

Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production

Hal Alper,^{1,3} Joel Moxley,¹ Elke Nevoigt,^{1,2} Gerald R. Fink,³ Gregory Stephanopoulos^{1*}

Global transcription machinery engineering (gTME) is an approach for reprogramming gene transcription to elicit cellular phenotypes important for technological applications. Here we show the application of gTME to *Saccharomyces cerevisiae* for improved glucose/ethanol tolerance, a key trait for many biofuels programs. Mutagenesis of the transcription factor Spt15p and selection led to dominant mutations that conferred increased tolerance and more efficient glucose conversion to ethanol. The desired phenotype results from the combined effect of three separate mutations in the *SPT15* gene [serine substituted for phenylalanine (Phe¹⁷⁷Ser) and, similarly, Tyr¹⁹⁵His, and Lys²¹⁸Arg]. Thus, gTME can provide a route to complex phenotypes that are not readily accessible by traditional methods.

The production of desirable compounds from microbes can often require a complete reprogramming of their innate metabolism. The evolution of such complex traits requires simultaneous modification in the expression levels of many genes, which may not be achievable by sequential multigene modifications. Furthermore, the identification of genes requiring perturbation may be largely unanticipated by conventional pathway analysis. The cellular engineering approach termed “global transcription machinery engineering” (gTME) alters [via error-prone polymerase chain reaction (PCR) mutations] key proteins regulating the global transcriptome and generates, through them, a new type of diversity at the transcriptional level.

This approach has already been demonstrated by engineering sigma factors in prokaryotic cells (1), but the increased complexity of eukaryotic transcription machinery raises the question of whether gTME can be used to improve traits in more complex organisms. For example, eukaryotic systems have more specialization—three RNA polymerase enzymes with separate functions, whereas only one exists in prokaryotes. Moreover, nearly 75 components have been classified as general transcription factors or coactivators of the RNA polymerase II (RNA Pol II) system (2), and loss of function for many of these components is lethal. Components of the general factor RNA Pol II transcription factor D (TFIID) include the

TATA-binding protein (*SPT15*) and 14 other associated factors (TAFs) that are collectively thought to be the main DNA binding proteins regulating promoter specificity in yeast (2–5). Mutations in a TATA-binding protein have been shown to change the preference of the three polymerases and to play an important role in promoter specificity (6).

Successful fermentations to produce ethanol using yeast require tolerance to high concentrations of both glucose and ethanol. These cellular characteristics are important because very high gravity (VHG) fermentations, which are common in the ethanol industry, give rise to high sugar concentrations (and thus high osmotic pressure), at the beginning of the process, and high ethanol concentration at the end of a batch (7, 8). As with ethanol tolerance in *Escherichia coli*, tolerance to ethanol and glucose mixtures does not seem to be a monogenic trait (9). Therefore, traditional methods of strain improvement have had limited success beyond the identification of medium supplementations and various chemical protectants (10–14).

To evaluate the approach of gTME in a eukaryotic system, two gTME mutant libraries were created from either *SPT15* (which encodes the TATA-binding protein) or one of the TATA-binding protein-associated factors, in this case, *TAF25* (15). The yeast screening and selection was performed in the background of the standard haploid *Saccharomyces cerevisiae* strain BY4741, which contains the endogenous, unmutated chromosomal copy of *SPT15* and *TAF25*. As such, this genetic screen uses a strain that expresses both the wild-type and mutated versions of the protein and, thus, permits the identification of dominant mutations that lead to novel functions in the presence of the unaltered chromosomal gene. These libraries were trans-

formed into yeast and were selected in the presence of elevated levels of ethanol and glucose. The *spt15* mutant library showed modest growth in the presence of 5% ethanol and 100 g/liter of glucose, so the stress was increased in the subsequent serial subculturing to 6% ethanol and 120 g/liter of glucose. After the subculturing, strains were isolated from plates, and plasmids containing mutant genes were isolated and retransformed into a fresh background, then tested for their capacity to grow in the presence of elevated glucose and ethanol levels. The best mutant obtained from each of these two libraries was assayed in further detail and sequenced.

