Coexpression of Linked Gene Pairs Persists Long after Their Separation

Guang-Zhong Wang1, Wei-Hua Chen2, and Martin J. Lercher*1

1Institute for Computer Science, Heinrich-Heine-University, 40225 Düsseldorf, Germany
2European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

*Corresponding author: E-mail: lercher@cs.uni-duesseldorf.de.

Accepted: 20 May 2011

Abstract

In many eukaryotes, physically linked gene pairs tend to be coexpressed. However, it is still controversial to what extent this neighbor coexpression is maintained by selection and to what extent it is nonselective, purely mechanistic “leaky expression.” Here, we analyze expression patterns of gene pairs that have lost their linkage in the evolution of Saccharomyces cerevisiae since its last common ancestor with Kluyveromyces waltii or that were never linked in the S. cerevisiae lineage but became neighbors in a related yeast. We demonstrate that coexpression of many linked genes is retained long after their separation and is thus likely to be functionally important. In addition, unlinked gene pairs that recently became neighbors in other yeast species tend to be coexpressed in S. cerevisiae. This suggests that natural selection often favors chromosomal rearrangements in which coexpressed genes become neighbors. Contrary to previous suggestions, selectively favorable coexpression appears not to be restricted to bidirectional promoters.

Key words: coexpression, genomic neighbors, chromatin regulation, yeast.

Introduction

A gene’s expression pattern is influenced by its genomic location, both in prokaryotes and eukaryotes. In prokaryotes, neighboring genes often form operons, resulting in tight coexpression of neighboring genes. In eukaryotes, physically linked gene pairs also show higher coexpression than randomly chosen gene pairs (Cohen et al. 2000; Kruglyak and Tang 2000; Lercher et al. 2002, 2003; Williams and Hurst 2002; Hurst et al. 2004; Singer et al. 2005; Lercher and Hurst 2006; Semon and Duret 2006; Batada et al. 2007; Kensche et al. 2008). For example, in yeast, adjacent gene pairs show correlated expression regardless of their relative orientation (Cohen et al. 2000; Kruglyak and Tang 2000), and this coexpression relationship spans up to 30 neighboring genes (Lercher and Hurst 2006). In the worm Caenorhabditis elegans, many coexpressed genes are organized into operons (Lercher et al. 2003). In the mouse genome, both immune system genes and tissue-specific genes are found to be expressed in clusters (Williams and Hurst 2002). In the human genome, housekeeping genes also show strong clustering (Lercher et al. 2002). Based on their apparent evolutionary conservation, it has been proposed that such coexpression clusters are selectively favorable in mammals (Singer et al. 2005). However, a later report found that highly coexpressed gene pairs are more likely to be broken up by rearrangements, concluding that neighbor coexpression is in fact generally disadvantageous in mammals (Liao and Zhang 2008).

The coexpression of neighboring genes in prokaryotic operons is conceptually simple. In eukaryotes, a range of mechanisms has been proposed to be responsible for the coexpression of closely spaced genes. Coexpressed neighboring genes in divergently transcribed orientation suggest that bidirectionally active promoters play a role in regulating coexpression (Cohen et al. 2000; Kruglyak and Tang 2000; Kensche et al. 2008), although such “bipromoters” may also serve to reduce stochastic gene expression noise (Wang et al. 2011). Chromatin structure also likely has an impact on the coexpression of closely located genes (Hurst et al. 2004; Batada et al. 2007; Chen et al. 2010). Finally, gene pairs that share the same transcription factors or that may be prone to a failure of transcription termination (“transcriptional read-through”) were also reported to be responsible for coexpression of neighboring genes (Semon and Duret 2006; Batada et al. 2007; Michalak 2008). Neighboring genes with similar functions have lead to the proposal...
that the coexpression of linked genes may be related to
gene function (Cohen et al. 2000; Michalak 2008).
Thus, neighboring eukaryotic genes tend to be coex-
pressed. But is this coexpression really selectively favorable,
or is it a nonselective, purely mechanistic by-product of ge-
nomic neighborhood, as suggested at least for mammals
(Liao and Zhang 2008)? If neighbor coexpression is indeed
functional, then coexpression should be maintained even if
the neighborhood is broken up by genomic rearrangements.
Here, we compare the effects of current and ancestral gene
order on current gene expression patterns in the yeast Sac-
charomyces cerevisiae. In particular, we show that gene
pairs that were genomic neighbors in the evolutionary past,
but are separated now, show higher coexpression than ran-
domly chosen gene pairs. As nonselective neighbor coexpres-
sion should seize after breaking up the neighborhood, our
results indicate a significant role of natural selection in the
coeexpression of linked yeast genes.

