Gains and Losses of Transcription Factor Binding Sites in *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*

Bernhard Schaefke[1,2,3], Tzi-Yuan Wang[2], Chuen-Yi Wang[2], and Wen-Hsiung Li[2,4,5,*]

[1]Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan

[2]National Yang-Ming University, Taipei, Taiwan

[3]Bioinformatics Program, Institute of Information Science, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan

[4]China Medical University Hospital, Taichung, Taiwan

[5]Department of Ecology and Evolution, University of Chicago

*Corresponding author: E-mail: whli@gate.sinica.edu.tw.

Accepted: July 9, 2015

Abstract

Gene expression evolution occurs through changes in *cis*- or *trans*-regulatory elements or both. Interactions between transcription factors (TFs) and their binding sites (TFBSs) constitute one of the most important points where these two regulatory components intersect. In this study, we investigated the evolution of TFBSs in the promoter regions of different *Saccharomyces* strains and species. We divided the promoter of a gene into the proximal region and the distal region, which are defined, respectively, as the 200-bp region upstream of the transcription starting site and as the 200-bp region upstream of the proximal region. We found that the predicted TFBSs in the proximal promoter regions tend to be evolutionarily more conserved than those in the distal promoter regions. Additionally, *Saccharomyces cerevisiae* strains used in the fermentation of alcoholic drinks have experienced more TFBS losses than gains compared with strains from other environments (wild strains, laboratory strains, and clinical strains). We also showed that differences in TFBSs correlate with the *cis* component of gene expression evolution between species (comparing *S. cerevisiae* and its sister species *Saccharomyces paradoxus*) and within species (comparing two closely related *S. cerevisiae* strains).

Key words: regulatory evolution, *cis* evolution, *trans* evolution, selective constraint, natural selection, yeast.

Introduction

The budding yeast *S. cerevisiae* has been used for fermentation by humans for millennia, including the production of bread and a large variety of alcoholic beverages, and is one of the most important eukaryotic model organisms. Various yeast strains are found in different environments (e.g., tree barks, wine or beer fermentation, laboratories, etc.) and display different metabolic characteristics, and growth and gene expression phenotypes (Spor et al. 2009; Skelly et al. 2013). Several studies have shown faster evolution for the laboratory strain S288C (Gu et al. 2005; Li et al. 2009) in comparison with other strains. A recent study supported the view that all "domesticated" strains (including laboratory, industrial, fermentation, clinical, and vineyard strains) have accumulated more deleterious recessive alleles than wild *S. cerevisiae* strains (Plech et al. 2014).

Phenotypic differences between closely related species or individuals of the same species are often caused by gene regulatory changes. These can be classified into changes in *cis*-regulatory elements or changes in *trans*-regulatory proteins, that is, transcription factors (TFs). Recent studies have addressed the relative contributions of *cis*- and *trans*-regulatory changes to gene expression evolution in yeast (Sung et al. 2009; Tirosh, Reikhav, et al. 2009; Emerson et al. 2010; Schaefke et al. 2013). The results suggested that the relative importance of *cis*-regulatory differences is higher between species than within species and that *trans*-regulatory changes are subjected to stronger selective constraints than *cis*-regulatory changes. The relationship between gene regulatory changes in *cis* and *trans* is complex. In general, *trans* changes can be expected to affect more downstream genes and have a larger impact on gene expression profiles, whereas changes in *cis*-regulatory elements allow for the fine-tuning of gene expression on a smaller scale. A study of mutation accumulation and natural isolate lines in *Caenorhabditis elegans* found that most *trans*-acting

mutations that resulted in expression changes of multiple genes were quickly removed by selection in natural populations (Denver et al. 2005). On the other hand, the gain of insulator proteins has played a significant role in arthropod evolution (Heger et al. 2013), and the duplication of TFs in *C. elegans* has contributed to extensive network rewiring (Reece-Hoyes et al. 2013). In addition, a recent study in *Paramecium* shows that the maintenance or loss of protein-coding genes after whole-genome duplication is largely determined by dosage constraints (Gout and Lynch 2015), conforming to a model under which the expression level of individual duplicated genes can evolve neutrally as long as they maintain a roughly constant summed expression.

The evolutionary dynamics of *cis*-regulatory elements have been studied in various organisms: Haygood et al. (2007) found that the promoter regions of many genes related to nutrition or to neural development and function have experienced positive selection in the human lineage. Liang et al. (2008) examined the core promoters of human and macaque, and found evidence of positive selection on the promoters of genes involved in metabolic and biosynthetic processes. A more recent study made use of large-scale Chip-seq data and found strong evidence of adaptive substitutions as well as weakly deleterious polymorphisms in human TF binding sites (TFBSs) (Arbiza et al. 2013). A study on TFBS gain and loss in two *Drosophila* species showed patterns of polymorphism and divergence that were inconsistent with neutral evolution, but supported a combination of positive selection and selective constraint (He et al. 2011). A comparison of regulatory networks in human, worm, and fly showed that structural properties of regulatory circuitry were largely conserved and that orthologous TF families recognize similar binding motifs (Boyle et al. 2014).

Other studies have investigated the specific evolutionary patterns of TFs (Coulombe-Huntington and Xia 2012) and of promoter sequences and TFBSs in the yeast genome (Doniger and Fay 2007; Tirosh et al. 2008; Tirosh, Barkai, et al. 2009; Chen et al. 2010; Tsai et al. 2012; Lin et al. 2013). A recent study found that regulatory motif variation is associated with gene expression differences between *S. cerevisiae* strains (Connelly et al. 2013). Moreover, previous studies have found that the majority of experimentally confirmed TFBSs in *S. cerevisiae* lies in between 200 and 80 bp upstream of the transcription starting site (TSS), with a peak at about 115 bp upstream of the TSS (Lin et al. 2010). Lin et al. also found that the regions from the TSS to −200 bp show fewer deletion polymorphisms among *S. cerevisiae* strains than expected, suggesting that the distance between TSS and TFBS has been under selective constraint. In contrast, they detected no relationship between SNP frequency and the distance from the TSS, but this could be because not all SNPs affect TFBSs. Therefore, our study compared the "proximal promoter" region, which is defined as the 200-bp region upstream of the TSS and is expected to contain more TFBSs, with the "distal promoter" region, which is defined as the intergenic region between 201 and 400 bp upstream of the TSS and is expected to be under weaker selective constraint. Finally, we investigated the patterns of TFBS evolution in different *S. cerevisiae* strains.

In summary, this study addresses the following issues: 1) Are TFBSs in the proximal promoter region more conserved than those in the distal promoter region? 2) How are differences in TFBSs between strains related to the *cis* and *trans* components of gene expression differences between strains? 3) Are the differences in the rates of TFBS gain and loss in diverse yeast strains and species related to differences in their ecological niches (laboratory, fermentation, wild)? We addressed these questions by investigating the gain and loss of TFBSs within and between *S. cerevisiae* and *S. paradoxus*, using *Saccharomyces mikatae* and *Saccharomyces bayanus* as outgroups.

Materials and Methods

Constructing Sets of Orthologous Sequences

We chose nine *S. cerevisiae* strains (BY4741, RM11-1a, YJM789, SK1, DBVPG6044, UWOPS03-461.4, Y12, YPS606, and UWOPS87-2421) and five *S. paradoxus* strains (CBS432, CBS5829, Y8.5, YPS138, and N-44) (table 1), representing different genetic clusters and geographic origins (Liti et al. 2009). Genome sequence data and annotations were downloaded from the *Saccharomyces* Genome Database (SGD, http://www.yeastgenome.org/download-data/sequence, last accessed: September 19, 2014) for BY4741, RM11-1a and YJM789, from the Yeast Resource Center (YRC, http://www.yeastrc.org/g2p-data/raw-data/genomes/, last accessed: September 19, 2014) for YPS606 and, from the *Saccharomyces* Genome Resequencing Project (SGRP, http://www.moseslab.csb.utoronto.ca/sgrp/download.html, last accessed: October 29, 2014) for the remaining strains (Bergstrom et al. 2014). The genome sequences of the *S. mikatae* and *S. bayanus* genomes (Kellis et al. 2003) were obtained from the BROAD institute website (http://www.broadinstitute.org/annotation/fungi/comp_yeasts/, last accessed: September 19, 2014).

