UNIVERSITY OF HAWAI'I LIBRARY
DNA SEQUENCES DIFFERENTIALLY REPRESENTED IN MALES AND
FEMALES OF THE ORIENTAL FRUIT FLY BACTROCERA DORSALIS

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 (GENETICS)

DECEMBER 2002

By
Janice S. Lai

Dissertation Committee:
David S. Haymer, Chairperson
Rebecca L. Cann
Steven N. Robinow
Donald O. McInnis
James L. Brewbaker
© Copyright 2002
by
Janice S. Lai
Acknowledgments

Many thanks to the members of my dissertation committee for their endless patience and generosity. You humble me with your wisdom and kindness. Thanks especially to Dr. Donald McInnis and Dr. Steven Robinow for their help with experimental reagents and equipment. Most importantly, thanks to Dr. David Haymer for being a gentle mentor, and for calmly wading through all of the experimental (and occasionally personal) crises during this process.

Thanks to past graduate students from the department for all of their help and encouragement. A special thanks goes out to Dr. Joan Kuh and Dr. Deborah Shearman (Fly Goddesses) for their advice about experimental techniques and their empathy with the frustrations of research and graduate school. Thanks also to Dr. Terrence Lyttle and Dr. Alan Lau for the valuable research experience gained in their labs.

Thanks to the Biology Program and UH CTE for all of their teaching and mentoring opportunities. Mahalo to Dr. Ronald Mau (UH CTAHR) and Dr. Lyle Wong (Hawaii State Dept. of Agriculture) and the USDA for funding support.

A BIG thanks to my dear friends. To Lefty and the Godfather - you are the Goddess Role Models! Most importantly, to the Haymer Lab Fly Girls (and occasional Fly Guy), especially the lab’s Supreme Beings - Pia, Chrystie, and Keao - my (separated at birth) sisters ... You’ve helped me through the best and worst, and I would be lost without your joyous laughter, silly stories (which we’ve heard before, but pretend that we haven’t), crazy “field trips” and lunch hours (and hours and hours..). You are all my guardian angels.

This dissertation is dedicated with gratitude and love to my “peeps” - Dad, Mom, and Tammy - who have put up with me for all of these years. You taught me the value of hard work and set the bar high, but lifted me up on your shoulders when I couldn’t reach it on my own.
Abstract

The objective of this dissertation is the isolation of DNA sequences that are differentially represented in males and females of the Oriental fruit fly *Bactrocera dorsalis*, specifically by initiating a molecular characterization of Y chromosome sequences in this species. Cytological observations have established the presence of a diminutive Y chromosome in *B. dorsalis* males. To isolate DNA sequences from the Y chromosome, a special method of genomic DNA isolation known as Representational Difference Analysis (RDA) was utilized to obtain DNA sequences unique to the *B. dorsalis* male genome. Genomic DNA from *B. dorsalis* males served as the “tester” DNA and female genomic DNA as the “driver” DNA. Six distinct RDA products were obtained following two complete rounds of DNA hybridization and difference enrichment via the Polymerase Chain Reaction (PCR). One of these products (RDA product 1) was used to isolate a genomic DNA clone (3.1a) from a *B. dorsalis* male genomic DNA minilibrary. This sequence shows similarity to the reverse transcriptase of R1 retrotransposable elements. The presence of R1 elements in the Tephritid insects has heretofore been undetected, although these elements have been previously described in the genomes of other Dipteran species.

Oligonucleotide primers for PCR were designed for the 3.1a clone. These primers consistently produce different amplification patterns in PCRs of genomic DNA from *B. dorsalis* males vs. females. Amplification using male genomic DNA produces 325 bp and 2.6 kb products while only a 2.6 kb product is obtained from female DNA. The amplification products obtained with these primers are also produced in PCRs of genomic
DNA from *B. dorsalis* embryos and third instar larvae, suggesting the ability of this method to infer sex at pre-adult stages of the *B. dorsalis* life cycle. Similar amplification products have also been obtained in other Bactrocera species.

Both the 325 bp male PCR product and the 2.6 kb products have regions of sequence similarity to R1 elements. The 2.6 kb product contains a putative 1.7 kb open reading frame (ORF) encoding 583 amino acids. Three amino acid motifs found in Drosophila R1 element reverse transcriptases are present in comparable locations within the hypothetical ORF product. Both of these sequences are also repetitively represented in the *B. dorsalis* male and female genomes. However, the 325 bp male product produces some bands that are male specific when used as a probe for Southern blots of *B. dorsalis* male and female genomic DNA.

The amplification pattern produced by the 3.1a primers is consistent with what would be expected if the 2.6 kb and 325 bp PCR products originated from the *B. dorsalis* X and Y chromosomes, respectively. Thus, the cloned male-specific sequence recovered here is potentially useful both as a gateway into the relatively uncharacterized *B. dorsalis* Y chromosome and as a tool for the characterization of other aspects of the *B. dorsalis* genome.
Table of Contents

Acknowledgments ................................................................. iv
Abstract ............................................................................. v
List of Tables .................................................................... xiii
List of Figures ..................................................................... xiv
List of Abbreviations ........................................................... xvi

Chapter 1: Introduction

1.1 Dissertation objectives...................................................... 1
1.2 The Family Tephritidae...................................................... 2
  1.2.1 The Tephritid life cycle............................................... 2
  1.2.2 *Bactrocera dorsalis*, the Oriental fruit fly..................... 3
  1.2.3 Strategies for the eradication of Tephritid fly pests.......... 3
    1.2.3.1 The Sterile Insect Technique (SIT)......................... 4
    1.2.3.2 Prospective alternatives and improvements to SIT..... 4
1.3 Y chromosomes................................................................. 5
  1.3.1 Basic characteristics of Y type chromosomes............... 6
  1.3.2 The Ys and wherefores of Y chromosome evolution....... 7
1.4 Human Y chromosome structure and function
  1.4.1 PARs for the course.................................................... 12
  1.4.2 Role of the human Y chromosome in male sex
determination................................................................. 13
  1.4.3 Human Y chromosome heterochromatin...................... 14
1.5 Drosophila Y chromosome structure and function

1.5.1 The Drosophila Y and male fertility.......................... 16

1.5.2 The structure and composition of the Drosophila Y chromosome......................................................... 17

1.5.3 Sex determination in Drosophila and other Dipterans:

Y not?................................................................. 19

1.5.3.1 Drosophila sex determination............................ 19

1.5.3.2 Sex determination in other Dipterans..................... 23

1.6 Tephritid Y chromosome structure and function

1.6.1 Sex chromosomes of the Medfly Ceratitis capitata....... 25

1.6.2 Bactrocera sex chromosomes............................... 28

1.6.3 The Medfly Y and male sex determination............... 29

1.6.3.1 Sex-specific gene expression in the Medfly......... 30

1.6.4 The B. dorsalis Y and male sex determination............ 31

1.7 Retrotransposable elements and the Y chromosome........... 32

1.7.1 General characteristics of retrotransposable elements... 32

1.7.2 LTR vs. non-LTR retrotransposable elements.............. 33

1.7.3 R1 and R1 elements.......................................... 34

1.7.3.1 The prevalence of R1 and R2 elements in insects... 35

1.7.3.2 A model for R2 element transposition.................. 37

1.7.3.3 Phenotypes associated with R1 and R2 element

insertion in Drosophila........................................... 38

1.8 Transposable elements in the Tephritids C. capitata and B. dorsalis... 39
1.9 Genomic DNA hybridization as a means of isolating specific types of sequences

1.9.1 Past work in the isolation of Y chromosome specific sequences

1.9.2 Representational Difference Analysis (RDA)

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Fruit fly strains

2.1.2 Reagents and systems

2.1.2.1 Enzymes

2.1.2.2 Chemical reagents

2.1.2.3 System products

2.1.2.4 PCR oligonucleotide primers

2.1.2.5 DNAs and probes

2.2 Methods

2.2.1 Genomic DNA isolation from *B. dorsalis* embryos

2.2.2 Genomic DNA isolation from single *B. dorsalis* third instar larvae and Bactrocera adult flies

2.2.3 Genomic DNA isolation from multiple adult flies

2.2.4 Representational Difference Analysis (RDA)

2.2.4.1 Creating *B. dorsalis* male and female "representations"

2.2.4.2 RDA primary hybridization
2.2.4.3 RDA secondary hybridization ....................... 63
2.2.4.4 Cloning RDA products ............................ 64

2.2.5 λgt10 genomic DNA library construction and screening
2.2.5.1 Library construction ................................ 64
2.2.5.2 Library screening ................................... 65
2.2.5.3 Phage DNA purification and subcloning ........... 67

2.2.6 Southern blots, probe hybridization, and probe detection ... 68

2.2.7 PCR amplification from B. dorsalis embryos, third instar
larvae, pupae, and adult flies ............................... 69

2.2.8 DNA sequencing and analysis .......................... 71

Chapter 3: Results

3.1 Initial attempts to isolate B. dorsalis Y chromosome specific
sequences ......................................................... 72

3.2 Representational Difference Analysis (RDA) products ............ 73
3.2.1 Characterization of RDA product clones .................... 74
3.2.1.1 Analysis of clones from RDA products 4 and 5 ..... 74
3.2.1.2 Analysis of clones from RDA products 1, 2, 3,
and 6 .......................................................... 81

3.3 Further characterization of RDA product 1 ..................... 89
3.3.1 B. dorsalis Southern blots probed with minilibrary
subclones ....................................................... 91

3.3.2 Sequence analysis of minilibrary subclones ................. 91
3.3.3 Sequencing the remainder of the 3.1a clon .................. 106

3.4 B. dorsalis PCR assays with 3.1aRev primers .................. 112

3.5 Sequence analysis of 3.1aRevF2-B2 PCR product clones ......... 113

3.5.1 Genbank database search with the 3.1aRev PCR products ... 117

3.5.2 Alignment of 3.1aRevF2-B2 PCR product sequences .......... 124

3.5.3 Alignment of B. dorsalis female 3.1a PCR product sequence with other R1 reverse transcriptase sequences .... 137

3.6 B. dorsalis genomic DNA Southern blots probed with 3.1aRevF2-B2 PCR products .............................................. 144

3.7 3.1aRev primer survey of various stages in the B. dorsalis life cycle ... 144

3.8 Sequence analysis of 3.1aRev PCR products from B. dorsalis embryos and larvae ...................................................... 149

3.9 Surveying other B. dorsalis strains and Bactrocera species with 3.1aRev primers ..................................................... 149

3.10 3.1aRevF2-B2 PCR product clones from B. papayae and B. curcurbitae ................................................................. 158

Chapter 4: Summary and Discussion ........................................... 163

4.1 Representational Difference Analysis ............................. 165

4.2 Characterization of RDA product 1 .............................. 168

4.3 Characterization of the 3.1a clone ............................... 168

4.4 Sex-specific PCR products in B. dorsalis adults .................. 169

4.5 Sex-specific PCR products in other stages of the B. dorsalis life cycle. .................................................. 172

4.6 PCR products in other Bactrocera species ....................... 174
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>Bactrocera</em> genomic DNA samples</td>
<td>47</td>
</tr>
<tr>
<td>2. 3.1a clone oligonucleotide PCR primer sequences</td>
<td>50</td>
</tr>
<tr>
<td>3. RDA oligonucleotide adaptor sequences</td>
<td>58</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</td>
<td>Human Y chromosome</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>The Representational Difference Analysis (RDA) method</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>3.1a clone oligonucleotide PCR primer diagram</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>Agarose gel results of <em>B. dorsalis</em> RDA</td>
<td>76</td>
</tr>
<tr>
<td>5</td>
<td>RDA product 4 clone GenBank search results</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>RDA product 5 clone GenBank search results</td>
<td>82</td>
</tr>
<tr>
<td>7</td>
<td><em>B. dorsalis</em> Southern blots probed with RDA product 4 and 5 clones</td>
<td>87</td>
</tr>
<tr>
<td>8</td>
<td><em>B. dorsalis</em> Southern blot probed with RDA product 1</td>
<td>88</td>
</tr>
<tr>
<td>9</td>
<td><em>B. dorsalis</em> male minilibrary probed with RDA product 1</td>
<td>90</td>
</tr>
<tr>
<td>10</td>
<td>2.1a clone GenBank search results</td>
<td>92</td>
</tr>
<tr>
<td>11</td>
<td>3.1aFor sequence GenBank search results</td>
<td>97</td>
</tr>
<tr>
<td>12</td>
<td>3.1aRev nucleotide sequence</td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>3.1aRev sequence GenBank search results</td>
<td>101</td>
</tr>
<tr>
<td>14</td>
<td><em>B. dorsalis</em> Southern blot probed with 3.1a clone</td>
<td>105</td>
</tr>
<tr>
<td>15</td>
<td><em>B. dorsalis</em> male and female PCRs with 3.1a long template PCR primers</td>
<td>107</td>
</tr>
<tr>
<td>16</td>
<td>883 bp sequence from one end (3.1aFor) of the 3.1a clone</td>
<td>108</td>
</tr>
<tr>
<td>17</td>
<td>2031 bp sequence from another end (3.1aRev) of the 3.1a clone</td>
<td>110</td>
</tr>
<tr>
<td>18</td>
<td>3.1aRevF1-B1 and F2-B2 primer diagram</td>
<td>114</td>
</tr>
<tr>
<td>19</td>
<td><em>B. dorsalis</em> 3.1aRevF1-B1 and 3.1aRevF2-B2 PCR products</td>
<td>116</td>
</tr>
<tr>
<td>20</td>
<td><em>B. dorsalis</em> 325 bp male PCR product sequence alignment</td>
<td>118</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>21. B. dorsalis female 2.6 kb PCR product sequence</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>22. 325 bp male product GenBank search results</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>23. 2.6 kb female product GenBank search results</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>24. Alignment of B. dorsalis 325 bp male product, 2.6 kb female product, and 3.1aRev sequence</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>25. Alignment of B. dorsalis female 2.6 kb product sequence with R1 reverse transcriptases from other insect species</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>26. B. dorsalis Southern blots probed with the 325 bp and 2.6 kb PCR products</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>27. 3.1aRevF2-B2 PCR amplifications in B. dorsalis embryos, larvae, and pupae</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>28. Alignment of 325 bp PCR products from B. dorsalis embryos and adult males</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>29. Alignment of 2.6 kb PCR products from B. dorsalis embryos and adult females</td>
<td>152</td>
<td></td>
</tr>
<tr>
<td>30. 3.1aRevF2-B2 PCR products from B. dorsalis (Taiwan and Okinawa strains), B. papayae, B. carambolae, and B. curcurbitae</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>31. Alignment of 325 bp PCR products from B. dorsalis and B. papayae adult males</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>32. Alignment of 325 bp PCR products from B. dorsalis adult males and B. curcurbitae adult males and females</td>
<td>161</td>
<td></td>
</tr>
</tbody>
</table>
List of Abbreviations

Frequently used acronyms

BLAST: Basic Local Alignment Search Tool
ORF: open reading frame
PAR: pseudoautosomal region
RDA: Representational Difference Analysis
SIT: Sterile Insect Technique
TDF: testis-determining factor
UV: ultraviolet

Chemicals

Amp: ampicillin
BCIP: 5-bromo-4-chloro-3-indolyl-phosphate, toluidine salt
CHISM: chloroform-isoamyl alcohol, 24:1
dATP: deoxyadenosine triphosphate
dCTP: deoxycytosine triphosphate
dGTP: deoxyguanosine triphosphate
ddH₂O: double-distilled water
Dig: digoxygenin
DNA: deoxyribonucleic acid
dNTP: deoxyadenosine triphosphate, deoxythymidine triphosphate,
deoxycytosine triphosphate, and deoxyguanosine triphosphate
dTTP: deoxythymidine triphosphate
Chemicals (continued)

EDTA: ethylenediaminetetraacetic acid
EPPS: N-(2-hydroxyethyl) piperazine-N’-(3-propanesulfonic acid)
HCl: hydrochloric acid
LB: Luria broth
MOPS: 3-(N-morpholino) propanesulfonic acid
NBT: Nitro blue tetrazolium chloride
NaCl: sodium chloride
NaOH: sodium hydroxide
SSC: sodium chloride-sodium citrate
SDS: sodium dodecyl sulfate
TB: Terrific broth
TBE: tris-borate EDTA
TE: tris-EDTA
Tris: tris(hydroxymethyl)aminomethane
Xgal: 5-bromo-4-chloro-3-indolyl-β-D-galactoside

Measurements

bp: base pair
°C: degrees Celsius
cc: cubic centimeter
cm: centimeter (10⁻² meter)
g: gram
kb: kilobase (10³ bases)
Measurements (continued)

M: moles per liter
Mb: megabase (10^6 bases)
min.: minutes
ml: milliliter (10^-3 liters)
mM: millimolar (10^-3 molar)
mg: milligram (10^-3 grams)
mU: milliunit (10^-3 units)
nt: nucleotide
pfu: bacteriophage plaque forming units
pmol: picomole (10^-12 moles)
pH: concentration of hydrogen ions
rpm: rotations per minute
µg: microgram (10^-6 grams)
µl: microliter (10^-6 liters)
V: volts
w/v: unit of weight per unit of volume
Symbols

p: plasmid vector

♂: male

♀: female

A: adenine

T: thymine

C: cytosine

G: guanine
Chapter 1

Introduction

1.1 Dissertation objectives

The overall objective of this dissertation is to isolate and characterize Y chromosome-specific DNA sequences from the Oriental fruit fly Bactrocera dorsalis using the RDA technique of genomic DNA hybridization. The isolation of Y-specific sequences in this species is significant for three major reasons. First, there is a general lack of information on the B. dorsalis genome. Therefore, any newly characterized sequences would contribute to the understanding of the genome in this species. Secondly, there is also a lack of information on the Y chromosome and the sex determination mechanism, particularly the role of the Y chromosome in sex determination in this species. The isolation of Y-specific sequences will provide an entryway into the complete characterization of the Y chromosome, and will also identify the existence of any male-determining regions on the Y. Third, because B. dorsalis is a significant agricultural pest in many parts of the world, male-specific (especially Y chromosome-specific) sequences are of particular interest as tools in the enhancement for the SIT technique as a more effective means of population control for B. dorsalis and other Tephritid species.
1.2 The Family Tephritidae

The Family Tephritidae, known as the “true fruit flies”, represents a diverse cast of characters from a wide variety of geographic origins. The family encompasses approximately 4000 different species, including some notorious for their devastating economical impact on agricultural crops. Tephritid fly species are characterized by the female ovipositor, which extends posteriorly and serves to penetrate the skin of host fruit and deposit fertilized eggs. In some species, the female ovipositor is used for egg deposition in flowers instead of fruit. Developing larvae then utilize the fruit or flower as a food source and for development through three instar stages. Fruit associated species leave the host fruit to pupariate in the soil, while flower associated species pupariate in the host flower (White and Elson-Harris 1992).

1.2.1 The Tephritid life cycle

Tephritid flies possess a life cycle similar to that of Drosophila, although the duration of each stage varies amongst different Tephritid species. Adult females deposit fertilized eggs into ripening fruit. For species such as B. dorsalis, the deposited embryos undergo development from 1-20 days, transforming into larvae. The larval stage lasts from 9-35 days, during which the larva undergoes three distinct periods of larval molting and development, called instars. The mature larvae develop into pupae, and develop as such for 10-30 days; adult flies emerge thereafter (reviewed by White and Elson-Harris 1992).
1.2.2 Bactrocera dorsalis, the Oriental fruit fly

The Oriental fruit fly Bactrocera (Dacus) dorsalis (Hendel) is the type species for a complex of approximately 52 different closely related Asian fruit fly species (Drew and Hancock 1994). The species Bactrocera carambolae and Bactrocera papayae, both members of the B. dorsalis complex, are mentioned later in this dissertation text. Once classified in the Tephritid genus Dacus, traditional taxonomy is particularly difficult within this complex when relying upon morphology alone; it has only been since the advent of various methods of genetic analysis, pheromone research, and studies of acoustic signals that species lines have been more clearly drawn (White and Elson-Harris 1992). The pronounced presence of B. dorsalis in southeast Asia has significantly impacted the tropical fruit industry in countries such as Thailand, where there is cultivation of host plants such as mango, guava, and rose apples (Baimai et al. 1995). Other known hosts include a wide variety of other fruit-bearing species, as well as coffee, chili pepper (Capsicum annuum), and orchids. B. dorsalis populations also presently exist in the Pacific (White and Elson-Harris 1992).

1.2.3 Strategies for the eradication of Tephritid fly pests

Because of their severe economic impact, vigorous efforts have been made to devise effective methods for eradicating Tephritid fly pests. These attempts at eradication are severely hampered when populations of flies infest regions that are a considerable distance away from their native area, prompting neighboring countries to enforce strict quarantine regulations for the import of commercial fruit from areas with existing fly populations to prevent their spread. However, quarantines are often unable to
halt the migration of flies, which are capable of circulating from regions such as South America to the United States on a regular basis. To combat infestations, techniques such as the release of sterile males have been employed (White and Elson-Harris 1992). Strict quarantine policies and eradication efforts place a significant burden upon the economics and work force of both fruit producing and consuming countries, making the development of effective prevention methods an even greater priority.

1.2.3.1 The Sterile Insect Technique (SIT)

The most popular pest control method with regard to Tephritid species is the environmentally friendly “Sterile Insect Technique” (SIT). The most effective SIT method to date involves the mass rearing and release of sterile males flies, eliminating any problems with “preferential mating among the released flies”, which could occur when releasing both sterile males and females. The strategy of exclusively rearing and releasing sterile males improves overall program monitoring by reducing the potential of additional damage to fruit by the release of female flies and minimizes the chance of releasing partially sterilized or unsterilized females (reviewed by Franz and Kerremans 1994).

1.2.3.2 Prospective alternatives and improvements to SIT

Much of the success of SIT lies in the efficient production of flies of a single sex. Presently, artificially reared fly populations are genetically manipulated in such a way that they can be induced to produce only male offspring. However, newer methods of genetic sexing are being developed for use in the mass rearing species such as the Medfly
and Oriental fruit fly for the purposes of population control. One such technique is RIDL ("Release of Insects carrying a Dominant Lethal"), which was demonstrated in Drosophila (Thomas et al. 2000). Two different transgenic Drosophila lines were created to test this technique - one carrying a construct with a cytotoxic gene that is inducible in females and another where females can be induced to express a dosage compensation gene usually active only in males. In both cases, females were completely eliminated from the population. When males from the first transgenic line were mated to nontransgenic females, no female progeny were produced, eliminating the need to sterileize males prior to their release in the wild.

Germline transformation via transposable elements is another possible approach for the genetic manipulation of these species. Handler and McCombs (2000) showed that the transposable element piggyBac, originally isolated in the moth Trichoplusia ni, is capable of germline transformation in B. dorsalis. The piggyBac vector has also been previously demonstrated to be capable of germline transformation in the Medfly (Handler et al. 1998). Medflies have also been successfully transformed by the Minos element from Drosophila hydei (Loukeris et al. 1995).

### 1.3 Y chromosomes

At the end of the 19th century, H. Henking, while observing spermatogenesis in the Hemipteran bug Pyrrhocoris apterus, noted the presence of a darkly-staining body in males that behaved differently than other chromosomes during meiosis. He referred to this chromosomal body as the "X" to reflect its mysterious nature. It was, from then on, known as the "X chromosome" (reviewed by Solari 1994). During the early 1900s, E.B.
Wilson observed differences in the chromosomes amongst the sexes of various Hemipteran insects (a group that includes cicadas, aphids, and the “true bugs”). Specifically, he observed two possible chromosome types in males. In the first type, males possessed a heteromorphic pair of chromosomes that, during meiosis, segregate into different daughter cells. In the second type, males had one of these heteromorphic chromosomes, while females had two. Wilson later referred to the heteromorphic chromosomes collectively as the “sex chromosomes”, and individually as either the “X chromosome” or the “Y chromosome” (reviewed by Mittwoch 1967). It has since been learned that, in some species, males are not the heterogametic sex, nor are the X and Y chromosomes the only heteromorphic chromosomes. In animals such as birds, females are the heterogametic sex, possessing a Z and a W chromosome. These chromosomes are heteromorphic, with the W chromosome being largely heterochromatic and smaller in size than the Z chromosome (reviewed by Solari 1994).

1.3.1 Basic characteristics of Y type chromosomes

Y type chromosomes are intriguing for several reasons. First, it is often the case with heteromorphic sex chromosomes such as the Y that the conventional rules of chromosome pairing and recombination, as exhibited by the autosomes, do not apply. Generally, there is a marked difference in the relative sizes and gene compositions of the sex chromosomes, so much so that, in most cases, only select regions of these chromosomes still maintain proper pairing and crossing over during meiosis. Second, in species where Y type chromosomes are found, they are typically inherited exclusively by
one sex. Third, despite the fact that they are often largely devoid of "traditional" genes, such as allelic counterparts of X-linked genes, Y chromosomes do often carry genes that are critical for functions such as sex determination and/or fertility. Finally, despite being typically highly divergent from its X chromosome counterpart, primarily because of the lack of recombination, Y type chromosomes somehow persist in many genomes.

1.3.2 The Ys and wherefores of Y chromosome evolution

The pronounced differences in sequence and gene composition between the well-described heteromorphic sex chromosomes in species such as humans and Drosophila collectively lead to questions such as: Why does such disparity exist between these two chromosomes? Theories abound in attempts to explain sex chromosome variability, yet none offer universal answers. The first model for sex chromosome variation, addressing in particular the degeneration of chromosomes such as the Y, is commonly known as Muller's ratchet. H.J. Muller theorized that the X and Y chromosomes originated as homologous chromosomes (reviewed by Charlesworth 1978). Muller's hypothesis asserts that at some point the Y failed to recombine with the X and that the continued failure to undergo recombination reduced the Y chromosome to a "permanently heterozygous" state (reviewed by Charlesworth 1978). The lack of recombination would
permit the fixation of mutations occurring on the Y, some of which cause a loss-of-
function. While in normally recombining chromosomes, detrimental mutations are
eliminated by recombination and selection, Muller’s ratchet predicts that the overall
number of mutations on a nonrecombining chromosome will continually increase within
a population of fixed size in the absence of recombination. With time, the number of
individuals with “mutant-free” chromosomes will decrease until they are permanently
lost from the population. Left are individuals with chromosomes containing one or more
mutant alleles. Muller theorizes that this is a continuous process – the class of
individuals with chromosomes containing one mutant allele will eventually be lost,
followed by the class with two mutant alleles, and so on. The permanent loss of entire
“classes” of individuals and an increase in established mutations on the chromosome turn
the ratchet (reviewed by Charlesworth 1978). This would eventually lead to the loss of
functioning wildtype genes and a gradual sequence divergence between the sex
chromosomes, ultimately leading to a completely inactive Y chromosome (reviewed by
Gordo and Charlesworth 2001).

Rice (1987), however, asserts that Muller’s ratchet cannot be solely responsible
for Y chromosome degeneration, citing two major principles of the ratchet that do not
appear to hold universally true. The first is that Y chromosome inactivation occurs in a
nonspecific fashion and that the loss of Y gene activity is compensated for by the
adjustment of X chromosome gene activity (i.e., X chromosome dosage compensation).
Rice asserts that this is not universally true, as dosage compensation is not found in birds
or butterflies. Furthermore, dosage compensation has been shown to be gene specific in
many cases. Secondly, for the influence of Muller’s ratchet to be significant, the Y
chromosome must be of a relatively large size. Rice cites an example in guppies, a species with small “differential segments” which would be assumed to be unaffected by the ratchet, that exhibit some degree of degeneration. Rice’s alternative hypothesis for chromosome degeneration is called the “genetic hitchhiking model”. In this model, mutations will potentially generate “deleterious alleles” on nonrecombining Y chromosomes. The model proposes that a change in the environment might favor some new “beneficial allele” on the Y. This beneficial allele will likely occur at low frequency on a “favorable background”. Rice hypothesizes that it is unlikely that the most favorable background will be one with the lowest number of mutations. When selection favors Y chromosomes with the beneficial allele in a favorable background over those with the allele in a background with the lower number of mutations, the allele can become established on the Y. With these conditions, Y chromosomes carrying the lowest number of mutations will be lost and it is possible for a single Y chromosome to be established in a population. Thus, deleterious mutations carried on this Y chromosome will also become fixed.

Peck (1994), in his “Ruby in the Rubbish” model, also attributes the degeneration of the Y to the introduction of beneficial mutations (those that confer some type of selective advantage). Peck’s model asserts that beneficial mutations will more likely be established in organisms reproducing sexually than those reproducing asexually. For a beneficial mutation to establish itself within an asexually reproducing species, it must arise within the fittest individuals (i.e., those with the fewest deleterious mutations), or it
will be lost. In sexually reproducing organisms, recombination creates the potential for beneficial mutations in separate individuals to be united in the same genome, increasing the likelihood of homozygosity and their perpetuation into later generations. Thus, even if beneficial mutations occur on the Y chromosome, they have a considerably lower likelihood of becoming established, due to the lack of recombination on the Y.

Orr and Kim (1998) proposed a model building on Rice’s premise that recombination is favorable for adaptive evolution of chromosomes. Because the Y chromosome is nonrecombining, mutations are less likely to become established. If some of these mutations are favorable to the individual but do not become established, the adaptive ability of the Y chromosome will be significantly less than that of the X. Selection will eventually favor increasing the expression of X-linked genes over the expression of Y-linked genes which do not recombine.

A number of examples of Y chromosome degeneration have been described in the literature. The *D. miranda* neo-X and neo-Y chromosomes are the products of fusions between autosomes and the X or Y. The neo-Y does not recombine, thus exhibiting degeneration that would be predicted for chromosomes in the absence of selective forces maintaining wildtype sequences. Bachtrog and Charlesworth (2002) showed through sequence analysis that genes found on both the neo-X and neo-Y vary in sequence from each other, and that none of the nucleotide variations that were observed were common to both chromosomes, as would be expected if there was a lack of crossing over between the two.
Steinemann and Steinemann (1992), in a study of the *D. miranda* larval cuticle protein (*Lcp*) gene cluster, showed via *in situ* hybridization an accumulation of the retrotransposons TRIM and TRAM and eight distinct inserted DNA sequences (ISYs) in the *Lcp* region on the neo-Y chromosome that were not present within the neo-X *Lcp* gene cluster. The gene *Lcp*-4, located within the neo-Y *Lcp* gene cluster, is also inactive due to a large deletion within its protein-encoding region.

In the dioecious plant *Silene latifolia*, females are XX while males are XY. Filatov et al. (2000) showed in their comparisons of the genes *SLX*-1 and *SLY*-1, a gene pair found on the X and Y chromosomes, respectively, that they do not share any of 54 nucleotide polymorphisms found. They also observed evidence of recombination occurring in the *SLX*-1 gene, but not in *SLY*-1.

Although a broadly applicable model to explain Y chromosome evolution is apparently still lacking, a considerable amount of information about Y chromosome structure and function has been amassed. The following sections review some of this information, focusing on humans and the Drosophilid and Tephritid flies.
1.4 Human Y chromosome structure and function

1.4.1 PARs for the course

The euchromatic human X chromosome is relatively large in size and thought to contain up to 4000 genes, while the heterochromatic Y chromosome is considerably smaller with only about 30 genes identified to date (reviewed by Graves et al. 1998). While the bulk of the Y chromosome consists of Y-specific sequences that do not interact with the X, Koller and Darlington found that a region of the Y recombined with a homologous region on the X during the pachytene subdivision of late prophase I. Genes located within this region mimic autosomal inheritance rather than display sex-linked inheritance, and the region was thus described as being “pseudoautosomal” by Burgoyne in 1982 (reviewed by Ellis and Goodfellow 1989). On average, one-third of the length of the Y chromosome is involved in human X-Y chromosome pairing (reviewed by Mohandas et al. 1992).

The human X and Y chromosomes contain two pseudoautosomal regions (PARs), one at each of their ends. The 2.6 Mb PAR1 is located at the tip of the X short arm (Xp) and Y short arm (Yp) and is known to contain 11 genes, while PAR2 is 320 kb in size and located at the tip of the X and Y chromosome long arm (Xq and Yq, respectively), and is known to contain four genes (reviewed by Graves and Shetty 2000). Outside of the PARs, the Y chromosome does not recombine with the X. The sequence within this nonrecombining region appears to have diverged significantly from X, except that there are at least 19 genes within this region of the Y that have counterparts on the X.
Interestingly, there is a lack of correspondence between the X chromosome and Y chromosome map order of genes. Genes within the nonrecombining regions of the X and Y are believed to be helpful clues in determining the origin of the sex chromosomes. The suggestion has been made that the human X and Y began to diverge when a series of inversions occurred on the Y chromosome (reviewed by Lahn and Page 1999).

