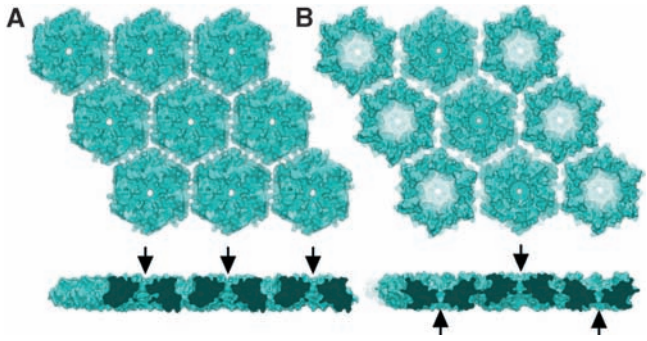

Fig. 4. Crystal packing of BMC domain proteins in molecular layers. (**A**) CcmK2 hexamers packed in uniform orientations (convex face shown). (**B**) CcmK4 hexamers (crystal form 2) packed in strips of alternating orientation. In the side view of each sheet, arrows mark the positions of the pores.

be present among a greater number of hexameric capsomeres (SOM text). These ideas are consistent with our studies of the BMC domain proteins. The first three crystal structures have revealed hexameric assemblies; sedimentation and native gel electrophoresis studies indicate that some of the subunits form pentamers in addition to hexamers (fig. S2, SOM text).

Likewise, similar but distinct protein subunits existing in quasi-equivalent environments or forms are required to construct virus capsids (*22*–*27*). Among the CcmK proteins, CcmK2 alone appears capable of forming closed shells under certain conditions, but these structures are much smaller than native carboxysomes and lack their polyhedral regularity (Fig. 1C). Apparently, as with many viruses, multiple distinct carboxysome subunit types appear to be required to achieve the correct architecture. The generally high conservation of amino acid residues among the CcmK paralogs at the interhexamer interfaces is consistent with the idea that hexameric (or possibly pentameric) units of different types could assemble together. The subtle differences in primary and tertiary structure could then govern the quasi-equivalent interactions necessary to create the native shell. The crystal structures of CcmK2 and CcmK4 give preliminary clues as to the origins of distinct assembly behavior, such as their disparate tendency to form sheets (Fig. 4). In CcmK4, clashes between C termini from adjacent hexamers could preclude the formation of flat sheets of the type seen with the CcmK2 hexamer. Interactions between the C-terminal tails of the BMC domain proteins could influence microcompartment domain assembly in the same way that the flexible termini of certain viral capsid proteins often participate as switches for distinct types of interactions in the mature viral capsid (*25*, *26*, *28*, *29*). The shape and assembly of the carboxysome could also be affected by other proteins that may be present in the shell (*30*) whose structures are not yet known.

As was the case for other large molecular machines such as viruses and ribosomes (*31*), fully elucidating the structure and function of bacterial microcompartments will require combining electron microscopy, biophysical experiments, and crystallographic studies on the multiple components in order to attain an understanding of the whole.

