Gene clustering pattern, promoter architecture, and gene expression stability in eukaryotic genomes

Yong H. Woo[a] and Wen-Hsiung Li[a,b,1]

[a]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637; and [b]Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan

Contributed by Wen-Hsiung Li, January 5, 2011 (sent for review November 5, 2010)

A balance between gene expression stability and evolvability is essential for the long-term maintenance of a living system. In this paper, we studied whether the genetic and epigenetic properties of the promoter affect gene expression variability. We hypothesized that upstream distance and orientation (head-to-head or head-to-tail) are important for the promoter architecture and gene expression variability. We found that in budding yeast genes with a short upstream distance tend to have low gene expression variability, and their promoter is flanked by strongly positioned nucleosomes and tends to have low nucleosome occupancy. These observations suggest that in vivo positioning of the flanking nucleosomes facilitates stable nucleosome depletion at the core promoter region and enhances gene expression stability. Head-to-head genes have, on average, lower gene expression variability, greater nucleosome depletion at the core promoter region, and more strongly positioned nucleosomes that flank the core promoter than do head-to-tail genes. These observations hold for diverse eukaryotes. In complex organisms such as mammals, only a small fraction of head-to-tail genes have retained a short upstream distance, probably because the promoter may not be flanked by a strongly positioned nucleosome on the upstream side.

bi-directional promoters | genome organization | nucleosome positioning

How genome organization relates to function and evolution of living systems is an ongoing area of research (1). Genes are organized on the genome by a combination of (*i*) gene order, (*ii*) intergenic distance, and (*iii*) adjacent gene orientation [tandem (++ or −−), convergent (+−), or divergent (−+) transcription configurations]. The importance of the three components is best illustrated by head-to-head clustering: two adjacent genes separated by a short intergenic distance and oriented in divergent (−+) transcription configuration. Head-to-head clusters are prevalent and conserved in many eukaryotes, including yeasts, plants, invertebrates, and vertebrates (2–5). Protein-coding genes also can form head-to-head clusters with noncoding genes (6). The most likely proposed function of such clustering is to coregulate two adjacent genes by a single bidirectional promoter. However, coexpression of two adjacent head-to-head genes does not occur much more often than would be expected by chance (7, 8). Thus, head-to-head clustering might have other regulatory roles.

A balance between reproducibly eliciting stable cellular processes and promoting variable responses under changing environments is fundamental for living systems (9). A previous study reported clustering of genes exhibiting low variability on highly accessible genomic regions (10). Although that study established the importance of gene order in gene expression variability, whether intergenic distance and orientation between adjacent genes are important for gene expression stability remains to be tested.

Gene expression can vary because of inherent stochasticity of transcription or as a response to changes in environment. Variations in gene expression can be deleterious for some genes; for example, high variations in gene expression among genes encoding protein complex subunits can cause dosage imbalance among the subunits of a protein complex (11). On the other hand, variations in gene expression can be a mechanism for responding to fluctuations in environments (12).

Gene-expression variability can arise from modification of chromatin structure (13). Recently, several studies pinpointed nucleosome occupancy at the core promoter, also called a "nucleosome-depleted region" (NDR), as a key determinant of gene expression variability; high nucleosome occupancy at the NDR is associated with highly variable gene expression, and low nucleosome occupancy is associated with stable gene expression (12, 14, 15). We posited that nucleosome occupancy at the NDR can be influenced by clustering with the upstream gene.

We studied the relationship between the upstream clustering pattern and gene expression variability in the budding yeast *Saccharomyces cerevisiae* using genome-wide datasets of gene expression and nucleosome occupancy. We found that genes with a short upstream distance exhibited significantly lower gene expression variability than would be expected by chance. They also showed strong nucleosome depletion at the core promoter; the depletion was greater in the head-to-head orientation than in the head-to-tail orientation. Our data support the view that the depletion was mediated by in vivo mechanisms rather than by intrinsic nucleosome-disfavoring sequences. Finally, the fractions of head-to-head and head-to-tail genes varied in different eukaryotes. In mammals, head-to-tail genes were ~10 times less frequent than head-to-head genes. We propose an explanation for these observations.

