An orthologous gene coevolution network provides insight into eukaryotic cellular and genomic structure and function

Jacob L. Steenwyk[1], Megan A. Phillips[1], Feng Yang[2,3], Swapneeta S. Date[1], Todd R. Graham[1], Judith Berman[2], Chris Todd Hittinger[4], & Antonis Rokas[1,*]

[1] Vanderbilt University, Department of Biological Sciences, Nashville, TN, United States of America

[2] Shmunis School of Biomedical and Cancer Research, Tel Aviv University, Ramat Aviv, Israel

[3] Department of Pharmacology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China

[4] Laboratory of Genetics, DOE Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, Center for Genomic Science Innovation, J.F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

*Correspondence should be addressed to: antonis.rokas@vanderbilt.edu

Running title: Budding yeast genetic coevolution network

125-character teaser: The budding yeast coevolution network captures cellular structure and function in the absence of functional data.

Keywords: Evolutionary rate covariation, DNA replication, DNA repair, genetic network, chromatin remodeler, Golgi transport

Abstract

The evolutionary rates of functionally related genes often covary. We present a gene coevolution network inferred from examining nearly three million orthologous gene pairs from 332 budding yeast species spanning ~400 million years of evolution. Network modules provide insight into cellular and genomic structure and function. Examination of the phenotypic impact of network perturbation using deletion mutant data from the baker's yeast *Saccharomyces cerevisiae*, which were obtained from previously published studies, suggests that fitness in diverse environments is impacted by orthologous gene neighborhood and connectivity. Mapping the network onto the chromosomes of *S. cerevisiae* and *Candida albicans* revealed coevolving orthologous genes are not physically clustered in either species; rather, they are often located on different chromosomes or far apart on the same chromosome. The coevolution network captures the hierarchy of cellular structure and function, provides a roadmap for genotype-to-phenotype discovery, and portrays the genome as a linked ensemble of genes.

Introduction

Genetic networks—diagrams wherein nodes represent genes and edges represent measured functional relationships between nodes—can elucidate how genes are organized into pathways and contribute to cellular functions, shedding light onto the relationship between genotype and phenotype (*1–4*). Given the rich information contained in or derived from genetic networks, numerous approaches that aim to capture some aspect(s) of functional relationships among genes in a genome (e.g., gene coexpression, genetic interaction) have been developed (*5–7*). While these networks are highly informative, their availability and applicability is typically limited to select model organisms and single extant species or strains. Application of information from the genetic network of one organism to understand the biology of another requires assuming that the networks of the two organisms are conserved, which is not always the case (*8*, *9*, *18*, *10–17*).

One complementary, but poorly studied, method for constructing genetic networks is by measuring the coevolution of orthologous genes, which can be done by calculating the covariation of relative evolutionary rates among orthologous genes (*19–22*). Briefly, by estimating an orthologous gene's phylogeny, one infers the rate (and changes in rate) of its evolution across the phylogeny; if the evolutionary rate values estimated for each branch of an orthologous gene's phylogeny are significantly correlated with those of another gene's phylogeny, the two orthologs are said to be coevolving. Note, coevolution of orthologous genes is distinct from organismal coevolution in which reciprocal evolutionary changes occur between interacting lineages—for example, insect pollinators impacting flowering plant diversification (*23*, *24*). By estimating coevolution for all pairs of orthologous genes in a clade, one can infer the clade's orthologous gene coevolution network, where nodes correspond to orthologs and edges correspond to the degree to which two orthologs coevolve (*22*). Genetic networks based on gene coevolution leverage evolutionary information, whereas standard genetic networks rely on the correlation of functional data such as gene expression or the presence of genetic interactions among genes within a single extant species or strain.

Orthologous gene coevolution is often observed among genes that share functions, are coexpressed, or whose protein products are subunits in a multimeric protein structure, and can yield insights into the genotype-to-phenotype map (*25*, *26*). For example, screening for genes

that have coevolved with genes in known DNA repair pathways across 33 mammals led to the identification of *DDIAS*, whose involvement in DNA repair was subsequently functionally validated (*26*). Furthermore, among 918 pairs of interacting proteins in the protein structural interactome map, a database of structural domain-domain interactions in the protein data bank (https://www.rcsb.org/), four out of five proteins exhibit signatures of gene coevolution (*27*). Although these and other studies have demonstrated that signatures of coevolution are a powerful method to detect functional associations among genes in the absence of functional data (*20*, *25*, *26*, *28–30*), the network biology principles of gene coevolution, especially between genes that have coevolved for hundreds of millions of years, remain unexplored.

To unravel general principles of orthologous gene coevolutionary networks, we constructed the coevolution network of a densely sampled set of orthologs from one-third of known budding yeast species (332 species) that diversified over ~400 million years. The inferred network provides a hierarchical view of cellular function from broad bioprocesses to specific pathways. Interpolation of the gene coevolution network with of fitness assay data from single- and digenic *S. cerevisiae* mutants (*1*, *2*, *31*, *32*) provides insight into subnetwork- and ortholog-specific potential to buffer genetic perturbations. Surprisingly, comparisons of genetic networks inferred from gene coevolution and genetic interactions yield similar functional insights; for example, hubs of genes tend to be functionally related and gene essentiality impacts gene connectivity wherein essential genes are more densely connected than non-essential genes. Unlike genetic interaction networks, gene coevolution networks can also provide evolutionary insights; for example, mapping the orthologous gene coevolution network onto the chromosomes of two model yeast genomes uncovers extensive inter-chromosomal and long-range intra-chromosomal associations, providing an 'entangled' view of the genome across evolutionary timescales. We anticipate these results will facilitate the generation, interpretation, and utility of these networks among other lineages in the tree of life.

Results

A gene coevolution network

We examined 2,898,028 pairs of orthologous genes from a dataset of 2,408 orthologous genes in 332 budding yeast species. Broad network properties were stable across a range of thresholds for

"significant" orthologous gene coevolution (Fig. S2). To conservatively define "significant" coevolution and therefore examine orthologous gene pairs with only robust signatures of coevolution, we implemented a high correlation coefficient threshold for significant orthologous gene coevolution ($r \geq 0.825$; Pearson correlation among relative evolutionary rates). This resulted in 60,305 significant signatures of orthologous gene coevolution; Fig 1A, 1B, and S1), which were used to construct a network where nodes are orthologous genes and edges connect orthologous genes that are significantly coevolving (Fig. 1C).

To determine how orthologous gene connectivity varied in the network, we examined patterns of dense and sparse connections for individual orthologous genes. Individual orthologous genes coevolved with a median of eight other orthologous genes, but connectivity varied substantially across the network (Fig. S3). For example, 1,091 orthologous genes have signatures of coevolution with five or fewer other orthologous genes and 601 orthologous genes are singletons, which we define as orthologous genes that are not significantly coevolving with any other orthologous genes in the dataset. In contrast, 420 orthologous genes have signatures of coevolution with 100 or more other orthologous genes, and 21 orthologous genes coevolve with 400 or more others.

Coevolving orthologous genes in the network tend to be functionally related. For example, *PEX1* and *PEX6* are one of the pairs of genes with the highest observed correlation coefficient in evolutionary rates (Fig. S4). In *S. cerevisiae*, the two orthologous genes encode a heterohexameric complex responsible for protein transport across peroxisomal membranes (*33*) and mutations in either gene can lead to severe peroxisomal disorders in humans (*34*). Functional enrichment among densely connected orthologous genes revealed that complex bioprocesses that require coordination among polygenic protein products are overrepresented (Fig. S5, Table S1). For example, *CHD1*, *INO80*, and *ARP5*, which encode proteins responsible for chromatin remodelling processes such as nucleosome sliding and spacing (*35*), are coevolving with 400 or more other orthologous genes (Fig. S5, Table S1). Taken together, these findings highlight that coevolution may be observed among orthologous genes that physically interact (e.g., *PEX1* and *PEX6*) or contribute to highly intricate biological processes (e.g., *INO80*). More broadly, these data support the hypothesis that coevolving orthologous genes tend to have similar functions.

To determine how connectivity varied within the network, we examined the properties of subnetworks across orthologous genes considered essential and nonessential in the model yeast *S. cerevisiae* or the opportunistic pathogen *C. albicans* (*36*, *37*). Essential genes are densely connected in the orthologous gene coevolutionary network, whereas nonessential genes exhibit sparser connections (Fig. 2A-D). To infer network orthologous gene communities—clusters of orthologous genes that have more connections between them than between orthologous genes of different clusters—we used a hierarchical agglomeration algorithm (Fig. 2A). Five large orthologous gene communities (clusters of more than 10 orthologous genes) were identified. Each orthologous gene community varied in size, orthologous gene community-to-orthologous gene community connectivity, and essential/nonessential orthologous gene composition. Specifically, the two largest orthologous gene communities, communities 1 and 2, share the most connections and belong to a higher-order cluster with the next two largest orthologous gene communities, communities 3 and 4 (Fig. 2E and S6). In contrast, the smallest orthologous gene community, community 5, does not cluster with the other orthologous gene communities. Similarly, essential genes are overrepresented in orthologous gene community 1 but are underrepresented in orthologous gene communities 2, 3, and in smaller communities of 10 or fewer orthologous genes (Fig. 2F; $p < 0.01$ for all tests; Fisher's exact test). The result that *S. cerevisiae* and *C. albicans* essential genes are central hubs in coevolution network constructed from orthologous genes that represent 400 million years of budding yeast evolution mirrors the finding that essential genes are central hubs in the *S. cerevisiae* genetic interaction network (*2*).

From processes to pathways: the budding yeast coevolution network captures the hierarchy of cellular function

To gain insight into the functional neighborhoods of the orthologous gene coevolution network, we examined via gene ontology (GO) enrichment analysis (*38*) the composition of each orthologous gene community. Among the highest-order cluster of orthologous gene communities (i.e., communities 1 through 4), we found that higher-order cellular processes including nucleic acid metabolism ($p = 0.040$; Fisher's exact test multi-test corrected using false discovery rate correction with Benjamini/Hochberg (FDR-BH)) and cellular anatomical entities ($p = 0.020$; Fisher's exact test multi-test corrected using FDR-BH) are enriched. At the individual

orthologous gene community level, we found that orthologous gene community 1 is enriched in orthologous genes with helicase activity ($p = 0.005$; Fisher's exact test multi-test corrected using FDR-BH), ligase activity ($p = 0.004$; Fisher's exact test multi-test corrected using FDR-BH), and translation initiation factors ($p = 0.024$; Fisher's exact test multi-test corrected using FDR-BH); orthologous gene community 2 is enriched in Golgi vesicle transport orthologous genes ($p = 0.009$; Fisher's exact test multi-test corrected using FDR-BH); whereas singletons are enriched in GTPase activity ($p = 0.016$; Fisher's exact test multi-test corrected using FDR-BH) and peroxiredoxin activity ($p = 0.036$; Fisher's exact test multi-test corrected using FDR-BH) (Fig. 2G-I, Table S3).

Functional neighborhoods of coevolving orthologous genes within and between biological functions as well as cellular compartments and complex categories are also captured by the network. For example, orthologous genes involved in the biological functions of ribosome biogenesis, rRNA processing, and translation, which represent different functional categories, are extensively coevolving with one another (Fig. S7A). This finding suggests that the complexity of protein biosynthesis, a process that requires coordination among diverse biochemical functions, is captured in the coevolution of the underlying orthologous genes. Similarly, orthologous genes involved in nuclear processes or located in the cytoplasm tend to coevolve with orthologous genes in the same cellular compartment, however, substantial signatures of coevolution between orthologous genes from different cellular compartments are also observed (Fig. S7B).

Finally, our network captures functional neighborhoods of coevolving orthologous genes at the level of pathways and complexes. We found strong signatures of coevolution among orthologous genes from specific pathways and complexes. For example, orthologous genes that encode proteins responsible for DNA replication coevolve with a larger number of other DNA replication orthologous genes than expected by random chance ($p < 0.001$; permutation test) (Fig. S8). Orthologous genes involved in DNA mismatch repair and nucleotide excision repair pathways, which participate in the repair of DNA lesions, have more signatures of coevolution than expected by random chance ($p < 0.001$ for each pathway; permutation test). Orthologous genes in the phosphatidylcholine biosynthesis pathway, which is responsible for the biosynthesis of the major phospholipid in organelle membranes, and orthologous genes in the tricarboxylic

acid cycle (also known as the Krebs cycle or citric acid cycle), a key component of aerobic respiration (Fig. S9), also have more signatures of coevolution than expected by random chance ($p < 0.001$ for each pathway; permutation test). Among complexes, orthologous genes that encode the minichromosome maintenance protein complex that functions as a DNA helicase, the DNA polymerase α-primase complex that assembles RNA-DNA primers required for replication, and DNA polymerase ε that serves as a leading strand DNA polymerase (Fig. 3) also coevolve with larger numbers of orthologs from the same complex than expected by random chance ($p < 0.001$ for each multimeric complex; permutation test). Note, certain gene categories (e.g., transposons and hexose transporters) are not represented in our dataset of orthologous genes and could not be examined (see *Methods*).

In summary, these findings reveal that functional aspects of the network can be viewed with varying degrees of specificity. For example, the highest-order insights (i.e., GO enrichment across orthologous gene communities 1, 2, 3, and 4) revealed coevolution among cellular anatomical entities whereas greater specificity—such as coevolution among orthologous genes responsible for Golgi vesicle transport—can be obtained by examining lower-order hubs of genes (e.g., GO enrichment in orthologous gene community 2). Furthermore, coevolutionary signatures can bridge distinct but related functional categories such as cellular compartments and complexes, highlighting the complex interplay of distinct functional modules over evolutionary time. Thus, the budding yeast coevolution network captures the hierarchy of cellular function from broad bioprocesses to specific pathways or multimeric complexes.

The coevolution network constructed from budding yeast orthologous genes is distinct, but complementary, to the *S. cerevisiae* genetic interaction network

To determine similarities and differences between our coevolution network inferred from orthologous genes in the budding yeast subphylum and the genetic interaction network inferred from digenic null mutants in the model organism *S. cerevisiae* (*1*, *31*), both data types were integrated into a single supernetwork (Fig. S10 and S11). In the genetic interaction network, nodes represent genes and edges represent non-additive genetic interactions between genes; in the supernetwork, nodes represent genes and edges connect two genes that have a significant signature of coevolution, of genetic interaction, or both. We hypothesize that there will be broad

similarities between the networks because they both capture functional associations; however, we also hypothesize that the connectivity of individual nodes between the networks will sometimes differ because one network is built from ~400 million years of orthologous gene coevolution whereas the other from genetic interactions in a single extant species.

Supporting this hypothesis, the orthologous gene community clustering observed in the gene coevolution network was also evident in the supernetwork and the two networks were found to be more similar for all metrics examined (i.e., mean distance and transitivity) than expected by random chance ($p < 0.001$ for both tests; permutation test); however, gene- / ortholog-wise connectivity at times differed suggesting each network harbors distinct and complementary insights (Fig. S10). For example, connectivity is similar for the gene / ortholog *CDC6*, which is required for DNA replication (*39*), between the two networks. Specifically, *CDC6* is connected to 96 genes / orthologs in both networks and 56 of the genes / orthologs are the same. This result suggests that the connectivity of the *CDC6* gene in *S. cerevisiae* is broadly conserved across species from the budding yeast subphylum. In contrast, different gene- / ortholog-wise connectivity was observed for the choline kinase *CKI1* (*40*, *41*); *CKI1* is coevolving with 87 orthologs, has a significant genetic interaction with 10 genes, and seven of these genes / orthologs are shared by both networks. This result suggests that the connectivity of the *CKI1* gene observed in *S. cerevisiae* is not broadly conserved across species from the budding yeast subphylum. This difference may be partially explained by the fact that *CKI1* has a paralog, *EKI1*, which arose from an ancient whole genome duplication event that affected some, but not all, species in the subphylum (*42*, *43*). These results reveal that orthologous gene coevolution networks inferred over macroevolutionary timescales and networks inferred from genetic interactions in single organisms offer complementary insights into functional relationships between genes.

Orthologous gene communities differ in capacity to compensate for perturbation

Examinations of gene dispensability in the model budding yeast *S. cerevisiae* and the opportunistic pathogen *Candida albicans* (*36*, *37*) suggest that single-organism genetic networks can buffer single gene losses as evidenced by the ability to maintain organismal viability. Thus, we sought to determine whether a gene's dispensability varies in an orthologous gene

community-dependent manner. To address this, we integrated information from the budding yeast orthologous gene coevolution network and genome-wide single-gene deletion fitness assays (or, in the case of essential genes, expression suppression) of *S. cerevisiae* in 14 diverse environments (*32*) (Fig. S12 and S13). Here, single-gene deletion fitness assays serve as a proxy for network perturbation in which deletion of a single gene is analogous to removing a node from the network. We found that fitness of *S. cerevisiae* gene knockouts in different environments was significantly dependent on orthologous gene community and the number of coevolving genes per gene (Fig. 4; $p < 0.001$ for both comparisons of an interaction between orthologous gene community:environment interaction and environment:number of coevolving genes, Multi-factor ANOVA). We also observed a significant fixed effect for orthologous gene community and environment ($p < 0.001$, Multi-factor ANOVA). These observations highlight the importance and role of the environment and the architecture of the underlying genetic network when evaluating the consequences of single-gene deletions on organismal fitness.

To further investigate the relationship between *S. cerevisiae* gene dispensability and structure of the coevolution network, we integrated *S. cerevisiae* genetic interaction data from double-gene or digenic deletion fitness assays, wherein positive and negative genetic interactions refer to positive and negative fitness effects in the digenic deletion mutants relative to those expected from the combined effects of the individual single-gene deletion mutants, respectively (*1*, *2*, *31*). We found that most gene pairs were associated with negative genetic interactions (Fig. S14). Furthermore, genetic interactions scores among different orthologous gene community combinations were not significantly different (p-value > 0.05; Kruskal-Wallis rank sum test) suggesting digenic losses negatively impacted fitness irrespective of orthologous gene community.

Finally, to examine evolutionary gene loss in the context of the gene coevolution network, we investigated orthologous gene community-wide patterns of gene losses among genes lost in a lineage of budding yeasts previously reported to have undergone extensive gene losses (*44*). These analyses revealed orthologous gene community 2 and singleton orthologs are more likely to be lost (Fig. S14B), which supports the hypothesis that gene losses do not occur stochastically

(*45*). In summary, the architecture of the coevolution network is significantly associated with a gene's dispensability.

An entangled genome: extensive inter- and long-range intra-chromosomal coevolution

Gene order is not random among eukaryotes and physically linked genes tend to be involved in the same metabolic pathway or protein-protein complex (*46*, *47*). Thus, we hypothesized that coevolving orthologous genes will likely be physically linked or clustered onto yeast chromosomes. To test this hypothesis, we projected the budding yeast gene coevolution network onto the one-dimensional genome structure of *S. cerevisiae* and *C. albicans*, which diverged ~235 million years ago (*48*). We chose the genomes of these two organisms because they both have complete and high-quality chromosome-level assemblies. The two organisms also have distinct evolutionary histories; the lineage that includes *S. cerevisiae* underwent whole-genome duplication, whereas *C. albicans* underwent intra-species hybridization (*42*, *49*). These processes have contributed to differences in chromosome number (16 in *S. cerevisiae* vs. eight in *C. albicans*) and a lack of macrosynteny (*50–54*) (Fig. 5A-B and Fig. S15-S16).

Contrary to our hypothesis, we observed extensive inter-chromosomal and long-range intra-chromosomal orthologous gene coevolution (Fig. 5 and Fig. S17-S23). Specifically, co-evolving orthologous gene pairs were commonly located on different chromosomes (Fig. 5C-D and Table S4). There was a near-perfect correlation between the number of intra-chromosomal signatures of coevolution (corrected by the number of genes on that chromosome in the dataset) and the number of inter-chromosomal signatures of coevolution (corrected by the number of genes on all other chromosomes in the dataset) ($r = 0.95$, $p < 0.001$ for *S. cerevisiae*; $r = 0.98$, $p < 0.001$ for *C. albicans*; Spearman correlation). This result suggests that orthologous genes located on the same or different chromosomes are equally like to be coevolving. Given the extensive coevolution among orthologous genes in the same or similar functional categories, these results support the notion that function, not chromosome structure, is the primary driver of coevolution over macroevolutionary timescales.

