References

1 Claverie, J-M. (2001) Gene number: what if there are only 30,000 human genes? *Science* 291, 1255–1257
2 Hartl, D.L. (2000) Molecular melodies in high and low C. *Nat. Rev. Genet.* 1, 145–149
3 Harrison, P.M. *et al*. (2002) A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* 30, 1083–1090
4 Gilbert, W. (1978) Why genes in pieces? *Nature* 271, 501
5 Long, M. *et al*. (1999) Origin of new genes and source for N-terminal domain of the chimerical gene, jingwei, in *Drosophila*. *Gene* 238, 135–141
6 Dupuy, D. *et al*. (2002) SCAN domain-containing 2 gene (*SCAND2*) is a novel nuclear protein derived from the zinc finger family by exon shuffling. *Gene* 289, 1–6
7 Kaessmann, H. *et al*. (2002) Signatures of domain shuffling in the human genome. *Genome Res.* 12, 1642–1650
8 de Souza, S.J. *et al*. (1998) Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5094–5099
9 Long, M. *et al*. (2003) Origin of new genes: evidence from experimental and computational analyses. *Genetica* 118, 171–182
10 Saxonov, S. and Gilbert, W. (2003) The universe of exons revisited. *Genetica* 118, 267–278
11 Fedorov, A. *et al*. (2001) Intron distribution difference for 276 ancient and 131 modern genes suggests the existence of ancient introns. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13177–13182
12 de Souza, S.J. *et al*. (1996) Intron positions correlate with module boundaries in ancient proteins. *Proc. Natl. Acad. Sci. U. S. A.* 93, 14632–14636
13 Doolittle, R.F. (1995) The multiplicity of domains in proteins. *Annu. Rev. Biochem.* 64, 287–314
14 Dorit, R.L. and Gilbert, W. (1991) The limited universe of exons. *Curr. Opin. Genet. Dev.* 1, 464–469
15 Kolkman, J.A. and Stemmer, W.P. (2001) Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.* 19, 423–428
16 Birney, E. *et al*. (2004) An overview of Ensembl. *Genome Res.* 14, 925–928
17 Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* 14, 755–763
18 Bateman, A. *et al*. (2002) The Pfam protein families database. *Nucleic Acids Res.* 30, 276–280
19 Blake, C.C. (1978) Do genes-in-pieces imply proteins-in-pieces? *Nature* 273, 267
20 Patthy, L. (1987) Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett.* 214, 1–7
21 van Rijk, A. and Bloemendal, H. (2003) Molecular mechanisms of exon shuffling: illegitimate recombination. *Genetica* 118, 245–249
22 Eickbush, T. (1999) Exon shuffling in retrospect. *Science* 283, 1465–1467
23 Roy, S.W. (2003) Recent evidence for the exon theory of genes. *Genetica* 118, 251–266

0168-9525/$ - see front matter
doi:10.1016/j.tig.2004.06.013

How much expression divergence after yeast gene duplication could be explained by regulatory motif evolution?

Zhongqi Zhang, Jianying Gu and Xun Gu

Department of Genetics, Development & Cell Biology, Center for Bioinformatics and Biological Statistics, Iowa State University, IA 50011, USA

We used the yeast genome sequences of gene families, microarray profiles and regulatory motif data to test the current wisdom that there is a strong correlation between regulatory motif structure and gene expression profile. Our results suggest that duplicate genes tend to be co-expressed but the correlation between motif content and expression similarity is generally poor, only ~2–3% of expression variation can be explained by the motif divergence. Our observations suggest that, in addition to the *cis*-regulatory motif structure in the upstream region of the gene, multiple *trans*-acting factors in the gene network can influence the pattern of gene expression significantly.

Predicting the transcriptional regulation network from genomic data is a major challenge [1–5]. Many computational methods are based on the expression-motif-conservation hypothesis that co-expressed genes are likely to share regulatory motifs that are conserved during evolution [6–14]. Several studies have tested this idea [15,16]. For instance, Yu *et al.* [15] found that (evolutionarily unrelated) yeast gene pairs targeted by shared transcription factors (TFs, also known as target gene pairs) are co-expressed more frequently than expected by chance. One would expect a similar co-expression between duplicated genes because they usually retain similar functions or regulatory patterns.

