THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

Fluorescent Protein Trapping in *Drosophila melanogaster*

ANNA VALERIE FOLTZ

SPRING 2023

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree in
Biochemistry and Molecular Biology
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Claire Thomas
Associate Professor of Biology
Associate Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Song Tan
Verne M. Willaman Professor of Molecular Biology
Honors Advisor

* Electronic approvals are on file.

ABSTRACT

Drosophila melanogaster is a popular model organism with tractable genetics. To track protein levels and localization *in vivo*, gene traps are popular tools used to tether a fluorescent protein to a protein of interest in the fly genome. Historically, Green Fluorescent Protein (GFP) has been used for this purpose. In Dr. Claire Thomas' 200-level Biology course Biol230M 'The Biology of Molecules and Cells', gene trapping has been used for several years to bring a research exercise into the laboratory classroom. This class used the existing Wee-P transposon, a GFP based gene trap. When the Wee-P transposon lands within introns at random across fly chromosomes, the Wee-P exon can be spliced to adjacent exons through splice sites located on either side of the protein coding sequence, generating a fluorescent protein fusion. Several years of experience in the lab has shown that the undergraduates consistently find it difficult to distinguish many GFP fusion events from the green-autofluorescence found in many *Drosophila* tissues. In this thesis I attempt to avoid this setback by making a new gene trapping transposon, based on the red fluorescent protein tdTomato. Reengineering this system also offers the opportunity to optimize several other aspects of the system. As well as improving trap detection with tdTomato, I also tried to optimize the trap identification process by designing better inverse PCR primers and more convenient restriction endonuclease sites to maximize the probability of success when using the iPCR gene identification strategy. Finally, the new design relocates the P[w^+] ammunition construct's selectable marker outside of the transposon to obviate the need to remove it after successfully trapping, simplifying the creation of a stable stock. We intend to make the new trap lines freely available to the *Drosophila* community, and I hope that some of the class-generated stocks will eventually prove useful to other fly labs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES CONT.

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Claire Thomas for her outstanding mentorship for the last three-and-a-half years of my thesis research. Without her guidance, I wouldn't have discovered my passion for Molecular Biology, as I entered Penn State through the College of Engineering. I thank her for the extensive time and effort put into my education, as well as the resources required to conduct necessary experiments for this thesis. I owe all my proficiency in the lab and bench skills to her.

Next, I would like to thank the members of the Thomas Lab who have made my time here so memorable: Chelsea Sarring, Kinsuk Shill, Ashley Swope, Maria Lovallo, Maria Elder, and Bryanna White. I have enjoyed every lab meeting we attended together and learned how wonderful collaborative science can be alongside all of you.

Lastly, I would like to thank my mother, father, and brother, who have supported me since day one and encouraged me to pursue everything I am curious about. Thank you all for your unwavering belief in me.

CHAPTER 1

Introduction

Fluorescent Protein Usage in Life Sciences

Green fluorescent protein (GFP) is a popular fluorescent marker utilized in biological laboratories. Discovered in 1994 by Martin Chalfie, Osamu Shimomura, and Roger Tsien, it was shown that GFP allows for tracking real-time development of biological structures *in vivo* (Swaminathan, 2009). The GFP gene is native to *Aequorea victoria*, a bioluminescent jellyfish (Swaminathan, 2009). When the GFP gene is fused to a flanking protein, the fluorescent protein follows its tethered neighbor protein wherever it is expressed and localized. This protein tracking mechanism is frequently used in *Drosophila melanogaster* for this very purpose: To determine how protein expression and location changes through the fly life cycle, or under mutant conditions. Figure 1 shows an example of GFP expression in developing *Drosophila*, from the embryonic stage to adulthood (Casso *et al.*, 1999). GFP has been fine-tuned and mutated over the years, to generate fluorescent proteins of other wavelengths. The discovery of other fluorescent proteins in other species and their modification has also greatly expanded the available range of wavelengths. For example, red and orange fluorescent proteins were first isolated from *Discosoma* species of sea anemones (Dietrich *et al.*, 2002). A comprehensive listing of all fluorescent proteins and their derivatives can be found at FPbase (https://www.fpbase.org/).

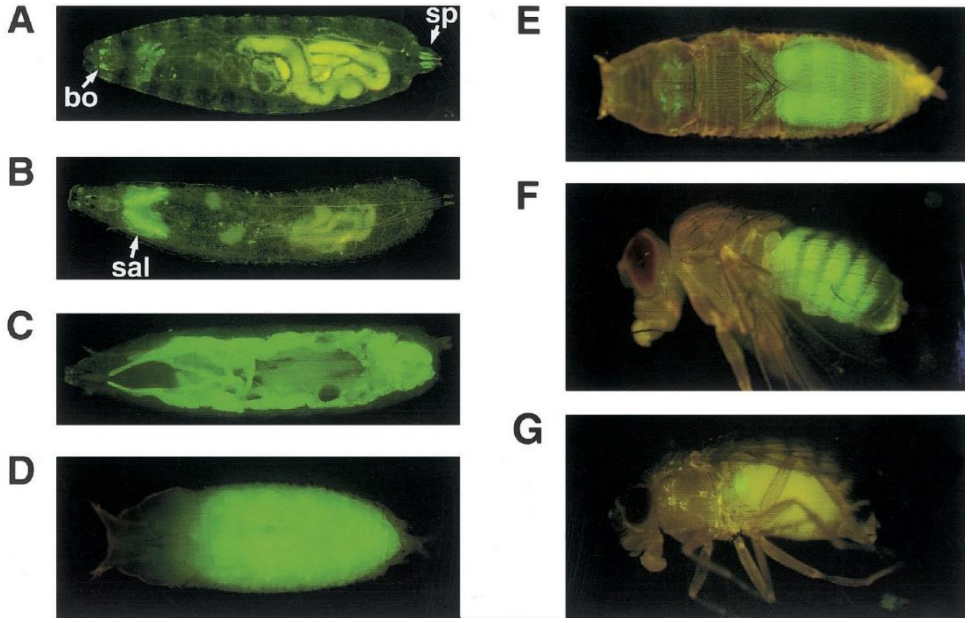

Figure 1: GFP Expression in developing *Drosophila melanogaster*.

A: late embryo, B,C: larvae; D, E: Pupae; F, G: Adults. Reproduced from Casso *et al.*, 1999 with permission.

Transposition in *Drosophila*

P-elements are naturally occurring mobile genetic elements in *Drosophila melanogaster*. These sequences were first discovered when wild-caught male *Drosophila* were mated to female laboratory strains, and it was observed that offspring of this cross were sterile and contained to significant chromosomal rearrangements (Ghanim *et al.*, 2020). In contrast, when wild-caught female *Drosophila* were crossed to laboratory strain males, offspring were healthy and able to reproduce. This phenomenon is named ‘dihybrid dysgenesis’ and can result in a significant number of mutations at the hands of mobile genetic elements with Long Terminal Repeats (LTR) called ‘P-elements’. Notably, some of the mutations were reversible. By focusing on eye color

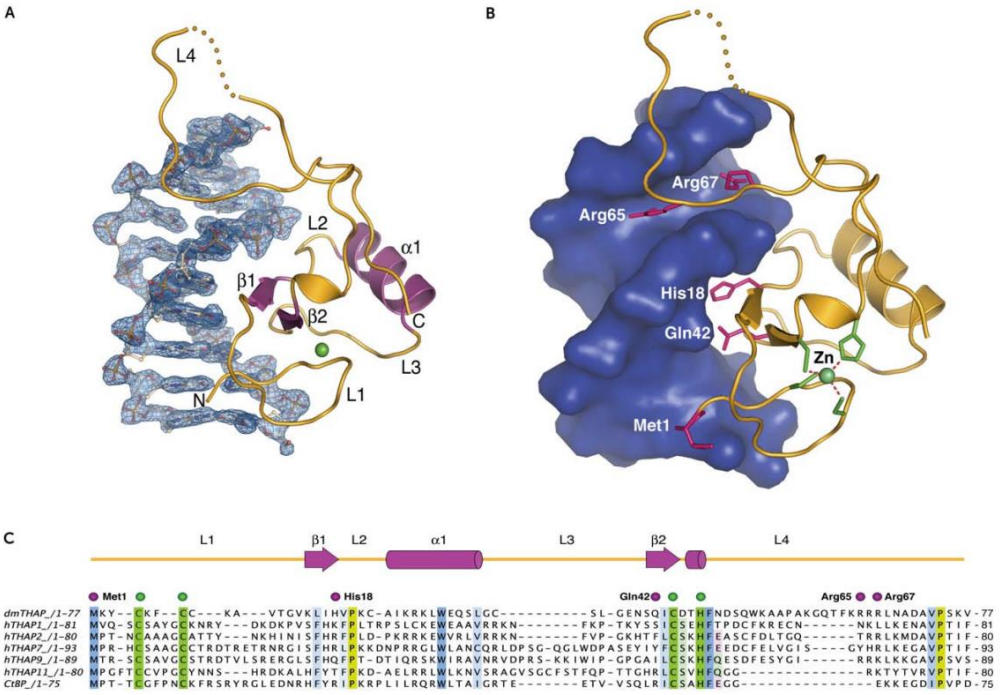

Figure 2: Structure of the transposase THAP Domain.

A,B,C, Reproduced from: Majumdar *et al.*, 2015 with permission.

mutants, scientists were able to pinpoint insertions in the *white* gene when eye color mutated to white and identify deletions when this mutation was reversed (Ghanim *et al.*, 2020). Subsequent experiments led to identification of the P-element as the mutagenic agent. The *white* alleles had an inserted P-element, but their remobilization was imprecise and could have a deletion at the site of insertion. Kaufman *et al.* (1989) showed that transposase is the enzyme responsible for mobility of P-elements and determined the P-element sequences that allowed for interaction with the mobilizing enzyme transposase.

The P-element transposase is a DNA binding protein that binds to a 10 base-pair consensus sequence located near the 31 base-pair LTRs within P-elements through its THAP (Thanatos Associated Protein) domains. Figure 2 displays the generally conserved β-α-β protein

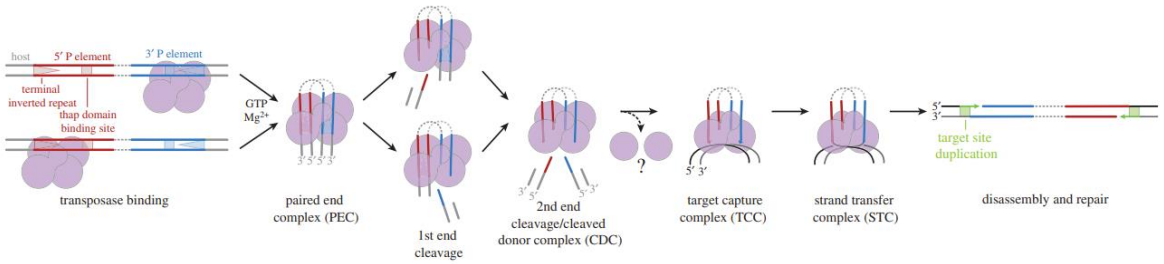

Figure 3: P-element Transposition Pathway

See text for discussion. Reproduced from Ghanim *et al.*, 2020 with permission.

structural fold of the THAP domain, that contains amino acid residues His18 and Gln42. These residues bind the major groove of target DNA, stabilized by a zinc-finger domain.

In the presence of substrate DNA, GTP (guanosine tri-phosphate), and Magnesium ions, bound subunits form a Pair-End Complex (PEC) causing the DNA contained between the P ends to fold. Figure 3 depicts how the THAP units dimerize to instigate transposon detachment. Inside the transposase enzyme, host DNA is cleaved away from the transposon complex, first 17 nucleotides into the 5’ site then directly at the end of the 3’ site. This frees the target capture complex (TCC) from the genome and can be interpreted as the initial ‘jump’. When landing arbitrarily into a different location, the TCC initiates integration of the DNA into the genome once more. This step is called the Strand Transfer Complex (STC), which results in reinsertion with an 8 bp (base pair) target site duplication at either end of the transposon. While there is the possibility that the transposon could land in any location, Ghanim *et al.* (2020), showed that STC are most likely to form on the same chromosome as the starting PEC, in a trend called local hopping. Furthermore, they determined that STC prefer euchromatin to densely packed heterochromatin, and especially regions near promoters and origins of replication (Ghanim *et al.*, 2020).

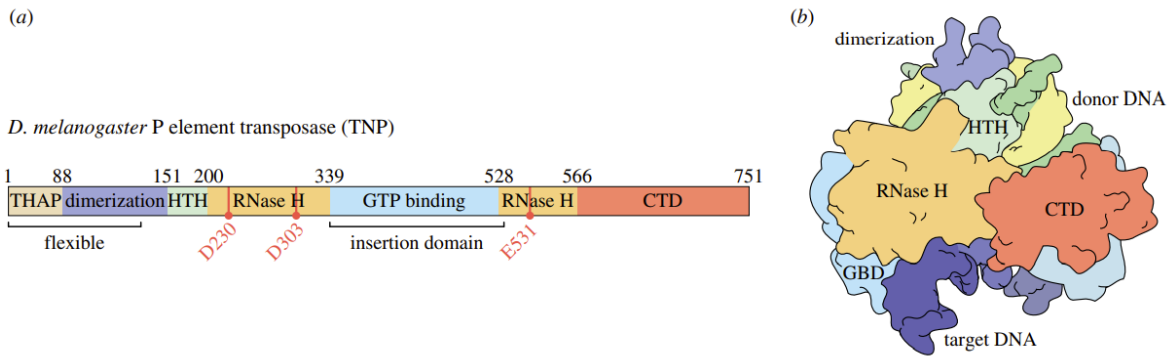

Figure 4: Illustrated Domains of Transposase

See text for discussion. Reproduced from**:** Ghanim *et al.*, 2020 with permission

Transposase Structure and Function

To reveal the structure and intricate mechanisms of transposase action, Ghanim *et al.* (2020) utilized cryo-electron microscopy to examine the STC, containing transposase bound to TCC. From this experiment, it was shown that the STC had six different key domains all illustrated in Figure 4: An N-terminal THAP DNA-binding domain; a leucine zipper-dimerization domain; a helix-turn-helix DNA-binding domain (HTH); and a split catalytic RNase H domain interrupted by a GTP-binding insertion domain (GBD) followed by a carboxy-terminal domain.

THAP recognizes sequence within the P-element end, upon which a zinc-binding motif initiates dimerization between THAP on the 5’ and 3’ P-element ends. Following the THAP domain’s initial folding/cleavage of the transposon, the helix-turn-helix (HTH) domain binds and stabilizes this TCC. The RNase-H like domain directs the amino acid residues D230, D303, and E531 act to initiate nucleophilic cleavage of the host DNA in the TCC. GTP is a required cofactor in the cleavage process, hence the GTP binding domain (GBD). Located near the

RNase-H domain, the GBD binds GTP upon which this molecule hydrogen-bonds to the final nucleotide of the transposon. Strangely, GTP hydrolysis cannot be observed at this location, as structures necessary for this functionality are located too far away in the enzyme structure for this to occur. It is suspected that this cofactor assists in the direction of the 3' OH- attack strand, when the STC is inserting itself into the landing site. The GBD also stabilizes the STC complex upon transposon insertion. Lastly, the Carboxy-terminal domain (CTD) contains a variety of DNA-interacting amino acid residues along an alpha-helical domain, and together with the GBD, remains near the target DNA in the transposon landing site.

P-element Structure

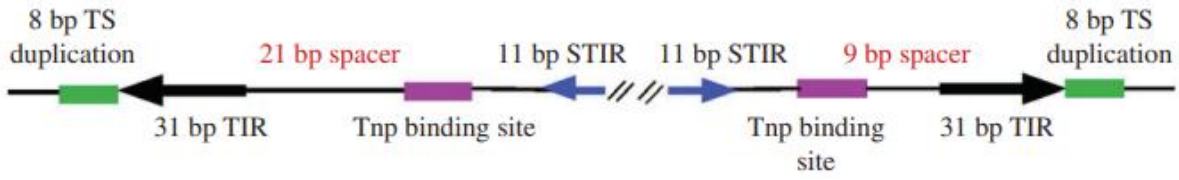

Figure 5: *Drosophila* P-element sequence structure:

Key: 8bp Target Site Duplication (green) ; 31 bp Terminal Inverted Repeat (black); 21 bp spacer (red) ; Transposase Binding Site (purple); 11bp Internal Inverted Repeat (blue). Reproduced from Ghanim *et al.*, 2020 with permission.

In order for transposase to recognize and form the PEC through its THAP domain, it is critical that a P-element contains a precise set of nucleotide sequences. Two classes of P-elements have been identified: autonomous transposase-coding elements, and smaller constructs that rely on an external source of transposase. Figure 5 contains a sequence description of the

latter. In a P-element sequence, there are 31 base-pair terminal inverted repeats as well as 11 base-pair internal inverted repeats that are shown to be responsible for THAP domain recognition (Madjumar *et al.,* 2015, Ghanim *et al.,* 2020). It is between these two regions within both P-element ends that THAP binds and bends the transposon. Once the eventual STC is inserted into the landing site intron, the 8 base-pair duplications that appear outside the transposon are in turn considering a portion of P-element structure.

