INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
MOLECULAR CHARACTERIZATION OF THE WHITE EYE GENE
OF THE ORIENTAL FRUIT FLY, BACTROCERA DORSALIS
(HENDEl), (DIPTERA: TEPHRITIDAE)

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ENTOMOLOGY

AUGUST 1995

By

Wendy X.Z. Chang

Dissertation Committee:

Stephen H. Saul, Chairperson
Ronald F.L. Mau
George K. Roderick
Donald O. McInnis
Gert H. De Couet
© Copyright 1995

by

Wendy X.Z. Chang
Dedicated to my parents, Zhaoxian Zheng and Xiasong Chen, and my husband Jingbo Chang. With your love and care, understanding and encouragement, and financial and emotional support I can be what I am today.
Acknowledgements

I would like to thank my advisor, Dr. Stephen H. Saul, for providing the opportunity to conduct fruit fly research in Hawaii. I thank him for his advise, understanding, the independent atmosphere he fostered, and the financial support throughout my Ph.D program.

Special thanks to my colleague, Dr. Susan D. McCombs, for her helpful discussion, reading the proposal and the dissertation, and the friendship in my daily life. Thank you to my colleagues Jingwei Xiao and Stephan Lee, and BMIBF manager Dr. Neil Reimer for their support to my research, and my friends Dr. Mei He and Qing Zhou for sharing their technical experience.

A sincere thank you to my committee members Drs. Gert De Couet, Ronald Mau, Diane Ullman, George K. Roderick, and Donald McInnis who have devoted their precious time and effort in preparing me for a professional career.
Abstract

The wild-type white eye gene of the oriental fruit fly, *B. dorsalis* (Hendel), has been cloned and partially sequenced. The 3859 bp DNA sequence includes exons 2, 3, 4a, 4b, 5, and 6, introns 2, 3, 4, 5, and part of intron 1, and the polyA addition site.

Compared to the white gene of *D. melanogaster*, there is an extra intron splitting exon 4 into two exons, 4a and 4b, in the white eye gene of the oriental fruit fly. The corresponding exons in the white eye genes of the two species are similar in size and the phases are the same. The two white eye genes are highly homologous to each other. The coding region sequence of the two genes has similarity of 93% at the amino acid level and 70.8% at the nucleotide level. The inconsistency of the two measures of similarity can be explained by high levels of silent mutation, 71.4%. Silent mutation shows different pattern of codon usage bias in the two white eye genes. *D. melanogaster* prefers GC base while the oriental fruit fly prefers AT base at the silent sites of the codon.
Table of Contents

ACKNOWLEDGEMENTS ........................................................................................................... v
ABSTRACT ................................................................................................................................. vi
LIST OF TABLES ......................................................................................................................... ix
LIST OF FIGURES ...................................................................................................................... x
LIST OF ABBREVIATIONS .......................................................................................................... xii

1. INTRODUCTION .................................................................................................................. 1
  1.1 INTRODUCTION TO THE ORIENTAL FRUIT FLY .......................................................... 1
    1.1.1 Distribution and host range ......................................................................................... 1
    1.1.2 Economic impact ......................................................................................................... 1
    1.1.3 Control methods ......................................................................................................... 2
  1.2 SIGNIFICANCE OF THE WORK ....................................................................................... 8
  1.3 CURRENT AND PREVIOUS RELATED RESEARCH ......................................................... 8
    1.3.1 White gene in D. melanogaster .................................................................................. 8
    1.3.2 White eye mutants of the oriental fruit fly and other insects .................................... 18
    1.3.3 Actin genes of the oriental fruit fly and the Mediterranean fruit fly ....................... 19

2. MATERIALS AND METHODS .............................................................................................. 20
  2.1 MATERIALS ....................................................................................................................... 20
    2.1.1 Oriental fruit fly strain ............................................................................................... 20
    2.1.2 Escherichia coli strains .............................................................................................. 20
    2.1.3 Plasmids ..................................................................................................................... 20
    2.1.4 Reagents and systems ............................................................................................... 20
  2.2 METHODS .......................................................................................................................... 21
    2.2.1 Fruit fly rearing ........................................................................................................... 21
    2.2.2 Genomic DNA isolation ............................................................................................ 21
    2.2.3 Genomic DNA library construction ........................................................................... 22
    2.2.4 Library screening ....................................................................................................... 22
    2.2.5 Recombinant λ-phage analysis ................................................................................... 24
    2.2.6 Sequencing ................................................................................................................ 29
    2.2.7 Sequence data analysis ............................................................................................. 30

3. RESULTS AND DISCUSSION ............................................................................................... 31
  3.1 GENOMIC DNA LIBRARY CONSTRUCTION ................................................................... 31
  3.2 GENOMIC LIBRARY SCREENING ................................................................................... 31
  3.3 SEQUENCING .................................................................................................................... 32
  3.4 SEQUENCE ANALYSIS ..................................................................................................... 32
List of Tables

Table                                                   Page
TABLE 1. EXON SIZE AND PHASE COMPARISON OF THE WHITE EYE GENES OF B. DORSALIS AND D. MELANOGASTER................................................................................. 44

TABLE 2. PAIRWISE COMPARISON OF THE WHITE EYE GENES AT THE DNA LEVEL AND THE PROTEIN........................................................................................................ 48

TABLE 3. CODON USAGE OF THE WHITE EYE GENES IN B. DORSALIS, D. MELANOGASTER, AND L. CUPRINA........................................................................................................ 55

TABLE 4. CODON BIAS AND GC CONTENT COMPARISONS FOR THE B. DORSALIS AND D. MELANOGASTER WHITE EYE AND ACTIN GENES. ........... 59
<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 1. D. MELANOGASTER EYE PIGMENTS BIOSYNTHETIC PATHWAY (SUMMERS ET AL., 1982; TEARLE, 1991)</td>
<td>10</td>
</tr>
<tr>
<td>FIGURE 2. UPSTREAM DNA REGULATORY ELEMENTS OF THE D. MELANOGASTER WHITE GENE</td>
<td>13</td>
</tr>
<tr>
<td>FIGURE 3. PROPOSED ACTION SITES OF THE WHITE PROTEIN (WE) (MODIFIED FROM TEARLE, 1991)</td>
<td>16</td>
</tr>
<tr>
<td>FIGURE 4. THE RELATIONSHIP OF THE PBLUE-SX4, PBLUE-ECOR3, PBLUE-CC5, PBLUE-EC1, PBLUE-DW-CLA, AND PSP-DW-CLA PLASMIDS</td>
<td>27</td>
</tr>
<tr>
<td>FIGURE 5. THE RELATIVE POSITIONS OF THE SEQUENCED SEGMENTS OF DIFFERENT SUBCLONES</td>
<td>33</td>
</tr>
<tr>
<td>FIGURE 6 DNA SEQUENCE AND THE POSSIBLE TRANSLATED PROTEIN SEQUENCE OF THE WHITE EYE GENE OF THE ORIENTAL FRUIT FLY</td>
<td>34</td>
</tr>
<tr>
<td>FIGURE 7 ALIGNMENT OF THE INTRON-EXON JUNCTION SEQUENCES</td>
<td>39</td>
</tr>
<tr>
<td>FIGURE 8. COMPARISON OF THE WHITE EYE GENE MODEL OF D. MELANOGASTER AND B. DORSALIS</td>
<td>42</td>
</tr>
<tr>
<td>FIGURE 9. PROTEIN SEQUENCE ALIGNMENT OF THE WHITE EYE GENES FROM B. DORSALIS (B.D), D. MELANOGASTER (D.M), AND L. CUPRINA (L.C)</td>
<td>46</td>
</tr>
<tr>
<td>FIGURE 10. CODING REGION DNA SEQUENCE ALIGNMENT OF THE WHITE EYE GENES FROM B. DORSALIS (B.D), D. MELANOGASTER (D.M), AND L. CUPRINA (L.C)</td>
<td>49</td>
</tr>
<tr>
<td>FIGURE 11. HYDROPATHY PROFILE AND THE SECONDARY STRUCTURE OF THE WHITE EYE PROTEIN OF B. DORSALIS</td>
<td>61</td>
</tr>
</tbody>
</table>
FIGURE 12. HYDROPATHY PROFILE AND THE SECONDARY STRUCTURE OF THE WHITE PROTEIN OF D. MELANOGASTER.................................................................63

FIGURE 13. HYDROPATHY PROFILE AND THE SECONDARY STRUCTURE OF THE BROWN PROTEIN OF D. MELANOGASTER.................................................................66

FIGURE 14. HYDROPATHY PROFILE AND THE SECONDARY STRUCTURE OF EXON 1 OF THE WHITE PROTEIN OF D. MELANOGASTER.................................................................69

FIGURE 15. ALIGNMENT OF THE AMINO ACID SEQUENCES OF THE ATP-BINDING SITES.................................................................................................................71

FIGURE 16. AMINO ACID SEQUENCE SIMILARITY OF THE WHITE (W) AND BROWN (BR) PROTEINS OF D. MELANOGASTER...........................................................................75

FIGURE 17: ATP-ASSOCIATED TRANSPORT SYSTEM EVOLUTION MODEL (MODIFIED FROM KANE, 1990)................................................................................................78
List of Abbreviations

°C: degree celsius
µg: microgram
µl: microliter
A: Adenosine
bp: base pair
C: Cytidine
CTAB: hexadecyltrimethylammonium bromide
dATP: deoxy-adenine triphosphate
dGTP: deoxy-guanosine triphosphate
EDTA: ethylenediamine tetraacetic acid
g: gram
G: Guanosine
h: hour
kb: kilobase
M: mole
min: minute
ml: milliliter
mM: mini mole
pH: potential of hydrogen
s: second
T: Thymidine

TE: 10 mM Tris-HCl (pH 7.4), 1mM EDTA buffer

Tris: Trishydroxymethylaminomethane

v/v: volume/volume

w/v: weight/volume
1. Introduction

1.1 Introduction to the Oriental fruit fly

1.1.1 Distribution and host range

The family Tephritidae includes about 4000 species arranged in 500 genera (White & Elson-Harris, 1992). There are four pest species in Hawaii. They are the oriental fruit fly (*Bactrocera dorsalis* (Hendel)), the Mediterranean fruit fly (*Ceratitis capitata* (Wiedemann)), the melon fly (*Bactrocera cucurbitae* (Coquillett)), and the malaysian fly (*Bactrocera latifrons* (Hendel)). The oriental fruit fly is of Indo-Malayan origin and is one of the 40 species in a complex that is distributed through Oriental Asia: Bhutan, southern mainland China, India, Myanmar, Thailand, Sri Lanka, Nepal, Laos, Vietnam, Cambodia, Taiwan, numerous Pacific islands, including Hawaii (Drew & Hancock, 1994; White, 1991), and occasionally California (Mitchell *et al.*, 1995). It was introduced to Hawaii in 1946 by armed services personnel returning from Saipan (Fullaway 1946). The Hawaii population has been recorded from over one hundred species of host plants, for example, orange, peach, guava, and papaya. Strawberry and common guava are the major reservoir hosts (Drew, 1982).

1.1.2 Economic impact

The oriental fruit fly can cause direct and indirect economic damage, which is connected to its life cycle. The female oviposits into the fruit where larvae develop until the mature 3rd instars pop out to pupate in the soil. Several days later the adults emerge to continue the cycle of damage (White & Elson-Harris, 1992).
The direct damage is caused by fruit flies depositing their eggs beneath the fruit skin. The oviposition puncture opens the fruit to microbes that can decrease the storage potential of the fruit or even render the fruit unmarketable.

The potential direct damage hinders agricultural development in endemic areas. Many crops have been seen as unsuitable because of the control cost and the expense of commodity treatment to meet quarantine regulations for export to non-infested areas.