Materials and Methods

Data Sources

The S. cerevisiae gene order as well as the ancestral gene
order were taken from the yeast gene order database (Byrne
and Wolfe 2005). We only retained genes with known po-
sitions in both data sets for further analysis.

Saccharomyces cerevisiae genome sequence data were
downloaded via ftp from the Saccharomyces Genome
Database (ftp://genome-ftp.stanford.edu/pub/yeast/). An-
cestral gene order from eight reconstructed chromosomes
was obtained from Gordon et al. (2009). Information on
gain and loss of neighborhood along the yeast phylogeny
was taken from Kensche et al. (2008).

The divergent gene pairs that share a promoter (biprom-
ooter pairs) were taken from genome-wide tiling array ex-
periments (Xu et al. 2009).

Expression Data

We employed the same expression data used by Batada,
Ut-
rutia, and Hurst to assess the influence of chromatin remod-
eling on the coexpression of neighboring yeast genes
(Batada et al. 2007). Briefly, coexpression was averaged
across 23 large-scale time course messenger RNA (mRNA)
expression data sets, each covering at least 10 different time
points. For each data set, Pearson’s product moment corre-
lation coefficient was calculated between the expression
vectors of any two genes; for a given gene pair, coexpression
was then defined as the mean value across data sets (for
details, see supplementary information 1 of Batada et al.
2007).

Some of these data sets were obtained using cDNA mi-
croarrays, which may have spotted chromosomal neigh-
ors onto neighboring microarray spots. It is hence possible that
coeexpression of genes neighboring in the current S. cerevi-
siae genome was overestimated due to experimental arti-
facts (Lercher and Hurst 2006). To address this issue, we
repeated part of our analyses using coexpression derived
from a set of 1,370 Affymetrix microarray experiments from
the National Center for Biotechnology Information Gene Ex-
pression Omnibus (GEO) database (GEO accession IDs are
listed in supplementary table S1, Supplementary Material
online). We renormalized the log2-transformed expression
coexpression between two genes was again calculated as Pear-
son’s product moment correlation coefficient across
experiments. Results based solely on Affymetrix microar-
rays were qualitatively very similar to those presented in
the main text (see supplementary results S1, Supplemen-
tary Material online).

Tandemly duplicated genes can lead to overestimation of
the coexpression of neighboring genes. To avoid biases
caused by tandem duplications, we removed all such pairs
from our analyses. Tandem duplicates were identified as
neighbors in the S. cerevisiae genome with Blast e value
<0.01 (Batada et al. 2007).

Evolutionary Conservation of Bipromoter Gene
Pairs

To test if bipromoter gene pairs are more conserved than other
divergent gene pairs, we employed the reconstructed ances-
tral gene order published by the Wolfe lab (http://wolfe.-
gen.tcd.ie/ygob/). Only genes annotated in both the
S. cerevisiae genome and the ancestor genome were used.
Divergent gene pairs in the S. cerevisiae genome were marked
as “ancestral” if they were direct chromosomal neighbors in
the reconstructed ancestor and as “new” otherwise.

A simple logistic regression model was utilized to deter-
mine if bipromoter gene pairs are still more conserved than
non-bipromoter pairs after controlling for coexpression level
and intergenic distance, which is the strongest known pre-
dictor for linkage breakup (Poyatos and Hurst 2007). We
used the following model:

\[ z \sim orient + coexpr + igd, z \sim orient + coexpr + igd, \]

with

\[ z: 1 = ancestral pair, 0 = new pair; \]

orient: 0 = bipromoter divergent pair, 1 = non-
bipromoter divergent pair;

coexpr: coexpression level;

igd: intergenic distance.

Intergenic distance was measured as the distance in base
pairs between the transcription start sites of genes. Calcu-
lations were performed in the R environment for statistical
computing. The model shows that after controlling for co-
expression and intergenic distance, bipromoter status still
has a highly significant effect on gene pair conservation \((P = 2.9 \times 10^{-7})\); the effect of orientation was indeed much stronger than the influence of the other two variables (data not shown).

### Dollo Parsimony Method to Calculate Ancestral States

Phylogenetic relationships of 19 yeasts and the neighboring gene pairs of orthologs in these fungi were downloaded from the supplementary file of Kensche et al. (2008). We used the Dollo parsimony method implemented in PAUP* (Wilgenbusch and Swofford 2003) to calculate the ancestral neighborhood state of each gene pair. This algorithm provided us with gain and loss information of gene neighborhood relationships at each internal node.

