REPORT

GENE EVOLUTION

Gene duplication can impart fragility, not robustness, in the yeast protein interaction network

Guillaume Diss,[1,2,3,4,5] Isabelle Gagnon-Arsenault,[1,2,3] Anne-Marie Dion-Coté,[1,2,3]* Hélène Vignaud,[1,2,3] Diana I. Ascencio,[1,2,3,6] Caroline M. Berger,[1,2,3] Christian R. Landry[1,2,3]†

The maintenance of duplicated genes is thought to protect cells from genetic perturbations, but the molecular basis of this robustness is largely unknown. By measuring the interaction of yeast proteins with their partners in wild-type cells and in cells lacking a paralog, we found that 22 out of 56 paralog pairs compensate for the lost interactions. An equivalent number of pairs exhibit the opposite behavior and require each other's presence for maintaining their interactions. These dependent paralogs generally interact physically, regulate each other's abundance, and derive from ancestral self-interacting proteins. This reveals that gene duplication may actually increase mutational fragility instead of robustness in a large number of cases.

Models of duplicated gene evolution posit that two paralogs avoid pseudogenization through changes in dosage effects, the partition of the ancestral functions, the evolution of new functions in one or both copies, or a combination of these changes (*1*–*3*). Over time, duplicate genes may diverge in sequence and regulation and become functionally independent (Fig. 1A, Fate 1). However, some duplicate pairs appear to maintain a functional overlap over macro-evolutionary time scales (*4*), which allows a paralog to compensate for any loss of function of its cognate copy, contributing to mutational robustness by buffering deleterious mutations (*5*–*7*) (Fig. 1A, Fate 2). There are also reports of functionally dependent paralogs (*8*, *9*) that may interfere with each other's evolutionary trajectories (Fig. 1A, Fate 3). Little is known regarding the origin, relative occurrence, and mechanistic bases of compensation and dependency and the extent to which they shape cellular networks.

We studied 56 pairs of paralogous proteins in yeast that are amenable to the study of protein-protein interactions (PPI) in vivo in different genetic backgrounds (fig. S1) (*10*). These duplicates share only homology with each other, cover a wide range of biological functions, and are either small-scale duplicates (SSDs) or ohnologs [from a whole-genome duplication (WGD)] (table S1). They therefore represent the diversity of paralogs found in yeast and other eukaryotic genomes. We mapped the PPI network of each of the 112 proteins in the wild-type (WT) background and in the deletion background of their cognate copy (Fig. 1B), for a total of 5688 highly replicated unique comparisons (figs. S1 and S2 and tables S2 and S3) (*10*). The assay performed is a protein-fragment complementation assay (PCA) in which proteins are fused to complementary moieties of the dihydrofolate reductase (DHFR) enzyme and expressed at endogenous levels in living cells. Interactions between two fusion proteins lead to DHFR complementation and colony growth on a restrictive medium (table S2 and fig. S1) (*10*, *11*). Because growth rate reflects the quantity of protein complex formed, this assay can be used to measure changes in PPI intensities in different genetic backgrounds (*12*, *13*).

Interaction scores derived from colony sizes (fig. S2) were correlated between the WT and deletion backgrounds (Fig. 1C), suggesting that the global PPI network is maintained in deletion strains. Nonetheless, we observed net and significant positive (increased PPIs) and negative perturbations (decreased PPIs) in response to the deletion of a paralog, representing, respectively, instances of PPI compensation and dependency (Fig. 1, B and C, and fig. S3). Cases of dependency appear to be as widespread as compensation (137 and 91 PPIs, involving 19 and 22 pairs out of the 56 tested, respectively). However, compensation and dependency rarely co-occur within a given pair (Fig. 1D), suggesting that the deletion of a paralog in a pair affects the other duplicate directly. Most pairs showed an asymmetric response to paralog deletion, with only one paralog compensating or being dependent on the other (19 of 22 and 14 of 19), indicating that one duplicate has a function in the PPI network that cannot be compensated by or that is not dependent on the other copy.

One of the mechanisms of compensation by paralogs is the increased expression of the remaining copy (*7*, *14*). Examining transcript abundance in yeast deletion strains (*15*), we did not find any compensating pairs with a significant change in mRNA expression in the paralog deletion background (fig. S4). We noted marginally significant changes in three cases when measuring protein levels and only one significant, but low magnitude, change (Fig. 2A). This finding agrees with previous observations that paralog up-regulation is not a general mechanism by which functional compensation takes place (*6*, *14*).

