# STUDYING F ELEMENT AND INSULIN SIGNALING EVOLUTION IN *DROSOPHILA* USING COMPARATIVE GENOMICS

Veronica Gomes '24 and Dr. Jennifer Kennell
Vassar College Department of Biology

## INTRODUCTION

Comparative gene annotations provide important data sets that can be further analyzed to address many scientific questions, especially those involving the effects of evolutionary time on chromosomes and genes. This study consisted of gene annotations and genomic analysis for two separate projects under the Genomics Education Partnership – **the F element** and the ***Drosophila* Pathways**.

### F ELEMENT

The Muller F element is the smallest chromosome in *Drosophila melanogaster*. Although it is heterochromatic, the F element has a high gene density and exhibits high gene expression.

Since the F elements of *D. bipectinata, D. ananassae, D. kikkawai,* and *D. takahashii* show substantial expansion compared to that of *D. melanogaster*, the F element project involves annotating genes on the F element of these species to examine factors that contributed to the expansion.

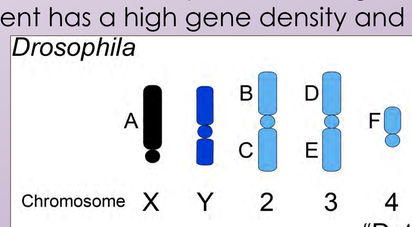


**Figure 1:** The chromosomes of *Drosophila*, including the F element (fourth chromosome)or the 'Dot' chromosome. From https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0141544.g001

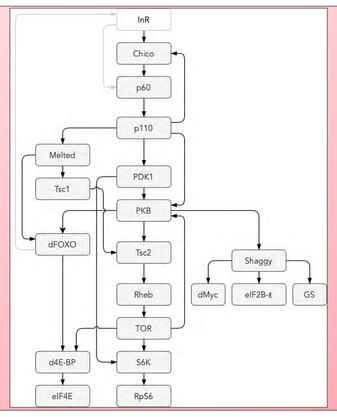


**Figure 2:** The insulin signaling pathway of *Drosophila melanogaster*. From https://thegep.org/pathways/

### PATHWAYS

The Pathways Project focuses on annotating genes in important biological pathways across different *Drosophila* species. This part of the project involved examining and annotating the gene, *Sdr*. The *secreted decoy of InR* (*Sdr*) is part of the insulin receptor signaling pathway.

## METHODS

Comparative annotations were produced, and gene models were constructed after evaluating different lines of evidence on the UCSC Genome Browser, such as RNA-Seq data, gene predictions, and repeats, and using other bioinformatic tools (e.g., NCBI BLAST). *D. melanogaster* served as a reference point when it came to comparing chromosomes, genes, or other genetic elements across Drosophila species because there is so much information known about it. The final gene models were checked using the GEP Genome Model Checker.

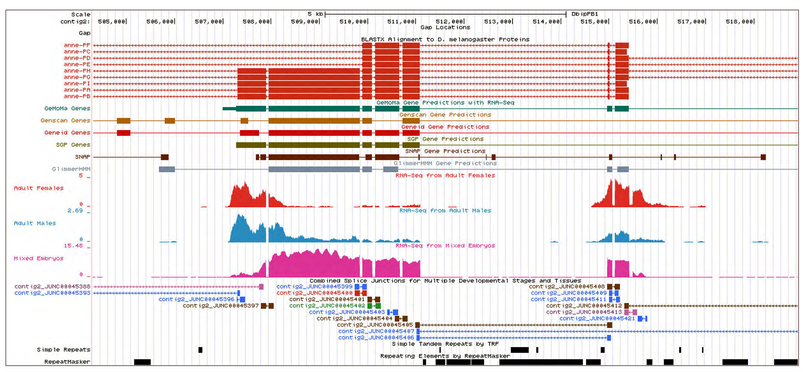


**Figure 3:** Examining the gene, *anne*, in *D.bipectinata* using the UCSC Genome Browser

**F element:** Annotations of the coding region exons (CDSs) of the genes in contigs 2, 7, and 11 of the F element of *Drosophila bipectinata*.

