A 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

ORCiDs

J.L. Steenwyk: 0000-0002-8436-595X

M.A. Phillips: 0000-0002-0781-3325

F. Yang: 0000-0001-5907-4539

S.S. Date 0000-0002-4086-110X

T.R. Graham 0000-0002-3256-2126

J. Berman: 0000-0002-8577-0084

C.T. Hittinger: 0000-0001-5088-7461

A. Rokas: 0000-0002-7248-6551

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

Running title: Budding yeast genetic coevolution network

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

Abstract

Gene coevolution—which refers to gene pairs whose evolutionary rates covary across speciation events—is often observed among functionally related genes. We present a comprehensive gene coevolution network inferred from the examination of nearly three million gene pairs from 332 budding yeast species spanning ~400 million years of eukaryotic evolution. Modules within the network provide insight into cellular and genomic structure and function, such as genetic pleiotropy, genes functioning in distinct cellular compartments, vesicle transport, and DNA replication. Examination of the phenotypic impact of network perturbation across 14 environmental conditions using deletion mutant data from the baker's yeast *Saccharomyces cerevisiae* suggests that fitness in diverse environments is impacted by gene neighborhood and gene connectivity. By mapping the network onto the chromosomes of *S. cerevisiae* and the opportunistic human pathogen *Candida albicans*, which diverged ~235 million years ago, we discovered that coevolving gene pairs are not clustered in either species; rather, they are most often located on different chromosomes or far apart on the same chromosome. The budding yeast gene coevolution network captures the hierarchy of eukaryotic cellular structure and function, provides a roadmap for genotype-to-phenotype discovery, and portrays the genome as an extensively linked ensemble of genes.

Introduction

Genetic networks can shed light on the functional associations between genes, the pathways they encode, and can elucidate genotype-to-phenotype relationships (Costanzo *et al.* 2010, 2016; Kuzmin *et al.* 2018). More broadly, genetic networks can also provide insight into complex polygenic relationships underlying cellular function (Costanzo *et al.* 2019). Given the rich information that can be derived from genetic networks, numerous approaches that capture some aspect(s) of the functional relationships of genes within a genome (e.g., gene coexpression, genetic interaction) have been developed to construct them (Lezon *et al.* 2006; Baryshnikova *et al.* 2013; Wisecaver *et al.* 2017). While these networks are highly informative, they depict functional relationships of genes within a single extant species or strain.

One exciting complementary, but relatively understudied, method for constructing genetic networks is gene coevolution. Gene coevolution refers to the covariation of relative evolutionary rates across speciation events (Sato *et al.* 2005; Clark *et al.* 2012). Genetic networks based on gene coevolution differ from standard genetic networks in that they rely on the correlation of evolutionary rates of genes across a lineage of species rather than on correlation of functional data (e.g., gene coexpression, fitness of gene deletion mutant vs. wild-type) in a single strain or species. These gene coevolution networks also capture functional associations between genes (e.g., 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 (Findlay *et al.* 2014; Brunette *et al.* 2019). 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 (Brunette *et al.* 2019). Furthermore, four out of five proteins in the protein structural interactome map—a database of structural domain-domain interactions in the protein data bank (https://www.rcsb.org/)—exhibit signatures of gene coevolution (Kim *et al.* 2004). Although these and other studies have demonstrated that signatures of coevolution are a powerful method to detect functional associations among genes (Clark *et al.* 2012; Findlay *et al.* 2014; Raza *et al.* 2019; Brunette *et al.* 2019; Huang *et al.* 2020; Talsness *et al.* 2020), 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 gene coevolutionary networks, we constructed a gene coevolution network from a densely sampled set of orthologs from one-third of known budding yeast species (332 species) that diversified over ~400 million years. The inferred gene coevolution network provides a hierarchical view of cellular function from bioprocesses to pathways. Interpolation of the gene coevolution network with of fitness assay data from single- and digenic *S. cerevisiae* mutants (Costanzo *et al.* 2010, 2016, 2021; Usaj *et al.* 2017) provides insight into subnetwork- and gene-specific potential to buffer genetic perturbations. Mapping the 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. Our findings reveal general principles of eukaryotic gene coevolution networks and 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 genes from 2,408 orthologs among 332 budding yeast species, identifying 60,305 significant signatures of coevolution ($r \geq 0.825$; Pearson correlation among relative evolutionary rates; Fig 1A, 1B, and S1). From these 60,305 pairs of significantly coevolving genes, we constructed a network where nodes are genes and edges connect genes that are significantly coevolving (Fig. 1C). Examination of network properties across a wide range of thresholds of significant coevolution (Pearson correlation coefficient range of [0.600-0.900] with a step of 0.005) revealed broad network stability (Fig. S2), suggesting that our conservative significance threshold does not significantly impact the inferred network.

To determine how gene connectivity varied in the network, we examined patterns of dense and sparse connections among single genes. Single genes coevolved with a median of eight other genes, but gene connectivity varied (Fig. S3). For example, 1,091 genes have signatures of coevolution with five or fewer other genes and 601 genes are singletons (i.e., genes with no signatures of coevolution with other genes). In contrast, 420 genes have signatures of coevolution with 100 or more other genes, and 21 genes coevolve with 400 or more other genes.

Coevolving gene pairs in the network are functionally related. For example, *PEX1* and *PEX6* encode a heterohexameric complex responsible for protein transport across peroxisomal membranes (Ciniawsky *et al.* 2015) and are significantly coevolving (Fig. S4). In humans, mutations in either gene can lead to severe peroxisomal disorders (Reuber *et al.* 1997), suggesting their coevolution may be driven by selection. More broadly, functional enrichment among densely connected 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 (Ayala *et al.* 2018), are coevolving with 400 or more other genes. Chromatin remodelers and other densely connected genes are coevolving with genes that span multiple function categories, including genes with known pleiotropic effects (Fig. S5, Table S1) (Rando and Winston 2012). These data support the hypothesis that coevolving genes tend to have similar functions and suggest that genes coevolving with hundreds of others exhibit are highly pleiotropic.

To determine how gene connectivity varied among groups of genes, we examined the properties of subnetworks across genes considered essential and nonessential in the model yeast *Saccharomyces cerevisiae* or the opportunistic pathogen *Candida albicans* (Winzeler *et al.* 1999; Segal *et al.* 2018). Essential genes are densely connected in the gene coevolutionary network, whereas nonessential genes exhibit sparser connections (Fig. 2A-D). To infer network communities—clusters of genes that have more connections between them than between genes of different clusters—we used a hierarchical agglomeration algorithm (Fig. 2A). Five large communities (clusters of 11 or more genes) were identified. Each community varied in size, community-to-community connectivity, and essential/nonessential gene composition. Specifically, the two largest communities, communities 1 and 2, share the most connections and belong to a higher-order cluster with the next two largest communities, communities 3 and 4 (Fig. 2E and S6). In contrast, the smallest community, community 5, does not cluster with the other communities. Similarly, essential genes are overrepresented in community 1 but are underrepresented in communities 2, 3, and in smaller communities of 10 or fewer genes (Fig. 2F; $p < 0.01$ for all tests; Fisher's exact test). The result that essential genes are central hubs in the

gene coevolution network of the budding yeast subphylum mirrors observations from the *S. cerevisiae* genetic interaction network (Costanzo *et al*. 2016). This suggests that essential genes' functions as central hubs in yeast species have been conserved for hundreds of millions of years.

