

**Genetic analysis and the evolutionary study of a pseudogene, *CR14033*, and the
corresponding parent gene, *CG9203***

by
G. Elizabeth Sperry

B.Sc. (Hons), Mount Allison University, 2009

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the Graduate Academic Unit of Biology

Supervisor: Denise Clark, Ph.D., Biology

Examining Board: Cheryl Patten, Ph.D., Biology, Chair
Adam Dyker, Ph.D., Chemistry

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

April, 2016

© G. Elizabeth Sperry, 2016

ABSTRACT

Gene duplication can lead to new gene function. If a duplicated gene holds no benefit, it will acquire mutations leading to a non-functional copy or pseudogene. The purpose of this study was to see if a transcribed gene duplicate in *Drosophila melanogaster*, *CR14033*, considered a pseudogene, is involved in regulating its parent gene (*CG9203*), to determine the function of the genes of interest, and to study the origin of *CR14033*. The regulation likely occurs through the endogenous small interfering RNA pathway. Mutations in *CR14033* and *CG9203* and transgenic flies expressing sense and antisense transcripts of each gene were generated in this study to test this hypothesis. From the analysis of the *CR14033* deletion and the overexpression of *CR14033* and *CG9203*, it appears as though there is an interaction between the two. To study function, sterility was also explored in these genotypes. A female sterile mutant phenotype was generated from *CG9203*. Sequence analysis of twelve *Drosophila* species showed that *CR14033* likely arose through a gene fusion event that occurred in the melanogaster subgroup ancestor.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor, Dr. Denise Clark, for all her guidance and understanding. I would also like to thank my mother and grandmother for all their support. I would also like to thank my lab mates Ryan O’Neill, Amy Botta, Eric Merzetti, Joseph Lazarev, and Ashley Dipasquale for all the laughs and guidance. I would also like to thank my committee members Dr. Shawn MacLellan and Dr. Dion Durnford for all their advice.

This research was supported by a NSERC Discovery Grant to Denise Clark.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
Table of Contents	iv
List of Tables	v
List of Figures	viii
List of Abbreviations	ix
Chapter 1: Introduction	1
Chapter 2: Materials and Methods	22
Chapter 3: Results	58
Chapter 4: Discussion	147
Bibliography	139
Appendix	168
Curriculum Vitae	

List of Tables

Table 2.1 Stocks used in this study.....	36
Table 2.2 Primers used for this study	39
Table 3.1 Summary of sequence characteristics for P-element excision strains generated from excising $P\{SUPor-P\}tkv^{KG01923}$, $P\{lacW\}l(2)k01302^{K01302}$, $P\{GSV6\}GS17311$	89
Table 3.2 Summary of sequence characteristics for the P-element excision strain generated excisions $P\{SUPor-P\}CG9203^{KG05829}$	91
Table 3.3 Summary of the complementation tests performed on the 2B-A and 8A-A partial excisions of $P\{lacW\}l(2)k01302^{k01302}$ and the original P-element insertion strain, 14403, using tkv^7 , $Df(2L)BSC693$, and $Df(2L)Exel6011$	93
Table 3.4 Summary of the complementation tests performed on the excisions of $P\{GSV6\}GS17311$ and the original P-element insertion strain, 200-010, using, tkv^1 , tkv^7 , $Df(2L)BSC693$, and $Df(2L)Exel6011$	94
Table 3.5 Sterility of male and female $CRI4033$ mutants $CRI4033^{18}$ and $P\{GSV6\}GS17311$; and $CG9203$ mutants, 24A-D and $P\{SUPor-P\}CG9203^{KG05829}$ as homozygotes	97
Table 3.6 Comparison between the region that produces siRNA from $CRI4033$ in orthologous parent genes	98
Table 3.7 Comparison between the region that produces siRNA in orthologous and paralogous parent genes and pseudogenes.....	100

Table 3.8. Comparison between the region that produces siRNA in orthologous pseudogenes	103
Table 3.9 Summary of phenotypic characterization of RNAi	105
Table 3.10 Male sterility in RNAi screen.....	106
Table 3.11 Over expressing <i>CG9203</i> and <i>CRI4033</i> did not produce a mutant phenotype	108
Table 3.12 The additional transcript of <i>CG9203</i> and <i>CRI4033</i> provided by expressing a UAS cDNA transgene did not cause sterility in males or females	111
Table 3.13 The lethal excision, 28B-D, was rescued by the duplication showing that the lethality was due to the P-element excision	113
Table 3.14. <i>P{lacW}l(2)k01302^{k01302}</i> complemented with the deficiency of <i>tkv</i> strains but not with <i>tkv⁷</i>	115

List of Figures

Figure 1.1 Gene duplication by unequal crossing over.....	12
Figure 1.2 Gene duplication by retrotransposition	14
Figure 1.3 Endo-siRNA pathway.....	16
Figure 1.4 <i>CG9203</i> and <i>CRI4033</i>	18
Figure 1.5 Small RNAs associated with <i>CRI4033</i> in larval brains and imaginal discs ...	20
Figure 2.1: Crossing scheme for P-element mutagenesis of the pseudogene, <i>CRI4033</i> ..	42
Figure 2.2: Crossing scheme for P-element mutagenesis of <i>CG9203</i> using the P-elements with a <i>white</i> and <i>yellow</i> marker gene	44
Figure 2.3: Balancing lethal <i>CRI4033</i> and <i>CG9203</i> P-element excision alleles with GFP- marked balancer chromosomes.....	46
Figure 2.4: Complementation test using a <i>tkv</i> allele crossing scheme.....	48
Figure 2.5: A) Complementation tests between deficiencies of <i>CRI4033</i> and lethal P- element excisions of <i>CRI4033</i> B) Complementation test using a duplication of <i>CG9203</i> regions and a lethal P-element excisions of <i>CG9203</i>	50
Figure 2.6: <i>CRI4033</i> cDNA (LD26673) inserted in the fly transformation plasmid pUASP	52
Figure 2.7: <i>CG9203</i> cDNA (MIP11944) inserted in the fly transformation plasmid pUASP	54
Figure 2.8: Induction of RNAi for <i>CRI4033</i> and <i>CG9203</i>	56
Figure 3.1 Location of the P-elements inserted in <i>CRI4033</i> that were used for P-element mutagenesis.....	116
Figure 3.2 Sequence alignment of 2B-A and 8A-A.....	118

Figure 3.3 Schematic for <i>CR14033</i> deletion mutant, <i>CR14033</i> ¹⁸	121
Figure 3.4 Location of the P-elements inserted in <i>CG9203</i> that were used for P-element mutagenesis.....	123
Figure 3.5 Sequence of 24A-D and schematic showing the 1 kb deletion resulting from the imprecise excision.....	125
Figure 3.6 Pairwise sequence alignment between CR14033 and CG9203 cDNAs MIP11944 and LD26673	127
Figure 3.7 Alignment of the sequence from the different reading frames of <i>CR14033</i> that has sequence similarity <i>CG9203</i>	129
Figure 3.8 Alignment of the sequence from the different reading frames of the <i>D. sechellia</i> <i>CR14033</i> and <i>CG9203</i> orthogs	131
Figure 3.9 Alignment of the sequence from the different reading frames of the <i>D. simulans</i> <i>CR14033</i> and <i>CG9203</i> orthogs	133
Figure 3.10 <i>CR14033</i> is conserved in the melanogaster subgroup	135
Figure 3.11 Expression of <i>CG9203</i> , <i>CR14033</i> , and <i>tkv</i> in <i>Dcr2</i> or <i>loqs</i> mutant testes...	137
Figure 3.12 Expression of <i>CG9203</i> , <i>CR14033</i> , and <i>tkv</i> in brains and discs from <i>Dcr2</i> mutants.....	139
Figure 3.13 Transcript changes in <i>CG9203</i> , <i>CR14033</i> , and <i>tkv</i> in testes flies expressing sense and antisense cDNA transgenes of <i>CR14033</i>	141
Figure 3.14 Transcript changes in <i>CG9203</i> , <i>CR14033</i> , and <i>tkv</i> in testes flies expressing sense and antisense cDNA transgenes of <i>CG9203</i>	143
Figure 3.15 Expression of <i>CG9203</i> and <i>tkv</i> in a <i>CR14033</i> deletion mutant.....	145

List of Abbreviations

BLAST	Basic Local Alignment Search Tool
bp	base pair
cDNA	complementary DNA
CIP	calf intestinal alkaline phosphatase
<i>CyO</i>	curly balancer chromosome
<i>D. ere</i>	<i>Drosophila erecta</i>
<i>D. mel</i>	<i>Drosophila melanogaster</i>
<i>D. sec</i>	<i>Drosophila sechellia</i>
<i>D. sim V2</i>	<i>Drosophila simulans</i> version 2
<i>D. sim</i>	<i>Drosophila simulans</i>
<i>D. yak</i>	<i>Drosophila yakuba</i>
Dcr2	Dicer 2
DEPCH ₂ O	Diethyl pyrocarbonate-treated H ₂ O
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphates
dsRNAs	double-stranded RNAs
endo-siRNA	endogenous small interfering RNA
ETBr	Ethidium Bromide
GFP	green florescent protein
miRNA	microRNA
nt	nucleotide
<i>P{*}</i>	P-element excision
pBSK	pBlueScript SK
PCR	polymerase chain reaction
PGAM3	phosphoglycerate mutase family 3
PTEN	phosphatase and tensin homolog
PTENPI	phosphatase and tensin homolog pseudogene 1
qPCR	quantitative reverse transcription polymerase chain reaction
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
siRNA	small interfering RNA
TLS	translesion synthesis
<i>tkv</i>	thickveins
UAS	upstream activation sequence
UTR	untranslated region

Chapter 1: Introduction

1.1 Gene duplication and its contribution to evolution

Genome evolution occurs in both prokaryotes and eukaryotes. Prokaryotes can acquire new genes and new gene function via gene duplication and horizontal gene transfer. With horizontal gene transfer, they can acquire genes that are new, genes that are similar to those already in their genome, or orthologs from another lineage that displaces the genes (Boto 2010, Koonin *et al.* 2001). The size of the prokaryotic genome stays roughly around the same size, meaning that the gain and loss of genes is balanced (Kunin and Ouzounis 2003).

Gene duplication is also a major driving force in eukaryotic evolution as it provides additional genetic information for mutation, drift, and selection to act upon (Ohno 1970). Gene duplication can occur through whole genome duplication or by changes in copy number. Genome duplication has occurred in many evolutionary lineages, sometimes happening multiple times, as seen in plants for example (Soltis and Soltis 1999). In *Drosophila melanogaster*, 5536 duplicated genes have been uncovered among its 13 601 genes (Rubin *et al.* 2000).

In addition to genome duplication, two mechanisms, unequal crossing over and retrotransposition, are known to produce duplicate genes. Unequal crossing over is a DNA-based mechanism that is a result of pairing between chromosomal segments at different loci (Fincham 2001). This mechanism generates gene duplicates that remain on the same chromosome in tandem, like the *Bar* gene (Bridges 1936) (Figure 1.1).

Depending on the chromosomal position when crossing over occurs, the duplicated region can be of the entire gene, several genes, or only part of the gene (Zhang 2003).

Gene duplication by retrotransposition is an RNA-based mechanism that occurs when mRNA is reverse transcribed into complementary DNA (cDNA), which is then integrated back into the genome (Figure 1.2). This method can result in gene duplicates that are located on a different chromosome than their parent gene (Vanin 1985), have a loss of introns and regulatory sequences, and have flanking short direct repeats, and, if recently duplicated, still have visible polyA tracts (Long 2001).

Newly duplicated genes tend to have relaxed functional constraints and as such may acquire mutations (Lynch and Conery 2000). As a result, there are three possible fates of a newly duplicated gene. First, the duplicate can gain a new function (neofunctionalization); second, the duplicate and its paralog can split the function provided previously by just the paralog (subfunctionalization); and, third, the duplicate can acquire mutations that result in no function as the duplicate does not provide a selective advantage (pseudogenization) (D'Errico *et al.* 2004 and Yano *et al.* 2004).

A gene duplicate can diverge from its paralog in its gene expression regulatory regions, and/or coding sequence, both delivering a different effect (Xu *et al.* 2011). If there is a change in the regulatory region, this can change the expression pattern of the duplicate while a change in coding sequence can result in the duplicate protein gaining a new function (Force *et al.* 1999 and Xu *et al.* 2011). However, the most common fate of a duplicated gene is pseudogenization (D'Errico *et al.* 2004 and Yano *et al.* 2004).

1.2 Pseudogenes

Pseudogenes are found in a wide range of prokaryotic and eukaryotic species. They resemble a functional gene in the same genome or an orthologous gene in the genome of another species, yet they have aberrations such as truncations, premature stop codons, insertions, and deletions that make them appear non-functional as these disrupt the coding region (Proudfoot 1980). Depending on the mechanism of duplication, pseudogenes are categorized into two types, processed and non-processed.

Processed pseudogenes are generated by retrotransposition. As aforementioned, these sequences are generated from mRNA and therefore they lack introns and a 5' regulatory region. If recently duplicated, they may still have remnants of 3' polyA tail, and flanking direct repeats, which are caused by insertion of the new gene at the target site where a staggered cut has been made. The overhang is then filled in by DNA polymerase creating direct flanking repeats on either end of the pseudogene. Due to the lack of a 5' regulatory region, unless these sequences insert in a region where the promoter from another gene can be used for transcription, the pseudogene will not be expressed (D'Errico *et al.* 2004, Yano *et al.* 2004, and Zheng *et al.* 2007). This is not the case for non-processed pseudogenes as these duplicates can maintain the original intron-exon structure as well as regulatory sequence (Darnell *et al.* 1990).

The number of pseudogenes detected in the *Drosophila* genome varies depending on how conservative the detection method is. One study reported 110 pseudogenes in the *Drosophila* genome, which is approximately 1 pseudogene for every 130 protein coding genes, by searching the genome for protein homology in the genomic DNA with disruptions such as premature stop codons and frameshift mutations in the sequence. Of

the 110 pseudogenes, 34 were processed pseudogenes randomly distributed throughout the genome (Harrison *et al.* 2003). Another study found 176 pseudogenes by using a less conservative method (Zdobnov *et al.* 2002). A more recent survey of the genome showed that there are 328 genes with “pseudogene attributes” in *D. melanogaster*. Thirty are annotated as rRNA or tRNA pseudogenes and therefore 298 pseudogenes correspond to protein coding genes. There are 15028 protein-coding genes that make up the *Drosophila* genome (dos Santos *et al.* 2014). This makes 1 pseudogene for every 50 protein coding genes.

Compared to other eukaryotes, the ratio of protein coding genes to pseudogenes in *Drosophila melanogaster* is much higher. For instance, there are 11, 580 pseudogenes in the human genome, an average of 1 for every 3 protein coding genes. Of these pseudogenes, 8,298 are processed (Harrow *et al.* 2012). The difference in the number of pseudogenes present in *Drosophila* compared to humans could be due to a high rate of DNA deletion from the *Drosophila* genome in unconstrained regions or selection for maintaining a small genome (Petrov *et al.* 1996).

1.2.1 Roles of pseudogenes

Due to all the aberrations in pseudogene sequences, it was previously believed that they were non-functional and as such considered junk (Sasidharan and Gerstein 2008). However, more recent studies challenge this view. From those pseudogenes that are transcribed, as evident by their appearance in the sequence data from next-generation sequencing, pseudogene functions are emerging (Li *et al.* 2013). These functions include

microRNA (miRNA) decoys, generating short polypeptides, and gene regulation through the endogenous small interfering (endo-siRNA) pathway.

Pseudogenes can function as a miRNA decoy. miRNA is a form of non-coding RNA that carries sequence complementarity to the 3' UTR of the target gene transcript. These miRNA genes are found throughout the genome and are transcribed like other genes into precursors that are processed into mature miRNA. The mature miRNA can then be incorporated into an RNA-silencing complex, which anneals to the target gene and decreases the amount of protein made for that target gene through post-transcriptional silencing (Chen 2010). The decrease in protein is a result of the protein-RNA complex disrupting translation of the mRNA or destabilizing the mRNA causing degradation (Bartel 2009, Baek *et al.* 2009, Tay *et al.* 2008, Lal *et al.* 2009). miRNA decoys, in this case, are transcribed pseudogenes that have seed sequences in the 3' UTR that are similar to that of the target gene. As a result, these sequences can compete for the miRNA-silencing complex effectively allowing more of the target gene transcripts to be translated into protein as fewer will be rendered disabled (Gu *et al.* 2009). Two examples of pseudogenes that function in this way are *PTENP1* and *Pbcas4*.

PTENP1 is a processed pseudogene for a human tumor suppressor gene, *PTEN*, with only 18 mismatched nucleotides between them. The pseudogene and parent gene carry a seed match for the same miRNAs in the 3' UTR region, therefore *PTENP1* can act as a decoy for *PTEN*. For instance, overexpression of *PTENP1* results in an increase in the *PTEN* transcript and protein (Poliseno *et al.* 2010). *Pbcas4* is another pseudogene involved in cancer, which acts as a miRNA decoy for breast carcinoma amplified sequence 4 (*BCAS4*) in neuroblastoma cells (Marques *et al.* 2012).

Some pseudogenes are able to produce polypeptides. The first one characterized was phosphoglycerate mutase family 3 (*PGAM3*) in primates (Betran *et al.* 2002). Even though the pseudogene possesses all the characteristics of a processed pseudogene and a number of mutations, it has a full-length open reading frame (762 bp) like that of the parent gene open reading frame. *PGAM3* is inserted in the open reading frame of another gene, the Menkes disease gene (Dierick *et al.* 1997). Betran *et al.* (2002) showed that there is an upstream promoter region for *PGAM3* and so it is to produce a functional protein. Another annotated pseudogene, *ΨCx43*, was also found to produce a protein in humans. The *Cx43* gene has an intron, which is not present in *ΨCx43* but *ΨCx43* still has the entire open reading frame intact along with regulatory elements (Willecke *et al.* 1990; Fishman *et al.* 1991). Because this “pseudogene” is able to produce protein, it should be reconsidered as to whether or not is truly a pseudogene.

Pseudogenes can also function through another small interfering pathway, the endo siRNA pathway (Figure 1.3). The endo-siRNA pathway is involved in transposon repression in somatic cells and the germline as well as gene regulation, using RNA interference (RNAi). Endo-siRNAs are generated from double-stranded RNAs (dsRNAs) that arise from convergent transcription or hairpin RNA formation (Czech *et al.* 2008). This pathway uses Dicer-2-mediated cleavage to generate the endo-siRNA from the dsRNA. Dicer-2 (*Dcr2*) is a ribonuclease III with two RNase III domains, a helicase domain, and a PAZ domain. The PAZ domain binds to the 3' end of the RNA positioning the RNase III domains for cleavage of the dsRNA (MacRae *et al.* 2007 and MacRae *et al.* 2006). The helicase domain of Dicer-2 is responsible for translocating Dicer-2 for processive generation of siRNA duplexes from a single substrate helix using ATP (Cenik

et al. 2011). It binds the RNA with help from loquacious, or loqs, a dsRNA-binding protein (Zhang *et al.* 2004; Liu *et al.* 2003; Zhou *et al.* 2009). The isoform used is loqs-PD (Zhou *et al.* 2009 and Hartig *et al.* 2009). Another dsRNA-binding protein in this pathway is R2D2, which is responsible for loading siRNA duplexes into RISC. It has also been shown to bind the more stable end of the duplex while Dcr2 binds the less stable end providing an orientation for loading the sequence into a RNA-induced silencing complex (RISC) (Tomari *et al.* 2004). The dsRNA is cut into 21 nucleotide pieces that are incorporated into and unwound in a RISC containing the slicer, Argonaute 2 (Ago2). This complex then finds complementary mRNA and cleaves it, preventing its translation (Hutvanger *et al.* 2008 and Meister *et al.* 2004).

The endo-siRNA pathway was discovered in several labs working on a variety of model organisms such as *Caenorhabditis elegans* (Han *et al.* 2009), mice (Tam *et al.* 2008 and Watanabe *et al.* 2008), and *Drosophila melanogaster* (Czech *et al.* 2008, Kawamura *et al.* 2008, Okamura *et al.* 2008a and b). The two important studies for my work are those of Czech *et al.* 2008 and Tam *et al.* 2008 because they demonstrate that pseudogenes can potentially have a function in gene regulation through the endo-siRNA pathway. For instance, Tam *et al.* (2008) demonstrated in mouse oocytes that some pseudogenes are transcribed and that their transcripts anneal to protein-coding mRNA generating double stranded RNA (dsRNA) that is made into endo-siRNA. They showed that there was an increase in the expression of the parent genes in *Dicer* mutants, indirectly indicating a regulatory function for the pseudogenes on the parent genes using this pathway.

In *Drosophila melanogaster*, endo-siRNA corresponding to pseudogenes has also been found. Czech *et al.* (2008) found an abundance of testis-specific endo-siRNA that arose from a gene, *CG14033*, now designated as pseudogene *CRI4033* (Figure 1.4). Marques *et al.* (2010) also found abundant endo-siRNA in larval brains and discs corresponding to *CRI4033* sequence (Figure 1.5). These reads that mapped to *CRI4033*, mapped here unambiguously even though there is similarity to *CG9203*. A protein-coding gene, *CG9203* (Figure 1.4), was found to have sequence similarity to *CRI4033* and, in *Dicer2* (*Dcr2*) and *Ago2* mutant testes, there was an increase in *CG9203* intact transcript (Czech *et al.* 2008). These results suggest that the pseudogene may have a regulatory function for expression of the parent gene through the endo-siRNA pathway (Czech *et al.* 2008).

1.3 Genes and pseudogene of interest

This study is concerned with two genes, *CG9203* and *tkv*, and one pseudogene, *CRI4033*. CG is used to denote protein-coding genes and CR is used for non-protein-coding genes. *CG9203*, a protein-coding gene, is located on the X chromosome and produces an 81.5 kDa protein with a zinc-finger and SprT-like domain, linking it to involvement in DNA repair (Tweedie *et al.* 2009). Its human ortholog, Spartan/C1orf124, is a nuclear protein involved in regulating post-replication repair by preventing mutations associated with replicating damaged DNA. It is recruited to sites of DNA damage that have resulted in stalled replication by a ubiquitinated proliferating cell nuclear antigen (Machida *et al.* 2012 and Ghossal *et al.* 2012).

CRI4033, a pseudogene, is located on chromosome 2. It overlaps with another gene, *tkv*, but is transcribed on the opposite strand. More precisely, *CRI4033* lies in the intron of transcript *tkv-RC* and it has a partial overlap with one of its exons at its 3' end (Figure 1.4). *CRI4033* appears to be a processed pseudogene of *CG9203* as it possesses high nucleotide sequence identity to *CG9203* and it is on a different chromosome (Okamura *et al.* 2008; Marques *et al.* 2008; Czech *et al.* 2008). The longest open reading frame for *CRI4033* encodes 74 amino acids (Okamura *et al.* 2008).

The second gene that is involved in this study is *tkv*. *Tkv* encodes a bone morphogenetic type 1 receptor (Moustakas and Heldin, 2009). It is involved in dorsoventral patterning of an early embryo as well as dorsal epidermis formation, wing development, dorsal medial integument development, oogenesis, and in leg development (Terracol and Lengyel 1994).

1.4 Regulation of *CG9203* by its pseudogene *CRI4033*

The purpose of this study is to see if a specific pseudogene, that bears sequence similarity to a protein-coding gene, is involved in regulating the expression of its parent gene in *Drosophila melanogaster*. *Drosophila melanogaster* is an excellent organism for this study because its genome is fully sequenced and, due to its history as a model organism, there are many techniques and tools available for extensive genetic analysis. It also requires little space to amass large numbers, it has a short generation time, and it is easy to culture. Also, numerous *Drosophila* species genomes that have been fully sequenced, 12 of which are used in this study, allow for cross-species comparison.

Few previous studies have undertaken the task to look at the effects of a specific pseudogene on its parent gene to determine, in a direct way, i.e. knock out mutations, if it does indeed regulate parent gene expression. This research seeks to do just that. The role that *CRI4033* plays in regulating the expression of what appears to be its parent gene, *CG9203*, is unknown.

There were four objectives in this study. The first was to determine the function of *CG9203* and *CRI4033* in *D. melanogaster*. The second was to determine if there was an interaction between the two and the third was to see if *CRI4033* is regulated by the endo-siRNA pathway. Finally, this study looked to determine if *CRI4033* is involved in regulating *tkv* as there is sequence overlap between *CRI4033* and one *tkv* exon. *CRI4033* and *tkv* are transcribed on opposite strands.

To determine the function of *CG9203* and *CRI4033* in *D. melanogaster*, deletion mutations were generated using P-element transposon mutagenesis. In addition, transgenic flies expressing a complementary DNA (cDNA) copy of *CG9203* or *CRI4033* were generated and their expression was induced to examine the effects of overexpressing these genes. Phenotypic characterization of these flies was also performed.

To determine if there is an interaction between *CG9203* and *CRI4033*, the relative transcript levels of the deletion mutants and the induced expression of transgenic flies were examined using quantitative reverse transcriptase polymerase chain reaction (qPCR).

To determine if *CRI4033* is regulated through the endo-siRNA pathway like *CG9203* (Czech *et al.* 2008), intact transcript levels of *CRI4033* were measured in *Dcr2*

and *loqs* mutant testes and larval brains and associated imaginal discs using qPCR.

These two mutants were used in this screen because both *Dcr2* and *loqs*, isoform-PD, have been found to be involved in the endo-siRNA pathway (Hartig *et al.* 2009, Miyoshi *et al.* 2010, Zhou *et al.* 2009).

The fourth objective was to determine if the level of *CRI4033* influenced *tkv* expression, since they are overlapping genes. This issue was addressed by seeing if there was a change in the intact transcript level of *tkv* in *CRI4033* mutants and in flies expressing an additional cDNA copy of *CRI4033*.

Figure 1.1 Gene duplication by unequal crossing over

Crossing over is a process that occurs during meiosis in prophase 1. It results in an exchange of genetic information between a homologous pair of chromosomes. Sometimes different segments of the chromosome align and when the crossing over takes place, instead of getting reciprocal exchange of chromosomal information, one chromosome gains sequence and the other one loses sequence (Fincham 2001). This figure shows an example of gene duplication resulting from unequal crossing over between two genes of similar sequence. One gene is represented by dark grey boxes while a gene similar to it is represented by the lighter grey boxes. The crossing over event is shown as an X. The result of this unequal crossing over is a chromosome with a third gene with sequences from the two original genes a chromosome that no longer carries the two original genes. Figure adapted from Zhang (2003).

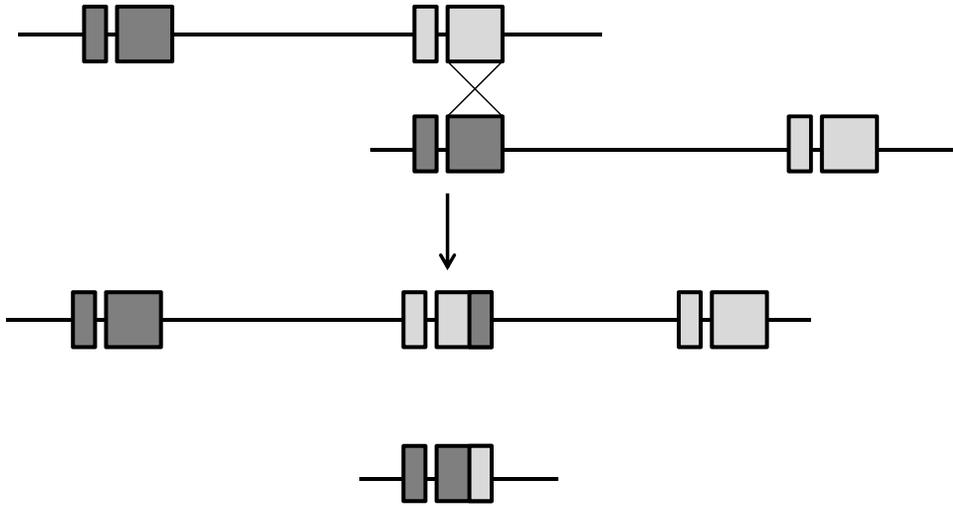


Figure 1.2 Gene duplication by retrotransposition

In this figure a gene is transcribed and processed into mRNA, which is then reverse transcribed into cDNA. The cDNA is then integrated back into the genome, at a site that can be on a different chromosome than the parent gene. Figure adapted from Zhang (2003).

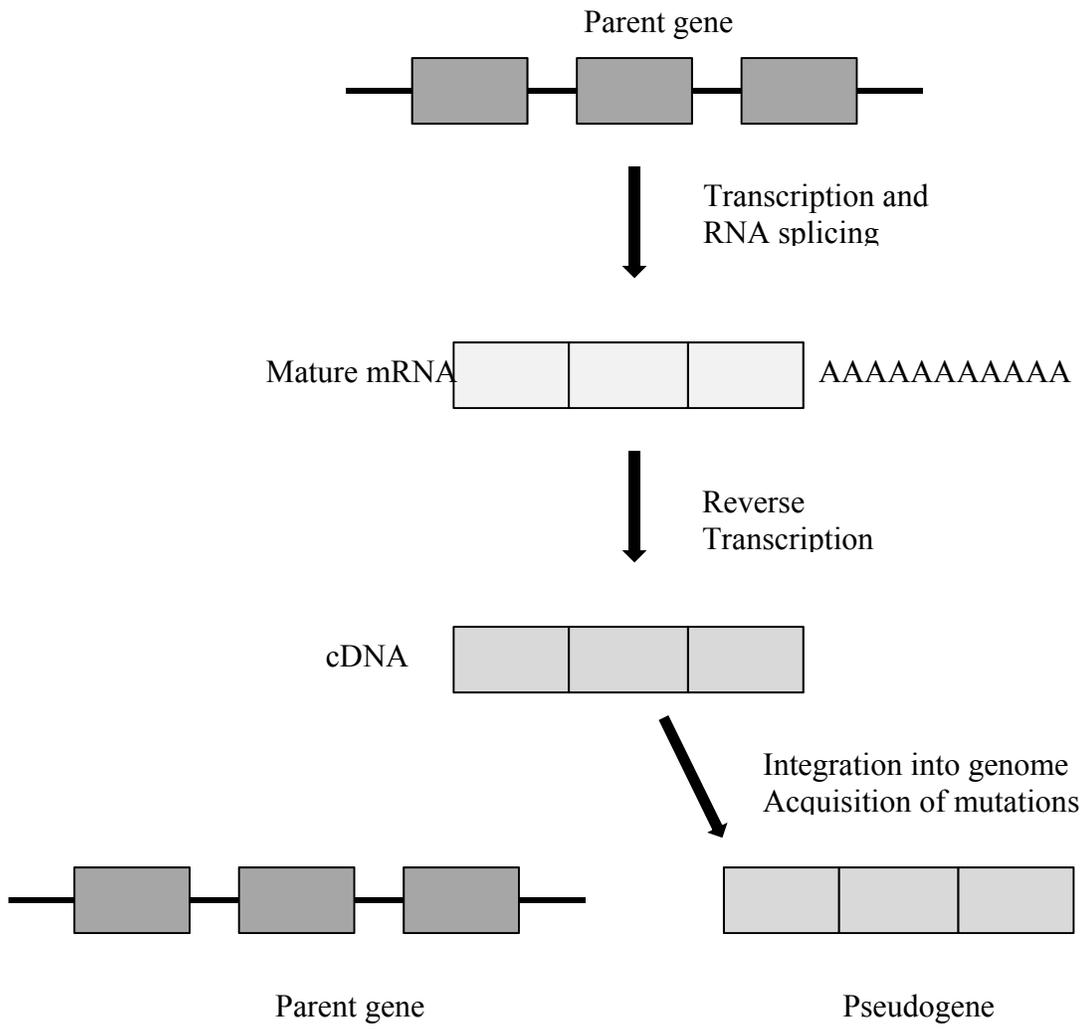


Figure 1.3 Endo-siRNA pathway

Endo-siRNA can be generated from hairpin RNA formation or through convergent transcription. Here it is shown through convergent transcription. The portions of the genes that are complementary to each other anneal, resulting in dsRNA. This triggers a ribonuclease III called Dcr2 to bind to the 3' end of the dsRNA with the help of its binding partners in this pathway, loqs or R2D2 (Zhang *et al.* 2004, Czech *et al.* 2008, and Liu *et al.* 2003). The dsRNA is cut into 21 nucleotide pieces (siRNA) that are unwound and incorporated into RISC containing the slicer, Ago2, which then finds complementary mRNA and cuts it, preventing its translation (Hutvanger *et al.* 2008 and Meister *et al.* 2004).

