



Duplicate genes increase gene expression diversity within and between species

Zhenglong Gu^{1,4}, Scott A Rifkin^{2,3}, Kevin P White³ & Wen-Hsiung Li¹

Using microarray gene expression data from several *Drosophila* species and strains, we show that duplicated genes, compared with single-copy genes, significantly increase gene expression diversity during development. We show further that duplicate genes tend to cause expression divergences between *Drosophila* species (or strains) to evolve faster than do single-copy genes. This conclusion is also supported by data from different yeast strains.

Different copies of duplicate genes can become specialized at different developmental stages (e.g., different copies of *Hox* genes^{1,2}). Therefore, duplicate genes should have more diversified expression profiles than single-copy genes during the development of an individual. Moreover, the redundancy conferred by duplicate genes for the regulation of a function may facilitate organismal adaptation to environmental changes, so that the expression patterns of duplicate genes are, in general, expected to diverge between species faster than those of single-copy genes. The latter hypothesis is particularly interesting because changes in gene expression may lead to important phenotypic evolution^{3,4}. A direct examination of these two hypotheses at the genomic level, however, was not possible until high-throughput gene expression data became available. In this study, we tested the two predictions using microarray data from fruit flies and yeast.

We first used the data on gene expression during the start of metamorphosis (between the late third-instar larval and white prepupal stages) in three species of the *D. melanogaster* subgroup. This data set included 11,723 genes from a study investigating the evolution of gene expression in four inbred strains of *D. melanogaster* (Canton S, Oregon R, Samarkand and Netherlands2), one inbred strain of *D. simulans* and one inbred strain of *D. yakuba*⁵; we compared larvae and prepupae eight times for each lineage to obtain more accurate estimates of gene expression (Supplementary Methods online). We estimated developmental changes in expression of each gene in each lineage separately using an ANOVA model⁵. We used the 95% confidence intervals to determine first whether a gene was differentially expressed across the start of metamorphosis in a particular lineage, and second, whether the developmental change differed between lineages (that is, whether it had evolved; see ref. 5 for methodology).

These lists of genes whose expression changed during development or differed between lineages were the starting point of this study.

We conducted two analyses. First, we examined whether expression levels of duplicate genes changed significantly across the start of metamorphosis more often than those of single-copy genes (a within-genome comparison). We used strict criteria to define duplicate and single-copy genes (Supplementary Methods online; relaxation of the criteria led to the same conclusions). Furthermore, to avoid cross-hybridization, we included in the analysis only those duplicate genes that had K_S values (number of substitutions per synonymous site) to their closest paralogs of >0.5 (Supplementary Methods online; all the following results were the same without this restriction). Because each gene family originated from one common ancestor, we regarded each gene family as a uniquely represented entry in our first analysis⁶. For this purpose, we grouped duplicate genes into families (see Supplementary Fig. 1 online for the family size distribution). We identified 3,332 duplicate genes belonging to 932 gene families in the data set. In 818 gene families, one or more family members had significant changes in expression across the start of metamorphosis in at least one strain (Table 1). The proportion of gene families that showed changes in gene expression was significantly higher (818 of 932; 88%) than that for single-copy genes (2,030 of 3,356; 60%). We obtained similar results in individual strains (Supplementary Table 1 online). Considering each duplicate gene individually, the proportion of duplicate genes (72%, 2,389 genes) that showed changes in expression across the start of metamorphosis was still much higher than that of single-copy genes (60%; Table 1 and Supplementary Table 2 online).

Table 1 Numbers of single-copy genes and gene families with changes in expression levels across the start of metamorphosis in fruit fly strains

Change across the start of metamorphosis?	Number of single copy genes	Number of gene families*
Yes	2,030	818
No	1,326	114
Total	3,356	932
Percent with changes in expression	60.5	87.8

*A gene family was considered to have changes in expression during the start of metamorphosis if one or more of its members had a significant change in expression across the start of metamorphosis in at least one of the six strains studied. For the comparison between single-copy genes and gene families, $\chi^2 = 242.1$, degrees of freedom (d.f.) = 1, $P \approx 0$. The results were essentially the same if the numbers of duplicate genes were used in the comparison, that is, a significantly higher proportion of duplicate genes than of single-copy genes had changes in gene expression during the start of metamorphosis ($\chi^2 = 93.2$, d.f. = 1, $P \approx 0$).

¹Department of Ecology & Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA. ²Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, Connecticut 06520-8106, USA. ³Department of Genetics, Yale University School of Medicine, PO Box 208005, New Haven, Connecticut 06520-8005, USA. ⁴Present address: Stanford Genome Technology Center, 855 California Avenue, Palo Alto, California 94304, USA. Correspondence should be addressed to W.-H.L. (whli@uchicago.edu).

These observations suggest that duplication of a developmentally regulated gene is more likely to be advantageous than duplication of a gene that is not involved in development. One possible reason for this is that gene duplication can provide regulatory diversity to an organism during development⁷. To determine whether this is the case, we investigated differences in gene expression patterns in each gene family. Of the 818 gene families that include genes with developmental changes in expression, we found that about 70% (545) had different expression profiles (increase, no expression change or decrease) between individual genes in each family. The actual proportion might be higher, because we excluded some members of certain gene families from the expression data set. These results indicate that duplicate genes significantly increase the gene expression diversity of an organism. This conclusion is strengthened by the fact that probable differences in spatial expression patterns between different tissues were not taken into account in the above analysis.

