The Role of *Cis*-Regulatory Motifs and Genetical Control of Expression in the Divergence of Yeast Duplicate Genes

Lindsey J. Leach, Ze Zhang,* Chenqi Lu,† Michael J. Kearsey,* and Zewei Luo*†*

*School of Biosciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom; and †Laboratory of Population and Quantitative Genetics, Institute of Biostatistics, Fudan University, Shanghai, China

Expression divergence of duplicate genes is widely believed to be important for their retention and evolution of new function, although the mechanism that determines their expression divergence remains unclear. We use a genetical genomics approach to explore divergence in genetical control of yeast duplicate genes created by a whole-genome duplication that occurred about 100 MYA and those with a younger duplication age. The analysis reveals that duplicate genes have a significantly higher probability of sharing common genetic control than pairs of singleton genes. The expression quantitative trait loci (eQTLs) have diverged completely for a high proportion of duplicate pairs, whereas a substantially larger proportion of duplicates share common regulatory motifs after 100 Myr of divergent evolution. The similarity in both genetical control and *cis* motif structure for a duplicate pair is a reflection of its evolutionary age. This study reveals that up to 20% of variation in expression between ancient duplicate gene pairs in the yeast genome can be explained by both *cis* motif divergence (~8%) and by *trans* eQTL divergence (~10%). Initially, divergence in all 3 aspects of *cis* motif structure, *trans*-genetical control, and expression evolves coordinately with the coding sequence divergence of both young and old duplicate pairs. These findings highlight the importance of divergence in both *cis* motif structure and *trans*-genetical control in the diverse set of mechanisms underlying the expression divergence of yeast duplicate genes.

Introduction

Expression divergence of duplicate genes has long been considered important both for understanding their retention and also for interpreting their functional divergence (Ohno 1970; Ferris and Whitt 1979; Force et al. 1999). With the advent of functional genomics, many data sets of genome-wide mRNA expression profiles and of *cis*-regulatory motifs identified by comparative genomics in the yeast *Saccharomyces cerevisiae* have accumulated in recent years (Aach et al. 2000; Pilpel et al. 2001; Kellis et al. 2003). These make it possible to depict the divergence of duplicate genes in *cis*-regulatory structure and expression at a genome scale.

Several studies have investigated how much expression divergence following yeast gene duplication could be explained by the evolution of regulatory motifs. They indicated that duplicate genes tend to be coexpressed, but the correlation between motif content and expression similarity is generally weak; only a trivial proportion of expression variation can be explained by motif divergence (Zhang et al. 2004; Gu et al. 2005). Therefore, it was postulated that, in addition to the *cis*-regulatory motif structure, multiple *trans*-acting factors in the gene network would play an important role in explaining the divergence pattern of expression of duplicate genes (Zhang et al. 2004), raising an interesting question of how much expression divergence following yeast gene duplication could be explained by the evolution of *trans*-acting factors.

The application of classical quantitative genetics approaches to genomics, the genetical genomics approach, has become a powerful new approach for exploiting the genetics of gene expression profiles (Brem et al. 2002; Cheung et al. 2003; Schadt et al. 2003; Morley et al. 2004). Genetical genomics applies molecular marker genotyping in combination with expression profiling to a segregating population to map gene expression variation to genetic loci known as expression quantitative trait loci (eQTL). In this approach, the gene expression profile is treated as a quantitative trait, allowing expression to be studied genetically using quantitative trait loci (QTLs) mapping methods. The mapped eQTLs can be classified as c*is*-acting, whereby the eQTL maps to the location of the gene transcript, implying that genetic variation in the vicinity of a gene affects the expression level of that gene; otherwise, the eQTL is *trans*-acting and locates elsewhere on the genome.

It has been shown that yeast duplicate genes share a greater expression similarity than expected by chance alone (Zhang et al. 2004). In this study, we investigate this observation by looking at the divergence of *cis*-regulatory motifs and of *cis* and *trans*-genetical control between duplicate copies. The focus is on yeast duplicate genes created by a whole-genome duplication event and retained in the genome (Kellis et al. 2004), which potentially removes the effect of different duplication ages on the inference of divergence patterns. We also investigate whether any associations exist among the expression profile of yeast duplicate genes and various regulatory mechanisms. The analysis has drawn a detailed picture of the mode and tempo of expression divergence in the evolution of yeast duplicate genes.

Materials and Methods

Identifying Study Genes in *S. cerevisiae*

We used the database (http://www.broad.mit.edu/seq/YeastDuplication/) containing full information of 457 pairs of duplicate genes in the *S. cerevisiae* genome, which have been rigorously verified to have derived from the whole-genome duplication event that occurred about 100 MYA (Kellis et al. 2004). An all-against-all BlastP search (Altschul et al. 1997) was conducted on the set of *S. cerevisiae* protein sequences (downloaded from the *Saccharomyces*

Key words: yeast, duplication, divergence, gene expression, *cis* motifs, genetic regulation.

E-mail: z.luo@bham.ac.uk.

Mol. Biol. Evol. 24(11):2556–2565. 2007
doi:10.1093/molbev/msm188
Advance Access publication September 10, 2007

Genome Database [SGD], http://genome-www.stanford.edu/Saccharomyces/). A duplicate pair was defined using the criteria that E value $<1 \times 10^{-10}$, and sequences were alignable over 100 amino acids with an identity score of >40%. A duplicate gene was determined as a member of a gene family if it was a duplicate to any member as judged using the above criteria. Excluding duplicate genes from genome duplication and including only genes available in the data set of Brem and Kruglyak (2005) yielded a set of 134 pairs with no other paralogues in the genome, referred to as the duplicate pairs from individual duplication (PSC), and 78 families (412 genes), referred to as the duplicate families from sequence comparison (FSC).

For each protein pair that met the homology criteria, the amino acid sequences were aligned using ClustalW (Thompson et al. 1994), and the corresponding coding sequences were aligned on the basis of the protein alignments. The rate of nonsynonymous substitution (K_A) and the rate of synonymous substitution (K_S) between duplicate pairs were estimated using the PAML software (Yang and Nielsen 2000) using default parameters. For each gene family, all $n(n-1)/2$ possible pairwise comparisons were considered, where n is the number of family members, yielding a set of 1,249 FSC pairs. The age of each FSC pair was defined arbitrarily based on the rate of synonymous substitution, K_S, as young ($0 \le K_S \le 0.25$), middle ($0.25 \le K_S \le 0.75$), or old ($0.75 \le K_S \le 1.5$). We considered pairs with $K_S \le 1.5$ because when K_S becomes large it is difficult to obtain a reliable estimate of evolutionary age due to repeated substitutions at the same site (Li 1997). In addition, we report correlations for $K_A \le 0.3$ according to Gu et al. (2002) because higher values of K_A may be recognized as being less reliable.

