THE USE OF SUPPRESSION SUBTRACTIVE HYBRIDIZATION IN THE IDENTIFICATION OF A NOVEL GENE ENCODING A PROTEIN CONTAINING A BTB-POZ DOMAIN IN THE MEDITERRANEAN FRUIT FLY, CERATITIS CAPITATA

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOMEDICAL SCIENCES (CELL AND MOLECULAR BIOLOGY)

DECEMBER 2002

By

Pia Marie Untalan

Dissertation Committee:

David S. Haymer, Chairperson
Rebecca L. Cann
David M. Jameson
Donald O. McInnis
Steven N. Robinow
I would first like to thank Dr. David Haymer for motivating me to pursue a career in graduate school. I thank my lucky stars that I found my way into your lab, and I appreciate the never-ending support you have provided me in the lab over the years.

I am grateful to Dr. John Hunt for allowing me to spend time in his lab honing my research skills. I am especially indebted to Dr. Rebecca Cann for always fostering an environment of independent, critical thinking throughout my graduate career. Thank you for being such a truly wonderful role model.

I am grateful to the Biology Program for providing me with invaluable teaching experience over the years. What started off as simply a means to fund graduate school has blossomed into a true love for teaching.

I honestly would not have been able to survive all of this if it weren’t for the phenomenal support from the Haymer Lab “Fly Girls” (and the occasional Fly Boy). I am especially grateful to Janice, Chrystie, and Keao—I’m convinced that were separated at birth!! I’ll miss the marathon lunch sessions, lab fieldtrips (woohoo!), and hours and hours of laughter. Thank you for being my family away from home.

Lastly, Si Yu’us Ma’ase to my family (and there are a LOT of them!). Your unconditional love and support has made me what I am today. My love to Rod for the friendship and kindness you’ve shown me over the years. Lastly, my love to Na and Gramps for being the best parents. My love to Peter for being not only a wonderful big brother and my inspiration, but my best friend as well.
ABSTRACT

Differential gene expression plays a key role in developmental pathways within an organism. Examples of such pathways include primary sex determination signaling and the formation of secondary sexual characteristics. This dissertation is focused on the use of suppression subtractive hybridization (SSH) to identify genes that are differentially expressed and involved in some aspect of sexual development in the Mediterranean fruit fly (medfly), Ceratitis capitata. In the course of this project, a method for sexing individual specimens from pre-adult stages was developed. This method was used to collect sex-specific RNAs at different developmental stages for use in SSH. A total of 25 subtraction products were obtained across all the stages examined. Analysis of these products revealed that approximately half were similar to cytoplasmic ribosomal proteins and mitochondrial ribosomal RNA. The remaining products represent putative medfly homologs of other previously identified genes or potentially novel genes. One of the subtraction products, representing a potentially novel gene, was characterized in detail.

This gene, named mapotige’, represents a novel medfly gene that appears to encode a polypeptide of 299 amino acids. The N-terminus of this polypeptide contains a BTB-POZ domain. This domain functions as a protein-protein interaction motif found in a wide range of organisms from humans to Drosophila that mediates protein dimerization and oligomerization. The temporal expression pattern of mapotige’ was determined using RT-PCR and Northern blot analysis. These revealed that the transcript is expressed throughout embryogenesis in both females and males, and in adult females that are >0.5 days post-eclosion. Minimal expression is observed in female and male third instar larvae, early pupae, and in adult males.

Studies were also initiated to characterize the representation of additional sequences containing a BTB-POZ domain in the medfly genome. This was performed using Southern blot analysis and degenerate primers for the polymerase chain reaction (PCR). These results indicate the presence of at
least three sequences in the medfly, in addition to *mapotge*, that contain a BTB-POZ domain.

Potential evolutionary relationships of the BTB-POZ domain sequences from the medfly and other insect species were also analyzed.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
</tr>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>List of Tables</td>
</tr>
<tr>
<td>List of Figures</td>
</tr>
<tr>
<td>List of Abbreviations</td>
</tr>
</tbody>
</table>

## Chapter 1: Introduction

1.1 Overview of the Differential Expression of Genes During Development | 2 |
1.2 Elements of Control of Gene Expression | 3 |
1.3 Differential Gene Expression in Insect Development | 9 |
1.4 Mediterranean Fruit Fly | 15 |

## Chapter 2: Materials and Methods

2.1 Materials | 21 |
2.2 Methods | 22 |

## Chapter 3: Application of Suppression Subtractive Hybridization (SSH) to Developmentally Staged Material from the Genome of the Medfly

3.1 Background | 43 |
3.2 Results | 47 |
3.3 Discussion | 71 |

## Chapter 4: Characterization of Subtraction Products

4.1 Background | 80 |
4.2 Results | 80 |
4.3 Discussion | 110 |
Chapter 5: Preliminary Characterization of the Representation of Additional Genes Encoding the BTB-POZ Domain in the Genome of the Medfly

5.1 Background ........................................ 116
5.2 Results ........................................... 118
5.3 Discussion ....................................... 126

Appendix: Oligonucleotide Primer Sequences ........................................ 131
References ............................................. 133
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Cumulative results of suppression subtractive hybridization conducted using adult, larval, and embryonic material</td>
<td>50</td>
</tr>
<tr>
<td>3.2 Results of subtractive hybridization using adult and third in-star larval material</td>
<td>55</td>
</tr>
<tr>
<td>3.3 Subtraction products representing mitochondrial ribosomal RNA sequences isolated using subtractive hybridization</td>
<td>57</td>
</tr>
<tr>
<td>3.4 Various medfly homolog sequences isolated from different stages of embryogenesis using subtractive hybridization</td>
<td>68</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Model for sex-specific regulation of $S_{xl}$</td>
<td>11</td>
</tr>
<tr>
<td>3.1</td>
<td>Principle of Suppression Subtractive Hybridization (SSH)</td>
<td>45</td>
</tr>
<tr>
<td>3.2</td>
<td>Y114 PCR Amplification</td>
<td>48</td>
</tr>
<tr>
<td>3.3</td>
<td>Subtractive hybridization products obtained using material from the adult and third instar larval stages</td>
<td>53</td>
</tr>
<tr>
<td>3.4</td>
<td>Amino acid sequence alignment of adult subtraction product sequences, APMS.1 and APMS.2, with respective homologous sequences</td>
<td>54</td>
</tr>
<tr>
<td>3.5</td>
<td>Amino acid sequence alignment of embryonic subtraction products 3-5PMS.4 and 5-18PMS.5 with representative 60S ribosomal protein L13a sequences</td>
<td>59</td>
</tr>
<tr>
<td>3.6</td>
<td>Amino acid sequence alignment of embryonic subtraction product 3-5PMS.5 with representative 60S ribosomal protein L35a sequences</td>
<td>60</td>
</tr>
<tr>
<td>3.7</td>
<td>Amino acid sequence alignment of embryonic subtraction product 5-18PMS.1 with representative 60S ribosomal protein L10 (QM) sequences</td>
<td>61</td>
</tr>
<tr>
<td>3.8</td>
<td>Amino acid sequence alignment of embryonic subtraction product 4-7PFS.4 with representative 60S ribosomal protein L18a sequences</td>
<td>62</td>
</tr>
<tr>
<td>3.9</td>
<td>Amino acid sequence alignment of embryonic subtraction product 5-18PMS.4 with representative 40S ribosomal protein S20 sequences</td>
<td>63</td>
</tr>
<tr>
<td>3.10</td>
<td>Amino acid sequence alignment of embryonic subtraction product 3-5PMS.2 with representative histone H2B sequences</td>
<td>64</td>
</tr>
<tr>
<td>3.11</td>
<td>Amino acid sequence alignment of embryonic subtraction product 5-18PMS.3 with representative 20S proteasome β2-subunit sequences</td>
<td>65</td>
</tr>
<tr>
<td>3.12</td>
<td>Amino acid sequence alignment of embryonic subtraction product 5-18PMS.6 with representative translation elongation factor eEF1-α sequences</td>
<td>66</td>
</tr>
<tr>
<td>3.13</td>
<td>Amino acid sequence alignment of embryonic subtraction product 3-5PFS.3 with representative nucleoside diphosphate kinase (NDK) sequences</td>
<td>67</td>
</tr>
<tr>
<td>3.14</td>
<td>RT-PCR analysis of 3-5PFS.1, 3-5PFS.2, and 3-5PFS.4</td>
<td>72</td>
</tr>
<tr>
<td>3.15</td>
<td>Northern blot analysis of cytoplasmic actin expression using variable amounts of unsexed embryonic material</td>
<td>79</td>
</tr>
<tr>
<td>4.1</td>
<td>Schematic overlap of clones used to deduce the sequence of CePFS.2</td>
<td>82</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>4.5</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>4.6</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>4.7</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>4.8</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>4.9</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>4.10</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>4.11</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>4.12</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>4.13</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>127</td>
<td></td>
</tr>
</tbody>
</table>

**Figure**

- 4.2 Schematic overlap of clones used to deduce the sequence of CcPFS.1
- 4.3 CcPFS.1A nucleotide sequence and conceptual translation
- 4.4 CcPFS.1A sequence alignment with representative transposon sequences
- 4.5 Southern Blot analysis of the genomic representation of CcPFS.1B
- 4.6 mapotge' nucleotide sequence, gene structure, and conceptual translation
- 4.7 mapotge' sequence alignment with representative BTB-POZ domain sequences
- 4.8 Identification of conserved residues within the BTB-POZ domain
- 4.9 RT-PCR analysis of the mapotge' transcript throughout development
- 4.10 RT-PCR analysis of the mapotge' transcript during embryogenesis
- 4.11 RT-PCR analysis of the mapotge' transcript in the thorax and abdomen of adult females and males
- 4.12 Analysis of the 3' end of mapotge' in males
- 4.13 Northern blot analysis of mapotge' expression
- 5.1 Domain architecture of proteins containing a BTB-POZ domain
- 5.2 Genomic representation of sequences potentially encoding a BTB-POZ domain in the medfly
- 5.3 Amino acid sequence alignment of BTB-POZ domain sequences
- 5.4 Nucleotide and amino acid sequence alignment of medfly tramtrack-like and Drosophila (Dm) tramtrack (ttk) BTB-POZ domain sequences
- 5.5 Amino acid sequence alignment of BTB7 and BTB8 with respective homologous sequences
- 5.6 Genomic representation of BTB5
- 5.7 Phylogenetic tree of arthropod BTB-POZ domain sequences
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEL</td>
<td>after egg laying</td>
</tr>
<tr>
<td>Arg</td>
<td>arginine</td>
</tr>
<tr>
<td>Asp</td>
<td>aspartic acid</td>
</tr>
<tr>
<td>BLASTN</td>
<td>nucleotide BLAST search algorithm</td>
</tr>
<tr>
<td>BLASTX</td>
<td>translated BLAST search algorithm</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BTB-POZ</td>
<td>bric-a-brac, tramtrack, Broad-complex/pox-virus zinc finger protein domain</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary deoxy ribonucleic acid</td>
</tr>
<tr>
<td>CSPD</td>
<td>disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2’-(5’-chloro)tricycle[3.3.1.1^3^7]decan}-4-yl)phenyl phosphate</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DIG-dUTP</td>
<td>digoxigenin-11-dUTP</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxy ribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetra-acetic acid, disodium salt</td>
</tr>
<tr>
<td>FST</td>
<td>female-specific transcript</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>His</td>
<td>histidine</td>
</tr>
<tr>
<td>hrs</td>
<td>hours</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pair</td>
</tr>
<tr>
<td>Leu</td>
<td>leucine</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>mins</td>
<td>minutes</td>
</tr>
<tr>
<td>ml</td>
<td>milliliter</td>
</tr>
<tr>
<td>mm</td>
<td>millimeter</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>mI</td>
<td>milliliter</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>oligo</td>
<td>oligonucleotide</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PFS</td>
<td>putative female specific sequence</td>
</tr>
<tr>
<td>pfu</td>
<td>plaque forming unit</td>
</tr>
<tr>
<td>PMS</td>
<td>putative male-specific sequence</td>
</tr>
<tr>
<td>poly (A)</td>
<td>polyadenosine</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNAPII</td>
<td>RNA Polymerase II</td>
</tr>
<tr>
<td>Ser</td>
<td>serine</td>
</tr>
<tr>
<td>SSH</td>
<td>suppression subtractive hybridization</td>
</tr>
<tr>
<td>TB</td>
<td>Terrific Broth</td>
</tr>
<tr>
<td>Tyr</td>
<td>tyrosine</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>μg</td>
<td>microgram</td>
</tr>
</tbody>
</table>
μl microliter

Zw glucose-6-phosphate dehydrogenase
Chapter 1

Introduction

In metazoans, the formation of a fully developed, multi-cellular organism, comprised of a multitude of different cell types that are intended to perform specific roles, involves an intricate progression of events that originates from a single, fertilized egg. This single cell undergoes a series of divisions (cleavages) to produce groups of cells that will eventually be committed to a particular fate (determination). The specialization of these cell-types (differentiation) is achieved primarily through the regulation of gene expression in the various cell lineages. Although nearly all cells within an organism contain the same genetic information (genomic equivalence), it is the orchestrated expression of different gene products in different cells at different times during development that produces the suite of proteins and other gene products necessary for the development of an organism.

Although several genes in the Mediterranean fruit fly (medfly), Ceratitis capitata, have been identified that are differentially expressed in females versus males at the adult stage, I am particularly interested in genes that may be differentially expressed and involved in some aspect of sexual development at the embryonic stage. Embryogenesis is a critical phase in which increasing complexity is generated by the differential expression of genes in different cell types. The identification of genes that play a role in causing the different developmental events to take place is of fundamental interest, biologically, in the medfly. Although aspects of development may be tightly conserved within insects, variability in regulatory pathways does exist. The most notable variation is seen in the different mechanisms by which sex is determined between insect species (Lauge, 1985) and even within the Diptera (reviewed in Schütz and Nöthiger, 2000).

The examination of differential gene expression during embryogenesis will expand the information available about how the mechanisms of gene regulation are utilized in the medfly. In addition, analysis of the different products that are expressed in females versus males is of interest.
with respect to broadening the arsenal of molecular tools that can be utilized in the medfly for prospective biological control methods.

The specific aims of this project include:

1) The development of a system to collect sex-specific pools of RNA from morphologically indistinguishable embryos,

2) The survey of whether a suppression-subtractive hybridization method can be utilized successfully to isolate differentially expressed sequences in female versus male medfly embryos, and

3) The characterization of sequences of interest that may arise from such a screen.

1.1 Overview of the Differential Expression of Genes During Development

Although a genome may contain tens of thousands of genes, only subsets of them are expressed in a cell at a given time (Liang and Pardee, 1992). The differential expression of specific genes is achieved in a variety of ways, including modification of chromatin structure, initiation of transcription, transcript processing and modification, translation initiation, and post-translational modification. At the transcriptional level alone, temporal and tissue-specific promoters that are present in many genes restrict expression to a particular stage in development or to a particular region within the developing organism by affecting the initiation of transcription (reviewed in Maniatis, et al., 1987; Harshman and James, 1998). A number of tissue-specific enhancers have also been identified (reviewed in Latchman, 1995b).

Alternative splicing, another mechanism used to regulate gene expression, involves the processing of complex transcription units in alternative ways to produce different mRNAs (Lodish, et al., 1999). This is achieved through the use of alternative 5' and 3' splice sites, optional exons, and retained introns (Lopez, 1998). This mechanism produces isoforms of proteins involved in cell function and
physiology, and is also used to create isoforms of transcription factors that are made in different
tissues and stages of development (reviewed in Lopez, 1995).

Modulating the length of the poly(A) tail is another mechanism by which genes can be
differentially expressed. Upon export of a processed mRNA from the nucleus into the cytoplasm, the
poly(A) tail is subject to removal and lengthening by cytoplasmic factors (Richter, 1999). The
resulting length of the poly(A) tail corresponds to the rate of translation initiation of certain mRNAs
such that increasing the length of the poly(A) tail leads to activation of translation and decreasing the
length of the poly(A) tail leads to translational repression (Richter, 1999; Johnston and Lasko, 2001).

1.2 Elements of Control of Gene Expression

1.2.1 Chromatin Remodeling

Eukaryotic DNA is packaged into chromatin, the primary structural unit of which is the
nucleosome. Approximately 146bp of DNA is wrapped around an octamer of core histone proteins
(H2A, H2B, H3, and H4) to form a nucleosome particle, all of which are connected by linker DNA on
which is situated another class of histone protein (H1). This string of particles coils first into a
solenoid then into a filament and, upon further compaction, into a supercoiled filament. It is this tight
packaging that can prevent access of the transcriptional machinery, as well as the replication
machinery, to the DNA molecule. The activation of transcription typically begins when gene-specific
transcriptional activator proteins bind to and de-compact the chromatin structure (Narlikar, et al.,
2002). The role of these activators is to remodel the chromatin surrounding a promoter sequence of
interest since these regions are essential for the docking of the RNA Polymerase. The activators also
recruit co-activators that encode subunits of enzymatic complexes involved in chromatin remodeling
and histone modification (McKenna and O’Malley, 2002).
1.2.2 Transcription Initiation

The initiation of transcription is a multi-step process that includes DNA-protein and protein-protein interactions and assembly. The initiation involves a suite of regulatory DNA sequences that are typically found near the gene (in cis) as well as transcription factors, which are proteins that are encoded by genes unrelated to the target gene (in trans).

1.2.2.1 Regulatory DNA Sequences

The regulatory DNA sequences that control gene expression include promoters, enhancers, and response elements. Promoter sequences are primarily located immediately upstream (5') of the start of transcription and can be divided into two groups, i.e. core/basal promoters and promoter proximal elements. The core promoter sequence is highly conserved in sequence, position, and function and often includes a sequence with a TATA nucleotide motif (the TATA “box”). The TATA binding protein (TBP), which serves as a point of interaction with various accessory proteins, typically interacts with the TATA box sequence. In place of a TATA box, some genes contain an initiator (Inr) element near the transcription start site that acts as an alternative promoter. These sequences have been identified in vertebrates (Javahery, et al., 1994) and an arthropod-specific initiator sequence has been recognized that appears to be present in a high percentage of arthropod promoters (Cherbas and Cherbas, 1993). Eukaryotic RNA polymerases require the help of a number of accessory proteins (general transcription factors [GTF]) to locate the proper promoter or initiator sequence and begin transcription. The collection of GTFs and the RNA Polymerase, together, form the transcription-initiation holoenzyme complex.

The core promoter typically provides a low level of gene expression. This rate is increased by the interaction of transcription factors at sites upstream of the TATA box that enhance the stability of the transcription complex. These additional sites are referred to collectively as promoter proximal elements and consist of sequences like the CCAAT box, the octamer sequence, and GC-rich promoter
elements. Enhancers are another class of regulatory sequences that bind transcription factors and influence transcription by increasing the rate at which it is initiated (Thompson and McKnight, 1992). Enhancers can be located upstream, downstream, or within a gene and can be positioned up to thousands of bases away from the target gene and still influence its expression. Enhancers can also activate promoters when placed in either orientation relative to the promoter. Enhancer-promoter interactions are regulated by two mechanisms: promoter competition and insulator DNAs (reviewed in Levine, 2001). In the promoter competition mechanism, a shared enhancer can activate multiple genes but prefers the promoter region associated with one of the genes. Insulator DNAs are sequences that prevent enhancers in one gene from interacting with promoters of neighboring genes.

A third class of regulatory sequences that are involved in controlling gene expression are response elements, which play a role in varying transcription in response to environmental factors or hormonal signals (reviewed in Harshman and James, 1998; Lodish, et al., 1999). In response to environmental stresses such as heat and exposure to heavy metals, eukaryotic cells respond by halting transcription and translation of most mRNAs and focusing on the production of proteins that promote cell survival, i.e. heat shock proteins and metallothioneins. In reaction to cellular stress, specific transcription factors are produced that bind at response elements upstream of the respective gene promoters to increase gene transcription. Hormones are another extracellular signal to which a cell responds. Lipid-soluble hormones with intracellular receptors are capable of traversing the plasma and nuclear membranes and interact with hormone receptors in the cytoplasm. Hormone receptors contain short DNA binding domains that then interact with hormone response elements in the nucleus to activate target genes “in response” to the hormone.

1.2.2.2 Transcription Factors

Transcription factors (TF) are integral components of the regulation of gene expression because they control the choice of initiation sites for transcription and are essential for directing the “on” or
states of genes. Brivanlou and Darnell (2002) proposed that TFs could be categorized according to their function into constitutive and regulatory classes. Based on this categorization, constitutive TFs, also referred to as GTFs, include factors that are necessary for the assembly of the RNA Polymerase II transcription-initiation complex and can interact with the promoter-proximal elements mentioned earlier to mediate their effect. Regulatory TFs comprise those that play a role in influencing the rate of gene transcription. This category can be further subdivided into those that can enter the nucleus and immediately affect transcription and those that require an appropriate intracellular or extracellular signal to be activated (Brivanlou and Darnell, 2002).

TFs contain two characteristic domains that facilitate their function: a DNA-binding domain and an activation domain (Lodish, et al., 1999). The structural motif that the DNA binding domain contains has traditionally been used to categorize TFs. TFs function via protein-DNA and/or protein-protein interactions. The more common motifs present in TFs that mediate DNA binding are the homeobox, the POU domain, zinc fingers, the helix-loop-helix (HLH), and leucine zippers, the first three of which bind to DNA molecules via an α-helical interaction at the DNA major groove.

The homeodomain is a conserved amino acid sequence found in homeotic genes, which play a role in segment identity along the body axis and in tissue patterning during development (reviewed in Bondos and Tan, 2001). The domain forms a helix-turn-helix (HTH) motif that interacts with DNA in a sequence-specific fashion and is present in many transcription factors. The homeodomain often occurs along with other domains, such as POU and LIM. The POU (Pit-Oct-Unc) domain mediates protein interaction and increases the affinity with which the homeodomain binds its DNA substrate (reviewed in Bondos and Tan, 2001) while the LIM domain, in contrast, represses DNA binding by the homeodomain; the LIM-homeodomain proteins are de-repressed when bound by proteins that interact at LIM (reviewed in Bondos and Tan, 2001).

The third type of motif is the zinc finger motif, the most common of which is the C2H2 arrangement. This is a motif characterized by a repeating unit that binds a zinc ion via two cysteine
(C2) and two histidine (H2) residues. This forms a base from which intervening amino acids form a "finger" composed of two anti-parallel β-sheets with an adjacent α-helix (Latchman, 1995a). The amino acids comprising this helix play a critical role in mediating sequence-specific binding.

The HLH and leucine zipper domains contain α-helices that mediate protein-protein dimerization rather than DNA binding. The helix-loop-helix (HLH) motif is characterized by two amphipathic α-helices, separated by a linker region (the loop), that mediate dimerization between proteins (Lewin, 1994). Some HLH-containing proteins contain a region adjacent to the helices that is highly basic and interacts with acidic DNA; these are referred to as bHLH proteins. DNA binding is mediated by the formation of homo- or hetero- dimers. A group of HLH-containing proteins lack the basic region (nonbasic-HLH) but can still form dimers. In this case, association of a nonbasic-HLH subunit with a bHLH subunit forms a dimer that is does not bind DNA specifically.

The leucine zipper motif is characterized by the presence of hydrophobic leucine at every seventh amino acid but, like the HLH motif described above, it does not directly interact with DNA. Instead, these regions are critical for dimerization of DNA-binding proteins (Latchman, 1995a). These regions form an α-helical structure in which leucine residues fall on the same side of a helix; one helix of a protein is capable of interacting with a similar helix of another protein to form a coiled-coiled structure ("zipper") that strengthens dimerization. This motif is also referred to as bZIP because adjacent regions of the dimerized protein that are rich in basic amino acids interact with acidic DNA to mediate DNA binding.

A variety of sequence motifs, apart from the DNA binding domain, are associated with TFs. One group of these sequences, protein-protein interaction motifs, regulates protein-protein interactions of TFs with each other and with other proteins. The KRAB domain is a conserved amino acid sequence that appears to be restricted to vertebrates, that is associated with transcription factors having multiple C2H2 zinc fingers at the carboxy terminus, and that is involved in transcription repression (reviewed in Collins, et al., 2001). Once the TF is bound to DNA via the zinc finger, the KRAB domain recruits
transcriptional co-repressors that mediate gene silencing. Another protein-protein interaction motif is the BTB-POZ domain, named after the Drosophila genes bric-a-brac, tramtrack, and Broad-complex and the pox-virus zinc finger protein, all of which have the domain in common. The domain is a conserved amino acid sequence found in a number of TFs containing C2H2 zinc fingers and which can be found in proteins containing actin binding motifs (reviewed in Collins, et al., 2001). The BTB-POZ domain mediates dimerization and oligomerization of proteins containing the domain and can recruit co-factors that mediate gene expression (Zollman, et al., 1994); the domain is not restricted to vertebrates. A characteristic shared by these domains is that they are always present at the amino terminus of the proteins in which they are found.

