



The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast

Gabriel Musso, Michael Costanzo, ManQin Huangfu, et al.

Genome Res. 2008 18: 1092-1099 originally published online May 7, 2008

Access the most recent version at doi:[10.1101/gr.076174.108](https://doi.org/10.1101/gr.076174.108)

References This article cites 52 articles, 19 of which can be accessed free at:
<http://genome.cshlp.org/content/18/7/1092.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2008, Cold Spring Harbor Laboratory Press

The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast

Gabriel Musso,^{1,2} Michael Costanzo,^{1,2} ManQin Huangfu,^{1,2} Andrew M. Smith,^{1,2} Jadine Paw,² Bryan-Joseph San Luis,¹ Charles Boone,^{1,2} Guri Giaever,^{1,3} Corey Nislow,^{1,2} Andrew Emili,^{1,2,4} and Zhaolei Zhang^{1,2,4}

¹Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada; ²Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada; ³Department of Pharmaceutical Sciences, University of Toronto, Toronto, Ontario M5S 3M2, Canada

Since complete redundancy between extant duplicates (paralogs) is evolutionarily unfavorable, some degree of functional congruency is eventually lost. However, in budding yeast, experimental evidence collected for duplicated metabolic enzymes and in global physical interaction surveys had suggested widespread functional overlap between paralogs. While maintained functional overlap is thought to confer robustness against genetic mutation and facilitate environmental adaptability, it has yet to be determined what properties define paralogs that can compensate for the phenotypic consequence of deleting a sister gene, how extensive this epistasis is, and how adaptable it is toward alternate environmental states. To this end, we have performed a comprehensive experimental analysis of epistasis as indicated by aggravating genetic interactions between paralogs resulting from an ancient whole-genome duplication (WGD) event occurring in the budding yeast *Saccharomyces cerevisiae*, and thus were able to compare properties of large numbers of epistatic and non-epistatic paralogs with identical evolutionary times since divergence. We found that more than one-third (140) of the 399 examinable WGD paralog pairs were epistatic under standard laboratory conditions and that additional cases of epistasis became obvious only under media conditions designed to induce cellular stress. Despite a significant increase in within-species sequence co-conservation, analysis of protein interactions revealed that paralogs epistatic under standard laboratory conditions were not more functionally overlapping than those non-epistatic. As experimental conditions had an impact on the functional categorization of paralogs deemed epistatic and only a fraction of potential stress conditions have been interrogated here, we hypothesize that many epistatic relationships remain unresolved.

[Supplemental material is available online at www.genome.org.]

Duplication is the primary mechanism for gene creation and consequent emergence of functional novelty in eukaryotic cells (Ohno 1970). While duplication typically involves one or few genes (small-scale duplication, or SSD), polyploidization events by whole-genome duplication (WGD) represent unique, abundant sources of functional diversity in the evolutionary histories of many eukaryotic species, accounting for ~15% of all genes in yeast (Wolfe and Shields 1997; Kellis et al. 2004) and sizeable fractions in plants (Blanc et al. 2000) and vertebrates (Jeffreys et al. 1980; Christoffels et al. 2004; Dehal and Boore 2005). Although WGD-resultant duplicates (also known as “ohnologs” in honor of Susumu Ohno [Wolfe 2000] but referred to herein simply as “paralogs”) have inherent differences from small-scale duplicates (Davis and Petrov 2005; Guan et al. 2007), they serve as a popular target for study since they represent a replete body of paralogs of identical age. The WGD-created genes facilitate comparative analysis of paralogs spanning diverse functional categories, while excluding partial duplicates and pseudogenes of dubious significance.

Initially following a WGD event, the resulting increased ploidy is thought to be genomically unstable (Mayer and Agu-

ilera 1990), leading to rapid and extensive gene loss. Two consequent questions are: how much functional overlap exists between that subset of duplicates that are retained over much longer periods, and does this functional buffering bestow any long-term selective advantage? Preservation of functional congruency is suspected to confer genomic robustness (Kirschner and Gerhart 1998; Conant and Wagner 2004), yet complete redundancy is unlikely from an evolutionary standpoint (Brookfield 1992), owing to relaxed selective pressure on one or both of the duplicates.

Previous examinations of the biological properties of extant paralogs based on the physical interactome (Baudot et al. 2004; Guan et al. 2007; Musso et al. 2007; Wapinski et al. 2007), metabolic networks (Papp et al. 2004; Kuepfer et al. 2005; Harrison et al. 2007), and single-gene deletion phenotypes (Gu et al. 2003) in the budding yeast *Saccharomyces cerevisiae* have implied extensive functional similarity among both WGD and SSD resultant paralogs, supporting an advantage to retaining substantive functional overlap. Conversely, synthetic genetic array analysis of epistasis (herein defined as the capacity for paralogs to buffer the phenotypic consequence of deletion of their sister gene) among a small subset of yeast WGD and SSD paralogs has led to an alternate hypothesis that while some paralog pairs have maintained the ability to buffer loss of a respective sister, this mechanism is limited in scope, not functioning over a wide range of compromising environmental conditions (Ihmels et al. 2007). This assertion

⁴Corresponding authors.

E-mail andrew.emili@utoronto.ca; fax (416) 978-8287.

E-mail zhaolei.zhang@utoronto.ca; fax (416) 978-8287.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.076174.108>.

contrasts with previous suggestions that duplicates may be preferentially retained to compensate for cellular stresses or perturbations (Gu et al. 2003; Musso et al. 2007). Consequently, the extent and context of functional buffering among WGD-resultant duplicates as well as the molecular properties of buffering paralogs remain to be resolved.

To address these issues directly, we have examined the relative fitness of haploid yeast strains bearing single and double deletions of all surveyable WGD-resultant paralog pairs in yeast. We find that more than one-third of surveyed WGD duplicates substantively buffer the loss of their respective sister genes under standard laboratory growth conditions. Further examination of epistasis under stressful conditions revealed additional instances of epistasis, demonstrating that the function of buffering paralogs (and by extension their within-species co-conservation and expression, which are inherently linked to function) is dependent on experimental conditions. As many of the innumerable environmental conditions remain unexplored, we submit that epistasis may be highly extensive among extant WGD-resultant paralogs.