From processes to pathways: a hierarchy of cellular function

To gain insight into the functional neighborhoods of the gene coevolution network, the functional composition of each community was examined via gene ontology (GO) enrichment analysis (GeneOntologyConsortium 2004). This analysis revealed that the network captures the hierarchy of cellular function. For example, among the highest-order cluster of communities (i.e., communities 1 through 4), higher-order cellular processes including nucleic acid metabolism ($p = 0.040$; Fisher's exact test) and cellular anatomical entities ($p = 0.020$; Fisher's exact test) are enriched. At the individual community level, community 1 is enriched in genes with helicase activity ($p = 0.005$; Fisher's exact test), ligase activity ($p = 0.004$; Fisher's exact test), and translation initiation factors ($p = 0.024$; Fisher's exact test); community 2 is enriched in Golgi vesicle transport genes ($p = 0.009$; Fisher's exact test); whereas singletons are enriched in GTPase activity ($p = 0.016$; Fisher's exact test) and peroxiredoxin activity ($p = 0.036$; Fisher's exact test) (Fig. 2G-I, Table S3).

Examination of genes involved in biological functions, cellular compartments, and complexes revealed that genes within each functional category exhibited extensive signatures of coevolution. For example, genes involved in ribosome biogenesis, rRNA processing, and translation, which represent different functional categories, are extensively coevolving (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 genes underlying these fundamental cellular processes. Similarly, genes involved in nuclear processes or located in the cytoplasm tend to coevolve with genes in the same cellular compartment, however, substantial signatures of coevolution between cellular compartments are also observed (Fig. S7B). These results expand our understanding of the complex interplay among biological functions or cellular compartments over evolutionary time.

Strong signatures of coevolution were also observed among genes from specific processes, pathways, and complexes. For example, DNA replication genes have more coevolving gene pairs than expected by chance ($p < 0.001$; permutation test) (Fig. S8). Genes involved in DNA mismatch repair and nucleotide excision repair pathways—which participate in the repair of DNA lesions—are also significantly coevolving ($p < 0.001$ for each pathway; permutation test). Genes in the phosphatidylcholine biosynthesis pathway, which is responsible for the biosynthesis of the major phospholipid in organelle membranes, and 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, 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 have more significant coevolving gene pairs 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 (see *Methods*). These results suggest that the coevolution network captures a hierarchy of cellular function from broad bioprocesses to specific pathways.

To determine similarities and differences between the gene coevolution network and the genetic interaction network inferred from digenic null mutants in *S. cerevisiae* (Costanzo *et al.* 2010; Usaj *et al.* 2017), both data types were integrated into a single network (Fig. S10 and S11). Broad patterns in the gene coevolution network were reinforced as evidenced by tighter community clustering; however, gene-wise connectivity at times differed suggesting each network harbors distinct and complementary insights (Fig. S10). For example, gene connectivity is similar in both networks for the gene *CDC6*, which is required for DNA replication (Hartwell *et al.* 1970). More specifically, *CDC6* is connected to 96 genes in both networks and 56 of the genes are the same. This result suggests that the connectivity of *CDC6* is important across macroevolutionary timescales and among extant yeast species. In contrast, different gene-wise connectivity was observed for *CKI1*, a gene that encodes a choline kinase that has a paralog, *EKI1*, that arose from an ancient whole genome duplication event in the lineage that includes *S. cerevisiae* and relatives (Hosaka *et al.* 1989; Kim *et al.* 1999). Specifically, *CKI1* is coevolving

with 87 genes, has a significant genetic interaction with 10 genes, and seven of these genes are shared by both networks. Differences in gene-wise connectivity of *CKI1* in the coevolution-based network may in part be driven by the whole genome duplication event and the innovation of a paralog, *EKI1*, that is distinct to the evolutionary history of *S. cerevisiae* and close relatives (Marcet-Houben and Gabaldón 2015; Wolfe 2015).

Communities differ in capacity to compensate for perturbation

Examinations of genome-wide gene dispensability in the model budding yeast *S. cerevisiae* and the opportunistic pathogen *Candida albicans* (Winzeler *et al.* 1999; Segal *et al.* 2018) suggest that single-organism genetic networks can buffer perturbations resulting from the deletion of individual genes. However, the degree of gene dispensability varies between genes and is dependent, in part, on gene function (Albalat and Cañestro 2016). We therefore hypothesized that community and gene connectivity will impact fitness. To test this, we integrated information from the budding yeast 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 (Costanzo *et al.* 2021) (Fig. S12 and S13). We found that fitness was significantly associated with specific communities and environments ($p < 0.001$ for both comparisons, Multi-factor ANOVA). Examination of the potential interaction between community, gene connectivity, and environment revealed that fitness in diverse environments is dependent on community and the number of coevolving genes per gene (Fig. 4). These results indicate that the relationship between fitness across environments in single-gene knockout strains is significantly associated with the network neighborhood (i.e., community) and how that gene is connected to other genes (i.e., the number of coevolving genes for the deleted gene).

To further investigate the impact of network perturbation, we integrated genetic interaction data from double-gene or digenic deletion fitness assays, wherein positive and negative genetic interactions refer to less and more severe fitness defects than expected from single-gene deletions, respectively (Costanzo *et al.* 2010, 2016; Usaj *et al.* 2017). In contrast to single-gene knockouts, digenic knockouts were rarely associated with positive phenotypic outcomes (Fig. S14A), suggesting digenic gene losses are more difficult to buffer. However, instances of positive phenotypic impact among digenic gene losses were observed among genes from the

same community (e.g., deleting two genes in community 1), but were less frequently observed among genes from different communities except for digenic gene losses in community 1 and 2, which are highly connected. These results suggest that digenic gene losses are difficult to buffer, however, if both genes belong to the same larger cluster of genes (or community) compensatory effects are more likely to be observed.

Finally, to examine evolutionary gene loss in the context of the gene coevolution network, we investigated community-wide patterns of gene losses among genes lost in a lineage of budding yeasts previously reported to have undergone extensive gene losses (Steenwyk *et al.* 2019). These analyses revealed community 2 and singletons are more likely to be lost (Fig. S14B), which supports the hypothesis that gene losses do not occur stochastically (Albalat and Cañestro 2016).

