Contrasting modes of macro- and micro-synteny evolution in a eukaryotic subphylum

Yuanning Li[1,*], Hongyue Liu[1], Jacob L. Steenwyk[2,3], Abigail L. LaBella[2,3], Marie-Claire Harrison[2,3], Marizeth Groenewald[4], Xiaofan Zhou[5], Xing-Xing Shen[6], Tao Zhao[7], Chris Todd Hittinger[8], and Antonis Rokas[2,3,9*]

[1] Institute of Marine Science and Technology, Shandong University, 72 Binhai Road, Qingdao 266237, China

[2] Department of Biological Sciences, Vanderbilt University, VU Station B#35-1634, Nashville, TN 37235, USA

[3] Vanderbilt Evolutionary Studies Initiative, Vanderbilt University, VU Station B#35-1634, Nashville, TN 37235, USA

[4] Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands

[5] Guangdong Laboratory for Lingnan Modern Agriculture, Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Centre, South China Agricultural University, 483 Wushan Road, Guangzhou 520643, China

[6] Key Laboratory of Biology of Crop Pathogens and Insects of Zhejiang Province, Institute of Insect Sciences, Zhejiang University, 866 Yuhangtang Rd, Hangzhou 310058, China

[7] State Key Laboratory of Crop Stress Biology for Arid Areas/Shaanxi Key Laboratory of Apple, College of Horticulture, Northwest A&F University, Taicheng Road 3, Yangling 712100, China

[8] Laboratory of Genetics, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, J.F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, 1552 University Avenue, University of Wisconsin-Madison, Madison, WI 53726-4084, USA

[9] Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118, Heidelberg, Germany

* Corresponding authors: yuanning.li@email.sdu.edu.cn and antonis.rokas@vanderbilt.edu

* Lead contact: antonis.rokas@vanderbilt.edu

Keywords: microsynteny, macrosynteny, budding yeasts, eukaryotes, *GAL* gene cluster, whole genome duplication

Twitter handles: @RokasLab; @YuanningLi1

Summary

Examination of the changes in order and arrangement of homologous genes is key for understanding the mechanisms of genome evolution in eukaryotes. Previous comparisons between eukaryotic genomes have revealed considerable conservation across species that diverged hundreds of millions of years ago (e.g. vertebrates[1–3], bilaterian animals[4,5], and filamentous fungi[6]). However, understanding how genome organization evolves within and between eukaryotic major lineages remains underexplored. We analyzed high-quality genomes of 120 representative budding yeast species (subphylum Saccharomycotina) spanning ~400 million years of eukaryotic evolution to examine how their genome organization evolved and to compare it to the evolution of animal and plant genome organization[7]. We found that the decay of both macrosynteny (the conservation of homologous chromosomes) and microsynteny (the conservation of local gene content and order) was strongly associated with evolutionary divergence across budding yeast major clades. However, whereas macrosynteny decayed very fast, within ~100 million years, the microsynteny of many genes – especially genes in metabolic clusters (e.g., in the *GAL* gene cluster[8]) – was much more deeply conserved both within major clades and across the subphylum. We further found that when genomes with similar evolutionary divergence times were compared, budding yeasts had lower macrosynteny conservation than animals and filamentous fungi but higher conservation than angiosperms. In contrast, budding yeasts had levels of microsynteny conservation on par with mammals, whereas angiosperms exhibited very low conservation. Our results provide new insight into the tempo and mode of the evolution of gene and genome organization across an entire eukaryotic subphylum.

Results and Discussion

Macrosynteny is conserved only in closely related budding yeast species

To examine the conservation of macrosynteny, we constructed Oxford dot plots comparing the chromosomal positions of homologous genes between the genomes of four representative species, *Saccharomyces cerevisiae* (Saccharomycetaceae clade), *Candida albicans* (CUG-Ser1 clade), *Ogataea parapolymorpha* (Pichiaceae clade), and *Yarrowia lipolytica* (Dipodascaceae/Trichomonascaceae clade) and all other 119 budding yeast species (Figures 1, S1, Data S1).

We found similar trends of decay of macrosynteny conservation in all four anchored species (Figures 1, S1). For example, the genome organization of both *S. cerevisiae* and *C. albicans* was nearly collinear when compared to their closest species relatives *Saccharomyces paradoxus* (16 chromosomes; macrosynteny conservation index (CI) = 0.99) and *Candida dubliniensis* (8 chromosomes; CI = 0.99), respectively, with most homologous genes lying on the diagonal of each chromosome (Figure 1). However, macrosynteny became less conserved as the evolutionary divergence between the species compared increased. For example, the Oxford dot plots between *C. albicans* and the more distantly related *Candida parapsilosis* (CI = 0.52) and *Spathaspora passalidarum* (CI = 0.72), two species estimated to have diverged ~60 and ~73 MYA, respectively, reveal multiple translocations and inversions and much more scrambled orders and locations of homologous genes (Figure 1A). This pattern suggests that homologous genes are still largely conserved within homologous chromosomes, but their gene order and location are diverging; this phenomenon has been previously observed in filamentous fungi and referred to as mesosynteny[6]. More strikingly, macrosynteny conservation appears to be almost completely lost once the evolutionary divergence of the budding yeast genomes compared reaches ~100 million years (CI < ~0.25) in all four anchored species (Figures 1, S1).

It is well known that macrosynteny can decay due to large-scale mutations that alter chromosome structure, such as chromosomal duplications and various types of rearrangements (e.g., inversions, translocations, etc.)[9]. However, our macrosynteny analysis suggests that budding yeast macrosynteny decays at a faster rate compared to other major eukaryotic lineages, such as bilaterians[2,5] and filamentous fungi[6], both lineages that also diverged more than 400 MYA. For example, we found a higher CI in filamentous fungi (using *Zymoseptoria tritici* as an anchor species) and bilaterian animals (using *Patinopecten yessoensis*, the scallop, as an anchor species) than in budding yeasts (Figure S2). For example, comparisons between scallop and amphioxus (*Branchiostoma lanceolatum*) genomes, which diverged more than 500 MYA, exhibited high macrosynteny conservation, with many large conserved chromosomal blocks (Figure S2A, CI = 0.65). These results contrast with the much lower levels of macrosynteny conservation observed between pairs to budding yeast species that diverged ~80–100 MYA (Figures 1, S1). The degree of macrosynteny conservation in filamentous fungi is also higher than that of budding yeasts. For example, the CI between *Z. tritici* and *Pseudocercospora fijiensis*, two species that diverged ~80 MYA[10], is 0.73 (Figure S2C), whereas that of *S. cerevisiae* and *Nakaseomyces castellii* is 0.26, even though the two species diverged around the same time (Figure 1). In contrast, the conservation of budding yeast macrosynteny is higher than that of angiosperm genomes (Figure S2B); for example, *Arabidopsis thaliana* and *Brassica rapa* diverged ~ 26 MYA but showed much lower macrosynteny conservation (CI = 0.39) that *C. albicans* and *C. parapsilosis* (~60 MYA divergence, CI = 0.52) (Figure 1A). The lower levels of macrosynteny conservation in angiosperms are probably due to multiple rounds of the large-scale gene or entire genome duplications[11].

The differences in the pace of macrosynteny decay might also be associated with differences in the generation time of organisms in these lineages. For example, the generation time of budding yeasts (e.g., 1.5 hours for *S. cerevisiae* and *C. albicans*[12]) is thought to be shorter than that of filamentous fungi in the subphylum Pezizomycotina (e.g., 2-3 hours for *Aspergillus nidulans*[13]). These results are also consistent with a recent study showing that the amino acid sequence substitution rate of budding yeast genomes is higher than that of filamentous fungi[14]. Interestingly, a previous study also suggested that the chromosome rearrangements (per Mb) are about 50-fold higher in budding yeasts than in vertebrate genomes[15]. Thus, the faster rate of macrosynteny decay of budding yeasts compared to filamentous fungi may be due to both their shorter generation times and higher mutation rates.

Conserved microsynteny within major clades and across the budding yeast subphylum

Previous results have suggested that macro- or microsynteny conservation is poor across fungal genomes, even between congeneric species[6]**.** To explore the evolution of microsynteny in major clades of budding yeasts, as well as across the entire subphylum, we examined the syntenic conservation of homologous genes across the genomes of 120 budding yeast species (Figure 2)**.** The entire microsynteny network is composed of all syntenic homologous genes, where genes are the nodes of the network, and the conservation of synteny between genes is the edges of the network. The budding yeast microsynteny network contains 566,379 nodes (genes) and 6,310,014 edges (instances of conservation of synteny between homologous genes). To identify homologous genes whose microsynteny has been conserved across or within budding yeasts, we decomposed the entire microsynteny network into 17,010 (number of nodes ≥ 3) nonoverlapping subnetworks (see Methods). These syntenic subnetworks varied with respect to the number of genes involved, from the minimum size of three genes to up to 743 genes (see Figshare Repository), reflecting the differences and dynamics of microsynteny conservation across gene families and yeast major clades. Subnetworks with larger gene sizes could correspond to genes that have undergone the whole genome and/or segmental duplication events7, tandem duplications, and/or genes that are highly conserved across the entire subphylum. For example, plasma membrane ABC transporters, ATPase, Rab family GTPase, Hsp70, and Hsp40 protein families were identified as the largest subnetworks in budding yeasts.

Even though macrosynteny is not conserved within major clades or across the budding yeast subphylum, we did identify 946 syntenic subnetworks (5.56% of all subnetworks) that were largely conserved (i.e., present in at least 80% of the genomes examined) across the budding yeast subphylum. The remaining ~95% of these 17, 010 subnetworks are mostly specific to individual major clades, indicating that a large proportion of yeast genomes are highly reshuffled in a lineage-specific manner, with many specific subnetworks for a particular major clade (e.g., Saccharomycetaceae, CUG-Ser1 clade, etc.) (Figure 2). Compared to the microsynteny networks of mammals and angiosperms, two lineages diverged much more recently than budding yeasts (~170 vs. ~400 MYA) (Figures S3A, S3B), we found that the overall pattern of conservation of microsynteny in budding yeasts is more similar to that of angiosperms (where ~8.7% of subnetworks are conserved across angiosperms) than to the mammal network (where ~66% of subnetworks are conserved across mammals) (Figure S3)[7].

To directly compare the rates of microsynteny decay between budding yeasts, mammals, and angiosperms, we plotted the patterns of microsynteny conservation for two budding yeast clades whose estimated times of origins are comparable to those of mammals and angiosperms: the clade of Saccharomycetaceae + Saccharomycodaceae (~170 MYA) and the CUG-Ser1 major clade (~200 MYA) (Figures S3C, S3D). We found that the overall microsynteny is more conserved in budding yeasts and mammals than angiosperms, suggesting that angiosperm genomes are highly fractionated and reshuffled.