References and Notes

1. T. E. Jensen, C. C. Bowen, *Proc. Iowa Acad. Sci.* **68**, 89 (1961).
2. E. Gantt, S. F. Conti, *J. Bacteriol.* **97**, 1486 (1969).
3. J. M. Shively, G. L. Decker, J. W. Greenawalt, *J. Bacteriol.* **96**, 2138 (1970).
4. J. M. Shively, R. S. English, *Can. J. Bot.* **69**, 957 (1991).
5. J. M. Shively, F. Ball, D. H. Brown, R. E. Saunders, *Science* **182**, 584 (1973).
6. G. D. Price, J. R. Coleman, M. R. Badger, *Plant Physiol.* **100**, 784 (1992).
7. A. K.-C. So, G. S. Espie, *Plant Mol. Biol.* **37**, 205 (1998).
8. A. K.-C. So *et al.*, *J. Bacteriol.* **186**, 623 (2004).
9. G. A. Codd, W. J. N. Marsden, *Biol. Rev. Camb. Philos. Soc.* **59**, 389 (1984).
10. G. D. Price, M. R. Badger, *Can. J. Bot.* **69**, 963 (1990).
11. G. C. Cannon *et al.*, *Appl. Environ. Microbiol.* **67**, 5351 (2001).
12. G. C. Cannon, S. Heinhorst, C. E. Bradburne, J. M. Shively, *Funct. Plant Biol.* **29**, 175 (2002).
13. J. M. Shively *et al.*, *Can. J. Bot.* **76**, 906 (1998).
14. E. Kofoid, C. Rappleye, I. Stojiljkovic, J. Roth, *J. Bacteriol.* **181**, 5317 (1999).
15. G. D. Havemann, T. A. Bobik, *J. Bacteriol.* **185**, 5086 (2003).
16. M. R. Badger, D. Hanson, G. D. Price, *Funct. Plant Biol.* **29**, 161 (2002).
17. L. Holm, C. Sander, *J. Mol. Biol.* **233**, 123 (1993).
18. J.-F. Gibrat, T. Madej, S. H. Bryant, *Curr. Opin. Struct. Biol.* **6**, 377 (1996).
19. P. Yuan *et al.*, *Nat. Struct. Biol.* **10**, 264 (2003).
20. J. M. Diprose *et al.*, *EMBO J.* **20**, 7229 (2001).
21. H. Naitow, J. Tang, M. Canady, R. B. Wickner, J. E. Johnson, *Nat. Struct. Biol.* **9**, 725 (2002).
22. D. L. D. Caspar, A. Klug, *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1 (1962).
23. J. M. Hogle, M. Chow, D. J. Filman, *Sci. Am.* **256**, 42 (March 1987).
24. M. G. Rossmann, J. E. Johnson, *Annu. Rev. Biochem.* **58**, 533 (1989).
25. S. C. Harrison, *Curr. Opin. Struct. Biol.* **5**, 157 (1995).
26. J. E. Johnson, J. A. Speir, *J. Mol. Biol.* **269**, 665 (1997).
27. S. C. Harrison, *Curr. Opin. Struct. Biol.* **11**, 195 (2001).
28. T. Stehle, S. J. Gamblin, Y. Yan, S. C. Harrison, *Structure* **4**, 165 (1996).
29. C. San Martin *et al.*, *Nat. Struct. Biol.* **9**, 756 (2002).
30. A. K. So *et al.*, *J. Bacteriol.* **186**, 623 (2004).
31. N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **289**, 905 (2000).
32. W. L. DeLano, www.pymol.org (2002).

33. We thank G. Cannon and M. Yeager for helpful discussions, D. Krogmann for the gift of *Synechocystis* sp. PCC 6803 genomic DNA, and J. Laidman and H. Adesiteyo for technical assistance. We also thank R. W. Roberson, W. F. J. Vermass, and A. van de Meene for providing transmission electron micrographs of *Syn* 6803 cells. C.A.K and T.O.Y acknowledge the U.S. Department of Agriculture, the National Institutes of Health, and the Office of Science Biological and Environmental Research Program, the U.S. Department of Energy. The coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 2A10 (CcmK4, Form 1); 2A18 (CcmK4, Form 2); and 2A1B (CcmK2).

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5736/936/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References and Notes

11 April 2005; accepted 22 June 2005
10.1126/science.1113397

Rewiring of the Yeast Transcriptional Network Through the Evolution of Motif Usage

Jan Ihmels,[1] Sven Bergmann,[1] Maryam Gerami-Nejad,[2] Itai Yanai,[1] Mark McClellan,[2] Judith Berman,[2] Naama Barkai[1]*

Recent experiments revealed large-scale differences in the transcription programs of related species, yet little is known about the genetic basis underlying the evolution of gene expression and its contribution to phenotypic diversity. Here we describe a large-scale modulation of the yeast transcription program that is connected to the emergence of the capacity for rapid anaerobic growth. Genes coding for mitochondrial and cytoplasmic ribosomal proteins display a strongly correlated expression pattern in *Candida albicans*, but this correlation is lost in the fermentative yeast *Saccharomyces cerevisiae*. We provide evidence that this change in gene expression is connected to the loss of a specific cis-regulatory element from dozens of genes following the apparent whole-genome duplication event. Our results shed new light on the genetic mechanisms underlying the large-scale evolution of transcriptional networks.