The sequence characteristics of these altered genes conferring the best properties (one Spt15p and one Taf25p) are shown in Fig. 1A. Each of these mutated genes contained three mutations, with those of *spt15* localized to the second repeat element, which consists of a set of β sheets (5, 16). These specific triple mutations in the *taf25* and *spt15* mutant genes are thus referred to as the *taf25-300* and *spt15-300* mutations.

The *spt15-300* mutant outperformed the control at all concentrations tested, with the strain harboring the mutant protein providing up to 13-fold improvement in growth yield at some glucose concentrations (Fig. 1B and fig. S1). The *taf25-300* mutant was unable to grow in the presence of 6% ethanol, consistent with the observations seen during the enrichment and selection phase. Despite these increases in tolerance, the basal growth rate of these mutants in the absence of ethanol and glucose stress was similar to that of the control. Furthermore, the differences in behavior between the *spt15-300* mutant and *taf25-300* mutant suggest that mutations in genes encoding different members of the eukaryotic transcription machinery are likely to elicit different (and unanticipated) phenotypic responses.

The remainder of this study focuses on the *spt15-300* mutant, because this triple mutation set, in which serine is substituted for phenylalanine (Phe¹⁷⁷Ser), and similarly, Tyr¹⁹⁵His, and Lys²¹⁸Arg (F177S, Y195H, and K218R, respectively), provided the most desirable phenotype with respect to elevated ethanol and glucose. At ethanol concentrations above 10%, the *spt15-300* mutant exhibited statistically significantly improved cellular viability (over the course of 30 hours of culturing) above that of the control, even at concentrations as high as 20% ethanol by volume (Fig. 2, A and B, and fig. S2).

Transcriptional profiling revealed that the mutant *spt15-300* exhibited differential expression of hundreds of genes [controlled for false discovery (17)] in the unstressed condition (0% ethanol and 20 g/liter glucose) relative to cells expressing the wild-type *SPT15* (18). This analysis mainly used the unstressed condition, rather than the stressed (5% ethanol and 60 g/liter

¹Department of Chemical Engineering, Massachusetts Institute of Technology, Room 56-469, Cambridge, MA 02139, USA. ²Department of Microbiology and Genetics, Berlin University of Technology, Seestr. 13, D-13353 Berlin, Germany. ³Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA.

*To whom correspondence should be addressed. E-mail: gregstep@mit.edu

glucose), because expression ratios were more reliable under this condition owing to the similarity of growth rates, which made gene ex-

pression profiles more comparable (SOM text, part c, and table S3). It is noted that the impact of the ethanol and glucose stress had a variable

effect on many of the genes, and often, the stress did not further affect many of the genes selected using unstressed conditions (SOM text, part c). Although this widespread alteration in transcription is similar to that observed in *E. coli* with an altered sigma factor, the majority of the genes with altered expression are up-regulated, unlike the balanced distribution seen with *E. coli* (SOM text, part b, and fig. S3). The transcriptional reprogramming in the *spt15-300* mutant was quite broad, yet it exhibited some enrichment of certain functional groups such as oxidoreductase activity, cytoplasmic proteins, amino acid metabolism, and electron transport (SOM text, part b, and fig. S4). Unclassified genes or genes with no known function were also found with higher levels of expression. An analysis of promoter-binding sites, as well as a search for active gene subnetworks using the Cytoscape (19) framework, failed to show that a particular pathway or genetic network was predominately responsible for the observed genetic reprogramming (15).