Different rates of microsynteny evolution in major eukaryotic lineages

Eukaryotic genomes differ substantially in their structure and organization across lineages. To assess the overall impact of evolutionary divergence on budding yeast microsyntenic conservation, we summarized the shared syntenic percentage of homologous genes for all pairwise comparisons into a heatmap matrix organized using the same species phylogenetic order as in Figure S3E. We found that budding yeast genomes show clear major clade-specific patterns of microsynteny conservation, with many syntenic homologous genes found between genomes within each major clade but few found between genomes that belong to different major clades. One exception to this pattern was *Hanseniaspora vineae*, which belongs to the family Saccharomycodaceae. *H. vineae* shares a higher syntenic percentage of homologous genes with genomes of species in the Saccharomycetaceae family (average = 50.89%) than it does with *H. valbyensis* and *H. uvarum* (average = 39.87%), two other members of the genus *Hanseniaspora* that also belong to the family Saccharomycodaceae (Figure S3F). Both *H. valbyensis* and *H. uvarum* lost many DNA repair genes, underwent rapid genome evolution, and have highly variable ploidies compared to other budding yeasts[16]. Furthermore, the genomes of *Hanseniaspora* species have been shown to be highly dynamic[17]. These results suggest the fast-evolving *Hanseniaspora* genomes also underwent extensive rearrangements, possibly driven by the loss of DNA repair genes.

To examine the relationship between synteny conservation and evolutionary divergence, we first calculated the pairwise syntenic percentage of homologous genes and the evolutionary distance (tip-to-tip distance in the phylogeny) between the *S. cerevisiae* genome and those of all other 119 species in our dataset (Figure 3A). We also performed the same analysis using *C. albicans*, *O. parapolymorpha*, and *Y. lipolytica* (6 chromosomes) (Figures 3B, 3C, 3D) as references. In all cases, we found that conservation of microsynteny decreases (and evolutionary distance increases) in relation to divergence time. The overall trend of the decay of microsynteny is very similar for all species (Pearson's correlation coefficient: $p < 2.2e\text{-}16$) (Figure 3). Interestingly, we found that closely related species exhibit high conservation of microsynteny, and the pairwise syntenic percentage decreases exponentially with increasing divergence time for divergence times below 200 MYA. Distantly related budding yeast species that diverged more than 200 MYA exhibit very low syntenic percentages of homologous genes that decrease very slowly with increasing evolutionary distance, indicating there is a small percentage of genes whose microsynteny is conserved across the subphylum, whose origin dates to 400 MYA (Figure 3). To examine if there are functional constraints associated with homologous genes whose microsynteny is conserved across the subphylum, we identified 301 subnetworks that are widely conserved across budding yeasts (see Methods). Gene ontology enrichment analysis of these genes shows that they are significantly enriched in metabolism-related terms (Data S2). These results are consistent with previous work suggesting that genes in the same metabolic pathway are significantly clustered in eukaryotic[18] and fungal[19,20] genomes.

We then compared the association between microsynteny conservation and evolutionary divergence between budding yeasts, angiosperms, and mammals (Figure 3E, Data S3). We found that angiosperms tend to show lower levels of microsynteny conservation than mammals and budding yeasts (Figure 3E). Moreover, we examined the association between gene gain/loss and microsynteny conservation. We used OrthoFinder to identify the numbers of shared orthologs within budding yeasts, mammals, and angiosperms and summarized the number of shared orthologs in each clade (Figure S3F). In general, we found similar numbers of shared orthologs across the three lineages, although their number decreases slightly as evolutionary distance increases (Figure S3F). Angiosperms did not exhibit a higher degree of gene gain/loss compared to budding yeasts and mammals (Figure S3F), so the lack of synteny conservation in angiosperms might be due to the repeated occurrence of WGD events and / or their higher content of transposable elements[9].

Large-scale gene duplication events are potentially widespread in budding yeasts

Gene and genome duplication are thought to have been key contributors to the evolution of biodiversity[21]. We next examined the evolution of all genes in our 120 budding yeast genomes with respect to different modes of gene duplication as part of our microsynteny pipeline. We identified duplicated genes using duplicate_gene_classfier employed in MCScanx and classified them into one of the five categories (Figure S4, Data S4): those being derived from whole genome or segmental duplication (WGD/SD), those from tandem duplication (TD), those from proximal duplication (PD), those from dispersed duplications (DD), and those that are singletons.

It is well known that *S. cerevisiae* and its close relatives (i.e., the WGD clade) arose from ancient whole-genome duplication caused by allopolyploidization and followed by massive gene loss[22–24]. As expected, we found higher percentages of WGD/SD-derived genes in the genomes of species from the WGD clade, including *S. cerevisiae* (10.8% WGD-derived genes). Moreover, we also identified other instances of homologous genes (350 subnetworks) whose microsynteny is conserved in a manner consistent with the WGD event in the Saccharomycetaceae WGD clade (colored in yellow) (Figure 2). Surprisingly, we found several species in the WGD clade that contained very few WGD-derived genes, such as the opportunistic pathogen *Candida (Nakaseomyces) glabrata* (0.27% WGD-derived genes) and its close relatives ("*glabrata* group") (Figure S4). Since WGD is often followed by extensive loss of duplicated genes[22,25], our results are consistent with previous work suggesting that the "*glabrata* group" experienced higher rates of gene loss after WGD events compared to other species in WGD clade[26]. This finding is also largely consistent with previous results suggesting that the "*glabrata* group" lineage reduced its set of protein-coding genes after separation from other post-WGD yeasts[27].

A higher frequency of predicted WGD/SD-derived genes is also observed in certain species in the Dipodascaceae/Trichomonascaceae clade, such as *Nadsonia fulvescens* (5.22%), *Geotrichum candidum* (7.26%), *Blastobotrys raffinofermentans* (6.28%), and *Wickerhamiella versatilis* (4.44%). Larger percentages of WGD/SD-derived genes are also identified in individual species in Lipomycetaceae, Phaffomycetaceae, and Pichiaceae clades (Figure 4). Although further analyses are warranted, these results suggest that segmental duplications and even whole genome duplication events might be more widespread in budding yeasts than previously recognized (see also[28]).

The *GAL* gene cluster may have originated in the *GAL10* genomic neighborhood

The conservation of macrosynteny decayed very fast in budding yeast genomes but the microsynteny of some genomic regions was much more deeply conserved both within major clades and across the subphylum. Studying the deep conservation of gene order can illuminate the relationship between genome architecture and organismal function and ecology[20,29,30]. For example, the physical linkage of the structural genes *GAL1*, *GAL7*, and *GAL10* of the *GAL*actose utilization pathway in diverse budding yeast genomes has been used as a model for understanding the evolution of metabolic gene clusters in eukaryotes[8,31,32]. By examining the microsynteny subnetworks and the gene organizations of the *GAL1*, *GAL7*, and *GAL10* genes across the 120 species (Figure 4), we found that *GAL10* genes show greater conservation of their microsynteny than *GAL1* and *GAL7* across budding yeast genomes. This raises the hypothesis that the *GAL* gene clusters of budding yeasts might have originated in the *GAL10* syntenic neighborhood (see Star Methods for more details).

Conclusion

In this study we examined the tempo and mode of evolution of genome organization within budding yeasts and compared it to those observed in other fungi, animals, and plants. We identified two distinct modes of evolution of genome organization in budding yeasts: (1) at the large-scale chromosome-level of organization, we found a faster decay of macrosynteny conservation compared to filamentous fungi and animals, which is corroborated by findings of rapid chromosome structure evolution in budding yeasts from the genus *Lachancea*[33]; (2) at the small-scale gene-level of organization, we identified both deeply conserved and lineage-specific instances of conservation of microsynteny across budding yeast genomes. The decay in microsynteny is generally correlated with evolutionary divergence, suggesting that it is most likely a neutral process[18]. In contrast, the microsynteny of certain genes is much more deeply conserved, suggesting that there are selective advantages to the evolutionary maintenance[20,34]. These results provide a robust framework to explore the evolution of fungal and eukaryotic genome organization.

Acknowledgments

We thank members of the Rokas laboratory and Y1000+ Project team (http://y1000plus.org) for discussions and comments. We thank Darrin Schultz for his help with the oxford dot plot analyses. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University and Yale Center for Research Computing (Farnam HPC cluster) for use of the research computing infrastructure. This work was supported by Shandong University Outstanding Youth Fund (62420082260514) to Y.L., the National Science Foundation (DEB-1442113 and DEB-2110404 to A.R.; DEB-1442148 and DEB-2110403 to C.T.H), in part by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02- 07ER64494), and the USDA National Institute of Food and Agriculture (Hatch project 1020204 to C.T.H.). C.T.H. is a H. I. Romnes Faculty Fellow, supported by the Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation, respectively. X.X.S. was supported by the National Natural Science Foundation of China (32071665). A.R. received additional support from the Burroughs Wellcome Fund, and the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R56 AI146096 and R01 AI153356). J. L. S. and A. R. received additional support by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program.

Author contributions

Y.L. and A.R. designed this study. Y.L., H.L., and J.L.S. conducted analyses and prepared figures. Y.L. and A.R. wrote the paper. All authors provided feedback on the experimental design, discussed the results and implications, and commented on the manuscript at all stages.

Inclusion and Diversity

We support inclusive, diverse, and equitable conduct of research.

Declaration of Interests

A. R. is a scientific consultant for LifeMine Therapeutics, Inc. and a member of the Advisory Board of *Current Biology*. J.L.S. is a scientific consultant for Latch AI Inc.

Figure Legends

Figure 1. The decay of macrosynteny conservation between *Saccharomyces cerevisiae*, *Candida albicans*, and their close relatives in the budding yeast subphylum

A. Oxford dot plots of homologous genes between *C. albicans* and three representative closely related species. The colored dots correspond to homologous genes from the chromosomes of *C. dubliniensis*, *C. parapsilosis*, or *Spathaspora passalidarum* and *C. albicans*, with chromosome boundaries indicated and sorted based on chromosomal size. The time-calibrated species tree on the left was obtained from a previous study of 332 budding yeast species[35]. B. An Oxford dot plot of homologous genes between *S. cerevisiae* and three representative closely related species. Note the lack of conservation of macrosynteny after ~100 million years of divergence in both lineages. C. Macrosynteny conservation index between *C. albicans* and all other 119 budding yeast genomes. D. Macrosynteny conservation index between *S. cerevisiae* and all other 119 budding yeast genomes. Related to Figures. S1, S2 and Data S1.

Figure 2. Microsynteny is conserved within major clades of budding yeasts, as well as across the entire subphylum

Phylogenomic microsynteny profiling of all budding yeast subnetworks (size ≥ 3 genes). The X-axis corresponds to the phylogeny of the 120 budding yeast species used in this study, which was taken from a previous study [35] . Gene copy numbers of orthogroups are labeled in different colors. Some of the blocks of orthogroups that display lineage-specific conservation of microsynteny are also labeled, including the block of homologous genes stemming from the whole genome duplication (WGD) event in Saccharomycetaceae and whose microsynteny is conserved. Note that microsynteny appears to be conserved for other instances of WGD or large-scale segmental duplications in the subphylum. Overall, the microsynteny of budding yeasts is less conserved than mammals but more conserved than angiosperms. Related to Figures S3, S4, Data S1, S2 and S4.