The potential damage also poses significant quarantine costs to both the importing and exporting countries. Importing countries have to impose quarantine regulations to limit establishment or further spread of the fly. If the introduction happens, it requires extensive and costly eradication programs. Quarantine restrictions imposed by an importing country can either deny a producing country a potential export market, or force the producer to carry out expensive disinfestation treatment.

1.1.3 Control methods

Due to the economic impact of the oriental fruit fly many control measures are used. These measures include quarantine regulations, cultural control, biocontrol, baits sprays, male annihilation, the sterile insect release method (SIRM), and post harvest disinfestation.

1.1.3.1 Biocontrol

Parasitic wasps are the major biocontrol agents and several of them have been introduced into Hawaii (Clausen et al., 1965; Haramoto & Bess, 1970). The primary natural enemy of the oriental fruit fly in Hawaii is the egg parasitic wasp, Biosteres arisanus (Sonan). B. vandenboschi (Fullaway) parasitizes the first instar larvae.
Diachasmimorpha longicaudata (Ashmead), Biosteres tryoni (Cameron), and Psyttalia incisi (Silvestri) attack the second and third instar larvae, (Wong et al., 1984a; Wong & Ramadan, 1987; Stark et al., 1991). The parasitoids complete their development in the host and emerge from the host puparium. The parasitization rates could reach a maximum of about 40% from each fruit host yearly and some times up to 60% in peach. Stark surveyed the parasitoids abundance in peach tree by the canopy fogging sampling method and found that B. arisanus accounts for up to 80% of the total parasitization, D. longicaudata 10%, P. incisi 10%, and B. vandenboschi 0.25% (Stark et al., 1991). The abundance of the parasitoids maybe different according to the sampling method and the host (Wong et al., 1984a; Wong & Ramadan, 1987).

Predators, pathogens, and nematodes are less effective biocontrol agents compared to parasitic wasps for tephritid fruit flies. Three species of ants, Pheidole megacephala (Willard, 1927), Iridomyrex humilis (Wong et al., 1984b), and Solenopsis geminata (Wong & Wong, 1988) attack the larva and pupae of the fruit flies. Ant predation can be considerable and important in local areas but inadequate to regulate the fruit fly populations. Two species of crabronine wasps, Crabro tumidoventris and C. unicolor (Willard, 1927) store adult fruit flies for their young. Earwigs have been observed as predators also (Willard, 1927). Birds, toads, geckos, and spiders also fed on fruit flies (Mitchell & Saul, 1990). Lab and field tests showed that the use of entomogenous nematode, Steinernema feltiae Filipjevr, would be cost prohibitive for oriental fruit fly larvae control because the concentration required is too high (Lindegren & Vail, 1986; Lindegren et al., 1990). Several pathogens, the bacterium Serratia marcescens (Moore &
Nadel, 1961), the fungus *Beauveria bassiana*, and a microsporidian *Nosema tephrititae* (Steiner & Mitchell, 1966; Fujii & Tamashiro, 1972), have been studied. None of them have been developed into biocontrol agents.

1.1.3.2 **Bait sprays**

Hagen (1953) found that protein hydrolysate is a feeding stimulant that provides nutrients necessary for sexual maturation of fruit flies. When mixed with a toxicant, usually malathion, the bait attracts and kills both sexes of the oriental fruit fly (Steiner 1952, 1954, 1957). The method was used extensively in the 1982 California medfly eradication program (USDA-APHIS, 1985). Environmental and human health concerns limit the use of bait sprays in Hawaii.

1.1.3.3 **Male annihilation**

Male oriental fruit flies are attracted and killed by the mixture of toxicant naled and synthetic lure, methyl eugenol (Steiner *et al.*, 1965; Koyama *et al.*, 1984). The wild population can be reduced effectively by eliminating the mate available for females. The effectiveness of the method was extensively proved in the oriental fruit fly eradication program carried out in Mariana Islands (Steiner *et al.*, 1970), Okinawa (Koyama *et al.*, 1984), and recently in California (CDFA, 1993). The use of the method will be limited in Hawaii because methyl eugenol is not registered for use in the state, and the potential carcinogenicity of the lure (Miller *et al.*, 1983). There is also concern about the side effect of the toxicant on insect pollinators and endangered and endemic species (USDA-APHIS, 1985).
1.1.3.4 Sterile insect release method (SIRM)

The sterile insect release method is the most attractive control measure because it is a nontoxic method that minimizes the ecological and environmental side effects (Steiner et al., 1970). It uses no pesticides and is species specific. The method is very effective against low level populations where high overflooding ratios are easier to sustain. The basic idea is to mass-release the laboratory-reared sterile male flies into the infested area to mate with the wild females so that no fertile offspring will be produced. The wild population will greatly decline over time and ultimately be eliminated (Knipling, 1979).

Improvement of SIRM.

One of the problems of the current SIRM technique is that it can not separate the male from the female. There are several advantages of releasing males only in eradication programs (Saul, 1986a; Franz & Kerremans 1994). First, it reduces the mass rearing and the monitoring costs of these programs. The removal of females early in development would result in an effective doubling of the rearing capacity of present facilities. In the process of the eradication program the size of the wild population can be estimated by monitoring the trapped females only. Second, it reduces intrapopulational mating, the mating between the sterile males and sterile females. Such kind of mating greatly reduces the effective population of the sterile male for the control program. Third, it eliminates sting damage caused by the female.

All these advantages can be achieved with an efficient means of genetic sexing. Genetic sexing is to separate the two sexes of the insects by using the genetic method. There are two principle components of the genetic sexing system, a selective marker and
the method to link the marker to the Y-chromosome or sex-specific genes. In such a system, males survive under selective condition while females are eliminated (Saul, 1984).

The current proposed genetic sexing systems are based on the traditional translocation strain using chemicals or physical traits as the selection agent. Several systems have been proposed for the Mediterranean fruit fly using expensive or toxic chemicals as selective agents. These systems include those based on insecticide resistance gene using dieldrin (Busch-Petersen & Wood, 1983; Wood et al. 1985; Rössler, 1986), on the alcohol dehydrogenase gene using ethanol (Robinson & Van Heemert, 1982a, 1982b; Rössler, 1985a), and on the xanthine dehydrogenase gene using purine (Saul, 1982). A pupal color dimorphism based genetic sexing system has been described in the Mediterranean fruit fly (Rössler, 1979a, b; Robinson & Riva, 1983; Robinson et al., 1986) and the oriental fruit fly (McCombs & Saul, 1995). The system requires expensive pupal sorting machines and a high input of energy and labor. It is not completely accurate in sorting and could cause slight damage to the pupae. The third type of genetic sexing system described is based on a translocation stock of a flightless mutant in which males are wild-type and females are flightless (Saul, 1990; McCombs & Saul 1992b; McCombs et al., 1993). Though the system removes the females and is easy to incorporate into current rearing facilities, the mass-rearing costs are still the same. The fourth system takes advantage of a temperature sensitive lethal mutation (Franz & Kerremans, 1994). The genetic sexing strain eliminates the females as embryos by elevating the temperature. This system is more cost effective than the others discussed previously.
Besides the difficulties in the above translocation based genetic sexing systems, there is also potential for translocation breakdown (Rossler, 1982a, b, 1985b; Robinson, 1984; Hooper et al., 1987). Genetic recombination in males between the marker gene and the translocation break point can destroy the base of the sex sorting. Basically, there are two methods to solve the problem. By bringing the translocation breakpoint close to the selective marker, recombination will be reduced to a frequency below 0.1% (Franz et al., 1994). Another method is to reduce the size of selectable markers to genes instead of a piece of chromosome. It can minimize the difficulties such as the chromosome instability, recombination, and mutant reversions (Handler & O’Brochta, 1991; Saul, 1986b). The genetic damage due to transformation would be negligible compared to translocations. The selectable marker can be linked to sex-specific genes, such as a chorion gene (Tolias et al., 1990) a vitellogenine gene (Rina & Savakis, 1991), or the sex-determination gene. In D. melanogaster, sex is determined by the ratio of the X-chromosome to the autosome through different splicing of the sex determination genes (Lewin, 1994). Therefore, in D. melanogaster, the resistant marker can be linked to the exon 6 of the double sex gene, which is present in males only. The dominant lethal gene can be inserted in the exon 4 of the sex lethal gene, which exists in female only. In both situations, females will be eliminated under restrictive conditions.

In order to manipulate the oriental fruit fly at the molecular level a functional transfer vector system is required. With the vector, we can introduce genetic material in vitro and integrate it into the host genome.
1.2 Significance of the Work

The objective of my research is to characterize the wild-type white eye gene of the oriental fruit fly. Through this work, we can have important experience in handling this pest at the molecular level, obtain some basic information on the genome, and provide an important marker gene for a future gene transformation system.

The use of gene transformation is significant in its potential to facilitate an understanding of the pest species genetics, biochemistry, development and behavior (Serano et al., 1994). A more complete understanding of insect biology, such as the chemical resistance mechanism, sex-determination, hybrid sterility, and hormone action and metabolism, would in turn certainly enhance current management programs, and promote development of new ones (Handler & O’Brochta, 1991).

1.3 Current and Previous Related Research

This work relies heavily on the previous genetic and molecular research on D. melanogaster Meigen, the Mediterranean fruit fly, and the oriental fruit fly.

1.3.1 White gene in D. melanogaster

1.3.1.1 Eye Pigments of D. melanogaster

There are two major types of eye pigments in the compound eye, the brown ommochromes and the red pteridines. The main ommochrome pigment, xanthommatin, is biosynthesized from tryptophan, and the red pteridine from guanosine triphosphate. These biosynthetic pathways are understood from studies of enzymes catalyzing each reaction in D. melanogaster. The two pathways interact with each other by the provision of enzyme cofactors for each other, and the common use or control of the uptake and storage of the
pigment precursors by the two pathway (Summers et al., 1982) (Fig. 1). White gene product maybe involved in transporting the substrates and the reaction intermediates across different membranes (Summers et al., 1982; Tearle, 1991).

1.3.1.2 **White Gene of *D. melanogaster*.**

1.3.1.2.1 **White mutants**

Morgan (1910) discovered the first white mutant, *w*′, of *D. melanogaster*. There are 109 polymorphisms found in the white locus with expression levels ranging from complete to partial loss of ommochromes and pteridines (Lindsley & Zimm, 1992; Miyashita & Langley, 1988). The mutant alleles do not affect the viability and fertility of the flies (Lindsley & Zimm, 1992). All mutants are positively phototactic but have no optomotor responses (Kalmus, 1943). Wild-type alleles are incompletely dominant over mutant alleles (Lindsley & Zimm, 1992).

1.3.1.2.2 **White gene**

Several white alleles have been cloned (Bingham et al., 1981; Levis & Rubin, 1982; Levis et al., 1982; Goldberg et al., 1982; Pirrotta et al., 1983). The white transcript, a 2.6 kb poly-(A) RNA, was found in embryos, larvae, pupa, and adults (O’Hare et al., 1983; Pirrotta et al., 1983; Fjose et al., 1984; Pirrotta & Bröckl, 1984). The nucleotide sequence of more than 14 kb of white DNA and the probable structure of the introns and exons were determined by O’Hare et al. (1984). There are six exons and five introns, and the first intron is about 3.1 kb long (O’Hare et al., 1984). A 9.9 kb segment with regulatory elements in the 5′ flanking region DNA sequences is sufficient for normal expression of the white gene (Levis et al., 1985; Pirrotta et al., 1985).
Figure 1. *D. melanogaster* eye pigments biosynthetic pathway (Summers *et al.*, 1982; Tearle, 1991).