Results

Frequent phenotypic buffering between WGD-resultant duplicates

We used two complementary experimental growth assays to systematically monitor the fitness of single and double mutants to determine the extent of phenotypic buffering among putative yeast WGD paralog pairs (Kellis et al. 2004) under standard culture conditions. We were unable to assess seven pairs because one or both paralogs was split into multiple open reading frames (Kellis et al. 2004), while 51 pairs were excluded from analysis because of the inviability of one or both of the single-deletion strains (see Supplemental Spreadsheet), leaving 399 surveyable pairs out of the initial set of 457.

Random-spore analysis (RSA) was first applied to measure the overall viability of the progeny of genetic crosses between individual single gene deletion strains. Haploid yeast strains containing deletions corresponding to either one or both paralogs of a WGD pair were grown on solid minimal media and selected for based on specific drug sensitivities (deleted genes were replaced by drug-resistance cassettes) (see Methods). Visual inspection conducted by two independent evaluators was ultimately used to define 51 obvious cases of synthetic sickness or lethality (see Supplemental Fig. S1; Supplemental Table S1). Tetrad dissection additionally confirmed 18 of the 31 pairs initially deemed non-obvious by either or both evaluators (see Methods), ultimately leading to the identification of epistasis among 69 WGD paralog pairs (17% of all pairs tested, 15% of all WGD paralog pairs). This frequency of epistasis for WGD paralog pairs is well beyond what would be expected for randomly selected gene pairs (<1% based on synthetic genetic array data) (Tong et al. 2004), and furthermore, beyond the eightfold to 10-fold increases in epistasis expected for gene pairs with similar or identical Gene Ontology (GO) annotations, respectively (Tong et al. 2004).

Next, growth-curve analysis (GCA) was applied as an alternate means to quantify growth rates to detect attenuated instances of epistasis among WGD paralogs. Unlike in RSA, growth for GCA is assayed in rich liquid media and culture growth is monitored through optical density, allowing more precise calculations of fitness, and hence, identification of more subtle growth

defects (see Methods). Through GCA, 119 double-mutant strains were identified as having growth decreased beyond that predicted by multiplying the fitness defects of the corresponding individual mutations (i.e., using a multiplicative model) (see Methods), 71 of which had not been witnessed through RSA. These data suggest that epistasis exists to some extent among 140 WGD paralog pairs (35% of those surveyed, 31% of all 457 WGD paralogs).

The RSA and GCA data were not completely overlapping (see Fig. 1), as nearly one-third of paralog pairs with obvious growth defects detected in RSA showed normal growth properties through GCA (listed in Supplemental Table S2). These remaining 21 pairs may have a function that specifically limited double-mutant growth in the minimal media of the RSA experiments (e.g., included in this list are *ENO1* and *ENO2*, which are known to function in concert only in low-glucose conditions; McAlister and Holland 1982) or the ability to grow on solid media.

Overall, these results indicate that epistasis occurs among approximately one-third of WGD paralog pairs. However, the difference in epistatic relationships as measured across liquid and solid media suggests a prevalence of condition-specific lethality, underscoring the importance of assaying genetic buffering under multiple conditions.

Further buffering relationships identified under simulated stress conditions

To further explore epistasis in alternate conditions, and to investigate previous hypotheses that buffering relationships may specifically be maintained to cope with cellular perturbations (Gu et al. 2003; Musso et al. 2007), we expanded our assay to include a

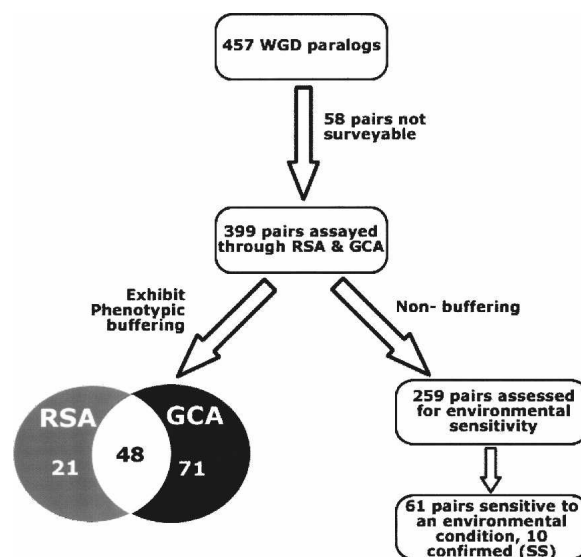


Figure 1. Experimental schema and summary of epistasis among WGD paralog pairs in yeast. The 399 WGD paralog pairs with viable constituent single-deletion strains (Giaever et al. 2002) were analyzed for genetic interaction using random-spore analysis (RSA) and growth-curve analysis (GCA). The overlap in terms of duplicates exhibiting significant phenotypic buffering is shown in the Venn diagram. The 259 WGD paralog pairs not exhibiting epistasis (and hence deemed to not buffer phenotypically) in rich growth media were further analyzed through environmental screening in five alternate media conditions designed to either mimic common cellular stress states or introduce competition (Giaever et al. 2002).

series of stresses. The relative fitness of the 259 double-deletion mutants corresponding to non-buffering WGD pairs (i.e., those not defined to be epistatic by either RSA or GCA; see Methods) were monitored for sensitivity in five media designed to induce cellular stress (alkaline pH, high salt, the presence of the antifungal agent Nystatin [decreases membrane permeability], high levels of Sorbitol, or with glycogen substituting for dextrose as the main carbon source) (Fig. 1) (Giaever et al. 2002). As a competitive growth assay suited to examining multiple media types, we measured the relative fitness of bar-coded double-mutant strains simultaneously by quantitative microarray hybridization (Giaever et al. 2002). Double-deletion strains exhibiting sensitivity within any of the five stress conditions were further analyzed using GCA conducted in the specific condition, allowing independent validation of these results.