Figure 3. Conservation of microsynteny decays at the same rate within budding yeasts but at different rates in budding yeasts, mammals, and angiosperms.

A. Plot of pairwise conservation of microsynteny (syntenic percentage; in turquoise color) and evolutionary distance (tip-to-tip distance in the phylogeny; in orange color) versus divergence time (in million years) between the *S. cerevisiae* genome and those of all other 119 other budding yeast species. We also performed the same analysis using (B) *C. albicans* and (C) *Ogataea parapolymorpha* (D) *Yarrowia lipolytica* as references. The Pearson correlation coefficient (Pearson's r) between the evolutionary distance and divergence time estimated was calculated using R. (E) Microsynteny conservation versus evolutionary distance for the lineages of budding yeasts, angiosperms, and mammals. Note that comparisons of genomes with similar levels of evolutionary distance (i.e., mammals and budding yeasts) tend to show higher levels of microsynteny conservation than angiosperms. (F) The correlation between the number of shared orthologs and microsynteny conservation is largely similar across mammals, angiosperms, and budding yeasts. Related to Data S3.

Figure 4. Conservation of microsynteny of the structural genes involved in *GAL*actose metabolism (*GAL1*, *GAL7*, and *GAL10*) suggests that the ancestor of the *GAL* gene cluster of *S. cerevisiae* and *C. albicans* likely originated in the *GAL10* genomic neighborhood.

A. The microsynteny subnetworks for the *GAL1*, *GAL7*, and *GAL10* genes in budding yeasts.

Nodes represent genes, and edges represent syntenic relationships between genes. Node colors represent different *GAL* genes: *GAL1* (dark purple); *GAL3* (light purple); *GAL7* (orange); *GAL10* (green). Yellow nodes represent duplicated *GAL* genes. B. Genomic organization of *GAL* metabolic cluster genes in different major clades of the budding yeast subphylum. Grey lines correspond to syntenic relationships between homologous genes. The rectangle dotted box represents *GAL* gene clusters of species that contain multiple *GAL1* genes. Only one copy of *GAL1* was identified in our synteny subnetwork of *Lipomyces starkeyi* and *Lipomyces mesembrius*, whereas three copies of *GAL1* were identified previously (LaBella et al. 2021; Harrison et al. 2021; Haase et al. 2021). Interestingly, the two copies of *GAL1* absent from our subnetwork are more similar in their sequences to the *GAL1* genes of filamentous fungi but are adjacent to the *GAL*7 and *GAL10* gene in genomes of *L. starkeyi* and *L. mesembrius* (See Results and Discussion section for more details). Thus, we labeled the two additional copies of *GAL1* in *L. starkeyi* and *L. mesembrius* as syntenic to other *GAL1* genes in dotted grey lines.

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Antonis Rokas (antonis.rokas@vanderbilt.edu).

Materials Availability

There are no materials to report.

Data and Code Availability

- All genome assemblies, data matrices, dop plot analyses and related figures and tables are deposited at a Figshare repository and are publicly available as of the date of publication. The public link to the repository is available via the link: https://doi.org/10.6084/m9.figshare.19508752. The DOI is listed in the key resources table.
- All original code is deposited at a Figshare repository and is publicly available as of the date of publication. The public link to the repository is available via the link: https://doi.org/10.6084/m9.figshare.19508752. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this study is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sequence data

To collect a high-quality set of genomes to study the evolution of budding yeast genome organization, we first retrieved the 332 publicly available Saccharomycotina yeast genomes, gene annotations, species trees, and Bayesian time-calibrated trees from a recent comprehensive genomic study of the Saccharomycotina yeasts[35]. To reduce the burden of computation but retain the breadth of genetic diversity of major yeast lineages where the genomes of more than 10 species are available, we retained higher-quality genomes based on their genome assembly statistics, including the number of contigs (< 100 contigs), N50 size (> 500 kb), and BUSCO completeness (> 90% completeness); this was the case for the major clades Saccharomycetaceae, Pichiaceae, Phaffomycetaceae, CUG-Ser1 clade, and Dipodascaceae/Trichomonascaceae. For major clades where the genomes of fewer than 10 species are available, we used a relaxed filtering strategy based on the number of contigs (<700 contigs) and N50 size (> 100 kb). The final dataset contained 120 budding yeast genomes; detailed information about these genomes can be found in Data S1. Each genome containing all protein sequences was searched against the Saccharomycotina_odb9 database[36].

METHOD DETAILS

Quantification and statistical analysis

Macrosynteny analyses. To examine the conservation of macrosynteny, we constructed Oxford dot plots comparing the chromosomal positions of homologous genes using the genomes of *S.* cerevisiae (from the Saccharomycetaceae major lineage), *C. albicans* (from the CUG-Ser1 clade)*, Ogataea parapolymorpha* (from Pichiaceae clade), *and Yarrowia lipolytica* (from Dipodascaceae/Trichomonascaceae clade) as the anchor species and all other genomes as the target species, respectively. Oxford dot plots are a common method for examining the conservation of macrosynteny between pairs of genomes. For example, a comparison of two perfectly collinear genomes (i.e., two genomes whose orthologous genes are 100% syntenic) gives a series of dots that lie on the main diagonal. The dense rectangular blocks of dots also imply conserved macrosynteny in which genes are conserved within homologous chromosomes but with randomized orders and orientations (also referred to as mesosynteny in fungi[6]). Chromosomal inversions and translocations can also be visualized on dot plots by diagonal lines on an opposite slope, and genes on a chromosome of one species are syntenic with two or more chromosomes, respectively.

Examination of macrosynteny was conducted using the *odp* pipeline (Figshare repository). Briefly, we looked for homologous chromosomes between the anchor and target genomes by plotting the protein coordinates of reciprocal best BlastP [37] hits (evalue < 1e-5). To avoid biasing our analyses due to linked paralogs (most of which are recent tandem duplications relative to the ancient chromosome-scale events of interest), we considered only a single paralog per chromosome/contig in our analyses.

For each of the four anchor species, we selected all other 119 species at increasing evolutionary distances based on a well-established, time-calibrated genome-scale budding yeast phylogeny[35]. To quantify the degree of conservation of macrosynteny, for each dotplot, we computed the conservation, an established quantitative measure of the degree of macrosynteny conservation index, across the budding yeast subphylum. This conservation index is calculated by counting the number of one-to-one orthologous gene pairs whose genes are in homologous chromosomes/scaffolds and dividing it by the number of one-to-one orthologs whose genes reside in non-homologous chromosomes/scaffolds[3,5]. The conservation index between two given genomes ranges from 0 (no macrosynteny conservation) to 1 (highly conserved macrosynteny). We used Fisher's exact test ($p < 0.05$) for the significance of the relatedness of homologous scaffolds described in ref. [2].

To compare the conservation of macrosynteny of budding yeasts to other major eukaryotic lineages, we also constructed Oxford dot plots between representative species of filamentous fungi (using the major plant pathogen *Zymoseptoria tritici* as an anchor species), bilaterians (using the scallop *Patinopecten yessoensis*, a well-established model for studies of macrosynteny conservation in animals, as an anchor species) and angiosperms (using *Arabidopsis thaliana* as anchor species).

Microsynteny network construction. To examine the evolution of microsynteny of the budding yeast subphylum, we used the pipeline from Zhao and Schranz[7]. Briefly, we used DIAMOND v.0.9.14.115[38] to perform all inter- and intra-pairwise all-vs.-all protein similarity searches using default parameters. In total, 14,280 whole-genome comparisons were conducted for 120 budding yeast genomes. Next, we used MCScanX [39] to identify pairwise synteny blocks between species; each synteny block must have at least four homologous genes within a set of 20 colinear genes in the two species compared. The syntenic percentage between each pair of species compared was calculated using the number of syntenic pairs relative to the total number of genes[38].

We merged syntenic gene pairs from all inter- and intra-species synteny blocks into one two-columned tabular-format file, which can serve as an undirected synteny network/graph and be further analyzed or visualized in various tools. In this synteny network, nodes are genes, edges stand for syntenic relationships between nodes, and edge lengths in this study have no meaning (unweighted). Further details can be found in the GitHub tutorial (https://github.com/zhaotao1987/SynNet-Pipeline).

The entire network, consisting of millions of nodes, was split into individual subnetworks (which can be thought of orthogroups or gene families whose synteny is conserved) using the Infomap method employed in igraph[40]. The sizes of individual subnetworks were determined by considering topological edge connections. The final microsynteny network of budding yeasts contains rows and columns. Each row represents a syntenic subnetwork, and each column represents a genome. The value for each cell represents the number of genes from each genome in a given subnetwork. All genomes are arranged based on phylogenetic relationships. The dissimilarity index of all subnetworks was calculated using the Jaccard method of the vegan package[41], then hierarchically clustered by "ward.D", and visualized by R package Pheatmap[42]. We only kept subnetworks that contained three or more genes. The correlation between the evolutionary distance and divergence time was calculated using Pearson's correlation coefficient using the cor.test function in R package stats v.3.6.2[43].

To gain insight into the functional categories of subnetworks whose microsynteny is conserved across the budding yeast subphylum, we first selected those subnetworks that contain genes from at least 80% of genomes or > 96 species and from at least 10 major clades but are also present in Lipomycetaceae, the major clade that is the sister group to all other clades. We then conducted gene ontology (GO) enrichment analysis using Goatools[44] using *S. cerevisiae* genes from these subnetworks as representatives.

To compare the dynamics and properties of the entire microsynteny networks in budding yeasts to other major eukaryotic lineages, we retrieved genomes and microsynteny networks from 87 mammalian and 107 angiosperm genomes from a previous study [7]). To examine the association between gene gain/loss and microsynteny conservation, we conducted OrthoFinder [45] analyses for each dataset to summarize the number of shared orthologs of each lineage against *S. cerevisiae*, *Homo sapiens*, and *Arabidopsis thaliana*, respectively.

Phylogenetic signal and tree reconstruction. To investigate if microsynteny information can be used in genome-scale phylogenetic reconstruction, we first tested whether individual subnetworks contain phylogenetic signal based on our time-calibrated tree using Phytools[46]. We quantified the information for each individual subnetwork by fitting three alternative models that describe different evolutionary dynamics: the Brownian-motion model (BM: describes a random motion of trait evolution along branches in the phylogeny, with an increase in trait variance centered around the initial value at the root of the tree[47]), the Ornstein-Uhlenbeck model (OU: describes that once traits have adaptively evolved, stabilizing selection pulls the trait values around an adaptive optimum for the trait[48]), the Early-Burst model (EB: describes exponentially increasing or decreasing rates of evolution over time-based on the assumption that niches are saturated by accumulating species within a lineage[49]), Comparisons of the goodness of fit for these models were performed through the Akaike Information Criterion (AIC)[50].