It appears that PARs are not universally essential for maintenance of the sex chromosomes. In marsupials, PARs are entirely absent from the sex chromosomes, and recent studies have confirmed that marsupial sex chromosomes do not pair or recombine during meiosis, apparently without impacting sex chromosome segregation during meiosis or male fertility (reviewed by Graves et al. 1998).

**1.4.2 Role of the human Y chromosome in male sex determination**

*SRY* (sex-determining region Y) is a gene located on the short arm of the human Y chromosome (Yp, Figure 1) and is believed to be the testis-determining factor (TDF). *SRY* was suspected of being the human TDF when XX males were found to have Y-specific markers translocated during meiosis to the paternal X chromosome, apparently as a result of faulty X-Y recombination (Koopman 2001). *SRY* is an intronless gene that encodes an 80 amino acid protein containing a sequence motif called the "HMG box" ("HMG" = "High Mobility Group", a protein group to which this domain is similar) and is believed to be a transcription factor involved in DNA binding and bending. Human XY females were found to have mutations in the *SRY* HMG box (Graves 1998). Two possible functions of SRY are currently being considered. The presence of a HMG box motif suggests that a possible function of SRY is to assist in gene transcription by
manipulating chromatin. The second possibility is that SRY forms a transcriptional protein complex with other proteins. The transcription factor Sox9 is currently considered the most promising target gene candidate. Located on 17q24, Sox9 has already been implicated in mammalian male determination, causing testis development ectopically expressed in XX individuals (reviewed by Koopman 2001). Sox9 has also been shown to be expressed in testes of XY male mouse embryos (reviewed by Graves 1998).

1.4.3 Human Y chromosome heterochromatin

While there are still a number of active genes on the human Y chromosome, the bulk of the Y is composed of heterochromatin. The general difficulties of characterizing heterochromatic DNA have impeded its characterization. To date, a number of repetitive DNA families have been characterized on the human Y chromosome. Among these include the DYZ1 family, which consists of repeated sequences that occur in 3.4 kb HaeIII fragment increments. The bulk of the repeats is found in the Y chromosome heterochromatin, and is estimated to occupy 25-30% of the Y and 50% of the large
Figure 1. An illustration of the human Y chromosome, showing the locations of the two pseudoautosomal regions PAR1 and PAR2 and SRY, the potential testis-determining factor (TDF), as depicted by Lahn and Page (1999) and Tilford et al. 2001). The majority of the human Y is nonrecombining, with a largely heterochromatic long arm (Yq). (Note: figure is not drawn to scale.)
heterochromatic block found in the Y long arm (reviewed by Willard 1985). The DYZ2 family consists of repeats of a 2.4 kb repeat found when HaeIII digestions of human genomic DNA are performed. It is estimated that 10-20% of the Y is composed of DYZ2 repeats (reviewed by Willard 1985). Other repetitive element families include DYZ3, which consists of short (340 or 680 bp) sequences found in larger EcoR1 fragments of pericentric Y DNA; the DXYZ1 family, found on both the X and Y chromosomes (reviewed by Daiger and Chakraborty 1985); the Kpn family, consisting of long repeats that have a relatively small representation on the Y; and alpha satellite DNA, consisting of repetitive pericentric sequences that are 340 bp in size (reviewed by Willard 1985).

1.5 Drosophila Y chromosome structure and function

1.5.1 The Drosophila Y and male fertility

The Drosophila Y is of particular interest because it does not participate in male sex determination. However, it has been shown to be essential for male fertility. Bridges first found that Drosophila XO males were sterile. Stern also observed that males deficient for even a portion of the submetacentric Y chromosome were sterile. He later concluded that both the Y short (YS) and long (YL) arms contain male fertility factors, and that duplications of one factor or complex of factors in one arm would not compensate for a deficiency of factor(s) in the other (reviewed by Hess and Meyer 1985). There are six known male fertility genes, as first identified by Brousseau: ks-1 and ks-2, located on the D. melanogaster YS, and kl-1, -2, -3, and -5, found on YL (reviewed by Hennig 1993). Goldstein et al. (1982) first showed that kl-2, kl-3 and kl-5 encoded high molecular weight proteins, and that deletion of the chromosomal regions containing
either \( kl-2 \), \( kl-3 \) or \( kl-5 \) result in the loss of dynein proteins in the sperm axoneme. Carvalho et al. (2000) confirmed that the fertility factors \( kl-2 \) and \( kl-3 \) encode dynein heavy chain proteins involved in sperm motility by constructing sequence alignments between unidentified genomic “scaffold” sequences isolated from the Drosophila Genome Project and those sequences already present in the GenBank database from the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov). The fact that all the known Y-linked male fertility factor genes encode proteins involved in sperm motility clearly indicate that loss of sperm motility is the primary reason for sterility in males lacking part or all of the Y chromosome. Carvalho et al. (2000) also theorize that sperm competition is the primary force driving Drosophila Y chromosome evolution. As evidence supporting their hypothesis, they cite the prevalence of male fertility factors exclusively involved in sperm motility, with the exception of very few other genes, residing on the Y. Carvalho et al. hypothesize that the influence of sperm competition would help to explain why the Y chromosome has evolved a number of motor proteins associated with sperm locomotion that are absent on the X.

1.5.2 The structure and sequence composition of the Drosophila Y chromosome

Besides the male fertility factors, other genes present on the Drosophila Y include ribosomal rRNA genes arrays and the \textit{Suppressor-of-Stellate [Su(Ste)]} locus, located on the Y long arm (Hardy et al. 1984). \textit{Su(Ste)} suppresses the \textit{Stellate} phenotype, the formation of proteinaceous crystals in Drosophila primary spermatocytes (reviewed by Livak 1990). Carvalho et al. (2000) also identified a new gene, \textit{PRY (Polycystine-related}}
Y), on the Y, located in the vicinity of kl-3. PRY is similar in sequence to the human polycystine gene and in turn, the sequence for a sea urchin gene encoding a receptor for egg jelly found in sperm. The function of the PRY gene has yet to be determined.

The majority of the Drosophila Y chromosome is composed of heterochromatin. Junakovic et al. (1998) identified the presence of a number of different transposable elements within the Drosophila Y heterochromatic regions, including multiple copies of representatives from the gypsy, copia and mdg1 element families through comparative Southern blot analyses of D. melanogaster and Drosophila simulans genomic DNA and in situ hybridizations of their Y chromosomes. Terrinoni et al. (1997) utilized the same approach to show the existence of copies of the non-long terminal repeat (LTR) elements G and jockey on the Drosophila Y. The Drosophila Y centromeric region is also a haven for repetitive sequences that are apparently been derived from the non-LTR retrotransposons Het-A and TART (Agudo et al. 1999). Carvalho et al. (2000) also note the frequent presence of a number of different transposable element sequences, particularly reverse transcriptase sequences, in the Drosophila genomic scaffold sequences screened during their identification and characterization of Y-linked male fertility factors. Interestingly, in situ hybridizations of Drosophila miranda chromosomes with the LTR-retrotransposon TRAM sequence has shown that it is enriched on the neo-Y chromosome but completely absent from its corresponding neo-X (X2) chromosome (Steinemann and Steinemann 1997).
1.5.3 Sex determination in Drosophila and other Dipterans: Y not?

Many Dipteran insects all share a basic overall mechanism for sex determination, yet not all of these mechanisms involve a Y chromosome. All begin with a "primary signal" that differs between males and females, which triggers the differential expression of other "key genes". The expression of these genes determines which of two pathways will be followed, each a cascade carried out by a subset of developmentally regulated genes, and results in the creation of male or female secondary sexual characteristics. However, it appears that different sets of regulatory genes have evolved to control different aspects of sex determination and there have been few homologs found between species (reviewed by Schütz and Nöthinger 2000). The sex determination mechanism in *Drosophila melanogaster* remains the most thoroughly described to date.

1.5.3.1 Drosophila sex determination

As previously stated, the Drosophila Y chromosome is unnecessary for male determination, but contains sperm motility genes important for male fertility (Carvalho et al. 2000). Instead, genes on the X chromosome and autosomes are implicated in the primary sex determination signal. In both somatic and germline cells, a primary signal is generated early in development, just before the formation of the blastoderm in embryogenesis. The "X:A" ratio, or the proportion of X chromosomes to autosomes generates this primary signal. An X:A ratio of 1.0 (two X chromosomes: two sets of autosomes) is the signal for female sex determination, while a ratio of 0.5 (one X
chromosome: two sets of autosomes) is the signal for male determination. In somatic cells, the X:A ratio directly affects the expression of the Sex-lethal (Sxl) gene. An X:A ratio of 1.0 will cause the expression of a functional SXL protein; thus, the expression functional SXL takes place only in females. The loss of Sxl expression has been demonstrated to be lethal in females, while the gain of Sxl expression is lethal in males (reviewed by Cline and Meyer 1996).

The genes controlling the determination of the X:A ratio all encode transcription factors. They can be classified into three distinct groups: "numerator" genes, "denominator" genes, and "maternal elements". The X-linked numerators include the sisterless genes (sisA, sisB/scute, sisC) and runt (run). The numerator protein products have been shown to bind to the Sxl early promoter Sxlpe following the establishment of a female X:A ratio. This prompts the production of "early SXL" protein. Sxlpe transcription stops and the Sxl maintenance promoter Sxlpm is transcribed in both sexes. Sxl then controls the splicing of its own transcript to create sex-specific SXL proteins. In females, the form of SXL protein produced is capable of perpetuating female-specific splicing of the Sxl transcript, while males lack early SXL and produce a truncated SXL protein that is incapable of autoregulation. Because of their X chromosome linkage, females usually have two copies of the numerator while males have only one. The presence of more than one numerator gene copy has also been shown to be lethal in XY zygotes, triggering the expression of Sxl (reviewed by Schütt and Nöthinger 2000).
The only known denominator gene is deadpan (dpn), located on the Drosophila chromosome 2. Like the numerator genes dpn encodes a protein containing a helix-loop-helix (HLH) motif that binds the Sxl promoter, but is thought to act as a negative regulator of Sxl gene activity in males. Altering the X:A ratio by increasing copies of the numerator sisb while decreasing copies of dpn is lethal in males (increasing the X:A ratio), while decreasing copies of sisb while increasing copies of dpn is lethal in females (decreasing the X:A ratio). Also, the absence of dpn expression in males results in the accumulation of SXL, resulting in lethality (Younger-Shepherd et al. 1992).

The maternal effect genes are those that are acquired through maternal inheritance and ensure the correct transduction of the X:A ratio signal to Sxl. Two of these genes, daughterless (da) and hermaphrodite (her), encode proteins that positively regulate Sxl gene expression. DA is a basic helix-loop-helix (bHLH) protein that forms heterodimers with SISB (also an HLH protein) in the zygote to activate transcription. HER is a zinc finger protein whose function is yet unknown, but appears to be important for controlling the expression of proteins further along the female determination cascade. Extramacrochaetae (emc) and groucho (gro) produce HLH proteins that negatively regulate Sxl expression. The proteins EMC and GRO downregulate Sxl expression by binding to DA and DPN respectively, thus is preventing them from activating transcription (reviewed by Cline and Meyer 1996).

SXL contains two RRM (RNA Recognition Motif) domains and activates the expression of the gene transformer (tra) in sex determination as well as participates in the control of dosage compensation and oogenesis. TRA works in concert with TRA2, the protein product of the constitutively active gene transformer2 (tra2).
of the Muscles of Lawrence (MOL), a pair of large dorsal muscles specifically found in
the fifth segment of the male abdomen with small counterparts found in the same location
within the female abdomen, is controlled by \textit{tra} and \textit{tra-2} (Lawrence and Johnston 1984,
Taylor 1992). These genes also control the splicing of \textit{doublesex (dsx)} in females. Males
normally lack both \textit{sxl} and \textit{tra} activity, creating a pattern of \textit{dsx} transcript splicing
different to that in females. Thus, \textit{dsx} is spliced in a sex-specific manner. This results in
two different versions of the DSX protein - DSXF in females and DSXM in males – that
differ from each other at their carboxy termini. Both DSXF and DSXM act as
transcription factors to regulate the expression of genes downstream in their respective
sex determination pathway (reviewed by Schüt and Nöthinger 2000).

While sex determination in Drosophila appears to be controlled by this primary
gene cascade, other genes acting in alternate pathways that also contribute to sexual
development and sex-specific behavior. One such gene is \textit{fruitless (fru)}, which directly
affects male mating behavior. Males that are mutant for \textit{fru} either exhibit altered
behavior during the latter stages of courtship, or these courtship stages are missing
altogether – that is, males will court females, but not copulate with them. This alteration
in behavior accounts for the lack of success of \textit{fru} males at mating, and essentially makes
them sterile. More interestingly, \textit{fru} males will also court other males (reviewed by Hall
1994). The expression of \textit{fru} is controlled by \textit{tra} and \textit{tra-2}, and its transcript is spliced in
a sex-specific manner. The $fru$ protein products are also similar in sequence to the BTB-ZF family of transcription factors (Ryner et al. 1996). The $fru$ gene product is implicated in multiple aspects of the development of the MOL in males, specifically in directing myoblasts to the intended site for the MOL and in other aspects of MOL development that have yet to be determined (Taylor and Knittel 1995).

The sex determination mechanism in the germline appears to differ greatly from that in somatic cells. The primary signal is still the X:A ratio, yet genes such as $Sxl$, $dsx$, and $tra$, while found in these cells, do not seem to be crucial for germline sex determination. Furthermore, somatic numerator genes do not control germline $Sxl$ gene expression, suggesting that different genes are responsible for interpreting the X:A ratio in this tissue. $Sxl$ has been shown to be important for the control of meiotic recombination in females (reviewed by Schütz and Nöthinger 2000).

1.5.3.2 Sex determination in other Diptera

Significantly less is known about mechanisms of sex determination in other Dipteran insects. Like Drosophila, a primary signal determines which of two developmental cascades will be chosen. The primary signal in the housefly Musca domestica is the male-determining gene ($M$) that, in different strains, can be found on different chromosomes, including the X, Y, and four of the 5 autosomes. The presence of at least one copy of $M$ is enough to inactivate the female-determining gene ($F$), resulting in male development. In the absence of $M$, $F$ prompts female determination. Two known types of $F$ mutations have been shown to behave much in the same way as $Sxl$ mutations – the dominant $F$ mutation ($F^D$), mirroring a gain-of-function of $Sxl$, promotes female
development, while the recessive $F$ mutation ($F^{man}$) mimics a loss-of-function of $Sxl$, promoting male development (reviewed by Hilfiker-Kleiner et al. 1994). Unlike Drosophila, $M. domestica$ germline sex determination follows that of the somatic cells and is "completely independent of the genotypic sex of the germ cells". In the presence of male somatic cells, even pole cells containing $F^{D}$, the dominant female-determining form of $F$, will form sperm. Pole cells homozygous for the male-determining gene $M$ will form eggs when in the presence of female somatic cells (Hilfiker-Kleiner et al. 1994). A gene similar in sequence to Drosophila $Sxl$ was also found in $M. domestica$ but does not appear to be sex-determining in this species (reviewed by Schüt and Nöthinger 2000).

In the blowfly $Chrysomya rufifacies$, the primary signal for sex determination does not originate from the embryo, but is instead maternally derived. There are two types of $C. rufifacies$ females: thelygenic females, which produce only female offspring, and arrhenogenic females, which produce only male offspring. Although both types of females appear alike in their morphology and cytology ($2n = 12$), the thelygenic females are heterozygous for $F'$, the dominant or epistatic form of a female sex determining element found on chromosome 5, while arrhenogenic females are homozygous for $f$, the recessive form of the element. A gene was found in $C. rufifacies$ that was similar in sequence and in gene structure to Drosophila $Sxl$ (Müller-Holtkamp 1995). However, it does not appear to undergo sex-specific splicing.
In addition to *M. domestica* and *C. rufifacies*, a gene with sequence similarity to *Sxl* has been isolated in the phorid fly, *Megaselia scalaris* (Saccone et al. 1998). However, like *M. domestica* and *C. rufifacies*, this gene does not appear to function in sex determination in this species. Genes similar in sequence to *dsx* that have been characterized in *M. domestica* and *M. scalaris* appear to be expressed in a sex-specific manner (reviewed by Schüt and Nöthinger 2000).

### 1.6 Tephritid Y chromosome structure and function

#### 1.6.1 Sex chromosomes of the Medfly *Ceratitis capitata*

Cytological techniques have been employed in the examination of the sex chromosomes of the Mediterranean fruit fly *Ceratitis capitata*. In Medfly mitotic chromosomes the X chromosome has been described as being subtelocentric and the longest of all of the chromosomes, while the Medfly Y is considerably smaller and also subtelocentric (Radu et al. 1975). Willhoeft et al. (1998) microdissected the sex chromosomes and used them as labeled DNA probes for FISH studies of Medfly mitotic chromosomes. Five different Y chromosome probes were generated, each containing material from the proximal region (approximately 15%) of Yq believed to contain the male-determining region (Willhoeft and Franz 1996). Material from the distal portion of Xq from five flies was pooled to create an X chromosome probe.
Two different FISH techniques were performed, one being a simple labeling of chromosomes with the probes described and the other being a chromosomal in situ suppression hybridization (CISS), where the probe DNA was annealed with a large excess of Medfly female genomic DNA prior to hybridization to screen out sequences common to both males and females, enriching the probe for male-specific DNA. With the standard FISH technique, the Y chromosome probes hybridized to regions on both of the sex chromosomes and on some of the autosomes. When CISS hybridization was performed with these probes, fluorescence was primarily seen on the Y chromosome male-determining region, as expected. From this, Willhoeft et al. (1996) presumed that the majority of the visible hybridization would be due to repetitive sequences because of the complexity of the Medfly genome. Unique sequences would probably be underrepresented in the FISH probes, and any hybridization occurring with these sequences would not be visible without some type of fluorescent signal enhancement. The FISH studies confirm that the Y chromosome is largely heterochromatic, and that Y repetitive sequences can also be found on the X and on autosomes.

FISH was also performed with the X chromosome probe. Hybridization was detected on the X chromosome and on some of the autosomes. CISS of the X chromosome probe was done with an excess of Medfly genomic DNA to filter out repetitive sequences. The resulting probe produced weaker signals on the sex chromosomes and autosomes compared to the unsubtracted probe. The combined FISH results indicate that the X chromosome is abundant in "X-enriched" repetitive DNA. Willhoeft et al. concluded that there are three distinct regions on the Medfly sex chromosomes. "Region I" includes the centromeric regions and Xp and Yp, which
contain rRNA genes and a relatively lower occurrence of repetitive DNA. "Region II" includes Xq and Yq, from which the probe DNA was made. These regions are abundant in repetitive DNA and X- or Y-chromosome enriched sequences. "Region III" consists of regions that did not hybridize to the given probes. An example is the constriction in Yq, which was presumed to be euchromatic but does not appear to be involved in male determination or male fertility, nor does it contain a high abundance of repetitive DNA. Willhoeft et al. hypothesize that this region was recently added to the Y.

The Y chromosome was found to be generally uniform in size among various Medfly strains. There were, however, differences observed in the FISH hybridization pattern of a Medfly Y chromosome-specific repetitive sequence (pY114) between the Y chromosomes of two Kenya wildtype strains and the Y chromosomes the other Medfly strains. Willhoeft and Franz (1996) also used both conventional chromosome staining and FISH in their comparison of the chromosomes of the Medfly to those of two related Tephritid species, *Ceratitis rosa* and *Trirhithrum coffeae*. The C-banding pattern of *C. rosa* autosomes resembled that of the Medfly. The *C. rosa* Y chromosome resembles that of the Medfly Y, but the X chromosomes differ in size, with the *C. rosa* X being the smaller of the two. The chromosomes of *T. coffeae* differ in size and C-banding pattern from those of *C. capitata* and *C. rosa*. FISH studies showed that, like the Medfly, ribosomal genes are found in Xp of *C. rosa* and *T. coffeae*. However, rRNA gene
hybridization patterns differed between the Y chromosomes of these species. Also, the pY114 repetitive DNA probe did not hybridize to the *C. rosa* or *T. coffeae* chromosomes. The C-banding results suggest a closer evolutionary relationship between *C. capitata* and *C. rosa* than *C. rosa* and *T. coffeae*. FISH of *C. rosa* and *T. coffeae* chromosomes with the Medfly male-determining region also seem to confirm the relationships between these species (Willhoeft et al. 1998).

A number of Y-specific repetitive sequences have been isolated and characterized from the Medfly Y (Anleitner and Haymer 1992). Differential hybridization was used to screen a lambda (λ) bacteriophage library of Medfly genomic DNA and isolated male-specific and male-enriched repetitive sequences. These were confirmed to reside on the Medfly Y via *in situ* hybridization. The Y-specific sequence Y114, originally isolated by Anleitner and Haymer, was found to be particularly rich in the nucleotides A and T (Zhou et al. 2000). This sequence was also used in FISH studies of the Medfly Y, described previously.

**1.6.2 Bactrocera sex chromosomes**

Very little, if anything, is known about the DNA sequence composition or gene content of the *B. dorsalis* sex chromosomes. What little is known is from the studies of Baimai et al. (1995), which cytologically characterized metaphase chromosomes of species within the *B. dorsalis* complex of fruit flies found in Thailand. The X chromosomes that were observed appeared as small and metacentric or large and submetacentric, and consist of one largely euchromatic arm and another predominantly heterochromatic arm. Cytological staining of the Y suggests that it is entirely
heterochromatic. The location of the Y centromere varies between species, creating a variety of forms of the Y ranging from telocentric, subtelocentric, or submetacentric. In all cases, the Y is significantly smaller than the X, appearing as a small, darkly staining dot on mitotic chromosome spreads. The combination of its heterochromatic nature and small size makes the \textit{B. dorsalis} Y a challenging candidate for further characterization through conventional molecular techniques.

1.6.3 The Medfly Y and male sex determination

Less is known about sex determination in Tephritid fruit flies than compared to some of the other Dipteran species mentioned previously. Of the Tephritids, most of the the work in this area has focused on the Mediterranean fruit fly \textit{Ceratitis capitata} (Wiedemann). The diploid number of chromosomes in this species is 12. Unlike \textit{Drosophila melanogaster}, maleness in the Medfly is determined by the presence of a Y chromosome. In Drosophila, XXY individuals would be female, while in the Medfly they would be male (Lifschitz and Cladera 1989). Zapater and Robinson (1986) described aneuploid Medfly adults that were XX22\textsuperscript{Y}, where a portion of the Y chromosome long arm proximal to the centromere was translocated to chromosome 2, as being phenotypically male. Like the Dipteran species \textit{M. domestica} and \textit{M. scalaris}, evidence indicates that the Medfly Y chromosome contains a male-determining region.
Willhoeft and Franz (1996) have since localized the male-determining region to the proximal region of Yq, approximately 15% of the total length of the Y, by utilizing existing Medfly Y-autosome translocation strains and fluorescence in situ hybridization (FISH). However, the male-determining locus has yet to be fully isolated and characterized. Willhoeft and Franz also observed that large portions of the Medfly Y could be absent without any apparent effect on sex determination or male fertility.

1.6.3.1 Sex-specific gene expression in the Medfly

Genes expressed in a sex-specific manner have also been isolated from the Medfly. Rina and Savakis (1991) cloned and characterized four different vitellogenin (egg yolk protein) genes, which are similar in structure to the Drosophila yolk protein genes, and mapped them to the Medfly chromosome 5. Thymianou et al. (1998) isolated a male-specific cDNA sequence by screening a λ bacteriophage library containing cDNA sequences from the Medfly adult male fat body. Further characterization of this male-specific cDNA showed it that likely encodes a protein belonging to the pheromone and general odorant-binding protein (PBP-GOBP) family. Rosetto et al. (2000) isolated four X chromosome-linked genes encoding ceratotoxin proteins, which are present in a substance with antibacterial properties that is secreted by the female onto her fertilized eggs during oviposition. FISH studies of Medfly mitotic chromosomes showed that these
genes reside on the X chromosome in a cluster approximately 26 kb in size. Screening of a Medfly cDNA libraries with a portion of Drosophila Sxl gene sequence identified a gene with sequence similarity to Drosophila Sxl, but it does not appear to undergo sex-specific splicing. Transcripts of this Medfly Sxl (CcSxl) were found to be present in both Medfly males and females, but sex-specific forms of CcSXL were not detected (Saccone et al. 1998).

1.6.4 The B. dorsalis Y and male sex determination

In the Oriental fruit fly Bactrocera dorsalis, the mechanism of sex determination has not yet been deciphered. A homolog of the Drosophila sex determining gene dsx has also been found in the Queensland fly Bactrocera tryoni. Like its counterpart in Drosophila, the B. tryoni Dsx undergoes sex-specific splicing to create two different forms of DSX (Shearman and Frommer 1998). Cytological characterization of this species has confirmed the presence of two X chromosomes in females and one X and Y in males, but the precise role of the Y chromosome in the process of sex determination has not yet been defined.
1.7 Retrotransposable elements and the Y chromosome

There are several examples that attest to the affinity of retrotransposable elements for the heterochromatic Drosophila Y chromosome (Terrinoni et al. 1997, Junakovic et al. 1998, Agudo et al. 1999, Carvalho et al. 2000) and the *D. miranda* neo-Y chromosome (Steinmann and Steinmann 1997). There is also evidence of a retrotransposable element inserted in the heterochromatic Y chromosome of *C. capitata* (Zhou and Haymer 1998). The insertion of these elements is by no means exclusive to Y chromosomes. A nested series of retrotransposable elements have been isolated from large portions of the nonrecombining W chromosomes from the silkworms *Bombyx mori* and *Bombyx mandarina*, where females are the heterogametic sex (with one Z and one W chromosome) and males are homogametic (ZZ, Abe et al. 2002).

1.7.1 General characteristics of retrotransposable elements

Retrotransposable elements, also commonly known as “retrotransposons”, are a class of mobile DNA element that differ from conventional transposable elements, or transposons, by their mechanism of integrating into their host genome. Specifically, retrotransposons integrate via an RNA intermediate. Retrotransposable elements fall into one of two major proposed groups based upon the sequence organization of the elements. Members of one group have long terminal repeat sequences at both ends of a central
region that contains protein-encoding ORFs. These elements are most closely related to retroviruses and are known as LTR (long terminal repeat) retrotransposable elements or LTR retrotransposons. The second group consists of retrotransposable elements lacking LTRs, which also called “non-LTR retrotransposons” or simply “retroposons”. This broad classification also reflects differences in the mechanisms of transposition existing between the two groups (reviewed by Robertson and Lampe 1995, McClure 1991).

1.7.2 LTR vs. non-LTR retrotransposable elements

LTR retrotransposons consist of a central region up to 9 kb in size. This internal region consists of multiple ORFs, which encode for proteins used for transposition, sandwiched between terminal repeat sequences up to 500 bp in length. Typically, the central region consists of three distinct ORFs, named after their homologies to retroviral genes – \textit{gag (group associated antigen)} encodes a protein required for an important protein-DNA structure required for transposition, \textit{pol (polymerase)} primarily encodes the reverse transcriptase and integrase enzymes in addition to other enzymes required for transposition, and \textit{env (envelope)} encodes a transmembrane protein associated with retroviral infection and may confer the potential for infectious activity to these retrotransposons (reviewed by Robertson and Lampe 1995).
Non-LTR retrotransposons also consist of a central region of one or two ORFs. Lacking LTRs, the ORF-containing region is instead flanked by 5’ and 3’ untranslated regions and a 3’ poly-A region. These elements integrate into their host genome with the assistance of an “integrase”, an endonuclease that is not related to the integrases found in LTR retrotransposons. These elements are widely represented in insects, and appear to be related to the vertebrate LINEs (Long Interspersed Nuclear Elements; reviewed by Robertson and Lampe 1995).

1.7.3 R1 and R2 elements

R1 and R2 are non-LTR retrotransposable elements found primarily within 28S ribosomal RNA (rRNA) genes. Once known as “Type I” and “Type II” insertions, they were recently renamed to reflect their high degree of sequence similarity with the R1Bm and R2Bm non-LTR retrotransposable elements that are also found inserted in the 28S rRNA genes in the silkmoth Bombyx mori (Jakubczak et al. 1990). R1 and R2 elements consist of a central region containing one or more protein-encoding ORFs flanked by 5’ and 3’ untranslated regions (UTRs). The abundance of R1 elements differs between the Drosophila sex chromosomes. Approximately 60% of X chromosome 28S rRNA genes and a considerably smaller percentage of Y chromosome 28S rRNA genes in the Oregon R strain of D. melanogaster contain R1 elements (Tartof and Dawid 1976). Drosophila R1 elements are typically 5 kb in size, with two overlapping open reading frames (ORFs). ORF 1 is similar in sequence to gag genes, encoding an endonuclease (integrase) and ORF 2 is similar in sequence to pol genes, encoding a reverse transcriptase. The presence of an R1 element in the 28S rRNA coding sequence appears to render that gene
copy inactive (reviewed by Jakubczak et al. 1991), but there does not appear to be a proportional increase in the accumulation of rRNA gene copies to compensate for the loss of gene activity by the copies interrupted by these elements (Jakubczak et al. 1992). Drosophila R2 elements are inserted 74 bp upstream of R1 elements within 28S rRNA genes, and are equally represented amongst the rRNA genes found on the X and Y chromosomes (Roiha et al. 1980, reviewed by Jakubczak et al. 1990). In contrast to R1, the 3.5 kb R2 elements contain a single ORF that appears to be the result of a gag-pol fusion (Jakubczak et al. 1990). The insertion of R1 and R2 elements appear to be independent of each other, with examples of Drosophila 28S rRNA gene copies containing either R1 or R2, or both (Jakubczak et al. 1992).

1.7.3.1 The prevalence of R1 and R2 elements in insects

R1 and R2 elements are found in the rRNA genes of a number of insect orders, including Diptera, Hymenoptera, Lepidoptera, and Coleoptera (Burke et al. 1993). Alignments involving representative sequences from each of these orders reveal a number of highly conserved regions. First, cysteine motifs were found in both the integrase (three motifs) and reverse transcriptase (one motif) encoding regions of these elements. Secondly, recovered copies of these elements are often truncated. The 3' ends of these elements appear to be important for the maintenance of the activity of these elements, and are thus the most conserved regions (Jakubczak et al. 1990). Lastly, comparisons of R1 and R2 insertion sites in 28S rRNA genes show that these are also highly conserved across different species (Jakubczak et al. 1991).
Detailed sequence comparisons of R1 and R2 elements show that several families of these elements can coexist within the same species. Examples of species harboring multiple families of R1 or R2 elements include Drosophila melanogaster (multiple R1 families), the beetle Popillia japonica (multiple R2 families), and the wasp Nasonia vitripennis (multiple R1 families; Lathe et al. 1995, Burke et al. 1993). Gentile et al. (2001) utilized degenerate primers, designed for amino acid motifs unique to different Drosophila R1 families, to further examine the distribution of each family among various Drosophila species. Their sequence comparisons and phylogenetic analyses confirmed the existence of two major Drosophila R1 families, designated R1A and R1B. The R1A family encompasses the R1 elements found in the bulk of Drosophila species, while the R1B family consists of R1 elements found in D. neotestacea whose nucleotide sequences differ greatly from those of the R1A family. The R1A family was further subdivided into three subfamilies – R1A1, R1A2, and R1A3 – to accommodate three different variations of the R1A element sequence found in Drosophila takahashii. Recently, Busseau et al. (2001) identified two new subfamilies of Drosophila non-LTR retrotransposons, Waldo-A and Waldo-B, which show similarity to R1 elements. Both elements, like R1 retrotransposons, contain two large overlapping ORFs, ORF1 encoding an endonuclease ("integrase") and ORF2 encoding a reverse transcriptase. Multiple copies of these elements were recovered from Drosophila, many of which were found to
be located near or within (CA)$_n$ repeats. The high degree of amino acid conservation within R1 families is believed to be due to the continual elimination of divergent R1 sequences from the rRNA genes via recombination. Retrotransposition of these elements is believed to be responsible for their retention and maintenance in the rRNA genes (Jakubczak et al. 1992).