We used *Saccharomyces cerevisiae* gene families to test this prediction and examined 202 yeast two-member gene families (Box 1). Using methods modified from Yu *et al.* [15], we have shown that 7% of duplicate genes pairs are co-expressed (Box 1), which is approximately seven times greater than random expectation (P-value$=1.9\times10^{-8}$). This can be compared with the results of Yu *et al.* [15] that 'overall, 3.3% TF-target gene pairs are co-expressed, which is four times greater than random expectation.' Our analysis indicates that duplicate gene pairs might have an even higher co-expression level compared with the TF-target gene pairs. Thus, we thought it was worthwhile exploring a new motif-search strategy by combining co-expressed cluster analysis and gene family phylogeny to increase the signal-to-noise ratio.

Corresponding author: Xun Gu (xgu@iastate.edu).

Box 1. Computational analysis for yeast genome datasets

Datasets

(i) The complete sequences for 43 genomes of bacteria, archaea and *Saccaromyces cerevisiae* are available at NCBI (http://www.ncbi.nlm.nih.gov/COG/), where gene families are classified as clusters of orthologous groups (COG). Amino-acid alignments of 202 COGs containing two yeast duplicate genes were downloaded.

(ii) The whole genome sequence of *S. cerevisiae* was downloaded from NCBI (http://www.ncbi.nlm.nih.gov/Ftp/index.html). The upstream regions (500 bp) of each yeast gene in the COG family were extracted. According to the list of 50 'known' regulatory motifs that were compiled by Kellis *et al.* [14] and verified by experimentation (among those 55 motifs, 50 are unique motif sequences), we used Perl to scan along the upstream region of each yeast gene to identify all of the motifs on both strands (Figure I). Kellis *et al.* had also predicted 72 'discovered' motifs, including 30 known motifs. The motif-expression (predicted) correlation is lower because of the apparent prediction errors (not shown).

(iii) A total of 276 microarray data points were collected from the online database [18–22]. As commonly suggested, we used the fold-change after normalization to representing the gene expression level. Because we are focused on the expression divergence between duplication pairs, potential cross hybridization [35] might give us a more conserved result.

(iv) Chromatin immunoprecipitation (ChIP) data [26] were downloaded from the Young laboratory homepage http://web.wi.mit.edu/young/regulator_network/. Each yeast gene in the 202 families was used as target gene to retrieve its corresponding transcription factor(s) from the dataset.

Association study for co-expression pattern of duplicates

The statistical significance of co-expression (top 1% of largest correlation coefficients of all possible gene pairs in the genome) between duplicate genes is calculated using a cumulative binomial distribution:

$$P_{c\geq c_0} = \sum_{c=c_0}^{N} \binom{N}{c} p^c (1-p)^{N-c} \qquad \text{(Eqn I)}$$

N is the total number of gene pairs of interest, c_0 is the number of co-expressed gene pairs and p is the probability of finding a co-expressed gene pair within the whole genome. (For more information, see Ref. [15].)

Weighted correlation

The model for motif (M)-expression (E) regression can be written as:

$$E = M_c\beta + \alpha, M_c = \frac{\sum_{k=1}^{K} \min(C_{1k}, C_{2k})}{\sum_{k=1}^{K} \max(C_{1k}, C_{2k})} \qquad \text{(Eqn II)}$$

where c_{1k} (c_{2k}) is the copy number of motif k in gene 1 (gene 2) of the duplicate pair. In particular, it becomes an unweighted correlation if $c_{ik}=1$, when motif k is found in the upstream region of a gene i or if $c_{ik}=0$, when a motif k is not found.

Identification of potential paralogous motifs

For each motif i, its position in the upstream sequence of gene X is denoted by L_{ix}. Two motifs located upstream of each duplicate X or Y are paralogous if: (i) they are the same type of motifs; (ii) the positions are sufficiently close, that is:

$$|L_{ix} - L_{iy}| < W \qquad \text{(Eqn III)}$$

where W is the window size; and (iii) when BLAST searches are performed with other yeast species [14], both upstream sequences around the binding sites show a high level of conservation (not shown). Typically, we set $W=50$ bp. Several values of window sizes were considered. Although different window sizes gave different numbers of paralogous motifs, few differences were found when computing the correlation between the fraction of paralogous motifs and the gene expression similarity (not shown).

Two-member gene families

We conducted the following analysis for each gene family. First, we computed the fraction of shared motifs between duplicates. Second, we calculated the standardized expression covariance between duplicate genes. Third, we estimated the amino-acid distance between duplicate genes using the poisson correction. Because many gene duplications are ancient, synonymous nucleotide distance might not be appropriate.