The "synteny-based tree" was then reconstructed using the Syn-MRL pipeline, which combines synteny network analysis, matrix representation, and maximum likelihood phylogenetic inference[51]. Briefly, Syn-MRL proceeded by encoding the phylogenomic synteny network obtained above into a binary data matrix, where rows represent species, columns represent subnetworks, and each cell was coded as a binary character (presence or absence of an individual subnetwork in a given species). Tree estimation was based on maximum-likelihood as implemented in IQ-TREE 2.1.2[52], using the binary MK+R+FO model (a Jukes-Cantor type model for discrete morphological data)[53]. The topological robustness of the topology was evaluated by 1,000 ultrafast bootstrap replicates[54]. We quantified the degree of incongruence for every internode by considering all prevalent conflicting bipartitions between “synteny-based tree” and “sequenced-based tree” derived from the previous analysis (sequence tree derived from 2408OG data matrix) using the “compare” function in Gotree version 1.13.6 (https://github.com/evolbioinfo/gotree).

It has been suggested microsynteny could be used as an additional marker for phylogenomic analyses[55,56]. We found that 98.8% of microsyntenic subnetworks (16,807 / 17,010) contain strong phylogenetic signal (p-value < 0.05) (Figshare repository). Inference and subsequent comparison of the “synteny-based tree” method to two standard approaches of phylogenomic inference (“sequence-based tree”), namely maximum likelihood (ML) analyses based on concatenation and coalescence, showed that the tree inferred using microsynteny information shared 88.03% of bipartitions with the concatenation tree and 87.29% with the coalescence tree (Figshare repository); for reference, the trees inferred from concatenation and coalescence approaches shared 97.4% of bipartitions. These results, together with other recent findings[57], suggest that microsynteny may be a useful, additional marker for phylogenomic studies.

To estimate the divergence time of previous mammalian and angiosperm datasets, we first retrieved the protein sequences from 87 mammalian and 107 angiosperm genomes (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BDMA7A)[7]. To obtain the “single-copy” orthologs for both mammalian and angiosperm genomes, we conducted Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.1.3 [58] analysis and mammalia_odb10 and embryophyta_odb10 databases for each genome, respectively. To minimize missing data and computational burden, we retained 3\00 single-copy BUSCO genes that are present in all taxa. For both mammalian and angiosperm datasets, we used the concatenation approach with a single model using IQ-TREE and used the r8s algorithm v. 1.70 [59] to conduct divergence time estimation without any fossil calibrations except for the root position (set as 170 MYA) based on the previous study [7].

Synteny network for the Galactose (*GAL*) clustering genes

To examine the evolution of the *GAL* gene cluster of budding yeasts, the *GAL1*, *GAL7*, and *GAL10* genes were obtained from the comparative analysis of the *GAL* pathway in budding yeasts [60]. All subnetworks containing *GAL* genes were extracted from the total network of 120 budding yeast genomes identified above. The subnetworks for *GAL1*, *GAL7*, and *GAL10* genes were then imported and visualized in Cytoscape 3.7.0[61].

The conservation of macrosynteny decayed very fast in budding yeast genomes but the microsynteny of some genomic regions was much more deeply conserved both within major clades and across the subphylum. Studying the deep conservation of gene order can illuminate the relationship between genome architecture and organismal function and ecology[20,29,30]. For example, the physical linkage of the structural genes *GAL1*, *GAL7*, and *GAL10* of the *GAL*actose utilization pathway in diverse budding yeast genomes has been used as a model for understanding the evolution of metabolic gene clusters in eukaryotes[8,31,32]. To further delve into an example of deep microsynteny conservation across the budding yeast subphylum, we examined the microsynteny subnetworks and the gene organizations of the *GAL1*, *GAL7*, and *GAL10* genes across the 120 species (Figure 4).

For *GAL1*, which is found in three subnetworks, we found that most *GAL1* genes reside in a single subnetwork that contains homologs from most budding yeast species; the *GAL1* genes from several early-diverging species (e.g., *Lipomyces*, *Trigonopsis*, *Blastobotrys*) are in another subnetwork that is loosely connected to the first subnetwork (Figure 4A). As expected, we also found that *GAL3* genes (which are paralogs of *GAL1* from the yeast WGD event) are still syntenic to *GAL1* in the Saccharomycetaceae WGD clade and are part of the largest *GAL1* subnetwork (Figure 4A). It should be noted that our analyses identified only one copy of *GAL1* from *Lipomyces starkeyi* and *Lipomyces mesembrius*, whereas three copies of *GAL1* gene were identified in our previous analyses[8,60,62] (Figure 4B). Interestingly, the two copies of *GAL1* genes absent from this analysis are adjacent to the *GAL7* and *GAL10* gene in genomes of *L. starkeyi* and *L. mesembrius*. We manually blasted these two *GAL1* genes and found that both genes are indeed *GAL1* genes predicted to encode galactokinases, but their best hits are from filamentous fungi (subphylum Pezizomycotina), instead of Saccharomycotina. Notably, we found that the *GAL1* genes in *Yarrowia* species formed a distinct third subnetwork, suggesting that the synteny of the genomic neighborhoods of these genes is not conserved in other budding yeasts. Moreover, *GAL7*, which is found in four subnetworks, exhibits a pattern of microsynteny conservation largely congruent with that of *GAL1*; the only difference is that the *GAL7* genes of *Trigonopsis* and *Blastobotrys* species, in addition to those of *Yarrowia*, also formed their own subnetworks (Figure 4B).

Finally, we found that *GAL10* is in two subnetworks, which are more conserved than those of *GAL1* and *GAL7* (Figure 4C). Most *GAL10* genes are part of a large subnetwork, but the *GAL10* genes of many species in the Dipodascaceae / Trichomonascaceae clade are part of a second subnetwork that is connected to the first (Figure 4A). In contrast to the *GAL1* and *GAL7* subnetworks, we found that the *GAL10* genes from *Yarrowia* and *Trigonopsis* species also reside in the major subnetwork (Figure 4A). These results suggest that *GAL10* genes show greater conservation of their microsynteny than *GAL1* and *GAL7* across budding yeast genomes, raising the hypothesis that the *GAL* gene clusters of budding yeasts might have originated in the *GAL10* syntenic neighborhood.

Supplemental Data

Data S1. Summary information of 120 budding yeast genomes used in this study. Related to Figures 1, 2.

Data S2. Gene ontology enrichment analysis of 301 conserved subnetworks across budding yeast genomes. Related to Figure 2.

Data S3. Summary of evolutionary distances, divergence time and microsyntenic percentage across mammal, angiosperm and budding yeast datasets. Related to Figure 3.

Data S4. Summary of predicted gene duplication categories of 120 budding yeast genomes. Related to Figure 2.

References

1. McLysaght, A., Enright, A.J., Skrabanek, L., and Wolfe, K.H. (2000). Estimation of synteny conservation and genome compaction between pufferfish (Fugu) and human. Yeast *17*, 22–36.

2. Simakov, O., Marlétaz, F., Yue, J.-X., O'Connell, B., Jenkins, J., Brandt, A., Calef, R., Tung, C.-H., Huang, T.-K., Schmutz, J., et al. (2020). Deeply conserved synteny resolves early events in vertebrate evolution. Nat Ecol Evol *4*, 820–830.

3. Simakov, O., Bredeson, J., Berkoff, K., Marletaz, F., Mitros, T., Schultz, D.T., O'Connell, B.L., Dear, P., Martinez, D.E., Steele, R.E., et al. (2022). Deeply conserved synteny and the evolution of metazoan chromosomes. Sci Adv *8*, eabi5884.

4. Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D., et al. (2013). Insights into bilaterian evolution from three spiralian genomes. Nature *493*, 526–531.

5. Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., et al. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat Ecol Evol *1*, 120.

6. Hane, J.K., Rouxel, T., Howlett, B.J., Kema, G.H.J., Goodwin, S.B., and Oliver, R.P. (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. Genome Biol. *12*, R45.

7. Zhao, T., and Schranz, M.E. (2019). Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. Proc. Natl. Acad. Sci. U. S. A. *116*, 2165–2174.

8. Harrison, M.-C., LaBella, A.L., Hittinger, C.T., and Rokas, A. (2021). The evolution of the GALactose utilization pathway in budding yeasts. Trends Genet. 10.1016/j.tig.2021.08.013.

9. Schmidt, R. (2000). Synteny: recent advances and future prospects. Curr. Opin. Plant Biol. *3*, 97–102.

10. Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., van de Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., et al. (2011). Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. Nat. Commun. *2*, 202.

11. Simillion, C., Vandepoele, K., Van Montagu, M.C.E., Zabeau, M., and Van de Peer, Y. (2002). The hidden duplication past of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U. S. A. *99*, 13627–13632.

12. Salari, R., and Salari, R. (2017). Investigation of the Best *Saccharomyces cerevisiae* Growth Condition. Electron. Physician *9*, 3592–3597.

13. Trinci, A.P. (1969). A kinetic study of the growth of *Aspergillus nidulans* and other fungi. J. Gen. Microbiol. *57*, 11–24.

14. Shen, X.-X., Steenwyk, J.L., LaBella, A.L., Opulente, D.A., Zhou, X., Kominek, J., Li, Y., Groenewald, M., Hittinger, C.T., and Rokas, A. (2020). Genome-scale phylogeny and contrasting

modes of genome evolution in the fungal phylum Ascomycota. Sci Adv *6*. 10.1126/sciadv.abd0079.

15. Drillon, G., and Fischer, G. (2011). Comparative study on synteny between yeasts and vertebrates. C. R. Biol. *334*, 629–638.

16. Steenwyk, J.L., Opulente, D.A., Kominek, J., Shen, X.-X., Zhou, X., Labella, A.L., Bradley, N.P., Eichman, B.F., Čadež, N., Libkind, D., et al. (2019). Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. PLoS Biol. *17*, e3000255.

17. Saubin, M., Devillers, H., Proust, L., Brier, C., Grondin, C., Pradal, M., Legras, J.-L., and Neuvéglise, C. (2019). Investigation of Genetic Relationships Between Hanseniaspora Species Found in Grape Musts Revealed Interspecific Hybrids With Dynamic Genome Structures. Front. Microbiol. *10*, 2960.

18. Koonin, E.V. (2009). Evolution of genome architecture. Int. J. Biochem. Cell Biol. *41*, 298–306.

19. Wisecaver, J.H., Slot, J.C., and Rokas, A. (2015). Correction: The Evolution of Fungal Metabolic Pathways. PLoS Genet. *11*, e1005449.

20. Rokas, A., Wisecaver, J.H., and Lind, A.L. (2018). The birth, evolution and death of metabolic gene clusters in fungi. Nat. Rev. Microbiol. *16*, 731–744.

21. Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., and Paterson, A.H. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biol. *20*, 38.

22. Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. Nature *387*, 708–713.

23. Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature *428*, 617–624.