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 (Hurst *et al.* 2004; Rokas *et al.* 2018). Thus, we hypothesized that coevolving 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 complete chromosome-level assemblies of *S. cerevisiae* and *C. albicans* (Goffeau *et al.* 1996; Jones *et al.* 2004), two species that diverged ~235 million years ago (Shen *et al.* 2018). The distinct evolutionary histories of *S. cerevisiae* and *C. albicans* are marked by whole-genome duplication and intra-species hybridization (Marcet-Houben and Gabaldón 2015; Mixão and Gabaldón 2020), respectively, which contribute to differences in chromosome number (16 in *S. cerevisiae* vs. 8 in *C. albicans*) and a lack of macrosynteny (Seoighe *et al.* 2000; Chibana *et al.* 2005; Wolfe 2006; Fitzpatrick *et al.* 2010; Dujon 2010) (Fig. 5A-B and Fig. S15-S16).

Contradictory to our hypothesis, we observed extensive inter-chromosomal and long-range intra-chromosomal gene coevolution (Fig. 5 and Fig. S17-S23). Specifically, co-evolving 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) and the number of inter-chromosomal signatures of coevolution (corrected by the number of genes on all other chromosomes) ($r = 0.95$, $p < 0.001$ for *S. cerevisiae*; $r = 0.98$, $p < 0.001$ for *C. albicans*; Spearman correlation). This result suggests that genes on the same or different chromosomes are equally like to be coevolving. Given the extensive coevolution among functionally related genes, 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 gene pair distances along the genome. Two genes coevolving on the same chromosome can be kilobase-to-megabase distances from one another (Fig. 5G-H). The distribution of the closest distance between a 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, 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 (Fig. S24). Examination of the distances between genes in our dataset and coevolving genes revealed that long-range intra-chromosomal coevolution was not an artifact of gene sampling (Fig. S24). Investigation of the interplay between gene coevolution and chromosomal contacts using a three-dimensional model of the *S. cerevisiae* genome (Duan *et al.* 2010) 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 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 (Gibbons *et al*. 2014, 2015). Our observations suggest extensive inter-chromosomal and long-range intra-chromosomal functional associations may be more common than previously appreciated.

Discussion

We constructed a network based on gene coevolution from a densely sampled set of orthologs across the budding yeast subphylum. Analyses of the global network, communities therein, and signatures of gene coevolution among bioprocesses, complexes, and pathways reveals that the network reflects a hierarchy of cellular function.

Based on these findings, we propose four general principles that influence coevolutionary network structure. First, hubs of genes tend to have similar function (Fig. 2, S4, and S7-9), which corroborates previous findings that genes with shared function, are co-expressed, or belong to the same multimeric complex tend to coevolve (Clark *et al.* 2012). Second, gene essentiality significantly impacts gene connectivity: essential genes are more densely connected than nonessential genes (Fig. 2), consistent with observations from single-organism genetic networks (Mnaimneh *et al.* 2004; Costanzo *et al.* 2010, 2016, 2021). This finding suggests that genes more important to cellular function are more likely central hubs in the coevolution network. Third, the impact of network perturbation (e.g., deletion of a gene) is dependent on community and gene-wise connectivity across diverse environments. For example, deletion of single genes from community 1 had the greatest negative fitness impact compared to single-gene deletions in other communities (Fig. 4A). Among digenic gene losses, compensatory potential is largely observed only when both genes belong to the same community (Fig. S12). Fourth, gene function, not location, drives gene coevolution over macroevolutionary timescales. To date, there have been few examples of inter-chromosomal gene associations (Gibbons *et al.* 2015), but mapping of the gene coevolution network onto budding yeast chromosomes reveals extensive signatures of inter- and long-range intra-chromosomal coevolution (Fig. 5, S21-S22). These results uncover a previously underappreciated degree of genome-wide coevolution that has been maintained over millions of years of budding yeast evolution, raising the hypothesis that eukaryotic genomes are best viewed as extensively linked ensembles of genes.

Our findings elucidate the general principles of coevolutionary networks and 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 (Shen *et al.* 2018). These 2,408 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 genes (Table S6), suggesting each chromosome is equally represented in our dataset.

Next, we calculated covariation of relative evolutionary rates of all 2,898,028 gene 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 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 (Pellegrini *et al.* 1999)—but instead uses correlations in genes' relative evolutionary rates (Pazos and Valencia 2001; Clark *et al.* 2012; de Juan *et al.* 2013).

To implement the CovER pipeline, single gene trees constrained to the species topology were first inferred using IQ-TREE, v1.6.11 (Nguyen *et al.* 2015) (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, branch lengths were transformed into relative rates by correcting the gene tree branch length by the corresponding branch length in the species phylogeny (Sato *et al.* 2005; Clark *et al.* 2012; Chikina *et al.* 2016). Outliers (defined as having corrected branch lengths greater than five) are known to cause false positive correlations and were removed (Clark *et al.* 2012). Branch lengths

were then Z-transformed and a Pearson correlation coefficient was calculated for each gene pair. The CovER algorithm has been integrated into PhyKIT, a UNIX toolkit for phylogenomic analysis (Steenwyk *et al.* 2020).

Network construction

Complex interactions between gene pairs were further examined using a network wherein nodes represent genes and edges connect pairs of genes that are coevolving. Following our previous work (Steenwyk *et al.* 2020), we considered 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 gene 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) (Fig. S2). In all cases, we found that the choice of threshold had little impact on network structure.

Network substructure is commonly referred to as community structure and describes a set of genes that are more densely connected with each other but more sparsely connected with other sets (or communities) of genes. To identify the community structure of our global gene coevolution network, a hierarchical agglomeration algorithm that conducts greedy optimization of modularity was implemented (Newman 2004).

Enrichment analysis

To determine functional category enrichment among sets of genes, gene ontology (GO) enrichment analysis was conducted. To do so, a background set of GO annotations were curated from the 2,408 orthologous genes (Shen *et al.* 2018). Specifically, for an orthologous group of genes, GO associations were mapped from the representative gene from *S. cerevisiae* (Goffeau *et al.* 1996). If an *S. cerevisiae* gene was not present, the annotation from the representative gene from *C. albicans* was chosen (Jones *et al.* 2004). 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 (Klopfenstein *et al.* 2018). 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 (Supek *et al.* 2011). Over and underrepresentation of essential genes across 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 (Kanehisa *et al.* 2016) 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 gene pairs 10,000 times and then determining the number of coevolving gene pairs among genes pairs for 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 gene coevolution network (Costanzo *et al.* 2010, 2016, 2021; Usaj *et al.* 2017). For example, the relationship between gene-wise community, connectivity, and fitness in diverse environments was evaluated. To determine if genes were equally likely to be lost across communities, we examined patterns of gene losses in *Hanseniaspora* spp., which have undergone extensive gene loss compared to other budding yeasts (Steenwyk *et al.* 2019).