To determine whether these up-regulated genes acted individually or as an ensemble to provide increased ethanol and glucose tolerance, we examined the effect of individual gene knockouts on the phenotype. Twelve of the most highly expressed genes in the mutant under the unstressed conditions of 0% ethanol and 20 g/liter of glucose were selected along with two additional genes (SOM text, part c, and tables S2 and S3). The results of the loss-of-phenotype assay are summarized in Fig. 3A. They show that deletion of the great majority of the overexpressed gene targets resulted in a loss of the capacity of the mutant *spt15-300* factor to impart an increased ethanol and glucose tolerance. All tested knockout strains not harboring the mutant *spt15-300* showed normal tolerance to ethanol and glucose stress, which indicated

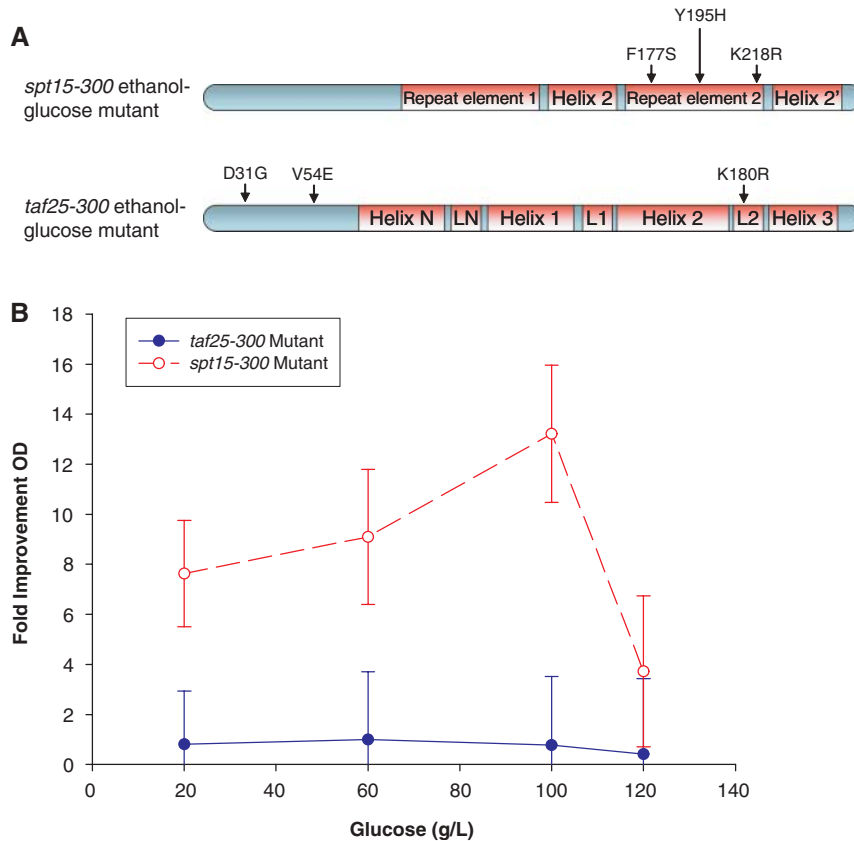


Fig. 1. Yeast gTME mutants with increased tolerance to elevated ethanol and glucose concentrations. **(A)** Mutations for the best clone isolated from either the *spt15* or *taf25* mutant library are shown mapped onto a schematic of critical functional components of the respective factor (SOM text, part a). **(B)** Growth yields of the clones from **(A)**, were assayed in synthetic minimal medium containing elevated levels (6% by volume) of ethanol and glucose after 20 hours. Under these conditions, the *spt15-300* mutant far exceeded the performance of the *taf25-300* mutant. Fold improvements of growth yields are compared with an isogenic strain that harbors a plasmid-borne, wild-type version of either *SPT15* or *TAF25*.