Examination of intra-chromosomal coevolution revealed variation in orthologous gene pair distances along the genome. Two coevolving orthologous genes on the same chromosome can be

kilobase-to-megabase distances from one another (Fig. 5G-H). The distribution of the closest distance between an orthologous gene and its coevolving partners revealed a positively skewed distribution with a similar range of kilobase-to-megabase associations (Fig. S23). In *S. cerevisiae*, the number of intra-chromosomal signatures of coevolution is correlated with the number of genes on a chromosome represented in the dataset, whereas in *C. albicans* the number of intra-chromosomal signatures of coevolution is correlated both with chromosome length and with the number of genes on a chromosome represented in the dataset (Fig. S24). Examination of the distances between orthologous genes in our dataset and their coevolving partners revealed that long-range intra-chromosomal coevolution was not an artifact of gene sampling (Fig. S24). Investigation of the interplay between orthologous gene coevolution and chromosomal contacts using a three-dimensional model of the *S. cerevisiae* genome (*55*) revealed signatures of coevolution occur independent of chromosomal contacts (Fig. S26).

Extensive inter- and intra-chromosomal associations are exemplified by *INO80*, which encodes a chromatin remodeler and has coevolved with 591 orthologous genes on all other chromosomes in both *S. cerevisiae* and *C. albicans* (Fig. 5I-J). To date, few examples of inter-chromosomal associations between loci are known. One example includes concerted copy number variation between 45S and 5S rDNA loci in humans; imbalance in copy number is thought to be associated with disease (*56*, *57*). Our observations suggest extensive inter-chromosomal and long-range intra-chromosomal functional associations may be more common than previously appreciated.

Discussion

We constructed a genetic network based on orthologous gene coevolution from a densely sampled set of orthologs across the budding yeast subphylum. These analyses are distinct from genetic interaction- and gene expression-based genetic networks in that they leverage evolutionary, rather than functional, data. Thus, coevolution networks infer functionally conserved relationships among orthologous genes across entire lineages, whereas genetic networks infer functional relationships among genes in a single extant species or strain (irrespective of whether these relationships are conserved in other species or not). Gene coevolution networks are also distinct from networks constructed from correlated presence and absence patterns of orthologs across a lineage (an approach known as phylogenetic profiling (*58*,

59)) in that coevolutionary networks depict relationships among orthologs conserved in the majority of taxa. Examination of the global coevolution network, orthologous gene communities therein, and signatures of orthologous gene coevolution among bioprocesses, complexes, and pathways reveals that the network reflects the hierarchy of cellular function. Moreover, the integration of network-based approaches provides new insights into coevolution among orthologous genes—for example, orthologous genes coevolving with hundreds of other orthologous genes, such as *INO80* (Figure 5I and 5J), are enriched in nucleosome mobilization (Figure S5).

Comparison of the budding yeast coevolution network to the genetic interaction-based network of *S. cerevisiae* revealed numerous notable similarities and differences. For example, both methods found that gene essentiality significantly impacts connectivity wherein essential genes / orthologous genes are more densely connected than nonessential genes / orthologous genes (Fig. 2). This finding suggests that genes with more essential cellular functions are more likely central hubs in the coevolution network (*1*, *2*, *5*, *32*, *60*). Similarities were also observed among genes with broadly conserved functions. For example, the majority of genes / orthologs connected to *CDC6*, a gene required for the fundamental and widely conserved process of DNA replication (*39*), in the orthologous gene coevolution network and the genetic interaction-based network were the same (*1*, *31*).

Similarities between genetic interaction and gene coevolution networks were also observed when examining the impact of gene deletion(s) on fitness in diverse environments. For example, integrating fitness data with data from the orthologous gene coevolution network revealed significant interactions between community and environment, environment and the number of coevolving genes, as well as fixed effects of community and environment (Figure 4). These results suggest that phenotype can be affected by genes coevolving with other genes and the environment—a finding that, to our knowledge, represents the first integration of orthologous gene coevolution information and cellular fitness across diverse environments. A similar observation was made in the genetic interaction network wherein phenotype was affected by genes interacting with other genes and the environment, a phenomenon known as differential genetic interaction (*32*). Taken together with insights discussed in the previous paragraph, these

striking similarities suggest that, despite using different data types to infer genetic interaction networks and gene coevolutionary networks (i.e., functional and evolutionary data, respectively), functional associations between genes, even those affected by environmental contexts, can be encoded in their coevolutionary histories; thus, functional insights can be inferred from gene coevolution networks. We find this observation particularly exciting because compared to genetic interaction analysis, which requires generating and phenotyping single and digenic knockouts for all pairwise gene combinations, orthologous gene coevolution analysis is potentially far less challenging technically and requires fewer resources. Notwithstanding these benefits, orthologous gene coevolution analysis does require the availability of well-annotated genome sequences of multiple species and knowledge of orthology relationships of their genes. Nonetheless, in the absence of physical interaction and genetic interaction data, co-evolution networks can provide similar insights into functional relationships among genes.

In contrast, differences between the two networks are likely driven by the fact that not all parts of the genetic interaction-based network of any single organism are conserved across an entire lineage (*8*, *9*, *18*, *10–17*). The more distinct the evolutionary histories of genes or pathways of species used to construct an orthologous gene coevolution network, the more divergent the topologies of the genetic interaction-based network of a species in that lineage will be from the coevolution network of the entire lineage. For example, *CKI1*, a choline kinase, gene connectivity substantially differed in the two networks. This may be in part driven by an ancient whole genome duplication event and retention of the duplicate gene copy in some, but not all, budding yeast species (*42*, *43*). Taken together, these results indicate that similarities and differences between networks inferred using orthologous gene coevolution from a lineage and networks inferred based on genetic interactions from a single organism are driven by divergence in individual organisms' genetic networks; thus, these methods offer distinct insights into functional associations among genes.

Another difference between the two networks is that the budding yeast coevolution network offers novel evolutionary insights, which cannot be inferred from genetic interaction networks in a single species. For example, hubs of genes do not only represent functionally related genes but also genes whose function has been maintained across long evolutionary timescales.

Furthermore, interpolation of the gene coevolution network and one-dimensional and three-dimensional chromosome structure offers novel insights into the interplay of chromosome structure and coevolution. Despite there being few known examples of inter-chromosomal gene associations (*56*), we find extensive signatures of inter- and long-range intra-chromosomal coevolution (Fig. 5, S21-S22), which suggests that gene function, not location, drives orthologous gene coevolution over macroevolutionary timescales. These results uncover a previously underappreciated degree of genome-wide coevolution that has been maintained over millions of years of budding yeast evolution, suggesting that the evolution and function of eukaryotic genomes is best viewed as extensively linked ensembles of genes.

The analyses presented herein enabled us to synthesize information from orthologous gene coevolution, genetic interactions, and cellular fitness among digenic knockout strains in a diverse panel of environments. Importantly, this data-rich case study of orthologous gene coevolution can be thought of as a proof-of-principle report that sets the stage for numerous exciting research opportunities and questions—such as comparisons of orthologous gene coevolutionary networks between lineages that exhibit key evolutionary differences. For example, in budding yeasts, such comparisons of orthologous gene coevolutionary networks could be performed for lineages that differ in their evolutionary rates (*44*), levels of horizontally acquired genes (*48*, *61*, *62*), genetic code (*63*, *64*), whole-genome duplication (*43*), or ecological niche (*65*). This approach may also be particularly powerful in lineages where genetically tractable models have yet to be established or in emerging model organisms that are ripe for functional examination.

In summary, we highlight complementary and novel insights that can be inferred using coevolutionary networks compared to other methods to infer genetic networks. Insights and methods used herein will facilitate the generation, interpretation, and utility of these networks for other lineages in the tree of life.

Methods

Inferring gene coevolution

To infer gene coevolution across ~400 million years of budding yeast evolution, we first obtained 2,408 orthologous sets of genes (hereafter referred to OGs) from 332 species (*48*). These 2,408 orthologous genes are from diverse GO bioprocesses but are underrepresented for gene functions known to be present in multiple copies, such as transposons and hexose transporters (Table S5). Thus, we conclude that the 2,408 orthologous sets of genes span a broad range of cellular and molecular functions. Examination of over and underrepresentation of genes from the various chromosomes of *S. cerevisiae* and *C. albicans* revealed no chromosome was over or underrepresented in the 2,408 orthologs (Table S6), suggesting each chromosome is equally represented in our dataset.

Next, we calculated covariation of relative evolutionary rates of all 2,898,028 pairs from the 2,408 orthologous sets of genes. To do so, we developed the CovER (Covarying Evolutionary Rates) pipeline for high-throughput genome-scale analyses of orthologous gene covariation based on the mirror tree principle (Fig. 1). The mirror tree principle is conceptually similar to phylogenetic profiling—wherein correlations in gene presence/absence patterns across a phylogeny are used to identify functionally related genes (*66*)—but instead uses correlations in orthologous genes' relative evolutionary rates (*20*, *67*, *68*).

To implement the CovER pipeline, single gene trees constrained to the species topology were first inferred using IQ-TREE, v1.6.11 (*69*) (Fig. 1). Thereafter, all pairwise combinations of gene trees were examined for significant signatures of coevolution (Fig. 1B). Differences in taxon occupancy between gene trees are accounted for by pruning both phylogenies to the set of maximally shared taxa. To mitigate the influence of factors that can lead to high false positive rates, such as time since speciation and mutation rate, and increase the statistical power of calculating gene coevolution, branch lengths were transformed into relative rates by correcting the gene tree branch length by the corresponding branch length in the species phylogeny (*19*, *20*, *70*). Single data point outliers (defined as having corrected branch lengths greater than five) are known to cause false positive correlations and were removed (*20*). Branch lengths were then Z-transformed and a Pearson correlation coefficient was calculated for each pair of orthologs. The

CovER algorithm has been integrated into PhyKIT, a UNIX toolkit for phylogenomic analysis (*22*).

Network construction

Complex interactions between orthologous gene pairs were further examined using a network wherein nodes represent orthologs and edges connect orthologs that are coevolving. Following our previous work (*22*), we considered orthologous gene pairs with a covariation coefficient of 0.825 or greater to have a significant signature of coevolution. This threshold resulted in 60,305 / 2,898,026 (2.08%) significant signatures of coevolution (Fig. S1). To explore the impact of our choice of a covariation coefficient threshold, we examined two measures that describe how densely the network is connected: edge density (the proportion of present edges out of all possible edges) and transitivity (ratio of triangles that are connected to triples); as well as two measures that describe how diffuse the network is: mean distance (average path length among pairs of nodes) and diameter (the longest geodesic distance). Across a wide range of thresholds of significant orthologous gene coevolution (Pearson correlation coefficient range of [0.600-0.900] with a step of 0.005), we found that the choice of threshold had little impact on network structure (Fig. S2).

Network substructure is commonly referred to as orthologous gene community structure and describes a set of orthologs that are more densely connected with each other but more sparsely connected with other sets (or orthologous gene communities) of orthologs. To identify the orthologous gene community structure of our global orthologous gene coevolution network, a hierarchical agglomeration algorithm that conducts greedy optimization of modularity was implemented (*71*).

To determine if the orthologous gene coevolutionary network and genetic interaction network were more similar than expected by random chance, we conducted a permutation test. To do so, we generated a null expectation of similarity between the orthologous gene coevolutionary network and 10,000 random networks. Random networks were generated by shuffling the edges of the genetic interaction network. In this way, the edge density (the ratio of the number of edges and the number of possible edges) is the same between the randomly generated network and the

genetic interaction network. This is a more conservative than a completely random (null) network that also alters edge density. Next, we took the absolute values of the differences between the descriptive statistics of the orthologous coevolutionary network and the 10,000 random networks to generate the null distribution. The absolute difference between the descriptive statistics of the orthologous coevolutionary network and the observed genetic interaction network were then examined along the null distribution to determine a p-value.

Enrichment analysis

To determine functional category enrichment among sets of orthologs, gene ontology (GO) enrichment analysis was conducted. To do so, a background set of GO annotations were curated from the 2,408 orthologous genes (*48*). Specifically, for an orthologous group of genes, GO associations were mapped from the representative gene from *S. cerevisiae* (*72*). If an *S. cerevisiae* gene was not present, the annotation from the representative gene from *C. albicans* was chosen (*73*). When neither species was represented in an orthologous group, we considered the function of the orthologous group to be uncertain and did not assign a GO term. Significance in functional enrichment was assessed using a Fischer's exact test with Benjamini Hochberg multi-test correction ($\alpha = 0.05$) using goatools, v1.0.11 (*74*). GO annotations were obtained from the Gene Ontology Consortium (http://geneontology.org/; release date: 2020-10-09). Higher-order summaries of GO term lists were constructed using GO slim annotations and REVIGO (*75*). Over and underrepresentation of essential genes across orthologous gene communities and genes on the various chromosomes were examined using the same approach in R, v4.0.2 (https://cran.r-project.org/).

Pathway analysis

To examine coevolution between genes in pathways, we first determined the genes belonging to pathways of interest. To do so, we leveraged pathway information in the KEGG database (*76*) and the *Saccharomyces* Genome Database (SGD; https://www.yeastgenome.org/). To determine if there are more signatures of coevolution within a pathway than expected by random chance, we conducted permutation tests. The null distribution was generated by randomly shuffling coevolution coefficients across all ~3 million orthologous gene pairs 10,000 times and then

determining the number of coevolving pairs among the pairs of the pathway of interest for each iteration.

Integrating gene loss information

To estimate the impact of network perturbation, fitness of single-gene deletions and genetic interaction scores inferred from digenic deletions from were combined with information from the orthologous gene coevolution network (*1*, *2*, *31*, *32*). For example, the relationship between gene-/ortholog-wise community, connectivity, and fitness in diverse environments was evaluated. To determine if genes / orthologs were equally likely to be lost across orthologous gene communities, we examined patterns of gene losses in *Hanseniaspora* spp., which have undergone extensive gene loss compared to other budding yeasts (*44*).

Projecting the network onto genome structure and organization

To gain insight into the relationship between genome structure and the orthologous gene coevolution network, we projected the network onto the complete chromosome genome assemblies of *S. cerevisiae* and *C. albicans* (*72*, *73*, *77*, *78*). Prior to mapping the network onto the genome assemblies, we investigated genome-wide synteny using orthology information from the Candida Gene Order Browser (*50*). Thereafter, the network was projected onto each genome assembly using Circos, v0.69 (*79*). Examination of the distance between coevolving orthologous genes and chromosomal contacts was conducted using a three-dimensional model of the *S. cerevisiae* genome (*55*).

Funding

JLS and AR were supported by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program. TRG was supported by NIH (1R01GM118452). MAP was partially supported by the Vanderbilt Undergraduate Summer Research Program and the Goldberg Family Immersion Fund. AR's laboratory received additional support from the Burroughs Wellcome Fund, the National Science Foundation (DEB-1442113 and DEB-2110404), and the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R56 AI146096 and R01 AI153356). This material is based upon work supported by the National Science Foundation under Grant No. DEB-1442148, in part by the

DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409), and the USDA National Institute of Food and Agriculture (Hatch Project 1020204). C.T.H. is a Pew Scholar in the Biomedical Sciences and an H. I. Romnes Faculty Fellow, supported by the Pew Charitable Trusts and Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation, respectively. JB was supported by the European Research Foundation Synergy Fungal Tolerance 951475.

Data Availability

To facilitate other researchers to explore the gene coevolution information, we created a web application, *the budding yeast coevolution network* (https://github.com/JLSteenwyk/budding_yeast_coevolution_network), written in the R programming language (https://cran.r-project.org/). All other supplementary information including single gene phylogenies used to examine coevolution and Pearson covariation coefficients among relative evolutionary rates for all pairwise combinations of orthologous groups of genes will be available on figshare upon publication (doi: 10.6084/m9.figshare.14501964). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Competing Interests

AR is a scientific consultant for LifeMine Therapeutics, Inc. JLS is a scientific consultant for Latch AI, Inc. The authors declare no other competing interests.

Author Contributions

JLS and AR designed the research; JLS, MAP, FY, and SSD performed analyses; JLS prepared the figures with input from JB and AR; JLS and AR wrote the paper; all authors contributed to the interpretation of the results and provided comments and input on the figures and manuscript.

References

1. M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L. Y. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R. P. St. Onge, B. VanderSluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z.-Y. Lin, W. Liang, M. Marback, J. Paw, B.-J. San Luis, E. Shuteriqi, A. H. Y. Tong, N. van Dyk, I. M. Wallace, J. A. Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pal, F. P. Roth, G. Giaever, C. Nislow, O. G. Troyanskaya, H. Bussey, G. D. Bader, A.-C. Gingras, Q. D. Morris, P. M. Kim, C. A. Kaiser, C. L. Myers, B. J. Andrews, C. Boone, The Genetic Landscape of a Cell. *Science (80-.).* **327**, 425–431 (2010).
2. M. Costanzo, B. VanderSluis, E. N. Koch, A. Baryshnikova, C. Pons, G. Tan, W. Wang, M. Usaj, J. Hanchard, S. D. Lee, V. Pelechano, E. B. Styles, M. Billmann, J. van Leeuwen, N. van Dyk, Z.-Y. Lin, E. Kuzmin, J. Nelson, J. S. Piotrowski, T. Srikumar, S. Bahr, Y. Chen, R. Deshpande, C. F. Kurat, S. C. Li, Z. Li, M. M. Usaj, H. Okada, N. Pascoe, B.-J. San Luis, S. Sharifpoor, E. Shuteriqi, S. W. Simpkins, J. Snider, H. G. Suresh, Y. Tan, H. Zhu, N. Malod-Dognin, V. Janjic, N. Przulj, O. G. Troyanskaya, I. Stagljar, T. Xia, Y. Ohya, A.-C. Gingras, B. Raught, M. Boutros, L. M. Steinmetz, C. L. Moore, A. P. Rosebrock, A. A. Caudy, C. L. Myers, B. Andrews, C. Boone, A global genetic interaction network maps a wiring diagram of cellular function. *Science (80-.).* **353**, aaf1420–aaf1420 (2016).
3. E. Kuzmin, B. VanderSluis, W. Wang, G. Tan, R. Deshpande, Y. Chen, M. Usaj, A. Balint, M. Mattiazzi Usaj, J. van Leeuwen, E. N. Koch, C. Pons, A. J. Dagilis, M. Pryszlak, J. Z. Y. Wang, J. Hanchard, M. Riggi, K. Xu, H. Heydari, B.-J. San Luis, E. Shuteriqi, H. Zhu, N. Van Dyk, S. Sharifpoor, M. Costanzo, R. Loewith, A. Caudy, D. Bolnick, G. W. Brown, B. J. Andrews, C. Boone, C. L. Myers, Systematic analysis of complex genetic interactions. *Science (80-.).* **360**, eaao1729 (2018).
4. M. Costanzo, E. Kuzmin, J. van Leeuwen, B. Mair, J. Moffat, C. Boone, B. Andrews, Global Genetic Networks and the Genotype-to-Phenotype Relationship. *Cell*. **177**, 85–100 (2019).
5. J. H. Wisecaver, A. T. Borowsky, V. Tzin, G. Jander, D. J. Kliebenstein, A. Rokas, A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic

Pathways in Plants. *Plant Cell*. **29**, 944–959 (2017).

6. A. Baryshnikova, M. Costanzo, C. L. Myers, B. Andrews, C. Boone, Genetic Interaction Networks: Toward an Understanding of Heritability. *Annu. Rev. Genomics Hum. Genet.* **14**, 111–133 (2013).

7. T. R. Lezon, J. R. Banavar, M. Cieplak, A. Maritan, N. V. Fedoroff, Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc. Natl. Acad. Sci.* **103**, 19033–19038 (2006).

8. A. H. Y. Tong, M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader, N. Pagé, M. Robinson, S. Raghibizadeh, C. W. V. Hogue, H. Bussey, B. Andrews, M. Tyers, C. Boone, Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. *Science (80-.).* **294**, 2364–2368 (2001).

9. B. Boucher, S. Jenna, Genetic interaction networks: better understand to better predict. *Front. Genet.* **4** (2013), doi:10.3389/fgene.2013.00290.

10. R. P. S. Onge, R. Mani, J. Oh, M. Proctor, E. Fung, R. W. Davis, C. Nislow, F. P. Roth, G. Giaever, Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* **39**, 199–206 (2007).

11. R. Mani, R. P. St.Onge, J. L. Hartman, G. Giaever, F. P. Roth, Defining genetic interaction. *Proc. Natl. Acad. Sci.* **105**, 3461–3466 (2008).

12. B. Lehner, Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27**, 323–331 (2011).

13. X. Pan, D. S. Yuan, D. Xiang, X. Wang, S. Sookhai-Mahadeo, J. S. Bader, P. Hieter, F. Spencer, J. D. Boeke, A Robust Toolkit for Functional Profiling of the Yeast Genome. *Mol. Cell*. **16**, 487–496 (2004).

14. S. J. Dixon, Y. Fedyshyn, J. L. Y. Koh, T. S. K. Prasad, C. Chahwan, G. Chua, K. Toufighi, A. Baryshnikova, J. Hayles, K.-L. Hoe, D.-U. Kim, H.-O. Park, C. L. Myers, A. Pandey, D. Durocher, B. J. Andrews, C. Boone, Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc. Natl. Acad. Sci.* **105**, 16653–16658 (2008).

