Parallel Concerted Evolution of Ribosomal Protein Genes in Fungi and Its Adaptive Significance

Alison Mullis,[†,1] Zhaolian Lu,[†,1] Yu Zhan,[†,1] Tzi-Yuan Wang,[2] Judith Rodriguez,[3] Ahmad Rajeh,[1,3] Ajay Chatrath,[1] and Zhenguo Lin [*,1]

[1]Department of Biology, Saint Louis University, St. Louis, MO

[2]Biodiversity Research Center, Academia Sinica, Nankang, Taipei, Taiwan

[3]Program of Bioinformatics and Computational Biology, Saint Louis University, St. Louis, MO

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: zhenguo.lin@slu.edu.

Associate editor: Claudia Russo

Abstract

Ribosomal protein (RP) genes encode structural components of ribosomes, the cellular machinery for protein synthesis. A single functional copy has been maintained in most of 78–80 RP families in animals due to evolutionary constraints imposed by gene dosage balance. Some fungal species have maintained duplicate copies in most RP families. The mechanisms by which the RP genes were duplicated and maintained and their functional significance are poorly understood. To address these questions, we identified all RP genes from 295 fungi and inferred the timing and nature of gene duplication events for all RP families. We found that massive duplications of RP genes have independently occurred by different mechanisms in three distantly related lineages: budding yeasts, fission yeasts, and Mucoromycota. The RP gene duplicates in budding yeasts and Mucoromycota were mainly created by whole genome duplication events. However, duplicate RP genes in fission yeasts were likely generated by retroposition, which is unexpected considering their dosage sensitivity. The sequences of most RP paralogs have been homogenized by repeated gene conversion in each species, demonstrating parallel concerted evolution, which might have facilitated the retention of their duplicates. Transcriptomic data suggest that the duplication and retention of RP genes increased their transcript abundance. Physiological data indicate that increased ribosome biogenesis allowed these organisms to rapidly consume sugars through fermentation while maintaining high growth rates, providing selective advantages to these species in sugar-rich environments.

Key words: **ribosomal proteins, gene duplication, gene conversion, concerted evolution, fermentation.**

Introduction

Gene duplication has served as a driving force for the evolution of new phenotypic traits and contributed to adaptation of organisms to their specific niches (Ohno 1970; Sidow 1996). Duplicate genes are mainly generated by chromosome or whole genome duplication (WGD), unequal crossing-over, and retroposition (Zhang 2013). Similar to other types of mutations, only a small portion of duplicate genes are eventually fixed in a population, and the survivors are usually advantageous to the organisms (Zhang 2003; Kondrashov and Kondrashov 2006). Highly diverse retention patterns of duplicate genes have been observed among gene families (Hahn et al. 2005). For instance, tens to hundreds of odorant receptor genes were found in metazoan genomes (Sanchez-Gracia et al. 2009). In contrast, many genes have been maintained as a single copy since the divergence of eukaryotes, such as the DNA repair genes *RAD51*, *MSH2*, and *MLH1* (Lin et al. 2006, 2007; Zeng et al. 2014).

Another notable example is the gene families encoding for cytosolic ribosomal proteins (RPs), which are the structural components of ribosomes. Ribosomes carry out one of the most fundamental processes of living systems by translating genetic information from mRNA into proteins. In eukaryotes, each ribosome consists of a small subunit and a large subunit. The two subunits comprise 78–80 different RPs and four types of ribosomal RNAs (rRNAs) (Wool 1979; Wimberly et al. 2000). RP genes are highly conserved in all domains of life (Korobeinikova et al. 2012). Each RP has unique amino acid sequences with very limited to no similarities between each other. For most animals studied, only a single functional gene is maintained in each RP family, though many processed pseudogenes may be found (Dudov and Perry 1984; Kuzumaki et al. 1987; Kenmochi et al. 1998). As structural components of the highly expressed macromolecular complex, the evolutionary constraints on duplicate RP genes were believed to be imposed by gene dosage balance (Birchler and Veitia 2012). Attributing to polyploidization or WGD events, multiple gene copies are usually present in each RP family in polyploid plants (Vision et al. 2000; Barakat et al. 2001). The retention of plant RP duplicates is probably because all RP genes were duplicated simultaneously by WGD, allowing maintenance of balanced RP dosages (Birchler and Veitia 2012).

Similar to polyploid plants, most RP families in a budding yeast *Saccharomyces cerevisiae* have duplicate copies due to a WGD event (Wolfe and Shields 1997; Kellis et al. 2004). Many RP ohnologs are more similar to each other than to their orthologous genes owing to interlocus gene conversion (Evangelisti and Conant 2010; Casola et al. 2012). During an interlocus gene conversion, one gene serves as a DNA donor that replaces the sequences of its paralogous gene (Chen et al. 2007). Gene conversion homogenized the sequences of paralogous genes so that the ancient duplicate events appear much more recent, which was called "concerted evolution" (Brown et al. 1972). One of the best-known examples of concerted evolution is the rRNA genes in both prokaryotes and eukaryotes (Arnheim et al. 1980; Schlotterer and Tautz 1994; Blattner et al. 1997).

According to the ribosomal protein gene database (RPG) (Nakao et al. 2004), three of ten fungal species listed have multiple gene copies in most RP families, including *S. cerevisiae*, a fission yeast *Sch. pombe*, and a pin mold *Rhizopus oryzae*. It was proposed that a WGD event has occurred during evolution of *R. oryzae* (Ma et al. 2009). The WGD might have contributed to the increased RP gene dosage. However, most RP families in *R. oryzae* have more than four gene copies, which cannot be explained by a single WGD. Unlike *S. cerevisiae* and *R. oryzae*, no WGD event has been detected during the evolution of *Sch. pombe* (Rhind et al. 2011), suggesting that each RP family might be duplicated independently by small-scale duplication events (SSDs). This observation is unexpected because the duplicates of genes encoding macromolecules generated by SSDs are much less likely to survive because they are sensitive to gene dosage balance (Li et al. 1996; Conant and Wolfe 2008). It remains obscure about how RP genes have been duplicated and maintained in fungi, particularly in the fission yeast *Sch. pombe*.

The expression of RP genes in yeast is tightly linked to growth and proliferation (Montagne et al. 1999; Jorgensen et al. 2002; Brauer et al. 2008). In rapidly growing yeast, ~50% of RNA polymerase II (Pol II) transcription events are devoted to RP expression (Warner 1999). Therefore, the duplication and retention of RP genes might have more functional impacts on these microorganisms than animals or plants. Like other types of mutations, the occurrence of gene duplication is mostly due to stochastic events, but the retention of duplicate genes would have been driven by natural selection (Panchy et al. 2016). A better understanding of evolutionary fates of RP duplicate genes could offer new insights into how gene duplication produced adaptive solutions to microorganisms.

To better understand the evolutionary patterns of RP genes and their adaptive significance, we conducted systematic identification and evolutionary analyses of all RP families in fungi. We searched for RP genes from 295 fungal species and identified independent duplications of most RP families in three distantly related fungal lineages. We inferred the timing and nature of gene duplication for each RP family in each fungal lineage. We found that a vast majority of RP paralogous genes have experienced repeated gene conversion events that have homogenized their sequences in each species. In aligning with integrative analyses of genomic, transcriptomic, and physiological data, we propose that the massive duplication, retention, and concerted evolution of RP genes have contributed to the evolution of fermentative lifestyle in these fungal species. This study offers a classic example illustrating the mechanisms and adaptive significance of maintaining duplicate genes encoding macromolecules.

Results

Massive Duplications of RP Genes Found in Three Distantly Related Fungal Lineages

To determine the prevalence of RP gene duplications in fungi, we first searched for RP homologous genes in all fungal species with NCBI Reference Sequence (RefSeq) protein data (supplementary table S1, Supplementary Material online). As of March 2019, 285 fungal species were annotated with RefSeq protein data, covering five of the seven fungal phyla. We conducted BLASTP searches against the 285 RefSeq protein data sets using amino acid sequences of RP genes from both *S. cerevisiae* and *Sch. pombe* as queries (see Materials and Methods section). Based on BLASTP search results, we calculated the gene copy numbers of each RP family for every examined species, and the total number of RP families with duplicate copies (supplementary table S1, Supplementary Material online).

We considered a species with massive RP duplications if more than 50% ($\geq$40) of RP families have duplicate copies. Among the 285 fungi examined, only ten species meet the criterion of massive RP duplications. The ten species distribute in three distantly related fungal lineages: three in the class of Saccharomycetes (budding yeasts), four in the class of Schizosaccharomycetes (fission yeasts), and three in the phylum of Mucoromycota (supplementary table S1, Supplementary Material online). Two Blast hits were found in most RP families in a budding yeast *Candida viswanathii*. Because the assembly type of the *C. viswanathii* genome is diploid, the two Blast hits represent different alleles instead of paralogous genes. Thus, *C. viswanathii* was not considered as a species with massive RP duplications.