In addition, to estimate the age of gene duplication event [36], we reconstructed the phylogeny of the gene family using the neighbor-joining (NJ) method (MEGA2, http://www.megasoftware.net/). After carefully excluding the lateral gene transfer events, the linearized NJ tree enabled us to compute the (average) duplication time relative to *Escherichia coli*–yeast split, or the relative age of duplication event (Figure II). Note that the yeast gene families used in our study include a significant portion of ancient gene duplication events (Figure IIa), whereas those events from the 'ancient' genome duplication appear relatively recent in our analysis (Figure IIb). Because the age estimation is approximate and is subject to the violations of molecular clock, lateral gene transfer and so on, we also used the amino-acid distance as an alternative measure for the duplication time. Both measures gave consistent results and support our conclusion.

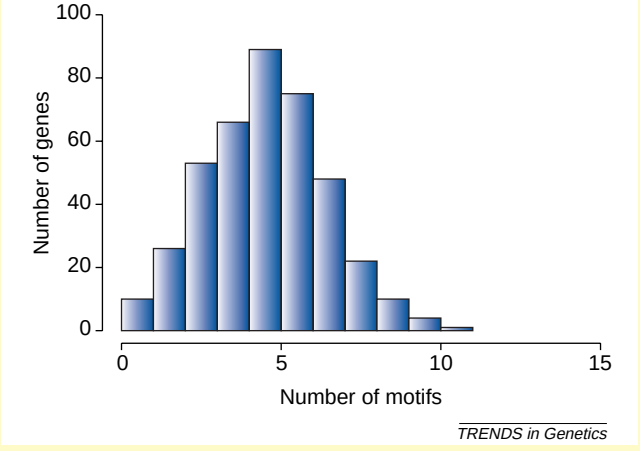

Figure I. The distribution of identified motifs (number of motifs per gene) in 404 yeast genes in our study.

According to the expression-motif-conservation hypothesis, a positive correlation is expected between the fraction of shared motifs and the gene expression similarity for duplicate gene pairs because both are negatively correlated with the age of the duplication event [16,17]. To minimize the error caused by falsely predicted TF-binding sites, we searched the upstream regions of duplicate genes for 50 'known' regulatory motifs that were compiled by Kellis *et al.* [14] and verified by experimentation (Box 1). The mean number of motifs appearing in the upstream region is 5.0 ($\pm$3.6) per gene, whereas the mean number of shared motifs between duplicate genes is 1.8 ($\pm$1.2) per pair. In addition, 183 gene families (91%) have at least one motif with more than one copy; on average, each regulatory motif is present in 1.3 copies in the upstream region. Meanwhile, the expression similarity between two duplicates for each gene family is computed based on 276 microarray data points under various experimental conditions [18–22].