Second, we made a between-genome comparison by examining differences in gene expression between species and between different strains within species. Of the 3,332 duplicate genes in the data set, 1,593 of them (~50%) had significantly different expression patterns between at least two of the six strains during the start of metamorphosis. This proportion is much higher than that of single-copy genes (1,202 of 3,356; 36%; **Table 2**). Individual pair-wise comparisons between strains also supported this conclusion (data not shown).

We also examined differences in expression of duplicated genes between two yeast strains and observed a similar pattern. A study dissecting transcriptional regulation in the budding yeast⁸ compared gene expression profiles between a laboratory strain (BY) and a wild strain (RM) of *Saccharomyces cerevisiae*. Differential expression between strains was detected for more than 1,500 genes (of 6,215 genes; **Supplementary Methods** online). Using these data, we found that a significantly higher proportion of duplicate genes than of single-copy genes had expression patterns that differed between the two strains of yeast (**Table 2**). Therefore, the results from both fruit flies and yeast indicate that duplicate genes cause expression divergences to evolve between species (or strains) faster than single-copy genes. This conclusion also holds true for very old duplicate genes (**Supplementary Table 3** online).

In both within-genome and between-genome comparisons, we found that duplicate genes were more likely than single-copy genes to show changes in expression profiles. These conclusions did not change when we used different criteria to define duplicate genes and single-copy genes (**Supplementary Table 4** online). Furthermore, we found that protein function, codon usage bias or gene evolutionary rate could not explain the observed patterns (data not shown). The observed association between gene duplication and increased gene expression diversity within and between species is important for two reasons. First, divergence in expression between duplicate genes may lead to functional specialization, which is a means of retaining both copies of duplicate genes in a genome^{9,10}. Second, we found that relatively old duplicate gene

Table 2 Distributions of single-copy and duplicate genes in comparisons between different strains of fruit fly and yeast

Organism	Differential expression? ^a	Number of single-copy genes	Number of duplicate genes
Fruit fly	Yes	1,201	1,593
	No ^b	2,155	1,739
	Total	3,356	3,332
	Percent with differential expression ^c	35.8	47.8
Yeast	Yes	541	392
	No	2,252	925
	Total	2,793	1,317
	Percent with differential expression ^c	19.4	29.8

^aYes, genes were differentially expressed between at least two strains within or between species; no, genes were similarly expressed between every pair of strains within and between species. ^bIf we considered only genes that showed changes in expression during the start of metamorphosis, the proportion of duplicate genes with different expression patterns between strains was still significantly higher than that of single-copy genes. ^cFor the comparison between fruit fly strains or species, $\chi^2 = 97.6$, d.f. = 1, $P = 0$. For the comparison between yeast strains, $\chi^2 = 54.5$, d.f. = 1, $P < 10^{-12}$.

pairs still contributed to expression diversity between strains, and earlier studies showed that functional redundancy may exist between distantly related duplicate genes^{11,12}. Thus, during speciation or adaptation, these functionally redundant copies could have more chances (less constraint) to adapt to new environmental and physiological conditions than single-copy genes. The results of this study are particularly interesting in light of the proposal that changes in gene expression can lead to important phenotypic changes in evolution^{13–15}, which might lead to species differentiation. As more intra- and inter-species gene expression data become available at the genome level, the relationships among gene duplication, gene expression evolution and phenotypic change will be better understood.

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

ACKNOWLEDGMENTS

We thank L. Zhang, C.-I. Wu, L.M. Steinmetz, M.H. Kohn, H. Wang, S. Hua and K. Thornton for discussions and comments. This work was supported by grants from the US National Institutes of Health to W.-H.L. and K.P.W.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 22 November 2003; accepted 2 April 2004

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

- McGinnis, W. & Krumlauf, R. *Cell* **68**, 283–302 (1992).
- Jozefowicz, C., McClintock, J. & Prince, V. *J. Struct. Funct. Genomics* **3**, 185–194 (2003).
- King, M.C. & Wilson, A.C. *Science* **188**, 107–116 (1975).
- Levine, M. & Tjian, R. *Nature* **424**, 147–151 (2003).
- Rifkin, S.A., Kim, J. & White, K.P. *Nat. Genet.* **33**, 138–144 (2003).
- Rubin, G.M. *et al. Science* **287**, 2204–2215 (2000).
- Cooke J., Nowak, M.A., Boerlijst, M. & Maynard-Smith, J. *Trends Genet.* **13**, 360–364 (1997).
- Brem, R.B., Yvert, G., Clinton, R. & Kruglyak, L. *Science* **296**, 752–755 (2002).
- Ohno, S. *Evolution by Gene Duplication* (Springer, Berlin, 1970).
- Force, A. *et al. Genetics* **151**, 1531–1545 (1999).
- Tautz, D. *BioEssays* **14**, 263–266 (1992).
- Gu, Z. *et al. Nature* **421**, 63–66 (2003).
- Wray, G.A. *et al. Mol. Biol. Evol.* **20**, 1377–1419 (2003).
- Carroll, S.B. *Cell* **101**, 577–580 (2000).
- Davidson, E.H. *Genomic Regulatory Systems: Development and Evolution* (Academic, San Diego, California, 2001).