STUDY OF THE EXPRESSION PATTERN OF A POSSIBLE PSEUDOGENE AND A FUNCTIONAL RETROGENE IN *DROSOPHILA*

by

ADITI BHARDWAJ

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in the Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN BIOLOGY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2005

ACKNOWLEDGEMENTS

I would like to thank Dr. Esther Betrán for guiding me throughout the research and my committee members Dr. James Robinson and Dr. Daniel Formanowicz for their support and guidance. I would also like to thank Javier Río for his help. Last but most importantly I would like to thank my family for their continued support and encouragement; without them this work would not have been possible. This work is dedicated to my family.

November 18, 2005

ABSTRACT

STUDY OF THE EXPRESSION PATTERN OF A POSSIBLE PSEUDOGENE AND A FUNCTIONAL RETROGENE IN *DROSOPHILA*

Publication No. -----------

Aditi Bhardwaj, M.S. Biology

The University Of Texas at Arlington, 2005

Supervising Professor: Esther Betrán

Gene duplication occurs when a mutation leads to the copying of a region of DNA that contains a functional gene. After duplication, there will be two copies of the gene in the genome. Gene duplication can have deleterious effects but it can also be the source of novelty. It has been acknowledged for a long time that selection may be relaxed for one copy rendering that locus free to mutate and discover new functions.

Another of the fates of duplication of special notice is the formation of a pseudogene. Pseudogenes have been defined as nonfunctional sequences of genomic DNA originally derived from functional genes. These are duplications that have undergone disabling mutations.

Pseudogenes exhibit features as premature stop codons and frameshift mutations that prevent them from coding for a functional protein. A particular mechanism of duplication is retroposition. This occurs when a new copy is generated trough an mRNA intermediate. We call these new copies retrogenes. Because of the way they originated, retrogenes will have to recruit new regulatory regions to express.

This work concentrates on the study of two *Drosophila* retrogenes: *CG12628* and *CG2222-like*. *CG12628* is a retrogene in *Drosophila melanogaster* that could be becoming a pseudogene because some alleles of this gene show disablements. *CG2222-like* is a functional retrogene in *Drosophila willistoni.* This work is concentrated in studying the pattern of expression of these genes and their parental genes, which are described in two different sections of this work. An additional section of this work is a compilation of pseudogenes in different species and of functions that have been suggested for these non-coding sequences.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

"Only by accumulating forbidden mutations at the active sites can a gene locus change its basic character and become a new gene locus. An escape from the ruthless pressure of natural selection is provided by the mechanism of gene duplication. By duplication a redundant copy of a locus is created. Natural selection often ignores such a redundant copy and by being ignored, it accumulates formerly forbidden mutations and is reborn as a new gene locus with a hitherto non-existent function".

These are words of Susumu Ohno (35) summarizing the importance of gene duplication for the generation of new functions in his influential book.

Gene duplication occurs when a mutation leads to the copying of a region of DNA that contains a functional gene. After duplication, there will be two copies of the gene in the genome. Gene duplication can have inherent deleterious effects due to the change of dose or it can also be the source of novelty (35). One of the most interesting aspects of gene duplication lies in the fact that when an extra copy of the gene is created it allows one of the gene copies to evolve while the second copy can retain the original function (35).

Generally this extra copy would be ignored but it is most likely to accumulate mutations over a period of time and probably become non-functional and fixed in the genome. In this way duplication provides the prime material from which new gene functions evolve from the already existing ones thus contributing to the diversification of life forms. Gene duplication is considered a major force of evolution. A number of geneticists including Haldane (1932) in his book pointed at the important role of gene duplication in evolution (20). However it was only after scientists sequenced genomes that the degree of importance of duplication in evolution became compellingly evident (35).

Duplication mechanisms

Duplications can occur in different ways and the different mechanisms will often lead to different size of duplications. Whole genome duplication can occur after non-disjunction. Duplications of small or big regions in tandem can occur as a result of unequal crossing over between homologous chromosomes during meiosis. These events will also give rise to the deletion of bases in the genomic sequences of one of the products (27).

A duplication mechanism of interest here is retroposition (Figure 1). The insertion of mRNA back into the genome at random sites after reverse transcription instead of getting translated into a protein is called retroposition (7). The resulting genes formed after duplication are called retrogenes and are characterized by lack of introns, presence of poly-A tail, direct repeats and absence of promoter region. Due to these features they would be expected to be non-functional since they can not be transcribed. However, functional retrogenes have been identified in the genome of most organisms (8).

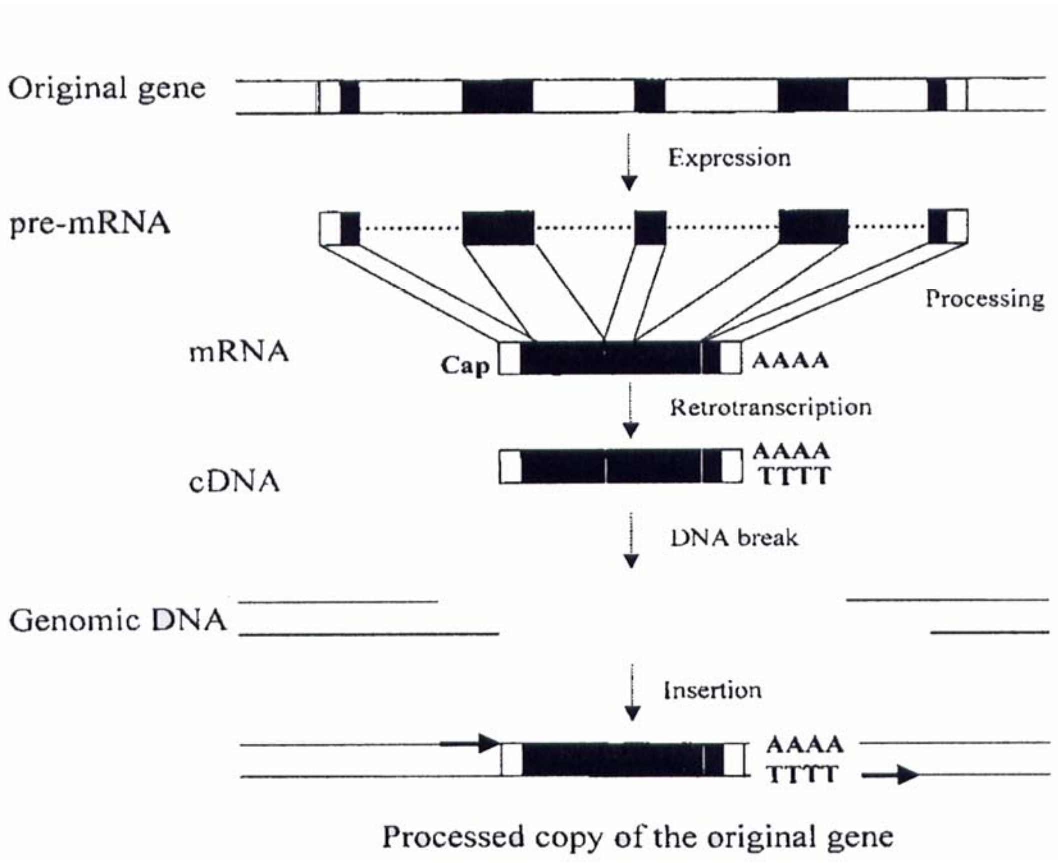


Fig 1: Retroposition mechanism of duplication (Figure courtesy 6)

Fates of gene duplications

Once a gene is duplicated in a genome it must fix in the population to have consequences for evolutionary purposes. Even after attaining a 100% frequency in the population, being preserved in time is also difficult. Many duplicates will degenerate becoming pseudogenes (35). Based on population genetics studies it’s very difficult for a gene to get fixed in the population even if it confers a positive effect to the population and stay there and models like the subfunctionalization model (see below) try to explain why so many duplicates are kept in the genomes. Once fixed, duplicates can meet several fates (31, 35) (Fig.2).

Nonfunctionalization: If after gene duplication, degenerative mutations occur that will usually result in formation of a pseudogene. These mutations are changes that lead to the introduction of a premature stop codon, change of frame or destruction of the protein (31).

A pseudogene can arise from genes that originated from any of the above mentioned duplication mechanisms. Processed pseudogenes are retroposed copies of genes which show disablements (4). Both prokaryote and eukaryote genomes show pseudogenes (4). I review in a section of this work the published work related to the amount of pseudogenes and their roles in the genome. Some of them have been shown to be transcribed and exhibit function.

Neofunctionalization: After duplication, one of the gene copies can change function. For that to occur, and the gene to gain a new function, an increased fixation of amino acid changes after duplication in one of the copies will occur. The signature of that will be a Ka/Ks (nonsynonymous substitutions per non-synonymous site/ synonymous substitutions per synonymous site) ratio bigger than 1. This process is known as neofunctionalization of the gene (31) (Fig.2). Neofunctionalization of a duplicate gene can also occur if one of the genes evolves a new pattern of expression, i.e. it begins expressing in tissues where the previous gene did not express.

Subfunctionalization: As the name suggests it involves partitioning of a particular gene function among the duplicates (31). This model most often refers to the partitioning of the pattern of expression and is an excellent way of optimizing the function of a gene without compromising its original function. Partitioning is advantageous in genes with multiple functions since it may increase fitness between the subfunctionalized duplicates by ending conflict between the functions (31), (Fig.2).

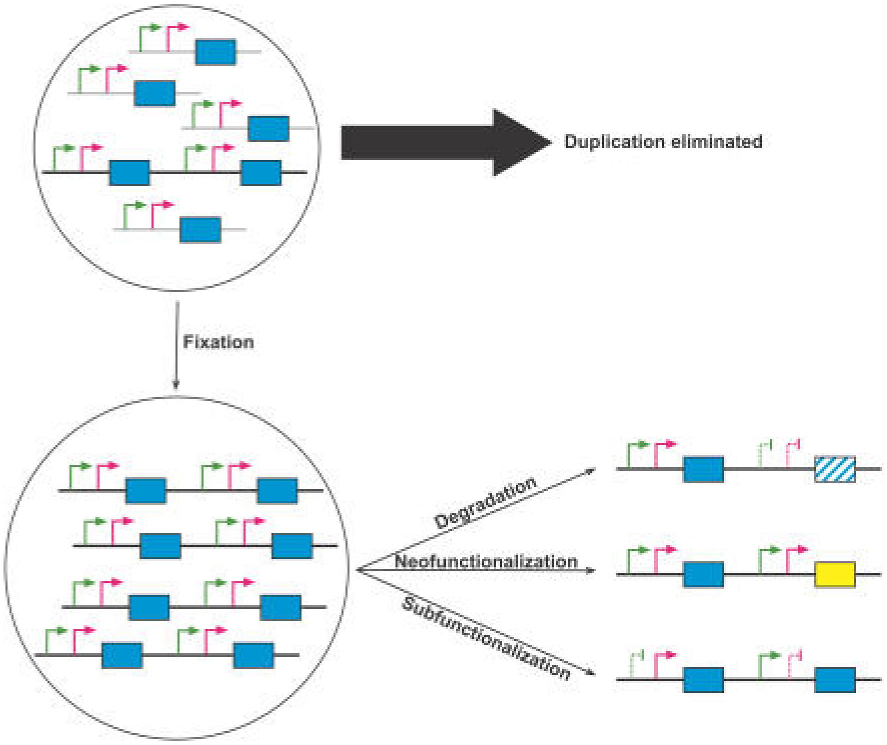


Fig 2: Fates of duplication. A new duplication in a gene (blue) with two tissue-specific promoters (arrows) arises in a population of single copy genes. Fixation within the population results in a minority of cases. After fixation, one gene is inactivated (degradation) or assumes a new function (neofunctionalization), or the expression pattern of the original gene is partitioned between the two duplicates as one promoter is silenced in each duplicate in a complementary manner (subfunctionalization) (Figure courtesy 31)

Duplication to produce more of the same:

New gene duplicates will be kept in the genomes if there is a need for more of the same product (35). There are very well known cases of this type of gene families. For example, there are tandemly organized genes for rRNA precursors.

Duplication to attain permanent heterozygous advantage:

There are examples of attainment of a permanent heterozygous advantage by the incorporation of two former alleles into the genome. One of the best documented examples is one on color vision genes. Positive selection is believed to have fixed duplication in order to achieve trichromatic vision in humans through which they can distinguish three colors. Humans possess three color sensitive pigment proteins; blue, red and green. The blue pigment is coded by an autosomal gene and red, green by X-linked gene. The red and green pigments are nearly identical (96% identical) and closely linked which indicates recent gene duplication (35), (30).

The organism of study: *Drosophila*

Drosophila is a genus composed by around 2000 different species. These are different species of fly present all over the world. Some of the species have cosmopolitan distribution and some have very narrow distribution like the Hawaiian *Drosophila*. In this thesis we work with two different species *D. melanogaster* and *D. willistoni* (38). A phylogeny showing the relationships between these two species and their divergence times is shown in figure 3**.**

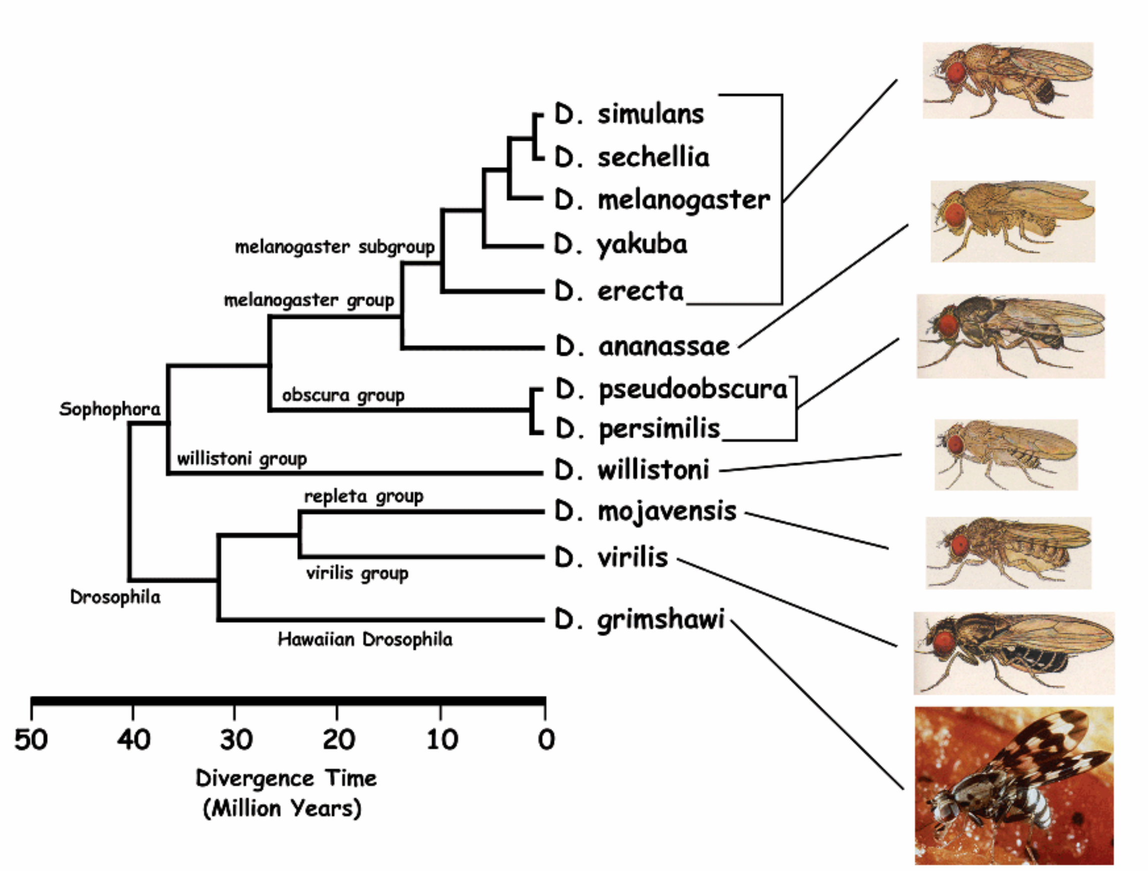


Fig. 3: Phylogeny of Drosophila species. (Figure courtesy 13)

Drosophila melanogaster

Drosophila melanogaster is a small fly of 3mm in length which is normally seen around spoiled fruits. *D. melanogaster* exhibits sexual dimorphism (49). On a visual level, females are larger, their abdomen is lighter compared to males and V-shaped. Males, on the other hand, are smaller than females, have sexual combs, one each on the anterior pair of legs which happens to be the most distinguishing mark between the male and female and their abdomen is darker and has round shape. Their genitalia is also different.

The fly is considered a model organism and has been valuable in biological research especially in genetics and developmental biology. Many *Drosophila* genes have homolog in humans which makes it a model organism for the study of some human diseases (48). Christiane Nusslein-Volhard, Ed Lewis and Eric Wieschaus received the Nobel Prize in 1995 in medicine/physiology for their study of "the genetic control of early embryonic development" in fruit fly further validating its importance in research (49).

The *Drosophila* genome has been sequenced (1), (33). It was the first eukaryotic organism in which the shotgun approach was used. The genome consists of four pairs of chromosomes, the X/Y sex chromosomes and the autosomes 2, 3 and 4 (Fig. 4). The size of the euchromatic portion of the genome is about 165 million bases and contains around 14,000 genes (4), (30).

