



## Unstable Tandem Repeats in Promoters Confer Transcriptional Evolvability

Marcelo D Vinces, Matthieu Legendre, Marina Caldara, Masaki Hagihara,  
Kevin J Verstrepen

### ► To cite this version:

Marcelo D Vinces, Matthieu Legendre, Marina Caldara, Masaki Hagihara, Kevin J Verstrepen. Unstable Tandem Repeats in Promoters Confer Transcriptional Evolvability. *Science*, 2009, 324 (5931), pp.1213-1216. 10.1126/science.1170097 . hal-02329186

**HAL Id: hal-02329186**

**<https://hal.science/hal-02329186>**

Submitted on 28 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Published in final edited form as:

*Science*. 2009 May 29; 324(5931): 1213–1216. doi:10.1126/science.1170097.

## Unstable Tandem Repeats in Promoters Confer Transcriptional Evolvability

Marcelo D. Vences<sup>1,2,3,\*</sup>, Matthieu Legendre<sup>1,4,\*</sup>, Marina Caldara<sup>1</sup>, Masaki Hagihara<sup>5</sup>, and Kevin J. Verstrepen<sup>1,2,3,†</sup>

<sup>1</sup>FAS Center for Systems Biology, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA.

<sup>2</sup>Laboratory for Systems Biology, Flanders Institute for Biotechnology (VIB), Katholieke Universiteit Leuven (K.U. Leuven), B-3001 Heverlee, Belgium.

<sup>3</sup>Genetics and Genomics Group, Centre of Microbial and Plant Genetics (CMPG), K.U. Leuven, Gaston Geenslaan 1, B-3001 Leuven (Heverlee), Belgium.

<sup>4</sup>Structural and Genomic Information Laboratory, CNRS-UPR 2589, IFR-88, Université de la Méditerranée Parc Scientifique de Luminy, Avenue de Luminy, FR-13288 Marseille, France.

<sup>5</sup>The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, 567-0047, Japan.

### Abstract

Relative to most regions of the genome, tandemly repeated DNA sequences display a greater propensity to mutate. A search for tandem repeats in the *Saccharomyces cerevisiae* genome revealed that the nucleosome-free region directly upstream of genes (the promoter region) is enriched in repeats. As many as 25% of all gene promoters contain tandem repeat sequences. Genes driven by these repeat-containing promoters show significantly higher rates of transcriptional divergence. Variations in repeat length result in changes in expression and local nucleosome positioning. Tandem repeats are variable elements in promoters that may facilitate evolutionary tuning of gene expression by affecting local chromatin structure.

The genomes of most organisms are not uniformly prone to change because they contain hotspots for mutating events. An abundant class of sequences that mutate at higher frequencies than the surrounding genome is composed of tandem repeats (TRs, also known as satellite DNA), DNA sequences repeated adjacent to one another in a head-to-tail manner (1). Errors during replication make TRs unstable, generating changes in the number of repeat units that are 100 to 10,000 times more frequent than point mutations (2). Variable TRs are often dismissed as nonfunctional “junk” DNA. However, some TRs located within coding regions (exons) have demonstrable functional roles. For example, TR copy numbers in genes such as *FLO1* in *Saccharomyces cerevisiae* generate plasticity in adherence to

Copyright 2009 by the American Association for the Advancement of Science; all rights reserved.

<sup>†</sup>To whom correspondence should be addressed. Kevin.Verstrepen@biw.vib-kuleuven.be.

\*These authors contributed equally to this work.

### Supporting Online Material

[www.sciencemag.org/cgi/content/full/324/5931/1213/DC1](http://www.sciencemag.org/cgi/content/full/324/5931/1213/DC1)

Materials and Methods

Figs. S1 to S14

Tables S1 to S6

References

Data Sets S1 to S3

substrates (3). In canines, variable repeats located in *Alx-4* and *Runx-2* confer variability to skeletal morphology, which may have facilitated the diversification of domestic dogs bred by humans (4). Thus, repeats located in coding regions may increase the evolvability of proteins.

