nature genetics

Exposing the fitness contribution of duplicated genes

Alexander DeLuna[1], Kalin Vetsigian[1], Noam Shoresh[1], Matthew Hegreness[1], Maritrini Colón-González[1], Sharon Chao[1] & Roy Kishony[1,2]

Duplicate genes from the whole-genome duplication (WGD) in yeast are often dispensable—removing one copy has little or no phenotypic consequence[1,2]. It is unknown, however, whether such dispensability reflects insignificance of the ancestral function or compensation from paralogs[3–7]. Here, using precise competition-based measurements of the fitness cost of single and double deletions, we estimate the exposed fitness contribution of WGD duplicate genes in metabolism and bound the importance of their ancestral pre-duplication function. We find that the functional overlap between paralogs sufficiently explains the apparent dispensability of individual WGD genes. Furthermore, the lower bound on the fitness value of the ancestral function, which is estimated by the degree of synergistic epistasis, is at least as large as the average fitness cost of deleting single non-WGD genes. These results suggest that most metabolic functions encoded by WGD genes are important today and were also important at the time of duplication.

Although several inferences of the extent and mechanisms of paralogous compensation have been suggested[1,3–7], direct measurements through double deletions of duplicate genes have not been done. The functional overlap between members of duplicate pairs manifests as deviations of the fitness cost of double deletions from the additive expectation based on the effects of single deletions[8–11] (**Fig. 1**). Measuring such synergistic epistasis between paralogs could therefore be used to estimate the fitness contribution of the overlapping function. Because this overlapping functionality masks the fitness consequence of deleting a single gene, we define the 'exposed' fitness contribution of a gene as the fitness effect of its deletion in the absence of its paralog. Further, by considering processes of functional gain and loss, the present-day synergy and the estimate of deleting both pair members can be used to estimate lower and upper bounds on the importance of the ancestral duplicated function.

To examine the phenotypic consequence of single and double deletions systematically, we developed a sensitive parallel technique for determining the fitness of *Saccharomyces cerevisiae* gene-deletion strains (**Fig. 2**). The method is based on a competition assay between strains labeled with two different variants of the green fluorescent protein—YFP (yellow fluorescent protein) and CFP (cyan fluorescent protein)—and provides precise measurements of relative fitness. Mutant and wild-type strains are grown in mixed cultures for ~50 generations through daily serial transfers, and the rate of change in genotype frequencies is estimated by bulk measurements of YFP and CFP fluorescence[12]. This rate of change gives the selection coefficient s, and so the fitness of the particular deletion strain, W, relative to the wild-type fitness, W_ϕ, is given by $W/W_\phi = 1-s$ (see **Supplementary Methods** online). Tag-swap experiments are done to control for variation in competition assays and in the construction of strains (**Supplementary Fig. 1** online). The sensitivity of our measurement allowed us to identify fitness effects of ~1%, which are too small to be detected by conventional measurements, but which are nonetheless evolutionary significant (large in comparison to the reciprocal of yeast effective population sizes)[13].

We concentrated our experimental analysis on genes related to metabolism, which provide a comprehensive model system for

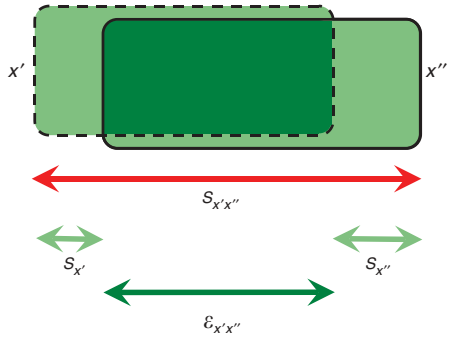

Figure 1 Measuring the extent of paralogous compensation. The rectangles symbolize the functions of two paralogous genes x' and x''. The fitness cost of a deletion is represented by the area of the corresponding function. When a member of the gene pair is deleted, the part of its function that is shared by the paralog is not lost, and the fitness cost ($s_{x'}$ or $s_{x''}$) reflects only the contribution of the non-overlapping part. When both genes are deleted, the functional loss is represented by the union of the rectangles, and the corresponding fitness cost, $s_{x'x''}$, is indicated. The overlapping area (dark green) is the source of synergistic epistasis ($\varepsilon_{x'x''}$) between the genes. The relation $\varepsilon_{x'x''} = s_{x'x''} - s_{x'} - s_{x''}$, implied by the figure, holds when the fitness effects are small.

[1]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. [2]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. Correspondence should be addressed to R.K. (roy_kishony@hms.harvard.edu).