In asymmetric pairs, the compensating paralog generally formed a weaker (or no) PPI in the WT background (fig. S5A). In addition, expression profiles of the two duplicates across environments and genetic backgrounds tend to be more correlated in compensating pairs than in non-compensating ones (fig. S5B). The two duplicates could therefore interact with some of the same partners, but PPIs with one would predominate over the other. This could be achieved through either of the two types of competitive binding parameters—i.e., higher abundance or higher affinity (Fig. 2B, center). Accordingly, compensation would result from a shift of the binding equilibrium upon deletion of one paralog without requiring a change in expression (Fig. 2B, left). We verified this hypothesis by overexpressing the compensating duplicate and measuring the interactions mediated by its sister copy (Fig. 2B, right). Testing also the overexpression on unrelated interactions allowed us to control for effects of overexpression on growth (*10*).

Assessing this model for five pairs that show global compensation identified 16 and 41 out of 79 PPIs between a paralog and its partners (using low-copy- and high-copy-number overexpression plasmids, respectively) that were decreased to a larger extent than a control interaction of similar strength [adjusted P value (P_{adj}) < 0.05, analysis of variance (ANOVA)] (table S4). For instance, the interaction of Sna4p with Yeh1p is decreased upon overexpression of *PMP3*, whereas the interaction of Yap1802p with Vma8p is not affected by the overexpression of *YAP1801* (Fig. 2, C and D). Overall, compensated PPIs were significantly more likely to be decreased by the overexpression of the compensating paralog (Fig. 2E and fig. S5C). This result shows that a shift in the binding equilibrium is one of the molecular mechanisms through which one paralog can compensate for the deletion of its cognate copy. Because we did not observe a significant difference in the protein abundance of the paralogs within pairs in WT cells (fig. S5D), the combined results suggest that, in most cases, mutual exclusion could be mainly driven by differential affinity.

[1]Département de Biologie, Université Laval, Québec, QC, Canada. [2]The Quebec Network for Research on Protein Function, Engineering, and Applications, Université Laval, Québec, QC, Canada. [3]Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, QC, Canada. [4]EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Doctor Aiguader 88, 08003 Barcelona, Spain. [5]Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. [6]Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, Mexico.
*Present address: Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853-2703, USA. †Corresponding author. Email: christian.landry@bio.ulaval.ca