Xanthomatin is synthesized from tryptophan in four steps. Drosopterin is synthesized from guanosine triphosphate in 5 steps.
1.3.1.2.3 Expression regulation

The expression of the *white* gene is controlled by trans- and cis-regulatory elements. There are five genes that have been described so far affecting the *white* gene: *zeste* (Qian *et al.*, 1992; Chen & Pirrotta, 1993), *su(wP)* (Davison *et al.*, 1985), *cut* (Liu & Jack, 1992), *Inr-a* (Rabinow *et al.*, 1991), and *Wow* (Birchler *et al.*, 1994). *Inr-a* and *Wow* also affect the expression of the *brown* and *scarlet* genes (Rabinow *et al.*, 1991; Birchler *et al.*, 1994). Multiple upstream DNA elements respond to the regulatory elements described above and are responsible for the *white* gene tissue-specific expression (Fig. 2). The DNA segment starting from position -400 to -216 regulates the *white* gene specific expression in larval and adult malpighian tubules. The fragment from -1084 to -1184 determines testis specific expression of *white* gene. Eye specific expression is controlled by the region from -1185 to -1455. The segment from -1084 to -1311 is responsible for the specific expression of the *white* gene in imaginal island of gut. The *zeste* protein binds the *white* locus in two areas, from -1185 to -1455 and from -110 to -140. An internal promoter sits in position -17 to +173. The distant enhancer of the *white* gene communicates with the internal promoter through the region from -113 to -17 (Levis *et al.*, 1985; Pirrotta *et al.*, 1985). The upstream region from -960 to -600 responds to the regulatory effect of gene *su(wP)* (Qian *et al.*, 1992).
Figure 2. Upstream DNA regulatory elements of the *D. melanogaster* white gene.
Position zero in the diagram is the transcription start site of the white gene of *D. melanogaster*. The function and the correlated DNA segments are listed below.
The function and the correlated DNA segment:

1 (-1455, -1185): \textit{zeste} binding site and for eye specific expression
2 (-1311, -1084): for imaginal island of gut specific expression
3 (-1184, -1084): for testis specific expression
4 (-960, -600): interacting with the \textit{su(w^{P})} protein
5 (-400, -216): for larval and adult malpighian tubule specific expression
6 (-140, -110): \textit{zeste} binding site
7 (-113, -17): distant enhancer interacting site
8 (-17, +173): internal promoter
1.3.1.3 *Function of the white protein.*

The *white* gene product is proposed to be involved in the production and distribution of pigments found in the compound eyes and ocelli of adult flies as well as the pigments in adult testis sheaths and larval malpighian tubules. It is a membrane-associated ATP-binding transport protein function in both the ommochrome and pteridine pathways (Sullivan and Sullivan, 1975; Mount, 1987; Dreesen *et al.*, 1988; Tearle *et al.*, 1989; Tearle, 1991).

A model of the action sites of the *white* and *scarlet* gene products has been proposed (Tearle, 1991). Together they transport tryptophan, the ommochromes precursor, across the plasma membrane into the cell. Then they transport kynurenine across the outer mitochondrial membrane to interact with kynurenine hydroxylase to form 3-hydroxykynurenine, which is helped to cross the pigment granule membrane into the pigment granule where xanthommatin is formed (Fig. 3).

There is extensive amino acid similarity of the putative translated protein of the *white* gene to the *brown* gene of the pteridine pathway (Dreesen *et al.*, 1988), and the *scarlet* gene of the ommochrome pathway (Tearle, 1989). There is high homology between the *white* gene isolated from blow sheep fly, *Lucilia cuprina* (Weidemann) and that of *D. melanogaster* (Bedo & Howells, 1987; Elizur *et al.*, 1990).

The *white* gene has been widely used as a marker in the P-element gene transformation system (Klemenz *et al.*, 1987; Serano *et al.*, 1994).
Figure 3. Proposed action sites of the white protein (we) (modified from Tearle, 1991).

The white protein (w) together with the scarlet (st) protein transport tryptophan (T), the xanthommatin substrate, across the plasma membrane, kynurenine (K) across the mitochondrial outer membrane and the 3-hydroxykynurenine (3HK) into the pigment granules. V is vermilion gene product and cn is cinnabar gene product.
1.3.2 *White eye* mutants of the oriental fruit fly and other insects

McCombs and Saul (1992a) isolated a *white eye* (*we*) mutant of the oriental fruit fly. It appeared as a spontaneous mutation in five males from a laboratory colony. The eye color is bright *white*. This allele is an autosomal recessive gene and the *we/we* homozygote is fully penetrant and completely expressed.

Many mutants in invertebrates are caused by the insertion of transposable elements (Finnegan, 1991). If it is the case in the oriental fruit fly, it is possible that we can isolate a transposable element which can be modified into a transformation vector. By using the wild-type *white eye* gene as a transformation reporter gene, the *white eye* mutant can serve as a testing host for the transformation system.

Mutants with a white-eyed phenotype have also been discovered in other insects, such as in the German cockroach, *Blattella germanica* (Linnaeus) (Ross & Cochran, 1974); the house fly, *Musca domestica* Linnaeus, (Tsukamoto *et al.*, 1961); the mosquito *Culex pipiens* Linnaeus, (Seal, 1966; Tadano, 1969) and *Culex molestus* (Gilchrist & Haldane, 1947); the sheep blowfly, *Lucilia cuprina* (Whitten *et al.*, 1974), the screwworm fly, *Cochliomyia hominivorax* (Coquerel), (Taylor & Martinez, 1986); the flour beetle, *Tribolium castaneum*, (Lemon & Blackman, 1967); the honeybee, *Apis mellifera*, (Tucker, 1986); the Mediterranean fruit fly (Saul & McCombs, 1992); and the melon fly (McCombs *et al.*, 1995). The mutants are lack of eye pigments. The *white* genes of screwworm, sheep blowfly, honey bee, flour beetle, oriental fruit fly, medfly, and melon fly are inherited as autosomal recessive genes, whereas the ones of *D. melanogaster*, *C. pipiens*, and *C. molestus* are sex-linked recessive genes.
1.3.3 Actin genes of the oriental fruit fly and the Mediterranean fruit fly

Muscle specific actin genes have been cloned in the oriental fruit fly (He & Haymer, 1994, 1991) and the Mediterranean fruit fly (He & Haymer, 1992; Haymer et al., 1990). Compared to D. melanogaster, the muscle specific actin genes of tephritid species are highly conserved in the coding regions and gene specific in the flanking regions. The deduced protein sequences from the coding regions have higher similarity than the nucleotide sequence of the actin genes. This inconsistency can be explained by the codon usage preference of T in Tephritidea vs. C in D. melanogaster. Information of the actin gene organization, temporal and spatial expression, intron position, and DNA sequence provide insight into the evolutionary relationship and genome structure difference between the Tephritidae and D. melanogaster.
2. Materials and Methods

2.1 Materials

2.1.1 Oriental fruit fly strain

The wild type oriental fruit fly stock was originally obtained from the USDA/ARS Tropical Fruit and Vegetable Laboratory Honolulu, Hawaii where it was maintained for several years as a mass rearing colony. It is maintained in the quarantine room in the Department of Entomology, University of Hawaii.

2.1.2 Escherichia coli strains

DH5α for the preparation of competent cell is from GIBCOBRL (Grand Island, NY). LE392 is from Promega (Madison, Wisconsin) as the host of λ bacteriophage.

2.1.3 Plasmids

Plasmids pBluescript II SK+ from Stratagen (La Jolla, CA) and pSp72 from Promega are used for the DNA subcloning.

2.1.4 Reagents and systems

2.1.4.1 Enzymes

All the restriction endonucleases and DNA modifying enzymes were purchased from Promega.

2.1.4.2 System products

The λ-GEM®-11 genomic cloning vector kit and the Erase-a-base® system are from Promega. The USBioclean™ MP kit is from United States Biochemical (Cleveland, OH). The Prep-A-Gene® plasmid purification kit is from Bio-rad (Hercules, CA). The
Genius™ nonradioactive DNA-labeling and detection kit is from the Boehringer Mannheim Biochemicals (Indianapolis, Indiana). The QIAprep-spin kit for plasmid mini preparations, QIAquick PCR purification kit, QIAGEN gel extraction kit and QIAGEN Lambda and M 13 kit are from QIAGEN (Chatsworth, CA). The PRISM™ ready reaction dyedideoxy™ terminator cycle sequencing kit is from Applied Biosystems (Foster city, CA).

2.2 Methods

2.2.1 Fruit fly rearing

The rearing method used was similar to that described by McCombs (1992c) for the oriental fruit fly. Female oriental fruit flies of at least two weeks old were egged with guava nectar as stimulant. The eggs are spread on the larval media. Pupa are sifted after about 18 d and placed in adult cages supplied with adult food and water. Adults emerge in about 20 d. The egg to egg generation time is about 34 d.

2.2.2 Genomic DNA isolation

Genomic DNA was isolated from the pupae of the oriental fruit fly by the method of Doyle & Doyle (Doyle & Doyle, 1991). Briefly, the fresh pupae (≈1 g) were frozen in dry ice and ground to a fine powder in a chilled mortar and pestle. The powder was resuspended in 7.5 ml of 60°C preheated CTAB isolation buffer [2% (w/v) CTAB(Sigma), 1.4 M NaCl, 0.2% (v/v) 2-mercaptoethanol, 20 mM EDTA, 100 mM Tris-HCl, (pH 8.0)] and incubated for 30 min at 60°C. The mixture was extracted with chloroform-isoamyl alcohol (24:1, v/v) and centrifuged (1600 g) at room temperature for 10 min. The aqueous phase was removed and mixed with 2/3 volumes of -4°C 100% isopropanol. The
genomic DNA was precipitated by centrifuging at 500 g for 5 min and washed in wash buffer [76% ethanol (v/v), 10 mM ammonium acetate] for at least 20 min. DNA was pelleted in 1600 g for 10 min and resuspended in 1 ml TE [10 mM Tris-HCl (pH 7.4), 1 mM EDTA] buffer. RNase A was added to the DNA solution to a final concentration of 100 µg/ml and incubated for 30 min at 37°C. The sample was diluted with 2 volumes of TE buffer and chloroform-isoamyl alcohol extraction was repeated. DNA was precipitated with 7.5 M ammonium acetate (pH 7.7) to a final concentration of 2.5 M and 2.5 volumes of -4°C 95% ethanol. DNA was recovered by centrifuging at 10,000 g for 10 min in a refrigerated centrifuge. The sample was air dried and resuspended in an appropriate amount of TE buffer.

2.2.3 Genomic DNA library construction

To construct the library, high molecular weight genomic DNA was partially digested with Sau3A. The digested DNA was partially filled in with dATP and dGTP. This step prevented the ligation between different inserted fragments. The partially filled-in vector, Lambda GEM®-11, and the partially filled-in inserts DNA were ligated, packaged, and used to infect E. coli strain LE 392, as described by the manufacturer (Promega). The library was amplified once, stored and screened immediately (Sambrook et al., 1989).

2.2.4 Library screening

2.2.4.1 Probe

The DNA probe, Dacd-we, used for the genomic DNA library screening was generously provided by Dr. Frank Collins and Dr. Nora Besansky (Centers for Disease
Control, Department of Health & Human Services, Atlanta). The probe is a PCR product of the white eye gene of the oriental fruit fly and is about 700 bp long, extending from near the 3' end of exon 4 to 5' end of exon 6 based on the white gene structure of D. melanogaster.