Received 20 August 2007; accepted 27 February 2008; published online 13 April 2008; doi:10.1038/ng.123

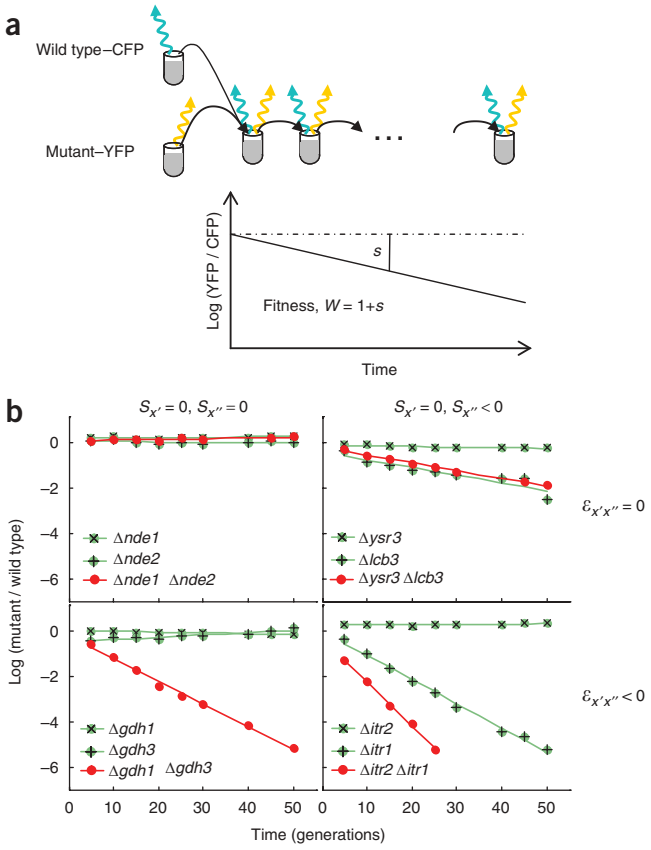

Figure 2 Competition-based high-resolution fitness profiling of single and double deletions. (**a**) Sensitive detection of the selection coefficient *s* of a given mutant relative to a neutral reference (wild type) is achieved by monitoring the rate of change of the population ratio of the two strains. Differential tagging of the two competing populations with CFP and YFP allows automated detection of the population ratio over several cycles of serial dilution. (**b**) Representative measurements of four WGD gene pairs, showing the ratios between the reference strain and each of the two single deletions ($\Delta x'$ and $\Delta x''$, green circles with cross and plus signs, respectively) as well as the double deletion ($\Delta x'\Delta x''$, red circles). Also shown is the linear least-squares fit of the log ratio of mutant to reference for each competition. The selection coefficient *s* is defined by the slopes of these fits. Synergistic epistasis whereby the double deletion has a higher fitness cost (more negative selection coefficient) than expected from the independent effects of the single deletions ($\varepsilon < 0$ in bottom panels) is indicative of paralogous compensation and functional overlap.

estimating the contribution of paralogs to mutational robustness[4,6,14,15]. We determined the fitness of most single deletions (557) out of the set of all metabolic genes (624). Included in this collection were deletions of 132 out of 144 genes that have duplicates originating from the WGD in yeast[16,17]. Combining deletions of these genes, we constructed and phenotyped 66 double deletions of paralogous pairs, as well as 298 strains in which the deleted WGD genes were paired at random. To increase the chance of revealing the phenotypic consequence of deletion of metabolic genes, we grew and assayed all strains in a nutrient-poor medium (see Methods). Our definition of fitness was restricted to this laboratory environment.

Although the fitness cost of deleting single WGD genes is typically small (in agreement with previous observations[1,2]), we found that the effect of deleting a paralogous pair is much larger than the added costs of the single deletions or the cost of deleting two unrelated WGD genes (**Fig. 3a,b**). The difference between the fitness of the double-deletion strain and the additive expectation yields the degree of epistasis, $\varepsilon_{x'x''} = W_{x'x''} - W_{x'}W_{x''}$ (refs. 9–11; see also Methods). The data showed a notably high occurrence of synergistic epistasis ($\varepsilon < 0$) between WGD genes and their paralogs: 55% of the pairs showed synergistic epistasis, which is much above the level suggested by genome-wide measurements of genetic interactions[8,9], or even that among genes with closely related functions[10,11]. In contrast, the data from our control set of randomly paired double deletions showed zero average epistasis (**Fig. 3c**), and their combined effect was fully predictable by adding their individual effects (agreement between the additive expectation and the measured random pairs in **Figure 3a**). These results confirm that the synergistic interactions that we found are not generic among the set of WGD metabolic genes; rather, the synergy is specific to paralogous pairing (**Fig. 3c**). Such synergy shows that the small fitness cost of deleting a WGD gene is only the tip of the iceberg: when a gene is deleted, most of the function