Gene Tagging and Trapping Strategies

Randomly tethering a fluorescent protein to endogenous proteins is only one example of the practice called 'Gene Trapping'. There are multiple gene trapping methods with differing mechanisms, all with the goal of 'trapping' some aspect of gene function. In addition, there are other ways to generate and express fluorescent fusion proteins. I will review select examples here.

1) *Trapping Promoters to Generate Driver Lines for The GAL4/UAS Binary Expression System*

Originating in yeast, GAL4 and its UAS were used to activate genes in the galactose metabolic pathway of *Saccharomyces cerevisiae*. Since then, it has been incorporated into a wide variety of model organisms to regulate transcription in a target tissue. The GAL4/UAS system is useful in that any UAS construct can be combined with any pre-existing GAL4 driver line. Many driver lines were generated by mobilizing a transposon containing a the GAL4 gene with a minimal to trap regulatory elements in promoters specific to the tissue/cell type of interest (Kanca *et al.,* 2017). These GAL4 lines can then be crossed to transgenic flies containing the

UAS sequence (upstream activating sequence) plus a reporter gene. In the offspring, tissue-specific promoter-led production of GAL4 triggers expression of the reporter gene through binding to UAS. Brand and Perrimon first demonstrated the viability of this method by crossing Rhodopsin 2-GAL4 lines to transgenic UAS-lacZ reporter lines (Brand *et al.*, 1993). Staining with x-gal (5-Bromo-4-chloro-3-indolyl beta-D-galactopyranoside) resulted in tissues containing Rh2-GAL4 as well as UAS-lacZ being marked with blue pigment. GFP can be used in replacement of lacZ to allow for *in vivo* tracking of the cell-type of interest, as GFP is visible in live organisms. By making a protein-fluorescent protein fusion in vitro and placing this in a UAS construct, the tagged protein will be expressed in a pattern reflecting GAL4 expression (which need not be reflective of the tagged protein's native promoter), and its subcellular location tracked.

Such enhancer trapping has resulted in a wide range of GAL4 drivers that allow for the expression of any UAS construct in many places. is a mechanism that was created to randomly generate useful GAL4 lines. The Bloomington *Drosophila* Stock Center currently lists over 8700 stocks [https://bdsc.indiana.edu/stocks/gal4/gal4_all.html], providing immense flexibility in the expression of any given UAS-expression responder construct.

Enhancer trapping only captures those enhancers that can interact with the minimal promoter in the GAL4 transposon. More recently the 'Trojan' exon trapping strategy has been developed to capture the complete expression pattern of specific genes using a GAL4 containing exon containing a site that permits independent (non-fusion) expression of GAL4 in the exact pattern of the gene's mRNA (Diao *et al.*, 2015)

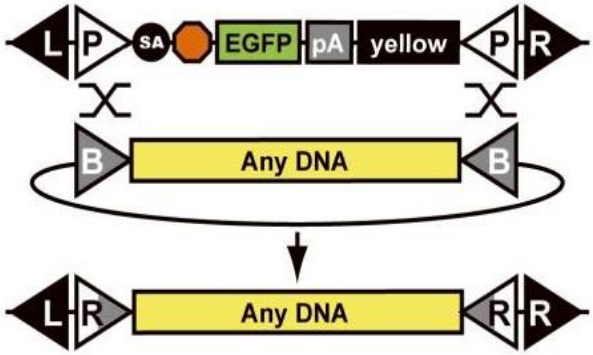

Figure 6: The MiMIC Integration Cassette

Foreign DNA flanked by inverted attB sites is exchanged with a MiMIC cassette at a pre-established locus.

Reproduced from Venken *et al.* with permission.

2)Minos Mediated Integration Cassette: MiMIC System

Possessing much flexibility as gene trapping system, MiMIC (Minos-Mediated Integration Cassette) fly lines offer the ability to introduce foreign nucleotide sequences into predetermined locations. In this construct, a gene cassette containing a splice acceptor site, GFP, a *yellow*$^{+}$ marker, and a stop codon for each ORF (open reading frame) is flanked by inverted attP sequences (Venken *et al.*, 2011). All listed components are then contained between Minos element inverted repeats, which function similarly to P-elements: Minos is considered a transposon. MiMIC versatility is demonstrated in Figure 6; injecting a plasmid containing an attB flanked exon allows for site specific recombination. This recombination is called RMCE (Recombinase Mediated Cassette Exchange), in which a recombinase enzyme (φC31 Integrase) recognizes specific nucleotides in both the MiMIC cassette attP sites and an injected plasmid's attB sites. The cassette is exchanged for any cDNA (complementary DNA) of choice, often this is a reporter gene such as GFP.

Unlike P-elements which exhibit a strong preference for 5' landing sites in genes, Minos transposons do not have any apparent insertional bias (Metaxakis *et al.*, 2005). If mid-gene insertion of a construct is critical to an experiment, or one gene in particular needs trapping, the MiMIC system is a better choice than traditional P-element transposition. A drawback of this method is that it involves an RMCE step, one not needed when the marker gene w^+ is not within the transposon. Furthermore, the goal of our experiments is not to trap a specific gene, rather it was to generate a large number of random insertions to be detected by a fluorescent larval screen.

3) Plug and Play Method: MiMIC and Crispr/Cas9 to Generate a Trojan Exon

While it is possible to produce a usable MiMIC insertion in the middle of a gene, MiMIC constructs located in the 5' UTR of a gene produce more reliably trapped genes: this way the coding sequence of the gene is not disrupted. Furthermore, 90% of genes in *Drosophila* do not have MiMIC trapped lines available since developing such lines is a lengthy process because Minos transposes at random (Diao *et al.*, 2015). To take advantage of MiMIC insertions into coding introns, a "Trojan Exon" containing GAL4 is introduced to flies and exchanged with the existing MiMIC cassette.

Due to the phenomena of ribosomal skipping, when a T2A sequence is translated, the ribosome responsible is unable to connect amino acids coded for by sequence outside T2A mRNA. Thus, including T2A before GAL4 "frees" it from the rest of the trojan exon (Diao *et al.*, 2015). As illustrated in Figure 7, GAL4 can be expressed from within an intron in a gene's coding sequence, and usage of MiMIC is not limited to 5' UTR sites. If desired, any other Trojan Exon can replace the initially used one at any time: hence the "play" in the construct name.

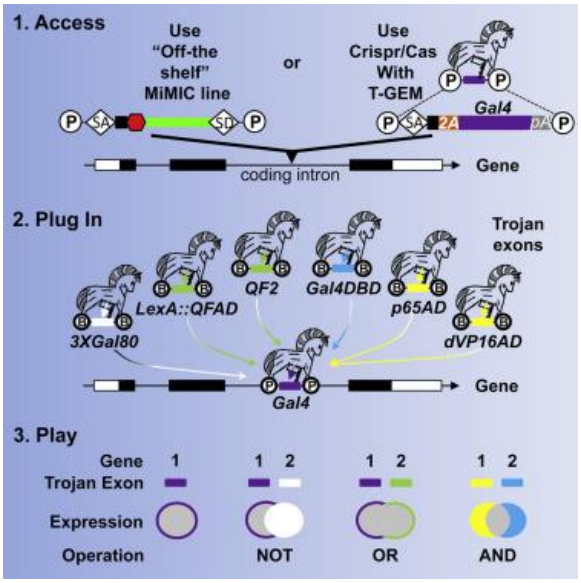

Figure 7: Plug and Play Trapping Method

CRIPSR/Cas9 selected integration sites or pre-existing MiMIC cassettes are used to express GAL4 individually from within an intron. Trojan Exons containing various reporters with a T2A site can replace the initial insertion through RMCE. Replicated from Diao *et al.*, 2015 with Open Access

To further improve the specificity of this system, CRIPSR/Cas9 can guide the initial cassette integration to any location, making any place in the genome available for trapping. Cas9 gRNA (guide ribonucleic acid) is used to generate double-stranded breaks in a specific location where insertion of the cassette is desired. 5' and 3' homology arms flanking an attP site on an externally introduced construct are used to repair Cas9 catalyzed double strand breaks, and hereby insert a φC31 integrase compatible recombination site at a chosen location (Gratz *et al.*, 2013). While an ingenious method of gene trapping, the Plug and Play Method is most useful when aiming to trap a specific gene: for the purposes of generating as many novel traps as possible, it is too specific.

4*) The Wee-P Construct: A Small Transposable Element Modified by FLP Recombinase*

Wee-P is a GFP-based gene trapping construct that is well established and has been widely used by the fly community to tag proteins in *Drosophila* (Clyne *et al.*, 2003) and for teaching purposes in the Biology 230 M (The Biology of Molecules and Cells) course by Dr. Thomas. Wee-P is based on the naturally occurring P-element transposon found in *Drosophila melanogaster*. In Figure 8, it is shown that between the two outlying P-element ends (LTRs), the Wee-P construct contains a *mini-white*$^{+}$ gene flanked by FRT recombination sites, as well as the GFP exon flanked by a splice acceptor and splice donor site. Figure 10 illustrates the need for

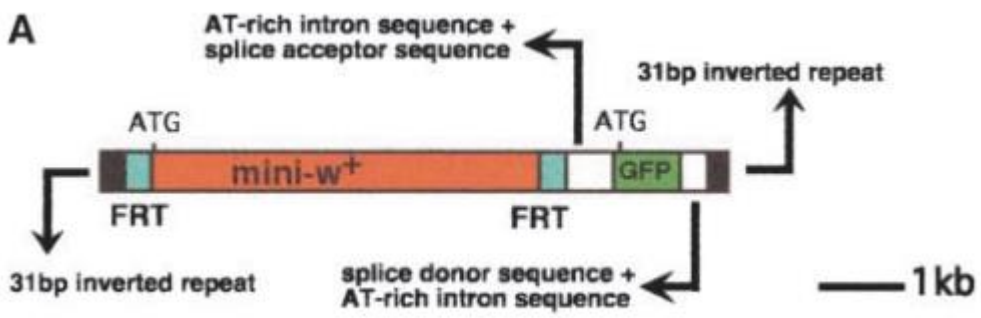

Figure 8: Wee-P Transposable Construct Schematic.

Wee-P construct container marker gene mini-w+ flanked by FRT recombination sites, as well as a GFP trapping exon flanked by a splice acceptor and donor sequence. 31 bp repeats are used to transpose Wee-P through transposase.

Reproduced from Clyne *et al.* (2003) with permission.

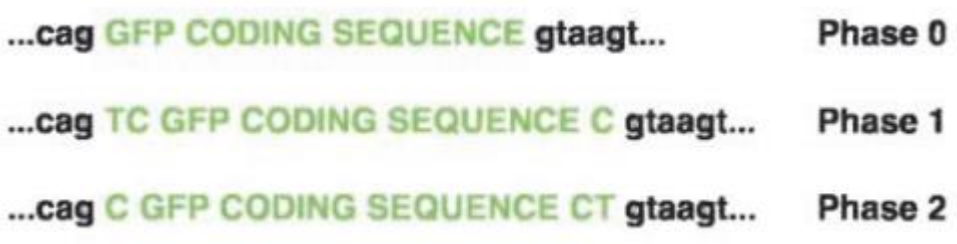

Figure 9: GFP Coding Sequence in Three Open Reading Frames

To take advantage of all ORF in the genome, three frames of Wee-P were constructed. This maximizes yield of visible GFP traps.

Reproduced from Clyne *et al.* (2003) with permission.

three ORF GFP exons each in their own construct, as this allows any intron to be succesfully trapped and translated to generate GFP tags. Open reading frames are continuously translated "bundles" of three nucleotides called codons; each codon combination codes for one of 20 amino acids. To be 'open' a reading frame must lack any stop codons. If, a coding exon ends in the middle of a codon as demonstrated in Figure 9 (Phase 1 and 2), additional nucleotides need to be added flanking the GFP coding sequence, so that the GFP is in-frame with the host protein on both sides.

Once introduced by injection to a *Drosophila* embryo, the construct is mobilized onto chromosome 2, ready for transposition. Flies in which this insertion has ultimately been successful will have red eyes due to the introduction of the *mini-white*$^{+}$ gene, because it rescues the white mutant background. After successful trap recovery, the *mini-white*$^{+}$ gene needs to be excised by FLP recombinase using two FRT sites flanking its sequence. To generate traps, adults are crossed to flies with a source of transposase (*Δ2-3;* Bloomington *Drosophila* Stock Center) and the exon between the P-element ends begins to 'jump' across the genome in the germline

(and elsewhere). To identify traps a second cross to *yellow white* (*yw*) flies is required to produce individuals that derive from individual sperm or eggs carrying an insertion, with each individual most likely representing an independent event. This cross can also serve to remove the transposase source in many cases to allow each new insert to remain stably integrated. However, it is not uncommon to have jumps to locations on the same chromosome as the Δ2-3 construct that will need to be separated by subsequent recombination steps. This would be achieved by taking *trap Δ2-3/+ +* females (meiotic recombination only occurs in females), crossing them to +/+ males and screening the F1 offspring for fluorescence (implying they have the trap). Individual F1 males would be crossed to +/+ females and, after offspring are visible, each male would be sacrificed and screened by PCR for the transposase gene. Crosses that derived from fluorescence-positive, PCR-negative F1 males would be kept. As confirmation, the eye color of the entire F2 generation should be uniform since continued somatic mobilization (ie Δ2-3 still there) is visible as a mottled eye color due to position effect variegation affecting the *mini-white*$^{+}$ gene in the trap.

Upon a successful landing within an intron matching the construct reading frame, GFP is transcribed alongside into pre-mRNA of flanking genes, and becomes fused to the protein following splicing and translation. This is a gene trap: The fluorescent protein is expressed whenever and wherever the tagged protein is translated. Again, to maximize the number of recovered traps, three different constructs matching any of the possible sites within a codon that an intron could be are needed. The reasoning for this step is pictured in Figure 9: the presence of all three reading frames of the transposable exon in one organism exposes any gene to the possibility of fluorescent trapping.

In Clyne *et al.*, Wee-P transposition and trapping was shown to be very successful. Wee-P inserts tend to land in introns averaging 2.9 kb in size, but have been recovered in introns as small as 183bp. This suggests that almost any size of intron is a viable landing site for a trap, increasing the chances of a positive fusion event.

The question was posed as to whether Wee-P fusions altered the tagged protein's function, since insertion events occur between exons that may not be structurally viable locations to maintain the function of the host protein. To verify normal protein functionality, through fly crosses chromosomes containing Wee-P fusion events were placed in trans over a deficiency or a null copy of the fused gene. As a result, offspring had to rely only on the copy of the Wee-P trapped gene. In none of the offspring were mutant phenotypes observed, leading to the conclusion that often Wee-P trapping did not harm its subjects. However, this remains a very real possibility.

Transgenesis: Site Specific Recombination into the Genome

While transposase is responsible for transposon relocation in trapping exercises, it and has also historically been used to mobilize constructs from plasmids into the genome (Rubin and Spradling, 1982) at random locations that needed to be mapped. This can also result in significant position effects depending on the landing site – making quantitative comparisons between constructs more difficult. As a result, site-specific recombination into a standard set of landing sites is now more commonly used to first introduce a construct into the genome. In *Drosophila*, this can be accomplished by using the φC31-integrase in accompaniment with attP/attB nucleotide sequences, or recombinase mediated cassette exchange (RMCE) using either

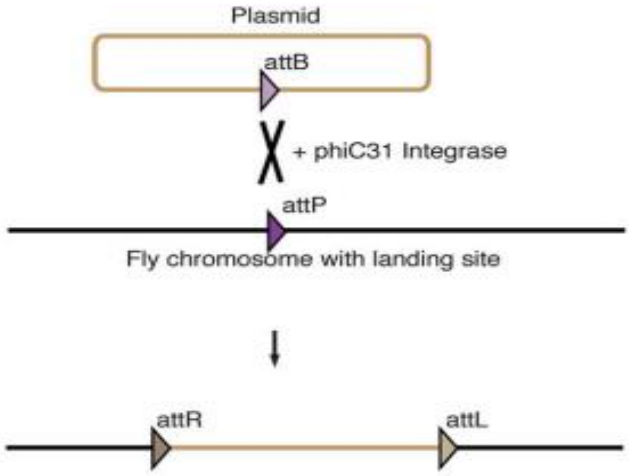

Figure 10: φC31-Integrase Mediated Recombination

An endogenous source of φC31-Integrase recombines an injected attB plasmid into a known attP site in the Drosophila genome. From the original attB site, attR and attL ends flank the recombined plasmid. φC31-Integrase is unable to recognize these sites, and the sequence is permanently integrated into the genome.

Reproduced from Kanca *et al.*, (2017) with permission

FLP or Cre Recombinase. In a strategy designed by Groth *et al.* (2004), attP sites were introduced to *Drosophila* with various transposable elements.