Large scale preparation of the Dacd-we probe containing plasmid DNA was carried out by the standard alkaline lysis method (Sambrook et al. 1989). The plasmid was digested with EcoR I and Xba I. The release insert was recovered from the agarose gel using the USBioclean™ MP kit. It was labeled by the random priming method using the Genius™ nonradioactive DIG DNA labeling and detection kit.

### 2.2.4.2 Identification of the white eye clone

Genomic library screening was conducted by the plaque lift method (Sambrook et al., 1989). NitroPlus nitrocellulose membranes (MSI, Inc., Westborough, Mass.) were used for the library screening. Fixation of the DNA to the membrane was carried out by baking for one h at 80°C. The Genius™ nonradioactive DIG DNA labeling and detection kit was used. Prehybridization, hybridization, wash, immunodetection and color reactions followed the manual provided by the manufacturer.

The basic idea of the method was to transfer the recombinant phages from the agarose plate onto the nitrocellulose membrane. After denaturation and neutralization, the released phage DNA was fixed onto the membrane by baking. Then the membrane was hybridized to the probe. All the hybridization signal positive clones were recovered from the plate. The positive clones were purified by repeating the plaque lifting screening two
to three times more. First time screening was carried out at 60°C, and the other purified screenings were carried out at 63°C.

2.2.5 Recombinant λ-phage analysis

2.2.5.1 Recombinant λ-phage DNA preparation

The positive recombinant phage λ-DW1A1 was prepared by the liquid method of high multiplicity infection (Sambrook et al., 1989). For large scale preparation of λ-DW1A1 DNA, λ-particles were first purified by the glycerol step gradient method and DNA was isolated by regular extraction methods (Sambrook et al. 1989). Small scale λ-DW1A1 DNA extraction was done by QIAGEN Lambda and M 13 kits.

2.2.5.2 Southern hybridization

The Southern hybridization technique used in this research was similar to the one described by Sambrook et al. (1989). The purpose was to localize the white eye gene fragment within a plasmid or phage DNA. The target DNA, plasmids or λ-phage DNA, was digested with one or more restriction enzymes, and the resulting fragments were separated by size through agarose gel electrophoresis. The DNA was then denatured, neutralized and transferred to the nitrocellulose membrane using the VacuGene XL vacuum blotting system (Pharmacia, Piscataway, NJ). The relative positions of the DNA fragments were preserved in the membrane that was hybridized to the Dacz-we probe later. Genius™ nonradioactive labeling and detection kit from Boehringer Mannheim was used for probe labeling, hybridization and immunodetection. The band that was complementary to the probe gave positive signal.
2.2.5.3 Subcloning of interested fragments into plasmid

The insert of \( \lambda{-}DW1A1 \) was subcloned into pBluescript II SK+ or pSp72 plasmids so as to sequence both strands of the insert by generating the nest-deleted subclones. The relative positions of different recombinant plasmids are depicted in figure 4.

2.2.5.3.1 Construction of pBlue-DW-Cla plasmid.

Recombinant bacteriophage \( \lambda{-}DW1A1 \) DNA was digested with Sac I DNA endonuclease (Sambrook et al. 1989). The insert was eluted from the agarose gel by QIAGEN gel extraction kit and digested with Cla I restriction endonuclease. The digestion mixture was electrophoresed and the fragment hybridized to the Dacd-we probe was retrieved from the agarose gel using the QIAGEN gel extraction kit. The retrieved fragment was treated with alkaline phosphatase from calf intestine (Sambrook et al., 1989), purified with QIAGEN PCR purification spin kit, and ligated to the pBlue-script II SK+ phagemid vector (Stratagene, La Jolla, CA), which was precut with Cla I and Sac I (Sambrook et al. 1989). The subclone is called pBlue-DW-Cla.

The ligation of the plasmid to the insert, and the transformation of competent cell DH5\( \alpha \) (GIBCOBRL, Grand Island, NY; Ausubel et al., 1989) by the ligation mixture followed the standard methods (Sambrook et al. 1989). Plasmid DNA was extracted and purified by QIAprep-spin Plasmid kit or the standard alkaline lysis method (Sambrook et al. 1989).

2.2.5.3.2 Construction of pSp-DW-Cla plasmid.

Plasmid pBlue-DW-Cla was digested with Sac I and Cla I endonuclease. The insert was isolated by gel electrophoresis and purified from the agarose gel by the
USBioclean™ MP kit. The free insert was then ligated to the Sac I and Cla I precut pSp72 plasmid (Promega) to form the recombinant plasmid pSp-DW-Cla. Similar technical steps were repeated as described above for the construction of pBlue-DW-Cla.

### 2.2.5.3.3 Construction of pBlue-DW-EcoR3 plasmid

Phage λ-DW1A1 and plasmid pBluescript II SK+ were digested separately with EcoRI and purified by the QIAquick PCR purification kit. A fragment about 3 kb long that hybridized to the probe Dacd-we was subcloned into precut pBluescript II SK+ following similar steps as described for pBlue-DW-Cla.

### 2.2.5.3.4 Construction of pBlue-DW-SX4 plasmid

Phage λ-DW1A1 and plasmid pBluescript II SK+ were double digested separately with Sac I and Xho I and purified by the QIAquick PCR purification kit. A fragment about 4 kb long that hybridized to the DNA probe Dacd-we was subcloned to pBluescript II SK+. All the techniques used here were similar to the construction of pBlue-DW-Cla.

### 2.2.5.3.5 Construction of pBlue-DW-EC1 and pBlue-DW-CC5 plasmid

Plasmid pBlue-DW-EcoR3 was digested with Cla I and EcoR I. Half of the digestion mixture was ligated to pBluescript II SK+ precut with Cla I resulting in plasmid pBlue-DW-CC5, and the other half ligated to pBluescript II SK+ precut with EcoR I and Cla I resulting in plasmids pBlue-DW-EC1 and pBlue-DW-EC2 (Fig. 4).
Figure 4. The relationship of the pBlue-SX4, pBlue-EcoR3, pBlue-CC5, pBlue-EC1, pBlue-DW-Cla, and pSp-DW-Cla plasmids. The restriction cloning sites of each clone are indicated on the ends. Dacd-we is the probe used in genomic DNA library screening and Southern hybridization.
2.2.6 Sequencing

2.2.6.1 Sequencing strategy

Overlapping subclones for DNA sequencing were generated by nested deletions using the Exo III/S 1 nuclease deletion kit (Promega) following their protocol except some modifications as described below. Recombinant pBlue-DW-Cla plasmid was double digested by Sal I and Kpn I and pSp-DW-Cla by Xba I and Sph I. Digested plasmids were purified by QIAGEN PCR purification spin kit and dissolved in TE buffer. Plasmid DNA was cut subsequently by Exo III and S 1 nuclease. DNA was purified again. Klenow DNA polymerase was added to flush the ends and then denatured by heating at 70°C for 10 min. The samples were ligated to circularize the deletion-containing vectors. Different batches of competent cells (DH5α) were transformed by different time point ligation mixtures. A number of subclones from each time point were screened to select for appropriate intervals between deletions. Plasmid DNA of suitable size was prepared by Prep-A-Gene® DNA purification kit or QIA-prep-spin mini preparation kit and dissolved in 10mM Tris buffer.

2.2.6.2 Sequencing

Automatic DNA sequencing was carried out by the DNA sequencer (type 373A) from Applied Biosystems (Foster City, CA) using the PRISM™ Ready Reaction DyeDeoxy™ Terminator Cycle Sequencing Kit in the biotechnology-molecular biology instrumentation and training facility, University of Hawaii. The method is similar to the one described by Sanger (Sanger et al., 1977). In a 0.5 ml eppendorf tube 1 µg purified plasmid was mixed with 9.5 µl Dyedideoxy terminator premix (ddNTP, Taq polymerase,
and sequencing buffer) and 3.2 pmol primer (synthesized or commercially available ones, SP6 promoter primer, pUC/M13 forward primer or pUC/M13 reverse primer). The tube was placed into a thermal cycler. The double stranded plasmid DNA was denatured at 96°C for 30 s and the primer annealed to the template when temperature ramped to 50°C for 15 s. At 60°C for 4 min the Taq polymerase synthesized a DNA strand complimentary to the DNA template just after the primer. The temperature cycle was repeated 25 times. The four types of dideoxynucleotide triphosphates linked to four different kinds of dye were randomly incorporated into the growing chain and terminated the elongation of the newly synthesized DNA chain. All the different length oligonucleotides were separated by polyacrylamide sequencing gel. The DNA sequence was read directly from the gel by the automated sequencer. Several primers synthesized by Dnagency (Aston, PA) were designed to sequence the gaps among the nonoverlapping deletion subclones.

2.2.7 Sequence data analysis

All the DNA sequences were edited, compiled and analyzed by the Genetic Group Computer program developed by the University of Wisconsin. The Blast, Geneid, Grail, and Netgene programs were also used in the construction of the gene model.
3. Results and Discussion

3.1 Genomic DNA library construction

The genomic library of the oriental fruit fly was constructed in the lambda-GEM®-11 vector. The size of the average insert of the library is $1.5 \times 10^4$ bp. It is supposed that the genome size of the oriental fruit fly is of the same order of magnitude as D. melanogaster, $1.4 \times 10^8$ bp. Therefore the size of the library that will represent the white eye gene with 99 percent probability is between $3 \times 10^4$ to $2.7 \times 10^5$ recombinant phages according to the following calculations. If the genome size is $1 \times 10^8$ bp, the library size should be: $\ln(1-0.99) / \ln(1-1.5 \times 10^4 / 1.0 \times 10^8) = 3 \times 10^4$ recombinant phages. If the genome size is $9 \times 10^8$ bp, the library size should be: $\ln(1-0.99) / \ln(1-1.5 \times 10^4 / 9.0 \times 10^8) = 2.7 \times 10^5$ recombinant phages. The genomic DNA library of the B. dorsalis is $3.0 \times 10^5$ plaques per microgram genomic DNA and represents the white eye gene with over 99% probability.

3.2 Genomic library screening

$10^5$ recombinant λ-phages from the oriental fruit fly genomic library were screened for the white eye gene-containing clone. The nonradio-labelled DNA probe, Dacd-we, and the plaque-lifting method were used. The screening was carried out under stringent conditions and one positive clone, λ-DW1A1, was found. The insert of the clone was digested with different restriction endonucleases and subcloned into both the pBlue-script® II SK+ and pSp-72 plasmids. The resulting recombinant plasmids are pBlue-DW-Cla, pSp-DW-Cla, pBlue-SX4, pBlue-EcoR3, pBlue-CC5, and pBlue-EC1 (Fig. 4). Results of
Southern hybridization indicated that the inserts of plasmids pBlue-DW-Cla, pSp-DW-Cla, pBlue-SX4, and pBlue-EcoR3 were homologous to the probe Dacd-we.

### 3.3 Sequencing

Each sequencing reaction can only obtain about 400 base DNA sequence information. Overlapping subclones were constructed by nested deletions. Both DNA strands in 28 subclones were sequenced. The sequenced segments of each subclone are depicted in figure 5. Total sequence information is 3859 bp long (Fig. 6).

### 3.4 Sequence Analysis

#### 3.4.1 Gene structure model

There are introns and exons in genomic DNA sequence. An intron is a segment of DNA that is transcribed, but removed within the transcript by splicing together the sequences on either side of it. An exon is the DNA segment that is transcribed and translated. The gene model is built on two analysis results. One is the results of gene model prediction computer programs Grail, Netgene, Geneid, and Blast. Another is the amino acid sequence homology comparison of the translated white eye protein of the oriental fruit fly to those of *D. melanogaster* and *L. cuprina*.