Evolution of gene expression plays a prominent role in generating phenotypic diversity (*1*–*3*), but little is known about the genetic basis underlying broad modulations of the genome-wide transcription program. Here we describe a rewiring of the yeast transcrip-

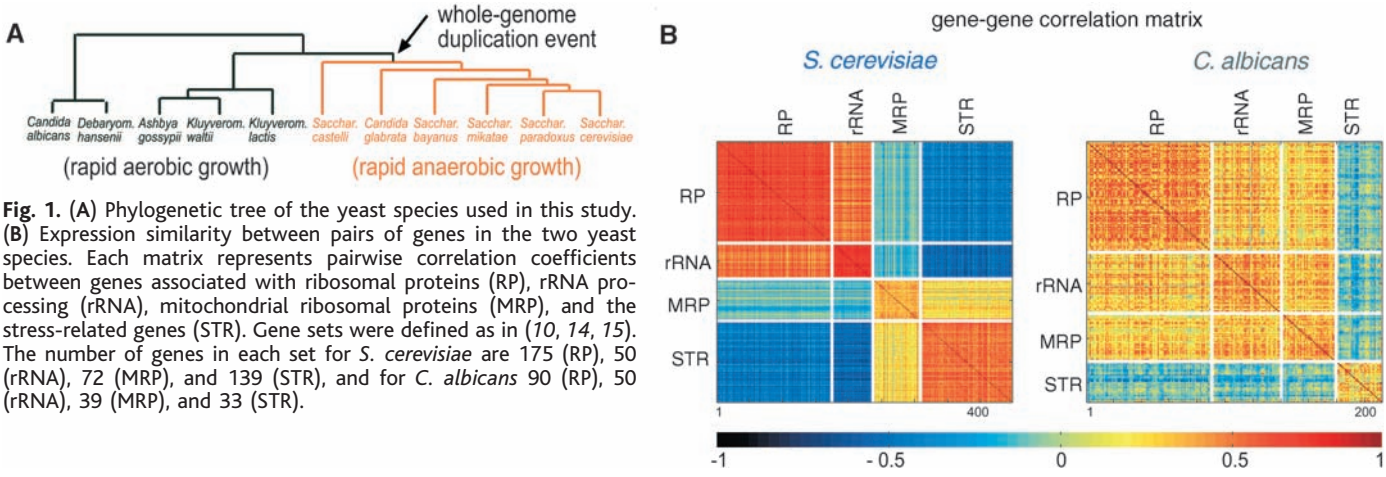

Fig. 1. (**A**) Phylogenetic tree of the yeast species used in this study. (**B**) Expression similarity between pairs of genes in the two yeast species. Each matrix represents pairwise correlation coefficients between genes associated with ribosomal proteins (RP), rRNA processing (rRNA), mitochondrial ribosomal proteins (MRP), and the stress-related genes (STR). Gene sets were defined as in (*10*, *14*, *15*). The number of genes in each set for *S. cerevisiae* are 175 (RP), 50 (rRNA), 72 (MRP), and 139 (STR), and for *C. albicans* 90 (RP), 50 (rRNA), 39 (MRP), and 33 (STR).

tional network that is linked to the massive loss of a conserved cis-regulatory element.

With the recent sequencing of many yeast genomes, the hemiascomycete lineage emerged as a central paradigm for studying the genetic basis of phenotypic diversity (*4*). Despite a strong conservation of gene content, yeast species exhibit major phenotypic differences. Most striking, perhaps, is their differential requirement for oxygen. Whereas growth of most yeast species requires oxygen, *Saccharomyces cerevisiae* grows rapidly in its absence and prefers to ferment glucose anaerobically even when oxygen is present. The emergence of this capacity is linked to the apparent whole-genome-duplication event (*5*–*7*), which allowed the specialization of enzymes to aerobic versus hypoxic conditions (*4*, *8*) (Fig. 1A).

