Where Are the Genes? Annotating a Genome Sequence from *Drosophila bipectinata*'s F Element

Franziska-Marie Ahrend and Christoph J. Hengartner

Barry University

Department of Biology
Barry University
11300 NE 2nd Ave.
Miami Shores, FL 33161
Phone: (305) 899 3744
chengartner@barry.edu

Barry University, Miami, FL

Figure 1. Drosophila melanogaster (fruit fly) [1]

Abstract

More than a century after Thomas Morgan used the fruit fly *Drosophila melanogaster* to demonstrate chromosomes contain the genetic material, Drosophila remains a vital research tool and model organism to investigate chromosome structure, genetics, development, and evolution. In 2000, its entire DNA was sequenced and ~60 percent of Drosophila genes are shared with humans. The Drosophila genome consists of four chromosomes that underwent rearrangements in some Drosophila species. According to the Muller nomenclature, each separate chromosome arm is labeled from A to F. The F element (aka dot chromosome) is usually the smallest chromosome; however, in some Drosophila species, the F element grew longer through expansion of repeated sequences. We have joined the Genomics Education Partnership (GEP), an association of over 150 research institutions that aims to integrate undergraduate students in original genetic and bioinformatics research projects. We join other GEP students and faculty in a crowd-sourcing approach to annotate the genes (coding region and transcription start site) in the F elements of *D. ananassae*, *D. bipectinata*, *D. kikkawai*, and *D. takahashii*. Our DNA annotation will use both experimental data (e.g. gene expression and conservation) and computational evidence (e.g. gene prediction algorithms) to identify gene elements such as exon and intron boundaries. The data of this comparative genome analysis will help us better understand the consequences of radical evolutionary changes in chromosome and gene structure. Studying the evolutionary forces that maintain and modify the chromosomes of these fruit flies may help us better understand how eukaryotic genomes grew so much larger than bacteria. For our project, we have chosen to examine a 650 kb region of *D. bipectinata*'s F element. We present here our methodology and approach for locating genes and critical gene elements within this DNA sequence, and we present any preliminary data we have accumulated.

Introduction

The tremendous 1000-fold expansion of chromosome size from bacteria to higher order organisms (such as humans) occurred over billions of years and the evolutionary forces that drove the expansion are still being investigated. The Genomics Education Partnership (GEP) is examining an example of chromosome expansion in Drosophila (fruit fly), a model organism that helped pioneer genetic studies over a century ago. In Figure 3, a phylogenetic tree shows how Drosophila species are related and indicates the structure of their chromosomes, including the dramatic expansion of *D. ananassae*'s and *D. bipectinata*'s F-Elements (in orange). In our research project, we are examining a 650kb DNA segment of *D. bipectinata*'s F-Element called contig 18. One essential part of the analysis is to find the coding regions within the genes in that segment. Drosophila's genes like most other eukaryotes, consist of coding (exons) and non-coding regions (introns). With splicing, introns in the RNA are later removed and the remaining exons are joined together to form the mRNA (Figure 2).

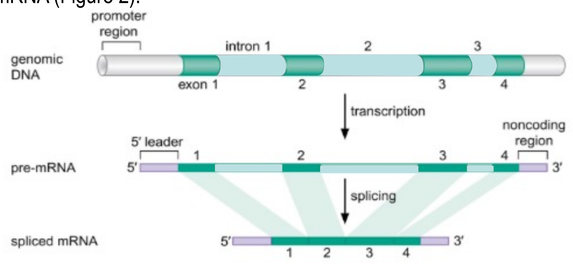

Figure 2. RNA splicing: removing introns (light green), attaching exons (dark green) [2]

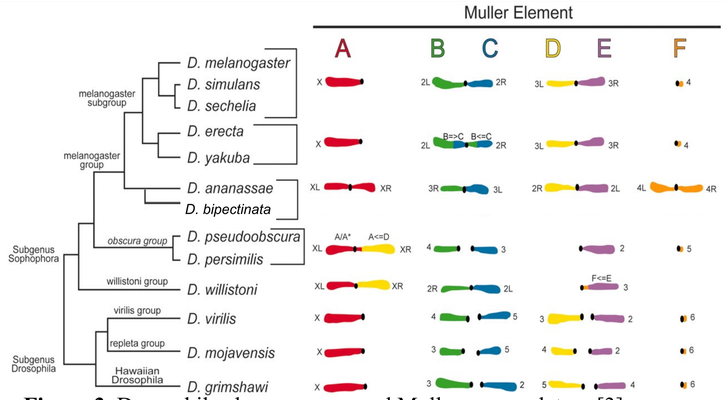

Figure 3. Drosophila chromosomes and Muller nomenclature [3]