A single copy gene (i.e., a singleton) was defined as coding for a protein that did not hit any other protein in the BlastP search with $E = 0.1$, giving a set of 1,606 singletons, 1,436 of which were available in the data set (Brem and Kruglyak 2005) and did not overlap with duplicate genes. We were able to analyze shared eQTLs for all 1,028,895 pairwise combinations of the 1,435 singleton genes for which eQTLs could be mapped. This loose similarity search criterion was used to make sure that a singleton is indeed a singleton. A criterion of $E = 0.01$ was also used to define a list of 2,161 singleton genes, 1,983 of which were available in the data set (Brem and Kruglyak 2005), although the results are qualitatively the same.

Data Set for Genetical Genomics Analysis

The mapping population comprised 112 haploid yeast segregants obtained from crossing 2 parental strains that differed in many morphological characters, a laboratory strain BY4716 (BY) and a wild-type strain RM11-1a (RM). These segregants were genotyped at each of 2,956 segregating genome-wide molecular markers and profiled for expression of 5,740 yeast open reading frames (ORFs) (5,727 genes) using cDNA microarrays (Brem and Kruglyak 2005). Combining the duplicate gene information (Kellis et al. 2004) with the data set of Brem and Kruglyak (2005), it is found that 448 of the 457 duplicate pairs have available data for genetical genomics analysis. These 448 pairs constituted the set of duplicates from genome duplication (DGD). Expression similarity between duplicate pairs and between pairs of singletons was estimated using a 1-way analysis of variance (ANOVA). The total variance was partitioned into variance between individuals (σ_b^2) and variance between the 2 copies of the pair (σ_w^2). The intraclass correlation (Snedecor and Cochran 1967) between the 2 copies was calculated as follows:

$$r = \left(s_b^2 - s_w^2\right) / \left\{s_b^2 + \left(n-1\right)s_w^2\right\}, \quad (1)$$

where r is the intraclass correlation and $n = 2$. For each duplicate pair, the ANOVA was performed separately using expression data of the 6 BY parent replicates and data of the 12 RM parent replicates to give intraclass correlations r_{BY} and r_{RM}, respectively; the 2 intraclass correlations were averaged to obtain an estimate of expression similarity for the ith pair as $r_i = (r_{BY} + r_{RM})/2$.

eQTL Detection Methods

On the basis of the marker and gene expression data sets, eQTL analysis in the present study was conducted for each of the 5,740 ORFs by mapping expression of each of these genes as a quantitative trait onto linkage maps of the genetic markers. The segregants were divided into 2 groups according to marker genotype and the gene expression levels compared between the 2 groups using the nonparametric Wilcoxon–Mann–Whitney (WMW) test (Conover 1980), as applied previously in Brem et al. (2002). In addition, eQTLs were mapped using a modified version of the composite interval mapping (CIM) method originally proposed by Zeng (1994). To select appropriate background markers for each expression trait, a multiple linear regression model was built using forward selection (Draper and Smith 1981) to select markers significantly associated with the trait. To avoid the problem of matrix singularity within the CIM algorithm, comarkers were filtered to exclude markers for which the individual genotypes are highly correlated (Pearson product moment correlation coefficient >0.8) with other comarkers. For each test interval, the flanking markers were also excluded as comarkers. The top 10 comarkers chosen by forward regression were used to provide the most effective background control.

eQTLs were then screened for each of the expression traits at a grid of 1 cM (or approximately a recombination frequency 0.01) across all 16 yeast chromosomes. The model to be fitted in the CIM analysis at the ith marker interval is as follows:

$$y_j = \mu + b^* x_j^* + \sum_{k \ne i, i+1} b_k x_{jk} + e_i, \quad \text{for } j = 1, 2, \ldots, n, \quad (2)$$

where y_j = phenotype of expression trait for the jth individual, μ = the overall population mean for the trait of interest, b^* = the genetic effect of the putative QTLs on the expression trait, x_j^* = the indicator of the QTL genotypes, $e_j \sim N\left(0, \sigma^2\right)$ is the effect of the environment on the trait, x_{jk} = genotype value of the kth comarker for the jth

individual, and b_k = partial regression coefficient of phenotype values regressed on the kth comarker.

It should be noted that estimation of the partial regression coefficients, b_k, is not straightforward when there is missing genotypic data at any comarkers in the model. To cope with the missing data problem, QTL Cartographer imputed the missing marker genotype with the mean of the genotypic values at the marker locus. It has been shown that this could result in serious bias in the estimate of the regression parameter, particularly when the proportion of missing data is large (Little 1992). In the QTL mapping setting, this will result in decreased statistical power for detecting the presence of QTL. Little (1992) suggested a Bayesian method for the imputation of missing data.

In the present study, we modified and reformulated the multiple regression analysis involved in CIM by implementing the Bayesian algorithm as suggested for regression with missing observations in explanatory variables in Little (1992). We first calculated the posterior probability distribution of each missing marker genotype from a Hidden Markov model and in turn calculated the probability distribution of marker genotypic values as the matrix $X = (x_{jk})$, where j represents the jth individual and k represents the kth marker. We considered all possible forms of X. For example, if missing information occurs at L marker loci, then the number of possible forms of the matrix will be 2^L in the present context because there are 2 possible genotypes at each missing marker locus. The expectation of matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is given by $E\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right] = \sum_{k=1}^{2^L} P_k\left[\left(X_k'X_k\right)^{-1}X_k'\right]$, with P_k being the multinomial probability of the kth possible form, $\boldsymbol{X}_k$, of matrix $\boldsymbol{X}$, under the assumption that the missing event occurs randomly. We compared the program with the modification with Windows QTL Cartographer 2.5 (Wang et al. 2006) by analyzing simulation data and found that the program we developed conferred an increased statistical power for detecting the simulated QTL in comparison with QTL Cartographer 2.5, particularly when the proportion of missing marker data is large.

A regulatory or eQTL was defined as an independent peak in the log-odds (LOD) score profile across a given chromosome. Peaks occurring within 20 cM of adjacent peaks were taken as a single eQTL peak because of insufficient evidence to declare the existence of multiple eQTL peaks over such narrow intervals. The eQTL location was defined as the location within the peak with the greatest LOD score. A 99% confidence interval (CI) for eQTL location was calculated according to the established 2 LOD support interval method (Lander and Botstein 1989).