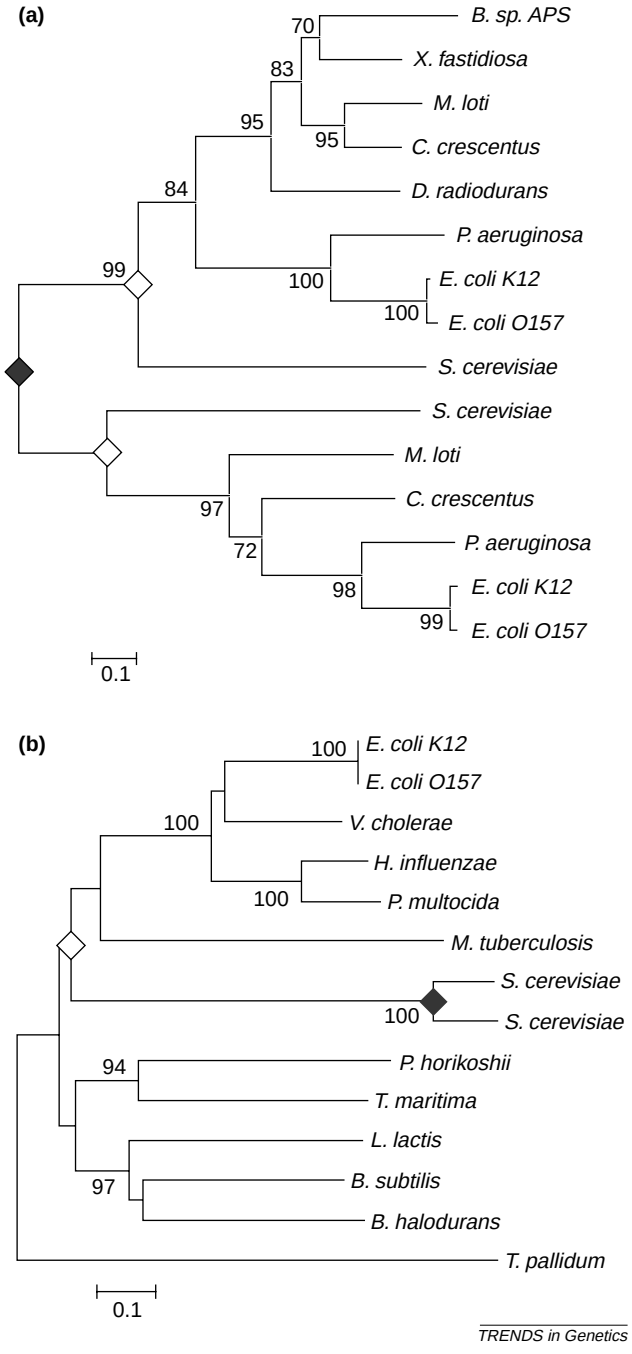

Figure II. The phylogenetic relationships in gene families with ancient or recent yeast gene duplication events. **(a)** The phylogenetic tree of the purine nucleoside phosphorylase gene family (COG0005), showing an ancient yeast gene duplication event. **(b)** The phylogenetic tree of the galactokinase gene family (COG0153), showing a recent yeast gene duplication event. The black diamonds refer to the gene duplication between the yeast genes, and the white diamonds refer to the *Escherichia coli*–yeast split.

Weak correlation between expression divergence and regulatory motif evolution

Figure 1a shows a weak positive correlation between the fraction of shared motifs and the expression similarity (Pearson correlation: $R=0.15$, P-value$=0.032$; Spearman rank correlation: $\rho=0.14$, P-value$=0.048$). We realized that the regulatory motif model might be too simplistic (e.g. one might consider whether the order and proximity of these motifs is important [23,24]). In addition, a small (but certain) portion of these short motifs could be the result of chance. In spite of its complexity, we have tried several approaches to see whether the motif-expression correlation can be improved.

Previous studies showed that multi-copied regulatory motifs are more likely to be active [4,25]. We have considered this effect by using the number of copies as a weighting factor (Box 1), resulting in a slightly higher statistical significance (Pearson correlation: $R=0.17$, P-value$=0.015$; Spearman rank: $\rho=0.15$, P-value$=0.032$).

Moreover, the motif-expression correlation becomes more meaningful for 'paralogous' motifs between duplicate pairs if the relative position of motifs in the regulatory region is taken into account. Tentatively, we defined two similar regulatory motifs located in each duplicate gene as 'paralogous' if: (i) their positions (relative to the transcription site) are close to each other; and (ii) they are in the evolutionary conserved yeast region (Box 1). In total, 93 shared paralogous motifs were detected within 72 gene families. The correlation (Figure 1b) between the fraction of shared paralogous motifs and expression similarity is improved slightly (Pearson correlation: $R=0.18$, P-value 0.013; Spearman rank correlation: $\rho=0.16$, P-value$=0.025$).

Because many potential TF-binding sites might not be functional, it would be beneficial to use those that are detected by large-scale chromatin immunoprecipitation (ChIP) experiments [26]. This complementary approach of determining functional interactions between TFs and upstream regions is not confounded by somewhat diverged motifs. We tested our result by using the dataset from Lee *et al.* (106 putative TFs from their ChIP experiment [26]) and found a higher motif-expression correlation (Pearson correlation $R=0.26$, P-value$=0.002$; Spearman rank correlation: $\rho=0.26$, P-value$=0.001$).

It would be desirable to exploit the assumption that paralogs possessed the same motif structures and expression profiles soon after gene duplication. To avoid the substantial rate variation of amino-acid distance among genes and the saturation of synonymous distance, we used the relative age of duplicates; the units used are defined by the *Escherichia coli*–yeast split (Box 1). Indeed, we have observed that young duplicates have stronger correlation between expression similarity and motif similarity, whereas a similar correlation for ancient duplicate pairs disappears. For instance, using 0.9 units as the cutoff, we divided the duplicate gene pairs into two same-size groups and found that the coefficient of motif-expression correlation for the 'recent' group is 0.19, whereas we found no correlation for the 'ancient' group ($R\approx0$). Note that this pattern is insensitive to the cutoff value (not shown). Adjacent pairs of genes often show expression similarity, even without sharing the same set of regulatory elements [27], because local chromatin structure can also modulate gene expression profiles. However, after excluding a few adjacent duplicate pairs in our analysis, we obtained almost the same pattern (not shown).

