Tempo and mode of promoter evolution as observed in the Paramecium aurelia species complex

R. Taylor Raborn,*[,1,2] Timothy Licknack,[1,2] Wanfeng Guo,[1,2] Shannon N. Snyder,[1,2] and Michael Lynch[1,2]

[1]Biodesign Institute Center for the Mechanisms of Evolution

[2]School of Life Sciences

Arizona State University, 797 E. Tyler Street, Tempe, AZ 85281

***Corresponding author:** E-mail: rtraborn@asu.edu

Associate Editor:

Abstract

Goes here

Key words: *cis*-regulatory regions, duplicate genes, paralogous genes, *Paramecium*, promoter evolution, TSS profiling.

Introduction

A major challenge in genomics is to decipher the precise instructions that regulate gene expression. The regulation of gene expression underpins fundamental processes within the cell, such as growth, development, the maintenance of homeostasis and metabolism. Transcription, a major step in gene expression, is controlled through *cis*-regulatory regions bound by *trans*-acting transcription factors. Investigations across select eukaryotes have drawn attention to the important role of *cis*-regulatory sequences in the evolution of gene expression differences (Siepel and Arbiza, 2014; Wittkopp and Kalay, 2012; Wittkopp *et al.*, 2008). Work over the previous two decades have emphasized the contribution of *cis*-regulatory sequence differences toward changes in gene expression (Wittkopp *et al.*, 2004). The fulcrum and first major step of gene expression is transcription initiation, which in eukaryotes begins with the mediated interaction of RNA polymerase II complex with the promoter, a *cis*-regulatory region located proximal to the gene (Kad, 2012). The locations of eukaryotic *cis*-regulatory sequences, including promoters, cannot be predicted from genome sequence alone with precision.

Comparative and functional genomic approaches have shed light on the contributions of *cis*-regulatory sequences to phenotypic evolution and species divergence. Investigating F1 hybrids of *D. melanogaster* and *D. simulans*, Wittkopp and colleagues (Wittkopp *et al.*, 2004) showed that interspecific expression differences were primarily caused by changes acting in *cis*, not *trans*. Evolutionary changes in gene expression patterns have been noted in a number of instances, including in expression timing (*i.e.*

© The Author 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please email: journals.permissions@oup.com

Article

heterochrony) (Wray and McClay, 1989), expression levels ((Crawford *et al.*, 1999)), spatial differences (Abzhanov, 2000), and sex-specific expression (Kopp *et al.*, 2000); reviewed in (Wray *et al.*, 2003). The structure and conservation of the regulatory regions of select genes, including the *hox* (Kmita *et al.*, 2002) and *pax* (Plaza *et al.*, 1999) gene complexes, were established. More recently, the advent of large-scale functional genomics methods has identified sequence variants within regulatory regions that contribute to expression differences (Lappalainen *et al.*, 2013; Pickrell *et al.*, 2010; Schor *et al.*, 2017); a number of these have been functionally characterized. Global functional genomics efforts have revealed the cellular mechanisms of transcriptome and regulatory variation, but have largely been confined to major model systems, especially human, and have not focused on exploring the changes to *cis*-regulatory regions over evolutionary time-scales. As it stands, the precise *cis*-regulatory sequence differences that accompany—or indeed underpin—species divergence remain largely obscure. Another potential explanation for this is the difficulty in predicting *cis*-regulatory regions from genomic sequence alone. As such, accurate estimation of promoter positions requires direct, experimental evidence. At present, the most efficient approach to identify promoters at genome-scale is transcription start site (TSS) profiling, which includes CAGE (Cap Analysis of Gene Expression) (**?**) and RAMPAGE (Bat, 2013), among others. While TSS profiling methods differ in technical between themselves, these protocols all capture the 5'-ends of capped mRNAs, sequence their corresponding cDNAs and align the reads to the genome to identify the TSSs present within a given transcriptome. Clustering gene-adjacent TSSs defines a promoter, transcription start region (TSR), at single base-pair resolution, thereby providing genomic locations for *cis*-regulatory regions within a species (Lenhard *et al.*, 2012; Rach, Elizabeth A *et al.*, 2009). TSS profling studies in a variety of model organisms has demonstrated that the shape of TSS distributions (*i.e.* promoter shape) at promoters is related to the promoter class and correlates with the function of the associated gene (Carninci *et al.*, 2006; Hoskins *et al.*, 2011; Raborn *et al.*, 2016; Rach, Elizabeth A *et al.*, 2009). In addition, recent work done in *Drosophila melanogaster* provides evidence that promoter shape is itself a quantitative trait (Schor *et al.*, 2017), a finding that carries valuable implications for the understanding of the evolution of *cis*-regulatory regions.