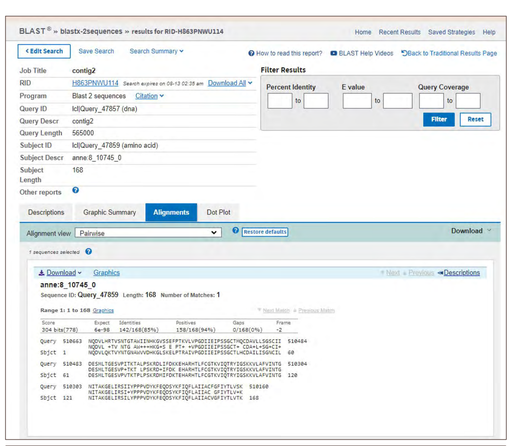


**Figure 4:** Using blastx and other BLAST tools to find the coordinates of the CDSs of a gene in the target species.

**Pathways:** Annotations of the coding regions and transcription start sites of *Sdr* in *D. ananassae, D. erecta, D. grimshawi, D. suzukii, D. yakuba, D. persimilis, D. mojavensis, D. virilis*, and *D. willistoni*.

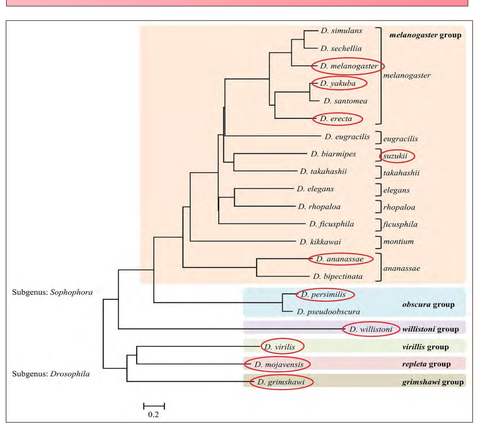


**Figure 5:** Phylogenetic tree depicting the hypothesized evolutionary relationships between the *Drosophila* species. The species annotated for the Pathways Project are highlighted in red. From https://peerj.com/articles/226/

## RESULTS AND INTERESTING FINDINGS

**Extra Exons and Introns**: Both genes annotated in contig 7 of *D. bipectinata*, *Ekar* and *Gat*, have an additional CDS compared to their *D. melanogaster* orthologs. These additional exons and introns were also found in the same genes in D. *ananassae*. Since these two species are more closely related to each other than to *D. melanogaster*, the gain of the extra CDSs might suggest that this change occurred in a common ancestor.

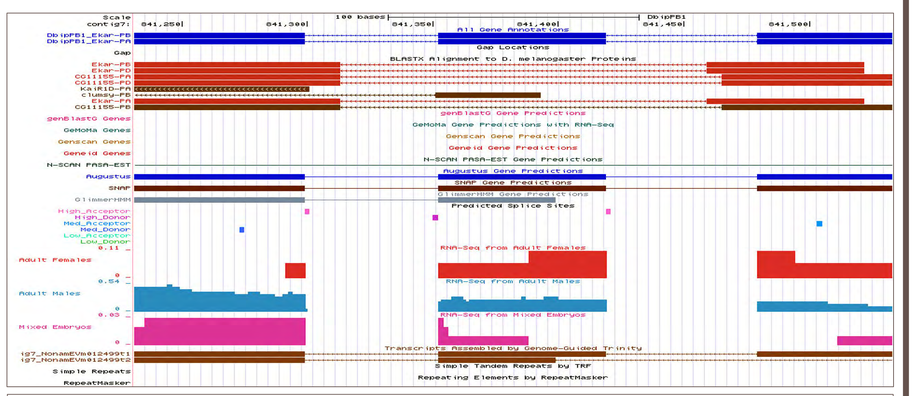


**Figure 6:** Genomic region and the accompanying evidence showing the position of the extra exon in *Ekar* in *D. bipectinata*.