Results

Gene Expression Variability and Upstream Configuration. We studied the relationship between a gene's expression variability and its orientation and distance to its upstream gene in *S. cerevisiae* (Fig. 1*A*). We obtained estimates of gene expression variability resulting from the stochastic nature of transcription (intrinsic variability), environmental perturbations (responsiveness to environmental perturbations), and processes that influence epigenetic state of the chromatin (sensitivity to chromatin regulation) (*Methods*). All three measures of gene expression variability decreased significantly as the upstream distance decreased (Fig. 1 *B–D*). These results underscore the importance of a short upstream distance for low gene expression variability. The trend was significant in both head-to-head and head-to-tail orientations. Genes in head-to-head orientation showed moderately lower variability than those in head-to-tail orientation (Fig. 1 *B–D*). Given the importance of epigenetic processes in gene expression variability, we thereafter focused on sensitivity to chromatin regulation for subsequent analyses. We define a head-to-head or head-to-tail gene as a gene with an upstream distance <300 bp.

Author contributions: Y.H.W. and W.-H.L. designed research; Y.H.W. performed research; Y.H.W. analyzed data; and Y.H.W. and W.-H.L. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: whli@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1100210108/-/DCSupplemental.

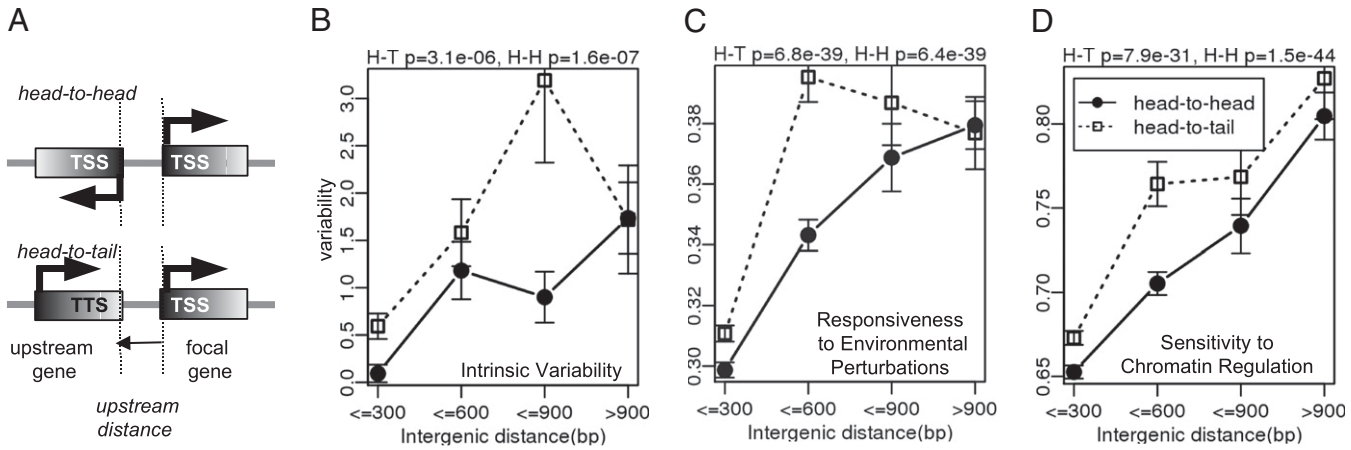

Fig. 1. Upstream orientation, upstream distance, and gene expression variability. (*A*) Clustering patterns with the upstream gene. From the perspective of the focal gene, the orientation with the upstream gene (the upstream orientation) is head-to-head when the 5′ end (head) of the focal gene is next to the 5′ end (head) of the upstream gene and is head-to-tail when 5′ end (head) of the focal gene is next to the 3′ end (tail) of the upstream gene. The upstream distance is the distance in bp between the 5′ end of the focal gene and the 5′ or 3′ end of the upstream gene. (*B–D*) Association between gene expression variability (*y* axis) and the upstream distance (*x* axis). Points and error bars indicate the mean and SE. Statistical significance is given for each orientation. (*B*) Intrinsic variability. (*C*) Responsiveness to environmental perturbations. (*D*) Sensitivity to chromatin regulation.

Promoter Architecture and Upstream Distance. The promoter architecture is the best-known determinant of gene expression stability. Especially high gene expression variability is observed for genes containing many transcriptional factor binding sites (TFBSs) or a TATA box at the promoter (12, 14, 15). Our own analysis confirmed that gene expression variability is correlated positively with the number of TFBSs and the presence of a TATA box (Table S1). Importantly, the number of TFBSs and the presence of a TATA box decreased as the upstream distance decreased, underscoring their importance in stable expression of genes with a short upstream distance (Fig. S1). After accounting for these features, gene expression variability still decreased as the upstream distance decreased, suggesting that other factors contribute to stable gene expression among genes with a short upstream distance (Table S1).