1.2.3 Post-Transcriptional Processing and Modification

Another level at which gene expression is controlled is through transcript processing and modification, both of which require RNAPII. Nascent transcripts produced are associated with heterogeneous ribonucleoprotein particles (hnRNPs) that play a role in the subsequent processing and transport of mRNAs (Lodish, et al., 1999). After 20-30 bases of RNA have been transcribed, the 5' end of the nascent transcript is modified by the addition of 7-methylguanosine (5' cap). The capping enzyme selectively binds to a region of RNAPII and this provides for the rapid targeting of the capping apparatus to RNAPII transcripts (reviewed in Hirose and Manley, 2000). In addition, an intact 5' cap is essential for efficient splicing and polyadenylation of the transcript (Hirose and Manley, 2000). Current evidence suggests that these processes occur co-transcriptionally (Hirose and Manley, 2000). Splicing is the mechanism by which introns are removed and exons are joined in RNA. This process is performed by the spliceosome, a complex consisting of five small nuclear RNA (snRNA) associated with small nuclear ribonucleoproteins (snRNPs). Evidence suggests that a region of RNAPII is required for targeting components of the spliceosome to transcription sites by facilitating the binding of snRNPs to proper 5' consensus splice sites to increase the efficiency of
splicing (reviewed in Hirose and Manley, 2000; Lodish, et al., 1999). Nearing the end of transcript production, RNAPII encounters a consensus sequence of the type AAUAAA that signals the initiation of polyadenylation. This process involves cleavage of the nascent RNA and addition of a ~200 nucleotide polyadenosine [poly(A)] tail at the 3’ end of the transcript.

1.2.4 Translation Initiation

Subsequent to polyadenylation, hnRNPs that contain processed mRNAs are shuttled from the nucleus to the cytosol through nuclear pore complexes. Once there, translation of mRNAs into polypeptides commences. Protein synthesis occurs at the ribosome, a complex comprised of ribosomal RNA and riboprotein subunits. The eukaryotic ribosome consists of a large ribosomal subunit (60S) and a small subunit (40S) that, together, form the 80S translation initiation complex (Lodish, et al., 1999). Briefly, the 5’ cap of the mRNAs is bound by a pre-initiation complex, which includes the 40S subunit, initiation factors, and a tRNA for methionine. The complex then slides along the mRNA until it reaches a translational initiator codon (AUG) codon for methionine), after which the large ribosomal subunit joins the complex and translation begins. The initiation of translation is affected by the concentration of initiation factors and the sequence of the 5’ untranslated region (5’UTR) preceding the initiator codon. Secondary structures that form near the 5’ cap can affect the ability of the ribosome to bind and the presence of any stem-loop structures in the 5’UTR can prevent ribosome scanning (reviewed in Curtis, et al., 1995). Additionally, sequences in the 3’UTR can repress or activate translation by modulating the length of the poly(A) tail.

1.3 Differential Gene Expression in Insect Development

Examples of the role that differential expression plays in development are available for many insect species. The D. melanogaster sex-determination pathway represents an ideal example of how
different mechanisms are utilized to regulate gene expression in an insect system to produce different proteins in different cells at different times during development.

The primary genetic signal for sex determination in *D. melanogaster* somatic cells is the ratio of X-chromosomes to sets of autosomes (X:A ratio) [reviewed in Schütz and Nathiger, 2000]. This ratio is cell-autonomously assessed so that an X:A ratio of 1.0 signals the female pathway, while an X:A ratio of 0.5 signals the male pathway. The target of the X:A signal is the binary-switch gene, *Sex-lethal* (*Sxl*), which becomes activated in chromosomally female (diplo-X) cells but remains inactive in chromosomally male (haplo-X) cells (Bopp, et al., 1991). The genes responsible for transmitting the X:A signal are activated at the syncytial blastoderm stage, prior to cellularization, and encode transcription factors that fall into 3 classes: numerators, denominators, and maternal elements (Figure 1.1).

Numerator genes are located on the X chromosome and are present in two copies in females (XX) and one copy in males (XY). Four of these genes have been identified—the sisterless genes (*sisA, sisB, and sisC*) and *runt* (Parkhurst, et al., 1993; Kramer, et al., 1999). The diplo-X dose of the numerator genes produces enough numerator protein to activate a *Sex-lethal early/establishment* promoter (*Sxlp*) that creates a pulse of *Sxl* mRNA; this signals the female determination pathway. In males, the haplo-X dose of numerator genes does not produce enough protein to activate *Sxlp*, and the male determination pathway is signaled. The *sis* genes are expressed as early as the third mitotic division and disappear by the 13th mitotic division (Parkhurst, et al., 1993); *runt* is zygotically expressed prior to the expression of *Sxlp* as evidenced by *runt* mutant embryos which lack *Sxlp* expression at the 12th mitotic division (Kramer, et al., 1999).

The only denominator gene identified, deadpan (*dpn*), is located on an autosome and negatively regulate *Sxl*. Its role in sex determination was demonstrated by the female-specific lethality that occurs upon increased doses of *dpn* in addition to the reduced male viability that was observed upon mutation of *dpn* (reviewed in Schütz and Nathiger, 2000). The earliest expression of *dpn* occurs
Figure 1.1 Model for sex-specific regulation of Sxl. (A) Genes responsible for transmitting the X: autosome signal (numerators, denominators, and maternal elements) are involved in the transcriptional regulation of Sxl. (B) X-chromosome counting region (XCR). Yang, et al. (2001) proposed that Sxlp activation is transduced by the number of DA/SISB binding sites that are occupied within the XCR, located proximal to Sxlp. Full occupancy of the sites occurs only in diplo-X females. The DA/SISB binding sites are represented by the shaded ovals. High-affinity sites are shaded black, moderate affinity sites are shaded gray, and low affinity sites are striped.
during the middle of the 13th mitotic division. \textit{dpn} appears, then, to be required for sex determination only during the later stages of X:A signaling since the numerator genes are expressed much earlier in development. Barbash and Cline (1995) suggest that \textit{dpn} is required later to prevent inappropriate activation of \textit{Sxl}_{P}\text{p} in the face of increasing \textit{sis} gene product levels.

Maternal elements help to interpret the X:A signal. Daughterless (\textit{da}) and hermaphrodite (\textit{her}) are known maternal elements that play a role in the positive regulation of \textit{Sxl}. Maternal elements that are negative regulators include \textit{extramachrochaetae (emc)} and \textit{groucho (gro)}.

The \textit{da}, \textit{sisb}, \textit{emc}, and \textit{dpn} gene products all encode transcription factors that contain an HLH DNA binding domain. The product of \textit{da} (DA) has been shown to dimerize with the numerator protein SISB to form a transcriptional complex that activates \textit{Sxl} production by binding to sites near \textit{Sxl}_{P}\text{p} that are located proximal and up to 3kb up-stream of the promoter (Estes, et al., 1995; Hoshijima, et al, 1995; Yang, et. al., 2001) [Figure 1.1]. Yang, et al. (2001) propose that the difference in concentration of the SISB protein between female and male embryos is transduced by the number of DA/SISB binding sites occupied within the X-chromosome counting region (XCR) located proximal to \textit{Sxl}_{P}\text{p}. Full occupancy of the sites occurs only in XX embryos. The remainder of the DA/SISB binding sites in the region 3kb up-stream of \textit{Sxl}_{P}\text{p} are believed to play a role in enhancing transcription of \textit{Sxl} once the female-specific XCR complex forms (Yang, et al., 2001).

Within the promoter proximal region of \textit{Sxl}_{P}\text{p}, binding sites for the \textit{dpn} gene product (DPN) have been identified that are in close proximity to DA/SISB binding sites. The product of \textit{gro}, a maternal product, is believed to dimerize with DPN as a co-repressor (Paroush, et al., 1994), interfering with the assembly of DA/SISB heterodimers at the XCR (Yang, et al., 2001). In addition, the product of \textit{emc} is believed to act as a negative regulator of \textit{Sxl} by binding to DA or SISB to prevent DA/SISB dimerization.

The determinative step in sexual pathway initiation is the sex-specific regulation of \textit{Sxl}_{P}\text{p} (Keyes, et. al., 1992). Production of early SXL from this promoter subsequently specifies the female-specific
splicing of Sxl pre-mRNAs produced from a late/maintenance promoter, Sxl pm. SXL is an RNA-binding protein that controls several alternative-splicing decisions. First, the activity of SXL maintains its expression in female flies by excluding an exon (the “male” exon) from its message that terminates translation. SXL also controls the alternative splicing of a gene downstream of it in the sex-determination pathway, transformer (tra). tra pre-mRNA is present in both sexes, but proper splicing to obtain an mRNA with an open reading frame requires SXL (Boggs, et al, 1987). TRA is a splice site regulator that further mediates the splicing of a gene downstream of it in the pathway, doublesex (d sx). Together with the constitutive gene product of tra2 (TRA2), TRA activates one of two competing 3’ splice sites in the d sx pre-mRNA. This results in a female-specific transcript (DSX F protein) with an alternative carboxyl terminal region. DSX F and DSX M control the expression/repression of genes necessary for sexual differentiation (reviewed in Schütt and Nöthiger, 2000). DSX proteins are transcription factors that bind to DNA in a sequence-specific manner through a zinc finger-like domain in the common amino terminal region of the proteins.

DSX M F plays a direct role in regulating the sex-specific expression of the yolk protein (yp) genes (yp1, yp2, and yp3) that are located on the X-chromosome. Expression of these genes occurs only in the fat body of adult females and in the ovarian follicle cells during particular stages of oogenesis (Ronaldson and Bownes, 1995). DSX M represses the expression of the yolk proteins while DSX F activates expression by interacting with co-activators at a fat-body enhancer element (An and Wensink, 1995).

The sex determination pathway of D. melanogaster provides an example of only one of the plethora of pathways in insects that manipulates the mechanisms of gene expression control to obtain temporal and spatial differential regulation of genes. Homeotic genes, which encode transcription factors that interact at homeobox sequences, are a class of proteins that also play critical roles in orchestrating the molecular events involved in determination of embryonic polarity and segmentation boundaries (reviewed in Akam, 1998) and developmental commitment within segments (reviewed in
Morata and Sanchez-Herrero, 1999; Bondos and Tan, 2001). The genes perform these roles by activation and/or repression of gene products in particular segments of the body plan. Adult appendages of *Drosophila* derive from the imaginal disc primordia; there are 19 imaginal discs in *Drosophila* larva represented by three pairs of head discs (labral, labial, and eye-antennal), three pairs of leg discs, three pairs of dorsal discs (prothoracic, wing, and haltere) and a genital disc (Campos-Ortega and Hartenstein, 1997). Limb and wing development from the leg and wing discs, respectively, are pathways that involve the temporal and spatial expression of homeotic genes in cells comprising the discs (Emerald and Cohen, 2001; Held, 1995; Vervoort, 2000).

The formation of sexually dimorphic abdominal pigmentation in *D. melanogaster* involves the cooperation of signals from both the sex determination pathway and the homeotic pathway (reviewed in Bopp, 2001). The gene involved in the integration of the pathways, *bric-a-brac* (*bab*), is one that contains the BTB-POZ protein-protein interaction motif and acts as a general repressor of pigmentation. In females, *bab* is expressed in all abdominal segments, repressing pigmentation, while in males, *bab* is not expressed in the most posterior segments allowing pigmentation of the posterior abdomen (Kopp, et al., 2000). The homeotic gene, *Abdominal-B* (*Abd-B*), is the primary activator of male pigmentation in posterior abdominal segments A5 and A6 of *D. melanogaster*, whereas the homeotic gene, *abdominal-A* (*abd-A*), is expressed in abdominal segments A2-A7 (reviewed in Kopp, et al., 2000).

Kopp, et al (2000) revealed that *bab* expression is repressed by the activity of *Abd-B*. In the posterior abdomen, where levels of *Abd-B* are high, *bab* expression is low/repressed and pigmentation of this region is produced, whereas in the anterior abdomen, where levels of *Abd-B* are low, *bab* is expressed and pigmentation is repressed. The production of DSX<sub>F</sub>, as part of the sex-determination pathway, overrides the affect of *Abd-B* in the posterior abdomen of females, allowing the expression of *bab* and the repression of pigmentation in this region. Production of DSX<sub>M</sub> is incapable of
repressing Abd-B in the posterior abdomen, bab expression is kept at a low level, and male pigmentation is produced.

Although there is a vast array of literature on the study of differential gene expression and its role in regulating developmental pathways in Drosophila, relatively few similar studies of non-Drosophilid insects have been carried out. In particular, insects that pose primary medical and/or agricultural problems have been used for gene regulation studies in order to identify genes that can be manipulated for control methods.

1.4 Mediterranean Fruit Fly

The Mediterranean Fruit Fly (medfly), C. capitata, is a major agricultural pest that is believed to have originated in sub-Saharan Africa and has spread worldwide via human transport of infested fruits and other commodities (Mitchell and Saul, 1990). It is considered a destructive pest because it attacks as many as 300 varieties of cultivated fruit and vegetable crops, among which half are economically important (Liquido, et al., 1991). Damage to fruit is caused by sting damage upon oviposition of eggs that expose the fruit to fungal and bacterial infections. Damage is also caused by the development of larvae, which live and feed on the flesh of the host plant. Methods currently developed for the control of the medfly include malathion and other bait sprays, the identification of braconid parasitoids of the medfly (reviewed in Mitchell and Saul, 1990; Vargas, et al., 2001), and the use of the Sterile Insect Technique (SIT). SIT is a method that involves the mass-rearing and release of large numbers of insects that have been sterilized, and is the most successful method of genetic control of insects (Robinson, 1998). Because of this, it is an ideal tool to manage the potential eradication of the medfly on an area-wide basis. Employment of SIT in the medfly has shown to be most effective through male-only releases of sterile flies (Hendrichs, et al., 1995).

The use of genetic mechanisms to produce unisexual or male-only progeny is called genetic sexing (Robinson, 1989). Different techniques can be used to produce such strains; in the medfly, radiation-
induced translocations of dominant selectable markers to the Y-chromosome have been primarily utilized. Translocation to the medfly Y-chromosome facilitates the inheritance of the mutation on the chromosome that is responsible for sex determination in males (Zapater and Robinson, 1986; Lifschitz and Cladera, 1989; Willhoeft and Franz, 1996). “First-generation” sexing strains of the medfly were developed using the white pupae (wp) pupal color mutation. The wild type allele (wp$^+$), located on chromosome 5, was linked to the Y-chromosome via radiation-induced translocation; the autosomes carry the mutant wp allele. Females, then, express the wp phenotype (white pupal color) and are separated from the wild-type males (brown pupal color) by a pupal color sorting machine.

The disadvantages of this system included the breakdown of the strain due to genetic recombination, the wasted diet resources used to raise the females to the pupal stage, and the economic burden of the expensive color-sorting machine that was only 95-98% accurate (Hendrichs, et al., 1995) but could be 99% accurate with two passes through the machine (McInnis, et al., 1994).

The “second generation” sexing strains of the medfly were developed using a temperature sensitive lethal (tsl) mutation that allows for the elimination of females at the embryonic or early larval stage as opposed to the pupal stage (Franz and Kerremans, 1994). Strains have been developed employing the wild-type allele of the tsl gene translocated to the Y-chromosome. Presently, the tsl translocation strains are being used in male-only release SIT programs (Fisher, 1998; Fisher, 2000), but there is evidence that even these strains can breakdown under mass rearing conditions (Franz, et al., 1994).

The development of “third generation” sexing strains, as introduced by Willhoeft, et al. (1996), would involve the manipulation of molecular mechanisms (genetic engineering) as opposed to the classical genetic approach taken thus far. Willhoeft, et al. (1996) proposed the necessity for a stable transformation system by which genetic constructs can be introduced into the medfly genome, for the isolation of inducible or sex-specific promoters, and for the identification of additional genes suitable for sexing in order to produce the “third generation”. Since the publication of Willhoeft, et al. (1996),
transformation systems have been identified that operate successfully in the medfly genome. These transformation systems include the Minos-based vector (Loukeris, et al., 1995), the piggyBac vector (Handler, et al., 1998), and the Hermes transposable element (Michel, et al., 2001).

1.4.1 Differential Expression of Genes in the Medfly

The identification of genes that are differentially expressed in the medfly has been achieved using various methodologies. In some cases, cross-hybridization of sequences known to be differentially expressed in Drosophila were used to identify homologues of these sequences in the medfly (Konsolaki, et al., 1990; Tolias, et al., 1990; Rina and Savakis, 1991). Differential screening was also used successfully for this purpose (Tolias, et al., 1990; Rosetto, et al., 1999). The method involves screening a library constructed from the cDNA of a particular tissue; the probes used in the screening are created from two different pools of material that represent the variable states/conditions, i.e. developmental stages, being compared. Those sequences that hybridize to the probe representing the stage of interest are then isolated and characterized.

Differential display is traditionally used to identify differences in mRNA sequences (expression) between pools of material being compared (Liang and Pardee, 1992). A variation of differential display using protein samples isolated from different developmental stages was also performed in the medfly to isolate differentially expressed sequences (Katsoris, et al., 1990).

1.4.1.1 Female-Specific Genes

Chorion genes, which encode eggshell proteins, are present in Drosophila as two clusters located on the X-chromosome (s36 and s38) and chromosome 3 (s18, s15, s19, and s16), the expression of which are restricted to choriogenic-stage follicle cells of the ovary (Spradling, 1981; Griffin-Shea, et al., 1982). The s36 and s38 proteins are expressed in the early stages of choriogenesis while the autosomal chorion cluster is expressed later in oogenesis. The genes are preferentially amplified in
ovaries but not in any other female or male tissue, thus representing a cluster of genes that are
differentially expressed in females and males. Preliminary attempts to isolate chorion gene sequences
from the medfly genome were based on cross-hybridization with Drosophila s36 and s38 sequences
(Konsolaki, et al., 1990; Tolias, et al., 1990). This resulted in the identification of medfly Ccs36 and
Ccs38 chorion genes, which were mapped to medfly chromosome 5; the medfly 5th chromosome is
equivalent to the Drosophila X-chromosome as it contains genes that are similar to other Drosophila
homologues of the Drosophila autosomal chorion locus using Drosophila chorion genes as probes in
low stringency hybridization experiments were unsuccessful (Vlachou, et al., 1997).

The differential screening of a medfly ovarian library using probes designed from prechoriogenic
and choriogenic-stage follicle cells was performed and only genes with expression restricted to
choriogenic-stage follicles were characterized (Tolias, et al., 1990). This resulted in the identification
of the medfly homologues Ccs18 and Ccs19. The remainder of the autosomal chorion locus genes
(Ccs15 and Ccs16) were identified, along with Ccs18 and Ccs19, using an indirect cloning strategy
that utilized the extensive conservation of chromosomal linkage groups in Diptera to initiate a
chromosome walk in the medfly from a gene known to be near the autosomal chorion locus in
Drosophila (Vlachou, et al., 1997).

The female reproductive accessory gland in insects secretes a substance that is involved in the
formation of ootheca (egg cases) in cockroaches, grasshoppers, and mantids and in the production of
an adhesive that allows eggs to stick to the surface on which they are laid (King and Buning, 1985).
The medfly female accessory gland secretion contains strong antibacterial properties and is laid on
the surface of the egg (Marchini, et al., 1997); the antibacterial property is attributed to the presence
of ceratotoxins in the secretion (Marchini, et al., 1991). The ceratotoxin genes—ceratotoxin A, A2,
B, C1, C2, and D—are located on the X-chromosome in the medfly, four of which are present in a
cluster (A, C1, C2, and D) (Rosetto, et al., 1997; Rosetto, et al., 2000). The genes are all exclusively
expressed in sexually mature females (7 days post-eclosion) with expression being slightly enhanced after mating (Marchini, et al., 1995; Rosetto, et al., 1996; Rosetto, et al., 1997). In addition, the genes are expressed in response to juvenile hormone stimulation (Manetti, et al., 1997).

A novel protein, designated female-specific transcript (FST), specific to the medfly female reproductive accessory gland, was identified by differential screening of a cDNA library constructed from sexually mature female accessory glands (Rosetto, et al., 1999). The library was screened using two different probes, one designed from newly emerged female flies and the other synthesized from sexually mature female flies. Clones that hybridized exclusively to the sexually mature female probe were further isolated and characterized, one of which represented a novel female-specific transcript that is expressed exclusively in the female reproductive accessory gland (Rosetto, et al., 1999). FST displays an expression pattern similar to that of the ceratotoxins with the exception that, in addition to being expressed in sexually mature females, FST is expressed in newly emerged females (Rosetto, et al., 1999).

The major egg yolk proteins (vitellogenins) of *Drosophila* are expressed exclusively in the fat body of adult females and in the follicle cells during oogenesis (Kunkel and Nordin, 1985). The isolation of vitellogenins from the medfly genome was initiated in order to obtain promoter sequences that are expressed in only one sex (Rina and Savakis, 1991). The genes (*Vg1* and *Vg2*) were isolated by cross-hybridization of a *Drosophila* yolk protein sequence with a medfly genomic library. Both *Vg1* and *Vg2* display conservation of sequence and structure with *D. melanogaster yp3* and *yp1/2*, respectively (Rina and Savakis, 1991). In addition, it has been shown that these proteins are expressed in the same pattern as the *Yp* genes of *D. melanogaster* (Rina and Mintzas, 1988).

1.4.1.2 Male-Specific Genes

In an attempt to isolate sequences that could be utilized for sex-specific gene expression studies, Katsoris, et al. (1990) compared the haemolymph profile of adult medfly females and males that were
5-days post-eclosion by electrophoretic analysis. Fragments that were unique to the male haemolymph were isolated and further characterized as male-specific serum proteins [MSSP]. The MSSPs comprise a gene family that consists of seven members, grouped as MSSP-α (2 genes), MSSP-β (3 genes), and MSSP-γ (2 genes) based on protein similarity, that represent putative members of the odourant binding protein family based on sequence analysis and that are expressed primarily in the male fat body with weak synthesis in the female fat body and midgut of both sexes (Christophides, et al., 2000a; Christophides, et al., 2000b). Christophides, et al. (2000b) demonstrated that a promoter associated with MSSPα-2 (MSSPα-2PS) drives the expression of a reporter gene specifically in the male fat body; the expression is temporally regulated, starting 72 hours after eclosion and reaching a maximum level 2-3 days later.
Chapter 2

Materials and Methods

2.1 Materials

2.1.1 General chemicals

Chemicals for molecular biology use were purchased from suppliers such as Sigma, Roche Applied Science, Bio 101, Inc., and Invitrogen and are identified within the Methods, where applicable. Recipes for general solutions and buffers used in the following experiments were made following standard recipes found in Molecular Cloning: A Laboratory Manual (Sambrook, et al., 1989) or as per the manufacturer’s protocol for reagents that were purchased pre-measured. Solutions were made using double-distilled water (ddH2O) and were subsequently sterilized by autoclaving.

2.1.2 Fly Stocks

The medfly strains used for these experiments include the Med(+) and Maui Coffee 1993 (MC '93) line. The MC'93 line was obtained from Don McInnis of the USDA facility in Manoa, Hawaii and maintained in our lab. The Med (+) line was obtained from Steven Saul at the University of Hawaii at Manoa and maintained in our lab; its geographic origin is unknown. Medfly adults were raised in population cages on a diet of sugar and yeast (hydrolysate) supplemented with casein and methionine at 25°C. To obtain embryonic material, an egger lined with peach juice attractant was placed into the cage and adults were allowed to lay eggs for fixed time periods of a few hrs. The egger used was a 1 inch (diameter) X 2 inch (length) polystyrene cylinder, with one open end, into which 1 mm holes were drilled along its entire length. The holes mimic sites for oviposition by adult females.

Developmental staged embryos were collected after 2, 3, and 13 hrs of egg laying and then removed from the cages and allowed to age for 3, 4 and 5 hrs at 25°C to obtain 3-5 hour, 4-7 hour, and 5-18 hour aged embryos, respectively. Third instar larval material was collected 12 days after egg
laying (AEL) and early pupal material was collected 14 days AEL. Whole adult material was collected 1, 2, 10, and 15 days post-eclosion for various experiments.

2.2 Methods

2.2.1 Agarose Gel Electrophoresis

Samples were electrophoresed on the desired percentage of agarose (molecular biology grade, Invitrogen) in 1XTBE (5XTBE: 0.445M Tris-borate, 0.445M boric acid, 0.01M EDTA pH 8.0; Research Organics) stained with 1μg/ml of ethidium bromide (Sigma). 1XTBE was also used as the running buffer for all gels. Loading dye (50% glycerol, 0.4% bromophenol blue) was added to each sample and subsequently loaded into individual wells. Gels were electrophoresed at 98V for 1.5 hrs or until the dye front had migrated three-fourths the length of the gel unit. Several standard molecular weight markers were available for electrophoresis alongside all samples: HindIII-digested lambda DNA (Marker II, Roche), HindIII/EcoRI-digested lambda DNA (Marker III, Roche), 100bp and 1kb ladder (New England BioLabs). Upon exposure to UV light, images of gels were obtained for archival purposes using a hand-held photo documentation camera (Fisher) with black and white film (Polaroid).