15. T. R. Sorrells, A. D. Johnson, Making Sense of Transcription Networks. *Cell*. **161**, 714–723 (2015).

16. A. L. Lind, J. H. Wisecaver, T. D. Smith, X. Feng, A. M. Calvo, A. Rokas, Examining the

evolution of the regulatory circuit controlling secondary metabolism and development in the fungal genus Aspergillus. *PLoS Genet.* **11**, e1005096 (2015).

17. B. Yang, P. J. Wittkopp, Structure of the Transcriptional Regulatory Network Correlates with Regulatory Divergence in Drosophila. *Mol. Biol. Evol.* **34**, 1352–1362 (2017).

18. G. Monaco, S. van Dam, J. L. Casal Novo Ribeiro, A. Larbi, J. P. de Magalhães, A comparison of human and mouse gene co-expression networks reveals conservation and divergence at the tissue, pathway and disease levels. *BMC Evol. Biol.* **15**, 259 (2015).

19. T. Sato, Y. Yamanishi, M. Kanehisa, H. Toh, The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. *Bioinformatics*. **21**, 3482–3489 (2005).

20. N. L. Clark, E. Alani, C. F. Aquadro, Evolutionary rate covariation reveals shared functionality and coexpression of genes. *Genome Res.* **22**, 714–720 (2012).

21. C.-S. Goh, A. A. Bogan, M. Joachimiak, D. Walther, F. E. Cohen, Co-evolution of proteins with their interaction partners 1 1Edited by B. Honig. *J. Mol. Biol.* **299**, 283–293 (2000).

22. J. L. Steenwyk, T. J. Buida, A. L. Labella, Y. Li, X.-X. Shen, A. Rokas, PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* (2021), doi:10.1093/bioinformatics/btab096.

23. M. Medina, D. M. Baker, D. A. Baltrus, G. M. Bennett, U. Cardini, A. M. S. Correa, S. M. Degnan, G. Christa, E. Kim, J. Li, D. R. Nash, E. Marzinelli, M. Nishiguchi, C. Prada, M. S. Roth, M. Saha, C. I. Smith, K. R. Theis, J. Zaneveld, Grand Challenges in Coevolution. *Front. Ecol. Evol.* **9** (2022), doi:10.3389/fevo.2021.618251.

24. K. M. Kay, R. D. Sargent, The Role of Animal Pollination in Plant Speciation: Integrating Ecology, Geography, and Genetics. *Annu. Rev. Ecol. Evol. Syst.* **40**, 637–656 (2009).

25. G. D. Findlay, J. L. Sitnik, W. Wang, C. F. Aquadro, N. L. Clark, M. F. Wolfner, Evolutionary Rate Covariation Identifies New Members of a Protein Network Required for Drosophila melanogaster Female Post-Mating Responses. *PLoS Genet.* **10**, e1004108 (2014).

26. G. J. Brunette, M. A. Jamalruddin, R. A. Baldock, N. L. Clark, K. A. Bernstein, Evolution-based screening enables genome-wide prioritization and discovery of DNA repair genes. *Proc. Natl. Acad. Sci.* **116**, 19593–19599 (2019).

27. W. K. Kim, D. M. Bolser, J. H. Park, Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). *Bioinformatics*. **20**, 1138–1150 (2004).
28. D. M. Talsness, K. G. Owings, E. Coelho, G. Mercenne, J. M. Pleinis, R. Partha, K. A. Hope, A. R. Zuberi, N. L. Clark, C. M. Lutz, A. R. Rodan, C. Y. Chow, A Drosophila screen identifies NKCC1 as a modifier of NGLY1 deficiency. *Elife*. **9** (2020), doi:10.7554/eLife.57831.
29. J.-W. Huang, A. Acharya, A. Taglialatela, T. S. Nambiar, R. Cuella-Martin, G. Leuzzi, S. B. Hayward, S. A. Joseph, G. J. Brunette, R. Anand, R. K. Soni, N. L. Clark, K. A. Bernstein, P. Cejka, A. Ciccia, MCM8IP activates the MCM8-9 helicase to promote DNA synthesis and homologous recombination upon DNA damage. *Nat. Commun.* **11**, 2948 (2020).
30. Q. Raza, J. Y. Choi, Y. Li, R. M. O'Dowd, S. C. Watkins, M. Chikina, Y. Hong, N. L. Clark, A. V. Kwiatkowski, Evolutionary rate covariation analysis of E-cadherin identifies Raskol as a regulator of cell adhesion and actin dynamics in Drosophila. *PLOS Genet.* **15**, e1007720 (2019).
31. M. Usaj, Y. Tan, W. Wang, B. VanderSluis, A. Zou, C. L. Myers, M. Costanzo, B. Andrews, C. Boone, TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3: Genes|Genomes|Genetics*. **7**, 1539–1549 (2017).
32. M. Costanzo, J. Hou, V. Messier, J. Nelson, M. Rahman, B. VanderSluis, W. Wang, C. Pons, C. Ross, M. Ušaj, B.-J. San Luis, E. Shuteriqi, E. N. Koch, P. Aloy, C. L. Myers, C. Boone, B. Andrews, Environmental robustness of the global yeast genetic interaction network. *Science (80-.).* **372**, eabf8424 (2021).
33. S. Ciniawsky, I. Grimm, D. Saffian, W. Girzalsky, R. Erdmann, P. Wendler, Molecular snapshots of the Pex1/6 AAA+ complex in action. *Nat. Commun.* **6**, 7331 (2015).
34. B. E. Reuber, E. Germain-Lee, C. S. Collins, J. C. Morrell, R. Ameritunga, H. W. Moser, D. Valle, S. J. Gould, Mutations in PEX1 are the most common cause of peroxisome biogenesis disorders. *Nat. Genet.* **17**, 445–448 (1997).
35. R. Ayala, O. Willhoft, R. J. Aramayo, M. Wilkinson, E. A. McCormack, L. Ocloo, D. B. Wigley, X. Zhang, Structure and regulation of the human INO80–nucleosome complex.

Nature. **556**, 391–395 (2018).

36. E. S. Segal, V. Gritsenko, A. Levitan, B. Yadav, N. Dror, J. L. Steenwyk, Y. Silberberg, K. Mielich, A. Rokas, N. A. R. Gow, R. Kunze, R. Sharan, J. Berman, Gene Essentiality Analyzed by In Vivo Transposon Mutagenesis and Machine Learning in a Stable Haploid Isolate of Candida albicans. *MBio*. **9** (2018), doi:10.1128/mBio.02048-18.

37. E. A. Winzeler, D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C. J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Veronneau, M. Voet, G. Volckaert, T. R. Ward, R. Wysocki, G. S. Yen, K. X. Yu, K. Zimmermann, P. Philippsen, M. Johnston, R. W. Davis, Functional characterization of the S-cerevisiae genome by gene deletion and parallel analysis. *Science (80-.).* **285**, 901–906 (1999).

38. GeneOntologyConsortium, The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, 258D – 261 (2004).

39. L. H. Hartwell, J. Culotti, B. Reid, Genetic control of the cell-division cycle in yeast. I. Detection of mutants. *Proc. Natl. Acad. Sci. U. S. A.* **66**, 352–9 (1970).

40. K. Hosaka, T. Kodaki, S. Yamashita, Cloning and characterization of the yeast CKI gene encoding choline kinase and its expression in Escherichia coli. *J. Biol. Chem.* **264**, 2053–9 (1989).

41. K. Kim, K. H. Kim, M. K. Storey, D. R. Voelker, G. M. Carman, Isolation and characterization of the Saccharomyces cerevisiae EKI1 gene encoding ethanolamine kinase. *J. Biol. Chem.* **274**, 14857–66 (1999).

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

43. K. H. Wolfe, Origin of the Yeast Whole-Genome Duplication. *PLOS Biol.* **13**, e1002221 (2015).

44. J. L. Steenwyk, D. A. Opulente, J. Kominek, X.-X. Shen, X. Zhou, A. L. Labella, N. P.

Bradley, B. F. Eichman, N. Čadež, D. Libkind, J. DeVirgilio, A. B. Hulfachor, C. P. Kurtzman, C. T. Hittinger, A. Rokas, Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLOS Biol.* **17**, e3000255 (2019).

45. R. Albalat, C. Cañestro, Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).

46. L. D. Hurst, C. Pál, M. J. Lercher, The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**, 299–310 (2004).

47. A. Rokas, J. H. Wisecaver, A. L. Lind, The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Microbiol.* (2018), doi:10.1038/s41579-018-0075-3.

48. X.-X. Shen, D. A. Opulente, J. Kominek, X. Zhou, J. L. Steenwyk, K. V. Buh, M. A. B. Haase, J. H. Wisecaver, M. Wang, D. T. Doering, J. T. Boudouris, R. M. Schneider, Q. K. Langdon, M. Ohkuma, R. Endoh, M. Takashima, R. Manabe, N. Čadež, D. Libkind, C. A. Rosa, J. DeVirgilio, A. B. Hulfachor, M. Groenewald, C. P. Kurtzman, C. T. Hittinger, A. Rokas, Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*. **175**, 1533-1545.e20 (2018).

49. V. Mixão, T. Gabaldón, Genomic evidence for a hybrid origin of the yeast opportunistic pathogen Candida albicans. *BMC Biol.* **18**, 48 (2020).

50. D. A. Fitzpatrick, P. O'Gaora, K. P. Byrne, G. Butler, Analysis of gene evolution and metabolic pathways using the Candida Gene Order Browser. *BMC Genomics*. **11**, 290 (2010).

51. K. H. Wolfe, Comparative genomics and genome evolution in yeasts. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 403–412 (2006).

52. B. Dujon, Yeast evolutionary genomics. *Nat. Rev. Genet.* **11**, 512–524 (2010).

53. H. Chibana, N. Oka, H. Nakayama, T. Aoyama, B. B. Magee, P. T. Magee, Y. Mikami, Sequence Finishing and Gene Mapping for Candida albicans Chromosome 7 and Syntenic Analysis Against the Saccharomyces cerevisiae Genome. *Genetics*. **170**, 1525–1537 (2005).

54. C. Seoighe, N. Federspiel, T. Jones, N. Hansen, V. Bivolarovic, R. Surzycki, R. Tamse, C. Komp, L. Huizar, R. W. Davis, S. Scherer, E. Tait, D. J. Shaw, D. Harris, L. Murphy, K. Oliver, K. Taylor, M.-A. Rajandream, B. G. Barrell, K. H. Wolfe, Prevalence of small inversions in yeast gene order evolution. *Proc. Natl. Acad. Sci.* **97**, 14433–14437 (2000).

55. Z. Duan, M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S.

Fields, C. A. Blau, W. S. Noble, A three-dimensional model of the yeast genome. *Nature*. **465**, 363–367 (2010).

56. J. G. Gibbons, A. T. Branco, S. A. Godinho, S. Yu, B. Lemos, Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2485–2490 (2015).

57. J. G. Gibbons, A. T. Branco, S. Yu, B. Lemos, Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. *Nat. Commun.* **5**, 4850 (2014).

58. M. Pellegrini, (2012), pp. 167–177.

59. S. Cokus, S. Mizutani, M. Pellegrini, An improved method for identifying functionally linked proteins using phylogenetic profiles. *BMC Bioinformatics*. **8**, S7 (2007).

60. S. Mnaimneh, A. P. Davierwala, J. Haynes, J. Moffat, W.-T. Peng, W. Zhang, X. Yang, J. Pootoolal, G. Chua, A. Lopez, M. Trochesset, D. Morse, N. J. Krogan, S. L. Hiley, Z. Li, Q. Morris, J. Grigull, N. Mitsakakis, C. J. Roberts, J. F. Greenblatt, C. Boone, C. A. Kaiser, B. J. Andrews, T. R. Hughes, Exploration of Essential Gene Functions via Titratable Promoter Alleles. *Cell*. **118**, 31–44 (2004).

61. P. Gonçalves, C. Gonçalves, P. H. Brito, J. P. Sampaio, The Wickerhamiella/Starmerella clade—A treasure trove for the study of the evolution of yeast metabolism. *Yeast*. **37**, 313–320 (2020).

62. C. Gonçalves, P. Gonçalves, Multilayered horizontal operon transfers from bacteria reconstruct a thiamine salvage pathway in yeasts. *Proc. Natl. Acad. Sci.* **116**, 22219–22228 (2019).

63. T. Krassowski, A. Y. Coughlan, X.-X. Shen, X. Zhou, J. Kominek, D. A. Opulente, R. Riley, I. V. Grigoriev, N. Maheshwari, D. C. Shields, C. P. Kurtzman, C. T. Hittinger, A. Rokas, K. H. Wolfe, Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. *Nat. Commun.* **9**, 1887 (2018).

64. A. L. LaBella, D. A. Opulente, J. L. Steenwyk, C. T. Hittinger, A. Rokas, Variation and selection on codon usage bias across an entire subphylum. *PLOS Genet.* **15**, e1008304 (2019).

65. D. A. Opulente, E. J. Rollinson, C. Bernick-Roehr, A. B. Hulfachor, A. Rokas, C. P. Kurtzman, C. T. Hittinger, Factors driving metabolic diversity in the budding yeast

802 subphylum. *BMC Biol.* **16**, 26 (2018).
803 66. M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, T. O. Yeates, Assigning
804 protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc.*
805 *Natl. Acad. Sci.* **96**, 4285–4288 (1999).
806 67. F. Pazos, A. Valencia, Similarity of phylogenetic trees as indicator of protein–protein
807 interaction. *Protein Eng. Des. Sel.* **14**, 609–614 (2001).
808 68. D. de Juan, F. Pazos, A. Valencia, Emerging methods in protein co-evolution. *Nat. Rev.*
809 *Genet.* **14**, 249–261 (2013).
810 69. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and
811 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol.*
812 *Biol. Evol.* **32**, 268–274 (2015).
813 70. M. Chikina, J. D. Robinson, N. L. Clark, Hundreds of Genes Experienced Convergent
814 Shifts in Selective Pressure in Marine Mammals. *Mol. Biol. Evol.* **33**, 2182–2192 (2016).
815 71. M. E. J. Newman, Fast algorithm for detecting community structure in networks. *Phys.*
816 *Rev. E*. **69**, 066133 (2004).
817 72. A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.
818 D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P.
819 Philippsen, H. Tettelin, S. G. Oliver, Life with 6000 Genes. *Science (80-.).* **274**, 546–567
820 (1996).
821 73. T. Jones, N. A. Federspiel, H. Chibana, J. Dungan, S. Kalman, B. B. Magee, G. Newport,
822 Y. R. Thorstenson, N. Agabian, P. T. Magee, R. W. Davis, S. Scherer, The diploid
823 genome sequence of Candida albicans. *Proc. Natl. Acad. Sci.* **101**, 7329–7334 (2004).
824 74. D. V. Klopfenstein, L. Zhang, B. S. Pedersen, F. Ramírez, A. Warwick Vesztrocy, A.
825 Naldi, C. J. Mungall, J. M. Yunes, O. Botvinnik, M. Weigel, W. Dampier, C. Dessimoz,
826 P. Flick, H. Tang, GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.*
827 **8**, 10872 (2018).
828 75. F. Supek, M. Bosnjak, N. Skunca, T. Smuc, REVIGO summarizes and visualizes long
829 lists of gene ontology terms. *PLoS One*. **6**, e21800 (2011).
830 76. M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference
831 resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
832 77. M. van het Hoog, T. J. Rast, M. Martchenko, S. Grindle, D. Dignard, H. Hogues, C.

Cuomo, M. Berriman, S. Scherer, B. Magee, M. Whiteway, H. Chibana, A. Nantel, P. Magee, Assembly of the Candida albicans genome into sixteen supercontigs aligned on the eight chromosomes. *Genome Biol.* **8**, R52 (2007).

78. D. Muzzey, K. Schwartz, J. S. Weissman, G. Sherlock, Assembly of a phased diploid Candida albicans genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. *Genome Biol.* **14**, R97 (2013).

79. M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, M. A. Marra, Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

Figure legends

Figure 1. Constructing the budding yeast orthologous gene coevolution network. (A) We determined coevolution in a set of 2,408 single gene trees in which branch lengths were inferred along the species tree topology. (B) Coevolution of orthologous genes was evaluated across all pairwise combinations of orthologous genes using the CovER function in PhyKIT, v0.1 (*22*). (C) Significantly coevolving pairs of orthologous genes were used to construct a global network of orthologous gene coevolution where nodes correspond to orthologous genes and edges connect orthologous genes that are significantly coevolving. The "ring" of nodes corresponds to the orthologous genes found to be coevolving with very few or no other (i.e., singletons) orthologous genes in our dataset.

Figure 2. Network modules reflect modules of bioprocesses. (A) Global network of orthologous gene coevolution and essential (B) and nonessential (C) orthologous gene networks in *S. cerevisiae* and *C. albicans*. The "ring" of nodes in each plot is comprised of orthologous genes that coevolve with very few or no other genes. (D) The essential gene subnetwork has higher transitivity and edge density values. The nonessential gene network has higher mean distance and diameter values. (E) There are five major subnetworks or orthologous gene communities illustrated by different colors; small communities (≤10 orthologous genes) are in gray. Edge width: number of co-evolving orthologous gene pairs between communities; node size: number of orthologous genes in a community. Orthologous gene communities 1–4 cluster together; community 5 is a singleton. (F) Orthologous gene community 1 is overrepresented with essential orthologous genes. (G-I) Orthologous gene communities differ in enriched terms. MF: molecular functions; BP: biological processes; circles: enriched GO terms; colors: $-\log_{10}$ p-values; size of circles: GO term uniqueness. Enrichment results for each orthologous gene community are reported in Table S3. The figure legend is to the right of panel F.

Figure 3. Extensive coevolution in DNA replication genes. Cartoon representation of DNA replication. Exemplary complex specific subnetworks are depicted in i, ii, and iii. (i) Extensive coevolution between orthologous genes that encode the helicase, minichromosome maintenance (MCM) complex, which functions as a helicase. (ii) Coevolution in the orthologous genes that encode the DNA polymerase α-primase complex and (iii) DNA polymerase ε complex, which

are responsible for RNA primer synthesis and leading strand DNA synthesis, respectively. Edges in blue connect orthologous genes that are significantly coevolving. Orthologous genes and complexes in bold have signatures of coevolution. Orthologous genes and complexes are colored according to orthologous gene community assignment. Complexes, such as the DNA polymerase α-primase complex, are depicted in multiple colors reflecting the multiple orthologous gene communities represented within the complex. There is significant coevolution across all DNA replication orthologous genes ($p < 0.001$; permutation test) as well as the multimeric complexes such as the MCM complex ($p < 0.001$ for each pathway; permutation test).

Figure 4. The impact of perturbing the orthologous gene coevolutionary network through single-gene deletion in diverse environments is dependent on orthologous gene community and gene connectivity. (A) Multi-factor ANOVA results indicate orthologous gene community, environment, the interaction between orthologous gene community and environment, and the interaction between environment and the number of coevolving orthologous genes per orthologous gene are significantly associated with the fitness of a single-gene deletion strain (relative to the wild-type strain). (B) Fitness of single-gene deletion strains in diverse environments is impacted by orthologous gene community. Here, the y-axis indicates mean fitness across all genes in a community (x-axis) regardless of node degree. (C) Fitness of single-gene deletion strains in diverse environments is impacted by the number of coevolving orthologous genes the deleted node is connected to. Here, the y-axis indicates fitness of all genes with a given node degree (x-axis) regardless of community status. These results indicate that fitness in diverse environments is impacted by orthologous gene neighborhood and connectivity in the network. In both panels, colors correspond to different environments that fitness was measured in. Df represents degrees of freedom; Sum of Sq. represents sum of squares; Mean of Sq. represents Mean of squares.

Figure 5. Extensive long range and inter-chromosomal gene coevolution. (A) *S. cerevisiae* and (B) *C. albicans* differ in chromosome number and size. (C & D) Numbers of genes with inter-chromosomal orthologous gene coevolution (blue), intra-chromosomal (green), or both (orange). (E & F) Intra-chromosomal signatures of orthologous gene coevolution corrected by number of genes on chromosome (x-axis) and number of inter-chromosomal signatures of

orthologous gene coevolution corrected by number of genes on other chromosomes (y-axis). Colors represent different chromosomes and the regression line of all chromosomes is in black. (G & H) Distances among intra-chromosomal signatures of orthologous gene coevolution. (I & J) *INO80*, an example of how orthologous genes can coevolve with others across the genome. Outermost track: chromosomes of either yeast with chromosome 1 at the 12 o'clock position; second track: genes on plus/minus strand; third track: genes colored according to orthologous gene community. Scatter plot shows the number of coevolving orthologous genes per orthologous gene; size reflects higher values. Links depict orthologous genes coevolving with *INO80* and are colored according to chromosomal location of the other orthologous gene. Colors in E-H and ideogram and link colors in J correspond to chromosomes (see panels A and B).