### Correlation between Age of Separation and Coexpression

To test if coexpression is higher for ancestrally linked gene pairs that were separated more recently, we also used the information on ancestral gene order and separation ages derived from Kensche et al. (2008) as described above. We did not find a significant correlation between separation age and coexpression level (Pearson’s product moment correlation coefficient: \(R = 0.087\), \(P = 0.14\); Spearman’s rank correlation coefficient: \(p = -0.044\), \(P = 0.46\).

### Neighborhood Conservation Predicts Coexpression only for Divergent Gene Pairs

We used the recently reconstructed gene order of the pre-whole-genome duplication yeast ancestor (Gordon et al. 2009), which is believed to be about 100–150 My old (Sugino and Innan 2005). We first compared the coexpression of gene pairs that are conserved between the ancestor and \(S.\ cerevisiae\) with the coexpression of gene pairs newly formed in \(S.\ cerevisiae\). Here, coexpression of two genes is defined as the correlation of gene expression values across a large data set of time series experiments (Kafri et al. 2005; Batada et al. 2007). To avoid potential biases caused by tandemly duplicated genes, such gene pairs were removed prior to all analyses.

Three possible scenarios exist: 1) If the conserved gene pairs are less likely to be coexpressed compared with newly formed gene pairs, then highly coexpressed neighboring gene pairs may be generally disadvantageous, as was observed recently in mammals (Liao and Zhang 2008). 2) If conserved gene pairs share similar coexpression profiles with newly formed gene pairs, then neighbor coexpression is likely to be largely selectively neutral. 3) If the conserved gene pairs generally show higher coexpression levels compared with newly formed gene pairs, then this indicates that neighbor coexpression is generally advantageous, as previously suggested (Singer et al. 2005).

Table 1 shows the results of this comparison. For divergently oriented \(S.\ cerevisiae\) gene pairs (\(\rightarrow\rightarrow\)), those that were already in this orientation in the ancestral genome show higher coexpression compared with newly formed divergent gene pairs. No such difference between conserved and new pairs was found for convergent or cooriented gene pairs. This indicates that in yeast, only divergent gene pairs are under selection for high coexpression. Surprisingly, there is no difference between the coexpression of newly formed divergent gene pairs and convergent gene pairs \((P = 0.59\) comparing new divergent gene pairs with conserved convergent gene pairs \(P = 0.59\) comparing new divergent gene pairs with newly formed convergent gene pairs, Brunner–Munzel tests). Thus, divergent gene pairs do not always show higher coexpression compared with other types of adjacent gene pairs in yeast. These results are not a consequence of variation in intergenic distance, which is known to be the strongest predictor of gene neighborhood conservation in yeast (Poyatos and Hurst 2007): the effect of neighborhood conservation status remains qualitatively unchanged when using both conservation status and intergenic distances as predictors in a logistic regression model (table 1).

The observed difference between ancestral and young divergent gene pairs is likely related to the activity of bidirectionally active promoters (up promoters) (Kruglyak and Tang 2000). Gene pairs newly formed by rearrangements will
Gene Pairs that Used to be Neighbors Are Still Coexpressed

We next analyzed the 2,765 ancestrally neighboring gene pairs that are located on different chromosomes in the current *S. cerevisiae* genome. On average, these separated pairs are significantly more coexpressed compared with 10,000 randomly chosen gene pairs \( (P = 5.3 \times 10^{-6}) \), Wilcoxon rank sum test. Except for shared cis-regulatory sites, none of the proposed mechanistic reasons for neighbor coexpression appear capable of explaining the persistence of coexpression after separation. Thus, coexpression is likely selectively favorable for these ancestrally linked gene pairs and was thus kept or restored after their separation.

We do not find significant differences between the coexpression of gene pairs that were ancestrally linked in different relative orientations \( (P = 0.084 \text{ between divergent and convergent pairs, } P = 0.13 \text{ between divergent and cooriented gene pairs, and } P = 0.70 \text{ between cooriented and convergent pairs; Brunner–Munzel tests}) \). Of the known factors that specifically affect the coexpression of genomic neighbors, only chromatin remodeling acts independently of gene orientation. That we find no differences between orientations thus appears consistent with the idea that ancestral coexpression of these pairs was caused by local chromatin remodeling and is now maintained by shared cis-regulatory sequences; these cis-regulatory sequences may affect transcription factor binding as well as local chromatin remodeling. This maintenance further suggests that low-level coexpression caused by chromatin remodeling may in many cases be selectively favorable.