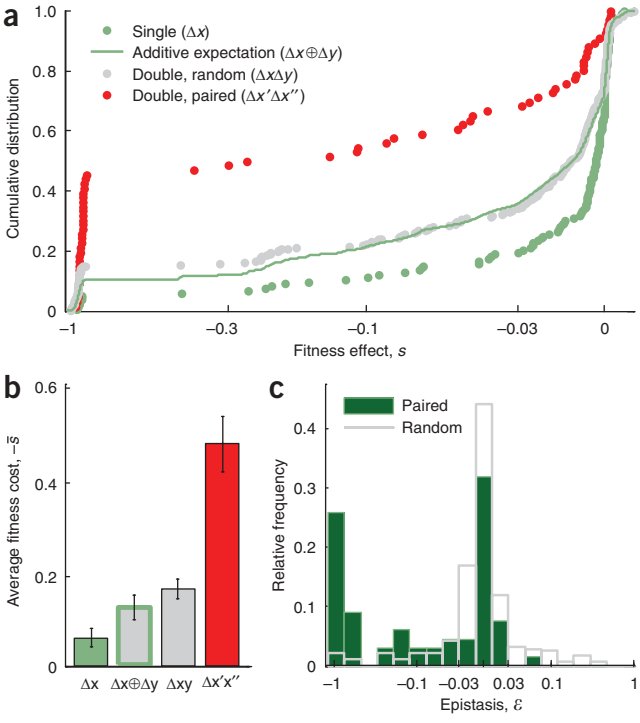

Figure 3 Dispensability of WGD genes is explained by paralogous compensation. (**a**,**b**) Cumulative distributions of fitness effects (**a**) and average fitness cost and standard error (**b**) of single deletions of WGD genes (light green circles, 'Single', $n = 132$); the combined additive effects of two random (nonparalogous) single deletions of WGD genes $s_{x\oplus y} \equiv (1+s_x)(1+s_y) - 1$ (green line, 'Additive expectation'); double deletions of randomly paired WGD genes (gray circles, 'Double, random', $n = 277$); and double deletions of paralogous pairs of WGD genes (red circles, 'Double, paired', $n = 66$). The distribution of fitness effects of the paired double deletions is significantly different from that of the additive expectation ($P < 10^{-6}$, Kolmogorov-Smirnov test). Data points at $s = -1$ have been slightly offset to distinguish between the overlaid curves. For better illustration of both strong and mild effects, the horizontal axis is shown on a log scale for $s < -0.03$. (**c**) Distribution of epistasis effects (ε) for paralogous pairs ($\Delta x'\Delta x''$, dark green line) and random pairs ($\Delta x\Delta y$, gray bars) of WGD genes.