24. Marcet-Houben, M., and Gabaldón, T. (2015). Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. PLoS Biol. *13*, e1002220.

25. Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science *290*, 1151–1155.

26. Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., et al. (2004). Genome evolution in yeasts. Nature *430*, 35–44.

27. Ahmad, K.M., Kokošar, J., Guo, X., Gu, Z., Ishchuk, O.P., and Piškur, J. (2014). Genome structure and dynamics of the yeast pathogen *Candida glabrata*. FEMS Yeast Res. *14*, 529–535.

28. Gabaldón, T. (2020). Hybridization and the origin of new yeast lineages. FEMS Yeast Res. *20*. 10.1093/femsyr/foaa040.

29. Zhao, T., Holmer, R., de Bruijn, S., Angenent, G.C., van den Burg, H.A., and Schranz, M.E. (2017). Phylogenomic Synteny Network Analysis of MADS-Box Transcription Factor Genes Reveals Lineage-Specific Transpositions, Ancient Tandem Duplications, and Deep Positional Conservation. Plant Cell *29*, 1278–1292.

30. Dewey, C.N. (2011). Positional orthology: putting genomic evolutionary relationships into context. Brief. Bioinform. *12*, 401–412.

31. Hittinger, C.T., Gonçalves, P., Sampaio, J.P., Dover, J., Johnston, M., and Rokas, A. (2010). Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature *464*, 54–58.

32. Slot, J.C., and Rokas, A. (2010). Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. Proc. Natl. Acad. Sci. U. S. A. *107*, 10136–10141.

33. Vakirlis, N., Sarilar, V., Drillon, G., Fleiss, A., Agier, N., Meyniel, J.-P., Blanpain, L., Carbone, A., Devillers, H., Dubois, K., et al. (2016). Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. Genome Res. *26*, 918–932.

34. McGary, K.L., Slot, J.C., and Rokas, A. (2013). Physical linkage of metabolic genes in fungi is an adaptation against the accumulation of toxic intermediate compounds. Proc. Natl. Acad. Sci. U. S. A. *110*, 11481–11486.

35. Shen, X.-X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase, M.A.B., Wisecaver, J.H., Wang, M., Doering, D.T., et al. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell *175*, 1533–1545.e20.

36. Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simão, F.A., Ioannidis, P., Seppey, M., Loetscher, A., and Kriventseva, E.V. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. *45*, D744–D749.

37. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

38. Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59.

39. Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49.

40. Csardi, G., Nepusz, T., and Others (2006). The igraph software package for complex network research. InterJournal, complex systems *1695*, 1–9.

41. Dixon, P. (2003). VEGAN, a package of R functions for community ecology. J. Veg. Sci. *14*, 927–930.

42. Raivo, K. (2019). Pheatmap: pretty heatmaps. R package version *1*.

43. Team, R.C., Team, M.R.C., Suggests, M., and Matrix, S. (2018). Package stats. The R Stats Package.

44. Klopfenstein, D.V., Zhang, L., Pedersen, B.S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., et al. (2018). GOATOOLS: A Python library for Gene Ontology analyses. Sci. Rep. *8*, 10872.

45. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome

comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157.

46. Revell, L.J. (2012). Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3: 217--223. http s. doi. org/10. 1111/j.

47. Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. *25*, 471–492.

48. Butler, M.A., and King, A.A. (2004). Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. Am. Nat. *164*, 683–695.

49. Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W., Kozak, K.H., McPeek, M.A., Moreno-Roark, F., Near, T.J., et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. Evolution *64*, 2385–2396.

50. Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Automat. Contr. *19*, 716–723.

51. Zhao, T., Zwaenepoel, A., Xue, J.-Y., Kao, S.-M., Li, Z., Schranz, M.E., and Van de Peer, Y. (2021). Whole-genome microsynteny-based phylogeny of angiosperms. Nat. Commun. *12*, 3498.

52. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 10.1093/molbev/msaa015.

53. Lewis, P.O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. *50*, 913–925.

54. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. *35*, 518–522.

55. Sankoff, D., Leduc, G., Antoine, N., Paquin, B., Lang, B.F., and Cedergren, R. (1992). Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. U. S. A. *89*, 6575–6579.

56. Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. *6*, 361–375.

57. Drillon, G., Champeimont, R., Oteri, F., Fischer, G., and Carbone, A. (2020). Phylogenetic Reconstruction Based on Synteny Block and Gene Adjacencies. Mol. Biol. Evol. *37*, 2747–2762.

58. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. *35*, 543–548.

59. Sanderson, M.J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics *19*, 301–302.

60. LaBella, A.L., Opulente, D.A., Steenwyk, J.L., Hittinger, C.T., and Rokas, A. (2021). Signatures of optimal codon usage in metabolic genes inform budding yeast ecology. PLoS Biol. *19*, e3001185.

61. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular

interaction networks. Genome Res. *13*, 2498–2504.

62. Haase, M.A.B., Kominek, J., Opulente, D.A., Shen, X.-X., LaBella, A.L., Zhou, X., DeVirgilio, J., Hulfachor, A.B., Kurtzman, C.P., Rokas, A., et al. (2021). Repeated horizontal gene transfer of GALactose metabolism genes violates Dollo's law of irreversible loss. Genetics *217*. 10.1093/genetics/iyaa012.

KEY RESOURCES TABLE

REAGENT or	SOURCE	IDENTIFIER
Deposited Data		
Genome assemblies	NCBI	See Table S1; Figshare repository: https://doi.org/10.6084/m9.figshare.19508752
Macrosynteny data	This study	Figshare repository: https://doi.org/10.6084/m9.figshare.19508752
Microsynteny data	This study	Figshare repository: https://doi.org/10.6084/m9.figshare.19508752
Software and Algorithms		
BUSCO v5.1.3	[58]	https://busco.ezlab.org/
MCScanX2	[39]	https://github.com/wyp1125/MCScanX
DIAMOND v0.9.14.115	[38]	https://github.com/bbuchfink/diamond
BLASTP	[37]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
SynNet-Pipeline	[7]	https://github.com/zhaotao1987/SynNet-Pipeline
Igraph 0.10.0	[40]	https://igraph.org/
Gotree v1.13.6	https://github.com/evolbioinfo/gotree	https://github.com/evolbioinfo/gotree
OrthoFinder v2.5.4	[45]	https://github.com/davidemms/OrthoFinder
R8s v 1.70	[59]	https://sourceforge.net/projects/r8s/
R package stats v3.6.2	[42]	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html
Cytoscape v3.7.0	[61]	https://cytoscape.org/
IQ-TREE v2.1.2	[52]	http://www.iqtree.org/
odp	https://doi.org/10.6084/m9.figshare.19508752	https://doi.org/10.6084/m9.figshare.19508752

Figure 1

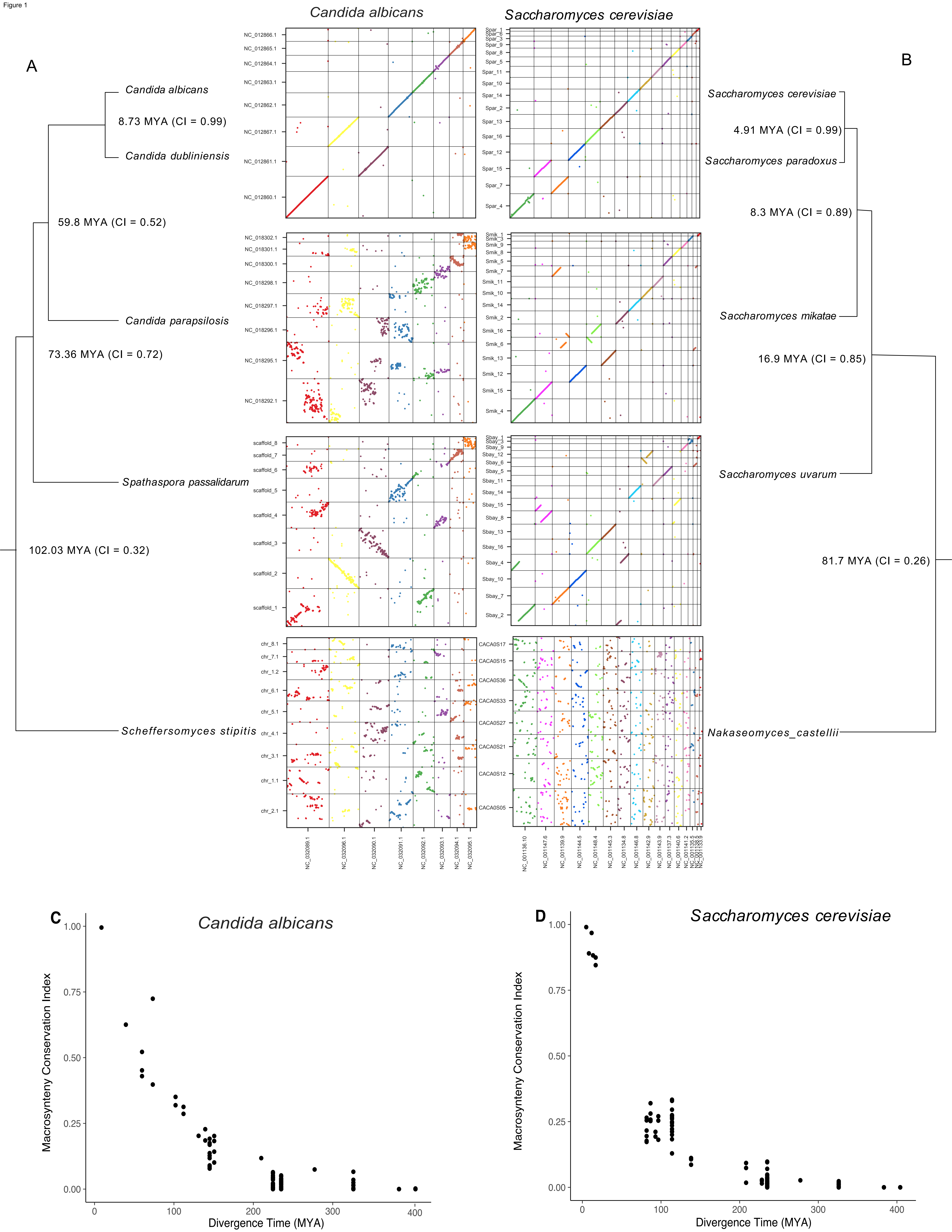
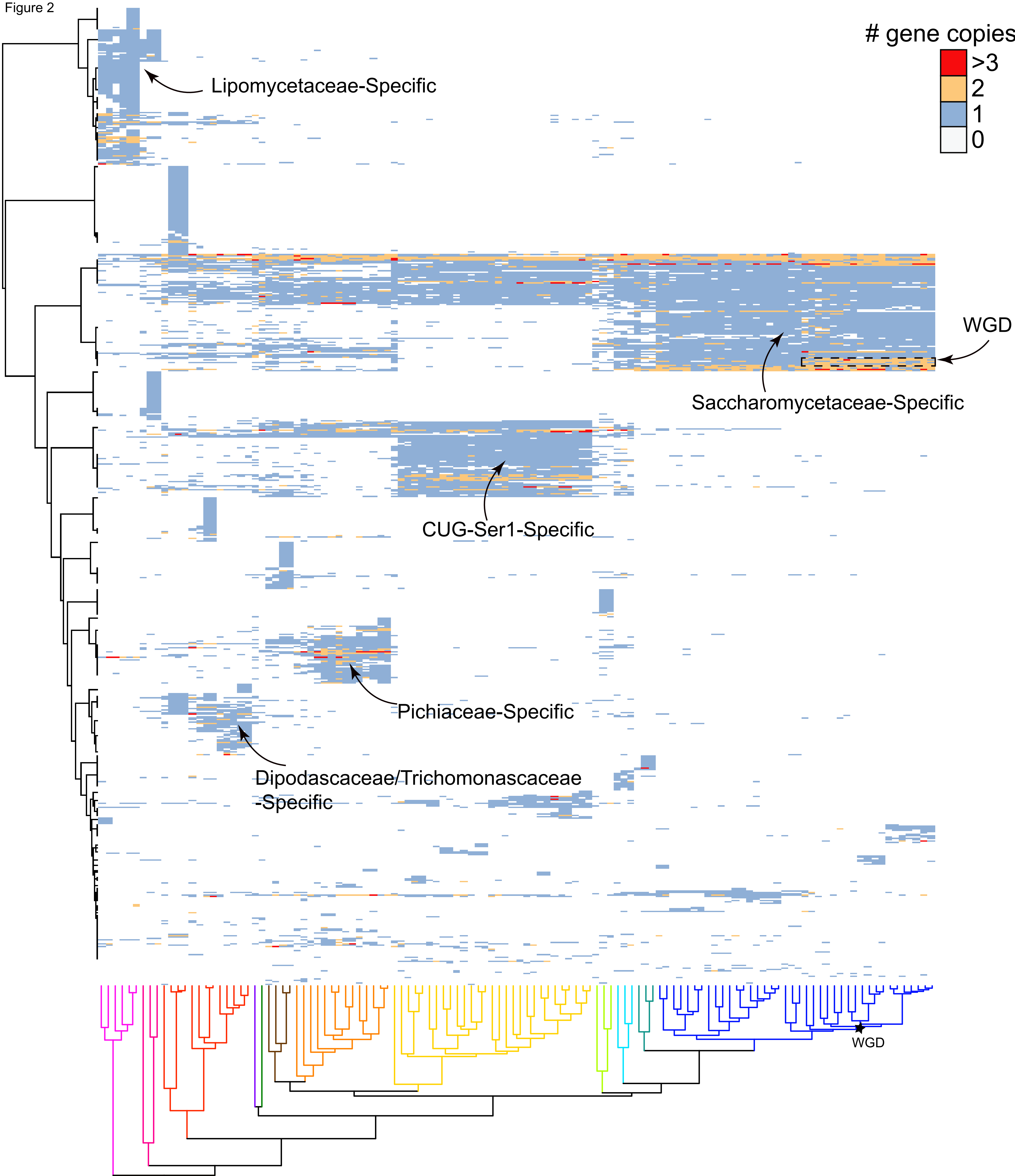

