by the Department of Energy (DOE) Great Lakes Bioenergy Research Center (DOE Biological and Environmental Research Office of Science DE–SC0018409). Research in the Rokas Lab is also supported by the NIH/National Institute of Allergy and Infectious Diseases (R01 AI153356). J.L.S. is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation.

Author Contributions

Conceptualization, KTD; data curation, KTD; formal analysis, KTD; investigation, LH, CG, JLS, AP, and PG; resources, LH, CG, CTH, and AR; software, KLD, JLS, visualization, KTD; writing – original draft, KTD; writing – review and editing, KTD, CTH, MP, and AR.

292 **Declaration of Interests**

293 J.L.S. is an advisor to ForensisGroup Inc. J.L.S. is a scientific consultant to FutureHouse Inc.

294 A.R. is on the Advisory Board of *Current Biology* and a scientific consultant for LifeMine

295 Therapeutics, Inc. All other authors declare no competing interests.

296

Figure and Table Legends

Figure 1. Three additional whole genome duplications in Saccharomycotina yeasts. Rates of gene duplication are mapped onto Saccharomycetales (left) and Dipodascales (right) phylogenies. Whole genome duplications indicated along the lineages in which they are predicted to occur. Trees have been pruned for visualization purposes. See Figure S1 for rates of gene duplication across the full backbone phylogeny and Figure S3 for a comparison between Orthofinder and AleRax.

Figure 2. Conserved gene order between duplicated segments (multiplicons) within a genome provides compelling evidence for WGD. A) Cumulative multiplicon size of each highly contiguous Saccharomycotina assembly, as measured by the number of unique paralogs. Genomes predicted to have undergone ancient WGD are highlighted. B) Gene order of select multiplicons. Gene order within contigs is represented by triangles, colored by orthogroup. Homology is indicated by lines drawn between genes. Orthogroups that only appear once are depicted in gray. For a full comparison of shared orthogroups between multiplicons see Data S2. For comparison of individual ohnolog pairs see Data S3.

Figure 3. Impacts of whole genome duplication in two exemplary KEGG pathways: glycolysis (left) and MAPK (right). Genes are colored by which event(s) ohnologs were retained. Transparent genes are those that were not identified in any genome in any amount. See Figure S2 for additional signaling pathways.

319 **Table 1. Chromosome count estimates.** Chromosome counts for available species within the
320 Dipodascales.

STAR Methods

Experimental model and study participant details

Dipodascus fermentans PYCC 3480^T (NRRL Y-1492) was obtained from the Portuguese Yeast Culture Collection (PYCC) and was routinely grown on solid yeast peptone dextrose (YPD) media at 25°C. *M. magnusii* (NRRL Y-17563) and *M. tetraspermus* (NRRL Y-7288) were routinely grown on solid yeast peptone dextrose (YPD) media at room temperature (22°C). Liquid cultures were inoculated from a single colony and grown in 25 ml YPD at room temperature in a 125 ml baffled flask shaking at 225 rpm.

Method details

***D. fermentans* extraction and assembly**

Genomic DNA from overnight grown cultures of *D. fermentans* was obtained using the Quick-DNA Fungal/Bacterial Miniprep Kit from Zymoresearch (cat no. D6005), following the manufacturer's protocol. Long-read data was obtained using Oxford Nanopore Technology, with a MinION flowcell. For de novo assembly, Canu v2.2⁸⁹ was used with default parameters, only adjusting the genome size flag to 25 m. The resulting contigs were corrected with two rounds of Racon v1.5.0⁹⁰, one with the Nanopore reads and the other with publicly available Illumina reads^{91,92} (SRR16988715). Afterwards, several rounds of Pilon v1.24⁹³ were performed using Illumina reads until no changes were seen on the change file. To further increase the contiguity of the assembly, LINKS v1.8.7⁹⁴ was implemented.

***M. magnusii* and *M. tetraspermus* extraction and assembly**

High molecular weight DNA from *M. magnusii* and *M. tetraspermus* was extracted using the Zymo Quick-DNA HMW MagBead Kit (cat no. D6060). The manufacturer's protocol was followed, but lysis was optimized for non-conventional yeast species. The yeast cells were pelleted by centrifugation at 5000 x *g* and resuspended in 1 ml of 1 M sorbitol with 50 mM dithiothreitol (DTT) and incubated at 30°C for 10 min. The cells were then washed in fresh 1 M sorbitol and resuspended in 200 ul 1 M sorbitol with 5 U of Zymolyase Ultra (Zymo Research, cat no. E1007-2) and incubated at 37°C for 2 hrs. Once cells were spheroplasted, 205 ul phosphate buffered saline, 20 ul of 10% SDS, and 10 ul proteinase K were added, and the spheroplasts were lysed at 55°C for 10 min with occasional inversion. The DNA in the lysate was then bound to the beads following the manufacturer's protocol. After extraction, the DNA was enriched for high molecular weight DNA using a bead cleanup with a custom buffer (10 mM Tris-HCl, 1 mM EDTA pH 8, 1.6 M NaCl, 11% PEG) as previously described⁹⁵. Genome sequencing was performed by Plasmidsaurus using Oxford Nanopore Technology. Both genomes were assembled with flye v2.9.6⁹⁶ using the `nano-hq` option.

Gene annotation

To infer gene boundaries for the newly generated genomes of *D. fermentans*, *M. magnussi*, and *M. tetraspermus* we used funannotate v1.8.16⁹⁷. To do so, each genome was first masked using tantan v40⁹⁸ using the funannotate mask function. Next, each genome was annotated using the funannotate predict function, which is a wrapper function to use multiple gene calling algorithms and creates a consensus set of gene boundaries. Prediction algorithms implemented include Augustus v3.3.2⁹⁹, SNAP v2006-07-28¹⁰⁰, and GlimmerHMM¹⁰¹ each algorithm was trained on gene models predicted using BUSCO v2.0¹⁰² gene models from the OrthoDB v9¹⁰³ database of

near-universally single-copy orthologs from fungi. For Augustus, the `optimize\_augustus` argument was used. Additional gene boundaries were predicted by mapping gene annotations from a clustered set of proteins from 332 Saccharomycotina proteomes⁹¹; clustering was done using CD-HIT v4.8.1¹⁰⁴ using default settings. The results from each gene prediction algorithm were used to create a consensus set of gene boundaries using EVidenceModeler v1.1.1¹⁰⁵ with the “repeats2evm” argument. All approaches were given the same weight, except high-quality gene annotations (defined as >90% exon evidence) predicted by Augustus, which were given twice the weight of other algorithms. Gene models less than 50 amino acids in length and putatively transposable elements were subsequently removed. Putative transposable elements were identified using sequence similarity searches conducted using DIAMOND v2.1.8¹⁰⁶ and the funannotate database of repeat sequences. The resulting gene models were functionally with both KEGG v114.0⁵⁹ and InterPro v106.0⁵⁸ databases. KEGG orthologs were identified through the KofamKOALA¹⁰⁷ web server. InterPro Gene Ontology annotations were assigned using InterProScan v5.74_105.0¹⁰⁸. Each gene model was annotated locally using the `disable-precalt` option.

Gene tree inference

Three comparative genomic datasets were used by this study. The first was based on the recently published Y1000+ Project dataset⁹². The full 1,154 genomes of this dataset proved computationally intractable and was subsampled down to 400 using the following procedure: a genome with the shortest terminal branch in the species tree was pruned at random, unless that genome had ≤ 50 contigs. This process was repeated iteratively until 400 genomes remained in the tree. This sampling strategy maximized phylogenetic breadth and depth while retaining

highly-contiguous genomes that could be used for synteny analysis. Gene trees were inferred with OrthoFinder v3.0.1b1²¹ using the Y1000+ species tree as a reference. OrthoFinder identified large spikes in duplication rate within the class Dipodascales, warranting further investigation. Therefore, a second dataset was assembled using all 184 available Dipodascales genomes, as well as a third dataset of 135 Saccharomycetales genomes where WGD1 was known to occur. OrthoFinder was run on both of these datasets as before. A full record of all genomes used in this study can be found at Data S1.