1.7.3.2 A model for R2 element transposition

Luan et al. (1993) has constructed a model for R2 element retrotransposition, a mechanism termed "RNA-mediated integration". First, the R2 ORF encodes a 120 kiloDalton (kDa) protein, which has been shown to associate with the 3' end of the R2 element transcript and also recognizes a 30 bp insertion site within the 28S rRNA gene. This protein then makes a nick in the coding strand of the double-stranded target. The exposed 3' end in the host genomic DNA resulting from this nick is used to prime reverse transcription, which is also performed by the 120 kDa protein, into the insertion site. The second strand is then nicked, and the element incorporated into the coding strand is used as a template for the replication of the sequence into the second strand, completing the integration of the R2 element into the 28S rRNA gene.
1.7.3.3 Phenotypes associated with R1 and R2 element insertion in Drosophila

Two different Drosophila phenotypes are either directly or indirectly associated with the levels of R1 and R2 element insertions in rRNA genes. One of these is the *bobbed (bb)* phenotype found in *D. melanogaster* and *D. hydei*. The *bb* phenotype is manifested in these Drosophila species as the shortening and thinning of scutellar bristles and is accompanied by delayed development. Correlations have been made between the severity of the *bb* phenotype and either the abundance of rRNA genes lacking R1 or R2 insertions or the decreasing levels of rRNA synthesis (reviewed by Franz and Kunz 1981). Because R1 and R2 element insertions inactivate rRNA gene copies, they are believed to be direct contributors to the *bobbed* phenotype.

R1 elements are also implicated in the *Drosophila mercatorum* phenotype *abnormal abdomen (aa)*. The *aa* phenotype closely resembles the *bb* phenotype, but differs in its effect on specific developmental stages. Both the level of R1-inserted rRNA genes and the locus under-replication (*ur*) are believed to be responsible for this phenotype. The *ur* locus controls the normal underreplication of 28S rRNA genes in *D. mercatorum* polytene chromosomes; those with R1 element insertions are even more underreplicated than the uninserted rRNA genes. Thus, the *aa* phenotype is believed to be the result of a high frequency of R1 element insertion occurring concurrently with the loss of normal levels of rRNA gene underreplication (reviewed by Jakubczak et al. 1992, Malik and Eickbush 1999). Malik and Eickbush (1999) critically reevaluated past
studies that attempted to elucidate the mechanisms involved in the \textit{aa} phenotype. They confirmed past results that showed that \textit{aa} flies underreplicate rRNA genes containing R1 element insertions, which is also seen in other Drosophila species besides \textit{D. mercatorum}. However, contrary to these earlier results, it does not appear that \textit{aa} flies underreplicate uninserted rRNA genes (which is seen in wildtype Drosophila) to compensate for the loss of activity of the inserted rRNA genes.

1.8 Transposable elements in the Tephritids \textit{C. capitata} and \textit{B. dorsalis}

Examples of transposable and retrotransposable elements have been isolated from the Medfly and Oriental fruit fly genomes. The LTR-retrotransposon \textit{yoyo}, belonging to the \textit{gypsy} family of retrotransposons, was found to have copies on both the autosomes and on the Medfly Y (Zhou and Haymer 1998). Handler and Gomez (1997) isolated the transposable element \textit{hopper} from \textit{B. dorsalis}, which has sequence similarity to the Drosophila transposon \textit{hobo}. The \textit{hobo} element has been shown to be similar in sequence to the \textit{Activator (Ac)} transposon found in maize and the \textit{Tam3} transposon in snapdragons; thus, sequence data prompted the grouping of these three elements to form a relatively new family of transposons (reviewed by Handler and Gomez 1997).
1.9 Genomic DNA hybridization as a means of isolating specific types of sequences

1.9.1 Past work in the isolation of Y chromosome specific sequences

In general, the Y chromosome, outside of rRNA genes, male-determining, or fertility factors, is largely composed of heterochromatin. An early group led by Kunkel was able to isolate repetitive DNA sequences specific for the human Y chromosome through hybridization studies (Kunkel et al. 1976). Hennig showed that the general difficulty in isolating Y chromosome DNA sequences, as illustrated in his work with Drosophila hydei, was predominantly due to the repetitive nature of the majority of the sequences. The isolation and localization of these sequences, via early recombinant technology, was a considerable challenge (Hennig 1985). With increasingly sophisticated DNA manipulation techniques, both Y chromosome specific and Y associated repetitive sequences have been successfully found and characterized in Drosophila. “Y specific” sequences have been defined by Vogt and Hennig (1983) as repetitive DNA sequences found only on the Y chromosome, but may include single copies present elsewhere in the genome. These Y specific sequences are generally short (approximately 200 bp) and high in copy number (up to 1000 copies). “Y associated” DNA sequences are not exclusive to the Y chromosome and can be several kilobases in length, but relatively low in copy number (10-100 copies). In Drosophila, Y specific sequences are generally species-specific, while Y associated sequences are “more conserved” (Hennig 1986). In this dissertation, the strategy for isolating Y-specific sequences in B. dorsalis was to utilize genomic DNA hybridization techniques.
The study of DNA reassociation kinetics has long been used as a technique for analyzing genomes. In several instances, methods of genomic DNA hybridization have been employed specifically to detect differences between two or more genomes. For example, Lamar and Palmer (1984) used a relatively simple method of DNA hybridization (referred to as “deletion enrichment”) to isolate several DNA sequences unique to the murine Y chromosome. Restriction enzyme digested male genomic DNA was hybridized to an excess of female genomic DNA sheared by sonication. Following hybridization, only male-male hybrids contained specific enzyme-generated ends, and only these hybrids would be successfully cloned into plasmid vectors, essentially generating a plasmid library of Y-specific DNA. Straus and Ausubel (1990) developed an early form of “genomic subtraction” and tested the technique by comparing yeast wildtype and deletion mutant strains. This genomic subtraction involves repeated rounds of hybridization between restriction enzyme digested wildtype genomic DNA with an excess of biotinylated and enzyme digested deletion mutant genomic DNA. Biotin-containing hybrids (representing wildtype-mutant hybrids or mutant-mutant hybrids) were removed by binding to avidin-coated beads. Wildtype-wildtype hybrids (lacking biotin labels) were then enriched by ligating short oligonucleotide adaptors and amplified via PCR.
1.9.2 Representational Difference Analysis (RDA)

RDA is a method of subtractive genomic DNA hybridization (Figure 2) used to isolate the differences between two complex genomes that is similar to the genomic DNA subtraction method originated by Straus and Ausubel (1990). Perhaps the greatest difference between the two techniques is that, because of a preliminary step enriching the two genomes to be used, RDA requires much less starting material. In RDA, one genome serves as the “tester” (the genome of interest, from which differences will be isolated) and the other as the “driver”. Both genomes are first digested with the same restriction enzyme. Then, short oligonucleotide adaptors are ligated to the genomic DNA fragments. The genomes are then reduced to “representations” via the Polymerase Chain Reaction (PCR) to decrease their complexity (the “total length of unique sequences”) to 2-15% (Lisitsyn 1999). Because PCR favors the amplification of shorter DNA sequences, representations are anticipated to contain DNA sequences that are 150-1500 bp in size (Lisitsyn 1999). The first set of adaptors is removed from both tester and driver, and a new set of adaptors is ligated to the tester but not the driver. The tester DNA is denatured, then undergoes hybridization with an excess of single-stranded driver DNA. After hybridization, PCR amplification is performed on the DNA pool. Because tester-tester hybrids will have adaptors on both strands of the hybrid, only these hybrids will be exponentially amplified with PCR. After each round of hybridization and amplification, unhybridized single-stranded DNAs are eliminated from the pool using mung bean endonuclease, and the adaptors are removed from the endonuclease-treated DNA pool with the restriction enzyme originally used to create the representations.
Figure 2. A schematic of the Representational Difference Analysis (RDA) technique, as illustrated by Lisitsyn et al. (1993) and Brown (1994).
Tester (male) representation

Driver (female) representation, in excess

Ligate 12-mer and 24-mer adaptors

Combine tester with excess driver, denature, anneal

- tester-tester hybrid
- tester-driver hybrid
- driver-driver hybrid
- single-stranded tester
- single-stranded driver

Fill in the ends

PCR amplify using 24-mer adaptor as primer

- tester-tester hybrids exponentially amplified
- tester-driver hybrids and single-stranded tester linearly amplified
- driver-driver hybrids and single-stranded driver not amplified

Remove adaptors
New adaptors are ligated to the DNA pool, and then the DNA is denatured and allowed to hybridize with a fresh excess of denatured driver DNA (Lisitsyn et al. 1993). Thus, two major factors in the RDA method increase the likelihood of recovering tester-specific sequences – the subtraction of sequences common to both tester and driver via repeated rounds of DNA hybridization combined with the enrichment of tester-specific sequences within the DNA pool via PCR (Lisitsyn et al. 1999).

RDA has been successfully used in a variety of applications. This technique was employed to isolate Y chromosome-specific DNA sequences from the dioecious plant *Silene latifolia*, useful as Y-specific DNA markers for the localization of male-determining genes on the Y chromosome (Donnison et al. 1996). RDA was also used to isolate two sex-specific cDNAs from mouse fetal brains. One of the isolated sequences was found to be the male-specific gene *Smcy* and the other was the female-expressed gene *Xist* (Eriksson et al. 1999). Michiels et al. (1998) modified the RDA technique to accommodate small amounts of tissue or cell culture (as little as 50 diploid cells) as starting material, making it possible to apply RDA to tissue samples such as tissue biopsies, where the sample may not contain a single cell type and microdissection techniques are capable of isolating only small samples of tumor cells.
Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Fruit fly strains

The Bactrocera species used in this dissertation are shown in Table 1. The B. *dorsalis* embryos, third instar larvae, and adult flies used in this dissertation are from the Puna, Hawaii strain maintained in our lab for several years. The Puna lab strain originated from wild flies collected in Puna, Hawaii circa 1985 by the USDA/ARS Fruit Fly Laboratory in Honolulu, Hawaii. Wild *B. dorsalis* adults originating from Taiwan were captured on the campus orchards of the National Pingtung University of Science and Technology in Pingtung, Taiwan using various fruit odorant lures and were kindly provided by Dr. Po-Yung Lai. *B. dorsalis* from Okinawa Island, Japan (reared at the Naha Plant Protection Office) were provided by Dr. Tadashi Teruya.

Wild *Bactrocera carambolae* and *Bactrocera papayae* adults originating from Malaysia were kindly provided by Suk Lin. The *Bactrocera curcurbitae* adults used in this dissertation are from the USDA/ARS Laboratory lab stock reared at their facilities in Honolulu, Hawaii. Wild *Bactrocera oleae* adults originating from Cuenca province, Spain were provided by Dr. Dolores Ochando.
Table 1. A summary of Bactrocera species used in this dissertation. Genomic DNA was extracted from single individuals. Mass adult genomic DNA extractions were done for Southern blots (described in Section 2.2.2).

<table>
<thead>
<tr>
<th>Species</th>
<th>Sample Origin</th>
<th>Developmental Stage and Sex (if identified)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bactrocera dorsalis</strong></td>
<td>Puna, Hawaii</td>
<td>Embryos</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 3 hour egg collection time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 24 hour maturation time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Third instar larvae</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pupae (1-2 days old)</td>
</tr>
<tr>
<td></td>
<td>Kalaheo, Hawaii</td>
<td>Adult ♂, ♀</td>
</tr>
<tr>
<td></td>
<td>Pingtung, Taiwan</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Okinawa Island, Japan</td>
<td></td>
</tr>
<tr>
<td><strong>B. papayae</strong></td>
<td>Hong Tan, Malaysia</td>
<td></td>
</tr>
<tr>
<td><strong>B. carambolae</strong></td>
<td>Hong Tan, Malaysia</td>
<td></td>
</tr>
<tr>
<td><strong>B. curcubitae</strong></td>
<td>USDA/ARS lab stock</td>
<td></td>
</tr>
<tr>
<td><strong>B. latifrons</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. occipitalis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. oleae</strong></td>
<td>Cuenca province, Spain</td>
<td></td>
</tr>
<tr>
<td><strong>B. philipinensis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. pyrifoliae</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.1.2 Reagents and systems

2.1.2.1 Enzymes

Proteinase K, RNase A, and all restriction endonucleases were obtained from Roche Molecular Biochemicals and were used under conditions recommended by the manufacturer. All other enzymes used were included in system kits as described in later sections.

2.1.2.2 Chemical reagents

Chemical reagents were purchased from chemical companies such as Bio 101, Inc., Boehringer Mannheim, Inc., Gibco BRL, Research Organics, Inc., Sigma, Inc., Stratagene, etc. Whenever possible, molecular biology grade reagents were purchased for use.

2.1.2.3 System products

The Dig High Prime DNA Labeling and Detection Starter Kit was purchased from Roche Molecular Biochemicals. This kit includes the reagents required for nonradioactive DNA labeling (with digoxigenin) ("Dig High Prime"), probing, and probe detection (10X blocking solution, Anti-dig-alkaline phosphatase conjugated antibodies, and NBT-BCIP tablets (substrate for the color detection reaction).

AmpliTaq\textsuperscript{R} DNA polymerase, 10X PCR Buffer II, and 25 mM MgCl\textsubscript{2} were obtained from Applied Biosystems. 10 mM dNTPs (containing dATP, dCTP, dTTP, dGTP) were obtained from Roche Molecular Biochemicals. Operon Technologies synthesized all RDA adaptor oligonucleotides and all PCR primer oligonucleotides.
The Expand Long Template PCR System from Roche Molecular Biochemicals was used to perform all long template PCRs. This kit contains buffers optimal for long template PCR and the Expand DNA polymerase.

The GeneClean Spin® DNA purification kit was purchased from Bio 101, Inc. This kit provides a rapid method of DNA purification from agarose gels or of DNA following various enzymatic manipulations.

The PCR-Script™ Amp Cloning Kit was purchased from Stratagene. The kit includes the requisite reagents and enzymes for the cloning of PCR amplified DNA into the plasmid vector pPCRSScript Amp SK (+).

The TA Cloning Kit® was purchased from Invitrogen, and was used as an alternative cloning method to the Stratagene PCR-Script™ Amp Cloning Kit. The cloning vector pCR2.1 is included as a kit component.

The Qiaprep kit was obtained from Qiagen, Inc. This kit was used for the rapid extraction and purification of high quality plasmid DNA from liquid cultures of *E. coli*.

The Turboblotter DNA transfer system was purchased from Schleicher and Schuell. The system contains a blotting tray and absorbent paper designed for downward DNA transfers from agarose gels onto nylon or nitrocellulose membranes.

Subcloning Efficiency DH5α *E. coli* cells were obtained from Invitrogen for heat shock transformation. DH10B electroporation cells and disposable Micro-Electroporation Chambers were used with the Cell Porator™ System electroporation apparatus, all from Gibco BRL.
2.1.2.4 PCR oligonucleotide primers

The sequence analysis program MacVector (Kodak) was used to design PCR oligonucleotide primer sequences from known sequence data. The primer pair 3.1aRevF1-B1 was designed to amplify most of the 3.1aRev sequence. The sequences of the primers are provided in Table 2, and the locations of the primer annealing sites are depicted in Figure 3.

Table 2. PCR Oligonucleotide primers designed for the 3.1a clone.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1aRevF1</td>
<td>GGC TAA ACG ACT TAG TCT GGG GTG</td>
</tr>
<tr>
<td>3.1aRevF2</td>
<td>AAT GGT GGA CGC ATG AGC TAA CCC</td>
</tr>
<tr>
<td>31aLongF1</td>
<td>CCA CGC TTT GTG TGG TAA ATC AC</td>
</tr>
<tr>
<td>31PasteF1</td>
<td>TTA GTG TCA CCA CGC TTT GTG TG</td>
</tr>
<tr>
<td>3.1aRevB1</td>
<td>GCT CCG TAA AAA CTG AAT CGT CC</td>
</tr>
<tr>
<td>3.1aRevB2</td>
<td>GGC TGT TCA AAC GCA ACT GGT AG</td>
</tr>
<tr>
<td>31LongB1</td>
<td>TGA GTG TGC AAG GGC TCT TCT AC</td>
</tr>
</tbody>
</table>

2.1.2.5 DNAs and probes

Inserts from phage clones were subcloned into the plasmid vector pUC19. PCR products were cloned into plasmid vectors using either the pPCRScript SK(+) Cloning Kit (Stratagene) or the TA Cloning Kit (Invitrogen). To generate nonradioactively labeled DNA probes, approximately 1 μg of DNA was diluted to a total volume of 16 μl with sterile water, then denatured at 95°C for 10 min., and cooled on ice. The DNA was labeled with digoxigenin by adding 4 μl of Dig High Prime Mix (Roche Molecular Biochemicals) to the cooled DNA and incubated at 37°C overnight.
Figure 3. Diagram of the 3.1a clone, showing the locations of the 3.1a clone primers.
E = EcoRI  
X = XhoI  
P = PstI

E = EcoRI  
X = XhoI  
P = PstI

region with amino acid sequence similarity to D. melanogaster R1Dm
2.2 Methods

2.2.1 Genomic DNA isolation from B. dorsalis embryos

B. dorsalis embryos were collected from mated adult females of the Puna, Hawaii lab strain at approximately two weeks of age. Eggs were collected for three hours. To isolate genomic DNA from these embryos, the collected embryos were first matured for 24 hours. Single embryos were then transferred to 1.5 ml tubes and mechanically ruptured with a small bore micropipette tip. The ruptured embryo and its contents were then thoroughly mixed with a 5 µl aliquot of sterile distilled water by pipetting up and down several times. InstaGene Matrix (Bio-Rad Laboratories) was added (100 µl) and the mixture was incubated at 56°C for 30 min., pausing once during the incubation to vortex the tube for 10-15 sec. The mixture was then placed in a boiling waterbath for 8 min., vortexed again for 10-15 sec., and centrifuged at 12,000 rpm for 3 min. at room temperature to pellet the matrix. The supernatant (containing the genomic DNA) was recovered and stored at -20°C.
2.2.2 Genomic DNA isolation from single *B. dorsalis* third instar larvae and *Bactrocera* adult flies

Genomic DNA was isolated from *B. dorsalis* third instar larvae, 1-2 day old pupae, and newly emerged adult flies from all species via the Lifton method described by Anleitner and Haymer (1992). For single adult flies preserved in ethanol, excess ethanol was blotted off using a clean paper towel and the fly was immersed in 300 μl of Lifton Grind Buffer (0.2M sucrose, 50 mM EDTA, 100 mM Tris pH 9.0, 0.5% SDS) overnight at 4°C. The residual buffer was blotted off with a clean paper towel and the Lifton extraction method was then performed. If freshly killed adult flies were used, females were harvested shortly (≤ 8 hrs.) after eclosion to ensure that they had not yet mated.

Freshly killed adults were prepared for homogenization by freezing at -20°C for 15-20 minutes. Single flies were placed in a blue Kontes 1.5 ml tube with 250 μl of Lifton Grind Buffer (0.2M sucrose, 50 mM EDTA, 100 mM Tris pH 9.0, 0.5% SDS) and thoroughly homogenized on ice using Kontes microhomogenizers. 100 μl of fresh Lifton Grind Buffer was used to rinse the homogenizer in the Kontes tube. The homogenate was strained though a 1 cc syringe lined with sterile polyester fiberfill into a 1.5 ml polypropylene tube on ice. 3 ul of 20 mg/ml Proteinase K (Roche Molecular Biochemicals) was added to the strained homogenate (for a final concentration of 0.2 mg/ml) and the mixture was incubated for one hour at 65°C. 38 μl of cold 8M potassium acetate was then added and the mixture was incubated on ice for at least one hour (up to overnight). The treated homogenate was centrifuged at 4°C for 15 min. at 10,000 rpm and 250 μl of the supernatant was transferred to a fresh 1.5 ml tube. Two volumes of
room temperature 95% ethanol were added to precipitate the genomic DNA, the tube contents were mixed, and the centrifugation was repeated at 4°C. The supernatant was discarded, leaving the pelleted genomic DNA. The pellet was resuspended in 200 ul TE, pH 8.0. 1 μl of 10 mg/ml RNase A was added to the resuspended pellet (for a final concentration of 50 μg/ml) and the mixture was incubated for 15-30 minutes at room temperature. An equal volume of phenol was added, the tube contents were mixed and briefly centrifuged to ensure a clear phase separation, and the upper aqueous phase was transferred to a fresh 1.5 ml tube. Two consecutive extractions with an equal volume of CHISM (chloroform: isoamyl alcohol, 24:1 mixture) were performed, each time recovering the aqueous phase. The genomic DNA was then precipitated from the aqueous phase by adding 0.1 volume of 3M sodium acetate, pH 6.0 and 2.5 volumes of ice-cold 95% ethanol. The precipitation was allowed to continue overnight at -20°C (minimum of one hour). The genomic DNA was recovered by centrifugation at 10,000 rpm and 4°C for 15 minutes. The supernatant was removed and the pellet was washed with 500 ul of ice-cold 70% ethanol, then centrifuged again for 10 minutes at 4°C. The ethanol wash was discarded and the pellet was vacuum dried for five minutes to remove any residual ethanol. The dried pellet was resuspended in 20 μl of TE, pH 8.0 and stored at 4°C. The quality and quantity of the isolated DNA was assessed via agarose gel electrophoresis.
2.2.3 Genomic DNA isolation from multiple adult flies

Mass genomic DNA isolations (genomic DNA extractions from more than one fly at a time) were done when larger amounts of DNA were required. As described in the previous section, genomic DNA was isolated using the Lifton method on a larger scale to accommodate the increased amount of starting material. Again, adult females were harvested shortly after eclosion to ensure that they had not yet mated.

Fresh flies were prepared for homogenization by freezing at −20°C for 15-20 min. Approximately 0.3 g (about 30 flies) of fly material was placed into a ground glass homogenizer with 5 ml of Lifton Grind Buffer (0.2M sucrose, 50 mM EDTA, 100 mM Tris pH 9.0, 0.5% SDS) and thoroughly homogenized on ice. The resulting homogenate was strained though a 10 cc syringe lined with sterile polyester fiberfill into a 6 ml polypropylene tube on ice. 50 μl of 20 mg/ml Proteinase K (Roche Molecular Biochemicals) was added to the strained homogenate (for a final concentration of 0.2 mg/ml) and the mixture was incubated for one hour at 65°C. 750 μl of 8M potassium acetate was then added and the mixture was incubated on ice for at least one hour. The treated homogenate was centrifuged at 4°C for 15 min. at 10,000 rpm and the supernatant was recovered in a 15 ml polypropylene tube. Two volumes of room temperature 95% ethanol were added to precipitate the genomic DNA, the tube contents were mixed, and the centrifugation was repeated at 4°C. The supernatant was discarded, leaving the pelleted genomic DNA. The pellet was resuspended in 500 μl TE, pH 8.0. 2.5 ul of 10 mg/ml RNase A was added to the resuspended pellet (for a final concentration of 50 μg/ml) and the mixture was incubated for 15-30 minutes at room temperature, then
transferred to a 1.5 ml microcentrifuge tube. An equal volume of phenol was added, then
the tube contents were mixed and briefly centrifuged to ensure a clear phase separation,
and the aqueous phase was recovered. A second extraction with CHISM was performed,
again recovering the aqueous phase. The genomic DNA was precipitated from the
aqueous phase by adding 0.1 volume of 3M sodium acetate, pH 6.0 and 2.5 volumes of
ice-cold 95% ethanol. The precipitation was allowed to continue overnight at -20°C
(minimum of one hour). The genomic DNA was recovered by centrifugation at 10,000
rpm and 4°C for 15 minutes. The supernatant was removed and the pellet was washed
with 500 ul of ice-cold 70% ethanol, then centrifuged again at 4°C. The ethanol wash
was discarded and the pellet was vacuum dried for five minutes to remove any residual
ethanol. The dried pellet was resuspended in 100 μl of TE, pH 8.0 and stored at 4°C.
The quality and quantity of the isolated DNA was analyzed via agarose gel
electrophoresis.

2.2.4 Representational Difference Analysis (RDA)

The RDA procedure largely follows that described by Lisitsyn (1999), illustrated
in Figure 2. Genomic DNA from a single B. dorsalis adult male and female was used as
the tester and driver DNA, respectively. Both tester and driver DNAs were digested
with the restriction enzyme BamHI (Roche Molecular Biochemicals) and the
oligonucleotide adaptor sequences for the BamHI RDA subtraction were used. Two complete rounds of hybridization and difference enrichment were performed.

Three different oligonucleotide adaptor sets (RBam, JBam, and NBam) were used, their sequences following those prescribed by Lisitsyn et al. (1993). Their sequences are given in Table 3.

Table 3. RDA oligonucleotides adaptor sequences, from Lisitsyn et al. (1993)

<table>
<thead>
<tr>
<th>Adaptor Set</th>
<th>Adaptor Name</th>
<th>Adaptor Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RBam24</td>
<td>AGC ACT CTC CAG CCT CTC ACC GAG</td>
</tr>
<tr>
<td></td>
<td>RBam12</td>
<td>GAT CCT CGG TGA</td>
</tr>
<tr>
<td>2</td>
<td>JBam24</td>
<td>ACC GAC GTC GAC TAT CCA TGA ACG</td>
</tr>
<tr>
<td></td>
<td>JBam12</td>
<td>GAT CCG TTC ATG</td>
</tr>
<tr>
<td>3</td>
<td>NBam24</td>
<td>AGG CAA CTG TGC TAT CCG AGG GAG</td>
</tr>
<tr>
<td></td>
<td>NBam12</td>
<td>GAT CCT CCC TCG</td>
</tr>
</tbody>
</table>

2.2.4.1 Creating *B. dorsalis* male and female “representations”

1 μg of *B. dorsalis* adult male (tester) or female (driver) genomic DNA was digested with the restriction enzyme BamHI. 5 μl of 20X EDTA/glycogen mix (from the cDNA Subtraction Kit from Clontech Inc.) was then added to the digested DNA to aid in DNA precipitation. Each mixture was extracted once with phenol:CHISM mix (Roche Molecular Biochemicals), then once with CHISM, and the aqueous phase (containing the digested genomic DNA was recovered. The digested DNA was
precipitated with 3M sodium acetate and 95% ethanol for 1 hour at 4°C. The precipitated DNA was recovered by centrifuging at 13,000 rpm for 20 minutes at 4°C. The resulting DNA pellet was washed once with cold 70% ethanol, then resuspended to an approximate concentration of 100 ng/µl with 1X TE, pH 8.0. 1 µl of the purified DNA was checked on a 0.8% agarose gel.

Eight µl of the purified BamHI-digested DNA was combined with 7.5 µl of each of the “R” set of adaptors (RBam24 and RBam12, 60 pmoles/µl), 4 µl of sterile double-distilled water (dd H2O), and 3 µl of 10X ligase buffer (New England Biolabs) in a 0.5 ml PCR reaction tube. A COY Tempcycler II machine was used to heat the reactions to 50°C, then cool them gradually to 10°C over one hour. The reactions were cooled on ice, then 1 µl of T4 DNA ligase (400 units/µl, New England Biolabs) was added and the reaction tubes were incubated at 14°C overnight (approximately 18 hours). The ligations were diluted 1:50 in sterile ddH2O.

To create the representations, two 25 µl PCR reactions were prepared for the tester DNA and 16-25 µl reactions were prepared for the driver, consisting of sterile ddH2O, 1X Buffer II (Applied Biosystems), 0.2 mM dNTPs, 1.5 mM MgCl2, 30 pmoles of RBam24, and 2.5 µl of the 1:50 diluted adaptor ligation reaction. The reactions were incubated in a Perkin-Elmer GeneAmp 2400 thermocycler preheated to 72°C for 3 minutes, then 1 unit of AmpliTaq DNA polymerase (5 units/µl) was added to the tubes. The tubes were then incubated again at 72°C for 5 minutes, then each of the tester and driver reactions were amplified by performing 30 cycles of 94°C for 1 minute and 72°C for 3 minutes, with a final extension at 72°C for 10 minutes (followed by a 4°C hold).
3 μl of each reaction were electrophoresed on a 2% agarose gel to check for amplified DNA. The two tester (male DNA adaptor ligation) PCRs were pooled and 16 driver (female DNA adaptor ligation) PCRs were pooled. Both DNA pools were purified using Microcon microconcentrators (Amicon Bioseparations) and eluted in a total volume of 30 μl of sterile ddH₂O. The male tester representation was 10 μg/ml, while the female driver representation was 1180 μg/ml, as measured by an Eppendorf Biophotometer.

30 ng of the male representation and 3 μg of the female representation were digested with BamHI to remove the RBam adaptor set. The digested DNA was purified using Micron microconcentrators. The purified digests were run on a 2% Nu-Sieve (BioWhittaker Molecular Applications) low-melt agarose gel with bromophenol blue loading dye until the dye front migrated approximately 3 cm from the gel wells. The DNA located above the dye front was excised from the gel using a razor blade, mixed with 3 ml ddH₂O, 400 μl of 0.5M MOPS, pH 7.0 (Sigma), and 400 μl of 5M NaCl, then melted at 72 °C. The melted agarose was passed through a Tip-20 column (Qiagen Inc.) to remove the RBam adaptors from the DNA representations. The DNA representations were concentrated using Microcon microconcentrators and eluted in 20 μl of 1X TE, pH 8.0. 3 μl of the purified DNA representations were checked on a 1.2% agarose gel.
2.2.4.2 RDA primary hybridization

Two reactions were created to ligate the male tester representation to the “JBam” adaptor set. In each ligation reaction, 10 µl of the male tester representation was combined with 7.5 µl of each of the “J” set of adaptors (JBam24 and JBam12, 60 pmoles/µl), 2 µl of dd H2O, and 3 µl of 10X ligase buffer (New England Biolabs) in a 0.5 ml PCR reaction tube. A COY Tempcycler II machine was used to heat the reactions to 50°C, then cool them gradually to 10°C over one hour. After the reactions were cooled on ice, 1 µl of T4 DNA ligase (400 units/µl, New England Biolabs) was added and the reaction tubes were incubated at 14°C overnight (approximately 18 hours).

Both male tester representation ligations were combined with 20 µl of the female driver representation (from section 2.2.4.1) into a single tube and concentrated using Microcon microconcentrators (Microcon Inc.), then eluted with 4 µl of 3X EE buffer [consisting of 30 mM EPPS, pH 8.0 (Sigma) and 3 mM EDTA, pH 8.0 (Sigma)]. The eluted DNA mixture was placed in a 0.2 ml PCR tube and overlayed with 10 µl of mineral oil, then heated in a Perkin-Elmer 2400 thermocycler at 95°C for 3 minutes to denature the DNA. 1 µl of 5M NaCl was added to the tube, and the DNA was hybridized at 67°C overnight (20 hours) in the Perkin-Elmer 2400. 1:10, 1:20, and 1:40 dilutions of the primary hybridization mixture were made in sterile ddH2O. Reactions to “fill in” the 3’ ends of the ligated adaptors consisted of 2.5 µl of the 1:10, 1:20, or 1:40 primary hybridization dilution, 1X Buffer II (Applied Biosystems), 0.2 mM dNTPs, and 1.5 mM
MgCl₂. The mixture was incubated at 72°C for 3 minutes. 1 unit of Taq DNA polymerase (5 units/μl) was added, and the mixture was incubated again at 72°C for 5 minutes. 38 pmoles of the JBam24 adaptor (60 pmoles/μl) were added, and the fill-in reaction consisted of 10 cycles of 95°C for 1 minute and 72°C for 3 minutes, with a final extension at 72°C for 10 minutes (followed by a 4°C hold).

Each of the three filled-in primary hybridization reaction products were amplified via PCR. The PCR mixture consisted of 2.5 μl of the fill-in reaction, 1X Buffer II (Applied Biosystems), 0.2 mM dNTPs, and 1.5 mM MgCl₂. 38 pmoles of JBam24 (60 pmoles/μl) were added to each tube, and the reactions were incubated in a Perkin-Elmer 2400 thermocycler at 72°C for 3 minutes. 1 unit of Taq DNA polymerase (5 units/μl) was added, and the mixture was incubated again at 72°C for 5 minutes. The PCR amplification consisted of 35 cycles of 95°C for 1 minute and 72°C for 3 minutes, with a final extension at 72°C for 10 minutes (followed by a 4°C hold). 5 μl of each primary hybridization PCR reaction was checked on a 2% agarose gel. The three primary hybridization PCR amplifications were pooled and purified using a Microcon microconcentrator (Microcon Inc.), to a total volume of 25 μl in sterile ddH₂O.
2.2.4.3 RDA secondary hybridization

Prior to the secondary hybridization, the JBam adaptors are removed from 1 μg of the concentrated primary hybridization PCR product and the RBam adaptors are removed from 3 μg of the female driver representation (from section 2.2.4.1) by digestion with BamHI, then the DNA was separated from the adaptors by running the digested DNA in a 2% Nu-Sieve agarose gel and purified via a Tip-20 column (Qiagen, Inc.), as described in section 2.2.4.1. The purified DNA was then concentrated using a Microcon microconcentrators (Microcon Inc.) to a total volume of 20 μl.

