Molecular Systems Biology 6; Article number 429; doi:10.1038/msb.2010.82
Citation: ***Molecular Systems Biology*** **6:429**

www.molecularsystemsbiology.com

molecular systems biology

Genetic interactions reveal the evolutionary trajectories of duplicate genes

Benjamin VanderSluis[1], Jeremy Bellay[1], Gabriel Musso[2,3], Michael Costanzo[2], Balázs Papp[4,5], Franco J Vizeacoumar[2], Anastasia Baryshnikova[2,3], Brenda Andrews[2,3], Charles Boone[2,3] and Chad L Myers[1,*]

[1] Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA, [2] Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada, [3] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada, [4] Institute of Biochemistry, Biological Research Center, Szeged, Hungary and [5] Department of Genetics, Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK

* Corresponding author. Department of Computer Science and Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455, USA. Tel.: +1 612 624 8306; Fax: +1 612 625 0572; E-mail: cmyers@cs.umn.edu

Received 4.5.10; accepted 27.9.10

The characterization of functional redundancy and divergence between duplicate genes is an important step in understanding the evolution of genetic systems. Large-scale genetic network analysis in *Saccharomyces cerevisiae* provides a powerful perspective for addressing these questions through quantitative measurements of genetic interactions between pairs of duplicated genes, and more generally, through the study of genome-wide genetic interaction profiles associated with duplicated genes. We show that duplicate genes exhibit fewer genetic interactions than other genes because they tend to buffer one another functionally, whereas observed interactions are non-overlapping and reflect their divergent roles. We also show that duplicate gene pairs are highly imbalanced in their number of genetic interactions with other genes, a pattern that appears to result from asymmetric evolution, such that one duplicate evolves or degrades faster than the other and often becomes functionally or conditionally specialized. The differences in genetic interactions are predictive of differences in several other evolutionary and physiological properties of duplicate pairs.
Molecular Systems Biology **6**: 429; published online 16 November 2010; doi:10.1038/msb.2010.82
Subject Categories: bioinformatics; functional genomics
Keywords: duplicate genes; functional divergence; genetic interactions; paralogs; *Saccharomyces cerevisiae*

Introduction

Gene duplication is a primary mechanism for generating functional novelty, because it allows for the relaxation of selective constraints and thus provides an opportunity for functional innovation or specialization (Ohno, 1970). Genome sequencing studies in several species have revealed that a sizable fraction of many genomes are duplicated and that paralogous genes retain a relatively high degree of sequence similarity (Kellis *et al*, 2004; Byrne and Wolfe, 2005). In addition to the similarity of nucleotide/amino-acid sequence, functional genomic studies have identified significant overlap between duplicate genes in terms of their physical interactions (Baudot *et al*, 2004; Guan *et al*, 2007; Musso *et al*, 2007; Wapinski *et al*, 2007), fitness effects (Gu *et al*, 2003), metabolic activity (Papp *et al*, 2004; Kuepfer *et al*, 2005) and gene expression patterns (Gu *et al*, 2002b), providing further evidence to suggest that functional similarity among duplicate gene families has been actively retained for over millions of years (Kellis *et al*, 2004; Kafri *et al*, 2006).

Genetic interaction analysis offers another means to assess functional relationships between duplicated genes. A genetic interaction refers to an unexpected phenotype not easily explained by combining the effects of the individual genetic variants (Dixon *et al*, 2009). This phenomenon is also generally referred to as epistasis by the statistical genetics and evolution communities and can refer to phenotypes that are either aggravated (synergistic combinations) or alleviated (antagonistic combinations) in combination with other variants. Synthetic lethality represents an extreme form of negative genetic interaction in which mutation of a single gene, although having little or no effect on the organism, results in cell death when combined with mutation of a second gene (Dobzhansky, 1946; Novick *et al*, 1989). Negative genetic interactions are often taken as evidence of a functional relationship and, as a result, can be used to directly assess

the extent of functional redundancy between genes. Indeed, a systematic survey identified negative interactions between 35% of gene pairs arising from the whole-genome duplication (WGD) event (Musso *et al*, 2008). This rate represents an approximately 20-fold enrichment over random pairs and confirms that functional redundancy is pervasive among duplicate pairs (DeLuna *et al*, 2008; Dean *et al*, 2008; Musso *et al*, 2008). Despite this wealth of data, we lack models that reconcile the long-term preservation of redundancy among duplicate genes with their patterns of functional divergence.

Synthetic genetic array (SGA) methodology enables large-scale analysis of genetic interactions in yeast (Tong *et al*, 2001, 2004; Costanzo *et al*, 2010), which can extend our view beyond individual duplicate pair interactions to systematically examine the subsets of genetic interactions between duplicate genes and the rest of the genome. Analogous to studies based on protein–protein interactions (PPIs), the number of negative genetic interactions for a given duplicate pair and the extent to which their interactions overlap should provide insight into functional similarities and relationships between duplicate gene pairs. Furthermore, genes belonging to the same biological pathway or protein complex often share similar profiles or patterns of genetic interactions (Tong *et al*, 2004). As a result, genes can be assigned into specific pathways or complexes by virtue of their genetic interaction profile similarity, as measured across a large fraction of the genome (Tong *et al*, 2004; Costanzo *et al*, 2010). This approach was adopted to examine the interaction profiles for 90 duplicate genes within a functionally biased subset of gene deletion mutants queried against itself (Ihmels *et al*, 2007). This analysis showed that even though duplicate genes display negative genetic interactions with each other, they also appear to behave like singleton genes, in that they exhibit numerous unique genetic interactions; the authors suggest that duplicates are functionally redundant but have divergent roles because they often fail to provide a genuine backup when another gene is deleted (Ihmels *et al*, 2007).

In the current work, we explore evidence for duplicate gene redundancy in their genetic interaction profiles and further explain the previously observed lack of similarity among the interaction profiles of duplicate gene pairs (Ihmels *et al*, 2007). Specifically, we propose that the established ability for many duplicate genes to buffer one another under certain conditions should cause genetic interactions related to common functions to be hidden from our experimental method. Furthermore, as duplicates evolve away from complete redundancy, non-overlapping genetic interactions should appear, reflecting their divergent roles. We find evidence to support these hypotheses in a genome-wide collection of quantitative genetic interactions in *Saccharomyces cerevisiae* (Costanzo *et al*, 2010). We show that exceptions to the model provide insight into evolutionary mechanisms of duplicate gene retention by distinguishing partially redundant genes maintained because of their functional divergence (Ohno, 1970; Hughes, 1994; Force *et al*, 1999; Conant and Wolfe, 2008; Marques *et al*, 2008) from those pairs retained because increased gene dosage is beneficial to the organism (Kondrashov and Kondrashov, 2006; Conant and Wolfe, 2007; Ihmels *et al*, 2007). Finally, we provide evidence based on genetic interaction profiles supporting an asymmetric model of divergence, and show a connection between genetic interaction asymmetry and other physiological and phylogenetic properties.

Results

A hypothesis about the buffering of genetic interactions after gene duplication

We hypothesize that immediately after a duplication event, duplicate genes are identical and presumably redundant, and thus, the only genetic interaction that either paralog exhibits should be with its sister gene (Figure 1A and B). Such a scenario cannot persist without selection pressure to maintain the now redundant copies (Brookfield, 1992). As the pair diverges, the selective pressures that maintained the ancestral gene will begin to act on each duplicate copy individually, creating unique genetic interactions (Figure 1C). Implicit in this hypothesis is the fact that genetic interactions are buffered and undetectable immediately after a duplication event, and then are gradually revealed in one sister duplicate or the other as the pair diverges (Figure 1C). The interactions that emerge after duplication may include the original ancestral genetic interactions that were buffered by the duplication or they may reflect a new function unique to one member of the pair, instances of sub- or neo-functionalization, respectively. On the basis of this hypothesis in which common functions are buffered, genetic interactions should reveal how paralogs have diverged, but seldomly reveal their common functions. Requisite to this reduction in common interactions is the ability of a duplicate gene to partially compensate for the loss of its sister, which has been well established in previous studies (Supplementary Figure 1; Gu *et al*, 2003; Ihmels *et al*, 2007; DeLuna *et al*, 2008; Dean *et al*, 2008; Musso *et al*, 2008).

Large-scale SGA data confirms an enrichment of negative genetic interactions among duplicates

To first affirm previous evidence for duplicate redundancy, we extracted genetic interactions for 576 duplicated *S. cerevisae* gene pairs (461 WGDs and 115 small-scale duplicates (SSD); see Materials and methods) from our recent quantitative and genome-scale SGA analysis (Costanzo *et al*, 2010). This study captures both negative interactions, those in which the double mutant was less fit than expected (synergism of mutation effects), and positive interactions, those in which the double mutant was more fit than expected (antagonism of mutation effects). Because our SGA study focused on only genetic interactions involving two genes, we restricted our analysis to two-gene duplicate families.