Quantification and statistical analysis

Gene/species tree reconciliation analysis was performed using Orthofinder's hybrid species overlap/duplication-loss-coalescent algorithm. This approach has been shown to maximize accuracy while remaining tractable for large comparative analyses²¹. As validation, we also performed a fully probabilistic analysis using AleRax v1.2.0¹⁰⁹ on a small subset of 30 genomes. Despite recovering far fewer duplication events due to the restricted sample size, relative peaks in duplication rate were largely congruent between the two approaches (Figure S3).

Synteny analysis was performed using the wgd v2 software package¹¹⁰. First, paralogous gene sets were identified within each of 83 Saccharomycotina genome assemblies with ≤ 50 contigs, in addition to the 2 new genomes sequenced by this study, using the ``wg dmd`` command with default parameters. Next, colinear segments (multiplicons) were called with the ``wg d syn`` command, also under default parameters. ``wg d syn`` is a wrapper function for i-ADHoRe v3.0, which detects statistically significant colinear regions from gene homology matrices constructed between each pair of contigs to identify and assemble multiplicons¹¹¹.

Enrichment analysis was performed using the clusterProfiler v4.10¹¹² package in R v4.3.2¹¹³. The function `enrichKEGG` was used for KEGG pathway enrichment (Data S5), while the function `enricher` was used for gene ontology enrichment (Data S4). False discovery rate¹¹⁴ was used to control for multiple testing in both cases. Target genes were ohnologs that occurred in duplicate across colinear segments within genomes from species predicted to have experienced a given WGD event, against a background of all genes in the genome(s). We acknowledge this working definition of ohnologs is conservative, as not all ohnologs are expected to remain in order over long timescales. As a result, our analysis likely underestimates the total number of ohnologs in each genome. For example, this approach recovers only 882 ohnologs in *S. cerevisiae*, which is thought to have at least 1,102¹¹⁵. As *M. magnusii* experienced two rounds of WGD, target genes in this species were further filtered to those estimated to have duplicated prior to the *M. magnusii* + *Sap. Suaveolens* split (for WGD3) and those duplicates specific to *M. magnusii* (for WGD4).

To test whether niche breadth or growth curves were significantly different in post-WGD species, we performed phylogenetic ANOVA using the `aov.phylo` function as implemented in the 21eiger v2.0.11¹¹⁶ package. Niche breadth data was taken from David et al. 2025⁸⁰, and growth curves were obtained from Opulente et al. 2024⁹².

Supplementary Information

Data S1. Details of all genome assemblies used in this study, related to STAR Methods.

Data S2. Pairwise comparison of orthogroup membership between multiplicons in *D. fermentans*, *G. candidum*, *Sap. Suaveolens*, *M. magnusii* genomes, related to Figure 2B.

Data S3. BLAST results between all colinear paralogs in post-WGD *Dipodascales* genomes, related to Figure 2B.

Data S4. Gene Ontology enrichment analysis results of each WGD, sorted by event. Related to STAR Methods. GeneRatio: the proportion of ohnologs with this GO annotation. BgRatio: the total number of genes associated with this GO annotation. P.adjust: p-value adjusted for false discovery rate.

Data S5. KEGG pathway enrichment analysis results of each WGD, sorted by event. Related to STAR Methods. GeneRatio: the proportion of ohnologs in this pathway. BgRatio: the total number of genes in this pathway. p.adjust: p-value adjusted for false discovery rate.

References

1. Otto, S.P. (2007). The evolutionary consequences of polyploidy. *Cell* 131, 452–462.
2. Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics* 18, 411.
3. Dehal, P., and Boore, J.L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS biology* 3, e314.
4. Ohno, S. (1970). *Evolution by gene duplication* (Springer Science & Business Media).
5. Yu, D., Ren, Y., Uesaka, M., Beavan, A.J.S., Muffato, M., Shen, J., Li, Y., Sato, I., Wan, W., Clark, J.W., et al. (2024). Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences. *Nat Ecol Evol* 8, 519–535.
<https://doi.org/10.1038/s41559-023-02299-z>.
6. Brunet, F.G., Crollius, H.R., Paris, M., Aury, J.-M., Gibert, P., Jaillon, O., Laudet, V., and Robinson-Rechavi, M. (2006). Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Molecular biology and evolution* 23, 1808–1816.
7. Glasauer, S.M., and Neuhauss, S.C. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular genetics and genomics* 289, 1045–1060.
8. Bodt, S.D., Maere, S., and Peer, Y.V. de (2005). Genome duplication and the origin of angiosperms. *Trends in Ecology & Evolution* 20, 591–597.
<https://doi.org/10.1016/j.tree.2005.07.008>.
9. Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B., and Harmon, L.J. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist* 207, 454–467.
10. Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Glöckner, G., Platts, A.E., Heckel, D.G., Der, J.P., Wafula, E.K., Tang, M., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences* 112, 8362–8366. <https://doi.org/10.1073/pnas.1503926112>.
11. Gregory, T.R., and Mable, B.K. (2005). Polyploidy in animals. In *The evolution of the genome* (Elsevier), pp. 427–517.
12. Otto, S.P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annual review of genetics* 34, 401–437.
13. Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., Ma, H., and Qi, J. (2018). Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms. *Molecular Plant* 11, 414–428.
<https://doi.org/10.1016/j.molp.2018.01.002>.

- 484 14. Clark, J.W., and Donoghue, P.C.J. (2018). Whole-Genome Duplication and Plant
485 Macroevolution. *Trends in Plant Science* 23, 933–945.
486 <https://doi.org/10.1016/j.tplants.2018.07.006>.
- 487 15. Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C., and Soltis, P.S.
488 (2018). Impact of whole-genome duplication events on diversification rates in angiosperms.
489 *American Journal of Botany* 105, 348–363. <https://doi.org/10.1002/ajb2.1060>.
- 490 16. Campbell, M.A., Ganley, A.R.D., Gabaldón, T., and Cox, M.P. (2016). The Case of the
491 Missing Ancient Fungal Polyploids. *The American Naturalist* 188, 602–614.
492 <https://doi.org/10.1086/688763>.
- 493 17. Naranjo-Ortiz, M.A., and Gabaldón, T. (2020). Fungal evolution: cellular, genomic and
494 metabolic complexity. *Biological Reviews* 95, 1198–1232.
495 <https://doi.org/10.1111/brv.12605>.
- 496 18. Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the
497 entire yeast genome. *Nature* 387, 708–713.
- 498 19. Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M., and Wolfe, K.H.
499 (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species
500 descended from a whole-genome duplication. *Proceedings of the National Academy of*
501 *Sciences* 104, 8397–8402. <https://doi.org/10.1073/pnas.0608218104>.
- 502 20. Kohanovski, I., Pontz, M., Vande Zande, P., Selmecki, A., Dahan, O., Pilpel, Y., Yona,
503 A.H., and Ram, Y. (2024). Aneuploidy can be an evolutionary diversion on the path to
504 adaptation. *Molecular biology and evolution* 41, msae052.
- 505 21. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for
506 comparative genomics. *Genome biology* 20, 1–14.
- 507 22. Yang, Y., Moore, M.J., Brockington, S.F., Mikenas, J., Olivieri, J., Walker, J.F., and Smith,
508 S.A. (2018). Improved transcriptome sampling pinpoints 26 ancient and more recent
509 polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytologist*
510 217, 855–870. <https://doi.org/10.1111/nph.14812>.
- 511 23. Li, Z., Baniaga, A.E., Sessa, E.B., Scascitelli, M., Graham, S.W., Rieseberg, L.H., and
512 Barker, M.S. (2015). Early genome duplications in conifers and other seed plants. *Science*
513 *advances* 1, e1501084.
- 514 24. Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E.,
515 Tomsho, L.P., Hu, Y., Liang, H., and Soltis, P.S. (2011). Ancestral polyploidy in seed plants
516 and angiosperms. *Nature* 473, 97–100.
- 517 25. Huerta-Cepas, J., Dopazo, H., Dopazo, J., and Gabaldón, T. (2007). The human phylome.
518 *Genome biology* 8, R109.