An eQTL was classified as *cis*-acting if the 99% CI for its location mapped to within 10 kb of the start site for the gene (obtained from the SGD); otherwise, the eQTL was classified as *trans*-acting. A shared *cis* eQTL between 2 copies was defined if both copies had a *cis*-regulator mapping to the same relative location (within x kb) to the start site of the duplicate gene. We tried several values of x between 1 and 5 kb, although the results were essentially the same. A shared *trans* eQTL between 2 duplicate genes was declared either if there was overlap between the 99% CI for the corresponding peaks or if the locations of the corresponding peaks were less than or equal to 10 kb apart.

Yeast *Cis*-Regulatory Motif Data Set

The present study utilized an extensive genome-wide annotation of regulatory sites in the yeast genome, produced by van Nimwegen and coworkers and available from the SwissRegulon database, accessible via http://www.swissregulon.unibas.ch (Erb and van Nimwegen 2006; Pachkov et al. 2007). The annotations were produced in 2 steps. In the first step, the authors collected a set of weight matrices (WMs) representing the binding motif of a yeast transcription factor (TF) or complex of TFs by combining several sources of information. First, the ChIP-on-chip binding data of Harbison et al. (2004) were analyzed using Phylogibbs (Siddharthan et al. 2005). Second, a binding site clustering algorithm was used to curate the promoter database of *S. cerevisiae*, SCPD (Zhu and Zhang 1999). Combining the motifs resulting from these 2 procedures led to a total of 72 high confidence WMs for over 150,000 putative binding sites, most of which correspond to the binding motif of a given yeast TF, whereas a small number correspond to a complex of yeast TFs.

In the second step, a newly developed MotEvo algorithm (Erb and van Nimwegen 2006) was used to identify binding sites for each of the WMs by scanning the multiple alignments of each *S. cerevisiae* intergenic region with orthologous regions from 4 other related *Saccharomyces* species. MotEvo exhaustively reports putative locations of regulatory binding sites and assigns a posterior probability to each reported site. Combining the binding data with conservation data significantly improves annotations of regulatory sites, and the resulting full set of predicted sites constitutes the most comprehensive annotation of regulatory sites in the *S. cerevisiae* genome to date.

Yeast Gene Expression Data Set

We downloaded microarray expression data compiled from public genome mRNA expression data sets totaling 213 experimental conditions from the Web site http://www.arep.med.harvard.edu (Aach et al. 2000). Pearson's correlation coefficient was calculated for the expression profiles between 2 duplicate copies and used as an estimate of their expression similarity.

Genomic Distribution of eQTL Linkages

The genomic distribution of eQTL linkages was analyzed as in Brem et al. (2002) by dividing the genome into 611 bins of size 20 kb each (the bins at the ends of chromosomes were smaller). For example, using CIM, we found 5,027 eQTLs (excluding *cis* eQTLs) at false discovery rate (FDR) 0.05 for the 890 duplicate genes from whole-genome duplication. If these were randomly distributed across the genome, then the number of eQTLs in any one bin would follow a Poisson distribution with a mean of 8.23; the probability that a given bin would contain

13 or more eQTLs is $P = 0.0414$; therefore, this would not be expected to occur by chance alone and can be used as a threshold.

GO Functional Comparisons

The set of genes linking to selected hot spots on each chromosome was classified using Gene Ontology (GO) biological process terms using the FatiGO tool (Al-Shahrour et al. 2005). A Fisher's exact test is used to determine significant overrepresentation of GO terms in the list using the remaining duplicate genes that do not link to the hot spot for comparison, and *P* values are returned adjusted according to the FDR.

All programs for eQTL mapping were written using Fortran 90 and are available upon request from the corresponding author.

Results

Two gene copies derived from a whole-genome duplication event are born equal; they will be identical in all aspects of *cis*-regulatory motif structure, regulatory factors, and function (Li 1997). Over evolutionary time, the accumulation of mutations in one or both copies will result either in functional loss or divergence between the 2 copies. Below we look at divergence in the first 2 aspects focusing on 448 yeast duplicate pairs verified from whole-genome duplication approximately 100 MYA (Kellis et al. 2004) and referred to as duplicates from genome duplication.

We present a comparison with a set of duplicate pairs predicted using BlastP (Altschul et al. 1997) to have derived from individual small-scale gene duplications. On the basis of the distribution of the rate of synonymous distance, K_S (supplementary fig. 1, Supplementary Material online), it is clear that the identified gene pairs fall into 2 subgroups. The first group contains 134 duplicate pairs from (individual duplication) (PSC) of more ancient origin than the DGD pairs and with no other paralogues in the genome. The second group contains 78 multigene families (412 genes) from sequence comparison of more recent origin than the DGD pairs. To control for codon usage bias, we also considered only those duplicates with low bias measured by the effective number of codons according to Papp et al. (2003), although the results remain the same (results not shown but available upon request). Within each family, we consider all possible pairwise combinations of genes.

eQTL Analysis

Controlling the FDR at level 0.05 according to Benjamini and Hochberg (1995), we detected eQTLs for only 83 of the 448 DGD pairs using the WMW test, in contrast to 442 pairs using the CIM method. The number of eQTLs mapped using CIM (supplementary fig. 2, Supplementary Material online) has a distribution with a mean number of 5.77 (±2.30), which is similar to the genome-wide distribution, with a mean of 5.63 (±2.31). As expected, the WMW method confers a lower statistical power than the CIM method because the latter enables use of map information. Furthermore, CIM provides more effective control of the background effect, which is important whenever there are linked eQTLs in the same chromosomal region (supplementary fig. 3, Supplementary Material online). We report here the results of the CIM analysis at an FDR level of 0.05 (average corresponding LOD score 6.143) derived separately for each expression trait.

Divergence in Genetical Control Explains Expression Divergence between Duplicates

Significant eQTLs were classified as *cis* if the gene transcript locates within 10 kb of the eQTL, or as *trans* otherwise. *Cis* eQTLs were apparent for 12.4% of DGD genes, a similar rate of *cis*-acting variation to the genome as a whole. A shared *cis* eQTL was defined if both copies of a duplicate pair had a *cis*-acting eQTL in the same location relative to the start site of the gene. However, in the majority of cases, *cis* regulation was detected only for one copy of a duplicate pair and we could not evidence any shared *cis* eQTLs. We made the same observations for both PSC and FSC duplicate pairs, implying there is a rapid divergence in *cis*-regulatory control between duplicate copies following duplication.