The preference for fermentation is reflected in the organization of the *S. cerevisiae* transcriptional network. For example, genes involved in oxidative phosphorylation are repressed in the presence of glucose (*9*). To examine the transcription program of *S. cerevisiae* in more detail, we compiled a database of over 1000 published genome-wide expression profiles (*10*). In accordance with its anaerobic growth phenotype, we found that the expression of *S. cerevisiae* genes encoding mitochondrial functions, most notably the mitochondrial ribosomal proteins (MRP), is not correlated (and perhaps is even weakly anticorrelated) with that of genes coding for cytoplasmic ribosomal proteins (RP) and with ribosomal RNA (rRNA) processing genes (*10*) (Fig. 1B). Instead, expression of the MRP genes exhibits a distinct correlation with that of genes induced in response to environmental stress conditions. At least in part, this correlation reflects the induction of the stress-related genes during the relatively slow growth in nonfermentable carbon sources, such as glycerol or ethanol, which requires mitochondrial function.

To examine the manifestation of this relationship between the expression of MRP genes and genes required for rapid growth in aerobic yeast species, we assembled a data set of 198 genome-wide expression profiles of *Candida albicans*, a human pathogen that primarily grows aerobically. In accordance with the requirement for respiration to support rapid growth, we found that in *C. albicans* the expression of MRP genes is strongly correlated with the expression of genes involved in ribsome assembly, including RP genes and rRNA processing genes (Fig. 1B). Thus, the organization of the *S. cerevisiae*

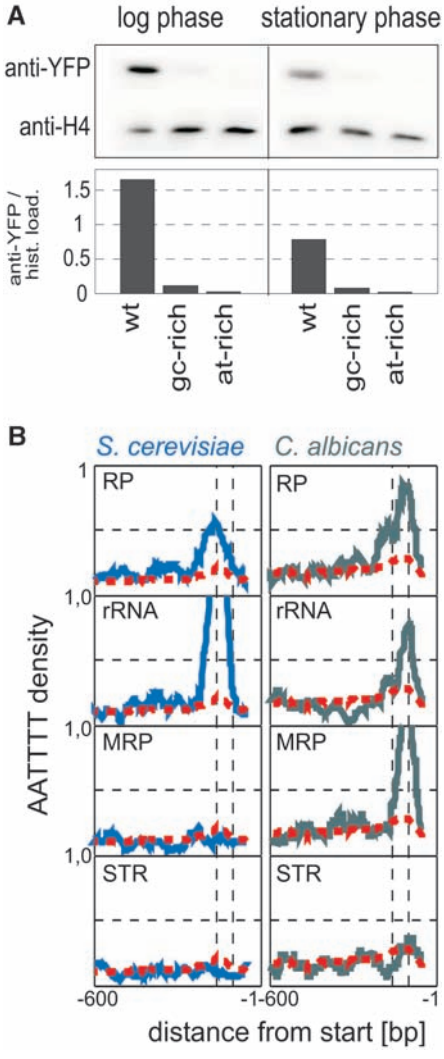

Fig. 2. (**A**) The AATTTT sequence is required to induce high-level transcription of a reporter protein fused to the *C. albicans* MRP7 promoter in exponentially growing cells. (**B**) Frequency of occurrence of RGE in a 50-bp window of the gene groups indicated. Background genomic frequency is indicated in red (*15*). Multiple occurrences of the RGE were included. The same quantitative result is obtained when only a single occurrence per gene is considered (fig. S2).

[1]Department of Molecular Genetics and Department of Physics of Complex systems, Weizmann Institute of Science, Rehovot, Israel. [2]Department of Genetics, Cell Biology and Development and Department of Microbiology, University of Minnesota, MN 55455, USA.

