Computational prediction and subsequent validation of *cis*-regulatory modules in the Zika vector mosquito, *Aedes aegypti*

Gregory M. George and Marc S. Halfon

University at Buffalo Department of Biochemistry, New York State Center of Excellence in Bioinformatics and Life Sciences

Introduction

Aedes aegypti is the mosquito species that transmits the Zika and Dengue viruses. The increasing prevalence of these viruses makes understanding the development and life cycle of *Aedes aegypti* a high priority.

Gene Regulation

Pre-transcriptional regulation is achieved in part by the activity of distal regulatory sequences, labeled *cis*-regulatory modules(CRMs). They serve to recruit transcription factors as well as stabilize interactions between transcription factors and RNA Pol II.
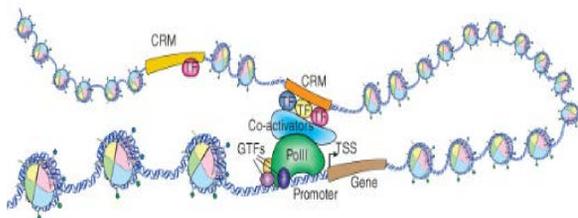

Figure 1. *cis*-Regulatory modules (CRMs). Adapted from Suryamohan and Halfon[1]

SCRMshaw

SCRMshaw is genome wide supervised *cis*-regulatory module prediction method developed by our lab that operates using machine learning algorithms. SCRMshaw predicts CRMs by identifying sequence "words", or "k-mers", from training sets of known *Drosophila* CRMs. These identified k-mers can subsequently be used to predict CRMs in *Drosophila*, as well as in other species over a large evolutionary range.

SCRMshaw: *Supervised CRM discovery*
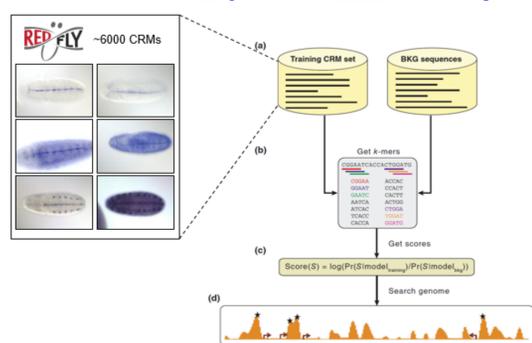
★★★ >80% true-positive rate in *Drosophila* by multiple measures ★★★

Figure 2. SCRMshaw method from Suryamohan and Halfon[1]

Training Set Selection

The training set of CRMs consists of *Drosophila melanogaster* enhancer sequences from the REDfly database that are expressed during late embryonic stages of development. The enhancers are involved with the development of the central nervous system of the fly and are expressed along the midline of the embryo. Midline expression was of interest as differences in expression patterns are noted between *Drosophila melanogaster* and *Aedes aegypti*.[1]
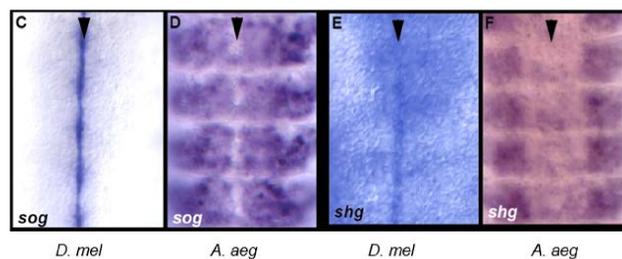

Figure 3. Expression patterns of midline *Drosophila* enhancers and *Aedes* homologs

Method of Validation

- Cloned into reporter gene construct
- *Drosophila* in β-green rabbit (GFP)
- *Aedes* in pLacZattb (βGalactosidase)
- Expression of enhancers viewed by antibody staining
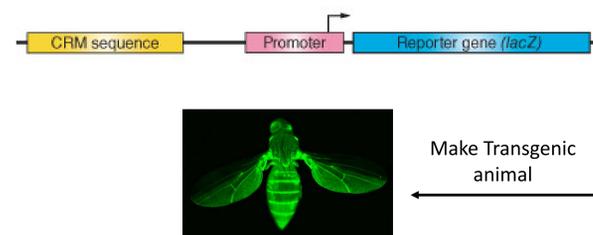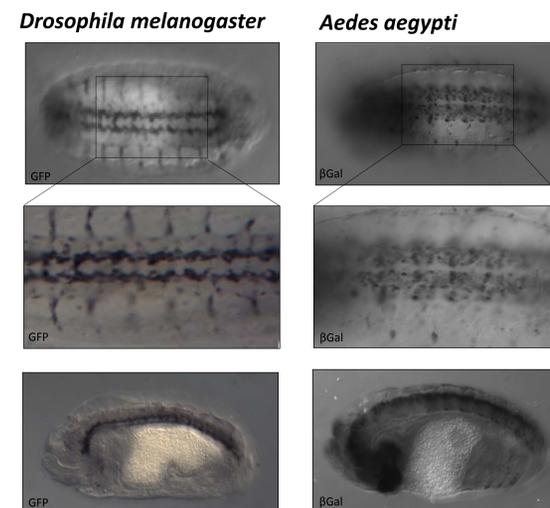 - α-GFP (*Drosophila*)
  - α-βGal (*Aedes*)