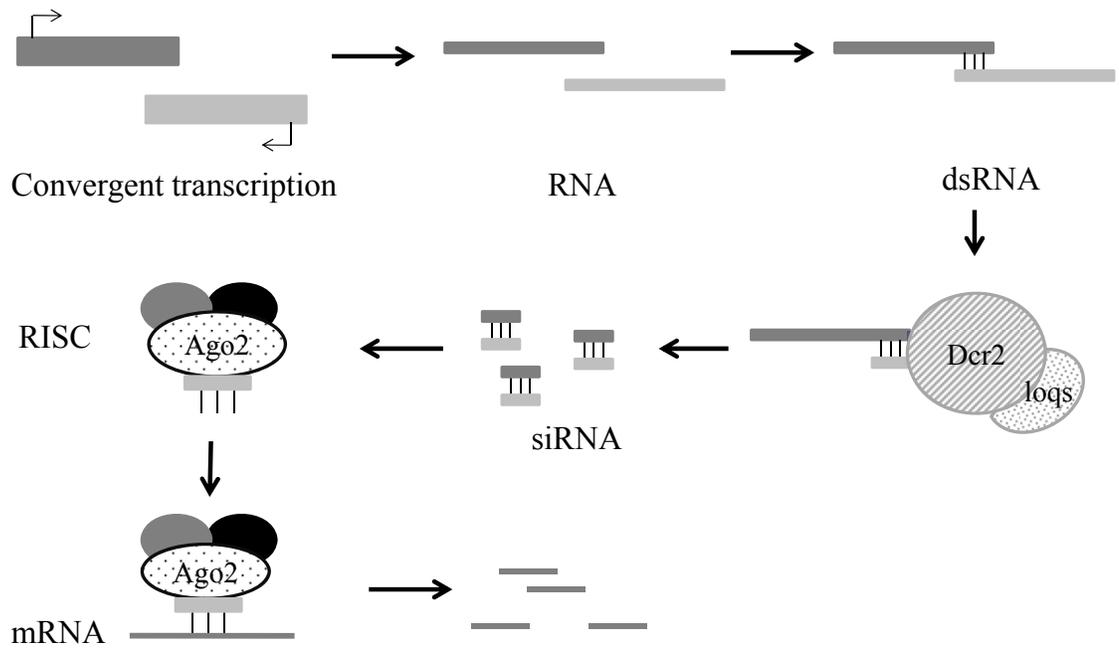
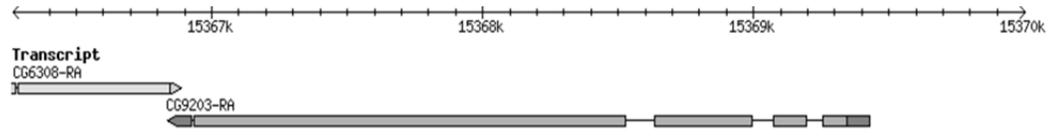


Figure 1.4 *CG9203* and *CRI4033*

Genome maps of the *CG9203* (A) and *CRI4033* (B) regions. The top portion in each figure shows the scale bar in kilobases. Transcripts are shown with the exons as thicker bars and the introns as thin lines. The lighter grey bars of the exons indicate amino acid coding regions and the darker grey bars indicate non-coding regions. A portion of a neighbouring gene is shown to the left of *CG9203*(A). The full length of the gene *tkv*, into which *CRI4033* is inserted, is shown, along with a higher resolution map of the *CRI4033* region. Also shown are other genes that overlap with *tkv* introns: two tRNA genes to the left of *CRI4033* and another pseudogene and 3 cytochrome P450 genes to the right (B).

Figure adapted from Flybase (dos Santos *et al.* 2014).

A.



B.

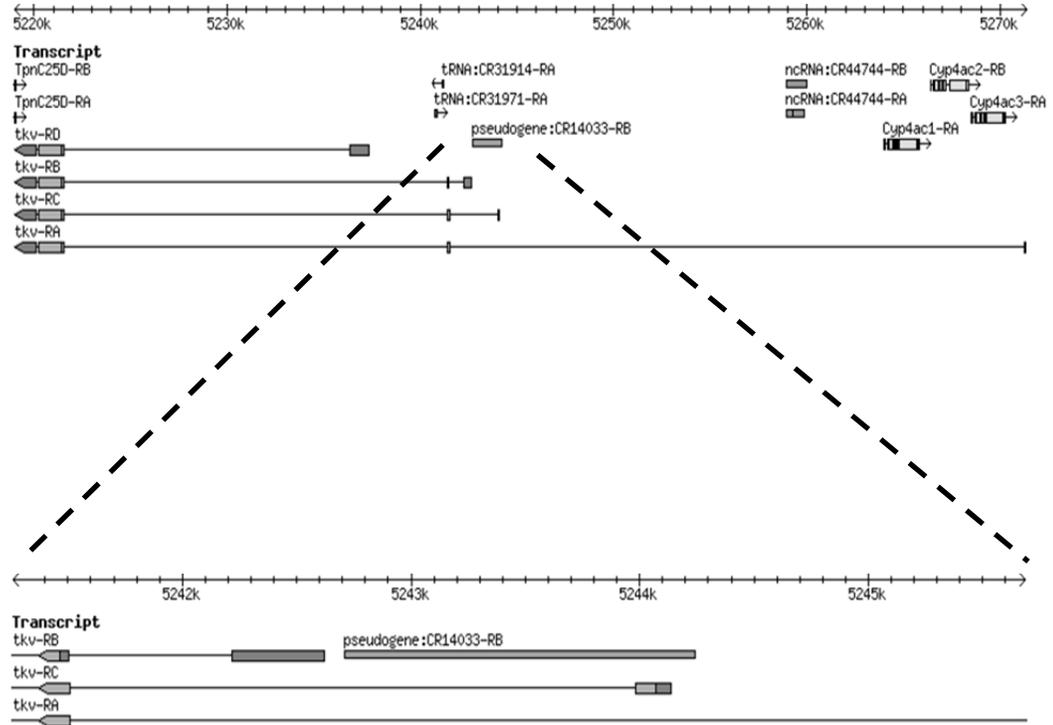
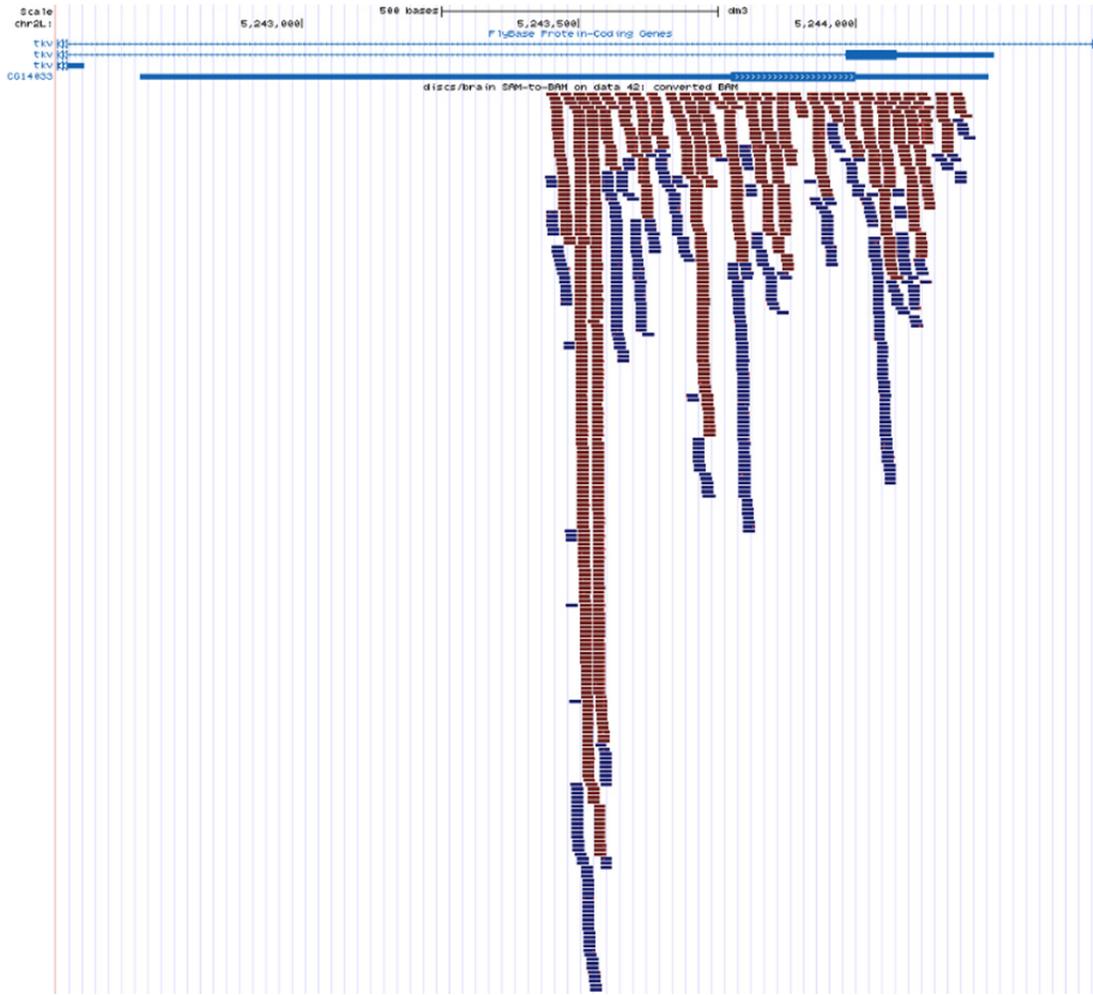


Figure 1.5 Small RNAs associated with *CR14033* in larval brains and imaginal discs

Small RNAs were isolated from larval brains and imaginal discs and strand-specific cDNAs were sequenced and aligned to the genome, including those for the region of *CR14033* that is similar to *CG9203* (Okamura *et al.* 2008). The blue and red lines represent these sequences, with + strand and – strand alignments, respectively. The siRNA data set (GSM399105) was aligned to the *Drosophila* genome (r. 5) using Galaxy server tools (Goecks *et al.* 2013) and visualized using the UCSC browser (Karolchik *et al.* 2012) (Denise Clark personal communication). Note that no siRNAs mapped to the 5' half of the *CR14033* transcribed region that does not share sequence similarity with *CG9203*.



Chapter 2: Materials and Methods

2.1 *Drosophila* stocks and culture

Fly stocks were maintained in vials at 25 °C and 60% relative humidity on a 12 hour light/dark cycle on cornmeal molasses agar media (Lakovaara 1969). Flies were either cultured in 37 mL vials (Fisher Scientific) or, for larger numbers, in 177 mL bottles (Fisher Scientific) on Instant *Drosophila* Media (Carolina Biological) supplemented with live yeast. Virgins were collected twice daily at 25 °C and left at 20 °C for 3 days to verify virginity. Experiments were conducted at 25 °C unless otherwise stated.

Strains used for P-element mutagenesis screens, complementation tests, RNAi screens, balancing, and sequence characterization are provided in Table 1.

2.2 P-element mutagenesis of *CRI4033*

P-element transposon insertions near *CRI4033* and *CG9203* were used to generate deletion mutations by screening for imprecise excisions of the P-elements. Briefly, the P-elements have a marker gene that allows one to detect them. The screen for deletions was done in two steps. The first step was to screen for individual flies that have lost the P-element as indicated by loss of the marker gene. The second step was to breed these flies and then screen a sample of genomic DNA using PCR to identify a deletion using primers that flank the insertion site of the P-element. PCR products smaller than wild-type indicate the presence of a deletion mutation. The details of the fly genotypes and crosses for *CRI4033* are described in this section and for *CG9203* in section 2.3.

Three strains carrying P-element transposons, $P\{SUPor-P\}tkv^{KG01923}$, $P\{lacW\}l(2)k01302^{k01302}$, and $P\{GSV6\}GS17311/SM1$, inserted 20 bp, 177 bp, and 251 bp downstream from the 5' end of *CR14033*, respectively, were used in screens for *CR14033* deletion mutations generated by imprecise excision of the P-element (Bellen *et al.* 2004). These P-elements are non-autonomous and, therefore, to excise the P-element, the P-element insertion strain was crossed to a strain producing transposase ubiquitously but lacking autonomous P-elements (Robertson *et al.* 1988) (Figure 2.1). Thus, the F1 progeny carried both the P-element as well as the gene for transposase. For the F1 cross, individual males carrying $P\{SUPor-P\}tkv^{KG01923}$ or individual males carrying $P\{lacW\}l(2)k01302^{k01302}$ were crossed to females carrying balancers that would allow lethal excision alleles to be maintained in subsequent generations. A P-element excision was indicated by a change from a wild-type eye color to white. For the F2 cross, two males with P-element excisions were isolated from each individual F1 cross based on eye color (white) and back-crossed to the same females as in the previous generation. The F3 cross consisted of crossing siblings with the same genotype from the F2 cross to produce a stock for each excision allele.

Excisions generated using insertion $P\{GSV6\}GS17311$ were generated by Dr. Denise Clark using the aforementioned method.

2.3 P-element mutagenesis of *CG9203*

Three strains carrying P-element insertions in or near the 5' end of *CG9203*, $P\{SUPor-P\}CG9203^{KG05829}$, $P\{XP\}d04265$, and, $y^* w^* P\{GawB\}NP0833/FM7c$, were used in screens for *CG9203* deletion mutations. $P\{SUPor-P\}CG9203^{KG05829}$ is located

332 bp into the 5' end of *CG9203* while $P\{XP\}d04265$ and $y^* w^* P\{GawB\}NP0833/FM7c$ are located 665 bp and 707 bp, respectively, upstream of *CG9203* (Bellen *et al.* 2004). Each transposon contains a visible marker to allow one to detect the P-element. The marker used for $P\{XP\}d04265$ and $y^* w^* P\{GawB\}NP0833/FM7c$ was *white* while *yellow* was used for $P\{SUPor-P\}CG9203^{KG05829}$ (Bellen *et al.* 2004).

Females from the aforementioned strains were crossed to males expressing a ubiquitous *transposase* gene. In the F1 generation, individual male progeny, carrying both the P-element and *transposase* chromosome, were crossed to virgin females with an *FM7c* X chromosome balancer, which carries the Bar eye dominant marker. In the F2 generation, excisions of $P\{XP\}d04265$ and $y^* w^* P\{GawB\}NP0833/FM7c$ were isolated by screening for virgin females with white Bar eyes. These females were crossed to Bar-eyed males to establish a stock in the F3 generation (Figure 2.2A).

$P\{SUPor-P\}CG9203^{KG05829}$ crosses were carried out using the same flies and crosses as for the other two P-element insertion strains above except that *yellow*, which is a body color marker, was used instead of *white*, an eye colour marker (Roseman *et al.* 1995). In this case, a P-element excision was indicated by a change from a wild-type body color to yellow. Thus, instead of white, Bar-eyed F2 females being chosen as candidates for carrying P-element excisions (Figure 2A), for $P\{SUPor-P\}CG9203^{KG05829}$, Bar-eyed yellow bodied females were chosen (Figure 2.2B).

2.4 Balancing lethal P-element excision alleles with a GFP-marked balancer

To allow identification of homo- or hemizygous mutant larvae carrying adult lethal P-element excision alleles from the P-element mutagenesis screens for both

CRI4033 and *CG9203*, strains carrying these alleles and a balancer chromosome marked with a green fluorescent protein (GFP) transgene were constructed. The GFP transgene marker allows one to distinguish mutant homo- and hemizygotes from balancer hetero- and hemizygotes during embryonic and larval development as the non-GFP-expressing individuals. Virgin females carrying *CRI4033* or *CG9203* lethal alleles were crossed to males carrying a GFP transgene associated with a balancer chromosome, *CyO* and *FM7c*, respectively (Casso *et al.* 2000). Siblings were then crossed to establish a stock (Figures 2.3A and 2.3B).

2.5 Genomic DNA extractions

Genomic DNA extractions (Huang *et al.* 2009) were performed on multiple whole females and whole males as well as single non-GFP larvae for P-element excision allele DNA characterization. 100 μ L of Buffer A (Appendix 7) was added to the flies. The flies were homogenized using Kontes pellet pestles (Fisher Scientific). Another 100 μ L of Buffer A was added and grinding continued until the flies or larvae were no longer visible. The tube and contents were then incubated at 65 °C for 30 minutes. 400 μ L of LiCl/KAc solution was added, mixed, and incubated on ice for 10 minutes. The tubes were then centrifuged at 14,000g for 15 minutes at room temperature. 500 μ L of the supernatant was transferred into a new tube with 300 μ L of isopropanol. The contents of the tube were mixed and spun for 15 minutes and then the supernatant was removed and the pellet was washed with 70% ethanol. The tube was then centrifuged for 5 minutes, the ethanol was removed, and the pellets were dried. The DNA was resuspended in 50 μ L of Tris-EDTA (TE) buffer (pH 8) and stored at -20 °C.

2.6 Complementation tests for excisions of P-elements associated with *CRI4033*

Duplicate crosses of 3-4 virgin females carrying lethal excisions from the P-element mutagenesis screens for *CRI4033* as well as those carrying the original P-element strains were crossed to 2-3 males carrying alleles of *tkv* or a deficiency of the *tkv* region in vials. Crosses were transferred to a new vial after four days. All adult progeny from the cross were examined for a presence of a phenotype and scored to completion. *Canton S* and a different *tkv* allele were used as negative and positive controls for non-complementation, respectively (Figure 2.4).

2.7 Duplication mapping for lethal excisions of $P\{SUPor-P\}CG9203^{KG05829}$

Two crosses of two virgin females from the lethal P-element excision allele stock, 28B-D, were crossed to two *Dp(1;3)DC300* males in vials at 25 °C. Progeny from this cross were scored to completion. Two *Canton S* virgin females crossed to two *Dp(1;3)DC300* males was used as a wild-type control. Two virgin females from the original P-element strain, 14125, were also crossed to two *Dp(1;3)DC300* males (Figure 2.5B).

2.8 Sequence characterization of *CRI4033* and *CG9203* excision alleles

For viable P-element excisions of $P\{SUPor-P\}tkv^{KG01923}$, five homozygous adult virgin females or single non-GFP larvae were collected for genomic DNA extraction. For excisions of $P\{SUPor-P\}CG9203^{KG05829}$, homozygous adult virgin females of variable number were collected. Genomic DNA from 2057 females was used as a wild-

type control. 2057 is the reference strain that was used for the *Drosophila* genome sequencing project (Adams *et al.* 2000) and is therefore the sequence on which PCR primers were designed. For lethal P-element excisions from *P{lacW}l(2)k01302^{k01302}*, 1-5 adult progeny generated during deficiency mapping or non-GFP larvae were collected for genomic DNA extraction (Huang *et al.* 2009). Adults and larvae were collected in this way because the *P{lacW}l(2)k01302^{k01302}* chromosome carries a second-site larval lethal mutation.

PCR products were generated using 14403, 10504, and 14125 primer pairs that flanked the original P-element insertion sites from a distance of 1 kb away to 8 kb (Table 2.2). These primers flank, *P{SUPor-P}tkv^{KG01923}*, *P{lacW}l(2)k01302^{k01302}*, and *P{SUPor-P}CG9203^{KG05829}* respectively. Long Amp Taq Polymerase (New England Biolabs) was used to amplify products from genomic DNA using the manufacturer's protocol. The PCR cycles consisted of an initial denaturation of 30 seconds at 94 °C followed by 35 cycles of denaturation at 94 °C for 15 seconds, primer annealing 60 °C for 15 seconds and elongation at 65 °C at 50 seconds/kb. The final extension time was set at 3 minutes or 10 minutes at 65 °C. PCR products were gel-purified (Qiagen) and sent for sequencing at The Center for Applied Genomics (TCAG) at the University of Toronto.

2.9 CG9203 and CRI4033 transgene construction

Transgenes for *in vivo* expression of *CG9203* and *CRI4033* cDNAs were constructed for germ-line transformation. Two cDNA clones, MIP11944 for *CG9203* and LD26673 for *CRI4033*, were obtained from the *Drosophila* Genomics Resource Center. The Celniker Lab and the modENCODE consortium already transformed

MIP11944, in plasmid vector pOTB7, into *E. coli*. LD26673, in pOT2, needed to be transformed. Both vectors carry an antibiotic resistance marker to chloramphenicol.

The first step in making the transgenic flies was to generate PCR products for MIP11944 and LD26673. Plasmid DNA was generated by selecting single colonies of each and growing them in 5 mL of LB broth with 3.6 μ L of 25 μ g/mL chloramphenicol at 37 °C in a shaker incubator. Plasmid DNA was extracted from the cultures by an Alkaline Lysis method (Sambrook *et al.* 1989). A PCR product of the cDNA insert was generated using LD26673F, LD26673R, MIP11944F, and MIP11944R primers (Table 2.2) and high fidelity Phusion DNA polymerase (New England Biolabs), which produces blunt-ended PCR products. The resulting PCR products were purified from a 0.7% agarose gel using QIAEX II (Qiagen).

The second step was to clone the PCR products into a standard cloning vector with more restriction sites to enable the final cloning step. Blunt end ligation was first used to insert the purified PCR products into the pBlueScript SK (pBSK) vector, and blue-white colony selection on X-gal and IPTG with 100 μ g/mL ampicillin was used to determine which colonies had the insert in the vector. Single colonies were grown in LB broth with ampicillin at 37 °C for 16 hours. Again, an Alkaline Lysis method was used to retrieve the plasmid DNA (Sambrook *et al.* 1989). The inserts were then sequenced to verify that no errors had been incorporated during PCR. Samples were sent to TCAG. After sequencing, the contigs were assembled and compared to Flybase sequence for these cDNAs. For the MIP11944 assembled from TCAG, a T was missing around 1480 bp. When looking more closely at the trace, it looks like two T's should have been called as the peak is twice of that of the T next to it. Therefore it appears to be a base calling

error rather than an actual mutation in the sequence. The LD26673 sequence was wild-type (data not shown).

The final step in the construction of the plasmid that allows insertion of the cDNA into the genome was to transfer the cloned cDNA sequences into the fly transformation vector, pUASp (Rorth 1998). Restriction digests were performed on LD26673 and MIP11944 in pBSK, as well as pUASP, with the restriction enzyme XbaI to retrieve the cDNA and to cut pUASp for future ligation. Subsequently, the ends of the pUASp vector were dephosphorylated with calf intestinal alkaline phosphatase (CIP) to prevent the vector from ligating to itself in the following steps using the supplier's protocol (New England Biolabs).

The digested pUASp plasmid DNA and cDNAs were purified and concentrated using phenol extraction and ethanol precipitation. The purified cDNA restriction fragments isolated from the pBSK vector were ligated into the CIP-treated pUASp vectors using T4 DNA ligase and following the supplier's protocol (New England Biolabs). The ligation products were then transformed into competent *E. coli* DH5 α and plated on LB agar plates with ampicillin. Plasmid preparations were used to recover the plasmid DNA (Sambrook *et al.* 1989).

To determine the orientation of the inserts in pUASp, additional restriction digests were carried out. For the clones containing LD26673, a KpnI digest was used and, for MIP11944 clones, a XbaI and BamHI digest and a BamHI digest were used. 50 mL cultures containing the four desired plasmids with inserts in both orientations (Figures 2.6 and 2.7) were then grown using LB broth with ampicillin and were extracted and highly purified using midiprep columns (Qiagen). This DNA was sent to Genetic Services for

P-element-mediated transformation into *D. melanogaster*. Larvae from w^{1118} embryos that had been injected with one of the four constructs were received from Genetic Services and incubated at 25 °C. The resulting progeny were backcrossed to w^{1118} of the opposite sex to allow identification of transformants, since pUASp carries the *white* gene marker, making their eyes pigmented.

For mapping transgene insertions, the transformants were crossed to *CyO/Sco*; *TM6B/MKRS* of the opposite sex. Siblings generated from this cross carrying either the balancers *CyO*; *TM6B* or *CyO*; *MKRS* were crossed together and independent assortment of markers was then used to determine chromosome linkage in the next generation (Greenspan 1997).

2.10 Inducing expression of *CG9203* and *CRI4033* transgenes

To induce expression of the cDNA transgenes, it is necessary to generate flies expressing the transcription factor GAL4, which binds to the UAS of the cDNA transgene and induces transcription (Duffy 2002). Transgenic flies were generated using the method described previously and crossed to flies that carry *tubP-GAL4*. The *tubP-GAL4* segment is controlled by the tubulin 1 α promoter where it can be expressed in any cell at any age (O'Donnell *et al.* 1994).

Two strains of transgenic flies, which carried a *CG9203* or *CRI4033* transgene inserted on either the X or 3rd chromosome, for each construct, were chosen to have expression of the transgene induced. This was achieved by crossing homozygous virgin females in the case of the X chromosome insertions, or males for chromosome 3 insertions, to flies of the opposite sex of the genotype w^* ; *tubP-GAL4/TM6B*. The w^{1118}

host strain crossed to the same strain as the transgenics was used as a negative control for induced expression. Crosses were done at 25 °C in vials.

2.11 Assessment of sterility in *CRI4033* and *CG9203* transgenics

Males and females expressing the cDNA transgene for two constructs of both *CRI4033* and *CG9203* were collected and individually crossed to a single virgin females or males, respectively, with the cDNA transgene but no *tubP-GAL4* chromosome. Presence of at least L1 larvae was deemed to show fertility. *w¹¹¹⁸* males or virgin females with the *tubP-GAL4* chromosome were used as a negative control and crossed to a single *w¹¹¹⁸* virgin females or males without the *tubP-GAL4* chromosome.

2.12 Collection of imaginal discs and brains and testes for RNA expression analysis

The flies carrying a UAS transgene were crossed to a strain with GAL4 (*w^{*}; tubP-GAL4/TM6B*). Resulting non-Tubby 3rd instar wandering larvae were collected, as were *loqs*, *Dcr2*, *w¹¹¹⁸*, and *w¹¹¹⁸; +; tubP-GAL4* 3rd instar wandering larvae. The brains and discs were dissected from the larvae in 1X PBS then placed in 100 µL of TRIzol™ Reagent and put at -80°C until needed for qPCR. Three sets of brains and discs were placed in each tube.

Males from the aforementioned crosses were collected between 0-24 hours after eclosion and placed in vials at 20 °C for 3 days. After exposure to CO₂, unconscious flies were placed in glass dishes on ice for dissection of testes. The dissected tissue was placed in 100 µL of TRIzol™ Reagent and stored at -80 °C. Three sets of testes were put in each tube.

2.13 RNAi of *CG9203* and *CRI4033*

Strains v23052 and v29075 carry transgenes that express hairpin RNA in the presence of *GAL4*. The sequence of the inverted repeats that generates the hairpin RNA in strain v23052 overlaps with the 3' end of the pseudogene, which is the portion that is similar to the parent gene. The inverted repeats that produce hairpin RNA in strain v29075 corresponds to a region of *CG9203* that is not found in *CRI4033* therefore only *CG9203* transcript should be affected.

Virgin females from v23052 and v29075 were crossed to *w**; *tubP-GAL4/TM6B*. The phenotype of the adult progeny that are expressing these hairpin RNAs was scored. In addition, males were checked for sterility by crossing individual flies to siblings with the RNAi transgene but no GAL4 driver (Figure 2.8).

2.14 RNA extraction

RNA was extracted from samples consisting of 10 whole flies, 1 pair of ovaries, 3 sets of testes, or larval brains and imaginal discs. Samples were removed from the freezer and an additional 250 μ L of TRIzolTM Reagent was added. The flies/tissues were homogenized using Kontes pellet pestles (Fisher Scientific). The tube was then centrifuged at 12,000g at 4 °C for 10 minutes. The supernatant was transferred to a new tube and left at room temperature for 5 minutes. 50 μ L of chloroform was then added and the solution was shaken for 15 seconds and then centrifuged again for 15 minutes. The top layer was pipetted into a new tube with 250 μ L of isopropanol to precipitate the RNA along with 1 μ L of glycogen (20 μ g). Extractions were left at -20 °C overnight and

then centrifuged at 12,000g at 4 °C for 15 minutes. The supernatant was removed and the pellet was washed with 500 µL of 75% ethanol with DEPC-treated H₂O and centrifuged at 7500g at 4 °C for 5 minutes. The ethanol was removed and the pellet was air-dried. The RNA was resuspended in 40 µL of DEPC ddH₂O and placed in a 65 °C water bath for 10 minutes. The RNA was then stored at -20 °C. The volumes of solutions used to extract the RNA from brains and discs as well as the transgenic flies with the reverse orientation of LD26673 and MIP11944 and the forward orientation of MIP11944 were halved.

RNA extraction quality was assessed on a 0.7% agarose gel by mixing 4µL of the sample with 16 µL of Ambion gel loading buffer II containing 10 µg/mL ethidium bromide (EtBr). The mixture was heated for 10 minutes at 65 °C prior to gel loading. The gel was run for approximately 1 hour at 100 volts (7 volts/cm) and visualized with BioRad Gel Doc imaging system with Quantity One[®] software.