- 519 26. Marcet-Houben, M., and Gabaldón, T. (2015). Beyond the Whole-Genome Duplication:
520 Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast
521 Lineage. *PLOS Biology* 13, e1002220. <https://doi.org/10.1371/journal.pbio.1002220>.
- 522 27. Zwaenepoel, A., Li, Z., Lohaus, R., and Peer, Y.V. de (2019). Finding Evidence for Whole
523 Genome Duplications: A Reappraisal. *Molecular Plant* 12, 133–136.
524 <https://doi.org/10.1016/j.molp.2018.12.019>.
- 525 28. Rasmussen, M.D., and Kellis, M. (2012). Unified modeling of gene duplication, loss, and
526 coalescence using a locus tree. *Genome research* 22, 755–765.
- 527 29. De Peer, and Van, Y. (2004). Computational approaches to unveiling ancient genome
528 duplications. *Nat Rev Genet* 5, 752–763. <https://doi.org/10.1038/nrg1449>.
- 529 30. Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny
530 and Collinearity in Plant Genomes. *Science* 320, 486–488.
531 <https://doi.org/10.1126/science.1153917>.
- 532 31. Glick, L., and Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number
533 evolution and the inference of polyploidy along a phylogeny. *Molecular biology and*
534 *evolution* 31, 1914–1922.
- 535 32. Deacon, J.W. (2005). *Fungal biology* (John Wiley & Sons).
- 536 33. Lue, N.F. (2010). Plasticity of telomere maintenance mechanisms in yeast. *Trends in*
537 *Biochemical Sciences* 35, 8–17. <https://doi.org/10.1016/j.tibs.2009.08.006>.
- 538 34. Peska, V., Fajkus, P., Bubeník, M., Brázda, V., Bohálová, N., Dvořáček, V., Fajkus, J., and
539 Garcia, S. (2021). Extraordinary diversity of telomeres, telomerase RNAs and their template
540 regions in Saccharomycetaceae. *Sci Rep* 11, 12784. [https://doi.org/10.1038/s41598-021-](https://doi.org/10.1038/s41598-021-92126-x)
541 [92126-x](https://doi.org/10.1038/s41598-021-92126-x).
- 542 35. Helsen, J., Ramachandran, K., Sherlock, G., and Dey, G. (2025). Centromeres evolve
543 progressively through selection at the kinetochore interface. Preprint at bioRxiv,
544 <https://doi.org/10.1101/2025.01.16.633479> <https://doi.org/10.1101/2025.01.16.633479>.
- 545 36. Haase, M.A.B., Lazar-Stefanita, L., Baudry, L., Wudzinska, A., Zhou, X., Rokas, A.,
546 Hittinger, C.T., Musacchio, A., and Boeke, J.D. (2025). Ancient co-option of LTR
547 retrotransposons as yeast centromeres. Preprint at bioRxiv,
548 <https://doi.org/10.1101/2025.04.25.647736> <https://doi.org/10.1101/2025.04.25.647736>.
- 549 37. Brejová, B., Lichancová, H., Brázdovič, F., Hegedúsová, E., Forgáčová Jakúbková, M.,
550 Hodorová, V., Džugasová, V., Baláž, A., Zeiselová, L., Cillingová, A., et al. (2019). Genome
551 sequence of the opportunistic human pathogen *Magnusiomyces capitatus*. *Curr Genet* 65,
552 539–560. <https://doi.org/10.1007/s00294-018-0904-y>.
- 553 38. Brejová, B., Lichancová, H., Hodorová, V., Neboháčová, M., Tomáška, L., Vinař, T., and
554 Nosek, J. (2019). Genome Sequence of an Arthroconidial Yeast, *Saprochaete fungicola* CBS

- 555 625.85. Microbiology Resource Announcements 8, 10.1128/mra.00092-19.
556 <https://doi.org/10.1128/mra.00092-19>.
- 557 39. Hodorová, V., Lichancová, H., Zubenko, S., Sienkiewicz, K., Penir, S.M.U., Afanasyev, P.,
558 Bocek, D., Bonnin, S., Hakobyan, S., Smyczynska, U., et al. (2019). Genome Sequence of
559 the Yeast *Saprochaete ingens* CBS 517.90. Microbiology Resource Announcements 8,
560 10.1128/mra.01366-19. <https://doi.org/10.1128/mra.01366-19>.
- 561 40. Lichancová, H., Hodorová, V., Sienkiewicz, K., Penir, S.M.U., Afanasyev, P., Bocek, D.,
562 Bonnin, S., Hakobyan, S., Krawczyk, P.S., Smyczynska, U., et al. (2019). Genome Sequence
563 of Flavor-Producing Yeast *Saprochaete suaveolens* NRRL Y-17571. Microbiol Resour
564 Announc 8, e00094-19. <https://doi.org/10.1128/MRA.00094-19>.
- 565 41. Gente, S., Desmasures, N., Jacopin, C., Plessis, G., Beliard, M., Panoff, J.-M., and Guéguen,
566 M. (2002). Intra-species chromosome-length polymorphism in *Geotrichum candidum*
567 revealed by pulsed field gel electrophoresis. International journal of food microbiology 76,
568 127–134.
- 569 42. Naumova, E.S., Smith, M.Th., Boekhout, T., de Hoog, G.S., and Naumov, G.I. (2001).
570 Molecular differentiation of sibling species in the *Galactomyces geotrichum* complex.
571 *Antonie Van Leeuwenhoek* 80, 263–273. <https://doi.org/10.1023/A:1013038610122>.
- 572 43. Filipp, D., Filipp, P., Nosek, J., and Hladk♦, M. (1995). Electrophoretic karyotype of
573 *Dipodascus (Endomyces) magnusii*: two main intraspecific chromosomal polymorphisms
574 associated with the difference in total genome size. *Curr Genet* 29, 81–87.
575 <https://doi.org/10.1007/BF00313197>.
- 576 44. Steenwyk, J.L., Lind, A.L., Ries, L.N.A., Reis, T.F. dos, Silva, L.P., Almeida, F., Bastos,
577 R.W., Silva, T.F. de C.F. da, Bonato, V.L.D., Pessoni, A.M., et al. (2020). Pathogenic
578 Allodiploid Hybrids of *Aspergillus Fungi*. *Current Biology* 30, 2495-2507.e7.
579 <https://doi.org/10.1016/j.cub.2020.04.071>.
- 580 45. Barros, K.O., Al-Oboudi, J., Freitas, L.F.D., Sousa, F.M.P., Batista, T.M., Santos, A.R.O.,
581 Morais, P.B., Sampaio, J.P., Lachance, M.-A., Hittinger, C.T., et al. (2025). Taxogenomic
582 analysis of *Pichia seneci* sp. nov. and new insights into hybridization events in the *Pichia*
583 *cactophila* species complex. *FEMS Yeast Res* 25, foaf037.
584 <https://doi.org/10.1093/femsyr/foaf037>.
- 585 46. Gordon, J.L., and Wolfe, K.H. (2008). Recent allopolyploid origin of *Zygosaccharomyces*
586 *rouxii* strain ATCC 42981. *Yeast* 25, 449–456. <https://doi.org/10.1002/yea.1598>.
- 587 47. Mira, N.P., Münsterkötter, M., Dias-Valada, F., Santos, J., Palma, M., Roque, F.C.,
588 Guerreiro, J.F., Rodrigues, F., Sousa, M.J., Leão, C., et al. (2014). The Genome Sequence of
589 the Highly Acetic Acid-Tolerant *Zygosaccharomyces bailii*-Derived Interspecies Hybrid
590 Strain ISA1307, Isolated From a Sparkling Wine Plant. *DNA Res* 21, 299–313.
591 <https://doi.org/10.1093/dnares/dst058>.

