

Washington University in St. Louis

Washington University Open Scholarship

Volume 12

Washington University
Undergraduate Research Digest

Spring 2017

Annotation of Contig19 in *Drosophila eugracilis*

Hang Xue

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/wuurd_vol12

Recommended Citation

Xue, Hang, "Annotation of Contig19 in *Drosophila eugracilis*" (2017). *Volume 12*. 212.
https://openscholarship.wustl.edu/wuurd_vol12/212

This Abstracts S-Z is brought to you for free and open access by the Washington University Undergraduate Research Digest at Washington University Open Scholarship. It has been accepted for inclusion in Volume 12 by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

ANNOTATION OF CONTIG19 IN *DROSOPHILA EUGRACILIS*

Hang Xue

Mentor: Sarah Elgin

Heterochromatin domains are regions found in the genome where DNA is densely packaged. In most cases, genes in these regions are silenced, presumably due to limited access of the transcriptional machinery. The small fourth chromosome of *Drosophila*, also known as the dot chromosome or F element, is unusual in exhibiting many characteristics of heterochromatic domains, and at the same time maintaining some euchromatic properties. There are 80 genes actively transcribed on the F element. To examine the regulation of these genes and the evolution of this unusual domain, a comparative study is being undertaken by manually curating gene models and identifying motifs specific to the F element using phylogenetic footprinting for species diverged recently (10-15 Myr) from *D. melanogaster*.

This project focuses on contig19, a 40 kb region of the dot chromosome in *D. eugracilis*. The *D. melanogaster* genome has been used as a reference to annotate all features. A variety of bioinformatics tools and databases were used during annotation, including UCSC Genome Browser GEP Mirror, FlyBase, BLAST, ab initio gene predictors, TopHat, RNA-seq data, DHS data, RAMPAGE data, BG3 and S2 data on DH sites, and RNA PolII CHIP-seq data. Four genes on this contig, *Zyx*, *PlexA*, *ATPsynbeta*, and *CaMKII*, were annotated. Comparative analysis indicates that *PlexA*, *ATPsynbeta*, and *CaMKII* are highly conserved between *D. eugracilis* and *D. melanogaster* while *Zyx* is less conserved and might have a new isoform. The transcription start site, repetitive elements, synteny and potential regulatory elements of each gene are being examined in this project, looking for any special characteristics of genes within the F element. These data can help us understand how genes in densely packaged DNA can be expressed when typical genes are generally silenced in such a heterochromatic environment.