As seen in Figure 10, a plasmid with the gene of interest containing an attB site is introduced to the genome by injection, upon which φC31-integrase mediates recombination between the attB and genomic attP site. This method is advantageous in that it is very efficient and it allows for integration of large constructs into the *Drosophila* genome. Within both the attP and attB sites, recombination initiates at a core 5'-TTG-3' sequence in these two sequences. Post recombination into the genome, the plasmid is flanked by two new sites: attL and attR. φC31-

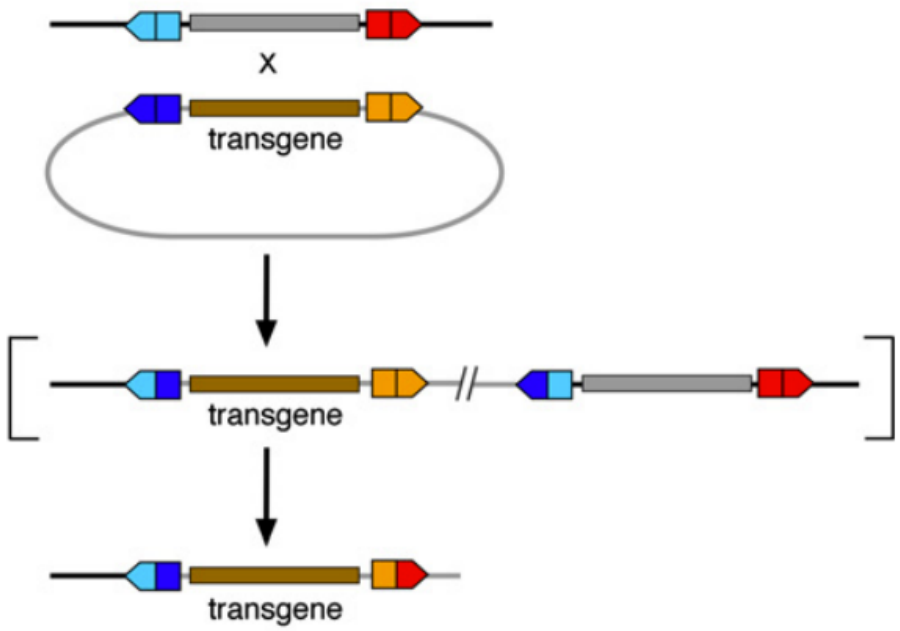

Figure 11: RMCE Mediated Recombination

After initial recombination of the transgene through the blue FRT sites, the gray cassette is exchange for the brown transgene through a second recombination between the orange and red FRT sites.

Reproduced from: Kanca *et al.*, (2017) with permission.

integrase is unable to recognize these sites, and unable to perform a reverse maneuver in which the plasmid is excised.

The RMCE method of integration involves replacing a gene native to *Drosophila* with a gene of choice: hence the term 'exchange.' Figure 11 depicts the mechanism of the cassette exchange mediated by FLP enzyme. A plasmid containing a transgene flanked by FRT recombination sites is recombined in entirety adjacent to a native gene using the left-hand FRT site, followed by a second recombination event led by FLP between the right-hand orange and red FRT sites. This second recombination removes the native gene, replacing it with the

transgene. Cre Recombinase is used in a similar manner, differing in the recognition sites. Discovered in *Escherichia coli* Bacteriophage 1, Cre recognizes loxP sites and achieves the same recombination effect as FLP (Kanca *et al.*, 2017). Due to the simplicity and high efficiency of this recombination method, the method chosen to first introduce the CTAF transposon to the genome was *via* the φC31-integrase mediated site-specific recombination.

Goals for This Thesis

To track protein localization in living organisms, a common strategy is to attach Green Fluorescent Protein (GFP) to a specific protein and observe the cells or tissue using a fluorescence microscope. In the undergraduate laboratory course Biology 230M (The Biology of Molecules and Cells), GFP within the Wee-P construct has been used to teach students basic genetic, molecular and cell biology laboratory skills. However, it was observed that *Drosophila* autofluorescence that is biased towards shorter blue/green wavelengths is a persistent complication in the identification of new GFP tags by novice students. The central goal of this project is to make a new transposable construct based on a bright red fluorescent protein tdTomato that is fluorescent at a longer wavelength where tissue autofluorescence is minimal. It is hypothesized that including an optimized gene trapping construct that codes for tdTomato will increase tagged protein visibility and provide the scientific community with a set of fly lines tagged at a different wavelength from GFP.

CHAPTER 2

DESIGNING THE NEW TRAPPING CONSTRUCT

General Design and Operation of The New Trap Construct

The Davis lab which created the original Wee-P construct could not find the original plasmid (Claire Thomas, personal communication). This meant we would have to rebuild everything from scratch. Completely rebuilding a Wee-P-like construct to contain the tdTomato coding sequences offered the opportunity to optimize a number of other aspects of the original construct which could be improved. The re-designed construct is shown in Figure 12. Specifically, these were the position and design of the primers for gene identification by inverse PCR (iPCR), and the restriction enzyme sites used for iPCR. There was also additional literature now available on designing the amino acid linker between the host protein and fluorescent protein (Chen *et al*., 2019), so this was also revised. Finally, Dr. Thomas' experience over the years with the GFP-based lab was that the trap fluorescence itself was a sufficient dominant marker to follow the trap insertions and so the P[w^+] eye color marker could be moved out of the trap transposon and would just be used to follow the launch construct. This would simplify the downstream genetics of each trap line by removing the need to FLP out the P[w^+] background to generate the more useful w^- background.

Starting with flies homozygous for the transposon pCTAF (w^+ phenotype), a series of crosses will introduce the exogenous transposase source (Δ2-3 transposase), the 5' and 3' P-element ends shown as black arrows in tandem with transposase will mobilize the trap transposon (Figure 13). As with Wee-P, after successfully landing within an intron in the correct

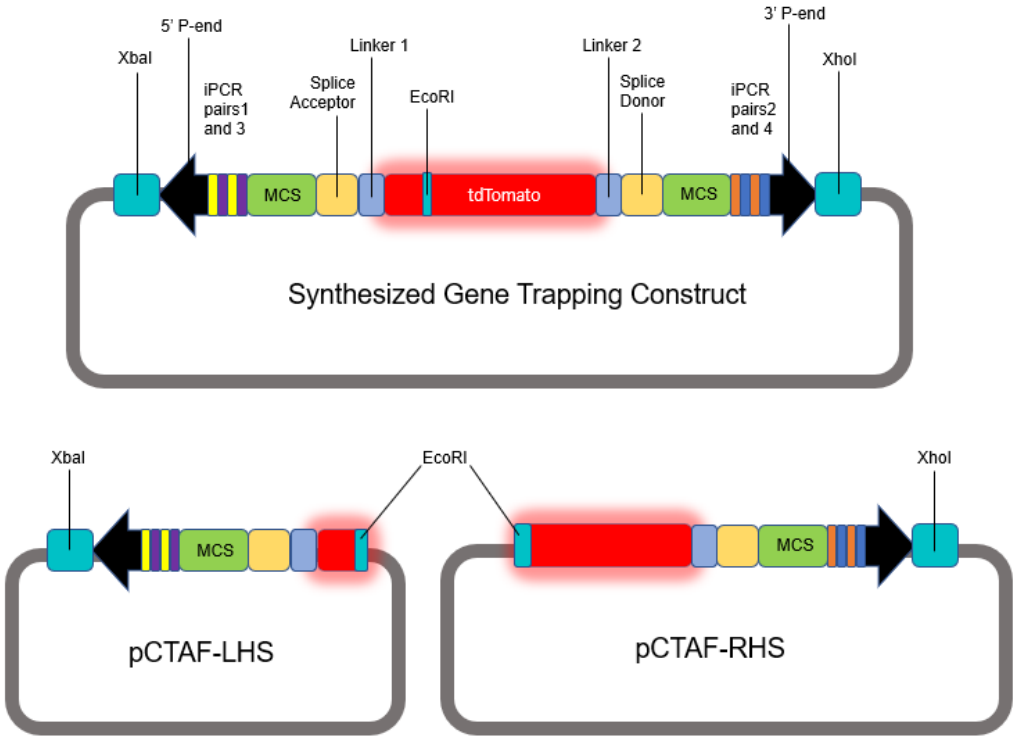

Figure 12: pCTAF, The Re-Designed Gene Trap

The synthesized trapping construct, pCTAF, was synthesized in two halves: pCTAF-LHS (left hand side) and pCTAF-RHS (right hand side). A native EcoRI site within tdTomato was used to reform the entirety of pCTAF from the LHS and RHS plasmids. This was achieved through restriction digests (XbaI, EcoRI, XhoI) and re-ligation of pCTAF DNA with ligase.

reading frame (Step 1), the entirety of the transposon will be transcribed into pre-mRNA (Step 2). Due to Splice Acceptor and Donor sites shown in yellow, the sequence for tdTomato flanked by two linkers will be spliced to adjacent exons (Step 3). From this point forward, anytime that mRNA is translated to protein, tdTomato 'traps' this protein. The flexible amino acid linkers will hopefully minimize any misfolding of the trapped protein, tethering tdTomato so it floats nearby (Step 4).

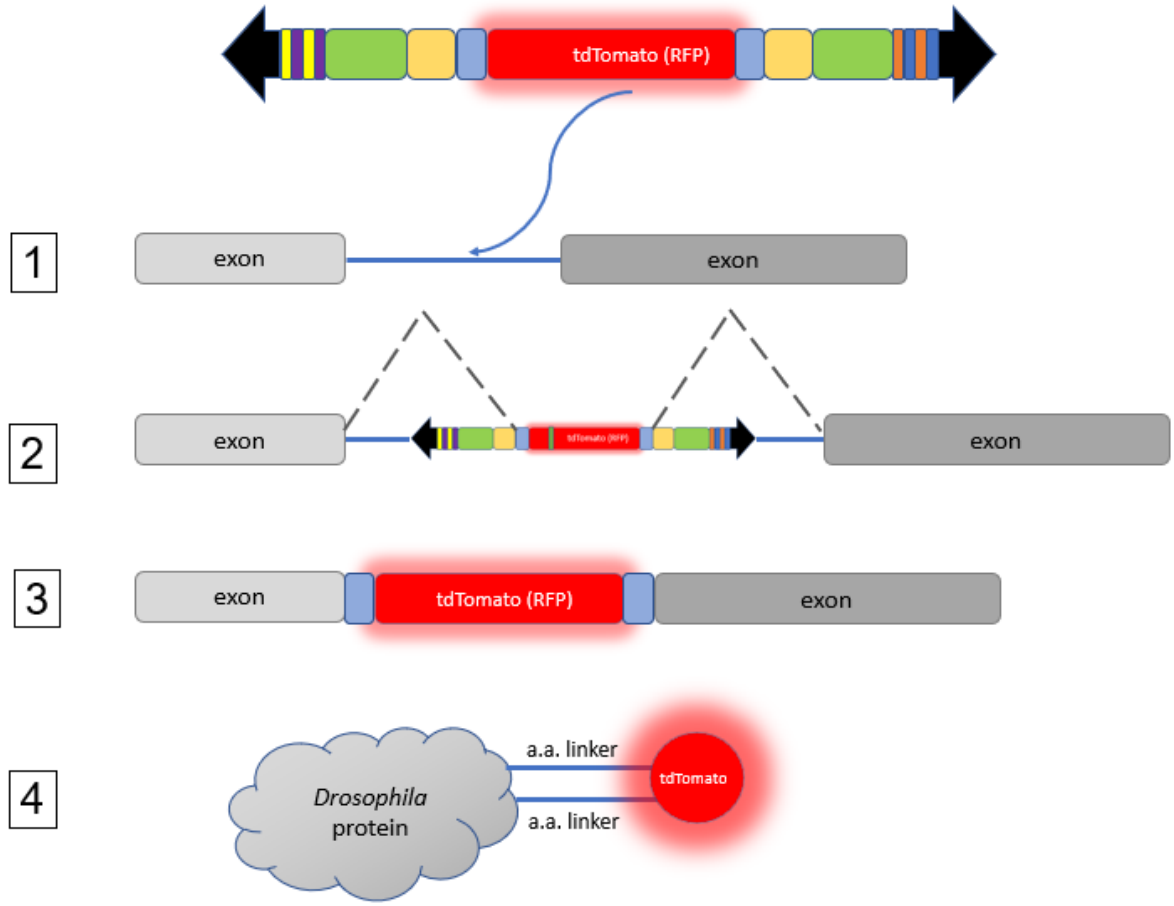

Figure 13: Transposition Mechanism of CTAF Construct

[1] In the presence of transposase, the CTAF construct is integrated between two exons. [2] Post transcription into pre-mRNA, endogenous Splice Acceptor and Splice Donor Sites (yellow) allow for connection of protein coding sequence (Linker1-tdTomato-Linker 2) to flanking exons. [3] Mature RNA now containing only coding sequence is translated, and an amino acid sequence for both the trapped protein and tdTomato is created. [4] Amino acid linkers prevent tdTomato from causing misfolding in the Drosophila protein, and tether it so it localizes in sync with the *Drosophila* protein.

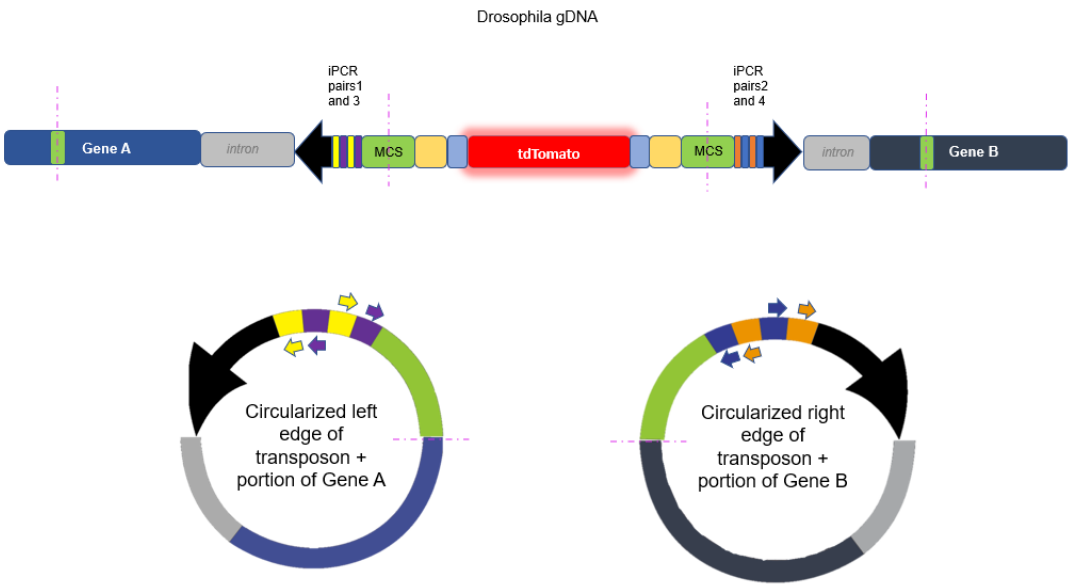

Figure 14: Using iPCR Primer Sites and MCS to Identify the Trapped Gene

Restriction enzymes cut the DNA at green locations marked with a dashed line: in the transposon and in the flanking genes. When the digested DNA is allowed to re-ligate into circular DNA pieces, iPCR primers bind to primer binding sites (each pair represented by yellow, purple, orange, or blue). The circular DNA is then amplified, generating many copies of a portion of Gene A and Gene B, which is consecutively sequenced.

To identify proteins tagged with tdTomato, genomic DNA is isolated from trapped flies , then digested with one of the following restriction enzymes: MboI, HpyCH4IV, BstUI, or HinP1I. As these sites contain four bases, it is likely that a matching restriction enzyme site lies nearby in the adjacent trapped gene. All four enzymes will cut within the construct at these sites, and in theory within ~2 k bp outside the transposon (Figure 14). When the fragment mixture is heavily diluted and ligated, the digested fragments form a circularized piece of DNA, upon

which added iPCR primers can now amplify a fragment that includes the flanking gene sequence. Sanger sequencing of the iPCR products gives an exact picture of what gene lies nearby either side of the trap- therefore, identifying the protein(s) tagged with tdTomato.

Many small sequence elements make up this overall design, so it made sense to use gene synthesis to put all of these together. Accordingly, we used the services of Integrated DNA Technologies (idtdna.com) to make the initial pCTAF-LHS and pCTAF-RHS plasmids. This process required that the construct be synthesized in two halves because the P-element inverted repeat was not compatible with their synthetic methodology. Figure 12 shows the nucleotide components of the synthesized construct along with their roles and the two 'half-constructs' that were put together to make this. Below I describe the role of each element.

New Gene Trap Sequence Elements

i) *Restriction Enzyme Sites for assembly and vector cloning*

For cloning purposes, restriction enzyme sites for XbaI and XhoI flank the 5' and 3' element ends. To avoid homologous recombination between the inverted repeats within the P-element ends during construct synthesis, an EcoRI restriction site was identified within the tdTomato sequence. This site serves as a reconnection point for the left and right insert halves, which needed to be synthesized in separate intermediate plasmids: pCTAF-LHS and pCTAF-RHS.

TABLE 1: **Restriction Enzyme Sites utilized during cloning process.**	
XbaI Restriction Site	5'-**TCTAGA**-3'
EcoRI Restriction Site	5'-**GAATTC**-3'
XhoI Restriction Site	5'-**CTCGAG**-3'

Restriction enzyme sites used to excise transposon halves out of synthesis plasmids (Left: XbaI and EcoRI, Right: EcoRI and XhoI), reform full transposon in pBluescript, and excise full transposon from pBSK to ligate into pUAST-attB.

ii) 5' and 3' P-element Ends

As used in the Wee-P construct, the 5' and 3' P-element end sequences were included to mobilize the contained exon. The component elements were described above when discussing P-element transposase action. Table 2 gives these exact sequences.