There is also evidence that repeats influence expression of certain genes (5–7). To investigate the involvement of TRs in gene expression variation, we first mapped and classified all repeats in the S288C yeast genome (8) (data set S1). TRs are enriched in yeast promoters (table S1). Of the ~5700 promoters in the genome, 25% (1455) contain at least one TR. Many TRs in promoters consist of short, A/T-rich sequences (table S2, fig. S1, and data set S2). Comparison of orthologous regions in genomes of different *S. cerevisiae* strains showed that many of the TRs are variable (data set S1). For example, 24.1% of orthologous TR loci in promoters differ in the number of repeat units between the two fully sequenced strains, S288C and RM11 (8). To confirm this, we sequenced 33 randomly chosen promoter repeats in seven *S. cerevisiae* genomes (Fig. 1A, figs. S2 and S3, and data set S3). Twenty-five of the 33 TRs differed in repeat units in at least one of the seven strains. The repeat variation frequency is 40-fold higher than the frequency of insertions and deletions (indels) and of point mutations in the surrounding nonrepetitive sequence ( $P < 10^{-15}$ ) (figs. S2 and S3).

To determine whether promoter TR variation affects gene expression, we compared repeat variability to expression divergence (ED), which represents how fast the transcriptional activity of each gene evolves (9–11). Promoters containing TRs showed significantly ( $P < 1.75 \times 10^{-4}$ ) higher amounts of ED than did promoters lacking TRs when comparing yeast species (*S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. kudriavzevii*) (Fig. 1, B to D, and fig. S4A) and *S. cerevisiae* strains (S288C and RM11) (Fig. 1, E to G, and fig. S4, B and C). This difference was independent of factors known to affect transcriptional divergence, for example, the presence of TATA boxes (fig. S5). Only promoters containing variable numbers of repeat units between strains or species showed the elevated ED (Fig. 1, D and G). Furthermore, when variable TRs were binned into variable and highly variable (10% most variable) groups, highly variable repeats displayed even higher ED. Hence, ED correlates not merely with TRs in promoters but more specifically with repeat number variation.

To directly test whether changes in promoter TRs affect transcriptional activity, we varied the TR repeat number in the promoters of yeast genes *YHB1*, *MET3*, and *SDT1* (Fig. 2 and fig. S6A). For each construct, expression increased as the length of the TR increased from zero, until a certain size was reached, after which expression dropped off. To determine whether natural variation between strains corresponded to similar changes in gene expression, we cloned promoters of several strains into the respective promoters of strain S288C. These transformants indeed mirrored the expression-level patterns that we observed with the engineered TR strains (fig. S6, B to D). Moreover, TR-mediated changes in expression have functional consequences. *SDT1* encodes a pyrimidine nucleotidase that confers resistance to the nucleotide analog 6-azauracil (6AU) (12), and we observed that strains with various *SDT1* promoter TR constructs show differences in growth that match the *SDT1* expression changes (fig. S7).

Because promoter TR length variation affects transcription, we speculated that it should be possible to exploit promoter TR instability to quickly select for changes in gene expression. We used strains with a relatively long *SDT1* TR tract, because repeat mutation rates increase with increasing tract length (13), and selected for variants having higher gene expression. The *SDT1* open reading frame was replaced with selectable markers, either *URA3* or yellow fluorescent protein (YFP), for selection in SC-ura (medium requiring yeast to express *URA3*

to grow in) or selection by sorting with flow cytometry, respectively. After a few rounds of selection, both regimes yielded mutants that showed increased growth on medium lacking uracil or increased YFP fluorescence. These mutants showed significant ( $P < 1 \times 10^{-15}$ ) changes in the length of the TR tract in the *SDT1* promoter. The most common tract lengths yielded by both selections was 13 repeats (26 nucleotides), corresponding to a size close to that of the engineered TR constructs with the highest expression (Fig. 2B). Construction of a 13-unit TR revealed that it had the highest expression of all the engineered strains (Fig. 3A). When strains containing the long TR tract were grown without any selective pressure, most strains remained at the initial 48 units, and the few mutants that arose show a broader TR size distribution. We were unable to obtain strains with higher expression or TR size changes when the repeat sequence in the promoter was replaced by a randomized (no repeat) sequence of equal length as the 48-unit repeat tract (Fig. 3B).