To be considered epistatic in a stress condition, the paralog pair had to satisfy two criteria. First, the pair had to not be epistatic in standard media conditions, indicating that if epistasis was detected in a stress condition, it was unique to that condition. Second, epistatic paralogs had to pass the multiplicative model in the given condition, meaning that either individual knockout could not be responsible for the stress sensitivity. We first detected candidate condition-specific fitness defects through microarray, as indicated by a >25% decrease in strain abundance for a given condition (see Supplemental Methods) among 61 paralog pairs not detected as buffering in standard media by either RSA or GCA (23.5% of those tested). To ensure that fitness defects were the result of the combined mutation and not the deletion of either single mutant, we conducted GCA for each of these 61 pairs in the appropriate sensitive condition. We found that 10 of the 61 pairs met the stringent confines of a multiplicative fitness model (see Methods), and thus were epistatic under the given stress conditions (see Supplemental Spreadsheet).

These 10 confirmed WGD pairs (herein referred to as “sensitive to stress,” or SS; see Table 1) both reaffirm previously described observations and suggest new functional synergies. Among those relationships confirmed are the hypersensitivity to high salt of the *BUL1/BUL2* double mutant (Yashiroda et al. 1998) and the hypersensitivity to carbon-source starvation of the *MSN2/MSN4* double mutant (Estruch and Carlson 1993). Interestingly, the *PYC1/PYC2* double mutant with known hypersen-

sitivity to glucose (Stucka et al. 1991) was found here to be epistatic in multiple conditions, but showed normal growth with galactose as the primary carbon source. Other characterizations of epistasis can also be explained given known protein functions. For example, *YGK3* and *MCK1* are members of the cell-wall integrity pathway (Griffioen et al. 2003), while *CHS6* and *BCH2* are involved in the transport of membrane proteins (Wang et al. 2006); both pairs were found here to be epistatic in the presence of Nystatin. Furthermore, *CNA1* and *CMP2*, which were epistatic here in 1 M NaCl, are involved in ion homeostasis and confer tolerance to high levels of sodium (Mendoza et al. 1996). Our inherent interpretation of these findings is that WGD paralog pairs that are non-epistatic in standard conditions may be epistatic in a condition that emphasizes their principle function, despite the fact that neither individual paralog is essential for survival in that function (i.e., epistasis can be evolved for specific circumstances).

Based on our observations of condition-specific epistasis among SS paralogs, we are able to propose functionality for some genes that are either uncharacterized or poorly understood. Specifically, as *YSP2* and its paralog *YHR080C* are epistatic in multiple conditions, we propose that the known pro-apoptotic protein *YSP2* may also exercise an inherent rescue function in conjunction with its paralog in response to cell wall or plasma membrane damage. Also, the epistasis of *SOL1* and *SOL2* in multiple conditions suggests that they may be involved in a general stress response. Lastly, we submit that the putative protein of unknown function, *YFR039C*, functions synergistically (either in conjunction or within an alternate functional pathway) with the glycosylphosphatidylinositol-anchored protein, *She10*, in response to cell stress, notably high salt and alterations in carbon source.

Experimental condition impacts the functional composition of epistatic paralogs

As the investigation above had revealed SS paralogs to generally have functions related to their incident condition (analogous to what has been previously reported for metabolic enzymes using flux balance analysis; Harrison et al. 2007), we next tested whether those paralogs epistatic only under standard conditions (i.e., those initially detected by RSA and GCA) had any common functional properties. Enrichment analysis (Maere et al. 2005) of functional categorization as assessed by the GO database (Tatusov et al. 2003) indicated that these paralogs were biologically disparate from their non-buffering counterparts. Specifically, WGD paralogs epistatic under standard laboratory conditions were more frequently involved in cell growth, protein metabolism, and division.

Many buffering paralogs (RSA and GCA) are annotated as ribosomal proteins or metabolic enzymes (23 pairs in each) (see Supplemental Spreadsheet). However, even upon removal of these proteins, functional GO enrichment analysis revealed that, although somewhat more functionally diverse, the remaining paralogs epistatic under normal growth conditions were still significantly enriched for growth-related processes. More specifically, while paralogs detected jointly by both RSA and GCA were mainly involved in protein production, those detected solely by GCA were notably enriched for involvement in the cell cycle (i.e., septin ring formation, cell cycle checkpoint, regulation of cell cycle), and those detected solely by RSA were seemingly enriched for metabolic processes and structure development. No paralog

Table 1. WGD paralogs exhibiting epistasis following environmental perturbation

Systematic Name 1	Systematic Name 2	Standard Name 1	Standard Name 2	Condition
YML111W	YMR275C	<i>BUL2</i>	<i>BUL1</i>	NaCl
YLR433C	YML057W	<i>CNA1</i>	<i>CMP2</i>	NaCl
YBR218C	YGL062W	<i>PYC2</i>	<i>PYC1</i>	Sorbitol, Nystatin
YPR074C	YBR117C	<i>TKL1</i>	<i>TKL2</i>	Nystatin
YJL099W	YKR027W	<i>CHS6</i>	<i>BCH2</i>	Nystatin
YNL307C	YOL128C	<i>MCK1</i>	<i>YGK3</i>	Nystatin
YDR326C	YHR080C	<i>YSP2</i>		Nystatin
YKL062W	YMR037C	<i>MSN4</i>	<i>MSN2</i>	Galactose
YCR073W-A	YNR034W	<i>SOL2</i>	<i>SOL1</i>	NaCl, Nystatin, galactose
YGL228W	YFR039C	<i>SHE10</i>		NaCl, galactose

Listed are the set of 10 non-buffering WGD paralog pairs assessed to be sensitive to at least one of the five experimental conditions examined (first identified through reduced hybridization of competitively grown bar-coded double mutants to a microarray, then subsequently confirmed using GCA).

pairs detected to be epistatic jointly by both RSA and GCA contained genes of unknown function. In contrast, non-epistatic paralogs were significantly enriched for elements of signal transduction (e.g., amino acid phosphorylation, signal transduction, cell communication; see Fig. 2), and frequently (25% of pairs) one or both paralogs were of unknown molecular function. As signal transduction is often associated with environmental or stress response and analysis of epistasis in stress conditions above suggests a distinct impact of survey condition on the functional properties of detected epistatic relationships, the possibility exists that some functional classes of genes (notably those that act in signal transduction or that are of currently unknown function) may have established yet-undetermined epistatic relationships. These relationships may only be evident in response to the stimulus that specifically stresses their functional pathway.