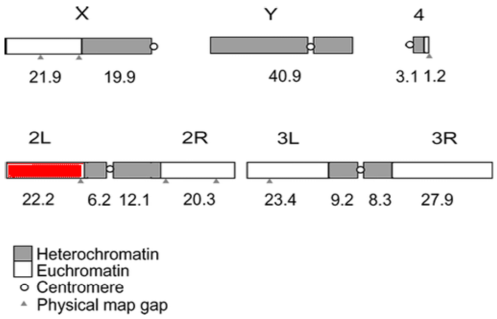


Fig 4: Chromosome structure of *Drosophila melanogaster*. (Figure courtesy 13)

Retrogenes in ***Drosophila***

Processed copies of genes (retrogenes) are rare in *Drosophila* with respect to other organisms, especially vertebrates. Several retroposed genes have been described in *Drosophila*. One of them is an interesting one which has been described: a new chimerical gene called *jingwei* (*jgw*; Long and Langley 1993). This gene is located on chromosome 3 in *D. teissieri* and *D. yakuba* and it is not present in their closest relatives.

The age of the gene has been estimated to be less than 2.5 My (Long et al. 1999). *Jgw* cDNA has been examined. *Jgw* has a 3' end *Adh*-like and a 5'end (3 exons) recruited from another gene *yellow emperor* (*ymp*; Long and Langley 1993; Wang et al. 2000). Another functional processed protein gene has been described in *Drosophila melanogaster* (Yuan et al. 1996): *Pros28.1A*. *Pros28.1A* is an intronless copy of *Pros28.1*. The new copy shows 74% identity at amino acid level with the parental copy but is located on a different chromosome. While *Pros28.1*maps in 14B4 on chromosome X, *Pros28.1A* is located in 60D7 on chromosome 2. *Pros28.1* transcription is detected during all *Drosophila* stages from embryo to adult (male and female). Unlike *Pros28.1, Pros28.1A* is expressed only in males and very specifically in germline during spermatogenesis. Recently, another chimeric gene derived from an *Adh* retrogene, *Adh-Twain*, was described. *Adh-Twain* is expressed and it is evolving under positive selection (Jones, Custer, Begun, 2005)

Betrán et al identified 24 young retrogenes and their parental genes with 70% amino acid identity or more and located on different chromosomes (8). These 24 pairs also showed typical features suggesting retrotransposition. This study involves research on one of these retrogenes Betrán et al. (8) identified in *D. melanogaster*: *CG12628* (see below). All of these genes with possible exception of *CG12628* (see results) are genes that are expressed and show constraint at the sequence level revealing that new genes were generated by retroposition.

The gene *CG12628* and its parental gene *Mgstl*

CG 12628 was described as a young retrogene by Betrán et al. (8) and it is only present in *D. melanogaster* genome. Figure 5 shows an alignment with the parental gene (*CG1742/ Mgstl*). Being the youngest retrogene described in this genome, it still keeps the features of the retroposition event: it is intronless, has a poly-A tail and it is flanked by short repeats (Figure 5). The direct repeats are clear: the region of homology with the parental gene is flanked by a 5 base pair direct repeat (AATCA), which also points towards a recent insertion event. Given the divergence times of *D. melanogaster* and the closest related species, *CG12628* should be less than 1.5-2 million years old.

Thirty-three alleles of this gene have been sequenced (8). Compared to the parental gene, twelve of the alleles of *CG12628* show a premature stop codon, (TGA) instead of Trp (TGG), and few other alleles show one base pair deletion. Despite this, *CG12628* was annotated in releases 1 and 2 of the *Drosophila* genome as a functional gene. Supporting this last view, a large proportion (60.61%) of alleles maintains an intact reading frame. Furthermore, nonsynonymous polymorphism is lower than synonymous polymorphism in both the normal alleles and the truncated alleles in which a shorter predicted open reading frame (ORF) remains. In addition, Ka/Ks is smaller than 1 (0.5370) although not significantly smaller. Thus, the functional role for this retrogene cannot be ruled out (8).

CG12628 and its parental gene (*CG1742*/ Microsomal glutathione S-transferase-like, *Mgstl*, gene) encode for microsomal glutathione S-transferases. Microsomal glutathione S-transferases have important protective functions and protect the cell from oxidative damages, and toxin response (52). *Mgstl* has homologue in humans (52). Experiments conducted by Toba and Aigaki (52) demonstrated that *Mgstl* is expressed in all developmental stages and in almost every part of the body including microsomes and outer membrane of mitochondria with the highest expression in the larval fat body, an insect organ which is supposed to be functionally corresponding to mammalian liver (52). It is not necessary for development as shown by the experiments on mutants which showed no defects in morphology. On the other hand, the life span of experimental flies was reduced compared to control flies, suggesting that the *Mgstl* protein is involved in aging processes (52). This is perhaps the first molecule identified in invertebrates that is similar to mammalian membrane-bound GSTs and has been localized to X chromosome 19E (52), (2).

Glutathione *S*-transferases (GSTs) are a family of enzymes that catalyze the conjugation of glutathione to a variety of electrophilic substances including carcinogenic, mutagenic, and toxic compounds (2). Isoenzymes are structurally related forms of enzyme which act in a similar way but have different chemical, physical or immunological characteristics. They can be either cytoplasmic (cytosolic) or membrane bound. The membrane-bound enzymes have an evolutionarily different origin from that of the cytosolic enzymes, and three enzymes, microsomal GST (MGST) -I, -II, and -III, have been identified in humans (2).

Drosophila willistoni

Drosophila willistoni is found in the New World tropics breeding in large and small rotting fruits (38). Males and females show size differences as in *D. melanogaster* but they are paler. This last feature makes differentiating both sexes more difficult. It diverged from *D. melanogaster* lineage around 40 million years ago (Figure 3). It is one of the 10 species of *Drosophila* whose genome is being sequenced in addition to the published genomes of *D. melanogaster* and *D. pseudoobascura.* At this point only traces of the genome are available but have been very useful for the study carried in the lab for this gene (see below).

The gene *CG2222* and its retrogene *CG2222-like*

In 2002, Bergman *et al.* while comparing the genomic sequences from 4 *Drosophila* species, *D. erecta*, *D. pseudoobscura, D. willistoni*, and *D. littoralis* with *D. melanogaster* as the reference species observed a sequence in the even skipped region (*eve*) unique to *D. willistoni* (5). This novel gene was called *CG2222*-like gene because it was inferred to have originated from a retrotransposition event as it lacks introns while its closest homolog, found on a different chromosome arm, X in the *D. melanogaster* genome (*CG2222*), has two introns (5). *CG2222-like* is located on chromosome arm 2R (Muller element B) in *D. willistoni*. The function of this gene or its parental gene is not known yet and they remain genes annotated by their CG number. *CG2222-like* in *D. willistoni* and *CG2222* in *D. melanogaster* show an identity at the amino acid level of 79.5%.

Betrán's laboratory has worked on this gene and it's parental. The following results are unpublished data from this project to which I am also contributing. Jamie L. Dunlop and Esther Betrán began to work with *CG2222-like* and *CG2222* in *D. melanogaster*. The gene *CG2222* is present in *D. melanogaster, D. yakuba, D. erecta, D. ananasae, D. pseudoobscura, D. mojavensis* and *D. virilis.* These species appear not to have *CG2222-like* copy (5 and Betrán lab data). This makes *CG2222-like* less than 40 million years old. These are data obtained from blast results in sequenced genomes.

Yongsheng Bai provided hits of *CG2222-like* in the traces of the genome sequence of *D. willistoni* that enable the finding of *CG2222* in *D. willistoni* (Figure 5 to 7) Jamie L. Dunlap and I studied the expression pattern of *CG2222* in *D. melanogaster* and *D. willistoni* and *CG2222-like* in *D. willistoni* using male and female RNA and revealed that all of them are expressed in male and in female. This reveals that the retrogene *CG2222* is expressed; i.e. it recruited new regulatory regions. Constraint in the sequences was analyzed using PAML software. If nonsynonymous sites (K_A) evolve significantly slower than synonymous sites (K_S) is support for functionality of the proteins. K_A/K_S ratio is 0.0377 in *CG2222* on average and 0.0900 in *CG2222-like* lineage supporting the functionality of both genes.

```
CG2222D.melanogasterGTAAACACCAAACGAATCATAAAGCTG-GGCCGCG----CAATGAACGGA
CG2222D.willistoni  GCAAATCTGTGTGGAACTTAGAAGCGAAATTT------TATCAGATGGCA
CG2222-like         --------------------------------------------------

CG2222D.melanogasterATGAACTATTTTCCGAACTATTACTCCATCGAGGACATATTCGTGACCCA
CG2222D.willistoni  GTCAATTATTTTCCCCATTACTACTCGATTGAAGATATTTTTGTCACCCA
CG2222-like         ATGAATTATTTTCCACATTACTACTCCATCGAGGACATTTTCGTGACGCA
                     * ** *  *****  * ** ** ** ** ** ** ** ** ** ** **

CG2222D.melanogaster GAGAAGGTGGAATGTCGGGTGAACACCAAGCTGCAGCGGATGGGTAAGT
CG2222D.willistoni   GAGAAAGTGGAATGCAAGGTTAACACAAAGCTACAGCGGATGGGTAAGT
CG2222-like          AGAAAAGGTGGAGTGCAGGGTTAACACAAGGCTGCAGCGCATG------
                        ** *  * *     **   *    ***            *

CG2222D.melanogasterGATTGGCTTGGAC------GGACA--CGCACATTCTGGCCAG-TAATCCG
CG2222D.willistoni  AAATGTAAACACTCT----------ATGAACAACAACAGCAAAGACGTTT
CG2222-like         ---------- ------------ --------------------------


CG2222D.melanogasterACAATATTGAGTTGCTTGCAGGATTCCTGGACTCCGGC---GCGGAATCG
CG2222D.willistoni  AAAACATTT-TTTTTGAACAGGGTTCCTAGATGCAGGC---TCTGAAACT
CG2222-like         ---------------------GATTCTTAGATGCGGGGGACTCGCAAACT
                                         * *** * **              ** *

CG2222D.melanogasterGATGACCTGGAGCCCGGCAGGACGGTCAATCTGCCGTTGTGGTACATCAA
CG2222D.willistoni  GACCATCTGGAACCGGGACGCACCGTCAACTTGCCTCTATGGTACATTAA
CG2222-like         GACCACTTGGAACCGGGACGCATTGTCAACTTGCCTCTCTGGTACATTAA
                    **  *  **** **     * *   * **  ****  * ***** ** **

CG2222D.melanogasterGGAGCTGAAGGTAAACAATGCCTACTTCACCGTCGCCGTGCCAGATATCT
CG2222D.willistoni  AGAACTCAAAGTTAGTAACGCCTACTTCACAGTTTGTGTGCCCGAAATCT
CG2222-like         AGAGCTAAAGGTAAACAATGCTTATTTCACCGTATGCATTCCGGAAATCT
                     ** ** ** ** *  **  * ** *****  *     * ** ** ** *

CG2222D.melanogasterATCGCAATGTGCATAAGGCCGTCTGCGAGGCGGAGACCACGCACATCGAG
CG2222D.willistoni  ATAAAAATGTGCACAAGGCCGTTTGCGAGGCAGAAACAACCCACATCGAG
CG2222-like         ACAAAAATGTACACAAGGCCGTCTGCGAGGCAGAAACAACTCACATTGAA
                     *    ** ** ** ******** ** ***** ** ** ** ***** **

CG2222D.melanogasterCTGGGACGCCTGCACCCGTACTTCTACGAGTTCGGTCGCTATCTGACGCC
CG2222D.willistoni  CTAGGCCGTTTACATCCCTATTTCTATGAATTCGGCCGCTATCTAACACC
CG2222-like         CTAGGGCGCTTACATCCTTATTTCTACGAATTCGGTCGCTATTTGACACC
                    ** ** **  * ** ** ** ***** ** ** ** **  * ** ** **
```

Fig 5: Sequence alignments of *CG2222* and *CG2222-like* in *D. willistoni* and *CG2222* in *D. melanogaster*. Sequence courtesy Y.Bai and J.Dunlap

```
CG2222D.melanogasterCTACGATCGCAACCATGTCATCGGCCGCATCATCTTCGAAACGCTGCGCC
CG2222D.willistoni  CTATGATAGCAATCATGTCATAGGCCGTATAATATTTGAGACGCTTCGCC
CG2222-like         CTATGATAGCAATCATGTCATAGGGCGAATAATCTTCGAGACACTGCGTC
                    *** **  **** ** **  * ** ** ** ** ** ** **  * ** *

CG2222D.melanogasterGTAAACACCAAACGAATCATAAAGCTG-GGCCGCG----CAATGAACGGA
CG2222D.willistoni  GCAAATCTGTGTGGAACTTAGAAGCGAAATTT------TATCAGATGGCA
CG2222-like         --------------------------------------------------
CG2222D.melanogasterATGAACTATTTTCCGAACTATTACTCCATCGAGGACATATTCGTGACCCA
CG2222D.willistoni  GTCAATTATTTTCCCCATTACTACTCGATTGAAGATATTTTTGTCACCCA
CG2222-like         ATGAATTATTTTCCACATTACTACTCCATCGAGGACATTTTCGTGACGCA
                     * ** *  *****  * ** ** ** ** ** ** ** ** ** ** **

CG2222D.melanogasterGGAGAAGGTGGAATGTCGGGTGAACACCAAGCTGCAGCGGATGGGTAAGT
CG2222D.willistoni  CGAGAAAGTGGAATGCAAGGTTAACACAAAGCTACAGCGGATGGGTAAGT
CG2222-like         AGAAAAGGTGGAGTGCAGGGTTAACACAAGGCTGCAGCGCATGG-----
                     ** ** ** ** **    ** ** **    ** ** **  ***

CG2222D.melanogasterGATTGGCTTGGAC------GGACA--CGCACATTCTGGCCAG-TAATCCG
CG2222D.willistoni  AAATGTAAACACTCT----------ATGAACAACAACAGCAAAGACGTTT
CG2222-like         ---------- ------------ --------------------------


CG2222D.melanogasterACAATATTGAGTTGCTTGCAGGATTCCTGGACTCCGGC---GCGGAATCG
CG2222D.willistoni  AAAACATTT-TTTTTGAACAGGGTTCCTAGATGCAGGC---TCTGAAACT
CG2222-like         ---------------------GATTCTTAGATGCGGGGGACTCGCAAACT
                                          *             *   *  *

CG2222D.melanogasterGATGACCTGGAGCCCGGCAGGACGGTCAATCTGCCGTTGTGGTACATCAA
CG2222D.willistoni  GACCATCTGGAACCGGGACGCACCGTCAACTTGCCTCTATGGTACATTAA
CG2222-like         GACCACTTGGAACCGGGACGCATTGTCAACTTGCCTCTCTGGTACATTAA
                    **  *  **** **     * *   * **  ****  * ***** ** **

CG2222D.melanogasterGGAGCTGAAGGTAAACAATGCCTACTTCACCGTCGCCGTGCCAGATATCT
CG2222D.willistoni  AGAACTCAAAGTTAGTAACGCCTACTTCACAGTTTGTGTGCCCGAAATCT
CG2222-like         AGAGCTAAAGGTAAACAATGCTTATTTCACCGTATGCATTCCGGAAATCT
                     ** ** ** ** *  **  * ** *****  *     * ** ** ** *

CG2222D.melanogasterATCGCAATGTGCATAAGGCCGTCTGCGAGGCGGAGACCACGCACATCGAG
CG2222D.willistoni  ATAAAAATGTGCACAAGGCCGTTTGCGAGGCAGAAACAACCCACATCGAG
CG2222-like         ACAAAAATGTACACAAGGCCGTCTGCGAGGCAGAAACAACTCACATTGAA
                     *    ** ** ** ******** ** ***** ** ** ** ***** **

CG2222D.melanogasterCTGGGACGCCTGCACCCGTACTTCTACGAGTTCGGTCGCTATCTGACGCC
CG2222D.willistoni  CTAGGCCGTTTACATCCCTATTTCTATGAATTCGGCCGCTATCTAACACC
CG2222-like         CTAGGGCGCTTACATCCTTATTTCTACGAATTCGGTCGCTATTTGACACC
                    ** ** **  * ** ** ** ** ** ** ** ** **  ** * ** **
```