TRs can contain transcription factor (TF) binding sites, so variation in the tract length may result in the removal or addition of binding sites. Of the 1455 TR-containing promoters, 113 contain known TF binding sites located within the repeats. Many of these TRs overlap stress-response TF binding sites (table S3). Investigation of one of these TRs, located in the promoter of *YKL107w*, indicated that, in this case, changes in the TR number may affect transcription through variation in the number of binding sites of the oxidative-stress-responsive TF, Yap1 (fig. S8).

Most promoter TRs, however, do not overlap known TF sites, indicating that another mechanism underlies repeat-related expression differences. To investigate whether distance between promoter elements affects transcription, we replaced the *SDT1* TR with different DNA sequences of the same length. These changes mostly led to severely reduced transcription, indicating that altered spacing alone is not sufficient to explain the effect of repeat variation on transcription (Fig. 4, A to C). Many promoter TRs are extremely A/T-rich, suggesting they may facilitate DNA melting. However, some of the constructs with the same high A/T content as the original repeat tract still show reduced transcription.

Most promoter TRs are located ~200 base pairs upstream from the translational start codon (Fig. 4D and fig. S9), corresponding to the nucleosome-free region of yeast promoters (14), suggesting a link between chromatin structure and TRs. We compared available genome-wide nucleosome positioning data with the positioning of TRs. Nucleosome density across promoter regions showed an inverse correlation with the presence of TRs (Fig. 4D and fig. S9A). Nucleosome depletion is especially pronounced around AT-rich repeats, which compose 80% of all repeats (fig. S9B). These results suggest either that TRs preferentially form in nucleosome-free regions or that nucleosomes cannot easily bind TRs. We also found that nucleosomes containing the histone variant H2AZ (15) tended to border the TRs (Fig. 4D and fig. S9). Although no simple rule governs nucleosome positioning, nucleosome positioning in the yeast genome is largely directed by DNA sequence (15–18), and a computational algorithm exists that uses DNA sequence information to predict where nucleosomes bind (16, 19). This algorithm predicts that TR regions in promoters are nucleosome-free (fig. S10), suggesting that promoter TR sequences intrinsically bind poorly to nucleosomes, presumably because the repeats affect biophysical properties of DNA (e.g., bendability) (20). In line with these predictions, an analysis of a transcriptome data set for a large group of chromatin remodeling mutants (21) showed that the expression of repeat-containing promoters is strongly influenced by regulators of chromatin structure and activity (fig. S11).

If repeats control nucleosome positioning, changing the repeat sequence should influence the local nucleosome structure. Deletion of the TR of *SDT1* resulted in binding of nucleosomes to this region and also disturbed the positioning of downstream nucleosomes

(Fig. 4E). Intermediate deletions of the TRs resulted in gradual changes in nucleosome occupancy (fig. S12), indicating that repeat variation directly affects the local chromatin structure. Replacing the repeat with a sequence predicted to have a high affinity for nucleosomes (19) resulted in a well-positioned nucleosome in the previously nucleosome-free region and in greatly reduced *SDTI* expression (Fig. 4, F and G).

Although several molecular mechanisms may underlie the effect promoter TRs have on gene expression, our data indicate that repeat-dependent changes in DNA sequence and chromatin structure play a role. Local chromatin structure is known to affect transcriptional activity (22). Most repeats in promoters do not contain the hallmark dinucleotide periodicities that are associated with nucleosome-binding DNA (19). As a result, these repeat tracts may help to establish variable nucleosome-free DNA structures and influence nucleosome positioning in nearby regions. Moreover, because of the high A/T content of most promoter TRs, it remains to be tested whether these nucleosome-free TRs may allow DNA melting for loading of the RNA polymerase.

Changes within coding regions and in protein sequence underlie much biological adaptation and innovation. However, changes in the regulation of genes may be equally important (23, 24). Our results presented here are consistent with a role for TRs as ubiquitous and adjustable “evolutionary tuning knobs” (25) for transcription that mediate rapid evolution of gene expression. Genes that respond to changing environmental conditions would be particularly suited for such variable genetic elements. Indeed, genes driven by repeat-containing promoters show elevated responsiveness to changing environmental conditions (fig. S13).