A
Candida albicans
8.73 MYA (CI = 0.99)
Candida dubliniensis
59.8 MYA (CI = 0.52)
Candida parapsilosis
73.36 MYA (CI = 0.72)
Spathaspora passalidarum
102.03 MYA (CI = 0.32)
Scheffersomyces stipitis
Candida albicans
Saccharomyces cerevisiae
NC_012866.1
NC_012865.1
NC_012864.1
NC_012863.1
NC_012862.1
NC_012867.1
NC_012861.1
NC_012860.1
NC_018302.1
NC_018301.1
NC_018300.1
NC_018298.1
NC_018297.1
NC_018296.1
NC_018295.1
NC_018292.1
scaffold_8
scaffold_7
scaffold_6
scaffold_5
scaffold_4
scaffold_3
scaffold_2
scaffold_1
chr_8.1
chr_7.1
chr_1.2
chr_6.1
chr_5.1
chr_4.1
chr_3.1
chr_1.1
chr_2.1
NC_032089.1
NC_032096.1
NC_032090.1
NC_032091.1
NC_032092.1
NC_032093.1
NC_032094.1
NC_032095.1
Spar_1
Spar_6
Spar_3
Spar_9
Spar_8
Spar_5
Spar_11
Spar_10
Spar_14
Spar_2
Spar_13
Spar_16
Spar_12
Spar_15
Spar_7
Spar_4
Smik_1
Smik_3
Smik_9
Smik_8
Smik_5
Smik_7
Smik_11
Smik_10
Smik_14
Smik_2
Smik_16
Smik_6
Smik_13
Smik_12
Smik_15
Smik_4
Sbay_1
Sbay_3
Sbay_9
Sbay_12
Sbay_6
Sbay_5
Sbay_11
Sbay_14
Sbay_15
Sbay_8
Sbay_13
Sbay_16
Sbay_4
Sbay_10
Sbay_7
Sbay_2
CACA0S17
CACA0S15
CACA0S36
CACA0S33
CACA0S27
CACA0S21
CACA0S12
CACA0S05
NC_001136.10
NC_001147.6
NC_001139.9
NC_001144.5
NC_001148.4
NC_001145.3
NC_001134.8
NC_001146.8
NC_001142.9
NC_001143.9
NC_001137.3
NC_001140.6
NC_001141.2
NC_001135.5
NC_001138.5
NC_001133.9
B
Saccharomyces cerevisiae
4.91 MYA (CI = 0.99)
Saccharomyces paradoxus
8.3 MYA (CI = 0.89)
Saccharomyces mikatae
16.9 MYA (CI = 0.85)
Saccharomyces uvarum
81.7 MYA (CI = 0.26)
Nakaseomyces_castellii
C
Candida albicans
Macrosynteny Conservation Index
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
Divergence Time (MYA)
D
Saccharomyces cerevisiae
Macrosynteny Conservation Index
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
Divergence Time (MYA)

Figure 2



gene copies
>3
2
1
0
Lipomycetaceae-Specific
WGD
Saccharomycetaceae-Specific
CUG-Ser1-Specific
Pichiaceae-Specific
Dipodascaceae/Trichomonascaceae
-Specific
WGD

Figure 3

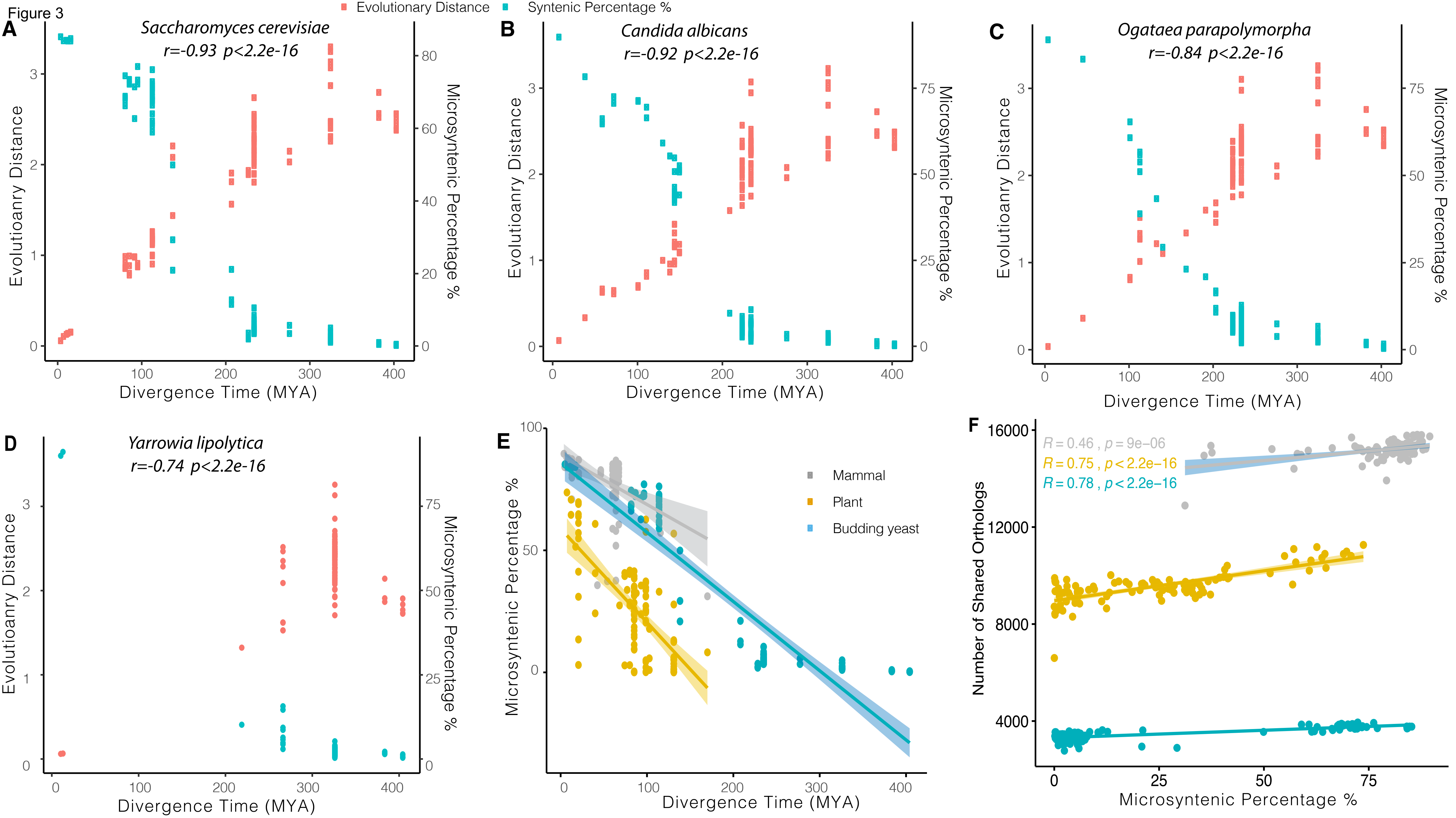
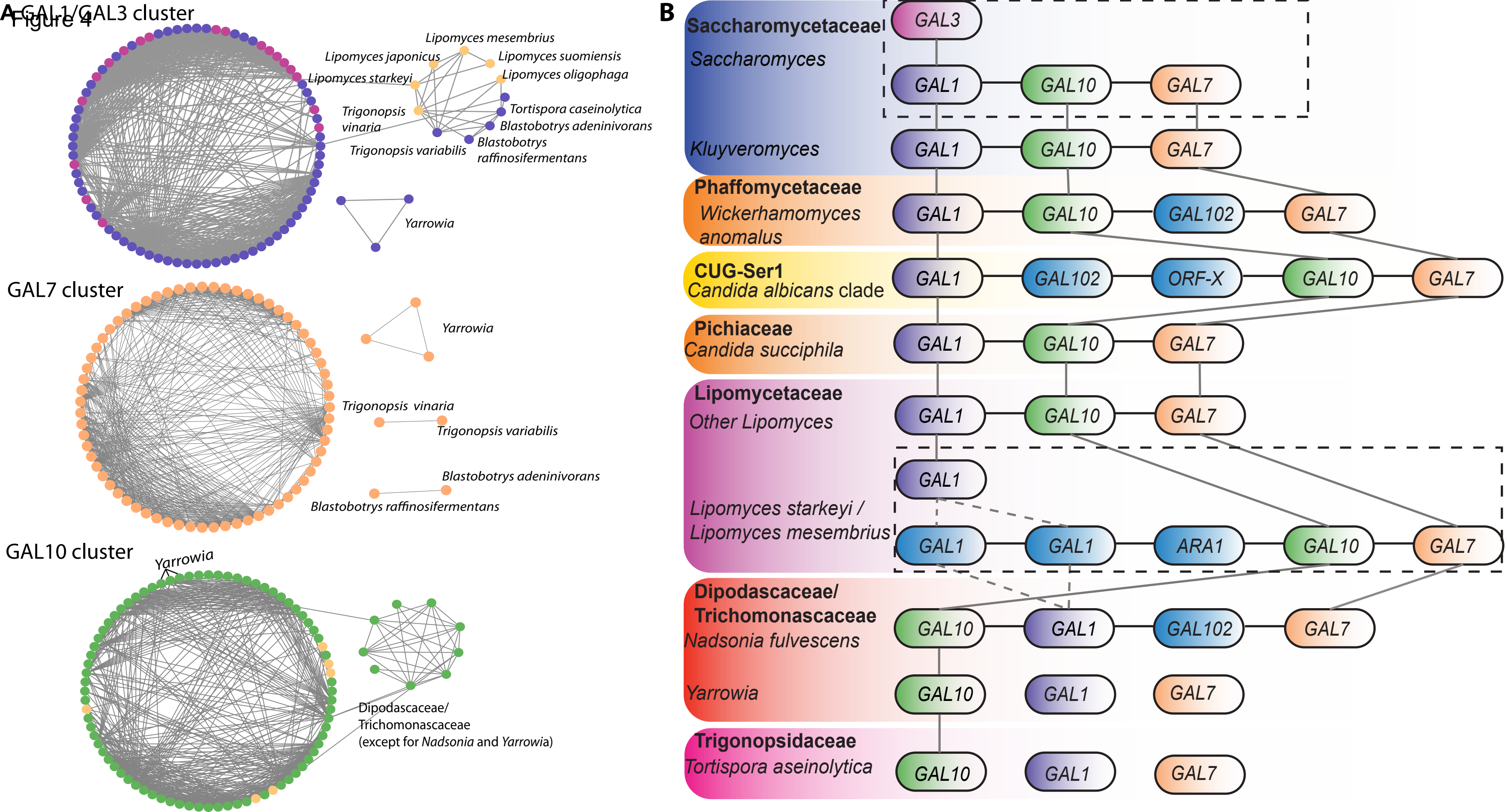