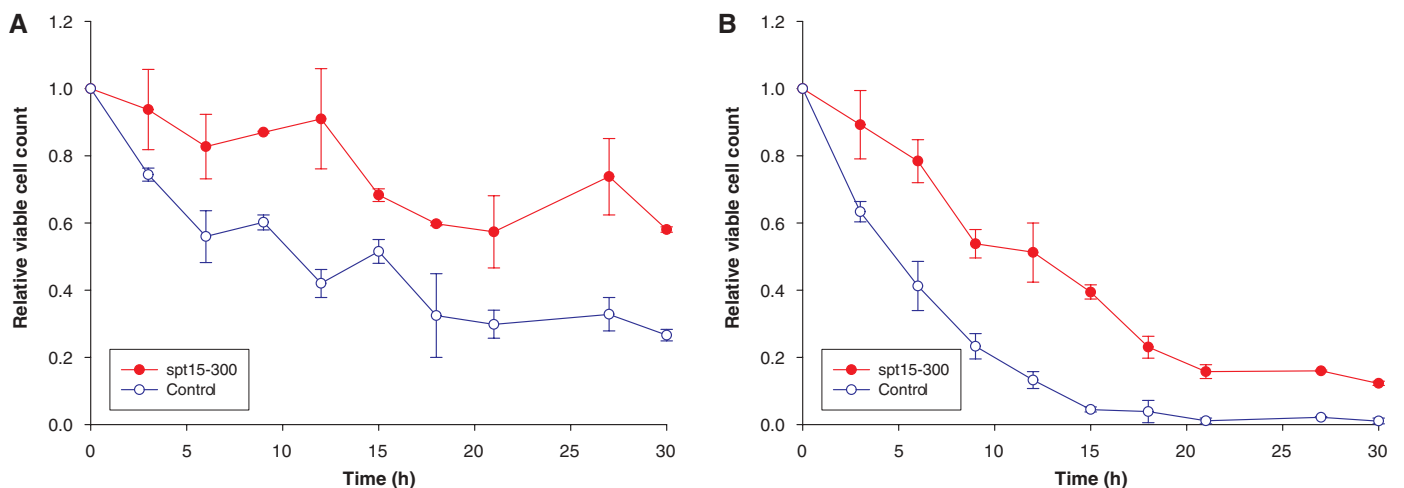


Fig. 2. Cellular viability curves to evaluate the tolerance of the mutant under ethanol stress. Viability of the *spt15-300* mutant strain compared with the control is measured as a function of time (hours) and expressed as the relative number of colony-forming units compared with colony count at 0 hours for stationary phase cells treated and incubated in standard

medium in the presence of **(A)** 12.5% and **(B)** 15% ethanol by volume. The *spt15-300* mutation confers a significantly enhanced viability at all concentrations tested above 10% ethanol by volume (fig. S2). Error bars represent the standard deviation between biological replicate experiments. Initial cell counts were $\sim 3.5 \times 10^6$ cells per ml.

that, individually, these genes are insufficient to constitute the normal tolerance to ethanol. Out of the 14 gene targets assayed, only loss of *PHM6* function did not reduce the novel phenotype. Thus, we hypothesize that each gene encodes a necessary component of an interconnected network, although there may be some redundancy of function (SOM text, part c).

Three genes that exhibited the greatest increase in expression level in the *spt15-300* mutant were investigated as overexpression targets in the control strain in a gain-of-function assay. *PHO5*, *PHM6*, and *FMP16* were independently and constitutively overexpressed under the control of the *TEF* promoter, and transformants were assayed for their capacity to

impart an ethanol- and glucose-tolerance phenotype. Overexpression of no single gene among the consensus, top-candidate genes from the microarray analysis produced a gain of phenotype similar to that of the mutant *spt15-300* (Fig. 3B).

We next constructed all possible single- and double-mutant combinations with the sites identified in the triple mutant (15). None of the single or double mutants came even close to achieving a phenotype similar to that of the isolated *spt15-300* triple mutant (SOM text, part d, and figs. S6 to S8). One could not predict the effect of these three mutations by a “greedy algorithm” search approach or select these by traditional selection for mu-

tations that cause incremental improvement, as many of these isolated mutations are independently relatively neutral in phenotype fitness. Consequently, such a multiple mutant is accessible only through a technique that specifically focuses on the in vitro mutagenesis of the *SPT15* gene followed by a demanding selection.