Fig 6: Sequence alignments of *CG2222* and *CG2222-like* in *D. willistoni* and *CG2222* in *D. melanogaster* continued. Sequence courtesy Y.Bai and J.Dunlap

```
CG2222D.melanogasterCTACGATCGCAACCATGTCATCGGCCGCATCATCTTCGAAACGCTGCGCC
CG2222D.willistoni  CTATGATAGCAATCATGTCATAGGCCGTATAATATTTGAGACGCTTCGCC
CG2222-like         CTATGATAGCAATCATGTCATAGGGCGAATAATCTTCGAGACACTGCGTC
                    *** **  **** ** **  * ** ** ** ** ** ** **  * ** *

CG2222D.melanogasterAGAGGGTACGTCATCTCCTGGACATCTCGAAGAGCGACGGA------CAG
CG2222D.willistoni  AGCGAGTACGTTATCTGTTAGACATATCGAAAAATGATGTTCAGAGCAGG
CG2222-like         AACGAGTGCGTCATTTGCTGGACATATCAAAGAATGATGACCAGAGCAGG
                     *  * ** **  *  *  * ***** ** ** *  **


CG2222D.melanogasterGCAGCCAAGGCGGAGCATCGACTGGATAACATCGAGGCCAAGCTGCACGA
CG2222D.willistoni  AGTAGTAAACCGGAACATCGTTTGGACAACATCGAGGCCAACCTGCATGT
CG2222-like         ATCAGCAAGACGGAACATCGACTGGATAACATTGAGGCCAATCTTCATAT
                           **    ** *  **  * ** ***** **    ** ** **

CG2222D.melanogasterGGCTGGAGTGCGCACCAACAGTCAGGTGGGT-------------------
CG2222D.willistoni  CGCAGGCACTAAAACAAATACTCAGGTAATG-------------------
CG2222-like         AGCGGGTATGAAAACTAATGCGCAG-------------------------
                     ** **       ** **    * *

CG2222D.melanogaster------GGCTGTACTACCAGTCAAGTGGGCA-----CGGAGT-----GTT
CG2222D.willistoni  -------AGTTTA--ACTGGAGTTACGCCCA----TCTCTCT-----ACT
CG2222-like         --------------------------------------------------


CG2222D.melanogasterATATTAATCACATA---ATTTCAGTACATCGAATGGCTGCAGATGACGGG
CG2222D.willistoni  TTATTTAAAATGTTC---TTTCAGTATATCAATTGGTTGCAGATGTGTAG
CG2222-like         ------------------------TATGTCAAATGGCTGCAGATGTCCAA
                                             *  * *** **** ***

CG2222D.melanogasterCAACAAGATTCGCACCTCGGAGCTGGTCGAGGAGCACCAAAAGAAGCGGC
CG2222D.willistoni  CAATAAGATATGCATCTCCCAATTAGTGGAGGAGCATCAGAAGAAGCGCA
CG2222-like         CAATAACATTTGCATCTCTGAATTGGTGGAGGAGCATCAGAGGAAACGTT
                     ***  *  *    * ***  *  * ** ******** ** * *** **

CG2222D.melanogasterGACGTGCGGATCGCAGCGATGACGAGGGCGATGCTCTGCCCAACAGCAAG
CG2222D.willistoni  AGCGTGCCGATTGCAGTGATGATGAAAACGACAGCCACCCTGCCAGTAAA
CG2222-like         TGCGGGTTGATCATAGCGATAAACAAATAGAAAGCTCACCTCCGTAA---
                       ** *   *    ** **  *  *    *        *

CG2222D.melanogasterCGGGCCACCTTGTAA-----------CAAGCAG-----------------
CG2222D.willistoni  CGAGCTACTTTGTGATGCTTGAAACTGAAATGTCTTATATACAAA---GT
CG2222-like         --------------------------------------------------
```

Fig 7: Sequence alignments of *CG2222* and *CG2222-like* in *D. willistoni* and *CG2222* in *D. melanogaster* continued. Sequence courtesy Y.Bai and J.Dunlap

CHAPTER II

AIMS OF THE WORK

2.1 Expression study of *CG12628* in stop codon and non-stop codon alleles

CG12628 is a young retrogene in which some alleles show a stop codon with respect to the parental gene *Mgstl*. One of the alleles is the one that was sequenced in the Drosophila genome and led to the annotation of this gene as a pseudogene in release 3 of the genome. Few other alleles show deletions but large proportions (60.61%) of alleles maintain an intact reading frame. These particular data raises a question. Is *CG12628* a functional gene? We have 3 different possibilities or hypothesis to work with. [1] It could be a functional protein coding gene. [2] It could be a functional non-coding gene. Or [3] it could be a dying gene; i.e. a gene that is becoming a pseudogene.

If it is a functional protein coding gene, there are again several possibilities. It could be producing a shorter protein compared to the parental gene; for example, by using another translation initiation site seven codons after the stop codon for example. That will mean that all the alleles would be functional even the ones containing stop codon. Or it could be producing the protein of the same length of the parental gene and in this case only 60% of the alleles are functional. Even if the mutant is recessive, this scenario leaves us with 16% (2x0.4.x0.6x100) of the individuals being homozygous for the mutant allele.

If there was strong selection against these homozygous individuals we could observe mutant alleles but only in heterozygosity. On the contrary we observe 8 homozygotes for the stop codon when sequencing PCR product from single fly. One aspect is important to mention. It is well known that functional genes can have nonsense mutations segregating in the populations (40). It is not very frequent but it can happen and it occurred for the parental gene in the Rinanga strain. Out of 15 alleles, the Rinanga allele shows an 11bp deletion. The gene is in the X chromosome and we sequenced males confirming the absence of the parental gene function as non-deleterious. However, the mutants of the parental gene are known to reduce life span (52). In this last scenario, functional full length protein, we expect phenotypic differences in the stop codon lines vs. non-stop codon. This would support obtaining these two types of lines: stop and non-stop codon lines. In any case, if one of these two options is what is happening, the gene should produce a transcript.

Another reason to obtain stop and non-stop codon lines is that there is nonsense mediated decay in Drosophila (40). This is a mechanism that detects premature stop codons and degrades the transcript. This is likely to happen in the stop codon allele if transcribed and will prevent us from seeing the transcript. Because of this it is advisable to look at expression separately in the different lines. The existence of this phenomenon also changes our expectations because it could be difficult to see the transcript even if it is produced in the stop codon line.

The second scenario is if the gene is a functional non-coding gene. There have been recent reports suggesting that pseudogenes can have regulatory functions. As you will see in the third section of this thesis I have compiled all the available data in this respect. This function will also require the gene to produce either a sense or an antisense transcript but the transcript does not need to be encompassing the complete coding region. However, it will need to be expressing in an overlapping pattern with the parental gene.

The last possibility is that the gene is becoming a pseudogene. In this scenario, it could be transcribing or not. If it is not transcribing, it will be evolving neutrally and the stop codon that we see could fix or not in the population. If it is expressed and becoming a pseudogene it could become a pseudogene under positive selection if it is beneficial to get rid of the gene or it could be evolving neutrally. Again knowing if the gene is expressed and its pattern of expression respect to the parental gene will be important for this hypothesis.

As introduced in the above hypothesis it would be interesting to obtain stop and non-stop codon lines and study in both the expression of the gene. Retrogenes are shown to express in adulthood with a bias for male and female germline and both the sexes at this life stage will be targeted. One additional comment is that we expect expression analysis to the difficult. The gene is, in general, difficult to work with since it has a 98% base pair identity with the parental gene (*CG1742*/*Mgstl*) (Fig. 8). Primers were designed with this in mind and assayed in genomic DNA to probe that they are working (Table 1).

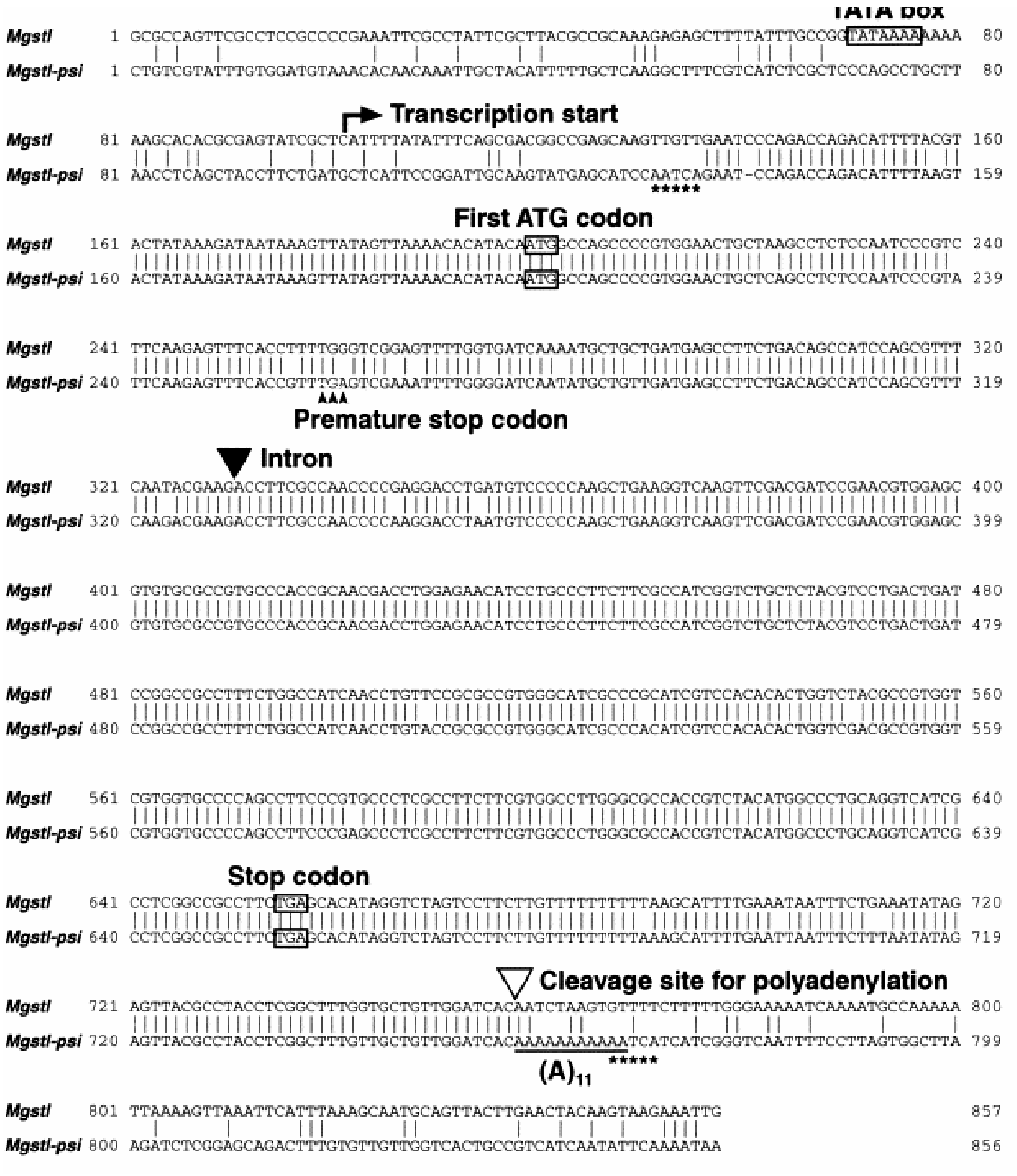


Fig 8: Sequences of *Mgstl* and *CG12628* compared. (Figure courtesy 52)

2.2 Expression study of *CG2222* and *CG2222-like*

CG2222 and *CG2222-like* are two functional genes as revealed by the sequence analysis and expression. Both show constraint and express in male and female. *CG2222-like* was derived from a retroposition event from *CG2222*. As outlined above, when a new duplicate originates its fate is diverse. Several things can be examined to explore what is the fate of *CG2222-like*: [1] how fast the protein has changed with respect to the parental copy and [2] how overlapping the pattern of expression is with the parental gene.

At this point it is known that the protein is changing at similar rates (see above) but little information is available about the detailed pattern of expression of both genes. It has often been seen for retrogenes that the parental gene is widely expressed and the new gene is expressed in male and/or female germline only (4), (5). I will study in more detail the pattern of expression of the two genes in several species to see if neofunctionalization or subfunctionalization could explain the fate of this gene.

2.3 Review of pseudogenes literature

As mentioned above there are recent compelling data that indicate that often pseudogenes are expressed and could produce transcripts that help regulate the parental gene. I review all this evidence in this section of the thesis (Table 2 and 3).

CHAPTER III

MATERIALS AND METHODS

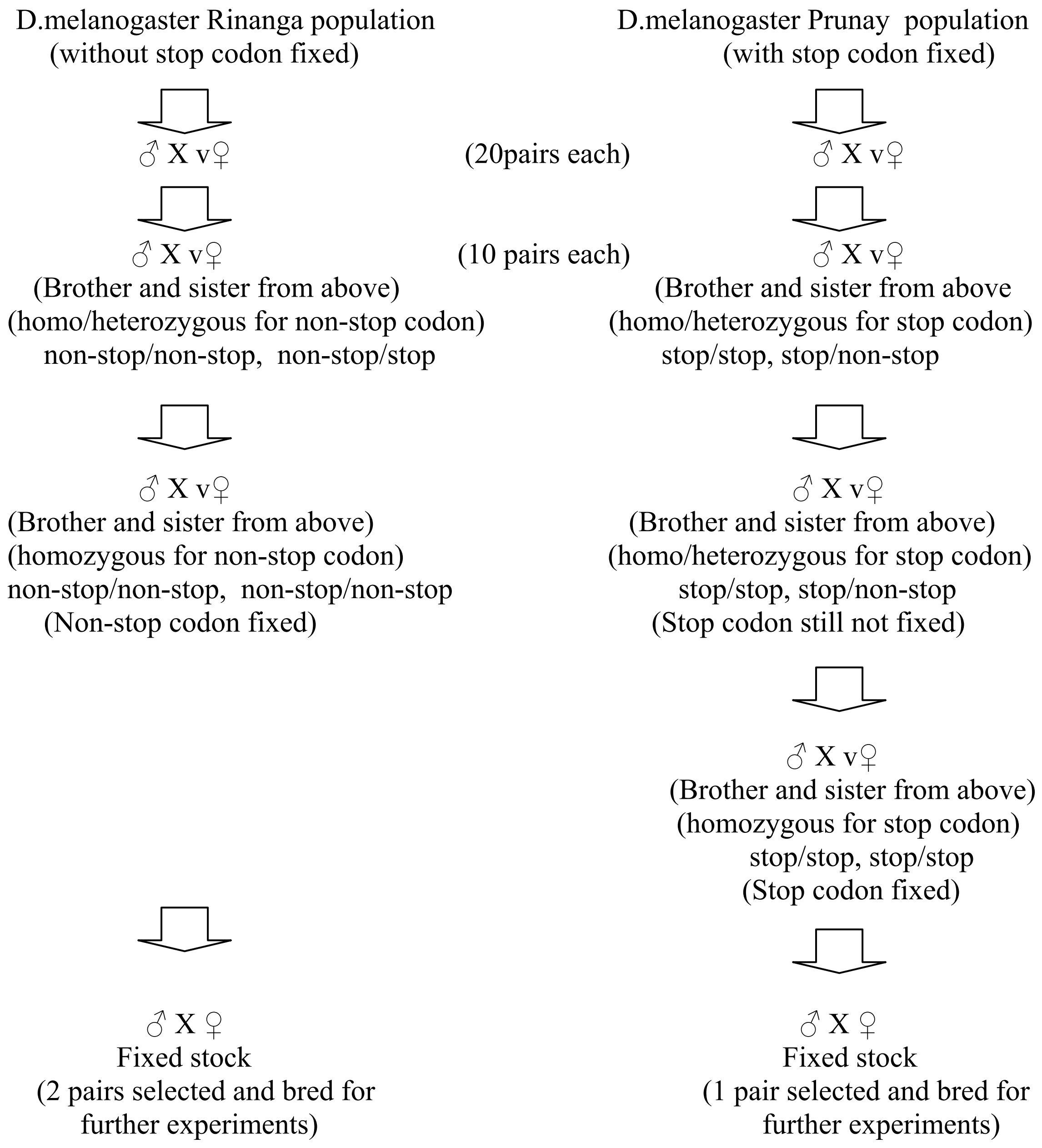
3.1 For retrogene *CG12628*

Drosophila melanogaster strains, crosses and growth conditions

D. melanogaster flies used in this study were obtained from Rinanga and Prunay regions of France by Patricia Gilbert. These are strains that come from pooling inseminated females from the particular location and can have a lot of variation. We use these two stocks to obtain stop and non-stop lines. An individual from Prunay strain was observed to be homozygote for stop codon (8) and an individual from Rinanga strain was observed to be homozygote for non-stop codon. So we decided to use these strains that were likely to be variable but where the two alleles had been observed to be fixed in homozygosis.

Brother-sister crosses were carried out to fix the alleles. We started with 20 pairs of each population by crossing males and virgin females. The strategy of obtaining crosses fixed for both stop and non stop codon is shown in Figure 9. A typical fly takes 72 hours to become a pupa from larvae and another 72 hours to hatch. The females require 7 hours to mature completely after which they can mate (49). Thus, care was taken to separate female flies' immediately after hatching or before they could mate.

Crosses and strains were routinely grown in corn media and the flies were maintained at room temperature and kept in the 20°C incubator at night. Every pair was allowed to lay eggs and then genotyped for the presence or absence of a stop codon (see below). Some pairs or single individuals were lost and could not be genotyped. After genotyping, offspring of the pairs enriched for the particular two alleles were used to produce from 10 to 20 next generation pairs of brother-sister crosses. These crosses were again genotyped and procedure was repeated until pairs with two homozygote individuals for the same alleles were found leading to the actual fixed lines.



D.melanogaster Rinanga population
(without stop codon fixed)

♂ X v♀

(20pairs each)

♂ X v♀

(10 pairs each)

(Brother and sister from above)
(homo/heterozygous for non-stop codon)
non-stop/non-stop, non-stop/stop

♂ X v♀
(Brother and sister from above)
(homozygous for non-stop codon)
non-stop/non-stop, non-stop/non-stop
(Non-stop codon fixed)

♂ X ♀
Fixed stock
(2 pairs selected and bred for
further experiments)

D.melanogaster Prunay population
(with stop codon fixed)

♂ X v♀

♂ X v♀
(Brother and sister from above
(homo/heterozygous for stop codon)
stop/stop, stop/non-stop

♂ X v♀
(Brother and sister from above)
(homo/heterozygous for stop codon)
stop/stop, stop/non-stop
(Stop codon still not fixed)

♂ X v♀
(Brother and sister from above)
(homozygous for stop codon)
stop/stop, stop/stop
(Stop codon fixed)

♂ X ♀
Fixed stock
(1 pair selected and bred for
further experiments)

Fig 9: Strategy for obtaining strains fixed for stop and non-stop codon. (v♀ stands for virgin female)

Genotyping

DNA extractions

Extractions were done using the Quick DNA isolation protocol (PUREGENE). The protocol was modified to one best suited for the extraction of a single fly. The sample fly was homogenized with 300 µl of cell lysis solution followed by incubation of the lysate at 65ºC for 15 minutes. The sample was then cool down to room temperature. Cell lysis was followed by RNAse treatment. 1.5 µl of RNAse was added to the lysate and mixed well. Later the mix was incubated at 37ºC for 30 minutes. 100µl of protein precipitation solution was added and vortexed vigorously for 20 seconds. Pellets were obtained by centrifugation (13,200 X g for 3 minutes). DNA was then precipitated with 300µl of isopropanol, after pipetting the supernatant. Pellets were obtained again by repeating the centrifugation step this time for 1 minute. Pellet was then washed with 200µl of 70% ethanol followed by centrifugation (13,200 X g for 1 minute). The pellet was then dried at room temperature for 15 minutes after pipetting out ethanol. The last step was to incubate the DNA at 65ºC for an hour after adding 25µl of DNA hydration solution. DNA obtained was stored at -20ºC.