| GENE | SPECIES | CODING SPAN/bp | TOTAL INTRON SIZE/ bp | TOTAL CDS SIZE/bp |
|---|---|---|---|---|
| ***Gat*** | *D. melanogaster* | 4,676 | 2,777 | 1,899 |
| | *D. bipectinata* | 85,704 | 83,793 | 1,911 |
| ***Ekar*** | *D. melanogaster* | 10,213 | 7,528 | 2,685 |
| | *D. bipectinata* | 6,46,586 | 643,946 | 2,640 |

**Table 1:** Table summarizing the differences in gene characteristics of the genes annotated in contig 7 of *D. bipectinata* and their *D. melanogaster* orthologs.

**Larger/Longer Genes:** The genes annotated in the contigs of *D. bipecintata* have much longer coding spans than their *D. melanogaster* orthologs. This is due to the presence of more repeat elements in the introns of the genes of *D. bipectinata* as described in the gene characteristics table (Table 1). This also suggests that the increased number of repeat elements might be a contributing factor to the overall expansion of the F element of D. *bipectinata*.

**Retrotransposed Psuedogene:** The multi-exon gene of *CG9922* in *D. melanogaster* has been copied as a single exon gene in contig 11 of *D. bipectinata*. This is supported by the results from blastx alignments of the *D. melanogaster* CDSs to contig 11 and several of the evidence tracks on the UCSC Genome Browser. Additionally, there are LTR retrotransposons of the same Gypsy family in the areas immediately before and after the "gene", suggesting that this is a retrotransposed pseudogene.

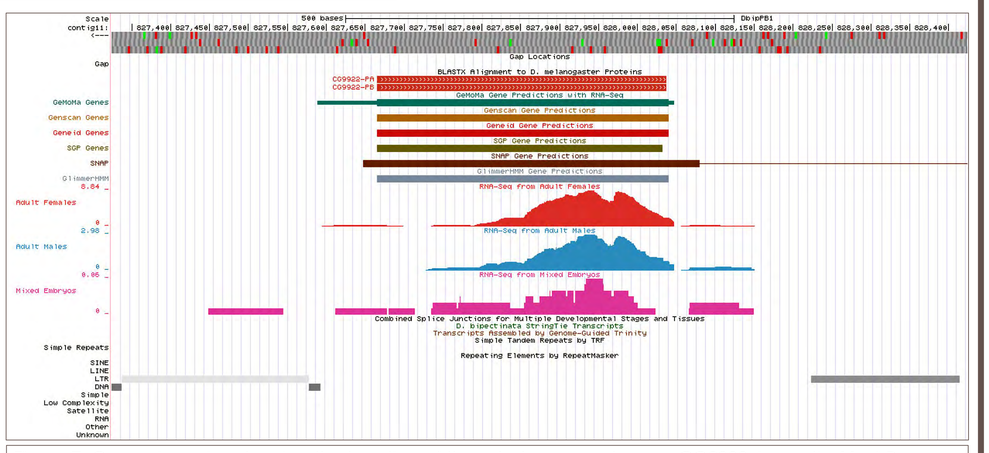


**Figure 7:** Genomic region showing the location of the retrotransposed gene, *CG9922*, in contig 11 of *D. bipectinata*, along with the RNA-Seq expression data as supporting evidence.