Genes previously documented as *SPT3*-dependent in expression (20, 21) were preferentially altered by our *spt15* mutant, as exhibited in the microarray data, with a Bonferroni-corrected *P* value of 1×10^{-12} . Furthermore, 7 of the 10 most highly expressed genes in the *spt15-300* mutant are *SPT3*-dependent genes. Genes that are down-regulated in *spt3* mutants were rela-

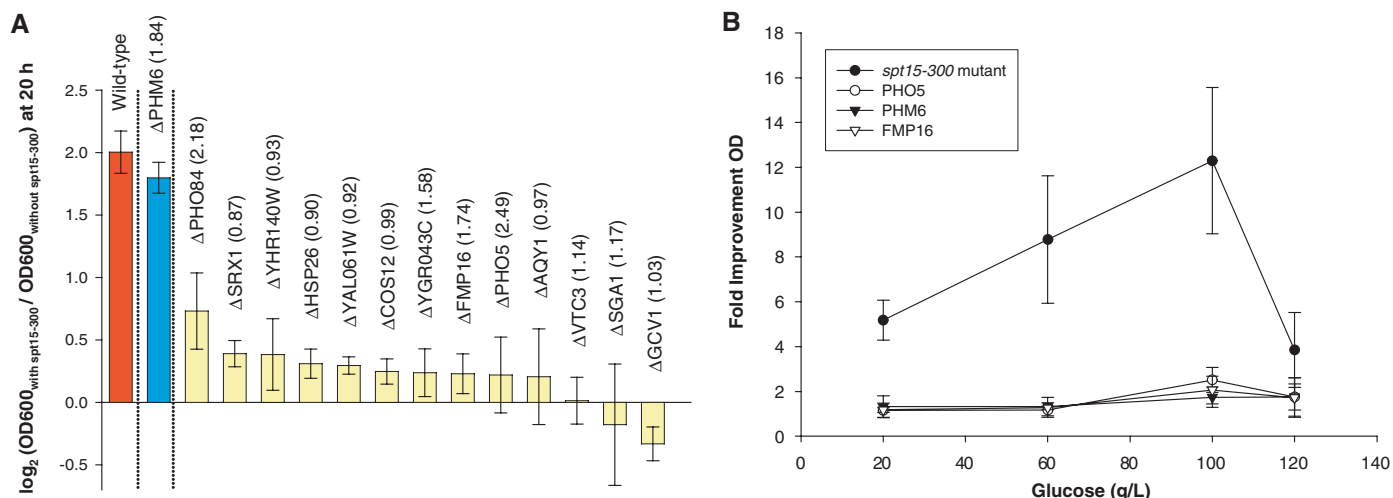
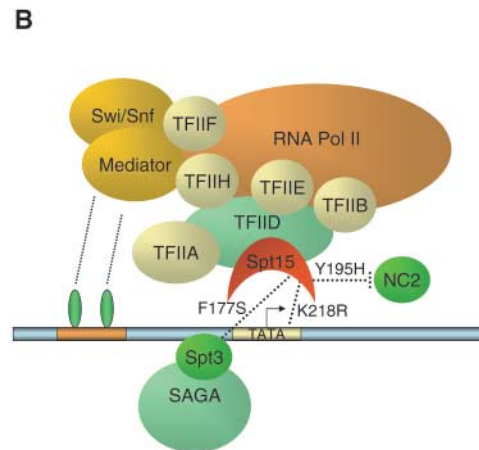
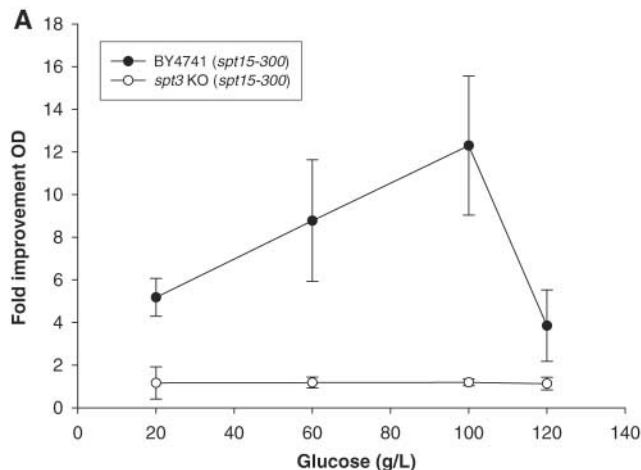


