Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

6-15-2021 2:00 PM

Fruitless (fru) and the Neural System of Rejection Behaviours in Female Drosophila melanogaster

Omar El-Deeb, *The University of Western Ontario*

Supervisor: Moehring, Amanda J., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Neuroscience

© Omar El-Deeb 2021

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Recommended Citation

El-Deeb, Omar, "Fruitless (fru) and the Neural System of Rejection Behaviours in Female Drosophila melanogaster" (2021). *Electronic Thesis and Dissertation Repository*. 7906.
https://ir.lib.uwo.ca/etd/7906

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

Among sexually reproducing organisms, species are separated by their inability to reproduce with each other and form viable, fertile offspring. One type of reproductive barrier is behavioural, whereby mating is not initiated between two species. In the model organism *Drosophila melanogaster*, a gene called *fruitless* regulates heterospecific courtship and rejection behaviours, however, its neural mechanisms of action remain unknown in females. In this work, I identified which *fruitless* splice variants are expressed in adult males and females, and created the plasmid vectors necessary to implement a genetic system for the purposes of identifying, silencing, and hyperactivating neuronal subsets that express distinct splice-variants of *fruitless.* These plasmid vectors can be used in the future to induce expression of exogenous genes of interest for further experimentation in *D. melanogaster*.

Key Words

Behavioural isolation, *Drosophila melanogaster*, female rejection, *fruitless*, RT-PCR, splice variants, PhiC31 recombination, CRISPR/Cas9, Trojan-Gal4, Gal4-UAS,

Summary for lay audience

Speciation, the process by which populations diverge to form separate species, is initiated and reinforced by the emergence of barriers that inhibit reproductive success between populations that have sexual reproduction. These barriers are often physical abnormalities, such as genetic mutations that create offspring with defects that make them inviable or infertile. However, sometimes populations can also accumulate behavioural differences, and become separate species because they are no longer attracted to each other as viable mates.

The mechanisms behind behavioural barriers are less understood than their physical counterparts. However, in the model organism *Drosophila melanogaster*, a gene called *fruitless* (*fru*) influences behavioural isolation with its sister species. This gene is transcribed into many different mRNA products, with functions largely determined by the first and last sections of each transcript. In female *D. melanogaster*, the less-studied sex, expression patterns of each transcript were lesser known, and therefore transcripts that potentially affect female rejection behaviours could not be identified. I identified each *fru* transcript and profiled its expression in adult male and female *D. melanogaster*. I also identified the presence of a new non-sex-specific transcript that has not previously been described in the literature.

I then developed a series of DNA constructs that can be microinjected into *Drosophila* embryos for the integration of a Trojan-Gal4 system at specific locations in the *fru* gene. The Trojan-Gal4 system can be paired with other genetic tools to drive expression of desired genes in the same cells as the transcript of interest. The desired genes can be used to visualize, hyperactivate, and silence the neurons that express them, allowing for in-depth analysis of a transcript's functions within the central nervous system, and on mating behaviours.

Acknowledgements

To my supervisor, Dr. Amanda Moehring, thank you for your patience, guidance, and insatiable passion for knowledge. You are a wonderful role model who constantly leaves me in awe of the things you manage to do.

To my advisors and program representatives, Dr. Anne Simon, Dr. Elizabeth MacDougall-Shackleton, Dr. Robert Cumming, Dr. Brian Corneil, and Dr. Sean Cregan; normally this would only be a list of three people, but life is definitely full of surprises, and I was blessed with a team of superheroes to rival The Avengers. Thank you for your kindness, encouragement, and insight.

Also, another special thank you goes to Dr. Robert Cumming, who acted as my interim supervisor for several months. Your leadership was truly comforting in a time of uncertainty.

To my examiners, Dr. Scott MacDougall-Shackleton, Dr. Nathalie Bérubé, and Dr. Tony Percival-Smith; thank you for challenging me to be a better scientist and helping me create an informative and compelling story.

To my labmates, Will Yeung, Joshua Isaacson, Brendan Charles, Sebastian Heine, Lucas Khodaei, Alannah Mattice, Tabashir Chowdhury, Olivia Siemons, Sara Kulinski, Jessie Jagiello, Mahsa Farmanbar, Sofia Savona-Jofre, and Pranjan Gandhi, it is a miracle that all thirteen of you are better friends than scientists, because you are all wonderful scientists. Thank you for making the long days feel shorter, and the short days feel fuller.

To the rest of my friends, who kept me smiling through it all; thank you for the memories I will cherish until the last image of a microinjection falls out of my mind.

Finally, to my family, who have unfailingly supported me in all of my goals. I'm at a loss for explaining how much I appreciate you. Thank you for being the wind at my back during my uphill struggles, and my biggest fans at the peak.

Table of Contents

Abstract ii

Summary for lay audience iii

Acknowledgements iv

Table of Contents v

List of Appendices vii

List of Figures viii

List of Abbreviations ix

1.1 Speciation 1

1.1.1 Mechanisms of Speciation 1

1.1.2 Genetic influences on mating barriers 3

1.2 *Drosophila* as models of mating barriers and speciation 3

1.2.1 Sexual selection in *Drosophila* 4

1.2.1.1 The male courtship ritual 5

1.2.1.2 Female behaviours of selection and receptivity 6

1.2.1.3 Heterospecific courtship and rejection 6

1.2.2 Candidate genes influencing female receptivity 8

1.2.2.1 Expression of *fruitless* in *Drosophila melanogaster* females 11

1.3.1 Gal4-UAS 12

1.3.1.1 Trojan-Gal4-UAS 13

1.4 Genetic tools for assessing transcripts affecting female receptivity 14

1.4.1.1 PhiC31-mediated recombination 14

1.4.1.2 CRISPR-Cas9/Homology-Directed Repair 15

1.5 Experimental objectives 17

2 Methods 18

2.1 Fly husbandry and stocks 18

2.2 Assessing *fru* transcript expression using RT-PCR 19

2.3 Plasmid preparation for microinjection 21

2.3.1 Plasmids for phiC31-mediated Trojan-Gal4 integration 5' of *fru* A and B exons 21

2.3.2 Plasmids for Trojan-Gal4 integration 5' of *fru* C exon via CRISPR/CAS9-mediated Homology-Directed Repair 22

2.3.2.1 Creating a suitable CRISPR gRNA plasmid vector 22

2.3.2.2 Creating a suitable Trojan-Gal4 Donor Vector 23

2.4 Embryonic microinjections 26
2.4.1 Microinjection preparations 26
2.4.2 Embryonic microinjection protocol 26
2.5 Crossing of survivors and screening of offspring for transgenesis 27
2.5.1 PhiC31-mediated recombination upstream of *fru* A and B exons 27
2.5.2 CRISPR-mediated HDR upstream of *fru* C exon 27
3. Results 28
3.1 *fru* expression in adult male and female *Drosophila melanogaster* 28
3.1.1 RT-PCR analysis of *fru* expression in adult male and female *Drosophila melanogaster* 28
3.1.2 Sequencing data of visualized cDNA bands 30
3.2 Microinjections for Trojan-Gal4 integration via PhiC31-mediated recombination 32
3.2.1 Efficiency of transgenesis upstream of the A and microinjection survivorship 32
3.2.2 Efficiency of transgenesis upstream of the B exon and microinjection survivorship 33
3.3 Microinjections of CRISPR/Cas9-mediated HDR 35
3.3.1 Preparation of plasmids used for CRISPR/Cas9 injections 35
3.3.2 CRISPR/Cas9-mediated Homology-Directed Repair survival rate and efficiency 38
3.3.2.1 Efficiency of transgenesis 38
3.3.2.2 Survival rate following microinjections 38
4. Discussion 39
4.1 Exploration of *fru* expression in male and female *Drosophila melanogaster* 39
4.2 Troubleshooting PhiC31-mediated recombination 43
4.3 Optimizing CRISPR/Cas9-mediated HDR 44
4.4 Future directions 45
5. Conclusions 46
Works Cited 47
Appendix: Crossing Scheme for PhiC31-mediated integration of Trojan Gal4 into MiMIC sites 56
Curriculum Vitae 57

List of Appendices

Appendix 1: Crossing Scheme for PhiC31-mediated integration of Trojan-Gal4 into MiMIC sites

List of Figures

Figure 1: Phylogenetic tree of the *Drosophila melanogaster* species subgroup 4
Figure 2: Schematic showing the possible mRNA transcripts derived from the *fruitless* gene 10
Figure 3: Mechanism of action of the traditional Gal4-UAS system 13
Figure 4: Schematic of PhiC31-mediated recombination of donor DNA into genomic MiMIC sites 15
Figure 5: Schematic of CRISPR-mediated HDR to integrate donor DNA into a genomic double-stranded DNA break 16
Figure 6: Annealing positions of primers used for *fru* RT-PCR 20
Figure 7: Sex-specific RT-PCR expression analysis of *fru* P1 and P2 transcripts in adult male and female *Drosophila melanogaster* 29
Figure 8: Sex-specific RT-PCR expression analysis of *fru* P1-S to P5 transcripts in adult male and female *Drosophila melanogaster* 30
Figure 9: cDNA sequence and amino acid analysis of secondary *fruitless* P2 transcript 31
Figure 10: Number of adult survivors following embryonic microinjections for PhiC31 integration of Trojan-Gal4 upstream of the *fru* A exon 32
Figure 11: Percent survival of embryonic microinjections for PhiC31 integration of Trojan-Gal4 upstream of the *fru* A exon 33
Figure 12: Number of adult survivors following embryonic microinjections for PhiC31 integration of Trojan-Gal4 upstream of the *fru* B exon 34
Figure 13: Percent survival of embryonic microinjections for PhiC31 integration of Trojan-Gal4 upstream of the *fru* A exon 35
Figure 14: Confirmation of gRNA sequence insertion into gRNA plasmid vector for CRISPR/Cas9-HDR upstream of *fru* C exon 36
Figure 15: Confirmation of homology arm insertion into Trojan-Gal4 donor vector for CRISPR/Cas9-HDR upstream of *fru* C exon 37
Figure 16: Number of adult survivors following embryonic microinjections for CRISPR/Cas9-HDR integration of Trojan-Gal4 upstream of the *fru* C exon 38
Figure 17: Percent survival of embryonic microinjections for CRISPR/Cas9-HDR integration of Trojan-Gal4 upstream of the *fru* C exon 39
Figure 18: Schematic showing the relative position of the newly identified exon in *fruitless* 41

List of Abbreviations

CRISPR/Cas9: clustered, regularly interspaced palindromic repeats/CRISPR-associated protein 9

fru: *fruitless*

gRNA: guide RNA

HDR: Homology-directed repair

MiMIC: *Minos* mediated integration cassette

RT-PCR: Reverse Transcriptional Polymerase Chain Reaction

TrpA1: Transient receptor potential cation channel subfamily A member 1

UAS: upstream activating sequence

1 Introduction

1.1 Speciation

Speciation refers to the process in which populations diverge and become distinct species (Sobel et al. 2010). The rich biodiversity on Earth results from a branching of countless speciation events separating homogenous groups into two or more distinct species. Consequently, in order to understand the origins and mechanisms of diverse physiological characteristics and traits, the study of speciation is critical.

There is some contention as to the definition of a "species," centered on an inability to define a trait that can be used to conclusively group or differentiate between all organisms (De Queiroz 2005; Mallet 2020; Sokal and Crovello 1970). For example, the commonly used "Biological species concept" describes distinct species as groups that can breed within themselves to create viable, fertile, offspring, but not with other groups (Löve 1964). Crucially, this concept excludes the majority of Earth's biodiversity that reproduces asexually, and so some biologists have been advocating for a genic view that groups organisms by genomic variation in critical loci (Baker and Bradley 2006; Harrison and Larson 2014; Mallet 2020). However, for the study of species that reproduce sexually, the biological species concept is prevalent and usually sufficient.

1.1.1 Mechanisms of Speciation

Under the biological species concept, populations must become reproductively isolated in order to be considered distinct species (Sokal and Crovello 1970). In organisms that reproduce sexually, this occurs via the gradual introduction of reproductive barriers that prevent populations from breeding with each other. These barriers must proceed beyond spatial isolation for divergence to occur, because groups that are physically distanced and do not naturally contact one another are not necessarily different species. Often, the barriers are categorized into two forms: pre-zygotic barriers, and post-zygotic barriers (Coyne and Orr 2004). As their names suggest, pre-zygotic barriers are mechanisms that prevent the fertilization of eggs, while post-zygotic barriers are mechanisms that prevent zygotes from maturing into viable, fertile adults.

Pre-zygotic barriers may manifest in many forms, including both behavioural and physical. Behavioural barriers are described as the manifestation of sexual selection for behaviours

specific to one's own species. Behavioural barriers may arise early in the speciation process and reinforce mating with one population over others, catalyzing divergence from other populations. For example, the European corn borer, *Ostrinia nubilalis*, is comprised of two different races (dubbed Z and E) differentiated by the sex pheromones produced by the females (Lassance et al. 2010; Wicker-Thomas 2011). The establishment of these two pheromone profiles are thought to be one of the reasons why males of each race naturally only mate with females of their own race, thus contributing to eventual speciation of the incipient populations (Wicker-Thomas 2011). Initially, physical barriers to reproduction were described using the work of Léon Dufour (Dufour 1844). He suggested that certain insect species use the shapes of their carapaces as a structural litmus test to exclusively allow mating with physically compatible individuals of their own species, similar to a "lock and key" (Masly 2012). However, modern research has introduced doubt into the ability of genital incompatibilities to independently stifle interspecific reproduction (LeVasseur-Viens et al. 2015; Masly 2012). Later studies showed other forms of physical incompatibilities in various species. For example, some species display deficiencies in the fusion of heterospecific eggs and sperm into viable zygotes. This may happen during attempted internal fertilization (Sweigart 2010), but is also commonly seen between species that utilize broadcast spawning (Lessios 2007; Vieira and Miller 2006). Nonetheless, these pre-zygotic mechanisms of reproductive isolation avert zygote formation, often by entirely preventing copulation. Therefore, they also prevent the use of crucial resources for the gestation and rearing of offspring that are incapable of continuing a lineage due to inviability or sterility.

Post-zygotic barriers are also expressed in a multitude of forms, with many resulting from unique genetic incompatibilities that prevent viability or fertility. Famously, mules, the offspring of male donkeys (*Equus asinus*) and female horses (*Equus caballus*), can grow into healthy adults; however, barring exceedingly rare exceptions, they are infertile. The infertility is a result of a chromosome mismatch originating from the inheritance of thirty-two chromosomes from the mother and thirty-one chromosomes from the father, which prevents the mules from creating functional sperm or eggs (Trujillo et al. 1962). Hybrid sterility and inviability are thought to originate from the accumulation of genetic differences in the separate lineages of two species. In many cases, there is a proportional relationship between genetic distance and the extent of incompatibility. For example, across pairs of frog species with hybrid sterility, there is a strong negative relationship between genetic distance and the percentage of embryos developing to the

larval stage, indicating the presence of mechanisms that terminate hybrid development (Sasa et al. 1998). These post-zygotic mechanisms prevent interbreeding among species, however, they are still wasteful for the parents (F0 generation; Noor 1995), because the resources used for mating and the development of inviable offspring come at a metabolic cost.

1.1.2 Genetic influences on mating barriers

In contrast to reproductive barriers that impair the development or viability of offspring, genetic influences on behavioural isolation between species are far less understood. There is evidence to suggest that genetic differences contribute to the sexual selection and resultant behavioural isolation in species pairs of frogs (*Physalaemus petersi*; Boul et al., 2007), flycatchers (*Ficedula sp.*; Saetre & Saether, 2010), and cichlids (*Pundamilia sp.*; Haesler & Seehausen, 2005), however, due to limited genetic tools in these species, mutagenic experiments are difficult to perform to identify or confirm potential candidate genes influencing sexual selection and enforcing behavioural isolation.

1.2 *Drosophila* as models of mating barriers and speciation

The *Drosophila* genus, and particularly the *Drosophila melanogaster* subgroup, has been extensively used to study the genetic basis of behavioural isolation and speciation (David et al., 2007; Kopp & True, 2002). The *D. melanogaster* subgroup is composed of nine closely-related species of fly (Figure 1.), some of which have post-zygotic reproductive barriers with sibling species from the group, and all of which are behaviourally isolated from each other to various extents (Cobb et al. 1988; Matute and Coyne 2010). In other words, all of these species preferentially mate with conspecific partners over heterospecific partners, and only some of these species pairs produce sterile or inviable hybrid offspring. Combined with a multitude of other factors that make *Drosophila* a widely-used model system, the genus becomes an exceptional candidate group for the study of speciation.

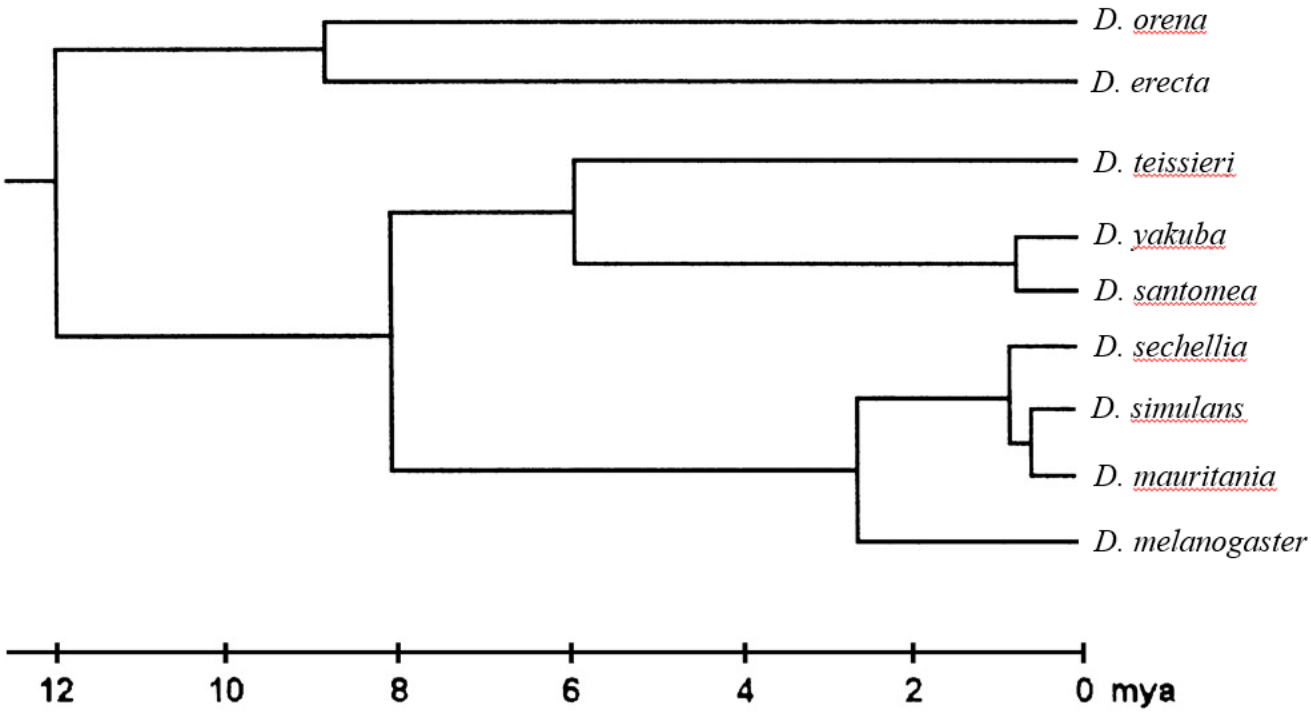

Figure 1: Phylogenetic tree of the *Drosophila melanogaster* species subgroup, including approximate divergence timeline in units of millions of years ago (mya; David et al. 2007).

1.2.1 Sexual selection in *Drosophila*

Species within the *Drosophila* genus exhibit stereotypical courtship and receptivity behaviours, which vary only slightly between species (Greenspan and Ferveur 2000). These shared behaviours allow Drosophilids to be used as model organisms in the study of behavioural isolation and speciation. In particular, *Drosophila melanogaster* is notable as a model organism that has been used extensively for decades, due primarily to short generational time, ease of maintenance, and most importantly, the development of genetic tools for mutagenic experiments (Roberts 2006; Yamaguchi and Yoshida 2018).