Nucleosome Occupancy, Gene Expression Stability, and Upstream Configuration. We examined the relationship between gene expression variability and nucleosome occupancy at the promoter. We focused on nucleosome occupancy at the core promoter region, the NDR (Fig. 2*A*) (12, 14, 15). Nucleosome occupancy can be influenced by intrinsic nucleosome sequence preferences and mechanisms that change nucleosome occupancy independent of sequence preferences (hereafter referred to as "in vivo mechanisms"). Nucleosome occupancy measured in vitro from a reconstitution experiment is influenced only by intrinsic nucleosome sequence preferences, whereas occupancy measured in vivo is influenced by both intrinsic nucleosome sequence preferences and in vivo mechanisms (Fig. 2*A*) (16, 17). Gene expression variability was associated positively with in vivo nucleosome occupancy but negatively with in vitro nucleosome occupancy (Fig. 2*B*). In other words, stably expressed genes tend to show low in vivo but high in vitro nucleosome occupancy at the NDR. These observations suggest that in stably expressed genes nucleosome depletion at the NDR tends to be mediated by in vivo mechanisms rather than by strong nucleosome-disfavoring sequences (*Discussion*).

We analyzed nucleosome occupancy at the NDR as a function of the upstream distance. As the upstream distance decreased, we detected a trend of decreasing nucleosome occupancy in vivo (Fig. 3*A* and Fig. S2 *A* and *D*). For in vitro nucleosome occupancy, the trend was not clear (Fig. 3*B* and Fig. S2 *B* and *E*). When in vitro nucleosome occupancy was subtracted from in vivo occupancy, the trend of decreasing occupancy was even stronger (Fig. 3*C* and Fig. S2 *C* and *F*). These observations suggest that in vivo mechanisms, rather than intrinsic nucleosome sequence preferences, mediate the upstream distance-dependent decrease in nucleosome occupancy. Interestingly, the nucleosome occupancy was lower in head-to-head than in head-to-tail orientation, suggesting that the vivo mechanisms for stable nucleosome depletion are stronger in the head-to-head orientation (Fig. 3 *A* and *C* and Fig. S2 *A*, *C*, *D*, and *F*).

How is nucleosome depletion in the NDR achieved in genes with a short upstream distance? In yeast promoters, an NDR generally is flanked by two strongly positioned nucleosomes, the +1 and −1 nucleosomes. Greater positioning strength of the two nucleosomes was associated significantly with greater in vivo nucleosome depletion at the NDR, consistent with the notion that strong in vivo positioning of the flanking nucleosomes that prevents encroachments facilitates nucleosome depletion at the NDR (18) (Fig. S3). Importantly, the positioning strength of the two nucleosomes increased with decreasing upstream distance (Fig. 4 and Fig. S4). The positioning was stronger for the head-to-head orientation; this observation makes biological sense, because the −1 nucleosome position of a head-to-head gene is close to or is occupied by the +1 nucleosome of the upstream gene, depending on the upstream distance, and because nucleosome positioning across a yeast gene is the strongest at +1 position (Fig. 4 and Fig. S4). These observations support the view that close clustering with the upstream gene facilitates precise positioning at the flanking regions and is preferable for stable nucleosome depletion at the NDR.

Upstream Distance and Orientation in Diverse Eukaryotes. We investigated the importance of the upstream configuration across six diverse eukaryotes: one yeast (*S. cerevisiae*), one plant (*Arabidopsis thaliana*), one fly (*Drosophila melanogaster*), and three mammals (*Mus musculus*, *Rattus norvegicus*, and *Homo sapiens*). We computed the proportion of genes with an upstream distance <300 bp for the head-to-head and the head-to-tail orientation. For both orientations, the proportion decreased in higher eukaryotes; the extent of decrease was less for the head-to-head orientation than for the head-to-tail orientation (Fig. 5*A*). As a result, the ratio of head-to-head genes to head-to-tail genes (<300 bp) was ≤1 in yeast and plant but >1 in fly and mammals; it was as high as ∼10 in mammals (Fig. 5*A*).