Fig. 3. Gene-knockout and overexpression analysis to probe the transcriptome-level response elicited by the mutant *spt15*. **(A)** Loss-of-phenotype analysis was performed using 12 of the most highly expressed genes in this mutant (\log_2 differential gene expression given in parentheses); two additional genes were chosen for further study (SOM text, part c). The tolerance (to 5% ethanol, 60 g/liter glucose) of 14 strains deleted in one of the 14 genes, respectively, was tested by comparing the knockout strain containing

the *spt15-300* mutation on a plasmid to a strain containing the wild-type *SPT15*. All gene knockouts, except *PHM6*, resulted in slight to full loss of phenotype. Control mutants for all of the gene knockout targets exhibited similar growth yields. **(B)** Gene overexpression studies are provided for the top three candidate genes from the microarray (*PHO5*, *PHM6*, and *FMP16*) and assayed under 6% ethanol by volume as previously assayed (see also fig. S5). The overexpression of these genes failed to impart a tolerance phenotype.

Fig. 4. Elucidation and validation of a mechanism partially mediated by the *SPT3*-SAGA complex. **(A)** The impact of an *spt3* knockout was evaluated through the introduction of the *spt15-300* mutant and assaying in the presence of 6% ethanol by volume. The incapacity of the mutant to impart the phenotype illustrates the essentiality of *SPT3* as a part of the mechanism provided. **(B)** The three mutations (F177S, Y195H, and K218R) are mapped on the global transcription machinery molecular mechanism proposed by prior studies, with each of these mutation sites (22–24, 27, 28). Collectively, these three mutations lead to a mechanism involving Spt3p.



tively up-regulated in the *spt15-300* mutant. The absence of negative cofactor 2 element (NC2) repression due to the Y195H mutation (22) may result in overrepresentation of up-regulated genes, because part of the negative regulation of the Spt15p can no longer take place. These data are consistent with previous work showing that the *spt15-21* mutation [a change from Ser to Leu or Arg at Phe¹⁷⁷ (F177L and F177R)] suppresses an *spt3* mutation as the result of an altered interaction between the Spt15p and Spt3p [part of the Spt-Ada-Gcn5-acetyltransferase (SAGA) complex] (21, 23, 24). As a further test of the link between Spt15p and Spt3p, it was found that an *spt15-300* mutant gene was unable to impart its ethanol- and glucose-tolerance phenotype to an *spt3* knockout strain (Fig. 4A).

From the results of the site-directed mutagenesis and mechanism depicted in Fig. 4B, it is conceivable that perturbations to the NC2 complex would also impact the ability of the *spt15-300* mutant to function; however, a null mutation in one of the genes in this heterodimer is inviable, which prevents such a follow-up experiment. Nevertheless, these results further underscore the importance of all three mutations acting in concert in order to create the complex phenotype mediated through an Spt3p-SAGA complex interaction. As a result, we posit that the mode of action is primarily a unique protein-protein-DNA interaction (Spt15-300p mutant-Spt3p-DNA), which leads to this transcriptional reprogramming of a large number of genes.

