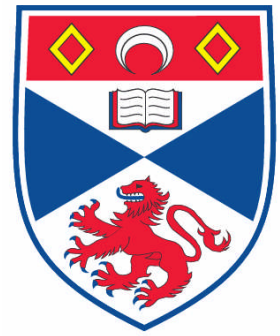
THE GENETICS OF SEXUALLY DIMORPHIC TRAITS IMPLICATED IN SEXUAL ISOLATION IN *DROSOPHILA* : QTLS AND CANDIDATE GENES

Robert Andrew James

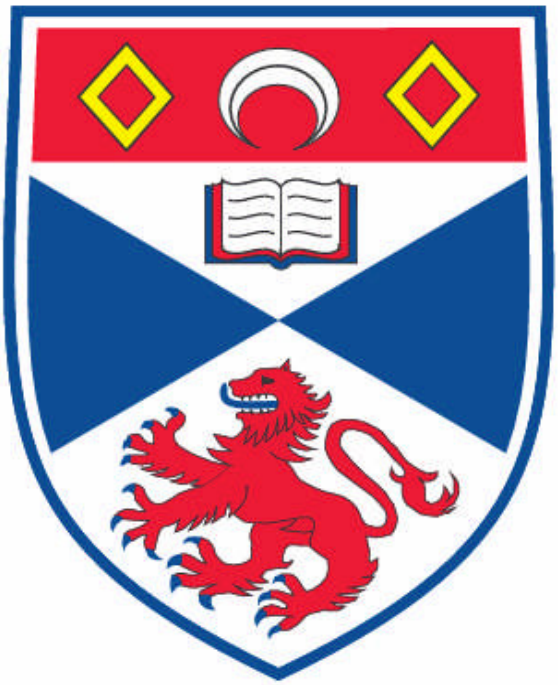
A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2008

The genetics of sexually dimorphic traits implicated in sexual isolation in *Drosophila*: QTLs and candidate genes.

Robert Andrew James



Thesis presented for the degree of Doctor of Philosophy

University of St Andrews

2008

“Time flies like an arrow; fruit flies like a banana”

Groucho Marx

TABLE OF CONTENTS

Declaration

I, R. Andrew James, hereby certify that this thesis, which is approximately 32000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date …… signature of candidate ……

I was admitted as a research student in September 2004 and as a candidate for the degree of PhD in August 2005; the higher study for which this is a record was carried out in the University of St Andrews between 2004 and 2008.

Date …… signature of candidate ……

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Date …… signature of candidate ……

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …… signature of supervisor ……..…

Acknowledgements

I thank the National Environment Research Council for funding this study. I thank Mike Ritchie, Anastasia Gardiner and Steven Goodwin for their sound scientific advice as well encouragement and support, Rosemary Bevan and Tanya Sneddon for their long term maintenance of the fly stocks, and help in the lab. I thank all the staff at the Harold Mitchell Building. I am also grateful to many more friends for much moral support during my PhD project, Gordon, Patty, Kit, Elisa, Gill, Siaf, Ingrid, John, Sylvie, Helena, Sarah, Charlie, Ben, Tino, Jeff, Ellie, Daniel, Christoph, Nicola, Claire and Mairi. I thank my family, especially my Mum and Dad and my three beautiful children Daniel, Joseph and India-Rose.

Thanks to all the kind folk of Fife for making my stay a pleasant one, cheers!

CHAPTER 1

Introduction

1.1 **Background on the development of modern evolutionary synthesis**

Charles Darwin established the concept of natural selection when he published "The Origin of Species: by means of natural selection or the preservation of favoured races in the struggle for life" in 1859, documenting the unique adaptations of many of the animals inhabiting the Galapagos Islands. One of his most famous observations was of the different species of finches inhabiting the islands. The size and beak morphology appeared to have an association with the type of food found within their respective environmental niche. His careful observations lead him to understand that naturally occurring variation within populations was selected upon and that particular advantageous traits would occur more frequently in future generations. Darwin coined this concept "descent with modification". Simultaneously during the 1860's an Augustinian monk named Gregor Mendel was carrying out pioneering work on genetic inheritance. It was Mendel's work focussing on several traits in *Pisum sativum* that lead to an understanding of the genetic assortment and segregation of parental alleles. He discovered that specific traits were passed on in units, now known as genes and that new combinations of the parental traits could be observed in subsequent filial generations (PURVES *et al.* 2000). Mendel and Darwin's early pioneering work highlighted the basic factors of evolution; phenotypic variation, genetic heritability and selection and consequently paved the way for the synthesis of modern evolutionary theory. It was from 1917 onward in the U.S.A that Sewall Wright (1932) constructed a comprehensive evolutionary theory, which included the

concepts of inbreeding, gene flow and random genetic drift. Fisher furthered the understanding of the principle tenets of the evolutionary synthesis, in his application of mathematical principles to the laws of natural selection (FISHER 1930). Dobzhansky (1937) went on to emphasise the importance of genetic mutation on species diversity. The major evolutionary tenet is that genetic mutation is the original source of genetic variation. Mutation results in changes within amino acid sequences that form the polypeptide chains of proteins. This in turn can lead to the alteration of protein function and consequent phenotypic variance. The total phenotypic variation is a product of genetic and environmental variance and can be written as $V_P = V_G + V_E$ (FALCONER and MACKAY 1996). Once variance at numerous loci has been established the process of sexual reproduction and random recombination of the different alleles promotes the rate of adaptation. This rate will of course depend on the environmental dynamics and the factors controlling gene flow within and between populations, which are the governing principles of speciation (GRIFFITHS *et al.* 2000).

1.2 The importance of sexual isolation in the process of speciation

Natural selection can cause evolutionary change within a single lineage, however the divergence of one ancestral species into one or more different species can only occur when one population becomes reproductively isolated from the other. Mayr (1963) stated that "species are groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups". The different selective pressures that operate on diverging populations can result in the accumulation of complementary sets of co-adapted genes within species that will reflect their respective adaptations and these same genes are thought to be responsible for interspecific hybrid dysfunction, which can result in both inviability and or

sterility (TURELLI and ORR 1995). Reproductive isolation can also operate before the formation of the zygote. The main types of reproductive isolation therefore fall into two main categories, postmating and premating.

The postmating factors that cause reproductive isolation are biological factors and can occur when the zygote fails to develop, or the first generation are unable to produce a viable second filial generation. The other factor is that the offspring are severely compromised in their rate of survivorship or reproductive fitness leading to F_2 hybrid breakdown (SINGH and HOENIGSBERG 2004). Experiments to investigate genes that influence speciation within animal populations have tended to concentrate more on those affecting male sterility. This is relatively easy to measure, and it is generally thought that the genetic incompatibilities affecting male sterility accumulate at a faster rate than those causing other types of reduced hybrid fitness (PRESGRAVES 2003).

Through the relatively recent progress made in genetic studies and the increased efficiency in the rescue of hybrid incompatibilities there is now a greater understanding of the biological reasons for hybrid inviability and why, during interspecific crosses, a general rule is that the heterogametic sex is the first to show dysfunction (WU and HOLLOCHER 1998). This rule was first formulated by Haldane (1922), and the alleles involved in these hybrid incompatibilities are thought to be, on average, partially recessive. This dominance theory neatly accounts for several phenomena (ORR 1995). A great deal of research has been carried out on the genetic factors contributing to male sterility. From experimental evidence it appears that numerous genes with an individual weak affect but strong interactions are involved

(PEREZ and WU 1995; WU and TING 2004). The *Odysseus* (*OdsH*) gene is one such gene, involved with sperm production (SUN *et al.* 2004), and it shows evidence of a history of strong directional selection between the two sibling species *D. simulans* and *D. mauritiana*, with an excess of non-synonymous to synonymous substitutions (TING *et al.* 1998). An experiment which co-introgressed the *OdsH* gene from *D. mauritiana* with the adjacent segment into *D. simulans* resulted in male sterility (PEREZ and WU 1995). The *OdsH* gene is also divergently regulated between these two species, and it appears that male sterility between *D. simulans* and *D. mauritiana* is caused by both the sequence divergence and the misregulation of the expression of this gene (WU and TING 2004). Genes that show male-biased expression have a greater ratio of non-synonymous to synonymous substitutions between species than within, indicating that they are under strong positive selection as opposed to relaxed constraint (ZHANG and PARSCH 2005). The "faster male theory" is one of the explanations for heterogametic hybrid dysfunction (TAO and HARTL 2003). However the theory does not account for infertility within female heterogametic taxa, for instance in the butterflies *Heliconius melpomene* and *H. cydno* female hybrids are completely sterile (NAISBIT *et al.* 2002). A study assessing hybrid female sterility within the *Drosophila* sibling species *D. simulans*, *D. mauritiana* and *D. sechellia* revealed that negative epistatic effects can cause female sterility between these three species. The experiment relied on constructing hybrid genotypes that allowed homozygous loci from each species to interact (DAVIS *et al.* 1994). It is possible that heterogametic hybrid dysfunction may well result from the composite effects of many factors.

The premating barriers which prevent gene flow can include spatial separation and temporal isolation, such as differing host plant species for different fly species, or

varying times at which mating takes place between species (SINGH and HOENIGSBERG 2004). However the most intriguing and important aspects of premating reproductive isolation are the ethological or behavioural aspects of an animal's courtship, that cause varying degrees of attractiveness between mates within a species group.

Behaviour is a particularly interesting component of speciation, since it is both subject to selection as well as an agent of selection (BOAKE *et al.* 2002). Female preferences for particular male traits between different races can contribute to species recognition (RYAN and STANLEY 1993) and act as a key factor in sexual isolation. The process of sexual selection may well be due to direct selection operating on species recognition systems to avoid reduced hybrid survivorship, caused by the genetic incompatibilities existing between the different species (COYNE *et al.* 2002). In 1942 Muller suggested that premating reproductive isolation occurred as a by-product of genetic divergence through allopatric speciation (MULLER 1942; SINGH and HOENIGSBERG 2004). There is a greater consensus that reproductive isolation is predominantly a result of allopatric speciation, when a geographical barrier separates populations preventing gene flow. This often results in differentiation across the entire genome (MAYR 1963). However the incidence of sympatric speciation in which diverging populations (Species) are not separated geographically is more debatable. Potential examples of sympatric speciation especially emphasise the important role of environmental selection and mate choice (SCHILTHUIZEN 2000). The significant interslope differences occurring between *Drosophila melanogaster* populations at Evolution Canyon Mt. Carmel, Israel provide evidence for the importance of ecological microsite heterogeneity as a factor influencing assortative mating and potentially sympatric speciation. The *D. melanogaster* population's sensitivity to desiccation

influences their choice of microhabitat. This has promoted an apparent microevolution of behavioural traits connected with habitat choice, facilitating population differentiation. A factor causing this example of speciation is genetically determined habitat selection resulting in reduced migration between populations (KOROL *et al.* 2006; WIENER and FELDMAN 1993). Ecological differentiation between incipient species is often associated with assortative mating limiting gene flow (YAWSON *et al.* 2007). It has been predicted that diverging species may have mosaic genomes, composed of highly differentiated and undifferentiated regions, due to selection operating on the regions associated with ecological adaptation and mating behaviour. Consequently, introgression will be much reduced within these same regions (GENTILE *et al.* 2002; MACHADO *et al.* 2002; YAWSON *et al.* 2007).

It has been greatly debated whether sexual selection can act as a driving force for speciation without the existence of ecological niche heterogeneity. Fisher's fundamental theorem of natural selection (FISHER 1958) established that at equilibrium the additive genetic variation of traits directly affecting fitness should be close to zero (BOAKE *et al.* 2002). However many theoretical models show that the "Fisherian" runaway process of sexual selection could instigate events of divergence, through behavioural differentiation (LANDE 1981). The genes associated with sexually dimorphic phenotypic expression are often the most divergent among species. This may be a result of the co-evolution of male and female sexually selected traits, as well as the lack of constraint the genes are under whilst in the genome of the sex where they are not expressed. However a recent study shows evidence that the higher amino acid polymorphism existing within male-biased genes is due to positive selection as opposed to a lack of constraint, with a significantly higher rate of non-synonymous to

synonymous changes between different species than within (PRÖSCHEL *et al.* 2006; ZHANG and PARSCH 2005). In order for female preference to drive speciation there must be a level of additive genetic variance in sexually selected traits within a population. It has been suggested that female preference for particular male traits, rather than depleting the genetic variance due to directional selection, can increase phenotypic variation by favouring extremes and supporting a higher mutation rate for these traits (PETRIE and ROBERTS 2007; POMIANKOWSKI and MØLLER 1995). Recent research also considers the indirect additive genetic variance derived from the conditional dependence associated with secondary sexual triats (HOLZER *et al.* 2003). The genetic basis of maternal behaviours, such as selection of offspring habitat and food provisioning may also indirectly influence the additive genetic variance associated with sexually selected traits in sons (MILLER and MOORE 2007).

If sexual selection acts as a major driver of the speciation process, there would be a greater diversity of species that show stronger signals of sexual selection. A robust study carried out on insects confirmed that species richness was significantly greater in polyandrous clades (ARNQVIST *et al.* 2000). However a large study on mammals, butterflies and spiders, found species richness was unrelated to size dimorphism (GAGE *et al.* 2002). The recent radiation of African Cichlid fish is one example where diverse male body colour acts as a target for female preference (SEEHAUSEN *et al.* 1999). However, analysis of the importance sexual selection has on this apparent rapid radiation remains unresolved due to difficulties in obtaining phylogenetic data. Also it has been argued that the large ecological niche differences, suggests sexual selection rather than being a driver of speciation, has more likely facilitated the process through assortative mating (KONDRASHOV and SHPAK 1998). The rapid

evolution of *Drosophila* on the Hawaiian oceanic islands has inspired theoretical models emphasising the influence of sexual isolation as a prominent driving force for the process of speciation (CARSON and TEMPLETON 1984; KANESHIRO and BOAKE 1987). The asymmetrical sexual isolation model (KANESHIRO 1989) proposes that founder-flush cycles may have resulted in the loss of elements involved in male courtship, as well as a decrease in female discrimination. This could then cause relaxed female preference and the simplification of courtship behaviour, compared to populations that retain the complexity and preference of their ancestral courtship behaviour. However the model has constraints, in that there must be little or no gene flow between diverging populations (KOEPFER and FENSTER 1991). The model has been criticised and many believe that the varied ecological range of niches present on the Hawaiian Oceanic Islands is sufficient to explain the vast species diversity (BARTON and CHARLESWORTH 1984). The influence of sexual isolation on facilitating the speciation process is accepted as an influential factor, however the direct affects are still unresolved (BOAKE 2005).