48. Pryszcz, L.P., Németh, T., Gácsér, A., and Gabaldón, T. (2014). Genome Comparison of *Candida orthopsilosis* Clinical Strains Reveals the Existence of Hybrids between Two Distinct Subspecies. *Genome Biol Evol* 6, 1069–1078. <https://doi.org/10.1093/gbe/evu082>.
49. Curtin, C.D., Borneman, A.R., Chambers, P.J., and Pretorius, I.S. (2012). De-Novo Assembly and Analysis of the Heterozygous Triploid Genome of the Wine Spoilage Yeast *Dekkera bruxellensis* AWRI1499. *PLOS ONE* 7, e33840. <https://doi.org/10.1371/journal.pone.0033840>.
50. Langdon, Q.K., Peris, D., Baker, E.P., Opulente, D.A., Nguyen, H.-V., Bond, U., Gonçalves, P., Sampaio, J.P., Libkind, D., and Hittinger, C.T. (2019). Fermentation innovation through complex hybridization of wild and domesticated yeasts. *Nat Ecol Evol* 3, 1576–1586. <https://doi.org/10.1038/s41559-019-0998-8>.
51. Clark, J.W., Hetherington, A.J., Morris, J.L., Pressel, S., Duckett, J.G., Puttick, M.N., Schneider, H., Kenrick, P., Wellman, C.H., and Donoghue, P.C.J. (2023). Evolution of phenotypic disparity in the plant kingdom. *Nat. Plants* 9, 1618–1626. <https://doi.org/10.1038/s41477-023-01513-x>.
52. Thomson, J.M., Gaucher, E.A., Burgan, M.F., De Kee, D.W., Li, T., Aris, J.P., and Benner, S.A. (2005). Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet* 37, 630–635. <https://doi.org/10.1038/ng1553>.
53. Lin, Z., and Li, W.-H. (2011). Expansion of Hexose Transporter Genes Was Associated with the Evolution of Aerobic Fermentation in Yeasts. *Molecular Biology and Evolution* 28, 131–142. <https://doi.org/10.1093/molbev/msq184>.
54. Hagman, A., Säll, T., Compagno, C., and Piskur, J. (2013). Yeast “Make-Accumulate-Consume” Life Strategy Evolved as a Multi-Step Process That Predates the Whole Genome Duplication. *PLOS ONE* 8, e68734. <https://doi.org/10.1371/journal.pone.0068734>.
55. Horianopoulos, L.C., Rokas, A., and Hittinger, C.T. (2025). Convergent evolution of aerobic fermentation through divergent mechanisms acting on key shared glycolytic genes. Preprint at bioRxiv, <https://doi.org/10.1101/2025.10.02.679963> <https://doi.org/10.1101/2025.10.02.679963>.
56. Uyeda, J.C., Zenil-Ferguson, R., and Pennell, M.W. (2018). Rethinking phylogenetic comparative methods. *Systematic Biology* 67, 1091–1109. <https://doi.org/10.1093/sysbio/syy031>.
57. Maddison, W.P., and FitzJohn, R.G. (2015). The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic Biology* 64, 127–136. <https://doi.org/10.1093/sysbio/syu070>.
58. Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research* 49, D344–D354. <https://doi.org/10.1093/nar/gkaa977>.

59. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28, 27–30.
60. Mdaini, N., Gargouri, M., Hammami, M., Monser, L., and Hamdi, M. (2006). Production of natural fruity aroma by *Geotrichum candidum*. *Appl Biochem Biotechnol* 128, 227–235. <https://doi.org/10.1385/ABAB:128:3:227>.
61. Boutrou, R., and Guéguen, M. (2005). Interests in *Geotrichum candidum* for cheese technology. *International Journal of Food Microbiology* 102, 1–20. <https://doi.org/10.1016/j.ijfoodmicro.2004.12.028>.
62. Barros, M.C., Quadros, J.S., Silva, M.A.P. da, and Nunes, R.S.C. (2022). Taxonomy and Characterization of Fungi Isolated from Cocoa Beans during Fermentation in Saf's Agroforestry System in the Amazon. *Open Access Library Journal* 9, 1–10. <https://doi.org/10.4236/oalib.1108825>.
63. Koné, M.K., Guéhi, S.T., Durand, N., Ban-Koffi, L., Berthiot, L., Tachon, A.F., Brou, K., Boulanger, R., and Montet, D. (2016). Contribution of predominant yeasts to the occurrence of aroma compounds during cocoa bean fermentation. *Food Research International* 89, 910–917. <https://doi.org/10.1016/j.foodres.2016.04.010>.
64. Grondin, E., Shum Cheong Sing, A., James, S., Nueno-Palop, C., François, J.M., and Petit, T. (2017). Flavour production by *Saprochaete* and *Geotrichum* yeasts and their close relatives. *Food Chemistry* 237, 677–684. <https://doi.org/10.1016/j.foodchem.2017.06.009>.
65. Conant, G.C., and Wolfe, K.H. (2007). Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Molecular Systems Biology* 3. <https://doi.org/10.1038/msb4100170>.
66. Peter Smits, H., Hauf, J., Müller, S., Hobley, T.J., Zimmermann, F.K., Hahn-Hägerdal, B., Nielsen, J., and Olsson, L. (2000). Simultaneous overexpression of enzymes of the lower part of glycolysis can enhance the fermentative capacity of *Saccharomyces cerevisiae*. *Yeast* 16, 1325–1334. [https://doi.org/10.1002/1097-0061\(200010\)16:14<1325::AID-YEA627>3.0.CO;2-E](https://doi.org/10.1002/1097-0061(200010)16:14<1325::AID-YEA627>3.0.CO;2-E).
67. Saito, H., and Posas, F. (2012). Response to Hyperosmotic Stress. *Genetics* 192, 289–318. <https://doi.org/10.1534/genetics.112.140863>.
68. Levin, D.E. (2011). Regulation of cell wall biogenesis in *Saccharomyces cerevisiae*: the cell wall integrity signaling pathway. *Genetics* 189, 1145–1175.
69. Gustin, M.C., Albertyn, J., Alexander, M., and Davenport, K. (1998). MAP Kinase Pathways in the Yeast *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* 62, 1264–1300. <https://doi.org/10.1128/mmbr.62.4.1264-1300.1998>.
70. Barbieri, M., Bonafè, M., Franceschi, C., and Paolisso, G. (2003). Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. *Am J Physiol Endocrinol Metab* 285, E1064–1071. <https://doi.org/10.1152/ajpendo.00296.2003>.