Many of the aforementioned mutagenic experiments involved the identification and manipulation of genes affecting sexual preferences. Further experimentation using other species in the *Drosophila melanogaster* subgroup has provided a wealth of information regarding heterospecific mating and speciation (Kopp and True 2002; Yassin and Orgogozo 2013). Perhaps the second most studied species in the subgroup is *Drosophila simulans*, a species of fly with which *D. melanogaster* can mate and form viable, but sterile, female hybrid offspring (David et

al. 2007) that are valuable in studying heterospecific mating behaviours. Hybrids contain homologous chromosomes from each of their parent species and provide the opportunity to use targeted mutations in a single relevant allele to explore mechanisms of interspecific behavioural isolation by assessing receptivity to male courtship (Barbash, 2010; M. Laturney & Moehring, 2012; Stern, 2014).

1.2.1.1 The male courtship ritual

The traditional male courtship ritual is by far the most researched and discussed set of behaviours associated with sexual selection in *Drosophila.* The success of courtship is reliant on both the presenter and potential mate's use and interpretation of signals arising from visual, chemical, tactile, and auditory cues (Anholt et al. 2020). Different species can display nuanced differences in the expression of courtship, however, the fundamental core behaviours are highly conserved (Greenspan and Ferveur 2000). Courtship is composed of several steps, which typically last for several seconds each and are repeated several times over the duration of courtship, but not necessarily in a fixed sequence (Spieth 1974). To initiate courtship, the male orients his head towards the female and begins tapping her abdomen with his fore tarsi (Spieth 1974). It is thought that the tapping action allows the male to sense his mate's pheromone profile and determine if she is a suitable mate (Bontonou and Wicker-Thomas 2014). The following courtship actions are not always performed in the order described here, but they are all ubiquitous courtship behaviours in the *D. melanogaster* subgroup (Spieth 1974). If his mate is deemed suitable, the male will vibrate one of his wings to create a distinct species-specific song determined by wing angles and ranges of motion (Spieth 1974). The south Asian species *Drosophila virilis,* for example, tend to extend their wings to an angle of 10-14 degrees from their abdomen before vibrating them with a small maximum wing stroke amplitude, while *Drosophila melanogaster* tends to extend their wings to an angle of 90 degrees before vibrating them with a much larger maximum amplitude (Spieth 1974). In addition, the male may vibrate his abdomen creating substrate-borne patterns that can be appraised by the female (Mazzoni et al. 2013). The male will then extend his proboscis and lick the female's genitalia, which is another transfer of chemical cues. If the female is receptive, he will then attempt to copulate by positioning himself between her wings using his fore and mid legs, before finally curling his abdomen underneath her (Bontonou and Wicker-Thomas 2014; Spieth 1974).

1.2.1.2 Female behaviours of selection and receptivity

Female *Drosophila* also have distinct behaviours of mate selection and receptivity (Cook and Connolly 2008; Manning 1967; Spieth 1974). They have a wide range of possible responses to male courtship. When a female accepts the courtship of a male, they affirmatively respond with three behaviours; wing spreading, genital spreading, and cessation of locomotion (Laturney and Billeter 2014; Spieth 1974). Wing spreading involves a female holding her wings in an outward position to allow a courting male to mount in between them. Genital spreading involves lower placement of the abdomen and slight extrusion of the genitalia to align with her mate's genitalia and allow copulation. It is important to note that mature *Drosophila* females must accept the advances of a courting male for copulation to be initiated (Spieth 1974).

The behaviours females use to reject male courtship are much more diverse than those for acceptance of males. They include efforts of physical distancing such as running, jumping, and flying (termed "decamping"), aggressive behaviours such as kicking and wing flicking and abdomen elevation and depression to prevent male courtship behaviours such as tapping of the abdomen or licking of the genitalia, and to prevent male mounting for copulation (Cook and Connolly 2008; Manning 1967; Spieth 1974).

1.2.1.3 Heterospecific courtship and rejection

To prevent the waste of resources in courting, copulation, and the production of non-viable or infertile offspring, species must be able to differentiate between themselves (conspecifics) and others (heterospecifics; Fan et al., 2013) Indeed, in the *Drosophila* genus, hybrid offspring are exceedingly rare in natural conditions (Barbash 2010; Spieth 1974) which suggests that individuals may employ biological mechanisms to determine which mates are suitable, while isolating themselves from unsuitable heterospecifics.

A wealth of research has confirmed that a variety of behavioural mechanisms inhibit mating between heterospecific *Drosophila* species. For example, it has been found that sister species *Drosophila yakuba* and *Drosophila santomea* are partially reproductively isolated by temperature preference (Matute et al. 2009). *D. yakuba* prefer to live in temperatures that are three to four degrees warmer than *D. santomea*. Additionally, at the threshold of 28°C, *D. yakuba* are much more fertile and long-lived than their counterparts (Matute et al. 2009). Thus,

climate preferences further reinforce reproductive isolation by limiting access to healthy heterospecific mates.

There is also evidence that *Drosophilids* can learn to discriminate between conspecifics and heterospecifics through experience (Spieth 1974). Naïve male *Drosophila persimilis* that face repeated rejection from female *Drosophila pseudoobscura*, become much more likely to court conspecific females in the future, reinforcing species isolation (Dukas 2008).

Drosophila females also show a sexual preference for conspecific courtship songs created by courting males. For example, when paired with wingless conspecific males who could not produce their own song, *D. melanogaster* females showed significantly higher receptivity to courtship attempts while listening to synthetic *D. melanogaster* song, as opposed to synthetic *D. simulans* song (Immonen and Ritchie 2012). In another experiment, three of five strains of *D. sechellia* were less receptive to courtship from intact-winged *D. melanogaster* males, than wingless ones who could not produce their own song (Tomaru and Oguma 2000). These results demonstrate conspecific courtship song biases in several *Drosophila* species.

Another major factor in mate discrimination are the cuticular hydrocarbons (CHCs), some of which act as species-specific sex pheromones, and allow individuals to distinguish conspecifics from heterospecifics (Bontonou and Wicker-Thomas 2014). All *Drosophila* express CHCs that are sensed via olfactory organs such as the antennae, and gustatory organs such as those located on the tarsi and proboscis (Ferveur 2005). Throughout the genus, CHCs vary in length, quantity, and saturation, with some species even having sex-specific differences. For example, males of the *Drosophila melanogaster* subgroup express relatively equivalent quantities of 7-tricosene and 7-pentacosene, but the females of many of these species express different hydrocarbons from their male counterparts and from each other, creating species-specific hydrocarbon profiles that can theoretically be used for mate discrimination (Bontonou and Wicker-Thomas 2014). Indeed, the relationship between hydrocarbon profiles and mate choice has been directly tested in several species. In one study, *Drosophila melanogaster* females had their oenocytes, the cells responsible for creating CHCs, ablated (dubbed oe^- females), and then were paired with males from three sibling species: *D. simulans, D. yakuba,* and *D. erecta* (Billeter et al. 2009). The males of the three species courted the oe^- *D. melanogaster* females at a much higher frequency than the oe^+ control females. Furthermore, coating the oe^- females with 7, 11-heptacosadiene, a

predominant CHC in the *D. melanogaster* female hydrocarbon profile, immediately halted further courtship attempts by the same males, highlighting the importance of CHCs in mate discrimination.

1.2.2 Candidate genes influencing female receptivity

Researchers have questioned the physiological origins of sexual behaviour phenotypes for decades. For *Drosophilids*, mate choice and sexual selection are at least partially derived from the expression of specific genes. The research involving genetic influences on mate choice in female *Drosophila* has been lacking compared to the depth of research on male mating behaviour. Neurons that express *doublesex* (*dsx*), a gene most-known for its role in somatic sex determination, have been explored in relation to female mate choice, with the reasoning that *dsx* may also regulate dimorphisms in sexual behaviour. Indeed, it was found that the activation of neuronal subsets expressing *dsx* increased female *D. melanogaster* receptivity to conspecific male courtship attempts, while silencing those same neurons had the opposite effect of depressing receptivity (Zhou et al. 2014).

The gene *painless* (*pain*), a homolog of the mammalian *TrpA1*gene that encodes for neuronal cation channels, also influences female receptivity. It was revealed that not only is *pain* responsible for physiological responses to aversive stimuli, but when knocked down, it causes female *Drosophila melanogaster* to mate with conspecific males much sooner after eclosion (Sakai et al. 2009). There were no differences in male courtship towards the *pain* mutant females, indicating that the increased mating success was exclusively a female response.

Indeed, research encompassing genetics and female receptivity is scarce and does not explain the conspecific preference found in females of all *Drosophila* species. To date, no “behavioural switch gene” candidate has been found for female mate preference. However, recent evidence indicates that natural alleles of a gene called *fruitless* (*fru*) do influence female mate rejection behaviours, both within and between species (Chowdhury et al. 2020). The *fru* gene has been studied for its effect on male courtship behaviours, but was previously thought to have no effect on female mating behaviours, so the literature regarding its effects on female sexual behaviours is comparatively lacking.

By using *Drosophila* as models for genetic and mutagenic experiments, researchers have characterized much of the structure of *fruitless* and its end products, as well as some of the most visible effects of these products on sex determination and mating behaviour (Chowdhury et al. 2020; Nojima et al. 2014; Ryner et al. 1996). Crucially, while it is most well-studied in *Drosophila*, *fruitless* has homologues in many insect species such as: mosquitoes (*Anopheles gambiae*), parasitic wasps (*Nasonia vitripennis*), and grasshoppers (*Chorthipuss spp*), indicating potentially conserved effects on sexual selection within the arthropoda phylum (Salvemini et al. 2010).

The *fruitless* gene encodes for transcription factors that regulate many biological processes, including sexual development and behaviour (Douglas and Levine 2006). The mRNA transcripts derived from *fru* begin with one of five first exons, dubbed P1-P5, followed by the common exons (C1-C5) present in almost all *fru* transcripts, and end with one of four terminal 3' exons, A, B, C, or D (Figure 2; Anand et al. 2001). The most researched transcripts are the sex-specifically spliced transcripts that begin with P1, which are involved in both sexual development and male sexual behaviour (Anand et al. 2001; Demir and Dickson 2005; Ryner et al. 1996). Due to sex-specific splicing, these P1 transcripts form functional *Fruitless* proteins in males (Fru^M), but have an early stop codon in females, creating truncated, presumably non-functional proteins (Von Philipsborn et al. 2014). In males, Fru^M can contain a 3' terminal domain derived from any one of the A, B, or C exons of *fru*, and each variant of Fru^M is involved in modulating distinct aspects of the male courtship ritual based on which terminal domain is expressed (Neville et al. 2014; Nojima et al. 2014).

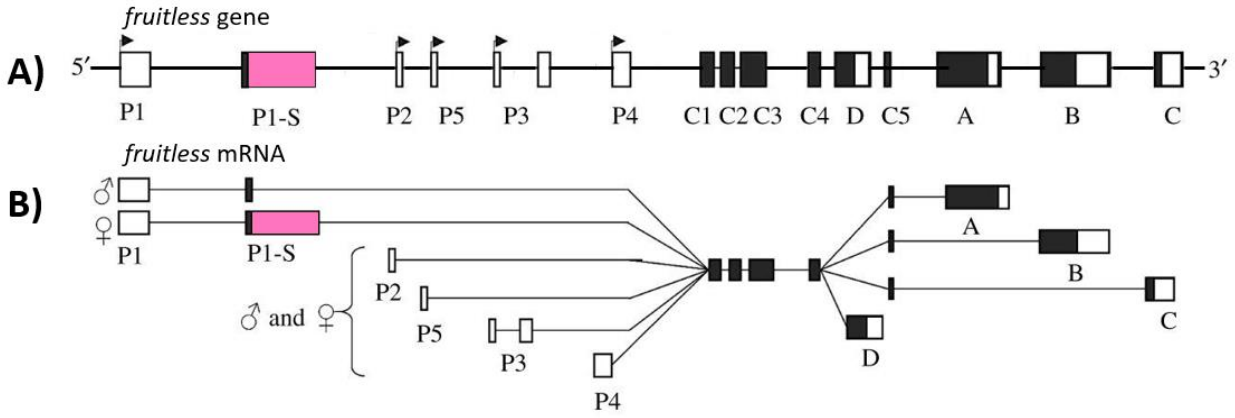

Figure 2: Schematic showing the (A) *fru* gene, and the (B) possible mRNA transcripts derived from it. Boxes represent exons; black boxes represent coding sequence; arrows indicate transcription start sites. Transcripts may contain any of the P1-P5 exons on the 5' end, the common exons in the centre, and any of the A-C exons on the 3' end; transcripts starting with P4 may also end with the D exon. The P1-S exon includes a female-specific portion (shown in pink) that contains a premature stop codon, preventing the translation of full-length P1 proteins. Figure modified from Chowdhury et al. (2020).

The Fru^M proteins also act as inhibitors of heterospecific courtship in *Drosophila melanogaster* males (Fan et al. 2013), thereby compounding the role of *fru* in regulating mate choice in males. However, while Fru^M provides a strong basis for the influence of genetics in male sexual behaviour, females also display distinct sexual behaviours (Cook and Connolly 2008; Spieth 1974) but their behaviour is not affected by *fru* P1 transcripts (Anand et al. 2001; Demir and Dickson 2005; Salvemini et al. 2010).

Unlike the *fru* P1 sex-specific transcripts that affect male courtship, the *fru* transcripts affecting female courtship are not sex-specifically spliced. Instead, research done in female *D. melanogaster* has revealed that a complete knockout of the *fru* P2 exon drastically reduces their receptivity to courtship by conspecific males (Chowdhury et al. 2020). In contrast, a hemizygous knockout of the same exon has no effect on conspecific receptivity, indicating that a single functional allele of P2 is sufficient to generate wildtype levels of receptivity in *D. melanogaster* females.

The *fru* P2 exon also affects heterospecific female rejection (Chowdhury et al. 2020). Hybrid *D. melanogaster*/*D. simulans* females have two wildtype *fru* alleles, one from each parent species. These hybrid females, display phenotypic levels of receptivity towards *D. melanogaster* male courtship that are more akin to the levels of pure *D. melanogaster* females, as opposed to the negligible receptivity shown by pure *D. simulans* females. In other words, the *D. melanogaster*

alleles for female receptivity towards *D. melanogaster* males are dominant or semi-dominant over the *D. simulans* alleles for female rejection of these males. By deleting the D. *melanogaster* allele of a particular candidate gene or transcript, the recessive *D. simulans* allele is 'unmasked' and its effects can potentially be observed. Hybrid females with a hemizygous deletion of their *D. melanogaster* P2 exon displayed reduced receptivity to *D. melanogaster* males, which is the more *D. simulans*-like female response. These same hybrid females showed no reduction in mating when paired with *D. simulans* males, indicating that this effect is not simply a reduction in overall mate receptivity. The loss of the *D. melanogaster* P2 exon unmasked the wildtype female *D. simulans* sexual behaviour phenotype in the hybrids. Therefore, the P2 isoforms of Fru appear to regulate female receptivity of conspecific and rejection of heterospecific males in *D. melanogaster* and *D. simulans*.

The implications are as follows: there is a dominance or semi-dominance of phenotypic expression in hybrids whereby the wildtype *D. melanogaster fru* P2 allele overrides the effects of the wildtype *D. simulans fru* P2 allele. In addition, because *fru* is a gene that is subject to alternative splicing, in the same way that Fru^M was found to affect distinct aspects of male courtship based on its 3' terminal domain (A-C; Nojima et al., 2014), the P2 isoforms of *fru* may also affect the distinct receptivity behaviours of females based on their 3' terminal domains (again, A-C).

1.2.2.1 Expression of *fruitless* in *Drosophila melanogaster* females

One major limitation on the study of *fru*'s effects on female receptivity is that there have been no comprehensive studies examining which alternatively spliced mRNA transcripts are expressed in adult female *Drosophila,* and where they are expressed. A few studies have ascertained where transcripts specifically beginning with each of the 5' exons P1-P4 are expressed in adult females (Leader et al. 2018). P3 and P4 transcripts were generally found in the whole body and reproductive organs, however, some sources suggest they may also be found in the central nervous system (Dornan et al. 2005; Salvemini et al. 2010). The behaviourally relevant P2 transcripts of *fru* seem to be primarily expressed in the adult female head, with limited expression of P2 transcripts that contain the 3' terminal A exon, in the central nervous system (Dornan et al. 2005; Salvemini et al. 2010). However, further categorizing those transcripts by expression of the various 3' terminal exons A-D has proven difficult (Dornan et al. 2005;

Salvemini et al. 2010). Essentially, the combinations of 5' and 3' ends that are actually expressed, and their expression patterns, remain unknown.

In the regulation of male courtship behaviours, the roles of each 3' variant of P1 *fruitless* transcript (P1-A, P1-B, P1-C) appear to be distinct from each other (Neville et al. 2014; Nojima et al. 2014). The effects of the transcripts also act additively and independently. This means that each one regulates specific sexual behaviours and the loss of an individual variant leaves other behaviours untouched in phenotype (Von Philipsborn et al. 2014). Furthermore, it is posited that each splice variant is expressed and functional in distinct (but partially overlapping) subsets of neurons (Neville et al. 2014). By extrapolation, if the alternatively spliced P2 mRNA (A, B, and C) transcripts are truly responsible for modulating female behaviours of receptivity, they may act in a similar fashion to how *fru* transcripts regulate behaviours in males. It is therefore imperative to be able to distinguish which splice variants are present in adult females, and in which cells they are expressed, to distinguish between variant-specific regulation of sexual behaviours.

1.3.1 Gal4-UAS

The location of gene expression can be manipulated and visualized using the Gal4-UAS system. The transgenic Gal4-UAS *Drosophila* expression system is derived from yeast (Strassburger and Teleman 2016). *GAL4* codes for a yeast transcriptional activator protein (Gal4) and can be placed under the control of an endogenous promotor (i.e. driver) to be expressed whenever the associated promotor initiates transcription (Figure 3; Strassburger and Teleman 2016). After translation, the functional Gal4 then binds to an enhancer sequence called the upstream activating sequence (UAS) which activates transcription of a linked gene of interest (i.e. a responder; Caygill & Brand, 2016). For example, the UAS could be linked to a reporter gene *GFP*, which would lead to expression of green fluorescent protein (GFP) in all cells where the promotor is active. The bipartite system means that any *GAL4* can be readily paired with any UAS by simply crossing together flies containing these separate components, making this a powerful system for controlling the temporal and spatial expression of genes.

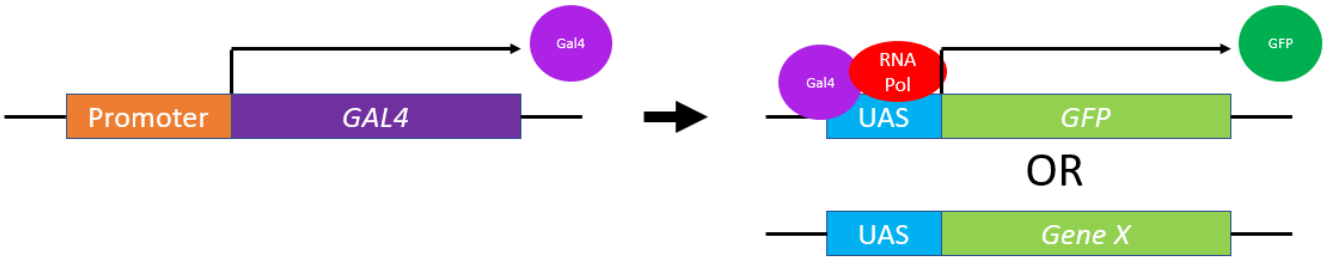

Figure 3: In the traditional Gal4-UAS system, exogenous *GAL4* (purple rectangle) is integrated into the genome and transcribed under the control of an endogenous promotor (i.e. driver, orange rectangle) for tissue specific expression. After translation, the functional Gal4 protein (purple circle), a transcriptional activator, binds to the UAS enhancer (blue rectangle) and recruits RNA polymerase (RNA Pol., red oval) to initiate transcription of a linked gene of interest (green rectangles), such as GFP, leading to the tissue specific production of GFP protein (green circle).