Running gels and subsequent purification methods

One percent agarose gel was run to check the presence of DNA in the extracted samples and after PCR (see below) gel was made by heating 99ml 1X TAE buffer with 1 gm agarose. Care was taken that the two mixed together well. After this 2 µl of Ethidium bromide were added; the contents mixed well and allowed to set for 30 minutes before the samples could be loaded. 2-Log DNA (1/10X) from New England Biolabs was used a marker. 5 µl of the samples and ladder were loaded with1 µl of 6X loading buffer (made from 0.25% Bromophenol, 0.25% Xylene cyanol, and 30% Glycerol).

The samples with visible high molecular weight band were identified. The next step was to identify crosses with stop codon and non-stop codon, to decide from what pair to obtain the next generation of crosses.

PCR method & Restriction Digest

Polymerase chain reaction (PCR) is a common method of creating copies of specific fragments of DNA. The method rapidly amplifies a single DNA molecule into many billions of molecules. I used PCR to amplify *CG12628* alleles present in the particular individual genome. A restriction digestion of the PCR product with an enzyme that differentiates the two alleles was used for genotyping (see details below).

Typical polymerase chain reactions (PCR's) contained 5 µl 10X Thermopol AMP buffer, 4 µl 2.5mM dNTP mixtures, 2 µl each of 5'(New F 33.5') and 3'(New F 33.3') primers, (10 pmol/µl concentration of primers was used) 0.5 µl *Taq* polymerase enzyme and Diuf water to bring the final volume to 25 µl. The amount of sample DNA template added varied from 5 µl to 16 µl, depending upon the strength/brightness of the bands seen on 1% gel. They were carried out in a PTC-100 Peltier thermal cycler (MJ research). Negative controls of the PCR were always run without template DNA. Primers of the PCR are listed in Table 2.

The standard reaction profile for all amplification reactions consisted of a initial denaturation for 2 minute at 94°C followed by 30 second denaturation incubation at 94°C, followed by 29 cycles of annealing (55°C for 30 seconds), extension (72°C for 1 minute/kilobase). Reactions were terminated by a 5 minute step at 72°C (15).

The PCR products were ran on 1% Agarose gel with 2-log DNA ladder from New England Biolabs as the marker, followed by subsequent purification of the products using the QIAquick PCR purification Kit protocol.

Products were cleaned using the QIAquick PCR purification Kit protocol using a microcentrifuge. The protocol included adding 5 volumes of Buffer PB to 1 volume of the sample and centrifuging it for 1 minute at 13,200 rpm. To wash, 0.75 ml of Buffer PE was added and the sample was again centrifuged at 13,200 rpm for a minute.

Finally the DNA was eluted with 30 µl of Buffer EB. The purified DNA was stored at -20°C for future use.

Following purification, restriction digests were performed. Restriction enzymes are DNA-cutting enzymes found in bacteria and harvested from them for use. Because they cut within the molecule, they are often called restriction endonucleases. The restriction enzyme used in this study was *Hinf* I which recognizes and cuts the sequence GANTC (5' end). Standard provider of restriction enzyme was New England Biolabs.

Typical digestion reactions contained 20µl total volume, including 10µl PCR product, 2 µl buffer, 0.5 µl of enzyme and 7.5 µl of DIUF water (Double ionized water). The purpose of digestion was to obtain products of specific base pair length which would in turn denote the genotype of the sample. So a sample which is homozygous for stop codon would show 2 bands of 606 and 154 base pairs, heterozygous for stop codon/non-stop would show 3 bands, 768, 606 and 154 base pairs and homozygous for non-stop codon would show just one band of 768 base pair length. The results were checked by extracting the genomic DNA, running on 2% Agarose gel. PCR products of the individuals that will produce the lines believed to be fixed were sequence (see protocol below).

Extraction & purification of DNA from gel

The samples with smeared bands on the gel were extracted and purified using the QIAquick Gel Extraction Kit protocol using a microcentrifuge. This process involved excision of the DNA fragment from the agarose gel with a clean, sharp scalpel. The gel slice was then weighed and 3 volumes of Buffer QG was added to 1 volume of the gel. The gel was incubated at 50ºC for 10 minutes to let the gel slice dissolve completely. To facilitate dissolution, the mix was vortexed. One volume isopropanol was added to the sample and mixed well, applied to a QIAquick spin column and centrifuged for a minute. 0.75 ml Buffer PE was added to the column and centrifuged again for a minute. Finally the DNA was eluted with 30 µl Buffer EB, centrifuged for a minute and the purified product was stored at -20ºC for further use.

Sequencing

DNA sequencing was done using the following protocol. DNA was dried first using the vacuum to concentrate the product. The centrifuge was set at 45ºC. It took approximately 60 minutes for 25µl of the sample. At the end of the drying process a

clear white pellet was seen which was then resuspended in 4µl of Diuf (Deionized ultra filtered) water from Fisher Scientific. The PCR reaction master mix for sequencing consisted of 0.3 µl of pallet paint from Novagen, 0.3 µl of primer (same as used in PCR reactions, either forward or reverse; 20 pmol/µl concentration of primers used was used) and 1 µl of Big dye Terminator v3.1 (from ABI). The order of adding products to the tubes was as follows. 1.3 µl of dried DNA was added first to which 15 µl of wax was added. This mixture was allowed to solidify at 4ºC in the thermal cycler for 3 minutes. 1.6 µl of the above mentioned master mix was then added on top of the above solidified DNA and wax. The reaction program consisted of a initial denaturation at 94C for 1 minute and then at 96C for 20 seconds, followed by 24 cycles of annealing (55°C for 30 seconds), extension (60°C for 1 minute/kilobase). Reactions were terminated by a 5 minute step at 60°C (15).

The nucleotide sequences were determined using ABI Prism 377 DNA Sequencer (Applied Biosystems). The sequences obtained were analyzed using Sequencher 4.2 version and were matched with the expected sequences (45). Clustal program was used to align sequences (10). The original sequences were obtained from NCBI sequence viewer AE003569 for the parental, AY150761 for *CG12628* Prunay strain and AY150760 for *CG12628* Rinanga and Flybase (14).

RNA extractions and RT-PCR

To check for expression, RNA extractions of whole male and virgin female flies were carried out using RNeasy Mini kit QIAGEN protocol. 30 whole flies were used every time. Following extraction, the total RNA obtained was digested since practically all RNA samples have some amount of DNA contamination. The best way to get rid of the traces of genomic contamination is DNase I digestion. This is a step that is needed to study intronless gene expression because the product after RT-PCR of cDNA and genomic DNA is of the same size.

Typical DNase digests consisted of 8 µl of RNA, mixed with 1 µl of DNase enzyme and 1 µl of 10X DNase buffer. The mix was kept at room temperature for 15 minutes followed by addition of 1 µl of 25mM EDTA. Subsequent incubation at 65°C was done for 10 minutes. Following digestion, the DNase digested RNA was reverse transcribed to produce cDNA by RT-PCR (Figure 10). To check for any kind of contamination, negative controls were also kept for RT-PCR. The mix contained 5 µl of DNase digested total RNA, 6 µl DPEC water, 1 µl oligo DT primer and 1 µl 10mM dNTPs. The mix was incubated at 65°C for 5 minutes. 2 µl 0.1 DTT and 4 µl 5X first strand buffer were added to the samples. The mix was heated at 45°C for 2 minutes after which superscript enzyme; SSII RT was added to the positive samples. The standard provider of the RT-PCR system was Invitrogen. After addition the samples were kept at the same temperature for 50 minutes more and then at 70°C for 15 minutes.

The cDNA formed was amplified by subsequent PCR. Typical PCR mix consisted of 2.5 µl Buffer number 2, 1 µl 2.5mM dNTPs, 1 µl *CG12628* without stop codon primer, 1 µl GSP2 for the non-stop codon samples and *CG12628* 3' RACE 1.2 & GSP2 primers for samples with stop codon (derived) and *CG1742* 3' RACE 1 and GSP2 primers for parental controls, 4 µl cDNA, 0.2 µl Expand Hi Fidelity *Taq* (EHF *Taq*) enzyme. The standard provider of the expand high fidelity PCR system was Roche

Applied Sciences. Diuf water was added to bring the total volume of the mix to 25 µl. *Gapdh2* 5’ AND *Gapdh2* 3’ primers were used in controls. Primers of the PCR are listed in Table 2. Negative controls of the PCR were always run without any template neither cDNA or RNA treated with DNAse I.

The PCR profile for amplification reactions consisted of a initial denaturation for 2 minute at 94°C and a 30 second denaturation incubation again at 94°C, followed by 34 cycles of annealing (55°C for 30 seconds), extension (72°C for 1 minute/kilobase). Reactions were terminated by a 5 minute step at 72°C (25). 1% agarose gel was ran to check for the expression and contamination.

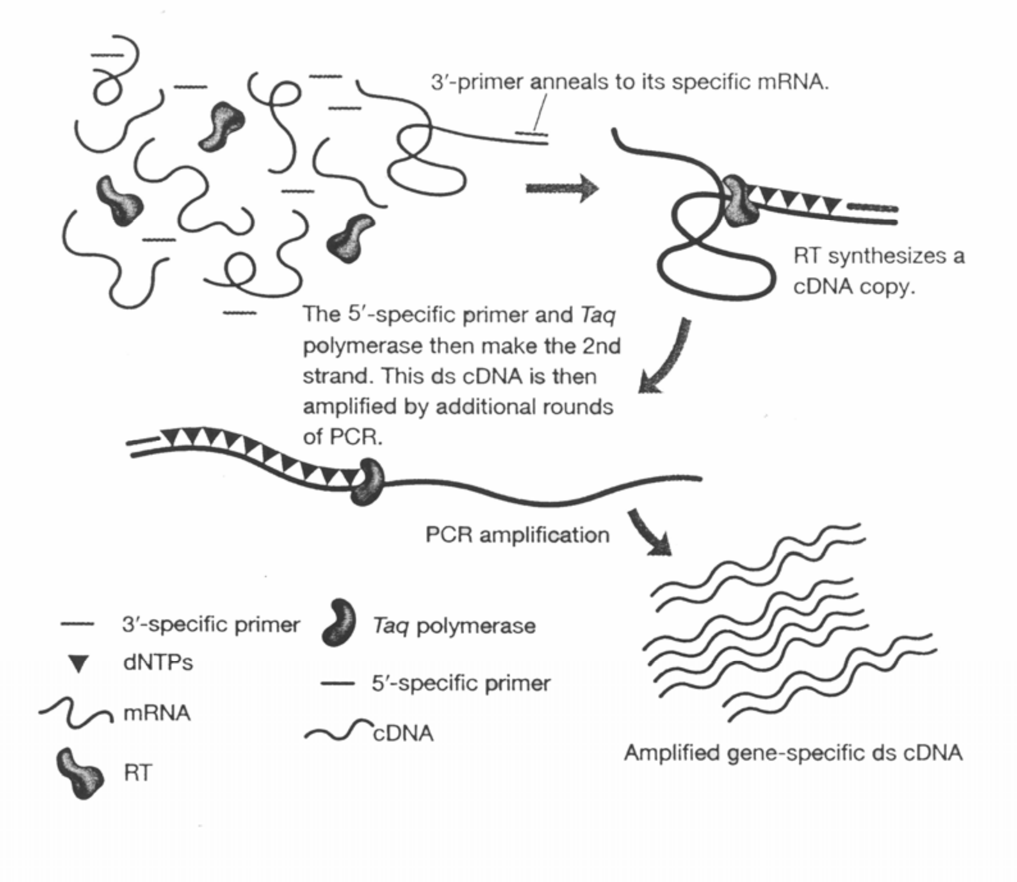

3′-primer anneals to its specific mRNA.
RT synthesizes a cDNA copy.
The 5′-specific primer and *Taq* polymerase then make the 2nd strand. This ds cDNA is then amplified by additional rounds of PCR.
PCR amplification
Amplified gene-specific ds cDNA
3′-specific primer
dNTPs
mRNA
RT
Taq polymerase
5′-specific primer
cDNA

Fig 10: Graphical illustration of the technique of RT-PCR (Figure courtesy 48)

3.2 For retrogene *CG2222-like*

Drosophila strains and growth conditions

The species studied for retrogene *CG2222-like* and its parental *CG2222* experiments were *Drosophila willistoni* and *D. melanogaster. CG2222-like* is not present in D. melanogaster but we want to compare patterns of expression to explore the fate of *CG2222-like* after duplication. The fly stock of *D. melanogaster* was the Rinanga stock described above. The *D. willistoni* was the stock 14030-0811.00 from Tucson Research Center, Arizona. The study started by growing the stock to obtain enough individuals. Corn media was routinely used for these stocks as described above. We decided to address the possible different pattern of expression between *CG2222-like* and its parental *CG2222* by looking at halves of flies. Males and female virgins were collected. As described above, the females require 7 hours to mature completely so we can separate both sexes before that. Individuals were cut in halves and RNA was inmediatly extracted as described below in both species and sexes.

RNA extractions and RT-PCR

To check for expression, RNA Extractions of male and virgin female flies (whole) was done using RNeasy Mini kit QIAGEN protocol. Flies were dissected into halves, abdomen and thorax. A minimum of 25 flies were used every time. Following extraction, the total RNA obtained was digested since practically all RNA samples have some amount of DNA contamination. The best way to get rid of the traces of genomic contamination is DNase I digestion. The reagents used and reaction profile was same as mentioned above.

The cDNA formed was amplified by subsequent PCR. Typical PCR mix consisted of 2.5 µl Buffer number 2, 1 µl 2.5mM dNTPs, 1 µl *CG2222* will exp 5A, *CG2222* will exp 3A *CG2222*mela exp 5', *CG2222*mela exp 3', *CG2222*-like 5A, *CG2222*-like 3A, 5'*ADH*willRT, 3'*ADH*willRT, 4 µl cDNA, 0.25 µl New England Biolabs *Taq* enzyme. Diuf water was added to bring the total volume of the mix to 25 µl. Specific primers were used for the particular samples, *D. melanogaster* or *D. willistoni*. *Adh* was the control for *D. willistoni* and *Gapdh2* for the *D. melanogaster* samples. Primers of the PCR are listed in Table 1. The PCR profile for amplification reactions consisted of a initial denaturation for 2 minute at 94°C and a 30 second denaturation incubation again at 94°C, followed by 34 cycles of annealing (56°C for 30 seconds), extension (72°C for 1 minute/kilobase). Reactions were terminated by a 5 minute step at 72°C (25). 1% agarose gel was run to check for expression and contamination.

Table 1. Oligonucleotide Sequences and Applications

Primer	**Sequence**	**Application**
GAPDH2 (forward)	5’-GGT GAT CAA CGA CAA CTT CGA G – 3’	PCR
GAPDH2 (reverse)	5’-GTC GTA CCA AGA GAT CAG CTT CAC -3’	PCR
New F 33- 5’	5’- ATT CCG GAT TGC AAG TAT GAG C – 3’	PCR & SEQUENCING
New F 33- 3’	5’-GAA CCC AAG ATC CGG ATT TAT TTT – 3’	PCR
3’RACE 1 *CG1742*	5’- TTC ACC TTT TGG GTC GGA G-3’	PCR & SEQUENCING
CG 1742 GSP2	5’- GAC CAG TGT GTG GAC GAT GC-3’	PCR & SEQUENCING
3’RACE 1.2 CG 12628	5’ – TTC ACC GTT TGA GTC GAA A – 3’	PCR & SEQUENCING
CG 12628 GSP2	5’- GAC CAG TGT GTG GAC GAT GT-3’	PCR & SEQUENCING
CG12628(without stop codon)	5’- TTC ACC GTT TGG GTC GAA A-3’	PCR & SEQUENCING
CG2222 will exp 5A	5’-GAG AAA GTG GAA TGC AAG GTT-3’	PCR & SEQUENCING
CG2222 will exp 3A	5’-TTC TGA TGC TCC TCC ACT AAT-3’	PCR
*CG2222*mela exp 5’	5’-GAA CTA TTA CTC CAT CGA GGA C-3’	PCR & SEQUENCING
*CG2222*mela exp 3’	5’-ATC TGC AGC CAT TCG ATG TAC T-3’	PCR
CG2222-like 5A	5’-CAA GAA AAG GTG GAG TGC AG-3’	PCR & SEQUENCING
CG2222-like 3A	5’-CAT ACC CGC TAT ATG AAG ATT G-3’	PCR
5’*ADH*willRT	5’-CTC TCA CCA ACA AGA ACA TC-3’	PCR
3’*ADH*willRT	5’-GCA ATG GAT TGA GTG AAG CT-3’	PCR

3.3 Review of pseudogenes literature

Pubmed entries were queried with the word pseudogene. The outcome was reviewed backwards down to the last 20 years. Summary tables of the numbers of pseudogenes and comments of their functions are given in section III of the results.