TABLE 2 P-element **end Sequences**	
5' P-element **Sequence**	5′-**CATGATGAAATAACATAAGGTGGTCCCGTCGATAGCCGAAGCTTACCG AAGTATACACTTAAATTCAGTGCACGTTTGCTTGTTGAGAGGAAAGGTTGT GTGCGGACGAATTTTTTTTTGAAAACATTAACCCTTACGTG**-3′
3' P-element **Sequence**	5′-**TTCAAACCCCACGGACATGCTAAGGGTTAATCAACAATCATATCGCTG TCTCACTCAGACTCAATACGACACTCAGAATACTATTCCTTTCACTCGCAC TTATTGCAAGCATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTA TGTTATTTCATCATG**-3′

5' and 3' P-element end sequences, where Transposase will bind *via* the THAP domain and initiate jumping.

iii) Coding Sequences and Splice Sites of the Trap

Most importantly, within the custom transposon lies the Red Fluorescent Protein tdTomato, flanked by two flexible linker sequences as well as a splice acceptor and splice donor sequence. Once spliced to adjacent sequence, the linkers are translated before and after the fluorescent protein- allowing it to 'float' behind the target protein it tags. This is to prevent any misfolding that could result if tdTomato was directly imbedded in its structure.

TABLE 3

Nucleotide Sequence of Splice Sites and Protein Coding Sequences

Splice Acceptor	5′-GGGATTATAATAATATTTCTCTATCTCTTGCATATAG-3′
Linker 1	5′-TCAGGAGGTGGCGGTTCTGGCGGCGGTGGGAGCGGTGGCGGTGGTTCT GGCGGTGGTGGTTCCGGCGGAGGCGGTA-3′ Amino Acid Seq: SGGGGSGGGGSGGGGSGGGGSGGGG
tdTomato	5′-TGGTGAGCAAGGGCGAGGAGGTGATCAAGGAATTCATGAGGTTCAAGG TGAGGATGGAGGGCTCCATGAATGGACATGAGTTTGAAATTGAAGGAGAGG GAGAGGGACGCCCTTATGAAGGCACCCAGACCGCCAAGCTGAAGGTGACCA AGGGCGGCCCCCTGCCCTTCGCCTGGGACATTCTGTCCCCTCAGTTTATGT ATGGATCTAAGGCTTATGTCAAACATCCTGCTGATATTCCCGACTACAAGA AGCTGAGCTTCCCCGAGGGCTTCAAGTGGGAGAGGGTTATGAACTTCGAAG ATGGAGGACTGGTCACAGTCACACAGGATTCCTCCCTGCAGGATGGAACAC TGATTTACAATGTGAAGATGAGGGGCACCAACTTTCCACCCGACGGCCCCG TGATGCAAAAGAAGACAATGGGATGGGAGGCTTCCACAGAACGCCTGTATC CTCGTGATGGAGTCCTGAAAGGAGAGATCCACCAGGCCCTGAAGCTGAAGG ACGGCGGCCACTACCTGGTGGAGTTTAAGACCATTTATATGGCTAAAAAAC CTGTCCAACTGCCTGGATATTATTATGTCGATACAAAACTGGACATCACCA GCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGAGGAGCGAGGGCC GCCATCATCTGTTCCTCTATGGAATGGATGAACTCTATAA-3′
Linker 2	5′-AGCGGTACTGGTGGTTCAGGAGGCACCGGTTCCGGCACAGGCGGTTCG GGCGGTACGGGTAGTGGCACTGGTGGGG-3′ Amino Acid Seq: SGTGGSGGTGSGTGGSGGTGSGTGG
Splice Donor	5′-TGAGCGGATCCATTGTTTTGGATTCCACTGCGTGCATTACACCATCTA TTATCC-3′

Protein coding sequences are represented by Linkers 1 and 2 (gray) as well as tdTomato (red). The EcoRI recognition site (needed for initial cloning purposes only) is shown in green to emphasize how the left and right fragment halves were reconnected.

iv) Trapped-Gene Identification Components (iPCR)

To determine exact trap locations in successfully integrated fly lines, two four-enzyme restriction enzyme site clusters as well as four sets of iPCR primer binding sites are included within the transposon. The choice of enzyme sites was dictated by the need to not have the site between the cluster site and the P-element end. In the original Wee-P construct the iPCR primers were designed against other sequence elements, but we decided to try and make robust predesigned primers. To do this we generated a random 100,000bp sequence and submitted this

to the NCBI primer design tool, then selected the top four primer pairs that did not have any significant homology to the fly genome as determined by an appropriate BLAST sequence search.

TABLE 4

Inverse PCR Involved Component Sequences

iPCR Pair 1 Site	Top: 5′-CTTATTCGGTTCCCGGTATAGC-3′ Bot: 5′-CCGGGCCTAAAGTAAGGTAATC-3′
iPCR Pair 3 Site	Top: 5′- CTTATATCCGCTTGACCGCTAG-3′ Bot: 5′-CCTCAAGGGCAACCAGATATAG-3′
Left Restriction enzyme site cluster	5′-ATCGCGCACGT-3′ MboI: 5′-/GATC-3′ BstUI: 5′-CG/CG-3′ HinP1I: 5′-GC/GC-3′ HpyCH41V: 5′-A/CGT-3′
Left Restriction enzyme site cluster	5′-ACGTGCGCGATC-3′ MboI: 5′-/GATC-3′ BstUI: 5′-CG/CG-3′ HinP1I: 5′-GC/GC-3′ HpyCH41V: 5′-A/CGT-3′
iPCR Pair 2 Site	Top: 5′-GATTAGAGTTCCCGTTCGCTG-3′ Bot: 5′-CCTGCTCTGCGTGTATCTTATC-3′
iPCR Pair 4 Site	Top: 5′-CTATTACACCTGGGCCGATTAG-3′ Bot: 5′-CTCGAATAGCGAGACAGCATAG-3′

Inverse PCR Sequences: iPCR Primer Binding Sites for iPCR Primer Pairs 1 and 3 identify gene left of the transposon, iPCR Pairs 2 and 4 identify gene right of transposon. Sequences for restriction enzyme clusters contain same cut sites, re-arranged to avoid recombination. Note: iPCR Pair 3 Bottom Primer Binding sequence shares a Guanine with the left restriction cluster (highlighted in green).

Post-Synthesis Assembly Plan

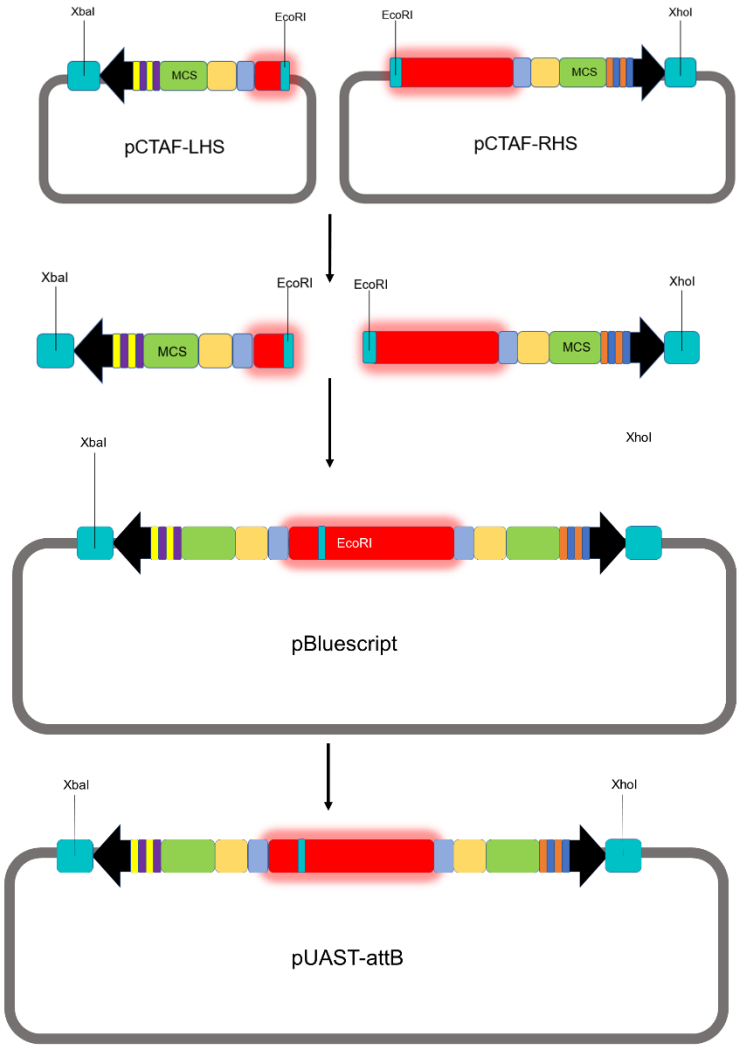

Figure 15: Cloning Workflow to Move Full Construct into Injection Vector

pCTAF halves are extracted with enzymes shown in teal (XbaI, EcoRI, XhoI) and reconnected at the cut sites in pBluescript (pBSK). Enzymes XbaI and XhoI excise pCTAF from pBSK and are used to move the construct into injection plasmid, pUAST-attB.

Upon receipt of the synthesized transposon halves from IDT (plasmids pCTAF-LHS and pCTAF-RHS), the constructs were transformed into *E. coli* XL1 Electrocompetent cells, inoculated into an ampicillin broth, and purified to extract a substantial quantity of plasmid. Figure 15 depicts the cloning workflow: pCTAF-LHS is digested with XbaI and EcoRI restriction enzymes. pCTAF-RHS is digested with EcoRI and XhoI. The reaction is run on an agarose gel, and the half-transposon fragments are excised and purified. The isolated left- and right-hand halves of the transposon are then 'triple ligated' into the plasmid pBluescript that has been pre-digested with XbaI and XhoI. This step assembles the whole transposon.

Following transformation into *E.coli* XLI blue and color selection, any colonies in which the insert is present are grown up and verified by digestion with XbaI and XhoI to identify clones containing the complete transposon. Clones are sequence-verified at this step. Using the same digestion, followed by purification of the complete transposon insert, the transposon is then ligated into pUAST-attB, the *Drosophila* transformation vector. That ligation reaction is transformed once again into XL1 cells, and colony PCR is used to screen for positives. A midiprep plasmid extraction is performed. Positive colonies are grown up and their plasmids sequence-verified both for the insert and for the presence of the P[w^+] selectable marker gene and the attB recombinase target sequence. Site directed mutagenesis is used to create all three reading frame versions required to generate a set of three plasmids: pUAST-attB-CTAF-FR1, pUAST-attB-CTAF-FR2, and pUAST-attB-CTAF-FR3. The complete, verified plasmids are injected into ϕC31 recombinase expressing embryos by a commercial service (Rainbow Transgenic Flies (rainbowgene.com). The company returns P[w^+] positive flies mated to the lab

Figure 16: Frameshift Mutations Made to Generate Frames 2 and 3

To create Frame 2, Tyrosine was added left of Linker 1 and Cytosine + Guanine were added right of Linker 2. To create Frame 3, Guanine + Cytosine were added left of Linker 1 and Tyrosine was added right of Linker 2.

wild-type stock yellow white for stock establishment. See Figure 18 for a description of the fly transformation process.

Site Directed Mutagenesis to Create Frames 2 and 3

In order to make all introns in the genome accessible to trapping transposons for all three splice-site codon positions are needed (i.e. to trap in any reading frame), so three independent fly lines are needed that will be transformed by pUAST-attB-CTAF-FR1, pCUAST-attB-CTAF-FR2, and pUAST-attB-CTAF-FR3.

To create these, point mutations were made between Splice Acceptor Site and Linker 1, as well as Linker 2 and Splice Donor Site (Figure 16). The intent here is to eventually recombine all three constructs into the same genome, as presence of mobilizable elements in three reading frames increases likelihood of a successful fluorescent protein fusion event, making this process more efficient at producing inserts and more rewarding for the Biol230M students. To do this point mutations will be made at the 5’ end of the upstream linker sequence just after the acceptor

splice site as well as at the 3’ end of the downstream linker just before the donor splice site. Figure 16 depicts the point mutation locations/nucleotides made to generate Frames 2 and 3 from the original construct, Frame 1.

Fly Transformation of Constructs into Syncytial Stage Embryos

To generate transformed lines containing pUAST-attB-CTAF-FRn (n=1,2, or 3) the plasmids must be injected into syncytial stage *Drosophila* embryos with the *white* eye phenotype that also express the ϕC31 recombinase. Due to the attB site on injection vector pUAST-attB and the presence of recombinase, recombination into chromosomes with an attP landing will occur. By doing this at the syncytial stage all nuclei are accessible to the DNA, but injection is done at the posterior end of the embryo to ensure that DNA enters the pole cells which will give rise to the gamete progenitor cells in the germline. With the mechanism depicted in Figure 17, pUAST-attB containing pCTAF and a *white+* gene recombines onto Chromosome 3. Figure 17 illustrates the chain of events following injection of pUAST-attB into the syncytium.

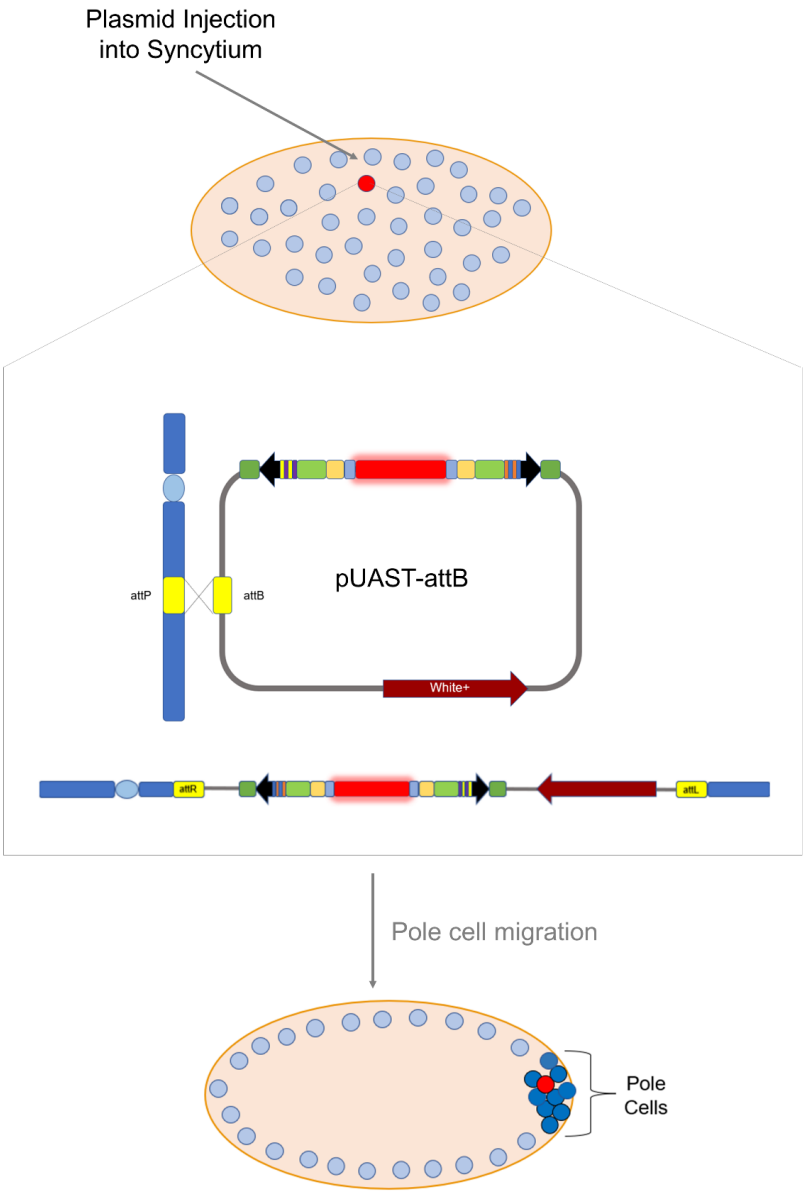

Figure 17: Recombination of pUAST-attB onto Chromosome 3.

φC31 mediates recombination of injected plasmid through attB and attP sites. If recombined into a pole cell nucleus, construct sequence is incorporated into gametes and passed on to offspring.

CHAPTER 3

Materials and Methods:

i. Bacterial Media

2xYT Liquid Media (500 mL): To 500 mL of double distilled water, 8 g Tryptone + 5 g Yeast Extract + 2.5 g NaCl were added and stirred until dissolved completely. Media was aliquoted as necessary in proper glassware depending on volume needed. Foil and autoclave tape was used to cover the flask opening and the components were autoclaved. Post-sterilization, Ampicillin was added to ensure a final Ampicillin concentration of 100 μg/ml.

2xYT Agar Plates (makes 20 plates): To 500 mL of double distilled water, 8 g Tryptone + 5 g Yeast Extract + 2.5 g NaCl were added and stirred until dissolved completely. 7.5 g of Bacteriological Agarose was added to solution and stirred until uniform. Foil and autoclave tape was used to cover the flask opening and the components were autoclaved. Post-sterilization 500 μl of 100mg/mL Ampicillin was added to cooled media and stirred. Media was then poured into 87 mm Petri Dishes.