2.15 DNase I Treatment of RNA

RNA samples were treated with RQ1 RNase-Free DNase I (Promega). 1 µL of DNase I, 1 µL of 10x reaction buffer and approximately 100 ng of RNA in an 8 µL volume were combined and incubated at 37 °C for 90 minutes. 1 µL of RQ1 DNase Stop Solution was added and incubated at 65 °C for 10 minutes to inactivate the DNase. Samples were checked for remaining DNA using quantitative-reverse transcription-PCR (qPCR) with *CG9203* 3L and 3R primers or *CR14033* 2L and 2R primers.

2.16 cDNA synthesis

The DNased RNA sample was made into cDNA using the Invitrogen SuperScript™ III Platinum Two-Step qPCR Kit or the Superscript® III First-Strand Synthesis SuperMix. These kits contain the necessary components to make cDNA, which includes reverse transcriptase, random hexamers, and deoxynucleoside triphosphates (dNTP). The recommended procedure was followed and for each sample a control was made by not including reverse transcriptase in the reaction.

2.17 Quantitative Reverse Transcriptase-PCR

cDNA samples were used for the qPCRs using the Corbett RG-6000 machine. qPCR was performed using the Invitrogen SuperScript™ III Platinum Two-Step qPCR Kit with SYBR Green or KAPASYBR® FAST Universal 2X qPCR Master Mix following the manufacturer's instructions except that 12.5 µL reactions in 0.2 mL tubes were used. Primer pairs used in the reactions can be found in Table 2.2. The Real-time PCR cycling protocol used for each run was a single cycle of 50 °C for 2 minutes, followed by a single cycle of 95 °C for 2 minutes, followed by 40 cycles of 95 °C for 15 seconds, then 60 °C for 30 seconds, and then 72 °C for 30 seconds. Finally, a melt curve was performed from 72 °C to 95 °C. Data were analyzed using LinReg to calculate Ct (cycling threshold) values and primer efficiencies and REST to determine relative expression levels (Ramakers *et al.* 2003 and Pfaffl 2001). REST calculates the difference in expression level of a gene in two samples by using a reference gene, in this case *rp49*, to normalize the expression level of the gene of interest. It compares differences between Ct values to determine the fold change in gene expression, 95% confidence intervals, and standard errors among replicates. A Ct value indicates the number of cycles it took in

order for a fluorescent signal to be registered. It is a measurement of how much cDNA was in the sample at the start of the run (Pfaffl *et al.* 2002). The fold change is also reported with the probability of a type 1 error, p(H). Three technical and biological replicates were done for each RNA sample in an experiment.

2.18 Comparative sequence analysis

For comparison of *D. melanogaster* *CR14033* with *CG9203* using sequence alignment, as well as examining appearance of *CR14033* orthologs in twelve other *Drosophila* species, various sequence analysis tools and data were used. The cDNA sequence for *CR14033* (LD26673) was obtained from Flybase (dos Santos *et al.* 2014). This sequence was used in a BLAST homology search using Flybase BLAST against the genomes of *D. simulans* (CH988051, CH983994), *D. sechellia* (CH480820, CH480888), *D. melanogaster* (AE014134, AE014298), *D. yakuba* (CM000157, CM000162), *D. erecta* (CH954177, CH954180), *D. anaassae* (CH902620), *D. pseudoobscura*, *D. persimilis* (CH479205, CH479180), *D. willistoni* (CH964239, CH 964182), *D. mojavensis* (CH933811), *D. virilis* (CH940653), *D. grimshawi* (CH916370, CH916372).

A second *D. simulans* genome sequence was obtained from the Andolfatto Lab at Princeton University (Hu *et al.* 2013). The sequence was used in the same manner as the *D. simulans* sequence available on Flybase.

Table 2.1: Stocks used in this study

The stock number, genotype, and stock center source are indicated.

Stock #	Genotype	Stock Center
1	<i>Canton S</i>	Bloomington Drosophila Stock Center
427	<i>tkv¹</i>	Bloomington Drosophila Stock Center
2057	<i>y¹; Gr22b¹ Gr22d¹ cn¹ CG33964^{R4.2} bw¹ sp¹; LysC¹ MstProx¹ GstD5¹ Rh6¹</i>	Bloomington Drosophila Stock Center
3242	<i>tkv¹ cn¹ bw¹ sp¹/CyO</i>	Bloomington Drosophila Stock Center
5138	<i>w[*]; P{tubP-GAL4}LL7/TM6B, Antp^{Hu} Tb¹</i>	Bloomington Drosophila Stock Center (This particular genotype was derived in the Clark lab)
5905	<i>w¹¹¹⁸</i>	Bloomington Drosophila Stock Center
5907	<i>w⁻; Sco/CyO</i>	Bloomington Drosophila Stock Center
7497	<i>w¹¹¹⁸; Df(2L)Exel6011, P{XP-U}Exel6011/CyO</i>	Bloomington Drosophila Stock Center
7707	<i>Df(1)Exel6233, w¹¹¹⁸ P{XP-U}Exel6233/FM7c</i>	Bloomington Drosophila Stock Center
8201	<i>y¹ w¹¹¹⁸; CyO, PBac{Delta2-3.Exel}2/amos^{Tft}</i>	Bloomington Drosophila Stock Center
10504	<i>y¹ w^{67c23}; P{lacW}l(2)k01302^{k01302}/CyO</i>	Bloomington Drosophila Stock Center
14125	<i>y¹ P{SUPor-P}CG9203^{KG05829}</i>	Bloomington Drosophila Stock Center
14403	<i>y¹ w^{67c23}; P{SUPor-P}tkv^{KG01923}</i>	Bloomington Drosophila Stock Center
18371	<i>w¹¹¹⁸, PBac{WH}loqs^{f00791}/CyO</i>	Bloomington Drosophila Stock Center
200-010	<i>y¹ w^{67c23}; P{GSV6}GS17311/SM1</i>	Drosophila Genomics Research Center
v23052	<i>w¹¹¹⁸; P{GD12463}v23052</i>	Vienna Drosophila RNAi Center
26545	<i>w¹¹¹⁸; Df(2L)BSC693, P+PBac{XP3.WH3} BSC693/SM6a</i>	Bloomington Drosophila Stock Center
v29075	<i>w¹¹¹⁸; P{GD14153}v29075</i>	Vienna Drosophila RNAi Center
30414	<i>w¹¹¹⁸; Dp(1;3)DC300, PBac{DC300}VK00033</i>	Bloomington Drosophila Stock Center
32064	<i>y^{d2}, w¹¹¹⁸ P{ey-FLP.N}2; Dcr-2^{L811jSX}; P{Dcr-2^{E1371K.t7.2}}3</i>	Bloomington Drosophila Stock Center
d04265	<i>w[*]; P{XP}d04265</i>	Exelixis at Harvard Medical School

Stock #	Genotype	Stock Center
5907	<i>w¹¹¹⁸/Dp(1;Y)y⁺; sna^{Sco}/SM6a</i>	Bloomington Drosophila Stock Center
3703	<i>w¹¹¹⁸/Dp(1;Y)y⁺; CyO/nub¹ b¹ sna^{Sco} lt¹ stw³; MKRS/TM6B, Tb¹, w¹¹¹⁸/Dp(1;Y)y⁺; sna^{Sco}/SM6a</i>	Bloomington Drosophila Stock center
	<i>ade2¹⁻⁶/CyO, KrGAL4, UASGFP</i>	Bloomington Drosophila Stock Center (Casso <i>et al.</i> 2000) and Holland <i>et al.</i> 2011
	<i>Df(1)Exel6233/FM7, Act-GFP</i>	<i>dm⁴/FM7, Act-GFP</i> was obtained from Bruce Edgar. Denise Clark used this strain in combination with 7707 to generate this deficiency train
103-781	<i>y[*] w[*] P{GawB}NP0833/FM7c</i>	Drosophila Genomics Research Center

Table 2.2 Primers used for this study

In the table are primers used for different aspects of this study. The primers at the beginning of the table that are named according to the strain are the primers used for PCR analysis and sequencing of P-element excisions. The next set of primers that are named for the genes were used for qPCR and the last section of primers were used for sequencing the cDNA insert when making the UAS transgene plasmid constructs. They are labeled according to the name of the cDNA. The direction of the primer is shown as L or R. The L represents left and the R represents right meaning that the left is going forward in the sense direction and the right is going in reverse in the antisense direction. The numbers before the L and R show how many kilobases away the primer is from the gene of interest. Int is for internal. These primers were used for sequencing and when trying to determine the reason for the lack of PCR product using other primer pairs.

Primer Name	Sequence
14125 int	TGTCCGGGGTGGGATCGACC
14125L	CCGCTTCTTGGCTGACGGCT
14125R	TGCCTTGACAAGCGTTGGGCC
14125_2L	CTTGCGACCAGCGCAGACGG
14125_2R	CGTTCGCCAGTGTTCGAAGG
14125_3L	TACCCCGGTTCCGACAGGC
14125_3R	TGCCAAGGTGATGGGGTGC
14125_4L	GGCCTGCTCGTACACCCTGG
14125_4R	ACACCGGCACATCGGGTAGG
14125_5L	AGGTGATCTTCCAGCGCGGC
14125_5R	ATCGGCCGTACGTTTCGAGG
14125_6L	CTCGAACTGCCGGCTCACGG
14125_6R	ATGTCCCGGGTAGCGGTGGG
14125_7L	CGATCGTCGAGCCCTTCGGC
14125_7R	CGATCGTCGAGCCCTTCGGC
10504 int	GGGTGGGCGAGGAAAGTCGC
10504L	GTGCCTCTCCGCAGACGGCT
10504R	TTGAGCAAGTGGGCTGCACGG
14403 int	ATGTGTGCCACTCGACGGGC
14403L	TCGCACCAACGGAGAATGGGC
14403R	GCTAGTTCGTTGGCGGCTGTT
14403-2L	TGCGAGTTCAGGCGCGTCAA
14403-2R	GGCCGCGTTTTTCTTTGGGCC
14403-3L	TCGCTCGAGTGGAAATGCGGG
14403-3R	ACTCGGCGCGTTGATGGACA
14403-4L	TGGTTCGGGGTCTGCTCTGCT
14403-4R	CTGTGTGGCCGTCGAGTTGCA
14403_5L	GAGAGCGTGGCGGAAAACGC
14403_5R	CTCTCGTGCAGCTGTCGGCG
14403_6L	TGGTGGCGTTGGTGCTCTGC
14403_6R	CGCAGATCGAGTGGGAGGCG
14403_7L	ATGGCCATGTCCGCCGATCC
14403_7R	CTCCACGCGCTTTTGCTGCC
14403_8L	AAGGGGTGCCCGGATTGTTGC
14403_8R	GGCTGCGTCTCCGTCTGCC

Primer Name	Sequence
14125 intR	TGGTCACACTGCGGCGAAGC
14125 L out	AGGCCCTGCTCATATCGTATCGCC
14125 R out	CCAGCTGATAGTCCCTGTCCTGC
14403 L out	CGCAAACCCCTATCGATATCATTGAAC
14403 R out	CCAGTGCAACCACGAAGTTAACCG
CG9203 L	AAGCGGACCAGCAACCGAGC
CG9203 R	ATGTCTTCCTTGGGCGCGGC
CR14033-2L	AGCGAGGACGAACTTACTCCG
CR14033-2R	AGAGATTGATGCGCCTCAGC
thick left	CCGTCGCTGCTACACACCCG
thick right	TGATGGTATCGGCGGCCGGA
rp49LqPCR	AATCTCCTTGCGCTTCTTG
rp49RqPCR	CTAAGCTGTCGCACAAATGG
M1911944F	GCTCTAGAATTTTCGAAGTATTTTGTGCACCTCG
M1P11944R	GCTCTAGAGGGCGAATCCTGTAACATGTTTT
LD26673F	GCTCTAGAAAGCAATATAATTTGTTAGCTGAAAGTGCC
LD26673R	GCTCTAGAGACTTTACAATCAGTTGTTTTATATATTTGC
ex200010-18L	CAAATAAATTCTCCGTACCCAGG
ex200010-18R	CCGAACCCACAACAAATGC

Figure 2.1: Crossing scheme for P-element mutagenesis of the pseudogene, *CR14033*

The P generation cross was between adult virgin females homozygous for the P-element and males carrying the ubiquitously expressed P-element transposase gene $\Delta 2-3$ on the *CyO* chromosome. This transposase causes pre-meiotic excisions in the F1 progeny germline (Engels 1996). F1 males of the necessary genotype were chosen based on orange eyes and the presence of Curly wings and a non-Tuft thorax and then individually crossed to virgin females with a *Curly* balancer chromosome (*CyO*). From this cross, single F2 males displaying white eyes were chosen. The white eyes are caused by the P-element excision (designated $P\{*\}$ here), since the *white* gene is a marker inside the P-element and, when the P-element is present, this *white* gene is responsible for the red or orange eye color. The *white* gene in its normal location on the X chromosome is carrying a mutation (w) so that the P element's *white* gene marker expression can be followed. These F2s were then individually crossed to virgin females with a *Curly* balancer chromosome (*CyO*) and, from this cross, non-Scutoid F3 males and females, lacking the Sco-marked chromosome, were chosen to establish a stock. The figure here shows the crossing scheme using the P-element for $P\{SUPor-P\}tkv^{KG01923}$ in strain 14403. The scheme was the same for $P\{lacW\}l(2)k01302^{k01302}$ and $P\{GSV6\}GS17311$ in strains 10504 and 200-010, respectively.

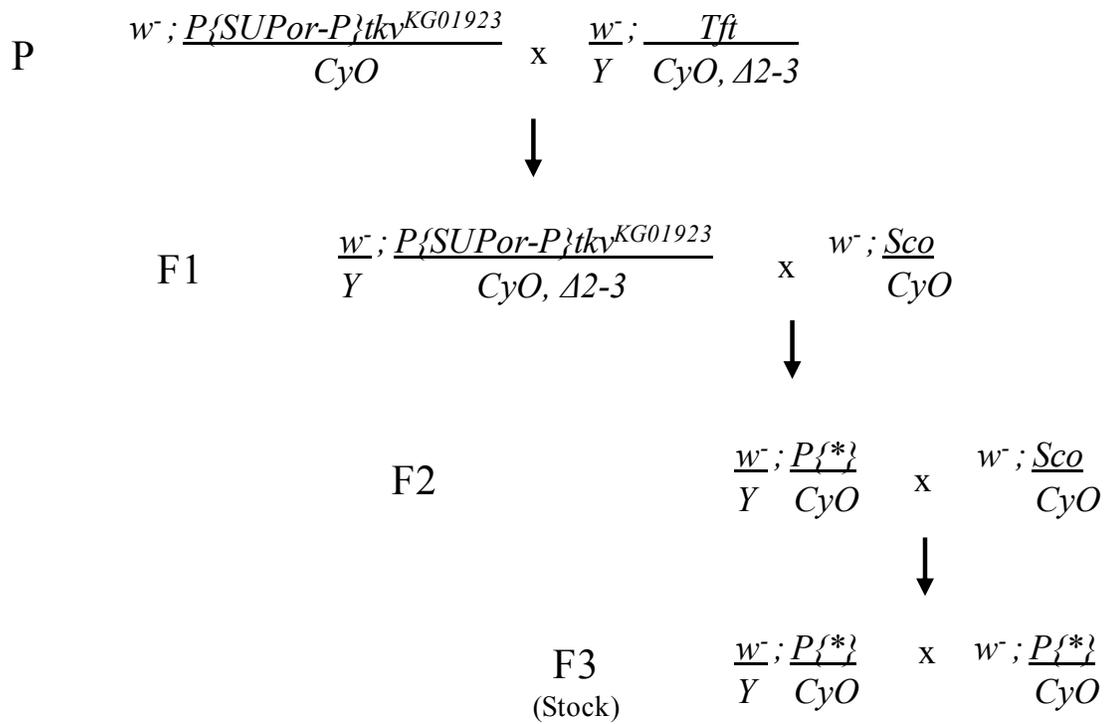


Figure 2.2: Crossing scheme for P-element mutagenesis of *CG9203* using the P-elements with a *white* and *yellow* marker gene

- A. For the two P-element insertions with a *white* marker gene, in strains d04265 and 103-781 (Table 2.1), virgin females from the P-element stock were crossed to males expressing transposase to cause pre-meiotic excisions in the germline of the F1 progeny. The male progeny chosen for the F1 cross had a Curly, non-Tuft phenotype, indicating they carried the $\Delta 2-3$ P-element transposase. These males were crossed to strain 7707 females carrying an X-chromosome balancer, *FM7c*, is marked with the dominant Bar-eye mutation. F2 females were chosen based on a white, Bar-eye phenotype. The white eyes show that the P-element excised and that a deletion of *CG9203* could have occurred. Individual F2 females were crossed to either sibling males or to 7707 males with Bar-eyes. The scheme for $P\{XP\}d04265$ in strain d04265 is shown. The same scheme was used for $P\{GawB\}NP0833$ in strain 103-781.
- B. For the P-element insertion $P\{SUPor-P\}CG9203^{KG05829}$ with a *yellow* marker gene, in strain 14125 (Table 2.1), the crossing scheme is laid out in the same manner as for the other P-element insertions in part A, except for the criteria used for screening for F2 females with a P-element excision. In this scheme, individual yellow-bodied bar-eyed females were chosen and crossed to Bar-eyed males.

A.

$$P \quad \frac{w^-, P\{XP\}d04265}{w^-, P\{XP\}d04265} \quad x \quad \frac{w^+; Tft}{Y \quad CyO, \Delta 2-3}$$

$$F1 \quad \frac{w^-, P\{XP\}d04265}{Y} ; \frac{+}{CyO, \Delta 2-3} \quad x \quad \frac{Df(1)Exel6233}{FM7c, w^B}$$

$$F2 \quad \frac{w^-, P\{*\}}{FM7c, w^B} \quad x \quad \frac{FM7c, w^B}{Y}$$

$$F3 \quad \frac{w^-, P\{*\}}{FM7c, w^B} \quad x \quad \frac{FM7c, w^B}{Y}$$

(Stock)

or

$$\frac{w^-, P\{*\}}{Y}$$

B.

$$P \quad y^l, P\{SUPor-P\}CG9203^{KG05829} \quad x \quad \frac{w^+; Tft}{Y \quad CyO, \Delta 2-3}$$



$$F1 \quad \frac{y^l, P\{SUPor-P\}CG9203^{KG05829}}{Y} ; \frac{+}{CyO, \Delta 2-3} \quad x \quad \frac{Df(1)Exel6233}{FM7c, w^B}$$



$$F2 \quad \frac{y^l, P\{*\}}{FM7c, w^B} \quad x \quad \frac{FM7c, w^B}{Y}$$



$$F3 \quad \frac{y^l, P\{*\}}{FM7c, w^B} \quad x \quad \frac{FM7c, w^B}{Y}$$

(stock)

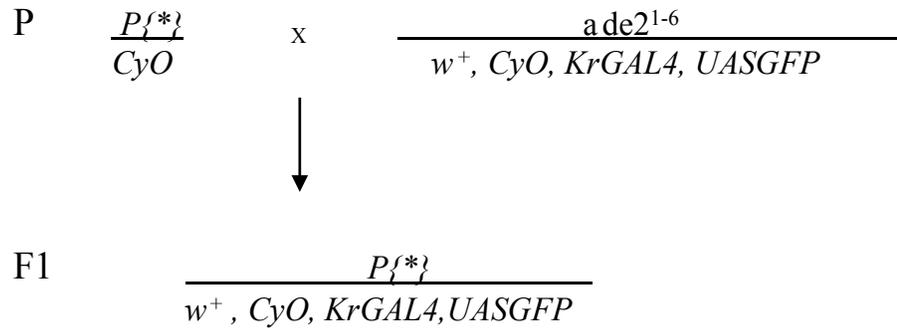
or

$$\frac{y^l, P\{*\}}{Y}$$

Figure 2.3: Balancing lethal *CR14033* and *CG9203* P-element excision alleles with GFP-marked balancer chromosomes

- A. Virgin females with a lethal P-element excision in *CR14033* balanced over *Curly* were crossed to males with a *Curly* balancer chromosome that also had GFP. Progeny were chosen based on eye color as the *Curly* GFP balancer chromosome also has a *white* gene giving the progeny with this chromosome an orange eye phenotype while the progeny inheriting the other *Curly* balancer chromosome have white eyes. The orange-eyed progeny, males and virgin females, were crossed together to make a stock.
- B. Virgin females with the lethal P-element excision in *CG9203* balanced over *FM7c* were crossed to males with the *FM7c* balancer chromosome with GFP and a *white* gene. Progeny with orange eyes were chosen to make a stock as these were the progeny to be carrying the *FM7c* balancer chromosome with GFP. The progeny inheriting the *FM7c* balancer had white eyes.

A.



B.

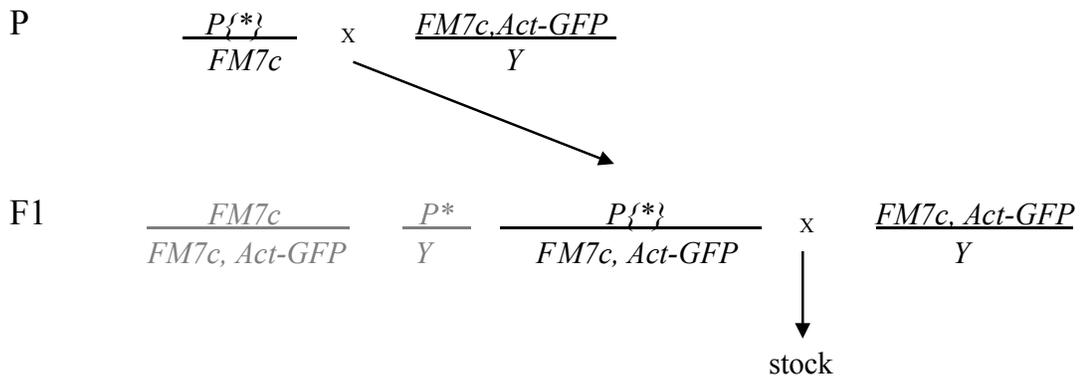


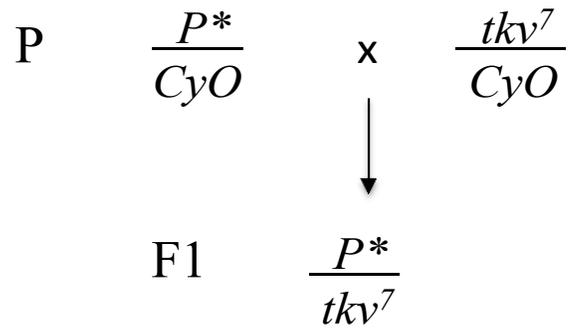
Figure 2.4: Complementation test using a *tkv* allele crossing scheme

- A. Complementation test cross between a lethal *CR14033* excision chromosome and the *tkv*⁷ allele. Both the male and female progeny without curly wings, therefore possessing both the 2nd chromosome with the lethal P excision and the 2nd chromosome with *tkv*⁷, were scored.

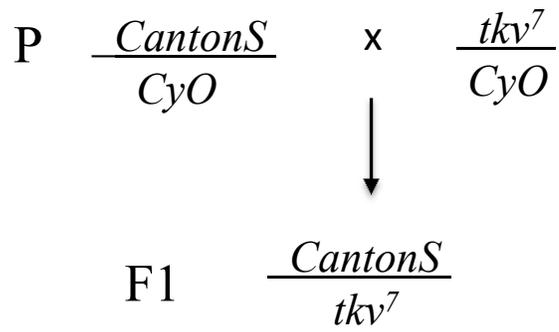
- B. Wild-type control cross, where females from a strain, *Canton S*, which is wild type for the *tkv* locus, were crossed to the same *tkv*⁷ allele males. Non-curly progeny were scored.

- C. Positive control cross, where another known *tkv* allele, *tkv*¹, tested for non-complementation with *tkv*⁷. Again, non-curly progeny were scored.

A.



B.



C.

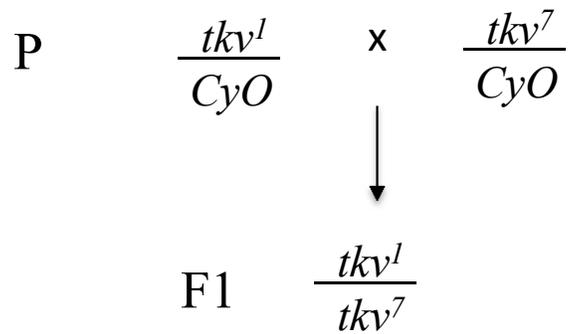
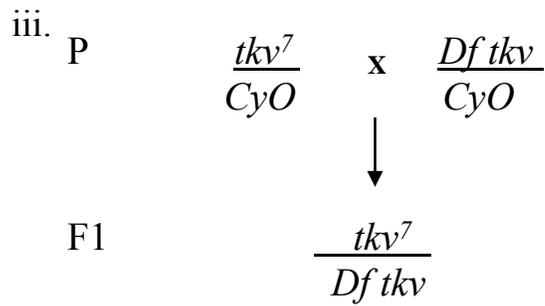
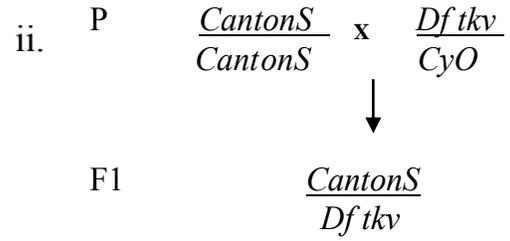
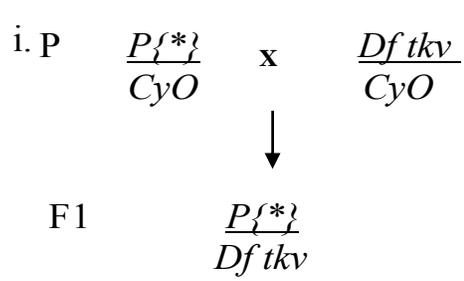


Figure 2.5: A) Complementation tests between deficiencies of *CR14033* and lethal P-element excisions of *CR14033* B) Complementation test using a duplication of the *CG9203* region and lethal P-element excisions of *CG9203*

- A. Virgin females from lethal P-element excisions strains generated from $P\{SUPor-P\}tkv^{KG01923}$, $P\{lacW\}l(2)k01302^{k01302}$, and $P\{GSV6\}GS17311$ (i) were crossed to males with deficiencies of the *tkv* region, $Df(2L)BSC693$ and $Df(2L)Exel6011$. In this figure these strains are denoted as $Df\ tkv$. The phenotype of the resulting progeny without the *Curly* balancer chromosome was scored. The deficiency of *tkv* strains were crossed to the tkv^7 (iii) strain as the positive control and *Canton S* (ii) as the negative control for non-complementation.
- B. Virgin females from lethal P-element excision strain, 28B-D, were crossed to males carrying a duplication inserted on chromosome 3 that has a region of the genome where *CG9203* is situated (iv). The presence of non-barred eyed males indicates that the duplication rescues the lethality.

A.



B. iv.

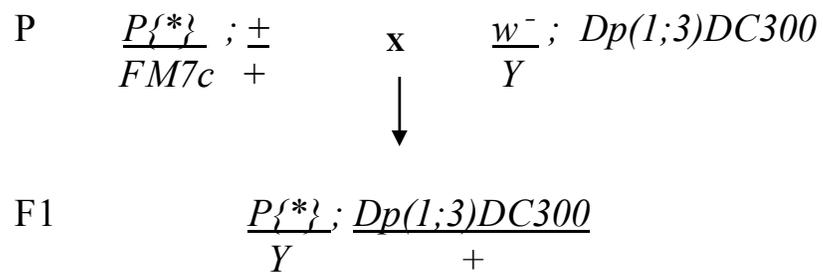
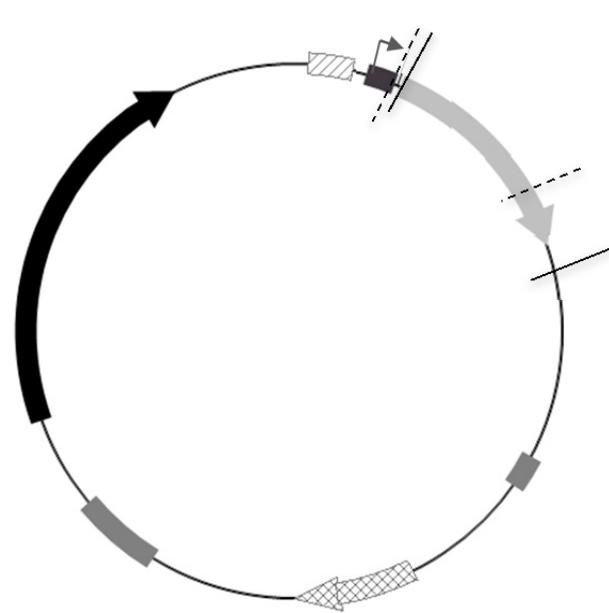


Figure 2.6: *CR14033* cDNA (LD26673) inserted in the fly transformation plasmid pUASP

- A. The plasmid is comprised of the cDNA and pUASp sequences. Here the cDNA is in the sense direction with respect to the transposase promoter. The 5' end of the P-element, farthest to the left and the 3' end of the P-element can be seen in a medium grey. The promoter for transposase is in dark grey with a bent arrow over it, the *ampicillin resistance* gene is the arrow with the hatching, and the mini *white* gene is the black arrow. The restriction enzyme cut sites used to cleave pUASp to insert the cDNA can be seen by the line upstream and downstream of the cDNA. The UAS, GAL4 binding site, is seen upstream to the transposase promoter as a grey and white box.
- B. The same elements as part A only the cDNA is in its antisense orientation.

A.



-  mini *white* gene
-  LD26673
-  amp^r gene
-  UAS
-  5' and 3' of P-element
-  Transposase promoter
-  XbaI cut site
-  BamHI cut site

B.