Projecting the network onto genome structure and organization

To gain insight into the relationship between genome structure and the gene coevolution network, we projected the network onto the complete chromosome genome assemblies of *S. cerevisiae* and *C. albicans* (Goffeau *et al.* 1996; Jones *et al.* 2004; van het Hoog *et al.* 2007;

Muzzey *et al.* 2013). Prior to mapping the network onto the genome assemblies, we investigated genome-wide synteny using orthology information from the Candida Gene Order Browser (Fitzpatrick *et al.* 2010). Thereafter, the network was projected onto each genome assembly using Circos, v0.69 (Krzywinski *et al.* 2009). Examination of the distance between coevolving genes and chromosomal contacts was conducted using a three-dimensional model of the *S. cerevisiae* genome (Duan *et al.* 2010).

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 the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R56AI146096). 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.

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).

References

Albalat R., and C. Cañestro, 2016 Evolution by gene loss. Nat. Rev. Genet. 17: 379–391. https://doi.org/10.1038/nrg.2016.39

Ayala R., O. Willhoft, R. J. Aramayo, M. Wilkinson, E. A. McCormack, *et al.*, 2018 Structure and regulation of the human INO80–nucleosome complex. Nature 556: 391–395. https://doi.org/10.1038/s41586-018-0021-6

Baryshnikova A., M. Costanzo, C. L. Myers, B. Andrews, and C. Boone, 2013 Genetic Interaction Networks: Toward an Understanding of Heritability. Annu. Rev. Genomics Hum. Genet. 14: 111–133. https://doi.org/10.1146/annurev-genom-082509-141730

Brunette G. J., M. A. Jamalruddin, R. A. Baldock, N. L. Clark, and K. A. Bernstein, 2019 Evolution-based screening enables genome-wide prioritization and discovery of DNA repair genes. Proc. Natl. Acad. Sci. 116: 19593–19599. https://doi.org/10.1073/pnas.1906559116

Chibana H., N. Oka, H. Nakayama, T. Aoyama, B. B. Magee, *et al.*, 2005 Sequence Finishing and Gene Mapping for Candida albicans Chromosome 7 and Syntenic Analysis Against the Saccharomyces cerevisiae Genome. Genetics 170: 1525–1537. https://doi.org/10.1534/genetics.104.034652

Chikina M., J. D. Robinson, and N. L. Clark, 2016 Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. Mol. Biol. Evol. 33: 2182–2192. https://doi.org/10.1093/molbev/msw112

Ciniawsky S., I. Grimm, D. Saffian, W. Girzalsky, R. Erdmann, *et al.*, 2015 Molecular snapshots of the Pex1/6 AAA+ complex in action. Nat. Commun. 6: 7331. https://doi.org/10.1038/ncomms8331

Clark N. L., E. Alani, and C. F. Aquadro, 2012 Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res. 22: 714–720. https://doi.org/10.1101/gr.132647.111

Costanzo M., A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, *et al.*, 2010 The Genetic Landscape of a Cell. Science (80-.). 327: 425–431. https://doi.org/10.1126/science.1180823

Costanzo M., B. VanderSluis, E. N. Koch, A. Baryshnikova, C. Pons, *et al.*, 2016 A global genetic interaction network maps a wiring diagram of cellular function. Science (80-.). 353: aaf1420–aaf1420. https://doi.org/10.1126/science.aaf1420

Costanzo M., E. Kuzmin, J. van Leeuwen, B. Mair, J. Moffat, *et al.*, 2019 Global Genetic

Networks and the Genotype-to-Phenotype Relationship. Cell 177: 85–100. https://doi.org/10.1016/j.cell.2019.01.033

Costanzo M., J. Hou, V. Messier, J. Nelson, M. Rahman, *et al.*, 2021 Environmental robustness of the global yeast genetic interaction network. Science (80-.). 372: eabf8424. https://doi.org/10.1126/science.abf8424

Duan Z., M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, *et al.*, 2010 A three-dimensional model of the yeast genome. Nature 465: 363–367. https://doi.org/10.1038/nature08973

Dujon B., 2010 Yeast evolutionary genomics. Nat. Rev. Genet. 11: 512–524. https://doi.org/10.1038/nrg2811

Findlay G. D., J. L. Sitnik, W. Wang, C. F. Aquadro, N. L. Clark, *et al.*, 2014 Evolutionary Rate Covariation Identifies New Members of a Protein Network Required for Drosophila melanogaster Female Post-Mating Responses, (J. Zhang, Ed.). PLoS Genet. 10: e1004108. https://doi.org/10.1371/journal.pgen.1004108

Fitzpatrick D. A., P. O'Gaora, K. P. Byrne, and G. Butler, 2010 Analysis of gene evolution and metabolic pathways using the Candida Gene Order Browser. BMC Genomics 11: 290. https://doi.org/10.1186/1471-2164-11-290

GeneOntologyConsortium, 2004 The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32: 258D – 261. https://doi.org/10.1093/nar/gkh036

Gibbons J. G., A. T. Branco, S. Yu, and B. Lemos, 2014 Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. Nat. Commun. 5: 4850. https://doi.org/10.1038/ncomms5850

Gibbons J. G., A. T. Branco, S. A. Godinho, S. Yu, and B. Lemos, 2015 Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proc. Natl. Acad. Sci. U. S. A. 112: 2485–2490. https://doi.org/10.1073/pnas.1416878112

Goffeau A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, *et al.*, 1996 Life with 6000 Genes. Science (80-.). 274: 546–567. https://doi.org/10.1126/science.274.5287.546

Hartwell L. H., J. Culotti, and B. Reid, 1970 Genetic control of the cell-division cycle in yeast. I. Detection of mutants. Proc. Natl. Acad. Sci. U. S. A. 66: 352–9. https://doi.org/10.1073/pnas.66.2.352

Hoog M. van het, T. J. Rast, M. Martchenko, S. Grindle, D. Dignard, *et al.*, 2007 Assembly of the Candida albicans genome into sixteen supercontigs aligned on the eight chromosomes.

Genome Biol. 8: R52. https://doi.org/10.1186/gb-2007-8-4-r52

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

Huang J.-W., A. Acharya, A. Taglialatela, T. S. Nambiar, R. Cuella-Martin, *et al.*, 2020 MCM8IP activates the MCM8-9 helicase to promote DNA synthesis and homologous recombination upon DNA damage. Nat. Commun. 11: 2948. https://doi.org/10.1038/s41467-020-16718-3

Hurst L. D., C. Pál, and M. J. Lercher, 2004 The evolutionary dynamics of eukaryotic gene order. Nat. Rev. Genet. 5: 299–310. https://doi.org/10.1038/nrg1319

Jones T., N. A. Federspiel, H. Chibana, J. Dungan, S. Kalman, *et al.*, 2004 The diploid genome sequence of Candida albicans. Proc. Natl. Acad. Sci. 101: 7329–7334. https://doi.org/10.1073/pnas.0401648101

Juan D. de, F. Pazos, and A. Valencia, 2013 Emerging methods in protein co-evolution. Nat. Rev. Genet. 14: 249–261. https://doi.org/10.1038/nrg3414

Kanehisa M., Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, 2016 KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44: D457–D462. https://doi.org/10.1093/nar/gkv1070