We have examined whether the expression similarity and the fraction of shared motifs are negatively correlated with the age of gene duplication [16,17,28]. Figure 1c

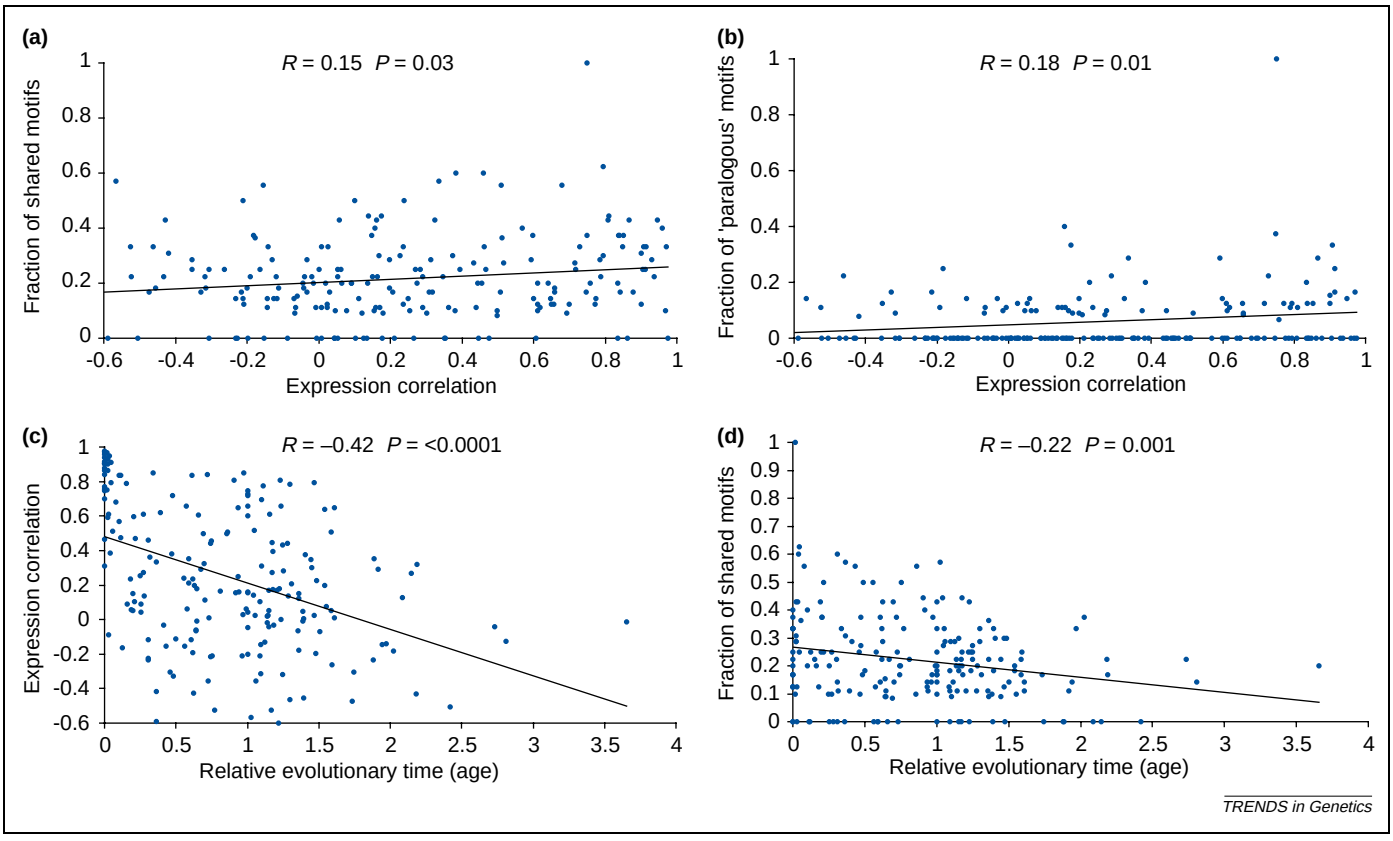

Figure 1. **(a)** A weak but significant (positive) correlation is observed between the fraction of shared motifs and gene expression correlation, showing that the two evolutionary processes are only weakly coupled. **(b)** A similar pattern between the fraction of 'paralogous motifs' and expression correlation. **(c)** A significant (negative) correlation between the gene expression correlation and the relative evolutionary time. **(d)** A significant (negative) correlation between the fraction of shared motifs and the relative evolutionary time.