Because protein annotations of a genome could be incomplete or inaccurate, manual curation is required for a more accurate survey of RP repertoire. It is necessary to carry out a second-round identification of RP genes with manual curation, focusing on the three fungal lineages. We selected 24 species from the three fungal lineages, including the ten species with massive RP duplication. To provide a more even distribution of taxonomic groups in each lineage, we included ten other species whose genomic data are available in NCBI Whole Genome Shotgun (WGS), Yeast Gene Order Browser (YGOB), and JGI (Byrne and Wolfe 2005; Maguire et al. 2013). In total, our second-round search examined 34 fungal species, consisting of 23 Saccharomycetes species, five Taphrinomycotina species (including the four fission yeasts), and six Mucoromycota species (fig. 1 and supplementary table S2, Supplementary Material online). The phylogenetic relationships of the 34 species were inferred using the amino

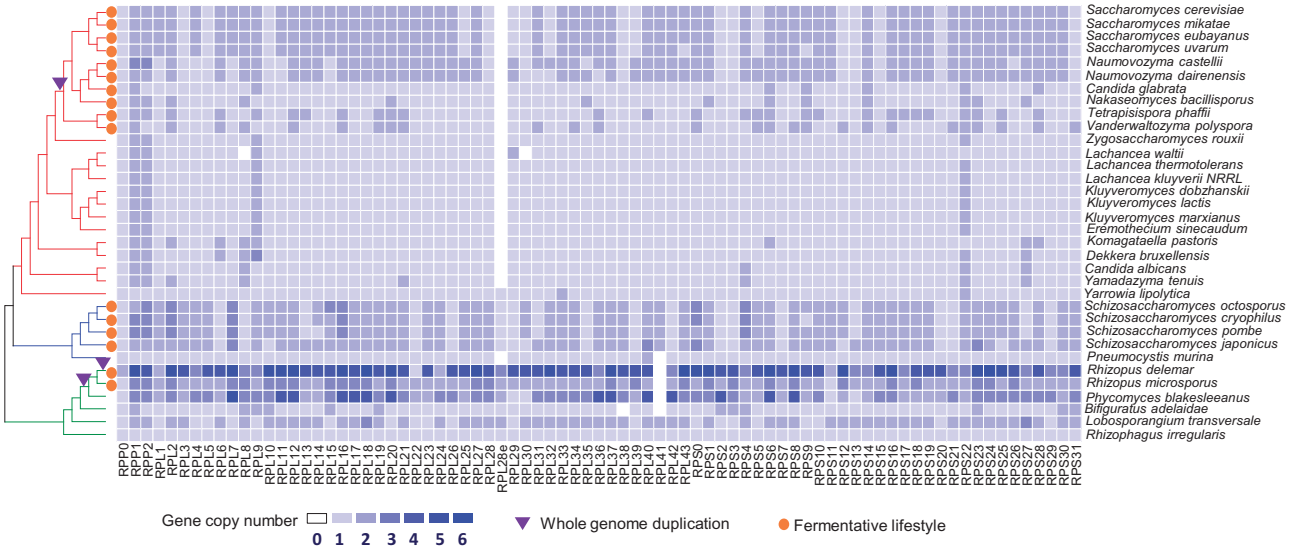

FIG. 1. Schematic illustration of gene duplication patterns of 79 RP families in 34 fungal species. Each row represents a fungal species, and each column represents an RP gene family. The color of a cell represents the numbers of gene copies identified in an RP family in a species. The evolutionary relationship for 34 species, inferred based on amino acid sequences of RNA Pol II were shown to the left side of the matrix. The species names were provided to the right of the matrix.

acid sequences of the largest subunit of RNA Pol II proteins (supplementary fig. S1, Supplementary Material online). Including the 285 species analyzed in our first-round analysis, we have examined a total number of 295 fungal genomes, representing the largest scale of RP repertoire survey in fungi to our knowledge.

To manually curate RP repertoire in a genome, we performed both BLASTP and TBlastN searches for each of the 34 fungal species. By comparing BLASTP and TBlastN search results, we identified discrepancies in the number of RP genes and aligned regions. We found that many TBlastN hits were absent in BLASTP searches, indicating the presence of unannotated RP genes. Thus, we have manually predicted 259 novel RP genes from 32 of the 34 species. We also revised the annotations of open reading frame (ORF) for 95 RP genes. In total, we identified 3,950 RP genes from the 34 fungal species (supplementary tables S2 and S3, Supplementary Material online).

We constructed maximum likelihood (ML) phylogenetic trees for each RP family (see Materials and Methods section). Similar tree topologies were observed among RP families with duplicate copies (supplementary file 1, Supplementary Material online). For instance, two copies of *RPL6* genes are present in all ten post-WGD budding yeasts and all four fission yeast species (fig. 2A). In Mucoromycota, 2–5 copies of *RPL6* genes are present in *Lobosporangium transversale*, *Phycomyces blakesleeanus*, *Rhizopus microsporus*, and *Rhizopus delemar* (former name *R. oryzae*). According to the ML tree (fig. 2A), the *RPL6* paralogous genes are more closely related to each other in each species than to their orthologous genes. Similar patterns are present in *RPS19* (fig. 2B) as well as many other RP families (supplementary file 1, Supplementary Material online). The tree topologies suggest that these RP paralogous genes have been independently duplicated in each species since their divergence. However, at least in the post-WGD budding yeasts, it has been documented that *RPL6* and *RPS19* were generated by the WGD event occurred prior to the divergence of *S. cerevisiae* and *Vanderwaltozyma polyspora* (Conant and Wolfe 2006). Therefore, the phylogenetic trees do not accurately depict the evolutionary history of *RPL6* and *RPS19* families in budding yeasts. It has been shown that *RPL6* and *RPS19* genes have experienced gene conversion during the evolution of *S. cerevisiae*, which explains the discrepancy (Evangelisti and Conant 2010; Casola et al. 2012). However, it is unknown whether it is the same case in the fission yeasts and Mucoromycota species. Resolving this problem requires accurate timing of duplication events in the two lineages. Because only a small number of species in the three fungal lineages have experienced massive RP duplications (fig. 1 and supplementary table S2, Supplementary Material online), the most parsimonious scenario is that the expansion of RP genes occurred independently in each lineage. In our subsequent analyses, we separately inferred the timing and nature of gene duplications for each RP family in each lineage and determined whether they have experienced gene conversion after gene duplication.

Duplication and Concerted Evolution of RP Genes in the Budding Yeasts

We manually identified all RP genes for the 23 representative budding yeasts. Fifty-nine RP families have duplicate copies in most post-WGD species (fig. 1 and supplementary table S2, Supplementary Material online). Fifty-five of them are ohnologs generated by the WGD (Conant and Wolfe 2006). The other four RP families, including *RPP1*, *RPP2*, *RPL9*, and *RPS22*, have duplicates in both post-WGD and non-WGD species, suggesting that they have been duplicated before the divergence of budding yeasts. In summary, most post-WGD budding yeasts have a significant increase in RP gene number,

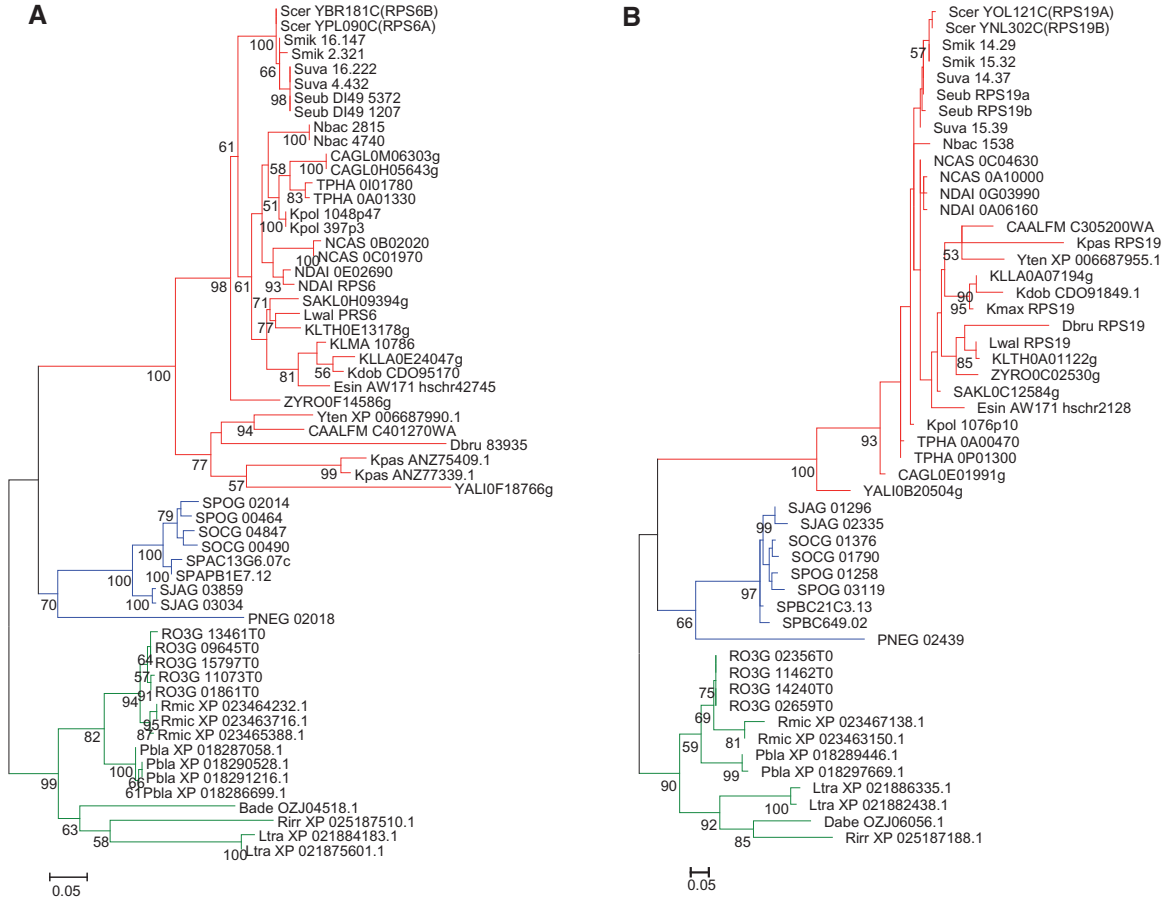

Fig. 2. Phylogenetic trees of representative RP gene families. Phylogenetic trees of the *RPS6* gene family (*A*) and the *RPS19* gene family (*B*) in 34 fungal species. The phylogenetic trees were inferred by ML method with 100 bootstrap replications. Only bootstrap values above 50 are shown next to each node. The branches of budding yeasts, fission yeasts, and Mucoromycota species are colored in red, blue, and green, respectively. The species names in taxa were provided in supplementary table S2, Supplementary Material online.

mainly due to WGD (fig. 1). Two post-WGD species, including a human opportunistic pathogen *Candida glabrata* (*Nakaseomyces glabrata*) and its closely related species *Nakaseomyces bacillisporus*, have only 85 and 97 RP genes, respectively, suggesting most RP ohnologs have been lost during their evolution.

