Expression Pattern of Tandem Gene Duplication in *Drosophila*

Harry Du and Santiago Ferris advised by David Loehlin

Tandem Gene Duplication in *Drosophila*

Tandem duplication is a type of gene sequence duplication where the duplicated copy is directly adjacent to the original sequence.
We focus on the *Adh* gene of the *Drosophila* genus, which is naturally duplicated in some species.

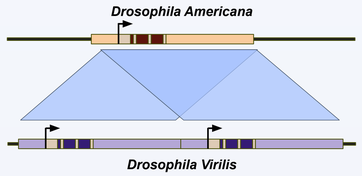

Figure 1: Comparison between Adh Gene of *Drosophila Americana* and *Drosophila Virilis*. The parallelogram shows that there is a match.

Expression Level of Tandem Duplications

Tandem gene duplications often result in expression levels higher than what's expected from simply doubling the gene copy number. This summer, ADH expression was measured in different *Drosophila* subspecies, including Virilis subspecies with significantly different duplicated structures. *virilis-09*, which is missing a large duplicated segment between its two copies containing a highly conserved, likely regulatory sequence, was significantly higher than every other strain.

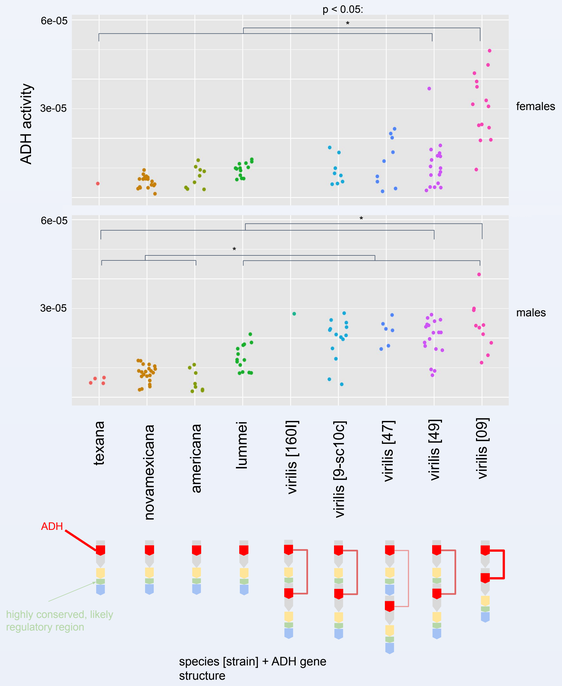

Figure 2. ADH activity across *Drosophila* subspecies with differing ADH gene structures.

Hypothesis 1: Cis-Regulatory Element

The over-expression primarily occurs at the RNA transcription level, suggesting a form of position effect where the tandem arrangement influences gene expression. One possible hypothesis is the duplication involves regulatory elements that enhance the binding of transcription factors.

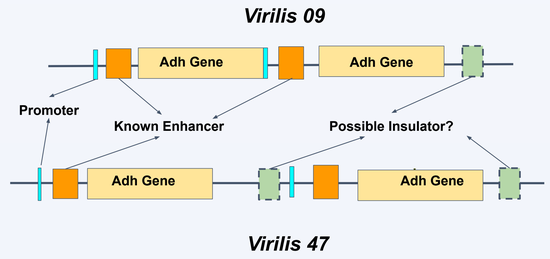

Figure 3: Structure of Adh Sequence in two Drosophila Virilis strain. We suspect that there may be a possible insulator downstream that stops the enhancer's effect on the second copy.

Alignment Algorithm to Locate Regulatory Elements

Analyzing regulatory elements in non-melanogaster Drosophila species is challenging because their genomes are less characterized.
We used a local alignment algorithm targeting tandem duplications, which highlights aligned regions and their divergence. Conserved regions revealed by this approach are strong candidates for regulatory sequences.

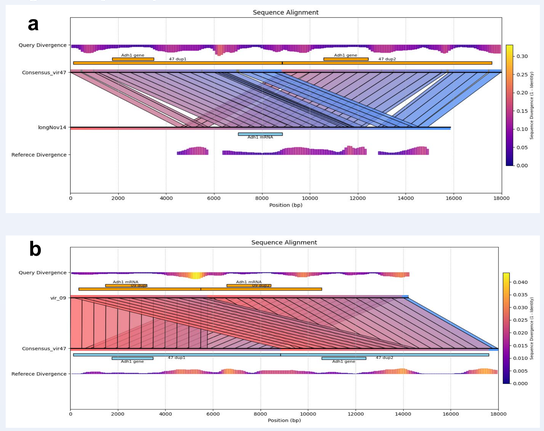

Figure 4: a. Alignments between *Drosophila Virilis 47* and *Drosophila Novamexicana.* **b.** Alignments between Virilis 09 and *Virilis 47* A parallelogram means that these sections are locally aligned. A Gaussian-smoothed Divergence Graph is shown on the top and the bottom of the sequence. Regulatory sequence are likely to locate in low-divergence regions.