A primary requisite of the duplicate buffering hypothesis is that sister duplicates should show negative genetic interactions with each other, indicating at least partial redundancy among paralogs (Figure 1C). We found a striking enrichment for negative genetic interactions between sister duplicates (67/205 pairs; 33%; Figure 2A; Supplementary Table 1), which was consistent with previous findings (35% (Musso *et al*, 2008); 34% (Dean *et al*, 2008); 55% (DeLuna *et al*, 2008)). This is substantially higher than the negative genetic interaction rate among randomly selected gene pairs (1.8%; Costanzo *et al*,

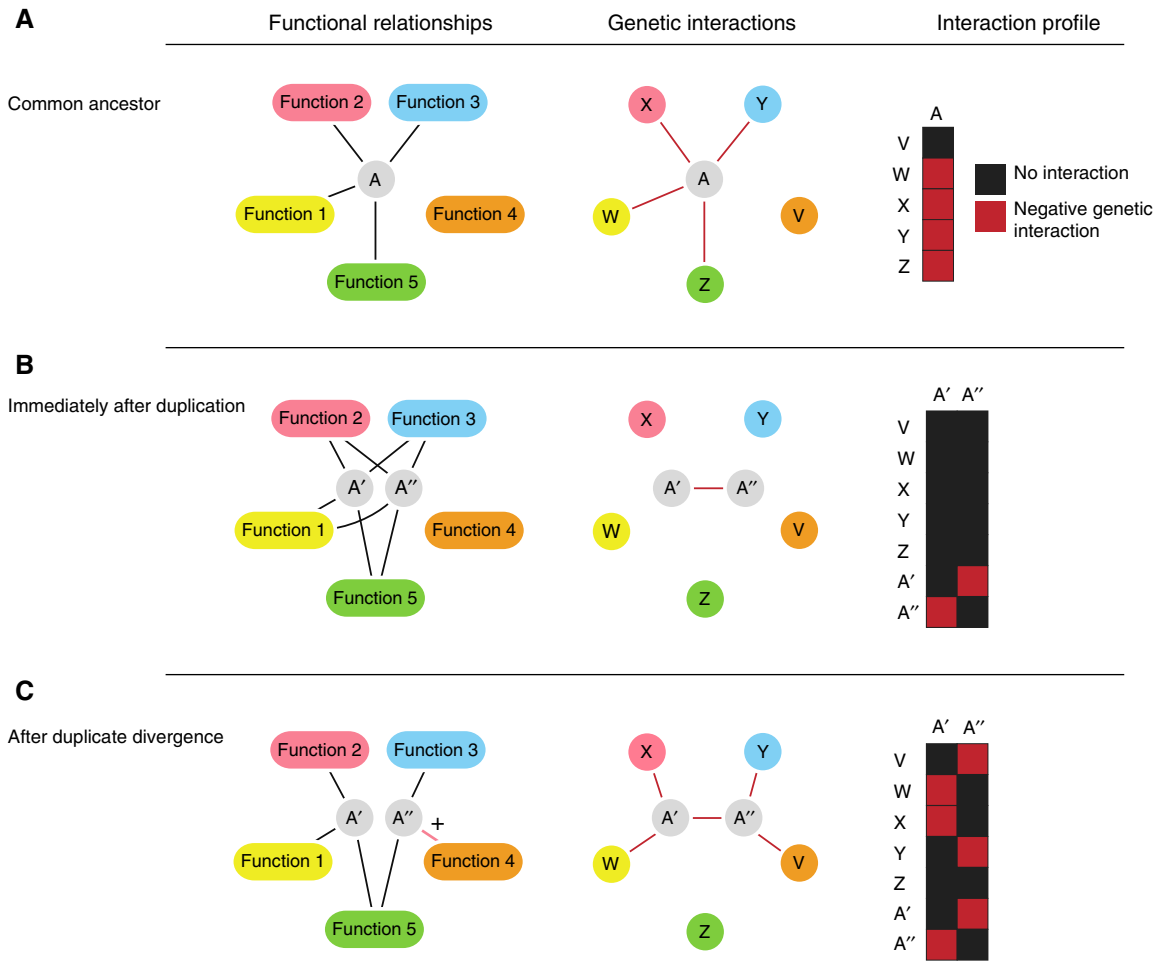

Figure 1 A model for the buffering of genetic interactions by partially redundant genes. The figure illustrates the relationship between a functional membership network, the observable genetic interaction network and the corresponding genetic interaction profiles, over the course of a duplication event and subsequent divergence. (**A**) Gene A has no redundant partner and its set of functional relationships is revealed through negative genetic interactions. The interaction profile for gene A is complete. (**B**) Immediately after duplication, genes A′ and A″ are fully redundant and their functional relationships are shared. Because each is capable of performing their common functions without the other, the deletion of A′ and A″ have negligible effects and do not exhibit negative interactions with any other genes. However, the simultaneous deletion of A′ and A″ reveals the original phenotype of their ancestor, and thus shows a negative genetic interaction. (**C**) A′ and A″ diverge, the redundancy becomes incomplete and unique deletion consequences emerge for each duplicate. Some of the negative genetic interactions observed for the ancestor gene A are not observed following duplication and divergence; for example, despite the functional relationship between A′ and A″ and Z, negative interactions are not observed with Z. A″ has evolved a new relationship with function 4(+). A′ lacks this ability and thus we see a genetic interaction between A″ and V.

2010), as well as the corresponding rate between physically interacting pairs (7%, $P<5\times10^{-23}$; Figure 2A; see Materials and methods) or pairs sharing specific functional annotations (4%; Myers *et al*, 2006). Although enrichment was observed for both WGD and SSD paralogs, the genetic interaction rate was significantly higher among WGD pairs ($P<5\times10^{-2}$; Figure 2B; see Materials and methods), supporting the greater retained functional overlap observed in general among WGD paralogs (Guan *et al*, 2007; Hakes *et al*, 2007). However, when ribosomal duplicates are removed from consideration, the difference between WGD and SSD is no longer significant (See Supplementary Note 1 for more information on ribosomal duplicates).

Genetic redundancy between duplicates causes disparate interaction profiles

Our hypothesis about duplicate gene buffering suggests that duplicate genes will show fewer genetic interactions with other genes, because they functionally buffer one another (Figure 1). Indeed, we found that duplicate genes, on average, exhibit 34 interactions compared with 55 interactions observed for singletons when assayed against a set of ~1700 functionally diverse query mutant strains ($P<6\times10^{-16}$; Figure 2C). Notably, the decrease in negative genetic interactions is more apparent on gene families consisting of more than two members. Only 5% (29/554; $P<1\times10^{-27}$; see Materials and methods) of duplicates belonging to large gene families exhibit negative genetic interactions with each other, illustrating the impact of higher-order buffering and/or condition specificity among repeatedly duplicated genes. To control for the tendency of certain classes of genes toward duplication (Marland *et al*, 2004; He and Zhang, 2006), we examined the number of genetic interactions (union) across a range of double-mutant fitness values, and confirmed that the deficit in genetic interactions is not due to a bias in duplicates toward gene pairs that are not important