We examined the number of *trans* eQTLs shared by DGD pairs after they have experienced the same divergent evolution period of about 100 Myr (Kellis et al. 2004). Of the 442 duplicate pairs mapped, 75.6% have completely diverged in terms of the genetical control of expression. This reveals that *trans* regulation of expression of duplicate genes diverged rapidly over 100 Myr of evolution. Among those 108 pairs that share genetical control of expression, the average proportion of shared *trans* eQTLs was only 28.7% (±1.92), showing how the regulation of expression of duplicate genes diverges over time for almost all pairs. However, 4 examples were found of complete conservation of *trans*-genetical control between the 2 copies of a duplicate pair, showing that some genes are completely constrained and do not diverge at all.

To pursue the analysis further, we questioned whether more recently duplicated gene pairs were less divergent in terms of shared genetical control than ancient pairs from whole-genome duplication. The 1,249 FSC pairs with eQTLs mapped were significantly more likely to share eQTLs than duplicates from genome duplication (table 2). Furthermore, the average proportion of shared eQTLs was significantly greater for FSC compared with DGD pairs, showing that younger duplicate pairs are much more likely to share *trans*-genetical control over their expression than older duplicate pairs. However, less than half of all FSC gene pairs shared any expression regulators at all, confirming the rapid nature of divergence in genetical control for the majority of duplicate pairs.

Given that the genetical control of expression had diverged completely for most duplicate gene pairs, we compared our findings with the shared genetical control for all possible pairwise combinations of singleton genes predicted from the entire genome using BlastP. Singleton pairs were highly significantly less likely to share eQTLs, and

Table 1
Correlations for DGD[a] and FSC[b] Duplicate Pairs

	Expression Similarity	Shared *Trans* eQTLs	Shared *Cis* Motifs
	(Pearson's Product Moment/Spearman Rank/Number of Gene Pairs)		
Amino acid sequence similarity	0.096 NS/0.013 NS/253[a]	0.344***/0.301*/104[a]	0.171 NS/0.152 NS/109[a]
	0.400***/0.314***/481[b]	0.545***/0.495**/481[b]	0.318***/0.172/216[b]
Expression similarity	—	0.318**/0.399***/108[a]	0.277*/0.224*/109[a]
		0.419***/0.418***/481[b]	0.241**/0.193*/216[b]
K_A	−0.707***/−0.700***/57[a]	−0.355*/−0.375*/57[a]	−0.409*/−0.421*/43[a]
	−0.256***/−0.241***/381[b]	−0.309***/−0.264***/381[b]	−0.440***/−0.385*/67[b]
K_S	−0.212***/−0.221***/401[b]	−0.326***/−0.289***/392[b]	−0.611***/−0.564***/53[b]

NOTE.—NS, not significant correlations.
*$P < 0.01$.
**$P < 0.001$.
***$P < 0.0001$.

where this occurred, the proportion of shared eQTLs was significantly lower than for all duplicate pairs of any age or origin (table 2). Therefore, yeast duplicate genes do still share a significantly greater level of genetical control of their expression than we would expect by chance alone, a particularly impressive finding for DGD pairs that have experienced over 100 Myr of divergent evolution.

Our comparison of ancient and recent duplicate genes suggested that divergence in genetical control may be coupled with evolutionary time. We used the rate of synonymous substitution, K_S, as a proxy of divergence time for the FSC duplicate pairs. A significant negative relationship was found between K_S and the fraction of shared eQTLs for a duplicate pair (table 1); in other words, the level of shared genetical control for a recently duplicated gene pair is a reflection of its evolutionary age. We classified the FSC duplicate pairs into 3 age groups (young, middle, and old) based on K_S and show how, as the age of a duplicate pair increases, the mean fraction of shared eQTLs decreases and approaches that for pairs of singleton genes (fig. 1). Note that this pattern is robust to the cutoff values chosen to form the age groups. This was logically plausible given that pairs of singletons shared on average a smaller proportion of their eQTLs than ancient DGD pairs, which in turn shared fewer eQTLs than more recently duplicated FSC pairs. A stronger correlation than that observed would not be expected because K_S is only a crude proxy of divergence time owing to considerable variation in synonymous rates among genes (Li 1997). Partial correlation analysis showed the negative correlation between K_S and expression similarity for FSC pairs (table 1) remains significant, but the magnitude of the correlation coefficient is markedly decreased ($r = -0.117$, $P = 0.02$, $n = 392$) when variation in shared *trans* eQTLs is controlled; therefore, the correlation between K_S and expression similarity can be explained at least partly by the correlation between K_S and divergence in genetical control.

The expression divergence over time between duplicate copies is well documented (Gu et al. 2002; Papp et al. 2003), although comparatively little is known about the mechanisms underlying the divergence. We obtained a highly significant positive correlation (table 1; supplementary fig. 4A, Supplementary Material online) between the expression similarity based on the parental expression data and the proportion of shared *trans* eQTLs between DGD pairs. This correlation was robust to the removal of 4 duplicate pairs with complete conservation of *trans*-genetical control (supplementary fig. 4B, Supplementary Material online). Shared *trans*-regulatory eQTLs account for ~10% (0.32^2) of the expression variation between 2 duplicate copies, confirming the relationship between expression similarity and the extent of shared *trans*-regulatory control between 2 duplicate genes from genome duplication. Furthermore, divergence in genetical control for

Table 2
Shared Regulatory Control for 3 Groups of Duplicates and Pairs of Singleton Genes

	FSC[a]	DGD[a]	PSC[a]	Singletons[b]	Comparison
					χ^2_1 (a/b)[c] = 26.56**
Percentage of pairs sharing eQTLs	38.5	24.4	25.4	21.6	χ^2_1 (d/e) = 179.35**
					t (a/b) = 4.60**, df = 148
Mean proportion of shared eQTLs	38.3 ± 0.81[d]	28.7 ± 1.92	24.0 ± 2.28	21.3 ± 0.02	t (d/e) = 19.55**, df = 623
					χ^2_1 (a/b) = 2.81 NS
Percentage of pairs sharing *cis* motifs	44	38	31	—	χ^2_1(a/c) = 4.32*
					t (a/b) = 2.07*, df = 217
Mean proportion of shared *cis* motifs	51.5 ± 3.45	46.2 ± 1.98	54.0 ± 4.29	—	t (b/c) = 1.65 NS df = 32

NOTE.—NS, not significant correlations; df, degrees of freedom.
[a] All duplicate genes collectively.
[c] Comparison of group a with group b using a chi square or *t*-test accordingly.
[d] Mean ± standard error.
*$P < 0.05$.
**$P < 0.0001$.