2.2.2 Restriction Enzyme Digestion.

All restriction enzymes used were purchased at a concentration of 10U/μl from Roche and were accompanied by an appropriate buffer to optimize the activity of the enzyme. Restriction enzyme digestion of genomic DNA was carried out using 1X supplied Reaction Buffer, 3μg genomic DNA, and 30-50U of the appropriate restriction enzyme in a volume of 30μl. Plasmid DNA digestion was performed using 1Xreaction Buffer, 250-500ng of plasmid DNA, and 10U of the appropriate
restriction enzyme in a volume of 20μl. All reactions were incubated in a 37°C air incubator for a minimum of 1.5 hrs up to 18 hrs.

2.2.3 Isolation of Nucleic Acids

2.2.3.1 Genomic DNA Isolation

Adult medfly genomic DNA was isolated using a method from Rick Lifton at Stanford University (the “Lifton” prep). Sex-specific DNA was prepared from adult males and females separated 12-24 hrs post eclosion. Adult flies were immobilized by incubating them at −80°C for 15 mins. Ten adults of each sex were placed, separately, into a chilled 25 ml ground glass homogenizer containing 2.5mls of Lifton grind buffer (0.2M sucrose, 0.05M EDTA, 0.1M Tris pH 9.0, 0.5% SDS) and were homogenized on ice. The sample was strained through a 10ml syringe (no needle) packed with 1ml of sterile 100% polyester fiber into a 6ml Falcon polypropylene tube. The homogenate was treated with Proteinase K at a final concentration of 0.2mg/ml and incubated at 65°C for 1 hour. After proteinase treatment, 375μl of 8M potassium acetate was added to the sample then incubated at −20°C for 1 hour or overnight. The cell debris was pelleted by centrifugation of the sample at 4°C for 15 mins at 10,000rpm. Two volumes of room temperature 95% ethanol were then added to the collected supernatant and the contents were immediately centrifuged at 4°C for 15 mins at 10,000rpm. The pellet was resuspended in 250μl of 1X TE (0.01M Tris pH 8.0, 0.001M EDTA), treated with 1.25μl of DNAse-free RNase A (stock concentration of 10mg/ml) and incubated at room temperature for 10 mins. The sample was extracted twice with 200μl of phenol/chloroform/isoamyl alcohol (25:24:1 v/v; Roche). The addition of 0.1 volume of 3M sodium acetate pH 5.2, two volumes of ice-cold 95% ethanol, and incubation at −20°C for 16-18 hrs induced precipitation of the DNA from the aqueous phase. The DNA was pelleted by centrifugation at room temperature for 15 mins at 10,000 rpm, and then washed with 500μl of ice-cold 70% ethanol. All pellets were resuspended in
80μl of either double-distilled water or 1X TE. For isolation of genomic DNA from single flies, the volumes used in the above protocol were scaled down.

2.2.3.2 Rapid isolation of genomic DNA from developmentally staged embryos, larvae, and pupae

The InstaGene® Matrix (BioRad) was used to successfully isolate genomic DNA from medfly embryos, larvae, and pupae for use in the Polymerase Chain Reaction (PCR). The Instagene matrix is a specially formulated chelating resin that adsorbs cell lysis products, i.e. proteins and lipids that may interfere with PCR, and prevents DNA degradation by binding cations, which are cofactors for DNAses.

Developmental staged embryos were placed individually into 1.5ml reaction tubes using a paintbrush. A pipet tip was used to rupture each embryo and 2μl of RNAse-free water was used to collect as much of the contents of the burst embryo as possible. The collected sample was then transferred to a fresh 1.5ml reaction tube, flash-frozen using dry ice, and stored at -80°C. The residual embryonic material remaining in the initial reaction tube was sufficient for DNA extraction using 100μl of InstaGene® Matrix. Individual larvae and pupae were collected, placed in a 1.5ml Kontes homogenizer, and homogenized in 50μl of RNase-free water. One-fifth of the homogenate was then used for DNA extraction with 100μl of the InstaGene® Matrix, while the remainder of the homogenate was flash-frozen using dry ice and stored at -80°C. All InstaGene® reactions were then incubated at 56°C for 30 mins, vortexed, boiled for 8 mins in a water bath, then centrifuged for 3 mins at 13,000rpm to collect the resin. The supernatant contained the DNA and was used in subsequent polymerase chain reaction (PCR) experiments.
2.2.3.3 Total RNA Isolation

Sex-specific total RNA was isolated from staged embryos (3-5 hour, 4-7 hour, and 5-18 hour), third instar larvae (12d AEL), early pupae (14d AEL), and adults (1, 2, 10, and 15 days post-eclosion) for subsequent RT-PCR and suppression subtractive hybridization experiments. Except for adult tissue, sexed material from each of the developmental stages was pooled from samples stored at -80°C. Fresh adult material was collected for each of the required time points. The TriPure® Isolation Reagent (Roche) was utilized for total RNA isolation from all tissues. The reagent is a clear, red solution of phenol and guanidine thiocyanate (GITC).

All samples were placed into a Kontes homogenizer and were briefly hand homogenized on ice in 400μl of TriPure® Isolation Reagent. An additional 400μl of TriPure® Isolation Reagent was then added and the samples were homogenized further. To assist in the precipitation of the RNA, 10μg of glycogen was added to each sample and incubated at room temperature for 5 mins. 160μl of chloroform was added to each sample, mixed by vigorous shaking for 20 seconds, and incubated at room temperature for 15 mins. The RNA, DNA, and protein-containing phases were then separated by centrifugation at 4°C for 15 mins at 12,000rpm, and the colorless, upper aqueous layer containing the RNA was transferred to a sterile reaction tube. The addition of 400μl isopropanol and incubation at room temperature for 10 mins induced the precipitation of RNA from this phase. The RNA was pelleted by centrifugation at 4°C for 10 mins at 12,000rpm and then washed with 800μl of 75% ethanol. The RNA pellet was stored at -20°C for at least 1 hour, centrifuged at 4°C for 5 mins at 7,500rpm, and air-dried prior to resuspension; pellets were not vacuum-dried. The RNA pellet was resuspended in 25μl of RNAse-free water, treated with 40U of RNAse-free DNAs (Roche), and incubated at room temperature for 20 mins. If not used immediately, the RNA pellet was stored at -80°C for up to six months.
2.2.3.4 mRNA Isolation

Poly (A+) RNA was isolated from DNase-treated total RNA using the PolyA Tract mRNA Isolation System (Promega). A biotinylated oligo(dT) primer was used to hybridize to the 3’ poly(A) region of mature mRNA. The hybrids were captured by incubation of the sample with streptavidin-paramagnetic particles (SA-PMP), washed in 0.5X SSC, and exposure to a magnet. After, the hybrids were washed at a high stringency (0.1X SSC) and the mRNA was eluted from the SA-PMPs using RNase-free water. The isolation was performed according to the manufacturer’s protocol for small-scale mRNA isolation. The mRNA was subsequently used as a template for the synthesis of first-strand cDNA as part of the suppression subtractive hybridization method.

2.2.4 Suppression Subtractive Hybridization (SSH).

SSH was performed using sex-specific mRNA isolated from staged embryos (3-5 hour, 4-7 hour, and 5-18 hour), third instar larvae (12d AEL), and adults (1d post-eclosion). Approximately 100 embryos of each sex (~4mg) were collected for all of the embryonic stages examined. Three third instar larvae of each sex and two adults of each sex were collected for SSH experiments at these stages. Equivalent amounts of material were used within each of the developmental stages based on trial experiments.

The subtractive hybridization was carried out using the PCR-Select™ cDNA Subtraction Kit (Clontech). Female and male material from the different developmental stages listed above was used as tester/driver genomes. Separate experiments were conducted alternating each sex as “driver” in order to isolate prospective male-specific and female-specific sequences.

Double-stranded cDNA was synthesized from separate pools of mRNA, in this case female and male samples from different developmental stages. Tester and driver cDNA were digested with RsaI to create short, blunt-ended fragments representing parts of transcribed genes and then two different adapters, Adaptor 1 (Ad1) and Adaptor 2R (Ad2R), were ligated to the 5’ ends of the tester cDNA.
Excess amounts of driver cDNA were hybridized separately with the Ad1-tester and the Ad-2R tester cDNAs for two rounds of subtractive hybridization. Only those hybrid molecules that were available for amplification using oligonucleotide primers designed from both of the adaptor sequences were of interest (Figure 3.1). All of the subtraction experiments were performed according to the manufacturer's protocol, with the following modifications:

a.) precipitation of cDNA was performed using 3M sodium acetate pH 5.2 instead of 4M ammonium acetate (steps IV.D.12 and IV.E.13),

b.) double-stranded cDNA was precipitated for 16-18 hrs (IV.D.12), and

c.) the first round of hybridization of tester and driver cDNA was incubated at 68°C for ten hrs, while the second round of hybridization was incubated at 68°C for 16 hrs.

Differentially expressed cDNAs were then selectively amplified in two rounds of PCR. In the first round (1°PCR), 0.005 volume of the hybridization mixture was used as a template with "Primer I" to exponentially amplify only those cDNAs with both the Ad1 and Ad2R adaptors at their ends. The second round (2°PCR) was a nested PCR using either no dilution or a 10−1 dilution of the 1°PCR amplification as a template along with primers that anneal to Ad1 and Ad2R; this was designed to enrich for the differentially expressed sequences present in the sample.

All amplifications were performed using a Perkin-Elmer GeneAmp PCR System 2400. The products amplified in the 2°PCR were analyzed on 2% agarose stained with ethidium bromide. All sequences for oligonucleotide primers used in the SSH experiments are reported in Appendix One.

A "hot start" method was performed for the 1°PCR by separating the Taq polymerase and magnesium chloride from the reaction, incubating the remainder of the reaction in the thermocycler at 75°C for 30 seconds, pausing the thermocycler to add the Taq/Mg mix, and then allowing the cycling parameters to commence. The cycling parameters used consisted of an initial incubation at 75°C for an additional 4.5 mins and an initial denaturation at 94°C for 25s followed by 32-37 cycles of 94°C,
10s; 66°C, 30s; and 72°C for 1.5mins. The initial pre-incubation at 75°C is necessary to create a binding site for Primer 1. An additional reaction containing all of the reagents in the absence of template were included in each experiment as a negative control. This reaction was stored at 4°C.

The 2°PCR was performed without using a "hot start" method and consisted of an initial pre-incubation at 75°C for 2 mins and an initial denaturation at 94°C for 25s followed by 35 cycles of 94°C, 30s and 72°C, 2 mins with no distinct annealing temperature. In all of the 2°PCR amplifications, the initial pre-incubation and denaturation remained the same. Slight modifications were made to the number of cycles and the annealing temperature used to amplify the differentially expressed sequences from the various developmental stages. For the 4-7 hour embryonic subtraction, the parameters used were 37 cycles of 94°C, 10s; 68°C, 30s; and 72°C, 1.5mins, and for the 5-18 hour embryonic subtraction, 30 cycles of 94°C, 30s; 68°C, 30s; and 72°C, 1.5mins were used.

A positive control template provided by the manufacturer was used in an SSH experiment and was subsequently used as template in a positive control reaction in initial 1° and 2° PCR experiments. Once it was verified that the procedure could identify subtraction product fragments using the experimental template, the positive control experiment was not included in later amplifications. Experimental templates to which adapters had been ligated, but that were not used in subtractive hybridization, were used as template in 1° and 2° PCR amplifications as well to ensure that a different pattern of fragments was produced when comparing un-subtracted and subtracted templates.

2.2.5 Genomic library screening

A medfly genomic library constructed previously (Haymer, et al., 1990) in EMBL4 was used for screening. The EMBL4 library titer was estimated at 6.86X10⁴ plaque forming units (pfu)/ml. Screening of the library was performed essentially as described in Molecular Cloning: A Laboratory Manual (Sambrook, et al., 1989).
Five mls of LB broth (LB broth base tablets, Sigma, 1 tablet/50 mls ddH₂O) supplemented with a final concentration of 0.02% maltose was inoculated with the *E.coli* strain LE392 and grown at 37°C with shaking for 14-16 hrs. The bacterial cells were pelleted by centrifugation at 4°C for 10 mins at 5,000 rpm and then resuspended in 1ml of 0.01M magnesium sulfate.

Various amounts of phage particles were incubated with a fixed volume (200µl) of LE392 bacterial cells to determine the amount of phage required to infect the bacteria and give ~40,000 pfu/150mm plate. The adsorption mixtures were incubated at 37°C for 20 mins, added to 8mls of 0.7% LB top agarose (LB broth + 0.7% agarose, Invitrogen), and then poured over 1.5% LB-agar (LB agar tablets, Sigma, 1 tablet/50mls ddH₂O) in 150mm diameter petri dishes. The top agarose was allowed to cool prior to incubation of the plates, inverted, at 37°C for 5-7 hrs or until plaques were visible on the bacterial lawn. Plates were stored at 4°C for a minimum of 1 hour before proceeding to plaque lift.

Phage DNAs from the plaques were transferred to a 132mm nylon membrane (Osmonics). Overlaying each plate with a nylon circle membrane for 1 minute and 5 mins was performed to create duplicates of every plate. All membranes were denatured in 1.5M sodium chloride/0.5M sodium hydroxide for 1 minute, neutralized in 1.5M sodium chloride/1M Tris pH 8.0 for 5 mins, rinsed briefly in 2XSSC (20X stock: 3M sodium chloride, 0.3M sodium citrate). The membranes were then exposed to 120,000µJ UV using a UV Stratalinker® 1800 (Stratagene).

The nylon circles were prehybridized for 1-16 hrs at 42°C in a solution containing a specially formulated 2% blocking reagent (Roche), 0.1% SDS, 5XSSC, 0.04% N-lauroylsarcosine, and 50% deionized formamide; the labeled probe was resuspended in the same mixture. Hybridization of the nylon circles with the labeled probe was performed at 42°C for 16-18 hrs. Plaques to which the probe hybridized on the membrane were localized on the original plate and subsequently isolated.

The isolated plaques were placed in 1ml of SM (0.1M sodium chloride, 0.01M magnesium sulfate, 0.05M Tris pH 7.5, and 0.01% gelatin) with 50µl of chloroform. Various dilutions of the
phage supernatant were then re-plated as described above for the initial library screen to give 200 pfu/100mm diameter petri dish. The isolated phage DNA was plated on 1.5% LB-agar at a low density to ensure that individual plaques that formed were isolated from one another. The lawn of plaques were transferred to 82mm nylon circle membranes and treated as described above for the initial library screening. Purified plaques were isolated from their respective plates and placed in 1ml of SM with 50μl of chloroform. The purified phage DNA was then amplified for preparation of confluent plate lysate to be used in the isolation of phage DNA (described below).

2.2.5.1 Isolation of Phage DNA

Isolation of phage DNA was performed using phage lysate collected from confluent plates. 200μl of an LE392 bacterial cell culture, resuspended in 0.01M magnesium sulfate, was infected with various amounts of purified phage DNA and incubated at 37°C for 20 mins for preparation of confluent plate lysates. The adsorption mixture was then added to 3mls of 0.7% LB top agarose and plated on 1.5% LB-agar to give ~5000pfu/100mm diameter petri dish. The plates were incubated, inverted, at 37°C until the edges of the plaques touched. These confluent plates were then overlayed with 5mls of SM and incubated at 4°C for 16-18 hrs. The plate lysate was collected and treated with 1/50 volume of chloroform. Phage DNA was extracted from the lysate using the Lambda DNA Purification Kit (Stratagene) according to the manufacturer’s protocol.

2.2.6 Polymerase Chain Reaction (PCR) Methods

2.2.6.1 Typical PCR conditions

All amplifications were performed using a Perkin Elmer GeneAmp 2400 Thermocycler. Generally, reactions were carried out in a 25μl volume containing 1X PCR Reaction Buffer (100mM Tris-HCl pH 8.3, 500mM KCl, Perkin Elmer), 1.5mM magnesium chloride (Perkin Elmer), 0.2mM
each dNTP (Roche), 5 picomoles each of the forward and reverse oligonucleotide primer, and 1U of Taq Polymerase (Perkin Elmer). Typical cycling parameters consisted of an initial denaturation at 94°C for 2 mins followed by 35 cycles of 94°C for 1 minute, 55°C for 30 seconds, and 72°C for 2 mins, after which the amplification was stored at 4°C. An additional reaction containing all of the reagents in the absence of template were included in each experiment as a negative control. Any variations to the cycling parameter are described within each PCR application below.

2.2.6.2 Method for sexing embryonic, larval, and pupal material

One-tenth of the supernatant from the InstaGene Matrix reaction (Section 2.2.3.2) was used as template in a PCR amplification to identify males. The primers used were designed from a medfly Y-specific repetitive sequence, Y114 (Zhou, et al., 2001). The reaction consisted of reaction buffer (1X), 1.5mM MgCl₂, 0.2mM each dNTP (Roche), 5 pmol of each primer, and 1U of Taq Polymerase (Applied Biosystems) in a final volume of 25µl. The forward and reverse primer sequences used were Y114F10: 5’-TGCCAAAGCACTATCTCTCGGAAG-3’ and Y114B13: 5’-GACGGTAACTGTCATTACGG-3’. The amplifications were performed using a Perkin-Elmer GeneAmp 2400 thermocycler and consisted of an initial denaturation at 94°C (one time) followed by 30 cycles of denaturation at 94°C for 30s, annealing at 58°C for 30s, and extension at 72°C for 2 mins. PCR products were visualized on 1.2% agarose gels stained with ethidium bromide. Amplification of template from each individual was conducted using both the Y114 primer pair and a primer pair designed from the medfly glucose-6-phosphate dehydrogenase (Zw) gene (He and Haymer, 1999) to ensure the presence of template in all samples. Samples that successfully amplified using the Y114 primer pair were designated males, while samples with no amplification product were designated females. This assay was used to infer the sex of the embryonic, larval, and pupal material collected and stored at -80°C.
2.2.6.3 RT-PCR analysis

First-strand cDNA was synthesized from DNase-treated total RNA using the RT-for-PCR Kit (Clontech). The reaction was performed using an Oligo(dT) primer according to the manufacturer’s protocol and was subsequently stored at -80°C. Second strand synthesis was carried out using gene specific oligonucleotide primers as listed in Appendix One. Genomic DNA was used as template alongside each experiment as a positive control, while separate reactions were included containing all of the reagents in the absence of template as a negative control. Primers designed from the Zw gene were used to amplify genomic DNA and all cDNA synthesized using this method as a control for the cDNA synthesis. The Zw primer pair is designed to amplify across an intron sequence (He and Haymer, 1999) and produces noticeably different fragment sizes when using genomic DNA as template (400 bp) and cDNA as template (200 bp).

2.2.6.4 5’ and 3’ RACE

The SMART™ RACE cDNA Amplification Kit (Clontech) was utilized to synthesize first-strand 5’ and 3’ RACE-Ready cDNAs. Poly(A) RNA isolated from adult females and males (two days post-eclosion) was used as template for separate 5’ RACE and 3’ RACE reactions. The SMART™ sequence is incorporated into both the 5’ and 3’ RACE reactions as part of the first-strand cDNA synthesis, which was performed as described in the manufacturer’s protocol. Second strand synthesis was carried out using 0.01 volume each of the 5’ or 3’ RACE Ready cDNA as template. A gene-specific primer (GSP) was used in combination with the Universal Primer Mix A (UPM) primer, specific to the SMART sequence, to direct amplification of the RACE cDNAs. An additional PCR was performed for those reactions in which a smear was produced upon amplification with the GSP/UPM primer pair. The additional PCR was nested and utilized 0.02 volume of GSP/UPM product as template for amplification with the GSP in combination with a Nested Universal Primer A.
(NUPM) primer. Sequences for the oligo primers utilized in this experiment can be found in Appendix One.

The Advantage 2 PCR Kit (Clontech) and the Expand Long Template PCR System (Roche) were both used to amplify the 5' and 3' RACE-Ready cDNAs following the manufacturer's protocol.

2.2.6.5 Long Distance PCR

Templates larger than 2.5 kb were amplified using the Expand Long Template PCR System (Roche). The reactions were conducted as outlined in the manufacturer's protocol. A negative control containing all of the reagents in the absence of DNA template was included alongside each experiment. All reactions were performed using the 10XPCR Buffer 3 that contained detergents and 22.5mM magnesium chloride. The cycling parameters used consisted of an initial denaturation at 94°C, 10 seconds followed by ten cycles of 94°C, 10 seconds; 62°C, 30 seconds; and 68°C for 4 mins. After, an additional twenty cycles of 94°C, 10 seconds; 62°C, 30 seconds; and 68°C, 4 mins with an additional twenty seconds per cycle ("AutoX" feature on GeneAmp 2400) were performed. A final extension of 7 mins at 68°C was carried out before the samples were stored at 4°C.

2.2.6.6 Degenerate Primed (DP) PCR

Degenerate primers were designed to amplify a portion of the BTB-POZ domain by focusing on conserved residues within the domain, as determined from an alignment of ten various BTB-POZ domain sequences (Chapter 5). The primers were designed to reflect codon usage in the medfly (He, and Haymer, 1992), and were intended to incorporate all relevant codons for the central sequence LSACSPY [5'-YTGTCGCTGYTCSCCHTAYTT-3', where Y is C or T; S is C or G; and H is A, C, or T] and for the C-terminal sequence DFMYTGE [5'-YTCRCCDGTRTACATRAARTC-3', where Y is C or T; R is A or G; and D is A, G, or T]. The primer pair was utilized to amplify female first-strand cDNA template from early adults (two days post-eclosion) and 3-5 hour embryos in a
25μl volume reaction containing 1X PCR Reaction Buffer (100mM Tris-HCl pH 8.3, 500mM KCl; Perkin Elmer), 1.5mM magnesium chloride (Perkin Elmer), 0.2mM each dNTP (Roche), 50 picomoles each of the forward and reverse degenerate oligonucleotide primer, and 2.5U of Taq Polymerase (Perkin Elmer). A negative control containing all of the reagents in the absence of DNA template was included alongside each experiment. The cycling parameters consisted of an initial denaturation at 94°C for 2 mins, followed by 35 cycles of 94°C, 1 minute; 53°C, 1 minute; and 72°C, 2 mins. The products were electrophoresed on a 2% agarose gel.

2.2.6.7 DIG-dUTP probe labeling

cDNA sequences CcPFS.1B and AORF, utilized as probes in Northern blot analyses, were labeled using the digoxigenen-11-dUTP (DIG-dUTP), alkali-stable base analog (Roche) in PCR amplification. A ratio of 4:1 (dTTP:DIG-dUTP) was used to successfully label templates ranging in size from 0.5kb to 1.5kb. Separate reactions containing the experimental template in the presence of dNTPs without the DIG-dUTP label (positive control) were included alongside each experiment, as were reactions containing all of the reagents in the absence of template (negative control). Reactions were carried out in a 25μl volume containing 1X PCR Reaction Buffer (100mM Tris-HCl pH 8.3, 500mM KCl, Perkin Elmer), 1.5mM magnesium chloride (Perkin Elmer), 0.2mM each of dCTP, dGTP, and dATP (Roche), 0.16mM dTTP, 0.04mM DIG-dUTP, 5 picomoles each of the forward and reverse oligonucleotide primer, and 1U of Taq Polymerase (Perkin Elmer). Typical cycling parameters consisted of an initial denaturation at 94°C for 2 mins followed by 35 cycles of 94°C for 1 minute, 55°C for 30 seconds, and 72°C for 2 mins, after which the amplification was stored at 4°C. Products were visualized on agarose to verify that the amplification product of a particular template using the DIG-dUTP was larger than the product amplified in the absence of DIG-dUTP.
2.2.7 Cloning and Transformation

2.2.7.1 Cloning products of SSH

Subsequent to electrophoresis of the 2°PCR amplification products, the DNA from the individual fragments produced was gel purified using the GeneClean Spin Kit-Isolation from Agarose Protocol (Bio101, Inc.). Each fragment represented a differentially expressed sequence from the respective subtraction experiment.

The purified fragments were then ligated into the pCR Script (SK+) vector using the pCR Script (SK+) Cloning Kit (Stratagene). The products were ligated according to the manufacturer’s protocol.