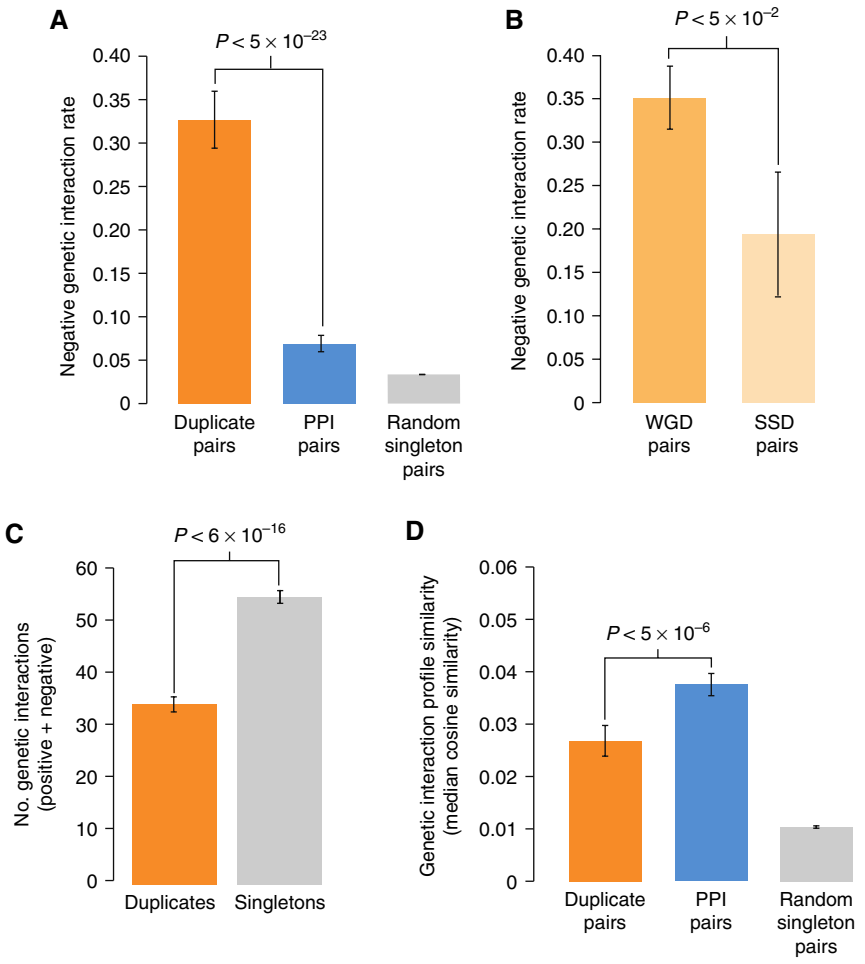

Figure 2 The distribution of genetic interactions supports the duplicate buffering hypothesis. (**A**) The proportion of negative interactions among screened pairs for duplicate pairs, singleton pairs with a protein–protein interaction (Materials and methods) and random singleton pairs. Error bars represent the error on a binomial proportion ($P < 5 \times 10^{-23}$; Binomial proportion test). (**B**) The proportion of negative interactions among duplicate pairs differs between modes of duplication. Whole-genome duplicates (WGD) exhibit a slightly higher rate of negative interaction than their small-scale duplication (SSD) counterparts ($P < 5 \times 10^{-2}$; Wilcoxon rank-sum). The rate of negative interactions within SSD pairs is still much higher than related singletons (Figure 2A), indicating that the functional overlap observed within duplicate pairs is not solely driven by WGD pairs. (**C**) The number of genetic interactions (both positive and negative) is plotted for all non-essential duplicates and singletons. Genes shown represent those found on the SGA deletion array and thus the counts represent the number of query genes with which a given array gene shows an interaction (see Materials and methods). Means are shown and error bars represent one standard deviation of the mean over 1000 bootstrapped samples of the distribution. ($P < 6 \times 10^{-16}$; Wilcoxon rank-sum), (**D**) Although duplicate genes show far greater profile similarity than random pairs, they show significantly less similarity than physically interacting pairs ($P < 5 \times 10^{-6}$; Wilcoxon rank-sum). Median cosine similarity is shown (Materials and methods). Error bars represent the standard deviation of the median over 1000 bootstrapped samples.

under the experimental conditions studied (Supplementary Figure 2).

In addition to fewer genetic interactions, our hypothesis suggests that sister duplicates should not share many interactions in common despite common function (Figure 1C). Indeed, we found that sister duplicates share an average of 1.2 negative genetic interaction partners, whereas genes encoding physically interacting proteins (a proxy for functionally related genes) share an average of 7.2 negative interactions (see Materials and methods). This trend extends beyond the counting of discrete interactions to more continuous measures of genetic interaction profile similarity. Duplicate pairs exhibit lower interaction profile similarity than functionally related gene pairs or genes encoding physically interacting proteins ($P < 5 \times 10^{-6}$; Figure 2D; Materials and methods; Supplementary Table 2). The lack of genetic interaction profile similarity among a number of partially redundant duplicate pairs was previously observed in Ihmels *et al* (2007), in which the authors attribute the phenomenon to incomplete buffering, that is, divergence. Differing genetic interactions certainly convey differentiation of function; however, our updated model (Figure 1) allows us to additionally explain how profile dissimilarity can also be a consequence of retained functional overlap. Thus, genetic interaction profiles for duplicate pairs are dissimilar, both for reasons of functional redundancy and divergence.

Dosage duplicates are exceptions to the buffering model

Assuming duplicate redundancy, our hypothesis about duplicate gene buffering suggests that only genetic interactions

resulting from functional divergence will be observable. However, this reasoning should not apply to an important class of duplicate genes, namely, those selected for increased protein product (Ohno, 1970; Ihmels *et al*, 2007). For example, Ihmels *et al*, noted that duplicates expressed in high abundance have retained very similar expression profiles, indicating the cell's need for both copies simultaneously. In general, if the cell benefits from higher gene dosage immediately on duplication, then the overlapping function of the duplicate copies is not truly redundant and should induce interactions in both sisters' profiles. Indeed Ihmels *et al* (2007), noted several examples of high-abundance duplicates with significantly correlated genetic interaction profiles. Thus, dosage duplicates appear to behave differently in the genetic interaction network than duplicates retained because of functional divergence.

To determine whether genetic interaction profiles could generally distinguish duplicates under dosage selection, we first compiled a set of likely dosage-related duplicates based on independent phylogenetic and genomic data (see Materials and methods). Using a combination of sequence and gene expression-related metrics, we defined a class of 80 putative 'dosage' duplicate pairs (Supplementary Table 1). Importantly, this class was enriched for known dosage-mediated paralogs (Kondrashov and Kondrashov, 2006; Conant and Wolfe, 2007; Ihmels *et al*, 2007). For example, 23 of the 80 pairs were ribosomal duplicates, which represents a significant enrichment ('Translation' GO term; $P<3\times10^{-5}$; hypergeometric cdf). Furthermore, deletion of one of the dosage paralogs resulted in a more severe fitness defect than other paralogs, suggesting that the dosage duplicates tend to lack the redundancy exhibited by other duplicates (Supplementary Figures 4, 5). The overall proportion of dosage pairs in our set is relatively low (~14%), but this is likely a conservative estimate for duplicates in general (Supplementary Figure 3).

Indeed, we found that dosage duplicates exhibit strikingly different characteristics in the genetic interaction network. Specifically, dosage duplicates show significantly greater genetic interaction profile similarity than other duplicates (Figure 3A). In fact, dosage duplicates are statistically indistinguishable from highly correlated singleton gene pairs that encode physically interacting proteins (Figure 3A; $P>0.4$; Wilcoxon rank-sum test; Materials and methods).

We speculated that the buffered interactions of non-dosage duplicates (for example, A′-Z and A″-Z in Figure 1C) could be present in the genetic interaction profiles of functionally related genes that lack a duplicated partner. To identify these functionally related 'proxy' genes, we focused on genes encoding proteins that exhibit physical interaction with both protein products of a duplicate gene pair (Figure 3B; Materials and methods). We reasoned that these proxy proteins may have physically interacted with the ancestor of the duplicates and, thus, have a genetic interaction profile resembling that of the ancestor gene. Subsequent to duplication, either these interactions were distributed uniquely between the modern copies (sub-functionalization) or new functions arose (neo-functionalization) as the pair diverged. Comparing the genetic interaction profiles of the duplicate genes with their corresponding proxy, we found that the large majority of divergent duplicate gene profiles are more similar to the proxy gene profile than to their corresponding sister's profile (Figure 3C). In contrast, dosage-mediated duplicates more often show higher profile similarity to each other than they do to the proxy gene (Figure 3C), suggesting that these genes tend not to buffer one another. Thus, genetic interaction profile similarity appears to be an effective way to distinguish dosage duplicates from duplicates undergoing functional divergence.

Duplicates exhibit asymmetric genetic interaction patterns

On the basis of the buffering model, genetic interaction profiles should reflect the unique roles of duplicate genes undergoing

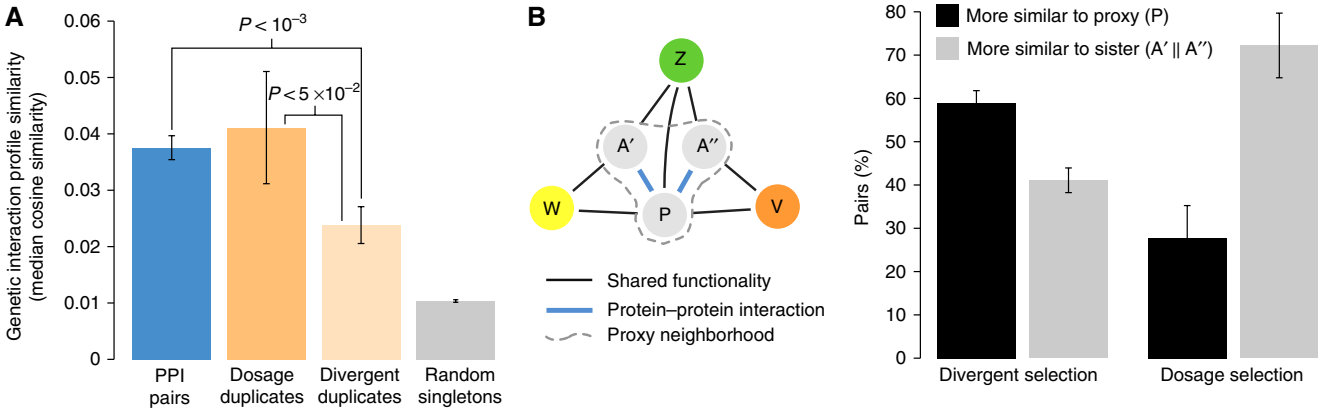

Figure 3 Global and local genetic interaction similarity comparisons support selection distinction. (**A**) Profile similarity shown as in Figure 2D. Duplicate pairs have been separated into dosage and non-dosage (divergent) classes (Materials and methods). Divergent duplicates show significantly less profile similarity than either dosage duplicates or singletons showing a physical interaction ($P<5\times10^{-2}$; $P<1\times10^{-3}$; Wilcoxon rank-sum test). Dosage duplicates are not statistically distinguishable from physically interacting singletons. (**B**) A hypothetical functional network is shown that contains a duplicate pair (A′/A″). A proxy gene (*P*) is identified by finding a protein that shares protein–protein interactions with both duplicates (see Materials and methods), and *P* is used to approximate the genetic interaction profile of the common ancestor (that is, A). The number of times a duplicate's similarity with its sister exceeded its similarity with *P* is shown as a percentage, and error bars represent error on a binomial proportion. Dosage and divergent pairs are counted separately. In terms of genetic interaction profiles, divergent pairs more closely resemble their common neighbor than they do each other. In contrast, dosage pairs more closely resemble each other. The probability that these two classes come from the same binomial distribution is small ($P<9\times10^{-5}$).

functional divergence. Ohno (1970) hypothesized that once a duplicate begins to accumulate mutations, the selection pressure will focus on the duplicate retaining the ancestral function and, therefore, most of the divergent changes should be confined to one copy. Although controversial (Wagner, 2002; Lynch and Katju, 2004; Fares *et al*, 2006; Byrne and Wolfe, 2007), evidence supporting such asymmetric divergence has been extracted from duplicate sequence data (Conant and Wagner, 2003; Zhang *et al*, 2003; Kellis *et al*, 2004; Scannell and Wolfe, 2008), PPIs (Wagner, 2002; He and Zhang, 2005) and expression patterns (Gu *et al*, 2002b; Tirosh and Barkai, 2007).