1.3.1.1 Trojan-Gal4-UAS

The Gal4-UAS system is efficient and convenient, but traditionally relies on the actions of an endogenous promotor to induce expression of Gal4 (Strassburger and Teleman 2016). For alternatively spliced genes, such as *fru*, it may be necessary to link Gal4 expression to the expression of specific exons rather than a gene in its entirety, meaning the traditional promotor-based system will not suffice. To achieve transcript-specific Gal4 expression, a technique called a Trojan-Gal4 Integration links Gal4 expression to a specific transcript (Figures 4, 5; Diao et al. 2015). For example, Trojan-Gal4 could be linked to the A exon of *fru*; when paired with a UAS-GFP, cells expressing fru-A transcripts can be visualized. The Trojan-Gal4 operates by inserting an exogenous DNA sequence, that encodes GAL4 and a 5' self-splicing polypeptide, into an intron between two endogenous coding exons (Diao et al. 2015). The *GAL4* gene construct is flanked by a 3' splice acceptor site and a 5' splice donor site which ensure incorporation of the Trojan exon's mRNA product into the mature endogenous *fru* mRNA. Finally, ribosomes are recruited to the Trojan mRNA and result in a fully functional Gal4 protein that can bind to a UAS (Diao et al. 2015; Diao and White 2012).

1.4 Genetic tools for assessing transcripts affecting female receptivity

1.4.1.1 PhiC31-mediated recombination

To get transcript-specific expression of Trojan-Gal4, it must be integrated into a precise position adjacent to the exon of interest. This site specificity can be achieved using phiC31-mediated recombination (Figure 4). Phic31-mediated recombination consists of injecting a plasmid into *Drosophila* embryos and swapping a gene of interest from that plasmid into a compatible genomic site in the embryo, using the actions of PhiC31 integrase (Venken et al. 2011). Libraries of *Drosophila* lines have been generated that contain compatible sites called *Minos* mediated integration cassette (MiMIC) sites, derived from transposable elements and flanked by PhiC31 recognition sequences (Nagarkar-Jaiswal et al. 2015). In theory, donor plasmids containing genes of interest flanked by the reciprocal PhiC31 recognition sites can be injected into the embryos to mediate a swap of desired genes into MiMIC sites in genomic loci (Venken et al. 2011). The PhiC31 enzyme is not native to *Drosophila*, so its expression may be induced by using a genetically-altered line that expresses the enzyme endogenously, or by simultaneously injecting plasmids with the gene's expression under the control of an *Actin*-promoter that is ubiquitously activated in *Drosophila* (Venken et al. 2011). If transgenesis occurs within the embryo's germline progenitors, it can pass the transgenic mutation to its offspring which can culminate in the creation of a stable transgenic line.

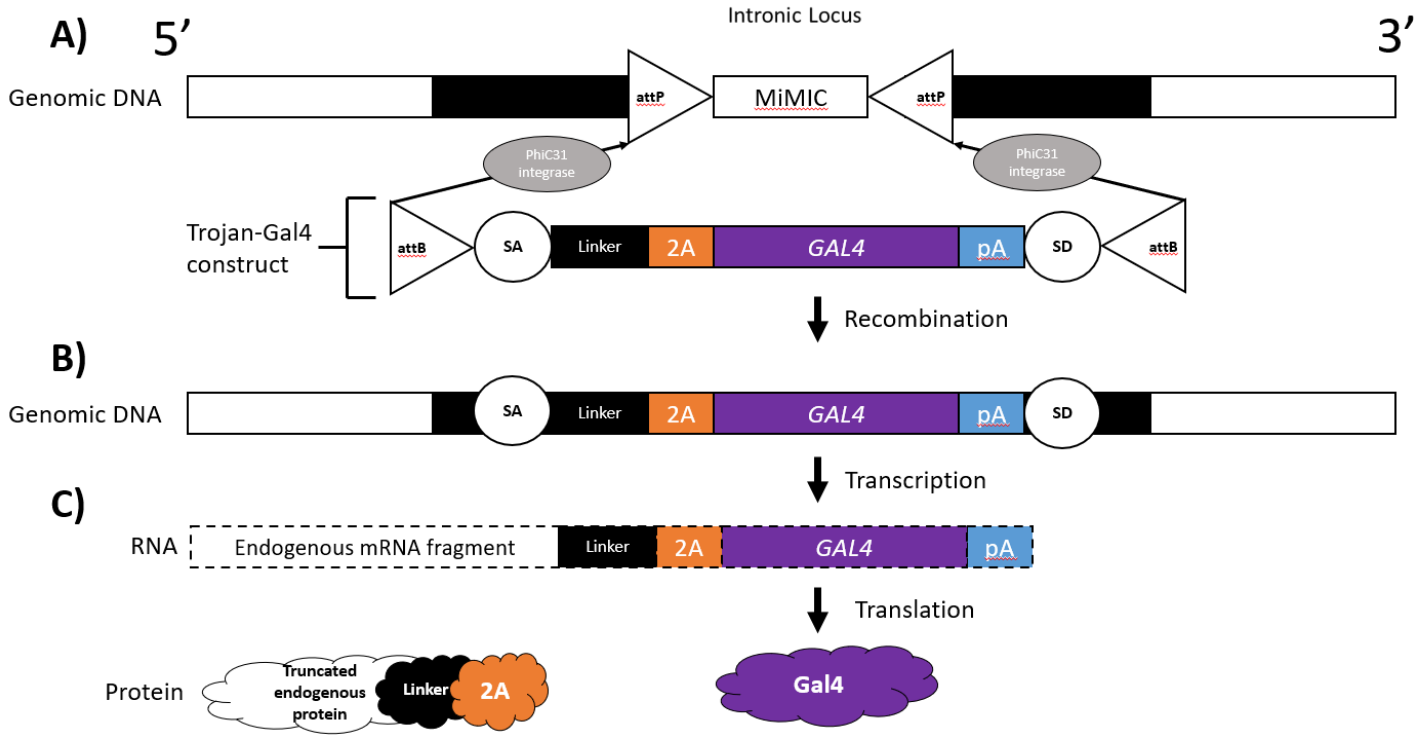

Figure 4: Trojan-Gal4 Integration system. (A) PhiC31 recombination integrates Trojan into the genome. The attB sites (attB, white triangles) flanking the Trojan-Gal4 construct and the attP sites (attP, white triangles) flanking the genomic MiMIC site are recognized by PhiC31 integrase, which then catalyzes a swap of the Trojan-Gal4 construct and the MiMIC site sequence. (B) The Trojan construct is incorporated into the transcript of interest. The splice acceptor and splice donor sites (SA and SD, white circles) ensure integration of the Trojan-Gal4 transcript into endogenous mRNA, leading to transcription whenever the endogenous gene is transcribed. The "linker" is a sequence of variable length designed for Trojan to be in the same functional reading frame as the transcript of interest. (C) During translation, the 2A sequence (orange square) codes for an amino acid motif that causes ribosomes to skip one of its own glycine codons before re-initiating translation at the following proline codon (Diao and White 2012), resulting in two unbonded peptides created from the same transcript, and thus separating the functional Trojan GAL4 protein from the endogenous protein.

1.4.1.2 CRISPR-Cas9/Homology-Directed Repair

A secondary mechanism of targeted mutagenesis uses a technique called CRISPR-Cas9/Homology-Directed Repair (Figure 5). This mechanism can be used for locations that do not have a pre-existing MiMIC site in the desired location for Trojan-Gal4 insertion. This technique involves co-opting bacterial defense mechanisms against viruses to induce genomic breaks and subsequent repair in eukaryotic cells using a gene template on a plasmid (Gratz et al. 2014; Port et al. 2014). This process culminates in the integration of the plasmid-donor gene into the embryonic genome at specifically targeted loci (Gratz et al. 2015). Essentially, a guide RNA (gRNA) plasmid is injected into the embryos. The guide RNA plasmid transcribes a gRNA

complementary in sequence to the genomic protospacer region that has the desired cut site embedded in it. The gRNA is then picked up by Cas9 nuclease proteins that are endogenously expressed, and its complementarity directs the Cas9-gRNA complex to the protospacer sequence (Figure 5A). Following recognition of an NGG nucleotide sequence called the Protospacer Adjacent Motif (PAM), the Cas9 introduces double stranded breaks in the genome three to four nucleotides upstream of it. A second plasmid containing a gene of interest flanked on either side by long sequences homologous to the cut-site-flanking sequences (homologous arms) is then recognized by the cell as a repair template for the break (Figure 5B; Lin and Potter 2016). The enzyme DNA polymerase is recruited by the cell and repairs the cut site with the gene of interest found between the homologous arms in the plasmid, thereby incorporating the gene from the plasmid-donor into the embryonic genome using homology-directed repair (HDR).

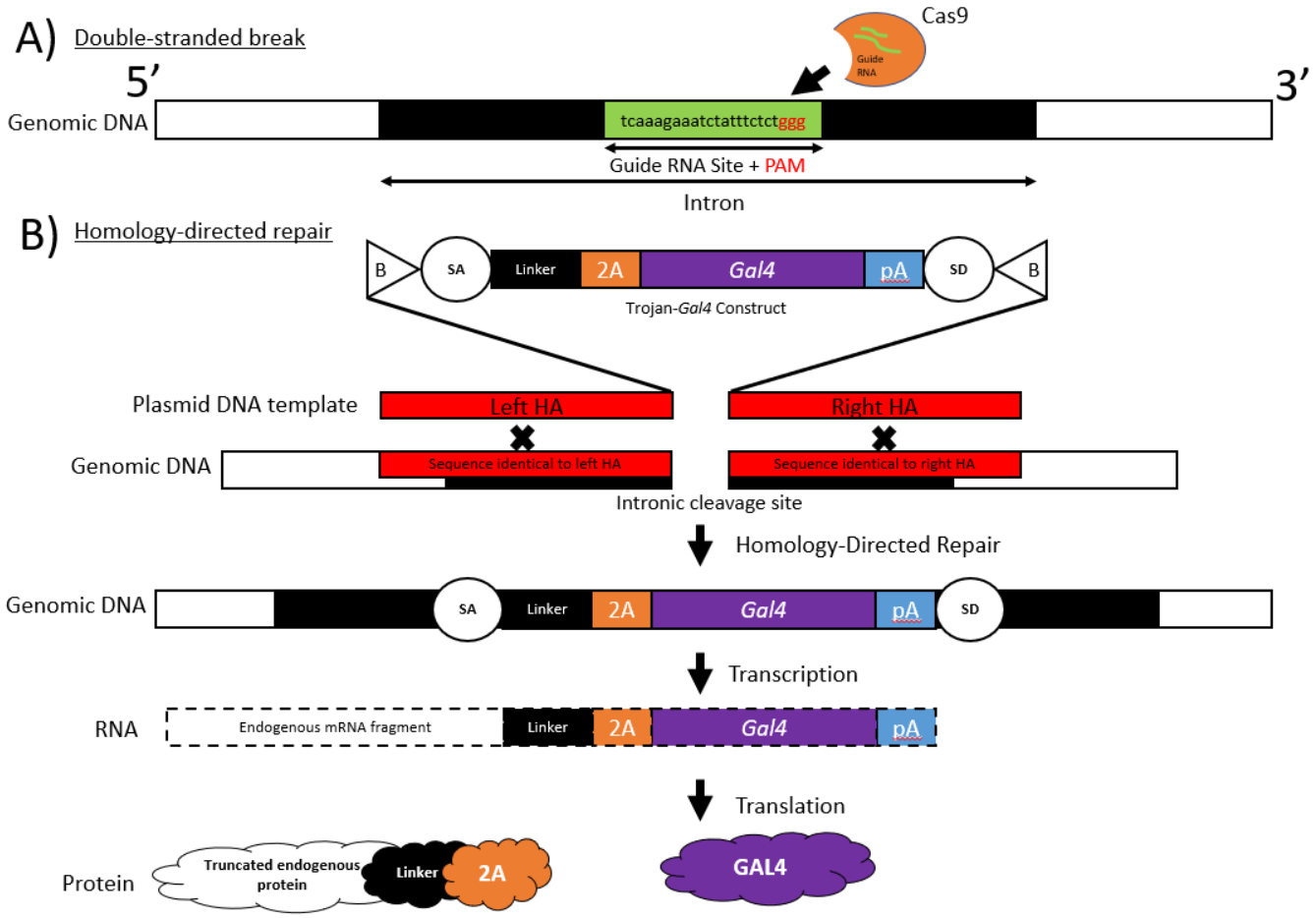

Figure 5: Schematic detailing CRISPR-HDR Trojan GAL4 integration into Cas9-expressing *Drosophila melanogaster*. (A) Endogenous Cas9 uses guide RNA to target the complementary genomic guide RNA

site, to create a double-stranded break in the intron between two exons (white rectangles). (B) Homologous arms (Left and Right HA, red rectangles) to the cut-site-flanking genomic DNA help the cell identify the Trojan GAL4 construct as a template for homology-directed repair of the now-cleaved intron. DNA polymerase is recruited and uses the Trojan construct as a nucleotide template to fill in the gap. Once the double-stranded break is filled with the Trojan sequence, transcription and translation proceed as normal. The splice acceptor and splice donor sites (SA and SD, white circles) ensure integration of the Trojan-Gal4 transcript into endogenous mRNA, leading to transcription whenever the endogenous gene is transcribed. The "linker" is a sequence of variable length designed for Trojan to be in the same functional reading frame as the transcript of interest. During translation, the 2A sequence (orange square) codes for an amino acid motif that causes ribosomes to skip one of its own glycine codons before re-initiating translation at the following proline codon (Diao and White 2012), resulting in two unbonded peptides created from the same transcript, and thus separating the functional Trojan GAL4 protein from the endogenous protein.

1.5 Experimental objectives

My thesis aims are:

1) To compare and contrast the *fru* mRNA transcripts present in male and female *Drosophila melanogaster*, and identify candidate transcripts influencing heterospecific female rejection behaviours. To catalogue the *fruitless* transcripts that are expressed, I will sequence the products of targeted RT-PCR (further described in Methods) produced from male and female *Drosophila melanogaster* RNA. Current knowledge of sex-specific splicing of *fruitless* focuses on the 5' end of the transcripts, with emphasis on transcripts containing the male-specific P1 exon splice variant. Therefore, my work aims to determine if there are any further sex-specifically-expressed transcripts in relation to the 3' terminal exons A, B, and C, which could be implicated in sexually-distinct behaviours. Particular emphasis will also be placed on the P2 variants which have been implicated in the regulation of female sexual behaviours.

2) To create a transgenic system that expresses Trojan-Gal4 in the pattern of A, B, or C transcripts. I will use phiC31 integrase and existing MiMIC sites to insert a Trojan-Gal4 adjacent to the A and B exons, and CRISPR-HDR to insert a Trojan-Gal4 next to the C exon, where a MiMIC site does not exist.

3) To use the Trojan-Gal4-UAS system to visualize the expression patterns of *fru* transcripts containing 3' exons A, B, or C, and determine their effects on female rejection of heterospecific mates. Once the Trojan-Gal4 constructs are integrated into their desired positions upstream of *fru* exons A, B, and C, these *Drosophila* lines can then be mated with

other lines containing UAS-linked genes of interest. The resulting offspring will consist of flies containing both the Trojan-Gal4, and a gene of interest that will consequently be transcribed in the Trojan-Gal4's exon-specific expression pattern. To identify the neuronal expression patterns of *fru* transcript variants containing the A, B, and C 3' terminal exons, allowing for exploration of neural networks that regulate sexual behaviours in female *D. melanogaster*. I will first pair the three Trojan-Gal4s (A, B, C) with UAS-GFP to visualize the location of expression of these transcripts. I will then assess the influence of each subset of neurons on heterospecific rejection behaviours in females by pairing the three Trojan-Gal4s with UAS-linked genes that silence, hyperactivate, or ablate neurons, then assess the resulting impact on female rejection behaviour.

2 Methods

2.1 Fly husbandry and stocks

All *Drosophila* stocks were maintained in 30 mL vials containing a standard cornmeal-based food source (Bloomington *Drosophila* Stock Center recipe). Mating pairs for the purpose of genetic crosses were kept in an incubator programmed at 24 °C, 70% humidity, and a 14h:10h light:dark cycle to simulate ideal climate preference. All other flies were contained in identical vials, however, they were simply stored in a temperature-controlled room at ~24 °C. For the purposes of identifying the *fru* transcripts expressed in *Drosophila melanogaster* via RT-PCR, a wildtype strain of *D. melanogaster*, Canton S, was obtained from Dr. Anne Simon. All other stocks were obtained from the Bloomington *Drosophila* Stock Center (BDSC; Bloomington, Indiana). For the purposes of Trojan-Gal4 integration upstream of the *fru* A and B exons, stocks containing *attP*-flanked MiMIC cassettes amenable to PhiC31-mediated recombination in the desired 3rd chromosome locations (Stock #42145: y^1w^*; $Mi\{y^{+mDint2}=MIC\}fru^{MI06350}/TM3, Sb^1 Ser^1$ and Stock #44345: y^1w^*; $Mi\{y^{+mDint2}=MIC\}fru^{MI07841}/TM3, Sb^1Ser^1$, respectively) were used. In order to proceed with Trojan-Gal4 integration upstream of the *fru* C exon using CRISPR-mediated HDR, a fly stock ubiquitously-expressing Cas9 under the control of an *Actin5C* promoter was obtained (Stock # 58492: y^1 $M\{Act5C\text{-}Cas9.P.RFP\text{-}\}ZH\text{-}2A$ w^{1118} $DNAlig4^{169}$).

Finally, a fly stock containing balancer chromosomes with phenotypically-dominant markers on the 3rd chromosome (Stock #3703: *w*^{1118}*/Dp(1;Y)y*$^{+}$*; CyO/nub*^{1} *b*^{1} *sna*^{Sco} *lt*^{1} *stw*^{3}*; MKRS/TM6B, Tb*^{1}) was obtained as a crossing line. The balancers maintain the desired genetic changes over generations through forced selection due to homozygous lethality of the balancer chromosomes and lethality of recombinant offspring.

2.2 Assessing *fru* transcript expression using RT-PCR

Two sets of twenty virgin females aged one week post-eclosion, were collected from the Canton-S strain of *D. melanogaster* in 1.5 mL microcentrifuge tubes and snap-frozen in liquid nitrogen prior to beginning the extraction protocols. Full-body total RNA was then extracted using a protocol (dx.doi.org/10.17504/protocols.io.fgtbjwn) adapted from the Invitrogen Life Technologies Trizol manual. The same RNA extraction method was then repeated using a set of twenty males and a set of twenty females aged two to three weeks post-eclosion. The different sets of RNA were used for biological controls, but herein, only the methods and results from the more-aged male and female samples are discussed.

The Maxima H Minus First Strand cDNA Synthesis Kit, with dsDNAse by Thermofisher Scientific (Waltham, Massachusetts) was used for all cDNA synthesis reactions, using 2 µg of RNA per sample, and following the manufacturer's protocol. In addition to cDNA pools derived from standard polyA tail primers which reverse transcribe all mRNA, cDNA pools from each sex of *Drosophila* were synthesized using primers specific to *fru*'s 3' terminal exons A, B, and C. The primer sequences are as follows: A: 5' GCTTGATCTTACAGTGCGCC; B: 5' TCAAGTGGTTGGCATTTGCG; C: 5' GGTTGAGTAGCCCTCATCCG. The resulting sex and exon-specific cDNA pools were then used for qualitative RT-PCR analysis of mRNA splice-variant expression.