The results presented above suggest that at least part of the high coexpression level of neighboring yeast gene pairs is due to natural selection on coexpression. Thus, genes still need to be coexpressed when pairs are separated through a genomic rearrangement. To further verify this hypothesis, we used recent data based on gene pair conservation across 19 fungi \( (\text{Kensche et al. 2008}) \). We reconstructed the gene order in the common ancestor of these species using Dollo parsimony as implemented in PAUP* \( (\text{Wilgenbusch and Swoford 2003}) \).

We only analyzed genes that were direct neighbors in the ancestral genome but that are now located on different chromosomes because genes located nearby on a yeast chromosome still show similar expression profiles even when separated by tens of genes \( (\text{Lercher and Hurst 2006}) \). As already observed in our first data set, separated gene pairs show slightly higher coexpression compared with random gene pairs \( (\text{fig. 1; } P = 0.00020, \text{ Brunner–Munzel test}) \). Again, there is no difference between the coexpression of divergent, convergent, and cooriented ancestral gene pairs after their separation \( (P = 0.18 \text{ between divergent and convergent pairs, } P = 0.32 \text{ between divergent and cooriented gene pairs, and } P = 0.76 \text{ between cooriented and convergent pairs; Brunner–Munzel tests}) \).

If gene neighborhood is under positive selection for genes that need to be coexpressed, then we would further...
expect that orthologs of non-neighboring coexpressed *S. cerevisiae* genes are more likely to become genomic neighbors in other yeast lineages; consequently, genes neighboring in at least one other yeast species but located on different chromosomes in both *S. cerevisiae* and in the common ancestor should show higher coexpression than random gene pairs. This is indeed the case (fig. 1; \( P = 3.3 \times 10^{-5} \), Brunner–Munzel test).

**Discussion**

Using ancestral gene order information gained from the yeast gene order browser (Byrne and Wolfe 2005), we confirmed that among neighboring gene pairs, divergently oriented pairs are the ones that were most likely to be conserved during genome evolution (Kensche et al. 2008). More specifically, this conservation is mostly due to bipromoter gene pairs. The conservation implicates stabilizing selection on the relative positioning of this subset of the divergently arranged gene pairs.

After separation of neighboring gene pairs through genomic rearrangements, we no longer found any difference between divergent and convergent or cooriented gene pairs; all three types of ancestrally neighboring gene pairs show higher than expected coexpression in *S. cerevisiae* after their separation through genomic rearrangements. It is possible that the two genes in these coexpressed separated pairs had part of their cis-regulatory apparatus in common even before their separation, so that coexpression could be partially maintained after the rearrangement; conversely, it may be that coexpression was initially lost in the rearrangement and was reinstated through cis-regulatory changes afterward.

The coexpression of ancestrally neighboring gene pairs that are now located on different chromosomes is sharply reduced compared with the pairs that are neighbors in the current yeast genome (fig. 1). This observation is expected, as factors such as chromatin remodeling are known to strongly influence the coexpression of linked genes in yeast (Batada et al. 2007). Thus, although part of the neighbor coexpression is likely maintained by natural selection, it is likely that a substantial component of neighbor coexpression is nonselective “leaky” expression of one or both neighbors.

Could it be that all coexpression is in fact nonselective (purely mechanistic), and separated pairs only show coexpression because part of the mechanistic apparatus shared between the two genes is maintained through the separation? In particular, many of the linkage losses may be a consequence of the whole-genome duplication experienced in the *S. cerevisiae* lineage after the common ancestor of the yeasts analyzed here. Assume that the common ancestor contained a neighboring gene pair A,B together with a cis-regulatory region c that affects both genes (c-A-B). The whole-genome duplication will duplicate the complete set, resulting in c1-A1-B1 and c2-A2-B2. If subsequently A1 and B2 (or A2 and B1) are lost, the now separated genes A and B retain their identical cis-regulatory region c.

However, it is unlikely that such a scenario explains our observations, for at least two reasons. First, if selection would play no role in the maintenance of coexpression, then coexpression should fade with increasing age of the separation. This is not the case (Spearman’s \( \rho = -0.044 \), \( P = 0.46 \); for details, see Materials and Methods). Second, and most importantly, a nonselective, purely mechanistic model cannot explain why unlinked gene pairs that only recently became neighbors in other yeast species are coexpressed in *S. cerevisiae*. That these pairs show coexpression very similar to ancestrally linked pairs (fig. 1) seems only compatible with the hypothesis that natural selection promotes chromosomal rearrangements that bring together coexpressed genes.