2.2.7.2 Shotgun subcloning of lambda phage DNA inserts into plasmids

Phage DNA fragments, produced by restriction enzyme digestion, were shotgun subcloned into the cloning site of the pUC19 plasmid vector. Shotgun subcloning relies on the random ligation of restriction fragments to the pUC19 plasmid vector, both of which are digested with the same restriction enzyme. Clones are then verified for the presence of a fragment of the desired size. All restriction digests were prepared for ligation by extraction with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) and precipitation of the phage DNA fragments by addition of 0.1 volume 3M sodium acetate, pH5.2 and 2 volumes of 95% ethanol. The precipitation was enhanced by incubation of the mixture at -20°C for 16-18 hrs. The phage DNA fragments were pelleted by centrifugation at room temperature for 15 mins at 10,000rpm, washed with 500µl of cold, 70% ethanol, and vacuum dried. The pellet was resuspended directly into the ligation mixture, which consisted of restriction enzyme digested vector DNA at a 1:1 ratio with the phage DNA fragments, 1X ligation buffer (Roche), and 1U of T4 DNA ligase (Roche). The ligation reaction was incubated at 4°C for 16-18 hrs.
2.2.7.3 Cloning products of PCR amplification

A majority of the PCR amplification products were cloned into the pCR Script (SK+) vector using the pCR Script (SK+) Cloning Kit (Stratagene). The products were ligated according to the manufacturer’s protocol. Alternately, products amplified using the Clontech Advantage 2 PCR Kit, the Expand System, and the degenerate primers were cloned into the pCR2.1 vector using the T/A Cloning System (Invitrogen) according to the manufacturer’s protocol.

2.2.7.4 Bacterial transformation

All products of ligation were introduced into *Escherichia coli* bacterial cells using either a “heat shock” method or electroporation. Heat shock transformation of DH-5α (Invitrogen) cells was performed by incubating half of the ligation reaction, irrespective of volume, with 100μl of cells on ice (−0°C) for 30 mins. The temperature of the sample was then increased significantly to 42°C for 30 seconds to induce the transformation. Electroporation of 1μl of a ligation reaction (either 10μl or 20μl volume) into 20μl of DH10B electroporator cells was performed by briefly incubating the ligation with the bacterial cells prior to delivering a voltage of 2.45kJ using a Cell-Porator—*E. coli* Pulser™ (GIBCO BRL). In both cases, the transformed bacterial cells were then incubated in 1ml of TB broth (growth media containing peptone and yeast extract) (Terrific Broth, Bio101, Inc.) at 37°C with vigorous shaking for 1 hour. 200μl and a concentrated volume of the transformed cells were plated onto 1.5% LB-agar supplemented with 0.04mg/ml X-gal (Sigma) and 0.1mg/ml ampicillin (Sigma) and incubated, inverted, at 37°C for 16-18 hrs.

Ligation products from the T/A cloning kit were transformed into TOP10 *E. coli* bacterial cells (Invitrogen) using the “heat shock” method. The transformed cells were then incubated in 250μl of S.O.C. growth media [peptone, yeast extract, glucose, and Mg⁺⁺ ions] (provided by manufacturer) and incubated at 37°C for 1 hour with vigorous shaking, after which 50μl and 150μl of the transformed
cells were plated onto 1.5% LB-agar supplemented with 0.04mg/ml X-gal and 0.1mg/ml ampicillin and incubated, inverted, at 37°C for 16-18 hrs.

White colonies, representing recombinant plasmids, were then re-streaked on a fresh petri dish of the same media to obtain isolated colonies.

2.2.7.5 Plasmid DNA isolation

2mls of TB growth media (Terrific Broth, Bio101 Inc.) supplemented with 100μg/ml ampicillin were inoculated with a single, white bacterial colony (from above) and shaken vigorously at 37°C for 16-18 hrs. Plasmid DNA was isolated from the bacterial culture using the QIAprep® Spin Miniprep Kit (QIAGEN) according to the manufacturer’s protocol. The plasmid DNA was eluted with 35μl of ddH₂O.

2.2.8 Nucleic Acid Blot Analysis

2.2.8.1 Non-Radiactive DNA labeling and detection

Random-primed, digoxigenin (DIG) labeled probes were synthesized using the DIG High Prime Labeling and Detection Kit (Roche). Alternatively, cDNA probes CcPFS.1B and AORF used for Northern blot analysis were synthesized using the DIG-dUTP nucleotide analog in a PCR labeling reaction, as described in Section 2.2.6.7. These probes were used for genomic library screening and for Southern and Northern blot analysis. The templates used for synthesis of DIG-labeled probes (PCR products, various restriction enzyme fragments) were purified using the Geneclean Spin Kit (Bio101, Inc.) prior to being labeled. Approximately 500ng-1μg of template in a 16μl volume were denatured at 95°C for 10 mins, quelled on ice, and then added to 4μl of DIG High Prime Mix (Roche). The reactions were incubated at 37°C for 16-18 hrs, regardless of the length of the probe. Prior to their use in experiments, the probes were hybridized with “dot blots” on which the same
template used to make the probe was bound. In addition, probe quantification assays were performed by comparison of the labeled experimental DNA with labeled control DNA, as described by the manufacturer.

Colorimetric detection of the probe molecules was performed as described in the manufacturer’s protocol for the DIG High Prime Labeling and Detection Kit. All the steps were performed with gentle agitation using a Red Rotor shaker (Hoefer Scientific Instruments). Subsequent to stringency washes designed to remove loosely bound probe molecules, the nucleic-acid bound membranes were briefly incubated in wash buffer (0.1M Tris pH 7.5, 0.15M sodium chloride) and then incubated in a specially formulated 1X blocking solution (10X blocking solution diluted in washing buffer; Roche) for 30 mins at room temperature. The solution was removed, replaced with DIG-alkaline phosphatase antibody conjugate (Roche) diluted 1:10,000 in 1X blocking solution, and incubated for 30 mins at room temperature. The membranes were then washed twice in washing buffer for 15 mins each, after which the membranes were equilibrated in detection buffer (1M Tris pH 9.5, 0.1M sodium chloride, and 0.05M magnesium chloride) for five mins. Colorimetric detection of the alkaline phosphatase conjugated DIG antibodies was performed by the addition of 10mls of NBT/BCIP staining solution [0.4mg/ml nitro blue tetrazolium chloride (NBT), 0.19mg/ml 5-bromo-4-chloro-3-indoyl-phosphate (BCIP), 0.1M Tris pH 9.5, and 0.05M magnesium sulfate]. The color reaction proceeded in a dark, flat space and was stopped by incubating the membranes in TE (0.01M Tris pH 8.0 and 0.001M EDTA) after the desired result was obtained.

For Northern analysis, detection of the probe molecules was conducted exactly as above using the DIG Wash and Block Buffer Set (Roche). The 10X solutions were diluted to 1X using DEPC-treated ddH2O. Chemiluminescent detection of the alkaline phosphatase conjugated DIG antibodies was conducted using CSPD, ready-to-use (Roche). After incubation in detection buffer, the membrane was placed on one leaf of a film-developing folder near the fold (GIBCO BRL). 1ml of CSPD was applied and the membrane was immediately covered with the second leaf of the folder in order to
spread the CSPD evenly across the membrane. The folder was incubated for 5 mins at room
temperature, after which the CSPD was squeezed out and the edges of the folder taped. The folder
was incubated at 37°C for 15 mins and exposed to an 8.5" X 11" X-ray film (Fuji) in a cassette. The
cassette was left in the dark at 25°C for variable lengths of time up to 36 hrs, after which the
luminescent signal diminishes.

2.2.8.2 Southern blot analysis

Southern blots were prepared essentially as described in Molecular Cloning: A Laboratory
Manual (Sambrook, et al., 1989). Medfly genomic DNA digested with various restriction enzymes
was electrophoresed on 0.8% agarose for 18 hrs at 32V. Southern blots of lambda phage DNA were
electrophoresed on 0.8% agarose for 2.5 hrs. Appropriate molecular weight markers were run
alongside the samples, and gels were documented using a hand-held Polaroid camera.

Agarose gels were denatured in 1.5M sodium chloride/0.5M sodium hydroxide for 1 hour,
neutralized in 1.5M sodium chloride/1M Tris pH 8.0 for 30 mins, and incubated in 20XSSC (transfer
buffer) for 30 mins. The DNA was transferred to a nylon membrane (Osmonics) of the appropriate
size using the Turboblotter Rapid Downward Transfer System (Schleicher & Schuell). Transfer of
DNA was initiated using 20XSSC transfer buffer and was allowed to proceed for a minimum of 3 hrs
(lambda DNA) or 16-18 hrs (genomic DNA). Following the transfer, the membranes were briefly
rinsed in 2XSSC and exposed to 120,000µJ UV using a UV Stratalinker® 1800 (Stratagene).
Membranes were pre-hybridized for 1-16 hrs at 42°C in a solution containing 2% blocking reagent
(Roche), 0.1% SDS, 5XSSC, 0.04% N-lauroylsarcosine, and 50% deionized formamide; the labeled
probe was resuspended in the same formula. Hybridization of the membranes with the labeled probe
was performed at 42°C for 16-18 hrs.

After hybridization, membranes were washed twice in 2XSSC and 0.1% SDS for 5 mins each at
room temperature. A second wash was performed, and repeated, at 0.5XSSC and 0.1% SDS for 15
mins. The second washes were performed at several different temperatures, depending on the
stringency required. For blots containing lambda DNA, all second washes were performed at 65°C.
For blots containing genomic DNA, various temperatures were utilized ranging from 37°C and 42°C
(low stringency) to 55°C, 60°C, and 65°C (high stringency).

2.2.8.3 Northern blot analysis

Northern blots were performed using poly(A) RNA isolated from staged female and male adults.
For each lane of the blot, poly(A) RNA from four individual flies was loaded. Preliminary blots were
performed using poly(A) RNA isolated from various amounts (100, 300, 600, 900) of unsexed 3-5
hour embryos, the entire volume of which was loaded into each lane. Labeled, cloned cDNA
products were used as probes (Section 2.2.6.7). All hardware and instruments utilized, i.e. gel-casting
trays, electrophoresis chambers, glass dishes, micro-pipets, were treated with RNase-Zap (Ambion)
and rinsed with DEPC-treated ddH₂O. All solutions were made using DEPC-treated ddH₂O.

A 1% denaturing agarose gel was prepared by dissolving 1.25g agarose in 12.5mls 10X MOPS
running buffer (10X MOPS [Research Organics], pH 7.0: 0.2M MOPS, 10mM EDTA, 50mM sodium
acetate) and 102.5mls DEPC-treated ddH₂O. The solution was cooled to 50°C prior to the addition of
6mls 37% formaldehyde (Sigma), and then poured into a gel-casting tray. Poly(A) RNA pellets were
resuspended in 25μl of Sample Buffer (2.5μl 10X MOPS, 12.5μl 100% formamide, 5μl 37%
formaldehyde, 0.125μl ethidium bromide [10mg/ml], and 5μl DEPC-treated ddH₂O) and incubated at
60°C for 10 mins for denaturation. 2μg of RNA molecular weight marker II (Roche; 5 fragments:
6.95kb, 4.74kb, 2.66kb, 1.82kb, and 1.52kb) was added to 14μl of sample buffer and treated just as
the experimental samples. The samples were quickly centrifuged to collect condensate and then
briefly quenched on ice. Prior to loading, 6μl of RNA loading dye (0.25% bromophenol blue, 25%
sucrose) was added to each tube. Splitting the ~30μl sample into two separate wells of the agarose

40
gel created identical blots that could be divided and probed independently. The RNA was electrophoresed in 1XMOPS running buffer at 110V until the bromophenol blue migrated the entire length of the gel. During electrophoresis, the running buffer was circulated using a Masterflex® pump controller (Cole Parmer).

Subsequent to electrophoresis, the gel was washed DEPC-treated water twice for 10 mins followed by twice for 5 mins each to remove as much formaldehyde as possible. The gel was then incubated in 10XSSC for 45 mins prior to transfer. The RNA was transferred to a nylon membrane (Osmonics) of the appropriate size using the Turboblotter Rapid Downward Transfer System (Schleicher & Schuell). Transfer of RNA was initiated using 10XSSC transfer buffer and was allowed to proceed for 16-20 hrs. Following transfer, the membranes were briefly rinsed in 2XSSC and exposed to 120,000μJ UV using a UV Stratalinker® 1800 (Stratagene). Membranes were pre-hybridized for 1-5 hrs at 42°C in DIG Easy Hyb (Roche); the labeled probe was resuspended in the same formula. The prehybridization solution was discarded and fresh hybridization solution containing the labeled probe was added to the membranes. For first use, the labeled cDNA was denatured at 95°-100°C for 10 mins, added to 10mls of DIG Easy Hyb, and then added to the membrane. For subsequent use of the probe resuspended in DIG Easy Hyb, the probe was denatured at 68°C for 10 mins. The hybridization was performed at 42°C for 16-18 hrs. After hybridization, membranes were washed twice in 2XSSC and 0.1%SDS for 5 mins each at room temperature. A second wash was performed, and repeated, in 0.5XSSC and 0.1% SDS for 15 mins at 42°C.

2.2.9 DNA sequencing and analysis

Cloned products of interest were sequenced at the Biotechnology/Molecular Biology Instrumentation and Training Facility (BMBITF) on the campus of the University of Hawai‘i using an ABI Prism sequencer. DNA and protein sequence similarity searches of the GenBank database were conducted using the BLASTN 2.2.3 and BLASTX 2.2.3 algorithms (Altschul, et al., 1997) with the
The BLASTN algorithm queries the nucleotide sequence of interest against the Genbank database.
The BLASTX algorithm converts a nucleotide query sequence into protein sequences in all six reading frames and then compares these protein sequences to the Genbank database. A cut-off value of $E < 0.001$ was used as a criteria to judge the significance of the sequence similarities reported by BLAST. At the time the sequence similarity searches were conducted, there were 1,275,557 nucleotide and 980,100 protein sequences available in the database.

The DNA Star computer package was used extensively for editing of sequences (EditSeq), restriction enzyme mapping (Map Draw). The MacVector Program (Macintosh) was used for designing all oligonucleotide primers used in PCR amplifications. Nucleotide and amino acid sequence alignments were generated using ClustalW (Thompson, et al., 1994), a program offered at the European Bioinformatics Institute (http://www.ebi.ac.uk/clustalw). This program was also used to construct a Neighbor-Joining Tree with 1000 bootstrap replications. TreeDraw was used to visualize the phylogenetic tree using tree files generated by ClustalW.
Chapter 3

Application of Suppression Subtractive Hybridization (SSH) to Developmentally Staged Material from the Genome of the Medfly

3.1 Background

Many methods based on the principle of subtractive hybridization have been developed to identify differentially expressed genes that play a role in various biological processes. These methods have become so widely used that a recent search of the PubMed database returned ~800 different papers based on its use, a majority of them published within the past five years. The technique of subtractive hybridization that involves the comparison of two populations of material and the subsequent isolation of sequences that are present/expressed in one population ("tester"), but not in the other ("driver"). Several PCR-based approaches that incorporate subtractive hybridization include, but are not limited to, representational difference analysis (RDA) and suppression subtractive hybridization (SSH).

RDA is a technique developed to identify differences between two complex genomes (Lisitsyn, et al., 1993) using a genome subtractive process (Lisitsyn, 1995). In RDA, subtractive hybridization is coupled with PCR to enrich for sequences present in the tester genome but absent from the driver. A variation of RDA, suppression subtractive hybridization (SSH) was designed to identify cDNA sequences that are differentially expressed in one genome versus another (Diatchenko, et al., 1996; Diatchenko, et al., 1999). SSH, then, is sensitive to different levels at which a particular gene may be expressed rather than simple presence or absence. Some advantages of the SSH method, relative to other subtractive hybridization approaches, are: i.) the equalization of overabundant sequences in the tester and similar sequences in the tester and driver populations and ii.) the potential for 1000-fold enrichment of differentially expressed sequences (Diatchenko, et al., 1999). It was these advantages of SSH that were intriguing and prompted the used of the method to identify genes that are differentially expressed during medfly embryogenesis.
The crux of SSH is the use of suppression PCR to selectively suppress amplification during PCR (Siebert, et al., 1995). The suppression is achieved by the use of complementary sequences attached at each end of a single-stranded (ss) cDNA. This results in the formation of a pan-like secondary structure, which prevents any oligonucleotide primer from properly annealing to its template (Siebert, et al., 1995).

To carry out the SSH method, poly(A+) RNA is first isolated from the two samples being compared. Double-stranded (ds) tester and driver cDNA is synthesized from the poly(A+) RNA template. These templates are restriction-enzyme digested using a 4-cutter restriction enzyme that creates blunt-ended fragments. The tester-cDNA is then separated into two fractions to which separate adapters—Adapter 1 (Ad1) and Adapter 2R (Ad2R)—are ligated at the 5' end. A portion of the Ad1- and Ad2R-tester cDNAs are mixed, denatured, and allowed to hybridize, separately, with excess amounts of driver-cDNA to produce a mixture of hybridization products (Figure 3.1). The less-abundant, differentially expressed sequences remain single-stranded while the more abundant cDNA molecules anneal more quickly during the hybridization to produce homo-hybrids (representing the sequences that are in abundance within the tester-cDNA pool) and hetero-hybrids (representing the sequences that are common between the tester and driver-cDNA pools). This results in the normalization of the tester-cDNA sample.

The Ad1- and Ad2R- hybridization mixtures are combined and hybridized with freshly denatured driver-cDNA to promote enrichment for differentially expressed sequences. Tester-cDNAs that are still ss have the potential to form additional homo-hybrids, but also have the potential to form new ds-hybrids consisting of ss-Ad1-tester cDNA and ss-Ad2R-tester cDNA (Figure 3.1). Sequences that are differentially expressed are then isolated from this mixture after two rounds of PCR. In the first round (1st) of PCR, a primer that is complementary to a segment of Ad1 and Ad2R is used to direct amplification. Those molecules that lack a binding site for this primer are not exponentially amplified (type a and d), while those molecules with only one primer binding site present are
Figure 3.1 Principle of Suppression Subtractive Hybridization (SSH). A flow chart representing the process by which different adapters are ligated to the ends of ds-cDNA and are used as a template to amplify sequences that are differentially expressed between the tester and driver genomes. After the first hybridization is performed, different pools of cDNAs are formed. Pool "a" represents those cDNAs that are less abundant within the tester genome and remain single-stranded because they are unique to the tester. Pool "b" corresponds to those cDNAs that are more abundant in the tester genome and that are further normalized. Those cDNAs that are common between the tester and driver genomes are represented by pool "c", while pool "d" represents those sequences in the driver genome that are less abundant (single-stranded) and more abundant (double-stranded). These pools are then combined for a second hybridization resulting in the production of pool "e", which represents those molecules that are unique to the tester genome and that have two different adapter sequences at their ends. Subsequent PCR amplification using oligonucleotide primers that anneal to the different adapter sequences results in the isolation of the subtraction products represented by pool "e".
amplified linearly (type e). Suppression PCR is utilized to suppress amplification of molecules that have only one adaptor at both of its ends (type b—either Ad1 or Ad2R) since these molecules form pan-like secondary structures. Only those molecules having both Ad1 and Ad2R at its ends (type e) can be successfully amplified in this first round. A nested-PCR approach is utilized in the second round of PCR (2°) with primers that are specific to Ad1 and Ad2R in order to reduce background PCR products and to enrich for the differentially expressed sequences represented by the type “e” hybrids.

3.1.1 Use of the SSH method in the medfly

SSH has been used previously in the medfly as a tool to remove female-specific sequences from a library of microdissected Y-chromosome sequences (IAEA, 1999), but has not been used as a tool to analyze differential gene expression in the species. In order to validate the use of the SSH method for this purpose, initial experiments were performed using medfly adult material. Since sex-specifically expressed sequences have been identified in medfly adults (Tolias, et al., 1990; Rina and Savakis, 1991; Rosetto, et al., 1999; Rosetto, et al., 2000; Christophides, et al., 2000a), the isolation of any of these sequences using SSH would provide validation for use of the method.

3.1.2 Molecular Sexing Assay

Although adult medfly individuals have secondary sexual characteristics that can be used to distinguish females from males, medfly embryos are morphologically indistinguishable. In order to determine the sex of these individuals, and of those from the larval stage, a PCR-based assay was utilized. This assay is designed to amplify sequences, designated Y114, originating from the Y-chromosome of the medfly (Zhou, et al., 2000).

In the course of this project it was demonstrated that Y114 PCR amplification, used previously to sex adult stage material, could be extended to material isolated from various embryonic, larval, and
pupal stages of development. Use of this assay was crucial to the collection of sex-specific RNAs from the embryonic and larval stages that were used as templates for the SSH experiments.

3.1.3 Chapter Goal

This chapter describes the results of SSH experiments performed using sexed material collected from 1.) medfly adults (1 day post-eclosion), 2.) third instar larvae, and 3.) developmentally staged embryos obtained 3-5, 4-7, and 5-18 hours after egg-laying (AEL). A total of 25 subtraction products were obtained across all the stages examined. Sequence analysis revealed that half of the products were similar to cytoplasmic ribosomal proteins and mitochondrial ribosomal RNA. The remainder of the products include medfly homologs of previously identified genes and potentially novel genes. The isolation of each of the sequences is discussed with regard to its potential for differential expression or its observed differential expression, and recommendations for improving the use of the SSH method in the medfly are provided.

3.2 Results

3.2.1 Method for sexing material from pre-adult stages

Primers designed from a region of the Y-chromosome specific repetitive sequence, YI14, were used in an assay for sexing individual, developmentally staged medfly embryos and third instar larvae (Figure 3.2). The ability to amplify Y-chromosome material was used as a basis for identifying the sex of an individual as a male. First-strand cDNAs collected from sex-specific pools of RNA for later use in RT-PCR were amplified using the YI14 primer pair in order to verify cDNA pools that were synthesized from male material. Amplification was observed to occur only using cDNAs synthesized from male samples (Figure 3.2).
Figure 3.2 Y114 PCR Amplification. (A) Genomic DNA template from individual early embryos, late embryos, 3\textsuperscript{rd}-instar larvae, and adults were used in PCR amplifications with the Y114 oligo primer pair. The same material was amplified using the Zw primer pair to ensure the presence of template in the Y114 amplification. The sex of each individual was identified based on the presence or absence of amplification, with amplification expected only when male material was used as template. (B) First-strand cDNA template from male early embryos (E), third-instar larvae (L), early pupae (P), and adults 2 days (A2) and 10 days (A10) post-eclosion was amplified using the Y114 primer pair. No amplification was observed using female cDNA as template.
3.2.2 Suppression Subtractive Hybridization

Individual subtraction products identified as fragments on agarose gels were subsequently cloned into the pCR Script (SK+) vector and sequenced. The products visualized on the gel contain adapter sequences (Ad1 and Ad2R), which add an additional 44 bases to the subtraction products. Sequence lengths reported here refer to the subtraction product without the adapter primer sequences.

The nomenclature used for all of the sequences is based on the premise that female-driven experiments will produce putative male-specific sequences while male-driven experiments will produce putative female-specific sequences. Hence, subtractive hybridization clones obtained using a female-driver are named PMS (putative male specific) and clones obtained using a male-driver are named PFS (putative female specific).

Although both the BLASTN and BLASTX algorithms were used to analyze subtraction product sequences, the identities of the subtraction clones reported below were assigned based on sequence similarity results using the BLASTX algorithm, unless otherwise noted. It is common practice to use the E-values reported in BLAST search results as a measure of statistical significance (Pertsemlidis and Fondon, 2001). An E-value represents the number of hits in the database that would expect (E) to be found by chance if there were no true matches in the database (Gibson and Muse, 2002). An E-value of 0.05 or less is generally considered to reflect a biologically significant alignment since it implies that the given alignment is not likely to occur by chance more than 5% of the time when there is no true match in the database. E-values near 1e-30 offer a high level of confidence that the sequence used for the search is evolutionarily related to the matched sequence in the database (Brenner, 1998). I have chosen an E-value cut-off of E< 0.001 to increase the likelihood that the identities assigned to each of the subtraction products is not erroneous. Subtraction sequences that aligned with the same gene from various organisms were tentatively identified as medfly homologs of that gene. Relevant sequence alignments for each of the subtraction products are included within the chapter as figures, and the assigned identities for each of the clones are summarized in Table 3.1.
Table 3.1 Cumulative results of suppression subtractive hybridization conducted using adult, larval, and embryonic material. (a) The subtraction sequences listed were isolated from a hybridization using female material as the “driver” and (b) male material as the “driver”. The different developmental stages analyzed are identified in bold italic. Material was utilized from three different stages of embryogenesis: 3-5 hour, 4-7 hour, and 5-18 hour AEL. The identity of each of the clones was assigned based on similarity to sequences in the Genbank database.