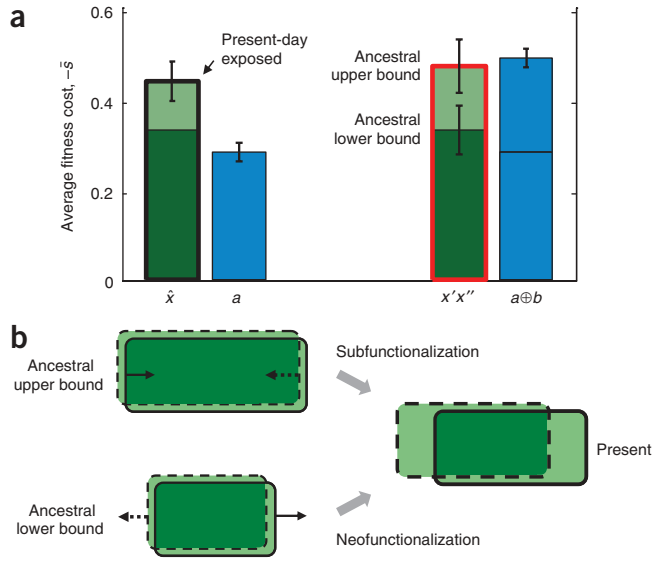

Figure 4 Exposed fitness of duplicated genes and bounds on fitness contribution of ancestral function. (**a**) Left: the average exposed fitness contribution of single WGD genes ($\hat{x}$, light and dark green bar, with the dark green area denoting only the contribution of the functional overlap, ε) is larger than the fitness cost of deleting an average non-WGD gene (*a*, blue bar). *t*-test: $P = 0.004$. Kolmogorov-Smirnov test for identical distributions: $P < 10^{-4}$; $\hat{x}$ is calculated as calculated as the average of $\hat{s}_{x'} = \frac{1+s_{x'x''}}{1+s_{x''}} - 1$. Right: the average cost of deleting a pair of WGD paralogs sets the upper bound on the ancestral function (total height of bar labeled *x′x″*), whereas the epistasis (dark green) sets the lower bound. The upper bound is close to the additive effect of two random non-WGD genes (blue bar, $a \oplus b$). Error bars, s.e.m. (**b**) The non-overlapping functionality of paralogous pairs is generated by two processes: subfunctionalization and neofunctionalization. If only the former takes place, each of the genes loses some of its original function, and the importance of the full original function can be seen by the effect of the double deletion ('Ancestral upper bound'). On the other hand, if new functions are the source of the differences between the pair members, the fitness contribution of the original function is equal to the functional overlap ('Ancestral lower bound').

it encodes is maintained by its paralog, and thus the large fitness cost of losing that function is hidden by paralogous compensation.

To isolate the fitness cost of losing the function encoded by a gene from the masking effects of paralogous compensation, we defined the exposed fitness of a gene as the fitness effect of its deletion in the absence of its paralog: $\hat{W}_{x'} \equiv W_{x'x''}/W_{x''}$. For small fitness effects, the exposed fitness of a gene approximately equals the sum of the cost of its deletion and the synergy with its paralog ($\hat{s}_{x'} = \hat{W}_{x'} - 1 \cong s_{x'} + \varepsilon_{x'x''}$). We found that the average exposed fitness contribution of single WGD genes is larger than that of non-WGD genes in metabolism (**Fig. 4a**, *t*-test, $P < 10^{-3}$; see **Supplementary Fig. 2** for cumulative distributions and statistical tests).

The starting point for every WGD gene pair is its ancestral form. Since the yeast whole-genome duplication, the fitness contribution of each gene has shifted as a result of neo- and subfunctionalization[18–20]. In neofunctionalization, new functions are acquired by either one of the pair members, increasing their combined fitness contribution. In subfunctionalization, on the other hand, partial loss of function by either one of the genes causes the fitness importance of the shared functions to shrink. It follows that the fitness contribution of the ancestral function is bounded from below by the inferred present-day fitness cost of deleting the shared functionality (as given by the epistasis) and from above by the cost of deleting both genes (**Fig. 4b**). These bounds, which provide an approximation of the fitness contribution of the ancestral function, depend on the implicit assumption that no functionality was jointly lost or jointly gained by the members of a paralogous pair. The lower and upper bounds that the data place on the fitness contribution of the pre-duplication genes are shown in **Figure 4a**. If neofunctionalization is the dominant process of functional divergence, then the ancestral function would be set at the lower bound, which we found to be roughly equal to the average fitness contribution of existing non-WGD metabolic genes. In contrast, if the main process is subfunctionalization, the ancestral function, approximated by the combined effect of the double deletion, would have a fitness contribution similar to the average additive effect of two random non-WGD genes (**Fig. 4a**, bar labeled $a \oplus b$). In either case, our bounds show that the ancestral metabolic function is at least as important as the function of a typical single non-WGD metabolic gene.

This conceptual framework can be used to contrast different functional categories and shed light on the factors determining the evolutionary fate of duplicate genes. We found that duplicate genes in the pathways for carbohydrate or amino acid metabolism participate in more crucial roles than their non-WGD counterparts, while WGD genes in lipid or nucleotide metabolism are as important as non-duplicate genes in the same categories (**Supplementary Fig. 3** online). Indeed, it has been suggested that genes with important glycolytic functions were preferentially retained after the WGD in yeast[21,22]. In addition, we compared the exposed fitness contribution of WGD genes that have other paralogs originating in smaller-scale duplication (SSD)[23] to those that have no duplicate besides their WGD paralog. SSD paralogs seem to have a much larger relative contribution to compensation of phenotypes for functions not encoded by WGD genes (**Supplementary Fig. 3**). The preferential enrichment for

