Nonrandom Survival of Gene Conversions among Yeast Ribosomal Proteins Duplicated through Genome Doubling

Annette M. Evangelisti[1], and Gavin C. Conant*[,2,3]

[1]Department of Biology, University of New Mexico

[2]Division of Animal Sciences, University of Missouri

[3]Informatics Institute, University of Missouri

*Corresponding author: E-mail: conantg@missouri.edu.

Abstract

By comparing the patterns of evolution in the coding and upstream noncoding regions of yeast ribosomal protein (RP) genes duplicated in a genome duplication, we find that although nonsynonymous sites in the coding sequences show strong evidence for the fixation of recent gene conversion events, similar patterns are less evident among the synonymous positions and noncoding regulatory elements. This result suggests a potential explanation for the somewhat puzzling fact that duplicated RP genes are not functionally redundant despite their very high protein sequence identity. An analysis of the patterns of regulatory network evolution after genome duplication also indicates that the duplicated proteins have diverged considerably in expression despite their similar protein sequences.

Key words: gene duplication, gene conversion, ribosomal proteins, expression divergence, yeast.

Introduction

With the completion of the genomic sequencing of numerous organisms, it has become evident that polyploidization (or whole-genome duplication [WGD]) events have occurred in diverse lineages, including flowering plants (Blanc et al. 2000; The Arabidopsis Genome Initiative 2000; Tuskan et al. 2006), amoeba (Aury et al. 2006), and vertebrates (Meyer and Van de Peer 2005). The first such event to be detected in a whole-genome sequence was in *Saccharomyces cerevisiae* (Wolfe and Shields 1997): striking confirmation of this event was found with the two-to-one mapping of chromosomal regions in *S. cerevisiae* to the genomes of other yeasts lacking the WGD (Dietrich et al. 2004; Dujon et al. 2004; Kellis et al. 2004).

Polyploidization events are often followed by substantial losses of duplicated genes (Sémon and Wolfe 2007). Which of the two duplicate copies is lost is generally thought to be selectively neutral: if two populations lose alternative copies such "reciprocal gene loss" can contribute to reproductive isolation and hence speciation (Werth and Windham 1991), events inferred to have occurred in yeast by Scannell et al. (2006). The nature of the duplicate genes that are retained is also of interest: functional classes of genes such as

transcription factors, kinases, and ribosomal proteins (RPs) commonly remain duplicated after WGD but, surprisingly, are not generally duplicated in smaller events (Seoighe and Wolfe 1999; Blanc and Wolfe 2004; Maere et al. 2005; Aury et al. 2006; Freeling and Thomas 2006; Amoutzias et al. 2010). Wolfe (2000) has proposed the name *ohnologs* (in honor of Susumu Ohno) for these duplicate genes surviving from WGD. By mapping the relative genome orders of *S. cerevisiae* and seven related species, Byrne and Wolfe (2005) have provided an essentially complete list of *S. cerevisiae* ohnologs.

In *S. cerevisiae*, approximately 10% of surviving ohnologs are in fact genes encoding RPs (Planta and Mager 1998; Byrne and Wolfe 2005; Kim et al. 2009) despite the fact that these genes represented only 3.5% of the preduplication genome (Gordon et al. 2009). It has been suggested that the RP genes may be among those genes for which there is selection to maintain (high) relative gene dosage after WGD (i.e., the dosage balance hypothesis; Papp et al. 2003; Koszul et al. 2004; Freeling and Thomas 2006; Birchler and Veitia 2007; Edger and Pires 2009). Given this hypothesis, it is suggestive that many of the RP ohnologs are very similar in sequence; in fact, it is thought that these genes have undergone one or more gene conversion events
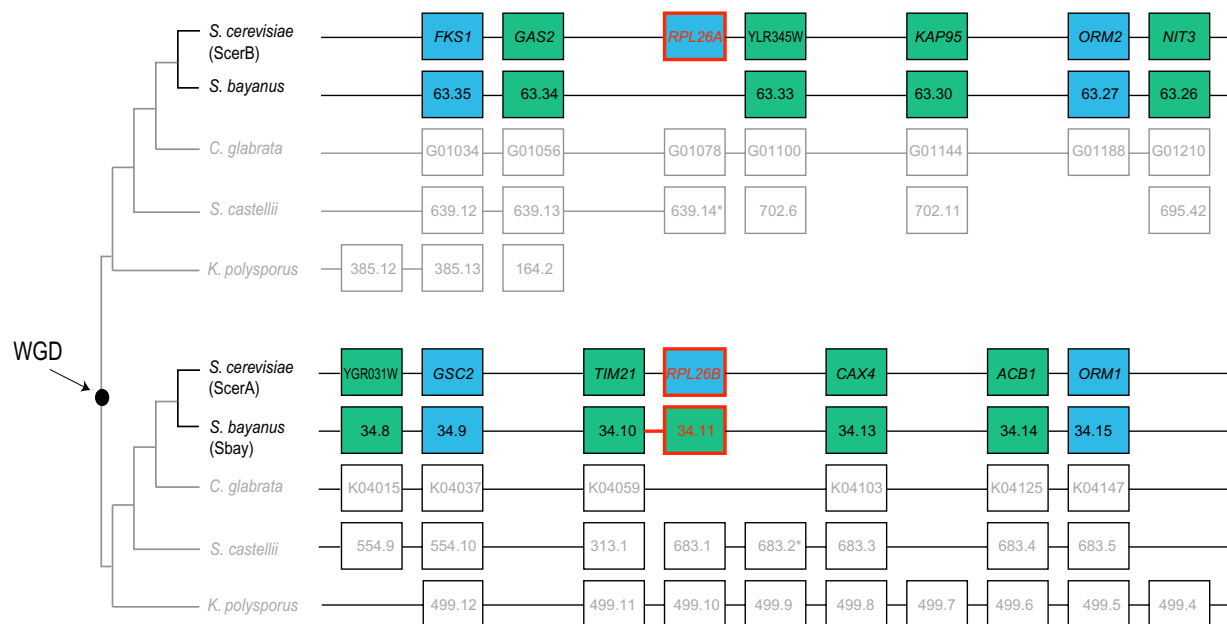
Fig. 1.—Illustration of the pattern of genome evolution after WGD in five yeast species in a region surrounding a pair of duplicated RP genes (RPL26A and RPL26B). The upper five tracks and the lower five tracks are inferred to be two orthologous groups. Lines connect genes that are adjacent on their respective contigs or chromosomes. Duplicate genes surviving from WGD are colored blue, green genes are cases where one member of the duplicate pair has been lost post-WGD. The orthology assignments between the paired Saccharomyces cerevisiae and S. bayanus genes on the upper and lower tracks are all inferred with greater than 99.99% confidence.