CHAPTER IV

RESULTS

4.1 *CG12628* results

DNA extractions

The stocks of the two strains in consideration, *D. melanogaster* Rinanga and Prunay were bred and multiple generations obtained. The project began by separating male and female flies after the first generation was obtained and DNA was extracted from them. The protocol involves obtaining genomic DNA. This step was particularly difficult and tricky, since the normal protocol followed is for 20-30 flies and single fly extraction was done for this set of experiments. Another reason for encountering difficulties in these set of experiments was that the flies after being separated were frozen at -20°C instead of -70°C, which is the ideal temperature for storing the flies for further use. Using live samples instead of frozen would have been a better option based on the brightness of the band obtained after some of the extractions were done with live samples.

One percent agarose gels demonstrated the presence of DNA. The brightness of the bands indicates the amount of DNA present in the sample (Fig. 10 and 11). Out of the 20 pairs of each strain extracted only the ones with clean, bright bands were chosen. A substantial number of generations were lost in both the strains due to the loss of either of the parents or due to the incapability to reproduce. Figure 9 shows the strategy of obtaining crosses that were carried out for Rinanga (A) and Prunay (B).

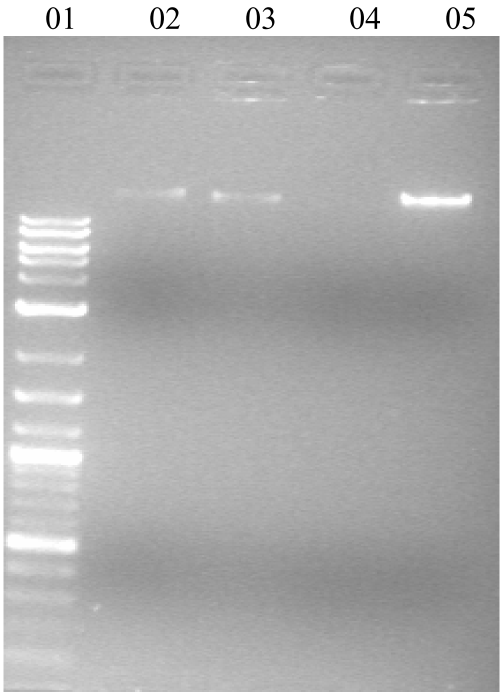


Fig 11: One percent gel picture of DNA extractions of single *D. melanogaster* Prunay and Rinanga fly. (01) shows ladder, (02) to (05) show single fly DNA samples. The various degrees of brightness indicate the amount of DNA that could be extracted from the single fly. (04) shows lack of DNA in one of the extractions. 5 µl of the sample and ladder were loaded along with 1 µl of the dye.

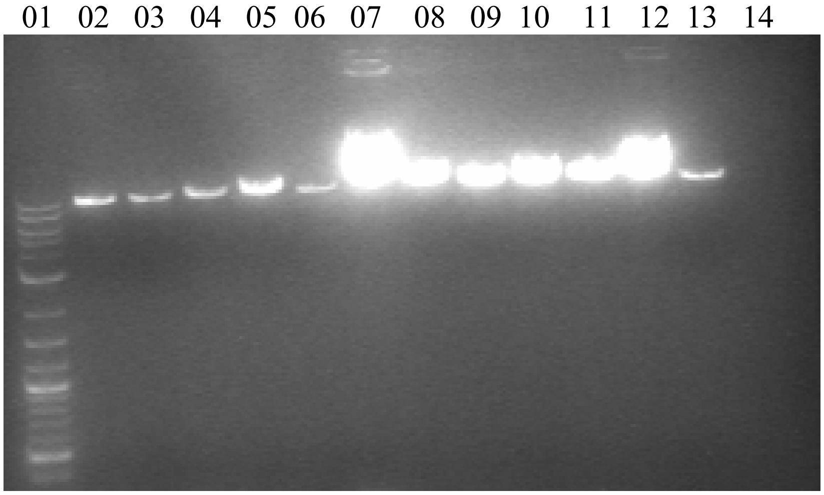


Fig 12: One percent gel picture of DNA extractions of single *D. melanogaster* Prunay and Rinanga fly. (01) indicates ladder, (02) to (14) are single fly DNA samples. The various degrees of brightness indicate the amount of DNA that could be extracted from the single fly. (14) shows the lack of DNA in one of the extractions. 5 µl of the sample and ladder were loaded along with 1 µl of the dye.

Genotyping and PCR

After DNA extractions, 6 pairs of *D. melanogaster* Prunay and 5 pairs of *D. melanogaster* which showed neat, bright bands in the 1% agarose gels were picked for further crosses. PCRs and restriction digestions allowed to differentiate parents with the desired genotypes, i.e., with homozygous or heterozugous for stop codon, heterozygous or homozygous for non-stop codon. This was done to get stocks with both the non-stop and stop codons fixed by carrying out brother sister crosses obtained from the parents. The 1% gel pictures of genotyping results are shown in Figure 13 through 16.

Amount of sample loaded in the 2% gel after digesting with *Hinf* I enzyme; which is used for recognizing and cutting the sequence at the stop codon depended on the brightness of the band obtained in 1% gel. Thus for some of the samples for which a light band was seen in 1% gel the amount loaded in 2% was increased so that even the slightest amount could show up after digestion. The samples with the desired genotype were chosen for further studies. The pairs were then allowed to mate to produce crosses between to same allele homozygous individuals and get a fix strain. At this stage I already obtained one pair of homozygous cross for *D. melanogaster* Prunay and four pairs of homozygous crosses for *D. melanogaster* Rinanga . Sequencing of the parental samples confirmed their genotypes and also that the stop and non-stop codon were indeed to be fixed in a sample of pooled flies from the strains (Fig. 17 and 18).

Overall, more than 30 extractions of samples including both Prunay and Rinanga strains were done and not all yielded positive results on 1% agarose gels. PCR of the samples with positive results was done. From these, it was inferred that four of the Prunay crosses had fix for stop codon (04th, 05th, 17th and 18th) and two crosses of Rinanga (11th, 15th) showed the non-stop codon fixed. Out of these 6 pairs, three were chosen for future experiments which included 04th cross from Prunay and 11th and 15th from Rinanga. This was due to the fact that the other pairs are not independent and we expect similar results for them. Before choosing the particular pairs for further experiments results were confirmed by repeating PCRs and running 1% gels again. Results were concurrent with the previous results and thus the crosses chosen previously were retained for further experiments.

Figure 13 and 14 show the 1% and 2% gels of genotyping results for *D. melanogaster* Rinanga strain. The 4 samples in the 1% gel were chosen for further experiments. The females show weak bands in 1% gel so the amount of sample loaded in the 2% gel after restriction digestion was increased accordingly. All four samples show a band in 700-800 bp region (~768) which indicates they are homozygous for non-stop codon. No contamination is seen since negatives are negatives in either of the gels.

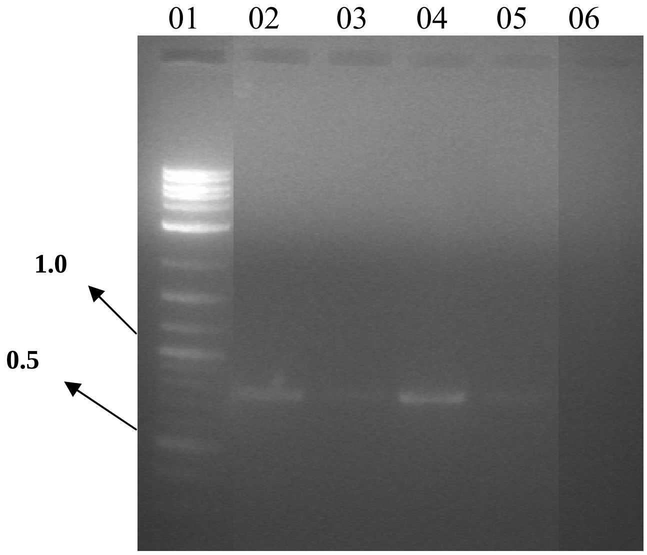


Fig 13: Genotyping results: 1% gel picture of genomic PCR products of *D. melanogaster* Rinanga pairs. (01) shows ladder, (02) and (04) show male and (03) and (05) show female samples respectively. The females show weak bands. (06) shows the negative control. There is no contamination as there is no band seen in (06). These samples were used to carry out further experiments. 5 µl of the sample and ladder were loaded along with 1 µl of the dye.

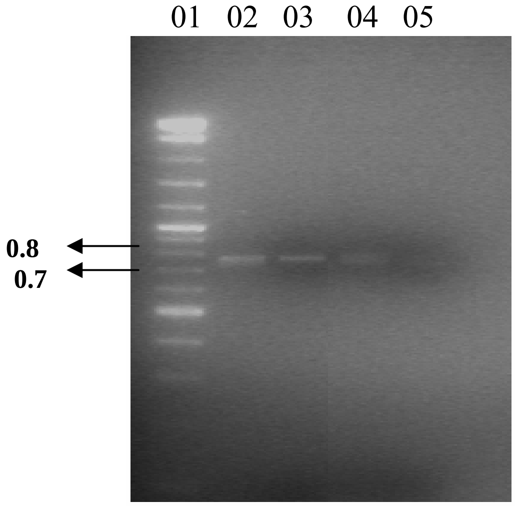


Fig 14: Genotyping results: 2% gel picture of the *D. melanogaster* Rinanga samples shown in figure 11 after digestion with *Hinf* I enzyme. The *Hinf* I enzyme is used for recognizing and cutting the sequence at the stop codon. (01) shows ladder. Sample (02), (04) show male and sample (03), (05) show female respectively. Sample (05) shows a weak band. All four samples show a band in 700-800 bp region (~768) which indicates they are homozygous for non-stop codon. The 4 samples were chosen for further experiments. Since the samples had shown light bands in 1% gel, 28 µl of the sample was loaded with 5 µl of dye. The ladder amount was kept same (5 µl+ 1 µl dye).

Figures 15 and 16 show the 1% and 2% gels of genotyping results for *D. melanogaster* Prunay strain. Very bright bands were seen for all 4 samples. After digestion the first pair turned out to be heterozygous for stop codon. The second pair however showed two bands in the 600-700 (~606) and 100-200 (~154) bp region indicating that it was homozygous for stop codon. No contamination is seen since negatives are negatives in either of the gels. Before going ahead with obtaining generations by mating the pairs the genotyping results obtained were confirmed by repeating the experiments again.

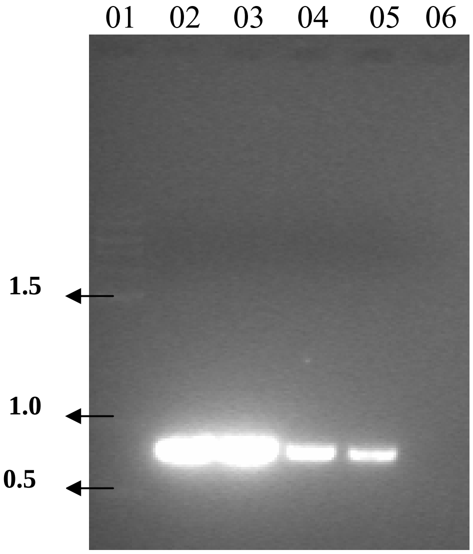


Fig 15: Genotyping results: 1% gel picture of genomic PCR products of two of the pairs of *D. melanogaster* Prunay. (01) shows ladder, (02) and (04) show male; (03) and (05) show female samples respectively. (06) shows the negative control. There is no contamination as there is no band seen in (06). 5 µl of the sample and ladder were loaded along with 1 µl of the dye. The 4 bands are quite bright indicating that a large amount of DNA was present initially.

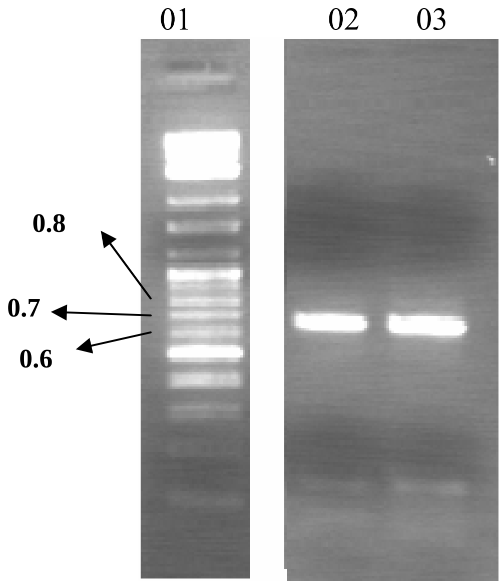


Fig 16: Genotyping results: 2% gel picture of two of the *D. melanogaster* Prunay samples shown in Figure 13 after digestion with *Hinf* I enzyme. The *Hinf* I enzyme is used for recognizing and cutting the sequence at the stop codon. (01) shows ladder. (02) shows the male sample and (03) shows female sample. Both samples are homozygous for stop codon since one band in 600-700(~606) is seen. The second band in the 100-200 (~154) bp region is light. Both the samples were chosen for further experiments. 5 µl of the sample and ladder were loaded along with 1 µl of the dye.

Sequencing

Sequencing of the selected parental pairs was done to make sure the stop and non-stop codons were indeed fixed and to carry out further experiments mostly expression analysis. The results are given below (Fig.17, 18).

```
DM_PRUNAY_C_18 ATGGCCAGCCCCGTGGAACTGCTCAGCCTCTCCAATCCCGTATTCAAGAGTTTCACCGTT 60
DM_RINANGA_C_15ATGGCCAGCCCCGTGGAACTGCTCAGCCTCTCCAATCCCGTATTCAAGAGTTTCACCGTT 60
DM_PRUNAY_C_05 ATGGCCAGCCCCGTGGAACTGCTCAGCCTCTCCAATCCCGTATTCAAGAGTTTCACCGTT 60
DM_RINANGA_C_11ATGGCCAGCCCCGTGGAACTGCTCAGCCTCTCCAATCCCGTATTCAAGAGTTTCACCGTT 60
DM_PRUNAY_C_17 ATGGCCAGCCCCGTGGAACTGCTCAGCCTCTCCAATCCCGTATTCAAGAGTTTCACCGTT 60
DM_PRUNAY_C_04 ---------------------------------------------------TTCACCGTT 9
                                                                  *********

DM_PRUNAY_C_18 TGAGTCGAAATTTTGGGGATCAATATGCTGTTGATGAGCCTTCTGACAGCCATCCAGCGT120
DM_RINANGA_C_15TGGGTCGAAATTTTGGGGATCAATATGCTGTTGATGAGCCTTCTGACAGCCATCCAGCGT120
DM_PRUNAY_C_05 TGAGTCGAAATTTTGGGGATCAATATGCTGTTGATGAGCCTTCTGACAGCCATCCAGCGT120
DM_RINANGA_C_11TGGGTCGAAATTTTGGGGATCAATATGCTGTTGATGAGCCTTCTGACAGCCATCCAGCGT120
DM_PRUNAY_C_17 TGAGTCGAAATTTTGGGGATCAATATGCTGTTGATGAGCCTTCTGACAGCCATCCAGCGT120
DM_PRUNAY_C_04 TGAGTCGAAATTTTGGGGATCAATATGCTGTTGATGAGCCTTCTGACAGCCATCCAGCGT 69
               ** *********************************************************

DM_PRUNAY_C_18 TTCAAGACGAAGACCTTCGCCAACCCCAAGGACCTACTGTCCCCCAAGCTGAAGGTCAAG180
DM_RINANGA_C_15TTCAAGACGAAGACCTTCGCCAACCCCAAGGACCTACTGTCCCCCAAGCTGAAGGTCAAG180
DM_PRUNAY_C_05 TTCAAGACGAAGACCTTCGCCAACCCCAAGGACCTACTGTCCCCCAAGCTGAAGGTCAAG180
DM_RINANGA_C_11TTCAAGACGAAGACCTTCGCCAACCCCAAGGACCTACTGTCCCCCAAGCTGAAGGTCAAG180
DM_PRUNAY_C_17 TTCAAGACGAAGACCTTCGCCAACCCCAAGGACCTACTGTCCCCCAAGCTGAAGGTCAAG180
DM_PRUNAY_C_04 TTCAAGACGAAGACCTTCGCCAACCCCAAGGACCTANTGTCCCCCAAGCTGAAGGTCAAG129
               ************************************ ***********************

DM_PRUNAY_C_18 TTCGACGATCCGAACGTGGAGCGTGTGCGCCGTGCCCACCGCAACGACCTGGAGAACATC240
DM_RINANGA_C_15TTCGACGATCCGAACGTGGAGCGTGTGCGCCGTGCCCACCGCAACGACCTGGAGAACATC240
DM_PRUNAY_C_05 TTCGACGATCCGAACGTGGAGCGTGTGCGCCGTGCCCACCGCAACGACCTGGAGAACATC240
DM_RINANGA_C_11TTCGACGATCCGAACGTGGAGCGTGTGCGCCGTGCCCACCGCAACGACCTGGAGAACATC240
DM_PRUNAY_C_17 TTCGACGATCCGAACGTGGAGCGTGTGCGCCGTGCCCACCGCAACGACCTGGAGAACATC240
DM_PRUNAY_C_04 TTCGACGATCCGAACGTGGAGCGTGTGCGCCGTGCCCACCGCAACGACCTGGAGAACATC189
               ************************************************************

DM_PRUNAY_C_18 CTGCCCTTCTTCGCCATCGGTCTGCTCTACGTCCTGACTGATCCGGCCGCCTTTCTGGCC300
DM_RINANGA_C_15CTGCCCTTCTTCGCCATCGGTCTGCTCTACGTCCTGACTGATCCGGCCGCCTTTCTGGCC300
DM_PRUNAY_C_05 CTGCCCTTCTTCGCCATCGGTCTGCTCTACGTCCTGACTGATCCGGCCGCCTTTCTGGCC300
DM_RINANGA_C_11CTGCCCTTCTTCGCCATCGGTCTGCTCTACGTCCTGACTGATCCGGCCGCCTTTCTGGCC300
DM_PRUNAY_C_17 CTGCCCTTCTTCGCCATCGGTCTGCTCTACGTCCTGACTGATCCGGCCGCCTTTCTGGCC300
DM_PRUNAY_C_04 CTGCCCTTCTTCGCCATCGGTCTGCTCTACGTCCTGACTGATCCGGCCGCCTTTCTGGCC249
               ************************************************************

DM_PRUNAY_C_18 ATCAACCTGTACCGCGCCGTGGGCATCGCCCACATCGTCCACACACTGGTCGACGCC-GT359
DM_RINANGA_C_15ATCAACCTGTACCGCGCCGTGGGCATCGCCCACATCGTCCACACACTGGTCGACGCT-GT359
DM_PRUNAY_C_05 ATCAACCTGTACCGCGCCGTGGGCATCGCCCACATCGTCCACACACTGGTCGACGCCCGT360
DM_RINANGA_C_11ATCAACCTGTACCGCGCCGTGGGCATCGCCCACATCGTCCACACACTGGTCGACGCT-GT359
DM_PRUNAY_C_17 ATCAACCTGTACCGCGCCGTGGGCATCGCCCACATCGTCCACACACTGGTCGACGCC-GT359
DM_PRUNAY_C_04 ATCAACCTGTACCGCGCCGTGGGCATCGCCCACATCGTCCACACACTGGTCGACGCC-GT308
               ********************************************************  **
```