The distribution of genetic interactions within each duplicate pair strongly supports a model of asymmetric evolution. We examined the ratio of unique negative genetic interactions for each pair of duplicates (max:min, see Materials and methods) and found that the ratio exceeds 4:1 for >30% of gene pairs surveyed (109/351), and more than 17% (60/351) of duplicate pairs exhibit a ratio greater than 7:1 (Figure 4A). The observed interaction ratios are significantly greater than expected under a null model of symmetric interaction ($P<1\times10^{-100}$; Wilcoxon rank-sum test; see Materials and methods), suggesting that genetic interactions tend to appear preferentially in one member of each duplicate pair.

We suspected that the asymmetric distribution of genetic interactions could be partially explained by asymmetric rates of sequence evolution, which provide an independent measure of selection pressure. Previous work showed a correlation

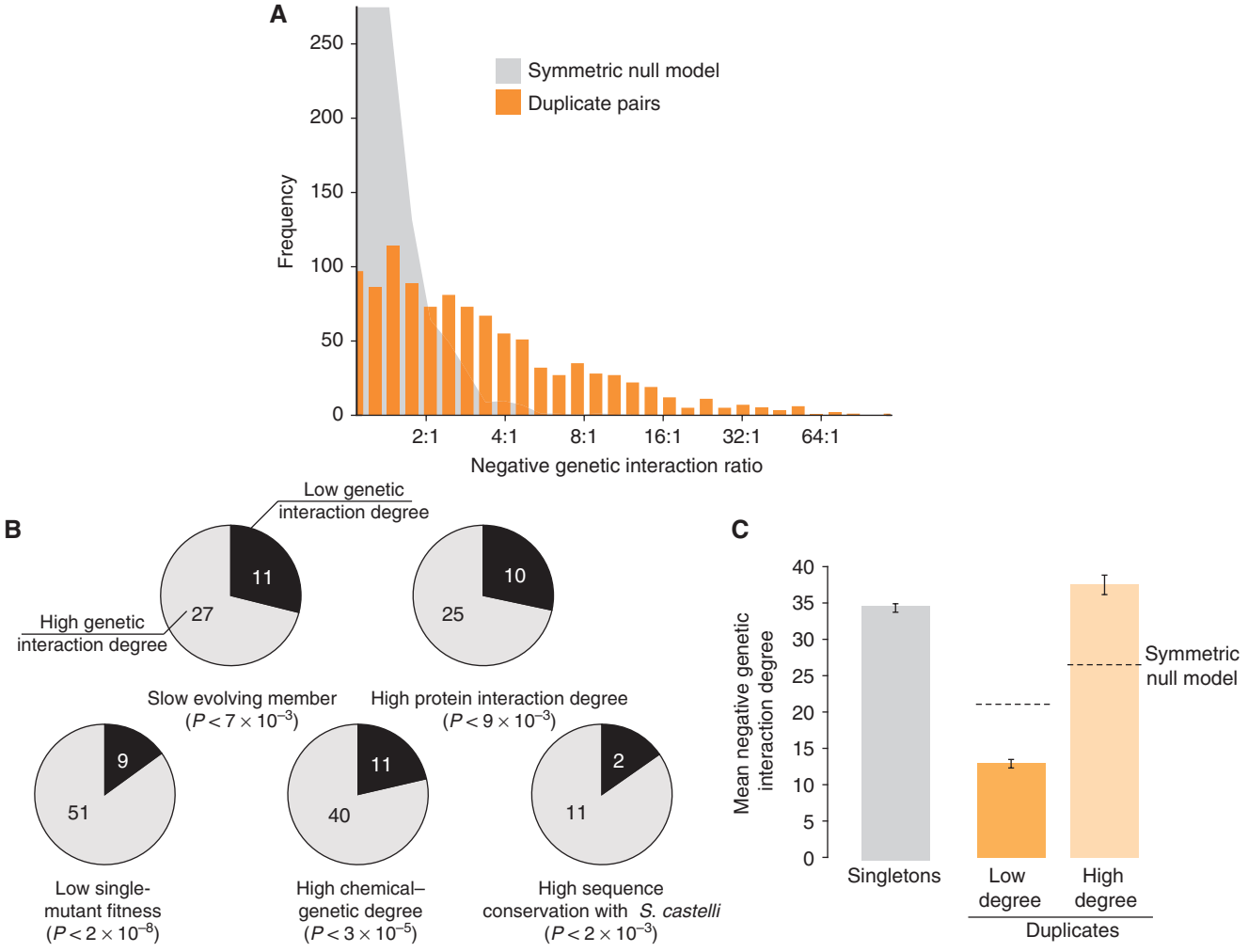

Figure 4 Genetic interactions provide evidence for asymmetric functional divergence. (**A**) A histogram of the duplicate interaction degree ratio. The ratio is defined for unique interactions with the higher degree in the numerator. Pairs included must have at least 10 total interactions between them, with each member having at least one interaction. Shown for comparison is another degree ratio histogram in which interactions for every duplicate pair are redistributed to either member with equal probability (symmetric null model). (**B**) Relating selection pressure measures on asymmetric duplicate pairs. Pairs with a unique interaction ratio exceeding 7:1 (60 pairs) are compared across several different sequence or functional genomic data sets. Each gene was put into the high or low interaction degree bin by comparison with its sister. Each pair was then examined for agreement in directionality with the indicated data set. For example, in 27 out of 38 pairs, the sister with higher genetic interaction degree also has a lower rate of sequence change. Comparisons with <60 pairs reflect missing pairs in the secondary data set. Also shown are *P*-values resulting from a binomial test in which genetic interaction degree is assumed independent of the other data type. (**C**) The number of negative genetic interactions for singletons and duplicates. Each duplicate pair was sorted by genetic interaction degree and means are shown. Dotted lines represent the same process applied to the simulated distribution from Figure 4A. The difference between high-degree duplicates and singletons is significant (34.9 versus 37.2; $P<5\times10^{-8}$; Wilcoxon rank-sum); however, the mean number of singleton interactions is reduced by a large portion of singletons with no measurable deletion effect, and the significant difference presented here subsides when controlling for gene importance (Supplementary Figure 7).

between protein dispensability and evolutionary rate among duplicate genes (Yang *et al*, 2003). A recent study of WGD pairs has also shown that both sisters undergo a period of accelerated change, but while one of them evolves much slower and is preferentially retained across different yeast species, the other evolves much faster and is preferentially lost (Byrne and Wolfe, 2007; Scannell and Wolfe, 2008). Interestingly, we found a related trend in which the rapidly evolving member had fewer genetic interactions than the more slowly evolving partner in 34/51 of previously defined asymmetric duplicate pairs (Kellis *et al*, 2004; $P<0.02$; binomial). The bias was more pronounced for pairs whose unique genetic interaction degree ratio exceeded 7:1. In this case, the rapidly evolving member was associated with a lower interaction degree for 27/38 pairs belonging to this group (Figure 4B; $P<7\times10^{-3}$). Furthermore, there was a significant correlation between the disparity in sequence evolution rates and the asymmetry of interaction degree (r=0.318, $P<0.03$), suggesting that the magnitude of asymmetry in genetic interaction degree was predictive of asymmetry in selection pressure acting on duplicate gene sequences. Interestingly, the set of duplicates with asymmetric evolution rates is significantly depleted for dosage-mediated pairs ($P<2\times10^{-3}$; hypergeometric cdf; Supplementary Note 2).

In searching for physiological evidence to corroborate the marked asymmetry in interaction degree, we examined PPIs involving gene pairs with the most extreme ratio of genetic interactions (7:1). Of these, 35 pairs exhibit at least one PPI for each member, and for 25/35 (71%) of these pairs, the partner with more genetic interactions also tended to have retained or gained more physical interactions ($P<9\times10^{-3}$; binomial; Figure 4B). Genetic interaction degree asymmetry as a measure of selection pressure is also predictive of measurements of single-mutant fitness, wherein we observed that the partner with more genetic interactions has a larger impact on fitness when deleted ($P<2\times10^{-8}$; binomial; Figure 4B). We observed a similar trend with the number of chemical environments in which each duplicate sister displays a phenotype (Hillenmeyer *et al*, 2008), wherein the duplicate sister with the higher genetic interaction degree generally had a higher chemical-genetic degree ($P<3\times10^{-5}$; binomial; Figure 4B; see Materials and methods). Interestingly, these trends between duplicate sisters mirror similar trends related to genetic interaction degree across the whole genome (Costanzo *et al*, 2010; Lehner, 2010).