Kim K., K. H. Kim, M. K. Storey, D. R. Voelker, and G. M. Carman, 1999 Isolation and characterization of the Saccharomyces cerevisiae EKI1 gene encoding ethanolamine kinase. J. Biol. Chem. 274: 14857–66. https://doi.org/10.1074/jbc.274.21.14857

Kim W. K., D. M. Bolser, and J. H. Park, 2004 Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). Bioinformatics 20: 1138–1150. https://doi.org/10.1093/bioinformatics/bth053

Klopfenstein D. V., L. Zhang, B. S. Pedersen, F. Ramírez, A. Warwick Vesztrocy, *et al.*, 2018 GOATOOLS: A Python library for Gene Ontology analyses. Sci. Rep. 8: 10872. https://doi.org/10.1038/s41598-018-28948-z

Krzywinski M., J. Schein, I. Birol, J. Connors, R. Gascoyne, *et al.*, 2009 Circos: An information aesthetic for comparative genomics. Genome Res. 19: 1639–1645. https://doi.org/10.1101/gr.092759.109

Kuzmin E., B. VanderSluis, W. Wang, G. Tan, R. Deshpande, *et al.*, 2018 Systematic analysis of

complex genetic interactions. Science (80-.). 360: eaao1729. https://doi.org/10.1126/science.aao1729

Lezon T. R., J. R. Banavar, M. Cieplak, A. Maritan, and N. V. Fedoroff, 2006 Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. Proc. Natl. Acad. Sci. 103: 19033–19038. https://doi.org/10.1073/pnas.0609152103

Marcet-Houben M., and T. Gabaldón, 2015 Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage, (L. D. Hurst, Ed.). PLOS Biol. 13: e1002220. https://doi.org/10.1371/journal.pbio.1002220

Mixão V., and T. Gabaldón, 2020 Genomic evidence for a hybrid origin of the yeast opportunistic pathogen Candida albicans. BMC Biol. 18: 48. https://doi.org/10.1186/s12915-020-00776-6

Mnaimneh S., A. P. Davierwala, J. Haynes, J. Moffat, W.-T. Peng, *et al.*, 2004 Exploration of Essential Gene Functions via Titratable Promoter Alleles. Cell 118: 31–44. https://doi.org/10.1016/j.cell.2004.06.013

Muzzey D., K. Schwartz, J. S. Weissman, and G. Sherlock, 2013 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. https://doi.org/10.1186/gb-2013-14-9-r97

Newman M. E. J., 2004 Fast algorithm for detecting community structure in networks. Phys. Rev. E 69: 066133. https://doi.org/10.1103/PhysRevE.69.066133

Nguyen L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh, 2015 IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32: 268–274. https://doi.org/10.1093/molbev/msu300

Pazos F., and A. Valencia, 2001 Similarity of phylogenetic trees as indicator of protein–protein interaction. Protein Eng. Des. Sel. 14: 609–614. https://doi.org/10.1093/protein/14.9.609

Pellegrini M., E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates, 1999 Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Proc. Natl. Acad. Sci. 96: 4285–4288. https://doi.org/10.1073/pnas.96.8.4285

Rando O. J., and F. Winston, 2012 Chromatin and Transcription in Yeast. Genetics 190: 351–

387. https://doi.org/10.1534/genetics.111.132266

Raza Q., J. Y. Choi, Y. Li, R. M. O'Dowd, S. C. Watkins, *et al.*, 2019 Evolutionary rate covariation analysis of E-cadherin identifies Raskol as a regulator of cell adhesion and actin dynamics in Drosophila, (C. Desplan, Ed.). PLOS Genet. 15: e1007720. https://doi.org/10.1371/journal.pgen.1007720

Reuber B. E., E. Germain-Lee, C. S. Collins, J. C. Morrell, R. Ameritunga, *et al.*, 1997 Mutations in PEX1 are the most common cause of peroxisome biogenesis disorders. Nat. Genet. 17: 445–448. https://doi.org/10.1038/ng1297-445

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

Sato T., Y. Yamanishi, M. Kanehisa, and H. Toh, 2005 The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. Bioinformatics 21: 3482–3489. https://doi.org/10.1093/bioinformatics/bti564

Segal E. S., V. Gritsenko, A. Levitan, B. Yadav, N. Dror, *et al.*, 2018 Gene Essentiality Analyzed by In Vivo Transposon Mutagenesis and Machine Learning in a Stable Haploid Isolate of Candida albicans, (A. Di Pietro, Ed.). MBio 9. https://doi.org/10.1128/mBio.02048-18

Seoighe C., N. Federspiel, T. Jones, N. Hansen, V. Bivolarovic, *et al.*, 2000 Prevalence of small inversions in yeast gene order evolution. Proc. Natl. Acad. Sci. 97: 14433–14437. https://doi.org/10.1073/pnas.240462997

Shen X.-X., D. A. Opulente, J. Kominek, X. Zhou, J. L. Steenwyk, *et al.*, 2018 Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell 175: 1533-1545.e20. https://doi.org/10.1016/j.cell.2018.10.023

Steenwyk J. L., D. A. Opulente, J. Kominek, X.-X. Shen, X. Zhou, *et al.*, 2019 Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts, (S. Kamoun, Ed.). PLOS Biol. 17: e3000255. https://doi.org/10.1371/journal.pbio.3000255

Steenwyk J. L., T. J. Buida, A. L. LaBella, Y. Li, X.-X. Shen, *et al.*, 2020 PhyKIT: a UNIX shell toolkit for processing and analyzing phylogenomic data. bioRxiv 2020.10.27.358143. https://doi.org/10.1101/2020.10.27.358143

Supek F., M. Bosnjak, N. Skunca, and T. Smuc, 2011 REVIGO summarizes and visualizes long

lists of gene ontology terms. PLoS One 6: e21800. https://doi.org/10.1371/journal.pone.0021800

Talsness D. M., K. G. Owings, E. Coelho, G. Mercenne, J. M. Pleinis, *et al.*, 2020 A Drosophila screen identifies NKCC1 as a modifier of NGLY1 deficiency. Elife 9. https://doi.org/10.7554/eLife.57831

Usaj M., Y. Tan, W. Wang, B. VanderSluis, A. Zou, *et al.*, 2017 TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. G3: Genes|Genomes|Genetics 7: 1539–1549. https://doi.org/10.1534/g3.117.040220

Winzeler E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, *et al.*, 1999 Functional characterization of the S-cerevisiae genome by gene deletion and parallel analysis. Science (80-.). 285: 901–906. https://doi.org/10.1126/science.285.5429.901

Wisecaver J. H., A. T. Borowsky, V. Tzin, G. Jander, D. J. Kliebenstein, *et al.*, 2017 A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. Plant Cell 29: 944–959. https://doi.org/10.1105/tpc.17.00009