- 667 71. Zaman, S., Lippman, S.I., Schneper, L., Slonim, N., and Broach, J.R. (2009). Glucose
668 regulates transcription in yeast through a network of signaling pathways. *Mol Syst Biol* 5,
669 245. <https://doi.org/10.1038/msb.2009.2>.
- 670 72. Thevelein, J.M., and De Winde, J.H. (1999). Novel sensing mechanisms and targets for the
671 cAMP–protein kinase A pathway in the yeast *Saccharomyces cerevisiae*. *Molecular*
672 *Microbiology* 33, 904–918. <https://doi.org/10.1046/j.1365-2958.1999.01538.x>.
- 673 73. Braicu, C., Buse, M., Busuioc, C., Drula, R., Gulei, D., Raduly, L., Rusu, A., Irimie, A.,
674 Atanasov, A.G., Slaby, O., et al. (2019). A Comprehensive Review on MAPK: A Promising
675 Therapeutic Target in Cancer. *Cancers* 11, 1618. <https://doi.org/10.3390/cancers11101618>.
- 676 74. Hakes, L., Pinney, J.W., Lovell, S.C., Oliver, S.G., and Robertson, D.L. (2007). All
677 duplicates are not equal: the difference between small-scale and genome duplication.
678 *Genome Biol* 8, R209. <https://doi.org/10.1186/gb-2007-8-10-r209>.
- 679 75. Freeling, M., and Thomas, B.C. (2006). Gene-balanced duplications, like tetraploidy, provide
680 predictable drive to increase morphological complexity. *Genome Res.* 16, 805–814.
681 <https://doi.org/10.1101/gr.3681406>.
- 682 76. Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S., and Van de Peer, Y.
683 (2006). The gain and loss of genes during 600 million years of vertebrate evolution. *Genome*
684 *Biol* 7, R43. <https://doi.org/10.1186/gb-2006-7-5-r43>.
- 685 77. Rice, A.M., and McLysaght, A. (2017). Dosage-sensitive genes in evolution and disease.
686 *BMC Biol* 15, 78. <https://doi.org/10.1186/s12915-017-0418-y>.
- 687 78. Makino, T., and McLysaght, A. (2010). Ohnologs in the human genome are dosage balanced
688 and frequently associated with disease. *Proceedings of the National Academy of Sciences*
689 107, 9270–9274. <https://doi.org/10.1073/pnas.0914697107>.
- 690 79. Singh, P.P., Arora, J., and Isambert, H. (2015). Identification of ohnolog genes originating
691 from whole genome duplication in early vertebrates, based on synteny comparison across
692 multiple genomes. *PLoS computational biology* 11.
- 693 80. David, K.T., Schraiber, J.G., Crandall, J.G., Labella, A.L., Opulente, D.A., Harrison, M.-C.,
694 Wolters, J.F., Zhou, X., Shen, X.-X., Groenewald, M., et al. (2025). Convergent expansions
695 of keystone gene families drive metabolic innovation in *Saccharomycotina* yeasts.
696 *Proceedings of the National Academy of Sciences* 122, e2500165122.
697 <https://doi.org/10.1073/pnas.2500165122>.
- 698 81. Scott, A.L., Richmond, P.A., Dowell, R.D., and Selmecki, A.M. (2017). The Influence of
699 Polyploidy on the Evolution of Yeast Grown in a Sub-Optimal Carbon Source. *Molecular*
700 *Biology and Evolution* 34, 2690–2703. <https://doi.org/10.1093/molbev/msx205>.
- 701 82. Zhu, Y.O., Sherlock, G., and Petrov, D.A. (2016). Whole Genome Analysis of 132 Clinical
702 *Saccharomyces cerevisiae* Strains Reveals Extensive Ploidy Variation. *G3*
703 *Genes|Genomes|Genetics* 6, 2421–2434. <https://doi.org/10.1534/g3.116.029397>.

704 83. Saada, O.A., Tsouris, A., Large, C., Friedrich, A., Dunham, M.J., and Schacherer, J. (2022).
705 Phased polyploid genomes provide deeper insight into the multiple origins of domesticated
706 *Saccharomyces cerevisiae* beer yeasts. *Current Biology* 32, 1350-1361.e3.
707 <https://doi.org/10.1016/j.cub.2022.01.068>.

708 84. Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., Meiri, S., Belmaker,
709 J., and Mayrose, I. (2019). The global biogeography of polyploid plants. *Nature Ecology &*
710 *Evolution* 3, 265–273.

711 85. David, K.T. (2022). Global gradients in the distribution of animal polyploids. *Proc. Natl.*
712 *Acad. Sci. U.S.A.* 119, e2214070119. <https://doi.org/10.1073/pnas.2214070119>.

713 86. Hahn, M.W. (2009). Distinguishing among evolutionary models for the maintenance of gene
714 duplicates. *Journal of Heredity* 100, 605–617.

715 87. David, K.T., Oaks, J.R., and Halanych, K.M. (2020). Patterns of gene evolution following
716 duplications and speciations in vertebrates. *PeerJ* 8, e8813.

717 88. Hittinger, C.T., and Carroll, S.B. (2007). Gene duplication and the adaptive evolution of a
718 classic genetic switch. *Nature* 449, 677–681. <https://doi.org/10.1038/nature06151>.

719 89. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017).
720 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat
721 separation. *Genome Res* 27, 722–736. <https://doi.org/10.1101/gr.215087.116>.

722 90. Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome
723 assembly from long uncorrected reads. *Genome Res* 27, 737–746.
724 <https://doi.org/10.1101/gr.214270.116>.

725 91. Shen, X.-X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase,
726 M.A.B., Wisecaver, J.H., Wang, M., Doering, D.T., et al. (2018). Tempo and Mode of
727 Genome Evolution in the Budding Yeast Subphylum. *Cell* 175, 1533-1545.e20.
728 <https://doi.org/10.1016/j.cell.2018.10.023>.

729 92. Opulente, D.A., LaBella, A.L., Harrison, M.-C., Wolters, J.F., Liu, C., Li, Y., Kominek, J.,
730 Steenwyk, J.L., Stoneman, H.R., VanDenAvond, J., et al. (2024). Genomic factors shape
731 carbon and nitrogen metabolic niche breadth across *Saccharomycotina* yeasts. *Science* 384,
732 eadj4503. <https://doi.org/10.1126/science.adj4503>.

733 93. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A.,
734 Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for
735 comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9,
736 e112963. <https://doi.org/10.1371/journal.pone.0112963>.

737 94. Warren, R.L., Yang, C., Vandervalk, B.P., Behsaz, B., Lagman, A., Jones, S.J.M., and Birol,
738 I. (2015). LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads.
739 *Gigascience* 4, 35. <https://doi.org/10.1186/s13742-015-0076-3>.

- 740 95. Rojas, J., Hose, J., Dutcher, H.A., Place, M., Wolters, J.F., Hittinger, C.T., and Gasch, A.P.
741 (2024). Comparative modeling reveals the molecular determinants of aneuploidy fitness cost
742 in a wild yeast model. *Cell Genomics* 4. <https://doi.org/10.1016/j.xgen.2024.100656>.
- 743 96. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone
744 reads using repeat graphs. *Nature biotechnology* 37, 540–546.
- 745 97. Palmer, J.M., and Stajich, J. (2020). Funannotate v1.8.1: Eukaryotic genome annotation.
746 (Zenodo). <https://doi.org/10.5281/zenodo.4054262> <https://doi.org/10.5281/zenodo.4054262>.
- 747 98. Frith, M.C. (2011). A new repeat-masking method enables specific detection of homologous
748 sequences. *Nucleic acids research* 39, e23–e23.
- 749 99. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006).
750 AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* 34,
751 W435–W439. <https://doi.org/10.1093/nar/gkl200>.
- 752 100. Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5.
753 <https://doi.org/10.1186/1471-2105-5-59>.
- 754 101. Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two
755 open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879.
756 <https://doi.org/10.1093/bioinformatics/bth315>.
- 757 102. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov,
758 G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO Applications from Quality
759 Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* 35, 543–548.
760 <https://doi.org/10.1093/molbev/msx319>.
- 761 103. Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simão, F.A.,
762 Ioannidis, P., Seppey, M., Loetscher, A., and Kriventseva, E.V. (2017). OrthoDB v9.1:
763 cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal,
764 bacterial and viral orthologs. *Nucleic Acids Research* 45, D744–D749.
765 <https://doi.org/10.1093/nar/gkw1119>.
- 766 104. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering
767 the next-generation sequencing data. *Bioinformatics* 28, 3150–3152.
- 768 105. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell,
769 C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using
770 EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biology* 9,
771 R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
- 772 106. Buchfink, B., Reuter, K., and Drost, H.-G. (2021). Sensitive protein alignments at tree-of-
773 life scale using DIAMOND. *Nature methods* 18, 366–368.

107. Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252.
108. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., and Nuka, G. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240.
109. Morel, B., Williams, T.A., Stamatakis, A., and Szöllősi, G.J. (2024). AleRax: a tool for gene and species tree co-estimation and reconciliation under a probabilistic model of gene duplication, transfer, and loss. *Bioinformatics* 40, btae162. <https://doi.org/10.1093/bioinformatics/btae162>.
110. Chen, H., Zwaenepoel, A., and Van de Peer, Y. (2024). wgd v2: a suite of tools to uncover and date ancient polyploidy and whole-genome duplication. *Bioinformatics* 40, btae272. <https://doi.org/10.1093/bioinformatics/btae272>.
111. Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., and Vandepoele, K. (2012). i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res* 40, e11. <https://doi.org/10.1093/nar/gkr955>.
112. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., and Zhan, L.I. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The innovation* 2.
113. Team, R.C. (2017). R: A Language and Environment for Statistical Computing.
114. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 289–300.
115. Byrne, K.P., and Wolfe, K.H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 15, 1456–1461. <https://doi.org/10.1101/gr.3672305>.
116. Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G., Alfaro, M.E., and Harmon, L.J. (2014). geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30, 2216–2218.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
<i>Dipodascus fermentans</i>	Portuguese Culture Collection	PYCC 3480
<i>Magnusiomyces magnusii</i>	ARS Culture Collection	NRRL Y-17563
<i>Magnusiomyces tetraspermus</i>	ARS Culture Collection	NRRL Y-7288
Chemicals, peptides, and recombinant proteins		
Zymolyase Ultra	Zymo Research	E1007-2
Critical commercial assays		
Quick-DNA HMW MagBead Kit	Zymo Research	D6060
Quick-DNA Fungal/Bacterial Miniprep Kit	Zymo Research	D6005
Deposited data		
Genomic and gene annotation data	Figshare	https://doi.org/10.6084/m9.figshare.29852876
Software and algorithms		
Canu v2.2	Koren et al. 2017 ⁸⁹	https://github.com/marbl/canu
Racon v1.5.0	Vaser et al. 2017 ⁹⁰	https://github.com/isovic/racon
Pilon v1.24	Walker et al. 2014 ⁹³	https://github.com/broadinstitute/pilon/wiki
LINKS v1.8.7	Warren et al. 2015 ⁹⁴	https://github.com/bcgsc/LINKS
flye v2.9.6	Kolmogorov et al. 2019 ⁹⁶	https://github.com/mikolmogorov/Flye
funannotate v1.8.16	Palmer & Stajich 2020 ⁹⁷	https://github.com/nextgenusfs/funannotate/tree/master
tantan v40	Frith 2011 ⁹⁸	https://gitlab.com/mcfrith/tantan
Augustus v3.3.2	Stanke et al. 2006 ⁹⁹	https://github.com/Gaius-Augustus/Augustus
SNAP v2006-07-28	Korf et al. 2004 ¹⁰⁰	https://hpc.nih.gov/apps/snap.html
GlimmerHMM	Majoros et al. 2004 ¹⁰¹	https://github.com/kbclin/glimmerhmm
BUSCO v2.0	Waterhouse et al. 2018 ¹⁰²	https://gitlab.com/ezlab/busco
OrthoDB v9	Zdobnov et al. 2017 ¹⁰³	https://www.ezlab.org/orthodb_v12_userguide.html
CD-HIT v4.8.1	Fu et al. 2012 ¹⁰⁴	https://github.com/weizhongli/cdhit
EvidenceModeler v1.1.1	Haas et al. 2008 ¹⁰⁵	https://github.com/EvidenceModeler/EvidenceModeler
DIAMOND v2.1.8	Buchfink et al. 2021 ¹⁰⁶	https://github.com/bbuchfink/diamond
KofamKOALA	Aramaki et al. 2020 ¹⁰⁷	https://www.genome.jp/tools/kofamkoala/

InterProScan v5.74_105.0	Jones et al. 2014 ¹⁰⁸	https://github.com/ebi-pf-team/interproscan
Orthofinder v3.0.1b1	Emms & Kelly 2019 ²¹	https://github.com/davidemms/OrthoFinder
AleRax v1.2.0	Morel et al. 2024 ¹⁰⁹	https://github.com/BenoitMorel/AleRax
wgd v2	Chen et al. 2024 ¹¹⁰	https://github.com/heche-psb/wgd
ADHoRe v3.0	Proost et al. 2012 ¹¹¹	https://github.com/VIB-PSB/i-ADHoRe
R v4.3.2	R Team 2017 ¹¹³	https://www.r-project.org/
clusterProfiler	Wu et al. 2021 ¹¹²	https://guangchuangyu.github.io/software/clusterProfiler/
geiger	Pennell et al. 2014 ¹¹⁶	https://cran.r-project.org/web/packages/geiger/index.html