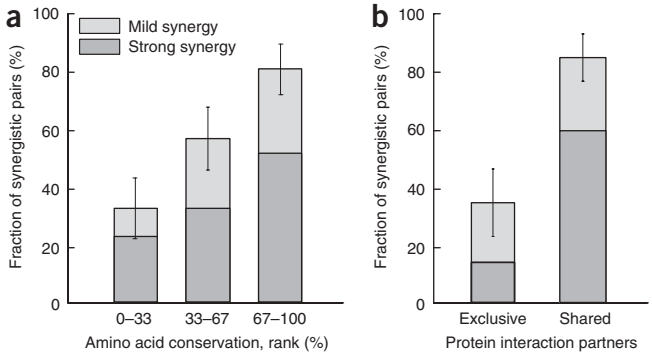

Figure 5 Synergistic epistasis between WGD pairs correlates with conservation of amino acid sequence and protein interaction partners. (**a**,**b**) Fraction of synergistic WGD pairs as a function of their amino acid conservation (**a**, rank of $1/K_A$) and divergence in their protein interaction pattern (**b**, 'exclusive', $n = 17$, no shared interaction partners; and 'shared', $n = 20$, with at least one shared interaction partner). K_A data are from the Yeast Gene Order Browser (see URLs section in Methods). Analysis in **b** is restricted to pairs that interact with four or more proteins. Data are from the Database of Interacting Proteins (see URLs section in Methods). The analysis was repeated using the literature-curated protein interaction data (see URLs section in Methods); the correlation between synergy and protein interaction divergence is also significant in this database ($P = 0.041$, Probit regression). Synergistic pairs are separated based on the strength of the epistatic interaction (mild synergy, light gray bars, $-0.5 \leq \varepsilon < -0.03$; strong synergy, dark gray bars, $\varepsilon < -0.5$). Error bars, s.e.m.

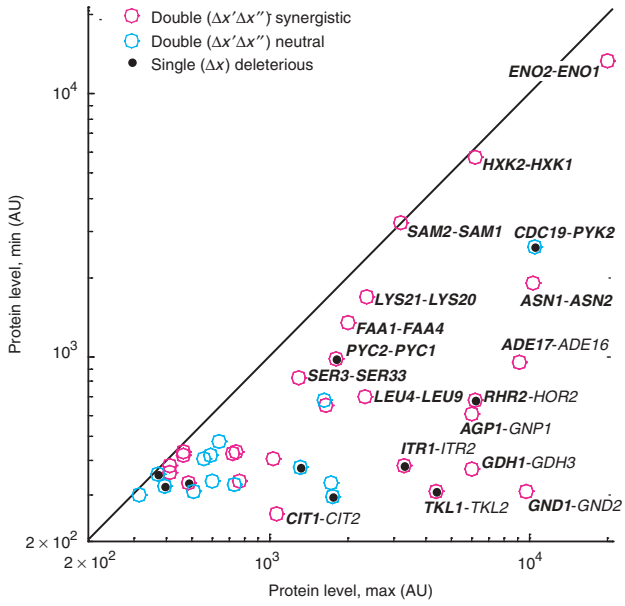

Figure 6 High occurrence of synergy between WGD gene pairs with at least one highly expressed member. Each circle represents the protein abundance of a duplicate pair measured as average fluorescence signal of GFP fusion to the individual pair members. For each pair, the higher signal is shown on the *x* axis. Phenotypic deletion data are noted for gene pairs with an indispensable member (black dot, single deleterious, $s < -0.03$), and for level of epistasis (synergy, magenta circles, $\varepsilon < -0.03$; and no epistasis, cyan circles, $\varepsilon \geq -0.03$). Gene names are shown for gene pairs with at least one highly expressed member (signal $> 10^3$); boldface for both genes indicates correlation in high expression, whereas boldface–plain font indicates differential expression.

neutrality in non-WGD genes with SSD paralogs, but not in WGD pairs with SSD paralogs, seems to explain much of the difference between the exposed fitness value of WGD genes in metabolism and the fitness contribution of their non-WGD counterparts.

