Redundancy and the Evolution of *Cis*-Regulatory Element Multiplicity

Tiago Paixão, Ricardo B. R. Azevedo*

Department of Biology and Biochemistry, University of Houston, Houston, Texas, United States of America

Abstract

The promoter regions of many genes contain multiple binding sites for the same transcription factor (TF). One possibility is that this multiplicity evolved through transitional forms showing redundant *cis*-regulation. To evaluate this hypothesis, we must disentangle the relative contributions of different evolutionary mechanisms to the evolution of binding site multiplicity. Here, we attempt to do this using a model of binding site evolution. Our model considers binding sequences and their interactions with TFs explicitly, and allows us to cast the evolution of gene networks into a neutral network framework. We then test some of the model's predictions using data from yeast. Analysis of the model suggested three candidate nonadaptive processes favoring the evolution of *cis*-regulatory element redundancy and multiplicity: neutral evolution in long promoters, recombination and TF promiscuity. We find that recombination rate is positively associated with binding site multiplicity in yeast. Our model also indicated that weak direct selection for multiplicity (partial redundancy) can play a major role in organisms with large populations. Our data suggest that selection for changes in gene expression level may have contributed to the evolution of multiple binding sites in yeast. We conclude that the evolution of *cis*-regulatory element redundancy and multiplicity is impacted by many aspects of the biology of an organism: both adaptive and nonadaptive processes, both changes in *cis* to binding sites and in trans to the TFs that interact with them, both the functional setting of the promoter and the population genetic context of the individuals carrying them.

Citation: Paixão T, Azevedo RBR (2010) Redundancy and the Evolution of *Cis*-Regulatory Element Multiplicity. PLoS Comput Biol 6(7): e1000848. doi:10.1371/journal.pcbi.1000848

Editor: Philip E. Bourne, University of California San Diego, United States of America

Received October 22, 2009; **Accepted** June 2, 2010; **Published** July 8, 2010

Funding: This work was supported in part by grants from the National Science Foundation (EF-0742803) and the James S. McDonnell Foundation to RBRA. TP was supported in part by a fellowship from the Fundação para a Ciência e a Tecnologia (Portugal). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: razevedo@uh.edu

Introduction

Promoters frequently contain multiple functional regulatory elements [1]. For example, the regulatory region for stripe 2 of *even-skipped* (*eve*) of the fruit fly *Drosophila melanogaster* comprises 17 binding sites for four transcription factors (TFs), including five binding sites (B1–B5) for the activator *bicoid* (*bcd*) [2]. How does *cis*-regulatory element multiplicity evolve? There are three possibilities. First, perhaps "more is better" when it comes to TF binding sites. Multiple binding sites may cause changes in the level of gene expression or in its robustness against variation in TF concentrations [1,3–5]. Second, multiplicity might be favored by selection, but independently of its functional consequences. For example, genotypes with many binding sites may be more likely to produce viable offspring after mutation or recombination with genotypes with fewer binding sites [6–9]. Third, *cis*-regulatory element multiplicity may arise by nonadaptive processes [9–11]. Stone and Wray [10] have shown that a population of 10^6 diploid individuals could evolve two identical copies of a 6 base pair (bp) binding site in a 200-bp promoter every 5.4×10^5 generations through random mutation and genetic drift alone. The intergenic regions of *Saccharomyces cerevisiae* are $\sim$400 bp long on average, whereas those of multicellular eukaryotes can be orders of magnitude longer.

The common thread to all the evolutionary scenarios listed above is redundancy, the ability of structurally identical elements to contribute to the same function [12–16]. Redundancy is thought to be widespread in biological systems. In eukaryotes, a large proportion of genes are duplicates, and deletion of one copy often has little or no phenotypic effect because the other copy can compensate for the loss of function [17]. Functionality and redundancy are more difficult to establish for the case of multiple *cis*-regulatory elements [1]. The five *bcd* binding sites in *eve* the stripe 2 enhancer are not fully redundant because loss-of-function mutations to B1, B2 or B3 cause reduced *eve* stripe 2 expression and gain-of-function mutations to B4 and B5 lead to increased expression [2,18]. However, redundancy was likely important in the evolution of these sites. When Ludwig and colleagues [3] compared the stripe 2 enhancers of different species of *Drosophila*, they found that some of them lacked the B3 site (Figure 1). This observation implies that the B3 site evolved recently in the lineage leading to the last common ancestor of *D. melanogaster* and *D. simulans*. Furthermore, the B3 site was probably redundant when it first appeared because the stripe 2 enhancers of three species lacking the B3 binding site were able to drive expression of a reporter gene in *D. melanogaster* embryos coincident with native *eve* stripe 2 (Figure 1). Thus, redundant transitional forms can, in principle, play an important role in the evolution of *cis*-regulatory element multiplicity [1,19]. In this paper we develop a model of binding site evolution and use it to evaluate the plausibility of different scenarios for the evolution of *cis*-regulatory element redundancy and multiplicity. We then test predictions obtained from our model using data from yeast.