A preliminary analysis of *Homo sapiens* promoters reveals a TR distribution comparable to that of yeast, suggesting that similar mechanisms are also at play in higher organisms (fig. S14).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

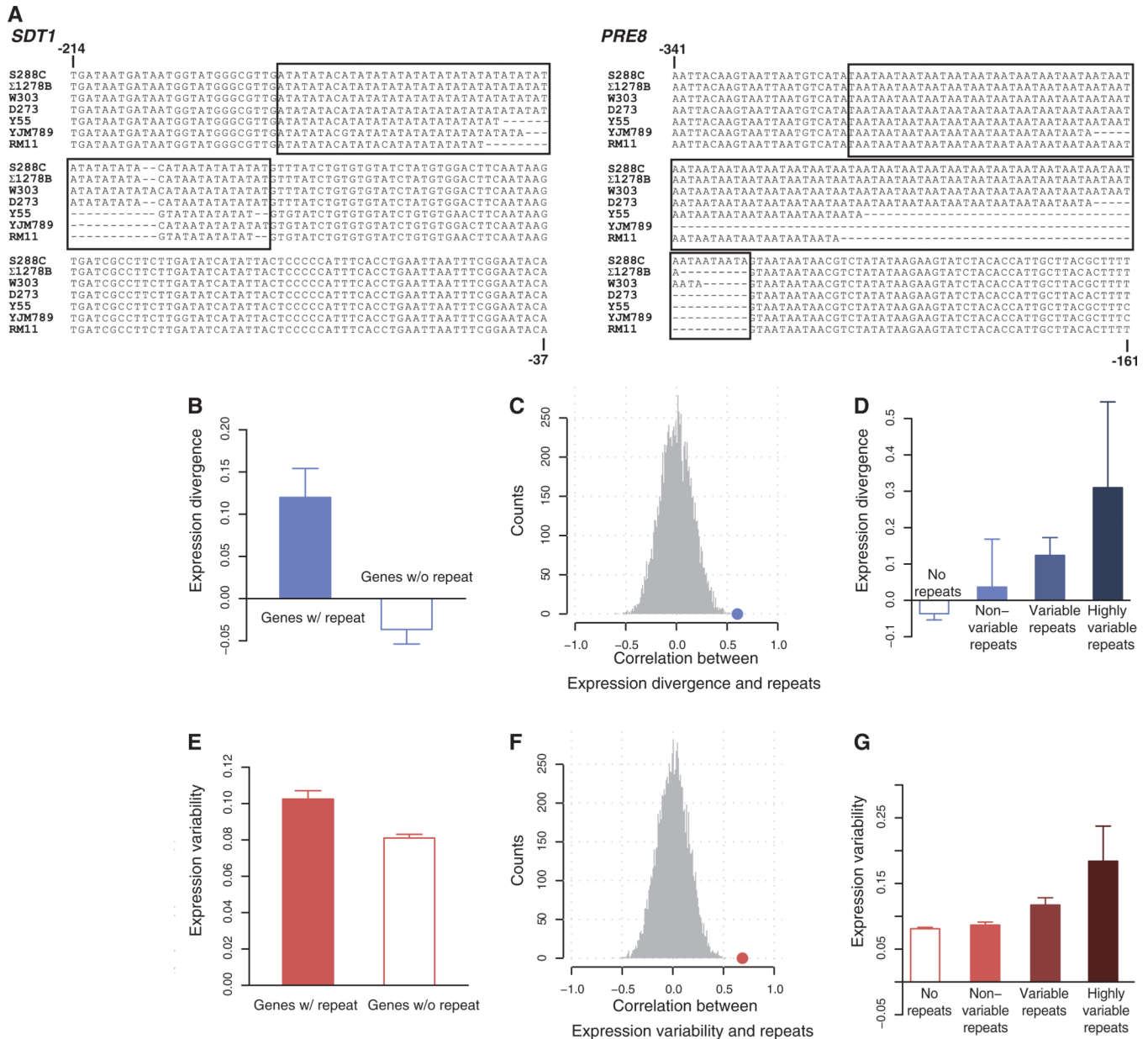
We thank B. Calderon, B. (Sze Ham) Chan, and B. Breaux for their contributions; B. Stern, E. O’Shea, A. Murray, A. Rowat, and A. New for critique of the manuscript; and the anonymous reviewers for their useful suggestions. Research in the lab of K.J.V. is supported by NIH National Institute for General Medical Studies grant P50GM068763, Human Frontier Science Program (HFSP) award RGY79/2007, the Flemish Institute for Biotechnology (VIB), and the Fonds Wetenschappelijk Onderzoek Vlaanderen–Odysseus program. M.D.V. acknowledges the Ford Foundation and the Belgian American Education Foundation.