To explore the basis of paralogous compensation, we considered the relation among synergy, sequence and function. If synergistic epistasis between paralogs reflects direct genetic compensation through shared biochemical functions, then it should be correlated with amino acid divergence[1] as well as with functional attributes such as protein–protein interactions[24]. Indeed, we found a significant correlation between the frequency of synergistic WGD pairs and the conservation of their amino acid sequences (**Fig. 5a**; $P < 10^{-3}$, Probit regression). Similarly, we observed an enrichment of synergy between WGD gene pairs sharing protein–protein interaction partners (**Fig. 5b**; $P = 0.002$, Probit regression). We note that the correlations of synergy with amino acid and protein interaction divergence are not strictly independent, as both these attributes are themselves correlated[25]; nevertheless, the correlation between epistasis and protein interaction divergence remained significant after subtraction of the effects of amino acid divergence ($P = 0.009$, partial correlation). Many of the duplicate genes we addressed encode isoenzymes catalyzing the same enzymatic reaction[26]; the synergy we measured merely reflects the lack of alternative genetic compensation for these reactions[6]. Nevertheless, paralogous compensation is maintained even among several WGD genes encoding enzymes that have significantly diverged in amino acid sequence (**Fig. 5a**) or that are not known to share common protein interactions (**Fig. 5b**). These phenotypic observations are consistent with the notion that duplicate genes may be only partially redundant and yet maintain the ability to compensate for each other's loss[27,28].

To relate the strength of synergy between paralogs to their activity in our growth environment, we measured expression for 99 WGD genes (see Methods). Using a library of yeast strains with proteins fused to the green fluorescent protein (GFP)[29], we measured the protein abundances of WGD genes in wild-type cells growing in our competition media (**Fig. 6** and **Supplementary Fig. 4** online). We observed strong correlation between expression and synergy ($P = 0.002$, Fisher's exact test)—virtually all gene pairs for which at least one of the proteins was present at high levels were synergistic (**Fig. 6**, magenta-labeled pairs). High expression of the two members of a duplicate pair could suggest dosage amplification as the factor determining their retention[2,4,21,22]. Indeed, several pairs of duplicates showed high expression in our study and are strongly deleterious when doubly deleted (**Fig. 6**, bold-bold labels). Note, however, that only one gene (*PYC1*) among these highly expressed pairs has a significant fitness effect when singly deleted. In addition, we observed synergistic epistasis within gene pairs with strong differential expression (**Fig. 6**, bold-unbold labels). For example, gene pairs like *GDH1-GDH3* interact synergistically (**Fig. 2b**), despite their different regulation and the unique biochemical properties of the glutamate dehydrogenase isoenzymes they encode[30].

The indispensability of functions encoded by duplicate genes in metabolism is obscured by the functional compensation that has persisted between WGD paralogs through 100 million years of evolution. Despite the small fitness cost associated with the deletion of individual WGD genes, their exposed fitness contribution is, on average, larger than (or, in certain gene subsets, at least equal to) the effect of single deletion of non-WGD metabolic genes. In addition, we have provided a novel method for inferring the fitness contribution of the ancestral, pre-duplication function by using the intersection and union of present-day functions of pair members as lower and upper bounds. These measurements address the question of whether less important genes are more likely to retain their duplicates, demonstrating that, at least as far as WGD gene pairs in metabolism are concerned, the average fitness contribution of an ancestral gene is no less than that of a non-WGD gene. Our method could also be applied to directly estimate the ancestral importance of other sets of paralogs[7]; in particular, different results could be expected for paralogs retained after smaller-scale duplication events[23]. Although the concepts presented here do not presume to capture the full complexity of biological function, regulatory feedback mechanisms or selection in a changing environment, we have shown that they can lead to useful insights. The exposed fitness and the ancestral bounds were obtained separately for each pair and can be used for elucidating the importance and evolutionary history of individual genes as well as for comparing and contrasting duplicates in different classes of genes. Using this approach, we find substantial differences in the functional fate of paralog pairs among different metabolic pathways. These methods are general and could be applied to additional gene sets, environments and species. The ability to estimate the exposed fitness and bound the importance of the ancestral function of duplicated genes may be a key to understanding the evolutionary fate and adaptive role of genetic redundancy.

METHODS

Strains, plasmids and media. We transformed yeast strain Y7092 (*MATα can1Δ::STE2pr-SP_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*) by PCR-based gene replacement of the neutral *HO* locus with the YFP (yEVenus) or CFP

(yECerulean) *natMX4* modules conferring resistance to nourseothricin; constitutive expression of fluorescent proteins is driven by the *TDH3* promoter. We used the resulting strains, Y7092-YFP and Y7092-CFP, for large-scale tagging of deletion strains. Competition medium was minimal low-fluorescence synthetic medium with 2% glucose, 2% proline as a nitrogen source, and methionine (25 mg/l), leucine (120 mg/l) and uracil (20 mg/l) supplements.