1.3 *Drosophila melanogaster*: Courtship behaviour and the sexually dimorphic traits influencing sexual isolation

Drosophila melanogaster is an outgroup of the *simulans* clade which includes *D. simulans*, *D. mauritiana* and *D. sechellia*. All of the species within the *simulans* clade differ from *D. melanogaster* by a large paracentric inversion on the right arm of the third chromosome. The ancestor of the *simulans* clade probably diverged from *melanogaster* ~2.5-3.4 mya, with *sechellia* diverging from *simulans* ~0.6-0.9 mya (HEY and KLIMAN 1993). *D. sechellia* is endemic to the Seychelles archipelago (LACHAISE *et al.* 1988), and unlike the other members of the clade is a specialist,

using *Morinda citrifolia* as its host plant (LOUIS 1986). The unique adaptation of the ancestor of *D. sechellia* to *M. citrofolia* may have been due to competition avoidance and the postzygotic isolation may have then followed as a consequence of pleiotropy (MACDONALD and GOLDSTEIN 1999). The crossing of these two species results in F1 male infertility (LEMEUNIER *et al.* 1986), and during copulation a low level of heterospecific sperm is transferred even during long copulations (PRICE *et al.* 2000).

The shape of the posterior lobe of the male genital arch is the most distinctive morphological difference between these two species (ASHBURNER 1989). The sex comb present on the prothoracic legs of the male is used to hold onto the female during mounting (COOK 1977; COYNE 1985). The phylogenetic patterns of the morphological differences of the sex comb between different Drosophilid lineages indicates it may play a role in species divergence (GRAZE *et al.* 2007). Experiments involving the removal of the sex comb results in severely affecting the male's ability to inseminate females (COOK 1977; COYNE 1985; SPEITH 1952).

Initial observational studies concluded that there was not a great deal of variation in the Mate Recognition Systems within *D. melanogaster* (HENDERSON and LAMBERT 1982). Further investigations have revealed that components of MRS vary within and between species (BUTLIN and RITCHIE 1994). The Mate Recognition Systems within *Drosophila* alternate in emphasis between visual, auditory and chemosensory signal detection (MARKOW and O'GRADY 2005). The difference in male courtship song and female pheromone blends (COBB and JALLON 1990; RITCHIE *et al.* 1999) are two important behavioural traits affecting prezygotic reproductive isolation between the sibling species *D. simulans* and *D. sechellia*.

The male fly detects a non-volatile pheromone produced by the female (AMREIN and THORNE 2005; COOK 1979; GREENSPAN and FERVEUR 2000; MANNING 1959; TOMPKINS *et al.* 1983; VENARD *et al.* 1989) by tapping her abdomen with a gustatory organ situated on his foreleg. The pheromones are composed of cuticular hydrocarbon chains (CHCs). Most species of the melanogaster group of *Drosophila* are sexually monomorphic for CHCs with high levels of monoenes, usually 7-tricosene (7-T). *D. sechellia* (like *D. melanogaster*) is sexually dimorphic with dienes in females, usually 7,11-heptacosadiene (7,11-HD) (COBB and JALLON 1990). The main behavioural importance of the pheromone effect is that males of monomorphic species do not court females with the wrong compounds, so the most important contributors to sexual isolation between *D. simulans* and *D. sechellia* is the change between 7-T, present in the cuticle of *D. simulans*, to 7,11-HD in *D. sechellia* females. The second important mate recognition factor is courtship song, which the male produces by wing vibration. The song is comprised of two components known as sine and pulse song. It is thought that sine song acts to prime the female for mating (VON SCHILCHER F. V. 1976), and that of the two, pulse song is more involved in mate choice (GLEASON *et al.* 2002; VON SCHILCHER F. V. 1976). The male produces a train of pulse song with intervals between each pulse known as the interpulse interval (IPI). The IPI is species-specific and females mate more quickly on hearing homospecific IPIs (BENNET-CLARK and EWING 1969 ; KAWANISHI and WATANABE 1980 ; RITCHIE *et al.* 1999; TOMARU *et al.* 2000; VON SCHILCHER F. V. 1976).

1.4 The role of the sex determination genes in the expression of sexual dimorphic traits within *Drosophila melanogaster*

The sex determination genes; *transformer* (*tra*), *doublesex* (*dsx*) and *fruitless* (*fru*) are master genes controlling regulatory cascades during development and are responsible for determining the sexual morphology and behaviour of the adult fly (BAKER *et al.* 2001; BURTIS 1993). In *Drosophila* the decision is made early on as to its sexual development and initially depends on the X-chromosome to autosome ratio, if it is a ratio of 1 then the *sex lethal* transcript is translated and the female sex differentiation pathway is switched on and the *tra* transcript is consequently translated into functional Tra protein (MCKEOWN 1992). The female *fru* transcript is altered due to the binding of the Tra protein, this causes a stop codon to be incorporated earlier on, and results in the deletion of a 101 amino acid tail, that appears to be specific to male transcripts (USUI-AOKI *et al.* 2000). *doublesex* regulates somatic sexual differentiation and is the major developmental gene controlling the sexual morphology of the adult fly. The sex specific splicing of *dsx* is also governed by *tra* and *tra-2*. In *D. melanogaster* the *fruitless* (*fru*) gene is involved in the expression of the male-specific muscle of Lawrence (MOL) and male courtship behaviour (GAILEY *et al.* 1991; HEINRICHS *et al.* 1998). The sex regulatory genes target the terminal differentiation genes, and their resulting products contribute to the sexually dimorphic characteristics of the adult (BAKER and RIDGE 1980; BURTIS 1993; BURTIS and BAKER 1989).

1.5 Candidate Genes and Quantitative Trait Loci Analysis (QTL)

Understanding the genetic determinants of sexually dimorphic traits influencing mate choice is essential for deciphering the genetic architecture of speciation, since sexual isolation is the cause of speciation in numerous taxa (BUTLIN and RITCHIE 1994;

PANHUIS *et al.* 2001). A candidate gene is a major affect gene that affects the same or a similar phenotype in more than one species (FITZPATRICK *et al.* 2005). Most candidate genes are usually identified through the "bottom up" approach which often detects the effects of single genes through mutational analysis (BOAKE *et al.* 2002). This method has been responsible for locating a number of major affect genes controlling morphology as well as behaviour. The mutations often cause major dysfunction, rather than the subtle differences you might observe in naturally occurring variation between populations and species (BOAKE *et al.* 2002).

Behaviours involved in speciation such as song and cuticular hydrocarbon chains are sex-limited, therefore the sex determination genes may influence their expression. The sex determination genes have all been implicated through mutational screening as having deleterious affects on a number of sexual dimorphic traits within *Drosophila*. *fruitless* is a candidate gene known to affect all aspects of male courtship behaviour including song (GOODWIN *et al.* 2000; RYNER *et al.* 1996; VILLELLA *et al.* 1997). *transformer* is a candidate gene affecting the production of female specific pheromones (SAVARIT *et al.* 1999), as well as the number of cycles per pulse in song (BERNSTEIN *et al.* 1992). *doublesex* plays a major role in the male genital development (SANCHEZ and GUERRERO 2001). It has been highlighted as a candidate gene affecting courtship song (GLEASON and RITCHIE 2004), and the production of female specific dienes in *D. melanogaster* (JALLON *et al.* 1988). *dsx* is also involved in the regulation and control of leg segmental identity (BAKER and RIDGE 1980; PATTATUCCI and KAUFMAN 1991) and mutations of the genes involved in segmentation result in deleterious affects on sex comb phenotypes (TOKUNAGA 1962).

The genes *desaturase1* (*desat1*), *desaturase2* (*desat2*), and *desturaseF* (*desatF*) have all been identified through mutagenesis as possible large affect candidate genes influencing variation in pheromone blend (DALLERAC *et al.* 2000; FANG *et al.* 2002; GLEASON *et al.* 2005; LABEUR *et al.* 2002). These genes affect the production of cuticular hydrocarbon chains (CHCs), which are long chain fatty acids located on the cuticle surface, which prevent desiccation and also function as contact pheromones (BLOMQUIST *et al.* 1987; GLEASON *et al.* 2005).

The assessment of the effects a known candidate gene has on naturally occurring variation can be made through Quantitative Trait Loci analysis. QTL analysis is used for the detection of genes responsible for traits that show a continuous variation. The first attempt at a mathematical model of the genetic architecture of quantitative traits, was Fisher's infinitesimal model (FARRALL 2004; FISHER 1930). This model proposed that an organism's genome was finely tuned with a number of small allelic changes accumulating over a long period of time, and large mutations would be deleterious and eliminated by natural selection. Kimura modified this theory suggesting that large favourable affect mutations had a high probability of fixation, and mutations of an intermediate size could lead to adaptation (KIMURA 1983). However the most robust explanation has recently been described in the "exponential" model, which predicts that a few major large affect genes control most of the genetic variation, and a large number of minor affect loci with increasingly smaller effects contribute to the rest of the variation (MACKAY 2001). This was inspired by Orr's prediction that larger mutations may be fixed in the early stages of adaptation, when there is more "adaptive space". And as an organism approaches the optimum the "adaptive space" decreases

and they are replaced by more numerous smaller effect mutations (FARRALL 2004; ORR 1998).

The detection of QTL depends on the segregation of numerous alleles within populations and between different species. QTL analysis relies on numerous markers scattered across the genome. The crossing of two different sibling species such as *D. simulans* and *D. sechellia* results in the reshuffling of alleles fixed in the parental strains. The markers indicate what genomic regions are present in a recombinant and a consequent assessment can be made of a markers association with a QTL within that particular section of the chromosome on the phenotypic trait. Therefore the essential factors for QTL assessment are; distinct polymorphic markers well distributed throughout the genome and a degree of variation of the specific trait between the different strains (FALCONER and MACKAY 1996). A simple method to test for a markers association with trait variation is by using a t-test or ANOVA, if a marker is close to a QTL and not segregating independently, the trait means will show a strong association with the marker's genotype. Another common approach is to assess intervals between adjacent markers. A strong association of both markers with trait variation is a good indicator of the presence of a QTL within the interval (KERSEY 1998). However the "interval mapping" approach does not take into account linkage with neighbouring markers. Within *Drosophila* different species have distinct chromosome inversions (GLEASON and RITCHIE 2004). The QTL found within such regions will have a low resolution to one specific gene, due to the low rate of recombination between the linked markers. The Composite Interval Mapping approach attempts to increase the resolution of QTL mapping. The analysis has to rely on inferred statistical probabilities, which attempt to estimate the effects of

neighbouring alleles (JANSEN and STAM 1994). The area being looked at has to be narrowed down as close as possible to the respective markers to attain a precise measure for each of the individual QTLs affect on the specific trait. Multiple regression is used in order to in narrow down the window of an area bound by two neighbouring markers. The use of additional markers as cofactors reduces the variation in the genetic background that may occur from other QTLs present within the genome (JANSEN and STAM 1994). The presence of epistatic interaction between genes within a single chromosome and between genes on different chromosomes also makes it difficult to assess with accuracy the true proportion of the effect an individual allele has on a trait (ZENG 1994). The most reliable method for mapping epistatic QTLs is Multiple Interval Mapping. It uses a search algorithm to analyse the overall genetic architecture of genetic traits within the complete genome. It incorporates a measure of the number, position, and epistatic interaction between the QTLs affecting a specific phenotypic trait (ZENG *et al.* 1999).

The candidate gene approach has tended to be more successful at detecting genes causing intraspecific trait variation than interspecific variation, and it appears that with increased phylogenetic distance the approach breaks down (HAAG and TRUE 2001). However the most recent Quantitative Loci study on the mean IPI of courtship song between *D. simulans* and *D. sechellia* found that the candidate gene *fru* was located within a QTL peak, which indicates that it may also contribute to the interspecific trait variation (GLEASON and RITCHIE 2004). Another recent QTL study implicated *dsx* as possible major affect candidate gene for the interspecific trait difference of sex comb tooth number between *D. mauritiana* and *D. sechellia* (GRAZE *et al.* 2007).

1.6 Aim and overview of project

Previous mutational analysis has implicated a number of candidate genes that may have the potential to influence traits involved in species-specific courtship behaviour. Furthermore recent QTL analysis has indicated a number of genomic regions associated with these candidate genes, which suggests they have the potential to affect naturally occurring interspecific trait variation. This study is primarily concerned with the contribution of the sex determination genes *fruitless, transformer* and *doublesex* on three sexually dimorphic traits that contribute to sexual isolation between different *Drosophila* species. In chapter two, a QTL analysis is carried out on the sex determination genes and the three desaturase candidate genes; *desat1*, *desat2* and *desatF*, assessing their affects on the cuticular hydrocarbon compound differences between *D. simulans* and *D. sechellia*. The desaturase loci have also been included since they are potential candidate genes that are likely to have a large affect on the different pheromone blends between these two *Drosophila* species. In chapter three *fru* and *dsx* are incorporated into a QTL analysis to assess the affects of these genes on interspecific variation of the mean IPI of courtship song, between *D. simulans* and *D. sechellia*. The QTL analysis for both song and pheromone blend difference incorporates Multiple Interval Mapping, which also tests for the complex epistatic interactions that may be taking place within the genetic background of these diverging Drosophilid species. The analysis of a number of Recombinant Inbred Lines is also carried out to test for the homozygous affects of the sex determination genes on the traits, sex comb tooth number and the mean interpulse interval (IPI). In chapter four a number of bioinformatic techniques are applied to test for positive selection within the exonic coding regions of the 13 different transcripts of *fruitless*. The assessment is made using a total of 10 different recently available Drosophilid genomes.

CHAPTER 2.

QTL analysis of candidate genes for cuticular hydrocarbon differences between *Drosophila simulans* and *D. sechellia*.

2.1 **Abstract**

The desaturase loci are candidate genes for cuticular hydrocarbon variation in *Drosophila*, which facilitate ecological adaptation and can influence sexual isolation. Here we score the sex determination genes and three desaturase loci and assess their affects on variation of six different cuticular hydrocarbon compounds present in *D. simulans* and *D. sechellia*. The three desaturase loci were previously implicated as potentially contributing to quantitative trait loci for 7-tricosene and 7,11-heptacosadiene in a backcross between *D. simulans* and *D. sechellia*. We find that *desat2* does not affect variation of 7-tricosene, even though this locus was previously implicated as affecting the same trait in *D. melanogaster*. *desat1* has a strong affect on the interspecific variation of a saturated hydrocarbon chain compound (Unbranched-23). The candidate gene, *desatF* potentially exerts an influence on the variation of 7,11-heptacosadiene through a large epistatic effect with unidentified loci, situated between the markers *pros* and *Mtn*. The candidate gene *eloF* is situated in this region, and is known to affect the elongation of unsaturated hydrocarbon chains. The QTL associated with the marker *desatF* influenced the variation of both diene compounds (7,11-HD and 7,11-PD), and intriguingly epistasis was only detected for the variation of these diene compounds. This highlights the potential involvement of two separate loci encoding for different enzymes important in female specific cuticular hydrocarbon synthesis. There was an extremely significant region on the X-chromosome situated near the marker *forked* involved in the interspecific variation of

the alkane linear compound (Unbranched-23) and all of the monoene compounds. This is intruiging since there are no known candidate genes affecting CHCs associated with this marker.