Paralogs buffering under standard conditions are highly co-conserved

We next investigated whether buffering paralogs had any additional unique evolutionary or genomic properties correlating

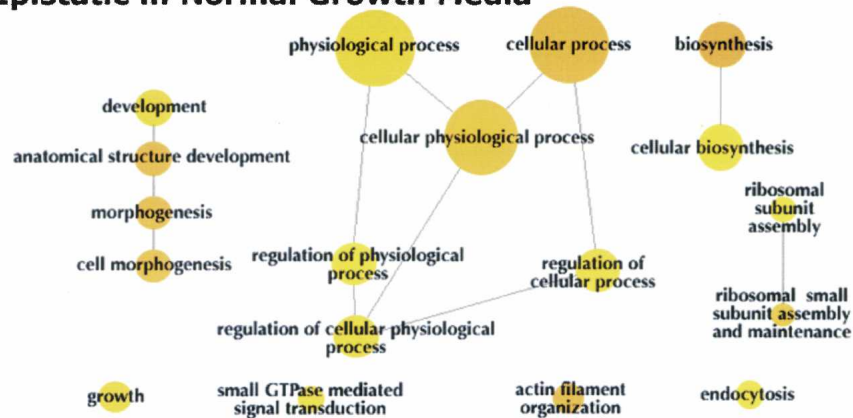
with their epistatic capacity. To eliminate any potential method-specific bias, when evaluating paralogs epistatic under normal laboratory conditions, we confirmed results using paralogs separated on the basis of joint detection by both the RSA and GCA (intersect group, IG), and using either of these two methods alone (union group, UG). First, we examined whether presence of additional (i.e., non-WGD-resultant) duplicates had an impact on the potential for phenotypic buffering. We noted that both buffering and non-buffering WGD paralog pairs showed equal enrichment for instances of multiple paralogy (i.e., having undergone additional, non-WGD, duplication events had no impact on epistasis). Specifically, 23% of epistatic WGD-paralog pairs had additional duplicates, compared to 20% for non-epistatic paralogs (for full analysis, see Supplemental Information and Supplemental Fig. S2).

Next, upon examining within-species sequence conservation, buffering paralogs (IG and UG) had significantly higher average sequence similarity than corresponding non-buffering paralogs (86% for IG, 77% for UG, versus 65% for non-buffering), and a consequent decrease in non-synonymous mutation rate

(tallied using *Kluyveromyces waltii* as the ancestral out-group) (see Methods), confirming increased within-species co-conservation. These findings were statistically robust ($P < 0.05$ for all comparisons, Mann-Whitney rank sum test; see Fig. 3A) against the removal of ribosomal proteins, which exhibit disproportionately high levels of sequence conservation. Despite these trends, it is interesting to note that this increased within-species co-conservation is not exclusive to buffering paralogs. For example, the UG contained three buffering paralog pairs sharing <40% sequence identity (*PHD1* and *SOK2*, *MGA2* and *SPT23*, and *BOI1* and *BOI2*), implying that co-conservation is not a prerequisite for phenotypic buffering. Furthermore, SS paralogs did not exhibit the same stringent conservation of sequence as IG and UG paralogs (the average sequence similarity among the 10 confirmed SS paralogs was 68%).

A breakdown of WGD paralogs by functional category as assigned by GO Slim (<http://geneontology.org/GO.slims.shtml>) demonstrates that epistasis under standard conditions is much more frequent among those paralogs involved in certain growth- and division-related processes (Fig. 4). Also, there is a direct linear correlation between the fraction of buffering paralogs within a functional group and the overall sequence similarity between paralogs in that group ($r = 0.64$, $P < 0.005$, linear regression $P < 0.05$; see Fig. 4). Therefore, as function and conservation appear to be intertwined, the possibility that functional bias is responsible for the observed increase in co-conservation of

Epistatic in Normal Growth Media



Non-epistatic in Normal Growth Media

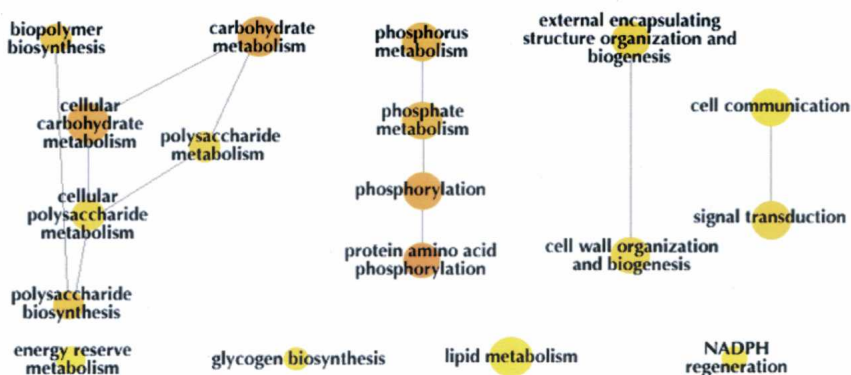


Figure 2. GO representation of epistasis under normal growth conditions. Depicted are the GO functional categories over-represented for paralogs both epistatic (*above*) and non-epistatic (*below*) under normal growth conditions (i.e., those detected by either RSA or GCA, the union group). The size of the node reflects the fraction of paralogs involved in that over-represented process, and darker coloring denotes higher significance. While the significance of each term was calculated independently (regardless of placement within the GO classification system), lines indicate continuance along the path of the GO directed acyclic graph. Increased statistical stringency ($P < 0.0005$) was used for depicted terms in order to account for increased representation of more general terms. The picture was created using the BinGO Cytoscape plug-in.