Xgal-IPTG Plates: These are made with the same recipe as the 2xYT Agar Plates, but in addition to Ampicillin, 250 μl of 40 mg/mL X-galactosidase and 500 μl of 100 μM IPTG (Isopropyl ß-D-1-thiogalactopyranoside) were stirred into media cooled to 55-60^{o}C. Media was then poured into 87 mm Petri Dishes.

ii. Transformations

1-10 ng of plasmid was added to 50 µl of Electroporation-Competent Cells in a 1mm electroporation cuvette. The cuvette was placed into the BioRad Pulser Machine, pulsed at 1.7 kV, immediately added to 1 mL of 2xYT that was prewarmed to 37ºC, and shaken at 250 rpm for 1 hr. at 37ºC. After recovery cells were plated onto 2xYT + Ampicillin agar plates and incubated overnight at 37ºC.

iii. DNA Extraction

Plasmid DNA Mini-Prep from Bacteria

2 µl of 100 mg/ml Ampicillin were added to each of 18 sterile 2 ml 2xYT broth aliquots. Typically, 18 well isolated colonies were selected from transformation plates with a sterile toothpick and streaked onto a fresh agarose plate. Each toothpick was dropped into a 2 mL 2xYT + Ampicillin and shaken at 250 rpm and 37ºC overnight. The plate was incubated overnight at 37ºC. Eppendorf Tubes were filled with 1.5 mL of overnight culture, spun for 2 minutes, and resuspended in 200 µl STET buffer (10mM Tris-HCl, 1mM EDTA, 100mM NaCl, and 5% Triton X-100) with a Lysozyme concentration of 1.25 µg/µl. After 5 minutes at room temperature the tubes were boiled for 45 seconds and spun for 10 minutes at 15,000 g. Pellets were removed and 5 µg of 1mg/ml RNase was added, then incubated at 37ºC for 15 minutes. DNA was precipitated by adding 20 µl CTAB buffer (5% CTAB, 0.5 M NaCl), then centrifuged at 15,000g for 5 minutes. Pellets were suspended in 300 µl of 1.2 M NaCl solution and 700 µl 100% Ethanol was added, followed by centrifugation at 15,000 g for 30 minutes. The pellet was then rinsed 3x with 1 ml 70% Ethanol and dried. Dry pellets were resuspended in 1 mL 1x 10mM Tris 1mM EDTA buffer.

Midi-Prep: Extracting Plasmid from Bacteria (Large Scale, High Purity)

This procedure used the QIAGEN® Plasmid Midi Kit. The following protocol was adapted from the QIAGEN® Plasmid Purification Handbook. Buffer compositions not given, as their recipes are confidential intellectual property of QIAGEN®.

A 50 ml 2xYT liquid broth was inoculated with a single colony on a sterile toothpick. The culture was grown 37°C for 12–16 h with vigorous shaking (approx. 300 rpm). Culture was transferred to centrifugation tubes and spun at 6000 x g for 15 min at 4°C. Bacterial pellets were resuspended in 4 ml Buffer P1 (Resuspension Buffer). 4 ml of Buffer P2 (Lysis Buffer) was added, and tubes were inverted 4-6 times, then incubated at room temperature for 5 minutes. Next 4 ml of chilled Buffer P3 (Neutralization Buffer) was added to tubes before mixture was inverted, then held on ice for 15 minutes. Samples were centrifuged at ~20,000 x g for 30 min at 4°C. For each culture, a QIAGEN-tip 100 was equilibrated with 4 ml Buffer QBT (Equilibration Buffer). Supernatant from sample tubes were added to the QIAGEN-tip and allowed to pass through the column resin. The QIAGEN-tip was washed with 2 x 10 ml Buffer QC (Wash Buffer). DNA was then eluted with 5 ml Buffer QF (Elution Buffer). To precipitate DNA, 3.5 ml (0.7 volumes) of 100% isopropanol was added to the elution. Samples were centrifuged at 15,000 x g for 30 min at 4°C. The DNA pellet was lightly washed with 2 ml of room-temperature 70% ethanol and centrifuged at 15,000 x g for 10 min. Without disturbing the pellet, supernatant was removed and the pellet air dried overnight. DNA was resuspended in 500 µl 10 mM Tris-Cl, pH 8.5.

Fly DNA Extraction

20 flies were anesthetized with carbon dioxide and placed in a 1.7 mL Eppendorf tube with 150 µl 'Flybuffer' (100mM Tris-HCl pH 7.5, 100mM EDTA, 100mM NaCl, 0.5% SDS). Once suspended flies were ground until the exoskeleton was well-fragmented, and the grinder was washed with 350 µl of Flybuffer over the tube. The Eppendorf tube was incubated at 65ºC for 30 minutes, then 70 µl of 8M Potassium Acetate was added to solution. After an additional incubation on ice for 30 minutes, the tube was centrifuged at 15,000 x g for 10 minutes and supernatant was moved to a fresh 1.7 mL tube. 570 µl of Phenol/Chloroform/IsoAmyl Alcohol (50:49:1) was mixed-in vigorously with supernatant and spun at 15,000 x g for 3 minutes. The aqueous phase was moved to a fresh tube and mixed with 570 µl of Phenol Chloroform once more, then centrifuged in the same fashion. At this stage the phenol phase was removed, and the aqueous phase was back extracted three times with 1 mL of Chloroform to remove residual Phenol. To the final aqueous phase, 63 µl (or 1/9 volume) of 3M Sodium Acetate pH 5.2 and 1.2 mL 100% Ethanol (or 2 volumes) were added, then mixed and centrifuged for 10 minutes. Supernatant was poured off and 1 mL of 70% Ethanol was gently added to the pellet without disturbing it. The tube was spun for 3 mins at 15,000 x g and the supernatant was poured off. To dry the pellet, the tube lid was opened and centrifuged in a Speedvac machine. Dried pellets were resuspended in 50 µl 1mM 10mM Tris 1mM EDTA buffer.

iv. Molecular Cloning, Nucleic Acid Manipulation Techniques

Agarose Gel Electrophoresis

To create an agarose gel, 17.5 mL of 20x TBE Buffer (Tris Base, Boric Acid, 0.5 M EDTA pH 8.0) was added to 332.5 mL ddH_2O. 15 µl Ethidium Bromide was added to solution

and mixed. 40 mL of this solution was poured into a 250 mL Erlenmeyer flask and boiled with 0.6 g Agarose to bring concentration to 1.5% (to create gels of different percentages [x%] : 40 mL * 0.X = grams of agarose). The mixture was then poured into an agarose gel mold and allowed to cool at 4ºC refrigerator for 30 minutes. Once the comb was removed, the gel was placed into the gel rig and remaining 310 mL of originally composed buffer is poured to cover the gel. A 4:1 ratio of DNA sample to DNA dye was established, and dyed samples were loaded onto the gel. Generally, gels were run between 90-100 V for 90 minutes.

Diagnostic Restriction Digests

Diagnostic digests were conducted after each step of the cloning process to verify each intermediate. 1.5 µg of plasmid DNA is digested per miniprep. To each digest of LHS plasmid, 2-µl of 10x CutSmart Buffer, 0.5 µl 100x BSA (bovine serum albumin), 0.5 µl XbaI, 0.5 µl EcoRI, 0.8 µl 0.5 mM Spermidine, 1.5 µg DNA, and ddH_2O were added to bring reaction volume to 20 µl. To each digest of RHS plasmid, 2 µl of 10x CutSmart Buffer, 0.5 µl 100x BSA, 0.5 µl EcoRI, 0.5 µl XhoI, 0.8 µl 0.5 mM Spermidine, 1.5 µg DNA, and ddH_2O were added to bring reaction volume to 20 µl. Digests were incubated at 37ºC for 1 hour. To verify insertion into pUAST after the triple ligation (below), 2 µl of 10x CutSmart Buffer, 0.5 µl 100x BSA, 0.5 µl XbaI, 0.5 µl EcoRI, 0.5 µl XhoI, 0.8 µl 0.5 mM Spermidine, 1.5 µg DNA, and ddH_2O were added to bring reaction volume to 20 µl.

Large Scale Restriction Digests

To extract a large amount of each transposon-half sequence for fragment purification and ligation, bulk restriction digests were conducted on midiprep pCTAF-FR1-LHS and pCTAF-

FR1-RHS respectively. 10 µg of DNA was digested per reaction. To each bulk digest of LHS, 10 µl of 10x CutSmart Buffer, 1 µl 100x BSA, 2 µl XbaI, 2 µl EcoRI, 4 µl 0.5 mM Spermidine, 10 µg DNA, and ddH_2O were added to bring reaction volume to 100 µl. To each bulk digest of RHS, 10 µl of 10x CutSmart Buffer, 1 µl 100x BSA, 2 µl EcoRI, 2 µl XhoI, 4 µl 0.5 mM Spermidine, 10 µg DNA, and ddH_2O were added to bring reaction volume to 100 µl. To digest the complete reformed transposon out of pBSK, 10 µl of 10x CutSmart Buffer, 1 µl 100x BSA, 2 µl XbaI, 2 µl EcoRI, 2 µl XhoI, 4 µl 0.5 mM Spermidine, 10 µg DNA, and ddH_2O were added to bring reaction volume to 100 µl.

Gel-DNA Recovery

This protocol is taken from the Zymoclean© Gel DNA Recovery Kit. The DNA fragment was excised from an agarose gel using a razor blade, weighed, and transferred to a 1.7 microcentrifuge tube. 3x volumes of ADB (agarose dissolving buffer) was added to the excised agar fragment (e.g., for 100 µl (mg) of agarose gel slice, 300 µl of ADB added) then incubated 5-10 minutes at 37ºC. Melted agarose solution was transferred to a Zymo-Spin™ Column in a Collection Tube and centrifuged for 30 seconds. 2x of 200 µl of DNA Wash Buffer was added to the column and each was centrifuged for 30 seconds. 10 µl of DNA Elution Buffer was added to the column and centrifuged for 60 seconds at 15,000 x g over a 1.7 mL microcentrifuge tube.

Triple Ligation into pBSK

EcoRI/XbaI digested LHS and EcoRI/XhoI digested RHS fragments were co-precipitated with XbaI/XhoI digested pBSK at a 3:3:1 molar ratio of LHS:RHS:plasmid. Since pBSK is ~3000 bp, LHS is approximately ~400 bp, and RHS is approximately ~1000 bp, 80 ng of LHS

fragment, 200 ng RHS fragment, and 200 ng digested pBSK were included in the ligation. In practice 6 µl of 33 ng/µl digested pBSK, 2 µl 40 ng/ µl LHS, 3 µl 66 ng/ µl RHS, 16 µl ddH_2O, 3 µl of 3 M Sodium Acetate, and 150 µl 100% Ethanol was mixed in a 1.7 mL Eppendorf tube. This mixture was centrifuged at 15,000 x g for 30 minutes, and supernatant was poured off. 1 mL of 70% Ethanol was used to gently rinse the pellet and centrifuged at 15,000 x g for 2 minutes. The pellet was dried and resuspended in 9 µl ddH_2O, 1 µl Ligase Buffer, and 0.5 µl Ligase. The solution was incubated at 16ºC overnight.

Ligation of completed transposon into pUAST-attB

To ligate the excised complete insert into pUAST-attB, a molar ratio of 10:1 fragment to plasmid ratio was utilized. As the insert size was 1.5 kb, and pUAST-attB predigested with XbaI and XhoI was 8.5 kb, 88 ng of insert was coprecipitated with 50 ng vector. 10 µl of 5 ng/ µl pUAST-attB, 18 µl of 5 ng/ µl insert, 3 µl of 3 M Sodium Acetate, and 78 µl of 100% Ethanol were mixed in a 1.7 mL Eppendorf tube. This mixture was centrifuged at 15,000 x g for 30 minutes, and supernatant was poured off. 1 mL of 70% Ethanol was used to gently rinse the pellet and centrifuged at 15,000 g for 2 minutes. The pellet was dried and resuspended in 9 µl ddH_2O, 1 µl Ligase Buffer, and 0.5 µl Ligase. The solution was incubated at 16 ºC overnight.

Colony PCR

Two primer Sequences:

To test for successful insert ligation in pUAST, 100 transformant colonies were tested via Colony PCR between the gene trap P-element end and sequences flanking the pUAST multiple cloning site. (P-End (OUT): 5’- CGACGGGACCACCTTATGTTATTTCATCATG-3’

UAST2: data sheet misplaced between synthesis date in 2007 and now). A 25x PCR Master Mix was created from 50 µl 10x CutSmart Buffer, 50 µl 2mM dNTP mix, 50 µl 4mM P-end(OUT) primer, 50 µl 4mM UAST2 primer, 5 µl undiluted Taq Polymerase, and 295 µl ddH_2O. 20 µl of Master Mix was aliquoted into 20 PCR tubes, upon which 5 colonies were streaked and inoculated into each tube (multiplexing). The following PCR protocol was used to amplify a small portion of P-element end:

Step 1	98 ºC, 30 seconds
Step 2	98 ºC, 10 seconds
Step 3	49 ºC, 30 seconds
Step 4	72 ºC, 40 seconds
Step 5	72 ºC, 10 minutes

Steps 2-4 are repeated for 34 cycles.

Site Specific Mutagenesis through PCR

To generate pCTAF-FR2 and pCTAF-FR3 from pCTAF-FR1 template, Agilent software was used to generate mutagenesis primers with point mutations. The Agilent Site-Specific Mutagenesis QuikChange II Kit was used to carry out reactions on pCTAF-FR1-LHS and pCTAF-FR1-RHS.

<u>Primers used for generating pCTAF-FR2-LHS and pCTAF-FR3-LHS:</u>

Frame 2 LHS Forward Primer: 5’- GCATATAGTTCAGGAGGTGGCGGTTCTG -3’

Frame 2 LHS Reverse Primer: 5’- TCCTGAACTATATGCAAGAGATAGAG- 3’

Frame 3 LHS Forward Primer: 5’- CATATAGCGTCAGGAGGTGGCGGTTCTG -3’

Frame 3 LHS Reverse Primer: 5'- TCCTGACGCTATATGCAAGAGATAGAG - 3'

Reaction Components: 40 ng (1 µl) FR1-LHS template, 5 µl 10x Reaction Buffer, 125 ng (4 µl) Frame 2 LHS or Frame 3 LHS Forward Primer, 125 ng (4 µl) Frame 2 LHS or Frame 3 LHS Reverse Primer, 1 µl dNTP mix, 1 µl Pfu Polymerase, and 34 µl ddH_2O was added to a PCR tube.

LHS Mutagenesis PCR Protocol:

Step 1	95 ºC, 30 seconds
Step 2	95 ºC, 30 seconds
Step 3	48.5 ºC, 1 minute
Step 4	68 ºC, 3 minutes

Steps 2-4 are repeated for 12 cycles.

<u>Primers used for generating pCTAF-FR2-RHS and pCTAF-FR3-RHS:</u>

Frame 2 RHS Forward Primer:5'-TGGTGGGGCGTGAGCGGATCCATTGTTTTGG-3'

Frame 2 RHS Reverse Primer: 5'-GCTCACGCCCCACCAGTGCCACTACCC -3'

Frame 3 RHS Forward Primer: 5'-CTGGTGGGCGTGAGCGGATCCATTGTTTTGG-3'

Frame 3 RHS Reverse Primer: 5'- GCTCACGCCCACCAGTGCCACTACCC -3'

Reaction Components: 40 ng (1 µl) FR1-RHS template, 5 µl 10x Reaction Buffer, 125 ng (3 µl) Frame 2 RHS or Frame 3 RHS Forward Primer, 125 ng (3 µl) Frame 2 RHS or Frame 3 RHS Reverse Primer, 1 µl dNTP mix, 1 µl Pfu Polymerase, and 36 µl ddH_2O was added to a PCR tube.

RHS Mutagenesis PCR Protocol:

Step 1	95 ºC, 30 seconds
Step 2	95 ºC, 30 seconds
Step 3	62 ºC, 1 minute
Step 4	68 ºC, 4 minutes

Steps 2-4 are repeated for 12 cycles.

Inverse PCR

To catalog launch-site and any non-specific bands generated by iPCR primers with *Drosophila* genomic DNA, DNA was extracted from flies with the construct P[w^+] and *yellow white* as a control. Both P[w^+] and *yw* DNA were digested in four separate diagnostic digests with 5 µl DNA : MboI, HinP1I, HpyCH4IV, and BstUI for a total of 8 digests. MboI, HinP1I, and HpyCH4IV digests were incubated at 37ºC for 1 hour and heat inactivated at 65ºC for 20 minutes. BstUI digests were incubated at 60ºC for 1 hour, and DNA was extracted with Phenol/Chloroform. Digests were resuspended in 900 µl of distilled water, and 100 µl 10x Ligase Buffer, and 20 µl Ligase. This was incubated at 16ºC overnight.

