Finding the switches that activate animal genes through a combined *in silico* and *in vivo* approach

Chad M. Jaenke[1], Michael L. Weinstein[1], Marc S. Halfon[2], and Thomas M. Williams[1]

[1]Department of Biology, University of Dayton; [2]Department of Biochemistry, University at Buffalo-SUNY

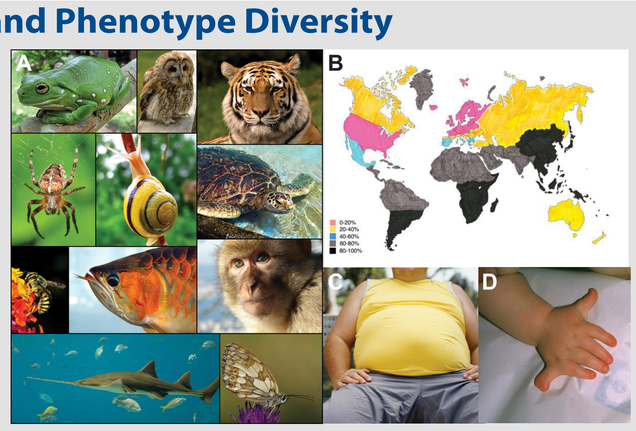
ABSTRACT

The DNA sequences of genomes encode the recipes for making functional cellular products, notably proteins, and switches that regulate when these products are made. While the genetic code for proteins has been known for decades, a similar code for the regulative switches is lacking. This presents a major challenge to understanding the genetic basis of life, as these switches (called ***cis*-regulatory elements** or **CREs**) may outnumber protein-coding genes by 20-50 fold. Both *in vivo* and *in silico* approaches exist to study CREs, but the former approaches are generally low throughput and not up to the scale of vast genomes, and the latter lack validation of predictions. We are merging *in silico* and *in vivo* approaches to identify the CREs controlling genes responsible for a fruit fly pigmentation trait. Here, we are leveraging the knowledge of six CREs that switch on the transcription of five different genes from a fruit fly tergite pigmentation **gene regulatory network** (**GRN**). We are using the SCRMshaw bioinformatic tool to identify novel predicted CREs controlling genes within this GRN based on underlying similarities in the DNA sequences of the known CREs. From this novel list, we will test 24 for CRE activity in *in vivo* reporter transgene assays. The results from these tests will reveal to what extent the *in silico* method succeeded. Novel validated CREs will be compared with the known six to reveal what the molecular functions are for the common DNA motifs as the next stage of this research project. The encoding of information in CREs is a universal feature of life, so these results bear upon life at every level, including the betterment of the human condition.

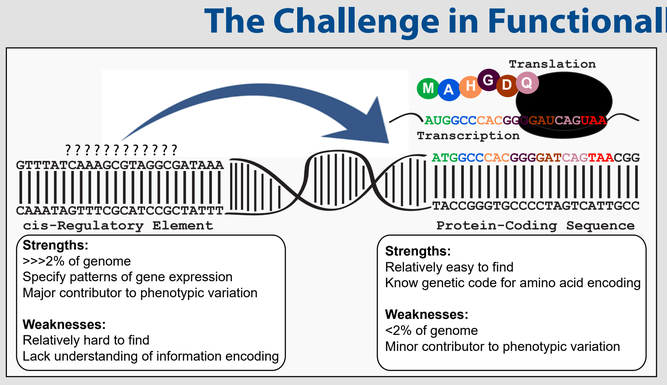
This project has received funding support from the Berry Summer Thesis Institute and the Biology Department at the University of Dayton, and the National Science Foundation (1555906).

CRE Evolution and Phenotype Diversity

CREs control gene expression patterns, and mutations in CREs are: (A) suspected to be a driver of morphological variation between diverse animals, (B) the cause of human lactose tolerance/intolerance, (C) a genetic cause of obesity, and (D) responsible for some cases of polydactyly, amongst other variations.