post-WGD (Kellis et al. 2004). Gene conversion occurs when recombination causes the overwriting of variations in one gene with the corresponding bases from another allele or paralog. Initial within-genome comparisons in the S. cerevisiae genome suggested reasonably high rates of gene conversion among gene duplicates (Drouin 2002); later work showed that conversion events among those duplicated genes tended to be confined to duplicate pairs retaining high sequence identity (Benovoy et al. 2005). In yeast, conversion can occur through a number of mechanisms, both during mitosis and meiosis (Chen et al. 2007). Interestingly, one of those mechanisms is mitotic conversion mediated by an mRNA or cDNA intermediate (Derr and Strathern 1993; Storici et al. 2007), as originally suggested by Baltimore (1985); notably, such a mechanism could impart biases in the location of conversion events.

One effect of gene conversion can be to erase the historical sequence divergence between paralogs, and one can plausibly argue that any functional differences between the two genes would be erased simultaneously. Curiously however, there are examples in yeast of paralogous RPs with high sequence identity (>97%) that nonetheless differ in their functional roles (Ni and Snyder 2001; Enyenihi and Saunders 2003; Kaeberlein et al. 2005; Komili et al. 2007; Kim et al. 2009).

Here, we examined the patterns of surviving gene conversions in the yeast RP ohnologs, finding strong evidence

for recent gene conversion at the nonsynonymous coding positions of these genes but little evidence of such conversion events in their upstream noncoding regions. An analysis of the RP ohnolog expression network also showed dissimilar expression patterns, consistent with regulatory divergence between the copies being responsible for the observed functional divergence.

Materials and Methods

Data Sources and Orthology Inference

A total of 55 previously described WGD-produced duplicate RPs (Planta and Mager 1998; Conant and Wolfe 2006) were analyzed. To this set, we added 84 pairs of enzyme genes duplicated at the WGD, identified by cross-referencing the list of metabolic genes of Kuepfer et al. (2005) to the set of S. cerevisiae ohnologs (Byrne and Wolfe 2005).