Make Transgenic animal

Figure 4. CRM incorporated into reporter construct adapted from Suryamohan and Halfon[1] and pmgbiology.com[2].

Predicted Enhancers from SCRMshaw

Predicted Construct	Activity *in vivo*
Pointed (*pnt*) - Dm	Present
Fasciclin 3 (Fas3) - Dm	Present
Roundabout 1 (robo1) - Dm	Present
Pointed (pnt) – Aae	Present
Fasciclin 3 (Fas3) – Aae	Present
Roundabout 1 (robo1) - Aae	Absent

Putative Pointed (pnt) Enhancer Expression

Drosophila melanogaster   **Aedes aegypti**

Putative Fasciclin (Fas3) Enhancer Expression

Drosophila melanogaster   **Aedes aegypti**

Putative Roundabout 1 (robo1) Enhancer Expression

Drosophila melanogaster   **Aedes aegypti**

Evolutionary Conservation of CRM Activity

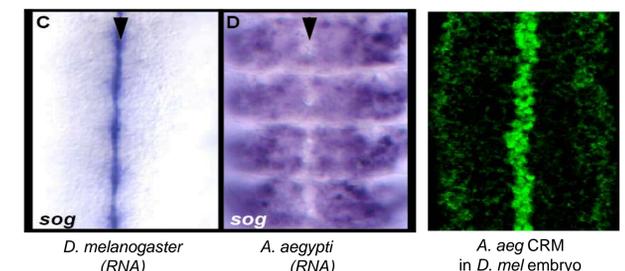*D. melanogaster* (RNA)   *A. aegypti* (RNA)   *A. aeg* CRM in *D. mel* embryo

Figure 5. *A. aegypti* enhancer drives expression in *D. Melanogaster* pattern in transgenic flies, demonstrating conservation of enhancer function

Conclusions

- Successfully validated five out of six SCRMshaw predicted enhancer sequences in *Drosophila melanogaster* and *Aedes aegypti*
- Expression in the central nervous system during late stages of embryonic development
- Expression patterns match those expected from the training set used

Future Work

- Directly compare expression of putative *Drosophila* and *Aedes* enhancers using fluorescent confocal microscopy
- Directly compare putative enhancer expression with endogenous gene expression using confocal microscopy
- Identification and verification of further sequences
  - Ventral veins lacking (vvl)
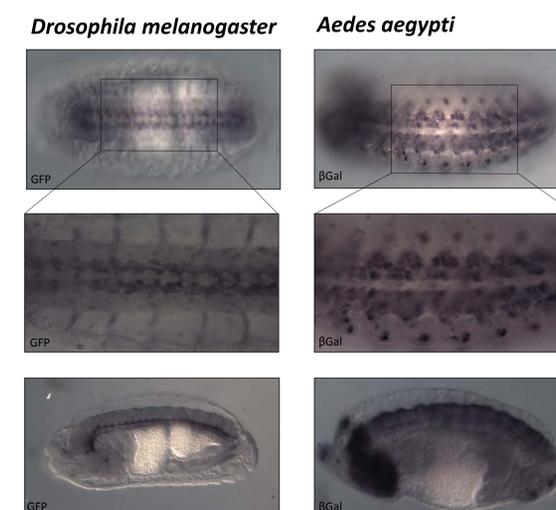 - SoxNeuro (SoxN)
  - Netrin A (NetA)

Significance

- Gene regulation is poorly understood in *Aedes aegypti*
- Increased understanding can lend insight into disease transfer

References

1. Suryamohan, K.; Halfon, M. S., Identifying transcriptional *cis*-regulatory modules in animal genomes. *Wiley Interdisciplinary Reviews: Developmental Biology* **2015**, *4* (2), 59-84.
2. GFP fly. pmgbiology.com.