| SPECIES | MOST UPSTREAM | NEAREST UPSTREAM | TARGET GENE | NEAREST DOWNSTREAM | MOST DOWNSTREAM |
|---|---|---|---|---|---|
| ***D. melanogaster*** | *RpS5b* (-) | *VhaPPA1-1* (-) | *Sdr* (+) | *CG14861* (+) | *RpL10Aa* (+) |
| ***D. ananassae*** | *RpS5b* (-) | *VhaPPA1-1* (-) | *Sdr* (+) | *CG14861* (+) | *RpL10Aa* (+) |
| ***D. erecta*** | *RpS5b* (-) | *VhaPPA1-1* (-) | *Sdr* (+) | *CG14861* (+) | *RpL10Aa* (+) |
| ***D. grimshawi*** | *VhaPPA1-1* (+) | *Pkd2* (+) | *Sdr* (-) | *Tak1* (-) | *CG14861* (-) |
| ***D. mojavensis*** | *VhaPPA1-1* (-) | *Pkd2* (-) | *Sdr* (+) | *CG14861* (+) | *dpr9* (-) |
| ***D. persimilis*** | *VhaPPA1-1* (-) | *CG11407* (+) | *Sdr* (+) | *TfIIFa* (+) | *CG14861* (+) |
| ***D. suzukii*** | *RpS5b* (-) | *VhaPPA1-1* (-) | *Sdr* (+) | cannot be determined | cannot be determined |
| ***D. virilis*** | *VhaPPA1-A* (+) | *Pkd2* (+) | *Sdr* (-) | *CG14861* (-) | *Cyp28a5* (-) |
| ***D. willistoni*** | *Hsc70-4* (-) | *VhaPPA1-1* (-) | *Sdr* (+) | *CG14861* (+) | *dpr-9* (-) |
| ***D. yakuba*** | *RpS5b* (+) | *VhaPPA1-1* (+) | *Sdr* (-) | *CG14861* (-) | *RpL10Aa* (-) |

**Table 2:** Table showing synteny between *D. melanogaster* and each target species. The symbols (+/-) indicate the strand on which the gene is present.

**Conservation of Sdr**: The coding regions of *Sdr* were found to be well-conserved across *the Drosophila* species that were annotated in this study, and synteny was almost always maintained, especially in species that are more closely related to *D. melanogaster*. This level of conservation suggests that this gene has an essential function in this biological pathway.

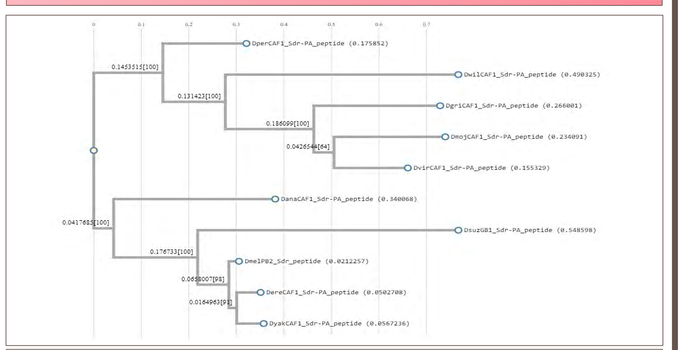


**Figure 8:** Phylogenetic tree generated for the protein sequences of *Sdr* in each target species.

## FUTURE DIRECTIONS

**F Element:** The F element of *D. bipectinata* shows substantial expansion, so the data from all these annotations is expected to facilitate comparative analyses that will help us to gain a more comprehensive understanding of the impact of evolution on chromosome and gene size. These datasets might also help us figure out why eukaryotic genomes become so large for mammals.

**Pathways:** It is anticipated that the comparative gene annotations of Sdr in different *Drosophila* species, in addition to network analysis methods, will be able to provide insight into the evolution and function of the important insulin signaling pathway.

## ACKNOWLEDGEMENTS

I would like to express my gratitude to my faculty mentor, Dr. Kennell, for her constant support, encouragement, and invaluable guidance throughout the project. I would also like to thank my fellow Student Fellows, Salome Ambokadze and Trina Chou, for their cooperation and helpfulness. This project has been possible because of the Genomics Education Partnership.