Open-reading frame sequences (ORFs) and coding sequences (CDSs) were extracted from the genome sequences according to the genome coordinates provided (SGD, YRC, and SGRP sequences) or obtained directly from the database (BROAD institute).

We considered two different promoter regions: 1) The proximal promoter region defined as the 200-bp region immediately upstream of the TSS, and 2) the distal promoter region defined as the region from 201 to 400 bp upstream of the TSS. Experimentally defined TSSs for the *S. cerevisiae* reference strain S288C were obtained from Nagalakshmi et al. (2008). In total, the proximal promoters of 1,400 genes with known TSSs and their orthologs in *S. mikatae*, *S. bayanus,* and

all selected *S. cerevisiae* and *S. paradoxus* strains were used in our study. We excluded those genes with a distal promoter region shorter than 100 bp or overlapping for 100 bp or more with the proximal promoter region of a neighboring (head-to-head) gene. This resulted in a set of 590 genes for the comparison between the proximal promoter and the distal promoter. All basic sequence analyses were performed using customized Python scripts (Sanner 1999) and Biopython (Cock et al. 2009; Talevich et al. 2012). Additionally, we compared three smaller regions, each with a length of 80 bp: 1) The 80-bp region immediately upstream of the TSS; 2) the 80 bp from 91 to 170 bp upstream of the TSS, which is expected to include the peak of the TFBS density; and 3) the region 201–280 bp upstream of the TSS.

We used the MUSCLE (MUltiple Sequence Comparison by Log-Expectation) tool (Edgar 2004) to align promoter sequences, ORFs, and CDSs. A maximum-likelihood (ML) tree (fig. 1) for the concatenated aligned ORFs and promoter sequences was constructed using PhyML (Guindon et al. 2009, 2010; Criscuolo 2011). For each internal node of the phylogenetic tree, a likelihood-based reconstruction of the ancestral promoter sequences (Yang et al. 1995; Koshi and Goldstein 1996) was obtained using the BASEML program in the PAML package (Yang 1997, 2007). The tree topology was fixed, the branch lengths estimated by PhyML were used as the initial values, and alignment gaps and ambiguous letters were removed for the ancestral sequence reconstruction, using the general time-reversible nucleotide substitution model (BASEML parameters: cleandata = 1, fix_blength = 1, model = 7). The removal of alignment gaps and ambiguous nucleotides limits our analysis to nucleotide substitutions between the different strains (and excludes insertions and deletions). After removal of regions overlapping with neighboring proximal promoters and of nonalignable sequences, the average length of proximal promoters was 163.1 bp (median: 168 bp) and that of distal promoters was 142 bp (median: 146 bp).

Identification of Putative TFBSs

We obtained position weight matrices for TFBSs from the JASPAR database (jaspar.genereg.net/html/DOWNLOAD/JASPAR_CORE/pfm/nonredundant/pfm_fungi.txt). The sources used by the JASPAR database were an in vitro binding screen (Badis et al. 2008), a protein-binding microarray experiment (Zhu et al. 2009), the compiled SCPD binding profile database (Zhu and Zhang 1999), the SwissRegulon computational reanalysis of multiple data collections (Pachkov et al. 2007), and a motif discovery-based collection from a widely used ChIP-chip data collection (MacIsaac et al. 2006), in the order of the curators' preference (Portales-Casamar et al. 2010).

We identified putative TFBSs using the FIMO tool of the MEME suite (Grant et al. 2011). We used the default parameters (threshold *P* value: 0.0001) and two different ways to determine the background nucleotide distribution: 1) We fixed

Table 1
The *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* Strains Used in This Study[a]

Strain	Geographic Origin	Ecological Niche	Genetic Cluster
Saccharomyces ***cerevisiae***			
Clade A			
YPS606	America	Wild	North American
YJM789	Europe	Clinical	Mosaic
BY4741	America	Laboratory	S288C-derived (mosaic)
RM11-1a	America	Fermentation (vineyard)	European/wine
Clade B			
UWOPS87-2421	America	Wild	Mosaic
Y12[b] (K12)	Asia	Fermentation (sake)	Sake
Clade C			
UWOPS03-461.4	Asia	Wild	Malaysian
SK1	America	Laboratory	(close to) West African
DBVPG6044	Africa	Fermentation (Bili wine)	West African
Saccharomyces paradoxus			
YPS138	America	Wild	America
N-44	Asia	Wild	Far East
Y8.5	Europe	Wild	Europe
CBS5829	Europe	Laboratory	Europe
CBS432	Europe	Laboratory	Europe

[a]The strains were chosen to represent different ecological niches, geographic origins, and genetic clusters of *S. cerevisiae* and *S. paradoxus* (Johnson et al. 2004; Liti et al. 2009).

[b]Y12 was previously reported as an African fermentation strain, isolated from palm wine in Ivory Coast (Liti et al. 2009), but was later identified as the Japanese sake strain K12 in a newer publication by the same group (Bergstrom et al. 2014).

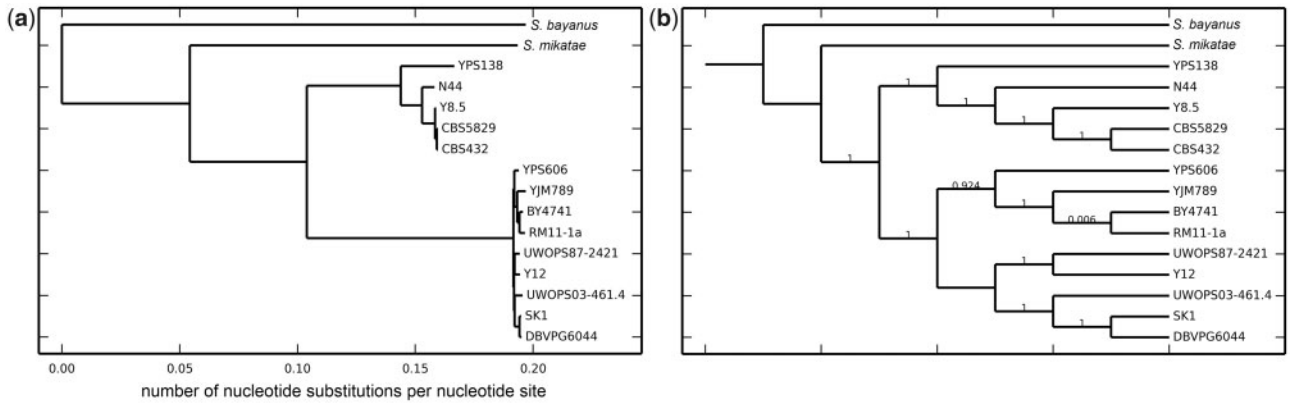

Fig. 1.—ML tree used for the reconstruction of ancestral sequences. (*a*) The branch lengths were drawn according to scale. The *x* axis indicates the number of nucleotide substitutions per nucleotide site. (*b*) Tree topology only. Branch labels indicate bootstrap values.

the background distribution equivalent to a GC content of 36%, and 2) we used a strain- and region-specific nucleotide distribution, according to the average base composition of either the proximal or the distal promoter regions of each strain (supplementary table S1, Supplementary Material online).