Figure 17: Sequencing results for *D. melanogaster* Prunay and Rinanga parental pairs. The numbers are parental cross number. Out of these, *D. melanogaster* Rinanga cross 15 and *D. melanogaster* Prunay cross 04 were chosen for further experiments (expression analysis). Sequence in red indicates the stop and non-stop codon fixed.

```
DM_PRUNAY_C_18 GGTCGTGGTGCCCCAGCCTTCCCGAGCCCTCGCCTTCTTCGTGGCCCTGGGCGCCACCGT419
DM_RINANGA_C_15GGTCGTGGTGCCCCAGCCTTCCCGAGCCCTCGCCTTCTTCGTGGCCCTGGGCGCCACCGT419
DM_PRUNAY_C_05 GGTCGTGGTGCCCCAGCCTTCCCGAGCCCTCGCCTTCTTCGTGGCCCTGGGCGCCACCGT420
DM_RINANGA_C_11GGTCGTGGTGCCCCAGCCTTCCCGAGCCCTCGCCTTCTTCGTGGCCCTGGGCGCCACCGT419
DM_PRUNAY_C_17 GGTCGTGGTGCCCCAGCCTTCCCGAGCCCTCGCCTTCTTCGTGGCCCTGGGCGCCACCGT419
DM_PRUNAY_C_04 GGTCGTGGTGCCCCAGCCTTCCCGAGCCCTCGCCTTCTTCGTGGCCCTGGGCGCCACCGT368
               ************************************************************

DM_PRUNAY_C_18 CTACATGGCCCTGCA-GGTCATCGCCTCGGCCGCCTTCTGAAGCACATAGGTCTA-GTCC477
DM_RINANGA_C_15CTACATGGCCCTGCATGGTCATCGCCTCGGCCGCCTTCTGAAGCAAC-------------466
DM_PRUNAY_C_05 CTACATGGCCCTGCA-GGTCATCGCCTCGGCCGCCTTCTNAA-CAN--------------464
DM_RINANGA_C_11CTACATGGCCCTGCA-GGTCATCGCCTCGGCCGCCTTCTGAG-CACATAGGTCTA-GTCC476
DM_PRUNAY_C_17 CTACATGGCCCTGCA-GGTCATCGCCTCGGCCGCCTTCTGAAGCACATAGGTCTA-GTCC477
DM_PRUNAY_C_04 CTACATGGCCCTGCA-NGTCATCGCCTCGGCCGCCTTCTGAGCAACATAGGTCTAAGTCC427
               ***************  ********************** *   *

DM_PRUNAY_C_18 TTCTTGTTTTTTTT-AAAGCATTTTGAAATAATTTCTTTAATATA-GAAGTTACGCCTAC535
DM_RINANGA_C_15------------------------------------------------------------
DM_PRUNAY_C_05 ------------------------------------------------------------
DM_RINANGA_C_11TTCTTGTTTTTTTT-AAAGCATTTTGAAATAATTTCTTTAATATA-GAAGTTACGCCTAC534
DM_PRUNAY_C_17 TTCTTGTTTTTTTT-AAAGCATTTTGAANTAATTTCTTTAATATAAGAAGTTACGCCTAC536
DM_PRUNAY_C_04 TTCTTGTTTTTTTTAAAGCATTTTGAAAATAATTTCTTTTAATATAGAAGTTACGCCTAC487


DM_PRUNAY_C_18 CTCGGCTTTG--TTGCTGTTGGATCACAAAAAAAAAA-------TATCATCGGGCAATTT586
DM_RINANGA_C_15------------------------------------------------------------
DM_PRUNAY_C_05 ------------------------------------------------------------
DM_RINANGA_C_11CTCGGCTTTG--TTGCTGTTGGATCACAAAAAAAAAAA------CATCATGGGGCCAATT586
DM_PRUNAY_C_17 CTCGGCTTTG--TTGCTGTTGGATCACAAAAAAAAAAA------CATCTCGGGGCAATTT588
DM_PRUNAY_C_04 CTCGGCTTTTGTTTGCTGTTGGATCACAAAAAAAAAAAAAACNTCCNCCGGGGGCCAANTT547


DM_PRUNAY_C_18 TCCTTAA-TGGCTTAAAATCTCGGAA--CAAANNTTTTG--TTGTTGGNCCCGCCGGCAT641
DM_RINANGA_C_15------------------------------------------------------------
DM_PRUNAY_C_05 ------------------------------------------------------------
DM_RINANGA_C_11TTCNTTT-TNGCTTTAAAATCNGGAA--NCAAACTTTTT--TTGTTGGCCACGGGCCGCA641
DM_PRUNAY_C_17 TCCTTTT-TGGCTTAAAATTNCNGAA--CAAAATTTTTTGTTTGTTGGCACCGCCCGGCC645
DM_PRUNAY_C_04 TCCTTNAGNGGNTTAANATTTCNGGAACCAAAAANTTTNTTTTNTTGGNTNCCTGCCCGG607
```

Figure 18: Sequencing results for *D. melanogaster* Prunay and Rinanga parental pairs continued. The numbers are parental cross number. Out of these, *D. melanogaster* Rinanga cross 15 and *D. melanogaster* Prunay cross 04 were chosen for further experiments (expression analysis). Sequence in red indicates the stop and non-stop codon fixed are shown in Figure 17.

Expression Analysis

After checking for the stop and non-stop codon being fixed in the parents, expression analysis was done. Before using the primers designed for expression analysis, they were checked for specificity, i.e., if they amplify the desired gene instead of the parental gene. This was done by extracting DNA of each of the two strains (Prunay 4 and Rinanga 15) and performing PCR with the primers that will be used for expression analysis. The primers used are shown in Table 1. They were designed in regions of distinct sequence between parental and derived genes. The bands obtained on the 1% gel corresponded to the 200-300 base pair region in the 2-log ladder which validated the specificity of the primers (Fig. 19).

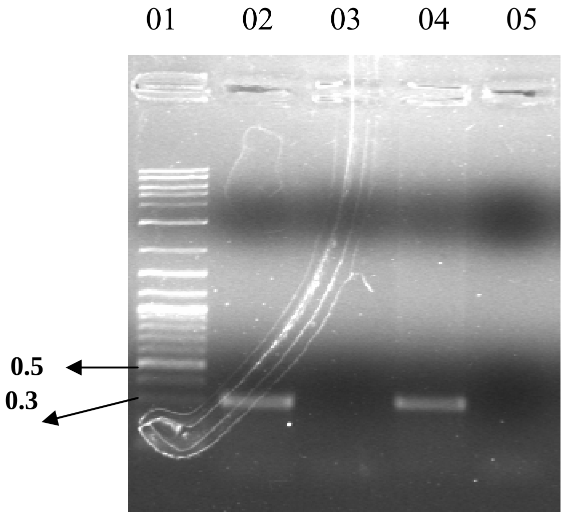


Fig 19: One percent gel picture to check primers designed for expression analysis. (01) shows ladder. (02) shows *D. melanogaster* Rinanga genomic DNA, (03) shows PCR negative of (02), (04) *D. melanogaster* Prunay genomic DNA, (05) shows PCR negative of (04). The bands were seen in desired regions (~300) showing that the primers were working.

After checking for the primers, RT and subsequent PCR of the selected Prunay and Rinanga male and female RNA for the genes *CG12628*, *Mgstl* and *Gapdh2* was performed. Each RT-PCR had a negative control for the RT and a negative control of the PCR to check for contaminations. *Gapdh2* is considered the positive gene control of the RT because is a gene that is expressed in every tissue.

The expression analysis was tricky since RNA is difficult to handle due to its susceptibility to contamination. Initially the RT-PCR yielded negative results. Troubleshooting was done, which included changes in the amount of reagents along with the amount of cDNA added, order of adding them and also the reaction profile of the PCR. The results of expression analysis are shown in Figures 20 through 23. The expression analysis results were counterchecked by performing RT-PCR again (Fig 24 and 25). Sequencing results confirmed the expression analysis results.

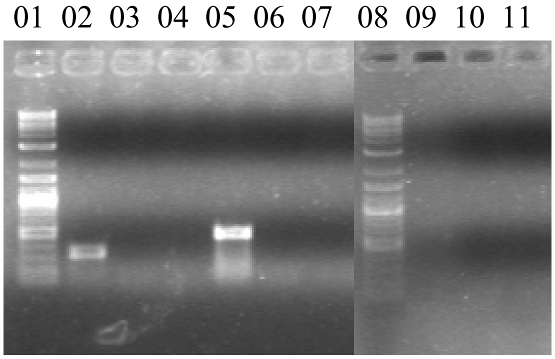


Fig 20: One percent gel picture of expression analysis of *CG12628* gene in *D. melanogaster* Rinanga male. (01) and (08) show ladder.

(05), (06), (07) show *Gapdh2* RT-PCR positive, negative and PCR negative respectively in D.melanogaster Rinanga male. The bright band indicates the high expression of the gene. *Gapdh2* was used as the positive control. No contamination is seen as negatives are negatives.

(02), (03), (04) show the parental gene *CG1742* RT-PCR positive, negative and PCR negative respectively in D.melanogaster Rinanga male. The bright band indicates the high expression of the gene. No contamination is seen as negatives are negatives.

(09), (10), (11) show the derived gene *CG12628* (without stop codon) RT-PCR positive, negative and PCR negative respectively in D.melanogaster Rinanga male. No band is seen indicating non-expression of the gene. No contamination is seen as negatives are negatives.

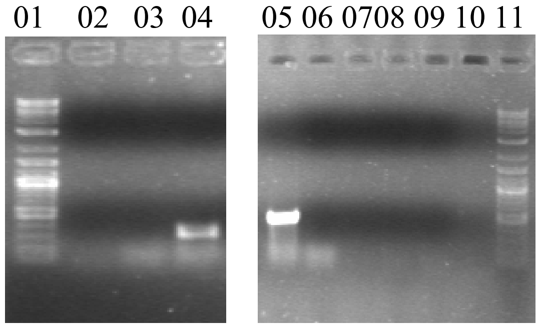


Fig 21: One percent gel picture of expression analysis of *CG12628* gene in *D. melanogaster* Rinanga female. (01) and (11) show ladder.

(05), (06), (07) show *Gapdh2* RT-PCR positive, negative and PCR negative respectively in D.melanogaster Rinanga female. The bright band indicates the high expression of the gene. *Gapdh2* was used as the positive control. No contamination is seen as negatives are negatives.

(04), (03), (02) show the parental gene *CG1742* RT-PCR positive, negative and PCR negative respectively in D.melanogaster Rinanga female. Band indicates expression of the gene. No contamination is seen as negatives are negatives.

(08), (09), (10) show the derived gene *CG12628* (without stop codon) RT-PCR positive, negative and PCR negative respectively in D.melanogaster Rinanga female. No band is seen indicating non-expression of the gene. No contamination is seen as negatives are negatives.

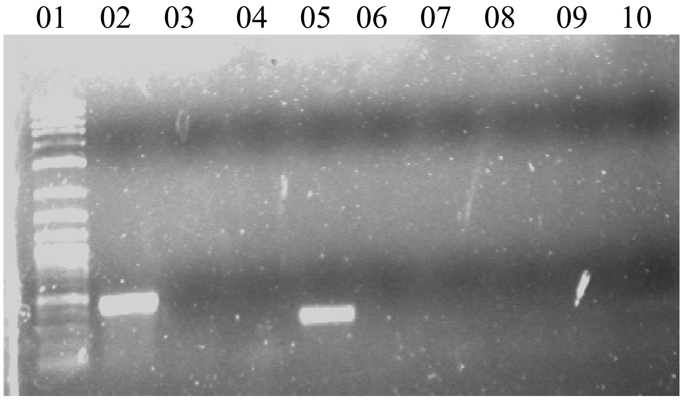


Fig 22: One percent gel picture of expression analysis of *CG12628* gene in *D. melanogaster* Prunay male. (01) shows ladder.

(02), (03), (04) show *GAPDH2* RT-PCR positive, negative and PCR negative.

(05), (06), (07) show *Mgstl* RT-PCR positive, negative and PCR negative.

(08), (09), (10) show *CG12628* RT-PCR positive, negative and PCR negative.

No contamination is seen because negatives are negatives.

The results confirmed the non-expression of the gene *CG12628*.

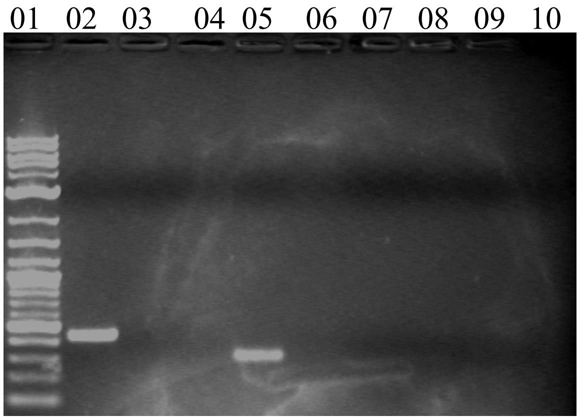


Fig 23: One percent gel picture of expression analysis of *CG12628* gene in *D. melanogaster* Prunay female. (01) shows ladder.

(02), (03), (04) show *Gapdh* RT-PCR positive, negative and PCR negative.

(05), (06), (07) show *Mgstl* RT-PCR positive, negative and PCR negative.

(08), (09), (10) show *CG12628* RT-PCR positive, negative and PCR negative.

No contamination is seen because negatives are negatives.

The results confirmed the non-expression of the gene *CG12628*.

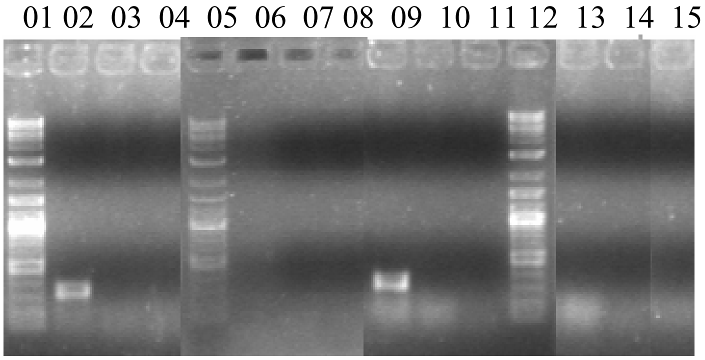


Fig 24: One percent gel picture of confirmation of expression analysis results of *D. melanogaster* Rinanga male and female. (01), (05), (12) show ladder.

(02), (03), (04) show *Mgstl* RT-PCR positive, negative and PCR negative in male. Band indicates expression of the gene.

(06), (07), (08) show *CG12628* RT-PCR positive, negative and PCR negative. Band indicates expression of the gene.

(09), (10), (11) show *Mgstl* RT-PCR positive, negative and PCR negative in female. No band is seen indicating non-expression of the gene.

(13), (14), (15) show *CG12628* RT-PCR positive, negative and PCR negative in female. No band is seen indicating non-expression of the gene.

No contamination is seen because negatives are negatives.

The results confirmed the non-expression of the gene *CG12628* in *D. melanogaster* Rinanga male and female.

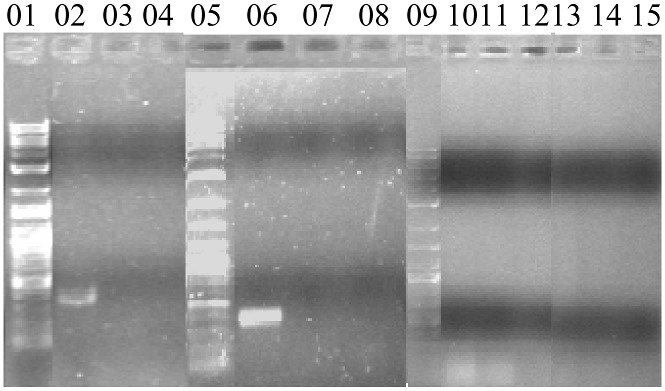


Fig 25: One percent gel picture of confirmation of expression analysis results of *D. melanogaster* Prunay male and female. (01), (05) and (09) show ladder.

(02), (03), (04) show *Mgstl* RT-PCR positive, negative and PCR negative in male. Band indicates expression of the gene. No contamination is seen as negatives are negatives.