Figure 1

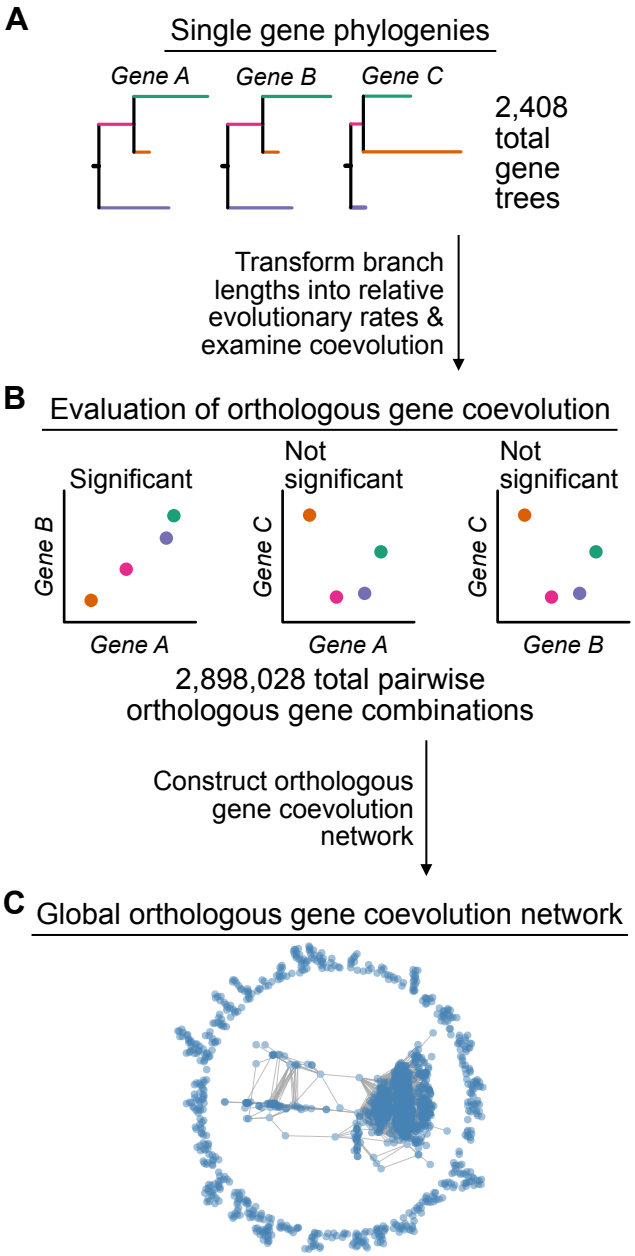

A
Single gene phylogenies
Gene A
Gene B
Gene C
2,408 total gene trees
Transform branch lengths into relative evolutionary rates & examine coevolution
B
Evaluation of orthologous gene coevolution
Significant
Not significant
Not significant
Gene B
Gene A
Gene C
Gene A
Gene C
Gene B
2,898,028 total pairwise orthologous gene combinations
Construct orthologous gene coevolution network
C
Global orthologous gene coevolution network

Figure 2

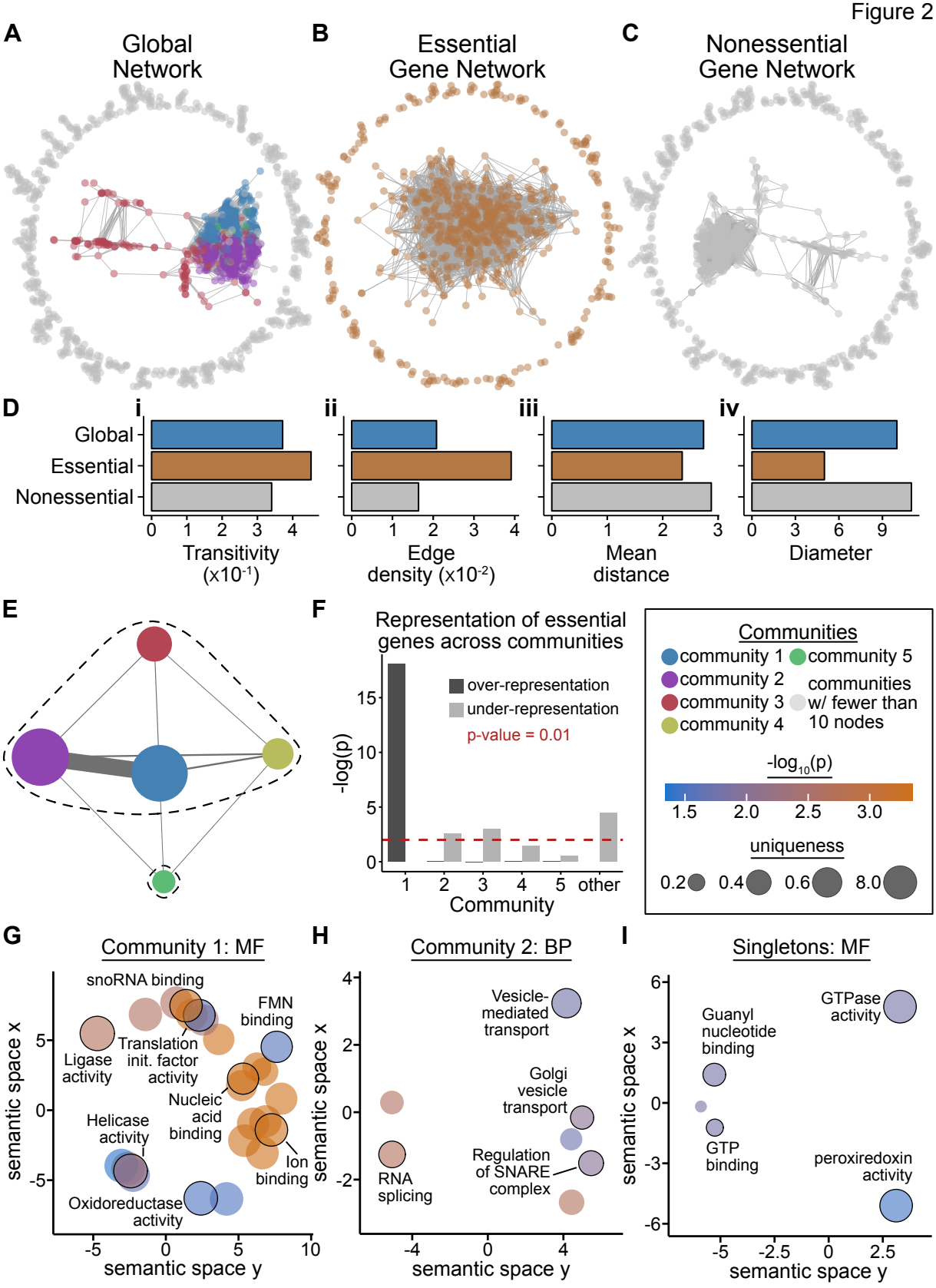
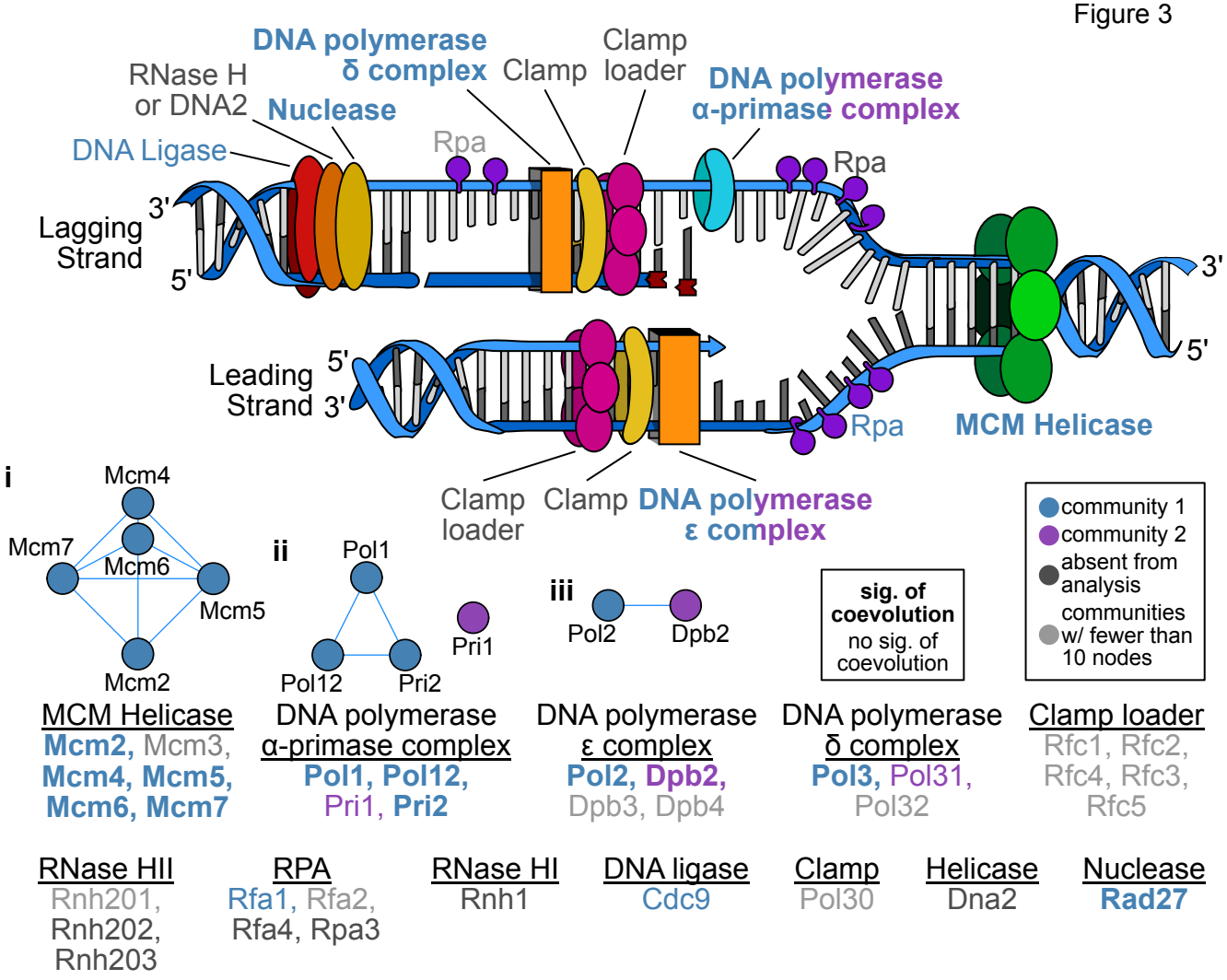

A
Global Network
B
Essential Gene Network
C
Nonessential Gene Network
D
i
Global
Essential
Nonessential
0 1 2 3 4
Transitivity (x10-1)
ii
0 1 2 3 4
Edge density (x10-2)
iii
0 1 2 3
Mean distance
iv
0 3 6 9
Diameter
E
F
Representation of essential genes across communities
over-representation
under-representation
p-value = 0.01
-log(p)
15
10
5
0
1 2 3 4 5 other
Community
Communities
community 1
community 2
community 3
community 4
community 5
communities w/ fewer than 10 nodes
-log10(p)
1.5 2.0 2.5 3.0
uniqueness
0.2 0.4 0.6 8.0
G
Community 1: MF
snoRNA binding
FMN binding
Ligase activity
Translation init. factor activity
Nucleic acid binding
Helicase activity
Ion binding
Oxidoreductase activity
semantic space x
semantic space y
H
Community 2: BP
Vesicle-mediated transport
Golgi vesicle transport
Regulation of SNARE complex
RNA splicing
semantic space x
semantic space y
I
Singletons: MF
GTPase activity
Guanyl nucleotide binding
GTP binding
peroxiredoxin activity
semantic space x
semantic space y

Figure 3



MCM Helicase
Mcm2, Mcm3, **Mcm4, Mcm5, Mcm6, Mcm7**

DNA polymerase α-primase complex
Pol1, Pol12, Pri1, **Pri2**

DNA polymerase ε complex
Pol2, Dpb2, Dpb3, Dpb4

DNA polymerase δ complex
Pol3, Pol31, Pol32

Clamp loader
Rfc1, Rfc2, Rfc4, Rfc3, Rfc5

RNase HII
Rnh201, Rnh202, Rnh203

RPA
Rfa1, Rfa2, Rfa4, Rpa3

RNase HI
Rnh1

DNA ligase
Cdc9

Clamp
Pol30

Helicase
Dna2

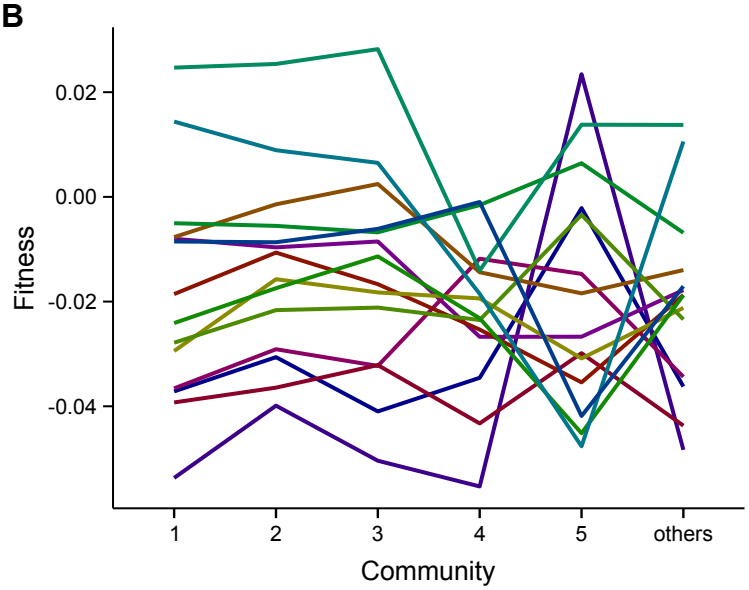
Nuclease
Rad27

Figure 4

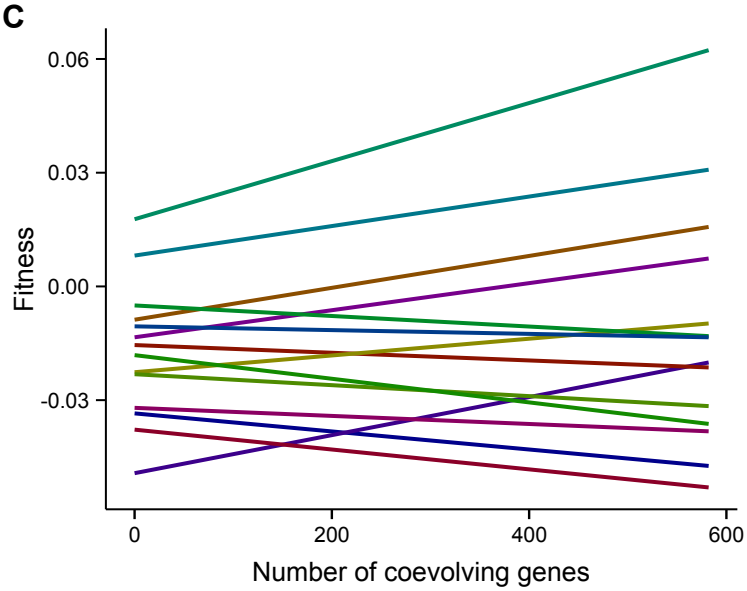
A

	Df	Sum of Sq.	Mean of Sq.	F value	P value	
Community	5	0.74	0.15	16.31	< 0.001	***
Environment	15	36.34	2.42	266.99	< 0.001	***
Number of coevolving genes	1	0.03	0.03	2.88	0.09	
Community:Environment	75	1.78	0.02	2.61	< 0.001	***
Community:Number of coevolving genes	2	0.13	0.03	2.82	0.02	
Environment:Number of coevolving genes	15	0.34	0.02	2.53	< 0.001	***
Community:Environment: Number of coevolving genes	75	0.58	0.01	0.85	0.82	

B



C



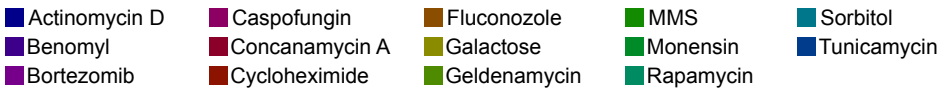
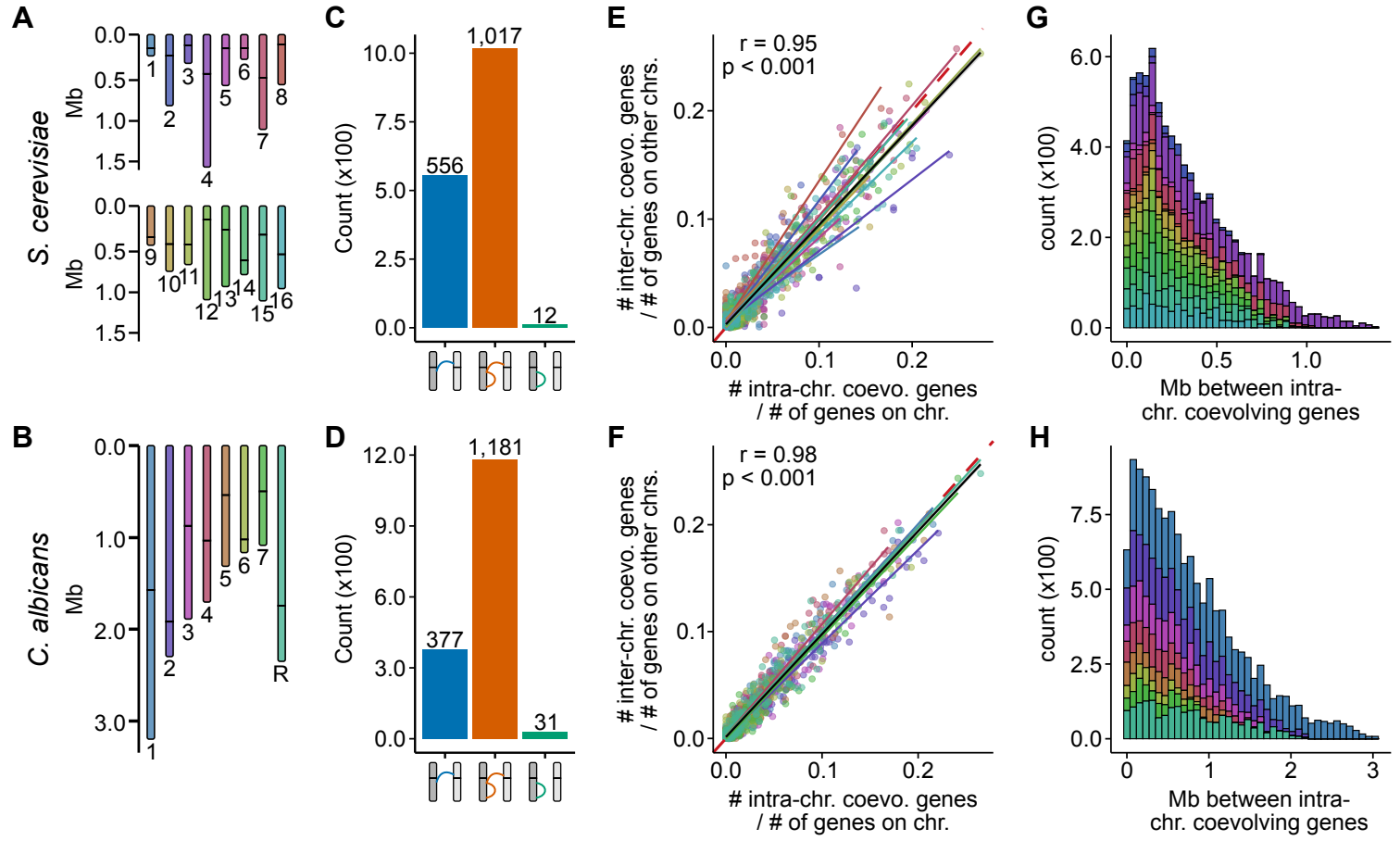
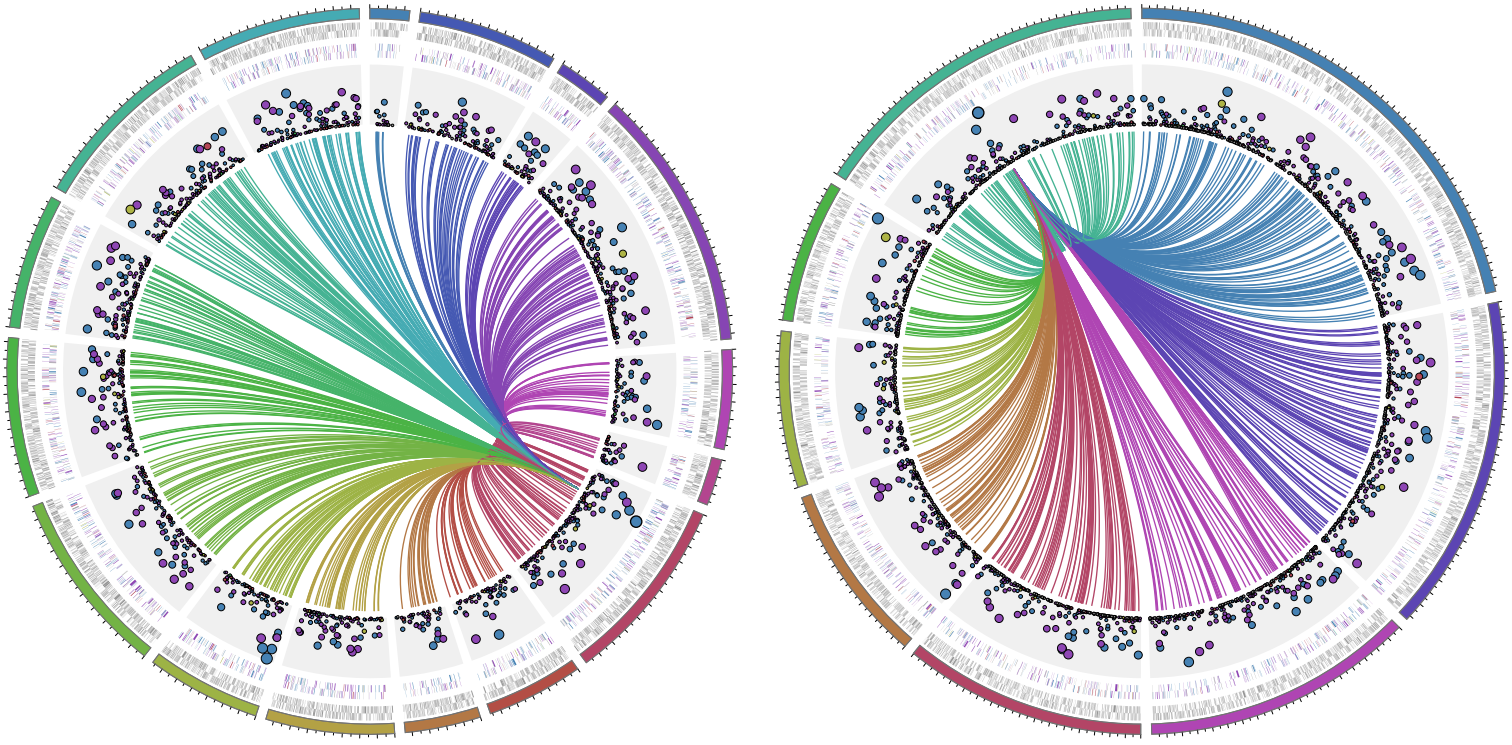