A limited number of studies have investigated the evolution of promoters using TSS profiling information. Frith et al. (Frith *et al.*, 2006) compared TSSs between human and mouse tissues using the first generation of CAGE data. The authors observed 1250 instances of shifting promoter usage within homologous sites as

· doi:10.1093/molbev/mst012 MBE

well as tissue-specific TSS differences between species. Frith and colleagues also found that, consistent with expectations, TSSs with high turnover exhibited less sequence conservation in its promoter region than those without. The authors propose a model of gradual shifting between TSSs via alternative initiation sites, as well as shifting of intiation along the sequence itself.

Main et al. (Main *et al.*, 2013), used TSS profiling information generated in *D. melanogaster* to identify putative orthologous TSSs in three other *Drosophila* taxa. -Add a paragraph introducing the Paramecium system -¿ the experiment

-Final sentence(s) of intro: what we found plus the novelty eg this is the first study of its kind to do x.

Demographic structure

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Subsection 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Subsection 2

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Paragraph header Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Methods

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Results

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1) Item 1
2) Item 2
3) Item 3

- Consider a fall in population induced by a decline in the number of births in the economy, taking as given mortality and migration.
- It is well known that a lower population growth raises the capital–labor ratio in the Solow–Swan growth model.
- The same property holds in Diamond's (1965) overlapping generations model, and it enhances welfare as long as the economy is dynamically efficient; i.e., when the interest rate exceeds the population growth rate.

A similar trend is observed in the United States and advanced European countries (Gustafsson and Kalwij, 2006), and also in Canada, Australia, and New Zealand (Sardon, 2006). Interestingly, as pointed out by Bongaarts and Feeney (1998), even when the cohort's lifetime fertility rate (the number of children a mother has in her lifetime) does not fall, the delayed childbearing alone leads to a decline in the number of childbirths, measured by the total period fertility rates (TPFRs).

Model

Demographic structure

i.e.:

$$\lambda_t = \begin{cases} 0, \, t < 0, \\ \lambda, \, t \geq 0. \end{cases} \tag{1}$$

· doi:10.1093/molbev/mst012 MBE

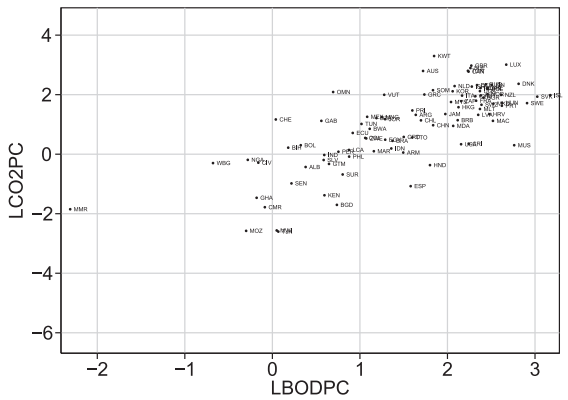

FIG. 1. Fluctuations in Cohort Size N_t over Generations.

Table 1. SH test results on nuclear and mitochondrial phylogenetic trees

Sequence data	Tree	$-\ln L$	SH test P-value
mtDNA	mtDNA	−109219.5	0.5
mtDNA	Nuclear	−61720.8	<0.00001
Nuclear	mtDNA	−113033.1	<0.00001
Nuclear	Nuclear	−60699.9	0.5

where C is a constant term defined as $C \equiv \beta \log \beta - (1+\beta)\log(1+\beta) + \beta \log A\alpha + (1+\beta)\log A(1-\alpha)$. Similarly, long-term welfare in the benchmark economy ($\lambda = 0$) can be written as:

$$U^* = (1+\beta)\log[A\alpha(k^*)^{2\alpha-1} + (k^*)^{\alpha}] - \beta(1-\alpha)\log k^* + C. \quad (2)$$

Supplementary Material

Supplementary tables S1?S7 and figures S1?S11 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors gratefully acknowledge the help of Robert Policastro and Gabriel Zentner at Indiana University Bloomington for technical assistance and use of laboratory facilities.

References

2012. Perspectives on the RNA polymerase II core promoter. 1(1): 40–51.

2013. RAMPAGE: Promoter Activity Profiling by Paired-End Sequencing of 5'-Complete cDNAs. pages 25B.11.1–25B.11.16. Current protocols in molecular biology / edited by Frederick M Ausubel [et al].

Abzhanov, A. 2000. Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. 127(11): 2239–2249.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A. M., Taylor, M. S., Engström, P. G., Frith, M. C., Forrest, A. R. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A., and Hayashizaki, Y. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, 38(6): 626–635.