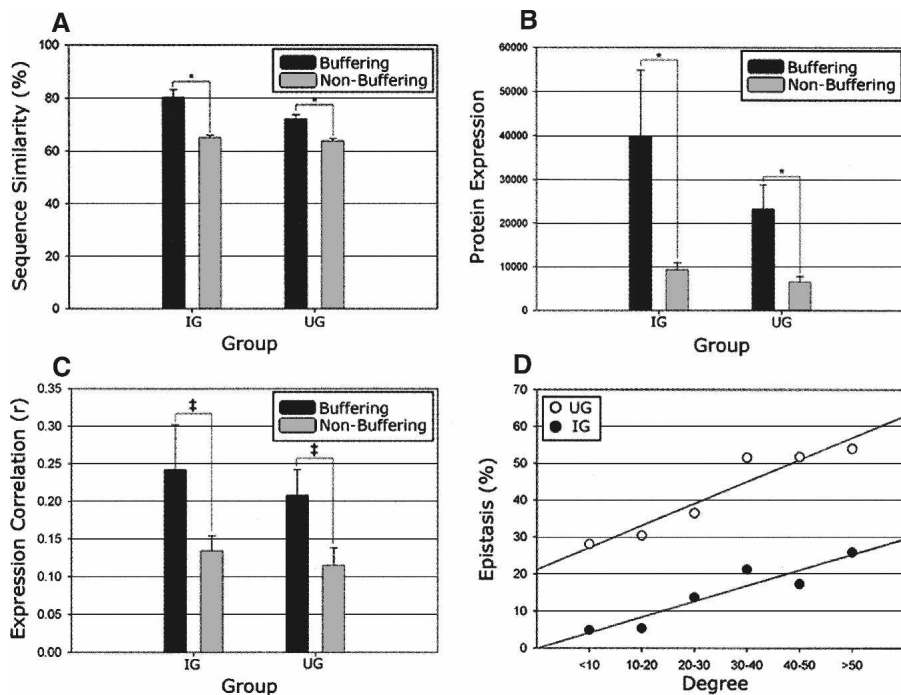


Figure 3. Expression and co-conservation of buffering paralogs. All relationships depict comparisons between epistatic paralogs (IG and UG) and respective non-epistatic following the removal of ribosomal proteins. (A) Buffering paralogs have significantly more co-conservation of sequence than non-buffering ($P < 0.001$ for IG and UG, indicated by asterisks). (B) Buffering paralogs are significantly more highly expressed than non-buffering (protein abundance depicted, same results noted for transcript abundance) and additionally exhibit a trend toward ($P < 0.1$, indicated by ‡) being more highly correlated throughout the cell cycle (displayed in C). (D) All WGD paralog pairs were binned based on their number of combined protein interactions (degree) within the BioGRID data set, with the respective percentage epistatic (both IG and UG) indicated, demonstrating a clear relationship between the number of physical interactions and the likelihood for epistasis under normal growth conditions.

paralogs epistatic under normal conditions cannot be immediately discounted.

On average, buffering paralogs (IG and UG) also exhibited significantly higher basal mRNA (Greenbaum et al. 2002) and protein (Ghaemmaghami et al. 2003) expression levels ($P < 0.05$, Mann-Whitney rank sum test; Fig. 3B) when compared to non-buffering (robust against removal of ribosomal proteins; see Fig. 3B). As expression magnitude is intrinsically linked to sequence conservation for yeast paralogs (Pal et al. 2001) and epistatic paralogs are more highly conserved, this finding is not unexpected. However, while IG and UG paralogs initially had more highly correlated mRNA expression both throughout the cell cycle ($P < 0.01$) (see Methods) and across varying cell conditions (Kafri et al. 2008), following the removal of ribosomal proteins, these relationships are no longer statistically significant. Therefore, expression correlation does not appear to be generally predictive of epistatic capacity for WGD-resultant paralogs.

Physical interactions are not predictive of phenotypic buffering

To test a recently observed correlation between the number of physical interaction partners (the “degree” of a protein) and the propensity to back up a sister paralog (Kafri et al. 2008), we next compared interactions of epistatic and non-epistatic paralogs.

Using the identical data set that this assertion had been based on (the full interaction data set contained at BioGRID) (Reguly et al. 2006), both IG and UG paralogs had significantly more interactions per pair than non-buffering ($P < 0.05$; Mann-Whitney rank sum test) (Fig. 3D). The significance of this relationship was robust against removal of ribosomal proteins that had a disproportionately high number of interactions. Alternately, SS paralogs were not significantly greater or lesser in degree than the remaining WGD paralogs, suggesting again that increased degree may be linked to the functional bias, increased expression, or increased conservation of IG and UG paralogs, and not necessarily an inherent property of epistatic relationships. The difference in degree between sister paralogs had no noticeable correlation with epistasis.

One could logically assert that paralog pairs capable of buffering deletion of their sister should be more congruent in function than those that do not. We next attempted to test this assertion using annotated physical interactions as a direct proxy of function (Baudot et al. 2004). We compared the extent of shared protein–protein interaction partners of buffering versus non-buffering WGD sister paralogs based on the results of two high-throughput proteomic screens (Gavin et al. 2006; Krogan et al. 2006). To eliminate potential biases created by high-throughput assay, we further confirmed all results using the aforementioned BioGRID data set (Reguly et al. 2006) following the manual removal of data resulting from high-throughput screens. Buffering and non-buffering WGD paralogs were compared in the following aspects: (1) propensity of the sisters to physically interact with each other, (2) propensity to be co-grouped into associative protein complexes, (3) the number of shared interacting partners, and (4) the proportion of non-shared partners (first two analyzed by the Fisher exact test, all else using Mann-Whitney rank sum test) (see Methods).

Only one comparison achieved marginal statistical significance. IG paralogs were more likely to interact with each other in the Krogan et al. (2006) interactome data set ($P < 0.05$), but this trend was not confirmed in either the Gavin et al. (2006) genome-wide study or using the BioGRID data set. Although our findings depend on the assumption that the current physical interaction data provide a fairly complete and consistent coverage of the WGD paralogs, these results perplexingly suggest that paralog pairs deemed to be buffering under standard laboratory conditions are not necessarily more functionally redundant. As previous studies have indicated a pervasive overlap in terms of the physical interaction partners of WGD-resultant duplicates (Baudot et al. 2004; Guan et al. 2007; Musso et al. 2007), one potential explanation for the lack of correlation between physical interaction data and epistatic capacity is that not all epistatic relationships have been revealed.