For these two lists (totaling 139 duplicate pairs), we next identified the corresponding orthologous genes in the genome of S. bayanus. Orthology inference in post-WGD species is challenging due to reciprocal gene loss, which can give rise to paired homologous genes that are paralogous rather than orthologous (fig. 1; Scannell et al. 2006, 2007). We have previously developed a maximum likelihood method that addresses this problem (Conant and Wolfe 2008). Briefly, the analysis begins with an inferred pre-WGD gene order (similar to that of Gordon et al. 2009).
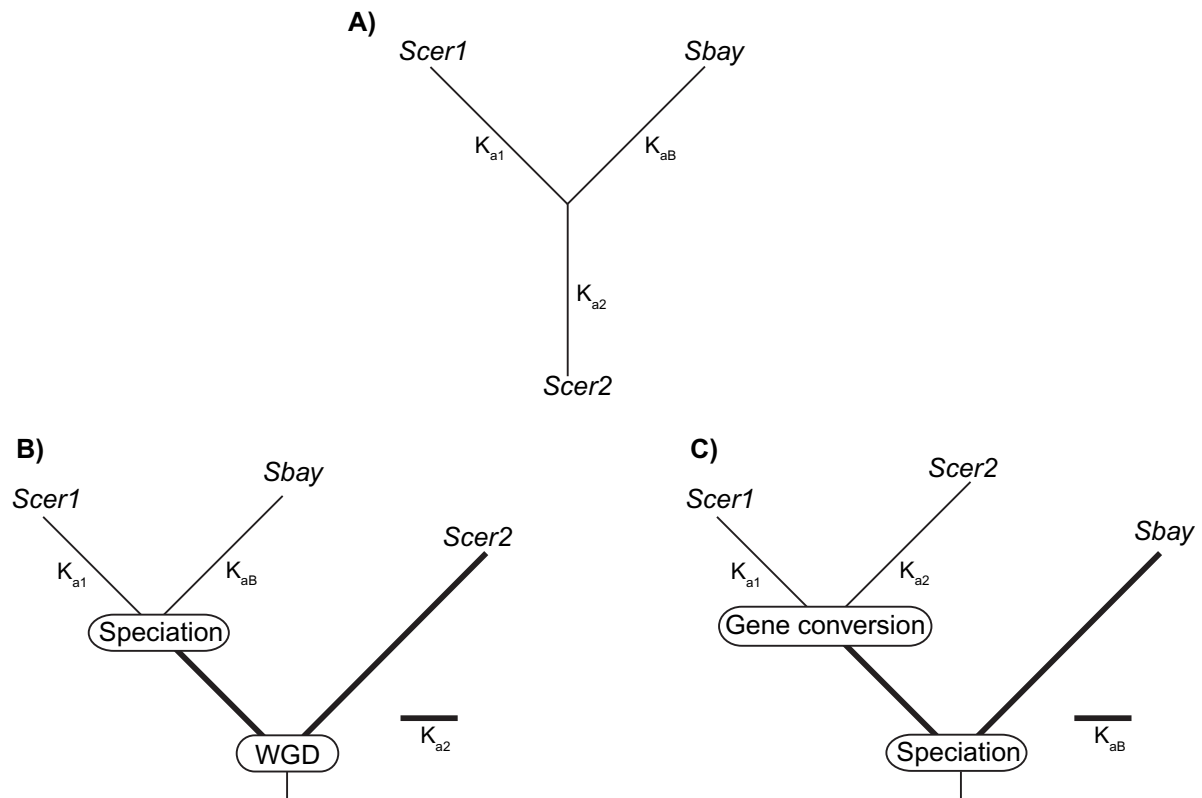
Fig. 2.—Analysis of duplicated *Saccharomyces cerevisiae* genes and a *S. bayanus* ortholog. (*A*) The format of our triplet-based sequence analysis. Because the models used are time reversible, only a single, three taxa tree is required. Independent estimates of K_a are made for each branch. (*B*) The expected pattern of branch lengths for the tree in *A* if the genes follow the known species tree. Note that we expect K_{a2} to be large as it represents both the divergence of the gene Scer2 as well as the shared divergence of Sbay and Scer1 post-WGD. (*C*) The expected gene tree if Scer1 and Scer2 have undergone recent gene conversion events. Here, we expect K_{aB} to be the largest of the three K_a values, under the same reasoning as in *B*.

A model of duplicate gene loss after WGD allows us to estimate the relative speciation times of the taxa analyzed and the probability of all possible orthology assignments. Thus, in figure 1, we estimate with greater than 99.99% confidence that *S. bayanus* gene number 34.11 is the ortholog of *S. cerevisiae* gene *RPL26B* as opposed to the alternative possible assignment that makes gene 34.11 the ortholog of *RPL26A*. Importantly, these inferences rest only on the relative gene orders: gene sequences are not considered.

From our list of 55 RP gene duplicates and 77 enzyme duplicates (metabolic protein [MP] genes) for which orthology estimation was possible, we selected the 29 RP gene pairs and 76 MP pairs for which the probability of the orthology assignment between *S. cerevisiae* and *S. bayanus* was >0.98. These genes represent a set for which we have high confidence orthology information independent of the sequences themselves.

Sequence Analyses

We next analyzed the sequence divergence in the coding regions of *S. cerevisiae* ohnolog pairs (fig. 2*A*). To do so, we aligned sequence triplets consisting of two ohnologs

from *S. cerevisiae* (Scer1 and Scer2 below) and the *S. bayanus* gene orthologous to Scer1 (Sbay below) using T-Coffee (Notredame et al. 2000). From these alignments, we estimated the number of nonsynonymous substitutions per nonsynonymous site (K_a) for each of the three branches in figure 2*A* by maximum likelihood, using our previously described software (Conant and Wagner 2003). Similar calculations are possible using the program HyPhy (Kosakovsky Pond et al. 2005). The same calculations were made for the synonymous sites (K_s). Note that for most *S. cerevisiae* ohnolog pairs, there are actually two possible triplets because the corresponding *S. bayanus* genes are also ohnologs. In such cases, we performed both comparisons (meaning that the identity of Scer1 and Scer2 was switched in the second case).

To test the statistical support for an inference of gene conversion between genes Scer1 and Scer2, we employed a likelihood ratio test (Sokal and Rohlf 1995). First, we identified cases where $K_{aB} > K_{a1}, K_{a2}$ (i.e., the signature of gene conversion; fig. 2*C*) and calculated the likelihood of the sequence alignment under this model (lnLH0). We then constrained the model such that $K_{a1} = K_{aB}$ and calculated the

likelihood under this alternative model (lnLHA). We compared 2·(lnLH0-lnLHA) to a chi-square distribution with 1 degree of freedom. An identical approach was used for the analysis of K_s.

To calculate pairwise noncoding region divergence, we first extracted the complete upstream intergenic region between each gene and the coding region of its 5′ neighbor. We then used the pairwise local alignment algorithm of Smith and Waterman (1981) to compute alignments between (Scer1 and Scer2), (Scer1 and Sbay), and (Scer2 and Sbay). Noncoding DNA tends to evolve rapidly (Lavoie et al. 2010), so to be sure that these alignments represent evolutionarily conserved regions and not simply statistical noise, we compared their local alignment scores against an expected distribution drawn from the genome at large. Briefly, using the complete collection of genes in the *S. cerevisiae* genome, we randomly selected 1,000 pairs and performed the same procedure of upstream region local alignment. We then compared the alignment scores obtained from these alignments with those seen for our orthologous and paralogous gene pairs. Scores in the upper 5% of this randomized distribution were inferred to show evolutionary conservation.