To avoid overestimating the number of TFBSs and of TFBS gain and loss events, we combined the results obtained with the two slightly different background nucleotide distributions in the following way: First, we built a core set of unambiguous TFBSs each of which was detected under both, the fixed background nucleotide distribution and the strain-specific background distribution, for a given promoter sequence. Second, if a TFBS was detected only under one of the two background nucleotide distributions, then it was counted as a TFBS only if the same sequence was already categorized as an unambiguous TFBS for the same TF in another strain (supplementary fig. S1, Supplementary Material online).

In cases where the software detected two overlapping TFBSs for the same TF, these were counted as one single TFBS if the overlap was larger than 50% of their average length. In contrast, if the binding sites of two different TFs overlapped, both were counted.

Identification of TFBS Gains and Losses

We determined TFBS gains, losses, and single-branch turnover events (i.e., loss of the ancestral TFBS and gain of a TFBS for the same TF at another position in the same promoter sequence) for each branch of the phylogenetic tree in a comparison of each ancestral sequence (internal node) and its daughter nodes.

In the case that a nucleotide difference between two strains affects two binding sites, it was counted for both TFBSs, because overlapping binding sites for different TFs were counted separately for each TF.

We classified each difference between two *S. cerevisiae* strains as follows: A putative TFBS loss had occurred in the strain under study if the TFBS was absent in this strain, but was present in the other strain as well as the reconstructed most recent common ancestor (MRCA) of the two strains (see supplementary fig. S2, Supplementary Material online). On the other hand, it was a putative gain if the TFBS was present in this strain, but in neither the other strain nor the reconstructed ancestral sequence. If a TFBS was present (absent) in the ancestral sequence and absent (present) in both strains, this was counted as a parallel loss (gain).

The total number of gain and loss events for a lineage was calculated by simply adding up the numbers of gain and loss events, separately, along the branch leading from the MRCA to the strain under study (as indicated in fig. 2). For example, the total number of gain events for DBVPG6044 in its comparison with UWOPS03-461.4 is the sum of the value for the branch leading from the MRCA of these two strains (supplementary fig. S3, Supplementary Material online, node with green circle) to the common ancestor of SK1 and DBVPG6044 (denoted as MRCA–SK1–DBVPG, supplementary fig. S3, Supplementary Material online, node with purple circle) and the number of gains on the branch leading from MRCA–SK1–DBVPG to DBVPG6044: 116 + 58 = 174.

The net number of TFBS gains and losses along a lineage from an ancestor X to a daughter sequence Z refers to the differences between sequences X and Z. This number was subtracted from the total number of gain and loss events on each lineage to estimate the number of multiple-branch TFBS turnovers. Multiple-branch TFBS turnover events are the changes between an ancestor X, an intermediate ancestor Y and a daughter sequence Z, where either a TFBS present in X was lost on the lineage leading to Y and regained on the branch(es) leading to Z (examples in supplementary fig. S2c, Supplementary Material online), or a TFBS was gained from X to Y, but subsequently lost on the lineage leading to Z.

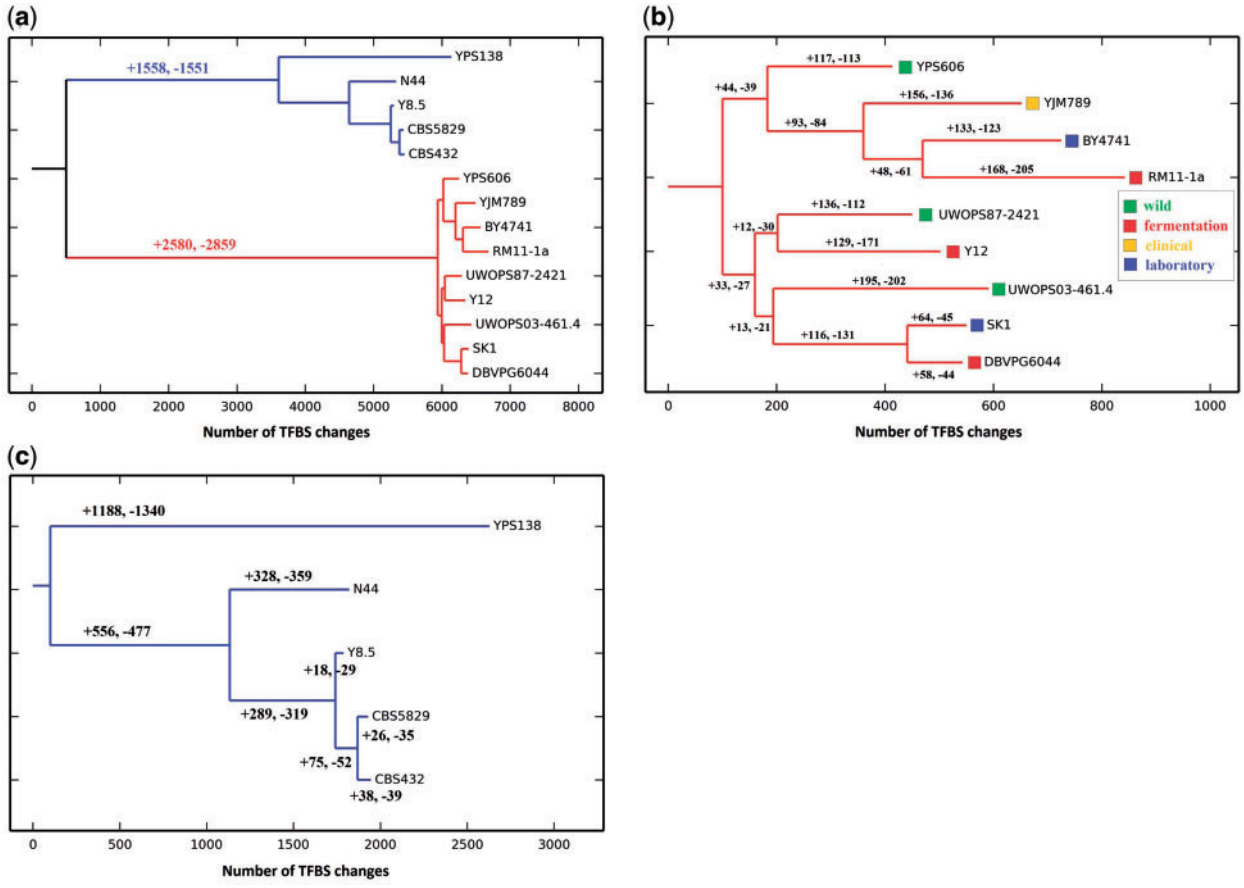

Fig. 2.—TFBS gains and losses along tree branches. (*a*) Gains and losses in the branches leading from the MRCA of *S. paradoxus* and *S. cerevisiae* to the common ancestor of the *S. cerevisiae* strains and to the MRCA of *S. paradoxus* strains, respectively. (*b*) The *S. cerevisiae* clade in red is enlarged: Gains and losses for the different *S. cerevisiae* strains are presented on the tree branches. (*c*) Gains and losses for the different *S. paradoxus* strains. Note that "+" indicates TFBS gains and "−" TFBS losses. Branch lengths represent the total number of changes.

We compared our approach using reconstructed ancestral sequences with one using an extant outgroup instead to infer the TFBS gains and losses between two taxa. In the latter method, we inferred TFBS gains and losses between two *S. cerevisiae* strains, using the *S. paradoxus* sequence as the outgroup (reference), and gains and losses between *S. cerevisiae* and *S. paradoxus*, using the *S. mikatae* sequence. If a TFBS was present in two of the strains but not the third, it was counted as a loss in the third lineage. If it was absent in all but one lineage, it was counted as a gain in this lineage. As parallel losses in two lineages are counted as gains in the third lineage with this outgroup-based approach (supplementary fig. S2*a* and *b*, Supplementary Material online), it leads to an overestimation of putative gains. Therefore, the gains inferred largely outnumbered the losses inferred when *S. paradoxus* was used as an outgroup for the pairwise comparisons of *S. cerevisiae* strains (supplementary table S2*b*, Supplementary Material online). As this method tends to overestimate gains and underestimate losses, we did not use it in our study.