Because the timings of these RP duplications in budding yeast have already been determined, it is possible to infer which RP paralogs have experienced gene conversion by comparing their gene trees with their duplication history. We can also use the tree topologies to infer when concerted evolution had terminated, which is the time when paralogous genes started to accumulate mutations independently. To simplify this process, we constructed phylogenetic trees for each duplicate RP family using five representative WGD species with different divergence times, including *S. cerevisiae*, *S. mikatae*, *S. eubayanus*, *Naumovozyma castellii* and *Tetrapisispora phaffii* (fig. 3 and supplementary file 2, Supplementary Material online). We found that at least 52 RP duplicate pairs in *S. cerevisiae* have experienced gene conversion, and 50 of them are ohnologs (fig. 3). Therefore, 88% (52 out of 59) of RP paralogous genes in *S. cerevisiae* have experienced gene conversion, which is more than that of previously identified (16 and 29) (Evangelisti and Conant 2010; Casola et al. 2012), suggesting that concerted evolution of RP genes in the budding yeasts is more prevalent than previously recognized.

Based on gene tree topologies, we inferred when concerted evolution of RP genes had terminated during evolution of budding yeasts. In 21 RP families, the RP duplicate genes in *S. cerevisiae* form a species-specific clade (fig. 3*A* and supplementary file 2, Supplementary Material online), suggesting that the concerted evolution is still ongoing or has recently terminated after its divergence from *S. mikatae*. In seven RP families, termination of concerted evolution occurred before the split between *S. cerevisiae* and *S. mikatae* (fig. 3*B*). Twenty-four RP families had ended their concerted evolution before the divergence of the *Saccharomyces sensu stricto* group, including *S. cerevisiae*, *S. mikatae*, *S. eubayanus* (fig. 3*C*). Only seven RP gene pairs do not show evidence of gene conversion (fig. 3*D*). A schematic summary of the concerted evolution of the 59 *S. cerevisiae* RP pairs is provided in figure 3*E*.

Duplication and Concerted Evolution of RP Genes in the Fission Yeasts

We identified all RP genes for five species in the subphylum of Taphrinomycotina, including the four fission yeasts and *Pneumocystis murina*. *Pneumocystis murina* belongs to the class of Pneumocystidomycetes, which is probably the most

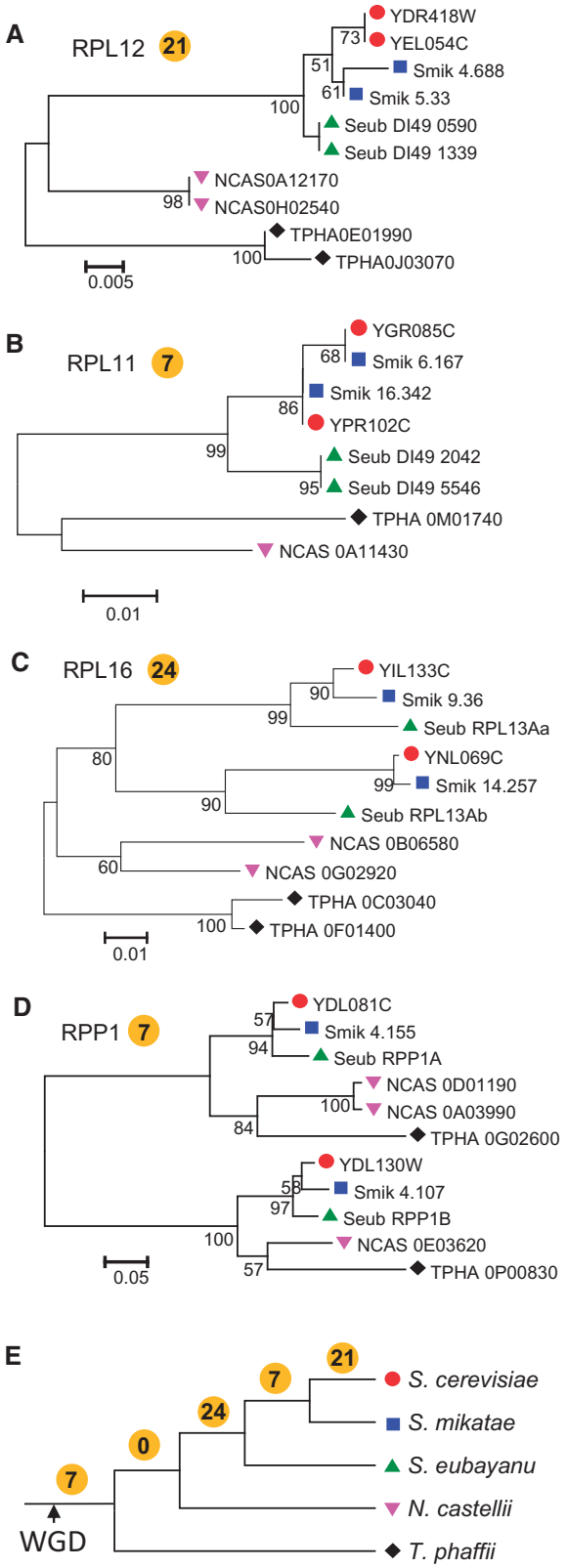

Fig. 3. Major types of tree topologies observed from 55 RP families with duplicates in budding yeasts. (*A*) Phylogenetic relationships of *RPL12* genes from five representative budding yeast species. In this case, the two copies of *RPL12* genes in *S. cerevisiae* form a species-specific clade. Twenty-one RP gene families demonstrate a similar tree topology, as indicated by number "21" in a yellow dot. (*B*) Phylogenetic relationships of *RPL11* genes. The two copies of *RPL11* genes in *S. cerevisiae* and *S. mikatae* form a well-supported clade, and the orthologous genes between *S. cerevisiae* and *S. mikatae* are more closely related to each other. Seven RP gene families have a similar tree topology. (*C*) Phylogenetic relationships of *RPL16* genes. Each of *RPL16* duplicate genes in *S. cerevisiae* is more closely related to their orthologous genes in *S. mikatae* and *S. eubayanus*. 24 RP gene families share a similar tree topology. (*D*) Phylogenetic relationships of RPP1 genes. Each of *RPP1* duplicate genes in *S. cerevisiae* is more closely related to their orthologous genes in the five WGD species. Seven RP gene families demonstrate a similar tree topology. (*E*) The distribution of RP families with different termination points of concerted evolution based on evolutionary relationships of RP duplicate genes in *S. cerevisiae*. The numbers on tree branches represent the numbers of RP families that have terminated concerted evolution in different evolutionary stages of *S. cerevisiae*.

closely related lineage to the fission yeasts, and it was used as an outgroup to infer the evolutionary history of RP genes in Taphrinomycotina. The number of RP families with duplicate copies ranges from 58 to 59 in the four fission yeasts (fig. 1 and supplementary table S2, Supplementary Material online). Most of them have two gene copies, but three gene copies are present in six RP families (fig. 1). In contrast, only one RP family (*RPL40*) has duplicates in *P. murina*. Thus, it is reasonable to assume that the massive expansion of RP genes in the fission yeasts occurred after their divergence from *P. murina*.

Similar to budding yeasts, most paralogous RP genes in fission yeasts are more similar to each other than to their orthologous genes (fig. 2 and supplementary file 3, Supplementary Material online). The tree topologies indicate that these RP genes were duplicated independently in each fission yeast after their divergence. However, we should consider the possibility of gene conversion. To infer when gene duplication occurred, we conducted gene collinearity (microsynteny) analysis for all duplicate RP genes in the four fission yeasts (see Materials and Methods section). If an RP gene was duplicated independently in each species after their divergence, the daughter genes are expected to be found in different genomic regions in these species. Under this scenario, only the parental RP genes in the four species share microsynteny. However, if a pair of RP duplicates share microsynteny by the four fission yeasts, they should be created by a single gene duplication event in their common ancestor.

We analyzed microsynteny for each pair of RP genes in fission yeasts by identifying their orthologous groups (supplementary table S4, Supplementary Material online) and gene orders. Herein, we defined a conserved region of microsynteny as a genomic block containing three or more conserved homologs within five genes downstream and upstream of an RP gene (fig. 4A). In *Sch. pombe*, 58 RP families have at least two gene copies. The duplicates of all 58 RP families in *Sch. pombe* share microsynteny with *Sch. cryophilus* and *Sch. octosporus* (supplementary table S5, Supplementary Material online), suggesting all duplicates were generated before the divergence of the three fission yeasts. We then inferred how many of them were duplicated even before the split of *Sch. japonicus* from the three fission yeasts. Thanks to