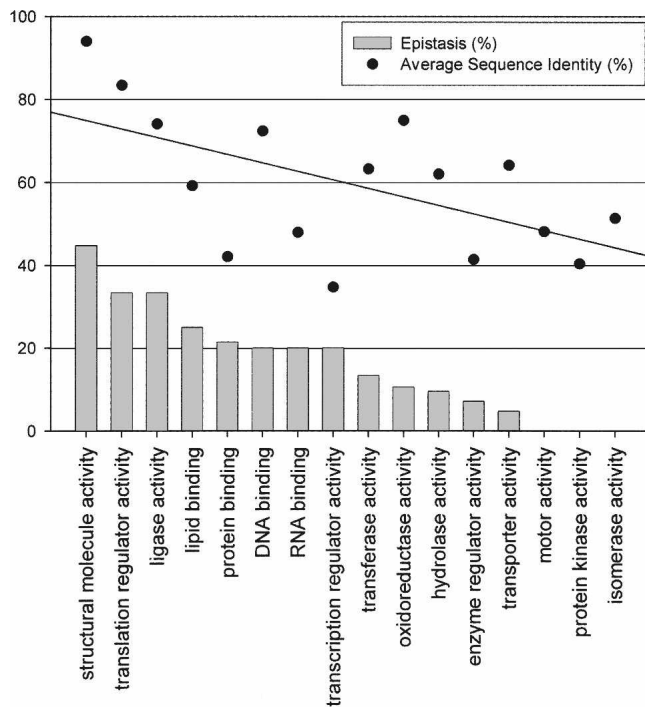


Figure 4. Buffering paralogs by GO slim annotation. WGD paralogs were grouped into functional categories based on the broad definitions of the GO slim hierarchy and ranked in decreasing order based on the percentage of paralogs found to be buffering (IG), indicated with bars. Only those paralog pairs with single, matching annotations were included, resulting in 315 depicted pairs (categories “other” and “molecular function” were removed). The juxtaposed dot plot indicates the average percent sequence identity of functional groups (buffering and non-buffering paralogs combined); the overlaid line indicates linear regression. There is a significant correlation ($r = 0.64$, $P < 0.005$) between the epistatic capacity of a functional group and the co-conservation of paralogs contained therein.

Discussion

The noted lack of cell morbidity when deleting individual yeast genes with at least one paralog has led to speculation that duplication is reserved for genes of limited functional importance (He and Zhang 2006). By extension, WGD-resultant paralogs (which have generally less consequence upon single-gene deletion than SSD-resultant) (Hakes et al. 2007) should be of even less functional importance than duplicates of other origin. However, this absent decrease in phenotypic consequence upon deletion of a WGD-resultant duplicate can alternately be explained by the elevated capacity of this group to buffer deletions. The degree of epistasis among duplicates is inevitably a compromise between the maintenance of mechanisms that could both protect against gene loss and confer stress resilience and the establishment of functional novelty. As we have witnessed here that epistasis is possible among even minimally co-maintained WGD paralogs, complete abolition of functional overlap would seem to have very little comparative benefit.

The frequency of WGD paralog pairs exhibiting phenotypic buffering under normal media conditions (~35% of those surveyed, or ~31% of all 457 WGD paralogs) observed in this study is similar to that previously determined through examination of a much smaller subset of duplicates (25 WGD- and 20 SSD-resultant) using a high-density genetic interaction map (Ihmels

et al. 2007). However, while the previous study suggested that there was little additional evidence of functional redundancy in alternate environmental conditions, our results conversely suggest there may be limitless potential to reveal buffering.

While paralogs epistatic in solid media and rich liquid media are notably more rigorously co-conserved and highly expressed, neither of these properties appear to be essential for epistasis. Furthermore, overlap in shared physical interaction partners, which serves as a proxy for overlapping function, showed no consistent difference between epistatic and non-epistatic paralogs. In a finding just as perplexing, the presence of additional, non-WGD-resultant duplicates appeared to have little to no impact on phenotypic buffering. As the nature of epistasis among WGD paralogs is highly dependent on the environmental context, that is, “environmental robustness” (Harrison et al. 2007), and as only a small fraction of the potential stresses (as well as the endless permutations of stress magnitudes and combinations) and alternate environmental conditions were explored here, many non-epistatic WGD paralog pairs (25% of which are of unknown function) may yet have preserved an unwitnessed buffering mechanism over ~100 million years of evolution. This is reminiscent of what had previously been observed in comprehensive phenotypic studies of single-gene deletion strains in *S. cerevisiae* wherein loss of only ~19% of genes results in morbidity (the so-called essential genes) (Giaever et al. 2002). Just as many single genes serve a function imperceptible under laboratory conditions but are required for viability specifically under cellular duress (Gasch et al. 2000; Giaever et al. 2002) or are suspected to function in non-laboratory growth conditions (Pena-Castillo and Hughes 2007), certain WGD paralogs may retain a condition-specific buffering capacity.

The observations regarding concerted expression of buffering paralogs presented herein are surprising given previous findings. Specifically, it had been demonstrated that dispensable (i.e., buffering) paralogs generally have both a higher degree and more disparate expression patterns than non-buffering (Kafri et al. 2005, 2008). While we do confirm here that WGD paralogs epistatic under standard conditions have a higher degree, they displayed no more disparity in expression than non-buffering (initially found to be significantly more correlated, although this was a byproduct of the influence of ribosomal proteins), illustrating potential differences between the epistatic mechanisms of WGD-resultant paralogs and duplicates of other origin (i.e., SSD).