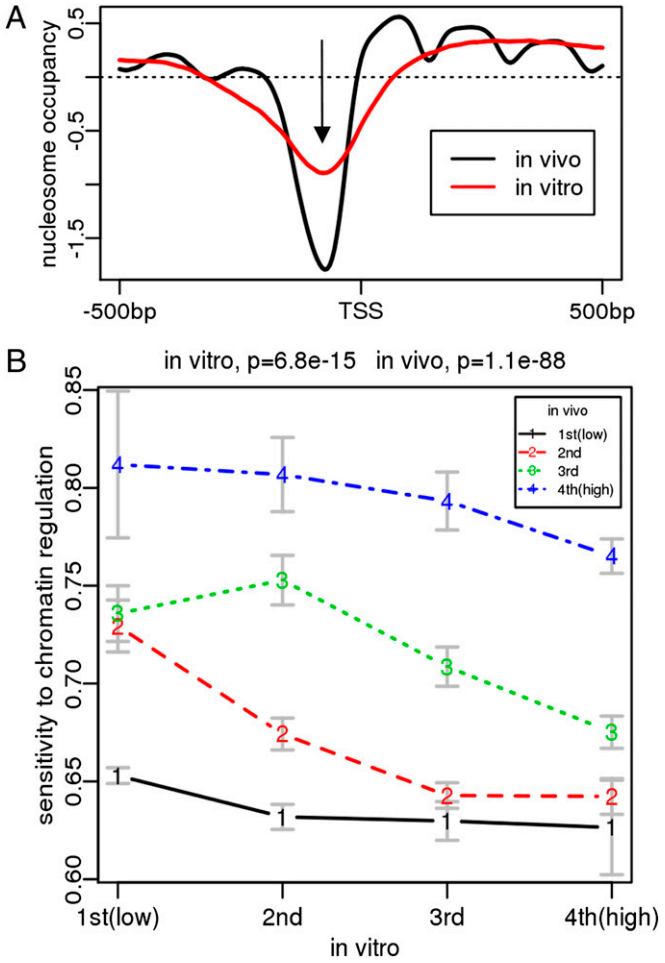

Fig. 2. Gene expression variability vs. nucleosome occupancy at the promoter. (*A*) In vivo and in vitro nucleosome occupancy across the promoter region. *X* axis: distance to the TSS (left: upstream; right: downstream). *Y* axis: nucleosome occupancy relative to the genome-wide average. Positive and negative values indicate enrichment and depletion, respectively. The NDR was defined as the region 65–90 bp upstream of the TSS (indicated by arrow). (*B*) Sensitivity to chromatin regulation as a function of in vivo and in vitro nucleosome occupancy at the NDR.

We investigated possible reasons for greater loss of head-to-tail genes in higher eukaryotes. First, we examined the gene expression variability of head-to-head and head-to-tail genes. In species other than *S. cerevisiae, w*e estimated intrinsic gene expression variability as variations in microarray gene expression level across biological replicates. In all species tested, head-to-head genes (upstream distance <300 bp) exhibited gene expression variability that was significantly lower than the genome-wide average (Fig. 5*B*). In contrast, head-to-tail genes exhibited lower gene expression variability in some but not all species (Fig. 5*B*). In support of this result, in *D. melanogaster* the nucleosome occupancy of the core promoter was significantly lower ($P < 10^{-12}$) in head-to-head genes than in head-to-tail genes (Fig. S5 *A* and *B*). These observations suggest that stably expressed genes tend to have the head-to-head orientation rather than the head-to-tail orientation. The difference in gene expression variability was not large, suggesting that other mechanisms contributed to the loss of head-to-tail genes. Interestingly, the length of 3′ UTR increased drastically in mammals, whereas the length of 5′ UTR was relatively constant in all six eukaryotes, raising the possibility that increased regulatory complexity at the 3′ end of genes in mammals contributed to a much greater loss of head-to-tail genes (Fig. 5*C*) (*Discussion*).

Discussion

In summary, we found a significant association between gene expression stability and a short upstream distance. We also found that as the upstream distance decreases, nucleosome occupancy at the NDR decreases, and the positioning strength of the nucleosomes that flank the NDR increases. Given the same upstream distance, genes in the head-to-head orientation tend to exhibit lower expression variability than genes in the head-to-tail orientation, providing an explanation for greater abundance of head-to-head genes than head-to-tail genes in complex organisms. We propose that flanking the NDR by two strongly positioned +1 nucleosomes results in a promoter architecture favorable for stable gene expression.

In this study, fewer TFBSs were found on promoters of genes with a short upstream distance. There are two possible reasons. First, when the number of TFBSs in a promoter is small, there is no need for a long upstream distance. Second, the short upstream distance constrains the complexity of the regulatory modules, contributing to stable gene expression. The latter possibility is important for the following reasons. Even after we accounted for the number of TFBSs, the upstream distance was a significant determinant of gene expression variability (Table S1). Also, the number of TFBSs did not increase linearly with the upstream distance. Interestingly, nucleosome occupancy around TFBSs decreased as the upstream distance decreased, probably because of the concentration of these motifs near the NDR, where nucleosome occupancy decreases as the upstream distance decreases (Fig. S6) (19). The upstream distance seems not to be merely a consequence of fewer TFBSs but to be important for the promoter architecture.