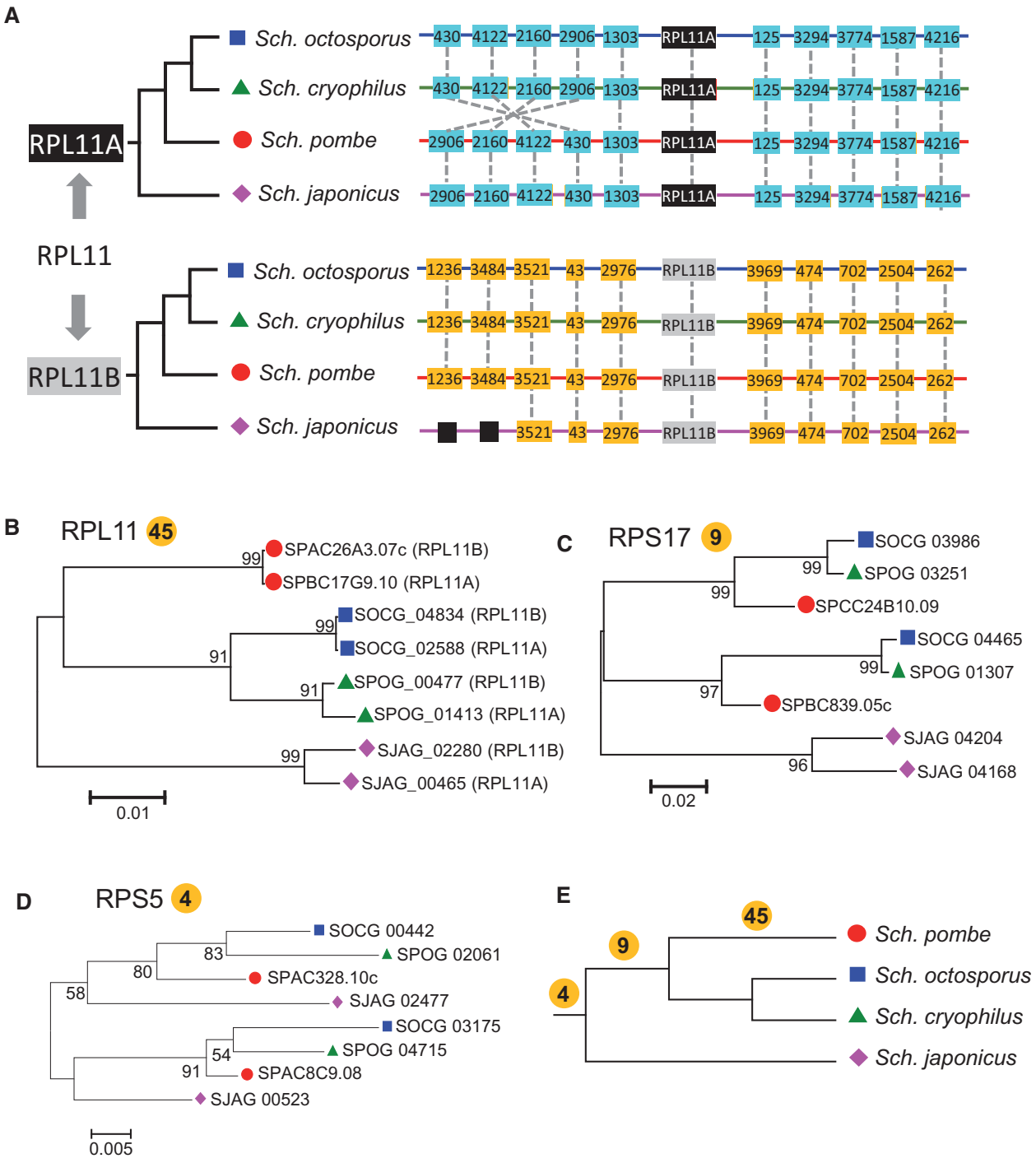

Fig. 4. The origin and evolution of RP genes in the fission yeasts. (*A*) A schematic illustration of microsynteny structures of *RPL11* genes in four fission yeasts. The microsynteny blocks of *RPL11A* are shared by all fission yeasts, so are the *RPL11B* genes, supporting that the duplication of *RPL11* occurred prior to the divergence of fission yeasts. The number in each box represents its orthologous group ID. (*B*) Phylogenetic relationships of *RPL11* genes in four fission yeasts. The *RPL11* duplicate genes in *Sch. pombe* are more closely related to each other than to their orthologous genes. Forty-five RP gene families demonstrate a similar tree topology. (*C*) Phylogenetic relationships of *RPS17* genes. Each of *RPL17* duplicate genes in *Sch. pombe* is more closely related to their orthologous genes in *Sch. octosporus* and *Sch. cryophilus*. Nine RP gene families demonstrate a similar tree topology. (*D*) Phylogenetic relationships of *RPS5* genes. Each copy of *RPS5* duplicates in *Sch. pombe* is more closely related to their orthologous genes. Four RP gene families demonstrate a similar tree topology. (*E*) The distribution of RP families with different termination points of concerted evolution in *Sch. pombe*. The numbers on each tree branch indicate the numbers of RP families that have terminated concerted evolution in different evolutionary stages of *Sch. pombe*.

the highly conserved gene order in fission yeasts (Rajeh et al. 2018), we detected shared microsynteny by both duplicates in 49 RP families among the four species. Therefore, at least 49 RP families have been duplicated before the divergence of the four fission yeasts (supplementary table S5, Supplementary Material online). For example, two *RPL11* genes are found in each fission yeast. Highly conserved regions of microsynteny surrounding *RPL11A* genes were found in all species, and so were the *RPL11B* genes (fig. 4A), supporting that *RPL11* was duplicated in their common ancestor and both copies have been maintained in each fission yeast after their divergence.

For nine RP families, we did not obtain conclusive evidence to determine whether they were duplicated before the split of *Sch. japonicus*. Four of them (*RPL10*, *RPL30*, *RPS12*, and *RPS25*) have only a single copy in *Sch. japonicus*. These genes could have been duplicated in their common ancestor, following by loss of one copy in *Sch. japonicus*. Alternatively, the duplication events have occurred after the split of *Sch. japonicus* from the other species. In the other five RP families (*RPL3*, *RPL17*, *RPL18*, *RPL21*, and *RPS19*), only one RP copy in *Sch. japonicus* shares microsynteny with the other three species. Similarly, the duplication events of these RP families could predate the divergence of fission yeasts, followed by genome rearrangements in *Sch. japonicus* that resulted in the loss of its gene collinearity. However, we cannot exclude the possibility that they were generated by independent duplication events in *Sch. japonicus*.

To determine which RP families have an incompatibility between gene phylogenetic tree and duplication history, we constructed a phylogenetic tree for each RP family with duplicates in the fission yeasts. In the case of *RPL11*, contradicting to the gene duplication history as inferred by microsynteny analysis (fig. 4A), the phylogenetic tree shows that *RPL11* paralogs form a species-specific clade in each fission yeast (fig. 4B). This contradiction suggests that gene conversion has occurred between *RPL11* paralogous genes in each fission yeast after their divergence. A total number of 45 RP families (77.6%) in fission yeasts have a similar tree topology to *RPL11* (fig. 4B and supplementary file 3, Supplementary Material online). In other families, such as *RPS17*, the two copies of RP genes from *Sch. pombe*, *Sch. octosporus*, and *Sch. cryophilus* form two clades and each clade consists of one RP gene from the three species. We observed nine RP families similar to *RPS17*, for which concerted evolution of these RP genes might have been terminated before the divergence of the three species (fig. 4C). However, we did not find evidence supporting gene conversion in only four RP families, including *RPL30*, *RPS5*, *RPS12*, and *RPS28* (fig. 4D and E).

Retroposition as a Major Mechanism for Massive Duplication of RP Genes in the Ancestral Fission Yeast

Because no WGD was detected during the evolution of *Sch. pombe* (Rhind et al. 2011), we then inferred other mechanisms that resulted in massive duplications of RP genes in fission yeasts, such as unequal crossing-over and retroposition. Unequal crossing-over typically generates segmental or tandem gene duplicates. If two genes were generated by segmental duplication, we expect to observe microsynteny between regions of paralogous RP genes within a species. However, we did not find any case of microsynteny in these RP families (supplementary table S5, Supplementary Material online). Furthermore, we did not detect tandemly arranged RP paralogous genes, suggesting that unequal crossing-over is not a major contributor for RP duplications in fission yeasts either.

Retroposition generates retroduplicates through random insertions of a retrotranscribed cDNA from parental source genes, resulting in intron-less retroduplicate genes (Kaessmann et al. 2009). We examined exon–intron structures for all RP paralogous genes in *Sch. pombe*. Among 21 singleton RP families in *Sch. pombe*, only 7 of them (33.3%) are intron-less (supplementary table S6, Supplementary Material online). In contrast, 33 of 58 duplicate RP families (56.9%) have at least one copy of intron-less gene, which is significantly higher than singleton RPs ($P = 0.006$, Fisher exact test). This ratio is also significantly higher than RP ohnologous pairs (27.3%) in *S. cerevisiae*. Thus, the enrichment of intron-less RP duplicate genes in fission yeasts suggests that they were likely generated by retroposition. For those RP duplicates with intron in both copies, we cannot exclude the possibility that they were created by retroposition followed by insertion of intron, because the locations and phases of introns between paralogous RP genes in *Sch. pombe* are usually different.

Duplication and Concerted Evolution of RP Genes in the Mucoromycota Species

Four Mucoromycota species examined demonstrate massive duplications of RP genes. Three of them belong to the order of Mucorales (pin molds) in subphylum of Mucoromycotina (311 RP genes in *R. delemar*, 182 in *R. microsporus*, and 217 in *P. blakesleeanus*). In a distantly related species within the same subphylum, *Bifiguratus adelaidae*, only 89 RP genes were found. Massive duplications of RP genes (137 RP genes) were also observed in *L. transversale*, a distantly related species belonging to another subphylum Mortierellomycotina. Seventy-eight RP genes are present in the earliest diverging species among Mucoromycota species *Rhizophagus irregularis* (fig. 1).

Based on RP gene copy numbers and the evolutionary relationships of these Mucoromycota species, it is most parsimonious to conclude that massive expansion of RP genes in the three pin mold species and *L. transversale* occurred independently. *Lobosporangium transversale* is a rare species that has only been reported by a few isolates in North American (Benny and Blackwell 2004). The genomic studies and physiological characterizations of *L. transversale* are scarce. Due to lack of genomic data from its closely related species, we cannot provide a systematic inference of the timings and nature of massive RP duplications in *L. transversale*. Thus, our subsequent analysis only focused on pin mold species.

A WGD event has been proposed in ancestral *R. delemar* (Ma et al. 2009). Another WGD was speculated to have occurred in *P. blakesleeanus* prior to its divergence from *R. microsporus* and *R. delemar* (Corrochano et al. 2016). Therefore, *R. delemar* might have experienced two rounds

of WGD, which correlates with the fact that it has the largest RP repertoire. Based on RP gene copy numbers, it is reasonable to conclude that the second WGD occurred after the divergence of *R. delemar* from *R. microsporus*.