Relationship between TFBS Changes and Gene Expression Differences in *cis* and *trans*

Gene expression differences between RM11-1a and BY4741 and estimates for their *cis* and *trans* components were obtained from Schaefke et al. (2013).

All statistical analyses were conducted using R (v. 3.0.2) (R-Core-Team 2013).

Mutagenesis of TFBS

To determine the effects of TFBS gain/loss on the expression level of the gene, we chose SNP (single nucleotide polymorphism) sites that lead to TFBS gain/loss events in a promoter for site-directed mutagenesis. The TFBS of interest in the BY4741 gene was first replaced by an *URA3* cassette and transformed by the LiOAc/SS Carrier DNA/PEG method. The inserted *URA3* was then replaced in a second transformation by the appropriate fragment of BY4741's polymerase chain reaction (PCR)-based SNP-mutated TFBS region in the *URA3*-inserted strain and screened by 5-Fluoroorotic Acid (5-FOA)

counter selection. Only the strains (called Swapped strains) that carried the desired sequence survived and formed colonies on the media with 5-FOA (2 μg/ml). The constructions in the TFBS region were confirmed by diagnostic PCR and sequencing.

Detecting Expression Level Shifts with Real-Time PCR

To compare the mRNA levels of genes in the mutagenesis group and in the control group, we used the SYBR green core reaction to perform quantitative PCR (Applied Biosystems model 7300 Real-Time PCR System, Carlsbad, CA). Real-time PCR analyses were carried out on a final volume of 25 μl containing 40 ng of the cDNA sample, 50 nM of each gene-specific primer, and 12.5 μl of the SYBER green taq premixture. The PCR conditions included enzyme activation at 50 °C for 2 min and 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 s, and annealing/extension at 60 °C for 1 min. To verify that a single product had been amplified, a dissociation curve was generated at the end of each PCR cycle using the software provided by the Applied Biosystems 7300 Real-Time PCR System (version 1.4). The relative expression of each gene was normalized to that of the *ACT1* gene (ΔCt) and quantified with the ΔΔCt relative quantification method. The relative expression ratio was calculated according to ABI's guideline, that is, ratio (Swapped/BY4741) = $2^{(-\Delta\Delta Ct)}$. The amplification efficiency of each primer pair was tested by using 2-fold serial dilutions of the templates as suggested by ABI; and the amplification efficiencies of the target gene and the reference gene were approximately equal. Finally, the mRNA levels of the candidate genes were compared, using a paired *t*-test.

Results

TFBS Numbers and Densities

We selected 1,400 genes with known TSSs (Nagalakshmi et al. 2008) and orthologs in all strains used for our study. TFBSs for the proximal promoter regions of these genes were predicted based on 177 yeast TF binding profiles, obtained from the JASPAR database (Portales-Casamar et al. 2010). The numbers of predicted TFBSs were similar among all *S. cerevisiae* strains and among *S. paradoxus* strains (table 2). *Saccharomyces paradoxus* strains have slightly higher numbers of predicted TFBSs than *S. cerevisiae* strains (Mann–Whitney *U* test *P* value: 9.99×10^{-4}).

In total, 590 of the 1,400 genes have an intergenic region of at least 300 bp upstream of the TSS without any major overlap to the promoter region of the neighboring gene (see Materials and Methods). We compared the average TFBS densities (number of TFBSs per 100 bp) of the proximal promoter regions and the distal promoter regions of these genes. Significantly more TFBSs were detected in the proximal promoter regions than in the distal promoter regions: The mean TFBS density in proximal promoters for all strains was 3.96, whereas that in the distal promoters was 3.46 (Wilcoxon signed-rank test *P* value: 1.85×10^{-6}).

We also examined the nucleotide compositions in the two regions (supplementary table S1, Supplementary Material online). In all strains, the distal promoter regions have a higher GC content than the proximal promoter regions (mean for all strains: 35.89% vs. 35.54%). However, this difference was not statistically significant (Wilcoxon rank-sum test *P* value: 0.2611).

TFBS Gains and Losses in Different *S. cerevisiae* and *S. paradoxus* Strains

We determined putative TFBS gains and losses for each branch of the phylogenetic tree (fig. 1) based on the reconstructed ancestral sequences of internal nodes (fig. 2). Within the *S. cerevisiae* clade (fig. 2*b*) and within the *S. paradoxus* clade (fig. 2*c*) gains and losses of TFBSs are relatively balanced on each branch. However, when we classify the differences between the *S. cerevisiae* strain BY4741 (which is closest to the *S. cerevisiae* reference strain S288C) and the *S. paradoxus* strain CBS432 (from which the *S. paradoxus* reference genome was derived) as gains or losses on either lineage (excluding parallel gains and losses), *S. paradoxus* appears to have experienced relatively fewer TFBS losses than *S. cerevisiae* and fewer TFBS changes in general (table 3; Fisher's exact test *P* value: 1.11×10^{-14}). The lower number of TFBS losses in *S. paradoxus* compared with *S. cerevisiae* is consistent with an earlier study on TFBS gains and losses in different *Saccharomyces* species (Doniger and Fay 2007).

Similarly, the wild *S. cerevisiae* strain UWOPS87-2421 appears to have undergone fewer losses and fewer TFBS changes in general than the fermentation strain Y12 (Fisher's exact test *P* value: 0.006; table 4). The vineyard strain RM11-1a exhibits slightly more TFBS losses than gains and more changes in general than the laboratory strain BY4741. However, this difference is not statistically significant due to the low total number of observed changes between the two strains (Fisher's exact test *P* value: 0.084; table 5).

These observations led us to ask whether fermentation strains have generally experienced more TFBS losses than gains from the common *S. cerevisiae* ancestor to the extant strains, in comparison with other *S. cerevisiae* strains. This was indeed the case (Wilcoxon rank-sum test: *P* value: 0.048). In addition to Y12, we also compared each of the other two fermentation strains with its closest wild relative (supplementary tables S3 and S4, Supplementary Material online). In the comparison between RM11-1a and YPS606 the same trend is visible, although not statistically significant because of the relatively low total number of differences (Fisher's exact test *P* value: 0.174; supplementary table S3, Supplementary Material online), but it is absent between DBVPG6044 and

Table 2
Numbers of Predicted TFBSs for the Proximal Promoter Regions of 1,400 Yeast Genes

	S. cerevisiae Strains								
	BY4741	RM11-1a	YJM789	SK1	DBVPG6044	UWOPS03-461.4	Y12	YPS606	UWOPS87-2421
Total no. of TFBSs	9,008	8,961	9,031	8,999	8,994	8,988	8,943	9,006	9,009
No. of TFBSs per 100 bp	3.94	3.92	3.95	3.94	3.94	3.94	3.92	3.94	3.94

	MRCA of all *S. cerevisiae* Strains[a]	MRCA of *S. paradoxus* and *S. cerevisiae*[a]	MRCA of All *S. paradoxus* Strains[a]	*S. paradoxus* Strains				
				Y8.5	YPS138	N-44	CBS5829	CBS432
Total no. of TFBSs	8,997	9,276	9,283	9,321	9,131	9,331	9,346	9,354
No. of TFBSs per 100 bp	3.94	4.06	4.06	4.08	3.99	4.09	4.09	4.09

[a]Reconstructed ancestral sequence.