DNA primers were obtained from BioCorp (Montreal, Quebec) for RT-PCR amplification of cDNA fragments spanning the distance from each of 5' *fru* exons P1-S, P2, P3, P4, and P5 (forward primers) to a common primer in the downstream common exon 3 (reverse primer). The forward primer sequences are as follows: P1-S: 5' TCAATCAACACTCAACCCGA; P2: 5' AATCGTCGCGGTCATAAAAT; P3: 5' TCATCAGCAAATGCCTCGT; P4: 5′ CCAAAAACTAAGCCCGTCAA; P5: 5' ACATAGACAGTGCCTCCTG. The common reverse

primer has a sequence of: 5’ AGTCGGAGCGGTAGTTCAGA. Each primer pair was used in RT-PCR reactions using exon-specific cDNA pools (A, B, and C) for each sex, as well as being used in the total cDNA pools for each sex as a form of positive control. Rather than designing primer pairs spanning the 2200-8400 bp length from each 5’ exon to each 3’ terminal exon of *fru* to be used in RT-PCR of total cDNA, this method of amplifying shorter fragments in pools of cDNA derived from exon-specific primers removes some of the cumbersome troubleshooting that may be prevalent in optimizing the amplification of lengthy DNA segments.

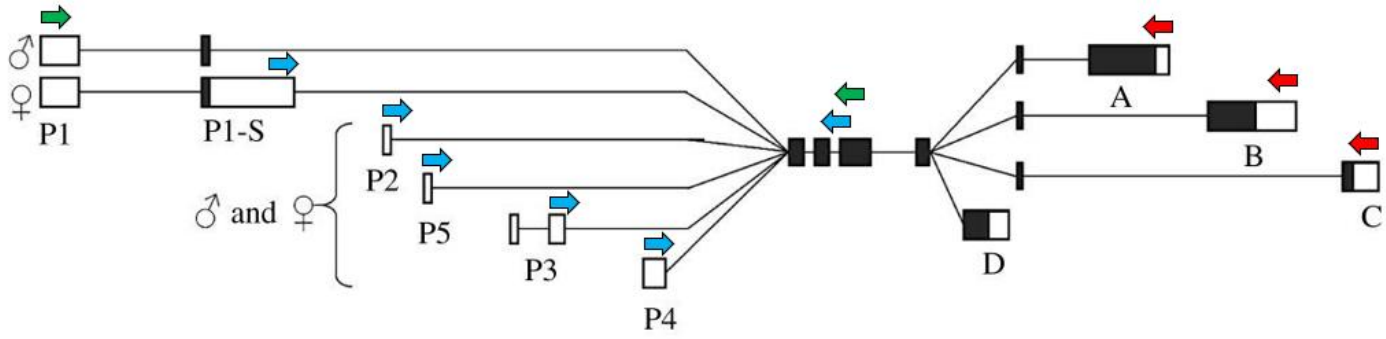

Figure 6: Schematic displaying the annealing positions of primers used in the RT-PCR experiments, relative to exons in the fru transcripts. cDNA synthesis: Red arrows indicate the positions of primers used for gene-specific cDNA synthesis, creating cDNA pools derived from transcripts containing only the A, B, or C exons. PCR: Blue arrows indicate the positions of forward primers in each of P1-S to P5 paired to the same reverse primer in common exon 3. The green arrows indicate the positions of a forward and reverse primer pair used exclusively for the amplification of fragments spanning from P1 to common exon 3. Each of these primer pairs was used to amplify fragments from each cDNA pool. Figure modified from Chowdhury et al. (2020).

RT-PCR was done using DreamTaq Green PCR Master Mix (2X) by Thermofisher Scientific. Initial samples for qualitative expression analysis were composed of 10 µL of master mix, 2 µL of 20uM forward and reverse primer mix, 1 µL of cDNA, and 7 µL of Ultrapure water for samples with a total volume of 20uL. The thermocycler protocol consisted of an initial 95 °C denaturation for 10 minutes, followed by a 15 cycle repeat sequence of: a 95 °C denaturation for 30 seconds, a 63 °C primer annealing step for 30 seconds (with the temperature decreasing by 1 °C per cycle in a touchdown pattern), and a 72°C elongation step for 1.5 minutes. Following the 15 cycle touchdown, the samples underwent 30-35 more identical PCR cycles, however, this time using constant annealing temperatures that were lower than their theoretical optimums according to the Thermofisher Scientific Tm calculator, in order to promote greater amplification. The final step was a 10 minute 72 °C final extension. The constant primer-

annealing temperatures that were used were: P1-S: 51 °C; P2: 49 °C; P3: 52 °C; P4: 50 °C; P5: 51 °C. Products were run on 1% agarose gels.

For sequencing, each RT-PCR was repeated with total sample volumes of 50 µL rather than 20 uL, to allow for the purification and collection of larger quantities of amplified DNA, and run on a 0.5% agarose gel.

DNA was extracted and purified using the GenepHIow Gel/PCR kit (DFH300) by FroggaBio (Concord, Ontario) according to the manufacturer's protocol. Sequencing was performed by the London Regional Genomics Centre (London, Ontario).

2.3 Plasmid preparation for microinjection

2.3.1 Plasmids for phiC31-mediated Trojan-Gal4 integration 5' of *fru* A and B exons

Two plasmids were obtained from the *Drosophila* Genomics Resource Center (NIH Grant 2P40OD010949; Bloomington, Indiana); *Act-phiC31-integrase*, and *pBS-KS-attB2-SA(1)-T2A-Gal4-Hsp70*. The *Act-phiC31-integrase* plasmid contains a sequence coding for the phiC31 enzyme under the control of a ubiquitous *actin* promoter. The *pBS-KS-attB2-SA(1)-T2A-Gal4-Hsp70* plasmid contains a Trojan-Gal4 sequence flanked by attB sites which would allow for recombination into an attP-flanked genomic site catalyzed by phiC31. In conjunction, these plasmids were injected into embryos of fly stocks #42145 and #44345, because the expression of phiC31 derived from the *Act-phiC31-integrase* plasmid should allow the Trojan-Gal4 construct from the donor plasmid to recombine into genomic attP-flanked MiMIC sites 5' of *fru* exons A and B, respectively. The plasmid injection mix was a total volume of 20 uL, containing 0.4 µg/µL *Act-phiC31-integrase*, 0.5 µg/µL *pBS-KS-attB2-SA(1)-T2A-Gal4-Hsp70*, and the remaining volume was standard blue food colouring. The mix was aliquoted into four separate tubes of 5 µL each, and stored at -20 °C.

2.3.2 Plasmids for Trojan-Gal4 integration 5’ of *fru* C exon via CRISPR/CAS9-mediated Homology-Directed Repair

2.3.2.1 Creating a suitable CRISPR gRNA plasmid vector

Using wildtype *Drosophila melanogaster* genomic data provided by FlyBase, potential 20-nucleotide protospacer sequences with adjacent PAM-sites were identified in the intronic region 5’ of the *fru* C exon and 3’ of the *fru* B exon. Protospacers were ranked by E-value to minimize off-target effects, and the sequence 5’ TCAAAGAAATCTATTTCTCT with a downstream PAM (5’ NGG) was selected as a suitable site. However, because I was planning to inject into a genetically modified stock of *D. melanogaster,* I accounted for the possibility of genomic variation in the intron by using PCR to amplify the region in stock #58492 and had it sequenced for any SNPs. The primers I used to amplify and sequence the region were as follows: Forward: 5’ ATCCCCTGTGTTAGTCACTC; Reverse: 5’ GATCCCTGATTGCCATAACC. The samples were then extracted and purified using the GenepHIow Gel/PCR kit (DFH300) by FroggaBio (Concord, Ontario) according to the manufacturer’s protocol, and sent to the London Regional Genomics Centre (London, Ontario) for sequencing.

Following confirmation of the desired sequence’s presence, the gRNA plasmid vector *pU6-3-gRNA* was obtained from the *Drosophila* Genomics Resource Center (NIH Grant 2P40OD010949). The gRNA plasmid was linearized with the restriction enzyme BsaI (New England Biolabs Inc.) in preparation for ligation of the gRNA sequence 24 base pairs 3’ of the U6 promoter. The restriction digestion protocol was as follows: 1 µg *pU6-3-gRNA*, 5 µL 10x NEBuffer 2.1, 1 µL BsaI, 1 µL Quick CIPhosphatase (New England Biolabs), and Ultra Pure water up to 50 µL, gently mixed and incubated at 37 °C for 4 hours, followed by a 20 minute enzyme inactivation at 65 °C.

The gRNA oligonucleotides were designed using the guide RNA site sequence and its complement as the base. However, research shows that reducing the length of a gRNA sequence from 20 nucleotides to 17 or 18 nucleotides increases specificity in the targeting of Cas9 nuclease (Fu et al. 2014). This is due to the tolerance of fewer nucleotide mismatches between the shorter gRNA and its complementary target sequence (Fu et al. 2014). Furthermore, the U6 promoter is substantially more efficient at initiating transcription of sequences that include a 5’ G, and so it is recommended to replace the first nucleotide of a gRNA sequence with a 5’ G if is

it not already this nucleotide (Cong et al. 2013). Finally, the 5' ends of the sense and antisense oligos must be complementary to the overhangs created by the restriction digest enzyme (in this case, BsaI) in the plasmid vector. Therefore, 5'-phosphorylated oligos with the following sequences were ordered from BioCorp: sense: 5' CTTCGAAGAAATCTATTTCTCT; antisense: 5' AAACAGAGAAATAGATTTCTTC. The oligos were then annealed together using the following protocol: 2 µg of each oligo, and Ultra Pure water added up to a total volume of 50 µL were mixed together in a 0.2 mL PCR tube and heated to 95 °C for 5 minutes before being allowed to cool to room temperature.

The annealed oligos were then ligated into the digested gRNA vector using T4 DNA Ligase (New England Biolabs) and the following protocol: 44.5 ng vector, 1 ng oligonucleotides, 2 µL 10x DNA Ligase Buffer, 1 µL T4 DNA Ligase, and Ultra Pure Water to 20 µL were gently mixed on ice before being incubated for 12 hours at 16 °C. The enzymes were heat inactivated at 65 °C for 10 minutes.

The ligated *pU6-3-gRNA* was added to NEB® 10-beta Competent E. coli (High Efficiency) cells (New England Biolabs) and transformed in accordance with the manufacturer's standard transformation protocol. One hundred microliters of the transformation mix was plated on LB+ampicillin plates (100 uL/mL) and grown overnight at 37 °C. Individual colonies were selected and grown in 1 mL of LB+ampicillin (100 uL/mL) for 24 hours. Plasmid was extracted from each cell culture using the Invitrogen PureLink™ Quick Plasmid Miniprep Kit according to the manufacturer's instructions, and each plasmid sample was sequenced at Robart's DNA Sequencing Facility (London, Ontario) to verify successful ligation of the oligos using the T7 and T3 primers (Figure 14).

2.3.2.2 Creating a suitable Trojan-Gal4 Donor Vector

The plasmid *pT-GEM(1)* was obtained from the *Drosophila* Genomics Resource Center (Bloomington, Indiana). This plasmid contains the sequence for a full attP-flanked Trojan-Gal4 expression construct, and has previously been used with CRISPR/Cas9-mediated homology-directed repair to create transgenic *Drosophila* that have exon-specific Gal4 expression (Diao et al. 2015).

The most crucial steps to preparing the plasmid for embryonic microinjections were flanking the Trojan construct with homologous arms identical to the approximately 800-900 base pair sequences on either side of the genomic cut site.

Based on wildtype *D. melanogaster* genomic data retrieved from the National Center for Biotechnology Information (Bethesda, Maryland), primer pairs were designed for amplifying the left homology arm and right homology arm. The primers were ordered with 5' additions that would result in complementary overhangs on the homologous arms for ligation into the plasmid vector following double restriction digestion. The restriction enzymes used to ensure the proper ligation orientation of the left homologous arm were AgeI-HF and NotI-HF (New England Biolabs), and those used for the right arm were AscI-HF and SpeI-HF (New England Biolabs). Therefore, the primer sequences were as follows: Left forward: 5' ACTGACACCGGTCAGCCCGAATTCGTTAAGTG; Left reverse: 5' ACTGACGCGGCCGCGAAATAGATTTCTTTGATTGTTTTGTTTTTTGTTT; Right forward: 5' ACTGACGGCGCGCCTCTGGGCTGATTTTCGTCCC; Right reverse: 5' ACTGACACTAGTCCACCTAGAACTGCAGCGAT.

Amplification of the homologous arms were conducted in a PCR reaction using total DNA from the injection stock # 58492 using Phusion™ High-Fidelity DNA Polymerase (2 U/µL) from Thermofisher Scientific (Waltham, Massachusetts) in the following protocol: 10 µL 5x Phusion™ HF buffer, 1 µL 10 mM dNTP mix, 2.5 µL 20 uM forward and reverse primer mix, 2.5 µL template DNA, 33.5 µL Ultra Pure water and 0.5 µL of Phusion™ High-Fidelity DNA Polymerase were added into PCR tubes and gently mixed for a total of 50 µL per tube. The tubes were then run in the following thermocycler protocol: A single 98 °C 30 second denaturation, followed by thirty cycles of 98 °C denaturation for 10 seconds, 72°C annealing for 30 seconds, 72 °C extension for 30 seconds, and a final single extension at 72 °C for ten minutes. The PCR products were finally purified using the GenepHlow™ Gel/PCR Kit (DFH100, DFH300) from FroggaBio (Concord, Ontario) according to the manufacturer's protocol.

The pTGEM vector and the left homologous arm were separately digested with AgeI-HF and NotI-HF (New England Biolabs), but using the same protocol: 1 µg DNA, 5 µL 10x Cutsmart Buffer, 1 µL AgeI-HF, 1 µL NotI-HF and Ultra Pure water to a total volume of 50 µL were added into a PCR tube, then incubated at 37 °C for 4 hours. The linearized vector was then 5'

dephosphorylated using Quick CIP (New England Biolabs) by mixing 2.5 µL into the sample and incubating it for a further 20 minutes at 37 °C before inactivation for 2 minutes at 80 °C. Both the linearized vector and the digested left homologous arm sample were then purified using the GenepHlow™ Gel/PCR Kit (DFH100, DFH300) from FroggaBio (Concord, Ontario) according to the manufacturer's protocol, in order to remove the excised DNA fragments from the solutions.

The Left homologous arm was ligated into the digested pTGEM vector using T4 DNA Ligase (New England Biolabs) and the following protocol: 100 ng vector, 35 ng Left homologous arm, 2 µL 10x DNA Ligase Buffer, 1 µL T4 DNA Ligase, and Ultra Pure Water to 20 µL were gently mixed on ice before being incubated for 12 hours at 16 °C. The enzymes were heat inactivated at 65 °C for 10 minutes.

The ligated pTGEM was added to NEB® 10-beta Competent E. coli (High Efficiency) cells (New England Biolabs) and transformed in accordance with the manufacturer's standard transformation protocol. One hundred microliters of the transformation mix were plated on LB+ampicillin plates (100 uL/mL) and grown overnight at 37 °C. Individual colonies were selected and grown in 1 mL of LB+ampicillin (100 uL/mL) for 24 hours. Plasmids were extracted from each cell culture using the Invitrogen PureLink™ Quick Plasmid Miniprep Kit according to the manufacturer's instructions.

Successful ligation of the left homologous arm was assessed in two ways. First, each plasmid sample was sequenced by Robart's DNA Sequencing Facility (London, Ontario) to verify successful ligation of the arm using T7 and T3 primers. Second, a small portion of each plasmid sample was single-digested with NotI-HF and compared to a linearized un-ligated control for the expected ~1 kbp size difference on a 1% agarose gel.

Successfully ligated pTGEM was collected and used to replicate the entire process from digestion to verification for the right homologous arm using the SpeI-HF and AscI-HF restriction enzymes (New England Biolabs) instead of AgeI-HF and NotI-HF. By running linearized samples through gel electrophoresis, the pTGEM plasmid with both arms ligated, approximately 10kbp in length, was easily distinguished from pTGEM containing only the left arm (~9kbp), and control pTGEM with neither arm in it (~8kbp; Figure 15A). The result (also verified by

sequencing at the London Regional Genomics Centre; Figure 15B, C) was an injection-ready stock of pTGEM(1) with both homologous arms flanking the Trojan-Gal4 construct.

The CRISPR/Cas9 injection mix was then prepared for injection into embryos from stock # 58492 and stored at -20 °C. The injection mix was made to a total volume of 20 uL, containing 0.1 µg/µL pU6-3-gRNA, 0.5 µg/µL ligated pTGEM(1), and the remaining volume was standard blue food colouring. The mix was aliquoted into four separate tubes of 5 µL each.

2.4 Embryonic microinjections

Embryonic microinjection entails using a light microscope and a microneedle manipulator to inject substances such as drugs or plasmids into *Drosophila* embryos, in order to assess their effects on development, or to induce mutagenesis in a line (Kiehart, Crawford, & Montague, 2007; Nagarkar-Jaiswal et al., 2015; Prokhorova et al., 1989).

2.4.1 Microinjection preparations

Approximately 500 flies were transferred to fly cages sealed with a standard cornmeal-based food plate, and acclimated for at least 24 hours in a 24 °C incubator set to 70% humidity and 14h:10h day:night cycle. The food plates were refreshed daily as needed. Loading needles, used to precisely fill injection needles with plasmid mix, were manually pulled from 1 mm x 0.7 mm borosilicate tubes over a Bunsen burner. Injection needles, used to inject plasmid mix into the embryos, were prepared from 0.5 mm borosilicate tubes using a Micropipette Puller P-97 needle puller (Sutter Instrument Company) obtained from Dr. Greg Gloor.

2.4.2 Embryonic microinjection protocol

Two hours prior to injections, the cornmeal food plates were swapped with apple juice agar plates topped with active yeast paste, and the cages were placed back into the incubator. The apple juice yeast plates were refreshed in 30 minute intervals for the duration of the two hours. During the wait, the injection needles were loaded with injection mix and inserted into the microinjector. Finally, the fly embryos were gently washed off the final plate using distilled water and a light paintbrush, and collected in baskets made of 1.5” wide mesh ribbon. The embryos in the basket were further washed off with distilled water, before gently being aligned in a vertical line with posterior ends to the left, on a 0.5 mm x 18 mm square silicone coverslip.

The embryos were then coated in olive oil and injected in their posterior ends while being viewed under a Nikon Stereo Microscope. Each round of injections concluded by washing the oil off of the embryos, and placing the edge of the coverslip into a food vial such that their anterior sides were in close proximity to the food. Each round of injections lasted for a maximum duration of 30 minutes, and was repeated as desired. The microinjected embryos were kept in a 24 °C incubator set to 70% humidity and a 14h:10h day:night cycle until pupation, upon which each pupae was placed in its own food vial awaiting eclosion.

2.5 Crossing of survivors and screening of offspring for transgenesis

2.5.1 PhiC31-mediated recombination upstream of *fru* A and B exons

Microinjection survivors from both the *fru* A (# 42145) and the *fru* B MiMIC stocks (#44345) were backcrossed in pairs to the same respective line, and offspring were screened for phiC31-mediated recombination of the Trojan-Gal4 construct into the genome. The screening was done visually under a basic optical microscope, as many transgenic offspring could be quickly identified by presence of the *yellow*-body phenotype as a result of the *yellow*-rescue, present in the original MiMIC site, being removed from the genome. This method of screening was possible because both MiMIC sites on the third chromosome appear to be homozygous lethal, similar to findings regarding MiMIC sites in other genomic locations (Venken et al. 2011), and the 3rd chromosome balancer, TM3, is also homozygous lethal (Nelson and Szauter 1992). Therefore, any *yellow*-phenotype offspring, or any offspring not expressing the TM3 visible marker *Stubble* should be transgenic.

2.5.2 CRISPR-mediated HDR upstream of *fru* C exon

Microinjection survivors from the *Actin*-Cas9 stock (# 58492) were crossed pair-wise to a balancing line (# 3703) that contains the 3rd chromosome balancers MKRS, and TM6B. This was done to ensure that Trojan-Gal4 integration found in the 3rd chromosome of any offspring, would be balanced, and thus potentially able to be maintained in a stable transgenic line.

Transgenic screening of offspring was done by visualizing late-stage pupae and adult flies for expression of Red Fluorescent Protein (RFP) in their eyes, derived from a gene-marker in the

Trojan-Gal4 construct. The Zeiss LUMAR stereoscope used for the screening was provided by the Biotron Facility (UWO, London, Ontario).