Two reactions were created to ligate the purified primary hybridization PCR product to the “NBam” adaptor set. In each ligation reaction, 10 μl of the primary hybridization PCR product was combined with 7.5 μl of each of the “N” set of adaptors (NBam24 and NBam12, 60 pmoles/μl), 2 μl of dd H2O, and 3 μl of 10X ligase buffer (New England Biolabs) in a 0.5 ml PCR reaction tube. A COY Tempcycler II machine was used to heat the reactions to 50°C, then cool them gradually to 10°C over one hour. The reactions were cooled on ice, then 1 μl of T4 DNA ligase (400 units/μl, New England Biolabs) was added and the reaction tubes were incubated at 14°C overnight (approximately 18 hours). The newly ligated primary hybridization product was combined with the 3 μg of purified female driver representation and concentrated using a Microcon microconcentrator (Microcon Inc.) into 4 μl of 3X EE buffer. The secondary hybridization and subsequent fill-in and PCR amplifications were carried out as described for the primary hybridization in section 2.2.4.2. 5 μl of each of the three secondary hybridization PCR products were electrophoresed on a 2% agarose gel.
2.2.4.4 Cloning RDA products

Prior to their cloning into plasmid vectors, each of the six RDA products were isolated by cutting them out of an agarose gel with a razor blade and purifying the DNA from the gel slice via the GeneClean Spin kit (Bio 101, Inc.). The purified DNA from each RDA product was then used as a template for a PCR reaction with the NBam set of oligonucleotides as primers to enrich the amount of each product. The reamplification PCR products were then cloned into the pPCRScript Amp SK(+) plasmid vector using the PCR-Script™ Amp Cloning Kit, according to the manufacturer's specifications, and transformed into DH10B *E. coli* bacteria via electroporation. Transformants were then plated on LB agar medium plates (100 mm x 15 mm) impregnated with ampicillin (final concentration is 100 μg/ml) and Xgal substrate (final concentration is 40 μg/ml). Potential transformants (grown into white colonies) were picked from the LB plates and grown as liquid cultures in Terrific Broth (TB, from Bio101, Inc.) supplemented with ampicillin. The liquid bacteria cultures were used for the extraction of plasmid DNA using the Qiaprep kit (Qiagen Inc.).

2.2.5 λgt10 genomic DNA library construction and screening

2.2.5.1 Library construction

1 μg of genomic DNA from a single *B. dorsalis* adult male fly was digested with the restriction enzyme EcoRI (Roche Molecular Biochemicals). Restriction fragments approximately 5 kb in size (as sized by comparison to standard DNA size ladders) were cut out of an agarose gel with a clean razor blade and purified from the gel slice using the
GeneClean Spin Kit (Bio101, Inc.). The *B. dorsalis* male genomic DNA λgt10 minilibrary was constructed by ligating the 5 kb *B. dorsalis* EcoRI-generated genomic DNA fragments to EcoRI-digested λgt10 phage arms (Promega) with T4 DNA ligase (Roche Molecular Biochemicals). The resulting ligation products were packaged into phage particles using the Gigapack III Gold Packaging Extract (Stratagene). The packaging reaction was then placed into 500 μl of SM buffer (100 mM NaCl, 8 mM MgSO₄, 50 mM Tris pH 7.5, 0.01% gelatin) with 20 μl of chloroform and stored at 4°C.

### 2.2.5.2 Library screening

Library screening was done using the plaque lift method of Benton and Davis (1977). A phage “adsorption mix” was prepared, combining the phage particles with the host bacteria to allow the phage to be incorporated by the host bacteria. The phage adsorption mix consisted of 800 plaque forming units (pfu) of the *B. dorsalis* male library and 200 μl of C600 *hfl rₚmₚ* host bacteria (from an overnight liquid culture of C600 *hfl rₚmₚ* bacteria grown in TB broth and resuspended in 10 mM MgSO₄). The phage adsorption mix was incubated for 20 minutes at 37°C. The adsorption mix was then combined with 3 ml of molten NZCYM top agar medium (first cooled to approximately 50°C) and poured onto a 100 mm x 15 mm NZCYM agar medium plate to form a thin layer of uniform thickness. Once the NZCYM top agar hardened, the plate was incubated at 37°C for 4-5 hours.
Duplicate plaque lifts were made of the library plate with MagnaGraph nylon membrane circles (82 mm diameter, from Osmonics Inc.). The "primary" lift was generated by leaving the membrane on the plate for 1 min. Next, a duplicate lift was generated by leaving the membrane on the plate for 5 min. The DNA on the membranes was denatured for 5 min. at room temperature by placing them DNA side up on filter paper (Fisher) amply moistened with denaturing buffer (0.5 M NaOH, 1.5 M NaCl). The membranes were neutralized in neutralizing buffer for 5 min. (0.5 M Tris pH 7.0, 1.5 M NaCl) in the same manner as the denaturing step and immersed in 2X SSC (0.3 M NaCl, 0.03 M Na citrate) for 3 min. The membranes were then air dried and the DNA was permanently bound to the membrane via UV crosslinking (Stratalinker® 1800 UV crosslinker from Stratagene, “autocrosslink” feature preset to 1200 microjoules x 100). The membranes were prehybridized in a sealed plastic bag with 20 ml of prehybridization solution (50% deionized formamide; 5X SSC; 2X blocking solution; 0.1% N-laurosarcosine, Na salt; 0.02% SDS) for 1 hour at 42°C. A "probe solution", consisting of digoxigenin-labeled probe DNA in 10 ml of prehybridization solution was denatured at 95°C for 10 min. The prehybridization solution was replaced with the probe solution and the membrane was incubated overnight at 42°C.

The following day, the probe solution was recovered, and the membranes were washed at high stringency as follows: twice for 15 min. in 2X SSC and 0.1% SDS at room temperature, then twice for 15 min. in 0.5X SSC and 0.1% SDS at 65°C. The hybridized probe was then detected using the Dig Nonradioactive DNA Hybridization and Detection kit (Roche Molecular Biochemicals). Colorimetric detection took place overnight at room temperature.
2.2.5.3 Phage DNA purification and subcloning

Phage plaques that showed a hybridization signal were punched out of the library plate using a wide bore micropipettor tip (1 ml capacity). These plaque "picks" were stored in 750 μl of SM buffer (100 mM NaCl, 8 mM MgSO₄, 50 mM Tris pH 7.5, 0.01% gelatin) and 20 μl of chloroform at 4°C. The phage from these picks were titered to determine the phage concentration in the number of plaque forming units per ml of phage suspension (pfu/ml). Approximately 100 pfu were plated onto fresh NZCYM agar plates (100 mm x 15 mm). Duplicate plaque lifts were generated as previously described, and the membranes were probed with the RDA product 1 digoxigenin-labeled probe to identify single positive plaques. Individual plaques were then picked from each of these purification plates and stored in 750 μl of SM buffer and 20 μl of chloroform at 4°C. One purified phage pick from each original library pick was plated on a fresh NZCYM agar plate to produce confluent plates.

Confluent plates were overlaid with 3 ml of SM buffer to collect the phage particles from the plaques, and incubated at 4°C overnight. The SM buffer was recovered the following day and phage DNA was purified from the recovered SM buffer via the Lambda DNA purification kit from Stratagene.

To subclone λgt10 cloned inserts of interest, DNA from each of the λgt10 clones was digested with the restriction enzyme EcoRI (Roche Molecular Biochemicals) to separate the inserted DNA from the phage arms. The digested DNA was then combined with EcoRI-digested pUC19 plasmid DNA. The mixture was precipitated (3M sodium acetate and 95% ethanol) overnight. The DNA was then pelleted via centrifugation
(14,000 rpm for 15 min.), washed with 70% ethanol, and air dried. The DNA was resuspended with sterile dH₂O and T4 DNA ligase (Roche Molecular Biochemicals) was added. This ligation mixture was incubated overnight at 4°C. The ligated DNA was transformed into DH10B cells via electroporation. Transformants were selected and plasmid DNA was purified as described in section 2.2.4.4 ("Cloning RDA products").

**2.2.6 Southern blots, probe hybridization, and probe detection**

Restriction enzyme-digested *B. dorsalis* male and female genomic DNA (from mass genomic DNA preparations) were run on a 0.8% agarose gel (15 cm x 20 cm) overnight at a low voltage (typically 30 V for approximately 18 hours). The DNA was denatured by immersing the gel in denaturing buffer (0.5 M NaOH, 1.5 M NaCl) for 45 min. at room temperature with gentle shaking. Neutralization was accomplished by immersing the gel in neutralizing buffer (0.5 M Tris pH 7.0, 1.5 M NaCl) for 45 min. with gentle shaking, followed by equilibration in 20X SSC (3 M NaCl, 0.3 M Na citrate) for 30 min. with gentle shaking. The DNA was then transferred to a nylon membrane sheet (Osmonics Inc.) via the Turboblotter DNA transfer system (Schleicher and Schuell) overnight according to the manufacturer's specifications. The DNA was UV crosslinked to the membrane as previously described. The membrane was prepared for probe hybridization by blocking in prehybridization solution (50% deionized formamide; 5X SSC; 2X blocking solution; 0.1% N-laurosarcosine, Na salt; 0.02% SDS) for 1 hour at 42°C in a sealed plastic bag. A probe solution (digoxigenin-labeled probe DNA in 10 ml of prehybridization solution) was denatured at 95°C for 10 min. The prehybridization solution was removed and replaced by the probe solution and the membrane was
incubated at 42°C overnight. The following day, the probe solution was recovered and two washes were performed at high stringency: twice for 15 min. in 2X SSC and 0.1% SDS at room temperature, then twice for 15 min. in 0.5X SSC and 0.1% SDS at 65°C. Detection of hybridized DNA probe was done using the Dig Nonradioactive DNA Hybridization and Detection Kit (Roche Molecular Biochemicals) as previously described. The duration of the colorimetric detection varied for each Southern blot, from overnight (approximately 18 hours) to a maximum of 36 hours.

2.2.7 PCR amplification from B. dorsalis embryos, third instar larvae, pupae, and adult flies

GeneAmp PCR reagents (Roche Molecular Biochemicals) were used for all PCR reactions except as indicated in long PCR reactions. Each PCR reaction was 25 μl, with approximately 10 ng of genomic DNA as the template (from larvae, pupae, and single adult flies) or 10 μl of InstaGene Matrix extract from embryo genomic DNA extractions, 1X PCR Buffer II (consisting of 50 mM KCl, 10 mM Tris-HCl, pH 8.3), 5 pmol of each primer, 0.2 mM of dNTPs, 1.5mM of MgCl₂, and 1 U of AmpliTaq DNA polymerase. The conditions for the PCR reactions consisted of an initial denaturation at 94°C for 1 min, followed by 30-40 cycles of denaturation at 94°C for 1 min., primer annealing at 56.5°C for 30 sec., and primer extension at 72°C for 1 min. PCR products were separated on 1.0% agarose gels and visualized by staining with ethidium bromide.
Long template PCR reactions were done using the Expand Long template PCR System (Roche Molecular Biochemicals). Each reaction was 50 μl, consisting of approximately 10 ng of genomic DNA (from larvae, pupae, and single adult flies) or 10 μl of InstaGene Matrix extract from embryo genomic DNA extractions, 1X PCR Buffer 1 from the Expand system kit (containing 1.75 mM MgCl₂), 0.3 mM of dNTPs (Applied Biosystems), and 0.75 μl of the Expand kit enzyme mix.

The conditions for the long PCR reactions (as recommended by the Expand Long Template PCR kit) consisted of an initial denaturation at 94°C for 2 min., followed by 10 cycles of denaturation at 94°C for 10 sec., primer annealing at 56.5°C for 30 sec., and primer extension at 68°C for 4 min., then 20 cycles of denaturation at 94°C for 10 sec., primer annealing at 56.5°C for 30 sec., and an initial primer extension at 68°C for 4 min., with 20 sec. added to the primer extension time after each cycle. PCR products were visualized by running them on 1.0% agarose gels stained with ethidium bromide.


2.2.8 DNA sequencing and analysis

Cloned DNAs were sequenced by an Applied Biosystems automated sequencer using the dye-terminator, cycle sequencing method with the M13 universal (-20) and reverse (-24) or T3 and T7 sequencing primers. Sequences were analyzed using the program MacVector (Kodak) to design oligonucleotide primers and the DNASTar software package (Lasergene), which includes the program EditSeq. EditSeq was used to perform routine editing of sequence data and ORF searches within sequences, and the PileUp sequence alignment program in the Wisconsin GCG software package was used to perform all nucleotide and amino acid sequence alignments included in this dissertation.

GenBank sequence database (http://www.ncbi.nlm.nih.gov) searches were performed using the BLASTN (for nucleotide sequence searches) and BLASTX (for amino acid sequence searches using a nucleotide sequence) search routines, using standard search parameters. BLAST (Basic Local Alignment Search Tool) generates local alignments between the submitted sequence (the “query”sequence) and database sequences. The BLASTN routine searches GenBank nucleotide sequences with a nucleotide query sequence. The BLASTX routine searches GenBank protein sequences with a nucleotide query sequence.
Chapter 3
Results

3.1 Initial attempts to isolate *B. dorsalis* Y chromosome sequences

Several attempts were made to isolate *B. dorsalis* Y chromosome DNA. One of the first emulated work done by Anleitner and Haymer (1992) to isolate male-specific repetitive DNA in *C. capitata*. Duplicate plaque lifts of a *B. dorsalis* λgt10 genomic DNA library were generated and differentially screened with probes of EcoRI digested *B. dorsalis* male vs. female genomic DNA. A number of clones were isolated that showed hybridization signal with the male genomic DNA probe but not the female genomic DNA probe. However, further characterization of the sequences isolated did not show any evidence for a male bias (results not shown).

A second approach was to use pulse field gel electrophoresis to separate *B. dorsalis* male genomic DNA digested with rare cutter restriction enzymes. Genomic DNA was extracted from the heads to obtain unpolytenized material of adult male and female *B. dorsalis*. The genomic DNA was digested with the restriction enzyme NotI and separated on a 0.7% agarose gel for one week in 1X TBE. The genomic DNA was then transferred to a nylon membrane via Southern blotting.
A probe consisting of digoxigenin-labeled *B. dorsalis* adult male genomic DNA (extracted from whole flies) was generated. The probe was initially used to probe a series of Southern blots of *B. dorsalis* female genomic DNA to eliminate DNA sequences common to both males and females. The probe was then applied to the pulse field gel Southern blot. The resulting probe hybridization signal appeared as smears in both the blotted male and female genomic DNA, without any sex-specific signal (results not shown).

Ultimately, a method of genomic DNA subtraction that mimicked a protocol designed for use with cDNA proved to be the most productive. Specifically, the technique of Representational Difference Analysis, tailored especially for the comparison of highly complex genomes, was successfully used.

### 3.2 Representational Difference Analysis (RDA) products

Following the first round of tester-driver hybridization and PCR difference enrichment, a number of distinct PCR products were present, ranging in size from just under 400 bp to approximately 550 bp (Figure 4a). Three different dilutions of the secondary (2°) hybridization product were made (1:10, 1:20, and 1:40 in sterile dH₂O) for the PCR difference enrichment step. The PCR difference enrichment from the 1:40 dilution gave the best resolution of the RDA product bands on the agarose gel. From this difference enrichment of the secondary hybridization product, six distinct RDA products were recovered (Figure 4b). Their sizes (as approximated by extrapolation from a standard 1 kb DNA ladder run concurrently on the gel) range in size from 360 to 550 bp, as depicted in Figure 4b.
3.2.1 Characterization of RDA product clones

Two different methods were used to clone RDA products 1-6. The first was to perform a ligation to the pPCRScript SK(+) plasmid vector using a mixture of all six. The second approach was to purify each of the six products individually, excising the individual bands from an agarose gel, purifying the DNA from the gel slice, and reamplifying the RDA product via PCR using the purified DNA as the template. Reamplified products were ligated to the pPCRScript SK(+) plasmid. Judging by the relatively large number of transformants, it appeared that the second method was more successful. From this, six to eight individual clones were isolated from each of the six original RDA products. Although a number of these clones appeared to contain inserts of comparable size to the original RDA product, an even greater number were not of the expected size. Sequencing was performed, using T3 and T7 plasmid vector sequencing primers, on a select group of these clones carrying inserts that appeared to be the same size as the corresponding original RDA product.

3.2.1.1 Analysis of clones from RDA products 4 and 5

Clones for RDA products 4 and 5 that appeared to carry inserts of the anticipated size (approximately 429 and 388 bp, respectively) were chosen for sequencing. Comparison of the sequence information obtained from these clones to the GenBank sequence database using the BLASTN and BLASTX search algorithms showed no significant matches to the RDA product 4 clone (Figure 5). A GenBank search with the
Figure 4. Agarose gel pictures of *B. dorsalis* adult RDA PCR difference enrichment following (a) the RDA primary hybridization and (b) the secondary hybridization. 1:10, 1:20, and 1:40 dilutions of the primary and secondary hybridization products were made prior to the PCR difference enrichment. Distinct products ranging from 350-600 bp are evident in both hybridization enrichment PCRs. The products of the second hybridization are labeled as RDA products 1-6, as indicated.
RDA product 1 (550 bp)
RDA product 2 (520 bp)
RDA product 3 (500 bp)
RDA product 4 (450 bp)
RDA product 5 (400 bp)
RDA product 6 (360 bp)
Figure 5. GenBank sequence database search results for the RDA product 4 plasmid clone using the BLASTN and BLASTX search algorithms. The sequence of the clone is also shown. The best matches are shown for those search results with a large number of matches, as indicated. The BLASTN or BLASTX alignments for the best matches are also shown below the search results.
RDA product 4 clone sequence (429 bp)

1   GATCCATAAA AGCGTCACAA ATTTGAAATG GATTTGCGCA CGTTTCGGTG
51  TTGGCAGCGC ACATCAAGTT ACTTATGTAC ATAAGTAGGT ATATATGTAC
101 ATACATTATA TGATTTTAGT TATATTGCTA TCGCCATGTC CACATGGTAT
151 TCAAGAATCG CATTCTTACA CGTAATTCTA CATTATTCTC ATCTAAGGAT
201 ACAGACAGGT CATAAGGTTCC ATAAAAAGTA TTGCAATTGT TCTGAAGGTA
251 TACATACTA TTAATACAA TATAGTGCTA TGGAGTGGA TGAAATCGGA
301 AGATAATCCC GATCACTGCC CTTACAACGG TACTGTTCAA AACTACTAAA
351 ACAAACGAGT CATGCGGTCC ATAAAAAGTA TTGCAATAGT TCTGAAGGTA
401 AGATAATCCC GAGGGACTT TATTTGATC

Base composition: A = 34.0%, T = 29.8%, G = 18.0%, C = 18.2%

RDA product 4 clone: GenBank BLASTN search results (six highest matches)

Sequences producing significant alignments:  

<table>
<thead>
<tr>
<th>Description</th>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>122832100</td>
<td>gb</td>
</tr>
<tr>
<td>gi</td>
<td>18030094</td>
<td>gb</td>
</tr>
</tbody>
</table>

Score = 46.1 bits (23), Expect = 0.039  
Identities = 23/23 (100%)  
Strand = Plus / Minus

Query: 90   tatatatgtacatacatatatag 112  
Sbjct: 148598 tatatatgtacatacatatatag 148576

Score = 46.1 bits (23), Expect = 0.039  
Identities = 23/23 (100%)  
Strand = Plus / Minus

Query: 90   tatatatgtacatatagatatag 112  
Sbjct: 21157 tatatatgtacatacatatatag 21135

78
**RDA product 4 clone: GenBank BLASTX search results (seven highest matches)**

<table>
<thead>
<tr>
<th>Sequence ID</th>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>325513</td>
<td>gb</td>
</tr>
<tr>
<td>gi</td>
<td>9628375</td>
<td>ref</td>
</tr>
<tr>
<td>gi</td>
<td>662096</td>
<td>gb</td>
</tr>
<tr>
<td>gi</td>
<td>8778093</td>
<td>gb</td>
</tr>
<tr>
<td>gi</td>
<td>9633142</td>
<td>ref</td>
</tr>
<tr>
<td>gi</td>
<td>11278244</td>
<td>pir</td>
</tr>
<tr>
<td>gi</td>
<td>13537662</td>
<td>emb</td>
</tr>
</tbody>
</table>

**gi|325513|gb|AAA65581.1| (M68963) origin-binding protein [Human herpesvirus 6]**

- Length = 703
- Score = 31.6 bits (70), Expect = 2.1
- Identities = 18/69 (26%), Positives = 36/69 (52%), Gaps = 6/69 (8%)
- Frame = -3

**Query:**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L349 LVVLNSTVVRAGIIFRPHPCTCIYILCNYVY--------LQnychtygphdfsfpdFLDE 188</td>
<td>349</td>
<td>2.1</td>
</tr>
<tr>
<td>VIAMDATLTHVEFFAFLKDPQILRNTFVSAMFSNRIAYFCDFGGKEFSPPARLE 229</td>
<td>349</td>
<td>2.1</td>
</tr>
<tr>
<td>KIVKFDCKM 161</td>
<td>187</td>
<td>2.1</td>
</tr>
<tr>
<td>DKLRWDKKL 230</td>
<td>230</td>
<td>2.1</td>
</tr>
</tbody>
</table>
gil9628375|ref|NP_042966.1| (NC_001664) origin binding protein [Human herpesvirus 6]
gi|1709431|sp|P52378|OBP_HSV6U Replication origin binding protein (OBP)
gi|854052|emb|CAA58365.1| (X83413) U73, origin binding protein [Human herpesvirus 6]
Length = 780
Score = 31.6 bits (70), Expect = 2.1
Identities = 18/69 (26%), Positives = 36/69 (52%), Gaps = 6/69 (8%)
Frame = -3

Query: 349 LVVLNSTVRAVIGFPLTCLYILCNMYVY------LQNYCNTFYGPHDFVPLDE 188
++ +++T+ R V+ F P T I ++ N +V + +C+TF+G SF E
Sbjct: 170 VIAMDATLTRHVVEFFA FKPTQIALIRNTFVSAMFNSR VAYFCDTFFGKEFSFARLE 229

Query: 187 KIVKFDCMK 161
+++D K+
Sbjct: 230 DKLRWDKEL 238

gi|662096|gb|AAA68464.1| (U13194) replication origin-binding protein [Human herpesvirus 6] Length = 659
Score = 31.6 bits (70), Expect = 2.1
Identities = 18/69 (26%), Positives = 36/69 (52%), Gaps = 6/69 (8%)
Frame = -3

Query: 349 LVVLNSTVRAVIGFPLTCLYILCNMYVY------LQNYCNTFYGPHDFVPLDE 188
++ +++T+ R V+ F P T I ++ N +V + +C+TF+G SF E
Sbjct: 49 VIAMDATLTRHVVEFFA FKPTQIALIRNTFVSAMFNSR VAYFCDTFFGKEFSFARLE 108

Query: 187 KIVKFDCMK 161
+++D K+
Sbjct: 109 DKLRWDKEL 117

gi|8778093|gb|AAF79202.1| (AF265225) CTD phosphatase-like protein [Emericella nidulans] Length = 409
Score = 31.6 bits (70), Expect = 2.1
Identities = 14/43 (32%), Positives = 25/43 (57%), Gaps = 3/43 (6%)
Frame = +2

Query: 203 KRVMRSIKSIAVL---KVYIHI+DIDTQWMKSEEDNPDCP 322
KR + + + +++V+ + IH D G+WM +DNP+H P
Sbjct: 48 KRRLANRKLSDLVLDLDQTI1HAADVPTIGEKNADKDNPHAP 90
RDA product 5 clone sequence showed significant matches to a number of Drosophila serine protease sequences, particularly with the BLASTX search algorithm (Figure 6). Among the top protein sequence matches were the Drosophila serine protease SER4 precursor (GenBank accession number AF006639 and a number of “unidentified” Drosophila gene products appearing to have identical sequences to the SER4 serine protease sequence.

Southern blots of *B. dorsalis* genomic DNA digested with the restriction enzymes EcoR1, PstI, and Hinfl were generated to determine the representation of these cloned sequences in the *B. dorsalis* genome and to detect any possible differences in their representation between males and females. A blot probed with the RDA product 4 clone showed the presence of single copies of the sequence in the *B. dorsalis* male and female genomes, without a difference in the hybridization pattern in males vs. females (Figure 7). A duplicate blot that was probed with the RDA product 5 clone shows a repetitive representation of the sequence in both the male and female genomes (Figure 7). As with the blots probed with the RDA product 4 clone, no discernable difference in the hybridization patterns was observed in males vs. females.

### 3.2.1.2 Characterization of RDA products 1, 2, 3, and 6

The representations of the RDA products 1, 2, 3, and 6 were also determined by using them to probe Southern blots. A blot containing *B. dorsalis* male and female genomic DNA digested with the restriction enzyme EcoRI was probed with RDA product 1. This produced two bands, estimated to be 5 and 10 kb in size, respectively, present in the male genomic DNA hybridization pattern that were not seen in the corresponding
Figure 6. GenBank sequence database search results for the RDA product 5 plasmid clone using the BLASTN and BLASTX search algorithms. The sequence of the clone is also shown. The best matches are shown for those search results with a large number of matches, as indicated. The BLASTN or BLASTX alignments for the best matches are also shown below the search results.
**RDA product 5 clone sequence (388 bp)**

```
1  GATCCACTCC GTTTCTTTCT CTGGCAACAT TCAAGCCATC AGACTGCCAT
51  CTTTCAACCA ATACAGCAAC TATGAGGGAC AATGGGCTAC CGCATCTGGC
101 TGGGGTGGCA CTTCAGGCAA CAACCAAGAC CATTTGCAAT ACGTCAGTGT
151 CCAGGTTGACT TCCAACCGCG AATGGGGCGG CATCTATGSC AGCAACACTG
201 TCACCGACAA CACCATCTGT GTTTCCACAA ACGGCCAGCG TTCAACTTG
251 GGTGGTGACG CTGGTTGCCC ATGGCTGTG GACAACAACC AAGTGTTGAT
301 TGTTTGACT TCATTCGTTG CTGCCGCTGG TTGCACCGCT GGTCTGCCAT
351 CTGGTTTCCA ACCTGTCAGT CGCCATTGCG ACTGGATC
```

Base composition: A = 21.9%, T = 25.0%, C = 28.4%, G = 24.7%

**RDA product 5 clone: GenBank BLASTN search results (five highest matches)**

<table>
<thead>
<tr>
<th>Sequence ID</th>
<th>Description</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>8037913</td>
<td>gb</td>
<td>AF233728.1</td>
</tr>
<tr>
<td>gi</td>
<td>4210801</td>
<td>emb</td>
<td>Y17904.1</td>
</tr>
<tr>
<td>gi</td>
<td>17384503</td>
<td>emb</td>
<td>AL56275.1</td>
</tr>
<tr>
<td>gi</td>
<td>2930981</td>
<td>gb</td>
<td>L15632.1</td>
</tr>
<tr>
<td>gi</td>
<td>184157021</td>
<td>ref</td>
<td>NM 110687.1</td>
</tr>
<tr>
<td>gi</td>
<td>119031223</td>
<td>gb</td>
<td>AY074713.1</td>
</tr>
</tbody>
</table>

**gi|8037913|gb|AF233728.1|AF233728 Agrotis ipsilon AiC2 chymotrypsinogen mRNA, complete cds**

Length = 945

Score = 44.1 bits (22), Expect = 0.14
Identities = 22/22 (100%)
Strand = Plus / Plus

Query: 296 ttgattggtgtgacttcattcg 317

Sbjct: 760 ttgattggtgtgacttcattcg 781

**gi|4210801|emb|Y17904.1|PC017904 Phaedon cochleariae mRNA for chymotrypsin**

Length = 918

Score = 44.1 bits (22), Expect = 0.14
Identities = 25/26 (96%)
Strand = Plus / Plus

Query: 246 cttgcggtgtgactctgtggccca 271

Sbjct: 688 cttgcggtgtgactctgtggccca 713
gi|17384503|emb|AL596275.10| Human DNA sequence from clone RP11-183G22 on chromosome 1, complete sequence [Homo sapiens] Length = 139273

Score = 44.1 bits (22), Expect = 0.14
Identities = 22/22 (100%)
Strand = Plus / Minus

Query: 283 caacaaccaagtgttgattggt 304

Sbjct: 113274 caacaaccaagtgttgattggt 113253

gi|293098|gb|L15632.1| LUCTRYPS4A Lucilia cuprina alpha-trypsin 3 (3' end) and 4 (complete cds) genes Length = 3109

Score = 44.1 bits (22), Expect = 0.14
Identities = 22/22 (100%)
Strand = Plus / Plus

Query: 254 ggtgactctgtggtggcccattgg 275

Sbjct: 1583 ggtgactctgtggtggcccattgg 1604

RDA product 5 clone: GenBank BLASTX search results (eleven high matches)

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>gi</th>
<th>Ref</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
</table>
| gi|17986193|ref|NP 524555.1| (NM_079831) Serine protease 3;... 103 6e-22
| gi|18079291|ref|NP 524554.1| (NM_079830) Serine protease 1;... 102 8e-22
| gi|85145|pir|JS0260| serine proteinase (EC 3.4.21.-) 1 prec... 102 8e-22
| gi|2463064|emb|CAA72952.1| (Y12273) chymotrypsin-like prote... 101 2e-21

84
gil213581031|ref|NP 648012.1| (NM_139755) CG6467 gene product [Drosophila melanogaster]
gi|7295397|gb|AAF50714.1| AE003564) CG6467-PA [Drosophila melanogaster]
gi|16769776|gb|AAL29107.1) (AY061559) LP10916p [Drosophila melanogaster]
Length = 271

Score = 113 bits (283), Expect = 4e-25
Identities = 59/133 (44%), Positives = 81/133 (60%), Gaps = 4/133 (3%)
Frame = +2

Query: 2 IHSVFSFGNIQAIRLPSE-NQYSNYEQQATASGWGTSGNNQD---HLQYVSQVISNS 169
I + S S I A++LPS N YS + G A ASGWG TS + ++LQYV + VI+N+
Sbjct: 133 IPATSSSSRISAVKLPISNSYSTFVGDVAVASGWGRTSDTSSGVATNLQYVDLTVITNT 192

Query: 170 ECASYGGSSTVTDTNCIVSTNXXXXXXPLAVDNQVLTGTSFVAAGCTAGLP 349
+CA YG+S VTD+T+CV+T PL + ++ IG+TSF A+AGC G P
Sbjct: 193 KCAQTYGTSVTVTDSTLCVATTDKSTCNGDSGPLVLKSSSEQIGLTSFGASAGCEKYP 252

Query: 350 SGFQRVSRLDLWI 388
+ F RV+ +LDWI
Sbjct: 253 AAPTRTVSYLDWI 265

gil|2353699|gb|AAC14351.1) (AF006639) serine protease SER4 precursor [Drosophila melanogaster] Length = 265

Score = 112 bits (279), Expect = 1e-24
Identities = 58/128 (45%), Positives = 78/128 (60%), Gaps = 2/128 (1%)
Frame = +2

Query: 11 VSFSGNIQAIRLPSEFNYSYEQQATASGWGTSGNNQDHLQYVSQVISNSCASV 184
V F + + LPSFN +Y+ Y+ WA A GWG T+G+ D ++ V +Q+ISNSEC+
Sbjct: 131 VDFWRMVKVELPSFNDRYMYNDNYWAVACGWGLTTAGSQPDWMECVDLQ11SNSCSRT 190

Query: 185 YGSSTVTDTNCIVSTNXXXXXXPLAVDNQVLTGTSFVAAGCTAGLPSEGFQR 364
YG T D +CVST+ PL + + L+GVT+S+V+ GCTAGLPSEF R
Sbjct: 191 YG--TQFDGILCVTSGGKSTCSGSGSGLVLHHDGGLVGVSNGCTAGLPSEFTR 248

Query: 365 VSRHLDWI 388
V+ LDWI
Sbjct: 249 VTNQLDWI 256
gi|17986193|ref|NP_524555.1|(NM_079831) Serine protease 3; serine protease [Drosophila melanogaster]

Score = 103 bits (256), Expect = 6e-22
Identities = 52/129 (40%), Positives = 74/129 (57%), Gaps = 3/129 (2%)
Frame = +2

Query: 11 VSFSGNIQAIRLPSFN-QYSNYEGQWATASGWGGT SGNNQDHLQYVSQVINSSECAS 181
V F + I LPS + +Y++YE W A+GWG N + L+ V ++VIS +EC +
Sbjct: 136 VDFYSLVKELPSLDDRNYENWQAAGWAIYDGNSVVELDRLKVISVABQA 195

Query: 182 VYGSSTVTNDTICVSTNXXXXXXXXXXPLAVDNSQVLIGVTSFVAAGCTAGLPGFQ 361
YG+ +NTIVC T PL LIG+TSFVA+GC G P+GF
Sbjct: 196 YYGTDTASENICTPETPGKATCQHSGGPLVTKEKDLIGITSFVSAYGCQVGPAGFT 255

Query: 362 RVSRHLDWI 388
RV+++L+WI
Sbjct: 256 RVTKYLEWI 264

---

gi|18079291|ref|NP_524554.1|(NM_079830) Serine protease 1; Serine protease 5 [Drosophila melanogaster]

Score = 102 bits (255), Expect = 8e-22
Identities = 54/129 (41%), Positives = 76/129 (58%), Gaps = 3/129 (2%)
Frame = +2

Query: 11 VSFSGNIQAIRLPSFN-QYSNYEGQWATASGWGGT SGNNQDHLQYVSQVINSSECAS 181
V F + +LPS+N +Y +Y G WA A+GWG + +D LQ V VQ+IS S+C+
Sbjct: 132 VDFWSLNKLPSLDDRNYENWQAAGWAIYDGNSVVELDRLKVISVABQA 191

Query: 182 VYGSSTVTNDTICVSTNXXXXXXXXXXPLAVDNSQVLIGVTSFVAAGCTAGLPGFQ 361
++DN IC+T+ PL +L+GVTSF AAGC G P+F
Sbjct: 192 TW---SLHDNMICINTDGGKSTCGDGGPPVTGDNRLVGVTSFGSAAGCQSGAPAVFS 248

Query: 362 RVSRHLDWI 388
RV+++L+WI
Sbjct: 249 RVTKYLEWI 257
Figure 7. Southern blots of *B. dorsalis* adult male and female genomic DNA digested with the restriction enzymes EcoRI, PstI and HinfI probed with (a) the RDA product 4 clone and (b) the RDA product 5 clone. The RDA product 4 probe produced single hybridization signals in both males and females, while the RDA product 5 probe produced three bands in both sexes. DNA sizes were estimated via comparison to standard DNA molecular weight markers.
Figure 8. Southern blot of *B. dorsalis* adult male and female genomic DNA digested with the restriction enzyme EcoRI. This blot shows the repetitive pattern of RDA product 1 probe hybridization, with two male-specific bands (estimated to be 5 and 10 kb in size).
female hybridization pattern (Figure 8). This Southern blot was replicated and resulted in
the appearance of the same male-specific bands. Duplicate blots were also probed with
RDA products 2, 3, and 6 (results not shown). All showed identical repetitive
hybridization patterns in males and females.