Wolfe K. H., 2006 Comparative genomics and genome evolution in yeasts. Philos. Trans. R. Soc. B Biol. Sci. 361: 403–412. https://doi.org/10.1098/rstb.2005.1799

Wolfe K. H., 2015 Origin of the Yeast Whole-Genome Duplication. PLOS Biol. 13: e1002221. https://doi.org/10.1371/journal.pbio.1002221

Figures and figure legends

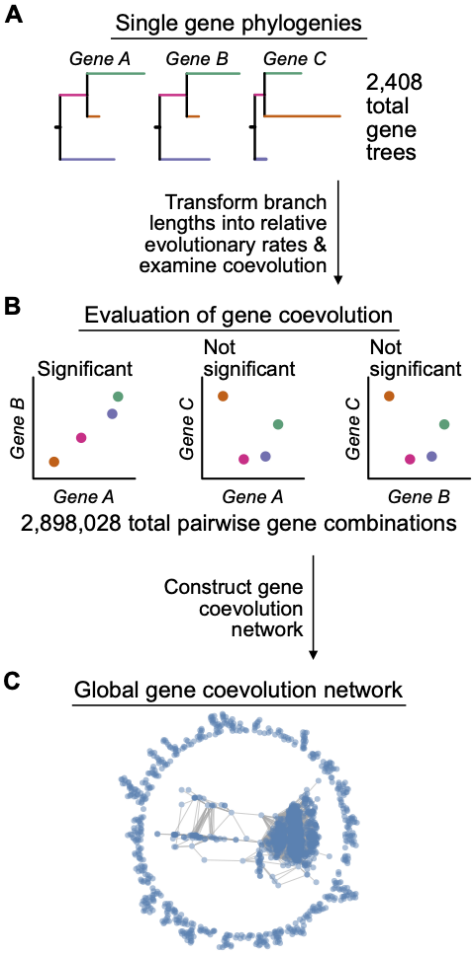

Figure 1. Constructing the budding yeast gene coevolution network. (A) We determined gene coevolution in a set of 2,408 single gene trees in which branch lengths were inferred along the species tree topology. (B) Gene coevolution was evaluated across all pairwise combinations of genes using the CovER function in PhyKIT, v0.1 (Steenwyk *et al.* 2020). (C) Significantly coevolving gene pairs were used to construct a global network of gene coevolution where nodes correspond to genes and significantly coevolving genes are connected by edges.

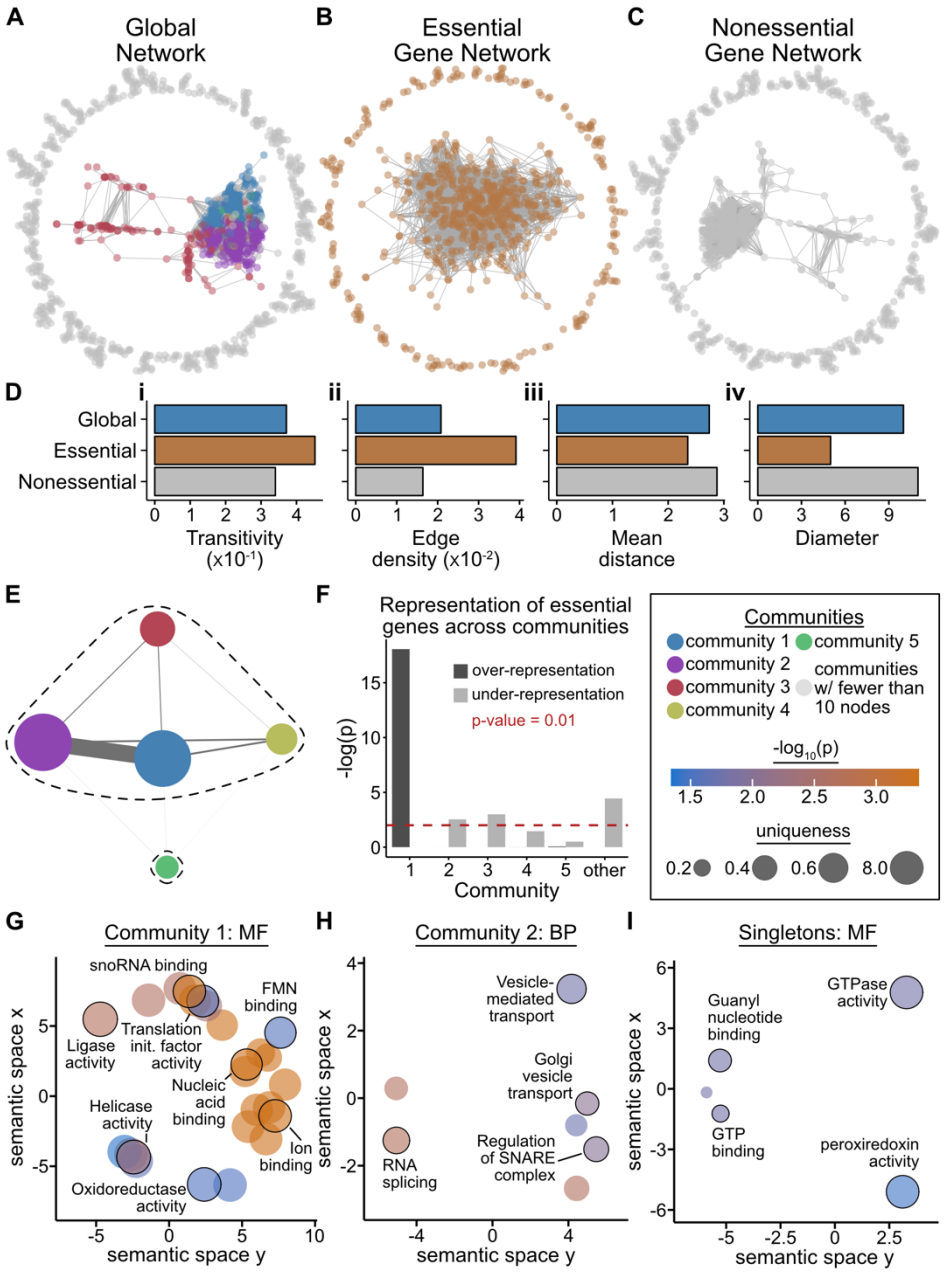

Figure 2. Network modules reflect modules of bioprocesses. (A) The global network of gene coevolution. (B) Essential gene and (C) nonessential gene subnetworks. (D) Examination of global, essential gene, and nonessential gene network properties reveals that the essential gene subnetwork has higher values for metrics of network density (i.e., transitivity and edge density), whereas the nonessential gene network has higher values for metrics that measure how diffuse the network is (i.e., mean distance and diameter). (E) Network community detection revealed five major subnetworks or communities. Genes from each community are depicted in the same

color in panel A and genes from small communities with 10 or fewer genes are depicted in gray. There are 804, 740, 161, 39, and 15 genes in the five largest communities (communities 1-5). Edge width reflects the number of co-evolving gene pairs between the two communities and node size reflects the number of genes in each community. Higher-order community clustering revealed communities 1 through 4 cluster together, whereas community 5 is a singleton as denoted by the dashed line. (F) Community 1 is overrepresented with essential genes whereas all other communities are underrepresented with essential genes. (G-I) Communities differ in enriched terms; for example, enriched molecular functions in community 1 includes genes associated with helicase, ligase, and translation initiation factor activities. MF and BP represent molecular functions and biological processes, respectively. Each enriched GO term is represented by a circle where circle color reflects $-\log_{10}$ p-value from a Fischer's exact test with Benjamini Hochberg multi-test correction and the size of each circle represents GO term uniqueness, a measure of GO term dissimilarity to other enriched GO terms wherein higher values reflect greater uniqueness. Complete enrichment results for each community are reported in Table S3. The box to the right of panel F is the legend for the whole figure.