shows a highly significant negative correlation between the expression correlation and the age of gene duplication (Pearson correlation: R=-0.42, P<0.0001; Spearman rank correlation: ρ=-0.44, P<0.0001). Similarly, Figure 1d reveals a negative correlation between the fraction of shared motifs and the age of gene duplication (Pearson correlation: R=-0.22, P=0.001; Spearman rank correlation: ρ=-0.19, P=0.008). Using amino-acid distance as an estimate of the time of divergence gives almost the same results (not shown).

Gene expression is influenced by many *trans*-factors

One might be puzzled by the discrepancy between the strong co-expression pattern and the relatively weak motif-expression correlation between duplicates. Actually, it can be understood in terms of how much co-expression can be explained by the correlation between expression and motif divergences. Recall that Yu *et al.* [15] claimed that 3.3% of gene pairs that share the same TF(s) are co-expressed, whereas our analysis showed that 7% of duplicate pairs are co-expressed. Meanwhile, the coefficient of correlation between expression divergence and regulatory motif divergence (R=0.15–0.18) suggests that ~2.3%–3.2% (i.e 0.15^2–0.18^2) of the expression variance can be explained by the regulatory motif divergence, even in the case of ChIP, it is only 6.7% (0.26^2). Therefore, the interpretation from these two studies is biologically consistent, although the significance level (P-value) might differ dramatically because the null hypotheses of two methods, and the sample size, are different. One explanation might involve the non-linear co-expression property or the shared motif combinations rather than individual motifs [16,29] – do the older duplicates have a lower expression similarity because they have lost particular motif combinations? We have studied a more general linear or non-linear (e.g. logit) regression model but we found no considerable improvement, partly because of the small sample size. One solution is to study the motif-expression correlation in the phylogeny of a large gene family, which can be tested by using more-sophisticated statistical methods [30].

Caveats notwithstanding, our analysis shows that experimental error is unlikely to be the only cause for the low explanatory power, supporting the notion that gene expression can be influenced by many *trans*-factors at different levels of gene networks. In addition to the *cis*-regulatory motif structure, other factors, such as motif–motif interactions, TF co-evolution and the chromatin structure [29,31,32], can affect the expression levels without changing the regulatory motifs.

The weak motif-expression correlation between duplicates might be the consequence of two phenomena. First, duplicates with divergent motif structures actually have similar expression profiles (e.g. among the ten most highly co-expressed duplicate gene pairs none of which has a fraction of shared motifs >50%). This could be due to the neutral turnover of binding sites (i.e. randomly generated binding sites might cause the shift of TF from its original

binding site) or convergent evolution (i.e. the binding between different TFs and their corresponding motifs might have a similar impact on gene expression regulation). Second, duplicates with a similar set of regulatory motifs have low expression similarities because of the context-dependence of *trans*-transcriptional regulation, which requires the combinatorial interaction of other bound proteins to function [29]. Further study is certainly needed when more data are available. However, the evolution of regulatory motifs might not necessarily result in expression divergence because of genetic robustness or because of an alternative regulatory pathway in which the original binding sites were not used. Furthermore, our analysis implies that ancient duplicate pairs might have more chances to be affected by the reshuffling of gene networks. Consequently, the motif-expression correlation might disappear because of the evolutionary reorganization of gene networks, resulting in considerable expression changes even when the same motifs are conserved [29]. In addition, there might be other *trans*-acting factors (other than the TFs) that have an influence on gene regulation [33].

Concluding remarks

If our interpretation is correct, some evolutionary models need to be revisited because the whole gene network could be involved in shaping the expression divergence after gene duplication. For example, the duplication-degeneration-complementation (DDC) model of gene duplication [34] assumes that the status (presence or absence) of regulatory motifs (or motif modules) dominates the status of gene expression, which seems to be oversimplified according to our analysis. (For more information, see Wagner [28].) The important information conveyed by our results is that the identification of motif structures is only one step towards discovering the mechanism underlying gene expression regulation and evolution, and the effect of gene networks other than the *cis*-factors should not be neglected.