Crawford, D. L., Segal, J. A., and Barnett, J. L. 1999. Evolutionary analysis of TATA-less proximal promoter function. *Molecular Biology and Evolution*, 16(2): 194–207.

Frith, M. C., Ponjavic, J., Fredman, D., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Hayshizaki, Y., and Sandelin, A. 2006. Evolutionary turnover of mammalian transcription start sites. *Genome research*, 16(6): 713–722.

Hoskins, R. A., Landolin, J. M., Brown, J. B., Sandler, J. E., Takahashi, H., Lassmann, T., Yu, C., Booth, B. W., Zhang, D., Wan, K. H., Yang, L., Boley, N., Andrews, J., Kaufman, T. C., Graveley, B. R., Bickel, P. J., Carninci, P., Carlson, J. W., and Celniker, S. E. 2011. Genome-wide analysis of promoter architecture in Drosophila melanogaster. *Genome research*, 21(2): 182–192.

Kmita, M., Fraudeau, N., Hérault, Y., and Duboule, D. 2002. Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs. *Nature*, 420(6912): 145–150.

Kopp, A., Duncan, I., and Carroll, S. B. 2000. Genetic control and evolution of sexually dimorphic characters in Drosophilia. *Nature*, 408(6812): 553–559.

Lappalainen, T., Sammeth, M., Friedländer, M. R., 't Hoen, P. A. C., Monlong, J., Rivas, M. A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P. G., Barann, M., Wieland, T., Greger, L., van Iterson, M., Almlöf, J., Ribeca, P., Pulyakhina, I., Esser, D., Giger, T., Tikhonov, A., Sultan, M., Bertier, G., MacArthur, D. G., Lek, M., Lizano, E., Buermans, H. P. J., Padioleau, I., Schwarzmayr, T., Karlberg, O., Ongen, H., Kilpinen, H., Beltran, S., Gut, M., Kahlem, K., Amstislavskiy, V., Stegle, O., Pirinen, M., Montgomery, S. B., Donnelly, P., McCarthy, M. I., Flicek, P., Strom, T. M., Geuvadis Consortium, Lehrach, H., Schreiber, S., Sudbrak, R., Carracedo, Á., Antonarakis, S. E., Häsler, R., Syvänen, A.-C., van Ommen, G.-J., Brazma, A., Meitinger, T., Rosenstiel, P., Guigo, R., Gut, I. G., Estivill, X., and Dermitzakis, E. T. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501(7468): 506–511.

Lenhard, B., Sandelin, A., and Carninci, P. 2012. Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, 13(4): 233–245.

Main, B. J., Smith, A. D., Jang, H., and Nuzhdin, S. V. 2013. Transcription start site evolution in Drosophila. *Molecular Biology and Evolution*, 30(8): 1966–1974.

Pickrell, J. K., Marioni, J. C., Pai, A. A., Degner, J. F., Engelhardt, B. E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y., and Pritchard, J. K. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. 464(7289): 768–772.

Plaza, S., Saule, S., and Dozier, C. 1999. High conservation of cis-regulatory elements between quail and human for the Pax-6 gene. *Development Genes and Evolution*, 209(3): 165–173.

Raborn, R. T., Spitze, K., Brendel, V. P., and Lynch, M. 2016. Promoter Architecture and Sex-Specific Gene Expression in Daphnia pulex. *Genetics*, 204(2): 593–612.

Rach, Elizabeth A, Yuan, Hsiang-Yu, Majoros, William H, Tomancak, Pavel, and Ohler, Uwe 2009. Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. *Genome Biology*, 10(7): R73.

Schor, I. E., Degner, J. F., Harnett, D., Cannavò, E., Casale, F. P., Shim, H., Garfield, D. A., Birney, E., Stephens, M., Stegle, O., and Furlong, E. E. M. 2017. Promoter shape varies across populations and affects promoter evolution and expression noise. *Nature Genetics*, 49(4): 550–558.

Siepel, A. and Arbiza, L. 2014. Cis-regulatory elements and human evolution. *Current opinion in genetics & development*, 29: 81–89.

Wittkopp, P. J. and Kalay, G. 2012. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, 13(1): 59–69.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. 2004. Evolutionary changes in cis and trans gene regulation. 430(6995): 85–88.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. 2008. Regulatory changes underlying expression differences within and between Drosophila species. *Nature Genetics*, 40(3): 346–350.

Wray, G. A. and McClay, D. R. 1989. Molecular Heterochronies and Heterotopies in Early Echinoid Development. *Evolution*, 43(4): 803–813.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A. 2003. The evolution of transcriptional regulation in

· doi:10.1093/molbev/mst012 MBE

eukaryotes. *Molecular Biology and Evolution*, 20(9): 1377–1419.