(a) Clone Name | Sequence Length (bp) | Identity (E-value)
--- | --- | ---
**Adult Stage**
APMS.1 | 128 | *D. melanogaster* RE56733 gene product NADH-ubiquinone oxidoreductase MLRQ subunit
APMS.2 | 115 | *C. capitata* male-specific serum polypeptide

**3rd Instar Larva Stage**
LPMS.1 | 131 | No significant similarity
LPMS.2 | 68 | Discarded (Identity uncertain)

**Embryonic Stage (3-5 hour AEL)**
3-5PMS.1 | 394 | Putative Novel Sequence
3-5PMS.2 | 168 | Histone H2B
3-5PMS.3 | 141 | Putative Novel Sequence
3-5PMS.4 | 108 | *D. melanogaster* CG1475 gene product 60S ribosomal protein L13a
3-5PMS.5 | 228 | *D. melanogaster* CG2099 gene product 60S ribosomal protein L35a

**Embryonic Stage (4-7 hour AEL)**
4-7PMS.1 | 368 | *C. capitata* 16S ribosomal RNA gene
4-7PMS.2 | 313 | *C. capitata* 16S ribosomal RNA gene
4-7PMS.4 | 188 | Putative Novel Sequence

**Embryonic Stage (5-18 hour AEL)**
5-18PMS.1 | 309 | 60S ribosomal protein L10 [QMI protein]
5-18PMS.2 | 287 | *C. capitata* 16S ribosomal RNA gene
5-18PMS.3 | 251 | 20S proteasome beta 2 subunit
5-18PMS.4 | 157 | 40S ribosomal protein S20
5-18PMS.5 | 108 | *D. melanogaster* CG1475 gene product 60S ribosomal protein L13a
5-18PMS.6 | 63 | Translation Elongation Factor eEF1-alpha
### Table 3.1 (continued)

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Sequence Length (bp)</th>
<th>Identity (E-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Embryonic Stage (3-5 hour AEL)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-5PFS.1</td>
<td>332</td>
<td>Putative Novel Sequence</td>
</tr>
<tr>
<td>3-5PFS.2</td>
<td>299</td>
<td>Putative Novel Sequence</td>
</tr>
<tr>
<td>3-5PFS.3</td>
<td>211</td>
<td>Nucleoside Diphosphate Kinase/abnormal wing disc protein</td>
</tr>
<tr>
<td>3-5PFS.4</td>
<td>187</td>
<td>Putative Novel Sequence</td>
</tr>
<tr>
<td><strong>Embryonic Stage (4-7 hour AEL)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-7PFS.1</td>
<td>570</td>
<td><em>C. capitata</em> 16S ribosomal RNA gene</td>
</tr>
<tr>
<td>4-7PFS.4</td>
<td>243</td>
<td>60S ribosomal protein L18a</td>
</tr>
<tr>
<td><strong>Embryonic Stage (5-18 hour AEL)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-18PFS.1</td>
<td>185</td>
<td><em>C. capitata</em> 16S ribosomal RNA gene</td>
</tr>
</tbody>
</table>
3.2.2.1 Subtractive hybridization results using adult material

Only female-driven subtractive hybridization was performed on material obtained from adults 1 day post-eclosion because of an interest in the identification of male-specific sequences. The 2° PCR amplification produced four main products, ranging in size from 150bp to 210bp, represented as fragments A.1, A.2, A.3, and A.4 (Figure 3.3a). Although several attempts were made to clone fragments A.1 and A.2, only fragments A.3 and A.4 were successfully cloned. These were renamed APMS.1 and APMS.2, respectively. Complete nucleotide sequences were determined for each of these.

APMS.1 is 127bp in length and displays sequence similarity to the RE56733 gene product of Drosophila melanogaster (E=5e-17) and to the MLRQ subunit of the NADH:ubiquinone oxidoreductase complex in Mus musculus (E=8e-07), Bos taurus (E=1e-06), and Homo sapiens (E=7e-05). APMS.2 is 115bp in length and displays sequence similarity to the male specific serum protein (MSSP) family (α1, α2, β1, β2, and β3) from Ceratitis capitata (E=3e-12). Alignment of these sequences is shown in Figure 3.4 and this information is summarized in Table 3.2.

3.2.2.2 Subtractive hybridization using third instar larval material

As in the adult stage, only female-driven subtractive hybridization was performed using material obtained from the third instar larval stage because of an interest in the identification of male-specific sequences. Two fragments, L.1 and L.2 (Figure 3.4b), resulted from this subtractive hybridization; these were later cloned and designated as LPMS.1 and LPMS.2, respectively. The DNA sequence of these inserts was determined.

LPMS.1 is 123bp in length and displays sequence similarity with the CG17288 gene product from D. melanogaster (E= 1.9), but the alignment is not statistically significant. LPMS.2 is 68bp in length and does not display significant similarity to any sequences in the database.
Figure 3.3 Subtractive hybridization products obtained using material from the adult and third instar larval stages. Subtraction products produced after 1° and 2° PCR amplification were electrophoresed on agarose and stained with ethidium bromide. The products depicted are 42bp larger than the reported size of the sequences due to the presence of the adapter primer sequence. (A) Adult subtraction products are labeled A.1, A.2, A.3, and A.4 (very light intensity). Fragments A.3 and A.4 were subsequently cloned and renamed APMS.1 and APMS.2, respectively. (B) Larval subtraction products are labeled L.1 and L.2, both of which were cloned and renamed LPMS.1 and LPMS.2, respectively.
Figure 3.4 Amino acid sequence alignment of adult subtraction product sequences, APMS.1 and APMS.2, with respective homologous sequences. (A) APMS.1 amino acid sequence alignment with the *D. melanogaster* RE56733 gene product and the MLRQ subunit of the NADH: ubiquinone oxidoreductase complex in *M. musculus*, *B. taurus*, and *H. sapiens*. (B) APMS.2 amino acid sequence alignment with the β1, β2, α1, and α2 members of the *C. capitata* male-specific serum protein family. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (.), while those in which a semi-conserved substitution is found are identified by a single dot (.)
Table 3.2 Results of subtractive hybridization using adult and third instar larval material. The subtraction sequences listed were isolated from a hybridization using adult and larval female material as the “driver”. The identity of each of the clones is included along with the size of each of the corresponding clones. The identity of each was assigned using the BLASTX algorithm at the NCBI and E-values are shown to identify the significance of the identity reported.

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Sequence Length (bp)</th>
<th>Identity (E-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APMS.1</td>
<td>128</td>
<td>D. melanogaster RE56733 gene product (5e-17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NADH-ubiquinone oxidoreductase MLRQ subunit (8e-07)</td>
</tr>
<tr>
<td>APMS.2</td>
<td>115</td>
<td>C. capitata male-specific serum polypeptide (2e-11)</td>
</tr>
<tr>
<td><strong>3rd In-Star Larva Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPMS.1</td>
<td>131</td>
<td>No significant similarity</td>
</tr>
<tr>
<td>LPMS.2</td>
<td>68</td>
<td>Discarded (Identity uncertain)</td>
</tr>
</tbody>
</table>
obtained for LPMS.2, though, was not ideal and further attempts to sequence the clone were unsuccessful. As a result, the clone was discarded. This information is summarized in Table 3.2.

### 3.2.2.3 Subtractive hybridization results using staged embryonic material

Subtractive hybridization was performed using material from three different stages of embryogenesis—3-5 hour, 4-7 hour, and 5-18 hour after egg laying (AEL). In *Drosophila*, genes that are involved in the pathway of sex determination are expressed prior to cellularization of the blastoderm. This stage is ideal for the examination of genes that are differentially expressed in females versus males. Medfly embryos at the 3-5 hour and 4-7 hour AEL stages represent those in the pre-cellularization stage of embryogenesis. Medfly embryos at the 5-18 hour AEL stage were selected to examine any genes that may be differentially expressed later during embryogenesis.

Separate experiments were performed at each stage alternating the female and male material as driver. The multitude of products that were isolated resulted from the use of non-diluted as well as a $10^1$ dilution of the 1°PCR product as template for the 2°PCR amplification. Overall, there were a total of 14 PMS products and 7 PFS products isolated from the embryonic stages. In order to identify the stage at which each product was isolated, the name of the clone is preceded by 3-5, 4-7, or 5-18 to indicate the appropriate embryonic time frame.

#### 3.2.2.3.1 Mitochondrial Ribosomal RNA

Approximately 25% (5/21) of the subtraction products identified exhibited sequence similarity with the medfly mitochondrial genome. These clones were not characterized further. Clones 4-7PMS.1 (368bp), 4-7PMS.2 (313bp), 5-18PMS.2 (287bp), 4-7PFS.1 (570bp), and 5-18PFS.1 (185bp) all display significant sequence similarity with the 16S ribosomal RNA gene using the BLASTN algorithm (E-values ranged from e-129 to e-137). These results are summarized in Table 3.3.
Table 3.3 Subtraction products representing mitochondrial ribosomal RNA sequences isolated using subtractive hybridization. The subtraction sequences listed were isolated from a hybridization using embryonic material from the 4-7 hour and 5-18 hour AEL stages of embryogenesis. Female and male material was alternated as “driver”. The identity of each of the clones is included along with the size of each of the corresponding clones. The identity of each was assigned using the BLASTN algorithm at the NCBI and E-values are included to identify the significance of the identity reported.

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Sequence Length (bp)</th>
<th>Identity (E-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-7PMS.1</td>
<td>368</td>
<td>C. capitata 16S ribosomal RNA gene <em>(5e-90)</em></td>
</tr>
<tr>
<td>4-7PMS.2</td>
<td>313</td>
<td>C. capitata 16S ribosomal RNA gene <em>(2e-83)</em></td>
</tr>
<tr>
<td>5-18PMS.2</td>
<td>287</td>
<td>C. capitata 16S ribosomal RNA gene <em>(3e-97)</em></td>
</tr>
<tr>
<td>4-7PFS.1</td>
<td>570</td>
<td>C. capitata 16S ribosomal RNA gene <em>(1e-127)</em></td>
</tr>
<tr>
<td>5-18PFS.1</td>
<td>185</td>
<td>C. capitata 16S ribosomal RNA gene <em>(2e-87)</em></td>
</tr>
</tbody>
</table>
3.2.2.3.2 Cytosolic Ribosomal Proteins

Six of the twenty-one embryonic subtractive hybridization products displayed sequence similarity with ribosomal proteins that comprise the large (60S) and small (40S) subunits of the cytosolic ribosome.

Five products, four from a female-driven and one from a male-driven subtractive hybridization, displayed similarity to ribosomal proteins residing in the 60S subunit. Homologs of 60S ribosomal proteins identified include L13a (3-5PMS.4 and 5-18PMS.5), L35a (3-5PMS.5), L10 (5-18PMS.1) and L18a (4-7PFS.4). One product from a female-driven subtraction (5-18PMS.4) displayed similarity with the 40S subunit riboprotein, S20. Alignments of each of the subtraction product sequences with the corresponding ribosomal protein sequences are shown in Figures 3.5, 3.6, 3.7, 3.8, and 3.9.

The sequences for L13a, L35a, and L10 were submitted to Genbank as medfly ribosomal protein homologs. This information is summarized in Table 3.4 along with corresponding E-values and Genbank accession identification numbers.

3.2.2.3.3 Medfly homologs of other previously identified genes

Four of the twenty-one subtraction products displayed sequence similarity to other types of sequences. Sequence similarity was identified to histone-H2B (3-5PMS.2), the 20S proteasome beta-2 subunit (5-18PMS.3), translation elongation factor 1-alpha (5-18PMS.6), and nucleoside diphosphate kinase (3-5PFS.3). Alignments of each of the subtraction product sequences with the corresponding sequences to which similarity was displayed are shown in Figures 3.10, 3.11, 3.12, and 3.13. As with the ribosomal proteins, these sequences were submitted to Genbank as medfly homologs of the respective genes. This information is summarized in Table 3.4 along with corresponding E-values and Genbank accession identification numbers.
Figure 3.5 Amino acid sequence alignment of embryonic subtraction products 3-5PMS.4 and 5-18PMS.5 with representative 60S ribosomal protein L13a sequences. The subtraction products are aligned with the *D. melanogaster* CG1475 gene product and L13a riboprotein sequences from *S. trutta*, *H. sapiens*, and *I. punctatus*. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (:), while those in which a semi-conserved substitution is found are identified by a single dot (.).
Figure 3.6  Amino acid sequence alignment of embryonic subtraction product 3-5PMS.5 with representative 60S ribosomal protein L35a sequences. The subtraction product is aligned with the *D. melanogaster* CG2099 gene product and L35a riboprotein sequences from *M. musculus*, *H. sapiens*, and *S. frugiperda*. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (.), while those in which a semi-conserved substitution is found are identified by a single dot (.)
Figure 3.7 Amino acid sequence alignment of embryonic subtraction product 5-18PMS.l with representative 60S ribosomal protein L10 (QM) sequences. The subtraction product is aligned with L10 riboprotein (QM) sequences from *H. virensens*, *B. mori*, *D. melanogaster*, and *G. gallus*. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (:), while those in which a semi-conserved substitution is found are identified by a single dot (.)
### Figure 3.8 Amino acid sequence alignment of embryonic subtraction product 4-7PFS.4 with representative 60S ribosomal protein L18a sequences.

The subtraction product is aligned with L18a riboprotein sequences from *D. melanogaster*, *S. frugiperda*, *H. sapiens*, and *M. musculus*. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (.), while those in which a semi-conserved substitution is found are identified by a single dot (.).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-7PFS.4</td>
<td>----------</td>
</tr>
<tr>
<td>Dmelanogaster_L18a</td>
<td>MRKGI.LHEYVGRKLFSEKEEFTFLYKMRIRAPDNIVAKSRFWYFLRLOLKKHKKTTGE 60</td>
</tr>
<tr>
<td>Sfrugiperda_L18a</td>
<td>MKKGI.KHEYVGRKLFSEKEEFKFLYKMRIRAPDNIVAKSRFWYFLRLOLKKHKKTTGE 60</td>
</tr>
<tr>
<td>Hsapiens_L18a</td>
<td>MKASGLRLEKVRKCLFTEKCTFHMYRMRFAPNHVYAKSRFWYFVSQLHMKKSSSE 60</td>
</tr>
<tr>
<td>Mmusculus_L18a</td>
<td>MKASGLRLEYVGRKCLFTEKCTFHMYRMRFAPNHVYAKSRFWYFVSQLHMKKSSSE 60</td>
</tr>
</tbody>
</table>

| 4-7PFS.4          | ----------|
| Dmelanogaster_L18a | IVSIKQYETSPVKNFISGRLYDGSHTNMGRDYRDLTVGAGTVQYDGMGARRHRAR 24 |
| Sfrugiperda_L18a  | IVSIKQYETSPVKNFISGRLYDGSHTNMGRDYRDLTVGAGTVQYDGMGARRHRAR 24 |
| Hsapiens_L18a     | IVSIREIPEKSPVKNFISGRLYDGSHTNMGRDYRDLTVGAGTVQYDGMGARRHRAR 120 |
| Mmusculus_L18a    | IVSIREIPEKSPVKNFISGRLYDGSHTNMGRDYRDLTVGAGTVQYDGMGARRHRAR 120 |

| 4-7PFS.4          | ----------|
| Dmelanogaster_L18a | APSIQIIKVEAIPASKTRRVHVQGSDSKIKFPLV----------59 |
| Sfrugiperda_L18a  | APSIQIIKVEAIPASKTRRVHVQGSDSKIKFPLV----------59 |
| Hsapiens_L18a     | APSIQIIKVEAIPASKTRRVHVQGSDSKIKFPLV----------59 |
| Mmusculus_L18a    | APSIQIIKVEAIPASKTRRVHVQGSDSKIKFPLV----------59 |

62
Figure 3.9 Amino acid sequence alignment of embryonic subtraction product 5-18PMS.4 with representative 40S ribosomal protein S20 sequences. The subtraction product is aligned with S20 riboprotein sequences from *D. melanogaster*, *S. frugiperda*, and *S. pombe*. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (:), while those in which a semi-conserved substitution is found are identified by a single dot (.)
<table>
<thead>
<tr>
<th>Dequinoxialis_histoneH2B</th>
<th>CAACAGGAAACTGGAGTCCAGCTCGGTTCGAGCG-AGACTTTGCCTTTCC 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dpavlovskiana_histoneH2B</td>
<td>CAACAGGAAACTGGAGCCTCGTGCAGCG-AGACTTTGCCTTTCC 99</td>
</tr>
<tr>
<td>Dinsularis_histoneH2B</td>
<td>CGACAGGAAACTGGAGTCCAGCTCGGTTCGAGCG-AGATTTTGCTTTCC 99</td>
</tr>
<tr>
<td>Dkikkawai_histoneH2B</td>
<td>CGACAGGAAACTGAGTCCAGCTCGGTTCGAGCG-AGATTTTGCTTTCC 99</td>
</tr>
<tr>
<td>3-5PMS.2</td>
<td>---------ACCTTGAGATATAATGCTGTTCTGCTTTCC 35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dequinoxialis_histoneH2B</th>
<th>CTTCCTTTGCCACCTTTGCCAGACGAGATTTGCATTTTCTCTTTCC 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dpavlovskiana_histoneH2B</td>
<td>CTTCCTTTGCCACCTTTGCCAGACGAGATTTGCATTTTCTCTTTCC 99</td>
</tr>
<tr>
<td>Dinsularis_histoneH2B</td>
<td>CTTCCTTTGCCACCTTTGCCAGACGAGATTTGCATTTTCTCTTTCC 99</td>
</tr>
<tr>
<td>Dkikkawai_histoneH2B</td>
<td>CTTCCTTTGCCACCTTTGCCAGACGAGATTTGCATTTTCTCTTTCC 99</td>
</tr>
<tr>
<td>3-5PMS.2</td>
<td>---------ACCTTGAGATATAATGCTGTTCTGCTTTCC 35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dequinoxialis_histoneH2B</th>
<th>TTT---TTCCCTTTTTACACACTGTTCACTTTTCAcACACAAATGTTGA 195</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dpavlovskiana_histoneH2B</td>
<td>TTT---TTCCCTTTTTACACACTGTTCACTTTTCAcACACAAATGTTGA 195</td>
</tr>
<tr>
<td>Dinsularis_histoneH2B</td>
<td>TTT---TTCCCTTTTTACACACTGTTCACTTTTCAcACACAAATGTTGA 195</td>
</tr>
<tr>
<td>Dkikkawai_histoneH2B</td>
<td>TTT---TTCCCTTTTTACACACTGTTCACTTTTCAcACACAAATGTTGA 195</td>
</tr>
<tr>
<td>3-5PMS.2</td>
<td>---------ACCTTGAGATATAATGCTGTTCTGCTTTCC 35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dequinoxialis_histoneH2B</th>
<th>TGTTGATGTGGTACATTGTCGTTTTATTTATACCTTTTCCAACGACATAC 245</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dpavlovskiana_histoneH2B</td>
<td>TGTTGATGTGGTACATTGTCGTTTTATTTATACCTTTTCCAACGACATAC 245</td>
</tr>
<tr>
<td>Dinsularis_histoneH2B</td>
<td>TGTTGATGTGGTACATTGTCGTTTTATTTATACCTTTTCCAACGACATAC 245</td>
</tr>
<tr>
<td>Dkikkawai_histoneH2B</td>
<td>TGTTGATGTGGTACATTGTCGTTTTATTTATACCTTTTCCAACGACATAC 245</td>
</tr>
<tr>
<td>3-5PMS.2</td>
<td>---------ACCTTGAGATATAATGCTGTTCTGCTTTCC 35</td>
</tr>
</tbody>
</table>

**Figure 3.10** Nucleotide sequence alignment of embryonic subtraction product 3-5PMS.2 with representative histone H2B sequences. The subtraction product is aligned with histone H2B sequences from *D. equinoxiaiis*, *D. pavlovskiana*, *D. insularis*, and *D. kikkawai*. An asterisk (*) identifies nucleotides found at a position in all sequences.
Figure 3.11 Amino acid sequence alignment of embryonic subtraction product 5-18PMS.3 with representative ZOS proteasome PZ-subunit sequences. The subtraction product is aligned with the *A. gambiae* agCP14390 gene product and PZ-subunit sequences from *D. melanogaster, M. musculus*, and *H. sapiens*. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (·), while those in which a semi-conserved substitution is found are identified by a single dot (.).
Figure 3.12 Amino acid sequence alignment of embryonic subtraction product 5-18PMS.6 with representative translation elongation factor eEF-1-a sequences. The subtraction product is aligned with eEF-1-a sequences from *Limnephilus* species, *B. mori*, *T. dorsalis*, *H. rubiginosa*, and *D. melanogaster*. Only the region of the representative eEF-1a sequences to which 5-18PMS.6 aligned is shown. This region represents the C-terminus of EF-1A. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (·).
Figure 3.13 Amino acid sequence alignment of embryonic subtraction product 3-5PFS.3 with representative nucleoside diphosphate kinase (NDK) sequences. The subtraction product is aligned with the NDK sequences from *C. parallela*, *D. melanogaster*, *B. taurus*, and *G. cirratum*. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (.), while those in which a semi-conserved substitution is found are identified by a single dot (.).
Table 3.4 Various medfly homolog sequences isolated from different stages of embryogenesis using subtractive hybridization. The subtraction sequences listed were isolated from a hybridization using embryonic material from the 3-5 hour, 4-7 hour and 5-18 hour AEL stages of embryogenesis. Female and male material was alternated as “driver”. The identity of each of the clones is included along with the size of each of the corresponding clones. The identity of each was assigned using the BLASTX algorithm at the NCBI and E-values are included to identify the significance of the identity reported. Accession IDs are provided for sequences that were submitted to Genbank.

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Sequence Length (bp)</th>
<th>Identity (E-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cytosolic Ribosomal Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-5PMS.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession ID: AF337889</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>5-18PMS.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession ID: AF337889</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>3-5PMS.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession ID: AF337890</td>
<td>228</td>
<td></td>
</tr>
<tr>
<td>5-18PMS.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession ID: AF337891</td>
<td>309</td>
<td></td>
</tr>
<tr>
<td>4-7PFS.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>5-18PMS.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>157</td>
<td></td>
</tr>
<tr>
<td>Clone Name</td>
<td>Sequence Length (bp)</td>
<td>Identity (E-value)</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td><strong>Other Homologs</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 3-5PMS.2 | 168 | **Histone H2B***:  
D. kikkawai (9e-19)  
D. equinoxialis (8e-16)  
D. pavlovskiana (8e-16)  
D. insularis (1e-14) |
| Accession ID: AF333963 | | |
| 5-18PMS.3 | 251 | Anoplophora gambiae CP14390 gene product (E=2e-37) |
| Accession ID: AF337892 | | |
| 5-18PMS.6 | 63 | 20S proteasome beta 2 subunit:  
D. melanogaster (E=5e-35)  
M. musculus (E=1e-27)  
H. sapiens (E=2e-27) |
| Accession ID: AA337893 | | |
| 3-5PFS.3 | 211 | Translation Elongation Factor eEF1-alpha:  
Tanyptera dorsalis (E=3e-04)  
Holorusia rubiginosa (E=3e-04)  
Limnephilus species (E=5e-04)  
D. melanogaster (E=0.001)  
B. mori (E=0.001)  
Choristoneura parallela (budworm) (E=3e-19)  
B. taurus (E=5e-16)  
Ginglymostoma cirratum (E=9e-16) |
| Accession ID: AF337894 | | |
3.2.2.3.4 Potentially Novel Genes

Approximately one-third of the subtraction products (6/21) identified did not display any significant similarity to sequences in the GenBank database using both the BLASTN and BLASTX algorithms. Clones 3-5PMS.1 (394bp), 3-5PMS.3 (141bp), 4-7PMS.4 (188bp), 3-5PFS.1 (332bp), 3-5PFS.2 (299bp), and 3-5PFS.4 (187bp) are all sequences that represent potentially novel genes.

3-5PMS.1 displays sequence similarity with a maturase protein from *Curtia verticillaris* (E=6.5), but the alignment is not statistically significant suggesting it may represent a novel sequence. A nucleotide search using the BLASTN algorithm produced alignments of 3-5PMS.1 to human DNA sequence BAC clones (E=0.44).

Searches using sequences from the remainder of the clones (3-5PMS.3, 3-5PFS.1, 3-5PFS.2, 3-5PFS.4, and 4-7PMS.4) did not reveal any significant similarity to sequences in the database.

3.2.3 Initial characterization of novel sequences

Preliminary characterization of the six products representing potentially novel sequences was initiated by designing oligonucleotide primers to amplify each subtraction product. Primers were designed for all but one of the sequences, 3-5PMS.3, due to the inability of the primer design program to identify a primer pair that could amplify the sequence. Instead, 3-5PMS.3 was labeled as a probe and used in a Southern blot of Med+/EcoRI restriction enzyme digested medfly genomic DNA to evaluate its representation in the genome. No fragments were observed to hybridize with the probe and further characterization of this sequence was not pursued.

The oligonucleotide primers that were designed for the remainder of the subtraction sequences were used in RT-PCR assays. These were intended to identify portions of transcripts as well as to estimate the level at which each of the potentially novel transcripts was expressed. The initial studies were performed to amplify first-strand cDNA isolated from females and males of the medfly Maui Coffee strain at the adult stage (1 day post-eclosion). After determination of the expression at the
adult stage, expression patterns were investigated at the 3rd instar larval, pupal (2 days), and eventually 3-5 hour embryonic stages.

The 3-5PMS.1 primers were designed to amplify 268bp out of the total of 394bp in the cloned product. The RT-PCR amplification revealed that a transcript of the expected size was expressed at all developmental stages in both females and males. Oligonucleotide primers designed from the 4-7PMS.4 sequence were designed to amplify 130bp of the 188bp subtraction product. Amplification of this small segment was observed in both sexes at all stages of development examined. No additional characterization of either of these sequences was performed since they did not appear to be differentially expressed between the sexes.