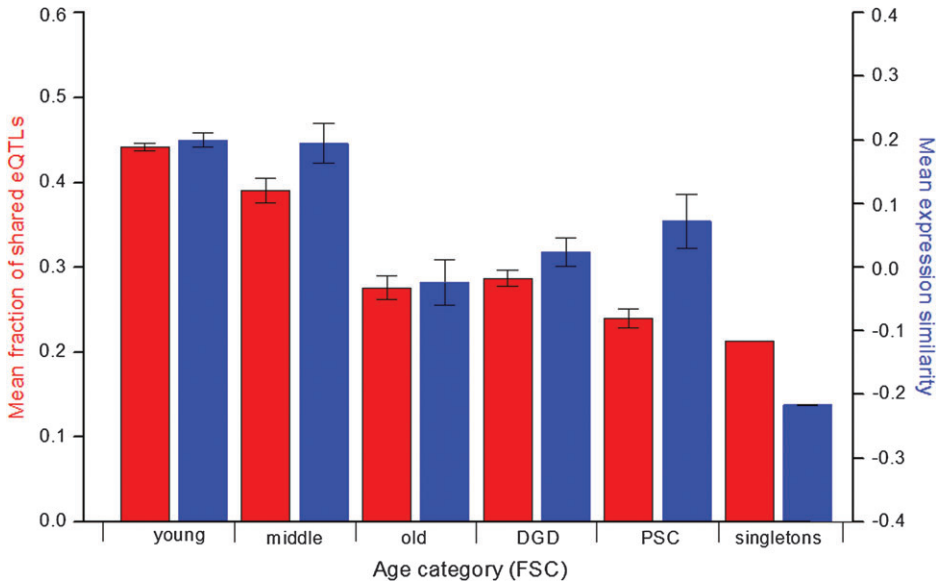

FIG. 1.—Comparison of shared eQTLs and expression similarity for duplicate and singleton pairs. The red segment shows the mean fraction of shared eQTLs, and the blue segment shows the mean expression similarity. The corresponding standard errors in each group are also shown. The FSC group (duplicate families from individual duplications) was divided into 3 age groups (young, middle, and old) on the basis of the rate of synonymous substitution, K_S.

FSC pairs is significantly positively correlated with expression similarity (table 1), and partial correlation analysis showed this correlation remains significant even when variation in K_S is controlled ($r = 0.279$, $P < 0.0001$, $n = 392$), supporting the significant relationship observed using DGD pairs that are all of exactly the same duplication age.

Genomic Distribution of eQTL Linkages

We asked the question whether there were many different *trans*-acting regulators, each controlling a few genes, or whether there were some master *trans* regulators responsible for regulating numerous genes. The yeast genome was divided into 20-kb bins, and the number of eQTL linkages was counted within each bin. Thresholds were calculated as the number of eQTLs in each bin expected by chance. An obviously nonrandom distribution of eQTL linkages was observed across the genome (fig. 2), with remarkable similarity for the 5 groups of genes (DGD, PSC, FSC, singletons, and the remaining genes in the genome). Indeed, there are very few linkage bins with an unequal contribution from the 5 groups (supplementary table 1, Supplementary Material online). The pattern observed is robust to the FDR level used for eQTL mapping (results not shown but available upon request). Therefore, it would seem that yeast duplicate genes are largely under the influence of the same master *trans* regulators as nonduplicate genes.

Detailed investigation showed the 3 most significant linkage hot spots (marked using arrows and enlarged in fig. 2) correspond to locations overrepresented by FSC duplicate genes. This was accounted for by a 30-member gene family containing known or predicted helicases. All but one of the family members with mapped eQTLs showed eQTLs in one or more of the 3 linkage hot spots. An analysis using the FatiGO tool (Al-Shahrour et al. 2005) confirmed that the FSC genes linking to each hot spot were significantly functionally enriched ($P < 0.05$) for telomere-independent telomere maintenance and for DNA helicase activity. We could identify possible master *trans* regulators responsible for 2 of these 3 hot spots using annotations in the SGD (supplementary table 2, Supplementary Material online). To further illustrate the existence of loci with widespread transcriptional effects on groups of genes sharing common functionality, we similarly analyzed the DGD genes linking to the richest eQTL hot spot on each chromosome. Four hot spots showed significant functional enrichment, and we identified possible master *trans* regulators responsible for 3 of these (table 3).

Divergence in *Cis* Motifs Explains Expression Divergence between Duplicates

On the basis of a comprehensive set of *cis*-regulatory motif predictions (Erb and van Nimwegen 2006; Pachkov et al. 2007), the distribution of the number of *cis*-regulatory motifs per gene from 717 yeast DGD genes showed a mean number of 3.04 (±2.12) (supplementary fig. 5, Supplementary Material online), which is significantly higher than the corresponding genome-wide mean of 2.84 (±2.03), as expected (He and Zhang 2005). We examined the number of motifs shared by each of the 290 duplicate pairs after they have experienced the same divergent evolution period of about 100 Myr. Of the 290 pairs, 181 (62%) did not share any common motifs. The remaining 109 pairs shared at least one motif, and the mean proportion of shared motifs was 46.2% (±1.98). This reveals that duplicate genes diverged rapidly in *cis*-regulatory structure, supporting the rapid divergence in *cis*-genetical control of expression for duplicate copies observed in the present study.

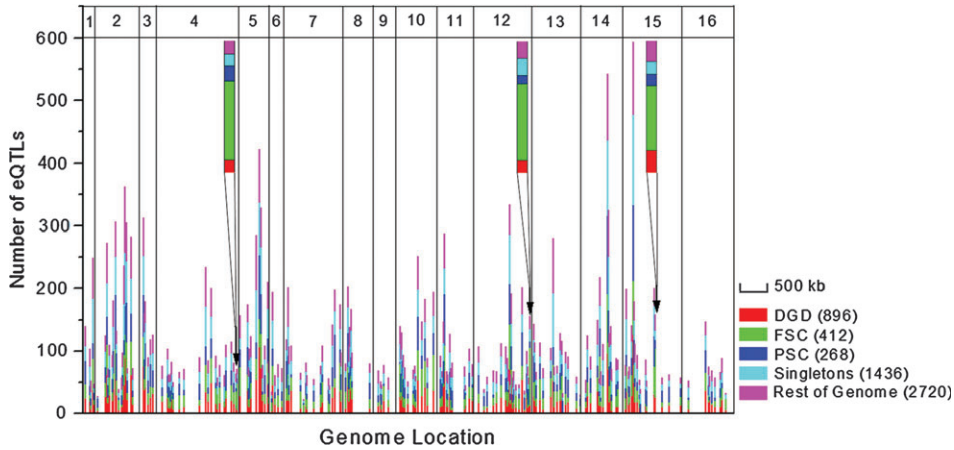

FIG. 2.—Number of eQTL linkages within 20-kb bins for linkage groups 1–16. The number of linkages within each bin is subdivided to show the relative contributions from 5 groups of genes: duplicate pairs from whole-genome duplication (DGD, red segment), duplicate families from individual duplicationsFSCs (FSC, green segment), duplicate pairs from individual duplications (PSC, blue segment), singletons (cyan segment), and all other genes (pink segment). The number of genes in each group is shown in brackets. Only linkage bins for which one or more of the 5 groups exceed the threshold level of linkages expected by chance are shown. Black arrows mark the locations of the 3 most significant (magnified) hot spots. Local (*cis*) linkages are excluded.