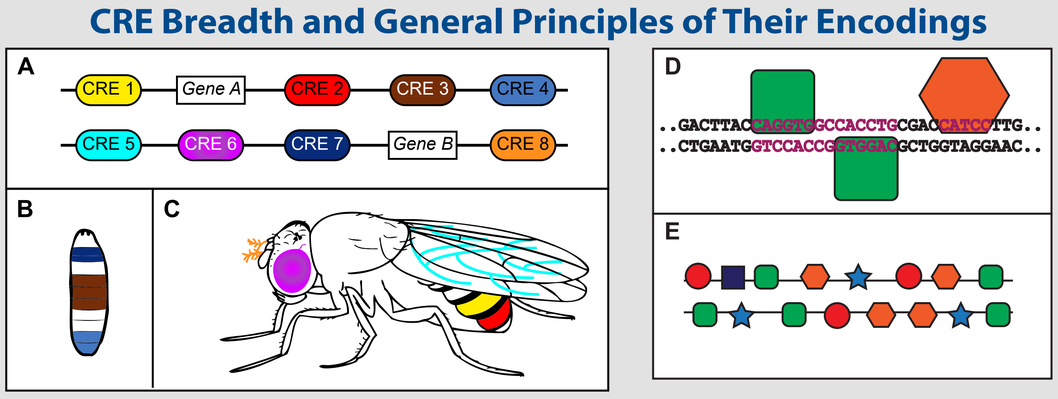


The Challenge in Functionally Annotating CREs



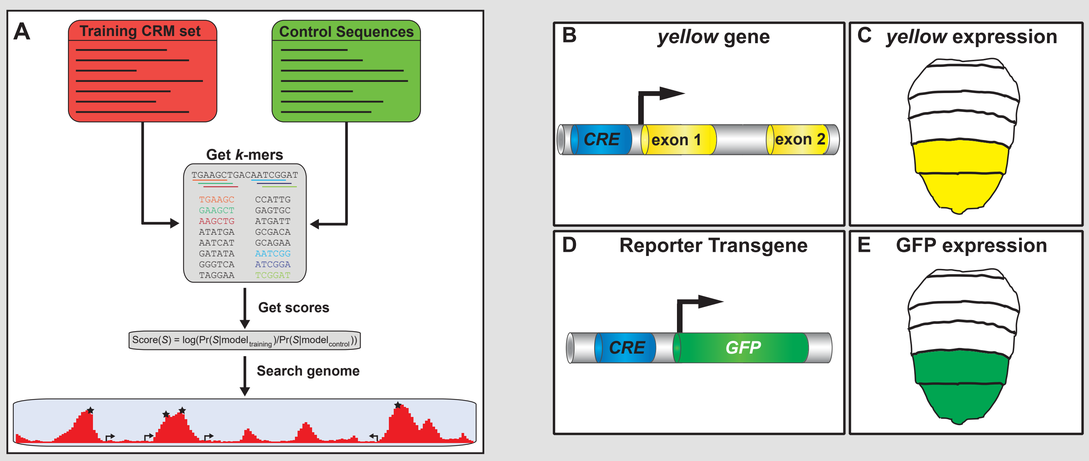
CREs activate transcription, though how particular CRE activities are encoded (???) in DNA sequence remains elusive, making these sequences more difficult to identify compared to protein-coding sequences.

CRE Breadth and General Principles of Their Encodings



(A) Genes are pleiotropic, and thus generally possess multiple CREs that each controls an aspect of the overall (B) temporal and (C) spatial expression pattern. (D) The fundamental unit of information in CREs are binding sites for a transcription factor protein. (E) Specific CRE activities are due to collaboration between two or more transcription factors binding to their sites.

A Combined *in silico* and *in vivo* Approach to Reveal More of the Pigmentation GRN



(A) We used SCRMshaw's machine-learning approach to evaluate statistics of short word counts in the six known CREs versus a background control set of sequences.

(B & C) We expect pigmentation GRN genes to have patterned expression controlled by a CRE.
(D & E) Regulatory activities can be seen *in vivo* by tests of reporter transgenes for which a known or putative CRE sequence is juxtaposed next to the *GFP* gene.

REFERENCES

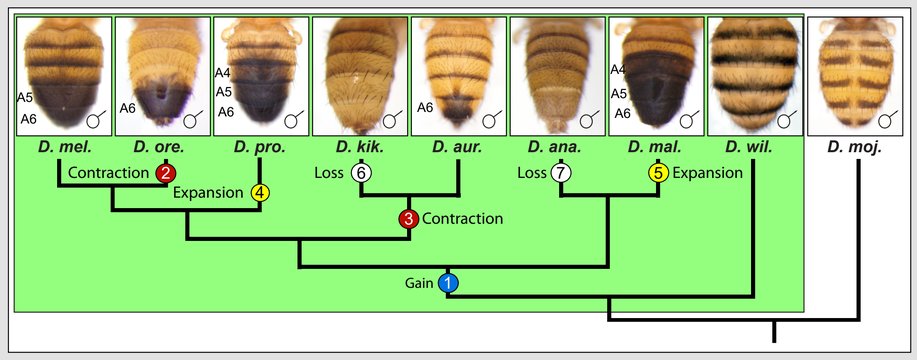
Kazemian, M., Zhu, Q., Halfon, M. S. and Sinha, S. (2011). Improved accuracy of supervised CRM discovery with interpolated Markov models and cross-species comparison. Nucleic Acids Res. 39, 9463–9472.

Rebeiz, M. and Williams, T. M. (2017). Using Drosophila pigmentation traits to study the mechanisms of cis-regulatory evolution. Curr. Opin. Insect Sci. 19, 1–7.

Finding the CREs for the Vast Networks of Genes that Build Traits

Morphological traits are constructed during development by the collective actions of scores to hundreds of genes, many whose identities are not known. The collaborations of binding sites for most CREs remain unknown. Thus, a major challenge for evo-devo research is to find the genes and their regulative CREs that comprise GRNs in order to facilitate investigations into how trait diversity evolved.