3. Results

3.1 *fru* expression in adult male and female *Drosophila melanogaster*

3.1.1 RT-PCR analysis of *fru* expression in adult male and female *Drosophila melanogaster*

The results of *fru* RT-PCR mostly matched the transcript combinations predicted via sequence analysis (Larkin et al. 2021). However, these predictions have not been empirically confirmed. Empirical studies have focused on which *fru* transcripts are expressed based on their 5' P exon, without assessing whether each 5' transcript class (P1-P5) contained all three possible 3' ends (A-C). Here, I detail which 3' exons are present for each *fru* 5' starting exon. Adult males express sex-specifically-spliced P1 transcripts with A, B, and C terminal exons (Figure 7). Females also express P1 transcripts of greater length than males due to the inclusion of the female-specific S-exon, which includes a stop codon prohibiting the formation of female P1 proteins. However, due to the greater transcript length, these transcripts were absent from the P1 RT-PCR results because the RT-PCR was optimized for smaller amplifications. Furthermore, both males and females express P2 transcripts ending in the A, B, and C exons, representative of their non-sex-specific splicing (Figures 7, 8). In addition to the expected P2 transcript sizes, I also found larger P2 transcripts that are also expressed in both males and females (Figures 7, 8, 9). These secondary P2 transcripts include each of the 3' terminal exons (A, B, and C; Figures 7, 8, 9).

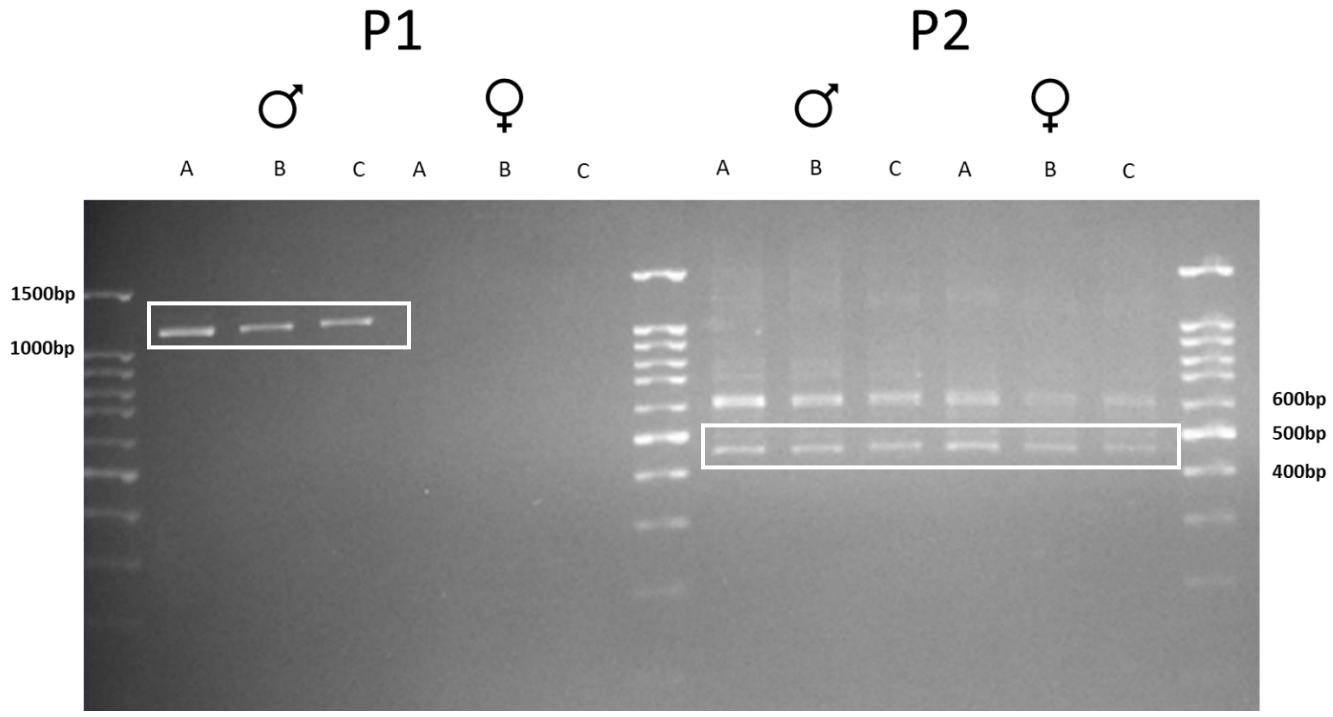

Figure 7: Sex-specific RT-PCR expression analysis of *fru* P1 and P2 transcripts containing 3' terminal exons A, B, and C. The mRNA was derived from full bodies of adult *Drosophila melanogaster*. Males express P1A, P1B, and P1C, while females do not. Males and females both express two forms of P2 transcripts, with variants of each ending in 3' terminal exons A, B, and C (compared to a 100 base pair ladder; expected fragment sizes (displayed in white boxes): P1: 1160bp, P2: 455bp).

As expected, P1-S transcripts were found in females, but not in males (Figure 8), because the majority of the P1-S exon's length is specific to females, and the forward primer used in this RT-PCR experiment was derived from that female-specific segment. As for the non-sex-specifically-spliced P3 and P4 transcripts, both were found expressed in males and females with variants of each terminating in the A, B, and C exons (Figure 8). Contrary to expectations, P5 transcripts which have been reported to be expressed in both adult males and females (Leader et al. 2018), were only found in males, with all three variants being represented (Figure 8). Aside from the P2 secondary bands, some samples such as P3B and P3C in males and females featured their own smaller, but prominent, secondary bands. In contrast to the P2 secondary bands, these were confirmed to be off-target amplifications of ribosomal cDNA via Sanger sequencing.

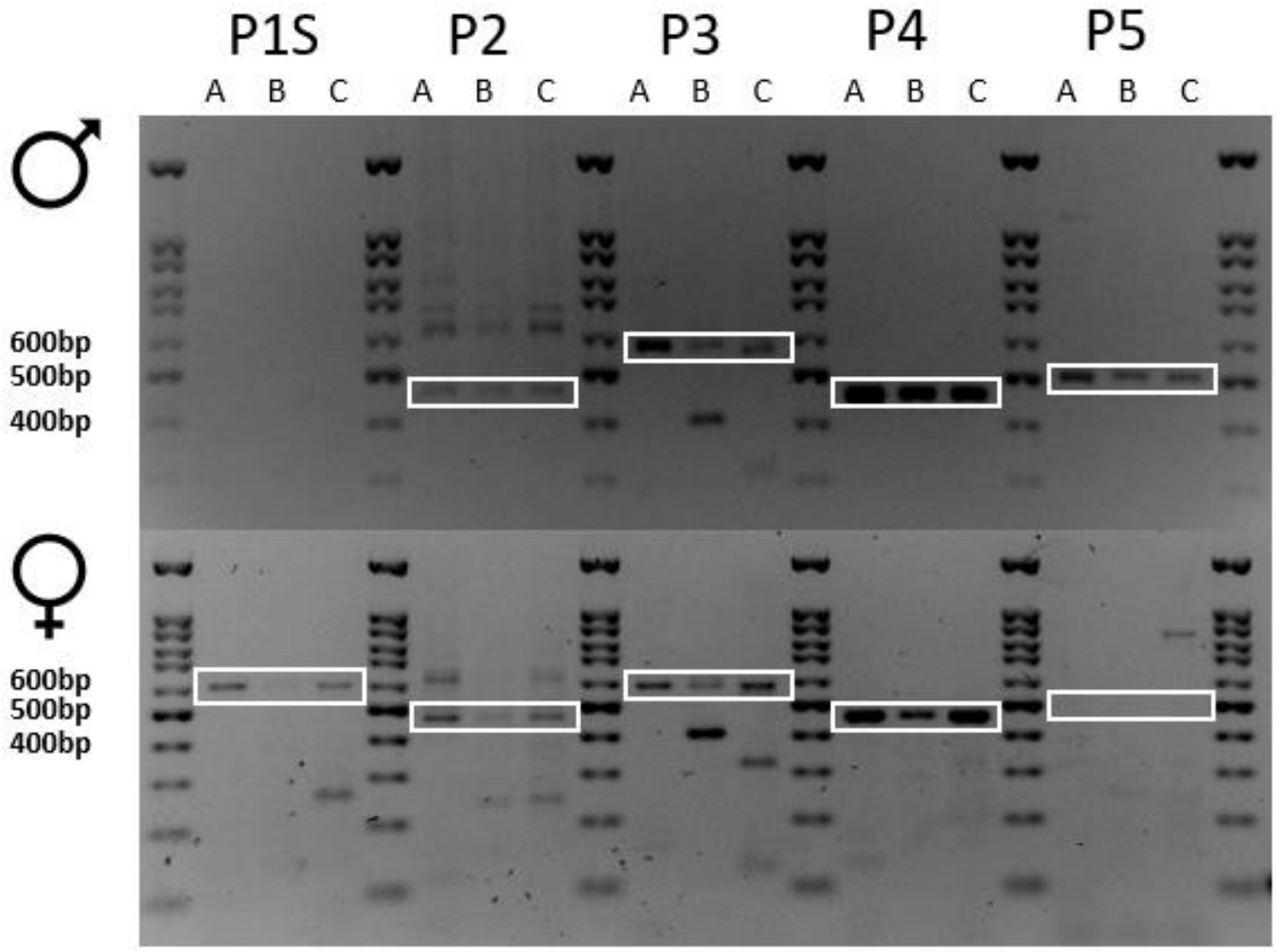

Figure 8: Sex-Specific RT-PCR expression analysis of fru P1-S-P5 transcripts containing 3' terminal exons A, B, C. The mRNA was derived from full bodies of adult *Drosophila melanogaster* (compared to a 100bp ladder; expected fragment sizes (displayed in white boxes): P1-S: 594bp, P2: 464 bp, P3: 573 bp, P4: 460bp, P5: 507 bp).

3.1.2 Sequencing data of visualized cDNA bands

In order to confirm the identities of each DNA band, I excised and purified one DNA band from each P1-S to P5 group and had them sequenced. By aligning the sequencing data of each band to the *fru* total RNA sequence, the identities of each sample were confirmed to be products from the expected first exons. For example, the P1-S transcript included a portion of the P1-S exon perfectly aligned to the RNA template in its expected location (this location being the most 5' of the P exons sequenced), followed by the P2 transcript following the same pattern (the second

most 5' P exon sequenced), then the P5, P3, and P4 transcripts in subsequent order. Furthermore, all of the fragments featured almost perfect alignment at the 3' end where their sequences were identical, representing the span of common region they all share and ending in the sequence of the common reverse primer used to amplify the fragments.

Alignment of the secondary, larger P2 band was more difficult because the fragment does not align well to the *fru* RNA template, or to any previously identified *fru* transcripts. This is due to the inclusion of an additional *fru* P2 exon (Figure 9B, C) that has not previously been reported in any *fru* transcripts. This additional P2 exon creates a larger P2 transcript but also contains a stop codon, making the protein produced from this secondary P2 transcript very small (Figure 9D).

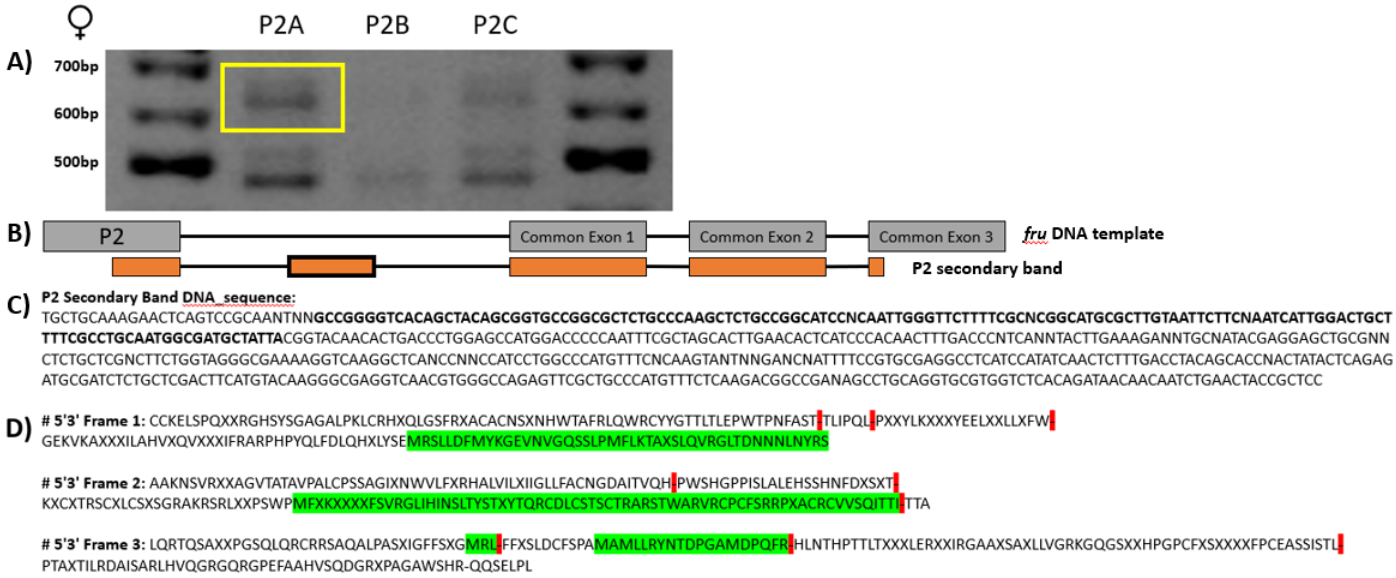

Figure 9: (A) Visualization of the secondary P2 transcript band (encompassed by the yellow rectangle) derived from RT-PCR using RNA extracted from adult female D. *melanogaster*. **(B)** Schematic of P2 secondary band sequence (orange) aligned to the *fru* sequence (grey). Each box represents an exon found in their respective sequences, and as shown, the secondary band contains an unidentified exon (bolded orange box) in between the canonical P2 and common exons. The P2 secondary band sequence only covers the regions in between the P2 forward primer and common exon 3 reverse primer, which explains why its sequence abruptly ends on either end. **(C)** DNA sequence of the *fru* P2 secondary band with the potential new *fru* exon shown in bold. **(D)** Amino acid translations in all three possible reading frames of this new *fru* P2 transcript. The green highlighted and underlined segments represent possible open reading frames, and the red highlights represent stop codons.

3.2 Microinjections for Trojan-Gal4 integration via PhiC31-mediated recombination

3.2.1 Efficiency of transgenesis upstream of the A and microinjection survivorship

Following the injection of 534 *D. melanogaster* embryos for PhiC31 recombination of the A MiMIC site, 103 survived to adulthood (Figure 10). This gives an overall survival percentage of 19.29% (Figure 11). Adult survivors were mated with a balancer stock for offspring screening. None of the offspring displayed transgenesis indicated by expression of the *yellow*-body phenotype, a mutant phenotype that adds yellow pigmentation to their wings and skin.

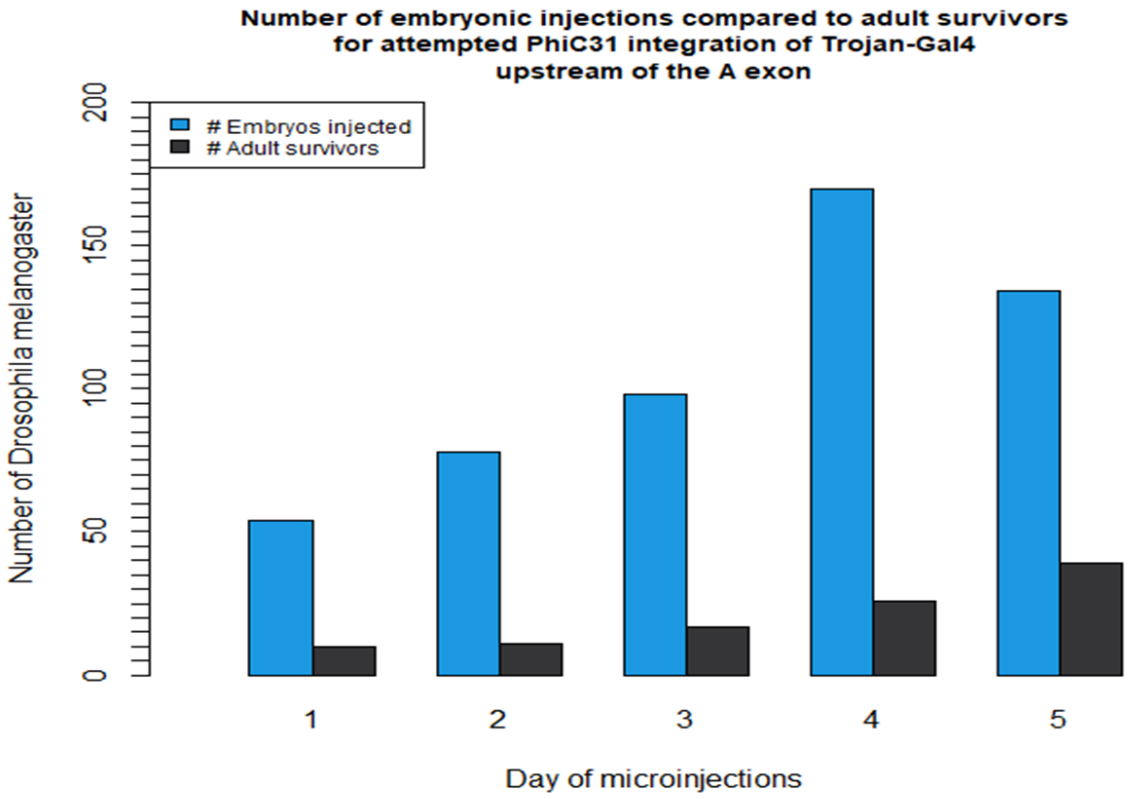

Figure 10: Bar plot depicting the number of *Drosophila melanogaster* embryos (Stock # 42145) microinjected by day, for the purposes of Trojan-Gal4 integration 5' of the *fru* A exon via PhiC31-mediated recombination. Blue bars show the number of embryos injected, and grey bars show the number of adult survivors from that day of injections. Days 1 and 2 were injected in March prior to COVID-19 lab shutdown. Day 3 to 5 embryos were injected in the subsequent fall and housed outside of their usual temperature, humidity, and light-controlled incubators due to sanitation concerns. Day 5 post-injection embryos were housed in cotton-sealed food vials rather than on food plates.