Significance of Ribosomal Gene Expression Network Partitioning

To determine if the RP gene expression data showed fewer crossing edges than would be expected by chance, we randomized the networks and recalculated the optimal partitioning. Randomization was performed by selecting every possible quartet of two pairs of duplicates. These four node subgraphs were replaced at random by another four node subgraphs with the same number of edges (Conant and Wolfe 2006). The probability of each such subgraph was calculated based on the inherent asymmetry in interaction degree between paralogs. Thus, we calculated the average fraction p of the total number of interactions for a paralog pair that belonged to the interaction-rich paralog. The probability of an interaction joining two interaction-rich genes is thus p^2, whereas the probability of an interaction joining an interaction-rich and interaction-poor gene is then $2p(1-p)$. Subgraph probabilities are calculated accordingly. The number of crossing edges in the original network was then compared with the distribution of number of crossing edges seen in 1,000 randomized networks.

Results

Strong Evidence for Numerous Gene Conversion Events among the Duplicated RPs

Using our previously described maximum likelihood model of gene loss following WGD (Conant and Wolfe 2008), we inferred orthologous chromosomal regions between two species sharing the WGD (*S. cerevisiae* and *Saccharomyces bayanus*) using only gene order information (fig. 1). As an aside, we note that we selected *S. bayanus* as an outgroup because it represents the most closely related yeast species for which the duplicated chromosomal segments created by the WGD have been mapped to their orthologous segments in *S. cerevisiae* (Byrne and Wolfe 2005; Conant and Wolfe 2008; Gordon et al. 2009). Our gene order-based approach allowed us to conclude with high confidence that all the duplicate gene loci we considered evolved according to the species tree in figure 2B (see Materials and Methods). Despite this fact, it is not necessarily the case that the sequences themselves will follow this set of relationships. In particular, a gene conversion event between Scer1 and Scer2 that occurred after the speciation of *S. cerevisiae* and *S. bayanus* would overwrite the historical signal in the sequences of the two genes and give rise to a gene tree of the form of figure 2C.

Using estimates of K_a for triplets of RP genes (see Materials and Methods), we asked whether the pattern of nonsynonymous divergence in each triplet was most compatible with divergence after WGD (i.e., $K_{a2} > K_{a1}, K_{aB}$, fig. 2B) or with a recent gene conversion event ($K_{aB} > K_{a1}, K_{a2}$, fig. 2C). Of the 29 pairs of duplicated RP genes in *S. cerevisiae*, two follow the pattern expected under WGD, two pairs present conflicting patterns of K_a values depending on the *S. bayanus* ortholog used, and the remaining 25 pairs have nonsynonymous divergences consistent with gene conversion. Intriguingly, when we use the same approach with the synonymous substitutions in the RP genes, we find many fewer cases where gene conversion needs to be invoked: only 10 of 29 duplicate pairs show any evidence of gene conversion at the synonymous sites. This difference is statistically significant ($P = 0.0001$, Fisher's exact test).

We next assessed whether the signature of gene conversion in the sequence data was strong enough to statistically reject the possibility that phylogenetic relationships within the triplet were simply ambiguous. Thus, we compared a model allowing gene conversion ($K_{aB} > K_{a1}, K_{a2}$) with an alterative model where K_{aB} was constrained to be equal to K_{a1}. Of the 25 RP duplicate pairs with signatures of gene conversion in K_a, 17 showed statistically significant improvement when a model allowing gene conversion was used ($P < 0.05$, likelihood ratio test). For the synonymous sites, only 6 of the 29 pairs showed significant evidence for gene conversion ($P < 0.05$).

Using a GenomeHistory homology search (Conant and Wagner 2002), we identified two pairs of duplicated RP genes in the *S. cerevisiae* genome that do not appear to derive from WGD: *RPL9A/B* and *RPS22A/B*. Using synteny data from the Yeast Genome Order Browser (Byrne and Wolfe 2005) to assign orthology between *S. cerevisiae* and *S. bayanus*, we find that *RPL9A* and *B* show significant evidence of gene conversion in K_a (but not K_s) following the split with

Table 1

Prevalence of Gene Conversion in Coding Regions and Noncoding Regions

| Gene Class | Coding Regions (K_a) | | Upstream Regions | | |
	Gene Conversion[a]	WGD[b]	Gene Conversion[a]	WGD[b]	P[c]
RP[d]	12	3	1	14	0.0002
MP[e]	1	38	5	34	0.2

[a] Cases where the two *Saccharomyces cerevisiae* paralogs share higher sequence identity to each other than either does to its respective ortholog (see text).

[b] Cases where at least one *S. cerevisiae* paralog shows higher sequence identity to its ortholog than to the other *S. cerevisiae* paralog.

[c] P value for the test of equal proportions of gene conversion events in the coding and upstream regions (Fisher's exact test).

[d] RP gene duplicates.

[e] MP gene duplicates.

S. bayanus ($P < 0.003$), but *RPL22A* and *B* show no signature of recent conversion.

Metabolic Genes Duplicated at WGD Do Not Show Similar Patterns of Gene Conversion

We applied the above approach to a similar set of WGD-duplicated MP genes. Among the 76 pairs considered, only three show any signs of gene conversion in K_a and only two of those have significant improvement when the gene conversion model is used ($P < 10^{-6}$, likelihood ratio test). The corresponding number of pairs showing signs of gene conversion for K_s are 6 and 4, ($P < 0.008$). These differences in the proportion of observed gene conversion events in K_a and in K_s between the RP and MP groups are highly significant ($P < 10^{-10}$ and $P < 0.002$, respectively; Fisher's exact test).