Nucleosome depletion at the core promoter region, a key determinant of stable gene expression, is influenced by both intrinsic nucleosome sequence preferences and by in vivo mechanisms (16, 17, 20). In this study, gene expression variability was associated positively with in vivo occupancy but negatively with in vitro occupancy. This paradoxical result might be understood from the evolutionary perspective: Nucleosome-disfavoring sequences, which mainly are repetitive sequences, lead not only to nucleosome depletion but also to increased evolvability, which would often be disadvantageous for stably expressed genes involved in fundamental cellular processes. This view is supported by a recent finding that genes containing tandem repeats, known to reduce nucleosome occupancy, in the promoter exhibit higher gene expression evolvability (21). We speculate that rarely would genes favoring evolutionary stability have a promoter in which the nucleosome depletion at the NDR is dominated by genetically unstable nucleosome-disfavoring sequences. Balanced use of the two mechanisms for nucleosome depletion is important for fine-tuning gene expression evolvability and stability.

How is in vivo nucleosome depletion at an NDR achieved for genes with a short upstream distance? In yeast, the +1 and −1 nucleosomes were proposed to stabilize nucleosome depletion at the NDR by preventing nucleosome encroachment (18). According to this view, positioning of both +1 and −1 nucleosomes would be important for nucleosome depletion at the NDR. Positioning at the +1 position usually is strongest across a yeast gene. We propose that nucleosome positioning at −1 position is strengthened by close juxtaposition of the 3′ or 5′ end of the upstream gene. Especially for head-to-head genes, where the upstream distance is short (i.e., <300 bp), the −1 nucleosome of the focal gene would be the +1 nucleosome of the upstream gene. The mechanistic relationship between upstream configuration and gene expression variability will be understood better as our understanding of how transcription is regulated by nucleosome positioning improves.

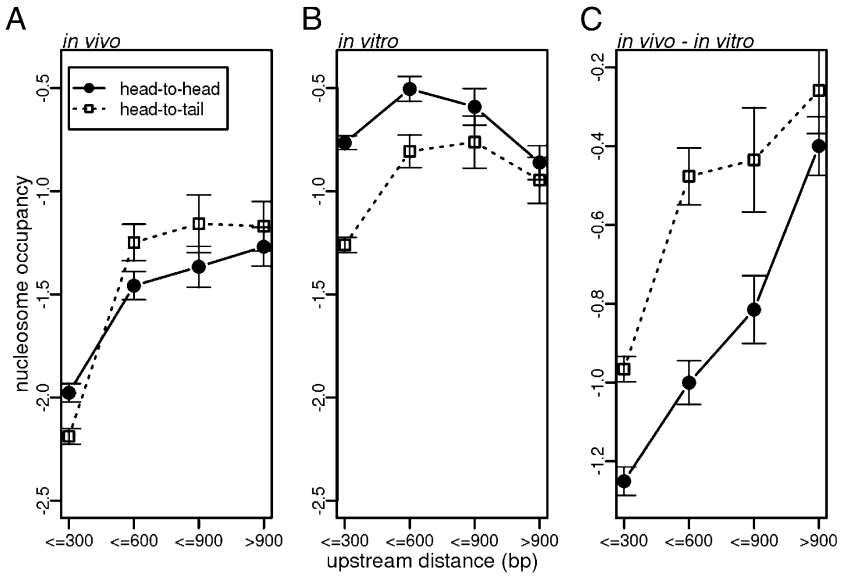

Fig. 3. Nucleosome occupancy at the NDR as a function of the upstream distance. (*A*) In vivo nucleosome occupancy. (*B*) In vitro nucleosome occupancy. (*C*) In vivo minus in vitro nucleosome occupancy.