2.2 Introduction

There is considerable interest in identifying genes which contribute to adaptive divergence and reproductive isolation between species as these are potential "speciation genes". A common cause of reproductive isolation in animals is sexual isolation but there are few known behavioural genes which influence speciation (NOOR 2003; ORR *et al.* 2004). One promising area in which to find genes of large effect on behaviour are pheromones, where single loci can cause large functional changes in important mating signals (GROOT *et al.* 2006; ROELOFS and ROONEY 2003).

The cuticular hydrocarbons (CHCs) of *Drosophila* are long chain fatty acids found on the cuticle surface (BLOMQUIST *et al.* 1987). Insects have developed an efficient system for the biosynthesis of pheromones, involving tissue-specific modified enzymes, which alter existing products of the original metabolism into pheromone compounds (TILLMAN *et al.* 1999). The synthesis of saturated and mono-unsaturated CHCs takes place on the first day of emergence, the female dienes (hydrocarbon chains containing two double bonds) appear after the first day and their levels become constant at about three days (TILLMAN *et al.* 1999). CHCs show geographic variation consistent with environmental selection, and influence desiccation, cold-tolerance and starvation resistance (ETGES and JACKSON 2001; GREENBERG *et al.* 2003; ROUAULT *et al.* 2000). In addition, some function as contact pheromones (COBB and JALLON 1990;

ETGES and AHRENS 2001; FERVEUR 2005). Directly "swapping" CHCs between species alters sexual attractiveness, for example, males of *D simulans* will normally ignore females of *D sechellia* but will court them vigorously if they have been coated with CHCs of *D simulans* females (COYNE *et al.* 1994).

Most species of the melanogaster group of *Drosophila* such as *D. simulans* and *D. mauritiana* are sexually monomorphic for CHCs with high levels of monoenes, usually 7-tricosene (7-T). *D. sechellia* (and cosmopolitan *D. melanogaster*) is sexually dimorphic with dienes in females, usually 7,11-heptacosadiene (7,11-HD) (COBB and JALLON 1990; JALLON and DAVID 1987). The main pheromone effect is that males of monomorphic species do not court females with the wrong compounds, so the most important effect on sexual isolation between *D. simulans* and *D. sechellia* is the change between 7-T in *D. simulans* females, to 7,11-HD in *D. sechellia* females. These compounds also vary geographically in *D. melanogaster*. Cosmopolitan females have 7,11-HD and African females 5,9-heptacosadiene, and these have been implicated in the well-characterised assortative mating seen between these races of flies (FANG *et al.* 2002), though this difference is unlikely to be the sole cause of this (COLEGRAVE *et al.* 2000; COYNE and ELWYN 2006).

A wide range of genes can influence the components of CHCs, and several have been identified which are potential candidate genes for species-specific variation (DALLERAC *et al.* 2000; JALLON and WICKER-THOMAS 2003). The enzyme activity of a number of desaturases are important for influencing the composition of long chain fatty acids, by introducing a double bond at specific sites along the hydrocarbon chain. This also causes an alteration of the pheromone blend (CHAN YONG and

JALLON 1986; ROELOFS and ROONEY 2003). The particular loci that influence the differences in CHCs within and between *Drosophila* species is unclear and a matter of some debate. When expressed in yeast, *desat2* was shown to insert a double bond at position five along the hydrocarbon chain (DALLERAC *et al.* 2000), and co-segregation analyses implicated *desat2* in the difference between 7,11-HD and 5,9-HD in Cosmopolitan and African *D. melanogaster* females (COYNE *et al.* 1999; TAKAHASHI *et al.* 2001). Sequence analysis implied that a non-functional allele (caused by a small deletion in the promoter region), had replaced the functional African one in Cosmopolitan strains. The non-functional allele results in increased levels of 7,11 HD. Greenberg *et al.* (2003) confirmed this by carrying out a precise targeted gene replacement study, replacing a cosmopolitan allele with an African one against a cosmopolitan background. These flies showed African-like CHC blends (with higher ratios of 5,9-HD/7,11-HD). Furthermore, they showed less cold-tolerance and greater starvation resistance, like African flies (but were not more desiccation resistant). However, Coyne and Elwyn (2006) failed to replicate some changes seen in these transgenic flies, so there is uncertainty about the precise contribution of *desat2* to these ecological adaptations in *D. melanogaster*. There may be variable expression in the transgenic flies (their sex-specific expression is altered), or there may be problems with standardising the genetic background (GREENBERG *et al.* 2006).

Another important fatty acid desaturase is encoded by the gene *desat1*, which in *D. melanogaster* is situated just 3.7 kb downstream of *desat2*. It is expressed in the same putative transmembrane region, but has different substrate specificity to that of *desat2*, and when expressed in yeast was shown to be responsible for the first desaturation step, leading to hydrocarbons bearing a double bond at the seventh

position (LABEUR *et al.* 2002). Induced mutations in *desat1* show reduced levels of all unsaturated CHCs (including 7,11-HD and 7-T) but increased levels of saturated hydrocarbons. Interestingly, one mutation of *desat1* also influences female perception of the pheromones (MARCILLAC *et al.* 2005). A clear relationship was found between *desat1* and a higher male ratio of 7-tricosene to 7-pentacosene which also caused increased mating efficiency (MARCILLAC *et al.* 2005) . It has been hypothesised that a higher level of 7-T is an indication of good genetic and physiological fitness. It has also been found that males experiencing high environmental stress have lower levels of 7-T (COBB and FERVEUR 1996; SAVARIT and FERVEUR 2002; SUREAU and FERVEUR 1999).

There is intraspecific geographic variation of the ratio between 7-T and 7-P in *D. simulans* and *D. melanogaster*. In *D. simulans* most strains produce high levels of 7-T except for within the Benin Gulf region in Africa. In *D. melanogaster* 7-T is gradually replaced by 7-P in strains nearing equatorial regions, and it has been suggested that this variation in the ratio of 7-T to 7-P may be due to climatic selection (ROUAULT *et al.* 2000). Chertemps *et al*. (2006) have recently characterised one desaturase locus, *desatF* (syn. *fad2*), which is expressed in females. RNAi knock-down of this gene reduces diene production and increases monoenes. Males with 'feminised' hydrocarbons show expression of *desatF* and aberrant courtship behaviour. *desatF* is also expressed in *D. sechellia* females (but not *D. simulans*), so it is a candidate gene for increased levels of 7,11-HD.

A number of the sex determination genes are also implicated as having an affect on CHCs. The *doublesex* gene (*dsx*) controls somatic sexual differentiation and is also a

candidate gene for CHCs as it influences the production of female dienes in *D. melanogaster* (BURTIS and BAKER 1989; JALLON *et al.* 1988). Both *transforme*r and *transformer-2* are responsible for the sex-specific splicing of *doublesex* and *fruitless*, and their expression is involved with the feminisation of pheromone compounds (SAVARIT *et al.* 1999).

Gleason *et al*. (2005) used a Quantitative Trait Locus (QTL) approach to examine differences in the amount of 7-T and 7,11-HD in females of *D. simulans* and *D. sechellia*. Composite Interval Mapping (CIM) found at least four QTLs for the monoene 7-T, three on the third chromosome and one on the X, and two QTLs for the diene 7,11-HD on chromosome three. When examining known candidate genes (Table 6 in (GLEASON *et al.* 2005) the only ones close to the QTLs were on chromosome three. These included *desat1* & *desat2* for 7-T and *desatF* for 7,11-HD. These did not map precisely to the QTL peaks and were described as being "on the edges". In addition, one gene involved in the sex determination pathway, *doublesex*, was on the edge of a QTL for 7-T.

Here we have developed markers within the loci *desat1*, *desat2*, *desatF* and the sex determination loci *doublesex (dsx)* and *fruitless (fru)*. We score these in the recombinant flies from the cross produced by Gleason *et al*. (2005), and assess the affects of these additional candidate markers on the six CHCs present within the cuticular hydrocarbon profiles of both species, by producing refined CIM mapping including these candidate genes, and test for epistasis with Multiple Interval Mapping.

2.3 Materials and methods

In the original experiment (GLEASON *et al.* 2005) fly culturing was standard, and the flies were kept at 25ºC and followed a 12hr light and 12hr dark cycle. A backcross experimental design had to be followed due to the male sterility that occurs within the F1 generation. Two strains were used one for each species, Jean R. David provided the pure bred strain of *D. sechellia* (inbred for 18 generations of brother-sister mating) and Jerry Coyne provided the *D. simulans* line containing five morphological markers (*f²*; *nt*; *pm*; *st*, *e*), one for each chromosome arm. The female *D. simulans* were crossed to male *D. sechellia* flies, and the resulting hybrid females were backcrossed to male *D. simulans*. The five morphological markers allowed an approximate equal selection of the 32 backcross genotypes. The cuticular hydrocarbon chains assessed in this study included the following compounds: 7-tricosene, 7,11-heptacosadiene, 23-Unbranched, 7,11-pentacosadiene, 7-pentacosene and 7-hexacosene. These were extracted from the resultant 487 recombinant females with hexane and the quantity of all six CHC compounds was calculated following gas chromatography. Hexacosane was used as a standard and all values adjusted, then log transformed (GLEASON *et al.* 2005). For the current study a further five candidate genes were scored and incorporated into the original map. The genomic sequences for each respective marker were obtained from FlyBase (http://flybase.bio.indiana.edu/). The *D. simulans* and *D. sechellia* sequences were aligned using ClustalW2 (http://ebi.ac.uk/Tools/clustalw2). The forward and reverse primers were designed in conserved regions flanking a target sequence where there was a distinguishable difference in sequence size, or enough sequence variation to allow the use of restriction enzymes to distinguish the alleles. These were found using RestrictionMapper3 (http://restrictionmapper.org). The Primer3 (Version 0.0.4, http://frodo.wit.mit.edu) program was used for the final

primer design. The markers were all PCR amplified for each of the remaining individual DNA samples, the PCR conditions can be seen in the supplementary material in the appendix section (Table A2.3). The different alleles were identified from different sized fragments for *D. sechellia* and *D. simulans* using 2% agarose. The details of the primers and restriction enzymes used can be seen in the appendix section (Tables A2.1 and A2.2)

A genetic linkage map was assembled from all data using MAPMAKER (LANDER *et al.* 1987). The new markers mapped in their expected genomic order, though map lengths resulting from this cross are unusually large compared to *D. melanogaster*. This may be caused by the deliberate selection of recombinant individuals, higher recombination in *D. simulans* than *D. melanogaster* and epistasis in viability interactions between the X chromosome of *D. sechellia* and the autosomes of *D. simulans* (GLEASON *et al.* 2005; JOLY 1997). The map obtained was subsequently imported into QTL Cartographer version 2.5 (BASTEN *et al.* 1997) and composite interval mapping carried out to identify QTL for all 6 of the CHC traits. CIM (JANSEN and STAM 1994; ZENG 1994) combines interval mapping with multiple regression and tests for a QTL within a region flanked by two markers, whilst simultaneously fitting partial regression coefficients for background markers. CIM settings used the Kosambi map function, a walking speed of 2 cM, window size of 5 cM and forward/backwards regression on all background markers. CIM uses the likelihood ratio (LR) as a test statistic calculated from the formula: $-2(\ln H_0 - \ln H_A)$. The null hypotheses (H_0), assumes a normal distribution and no linkage between the analysis point and the trait variation. The alternative hypothesis assumes a QTL is present and linked to a marker or interval between markers (CHURCHILL and DOERGE 1994). A

high LR indicates a strong association between the analysis point and the trait variation. The experiment-wide significance level was set at 0.05, calculated from 1000 permutations of the trait data among marker classes. The permutation test randomly reassigns trait values to different individual within the data set, and therefore measures the level of chance involved for the occurrence of a significant QTL (CHURCHILL and DOERGE 1994). We also carried out Multiple Interval Mapping since this allows testing for epistasis. MIM incorporates a measure of the number, position, and interactions between all the QTLs affecting a specific phenotypic trait (BASTEN *et al.* 1997). Due to the backcross design, only heterozygous effects could be detected (and additive effects fitted).

2.4 Results

This results section firstly focuses on the two main CHCs associated with sexual isolation between *D. simulans* and *D. sechellia*. This is essentially a follow up from the original study (GLEASON *et al.* 2005), with addition of novel candidate genes as markers and the use of MIM analysis (to test for epistasis). Figure 1 depicts the CIM LOD plot for chromosome 3 for 7,11-HD and 7-T. This is similar to the previous study, identifying a region influencing both traits on the right arm of the chromsome and a smaller peak for 7,11-HD on the left arm. *desat1* and *desat2* now lie outside of the QTL peak on the right arm. *fru* was not implicated, however there is a substantial peak over the *dsx* locus (slightly below the significance threshold), indicating that it may affect the interspecific variation of 7-T.

