

Title: **Evolution of the Dorsal-Ventral gene regulatory network in *Drosophila* species**

PIs: **D. Arnosti (MSU) and C. Adami (MSU)**

**Summary:** Gene regulation in animals is arguably at least as important as the genes that are being regulated. Animal body plans, their structures and in particular the functions that the animal morphology provides, are the consequence over time and space of successive regulatory and developmental processes. Gene regulation in animals is a highly complex process, and can be likened to a computation that the regulatory machinery performs [1]. Often, single genes are regulated by a complex network of genes with activators, repressors, attenuators and the like, and the elucidation of these networks has taken molecular and developmental biologists decades. But while we know a tremendous amount about how genes and their associated proteins evolve, much less is known about how regulatory systems evolve. We know the basic building blocks: multiple transcription factor binding sites that regulate the expression of other transcription factors that ultimately lead to the expression of the regulated gene. Each transcription factor has a specific affinity to its binding site, and binding sites can interact either synergistically or antagonistically. If we compare the regulatory regions for the same gene across species in the same family, we can see sometimes significant differences in the regulatory sequence. Are these differences adaptive? How do regulatory networks change in response to changes, either in the environment or in response to a change in body size? In the proposed work, we will analyze the gene regulatory network or “cis-regulatory module” (CRM) that regulates the patterning of a fly embryo in the dorsal-ventral axis. This is a well-studied system for which expression and sequence data is available from the Arnosti lab, and aligned homologous CRMs have been collected [2]. Yet, we do not know in detail how evolution affects such systems. What are the “operators” that evolution uses to change these networks? A computational analysis of the regulatory region of 12 *Drosophila* species will help us move towards a better understanding of how regulatory systems evolve.

**Methods:** Here we propose to test whether there have been lineage-specific adaptations within the physical rules that govern the regulatory sequences of the Dorsal-Ventral (DV) CRM, which is regulated by three transcription factors: Dorsal, Twist, and Snail. The physical rules that determine how CRMs pattern the DV axis comprise the binding affinity of the transcription factor to the DNA binding site and factor-factor interactions (cooperativity, both synergistic and antagonistic), and the CRM’s interactions with the basal transcription machinery [3] [4]. A number of labs have developed thermodynamic models [5] [3] [6] that can effectively fit the logic of interactions, thereby determining if one lineage has increased, for example, the cooperativity between two factors. We have previously created a thermodynamic model that allows for the most consistent annotation of the binding sites within a CRM for a narrow window of the target gene expression. This algorithm is a function of the concentrations of the transcription factors and the parametrization of the physical rules of the CRM. The output of the model is visualized in part A of Fig. 1 as a list of annotated binding sites (d=Dorsal, t=Twist, s=Snail), and a list of spacers labeled as empty sites (e) from the CRM sequences that activate the rhomboid gene in the species *D. melanogaster* and *D. virilis* (two example species from the 12 we will consider). In part B of the figure, the annotated binding sites of *D. virilis* are aligned to the sites of *D. melanogaster* by extracting four basepairs upstream and downstream of each binding site, then all possible pairwise alignments are calculated to determine which sites best match one another (how to do this best is still an open research problem). In the thermodynamic model, a parameter is set that determines the maximum number of sites any CRM can have in the DV network. In the example case we have set this to three sites for all factors, hence the Dorsal sites are labeled d1, d2, and d3 where the numerical order follows the order of the sites within the *D. melanogaster* species, and so on. In part C of the figure, the aligned sites are transformed into a symbolic sequence (a “language”). In part D of the figure we apply rules of the language to transform the *D. virilis* symbolic sequence into the *D. melanogaster* symbolic sequence.

The rules of the language are effectively operators that transform one CRM into another, such as for example: “insert binding site”, “delete binding site”, “duplicate binding site”, “add spacer”, “delete spacer”,

**A**

**Apply thermodynamic model to sequence and expression data**

```

rhomel      aaacgaaactctgggaagcggaaaaaggacacctgtctgtcggcgggaagcgcaagtggcggcggaatttctgattcgcatgccai
            gaggcactcgcataatgttgagcacatgttttgggggaaattcccgggcgacgggccaggaatcaacgtctctgtcgtgggaaaagcc
            cacgtctaccacagcccactcgggttac

            eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeed eeeeeeeeeeeeeeeeeeeee
            eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeed eeeeeeeeeeeeeeeeeeeee

rhovir      cgagtgtgtgtcgtgctgtgtgtgcaacaagtgcggaaattctgaaatgcacatgtggcagcacatgtcgagcgggaaaaaacgc
            tcgatgctcggccaaggaattcccagccaaggaagtccgccaaaacacgcccaaaagggccaattatgattac

            eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeed eeeeeeeeeeeeeet eeeeeeeeeet eeeeeeeeeed eeeeeeeee
            eeeeeeeeeeeeeed eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee
    
```