The distribution of head-to-head and head-to-tail genes varies across eukaryotic taxa. Both are relatively abundant in yeasts and plants. In contrast, in mammals, head-to-tail genes can be as much as 10 times fewer than head-to-head genes. We list several possible reasons for the much greater loss of head-to-tail genes in mammals. First, head-to-tail genes showed higher gene expression variability than head-to-head genes in diverse eukaryotes, suggesting that the higher variability posed weaker constraints on the upstream distance in the head-to-tail orientation. Second, length and regulatory complexity increased faster at the 3′ UTR than at the 5′ UTR in more complex organisms (Fig. 5*C*) (22). Possibly, it has become increasingly difficult for the 5′ end of the focal gene to cluster tightly with the 3′ end of another gene. Consistent with this notion, nucleosome occupancy at the −1 position is less in *D. melanogaster* than in *S. cerevisiae* (23). In our preliminary analysis in *D. melanogaster*, nucleosome occupancy at the −1 position was close to the genome-wide average for head-to-tail genes but was high for head-to-head genes, suggesting that the stabilizing effects of the upstream gene's +1 nucleosome is even more important, constraining the upstream distance for head-to-head genes (Fig. S5 *C* and *D*) (24). These questions will be understood better with further elucidation of gene regulation by 3′ UTR processing and nucleosome positioning.

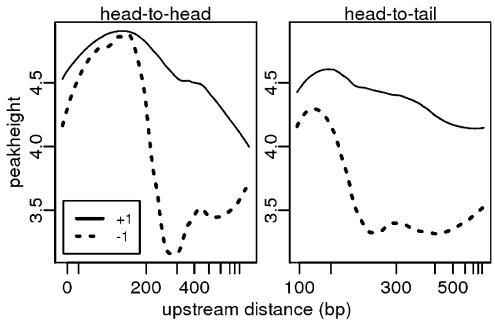

Fig. 4. Positioning strength of nucleosomes as a function of the upstream distance. (*Left*) Head-to-head orientation. (*Right*) Head-to-tail orientation. *Y* axis: the maximum occupancy ("peak height") for each positioned nucleosome (30). Nonparametric smoothing was performed by the loess algorithm, using the rank of upstream distance as the predictor variable. For the head-to-tail orientation, we used the distance from the TSS to the ORF end of the upstream gene, because nucleosomes tend to be highly positioned at the ORF end (16, 20). We detected the same trend when using the TTS instead of the ORF end and when using positioning fuzziness instead of peak height (Fig. S4).

This study examined various measures of gene expression variability. In lieu of direct measurement of variations, microarray-based gene expression variations were used as a surrogate measure of intrinsic variability. Because microarrays measure the average expression level in pooled cells, microarray-related artifacts, average population-wide behaviors, or cell-type heterogeneity in a tissue could influence our conclusion. These possibilities are unlikely for the following reasons. In *S. cerevisiae*, we detected the same trend for single cell-based variability (25). Genes encoding protein complex subunits, where high gene expression variability could be deleterious because of dosage imbalance among subunits, tend to keep short upstream distances in both yeast and human, further corroborating our conclusion that short upstream distances are important for stable gene expression (Fig. S7) (11). As the ability to measure gene expression level in single cells advances, difference between different types of variability will be better elucidated.

The association between gene expression variability and upstream distance does not reveal a causal relationship between the two. However, our preliminary analysis suggests that a short upstream distance is important for stable gene expression; of the genes with a short upstream distance in *S. cerevisiae*, those with low gene expression variability were more likely to have a short upstream distance in *Saccharomyces bayanus*, suggesting that they are maintained selectively for stably expressed genes during yeast evolution (Fig. S8). The causal contribution of short upstream distances to gene expression variability could be tested experimentally, for example, by inserting or deleting nonfunctional sequences in the promoter regions. Successful manipulation of the upstream distance without disturbing transcription and measuring gene expression variability with high precision are serious challenges and warrant future investigation.

The upstream distance could also be important for the gene's evolvability. We conjecture that short upstream distances constrain mutable target sizes and restrict the use of genetically unstable sequences for nucleosome depletion, thus lowering evolvability. This restriction would be advantageous for stably