Table 3
TFBS Gains and Losses in the Proximal Promoter of *Saccharomyces cerevisiae* (BY4741) and *Saccharomyces paradoxus* (CBS432)

Type of Event	Total Changes[a] along the Branches Leading to BY4741 (*S. cerevisiae*) and CBS432 (*S. paradoxus*) from Their MRCA		Differences[b] between BY4741 and CBS432 that Were Assigned to a Lineage		Parallel Events between the BY4741 and CBS432 Lineages
	BY4741	CBS432	BY4741	CBS432	
Gains	2,702	2,106	2,596	2,000	106
Losses	2,970	2,028	2,013	1,071	957
Turnovers	196[c] + 57[d]	410[c] + 25[d]			

[a]Total numbers of TFBS changes in the proximal promoters of 1,400 genes along the lineages leading to the *S. cerevisiae* strain BY4741 or the *S. paradoxus* strain CBS432 compared with their MRCA (including parallel gains and losses). TFBS turnover events are classified into two categories: multiple-branch turnovers and single-branch turnovers.
[b]Pairwise differences between the BY4741 strain and the *S. paradoxus* strain CBS432 were classified as gains or losses with regard to the reconstructed sequence for the common ancestor of *S. cerevisiae* and *S. paradoxus*.
[c]Multibranch turnovers.
[d]Single-branch turnovers.

UWOPS03-461.4 (supplementary table S4, Supplementary Material online, Fisher's exact test *P* value: 0.88).

Evolutionary Conservation of Proximal Promoters versus Distal Promoters

We investigated whether TFBSs in proximal promoters are not only more frequent but also better conserved than those in distal promoters:

1. As the number of TFBS losses which can occur on a branch is dependent on the number of TFBSs in the ancestral sequence, we divided the total number of TFBS losses on all branches by the total number of TFBSs of all ancestral sequences for each gene, and compared the two regions. There were indeed significantly more losses per ancestral TFBS in the distal promoter regions than in the proximal promoter regions (0.23 vs. 0.19 mean losses per ancestral TFBS; Wilcoxon signed-rank test *P* value: 7.53×10^{-5}).
2. As the number of TFBSs which can be gained on a branch depends on the length of the sequence under consideration, we divided the total number of TFBS gains by the length of the region for the proximal promoters and for the distal promoters. There were slightly more TFBS gains in distal promoters than in proximal promoters; the difference was not significant (0.09 vs. 0.086 mean total gains per base pair; Wilcoxon signed-rank test *P* value: 0.21).

To assess the impact of the peak region around −115 bp on the observed TFBS conservation differences between proximal and distal promoters and to control for the variable lengths of distal promoter regions, we compared a set of three 80-bp regions:

1. Region 1: The 80-bp region immediately upstream of the TSS.
2. Region 2: From 91 to 170 bp upstream of the TSS.
3. Region 3: From 201 to 280 bp upstream of the TSS.

Region 2, which includes the peak of the TFBS density, is expected to contain more functional TFBSs than the other two regions (Lin et al. 2010). As expected, Region 2 was found to have a significantly higher TFBS density (4.16 TFBS per 100 bp) than Region 1 (1.93 TFBS per 100 bp; Wilcoxon signed-rank test *P* value: 2.2×10^{-16}) and Region 3 (2.23 TFBS per 100 bp; Wilcoxon signed-rank test *P* value: 2.2×10^{-16}). Also, the number of TFBS gains per base pair was significantly higher

Table 4

TFBS Gains and Losses in the Proximal Promoter of the *Saccharomyces cerevisiae* UWOPS87-2421 and Y12 Lineages

	Total Number of Changes[a] along the Y12 and UWOPS87-2421 Branches		Differences[b] between Y12 and UWOPS87-2421		Parallel Gains and Losses[c] between the Y12 and UWOPS87-2421 Lineages
	Y12	UWOPS87-2421	Y12	UWOPS87-2421	
Gains	129	136	129	136	0
Losses	171	112	167	108	4
Turnovers	2	0			

[a]Total numbers of TFBS gain and loss events along the lineages leading to the *S. cerevisiae* strains Y12 and UWOPS87-2421 from their MRCA.
[b]Pairwise differences in TFBS numbers between the wild strain UWOPS87-2421 and the fermentation strain Y12, classified into gains and losses.
[c]Parallel TFBS gains and losses in the two strains, with regards to the reconstructed ancestral sequence.

Table 5

TFBS Gains and Losses in the Proximal Promoters of BY4741 and RM11-1a

	Total Number of Changes[a] along the BY4741 and RM11-1a Branches		Differences[b] between BY4741 and RM11-1a		Parallel Gains and Losses[c] between the BY4741 and RM11-1a Lineages
	BY4741	RM11-1a	BY4741	RM11-1a	
Gains	133	168	129	164	4
Losses	123	205	118	200	5
Turnovers	1	2			

[a]Total number of TFBS gain and loss events along the lineages leading to the *Saccharomyces cerevisiae* strains BY4741 and RM11-1a from their MRCA.
[b]Pairwise differences in TFBSs between the laboratory strain BY4741 and the vineyard strain RM11-1a, classified into gains and losses.
[c]Parallel TFBS gains and losses in the two strains, with regards to the reconstructed ancestral sequence.

for Region 2 (mean: 0.088 gains per base pair) than for Region 1 (mean: 0.052 gains per base pair; Wilcoxon signed-rank test P value: 2.2×10^{-16}) and Region 3 (mean: 0.056 gains per base pair; Wilcoxon signed-rank test P value: 2.2×10^{-16}). The number of TFBS losses per ancestral TFBS was lower in Region 2 (mean: 0.24) than in Region 1 (mean: 0.31) and Region 3 (0.27). However, the difference was statistically significant only for the comparison between Region 2 and Region 1 (Wilcoxon signed-rank test P value: 0.001).

Comparison of Different Constraint Categories

According to the analyses above and a previous study (Lin et al. 2010), the proximal promoter regions are expected to contain more TFBSs than the distal promoter regions. For this reason, we focused on the proximal promoters in the following analyses.

We compared the numbers of TFBS gain and loss events in the proximal promoter of genes in different categories of selective or functional constraint. Essential genes are found to have experienced significantly fewer loss events than nonessential genes (the mean number of losses per ancestral TFBS: 0.158 for essential genes vs. 0.184 for nonessential genes; Wilcoxon rank-sum test P value: 0.015), whereas the numbers of gain events per base pair were similar for both groups (0.081 vs. 0.084).

Similarly, the proximal promoter regions of genes with no known protein–protein interaction (PPI) partners have experienced more TFBS losses than those with five (median) or more PPI partners (mean: 0.186 vs. 0.161; Wilcoxon rank-sum test P value: 0.026).

Additionally, we compared proximal promoters with a well-defined nucleosome-free region close to the TSS, called depleted proximal-nucleosome promoters, with those without such a region, called occupied proximal-nucleosome promoters (Tirosh and Barkai 2008). There were no significant differences in TFBS densities (3.73 vs. 4.33 TFBSs per 100 bp) and in the number of gain or loss events (0.08 vs. 0.082 gains per 100 bp; 0.177 vs. 0.168 losses per ancestral TFBS) between these two groups (Wilcoxon rank-sum test P values: 0.526, 0.505, and 0.751, respectively). In contrast, for the comparison between TATA box containing and TATA-less promoters (Basehoar et al. 2004), there were significant differences in the mean density of predicted TFBSs (5.01 vs. 3.85 TFBSs per 100 bp; Wilcoxon rank-sum test P value: 1.56×10^{-5}) and in the mean number of TFBS losses per ancestral TFBS (0.156 vs. 0.182; Wilcoxon rank-sum test P value: 7.31×10^{-3}).