RP Noncoding Regions Do Not Show Evidence of Gene Conversion

For each of the RP gene pairs considered above, we measured the sequence identity in their upstream noncoding regions. Among the pairs considered, 15 RP ohnolog pairs had local alignment scores significantly larger than would be expected for unrelated regions. For these pairs, we compared the alignment score $S_{1,2}$ of the ohnolog pair (Scer1, Scer2) to the scores from the comparison of each paralog to Scer1's ortholog in *S. bayanus* (i.e., $S_{1,B}$ for Scer1, Sbay and $S_{2,B}$ for Scer2, Sbay). Cases where $S_{1,2} > S_{1,B}, S_{2,B}$ were interpreted as evidence of upstream conversion. We found that only 1/15 (6%) of the triplets of pairwise noncoding alignments showed evidence of gene conservation compared with the 12/15 (80%) for K_a in the coding regions (from the analysis above; table 1). This difference in the prevalence of conversion events between the two locations is highly significant ($P = 0.0002$, Fisher's exact test; table 1). Interestingly, when this same approach is applied to 39 MP genes, we find very few instances of gene conversion in

either region (<10%) and no significant difference in the proportion of conversion events between the noncoding and coding regions (table 1).

To further test for gene conversion in the RP gene upstream regions, we asked whether the RP ohnologs share more common upstream regulatory motifs than do the *S. cerevisiae–S. bayanus* orthologs. Using the motifs defined in Kellis et al. (2003), we counted the number of shared motifs between genes Scer1 and Sbay and between Scer1 and Scer2. Contrary to the prediction of gene conversion, the average number of shared motifs was higher for the orthologs than for the ohnologs (0.51 vs. 0.33), although this difference was not statistically significant ($P = 0.28$, likelihood ratio test). Similar results were seen for the MP genes (data not shown).

Analysis of Duplicated RP Gene Expression Networks

We have previously described an algorithm for detecting network partitioning among WGD-produced duplicate genes (Conant and Wolfe 2006). As is illustrated in figure 3, paralogs are divided into two columns with ohnologs opposite each other. Gene expression data for 51 pairs of RP ohnologs were obtained from the expression compendia of Hughes et al. (2000) and overlaid as graph edges. We defined an edge between any two genes if they shared a correlation (Pearson's r) in gene expression of 0.8 or greater across the set of more than 300 experiments (a threshold of 0.75 produced similar results; data not shown). We divide these edges into internal edges, connecting nodes in the same column (arcs or vertical lines in fig. 3), and crossing edges, joining nodes in opposite columns (diagonal lines in fig. 3). Note that the initial assignment of a particular paralog to the first or second column is arbitrary, meaning that there are 2^{n-1} possible unique partitioning of the duplicates into columns. We thus calculated the network partitioning that resulted in the fewest number of crossing edges (see Materials and Methods). For these data, the optimal partitioning had 102 crossing edges, which, although it appears to be a large number, is significantly smaller than the minimum number of crossing edges seen in the randomized networks ($P < 0.001$). It is also relevant to note the extreme degree of asymmetry evident in this figure: the paralogous RPs, despite their sequence similarity, do not have identical expression patterns.