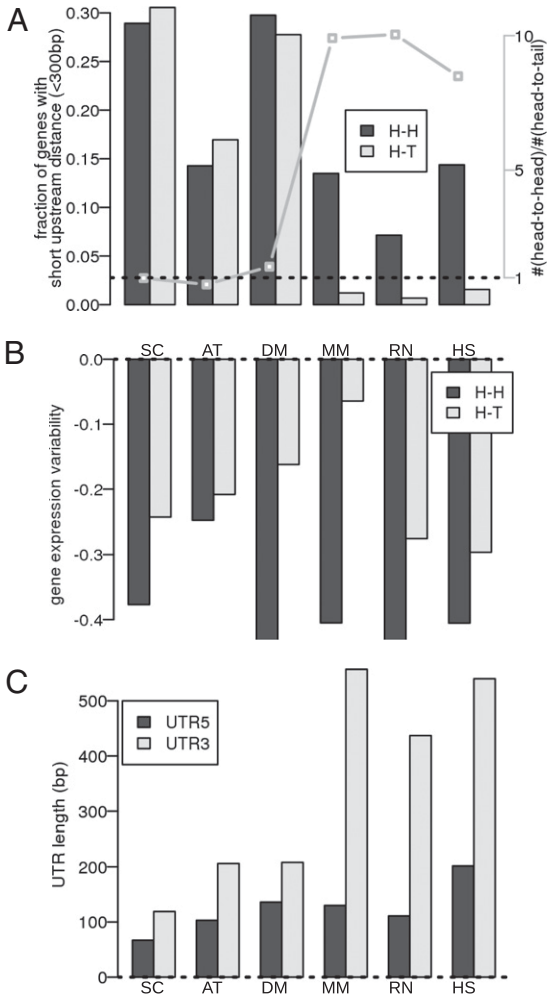

Fig. 5. Upstream orientation and distance in six eukaryotes: *S. cerevisiae* (SC), *A. thaliana* (AT), *D. melanogaster* (DM), *M. musculus* (MM), *R. norvegicus* (RN), and *H. sapiens* (HS) (*A*) Bars show the fraction of genes with a short upstream distance (<300 bp) in the head-to-head (H–H, dark gray bars) and head-to-tail (H–T, light gray bars) orientation. The ratio of head-to-head to head-to-tail genes is shown as a line; a higher value indicates greater abundance of head-to-head genes. (*B*) The average intrinsic gene expression variability for head-to-head (H–H, dark gray bars) and head-to-tail (H–T, light gray bars) genes compared with the remaining genes in the genome. (*C*) Median lengths of 5′ UTR (dark gray bars) and 3′ UTR (light gray bars).

expressed genes involved in fundamental cellular processes. The upstream configuration, together with the dynamic nature of the genome, may be a mechanism for constantly tinkering with stabilizing effects or fine-tuning the balance between stability and evolvability (26).

Methods

Genome-Wide Datasets and Data Processing. ***Gene expression.*** For *S. cerevisiae*, we obtained various measures of gene expression variability. For intrinsic variability, we obtained gene expression variation between single cells, measured using GFP and flow cytometry; we used a normalized version in which the coefficient of variation was converted to a distance-to-median metric to account for abundance-dependent variability (15, 25). For responsiveness to environmental perturbation, we averaged absolute gene expression changes across ~1,000 experiments (27). For sensitivity to chromatin regulation, we obtained an average of absolute expression changes upon deletion of various chromatin modifiers (15, 28). For other species, we obtained microarray gene expression data sets (Table S2) and performed ANOVA to estimate gene expression variations between biological replicates. The resulting variances were log-transformed and normalized for gene expression intensity-dependence by the loess algorithm (29).

Nucleosome occupancy. Reconstituted (in vitro) and in vivo (YPD medium) nucleosome occupancy maps were obtained from ref. 16. The individual nucleosome-positioning data were obtained from ref. 30. To calculate nucleosome occupancy independent of intrinsic nucleosome sequence preferences, we subtracted in vitro occupancy from in vivo occupancy. The subtracted occupancy was scaled to have a genome-wide average of zero.

Transcriptional factor binding sites. TFBS information was obtained from a published resource (http://fraenkel.mit.edu/improved_map) (31). The number of TFBSs was counted from the ORF start to the upstream ORF boundary, truncated at 2,000 bp upstream of the transcriptional start site (TSS). A genome-wide map of TATA boxes was obtained from a published resource (20).

Genomic annotations. For *S. cerevisiae,* we used genomic coordinates of ORFs from the Saccharomyces Genome Database (http://www.yeastgenome.org). Dubious ORFs were removed. Experimentally determined genome-wide TSSs and transcription termination sites (TTSs) were obtained from ref. 6. For *S. bayanus,* we obtained the ORF location from Yeast Gene Order Browser (v4) (32). For nonyeast species, we obtained longest "gene start" and "gene end" coordinates from the Ensembl, Ensembl plant, and Ensembl metazoa (August 2010); the analysis was restricted to protein-coding genes on major autosomes with NCBI Homologene ID. The 5′ (and 3′) UTR lengths were calculated as the difference between the TSS (and TTS) and the ORF start (and end) location for *S. cerevisiae*; for other species they were obtained from the UTRdb database (http://utrdb.ba.itb.cnr.it/) (33).