Hypothesis 2: Distance Dependence

This over-expression can also be attributed to the distance between the two tandem duplicates. We can see this distance differ between some of the *virilis* strains we used (figure 2). We also used CRISPR to create *melanogaster* strains that differ from each other only in the length of an inserted binding-site free sequence to test this hypothesis.

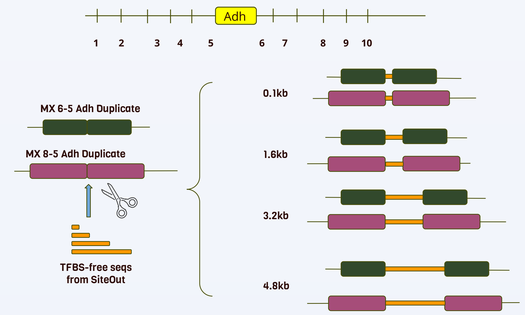

Figure 5: Distance Dependence experiment setup. The TFBS-free sequence are generated using SiteOut program from the DePace Lab. The sequences are inserted into Drosophila Melanogaster using CRISPR.

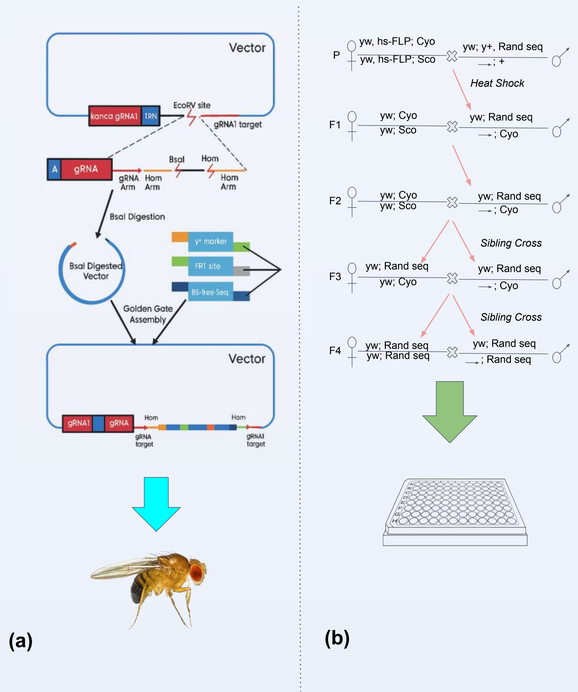

Figure 6: a. CRISPR plasmid components and assembly details. **b.** Crossing scheme to remove genetic marker and make the flies homozygous.

Expected Result

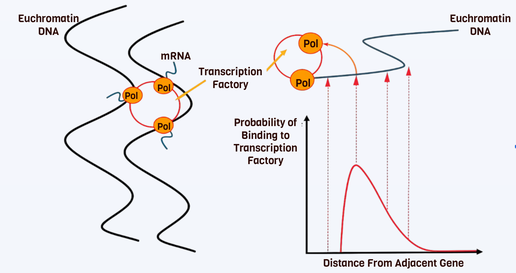

Figure 7: Transcription Factory Hypothesis, first proposed by Dr. Cook in 2006. In our case, we expect the genes separated by the correct distance will have the highest expression.

Future Directions

Current preliminary results from the CRISPR experiment do not support the distance-dependence hypothesis. A more rigorous test will be conducted after completing the genotyping of all flies.

The relatively high expression level of *virilis-09* compared to other double-copy strains modestly supports both hypotheses due to *virilis-09* not only having two close copies but also missing a likely regulatory region between them. Now that we can identify potential regulatory elements, the next step is to verify whether they influence Adh gene expression using CRISPR.

Acknowledgements

This research was supported by the Williams College Biology Department. We would like to thank Prof. Loehlin for his mentorship, and Olivia LaBelle for improving the fly DNA extraction procedure and for her help with fly genotyping.

Reference

Loehlin, D.W., McClain, G.L., Xu, M., Kedia, R. and Root, E., 2023. Demonstration of in vivo engineered tandem duplications of varying sizes using CRISPR and recombinases in Drosophila melanogaster. G3: Genes, Genomes, Genetics, 13(10), p.jkad155.

Estrada, J., Ruiz-Herrero, T., Scholes, C., Wunderlich, Z. and DePace, A.H., 2016. SiteOut: An online tool to design binding site-free DNA sequences. PLoS One, 11(3), p.e0151740.

Bon, M., Marenduzzo, D. and Cook, P.R., 2006. Modeling a self-avoiding chromatin loop: relation to the packing problem, action-at-a-distance, and nuclear context. Structure, 14(2), pp.197-204.