3.3 Further characterization of RDA product 1

To isolate the genomic segment corresponding to the 5 kb male-specific genomic
DNA fragment seen when RDA product 1 was used to probe *B. dorsalis* genomic DNA, a
male-specific “minilibrary” was constructed. Genomic DNA from a single adult *B.
dorsalis* male was digested with EcoRI and separated on an agarose gel. DNA
approximately 5 kb in size was excised and purified from the gel, then used to construct
the λgt10 minilibrary. The phage titer for this library was estimated to be $9.8 \times 10^4$
plaque forming units per milliliter (pfu/ml).

Approximately 800 pfu of the library were plated onto a 100 mm x 15 mm
NZCYM plate and plaque lifts were probed with RDA product 1. The probe hybridized
to eight distinct plaques in all, varying in hybridization signal intensity (Figure 9).
Plaque lifts of each of the eight phage clones were made and probed to confirm the
positive hybridization. Of the eight purified phage clones (designated 1.1 to 8.1), clones
3.1, 5.1, 6.1, and 8.1 appeared to be the most promising, as they all contained inserts that
were closest to 5 kb in size. Clone 2.1 carried an insert approximately 4.5 kb in size.
Clones 1.1 and 4.1 were not pursued further, as there were difficulties in purifying phage
dNA from these clones. Inserts from the phage clones, except for 1.1 and 4.1, were
subcloned into the plasmid vector pUC19 for further analysis.
Figure 9. Enlarged view of *B. dorsalis* adult male minilibrary plaque lift probed with RDA product 1. Approximately 800 plaques were plated, eight of which (designated 1.1 to 8.1) showed probe hybridization.
3.3.1 Sequence analysis of minilibrary subclones

Partial DNA sequences were obtained from both ends of the plasmid clones 2.1a (the plasmid subclone of phage clone 2.1) and 3.1a (the plasmid subclone of phage clone 3.1). Sequences from the 2.1a clone and one end of the 3.1a clone (517 bp in size, designated “3.1aFor”) showed no significant matches sequences in GenBank using either the BLASTN or BLASTX search routines (Figures 10 and 11, respectively). A BLASTX search of GenBank using the sequence from the other end of 3.1a (550 bp in size and designated “3.1aRev”, sequence shown in Figure 12) showed similarity to the reverse transcriptase sequence of the retrotransposable element R1Dm, a non-LTR retrotransposon present in the 28S rRNA genes of *Drosophila melanogaster* (Jakubczak et al. 1990). The results of this search are shown in Figure 13. Other BLASTX matches included R1 element reverse transcriptases from the fungus gnat, *Sciara coprophila*, (L00945.1) and *Drosophila mercatorum* (AF015277).

3.3.2 *B. dorsalis* Southern blots probed with minilibrary subclones

The insert sequences from phage subclones 2.1a and 3.1a were each used to probe Southern blots of EcoRI-digested genomic DNA to help determine their representation in the *B. dorsalis* males and females. Both blots showed a repetitive representation of the 2.1a and 3.1a sequences in the male and female genomes without any apparent sex-specific bands (2.1a probed blot not shown, 3.1a probed blot shown in Figure 14). These blots were replicated and the results confirmed the identically repetitive representations of the 2.1a and 3.1a sequences in the *B. dorsalis* male and female genomes.
Figure 10. GenBank sequence database search results for the 2.1a plasmid clone using the BLASTN and BLASTX search algorithms. The sequenced ends of the 2.1a clone are designated “2.1aFor” (515 bp in size) and “2.1aRev” (589 bp). The sequences of these ends are also given in the figure. The best matches are shown for those search results with a large number of matches, as indicated. The BLASTN or BLASTX alignments for the best matches are also shown below the search results. There was no sequence similarity found for a BLASTX search of GenBank with the 2.1aRev sequence.
2.1 For sequence (515 bp)

AATTCAATGC TCAACCTGTT ACTCGAGAAG TATGTATAAT GGTGTTCGTA
CGAAAAACAG AGATGTTTGA TATGAAAAAC GCGTGGTTTG AAGACAACTA
ACGAAGTTGA TAATGATGCT TAGAAATGCC CGGTATTTGA ATGAAGATAA
GAGTGGTTCA CATCCTAAG AAGACAAAAA ATAAATAATG ACAACATCTA
CTAAATGCAATTTGATT TTATGCTAT CTCGGCAAT TTTTTATAGG
ACGTGGTTCA CATCCTAAG AAGACAAAAA ATAAATAATG ACAACATCTA
TTATTTGACT GATTT

Base composition: A = 33.2%, T = 32.2%, C = 14.4%, G = 20.2%

2.1 For sequence: GenBank BLASTN search results (six highest matches)

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>229538161</td>
<td>gb</td>
</tr>
<tr>
<td>gi</td>
<td>14578098</td>
<td>gb</td>
</tr>
</tbody>
</table>

Score = 44.1 bits (22), Expect = 0.18
Identities = 22/22 (100%)
Strand = Plus / Plus

Query: 293  aaagtttccatttctctgattt 314
Sbjct: 146461 aaagtttccatttctctgattt 146482

Score = 44.1 bits (22), Expect = 0.18
Identities = 22/22 (100%)
Strand = Plus / Minus

Query: 293  aaagtttccatttctctgattt 314
Sbjct: 105146 aaagtttccatttctctgattt 105125
gi|3628576|gb|AC004371.1|AC004371 Drosophila melanogaster DNA sequence (P1 DS06482 (D76)), complete sequence Length = 87124

Score = 44.1 bits (22), Expect = 0.18
Identities = 22/22 (100%)
Strand = Plus / Plus

Query: 293 aaagttcatttctctgattt 314

Sbjct: 56688 aaagttcatttctctgattt 56709

2.1a For sequence: GenBank BLASTX search results

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>1.7</td>
</tr>
</tbody>
</table>

gi|4704716|gb|AAD28229.1|AF121254 11 (AF121254) transcripti...
Score = 33.1 bits (74), Expect = 1.7
Identities = 17/51 (33%), Positives = 31/51 (60%)
Frame = -3

Query: 333 EIDMENQIREMETFATSQFSTEVNVITHRSITQKSILTQVLVDVNYY 181
Sbjct: 166 QVNFDEQIEDMETLRDM-----LNIIKYHSIDEESINYMLVTHLQYF 211
2.1a Rev sequence (589 bp)

```
1  GATTCCGACT TTGGATTTTT GCCGCAGATA ACGTGGCAGC TGATTTCTCT
51  CTCATAACGC AGGGTTGCTA ATATCGATAT ACTGCAGTAA TTATCAACTT
101 TTATAATATA ATAAACTATT TTGTTGTTT TTGTTACGTT TTGATATTT
201 TATATTGATA AAAATTTGTA TGGAACAAA AGATGTTGCT AAATACATT
251 ATGCGTATTG GCAGAATGGCA ACATTGTTCA AATGAAAACA AAGCAAAGAT
301 TGGCAGGGTT TGCGTTTAGA CCACGTTTTA CACTGTTCAC ACATTTTGGCC
351 GTATATCGAAA GAAAAGGAAA GAAGCAATA GCAGTTTCGA CTGGTGTGGC
401 TATTATGAAA GTAAATTATT CGTTTTTCAA GAAATTTTGC GTTATACGAG
451 AAAACATGTG TAACTAATGT TCTCTCAGTT AGAAATTTAC ATTAATTTTA
501 TGTTTAGAAT ACAGACAGAT ATATATCTAT AAATATTTAC ATATATGTAT
551 ATGGTATATG TGTATATATA TGAATGCGCT ATCAGTATG
```

Base composition: A = 33.5%, T = 37.0%, C = 13.4%, G = 16.1%

2.1a Rev sequence: GenBank BLASTN search results (ten highest matches)
(Note: there was no sequence similarity found with the BLASTX search method)

Sequences producing significant alignments:

| gi|23306055|gb|AC022296.34| Homo sapiens 3 BAC RP11-91K8 (Roswell Park Cancer Institute Human BAC Library) complete sequence Length = 153650 |
| gi|11119726|gb|AF195044.1|AF19584982 Homo sapiens beaded fi... |
| gi|22795317|gb|AC129980.6| Homo sapiens chromosome 15, clone... |
| gi|18308593|gb|AC104420.21| Homo sapiens chromosome 15, clone... |
| gi|21717102|gb|AF165425.41| Homo sapiens chromosome 8 clone... |
| gi|116751901|gb|AC092814.21| Homo sapiens chromosome 1 clone... |
| gi|12324973|gb|AC018848.6|AC018848 Arabidopsis thaliana chr... |
| gi|16730722|gb|AC018849.2|AC018849 Arabidopsis thaliana chr... |
| gi|3449313|db|AB016872.1| Arabidopsis thaliana genomic DNA... |
| gi|23494865|gb|AE014829.1| Plasmodium falciparum 3D7 chromo... |

Score = 50 bits (25), Expect = 0.003
Identities = 25/25 (100%)
Strand = Plus / Minus

Query: 147 ttatttataaatatataaactattt 171
Sbjct: 104040 ttatttataaatatataaactattt 104016
gi|11119726|gb|AP195044.1|AP195849S2 Homo sapiens beaded filament component protein (CP49) gene, partial cds Length = 12473

Score = 50.1 bits (25), Expect = 0.003
Identities = 25/25 (100%)
Strand = Plus / Plus

Query: 147 ttatttataaatatataaactattt 171

Sbjct: 6425 ttatatttataatataaactatttt 6449

---

gi|22795317|gb|AC129980.6| Homo sapiens chromosome 15, clone CTD-2529M12, complete sequence Length = 236391

Score = 48.1 bits (24), Expect = 0.014
Identities = 27/28 (96%)
Strand = Plus / Plus

Query: 148 tatttataaatataaatatatttttag 175

Sbjct: 17066 tatttataaatataaatataattttag 17093
Figure 11. GenBank sequence database search results for one sequenced end of the 3.1a plasmid clone ("3.1aFor", 517 bp) using the BLASTN and BLASTX search algorithms. The 3.1aFor sequence is also given in the figure. The best matches are shown for those search results with a large number of matches, as indicated. The BLASTN or BLASTX alignments for the best matches are also shown below the search results.
3.1a For sequence (517 bp)

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>AATTCGGCTC CTCGTTACAC GCACTGGAAT TTCCCCGTGT CCTCTCACAG</td>
</tr>
<tr>
<td>51</td>
<td>101</td>
<td>AAATTCTAAC TACCTGTGTT AGGTAGTCCA CACAAACACT TAGGTCACTG</td>
</tr>
<tr>
<td>101</td>
<td>151</td>
<td>TTTAGTGCTT AGTGTAGCAC ACACACACTA TTTTCAGGGT TACTACCCTT</td>
</tr>
<tr>
<td>151</td>
<td>201</td>
<td>TTTTGCCACT ATTGTTGTAC ACTCTATTGT ACTACTTTCA GGGTCACTAA</td>
</tr>
<tr>
<td>201</td>
<td>251</td>
<td>CTTTTTTTGC CACTATTGCC CTTTGTGCTT TGTGTAGCAC GCACGCAAC</td>
</tr>
<tr>
<td>251</td>
<td>301</td>
<td>TATTATCTCA CTTGACATT AACTCTATTGT ACTACTTTCA GGGTCACTAA</td>
</tr>
<tr>
<td>301</td>
<td>351</td>
<td>CCTTTTTTGC CACTATTGCC CTTTGTGCTT TGTGTAGCAC GCACGCAAC</td>
</tr>
<tr>
<td>351</td>
<td>401</td>
<td>TATTATCTCA CTTGACATT AACTCTATTGT ACTACTTTCA GGGTCACTAA</td>
</tr>
<tr>
<td>401</td>
<td>451</td>
<td>CCTTTTTTGC CACTATTGCC CTTTGTGCTT TGTGTAGCAC GCACGCAAC</td>
</tr>
<tr>
<td>451</td>
<td>501</td>
<td>CTTTTTTTGC CACTATTGCC CTTTGTGCTT TGTGTAGCAC GCACGCAAC</td>
</tr>
</tbody>
</table>

3.1a For sequence: GenBank BLASTN search results (five highest matches)

<table>
<thead>
<tr>
<th>Score</th>
<th>E Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>0.73</td>
<td>Geobacter metallireducens...</td>
</tr>
<tr>
<td>40</td>
<td>2.9</td>
<td>Homo sapiens chromosome 1 clone ...</td>
</tr>
<tr>
<td>40</td>
<td>2.9</td>
<td>Homo sapiens chromosome 1 BAC RP11-67L3, complete sequence Length = 155925</td>
</tr>
<tr>
<td>40</td>
<td>2.9</td>
<td>Homo sapiens PAC clone RP...</td>
</tr>
<tr>
<td>40</td>
<td>2.9</td>
<td>Human DNA sequence from clone...</td>
</tr>
</tbody>
</table>

Query: 326 ctatcctttttgggactattgt 346

Sbjct: 370522 ctatcctttttgggactattgt 370502

Query: 191 gggtcactaacctttttttgc 210

Sbjct: 10980 gggtcactaacctttttttgc 10961
gi|13162257|gb|AC024908.22| Homo sapiens 10 BAC RP11-446F3 (Roswell Park Cancer Institute Human BAC Library) complete sequence Length = 158349

Score = 40.1 bits (20), Expect = 2.9
Identities = 20/20 (100%)
Strand = Plus / Minus

Query: 115 tagcacacacacactatttt 134
Sbjct: 140731 tagcacacacacactatttt 140712

3.1a For sequence: GenBank BLASTX search results

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>17158198</td>
</tr>
<tr>
<td>gi</td>
<td>15897900</td>
</tr>
<tr>
<td>gi</td>
<td>17158198</td>
</tr>
<tr>
<td>gi</td>
<td>17016492</td>
</tr>
<tr>
<td>gi</td>
<td>19481742</td>
</tr>
</tbody>
</table>

Length = 163
Score = 35.0 bits (79), Expect = 0.45
Identities = 13/36 (36%), Positives = 17/36 (47%)
Frame = +1

Query: 154 CHYCCTLYCTTFRVTNLFCYCALCFV*HARTLLSH 261
Sbjct: 51 CYYCCCYCCYYCCYYYCCYYCCCCY--HCRCHYLH 84

<table>
<thead>
<tr>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>15897900</td>
</tr>
<tr>
<td>gi</td>
<td>13814215</td>
</tr>
</tbody>
</table>

Length = 265
Score = 31.6 bits (70), Expect = 5.0
Identities = 25/62 (40%), Positives = 31/62 (49%), Gaps = 4/62 (6%)
Frame = +2

Query: 95 SLFSAC*CTSTHLFGPPLPPFFATIVVHSVYVLLS-SLTFFA-TIALCALCSTHAYHLT 262
Sbjct: 19 SLLLALPTLYLYFYGYGPFFVKSASFSSLLSSIGLSFFASALSIVTLIFTPLLISYLS 78

Query: 263 WH 268
Sbjct: 79 RH 80
Figure 12. The 550 bp 3.1aRev sequence. The 102 bp sequence showing similarity to the reverse transcriptase of the Drosophila R1 retrotransposon R1Dm, as determined by a BLASTX search of the GenBank sequence database, is highlighted in boldfaced type.

Base composition: A = 27.8%, T = 24.0%, C = 20.2%, G = 28.0%
Figure 13. GenBank sequence database search results for the sequenced end of the 3.1a plasmid clone designated 3.1aRev (550 bp) using the BLASTN and BLASTX search algorithms. The best matches are shown for those search results with a large number of matches, as indicated. The BLASTN or BLASTX alignments for the best matches are also shown below the search results.
### 3.1a Rev sequence: GenBank BLASTN search results (nine highest matches)

**Sequences producing significant alignments:**

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Description</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>21844714</td>
<td>gb</td>
<td>AC122868.2</td>
</tr>
<tr>
<td>gi</td>
<td>9944154</td>
<td>emb</td>
<td>AL353897.7</td>
</tr>
<tr>
<td>gi</td>
<td>22798355</td>
<td>emb</td>
<td>AL807760.6</td>
</tr>
<tr>
<td>gi</td>
<td>24196877</td>
<td>gb</td>
<td>AE011464.1</td>
</tr>
<tr>
<td>gi</td>
<td>23306279</td>
<td>ref</td>
<td>NT_036314.1</td>
</tr>
<tr>
<td>gi</td>
<td>21957778</td>
<td>gb</td>
<td>AE013707.1</td>
</tr>
<tr>
<td>gi</td>
<td>16328260</td>
<td>gb</td>
<td>AC011373.4</td>
</tr>
<tr>
<td>gi</td>
<td>15980975</td>
<td>emb</td>
<td>AJ414155.1</td>
</tr>
<tr>
<td>gi</td>
<td>20805083</td>
<td>dbj</td>
<td>AP003722.3</td>
</tr>
</tbody>
</table>

**gi|21844714|gb|AC122868.2| Mus musculus clone RP23-148H17, complete sequence**

**Length = 202374**

- **Score = 44.1 bits (22)**, **Expect = 0.20**
- **Identities = 22/22 (100%)**
- **Strand = Plus / Plus**

**Query:** 244 tacaaagacatgctcctcagaa 265

**Sbjct:** 5954 tacaaagacatgctcctcagaa 5975

- **Score = 44.1 bits (22)**, **Expect = 0.20**
- **Identities = 22/22 (100%)**
- **Strand = Plus / Minus**

**Query:** 244 tacaaagacatgctcctcagaa 265

**Sbjct:** 168059 tacaaagacatgctcctcagaa 168038

**gi|9944154|emb|AL353897.7| Human DNA sequence from clone RP5-859D17 on chromosome 11p13-14.3**

Contains ESTs, STSs and GSSs. Contains part of a novel gene, complete sequence

[Homo sapiens] **Length = 109891**

- **Score = 42.1 bits (21)**, **Expect = 0.78**
- **Identities = 21/21 (100%)**
- **Strand = Plus / Plus**

**Query:** 419 ctcttctacttttttaaaaca 439

**Sbjct:** 39019 ctcttctacttttttaaaaca 39129
### 3.1aRev sequence: GenBank BLASTX search results (11 highest matches)

<table>
<thead>
<tr>
<th>GenBank Accession</th>
<th>Description</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>103359</td>
<td>pir</td>
<td>1S09111</td>
</tr>
<tr>
<td>gi</td>
<td>140023</td>
<td>sp</td>
<td>P16425</td>
</tr>
<tr>
<td>gi</td>
<td>123542</td>
<td>gb</td>
<td>AAB94027.1</td>
</tr>
<tr>
<td>gi</td>
<td>548540</td>
<td>sp</td>
<td>P03277</td>
</tr>
<tr>
<td>gi</td>
<td>220040</td>
<td>dbj</td>
<td>BAC06460.1</td>
</tr>
<tr>
<td>gi</td>
<td>212960</td>
<td>gb</td>
<td>EAA08234.1</td>
</tr>
<tr>
<td>gi</td>
<td>223315</td>
<td>ref</td>
<td>NP 683657.1</td>
</tr>
<tr>
<td>gi</td>
<td>224877</td>
<td>dbj</td>
<td>EAA2147.1</td>
</tr>
<tr>
<td>gi</td>
<td>129204</td>
<td>np</td>
<td>NP 68396.1</td>
</tr>
<tr>
<td>gi</td>
<td>205212</td>
<td>dbj</td>
<td>BAB91973.1</td>
</tr>
</tbody>
</table>

**gi|103359|pir|1S09111 hypothetical protein 2 - fruit fly (Drosophila melanogaster) transposable element R1**

Length = 1021

Score = 79.7 bits (195), Expect = 2e-14

Identities = 46/143 (32%), Positives = 72/143 (50%), Gaps = 6/143 (4%)

Frame = +1

Query: 16 VDDQIARLNDLVWGVNDSLL-RKSESYRGGKKLKWTHEXXXX XXXXXXXX XXXXXXXX SXX 192

+D Q+ L +V V D+ L RK ++ +WWT +

Sbjct: 272 LDQQVSTLRISVHSVCDTALGRKLTTRSPRARRWWTADLCAARREVRLRLRLLQGRRR 331

Query: 193 SDRLAQIRHDLSTEMERE----YKDMLLRRIEEDWRSFVGQKDDPVGQYIKICGRRRS-Y 357

D ++ +E+R YK ++ R K +DW+ FVG + DDPWG++YKICGRR

Sbjct: 332 DDAAVHEL---VVVELRASASYYKHLGRAMDDWKRFVGDHADDwpGRYKICGRRKCT 388

Query: 358 DITSLRVDGLTMLSTSECARALL 426

+I LRV +++ +CAR LL

Sbjct: 389 EIGCRLVNGELITDQWDCARVLL 411

**gi|140023|sp|P16425|Y2R2 DROME**

Hypothetical 115 kDa protein in type I retrotransposable element RdRm (ORF 2)

[GenBank:emb|CAA36227.1](X51968) open reading frame II (AA 1-1021)

[GenBank:Drosophila melanogaster] Length = 1021

Score = 79.7 bits (195), Expect = 2e-14

Identities = 46/143 (32%), Positives = 72/143 (50%), Gaps = 6/143 (4%)

Frame = +1

Query: 16 VDDQIARLNDLVWGVNDSLL-RKSESYRGGKKLKWTHEXXXX XXXXXXXX XXXXXXXX SXX 192

+D Q+ L +V V D+ L RK ++ +WWT +

Sbjct: 272 LDQQVSTLRISVHSVCDTALGRKLTTRSPRARRWWTADLCAARREVRLRLRLLQGRRR 331

Query: 193 SDRLAQIRHDLSTEMERE----YKDMLLRRIEEDWRSFVGQKDDPVGQYIKICGRRRS-Y 357

D ++ +E+R YK ++ R K +DW+ FVG + DDPWG++YKICGRR

Sbjct: 332 DDAAVHEL---VVVELRASASYYKHLGRAMDDWKRFVGDHADDwpGRYKICGRRKCT 388

Query: 358 DITSLRVDGLTMLSTSECARALL 426

+I LRV +++ +CAR LL

Sbjct: 389 EIGCRLVNGELITDQWDCARVLL 411
**gi|2735949|gb|AAB94027.1| (AF015277) reverse transcriptase [Drosophila mercatorum mercatorum]** Length = 664

Score = 73.9 bits (180), Expect = 1e-12
Identities = 39/142 (27%), Positives = 70/142 (48%), Gaps = 1/142 (0%)
Frame = +1

Query:
```
RALGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
RAMTSDEQVTAUXSVHQVSDAVGQRQOLRAKRRVSWTALTDARLRRARRLQHAR
```

**gi|548540|sp|Q03277|PO11 SCICO Retrovirus-related POL polyprotein from type I retrotransposable element RI [Contains: Reverse transcriptase; Endonuclease]**

Score = 62.8 bits (151), Expect = 2e-09
Identities = 36/141 (25%), Positives = 65/141 (45%), Gaps = 2/141 (1%)
Frame = +1

Query:
```
LGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
VSVDEKVDLLTEWIIYGAENVNRRHTAVRTFQNEWSVLAEKRSELRRRRAQRQIRNA
```

**gi|160852|gb|AAA29813.1| (L00945) reverse transcriptase [Bradysia coprophila]**

Score = 73.9 bits (180), Expect = 1e-12
Identities = 39/142 (27%), Positives = 70/142 (48%), Gaps = 1/142 (0%)
Frame = +1

Query:
```
RALGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
RAMTSDEQVTAUXSVHQVSDAVGQRQOLRAKRRVSWTALTDARLRRARRLQHAR
```

**gi|2735949|gb|AAB94027.1| (AF015277) reverse transcriptase [Drosophila mercatorum mercatorum]** Length = 664

Score = 73.9 bits (180), Expect = 1e-12
Identities = 39/142 (27%), Positives = 70/142 (48%), Gaps = 1/142 (0%)
Frame = +1

Query:
```
RALGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
RAMTSDEQVTAUXSVHQVSDAVGQRQOLRAKRRVSWTALTDARLRRARRLQHAR
```

**gi|548540|sp|Q03277|PO11 SCICO Retrovirus-related POL polyprotein from type I retrotransposable element RI [Contains: Reverse transcriptase; Endonuclease]**

Score = 62.8 bits (151), Expect = 2e-09
Identities = 36/141 (25%), Positives = 65/141 (45%), Gaps = 2/141 (1%)
Frame = +1

Query:
```
LGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
VSVDEKVDLLTEWIIYGAENVNRRHTAVRTFQNEWSVLAEKRSELRRRRAQRQIRNA
```

**gi|160852|gb|AAA29813.1| (L00945) reverse transcriptase [Bradysia coprophila]**

Score = 73.9 bits (180), Expect = 1e-12
Identities = 39/142 (27%), Positives = 70/142 (48%), Gaps = 1/142 (0%)
Frame = +1

Query:
```
RALGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
RAMTSDEQVTAUXSVHQVSDAVGQRQOLRAKRRVSWTALTDARLRRARRLQHAR
```

**gi|2735949|gb|AAB94027.1| (AF015277) reverse transcriptase [Drosophila mercatorum mercatorum]** Length = 664

Score = 73.9 bits (180), Expect = 1e-12
Identities = 39/142 (27%), Positives = 70/142 (48%), Gaps = 1/142 (0%)
Frame = +1

Query:
```
RALGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
RAMTSDEQVTAUXSVHQVSDAVGQRQOLRAKRRVSWTALTDARLRRARRLQHAR
```

**gi|548540|sp|Q03277|PO11 SCICO Retrovirus-related POL polyprotein from type I retrotransposable element RI [Contains: Reverse transcriptase; Endonuclease]**

Score = 62.8 bits (151), Expect = 2e-09
Identities = 36/141 (25%), Positives = 65/141 (45%), Gaps = 2/141 (1%)
Frame = +1

Query:
```
LGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
VSVDEKVDLLTEWIIYGAENVNRRHTAVRTFQNEWSVLAEKRSELRRRRAQRQIRNA
```

**gi|160852|gb|AAA29813.1| (L00945) reverse transcriptase [Bradysia coprophila]**

Score = 73.9 bits (180), Expect = 1e-12
Identities = 39/142 (27%), Positives = 70/142 (48%), Gaps = 1/142 (0%)
Frame = +1

Query:
```
RALGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
RAMTSDEQVTAUXSVHQVSDAVGQRQOLRAKRRVSWTALTDARLRRARRLQHAR
```

**gi|2735949|gb|AAB94027.1| (AF015277) reverse transcriptase [Drosophila mercatorum mercatorum]** Length = 664

Score = 73.9 bits (180), Expect = 1e-12
Identities = 39/142 (27%), Positives = 70/142 (48%), Gaps = 1/142 (0%)
Frame = +1

Query:
```
RALGVDDQIARLNDLVGWVNDSSLRLKSESYSRGGKLKWWTDEXXXXXXXXXXX
```
Sbjct:
```
RAMTSDEQVTAUXSVHQVSDAVGQRQOLRAKRRVSWTALTDARLRRARRLQHAR
```
Figure 14. Southern blot of EcoRI-digested *B. dorsalis* male and female genomic DNA probed with the 3.1a clone sequence. The male and female hybridization patterns are both repetitive, without any apparent sex-specific bands.
3.3.3 Sequencing the remainder of the 3.1a clone

The 3.1a clone was chosen for further characterization because of its similarity to R1 retrotransposable elements. The insert of clone 3.1a was estimated to be 4.8 kb in size. The first approach used to determine its complete sequence was primer walking. The 3.1aFor and 3.1aRev sequences were used to design primers to amplify the region in between. The amplification product would then be cloned and sequenced at both ends, and the new sequence data would be used to design new primers for further amplification, cloning, and sequencing. The process would be repeated until the entire ~4.8 kb insert was completely sequenced.

PCR primers were designed from the sequenced ends of 3.1a to amplify the 3.1a clone insert using long template PCR, a method used to amplify larger target sequences that are longer than the amplification capabilities of conventional Taq DNA polymerase (Figure 15). Additional sequence data was obtained by sequencing from the 3.1aFor end of the long template PCR clone to obtain a total sequenced area of 883 bp from this end of the 3.1a clone (Figure 16). This 883 bp sequence contains a potential open reading frame (ORF) 306 bp in size spanning nucleotides 578 to 883. The putative ORF encodes 102 amino acids. A GenBank search using the BLASTN and BLASTX routines gave results similar to searches with the 3.1aFor sequence (Figure 11).