To amplify circularized DNA, the following iPCR primers were used:

<u>iPCR Pair 1:</u>

1Top: 5’- GCTATACCGGGAACCGAATAAG- 3’

2Bot: 5’- GGCCCGGATTTCATTCCATTAG- 3’

<u>iPCR Pair 2:</u>

3Top: 5’- CAGCGAACGGGAACTCTAATC -3’

4Bot: 5’-GGACGAGACGCACATAGAATAG -3’

iPCR Pair 3:

5Top: 5’- CTAGCGGTCAAGCGGATATAAG- 3’

6Bot: 5’- GGAGTTCCCGTTGGTCTATATC- 3’

iPCR Pair 4:

7Top: 5’-CTAATCGGCCCAGGTGTAATAG -3’

8Bot: 5’-GAGCTTATCGCTCTGTCGTATC -3’

A 40x PCR Master Mix was made from 240 µl ddH_2O, 80 µl of Standard Taq Buffer, 4 µl undiluted Taq Polymerase, and 80 µl of 2 mM dNTP mix. 11 µl of Master Mix was added to 32 PCR tubes. 5 µl of each circularized ligation was added to 4 PCR tubes each (8 groups of 4). Within each ligation group of 4, 2 µl of corresponding top and bottom primers (1Top+2Bot, 3Top+4Bot, 5Top+6Bot, 7Top+8Bot) were added. The following iPCR protocol was used:

Step 1	98 ºC, 30 seconds
Step 2	98 ºC, 10 seconds
Step 3	49 ºC, 30 seconds
Step 4	72 ºC, 40 seconds
Step 5	72 ºC, 10 minutes

Steps 2-4 are repeated for 34 cycles.

v. Fly Husbandry

Fly Food Formula: The following dry ingredients were combined in a steam kettle: 70 g Agar, 1,100 g Sugar, 274 g Yeast, and 520 g of Cornmeal. 23.8 g of Tegosept was dissolved in 92 mL of 95% ethanol. Dry ingredients (Agar, Sugar, Yeast, and Cornmeal) were added with a sieve to 8.5 mL of vigorously stirring water. Heat was applied and the mixture was allowed to boil at which point heat was turned off, and the mixture continued to stir for 15 minutes. Water levels were brought back to 8.5 L due to evaporation. If necessary, heat was applied until the mixture maintained, but did not exceed 80ºC. Tegosept/Ethanol solution as well as 15 mL of Propionic Acid were added to food and mixed. Once the food had thickened, the mixer was turned off and food pumped into vials and bottles.

Grape Plate Formula: 136 mL of distilled water and 114 mL of grape juice were brought to a boil within a microwave. Once transferred to a heated stirrer, 21.75 g granulated sugar and 4.5 g dry yeast were added to solution. 5.5 g Technical Agar was added in small increments to avoid clumping. This mixture has allowed to heat/stir for 10 minutes, or until no clumps were visible. Solution was then removed from the heated stirrer and allowed to cool until reaching 55-66ºC. Once cooled, the mixture was returned to the stirrer and 2.75 mL 1.25M NaOH and 2.8 mL Acid Mix A (250 mL distilled water, 209 mL Propionic acid, 20.8 mL Phosphoric acid) were added. The mixture was then poured into 37mm petri dishes.

Female Virgin Collection: Once a bottle with no adult flies (with visible eggs ready to hatch) has been cleared, any female animal born within a 16 hour window is guaranteed to be a virgin. Collection of female virgins occurred at least once every 16 hour window, with the clock

being ‘reset’ every time the bottle was cleared post-collection. Males were not collected by a rigorous timetable, as virginity status did not impact offspring of the future crosses.

Fly Cages: 20 males and 40 females destined for crossing are pre-sorted and placed in vials (See Cross 2: Figure 20). Granulated yeast was sprinkled onto 25 Grape Plates. Once prepared, a vial containing approximately 60 flies was dumped into the cage (a beaker with netting to allow for airflow) and a Grape Plate was wedged over the opening to prevent fly escape. Grape Plates were replaced every three days, while original 60 flies remained in the cage.

Larval Screening: Approximately 6-7 days after egg laying (see Figure 19), grape plates containing larvae was thoroughly masticated and could be gently scooped with a spatula and dispersed into room temperature water. Once media dissolved, the mixture was poured over nylon netting to filter out the larvae. The larvae were directly screened on this netting using a dissection microscope equipped for red fluorescence from NIGHTSEA. Candidate fluorescent larvae were removed from the net and accumulated in a petri dish containing water. Larvae were then gently introduced to a regular culture vial and allowed to mature. To generate lines stemming from fluorescent larvae, *y w* flies of the opposite sex were introduced to the vial and allowed to mate with individual candidate fluorescent flies. If the fly was a genuine positive this will generate a 50:50 Mendelian ratio of fluorescent: non-fluorescent offspring, the fluorescent ones of which can be used to establish a stock.

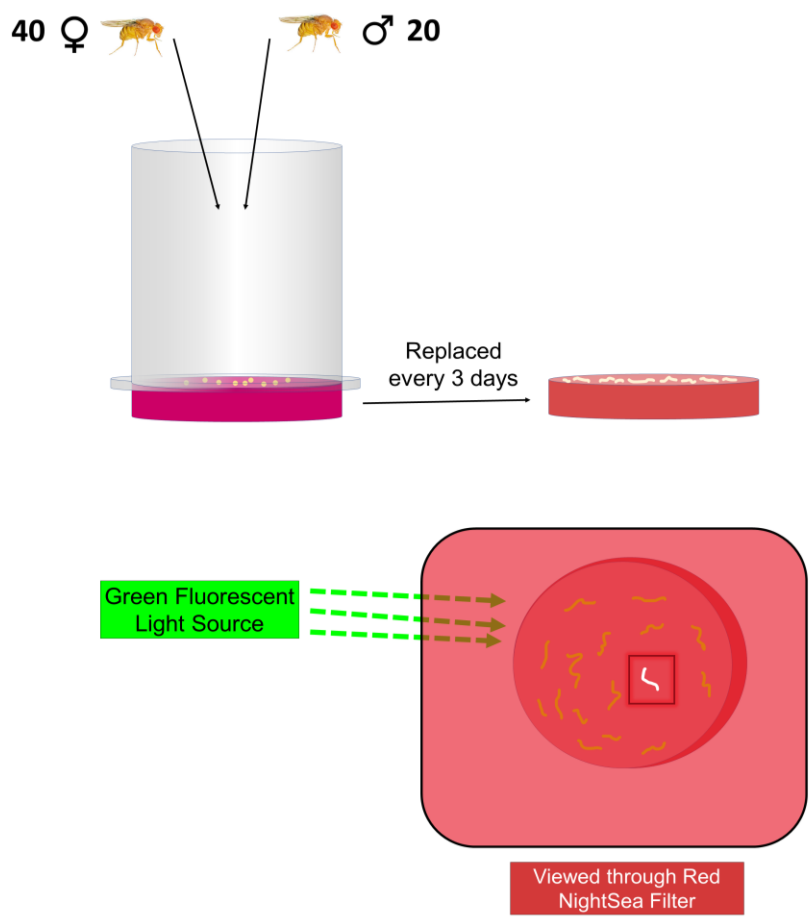

Figure 18: Fly Cage set-up to generate and easily screen fluorescent larvae.

Grape plates were wedged into tricorner beakers containing Cross 2 flies. Every three days, a fresh plate was inserted. Plates containing larvae were allowed to mature another 6-7 days (or until larvae reached 3rd instar stage). A green fluorescent light source excites tdTomato in positive larvae, causing bright red light emission, observable through a red filter.

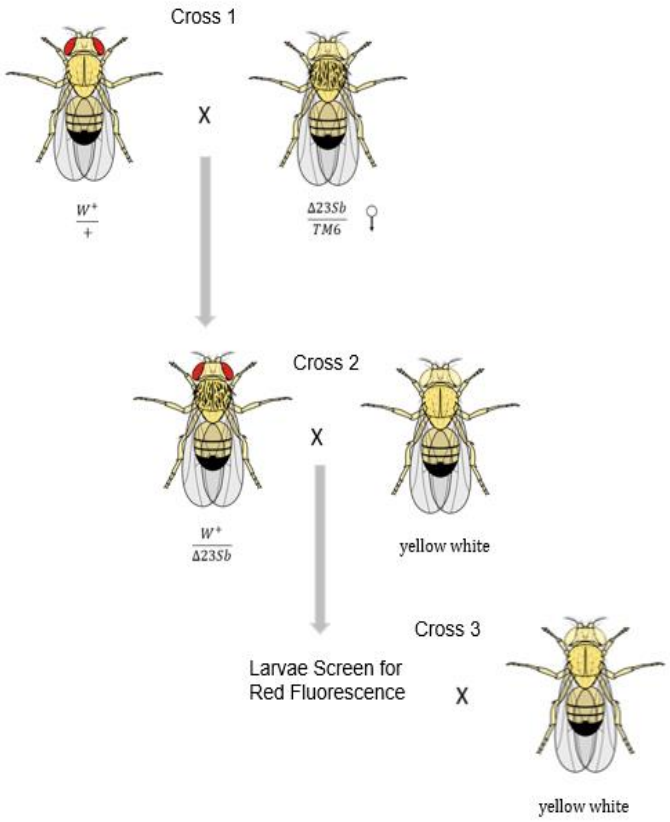

Figure 19: Crosses to mobilize transposon.

Flies with the P[w^+] gene (red eyes) are known to contain the transposon, as this is a marker phenotype induced by presence of pUAST-attB-pCTAF. In cross 1, red eyed females were mated to males containing a transposase source (Δ2-3) linked to the *Stubble* (*Sb*) phenotypic marker. Offspring were selected based on red eyes and Stubble hairs; the transposon will be mobilized in these flies due to the Δ2-3 transposase. To produce offspring with a single trap, males with the mobilized transposon were crossed to virgin *yellow white* females. Larvae produced from Cross 2 were screened for fluorescence. If detected, the fluorescent larvae were crossed to *yellow white* flies of the appropriate sex to start to generate a line containing an established gene trap.

CHAPTER 4

Results and Discussion

In this chapter I display photographs and explanations of agarose gels used to monitor the success of the cloning process, from handling of initial synthesized transposon halves in synthesis plasmids, to purification of pUAST-attB-pCTAF injection compatibility plasmid. I will also discuss the outcome of crosses used to mobilize and establish the transposon CTAF as a gene trap. To verify plasmid presence in a colony selected from pCTAF-LHS and pCTAF-RHS transformations, minipreps of these colonies were ran on a 1.5% Agarose gel. An example

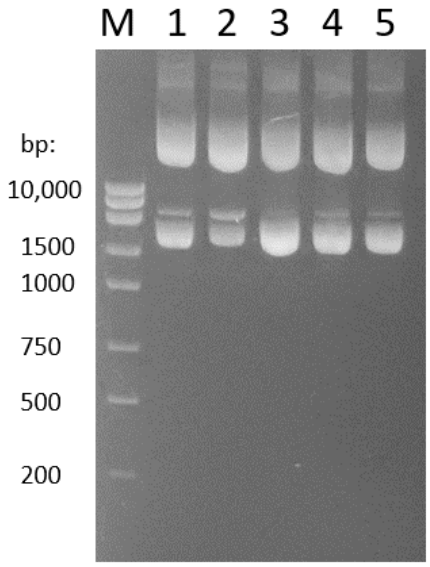

Figure 20: Successful transformation of pCTAF-LHS and pCTAF-RHS

The synthesized constructs from IDT were transformed and Colony Miniprep Samples were run on Agarose Gel. The bands on this gel indicate that this was successful. DNA from one of these colonies for each half was sequence verified (see text).

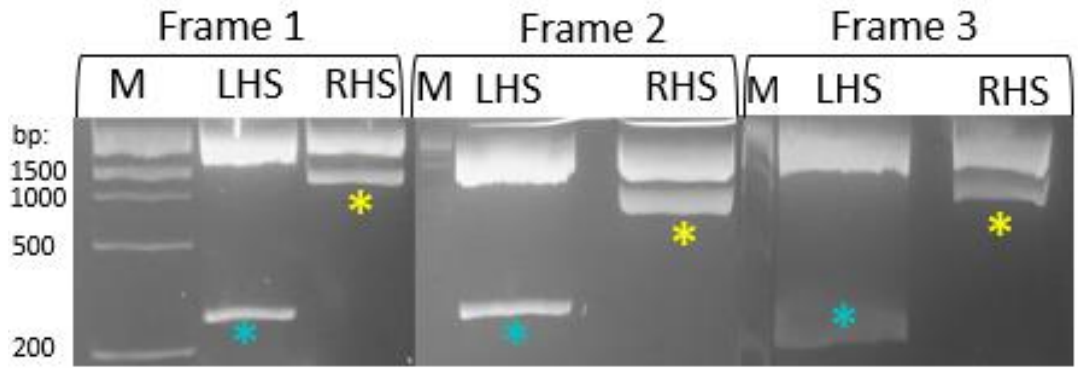

Figure 21: Isolated Left-Hand Side and Right-Hand Side Fragments

LHS is denoted by the blue star (393 bp), while RHS is denoted by the yellow star (1072 bp). As ladders in the Frame 2 and 3 gels are hard to distinguish, gel photos were included alongside Frame 1 to show how distance traveled is indicative of fragment size.

miniprep test gel is shown in Figure 20. From positive colonies, a midiprep was created and sequence verified to that the synthesis had gone correctly.

To extract the insert halves, bulk restriction enzyme digests were used to excise insert portions out of synthesis plasmid. For the left half of CTAF, XbaI and EcoRI were used to cut, and for the right side, EcoRI and XhoI were used to cut. The expected band size for the LHS was 393 bp, for the RHS it was 1072 bp. Fragment digests were conducted for the halves of all three frames: Frame 1, Frame 2, and Frame 3, as shown in Figure 21. Fragment DNA was then extracted and purified from the gel.

Next, the LHS and RHS fragments were joined to complete the transposon in the plasmid pBluescript for each of the three frames in a triple-ligation and using blue/white colony selection. White colonies were selected and miniprepped to test for successful insert creation. Figure 22 shows a positive colony digested in a diagnostic assay to establish presence of fragment halves. This proved that the triple-ligation was successful and sequencing subsequently confirmed

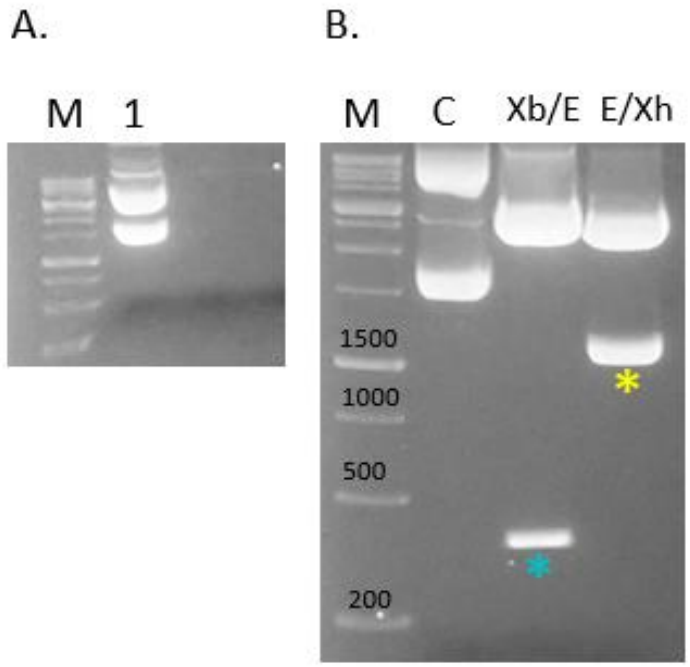

Figure 22: Verifying Success of Frame 1 Triple Ligation

Panel A shows the miniprep created from a positive Frame 1 triple-ligation transformant: Colony 1 (1) was chosen. Panel B shows the diagnostic digest of Colony 1 miniprep, with an undigested control (C), XbaI/EcoRI digest (Xb/E) to establish LHS, and EcoRI/XhoI digest (E/Xh) to establish RHS.

presence the full transposon. Miniprep checks and diagnostic digests were repeated for Frame 2 triple ligation transformants as seen in Figure 23. At the time of writing, Frame 3 transformants have been purified, but have not yet been digested to verify complete insert presence.

Following reformation of the construct in pBSK, the entire fragment was excised from the vector with XbaI and XhoI, restriction sites that flank the 5' and 3' P-element end sequences of the construct. Figure 24 displays the gel from which the reconnected fragment was excised.

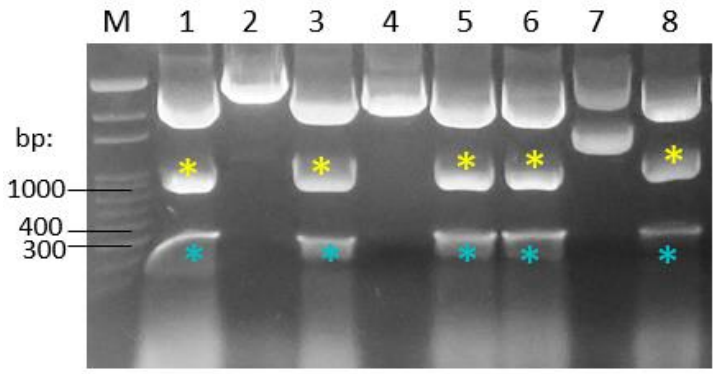

Figure 23: Verifying Success of Frame 2 Triple Ligation

Minipreps containing Frame 2 plasmid were digested with XbaI, EcoRI, and XhoI. Colonies 1, 3, 5, 6, and 8 were shown to contain both the Left-Hand Side (blue star) and the Right-Hand Side (yellow star).