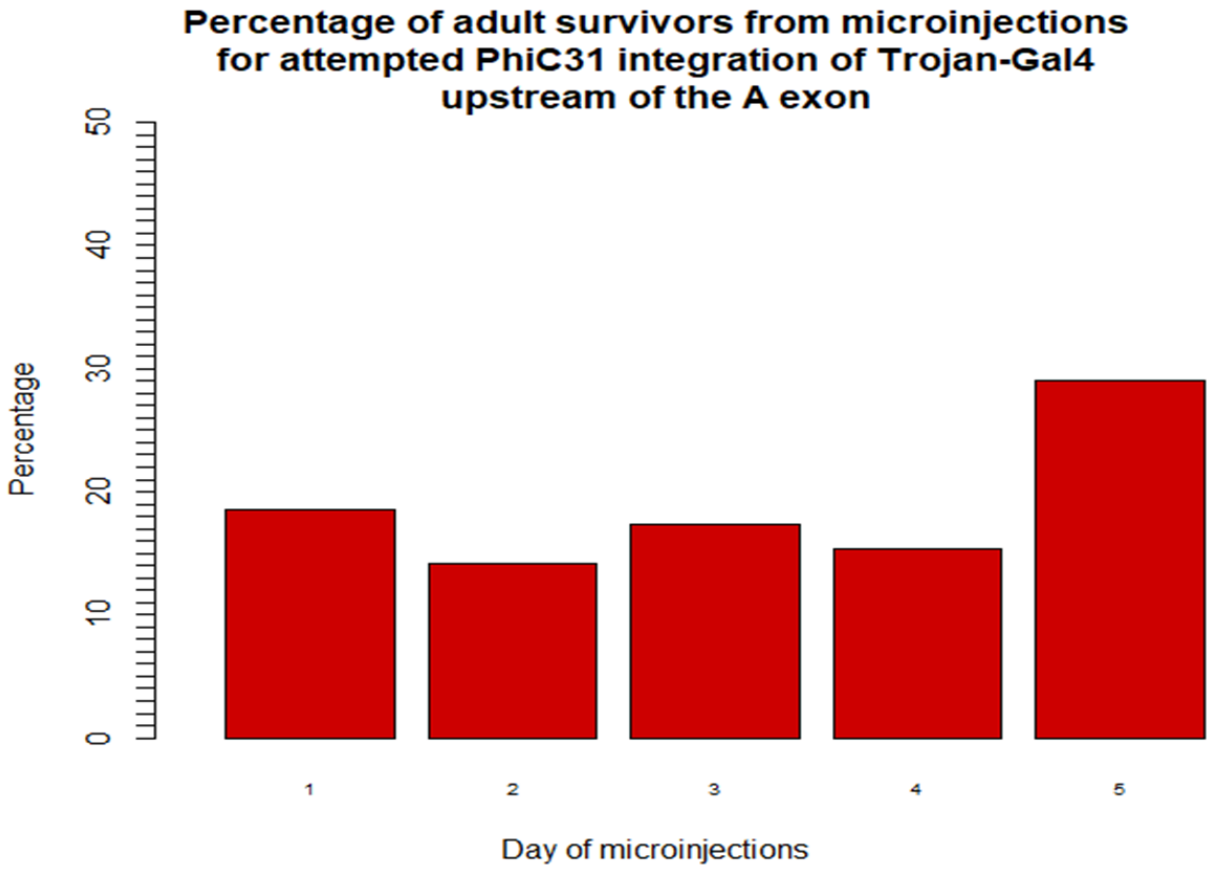

Figure 11: Bar plot depicting the percent-survival of *Drosophila melanogaster* embryos (Stock # 42145) microinjected by day, for the purposes of Trojan-Gal4 integration 5' of the fru A exon via PhiC31-mediated recombination. Days 1 and 2 were injected in March prior to COVID-19 lab shutdown. Day 3 to 5 embryos were injected in the subsequent fall and housed outside of their usual temperature, humidity, and light-controlled incubators due to sanitation concerns. Day 5 post-injection embryos were housed in cotton-sealed food vials rather than on food plates.

3.2.2 Efficiency of transgenesis upstream of the B exon and microinjection survivorship

Following injection of 416 *D. melanogaster* embryos for PhiC31 recombination of the B MiMIC site, 84 survived to adulthood (Figure 12). This gives an overall survival percentage of 20.19% (Figure 13). Adult survivors were mated with a balancer stock for offspring screening. None of the offspring displayed transgenesis indicated by expression of the *yellow*-body phenotype.

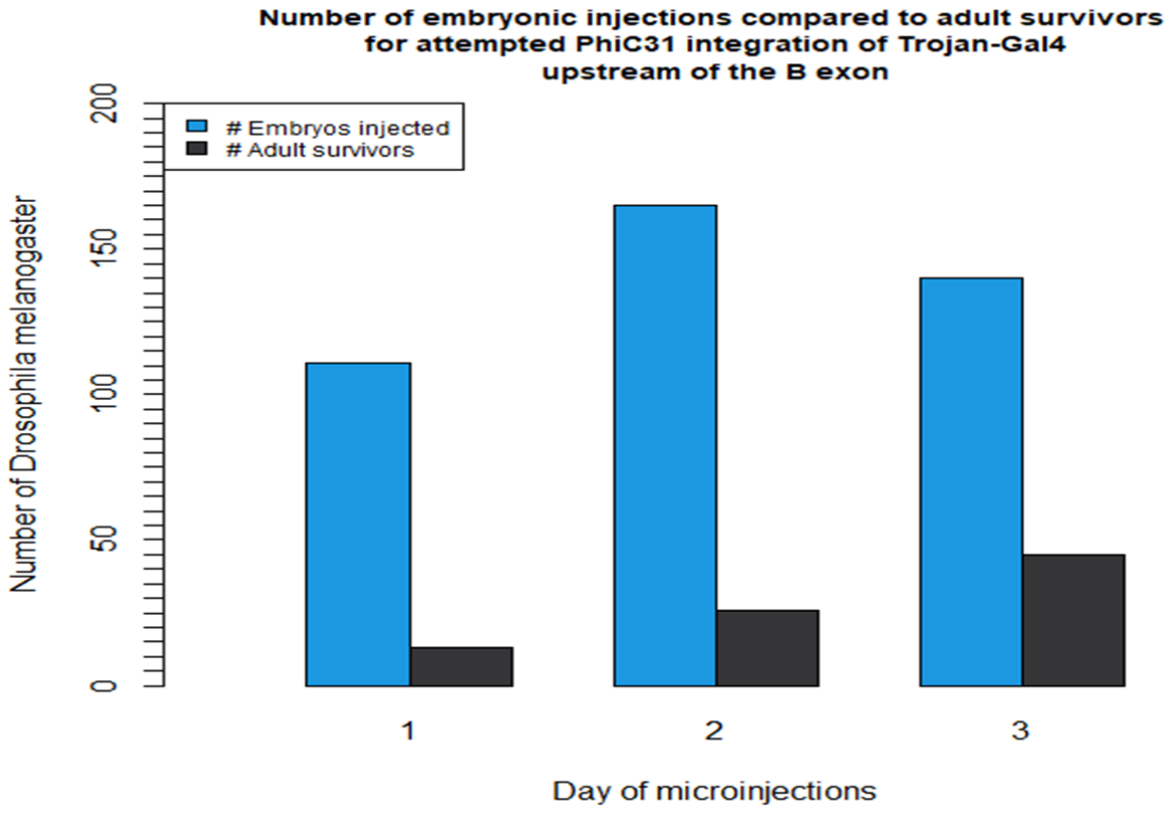

Figure 12: Bar plot depicting the number of *Drosophila melanogaster* embryos (Stock # 44345) microinjected by day, for the purposes of Trojan-Gal4 integration 5' of the *fru* B exon via PhiC31-mediated recombination. Blue bars show the number of embryos injected, and grey bars show the number of adult survivors from that day of injections. All injection days were in the Fall of 2020. Day 3 embryos were housed in cotton-sealed food vials rather than on food plates, and outside of their usual temperature, humidity, and light-controlled incubators due to sanitation concerns.

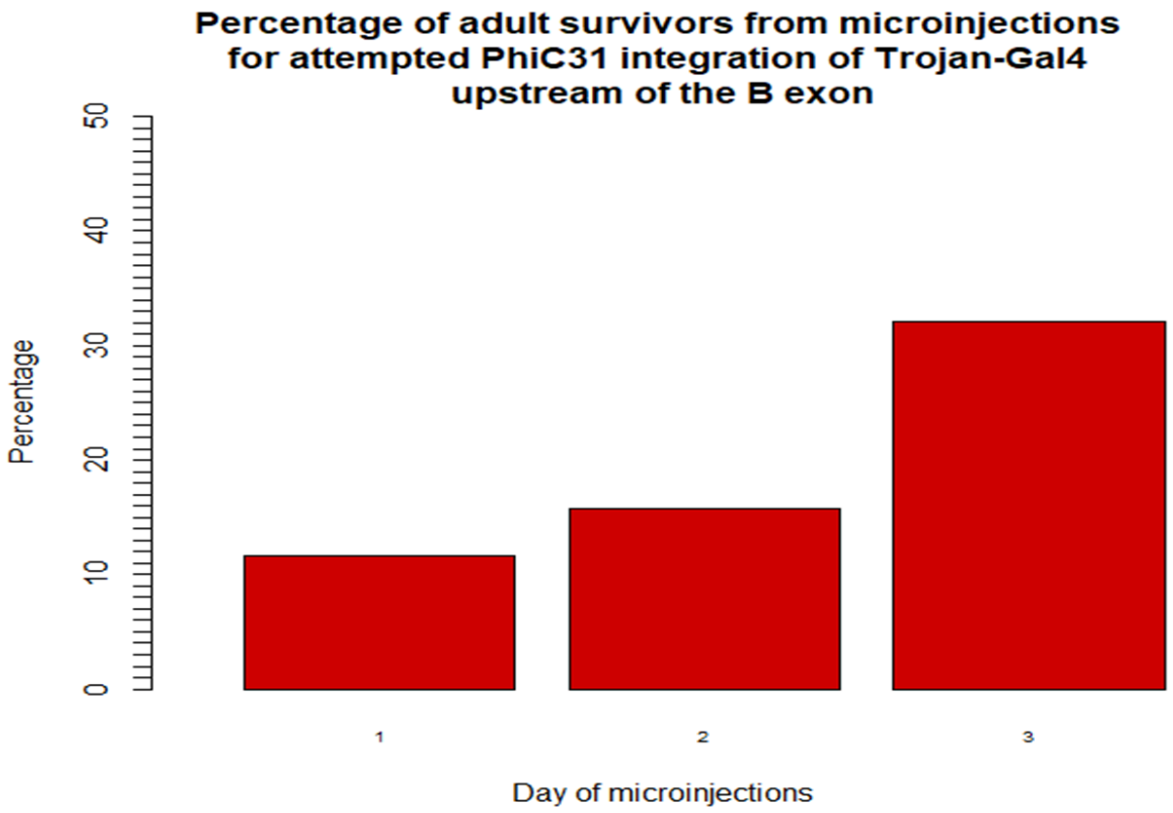

Figure 13: Bar plot depicting the percent-survival of *Drosophila melanogaster* embryos (Stock # 44345) microinjected by day, for the purposes of Trojan-Gal4 integration 5' of the fru B exon via PhiC31-mediated recombination. All injection days were in the Fall of 2020. Day 3 embryos were housed in cotton-sealed food vials rather than on food plates, and outside of their usual temperature, humidity, and light-controlled incubators due to sanitation concerns.

3.3 Microinjections of CRISPR/Cas9-mediated HDR

3.3.1 Preparation of plasmids used for CRISPR/Cas9 injections

The desired protospacer sequence was confirmed via Sanger sequencing to be present in the genome of the two possible Cas9-expressing injection strains of *D. melanogaster* (Figure 14A, B). The gRNA vector (PU6:3 gRNA) was successfully cloned with this necessary protospacer sequence, confirmed via Sanger Sequencing (Figure 14C).

Figure 14: Protospacer integration into pU6 guide RNA plasmid. Geneious software analysis showing a 100% identity match (indicated by solid green bar above paired base-pair sequences) between theoretical protospacer sequence and actual genomic protospacer sequence in (A) nanos-Cas9 (a back-up line that exclusively expresses Cas9 during embryonic development) and (B) Actin-Cas9-expressing *Drosophila melanogaster*. Successful integration of protospacer sequence (shown by green bar labeled "PU6 C Intron Target 1 Insert") into pU6 guide RNA plasmid (C) confirmed by Sanger sequencing of insert between BSAI restriction digest cut sites.

The homology arms necessary for homology-directed repair were successfully cloned into Trojan-Gal4 donor plasmids of all three phases, encompassing all six possible reading frames of the gene (Figure 15A). I had already identified phase one as the correct plasmid for the desired reading frame of my gene, however, I created plasmids of the other two phases as well in case they were needed. Successful insertion of the left (Figure 15B) and right (Figure 15C) homology arms flanking the Trojan-Gal4 construct was confirmed via Sanger Sequencing.

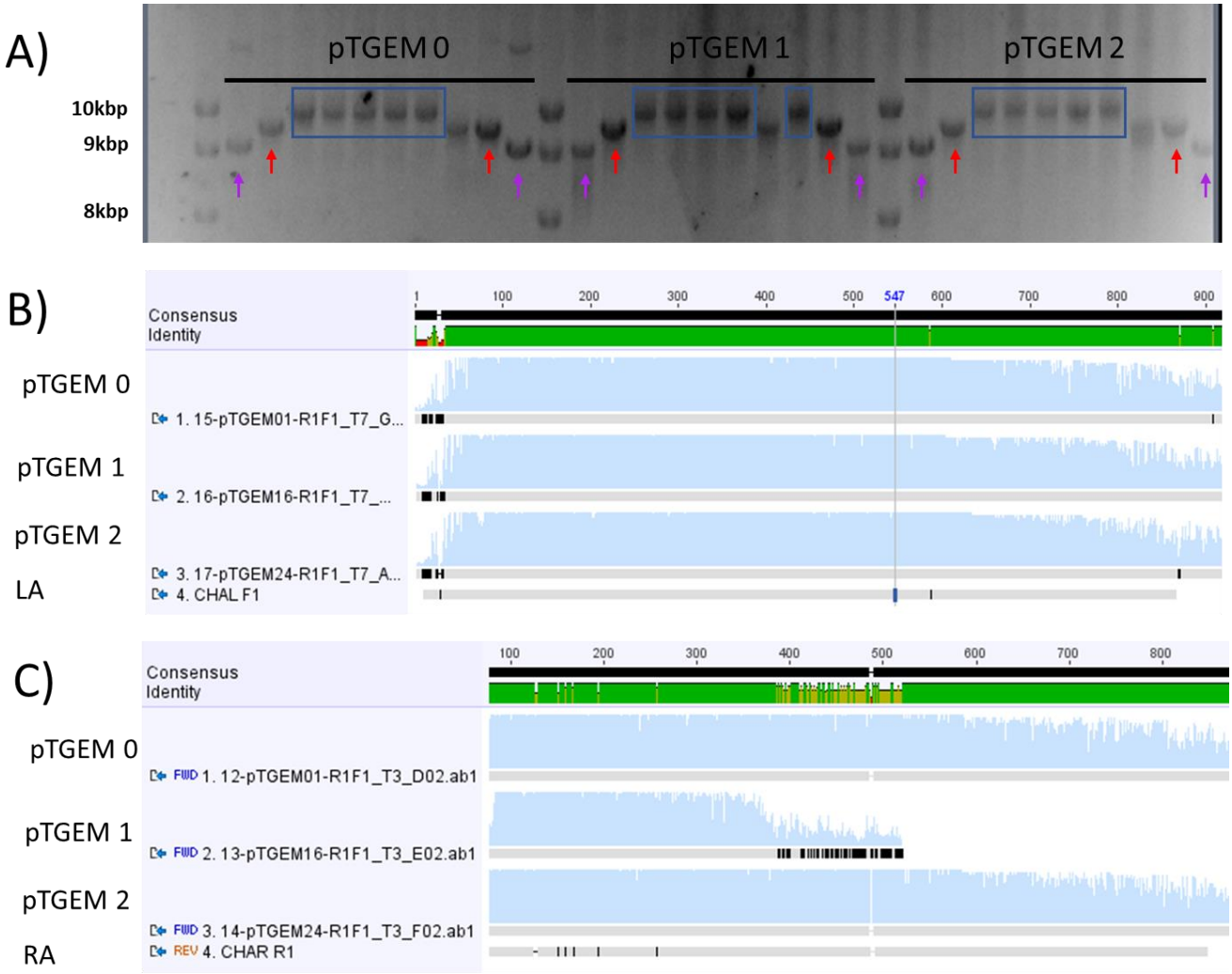

Figure 15: *Drosophila melanogaster* homology arm insertion into Trojan-Gal4 expression module plasmid (pTGEM) of all three phases. (A) Gel electrophoresis displaying a successful insertion of both the left (953 base pairs) and right (834 base pairs) homology arms derived from Cas9-expressing flies into pTGEM phase 0, 1, and 2 vectors (DNA bands within the blue squares that measure around 10kbp) Red arrows indicate control plasmids (measuring approximately 9kbp in total) with only the right arm inserted, and purple arrows indicate control plasmids (measuring approximately 8kbp in total) with no homology arms inserted for comparison. There are 1kb DNA ladders separating each phase of pTGEM (0, 1, and 2), as well as in the left-most lane of the gel with key sizes indicated. Geneious software was used to align sequencing data with the plasmid vector template and confirm homology arm integration into each phase of pTGEM. Each phase was sequenced for insertion of the left (B) and then the right (C) homology arms flanking the TGEM construct, and verified by comparing their actual sequences to the theoretical homology arm sequences (Left Arm: LA, Right Arm: RA) derived from the wildtype genome of *Drosophila melanogaster* of the canton S strain. The sequence identities of the inserted homology arms from the Cas9-expressing D. *melanogaster* were almost identical to the sequences derived from wildtype D. *melanogaster* (indicated by the green bars at the top of each image) except for several potential single nucleotide polymorphisms.

3.3.2 CRISPR/Cas9-mediated Homology-Directed Repair survival rate and efficiency

3.3.2.1 Efficiency of transgenesis

Following the injection of 3957 actin-Cas9-expressing *D. melanogaster* embryos, 646 survived to adulthood and were paired with a balancer stock for offspring screening. None of the offspring displayed transgenesis indicated by Red Fluorescent Protein expression in their eyes.

3.3.2.2 Survival rate following microinjections

Over the course of 28 days spent microinjecting (Figure 16), the total number of injected embryos was 3957. Of these 3957 embryos, 646 of them survived to adulthood which gives an overall survival rate of 16.33% (Figure 17). However, survival rate varied greatly by day of injection, depending on the post-injection conditions of the embryos, as well as the amount of practice I had (Figure 17).

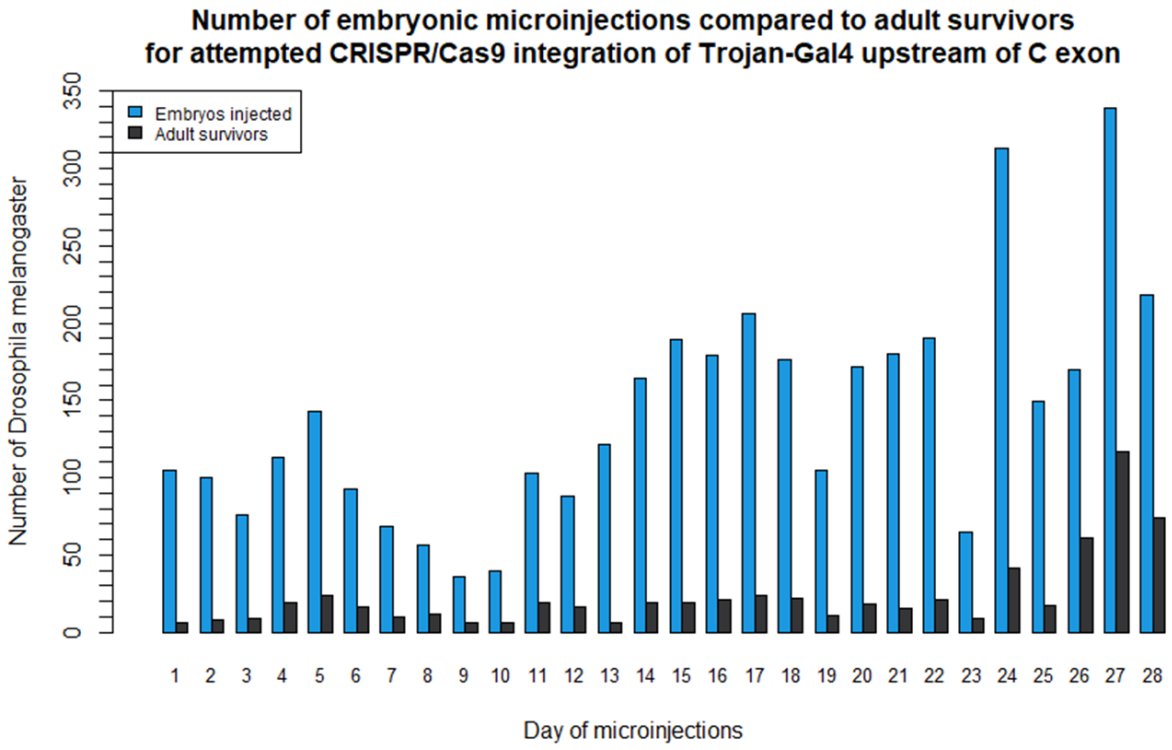

Figure 16: Bar plot depicting the number of *Drosophila melanogaster* embryos (Stock # 58492) microinjected by day, for the purposes of Trojan-Gal4 integration 5' of the fru C exon via CRISPR/Cas9-

mediated Homology-directed repair. Blue bars show the number of embryos injected, and grey bars show the number of adult survivors from that day of injections.