The software results provided information on where the potential intron-exon splicing sites and the polyA addition site were located. The white eye gene splicing sites were aligned to those of the actin genes (Fig. 7). The consensus sequence of the intron-exon junction derived from the alignment of the actin gene and the white eye gene of *B. dorsalis* differs from that of the alignment of the white eye genes of *B. dorsalis* and *D.*
Figure 5. The relative positions of the sequenced segments of different subclones
Figure 6 The DNA and the translated amino acid sequence of the white eye gene of *Bactrocera dorsalis.*

The intron-exon junction: "I". The ATP-binding site: "-". The consensus ATP-binding sequence "=". The poly-A addition site: □.
Figure 6 (Continued) The DNA and the translated amino acid sequence of the *white eye* gene of *Bactrocera dorsalis*.
Figure 6 (Continued) The DNA and the translated amino acid sequence of the white eye gene of *Bactrocera dorsalis.*
Figure 6 (Continued) The DNA and the translated amino acid sequence of the white eye gene of *Bactrocera dorsalis*.
Figure 6 (Continued) The DNA and the translated amino acid sequence of the white eye gene of Bactrocera dorsalis.

38
Figure 6 (Continued) The DNA and the translated amino acid sequence of the white eye gene of *Bactrocera dorsalis*.
<table>
<thead>
<tr>
<th></th>
<th>TACCCAG</th>
<th>GTTAGT</th>
<th>TTATAG</th>
<th>GTAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BdA5</td>
<td>TACCCAG</td>
<td>GTTAGT</td>
<td>TTATAG</td>
<td>GTAT</td>
</tr>
<tr>
<td>BdA1</td>
<td>TACCCAG</td>
<td>GTTAAT</td>
<td>TTACAG</td>
<td>GTAT</td>
</tr>
<tr>
<td>BdA2</td>
<td>GATAATG</td>
<td>GTGCGT</td>
<td>TTTTAG</td>
<td>GCTC</td>
</tr>
<tr>
<td>Con1.</td>
<td>NHNVMHD</td>
<td>GTDMDY</td>
<td>BWNYAG</td>
<td>NNNN</td>
</tr>
<tr>
<td>BdII</td>
<td>TACCCAG</td>
<td>AAACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BdI2</td>
<td>TTCCAG</td>
<td>AACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BdI3</td>
<td>TCTCATA</td>
<td>GTGAGT</td>
<td>CTCTAG</td>
<td>CATC</td>
</tr>
<tr>
<td>BdI4a</td>
<td>TGGAAAA</td>
<td>GTAAAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BdI4b</td>
<td>ATGGG</td>
<td>GTAAAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BdI5</td>
<td>CATAACT</td>
<td>GTAGCT</td>
<td>TTACAG</td>
<td>GTCG</td>
</tr>
<tr>
<td>Con2.</td>
<td>NHNVMHD</td>
<td>GTRMRY</td>
<td>NHNYAG</td>
<td>NDBN</td>
</tr>
<tr>
<td>DmI1</td>
<td>GAACAA</td>
<td>GTGAGT</td>
<td>TTGTAG</td>
<td>TTGA</td>
</tr>
<tr>
<td>DmI2</td>
<td>AAGAACC</td>
<td>GTGAGT</td>
<td>ACTCAG</td>
<td>TTGG</td>
</tr>
<tr>
<td>DmI3</td>
<td>TTTTCTA</td>
<td>GTGAGT</td>
<td>ATCCAG</td>
<td>CTGT</td>
</tr>
<tr>
<td>DmI4</td>
<td>GACAA</td>
<td>GTGAGT</td>
<td>TTGAG</td>
<td>ATGG</td>
</tr>
<tr>
<td>DmI5</td>
<td>GATAAA</td>
<td>GTGAGT</td>
<td>ATTTAG</td>
<td>GTGT</td>
</tr>
</tbody>
</table>

**Intron**

Figure 7 Alignment of the intron-exon junction sequences

The junction sequences from three *B. dorsalis* actin genes (BdA5, BdA1, and BdA2) are named as the genes respectively. The junction sequences of intron 1, 2, 3, 4a, 4b, and 5 of the *B. dorsalis white eye* gene are abbreviated as BdI1, BdI2, BdI3, BdI4a, BdI4b, and BdI5. DmI1, DmI2, DmI3, DmI4, and DmI5 are the junction sequences of intron 1, 2, 3, 4, and 5 of the *D. melanogaster white* gene. Con1 is the consensus sequence derived from the comparison between the actin genes and the white eye gene of *B. dorsalis*. Con2 is the consensus sequence derived from the comparison between the white eye genes of *B. dorsalis* and *D. melanogaster*. M: A or C, R: A or G, W: A or T, Y: C or T, V: A or C or G, H: A or C or T, D: A or G or T, B: C or G or T, N: A or T or G or C.
melanogaster. The nucleotides at a specific position of the intron-exon junction vary in different introns except that GT is present at the 5' end and AG at the 3' end of the intron.

Further analysis to differentiate the most likely splicing sites from the others was done by comparing the amino acid sequence similarity of the translated white eye protein of the oriental fruit fly to those of D. melanogaster and L. cuprina. There are six exons, five introns and a potential polyA addition site in the white eye gene model of the oriental fruit fly (Fig. 8). Exon four of the white gene of D. melanogaster is split into two exons in B. dorsalis (Fig. 8). The size of the corresponding exons of the two white eye genes is similar and the phases are the same (Table 1).

The promoters known so far can be categorized into three types. The first type is the well-studied TATA box containing promoter. The key elements are the TATA box only or a TATA box and a transcription initiation site. Transcription starts about 30 bp downstream of the TATA box. A discrete start site has not been identified nor has its role in the accurate initiation of transcription been determined (Lewin, 1990; Buratowski et al., 1988, Dynan, 1989). The second class contains no obvious TATA box but is GC rich. This type of promoter is found primarily in the housekeeping genes (Sehgal et al., 1988; Dynan, 1989). The third class contains neither a TATA box nor a GC rich region (Smale & Baltimore, 1989; Arkhipova & Iiyin, 1991). A conservative DNA sequence, ACAG/C or CGTG, about 40 bp downstream from the transcription start site is important for the correct initiation. The D. melanogaster white and zeste genes belong to this class.

There are several short fragments of ACAG/C and CGTG in the oriental fruit fly white eye gene DNA sequence presented here, but they do not look like the internal
Figure 8. Comparison of the white eye gene model of *D. melanogaster* and *B. dorsalis*.

