Evolutionary Path of a Non-canonical Start Codon in the *Drosophila Fmr1* Gene

Meagan Hood and Dr. Jack Vincent
Division of Sciences & Mathematics, University of Washington-Tacoma

Genomics Education Partnership

INTRODUCTION

Non-canonical start codons are understood to be an alternative mechanism for gene expression, allowing for the diversification of protein isoforms.

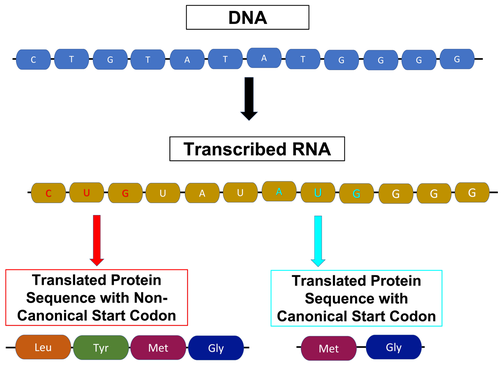


Figure 1. Translation starting at non-canonical and standard start codons can alter the coded amino acid sequence of a gene

The *Fmr1* gene encodes a protein, FMRP, known to play a role in behavior and the development of neural anatomy and has a well studied ortholog in the model species *Drosophila melanogaster*. Previous research has found the *Fmr1* gene in *Drosophila melanogaster* utilizes non-canonical start codons in a small subset of larger isoforms (Beerman and Jongens 2011). Closely related species to *Drosophila melanogaster* have also been identified as using this alternative mechanism for gene translation, raising questions about the evolutionary history of this non-canonical start codon in the *Drosophila* genus (Beerman and Jongens 2011). Our project aimed to further the understanding of the evolutionary path of the non-canonical start site used in the translation of the *Fmr1* gene in the *Drosophila* genus through annotation of the *Fmr1* orthologs in those species.

METHODS

- We annotated the *Fmr1* ortholog in *D. annassae, D. eugracilis, D. busckii,* and *D. virilis,* utilizing RNA seq data, splice junction data, and localized homology searching (Figure 2).
- We used the same evidence tracks to search for evidence of the non-canonical start site in the PG and PH isoforms of the *Fmr1* gene in these four species as well as previously annotated orthologs in ten other species.

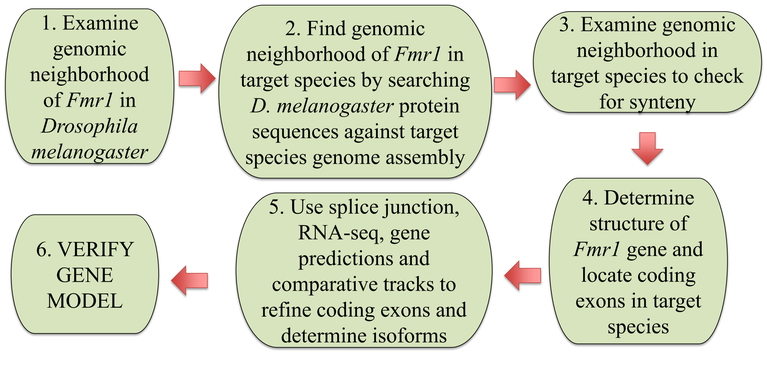


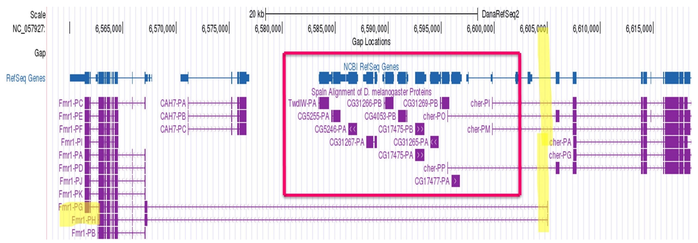
Figure 2. Flowchart detailing workflow of gene annotations through the Pathways Project by the Genomics Education Partnership (GEP).

RESULTS

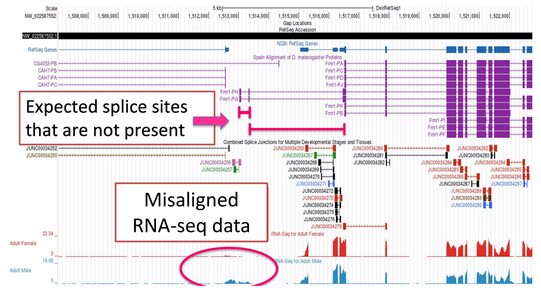
To find evidence for *Fmr1* isoforms that utilize non-canonical start codons, we utilized Spaln and BLAT alignments based on *D. melanogaster Fmr1* to identify potential homology in protein coding sequence upstream of the standard start codon. *If* this local homology was present, we then…