We also found that WGD sisters with more genetic interactions tend to have higher sequence similarity to the remaining member of the pair in other WGD species (*S. castellii*, $P<2\times10^{-3}$; *Candida glabrata*, $P<1\times10^{-2}$; binomial; Figure 4B; see Materials and methods). Specifically, in 11 of 13 instances in *S. castellii* and in 12 of 16 such cases in *C. glabrata*, the higher degree sister showed higher sequence identity to the single remaining WGD sister. Additionally, the duplicate sister with more genetic interactions tended to have a greater mRNA expression level (Holstege, 1998) for 32 out of the 51 pairs (63%; $P<0.046$; binomial), although this difference was not significant in an independent expression level study (Nagalakshmi *et al*, 2008). Interestingly, we found that the rate of negative interactions between sisters in the asymmetric set was 46%, which is no less than the background rate for duplicates (Supplementary Figure 6), indicating retained functional overlap for even these highly skewed pairs.

The asymmetric distribution of genetic interactions among duplicate pairs motivated us to question whether the overall deficit of genetic interactions among duplicate genes is a result of buffered interactions distributed in both duplicate copies evenly or rather in only one paralog. Strikingly, we found that, on average, one of the two duplicates had a comparable or larger number of interactions than singletons while the sister has significantly fewer interactions (Figure 4C). The slightly higher number of interactions for the high-degree duplicate gene appears to be a result of an important bias among the ancestors of the duplicates, as they became statistically indistinguishable from singleton genes after controlling for gene importance (Supplementary Figure 7). Thus, the overall deficiency of duplicate genes for genetic interactions (Figure 2C) as well as the asymmetric distribution of modern interactions (Figure 4A) suggests that the majority of the interactions of the common ancestor are associated with a single member of the pair.

Dissecting the divergent functions of duplicates through genetic interaction profiles

Genes belonging to the same biological pathway or protein complex tend to share similar patterns of genetic interactions, and similarity between genetic interaction profiles has proven effective for predicting gene function and defining pathway and complex membership (Costanzo *et al*, 2010). In this study, we exploited genome-wide genetic interaction profiles along with specific interactions to identify the functional differences that distinguish divergent gene pairs. For example, *SSO1* and *SSO2* encode SNARE proteins, core components critical for the specificity of membrane fusion and intracellular transport in eukaryotic cells (Jahn and Scheller, 2006; Yang *et al*, 2008). Although vesicle fusion with the plasma membrane is dependent on either *SSO1* or *SSO2* gene function, previous studies have shown an *SSO1*-specific requirement for prospore membrane formation during sporulation (Jantti *et al*, 2002; Yang *et al*, 2008). We noticed that genes involved in chitin biosynthesis (*CHS3*, *CHS5* and *SKT5*) and polarized cell growth (*BUD6*, *BEM3* and *AXL2*) shared genetic interactions in common with *SSO1* ($r>0.14$; Supplementary Table 4; see Materials and methods) but not with *SSO2* ($r<0.04$), suggesting a specific role for *SSO1* in these processes during vegetative growth. These genetic interaction profile similarities support previous observations from high-content screening experiments, indicating that *SSO1* is important for normal actin localization, and deletion of *SSO1* results in more severe actin mis-localization (21%) compared with a *sso2*Δ mutant strain (Ohya *et al*, 2005; 4%; Supplementary Figure 8).

We found that *SSO1* and *SSO2* also varied extensively in terms of their interaction degree. In fact, the ratio of *SSO1*:*SSO2* interactions was among the most asymmetric, with 149 negative interactions for *SSO2* compared with only 15 negative interactions involving *SSO1* (Supplementary Table 3). Consistent with evolution of a condition-specialized function, previous studies suggest that functional divergence has led to a

more prominent sporulation-specific function for *SSO1* (Jantti *et al*, 2002; Yang *et al*, 2008). The reduced number of interactions observed for *SSO1* may reflect its specialized function, in part, because genetic interactions were mapped under vegetative conditions when sporulation is not required. In a similar example, highly asymmetric genetic interaction degree may reflect sporulation or meiosis-specialized function for cell wall assembly duplicates *GAS1* and *GAS2*, suggesting that this may be a common basis for imbalances in genetic interaction degree (Supplementary Note 3).

Genetic interaction profile examination yielded another interesting example in duplicate pair *CIK1*/*VIK1*. Comparison of profile similarity and interaction degree of *CIK1* and *VIK1* demonstrates the ability of genetic interaction analysis to distinguish subtle functional differences between paralogous genes. *CIK1* and *VIK1*, which arose from the WGD event, encode kinesin-associated proteins that form separate heterodimeric complexes with Kar3, a minus-end-directed microtubule motor protein, to mediate a diverse set of microtubule-dependent processes (Manning *et al*, 1999). Despite strong sequence and structural similarities, *CIK1* and *VIK1* exhibit different genetic interaction profiles, suggesting that these proteins have specialized functional roles. Although both proteins depend on physical interaction with Kar3 for proper function, *CIK1* has more genetic interactions in common and is more closely correlated to the *KAR3* interaction profile (*CIK1–KAR3*; r=0.5; see Materials and methods) compared with its duplicate *VIK1* (*VIK1–KAR3;* r=0.3). Consistent with closely related interaction profiles (Figure 5A), *kar3*Δ and *cik1*Δ deletion mutants share several phenotypes including abnormally short spindles, chromosome loss and delayed cell cycle progression (Page *et al*, 1994; Manning *et al*, 1999). In contrast, a *vik1*Δ mutant strain does not exhibit any overt phenotype (Manning *et al*, 1999).

In addition, *VIK1* and *CIK1* differ in their gene expression and protein localization (Manning *et al*, 1999). Interestingly, we found that *CIK1* and *KAR3* interaction profiles more closely resemble the profiles of genes involved in chromosome cohesion and segregation (GO:0000070; $P<8\times10^{-8}$; hypergeometric cdf; Figure 5A), whereas *VIK1* was more correlated

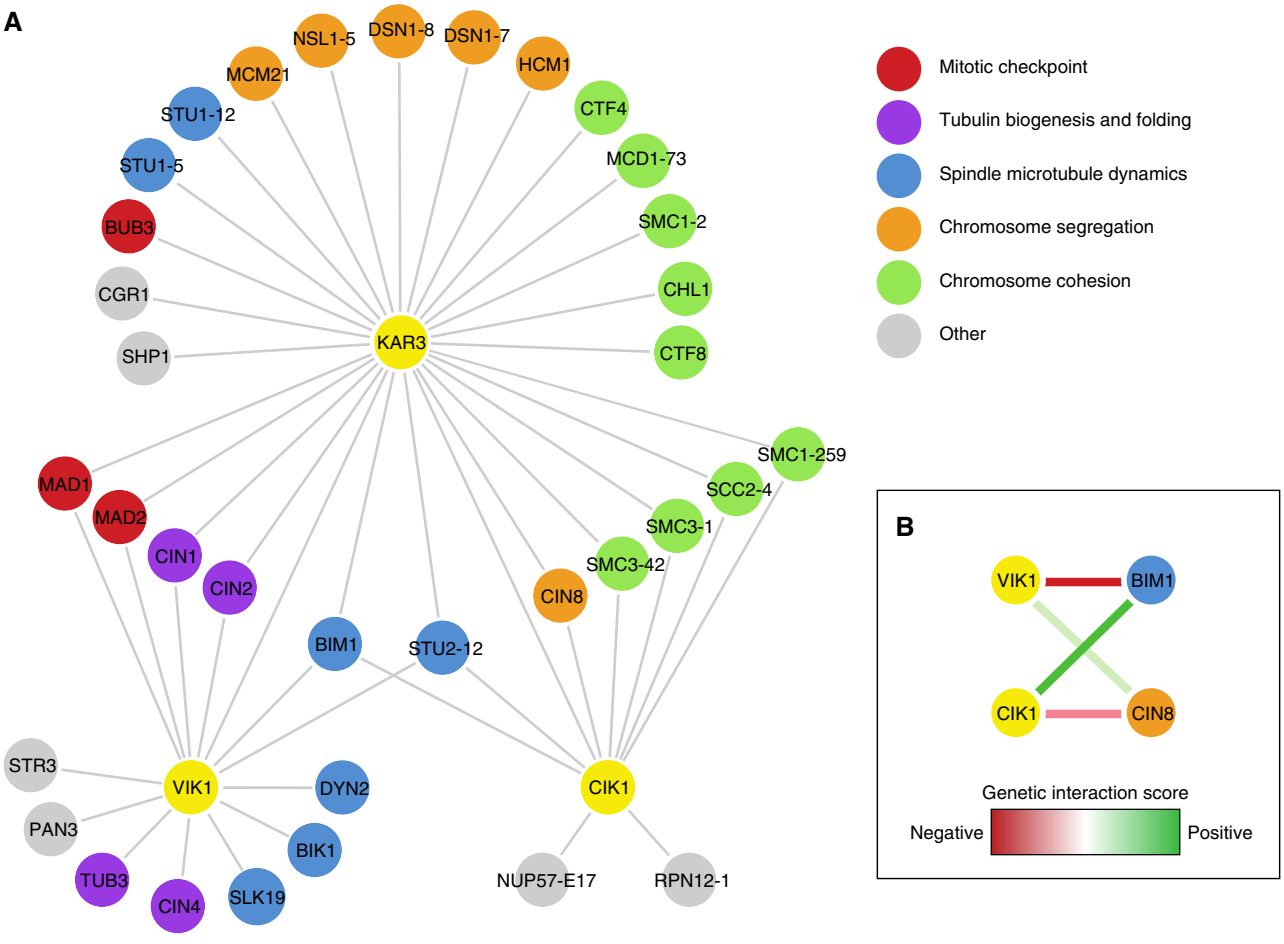

Figure 5 Functional analysis of duplicate pair *CIK1–VIK1* (**A**) Genetic interaction profile similarity. Similarity scores were taken from Costanzo *et al* (2010) and represent a combination of array side and query side correlations (Materials and methods). Nodes shown include all first neighbors of the three primary genes of interest (*CIK1, VIK1* and *KAR3*). A threshold of 0.2 was used as in Costanzo *et al* (2010) and edges between first neighbors of genes of interest have been removed for clarity. (**B**) Genetic interactions. SGA genetic interaction scores from Costanzo *et al* (2010) highlight differences between *CIK1* and *VIK1*. Green lines represent positive interactions, whereas red lines represent negative interactions. The opacity of the line is proportional to the strength of the interaction.