*To whom correspondence should be addressed. E-mail: naama.barkai@weizmann.ac.il

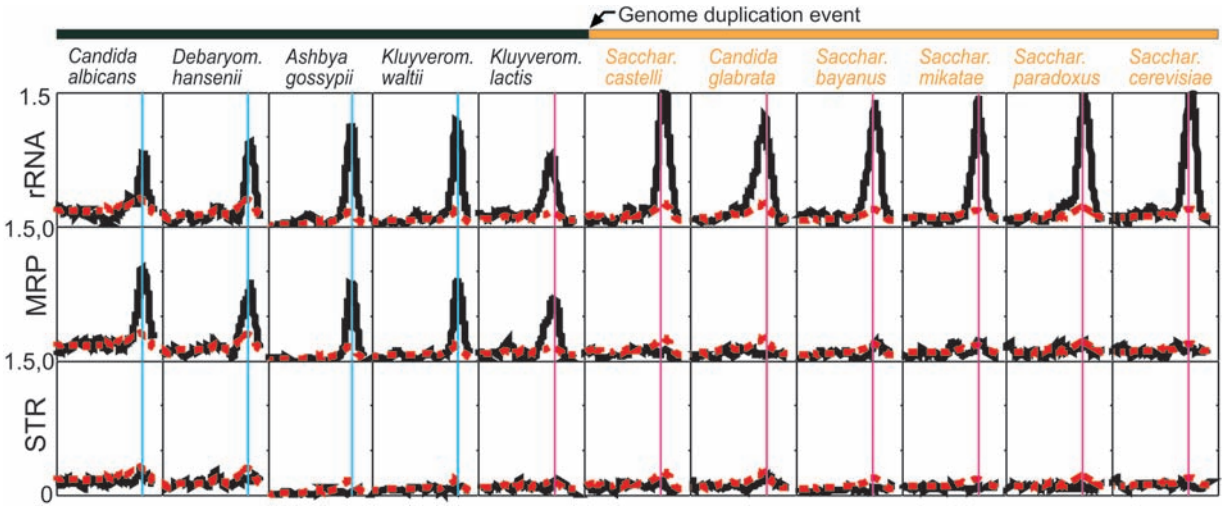

Fig. 3. Frequency of occurrence of RGE in intermediate yeast species (*15*). The gene sets for *S. cerevisiae* were defined as in Fig. 1B. Gene sets for the remaining species consist of orthologs of the *S. cervevisiae* genes.

and *C. albicans* transcriptional networks reflects their respective physiologies.

Gene expression can evolve through mutations in trans-acting regulatory proteins or in cis-acting DNA elements (compare fig. S4). For example, the regulation of a common transcription factor could diverge between the species. Alternatively, modifications of the individual cis-regulatory elements of each downstream gene could change their regulation, potentially coopting or losing regulation by a transcription factor. To distinguish between these scenarios, we searched the regions 600 base pairs (bp) upstream of open reading frames of both organisms for overrepresented sequence elements. Specific regulatory sequences that appear to be conserved between *S. cerevisiae* and *C. albicans* were identified (*11*, *12*). Most notably, the PAC (GATGAG) sequence (*13*) was associated with rRNA processing genes in both species. In agreement with previous reports (*10*), no overrepresented sequence was associated with the MRP genes in *S. cerevisiae*.

In contrast, in *C. albicans*, the MRP genes were clearly associated with an overrepresented upstream sequence motif (AATTTT). This sequence was previously implicated in the regulation of rRNA processing genes in *S. cerevisiae* (*13*), although its functional role was not demonstrated experimentally. To examine whether this motif contributes to the regulation of MRP genes in *C. albicans*, we fused the promoter of the MRP7 gene to a yellow fluorescent protein (YFP) reporter. As expected, strong induction of the YFP reporter was observed during exponential growth, whereas expression in stationary phase was much weaker (Fig. 2A). Mutating the AATTTT sequence in the MRP7 promoter to either a GC-rich or to an AT-rich sequence drastically reduced expression of the YPF (Fig. 2A), indicating that it is the AATTTT sequence itself, and not its AT-rich nature, that is important for its function. Similar results were obtained with two additional mitochondrial ribosomal protein gene promoters (fig. S5). We conclude that the AATTTT sequence functions as a cis-regulatory element of MRP genes in *C. albicans*.