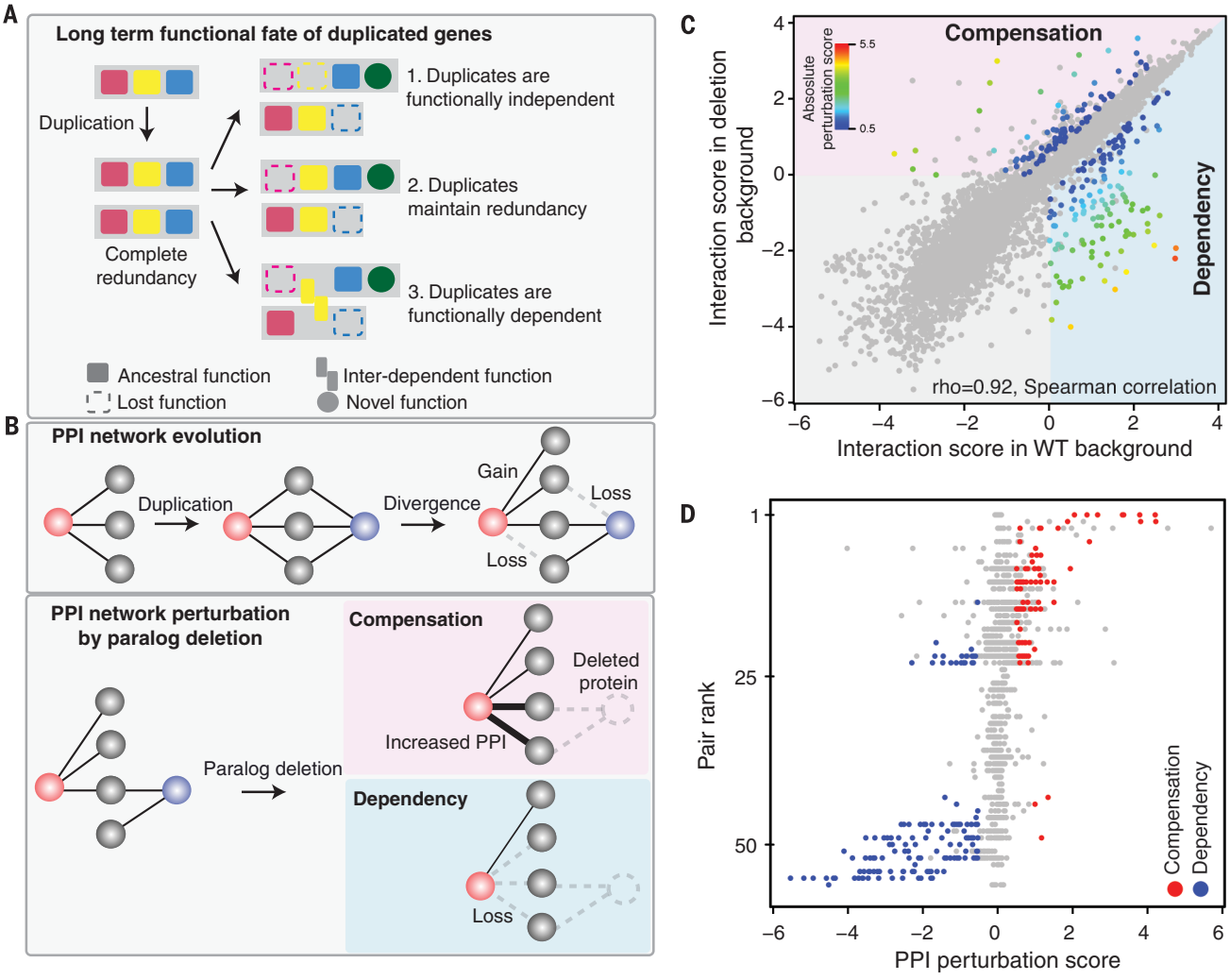

Fig. 1. Functional fates of paralogous genes and their response to each other's deletion on the PPI network. (**A**) Paralogs diverge until they become functionally unrelated (Fate 1), but they can maintain functional overlap (Fate 2) or remain functionally interdependent (Fate 3). (**B**) Paralogs gain and lose PPIs after divergence. Upon deletion of a paralog, PPIs are compensated by the remaining paralog. PPIs can also be lost, indicating functional dependency between paralogs. (**C**) Protein complex formation in WT (*x* axis) and paralog deletion strains (*y* axis). The color scale represents perturbation scores when significant (Student's *t* test, P_{adj} < 0.01; absolute perturbation score > 0.5; and interaction score > 0 in at least one of the two backgrounds). (**D**) Paralog pairs tend to exhibit either compensation or dependency but rarely both. Dots on the same horizontal axis represent all interactions mediated by either member of the corresponding pair of paralogs.

We next examined whether physical interactions between paralogs (*8*, *9*) could contribute to dependency by stabilizing or regulating the partners. Supporting this model, we noticed that dependent pairs are enriched for paralogous heteromers (paralogs that physically interact) (Fig. 3A). In asymmetrical pairs, the dependent duplicate generally has lower protein abundance than its cognate copy (Fig. 3B), which further decreases upon deletion of the latter (Fig. 3C). These protein changes are unlikely to derive from changes in mRNA levels (fig. S6). We also found that the deletion of the independent duplicate impairs fitness more strongly than that of the dependent one (fig. S7A). These results are consistent with a model of dependency through protein stabilization of one paralog upon interaction with its cognate copy.

Exploring a larger set of paralogs, we found that across all paralogs in budding yeast, the deletion of heteromer-forming SSDs reduces fitness more than that of singletons (single-copy genes), whereas ohnolog deletion is slightly less deleterious than that of singletons (Fig. 3D). This discrepancy could be the consequence of SSDs having more divergent functions than ohnologs (*16*). Overall, for both SSDs and ohnologs, the deletion of heteromeric paralogs has a larger cost than that of nonheteromeric paralogs (fig. S7B). We found the same trend in four human cell lines (fig. S7, C to F), suggesting that the increased fragility related to heteromeric paralogs is not limited to yeast. Paralogous heteromers could therefore be working as functional units. Supporting this, we found that they are more similar in amino acid sequences—which suggests that they could be younger—and are more likely to share molecular functions, as shown by their similarities in PPI and genetic interaction profiles, and gene ontology annotations (fig. S8, A to D). Paralogous heteromers work as functional units in several biological processes, including protein stability and proteolysis (fig. S8E).