## References and Notes

1. Rando OJ, Verstrepen KJ. Cell. 2007; 128:655. [PubMed: 17320504]
2. Lynch M, et al. Proc. Natl. Acad. Sci. U.S.A. 2008; 105:9272. [PubMed: 18583475]
3. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Nat. Genet. 2005; 37:986. [PubMed: 16086015]
4. Fondon JW 3rd, HR Garner. Proc. Natl. Acad. Sci. U.S.A. 2004; 101:18058. [PubMed: 15596718]
5. Streelman JT, Kocher TD. Physiol. Genomics. 2002; 9:1. [PubMed: 11948285]
6. Martin P, Makepeace K, Hill SA, Hood DW, Moxon ER. Proc. Natl. Acad. Sci. U.S.A. 2005; 102:3800. [PubMed: 15728391]
7. Rockman MV, Wray GA. Mol. Biol. Evol. 2002; 19:1991. [PubMed: 12411608]
8. Material and methods are available as supporting material on *Science* Online.
9. Choi JK, Kim YJ. Nat. Genet. 2008; 40:141. [PubMed: 18227874]

10. Ronald J, Brem RB, Whittle J, Kruglyak L. PLoS Genet. 2005; 1:e25. [PubMed: 16121257]
11. Tirosh I, Weinberger A, Carmi M, Barkai N. Nat. Genet. 2006; 38:830. [PubMed: 16783381]
12. Nakanishi T, Sekimizu K. J. Biol. Chem. 2002; 277:22103. [PubMed: 11934891]
13. Legendre M, Pochet N, Pak T, Verstrepen KJ. Genome Res. 2007; 17:1787. [PubMed: 17978285]
14. Yuan GC, et al. Science. 2005; 309:626. published online 16 June 2005 (10.1126/science.1112178). [PubMed: 15961632]
15. Albert I, et al. Nature. 2007; 446:572. [PubMed: 17392789]
16. Field Y, et al. PLOS Comput. Biol. 2008; 4:e1000216. [PubMed: 18989395]
17. Ioshikhes IP, Albert I, Zanton SJ, Pugh BF. Nat. Genet. 2006; 38:1210. [PubMed: 16964265]
18. Lee W, et al. Nat. Genet. 2007; 39:1235. [PubMed: 17873876]
19. Segal E, et al. Nature. 2006; 442:772. [PubMed: 16862119]
20. Drew HR, Travers AA. J. Mol. Biol. 1985; 186:773. [PubMed: 3912515]
21. Steinfeld I, Shamir R, Kupiec M. Nat. Genet. 2007; 39:303. [PubMed: 17325681]
22. Lam FH, Steger DJ, O'Shea EK. Nature. 2008; 453:246. [PubMed: 18418379]
23. Pennisi E. Science. 2008; 321:760. [PubMed: 18687933]
24. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Science. 2007; 317:118. [PubMed: 17525304]
25. King D, Soller M, Kashi Y. Endeavour. 1997; 21:36.
26. Sekinger EA, Moqtaderi Z, Struhl K. Mol. Cell. 2005; 18:735. [PubMed: 15949447]