Protein complex. We obtained complex subunit data for *S. cerevisiae* from ref. 34 and for *H. sapiens* from the CORUM database (September 2009) (35).

Intergenic Distance. We calculated the intergenic distance as the distance between the transcript boundaries of two adjacent genes. We removed ORFs completely nested within another ORF. Nonoverlapping intergenic distances were analyzed in the study.

Statistical Analysis. Analyses were performed in the R environment (http://www.r-projects.org). We used a linear model to test the relationship between intergenic distance and/or orientation and the various gene properties in the study. When considering effects of several factors, we used type III ANOVA tests to assess the statistical significance of each factor after accounting for the other factors.

ACKNOWLEDGMENTS. We thank Z. Lin for helpful discussions. This work was supported by National Institutes of Health Grants GM30998 and GM081724.

1. Hurst LD, Pál C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5:299–310.
2. Herr DR, Harris GL (2004) Close head-to-head juxtaposition of genes favors their coordinate regulation in *Drosophila melanogaster*. *FEBS Lett* 572:147–153.
3. Trinklein ND, et al. (2004) An abundance of bidirectional promoters in the human genome. *Genome Res* 14:62–66.
4. Li YY, et al. (2006) Systematic analysis of head-to-head gene organization: Evolutionary conservation and potential biological relevance. *PLOS Comput Biol* 2: e74.
5. Wang Q, et al. (2009) Searching for bidirectional promoters in *Arabidopsis thaliana*. *BMC Bioinformatics* 10(Suppl 1):S29.
6. Xu Z, et al. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457:1033–1037.
7. Takai D, Jones PA (2004) Origins of bidirectional promoters: Computational analyses of intergenic distance in the human genome. *Mol Biol Evol* 21:463–467.
8. Lin JM, et al. (2007) Transcription factor binding and modified histones in human bidirectional promoters. *Genome Res* 17:818–827.
9. Wagner A (2007) *Robustness and Evolvability in Living Systems* (Princeton Univ Press, Princeton).
10. Batada NN, Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* 39:945–949.
11. Yang J, Lusk R, Li WH (2003) Organismal complexity, protein complexity, and gene duplicability. *Proc Natl Acad Sci USA* 100:15661–15665.
12. Tirosh I, Barkai N (2008) Two strategies for gene regulation by promoter nucleosomes. *Genome Res* 18:1084–1091.
13. Boeger H, Griesenbeck J, Kornberg RD (2008) Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell* 133: 716–726.
14. Field Y, et al. (2008) Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLOS Comput Biol* 4:e1000216.

15. Choi JK, Kim YJ (2009) Intrinsic variability of gene expression encoded in nucleosome *Nat Genet* 41:498–503.
16. Kaplan N, et al. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458(7236):362–366.
17. Zhang Y, et al. (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol* 16:847–852.
18. Cairns BR (2009) The logic of chromatin architecture and remodelling at promoters. *Nature* 461:193–198.
19. 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.
20. Mavrich TN, et al. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* 18:1073–1083.
21. Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ (2009) Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* 324: 1213–1216.
22. Mazumder B, Seshadri V, Fox PL (2003) Translational control by the 3′-UTR: The ends specify the means. *Trends Biochem Sci* 28:91–98.
23. Mavrich TN, et al. (2008) Nucleosome organization in the *Drosophila* genome. *Nature* 453(7193):358–362.
24. Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K (2009) Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* 19:460–469.
25. Newman JRS, et al. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441:840–846.
26. Poyatos JF, Hurst LD (2006) Is optimal gene order impossible? *Trends Genet* 22: 420–423.
27. Ihmels J, et al. (2005) Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science* 309:938–940.
28. Steinfeld I, Shamir R, Kupiec M (2007) A genome-wide analysis in Saccharomyces cerevisiae demonstrates the influence of chromatin modifiers on transcription. *Nat Genet* 39:303–309.
29. Kerr MK, Churchill GA (2007) Statistical design and the analysis of gene expression microarray data. *Genet Res* 89:509–514.
30. Jiang C, Pugh BF (2009) A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome. *Genome Biol* 10:R109.
31. MacIsaac KD, et al. (2006) An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* 7:113.
32. Byrne KP, Wolfe KH (2005) The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 15:1456–1461.
33. Grillo G, et al. (2010) UTRdb and UTRsite (RELEASE 2010): A collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 38(Database issue):D75–D80.
34. Gavin AC, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440:631–636.
35. Ruepp A, et al. (2008) CORUM: The comprehensive resource of mammalian protein complexes. *Nucleic Acids Res* 36(Database issue):D646–D650.