Acknowledgements

We thank Molley Vanhouten for assistance in collecting the microarray gene expression data. This work was supported by an NIH grant to X.G.

References

1 Ren, B. *et al*. (2000) Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309
2 Ioshikhes, I.P. and Zhang, M.Q. (2000) Large-scale human promoter mapping using CpG islands. *Nat. Genet.* 26, 61–63
3 Workman, C.T. and Stormo, G.D. (2000) ANN-Spec: A method for discovering transcription factor binding sites with improved specificity. *Pac. Symp. Biocomput.* 5, 464–475
4 Guhathakurta, D. and Stormo, G.D. (2001) Identifying target sites for cooperatively binding factors. *Bioinformatics* 17, 608–621
5 Blanchette, M. and Tompa, M. (2002) Discovery of regulatory elements by a computational method for phylogenetic footprinting. *Genome Res.* 12, 739–748
6 Roth, F.R. *et al*. (1998) Finding DNA regulatory motifs within unaligned non-coding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* 16, 939–945
7 Bussemaker, H.J. *et al*. (2000) Building a dictionary for genomes: identification of presumptive regulatory sites by statistical analysis. *Proc. Natl. Acad. Sci. U. S. A.* 97, 10096–10100
8 Helden, J. *et al*. (2000) Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic Acids Res.* 28, 1808–1818
9 Thijs, G. *et al*. (2001) A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. *Bioinformatics* 17, 1113–1122
10 Pilpel, Y. *et al*. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* 29, 153–159
11 McCue, L.A. *et al*. (2001) Phylogenetic footprinting of transcription factor binding sites in proteobacterial genomes. *Nucleic Acids Res.* 29, 774–782
12 Thijs, G. *et al*. (2002) A Gibbs sampling method to detect over-represented motifs in the upstream regions of coexpressed genes. *J. Comput. Biol.* 9, 447–464
13 Liu, X.S. *et al*. (2002) An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat. Biotechnol.* 20, 835–839
14 Kellis, M. *et al*. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241–254
15 Yu, H. *et al*. (2003) Genomic analysis of gene expression relationship in transcriptional regulatory networks. *Trends Genet.* 19, 422–427
16 Papp, B. *et al*. (2003) Evolution of cis-regulatory elements in duplicated genes of yeast. *Trends Genet.* 19, 417–422
17 Gu, Z. *et al*. (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* 18, 609–613
18 DeRisi, J.L. *et al*. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686
19 Chu, S. *et al*. (1998) The transcriptional program of sporulation in budding yeast. *Science* 282, 699–705
20 Spellman, P.T. *et al*. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces* cerevisiae by microarray hybridization. *Mol. Biol. Cell.* 9, 3273–3297
21 Gasch, A.P. *et al*. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell.* 11, 4241–4257
22 Lyons, T.J. *et al*. (2000) Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7957–7962
23 Keleş, S. *et al*. (2002) Identification of regulatory elements using a feature selection method. *Bioinformatics* 18, 1167–1175
24 Mose, A.M. *et al*. (2003) Position specific variation in the rate of evolution in transcription factor binding sites. *BMC Evol. Biol.* 3, 19
25 Wagner, A. (1999) Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. *Bioinformatics* 15, 776–784
26 Lee, T.I. *et al*. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804
27 Cohen, B.A. (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* 26, 183–186
28 Wagner, A. (2000) Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6579–6584
29 Wray, G.A. *et al*. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419
30 Gu, X. (2004) Statistical framework for phylogenetic analysis of gene family expression profiles. *Genetics* 167, 531–542
31 Kel, O.V. *et al*. (1995) A compilation of composite regulatory elements affecting gene transcription in vertebrates. *Nucleic Acids Res.* 23, 4097–4103
32 Zhou, Y-H. *et al*. (2002) Novel PAX6 biding sites in the human genome and the role of repetitive elements in the evolution of gene regulation. *Genome Res.* 12, 1716–1722
33 Yvert, G. *et al*. (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* 35, 57–64
34 Force, A. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545
35 Evertsz, E.M. *et al*. (2001) Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. *Biotechniques* 31, 1182–1192
36 Gu, X. *et al*. (2002) Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat. Genet.* 31, 128–129

0168-9525/$ - see front matter
doi:10.1016/j.tig.2004.07.006