We conducted microsynteny analysis to infer which RP gene pairs were generated by the two rounds of WGD in pin molds. The estimated divergence time between *Phycomyces* and *Rhizopus* is over 750 Ma (Mendoza et al. 2014). Most, if not all, microsynteny blocks generated by the first WGD might have been lost during the evolution of pin molds. Even though we have used a less strict definition of microsynteny (a minimum of 3 shared homologs in a block of ±10 neighboring genes surrounding RP), we only identified 3 and 10 pairs of microsynteny blocks between paralogous RP genes in *R. microsporus* and *P. blakesleeanus*, respectively. In contrast, we detected microsynteny for 63 pairs of *R. delemar* RP paralogous genes (supplementary table S7, Supplementary Material online), supporting the recent WGD event as a major contributor to the expansion of RP genes in *R. delemar*.

We attempted to identify microsynteny for orthologous RP genes to infer the evolutionary history of each RP family in pin molds (supplementary table S8, Supplementary Material online). Due to the large divergence times between these species, shared microsynteny regions are absent among most RP orthologous genes. The most well-supported example is probably the *RPL3* family (fig. 5A). Based on the shared gene orders between paralogous and orthologous *RPL3* genes, it is reasonable to infer that *RPL3* had been duplicated before the divergence of the three pin mold species, probably due to the first WGD. In *R. delemar*, the two *RPL3* copies have been further duplicated by the recent WGD, generating four copies. However, their gene tree (fig. 5B) demonstrates that *RPL3* paralogous genes in each species form a species-specific clade, suggesting they have experienced gene conversion in each species. A total number of 57 RP families have a similar tree topology (supplementary file 4, Supplementary Material online). Although there is no conclusive microsynteny evidence to support that these RP families have the same evolutionary history as *RPL3*, it is tempting to assume that it would be the most likely scenario. In some RP families, such as *RPL38* (fig. 5C), the genes from *R. delemar* and *R. microsporus* form two clades, and each clade has members from both species. A total number of 17 RP families have a similar tree topology to *RPL38*. Assuming these paralogous RP genes were generated by the ancient WGD, their concerted evolution had terminated prior to the divergence of the two *Rhizopus* species. The last type of tree topology, such as *RPS20* (fig. 5D), whose members form two clades, and each clade includes genes from three pin mold species. Such tree topology does not support the occurrence of gene conversion. Only two RP families demonstrate such type of tree topology (fig. 5E). In summary, our results imply that most RP paralogous genes in pin molds might have also experienced gene conversion, similar to that of budding yeasts and fission yeasts.

cDNA as the Probable Donor for Gene Conversion between RP Paralogous Genes

During gene conversion, the genomic sequence of the "acceptor" locus is replaced by a "donor" sequence through recombination (Chen et al. 2007). The donor can be genomic DNA or cDNA derived from an mRNA intermediate (Derr and Strathern 1993; Storici et al. 2007). If genomic DNA is the donor, the sequences of both intron and exon can be homogenized. In contrast, if cDNA is the donor, only the exon sequences in the acceptor are replaced. Considering that synonymous mutations are largely free from natural selection, it is possible to determine the donor of gene conversion by comparing the substitution rates between introns and synonymous sites. If the synonymous substitution rates (d_S) are significantly lower than intron substitution rates (μ_{intron}), supporting cDNA as a donor. We calculated d_S and μ_{intron} for all RP duplicate genes for one representative species from each fungal lineage: *S. cerevisiae*, *Sch. pombe*, and *R. microsporus* (supplementary table S9, Supplementary Material online). Overall, the d_S values of all paralogous RP genes are significantly lower than μ_{intron} for representative species (fig. 6A–C, Student's *t*-test, $P < 0.01$). Considering that different genomic regions might have different substitution rates, we then compared the d_S and μ_{intron} between each pair of RP paralogous genes (fig. 6D and E). Consistently, most of RP duplicate gene pairs have lower d_S values than μ_{intron}. In a small number of cases, high d_S values were observed, probably because the concerted evolution had terminated far in the past, resulting in accumulation of many synonymous mutations. In contrast, such patterns are not present among RP paralogs without gene conversion (supplementary fig. S2, Supplementary Material online). These results suggest that, in most cases, only the coding sequences (CDS) have been homogenized by gene conversion, supporting cDNA as the main gene conversion donor.

The Retention of RP Gene Duplicates Was Associated with the Evolution to Fermentative Ability in Fungi

Most eukaryotic species fully oxidize glucose, their primary carbon and energy source, through mitochondrial oxidative phosphorylation in the presence of oxygen for maximum energy production. In contrast, post-WGD budding yeasts and fission yeasts predominantly ferment sugar to ethanol in the presence of excess sugars, even under aerobic conditions, which was called aerobic fermentation (Alexander and Jeffries 1990; Lin and Li 2011a). Aerobic fermentation has independently evolved in the budding yeasts and fission yeasts (de Jong-Gubbels et al. 1996). Among pin molds, the domesticated form of *R. microsporus* has been widely used as a starter culture for the production of tempeh from fermented soybean (Hachmeister and Fung 1993). Its close relative, *R. delemar*, was also well known as efficient ethanol and fumaric acid producer by fermentation (Kito et al. 2009; Straathof and van Gulik 2012). *Phycomyces blakesleeanus* was known to be capable of fermenting sugar into β-carotene at an industrial scale, which is derived from the end product of glycolysis (Kaessmann et al. 2009).

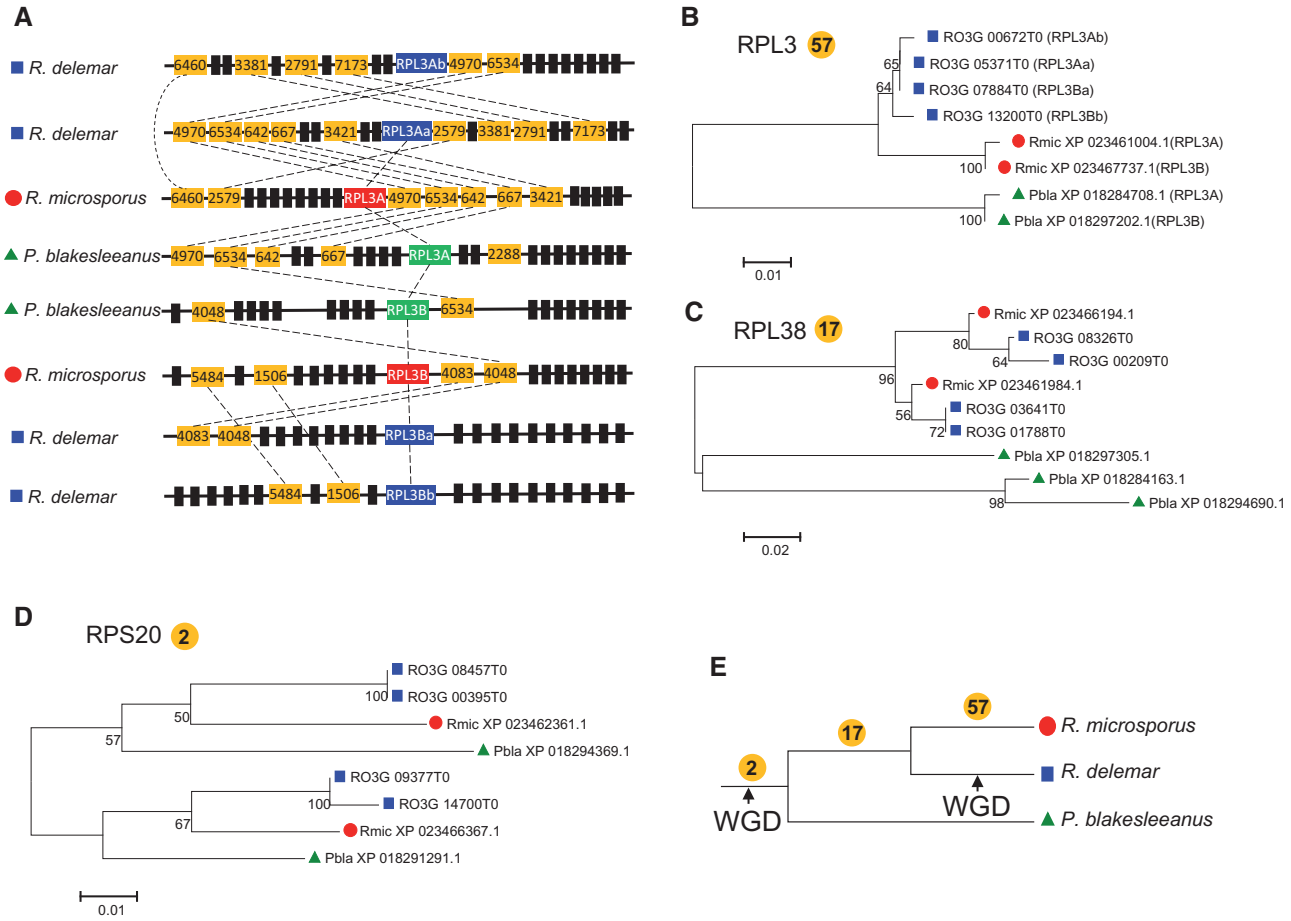

FIG. 5. The origin and evolution of RP genes in pin molds. (*A*) A schematic illustration of microsynteny structures of *RPL3* genes in three pin mold species. The shared microsynteny structure suggested that the first duplication event of *RPL3* genes have occurred prior to the divergence of the pin molds. The two *RPL3* copies have experienced a second round of duplication by WGD in *R. delemar*. (*B*) Phylogenetic relationships of *RPL3* genes in three pin mold species. The *RPL3* paralogous genes in *R. microsporus* are more closely related to each other than to their orthologous genes. Fifty-seven RP gene families demonstrate a similar tree topology. (*C*) Phylogenetic relationships of *RPL38* genes in Mucorales. Each of *RPL38* duplicate genes in *R. microsporus* is more closely related to their orthologous genes in *R. delemar*. Seventeen RP gene families demonstrate a similar tree topology. (*D*) Phylogenetic relationships of *RPS20* genes in three pin molds. Each *RPS20* duplicate gene in *R. microsporus* is more closely related to their orthologous genes. Two RP gene families demonstrate a similar tree topology. (*E*) The distribution of RP families with different termination points of concerted evolution during the evolution of *R. microsporus*. The numbers on each tree branch represent the estimated numbers of RP families that have terminated concerted evolution in different evolutionary stages of *R. microsporus*.