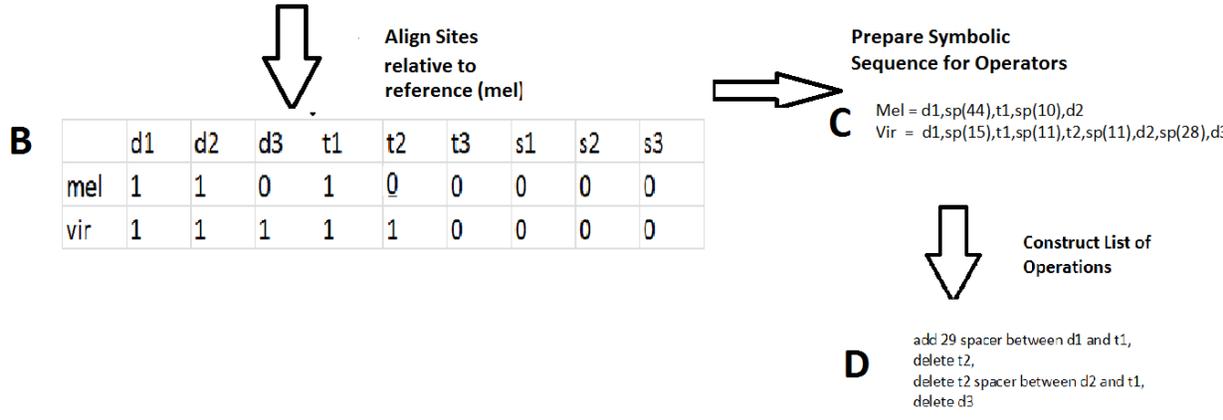


Figure 1: Workflow of the steps necessary to construct the evolutionary trajectories. A. The output of thermodynamic modeling leads to an annotation for the CRMs in terms of binding sites (d=Dorsal, t=Twist, s=Snail, e= empty). B. Sites are aligned relative the *D. melanogaster* reference binding sites. C. A symbolic sequence is prepared for operator action. D. The most parsimonious list of operations is applied to transform *D. virilis* into *D. melanogaster* reference.

”increase activity of site”, ”decrease activity of site”, and perhaps a few more. Once we defined our operators (each one of them can roughly be identified with a particular mutational “event”), **we can construct the most parsimonious transformation from one symbolic sequence to another.** This transformation will be guided by always using the *D. melanogaster* CRM as the reference (all symbolic sequences are to be transformed into the corresponding *D. melanogaster* symbolic sequence). By clustering *Drosophila* species that are closest to each other in this symbolic space, we can create a phylogeny of the *Drosophila* within the space of CRMs. By comparing this phylogeny to the genetic phylogeny, we can infer whether the changes to the CRM were adaptive in nature, or instead neutral. If the changes were adaptive, the inferred phylogeny can then be used to *reconstruct the evolutionary line of descent* of a given CRM.

Using this information, we can ask: Are the changes that we see in the DV gene regulatory network adaptive? Are some of the features monophyletic (in this ”distance phylogeny”)? Can we understand the sequence of genetic changes along reconstructed lines of descent? (For example, duplication followed by modification of activity and then fine-tuning spacing). Are certain types of changes common for a particular adaptive purpose? This way, perhaps we can come up with better models of how gene regulatory networks actually evolve.

## References

- [1] E. Davidson, *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. Burlington, MA: Academic Press Elsevier, 2006.
- [2] D. Papatsenko, Y. Goltsev, and M. Levine, “Organization of developmental enhancers in the drosophila embryo,” *Nucleic acids research*, vol. 37, no. 17, pp. 5665–5677, 2009.
- [3] R. P. Zinzen, K. Senger, M. Levine, and D. Papetatsenko, “Computational models for neurogenic gene expression in the Drosophila embryo,” *Curr. Biol.*, vol. 16, pp. 1358–1365, Jul 2006.
- [4] Y. T. Ip, R. E. Park, D. Kosman, E. Bier, and M. Levine, “The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the Drosophila embryo,” *Genes Dev.*, vol. 6, pp. 1728–1739, Sep 1992.
- [5] X. He, M. A. Samee, C. Blatti, and S. Sinha, “Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression,” *PLoS Comput. Biol.*, vol. 6, Sep 2010.
- [6] W. D. Fakhouri, A. Ay, R. Sayal, J. Dresch, E. Dayringer, and D. N. Arnosti, “Deciphering a transcriptional regulatory code: modeling short-range repression in the drosophila embryo.,” *Mol Syst Biol*, vol. 6, p. 341, 2010.