To obtain sequence data from the 3.1aRev end, the entire 3.1a insert was digested with the restriction enzymes EcoRI, PstI, and XhoI. The restriction fragments were cloned and sequenced. In all, 2031 bp of sequence data was obtained from the 3.1aRev end of the 3.1a clone (Figure 17). A BLASTX-facilitated GenBank
Figure 15. (a) 31aLongF1-B1 and (b) 31PasteF1-3.1aRevF2 long template PCR amplification product sizes from *B. dorsalis* males and females. Both primer pairs produce the same amplification patterns, with two male-specific amplification products, estimated to be 4.5 and 5 kb in size. The locations of the primers within the 3.1aFor and 3.1aRev sequences are indicated by the colored arrows.
Figure 16. Diagram of 3.1a clone showing the 883 bp sequenced region (including 3.1aF or sequence) shaded in gray. A potential ORF, 306 bp in size, is highlighted in purple type. This putative ORF encodes 102 amino acids.
883 bp sequenced region

M13 For

E = EcoRI
X = XhoI
P = PstI

3.1a (517 bp)

1 kb

Base composition: A = 24.6%, T = 29.3%, C = 26.6%, G = 19.4%
Figure 17. Diagram of 3.1a clone showing the 2031 bp sequenced region (including 3.1a Rev sequence) shaded in gray. The region within this sequence showing sequence similarity to R1 retrotransposon reverse transcriptases is underlined in black. A potential ORF, 609 bp in size, is highlighted in purple type. This putative ORF encodes 170 amino acids.
### 2031 bp sequenced region

<table>
<thead>
<tr>
<th>Base Composition</th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Composition</td>
<td>26.8%</td>
<td>29.8%</td>
<td>26.1%</td>
<td>17.2%</td>
</tr>
</tbody>
</table>

### 3.1a (-4.8 kb)

- **E**: EcoRI
- **X**: Xhol
- **P**: PstI

### M13 For

| 1 | TCGAGCTCCCT CCCCCCGAA CACCATGCGG CTTAACTCAA GAGTTAAGTC CGACATAGCA TTAAGGTGGC TCTTAGCATC AGGCGATATA |
|---|---|---|---|---|---|---|---|---|---|
| 91 | AACCTACCA CACCCCAAC AGAAATACC CTCATTTGCCA AAGCGGAATT TGGCTCTACA ATTACCTCCT CTTTGCAGGA CGTCTCAGCT |
| 81 | TTGTTCCTTCT CTCTCCGGTC GACTGCCTTT CAAGCTTTTC CAAAGGGCAC TGTTGGAGCT GTTATCCAGA CTGATCAGCA |
| 271 | CAGTGAATAT CAGGGATAAC AGAATTACCC TCATTTCCGC AAGCGGAATT TGGCTCTACA ATTACCTCCT CTTTGCAGGA CGTCTCAGCT |
| 361 | AGAGATGATG ATGGAAGAAAG |

### 3.1a Rev (550 bp)

| Base composition: A = 26.8%, T = 29.8%, C = 26.1%, G = 17.2%|
|---|---|---|---|---|
| TCGAGCTCCCT CCCCCCGAA CACCATGCGG CTTAACTCAA GAGTTAAGTC CGACATAGCA TTAAGGTGGC TCTTAGCATC AGGCGATATA|
| AACCTACCA CACCCCAAC AGAAATACC CTCATTTGCCA AAGCGGAATT TGGCTCTACA ATTACCTCCT CTTTGCAGGA CGTCTCAGCT|
| TTGTTCCTTCT CTCTCCGGTC GACTGCCTTT CAAGCTTTTC CAAAGGGCAC TGTTGGAGCT GTTATCCAGA CTGATCAGCA |
| CAGTGAATAT CAGGGATAAC AGAATTACCC TCATTTCCGC AAGCGGAATT TGGCTCTACA ATTACCTCCT CTTTGCAGGA CGTCTCAGCT |
| AGAGATGATG ATGGAAGAAAG |

### 1 kb

- **E**: EcoRI
- **X**: Xhol
- **P**: PstI
search with this 2031 bp sequence confirmed the similarity to Drosophila R1 reverse transcriptase originally seen with the 550 bp 3.1aRev sequence. Also located within this 2031 bp end is a potential ORF spanning nucleotides 25 to 634. This putative 609 bp ORF encodes 170 amino acids. A GenBank search using the BLASTN and BLASTX routines gave results similar to searches with the 3.1aRev sequence (Figure 13), showing similarity with the BLASTX routine to R1 element reverse transcriptase sequences.

3.4 *B. dorsalis* PCR assays with 3.1aRev primers

To further characterize the clone 3.1a, PCR primers were designed to amplify the region within the 3.1aRev sequence containing the predicted amino acid sequence similarity to R1Dm (Figure 18). Two different sets of PCR primers were designed. The pair 3.1aRev F1-B1 was designed to amplify most of the 550 bp 3.1aRev sequence. The 3.1aRevF2-B2 pair was designed to flank the 102 bp region of the 3.1aRev sequence showing similarity to R1Dm reverse transcriptase. *B. dorsalis* adult male and female genomic DNAs were used as templates for both primer pairs.

These primer pairs produced different amplification patterns in *B. dorsalis* males and females. In PCR amplifications where conventional Taq polymerase was used, the 3.1aRevF1-B1 primer pair generated a male-specific PCR product approximately 500 bp in size. The F2-B2 pair amplified a male-specific PCR product approximately 350 bp
(~350 bp) in size (Figure 19a). To confirm that the observed lack of amplification in females was not due to the target sequence size limitations of Taq polymerase, the PCR reactions were repeated using DNA polymerase for long template PCR. Long template PCR amplifications of male genomic DNA produced the ~350 bp product, along with a new product approximately 3 kb (~3 kb) in size. Long PCRs with female genomic DNA template also produced a ~3 kb product (Figure 19b).

The long template PCR primer pairs 31aLongFl-B1 and 31PasteF1-3.1aRevF2 also produced different-sized amplification products in B. dorsalis males and females. Both primer pairs amplified products approximately 3.0, 4.5, 5.0 and 6.0 kb in males and 3.0, 4.5, and 6.0 kb products in females (Figure 15).

3.5 Sequence analysis of 3.1aRevF2-B2 PCR product clones

The B. dorsalis male ~350 bp and the female ~3 kb products were cloned into plasmid vectors and sequenced. The ~3 kb female product was completely sequenced using primer walking. Complete sequences of these two products showed that the short male product was 325 bp in size (Figure 20), while the female ~3 kb PCR product was actually 2.6 kb in size (Figure 21).
Figure 18. Diagram of 3.1a clone showing the location of the 3.1aRevF1-B1 and F2-B2 primer pairs within the 3.1aRev sequence.
B. dorsalis 3.1aRevF2-B2 PCR products:

♂ = 325 bp, 2.6 kb
♀ = 2.6 kb

E = EcoRI
X = XhoI
P = PstI

= region with amino acid sequence similarity to D. melanogaster R1Dm

= 3.1aRev PCR primers
Figure 19. (a) PCR amplification products of *B. dorsalis* male and female genomic DNA (Puna, Hawaii strain) using conventional Taq DNA polymerase and long template PCR amplification products of *B. dorsalis* adult female genomic DNA, all with the 3.1aRev primers (3.1aRevF1-B1 or 3.1aRevF2-B2). As expected from the existing 3.1aRev sequence data, the male F1-B1 product was approximately 500 bp, while the male F2-B2 product was 325 bp. PCRs using female genomic DNA showed no amplification with regular Taq polymerase. However, when long template PCRs were done, an amplification product approximately 3 kb in size was produced with both primer sets. (b) Long template PCR amplifications of *B. dorsalis* adult male and female genomic DNA with the 3.1aRevF2-B2 primers, showing two products in the male amplifications (325 bp and 2.6 kb) and only one product (2.6 kb) in the female amplifications.
The 325 bp product has been sequenced from five different *B. dorsalis* adult males, while the 2.6 kb product has been sequenced from one adult female. Alignments show a small degree of sequence variation between the 325 bp product sequences amongst the five adult males, consisting of occasional single nucleotide differences (Figure 20). *B. dorsalis* males 4 and 5 both have a G instead of an A at their nucleotide position 38, and male 4 also has a T instead of a C at nucleotide position 177.

3.5.1 GenBank database search with the 3.1aRev PCR products

BLASTN and BLASTX searches of the GenBank database were performed with the 325 bp male product and the 2.6 kb female product. The 325 bp male product showed a high degree of sequence similarity, particularly with the BLASTX search, to the reverse transcriptases of R1 retrotransposable elements from a variety of different insect species (Figure 22), including *D. melanogaster* (S09111), *D. mercatorum* (AAB94027.1), and *S. coprophila* (Q03277). These alignments are similar to those obtained using the original 3.1aRev sequence (Figure 13).
Figure 20. Nucleotide sequence alignment of the 325 bp 3.1aRevF2-B2 PCR products from five \textit{B. dorsalis} adult males (Puna, Hawaii strain). Nucleotide differences are shown in color and their positions are indicated by asterisks. \textit{B. dorsalis} males 4 and 5 both have a G instead of an A at their nucleotide position 38, and male 4 also has a T instead of a C at nucleotide position 177.
Bdor M1 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAA AG CCCTTGCAACA
Bdor M2 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAA AG CCCTTGCAACA
Bdor M3 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAA AG CCCTTGCAACA
Bdor M5 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAAGAG CCCTTGCAACA
Bdor M4 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAAGAG CCCTTGCAACA

Base composition for Bdor M1: A= 24.6%, T= 29.5%, C= 26.4%, G= 19.4%

Bdor M1 CTGACTCTAT TTGATAACA TGTTGTCAAC AACACGAAGA GAGGTTATAT
Bdor M2 CTGACTCTAT TTGATAACA TGTTGTCAAC AACACGAAGA GAGGTTATAT
Bdor M3 CTGACTCTAT TTGATAACA TGTTGTCAAC AACACGAAGA GAGGTTATAT
Bdor M5 CTGACTCTAT TTGATAACA TGTTGTCAAC AACACGAAGA GAGGTTATAT
Bdor M4 CTGACTCTAT TTGATAACA TGTTGTCAAC AACACGAAGA GAGGTTATAT

Base composition for Bdor M1: A= 24.6%, T= 29.5%, C= 26.4%, G= 19.4%

Bdor M1 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAA AG CCCTTGCAACA
Bdor M2 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAA AG CCCTTGCAACA
Bdor M3 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAA AG CCCTTGCAACA
Bdor M5 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAAGAG CCCTTGCAACA
Bdor M4 ATAGGGACAA GTCCGGTGTT GTTTTAAAAA AGTAGAAGAG CCCTTGCAACA

Base composition for Bdor M1: A= 24.6%, T= 29.5%, C= 26.4%, G= 19.4%
Figure 21. The complete sequence of the *B. dorsalis* adult female 2.6 kb PCR product using the 3.1aRevF2-B2 primer pair. The amino acid translation of the putative open reading frame, encoding 583 amino acids, is also shown. Amino acid motifs common to Drosophila R1 retrotransposable elements reverse transcriptase sequences are underlined and in blue type.
1 TAAAGGGAGGGCTGTCAGGCGTTTGAGACGGAAGTTTCAACGTGCTAGGCGGTCCAATTCCGATAGGTTGGCCCAGATT 80

81 CGGCATGATCCTAGTACTGAGatgagagggtacaaagacatgctcctcagaatcaaagaagactggagaagcttcgt 160

1 M R G Y K D M L L R IKE E D W R S F V 20

161 ggggcaaaaacaagacgaccctggggcagatctataaagatctgcaggggtcgtgagaagctatagtagataaatctcccaggtcggaaagca 240

21 G Q N K D D P W G Q I Y K I C R G R R S Y D I T S L R 47

241 gggtttggtgacacccatgcttagtcaacatggagtgagtgcaagggctcttttagtgctttttttcccaaggtcggaagca 320

48 V G D T M L S T W SEC A R ALL S A F F P R SEA 73

321 cagggcactctctagcacaagaggctacagctccccactagatgacgaggttgaggtacctggctgttagatgtc 400

74 Q A P L A Q E V P V P P L D D G E L G Y A F G L I R S 100

401 taaagatccccagctcttgatggtttagtggaacgagatctgtgtaagcttcacccaggaatactcggaag 480

101 K R S P G L D G LNG E I C K S L W K F I P E Y L E A 127

481 ctatctatgataaatgctgtcctggagggatttttcacccacgtagtggaaaaaagctaggggtgtctcctcctgtaag 560


561 ccttagaaaaatgacagcaatactcctgcctctctatcgcggcatacgctccccactagttgtcagttcggaaagct 640


641 aatggtaggaacgcgtcctgcagagctacgacggtgttggtgctgatgacacaatctgcggccatattttgtgtgctctcag 720

181 M V E R L Q E L T S G M W S D R Q Y G F R R G R S V K 207

721 agggccctgtgattgtgatgttctgtcattgtggaaggtggataatgtgaaacaataatattgtgctcgacatattttgtgtgctctcaag 800

208 D A W M Y V L N A V R N V N K Y V L G I F V D F K 233

801 gggacatcctgactcctgacgagagtgcggtgatttgacagagcaataaagagcattacgtcagcagaaatctccttttggag 880

234 G A F D Y L I W D R V I E R L K E L N C P E I S L W R 260

881 gatgtttttcggcagcaaaaggctctctctctttggaatgtgagttggagcggctgtgctgtgcttcac 960

261 S Y F S D R K A S I V G M S G S V E T S V T R G C P Q 287

961 agggatctctctgtgctctatatatggaaccccctgatggaacccctctgtgctgcgtggtgcagcatttttgtgaatgt 1040

288 G S I C G P Y I W N L M N S L L A Q L E P L C E C 313
1041 tctgcatacggatgcacctttctcattttamtgggggtgctctctgtctaacagggggtctcaattct 1120
314 S AY ADD L L I L M G R S R N E I E R A G S Q F L 340
1121 cgaattttccacaaattggggggcctgccctggtggtgtggggtgacctatcactggtgtggataaaacggtgcaaatagctgaaagga 1200
341 E I V H N W G V D I S M K T V T M L K G K 367
1201 aacactgtcgcctgctcaccaacatcgtcctggttaaatggggtcagcatcaggtatgtgacgcaagtcaaatatctggga 1280
368 L S P V R S P I V R L N G V S I R Y V T Q V K Y L G 393
1281 ctgaccatgatgaaagtaattggttctcactccacatttagcaattgtgaaggagcactgcagctgaaaggca 1360
394 E I V H N W G V G V G V DIS M D K T V T M M L K G K 420
1361 aacgacgcatattttgaggagtgattggggtctcgggcgtcgtgcgtgctgttcgccaccatatatcggggtttgttgttggtgcttg 1440
421 R R I L I R S D W G L G R A V R T I Y R G L F V A C A 447
1441 ccacttacgggctgcctgctaatggtgggaacaacacacgacagtctctctggggtcagttctctctcataacacgctctt 1520
448 L T M S E R M S F H L I V K E R L Q G T V G K I 473
1521 ataattttcatttttcgctctgtctacggcaccctctcaacgggaagcaatgctgtatgggagctcccccggcttcg 1600
474 I M L A C L P V C R T V S T E A M Q V L G A P P L D 500
1601 cctgatgtttataacagggagctgtgccatcaggttaaggaggggtgtgtgtcattttctccagaatgtgattgatttt 1680
501 L I V I Q R A V A F R L R G L S V S L F Q N D W I S 527
1681 ctgaccaatgatgcaagggagattatatcttaggaagaagctgtgctgtgtgctctagctcctctgctgctggctg 1760
528 D N D A R E L G S K R L L D E I V R N K W Q N 553
1761 cgtgggaacaatggcaaaaaacacgctcgggattattctacccacattttccagagttatagtttgggaagaaaccccccagcttc 1840
554 R W N N S N N V E L L T N T S R M L G L W K K T P S 580
1841 agattctgtctgagctccccggctcccctctcctcctccgttcaatatctctctacatcaatgtgagctgccttcaggaacttagcaataaat 1920
581 D S V * 583
1921 AACACCCGAATGCAGGTGGCTTCCTACTCAGGGACATGGGCCTCTGAAATGCGATTCTCTGCATAGGAGGAACCTGTCGGA 1990
2001 GGAGTTTTAGGGGATGGGGGATAAGCTATTAGTTAGTTAGTAAGAAGCAAGCTAGATGTTGACCGCTGTGATCTCCACTAGCTGGAAATGCC 2080
2081 GAACTCTTAGAGCGTCTTAATGRTTTCGCGGTGAGCCCATTAAAGCGGAGGATGAAATGACGTATTAGTTAAGT
2160
2161 CAGGTTAGGGTATGGAATTGCTGGTGAGGTGCATGGTACCCACGAGCAGTATGGAAGCTC
2240
2241 AACTGGTAGTAGTTTCCGGCTACGAAGTCTGACGGAGACTTAATTCTGGTACCAGGGAGCCAGGAGCCTTGAGTTG
2320
2321 CTCAACACTGCTATGCGAGCCCGCCTTGAGGATCAGGTGTAGTTTATACCCAAATCGCGGAAGAGTGGT
2400
2401 TCGATCTCAATGGAAGTGCCATGGCAACCCGAGGAGAGATAGCTACGACGAGGGTGTGGTTGCGGGTCTCAACC
2480
2481 CTACCAGGGTAGGTGTGCTCCATCCACACTGCAAATTTGACTGAAAATGCTTAGATTCAGGGCCGCTGATCAAC
2560
2561 GCATTGATTGTGCCAGGGTACTTTTTGAGTACCGTAGAGTTTGGTCATCGACAGGTTACCTACGTATTTAACACCCGA
2640
2641 CGTTGTCCTAT 2652

Base composition: A = 25.0%, T = 25.6%, C = 20.0%, G = 29.2%
A BLASTX search of the GenBank database also showed that the 2.6 kb female product had a high degree of sequence similarity to a variety of R1 retrotransposable element reverse transcriptase sequences (Figure 23), including *D. melanogaster* (S09111, also the top BLASTX sequence match to the male 325 bp product translated sequence), *Drosophila neotestacea* (U23192), and *S. coprophila* (Q03277, also reported by BLASTX as a top sequence match to the 325 bp male PCR product). Analysis of the sequence also identified a large putative ORF, 1752 nucleotides in length, encoding a predicted protein of 548 amino acids. Within this translated ORF there are three distinct amino acid motifs common to Drosophila R1 retrotransposon reverse transcriptases (reviewed by Gentile et al. 2001), in locations comparable to their Drosophila counterparts (Figure 21). These are AYADD, located at amino acid positions 315-319 within the translated ORF; QVKYLG, at amino acid positions 388-393; and CLPVCRT, located at amino acid positions 478-484.

### 3.5.2 Alignment of 3.1aRevF2-B2 PCR product sequences

Sequence alignments were also constructed between the male-specific 325 bp PCR product, the female 2.6 kb PCR product, and the end of the 3.1a clone (3.1aRev) that was used to design the 3.1aRevF2-B2 primers, to determine the relationship between these three sequences (Figure 24). The male-specific PCR product aligns almost exactly with nucleotides 134 to 462 of the 3.1aRev sequence (with single nucleotide difference position 288 of the 325 bp sequence), as was expected from the locations of the
Figure 22. GenBank sequence database search results for the 325 bp *B. dorsalis* male
PCR product sequence using the BLASTN and BLASTX search algorithms. The best
matches are shown for those search results with a large number of matches, as indicated.
The BLASTN or BLASTX alignments for the best matches are also shown below the
search results. The 325 bp PCR product shows similarity to R1 element reverse
transcriptases in the BLASTX search method.
325 bp PCR product: GenBank BLASTN search results (five highest matches)

Sequences producing significant alignments: (bits) Value

| gi|21844714|gb|AC122868.2| Mus musculus clone RP23-148H17, complete sequence |
|---|---|---|
| Length | 202374 |
| Score | 44.1 bits (22) | Expect = 0.11 |
| Identities | 22/22 (100%) |
| Strand | Plus / Minus |

Query: 194 ttctgaggagcatgtctttgta 215

Sbjct: 5975 ttctgaggagcatgtctttgta 5954

Score = 44.1 bits (22), Expect = 0.11
Identities = 22/22 (100%)
Strand = Plus / Plus

Query: 194 ttctgaggagcatgtctttgta 215

Sbjct: 168038 ttctgaggagcatgtctttgta 168059

| gi|24196877|gb|AE011464.1| Leptospira interrogans serovar lai str. 56601 chromosome I, section 273 of 397 of the complete sequence |
|---|---|---|
| Length | 10911 |
| Score | 42.1 bits (21) | Expect = 0.45 |
| Identities | 21/21 (100%) |
| Strand | Plus / Minus |

Query: 18 tgtgttttaaaaaagtagaaa 38

Sbjct: 8044 tgtgttttaaaaaagtagaaa 8024

| gi|20340490|gb|AC098653.2| Homo sapiens chromosome 1 clone RP11-393P15, complete sequence |
|---|---|---|
| Length | 178316 |
| Score | 42.1 bits (21) | Expect = 0.45 |
| Identities | 24/25 (96%) |
| Strand | Plus / Minus |

Query: 21 gttttaaaaaagtagaaaaagccctt 45

Sbjct: 33403 gttttaaaaaagtagaaaaagccctt 33379
**325 bp PCR product: GenBank BLASTX search results (seven highest matches)**

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>103359</td>
<td>pir</td>
<td>S09111</td>
</tr>
<tr>
<td>gi</td>
<td>140023</td>
<td>sp</td>
<td>P16425</td>
</tr>
<tr>
<td>gi</td>
<td>1273594</td>
<td>gb</td>
<td>AAB94027.1</td>
</tr>
<tr>
<td>gi</td>
<td>548540</td>
<td>sp</td>
<td>Q03277</td>
</tr>
<tr>
<td>gi</td>
<td>23331529</td>
<td>ref</td>
<td>NP 683657.1</td>
</tr>
<tr>
<td>gi</td>
<td>23487793</td>
<td>gb</td>
<td>EAA21147.1</td>
</tr>
<tr>
<td>gi</td>
<td>22004017</td>
<td>dbj</td>
<td>S62244</td>
</tr>
</tbody>
</table>

**gi|103359|pir|S09111** hypothetical protein 2 - fruit fly (Drosophila melanogaster) transposable element R1

Length = 1021

Score = 67.8 bits (164), Expect = 3e-11

Identities = 30/62 (48%), Positives = 42/62 (67%), Gaps = 1/62 (1%)

Frame = -3

Query: 215 YKDMLRLIKEEDWRFSVGQNKDDDPWQIYKICGRGRS-YDITSRLVGDMLST*SECARA 39

YK ++ R K +DW+ FVG + DDPW+ +YKICGRG +I LRV +++ +CAR

Sbjct: 350 YKKLIGRAKMDDWKRFVGDHADDNPWGRVYKICGRGRKCTEIGCLRVENELTIDWGDACRV 409

Query: 38 FL 33

L

Sbjct: 410 LL 411

**gi|140023|sp|P16425|Y2R2 DROME** Hypothetical 115 kDa protein in type I retrotransposable element R1DM (ORF 2)

Length = 1021

Score = 67.8 bits (164), Expect = 3e-11

Identities = 30/62 (48%), Positives = 42/62 (67%), Gaps = 1/62 (1%)

Frame = -3

Query: 215 YKDMLRLIKEEDWRFSVGQNKDDDPWQIYKICGRGRS-YDITSRLVGDMLST*SECARA 39

YK ++ R K +DW+ FVG + DDPW+ +YKICGRG +I LRV +++ +CAR

Sbjct: 350 YKKLIGRAKMDDWKRFVGDHADDNPWGRVYKICGRGRKCTEIGCLRVENELTIDWGDACRV 409

Query: 38 FL 33

L

Sbjct: 410 LL 411
gi|2735949|gb|AAAB94027.1| (AF015277) reverse transcriptase [Drosophila mercatorum mercatorum]
Length = 664
Score = 52.8 bits (125), Expect = 1e-06
Identities = 26/81 (32%), Positives = 46/81 (56%), Gaps = 1/81 (1%) Frame = -3

Query: 272 SNSDRLAQIRHDLSTEMREYKDMLLRIKEEDWRSFVG-QNKDDPWPQIYKICRGRRS-YDI 96
+++SD + +EY+ M+L K +W+ +VG+++ PWG +Y+ICRGR+ D+
Sbjct: 331 THSDSATVLASYFRIARKEYERMMLHEKR-NWKRYVGEHQRHPWGSVYRICRGRKCTDL 389

Query: 95 TSLRVGDTMLST*SECARAFL 33
LR ++ T + CA L
Sbjct: 390 GCLRWNNELVTTWAACANVLL 410

---

gi|548540|sp|Q03277|POL1 SCICO Retrovirus-related POL polyprotein from type I retrotransposable element RI [Contains: Reverse transcriptase; Endonuclease]

Score = 47.8 bits (112), Expect = 3e-05
Identities = 22/55 (40%), Positives = 34/55 (61%), Gaps = 2/55 (3%) Frame = -3

Query: 218 EYKDMLLRIKEEDWRSFVG-QNKDDPWPQIYKICRGRRS-YDTSLRVGDTMLST 60
EYK ML K W+ FV ++ ++PVG+++K+CRGR+ D+ S++V T
Sbjct: 314 EYKRMICAEKLRWQCEFVASESNENPWPGRVFKLCRGRKPVDCVKVDGVYDTDL 368

---

gi|22331529|ref|NP_683657.1| (NM_148815) similar to putative transposon protein; protein id: At3g42916.1 [Arabidopsis thaliana]
Length = 517
Score = 31.6 bits (70), Expect = 2.4
Identities = 22/66 (33%), Positives = 36/66 (54%), Gaps = 1/66 (1%) Frame = +1

Query: 55 LYVDNMVSPTREVIS*LLRPLQIL*ICPQGSSLFCPTKLQSSSLLLRMSLSYS-LISV 231
+ +D ++ PTRRE + RP+Q + + ++ T LL SL S+S+Y S+L+S+S
Sbjct: 438 IQLDPLLYPTTPREREVHSTRPVQSIQAAIEEAI-H-STHLLDGSSLRESSRTSIYSPLLSI 496

Query: 232 LRSCRI 249
L S +
Sbjct: 497 LYSTHL 502
Figure 23. GenBank sequence database search results for the 2.6 kb *B. dorsalis* female PCR product sequence using the BLASTN and BLASTX search algorithms. The best matches are shown for those search results with a large number of matches, as indicated. The BLASTN or BLASTX alignments for the best matches are also shown below the search results. The 2.6 kb product shows sequence to R1 element reverse transcriptases when both the BLASTN and BLASTX search methods are used.
2.6 kb PCR product: GenBank BLASTN search results (seven highest matches)

Sequences producing significant alignments:  

<table>
<thead>
<tr>
<th>gi</th>
<th>gb</th>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1733459</td>
<td></td>
<td>72</td>
<td>4e-09</td>
</tr>
<tr>
<td>1733469</td>
<td></td>
<td>64</td>
<td>1e-06</td>
</tr>
<tr>
<td>1732437</td>
<td></td>
<td>62</td>
<td>4e-06</td>
</tr>
<tr>
<td>1732443</td>
<td></td>
<td>60</td>
<td>2e-05</td>
</tr>
<tr>
<td>1732439</td>
<td></td>
<td>60</td>
<td>2e-05</td>
</tr>
<tr>
<td>158262</td>
<td></td>
<td>56</td>
<td>3e-04</td>
</tr>
<tr>
<td>8469</td>
<td></td>
<td>56</td>
<td>3e-04</td>
</tr>
</tbody>
</table>

gi1733459|gbjU23167.1|DAU23167 Drosophila ananassae 28S ribosomal RNA gene, partial sequence, and R1 retrotransposable element reverse transcriptase gene, partial cds  
Length = 2257  
Score = 71.9 bits (36), Expect = 4e-09  
Identities = 45/48 (93%)  
Strand = Plus / Plus

Query: 2343 gccccttgggagtatcgtggtggttgtggtttatacccaaatcgcgg 2390  
Sbjct: 1352 gccccttgggagtatcgtggtgctgtggttgatacccaaatcgcgg 1399

gi1733469|gbjU23192.1|DNU23192 Drosophila neotestacea 28S ribosomal RNA gene, partial sequence, and R1 retrotransposable element reverse transcriptase gene, partial cds  
Length = 2329  
Score = 63.9 bits (32), Expect = 1e-06  
Identities = 41/44 (93%)  
Strand = Plus / Plus

Query: 960 cagggatctatctgtggtccatatatatggaacctcatgatgga 1003  
Sbjct: 16 cagggatctatctgtggtccatatatatggaacctaatgatgga 59

Score = 46.1 bits (23), Expect = 0.25  
Identities = 26/27 (96%)  
Strand = Plus / Plus

Query: 2350 ggggagtatcgtggtggttgtggtttta 2376  
Sbjct: 1408 ggggagtatcgtggtggtggtttta 1514
gi|732437|gb|U13026.1|DOU13026 Drosophila orena reverse transcriptase (Dore\Rl-element) gene, partial cds
Length = 1715
Score = 61.9 bits (31), Expect = 4e-06
Identities = 43/47 (91%)
Strand = Plus / Plus

Query: 2351 gggagtacgctggtgtgtgtgtttatatcccaaatcgccggaagagt 2397
Sbjct: 1357 gggagtacgctggtgtgtgtgtttatatcccaaatcgccggaagagt 1403

---

gi|732443|gb|U13029.1|DTU13029 Drosophila teissier reverse transcriptase (Dtei\Rl-element) gene, partial cds
Length = 1704
Score = 60.0 bits (30), Expect = 2e-05
Identities = 42/46 (91%)
Strand = Plus / Plus

Query: 2351 gggagtacgctggtgtgtgtgtttatatcccaaatcgccggaagag 2396
Sbjct: 1363 gggagtacgctggtgtgtgtgtttatatcccaaatcgccggaagag 1408

---

gi|732439|gb|U13027.1|DSU13027 Drosophila sechellia reverse transcriptase (Dsec\Rl-element) gene, partial cds
Length = 1723
Score = 60.0 bits (30), Expect = 2e-05
Identities = 42/46 (91%)
Strand = Plus / Plus

Query: 2351 gggagtacgctggtgtgtgtgtttatatcccaaatcgccggaagag 2396
Sbjct: 1356 gggagtacgctggtgtgtgtgtttatatcccaaatcgccggaagag 1401

---

gi|158262|gb|J01118.1|DRORGM102 d.melanogaster 28s rRNA 3' region, 5 kb type I insertion, clone cdm103
Length = 1161
Score = 56.0 bits (28), Expect = 3e-04
Identities = 53/60 (88%), Gaps = 1/60 (1%)
Strand = Plus / Plus

Query: 2338 ggtaggccccccttg-gggagtacgctggtgtgtgtgtttatatcccaaatcgccggaagag 2396
Sbjct: 744 ggtaggccccccttg-gggagtacgctggtgtgtgtgtttatatcccaaatcgccggaagag 803

131
2.6 kb PCR product: GenBank BLASTN search results (11 highest matches)

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>549</td>
<td>e-178</td>
</tr>
<tr>
<td>548</td>
<td>e-177</td>
</tr>
<tr>
<td>335</td>
<td>e-122</td>
</tr>
<tr>
<td>419</td>
<td>e-115</td>
</tr>
<tr>
<td>264</td>
<td>e-100</td>
</tr>
<tr>
<td>263</td>
<td>e-100</td>
</tr>
<tr>
<td>263</td>
<td>e-100</td>
</tr>
<tr>
<td>367</td>
<td>e-100</td>
</tr>
<tr>
<td>278</td>
<td>e-96</td>
</tr>
<tr>
<td>272</td>
<td>1e-95</td>
</tr>
<tr>
<td>268</td>
<td>3e-94</td>
</tr>
</tbody>
</table>

Sequences producing significant alignments:

- gi|103359|pir|S09111 hypothetical protein 2 - fruit fly (Drosophila melanogaster) transposable element R1
Length = 1021
Score = 549 bits (1414), Expect(2) = e-178
Identities = 278/547 (50%), Positives = 374/547 (67%), Gaps = 5/547 (0%)
Frame = +3

Query: 111
YKDMLLRIKEEDWRPSFVQGNKDDPFWGYQ1YKICRGRRSY-DITSLRVGDMLSTWSECARA 287
YK ++ R K ++DW ++ FGV + DDFW+++YKICRGRR +I LRV +++ W +CAR

Sbjct: 350
YKDMLLRIKEEDWRPSFVQGNKDDPFWGYQ1YKICRGRRSY-DITSLRVGDMLSTWSECARA 287
YK ++ R K ++DW ++ FGV + DDFW+++YKICRGRR +I LRV +++ W +CAR

Query: 288
ILLSAFFPFRSEAP--LAEQVFVPFLDGEDGFLYAFGLIRSKRPGLDGGLNGEICKSKLWF 461
LL FFP ++E++AP +A+EV P L+ E++ +S+RSPGDG+NG ICK+W+

Sbjct: 410
ILLSAFFPFRSEAP--LAEQVFVPFLDGEDGFLYAFGLIRSKRPGLDGGLNGEICKSKLWF 461
LL FFP ++E++AP +A+EV P L+ E++ +S+RSPGDG+NG ICK+W+

Query: 462
IFPELEAIYDKCVWGEGFPRERQGARVPPLEKSPDKNRSPSYRGGISLLPVGLKLRV 641
IFP++L ++ ++ C+ GYFP EWK RVV LLK PDK + P SYRG1 LLPV GKLVE +