As the existence of duplicated genes can confound predictions of epistasis (Harrison et al. 2007), a comprehensive understanding of the buffering capacity of paralogs is essential before extrapolating the knowledge gleaned from model organisms, such as budding yeast to higher eukaryotes. Since the human genome has arguably been profoundly influenced by genome duplication events (Dehal and Boore 2005), it will ultimately be the understanding of the adaptive epistatic mechanisms evolved to cope with the abnormal states of stress and perturbation that will lend the most insight into disease-related maladaptive processes.

Methods

Assessment of synthetic lethality

The method for screening of synthetic lethality through random spore analysis (RSA) was based on that previously used (Tong et al. 2001, 2004). For growth curve analysis (GCA), MATa spore

progeny obtained as described for RSA were grown in 96-well format in media selecting for double mutants (SD + Canavanine + G418 + NAT). The growth rate was monitored for five generations using a TECAN reader, and corresponding curves were analyzed visually. For RSA and GCA scoring scheme and results, please see the Supplemental material.

Environmental screening

The combined lethality of double-mutant haploid yeast strains, grown as described above and combined into a pool of 499 members (399 WGD paralog pairs, 100 double-mutant control strains; slow-growers supplemented to obtain equal starting concentrations of each double-mutant strain), was examined in five media conditions previously designed (Giaever et al. 2002) to either introduce competition or mimic common stress states (1 M NaCl, 1.5 M sorbitol, YPGal [where 2% galactose was substituted for dextrose], 10 μ M Nystatin, and YPD at pH 8). Equalized concentrations of cells were grown in respective media conditions for five generations, at which time cells were lysed using a QIAGEN DNeasy kit and DNA was isolated. Both UPTAG and DNTAG barcodes were amplified via PCR and hybridized to high-density oligonucleotide Affymetrix arrays (as described previously [Giaever et al. 2002], except with use of Tag4 arrays [Pierce et al. 2006]). For determination of condition-specific lethality, please see the Supplemental material.

Analysis of physical interactions

The physical interactions of genetic interactors and noninteractors were compared as described previously (Musso et al. 2007) using two published tandem affinity purification interaction data sets (Gavin et al. 2006; Krogan et al. 2006), and the data compiled in BioGRID (Reguly et al. 2006) following removal of all data generated using a high-throughput assay (performed manually). Lists of protein complexes were also obtained from Krogan et al. (2006) and Gavin et al. (2006) publications, as well as the MIPS database (Mewes et al. 2004). Briefly, genetic interactors (IG and UG) were compared against non-interactors in five categories: number of shared interactions, interaction overlap score (number of shared interactions per unshared interactions), non-shared interaction score (previously described as the negative logarithm of the fraction of non-shared interactions) (de Lichtenberg et al. 2005), propensity to interact with each other, and propensity to co-cluster (the latter two analyzed by Fisher exact test; all else analyzed using the Mann-Whitney rank sum test).

Comparisons of conservation and expression

Sequence identity among paralogs was evaluated through application of a global alignment algorithm (Needle as part of the EMBOSS suite of programs) (Rice et al. 2000) calculated on amino acid sequences. The nonsynonymous substitution rate (K_a) was calculated using Codeml (Yang 1997) (again through EMBOSS; Rice et al. 2000) against the appropriate *K. waltii* ancestor (Kellis et al. 2004). While the ratio of non-synonymous to synonymous substitution rate is the preferred metric when assessing evolutionary conservation, the synonymous substitution rates were saturated and therefore unusable. Protein (Ghaemmaghami et al. 2003) and mRNA (Greenbaum et al. 2002) expression data were obtained as published. To compare temporal expression patterns of paralog pairs, expression data were compiled at various time points throughout the cell cycle using two previously published data sets (Cho et al. 1998; Kafri et al. 2008). These Cho et al. data were first normalized based on the logarithm of each gene's median intensity value, then the Pearson correlation coefficient for every paralog pair over the first eight time points was determined

(internal analysis revealed that all points correlated poorly beyond the first eight time points; data not shown). The Kafri et al. data set was used as published.

Statistical analysis

All statistical calculations and correlations were performed using SigmaStat 3.1. The representation of statistically enriched GO terms was drawn using the BinGO plug-in (Maere et al. 2005) for the Cytoscape software environment (Shannon et al. 2003).

Acknowledgments

We thank Gary Bader for input on analysis of the protein interaction network and interpretation of previous SGA data. We also thank J. Javier Diaz Mejia for insightful comments on the manuscript, and Marinella Gebbia and Larry Heisler for assistance in conducting microarray analysis. This work was supported in part by new faculty start-up funds from the University of Toronto (Z.Z.) and grants from Genome Canada through the Ontario Genomics Institute.

References

- Baudot, A., Jacq, B., and 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. doi: 10.1186/gb-2004-5-10-r76.
- Blanc, G., Barakat, A., Guyot, R., Cooke, R., and Delseny, M. 2000. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**: 1093–1101.
- Brookfield, J. 1992. Can genes be truly redundant? *Curr. Biol.* **2**: 553–554.
- Cho, R.J., Campbell, M.J., Winzler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., et al. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**: 65–73.
- Christoffels, A., Koh, E.G., Chia, J.M., Brenner, S., Aparicio, S., and Venkatesh, B. 2004. *Fugu* genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* **21**: 1146–1151.
- Comant, G.C. and Wagner, A. 2004. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc. Biol. Sci.* **271**: 89–96.
- Davis, J.C. and Petrov, D.A. 2005. Do disparate mechanisms of duplication add similar genes to the genome? *Trends Genet.* **21**: 548–551.
- Dehal, P. and Boore, J.L. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**: e314. doi: 10.1371/journal.pbio.0030314.
- de Lichtenberg, U., Jensen, L.J., Brunak, S., and Bork, P. 2005. Dynamic complex formation during the yeast cell cycle. *Science* **307**: 724–727.
- Estruch, F. and Carlson, M. 1993. Two homologous zinc finger genes identified by multicopy suppression in a SNF1 protein kinase mutant of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**: 3872–3881.
- Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**: 4241–4257.
- Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., et al. 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**: 637–643.
- Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. 2003. Global analysis of protein expression in yeast. *Nature* **425**: 737–741.
- Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387–391.
- Greenbaum, D., Jansen, R., and Gerstein, M. 2002. Analysis of mRNA expression and protein abundance data: An approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics* **18**: 585–596.
- Griffioen, G., Swinnen, S., and Thevelein, J.M. 2003. Feedback