to genes involved in microtubule assembly and stabilization (GO:0007017; $P<2\times10^{-8}$; Figure 5A). Our findings support a previous hypothesis (Manning *et al*, 1999) and suggest that the Cik1–Kar3 and Vik1–Kar3 heterodimers serve distinct, yet related, roles during cell division. In addition to profile similarity, examination of individual genetic interactions also highlight potential functional differences between these microtubule motor-associated proteins. We noticed strong asymmetry in the ratio of *CIK1:VIK1* interaction degree and, consistent with a more severe deletion phenotype, we found that *CIK1* has 4.5-fold more negative genetic interactions than *VIK1* (Supplementary Table 3). Interestingly, several genetic interactions connecting *VIK1* and *CIK1* to common partners differ in their type. In particular, the plus-end microtubule motor-encoding gene, *CIN8*, shares a modest positive genetic interaction with *VIK1*, whereas a *cik1*Δ-*cin8*Δ double mutant displayed a synthetic sick/lethal phenotype (Figure 5B). Findings derived from our large-scale survey of genetic interactions support previous observations that disruption of *VIK1*, but not *CIK1*, partially suppresses the temperature-sensitive growth defect of a *cin8-3 kip1*Δ double mutant (Manning *et al*, 1999). One role for the Kar3 microtubule motor during vegetative growth is thought to involve opposing the action of the Cin8 and Kip1 motor proteins. The *VIK1*-specific positive genetic interactions reported here and elsewhere (Manning *et al*, 1999) suggest that a *CIN8* and *KIP1* antagonistic function may be unique to the Vik1–Kar3 heterodimer, thus distinguishing between Vik1–Kar3- and Cik1–Kar3-related functions. In another example, we found that *BIM1* shared a positive interaction with *CIK1* (*bim1*Δ suppressed the *cik1*Δ growth defect) and a negative interaction with *VIK1* (Figure 5B). Bim1 is a microtubule-binding protein that localizes to the plus end of the microtubules where it is required for proper positioning of the nucleus during nuclear migration (Tirnauer *et al*, 1999; Lee *et al*, 2000). Recent studies have shown that Bim1 also localizes to the spindle midzone to stabilize microtubules during anaphase (Gardner *et al*, 2008). Interestingly, Kar3 also exhibits different sub-cellular localization patterns that are dependent on physical interaction with Vik1 or Cik1. During vegetative growth, Kar3 associates with the spindle midzone in a Cik1-dependent manner (Sproul *et al*, 2005), whereas the Kar3–Vik1 heterodimer localizes to the spindle poles (Manning *et al*, 1999; Allingham *et al*, 2007). Although the nature of the genetic interactions is unclear, the negative interaction between *BIM1* and *VIK1* might reflect the failure in nuclear positioning due to unstable microtubules while positive interaction observed between *BIM1–CIK1* might reflect opposing functions involved in stabilizing and destabilizing the microtubules (Sproul *et al*, 2005; Gardner *et al*, 2008).

In both pairs of duplicates we investigated in detail (*SSO1–SSO2* and *CIK1–VIK1*), the duplicate genes exhibited a strong negative interaction between sisters. This suggests that despite evidence for functional specialization and dramatic asymmetry in their overall interaction degree, sister duplicates retain the ability to partially compensate for the loss of one another, and this trend appears to be relatively common across duplicates in yeast (Supplementary Figure 6). We also noted that, although genetic interactions can resolve functional differences between sisters, in these cases, the differences appear to be relatively subtle: context or conditional specialization in the case of *SSO1–SSO2* and localization specialization in the case of *CIK1–VIK1*.

Discussion

We examined how partial redundancy and the functional divergence of duplicate gene pairs relates to their genetic interaction profiles. We found evidence for the hypothesis that immediately after duplication, duplicated gene pairs will mask each other's interactions with other genes, and that as the pair evolves apart, interactions reappear, highlighting functional differences between them. We have also shown that genome-wide genetic interaction profiles provide insight into the mechanisms of duplicate gene evolution by distinguishing duplicate pairs maintained for gene dosage effects from those retained because of functional divergence. These findings clarify previous observations about the surprising prevalence of genetic interactions for apparently redundant duplicate genes (Ihmels *et al*, 2007), and provide evidence that they do indeed reflect functional redundancy as well as functional divergence. Finally, we also showed that a disproportionate distribution of genetic interactions among gene pairs supports the asymmetric evolution of duplicate genes whereby one member of a duplicate pair is under stronger selective pressure. The skewed distribution is correlated with differences in rates of sequence evolution, PPI degree, single-mutant fitness defects and sensitivity to a variety of chemical environments, suggesting that one member of the gene pair assumes a predominant role under standard vegetative growth conditions.

Previous studies suggest that the asymmetric accumulation of loss-of-function mutations in many duplicate pairs is established quickly based on sequence evidence from the WGD event that indicates that the identity of the quickly evolving sister is consistent across several yeast species (Fares *et al*, 2006; Byrne and Wolfe, 2007; Scannell and Wolfe, 2008). On the basis of these observations combined with results from this study, we propose a refined model of duplicate evolution (Figure 6). Following a duplication event that does not provide a dosage-dependent fitness advantage, we argue that one member of a duplicate pair should accumulate loss-of-function mutations more quickly due to relaxed purifying selection alone (Supplementary Note 4; Supplementary Figures 9–11). In essence, a degenerate paralog is more accommodating of mutations and stands a higher chance of sustaining a mutation affecting any remaining redundant functions (Supplementary Note 4). In many cases, the fast evolving duplicate meets the common fate of non-functionality and eventual gene loss. If early function loss is complementary, the pair is put on a path toward functional partition. Gene properties that are necessary for multiple functions may be preserved in both copies if previous mutations caused these functions to fall to different sisters. Such an arrangement would render a complete functional divergence impossible. We note that this natural progression of asymmetry should occur for any duplication event, either whole-genome or small-scale, although the means of preservation of a duplicate pair might be distinct depending on the context. Presumably, in some

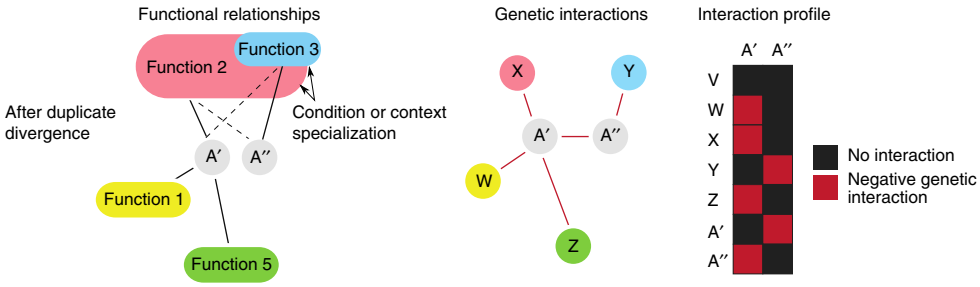

Figure 6 Updated model of asymmetric duplicate genetic interaction evolution. Asymmetry is rapidly established through the absence of purifying selection on a duplicate pair, but in rare cases, the quickly evolving duplicate confers a fitness advantage through functional or context specialization (Function 3). Subsequent selection on Function 3, however, also maintains a limited capacity of duplicate A″ to carry out Function 2 (dotted lines). In this scenario, there is overlap in function, but the efficacy of the duplicate pair with respect to a particular function differs, and so the buffering is asymmetric. Fewer genetic interactions are observed for A″ either because of its less constrained function or because of its role in other environmental or developmental contexts.

cases, sister duplicates simply maintain complementary but essential roles despite their asymmetry, whereas in other cases, the asymmetric configuration provides some fitness advantage that ultimately enables a selective sweep.