It is tempting to propose that the massive duplication and retention of RP genes have contributed to the evolution of fermentative ability in these species. Increased gene dosage could lead to a quantitative increase in transcript and protein production. To determine the impact of gene duplication on the production of RP transcripts, we calculated the total transcript abundance of all RP genes using our transcriptomic data generated by Cap Analysis of Gene Expression (CAGE) (McMillan et al. 2019). The CAGE technique captures and sequences the first 75 bp of transcripts, which quantifies the transcript abundance based on numbers of mapped reads (Murata et al. 2014). Nine budding yeasts and two fission yeasts examined have CAGE data that were obtained from cells grown in YPD rich medium (supplementary table S10, Supplementary Material online). As shown in figure 7A, the RP copy numbers are positively correlated with total transcript abundance values of RP genes (supplementary table S10, Supplementary Material online, Pearson correlation $r = 0.72$), supporting that the increased RP gene dosage might have increased ribosome biogenesis by generating more RP transcripts.

We then infer whether the increased RP gene dosage is associated with better fermentative ability. A previous study has measured various physiological characteristics for over 40 yeast species (Hagman et al. 2013), including 19 species examined in this study. We observed a positive correlation between RP gene copy numbers and ethanol production efficiency ($r = 0.80$), and glucose consumption rates ($r = 0.76$) (fig. 7B and C). We also observed a significant positive correlation between total transcript abundance of RP genes and ethanol production efficiency ($r = 0.87$), as well as glucose consumption rates ($r = 0.88$) (supplementary fig. S3, Supplementary Material online). These results suggest that the increased RP expression by gene duplication might have enhanced these organisms' ability to rapidly consuming glucose through the fermentation pathway.

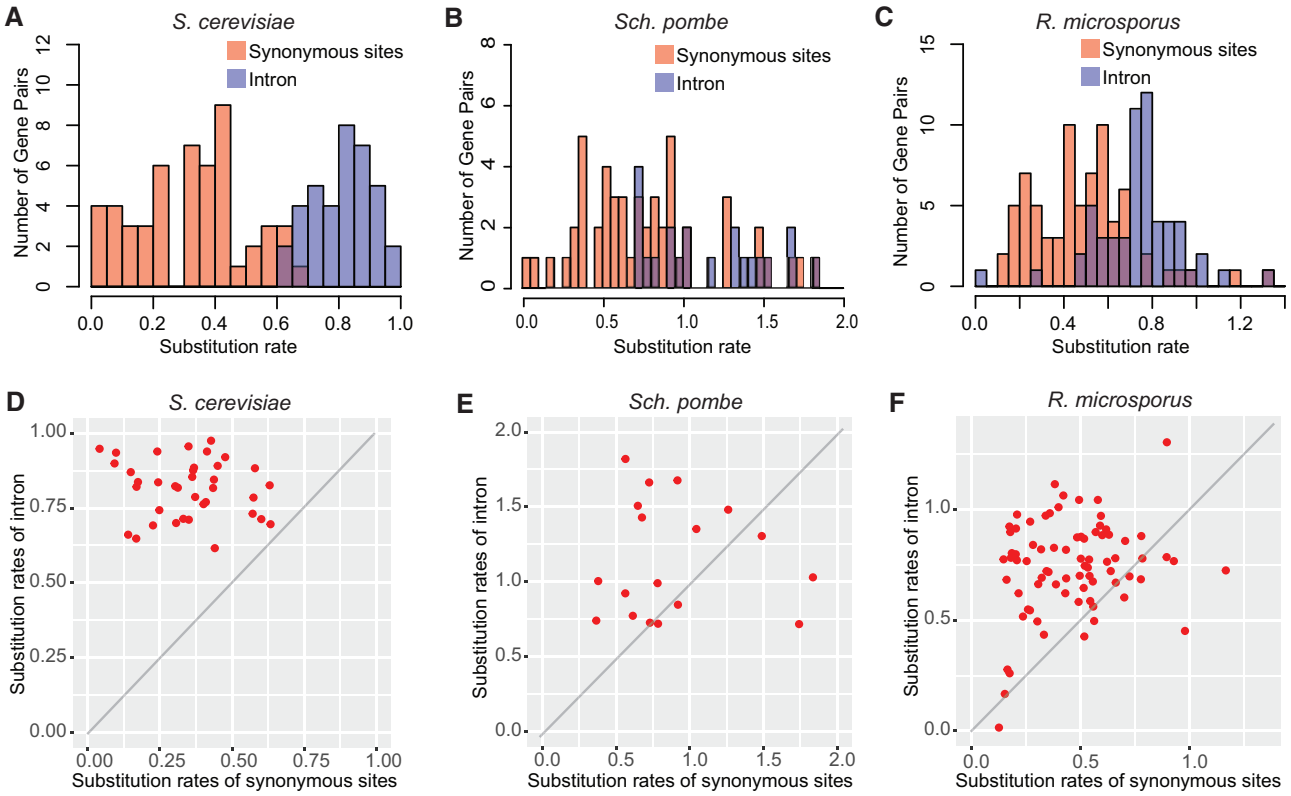

FIG. 6. Distinct substitution rates in substitution sites and introns between RP paralogous genes. The distributions of substitution rates in introns and synonymous sites between RP paralogous genes in *S. cerevisiae* (*A*), *Sch. pombe* (*B*), and *R. microsporus* (*C*). Scatter plots of substitution rates in introns against synonymous sites between RP paralogous genes in *S. cerevisiae* (*D*), *Sch. pombe* (*E*), and *R. microsporus* (F).

Discussion

The Preferential Retention of RP Duplicate Genes Was Selection-Driven

Our survey of 295 fungal genomes revealed that massive duplications of RP genes are not prevalent. However, significant increases in RP gene copy numbers had independently occurred in budding yeasts, fission yeasts, and pin molds. WGD events have played an important role in the expansion RP repertoire in budding yeasts and pin molds. In budding yeasts, only ~10% of WGD ohnologs have survived, whereas 70.5% of RP duplicates generated by WGD have been maintained. As indicated by previous studies, the survival rate of RP ohnologs is significantly higher than the other WGD ohnologs (Papp et al. 2003).

Our results suggest that RP genes in fission yeasts were likely individually duplicated by SSDs, such as retroposition. In general, the retention rate of duplicate genes generated by SSDs is much lower than ohnologs (half-life of 4 My vs. 33 My) (Hakes et al. 2007). It is even lower for genes encoding macromolecular complexes due to evolutionary constraints imposed by gene dosage balance (Li et al. 1996; Conant and Wolfe 2008). Similar to the fission yeasts, 99.8% of RP duplicates in mammals were found to be generated by retroposition (Dharia et al. 2014). However, almost all RP retroduplicates in mammals became pseudogenes (Dharia et al. 2014). Therefore, the high retention rates of functional RP duplicates generated by SSDs in each fission yeast are indeed striking. The increases in RP gene dosage should have provided some selective advantages to these species so that the retention of RP duplicates had been favored by natural selection.

There is another line of evidence supporting that the retention of RP duplicate genes in fission yeasts is better explained by natural selection. The fission yeasts have been known to maintain a single gene copy in most families (Rhind et al. 2011; Rajeh et al. 2018). Based on our orthologous group data (supplementary table S4, Supplementary Material online), 86% (4,069/4,734) of fission yeast ortholog groups maintained only a single gene copy in each species (or 1:1:1:1 ortholog). Of the gene families with gene duplication or loss in at least one fission yeast, RP genes account for 9.3% (62/665) of them, which is significantly overrepresented in this group ($P < 10^{-5}$, Fisher exact test).

How Duplication and Retention of RP Genes Contributed to the Evolution of Fermentative Ability in Fungi

The association between retention of RP duplicate genes and fermentative ability in the three distantly related fungal lineages suggests that increased RP gene dosage might have contributed to the independent evolution of strong fermentative ability. The fermentative yeasts were believed to have gained a growth advantage through rapid glucose fermentation in the presence of excess sugars (Piskur et al. 2006). It was found that

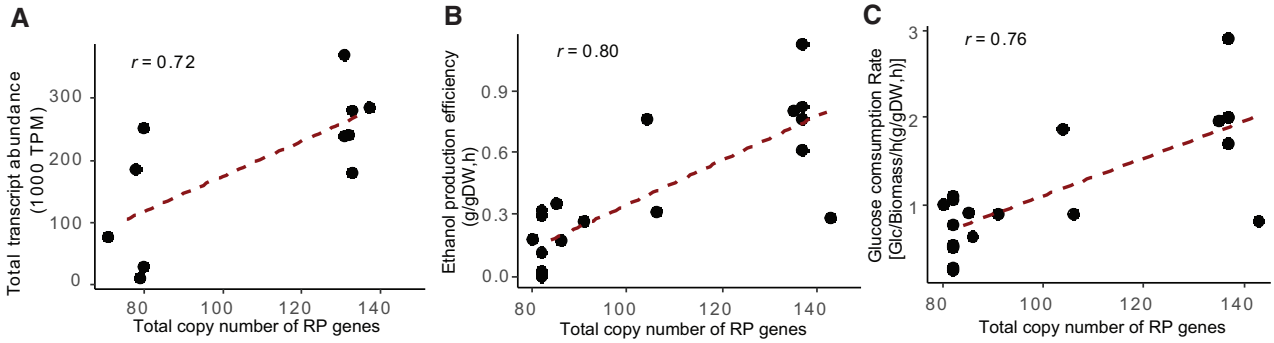

FIG. 7. RP gene copy numbers are positively correlated with RP transcript abundance, ethanol production efficiency, and glucose consumption rates. (*A*) A scatter plot between the RP copy numbers and total RP transcript abundance in 11 yeast species. (*B*) A scatter plot between the RP copy numbers and ethanol production efficiency in 19 yeast species. (*C*) A scatter plot between the RP copy numbers and glucose consumption rate in 19 yeast species.