1. Examined transcription of those regions of homology by comparing to RNA seq data
2. Examined other positional evidence (presence of in-frame CTG sequence upstream of this homology, predicted splice junctions, presence of nested genes) that would indicate if expression of these regions occurs

(a) Drosophila ananassae Fmr1: Too many nested genes disrupting homology upstream of standard start codon



(b) Drosophila virilis Fmr1: Lacks splice site or RNA seq data supporting expression of homology upstream of standard start codon



(c) Drosophila ficusphila: CUG start codon supported by high quality RNA seq data

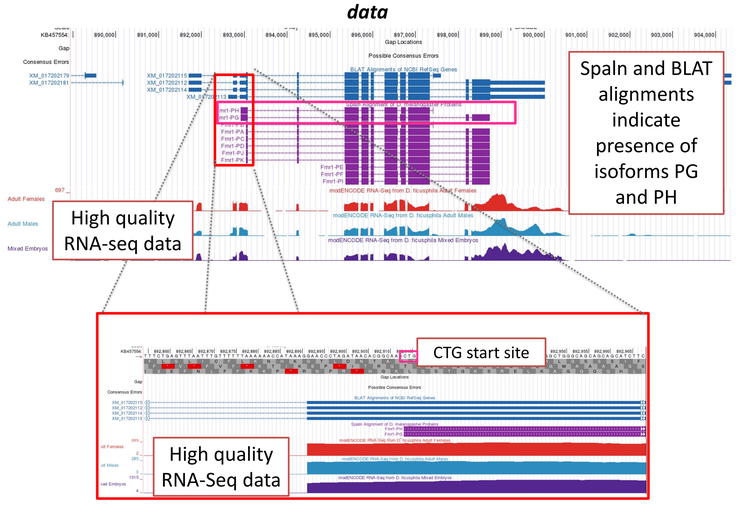


Figure 3. Evidence tracks viewed within a genome browser that either contradict (a, b) or support (c) the usage of a non-canonical start codon within isoforms of *Fmr1* orthologs. RNA seq data, splice junction data, and prediction tracks were all used as evidence for the determination of non-AUG start codon usage.

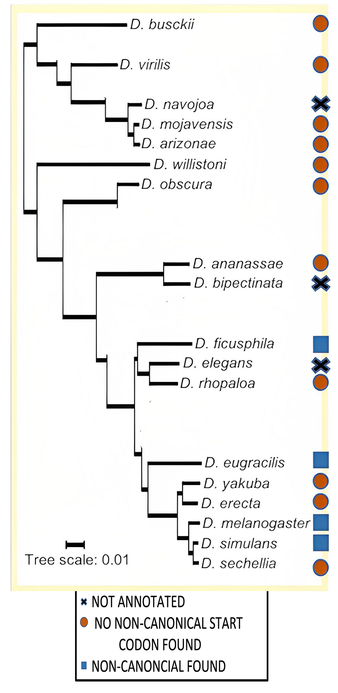


Figure 4. Phylogenetic tree of *Drosophila* species with annotated *Fmr1* ortholog indicating evidence supportive of non-canonical start codon presence.

- After annotating four species within the *Drosophila* genus, no evidence for a non-canonical start codon was present in *D. virilis, D. busckii, D. ananassae.* One species, *D. eugracilis* showed conservation of the stretch of amino acids before the standard start codon as seen by the SPALN alignments and high-quality RNA-Seq data consistent with a non-canonical start codon within the PG and PH isoforms.
- The investigation into previous annotations indicated the utilization of the non-canonical start codon within two other species, *D. ficusphila* and *D. simulans*.

CONCLUSIONS

- Our results show that expression from non-canonical start codon in *Fmr1* is more likely conserved in species more closely related to *D. melanogaster* and is less likely conserved in distantly related species, indicating its likely recent evolution as a gene expression mechanism.
- Our results raise additional questions about translation initiation of, and the diversification of, FMRP within eukaryotic genomes. Continued annotation will be supportive of any mechanistic studies into translation initiation of the *Fmr1* gene.

ACKNOWLEDGMENTS

This project is associated with the Genomics Education Partnership (GEP) Pathways Project, designed to train undergraduates in gene annotation then annotate genes the Insulin Signaling Pathway in *Drosophila* species.

Special thanks to the UWT Biomedical Science students who completed previous annotations on the *Fmr1* gene. Your work inspired us to look further into this topic.

BIBLIOGRAPHY

- Beerman RW, Jongens TA. 2011. A non-canonical start codon in the Drosophila fragile X gene yields two functional isoforms. Neuroscience. 181:48–66.
- Genomics Education Partnership (GEP). Pathways Project: Annotation Workflow. https://thegep.org/lessons/pathways-project-annotation-workflow/.