Discussion

We have found that although duplicated RP genes created by WGD show strong evidence of surviving recent gene conversion in their coding regions, the same is not true of the upstream noncoding regions. Moreover, even in the coding regions, there is a bias toward detecting conversion events at nonsynonymous positions. One role of
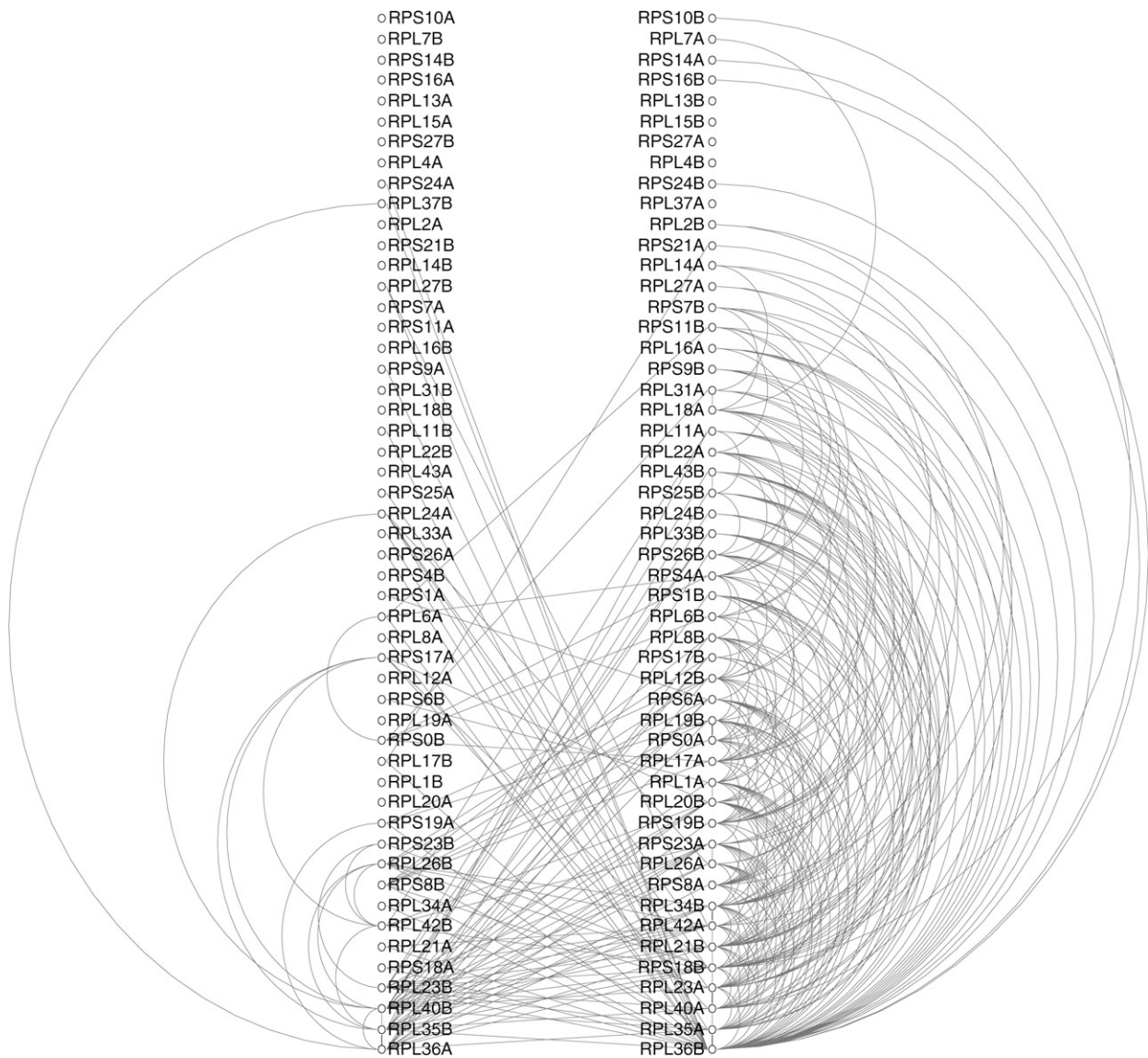
Fig. 3.—Gene expressions networks of duplicated RPs have diverged since WGD. Pairs of duplicated RPs are arranged opposite each other. Edges connect pairs of genes with correlation in gene expression >0.8. We searched among the 2^{n-1} permutations of the column arrangements to find this arrangement, which has the minimal number of edges (102) crossing between the two partitions. The minimal number of crossing edges seen in randomized networks was 107, the mean was 116. Note also the high degree of asymmetry in the number of interactions seen between duplicated RPs.

these results is to shed light on previous analyses. For instance, Gao and Innan (2004) argued that high rates of gene conversion had resulted in overestimates of the rate of yeast gene duplication. However, the data set used by these authors was heavily biased toward RP genes: of the 68 duplicate pairs considered, fully 50 of them fall into the set of the 55 WGD-produced RP duplicate genes. Our results imply that RPs are likely to be somewhat unique in both their patterns of duplication and of gene conversion and therefore probably should not be used as a proxy for the genome at large.

The conservation in the sequences of the encoded RPs induced by gene conversion is not unexpected as these proteins are highly conserved across a wide range of taxa (Alksne et al. 1993; Manuell et al. 2005; Ross et al. 2007). RP genes are also somewhat unusual in their response to genome duplication: they have survived in excess after WGDs in a variety of genomes including yeast (Seoighe and Wolfe 1999; Blanc and Wolfe 2004; Maere et al. 2005; Aury et al. 2006).

Given the existence of cDNA or mRNA-based gene conversion (Derr and Strathern 1993; Storici et al. 2007), events

that would be naturally limited to transcribed regions, one might ask if the patterns observed here might simply represent a mutational bias in the conversion events themselves. This model is attractive because RP genes are highly expressed and recombination and gene conversion are associated with high levels of transcription (Aguilera and Gómez-González 2008). However, a purely mutational bias explanation for these repeated gene conversion events is unsatisfying for three reasons. First, no bias against upstream conversion is evident for meiotic recombination. Given that the per-cell-cycle rate of conversion is roughly four orders of magnitude greater for meiosis than for mitosis (Barbera and Petes 2006), even the roughly 1,000-fold excess of mitotic to meiotic cell divisions seen in wild relatives of *S. cerevisiae* (Tsai et al. 2008) is insufficient to give an overwhelming signature of coding region gene conversion. Secondly, a mutational bias does not explain the absence of gene conversion among the metabolic genes: of the nine MP gene pairs with expression levels (Holstege et al. 1998) that are as high as those of the RP gene pairs, only one shows evidence of gene conversion (data not shown). Finally, such a bias does not explain the fact that synonymous sites in the RP genes show much less evidence for conversion events than do nonsynonymous sites. Thus, we argue that in addition to any mutational biases, some intervening process of selection is helping to fix nonsynonymous conversion events.

Another partial explanation for the similarity in yeast RP coding sequences is selection for high dosages of these proteins. There is evidence for dosage benefits from RP gene duplication (Koszul et al. 2004). However, this explanation does not wholly explain the biology of these ohnologs, as we will discuss below. Moreover, such conservation does not appear to be a general response to dosage selection: metabolic genes also likely survived in duplicate partly due to dosage selection (Blank et al. 2005; Piškur et al. 2006; Conant and Wolfe 2007; Merico et al. 2007; van Hoek and Hogeweg 2009) and yet do not have strong signatures of conversion.

