

Regulatory evolution across the protein interaction network

Bernardo Lemos, Colin D Meiklejohn & Daniel L Hartl

Protein-protein interactions may impose constraints on both structural and regulatory evolution. Here we show that protein-protein interactions are negatively associated with evolutionary variation in gene expression. Moreover, interacting proteins have similar levels of variation in expression, and their expression levels are positively correlated across strains. Our results suggest that interacting proteins undergo similar evolutionary dynamics, and that their expression levels are evolutionarily coupled. These patterns hold for organisms as diverse as budding yeast and fruit flies.

Recent studies on variation in gene expression in natural populations indicate that a number of factors influence regulatory evolution, including membership in a particular functional class or biological process¹ and the pattern of sex-biased expression^{2,3}. Protein-protein interactions may also be relevant to regulatory evolution if they mean that the interacting partners of a given protein impose stricter stoichiometric requirements. For proteins that participate in complexes, such requirements may result in increased selection pressure to maintain evolutionarily stable expression levels and might be manifested in a negative association between evolutionary variation in gene expression and the number of protein interactions.

We analyzed variation in expression of 5,978 yeast (*Saccharomyces cerevisiae*) genes among four natural isolates grown under identical laboratory conditions⁴, of 4,439 fly (*Drosophila melanogaster*) genes in adult males from eight wild-type strains of geographically diverse origin² and of a similar number of genes assayed in both sexes of two species of *Drosophila*³. We matched these data to protein-protein interactions that could be assigned with high confidence (confidence score > 0.55): 8,728 in yeast⁵ and 3,964 in flies⁶. Choosing interactions with more stringent selection criteria produced similar results (Supplementary Note online).

The breadth of variation of expression for a given gene in both flies and yeast (gene expression polymorphism) was negatively correlated with the number of protein-protein interactions found for the product of that gene (yeast: $\rho = -0.11$, $P < 0.0001$, $n = 3,536$; flies: $\rho = -0.15$, $P = 0.022$, $n = 225$; Fig. 1). Gene expression divergence between two *Drosophila* species was also negatively associated with the number of protein-protein interactions (males: $\rho = -0.18$, $P = 0.012$, $n = 205$; females: $\rho = -0.19$, $P = 0.007$, $n = 205$). Furthermore, the connectivity of a protein was an excellent predictor of the mean variation in expression in classes of genes with a similar number of interactions

(Supplementary Fig. 1 online). Taken together, these results suggest that protein interactions might be a constraint that reduces evolutionary variation in gene expression both within and between species.

Among yeast proteins, the rate of accumulation of replacement substitutions decreases as the number of protein-protein interactions increases^{7,8}. The importance of this conclusion has been disputed on the grounds that the effect may be confounded with gene expression level⁹ (absolute transcript abundance) and is driven disproportionately by highly connected proteins¹⁰. In our data, both of these hypotheses can be ruled out (Supplementary Note online). First, the bivariate associations between evolutionary variation in gene expression, number of protein-protein interactions and gene expression level indicate that gene expression level may obscure the negative association between the number of protein-protein interactions and gene expression polymorphism and divergence. Accordingly, when we include gene expression level as a covariate, the partial Spearman rank correlation between evolutionary variation in gene expression and number of protein-protein interactions not only remains statistically significant but tends to become stronger. Second, the correlation between gene expression variation and number of protein-protein interactions remains significant in both yeast and flies when removing proteins with more than four interactions, whereas it is lessened by removing minimally connected proteins or proteins that do not participate in complexes. This is in agreement with the observation that the disruption of genes participating in protein complexes has a stronger effect on fitness than the disruption of genes not involved in complexes¹¹.

Quantitative genetic models identify a number of forces that may influence the breadth of variation in quantitative traits in natural populations, including population size, the mutation rate for the trait and the strength of stabilizing selection. In particular, mutation-selection models¹² and the neutral theory of evolution¹³ indicate that the strength of stabilizing selection is an important determinant of standing levels of variation. All else being equal, weaker stabilizing selection is expected to maintain higher levels of variation than

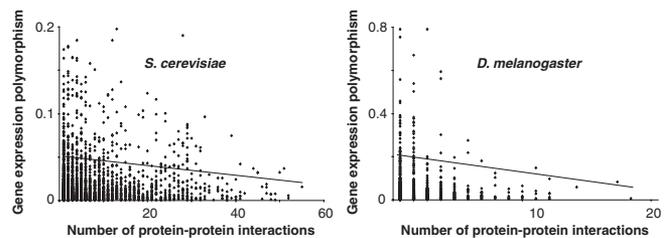


Figure 1 Relationship between the population genetic variation in gene expression (gene expression polymorphism) in *S. cerevisiae* and *D. melanogaster* and the number of protein-protein interactions.

Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Hartl Lab, Room 2105, Cambridge, Massachusetts, 02138, USA. Correspondence should be addressed to B.L. (blemos@oeb.harvard.edu).