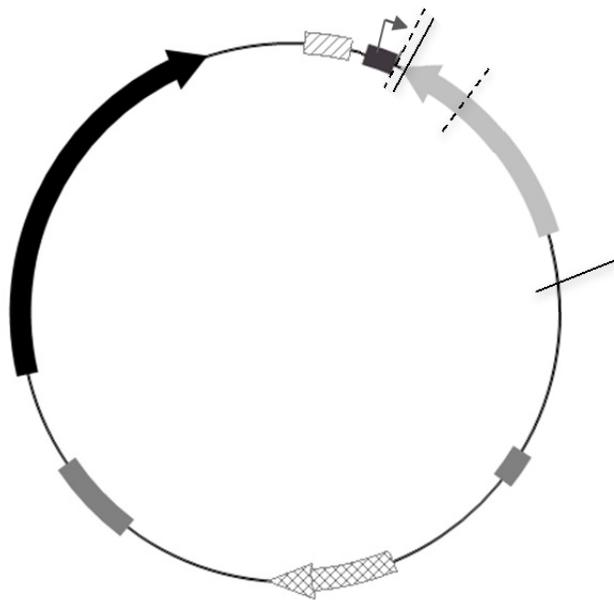
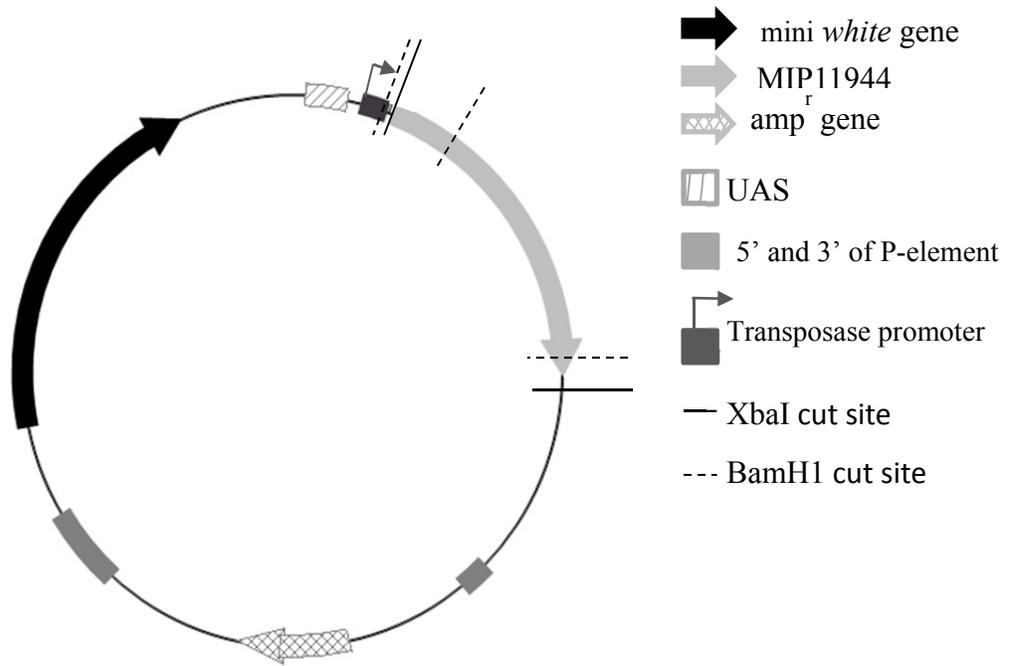


Figure 2.7: CG9203 cDNA (MIP11944) inserted in the fly transformation plasmid pUASP

- A. The figure shows the cDNA, MIP11944, as a light grey arrow while the remainder of the plasmid is pUASp sequence. The cDNA is in the sense direction with respect to the transposase promoter. The 5' end of the P-element, farthest to the left and the 3' end of the P-element can be seen in a medium grey. The promoter for transposase is in dark grey with a bent arrow over it and the *ampicillin resistance* gene is the arrow with the hatching and the mini *white* gene is the black arrow. The restriction enzyme cut site used to cleave pUASp to insert the cDNA can be seen by the line upstream of the cDNA. The UAS/GAL4 binding site is seen upstream to the transposase promoter as a grey and white box.
- B. Part B shows the same elements as Part A only the cDNA is in its antisense orientation.

A.



B.

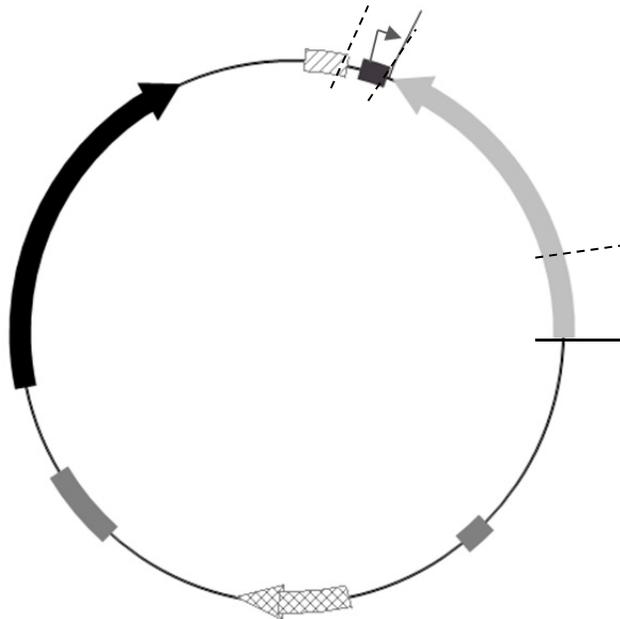


Figure 2.8: Induction of RNAi for *CRI4033* and *CG9203*

Strains with a transgene containing a UAS enhancer that binds GAL4 and drives the expression of inverted repeats, resulting in hairpin RNA, were used to cause RNAi of *CRI4033* and *CG9203*. Virgin females from each of these strains were crossed to males with the ubiquitously expressed *tubP-GAL4* driver. This crossing scheme is the same for both *CRI4033* and *CG9203* except that the sequence of the inverted repeats is different. When the hairpin is formed and cleaved by components of the silencing RNA pathway, these pieces of RNA will cause silencing of *CRI4033* with w^{1118} ; $P\{GD12463\}v23052$ and *CG9203* with w^{1118} ; $P\{GD14153\}v29075$ (A and B). 5905 was used as a control as it had w^{1118} as for the transgenic strains but was otherwise wild-type (C). The F1 females were chosen based on the absence of the Tb and Hu dominant markers on the *TM6B* balancer chromosome.

A.

P $w^{1118} ; P\{GD12463\}v23052 ; +$ x $w^* ; + ; \frac{tubP-GAL4}{TM6B}$



F1
(females) $\frac{w^{1118}}{w^*} ; \frac{P\{GD12463\}v23052}{+} ; \frac{tubP-GAL4}{+}$

B.

P $w^{1118} ; P\{GD14153\}v29075 ; +$ x $w^* ; + ; \frac{tubP-GAL4}{TM6B}$



F1
(females) $\frac{w^{1118}}{w^*} ; \frac{P\{GD14153\}v29075}{+} ; \frac{tubP-GAL4}{+}$

C.

P $w^{1118} ; + ; +$ x $w^* ; + ; \frac{tubP-GAL4}{TM6B}$



F1
(females) $\frac{w^{1118}}{w^*} ; \frac{+}{+} ; \frac{tubP-GAL4}{+}$

Chapter 3: Results

3.1 One deletion of *CRI4033* and one partial deletion of *CG9203* were generated using P-element mutagenesis

P-element mutagenesis was used to generate deletion mutations in *CRI4033* and *CG9203*. The purpose of generating deletion mutations in *CRI4033* was to use a direct method to determine if the pseudogene has a role in the parent gene's expression and to see if there was a phenotype associated with the lack of *CRI4033* to provide further insight into its function. If there were an effect on *CG9203*'s expression, through an endo-siRNA-mediated pathway, one would expect to see an increase in *CG9203* transcript when *CRI4033* is knocked out; aspects of its function could be determined based on a phenotypic change. The purpose for generating a deletion of *CG9203* was to determine its function based on mutant phenotype and to see if there is a corresponding increase in *CRI4033* intact transcript levels to provide evidence for an interaction between the two.

A useful tool to study gene function is P-element mutagenesis as it generates a mutation in a gene of interest. One way it does this is through an imprecise excision of a P-element inserted in or close to the gene of interest, potentially leading to deletion of sequences flanking the P-element insertion site (Hummel and Klambt 2008). This technique has been used previously to study numerous genes. For instance, Holland *et al.* (2011) used it to create loss of function mutations in *ade2*, a gene involved in purine *de novo* synthesis. Here, this technique was used with three different P-element-containing strains for each gene of interest to conduct P-element excision screens with the goal of generating deletions of *CRI4033* and *CG9203*. The six P-element insertion strains used

in this study were generated during the Berkeley Gene Disruption Project (Bellen *et al.* 2004).

The insertion of a P-element into a gene is in itself a mutation (insertional mutagenesis), which can lead to phenotypic changes. However, not all P-element insertions lead to a loss of function phenotype, especially if they are in non-coding sequence. Deletion mutations were created in order to examine the effects of the complete loss of transcription of *CRI4033* and *CG9203*. Deletion mutants were important to this study, as it is believed that *CRI4033* and *CG9203* interact using the endo siRNA pathway. In order for this pathway to run it is necessary to have RNA of the genes involved. Deletion mutants do not produce RNA and therefore there should be a change in transcript levels of the gene that this RNA would normally interact with if the endo siRNA pathway were allowed to progress. For this reason, deletion mutants were generated in order to answer if there was a phenotype associated with the lack of *CRI4033* and *CG9203* and to see if there is an interaction between *CRI4033* and *CG9203*. Because Czech *et al.* found *CG9203* was regulated through the endo siRNA, it is likely *CRI4033* is as well because they have a similar sequence.

Three strains carrying a P-element, $P\{SUPor-P\}tkv^{KG01923}$, $P\{lacW\}l(2)k01302^{k01302}$, and $P\{GSV6\}GSI7311$, inserted in *CRI4033*, were chosen to generate imprecise excision mutations. These P-elements were chosen because they are inserted in the *CRI4033* region and are not in the coding region of *tkv* (Figure 3.1). After isolation of flies carrying independent P-element excisions based on a white-eyed phenotype and establishment of strains, the P-element excision strains were characterized based on sequence and phenotype. For sequence characterization, PCR products were

first generated from the genomic DNA of the P-element excision strains and were compared in size to the product generated from the reference genome strain, 2057 (Adams *et al.* 2000). If the PCR products sizes were different from 2057, the region across the original site of the P-element insertion was sequenced.

15 independent excisions, one kept in duplicate, were generated from 35 single F1 males collected from the mutagenesis crosses for $P\{SUPor-P\}tkv^{KG01923}$ (Figure 2.1; Table 3.1). Of the 15 independent excisions, 8 were found to have precise excision of the P-element as PCR and sequencing results showed the sequence was wild-type when compared to 2057, while three that were not sequenced also appeared to have excisions that did not delete *CRI4033* as they produced wild-type sized PCR products (data not shown). Two excisions, 2B-A and 8A-A, showed PCR products larger than the wild-type size. Sequencing revealed that they still contained portions of P-element but the entire sequence of *CRI4033* was intact (Figure 3.2).

For P-element $P\{lacW\}l(2)k01302^{k01302}$, a total of 26 independent excisions were generated from 30 single F1 male mutagenesis crosses (Table 3.1). Two excision sublines were isolated from each F1 male, as there is a possibility that they are independent of each other and both sublines were characterized. All of the PCR products generated for these strains using 14403 4L and 4R and 10504 L and R primers were the equivalent size of 2057 except for 2A-A and 23-B, which did not produce products with 14403 8L and 8R and 14403 4L and 4R, respectively. This indicates that they were likely large deletions that had lost the primer annealing sites and that delete flanking gene sequences. Based on the wild-type size of all of the PCR products obtained (data not shown), no imprecise excisions were generated from excising this P-element.

For P-element *P{GSV6}GS17311*, 37 independent excisions were generated from 40 F1 male mutagenesis crosses. The genomic DNA from ten excision heterozygotes was pooled for each PCR and, among 4 PCRs using primers 14403 4L and 4R, only one showed a smaller PCR product in addition to the wild-type product. Subsequent PCR analysis identified the strain carrying the deletion, *CR14033*¹⁸, which was 3318 bp (Clark, personal communication). At the same time, a strain carrying a wild-type excision allele was kept to use as a control for genetic background, *CR14033*³ (D. Clark, personal communication) (Figure 3.3). Phenotypic and gene expression analyses were then conducted as part of this thesis research (Table 3.1).

The P-elements used for generating *CG9203* deletion mutations were *P{SUPor-P}CG9203*^{KG05829}, *P{XP}d04265*, and *P{GawB}NP0833* (Figure 3.4). *P{SUPor-P}CG9203*^{KG05829} has a *yellow* marker gene while *P{XP}d04265*, and *P{GawB}NP0833* have a *white* marker gene (Figure 2.2). Of the three P-elements in *CG9203*, only *P{SUPor-P}CG9203*^{KG05829} was found to produce excisions. The other two P-elements may have mutations in their inverted repeats, which would prevent them from excising.

19 independent excisions were generated from 33 single male mutagenesis crosses. Duplicates were kept for each independent excision if possible and both were sequenced. Sequencing results showed that 5 precise excisions, 18 partial excisions, 2 imprecise excisions, and an uncharacterized recessive lethal mutant were generated (Table 3.2). For 28B-D, a PCR product could not be generated consistently, even with the 14 kb-spanning primer pair 14125 7L and 7R, which amplified but did not yield as much product as 2057. The lower yield of product could be because in the stomach of the non-GFP larva chosen there could have been DNA of a GFP sibling that it consumed.

The GFP sibling would have had *FM7c* and that is what was being amplified. It is possible that the 28B-D excision is a large deletion affecting neighbouring genes that I was not able to amplify with primers as far apart as 14 kb. One imprecise excision, 24A-D, is a 1 kb deletion that did not delete the *CG9203* sequence that is similar to *CRI4033* (3' end). It deleted the 5' end of the *CG9203* (Figure 3.5A and B). Another strain, 20A-D, had a 9 bp deletion that is also not in the region of *CG9203* that has sequence similarity to *CRI4033* (data not shown).

In summary, for *CRI4033*, one deletion mutant was generated. For *CG9203*, none of the independent excisions led to a small deletion mutation of the entire region of *CG9203* containing the sequence similar in *CRI4033*. However, a putative large deletion >14kb (28B-D) and a small deletion of the 5' end of the gene (24A-D) were generated.

3.2 Genetic analysis of *CRI4033* mutations

Complementation tests are used to determine if two recessive loss-of-function mutations are alleles or, in other words, are affecting the same gene (Greenspan 2004). Two forms of complementation tests were used when looking at P-element excisions associated with *CRI4033*. One was using a *tkv* allele to determine if the excisions of *CRI4033* are disrupting *tkv* expression, as *CRI4033* is nested in a *tkv* intron and overlaps an exon of one of the *tkv* transcript isoforms, *tkv-RC* (Figure 3.1). It is important to be able to discern if any change in phenotype or transcript amount is associated with *CRI4033*, and not *tkv*, to gain an idea about the function and habits of *CRI4033*.

The other complementation test was with a deficiency of the *tkv* region to see if an additional recessive mutant phenotype would be uncovered that might be associated

with the pseudogene. In addition, since the chromosomes carrying P-element excisions could carry recessive second-site mutations in the *tkv* region that were not observed by PCR or sequencing (Greenspan 2004), looking at the phenotype of the excision allele over a deficiency would show whether that phenotype maps to *CR14033* region. One method for the generation of second site mutations is by a P-element jumping locally, which they have a tendency to do (Tower *et al.* 1993). Because of the local jumping, it is possible that they could insert and excise leaving behind a deletion in that site via imprecise excision. The complementation tests were carried out by crossing virgin females heterozygous for the P-element excision to males heterozygous for either *tkv*⁷, *Df(2L)BSC693* or *Df(2L)Exel6011*.

For this study, the areas of interest in terms of phenotypes associated with *tkv* mutations include both wing and thorax changes and as such these two areas were examined for each complementation test. Phenotypes seen in the wings are thickening and abnormal branching of the wing veins in the region of the intersection with the cross veins and near the ends of the longitudinal veins and a blister near the posterior cross vein (Szidony and Reuter 1988, Lindsley and Zimm 1992, Terracol and Lengyel 1994). Also, there can be a reduction in the length of L4 and L5 (Adachi-Yamada 1999). As for the thorax, some phenotypes associated with mutations in *tkv* include a midline furrow in the thorax (Terracol and Lengyel 1994).

Data from sections 3.2.1 and 3.2.2 are summarized in Table 3.3, data from sections 3.2.3 and 3.2.4 are summarized in Table 3.14, and data from sections 3.2.5 and 3.2.6 are summarized in Table 3.4.

3.2.1 Complementation analysis of the P-element $P\{SUPor-P\}tkv^{KG01923}$ insertion shows it disrupts tkv

Crossing the original P-element strain, 14403, carrying the $P\{SUPor-P\}tkv^{KG01923}$ insertion allele, to the tkv^7/CyO strain was done to determine if there was a phenotype associated with $CR14033$ due to the P-element insertion and to see if tkv was disrupted by its insertion (Figure 2.5A). Characterization of the original insertion would then allow me to determine if excisions I generated caused a reversion of the phenotype or a new one (Section 3.2.2).

In the adult progeny from this cross, wing vein patterns as well as thoracic phenotype were examined. The wings of the adult progeny displayed a tkv wing phenotype like that observed in tkv^7/tkv^1 adults as well as an additional thorax phenotype characteristic of severe tkv alleles (Terracol *et al.* 1994). The wing phenotype seen in both instances showed a reduction in veins L4 and L5, which is similar to that described previously by Adachi-Yamada and colleagues, only in this instance the crossvein between the two remained (1999). The thorax phenotype observed in $P\{SUPor-P\}tkv^{KG01923}/tkv^7$ adults was a furrow running from the dorsal anterior portion of the thorax to the posterior portion as described by Terracol *et al.* (1994) as well as displaced scutellar bristles.

When 14403 was crossed to two different strains carrying deficiencies of the tkv region, phenotypes emerged that were not seen in the cross with tkv^7/CyO . Both the male and female progeny heterozygous for the P chromosome and the deficiency chromosome, $Df(2L)Exel6011$, had an ectopic wing vein extending from the posterior crossvein, enlarged wing vein ends, and missing posterior scutellar bristles with extra anterior

scutellar bristles. A similar phenotype was seen with those progeny carrying *Df(2L)BSC693* with the exclusion of the ectopic vein extending from posterior crossvein but rather from the anterior crossvein (Appendix 2).

The negative control, *tkv*⁷/*Df*, with *Df* meaning either deficiency of thickveins, was lethal, as expected (Christensen *et al.* 2008, Ryder 2004).

The insertion of *P{SUPor-P}tkv*^{KG01923}/*tkv*⁷ into *CRI4033* caused disruption of *tkv* as evident by the phenotypes generated in the crosses. The difference in phenotypes seen between 14403 crossed to *tkv*⁷ versus either deficiency of *tkv* could be due to a disruption of another gene in the region or a disruption of *CRI4033*. In order to decipher if it is another gene or *CRI4033* that is causing the additional phenotypes, phenotypic analysis of a deletion mutant of *CRI4033* is needed.

3.2.2 Complementation analysis of the P-element *P{SUPor-P}tkv*^{KG01923} excisions

Complementation tests were performed on the P-element excision strains generated from excision of *P{SUPor-P}tkv*^{KG01923} to see if a recessive phenotype associated with *CRI4033* could be uncovered and to see if they were alleles of *tkv* and not *CRI4033*.

Of the 16 *P{SUPor-P}tkv*^{KG01923} excisions, 11 precise excisions complemented *tkv*⁷ while 2 of these, 25A-A and 29B-A, produced progeny where a few flies had a blunt bristle phenotype. These precise excisions when crossed to *Df(2L)BSC693* resulted in all of the *P*/Df(2L)BSC693* progeny with a wing and/or bristle phenotype consisting of thickened wing vein ends, an ectopic wing vein extending from posterior crossvein, and extra anterior scutellar bristles with or without missing posterior bristles. The phenotypes

shown are consistent with a *tkv* mutant. It is possible that these excisions actually have small deletions of *tkv* that were not detected on gel electrophoresis as they appeared to be wild-type size. In this study any excision that appeared as wild-type in size was not sequenced. Other possibilities for the *tkv* phenotype include a second site mutation near *tkv* and a rearrangement that is affecting *tkv*.

The two P-element partial excision alleles that still had some *P{SUPor-P}{tkv^{KG01923}}* sequence remaining, 2B-A and 8A-A, when crossed to *tkv⁷*, exhibited different phenotypes. 2B-A retained a *tkv* phenotype consisting of shortened wing veins, additional scutellar bristles, as well as a lack of more posterior scutellar bristles; 8A-A showed no mutant phenotype. Also, when both excisions were crossed to *Df(2L)BSC693* and *Df(2L)Exel6011*, 2B-A had a *tkv* phenotype with extra anterior scutellar bristles and a lack of the more posterior bristles along with widened wing veins ends, and an ectopic wing vein at the posterior crossvein. The 8A-A cross produced four flies with additional scutellar bristles and two with an ectopic wing vein extending posteriorly from the posterior crossvein when crossed to *Df(2L)BSC693* but no phenotype with *Df(2L)Exel6011* (Table 3.3).

From these results, it would appear as though *tkv* is disrupted in 2B-A, 25A-A and 29B-A. 8A-A is most likely a weak allele, which is why *tkv* phenotypes were not consistently seen. When sequenced, 8A-A still had P-element sequence and we know from complementation tests with the P-element insertion strain that the P-element disrupts *tkv* (Figure 3.2).

Two of the independent excision strains were lost before they could be characterized.

3.2.3 Complementation analysis of the P-element $P\{lacW\}l(2)k01302^{k01302}$ insertion

In order to determine if the insertion of the P-element, $P\{lacW\}l(2)k01302^{k01302}$, disrupts *tkv* expression or causes a recessive phenotype, complementation tests using $Df(2L)BSC693$ and tkv^7 were performed.

The insertion of $P\{lacW\}l(2)k01302^{k01302}$ does not have a recessive phenotype as it complemented $Df(2L)BSC693$. The positive control showed a small number of males with extra scutellar bristles and blunt scutellar bristles, which are most likely developmental abnormalities. The strain used for the negative control, tkv^7 , did not complement the deficiency as expected (Table 3.14B).

When the original P-element insertion strain, 10504, was crossed to tkv^7 , four females were found to have all or just blunt scutellar bristles, additional crossveins and/or ectopic wing veins. As a result, it fails to complement tkv^7 and therefore the insertion disrupts this gene (Table 3.14A).

3.2.4 Complementation analysis of the P-element $P\{lacW\}l(2)k01302^{k01302}$ excisions

All 47 excision chromosomes from $P\{lacW\}l(2)k01302^{k01302}$ appeared recessive lethal as no strains were homozygous viable. When crossed to tkv^7 , all but 7 excision alleles complemented (data not shown). The phenotypes seen in those that did not complement showed the posterior crossvein not extending the whole way from L4 to L5, an ectopic wing vein extending from L2 and extra scutellar bristles, which are all consistent with a mutation the *tkv*. The product size of these excision strains was determined using gel electrophoresis where wild-type product size was seen. Even

though this was the case, there is a possibility that a small deletion in *tkv* exists that does not decrease product size to the point where it is visible as decreased on gel electrophoresis and is only seen as a lack of complementation to *tkv* (Table 3.14A).

When crossed to *Df(2L)BSC693, P*/Df* was viable, as was the original P-element insertion over the deficiencies but the P-element excision strains were not homozygous viable. This demonstrated that the lack of homozygous viable flies in all excision strains was most likely due to a second site lethal mutation (Table 3.14B).

The progeny generated from a second complementation test using a deficiency of *tkv, Df(2L)Exel6011*, displayed a mutant phenotype, which was not seen in the positive control. These progeny had an ectopic wing vein extending from the posterior crossvein and blunt thoracic bristles or extra scutellar bristles. All but the blunt bristles are seen with *tkv* mutations (Table 3.14B).

Based on these results, it can be concluded that these excisions do not produce a deletion mutant of *CRI4033* and that the excisions produced disrupt *tkv*.

3.2.5 Complementation analysis of the P-element *P{GSV6}GS17311* insertion

To assess if the insertion of *P{GSV6}GS17311* disrupts *tkv* expression, the original P-element strain, 200-010, was crossed to *tkv¹*, which resulted in complementation. When 200-010 was crossed to *tkv⁷* there was no complementation (Table 3.4 A, B). The phenotypes seen with the lack of *tkv⁷* complementation were extra anterior scutellar bristles and a lack of posterior scutellar bristles, a small ectopic wing vein on the medial side of L2, and a thoracic furrow, like that seen with 14403 when crossed to *tkv⁷*, only in a milder form.

When crossed to the *Df(2L)Exel6011* and *Df(2L)BSC693* there was lack of complementation (Table 3.4 C and D). The phenotypes seen from the lack of complementation between *P{GSV6}GS17311* and the two deficiencies were additional anterior and lack of posterior scutellar bristles, except in one male, which was classified as wild-type.

Therefore, it appears that *P{GSV6}GS17311* disrupts *tkv* based on the cross with *tkv⁷*, which is a null allele (Horsfield *et al.* 1998) whereas *tkv¹* is a hypomorphic allele, a partial loss of function (Terracol and Lengyel 1994). A recessive phenotype present in the strain consisting of supernumerary anterior scutellar bristles, associated with *tkv*, and a lack of posterior scutellar bristles, which is not. This is a novel phenotype, which could be associated with disruption of *CRI4033* or another gene in this region.

3.2.6 Complementation analysis of the P-element *P{GSV6}GS17311* excisions

When *CRI4033¹⁸* was crossed to *tkv¹* and *tkv⁷*, it complemented showing that the deletion caused by an imprecise excision of the P-element does not disrupt *tkv* expression and therefore it is not an allele of *tkv* (Table 3.4 A and B). A couple of female progeny from *CRI4033³* did have one extra anterior scutellar bristle but because this exact phenotype was also seen in the negative control, *w¹¹¹⁸*, it was deemed not to be a result of lack of complementation with *tkv* but due to a developmental abnormality in this case.

When crossed to both deficiencies of *tkv*, *Df(2L)Exel6011* and *Df(2L)BSC693*, *CRI4033¹⁸* complemented just like the negative controls, *CRI4033³* and *w¹¹¹⁸* (Table 3.4 C and D). This means that there is no visible recessive phenotype associated with *CRI4033¹⁸*.

3.3 Sterility analysis of the *CR14033 P{GSV6}GS17311* insertion and excisions

Since Benjamin Loppin found *CG9203* mutants that were female sterile (personal communication), the sterility of *CR14033¹⁸* was tested (Table 3.5A). The females from the deletion mutants, *CR14033¹⁸*, as well as those from the wild-type excision, *CR14033³*, which was used as a negative control, were all fertile. The *CR14033¹⁸* males showed 18% sterility while the males of the negative control were all fertile. The original P-element strain *P{GSV6}GS17311* was also tested for sterility. Both the males and females from this strain showed some sterility. In the females, 3 of the 14 individual females crossed were sterile and 9 of the 15 individual males were sterile.

Using a Fisher's exact test, it is not possible to reject the null hypothesis that there is no relationship between the sterility of *CR14033¹⁸* and *CR14033³* with a p-value of 0.126. Using the same test to assess the results for the sterility analysis of *P{GSV6}GS17311* males, it is possible to reject the null hypothesis that there is no relationship between sterility and the P-element insertion into *CR14033* (p= 0.0000952) (Preacher and Briggs 2001).

3.4 Genetic analysis of *CG9203* mutations

3.4.1 Duplication mapping

Another form of complementation test, duplication mapping, was used to see if the lethal mutation generated by excision of *P{SUP^{or}-P}CG9203^{KG05829}*, 28B-D, was due

to the excision of the P-element specifically at *CG9203* rather than a second site lethal in another region. Duplication mapping must be employed for complementation mapping on the X chromosome since it is not possible to generate a male carrying a lethal mutation on its single X chromosome (Venken *et al.* 2010). Instead, one can determine if the mutation lies in the region of the duplication by seeing if the duplicated region of chromosome rescues the mutant phenotype, which in this case is the recessive lethality. *Dp(1;3)DC300* contains a 90,961bp fragment of sequence from the X chromosome that includes *CG9203* inserted on chromosome 3 (Venken *et al.* 2010). Duplication mapping was performed (Figure 2.5B) with 28B-D and *Dp(1;3)DC300* and the lethality was rescued by the duplication (Table 3.13). Therefore, the lethality of 28B-D is caused by an excision of the P-element and not a mutation outside of the region covered by the duplication.

3.4.2 Sterility phenotype of *CG9203* mutations

Because of the previously mentioned correspondence with Benjamin Loppin, I checked for sterility of 24A-D and *P{SUPor-P}CG9203^{KG05829}* homozygotes. Crosses were set up to test for sterility of individual virgin females and males from the aforementioned strains with the *w¹¹¹⁸* strain as a wild-type control. All of the 24 females homozygous for the P-element excision 24A-D were sterile when crossed to two *CantonS* males. Of the 11 24A-D males tested, 6 were sterile (55%). 10 single *P{SUPor-P}CG9203^{KG05829}* virgin females crossed to 2 *CantonS* males were all fertile while 63% of the single *P{SUPor-P}CG9203^{KG05829}* males crossed to *CantonS* virgin females were sterile. The crosses using single *w¹¹¹⁸* males and females showed a small amount of

sterility (20% in females and 30% in males) when crossed to two *CantonS* flies of the opposite sex (Table 3.5B). With the sterility seen in the males it was not possible to reject the null hypothesis that there is no relationship between the sterility seen between 24A-D and *w¹¹¹⁸* ($p=0.659$). The sterility seen in the females was found to be significant ($p= 0.00000248$), therefore it is possible to reject the null hypothesis.

3.5 *CRI4033* appears to be a pseudogene of *CG9203* and the 3' end of *CG9203* and *CRI4033* have sequence similarity in *Drosophila melanogaster*

Pseudogenes are often gene duplicates and they are detected in genome sequence because their nucleotide sequence is similar to the gene that they arose from. Czech *et al.* isolated testis-specific siRNA from a *tkv* intron corresponding to a portion of *CRI4033* with sequence similarity to *CG9203* and reported that the 3' regions of both genes were similar (Czech *et al.* 2008). To determine just how similar these sequences were to each other, a local alignment of *CG9203* and *CRI4033* cDNAs (MIP11944 and LD26673, respectively) was performed. From this alignment, it was determined that the 3' end of the two sequences aligned with 73% nucleotide identity over 41% of the pseudogene transcript (Figure 3.6). This entire region corresponds to the region that generates siRNA (Figure 1.3). The sequences of the siRNA-generating region of *CRI4033* and *CG9203* are different enough even with the extensive sequence similarity that they map unambiguously to one locus (Czech *et al.* 2008).

Pseudogenes carry several characteristics that cause them to differ from their parent gene. These differences result in the majority of transcribed pseudogenes not being translated into a protein. These changes include premature stop codons,

deletions/insertions, and frameshift mutations (Poliseno *et al.* 2010). I further characterized the relationship between *CRI4033* and *CG9203* by examining the amino acid sequence from the three forward reading frames of *CRI4033* for the aforementioned characteristics of a pseudogene. This analysis helps to further verify that *CRI4033* is a pseudogene of *CG9203*. The three reading frames were generated by submitting LD26673 to the showorf application from EMBOSS (Rice *et al.* 2000). The three forward reading frames were compared to the amino acid sequence of *CG9203*, which was obtained by submitting the MIP11944 to showorf. The majority of amino acid (aa) sequence of *CG9203*, which is similar to *CRI4033*, was identified when searching the three forward reading frames of *CRI4033*. The amino acid sequence of *CG9203* was found in two different reading frames of *CRI4033*, switching between the two reading frames twice, indicating two frameshift mutations along with two insertions of 13 aa and 6 aa, four deletions between 4 aa and 11 aa, and 3 nonsense mutations (Figure 3.7).