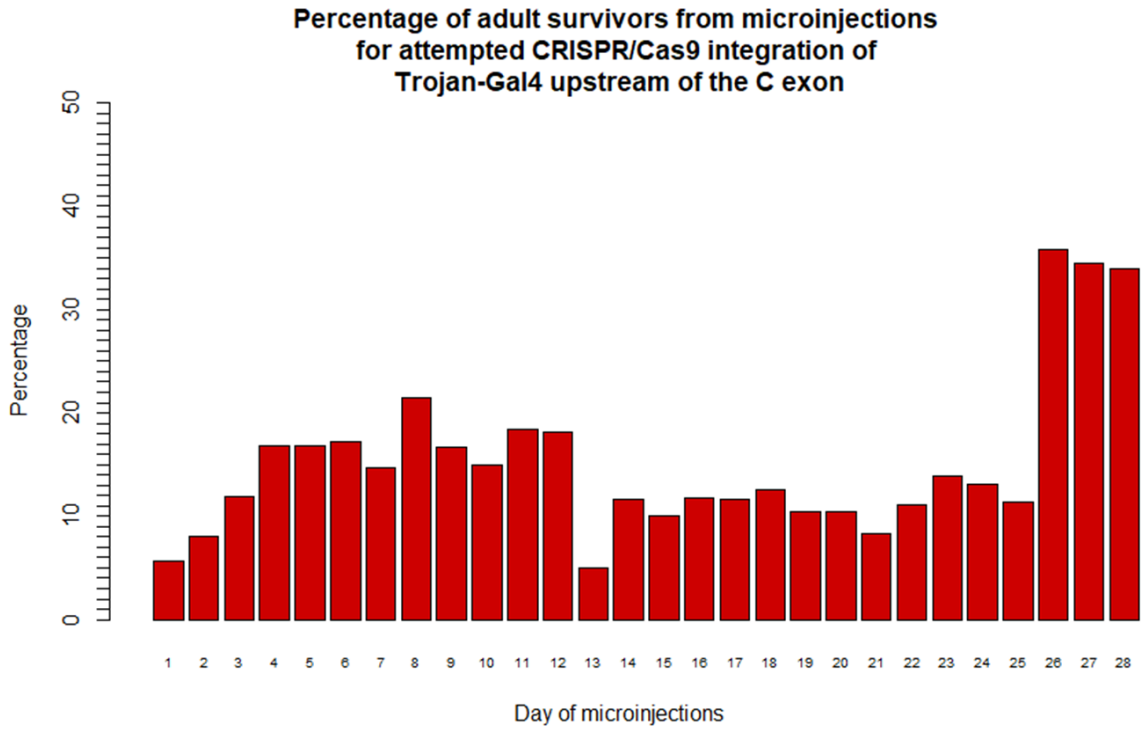

Figure 17: Bar plot depicting the percent-survival of *Drosophila melanogaster* embryos (Stock # 58492) microinjected by day, for the purposes of Trojan-Gal4 integration 5' of the fru C exon via CRISPR/Cas9-mediated Homology-directed repair. Day 13 represents a possible outlier for survivorship, as these embryos were accidentally stored in the fumes of paper that is poisonous to mites and perhaps other organisms. Consequently, to avoid these fumes, Day 14 marks the first day that embryos began to be housed on food plates outside of their usual temperature, humidity, and light-controlled incubators. Day 26 marks the day that embryos began to be housed in cotton-sealed food vials rather than on food plates, but remained outside of their incubators.

4. Discussion

4.1 Exploration of *fru* expression in male and female *Drosophila melanogaster*

I generated a reliable profile of the *fru* transcripts expressed in adult *D. melanogaster* (Figures 7, 8, 9). As expected, I confirmed that the canonical P1 and P1-S transcripts are sex-specifically spliced. I found that all five first exon classes (P1-P5) produced transcripts with all three 3' ends (A-C). However, P5 transcripts were absent in females but present in males, suggesting a potential sex-specific role for this transcript class. The P5 class of *fru* transcripts were only recently discovered, and therefore their functions remain relatively unknown. Indeed, the P5 exon was previously thought to be a second exon of P2 until research showed that they are

expressed entirely independently of each other in separate *fru* transcripts (Larkin et al. 2021). High-throughput assays show that despite overall low expression, P5 transcripts are mostly found in the male head, eyes, and testes and in the female brain (FlyAtlas2: Leader et al. 2018). Other areas of the body in both sexes show low to negligible expression. This may explain why I was unable to amplify P5 transcripts from females (Figure 8), as their low expression levels make them unsuitable targets for exon-specific RT-PCR amplification. Nonetheless, the differential expression of P5 transcripts between sexes, combined with their unknown functions, may make them a worthwhile topic for future studies of genetic influences on sexual dimorphism and behaviour.

Surprisingly, I discovered a new exon of *fru*, specifically associated with transcripts derived from the P2 first exon (Figure 9). Transcripts starting with the P2 exon are particularly interesting since the P2 transcripts have been shown to affect female receptive behaviour (Chowdhury et al. 2020). It is currently unknown if or how this newly discovered splice variant influences phenotypes related to sexual behaviour, or development, as other *fru* transcripts do. Furthermore, my sequencing results only go so far as to determine the nucleotide composition from the transcripts' initial 5' exon to the second common exon. It is therefore unknown whether beyond these two common exons, there is more splicing that occurs in the latter 3' segments that I have not explored. It would be worthwhile to pursue the amplification and eventual sequencing of this transcript in its entirety, using a high-fidelity DNA polymerase capable of amplifying long sequences. In addition to the new P2 exon I describe above, two alternatively spliced *fru* microexons (exons under 30 nucleotides) were also recently discovered that are expressed under different circumstances (Pang et al. 2021). The first, ME756, is expressed exclusively in the head and codes for a premature stop codon that decreases the quantity of FRU protein expression. The second, ME177, has currently unknown function, but is male-specific. Although the *fru* exon I identified is longer than these microexons, and not sex-specific, it, too, may provide an unidentified alternative function for the *fru* gene.

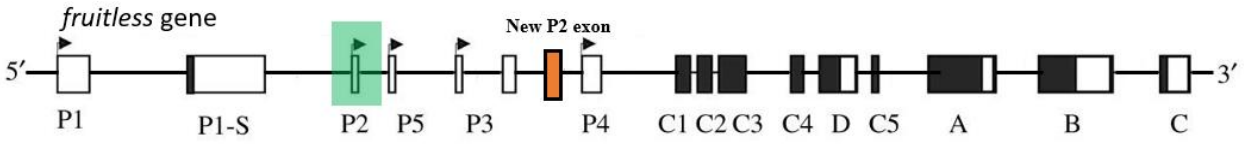

Figure 18: Schematic showing the relative positions of the exons of the *fru* gene within the genome, including the newly identified exon (orange box). Boxes represent exons, and black portions represent coding regions. Transcripts derived from *fru* contain one of the 5' P1-P5 exons, the common exons, and then one of the 3' A-C exons. Transcripts that contain the 5' P4 exon may include the 3' D exon. The newly identified exon is being associated with the 5' P2 exon (highlighted in green) because it was identified within transcripts that contain the P2 exon in their 5' end. It is currently unknown if the newly identified P2 exon is coding.

New splice variants of *fru* unveil new possibilities for the exploration of the gene's functions in sexual behaviour, development, or other unknown traits. The functions of these secondary P2 transcripts cannot currently be speculated on, but preliminary analysis has revealed some pertinent information.

By passing the nucleotide sequence of the transcript fragment through an amino acid convertor (ExPASy Translate tool), the amino acid sequences of theoretical peptide products were examined (Figure 9D). Due to the prevalence of stop codons, any possible open reading frames are relatively short in length compared to the reading frames found in canonical *fru* transcripts. This makes it unlikely, but not impossible, that these secondary transcripts create proteins influencing behavioural or developmental phenotypes, or at least suggests an alternative mechanism of action for any proteins created. Generally, proteins comprised of less than one hundred amino acids are categorized as "small proteins" (Su et al. 2013). Most of these proteins have been largely ignored due to their lack of obvious function; however, there are examples of small proteins with important roles in cell signaling and structure (Hobbs et al. 2011; Storz et al. 2014). Therefore, despite the differences from canonical *fru* transcripts, the possibility remains that the secondary P2 transcripts could potentially translate into functional proteins.

Nucleotide alignments to the *fruitless* genomic sequence and its known transcripts (NCBI Blast) show that a portion of these new transcripts are mRNA derived from what was thought to be a non-coding intron (Figure 9B, C). High-throughput whole-genome expression analyses reveal that this novel mRNA segment appears to be expressed moderately in the male eye, and to a much lesser extent in the female brain and head (FlyAtlas2: Leader et al. 2018). My targeted transcript analysis identifies this region of the genome as an alternative exon of the P2 transcript

class of *fru.* Furthermore, sexually dimorphic expression patterns may be an indicator of sex-specific transcript use, and add incentive for continued investigation. However, further study would be needed to verify whether this is the discovery of one or more new exons in the *fru* gene, or if these segments, or the transcripts in their entirety, operate as non-coding RNA (ncRNA) with unknown functions.

Non-coding RNA is simply defined as an RNA transcript that is not coded into a protein (Mattick and Makunin 2006). A growing body of literature is currently attempting to delineate the importance and functions of ncRNA within a plethora of species, including humans (Mattick and Makunin 2006; Liu et al. 2015; Guttman and Rinn 2012). Non-coding RNAs are ubiquitous throughout eukaryotic species, and yet there is debate on their importance. Some researchers proclaim that because of their prevalence, but comparative lack of established function, the default assumption upon discovering a ncRNA is that it is genetic "junk" with little biological purpose (Palazzo and Lee 2015; Hüttenhofer et al. 2005). In contrast, some researchers extol the study of ncRNAs as many have been implicated in the regulation of crucial functions such as: gene expression, RNA editing and splicing, chromatin remodeling, and epigenetic regulation (Costa 2008; De Lucia and Dean 2011; Mattick and Makunin 2006).

In the *Drosophila* genus, ncRNAs are widespread (Brown et al. 2014). Functional ncRNA transcripts derived from newly identified introns are constantly being identified, partially due to copious possible permutations of promoter regions and splice sites (Brown et al. 2014). Furthermore, among *Drosophila* species, many non-coding RNA appear to be under strong selective pressure, supporting the hypothesis that many ncRNA have critical functions, leading to their evolutionary conservation (Clark et al. 2007).

Reflecting on prior work assessing genomic variability within the *Drosophila* genus, an enlightening next step regarding the discovery of these secondary P2 transcripts would be to use the same RT-PCR methods to see if the conserved transcripts are also present in other species. These results may elucidate whether there is any relationship between these new transcripts and rejection behaviours in *Drosophila* females, and perhaps will aid with the assessment of whether these transcripts act as ncRNA, or as translated proteins.

Normally, one of the traditional methods of determining if an RNA is translated into protein is via a Western Blot. Western blot assays rely on the creation of fluorescent antibody probes that

bind to proteins of interest and provide semi-quantitative expression results (Mahmood and Yang 2012). This method could not be used to screen for proteins derived from our potential new exon, because the exon's short length is prohibitive for the development of a suitable antibody probe for its theoretical amino acid sequence.

Confirming the identity of an ncRNA is a rigorous process, however, it has been made easier in recent years. Contemporary research often uses computational analyses of a genome to identify candidate sequences that share characteristics with known ncRNAs (Weinberg et al. 2007, 2017). The most promising candidates are then tested using experimental methods for their functions in an organism. After software analysis, if this new transcript is deemed a likely ncRNA, its potential functions can be explored in more detail within *D. melanogaster* via methods such as a targeted knockdown using RNAi.

4.2 Troubleshooting PhiC31-mediated recombination

For a variety of different MiMIC sites, the efficiency of PhiC31-mediated recombination was found to range from 3-33% (Diao et al. 2015). In theory, it is extremely unlikely, but still possible, that I did not end up with enough adult microinjection survivors to guarantee at least one instance of successful Trojan-Gal4 integration in the A and B exon lines. However, due to this limited possibility, I explored other potential issues in the protocol.

I first used PCR to verify that the PhiC31 expression vector indeed contained the gene for PhiC31 integrase via PCR (unpictured). In addition, I had prior work attempting to create these transgenic stocks using multi-generational crosses of genetically-modified *D. melanogaster*, that contained a different Trojan-Gal4, and that construct also failed to integrate (Appendix Figure 1). Since neither Trojan-Gal4 construct was integrated, the Trojan constructs are unlikely culprits for unsuccessful transgenesis.

The most likely remaining possibility for error is the attP sites flanking the MiMIC construct, which are responsible for recognition and catalyzation of the PhiC31-mediated gene swap. If these recognition sites are mutated in any way, the PhiC31 integrase will no longer recognize the MiMIC site as a potential region for cassette exchange, and no recombination will occur. This should be checked via PCR amplification of the attP sites and subsequent Sanger sequencing to verify that no mutations have occurred, by injecting a construct known to successfully integrate

into an attP site into these flies, and/or by injecting my construct into a fly strain containing a MiMIC site known to have functioning attP sites.

4.3 Optimizing CRISPR/Cas9-mediated HDR

Targeting a specific cut site with CRISPR/Cas9 can be imperfect. The guide RNA produced from the gRNA plasmid usually leads the Cas9 protein to its complementary genomic site; however, there is a tolerance of base pair mismatches that may allow the gRNA to guide the Cas9 protein to an undesired off-target site (Mali et al. 2013). Among other mutagenic results related to DNA breaks, this can result in the integration of a sequence of interest into an off-target position, and cause false positives when screening the animals for your desired integration.

In order to avoid off-target Cas9 cuts, I truncated my gRNA sequence from a typical 20-nucleotide sequence to a truncated 18-nucleotide sequence, and included a 5' G as detailed in the methods. However, there is conflicting literature debating whether increased specificity due to a truncated gRNA is traded for efficiency in on-target Cas9 cleavage. For example, this trade-off has been reported in human embryonic kidney cells (Pattanayak et al. 2013), while in bovine cells, cleavage efficiency decreased proportionally to truncation of the gRNA, with some notable exceptions (Zhou et al., 2019). One study in yeast (*Saccharomyces cerevisiae*) even found that truncated gRNAs reduce efficacy of CRISPR modifications, while simultaneously not providing benefits to specificity of cleavage (Smith et al. 2016). In contrast, another study using human bone osteosarcoma epithelial cells found that truncated gRNAs of 17 or 18 nucleotides in length were equally- or more efficient than 20 nucleotide gRNAs in catalyzing genomic edits (Fu et al. 2014). Further studies stress the importance of cut-site location and cell-type on efficient cleavage using truncated gRNAs, with regions closer to promoters being more efficient, and diverse cell groups showing diverse efficacies (Zhang et al. 2016).

Essentially, gRNAs play a large role in successful CRISPR/Cas9 modifications, and there is still doubt as to the design of the perfect gRNA for any individual situation. This may be one of the reasons my CRISPR/Cas9-mediated HDR failed. As a result, I have created a second gRNA plasmid using the same method as the first, however, this one has a 20-nucleotide sequence instead of an 18-nucleotide sequence. This new plasmid can be used in the future to re-try my CRISPR/Cas9 modification experiment.

4.4 Future directions

Once *Drosophila melanogaster* lines are created that contain the Trojan-Gal4 constructs integrated into their respective desired loci, behavioural analyses of the functions of *fruitless* P2 transcripts may begin. Using the UAS system described in the introduction, linked response genes can be expressed simultaneously with the *fru* A, B, or C exons, through the actions of functional Gal4 proteins. For example, a response gene expressing a fluorophore can allow for easy visualization of the neurons expressing the A, B, or C transcripts. This can be paired with a separate gene or neuron of interest with a different fluorophore, providing a relatively simple means of screening for overlap.

Once the *fru* A, B, and C neurons are visualized, further crosses can be done to pair the Gal4 system with UAS-linked *shibire*, which silences neurons (Gonzalez-Bellido et al. 2009), and *TrpA1*, which hyper-activates neurons. *Shibire* is a gene that encodes for a protein similar to the mammalian protein dynamin, which inhibits endocytosis of post-synaptic vesicles and thus prevents neurons from recycling them for continued use (Poodry and Edgar 1979; Van Der Bliek and Meyerowrtz 1991). A temperature-sensitive allele of this gene can be used to temporally control the silencing effect of *Shibire*. When activated by the Gal4 system in its permissive temperature, *Shibire* expression will silence the neurons expressing the *fru* A, B, or C isoforms, as they will eventually cease to release neurotransmitters into their associated synaptic junctions (Gonzalez-Bellido et al. 2009). *TrpA1* is a gene that encodes the temperature-sensitive TRPA1 ion channel located on the plasma membrane (Berni et al. 2010). Over-expression of TRPA1 channels in the relevant neurons will stimulate more frequent action potentials in warm conditions leading to overexcitation (Kilian et al. 2017)

Behavioural studies using the *D. melanogaster/D. simulans* hybrid system described in the introduction and methods should be used to assess the influence of each *fru*-expressing neuronal subset on female heterospecific rejection. *GAL4*-UAS *D. melanogaster* females can be crossed to wildtype *D. simulans* males to form female *GAL4-Shibire*, and *GAL4-TrpA1* hybrids. As a consequence of TrojanGal4's self-splicing ability, the *D. melanogaster fru* allele will form truncated, non-functional protein whenever the Gal4 is expressed (Diao et al. 2015), allowing for the unmasking of isoform-specific *D. simulans* behaviour, as previously discussed. To determine the effects of silencing and over-excitation of the *fru* A, B, and C-expressing neurons, hybrids

should be subjected to comprehensive behavioural assays assessing their responses to male *D. melanogaster* courtship attempts and encompassing factors such as egg-laying, courtship duration, and active female rejection behaviours (Chowdhury et al. 2020; Demir and Dickson 2005; Neville et al. 2014). These experiments may greatly expand our understanding of genetic influences on heterospecific female rejection behaviours, and behavioural barriers of species isolation.

5. Conclusions

I have created an expression profile of *fruitless* transcripts in both male and female adult *Drosophila melanogaster* based on both the 5' initial exon and 3' terminal exon. My work complements the existing literature which generally focuses on the expression patterns of transcripts based on either the 5' or 3' exons alone. Furthermore, I have identified a new P2 transcript with unknown function, containing DNA that was formerly thought to be non-transcribed. In addition, I have created injection mixes of plasmid vectors that can be microinjected into embryos to integrate Trojan-Gal4 constructs into the *Drosophila melanogaster* genome via either PhiC31 recombination, or CRISPR/Cas9-mediated HDR. Although my own Trojan-Gal4 integrations were unsuccessful to date, my efficiency and skill with the microinjection protocol has increased over time, preparing me for future endeavours.

Works Cited

Anand A, Villella A, Ryner LC, Carlo T, Goodwin SF, Song HJ, Gailey DA, Morales A, Hall JC, Baker BS, et al. 2001. Molecular genetic dissection of the sex-specific and vital functions of the Drosophila melanogaster sex determination gene fruitless. *Genetics* **158**: 1569–1595.

Anholt RRH, O'Grady P, Wolfner MF, Harbison ST. 2020. Evolution of reproductive behavior. *Genetics* **214**: 49–73.

Baker RJ, Bradley RD. 2006. Speciation in mammals and the genetic species concept. *J Mammal* **87**: 643–662.

Barbash DA. 2010. Ninety years of Drosophila melanogaster hybrids. *Genetics* **186**: 1–8.

Berni J, Muldal AM, Pulver SR. 2010. Using neurogenetics and the warmth-gated ion channel TRPA1 to study the neural basis of behavior in Drosophila. *J Undergrad Neurosci Educ* **9**: A5–A14.

Billeter JC, Atallah J, Krupp JJ, Millar JG, Levine JD. 2009. Specialized cells tag sexual and species identity in Drosophila melanogaster. *Nature* **461**: 987–991.

Bontonou G, Wicker-Thomas C. 2014. Sexual Communication in the Drosophila Genus. *Insects* **5**: 439–458.