Evolutionary Distance
Syntenic Percentage %
A
Saccharomyces cerevisiae
r=-0.93 p<2.2e-16
B
Candida albicans
r=-0.92 p<2.2e-16
C
Ogataea parapolymorpha
r=-0.84 p<2.2e-16
D
Yarrowia lipolytica
r=-0.74 p<2.2e-16
Evolutioanry Distance
Microsyntenic Percentage %
Divergence Time (MYA)
0
100
200
300
400
0
1
2
3
0
20
40
60
80
0
25
50
75
E
Mammal
Plant
Budding yeast
Microsyntenic Percentage %
100
50
0
Divergence Time (MYA)
F
R = 0.46 , p = 9e−06
R = 0.75 , p < 2.2e−16
R = 0.78 , p < 2.2e−16
Number of Shared Orthologs
16000
12000
8000
4000
Microsyntenic Percentage %
0
25
50
75

Figure 4


A
GAL1/GAL3 cluster
Lipomyces mesembrius
Lipomyces japonicus
Lipomyces suomiensis
Lipomyces oligophaga
Lipomyces starkeyi
Trigonopsis vinaria
Tortispora caseinolytica
Blastobotrys adeninivorans
Blastobotrys raffinosifermentans
Trigonopsis variabilis
Yarrowia
GAL7 cluster
Yarrowia
Trigonopsis vinaria
Trigonopsis variabilis
Blastobotrys adeninivorans
Blastobotrys raffinosifermentans
GAL10 cluster
Yarrowia
Dipodascaceae/
Trichomonascaceae
(except for Nadsonia and Yarrowia)
B
Saccharomycetaceae
Saccharomyces
Kluyveromyces
Phaffomycetaceae
Wickerhamomyces anomalus
CUG-Ser1
Candida albicans clade
Pichiaceae
Candida succiphila
Lipomycetaceae
Other Lipomyces
Lipomyces starkeyi / Lipomyces mesembrius
Dipodascaceae/ Trichomonascaceae
Nadsonia fulvescens
Yarrowia
Trigonopsidaceae
Tortispora aseinolytica
GAL3
GAL1
GAL10
GAL7
GAL102
ORF-X
ARA1

Supplemental Figures

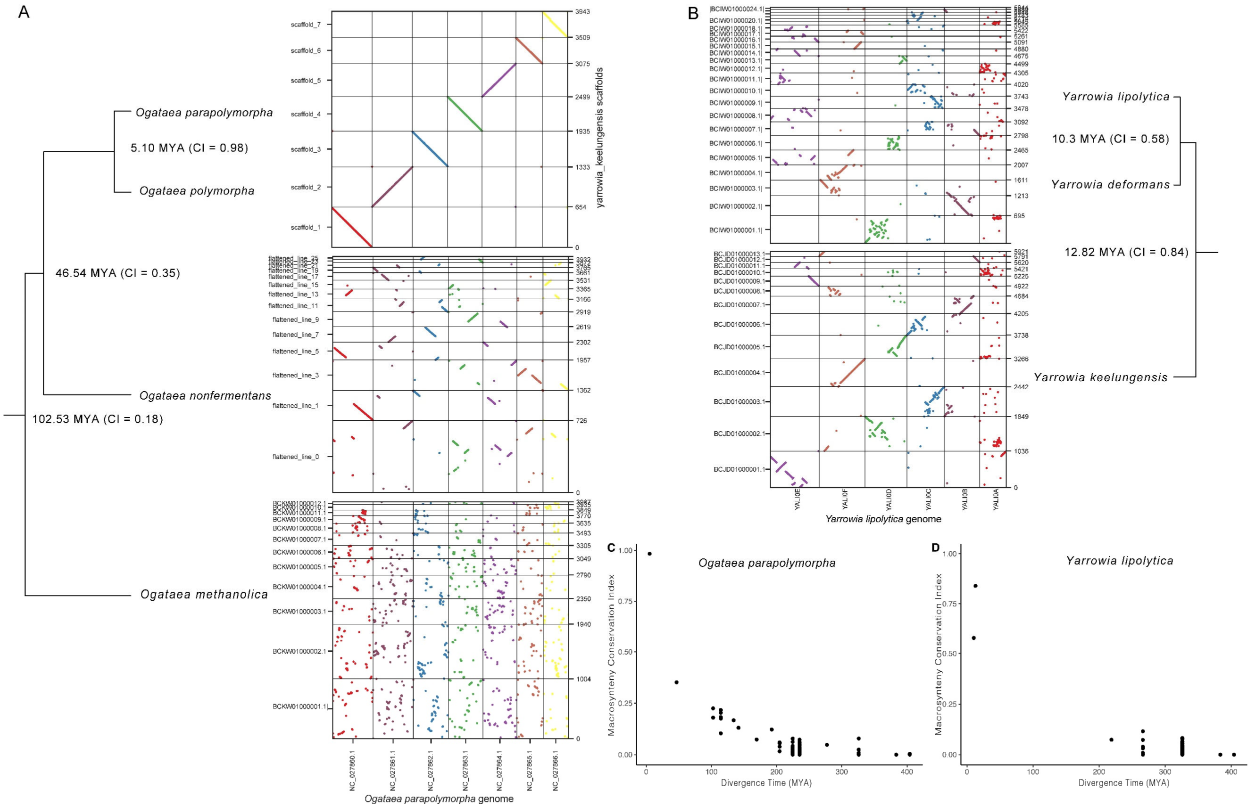


Figure S1. The decay of macrosynteny conservation between *Ogataea parapolymorpha*, *Yarrowia lipolytica*, and their close relatives in the budding yeast subphylum. Related to Figure 1.

(A) Oxford dot plots of homologous genes between *O. parapolymorpha* and three representative closely related species. The colored dots correspond to homologous genes from the chromosomes, with chromosome boundaries indicated and sorted based on chromosomal size. The time-calibrated species tree on the left was obtained from a previous study of 332 budding yeast species^{S1}. (B) An Oxford dot plot of homologous genes between *Y. lipolytica* and two representative closely related species. (C) Macrosynteny conservation index between *O. parapolymorpha* and all other 119 budding yeast genomes. (D) Macrosynteny conservation index between *Y. lipolytica* and all other 119 budding yeast genomes. Note the lack of conservation of macrosynteny after ~100 million years of divergence in both lineages.