Sbjct: 469
IFPELEAIYDKCVWGEGFPRERQGARVPPLEKSPDKNRSPSYRGGISLLPVGLKLRV 641
IFP++L ++ ++ C+ GYFP EWK RVV LLK PDK + P SYRG1 LLPV GKLVE +

Query: 642
MVERLQE-LSGGM-WSDRQYGFFRGGRSVKDAWYVLNAVRVNVNKKLGILGFDFDFGDVY 815
MV R++E L G W +GFR+GR V+DAV ++V + +YVLG FVD Pathfinder

Sbjct: 529
MVERLQE-LSGGM-WSDRQYGFFRGGRSVKDAWYVLNAVRVNVNKKLGILGFDFDFGDVY 815
MV R++E L G W +GFR+GR V+DAV ++V + +YVLG FVD Pathfinder

Query: 816
LIWDRVIERLKLNCPEISLWYFSDKASIVGMSQSVETSTVRCPQG9SICGYPYIYNL 995
W + + R +L +C E+ LIW+FS++ R+ A+ I SG+VE TVRCPQG9SI GP+I++

Sbjct: 586
LIWDRVIERLKLNCPEISLWYFSDKASIVGMSQSVETSTVRCPQG9SICGYPYIYNL 995
W + + R +L +C E+ LIW+FS++ R+ A+ I SG+VE TVRCPQG9SI GP+I++

Query: 996
MDMGMLAELEPIECECSAYADDDBLLLXGGRSXRKIEIRAGQFELIVNHNGGVGVGDI1MDK 1175
MDMGMLAELEPIECECSAYADDDBLLLXGGRSXRKIEIRAGQFELIVNHNGGVGVGDI1MDK 1175

Sbjct: 646
MDMGMLAELEPIECECSAYADDDBLLLXGGRSXRKIEIRAGQFELIVNHNGGVGVGDI1MDK 1175
MDMGMLAELEPIECECSAYADDDBLLLXGGRSXRKIEIRAGQFELIVNHNGGVGVGDI1MDK 1175

Query: 1176
TVTXMNLGGDVSFPVSLVNLGVSQIRYTVQKYLGLTMERSMSSPFPHTAVLRQKTVGTGTV 1355
TV M+LKG L R+P VR G ++ YV +YLG+T+SE M F H+A +++R+ G VG

Sbjct: 706
TVTXMNLGGDVSFPVSLVNLGVSQIRYTVQKYLGLTMERSMSSPFPHTAVLRQKTVGTGTV 1355
TV M+LKG L R+P VR G ++ YV +YLG+T+SE M F H+A +++R+ G VG

Query: 1356
KIRRLSDNGLRGRAREVRTRIYRLFVACATAYGAAVA+WETATTVSQGKLLSIQRLM1AC 1535
+ R+L+R+D+W+ RA RTYI GL C +GA VW+++T+V+ ++A L S QRLI+L C

Sbjct: 764
KIRRLSDNGLRGRAREVRTRIYRLFVACATAYGAAVA+WETATTVSQGKLLSIQRLM1AC 1535
+ R+L+R+D+W+ RA RTYI GL C +GA VW+++T+V+ ++A L S QRLI+L C

Query: 1536
LPVCRVTSTEAMQVLGAPFGLLVDIVIQRVAFAVLRLRGLSILFQNDWSDNADAEVCL1G 1715
LPVCRVTSTEAMQVLGAPFGLLVDIVIQRVAFAVLRLRGLSILFQNDWSDNADAEVCL1G 1715

Sbjct: 824
LPVCRVTSTEAMQVLGAPFGLLVDIVIQRVAFAVLRLRGLSILFQNDWSDNADAEVCL1G 1715
LPVCRVTSTEAMQVLGAPFGLLVDIVIQRVAFAVLRLRGLSILFQNDWSDNADAEVCL1G 1715

Query: 1716
SKRLLDZ 1736
K L+E

Sbjct: 882
SKRLLDZ 1736
K L+E

132
Score = 100 bits (249), Expect(2) = e-178
Identities = 48/119 (40%), Positives = 68/119 (56%)
Frame = +2

Query: 1784
RVTQYIQDVRFVNEPDDFRFCSLGFSTLHGPFLNAILHRLNLETPECCAGARNEDMV 1963
RVT+++I V +P E F + + FLLTHG NAFLH R LS+T CACG EDW+
Sbjct: 906
RVTTHFIPYVTLAYRDFSFGMRSTFSLGHTGSHFSNAILHRLGALSDTACACGDPAFWM 965

Query: 1964
HVIACEMYS附加值GYNQLDVSFGVISTWNAELLETFNRAFEAKRRSQL 2140
H++ C +Y+D+R G+G+ + E G++ + E + FA E F RR L
Sbjct: 966
HILCACPLYADLRDDGLVGQRGENWIGELD---DQEKTRILMAFEEVFLRRRAL 1021

gi|140023|sp|P61425|Y2R2 DROME Hypothetical 115 kDa protein in T retrotransposable element R1Dm (ORF 2)
gi|1335645|emb|CA36227.1 (X51968) open reading frame II (AA 1-1021)
[Drosophila melanogaster]
Length = 1021
Score = 548 bits (1411), Expect(2) = e-177
Identities = 278/547 (50%), Positives = 374/547 (67%), Gaps 5/547 (0%)
Frame = +3

Query: 111
YKDLMLRIKKEEDWSFVGQKNDDPDGQKYYKICRGGRSY-DITSLRVGDTMLSTWSECARA 287
YK ++ R K +DW FVG + DDPG++YYKICRGGR +I LRV +++ W +CAR
Sbjct: 350
YKKGILGAGGMDDKWDYDIFKYKICRGGRKCTEIIGCVNLIGELIGFDCAV 409

Query: 288
LLSAPFFPSQFAAF-LLQEFVVFPLLGDHELGAFYGLRKLSPGDGLGNEICKSLWKF 461
LL FPP +E++AP +A+EVP P L+ E+ +S+PQGDG+NG ICK+W+
Sbjct: 410
LLRNFFPVAESEAPTAIAEVP-PALEVFEVDTCVARKSRSSRGPDGLDGINTICKAVWA 468

Query: 462
IEYELAAYDKCCWEGYFPPPEKAAVPLKSDPISRNSPSPYRSGILPLPLFVLGKVLEVR 641
IEP+E + +C+ GYFP EWK RLK LLK P +K + + + + +VLYG FVDFKA PAFD
Sbjct: 469
IEPHLSASFSRCIRLGYFPAEWHCPVRVSAHLDKPDKEPSSYRGICLPLPVFGKVEAI 528

Query: 464
MVEKLE-QE-ITSIDM-WSDRQYGFRRGGRVQDAWVMYVLNAVRNVNKLGVILFDFKGAPFDY 815
MV R++E L G W Q+GR+GR V+DAW +V ++V + +VYLG FVDFKAFD
Sbjct: 529
MVNRVREVLFPEGCRW---QGFQRGFCVEDARHWKVSSGAASYQVLGTFVDFKAGFDN 585

Query: 816
LIWDRVIERLKECNPEISLWRSYFSDRASIVMSGSVETSVTIRCPQGGSGQPYYWNL 995
+W + R + L + C E+ LW+5+FS R A+ A SLG+VE VTIRCQPQGI GP+I+WW+
Sbjct: 586
VBWASAASLRLADLGMEQLWQGSGRRGARRSSTSSGTVEVPQVCRQPGSGISGGPIWDI 645

Query: 996
MMDSSLLAQELPCECSAYADDILLLLXGGRSRXEIERAGSQFLEIVHNVGWGVGVDISMDK 1175
+MD LL +L+P C+ SADDELL+L G SR +E G+ + IV WG VG +S K
Sbjct: 646
LMVSLILQRLQQPGSCMSAYADDLHLILVGEINSRAVLEEKQALQMSIETVWAGVDCISTK 705

Query: 1176
TVTMMLGKLFSVPISIVRNLGINMIVYRTTQKLYLTMSMSEPMTHLAVKILQGGTVG 1355
TV M+LKG L +P VR G ++ YV +YLG+T+SE M F H+AA ++R+ G VG
Sbjct: 706
TVMTMLGA--R-APTRVAFGRNLPLLVPVRSCKYLGITVSEQKFDHISLRAVRMTGVG 763

Query: 1356
KIRRISDWDGGLGAVRVTIYRKLVFCATYGAHAVWETATTVSGQKLLSISQLRMLAC 1535
+R+LR+DDG RA RTY GL C +GA WY+TA V+ ++L S QRRI+L C
Sbjct: 764
ALARVLRADWGSFLPARRTITYIDGIMAPCVLFGAPWYDTEAQVAQRRLLASCQRLILLC 823

Query: 1536
LPCVTRTVSTEAMQQLGAPPDDLILVIVQAVAFRRRLRLSGLVLFSQVNDSNWDAEREGYGLG 1715
L VCRTST A-QYL GAPPD+ A+ ++L+RG L +NDW+ D +
Sbjct: 824
LSCVTRTVSTVALQGLLGAPAPLTLAAMQYKLLAIKYLKRGF---PLEENWDYGEIACLSQEV 881

Query: 1716
SKRLDE 1736
K L+E
Sbjct: 882
RKTRLEE 888

Score = 100 bits (249), Expect(2) = e-177
Identities = 48/119 (40%), Positives = 68/119 (56%)
Frame = +2

Query: 1784 RVTYQYIQDVFVVEENPDRFCLSSQFGFLTTGHGPLNAFLHRNLSETPECAGARNEDWV 1963
RVT++I V ++P E F ++ +PLFTHG NAFHL R LS+T CACG EDW+
Sbjct: 906 RVTFRFIPFYVTLAYRDFSFGMSRTSSLTHGSHSNAFLHSGRLSDTATACGDPYEDWM 965

Query: 1964 HVIAECAMYSDIRSFQMGISYVENQLDVGSVISTSWNAELETNFRAEFKRRSQL 2140
H++ C +Y+D+R G++ + E G++ + + E + FA E F RR L
Sbjct: 966 HILCACPLYADRLDGLGVRQGLENWIFEGILD---DQEKTLARMFAEEVFLRRRAL 1021

gi|733470|gb|AAA90996.1| (U23192) reverse transcriptase [Drosophila neotestacea]
Length = 400

Score = 335 bits (859), Expect(2) = e-122
Identities = 162/264 (61%), Positives = 201/264 (75%)
Frame = +3

Query: 951 GCPQGSICGPIWNLMDSSLLAQLPLECECLASAYADDLLILXGGRSRXEIERAGSQUALIV 1130
GCPQGSICGPIWNLMDSSLLAQLPLECECLASAYADDLLILXGGRSRXEIERAGSQUALIV
Sbjct: 3 GCPQGSICGPIWNLMDSSLLAQLPLECECLASAYADDLLILXGGRSRXEIERAGSQUALIV 62

Query: 1131 HNWGVGVDISMDKTVTMMLKGKLSPVRSI VLYVTQKYLGLTMERSEMTF 1310
+WG VGV ++MDKT TM+LKG+LS R P + LNG +RYVT+VKYLG+T ERMT 62
Sbjct: 63 CDWNSVGVSLAMDHTMLMLKDRLSASRHPSILGLAFYLYTEWYVEKGLTPFGERMCFT 122

Query: 1491 RQKLLSIOQLMLACLPLVCRVSTEAMQVLGLPPDLDIVIQRAVAFRRLGRGLSVSLFQN 1670
R+K+L+ QR++ M+ C+PVCRVSTEAMQVLGLPPDLDIVIQRAVAFRRLGRGLSVSLFQN
Sbjct: 183 RKKVLACQVRTMGCMPVCRVSTEAMQVLGLPPDLDLEVRARRLAFKVR---RIPLLQG 240

Query: 1671 DWISDNAEREGYLSKRLLEDIEV 1742
+W++D + E G K+LLDE V
Sbjct: 241 EWLADRNVESLGSVCKKLDECV 264

Score = 127 bits (318), Expect(2) = e-122
Identities = 66/120 (55%), Positives = 75/120 (62%)
Frame = +2

Query: 1784 RVTYQYIQDVFVVEENPDRFCLSSQFGFLTTGHGPLNAFLHRNLSETPECAGARNEDWV 1963
R VT+Y+I+DV FV PDF F LS GFLITGHG LNAFLH+R LS+T EC CG E W
Sbjct: 279 RDTYRIRIDTVFSGPDSFGNLSFLITGHGSINAFHQLQRRLSDTQECHGGLSEETWE 338

Query: 1964 HVIAECAMYSDIRSFQMGISYVENQLDVGSVISTSWNAELETNFRAEFKRRSQL 2143
HV+ EC Y D+R G+ + E DVS +STS LL N FAB AF RR L+
Sbjct: 339 HVLCECPSYEDRLSLSAFGVRQGFDVSQALSTSDRVRLL---NEFARAAFARRVLT 395
Retrovirus-related POL polyprotein from type I retrotransposable element R1 [Contains: Reverse transcriptase; Endonuclease]
gi|160852|gb|AA29813.1 (L00945) reverse transcriptase [Bradysia coprophila]
Length = 1004

Score = 419 bits (1077), Expect = e-115
Identities = 234/543 (43%), Positives = 328/543 (60%), Gaps = 10/543 (1%)
Frame = +3
Score = 114 bits (285), Expect = 7e-24
Identities = 57/122 (46%), Positives = 75/122 (60%)
Frame = +2

Query: 1775 QQRVTYQYIQDVRFVVEENFDPRFCLSGFLTGHGPNLHNRHNLSETEPCACGARNE 1954 ++ R++T+++I VRFVF EN F L G+LTHG +N FLH+R LS TP C GA NE
Sbjct: 878 EKGRILTHEFIPSVRFVENEWVAFGICLGYVLTHGPDMLHRLGSLNTPVCMCGAFNE 937

Query: 1955 DWVHVIAECAMSDRSFRGMSGIYYVNEQLDVSGVISTSWNAILTEFNRFAREAFKRRS 2134 D H++ EC +Y D+R G G+ N LDVSG +S E N+FA F RRS
Sbjct: 938 DVKKHLLGECPLAYDLRDLNGGCLLLRNSLDVGALS----EIGAVEKLNQFAVSFLGRRS 994

Query: 2135 QL 2140 +L
Sbjct: 995 RL 996

gi|8489513|gb|AAFB7569.1|AF248070.1 (AF248070) reverse transcriptase
[Drosophila orientacea]
Length = 358

Score = 264 bits (674), Expect(2) = e-100
Identities = 130/222 (58%), Positives = 166/222 (74%)
Frame = +3

Query: 1077 GRSRXIERAGSQFLIEHVNWGCGVGVDISMDKTVMGLKGLSVPVPVRLNGV5MYR 1256 G++SR EIE L V +WG VWG +++MDKT TM+LKG+LS R P + LNGV+++PY
Sbjct: 3 GQSRAIEIALAGAHRLRTCDVGANSGLAMDKTTTMLKGLRASRHPSGLGNGPLRY 62

Query: 1257 VTOVYKGLLGSRMSFTPLAHVRLQGTVKIRLRSWGLGGRARAVTTYQRLFVA 1436 V+TKYG+T ERM FTPH +KR RL G VG++RRIR++WGL RRAVTRITYGLFVA
Sbjct: 63 VTEVYKLITGFGERMCFTPHFTGLKRRLLGVGGQVRMRILRNEWGSLRRAVTRITYGLFVA 122

Query: 1437 CATYGAAVWETAVTSRGKLLSSLQRMLACLPIVCRVTSMQVLLGAPPLDLIVIQ 1616 CATY+G+V + TTV GR+K+L+ QR+ M+ C+PTVCTSMQVLLGPLDL V +
Sbjct: 123 CATYGGSSVWCDTVTVGRKVLACQRTVMGMPCTVSMQVLLGVPPLDLERVR 182

Query: 1617 RAVAFRLRGLSLVLFQNDWSDNDAEGERYLSKRLLESIV 1742 RAV F+++R + L Q ++W++D + E G K++L+E V
Sbjct: 183 RAVLFWKVR--RIPLLQGEWALDRNVESLGLSVCKKLNECV 222

Score = 127 bits (319), Expect(2) = e-100
Identities = 66/120 (55%), Positives = 75/120 (62%)
Frame = +2

Query: 1784 RVTQYQYIQDVRFVVEENFDPRFCLSGGFLTGHGPNLHNRHNLSETEPCACGARNE 1963 R TY+Y+I+DV F V PDF F LS GPILHG LNAFLH+R LS+T EC CG E W
Sbjct: 237 RDTRYIRDVTYGSRDPFGLNLGLITTGHLGGINAFHLQHQLRSSTQDTCHELCSLSETEWE 296

Query: 1964 HVIAECAWMYSRDRSTGMSGIYYVNEQLDVSGVISTSWNAILTEFNRFAREAFKRSQLS 2143 H+++ V CEL Y +R S G+ V DVS +STS LL N PAB AF BR L+
Sbjct: 297 HVICECPSEYDLRLSALAFGVRQVFGRFVDVSQALSTSDRVL---NEFARSAFRRRTVIT 353
3.1aRevF2-B2 PCR primer sequences. The 2.6 kb female product sequence aligned with nucleotides 134 to 424 of the 3.1aRev sequence and nucleotides 1 to 291 of the 325 bp male product. However, the remainder of the 2.6 kb female product sequence does not align well with the remainder of the 325 bp male-specific product or the 3.1aRev sequences.

3.5.3 Alignment of *B. dorsalis* female 2.6 kb PCR product sequence with other R1 reverse transcriptase sequences

The amino acid sequence encoded by the potential 1.7 kb ORF was aligned with a number of R1 element reverse transcriptase amino acid sequences from different insect species that were initially found to be highly similar by BLASTX, including *D. melanogaster* (accession number S09111), *D. mercatorum* (AAA90994.1), *D. neotestacea* (AAA90996.1), *D. testacea* (AAA91000.1), *S. coprophila* (A44490), and *B. mori* (BAA07647.1). This alignment (Figure 25) shows that the locations of the Drosophila R1 element reverse transcriptase amino acid motifs described previously (AYADD, QVKYLG, CLPVCRT) within the 2.6 kb product ORF correspond to the locations of these motifs within the R1 reverse transcriptase encoding sequences from the other Dipteran species with few gaps occurring within the region containing these sequences.
Figure 24. Alignment of the nucleotide sequence 3.1aRev, the *B. dorsalis* adult male 325 bp 3.1aRevF2-B2 PCR product ("Bdor_M1", reverse complement of the sequence shown in Figure 20), and the *B. dorsalis* adult female 2.6 kb 3.1aRevF2-B2 PCR product ("2.6 kb"). The latter portion of the 2.6 kb product sequence has been excluded from the figure. Nucleotide differences shown among all three sequences are in color, their locations indicated by asterisks.
1 50
3.1aRev ATTAGGGCTC TGGGTGGTTG CGATCAAATC GCTCGGCTAA ACGACTTAGT
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

51 100
3.1aRev CTGGGGGTGT TACCGAAGCT TACCGAAGCT TACCGAAGCT TACCGAAGCT
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

101 150
3.1aRev AAAAACTCAA ATGGTGGACG CATGAGCTAA
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

151 200
3.1aRev AGGCCTTTGA GACGGAAGTT TCAACGTGCT
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

201 250
3.1aRev GTTGCCCAAT ATCCGGATG TACCTAGTAC TGAGATGAGA
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

251 300
3.1aRev ACATGCTCCT CAGAATCAAA GAAGAAGACT GGAGAAGCTT
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

301 350
3.1aRev AGATCTATGAT ATAACCTCTC TTCGTGTTGG TGACACCATG
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

351 400
3.1aRev AGAGTGAGTG
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

401 450
3.1aRev AGGTGAGCT TGCAAGGGCT CTTC....TA CTTTTTTAAA ACACACCGGA
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

451 500
3.1aRev CTGTTCCCT ATCTACAGTT TGCGGTGGAA CAGCCGAAGC GAACGGACGT
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

501 550
3.1aRev GTGCCCGGAC GATTCAGTTT TTACGGAGCG TTGATATTGA TGACGGTGTT
Bdor_Ml ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
2.6 kb ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
3.1aRev

Bdor_M1

2.6 kb TTGATGTTTT GAACGGAGAG ATCTGTAAGA GCTTGTGGAA GTTCATTCCA
Figure 25. Alignment of the putative amino acid sequence from the *B. dorsalis* female 2.6 kb PCR product ORF with R1 element reverse transcriptase encoding sequences found by BLASTX to show high sequence similarity to the ORF sequence. Amino acid identities are in color. The locations of Drosophila R1 element reverse transcriptase amino acid sequence motifs (AYADD, QVKYLG, CLPVCRT) in the 2.6 kb product ORF are indicated by braces (\(\_\_\_\_\_\) ).
| D.melano | VCRTVSTVAL QVLGAAPPLDLAAKLAILEKLQKG..EPL EENDWLYQED |
| D.testacea | TCRTVSTVAL QVLGAAPPLDLAAKLAILEKLQKG..EPL EENDWLYQED |
| D.mercatorum | TVCQVSTVQL VLAGAPPFLLVAMRTAMQFLKRRN..YPL EEDWLYQED |
| D.neotest | TVCQVSTEAM QVLGAAPFLDLAAKLAILEKLQKG..EPL EENDWLYQED |
| B.dor 2.6 | TVCQVSTVQL VLAGAPPFLLVAMRTAMQFLKRRN..YPL EEDWLYQED |
| S.coprophila | TVCQVSTVQL VLAGAPPFLDLAAKLAILEKLQKG..EPL EENDWLYQED |
| B.mori | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.melano | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.testacea | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.mercatorum | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.neotest | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| B.dor 2.6 | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| S.coprophila | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| B.mori | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.melano | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.testacea | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.mercatorum | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.neotest | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| B.dor 2.6 | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| S.coprophila | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| B.mori | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.melano | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.testacea | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.mercatorum | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| D.neotest | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| B.dor 2.6 | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| S.coprophila | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
| B.mori | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |

D.melano  | 851  | VCRTVSTVAL QVLGAAPPLDLAAKLAILEKLQKG..EPL EENDWLYQED |
D.testacea  | 900  | VCRTVSTVAL QVLGAAPPLDLAAKLAILEKLQKG..EPL EENDWLYQED |
D.mercatorum  | TCRTVSTVQL VLAGAPPFLLVAMRTAMQFLKRRN..YPL EEDWLYQED |
D.neotest  | TVCQVSTVQL VLAGAPPFLDLAAKLAILEKLQKG..EPL EENDWLYQED |
B.dor 2.6  | TVCQVSTVQL VLAGAPPFLDLAAKLAILEKLQKG..EPL EENDWLYQED |
S.coprophila  | TVCQVSTVQL VLAGAPPFLDLAAKLAILEKLQKG..EPL EENDWLYQED |
B.mori  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |

D.melano  | 951  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
D.testacea  | 1000 | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
D.mercatorum  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
D.neotest  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
B.dor 2.6  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
S.coprophila  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
B.mori  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |

D.melano  | 1001 | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
D.testacea  | 1050 | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
D.mercatorum  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
D.neotest  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
B.dor 2.6  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
S.coprophila  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |
B.mori  | IACLSWEQRK TRLEECLQS WQNRWDDD..SEPGRVTHRF IPYVTLAYRD |

---

### Sequence Comparison

- **D.melano**
- **D.testacea**
- **D.mercatorum**
- **D.neotest**
- **B.dor 2.6**
- **S.coprophila**
- **B.mori**

### Additional Information

- GenBank Accession Numbers:
  - D.melano: [Accession Number](https://www.ncbi.nlm.nih.gov/)
  - D.testacea: [Accession Number](https://www.ncbi.nlm.nih.gov/)
  - D.mercatorum: [Accession Number](https://www.ncbi.nlm.nih.gov/)
  - D.neotest: [Accession Number](https://www.ncbi.nlm.nih.gov/)
  - B.dor 2.6: [Accession Number](https://www.ncbi.nlm.nih.gov/)
  - S.coprophila: [Accession Number](https://www.ncbi.nlm.nih.gov/)
  - B.mori: [Accession Number](https://www.ncbi.nlm.nih.gov/)

---

### References

2. [Research Article](https://www.ncbi.nlm.nih.gov/pmc/articles/)
3. [Conference Proceeding](https://www.ncbi.nlm.nih.gov/books/)

---

### Acknowledgments

This research was supported by the National Institutes of Health grant [Grant Number](https://www.nih.gov/grants/)

---

*This is a sample document with fictional content for demonstration purposes.*
3.6 *B. dorsalis* genomic DNA Southern blots probed with 3.1aRevF2-B2 PCR products

The 325 bp male and 2.6 kb female PCR products were used as probes for duplicate Southern blots of adult *B. dorsalis* male and female genomic DNA digested with the restriction enzymes EcoRI, PstI, and HindIII. Both PCR product sequences appeared to be repetitively represented in the *B. dorsalis* genome. The blot probed with the 2.6 kb PCR product hybridized in the same repetitive fashion in both males and females without any apparent sex-specific probe hybridization pattern (Figure 18b). In contrast, an identical Southern blot probed with the 325 bp male PCR product showed a similar repetitive pattern of probe hybridization in both males and females, but with male-specific bands appearing in the genomic DNA digested with PstI and HindIII (Figure 18a). The PstI male-specific band was estimated to be 4 kb, while the HindIII male-specific band was estimated to be 5 kb in size.

3.7 3.1aRev primer survey of various stages in the *B. dorsalis* life cycle

To assess the utility of the 3.1aRev primer amplification patterns as a tool for inferring sex at all stages of the *B. dorsalis* life cycle, genomic DNAs from *B. dorsalis* embryos and third instar larvae were used as templates for PCR reactions with these primers. Amplification products from PCRs of genomic DNA derived from *B. dorsalis* embryos appeared to differ slightly in size, approximately 350 to 400 bp (Figure 19). Three of the 10 embryo PCRs also produced amplification products that were approximately 800 bp and 2.6 kb in size.
Figure 26. Southern blots of *B. dorsalis* male and female genomic DNA digested with EcoRI, PstI, and HindIII (a) probed with the 325 bp *B. dorsalis* male 3.1aRevF2-B2 PCR product, with male-specific bands in the PstI and HindIII-digested DNA hybridization patterns (PstI band is approximately 4 kb, HindIII band is approximately 5 kb); (b) probed with the 2.6 kb *B. dorsalis* female 3.1aRevF2-B2 PCR product, without an apparent sex-specific probe hybridization pattern.
The PCR amplifications of genomic DNA extracted from third instar larvae closely mimicked the adult PCR amplification patterns. Of the 10 individuals that were surveyed, five had PCR products that appeared to be similar in size to the adult male 325 bp product (Figure 19). One of these individuals had what appeared to be the 325 bp product and another that appeared to be the 2.6 kb product. The remaining five larval PCRs gave rise to products similar in size to the 2.6 kb adult product.

PCRs with genomic DNA derived from *B. dorsalis* pupae all appeared to amplify a single product that was similar in size to the adult 2.6 kb product (Figure 19). A number of attempts were made to perform long template PCR with these embryonic, larval, and pupal genomic DNA samples in order to confirm the PCR amplification patterns seen with Taq polymerase, but these long template PCR attempts did not result in the amplification of DNA. Thus, the amplification patterns in *B. dorsalis* pupae differed from the patterns observed in both embryos and third instar larvae.
Figure 27. 3.1aRevF2-B2 primer PCR amplifications using genomic DNA extracted from *B. dorsalis* embryos (24 hours old), third instar larvae, or pupae as templates. PCR reactions were performed using conventional Taq polymerase. The embryo and third instar larval PCR amplifications produced 325 bp (E1-E6, L1, L2) and 2.6 kb products (E4, L1, L3-L5), while the PCR reactions with genomic DNA from pupae only produced the 2.6 kb product. Long template PCR reactions were attempted, but without successful DNA amplification.
3.8 Sequence analysis of 3.1aRev PCR products from *B. dorsalis* embryos and larvae

The short (putative male-specific) PCR product from *B. dorsalis* embryo 6 (E6, PCR result in Figure 19) was cloned and sequenced. The sequence data showed that this PCR product was actually 325 bp in size. An alignment of the embryonic PCR product with the 325 bp products from the five adult males that were previously described (and were also aligned in Figure 14) showed that the embryo PCR product is completely alike in sequence to the 325 bp PCR products from adult males 1-3 (Figure 20). The ends of the 2.6 kb products from two different embryos were also sequenced. These sequenced ends align with the ends of the adult female 2.6 kb PCR product, but with single nucleotide differences (Figure 21).

3.9 Surveying other *B. dorsalis* strains and Bactrocera species with 3.1aRev primers

Genomic DNAs from other wild *B. dorsalis* individuals were assayed with the 3.1aRevF2-B2 primer pair to assess whether their amplification patterns coincided with the sex-specific amplification patterns seen consistently in the males and females of the Puna, Hawaii *B. dorsalis* strain (Figure 22). The DNA from two *B. dorsalis* adult males and females from each of the Taiwan and Okinawa wild strains was surveyed with these primers. In long template PCR amplifications of the Taiwan and Okinawa samples, the males produced both the 325 bp and 2.6 kb amplification products, while the females produced only the 2.6 kb product. Thus, the amplification pattern seen in these strains is identical to the patterns seen in males and females from the Puna, Hawaii strain.
Figure 28. Nucleotide sequence alignment of the 325 bp 3.1aRevF2-B2 PCR products from five adult *B. dorsalis* males and one 24 hour-old *B. dorsalis* embryo (Bdor_embryo6). The embryo product sequence is identical to the 325 bp product sequences obtained from *B. dorsalis* adult males. The nucleotide differences are in color and the locations are indicated by asterisks.
Figure 29. Nucleotide sequence alignment of one end of two different *B. dorsalis* embryo (embryo 6, embryo 8) 3.1aRevF2-B2 2.6 kb PCR products with the *B. dorsalis* adult female 2.6 kb product. The nucleotide differences are in color and the locations are indicated by asterisks.
embryo 6 TAAAGAGGAG GACTGTCAGG CGTTTGAGAC GGAAGTTTCA ACGTGCTAGA
embryo 8 TAAAGAGGAG GACTGTCAGG CGTTTGAGAC GGAAGTTTCA ACGTGCTAGA
2.6 kb TAAAGAGGAG GCTGTCAGG GTTTTGAGAC GGAAGTTTCA ACGTGCTAGA

embryo 6 CGGTCCAATT CCGATAGGTT GGCCCAGATT CGGCATGATC TTAGTACTGA
embryo 8 CGGTCCAATT CCGATAGGTT GGCCCAGATT CGGCATGATC TTAGTACTGA
2.6 kb CGGTCCAATT CCGATAGGTT GGCCCAGATT CGGCATGATC TTAGTACTGA

embryo 6 GATGAGAG G TACAAAGACA TGCTCCTCAG AATCAAAGAA GAAGACTGGA
embryo 8 GATGAGAGAG TACAAAGACA TGCTCCTCAG AATCAAAGAA GAAGACTGGA
2.6 kb GATGAGAGGG TACAAAGACA TGCTCCTCAG AATCAAAGAA GAAGACTGGA

embryo 6 GAAGCTTCGT GGGGCAAAAC AAAGACGACC CCTGGGGGCA GATCTATAAG
embryo 8 GAAGCTTCGT GGGACAAAAC AAAGACGACC CCTGGGGGCA GATCTATAAG
2.6 kb GAAGCTTCGT GGGGCAAAAC AAAGACGACC CCTGGGGGCA GATCTATAAG

embryo 6 ATCTGCAGGG GTCGTAGAAG CTATGACATA ACCTCTCTTC GTGTTGGTGA
embryo 8 ATCTGCAGGG GTCGTAGAAG CTATGACATA ACCTCTCTTC GTGTTGGTGA
2.6 kb ATCTGCAGGG GTCGTAGAAG CTATGACATA ACCTCTCTTC GTGTTGGTGA

embryo 6 CACCATGTTA TCAACATGGA GTGAGTGTGC AAGGGCTCTT CTTAGTGCGT
embryo 8 CACCATGTTA TCAACATGGA GTGAGTGTGC AAGGGCTCTT CTTAGTGCGT
2.6 kb CACCATGTTA TCAACATGGA GTGAGTGTGC AAGGGCTCTT CTTAGTGCGT

embryo 6 TTTTTCCCAG GTCGGAAGCA CAGGCACCTC TAGTACAAGA GGCACCAGTT
embryo 8 TTTTTCCCAG GTCGGAAGCA CAGGCACCTC TAGTACAAGA GGCACCAGTT
2.6 kb TTTTTCCCAG GTCGGAAGCA CAGGCACCTC TAGTACAAGA GGCACCAGTT

embryo 6 TGTGGAAGTT CATTCCAGAA TACCTGGAAG CTATCTACGA TAAGTGCGTC
embryo 8 TGTGGAAGTT CATTCCAGAA TACCTGGAAG CTATCTACGA TAAGTGCGTC
2.6 kb TGTGGAAGTT CATTCCAGAA TACCTGGAAG CTATCTACGA TAAGTGCGTC

embryo 6 TAAACGATCC CCAGGTCTTG ATGGCTTGAA CGGAGAGATC TGTAAGAAGCT
embryo 8 TAAACGATCC CCAGGTCTTG ATGGCTTGAA CGGAGAGATC TGTAAGAAGCT
2.6 kb TAAACGATCC CCAGGTCTTG ATGGCTTGAA CGGAGAGATC TGTAAGAAGCT

embryo 6 TGTGGGAAGTT CATTCGGAAA TACCTGGAAG CTATCTAAGA TAAAGTGCGTC
embryo 8 TGTGGGAAGTT CATTCGGAAA TACCTGGAAG CTATCTAAGA TAAAGTGCGTC
2.6 kb TGTGGGAAGTT CATTCGGAAA TACCTGGAAG CTATCTAAGA TAAAGTGCGTC
embryo 6  TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT
embryo 8  TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT
         2.6 kb TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT

embryo 6  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
embryo 8  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
         2.6 kb CCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA

embryo 6  TCACTCCTCT TCCAGTGCTT GGAAAAGTGC TGGAAAGAGA AATGGTGGAA
embryo 8  TCACTCCTCT TCCAGTGCTT GGAAAAGTGC TGGAAAGAGA AATGGTGGAA
         2.6 kb TCACTCCTCT TCCAGTGCTT GGAAAAGTGC TGGAAAGAGA AATGGTGGAA

501 550  
embryo 6  TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT
embryo 8  TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT
         2.6 kb TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT

embryo 6  TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT
embryo 8  TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT
         2.6 kb TGGGAGGGTT ATTTTCCACG TGAATGGAAA AAAGCTAGGG TTAT

551 600  
embryo 6  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
embryo 8  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
         2.6 kb CCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA

embryo 6  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
embryo 8  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
         2.6 kb CCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA

601 650  
embryo 6  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
embryo 8  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
         2.6 kb CCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA

2.6 kb  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA
         2.6 kb  TCTGAAGTCC CCTGATAAAA TCAGGAGCAA TCCTCGCTCC TATCGGGGCA

154
The 3.1aRevF2-B2 primers were also used to survey other Bactrocera species. *Bactrocera carambolae, Bactrocera papayae, Bactrocera philipinensis,* and *Bactrocera pyrifoliae* are considered to be closely related to *B. dorsalis,* all of which are members of the *B. dorsalis* complex of flies (Drew and Hancock 1994). Other Bactrocera species, such as *Bactrocera correcta,* the melon fly *Bactrocera curcurbitae,* *Bactrocera latifrons,* and *Bactrocera occipitalis* are not considered to be members of the *B. dorsalis* species complex.