Boul KE, Funk WC, Darst CR, Cannatella DC, Ryan MJ. 2007. Sexual selection drives speciation in an Amazonian frog. *Proc R Soc B Biol Sci* **274**: 399–406.

Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM, et al. 2014. Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**: 393–399.

Chowdhury T, Calhoun RM, Bruch K, Moehring AJ. 2020. The fruitless gene affects female receptivity and species isolation. *Proc R Soc B Biol Sci* **287**.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**: 203–218.

Cobb M, Burnet B, Connolly K. 1988. Sexual isolation and courtship behavior in Drosophila simulans, D. mauritiana, and their interspecific hybrids. *Behav Genet* **18**: 211–225.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science (80-)* **339**: 819–823.

Cook R, Connolly K. 2008. Rejection Responses By Female Drosophila Melanogaster : Their Ontogeny, Causality and Effects Upon the Behaviour of the Courting Male. *Behaviour* **44**: 142–165.

Costa FF. 2008. Non-coding RNAs, epigenetics and complexity. *Gene* **410**: 9–17.

Coyne JA, Orr HA. 2004. *Speciation*. Oxford University Press, Oxford.

David JR, Lemeunier F, Tsacas L, Yassin A. 2007. The historical discovery of the nine species in the Drosophila melanogaster species subgroup. *Genetics* **177**: 1969–1973.

De Lucia F, Dean C. 2011. Long non-coding RNAs and chromatin regulation. *Curr Opin Plant Biol* **14**: 168–173.

De Queiroz K. 2005. Ernst Mayr and the modern concept of species. *Proc Natl Acad Sci U S A* **102**: 6600–6607.

Demir E, Dickson BJ. 2005. fruitless splicing specifies male courtship behavior in Drosophila. *Cell* **121**: 785–794.

Diao F, Ironfield H, Luan H, Diao F, Shropshire WC, Ewer J, Marr E, Potter CJ, Landgraf M, White BH. 2015. Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes. *Cell Rep* **10**: 1410–1421.

Diao F, White BH. 2012. A novel approach for directing transgene expression in Drosophila: T2A-Gal4 in-frame fusion. *Genetics* **190**: 1139-44.

Dornan AJ, Gailey DA, Goodwin SF. 2005. GAL4 enhancer trap targeting of the Drosophila sex determination gene fruitless. *Genesis* **42**: 236–246.

Douglas SJ, Levine JD. 2006. Sex Cells: Dissecting the Functions of Fruitless Isoforms. *Curr Biol* **16**: R405-7.

Dufour L. 1844. Anatomie generale des dipteres. *Ann des Sci Nat* **1**: 244–264.

Dukas R. 2008. Learning decreases heterospecific courtship and mating in fruit flies. *Biol Lett* **4**: 645–647.

Fan P, Manoli DS, Ahmed OM, Chen Y, Agarwal N, Kwong S, Cai AG, Neitz J, Renslo A, Baker BS, et al. 2013. Genetic and neural mechanisms that inhibit drosophila from mating with other species. *Cell* **154**: 89–102.

Ferveur JF. 2005. Cuticular hydrocarbons: Their evolution and roles in Drosophila pheromonal communication. *Behav Genet* **35**: 279–295.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**: 279–284.

Gonzalez-Bellido PT, Wardill TJ, Kostyleva R, Meinertzhagen IA, Juusola M. 2009. Overexpressing temperature-sensitive dynamin decelerates phototransduction and bundles microtubules in Drosophila photoreceptors. *J Neurosci* **29**: 14199–14210.

Gratz SJ, Rubinstein CD, Harrison MM, Wildonger J, O'Connor-Giles KM. 2015. CRISPR-Cas9 genome editing in Drosophila. *Curr Protoc Mol Biol* **2015**: 31.2.1-31.2.20.

Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, Oconnor-Giles KM. 2014. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* **196**: 961–971.

Greenspan RJ, Ferveur JF. 2000. Courtship in Drosophila. *Annu Rev Genet* **34**: 205–232.

Guttman M, Rinn JL. 2012. Modular regulatory principles of large non-coding RNAs. *Nature* **482**: 339–346.

Haesler MP, Seehausen O. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: Implications for speciation. *Proc R Soc B Biol Sci* **272**: 237–245.

Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. *J Hered* **105:** 795-809.

Hobbs EC, Fontaine F, Yin X, Storz G. 2011. An expanding universe of small proteins. *Curr*

Opin Microbiol **14**: 167–173.

Hüttenhofer A, Schattner P, Polacek N. 2005. Non-coding RNAs: Hope or hype? *Trends Genet* **21**: 289–297.

Immonen E, Ritchie MG. 2012. The genomic response to courtship song stimulation in female Drosophila melanogaster. *Proc R Soc B Biol Sci* **279**: 1359–1365.

Kilian JG, Hsu HW, Mata K, Wolf FW, Kitazawa M. 2017. Astrocyte transport of glutamate and neuronal activity reciprocally modulate tau pathology in Drosophila. *Neuroscience* **348**: 191–200.

Kopp A, True JR. 2002. Evolution of male sexual characters in the Oriental Drosophila melanogaster species group. *Evol Dev* **4**: 278–291.

Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati P V., Goodman JL, Sian Gramates L, Millburn G, Strelets VB, et al. 2021. FlyBase: Updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Res* **49:** D899-D907.

Lassance JM, Groot AT, Liénard MA, Antony B, Borgwardt C, Andersson F, Hedenström E, Heckel DG, Löfstedt C. 2010. Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature* **466**: 486–489.

Laturney M, Billeter JC. 2014. Neurogenetics of female reproductive behaviors in drosophila melanogaster. *Adv Genet* **85:** 1-108.

Laturney M, Moehring AJ. 2012. Fine-scale genetic analysis of species-specific female preference in Drosophila simulans. *J Evol Biol* **25**: 1718–1731.

Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT. 2018. FlyAtlas 2: A new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res* **46**: D809–D815.

Lessios HA. 2007. Reproductive isolation between species of sea urchins. *Bull Mar Sci* **81**: 191–208.

LeVasseur-Viens H, Polak M, Moehring AJ. 2015. No evidence for external genital morphology affecting cryptic female choice and reproductive isolation in Drosophila. *Evolution (N Y)*

69: 1797–1807.

Lin CC, Potter CJ. 2016. Editing transgenic DNA components by inducible gene replacement in Drosophila melanogaster. *Genetics* **203**: 1613–1628.

Liu X, Hao L, Li D, Zhu L, Hu S. 2015. Long Non-coding RNAs and Their Biological Roles in Plants. *Genomics, Proteomics Bioinforma* **13**: 137–147.

Löve A. 1964. the Biological Species Concept and Its Evolutionary Structure. *Taxon* **13**: 33–45.

Mahmood T, Yang PC. 2012. Western blot: Technique, theory, and trouble shooting. *N Am J Med Sci* **4**: 429–434.

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**: 833–838.

Mallet J. 2020. Alternative views of biological species: Reproductively isolated units or genotypic clusters? *Natl Sci Rev* **7**: 1401–1407.

Manning A. 1967. The control of sexual receptivity in female Drosophila. *Anim Behav* **15**: 239–250.

Masly JP. 2012. 170 Years of "Lock-and-Key": Genital Morphology and Reproductive Isolation. *Int J Evol Biol* **2012**: 1–10.

Mattick JS, Makunin I V. 2006. Non-coding RNA. *Hum Mol Genet* **15 Spec No**: R17-29.

Matute DR, Coyne JA. 2010. Intrinsic reproductive isolation between two sister species of drosophila. *Evolution (N Y)* **64**: 903–920.

Matute DR, Novak CJ, Coyne JA. 2009. Temperature-based extrinsic reproductive isolation in two species of Drosophila. *Evolution (N Y)* **63**: 595–612.

Mazzoni V, Anfora G, Virant-Doberlet M. 2013. Substrate vibrations during courtship in three Drosophila species. *PLoS One* **8**: e80708.

Nagarkar-Jaiswal S, Lee PT, Campbell ME, Chen K, Anguiano-Zarate S, Gutierrez MC, Busby T, Lin WW, He Y, Schulze KL, et al. 2015. A library of MiMICs allows tagging of genes

and reversible, spatial and temporal knockdown of proteins in Drosophila. *Elife* **31:** e05338.

Nelson CR, Szauter P. 1992. Cytogenetic analysis of chromosome region 89A of Drosophila melanogaster: isolation of deficiencies and mapping of Po, Aldox-1 and transposon insertions. *MGG Mol Gen Genet* **235**: 11–21.

Neville MC, Nojima T, Ashley E, Parker DJ, Walker J, Southall T, Van De Sande B, Marques AC, Fischer B, Brand AH, et al. 2014. Male-specific fruitless isoforms target neurodevelopmental genes to specify a sexually dimorphic nervous system. *Curr Biol* **24**: 229–241.

Nojima T, Neville MC, Goodwin SF. 2014. Fruitless isoforms and target genes specify the sexually dimorphic nervous system underlying Drosophila reproductive behavior. *Fly (Austin)* **8**: 95–100.

Noor MA. 1995. Speciation driven by natural selection in Drosophila. *Nature* **375**: 674–675.

Palazzo AF, Lee ES. 2015. Non-coding RNA: What is functional and what is junk? *Front Genet* **6**.

Pang TL, Ding Z, Liang SB, Li L, Zhang B, Zhang Y, Fan YJ, Xu YZ. 2021. Comprehensive Identification and Alternative Splicing of Microexons in Drosophila. *Front Genet* **12:** 642602.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**: 839–843.

Poodry CA, Edgar L. 1979. Reversible alterations in the neuromuscular junctions of Drosophila melanogaster bearing a temperature-sensitive mutation, shibire. *J Cell Biol* **81**: 520–527.

Port F, Chen HM, Lee T, Bullock SL. 2014. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc Natl Acad Sci U S A* **111**: E2967–E2976.

Roberts DB. 2006. Drosophila melanogaster: The model organism. *Entomol Exp Appl* **121**: 93–103.

Ryner LC, Goodwin SF, Castrillon DH, Anand A, Villella A, Baker BS, Hall JC, Taylor BJ, Wasserman SA. 1996. Control of male sexual behavior and sexual orientation in Drosophila by the fruitless gene. *Cell* **87**: 1079–1089.

Sætre GP, Sæther SA. 2010. Ecology and genetics of speciation in Ficedula flycatchers. *Mol Ecol* **19**: 1091–1106.

Sakai T, Kasuya J, Kitamoto T, Aigaki T. 2009. The Drosophila TRPA channel, Painless, regulates sexual receptivity in virgin females. *Genes, Brain Behav* **8**: 546–557.

Salvemini M, Polito C, Saccone G. 2010. fruitless alternative splicing and sex behaviour in insects: An ancient and unforgettable love story? *J Genet* **89**: 287–299.

Sasa MM, Chippindale PT, Johnson NA. 1998. Patterns of postzygotic isolation in frogs. *Evolution (N Y)* **52**: 1811–1820.

Smith JD, Suresh S, Schlecht U, Wu M, Wagih O, Peltz G, Davis RW, Steinmetz LM, Parts L, St.Onge RP. 2016. Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biol* **17**.

Sobel JM, Chen GF, Watt LR, Schemske DW. 2010. The biology of speciation. *Evolution (N Y)* **64**: 295–315.

Sokal RR, Crovello TJ. 1970. The Biological Species Concept: A Critical Evaluation. *Am Nat* **104**: 127–153.

Spieth HT. 1974. Courtship behavior in Drosophila. *Annu Rev Entomol* **19**: 385–405.

Stern DL. 2014. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends Genet* **30**: 547–554.

Storz G, Wolf YI, Ramamurthi KS. 2014. Small proteins can no longer be ignored. *Annu Rev Biochem* **83**: 753–777.

Strassburger K, Teleman AA. 2016. Protocols to study growth and metabolism in drosophila. *Methods Mol Biol* **1478**: 279–290.

Su M, Ling Y, Yu J, Wu J, Xiao J. 2013. Small proteins: Untapped area of potential biological importance. *Front Genet* **4**.

Sweigart AL. 2010. The genetics of postmating, prezygotic reproductive isolation between Drosophila virilis and D. americana. *Genetics* **184**: 401–410.

Tomaru M, Oguma Y. 2000. Mate choice in Drosophila melanogaster and D. sechellia: Criteria and their variation depending on courtship song. *Anim Behav* **60**: 797–804.

Trujillo JM, Stenius C, Christian LC, Ohno S. 1962. Cromosomes of the horse, the donkey, and the mule. *Chromosoma* **13**: 243–248.

Van Der Bliek AM, Meyerowrtz EM. 1991. Dynamin-like protein encoded by the Drosophila shibire gene associated with vesicular traffic. *Nature* **351**: 411–414.

Venken KJT, Schulze KL, Haelterman NA, Pan H, He Y, Evans-Holm M, Carlson JW, Levis RW, Spradling AC, Hoskins RA, et al. 2011. MiMIC: A highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nat Methods* **8**: 737–747.

Vieira A, Miller DJ. 2006. Gamete interaction: Is it species-specific? *Mol Reprod Dev* **73**: 1422–1429.

Von Philipsborn AC, Jörchel S, Tirian L, Demir E, Morita T, Stern DL, Dickson BJ. 2014. Cellular and behavioral functions of fruitless isoforms in Drosophila courtship. *Curr Biol* **24**: 242–251.

Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, et al. 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35**: 4809–4819.

Weinberg Z, Lünse CE, Corbino KA, Ames TD, Nelson JW, Roth A, Perkins KR, Sherlock ME, Breaker RR. 2017. Detection of 224 candidate structured RNAs by Comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res* **45**: 10811–10823.

Wicker-Thomas C. 2011. Evolution of insect pheromones and their role in reproductive isolation and speciation. *Ann la Soc Entomol Fr* **47**: 55–62.

Yamaguchi M, Yoshida H. 2018. Drosophila as a model organism. *Adv Exp Med Bio* **1076:** 1-10.

Yassin A, Orgogozo V. 2013. Coevolution between Male and Female Genitalia in the Drosophila melanogaster Species Subgroup. *PLoS One* **8:** e57158.

Zhang JP, Li XL, Neises A, Chen W, Hu LP, Ji GZ, Yu JY, Xu J, Yuan WP, Cheng T, et al. 2016. Different Effects of sgRNA Length on CRISPR-mediated Gene Knockout Efficiency. *Sci Rep* **6:** 28566.

Zhou C, Pan Y, Robinett CC, Meissner GW, Baker BS. 2014. Central brain neurons expressing doublesex regulate female receptivity in Drosophila. *Neuron* **83:** 149-163.

ZHOU Z wei, CAO G hua, LI Z, HAN X jie, LI C, LU Z yu, ZHAO Y hang, LI X ling. 2019. Truncated gRNA reduces CRISPR/Cas9-mediated off-target rate for MSTN gene knockout in bovines. *J Integr Agric* **18**: 2835–2843.

Appendix: Crossing Scheme for PhiC31-mediated integration of Trojan Gal4 into MiMIC sites

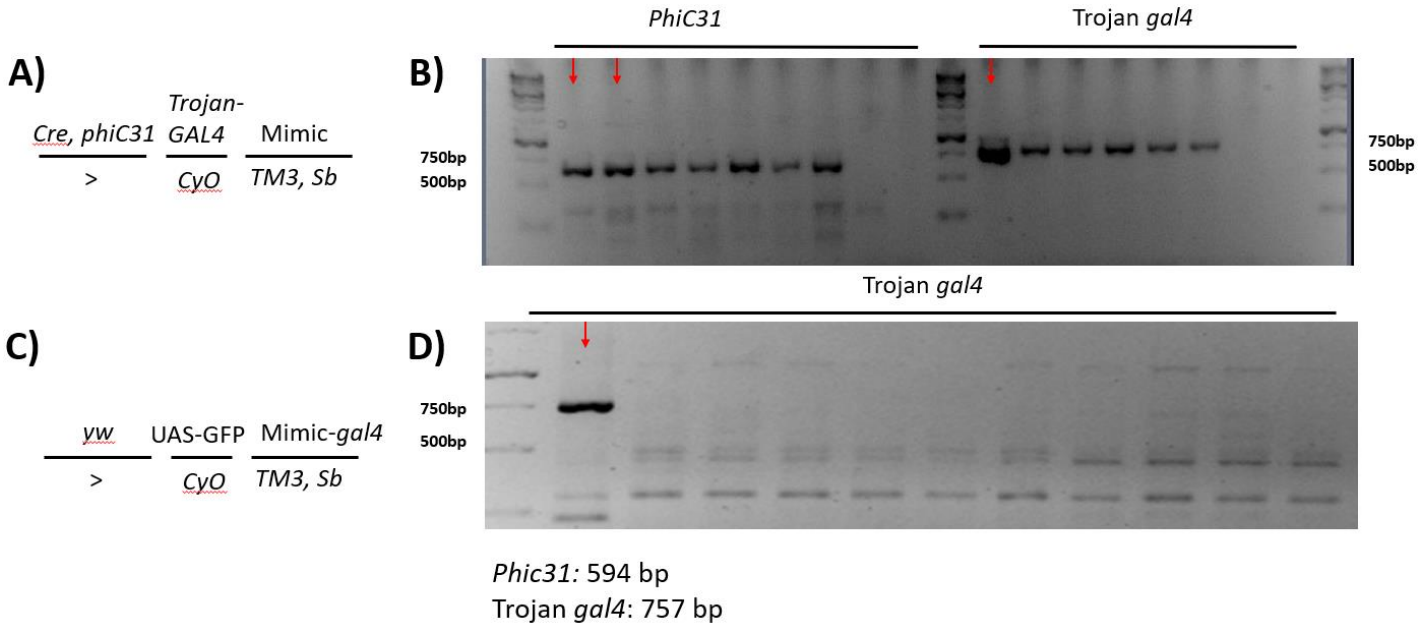

Figure 1. PCR of required genetic components for Trojan-Gal4 integration via crosses. **(A)** Necessary genotype of fourth generation *Drosophila melanogaster* males in the crosses for successful PhiC31 recombination. **(B)** Gel electrophoresis of PCR products confirming the necessary presence of *phiC31* integrase (630 base pair fragments) and Trojan *GAL4* (757 base pair fragments) in 4th generation *Drosophila melanogaster* males for Trojan-Gal4 integration adjacent to A and B 3' terminal exons. **(C)** Desired genotype of fifth generation *Drosophila melanogaster* males expressing successful Trojan-Gal4 integration via the multigenerational genetic crossing scheme. **(D)** Gel electrophoresis of PCR products displaying the unsuccessful integration of Trojan-Gal4 (757 base pairs) into attB swap sites in 5th generation offspring, despite the parental generation having the necessary genetic components. DNA is from a single fly per lane. Red arrows indicate positive DNA controls.

Curriculum Vitae

Name: Omar El-Deeb

Post-Secondary Education and Degrees:

Master's of Neuroscience
Western University
London, Ontario, Canada
2018-Present

Honour's Bachelor's of Medical Sciences
Western University
London, Ontario, Canada
2013-2017

Related Work Experience:

Teaching Assistant
Biology 2601A Organismal Physiology
Biology 1001A/1201A Introduction to Biology
Biology 2290 Scientific Methods – Genetics Unit
Western University
2018-Present

Publications:

Mushtaha, N. F., Kuehn, T. K., El-Deeb, O., Rohani, S. A., Helpard, L. W., Moore, J., Ladak, H., Moehring, A., Baron, C. A, and Khan, A. R. Microstructural characterization and validation of a 3D printed axon-mimetic phantom for diffusion MRI. Magn Reson Med. DOI: 10.1002/mrm.28886.

El-Deeb, O. 2020. Digest: The evolutionary relationship between environment and size in Australian rodents. Evolution. DOI: 10.1111/evo.13909

El-Deeb, O. 2019. Digest: The evolution of sex-specific mating preferences in *Drosophila*. Evolution. **73** (9): 2024–2025.

El-Deeb, O. 2019. Digest: Fitness costs of Spiroplasma infection in pea aphids. Evolution. **73** (7): 1490–1491.

El-Deeb, O. 2019. Digest: A clinal polymorphism and life-history adaptations in *Drosophila*. Evolution. **73** (9): 2026–2027.