Relationship between TFBS Changes and Gene Expression Differences in *cis* and *trans*

We examined the relationship between changes in TFBSs and gene expression differences between RM11-1a and BY4741 (Schaefke et al. 2013). The proximal promoter regions of genes with a significant *cis* component in expression evolution

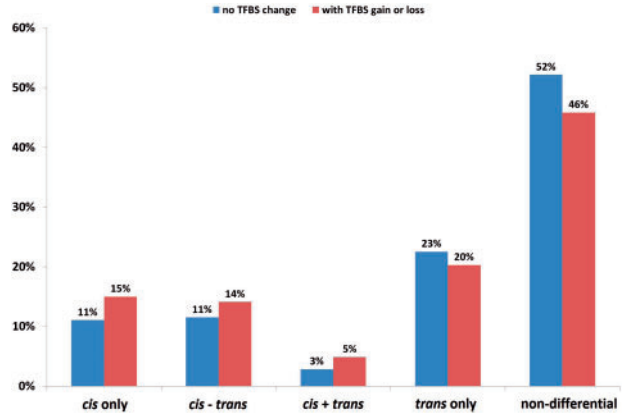

Fig. 3.—Distributions of genes with TFBS gains/losses (227 genes) or without TFBS gains/losses (888 genes) between the *S. cerevisiae* strains BY4741 and RM11-1a in different *cis/trans* categories. Genes with TFBS changes in the proximal promoter regions are more frequently found in the categories with a significant *cis* component ("*cis* only," "*cis* + *trans*," and "*cis* – *trans*").

have experienced significantly more TFBS changes (mean: 0.76 differences per gene) between the two strains than those of genes without a significant *cis* component (mean: 0.38; Wilcoxon rank-sum test *P* value: 6.67×10^{-3}). Interestingly, this appears to be largely due to the effect of TFBS losses in RM11-1a: Genes with a significant *cis* component have a mean of 0.33 losses in RM11-1a, whereas the other genes have significantly fewer losses in RM11-1a (mean: 0.10) (Wilcoxon rank-sum test *P* value: 3.5×10^{-4}). Moreover, the numbers of TFBS gains in BY4741 and in RM11-1a were also higher for genes with a significant *cis* component than for other genes in our data set (mean: 0.126 vs. 0.093 and 0.215 vs. 0.087, respectively), but this difference is not statistically significant (Wilcoxon rank-sum test *P* values: 0.53 and 0.08, respectively).

Grouping genes into those with or without TFBS changes between RM11-1a and BY4741, we observe the same relationship: Genes with TFBS differences have a mean *P* value of 0.26 (median: 0.12) for the *cis* component of gene expression differences, whereas genes without predicted TFBS changes between the two strains have a mean *cis* *P* value of 0.32 (median: 0.24; Wilcoxon rank-sum test *P* value: 8.5×10^{-4}). This difference is also reflected in the distribution of genes with or without TFBS changes among the different *cis/trans* categories (fig. 3). In general, genes with a significant *cis* component are more likely to have experienced TFBS gains or losses (Fisher's exact test *P* value: 0.012; supplementary table S5, Supplementary Material online).

We also related the changes in TFBSs to gene expression data from an interspecific comparison between the *S. cerevisiae* strain BY4741 and the *S. paradoxus* strain CBS432 (Tirosh, Reikhav, et al. 2009). The mean number of TFBS differences for genes with a significant *cis* component was 5.82, compared with 5.19 for genes without a significant *cis* component (Wilcoxon rank-sum test *P* value: 0.051). Only the number of changes in the lineage leading to CBS432 showed a significant difference between genes with a significant *cis* component (mean number of gains: 1.51; mean number of losses: 0.86) and those without (mean number of gains: 1.30; mean number of losses: 0.73; Wilcoxon rank-sum test *P* values: 0.049 and 0.043, respectively). Genes with a significant *cis* component are more likely to have experienced a larger number of TFBS gains or losses, but reaching statistical significance only for changes on the *S. paradoxus* lineage (Fisher's exact test *P* value: 0.019 for TFBS gains or losses to CBS432; Fisher's exact test *P* value: 0.24 for TFBS gains or losses to BY4741; supplementary table S5*b*, Supplementary Material online).

No significant difference in TFBS changes between genes with a significant *trans* component in expression evolution and those without a significant difference in *trans* was detected in either the intraspecific or the interspecific comparison.

To provide an example of gene expression influenced by TFBS gain/loss events, we constructed two mutant BY4741 strains, targeting the Ndt80p and Reb1p binding sites in the proximal promoter region of the *MCT1* gene (YOR221C) with site-directed mutagenesis. Ndt80p is a positive regulator of transcription, whereas Reb1p bound to DNA acts to block RNA polymerase II read-through transcription. In each case, one nucleotide was changed to obtain the same nucleotide as in RM11-1a (in which the respective TFBS was not predicted in our scan). Both substitutions significantly changed expression levels of the *MCT1* gene. As expected, this change was in opposite directions (supplementary table S6, Supplementary Material online): Expression of *MCT1* increased after turning the Reb1p binding site nonfunctional, but decreased after targeting the Ndt80p binding site.

Comparison of TFBS Changes in Different Lineages

We examined whether the promoter regions of specific genes or the binding sites of specific TFs have undergone more TFBS changes in the lineages leading from wild strains to fermentation or lab strains.

For this purpose we divided our phylogenetic tree of *S. cerevisiae* strains into three clades, each of which included at least one fermentation and one wild strain (supplementary fig. S3, Supplementary Material online): 1) Clade A: BY4741 (lab), RM11-1a (fermentation), and YPS606 (wild); (2) clade B: Y12 (fermentation) and UWOPS87-2421 (wild), and (3) clade C: SK1 (lab), DBVPG6044 (fermentation), and UWOPS03-461.4 (wild). Then, we conducted three groups of pairwise comparisons (with each comparison considering differences classified as gains and losses with respect to the MRCA of the two strains): 1) Each lab strain versus its closest wild relative (clade A: BY4741 vs. YPS606, clade C: SK1 vs. UWOPS03-461.4), 2) each lab strain versus the closest fermentation strain

(clade A: BY4741 vs. RM11-1a, clade C: SK1 vs. DBVPG6044), and 3) each fermentation strain versus its closest wild relative (clade A: RM11-1a vs. YPS606, clade B: Y12 vs. UWOPS87-2421 and clade C: DBVPG6044 vs. UWOPS03-461.4). In every group (niche comparison), we obtained the sum of TFBS gains or losses from the respective clade ancestor for the strains belonging to the same ecological niche (e.g., gains for RM11-1a vs. YPS606 + gains for Y12 vs. UWOPS87-2421 + gains for DBVPG6044 vs. UWOPS03-461.4 = total fermentation strain gains). We pooled the data for each TF with binding site gains or losses in the same way. Then, we performed Fisher's exact test for each gene/TF and calculated the difference between gains and losses for each gene/TF (total gains per gene/TF – total losses per gene/TF).

1. Lab versus wild: Out of 1,400 proximal promoter regions in our analysis, 322 showed TFBS differences between the lab strain BY4741 and the wild strain YPS606, and 379 genes showed differences between the lab strain SK1 and the wild strain UWOPS03-461.4. A total of 118 genes showed differences in both comparisons (81 expected by chance, hypergeometric test *P* value: 8.55×10^{-8}).
2. Lab versus fermentation: Out of the 1,400 proximal promoter regions 259 showed TFBS differences between BY4741 and the vineyard strain RM11-1a, whereas 92 showed differences between SK1 and the fermentation strain DBVPG6044. A total of 21 genes showed TFBS differences in both comparisons (17 expected by chance; hypergeometric test *P* value: 0.11).
3. Fermentation versus wild: 359 proximal promoter regions showed differences in TFBSs between RM11-1a and YPS606, 264 between Y12 and UWOPS87-2421, and 377 between DBVPG6044 and UWOPS03-461.4. In total, 126 genes had experienced changes between the fermentation strain and the wild strain in clades A and B (expected: 68; hypergeometric test *P* value: 8.03×10^{-19}), 119 were different between the two niches in both clade A and clade C (expected: 97; hypergeometric test *P* value: 9.34×10^{-4}), and 95 genes were in the intersection between clade B and clade C (expected: 71; hypergeometric test *P* value: 1.16×10^{-4}). Finally, a total of 51 genes (supplementary table S7, Supplementary Material online) showed TFBS changes in all three pairwise fermentation versus wild comparisons (multivariate hypergeometric test *P* value: 1.78×10^{-4}).