* stands for the poly A addition site
$D.\ melanogaster$

$B.\ dorsalis$

```
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>4a</th>
<th>4b</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>exon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intron</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 bp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
Table 1. Exon size and phase comparison of the *white eye* genes of *B. dorsalis* and *D. melanogaster*.

<table>
<thead>
<tr>
<th>Gene exon</th>
<th><strong>B. dorsalis</strong></th>
<th><strong>D. melanogaster</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>size(^a)</td>
<td>phase</td>
</tr>
<tr>
<td>1</td>
<td>82</td>
<td>0,0</td>
</tr>
<tr>
<td>2</td>
<td>218</td>
<td>0,1</td>
</tr>
<tr>
<td>3</td>
<td>106</td>
<td>1,2</td>
</tr>
<tr>
<td>4</td>
<td>44</td>
<td>2,0</td>
</tr>
<tr>
<td>5</td>
<td>204</td>
<td>0,0</td>
</tr>
<tr>
<td>6</td>
<td>205</td>
<td>0,0</td>
</tr>
</tbody>
</table>

\(^a\) The size is the number of amino acids.
promoter for two reasons. First, the short peptides starting with methionine behind the conservative DNA sequence of the possible internal promoter do not share the protein secondary structure properties with the exon 1 of the white gene of D. melanogaster. Second, intron 1 of the D. melanogaster white gene is about 3.1 kb long and the white eye sequence of B. dorsalis from exon 2 to the sequencing start site is 1.1 kb long. It is possible that the sequence presented here is not long enough to include exon 1, the transcription starting site, and the internal promoter.

3.4.2 Coding region sequence comparison

The deduced partial cDNA sequence of the oriental fruit fly white eye gene was translated into protein, and the amino acid sequence was aligned with the white proteins of D. melanogaster and L. cuprina (Feng & Doolittle, 1987) (Fig. 9). These white eye genes are highly conserved at the protein level with a 93% similarity (Table 2).

The coding region nucleotide sequence alignment of the white eye genes of the three species shows a higher degree of diversity (Feng & Doolittle, 1987) (Fig. 10). The white eye gene of B. dorsalis has 70.8% similarity to that of D. melanogaster and 72.6% similarity to that of L. cuprina (Table 2). This inconsistency of similarity at the DNA and protein levels suggests a high rate of silent substitutions and selective constraints at the protein level. There are 71.4% silent mutations and 28.6% replacement mutations (method by Perler et al., 1980) between the white eye genes coding region sequence of B. dorsalis and D. melanogaster. This reflects stronger constraints in amino acid replacement mutation than the silent substitution.
Figure 9. Protein sequence alignment of the white eye genes from *B. dorsalis* (B.d), *D. melanogaster* (D.m), and *L. cuprina* (L.c).

Only the amino acids that are different from the consensus sequence are shown in small case. The "." symbol stands for no corresponding sequence data at a specific position.
<table>
<thead>
<tr>
<th>401</th>
<th>450</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bd.pep</td>
<td>--------s-</td>
</tr>
<tr>
<td>451</td>
<td>500</td>
</tr>
<tr>
<td>Bd.pep</td>
<td>--------</td>
</tr>
<tr>
<td>501</td>
<td>550</td>
</tr>
<tr>
<td>Bd.pep</td>
<td>--------</td>
</tr>
<tr>
<td>551</td>
<td>600</td>
</tr>
<tr>
<td>Bd.pep</td>
<td>--------</td>
</tr>
<tr>
<td>601</td>
<td>650</td>
</tr>
<tr>
<td>Bd.pep</td>
<td>--------</td>
</tr>
<tr>
<td>651</td>
<td>689</td>
</tr>
<tr>
<td>Bd.pep</td>
<td>--------</td>
</tr>
</tbody>
</table>

Figure 9. (Continued) Protein sequence alignment of the *white eye* genes from *B. dorsalis* (B.d), *D. melanogaster* (D.m), and *L. cuprina* (L.c).
Table 2. Pairwise comparison of the *white eye* genes at the DNA level and the protein level from *B. dorsalis*, *D. melanogaster* and *L. cuprina*.

<table>
<thead>
<tr>
<th></th>
<th><em>D. melanogaster</em>&lt;sup&gt;a&lt;/sup&gt;</th>
<th><em>L. cuprina</em></th>
<th><em>B. dorsalis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. melanogaster</em>&lt;sup&gt;b&lt;/sup&gt;</td>
<td>---</td>
<td>91.6</td>
<td>93.4</td>
</tr>
<tr>
<td><em>L. cuprina</em></td>
<td>71.9</td>
<td>---</td>
<td>93.7</td>
</tr>
<tr>
<td><em>B. dorsalis</em></td>
<td>70.8</td>
<td>72.6</td>
<td>---</td>
</tr>
</tbody>
</table>

<sup>a</sup> Upper right portion: amino acid sequence
<sup>b</sup> Bottom left portion: DNA sequence
<sup>c</sup> Similarity in percentage.
**Figure 10.** Coding region DNA sequence alignment of the white eye genes from *B. dorsalis* (B.d), *D. melanogaster* (D.m), and *L. cuprina* (L.c).

Only the nucleotides that are different from the consensus sequence (consen) are shown in small case. The ‘.’ symbol stands for no corresponding sequence data at a specific position.
Figure 10. (Continued) Coding region DNA sequence alignment of the white eye genes from *B. dorsalis* (B.d), *D. melanogaster* (D.m), and *L. cuprina* (L.c).
Figure 10. (Continued) Coding region DNA sequence alignment of the white eye genes from *B. dorsalis* (B.d), *D. melanogaster* (D.m), and *L. cuprina* (L.c).
Figure 10. (Continued) Coding region DNA sequence alignment of the white eye genes from B. dorsalis (B.d), D. melanogaster (D.m), and L. cuprina (L.c).
Figure 10. (Continued) Coding region DNA sequence alignment of the white eye genes from B. dorsalis (B.d), D. melanogaster (D.m), and L. cuprina (L.c).
3.4.3 Codon usage

The GC content of the coding region sequence of white eye gene varies in different species, 46.2% in B. dorsalis, 55.5% in D. melanogaster, and 43% in L. cuprina. The difference is caused by the nonrandom use of different codons (Table 3). The GC content of the white eye gene in different species correlates well with the nucleotide preference pattern of the codon silent sites, preference for AT in the oriental fruit fly, GC in D. melanogaster, and AT in L. cuprina (Table 3). The GC content at the silent sites of the white eye genes is 41.1% in B. dorsalis and 71.9% in D. melanogaster.

The codon bias and GC content of the white eye gene data reported here agrees with the published actin gene data (Table 4). The actin and white eye genes of B. dorsalis have a lower GC content than those of D. melanogaster. The codon usage differences between species are determined by several factors, such as the amounts and availability of tRNA species, GC content of the genome or of the local area (Ikemura, 1985), the DNA polymerase system (Cox & Yanofsky, 1967), and the relative abundance of certain nucleotide synthetase (Honess et al., 1986). The consistent tendency of the codon usage pattern and the GC content data of the silent positions and coding region sequence of the white eye gene and the actin genes may indicate the genome structure differences between B. dorsalis and D. melanogaster.

Generally, there is high codon usage bias in the highly expressed genes caused by selection pressure, whereas genes expressed at a lower level have low codon usage bias determined by mutation (Sharp & Li, 1986). Genes with high proportions of optimal codons evolve more slowly (Ikemura, 1985). The codon bias data presented here is
Table 3. Codon usage of the white eye genes in *B. dorsalis*, *D. melanogaster*, and *L. cuprina*

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Codon</th>
<th>B. dorsalis</th>
<th>D. melanogaster</th>
<th>L. cuprina</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Number</td>
<td>Fraction</td>
<td>Number</td>
</tr>
<tr>
<td>Gly</td>
<td>GGG</td>
<td>2.00</td>
<td>0.04</td>
<td>5.00</td>
</tr>
<tr>
<td>Gly</td>
<td>GGA</td>
<td>7.00</td>
<td>0.15</td>
<td>12.00</td>
</tr>
<tr>
<td>Gly</td>
<td>GGT</td>
<td>21.00</td>
<td>0.45</td>
<td>9.00</td>
</tr>
<tr>
<td>Gly</td>
<td>GGC</td>
<td>17.00</td>
<td>0.36</td>
<td>26.00</td>
</tr>
<tr>
<td>Gly</td>
<td>GAG</td>
<td>15.00</td>
<td>0.52</td>
<td>25.00</td>
</tr>
<tr>
<td>Gly</td>
<td>GAA</td>
<td>14.00</td>
<td>0.48</td>
<td>7.00</td>
</tr>
<tr>
<td>Asp</td>
<td>GAT</td>
<td>14.00</td>
<td>0.67</td>
<td>8.00</td>
</tr>
<tr>
<td>Asp</td>
<td>GAC</td>
<td>7.00</td>
<td>0.33</td>
<td>13.00</td>
</tr>
<tr>
<td>Val</td>
<td>GTG</td>
<td>15.00</td>
<td>0.30</td>
<td>28.00</td>
</tr>
<tr>
<td>Val</td>
<td>GTA</td>
<td>11.00</td>
<td>0.22</td>
<td>6.00</td>
</tr>
<tr>
<td>Val</td>
<td>GTT</td>
<td>15.00</td>
<td>0.30</td>
<td>5.00</td>
</tr>
<tr>
<td>Val</td>
<td>GTC</td>
<td>9.00</td>
<td>0.18</td>
<td>12.00</td>
</tr>
<tr>
<td>Ala</td>
<td>GCG</td>
<td>13.00</td>
<td>0.28</td>
<td>9.00</td>
</tr>
<tr>
<td>Ala</td>
<td>GCA</td>
<td>8.00</td>
<td>0.17</td>
<td>3.00</td>
</tr>
<tr>
<td>Ala</td>
<td>GCT</td>
<td>12.00</td>
<td>0.26</td>
<td>3.00</td>
</tr>
<tr>
<td>Ala</td>
<td>GCC</td>
<td>14.00</td>
<td>0.30</td>
<td>40.00</td>
</tr>
</tbody>
</table>
Table 3. (Continued) Codon usage of the white eye genes in *B. dorsalis*, *D. melanogaster*, and *L. cuprina*

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>AGG</td>
<td>1.00</td>
<td>0.03</td>
<td>7.00</td>
<td>0.19</td>
<td>0.00</td>
</tr>
<tr>
<td>Arg</td>
<td>AGA</td>
<td>3.00</td>
<td>0.09</td>
<td>1.00</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Ser</td>
<td>AGT</td>
<td>2.00</td>
<td>0.04</td>
<td>4.00</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td>Ser</td>
<td>AGC</td>
<td>4.00</td>
<td>0.08</td>
<td>11.00</td>
<td>0.21</td>
<td>1.00</td>
</tr>
<tr>
<td>Lys</td>
<td>AAG</td>
<td>6.00</td>
<td>0.20</td>
<td>15.00</td>
<td>0.60</td>
<td>0.00</td>
</tr>
<tr>
<td>Lys</td>
<td>AAA</td>
<td>24.00</td>
<td>0.80</td>
<td>10.00</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td>Asn</td>
<td>AAT</td>
<td>19.00</td>
<td>0.63</td>
<td>12.00</td>
<td>0.41</td>
<td>2.00</td>
</tr>
<tr>
<td>Asn</td>
<td>AAC</td>
<td>11.00</td>
<td>0.37</td>
<td>17.00</td>
<td>0.59</td>
<td>0.00</td>
</tr>
<tr>
<td>Met</td>
<td>ATG</td>
<td>14.00</td>
<td>1.00</td>
<td>16.00</td>
<td>1.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Ile</td>
<td>ATA</td>
<td>13.00</td>
<td>0.35</td>
<td>5.00</td>
<td>0.14</td>
<td>2.00</td>
</tr>
<tr>
<td>Ile</td>
<td>ATT</td>
<td>20.00</td>
<td>0.54</td>
<td>12.00</td>
<td>0.33</td>
<td>3.00</td>
</tr>
<tr>
<td>Ile</td>
<td>ATC</td>
<td>4.00</td>
<td>0.11</td>
<td>19.00</td>
<td>0.53</td>
<td>1.00</td>
</tr>
<tr>
<td>Thr</td>
<td>ACG</td>
<td>6.00</td>
<td>0.16</td>
<td>12.00</td>
<td>0.33</td>
<td>0.00</td>
</tr>
<tr>
<td>Thr</td>
<td>ACA</td>
<td>14.00</td>
<td>0.37</td>
<td>5.00</td>
<td>0.14</td>
<td>0.00</td>
</tr>
<tr>
<td>Thr</td>
<td>ACT</td>
<td>6.00</td>
<td>0.16</td>
<td>2.00</td>
<td>0.06</td>
<td>4.00</td>
</tr>
<tr>
<td>Thr</td>
<td>ACC</td>
<td>12.00</td>
<td>0.32</td>
<td>17.00</td>
<td>0.47</td>
<td>4.00</td>
</tr>
<tr>
<td>Trp</td>
<td>TGG</td>
<td>8.00</td>
<td>1.00</td>
<td>8.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>End</td>
<td>TGA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cys</td>
<td>TGT</td>
<td>6.00</td>
<td>0.55</td>
<td>3.00</td>
<td>0.23</td>
<td>2.00</td>
</tr>
<tr>
<td>Cys</td>
<td>TGC</td>
<td>5.00</td>
<td>0.45</td>
<td>10.00</td>
<td>0.77</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 3. (Continued) Codon usage of the white eye genes in *B. dorsalis*, *D. melanogaster*, and *L. cuprina*

<table>
<thead>
<tr>
<th></th>
<th>End</th>
<th>TAG</th>
<th>0.00</th>
<th>0.00</th>
<th>1.00</th>
<th>1.00</th>
<th>0.00</th>
<th>0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>End</td>
<td>TAA</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Tyr</td>
<td>TAT</td>
<td>10.00</td>
<td>0.56</td>
<td>8.00</td>
<td>0.42</td>
<td>4.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Tyr</td>
<td>TAC</td>
<td>8.00</td>
<td>0.44</td>
<td>11.00</td>
<td>0.58</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>TTG</td>
<td>20.00</td>
<td>0.26</td>
<td>12.00</td>
<td>0.14</td>
<td>3.00</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>TTA</td>
<td>14.00</td>
<td>0.18</td>
<td>1.00</td>
<td>0.01</td>
<td>7.00</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Phe</td>
<td>TTT</td>
<td>25.00</td>
<td>0.61</td>
<td>16.00</td>
<td>0.43</td>
<td>5.00</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Phe</td>
<td>TTC</td>
<td>16.00</td>
<td>0.39</td>
<td>21.00</td>
<td>0.57</td>
<td>4.00</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Ser</td>
<td>TCG</td>
<td>18.00</td>
<td>0.35</td>
<td>14.00</td>
<td>0.26</td>
<td>1.00</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Ser</td>
<td>TCA</td>
<td>12.00</td>
<td>0.23</td>
<td>5.00</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Ser</td>
<td>TCT</td>
<td>7.00</td>
<td>0.13</td>
<td>3.00</td>
<td>0.06</td>
<td>2.00</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Ser</td>
<td>TCC</td>
<td>9.00</td>
<td>0.17</td>
<td>16.00</td>
<td>0.30</td>
<td>3.00</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Arg</td>
<td>CGG</td>
<td>2.00</td>
<td>0.06</td>
<td>11.00</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Arg</td>
<td>CGA</td>
<td>3.00</td>
<td>0.09</td>
<td>10.00</td>
<td>0.27</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>Arg</td>
<td>CGT</td>
<td>13.00</td>
<td>0.41</td>
<td>3.00</td>
<td>0.08</td>
<td>2.00</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Arg</td>
<td>CGC</td>
<td>10.00</td>
<td>0.31</td>
<td>5.00</td>