Figure 5



I Genes coevolving with *INO80* in *S. cerevisiae*

J Genes coevolving with *INO80* in *C. albicans*



Supplementary Figures

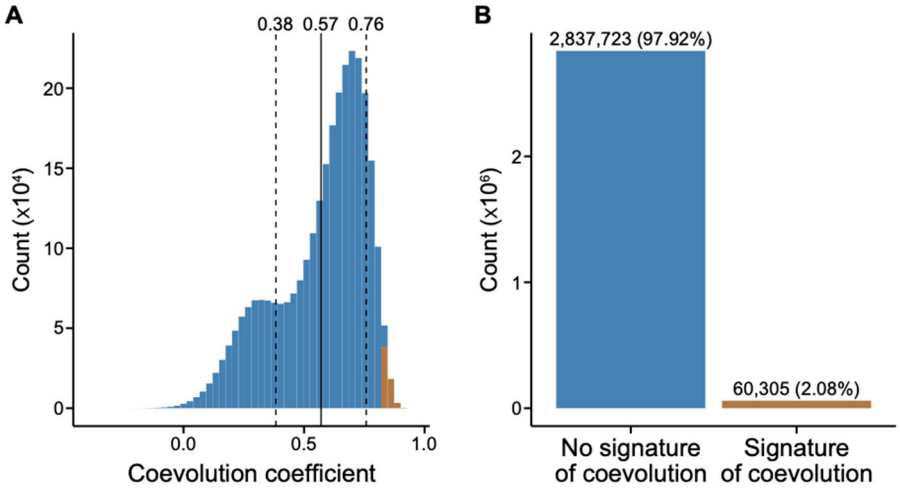


Figure S1. A conservative threshold for signatures of gene coevolution was used to construct the gene coevolution network. (A) Distribution of coevolution coefficients. The average coevolution coefficient is 0.57 and is represented by a solid line. One standard deviation away from the average is depicted using a dashed line. Insignificant coevolution coefficients are shown in blue whereas significant coevolution coefficients, which are defined as having values greater than or equal to 0.825, are shown in gold. (B) There are 60,305 significant signatures of gene coevolution whereas there are 2,837,723 insignificant signatures of gene coevolution.

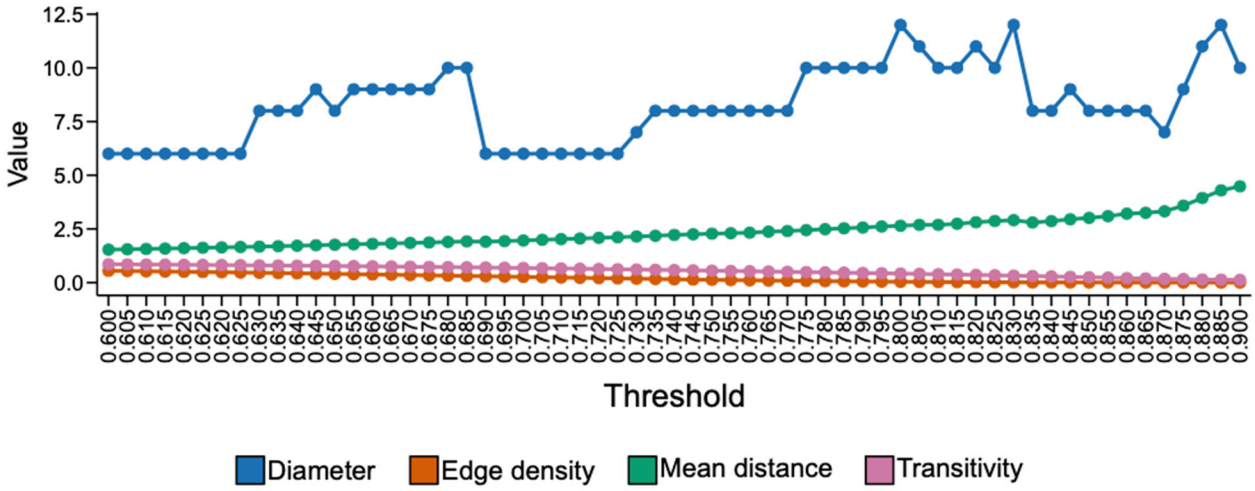


Figure S2. Various thresholds of significant gene coevolution had little impact on overall network features. To determine the impact of various thresholds of coevolution coefficients on the resulting network, we examined network diameter, edge density, mean distance, and transitivity. Network properties were similar regardless of the threshold used.

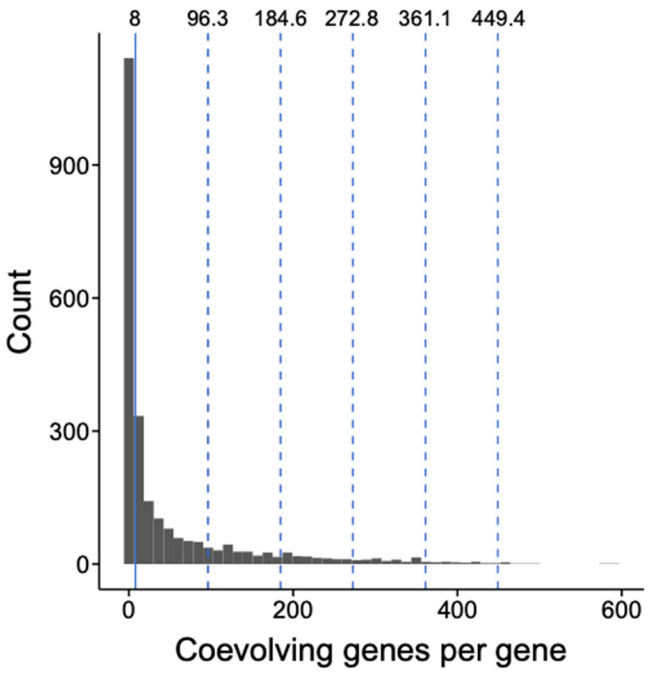


Figure S3. Distribution of node degrees. The median number of node degrees is eight (solid line). The dashed lines represent the median plus one, two, three, four, or five standard deviations.

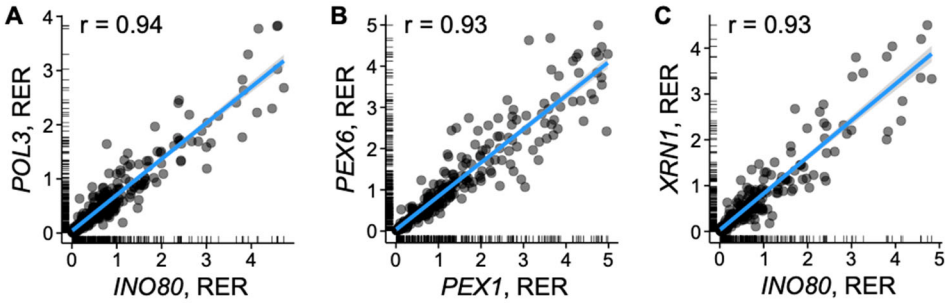


Figure S4. Three gene pairs with the strongest signatures of gene coevolution. (A) *INO80*, a gene that encodes a nucleosome spacing factor, and *POL3*, a gene that encodes the catalytic subunit of DNA polymerase delta, are significantly coevolving. *PEX1* and *PEX6*, which form a heterodimer, are also coevolving. Similarly, *INO80* and *XRN1*, which encodes an exoribonuclease, have significant signatures of evolution. Each dot represents a branch in the gene tree phylogeny.

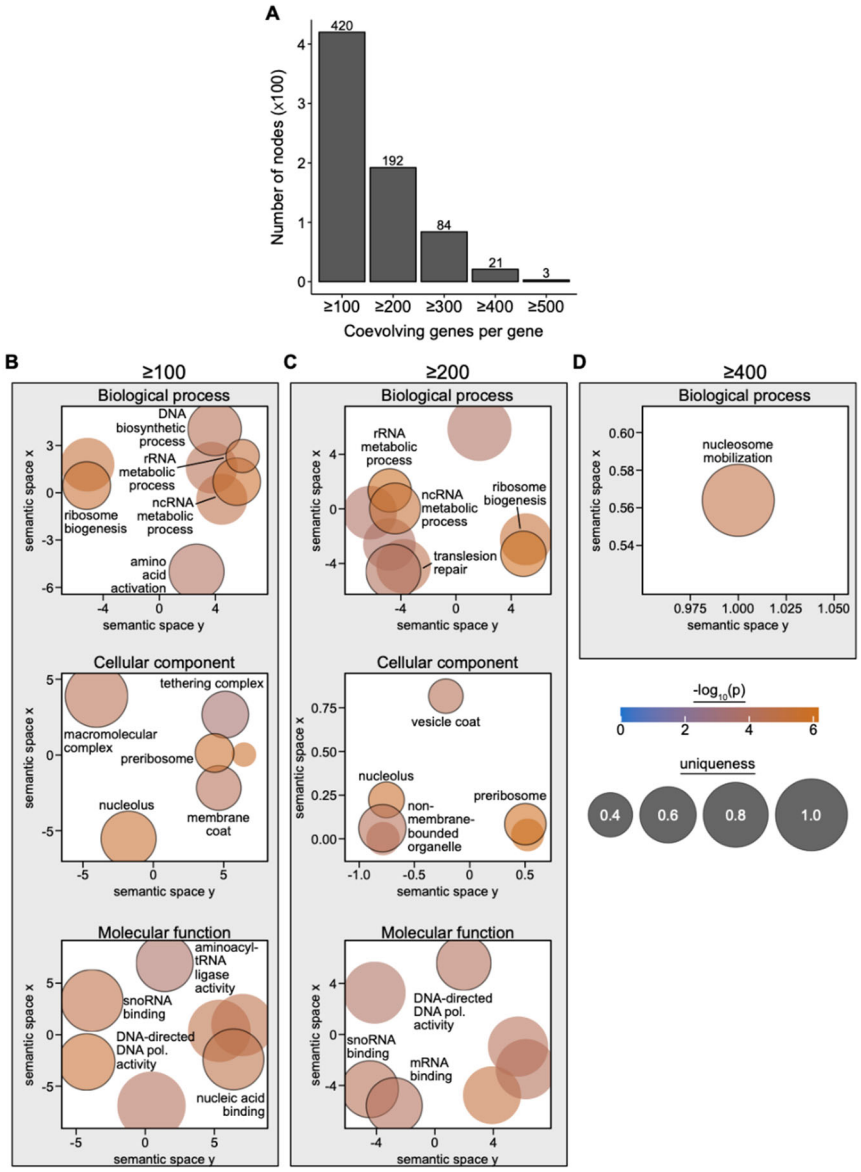


Figure S5. Gene ontology enrichment of genes with high degrees reveals functional categories associated with highlight coordinated processes. To determine what functional categories of genes are densely connected to other genes, we conducted GO enrichment analysis. (A) To do so, we first binned genes into groups having ≥100, ≥200, ≥300, ≥400, and ≥500 coevolving genes per gene. (B, C, and D) Enriched terms were observed among genes coevolving with ≥100, ≥200, and ≥400 genes. Enriched terms among genes coevolving with ≥100 and ≥200 genes included those involved in ribosome biogenesis and processes involving nucleic acids such as their binding and synthesis. (D) Among genes coevolving with ≥400 genes, there was one enriched term, nucleosome mobilization, which is associated with chromatin remodelling. In B-D, significantly enriched terms ($\alpha = 0.05$) are represented as circles in

semantic space. Uniqueness, a measure of GO term dissimilarity to all other enriched terms, is represented by circle size. Circle color is representative of $-\log_{10}$ transformed p-value. Highlighted enriched terms are written within the figure and their corresponding circle has a black outline. Complete GO enrichment analyses are reported in Table S1.

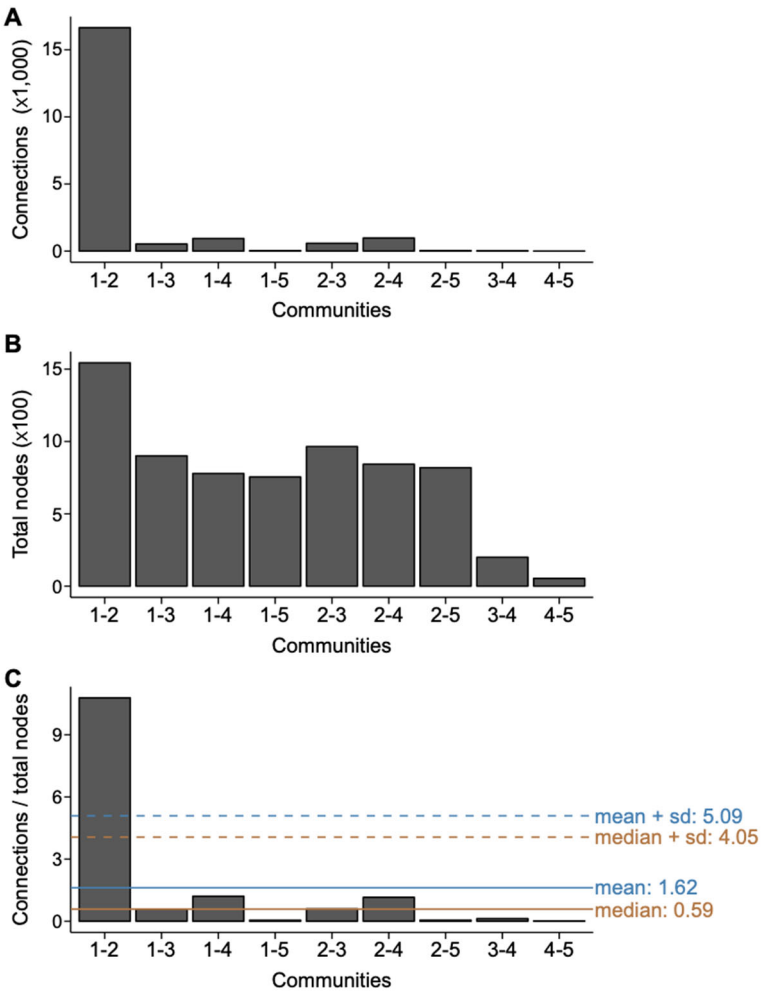


Figure S6. Orthologous gene communities one and two are highly connected. (A) Examination of the total number of connections between orthologous gene communities reveals orthologous gene community one and two are highly connected. Correcting the number of connections between orthologous gene communities by (B) the total number of genes in each orthologous gene community reveals that (C) orthologous gene communities one and two are exceptionally interconnected. Mean and median connections between orthologous gene communities corrected by the total number of genes in each orthologous gene community is represented in a blue and orange solid line, respectively. The mean or median plus one standard deviation is shown in a dashed line.

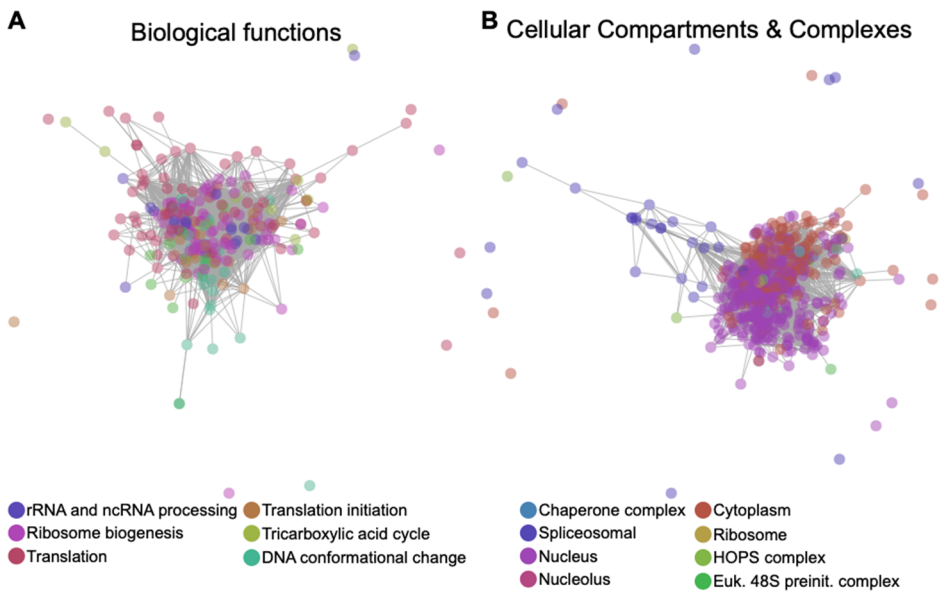


Figure S7. Subnetworks of broad categories reveal connections and bridges between biological functions and cellular compartments and complexes. (A) Diverse biological functions derived from enriched terms across orthologous gene communities reveal a high degree of interconnectedness. (B) Examination of cellular compartments and complexes uncovers cellular structure. For example, the nucleus (purple) and cytoplasm (red) are adjacent and intertwined in the network and are bridged by genes that perform spliceosome-related functions (dark blue). Similarly, transcripts are created and processed by the spliceosome in the nucleus and before being transported to the cytoplasm for translation.

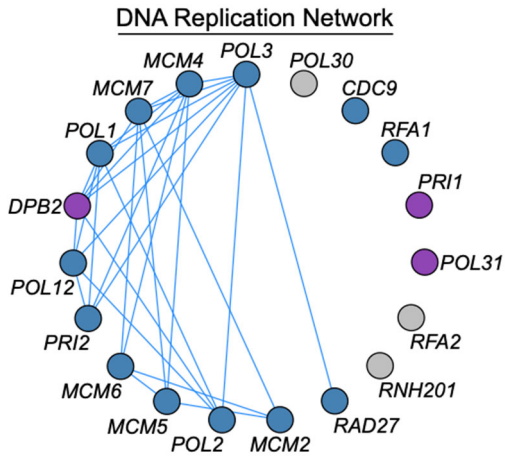


Figure S8. DNA replication as an exemplary pathway with signatures of gene-gene coevolution. Gene-gene coevolution network among genes involved in DNA replication. Nodes represent genes and edges connect coevolving genes. Genes are arranged counter-clockwise according to decreasing numbers of degrees, or coevolving genes per gene, in the subnetwork.