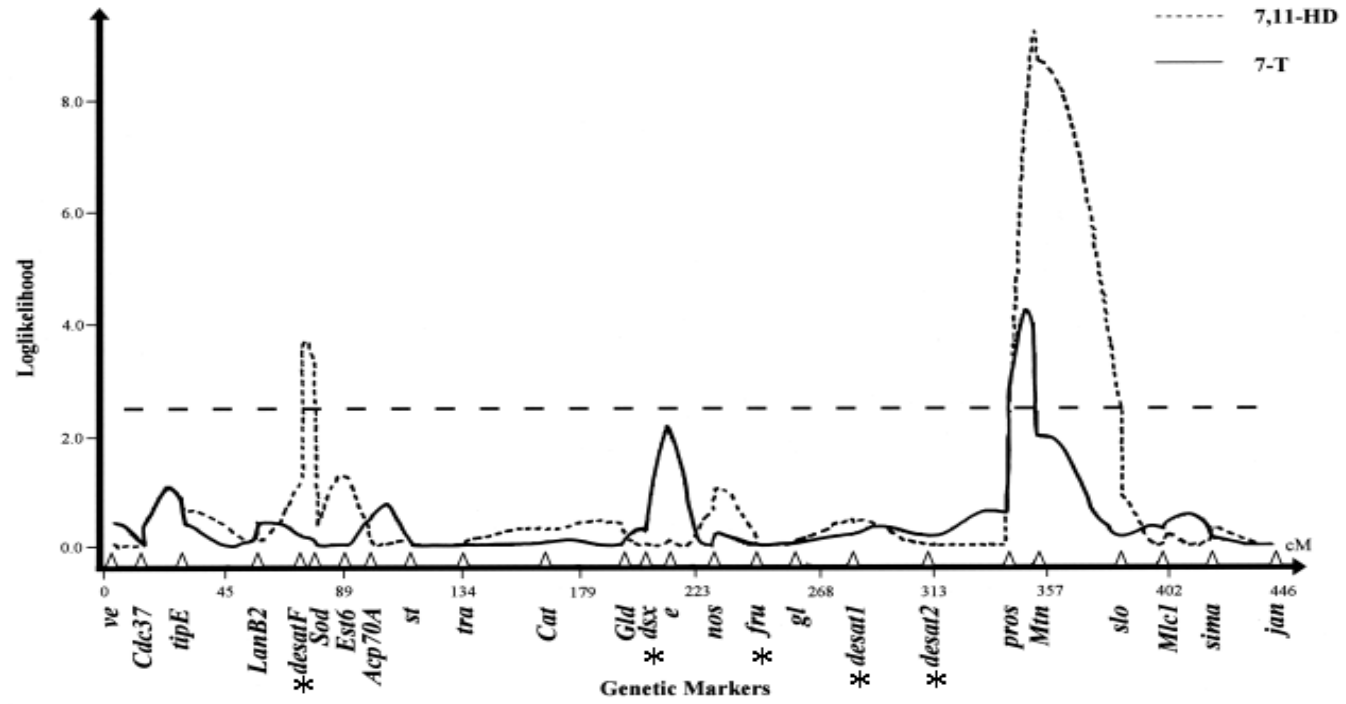


Figure 1. QTL plot of chromosome 3 from composite interval mapping for 7-T and 7,11-HD. The location of the marker loci is indicated on the X axis and the dashed vertical line is set at a confidence threshold P <0.05 (LOD = 2.4). The new marker loci are indicated with an asterix.

Table 1. Composite Interval Map (CIM) and Multiple Interval Map (MIM) results for 7-T and 7,11-HD. Showing the estimates of the positions and effects for all significant QTLs. CIM LOD threshold of 2.4, corresponds to a confidence value of P <0.05 and MIM LOD threshold of 0, corresponds to a confidence value of P <0.05 (there were no epistatic effects for 7-T).

Analysis	Pheromone	QTL	Chrom	Position (Cm)	Nearest marker	LOD score	Additive effect	% Genotypic effect
CIM	**7-T**	1	1	128.85	*forked*	6.88	0.268	31.14
		2	1	137	*forked*	6.1	0.1808	14.08
		3	3	349.95	*pros*	4.25	0.1838	14.55
	7,11-HD	1	3	75	*desatF*	3.62	-0.0405	16.34
		2	3	353.95	*Mtn*	9.19	-0.0511	26.01
		3	3	357.24	*Mtn*	8.76	-0.0563	31.57
MIM	**7-T**	1	1	132.85	*forked*	2.25	0.5406	34.9
		2	3	103.8	*Acp70A*	2.27	0.1125	3.6
		3	3	348.95	*pros*	8.56	0.2363	16
	7,11-HD	1	2	25.71	*odd*	3.59	-0.021	1.7
		2	3	74	*desatF*	27.41	-0.077	25.1
		3	3	171.61	*cat*	1.75	-0.018	3.9
		4	3	356.24	*Mtn*	29.24	-0.075	23
Epistasis	**7,11-HD**	1x2				2.99	0.038	1.8
		2x3				3	0.049	1.5
		1x4				3.12	0.038	1.7
		2x4				13.58	0.124	18.4
		3x4				2.44	0.044	3.1

Table 1 shows the position, affect size and interactions of significant QTLs detected using CIM and MIM. The position of each QTL is given as well as the nearest marker. *desatF* is clearly implicated as contributing to the QTL peak on the left arm of chromosome three for 7,11-HD. The CIM has detected multiple QTL associated with the markers *forked* and *Mtn*. MIM analysis shows that for 7,11-HD, QTL 2 (*desatF*), exerts its influence largely through an epistatic interaction with the peak on the right arm (QTL 4: *Mtn*) such that large amounts of 7,11-HD are produced only when both regions include alleles from *D. sechellia*. The two loci *forked* and *Mtn*, which are associated with multiple QTL in the CIM, are associated with just one large QTL effect in the MIM analysis.

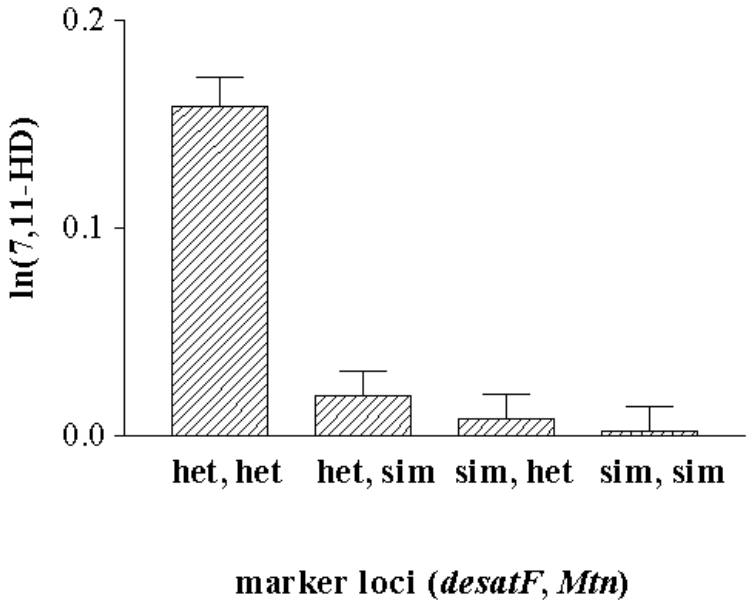


Figure 2. Epistasis plot showing the interaction between the allelic state (heterozygote sechellia/simulans or homozygote simulans) at the *desatF* and *Mtn* loci and the amount of 7,11-HD present in recombinant females. Error bars are 95% confidence intervals.

The presence of a *D. sechellia* allele at both the *desatF* and *Mtn* loci results in a large increase in the amount of 7,11-HD (Figure 2). This indicates that the production of 7,11-HD depends on an interaction between both QTL associated with these markers.

Table 2 shows the novel CIM and MIM results of all the significant QTLs ($P<0.05$) detected for 4 additional cuticular hydrocarbon chain compounds, which have been included in this study (23-Unbranched, 7,11-pentacosadiene, 7-pentacosene and 7-hexacosene). These have been included separately since they are of less importance to the sexual isolation between *D. simulans* and *D. sechellia*. However their inclusion allows an assessment on the association of particular markers with the two types of unsaturated CHCs (monoenes and dienes), and their potential interaction.

Table 2. Composite Interval Map (CIM) and Multiple Interval Map (MIM) results for 23-UB, 7-P, 7-H and 7,11-PD. Showing the estimates of the positions and effects for all significant QTLs. (CIM LOD threshold of 2.4, corresponds to a confidence value of P <0.05 and MIM LOD threshold of 0, corresponds to a confidence value of P <0.05).

Analysis	Pheromone	QTL	Chrom	Position (Cm)	Nearest marker	LOD score	Additive effect	% Genotypic effect
CIM	**23-UB**	1	1	130.85	*forked*	6.1	0.0831	17.1
		2	1	139	*forked*	30.8	0.1013	25.4
		3	2	128.49	*Su*	3.2	-0.0326	2.6
		4	3	105.8	*Acp70A*	2.8516	0.0308	2.3
		5	3	210.76	*e*	7.3269	0.0554	7.6
		6	3	283.9	*desat1*	13.642	0.0763	14
		7	3	288.09	*desat1*	13.938	0.0822	16
	7-P	1	1	128.85	*forked*	19.197	0.1761	9
		2	3	340.99	*pros*	9.046	-0.1981	11.4
		3	3	349.95	*pros*	58.289	-0.3473	35
		4	3	365.24	*Mtn*	14.082	-0.3843	43
	7-H	1	1	130.85	*forked*	6.794	0.0249	12.2
		2	3	336.99	*pros*	3.806	-0.0321	20.2
		3	3	349.95	*pros*	25.953	-0.0511	51.3
		4	3	357.24	*Mtn*	2.878	-0.0262	13.5
	7,11-PD	1	2	33.53	*odd*	6.097	-0.059	1.97
		2	3	24.25	*tipE*	4.108	-0.0565	1.8
		3	3	35.34	*tipE*	4.276	-0.0678	2.6
		4	3	68.64	*desatF*	13.747	-0.2433	33.6
		5	3	75	*desatF*	12.398	-0.1659	15.6
		6	3	84.67	*Est6*	7.767	-0.2442	33.9
		7	3	91.91	*Est6*	4.8	-0.0939	5
MIM	**23-UB**	1	1	138.003	*forked*	15.56	0.1012	19.5
		2	2	124.49	*Su*	2.18	-0.0352	1.4
		3	3	214.04	*e*	2.04	0.0456	8.3
		4	3	347.95	*pros*	5.58	0.0611	10
	7-P	1	1	131.854	*forked*	4.08	0.1761	9
		2	1	136.003	*forked*	2.34	-0.253	-15.8
		3	3	350.955	*Mtn*	3.27	-0.4238	59
	7-H	1	1	135.003	*forked*	2.83	0.0183	4.2
		2	3	353.955	*Mtn*	1.73	-0.056	25.2
	7,11-PD	1	2	37.53	*odd*	9.42	-0.0801	3.3
		2	3	68.64	*desatF*	16.53	-0.2995	62.8
		3	3	344.955	*pros*	8.57	0.2829	19
Epistasis	**7,11-PD**	1x2				8.598	0.1505	4.2

Table 3. A summary of the 12 markers most strongly associated with the QTL positions affecting the trait variation of all 6 cuticular hydrocarbon compounds. The black shaded areas indicate the markers associated with QTL affecting each of the CHCs included in this study.

	forked	*odd*	*Su*	*tipE*	*desatF*	*Est6*	*Acp70A*	*Cat*	*e*	*desat1*	*pros*	*Mtn*
23-UB (Saturated)	■		■				■		■	■	■	
7-T (Monoene)	■						■				■	
7-P (Monoene)	■										■	■
7-H (Monoene)	■										■	■
7,11-HD (Diene)		■			■			■				■
7,11-PD (Diene)		■		■	■	■		■			■	

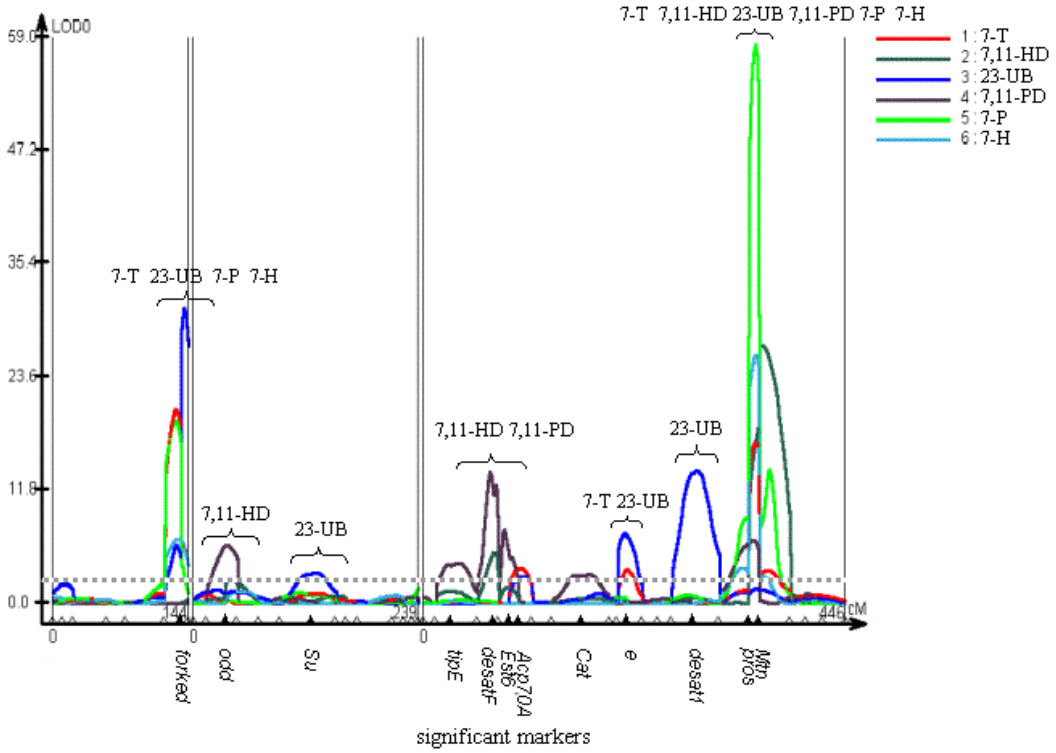


Figure 3. CIM QTL plot showing an overview of all the QTL positions and associated markers affecting all 6 of the cuticular hydrocarbon chain compounds, on all the chromosomes (the double vertical lines separate each chromosome, X, 2nd and 3rd). The location of the markers associated with significant QTL, are indicated on the X-axis. The dashed horizontal line is set at a confidence threshold P <0.05 (LOD = 2.4 on the Y-axis).

Table 3 and figure 3 show a summary of the chromosomal regions affecting each one of the six CHC compounds included in this study. The two CHCs (7-T and 7,11-HD) shown in the initial CIM plot have also been included in this second CIM plot to show the relationship that exists between specific chromosomal regions and different types

of CHCs. The chromosomal region situated between the markers *pros* and *Mtn* appears to affect all CHCs, whereas the influential region on the X-chromosome nearest the marker *forked* affects monoenes and the saturated CHC (23-Unbranched). The chromosomal regions affecting dienes are exclusively associated with only these compounds, which are most strongly associated with the markers *desatF*, *odd* and *Cat*.