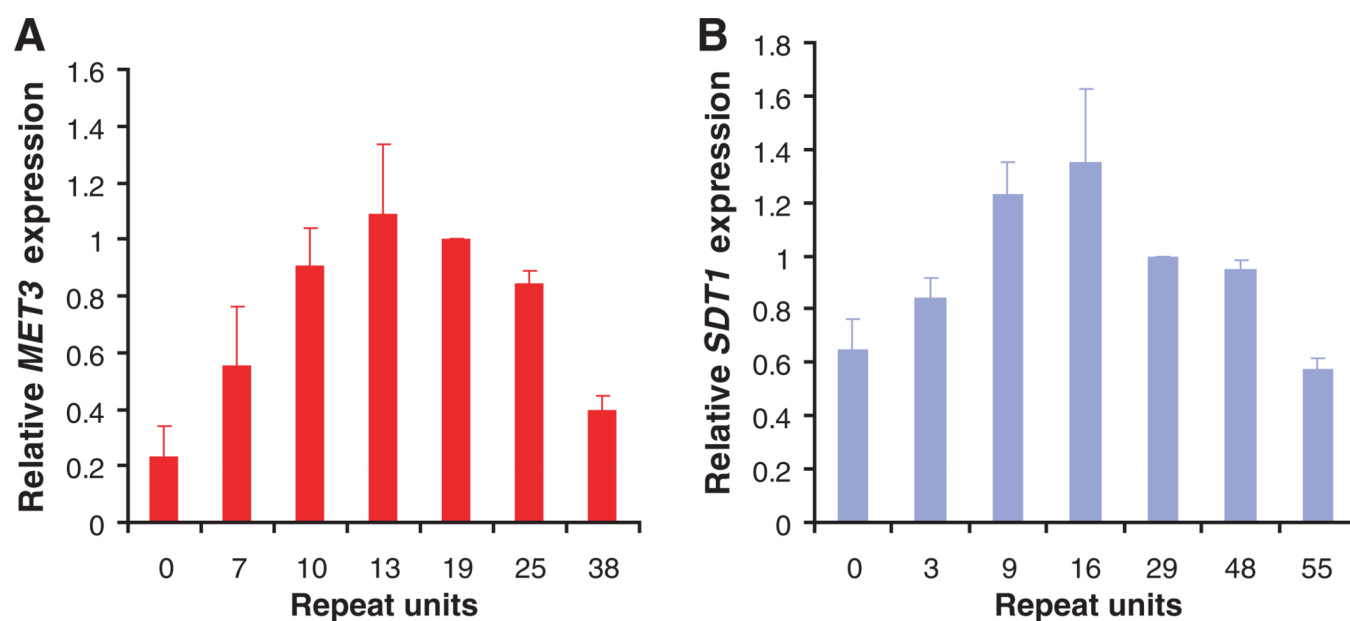




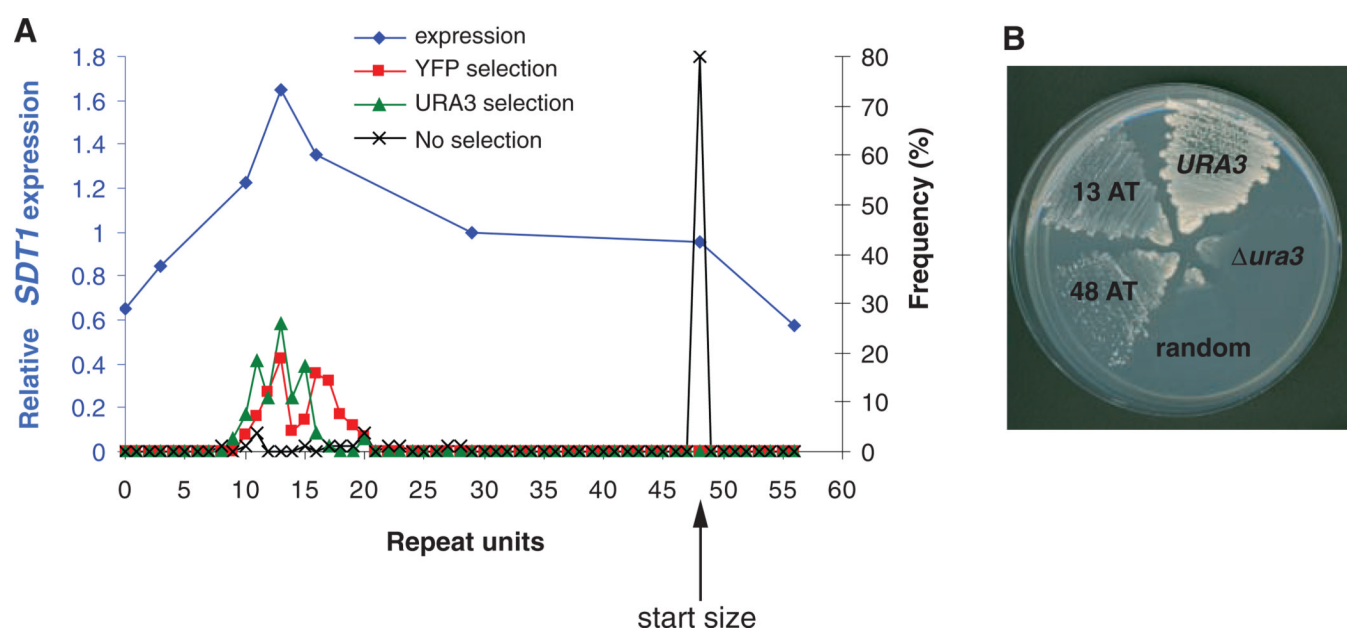
**Fig. 1.** Promoters containing variable TR tracts show elevated expression divergence. **(A)** Nucleotide alignments of two gene promoters from seven yeast strains (indicated on left side and in table S4) showing variable TRs (boxed). More examples in fig. S2 and data sets S1 and S3. **(B)** ED is associated with presence of TRs. Promoters were divided into those containing TRs and those lacking TRs. The average ED calculated from expression data from four *Saccharomyces sensu strictu* yeast species (*S. cerevisiae*, *S. paradoxus*, *S. mikatae*, and *S. kudriavzevii*) (11) is greater for promoters that contain TRs than for those that do not ( $P < 1.75 \times 10^{-4}$ ; error bars represent SE). **(C)** Promoters analyzed in (B) were sorted by their EDs and split in bins of 100 promoters (fig. S4A). The number of repeat-containing promoters in each bin and ED were significantly correlated. Shown is the correlation coefficient of the data set (blue dot; Z score = 2.70, empirical  $P$  value =  $6.75 \times 10^{-3}$ ) and of randomly shuffled data (gray histogram). **(D)** Only repeats that vary between