The capacity of the *spt15-300* mutant to use and ferment glucose to ethanol under a variety of conditions was assayed in simple batch shake-flask experiments of low and high cell density under an initial concentration of 20 or 100 g/liter of glucose (SOM text, part e, and

figs. S9 to S11). In each of these cases, the mutant has growth characteristics superior to those of the control with a prolonged exponential growth phase that allows for a higher, more robust biomass production and a higher ethanol yield. Specifically, in high-cell density fermentations, with an initial optical density at 600 nm (OD₆₀₀) of 15, the mutant's performance far exceeds that of the control, with more rapid utilization of glucose, improved biomass yield, and higher volumetric ethanol productivity (2 g/liter of ethanol per hour) relative to the control strain (Table 1). In addition, sugars were rapidly and fully used at a yield that exceeds that of the control and approaches the theoretical value when taking into account the amount of glucose consumed for cell growth.

These results demonstrate the applicability of gTME to alter cellular eukaryotic phenotypes. The isolation of dominant mutations permits the modification of vital functions for novel tasks, whereas the unmodified allele carries out the functions critical for viability. An examination of further modifications of other transcription factors through gTME could additionally have the potential for drastically improving ethanol fermentations and for improving the prospects of ethanol production. For the mutants analyzed, altered fermentation conditions and additional pathway engineering are likely to further increase ethanol production (25, 26). Furthermore, the strain used in this study is a standard laboratory yeast strain, and this method could be explored in industrial or isolated yeast exhibiting naturally higher starting ethanol tolerances. Finally, we note that the transcription factors modified in this study have similarity to those in more complex eukaryotic systems including those of mammalian cells, which raises the possibility of using this tool to elicit complex phenotypes of both

biotechnological and medical interest in these systems as well.

