Genes Encoding Subunits of Stable Complexes Are Clustered on the Yeast Chromosomes: An Interpretation From a Dosage Balance Perspective

Sarah Amalia Teichmann* and Reiner Albert Veitia†,1

*MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, England and †Université Denis Diderot/Paris VII, INSERM U361 Reproduction et Physiopathologie Obstétricale, Hôpital Cochin, 75014 Paris, France

Manuscript received November 12, 2003
Accepted for publication February 1, 2004

ABSTRACT

Genomic evidence for colocalization of functionally related genes on eukaryote chromosomes is mounting. Here we show that a statistically significant fraction of yeast genes coding for subunits of stable complexes are located within 10–30 kb of each other. Clustering of genes encoding subunits of complexes may ensure better coregulation and maintain the right stoichiometry of complexes upon duplication of chromosomal segments.

CLUSTERING of functionally related genes on chromosomes is widespread in prokaryotes due to the existence of operons, which are rare in eukaryotes (Blumenthal and Spieth 1996). Nevertheless, there is increasing genomic evidence for clusters of functionally related genes in eukaryotes. For instance, in yeast, there is a correlation between the expression patterns of adjacent, as well as nearby nonadjacent pairs of genes, and a significant proportion of adjacent pairs fall into the same functional category (Cohen et al. 2000). Moreover, genes with similar expression profiles are more likely to encode physically interacting proteins (Ge et al. 2001; Jansen et al. 2002). Eukaryotic pathways in general exhibit significantly higher gene clustering than expected by chance (Lee and Sonnhammer 2003). Finally, it is known that two interacting polypeptides in one organism can exist as a single polypeptide in another. A great number of composite proteins in eukaryotes including yeast and higher organisms have been described (Enright and Ouzounis 2001). They can be considered as examples of the maximum degree of clustering (fusion) of the underlying coding sequences.

According to the specific topology of a complex, increased or decreased amounts of one component can diminish the yield of the whole complex, decrease the fitness of the organism, and lead to dominant phenotypes (Veitia 2002, 2003). There is genomic evidence for the dosage balance notion in yeast, as both under- and overexpression of the protein subunits of complexes lower fitness more than that for genes not involved in complexes. Accordingly, when the degree of coexpression is higher, the heterozygote fitness is lower. Moreover, as noted by Papp et al. (2003), the balance theory predicts that many single-gene duplication events can be harmful, as this leads to imbalance. Consequently, members of large gene families are rarely involved in complexes. Here we use the yeast Saccharomyces cerevisiae to explore the possibility that proteins involved in the same stable complex are encoded by chromosomally clustered genes as this is likely to ensure better coregulation.

Clustering of genes encoding complex components is conceivable in evolutionary terms. Selection can establish and conserve a new gene order resulting in clustering. In a sexual population, fixation of a beneficial combination of alleles at two loci (i.e., haplotype) takes place if individuals with this haplotype are the fittest and if both loci are closely linked (Kimura 1956; Bodmer and Parsons 1962). In our context, the advantage can be provided by a tighter coregulation between the subunits. Chromosomal inversions, which are frequent in yeast (Seoighe et al. 2000), can generate a new gene order putting the relevant loci in proximity. This increases linkage and the new gene order can replace the previous one. Note also that if duplication of a multisubunit complex is advantageous, fixation of new gene order (cluster) is accelerated. An alternative mechanism that is worth considering in this context is the nonrandom deletion of genes from paralogous chromosomes after whole-genome duplication (Langkjaer et al. 2003),

1Corresponding author: Université Denis Diderot/Paris VII, INSERM U361 Reproduction et Physiopathologie Obstétricale, Hôpital Cochin, Pavillon Baudelocque, 129 Bd de Port Royal, 75014 Paris, France. E-mail: veitia@cochin.inserm.fr
leading to preferential linkage of interacting gene pairs to better adjust dosage balance.

To investigate whether there is a tendency for genes coding for subunits of stable protein complexes to cluster (according to the MIPS catalog; Mevès et al. 2002), we have extracted and statistically analyzed their chromosomal locations. We also explored genes encoding proteins engaged in transient interactions. We expect the evolutionary effects described above to be more pronounced in stable complexes than in common transient genes and all possible pairs of the whole set of yeast proteins engaged in transient interactions. We expect the bias for genes coding for subunits of stable complexes, located within 10 or 30 kb, and this trend is stronger than that for genes encoding subunits of transient complexes as illustrated in Figure 1 and Table 1.