The 3-5PFS.1 primers were designed to amplify 260bp of the 332bp subtraction product, the 3-5PFS.2 primers were designed to amplify 190bp of the 299bp subtraction products, and primers for the 3-5PFS.4 sequence were designed to amplify 150bp of the 187bp subtraction product. The RT-PCR amplification results showed that the transcripts of all three sequences are present at all developmental stages in both females and males, but that the transcript is either limited to or at a higher level in females versus males at the adult stage (Figure 3.14). In the case of 3-5PFS.2, the absence of amplification in the adult males appears to be more evident than when the 3-5PF.1 primer pair is used to amplify the same template; the difference in the intensity of amplification in females versus males using the 3-5PFS.4 primers is less extreme than that seen for the previous 2 primer pairs, but the difference is still present. These results were intriguing since they suggested that the sequences were exhibiting some form of sex-specific expression.

3.3 Discussion

The successful collection of sex-specific RNAs from pre-adult stages was made possible using a molecular sexing assay based on the amplification of Y-chromosome specific repetitive sequences (Zhou, et al., 2000). The work presented here represents the first demonstration that the sexing assay
Figure 3.14 RT-PCR Analysis of 3-5PFS.1, 3-5PFS.2, and 3-5PFS.4. First-strand cDNA from female and male 3-5 hour embryos, third instar larvae, early pupae, and 2-day and 10-day adults post-eclosion were amplified using primers designed from the subtraction product sequences. The transcripts appear to be present in both sexes in all of the pre-adult stages. At the adult stage, the difference in amplification produced when using female versus male template suggests that the transcripts may be present at different levels in females versus males. Primers that amplify the Zw gene were used to ensure that equal amounts of template were present in the above amplifications and were also used as a control for cDNA synthesis as these primers are designed to direct amplification across a 200bp intron.
could be applied to embryonic, larval, and pupal material as well as adult material. The assay provides a new tool for conducting molecular experiments at early stages of medfly development, such as subtractive hybridization, that require sex-specific material.

The ability of the SSH method to successfully isolate differentially expressed sequences from the medfly genome was demonstrated by the identification of a product (APMS.2) similar to a previously characterized medfly male-specific sequence using material from the adult stage. This sequence was identified using adult female material as the driver and is similar to the male-specific serum protein (MSSP) product, reported by Christophides, et al. (2000), which exhibits male-specific expression at the adult stage. An additional adult subtraction sequence product, APMS.1, was isolated that displays significant sequence similarity with the NADH-ubiquinone reductase complex of the mitochondrial electron transport chain. The SSH method was also used to isolate one sequence from the larval stage of development, LPMS.1, which represents a potentially novel sequence, at least to the extent that it does not display similarity to any sequences presently in the Genbank database. The experiment using adult material was performed first and its results provided an incentive to perform the SSH method on sexed material from different stages of embryogenesis. An extensive discussion of the sequence obtained from these experiments follows.

3.3.1 Mitochondrial Ribosomal RNA (rRNA)

The identification of 16S rRNA sequences in the subtraction products reflects the inability of the polyA(+) isolation technique to selectively remove rRNAs from the mRNA pools collected. Since mRNAs represent only 1-5% and rRNA represents ~80% of the total RNA population, it is not entirely unexpected to see ribosomal RNA contamination in such a procedure.
3.3.2 Identification of various medfly homologs

Application of SSH at the various embryonic stages enabled the isolation of medfly homologs of numerous cytosolic ribosomal proteins as well as of histone H2B, the 20S proteasome β-2 subunit, translation elongation factor 1-alpha, and nucleoside diphosphate kinase. All of the homologs represent portions of genes that have not been previously isolated from the medfly genome, thus this data contributes to the limited information available about these particular genes in the species.

Eukaryote homologs of the above gene products are typically transcribed at considerable levels during embryogenesis and play a role in mechanisms that regulate development, i.e. translation, chromatin packaging, proteolysis. Stage-specific differential expression has been reported for the five genes and is reflective of the various requirements of the organism for protein synthesis during development. The isolation of these sequences using subtractive hybridization suggests that they are potentially differentially expressed in females versus males during embryogenesis. There does not appear to be evidence in the literature to support sex-based differential expression of all of these sequences, with the exception of ribosomal proteins (riboproteins).

Riboproteins are recruited during ribosome assembly to initiate the folding of specific rRNA segments of the 60S and 40S subunits as well as to stabilize the tertiary structure of the rRNA (Ramakrishnan and White, 1998). Riboproteins are classified as L or S based on whether they are located in the large or small subunit, respectively. In the medfly, there are only three ribosomal protein genes that have been isolated—CcPO, CcP1, and CcP2 (Gagou, et al., 1999; Gagou, et al. 2000). The genes encode the acidic ribosomal P proteins, which comprise the "stalk" of the large eukaryotic ribosomal subunit. The stalk is located in the active part of the ribosome interacting with mRNA, tRNA, and translation factors involved in protein synthesis (Tchorzewski, 2002).

The isolation of five ribosomal protein sequences will add to the riboprotein sequence information available for insect species, which is presently dominated by riboprotein sequences from D.
melanogaster and Anopheles. Their isolation is of additional importance because it suggests that they are potentially differentially expressed in female versus male medfly during embryogenesis.

The human S4 riboprotein is transcribed by two different genes, one on the X-chromosome and one on the Y-chromosome (Fisher, et al., 1990). Translationally active ribosomes are observed to contain about 90% S4-X and 10% S4Y and the S4Y gene is on the region of the Y-chromosome that has been associated with Turner's Syndrome. Haploinsufficiency of riboprotein S4 (lack of S4Y) has been implicated in manifestation of Turner females (Fisher, et al., 1990), but Wool (1996) suggests that S4Y could serve an extraribosomal function and act as a transcription factor that regulates the expression of genes necessary for maleness. Based on this, the riboproteins identified may be expressed at different levels in medfly females versus males as part of the pathway, though not the actual targets, by which genes necessary for maleness are regulated.

3.3.4 Isolation of Novel Medfly Sequences

The classification of six subtraction sequences—3-5PMS.1, 3-5PMS.3, 4-7PMS.4, 3-5PFS.1, 3-5PFS.2, and 3-5PFS.4—as novel is based primarily on E-values obtained upon sequence alignments using the BLAST algorithm. Although “hits” to the database were obtained for some of the sequences, the E-values were much greater than 0.001 and were discarded as insignificant alignments. Since an enormous amount of sequence information has been produced as a result of various genome-sequencing projects, it was not unexpected to identify these types of alignments. Additionally, the BLAST algorithm was designed to identify regions of local alignment and thus has a propensity to identify short non-significant regions of similarity, especially when utilizing BLASTN.

For a majority of the putative novel sequences, minor additional characterization was performed. Because my interest is in identifying genes that are differentially expressed in females versus males, any preliminary data that suggested the contrary was used as a basis for not pursuing further characterization of the sequence.
RT-PCR results using primers designed to amplify the subtraction sequences revealed that the transcripts for 3-5PMS.1 and 4-7PMS.4 were present at all stages of development examined (3-5 hour embryos, 3rd instar larvae, early pupae, and early adults 1 day post-eclosion), but that they did not exhibit any differential expression between the sexes.

Although primers were unable to be designed to amplify 3-5PMS.3, a Southern blot was performed that revealed no fragments were hybridizing to the labeled 3-5PMS.3 probe. A failure to identify a fragment could reflect the concentration of DNA used for the blot or even the success of the transfer, but does not necessarily mean that the sequence is not present in the genome. For example, it is possible that the sequence is present in a very high molecular weight fragment that was not resolved well on the blot. Since only EcoRI was used to digest the genomic DNA, performing additional restriction digests of the DNA with alternative enzymes may produce a fragment of a smaller size to which the 3-5PMS.3 probe can hybridize.

Using RT-PCR analysis, it was observed that transcripts of the 3-5PFS.1, 3-5PFS.2, and 3-5PFS.4 subtraction products were present at the adult stage and the amplification products were of different intensities in females versus males; this difference in intensity suggested that the respective transcripts were differentially expressed between the sexes. RT-PCR experiments performed with the same sets of primers using sexed material from early embryos, third instar larvae, and early pupae revealed that there was no apparent sex-specific difference in the presence of the transcript. The 3-5PFS.4 sequence was not characterized any further since the observed difference in transcript levels was more striking for the 3-5PFS.1 and 3-5PFS.2 sequences. Since the latter two represent potentially novel genes and they display a very intriguing expression pattern, I chose to pursue an in-depth characterization of both sequences (Chapter Three).
3.3.5 Use of SSH to Isolate Differentially Expressed Sequences in the Medfly

Subtractive hybridization has been used in insect species as a tool to isolate genes that are up-regulated at a particular developmental stage (Meszaros and Morton, 1994; Meszaros and Morton, 1996; Daibo, et al., 2001), genes that are preferentially expressed in specific tissues or structures (Smith, et al., 1990; McKenna, et al., 1994; Taylor, 2000), and sequences that can be used to differentiate between species of a genus (Clapps, et al., 1993; Crabtree, et al., 1997). Based on the precedence of its use in identifying differentially expressed sequences, subtractive hybridization was used for this purpose in the medfly. In particular, a suppression subtractive hybridization (SSH) method was utilized because of its advantages with respect to the normalization of abundant transcripts and the enrichment for differentially expressed sequences.

A diverse array of products were isolated when the SSH method was applied to material from the various embryonic stages. Because the medfly homolog sequences identified represent regions of genes known to be expressed at elevated levels during embryogenesis, it cannot be ruled out that these sequences represent false-positives; this possibility is supported by the isolation of mitochondrial ribosomal RNA sequences using the SSH method. Also, RT-PCR analysis of the subtraction products representing novel sequences suggested that they were not differentially expressed, with the exception of 3-5PFS.1, 3-5PFS.2, and 3-5PFS.4. Since RT-PCR is used only to estimate levels of expression, performing quantitative methods would be desirable in order to comment on the actual levels of expression of these sequences.

A critical component to the SSH method is the amount of starting material used since two separate pools of transcript are being compared and any differences in the initial concentration may be reflected as different levels of expression using SSH. It was assumed that an equivalent number of embryos would be sufficient to reflect equal concentrations of material, but the embryos are harvested individually and minor differences in their collection from day to day may affect concentration and quality of the RNA, despite efforts to keep the collections as controlled as possible. The
identification of the potentially novel sequences provides some hope that the SSH method worked since it is designed to isolate rare transcripts, but additional characterization will need to be performed to state this with confidence. This characterization would involve isolation of genomic clones containing these subtraction products and verifying, by sequence analysis, whether they have sequence similarity to the Genbank database. Absence of sequence similarity would suggest the sequences represent novel genes.

SSH is a powerful method that has the potential to successfully isolate differentially expressed sequences from the medfly, but it is essential that its application be coupled with a method to efficiently screen whether the products are indeed differentially expressed. On a larger scale, microarray technology could be utilized to efficiently screen through subtraction products since it allows the expression levels of thousands of genes to be monitored simultaneously (Melson and Denny, 1999). On a smaller scale, this can be accomplished by performing RNA slot blots in which sex-specific RNA pools isolated from the appropriate embryonic stages are hybridized to a membrane and probed with the individual subtraction products. Differences in levels of expression would be reflected in the hybridization intensities between the two pools. In addition, since the mitochondrial genome of the medfly has been sequenced (Spanos, et al., 2000), it can be used as a probe to remove "false-positive" clones representing mitochondrial sequences.

The isolation of high concentration, sex-specific, embryonic RNA pools from three different stages of embryogenesis is quite laborious. Preliminary Northern blot experiments using various numbers of un-sexed embryos revealed that mRNA from at least 900 embryos is necessary to perform a single Northern hybridization experiment (Figure 3.15). Since these preliminary experiments were performed using optimal conditions, i.e. no processing steps used when embryos are collected for sexing that could degrade the RNA, it’s probable that at least 1100 embryos of each sex would have to be collected to provide enough high-quality RNA. Because of this limitation, steps to screen the subtraction products for differential expression, as they were isolated, was not performed.
Figure 3.15 Northern blot analysis of cytoplasmic actin expression using variable amounts of unsexed embryonic material. The blot was generated to determine the number of individual embryos necessary for minimal detection of a transcript that is expressed at a high level, i.e. cytoplasmic actin. Minimal expression was observed using 600 embryos, while a higher level of expression was observed using 900 embryos.
Chapter 4

Characterization of Subtraction Products

4.1 Background

This chapter describes the isolation and characterization of genomic clones, designated CcPFS.1 and CcPFS.2, to further characterize potentially interesting subtraction products obtained using SSH. SSH experiments performed using developmentally staged embryos resulted in the identification of two subtraction products, 3-5PFS.1 and 3-5PFS.2. These sequences appear to represent differentially expressed genes based on preliminary RT-PCR experiments. Database searches using these sequences did not produce significant alignments, suggesting that they could be derived from novel genes in the medfly genome.

4.2 Results

4.2.1 Characterization of 3-5PFS.2

Isolation of CcPFS.2, a genomic clone containing the 3-5PFS.2 sequence

The 3-5PFS.2 subtraction product was used as a probe to screen 40,000 phage of an EMBL4 medfly genomic library. Six phage plaques hybridized with the probe. These were designated λYD2.1, 2.2, 2.3, 2.4, 2.5, and 2.6. Two of these (λYD2.1 and λYD2.2) displayed a more intense hybridization than the rest. The λYD2.1 and λYD2.2 plaques were purified and grown to obtain phage DNA. The respective phage DNAs were then digested with EcoRI to separate the phage arms from the insert sequences. Restriction digestion of λYD2.1 phage DNA produced five insert fragments of the following sizes: 20kb, 6kb, 5kb, 3.5kb, and 2.5kb in addition to the phage arms. Restriction digestion of λYD2.2 produced five fragments. These were identical in size to the other phage with the exception of a 4kb fragment instead of the 5kb fragment.

Southern blot hybridization of EcoRI digests of these phage DNA were performed using the 3-5PFS.2 subtraction product as a probe. The probe hybridized to the 6kb fragment in both digests,
suggesting that the fragment contained the 3-5PFS.2 sequence. Because both phage clones appeared to contain the same fragment of interest, only the insert from λYD2.2 was used in subsequent experiments.

The presence of restriction sites within the 6kb fragment were identified that could produce a more manageable fragment containing the region of 3-5PFS.2 similarity. A BamHI site within the 6kb fragment was identified that, when combined with EcoRI, produced a 2.8kb EcoRI/BamHI fragment. Using Southern blot hybridization, this fragment was verified to contain the 3-5PFS.2 sequence. The 2.8kb fragment was isolated and cloned into the EcoRI/BamHI site of pUC19. Presence of the 3-5PFS.2 sequence in the clone was then verified by PCR amplification using the primers designed from the original subtraction product (3-5PFS.2_F1/3-5PFS.1_B13). This 2.8kb genomic clone was designated CcPFS.2.

**Sequencing of CcPFS.2**

Sequencing of CcPFS.2 was initiated from both ends of the clone. This provided approximately 550 bases of reliable sequence on either end (1100 bases total), none of which contained identifiable 3-5PFS.2 sequence. Two approaches were pursued to facilitate sequencing. The first relied on the identification of restriction sites within the clone to possibly sub-clone smaller fragments containing the sequence of interest. The CcPFS.2 clone contains restriction sites for XhoI, PstI, EcoRV, and HindIII (Figure 4.1). The second approach relied on amplification using one primer from the pUC19 plasmid (M13F or M13R) in combination with one of the original subtraction product primers located within the insert. One product was amplified using the M13F/3-5PFS.2_F1 primer pair and a second product was amplified using the M13R/3-5PFS.2_B13 primer pair (Figure 4.1). These PCR products were then cloned into the pCR-Script (SK+) plasmid vector and sequenced from both ends. All together, this provided all but 1kb of sequencing information residing at the center of the clone. To obtain information from the remaining sequence, primers were designed from the existing sequence...
GAATTCACGTGTTGAGACTTCCAGCTGCTACATAGACAACTCTTCCACATAGACACGAGCTGTTTATGCGTGTGCTGATACGCGGCTGAATTTTATTCTCTGCTCTGTTCGTTGAAGTGCGCAAAATTGCATTAATCGTCATGATAGCCOAGTOGTIAGAGGTCATCGCTGTAACGCGTTAA
TACCCCOGCTCOAATCCCGCTAOAGGCATATIATTTCTCAATn"CAAIIIIIIIIIIAATArn'GGTGAAATTTAATAATCTGAATTTAAT'l111 II IIIGGTGAAACAOA
AATTTTCACCTCCACGTCATTCGTATTCATTCGTITCATCTCCTAAAAAOTGCTOTTGAAGGGCGCAGATTTCGTTTCTCCACATACAATAACGAAAOTCTGGTGGTTAG
AGGTCnCACTATGGTGCGCTAOGnGTCAGTI AACTCTCGCAGCCGOCAAACTGAA TTTT A Tn'GA TIT AGTT A TT A TIT
ACrrrGmGAA1itAGAAATTCAACATOOc
TAAATGAACTAGTAnOAACCAGTATGAAOCTGATGCACAATAGGATAATTOTCAGCTCTTAGOATATCGCGATGTTAAAAATAACAGTTCCTTCCATGOAAAATGGA
TGATGATGAACT ACTIGGTGGAACAA TGT AAACT AGTTGGA TTCCAGTTOACAACTAGT A
l1TTCTTCTGCAGGGrn'('T
AATGCACTGT AGOC AAAAAA rrCTTGOGC
TCGCACTAAAGGTTCCOTACAAAArITTArntTAGTTGATTTTAAATOGAGGACAAGCGCnCTCTOOTGCATAAOCCAATTTOTGACACTAAATACCGTTATTACAAT
GOTAACGAAAGOTCACCTCGTGGTATATATCAGGTGTACAACTTTGCTICCACCGTTTTTn'CTAAAAnAAAAGCTTTATTGCOAAAAAGTGCnACAGATATATIATT
CAAAOTATTGTCCTTCGCTAGCAACAACrrrCTCCCAl'CrrrCCGACAATA'ITCGGATCC

350 bases
Figure 4.1 Schematic overlap of clones used to deduce the sequence of CcPFS.2. The sequence is represented by the solid black box above and single letters identify restriction enzyme sites. The grey boxes below represent overlapping clones generated by PCR amplification using oligo primers identified at each end; these oligos are identified by blue and pink arrows. The nucleotide sequence of CcPFS.2 is reported in the box below and, within this sequence, shaded residues in the respective color of the filled arrows identify the primer sequences. E: EcoRI; X: XhoI; P: PstI; RV: EcoRV; H: HindIII; B: BamHI. M13F/M13R: primer binding sites on the pUC19 plasmid vector.
data to amplify the internal unsequenced regions. The product produced was then cloned into the pCR-Script (SK+) plasmid vector and sequenced from both ends. The sequencing information obtained from these sets of experiments was then combined to create a complete contiguous CcPFS.2 genomic sequence 2.781kb in length (Figure 4.1).

CcPFS.2, a novel sequence from the medfly genome

Using the BLASTN and BLASTX search algorithms, the entire sequence was submitted to identify any sequence similarity with entries in the Genbank database. No significant alignments were produced. Smaller fragments of the whole sequence were also submitted and similar results were obtained. Five ORFs that could encode greater than 50 amino acids were identified from the sequence data. Two ORFs were identified in the EcoRI→BamHI orientation—ORF1: 68 residues and ORF2: 57 residues. Three ORFs were identified in the BamHI→EcoRI orientation—ORF3: 58 residues, ORF4: 50 residues, and ORF5: 91 residues. These ORF sequences were queried using BLASTP, but here again no significant alignments were identified.

Additional Characterization of CcPFS.2

To identify whether putative introns were present within the CcPFS.2 genomic sequence, a comparison was made of cDNA and genomic DNA amplification products. Two pairs of primers located within the CcPFS.2 sequence were used for this purpose (Figure 4.1). Primer pairs were used to amplify genomic and cDNA templates isolated from adult female and male material, 2 days post-eclosion and a genomic product of the expected size was amplified in each case.

Rapid Amplification of cDNA Ends (RACE) experiments were performed using 5’ and 3’ RACE-Ready cDNAs that were produced using adult female material. 5’RACE experiments were conducted using the 5’ RACE Ready cDNA with 3-5PFS.2_F1 as a gene-specific primer along with the manufacturer’s universal primer (UPM) [see Methods]. This amplification produced fragments that
were 750bp and 850bp in size. The same gene-specific primer was used in a 3'RACE experiment and did not amplify a product, confirming that the proper orientation was being examined. 3'RACE experiments were then performed using the 3'RACE Ready cDNA with 3-5PFS.2_B13 as a gene-specific primer. This amplified a 1300bp and a 1500bp product. Again, the gene-specific primer was used in a 5'RACE experiment, but did not amplify a product.

The amplification of two distinct products in both experiments suggests that there may be alternative transcripts produced from the CcPFS.2 gene product. In addition, the results suggest the location of the 5' end of the CcPFS.2 gene (near one end of the CcPFS.2 genomic clone), although analysis of this region does not indicate the presence of any hallmark 5' untranslated region (UTR) sequences, i.e. A-T rich region suggestive of a TATA box. Analysis of the CcPFS.2 genomic sequence in the presumed 3' region reveals the presence of a consensus polyadenylation signal sequence (AATAAA) located approximately 400bp from the putative end. The ORFs capable of encoding greater than 50 amino acid residues were not identified from this region of the clone. The genomic sequence is relatively A-T rich, composed of 63% A-T and 37% G-C base pairs.

4.2.2 Characterization of 3-5PFS.1

Isolation of CcPFS.1, a genomic clone containing the 3-5PFS.1 sequence

The 3-5PFS.1 subtraction product was used as a probe to screen 80,000 phage of an EMBL4 medfly genomic library. Three phage plaques hybridized with the probe (λYD1.1, 1.2, and 1.3) with the last phage displaying a more intense hybridization signal than the others. All three phage picks were purified and re-screened, but only the λYD1.3 phage displayed hybridization with the probe. Three plaques were then isolated from the λYD1.3 plaque purification and another round of plaque purification was performed on these phage picks. This phage was grown to obtain phage DNA and was then digested with EcoRI to separate the phage arms from the insert sequences. Restriction
digestion of λYD1.3 DNA produced four fragments of the following sizes: 9kb, 4.5kb, 2kb, and 1.75kb in addition to the phage arms.

Southern blot hybridization of λYD1.3 phage DNA digested with EcoRI was performed using the 3-5PFS.1 subtraction product as a probe. The probe hybridized to the 9kb fragment and suggested that this fragment contained the 3-5PFS.1 sequence. Attempts to clone the 9kb fragment as an EcoRI fragment were made difficult by the fact that it migrated along with one of the phage arms. Restriction sites within the 9kb fragment were identified that could produce a more manageable fragment. A 6.8kb EcoRI/Sall fragment was produced that was verified, using Southern blot hybridization, to contain the 3-5PFS.1 sequence. This fragment was then cloned into the EcoRI/Sall site of pUC19 (Figure 4.2). Presence of the 3-5PFS.1 sequence in the clone was then verified by PCR amplification using the primers originally designed to amplify the 3-5PFS.1 subtraction product (3-5PFS.1_F1/3-5PFS.1_B1). This genomic clone was named CcPFS.1.

**Sequencing Strategy for CcPFS.1**

The approaches used to sequence CcPFS.2 were also utilized to sequence CcPFS.1. Primary sequence information was obtained by sequencing in from the ends of the clone. This provided approximately 625 bases of reliable sequence information from each end. The end bordering the EcoRI site (Figure 4.2) contained a small portion with sequence similarity to the D. melanogaster S-element transposase (E=6e-20) and to the D. virilis Tc-1-like transposase (E=3e-18), while the end bordering the Sall site contained the primer sequence for 3-5PFS.1_F1. A restriction mapping approach was taken to identify any sites within the clone that could be used to isolate only that region. CcPFS.1 contained restriction sites for ClaI, EcoRV, HindIII, SacI, and XhoI (Figure 4.2). These restriction sites were used in combination with restriction sites found in the vector cloning site to subclone smaller fragments of CcPFS.1. Smaller fragments were cloned into the pBluescript (pBS) vector. A 2kb XhoI/Pst fragment, a 1.5kb ClaI/PstI fragment, and a 1.3kb EcoRV/HindIII fragment
Figure 4.2 Schematic overlap of clones used to deduce the sequence of CcPFS.1. The genomic sequence was divided into two regions, A and B, based on the identification of separate genes in these areas of the clone. (A) CcPFS.1A, represented by the solid black box, contains an ORF encoding a polypeptide with sequence similarity to an insect transposase sequence. The purple arrow represents the direction of the reading frame and the grey section of the box below it represents a clone used to obtain sequence information about CcPFS.1A. The nucleotide sequence of CcPFS.1A is reported in the text box below in the direction of the ORF, with the ORF identified in purple. (B) CcPFS.1B, represented by the solid yellow box, contains an ORF encoding a novel polypeptide containing a region with sequence similarity to a BTB-POZ domain. The green arrow represents the direction of the reading frame and the grey boxes below it represent the clones used to obtain sequence information about CcPFS.1B. The nucleotide sequence of CcPFS.1B is reported in the text box below in the direction of the ORF, with the ORF identified in green. Filled arrows in blue and red, below each graphical representation, identify oligo primers designed in these regions. Within the reported nucleotide sequence, shaded residues in the respective color of the filled arrows identify the primer sequences. Letters below the graphic identify restriction sites within the clone and are R: EcoRI, H: HindIII, Sc: SacI, X: XhoI, C: ClaI, RV: EcoRV, S: SalI, and P: PstI. The PstI and HindIII sites designated to the right of SalI are located within the multiple cloning site of the pUC19 plasmid.
present on the *Sall* end of CcPFS.1 were subcloned to provide overlapping fragments to verify any sequencing information obtained. The fragments were successfully sequenced and used to create a 2kb contig containing the 3-5PFS.1 sequence (Figure 4.2).