These results provide evidence that *CRI4033* is a transcribed pseudogene of *CG9203* because *CRI4033* carries the characteristics of a pseudogene, namely, insertions, deletions, frameshift and nonsense mutations while having a similar sequence to *CG9203* over 41% of its transcript. Its known transcription is based on cDNA isolation by the genome project (Rubin *et al.* 2000) and RNA-seq data in modENCODE (Graveley *et al.* 2011).

3.6 Orthologs for *CRI4033* were found in *D. sechellia* and *D. simulans*

The aforementioned technique used in *Drosophila melanogaster* to compare the sequence of *CRI4033* to *CG9203* was applied to the gene orthologs in two other species in the melanogaster subgroup, *D. sechellia* and *D. simulans*.

It would appear as though the sequence corresponding to LD26678 in *D. sechellia*, scaffold 5, is the pseudogene paralog for the MIP1944 ortholog, scaffold 73. When aligned, the pseudogene has frameshift mutations, deletions between 1 aa and 28 aa long, and insertions compared to the parent gene. There are also three premature stop codons as well as amino acid changes (Figure 3.8). These changes in scaffold 5 when compared to scaffold 73 are consistent with the definition of a pseudogene.

Like *D. sechellia*, the pseudogene ortholog in *D. simulans* found during the search with BLAST Stand Alone (Camacho *et al.* 2009) has the same features, premature stop codons, deletions, insertions, and amino acid changes, that provide evidence for it being a pseudogene of the *CG9203* ortholog in *D. simulans* (Figure 3.9).

3.7 The gene duplication event that gave rise to *CRI4033* appears to have occurred in the melanogaster subgroup

To determine when *CRI4033* first arose in the *Drosophila* lineage, a BLAST homology search with the nucleotide sequence of cDNA LD26673 using the Flybase BLAST tool was performed to identify similar sequence in the genomes of 12 *Drosophila* species (*D. simulans*, *D. sechellia*, *D. melanogaster*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis*, *D. grimshawi*) (*Drosophila* 12 genome consortium *et al.* 2007).

The BLAST report showed the two other species in the melanogaster subgroup, *D. simulans*, and *D. sechellia*, have *CR14033* orthologs close to the full length of *CR14033*. In the case of the *D. simulans* sequence (r1.4) on Flybase, the *CR14033* ortholog is on an unanchored sequence and is truncated with a tandem repeat and there is no sequence similar to it at the orthologous *tkv* locus. However, a second, more recently generated, version of the *D. simulans* genome was obtained (Hu *et al.* 2013) and was searched for a *CR14033* ortholog using BLAST Stand Alone from NCBI. The result of this BLAST search with LD26673 contradicted the previous finding of a lack of *CR14033* in the *tkv* intron of *D. simulans* (r1.4). In the Hu *et al.* (2013) sequence, a *CR14033* ortholog was located in the orthologous *tkv* locus on chromosome 2. These two results suggest that the older assembly is likely incomplete and that *D. simulans* does have a *CR14033* ortholog in the intron of a *tkv* ortholog, like *D. melanogaster*. The *D. sechellia* genome, like the second version of *D. simulans* genome and the *D. melanogaster* genome, also has a full length *CR14033* ortholog on chromosome 2 in the orthologous *tkv* locus (Figure 3.10). Two other species in the melanogaster group, *D. erecta* and *D. yakuba*, also had evidence of *CR14033*. However, these two species did not have a full-length ortholog in *tkv*. Only the 5' end of the pseudogene that does not correspond to *CG9203* was present in the intron of *tkv*.

Using the same method, the genomes of *Drosophila* species outside of the melanogaster subgroup were examined for a *CR14033* ortholog. *D. persimilis*, *D. pseudoobscura*, and *D. ananasssae* possess orthologous *CR14033* sequence in a *tkv* intron. However, the portion of the *CR14033* sequence in the *tkv* intron of these species

only had sequence similar to the 5' end of the pseudogene and not to *CG9203* (Figure 3.10).

3.8 *CRI4033* orthologs have more similarity to each other than to their respective *CG9203* paralogs

The 3' end of *Drosophila melanogaster CRI4033*, from nucleotide 5, 244, 216 to nucleotide 5, 243, 439, was the region from which siRNAs were generated (Okamura *et al.* 2008). The targeting of siRNAs to a transcript requires nearly perfect nucleotide sequence similarity (Golden *et al.* 2008). If siRNA targeting occurs between *CG9203* and *CRI4033* transcripts, then one would predict that the paralogs would be more similar to each other than their orthologs as there would be selection to preserve nucleotide sequence similarity. To address this prediction, the sequence from this region of the 3' end of *CRI4033* was extracted as a FASTA file from Flybase (dos Santos *et al.* 2014) and local sequence alignments to twelve *Drosophila* species' genomes (*D. simulans*, *D. sechellia*, *D. melanogaster*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis*, *D. grimshawi*) were generated using this sequence and the BLAST tool on Flybase with the low complexity filter off. The sequence identities for each hit were recorded. If there were multiple contiguous hits, the sequence identities were added together to determine the overall identity between the query and subject for that region. The nucleotide sequence for this region was acquired as a FASTA file from Flybase. These sequences were then used as BLAST queries to determine their identity to parent genes and orthologous sequences in other species. Using these sequences, more BLAST searches were conducted to determine percent

nucleotide identity between *Drosophila* sp. pseudogene sequences, parent gene sequences, and pseudogene and parent genes.

Contrary to the above prediction, the majority of the data shows that the region in the parent genes corresponding to the region of the pseudogene that produces endo siRNA is more similar to other parent gene sequences than to their paralogous pseudogene sequence (Tables 3.6 and 3.7). The same trend can be seen when looking at the pseudogene sequence. The pseudogene sequences between species were more similar than the pseudogene and parent gene sequence within a species (Table 3.7 and 3.8).

In the case of *D. erecta* and *D. yakuba*, the only sequence that had homology to this region of pseudogene was found on the X chromosome with no sequence corresponding to a second gene on any chromosome. Even though *D. erecta* and *D. yakuba* do not possess an ortholog of *CRI4033*, I was curious to see if the parent gene was still under the same constraints as species with both the pseudogene and parent gene and if their *CG9203* ortholog would still be similar. When using the *D. erecta* and *D. yakuba* parent gene sequence to search for similar parent gene sequence in other species, there was greater nucleotide sequence identity between *D. erecta* and *D. yakuba* than there was to other species. Even though this was the case, there was still a high percentage of nucleotide identity between these two species and the other *Drosophila* species. The parent gene in *D. erecta* and *D. yakuba* could be more similar to each other because they are more closely related to each other than they are to the other three species that it was compared to.

D. persimilis, *D. pseudoobscura*, and *D. ananasssae* also have the 5' end of *CRI4033* but, like *D. erecta* and *D. yakuba*, are missing the 3' end that is similar to

CG9203. The other species examined do not have evidence of *CR14033* in their *tkv* intron (Figure 3.10).

3.9 Mutant phenotypes were observed in females expressing a transgene causing RNAi for *CG9203* and *CR14033* but no mutant phenotypes were observed in the fertile males

RNAi is a technique used for gene specific knock down and is a useful method for studying gene function. I used this technique to study the function of *CG9203* and *CR14033*. This technique could not be used to look at effects of the pseudogene on parent gene expression as the method by which this test works is the pathway likely to be used by the pseudogene to affect parent gene expression. Therefore, I could only use this technique to potentially give insight into a phenotype associated with loss of *CR14033* and *CG9203* expression. Each strain contains a transgene (Bellen *et al.* 2007) with inverted repeats derived from a portion of cDNA that produces a hairpin RNA. The transgene is expressed under the control of a UAS enhancer, which is induced when bound to the GAL4 transcription factor. Thus the transgene is expressed in the progeny of a cross between a strain carrying the RNAi transgene and a strain carrying a GAL4 driver. This hairpin RNA then enters a small RNA pathway and produces small pieces of RNA complementary to either the pseudogene or parent gene, depending on the strain, and would then cause a decrease in transcript. The RNAi strains, w^{1118} ; $P\{GDI2463\}v23052$ and w^{1118} ; $P\{GDI4153\}v29075$, were crossed to a *tubP-GAL4* strain to induce expression of the transgene to produce a hairpin RNA. Progeny were examined for phenotypic changes. Males were individually crossed to virgins with the *tubP-GAL4* driver but no transgene as a wild-type control (Figure 2.8). It is important to note that the

promoter used for both RNAi transgenes is the *hsp70* promoter (Dietzl *et al.* 2007).

Since this promoter is not expressed in the female germline, it was not possible to look at female sterility (Rorth 1998).

Phenotypic characterization of the progeny revealed no visible mutant phenotypes in the male progeny expressing the RNAi construct but the females had an ectopic wing vein protruding from the cross vein between L4 and L5 that was not seen in the control cross. Also, found in both the experimental and control progeny was a wing vein close to the lateral exterior margin of the wings. This small additional vein was deemed to be a result of the *tubP-GAL4* chromosome since it was seen in the control and thus not a result of a reduction in *CG9203* and *CRI4033* expression (Table 3.9).

As for sterility all three males surveyed for *P{GD12463}v23052/tubP-GAL4* and *P{GD14153}v29075/tubP-GAL4* were found to be sterile but one out of each of the two *w¹¹¹⁸/tubP-GAL4* for the control were also sterile (Table 3.10). The sample size is small and in order to elucidate if the sterility was by chance or due to RNAi, this experiment would have to be repeated with a larger sample size.

3.10 Over expressing *CG9203* or *CRI4033* does not result in a mutant phenotype nor does it cause sterility

3.10.1 Phenotypic characterization

Adding and expressing a cDNA copy of *CG9203* or *CRI4033* was used in this study to see if there would be a visible mutant phenotype produced as well as to see if there would be a change in the expression level of the paralogous gene as well as *tkv* in

those flies with the additional copy of *CRI4033*. The change in the expression level will be discussed in section 3.11.

The majority of the progeny that were *UAS-cDNA/tubP-GAL4* and *w¹¹¹⁸/tubP-GAL4* had a wild-type thorax, wing vein, and scutellar bristle pattern but there were exceptions (Table 3.11). In flies expressing the cDNA copy of *CG9203* and *CRI4033* in either orientation as well as in the wild-type control cross using *w¹¹¹⁸*, there were some wing vein and scutellar bristle abnormalities (Table 3.11). Additional scutellar bristles were observed in some cases as well as an additional wing vein found on the lateral margin of the wing and a protrusion off the crossvein between L4 and L5. Because these abnormalities were seen in both the experimental and control crosses, they cannot be attributed to expression of the cDNAs.

3.10.2 Assessment of Sterility

Testing for sterility of the transgenic flies with the UAS transgene for a cDNA copy of *CG9203* and *CRI4033* was performed to see what the effects of additional transcript for *CG9203* or *CRI4033* would be on the germ line. This aspect was inspected due to personal communication with Dr. Benjamin Loppin who believed he had a homozygous mutant of *CG9203* that showed female sterility.

Expression of the UAS transgene for either *CG9203* or *CRI4033* in the sense or antisense orientation was induced in progeny by crossing virgin females carrying the cDNA transgene to males carrying the *tubP-GAL4* chromosome. When virgin females and males from the pseudogene (2-7A and 2-24A) and parent gene (3-31A and 3-38B) in the antisense direction were crossed to *w¹¹¹⁸* all were fertile (Table 3.12B and C). This

was also seen with the negative control, w^{1118} , as well as the virgin females and males for the transgene of the pseudogene (1-7A) and parent gene (4-2A) in the sense direction. The females for the other strain with the pseudogene transgene in the sense direction (1-8A) were all fertile, as well (Table 3.12A and D).

The induced expression of the parent gene in the sense direction (4-42B) led to one virgin female being sterile while the remaining 19 were fertile. The male progeny for this strain as well as the males from 1-8A were all fertile but each cross had one replicate that died before mating (Table 3.12D).

Based on these results, overexpression of the sense or antisense transcripts for *CR14033* and *CG9203* does not cause sterility.

3.11 Determining the relative amounts of *CR14033* and *CG9203* transcripts under different conditions using qPCR

3.11.1 Mutation of *Dcr2* leads to increased *CG9203* and *CR14033* transcript in testes

In previous studies, an increase in transcript levels for *CG9203* and *CR14033* were found in *Dcr2* mutant testes (Czech *et al.* 2008 and Marques *et al.* 2010) but no significant change in *tkv* had been detected (Marques *et al.* 2010). With *loqs* mutant testes, levels comparable to wild-type were seen for the three genes of interest for this study (Marques *et al.* 2010). To see if the effects seen were *Dcr2*- or *loqs*-dependent, and if it was possible to reproduce these results in testes from *Dcr2* and *loqs* mutants, *Dcr2*^{L811fsX} and *loqs*^{f00791} mutants, respectively, were examined for changes in the expression levels of *CG9203*, *CR14033*, and *tkv*.

Testes from *Dcr2* mutants showed an increase in intact transcript levels ranging from 2.8-3.8-fold for *CG9203*, with only one replicate being significant. Intact transcript also increased in *CRI4033*, which ranged from 3.7 to 5.3, with two replicates being significant. Like the results from previous studies, there was no significant change observed in *tkv* transcript level (Figure 3.11).

Testes from *loqs* mutants did not show a consistent trend in fold change. For *CG9203*, it ranged from 0.403- 1.676, *CRI4033* ranged from 0.86 to 1.883, while *tkv* ranged from 0.703- 1.8 (Figure 3.11).

From these results I conclude that *Dcr2* does indeed have an effect on *CG9203* and *CRI4033* while any involvement from *loqs* is subtle to non-existent. These results are consistent with previously published work (Czech *et al.* 2008 and Marques *et al.* 2010)

3.11.2 No consistent change was seen in *CG9203*, *CRI4033*, and *tkv* transcript in larval brains and discs

Okamura *et al.* observed abundant amounts of siRNAs from the *tkv* region in mass isolated imaginal discs and brains as well as testes (2008). To see if there was an effect on *CG9203* and *CRI4033* caused by knocking out *Dcr2*, qPCR was performed on larval brains and their associated imaginal discs from 3rd instar larvae.

CG9203, *CRI4033*, and *tkv* transcripts were all measured. In the first replicate, less transcript was seen in all three genes but in the second replicate an increase in the transcript of *CG9203* and *CRI4033* was seen and there was a decrease seen in *tkv* transcript (Figure 3.12). These results are very inconsistent. More work is necessary to

elucidate if there is an effect on the genes of interest in larval brains and discs with an absence of *Dcr2*.

3.11.3 No consistent results were seen between two independent strains expressing increased amounts of *CR14033*

Two independent insertions strains that possess an additional copy of the pseudogene in the sense orientation were used to observe the effects of having additional amounts of the pseudogene transcript on *CG9203* and *tkv* in testes and imaginal discs and brains. With an increased level of *CR14033* sense transcript, one could expect to see an increased amount of transcript of *CG9203* if *CR14033* acts as an mRNA decoy or less *CG9203* if there is interaction via the endo siRNA pathway. Less *tkv* transcript is predicted if *CR14033* is transcribed at the same time as *tkv* leading to dsRNA formation and cleavage by the endo siRNA pathway.

The data obtained for the testes from both strains were at odds with each other. With the first strain, 1-7A, a decrease in transcript was observed for *CG9203* ranging from 0.495- 0.936. *CR10433* went up in expression level with 2.943, 4.047, and 4.344. The latter two increases were significant. *tkv* was inconsistent showing a downward trend in one replication, 0.703, and an upward trend in the other two, 1.049 and 1.833, none of which was significant (Figure 3.13).

For 1-8A, a significant increase in intact *CR14033* transcript was again seen ranging from 3.22 to 7.085. However, the results for *CG9203* are less clear, with one replication showing decrease in intact transcript level at 0.752 and two replications

showing a small increase at 1.15 and 2.03. None of these results was significant. There was no real fold change observed in *tkv* (Figure 3.13).

The increase in *CR14033* transcript indicates that transcription of the transgene was induced to approximately the same level in both 1-7A and 1-8A. The extra transcript appears to have no effect on *tkv* while there may be a subtle reduction of *CG9203* transcript.

Two independent insertion strains with a transgene for an antisense transcript of the pseudogene were examined for fold changes in levels of *CG9203* and *tkv*. With the antisense pseudogene sequence, it is predicted that there will be less *CG9203* transcript as it is complementary to *CG9203* and can lead double stranded RNA formation. The double stranded RNA would be processed via the endo siRNA pathway and lead to increased endo siRNA fragments that can then lead to processing of *CG9203* mRNA using the RISC complex further decreasing the amount of transcript. For *tkv*, it is predicted that there would not be a change in transcript level.

The response of *CG9203* in the two strains differed. In 2-24A, a significant decrease in *CG9203* transcript was seen, 0.155-0.876, but in 2-7A testes, there was no significant change in *CG9203* transcript with expression levels of 0.696, 1.098, and 0.889. In 2-7A, *CR14033* transcript levels significantly increased (2.770, 1.592, 1.718), showing that the *CR14033* cDNA transgene was being transcribed. For 2-24A the *CR14033* and *tkv* expression levels were inconsistent and not significant (Figure 3.13). *tkv* expression in two replicates for 2-7A both increased, 1.512 and 1.284, with 1.512 being significant.

The amount of *CRI4033* transcript increased showing that the *CRI4033* cDNA transgene was being transcribed. With this increase in transcript an increased amount of *CRI4033* transcript in the antisense direction appears to have an affect on *tkv* in one strain and a decrease in *CG9203* in the other strain tested. The decrease in *CG9203* is what would be expected with increased antisense *CRI4033* transcript if regulation was performed via the endo siRNA pathway. The increase in *tkv* was unexpected and reasons for this will be considered in the discussion. Also, the possible reasons for the differences between the two strains are considered in the discussion.

3.11.4 No consistent results were seen between two independent strains expressing increased amounts of *CG9203* sense and antisense transcripts

In an attempt to determine if there is an interaction between *CG9203* and *CRI4033* in testes, two independent strains expressing *CG9203* sense and antisense transgenes were examined by looking for changes in *CRI4033* transcript amounts when there is an increased amount of *CG9203* in the sense and antisense orientation.

With increased amounts of *CG9203* in the sense direction, you would expect to see no change to an increase in *CRI4033* if it is acting as an mRNA decoy or a decrease if they interact through the endo siRNA pathway. With increased amounts of *CG9203* in the antisense direction you would expect to see decreased amounts of *CRI4033* due to RNA silencing via the endo siRNA pathway.

For one independent strain expressing an antisense cDNA transcript of *CG9203*, 3-31A, showed inconsistent expression of *CG9203* with one replicate decreasing, 0.558 fold, and two replicates increasing with a fold change of 2.801 and 7.526, as expected.

All of these fold changes were significant. In all three replications, *CRI4033* transcripts decreased, 0.689, 0.907, and 0.048 fold, but none was significant (Figure 3.14).

With the other independent strain, 3-38B, only two replicates were obtained for *CG9203* transcript. Both had an increase in fold change for *CG9203*, 1.500 and 2.667, showing that the cDNA transgene was being transcribed but neither was shown to be significant. The three replicates of *CRI4033* had mixed results with the fold changes, one increased slightly, 1.116, while the other two decreased, 0.106 and 0.764, but again neither was significant (Figure 3.14). The decrease in fold change for the first replicate of *CRI4033* (0.106) does not have a replicate of *CG9203* so it is not possible to know if there was a corresponding increase in *CG9203* that could be responsible for this decrease.

The strains with the sense *CG9203* UAS transgene had the expression significantly induced in all cases, as expected. For strain 4-2A, only two replicates were attained but in both cases *CG9203* was increased in the testes but the results for *CRI4033* were inconsistent. One showed a significant fold change decrease at 0.238, while the other one showed a small increase in *CRI4033* intact transcript, 1.356, but was not significant. For strain 4-42B, three replicates were attained. Two showed a significant decrease in *CRI4033* transcript amount, 0.531 and 0.838 fold. The third replicate also showed a decrease in transcript level but it was not seen to be significant, 0.921 fold. All three replicates had an increase in *CG9203* (Figure 3.14).

In summary, it would appear as though an increase in *CG9203* in the sense direction leads to decreased amounts of *CRI4033* and increased amounts of *CG9203* in the antisense direction did not yield changes in *CRI4033* that were found to be significant. The decrease in *CRI4033* when *CG9203* is over expressed in the sense

direction could be due to increased amounts of RISC complexes formed from the dsRNA is made from annealing with the *tkv* intron. This would only occur if the *tkv* intron is still around at the time of *CG9203* expression. The RISC complexes would then cleave complementary mRNA, *CR14033*, leading to a decrease the amount of *CR14033* transcript.

3.11.5 The deletion of *CR14033* appears to affect *CG9203* and *tkv* expression

In the *CR14033* deletion strain, *CR14033*¹⁸, the amount of *CG9203* and *tkv* transcript was measured. The expected result was for an increase in *CG9203* transcript because there would be no interaction with *CR14033* via the endo siRNA pathway or a decrease in *CG9203* if *CR14033* to act as an mRNA decoy. Two of the three replicates showed an increase in *CG9203* transcript, 2.446 and 2.341 fold, while *tkv* transcript decreased, 0.445 and 0.631 fold. The third replicate had no change in *CG9203* or *tkv*, 1.016 and 1.101 fold, respectively (Figure 3.15).

The lack of *CR14033* has an effect of *CG9203* and *tkv*. The decrease in *tkv* was an unexpected result. It could reflect an interaction between *tkv* and *CR14033* that was not evident in the overexpression experiments or other experiments in different laboratories. The reason for the decrease will be addressed in the discussion. The increase in *CG9203* was as expected if the interaction occurred via the endo siRNA pathway.

Table 3.1 Summary of sequence characteristics for P-element excision strains generated from excising $P\{SUPor-P\}tkv^{KG01923}$, $P\{lacW\}l(2)k01302^{k01302}$, $P\{GSV6\}GS17311$

The number of excisions that were precise, partial, and imprecise are shown for each P-element insertion strain, in addition to the number lost before sequence characterization. The excisions generated for $P\{GSV6\}GS17311$ were generated by Denise Clark and then were characterized with respect to phenotype and gene expression in this thesis.

P-element	Precise excisions (#)	Partial excisions (#)	Imprecise excisions (#)	Lost (#)	Total (#)
<i>P{SUP_{or}-P}tkv^{KG01923}</i>	11	2	0	3	16
<i>P{lacW}l(2)k01302^{k01302}</i>	44	0	2	1	47
<i>P{GSV6}GS17311</i>	≥4*	n/a	1	0	37

*only 5 excisions were characterized by DNA sequencing (D. Clark, personal communication).

Table 3.2 Summary of sequence characteristics for the P-element excision strain generated excisions *P{SUP^{or}-P}CG9203^{KG05829}*

The numbers of strains that generated precise, partial, unknown, and imprecise excisions are shown for the P-element insertion strain.

Precise excisions (#)	Partial excisions (#)	Imprecise excisions (#)	Unknown excisions (#)	Lost before characterization (#)	Total (#)
5	18	2	1	0	26

Table 3.3 Summary of the complementation tests performed on the 2B-A and 8A-A partial excisions of *P{lacW}l(2)k01302^{k01302}* and the original P-element insertion strain, 14403, using *tkv⁷*, *Df(2L)BSC693*, and *Df(2L)Exel6011*

The results for the complementation tests with *tkv⁷* (A), *Df(2L)BSC693* (B), and *Df(2L)Exel6011* (C) are shown for 2B-A, 8A-A, and 14403. In each case wing vein pattern, thorax appearance, and scutellar bristles of males (M) and females (F) were examined for phenotypic change from wild-type. Any change from wild-type seen in any or all of these three parts resulted in that fly being scored as mutant. The double columns for each genotype/phenotype class denote replicates 1 and 2.

A.

Strain	<i>CyO</i>		<i>P*/tkv⁷</i>		<i>P*/tkv⁷</i>		<i>P*/tkv⁷</i>		<i>P*/tkv⁷</i>		Complement (+) or (-)	
			WT	F	WT	M	mut	F	mut	M		
2B-A	38	55	12	16	0	5	0	10	3	17	(-)	(-)
8A-A	42	23	19	4	19	4	0	0	0	0	(+)	(+)
14403	79	97	0	0	0	1	13	10	15	24	(-)	(-)

B.

Strain	<i>CyO</i>		<i>P*/df</i>		<i>P*/df</i>		<i>P*/df</i>		<i>P*/df</i>		Complement (+) or (-)	
			WT	F	WT	M	mut	F	mut	M		
2B-A	51	14	0	0	0	0	17	25	20	4	(-)	(-)
8A-A	50	31	25	11	26	7	0	3	0	1	(-)	(-)
14403	43	69	9	0	3	0	26	9	24	9	(-)	(-)

C.

Strain	<i>CyO</i>	<i>P*/df</i> WT		<i>P*/df</i> mut		Complement (+) or (-)
		F	M	F	M	
2B-A	16	0	0	25	12	(-)
8A-A	21	7	8	0	0	(+)
14403	49	0	1	8	8	(-)

Table 3.4 Summary of the complementation tests performed on the excisions of *P{GSV6}GS17311* and the original P-element insertion strain, 200-010, using, *tkv¹*, *tkv⁷*, *Df(2L)BSC693*, and *Df(2L)Exel6011*

The results for the complementation tests with *tkv¹* (A), *tkv⁷* (B), *Df(2L)BSC693* (C), and *Df(2L)Exel6011* (D) are shown for the deletion *CRI4033¹⁸*, *CRI4033³*, and 200-010. In each case wing vein pattern, thorax appearance, and scutellar bristles were examined for phenotypic change from wild-type. Table E shows the results of the negative controls used in each complementation test.

A.

Strain	<i>P*/tkv¹</i> WT F	<i>P*/tkv¹</i> WT M	<i>P*/tkv¹</i> mut F	<i>P*/tkv¹</i> mut M	Complement (+) or (-)
<i>CR14033³</i>	*12	20	0	0	(+)
<i>CR1033¹⁸</i>	43	36	0	0	(+)
200-010	29	13	0	0	(+)
<i>w¹¹¹⁸</i>	*30	38	0	0	(+)

*1 *CR14033³* female with extra bristle, 1 *w¹¹¹⁸* female with extra bristles

B.

Strain	<i>P*/tkv⁷</i> WT F	<i>P*/tkv⁷</i> WT M	<i>P*/tkv⁷</i> mut F	<i>P*/tkv⁷</i> mut M	Complement (+) or (-)
<i>CR14033³</i>	*22	15	0	0	(+)
<i>CR1033¹⁸</i>	17	13	0	0	(+)
200-010	0	0	3	6	(-)
<i>w¹¹¹⁸</i>	*40	*57	0	0	(+)

*1 *CR14033³* female with extra bristle, 3 *w¹¹¹⁸* females and 1 male with extra bristles

C.

Strain	<i>P*/Df</i> WT F	<i>P*/Df</i> WT M	<i>P*/Df</i> mut F	<i>P*/Df</i> mut M	Complement (+) or (-)
<i>CR14033³</i>	45	29	0	0	(+)
<i>CR1033¹⁸</i>	35	18	0	0	(+)
200-010	0	0	6	5	(-)
<i>w¹¹¹⁸</i>	34	8	0	0	(+)

D.

Strain	<i>P*/Df</i> WT F	<i>P*/Df</i> WT M	<i>P*/Df</i> mut F	<i>P*/Df</i> mut M	Complement (+) or (-)
<i>CR14033³</i>	29	13	0	0	(+)
<i>CR1033¹⁸</i>	14	18	0	0	(+)
200-010	0	1	5	2	(-)
<i>w¹¹¹⁸</i>	20	20	0	0	(+)

E.

Cross	<i>CyO</i>	<i>P*/df</i> WT F	<i>P*/df</i> WT M	<i>P*/df</i> mut F	<i>P*/df</i> mut M	Complement (+) or (-)
<i>tkv⁷</i> x <i>tkv¹</i>	61	0	0	12	14	(-)
<i>tkv¹</i> x Df(1)	32	1	5	4	15	(-)
<i>tkv¹</i> x Df(2)	60	0	0	11	12	(-)
<i>tkv⁷</i> x Df(1)	94	0	0	0	0	(-)
<i>tkv⁷</i> x Df(2)	42	0	0	0	0	(-)

Table 3.5 Sterility of male and female *CRI4033* mutants *CRI4033*¹⁸ and *P{GSV6}GS17311*; and *CG9203* mutants, 24A-D and *P{SUPor-P}CG9203*^{KG05829} as homozygotes

Single males or females homozygous for the P-element deletion or insertion, *CRI4033*¹⁸ and *P{GSV6}GS17311* (A) and 24A-D and *P{SUPor-P}CG9203*^{KG05829}, were crossed to *w*¹¹¹⁸ females or males, respectively, to test for sterility. Absence of at least one first instar larva by day 10 was used to denote sterility whereas the presence of these larvae or further life stages denoted fertility. The number of adult males and females that were sterile or fertile are shown.

A.

Strain	Sterile	Fertile	Total
<i>CR14033</i> ¹⁸ females	0	9	9
<i>CR14033</i> ¹⁸ males	2	9	11
<i>CR14033</i> ³ females	0	12	12
<i>CR14033</i> ³ males	0	19	19
<i>P{GSV6}GS17311</i> females	3	11	14
<i>P{GSV6}GS17311</i> males	9	6	15

B.

Strain	Sterile	Fertile	Total
24A-D females	24	0	24
24A-D males	5	6	11
<i>P{SUPor-P}CG9203</i> ^{KG05829} females	0	10	10
<i>P{SUPor-P}CG9203</i> ^{KG05829} males	5	3	8
<i>w</i> ¹¹¹⁸ females	2	8	10
<i>w</i> ¹¹¹⁸ males	3	7	10

Table 3.6 Comparison between the region that produces siRNA from *CRI4033* in orthologous parent genes

The query sequence for each species was derived from searching the genomes with the region of *CRI4033* that produces siRNA. Each sequence was extracted and a BLAST homology search was performed to determine the percent nucleotide identity between these sequences for the hits from other species. The percent nucleotide identity was derived by dividing the number of nucleotides in the hit that matched the query by the number of nucleotides in the query.