As a comparison to chromosomal clustering observed for proteins in stable and transient complexes, we calculated the equivalent trends for all possible pairs of essential genes and all possible pairs of the whole set of yeast genes. Pal and Hurst (2003) have shown that essential genes tend to cluster. However, as shown in Table 1 and Figure 1, the trend for chromosomal clustering of subunits of stable complexes is stronger than the effect of clustering of essential gene pairs, which is roughly equal to that for proteins in transient interactions. In the set of 32 pairs of genes within 10 kb that code for proteins in part of the same stable complex, there are 7 pairs in which both genes are essential, showing that the bias for genes coding for subunits of the same complex is not explained by essentiality. Furthermore, none of the pairs in question (essential or not) had >60% identity to each other, so they cannot be considered recent duplicates (neo-homodimers), because they have diverged too far in sequence to be equivalent. Thus, the clustering of genes in stable complexes cannot be attributed to the effect of essentiality or to tandem duplication.

We next tested whether the fraction of pairs in stable complexes and transient interactions that are on the same chromosome, within 10 kb and within 30 kb, is

### Table 1

Percentages of linked gene pairs of *S. cerevisiae* encoding subunits of the same macromolecular complexes

<table>
<thead>
<tr>
<th>Members of stable complexes*</th>
<th>Transient pairwise interactions*</th>
<th>Essential genes†</th>
<th>All genes: random expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairs</td>
<td>P-values</td>
<td>Pairs</td>
<td>P-values</td>
</tr>
<tr>
<td>Total no. of possible pairs</td>
<td>11,288</td>
<td>57,289</td>
<td>1,227,664</td>
</tr>
<tr>
<td>% pairs on SC</td>
<td>8%</td>
<td>7.8%</td>
<td>4%</td>
</tr>
<tr>
<td>% pairs within 10 kb of SC</td>
<td>3.4% (32)</td>
<td>2.6% (119)</td>
<td>2.5%</td>
</tr>
<tr>
<td>% pairs within 30 kb of SC</td>
<td>9% (83)</td>
<td>6.8% (306)</td>
<td>6.8%</td>
</tr>
<tr>
<td>% pairs within 10 kb of SC</td>
<td>0.02</td>
<td>0.08</td>
<td>0.2</td>
</tr>
<tr>
<td>% pairs within 30 kb of SC</td>
<td>&lt;10^−4</td>
<td>0.08</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Genes encoding proteins that are part of the same stable complex have the largest tendency to be clustered on the yeast chromosomes, followed by proteins in transient interactions and essential genes (departures of the corresponding percentages from random were all significant, as shown in the text). Absolute numbers of gene pairs are in parentheses. SC, same chromosome.

*In the MIPS Comprehensive Yeast Genome Database "catalog of complexes" (Mewès et al. 2002), there are 1185 proteins in 216 complexes ranging in size from 2 to 81 proteins. A total of 123 proteins occur in 2 or 3 different complexes. This data set represents the set of small, specific complexes that are at the lowest level within this hierarchical catalog. The pairs we consider are all possible pairs within each complex. Therefore, the set of all pairs in stable complexes may also include pairs that are not physically in contact, but are members of the same complex. This means that the fraction of clustered pairs could actually be an underestimate.

†There are 2598 proteins in 57,289 pairs of transient protein-protein interactions found in the MIPS Comprehensive Yeast Genome Database table of physical interactions together with the complexes identified by large-scale mass spectrometry experiments within the MIPS catalog of complexes. These exclude interactions identified by the error-prone yeast-two-hybrid method, as well as any interactions in the data set of stable complexes. Even so, there may be some stable interactions in this data set that are not in the group of proteins in stable complexes, so the chromosomal clustering of the set of proteins in stable complexes could be even stronger.

+ A total of 1106 genes in the data set were provided by the "Saccharomyces Genome Deletion Project" (http://www-sequence.stanford.edu/group/yeast_deletion_project/essential).
that are part of the same stable complex supports our idea that the proximity in chromosomal location is connected to the physical interaction of their gene products. Clustering may be advantageous to the organism, in the short term, as it may ensure transcriptional coregulation. In line with this, Kepes (2003) has shown that in yeast, genes that are controlled by the same transcription factor tend to be regularly spaced with periods that can go up to 50 kb.

As shown above, the driving force of clustering would be ensuring tighter coregulation among complex components. However, the formation of a cluster may have further evolutionary consequences. Within the context of the dosage balance notion (Veitia 2003) consider a complex \( A-B-C \), where subunit \( B \) forms a single bridge between two or more separable parts (\( A \) and \( C \)). A modest increase of \( B \) concentration may drastically inhibit complex formation as inactive subcomplexes \( AB \) and \( BC \) may form. \( B \) is said to be titrating (Veitia 2003). When subunits form multiple bonds within the body of the complex, increased concentrations during irreversible binding can also lead to inactive complexes (Bray and Lay 1997; Veitia 2003). In complexes such as \( A,B \), where \( B \) is a bridging factor, a decrease in concentration of \( A \) may lead to a disproportional reduction in complex formation, especially during irreversible reactions. Moreover, in larger complexes (i.e., \( n > 2 \)), the nonlinear effects of either increasing \( B \) or decreasing \( A \) sharpen as \( n \) increases (Figure 2).