2.5 Discussion

In total 12 markers were associated with interspecific variation of the hydrocarbon compounds present in *D. sechellia* and *D. simulans* (Table 3). A number of QTLs are associated with more than one trait. Figure 3 shows that specific chromosomal regions are influential on particular types of hydrocarbon compounds. The significant QTL on the X-chromosome situated nearest to the marker *forked* (15F4) affects the variation in all of the hydrocarbon monoenes as well as the linear alkane 23-Unbranched. The significant region on the left arm of the third chromosome situated between the markers *desatF* (68A1) and *Est6* (69A1) appears to influence diene production, and may be involved with the second desaturation step, since flies carrying deficiencies within this chromosomal region produce lower levels of female dienes (WICKER-THOMAS and JALLON 2000). *desatF* is thought to be responsible for the female specific second desaturation leading to diene production (CHERTEMPS *et al.* 2006). Epistatic interactions were found for the diene pheromone compounds, 7,11-HD and 7,11-PD. Both epistatic interactions involved the QTL on the left arm of the third chromosome associated with the marker *desatF*. The production of 7,11-PD was dependent on this specific region, and involved an epistatic interaction with a region on the second chromosome near the marker *odd*. There is also a very large epistatic

interaction between two QTL contributing to the variation of 7,11-HD, involving the QTL nearest to the marker *desatF* on the left arm, with the marker *Mtn* situated on the right arm of the third chromosome. The markers *desatF*, *Est6*, *Mtn, odd* and *tipE* all appear to be linked to QTLs which are influential on diene production, and they all show negative additive affects. Thus *D. simulans* homozygous for these particular markers produce lower levels of dienes. The significant peak associated with *tipE* may implicate the involvement of the near by candidate gene *ecdysonless* (*ecd*). This gene is responsible for the production of ecdysteroids, which are produced at the cuticle surface and linked with the biosynthesis of female specific pheromones (dienes). Mutations at this specific loci cause a decrease in diene production and an increase in the level of monoenes (GLEASON *et al.* 2005; WICKER and JALLON 1995)

The epistatic interaction between the markers *desatF* and *Mtn* shown in Figure 2 indicates that both QTL may be involved in different stages of diene production and account for a large amount of the interspecific variation of 7,11-HD. The chromosome region situated on the right arm of the third chromosome between the markers *pros* and *Mtn* appears to be influential on variation in all of the hydrocarbon compounds. Many enzymes, including reductases, carboxylases and elongases are involved in the production of long chain fatty acids, and recently Chertemps *et al*. (2007) characterised an elongase (*eloF*), expressed in female *D. melanogaster* and *D. sechellia. eloF* is believed to play a role in electron transfer during the redox reactions involved in the elongation process (SHANKLIN *et al.* 1994). The expression of this gene is important for the elongation of hydrocarbon chains. RNAi knockdown of *eloF* in *D. melanogaster* resulted in an alteration of the CHC profiles, females had significantly less C29 dienic hydrocarbons and increased C25 dienes. The longer

monoene CHCs were also reduced showing a decrease in C27 monoenes and an increase in C23 monoenes (CHERTEMPS *et al.* 2007). Its cytological position in *D. melanogaster* is 85E10, close to *Mtn*. The elongation process is important in the formation of all CHCs, but appears to have a stronger influence on unsaturated CHCs. Experimental evidence has also indicated that *eloF* expression is under control of the sex determination hierarchy. *transformer* expression was targeted in *D. melanogaster* by using the *ok72-GAL* driver, which resulted in the feminisation of hydrocarbon profiles, increasing diene production in males. *eloF* is not expressed in the monomorphic species *D. simulans*, and is responsible for the elongation process in both monoene and diene CHCs present in the dimorphic species (CHERTEMPS *et al.* 2007). However there are also positive additive affects found for the production of the diene compounds within the region between *pros* and *Mtn*, and it appears that a second important locus is situated within this region. This second QTL may be involved in a more general role with the production of cuticular hydrocarbon chains. A QTL situated near the marker *desat1* has a significant affect on the levels of the linear alkane, 23-Unbranched. The marker *desat2* is linked with variation of both 7,11-HD and 5,9-HD in *D. melanogaster* (COYNE *et al.* 1999; TAKAHASHI *et al.* 2001), but is not implicated as having an influence on any of the hydrocarbon compounds differing between *D. simulans* and *D. sechellia*. CHCs with a desaturation at the five carbon position are not involved in this study, and therefore *desat2* may not be influential on the variation of 7,11-HD since it is not competing for the cuticular hydrocarbon substrate that contribute to the pheromone profiles of these sibling species.

In the original study of Gleason *et al*. (2005) *desatF* was described as being close to the left arm QTL for 7,11-HD, and this study implicates this gene further, confirming this locus as an ideal candidate for additional study. There are relatively few studies that have succeeded in moving from QTL peaks to underlying loci for genes of large effect on ecologically important traits e.g. (GRATTEN *et al.* 2007), but confirming the role of *desatF* requires additional work, and additional, possibly linked loci may also affect the trait.

The MIM analysis reveals that the potential effect of *desatF* is largely through an epistatic interaction involving unidentified loci on the right arm. We know that regulatory changes must be important because 7,11-HD is both sex and species-specific in expression. The sex determination loci *fru, tra* and *dsx* seem not to be of major importance here so the sex-specificity must lie further down the sex determination cascade. *desatF* is not expressed in *D. simulans* (CHERTEMPS *et al.* 2006) so an epistatic interaction where two regions must be heterozygous for high levels of 7,11-HD may indicate that the region on the right arm of the third chromosome situated near *Mtn* is involved in modulating expression of *desatF*. Alternatively it may indicate that two interacting enzymes are necessary to produce 7,11-HD.

It appears that the production of 7,11-HD within the females of the dimorphic species relies on the composite interaction of a number of enzymes. These results provide a cautious counter-example to the "candidate gene" approach to studying the genetics of complex traits in behavioural biology and highlight the importance of the species-specific background in the functional role of traits involving composite genetic

interactions. Candidate genes are genes identified to influence a trait in a model organism, and several have been found to also contribute to variation in apparently homologous behaviours in a range of other species, contributing to the promotion of the candidate gene approach (FITZPATRICK *et al.* 2005). Examples of spectacularly successful use of candidate genes for ecologically important behavioural variation include the *foraging* gene of *D. melanogaster*, which influences larval feeding strategies and has been shown to influence similar behaviours in species as phylogenetically remote as bees and ants (BEN-SHAHAR *et al.* 2002; INGRAM *et al.* 2005). *desat2* and *desat1* have been shown to influence the relative amount of 7,11-HD and 5,9-HD in *D. melanogaster*, (DALLERAC *et al.* 2000; GREENBERG *et al.* 2003) yet here we find that one of these traits in a sibling species is not influenced by these but by linked loci. If we had only scored these candidate genes we would have been given a false impression of their likely role (in single marker regression both loci covaried significantly with 7,11-HD, with P values < 0.001). These potential problems seem especially relevant when examining gene families or interacting loci. Confusion about the influence of the genetic background in studies of *desat2* in *D. melanogaster* (COYNE and ELWYN 2006; GREENBERG *et al.* 2006) suggests that even precise gene replacement might not conclusively identify the role of individual loci. The fact that *desatF* shows a strong epistatic effect indicates how the genetic background into which a candidate gene is introduced in a manipulation experiment may have a very large influence on the magnitude of the effect seen. However, this study has detected a very significant epistatic interaction between both QTL involved in the interspecific variation of 7,11-HD, and indicates that the candidate genes *desatF* and possibly *eloF* may be involved in the sexual isolation of these sibling species.

CHAPTER 3.

Quantative Trait Locus analysis of two sexually dimorphic traits between *Drosophila simulans* and *D. sechellia*: the mean interpulse interval of courtship song, and the prothoracic leg sex comb tooth number.

3.1 **Abstract**

The sex determination genes *fruitless* (*fru*) and *doublesex* (*dsx*) were incorporated into a study of the Quantitative Trait Loci (QTL) for a courtship song difference between *Drosophila simulans* and *D. sechellia* (GLEASON and RITCHIE 2004), in which a backcross analysis was carried out using a total of 45 markers spread evenly across the genome. Quantitative traits are derived from the accumulation of several loci that have an affect on a specific phenotype. The analysis of such traits depends on the use of molecular as well as morphological markers which through recombination can indicate the regions of the chromosome that influence the trait variation. The original study discovered six QTL that explained 40.7% of the phenotypic variation. The QTL coincided with three candidate genes, *maleless, fruitless* and *croaker*. These genes had been implicated through previous mutational analysis to have an affect on courtship song. *maleless* and *fruitless* in particular were both directly associated with the specific trait mean interpulse interval. A revised Composite Interval Mapping indicated *nanos (nos)* and *fru* as the most significant genes, with *nos* scoring a slightly higher additive effect than *fru*. The CIM also detected a significant QTL situated on the X-chromosome near the marker *forked*. Multiple Interval Mapping analysis was also carried out (a more advanced option for QTL detection) and found a specific position situated at 226.308 cM. *fru* was the closest marker associated with this significant QTL on the third chromosome. MIM also detected a significant QTL associated with the marker *dgα* situated on the second chromosome. Significant

epistatic interactions were detected between a further QTL situated nearest the marker *forked* on the X-chromosome with both of the other significant QTL situated on the third and second chromosomes.

The sex determination genes; *fru*, *dsx*, *tra* as well as the five morphological markers f^2 (*forked*), *nt* (*net*), *st* (*scarlet*) and *e* (*ebony*) were also scored in a number of recombinant inbred (RI) lines. These lines contain a mosaic of the *Drosophila simulans* and *D. sechellia* parental genomes. The RI lines were predominantly *D. simulans* in background, however a number of RI lines proved positive for the presence of homozygous *D. sechellia* alleles at one or more of the sex determination loci. General Linear Model analysis was used to test for their covariance with mean interpulse interval (IPI) and sex comb tooth number. The presence of *D. sechellia* homozygotes at the *st* and *fru* loci caused a significant increase in mean IPI. *dsx* was also associated with a significant influence on trait variation but had a negative affect. In the analysis of sex comb tooth variation, it appears that all RI lines homozygous for *D. sechellia* alleles at the sex determination loci had significantly higher numbers of sex comb teeth. The chromosomal region associated with the marker *fru* appeared to exert the greatest influence.

3.2 Introduction

Traits influencing aspects of reproduction can have a significant effect on the speciation process. The traits involved in male reproductive functions show a higher rate of divergence between species, compared to non-sexual traits (CIVETTA and SINGH 1998). This chapter focuses on two sexually dimorphic trait differences between *D. simulans* and *D. sechellia*; sex comb tooth number and the mean interpulse interval (IPI) of courtship song.

3.2.1 Sex comb

The sex comb present on the prothoracic legs of the male (Figure 1) is used to hold onto the female during mounting (COOK 1977; COYNE 1985). Experiments involving the removal of the sex comb severely affect the males' ability to inseminate females (COOK 1977; COYNE 1985; SPEITH 1952). The two main genes implicated in sex comb variation are *Sex combs reduced* (*Scr*) and *doublesex* (*dsx*). *Scr* is primarily involved in the leg segmentation development (PATTATUCCI and KAUFMAN 1991) and *dsx* determines male sexual morphology (BAKER and RIDGE 1980). An interspecific study carried out between *D. simulans* and *D. mauritiana* on the trait variation of sex comb teeth found significant affects associated with a region on each chromosome arm (COYNE 1985). A refined analysis found two QTL on the third chromosome, one situated between *tra* (cytological position on the *D. melanogaster* map 73A), and *Antp* (84BD), contributing to 53.6% of the trait variation (TRUE *et al.* 1997). The second QTL was

Figure 1. *Drosophila melanogaster* sex comb (GRAZE *et al.* 2007).

situated between *Ald* (97B) and *janA* (99D). A further two weaker effect QTLs were found on the X chromosome. The larger QTL was also linked to one affecting the number of anal plate bristles, clasper bristles as well as the number of bristles on the 5th sternite (GRAZE *et al.* 2007). Sex comb teeth are modified bristles and it has been suggested that they are linked to the bristle formation pathway (MACKAY 1995). The most recent study carried out a further QTL analysis focussing on the significant region situated between 73A (*tra*) and 84AB (*Antp*). The study found at least two QTL either side of the third chromosome centromere, associated with 77B and 83B-84B (GRAZE *et al.* 2007). Fine scale mapping was then applied to the region in order to eliminate positive association of markers due to linkage disequilibrium. The study used *D. mauritiana* introgression lines marked with P-element {lacW} 76C, these lines were then crossed to *D. simulans* lines with visible mutations, this created a set of lines with recombination breakpoints between P {lacW} 76C and the right flanking marker *ebony* (93C7). The results indicated that the single large affect QTL situated on the third chromosome found in the previous study (TRUE *et al.* 1997), is possibly due to the composite affect caused by a linked complex of several QTLs (GRAZE *et al.* 2007). Two large positive main affect QTLs were found at 84A (*Scr*) and 93C7 (*ebony*) and a further two QTLs with transgressive effects were also found in the regions 76C-79D and 80A-84A. The transgressive QTLs cause negative effects with respect to the *D. mauritiana* phenotype. The negative effect of the QTL associated with the marker *Dip2* is in part due to an epistatic interaction, since *D. mauritiana* alleles present at both markers *CG11367* (80A) and *Dip2* (82C) cause a reduction in this negative affect. It has been postulated that the apparent negative QTL affect depends on the *D. simulans* or *D. mauritiana* genetic backgrounds (GRAZE *et al.* 2007).

Previous evidence has shown that sexually dimorphic traits, particularly relating to male fertility, have a faster rate of divergence than other genes and depend strongly on background epistatic interactions (MEIKLEJOHN *et al.* 2003; NUZHDIN *et al.* 2004; ORR and IRVING 2001; ZHANG *et al.* 2004).

Despite both *dsx* and *Scr* being identified as major candidate genes within *D. melanogaster* involved in sex comb function, there was no evidence of divergence of their expression in the prothoracic and mesothoracic legs in *D. simulans* and *D. mauritiana*. This is understandable since many genes are involved in the complex developmental pathways of sex comb morphology and the trait has evolved a great deal between *D. melanogaster* (GRAZE *et al.* 2007). However, the three genes associated with significant QTL showing strong interspecific expression differences in male prothoracic and mesothoracic legs were *CG2791*, *CG15186* and *CG2016*. Not a lot is known of the functional role of these genes, except that *CG2016* is a member of the *takeout* gene family and a number of genes within this family are expressed in a sex-specific manner. However analysis has indicated that *CG2016* is not expressed in a sex-specific manner (DAUWALDER *et al.* 2002). It is possible that the candidate genes *Scr* and *dsx* exert an influence on interspecific difference through variation in coding sequence, or regulatory differences, that may affect expression at earlier stages in the development of this particular trait. It has also been speculated that the interspecific difference in sex comb may be due to the evolution of these modifier genes rather than major regulatory genes (GRAZE *et al.* 2007).