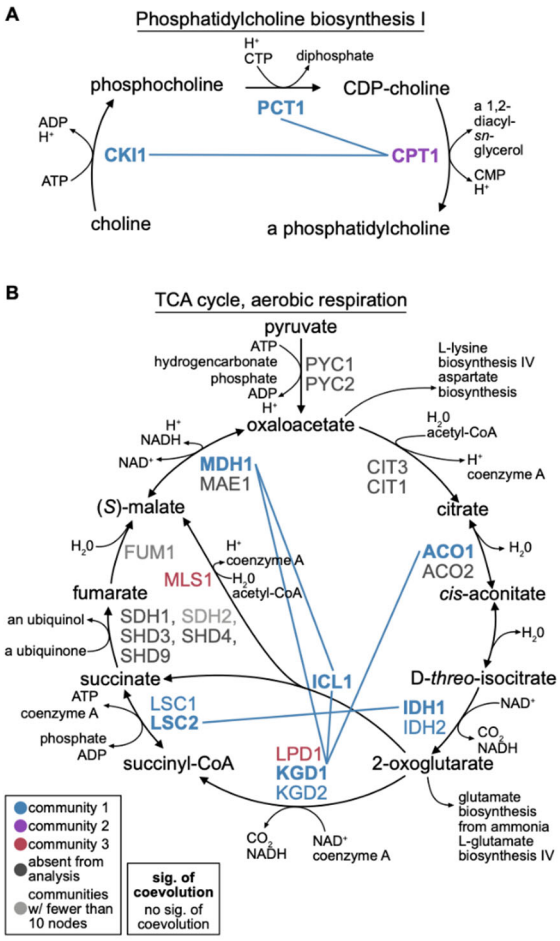


Figure S9. Exemplary pathways that are coevolving. (A) Phosphatidylcholine, the major phospholipid in organelle membranes, is synthesized by the genes *CKI1*, *PCT1*, and *CPT1*. *CPT1* is coevolving with *CKI1* and *PCT1*. (B) The tricarboxylic acid cycle (TCA cycle; also referred to as the Krebs cycle or citric acid cycle) is a key component of aerobic respiration in cells. Many genes in the TCA cycle are coevolving with one another.

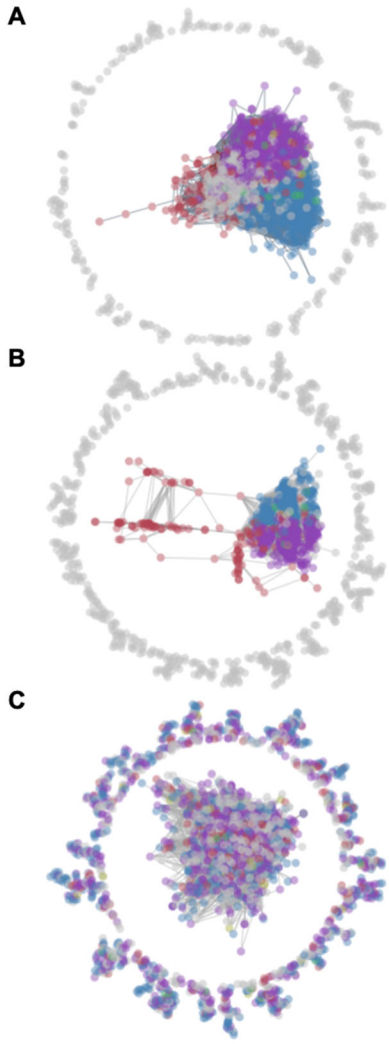


Figure S10. The coevolution genetic network and genetic interaction network differ. (A) Edges identified in the coevolution genetic network and genetic interaction network are combined into one super network. (B) The global network inferred using gene-gene coevolution. (C) The genetic interaction network among only significant genetic interaction scores. Nodes are genes and edges connect genes with significant signatures of coevolution or genetic interactions. Genes in orthologous gene community one, two, three, four, five, and all other small orthologous gene communities are depicted in blue, purple, red, yellow, green, and grey respectively.

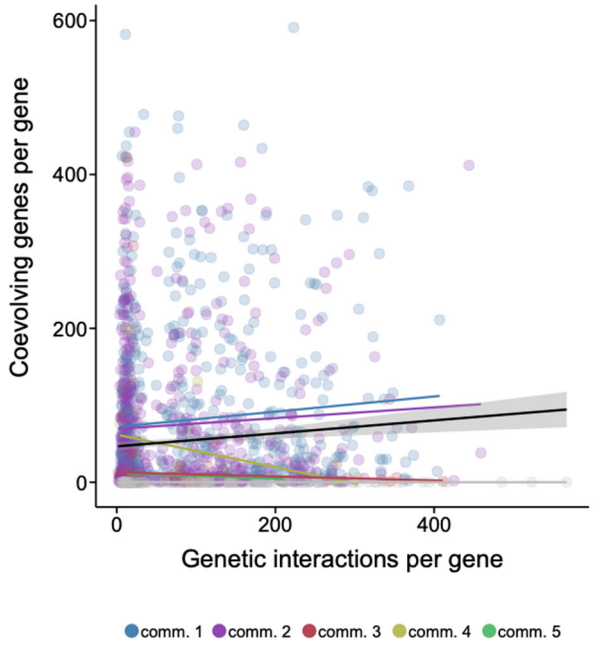


Figure S11. The coevolution genetic network and genetic interaction network differ in gene connectivity. Degrees in the coevolutionary network and the genetic interaction network are not related to one another. Each circle represents a gene and their color reflects the orthologous gene community they belong to. Lines represent linear model regressions between the degrees of each network and their color reflects the orthologous gene community they represent; a black line with a 95% confidence in grey represents the linear model regression across all orthologous gene communities.

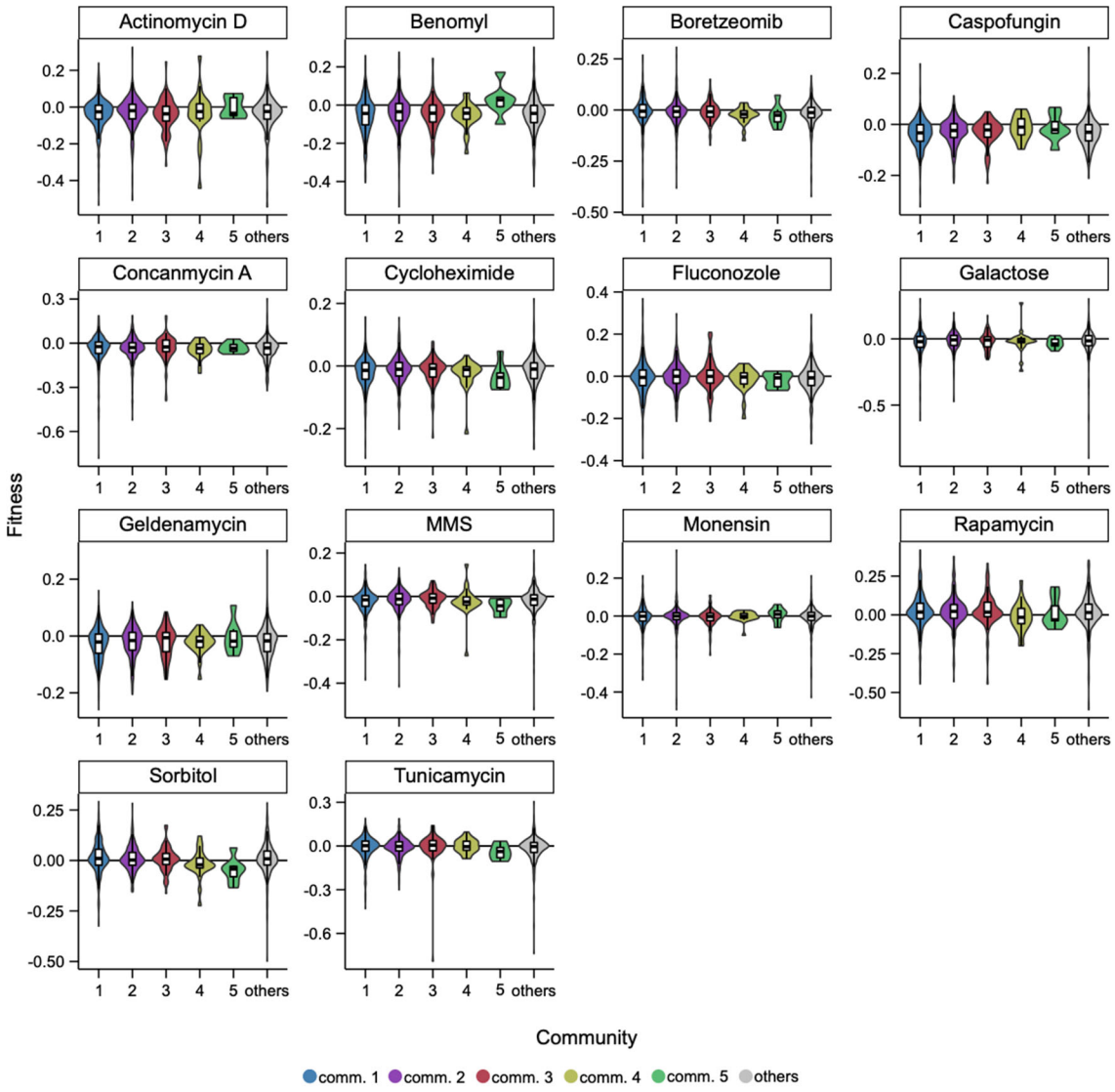
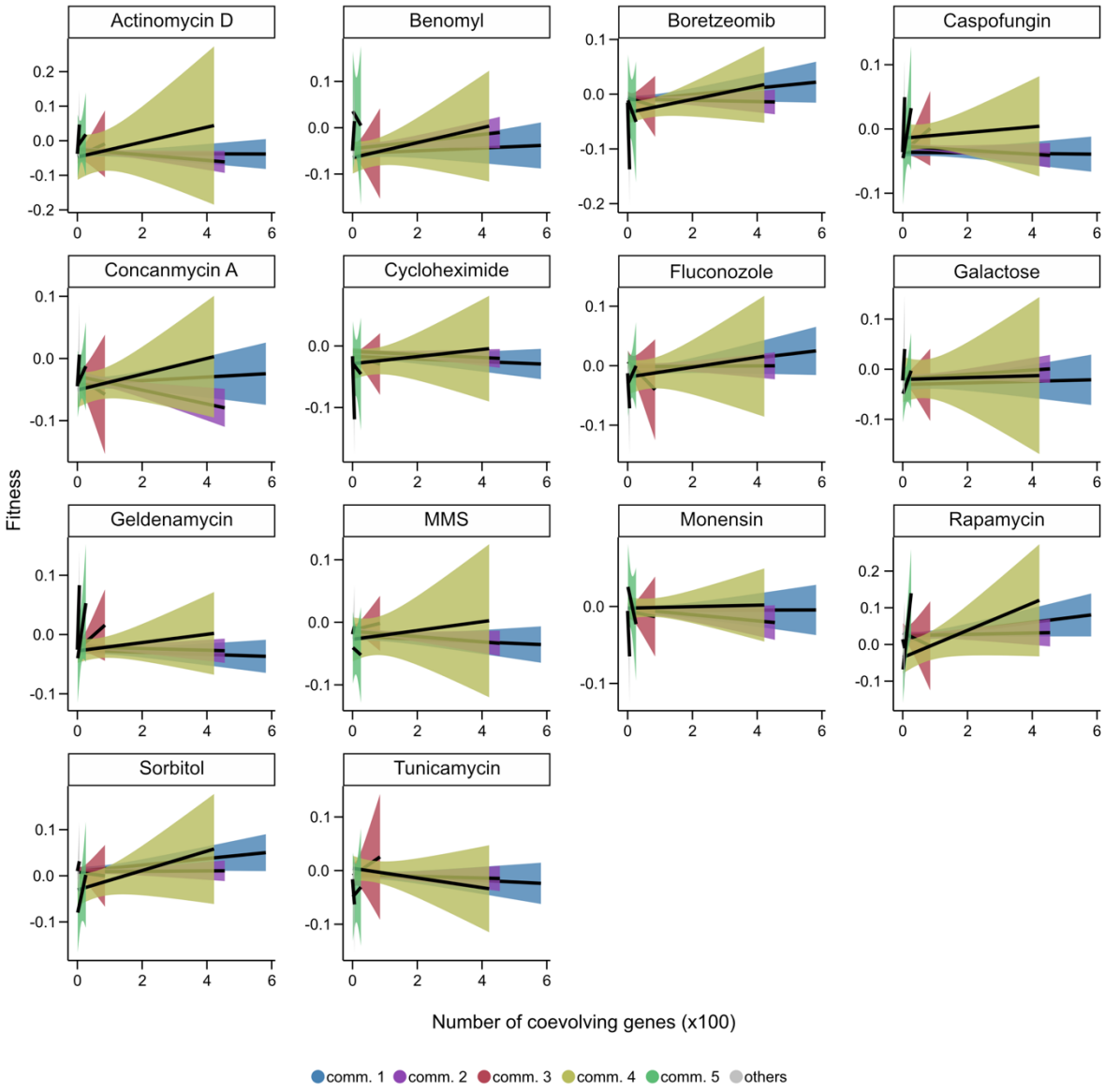


Figure S12. Orthologous gene community-dependent variation in fitness of single-gene knockouts across environments. Fitness (y-axis) among single-gene knockouts is in part dependent on the orthologous gene community (x-axis) to which the gene belongs to across 14 environments.



104

105 **Figure S13. Orthologous gene community-dependent variation in fitness of single-gene**
106 **knockouts is in part dependent on gene connectivity across environments.** Fitness (y-axis)
107 among single-gene knockouts is in part dependent on the orthologous gene community to which
108 the gene belongs to across 14 environments as well as gene connectivity as measured by the
109 number of coevolving genes per gene (x-axis).

110

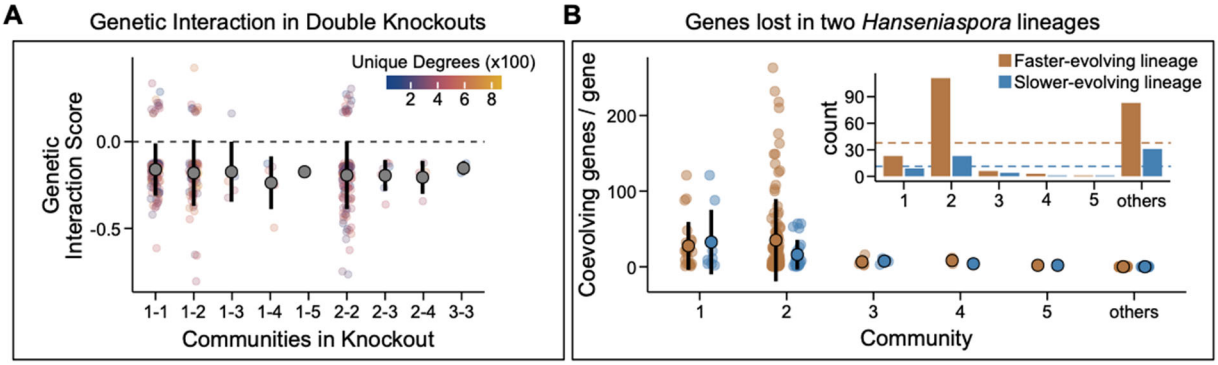


Figure S14. Digenic gene losses greatly impact cellular fitness and genes lost in the yeast genus *Hanseniaspora* are lost asymmetrically across orthologous gene communities. (A) Negative genetic interaction scores reflecting a worse phenotypic impact when two genes are deleted are more frequently observed across double knockouts. Genetic interaction scores on double knockouts of genes from different orthologous gene community combinations were not significantly different (p-value > 0.05; Kruskal-Wallis rank sum test). (B) Gene loss occurs asymmetrically across orthologous gene communities in two *Hanseniaspora* lineages. The inset depicts total counts of losses per orthologous gene community and dashed lines reflect average losses for each lineage.

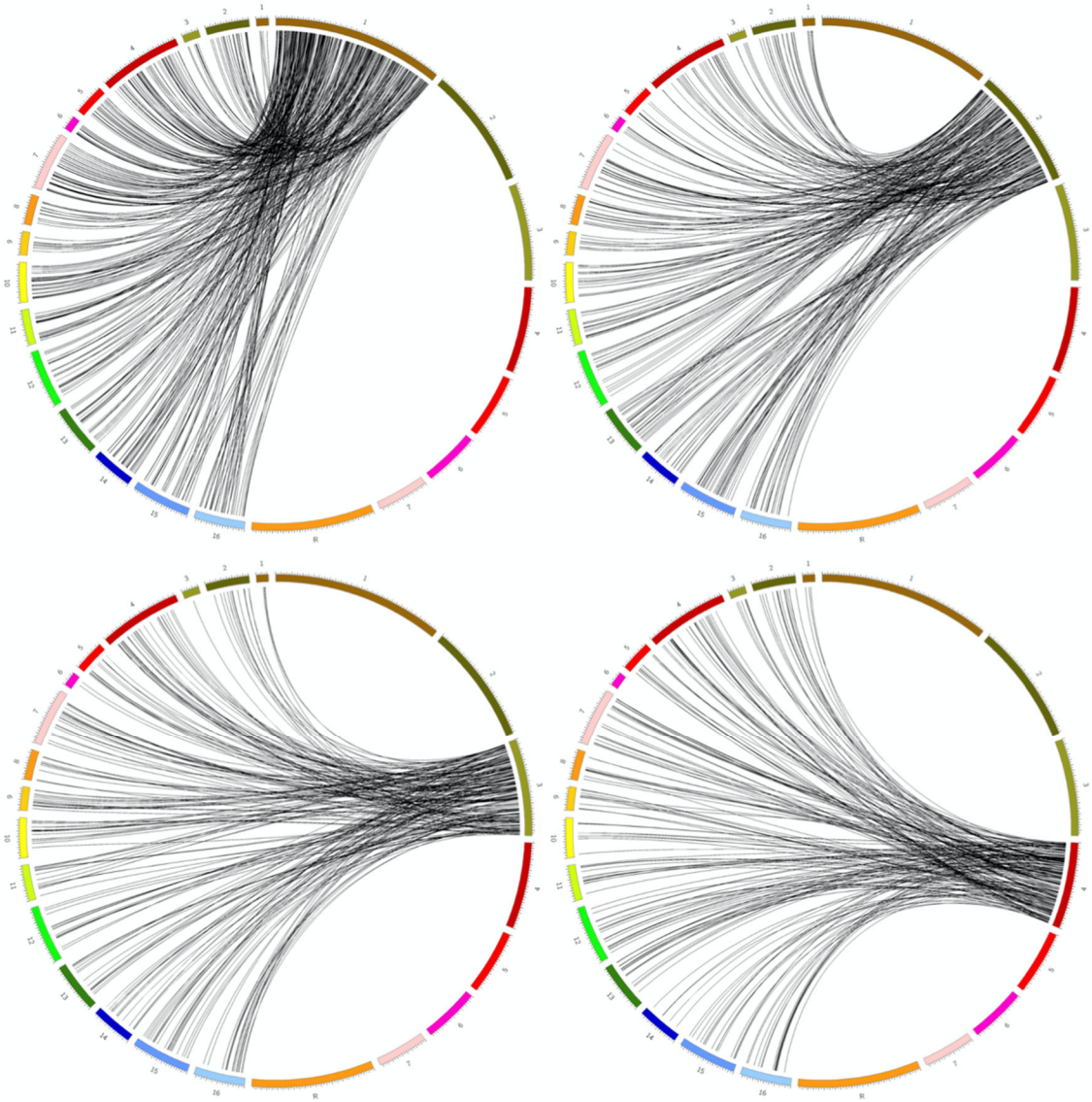


Figure S15. Lack of synteny between the *Candida albicans* chromosomes 1, 2, 3, and 4 and *Saccharomyces cerevisiae* chromosomes. The left 16 chromosomes are the *S. cerevisiae* chromosomes. The right eight chromosomes are the *C. albicans* chromosomes. Links start from a specific *C. albicans* chromosome and are connected to their orthologous gene in the *S. cerevisiae* genome. Gene orthology reveals a lack of synteny between the two genomes. Gene orthology information was obtained from the *Candida* genome browser http://www.candidagenome.org/download/homology/orthologs/C_albicans_SC5314_S_cerevisiae_by_CGOB/.