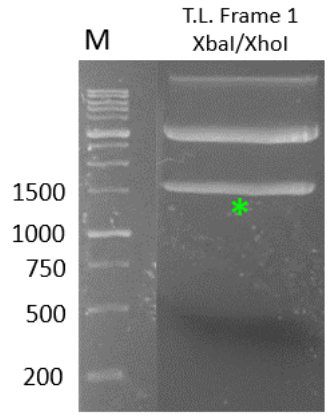

Figure 24: Cutting out Frame 1 from pBSK Triple Ligation

The full Frame 1 fragment (green star) of 1465 bp was isolated and purified

The full insert (FR1-CTAF) was then subcloned into the recombination vector pUAST-attB. To do so, pUAST-AttB was predigested with XbaI and XhoI, to match the restriction enzyme cut ends of FR1-CTAF. Transformants with this ligation were identified with Colony PCR with 5-fold multiplexing to increase the number screened, as shown in Figure 25. The

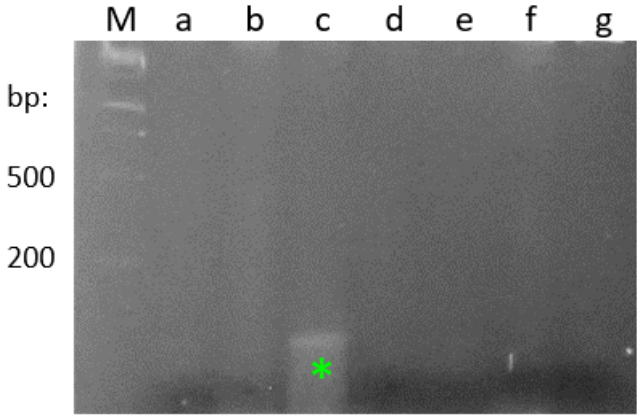

Figure 25: Multiplexed Colony PCR amplification of pUAST-attB-FRI-CTAF

Positive amplicon denoted with a green star in multiplex group c (colonies 11-15). From this data, it is hypothesized that one or more of the colonies within group c (11, 12, 13, 14, 15) contains the full insert as excised from pBSK.

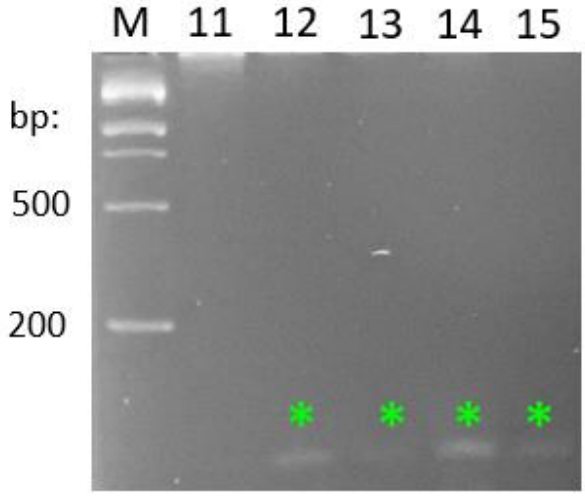

Figure 26: Colony PCR from individual colonies within Group C

Amplicons are denoted with a green star, indicative that a 5' P-element end is present in transformants.

product size was expected to be less than 200 bp, as the amplification region is from right outside the MCS to the edge of the 5' P-element end. A positive band was clearly identified in one mix.

To verify that at least one colony was indeed a positive transformants, the original 5 colonies from the mix were individually amplified with the same primers (UAST2 and Pend) in 5 separate PCR reactions. It was observed that colonies 12, 13, 14, and 15 contained the transformants with a 5' P-element end, from the gel pictured in Figure 26. As the statistical chance of this happening in only one multiplex reaction is low, it seems likely that many more multiplexed PCR reactions contained positive transformants. Smaller DNA fragments require higher agarose density to be best visualized, so positives from the other PCR reactions would have likely been more visible on a higher percent gel (4%) as opposed to the 2% that was used. Also, the small size of the target band made it hard to see with ethidium bromide, so perhaps many were missed because they only had 1/5 or 2/5 positive colonies in the mix, and it was the coincidence of 4/5 that permitted us to spot this one set.

Plasmid was extracted from both Colonies 12 and 14, to generate large amounts of pure pUAST-attB-FR1-pCTAF. These were diagnostically digested to verify presence of all insert components. From the gel shown in Figure 27, it was proven that the full insert was present in both transformants. Both purified plasmids were sequence verified, and pUAST-attB-FR1-attB was sent to Rainbow Transgenic Flies for fly transformation.

Once flies with established construct recombination in the germline (evident from red eyes) were obtained, launch site and nonspecific iPCR products for each primer pair/enzyme cluster site were established by comparing P[w^+] digested DNA to a *yw* control. iPCR product bands using the same enzyme/primer pair were compared to identify bands present in P[w^+] DNA, but not *yw.* Specific bands were determined using the gel shown in Figure 29.

Crosses depicted in Figure 19 (mobilization of transposon and generation of trap lines) were carried out according to plan, but no fluorescent larvae were detected from initial pilot

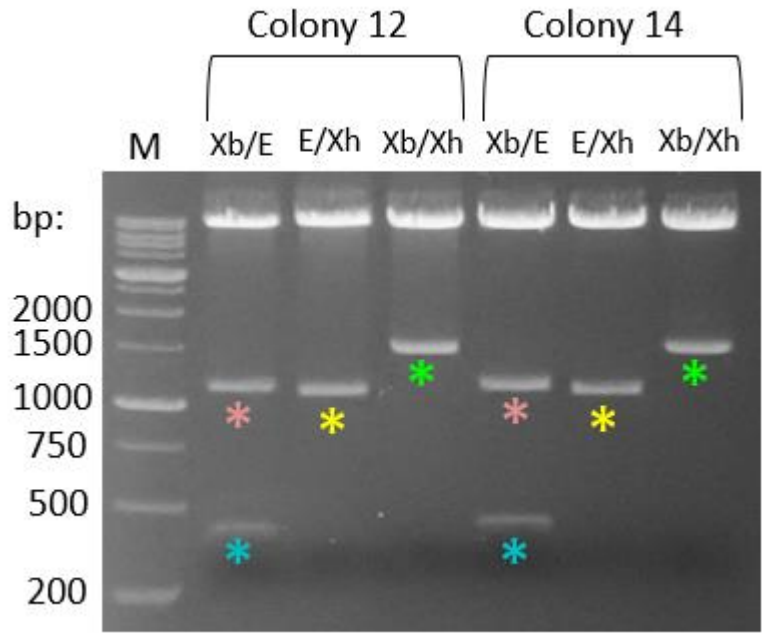

Figure 27: Diagnostic digests of purified pUAST-attB-FR1-CTAF

Plasmid from Colony 12 and Colony 14 transformants was digested in three separate reactions: XbaI/EcoRI, EcoRI/XhoI, and XbaI/XhoI. The left-side presence is confirmed by the 393 bp band (blue star). Bands marked with a pink star represent the remainder of the insert + MCS sequence (pUAST has an EcoRI site). The right-side presence is confirmed by the 1072 band (yellow star). A final digest with XbaI/XhoI verified complete fragment presence (green star) of 1465 bp.

screen. Only Frame 1 (pCTAF-FR1) was used to conduct the pilot screen, Frames 2 and 3 (pCTAF-FR2, pCTAF-FR3) were held at the Triple Ligation in pBSK stage to see if the pilot was successful. As it appears that Frame 1 did not mobilize, the cloning of Frames 2 and 3 was paused in order to diagnose the problem.

While no tdTomato traps were observed in offspring of Cross 2 when using pCTAF-FR1 flies, I noticed bacterial pellets during the pCTAF-FR2 cloning process were pink under white light. When placed under a green fluorescent light source and viewed through a Red NIGHTSEA

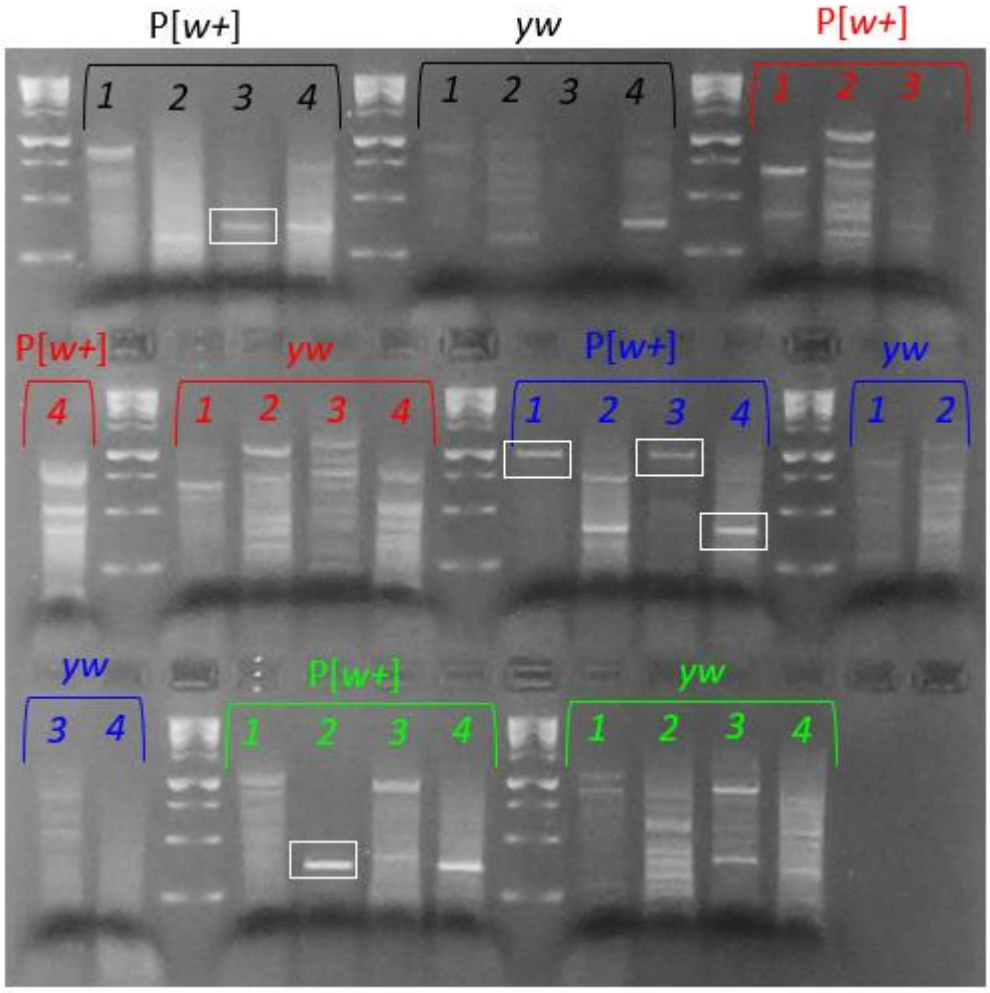

Figure 28: iPCR products from P[*w*+] and *yw* gDNA

Reaction markings were color coded by enzyme: MboI (black). BstUI (red), HpyCH4IV (blue), and HinPI1 (green). Each digest was religated and amplified with iPCR Pair 1, Pair 2, Pair 3, or Pair 4. Bands specific to P[w+] flies are boxed in white: MboI Pair 3, HpyCH4IV Pair 1, 3, 4, HinPI1 Pair 2.

filter, tdTomato within the pellet fluoresced brightly (Figure 29), demonstrating that the tdTomato sequence is correctly assembled by this strategy in CTAF-FR2 in pBSK. This leads to

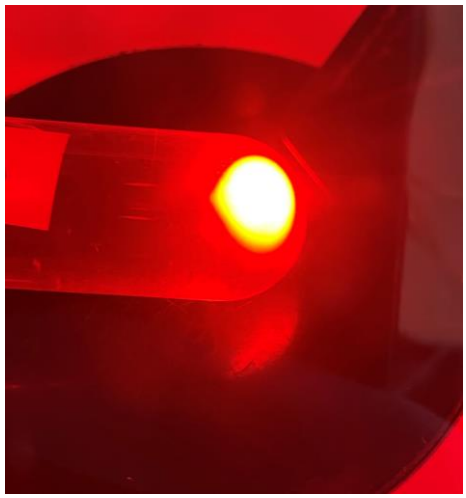

Figure 29*: E. coli* bacterial pellet expressing transformed tdTomato

When observed under green light through a red filter, tdTomato is clearly expressed correctly from the pCTAF-FR2 clone transformation.

the conclusion that the failure (or unexpectedly low success rate) of the gene trap was likely due to a problem with transposition and/or the linker and/or the splice donor/acceptor sequences.

CHAPTER 5

Future Direction

It has been observed that 8 base pair duplications appear on either side of new P-element insertions into the genome that arises from the mechanism of P-element insertion (Figure 30). It was noted that "…the 8-bp target-site duplication appeared to play no role in forward transposition," (Mullins, Rio, and Rubin, 1989). A review of the literature found that several papers

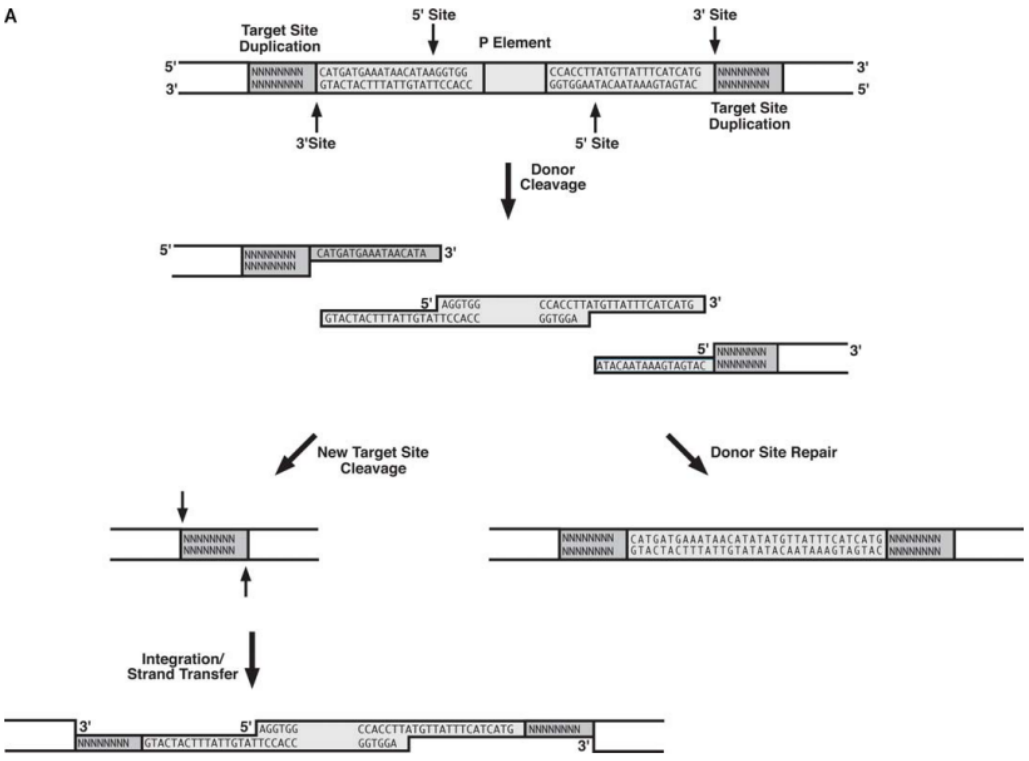

Figure 30: Illustrated Target Site Duplication

As a result of transposon insertion, duplications are produced that flank the insertion site. In the diagram, they are labeled NNNNNN, indicating that this duplication varies and is an irrelevant sequence. Reproduced from Madjumar *et al*., 2015, with permission.

detailing P-element transposition also mention this repeat, and claim it has no observed impact on transposition. However, an extensive analysis of >20,000 P-element insertion sites done by Madjumar et al. (2015) uncovered an apparent consensus in the 8bp duplications (shown in Figure 31). Mullins, Rio, and Rubin observed that the 8 bp target duplication sequences illustrated in Figure 31 had no impact on transposition; however, they did not have to synthesize P-elements from scratch. We hypothesize that duplications flanking the transposon may in fact be required for the initial "jump" out of the pUAST-attB plasmid recombined onto chromosome 3.