References and Notes

- H. Alper, thesis, Massachusetts Institute of Technology (2006).
- S. Hahn, *Nat. Struct. Mol. Biol.* **11**, 394 (2004).
- M. Hampsey, *Microbiol. Mol. Biol. Rev.* **62**, 465 (1998).
- J. L. Kim, D. B. Nikolov, S. K. Burley, *Nature* **365**, 520 (1993).
- D. I. Chasman, K. M. Flaherty, P. A. Sharp, R. D. Kornberg, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8174 (1993).
- M. C. Schultz, R. H. Reeder, S. Hahn, *Cell* **69**, 697 (1992).
- R. Thatipamala, S. Rohani, G. Hill, *Biotechnol. Bioeng.* **40**, 289 (1992).
- F. W. Bai, L. J. Chen, Z. Zhang, W. A. Anderson, M. Moo-Young, *J. Biotechnol.* **110**, 287 (2004).
- F. van Voorst, J. Houghton-Larsen, L. Jonson, M. C. Kielland-Brandt, A. Brandt, *Yeast* **23**, 351 (2006).
- Z. P. Cakar, U. O. Seker, C. Tamerler, M. Sonderegger, U. Sauer, *FEMS Yeast Res* **5**, 569 (2005).
- K. Furukawa, H. Kitano, H. Mizoguchi, S. Hara, *J. Biosci. Bioeng.* **98**, 107 (2004).
- M. Nozawa, T. Takahashi, S. Hara, H. Mizoguchi, *J. Biosci. Bioeng.* **93**, 288 (2002).
- Y. Ogawa *et al.*, *J. Biosci. Bioeng.* **90**, 313 (2000).
- H. Takagi, M. Takaoka, A. Kawaguchi, Y. Kubo, *Appl. Environ. Microbiol.* **71**, 8656 (2005).
- Materials and methods are available as supporting material on Science Online.
- J. H. Geiger, S. Hahn, S. Lee, P. B. Sigler, *Science* **272**, 830 (1996).
- J. D. Storey, R. Tibshirani, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9440 (2003).
- For each gene, a *P* value for differential expression between the two conditions was calculated by a *t* test. To simultaneously test multiple hypotheses, *P* values were corrected in a false-discovery rate analysis (17). False-discovery rates are commonly used for the analysis of large data sets (such as microarrays), which limits false-positives, akin to a Bonferroni correction. In this case, 366 genes were found to be significantly differentially expressed, at a false-discovery rate of 1%.
- P. Shannon *et al.*, *Genome Res.* **13**, 2498 (2003).
- K. L. Huisinga, B. F. Pugh, *Mol. Cell* **13**, 573 (2004).
- T. I. Lee *et al.*, *Nature* **405**, 701 (2000).
- Y. Cang, D. T. Auble, G. Prelich, *EMBO J.* **18**, 6662 (1999).
- H. Kou, J. D. Irvin, K. L. Huisinga, M. Mitra, B. F. Pugh, *Mol. Cell. Biol.* **23**, 3186 (2003).
- D. M. Eisenmann, K. M. Arndt, S. L. Ricupero, J. W. Rooney, F. Winston, *Genes Dev.* **6**, 1319 (1992).
- T. L. Nissen, M. C. Kielland-Brandt, J. Nielsen, J. Villadsen, *Metab. Eng.* **2**, 69 (2000).
- P. Slininger, B. Dien, S. Gorsich, Z. Liu, *Appl. Microbiol. Biotechnol.* [10.1007/s00253-006-0435-1](https://doi.org/10.1007/s00253-006-0435-1) (2005).
- M. P. Klejman, X. Zhao, F. M. van Schaik, W. Herr, H. T. Timmers, *Nucleic Acids Res.* **33**, 5426 (2005).
- D. K. Lee, J. DeJong, S. Hashimoto, M. Horikoshi, R. G. Roeder, *Mol. Cell. Biol.* **12**, 5189 (1992).
- G. M. O'Connor, F. Sanchez-Riera, C. L. Cooney, *Biotechnol. Bioeng.* **39**, 293 (1992).
- We acknowledge financial support of this work by the DuPont-MIT Alliance, the Singapore-MIT Alliance (SMA1), the NIH grant GM035010, and Department of Energy grant DE-FG02-94ER14487. We thank F. Winston for a thoughtful discussion regarding experimental design. Microarray data deposited to the Gene Expression Omnibus database under the accession number GSE5185.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5805/1565/DC1
Materials and Methods
SOM Text
Figs. S1 to S11
Tables S1 to S6
References

30 June 2006; accepted 1 November 2006
10.1126/science.1131969

Table 1. Fermentation results to evaluate the ethanol production potential of the *spt15* mutant. Cells were cultured in biological replicate in 100 g/liter of glucose with a high inoculum of initial cell optical density of (OD₆₀₀) of 15 [~4 g DCW(dry cell weight)/liter]. Fermentation profiles for the high-cell density fermentation are provided and illustrate the capacity of this mutant to produce higher productivities of ethanol at the theoretical yield, surpassing the function of the control. Biomass yield from glucose is from reported values (29). Results represent the average between biological replicate experiments (SOM text, part e, and figs. S9 to S11). EtOH, ethanol.

Parameter	<i>spt15-300</i> mutant	Control	Percent improvement
Initial DCW (g/liter)	4.06	4.10	—
Final DCW(g/liter)	6.46	5.39	+20%
Volumetric productivity (g/liter h ⁻¹)	2.03	1.20	+69%
Specific productivity (g/DCW h ⁻¹)	0.31	0.22	+41%
Conversion yield calculated between 6 and 21 hours	0.36	0.32	+14%
True EtOH yield accounting for biomass production (Percentage of 0.41 g/g, which represents the theoretical maximum)	0.40 (98%)	0.35 (86%)	+15%
EtOH produced (g/liter)			
$\text{glucose used (g/liter)} - \left(\frac{1 \text{ g glucose}}{0.5 \text{ g DCW}}\right) \text{DCW produced (g/liter)}$			