Instead, these results provide evidence that even if the coding regions of the duplicated RP genes are still being homogenized by gene conversion, their expression patterns have diverged considerably since the WGD. In fact, a number of recent analyses that have demonstrated that the duplicated yeast RP genes are not, in fact, functionally interchangeable. Thus, several RP genes, but not their paralogs, have been shown to be essential for determining bud location in *S. cerevisiae* (Ni and Snyder 2001) and for localizing proteins to that bud (Komili et al. 2007). Similar patterns have been seen in *Brassica napus*, where RP genes show tissue-specific expression despite high amino acid sequence identity (Whittle and Krochko 2009).

An equally intriguing case is the difference in protein localization between the RP paralogs Rpl7a and Rpl7b. Rpl7a is much more highly expressed than is Rpl7b (Ghaemmaghami et al. 2003) but while Rpl7a is only found in the cytoplasm, Rpl7b, despite its lower abundance, is found both in the cytoplasm and in the nucleolus (Kim et al. 2009). This difference does not appear to be caused by variations in the coding sequences of the two genes: replacing the *RPL7B* sequence with that from *RPL7A* does not alter the cellular localization of the protein encoded at that locus (Kim et al. 2009). These authors propose that the localization difference is instead driven by preferential incorporation of Rpl7a into ribosomal subunits, meaning that the free protein is rarely present at the site of ribosome subunit assembly in the nucleolus. However, the origins of this difference in incorporation rate remain unclear given the apparent equivalence of the two protein sequences seen after sequence replacement.

What combination of phenomena might give rise to expression divergence coupled to strong protein sequence homogenization? One possibility is expression subfunctionalization. This hypothesis is partly supported by our network analysis, which indicates partial expression isolation between two groups of RP duplicate genes. Because the ribosome represents a tightly integrated functional module, expression subfunctionalization might still require very high degrees of protein identity between the RP paralogs, such that the proteins encoded by these paralogs are able to substitute for each other under the different expression conditions. This hypothesis of strong purifying selection acting on RP coding sequences is supported by the observation that these genes show fewer single-nucleotide polymorphisms per base pair than do the MP genes (0.002 vs. 0.007; polymorphism data taken from Schacherer et al. 2009). Likewise, an increased frequency of gene-expression coupled gene conversion events, as discussed above, could very well improve the ability of natural selection to maintain such coding sequence conservation. On the expression front, the subfunctionalization itself might be either quantitative (only expression of both paralogs gives sufficient protein product) or qualitative (the expression of the two paralogs varies with respect to each other temporally). Similarly, the process might follow either the purely neutral DDC model originally proposed by several authors (Force et al. 1999; Stoltzfus 1999) or involve other selective forces, including adaptive ones (Des Marais and Rausher 2008): for review, see Innan and Kondrashov (2010). Neofunctionalization is another possible explanation for the expression divergence among the RP genes, but we are skeptical that it would occur on this scale (more than 50 genes).

Finally, we note that a broader perspective argues that gene dosage and subfunctionalization are not mutually exclusive explanations for the fate of a duplicate gene pair: He and Zhang (2005) propose that a duplication might be initially preserved by subfunctionalization and then might later undergo neofunctionalization. In the case of the RPs, we

suspect that the initial preservation of the duplicated RP genes was for reasons of gene dosage; this process was likely followed by subfunctionalization (He and Zhang 2005; Innan and Kondrashov 2010).

The key prediction of our model is that gene conversion among the RPs helps to maintain a coadapted functional module: the ribosome. In future, it would be very interesting to test if this prediction of coadaptation holds.

Acknowledgments

Literature Cited

Aguilera A, Gómez-González B. 2008. Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet. 9:204–217.

Alksne LE, Anthony RA, Liebman SW, Warner JR. 1993. An accuracy center in the ribosome conserved over 2 billion years. Proc Natl Acad Sci U S A. 90:9538–9541.

Amoutzias GD, et al. 2010. Posttranslational regulation impacts the fate of duplicated genes. Proc Natl Acad Sci U S A. 107:2967–2971.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 408:796–815.

Aury JM, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature. 444:171–178.

Baltimore D. 1985. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. Cell. 40:481–482.

Barbera MA, Petes TD. 2006. Selection and analysis of spontaneous reciprocal mitotic cross-overs in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 103:12819–12824.

Benovoy D, Morris RT, Morin A, Drouin G. 2005. Ectopic gene conversions increase the G + C content of duplicated yeast and Arabidopsis genes. Mol Biol Evol. 22:1865–1868.

Birchler JA, Veitia RA. 2007. The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell. 19:395–402.

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M. 2000. Extensive duplication and reshuffling in the Arabidopsis genome. Plant Cell. 12:1093–1101.

Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell. 16:1679–1691.

Blank LM, Lehmbeck F, Sauer U. 2005. Metabolic-flux and network analysis of fourteen hemiascomycetous yeasts. FEMS Yeast Res. 5:545–558.

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

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

Conant GC, Wagner A. 2002. GenomeHistory: a software tool and its application to fully sequenced genomes. Nucleic Acids Res. 30:3378–3386.

Conant GC, Wagner A. 2003. Asymmetric sequence divergence of duplicate genes. Genome Res. 13:2052–2058.

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

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

Conant GC, Wolfe KH. 2008. Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. Genetics. 179:1681–1692.