(06), (07), (08) show *Mgstl* RT-PCR positive, negative and PCR negative in female. No contamination is seen as negatives are negatives.

(10), (11), (12) show *CG12628* RT-PCR positive, negative and PCR negative in male. No contamination is seen as negatives are negatives. No band is seen indicating non-expression of the gene.

(13), (14), (15) show *CG12628* RT-PCR positive, negative and PCR negative in female. No contamination is seen as negatives are negatives. No band is seen indicating non-expression of the gene.

The results reconfirmed the non-expression of the gene *CG12628* in *D. melanogaster* Prunay male and female.

4.2 *CG2222* & *CG2222-like* results

DNA extraction and PCR of the species in question, *D. melanogaster* (Rinanga) and *D. willistoni* was done. This project mostly involved the expression analysis for *CG2222* and *CG2222-like* in *D. melanogaster* and *D. willistoni* species. RNA extraction of fly halves followed by RT-PCR and subsequent PCR was done. The gene *Adh* was used as positive control for *D. willistoni* samples and *Gapdh2* for *D. melanogaster* samples. Negatives were used to check for contamination. The results are shown in figures 26 through 32. Based on the results expression of *CG2222*-like is seen in the thorax and abdomen of *D. willistoni* both male and female.

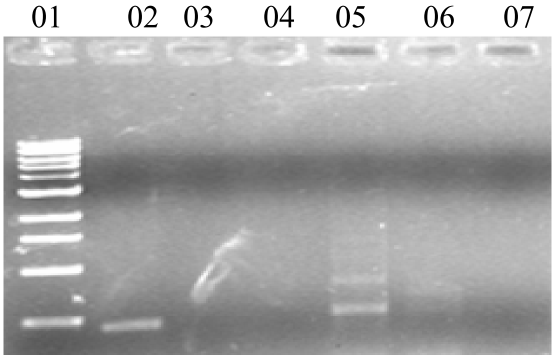


Fig 26: One percent gel picture of expression analysis of *CG2222* gene in *D. melanogaster* male abdomen. (01) shows ladder.

(02), (03), (04) show *Gapdh2* RT-PCR positive, negative and PCR negative of *D. melanogaster* male abdomen respectively. No contamination is reported as negatives are negatives.

(05), (06), (07) show *CG2222* RT-PCR positive, negative and PCR negative of *D. melanogaster* male abdomen respectively.

No contamination is reported as negatives are negatives.

Expression is manifested by the presence of bands.

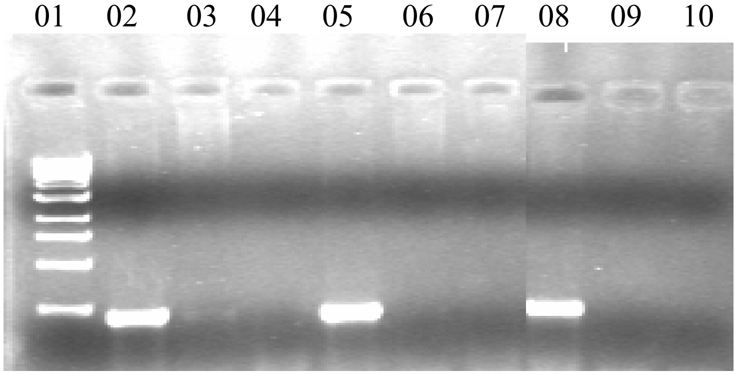


Fig 27: One percent gel picture of expression analysis of *CG2222* gene in *D. melanogaster* male thorax. (01) shows ladder.

(02), (03), (04) *Gapdh2* RT-PCR positive, negative and PCR negative of *D. melanogaster* male abdomen respectively.

(05), (06), (07) *Gapdh2* RT-PCR positive, negative and PCR negative of *D. melanogaster* male thorax respectively.

(08), (09), (10) *CG2222* RT-PCR positive, negative and PCR negative of *D. melanogaster* male thorax respectively.

No contamination is seen since negatives are negatives.

Expression is manifested in the presence of bands.

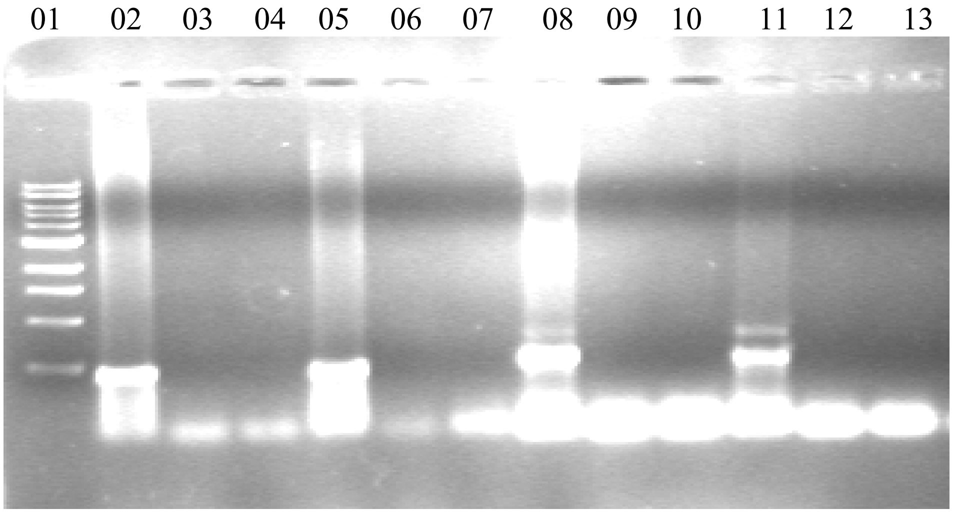


Fig 28: One percent gel picture of expression analysis of *CG2222* gene in *D. melanogaster* female.

(01) shows ladder. (02), (03), (04) show *Gapdh2* RT-PCR positive, negative and PCR negative samples from *D. melanogaster* female abdomen respectively. No contamination is reported as negatives are negatives.

(05), (06), (07) show *Gapdh2* RT-PCR positive, negative and PCR negative samples from *D. melanogaster* female thorax respectively. No contamination is reported as negatives are negatives.

(08), (09), (10) show *CG2222* RT-PCR positive, negative and PCR negative samples from *D. melanogaster* female abdomen respectively. No contamination is reported as negatives are negatives. High expression of the gene is seen as indicated by the bright band.

(11), (12), (13) show *CG2222* RT-PCR positive, negative and PCR negative samples from *D. melanogaster* female thorax respectively. No contamination is reported as negatives are negatives. High expression of the gene is seen as indicated by the bright band.

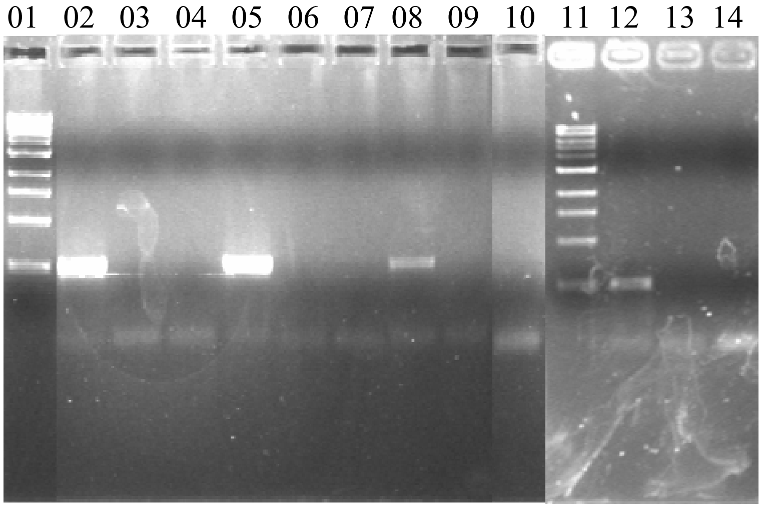


Fig 29: One percent gel picture of expression analysis of *CG2222* gene in *D. willistoni* male. (01) shows ladder.

(02), (03), (04) show *Adh* RT-PCR positive, negative and PCR negative of *D. willistoni* male abdomen respectively. No contamination is reported as negatives are negatives.

(05), (06), (07) show *Adh* RT-PCR positive, negative and PCR negative of *D. willistoni* male thorax respectively. No contamination is reported as negatives are negatives.

(08), (09), (10) show *CG2222* RT-PCR positive, negative and PCR negative of *D. willistoni* male abdomen respectively. No contamination is reported as negatives are negatives. Expression is manifested in the presence of band.

(11) shows ladder. (12), (13), (14) show *CG2222* RT-PCR positive, negative and PCR negative in *D.willistoni* male thorax respectively. No contamination is reported as negatives are negatives. Expression is manifested in the presence of band.

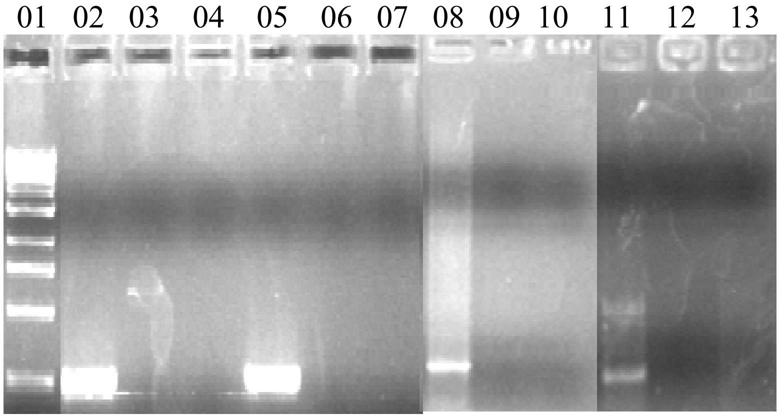


Fig 30: One percent Gel picture of expression analysis of *CG2222-like* gene in *D. willistoni* male. (01) shows ladder.

(02), (03), (04) show *Adh* RT-PCR positive, negative and PCR negative of *D. willistoni* male abdomen respectively. No contamination is reported as negatives are negatives.

(05), (06), (07) show *Adh* RT-PCR positive, negative and PCR negative of *D. willistoni* male thorax respectively. No contamination is reported as negatives are negatives.

(08), (09), (10) show *CG2222-like* RT-PCR positive, negative and PCR negative of *D.willistoni* male abdomen respectively. No contamination is reported as negatives are negatives. Expression is manifested in the presence of a strong bright band.

(11), (12), (13) show *CG2222-like* RT-PCR positive, negative and PCR negative of *D.willistoni* male thorax respectively. No contamination is reported as negatives are negatives. Expression is manifested in the presence of band.

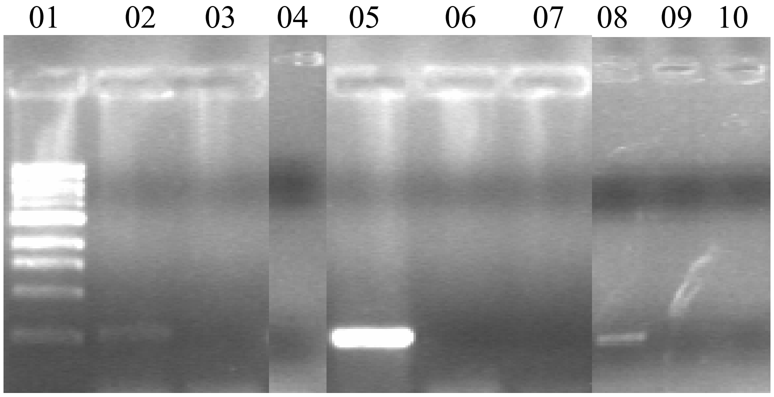


Fig 31: One percent gel picture of expression analysis of *CG2222* gene in *D. willistoni* female. (01) shows ladder.

(02), (03), (04) show *Adh* RT-PCR positive, negative and PCR negative samples from *D. willistoni* female abdomen respectively. No contamination is reported as negatives are negatives.

(05), (06), (07) *Adh* RT-PCR positive, negative and PCR negative samples from *D. willistoni* female thorax respectively. No contamination is reported as negatives are negatives.

(08), (09), (10) show *CG2222* RT-PCR positive, negative and PCR negative samples from *D. willistoni* female abdomen respectively. No contamination is reported as negatives are negatives.

Expression is manifested in the presence of bands.

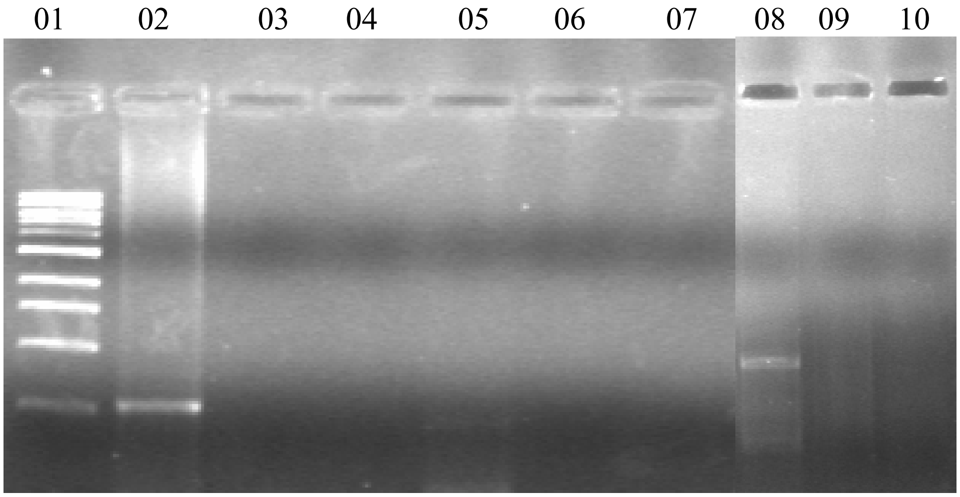


Fig 32: One percent gel picture of expression analysis of *CG2222-like* gene in *D. willistoni* female.

(01) shows ladder. (02), (03), (04) show *CG2222* RT-PCR positive, negative and PCR negative from *D. willistoni* female thorax. No contamination is reported as negatives are negatives.

(05), (06), (07) show *CG2222-like* RT-PCR positive, negative and PCR negative from *D. willistoni* female abdomen. A very light band is seen in (05) which indicates low expression of the gene *CG2222-like* in female abdomen. No contamination is reported as negatives are negatives.

(08), (09), (10) show *CG2222-like* RT-PCR positive, negative and PCR negative from *D. willistoni* female thorax.

No contamination is reported as negatives are negatives.

Expression is manifested in the presence of a band.

4.3 Review of pseudogenes literature

In the last couple of years efforts have been made to identify and characterize pseudogenes in sequenced organisms especially after the human genome was sequenced. The table below shows the number of pseudogenes present in some of the major organisms whose genome has been completely sequenced.

Table 2: Number of genes and pseudogenes in completely sequenced genomes

Organism	Genome size	No. of genes	No. of pseudogenes
R. prowazekii	1.1	834	241
M. leprae	3.3	1604	1116
Y. pestis	4.6	4061	160
E. coli K-12 strain	4.6	1100	95
S. cerevisiae	12.1	6340	241
C. elegans	102.9	20,009	2168
D. melanogaster	128.3	14,332	110
A. thaliana	115.4	25,464	>700
H. sapiens	3040	22,000-39,000	13,398 (19,929)[a]
M. musculus	2493	22,011	14,000 (∼10,000)[b]

a Number in the parenthesis are pseudogenic fragments

b Unpublished results by the authors

(Table data from 18)

Pseudogenes are common but in general it is not possible to tell the exact numbers until the organism genome has been completely sequenced. After the sequencing of genomes of living organisms (yeast, worm, human and so on) efforts have been made in the last several years to identify and characterize their pseudogene populations. According to Harrison et al about one pseudogene for every eight functional genes is present in *C. elegans* (24). Large number of pseudogenes also exist in prokaryotes like bacteria (11). Thus looking at the large number of pseudogenes present in sequenced genome their importance cannot be ruled out. Because of the large size of human genome, new pseudogenes are being discovered everyday. The first chromosomes to be looked at in humans were chromosomes 21 and 22 due to their small size which revealed more than 400 pseudogenes (21). Other genes that were looked at were cytoplasmic and mitochindria ribosomal genes (58), nuclear mitochondrial pseudogenes (54) and olfactory receptors (19).

The importance of identifying and characterizing pseudogenes is huge especially in humans because of their high sequence similarity with the corresponding functional genes they often interfere with RT-PCR or in-situ hybridization experiments and may be mistaken for the parental gene itself. (see CK19 below) if they are transcribed further complicating the interpretation of experimental results. Pseudogenes are also compared to “fossils” which provide a molecular record useful in studying evolution especially since the rate of nucleotide substitution and DNA loss can be inferred from their study. Also since pseudogenes often induce errors, their systematic and exact identification and characterization would in turn improve the information about the already predicted gene sequences, as seen in *C .elegans* genome (34).