Published online 19 September 2004; doi:10.1038/ng1427

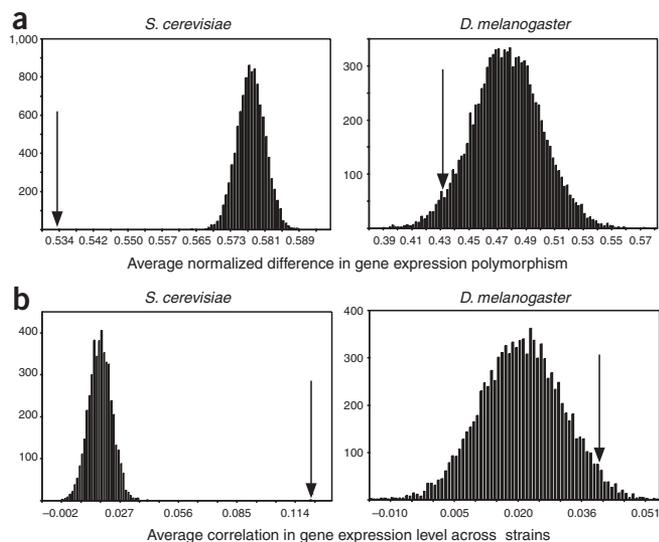


Figure 2 Interacting proteins have similar levels of population genetic variation in expression (gene expression polymorphism) and expression levels that are positively correlated across strains. **(a)** Distributions of the average normalized difference in gene expression polymorphism in sets of proteins randomly paired in yeast and fruit flies. The average normalized difference in gene expression polymorphism between proteins that interact is indicated by arrows. **(b)** Distributions of the average Spearman rank correlation of proteins randomly paired in yeast and fruit flies. The average Spearman rank correlation between proteins that interact is indicated by arrows.

stronger stabilizing selection. Therefore, differences in the breadth of population genetic variation in expression levels (gene expression polymorphism) may reflect differences in the strength of stabilizing selection across genes. In yeast, deletion of individual interacting proteins has similar fitness consequences⁸. This suggests that proteins that interact with each other may have similar effects on overall fitness and may, therefore, be under similar strengths of stabilizing selection.

Data on gene expression polymorphism and absolute transcript abundance were available for both members of 8,607 (yeast) and 550 (flies) interacting protein pairs. In these data sets, we calculated the normalized difference in gene expression polymorphism between interacting proteins averaged across all pairs (**Supplementary Methods** online). We calculated null distributions from 10,000 samples generated by shuffling the list of interacting partners. The amount of expression polymorphism among genes that encode interacting proteins was more similar than that among random pairs of genes (yeast, $P < 10^{-4}$; flies, $P = 0.0992$ in genes with more than one allele in both interacting proteins), and these results hold true when gene expression polymorphism is standardized by absolute expression level (yeast, $P < 10^{-4}$; flies, $P = 0.0715$; **Fig. 2a**). These results confirm our prediction that interacting proteins have similar levels of gene expression polymorphism. This observation suggests that the two proteins in an interacting pair undergo similar

evolutionary dynamics and, more specifically, may be subject to similar strengths of stabilizing selection.

Expression levels of interacting proteins may covary across genotypes if both proteins are regulated by the same segregating genetic variants or if there is coevolution to minimize the deleterious effects of imbalance in protein concentration between members of an interacting protein pair^{7,14}. In either case, we would expect the gene expression levels of interacting proteins to be positively correlated across strains. To test this prediction, we calculated the Spearman rank correlation in gene expression between interacting proteins averaged across all pairs (yeast, $\rho = 0.12$; flies, $\rho = 0.05$). Expression of genes that encode interacting proteins was significantly more positively correlated across strains than that of randomly paired proteins (yeast, $P < 10^{-4}$; flies, $P = 0.0580$; **Fig. 2b**).

In yeast, genes whose products are involved in complexes have larger fitness consequences when deleted¹¹, are precisely coregulated with their interacting partners¹¹ and have transcription and translation properties that minimize noise in the final protein concentration¹⁵, compared with genes whose products do not participate in protein complexes. Our results on variation in gene expression within and between species indicate that protein-protein interactions may constrain regulatory evolution, with such constraints presumably serving to maintain the protein's stoichiometric balance with its interacting partners^{11,14}. These results indicate that the protein interaction network may provide a common ground for understanding regulatory evolution in disparate organisms.

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

ACKNOWLEDGMENTS

We thank D. Hewitt, R. Frankham, J. Blumenstiel, J. Parsch and members of the laboratory of D.L.H. for suggestions. This work was supported by grants from the US National Institutes of Health to D.L.H.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 11 May; accepted 18 August 2004

Published online at <http://www.nature.com/naturegenetics/>

- Rifkin, S.A., Kim, J. & White, K.P. *Nat. Genet.* **33**, 138–144 (2003).
- Meiklejohn, C.D., Parsch, J., Ranz, J.M. & Hartl, D.L. *Proc. Natl. Acad. Sci. USA* **100**, 9894–9899 (2003).
- Ranz, J.M., Castillo-Davis, C.I., Meiklejohn, C.D. & Hartl, D.L. *Science* **300**, 1742–1745 (2003).
- Townsend, J.P., Cavalieri, D. & Hartl, D.L. *Mol. Biol. Evol.* **20**, 955–963 (2003).
- Bader, J.S., Chaudhuri, A., Rothberg, J.M. & Chant, J. *Nat. Biotechnol.* **22**, 78–85 (2004).
- Giot, L. *et al. Science* **302**, 1727–1736 (2003).
- Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C. & Feldman, M.W. *Science* **296**, 750–752 (2002).
- Fraser, H.B., Wall, D.P. & Hirsh, A.E. *BMC Evol. Biol.* **3**, 11 (2003).
- Pal, C., Papp, B. & Hurst, L.D. *Nature* **421**, 496–497 (2003).
- Bloom, J.D. & Adami, C. *BMC Evol. Biol.* **3**, 21 (2003).
- Papp, B., Pal, C. & Hurst, L.D. *Nature* **424**, 194–197 (2003).
- Lande, R. *Genet. Res.* **26**, 221–235 (1975).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, Massachusetts, 1983).
- Veitia, R.A. *BioEssays* **24**, 175–184 (2002).
- Fraser, H.B., Hirsh, A., Giaever, G., Kumm, J. & Eisen, M.B. *PLoS Biology* **2**, 834–838 (2004).