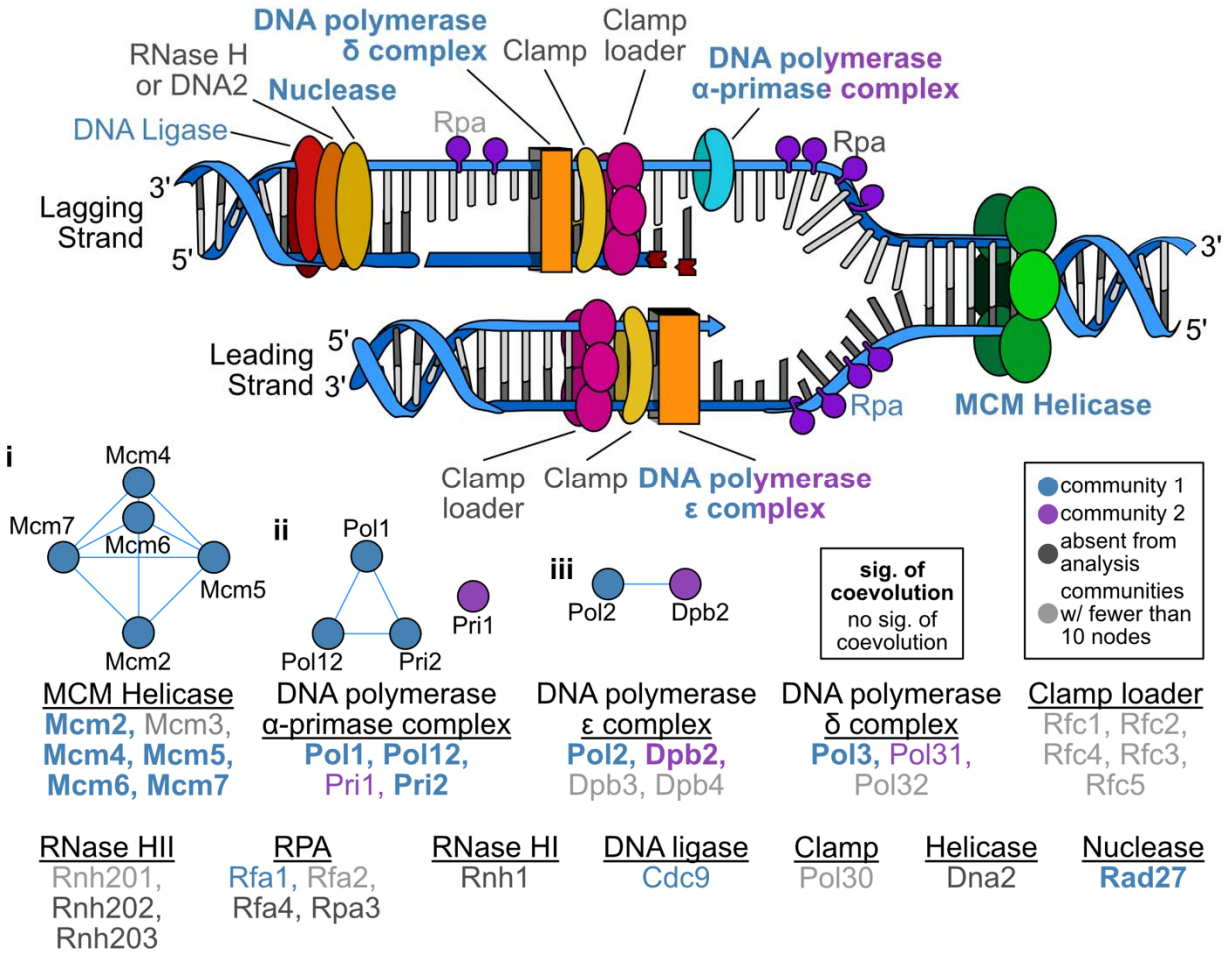

Figure 3. Extensive coevolution in DNA replication genes. Cartoon representation of DNA replication. Exemplary complex specific subnetworks are depicted in i, ii, and iii. (i) Extensive coevolution between genes that encode the helicase, minichromosome maintenance (MCM) complex, which functions as a helicase. (ii) Coevolution in the 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 genes that are significantly coevolving. Genes and complexes in bold have signatures of coevolution. Genes and complexes are colored according to community assignment. Complexes, such as the DNA polymerase α-primase complex, are depicted in multiple colors reflecting the multiple communities represented within the complex. There is significant coevolution across all DNA replication 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).