Additional information from the 4kb region at the center of CcPFS.1 was obtained by using the sequence information from the ends of the clone to identify possible primers for amplification of the internal fragment. A primer pair was used to successfully amplify a 4kb region which was then cloned into the pCR-Script (SK+) plasmid vector and sequenced from both ends. This provided an additional 625bp of sequence information to the *EcoRI* end and 470bp to the *Sall* end of the CcPFS.1 clone. Overall, 1250bp of sequence information was obtained from the *EcoRI* end; this region was named CcPFS.1A. An additional 2470bp of information was obtained from the *Sall* end; this region was named CcPFS.1B (Figure 4.2). The remaining 3kb of sequence at the center of CcPFS.1 was not determined.

*Database Analysis of CcPFS.1A and CcPFS.1B*

The two regions of CcPFS.1, A and B, appear to represent two different genes that are transcribed in opposite directions.

CcPFS.1A (1250bp) contains an ORF that is 144 amino acid residues in length (Figure 4.3). A database search using the amino acid sequence revealed that it displayed significant sequence similarity to the *D. virilis Tc-l* like transposase (*E*=4e-56), the S element and Uhu element transposases of *D. melanogaster*, and a transposase from *P. platessa* (plaice) (Figure 4.4). Because this region did not contain the 3-5PFS.1 sequence, its characterization was not pursued any further.

CcPFS.1B (2470bp) contains an ORF that is 237 amino acid residues in length. A database search using the amino acid sequence revealed that a region within the predicted translated product displayed significant similarity to the BTB-POZ domain within a kelch-like swinepox virus protein (*E*=2e-09) and the *D. melanogaster* broad-complex (*BR-C*) core protein (*E*=5e-07). The swinepox virus protein
Figure 4.3 CcPFS.1 A nucleotide sequence and conceptual translation. A 144 amino acid product is shown as single letter residues. The pink residues within the conceptual translation represent the transposase region of the polypeptide.
Figure 4.4 CcPFS.1A sequence alignment with representative transposon sequences. Transposase sequences shown are from *D. virilis* (Tc-1 transposon), *D. melanogaster* (1: S element transposon and 2: Uhu transposon), and *P. platessa*. Only the regions of the transposon sequences to which CcPFS.1A displays similarity are shown. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (:), while those in which a semi-conserved substitution is found are identified by a single dot (.).
exhibits sequence similarity to the *Drosophila* kelch gene that encodes a BTB-POZ domain in addition to the kelch protein-protein interaction motif, while *BR-C* is a *Drosophila* gene that has been characterized as encoding a BTB-POZ domain. BTB-POZ is a domain that is named after proteins in *Drosophila* (*broad-complex, tramtrack, and bric-a-brac*) and a poxvirus protein (*pox-virus zinc finger protein*) that share the region in common.

*CcPFS.1B* represents a single copy gene that encodes a novel protein containing a BTB-POZ domain

The representation of *CcPFS.1B* in the genome was examined by Southern hybridization. Adult medfly genomic DNA was restriction-enzyme digested using *EcoRI* and *PstI* then blotted and probed with the *XhoI/PstI* fragment of *CcPFS.1B* (Figure 4.2b). The probe hybridized to the same single *EcoRI* fragment (15.5kb) and to the same *PstI* fragment (15.2kb) in females and males (Figure 4.5) suggesting that *CcPFS.1B* is a single copy gene.

RACE experiments were performed using 5' and 3' RACE-Ready cDNAs from adult female material, 2 days post-eclosion. 5'RACE experiments were conducted using the 5'cDNA with 3-5PFS.1_F1 as a gene-specific primer along with the manufacturer’s universal primer (UPM) [see Methods]. This amplification produced a single fragment 1100bp in length. The same gene-specific primer was used in a 3'RACE experiment, but did not amplify a product. 3'RACE experiments were then performed using the 3' cDNA with 3-5PFS.1_B2 as a gene-specific primer along with the manufacturer’s universal primer (UPM). This amplified a single 500bp product. Again, this same primer was used in a 5'RACE experiment, but did not amplify a product.

The 5' and 3' RACE products were then cloned into the pCR Script (SK+) vector for sequencing. The overlapping data sets were used to create a contiguous cDNA sequence. The entire sequence is 1576bp in length and includes the 5' and 3' untranslated regions (UTR). The coding region spans 897bp, with the ATG translation initiation start codon at position 240 and the TAA translation stop codon located at position 1134 of the cDNA sequence. The ATG at this position is believed to be the
correct translation start because it facilitates the longest ORF and represents a consensus translation initiation sequence, AAAAATG (Cavener, 1987). The cDNA encodes a polypeptide that is 299 amino acid residues in length when conceptually translated (Figure 4.6). The BTB-POZ domain represents 105 residues of this polypeptide and is present from residues 27-132. The genomic and cDNA sequences are relatively A-T rich, composed of 65% A-T and 35% G-T base pairs. There are two putative TATA box sequences located in the 5'UTR between bases -60 and -66 and -109 and -115, respectively, relative to the translation initiation codon ATG. There also appears to be a putative arthropod-specific transcription initiator sequence, TCAGT, at the immediate start of the cDNA sequence. This consensus sequence is found in 25% of all arthropod promoters within 10bp of the transcription start site (Cherbas and Cherbas, 1993).

Alignment of the cDNA sequence with the genomic sequence provided information about exon/intron structure of CcPFS.IB (Figure 4.6). The gene contains 3 exons that are 166bp, 707bp, and 489bp in length and 3 introns, two of which are located within the transcribed region (98bp and 175bp). The third intron (87bp) is in the 5'UTR located 28 bases from the translational start site. The introns within the transcription region conform to the GT-AG exon/intron splice rule (Breathnach and Chambron, 1981), but the intron in the 5'UTR does not. Two putative polyadenylation signals, AATAAA, at positions 1386-1391 and 1542-1547 are present in the 3'UTR; these span residues 1938-1943 and 2094-2099 of the genomic sequence. The second polyadenylation signal is believed to be the correct one since it is followed by a stretch of A nucleotides in the cDNA sequence that represents the polyA-tail.

The CcPFS.IB conceptually translated polypeptide (299 amino acid residues) was queried against GenBank and it was revealed that, with the exception of the BTB-POZ domain, CcPFS.IB did not exhibit sequence similarity with any other sequence in the database. To ensure that there was no bias formed for the BTB-POZ domain, separate database searches were performed with the region of BTB-POZ (residue 133-299) similarity removed. No significant alignments were produced.
Transcription Initiator (TCAGT)
Translation Initiation

87bp
98bp
176bp

5'UTR intron

RACEFa
RACEFb

3-5PFS.1_B1
3-5PFS.1_F1
MD1endFb

AATAAA
AATAAA

TAA-stop
Figure 4.6 *mapotge*’ nucleotide sequence. (A) Graphical representation of the *mapotge*’ gene structure identifying important consensus sequences and locations of oligo primers designed within the gene. (B) Nucleotide sequence of *mapotge*’ genomic clone and surrounding area. The conceptual translation is depicted above the DNA sequence and numbers refer to amino acid and nucleotide positions. The putative arthropod initiator sequence is shaded in purple, putative TATA boxes are shaded in yellow, and the 5’UTR intron is underlined. The consensus translation initiation sequence is shaded in blue and polyadenylation signals are shaded in pink. Green residues within the conceptual translation represent the region identified as the BTB-POZ domain.
on this information, it was concluded that CcPFS.1B represents a novel gene containing a BTB-POZ
domain from the genome of the medfly. The gene was named mapotge'. Mapotge' is a Chamorro
word that means "pregnant". It was chosen to reflect embryogenesis, as there is no word in the
Chamorro language that translates as “embryo”.

The BTB-POZ domain of mapotge' aligns extensively to the BTB-POZ domain sequences of the
SPV136 kelch-like swinepox virus protein (E=2e-09), the D. melanogaster BR-C core protein (E=5-
07), and the actinfilin protein of Rattus norvegicus (E=4e-07) [Figure 4.7].

Conservation of structural features within the BTB-POZ domains

The crystal structure of the BTB-POZ domain has been reported for the promyelocytic zinc finger
(PLZF) protein (Ahmad, et al., 1998). This revealed that the domain is arranged into a series of α-
helices that are flanked by β sheets at its top and bottom (Figure 4.8). The crystal structure of the
potassium channel, a representative protein containing a BTB-POZ domain, has also been solved and
has been used for functional studies of the role of the domain (Kreusch, et al., 1998). Based on these
studies, it was determined that BTB-POZ domain is present as an obligate dimer, capable of forming
both homo- and hetero- dimers with other proteins containing BTB-POZ domains (Ahmad, et al.,
1998; Bardwell and Treisman, 1994). The BTB-POZ monomers interact at a hydrophobic interface
comprised of residues that lie along the length of the protein (Ahmad, et al., 1998). Upon
dimerization, there is extensive interaction between α helices and β sheets of each monomer (Figure
4.8). Several important regions of the domain—the monomer core, the charged pocket, and the
hydrophobic face—have been recognized by identifying conserved residues within the domain using
sequence alignment results and comparing this information with the available crystal structures

The monomer core does not appear to contribute to actual protein-protein interaction, but to
stabilization of the domain. The core is comprised of conserved residues (His-48, Leu-52, Scr-56,
Figure 4.7 *mapotge' sequence alignment with representative BTB-POZ domain sequences. BTB-POZ domain sequences shown are from the SPV136 swinepox virus, R. norvegicus actinfilin, and D. melanogaster BR-C. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (·), while those in which a semi-conserved substitution is found are identified by a single dot (.).
Figure 4.8 Identification of conserved residues within the BTB-POZ domain. The BTB monomer ribbons are shown in green and blue. The gray arrows identify the hydrophobic surface region while the purple arrow indicates the conserved charged pocket. Amino acid positions referred to in the text are identified above the alignment. Conserved residues within the monomer core and the charged pocket are shaded gray. An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (:), while those in which a semi-conserved substitution is found are identified by a single dot (.).

101
and Tyr-88) within the α2, α3, and α4 helices, all of which are also present within the BTB-POZ domain of *mapotge* (Figure 4.8).

The charged pocket region of the domain is believed to be a site of protein interaction characterized by conserved Asp and Arg residues within each of the monomers that are symmetrically located upon dimerization. The Asp-35 residue is conserved within almost all BTB-POZ proteins, with the exception of the potassium channel tetramerization domain (Aravind and Koonin, 1999). The conserved Asp-35 is present in *mapotge*.

The hydrophobic face of the domain is situated on the side of the molecule opposite the charged pocket and is comprised of the α5 and α6 helices as well as the anti-parallel β1 and β5 sheets from both monomers (Ahmad, et al., 1998). The β-sheets contribute to dimer stability and conserved Leu residues at positions 11, 92, and 114 in the β1, β5, and α6 chains, respectively, are believed to play a role in this stabilization (Melnick, et al., 2000). Ahmad, et al. (1998) report that there are several BTB-POZ domain containing proteins, including most of the poxvirus proteins, that have a shortened N-terminus and do not form the β1-sheet. In these cases, residue conservation is seen in the β5 sheet. *mapotge* encodes a representative protein with a shortened N-terminus and conservation of Leu-92, which resides in the β5 sheet.

Another structure commonly found in genes encoding the BTB-POZ domain is the presence of an intron within the domain coding region. Two introns in the *fru* gene (Davis, et al., 2000) and one intron in the *abrupt* gene (inferred based on Genbank entry) disrupt the BTB-POZ coding region. These introns are relatively short (<200bp). In *mapotge*, one intron (98bp) is present in the region encoding the BTB-POZ domain at position 658-756 of the genomic sequence (Figure 4.6).

*mapotge* is expressed throughout embryogenesis and in adult females

For expression studies using RT-PCR, primers were designed to span the entire length of the cDNA product. One forward primer was positioned before the 5'UTR intron (RACEFa) and another
was positioned after the 5'UTR intron (RACEFb) in order to determine whether the 5'UTR intron is spliced out (Figure 4.6a). Amplification of first-strand cDNA template (RT-PCR) from sexed 3-5 hour embryos, 3rd in-star larvae, early pupae, and adults 2 days post-eclosion was performed using the RACEFa/3-5PFS.1_F1 primer pair. This primer pair amplified a 1587bp product using genomic DNA as template, while a product of 1226bp was amplified when the above cDNAs were used as template. The size of the cDNA amplification product corresponds to the expected product size if the 5'UTR intron were in fact spliced out, despite the absence of the consensus GT-AG exon/intron splice sequence in this intron.

The RT-PCR experiments revealed that the mapotge' transcript is present in both sexes at the 3-5 hour embryonic stage and in adult females (2 days post-eclosion) (Figure 4.9). Although the transcript appears to be present in both sexes at the larval, pupal, and in adult males, the amplification products are much less intense than that seen in embryos and adult females. When larval, pupal, and adult male material was used as template, two amplification products were consistently produced that are 1587bp and 1226bp in size. These appear to reflect the presence of un-processed and processed transcripts, respectively. The 1587bp product is not seen using embryonic and adult female material as template.

These results suggest that the levels of mapotge' transcript decrease dramatically during the transition from the embryonic to the larval stage. To determine whether the mapotge' transcript is present throughout embryogenesis, first-strand cDNA from unsexed embryonic material was synthesized using material from the 5-9 hour and 12-24 hour stages of embryogenesis. These templates were amplified using the RACEFa/3-5PFS.1_F1 primer pair (Figure 4.6a). The transcript is expressed at the 5-9 hour and 12-24 hour stages of embryogenesis (Figure 4.10).
Figure 4.9 RT-PCR analysis of the *mapotge* transcript throughout development. (A) First-strand cDNA isolated from female and male 3-5 hour embryos, third instar larvae, early pupae, and 2-day and 10-day post-eclosion adults was used as template for RT-PCR. The *mapotge* transcript is present at the embryonic stage in both females and males and in adult females. Minimal transcript levels appear to be present at the larval, pupal, and adult male stages. (B) The same material was also used as template for amplification with primers designed from a housekeeping gene, *Zw*, to ensure equivalent amounts of template were used in (A).
Figure 4.10 RT-PCR analysis of the *mapotge*’ transcript during embryogenesis. First-strand cDNA from unsexed, developmentally staged embryos was used as template for RT-PCR. Embryos were collected from the 3-5 hour, 5-9 hour, and 12-24 hour AEL stages of embryogenesis. The *mapotge*’ transcript is present at all stages investigated.
mapotge' expression in adult female abdomen material

The 3-5PFS.1 subtraction product was originally isolated from a screen of early embryonic material. The apparent female-specific expression at the adult stage coupled with the apparent lack of differential expression between the sexes at the investigated embryonic stage (RT-PCR results) was initially puzzling. To examine this, first-strand cDNA was synthesized using material from separate head, thorax, and abdomen regions of female and male adults (10 days post-eclosion); this template was used in RT-PCR amplification with the RACEFa/3-5PFS.1_F1 primer pair (Figure 4.11). The mapotge' transcript appears to be expressed in females nearly exclusively in the abdomen as evidenced by the intense amplification product. The transcript is also present in the abdomens of adult males, but at low levels in comparison with females. The sizes of the fragments in the male may reflect unprocessed and processed transcripts. Amplification of the transcript in the female abdomen material, together with the fact the original sequence was isolated from embryonic material suggests that the products are a reflection of the embryos carried in the female abdomen. The amplification in males is consistent with the results seen using larval and pupal material and indicates that there is some basal level of transcript present.

When the original subtraction product primers were used to amplify first-strand cDNA synthesized from adult males, successful amplification appeared to depend on the annealing temperature used. This temperature effect was not seen in females. Based on this, an oligo primer (MD1endFb) was designed 5bp downstream of the second polyadenylation signal (Figure 4.6) and used in combination with the 3-5PFS.1_F1 primer to test whether there was a difference at this end of mapotge' that might be affecting amplification in adult males. Here, the expected amplification product was 463bp in size. Adult female and male first-strand cDNA were amplified at different annealing temperatures to determine the threshold temperature at which amplification using the male template was lost. Using male material as template, amplification was observed using an annealing temperature of 56°C, but was not observed at 59°C (Figure 4.12). Using female material as template,
Figure 4.11 RT-PCR analysis of the *mapotge* transcript in the thorax and abdomen of adult females and males. (A). First-strand cDNA isolated from the thorax and abdomen of adult females and males was used as template to amplify the *mapotge* transcript. The transcript is predominantly present in the abdomen of females, reflecting expression within embryos carried in the abdomen. The same material was used as template in amplification with primers designed from a housekeeping gene, *Zw*, to ensure that equal amounts of template were used in the *mapotge* amplification.
Figure 4.12 Analysis of the 3’ end of *mapotge’* in males. Primers designed from the 3’ end of *mapotge’* were used to amplify female and male adult cDNA templates (2 days post-eclosion). (A) An increase in three degrees results in the abolishment of amplification in males and (B) nucleotide sequence analysis of the last 110 bases of the amplification products reveals that this may be attributed to a truncation of the respective cDNAs in males. The sequence of the male product amplified at 56°C is identified as Male (+), while the sequences of the male products that are undetectable at 59°C are identified as Male (-) 1 and Male (-) 2. The polyadenylation signal is shaded in gray.
amplification was still observed at 59°C (Figure 4.12). These cDNA templates were also amplified using primers designed from the Zw gene to ensure that equal amounts of template were being used in the experiments with 3-5PFS.1_F1/MD1endFb.

Despite no detectable amplification product at 59°C using the male template, the reaction was used for cloning into the pCR Script (SK+) vector. Two clones were isolated, designated MD1end (male-)1 and MD1end (male-)2, and sequenced. The male product amplified at 56°C was cloned into the same vector and sequence information was obtained from three different clones, MD1end(male+)1, 2, and 3. Finally, the female product amplified at 59°C was cloned into the same vector and sequence information was obtained from one female clone.

Sequence alignment of the clones revealed that the difference between them existed at the end near the polyA signal (Figure 4.12). MD1end (male-)1 was truncated after the first base of the polyA signal, while MD1end (male-)2 was truncated 7bp downstream of the polyA signal. No truncation was apparent in any of the (male+) clones and the female clone. This truncation abolishes the template required for annealing of the MD1endFb primer, explaining why no amplification is observed.

Northern hybridization analysis of mapotge' expression

The prevalence of the mapotge' transcript was examined in adult females and males of various ages by Northern blot hybridization. Poly(A+) RNA was isolated from sexed adults that were 0-0.5 days, 10 days, and 16 days post-eclosion and these were probed using the mapotge' cDNA sequence as well as the medfly cytoplasmic actin sequence (AORF) to ensure that equal amounts of product were loaded on the gel. The AORF probe hybridized to a 1.6kb transcript in all stages and sexes of adulthood examined. The mapotge' probe hybridized to a 1.2kb transcript in adult females 10d and 16d post-eclosion, but did not hybridize with transcripts from 0-0.5 day adults of both sexes and from
adult males 10 days and 16 days post-eclosion (Figure 4.13). Since 0-0.5 day females are not sexually mature, it is assumed that the embryos they carry are unfertilized.

4.3 Discussion

Because of the integral role that differential gene expression plays in the development of an organism, an attempt to isolate such genes was initiated in this project by the application of SSH to embryonic material from the medfly. Two novel genes (CcPFS.2 and mapotge') have been identified and characterized here as a result of their isolation from the SSH method.

CcPFS.2 was identified as a novel sequence based on the lack of sequence alignment with the Genbank database. RACE experiments indicate that the gene represented by CcPFS.2 may produce alternative transcripts. Further sequence analysis of the RACE products could provide additional information about this prospect and subsequent Northern blot hybridization of the cDNA products to adult material would help to elucidate this.

mapotge' (CcPFS.1B) is a novel gene from the medfly harboring a BTB-POZ domain. The identification of numerous conserved, critical residues within the BTB-POZ domain of mapotge' supports this characterization. The BTB-POZ domain is comprised of 120 conserved amino acid residues that are always found at or near the N-terminus of the protein. The domain functions as a protein-protein interaction module by mediating the dimerization and oligomerization of proteins containing the domain, and it is often found in proteins that are involved in critical aspects of development (Albagli, et al., 1995). Different classes of these proteins are categorized according to the presence of additional motifs, i.e. zinc fingers and other protein-protein interaction motifs (Aravind and Koonin, 1999). mapotge' appears to belong to a class comprised of genes containing a BTB-POZ domain in the absence of any additional motifs. Representative genes from this category include bric-a-brac (bab) and germ-cell-less (gel). The bab gene functions as an integrator of signals from the sex determination and homeotic pathways in the formation of the sex-specific abdominal
Figure 4.13 Northern blot analysis of *mapotge* expression. Poly(A+) RNA from female and medfly adults 0-0.5 days, 10 days, and 16 days post-eclosion were blotted and subsequently probed with the *mapotge* cDNA. The transcript is expressed in 10 day and 16 day females, but is absent from 0-0.5 day adults of both sexes and from 10 day and 16 day males. The material was also probed with the cDNA of cytoplasmic actin to ensure that equal amounts of material were loaded in the female and male lanes within each adult stage examined. These results suggest that *mapotge* is not expressed within embryos carried by newly emerged adult females.
pigmentation pattern in *Drosophila* (reviewed in Bopp, 2001). The *gel* gene is a maternally supplied product that is necessary for germ-cell development (Jongens, et al., 1994).

Temporal expression analysis using RT-PCR indicated that the *mapotge* transcript is present in 3-5, 5-9, and 12-24 hour developmentally staged embryos. In addition, Northern blot analysis of 0-0.5 day, sexually immature adult females revealed that the transcript is not present in embryos carried by females this early in development. In order to confirm that the *mapotge* expression observed in adult females is attributed to expression within embryos carried in its abdomen, it would be of interest to conduct Northern blot analysis using RNA isolated from sterile adult females with no ovaries and eggs.

Overall, the *mapotge* transcript appears to be expressed throughout embryogenesis. The abundance of the transcript from the embryonic to the larval and pupal stages appears to decrease dramatically. The persistence of the *mapotge* transcript throughout embryogenesis suggests that it is not maternally supplied since the presence of maternal transcripts is typically restricted to early embryogenesis. Prevalence of the transcript also implies that it may play a housekeeping role at this stage of development, but, coupled with the presence of the BTB-POZ domain, it seems more probable that the gene performs a regulatory role at this stage. This presumption is based solely on the functional roles that have been elucidated for various proteins containing a BTB-POZ domain.

The precise function of *mapotge* remains unclear, but efforts to deduce its role incorporate hybridizing the transcript to whole embryos via *in situ* hybridization in order to identify specific tissues in which the transcript is expressed or whether the transcript is expressed ubiquitously. Established protocols for the successful application of this procedure in the medfly are sparse and variations of protocols used effectively in *Drosophila* and zebrafish are currently being tested.

At the larval, pupal, and adult male stages, RT-PCR experiments showed that there appeared to be two transcripts from *mapotge*'. One of the transcripts was the same size as the expected genomic product, suggesting that it represents an un-processed transcript, while the other was the size of an
expected processed transcript. DNAse-treated total RNA was the template for synthesis of the first-strand cDNA used in these experiments. Primers designed to amplify across an intron of the Zw gene were used to amplify the cDNA template (He and Haymer, 1999) in order to verify that there was no genomic DNA contaminant present. In this case, no products representing genomic DNA were produced, indicating that the amplification of an apparent un-processed transcript is genuine. Northern blot hybridization of adult male material from 0-0.5 days, 10 days, and 16 days did not identify any expression of mapotge'. It remains unclear whether the presence of unprocessed transcripts plays any role at these stages.