Figure 1

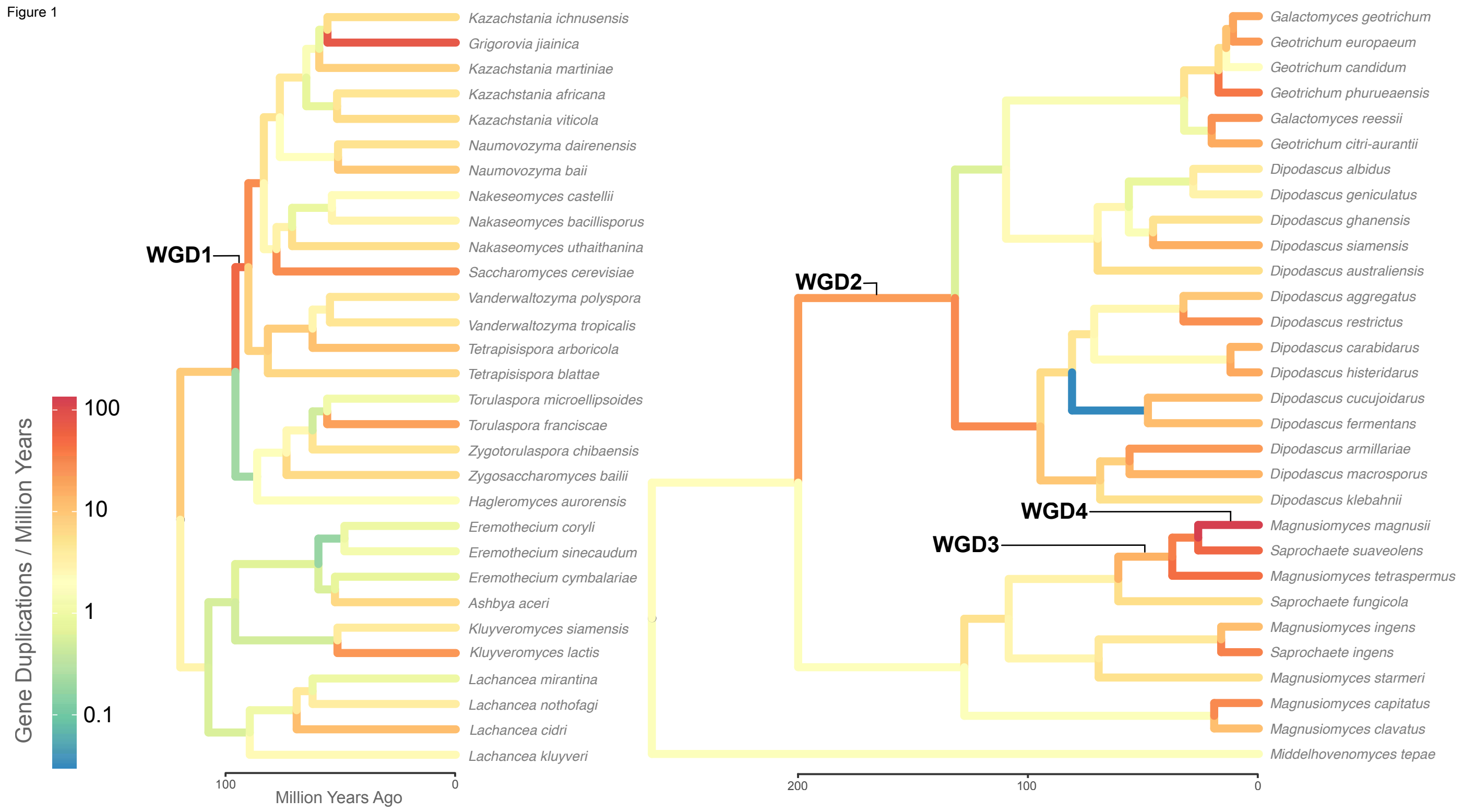
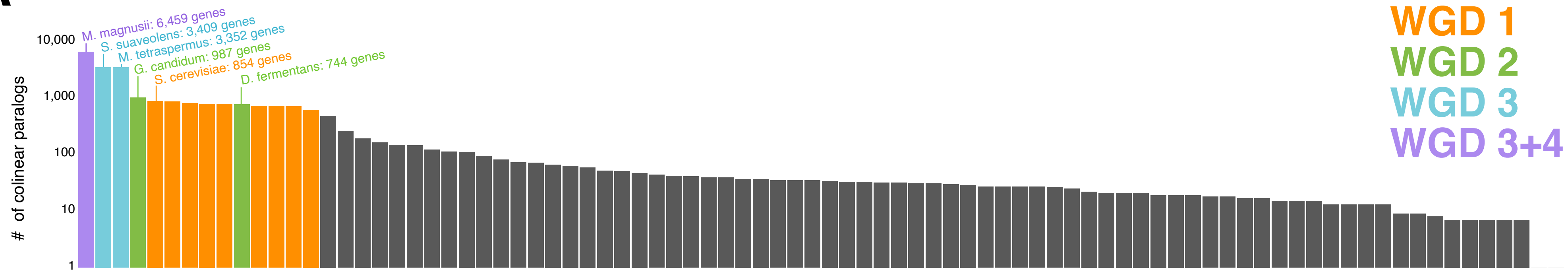
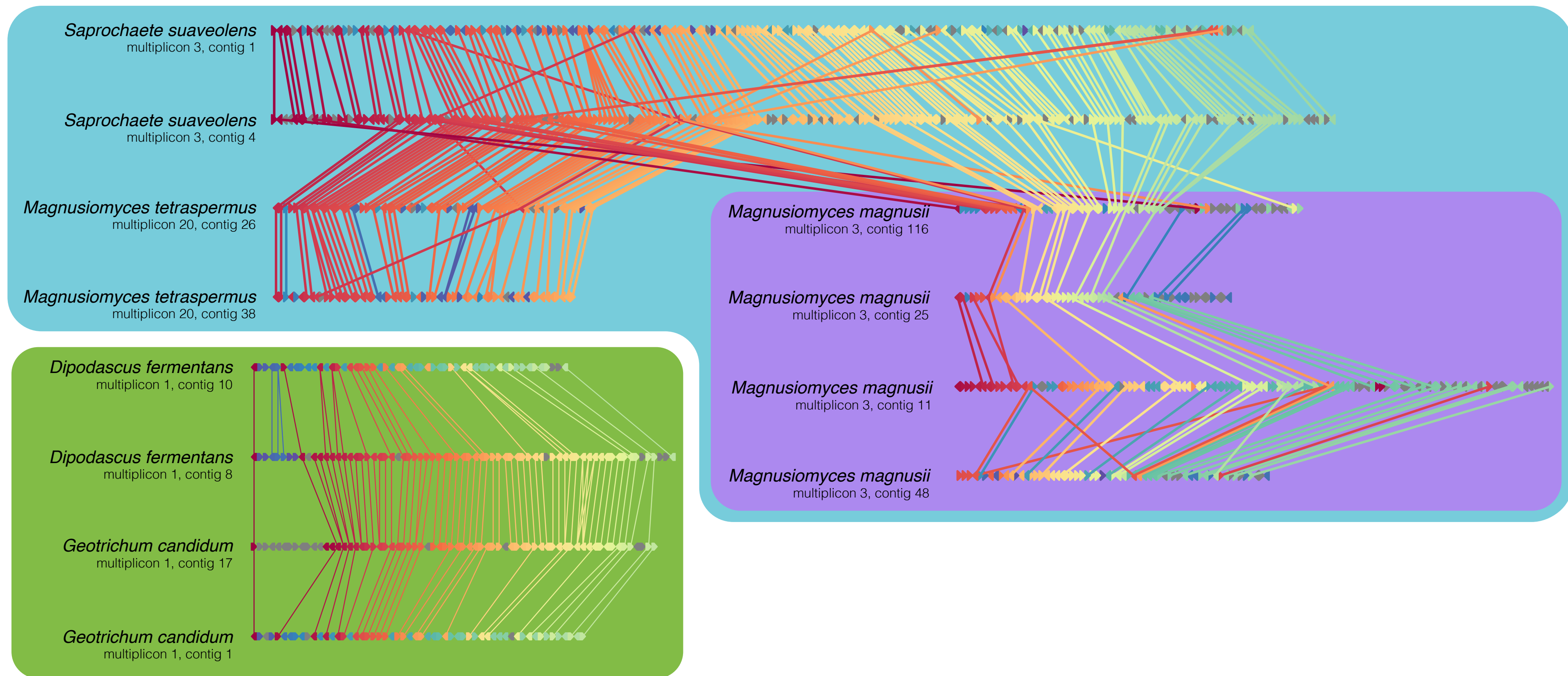