Examining the appearance of this motif in more detail, we found that its position relative to that of the ORF start codon is highly confined in both *S. cerevisiae* and *C. albicans*, although the precise position is somewhat different (160 versus 110 bp in *S. cerevisiae* and *C. albicans*, respectively). In both organisms, this sequence is significantly overrepresented in genes involved in rRNA processing and ribosomal proteins, but not in genes associated with the environmental stress response. In addition, only in *C. albicans* is it overrepresented also in MRP genes (Fig. 2B). Because this conserved element is associated with genes required for rapid growth in both organisms, we denote it as a rapid growth element (RGE).

This RGE could have been coopted to *C. albicans* MRP genes, or alternatively, could have been lost in *S. cerevisiae*. To distinguish between these alternatives, we examined the appearance of the RGE in nine sequenced yeast species that are considered to be intermediate in the evolution between *S. cerevisiae* and *C. albicans* (Fig. 3). In all species examined, the RGE is significantly overrepresented in genes involved in rRNA processing. Notably, overrepresentation of the RGE in MRP promoters was found in all genomes that diverged from the *S. cerevisiae* lineage before the whole-genome duplication event (Fig. 3). This indicates a widespread loss of the RGE in MRP genes following the genome duplication.

Taken together, it appears that the emergence of anaerobic growth capacity in yeast is associated with a global rewiring of the yeast transcriptional network. This rewiring involved changes in the promoter regions of dozens of genes, manifested by the loss of a specific regulatory motif from MRP gene promoters. It would be interesting to examine which transcription factor binds to the RGE site and whether its coding sequence was also modified in association with the emergence of anaerobic growth. Intriguingly, the most parsimonious scenario for the described promoter adaptation appears to be intimately linked with the whole-genome duplication event. Our results suggest that gene duplication can facilitate the evolution of new function not only by specialization of coding sequences but also by facilitating the evolution of gene expression.

References and Notes

1. M. C. King, A. C. Wilson, *Science* **188**, 107 (1975).
2. S. B. Carroll, *Cell* **101**, 577 (2000).
3. E. H. Davidson, *Genomic Regulatory Systems: Development and Evolution* (Academic Press, San Diego, CA, 2001).
4. J. Piskur, R. B. Langkjaer, *Mol. Microbiol.* **53**, 381 (2004).
5. K. H. Wolfe, D. C. Shields, *Nature* **387**, 708 (1997).
6. M. Kellis, B. W. Birren, E. S. Lander, *Nature* **428**, 617 (2004).
7. F. S. Dietrich *et al.*, *Science* **304**, 304 (2004).
8. K. Wolfe, *Curr. Biol.* **14**, R392 (2004).
9. H. J. Schuller, *Curr. Genet.* **43**, 139 (2003).
10. J. Ihmels *et al.*, *Nat. Genet.* **31**, 370 (2002).
11. A. P. Gasch *et al.*, *PLoS Biol.* **2**, 398 (2004).
12. A. Tanay, A. Regev, R. Shamir, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7203 (2005).
13. S. Tavazoie, G. M. Church, *Nat. Biotechnol.* **16**, 566 (1998).
14. S. Bergmann, J. Ihmels, N. Barkai, *PLoS Biol.* **2**, E9 (2004).
15. **Methods** are available as supporting material on *Science* Online. See also www.weizmann.ac.il/home/barkai/Rewiring.

16. We thank the following for providing transcription profiling data sets: R. Bennett, A. Tsong, A. Johnson, M. Lorenz, C. D'enfert, G. Fink, M. Whiteway, A. Nantel, P. D. Rogers, and especially D. Davis, D. Harcus, B. Hube, and D. Sanglard, for providing transcription profile data sets before publication. We also thank A. Nantel, C. D'enfert, and A. Tsong for providing gene name information that allowed assignment of genes on different arrays to their orf19 gene identities. This work was supported by NIH grants A150562 (N.B.) and R01 DE/AI 14666 (J.B.) and a grant from the Kahn Fund for Systems Biology at the Weizmann Institute of Science (N.B.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5736/938/DC1
Methods
Figs. S1 to S5
Tables S1 to S4

20 April 2005; accepted 23 June 2005
10.1126/science.1113833