the four yeast species show elevated ED. ED was calculated for promoters divided into four categories as shown (error bars represent SE). **(E to G)** Similar analyses as in (B) to (D), except that ED values were calculated by using data from two *S. cerevisiae* strains (S288c and RM11) (9). The results show a similar correlation of variable repeats and ED (all *P* values are  $< 10^{-3}$ ). The data set for (F) is in fig. S4B.

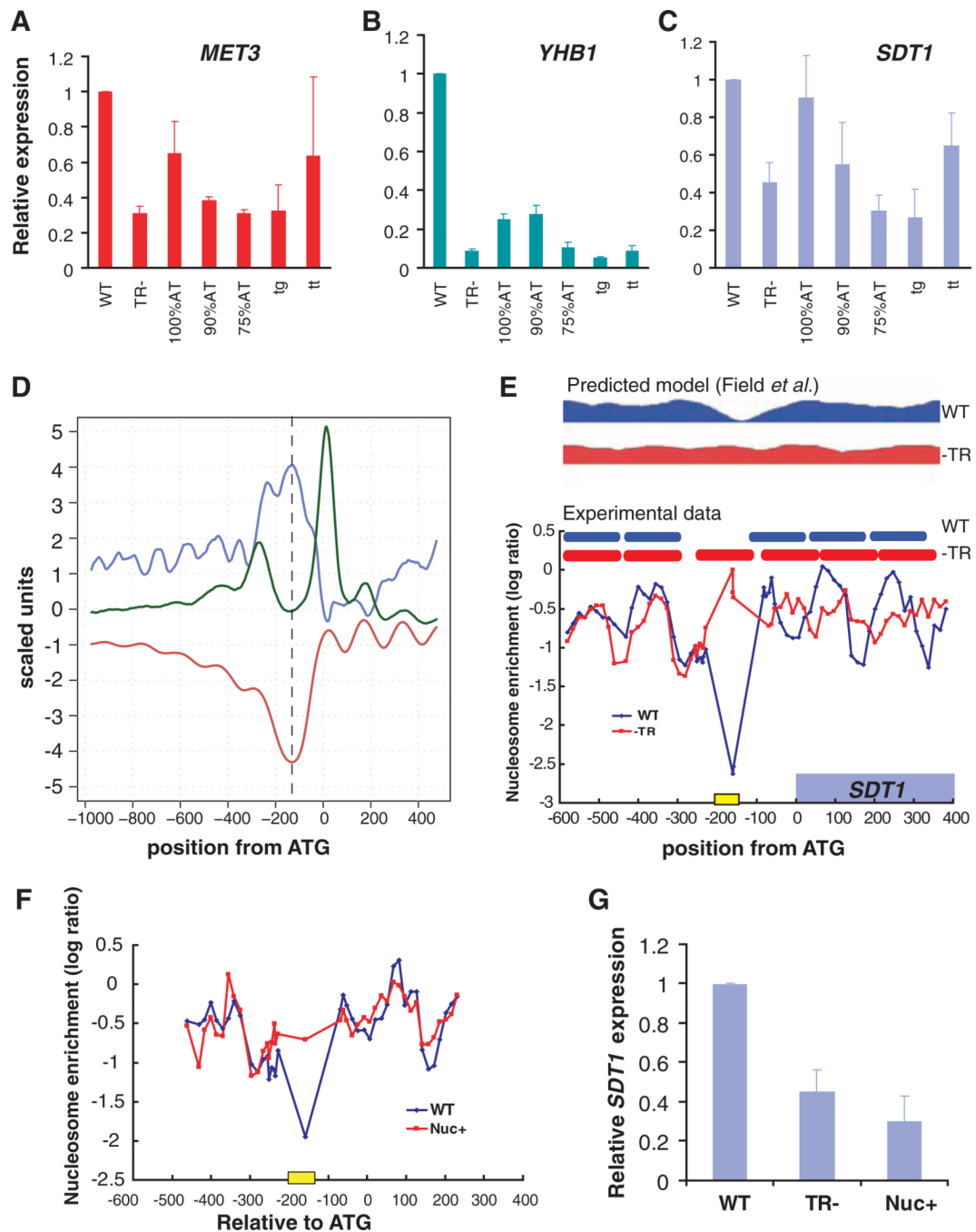




**Fig. 2.** Mutation of TRs in promoters alters gene expression. Sets of strains with varying numbers of repeat units were constructed for two yeast gene promoters: (A) *MET3* and (B) *SDT1*. Expression was determined by quantitative polymerase chain reaction (PCR). All values are normalized to transcript levels of housekeeping gene *ACT1* and to wild-type (WT) expression levels; error bars indicate SE,  $n = 3$ . Data for *YHB1* are in fig. S6A.

**Fig. 3.**

Experimental evolution of gene expression mediated by TRs. **(A)** Selection for higher expression of the *SDT1* promoter results in changes in the number of TR units in the promoter. The starting strain (48 repeat units, indicated by arrow) has relatively low expression (blue line, left axis). The green (*URA3*) and red (YFP) lines represent the final size distributions of TRs after selection for higher expression by using *URA3* or YFP reporters. TR size distribution without selection (black line) remains mostly at the initial 48 units. See (8) for details. **(B)** Growth on SC-ura medium of strains with native *URA3* (positive control); with *ura3* deleted (negative control); and with *URA3* fused to the *SDT1* promoter containing a long nonrepetitive (scrambled, random) sequence, a 48-AT repeat, or a 13-AT repeat.

