UNIVERSITY of WASHINGTON TACOMA

Gene Annotation and Evolution of the PolyQ Regions in Fmr1 in Drosophila Species

Alexis Heath and Dr. Jack Vincent

INTRODUCTION

Expansion of Polyglutamine (PolyQ) tracts in the *Fmr1* gene are associated with Fragile X syndrome (Cohen-Carmon and Meshorer 2012). The Fmr1 protein is critical for normal brain development in metazoans. Expansions of the Polyglutamine tract within this protein results in structural instability (Totzeck et al. 2017). It is hypothesized that this structural instability will create conformational changes within the protein and affect its function (Robertson et al. 2008). When the instability of the structure reaches the threshold, this results in neurodegenerative diseases, such as Fragile X Syndrome. PolyQ related disorders affect up to 700,000 individuals worldwide (Fan et al. 2014). The severity of these diseases is dependent on the length of the PolyQ stretches in the protein, the longer the stretch, the more severe (Totzeck et al. 2017). Research also suggests that the location of these PolyQ stretches is also important in determining disease severity (Robertson et al. 2008).

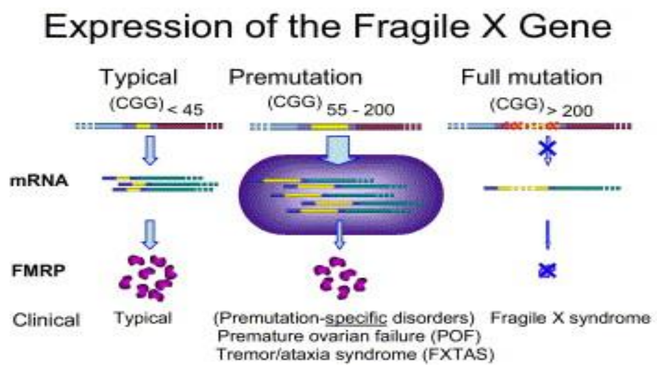

Figure 1. A depiction of how the mutations in the *Fmr1* gene can lead to disorders (Wittenberger et al. 2007). CGG is another sequence of nucleotides that can lead to a glutamine amino acid, when the Q tracts have upwards of 55 Q amino acids in a single tract, this generally always leads to mutation.

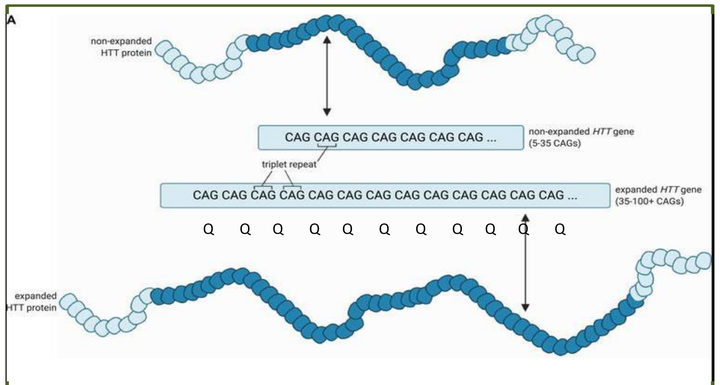

Figure 2. A deeper visualization of the PolyQ tracts within genes (Donaldson et al. 2021). Each CAG repeat corresponds to one glutamine amino acid, which has the symbol "Q." Hence the name, PolyQ tract.

OBJECTIVES

We aimed to determine whether PolyQ tracts because they are encoded by repetitive DNA, are hypervariable between species when looking at the *Fmr1* gene. This study aims to determine the evolution of Polyglutamine (PolyQ) tracts within the *Fmr1* gene within *D. sechellia, D. rhopaloa*, *and D. willistoni* through comparison and annotation. These three species were chosen because they are in different branches of the Drosophila species, at varying degrees of separation in their evolutionary tree (Figure 3), providing adequate information on the evolution of PolyQ tracts in relation to the model organism *D. melanogaster.*

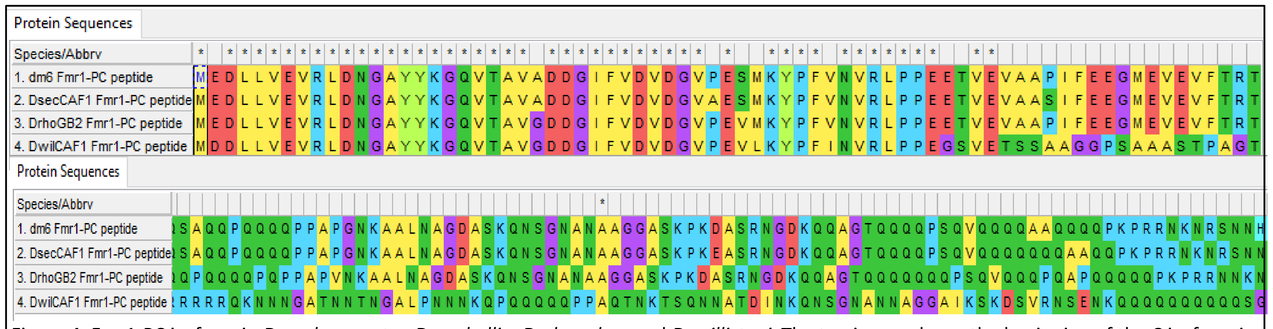

Figure 4. Fmr1-PC isoform in *D. melanogaster, D. sechellia, D. rhopaloa,* and *D. willistoni.* The top image shows the beginning of the C isoform in all species, this shows high conservation across all species. The bottom shows the last third of the C isoform, many PolyQ tracts exist here, and the variability is significant.