We speculate that linkage of genes coding for dosage-sensitive subunits in complexes may help maintain the right stoichiometry of complexes on two scales: “here and now” and upon duplication of chromosomal segments. Elsewhere, we have carried out computer simulations to assess the impact of a concerted increase or diminution of a titrating monomer and one or multiple partners (as described by Veitia 2003). When a strongly titrating factor (such as \( B \)) is coinced with a separable component directly linked to it, its titrating power diminishes under irreversible conditions. For instance, suppose that to form complex \( A-B-C \), the initial relative amounts/rate of synthesis of \( A, B, \) and \( C \) are as in the formula and that all rate constants are identical. An excess of \( B \) lowers \( ABC \) formation but a parallel increase of \( B \) and another partner (\( A \) or \( C \)) leads to a normal amount of trimer. Interestingly, a parallel diminution of both monomers does not have a higher effect than deleting one of them alone (50% of normal). The titrating power of a bridging subunit (i.e., \( B \)) is higher in a complex \( ABCD \) (\( A, C, \) and \( D \) are separable components). For instance, \( 1.5 \times B \) may lead to a yield of <40% of \( ABCD \). In this case, coincrease of \( B \) with one component diminishes titration (yield \( ABCD > 65\% \)) and coincrease with two components removes the problem. In a complex \( A-B-A \), increasing \( A \) has no effect but an increase of \( B \) (1.5\( \times \)) diminishes the yield of \( ABA \) (i.e., \( \sim 77\% \) instead of 100%). Here, the advantage of clustering is
obvious and applies also to deletions. That is, deletion of one copy of gene A alone leads to <50% of complex yield while heterozygous deletions of both A and B lead to a 50% reduction of trimer yield. Linkage is also advantageous when the subunits can form alternative complexes (i.e., AA, AB, and BB) that must respect the relative molar concentrations. For equal starting concentrations and specific rates one will normally have 1AA:2AB:1BB. However, halving A alone leads to 1AA:4AB:4BB while trisomy of A yields 9AA:12AB:4BB (Figure 2). Another example is when two or more partners interact with the same subunit; e.g., A, B, and C yield AC and BC. Under the same assumptions as above, under normal conditions the concentrations respect the proportion, for instance, 1AC:1BC. However, halving A yields 1AC:2BC while increasing A (1.5×) leads to 3AC:2BC. In absolute terms, if the common partner C is in excess, the amount of BC will be normal for any change of A. However, if C is limiting, an increase (decrease) of A translates into a decrease (increase) of BC. Under this condition not only are the molar ratios altered but also the absolute amounts of each product are altered. Covariation of A and B solves the problem.

The advantage of a better coregulation of dosage-sensitive genes can be the main force that drives cluster-

ing. Note that the stoichiometric balance principle outlined here is general and applicable to other configurations of linked genes such as those described elsewhere in the context of dosage compensation (Birchler et al. 2001) and very likely to those involved in signaling and metabolic pathways. Moreover, conservation of syntenic associations from one organism to another is not expected in all cases as dosage sensitivity can vary according to the organisms. Besides, as shown above, a dosage-sensitive gene can be linked to different partners to obtain similar results. Interestingly, in the long run, gene clusters can be considered as functional modules. Indeed, genes in close proximity (clustered) are more likely to be duplicated together, avoiding imbalance. One may argue that coduplication frequency is probably comparable to the frequency of rearrangements breaking up the cluster. If such cluster-destroying translocations leave both genes functional, they would not be penalized by natural selection. However, if there is a selection pressure for increased quantity of some gene product (especially the dosage-sensitive one), then duplication would be favored and may include adjacent genes in the process. Those duplications will be more likely to survive if the genomic configurations are such that molecular complexes, encoded by the linked genes, maintain their balance. In essence, such a “hitchhiking effect” might be an additional selective pressure for this type of cluster. Kondrashov et al. (2002) have hypothesized that “gene duplications that persist on an evolutionary scale are beneficial from the time of their origin, due primarily to a protein dosage effect in response to variable environmental conditions.” Note that coduplication would be the only way to duplicate (on a small genomic scale) a dosage-sensitive subunit without an immediate decrease of fitness of the aneuploid. The coduplication hypothesis we present above implies linkage of duplicated gene pairs, but the syntenic association does not have to be preserved or detectable in every single case. It is known that if a functional allele raises its frequency at a duplicated locus, one of the copies can become a pseudogene and even disappear, giving the false appearance that the “companion” gene is a singleton.

All in all, we show that a significant fraction of genes encoding subunits of stable complexes are close to each other on yeast chromosomes. We interpret this as part of a strategy ensuring better coregulation and coduplication.

We are grateful to Jose Leal for helpful discussions and Cei Abreu for parsed data sets of protein complexes and pairwise interactions. We are indebted to Sir Walter Bodmer and Andreas Wagner for interesting discussions about the topic treated here. We thank James Birchler for insightful comments on the manuscript.

LITERATURE CITED


Communicating editor: J. A. Birchler