Project 1: Studying the Expansion of Muller F Element in Drosophila by Annotating the Coding Spans of its Genes

- The 4th chromosome, or the F element, exhibits unique properties in Drosophila melanogaster: it is packaged as heterochromatin (high repeat and methylation content, late replication), but exhibits high gene expression levels of euchromatin.
- The F element is known to have disproportionately expanded in at least 4 Drosophila species: D. melanogaster, D. bipectinata, D. kikkawai, D. takahashii.
- In collaboration with GEP/Genomics Education Partnership, we have annotated the coding spans in two genomic regions (contigs) of the F element: one in D. ananassae and one in D. bipectinata.
- Goal is to identify precise intron/exon boundaries of the genes producing coherent coding region models (along with their transcript and peptide files).
- Annotations gathered by GEP will provide insight into the evolutionary impacts of chromosome and gene expansion, as well as how genes function within heterochromatin.

Results:

Coding regions of Contig 2 in D. ananassae and Contig 12 in D. bipectinata have been fully annotated

Methods:

1. Inspecting genomic models:
   - Gene Model Checker.
2. Finding the orthology in known Drosophila genomes:
   - BLAST.
3. Mapping CDS against the genome:
   - CLUSTALW.
4. Realigning gene models:
   - GeneMark.

Additional introns have been found in D. ananassae Ekar and Gaf genes

Results:

There are two coding sequences (Fig. 2c-d) within the D. ananassae Gaf gene in a single 79 kb CDS that produce a protein sequence with high homology to D. melanogaster Gaf1 (207 aa) and D. yakuba Gaf2 (201 aa). Both protein sequences are absent from D. melanogaster.

There are two coding sequences (Fig. 2c-d) within the D. ananassae Gaf gene in a single 79 kb CDS that produce a protein sequence with high homology to D. melanogaster Gaf1 (207 aa) and D. yakuba Gaf2 (201 aa). Both protein sequences are absent from D. melanogaster.

Conclusion: Pathways Project: Project conducted by GEP, aims to better understand the evolution of Drosophila genes in the context of their role within a biological pathway.

Methods:

1. GEP genome browser view of Tsc1 genomic neighborhood, chromosome 3R. Narrow bands are UTRs. Thick bands are CDS.
2. Full TSS and transcript annotations of Drosophila species: Synteny and gene structure are highly conserved

References:


Acknowledgements:

I would like to express my thanks to Professor Kennell for her support and guidance throughout this project. Despite being fully remote, she managed to bring in enthusiasm and sequence of team work meetings, which I would like to express my gratitude to Ms. Flora, Mr. A. Davis for funding this research opportunity and allowing me to pursue my interest in scientific investigation.

Project 2: Evolutionary Analysis of Drosophila Gene Tsc1 in the Context of Insulin-Signaling Pathway

- The Pathways Project, conducted by GEP, aims to better understand the evolution of Drosophila genes in the context of their role within a biological pathway.
- Our focus is the insulin-signaling pathway of Drosophila, which is highly conserved among species and crucial to growth, and metabolic homeostasis.
- This study focuses on structural variations of a single gene within this pathway, Tsc1, across 9 different species.
- Entire transcript spans (including the Transcription Start Sites, TSS) of the Tsc1 gene have been annotated in this project.
- Annotations gathered by GEP will provide insight into the evolutionary regulation of the regulatory region in the context of their position in a larger biological network.

Methods:

1. GFP genome browser view of Tsc1 genomic neighborhood, chromosome 3R. Narrow bands are UTRs. Thick bands are CDS.
2. Full TSS and transcript annotations of Drosophila species: Synteny and gene structure are highly conserved

Results:

- Syntaxin has been shown to be fully conserved in 8 of the 9 annotated species. Blast search against D.mel protein has found matches to Sec10 and Npc2p downstream, and GafB and Root upstream of Tsc1 ortholog (D. suz, shows ortholog as Tsc1). Conservation is likely driven by expression data and have been fine-mapped. All species show evidence supported TSS and 5' UTR as well.
- Phylogenetic tree based on the translated protein sequence of Tsc1 across 9 species has been produced by CLUSTALW. evolutionary relationship based on the protein region. Conservation reflects what has been hypothesized by GEP.

Conclusion: Pathways Project: Project conducted by GEP, aims to better understand the evolution of Drosophila genes in the context of their role within a biological pathway.

Methods:

1. GEP genome browser view of Tsc1 genomic neighborhood, chromosome 3R. Narrow bands are UTRs. Thick bands are CDS.
2. Full TSS and transcript annotations of Drosophila species: Synteny and gene structure are highly conserved

Results:

- Syntaxin has been shown to be fully conserved in 8 of the 9 annotated species. Blast search against D.mel protein has found matches to Sec10 and Npc2p downstream, and GafB and Root upstream of Tsc1 ortholog (D. suz, shows ortholog as Tsc1). Conservation is likely driven by expression data and have been fine-mapped. All species show evidence supported TSS and 5' UTR as well.
- Phylogenetic tree based on the translated protein sequence of Tsc1 across 9 species has been produced by CLUSTALW. evolutionary relationship based on the protein region. Conservation reflects what has been hypothesized by GEP.

Conclusion: Pathways Project: Project conducted by GEP, aims to better understand the evolution of Drosophila genes in the context of their role within a biological pathway.

Methods:

1. GEP genome browser view of Tsc1 genomic neighborhood, chromosome 3R. Narrow bands are UTRs. Thick bands are CDS.
2. Full TSS and transcript annotations of Drosophila species: Synteny and gene structure are highly conserved

Results:

- Syntaxin has been shown to be fully conserved in 8 of the 9 annotated species. Blast search against D.mel protein has found matches to Sec10 and Npc2p downstream, and GafB and Root upstream of Tsc1 ortholog (D. suz, shows ortholog as Tsc1). Conservation is likely driven by expression data and have been fine-mapped. All species show evidence supported TSS and 5' UTR as well.
- Phylogenetic tree based on the translated protein sequence of Tsc1 across 9 species has been produced by CLUSTALW. evolutionary relationship based on the protein region. Conservation reflects what has been hypothesized by GEP.

Conclusion: Pathways Project: Project conducted by GEP, aims to better understand the evolution of Drosophila genes in the context of their role within a biological pathway.

Methods:

1. GEP genome browser view of Tsc1 genomic neighborhood, chromosome 3R. Narrow bands are UTRs. Thick bands are CDS.
2. Full TSS and transcript annotations of Drosophila species: Synteny and gene structure are highly conserved

Results:

- Syntaxin has been shown to be fully conserved in 8 of the 9 annotated species. Blast search against D.mel protein has found matches to Sec10 and Npc2p downstream, and GafB and Root upstream of Tsc1 ortholog (D. suz, shows ortholog as Tsc1). Conservation is likely driven by expression data and have been fine-mapped. All species show evidence supported TSS and 5' UTR as well.
- Phylogenetic tree based on the translated protein sequence of Tsc1 across 9 species has been produced by CLUSTALW. evolutionary relationship based on the protein region. Conservation reflects what has been hypothesized by GEP.