From the above analyses, we found significantly more TFBS gains per TF in fermentation strains than in wild strains (mean: 3.71 vs. 2.8; Wilcoxon rank-sum test *P* value: 4.33×10^{-4}), and also significantly more losses (mean: 4.24 vs. 2.67; Wilcoxon rank-sum test *P* value: 5.66×10^{-7}). The difference in losses is larger, so that the mean number of gains–losses per TF is negative in the fermentation strains (−0.525), but positive in wild strains (0.132; Wilcoxon rank-sum test *P* value: 0.046; fig. 4).

The three TFs with the highest fermentation versus wild loss ratios were *DAL80* (YKR034W), *ADR1* (YDR216W), and *AAF1* (YML081W) (supplementary table S8, Supplementary Material online). Dal80p is a negative regulator of genes in multiple nitrogen degradation pathways (Chisholm and Cooper 1982; Cunningham and Cooper 1991). Adr1p is required for the transcription of the glucose-repressed gene *ADH2* and of genes required for ethanol, glycerol, and fatty acid utilization (Simon et al. 1991; Tachibana et al. 2005). Aaf1p is a TF that regulates acetate production (Walkey et al. 2012).

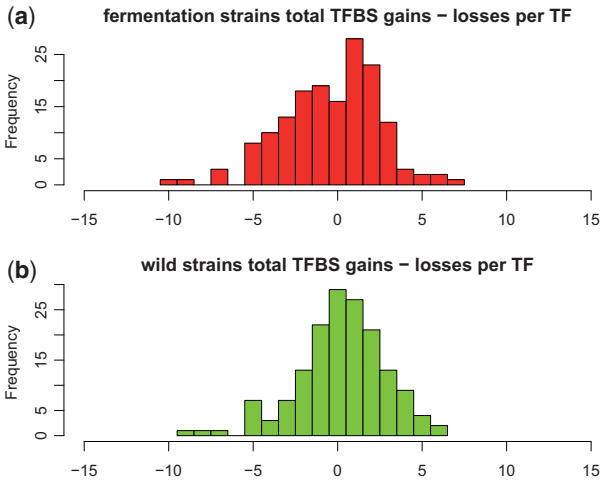

Fig. 4.—TFBS gains–losses per TF in (*a*) fermentation versus (*b*) wild strains.

Relationship between TFBS Changes and TF Evolution

The TFBS matrices used in our study were derived from experimental data in laboratory strains of *S. cerevisiae*. Thus, not all detected putative gains of TFBSs in BY4741 or losses in other strains or species necessarily reflect real changes of TF binding, but might be the consequence of changes in the DNA binding region of specific TFs and compensatory evolution in the TFBS. To investigate this issue, we searched for TFs which were overrepresented among TFBS gains in *S. cerevisiae* and/or losses in *S. paradoxus,* because we expected TFs with different DNA binding specificity between *S. cerevisiae* and *S. paradoxus* predominantly to fall into this category. Then, we examined whether these TFBSs showed amino acid differences in their DNA binding regions and consistent patterns in the other species *S. mikatae* and *S. bayanus*. Among the three TFs with the highest ratio of *S. cerevisiae* gains and *S. paradoxus* losses to *S. cerevisiae* losses and *S. paradoxus* gains (Smp1p, Arg81p, and Stb5p), only the Stb5p protein shows major amino acid diversity in its DNA binding region (protein coordinates 16–60; supplementary fig. S4, Supplementary Material online) among the *Saccharomyces* species, whereas the others are conserved.

Discussion

TFBSs are an important nexus for the interaction of *trans*-regulatory factors with *cis*-regulatory regions. Our results confirm previous work that most functional TFBSs in *Saccharomyces* yeasts can be found in a narrow region up to 200 bp upstream of the TSS (Venters and Pugh 2009; Lin et al. 2010). The density and the evolutionary conservation of TFBSs in this region are higher than in the distal promoter region (table 2 and supplementary table S2, Supplementary Material online).

We show a strong relationship of TFBS changes in proximal promoters between the two *S. cerevisiae* strains BY4741 and RM11-1a and the *cis* component of gene expression differences between these two strains (supplementary table S5, Supplementary Material online, and fig. 3). We also present experimental proof that a single nucleotide change in a predicted TFBS, which has undergone a gain/loss event between the two strains, can lead to significant expression differences (supplementary table S6, Supplementary Material online). This shows the possibility that a combination of hybrid experiments between closely related strains of interest and computational prediction of TFBS gains and losses could be utilized to pinpoint the evolutionary events which have led to expression differences and phenotypical divergence between these strains.

The relationship between TFBS changes and the *cis* component of gene expression differences between the species *S. cerevisiae* and *S. paradoxus* is less pronounced. The reason for this could be the different nature of the data: High-throughput sequencing in the case of the within-species comparison (Schaefke et al. 2013) and microarray data in the case of the between-species comparison (Tirosh, Reikhav, et al. 2009). Another possible explanation would be that different evolutionary forces (He et al. 2011) have shaped the gain and loss of TFBSs on the two lineages: Although the *S. cerevisiae* lineage might have experienced more mutation-driven losses of TFBSs and a partially compensatory gain of different binding sites with similar regulatory effect, the *S. paradoxus* lineage might have been under stronger selective constraint, experiencing less TFBS changes, but with a higher proportion of these changes reflecting positive selection and contributing to gene expression changes (reflected in our result that there is a significant relationship between TFBS changes and the *cis* component of gene expression differences between BY4741 and CBS432 on the *S. paradoxus* lineage, but not on the *S. cerevisiae* lineage).

In agreement with this hypothesis, we found that *S. paradoxus* has experienced comparatively fewer TFBS losses than *S. cerevisiae*. Within *S. cerevisiae*, fermentation strains appear to have evolved more rapidly and lost TFBSs more frequently than strains from other environments. A possible explanation for this phenomenon might be that *S. cerevisiae* has shifted from tree barks to more sugar-rich environments, which reduced the need of fine-regulation of gene expression. This trend has been continued in the diversification of domesticated *S. cerevisiae* strains, especially those used in alcoholic fermentation, from those living in environments more similar to *S. paradoxus*. Future studies to compare larger numbers of different strains, especially wild strains isolated from tree bark and fermentation strains, might be able to validate or reject this hypothesis.

Possible candidates for TFs which have experienced losses of TFBSs in their target genes as result of this environmental shift in fermentation strains are *AAF1* and *ADR1*. Aaf1p regulates the production of acetic acid during wine fermentation, which is an important factor in determining the sensory quality of the product (Walkey et al. 2012). Adr1p is involved in the response to carbon-source availability and has been shown to mediate glucose derepression (Simon et al. 1991; Tachibana et al. 2005), which "acts as an impediment to effective alcoholic fermentation under glucose-rich fermentative conditions" (Watanabe et al. 2013). Again, here a larger-scale analysis is necessary to evaluate the roles of *ADR1* and *AAF1* TFBSs and whether their increased losses in fermentation strains have occurred because of a relaxation of selective constraint or as an adaptation to the new environment. It will also be of interest to examine how differences in *ADR1* and *CAT8* TFBSs contribute to the marked differences between the lab strain BY4741 and the fermentation strain RM11-1a with regards to glucose consumption rate and transition to nonfermentative metabolism (Chang et al. 2008). Competition experiments can help to determine whether the loss of Aaf1p or Adr1p binding sites confers a competitive advantage over wild-type strains in sugar-rich environments or might have been the result of human selection for specific qualities of alcoholic beverages, for example, ethanol content or acidity.