It is generally believed that duplicate genes are likely to share common regulatory motifs (Papp et al. 2003; Zhang et al. 2004; Li et al. 2005); in this respect, we obtained a significant positive correlation (table 1) between expression similarity (based on a microarray expression database from 213 different experimental conditions) and the proportion of shared *cis* motifs between DGD pairs. Shared regulatory motifs account for a noteworthy ~8% (0.27^2) of expression variation between duplicate copies. Interestingly, there is no significant relationship between the fraction of shared *cis* motifs and the fraction of shared *trans* eQTLs between DGD copies (Pearson correlation: $r = -0.151$, $P = 0.444$, degrees of freedom = 27).

Given that divergence in *trans*-genetical control was coupled with evolutionary time, we investigated whether FSC pairs would be more likely to share *cis* motif structure than more ancient duplicates. We found 44% of FSC pairs shared *cis* motifs, which is similar to the proportion for DGD pairs (table 2). However, for those pairs that shared motifs, the motif similarity was significantly greater for FSC compared with DGD pairs. A highly significant negative correlation was found between K_S and the fraction of shared *cis* motifs for FSC duplicate pairs (table 1), showing that the extent of shared *cis* motif structure for a recently duplicated gene pair is a reflection of its evolutionary age. We note that FSC pairs have a significantly higher rate of nonsynonymous to synonymous substitution (K_A/K_S) than DGD pairs (supplementary fig. 6, Supplementary Material online). A greater rate of sequence evolution for FSC pairs may explain, at least in part, why FSC pairs were not more likely to share *cis* motif structure despite having a more recent origin than DGD pairs.

Relationship between Expression Divergence and Coding Sequence Divergence

Numerous studies have attempted to unravel the relationship between expression divergence and coding sequence divergence using a combination of duplicate genes both from whole-genome and small-scale duplication events (Wagner 2000; Maslov et al. 2004; Zhang et al. 2004; Gu et al. 2005). However, several more recent studies (Davis and Petrov 2005; Guan et al. 2007; Musso et al. 2007) have employed functional analyses of gene duplications to show that there is a potential of distinct evolutionary scenarios for paralogues that arose through different duplication mechanisms, emphasizing the importance of discriminating between different groups of duplicate genes.

In the present analysis, the amino acid sequence similarity between younger duplicate pairs (FSC) is significantly positively correlated with all 3 aspects of divergence in *cis* motif structure, *trans* eQTLs, and expression (table 1). As expected, corresponding significant negative correlations are observed using the nonsynonymous distance, K_A (table 1), as a measure of coding sequence divergence. As

Table 3
eQTL Hot spots Enriched ($P < 0.05$) for Functional Categories of Yeast DGD Genes

Linkage Bins[a]	Number of Genes	Common Function	Putative Regulators
	88	Nucleosome assembly	SWI4, SLX8
V:390		Protein biosynthesis	Unknown
VIII:90	39	Nucleosome assembly	YAP3
XII:670	64	Electron transport	HSP60
XIV:490	79	Intracellular transport	Unknown

[a] The location of the center of each linkage bin is shown as chromosome: kilobase pair.

the sequence divergence of a young duplicate pair increases, as indicated by either measure, both the fraction of shared *trans* eQTLs and expression similarity decrease and approach the significantly lower levels observed for pairs of singleton genes (supplementary fig. 7, Supplementary Material online). Thus, initially at least, divergence in genetical control, *cis* motif structure, and expression are all moderately correlated with coding sequence divergence of duplicate genes.

In contrast, analysis of more ancient DGD pairs does not support a significant association between the percent identity of amino acid sequences between 2 duplicate copies and the 3 aspects of divergence in *cis* motif structure, *trans* eQTLs, and expression. However, a significant correlation was revealed with divergence in *trans* eQTLs when excluding those 4 duplicate pairs with complete conservation of eQTLs (table 1). In fact, unlike previous studies that have considered only sequence identity (Guan et al. 2007), the present study reveals a significant negative correlation between the nonsynonymous distance, K_A (table 1), and all 3 aspects of divergence, indicating a significant relationship between coding sequence divergence and expression divergence, at least in the early stages following the duplication event. Use of different measures for protein sequence divergence may have a marked effect on statistical power in detecting association between similarity in protein sequence of ancient DGD pairs and their genetical control.

Discussion

It is widely believed that expression divergence of duplicate genes is a key step in their retention and evolution of new function. However, the current literature shows relatively little progress has been made to unravel the complexity of the evolutionary mechanisms underlying expression divergence of duplicate genes despite the accumulation of extensive genomics data sets in the yeast *S. cerevisiae*. In particular, Brem and Kruglyak (2005) generated genome-wide gene expression and molecular marker genotype data for a population of segregants derived from a cross between 2 yeast strains. Furthermore, Kellis et al. (2004) have rigorously verified the occurrence of a whole-genome duplication event in the yeast genome approximately 100 MYA. In combination, these data sets presented a unique opportunity to investigate the role of both *cis*-regulatory motifs and genetical control of expression in the global expression divergence of yeast duplicate genes.

We demonstrated that approximately one quarter of duplicate pairs from genome duplication share varying numbers of *trans* regulators, even after having experienced 100 Myr of divergent evolution. A higher proportion (38%) shares varying numbers of *cis*-regulatory motifs, showing that duplicate genes diverge more slowly in motif structure than in terms of *trans*-acting expression regulators. By implication, only small changes in *cis*-regulatory structure may be sufficient to lead to considerable or even complete divergence in expression regulation. Interestingly, however, divergence in *cis* motif structure and divergence in genetical control of expression do not evolve in the same way in duplicate genes of any age or origin. As discovered by Yvert et al. (2003), *trans*-regulatory variation is broadly dispersed across classes of genes with different molecular functions, with no enrichment for TFs. Many *trans* regulators are likely to indirectly exert effects on the transcriptional regulatory complex that assembles in the region of *cis*-regulatory motifs, explaining the absence of correlation between the 2 aspects of divergence.