Paralogous heteromers are generally frequent in eukaryotic PPI networks, representing between 6 and 27% of all paralogous pairs (fig. S9A). One way by which duplicates could evolve into a heteromer is by the duplication of an ancestral self-interacting protein (homomer) (Fig. 4A) (*17*). We compared ohnologs in budding yeast with their orthologs from *Schizosaccharomyces pombe* to infer the history of duplication and use PPI data for all ohnologs (Fig. 4B). Ohnologs that form paralogous heteromers are more likely to have a homomeric ortholog than ohnologs in general (Fig. 4C). This observation, coupled with our data indicating dependency of paralogs, suggests that some pairs of paralogs exhibit dependency as a consequence of the homomeric status of the ancestral protein. Additional analyses of the

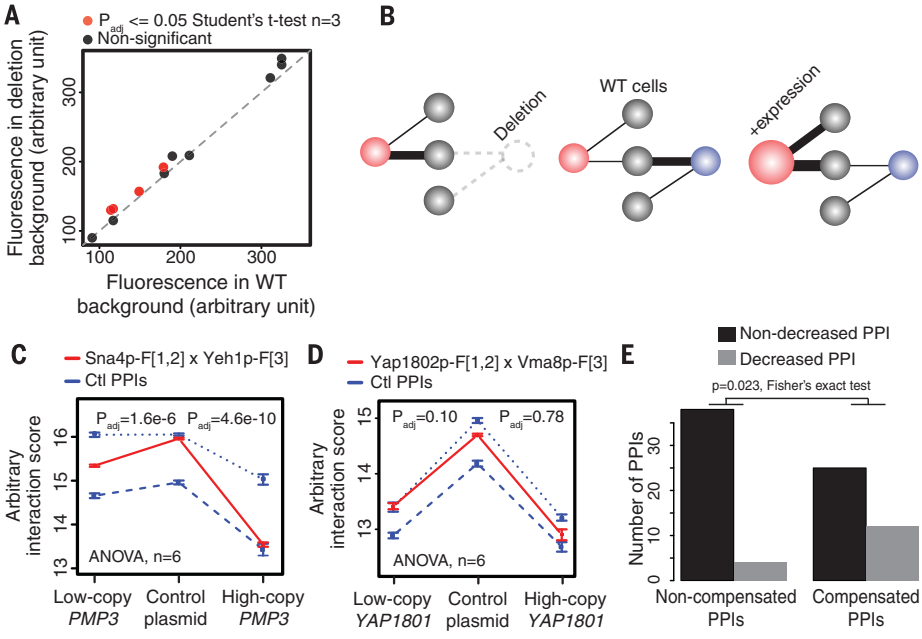

Fig. 2. Compensation may result from the removal of the physical exclusion by paralogs. (**A**) Protein abundance as measured by flow cytometry of green fluorescent protein (GFP)–tagged proteins in paralog deletion strains. Significant changes occur for one protein (*t* test, P_{adj} = 0.02), and three others are marginally significant (*t* test, P_{adj} = 0.05). (**B**) Compensation may result from mutual exclusion in WT cells caused by differential affinities or abundance (middle). Deletion of one paralog (blue) alleviates the exclusion, resulting in functional compensation (left). If they exclude each other, overexpression of the compensating paralog (red) should decrease the interaction involving the other paralog (blue, right). (**C** and **D**) The balance of protein interactions is shifted by overexpressing paralogs on a low- or high-copy-number plasmid, showing that paralogs can mutually exclude each other when interacting with their partners. Examples of affected (C) and unaffected (D) PPIs. F[1,2] and F[3] relate to gene fusion with DHFR F[1,2] and DHFR F[3], respectively. Two control interactions are shown in each case (blue). (**E**) Compensated PPIs are more likely to be decreased by overexpression of the compensating paralog (low-copy-number plasmid) than noncompensated ones.