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	

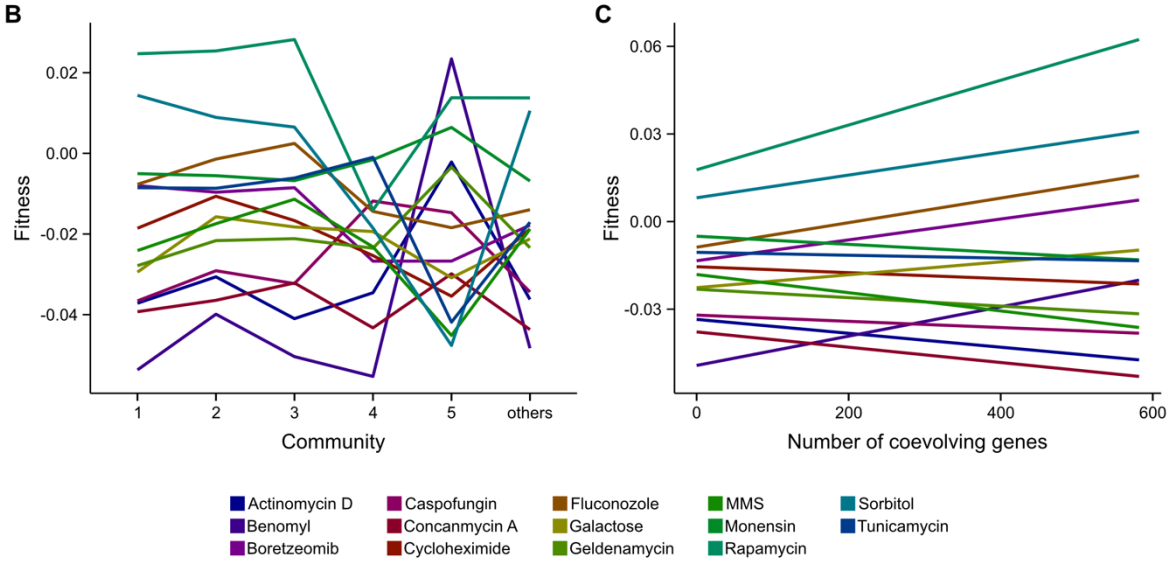

Figure 4. The impact of perturbing the gene coevolutionary network through single-gene deletion in diverse environments is dependent on community and gene connectivity. (A) Multi-factor ANOVA results indicate community, environment, the interaction between community and environment, and the interaction between environment and the number of coevolving genes per 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 community. (C) Similarly, fitness of single-gene deletion strains in diverse environments is impacted by the number of coevolving genes the deleted gene is connected to. These results indicate that fitness in diverse environments is impacted by the gene neighborhood and connectivity in the network. In both panels, each color corresponds to a different environment that fitness was measured in. Df represents degrees of freedom; Sum of Sq. represents sum of squares; Mean of Sq. represents Mean of squares.

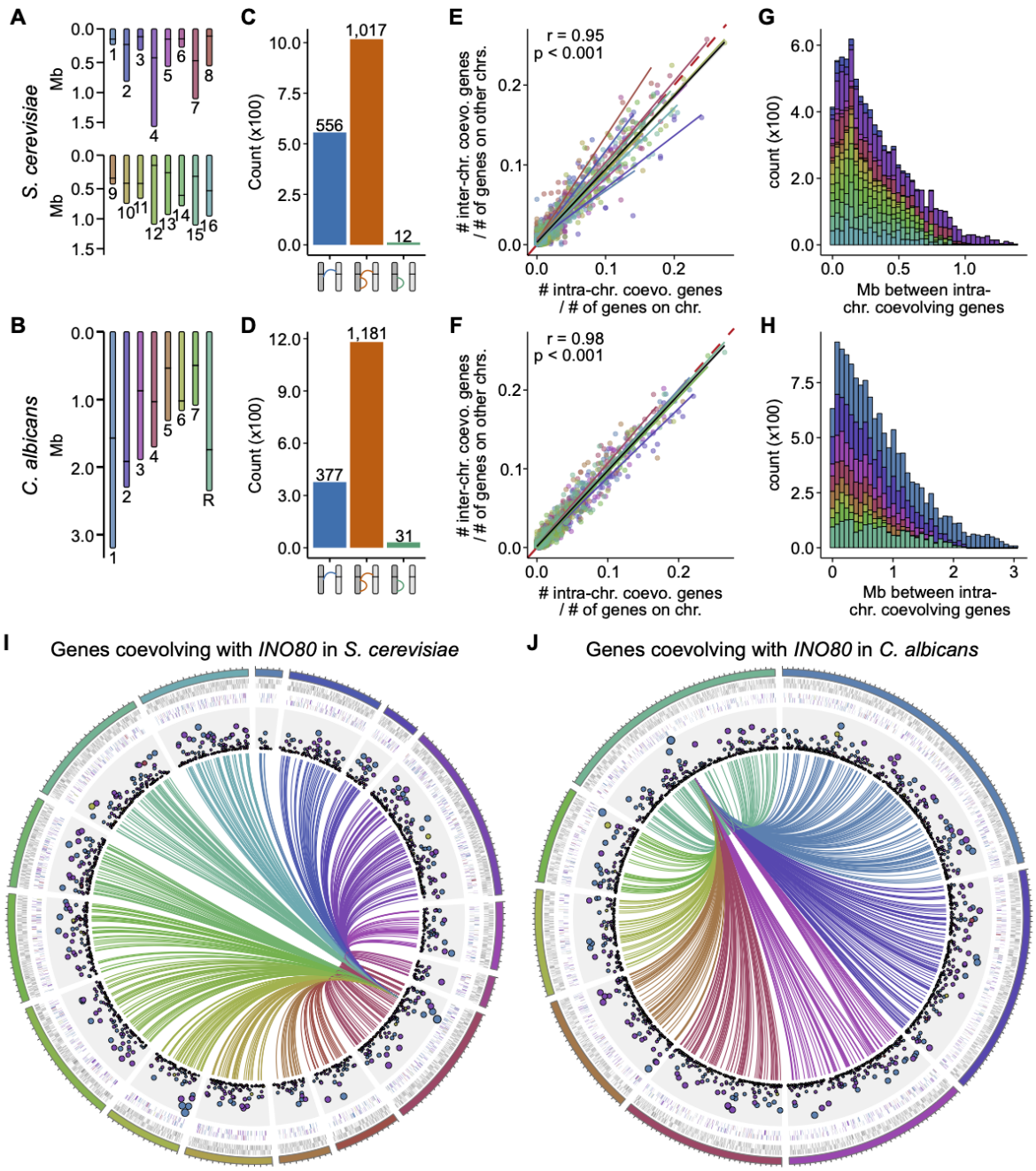

Figure 5. Extensive long range and inter-chromosomal gene coevolution. (A & B) The number and size of chromosomes differ in (A) *S. cerevisiae* and (B) *C. albicans*. Black lines indicate locations of centromeres. (C & D) The numbers of genes with only inter-chromosomal (blue), both inter- and intra-chromosomal (orange), or only intra-chromosomal (green) signatures of gene coevolution reveal substantially more inter-chromosomal associations than intra-chromosomal ones. (E & F) The relationship between the number of intra-chromosomal signatures of gene coevolution corrected by the number of genes on the same chromosome (x-

axis) and the number of inter-chromosomal signatures of gene coevolution corrected by the number of genes on other chromosomes (y-axis) reveals intra- and inter-chromosomal associations are equally likely, suggesting function, rather than genetic neighborhood, is the primary driver of gene coevolution. Each color represents a different chromosome. Regression lines are depicted for each chromosome using a linear model. A summary regression line is depicted in black. (G & H) Distribution of distances among intra-chromosomal signatures of gene coevolution reveals long-range coevolution is common. (I & J) An exemplary gene, *INO80*, reveals how a single gene can be coevolving with other genes on the same or other chromosomes. The first, innermost track depicts the various chromosomes of either yeast in which chromosome 1 is shown at the 12 o'clock position and increasing chromosome number is depicted in a clock-wise manner. The second track shows genes on the plus or minus strand. The third track shows the same data for the genes present in our dataset and each gene is colored according to the community they are part of. The scatter plot shows the number of coevolving genes per gene. Larger circles represent genes that are connected to more genes in the network. The links depict genes that are coevolving with *INO80* and each link is colored according to the chromosomal location of the other gene it is coevolving with. Colors in E-H as well as ideogram and link colors in J correspond chromosomes as depicted in A and B.