<td>0.14</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Gin</td>
<td>CAG</td>
<td>12.00</td>
<td>0.40</td>
<td>24.00</td>
<td>0.73</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Gin</td>
<td>CAA</td>
<td>18.00</td>
<td>0.60</td>
<td>9.00</td>
<td>0.27</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>His</td>
<td>CAT</td>
<td>6.00</td>
<td>0.67</td>
<td>4.00</td>
<td>0.36</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 3. (Continued) Codon usage of the white eye genes in *B. dorsalis*, *D. melanogaster*, and *L. cuprina*

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>His</td>
<td>CAC</td>
<td>3.00</td>
<td>0.33</td>
<td>7.00</td>
<td>0.64</td>
<td>1.00</td>
</tr>
<tr>
<td>Leu</td>
<td>CTG</td>
<td>13.00</td>
<td>0.17</td>
<td>38.00</td>
<td>0.44</td>
<td>3.00</td>
</tr>
<tr>
<td>Leu</td>
<td>CTA</td>
<td>8.00</td>
<td>0.10</td>
<td>7.00</td>
<td>0.08</td>
<td>2.00</td>
</tr>
<tr>
<td>Leu</td>
<td>CTT</td>
<td>9.00</td>
<td>0.12</td>
<td>10.00</td>
<td>0.11</td>
<td>2.00</td>
</tr>
<tr>
<td>Leu</td>
<td>CTC</td>
<td>14.00</td>
<td>0.18</td>
<td>19.00</td>
<td>0.22</td>
<td>1.00</td>
</tr>
<tr>
<td>Pro</td>
<td>CCG</td>
<td>9.00</td>
<td>0.29</td>
<td>17.00</td>
<td>0.47</td>
<td>1.00</td>
</tr>
<tr>
<td>Pro</td>
<td>CCA</td>
<td>13.00</td>
<td>0.42</td>
<td>10.00</td>
<td>0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>Pro</td>
<td>CCT</td>
<td>6.00</td>
<td>0.19</td>
<td>2.00</td>
<td>0.06</td>
<td>4.00</td>
</tr>
<tr>
<td>Pro</td>
<td>CCC</td>
<td>3.00</td>
<td>0.10</td>
<td>7.00</td>
<td>0.19</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 4. Codon bias and GC content comparisons for the *B. dorsalis* and *D. melanogaster* white eye and actin genes.

<table>
<thead>
<tr>
<th>Gene <em>a</em></th>
<th>B. d white</th>
<th>D. m white</th>
<th><em>B. dorsalis</em> Actin</th>
<th><em>D. melanogaster</em> Actin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B. d</td>
<td>D. m</td>
<td>BdA1</td>
<td>BdA2</td>
</tr>
<tr>
<td></td>
<td>white</td>
<td>white</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>GC content of silent site (%)</td>
<td>0.411</td>
<td>0.719</td>
<td>0.554</td>
<td>0.558</td>
</tr>
<tr>
<td>GC content of CDS <em>b</em> (%)</td>
<td>0.462</td>
<td>0.555</td>
<td>0.490</td>
<td>0.510</td>
</tr>
</tbody>
</table>

*a* Sources of sequences: *D. melanogaster* Act79B, Act87E, and Act88F have Genbank nucleic acid database accession numbers J01064, K00674, and J01065 respectively. Act 57A has the Genbank accession number K00672 for exon 1 and K00673 for exon 2. The accession numbers for the *B. dorsalis* actin genes are L12253, L12254, L12255, and L12256 respectively from Genbank database. The *D. melanogaster* white gene is from Genbank entry X51749. B.d: *B. dorsalis*, D.m: *D. melanogaster*.

*b* CDS: coding region sequence.
compatible with the expression pattern (Fjose et al., 1984) and the evolutionary conservativeness of the white genes (Fig. 9).

3.4.4 Protein structure prediction

Results of the DNA sequence analysis of the oriental fruit fly white eye gene predicts a polypeptide of 654 amino acids encoded by exons 2-6. It is a membrane-associated ATP-binding transport protein.

The carboxy-terminal of the white eye protein, from position 406 to the end, is encoded by exons 5 and 6. It is highly hydrophobic and has six lipid bilayer spanning segments (Fig. 11). The amino-terminal, from the beginning to position 407, is encoded by exons 2, 3, 4a, and 4b. It is relatively hydrophilic and contains a ATP-binding site. Possible protein secondary structure were predicted using the Garnier et al. (1978) and the Chou-Fasman (1978) methods. Though glycosylation has not been studied in the white protein of D. melanogaster, it is present in several ATP-binding transport proteins, such as the human multidrug resistance protein (Endicott & Ling, 1989) and the cystic fibrosis protein (Riordan et al., 1989). The multi-drug resistant protein mediates one type of multi-drug resistance in the cells by decreasing the accumulation of drugs. Cystic fibrosis is an autosomal recessive genetic disorder. It affects lung airway, pancreas and sweat glands. There is high electrical potential difference across the epithelial surfaces of the cystic fibrosis respiratory tract. The oligosaccharides of a glycoprotein may involve in the cell-cell and protein-protein recognition and interaction. The potential glycosylation sites of the white eye protein of the oriental fruit fly are also predicted.
Figure 11. Hydropathy profile and the secondary structure of the white eye protein of *B. dorsalis*.

The mean hydropathy index is composed of amino acid residues i-9 to i+9 plotted against i, where i represents amino acid number (Kyte & Doolittle, 1982). The secondary structure of the corresponding positions is predicted using the method by Chou-Fasman or by Garnier *et al.* Six hydrophobic lipid bilayer spanning segments are labelled from 1 to 6.

Figure 12. Hydropathy profile and the secondary structure of the white protein of *D. melanogaster*.

The mean hydropathy index is composed of amino acid residues i-9 to i+9 plotted against i, where i represents amino acid number (Kyte & Doolittle, 1982). The secondary structure of the corresponding positions is predicted using the method by Chou-Fasman or by Garnier *et al.* Six hydrophobic lipid bilayer spanning segments are labelled from 1 to 6. GOR: the Garnier *et al.* method (Garnier *et al.*, 1978), CF: Chou-Fasman method (Chou & Fasman, 1978), T: turns, H: a-helices, S: b-sheets, Gly. sites: glycosylation sites.
The *white eye* protein of the oriental fruit fly is highly homologous to that of *D. melanogaster* at the protein sequence level with 93% similarity (Fig. 9). High homology can also be expected at the protein-secondary structure level. The ATP-binding amino-terminal of the *white* protein is from position 24 to 439 (Fig. 12). It is encoded by exons 2, 3, and 4. The transmembrane carboxy-terminal is from position 440 to the end. It is encoded by exon 5 and 6. The two *white eye* genes share similar hydropathy, α-helices, β-sheets and turns profile (Fig. 11 and 11). It is interesting that the *brown* protein has the similar secondary structure (Fig. 13). All these possible secondary structures can only be verified when the proteins crystallographic data become available.

**3.4.4.1 Exon 1 of the white eye protein**

Exon 1 is important for the function of the *white eye* protein. Without it there was no pigmentation in the compound eye, ocelli, Malpighian tubules and testis of *D. melanogaster* (Levis et al., 1985). It is proposed to be a leading peptide that directs the protein to the right location. The oligopeptide permease in *S. typhimurium*, another ATP-binding transport system, is homologous to the *white* protein of *D. melanogaster*. Hiles et al. (1987) demonstrated that the oligopeptide permease subunits had no typical signal sequences, however, the hybrid protein could be directed to the inner membrane by fusing less than 50 amino acid of the N-terminus of OppB subunit to the β-galactosidase.

The sequence of the leading peptide is variable among different species and different proteins, so it is hard to identify based on the genomic DNA sequence. The cDNA sequence is needed for its identification, however this is outside the scope of the present work. It is possible that exon 1 is there in the genomic DNA sequence, but it
Figure 13. Hydropathy profile and the secondary structure of the brown protein of D. melanogaster.

The mean hydropathy index is composed of amino acid residues i-9 to i+9 plotted against i, where i represents amino acid number (Kyte & Doolittle, 1982). The secondary structure of the corresponding positions is predicted using the method by Chou-Fasman or by Garnier et al. Six hydrophobic lipid bilayer spanning segments are labelled from 1 to 6. GOR: the Garnier et al. method (Garnier et al., 1978), CF: Chou-Fasman method (Chou & Fasman, 1978), T: turns, H: a-helices, S: b-sheets, Gly. sites: glycosylation sites.
Hydrophobic

Hydrophilic

Gly. Sites

CFT
CF H
CF S
GOR T
GOR H
COR S
cannot be identified at present. The white eye protein of B. dorsalis is highly homologous to that of D. melanogaster. Therefore we can expect that the hydrophobicity and the secondary structure profile of the-exon 1 of B. dorsalis white eye protein is similar to that of D. melanogaster. The D. melanogaster exon 1 profile (Fig. 14) can serve as a blueprint in identifying the B. dorsalis one in future research.

3.4.4.2 C-terminal of the white eye protein

The carboxy-terminal of the oriental fruit fly white eye gene consists of six hydrophobic helices that are capable of spanning a lipid bilayer (Fig. 11). This result is based on the research done by Kyte and Doolittle (1982), which showed that when the average hydrophobicity of a 19-residue peptide segment was greater than +1.6, it was highly possible that the segment spanned the lipid bilayer membrane. Comparison of the different transmembrane proteins reveals higher structural similarity among these proteins regardless of heterogeneity of the amino acid sequence. Each protein has one to several transmembrane domains and each domain has six transmembrane segments. For example, sodium and calcium channels contain 24 transmembrane segments (Noda et al., 1984; Tanabe et al., 1987; Catterall, 1988), multidrug resistant and cystic fibrosis proteins have 12 (Riordan et al., 1989; Endicott & Ling, 1989), and the B and C subunits of the oligopeptide permease of S. typhimurium each contain 6 (Hiles et al., 1986). It was hypothesized that the transmembrane hexamers interact with each other to form higher-order structure (Kane et al., 1990). Substrate may be transported through the pore formed by the interaction of the hexamers just as in the sodium and calcium channels (Noda et al., 1984; Tanabe et al., 1987; Catterall, 1988).
Figure 14. Hydropathy profile and the secondary structure of exon 1 of the white protein of *D. melanogaster*.

The mean hydropathy index is composed of amino acid residues i-2 to i+3 plotted against i, where i represents amino acid number (Kyte & Doolittle, 1982). The secondary structure of the corresponding positions is predicted using the method by Chou-Fasman or by Garnier *et al.*. GOR: the Garnier *et al.* method (Garnier *et al.*, 1978), CF: Chou-Fasman method (Chou & Fasman, 1978), T: turns, H: α-helices, S: β-sheets, Gly. sites: glycosylation sites.
Figure 15. Alignment of the amino acid sequences of the ATP-binding sites.

Gaps are introduced to maximize the alignment. Only the amino acids that are different from the consensus sequence are shown in small case. The protein sequences are the white eye gene of *B. dorsalis* (Bdatp), white gene (Genbank accession number X51749) of *D. melanogaster* (Dmatp), brown gene (Genbank accession number M20630) of *D. melanogaster* (brownatp), Oppd (Oppdatp) and Oppf (Oppfatp) subunit of the oligopeptide permease of *S. typhimurium* (Genbank accession number X05491), human multidrug resistance gene (Genbank accession number M14758) carboxyl-terminal half (Mdrcatp) and the amino-terminal half (Mdrnatp), and the human cystic fibrosis gene (Genbank accession number M28668) carboxyl-terminal half (Cfcatp) and the amino-terminal half (Cfnatp). The "." symbol stands for no corresponding sequence data at a specific position. The secondary structure is predicted by the method of Chou and Fasman (1978). α: α-helices, β: β-sheets.
3.4.4.3 N-terminal of the white eye protein

The amino-terminal of the white eye protein consists of a conservative ATP-binding site, which is also present in the periplasmic binding protein-dependent transport systems in prokaryotes (Higgins et al., 1990a) and some eukaryote transport systems (Riordan et al., 1989; Endicott & Ling, 1989). Each transport protein has one to two ATP-binding domains. The whole transport system requires two ATP-binding domains or subunits (Higgins et al., 1990b). The energy of the transport process comes from the hydrolysis of ATP. The ATP-binding domain has no membrane-spanning helix and is accessible by the protease only from the cytoplasmic side of the membrane (Gallagher et al., 1989). It indicates that the domain may be peripherally associated with the inner membrane, which is compatible with the proposed ATP hydrolysis function (Higgins et al., 1990b).

There is extensive similarity of the amino acid sequence (Fig. 15) and protein secondary structure (Fig. 11, 11, 12) in the ATP-binding domains in different transport systems from different organisms. Based on the primary and secondary structure of the ATP-binding domains, Hyde et al. (1990) constructed the tertiary structure model. The amino acid sequences aligned (Fig. 15) are only the segments corresponding to those used in the tertiary structure construction for the ATP-binding cassette (Hyde et al., 1990). Loop 2 interacts with the membrane-associated domain or subunit to couple the ATP hydrolysis to the conformational change of the whole system to facilitate the substrate transport. Homology among the ATP-binding cassettes indicate that they may derive from a common evolutionary origin (Higgins et al., 1990b).
3.4.4.4 Evolution of the white eye protein structure