Figure 2

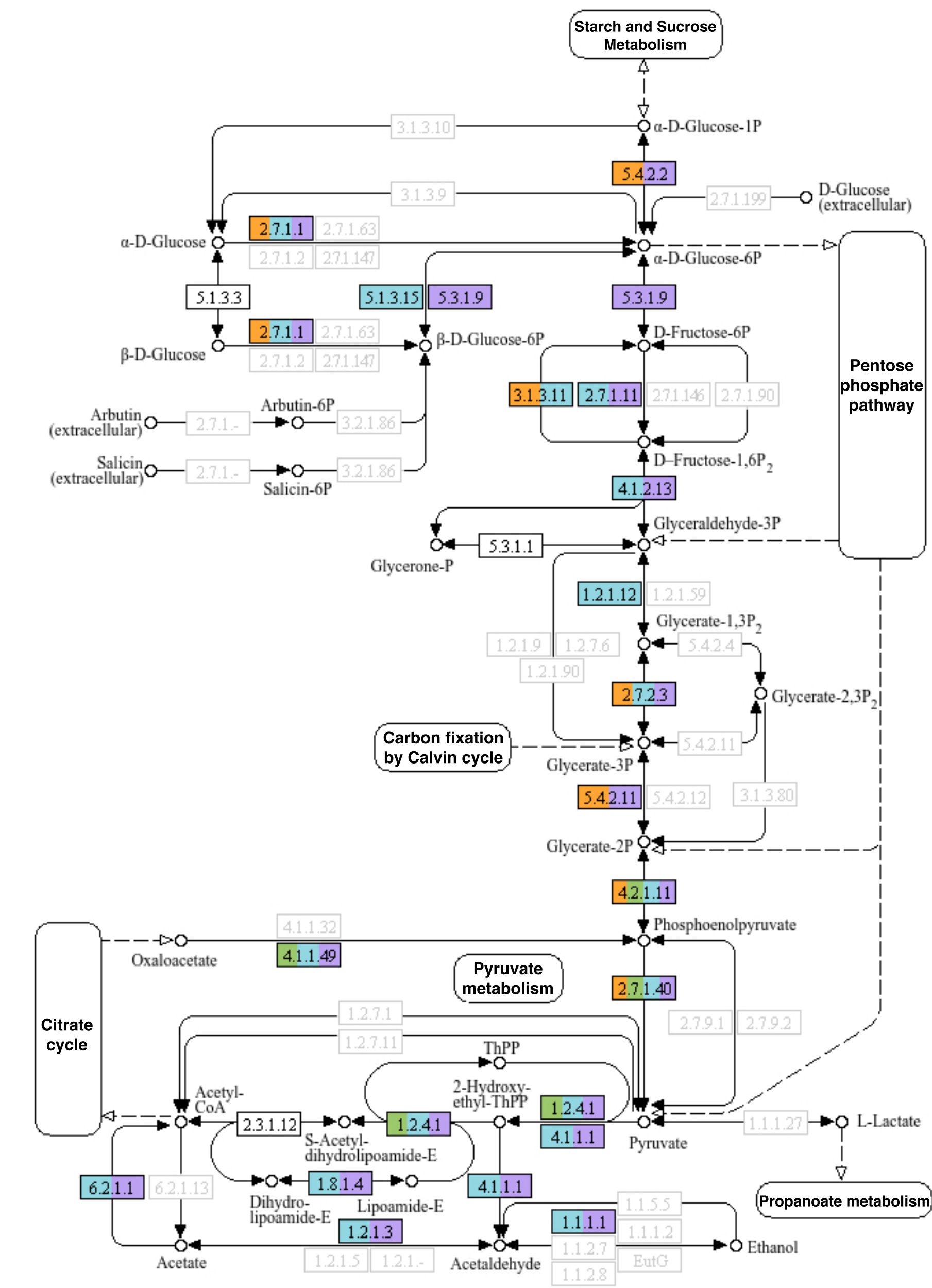
A



B



Glycolysis / Gluconeogenesis



WGD 1 WGD 3
WGD 2 WGD 4

MAPK Signaling Pathway - Yeast

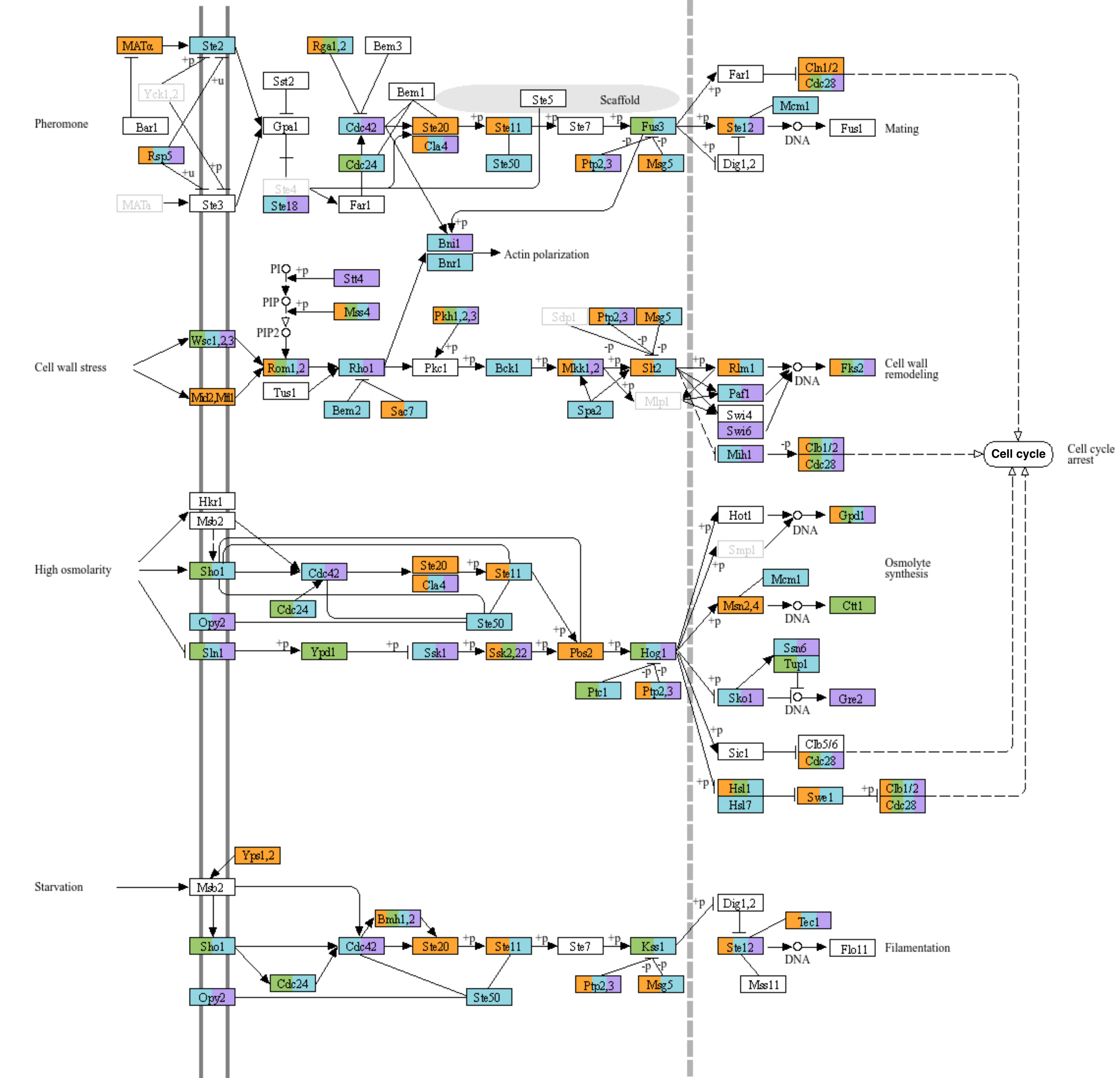


Table 1

Species	Chromosome Count	Method	Reference
<i>Magnusiomyces capitatus</i>	4	PFGE	Brejova et al 2019a ³⁷
<i>Magnusiomyces ingens</i>	4	PFGE	Brejova et al 2019a ³⁷
<i>Saprochaete fungicola</i>	5	long-read sequencing	Brejova et al 2019b ³⁸
<i>Saprochaete ingens</i>	5	long-read sequencing	Hodorova et al 2019 ³⁹
<i>Saprochaete suaveolens</i>	7	long-read sequencing	Lichancova et al 2019 ⁴⁰
<i>Geotrichum candidum</i>	8	PFGE	Gente et al 2002 ⁴¹
<i>Galactomyces reessii</i>	9	PFGE	Naumova et al 2001 ⁴²
<i>Magnusiomyces magnusii</i>	13	PAGE	Filipp et al 1995 ⁴³