We cannot rule out the possibility that neo-functionalization may have a role in the preservation of some duplicate pairs and their subsequent asymmetric evolution, but if that is the case, the quickly evolving duplicate appears to take on a more inconspicuous functional role in most pairs. Our data argues against dramatic neo-functionalization and instead suggests that the rapidly evolving duplicate retains a subset of the ancestral function for which it has become optimized (Figure 6). Importantly, despite specialization, the high rate of negative genetic interactions observed between asymmetric duplicate pairs (Supplementary Figure 6) indicates that the lower degree sister often retains some ability to compensate for the loss of the more constrained sister. We do not interpret this as evidence for selection on their redundancy, rather that the function or context for which the quickly evolving duplicate has been specialized allows or requires it to at least partially maintain the ancestral role (Supplementary Figure 9).

Our observations are consistent with previously proposed models of sub-functionalization, including the Duplication–Degeneration–Complementation and Escape from Adaptive Conflict models (Hughes, 1994; Des Marais and Rausher, 2008; Innan and Kondrashov, 2010). Both these schemes describe ancestral functions being split between duplicates, the latter allowing for optimizations previously constrained by other functions. Indeed, we identified several gene pairs in the yeast genetic interaction network that support specialization driven by adaptation to different environmental or developmental conditions, leading us to speculate that a special case of the Escape from Adaptive Conflict or Duplication–Degeneration–Complementation models may apply to a large fraction of duplicates in *S. cerevisiae*, in which this specialization is driven by adaptation to different environmental or developmental conditions. For example, several of the most asymmetric pairs involve a gene specialized for sporulation or meiosis. Sporulation requires formation of a membrane structure known as the prospore membrane, which is dependent on the Sso1–Spo20 t-SNARE complex. Although *in vitro* experiments indicate that both Sso1 and Sso2 can bind to Spo20 to form a functional t-SNARE, the Sso2–Spo20 complex exhibits much weaker membrane-fusion capacity and, thus, may explain why only Sso1 is able to support sporulation (Liu *et al*, 2007). Furthermore, studies have shown that Sso1 can interact with phosphatidic acid, which is necessary for Spo20 localization and function (Liu *et al*, 2007). Although the exact cause of functional divergence remains unclear, it is possible that the *SSO1* gene product acquired a specialized role after duplication, which is important for modulating *SPO21* function in non-dividing cells. This example supports our model illustrating that changes in protein function are often relatively subtle, and condition or developmental specialization may instead be the driving force behind duplicate gene retention.

Although genome sequences provide a wealth of information about gene ancestry, they fail to address the functional efficacy of genes on which selection ultimately acts. Network analysis of PPIs (Presser *et al*, 2008) provide a complementary view, but common physical interactions shared by a duplicate pair still do not reveal whether interaction with a specific member of a duplicate pair has a functional consequence to the cell under a given experimental condition. Genetic interactions address both of these shortcomings by revealing exactly which relationships have an impact on fitness, and which do not, and thus provide a powerful perspective for understanding duplicate gene evolution.

Materials and methods

Definition of duplicates and singletons

The full list of duplicate pairs consists of those identified as the result of the WGD event, as reconciled from several sources (Byrne and Wolfe, 2005). Additionally, any pair of genes fulfilling established similarity requirements (Gu *et al*, 2002a) was reasoned to be a duplicate pair resulting from a SSD event. Specifically, the gene pair must have a sufficient sequence similarity score (FASTA Blast, E=10) and sufficient protein alignment length(>80% of the longer protein). The pair must also have an amino-acid level identity of at least 30% for proteins with aligned regions longer than 150 amino acid, and for shorter proteins, the identity must exceed $0.01n + 4.8\,L^{(-0.32(1+\exp(-L/1000)))}$, where L is the aligned length and n=6 (Rost, 1999; Gu *et al*, 2002a).

After combining pairs from the WGD event, with pairs determined through sequence alone (SSD), families with more than two members as a result of multiple pairings were completely removed from analysis to control for potential buffering from a third member affecting the interactions of the first two, and any gene not involved in any pairings was deemed an unambiguous singleton.

Functionally related pairs

As a proxy for non-duplicated yet functionally related gene pairs, we have used pairs that exhibited a PPI in at least one of two high-throughput TAP-MS studies (Gavin *et al*, 2006; Krogan *et al*, 2006). To increase the number of duplicate pairs considered in the analysis relating sister–sister profile similarity to sister–proxy similarity, we did not limit PPI interactions to TAP-MS (see next section). Interactions for this analysis were included from BioGrid if they fell into one of the following categories: affinity capture-RNA, affinity capture-Western, two-hybrid, PCA, affinity capture-MS, co-fractionation, biochemical activity, co-crystal structure, co-purification, far western, FRET, protein–peptide, protein–RNA or reconstituted complex.

Significance of binomial proportions (synthetic sick/lethal or dosage membership)

Synthetic sick/lethal proportion rates were tested under using the following normally distributed random variable:

$$Z_0 = \frac{P_1 - P_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

where P_1 and P_2 are the binomial proportions in the respective classes and $\hat{P}$ is the binomial proportion of the combined set.

Genetic interaction data and profile similarity calculations

Genetic interaction data were taken from a recent global genetic interaction study (Costanzo *et al*, 2010). For the presence or absence of individual interactions, such as calculating the proportion of synthetic lethal duplicates, or counting interaction degree for a given gene magnitude, *P*-value thresholds were used ($e > 0.08$ and $P < 0.05$). When counting discrete interactions, column degree was used. Thus, only genes in the deletion array (3885 genes) have valid degrees. This dimension was chosen to maximize the number of covered genes, as fewer genes (1712) have been screened as queries. For assessing profile similarity, we first normalized the (unthresholded) data along both rows and columns and then used inner product between any pair of array genes as their profile similarity (Rost, 1999; Gu *et al*, 2002a).

Definition of dosage class

A duplicate pair was labeled as a 'dosage' pair if it met two of the following three conditions: (1) The pair's representative ortho-group had a volatility score (Wapinski *et al*, 2007) in the top quartile. (2) The pair had a scaled difference in transcript quantity in the bottom quartile. Absolute expression data is taken from Holstege (1998) and scaled expression difference is defined as in Ihmels *et al* (2007):

$$\text{Scaled difference}\,(a,\,b) = \frac{|a-b|}{a+b}$$

(3) The pair had a scaled difference in expression stability in the bottom quartile, wherein stability for each gene is defined as the number of data sets (out of a possible 127 from Hibbs *et al*, 2007) in which the expression of the given gene is in the bottom 2% for variance.

Ancestral proxies on the PPI network

To find suitable proxy genes for a given duplicate pair, we isolated the common interaction partners on the expanded physical PPI network for each pair with the assumption that interactions common to both paralogs are not likely to have evolved independently, and are therefore tied to one or more of the pair's ancestral functions. We then measured genetic interaction profile similarity between each paralog and the neighbor for comparison with profile similarity between the duplicates themselves. Results were averaged across all common partners for a given duplicate pair.

Genetic interaction degree asymmetry

To compare genetic interaction degree and rates of evolution, we used the original rates provided in the supplement to Kellis *et al*, 2004. This ratio was defined as the rate of the quickly evolving or 'derived function' member divided by that of the slowly evolving or 'ancestral function' member. To test for bias in which member of the pair had more interactions, we assumed a null model in which either gene was equally probable to have the most interactions. We obtained a *P*-value for this hypothesis using MATLAB's binomial cumulative distribution function binocdf(). The proposed ancestral gene generally has a higher degree; hence, the genetic interaction ratio for the pair was calculated with the 'ancestral function' member's property in the numerator.

Chemical–genetic degree

To ascertain the number of chemical environments under which a gene displayed a significant phenotype, we used the original data from Hillenmeyer *et al* (2008). We counted the number of conditions in which the homozygous deletion displayed a significant *P*-value ($P < 0.05$) out of a possible 1144. As above, we then used a binomial cumulative distribution to test whether the correspondence between the two data sets (the number of times the gene with more genetic interactions also had more chemical interactions) could be attributed to chance.

Phylogenetic comparison for asymmetric pairs

We compared the sequence similarity of the WGD pairs in *S. cerevisiae* with orthologs in other post-WGD species (*S. castellii*, *C. glabrata and S. bayanus*) in which one WGD copy had been lost as annotated in the Yeast Genome Order Browser (Byrne and Wolfe, 2005). For each such case, we produced an amino-acid sequence alignment between each *S. cerevisiae* gene and the out-group ortholog using the BLAST algorithm (Johnson *et al*, 2008). We then compared the percent identity score for each duplicate with the out-group ortholog. For every pair identified as asymmetric, we used a binomial test to ascertain whether the gene with more interactions was more similar to the orthologous gene, the null hypothesis being that the lower degree and higher degree genes have equal chance of a higher percent identity score with the orthologous gene. In *S. bayanus*, we found only three single orthologs to asymmetric WGD pairs in *S. cerevisiae*, and as such that data is not included.