Query Hit	<i>D. mel</i> parent gene	<i>D. sec</i> parent gene	<i>D. sim</i> parent gene	<i>D. sim</i> V2 parent gene	<i>D. yak</i> parent gene	<i>D. ere</i> parent gene
<i>D. mel</i> parent gene		652/723= 90.2%	637/723= 88.1%	671/732= 91.7%	503/599= 84.0%	563/675= 83.4%
<i>D. sec</i> parent gene	599/660= 90.8%		683/743= 91.9%	744/764= 97.4%	487/571= 85.3%	560/675= 83.0%
<i>D. sim</i> parent gene	584/660= 88.5%	672/731= 86.0%			470/562= 83.6%	553/675= 81.9%
<i>D. sim</i> V2 parent gene	511/551= 92%	532/537= 99%			569/669= 85%	548/631= 86.8%
<i>D. yak</i> parent gene	509/599= 85.0%	493/571= 86.3%	476/562= 84.7%	575/669= 85.9%		670/738= 90.8%
<i>D. ere</i> parent gene	563/675= 83.4%	560/675= 83.0%	553/675= 81.9%	571/675= 84.6%	670/738= 90.8%	

Table 3.7 Comparison between the region that produces siRNA in orthologous and paralogous parent genes and pseudogenes

The query pseudogene sequence for each species was derived from searching the genomes with the region of *CRI4033* that produces siRNA. A BLAST homology search was performed to determine the percent nucleotide identity between these sequences and the orthologous and paralogous parent gene hits. *D. erecta* and *D. yakuba* were included even though they do not have a *CRI4033* paralog because I was curious to see if the pseudogenes were under the same selection pressure as *CG9203* orthologs in species without *CRI4033* orthologs. The percent nucleotide identity was derived by dividing the number of nucleotides in the hit that matched the query by the number of nucleotides in the query. Grey boxes indicate comparisons between paralogs. There are two empty boxes because the two versions of *D. simulans* were not compared to each other.

Query Hit	<i>D. mel</i> Pseudogene	<i>D. sec</i> Pseudogene	<i>D. sim</i> pseudogene	<i>D. sim</i> V2 pseudogene
<i>D. mel</i> parent gene	413/480= 86%	420/504= 83.0%	190/222= 85.6%	471/496= 94.9%
<i>D. sec</i> parent gene	405/471= 85.0%	489/586= 83.4%	211/239= 88.3%	625/728= 85.9%
<i>D. sim</i> parent gene	387/437= 88.6%	332/386= 86%	190/224= 84.8%	
	218/262= 83.2%	213/263= 81%	189/214= 88.3%	
<i>D. sim</i> V2 parent gene	516/568= 90.8%	476/538= 88.5%		612/686= 89.2%
<i>D. yak</i> parent gene	391/435= 89.9%	447/530= 84.3%	179/211= 84.8%	565/653= 86.5%
<i>D. ere</i> parent gene	482/563= 85.6%	451/538= 83.8%	179/214= 83.6%	594/684= 86.8%

Table 3.8. Comparison between the region that produces siRNA in orthologous pseudogenes

A BLAST homology search was performed to determine the percent nucleotide identity between orthologous pseudogenes. The percent nucleotide identity was derived by dividing the number of nucleotides in the hit that matched the query by the number of nucleotides in the query. There are empty boxes because the pseudogenes in each species were not compared to themselves.

Query Hit	<i>D. mel</i> Pseudogene	<i>D. sec</i> pseudogene	<i>D. sim</i> pseudogene	<i>D. sim</i> V2 pseudogene
<i>D. mel</i> pseudogene		560/613= 91.4%	140/152= 92.1%	1357/1476= 91.9%
<i>D. sec</i> pseudogene	541/594= 91.1%		197/210= 93.8%	650/685= 94.9%
<i>D. sim</i> pseudogene	236/277= 85.2%	255/275= 92.7%		
	225/267= 84.3%			
<i>D. sim</i> V2 pseudogene	771/885= 87%	659/682= 96.6%		

Table 3.9 Summary of phenotypic characterization of RNAi

Phenotypic characterization of activated RNAi strains targeting *CRI4033* (*P{GD12463}v23052*) and *CG9203* (*P{GD14153}v29-75*) was performed with *w¹¹¹⁸* crossed to *tubP-GAL4/TM6B* as a negative “no RNAi” control so that any phenotype seen in the experimental crosses was as a result of RNAi and not a product of the *tubP-GAL4* chromosome. The wing veins, thorax and scutellar bristles were examined for phenotypic changes. The mutant phenotype seen in females expressing the UAS transgene was an ectopic wing vein protruding from the crossvein between L4 and L5. An additional change to the wing was seen, a small extra wing vein located close to the lateral exterior boarder of the wing, in the experimental crosses but these were also seen in the control cross and deemed to be a consequence of *tubP-GAL4* chromosome. The crosses were performed in replicate, shown here by the two columns under each genotype.

Progeny genotype	<i>P{GD14153}v29075</i>		<i>P{GD12463}v23052</i>		<i>w¹¹¹⁸</i>	
<i>UAS/tubP-GAL4</i> females WT	26	17	50	39	16	8
<i>UAS/tubP-GAL4</i> males WT	44	16	56	47	8	5
<i>UAS/tubP-GAL4</i> females mut	11	10	20	15	0	0
<i>UAS/tubP-GAL4</i> males mut	0	0	0	0	0	0
Total	157	89	251	188	43	18

Table 3.10 Male sterility in RNAi screen

The sterility of the male progeny (+/+ ; *UAS/tubGAL4*) was examined with *w¹¹¹⁸/Y* ; *tubGAL4/+* males as a negative control. Three crosses to *UAS* ; + ; *TM6B* virgin females were performed for each.

	<i>P{GD14153}v29075</i> (<i>CG9203</i> UAS- dsRNA transgene)		<i>P{GD12463}v23052</i> (<i>CR14033</i> UAS dsRNA transgene)		<i>w¹¹¹⁸</i>	
Fertile	3	3	3	3	2	2
Sterile	0	0	0	0	1	1

Table 3.11 Over expressing *CG9203* and *CR14033* did not produce a mutant phenotype

Transgenic flies with a UAS transgene consisting of a cDNA copy of *CR14033* or *CG9203* in the forward or reverse orientation had the expression of this transgene induced by exposure to *tubP-GAL4*. Phenotypic characterization of the wing veins, thorax as well as scutellar bristles was performed. Flies were categorized based on genotype and presence (mut) or absence (WT) of mutant phenotype. Two constructs of the sense (A) and the antisense orientation (B) for the cDNA corresponding to *CR14033* and two constructs of the forward orientation (C) and reverse orientation (D) for the cDNA corresponding to *CG9203* were examined. *w¹¹¹⁸* was used as a negative control (E).

A

Progeny Genotype	1-7A (rep1)	1-7A (rep2)	1-8A (rep1)	1-8A (rep2)
<i>UAS/TM6B</i> Females	13	30	27	42
<i>UAS/TM6B</i> Males	16	28	30	36
<i>UAS/tubP-GAL4</i> Females WT	15	23	35	20
<i>UAS/tubP-GAL4</i> Males WT	19	30	42	29
<i>UAS/tubP-GAL4</i> Females mut	0	7	7	5
<i>UAS/tubP-GAL4</i> Males mut	2	4	1	3
Total	65	122	142	135

B

Progeny Genotype	2-7A (rep1)	2-7A (rep2)	2-24A (rep1)	2-24A (rep2)
<i>UAS/TM6B</i> Females	11	38	60	42
<i>UAS/TM6B</i> Males	13	41	38	57
<i>UAS/tubP-GAL4</i> Females WT	12	20	35	25
<i>UAS/tubP-GAL4</i> Males WT	12	55	30	40
<i>UAS/tubP-GAL4</i> Females mut	0	8	31	25
<i>UAS/tubP-GAL4</i> Males mut	1	5	11	26
Total	49	167	205	215

C

Progeny Genotype	3-31A (rep1)	3-31A (rep2)	3-38B (rep1)	3-38B (rep2)
<i>UAS/TM6B</i> Females	55	25	25	40
<i>UAS/TM6B</i> Males	55	35	25	56
<i>UAS/tubP-GAL4</i> Females WT	34	27	16	21
<i>UAS/tubP-GAL4</i> Males WT	56	29	14	34
<i>UAS/tubP-GAL4</i> Females mut	12	1	5	20
<i>UAS/tubP-GAL4</i> Males mut	1	0	2	9
Total	213	117	87	180

D

Progeny Genotype	4-2A (rep1)	4-2A (rep2)	4-42B (rep1)	4-42B (rep2)
<i>UAS/TM6B</i> Females	22	11	30	20
<i>UAS/TM6B</i> Males	25	10	15	29
<i>UAS/tubP-GAL4</i> Females WT	9	7	24	12
<i>UAS/tubP-GAL4</i> Males WT	17	5	13	27
<i>UAS/tubP-GAL4</i> Females mut	4	4	9	7
<i>UAS/tubP-GAL4</i> Males mut	0	2	0	0
Total	77	39	91	95

E

Progeny Genotype	Rep 1	Rep2
<i>w¹¹¹⁸/TM6B</i> Females	46	40
<i>w¹¹¹⁸/TM6B</i> Males	59	27
<i>w¹¹¹⁸/tubP-GAL4</i> Females WT	42	8
<i>w¹¹¹⁸/tubP-GAL4</i> Males WT	63	31
<i>w¹¹¹⁸/tubP-GAL4</i> Females mut	17	17
<i>w¹¹¹⁸/tubP-GAL4</i> Males mut	2	3
Total	229	126

Table 3.12 The additional transcript of *CG9203* and *CR14033* provided by expressing a UAS cDNA transgene did not cause sterility in males or females

Expression of the UAS transgene for a cDNA of *CG9203* or *CR14033* in either the sense or antisense orientation was induced and the male and female progeny were examined for sterility. The number of individuals tested for each strain varied. Each vial was checked for the presence of progeny and scored accordingly as fertile or sterile. The other category was reserved for circumstances that did not produce progeny due to death of the parents before mating or eggs could be laid. Two constructs for *CR14033* in the sense orientation (A) and two in the antisense orientation (B) were used and two for each orientation for *CG9203* were also used (C and D). A positive control was also done using *w¹¹¹⁸* virgin females crossed to *tubP-GAL4* males (E).

A

	Sense Females (1-7A)	Sense Males (1-7A)	Sense Females (1-8A)	Sense Males (1-8A)
Fertile	10	7	15	12
Sterile	0	0	0	0
Total	10	7	14	12

B

	Antisense Females (2-7A)	Antisense Males (2-7A)	Antisense Females (2-24A)	Antisense Males (2-24A)
Fertile	7	5	13	9
Sterile	0	0	0	0
Total	7	5	13	9

C

	Sense Females (4-2A)	Sense Males (4-2A)	Sense Females (4-42B)	Sense Males (4-42B)
Fertile	10	6	19	12
Sterile	0	0	1	0
Total	10	6	20	12

D

	Antisense Females (3-31A)	Antisense Males (3-31A)	Antisense Females (3-38B)	Antisense Males (3-38B)
Fertile	17	20	6	7
Sterile	0	0	0	0
Total	17	20	6	7

E

	<i>w¹¹¹⁸/tubP-GAL4</i> Females	<i>w¹¹¹⁸/tubP-GAL4</i> Males
Fertile	13	13
Sterile	0	0
Total	13	13

Table 3.13 The lethal excision, 28B-D, was rescued by the duplication showing that the lethality was due to the P-element excision

The progeny are shown for the *Dp(1;3)DC300* x *28B-D/FM7c* cross. They are characterized based on eye morphology as the *FM7c* balancer carries the dominant Bar-eye marker. The *28B-D/FM7c* stock only has flies with the balancer showing the *28B-D* chromosome is recessive lethal. The male progeny from this cross that do not have the balancer are of genotype *28B-D/Y ; Dp(300)/ +*, showing that the lethality was rescued by the duplication.

Cross	non <i>FM7c</i> female		non <i>FM7c</i> male		<i>FM7c</i> female		<i>FM7c</i> males	
	<i>Dp(1;3)DC300</i> x 29B-D	24	26	31	29	29	19	2

Table 3.14. $P\{lacW\}l(2)k01302^{k01302}$ complemented with the deficiency of tkv strains but not with tkv^7

Canton S was used as a positive control for complementation in each complementation test while tkv^7 crossed to $Df(2L)BSC693$ was used as a negative control for complementation in both complementation test. WT stands for wild-type, mut for mutant, F for female, and M for males. (-) mean that there was lack of complementation while (+) means that they complemented. Table A shows the results of the complementation test between $P\{lacW\}l(2)k01302^{k01302}$ and tkv^7 and Table B shows the results of the complementation test between $P\{lacW\}l(2)k01302^{k01302}$ and $Df(2L)BSC693$.

A.

Strain	WT F	WT M	mut F	mut M	CyO	Complement (+) or (-)
<i>P{lacW}l(2)k01302^{k01302}</i>	7	11	0	0	34	(+)
CantonS	25	31	0	3*	76	(-)
<i>tkv⁷</i>	0	0	0	0	18	(-)

* extra bristles, blunt bristle

B.

Strain	WT F	WT M	mut F	mut M	CyO	Complement (+) or (-)
<i>P{lacW}l(2)k01302^{k01302}</i>	12	15	4*	0	72	(-)
CantonS	7	8	0	0	13	(+)
<i>Df(2L)BSC693</i>	0	0	0	0	18	(-)

* blunt bristles, additional cross vein(s), ectopic veins

Figure 3.1 Location of the P-elements inserted in *CRI4033* that were used for P-element mutagenesis

CRI4033 is shown as a grey rectangle with its transcript oriented left to right from 5' to 3'. The three P-elements used in the P-element mutagenesis screens are shown as dark grey triangles. All three P-elements, $P\{SUPor-P\}tkv^{KG01923}$, $P\{lacW\}l(2)k01302^{k01302}$, $P\{GSV6\}GS17311$ are located at the 5' end of the pseudogene 20 bp, 177bp, and 251 bp, respectively downstream from the transcription start of *CRI4033*. Figure adapted from Flybase (dos Santos *et al.* 2014).

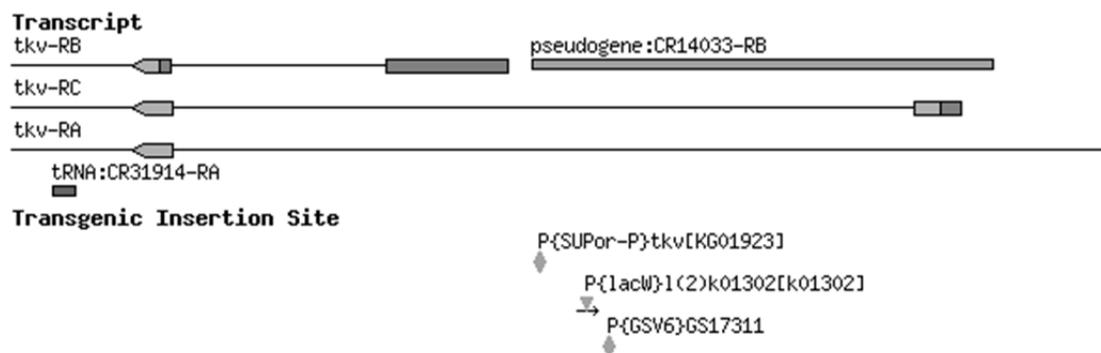


Figure 3.2 Sequence alignment of 2B-A and 8A-A

Multiple sequence alignment of the two partial excisions from $P\{SUPor-P\}tkv^{KG01923}$ that includes the sequence of the DNA flanking the P-element insert and *CRI4033* sequence. Flanking means the section of genomic DNA that is adjacent to the P-element insertion site as determined by the Berkley *Drosophila* Genome Project (Bellen *et al.* 2004). Alignments were done manually using Geneious 8.1.2

A. A schematic of the alignment between the $P\{SUPor-P\}$ element sequence (Roseman *et al.* 1995) and sequencing reactions across its insertion site in the reference genome strain (2057) and in the 2B-A strain from the right (2B-22) and the left (2B-12int) of the insertion site. The grey areas are regions of overlap between the sequences. The internal part of the $P\{SUPor-P\}$ element in 2A was not determined, but the large PCR product size indicated that it has a small deletion (data not shown).

B. Alignment between part of the $P\{SUPor-P\}$ element that is retained in 8A-A and sequencing reactions across its insertion site in the reference genome strain (2057) and the 8A-A strain. The data show that only 168 bp of the $P\{SUPor-P\}$ element (plus 8bp target site duplication) are retained in 8A-A. The red zig-zags represent trimmed low-quality sequence data and site of the deletion in the 12 kb $P\{SUPor-P\}$ element.

A



Figure 3.3 Schematic for *CR14033* deletion mutant, *CR14033*¹⁸

The deleted region of the pseudogene, 3318 bp in size, is seen as a black bar. The red arrow denotes the pseudogene and thickveins is seen as the blue bar. The three splice variants of *tkv* are seen as thick orange/grey arrows (exons) with thin black lines (introns) connecting them. Grey indicates 5'-untranslated and orange indicates amino acid coding sequence. P-elements inserted in this region are shown as blue triangles or diamonds and there are also two tRNA genes in this section of the genome.

To determine the breakpoints, primers 14033 2L and 5R were used for PCR to generate the template for sequencing. This PCR product was sequenced in both directions using the ex200010-18L and R primers. The grey shaded vertical bars in the figure represent the areas of the PCR product sequence that aligned in a BLAST search using the sequencing data as a query to the whole genome, aligning to the regions flanking the deletion.

The grey shading allows you to follow where the BLAST hits align.

Figure 3.4 Location of the P-elements inserted in *CG9203* that were used for P-element mutagenesis

CG9203 is represented by the series of rectangles (exons) and thin line connecting them (introns). The three P-element insertions, *P{SUPor-P}CG9203^{KG05829}*, *P{XP}d04265*, and *P{GawB}NP0833* used for P-element mutagenesis are shown as grey triangles. The three P-elements are inserted 322 bp into the 5' end of *CG9203* and 665 bp and 707 bp upstream of *CG9203*, respectively. Figure adapted from Flybase (dos Santos *et al.* 2014).

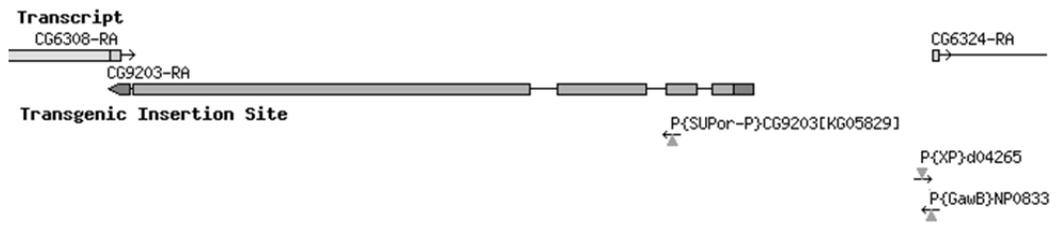


Figure 3.5 Schematic showing the 1 kb deletion resulting from the imprecise excision in 24A-D

A schematic showing the sequenced regions of 24A-D aligned to *CG9203* (*mh*), allowing visualization of the deleted region for this allele. The grey shading allows you to follow where on *CG9203* and which base pairs the BLAST hits (HSP (1) and HSP (2)) match. The space between the BLAST hits (non-shaded area) is the deleted region of *CG9203*.

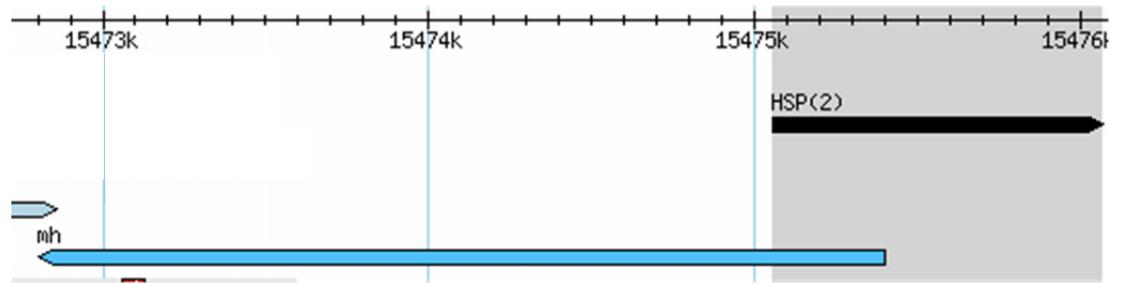
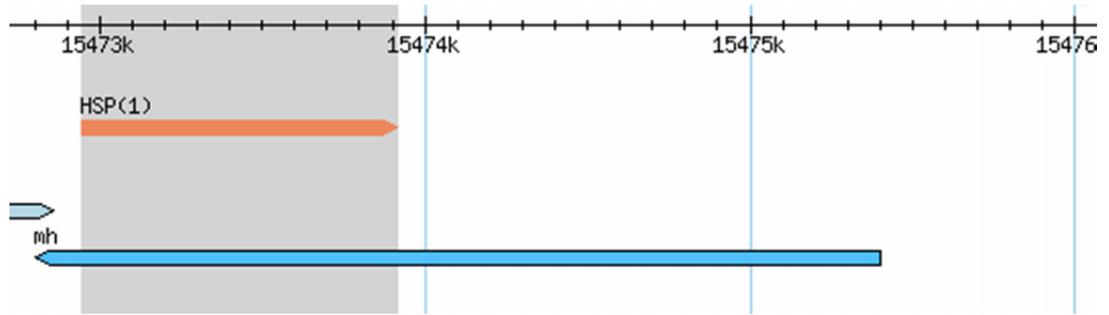


Figure 3.6 Pairwise sequence alignment between *CR14033* and *CG9203* cDNAs

The query sequence is the *CR14033* cDNA (LD26673) and the subject sequence is the *CG9203* cDNA (MIP11944), which are 1526 and 2357 nucleotides in length, respectively, without their poly-A tails. Alignment was performed using the Needleman-Wunsch global pairwise alignment program (Needleman and Wunsch 1970) as implemented in EMBOSS (Rice *et al.* 2000).

Figure 3.7 Alignment of the sequence from the different reading frames of *CR14033* that has sequence similarity *CG9203*

A composite of the sequences found in one of the three forward reading frames of *CR14033* that have similarity to *CG9203*.

The black letters indicate conserved amino acids, purple letter indicate an amino acid change, red stars and letters are for nonsense mutations, blue letters indicate sequence deleted in the pseudogene and green letters indicate an insertion. The underlined letters denote sequence found on reading frame 3 while the sequence that is not underlined is for sequence found in the 2nd reading frame.

CR14033 1 KMPACRNTMSSQERTQNIKR*IMEDELTLCEVNTLSIDYEAGK*PCARSIPFOSIMRRP
 CG9203 514 KMPAGRKTMTSQERTQIIKRQVMEDELTLCEDDILLIDDEYDDE-----VA

CR14033 ANDSLTAATELADOSIIDDVFGEDILMKDLORKNTVOPTCSTTTSYLHLLNNDIVSCPIC
 CG9203 ANESLTAATELADQSIIDDIFGEDTLLEKFQRENDVQPSCS---SYAKRIENDIVSCPIC

CR14033 FEKMKRG-----TIMVCLEPPSF-----TSKDLDRKSSSGSSTSRCTKS
 CG9203 FEKMRTELSNHFDGCAIMVRLPEPPSFKPKNRRTTASLTSKD-----SSGASTSRAKS

CR14033 KNSSSKRILRSSGYTE-----LNSNSTT-----LSSSEDELTPRQRL*CNLFKQTVGCPRC
 CG9203 KAKSSKGILRSSGYTEGEIDKLNLSSTDSSTLPSSSEDELTPRQRQRNLFKETVGCPRC

CR14033 GLEFKGHQLKAHQSLCQGRKKH* 245
 CG9203 GLEIMPLHLKVHGSVCAGRKKR* 751

Figure 3.8 Alignment of the sequence from the different reading frames of the *D. sechellia* CR14033 and CG9203 orthologs

A composite of the sequences found in one of the three forward reading frames of the CR14033 ortholog in *D. sechellia*, scaffold 5, that have similarity to *D. sechellia* CG9203 ortholog, scaffold 73.

The black letters indicate conserved amino acids, purple letter indicate an amino acid change, red stars and letters are for nonsense mutations, blue letters indicate sequence deleted in the pseudogene and green letters indicate an insertion in the pseudogene. The underlined letters denote sequence found on reading frame 3 while the sequence that is not underlined is for sequence found in the reading frame 2.

scaffold 5 1 EVKMPASRNTMSSQERTQNI----MEDELTLC**EVKIL***FHEE*EEEEAAV
 scaffold 73 305 EMKMPAGRNTMSSQERTQNI**KRQV**MEDELTLCEDDIL**LIDDEYDDEVAA-**

 scaffold 5 NDSLTAATELADQSI**IKD**VFGE**DILMKEWQLENAVHPTC-SYAK-T----**
 scaffold 73 NDSLTAATELADQSI**IDDIF**GED**TLLKEFQRENDVQPTCSSYGKNTGNDI**

 scaffold 5 -----MVCLEPPSF**TSKMDSSSSSLRP----**
 scaffold 73 **VSCPICFEKMKRTELSNHFDGCAIMVRL**EPSP**FKPKNRRPRASFTSKDSS**

 scaffold 5 --STSR**CAKSKA--SKRILRNSGNTE-EIAELHLS**SSSQ**LDYSAVQRGRT**
 scaffold 73 **GFSTSRGAKSKAKS**SKRILR**SSGYTEKEIDELFHR----**LVC**SALQRGRT**

 scaffold 5 YSASAPPAQ**SLQ**TD**RRLS***MWPE**VQ**PPAEGAS**ISLPGSQEALQ**²⁴⁴
 scaffold 73 YSASAPAA**QPLQ**TD**RRLSQ**MWP**GDH**AP**PEGASICLRGLQEALG** 543

Figure 3.9 Alignment of the sequence from the different reading frames of the *D. simulans* CRI4033 and CG9203 orthologs

A composite of the sequences found in one of the three forward reading frames of the *D. simulans* CRI4033 ortholog that have similarity to the CG9203 ortholog.

The black letters indicate conserved amino acids, purple letter indicate an amino acid change, red stars and letters are for nonsense mutations, blue letters indicate sequence deleted in the pseudogene and green letters indicate an insertion in the pseudogene. The underlined letters denote sequence found on reading frame 3, while the sequence that is not underlined is for sequence found in the reading frame 2.

sim CR14033 1 DAS*QEHHEQSGAHAHQARGNGG*IDLVRGQYPLIS*RAGGGRAGRK*QFNGRNGTS
sim CG9203 788 DASWQEHHEQSRHAHAEQAPGNGG*VDLV*GRYSLN***IRR*D-GRK*LFNSRHGTS

sim CR14033 *PIHH*GCF*KXI-----HLLNLRQYEHRHRILSHLL*EDEAHSVGQSI
sim CG9203 *PIHH*YFWRGHPAEGVPA*ERRAAHLLKLRQYEQRHCIILSHLL*EDEAH*VVQSL

sim CR14033 RERLKMPACRNTIHSYLAGIYAWSLALKSLATEPSD*HNS*RPGRAQLRRRGLGEGP
sim CG9203 RWLRHYGALGTAIV*A*KPSTKGFLLY*GLLRI-----

sim CR14033 TTIGEKNNTSSWGPFFLCNRTRREP*AEVNNGPPPPIVIRVCFWRGHPDERVAA*ERRA
sim CG9203 -----

sim CR14033 AHLLKLRQNYGVLGTAIVYFKGHGLQFLLLGTFHITLRQIESH-IKANSSQLR*HRGG
sim CG9203 -----FHITRRQIESQVIKTNSSQLRLHRKG

sim CR14033 DRRAPPELLQQLDYSAVQRGRTYASAPPAQSLQTDRRLSQMWPEVQPPAEGASISL
sim CG9203 DRRAQPELLH*LVCSALQRGRTYASAPAAQPLQTDRRLSQMWPGDHAPPPEGASICL

sim CR14033 PGSQEALGGHLIQTIIH*I 365
sim CG9203 RGLQEALGRHLLQTIIR*I 103

Figure 3.10 *CRI4033* is conserved in the melanogaster subgroup

CG9203 is found on the X chromosome. The region of *CG9203* that shows sequence similarity to *CRI4033* is the 3' end. It is represented in this figure as a dark green arrow. The black portion of the arrow represents the 5' end of *CG9203* that does not share any sequence similarity to the *CRI4033*. A full black arrow is present in the species that do not possess any pseudogene sequence.

CRI4033 is located on chromosome 2 and is nested in a *tkv* intron. The region of *CRI4033* that shares sequence similarity to *CG9203* is shown as a light green arrowhead and the region that does not share similarity is shown in light blue. If the light green and blue are not found in the species this means that no *CRI4033* paralog exists in that species. *tkv* is represented by the purple arrow that has a peak in it representing an intron and the exons are represented by the horizontal lines.

For *D. simulans*, the upper arrows represent the results with the Flybase genome assembly and the lower arrows represent the results with the assembly by Hu *et al.* (2013). The phylogenetic tree is based on one provided to Flybase by Nicolas Gompei (dos Santos *et al.* 2014). The melanogaster group is indicated with the larger bracket on the right and the melanogaster subgroup is indicated by the smaller bracket. Branch lengths are not to scale.

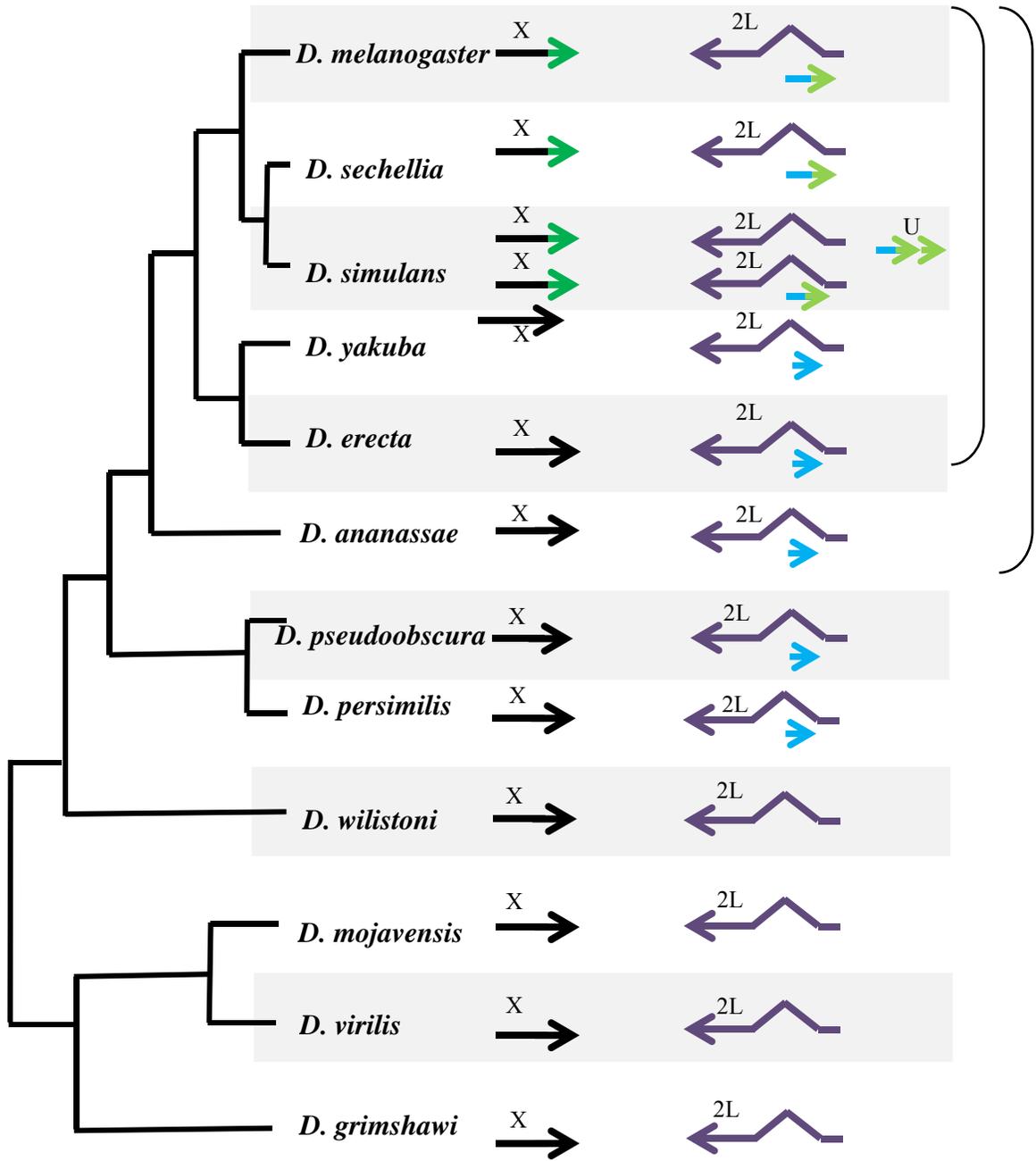


Figure 3.11 Expression of *CG9203*, *CRI4033*, and *tkv* in *Dcr2* or *loqs* mutant testes

Fold changes in transcript levels in testes from *Dcr2* or *loqs* mutants compared to *w¹¹¹⁸*. Each replicate represents three sets of testes from three *Dcr2* or *loqs* mutants. Each were tested for amount of intact transcript for *CG9023*, *CRI4033*, and *tkv* was measured using

Fold change is on the y-axis and the replicates for each mutant is on the x-axis. The dark grey bars show fold change of *CG9023* transcript, the light grey bars show *CRI4033*, and the medium grey bars represent *tkv* fold change. Standard error is also depicted.