The apparent bidirectional QTL effects for sex comb variation between *D. mauritiana* and *D. simulans* agree with the findings of a quantitative genetic analysis of male

sexual trait differences between *D. simulans* and *D. sechellia*. The study included the following traits; sex comb tooth number, posterior lobe of the genital arch, hybrid male sterility, sperm and testis size, and of all these only sex comb tooth number showed evidence for QTL acting in both directions (MACDONALD and GOLDSTEIN 1999). It is also possible that sex comb tooth number variation may be a by-product of selection operating on other traits controlled by pleiotropic genes. It is debatable as to how much the sex comb morphology depends on specificity to the female and it has been suggested that the number of sex comb teeth is not important, since this trait also exhibits large environmental variation (MACDONALD and GOLDSTEIN 1999). However more recent evidence indicates that the trait may well have co-evolved between males and females of different species, and the male's ability to grasp the female may depend on the size and shape of the sex comb. In *Drosophila bipectinata* it has been observed that males with larger sex combs have higher mating success (POLAK *et al.* 2004). An experiment involving the ablation of the sex comb, through the expression of *transformer* in the tarsal segments of male's legs, showed that this did not alter male mating behaviour, but did change the female's behaviour towards the male. This indicates that mating success depends on sex comb morphology, however this may be due to female preference, or possibly the co-evolution of its mechanical specificity to the female (NG and KOPP 2008).

3.2.2 Mean IPI

Within *D. melanogaster* song is produced by vibration of one wing, usually the one closest to the female's head (SPEITH 1952). The males emit a sound pulse (pulse song) with an interval between pulses known as the interpulse interval (SHOREY 1962). This trait has been found to be species-specific (KAWANISHI and WATANABE 1980) The

courtship song of *D. melanogaster* is comprised of two components known as sine and pulse song. It is thought that sine song acts to prime the female for mating (VON SCHILCHER F. V. 1976) and that of the two, pulse song is more important in mate choice (GLEASON *et al.* 2002; VON SCHILCHER F. V. 1976). The pulse song consists of the mean interpulse interval (IPI) and a patterned cycle in IPI (BENNET-CLARK and EWING 1969). The length of the song IPI cycle of *D. melanogaster* and *D. simulans* appears to be determined by a single allele, the sex-linked *period* gene (WHEELER *et al.* 1991).

Females within the *D. melanogaster* complex show a strong homospecific preference for IPI and females will mate more readily in the presence of their own species specific mean IPI (RITCHIE *et al.* 1999; VON SCHILCHER F. V. 1976). It is worth noting however that heterospecific song does have the effect of stimulating mating between different species. This appeared to be true for females of *D. melanogaster*, *D. simulans*, and *D. mauritiana.* These all mated with heterospecific species more often when heterospecific song was present. *Drosophila sechellia* appeared to be the most species specific with respect to song, and mated with heterospecific species more often when no song was present (TOMARU *et al.* 2000).

The difference in mean interpulse interval of pulse song may act as an important factor influencing the level of reproductive isolation between different species (GLEASON and RITCHIE 2004; RITCHIE *et al.* 1999), and genes that affect this trait may be a major influence on speciation within the *Drosophila melanogaster* species complex. Previous experimental evidence has found that mutations of the genes involved in the sex determination pathway; *fruitless*, *doublesex* and *transformer* can

cause strong effects on courtship song (BERNSTEIN *et al.* 1992; GLEASON and RITCHIE 2004; GOODWIN *et al.* 2000; RYNER *et al.* 1996; VILLELLA *et al.* 1997).

Through P-element insertional mutation the *fru* allele was discovered to be a major gene involved in determining male sexual behaviour. *fru* controls the development of the neural cell network within the CNS, and is expressed within regions of the mesothoracic ganglia (BILLETER *et al.* 2007; GOODWIN *et al.* 2000). The *fru* gene is associated with direct affects on innate sexual behaviour. The *fru* mutations are influential on all aspects of male specific courtship behaviour as well as the formation of a pair of male specific abdominal muscles, known as the muscle of Lawrence (MOL) (BILLETER *et al.* 2007; GLEASON and RITCHIE 2004; GOODWIN *et al.* 2000; RYNER *et al.* 1996; VILLELLA *et al.* 1997). The *dsx* gene is an important gene for sexual morphological development, particularly the formation of the male genital disc, as well as courtship behaviour (CHEN and BAKER 1997). Mutations of the *dsx* allele results in the elimination of pulse song (VILLELLA and HALL 1996) and *tra* is known to effect the structure of IPI cycles (VON SCHILCHER 1977).

tra and *tra-2* regulate the sex specific splicing of *fru* and *dsx* (BAKER and RIDGE 1980) and both *dsx* and *fru* are genes that transcribe putative transcription factors, which regulate numerous genes controlling male morphology and behaviour (ANAND *et al.* 2001; BILLETER *et al.* 2007; GOODWIN *et al.* 2000; USUI-AOKI *et al.* 2000). Tra proteins are expressed only in females, which bind to pre-mRNA consensus binding sites, causing alternative splicing of these transcripts. The male sex specific transcripts are produced from the P1 distal promoter, and are spliced near the 5' termini (GOODWIN *et al.* 2000; RYNER *et al.* 1996). The male specific splicing of *fru* can be

induced in *tra* and *tra-2* mutant females and anti-*fru* antibody reactive neurons have shown Fru proteins to be present in these mutant females. The hypothesis that male courtship behaviour is governed by the specific splicing of the *fru* P1 transcripts, was tested by Demir and Dickson (2005). Homologous recombination was used to create a number of *fru* alleles that prevented male splicing in male flies and induced male splicing in females. The induced male splicing in female flies resulted in females displaying all stages of male courtship, except for copulation. However subtle differences were observed, in that the mutant females spent less time extending and vibrating their wing during courtship song (DEMIR and DICKSON 2005).

It has been established that the sex-specific splicing of *dsx* controls the sexual morphological differentiation within *D. melanogaster* (BAKER 1989; BURTIS and BAKER 1989; WALTHOUR and SCHAEFFER 1994; WATERBURY *et al.* 1999). Similar functional homologs can be found in a number of different species and phyla (CLINE and MEYER 1996; RAYMOND *et al.* 1998). The expression of *tra* and *tra-2* triggers the female specific splicing of *dsx*, which is essential for the development of the correct female morphology within *D. melanogaster*. Females with homozygous mutations at these loci develop male morphology (BAKER and RIDGE 1980; NAGOSHI *et al.* 1988; NAGOSHI and BAKER 1990). The alternative splicing of *dsx* results in the sex-specific transcripts *dsx*-M and *dsx*-F, which have different DNA binding properties. These compete for the regulation of target genes directly linked with sex-specific differentiation (WATERBURY *et al.* 1999). However it appears that both *dsx* and *fru* play a role in the development of the CNS, since *dsx* has an important role in the differentiation of sex specific neuroblasts in the abdominal ganglion (BILLETER *et al.* 2006). Their correct sexually dimorphic pathway is dependent on *dsx* expression, and

the presence of functional *dsx* is essential for the post-embryonic division of sex specific neuroblasts, into male and female neurons (RAYMOND *et al.* 1998; TAYLOR and TRUMAN 1992). Recent experimental findings have established that courtship song depends on the apparent co-expression of *dsx* and *fru* within regions of the mesothoracic ganglia (BILLETER *et al.* 2007; RIDEOUT *et al.* 2007).

The assessment of the effects a known candidate gene has on naturally occurring variation can be made through Quantitative Trait Loci analysis. Previous QTL analysis on the interspecific trait variation of mean IPI between *D. simulans* and *D. sechellia* found six QTL accounting for 40.7 % of the phenotypic variation (GLEASON and RITCHIE 2004). All the significant QTL regions were situated on the third and second chromosomes and *fru* fell within a significant QTL region on the third chromosome. None of the QTL detected had a large effect on the interspecific trait variation, or coincided with the QTL from a previous study carried out on the same trait (mean IPI) between *D. melanogaster* strains (GLEASON *et al.* 2002), in which three QTL were found to account for 54 % of the variation between two inbred lines. However studies on female abdominal pigmentation and sex comb tooth number, found the QTL responsible for intraspecific trait differences did coincide with the same QTL regions causing interspecific trait variation (KOPP *et al.* 2003; NUZHDIN and REIWITCH 2000). In general the QTL for the intraspecific and interspecific studies do not coincide, which may indicate that the alleles contributing to trait variation within species are transient mutations not yet fixed (GLEASON and RITCHIE 2004). The more numerous QTL found for the trait variation between species complies with general findings that interspecific trait variation is of a more polygenic nature (GLEASON *et al.* 2002; KIM and RIESEBERG 1999) and the greater time scale of

divergence between species results in the detection of more numerous QTL (ORR 1998).

Candidate genes that have been detected through mutational analysis to have severe dysfunctional effects on trait variation do not always appear as significant QTL contributing to natural interspecific variation. By transformation of the cloned *nonA* allele from *D. virilis* to *D. melanogaster* (CAMPESAN et al. 2001), it was found that courtship song variation was influenced by this allele. However the same allele has not been detected as influential on naturally occurring trait variation within *D. virilis* or between the species of *D. virilis* group and the levels of nucleotide variability showed no signs of deviation from neutrality (HUTTENEN *et al.* 2002).

This QTL analysis primarily attempts to assess whether *fru*, along with the other sex determination genes *dsx* and *tra*, contribute significantly to natural interspecific trait variation of mean IPI and tests the magnitude of these effects. The additional analysis using the RI lines also attempts to test the association of the sex determination genes with mean IPI and sex comb tooth number. Previous evidence indicates that expression of both *fru* and *dsx* are involved in the trait variation of mean IPI (BILLETER *et al.* 2007; BILLETER *et al.* 2006; RIDEOUT *et al.* 2007), whereas the morphological trait variation of sex comb tooth number is associated with the *dsx* locus (BAKER and RIDGE 1980; GRAZE *et al.* 2007).

3.3 Materials and methods

3.3.1 QTL analysis

In the original experiment of Gleason *et al.* (2005), one inbred strain of each species was used, the *D. sechellia* strain was provided by David and the *D. simulans* strain by Coyne. A backcross experimental design was used due to the male sterility that occurs within the F1 generation. Female *D. simulans* were crossed to male *D. sechellia* flies and the resulting hybrid females were backcrossed to male *D. simulans*. The resulting backcrossed progeny would be expected to appear in a 1:1 ratio of male flies heterozygous for *D. simulans* and *D. sechellia* alleles to those homozygous for *D. simulans* alleles. The *D. simulans* strain incorporated 5 morphological markers, one for each chromosome arm, *forked* 15F7-9; *net* 21A5; *plum* 59E2-3; *scarlet* 73A3; *ebony* 93D1. These were used to pick out 32 backcross genotypes so that the song recordings included a good cross section of all the chromosomal combinations.

3.3.2 Song recording

The males were isolated in separate vials and recorded at 8 - 10 days old, by placing them in a recording chamber with an immature newly eclosed female. The recordings were made using an insectavox microphone (GORCZYCA and HALL 1987), for a duration of 5 minutes. The song recordings were then digitised using a Cambridge Electronic Design 1401 A/D converter (at 2 kHz after bandpass filtering at ~100 Hz-1kHz). The histograms of the IPIs were analysed using the "Spike2" language (RITCHIE and KYRIACOU 1994; RITCHIE and KYRIACOU 1996). A measurement of the recorded mean IPIs could then be made. Previous evidence found that mean IPIs are temperature dependant (SHOREY 1962). The temperature was recorded before and after each recording and all IPIs were adjusted to a common temperature of 25˚C,

using the formula $-1.6(25-T) + I$; T is the mean temperature of the recording and I is the mean IPI. The coefficient 1.6 was derived from previous studies (RITCHIE and KYRIACOU 1994; RITCHIE and KYRIACOU 1996).

3.3.3 Marker scoring

The male flies were frozen and the DNA was isolated (GLOOR and ENGELS 1992). In the original study Gleason et al (2005) 45 markers were scored in total on 433 individuals (including the 5 morphological markers). In this study the additional markers *doublesex* and *fruitless* were added. The markers were all PCR (see table A2.3 in the appendix section for PCR conditions) amplified and the different alleles were identified from different sized fragments for *D. sechellia* and *D. simulans* on a 2% agarose gel. The size differences were caused by natural variation in sequence length or by restriction enzyme cut sites (for the details of the primers and restriction enzymes used in this study, see tables A2.1 and A2.2 in the appendix section). The mapping of the order of the genes to their respective chromosomes was done using Mapmaker (LANDER *et al.* 1987). This information could then be used for QTL analysis, using QTL Cartographer version 1.16 (BASTEN *et al.* 1997).

3.3.4 Recombinant inbred lines: song recording and sex comb tooth analysis

Pure parental strains of *D. simulans* and *D. sechellia* were inbred for 18 generations. The infertility present in F1 males meant that the F1 heterozygote females had to be backcrossed to *D. simulans* males. Subsequent inbreeding and repeated sibling mating resulted in a number of RI lines with a mosaic of the parental genomes. The backcrossed progeny had a bias for homozygous *D. simulans* alleles. Any *D. sechellia* homozygous loci would therefore be present against an otherwise predominantly *D. simulans* background. The song recordings were then made of the parental lines and

all the RI lines that were *D. sechellia* homozygotes for one or more of the sex determination genes. The same methodology for the QTL song recording and marker scoring was used for the RI lines. Approximately 300 IPIs were recorded for each line. For the assessment of sex comb tooth number, the male flies were anesthetized and the prothoracic legs were removed from each male. The sex comb teeth of 15 males were counted using a microscope with a graticule inserted in the eyepiece. A measurement of tibia length was also taken to factor out the contribution body size might have on the number of sex comb teeth (MACDONALD and GOLDSTEIN 1999). In this study a correlation of tibia length and sex comb tooth number indicated that the trait did not correlate with tibia length variation. The *D. simulans* strain incorporated 5 morphological markers, one for each chromosome arm *forked* 15F7-9; *net* 21A5; *plum* 59E2-3; *scarlet* 73A3; *ebony* 93D1. The morphological markers were included to test for the possible influence that their respective chromosomal regions had on trait variation. All of the morphological markers along with the sex determination genes *fru*, *dsx* and *tra* were scored. The effects of all markers on both traits were assessed using a General Linear Model (GLM).