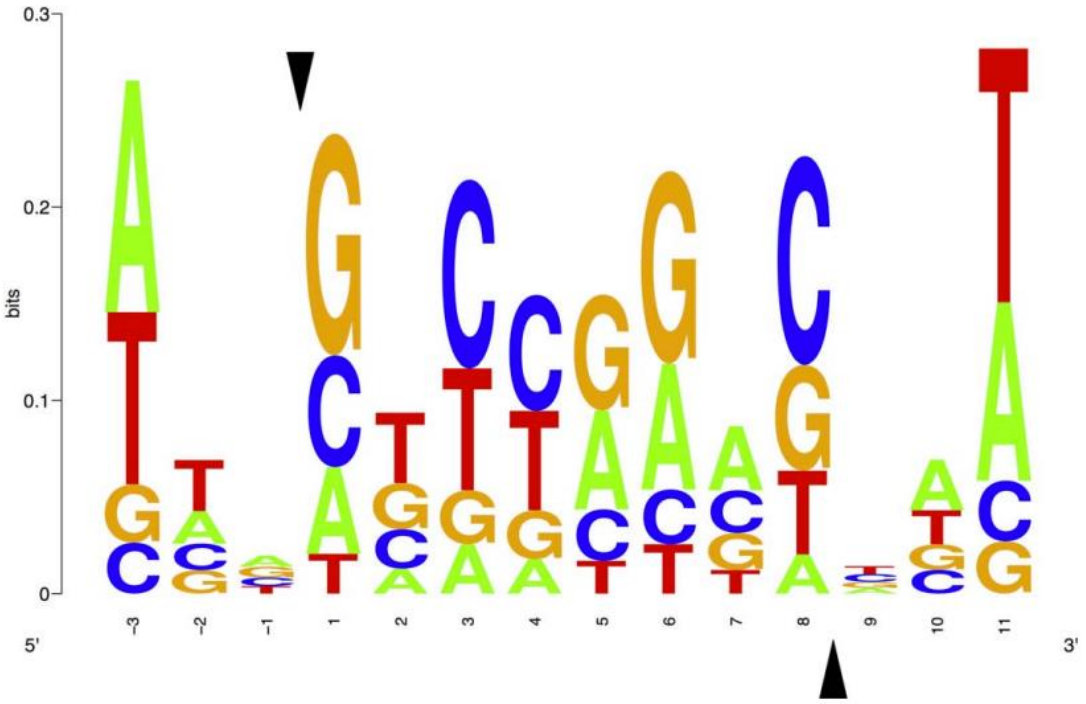

Figure 31: Duplication Consensus Sequence

Black arrows denote the start and end of the duplication consensus sequence (3' and 5' respectively). Reproduced from Madjumar *et al*., 2015, with permission.

Mullins, Rio, and Rubin observed that the 8 bp target duplication sequences illustrated in Figure 30 had no impact on transposition- however, they did not have to synthesize P elements from scratch. When designing transposable constructs, previous scientists have used 5' and 3' P element sequences from the plasmids pCasPeR as well as pCarnegie4 which contain an average of 40 more nucleotides than P element sequence suggested by Mullins, Rio, and Rubin. Mullins, Rio, and Rubin narrowed down the precise nucleotide sequence required for transposition from these P elements; this is what was used in the designed transposon CTAF Frames 1, 2, and 3. As the 8 bp duplications were not included in the P element ends of pCTAF, we hypothesize that duplications flanking the transposon may be required for the initial "jump" out of pUAST-attB-CTAF plasmid recombined onto chromosome 3.

To test this hypothesis, I referenced sequence of the plasmid pCarnegie4 as well as pCasPeR. From the FASTA format DNA sequence, I checked sequence homology 15 bp outside of the 5' and 3' P elements on both pCasPeR and pCarnegie4. The sequences of both were an exact match; from this I gathered that in multiple successful transposition events (conducted by other labs) that 15 bp sequence outside both P ends has been present. As Madjumar *et al*. (2015) listed 11 consensus nucleotides with regard to target site duplications, I added 11 nucleotides from the 15 bp sequences flanking 5' P end and 3' P end in pCarnegie4 and pCasPeR to the 5' P end and 3' P end of pCTAF Frame 1. In front of the 5' P end, the following 11 bp sequence was added: 5'-TGGGTCTGGCC-3'. After the 3' P end, the following 11 bp sequence was added: 5'-GTCTGGCCATT-3'. This addition is illustrated in Figure 32 shown below.

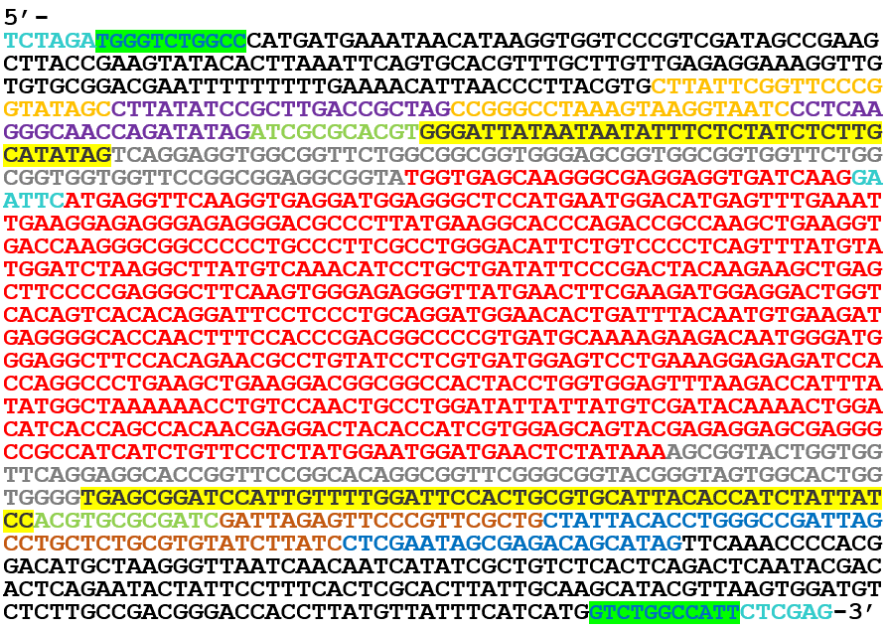

Figure 32: Nucleotide Additions from Referencing pCasPeR and pCarnegie4

11 base pairs were added (green highlighted text) outside of the 5' P end (top black sequence) and 3' P end (bottom black sequence). The rest of pCTAF remains unchanged, and is color coded according to information shown in Tables 1, 2, 3, and 4 (pgs. 25-28).

In addition to the first modified construct (Figure 32), I have designed a second alternative-pCTAF construct containing consensus sequence described in Madjumar *et al*. (Figure 31) outside the 5' and 3' P ends. Instead of green-highlighted nucleotides shown in Figure 32, the 11 nucleotides 5'-ATAGTCCGGAC-3' were added outside the 5' P end and 5'-GTCCGGACTAT-3' was added outside the 3' P end of CTAF-Frame 1 transposon. This addition is illustrated in Figure 33.

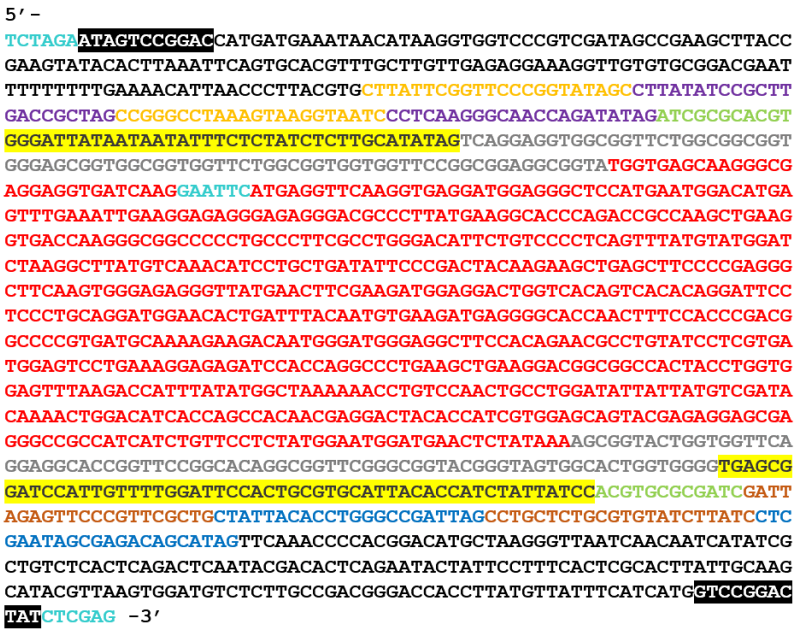

Figure 33: Nucleotide Additions from Referencing Madjumar Consensus

11 base pairs were added (black highlighted/white text) outside of the 5' P end (top black sequence) and 3' P end (bottom black sequence). The rest of pCTAF remains unchanged, and is color coded according to information shown in Tables 1, 2, 3, and 4 (pgs. 25-28).

When the pCarnegie4 and Madjumar- influenced P end additions were aligned, I noticed a significant homology between the sequences. This is to be expected, as the consensus sequence in Madjumar *et al*. (2015) was made through referencing thousands of transposition events. An alignment shown in Figure 34 pertaining to both 5' and 3' P end additions displays the sequence similarity.

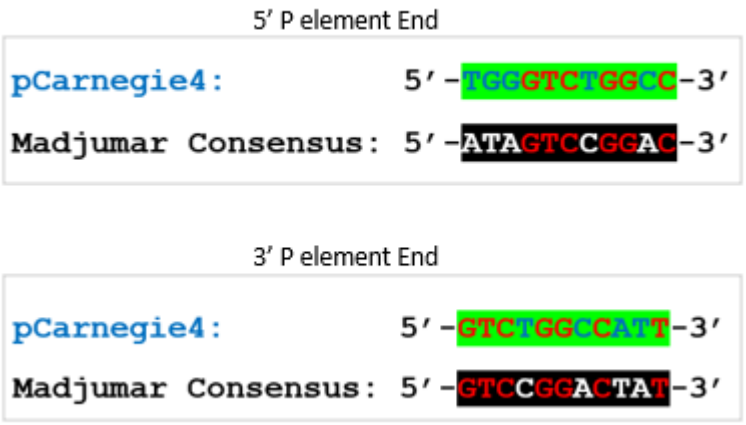

Figure 34: Alignment of 5' and 3' P End Nucleotide Additions

Nucleotide additions taken from pCarnegie4 (green) and the Madjumar Consensus (black) are aligned: the top compares 5' P end additions, the bottom compares 3' P end additions. Homologous nucleotides are shown in red to emphasize similarity between both.

Once both of the modified versions of pCTAF-Frame 1 are ordered from IDT (Integrated DNA Technologies©) cloning will proceed according to procedure listed in Materials and Methods. Depending on the time frame, this will be initiated by myself or Dr. Claire Thomas. Ideally, a new undergraduate will inherit the project in September 2023 and complete cloning/screening for both new constructs. If tdTomato expression is observed and confirmed in at least one Drosophila larvae, iPCR will be used to identify the trapped gene, and a fly stock established. This line will be made available to other institutions, in addition to an open-access curriculum for an undergraduate sophomore biology lab course.

Bibliography

[1] Swaminathan, S. GFP: the green revolution. Nat Cell Biol 11 (Suppl 1), S20 (2009). https://doi.org/10.1038/ncb1953

[2] David Casso, Felipe-Andrés Ramírez-Weber, Thomas B Kornberg. GFP-tagged balancer chromosomes for Drosophila melanogaster. *Mechanisms of Development*, Volume 88, Issue 2, 1999, Pages 229-232,ISSN 0925-4773, https://doi.org/10.1016/S0925-4773(99)00174-4.

[3] Dietrich C, Maiss E. Red fluorescent protein DsRed from Discosoma sp. as a reporter protein in higher plants. Biotechniques. 2002 Feb;32(2):286, 288-90, 292-3. doi: 10.2144/02322st02. PMID: 11848404.

[4] Ghanim GE, Rio DC, Teixeira FK. 2020 Mechanism and regulation of P-element transposition. Open Biol. 10: 200244. https://doi.org/10.1098/rsob.200244

[5] Madjumar S, Rio DC. 2014. P Transposable elements in Drosophila and other eukaryotic organisms. Microbiol Spectrum 3(2):MDNA3-0004-2014. doi:10.1128/microbiolspec.MDNA3- 0004-2014.

[6] Kaufman PD, Dole RF, Rio DC. Drosophila P-element transposase recognizes internal P-element sequences. Cell. 1989 October, https://doi.org/10.1016/0092-8674(89)90297-3

[7] Kanca O, Bellen H, Schnorrer F. 2017 Gene tagging strategies to assess protein expression, localization, and function in Drosophila. Genetics, Vol. 207, 389–412. https://doi.org/10.1534/genetics.117.199968

[8] Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993 Jun;118(2):401-15. doi: 10.1242/dev.118.2.401. PMID: 8223268.

[9] Diao *et al.*, 2015, Cell Reports 10, 1410–1421, March 3, 2015 ª2015 The Authors http://dx.doi.org/10.1016/j.celrep.2015.01.059

[10] Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics. 2013 Aug;194(4):1029-35. doi: 10.1534/genetics.113.152710. Epub 2013 May 24. PMID: 23709638; PMCID: PMC3730909.

[11] Clyne PJ, Brotman JS, Sweeney ST, Davis G. Green fluorescent protein tagging Drosophila proteins at their native genomic loci with small P-elements. Genetics. 2003 Nov;165(3):1433-41. doi: 10.1093/genetics/165.3.1433. Erratum in: Genetics. 2004 Aug;167(4):2143. PMID: 14668392; PMCID: PMC1462835.

[12] Rubin GM, Spradling AC. Genetic transformation of Drosophila with transposable element vectors. Science. 1982 Oct 22;218(4570):348-53. doi: 10.1126/science.6289436. PMID: 6289436.

[13] Venken KJ, Schulze KL, Haelterman NA, Pan H, He Y, Evans-Holm M, Carlson JW, Levis RW, Spradling AC, Hoskins RA, Bellen HJ. MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat Methods. 2011 Sep;8(9):737-43. doi: 10.1038/nmeth.1662. PMID: 21985007; PMCID: PMC3191940.

[14] Diao F, Ironfield H, Luan H, Diao F, Shropshire WC, Ewer J, Marr E, Potter CJ, Landgraf M, White BH. Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes. Cell Rep. 2015 Mar 3;10(8):1410-21. doi: 10.1016/j.celrep.2015.01.059. Epub 2015 Feb 26. PMID: 25732830; PMCID: PMC4373654.

[15] Metaxakis A, Oehler S, Klinakis A, Savakis C. Minos as a genetic and genomic tool in Drosophila melanogaster. Genetics. 2005 Oct;171(2):571-81. doi: 10.1534/genetics.105.041848. Epub 2005 Jun 21. PMID: 15972463; PMCID: PMC1456772.

[16] Groth AC, Fish M, Nusse R, Calos MP. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics. 2004 Apr;166(4):1775-82. doi: 10.1534/genetics.166.4.1775. PMID: 15126397; PMCID: PMC1470814.

[17] Mullins MC, Rio DC, Rubin GM. cis-acting DNA sequence requirements for P-element transposition. Genes Dev. 1989 May;3(5):729-38. doi: 10.1101/gad.3.5.729. PMID: 2545527.

[18] Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013 Oct;65(10):1357-69. doi: 10.1016/j.addr.2012.09.039. Epub 2012 Sep 29. PMID: 23026637; PMCID: PMC3726540.

Academic Vitae

Anna V. Foltz

avf5465@psu.edu

Education

B.S. with Honors in Biochemistry and Molecular Biology (2019-2023)

Molecular Biology Option.

The Pennsylvania State University, Schreyer Honors College

Practical Skills

Undergraduate Researcher at the Pennsylvania State University, Department of Biology (2020-2023, including summers)

Honors Thesis: "Fluorescent protein trapping in *Drosophila melanogaster*" *In Progress* PI: Dr. Claire Thomas

- Designed transposable DNA sequence.
- Cloned fluorescent *Drosophila* genes into multiple plasmids, often triple ligations.
- Transformed foreign DNA into *Escherichia coli* to generate large quantities plasmid.
- Coordinated fly mating crosses to achieve phenotypes of interest.
- Obtained significant experience in traditional and fluorescent light microscopy.
- Navigated online genetic databases (NCBI BLAST, Flybase) to select sequences cloning.

Relevant Experience

Teaching Assistant, BIOL 230M: *Honors Biology, Molecules and Cells* (2022)

- Incorporated my honors thesis work to teach elementary genetics to undergraduate students.
- Introduced students to DNA purification, restriction enzymology, PCR, iPCR, larvae screening, fly husbandry.

Undergraduate Research Assistant, the Pennsylvania State University Department of Environmental Engineering (2019-2020)

Penn State Eco-Machine™: Carbon and Energy Balance, Dr. Rachel Brennan

- Cared for waste-water filtration organisms (canna lily, taro plant, duckweed)
- Constructed/maintained a vertical duckweed farm, with intention to measure nitrogen fixation levels in *Lemnoideae* (duckweed)

Awards/Certifications

NASA Pennsylvania Space Grant Consortium WISER Grant (2020 (2x), 2021)

Rodney A. Erickson Discovery Grant (2021)

Schreyer Honors College Academic Excellence Scholarship, (2019-2023)

64th Annual Drosophila Research Conference, Poster Presenter (2023)

BioGSA Biology Student Research Showcase, 2nd Place Poster Presentation (2022)

Dean's List, Eberly College of Science Spring 2022 and Fall 2023

CITI Yearly Biosafety Training- Recombinant DNA User (2020-Present)

Relevant Coursework

Molecular Cloning, Immunology, Cell Growth and Differentiation, Molecular Biology, Microbiology, General Biochemistry, Genomics, Organic Chemistry, Statistics