- inhibition on cell wall integrity signaling by Zds1 involves Gsk3 phosphorylation of a cAMP-dependent protein kinase regulatory subunit. *J. Biol. Chem.* **278**: 23460–23471.
- Gu, Z., Steinmetz, L.M., Gu, X., Scharfe, C., Davis, R.W., and Li, W.H. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**: 63–66.
- Guan, Y., Dunham, M.J., and Troyanskaya, O.G. 2007. Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* **175**: 933–943.
- Hakes, L., Pinney, J.W., Lovell, S.C., Oliver, S.G., and Robertson, D.L. 2007. All duplicates are not equal: The difference between small-scale and genome duplication. *Genome Biol.* **8**: R209. doi: 10.1186/gb-2007-8-10-r209.
- Harrison, R., Papp, B., Pal, C., Oliver, S.G., and Delneri, D. 2007. Plasticity of genetic interactions in metabolic networks of yeast. *Proc. Natl. Acad. Sci.* **104**: 2307–2312.
- He, X. and Zhang, J. 2006. Higher duplicability of less important genes in yeast genomes. *Mol. Biol. Evol.* **23**: 144–151.
- Ihmels, J., Collins, S.R., Schuldiner, M., Krogan, N.J., and Weissman, J.S. 2007. Backup without redundancy: Genetic interactions reveal the cost of duplicate gene loss. *Mol. Syst. Biol.* **3**: 86. doi: 10.1038/msb4100127.
- Jeffreys, A.J., Wilson, V., Wood, D., Simons, J.P., Kay, R.M., and Williams, J.G. 1980. Linkage of adult alpha- and beta-globin genes in *X. laevis* and gene duplication by tetraploidization. *Cell* **21**: 555–564.
- Kafri, R., Bar-Even, A., and Pilpel, Y. 2005. Transcription control reprogramming in genetic backup circuits. *Nat. Genet.* **37**: 295–299.
- Kafri, R., Dahan, O., Levy, J., and Pilpel, Y. 2008. Preferential protection of protein interaction network hubs in yeast: Evolved functionality of genetic redundancy. *Proc. Natl. Acad. Sci.* **105**: 1243–1248.
- Kellis, M., Birren, B.W., and Lander, E.S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617–624.
- Kirschner, M. and Gerhart, J. 1998. Evolvability. *Proc. Natl. Acad. Sci.* **95**: 8420–8427.
- Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**: 637–643.
- Kuepfer, L., Sauer, U., and Blank, L.M. 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* **15**: 1421–1430.
- Maere, S., Heymans, K., and Kuiper, M. 2005. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**: 3448–3449.
- Mayer, V.W. and Aguilera, A. 1990. High levels of chromosome instability in polyploids of *Saccharomyces cerevisiae*. *Mutat. Res.* **231**: 177–186.
- McAlister, L. and Holland, M.J. 1982. Targeted deletion of a yeast enolase structural gene. Identification and isolation of yeast enolase isozymes. *J. Biol. Chem.* **257**: 7181–7188.
- Mendoza, I., Quintero, F.J., Bressan, R.A., Hasegawa, P.M., and Pardo, J.M. 1996. Activated calcineurin confers high tolerance to ion stress and alters the budding pattern and cell morphology of yeast cells. *J. Biol. Chem.* **271**: 23061–23067.
- Mewes, H.W., Amid, C., Arnold, R., Frishman, D., Guldener, U., Mannhaupt, G., Munsterkötter, M., Pagel, P., Strack, N., Stumpflen, V., et al. 2004. MIPS: Analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32**: D41–D44.
- Musso, G., Zhang, Z., and Emili, A. 2007. Retention of protein complex membership by ancient duplicated gene products in budding yeast. *Trends Genet.* **23**: 266–269.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, Berlin.
- Pal, C., Papp, B., and Hurst, L.D. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* **158**: 927–931.
- Papp, B., Pal, C., and Hurst, L.D. 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**: 661–664.
- Pena-Castillo, L. and Hughes, T.R. 2007. Why are there still over 1000 uncharacterized yeast genes? *Genetics* **176**: 7–14.
- Pierce, S.E., Fung, E.L., Jaramillo, D.F., Chu, A.M., Davis, R.W., Nislow, C., and Giaever, G. 2006. A unique and universal molecular barcode array. *Nat. Methods* **3**: 601–603.
- Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B.J., Hon, G.C., Myers, C.L., Parsons, A., Friesen, H., Oughtred, R., Tong, A., et al. 2006. Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *J. Biol.* **5**: 11. doi: 10.1186/jbiol36.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**: 276–277.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–2504.
- Stucka, R., Dequin, S., Salmon, J.M., and Gancedo, C. 1991. DNA sequences in chromosomes II and VII code for pyruvate carboxylase isoenzymes in *Saccharomyces cerevisiae*: Analysis of pyruvate carboxylase-deficient strains. *Mol. Gen. Genet.* **229**: 307–315.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. 2003. The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* **4**: 41. doi: 10.1186/1471-2105-4-41.
- Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**: 2364–2368.
- Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. 2004. Global mapping of the yeast genetic interaction network. *Science* **303**: 808–813.
- Wang, C.W., Hamamoto, S., Orci, L., and Schekman, R. 2006. Exomer: A coat complex for transport of select membrane proteins from the trans-Golgi network to the plasma membrane in yeast. *J. Cell Biol.* **174**: 973–983.
- Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. 2007. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**: 54–61.
- Wolfe, K. 2000. Robustness—It's not where you think it is. *Nat. Genet.* **25**: 3–4.
- Wolfe, K.H. and Shields, D.C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708–713.
- Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- Yashiroda, H., Kaida, D., Toh-e, A., and Kikuchi, Y. 1998. The PY-motif of Bul1 protein is essential for growth of *Saccharomyces cerevisiae* under various stress conditions. *Gene* **225**: 39–46.

Received February 7, 2008; accepted in revised form April 22, 2008.