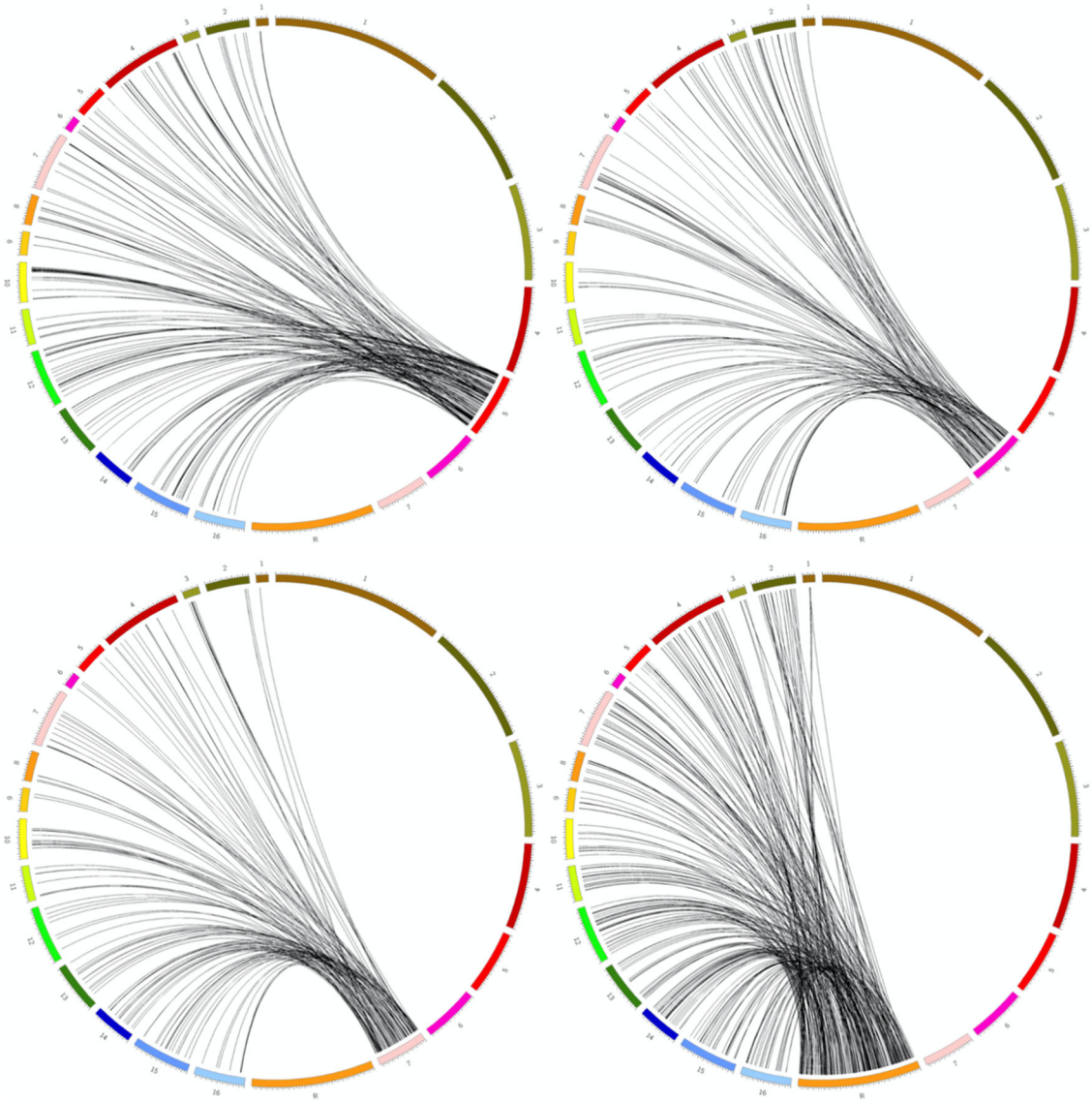


Figure S16. Lack of synteny between the *Candida albicans* chromosomes 5, 6, 7, and R and *Saccharomyces cerevisiae* chromosomes. The left 16 chromosomes are the *S. cerevisiae* chromosomes. The right eight chromosomes are the *C. albicans* chromosomes. Links start from a specific *C. albicans* chromosome and are connected to their orthologous gene in the *S. cerevisiae* genome. Gene orthology reveals a lack of synteny between the two genomes Gene orthology information was obtained from the *Candida* genome browser http://www.candidagenome.org/download/homology/orthologs/C_albicans_SC5314_S_cerevisiae_by_CGOB/.

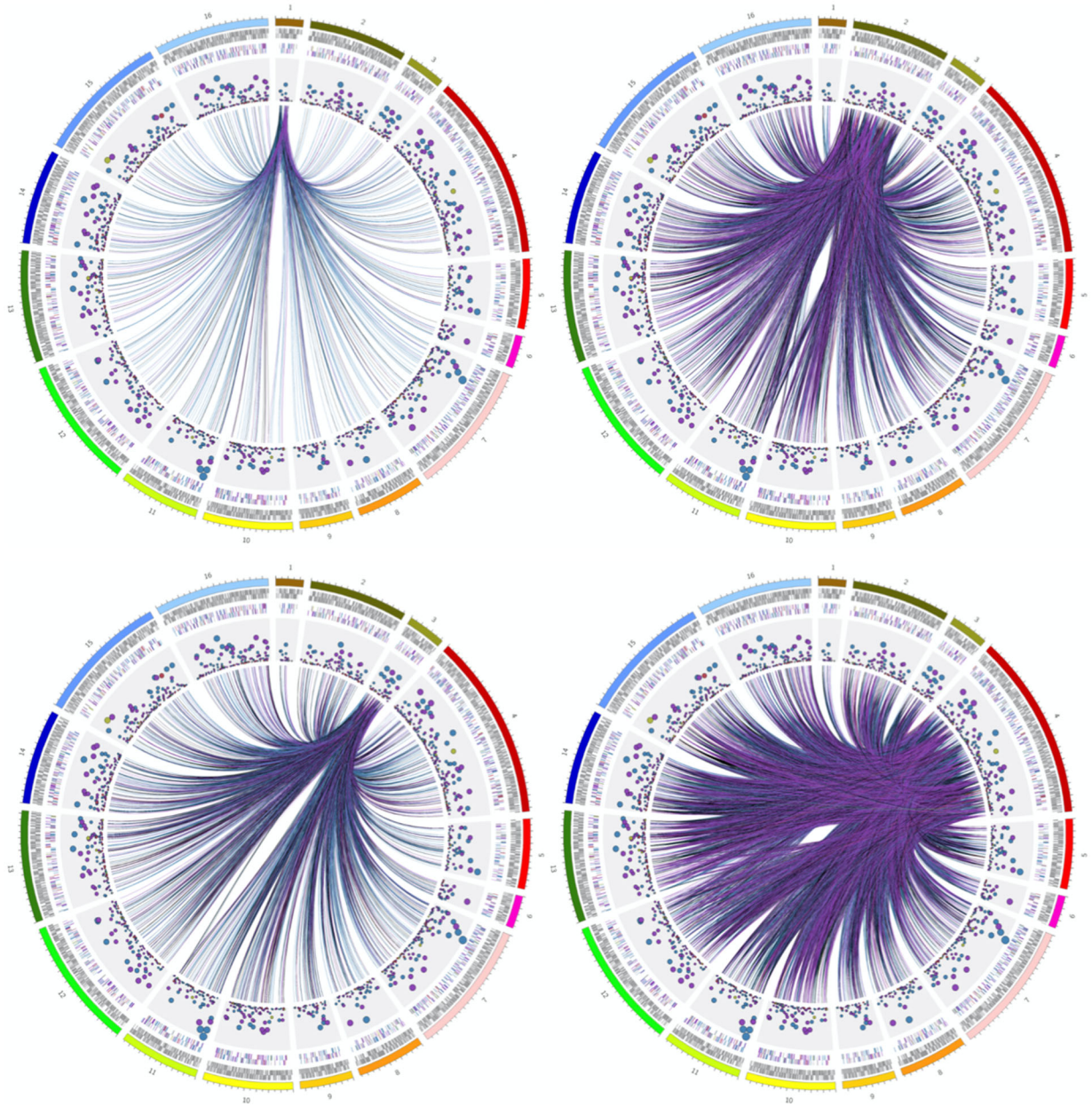


Figure S17. Genes on chromosomes one through four are coevolving with genes on all other chromosomes in *Saccharomyces cerevisiae*. The first track depicts the 16 chromosomes of *S. cerevisiae*. The next track depicts the location of all genes on the plus strand followed by the minus strand. The next tracks depict genes present in the dataset and are colored according to the orthologous gene community they belong. The scatter plot depicts the number of coevolving genes a gene has and the larger the circle represents more coevolving genes. The last track depicts links between genes on a highlighted chromosome that are coevolving with other genes

across all chromosomes. Plots are arranged in ascending chromosome order from top left to bottom right.

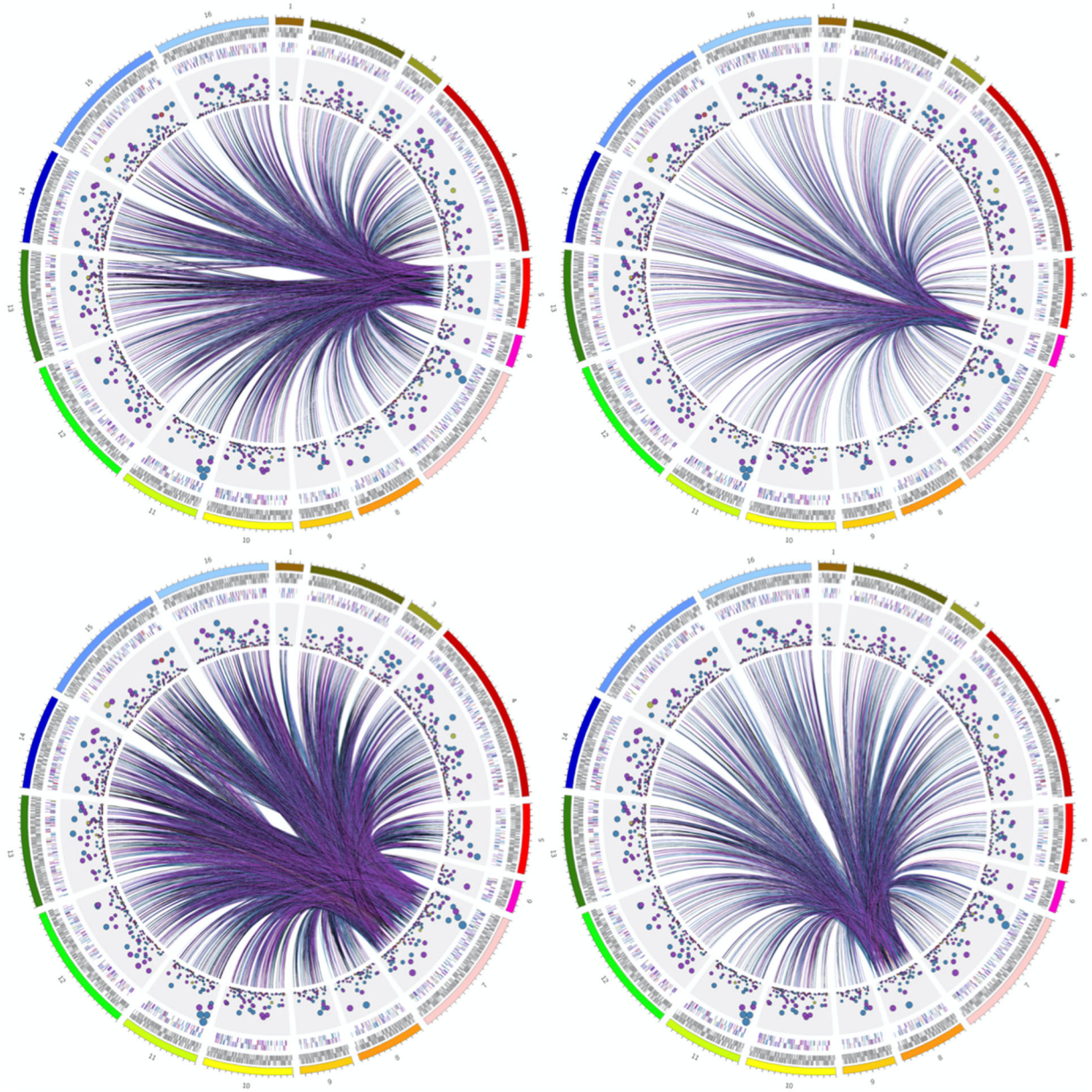


Figure S18. Genes on chromosomes five through eight are coevolving with genes on all other chromosomes in *Saccharomyces cerevisiae*. The first six tracks depict the same data as in Figure *N* (Genes on chromosomes one through four are coevolving with genes on all other chromosomes). Links depict genes that are coevolving with genes on a highlighted chromosome. Plots are arranged in ascending chromosome order from top left to bottom right.

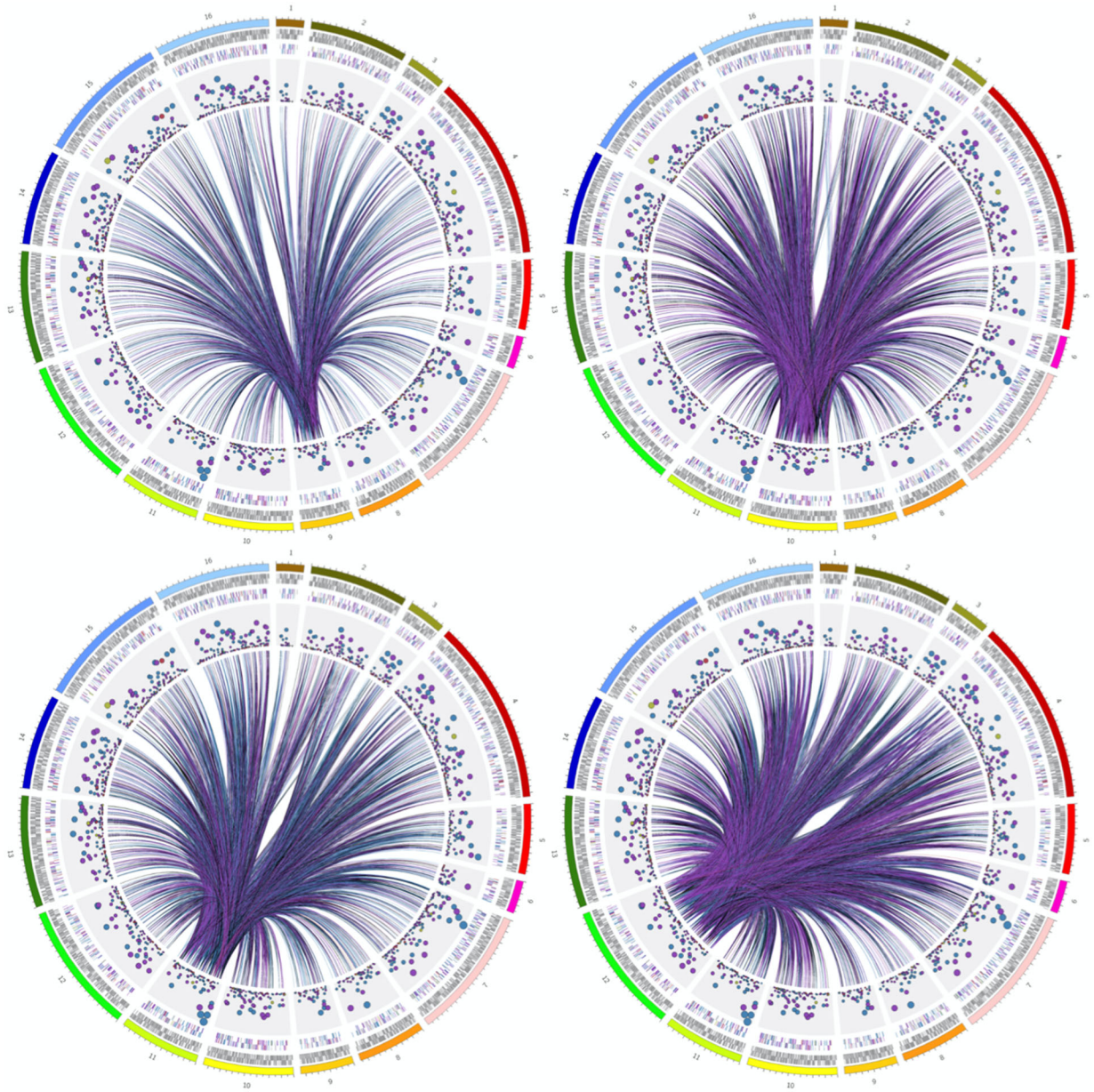


Figure S19. Genes on chromosomes nine through twelve are coevolving with genes on all other chromosomes in *Saccharomyces cerevisiae*. The first six tracks depict the same data as in Figure *N* (Genes on chromosomes one through four are coevolving with genes on all other chromosomes). Links depict genes that are coevolving with genes on a highlighted chromosome. Plots are arranged in ascending chromosome order from top left to bottom right.

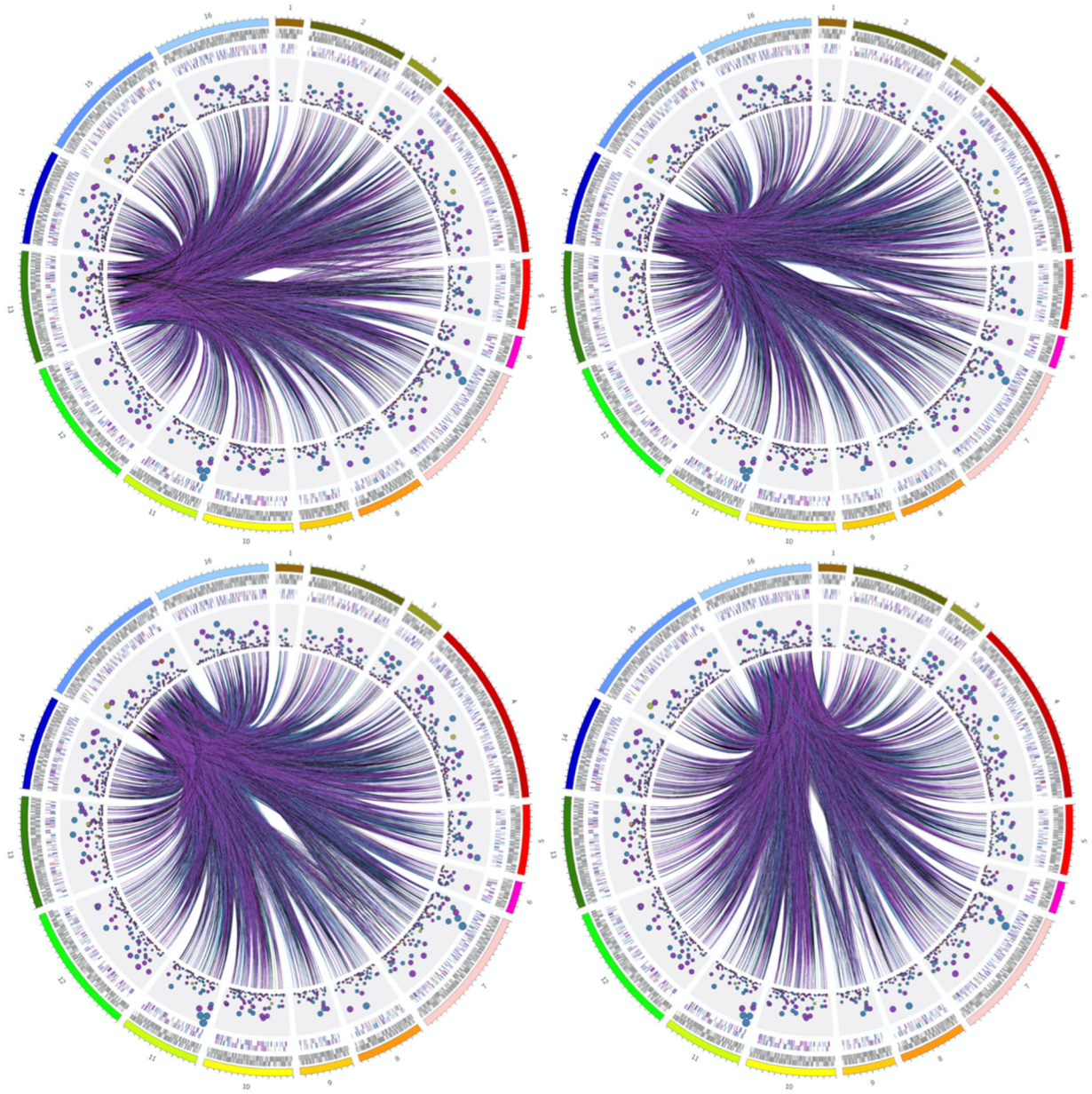


Figure S20. Genes on chromosomes thirteen through sixteen are coevolving with genes on all other chromosomes in *Saccharomyces cerevisiae*. The first six tracks depict the same data as in Figure *N* (Genes on chromosomes one through four are coevolving with genes on all other chromosomes). Links depict genes that are coevolving with genes on a highlighted chromosome. Plots are arranged in ascending chromosome order from top left to bottom right.