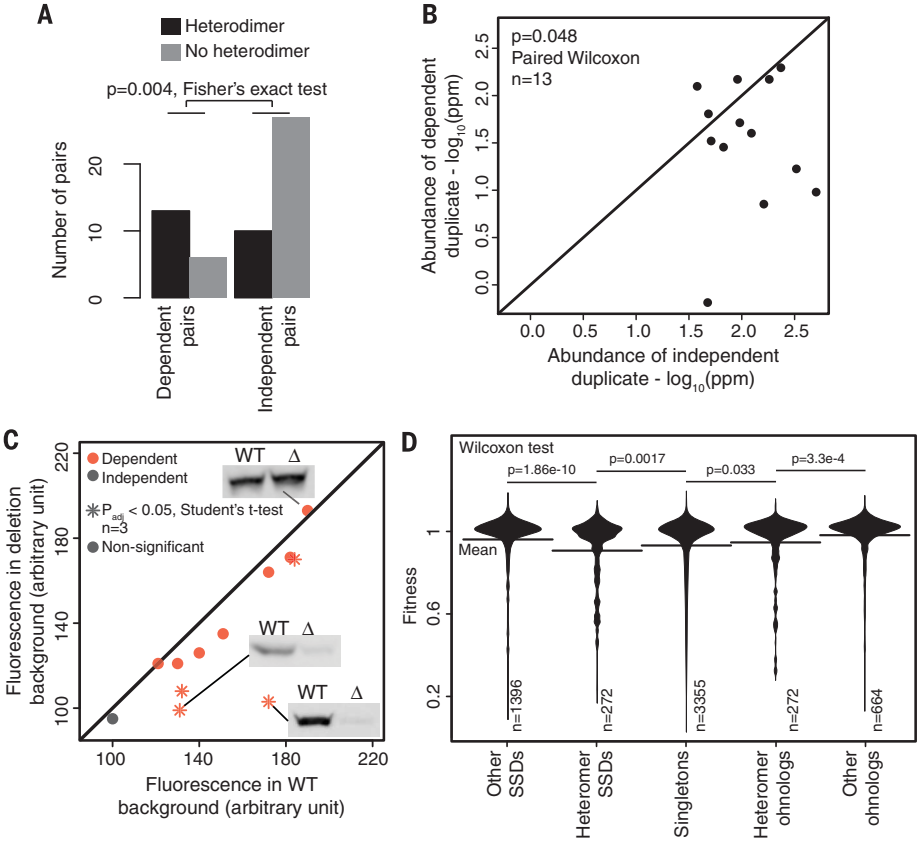

Fig. 3. Paralog dependency may be the consequence of the destabilization of one paralog after deletion of its sister copy. (**A**) Dependent paralog pairs are enriched for heteromers (interactions between paralogs). (**B**) In asymmetrically dependent pairs, the independent paralog is generally the most abundant of the pair. (**C**) Abundance of paralogs in the WT background (*x* axis) and after deletion of its sister copy (*y* axis), measured by flow cytometry of GFP–tagged proteins (scatter plot) or by Western blot (inserts). Shapes indicate statistical significance and colors whether paralogs are dependent or not. (**D**) Distribution of relative fitness for different classes of gene deletions.

Arabidopsis and human PPI networks show similar patterns (fig. S9, B and C), suggesting that the duplication of homomers is a general mechanism for the evolution of paralogous heteromers. The functional dependency of heteromers could make duplicates originating from a homomeric ancestor more likely to be maintained as a pair, because the loss of either copy could be functionally equivalent to losing both copies. The fact that ohnologs whose ancestral ortholog formed homodimers have indeed been retained in a greater number of post-WGD species