When discussing the properties of neighboring gene pairs, these are usually classified by their relative orientation into three categories—divergent gene pairs (head to head), convergent gene pairs (tail to tail), and cooriented gene pairs. Those three types of gene pairs appear to have different properties—divergent gene pairs are the most conserved and show stronger coexpression than the other two orientations (Kensche et al. 2008). Here we show that as far as coexpression is concerned, there are essentially only two types of neighboring gene pairs in the genome—bipromoter gene pairs and non-bipromoter gene pairs. Bipromoter gene pairs show strong signals of conservation and coexpression, whereas non-bipromoter gene pairs do not. After separation through genomic rearrangements, ancestral divergent gene pairs no longer exhibit higher coexpression compared with other ancestral gene pairs, supporting the view that chromatin remodeling dominates the coexpression of most neighboring gene pairs (Batada et al. 2007).

In conclusion, we have shown that not only gene neighborhood in the current *S. cerevisiae* genome but also gene order in the ancestral genome and gene order in related yeasts are predictive of coexpression. These results support a role for natural selection in the establishment and maintenance of neighbor coexpression in yeast and argue against a purely mechanistic view that considers neighbor coexpression as a neutral (or even slightly deleterious) phenomenon.

**Supplementary Material**

Supplementary results, table S1, and figure S1 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

**Acknowledgments**

We thank Araxi O. Urrutia for giving us access to her coexpression data.
Literature Cited

Batada NN, Urrutia AO, Hurst LD. 2007. Chromatin remodelling is a major source of coexpression of linked genes in yeast. Trends Genet. 23:480–484.

Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 15:1456–1461.

Chen WH, de Meaux J, Lercher MJ. 2010. Co-expression of neighbouring genes in Arabidopsis: separating chromatin effects from direct interactions. BMC Genomics. 11:178.

Cohen BA, Mitra RD, Hughes JD, Church GM. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet. 26:183–186.

Gentleman RC, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5:R80.

Gordon JL, Byrne KP, Wolfe KH. 2009. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern Saccharomyces cerevisiae genome. PLoS Genet. 5:e1000485.

Hurst LD, Pal C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet. 5:299–310.

Kafri R, Bar-Even A, Pilpel Y. 2005. Transcription control reprogramming in genetic backup circuits. Nat Genet. 37:295–299.

Kensche PR, Oti M, Duret L. 2006. Conservation of divergent transcription in fungi. Trends Genet. 24:207–211.

Kruglyak S, Tang H. 2000. Regulation of adjacent yeast genes. Trends Genet. 16:109–111.

Lercher MJ, Blumenthal T, Hurst LD. 2003. Coexpression of neighboring genes in Caenorhabditis elegans is mostly due to operons and duplicate genes. Genome Res. 13:238–243.

Lercher MJ, Hurst LD. 2006. Co-expressed yeast genes cluster over a long range but are not regularly spaced. J Mol Biol. 359:825–831.

Lercher MJ, Urrutia AO, Hurst LD. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet. 31:180–183.

Liao BY, Zhang J. 2008. Coexpression of linked genes in mammalian genomes is generally disadvantageous. Mol Biol Evol. 25:1555–1565.

Michalak P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics. 91:243–248.

Poyatos JF, Hurst LD. 2007. The determinants of gene order conservation in yeasts. Genome Biol. 8:R233.

Semon M, Duret L. 2006. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. Mol Biol Evol. 23:1715–1723.

Singer GA, Lloyd AT, Hummiecki LB, Wolfe KH. 2005. Clusters of coexpressed genes in mammalian genomes are conserved by natural selection. Mol Biol Evol. 22:767–775.

Sugino RP, Innan H. 2005. Estimating the time to the whole-genome duplication and the duration of concerted evolution via gene conversion in yeast. Genetics. 171:63–69.

Wang GZ, Lercher MJ, Hurst LD. 2011. Transcriptional coupling of neighbouring genes and gene expression noise: evidence that gene orientation and non-coding transcripts are modulators of noise. Genome Biol Evol. 3:320–331.

Wilgenbusch JC, Swoford D. 2003. Inferring evolutionary trees with PAUP*. Curr Protoc Bioinformatics. Chapter 6: Unit 6.4. doi: 10.1002/0471250953.bi0604s00

Williams EJ, Hurst LD. 2002. Clustering of tissue-specific genes underlies much of the similarity in rates of protein evolution of linked genes. J Mol Evol. 54:511–518.

Xu Z, et al. 2009. Bidirectional promoters generate pervasive transcription in yeast. Nature. 457:1033–1037.

Associate editor: George Zhang