Large-scale tagging of single-deletion strains. Starter strains Y7092-YFP and Y7092-CFP were mated to 521 strains from the BY4741 yeast deletion collection (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* xxxΔ::*kanR*) using synthetic genetic array (SGA) technology[8] (200 μg/ml geneticin, 100 μg/ml nourseothricin; colony arrays were transferred manually with a 384-head pin tool, V&P Scientific). We successfully tagged 454 (87%) of the strains. The average tagging efficiencies of single deletions of WGD and nonduplicate genes were the same, despite the lower average fitness of the former; suggesting that the strains missing in our single-deletion collection were not significantly biased to slow-growing strains. The collection of single deletions is described in **Supplementary Table 1** online.

Construction of double-deletion strains and synthetic lethal screen. We used starter Y7092-YFP and Y7092-CFP strains to construct two sets (YFP and CFP) of WGD-deletion starter strains by direct PCR-based replacement, using the *CaURA3-MX4* cassette, of one gene of each of the 72 WGD pairs. In addition, 72 target WGD-deletion strains, paralogs of the deletion starter strains, were generated *de novo* by direct transformation of BY4741 with the *kanMX4* cassette (pairs of WGD taken from a previous study[17]). The deletion starter array was mated to the target array using SGA[8]; double deletions were selected in synthetic complete medium minus uracil plus geneticin and nourseothricin. We constructed isogenic single-deletion references by mating the starter and the target arrays to neutral *his3*Δ::*kanR* and *his3*Δ::*URA3* deletion strains. To allow for generation of a nonconstrained set of random double knockouts, we expanded the target array to include the whole set of 144 WGD deletions. Each YFP-tagged starter strain was then mated with four to six random strains (obtained by random permutation in Matlab) from the expanded target array. The two collections of double mutants (paralogous and randomly paired) are described in **Supplementary Table 1**.

We scored the synthetic lethal pairs of deletions, taking advantage of the fluorescence-labeled strains (**Supplementary Fig. 5** online). A synthetic lethal interaction was scored for each double mutant failing to grow in at least six out of eight crosses per double-deletion strain (four YFP and four CFP replicates).

For quality control, we verified all parental deletion strains by PCR using marker and upstream primers. Further, a self cross was carried out as positive synthetic lethal control for the complete set of query genes. Genes that failed these tests were discarded from further analysis, leaving a total of 66 gene pairs.

Fitness measurements by automated competition experiments. YFP- or CFP-tagged mutants were competed against common reference strains: Y7092-CFP *his3*Δ::*kanR* or Y7092-YFP *his3*Δ::*kanR*, respectively. CFP and YFP swap experiments were done in parallel to account for variation due to strain-tagging procedures and measurement errors (**Supplementary Fig. 1**). Overnight cultures of mutant and reference strains were mixed 1:1 and inoculated into 150 μl of competition medium in 96-well microtiter plates (Costar). We diluted the competition cultures 32-fold into sterile fresh medium every 36 h; doubling time of the reference strain was 3.5 h. Cultures were propagated in parallel for approximately 50 generations (15 d) in a fully automated robotic system (Staccato Sciclone Cell Station, Caliper LifeSciences) consisting of a microplate shaker (Liconic), a liquid handler (Sciclone) and a plate reader (Victor3). We monitored bulk fluorescence signal and absorbance at 600 nm (OD_{600}) every ~2.5 h (Chroma filters HQ500/20x and HQ535/30m for YFP; D430/20x and D486/30m for CFP). The equipment was maintained in an environmental room at constant temperature (30 °C) and relative humidity (70%). The system was controlled by a Matlab code to determine and keep track of reading and dilution times, reading output, dilution volumes and plate identities. We carried out a total of 1,540 competition cultures, consisting of replicate experiments of deletions of 454 single genes, 66 pairs of WGD genes, 132 WGD single genes combined with a neutral insertion, and single experiments for 277 randomly paired deletions (**Supplementary Fig. 6** online). Synthetic-lethal double deletions in our screen, auxotroph strains that failed to grow in competition medium and mutants whose frequency dropped below the detection limit after one competition cycle were defined to have a fitness cost $s = -1$. For all other strains, the YFP, CFP and OD measurements were used to determine the selective coefficient (s) and epistasis level ($\varepsilon = W_{x'x''} - W_{x'}W_{x''}$); see detailed data analysis in **Supplementary Methods**. Average selection coefficients and standard errors for single- and double-deletion mutants are provided in **Supplementary Table 1**.

Flow cytometry. To estimate the protein abundances of WGD genes during exponential growth in the competition medium, we used a library of GFP(S65T)-His3MX–tagged strains[29]. A collection of 99 wild-type strains with different GFP fusions on WGD genes were grown to OD 0.4–0.45 and delivered to an analytical cytometer using an autosampler device (LSR-II, BD). Reported GFP signals are the average fluorescence of each strain. Raw measurements are shown in **Supplementary Figure 4**.