Table 3: Transcribed pseudogenes in completely sequenced genomes

Organism	Transcribed pseudogene	Parental gene	Function	Reference
Human	Human Prohibitin gene	A Zn-finger containing protein	Inhibition of DNA synthesis	22
E. coli	Mitochondrial 2-amino 3-ketobutyrate coenzyme A	-	Precursor to a Mitochondrial enzyme	22
Human	L-Threonine 3-Dehydrogenase	-	Indispensable amino acid	15
Human	TOP1 pseudogene	TOP1	DNA unwinding enzyme	60
Human	5HT7 receptor psudogene	5-HT7	Adenylate cyclase activity	36
Lymnaea stagnalis	NOS pseudogene	NOS(nitric oxide synthase)	Intercellular signaling molecule in NS(NO) is generated by this enzyme(NOS)	28
Mice	Makorin1-p1	Makorin 1	Kidneys and bones	26
Mice	ΨFgfr-3	Fgfr-3	Fetal tissues	55
Human	hHaA	hHa3	Hair	42
Human	TPT 1	-	Tumor protein	51
Human	MY015BP	-	Myosin	9
Human	psGBA	GBA	Fibroblasts	46
Cow	Aromatose pseudogene, ΨCyp19	Cyp19	Placenta	17
Sheep	Ribonuclease pseudogene BSRNAase	RNAase	Digestive enzyme secreed by pancreas	12
Arabidopsis thaliana	TGG pseudogene	TGG	Stamen and petal	57
Arabidopsis thaliana	*rps14* pseudogene	Ribosomal protein gene	Mitochondria	3

Only pseudogenes with disablements (for example deletions, insertions and stop codon) and are transcribed were taken into consideration. A comprehensive review of transcribed pseudogenes is shown in Table 3. Most of them are present in humans since there are approximately 20,000 pseudogenes present. Harrison et al found about 166-233 transcribed pseudogenes which comprise about 4-6% of peudogenes in humans (22). Only the ones which are well identified and characterized are mentioned in the table.

Harrison et al were the first people who developed an initial strategy for annotating pseudogenes in human genome. Chromosomes 21 and 22 were the first ones to be considered because of their small size and still being discovered. (32). The review presents some interesting examples of transcribed pseudogenes. Human topoisomerase 1 (*TOP1*) was the first example of a naturally occurring antisense RNA transcript produced from a pseudogene (60). Its murine homolog ΨFgfr-3 is also transcribed in an antisense direction (55). Sometimes the transcribed pseudogene is detected in tissues it is not normally found to express for example the 5-HT7 receptor pseudogene which is seen in liver and kidney whereas the parental gene is expressed only in brain (36). In other case, the pseudogene is transcribed more than the parental gene itself as seen in tumor repressor pseudogene *ΨPTEN.* One other example is the human glucocerbrosidase pseudogene psGBA (4).

Some pseudogenes that are transcribed but not translated may have medical implications both are tumor related genes and may interfere with the tumor diagnosis assays since they may get mutated and interfere with the interpretation of the RT-PCR results.. Examples are PTEN and CK19 pseudogenes (16). Sometimes a transcribed pseudogene may regulate the parental gene as seen the mouse Makorin1-p1 pseudogene (26) and the pNOS pseudogene in snail *Lymnaea stagnalis* (28). Other undiscovered or discovered pseudogenes may have important pregulatory functions like them (26). Since pseudogenes assemble more mutations than translated genes, this discovery of gene regulation by pseudogenes would thus allow more rapid diversification of functions than of protein coding genes. Some other transcribed pseudogenes have been found in plants. These include the rps19 pseudogene in Oenothera berteriana (44) and N*ADH* pseudogene in liverwort *Machantia polymorpha* (47).

There is also a set of pseudogenes which are presently considered non-functional but can be investigated in the future for being expressed and having a function. These include LAMRL5 belonging to the human laminin receptor (LAMR) gene family (43) and human ΨCK2ά (41).

No data currently suggests that they are transcribed but looking at the sequence arrangement the possibility of them being transcribed cannot be ruled out (32). One of the recently discovered doubtful cases of transcribed pseudogenes is seen in cat. The sweet receptor gene in mammals is formed by the dimerization of 2 proteins produced by Tas1r2 and Tas1r3 recpetor gene. According to Li et al the Tas1r3 gene is functional and expressed receptor while Tas1r2 is an unexpressed pseudogene. Due to this a functional receptor heteromer cannot form since Tas1r2 is not transcribed or if it does, it rapidly degrades perhaps through non-sense mediated mRNA decay and thus the cats either do not prefer or are not able to detect sweetness because the receptor is never expressed. This molecular change might be an important turnaround in the evolution of cat's carnivorous behavior (29).

As different genomes are sequenced and more pseudogenes are identified and characterized, their impact in molecular genetics and evolutionary biology will increase.

CHAPTER V

DISCUSSION

In this thesis the expression of two newly originated genes has been studied. One of them *CG12628* is very young and is present only in one species and the other one, *CG2222-like*, is older. A comprehensive review of literature with respect to the number of pseudogenes in the genomes and their possible functions has been done because it is relevant for the questions posed by the features observed in *CG12628*.

***CG12628*: possibly a non-transcribed retropseudogene**

CG12628 is the youngest and one of the most intriguing of the 24 retrogenes described by Betrán et al. (8). Its age is estimated to be approximately 2 million years based on sequence divergence and in the fact that the pseudogene it is present only in *D. melanogaster* species in which divergence between species is 2-3 million years old. To study its expression we set the goal of doing it in two different alleles of this gene to control for nonsense mediated decay a phenomenon that leads to the degradation of transcripts with premature stop codons and that could interfere with our analyses.

After brother-sister consanguineous crosses fixed lines for stop and non-stop codons were obtained. Three generations were needed to fix the non-stop codon from the Rinanga and about four generations to obtain a line fixed for the stop codon from the Prunay strain.

RT-PCR was conducted from total RNA from females and males in both lines. While other genes (*Mgstl* and *Gapdh2*) were shown to be expressed in both sexes and at high levels, no RT-PCR product was observed in any sex despite repeating the RNA extractions and using more cDNA for the PCR. This reveals that our primers are being very specific, i.e. they do not amplify the parental gene (*Mgstl*), and points to the fact that *CG12628* is not expressed in *D. melanogaster* adults at least not at a detectable level in either allele. I focus on the adult stage because many retrogenes have been shown to specifically express in male and/or female germline but other stages of the life cycle (embryos, larvae and pupae) remain to be studied. If expression is observed at all in these stages it will be exciting to keep the stop and non-stop lines to study possible phenotypic effects of this gene. If expression is detected, it will also have to be determined if it is sense or antisense. The interpretation of the role of the gene would be different if a sense or an antisense is shown to express as introduced in section III of the results.

As shown when reviewing pseudogene features and presenting the questions, most of the pseudogenes are just dying non-transcribed copies of genes. Those are very abundant in vertebrate mammals but not in *Drosophila*. At this point the negative result for the expression of *CG12628* supports that it is a non-transcribed retropseudogene.

CG2222-like: a new gene with overlapping expression with its parental gene

CG2222-like is a gene that originated through retroposition less than 40 million years ago given its phylogenetic distribution. K_S between parental and derived gene, i.e. between *CG2222* and *CG2222-like* using Nei-Gojobori method with Jukes and Cantor correction for multiple hits was estimated to be 0.8815. If we assume 1% divergence per million years per lineage as has been proposed for *Drosophila* (38), the time of duplication would be ~ 44 million years ago consistent with the phylogenetic distribution and a quite old age of *CG2222-like*.

Many features support the functionality of *CG2222* as presented in the introduction: its old age, as shown above, and intact open reading frame, protein constrain and expression in males and females. In this thesis we address the possible role of this duplicate in the genome.

The comparison of the pattern of expression between parental gene in *D. melanogaster* and *D. willistoni* and the derived gene in *D. willistoni* could reveal if the gene is kept in the genome because it shows a different pattern of expression (neofunctionalization), a complementary pattern of expression with respect to the ancestral (subfunctionalization) or the same pattern of expression (need of more of the same).

My data reveals that *CG2222*, the parental gene, is expressed in *D. melanogaster* and *D. willistoni* in both thorax and abdomen of males and females and that *CG2222-like* in *D. willistoni* is also expressed in both thorax and abdomen of males and females. While this is not a detail tissue analysis of expression it reveals that *CG2222-like*, unlike other known retrogenes (6), (7) does not express specifically in male and/or female germline. Given the complete overlap of expression between parental and derived gene, one interpretation, given the experimental results till date would be that *CG2222-like* was kept in the genome to produce a higher amount of this type of protein. The level of identity between the two proteins (85%) even after 40 million years further supports this view. However this 15% difference between the two proteins is substantial given the fact that sometimes even a single amino acid change can impart a new function to the gene. This possibility (neofunctionalization) cannot be ruled out in this particular case.

REFERENCES

1. Adams et al. The genome sequence of Drosophila melanogaster. Science, March 2000. 287(5461):2185-95

2. Armstrong, R. N. (1993). Glutathione s-transferases: structure and mechanism of an archetypical detoxification enzyme. Advanced enzymology. 69: 1-44

3. Aubert et al. Mitochondrial *rps14* is a transcribed gene and edited pseudogene in *Arabidopsis thaliana*. Plant Molecular Biology, 1992. 20:1169-1174

4. Balakirev, Evgeniv S., Ayala, Francisco J. Pseudogenes: Are They "Junk" or Functional DNA? Annual Review of Genetics, December 2003, Vol. 37, Pages 123-151

5. Bergman et al. Assessing the impact of comparative genomic sequences data on the functional annotation of the *Drosophila* genome.Genome Biology 2002, Vol. 3, Issue 12

6. Betrán, Esther, Long, Manyuan**.** Dntf-2r, a Young Drosophila Retroposed Gene With Specific Male Expression Under Positive Darwinian Selection. Genetics, Vol. 164, 977-988, July 2003

7. Betran, Esther, Long, Manyuan. Expansion of genomic coding regions by acquisition of new genes. Genetica, 2002. 115:65-80

8. Betran, Esther, Thornton, Kevin, Long, Manyuan. Retroposed New Genes out of the X in Drosophila. Genome research, 2002, Vol. 12: 1854-1859

9. Boger et al. Human myosin XVBP is a transcribed pseudogene. J. Muscle Res. Cell Motil, 2001. 22:477-483

10. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., et al. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res, 31(13), 3497-3500

11. Cole et al. Massive gene decay in the leprosy bacillus. Nature, 2001. 409:1007-1011

12. Confalone et al. Molecular evolution of genes encoding ribonucleases in ruminant species. J. of Molecular evolution, 1995. 41:850-858

13. Drosophila @ NCBI domain page.
http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/7227.html
Accessed August 15, 2005

14. Drysdale, R. A., & Crosby, M. A. (2005). FlyBase: genes and gene models. Nucleic Acids Res, 33(Database issue), D390-395

15. Edgar, Alasdair J. The human L-threonine 3-dehydrogenase gene is an expressed pseudogene. BMC Genetics, 2002. Vol 3:18

16. Fuji, G.H., Morimoto A.M., Berson A.E., Bolen J.B. Transcriptional analysis of the PTEN/MMAC1 pseudogene, ΨPTEN. Nature, 1998. Vol 18, 1765-1769

17. Furbass R, Vanselow, J. An aromatose pseudogene is transcribed in the bovine placenta. Gene, 1995. 154:287-291

18. Gerstein, Mark, Harrison, Paul, Zhang Zhaolei. Millions of Years of Evolution preserved: A Comprehensive catalog of the Processed Pseudogenes in the Human Genome. Genome Research, 2003. Vol 13:2541-2558

19. Glusman et al. The complete human olfactory subgenome. Genome research, 2001. 11:685-702

20. Haldane, J.B.S. The causes of evolution. Princeton University Press, 1996. Pg 19-78

21. Harrsion et al. Molecular Fossils in the Human Genome: Identification and Analysis of the Pseudogenes in Chromosomes 21 and 22. Genome research, 2002. 12:272-280

22. Harrison et al. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic acids research, 2005. Vol. 33. No.8

23. Harrison, Paul M., Milburn Duncan, Zhang Zhaolei, Bertone, Paul, Gerstein, Mark. Identification of Pseudogenes in the Drosophila melanogaster genome. Nucleic Acids Research, 2003, Vol.31:1033-1037

24. Harrsion, P., Echlos, N., Gerstein, M. Digging for Dead Genes: An Analysis of the Characteristics of the Pseudogene Population in *C.elegans*. Genome. Nucleic Acids Research, 2001. 29:818-830

25. Hillis D.M, Monitz C., Mable B.K. Molecular Systematics, Sinauer. 1996. Pg 205-232

26. Hirotsune, S, Yoshida N et al. An expressed pseudogene regulates the mRNA stability of its homolog coding gene. Nature, 2003. 423:91-96

27. Hurles, Matthew. Gene Duplication: The Genomic Trade in Spare Parts. PLoS Biol. 2004 July; 2(7): e206

28. Korneev et al. Neuronal Expression of Neural Nitric Oxide Synthase (nNOS) Protein Is Supressed by an Antisense RNA Transcribed from an NOS pseudogene. The Journal of Neuroscience, 1999, 19(18):7711-7720

29. Li et al. Pseudogenization of a Sweet-Eeceptor Gene Accounts for the Cats Indifference toward Sugar. PLoS Genetics, 2005. Vol 1, Issue 1: e3

30. Li, Wen-Hsiung. Molecular evolution. Sinauer associates, Inc. publishers. 1997. Pg 269-285

31. Lynch M, Conery J. S. The evolutionary fate and consequences of duplicate genes. Science. 2000; 290:1151–1155

32. Mighell,A.B, Smith, N.R, Robinson,P.A., Markham,A.F. Vertebrate pseudogenes. FEBS letters, 2000. 468, 109-114

33. Misra et al. Annotation of the Drosophila melanogaster euchromatic genome: A systematic review. Genome Biology 2002 Vol 3, Issue 12

34. Mounsey et al. Evidence suggesting that a fifth of annotated C.elegans genes may be pseudogenes. Genome Research, 2002. 12:770-775

35. Ohno, Sushmo. Evolution by gene Duplication. Springer-Verlag. 1970. Pg 1-65.

36. Olsen M.A., Schechter L.E. Cloning, mRNA localization and evolutionary conservation of human 5-HT7 receptor pseudogene. Gene,1999. 227. Pg 63-69

37. Petrov, D.A, Hartl D.L. Pseudogene Evolution and Natural Selection for a compact genome. The American Genetic Association , 2000. Vol 91, Pg 221-227

38. Powell, Jeffery R. Progress and Prospects in Evolutionary Biology: The Drosophila Model. Oxford Press, 1997. Pg 1-20

39. Qian, I.H et al. A human serotonin-7-receptor pseudogene. Mol. Brain res., 1998. 53(1-2), 339-343

40. Rehwinkel et al. Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. RNA, 2005; 11: 1530-1544

41. Richardson et al. Molecular cloning and characterization of a highly conserved human 67-kDa laminin receptor pseudogen mapping to Xq21.3. Gene 1998. 206:145-150

42. Rogers, M.A, Winter, H, Wolf C, Heck M, Schweizer J. characterization of a 190-kilobase pair domain of human type I hair keratin genes. J.Biol.Chem., 1998.273:26683-91

43. Ruud et al. Identification of a novel cytokeratin 19 pseudogene that may interfere with reverse transcriptase-polymerase chain reaction assays used to detect micrometastatic tumor cells. Int. J. of, 1999. 80:119-125

44. Schuster W, Brennicke A. RNA editing makes mistakes in plant mitochondria:editing loses sense in transcripts *rps19* pseudogene and in creating stop codon in *cox1* and *rps3* mRNAs in *Oenothera*. Nucleic Acid Research, 1991. 19:6923-6928

45. Sequencher – Gene Codes Corporation http://www.genecodes.com/sequencher/index.html

46. Sorge, J, Gross E, West C, Beutler,E. high level transcription of the glucocerebrosidase pseudogene in normal subjects and patients with gaucher disease. J.Clinical Invest, 1990. 86:1137-41

47. Takemura et al. Active transcription of the pseudogene for subunit 7 of the NADH dehydrogenase in *Marchantia polymorpha* mitochondria. Mol. Gen. Genet. 1995. 247:565-570

48. The College of New Jersey website. bricker.tcnj.edu/ tech/Geneiso.html Accessed October 18th, 2005

49. The Drosophila Virtual Library page. http://www.ceolas.org/fly/ Accessed August 26, 2005

50. The national health museum webpage.
http://www.accessexcellence.org/RC/VL/GG/polymerase.html
Accessed July 22, 2005

51. Theile et al. Expression of the gene and processed pseudogene encoding the human and rabbit translationally controlled tumor protein (TCTP) . Eur. J. of Biochemistry, 2000. 267:5473-81

52. Toba, Gakuta, Aigaki, Toshiro. Disruption of Microsomal glutathione S-transferase-like gene reduces life span of *Drosophila melanogaster*. Gene 253:179-187, 2000

53. Torrents D, Suyama M, Zdobnov E, Bork P. A genome-wide survey of human pseudogenes. Genome Res. 2003; 13:2559–2567

54. Tourmen et al. Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics, 2002. 80:71-77

55. Weil et al. Antisense transcription of a murine FGFR-3 pseudogene during fetal development. Gene, 1997. Issue 187, 115-122

56. Zhang et al. Comparative analysis of the processed pseudogenes in Mouse and humans. TRENDS in Genetics Vol.20 No.2 February 2004

57. Zhang J, Pontoppian B. The third myrosinase gene TGG3 in Arabidopsis thaliana is a pseudogene specifically expressed in stamen and petal. Phsiol. Plant, 2002. 115:25-34

58. Zhang P., Gerstein, M. Identification and characterization of over 100 mitochondrial ribosomal protein pseudogenes in human genome. Genomics, 2003. 81:468-480

59. Zhang P, Gu Z, Li WH. Different evolutionary patterns between young duplicate genes in the human genome. Genome Biol. 2003; 4:R56

60. Zhou, B.S., Beidler D.R., Cheng Y. C. Identification of antisense RNA transcripts from a human DNA topisomerase I pesudogene. Cancer research, 1992. Vol 52, Issue 15, 4280-4285

BIOGRAPHICAL INFORMATION

Aditi Bhardwaj had a passion for biology which incited her to earn her Undergraduate and Graduate degree in Biosciences and Plant Sciences respectively from India. Later, she joined the University of Texas at Arlington in August 2003 and completed her degree Master of Science in Biology in December 2005. In future she is determined to pursue her career in research.