PCR reactions using genomic DNA samples from adult males and females of the species *B. correcta, B. latifrons, B. occipitalis, B. philipinensis,* and *B. pyrifoliae* as templates all showed no observable amplification products with the 3.1aRevF2-B2 primers under PCR cycling conditions identical to those used in *B. dorsalis* (results not shown). Products were obtained using the 3.1aRev primers in PCRs with genomic DNA from adult males and females from *B. carambolae, B. curcurbitae,* and *B. papayae.*

The 3.1aRevF2-B2 primers produced a single amplification product in both *B. carambolae* males and females that was approximately 2.6 kb in size when long template PCR was performed (Figure 22). This amplification pattern was observed in PCRs with genomic DNA from two adult males and females. The amplification patterns in these two *B. carambolae* males differs from the pattern observed in *B. dorsalis* adult males.
Figure 30. 3.1aRevF2-B2 PCR amplifications using genomic DNA from *B. dorsalis* wild strains from Taiwan and Okinawa (long template PCR), Malaysian *B. carambolae* and *B. papayae* (long template PCR), or *B. curcurbitae* (USDA lab strain, amplified using Taq polymerase) as templates. The amplification products from the *B. dorsalis* wild strains and *B. papayae* are consistent with those seen from the Puna, Hawaii *B. dorsalis* strain (i.e., males have 325 bp and 2.6 kb product, females only the 2.6 kb product). PCRs with *B. carambolae* DNA only amplified the 2.6 kb product in both sexes, while the *B. curcurbitae* amplifications produced only the 325 bp product in both sexes.
The PCR amplification results of *B. curcurbitae* males and females were the exact opposite of those for *B. carambolae*. While PCR reactions with *B. carambolae* males and females gave a 2.6 kb PCR amplification product, PCRs of genomic DNA from three *B. curcurbitae* males and females both amplified only a 325 bp PCR product. The results of PCR reactions with Taq polymerase are shown in Figure 22. Long template PCRs were attempted with these genomic DNAs, but did not show any observable amplification.

3.1aRevF2-B2 PCR amplification patterns in *B. papayae* adult males and females mirrored patterns seen in *B. dorsalis* adult males and females. *B. papayae* males appeared to have only one PCR product, 325 bp in size, when genomic DNA was amplified using *Taq* polymerase. When long template PCR was performed, two products, 325 bp and 2.6 kb in size, were amplified. The long template PCR reaction with genomic DNA from one male (Figure 22) produced an additional amplification product that was approximately 650 bp in size. Both of the *B. papayae* females that were surveyed produced a single amplification product approximately 2.6 kb in size. These results were replicated when long template PCR was performed (Figure 22).
3.10 3.1aRevF2-B2 PCR product clones from *B. papayae* and *B. curcurbitae*

The 325 bp PCR products from four different *B. papayae* adult males were cloned and sequenced. All four products were nearly identical in sequence. Two single nucleotide differences were identified (Figure 23). The 325 bp product from *B. papayae* male 5 contains an A at nucleotide position 38 instead of a G seen in the other three males, and a T instead of a C in nucleotide position 208. Alignments between these *B. papayae* 325 bp product sequences and the *B. dorsalis* 325 bp male PCR product show that the *B. papayae* and *B. dorsalis* PCR products are almost exactly alike in sequence, but also with single nucleotide differences.

The 325 bp PCR products from two different *B. curcurbitae* adult males and females were also cloned and sequenced. The sequence data showed that all of these sequences were virtually identical to each other, with the exception of two different single nucleotide differences (Figure 24). The 325 bp product from *B. curcurbitae* female 2 contained two single nucleotide differences from the rest of all of the *B. curcurbitae* 325 bp products, an A instead of a G at nucleotide position 118 and a G instead of an A at position 199. Both the male and female *B. curcurbitae* PCR products were also virtually identical in sequence to the *B. dorsalis* 325 bp male PCR product. The A present at position 118 of the *B. curcurbitae* female 2 product corresponds to an A present at that position in all five of the *B. dorsalis* 325 bp male products that were sequenced.
Figure 31. Nucleotide sequence alignment of the 325 bp 3.1aRevF2-B2 PCR products from two *B. dorsalis* adult males (Bdor_M1, M4) and four *B. papayae* adult males (Bpap_M1-M5). The nucleotide differences are in color and their locations are indicated by asterisks.
Bpap_M1 ATAGGGACAA CGTCCGGTGT GTTTTAAAAA AGTAGAAGAG CCCTTGCACA
Bpap_M3 ATAGGGACAA CGTCCGGTGT GTTTTAAAAA AGTAGAAGAG CCCTTGCACA
Bpap_M4 ATAGGGACAA CGTCCGGTGT GTTTTAAAAA AGTAGAAGAG CCCTTGCACA
Bdor_M1 ATAGGGACAA CGTCCGGTGT GTTTTAAAAA AGTAGAAGAG CCCTTGCACA
Bpap_M5 ATAGGGACAA CGTCCGGTGT GTTTTAAAAA AGTAGAAGAG CCCTTGCACA
Bdor_M4 ATAGGGACAA CGTCCGGTGT GTTTTAAAAA AGTAGAAGAG CCCTTGCACA

Bpap_M1 CTCACTCTAT GTTGATAACA TGGTGTCACC AACACGAAGA GAGGTTATAT
Bpap_M3 CTCACTCTAT GTTGATAACA TGGTGTCACC AACACGAAGA GAGGTTATAT
Bpap_M4 CTCACTCTAT GTTGATAACA TGGTGTCACC AACACGAAGA GAGGTTATAT
Bdor_M1 CTCACTCTAT GTTGATAACA TGGTGTCACC AACACGAAGA GAGGTTATAT
Bpap_M5 CTCACTCTAT GTTGATAACA TGGTGTCACC AACACGAAGA GAGGTTATAT
Bdor_M4 CTCACTCTAT GTTGATAACA TGGTGTCACC AACACGAAGA GAGGTTATAT

Bpap_M1 CATAGCTTCT ACGACCCCTG CAGATCTTAT AGATCTGCCC CCAGGGATCG
Bpap_M3 CATAGCTTCT ACGACCCCTG CAGATCTTAT AGATCTGCCC CCAGGGATCG
Bpap_M4 CATAGCTTCT ACGACCCCTG CAGATCTTAT AGATCTGCCC CCAGGGATCG
Bdor_M1 CATAGCTTCT ACGACCCCTG CAGATCTTAT AGATCTGCCC CCAGGGATCG
Bpap_M5 CATAGCTTCT ACGACCCCTG CAGATCTTAT AGATCTGCCC CCAGGGATCG
Bdor_M4 CATAGCTTCT ACGACCCCTG CAGATCTTAT AGATCTGCCC CCAGGGATCG

Bpap_M1 TCTTTGTTTT GCCCCACGAA GCTTCTCCAG TCTTCTTCTT TGATTCTGAG
Bpap_M3 TCTTTGTTTT GCCCCACGAA GCTTCTCCAG TCTTCTTCTT TGATTCTGAG
Bpap_M4 TCTTTGTTTT GCCCCACGAA GCTTCTCCAG TCTTCTTCTT TGATTCTGAG
Bdor_M1 TCTTTGTTTT GCCCCACGAA GCTTCTCCAG TCTTCTTCTT TGATTCTGAG
Bpap_M5 TCTTTGTTTT GCCCCACGAA GCTTCTCCAG TCTTCTTCTT TGATTCTGAG
Bdor_M4 TCTTTGTTTT GCCCCACGAA GCTTCTCCAG TCTTCTTCTT TGATTCTGAG

Bpap_M1 GAGCATGCCT TTGTACTCTC TCATCTCAGT ACTAAGATCA TGCCGAATCT
Bpap_M3 GAGCATGCCT TTGTACTCTC TCATCTCAGT ACTAAGATCA TGCCGAATCT
Bpap_M4 GAGCATGCCT TTGTACTCTC TCATCTCAGT ACTAAGATCA TGCCGAATCT
Bdor_M1 GAGCATGCCT TTGTACTCTC TCATCTCAGT ACTAAGATCA TGCCGAATCT
Bpap_M5 GAGCATGCCT TTGTACTCTC TCATCTCAGT ACTAAGATCA TGCCGAATCT
Bdor_M4 GAGCATGCCT TTGTACTCTC TCATCTCAGT ACTAAGATCA TGCCGAATCT

Bpap_M1 GGGCCAACCT ATCGGAATTG GACCGTCTAG CACGTTGAAA CTTCGGTCTC
Bpap_M3 GGGCCAACCT ATCGGAATTG GACCGTCTAG CACGTTGAAA CTTCGGTCTC
Bpap_M4 GGGCCAACCT ATCGGAATTG GACCGTCTAG CACGTTGAAA CTTCGGTCTC
Bdor_M1 GGGCCAACCT ATCGGAATTG GACCGTCTAG CACGTTGAAA CTTCGGTCTC
Bpap_M5 GGGCCAACCT ATCGGAATTG GACCGTCTAG CACGTTGAAA CTTCGGTCTC
Bdor_M4 GGGCCAACCT ATCGGAATTG GACCGTCTAG CACGTTGAAA CTTCGGTCTC

Bpap_M1 AAACGCCTGA CAGTCCTCTT CTTTA
Bpap_M3 AAACGCCTGA CAGTCCTCTT CTTTA
Bpap_M4 AAACGCCTGA CAGTCCTCTT CTTTA
Bdor_M1 AAACGCCTGA CAGTCCTCTT CTTTA
Bpap_M5 AAACGCCTGA CAGTCCTCTT CTTTA
Bdor_M4 AAACGCCTGA CAGTCCTCTT CTTTA
Figure 32. Nucleotide sequence alignment of the 325 bp 3.1aRevF2-B2 PCR products from two *B. dorsalis* adult males (Bdor_M1, M4), two *B. curcurbitae* adult males (Bcur_M1, M2), and two *B. curcurbitae* adult females (Bcur_F1, F2). The nucleotide differences are in color and their locations are indicated by asterisks.
Chapter 4

Summary and Discussion

A number of interesting characteristics distinguish Y type chromosomes from other chromosomes. One such feature is that Y type chromosomes often undergo recombination less frequently than other chromosomes. The human Y contains two "pseudoautosomal" regions at either end of the chromosome that recombine with counterparts on the X chromosome. Otherwise, the majority of the human Y chromosome does not undergo crossing over (reviewed by Lahn and Page 1999). Secondly, Y type chromosomes often contain fewer active genes than other chromosomes. The Y chromosomes of humans, Drosophila, and Tephritids have been characterized as being largely composed of heterochromatin, which is considered to be genetically inactive. Several theories attribute this lack of genetic activity in part to the absence of recombination, which is thought to facilitate the fixation of mutations. The rapid sequence evolution is considered a significant factor in the degeneration of the Y chromosome (Charlesworth 1978). A third characteristic of Y type chromosomes is that they are inherited only by males. In humans and the Tephritid fly Ceratitis capitata, a region of Y chromosome has been implicated in male sex determination (reviewed by Koopman 2001, Willhoeft and Franz 1996). In contrast, the Drosophila Y chromosome, while present only in males, is important for male fertility, but does not participate in sex determination. The diversity of functions attributable to these chromosomes, combined with their heterochromatic composition and overall lack of genetic activity, make them intriguing subjects for further study.
The objective of this dissertation was the isolation and characterization of male-specific (and Y chromosome specific) DNA sequences in the Oriental fruit fly, *Bactrocera dorsalis*. Isolating Y chromosome specific sequences from *B. dorsalis* is of considerable interest for many reasons. One is that it would provide an overall better understanding of the *B. dorsalis* genome. The *B. dorsalis* Y chromosome is described as being particularly small in size when compared to the X chromosome and the other autosomes, appearing as a small darkly-staining dot in mitotic metaphase chromosome spreads (Baimai et al. 1995). Like other Y type chromosomes, it is also largely heterochromatic in nature (Baimai et al. 1995). The diminutive size and heterochromatic nature of the Y in this species makes its characterization all the more challenging.

Secondly, there is interest in investigating if the *B. dorsalis* Y chromosome plays a role in male sex determination. The Y chromosome is implicated in male determination in other Tephritids, such as the Medfly. Lifshitz and Cladera described examples of Medflies with sex chromosome aneuploidies that were phenotypically male. There is also evidence of a male determining element on the Y chromosome in Tephritids (Zapater and Robinson 1986, Willhoeft and Franz 1996). In *B. dorsalis*, both the small size and heterochromatic nature of the Y, combined with an overall lack of knowledge of the genome of this species, have made it difficult to investigate these questions using conventional molecular techniques.
Previous attempts were made to isolate male-specific sequences from *B. dorsalis* by using methods that have been productive in other species. In the end, a method of genomic DNA subtraction called Representational Difference Analysis (RDA), designed especially for the comparison of complex genomes, proved to be the most productive.

### 4.1 Representational Difference Analysis

Representational Difference Analysis (RDA) is a method of comparing complex genomes (Lisitsyn et al. 1993). Restriction enzyme digested DNA from the genome of interest (the "tester") and the genome for comparison (the "driver") is first enriched for sequences less than 1 kb (the enriched DNA pools are called "representations") via PCR. The tester DNA representation is then denatured and hybridized to an excess of the driver representation, producing different hybridization products. The RDA method is designed so that only tester-tester hybrids, representing sequences present only in the genome of interest, are exponentially amplified via PCR. RDA was previously used to isolate Y chromosome-specific DNA from plants (Donnison et al. 1996), a male specific gene from mice (Eriksson et al. 1999), and a variety of male-specific sequences from the Queensland fly *B. tryoni* (Shearman 1999).

RDA was carried out with genomic DNA from *B. dorsalis* adult males and females. Six distinct products were obtained (Figure 4b), ranging from 350 to 600 bp in size. Each of these was characterized, and one (RDA product 1) showed a male-specific hybridization pattern when used as a labeled probe for Southern blotted adult male and female genomic DNA, with male-specific bands that were estimated to be 5 and 10 kb in size (Figure 8). Donnison et al. (1996) characterized four different products resulting
from RDA using the genomic DNA from male and female plants of the dioecious species 
*Silene latifolia* as the tester and driver, respectively. Southern blots of genomic DNA that 
were probed with the individual RDA products revealed the presence of male-specific 
bands, many of whom were several kilobases in size. In the case of at least one of these 
RDA product probes, the male-specific bands appeared within a highly repetitive 
hybridization pattern present in both males and females.

It would be worthwhile to alter the RDA procedure for future attempts at isolating 
*B. dorsalis* Y chromosome specific sequences. Lisitsyn et al. (1993) noted that the success 
of the RDA procedure is entirely dependent upon simplifying the complexity of the 
genomes involved. Performing several RDA hybridizations simultaneously, each using a 
different restriction enzyme to digest the genomic DNA, would be a more effective 
means of isolating multiple male-biased sequences. Lisitsyn et al. (1993) performed 
three different RDA hybridizations with the same tester and driver genomic DNA 
digested with three different restriction enzymes – BamHI, BglII, and HindIII. Shearman 
(1999) performed RDA using all three of these enzymes in *B. tryoni*, with *B. tryoni* male 
genomic DNA as the tester DNA and female genomic DNA as the driver. No RDA 
products were obtained with the HindIII-digested RDA hybridization. However, a 
number of different products were obtained from both the BamHI and BglII-digested 
RDA hybridizations following three complete rounds of hybridization and difference
enrichment. A majority of these RDA products (14 from the BamHI RDA, 15 from the BgIII RDA) were tentatively identified as putative microsatellites and transposon transposase or retrotransposon reverse transcriptase sequences. Donnison et al. (1996) also isolated a greater number of male-specific RDA products from Silene latifolia by using multiple restriction enzymes. A single S. latifolia male-specific product was isolated when RDA was performed using BamHI-digested genomic DNA. 36 RDA products were isolated when RDA was performed with BgIII-digested DNA.

The complete sequence of RDA product 1 has yet to be determined due to the difficulties that were encountered in its cloning. One such difficulty is attributable to the small size range of these products, between 350 and 600 bp. The DNA from other RDA products sizes was often carried over while excising individual products from an agarose gel for purification, and might be preferentially incorporated into the plasmid cloning vector. Moreover, the PCR difference enrichment of the secondary hybridization product shows six distinct RDA products within a larger DNA smear, perhaps representing single-stranded DNA that was not removed when the second mung bean nuclease digestion was omitted. The presence of these products after the secondary hybridization could have interfered with the successful isolation of the RDA products as well.
4.2 Characterization of RDA product 1

A bacteriophage genomic DNA minilibrary, with phage clones containing fragments of *B. dorsalis* genomic DNA approximately 5 kb in size, was constructed to isolate the 5 kb male-specific band seen in the Southern blot of *B. dorsalis* male genomic DNA probed with RDA product 1 (Figure 8). The minilibrary was also screened by probing with RDA product 1. This led to the isolation of eight λ phage clones (Figure 9), each of which representing *B. dorsalis* male genomic sequences with some sequence similarity to the RDA product 1. One of these clones, 3.1a, exhibited sequence similarity at one end ("3.1aRev", Figure 8) to the amino acid sequences known R1 retrotransposable element reverse transcriptase sequences from other insect species, particularly those in the genus Drosophila (Figure 12 and 13). To date, this represents the first time that the presence of R1 retrotransposable elements has been reported in Tephritid flies.

4.3 Characterization of the 3.1a clone

A Southern blot of *B. dorsalis* male and female genomic DNA was probed with the 3.1a clone to determine its representation in the genome. The 3.1a sequence produced identical repetitive patterns in both males and females (Figure 14). It would be worthwhile to repeat this Southern blot using genomic DNA digested by other enzymes to verify its equal representation in males and females.

Additional work was done to sequence the remainder of the 3.1a clone, for two primary reasons. First, additional sequence data would potentially reveal if the clone contained part of an intact R1 element. Secondly, as R1 elements are known to be
differentially represented in the 28S rRNA genes found on the X and Y chromosomes in Drosophila (Tartof and Dawid 1976), the 3.1a clone could serve as a gateway into either of these chromosomes in *B. dorsalis*. In all, 2031 bp of sequence was determined from the 3.1aRev end, which exhibits sequence similarity to R1 element reverse transcriptases. 883 bp was sequenced from the other end ("3.1aFor"), using a combination of restriction enzyme digestion and primer walking (Figures 16 and 17). The additional sequence data, with the exception of the 3.1aRev sequence data, does not show significant similarity to sequences in the GenBank database. PCR primers that were designed for the 3.1aFor end of the clone also do not exhibit sex-specific amplification patterns, while PCR primers designed for the 3.1aRev end of the clone and for the long template PCR amplification of most of the 3.1a clone insert display sex-specific amplification patterns.

### 4.4 Sex-specific PCR products in *B. dorsalis* adults

PCR primers designed for one end of the 3.1a clone (3.1aRev, which shows sequence similarity to R1 element reverse transcriptases) amplify in a sex-specific manner in *B. dorsalis* adults, producing a 325 bp and 2.6 kb product in males and only a 2.6 kb product in females (Figure 19). The 325 bp products from five different *B. dorsalis* adult males are consistent in sequence (Figure 20), with occasional single nucleotide differences. These single nucleotide differences could be attributed to PCR amplification errors generated by the DNA polymerase.
The 2.6 kb product from a single *B. dorsalis* adult female shows the presence of a putative ORF that is 1752 bp in size and encodes 583 amino acids (Figure 15). When a BLASTX search of the GenBank database was performed, both of these PCR products exhibited sequence similarity to a number of R1 reverse transcriptase sequences in other insects, including *Drosophila*. The translated 2.6 kb product ORF also contains several amino acid motifs common to *Drosophila* R1 element reverse transcriptases at locations comparable to their *Drosophila* counterparts (Figure 21).

Alignments of the 325 bp male-specific product and the 3.1aRev sequence (from one end of the 3.1a clone) confirm that this product is the result of PCR amplification within the 3.1aRev sequence (Figure 24). While there is a great deal of sequence similarity between most of the 325 bp male and one end of the 2.6 kb female product, the lack of correspondence between the latter portion of the two sequences shows that the 325 bp product is not the result of amplification from within the 2.6 kb sequence. Therefore, the 325 bp product originates from the 3.1aRev sequence and not the 2.6 kb product. The 2.6 kb product sequence also does not correspond well to the 3.1aRev sequence. These alignment results suggest that the 325 bp and 2.6 kb products are sequences that have been amplified from different locations within the *B. dorsalis* genome.

When used as a probe for Southern blots of *B. dorsalis* genomic DNA, both of the PCR products produced repetitive hybridization patterns in both males and females (Figure 26). However, the 325 bp male-specific sequence produced additional bands in the male genomic DNA hybridization pattern that were not seen in the female genomic DNA hybridization pattern. This repetitive pattern of probe hybridization, containing
single male-specific bands, is consistent with the published results of Drosophila male and female genomic DNA blots probed with sequences from the non-LTR retrotransposons \textit{G} and \textit{jockey}. \textit{In situ} hybridizations of sequences from these elements to Drosophila mitotic chromosomes confirmed that the male-specific copies of these elements observed in the Southern blots were indeed present on Drosophila Y chromosome (Terrinoni et al. 1997). Pimpinelli et al. (1995) also identified copies of the retrotransposable elements \textit{copia}, \textit{gypsy}, \textit{mgd-1}, \textit{blood}, \textit{Doc}, \textit{I}, and \textit{F} that were present throughout the heterochromatic Drosophila Y chromosome via \textit{in situ} hybridization. A Y chromosome origin of copy of the \textit{gypsy}-like LTR retrotransposon \textit{yoyo} in the Medfly was also initially identified through Southern blotting (Zhou and Haymer 1998).

The 3.1aRev primers also produce sex-specific amplification patterns in \textit{B. dorsalis} strains from Taiwan, Okinawa, and Kalaheo, Hawaii that are consistent with the Puna, Hawaii lab strain. The consistency of the sex-specific amplification pattern, combined with the alignment and Southern blot hybridization results, lead to the present working hypothesis that these PCR products originate from the \textit{B. dorsalis} sex chromosomes – a 2.6 kb “X chromosome” copy, and a 325 bp “Y chromosome” copy. This would explain the presence of two products in males (a 2.6 kb “X chromosome” copy and a 325 bp “Y chromosome” copy) and only one in females (which would presumably have two-2.6 kb “X chromosome” copies). Work has already begun to confirm of the location of the 325 bp male-specific product to the \textit{B. dorsalis} Y chromosome via \textit{in situ} hybridization. It would be expected from the Southern blot results (Figure 26) that the 325 bp product would be located on both the Y chromosome and elsewhere in the genome. It would also be worthwhile to repeat the \textit{in situ}
hybridizations with the 2.6 kb product. Southern blot results confirm a repetitive representation of this sequence (Figure 26). At least two possible representations of this sequence could exist in the *B. dorsalis* genome. The first possibility is that the 2.6 kb product sequence is found on the *B. dorsalis* X chromosome and elsewhere in the genome. The second possibility is that it is repetitively represented on the autosomes and is not found on the X chromosome.

4.5 Sex-specific PCR products in other stages of the *B. dorsalis* life cycle

Currently, the preferred method of implementing the Sterile Insect Technique (SIT), as a means of population control in Tephritids, centers around the mass rearing and release of sterile males (reviewed by Franz and Kerremans 1994). To expedite the process of male mass rearing, various “genetic sexing strains” have been developed in the Medfly, where flies have been genetically altered in such a way that their propagation results in the exclusive production of males. *B. dorsalis* is at somewhat of a disadvantage, relative to species such as the Medfly, because its overall lack of known molecular markers for the development of new genetic sexing strains (reviewed by Willhoeft et al. 1996). Thus, the development of new sexing techniques in this species, utilizing male-specific markers that can be identified early in the *B. dorsalis* life cycle, would be of great interest.
The consistency of the sex-specific amplification patterns produced by the 3.1aRev primers in *B. dorsalis* adults make them natural candidates for use as tools for distinguishing between the sexes at different stages of the *B. dorsalis* life cycle. Initial surveys of *B. dorsalis* embryos and third instar larvae show that the amplification patterns of the 3.1aRev primers resemble those seen in adults, namely the amplification of a 325 bp and a 2.6 kb product in some individuals and a 325 bp product in others (Figure 27). Alignment of sequences of the 325 bp and the 2.6 kb products from 24 hour old embryos show that they are nearly identical to their adult counterparts (Figures 28 and 29). It would also be useful to confirm that these amplification patterns are indeed sex-specific in these life cycle stages, perhaps by identifying the gonads in third instar larvae prior to processing the tissue for the extraction of genomic DNA.

PCR-based sexing can be an efficient alternative to other sexing methods. A PCR-based sexing method has also been devised and tested in chickens, amplifying repetitive sequences and other DNA markers associated with the W chromosome, and has been found to be fairly reliable (Klein and Ellendorff 2000). Although this type of PCR-based sexing could not be applied to live embryos or larvae, the 3.1aRev primers provide a rapid method of sexing tissue from these developmental stages for other molecular experiments.
4.6 PCR products in other Bactrocera species

Yong (1995) compared the allele frequencies of 17 different enzyme loci amongst five species belonging to the *B. dorsalis* complex (including *B. dorsalis*, *B. carambolae*, and *B. papayae*) and concluded that *B. papayae* was most closely related to *B. dorsalis* when compared to the other Bactrocera species included in the survey. Interestingly, the amplification patterns from *B. dorsalis* and *B. papayae* are identical to each other (males have 325 bp and 2.6 kb product, females have only the 2.6 kb product), while the patterns observed in *B. carambolae* (both males and females have only the 2.6 kb product) and *B. curcurbitae* (both males and females have only the 325 bp product) each differ from the patterns from the other three species. However, alignments of the 325 bp PCR product sequences from these four species show that they are virtually indistinguishable from each other. This is not surprising, considering that the presence of R1 elements is widespread in insects (Jakubczak et al. 1991).

While the 3.1aRev primer amplification patterns may be quite limited in its ability to distinguish between Bactrocera species, other PCR-based methods have been effectively used to differentiate between different populations within a single Tephritid species, or to differentiate between two Tephritid species altogether. One example is the method combining exon-primed intron-crossing PCR (EPIC-PCR), amplifying introns within actin genes, with restriction enzyme digestion to identify different populations of
B. dorsalis (He and Haymer 1997). This combination of techniques is capable of making fine distinctions, even between different Hawaiian B. dorsalis populations. Another EPIC-PCR approach, amplifying internal transcribed spacer (ITS) regions present between ribosomal rRNA genes, produces differently-sized products in C. capitata and C. rosa. There is also a difference in the EPIC-PCR product sizes that distinguishes among C. rosa populations (Douglas and Haymer 2001).

4.7 Conclusions

The nonrecombining Y chromosome has long been considered a haven for transposable elements. The RNA transcripts of male fertility factors on the D. hydei Y chromosome, called "lampbrush loops", are known to contain degenerate copies of transposable elements as a number of other repetitive sequences (reviewed by Hackstein and Hochstenbach 1995). Steinemann and Steinemann (2000) hypothesize that the degradation of the Y chromosome is initiated by the accumulation of transposons, particularly retrotransposons. The Y chromosomes of D. melanogaster and D. miranda, while sharing few functional genes, are still similar in structure and harbor many of the same transposons. They theorize that the transposons amassed on the Drosophila Y chromosome are responsible for the inactivation of functional Y-linked genes. In the D. miranda neo-Y chromosome, evidence of gene inactivation by transposable elements exists in the larval cuticle protein gene Lcp4, which has been shown to be inactive due to a number of sequence insertions, including inserted retrotransposons, while the neo-X chromosome contains an active copy of Lcp4 (Steinemann and Steinemann 1992). Losada et al. (1997) characterized a region near the Drosophila Y chromosome
centromere approximately 200 kb in size and found that it contained a number of HeT-A retrotransposable elements that had previously been found in the telomeres of chromosomes. A possible correlation has been established between an abundance of transposable elements and a reduced rate of recombination. Transposable elements have been found to be in significantly higher concentrations around regions of chromosomes containing inversions, where recombination is suppressed. Retrotransposons have also been found to be plentiful in the X and Y chromosomes of deer mice, where recombination is greatly diminished in comparison to the autosomes (reviewed by Charlesworth et al. 1994).

The DNA sequences isolated in this dissertation are repetitively represented in the B. dorsalis genome. One of these sequences, 325 bp in size, appears to be enriched in the male genome. PCR amplification of this DNA also appears to occur only in B. dorsalis males. These data suggest that it might be located on the Y chromosome. Vogt and Hennig (1983) described Y-associated sequences as those found on the Y chromosome and other chromosomes. If the 325 bp sequence truly is a Y chromosome-associated sequence, then it has great potential as a tool to gain access into other regions of the Y chromosome. Confirmation that this 325 bp sequence is located on Y chromosome via in situ hybridization would imply that the 3.1a clone, from which this sequence originates, is also a cloned portion of the B. dorsalis Y.
The diminutive *B. dorsalis* Y chromosome is indeed a curiosity. It persists despite of all known evolutionary forces governing genome maintenance that would favor its elimination from the *B. dorsalis* male genome. If the Y is so small and heterochromatic, why is it still retained in *B. dorsalis* males? The logical answer is that its presence is essential for males of this species. If so, does the Y chromosome contain male-determining or male fertility genes? With the use of more sophisticated molecular techniques, the answers to these questions feel close within reach.
Bibliography


Jakubczak J.L., Xiong Y., and Eickbush T.H. 1990. Type I (R1) and Type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. *J. Mol. Biol.* 212: 37-52.