3.4 Results 1: QTL analysis of mean IPI

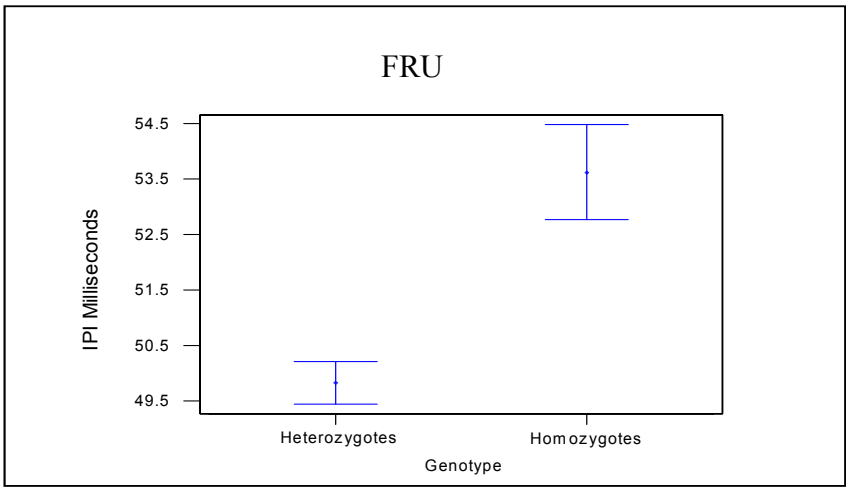


Figure 2. Represents a one-way anova, which indicated *fruitless* as the most significant marker associated with mean IPI on the third chromosome.

Single marker regression analysis was carried out on each of the 47 genes to test their potential linkage to a QTL. The most significant marker associated with mean IPI was *fru* (Figure 2). The second most significant one was *nanos* (*nos*). For all of the significant genes situated on the third chromosome, the *D. simulans* homozygotes showed longer mean IPIs whereas *D.simulans/D. sechellia* heterozygotes had shorter mean IPIs, this was the unexpected direction. The opposite relationship was seen for all the significant associated markers situated on the 2nd chromosome, which is the correct direction with respect to the *D. sechellia* phenotype. In relation to the single marker regression results (Table 1) over all chromosomes, the markers with the strongest association with the trait variation were *Pgi* and *Dga*. *tra* was not significantly associated with trait variation.

Table 1. The single marker regression results showing the most significant associated markers, a high F-value indicates a strong regression between each marker and the mean IPI.

Marker	Chromosome	DF	F-value	P-value
fru	3	1, 431	20.65	0.0001
nos	3	1, 431	19.26	0.0001
gl	3	1, 431	18.03	0.0001
e	3	1, 431	12.25	0.001
dsx	3	1, 431	7.44	0.001
Gld	3	1, 431	8.91	0.003
Pgi	2	1, 431	34.69	0.0001
Dga	2	1, 431	30.71	0.0001
sli	2	1, 431	27.33	0.0001
cad	2	1, 431	17.61	0.0001
f²	1	1, 431	9.08	0.003
nonA	1	1, 431	6.9	0.009

The single marker regression results highlight regions of each chromosome that may contain a QTL influencing this trait. *fru* and *dsx* were significant, however situated within a cluster of 6 significant genes (Table 1). The next stage in the analysis was to map the QTL using CIM. *fru* and *dsx* were then incorporated into the genetic map using MAPMAKER (LANDER *et al.* 1987), both genes mapped in their expected

order. Quantitative Trait Loci analysis was then carried out using QTL Cartographer (BASTEN *et al.* 1997). This program incorporates Composite Interval Mapping to assess the contribution of numerous genes on a specific trait or traits, by calculating a test statistic that accounts for the contribution of neighbouring QTL effects. This is important when assessing a phenotype that may involve numerous QTLs. The CIM test refines the study by testing for a QTL whilst assessing the influence of background markers, therefore testing more accurately for the actual effect of the individual QTL.

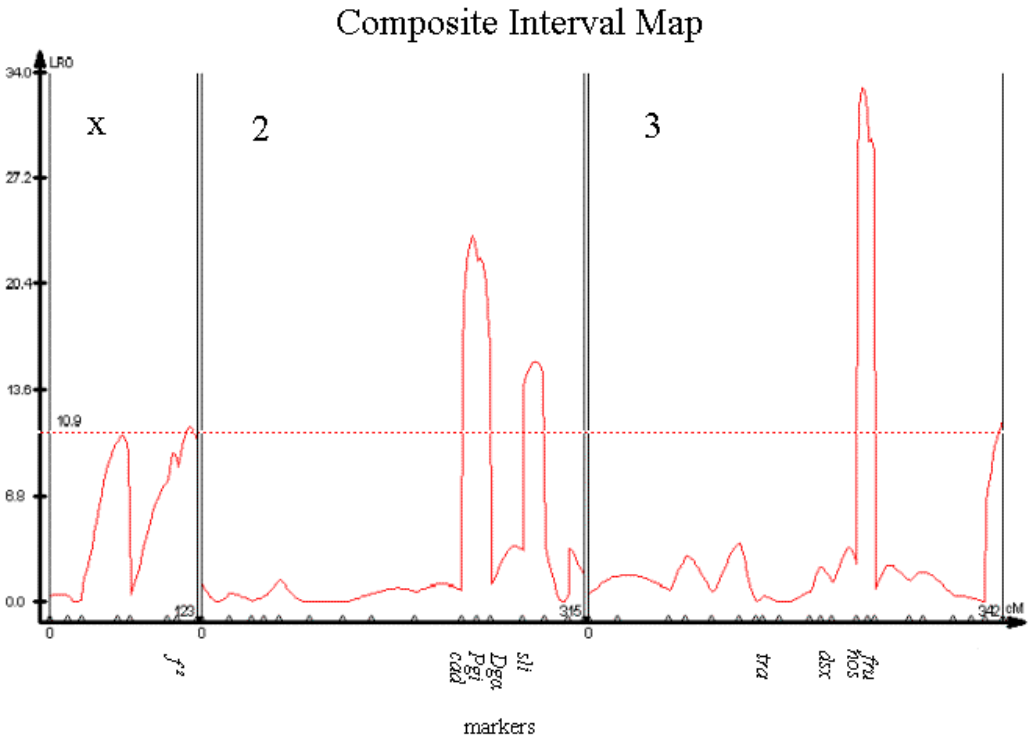


Figure 3. A Composite Interval Map using parameters as in the previous QTL study (GLEASON and RITCHIE 2004). The X-axis shows the mapped positions in cM (centimorgans) of all the markers (triangles). The sex determination genes and the markers associated with significant QTL are written in italics below. A likelihood ratio [LR] greater than 10.9 (LOD 2.4) corresponds to a significant P value 0.05.

The genomic regions influencing mean IPI variation within this study are very similar to those found in the initial QTL analysis (GLEASON and RITCHIE 2004). When using 7 background markers as cofactors the CIM test identified two significant (LR > 10.9) QTL associated with the markers *nos* and *fru* on the 3rd chromosome (Figure 3).

Table 2. Shows the position (cM), nearest marker and effects of the significant QTL, on all chromosomes (Chrom).

QTL	Chrom	Marker	Position	LOD	Additive	% Effect
1	1	*decI*	60.37	2.3369	0.0416	2.86
2	1	*nonA*	101.43	2.0806	0.0414	2.83
3	1	*f²*	115.82	2.4454	0.0451	3.35
4	2	*cad*	221.77	5.1051	-0.0688	7.8
5	2	*Pgi*	229.25	4.793	-0.0657	7.12
6	2	*sli*	274.75	3.3599	-0.0543	4.86
7	3	*nos*	225.32	7.1859	0.0761	9.55
8	3	*fru*	233.14	6.4763	0.0686	7.76

The presence of a *D. sechellia* allele at the markers associated with QTL on the 2nd chromosome; *cad*, *Pgi*, and *sli* resulted in a more *D. sechellia* like mean IPI. *D. simulans* homozygotes at these markers resulted in shorter mean IPIs, therefore with respect to *D. simulans* the affect was a negative value (Table 2). This expected directional affect was only apparent with loci situated on the 2nd chromosome. There was a significant QTL found on the X-chromosome situated near to *forked*. However the study does differ in that it has detected a significant QTL situated on the X-chromosome, and has identified *fru* along with *nos* as the markers on the 3rd chromosome with the strongest association with mean IPI.

The next stage was to use Multiple Interval Mapping. This simultaneously maps multiple marker intervals to QTL, and as a consequence takes into account all genetic interactions, allowing the detection of epistasis between the putative QTL (ZENG *et al.* 1999).

Table 3. Multiple Interval Mapping Results.

QTLs and Interactions	Type	Chrom	Nearest marker	Position (cM)	LOD	Effect	% Effect
1	A	2	*Dgα*	250.2636	5.49	-0.098	12.8
2	A	3	*fru*	226.208	3.94	0.0785	7.4
3	A	1	*f²*	114.8055	2.82	0.0688	4.3
Epistatic 1x3	AA				0.393	-0.054	1.2
Epistatic 2x3	AA				1.481	0.1023	2.8

MIM (Table 3) has identified fewer QTLs. The genomic region influential on IPI variation on the 3rd chromosome is situated between *nos* and *fru*, but appears to be more closely associated with *fru*. The MIM has also identified *Dgα* as the only significant QTL on the 2nd chromosome, and the magnitude of the individual QTL appears to have increased. *Dgα* was not detected as a significant marker in the CIM, and is situated in the interval between the two major peaks on the 2nd chromosome (Figure 3). MIM also agrees with the CIM findings in that there is a significant QTL situated on the X-chromosome associated with the marker *forked*. MIM has also detected that the QTL near *forked* on the X-chromosome is involved in epistatic interactions with both of the significant QTL situated on the 2nd and 3rd chromosome.

3.4.1 Results 2: Recombinant inbred lines

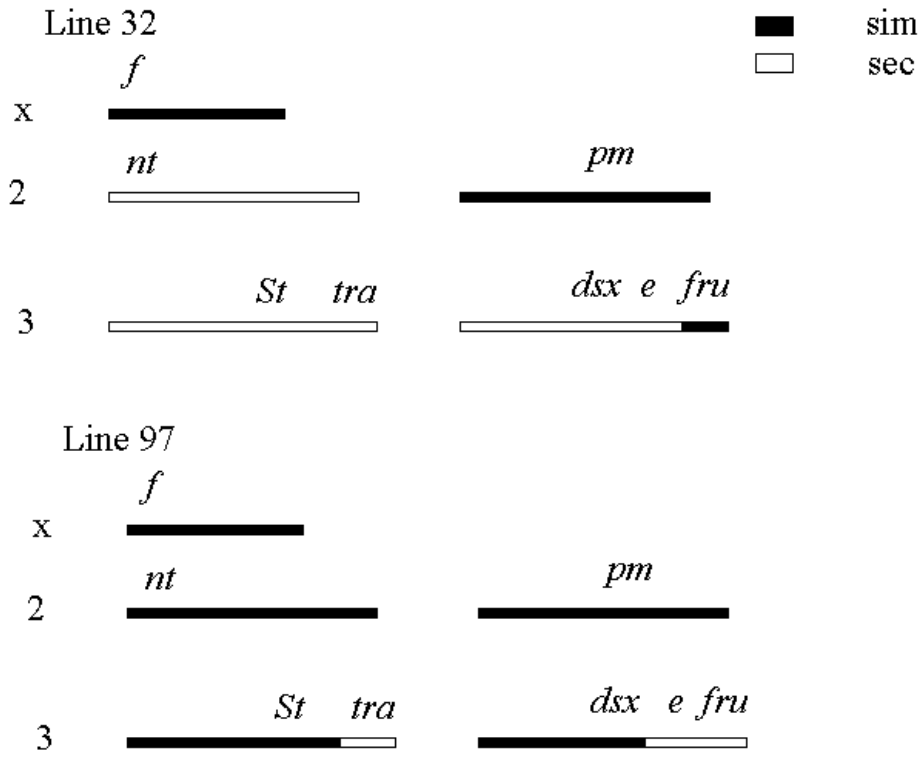
Table 4. Homozygous species specific alleles present within each RI line.

RI LINE	*f²* X	*nt* 2L	*pm* 2R	*st* 3L	*tra* 3L	*dsx* 3R	*e* 3R	*fru* 3R
Simulans	0	0	0	0	0	0	0	0
Sechellia	1	1	1	1	1	1	1	1
32	0	1	0	1	1	1	1	0
88	0	0	0	1	0	1	1	0
91	0	1	0	0	0	0	0	1
97	0	0	0	0	1	0	1	1
11	0	1	0	1	1	0	0	0
119	0	1	0	0	0	0	0	1
107	0	1	0	0	1	0	0	1
102	0	1	0	0	1	0	0	1

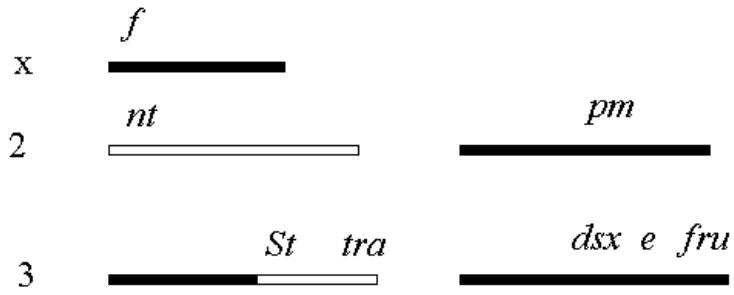
0 = D. simulans homozygotes 1= D. sechellia homozygotes

Table 4 shows the genotypes for all the markers scored in the recombinant inbred lines. There are no *D. sechellia* alleles present in all RI lines for the markers *forked*(*f*[2]) on the X chromosome and *plum* (*pm*) situated on the right arm of the 2[nd] chromosome. Therefore they could not be included in the GLM analysis. The overall level of recombination occurring in each RI line is represented diagrammatically in figure 4.

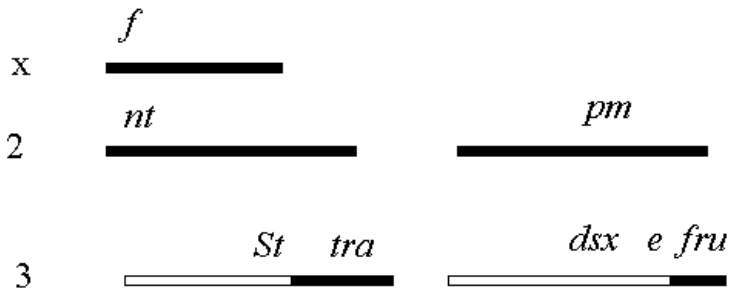
Figure 4. Schematic Diagram showing the recombination of *D. simulans* and *D. sechellia* markers for each RI line.