The ATP-associated transport systems in prokaryotes and eukaryotes consist of two basic units, each of which comprises of one ATP-binding site and a hydrophobic membrane-interacting hexamer (Higgins et al., 1990b). The basic unit is abbreviated as ATP-TM. The two basic units are homologous to each other at the high protein structural levels and at the primary sequence level. In some prokaryote systems, different subunits are encoded by different genes (Hiles et al., 1987; Higgins et al., 1990a) and frequently subunits are fused with each other. Evolution favors the gene fusion model because it improves the coordinate regulation of the different peptides. For example, two ATP-binding subunits of the E. coli ribose transport system are fused together and all four domains are encoded by one gene in the multi-drug resistant protein (Endicott & Ling, 1989; Chen et al., 1986; Ueda et al., 1987; Kioka et al., 1989). This is also observed in the cystic fibrosis protein (Riordan et al., 1989).

The D. melanogaster white and brown proteins each can be viewed as the combination of one hydrophobic domain and one ATP-binding site or just one basic unit (ATP-TM). White and brown proteins share extensive similar amino acid sequence (Fig. 16) and protein secondary structure (Fig. 12 and 13). It is proposed that each transport system requires two basic-units (ATP-TM). The model is consistent with the proposed joint function of the two proteins being involved in transport of the eye pigment synthesis chemicals through the membranes (Tearle, 1991).

Kane et al. (1990) proposed that different ATP-associated transport systems were assembled from the same ATP-binding cassette and different transmembrane domains
Figure 16. Amino acid sequence similarity of the white \textit{(w)} and brown \textit{(br)} proteins of \textit{D. melanogaster}.
The identical amino acids are paired by a "1", the similar ones with comparison value equal or greater than 0.50 by a ":", and the ones with comparison value equal or greater than 0.10 by a ".". The "\" symbol stands for no corresponding sequence data at a specific position. (Needleman & Wunsch, 1970).
The ATP-binding cassettes in different ATP-associated transport systems may have evolved from the same ancestor by exon shuffling. The alignment of amino acid sequence of the ATP binding domains indicates that they share about 30% amino acid sequence similarity (Fig. 15).

Intron phase refers to the codon position where an intron splits an exon. For example, if an intron sits between the first and second nucleotides of a codon, it is in phase 1. If the intron lies between the second and third nucleotides of a codon, it is in phase 2. If intron lies between two codons, it is in phase 0. Introns with the same phase are considered homologous. The nucleotide sequence of the homologous introns can drift rapidly. The position and even the number of introns can change by deletion or insertion, but the phase in which they split the reading frame is preserved during evolution (Patthy, 1987).

The same concept is applied to the definition of exon phase. The exon phase depends on the phase of the introns at its two ends. If the intron at its 5' end is 1 and 3' end is 2, the exon phase will be (1,2). Symmetrical exons are those with the same phase of introns at both ends. Symmetrical exons are favored in the exon shuffling evolution for they are easily undergo duplication and insertion into the other intron (Patthy, 1994).

The ATP-binding domains of the white eye genes of B. dorsalis and D. melanogaster have similar exon-intron junction phase. Both of them are encoded by three exons with the phase of (0,1), (1,2) and (2,0), respectively. The phase for the whole ATP-binding domain is (0,0). In another words, it is a big symmetrical exon, which is favored in gene evolution by exon shuffling. Exon shuffling, a school of thought on gene
Figure 17. ATP-associated transport system evolution model (modified from Kane, 1990).
The shaded rectangles stand for the transmembrane domains. The open rectangle represents the ATP-binding domain (ATP). The ATP-binding domain and the transmembrane domains (TM) function as separate subunits in the oligopeptide permease (S. typhimurium), or fuse together in the configuration of TM-ATP in the white and brown proteins (D. melanogaster) and TM-ATP-TM-ATP in the cystic fibrosis protein (C. hominidae).
Transmembrane domains

ATP binding domain

Oligopeptide permease

Cystic fibrosis

white

Brown
evolution, proposes that all proteins have been derived from a small number of exons by duplication and juxtaposition (Patthy, 1994). The conservation of the exon phase and the 30% amino acid sequence similarity in the ATP-binding cassettes may indicate that they derive from the same ancestor.

3.5 Intron

The intron sequence was determined by the relative canonical intron-exon junction sequence and the conservative protein sequence. An extra intron was found in the oriental fruit fly white eye gene splitting exon 4 of the D. melanogaster white gene into two exons just before an α-helics structure. As white eye gene sequences are determined in more insect species, we can determine whether the B. dorsalis gene has gained an extra intron or the D. melanogaster has lost one during evolution.

The comparison of the correlated introns between the white eye gene of the oriental fruit fly and that of D. melanogaster shows low similarity, around 40%. It may be caused by the lack of selective pressure.

3.6 Future research

3.6.1 Isolation of the white eye gene cDNA clone

To finely characterize the white eye gene of the oriental fruit fly, the cDNA sequence is needed. The exon-intron junctions can be determined from this sequence and the transcription start site can be identified by the reverse primer extension method.

3.6.2 White eye gene used as a marker gene

Using the white eye mutant as a host the wild-type white eye gene can be used as a functional marker gene for genetic transformation. With the obvious eye color change
from white to deep purple the activity of the transformation system can be traced easily. The transformed individual can be selected.

3.6.3 Localization of the white eye gene

3.6.3.1 White eye gene as an autosome marker

The white eye gene can be localized on the chromosome by in situ hybridization. Five linkage groups with eye color and morphology, puparium color and morphology, and wing morphology markers have been defined in B. dorsalis (McCombs, 1992c). It would be interesting to pair the linkage groups with the chromosomes. This can provide a molecular marker for an autosome, support the evolutionary study of chromosome elements in different species, and indirectly aid in the development of a stable genetic sexing strain.

3.6.3.2 Chromosome evolution study

It has been observed that some gene clusters are linked persistently in different species (Loukas & Kafatos, 1986, 1988; Zacharopoulou et al., 1992; Heckel, 1993). For example, chorion, vitellogenin, and integrin are localized on 5L chromosome in Mediterranean fruit fly and on the X chromosome in D. melanogaster. The molecular evidence directly proved that at least part of the chromosome 5L arm of C. capitata was homologous to the X chromosome of D. melanogaster (Zacharopoulou et al., 1992). This makes it possible to predict the location of similar genes in other species and study the evolutionary relationship of the chromosomes among species. So far, the actin genes of B. dorsalis have been defined to a specific location on different chromosomes (He,
1993). With the increasing number of molecular markers, the cytogenetics of this agricultural pest can be extensively studied.

3.6.3.3 Development of a stable-genetic sexing strain

Currently, genetic sexing strains are constructed by translocating the marker containing autosome to the Y-chromosome. When the autosomal segment is large, the recombination between the autosome translocation breakpoint and the marker cause the breakdown of the genetic sexing strain. One way to overcome the problem is to bring the chromosome breakpoint close to the marker to limit the recombination rate. With the precise chromosome position of the white eye gene we can determine the distance between the white eye gene and the breakpoint by looking at the chromosome polytene pattern (Kerremans et al., 1991). If the autosome marker used is in the white eye gene linkage group and the map distance of the white eye gene to the marker is available, it is easier to estimate the distance of the translocation breakpoint to the marker.

3.6.4 Localization of the white eye protein

The proposed functional sites of the white protein of D. melanogaster are based on the mutant phenotype in different organs, electron microscope studies, eye pigment biosynthetic pathway, and cross relationships among different eye mutants (Summers et al., 1982; Tearle, 1991). The possible function of the white eye protein of B. dorsalis is deduced from similar studies and its sequence homology to the D. melanogaster white protein. The white eye gene is epistatic to the yellow eye and the mandarin red genes (McCombs, 1992c). The epistatic effect may correspond to the sequence of the eye pigments biosynthetic steps the mutant genes involved in. It supports the proposed
function of white eye gene product transporting the substrates and reaction intermediates of the eye pigment biosynthetic pathway. The white eye mutant phenotype also shows up on the head and Malpighian tubules: It may be involved in other important physiological processes.

Precise localization of the white eye protein will be the direct evidence of the function of the protein. In D. melanogaster the expression of the white protein is low with a peak level of 0.003% in the mature third instar larva when the eye pigments are synthesized (Fjose et al., 1984). To overcome the low expression obstacle, we could take advantage of the hybrid protein by fusing the white eye protein to β-galactosidase. The hybrid protein would be expected in the cell fractions of mitochondria and membrane. It can be detected by immunoblotting using the anti-β-galactosidase antibodies.
4. Summary

The genomic DNA library of the oriental fruit fly has been constructed in λ-GEM-11 vector. The size of the library is $3 \times 10^5$ recombinant phage per microgram genomic DNA. One positive clone λ-DW-1A1, which hybridized to the white eye gene probe of the oriental fruit fly, was found.

A DNA fragment of 3859 bp long was sequenced. Analysis results indicate that this fragment includes introns 2, 3, 4a, 4b, 5 and partial 1; exons 2, 3, 4a, 4b, and 5, and the polyA addition site. Compared to the white gene of D. melanogaster, there is one extra intron (4b), which splits exon 4 into 4a and 4b in the white eye gene of the oriental fruit fly. The corresponding exons are of similar size and phase. The two white eye genes are highly homologous. The coding region sequence of the two genes has 93% similarity at the amino acid level and 70.8% at the nucleotide level. There is a high level of silent mutation, 71.4%, which could explain the apparent inconsistency of these similarity comparisons. The two white eye genes have a different pattern of codon usage bias. D. melanogaster prefers GC base while the oriental fruit fly prefers AT base at the silent site of the codon. GC content of the silent sites correlates well with the GC content of the coding region sequence. That is, D. melanogaster has a higher GC content than the oriental fruit fly at the codon silent sites and as a whole gene. The codon usage bias and GC content of the coding region sequence of the white eye genes is consistent with that of the actin genes of the two species.
The carboxy-terminal of the white eye protein, from position 406 to 654, is encoded by exons 5 and 6. It is highly hydrophobic and consists of six lipid bilayer spanning segments. The amino-terminal, from position 1 to 405, is encoded by exon 2, 3, 4a, and 4b. It is hydrophilic and contains an ATP-binding site. White and brown proteins of D. melanogaster share similar protein secondary structure and hydropathy profile with the white eye protein. All three can be viewed as one ATP-binding site fusing to one transmembrane domain. Comparison among the ATP-binding transport systems across different species of prokaryotes and eukaryotes indicates that there are two ATP-binding domains and two transmembrane domains for each system. This supports the proposed joint function of white with brown and white with scarlet to transport substrates, reaction intermediates, and products of the eye pigment synthetic pathways.

The ATP-binding transport systems may be assembled from the same ATP-binding domain and different transmembrane domains. Evidence for a common ancestral ATP-binding in B. dorsalis and D. melanogaster comes from two sources. The alignment of the ATP-binding domain amino acid sequences shows about 30% similarity. Also the ATP-binding domains of the white eye genes of B. dorsalis and D. melanogaster share the same symmetrical exon phase (0,0), which is favored in the exon shuffling process.
5. Bibliography


Moore, I. & D. Nadel. 1961. A bacterium present within the eggs of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann). Ktavim 11: 133.