S. cerevisiae often outgrew its nonfermentative competitors in coculture experiments (Pérez-Nevado et al. 2006; Williams et al. 2015). Fermentation is much less efficient in generating energy and biomass, but fermentative organisms overcome this disadvantage by rapidly consuming sugars through fermentation pathway under sugar-rich environments, providing selective advantages (Pfeiffer et al. 2001; Pfeiffer and Morley 2014). The rapid glucose consumption had been facilitated by the increased dosages of genes involved in glycolysis flux (Conant and Wolfe 2007) and transporting glucose across cellular membranes (Lin and Li 2011b). In this study, we observed independent increases in RP gene dosages in three distantly related fungal lineages. These fungal species share a common physiological feature of good fermentative ability. Two post-WGD budding yeasts (*N. glabrata* and *N. bacillisporus*) have lost most RP ohnologs. Consistently, they have only 50% of glucose consumption rate and ethanol production efficiency compared with *S. cerevisiae* (Hagman et al. 2013).

Ribosome biosynthesis comes with the opportunity cost of higher expression of other cellular processes needed for cell viability and function. Maintaining one RP gene per family may be advantageous for most species to allow for greater Pol II transcription potential for other genes. During evolution of fungi, some ancestral fungal organisms had gained the ability to rapidly consume glucose through fermentative pathways. The fermentation products could serve as carbon substrates after depletion of sugars or inhibit the growth of competing organisms, providing selective advantages under sugar-rich environments (Piskur et al. 2006). Therefore, it is reasonable to propose that the increased RP gene dosage had facilitated the evolution of fermentative ability. Our data suggested the contribution of increased RP gene dosage was likely achieved by increasing transcript abundance, which could lead to increased biogenesis of ribosomes, glycolysis enzymes, glucose transporters, and other building blocks for cell growth and proliferation. This strategy allowed these organisms to maintain a high growth rate while conducting the low-efficient fermentation by rapidly consuming sugars.

Gene Duplication Allows Further Increase in Transcript Abundance from Highly Expressed RP Genes

One may argue that the increase in ribosome biosynthesis can also be achieved by elevated transcription activities of RP genes. It is probably true because we also observed elevated expression levels of RP genes in two non-WGD yeasts that demonstrate an intermediate level of ethanol fermentation ability: *Lachancea thermotolerans* and *Lachancea waltii* (Hagman et al. 2013). RP genes are among the most abundantly transcribed genes in eukaryotic cells, accounting for 50% of RNA Pol II transcription (Warner 1999). RP genes have the highest density of bound Pol II. In *S. cerevisiae*, 100 RP genes have on average >60% of the maximum Pol II occupancy, whereas a majority of the genome only has <5% of the maximum Pol II density (Venters and Pugh 2008). Therefore, there is limited room for further increasing the transcription of RP genes. Duplication of RP genes provides additional substrates on which Pol II can transcribe into RP mRNAs, so that transcription is no longer a rate-limiting step in the process of ribosome biogenesis in these organisms.

Gene Conversion Facilitated Retentions of RP Duplicate Genes

The sequences of all RP families are highly conserved during the evolution of eukaryotes due to their vital roles in many cellular functions (Korobeinikova et al. 2012). Because misfolding and misinteractions of highly abundant proteins can be more costly, proteins like RPs should have been under more functional constraints and evolved even slower than other important proteins (Zhang and Yang 2015). It has shown that there is a strong evolutionary constraint posed on the duplicability of genes encoding core components of protein complexes (Li et al. 2006). Accumulation of new mutations in duplicate genes could impact the stability of protein complexes, posing selective disadvantages.

Our study demonstrates that gene conversion following the duplication of RP genes appears to be a universal path in fungal species. Through gene conversion, the sequences of paralogous genes are homogenized, easing the new

mutations accumulated in one of the paralogous genes. It has been shown that highly expressed genes are more likely to experience mRNA-mediated gene conversion (Weng et al. 2000; Schildkraut et al. 2006). Thus, as a group of most actively transcribed genes (Warner 1999), the repeated occurrence of gene conversion between RP paralogous genes is expected. It was also found that the promoters or flanking genomic sequences between paralogous RP genes are much more divergent than their CDS (Evangelisti and Conant 2010), further supporting that gene conversion in RPs was mediated by cDNA. Gene conversion could provide an additional layer of protection on top of purifying selection to remove newly accumulated mutations (Evangelisti and Conant 2010; Scienski et al. 2015). Increased RP gene copies could be advantageous to these fungal species living in sugar-rich environments through fermentative growth. The repeated occurrence of gene conversion between RP paralogous genes had contributed to the maintenance of functional RP duplicates in these organisms by removing newly accumulated mutations.

Materials and Methods

Data Sources, Identification, and Manual Curation of RP Repertoire

We obtained a complete list of RP genes in *S. cerevisiae* and *Sch. pombe* from the RPG (Nakao et al. 2004). We downloaded RefSeq protein sequence data of 285 fungal genomes from NCBI (supplementary table S1, Supplementary Material online). In the first round of homologous sequences, we used RP sequences from *S. cerevisiae* and *Sch. pombe* as queries to run BLASTP search against the 285 proteomic data (Camacho et al. 2009). For BLASTP search, we used an *e* value cutoff of 1e-10, and only hits with a minimum alignment length of 50% of query sequences were considered as homologous RP genes.

In the second round of homologous searches, we obtained the protein and genome sequences of 34 fungal species from NCBI WGS, JGI, and YGOB (Byrne and Wolfe 2005; Maguire et al. 2013) (supplementary table S2, Supplementary Material online). We first used 79 RP sequences from *S. cerevisiae* and *Sch. pombe* as queries to search for homologous sequences from the ten newly added fungal species using BLASTP. To identify RP sequences not predicted by existing genome annotations, we conducted TBlastN searches against all 34 genomic sequences. Manual inspections were performed to compare the BLASTP and TBlastN results to identify discrepant hits. For hits obtained by TBlastN by not BLASTP, we predicted the CDS based on six-frame translations of genomic sequences. The exon–intron boundaries were determined based on the TBlastN alignments and the presence of GT/AG splice sites in flanking intron sequences. We also revised the predicted protein sequences if there is a discrepancy in aligned regions between BLASTP and TBlastN results. The same gene prediction method was used to revise misannotated ORF.

Construction of Phylogenetic Tree for RP Genes

We inferred the phylogeny for each RP family using RP sequences collected from the 34 representative species. Sequences were aligned through MUSCLE (Edgar 2004). The molecular phylogenetic trees were inferred by the ML method using RAxML with 100 bootstrap pseudo-replicates (Stamatakis 2006). The best-fit substitution model was inferred by using ProtTest (Abascal et al. 2005). As the LG model was the best substitution model identified for the majority of RP families, it was used in our phylogenetic reconstruction. A discrete Gamma distribution [+G] and invariable sites [+I] was used to model evolutionary rate differences among sites. We also constructed lineage-specific phylogenetic trees for duplicate RP families using representative species from each fungal lineage using neighbor-joining method using MEGA 7 (Kumar et al. 2016). For those RP families with almost identical amino acid sequences, we used nucleotide sequences of their CDS for construction of gene trees to obtain better resolved tree topology (supplementary files 2–4, Supplementary Material online).

Homology Microsynteny Analysis

We conducted microsynteny analysis for each RP gene family in four fission yeasts and three pin mold species, including *R. delemar*, *R. microsporus*, and *P. blakesleeanus*. Gene order information was retrieved from genome annotations of each species obtained from NCBI. The orthologous gene groups in fission yeasts and pin mold species were respectively identified using the OrthoDB (Kriventseva et al. 2015). For fission yeasts, we obtained a list of ten genes surrounding each RP gene (five upstream and five downstream of RP gene). For the pin mold species, we extended our microsynteny analysis to a block of ten genes upstream and ten downstream of RP gene due to their divergent genome structures.

Estimation of Substitution Rates in Intron and Synonymous Sites

We calculated the substitution rates for every pair of duplicate RP genes for three species representing the three fungal lineages with massive RP duplications, including *S. cerevisiae*, *Sch. pombe*, and *R. microsporus*. The CDS and intron sequences were retrieved from NCBI and were aligned using MUSCLE. Synonymous substitution rate was calculated using Li–Wu–Luo method with Kimura 2-parameter model (Li et al. 1985) in MEGA 7 (Kumar et al. 2016). Nucleotide substitution rates in intron sequences were calculated using the Kimura 2-parameter model in MEGA 7.

Analysis of RP Gene Transcriptomic and Physiological Data

The transcriptomic data of nine budding yeasts and two fission yeasts examined were obtained from McMillan et al. (2019) based on CAGE. The expression abundance of an RP gene was defined as the sum of transcripts initiated from all core promoters within 500 base pairs upstream of its annotated start codon, which was normalized as TPM (tags per million mapped reads, and each tag represent one sequenced transcript). The total transcript abundance of RP genes in a species was calculated as the sum of TPM of all RP genes identified in this species. For newly predicted RP genes that were not annotated in CAGE data sets, we performed

TBlastN searches to determine their genomic locations of CDS and obtained the expression abundance data using the same criteria. The ethanol production efficiency and glucose consumption rate of 19 budding and fission yeast species were obtained from Hagman et al. (2013). The ethanol production efficiency was measured as grams of ethanol produced per gram of biomass per gram of glucose consumed. The glucose consumption rate was measured as grams of glucose consumed per gram of biomass per hour. If multiple biological replicates were measured for a single species, their average values were used for our analysis.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This study was supported by the start-up fund and President's Research Fund from Saint Louis University to ZL. We would like to thank Dr Zhenglong Gu and Dr Dapeng Zhang for valuable discussions and suggestions for this work.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104–2105.

Alexander MA, Jeffries TW. 1990. Respiratory efficiency and metabolite partitioning as regulatory phenomena in yeasts. *Enzyme Microb Technol*. 12(1):2–19.

Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E. 1980. Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. *Proc Natl Acad Sci U S A*. 77(12):7323–7327.