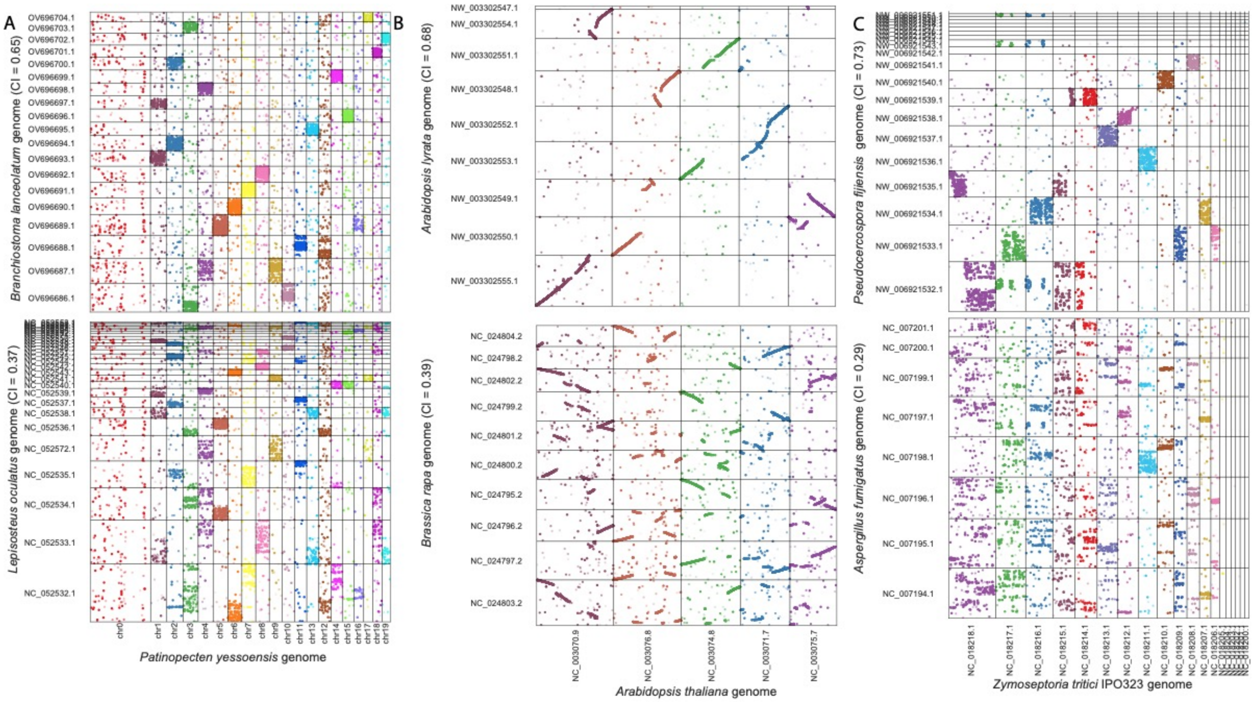


Figure S2. Macrosynteny conservation between representative species within bilaterian animals, angiosperm plants, and filamentous fungi. Related to Figure 1.

(A) Oxford dot plots of homologous genes between scallop (*Patinopecten yessoensis*) and two representative bilaterian species: amphioxus (*Branchiostoma floridae*) and spotted gar (*Lepisosteus oculatus*). (B) Oxford dot plots of homologous genes between *Arabidopsis thaliana* and two other flowering plants. (C) Oxford dot plots of homologous genes between *Zymoseptoria tritici* and two representative filamentous fungal species. The colored dots correspond to homologous genes from the chromosomes, with chromosome boundaries indicated and sorted based on chromosomal size.

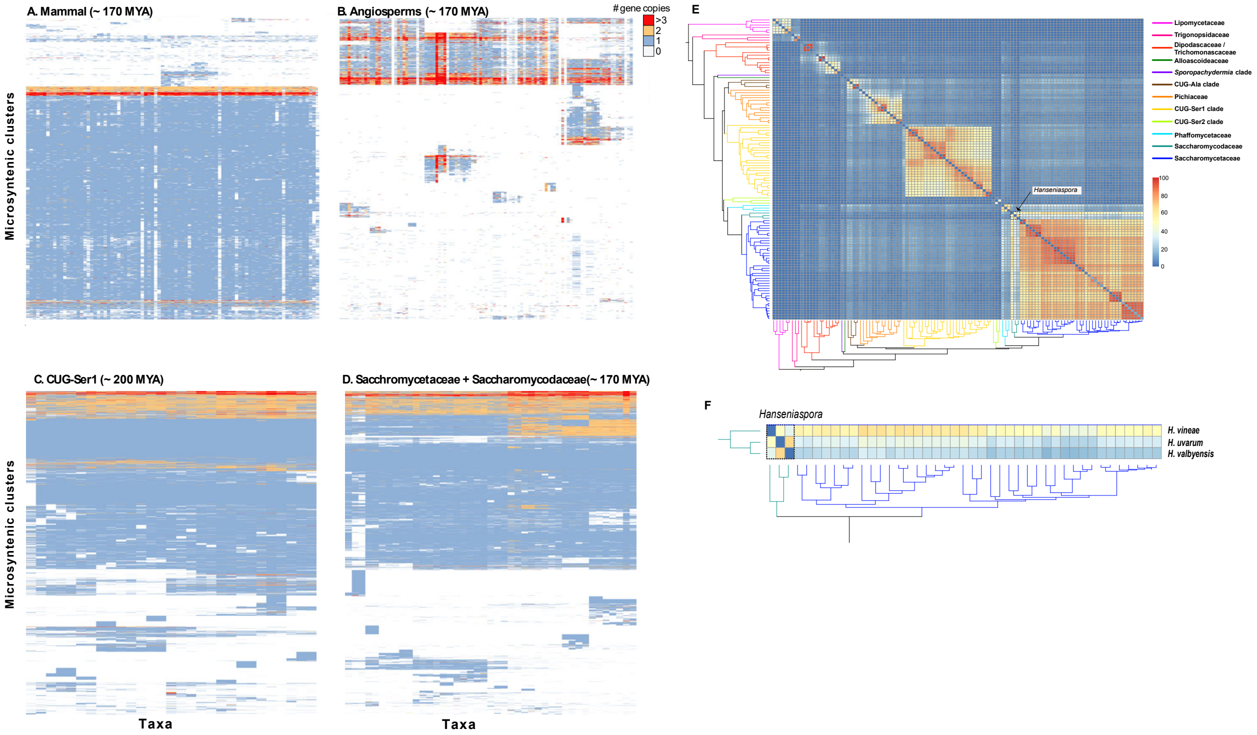


Figure S3. Microsynteny conservation of budding yeast, mammalian, and angiosperm genomes. Related to Figure 2.

(A) Microsynteny conservation of 87 mammalian genomes[S2]. (B) Microsynteny conservation of 107 angiosperm genomes[S2]. (C) Microsynteny conservation across 29 genomes of the CUG-Ser1 major clade of budding yeasts. (D) Microsynteny conservation across 43 genomes of clade including the Saccharomycetaceae + Saccharomycodaceae major clades of budding yeasts. The Y axis in panels A through D shows orthogroups with conserved microsynteny; different gene copy numbers of orthogroups with conserved microsynteny are labeled in different colors. The X axis corresponds to species within different clades. (E) Pairwise microsynteny comparisons of 120 budding yeast genomes. Species are arranged according to the species phylogeny derived from the previous study[S1]; the branches corresponding to different major clades of budding yeasts are shown in different colors. Overall, the percentage of syntenic homologous genes varied in a lineage-specific manner, with higher percentages observed in comparisons of closely related species. (F) The dotted box represents *Hanseniaspora*, a genus of budding yeasts previously shown to exhibit extensive variation in its evolutionary rates among its species. Note that the slow-evolving *H. vineae* exhibits higher microsynteny conservation percentages with genomes from the sister lineage Saccharomycetaceae than fast-evolving *Hanseniaspora* species.

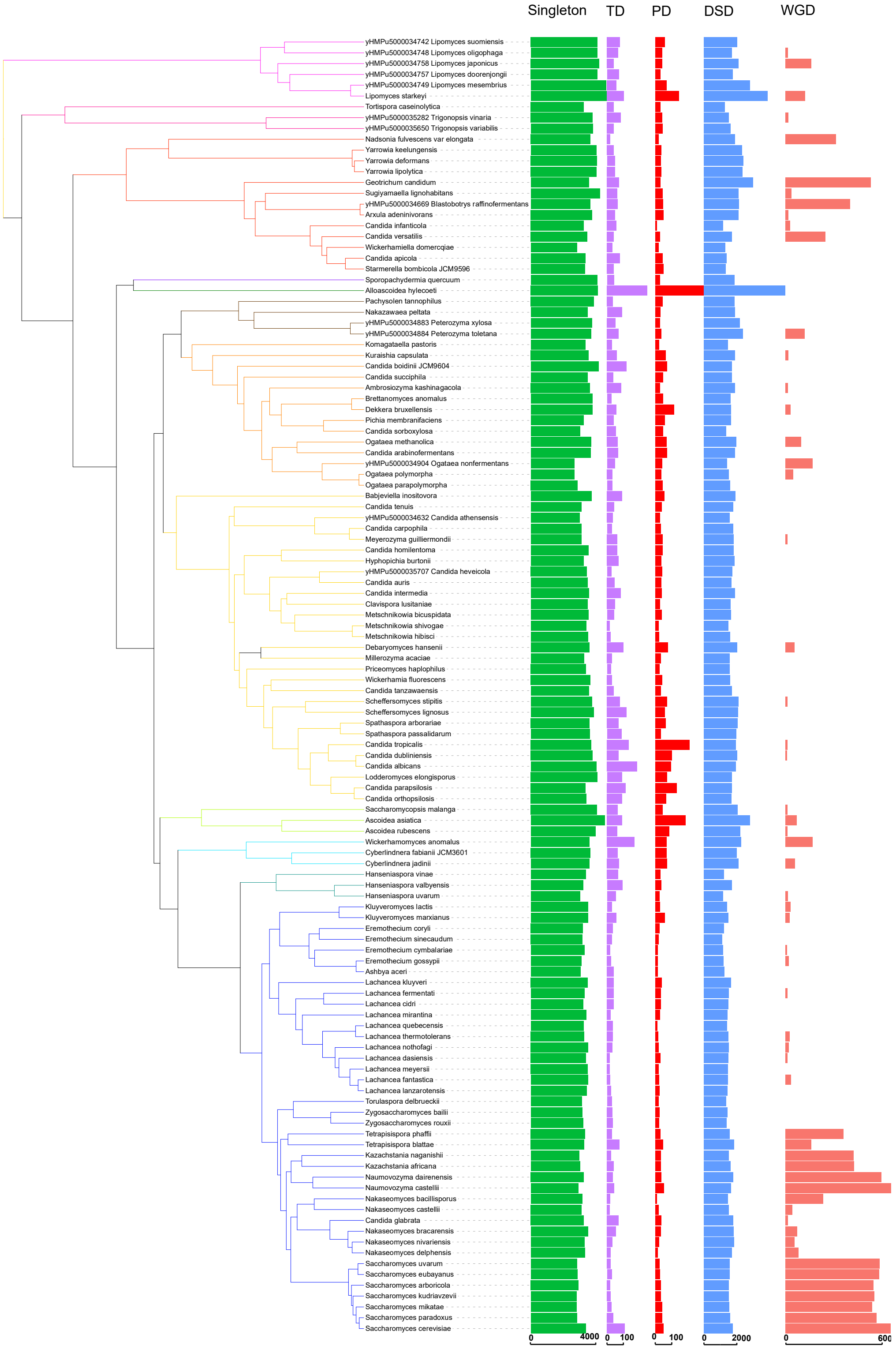


Figure S4. The number of genes predicted to be derived from different modes of duplication in 120 budding yeast genomes. Related to Figure 2.

WGD: whole-genome or segmental duplication, TD: tandem duplication, PD: proximal duplication, and DSD: dispersed duplication. Note that the scales of different duplication modes are different for illustration purposes.

Supplemental References

S1. Shen, X.-X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase, M.A.B., Wisecaver, J.H., Wang, M., Doering, D.T., et al. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell *175*, 1533–1545.e20.

S2. Zhao, T., and Schranz, M.E. (2019). Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. Proc. Natl. Acad. Sci. U. S. A. *116*, 2165–2174.