Previous attempts to unravel the expression divergence of yeast duplicate genes (Papp et al. 2003; Zhang et al. 2004) have shown that expression similarity between duplicate pairs is marginally correlated with shared *cis* motif content; however, these studies could not distinguish between duplicate pairs derived from individual duplication events and those derived from genome duplication. We have gone one step further to divide yeast duplicates accordingly and also to distinguish duplicate pairs with no other paralogues in the genome from multigene families. Intriguingly, we find that despite a more recent origin, duplicate pairs within families are not more likely to share *cis* motif structure than duplicate pairs from ancient genome duplication. For some gene families, there may be less selective constraint because there are more gene copies and therefore a greater plasticity for functional compensation in the event of mutation (Gu and Steinmetz 2003). Conversely, duplicate pairs with no other paralogues may diverge comparatively slowly, and so they remain moderately similar in terms of *cis* motif structure despite a more ancient origin, which can favor genetic robustness against mutations (Kafri et al. 2005). The nature of the duplication event in addition to its age may be an important factor in the divergence variation between duplicate pairs from whole-genome duplication (DGD) and pairs from small-scale duplication events (Guan et al. 2007).

Our study has shown that the expression variation between duplicate copies from genome duplication is explained in part by *cis* motif divergence (~8%) and by *trans*-regulatory divergence (~10%). Our estimates are expected to be robust because the duplicate genes have been rigorously verified to have derived from a genome duplication event approximately 100 MYA (Kellis et al. 2004) and are thus of exactly the same evolutionary age. Therefore, the observed patterns of duplicate gene divergence do not suffer from the potential noise that has thwarted previous studies (Papp et al. 2003; Maslov et al. 2004; Gu et al. 2005). The sequence-based nature of previous analyses may be questionable in dating gene duplications because estimates of duplication age can be biased by many factors, including codon usage, functional constraints, and gene conversion (Lin et al. 2006). We therefore expect that our analysis has provided more stringent information on the divergence pattern of *cis*-regulatory motifs and expression of duplicate genes at a genome scale.

There are various reasons why the expression variation of duplicate copies cannot be fully explained by the 2 aspects of *cis* motif divergence and divergence in *trans*-genetical control. Duplicate genes with divergent regulatory structures may have similar expression profiles due to convergent evolution of TFs. Similarly, duplicate pairs with divergent regulation may have similar expression profiles due to different regulators having a similar impact on gene expression. At the opposite extreme, duplicates with a similar set of motifs

and/or expression regulators may have low expression similarities because of the context dependence of transcriptional regulation, in which motif–motif interactions may be a crucial factor (Wray et al. 2003). Furthermore, compensatory mutational changes in other regulatory elements could mean that regulatory regions gradually change in the duplicates without substantial loss of regulatory binding sites and regulators (Ludwig et al. 2000). In respect of these issues, searching for epistatic interactions among genetic loci, observed for over half of all yeast transcripts (Brem and Kruglyak 2005), could fruitfully be incorporated into the analysis strategies applied in the present study.

The divergence patterns for the genetical control of duplicate genes observed in the present study are dependent on the gene expression data used, which related only to a single growth condition. Indeed, it is possible that some genes may only show divergent gene expression patterns under particular conditions. In this respect, the yeast data set analyzed here was ideally suited to our purposes because it measured gene expression under logarithmic phase growth, in which we would expect most genes to show their most abundant expression. A small mapping population of 112 segregants might be recognized as a limit to achieving sufficient statistical power for detecting regulatory QTL for some genes. However, a population of this size is expected to have adequate power to detect eQTLs given the high levels of heritability (on average 84%) for parental differences in expression estimated for yeast transcripts (Brem et al. 2002). Indeed, the CIM algorithm that we modified to take appropriate account of missing marker data could identify eQTLs for >98% of duplicate pairs from whole-genome duplication.

Yeasts provide a powerful model system for comparative genomics. Previous studies have reconstructed the phylogeny of each duplicate gene family using the sequences of bacteria and archae (Zhang et al. 2004; Gu et al. 2005) to compute the relative age of the duplication event for each pair of duplicate genes, with the bacteria/yeast split as the time unit. With the recent availability of genome sequence data of other *Saccharomyces* species, one could use the genome sequence of *S. cerevisiae* and its relatives to time the small-scale duplication events by "mapping" the timings of recent duplication events onto the species tree (Gao and Innan 2004), for example, using the Yeast Gene Order Browser (Byrne and Wolfe 2005), and investigate alternative proxies for duplication age.

In conclusion, we have shown that divergence in both *cis* motif structure and *trans*-genetical control of expression is an important factor underlying the expression divergence of yeast duplicate genes arising from a whole-genome duplication event. However, it would be over simplistic to assume that the divergence pattern of duplicate gene expression is dominated by the status (presence or absence) of *cis*-regulatory motifs or *trans*-acting expression regulators. Although the present study has revealed a statistically insignificant correlation between these 2 genetic factors in affecting expression divergence of duplicate genes, this does not suffice to neglect the significance of a *cis*-by-*trans* term in explaining the divergence. However, to achieve insightful knowledge about how the *cis* or *trans* TFs regulate duplicate genes independently or interactively, one needs first to dissect the eQTLs on a much finer scale such that actual regulators can be identified and, second, to integrate the genetical genomics analysis with analysis of expression network and information of functional pathways involving duplicate paralogues. In addition, other factors such as mRNA stability may be equally important, as may the local chromatin environment (Cohen et al. 2000), given that the 2 copies of a duplicate pair from genome duplication are usually located on separate chromosomes, which may differ in both chromatin structure and recombination rate (Zhang and Kishino 2004a, 2004b). In future, it will be necessary to identify and account for such factors for a more complete understanding of the expression divergence of yeast duplicate genes.

Supplementary Material

Supplementary figures 1–7 and tables 1 and 2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Dr Leonid Kruglyak for allowing us to share the data sets of expression profiled on a segregating yeast population and their parental strains, Dr Rachel Brem in the Kruglyak group for her patience in explaining the data development, and Dr Joshua Rest for the updated *cis*-regulatory motif data. We also thank one of 2 anonymous reviewers for guiding us to consider use of K_A as an alternative metric for protein sequence similarity in investigating its association with *trans* eQTLs shared by DGD pairs and the other for highlighting a possible role of *cis*-by-*trans* interactions in explaining the divergence pattern. This study is supported by research grants from the Biotechnology and Biological Sciences Research Council and the National Environment Research Council of the United Kingdom to Z.W.L. and M.J.K. Z.W.L. is also supported by the National Natural Science Foundation (30430380) and Basic Research Program of China (2004CB518605). Z.Z. is supported by the Chinese "863" Hi-tech program.