Biological example profile similarity

Profile correlations for specific biological examples (*SSO1*, *SSO2*; *GAS1*, *GAS2*; *CIK1*, *VIK1*) were taken from the supplement to Costanzo *et al* (2010). It represents a composite score using information from both array and query profiles in an attempt to give a uniform similarity score across all pairs of genes. Figure 5A shows edges from this composite network involving *CIK1*, *VIK1* and *KAR3* using a correlation threshold of 0.2.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We would like to thank Dr Tamar Lahav and Dr Judith Berman for their valuable insights and helpful comments regarding this work and also Dr Nathan Springer for his helpful comments on the manuscript. BV, JB and CLM are partially supported by funding from the University of Minnesota Biomedical Informatics and Computational Biology program, and a seed grant from the Minnesota Supercomputing Institute. CLM and JB are supported by the National Institutes of Health (1R01HG005084-01A1) and CLM and BV are supported by the National

Science Foundation (DBI 0953881). BP is supported by The International Human Frontier Science Program Organization, by the Hungarian Scientific Research Fund (OTKA) and by the 'Lendület Program' of the Hungarian Academy of Sciences. CB and BA are supported by Genome Canada through the Ontario Genomics Institute (2004-OGI-3-01), the Canadian Institutes of Health Research (GSP-41567) and the Canadian Institute for Advanced Research.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Allingham JS, Sproul LR, Rayment I, Gilbert SP (2007) Vik1 modulates microtubule-Kar3 interactions through a motor domain that lacks an active site. *Cell* **128:** 1161–1172

Baudot A, Jacq B, Brun C (2004) A scale of functional divergence for yeast duplicated genes revealed from analysis of the protein-protein interaction network. *Genome Biol* **5:** R76

Brookfield J (1992) Can genes be truly redundant? *Curr Biol* **2:** 553–554

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

Byrne KP, Wolfe KH (2007) Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. *Genetics* **175:** 1341–1350

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

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

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

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St. Onge RP, Vander Sluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M *et al* (2010) The genetic landscape of a cell. *Science* **327:** 425–431

Dean EJ, Davis JC, Davis RW, Petrov DA, McVean G (2008) Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet* **4:** e1000113

DeLuna A, Vetsigian K, Shoresh N, Hegreness M, Colon-Gonzalez M, Chao S, Kishony R (2008) Exposing the fitness contribution of duplicated genes. *Nat Genet* **40:** 676–681

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

Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C (2009) Systematic mapping of genetic interaction networks. *Ann Rev Genet* **43:** 601–625

Dobzhansky T (1946) Genetics of natural populations. XIII. Recombination and variability in populations of *Drosophila pseudoobscura*. *Genetics* **31:** 269

Fares MA, Byrne KP, Wolfe KH (2006) Rate asymmetry after genome duplication causes substantial long-branch attraction artifacts in the phylogeny of Saccharomyces species. *Mol Biol Evol* **23:** 245–253

Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151:** 1531–1545

Gardner MK, Haase J, Mythreye K, Molk JN, Anderson M, Joglekar AP, O'Toole ET, Winey M, Salmon E, Odde DJ, Bloom K (2008) The microtubule-based motor Kar3 and plus end binding protein Bim1 provide structural support for the anaphase spindle. *J Cell Biol* **180:** 91–100

Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier M-A, Hoffman V, Hoefert C, Klein K, Hudak M, Michon A-M, Schelder M, Schirle M, Remor M *et al* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440:** 631–636

Guan Y, Dunham MJ, Troyanskaya OG (2007) Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* **175:** 933–943

Gu Z, Cavalcanti A, Chen F, Bouman P, Li W (2002a) Extent of gene duplication in the genomes of drosophila, nematode, and yeast. *Mol Biol Evol* **19:** 256–262

Gu Z, Nicolae D, Lu HH, Li WH (2002b) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet* **18:** 609–613

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li W (2003) Role of duplicate genes in genetic robustness against null mutations. *Nature* **421:** 63–66

Hakes L, Pinney J, Lovell S, Oliver S, Robertson D (2007) All duplicates are not equal: the difference between small-scale and genome duplication. *Genome Biol* **8:** R209

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

He X, Zhang J (2006) Higher duplicability of less important genes in yeast genomes. *Mol Biol Evol* **23:** 144–151

Hibbs MA, Hess DC, Myers CL, Huttenhower C, Li K, Troyanskaya OG (2007) Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics* **23:** 2692–2699

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320:** 362–365

Holstege F (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95:** 717–728

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* **256:** 119–124

Ihmels J, Collins SR, Schuldiner M, Krogan NJ, Weissman JS (2007) Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Mol Syst Biol* **3:** 86

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

Jahn R, Scheller RH (2006) SNAREs [mdash] engines for membrane fusion. *Nat Rev Mol Cell Biol* **7:** 631–643

Jantti J, Aalto MK, Oyen M, Sundqvist L, Keranen S, Ronne H (2002) Characterization of temperature-sensitive mutations in the yeast syntaxin 1 homologues Sso1p and Sso2p, and evidence of a distinct function for Sso1p in sporulation. *J Cell Sci* **115:** 409–420

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden T (2008) NCBI BLAST: a better web interface. *Nucl Acids Res* **36:** W5–W9

Kafri R, Levy M, Pilpel Y (2006) The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc Natl Acad Sci USA* **103:** 11653–11658

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

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

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrín-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B *et al* (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440:** 637–643

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

Lee L, Tirnauer JS, Li J, Schuyler SC, Liu JY, Pellman D (2000) Positioning of the mitotic spindle by a cortical-microtubule capture mechanism. *Science* **287:** 2260–2262

Lehner B (2010) Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PLoS ONE* **5:** e9035

Liu S, Wilson KA, Rice-Stitt T, Neiman AM, McNew JA (2007) *In vitro* fusion catalyzed by the sporulation-specific t-SNARE light-chain Spo20p is stimulated by phosphatidic acid. *Traffic* **8:** 1630–1643

Lynch M, Katju V (2004) The altered evolutionary trajectories of gene duplicates. *Trends Genet* **20:** 544–549

Manning BD, Barrett JG, Wallace JA, Granok H, Snyder M (1999) Differential regulation of the Kar3p kinesin-related protein by two associated proteins, Cik1p and Vik1p. *J Cell Biol* **144:** 1219–1233

Marland E, Prachumwat A, Maltsev N, Gu Z, Li W (2004) Higher gene duplicabilities for metabolic proteins than for nonmetabolic proteins in yeast and E coli. *J Mol Evol* **59:** 806–814

Marques AC, Vinckenbosch N, Brawand D, Kaessmann H (2008) Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. *Genome Biol* **9:** R54–R54

Musso G, Costanzo M, Huangfu M, Smith AM, Paw J, San Luis B, Boone C, Giaever G, Nislow C, Emili A, Zhang Z (2008) The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Res* **18:** 1092–1099

Musso G, Zhang Z, Emili A (2007) Retention of protein complex membership by ancient duplicated gene products in budding yeast. *Trends Genet* **23:** 266–269

Myers C, Barrett D, Hibbs M, Huttenhower C, Troyanskaya O (2006) Finding function: evaluation methods for functional genomic data. *BMC Genomics* **7:** 187

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320:** 1344–1349

Novick P, Osmond BC, Botstein D (1989) Suppressors of Yeast actin mutations. *Genetics* **121:** 659–674

Ohno S (1970) *Evolution by Gene Duplication*. Berlin, New York: Springer-Verlag

Ohya Y, Sese J, Yukawa M, Sano F, Nakatani Y, Saito TL, Saka A, Fukuda T, Ishihara S, Oka S, Suzuki G, Watanabe M, Hirata A, Ohtani M, Sawai H, Fraysse N, Latgé J, François JM, Aebi M, Tanaka S *et al* (2005) High-dimensional and large-scale phenotyping of yeast mutants. *Proc Natl Acad Sci USA* **102:** 19015–19020

Page B, Satterwhite L, Rose M, Snyder M (1994) Localization of the Kar3 kinesin heavy chain-related protein requires the Cik1 interacting protein. *J Cell Biol* **124:** 507–519

Papp B, Pal C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429:** 661–664

Presser A, Elowitz MB, Kellis M, Kishony R (2008) The evolutionary dynamics of the Saccharomyces cerevisiae protein interaction network after duplication. *Proc Natl Acad Sci USA* **105:** 950–954

Rost B (1999) Twilight zone of protein sequence alignments. *Protein Eng* **12:** 85–94

Scannell D, Wolfe K (2008) A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Res* **18:** 137–147

Sproul LR, Anderson DJ, Mackey AT, Saunders WS, Gilbert SP (2005) Cik1 targets the minus-end kinesin depolymerase Kar3 to microtubule plus ends. *Curr Biol* **15:** 1420–1427

Tirnauer JS, O'Toole E, Berrueta L, Bierer BE, Pellman D (1999) Yeast Bim1p promotes the G1-specific dynamics of microtubules. *J Cell Biol* **145:** 993–1007

Tirosh I, Barkai N (2007) Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. *Genome Biol* **8:** R50

Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CWV, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294:** 2364–2368

Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L *et al* (2004) Global mapping of the yeast genetic interaction network. *Science* **303:** 808–813

Wagner A (2002) Asymmetric functional divergence of duplicate genes in yeast. *Mol Biol Evol* **19:** 1760–1768

Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449:** 54–61

Yang H, Nakanishi H, Liu S, McNew JA, Neiman AM (2008) Binding interactions control SNARE specificity *in vivo*. *J Cell Biol* **183:** 1089–1100

Yang J, Gu Z, Li W (2003) Rate of protein evolution versus fitness effect of gene deletion. *Mol Biol Evol* **20:** 772–774

Zhang P, Gu Z, Li W (2003) Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol* **4:** R56–R56