**Fig. 4.**

TRs act as nucleosome positioning elements in promoters. Replacement of TRs with various sequences of the same size does not restore normal gene expression of (A) *MET3*, (B) *YHB1*, or (C) *SDT1*. TRs were replaced with other sequences of the same length: random (nonrepetitive) sequences (% A/T content indicated) and dinucleotide repeats (TG and TT repeats). The TR-strain lacks the entire TR. Error bars represent standard errors. (D) Genome-wide nucleosome position is largely anticorrelated with positions of TRs. Positions of the TR tracts are marked in blue; nucleosome positions, red (18); H2AZ nucleosome positions, green (15); and location of greatest TR enrichment, dashed vertical line. All repeat-containing promoters were aligned relative to the translational start site (ATG) or to

the transcriptional start site (fig. S9A). **(E)** Deletion of TRs in the *SDT1* promoter interferes with maintenance of the nucleosome-free region. *SDT1* coordinates on the  $x$  axis are relative to the ATG. The  $y$  axis represents the signal from a PCR-based nucleosome positioning assay (26). The yellow box represents TR location. Inferred nucleosome positions based on our experimental data in the graph are depicted above (WT, blue ovals; -TR mutant, red ovals). Above these are the predicted nucleosome profiles (16). **(F)** Nucleosome positioning when *SDT1* promoter TRs are replaced with a sequence predicted to position nucleosomes (Nuc+). **(G)** *SDT1* expression in the Nuc+ strain compared with those of WT and TR- strains.