Supplementary Material

Supplementary figures S1–S4 and tables S1–S8 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank J.J. Emerson, Daryi Wang, James Lin, Krishna Swamy and Sean Chun-Chang Chen for helpful discussions and advice, and Li-Chuan Weng for experimental assistance. This study was supported by intramural grants from Academia Sinca, Taiwan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Literature Cited

Arbiza L, et al. 2013. Genome-wide inference of natural selection on human transcription factor binding sites. Nat Genet. 45:723–729.

Badis G, et al. 2008. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell. 32:878–887.

Basehoar AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA box-containing genes. Cell 116:699–709.
Bergstrom A, et al. 2014. A high-definition view of functional genetic variation from natural yeast genomes. Mol Biol Evol. 31:872–888.
Boyle AP, et al. 2014. Comparative analysis of regulatory information and circuits across distant species. Nature 512:453–456.
Chang YW, et al. 2008. Roles of *cis*- and *trans*-changes in the regulatory evolution of genes in the gluconeogenic pathway in yeast. Mol Biol Evol. 25:1863–1875.
Chen K, van Nimwegen E, Rajewsky N, Siegal ML. 2010. Correlating gene expression variation with *cis*-regulatory polymorphism in *Saccharomyces cerevisiae*. Genome Biol Evol. 2:697–707.
Chisholm G, Cooper TG. 1982. Isolation and characterization of mutants that produce the allantoin-degrading enzymes constitutively in *Saccharomyces cerevisiae*. Mol Cell Biol. 2:1088–1095.
Cock PJA, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423.
Connelly CF, Skelly DA, Dunham MJ, Akey JM. 2013. Population genomics and transcriptional consequences of regulatory motif variation in globally diverse *Saccharomyces cerevisiae* strains. Mol Biol Evol. 30:1605–1613.
Coulombe-Huntington J, Xia Y. 2012. Regulatory network structure as a dominant determinant of transcription factor evolutionary rate. PLoS Comput Biol. 8:e1002734.
Criscuolo A. 2011. morePhyML: improving the phylogenetic tree space exploration with PhyML 3. Mol Phylogenet Evol. 61:944–948.
Cunningham TS, Cooper TG. 1991. Expression of the *DAL80* gene, whose product is homologous to the GATA factors and is a negative regulator of multiple nitrogen catabolic genes in *Saccharomyces cerevisiae*, is sensitive to nitrogen catabolite repression. Mol Cell Biol. 11:6205–6215.
Denver DR, et al. 2005. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. Nat Genet. 37:544–548.
Doniger SW, Fay JC. 2007. Frequent gain and loss of functional transcription factor binding sites. PLoS Comput Biol. 3:e99.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.
Emerson JJ, et al. 2010. Natural selection on *cis* and *trans* regulation in yeasts. Genome Res. 20:826–836.
Gout JF, Lynch M. 2015. Maintenance and loss of duplicated genes by dosage subfunctionalization. Mol Biol Evol. 32:2141–2148.
Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics 27:1017–1018.
Gu Z, et al. 2005. Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 102:1092–1097.
Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol. 537:113–137.
Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59:307–321.
Haygood R, Fedrigo O, Hanson B, Yokoyama KD, Wray GA. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. Nat Genet. 39:1140–1144.
He BZ, Holloway AK, Maerkl SJ, Kreitman M. 2011. Does positive selection drive transcription factor binding site turnover? A test with *Drosophila cis*-regulatory modules. PLoS Genet. 7:e1002053.
Heger P, George R, Wiehe T. 2013. Successive gain of insulator proteins in arthropod evolution. Evolution 67:2945–2956.
Johnson LJ, et al. 2004. Population genetics of the wild yeast *Saccharomyces paradoxus*. Genetics 166:43–52.
Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241–254.
Koshi JM, Goldstein RA. 1996. Probabilistic reconstruction of ancestral protein sequences. J Mol Evol. 42:313–320.
Li YD, et al. 2009. Detecting positive selection in the budding yeast genome. J Evol Biol. 22:2430–2437.
Liang H, Lin YS, Li WH. 2008. Fast evolution of core promoters in primate genomes. Mol Biol Evol. 25:1239–1244.
Lin Z, et al. 2013. Identifying *cis*-regulatory changes involved in the evolution of aerobic fermentation in yeasts. Genome Biol Evol. 5:1065–1078.
Lin Z, Wu WS, Liang H, Woo Y, Li WH. 2010. The spatial distribution of *cis* regulatory elements in yeast promoters and its implications for transcriptional regulation. BMC Genomics 11:581.
Liti G, et al. 2009. Population genomics of domestic and wild yeasts. Nature 458:337–341.
MacIsaac KD, et al. 2006. An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. BMC Bioinformatics 7:113.
Nagalakshmi U, et al. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349.
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.
Plech M, de Visser JA, Korona R. 2014. Heterosis is prevalent among domesticated but not wild strains of *Saccharomyces cerevisiae*. G3 (Bethesda) 4:315–323.
Portales-Casamar E, et al. 2010. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. 38:D105–D110.
R-Core-Team. 2013. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
Reece-Hoyes JS, et al. 2013. Extensive rewiring and complex evolutionary dynamics in a C. *elegans* multiparameter transcription factor network. Mol Cell. 51:116–127.
Sanner MF. 1999. Python: a programming language for software integration and development. J Mol Graph Model. 17:57–61.
Schaefke B, et al. 2013. Inheritance of gene expression level and selective constraints on *trans*- and *cis*-regulatory changes in yeast. Mol Biol Evol. 30:2121–2133.
Simon M, Adam G, Rapatz W, Spevak W, Ruis H. 1991. The *Saccharomyces cerevisiae ADR1* gene is a positive regulator of transcription of genes encoding peroxisomal proteins. Mol Cell Biol. 11:699–704.
Skelly DA, et al. 2013. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. Genome Res. 23:1496–1504.
Spor A, et al. 2009. Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of *Saccharomyces cerevisiae*. BMC Evol Biol. 9:296.
Sung HM, et al. 2009. Roles of *trans* and *cis* variation in yeast intraspecies evolution of gene expression. Mol Biol Evol. 26:2533–2538.
Tachibana C, et al. 2005. Combined global localization analysis and transcriptome data identify genes that are directly coregulated by Adr1 and Cat8. Mol Cell Biol. 25:2138–2146.
Talevich E, Invergo BM, Cock PJA, Chapman BA. 2012. Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. BMC Bioinformatics 13:209.
Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. Genome Res. 18:1084–1091.
Tirosh I, Barkai N, Verstrepen KJ. 2009. Promoter architecture and the evolvability of gene expression. J Biol. 8:95.

Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. Science 324:659–662.
Tirosh I, Weinberger A, Bezalel D, Kaganovich M, Barkai N. 2008. On the relation between promoter divergence and gene expression evolution. Mol Syst Biol. 4:159.
Tsai ZT, et al. 2012. Evolution of *cis*-regulatory elements in yeast de novo and duplicated new genes. BMC Genomics 13:717.
Venters BJ, Pugh BF. 2009. A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. Genome Res. 19:360–371.
Walkey CJ, Luo Z, Madilao LL, van Vuuren HJ. 2012. The fermentation stress response protein Aaf1p/Yml081Wp regulates acetate production in *Saccharomyces cerevisiae*. PLoS One 7:e51551.
Watanabe D, et al. 2013. Accelerated alcoholic fermentation caused by defective gene expression related to glucose derepression in *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem. 77:2255–2262.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yang Z, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650.
Zhu C, et al. 2009. High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. 19:556–566.
Zhu J, Zhang MQ. 1999. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. Bioinformatics 15:607–611.

Associate editor: Takashi Gojobori