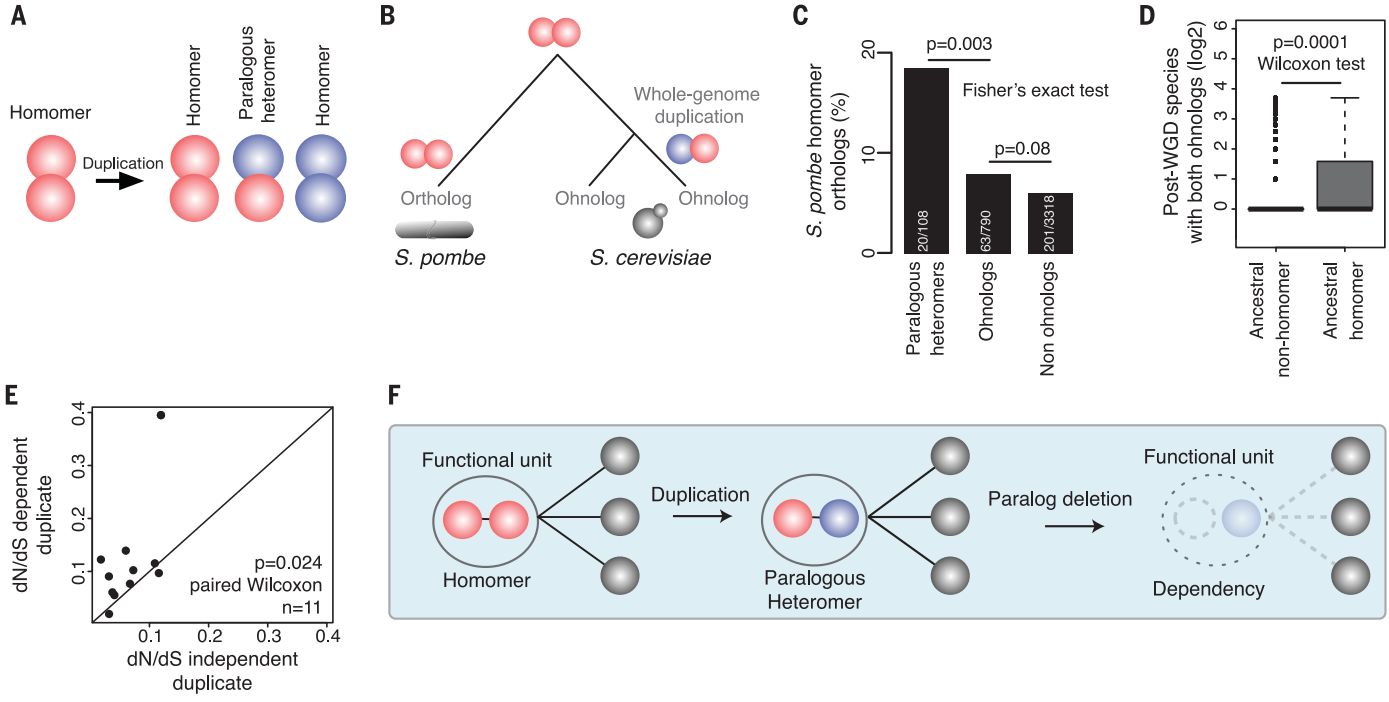

Fig. 4. Paralogous heteromers mostly evolve by the duplication of homomers, leading to functional dependency. (**A**) Homomers give rise to three potential homomers and heteromers after duplication. (**B**) The ancestral state of ohnolog heteromers can be inferred from PPI data from a species that diverged before the budding yeast whole-genome duplication. (**C**) Paralogous heteromers are more likely to derive from orthologs that form homomers. (**D**) Proteins that form homomers in *S. pombe* were maintained as a pair in a larger number of species after the WGD. (**E**) Dependent duplicates accumulate more nonsynonymous substitutions (dN) relatively to synonymous substitutions (dS), compared with the independent duplicate among species. (**F**) Homomers lead to heteromers of paralogs that form a functional unit and are thus dependent. The deletion of one affects the function, and thus the PPIs, of the other, increasing fragility rather than robustness.

supports this model (Fig. 4D), although direct selection on increased dimer dosage could also contribute (*18*). Dependency also has consequences on the rate of evolution. The dependent duplicate generally evolves faster among species than its stabilizing sister copy (Fig. 4E). This could be the consequence of stabilization by the duplicate that would buffer the effect of mutations and hence facilitate the exploration of sequence space, leading to accelerated gain of functions or degenerative evolution (*19*, *20*).

Our results show that paralogous genes can compensate each other's loss in the yeast PPI network, supporting models in which they contribute to the robustness of cellular functions (*5*, *6*). We also observed that many paralogs have an opposite effect and actually increase mutational fragility when the deletion of a paralog acts in a dominant manner to affect its cognate copy. Human genetics has shown that paralogs can produce a genetic background that mitigates the effects of mutations (*21*), while at the same time being rich in nominally dominant disease mutations (*22*). Our results offer at least a partial mechanism to explain this contradiction, where the duplication of homomers increases the complexity of protein interaction networks in a non-reversible manner, often at the cost of an increased susceptibility to mutations (Fig. 4F). By expanding the scale of previous findings showing that physical interaction between paralogs could interfere with their evolution (*8*, *9*), our work changes the scope of functional dependency from isolated events to a potentially widespread phenomenon that could negatively affect the robustness of PPI networks. Further studies testing larger cohorts of paralogs and in other species will allow assessment of the extent of functional dependency as a feature of the evolution of protein interaction networks.