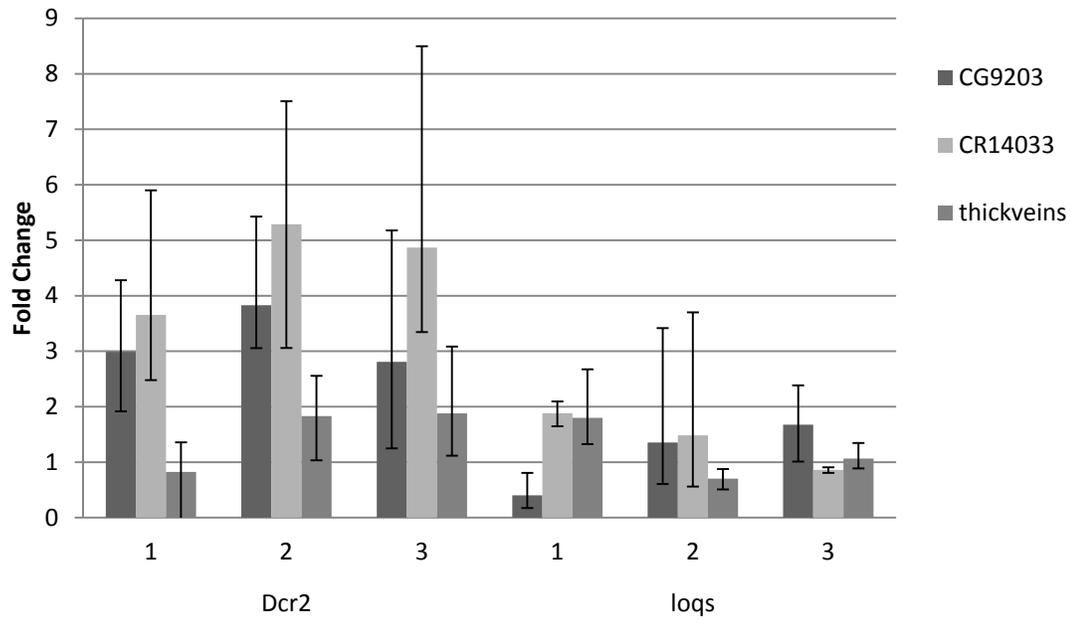


Figure 3.12 Expression of *CG9203*, *CRI4033*, and *tkv* in brains and discs from *Dcr2* mutants.

Fold changes in transcript levels in brains and discs from *Dcr2* mutants compared to *w¹¹¹⁸*. Each replicate represents three sets of brains and discs from three *Dcr2* mutants. Each were tested for amount of intact transcript for *CG9023*, *CRI4033*, and *tkv* was measured using qPCR

Fold change is on the y-axis and the replicates for each mutant are on the x-axis. The dark grey bars show fold change of *CG9023* transcript, the light grey bars show *CRI4033*, and the medium grey bars represent *tkv* fold change. Standard error is also depicted for each replicate. The numbers 1 and 2 denote replicates.

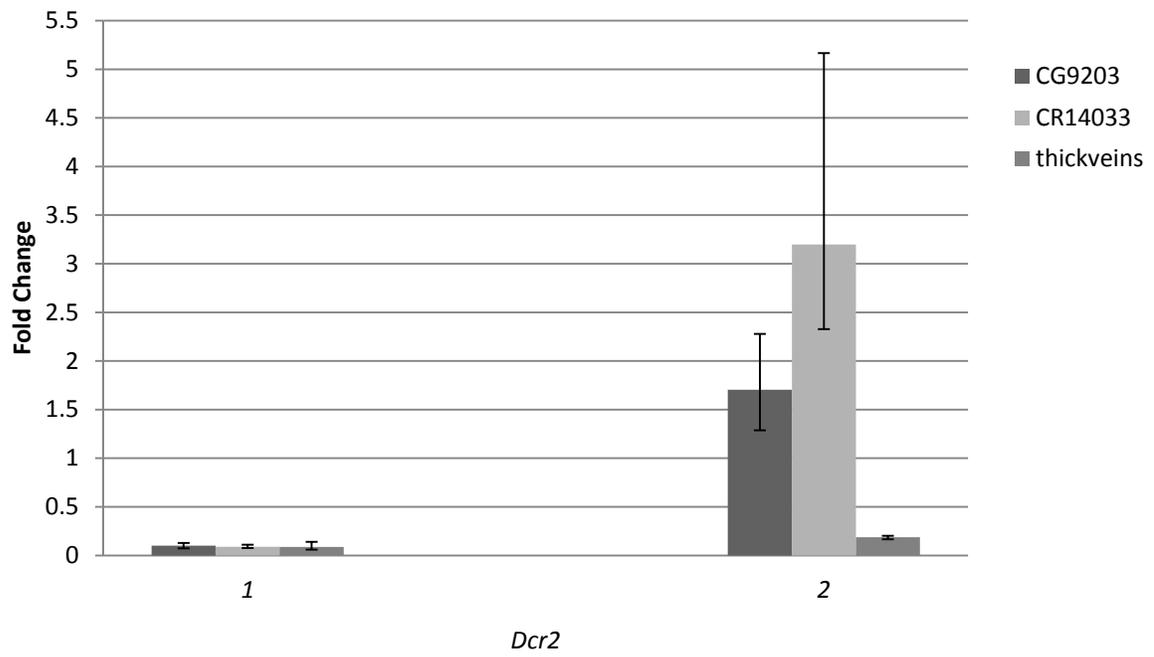


Figure 3.13 Transcript changes in *CG9203*, *CRI4033*, and *tkv* in testes from flies expressing sense and antisense cDNA transgenes of *CRI4033*

Fold change on a logarithmic scale is on the y-axis. On the x-axis are the replicates for the two pseudogene cDNA transgenes. 1-7A and 1-8A are the strains with the pseudogene oriented in the sense direction and 2-7A and 2-24A in the antisense direction. Each of the replicates has three sets of testes (X-axis labels 1-3). *CG9203* is the dark grey bar, *CRI4033* is a light grey bar and *tkv* is a medium grey bar. The error bars reflect standard error. Two asterisks (**) denote p-values <0.05 while three asterisks (***) denote p-values of 0.

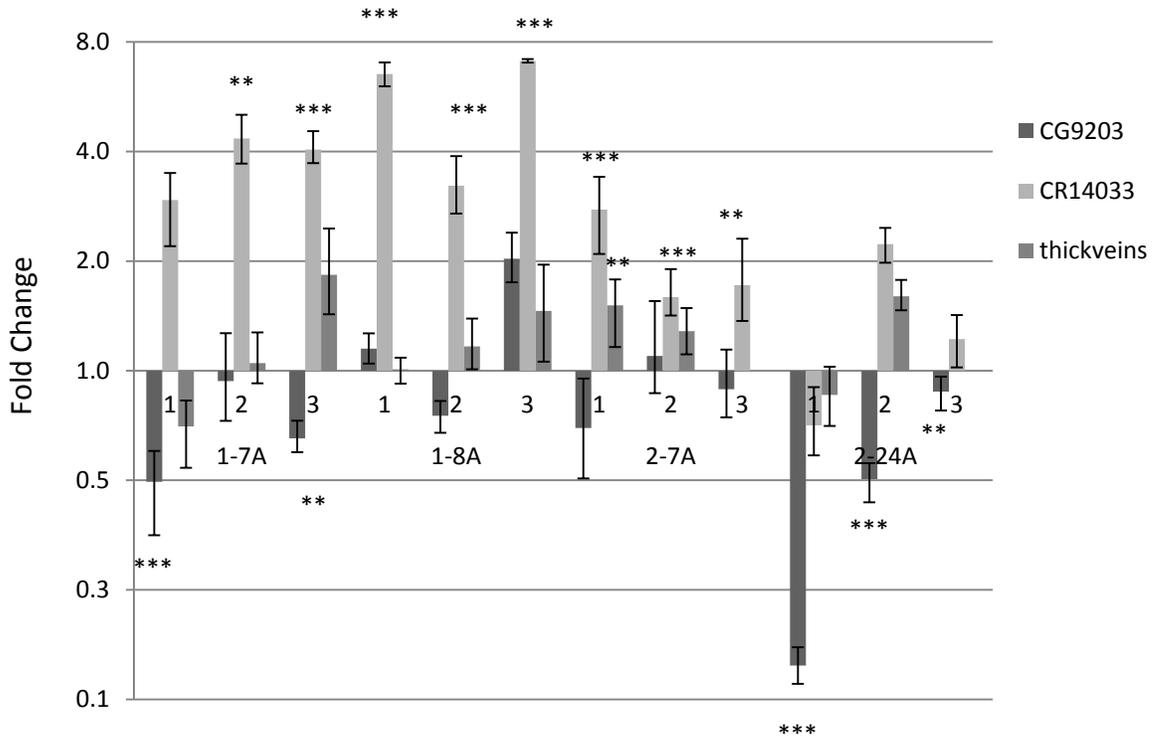


Figure 3.14 Transcript changes in *CG9203* and *CRI4033* in testes from flies expressing sense and antisense cDNA transgenes of *CG9203*

Fold change on a logarithmic scale is on the y-axis. On the x-axis are the replicates for the two parent gene cDNA transgenic flies. 3-31A and 3-38B are the parent gene in the antisense direction and 4-2A and 4-42B are the parent gene in the sense direction. Each of the replicates has three sets of testes. *CG9203* is the dark grey bar and *CRI4033* is a light grey bar. The error bars reflect standard error. Two asterisks (**) denote p-values <0.05 while three asterisks (***) denote p-values of 0.

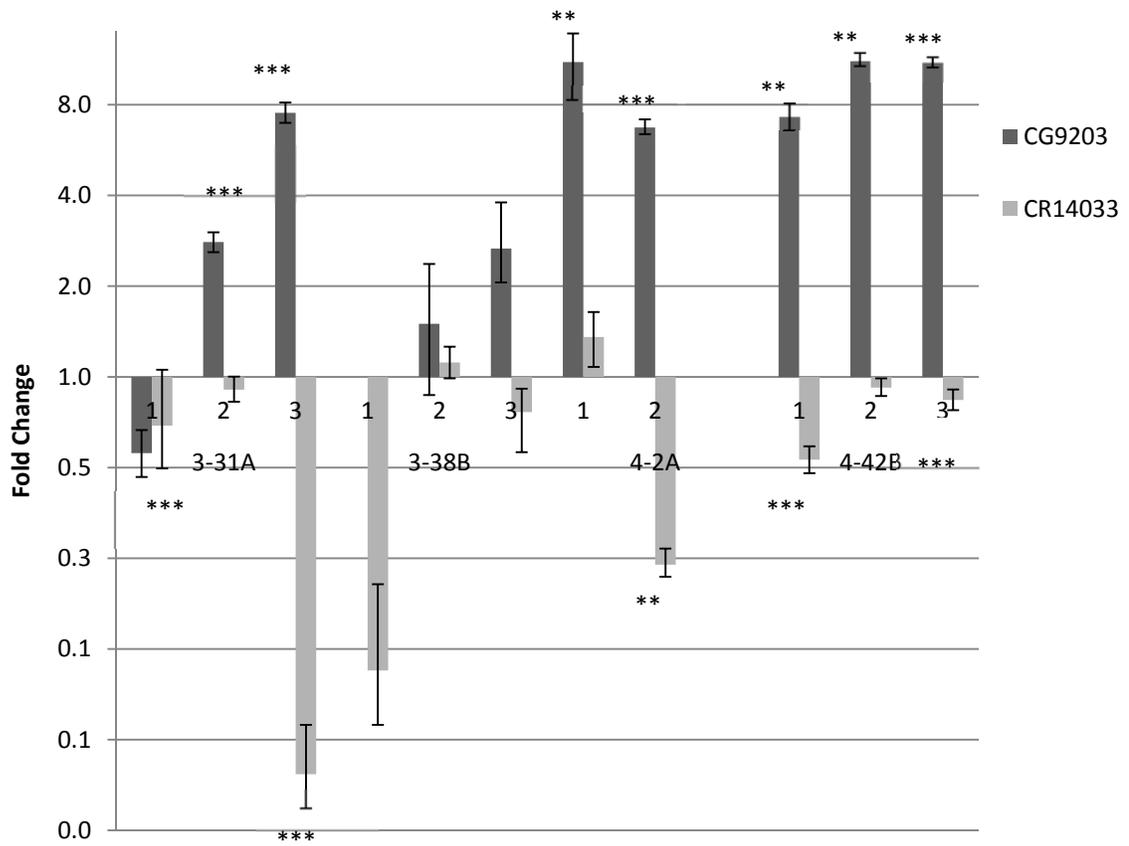
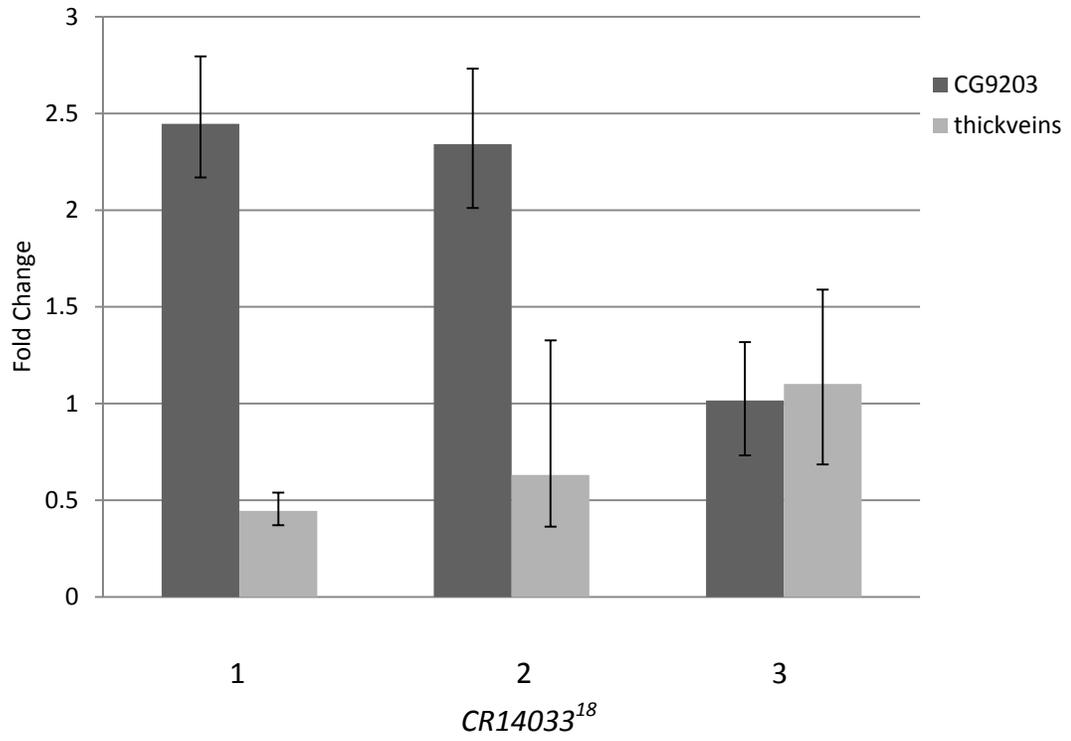


Figure 3.15 Expression of *CG9203* and *tkv* in a *CRI4033* deletion mutant

The fold change in intact transcript levels of testes from *CRI4033*¹⁸ deletion mutants relative to *w*¹¹¹⁸. Each replicate represents three sets of testes from three *CRI4033* deletion mutants. Intact transcript was measured using qPCR.

Fold change is on the y-axis and the replicates for each mutant are on the x-axis. The dark grey bars show fold change of *CG9023* transcript and the medium grey bars represent *tkv* fold change. Standard error is also depicted for each replicate.



Chapter 4: Discussion

The function of *CR14033* was not completely elucidated

Mutations are an important method for analyzing gene function (Bellen *et al.* 2004). In an attempt to determine the function of *CR14033*, P-element mutagenesis was performed on P-element insertion strains $P\{SUPor-P\}tkv^{KG01923}$, $P\{GSV6\}GS17311$, and $P\{lacW\}l(2)k01302^{k01302}$. The phenotypes of the original insertions and various excision alleles were characterized. The only deletion mutant generated for *CR14033*, $CR14033^{18}$, was from an imprecise excision of $P\{GSV6\}GS17311$.

Because a dominant phenotype was not seen for $CR14033^{18}$, a complementation test with a deficiency of *thickveins* was used to reveal any recessive ones. No recessive phenotypes were seen in the progeny. It is possible that the function of *CR14033* does not necessarily change the outward appearance of the mutant or at least not to a degree that is overtly apparent. Also, it is possible that its function is reproductively oriented therefore it would only be visible as sterile flies or flies with decreased fertility.

When sterility was checked in males and females, all of the females were fertile while the males showed some sterility but this was also seen in the control. Because the p-value was >0.05 this could have been due to random chance (Section 3.3). A factor that could lend to the uncertainty of whether this sterility is due to random chance or true is the low sample size in this experiment. In addition, in order to ensure that these males are not partially sterile as opposed to completely sterile, it would be beneficial to perform an experiment and determine the number progeny the males produce and compare it to a control.

The P-element insertion *P{GSV6}GS17311* strain, 200-010, also showed partial male sterility. This sterility was shown to be significant and not by chance with a p-value of 0.0000952 using the Fisher's exact test. The reason for this sterility is unclear. There is no precedence for this in literature in regards to loss of function *tkv*. The only study that showed a sterility phenotype for *tkv* was looking at a constitutively active mutant form of *tkv* (Xia *et al.* 2010).

As overexpression of the pseudogene could be predicted to cause a greater reduction in the *CG9203* transcript and, hence a mutant phenotype, an increased amount of the pseudogene in the sense or antisense direction was generated by inducing transcription of the cDNA transgenes. The strains with a pseudogene cDNA in the sense direction, 1-7A and 1-8A, as well as the stains with the antisense cDNA, 2-7A and 2-24A, were induced and the progeny were examined for changes in phenotype. No phenotypic changes were noted. Sterility tests were also performed on these strains. No changes in sterility were noted. Therefore, *CRI4033* does not affect *CG9203* in the ways that I tested for here.

CG9203 is involved in UV resistance but most importantly it is involved in paternal chromosome integrity during zygote formation

In this study, it was not possible to create a complete deletion mutant of *CG9203* using P-element mutagenesis. However, a partial deletion (24A-D) of *CG9203* was generated using *P{SUPor-P}CG9203^{KG05829}*. For the purpose of determining if there is an interaction between *CG9203* and *CRI4033*, 24A-D was not useful as it still has a transcription start site (Figure 3.5) and it is unknown if 24A-D was transcriptionally

active. In order to determine this qPCR could have been used to see if a transcript could be detected for the region.

Since these screens were performed, another lab generated a *CG9203* deletion mutant using $P\{SUPor-P\}CG9203^{KG05829}$ and using this deletion mutant they were able to determine the function of *CG9203*. They have since renamed *CG9203*, *maternal haploid (mh)*, based on its function. The deletion mutant generated by Delabaere *et al.*, mh^2 , had a 903 bp deletion and a premature stop codon at the end of the second exon and therefore only 65 residues remained of the 724 aa that were predicted (2014). Like 24A-D, this mutant was found to be female sterile where the females produced normal amounts of eggs but these eggs did not hatch. What they were able to determine was that these eggs did not hatch because the rounds of mitosis that occur post-fertilization did not progress normally leading to arrested development of the embryos where only 20% were able to survive and developed as nonviable gynogenetic haploid embryos. They suspect that *mh* is involved in maintaining the integrity of the male DNA during nuclear decondensation because during the first S phase, MH proteins gather in the male pronucleus explaining why in its absence mitosis in the embryo does not progress normally. (Delabaere *et al.* 2014).

Delabaere and colleagues were also able to show that MH, like its human homolog Spartan, is involved in the regulation of translesion synthesis (TLS) in response to UV-induced DNA damage (2014). Exposing mh^2 larvae and larvae deficient for TLS polymerase DNA pol-eta to UV radiation had the same result. They both showed the same sensitivity. They were able to rescue this phenotype using two copies of *V5:mh*

transgene (Delabaere *et al.* 2014). From these experiments they concluded that MH is involvement in repair of UV-induced DNA damage.

An interaction between *CG9203* and *CRI4033* but not between *CRI4033* and *tkv*

*CRI4033*¹⁸ allowed me to determine if there was an interaction between *CRI4033* and *CG9203* through the change in the amount of intact transcript seen using qPCR. With two of the three replicates for *CRI4033*¹⁸, a significant increase in *CG9203* transcript was observed. This is consistent but inconclusive with the idea that *CRI4033* has a role for regulating the expression of level of *CG9203*. In the absence of *CRI4033* there appears to be more available *CG9203* transcript, presumably because it is not being cleaved by the endo-siRNA pathway. In order to further establish that this is the case, the amount of cleaved transcript should be measured using a northern blot.

To look for additional evidence for an interaction between *CG9203* and *CRI4033*, the expression of *CG9203* was examined when there was additional ubiquitous production of *CRI4033* transcript using cDNA transgenes of LD26673. Results suggest that changes in the amount of *CG9203* transcript are dependent on the amount of *CRI4033* transcript present. In two strains expressing the sense cDNA transgene, there is an increase in *CRI4033* transcript, indicating that the transgene transcription was induced, and there was a decrease in *CG9203*. However, in one strain, 1-8A, the second replicate shows an increase in *CG9203*. The difference seen could be due to changes in expression level of the genes of interest if the flies were of different ages or if the testes dissected out were contaminated with another tissue.

From all of these results in this section as well as those previously mentioned, it would appear as though the function of *CRI4033* is not clear-cut. However, this *D. melanogaster* pseudogene that is conserved among three *Drosophila* species does appear to interact with *CG9203* by decreasing the amount of intact transcript through the siRNA pathway. However, the exact function of this decrease or other possible functions of this pseudogene are unknown.

tkv does not appear to be regulated through the endo-siRNA pathway whereas *CRI4033* is

In an effort to determine if the endo-siRNA pathway regulates *CRI4033*, like *CG9203*, the amount of intact transcript was examined in *Dcr2* and *loqs* mutants in testes and in imaginal discs and brains.

In *Dcr2* testes *CRI4033* transcript increased, as predicted, where the increase in two of the three replicates was significant. The reason for the increase is because *Dcr2* is responsible for cleaving the dsRNA into 21 nt segments, which then find complementary RNA as part of an RNA-induced silencing complex to cleave complementary mRNA (MacRae *et al.* 2006 and 2007). Without *Dcr2* this is not possible and thus the mRNA for *CRI4033* remains intact.

In *loqs* testes the results were very inconsistent. There were both increases and decreases with one of each being significant. This was the case with all genes tested in *loqs* mutants including *tkv*. The function of *loqs* is to help the 21 nt fragments to bind to *Dcr2* (Zhang *et al.* 2004; Liu *et al.* 2003; Zhou *et al.* 2009). The inconsistency seen with the results could be because Loqs-PD and R2D2 have some overlapping function

(Mirkovic-Hosle and Forstemann 2014), which could have compensated for the loss of Loqs-PD. It may be beneficial to look at testes from *r2d2* mutants to see if the same results are seen. From the results obtained from the three replicates, it would appear as though *loqs* has no effect on the genes of interest.

In imaginal discs and brains from *Dcr2* mutants, only two replicates were obtained and they are equivocal. These results suggest that more work in this area needs to be done to elucidate whether *tkv* is a target of the endo-siRNA pathway. For instance, the sample collection times may need to be more precisely controlled. For 3rd instar larvae 24 hours is a long time and some genes could fluctuate in the amount of transcript being produced at different times during that 24 hours.

In conclusion, in both *Dcr2* mutants and *loqs* mutants, no significant effects on *tkv* expression were observed. .

Covariation was found between *CG9023* and *CR14033*

A novel program, CovaRNA, was generated by Bindewalk and Shapiro (2013) that allowed them to detect long-range nucleotide covariation from genomic multiple sequence alignment. Long-range nucleotide covariation in this study either indicated covariation clusters that could potentially have long range interaction between two genomic regions either on the same chromosome and be at least 6000 nt apart or on two different chromosomes. They found covariations that related pseudogenes with protein-coding genes. What is really interesting is that covariation was found between the *tkv* locus, which has *CR14033* as its antisense transcript and the locus corresponding to *CG9203*. This result indicates that there could be an interaction occurring and that the

sequences are under selection to maintain their similarity. Their hypothesis of interaction between the three genes is as follows. When *CRI4033* is expressed it forms endo siRNA with *tkv*, which leads to the down regulation of *CG9203*. A change in *tkv* was not seen in my study when there was an increase in *CRI4033* transcript but the primers I used for *tkv* transcript were detecting *tkv* mature mRNA. Since the region of *CRI4033* that has similarity to *CG9203* is in the *tkv* intron, the qPCR of *tkv* mRNA did not directly measure the abundance of this intron RNA, which would be needed to predict if it is used for generation of siRNA that is targeted to *CG9203*.

CRI4033 is the product of a gene fusion event

Gene fusion occurs in many organisms. It occurs as a result of chromosomal rearrangements and abnormal transcription. A few examples of fused genes are the fatty acid synthase genes in (McCarthy and Hardie 1984), tryptophan synthase gene in fungi (Burns *et al.* 1990), the Sp100-rs in *M. musculus* (Weichenhan *et al.* 1998), and the Kua and UEV genes in humans (Thompson *et al.* 2000). Gene fusion can result in new functions for proteins such as those seen with *Kua* and *UEV*. The non-fused form of UEV proteins normally localize in the nucleus whereas the KUA proteins localize in endomembranes. When KUA and UEV are combined they are found with cytoplasmic structures so that UEV proteins work in a new location and therefore have a new function. A duplicate copy of *UEV*, *UEV2*, has been found suggesting that there is pressure to maintain the original function of *UEV*, along with its new fused function (Thomson *et al.* 2000).

When searching for orthologs of *CR14033* in 12 of the other *Drosophila* species, the longest regions of sequence similarity to the pseudogene were found in two species, *D. simulans* and *D. sechellia*. These two species plus *D. melanogaster*, all belong to the melanogaster subgroup, which shared a common ancestor with the obscura group (*D. pseudoobscura/D. persimilis*) around 55 million years ago (Tamura *et al.* 2004). Sequences similar to the 5' end of *CR14033* are present in *D. erecta* and *D. yakuba*. They were also seen in one of the species in the obscura group, *D. pseudoobscura*, and the melanogaster group, *D. annanasae*. No other species appeared to possess the 5' end or 3' end of the pseudogene. From this analysis it appears that the 3' end of *CR14033* inserted next to the 5' end in the melanogaster subgroup while the 5' end, which does not have similarity to *CG9203*, appears to have been there already. The species that are missing the 5' end could have had it once but there was no selection pressure to keep it and therefore it acquired a number of mutations to the point where it no longer is detectable in a BLAST search.

When comparing *CR14033* to its orthologs, and *CG9203* to its orthologs, and then comparing the parent genes and pseudogenes (paralogs) in each species, the pseudogene orthologs were all more similar to each other than they were to their respective parent gene paralogs. This could be because there is a function that the pseudogene provides other than regulation of the parent gene and that there is selection to preserve the sequence.

Bibliography

- Adachi-Yamada, T., K. Fujimura-Kamada, Y. Nishida and K. Matsumoto, 1999
Distortion of proximodistal information causes JNK-dependent apoptosis in
Drosophila wing. *Nature* **400**: 166-169.
- Adams, M. D., S.E. Celniker, R.A. Holt, C.A. Evans, J. D. Gocayne, and P.G.
Amanatides, 2000 The genome sequence of *Drosophila melanogaster*. *Science*
287: 2185-2195
- Baek, D., J. Villen, S. Chanseok, F. Camargo, S. Gygi, and D. Bartel. 2009. The impact
of microRNAs on protein output. *Nature*. **45**: 64-71.
- Bartel, D. P., 2009 MicroRNAs: target recognition and regulatory functions. *Cell* **136**:
215-233.
- Bellen, H. J., R. W. Levis, G. Liao, Y. He, J. W. Carlson *et al.*, 2004 The BDGP gene
disruption project: single transposon insertions associated with 40% of Drosophila
genes. *Genetics* **167**: 761-781.
- Betran, E., W. Wang, L. Jin and M. Long, 2002 Evolution of the phosphoglycerate
mutase processed gene in human and chimpanzee revealing the origin of a new
primate gene. *Mol Biol Evol* **19**: 654-663.
- Bindewalk, E., and B. A. Shapiro, 2013 Computational detection of abundant long range
nucleotide covariation in Drosophila genomes. *RNA*. **19**: 1171-1182
- Boto, L, 2010 Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci*
277: 819-827.
- Bridges, C. B., 1936 The Bar "Gene" a Duplication. *Science* **83**: 210-211.

