Volume 12

Washington University
Undergraduate Research Digest

Spring 2017

Reconciliation of Protein Coding Genes on the Drosophila elegans Muller F and D Elements

Jacob Cantrell
Washington University in St. Louis

Emily Chi
Washington University in St. Louis

Guanlan Dong
Washington University in St. Louis

Ben French
Washington University in St. Louis

Recommended Citation
Cantrell, Jacob; Chi, Emily; Dong, Guanlan; and French, Ben, "Reconciliation of Protein Coding Genes on the Drosophila elegans Muller F and D Elements" (2017). *Volume 12*. 24.
https://openscholarship.wustl.edu/wuurd_vol12/24

Reconciliation of Protein Coding Genes on the *Drosophila elegans* Muller F and D Elements

Jacob Cantrell, Emily Chi, Guanlan Dong,
Ben French, and Monica Perumattam

Mentors: Sarah C.R. Elgin and Christopher D. Shaffer

The Muller F element is an unusual domain within the *Drosophila melanogaster* genome because it is packaged mostly as heterochromatin, but also contains 78 protein-coding genes. To understand the factors that enable F element genes to be expressed within a heterochromatic environment, students participating in the Genomics Education Partnership (GEP) produced gene annotations for the F element and the base of the D element for multiple Drosophila species. Manual gene annotation is a meticulous process that requires students to construct gene models based on multiple potentially contradictory lines of evidence (e.g., sequence homology, gene predictions, and RNA-Seq data). For quality control purposes, each project was annotated by at least two independent groups. These annotations must be reconciled before they can be used in subsequent comparative analyses (e.g., gene and repeat characteristics, identification of conserved regulatory motifs). We used WebApollo to reconcile 856 gene models from the *Drosophila elegans* F element (~1.9Mb) and the base of the D element (~2.1Mb). We find that 58% of the gene models submitted by GEP students are in congruence with the reconciled gene models. The most common annotation errors made were the selection of incorrect splice sites (44%) and gene models with extra or missing exons (20%). As part of the reconciliation process, we also produced a report for each GEP faculty member that summarizes the mistakes and provides feedback so that faculty can improve their pedagogical approaches in subsequent years. We find that only 17% (21/127) of the reconciled genes in the *D. elegans* F element matched completely with a computer-based Genscan gene prediction. Hence the reconciled gene models will provide a more robust platform for the subsequent investigations into the unusual characteristics of the F element.