Literature Cited

Aach J, Rindone W, Church G. 2000. Systematic management and analysis of yeast gene expression data. Genome Res. 10:431–435.

Al-Shahrour F, Minguez J, Vaquerizas L, Conde L, Dopazo J. 2005. Babelomics: a suite of web-tools for functional annotation and analysis of groups of genes in high-throughput experiments. Nucleic Acids Res. 33:W460–W464.

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 57:289–300.

Brem R, Kruglyak L. 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci USA. 102:1572–1577.

Brem R, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. Science. 296:752–755.

Byrne K, Wolfe K. 2005. The yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 15:1456–1461.
Cheung V, Conlin L, Weber T, Arcaro M, Jen K-Y, Morley M, Spielman R. 2003. Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet. 33:422–425.
Cohen B, Mitra R, Hughes J, Church G. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet. 26:183–186.
Conover W. 1980. Practical nonparametric statistics. New York: John Wiley & Sons.
Davis J, Petrov D. 2005. Do disparate mechanisms of duplication add similar genes to the genome? Trends Genet. 21:548–551.
Draper NR, Smith H. 1981. Applied regression analysis. New York: John Wiley & Sons.
Erb I, van Nimwegen E. 2006. Statistical features of yeast's transcriptional regulatory code. IEE Proceedings first International Conference on Computational Systems Biology (ICCSB 2006), Shanghai, China. p. 111–118.
Ferris S, Whitt G. 1979. Evolution of the differential regulation of duplicate genes after polyploidization. J Mol Evol. 12:267–317.
Force A, Lynch M, Pickett F, Amores A, Yan Y-L, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531–1545.
Gao L-Z, Innan H. 2004. Very low gene duplication rate in the yeast genome. Science. 306:1367–1370.
Gu Z, Nicolae D, Lu H, Li W-H. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet. 18:609–613.
Gu Z, Steinmetz L. 2003. Role of duplicate genes in genetic robustness against null mutations. Nature. 421:63–66.
Gu X, Zhang Z, Huang W. 2005. Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proc Natl Acad Sci USA. 102:707–712.
Guan Y, Dunham M, Troyanskaya O. 2007. Functional analysis of gene duplications in *Saccharomyces cerevisiae*. Genetics. 175:933–943.
Harbison C, Gordon D, Lee T, et al. (20 co-authors). 2004. Transcriptional regulatory code of a eukaryotic genome. Nature. 431:99–104.
He X, Zhang J. 2005. Gene complexity and gene duplicability. Curr Biol. 15:1016–1021.
Kafri R, Bar-Even A, Pilpel Y. 2005. Transcription control reprogramming in genetic backup circuits. Nat Genet. 37:295–299.
Kellis M, Birren B, Lander E. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature. 428:617–624.
Kellis M, Patterson N, Endrizzi M, Birren B, Lander E. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature. 423:241–254.
Lander E, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121:185–199.
Li W-H. 1997. Molecular evolution. Sunderland (MA): Sinauer Associates.
Li W-H, Yang J, Gu X. 2005. Expression divergence between duplicate genes. Trends Genet. 21:602–607.
Lin Y-S, Byrnes J, Hwang J-K, Li W-H. 2006. Codon-usage bias versus gene conversion in the evolution of yeast duplicate genes. Proc Natl Acad Sci USA. 103:14412–14416.
Little R. 1992. Regression with missing X's: a review. J Am Stat Assoc. 87:1227–1237.
Ludwig M, Bergman C, Patel N, Kreitman M. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature. 403:464–467.
Maslov S, Sneppen K, Eriksen K, Yan K-K. 2004. Upstream plasticity and downstream robustness in evolution of molecular networks. BMC Evol Biol. 4:9.
Morley M, Molony C, Weber T, Devlin J, Ewens K, Spielman R, Cheung V. 2004. Genetic analysis of genome-wide variation in human gene expression. Nature. 430:743–747.
Musso G, Zhang Z, 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. London: Allen and Unwin.
Pachkov M, Erb I, Molina N, van Nimwegen E. 2007. SwissRegulon: a database of genome-wide annotations of regulatory sites. Nucleic Acids Res. 35:D127–D131.
Papp B, Pal C, Hurst L. 2003. Evolution of *cis*-regulatory elements in duplicated genes of yeast. Trends Genet. 19:417–422.
Pilpel Y, Sudarsanam P, Church G. 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet. 29:153–159.
Schadt E, Monks S, Drake T, et al. (14 co-authors). 2003. Genetics of gene expression surveyed in maize, mouse and man. Nature. 422:297–302.
Siddharthan R, Siggia E, van Nimwegen E. 2005. Phylogibbs: a gibbs sampling motif finder that incorporates phylogeny. PLoS Comput Biol. 1:e67.
Snedecor G, Cochran W. 1967. Statistical methods. Aimes (IA): Iowa State University Press.
Thompson J, Higgins G, Gibson T. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
Wagner A. 2000. Decoupled evolution of coding region and mRNA expression patterns after gene duplication. Proc Natl Acad Sci USA. 97:6579–6584.
Wang S, Basten C, Zeng Z. 2006. Windows QTL Cartographer 2.5. Raleigh (NC): Department of Statistics, North Carolina State University.
Wray G, Hahn M, Abouheif E, Balhoff J, Pizer M, Rockman M, Romano L. 2003. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol. 20:1377–1419.
Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 17:32–43.
Yvert G, Brem R, Whittle J, Akey J, Foss E, Smith E, Mackelprang R, Kruglyak L. 2003. *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. Nat Genet. 35:57–64.
Zeng Z. 1994. Precision mapping of quantitative trait loci. Genetics. 136:1457–1468.
Zhang Z, Gu J, Gu X. 2004. How much expression divergence after yeast gene duplication could be explained by regulatory motif evolution? Trends Genet. 20:403–407.
Zhang Z, Kishino H. 2004a. Genomic background predicts the fate of duplicated genes: evidence from the yeast genome. Genetics. 166:1995–1999.
Zhang Z, Kishino H. 2004b. Genomic background drives the divergence of duplicated amylase genes at synonymous sites in *Drosophila*. Mol Biol Evol. 21:222–227.
Zhu J, Zhang M. 1999. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. Bioinformatics. 15:607–611.

Jeffrey Thorne, Associate Editor

Accepted September 5, 2007