Derr LK, Strathern JN. 1993. A role for reverse transcripts in gene conversion. Nature. 361:170–173.

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature. 454:762–765.

Dietrich FS, et al. 2004. The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. Science. 304:304–307.

Drouin G. 2002. Characterization of the gene conversions between the multigene family members of the yeast genome. J Mol Evol. 55:14–23.

Dujon B, et al. 2004. Genome evolution in yeasts. Nature. 430:35–44.

Edger PP, Pires JC. 2009. Gene and genome duplications: the impact of dosage sensitivity on the fate of nuclear genes. Chromosome Res. 17:699–717.

Enyenihi AH, Saunders WS. 2003. Large-scale functional genomic analysis of sporulation and meiosis in Saccharomyces cerevisiae. Genetics. 163:47–54.

Force A, et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531–1545.

Freeling M, Thomas BC. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. 16:805–814.

Gao LZ, Innan H. 2004. Very low gene duplication rate in the yeast genome. Science. 306:1367–1370.

Ghaemmaghami S, et al. 2003. Global analysis of protein expression in yeast. Nature. 425:737–741.

Gordon JL, Byrne KP, Wolfe KH. 2009. Additions, Losses, and Rearrangements on the Evolutionary Route from a Reconstructed Ancestor to the Modern Saccharomyces cerevisiae Genome. PLoS Genet. 5(5):e1000485.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics. 169:1157–1164.

Holstege FCP, et al. 1998. Dissecting the regulatory circuitry a eukaryotic genome. Cell. 95:717–728.

Hughes TR, et al. 2000. Functional discovery via a compendium of expression profiles. Cell. 102:109–126.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 11:97–108.

Kaeberlein M, et al. 2005. Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. Science. 310:1193–1196.

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

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature. 423:241–254.

Kim T-Y, Ha CW, Huh W-K. 2009. Differential subcellular localization of ribosomal protein L7 paralogs in Saccharomyces cerevisiae. Mol Cells. 27:539–546.

Komili S, Farny NG, Roth FP, Silver PA. 2007. Functional specificity among ribosomal proteins regulates gene expression. Cell. 131:557–571.

Kosakovsky Pond SL, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21:676–679.

Koszul R, Caburet S, Dujon B, Fischer G. 2004. Eukaryotic genome evolution through the spontaneous duplication of large chromosomal segments. EMBO J. 23:234–243.

Kuepfer L, Sauer U, Blank LM. 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. Genome Res. 15:1421–1430.

Lavoie H, et al. 2010. Evolutionary tinkering with conserved components of a transcriptional regulatory network. PLoS Biol. 8: e1000329.

Maere S, et al. 2005. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A. 102:5454–5459.

Manuell AL, Yamaguchi K, Haynes PA, Milligan RA, Mayfield SP. 2005. Composition and structure of the 80S ribosome from the green alga *Chlamydomonas reinhardtii*: 80S ribosomes are conserved in plants and animals. J Mol Biol. 351:266–279.

Merico A, Sulo P, Piskur J, Compagno C. 2007. Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. FEBS J. 274:976–989.

Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). BioEssays. 27:937–945.

Ni L, Snyder M. 2001. A genomic study of the bipolar bud site selection pattern in *Saccharomyces cerevisiae*. Mol Biol Cell. 12: 2147–2170.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol. 302:205–217.

Papp B, Pal C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. Nature. 424:194–197.

Piškur J, Rozpedowska E, Polakova S, Merico A, Compagno C. 2006. How did *Saccharomyces* evolve to become a good brewer? Trends Genet. 22:183–186.

Planta RJ, Mager WH. 1998. The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. Yeast. 14:471–477.

Ross CL, Patel RR, Mendelson TC, Ware VC. 2007. Functional conservation between structurally diverse ribosomal proteins from *Drosophila melanogaster* and *Saccharomyces cerevisiae*: fly L23a can substitute for yeast L25 in ribosome assembly and function. Nucleic Acids Res. 35:4503–4514.

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature. 440:341–345.

Scannell DR, et al. 2007. Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Natl Acad Sci U S A. 104:8397–8402.

Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. 2009. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. Nature. 458:342–345.

Sémon M, Wolfe KH. 2007. Consequences of genome duplication. Curr Opin Genet Dev. 17:505–512.

Seoighe C, Wolfe KH. 1999. Yeast genome evolution in the post-genome era. Curr Opin Microbiol. 2:548–554.

Smith TF, Waterman MS. 1981. Identification of common molecular subsequences. J Mol Biol. 147:195–197.

Sokal RR, Rohlf FJ. 1995. Biometry, 3rd ed. New York: W.H. Freeman and Company.

Stoltzfus A. 1999. On the possibility of constructive neutral evolution. J Mol Evol. 49:169–181.

Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. 2007. RNA-templated DNA repair. Nature. 447:338–341.

Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. Proc Natl Acad Sci U S A. 105:4957–4962.

Tuskan GA, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 313:1596–1604.

van Hoek MJ, Hogeweg P. 2009. Metabolic adaptation after whole genome duplication. Mol Biol Evol. 26:2441–2453.

Werth CR, Windham MD. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am Nat. 137:515–526.

Whittle CA, Krochko JE. 2009. Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in Brassica napus. Plant Cell. 21:2203–2219.

Wolfe KH. 2000. Robustness: it's not where you think it is. Nat Genet. 25:3–4.

Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 387:708–713.

Associate editor: Laurence Hurst