URLs. Yeast Gene Order Browser, http://wolfe.gen.tcd.ie/ygob/; Database of Interacting Proteins, http://dip.doe-mbi.ucla.edu/; literature-curated protein interaction data, http://www.thebiogrid.org/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank E. Angelino, N.Q. Balaban, N. Barkai, R. Chait, M. Elowitz, M. Ernebjerg, R. Kafri, E. Mancera, R. Milo, M. Vibranovski, D.R. Wagner, I. Wapinski, R. Ward and P. Yeh for helpful discussions and critical reading of the manuscript, M. Springer for help with flow cytometry, and C. Boone (University of Toronto) and J.S. Weissman (University of California San Francisco) for their gifts of strains and plasmids. This work was supported by a grant from the Human Frontiers Science Program to R.K.; A.D. is a Fellow of the Pew Program in the Biomedical Sciences.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Gu, Z.L. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
2. Seoighe, C. & Wolfe, K.H. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2**, 548–554 (1999).
3. Wagner, A. Robustness against mutations in genetic networks of yeast. *Nat. Genet.* **24**, 355–361 (2000).
4. Papp, B., Pal, C. & Hurst, L.D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661–664 (2004).
5. Kafri, R., Bar-Even, A. & Pilpel, Y. Transcription control reprogramming in genetic backup circuits. *Nat. Genet.* **37**, 295–299 (2005).
6. Kuepfer, L., Sauer, U. & Blank, L.M. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* **15**, 1421–1430 (2005).
7. He, X.L. & Zhang, J.Z. Higher duplicability of less important genes in yeast genomes. *Mol. Biol. Evol.* **23**, 144–151 (2006).
8. Tong, A.H.Y. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).
9. Segre, D., DeLuna, A., Church, G.M. & Kishony, R. Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77–83 (2005).
10. Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
11. St Onge, R.P. *et al.* Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* **39**, 199–206 (2007).
12. Hegreness, M., Shoresh, N., Hartl, D. & Kishony, R. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* **311**, 1615–1617 (2006).
13. Thatcher, J.W., Shaw, J.M. & Dickinson, W.J. Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl. Acad. Sci. USA* **95**, 253–257 (1998).
14. Deutscher, D., Meilijson, I., Kupiec, M. & Ruppin, E. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat. Genet.* **38**, 993–998 (2006).
15. Harrison, R., Papp, B., Pál, C., Oliver, S.G. & Delneri, D. Plasticity of genetic interactions in metabolic networks of yeast. *Proc. Natl. Acad. Sci. USA* **104**, 2307–2312 (2007).
16. Wolfe, K.H. & Shields, D.C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
17. Kellis, M., Birren, B.W. & Lander, E.S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
18. Lynch, M. & Force, A. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473 (2000).

19. He, X. & Zhang, J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164 (2005).
20. Tirosh, I. & Barkai, N. Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. *Genome Biol.* **8**, R50 (2007).
21. Vitkup, D., Kharchenko, P. & Wagner, A. Influence of metabolic network structure and function on enzyme evolution. *Genome Biol.* **7**, R39 (2006).
22. Conant, G.C. & Wolfe, K.H. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol. Systems Biol.* **3**, 129 (2007).
23. Guan, Y., Dunham, M.J. & Troyanskaya, O.G. Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* **175**, 933–943(2007).
24. Presser, A., Elowitz, M.B., Kellis, M. & Kishony, R. The evolutionary dynamics of the *Saccharomyces cerevisiae* protein interaction network after duplication. *Proc. Natl. Acad. Sci. USA* **105**, 950–954 (2008).
25. Makino, T., Suzuki, Y. & Gojobori, T. Differential evolutionary rates of duplicated genes in protein interaction network. *Gene* **385**, 57–63 (2006).
26. Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54–61 (2007).
27. Kafri, R., Levy, M. & Pilpel, Y. The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc. Natl. Acad. Sci. USA* **103**, 11653–11658 (2006).
28. Ihmels, J., Collins, S.R., Schuldiner, M., Krogan, N.J. & Weissman, J.S. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Molecular Systems Biology* **3**, 86 (2007).
29. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
30. DeLuna, A., Avendano, A., Riego, L. & Gonzalez, A. NADP-glutamate dehydrogenase isoenzymes of *Saccharomyces cerevisiae*—purification, kinetic properties, and physiological roles. *J. Biol. Chem.* **276**, 43775–43783 (2001).