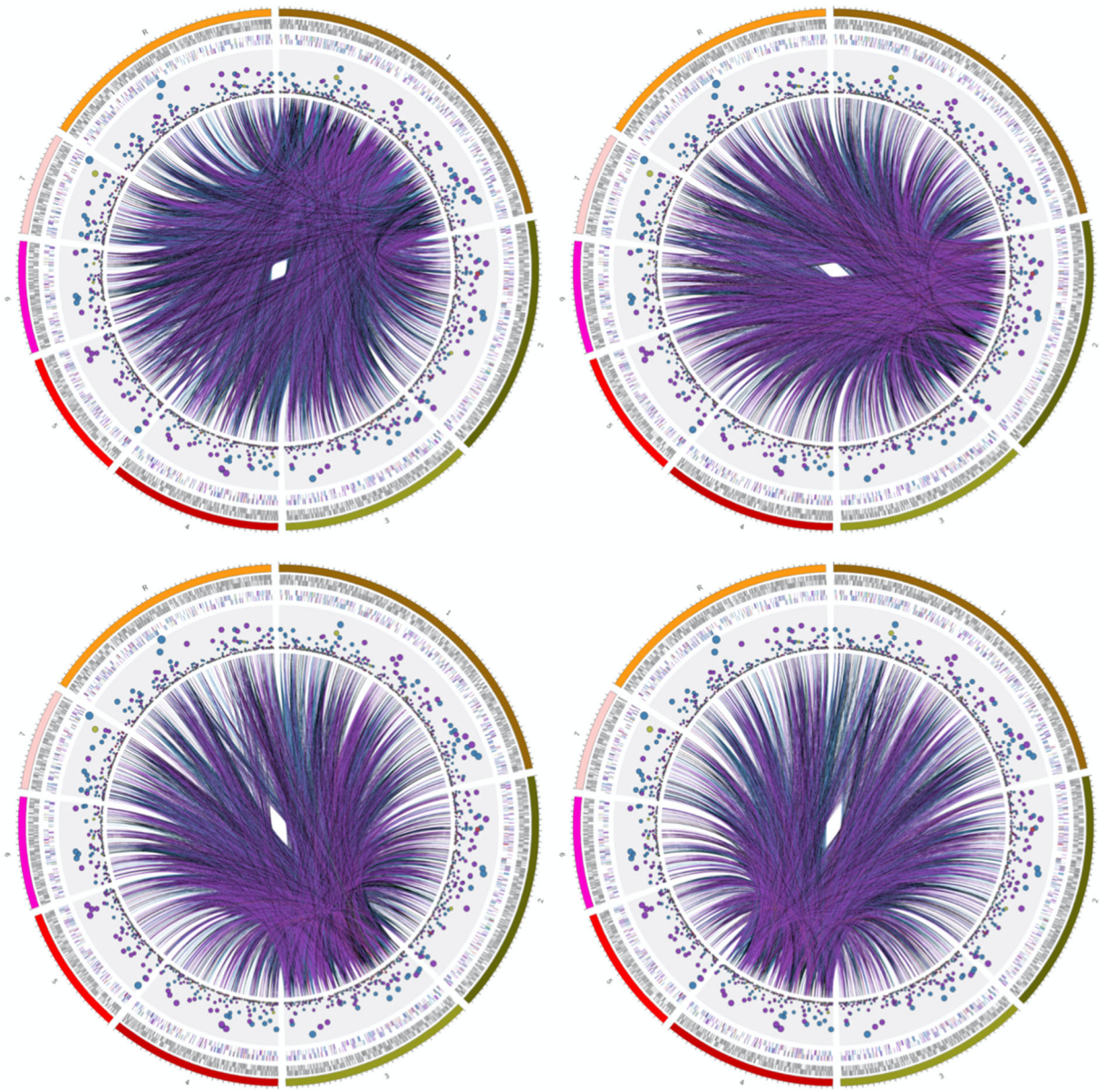


Figure S21. Genes on chromosomes one through four are coevolving with genes on all other chromosomes in *Candida albicans*. The first six tracks depict the same data as in Figure *N* (Genes on chromosomes one through four are coevolving with genes on all other chromosomes). Links depict genes that are coevolving with genes on a highlighted chromosome. Plots are arranged in ascending chromosome order from top left to bottom right.

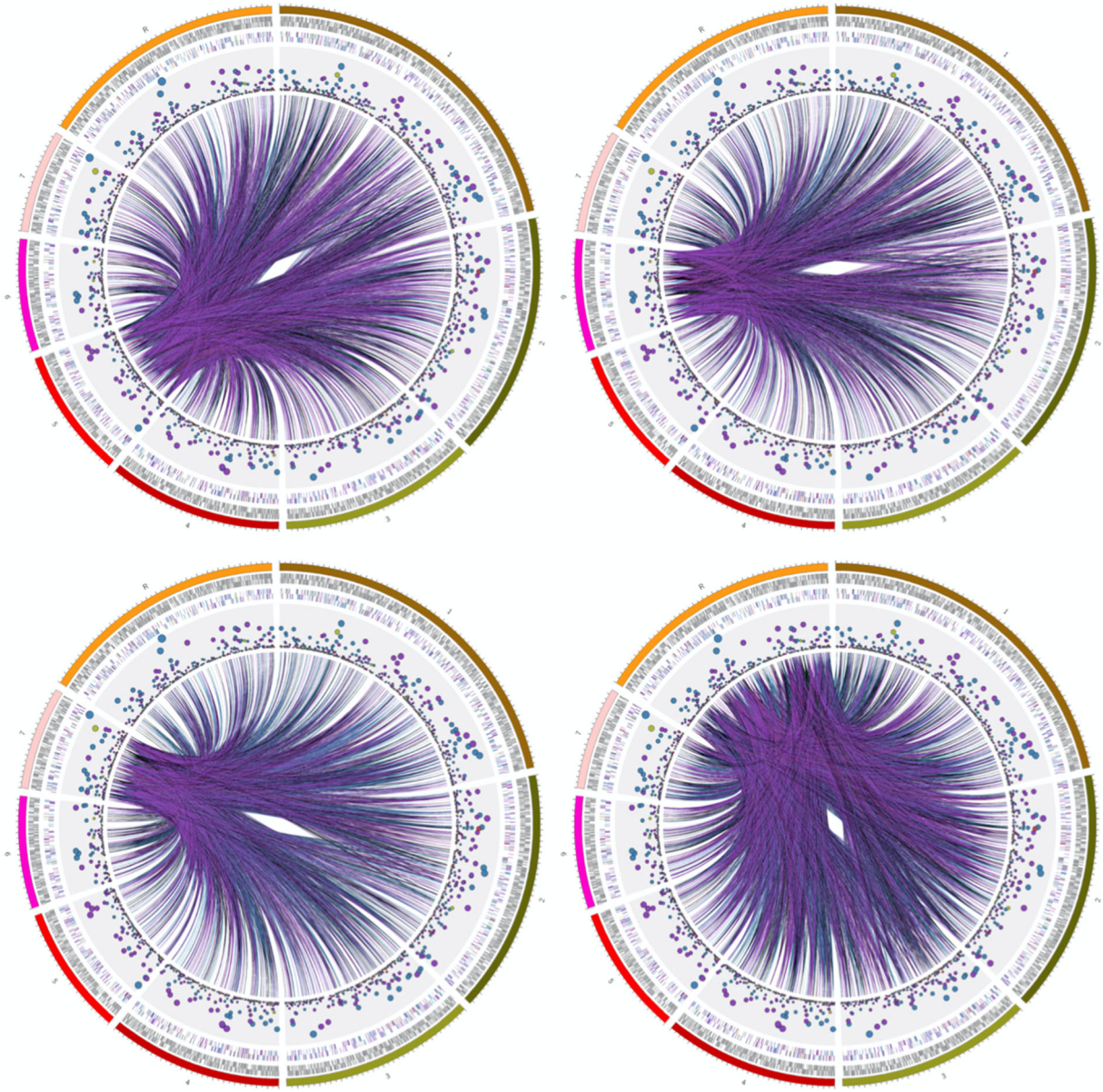
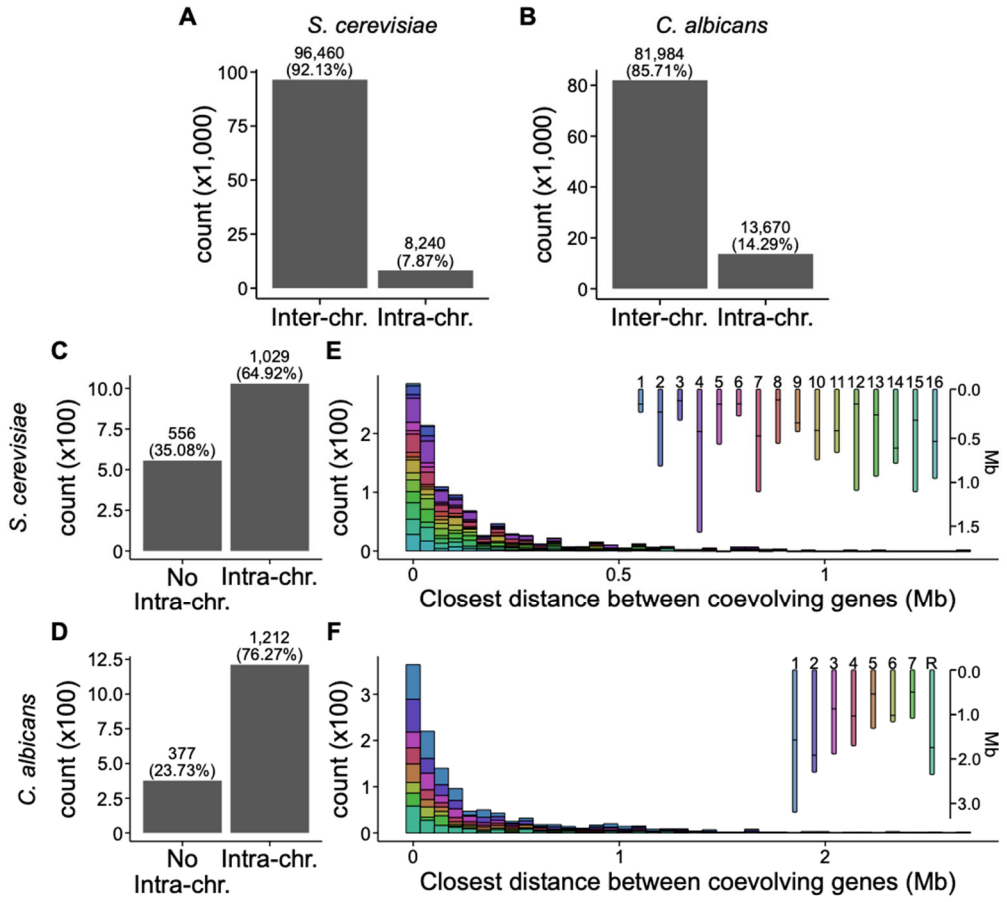


Figure S22. Genes on chromosomes five through seven and chromosome R are coevolving with genes on all other chromosomes in *Candida albicans*. The first six tracks depict the same data as in Figure *N* (Genes on chromosomes one through four are coevolving with genes on all other chromosomes). Links depict genes that are coevolving with genes on a highlighted chromosome. Plots are arranged in ascending chromosome order from top left to bottom right.



186

187 **Figure S23. There are many inter-chromosomal associations and intra-chromosomal**
188 **associations can be long range.** (A and B) The total number of inter- and intra-chromosomal
189 gene coevolutionary signatures reveal substantially more inter-chromosomal coevolution in
190 comparison to intra-chromosomal coevolution in *S. cerevisiae* and *C. albicans*. (C and D)
191 Examination of the number of genes that have no signatures of intra-chromosomal evolution and
192 the number of genes with signatures of intra-chromosomal coevolution reveal a substantial
193 portion of genes are not coevolving with genes on the same chromosome in either species of
194 yeast. (E and F) The distribution of every gene and their closest coevolving partner among genes
195 with evidence of intra-chromosomal coevolution reveals some genes may be coevolving despite
196 substantial distances between them.

197

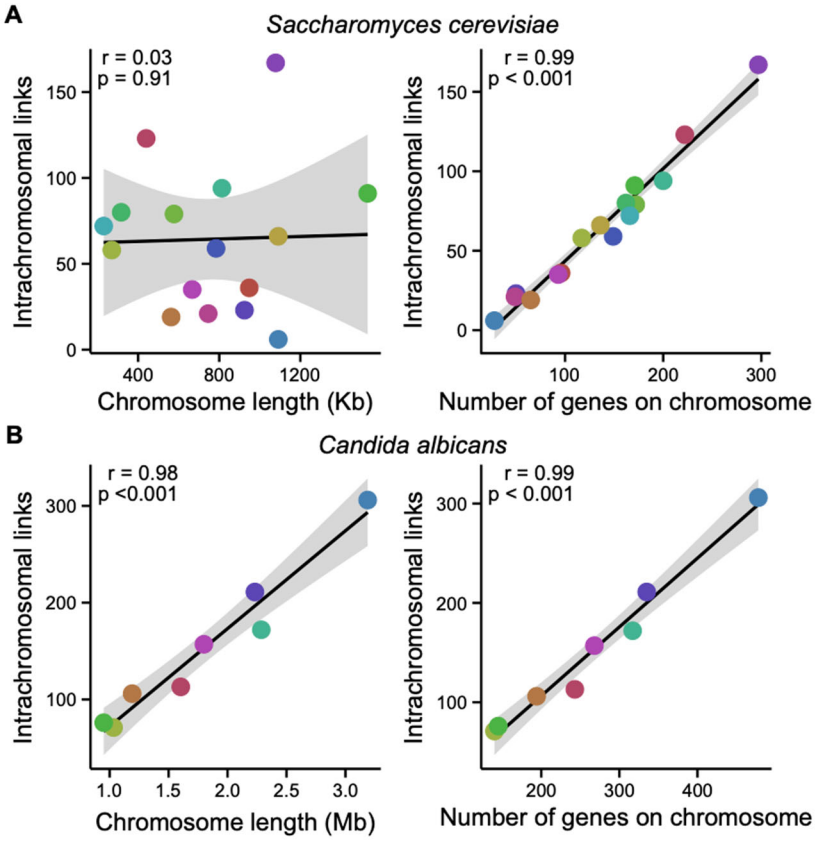


Figure S24. The number of genes on a chromosome is the primary driver of the number of intrachromosomal links. (A and B left panels) Chromosome length (x-axis) and the number of signatures of intrachromosomal gene coevolution (y-axis) are not significantly associated in (A) *S. cerevisiae* but are in (B) *C. albicans* (r = 0.03, p = 0.91 and r=0.98, p<0.001, respectively; Pearson Correlation). (A and B right panels) The number of genes on a chromosome (x-axis) and the number of signatures of intrachromosomal gene-gene coevolution (y-axis) are significantly associated for both (A) *S. cerevisiae* and (B) *C. albicans* (r = 0.99, p < 0.001 for both species; Pearson Correlation). Colors of each dot correspond to one chromosome. The color scheme is the same as depicted in Figure S21.

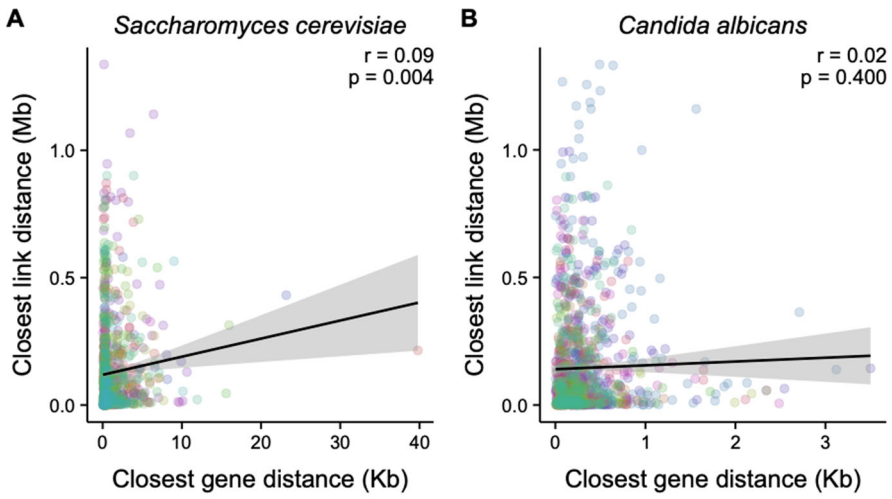
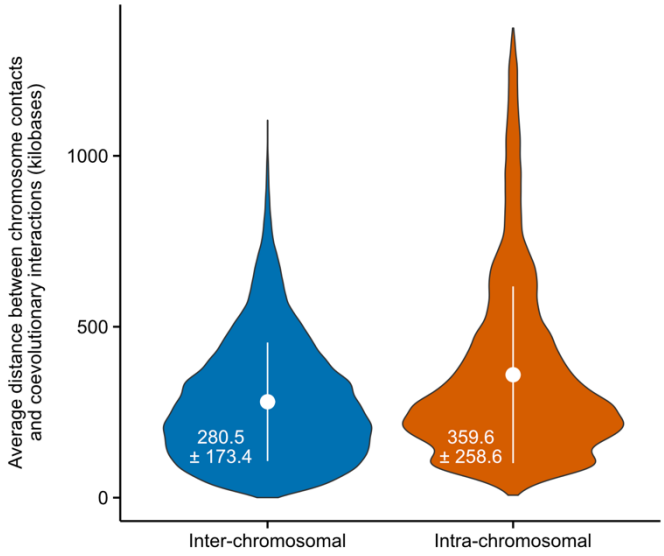


Figure S25. The genomic distribution of orthologous genes is not driving the signature of long distance intra-chromosomal coevolution. (A) In the genome of *S. cerevisiae*, there is no substantial association between a gene and the gene that it is most closely coevolving with (y-axis) and the closest gene in the data set (x-axis) (r = 0.09, p=0.004; Pearson Correlation). (B) Similarly, in *C. albicans*, there is not a significant association between the closest distance of a gene and an intra-chromosomal coevolving gene (y-axis) and the distance between the closest gene in the dataset (x-axis). Each data point represents a gene and the distance between either the closest intra-chromosomal gene that it is coevolving with (y-axis) and the distance between it and the closest gene in the data set (x-axis). Colors of each dot correspond to one chromosome. The color scheme is the same as depicted in Figure S21.



222

223 **Figure S26. Chromosome interactions do not influence signatures of coevolution.** For every
224 chromosomal interaction in *Saccharomyces cerevisiae*, the closest gene pair with a
225 coevolutionary signature was identified. The average distance of the closest chromosomal
226 interaction and coevolutionary signature was calculated and that distribution is shown here for
227 interchromosomal (left) and intrachromosomal (right) interactions. White circles depict the mean
228 and error bars are plus or minus one standard deviation. Mean and standard deviation values are
229 shown in white text. Although some distances between chromosomal interactions and coevolving
230 genes can be small, the average values observed are so great that chromosomal interactions do
231 not appear to influence signatures of gene-gene coevolution.

232

Supplementary Table Legends

Table S1. GO enrichment analysis for highly connected genes.

BP = biological processes; CC = cellular component; MF = molecular functions; e = enrichment; p = purifying; fdr_bh = false discovery rate, Benjamini-Hochberg.

Table S2. GO enrichment analysis across orthologous gene communities 1, 2, 3, and 4.

BP = biological processes; CC = cellular component; MF = molecular functions; e = enrichment; p = purifying; fdr_bh = false discovery rate, Benjamini-Hochberg.

Table S3. GO enrichment analysis per orthologous gene community.

BP = biological processes; CC = cellular component; MF = molecular functions; e = enrichment; p = purifying; fdr_bh = false discovery rate, Benjamini-Hochberg.

Table S4. Genes with only signatures of intra-chromosomal evolution.

Table S5. GO enrichment analysis of the 2,408 orthologous genes using *S. cerevisiae* and *C. albicans*.

BP = biological processes; CC = cellular component; MF = molecular functions; e = enrichment; p = purifying; fdr_bh = false discovery rate, Benjamini-Hochberg. Background is either *Saccharomyces cerevisiae* or *Candida albicans* genes, as specified by the column "Species."

Table S6. Enrichment of gene representation per chromosome in *S. cerevisiae* and *C. albicans*.

Benjamini-Hochberg multi-test correction was used for multi-test correction.