- Burns, D.M., V. Horn, J. Paluh, and C. Yanofsky, 1990 Evolution of tryptophan synthetase of fungi: Analysis of experimentally fused *Escherichia coli* tryptophan synthetase alpha and beta chains. *J. Biol Chem.* **4**: 2060-2069
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, *et al.*, 2009 BLAST+: architecture and applications. *BMC Bioinformatics.* **10**: 421
- Casso, D., F. Ramirez-Weber And T. B. Kornberg, 2000 GFP-tagged balancer chromosomes for *Drosophila melanogaster*. *Mech Dev* **91**: 451-454.
- Cenik, E.S., R. Fukunga, G. Lu, R. Dutcher, Y. Wang, T.M. Tanaka Hall, and P.D. Zamore. 2011 Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol Cell.* **2**: 172-84.
- Chen, M, 2010 Smoke and miRrors: pseudogenes tricking miRNAs. 2010 *Pigment Cell Melanoma Res.* **5**: 583-584
- Christensen, S., Cook, K., Cook, K. (2008.12.28). Isolation and characterization of Df(2L)BSC693.
- Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde *et al.*, 2008 An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**: 798-802.
- Darnell, J. H. Lodish, and D. Baltimore, 1990 *Molecular Cell Biology*, Scientific American Books, New York, NY.
- Delabaere, L., G. A. Orsi, L. Sapey-Triomphe, B. Horard, P. Couble *et al.*, 2014 The Spartan ortholog maternal haploid is required for paternal chromosome integrity in the *Drosophila* zygote. *Curr Biol* **24**: 2281-2287.
- D'errico, I., G. Gadaleta and C. Saccone, 2004 Pseudogenes in metazoa: origin and features. *Brief Funct Genomic Proteomic* **3**: 157-167.

- Dierick, H. A., A. N. Adam, J. F. Escara-Wilke and T. W. Glover, 1997
Immunocytochemical localization of the Menkes copper transport protein
(ATP7A) to the trans-Golgi network. *Hum Mol Genet* **6**: 409-416.
- Dietzl, G., D. Chen, F. Schnorrer, K.C. Su, Y. Barinova, *et al.* 2007 A genome-wide
transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*
448: 151-156
- dos Santos G., A.J. Schroeder, J.L. Goodman, V.B Strelets, and M.A. Crosby, 2014
FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference
genome assembly and large-scale migration of genome annotations. *Nucleic
Acids Res* **43**: D690-D697
- Drosophila* 12 genome consortium, A.G. Clark, M.B. Eisen, D.R. Smith, C.M. Bergman
et al. 2007. Evolution of gene and genomes on the *Drosophila* phylogeny.
Nature. **450**: 203-18
- Duffy. J.B., 2002 GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife.
Genesis **34**: 1-15
- Engels, W.R. 1996 P elements in *Drosophila*. *Transposable elements*. Springer-Verlag
Berlin 103-123
- Evans, T.A., H, Haridas, and J.B. Duffy, 2009 Kekk5 is an extracellular regulator of
BMP signaling. *Dev. Biol.* **326**: 36-46
- Fincham, J. R.S., 2000 Unequal Crossing Over, 2001 Academic Press: 2095-2096
- Fishman, G. I., R. L. Eddy, T. B. Shows, L. Rosenthal and L. A. Leinwand, 1991 The
human connexin gene family of gap junction proteins: distinct chromosomal
locations but similar structures. *Genomics* **10**: 250-256.

- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545.
- Ghosal, G., J. W. Leung, B. C. Nair, K. W. Fong and J. Chen, 2012 Proliferating cell nuclear antigen (PCNA)-binding protein C1orf124 is a regulator of translesion synthesis. *J Biol Chem* **287**: 34225-34233.
- Goecks, J., C. Eberhard, T. Too, A. Nekrutenko and J. Taylor, 2013 Web-based visual analysis for high-throughput genomics. *BMC Genomics* **14**: 397.
- Golden, D.E., V.R. Gerbasi, and E.J. Sontheimer. 2008 An Inside Job for siRNAs. *Molecular Cell* **31**: 309-312.
- Graveley, B.R., A.N. Brooks, J.W. Carlson, M.O. Duff, and J.M. Landolin, 2011 The developmental transcriptome of *Drosophila melanogaster*. *Nature*. **471**: 473-479.
- Greenspan, R.J, 1997 Fly Pushing: The Theory and Practice of *Drosophila* Genetics. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.
- Greenspan, R.J, 2004 Fly Pushing: The Theory and Practice of *Drosophila* Genetics. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.
- Gu, S., L. Jin, F. Zhang, P. Sarnow and M. A. Kay, 2009 Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol* **16**: 144-150.
- Han, T., A. P. Manoharan, T. T. Harkins, P. Bouffard, C. Fitzpatrick *et al.*, 2009 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* **106**: 18674-18679.

- Harrison, P. M., D. Milburn, Z. Zhang, P. Bertone and M. Gerstein, 2003 Identification of pseudogenes in the *Drosophila melanogaster* genome. *Nucleic Acids Res* **31**: 1033-1037.
- Harrow, J., A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, *et al*, 2012 GENCODE: the reference human genome annotation for THE ENCODE Project. *Genome Res.* **22**: 1760-1774
- Hartig, J. V., S. Esslinger, R. Bottcher, K. Saito and K. Forstemann, 2009 Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *EMBO J* **28**: 2932-2944.
- Holland, C., D. B. Lipsett and D. V. Clark, 2011 A link between impaired purine nucleotide synthesis and apoptosis in *Drosophila melanogaster*. *Genetics* **188**: 359-367.
- Horsfield, J., A. Penton, J. Secombe, F.M. Hoffman, and H. Richardson, 1998 Decapentaplegic is required for arrest in G1 phase during *Drosophila* eye development. *Development* **125**: 5069-5078.
- Hu, T.T., M.B. Eisen, K.R. Thornton, and P. Andolfatto. 2013 A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. *Genome Res.* **23**: 89-98.
- Huang, A. M. E. J. Rehm and G.M. Rubin, 2009 Quick Preparation of Genomic DNA from *Drosophila*. *Cold Harbour Protocols* **2009**: pdb.prot5198
- Huang, A. M., E. J. Rehm and G. M. Rubin, 2009 Recovery of DNA sequences flanking P-element insertions in *Drosophila*: inverse PCR and plasmid rescue. *Cold Spring Harb Protoc* **2009**: pdb prot5199.

- Hummel, T., and C. Klambt, 2008 P-element mutagenesis. *Methods Mol Biol* **420**: 97-117.
- Hutvagner, G., and M. J. Simard, 2008 Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**: 22-32.
- Karolchik, D., A. S. Hinrichs and W. J. Kent, 2012 The UCSC Genome Browser. *Curr Protoc Bioinformatics Chapter 1: Unit1 4*
- Kawamura, Y., K. Saito, T. Kin, Y. Ono, K. Asai *et al.*, 2008 Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**: 793-797
- Klug, K. M., D. Alvarado, M. A. Muskavitch and J. B. Duffy, 2002 Creation of a GAL4/UAS-coupled inducible gene expression system for use in Drosophila cultured cell lines. *Genesis* **34**: 119-122.
- Koonin, E. V., K. S. Makarova and L. Aravind, 2001 Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**: 709-742.
- Kunin, V., and C. A. Ouzounis, 2003 The balance of driving forces during genome evolution in prokaryotes. *Genome Res* **13**: 1589-1594.
- Lal, A., F. Navarro, C. Maher, L. Maliszewski, N. Yan, *et al.* 2009. miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. *Mol Cell*. **35**:610-625
- Li, W., W. Yang and X. J. Wang, 2013 Pseudogenes: pseudo or real functional elements? *J Genet Genomics* **40**: 171-177.
- Liu, Q., T.A. Rand, S. Kalidas, F. Du, H.E. Kim, *et al.*, 2003 R2D2, a bridge between the initiation and effect steps of the Drosophila RNAi pathway. *Science*. **301**: 1921-1925

- Long, M., 2001 Evolution of novel genes. *Curr Opin Genet Dev* **11**: 673-680.
- Long, M., 2000 A new function evolved from gene fusion. *Genome Research*. 1655-1657.
- Lindsley, D.L., and G.G. Zimm, 1992 *The Genome of Drosophila melanogaster*. Academic Press, NY.
- Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151-1155.
- Machida, Y., M. S. Kim and Y. J. Machida, 2012 Spartan/C1orf124 is important to prevent UV-induced mutagenesis. *Cell Cycle* **11**: 3395-3402.
- MacRae, I. J., K. Zhou and J. A. Doudna, 2007 Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* **14**: 934-940.
- MacRae, I. J., K. Zhou, F. Li, A. Repic, A. N. Brooks *et al.*, 2006 Structural basis for double-stranded RNA processing by Dicer. *Science* **311**: 195-198.
- Marques, A. C., J. Tan, S. Lee, L. Kong, A. Heger *et al.*, 2012 Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. *Genome Biol* **13**: R102.
- Marques, J. T., K. Kim, P. H. Wu, T. M. Alleyne, N. Jafari *et al.*, 2010 Loqs and R2D2 act sequentially in the siRNA pathway in *Drosophila*. *Nat Struct Mol Biol* **17**: 24-30.
- McCarthy, C. M., 1984 Fatty acid synthase- an example of protein evolution by gene fusion. *Trends in Biochemical Sciences* **9**: 60-63.

- Meister, G., M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng *et al.*, 2004 Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**: 185-197.
- Miyoshi, K., T. Miyoshi, J. V. Hartig, H. Siomi and M. C. Siomi, 2010 Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA* **16**: 506-515.
- Moustakas, A., and C. H. Heldin, 2009 The regulation of TGF β signal transduction. *Development* **136**: 3699-3714.
- Needleman S.B., and C.D. Wunsch, 1970 A general method applicable to the search for similarities in amino acid sequence of two proteins. *J.Mol Biol.* **48(3)**: 443-453.
- O'donnell, K. H., and P. C. Wensink, 1994 GAGA factor and TBF1 bind DNA elements that direct ubiquitous transcription of the *Drosophila* alpha 1-tubulin gene. *Nucleic Acids Res* **22**: 4712-4718.
- Ohno, S, 1970 *Evolution by Gene Duplication*. Springer-Verlag NY.
- Okamura, K., S. Balla, R. Martin, N. Liu and E. C. Lai, 2008 Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol* **15**: 581-590.
- Okamura, K., W. J. Chung, J. G. Ruby, H. Guo, D. P. Bartel *et al.*, 2008 The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**: 803-806.
- Petrov, D. A., E. R. Lozovskaya and D. L. Hartl, 1996 High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384**: 346-349.

- Pfaffl, M. W., 2001 A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: e45.
- Pfaffl, M. W., G. W. Horgan and L. Dempfle, 2002 Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* **30**: e36.
- Poliseno, L., L. Salmena, J. Zhang, B. Carver, W. J. Haveman *et al.*, 2010 A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**: 1033-1038.
- Preacher, K.J., and N.E. Briggs, 2001. Calculation for Fisher's Exact Test: An Interactive calculation tool for Fisher's exact probability test for 2 x2 tables. [Computer software] <http://quantpsy.org>.
- Proudfoot, N., 1980 Pseudogenes. *Nature* **286**: 840-841.
- Ramakers, C., J. M. Ruijter, R. H. Deprez and A. F. Moorman, 2003 Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* **339**: 62-66.
- Rice, P., I. Longden and A. Bleasby, 2000 EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276-277.
- Robertson, H. M., C. R. Preston, R. W. Phillis, D. M. Johnson-Schlitz, W. K. Benz *et al.*, 1988 A stable genomic source of P element transposase in *Drosophila melanogaster*. *Genetics* **118**: 461-470.
- Rorth, P., 1998 Gal4 in the *Drosophila* female germline. *Mech Dev* **78**: 113-118.

- Roseman, R. R., E. A. Johnson, C. K. Rodesch, M. Bjerke, R. N. Nagoshi *et al.*, 1995 A P element containing suppressor of hairy-wing binding regions has novel properties for mutagenesis in *Drosophila melanogaster*. *Genetics* **141**: 1061-1074.
- Rubin, G. M., M. D. Yandell, J. R. Wortman, G. L. Gabor Miklos, C. R. Nelson *et al.*, 2000 Comparative genomics of the eukaryotes. *Science* **287**: 2204-2215.
- Ryder, E.J. (2004.4.26). Exelixis CG deletion data.
- Sambrook, J., E.F, Fritsch, and T. Maniatis, 1989 *Molecular Cloning Laboratory Manual*. Cold Harbour Laboratory Press, Cold Spring Harbor. NY.
- Sasidharan, R., and M. Gerstein, 2008 Genomics: protein fossils live on as RNA. *Nature* **453**: 729-731.
- Soltis, D. E., and P. S. Soltis, 1999 Polyploidy: recurrent formation and genome evolution. *Trends Ecol Evol* **14**: 348-352.
- Szidony, J., and G. Reuter, 1988 Cytogenic analysis of the echinoid (ed), dumpy (dp) and clot (cl), region in *Drosophila melanogaster*. *Genet. Res.* **51**: 197-208.
- Tam, O. H., A. A. Aravin, P. Stein, A. Girard, E. P. Murchison *et al.*, 2008 Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**: 534-538.
- Tamura, K., S. Subramanian, and S. Kumar, 2004 Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation close. *Mol Biol Evol.* **21**: 36-44
- Tay, Y., J. Zhang, A.M. Thomson, and I. Rigoutsos, 2008 MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**: 1124-1128

- Terracol, R., and J. A. Lengyel. 1994 The thick veins gene of *Drosophila* is required for dorsoventral polarity of the embryo. *Genetics* **138**: 165-178.
- Thompson, T.M., J.J. Lozano, N. Loukili, R. Carrio, F. Serras, *et al.*, 2000 *Genome Res.***10**: 1743-1756.
- Thomson, T.M., J.J. Lozano, N. Loukii, R. Carrio, F. Serras, B. Cormand, M. Valeri, V.M. Diaz, J. Abril, M. Buset, J. Merino, A. Macaya, M. Corominas, and R. Guigo 2000 Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene. *Genome Res.* **10**: 1743-1756
- Tomari, Y., C. Matranga, B. Haley, N. Martinez, P.D. Zamore, 2004 A protein sensor for siRNA asymmetry. *Science.* **306**: 1377-1380.
- Tower, J., G.H. Karpen, N. Craig, and A.C. Spradling, Preferential transposition of *Drosophila* P elements to nearby chromosomal sites. *Genetics* **133**: 347-358
- Tweedie, S., M. Ashburner, K. Falls, P. Leyland, P. Mcquilton *et al.*, 2009 FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic Acids Res* **37**: D555-559.
- Vanin, E. F., 1985 Processed pseudogenes: characteristics and evolution. *Annu Rev Genet* **19**: 253-272
- Venken, K. J., E. Popodi, S. L. Holtzman, K. L. Schulze, S. Park *et al.*. 2010 A molecularly defined duplication set for the X chromosome of *Drosophila melanogaster*. *Genetics* **186**: 1111-1125.
- Watanabe, T., Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa *et al.*, 2008 Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**: 539-543.

- Weichenhan, D., B. Kunze, W. Traut and H. Winking, 1998 Evolution by fusion and amplification: the murine Sp100-rs gene cluster. *Cytogenet Cell Genet* **80**: 226-231.
- Willecke, K., S. Jungbluth, E. Dahl, H. Hennemann, R. Heynkes *et al.*, 1990 Six genes of the human connexin gene family coding for gap junctional proteins are assigned to four different human chromosomes. *Eur J Cell Biol* **53**: 275-280.
- Xia, L., S. Jia, H. Wang, Y. Zui, Y. Mu, *et al.* 2010 The Fused/Smurf complex controls the fate of *Drosophila* germline stem cells by generating a gradient BMP response. *Cell* **143**: 978-990.
- Xu, G., C. Guo, H. Shan and H. Kong, 2011 Divergence of duplicate genes in exon-intron structure. *Proc Natl Acad Sci U S A* **109**: 1187-1192.
- Yano, Y., R. Saito, N. Yoshida, A. Yoshiki, A. Wynshaw-Boris *et al.*, 2004 A new role for expressed pseudogenes as ncRNA: regulation of mRNA stability of its homologous coding gene. *J Mol Med (Berl)* **82**: 414-422.
- Zdobnov, E. M., C. Von Mering, I. Letunic, D. Torrents, M. Suyama *et al.*, 2002 Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**: 149-159.
- Zhang, K., J. M. Akey, N. Wang, M. Xiong, R. Chakraborty *et al.*, 2003 Randomly distributed crossovers may generate block-like patterns of linkage disequilibrium: an act of genetic drift. *Hum Genet* **113**: 51-59.
- Zhang, L., J. Huang, N. Yang, J. Greshock, M.S. Megraw *et al.*, 2006 microRNAs exhibit high frequency genomic alterations in human cancer. *Proc Natl Acad Sci USA*. **103**: 91336-9141.

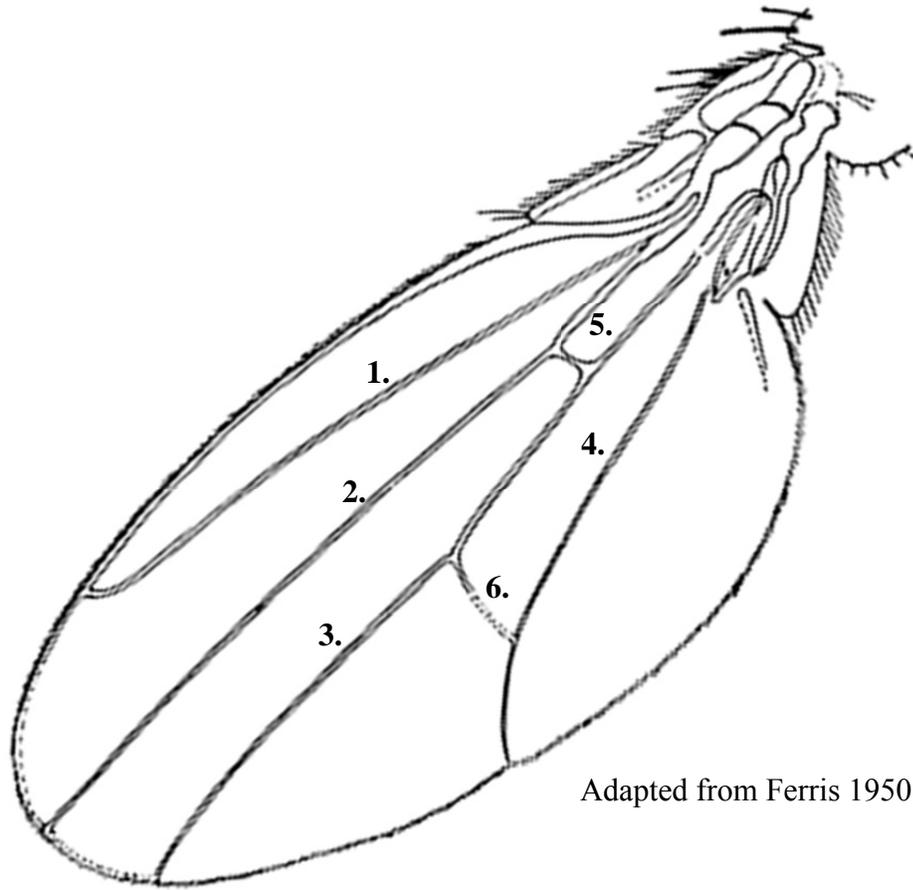
Zheng, D., and M. B. Gerstein, 2007 The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends Genet* **23**: 219-224.

Zhou, R., B. Czech, J. Brennecke, R. Sachidanandam, J. A. Wohlschlegel *et al.*, 2009 Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* **15**: 1886-1895.

Appendix

Appendix 1. Relevant parts of the *Drosophila melanogaster* wing

The relevant anatomy of the *Drosophila melanogaster* wing is shown (Adapted from Ferris 1950). The numbers adjacent to the wing veins correspond to the legend under the figure.



Adapted from Ferris 1950

- 1. wing vein 2
- 2. wing vein 3
- 3. wing vein 4
- 4. wing vein 5
- 5. anterior crossvein
- 6. posterior crossvein

Appendix 2. Thoracic and wing phenotypes associated with complementation tests for 14403

An example of a wild-type wing vein pattern and thorax can be seen in the picture of Canton S (i). Also shown the wing and thorax of a 14403 homozygote (ii). Although it appears as though the L5 wing vein does not connect it does. This effect is due to the lighting conditions and small size of the picture. The other pictures and schematic drawings are the phenotypes associated with crossing 14403 to a *tkv*⁷ mutant strain (3242) (iii) and to two deficiencies of *tkv*, *Df(2L)Exel6011* and *Df(2L)BSC693* (iv and v). The phenotypes seen from progeny of the 14403 crossed to *Df(2L)BSC693* are the same as those seen when crossed to *Df(2L)Exel6011* except for the location of the ectopic wing vein seen at the anterior crossvein rather than the posterior one.

i.



ii.



iii.



iv.



v.



Appendix 3. REST data for the *CRI4033* transgenic flies with LD26673 in the sense direction

For each replicate and gene of interest the data concerning the reaction efficiency, expression, standard error, confidence interval, and p-value is shown. All of the data was determined using the Relative Expression Software Tool (REST) (Pfaffl *et al.* 2002). In the column with the heading “Results” “Up” or “Down” indicate a significant result where $P(H1) \leq 0.05$.

Strain	Replication #	Gene	Reaction Efficiency	Expression	Std. Error	95% C.I.	P(H1)	Result
1-7A testes	1	<i>CG9203</i>	0.948	0.495	0.353 - 0.602	0.339 - 0.630	0	Down
		<i>CRI4033</i>	0.959	2.943	2.197 - 3.490	2.122 - 3.543	0.1	
		<i>tkv</i>	0.932	0.703	0.541 - 0.828	0.516 - 0.837	0.1	
	2	<i>CG9203</i>	0.897	0.936	0.728 - 1.268	0.706 - 1.481	0.756	
		<i>CRI4033</i>	0.854	4.344	3.704 - 5.043	3.603 - 5.627	0.019	Up
		<i>tkv</i>	0.847	1.049	0.923 - 1.274	0.793 - 1.321	0.684	
	3	<i>CG9203</i>	0.943	0.652	0.597 - 0.729	0.537 - 0.756	0.048	Down
		<i>CRI4033</i>	0.961	4.047	3.717 - 4.549	3.572 - 4.737	0	Up
		<i>tkv</i>	0.928	1.833	1.429 - 2.458	1.111 - 2.496	0.063	
1-8A testes	1	<i>CG9203</i>	0.948	1.15	1.046 - 1.266	0.976 - 1.305	0.105	
		<i>CRI4033</i>	0.959	6.526	6.041 - 7.024	5.878 - 7.109	0	Up
		<i>tkv</i>	0.944	1.009	0.921 - 1.085	0.891 - 1.114	0.822	
	2	<i>CG9203</i>	0.95	0.752	0.675 - 0.826	0.640 - 0.890	0.053	
		<i>CRI4033</i>	0.854	3.22	2.700 - 3.885	2.531 - 4.177	0	Up
		<i>tkv</i>	0.847	1.166	1.009 - 1.390	0.883 - 1.481	0.279	
	3	<i>CG9203</i>	0.95	2.03	1.748 - 2.393	1.612 - 2.522	0.088	
		<i>CRI4033</i>	0.955	7.085	7.016 - 7.172	6.947 - 7.198	0	Up
		<i>tkv</i>	0.928	1.459	1.058 - 1.955	0.915 - 2.067	0.224	

Appendix 4. REST data for the *CRI4033* transgenic flies with LD26673 in the antisense direction

For each replicate and gene of interest the data concerning the reaction efficiency, expression, standard error, confidence interval, and p-value is shown. All of the data was determined using the REST program (Pfaffl *et al.* 2002). In the column with the heading “Results” is either an Up, Down or blank. Up and Down signify a significant result where $P(H1) \leq 0.05$.

Strain	Replication #	Gene	Reaction Efficiency	Expression	Std. Error	95% C.I.	P(H1)	Result
2-7A testes	1	<i>CG9203</i>	0.948	0.696	0.506 - 0.951	0.429 - 1.051	0.228	
		<i>CRI4033</i>	0.959	2.77	2.091 - 3.408	1.924 - 3.511	0	Up
		<i>tkv</i>	0.932	1.512	1.161 - 1.781	1.111 - 1.802	0.029	Up
	2	<i>CG9203</i>	0.807	1.098	0.867 - 1.553	0.815 - 1.639	0.802	
		<i>CRI4033</i>	0.864	1.592	1.417 - 1.901	1.390 - 1.949	0	Up
		<i>tkv</i>	0.828	1.284	1.109 - 1.486	0.996 - 1.629	0.107	
	3	<i>CG9203</i>	0.951	0.889	0.744 - 1.143	0.702 - 1.226	0.458	
		<i>CRI4033</i>	0.97	1.718	1.369 - 2.305	1.324 - 2.378	0.019	Up
		<i>tkv</i>	0.904	0.858	0.705 - 1.025	0.627 - 1.155	0.315	
2-24A testes	1	<i>CG9203</i>	0.954	0.155	0.138 - 0.174	0.127 - 0.191	0	Down
		<i>CRI4033</i>	0.893	0.708	0.585 - 0.901	0.471 - 0.984	0.14	
		<i>tkv</i>	0.904	0.858	0.705 - 1.025	0.627 - 1.155	0.315	
	2	<i>CG9203</i>	0.93	0.504	0.435 - 0.557	0.422 - 0.602	0	Down
		<i>CRI4033</i>	0.926	2.226	1.979 - 2.468	1.882 - 2.523	0.06	
		<i>tkv</i>	0.85	1.601	1.465 - 1.776	1.364 - 1.873	0.063	
	3	<i>CG9203</i>	0.945	0.876	0.777 - 0.963	0.756 - 0.987	0.036	Down
		<i>CRI4033</i>	0.952	1.221	1.046 - 1.421	1.000 - 1.553	0.105	

Appendix 5. REST data for the *CG9203* transgenic flies with MIP11944 in the antisense direction

For each replicate and gene of interest the data concerning the reaction efficiency, expression, standard error, confidence interval, and p-value is shown. All of the data was determined using the REST program (Pfaffl *et al.* 2002). In the column with the heading “Results” is either an Up, Down or blank. Up and Down signify a significant result where $P(H1) \leq 0.05$.

Strain	Replication #	Gene	Reaction Efficiency	Expression	Std. Error	95% C.I.	P(H1)	Result
3-31A testes	1	<i>CG9203</i>	0.954	0.558	0.465 - 0.666	0.444 - 0.751	0	Down
		<i>CR14033</i>	0.893	0.689	0.497 - 1.055	0.423 - 1.254	0.223	
	2	<i>CG9203</i>	0.913	2.801	2.594 - 3.020	2.594 - 3.020	0	Up
		<i>CR14033</i>	0.919	0.907	0.827 - 1.002	0.827 - 1.002	0.501	
	3	<i>CG9203</i>	0.945	7.526	6.967 - 8.140	6.967 - 8.140	0	Up
		<i>CR14033</i>	0.952	0.048	0.037 - 0.070	0.037 - 0.070	0	Down
3-38B testes	1	<i>CR14033</i>	0.893	0.106	0.070 - 0.205	0.069 - 0.206	0.074	
		<i>CG9203</i>	0.869	1.5	0.870 - 2.370	0.786 - 3.673	0.352	
		<i>CR14033</i>	0.877	1.116	0.989 - 1.259	0.959 - 1.375	0.385	
	3	<i>CG9203</i>	0.923	2.667	2.059 - 3.795	1.811 - 4.086	0.056	
		<i>CR14033</i>	0.952	0.764	0.562 - 0.913	0.537 - 0.992	0.164	

Appendix 6. REST data for the *CG9203* transgenic flies with MIP11944 in the sense direction

For each replicate and gene of interest the data concerning the reaction efficiency, expression, standard error, confidence interval, and p-value is shown. All of the data was determined using the REST program (Pfaffl *et al.* 2002). In the column with the heading “Results” is either an Up, Down or blank. Up and Down signify a significant result where $P(H1) \leq 0.05$.

Strain	Replication #	Gene	Reaction Efficiency	Expression	Std. Error	95% C.I.	P(H1)	Result
4-2A testes	1	<i>CG9203</i>	0.954	11.092	8.269 - 13.802	7.383 - 15.164	0.032	Up
		<i>CRI4033</i>	0.959	1.356	1.079 - 1.642	0.956 - 1.796	0.164	
	2	<i>CG9203</i>	0.913	6.727	6.390 - 7.154	6.066 - 7.242	0	Up
		<i>CRI4033</i>	0.934	0.238	0.217 - 0.269	0.194 - 0.280	0.033	Down
4-42B testes	1	<i>CG9203</i>	0.954	7.284	6.578 - 8.071	6.463 - 8.231	0.036	Up
		<i>CRI4033</i>	0.959	0.531	0.479 - 0.588	0.452 - 0.629	0	Down
	2	<i>CG9203</i>	0.93	11.147	10.749 - 11.861	10.710 - 11.955	0.029	Up
		<i>CRI4033</i>	0.926	0.921	0.864 - 0.988	0.826 - 1.019	0.254	
	3	<i>CG9203</i>	0.945	11.024	10.638 - 11.482	10.447 - 11.763	0	Up
		<i>CRI4033</i>	0.952	0.838	0.775 - 0.908	0.746 - 0.944	0	Down

Appendix 7. Buffer A Recipe

100mM Tris-Cl (pH 7.5)

100 mM EDTA

100 mM NaCl

0.5% SDS

(Huang *et al.* 2009)

Curriculum Vitae

Education

2012 – Present	Atlantic Veterinary College Class of 2016
2010 – Present	University of New Brunswick, Fredericton Master of Science Candidate
2005 - 2010	Mount Allison University Bachelor of Science in Biology with Honors

Research Interests

- Masters of Science Research
 - My research area is based in evolutionary genetics. My project is to determine if a pseudogene is involved in the regulation of the parent gene function in *Drosophila melanogaster*.
- Honors in Biology Research
 - My project was based in genomic imprinting and it was to determine if *TOP2*, *BEEF*, or *CP190* were involved in the paternal imprint in *Drosophila melanogaster*.

Conferences and Presentations

- 2012 American Drosophila Research Conference
 - Poster Presentation
 - “Genetic analysis of a pseudogene and its parent gene in *Drosophila melanogaster*”
- 2011 Biennial Canadian Drosophila Research Conference
 - Poster Presentation
 - “The role of pseudogenes in regulating parent gene expression in *Drosophila melanogaster*”

Teaching and Laboratory Experience

- Tutoring VBS 101 and 112- Macroscopic Anatomy
- Tutoring VPM 152- General Pathology
- Teaching Assistant for BIOL 1782 and 2721- Human Physiology I and II
- Laboratory Prep Assistant for Biology 2601 (Genetics), Mount Allison University

Publications

MacDonald, W., M. Debashish, N. Bartlett, G. Sperry, V. Rasheva, V. Meller, and V. Lloyd, 2010 The *Drosophila* homolog of the mammalian imprint regulator CTCF, maintains the maternal genomic imprint in *Drosophila melanogaster*. *BMC Biology* 8:105