REFERENCES AND NOTES

1. A. Force *et al.*, *Genetics* **151**, 1531–1545 (1999).
2. X. He, J. Zhang, *Genetics* **169**, 1157–1164 (2005).
3. F. A. Kondrashov, I. B. Rogozin, Y. I. Wolf, E. V. Koonin, *Genome Biol.* **3**, research0008.1 (2002).
4. T. Vavouri, J. I. Semple, B. Lehner, *Trends Genet.* **24**, 485–488 (2008).
5. A. DeLuna *et al.*, *Nat. Genet.* **40**, 676–681 (2008).
6. G. Diss, D. Ascencio, A. DeLuna, C. R. Landry, *J. Exp. Zool. B Mol. Dev. Evol.* **322**, 488–499 (2014).
7. R. Kafri, M. Levy, Y. Pilpel, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11653–11658 (2006).
8. C. R. Baker, V. Hanson-Smith, A. D. Johnson, *Science* **342**, 104–108 (2013).
9. J. T. Bridgham, J. E. Brown, A. Rodríguez-Marí, J. M. Catchen, J. W. Thornton, *PLOS Genet.* **4**, e1000191 (2008).
10. Materials and methods are available online as supplementary materials.
11. K. Tarassov *et al.*, *Science* **320**, 1465–1470 (2008).
12. L. Freschi, F. Torres-Quiroz, A. K. Dubé, C. R. Landry, *Mol. Biosyst.* **9**, 36–43 (2013).
13. G. Diss, A. K. Dubé, J. Boutin, I. Gagnon-Arsenault, C. R. Landry, *Cell Rep.* **3**, 2155–2167 (2013).
14. A. DeLuna, M. Springer, M. W. Kirschner, R. Kishony, *PLOS Biol.* **8**, e1000347 (2010).
15. P. Kemmeren *et al.*, *Cell* **157**, 740–752 (2014).
16. M. A. Fares, O. M. Keane, C. Toft, L. Carretero-Paulet, G. W. Jones, *PLOS Genet.* **9**, e1003176 (2013).
17. A. Wagner, *Proc. Biol. Sci.* **270**, 457–466 (2003).
18. R. A. Veitia, S. Bottani, J. A. Birchler, *Trends Genet.* **29**, 385–393 (2013).
19. M. W. Gray, J. Lukes, J. M. Archibald, P. J. Keeling, W. F. Doolittle, *Science* **330**, 920–921 (2010).
20. P. D. Dixit, S. Maslov, *PLOS Comput. Biol.* **9**, e1003023 (2013).
21. T. L. Hsiao, D. Vitkup, *PLOS Genet.* **4**, e1000014 (2008).

ACKNOWLEDGMENTS

The data reported in this article is available in the supplementary materials. Funding for this project comes from Canadian Institutes of Health Research grants 299432 and 324265 to C.R.L. C.R.L. holds the Canada Research Chair in Evolutionary Cell and Systems Biology. G.D. is a Marie-Curie Fellow and received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement 608959. We acknowledge support of the Spanish Ministry of Economy and Competitiveness, "Centro de Excelencia Severo Ochoa 2013-2017" and of the Centres de Recerca de Catalunya Programme/Generalitat de Catalunya. The authors thank E. D. Levy, S. Peisajovich, B. Lehner, A. Deluna, A. M. New, and N. Aubin-Horth for comments and K. Adams for sharing data files. G.D. and C.R.L. designed the research, analyzed the results, and wrote the manuscript, with help from I.G.-A. G.D., I.G.-A., A.-M.D.-C., H.V., D.I.A., and C.M.B. performed experiments. A.-M.D.-C., H.V., D.I.A., and C.M.B. provided comments on the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/355/6325/630/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S4
References (*22*–*52*)

13 September 2016; accepted 13 January 2017
10.1126/science.aai7685

Editor's Summary

Robustness of protein networks

It is thought that gene duplication helps cells maintain genetic robustness, but this seems not to be the whole story. Diss *et al*. investigated the fate of protein-protein interactions among duplicated genes in yeast. Some interacting duplicates evolved mutual dependence, resulting in a more fragile system. This finding helps us understand the evolutionary trajectories of gene duplications and how seemingly redundant genes can increase the complexity of protein interaction networks.

Science, this issue p. 630