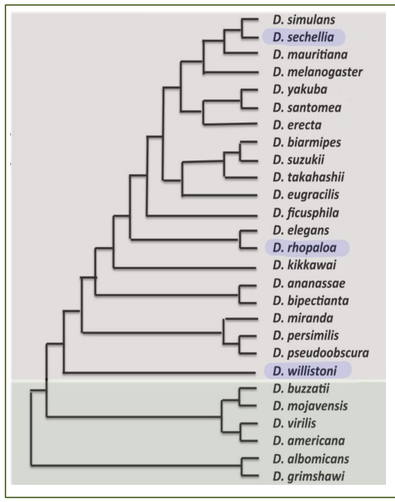

Figure 3. Image of a phylogenetic tree of Drosophila species (Markow 2015).

For this project, we will be focusing on *D. sechellia, D. rhopaloa,* and *D. willistoni*, which are highlighted in blue. As can be seen, they are at varying intervals of separation.

These varying intervals of separation are critical to observe differences in sequence.

Species, Isoform	Amino acid differences per site between sequences
D. melanogaster, Isoform C	
D. sechellia, Isoform C	0.0731
D. rhopaloa, Isoform C	0.3070
D. willistoni, Isoform C	2.0018

Table 1. Data from my comparative analysis of the Fmr1-PC sequences using MEGA-X. This allows us to determine the genetic distance (and similarities) between each species to melanogaster quantitatively. As can be observed, sechellia's sequence is minimally different to melanogaster at 0.0731 amino acid differences per site, and willistoni's sequence is significantly different than melanogaster's at 2.0018 amino acid differences per site.

METHODS AND MATERIALS

- The Genomics Education Partnership (GEP) pathways annotation walkthrough was used to annotate the *Fmr1* orthologs for *D. sechellia*, *D. rhopaloa*, and *D. willistoni*.
- Analyze the genomic neighborhood of *Fmr1* using the GEP UCSC Genome Browser.
- Identify *Fmr1* in *D. melanogaster* to provide the coding exons, isoforms, and UTR's, using the Gene Record Finder,.
- Using each of the orthologs' coding-exon sequence information in the model organism, a tBLASTn alignment search was used to determine the location of the orthologs' coding exon sequences and the subject frame in the target species' *Fmr1* gene scaffold.
    - These species are related through evolution, so they'll likely have similar exon and intron structures; finding where they are located is important to determine the splice site locations.
- These coordinates were refined using the GEP UCSC Genome Browser.
- Using the refined coordinates, the gene can then be verified for accuracy using the GEP Gene Model Checker.
- Using the sequence alignments for the orthologs from the target species' *Fmr1* gene, a comparative analysis will be performed using the Molecular Evolutionary Genetic Analysis (MEGA) program.

RESULTS/FURTHER DIRECTION

Successful annotation of Fmr1 in *D. sechellia, D. rhopaloa*, and *D. willistoni*.

Comparative analysis results:

- Last third of the sequence in all isoforms are most variable.
- PolyQ tracts are hypervariable between all species.
- Most variability found in *D. willistoni.*

Implications/Future Direction:

- PolyQ tract length polymorphisms present most commonly in *D. willistoni.*
- Supports the implications of PolyQ impacts on Fmr1

ACKNOWLEDGEMENTS

This project is connected to the Genomic Education Partnership (GEP), which conducts multiple projects relating to Drosophila and gene pathways.

Special thanks to my mentor Dr. Jack Vincent who diligently and thoughtfully directed this capstone project; and who created a wonderful environment to research and study.

REFERENCES

Cohen-Carmon D, Meshorer E. 2012. Polyglutamine (polyQ) disorders: the chromatin connection. Nucleus. 3(5):433-41. DOI: 10.4161/nucl.21481.

Donaldson J, Powell S, Rickards N, Holmans P, Jones L. 2021. What Is the Pathogenic CAG Expansion Length in Huntington's Disease? Journal of Huntington's Disease. 10(1): 175–202

Fan HC, Ho LI, Chi CS, Chen SJ, Peng GS, Chan TM, Lin SZ, and Harn HJ. 2014. Polyglutamine (PolyQ) Diseases: Genetics to Treatments. Cell Transplantation. 23(4-5):441–58.

[HDSA] Huntington's Disease Society of America. c2021. Overview of Huntington's Disease [Internet]; [cited 2021 Aug 06]. Available from https://hdsa.org/what-is-hd/overview-of-huntingtons-disease/

Markow, TA. 2015. The Secret Lives of Drosophila Flies. ELife. 4(1): e06793. https://doi.org/10.7554/eLife.06793.

Robertson AL, Horne J, Ellisdon AM, Thomas B, Scanlon MJ, Bottomley SP. 2008. The structural impact of a polyglutamine tract is location-dependent. Biophys J. 95(12):5922-30. DOI: 10.1529/biophysj.108.138487.

Totzeck F, Andrade-Navarro MA, Mier P. 2017. The Protein Structure Context of PolyQ Regions. PLoS One. 12(1):e0170801. DOI: 10.1371/journal.pone.0170801.

Wittenberger MD, Hagerman RJ, Sherman SL, McConkie-Rosell A, Welt CK, Rebar RW, Corrigan EC, Simpson JL, Nelson LM. 2007. The FMR1 Premutation and Reproduction. Fertility and Sterility. 87(3): 456–65.