5' and 3' UTR Sequence Analysis

In addition to the presence of the BTB-POZ domain, an interesting property of mapotge' is the presence of an intron in the 5'UTR. Deutsch and Long (1999) surveyed the distribution of exon-intron structure in eukaryotic genes by obtaining sequences representing ten model organisms. It was discovered that of the 74 genes containing introns/exons in the database, seven contained a 5'UTR intron. In Gossypium, an 1100bp intron present in the 5'UTR of the FAD2-1 fatty acid desaturase gene was used to infer phylogenetic relationships between Gossypium species, but no functional role of the intron was implied (Liu, et al., 2001). A role for a 5'UTR intron was suggested by studies of male-specific lethal 2 (msl-2) in Drosophila. Msl-2, an immediate target of Sxl, is expressed exclusively in males and helps regulate dosage compensation. Splicing of the gene's 133bp 5'UTR intron usually occurs in males and not in females, but the splicing does not affect the reading frame in either case and is not required for the male-specific expression of the gene (Kelley, et al., 1997). Rather, the intron provides binding sites for SXL and this leads to subsequent repression of translation of msl-2 in females.

RT-PCR experiments indicated that the 5'UTR intron of mapotge' is spliced out at all stages of embryogenesis. Analysis of the intron sequence did not identify any potential protein or RNA
binding sequences. This suggests that, although it is an interesting property of \textit{mapotge}', any role of the 5'UTR intron in gene regulation is not clear.

Additional examination of sequences at the 3'end of \textit{mapotge}' were performed using a primer pair that amplified past the second polyadenylation signal (MD1endFb). This examination was prompted by a difference in amplification that appeared to be regulated by temperature. Male adult transcripts appear to be truncated near the second polyadenylation signal at certain temperatures. These temperature effects were not seen when female cDNA was used as template. Several possible scenarios that could explain this are: i.) the presence of different populations of transcripts in females and males, with the male transcripts comprised of full-length and truncated 3' ends and the female transcripts represented by full-length ends, or ii.) an actual difference in sequence at the 3'end that, upon increase in annealing conditions, does not permit the annealing of the MD1endFb primer and prevents isolation of the appropriate sequence from the region. In the second case, the difference in sequence would have to exist in the 3'UTR of the cDNA since the MD1endFb/3-5PFS.1_F1 primer pair can successfully amplify female and male genomic DNA templates, even at the more restrictive temperatures used.

\textbf{4.4 \textit{mapotge}' as a product of suppression subtractive hybridization (SSH)}

Based on the molecular characterization of \textit{mapotge}', it appears that it represents a developmentally regulated gene but not one that is expressed in a sex-specific manner. The SSH method is designed to enrich for sequences that display a 5-fold to 1000-fold difference in the level at which they are expressed. It is possible that the level of differential expression of these sequences is small and undetectable by expression analysis experiments performed here. Since the BTB-POZ domain can mediate the formation of heterodimers, \textit{mapotge}' could potentially interact with both female-specific or male-specific transcription factors that contain a BTB-POZ domain. This
interaction could potentially explain the isolation of *mapotge* in a screen for genes that are differentially expressed in females versus males during embryogenesis.

In combination with the overall results of the SSH experiments that were performed using embryonic material (Chapter 3), it is more likely that *mapotge* is a novel gene that is expressed during embryogenesis and that plays a role in a general mechanism regulating development at this stage. This indicates that application of SSH to obtain sex-specific expression products at the embryonic stage in the medfly require modifications of the protocol used here. Primarily, the concentration of the material used for SSH is of significance. 100 embryos of each sex were used as starting material for the experiment, but perhaps double or triple this amount would be more advantageous to account for any degradation of the template that may occur during the sexing process. The development of a more efficient assay to screen subtraction products for differential expression would be beneficial. The assay would ideally eliminate subtraction products that are recovered simply because they are expressed at a high level.

The isolation of *mapotge* from the medfly genome was intriguing because it represents a novel gene and because this characterization contributes to the biological information available about this pest species. The properties of *mapotge* in combination with its expression profile suggest that it plays a role in regulating embryonic development. *mapotge* represents the first gene containing a BTB-POZ domain to be completely characterized from the medfly.
Chapter 5

Preliminary Characterization of the Representation of Additional Genes Encoding the BTB-POZ Domain in the Genome of the Medfly

5.1 Background

At least 64 genes encoding a BTB-POZ domain are thought to be present in Drosophila based on information obtained from the Drosophila Genome Project (Rubin, et al., 2000). In humans, 113 BTB-POZ coding sequences were estimated to be present as discovered by the Human Genome Sequencing Project (Lander, et al., 2001). Members of these families, all of which contain the conserved BTB-POZ domain, have been loosely categorized according to the presence of other motifs in addition to BTB-POZ (Aravind and Koonin, 1999) [Figure 5.1]. The largest class of these proteins includes those that have the domain in the presence of the C2H2-type zinc finger DNA binding domain. These proteins are usually involved in transcriptional regulation and chromatin remodeling. In this context, it has been proposed that the BTB-POZ domain could mediate the formation of higher-order zinc-finger protein oligomers, i.e. tetramers, hexamers, which enhances the interaction of the zinc-finger proteins with their DNA target. Examples of these proteins are produced by the Drosophila genes tramtrack (ttk), involved in the development of photoreceptor and cone cells in the developing eye (Wen, et al., 2000), fruitless, which regulates male sexual courtship behavior (Ryner, et al., 1996), and trithorax-like, which produces a GAGA protein (Granok, et al., 1995). In addition, the BTB-POZ domain appears to play a role in mediating the tetramerization of subunits in various subfamilies of animal potassium channels (Aravind and Koonin, 1999).

Proteins containing a BTB-POZ domain are also characterized by the additional presence of other protein-protein interaction motifs, such as kelch repeats, ankyrin repeats, and MATH domains (reviewed in Aravind and Koonin, 1999). Lastly, some proteins contain a BTB-POZ domain in the absence of any other structural motif. Examples of these are the bab, gc (Jongens, et al., 1994), and the mapo gene isolated in this project.
Figure 5.1 Domain architecture of proteins containing a BTB-POZ domain. Representatives from *D. melanogaster* of the different genes encoding proteins containing a BTB-POZ domain in the presence of the C2H2 zinc finger motif (A), kelch repeats (B), and in the absence of accessory motifs (C). *mapote* is included as a member of the last category.

Adapted from Aravind and Koonin (1999)
The identification of a gene producing a member of the BTB-POZ family of proteins in the medfly was intriguing primarily because of the fundamental importance of these genes in developmental pathways. This is also the first completely characterized gene of this type to be isolated from the species. The fact that genes of this type are usually members of a family of genes also prompted the investigation of the representation of additional genes containing the domain within the medfly genome. Southern blot analysis suggests that there are at least four genes, including *mapotge’*, containing this domain in the medfly. Degenerate-primed PCR was used as an alternative approach to obtain sequences representing the domain from the medfly genome.

5.2 Results

5.2.1 Representation of additional sequences containing a BTB-POZ domain in the medfly

The representation of related sequences in the medfly genome was analyzed using a probe representing a 266 bp fragment from the BTB-POZ domain of the medfly gene *mapotge’*. This product, generated by PCR amplification, was then used to probe female and male medfly genomic DNAs restriction digested with *BamHI*. The blot was treated using very low stringency conditions to enable the identification of as many representative fragments as possible. The probe hybridized to four fragments that are 13.7kb, 10.8kb, 8.4kb, and 7.3kb in size (Figure 5.2). The CcPFS.1 genomic fragment that contained *mapotge’* was initially identified as a 9kb *EcoRI* fragment that did not contain any restriction sites within the sequence for *BamHI*. It is possible that *mapotge’* is represented on this Southern blot as either the 10.8kb or the 13.7kb *BamHI* fragment. These results suggest that, in addition to *mapotge’*, there are at least three other genes containing the BTB-POZ domain within the medfly genome.
Figure 5.2 Genomic representation of sequences potentially encoding a BTB-POZ domain in the medfly. Medfly genomic DNA was restriction enzyme digested with *BamHI*, blotted, and subsequently probed with a fragment representing the BTB-POZ domain of *mapotge*. The probe hybridized to four fragments, identified above by arrows. The approximate size of each fragment is identified.
5.2.2 Identification of BTB-POZ domain sequences in the medfly

An alternative method used to obtain sequences encoding the BTB-POZ domain in the medfly was carried out using degenerate primed PCR. Alignment of sequences from various insect and human BTB-POZ domain representatives has identified the presence of conserved residues within the domain (Zollman, et al., 1994; Aravind and Koonin, 1999; Melnick, et al., 2000). An alignment was produced comparing ten genes that contain a BTB-POZ domain, nine from Drosophila and one from humans. Degenerate primers were then designed, using medfly codon bias information (He and Haymer, 1992), to amplify sequences corresponding to the conserved amino acid residue regions, LSACSPY and DFMYTGE, of the domain (Figure 5.3). These regions are known to contain residues that are conserved within the monomer core of the domain (Ahmad, et al., 1998; Melnick et al., 2000).

These primers were used to amplify cDNA templates derived from first-strand cDNA from female 3-5 hour embryos and female adults 2-days post-eclosion. Fragments ranging in size from 128-132 bp were identified. These were cloned and sequenced, and the tentative identity of each was assigned using the BLASTX algorithm. Sequences obtained from seven clones displayed significant similarity to the BTB-POZ domain of sequences in the database. Of these, four (BTB2, BTB5, BTB9, and BTB10) exhibited 93% identity to the BTB-POZ domain of Drosophila ttk (E=4e-17). Nucleotide sequence alignment of the medfly ttk-like and Drosophila ttk domains reveals variability between the medfly sequences (Figure 5.4). The sequence BTB9 displays more similarity to the Drosophila ttk at the nucleotide level than do the other medfly sequences.

BTB7 (132 bp) displays 92% identity to the CG8924 gene product of Drosophila (E=6e-16), 88% identity to the agCP1249 gene product of A. gambiae (E=5e-10), and 79% identity to the BTB-III protein of Drosophila (E=5e-10). A query of the Genbank database was also conducted using the BTB7 sequence with the oligo primer sequence removed. This search identified significant sequence similarity to the sequences listed above. There is no information listed in the description of the first
Figure 5.3 Amino acid sequence alignment of BTB-POZ domain sequences. The regions from which the degenerate primers were designed are shaded in gray. Gaps are identified by dashes. The residues are colored based on their biochemical properties (Red: small, hydrophobic; Blue: acidic; Pink: basic; and Green: hydrophilic and H,C,G). An asterisk (*) identifies residues found at a position in all sequences. Those residues in which a conserved substitution is located are identified by two dots (;), while those in which a semi-conserved substitution is found are identified by a single dot (.).
Figure 5.4 Nucleotide and amino acid sequence alignment of medfly
tramtrack-like and *Drosophila* (Dm) tramtrack (tk) BTB-POZ domain sequences. (A) Nucleotide sequence alignment of medfly
ttk-like sequences, BTB2, BTB5, BTB9, and BTB10, with the
*D. melanogaster* sequence encoding the BTB-POZ domain of *tk*. Bases shaded in gray are those that differ from the medfly in
*Drosophila*. Bases shaded in purple are those that differ in the medfly from the other medfly
sequences and from *Drosophila*. (B) Amino acid sequence alignment of the translated BTB2
sequence with the BTB-POZ domain of *ttk*. Despite the variability in nucleotide sequence, 93%
amino acid identity between the medfly sequence and *Drosophila* *ttk* is observed. An asterisk (*)
represents bases that are identical in all sequences while those in which a conserved substitution is
located are identified by two dots (:).
two sequences to suggest that they represent sequences encoding BTB-POZ domains. But, a search of the database using BLASTP with the conceptual translation of BTB7 identified the sequence as a putative conserved BTB-POZ domain. Additional significant alignments of BTB7 to sequences in the database include the BTB-POZ domain containing Drosophila genes *pipsqueak* (*psq*) (E=8e-10), *bab* (E=3e-08), and *lola* (E=3e-08) to which BTB7 has 78%, 83%, and 78% identity, respectively (Figure 5.5).

BTB8 (132 bp) exhibits 97% identity to the ebiP8749 gene product of *A. gambiae* (E=8e-15) and to the *Drosophila* genes *ribbon* (E=3e-11), *lola* (E=3e-05), and BTBII (E=3e-05) to which BTB8 has 82%, 69%, and 66% identity, respectively (Figure 5.5). A query of the Genbank database was also conducted using the BTB8 sequence with the oligo primer sequence removed. This search identified significant similarity to the *A. gambiae* and *ribbon* sequences listed above. All of the *Drosophila* genes identified contain a BTB-POZ domain and, as with the case of BTB7, a search of the database using the BTB8 conceptual translation revealed that the sequence represents a putative conserved BTB-POZ domain.

### 5.2.3 BTB5 is present as a single-copy sequence in the medfly

The *ttk* gene is present as a single copy in the *Drosophila* genome. It is alternatively spliced to produce two proteins that are identical at the N-terminus, but that differ at the C-terminus with respect to the arrangement of zinc-finger domains (Read and Manley, 1992). Since BTB2, 5, 9, and 10 display 93% identity to the BTB-POZ domain of *ttk*, it is highly probable that these sequences represent the BTB-POZ domain of a *ttk*-like medfly homolog. To investigate whether the *ttk*-like domain is also present as a single copy in the medfly genome, BTB5 was used to probe medfly female and male genomic DNA digested with *EcoRI*, *HindIII*, and *PstI*. The probe hybridized to a single 6.815kb *EcoRI* fragment, a single 5.10kb *HindIII* fragment, and a single 10.4kb *PstI* fragment. This suggests that the *ttk*-like homolog is also single-copy in the medfly (Figure 5.6).
Figure 5.5 Amino acid sequence alignment of BTB7 and BTB8 with respective homologous sequences. (A) BTB7 amino acid sequence alignment with the D. melanogaster CG8924 gene product, the A. gambiae agCP1249 gene product, and the D. melanogaster BTB-III domain sequence. (B) BTB8 amino acid sequence alignment with the A. gambiae ebiP8749 gene product and the BTB-POZ domain sequences of the Drosophila genes ribbon and lola, and the Drosophila BTB-II domain.
Figure 5.6 Genomic representation of BTB5. Medfly genomic DNA was restriction enzyme digested with EcoRI (E), HindIII (H), and PstI (P), blotted, and subsequently probed with the BTB5 nucleotide sequence. The probe hybridized to a single fragment in each lane, suggesting that the gene containing BTB5 is present as a single copy in the medfly.
Possible evolutionary relationships of the medfly BTB-POZ sequences with domain sequences from other insect species

A CLUSTALW alignment was produced comparing twenty-seven BTB-POZ domain sequences from arthropod species, including the medfly sequences identified in this project. Eighteen of the sequences were identified in *Drosophila*, while the remainder were identified in other arthropod species. A Neighbor-Joining tree was then constructed using the aligned sequences above with 1000 bootstrap replications (Figure 5.7).

This analysis identified several clusters of POZ domains, including the domain sequences of the insect potassium channels (Shak, Shal, and Shab), the domain sequences of kelch-type proteins (*kelch* and *diablo*), and the domain sequences of zinc-finger proteins (*GAGA, BR-C, ttk*, and *fru*). These clusters were supported by bootstrap values. The placement of other sequences within the tree, including domain sequences that are not associated with other structural motifs (*MOD_MDG4, gcl*, and *bab*) and some domain sequences associated with zinc-fingers (*ken* and *lola*) were not supported by bootstrap values. The medfly BTB2 sequence formed a cluster with *ttk* (100% bootstrap). Medfly BTB7 and BTB8 were grouped with BTB-II and *lola*, respectively. *mapotge* formed a sister group to the potassium channel proteins, but this was not supported by bootstrap values at the 50% level.

Discussion

5.3.1 BTB-POZ domain sequences in the medfly genome

The numerous proteins containing a BTB-POZ domain encompass a protein family, the members of which have been loosely categorized according to the presence of other structural motifs in addition to BTB-POZ (Aravind and Koonin, 1999). By combining the data from Southern blot analysis and DP-PCR, four unique sequences encoding a BTB-POZ domain were identified in the medfly that represent four different genes containing the domain. One of the four medfly sequences identified has a high percent similarity with the BTB-POZ domain of *ttk*, suggesting that it is a
Figure 5.7 A phylogenetic tree of arthropod BTB-POZ domain sequences. The figure shows a neighbor-joining tree constructed using 27 aligned sequences of arthropod BTB-POZ domains with 1000 bootstrap replications. Only bootstrap values >50% are shown. The blue bar indicates the cluster of BTB-POZ domains associated with zinc-finger proteins. The green bar identifies the domains within the zinc-finger protein cluster that are not associated with any accessory structural motifs. The red bar identifies the tetramerization domain of the insect potassium channels, and the purple bar identifies domains associated with kelch-containing proteins. Domain sequences presented as only the gene name were identified in *Drosophila*, while domain sequences preceded by a species name were identified in that particular species. Yellow arrows identify sequences isolated from the medfly as part of this project.
medfly clone encoding a ttk-like BTB-POZ domain sequence. BTB7 and BTB8 exhibit the highest percent identity to gene products that are uncharacterized, but residues conserved within BTB7 and BTB8 suggest that they can be identified as representative domain sequences.

In the medfly, a sequence encoding a fru-like BTB-POZ domain was previously identified as part of an evolutionary analysis of the conservation of the domain within fru genes (Davis, et al., 2000). Davis, et al. (2000) also identified sequences encoding the fru-like BTB-POZ domain from the related Tephritids, B. cucurbitae and B. dorsalis. Full characterization of these genes was not pursued past identification of the domain. This information, in combination with the data obtained in this project, provides sequence information for at least five genes encoding the BTB-POZ domain in the medfly genome.

5.3.2 Evolutionary relationship of the BTB-POZ domain in insects

Aravind and Koonin (1999) examined the evolutionary relationship of the BTB-POZ domain within eukaryotes, including vertebrates, nematodes, and arthropods, by comparing domain sequences from these organisms. In this study, zinc finger-containing and related proteins in Drosophila formed a monophyletic clade as did the tetramerization domain of the Shal, Shak, and Shab subfamily of potassium channel proteins (Aravind and Koonin, 1999).

Examination of the distribution of the BTB-POZ domain within insect species has remained slightly biased, as domain sequences have predominantly been isolated from Drosophila species, particularly from D. melanogaster. There are a limited number of other insect species from which BTB-POZ domain sequences have been identified. In addition to the medfly sequences isolated here, domain sequences are also present in Broad-complex from Manduca sexta (tobacco moth), pipsqueak from Apis mellifera (honeybee), and fru from B. cucurbitae (melon fly), B. dorsalis (oriental fruit fly), and Zaprionus tuberculatus (Drosophilidae). The domain sequences reported for the last three were isolated strictly to examine the domain and not the gene product.
An evolutionary analysis was performed using the BTB-POZ sequence information available for arthropods. This analysis identified clusters of POZ domains that appear to correspond with the loose categorization proposed by Aravind and Koonin (1999). The tetramerization domain sequences of the insect potassium channels form a cluster (bootstrap support, 99%; Figure 5.6). This cluster forms a sister group to the BTB-POZ domain of mapotge’, but bootstrap support was minimal (32%, not shown). The kelch-type domains also formed a distinct group (bootstrap support, 99%; Figure 5.7). The third cluster was comprised of domains known to be associated with zinc-fingers. Two of the BTB-POZ domains grouped together with this cluster, PSQ and BAB, are found in proteins that do not have any accessory structural motifs.

As expected from sequence data, the domain from BTB2 formed a cluster with the domain from ttk (bootstrap support, 100%; Figure 5.6). Low bootstrap support for the placement of BTB7 and BTB8 was anticipated, as the percent identity of these domains was highest to uncharacterized gene products in the database.

5.3.3 Sequences encoding BTB-POZ domains in the medfly

The identification of several sequences encoding unique BTB-POZ domains in the medfly suggests that there is a family of these proteins in the species. This is intriguing given the role that proteins containing the domain have in regulating development, particularly the role in the formation of secondary sexual characteristics. A primary example of this is bab, which has been identified as playing a role in the formation of sexually dimorphic abdominal pigmentation in Drosophila melanogaster (Kopp, et al., 2000). The pigmentation pattern is regulated by the different modification of bab levels along the abdomen in females versus males. The absence of this sexually dimorphic pigmentation (monomorphic) in the medfly suggests that, if a bab homolog exists in the medfly, it would be structured or regulated differently than it is in Drosophila. This is supported by evidence that bab levels are similar in females and males of some no-sexually dimorphic Drosophila
species (Kopp, et al., 2000). In addition to the role that it plays in abdominal pigmentation, bab levels are also regulated in the formation of the sexually dimorphic pattern of bristles and trichomes (cuticle projections) in D. melanogaster [Kopp, et al., 2000]. But in some melanogaster subgroup species, the bristle pattern and trichomes vary irrespective of bab levels. This suggests that the function of bab has evolved, even within Drosophila (Kopp, et al., 2000).

Further detailed comparisons of medfly genes that are homologs of these Drosophila genes will no doubt yield important information about the evolution of this gene family.
Appendix

Oligonucleotide Primer Sequences

The primer sequences provided are grouped according to their use in different sets of experiments.

### Amplification of Y-chromosome sequence, Y114

**Target:** Medfly male genomic DNA and first-strand cDNA

- **Y114F10:** 5'-TGCCAAAGCCTATCTTCCTCAGAAG-3'  
- **Y114B13:** 5'-GACCGTAAATGTCATTCCTCAGG-3'

### Suppression Subtractive Hybridization Primers

**Adaptor 1:**

5' CTAATACGACTCATATAGGGC

**Adaptor 2R:**

5' CTAATACGACTCATATAGGGAGCGTGGTCGCGGCCGAGGT-3'

### Amplification of Subtraction Products Isolated using SSH

**Target:** 3-5PMS.1

- **3-5PMS.1_F1:** 5'-CAGCCCGATCAGATTCTCTCGT-3'  
- **3-5PMS.1_B1:** 5'-GCTGCGAAAACGTAACTTTC-3'

**Target:** 4-7PMS.4

- **4-7PMS.4_F1:** 5'-CAGCCCGATCAGATTCTCTCGT-3'  
- **4-7PMS.4_B1:** 5'-GCTGCGAAAACGTAACTTTC-3'

**Target:** 3-5PFS.1

- **3-5PFS.1_F1:** 5'-GCCTTTGCGTCTTCTTTC-3'  
- **3-5PFS.1_B2:** 5'-CAGGGTAGATAACATITGGG-3'

**Target:** 3-5PFS.2

- **3-5PFS.2_F1:** 5'-CAAGTCGTCCGAGATTATTG-3'  
- **3-5PFS.2_B13:** 5'-CCATTAGACCCCATACGTCCTT-3'

**Target:** 3-5PFS.3

- **3-5PFS.3_F1:** 5'-CAAGTCGGTCGCAATATCATCC-3'  
- **3-5PFS.3_B14:** 5'-TGCCAAGAAMCACTGCTTACTC-3'

**Target:** 3-5PFS.4

- **3-5PFS.4_F2:** 5'-CCACAAAGCAGGCTTATTGC-3'  
- **3-5PFS.4_B10:** 5'-AAAGTCTGTATGTGCTGCGTGTTGCAC-3'
Rapid Amplification of cDNA Ends (RACE)

Manufacturer Primers
Universal Primer (UPM):
5’-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3’
Nested UPM: 5’-AAGCAGTGGTATCAACGCAGAGT-3’

Amplification of CePFS.2

Obtain sequence information
PFS.2_F2: 5’-GAAAAAAATGCTAAGCGG-3’
PFS.2_b1: 5’-GCATGGTGAAATTCTACTTCAC-3’

Amplification of CePFS.1B

5’ UTR region
RACEFa: 5’-ATCCCTACTTCCCGCTTGAC-3’

3’ region near polyA signal
MD1endFb: 5’-AGTTTTCTTTCAAAACACAT-3’

Obtain sequence information
MD1contigart_f1: 5’-CAACCCCAACCCATTACAGAAG-3’
MD1contigart_b2: 5’-TCGGACATTTGAGACGAATC-3’

Degenerate Primers Designed to Amplify BTB-POZ Domain Sequences
Target: Medfly female embryonic and adult first-strand cDNA
LSACSPY_For: 5’-YTGTCSGCYTGYTCSCCHfAYTT-3’
DFMYTGE_Rev: 5’-YTCRCCDGTRTACATRAARTC-3’

Primers for Sequencing Plasmid Clones
Template: pUC19, pCR Script (SK+), pCR2.1
M13(-20): 5’-GTAAAACGACGGCCAGT-3’
M13(-24): 5’-AACAGCTATGACCATG-3’
T3: 5’-ATTAACCCTCACTAAAG-3’
T7: 5’-AATACGACTCACTATAG-3’

Miscellaneous Primers
DdAI_F4 and DdAI_R4 (actin)
DdAI_F4: 5’-GTTGCTTTGGACTTTGAGCAGG-3’
DdAI_R4: 5’-AGAACACCAATCCAACAGGAG-3’

CNZwF and CNZwB (glucose-6-phosphate dehydrogenase)
CNZwF: 5’-CATTTGCATTTGAGCGAG-3’
CNZwB: 5’-CTCATATTTGCTGCTG-3’
References


IAEA. Research on medfly genetic sexing strains. Entomology Unit 1999 Annual Report


137