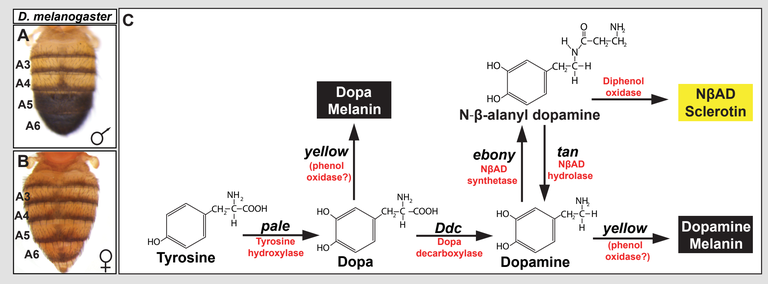
Fruit Fly Pigmentation as an Evo-Devo Model for CRE Function and Evolution



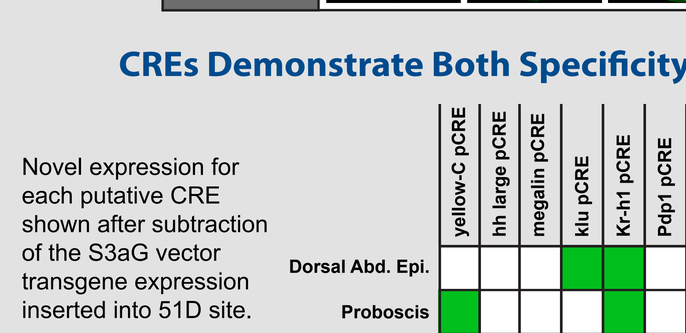
- Green background indicates the *Sophophora* subgenus. In *Sophophora* ancestry, male-specific tergite pigmentation was initially (1) gained, and this pattern (4 & 5) expanded, (2 & 3) contracted, and was (6 & 7) repeatedly lost.
- These species largely possess the same genes, thus this diversity was likely driven by CRE evolution. However, for which GRN genes and which CREs?

Spatial, Temporal, and Sex-specific Expression of Pigmentation Enzymes Underlie *D. melanogaster* Phenotype

(A & B) The *D. melanogaster* trait requires patterned expression of (C) pigment metabolism genes such as *tan*, *ebony*, *yellow*, and *Ddc*.



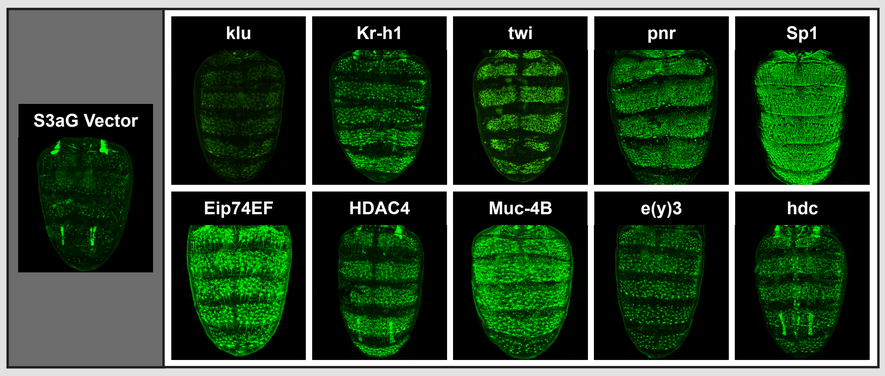
Elucidating the GRN for Tergite Pigmentation



20 years of research has revealed a rudimentary GRN for *D. melanogaster* tergite pigmentation, which includes pigment metabolism genes and several expression-regulating transcription factor genes. Six CREs were identified that can drive patterned expression of the *Green Fluorescent Protein* (*GFP*) reporter gene.

A major impediment to progression of this evo-devo model is finding the other genes and CREs that comprise the full GRN at a time when we still do not understand the encodings of the first six CREs.

10 of 18 Putative CREs Demonstrate Abdominal Enhancer Activity



CREs Demonstrate Both Specificity and Modest Pleiotropy

Novel expression for each putative CRE shown after subtraction of the S3aG vector transgene expression inserted into 51D site.

	yellow-C pCRE	hh large pCRE	megalin pCRE	klu pCRE	Kr-h1 pCRE	Pdp1 pCRE	CG9650 pCRE	Abd-B pCRE	yellow-h pCRE	Eip93F pCRE	hdc pCRE	HDAC4 pCRE	Eip74EF pCRE	Sp1 pCRE	pnr pCRE	e(y)3 pCRE	twi pCRE	Muc4B pCRE
Dorsal Abd. Epi.				■	■						■	■	■	■	■	■	■	■
Proboscis	■				■		■				■	■	■	■	■	■		■
Dorsal Head	■								■		■		■	■	■	■		

Future Directions

- Identify endogenous target gene of novel CREs and their potential function in an abdominal pigmentation GRN
- Test whether a larger set of training data improves SCRMshaw's ability to find pigmentation GRN CREs
- Investigate whether the CREs and their target genes of regulation palyed a role in the origin, diversification, and loss of sexually dimorphic pigmentation

ACKNOWLEDGEMENTS

CMJ was supported by the University of Dayton's Berry Summer Thesis Institute and the Department of Biology. TMW was supported through funding from the National Science Foundation (IOS-1146373 and IOS-1555906).