Barakat A, Szick-Miranda K, Chang F, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J. 2001. The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol*. 127(2):398–415.

Benny GL, Blackwell M. 2004. Lobosporangium, a new name for Echinosporangium Malloch, and Gamsiella, a new genus for *Mortierella multidivaricata*. *Mycologia* 96(1):143–149.

Birchler JA, Veitia RA. 2012. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci*. 109(37):14746–14753.

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453–1462.

Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. 2008. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell*. 19(1):352–367.

Brown DD, Wensink PC, Jordan E. 1972. A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J Mol Biol*. 63(1):57–73.

Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res*. 15(10):1456–1461.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinform*. 10(1):421.

Casola C, Conant GC, Hahn MW. 2012. Very low rate of gene conversion in the yeast genome. *Mol Biol Evol*. 29(12):3817–3826.

Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet*. 8(10):762–775.

Conant GC, Wolfe KH. 2006. Functional partitioning of yeast co-expression networks after genome duplication. *PLoS Biol*. 4(4):e109.

Conant GC, Wolfe KH. 2007. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol Syst Biol*. 3(1):129.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet*. 9(12):938–950.

Corrochano LM, Kuo A, Marcet-Houben M, Polaino S, Salamov A, Villalobos-Escobedo JM, Grimwood J, Alvarez MI, Avalos J, Bauer D, et al. 2016. Expansion of signal transduction pathways in fungi by extensive genome duplication. *Curr Biol*. 26(12):1577–1584.

de Jong-Gubbels P, van Dijken JP, Pronk JT. 1996. Metabolic fluxes in chemostat cultures of *Schizosaccharomyces pombe* grown on mixtures of glucose and ethanol. *Microbiology* 142(6):1399–1407.

Derr LK, Strathern JN. 1993. A role for reverse transcripts in gene conversion. *Nature* 361(6408):170.

Dharia AP, Obla A, Gajdosik MD, Simon A, Nelson CE. 2014. Tempo and mode of gene duplication in mammalian ribosomal protein evolution. *PLoS One* 9(11):e111721.

Dudov KP, Perry RP. 1984. The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. *Cell* 37(2):457–468.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.

Evangelisti AM, Conant GC. 2010. Nonrandom survival of gene conversions among yeast ribosomal proteins duplicated through genome doubling. *Genome Biol Evol*. 2(0):826–834.

Hachmeister KA, Fung DY. 1993. Tempeh: a mold-modified indigenous fermented food made from soybeans and/or cereal grains. *Crit Rev Microbiol*. 19(3):137–188.

Hagman A, Sall T, Compagno C, Piskur J. 2013. Yeast "make-accumulate-consume" life strategy evolved as a multi-step process that predates the whole genome duplication. *PLoS One* 8(7):e68734.

Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res*. 15(8):1153–1160.

Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL. 2007. All duplicates are not equal: the difference between small-scale and genome duplication. *Genome Biol*. 8(10):R209.

Jorgensen P, Nishikawa JL, Breitkreutz B-J, Tyers M. 2002. Systematic identification of pathways that couple cell growth and division in yeast. *Science* 297(5580):395–400.

Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet*. 10(1):19–31.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428(6983):617–624.

Kenmochi N, Kawaguchi T, Rozen S, Davis E, Goodman N, Hudson TJ, Tanaka T, Page DC. 1998. A map of 75 human ribosomal protein genes. *Genome Res*. 8(5):509–523.

Kito H, Abe A, Sujaya IN, Oda Y, Asano K, Sone T. 2009. Molecular characterization of the relationships among *Amylomyces rouxii*, *Rhizopus oryzae*, and *Rhizopus delemar*. *Biosci Biotechnol Biochem*. 73(4):861–864.

Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. *J Theor Biol*. 239(2):141–151.

Korobeinikova AV, Garber MB, Gongadze GM. 2012. Ribosomal proteins: structure, function, and evolution. *Biochemistry (Moscow)* 77(6):562–574.

Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM. 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res*. 43(D1):D250–256.

Kumar S, Stecher G, Tamura K. 2016. MEGA 7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33(7):1870–1874.

Kuzumaki T, Tanaka T, Ishikawa K, Ogata K. 1987. Rat ribosomal protein L35a multigene family: molecular structure and characterization of three L35a-related pseudogenes. *Biochim Biophys Acta* 909(2):99–106.
Li B, Vilardell J, Warner JR. 1996. An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness. *Proc Natl Acad Sci*. 93(4):1596–1600.
Li L, Huang Y, Xia X, Sun Z. 2006. Preferential duplication in the sparse part of yeast protein interaction network. *Mol Biol Evol*. 23(12):2467–2473.
Li WH, Wu CI, Luo CC. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol*. 2(2):150–174.
Lin Z, Kong H, Nei M, Ma H. 2006. Origins and evolution of the *recA*/*RAD51* gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc Natl Acad Sci U S A*. 103(27):10328–10333.
Lin Z, Li WH. 2011a. The evolution of aerobic fermentation in *Schizosaccharomyces pombe* was associated with regulatory reprogramming but not nucleosome reorganization. *Mol Biol Evol*. 28(4):1407–1413.
Lin Z, Li WH. 2011b. Expansion of hexose transporter genes was associated with the evolution of aerobic fermentation in yeasts. *Mol Biol Evol*. 28(1):131–142.
Lin Z, Nei M, Ma H. 2007. The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution. *Nucleic Acids Res*. 35(22):7591–7603.
Ma L-J, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, Elias M, Idnurm A, Lang BF, Sone T, et al. 2009. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet*. 5(7):e1000549.
Maguire SL, ÓhÉigeartaigh SS, Byrne KP, Schröder MS, O'Gaora P, Wolfe KH, Butler G. 2013. Comparative genome analysis and gene finding in Candida species using CGOB. *Mol Biol Evol*. 30(6):1281–1291.
McMillan J, Lu Z, Rodriguez JS, Ahn TH, Lin Z. 2019. YeasTSS: an integrative web database of yeast transcription start sites. *Database (Oxford)* 2019: baz048.
Mendoza L, Vilela R, Voelz K, Ibrahim AS, Voigt K, Lee SC. 2014. Human fungal pathogens of Mucorales and Entomophthorales. *Cold Spring Harb Perspect Med*. 5:a019562.
Montagne J, Stewart MJ, Stocker H, Hafen E, Kozma SC, Thomas G. 1999. Drosophila S6 kinase: a regulator of cell size. *Science* 285(5436):2126–2129.
Murata M, Nishiyori-Sueki H, Kojima-Ishiyama M, Carninci P, Hayashizaki Y, Itoh M. 2014. Detecting expressed genes using CAGE. *Methods Mol Biol*. 1164:67–85.
Nakao A, Yoshihama M, Kenmochi N. 2004. RPG: the ribosomal protein gene database. *Nucleic Acids Res*. 32(Database issue):D168–170.
Ohno S. 1970. Evolution by gene duplication. Berlin (New York): Springer-Verlag.
Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. *Plant Physiol*. 171(4):2294–2316.
Papp B, Pal C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424(6945):194–197.
Pérez-Nevado F, Albergaria H, Hogg T, Girio F. 2006. Cellular death of two non-*Saccharomyces* wine-related yeasts during mixed fermentations with *Saccharomyces cerevisiae*. *Int J Food Microbiol*. 108(3):336–345.
Pfeiffer T, Morley A. 2014. An evolutionary perspective on the Crabtree effect. *Front Mol Biosci*. 1:17.
Pfeiffer T, Schuster S, Bonhoeffer S. 2001. Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292(5516):504–507.
Piskur J, Rozpedowska E, Polakova S, Merico A, Compagno C. 2006. How did *Saccharomyces* evolve to become a good brewer? *Trends Genet*. 22(4):183–186.
Rajeh A, Lv J, Lin Z. 2018. Heterogeneous rates of genome rearrangement contributed to the disparity of species richness in *Ascomycota*. *BMC Genomics* 19(1):282.
Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF, Heiman DI, et al. 2011. Comparative functional genomics of the fission yeasts. *Science* 332(6032):930–936.
Sanchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major chemosensory gene families in insects. *Heredity (Edinburgh)* 103:208–216.
Schildkraut E, Miller CA, Nickoloff JA. 2006. Transcription of a donor enhances its use during double-strand break-induced gene conversion in human cells. *Mol Cell Biol*. 26(8):3098–3105.
Schlotterer C, Tautz D. 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr Biol*. 4:777–783.
Scienski K, Fay JC, Conant GC. 2015. Patterns of gene conversion in duplicated yeast histones suggest strong selection on a coadapted macromolecular complex. *Genome Biol Evol*. 7(12):3249–3258.
Sidow A. 1996. Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev*. 6(6):715–722.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. 2007. RNA-templated DNA repair. *Nature* 447(7142):338.
Straathof AJ, van Gulik WM. 2012. Production of fumaric acid by fermentation. *Subcell Biochem*. 64:225–240.
Venters BJ, Pugh BF. 2008. A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. *Genome Res*. 19(3):360–371.
Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290(5499):2114–2117.
Warner JR. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci*. 24(11):437–440.
Weng Y-S, Xing D, Clikeman JA, Nickoloff JA. 2000. Transcriptional effects on double-strand break-induced gene conversion tracts. *Mutat Res/DNA Repair* 461(2):119–132.
Williams KM, Liu P, Fay JC. 2015. Evolution of ecological dominance of yeast species in high-sugar environments. *Evolution* 69(8):2079–2093.
Wimberly BT, Brodersen DE, Clemons WM, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature* 407(6802):327–339.
Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387(6634):708–712.
Wool IG. 1979. The structure and function of eukaryotic ribosomes. *Annu Rev Biochem*. 48(1):719–754.
Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun*. 5(1):4956.
Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol*. 18:292–298.
Zhang J. 2013. Gene duplication. In: Losos J, editor. The Princeton guide to evolution. Princeton (NJ): Princeton University Press. p. 397–405.
Zhang J, Yang JR. 2015. Determinants of the rate of protein sequence evolution. *Nat Rev Genet*. 16(7):409–420.