Materials & Methods

GEP Annotation Workflow

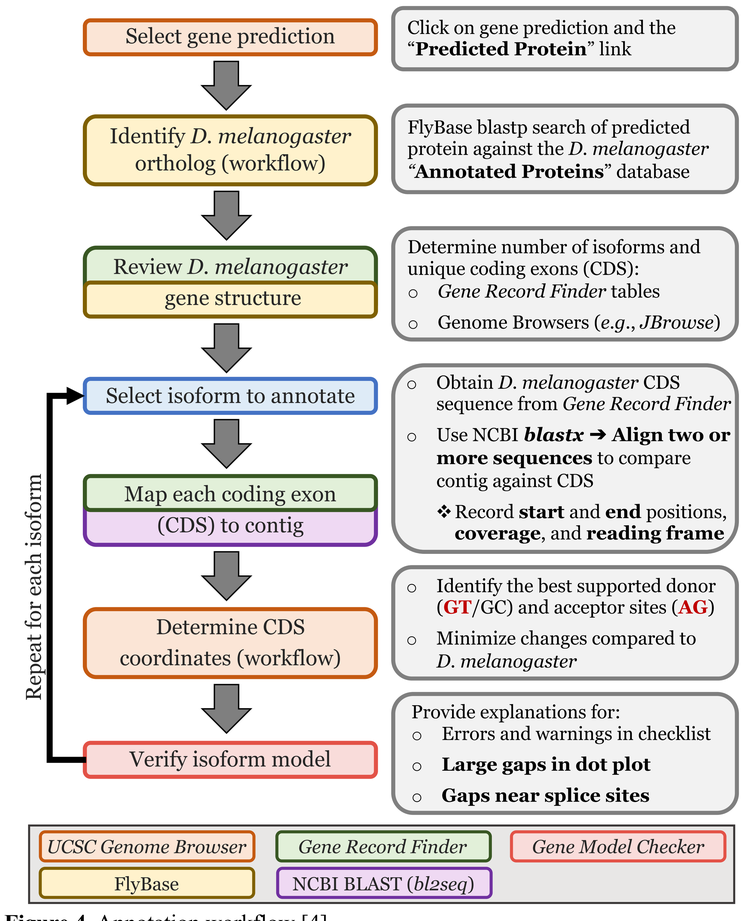

Figure 4. Annotation workflow [4]

Preliminary Results

The GEP association is providing students with contigs of different lengths and difficulties to annotate. The annotation starts with using the BLAST (Basic Local Alignment Search Tool) algorithm to identify Transcription Start Sites (TSS) in the *D. melanogaster* ortholog with the help of the Genome Browser by Washington University of St. Louis (https://gander.wustl.edu/cgi-bin/hgGateway) and can identify the contig's isoforms. Gene location predictions like Genscan or N-Scan can help us find gene exons and introns within the contig as visualized in Figure 5.

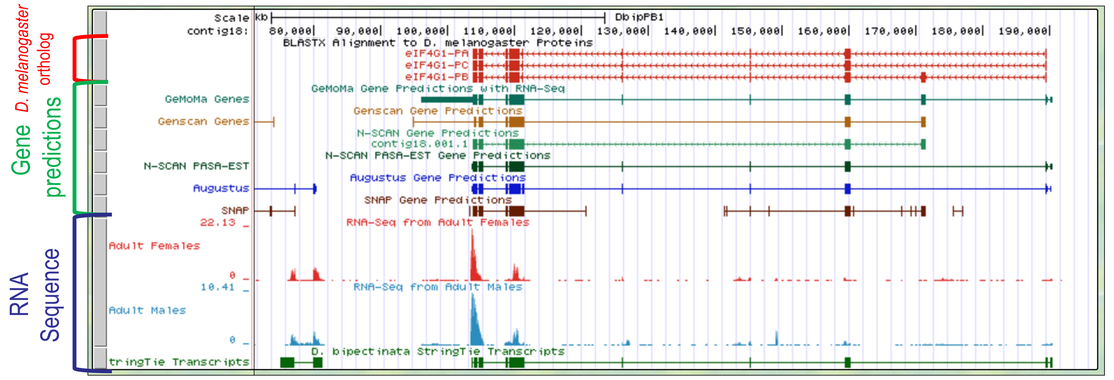

Figure 5. BLAST Alignment with *D. melanogaster* ortholog (red track), beneath various gene predictions

Those gene prediction cannot be used to assign the ortholog, however it gives an idea about the location of the gene and we can use the predicted protein sequence to BLAST it on the website http://flybase.org which results in many matches from which we choose those with a low e-value, as it makes the match statistically more significant. The gene record finder (https://thegep.org/felement/) shows isoforms and their number of coding exons. Furthermore, it is possible to approximate the individual locations of exons based on the knowledge of their splicing rules.

Current conclusion and future directions

We started working on the PA isoform of the eIF4G1 genes and have located the translation start site (beginning of CDS) and exon-intron boundaries of the first four exons. We have discovered that the D. bipectinate ortholog has undergone changes in the location of the splice sites. See example in Figure 6 below.

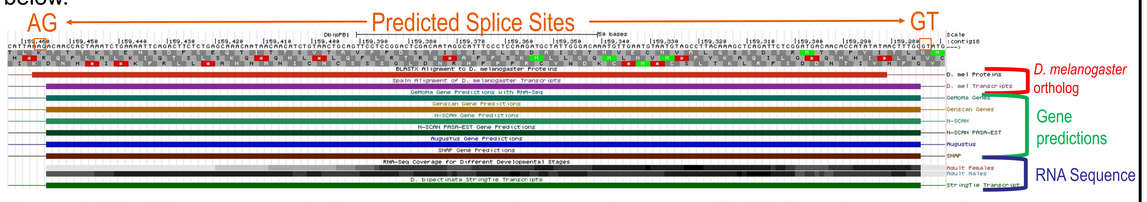

Figure 6. BLAST Alignment with *D. melanogaster* and gene predictions of eIF4G1-Exon 3 of isoform PA

References

[1] https://kxci.org/wp-content/uploads/2017/09/single-fruit-fly-drosophila-melanogaster-on-white-background-cropped-620x344.jpg
[2] https://o.quizlet.com/AYc3sm4lXU0lk9zNIhVpRQ.png
[3] Schaeffer SW *et al*, 2008. Polytene Chromosomal Maps of 11 *Drosophila* Species: The Order of Genomic Scaffolds Inferred From Genetic and Physical Maps. Genetics. 2008 Jul;179(3):1601-55
[4] Leung, W. (2014). Annotation of a Drosophila Gene. Genomics. *Genomics Education Partnership*. *https://thegep.org/lessons/wleung-walkthrough-annotation_drosophila_gene/*

Acknowledgements

We acknowledge the Genomics Education Partnership.