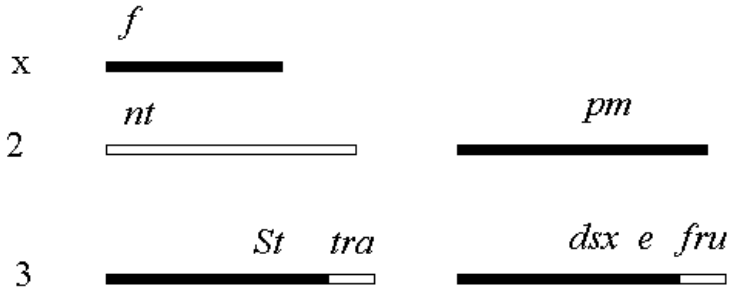
Line 11



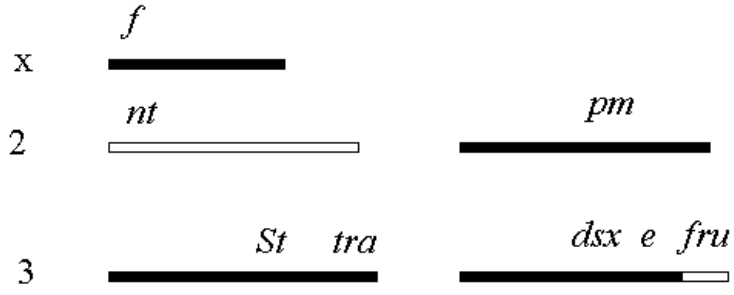
Line 88

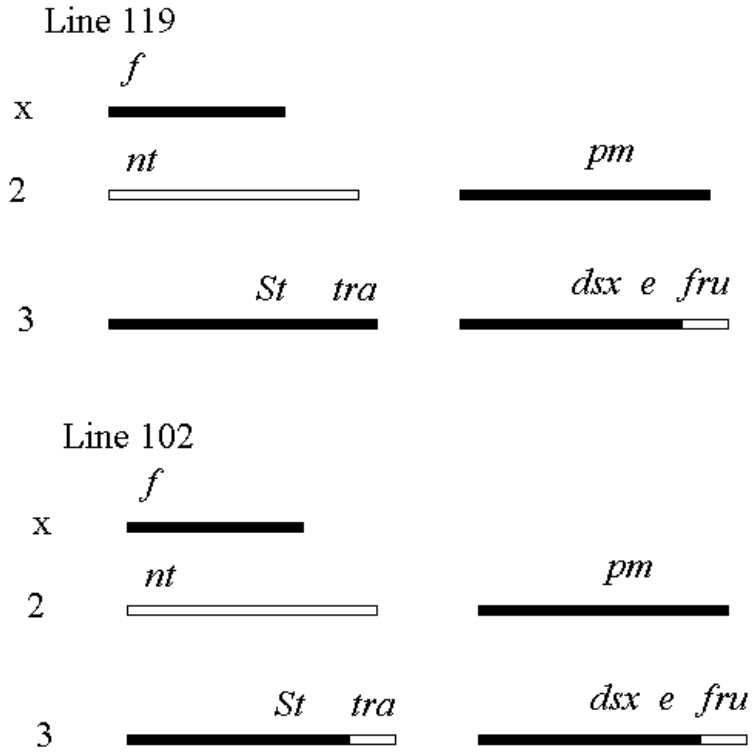


Line 107



Line 91





3.4.2 Results 2(a): Mean IPI: Recombinant inbred lines.

Figure 5. Mean IPI Interval Plot

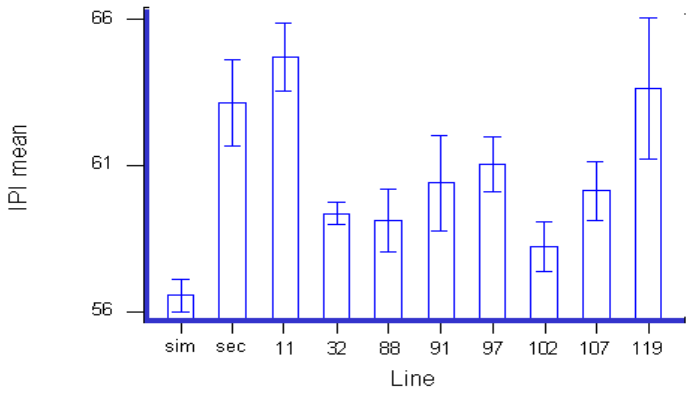


Figure 5 shows the variation in mean IPI (the bars represent the standard error) for both parental strains and each RI line. Both line 11 and 119 have significantly higher mean IPIs than all other lines.

Table 5. The General Linear Model (GLM), showing the direction of effect and the significance each chromosome (Chrom) region had on the variation of mean IPI.

Marker	DF	Chrom	Sim Mean IPI	Sec Mean IPI	F-value	P-value
nt	1, 88	2L	60.25	59.07	0.46	0.499
st	1, 88	3L	55.09	64.23	25.67	0.0001
tra	1, 88	3L	59.58	59.73	0.02	0.149
dsx	1, 88	3R	62.4	56.92	6.1	0.015
e	1, 88	3R	59.23	60.08	0.17	0.684
fru	1, 88	3R	57.18	62.13	12.46	0.001

RI Lines homozygous for *D. sechellia* alleles at the *fru* locus were associated with significantly longer mean IPIs and this is the expected direction, unlike the QTL direction of affect associated with this marker when *fru* was heterozygous (*D. sechellia/ D. simulans*) at this locus. The region associated with the *st* locus had the strongest association with mean IPI variation. Curiously this region is not associated with any known candidate gene affecting mean IPI. The *dsx* locus was associated with significantly shorter mean IPIs.

3.4.3 Results 2(b): Sex comb tooth number

Figure 6. Sex comb tooth number Interval Plot.

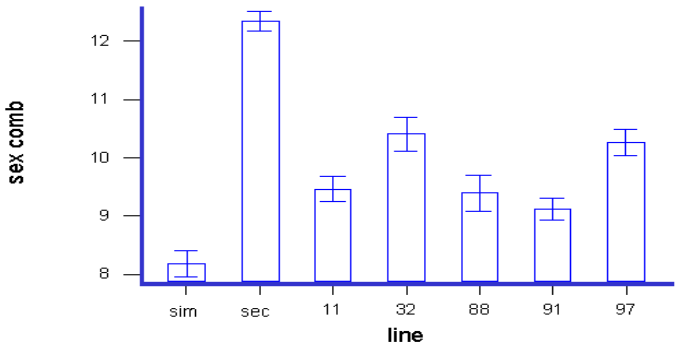


Figure 6 shows the variation in sex comb tooth number for both parental strains and each RI line. All RI lines appear to have a significant increase in sex comb teeth

compared to *D. simulans*. Line 32 has the highest mean for sex comb tooth number and has the highest number (5:3) of markers homozygous for *D. sechellia*. Line 97 also has a high mean for sex comb teeth, but only a ratio of 3:5 for *D. sechellia* homozygous markers.

Table 6. GLM results showing the direction of effect and the significance each chromosome (Chrom) region had on the variation of sex comb tooth number (SC).

Marker	DF	Chrom	Sim Mean SC	Sec Mean SC	F-value	P-value
nt	1, 97	2L	10.848	9.76	4.03	0.047
st	1, 97	3L	10.178	10.43	0.28	0.597
tra	1, 97	3L	9.267	11.341	15.76	0.0001
dsx	1, 97	3R	8.794	11.815	8.22	0.005
e	1, 97	3R	11.328	9.28	4.33	0.04
fru	1, 97	3R	9.29	11.319	21.17	0.0001

All the sex determination gene markers are associated with a significant increase in sex comb teeth. However a number of loci situated between the markers *tra* (73A10) and *fru* (91A7-91B3) may be involved with causing this affect.

3.5 Discussion

Multiple Interval Mapping (ZENG *et al.* 1999) is the most advanced option for detecting the presence of QTLs, and this method implicated *fru* as the marker most strongly associated with the significant QTL on the third chromosome affecting mean IPI variation. However surprisingly the presence of a *D. sechellia* allele at the *fru* locus was associated with significantly shorter mean IPIs, this is the opposite direction since *D. sechellia* has a longer mean IPI than *D. simulans*. The CIM indicated the region on the second chromosome between the markers *cad* (38E9) and *sli* (52C9) as most influential on mean IPI. The presence of *D. sechellia* alleles at all marker loci within this region resulted in longer mean IPIs. This conforms to the *D. sechellia* phenotype for this trait. This region includes the candidate genes *croaker* and *maleless*

(*mle*), both of which are known to be influential on courtship song (GLEASON 2005). Both the CIM and MIM analysis detected a significant QTL on the proximal end of the X-chromosome associated with the marker *forked*. This is not in agreement with a previous study (GLEASON and RITCHIE 2004) on mean IPI within *D. melanogaster*, in which influential QTLs were found only on the second and third chromosomes. However within *D. virilis* the *dissonance* (*diss*) allele of the *nonA* gene situated on the X-chromosome, in close proximity to *forked*, has been associated with alterations in pulse train and reduced mean IPI (CAMPESAN et al. 2001). The *diss* marker is close to a QTL in the CIM analysis of this study, but just below the LR significance threshold. The QTL associated with the marker *forked* was also involved in epistatic interactions with both the significant QTLs situated on the second and third chromosome. The epistatic effects detected between the QTL on the second chromosome with the QTL near *forked* on the X-chromosome may be a result of the interaction of the near by candidate gene *mle*, which interacts with *sex lethal* in relation to the hyper-transcription of X-chromosome gene products (LEE *et al.* 1997). This is purely speculative, however the epistatic interactions involving *forked* on the X-chromosome, may be in part responsible for the species-specific expression of this sexually dimorphic trait. It has been theorised that genes associated with sexually dimorphic traits would be more likely to be found on the sex chromosomes (RICE 1984; FISHER 1931), and a number of studies have found significant sex linkage associated with a number of sexually selected genes (LINDHOLM and BREDEN 2002; REEVE and PFENNIG 2002). However a more recent study assessing the percentage of pleiotropic genes (multiple phenotypes associated with one locus) within *D. melanogaster*, found 73 % of sexually selected genes to be pleiotropic, furthermore they did not show a significant sex linkage bias (FITZPATRICK 2004).

In this study there is a significant contribution from the sex chromosome, however it is not proportionally greater than the contribution from the significant regions found on each autosome. This study implicates the same regions of the genome as the previous QTL study on mean IPI variation between these two sibling species (GLEASON and RITCHIE 2004), however this study differs in that it also detects a significant affect from the QTL situated on the X-chromosome. However there are a few QTL studies on important parameters in courtship song that have detected significant contributions from the X-chromosome. A QTL study on IPI variation between *D. psuedoobscura* and *D. persimilis*, found two major affect QTLs, one situated on the X-chromosome and another on the second chromosome, both of these QTLs accounted for 95.8% of the genetic variation. It is important to note however that these QTLs were detected on non-recombining regions of each chromosome, and the individual QTLs may represent the affects of other undetected genes (WILLIAMS *et al.* 2001). Studies investigating the QTLs associated with the most variable courtship song parameter within different strains of *D. virilis* (the number of pulses) have found eight significant QTLs from all chromosomes most of which were situated on the third chromosome, which accounted for the largest proportion of the affect on the variation of this trait (HUTTUNEN *et al.* 2004). However studies investigating the QTLs affecting the same triat variation between members of the *D. virilis* group have found that the QTLs associated with regions on the X-chromosome contribute the most to the apparent interspecific variation (HOIKKALA *et al.* 2000).

Here the MIM analysis shows that more complex epistatic interactions, underlies the final expression of this sexually dimorphic trait. The QTL analysis does not detect *dsx* as a potential influence on natural trait variation and MIM shows that the

interspecific trait variation may well depend on the interaction of at least three relatively large affect QTLs involving the sex chromosome and autosomes. However none of these QTLs reach the affect level of 25%, attributed to "major affect" genes (BRADSHAW *et al.* 1998).

The RI analysis is limited in comparison to the QTL analysis of mean IPI, in that it uses fewer markers and smaller sample size. However an advantage of the RI line analysis is that it tests the effects of markers homozygous for the parental alleles for a specific trait. The RI line results (Table 5) indicate that both regions represented by the markers *fru* and *dsx* are associated with significant affects on mean IPI. However they show conflicting directional effects. Alleles homozygous for *D. sechellia* at the *fru* locus are associated with a significant increase in mean IPI (57.18 to 62.13 milliseconds). This is the predicted direction with respect to the *D. sechellia* phenotype, and this may indicate underdominance is occurring at the QTL associated with *fru*, in that the shorter mean IPIs were recorded for flies heterozygous at the *fru* locus in the QTL analysis (Figure 2). The opposite appears to be true for the QTL associated with the *dsx* locus and RI lines homozygous for *D. sechellia* at this locus have significantly shorter mean IPIs. The bidirectional affects detected may depend on the genetic background. In this study none of the RI lines were homozygous for *D. sechellia* at both of the *dsx* and *fru* loci simultaneously therefore the combined affects of both loci could not be assessed. This would have been interesting since recent findings have shown that the expression of both *dsx* and *fru* within the mesothoracic ganglia are necessary for song production in *Drososphila* (BILLETER et al. 2007; RIDEOUT *et al.* 2007). Intriguingly *st* (*scarlet*), situated on the left arm of the third

chromosome shows the greatest positive affect on the trait variation of mean IPI. This marker is not associated with any specific candidate gene affecting mean IPI.

The RI line analysis indicates that sex comb tooth number may depend on the additive effects of numerous loci and that the large QTL situated on the third chromosome associated with 53% of the trait variation between *D. simulans* and *D. mauritiana* may indeed be due to the composite effects of several QTL (GRAZE *et al.* 2007). All RI lines containing a proportion of the *D. sechellia* genome (roughly 15%) showed an increase in sex comb teeth compared to the mean of *D. simulans*. Interestingly, of the sex determination genes *fru* was the marker associated with having the most significant effect followed by *tra* and then *dsx* (Table 6). All of the sex determination markers were associated with a significant increase in sex comb tooth number when homozygous for *D. sechellia* alleles, whereas alleles homozygous for *D. sechellia* at the loci *nt* and *e* were associated with a decrease in the numbers of sex comb teeth.

The bidirectional effects detected for both traits, indicates the importance of the genetic background, which may influence the epistatic interactions controlling their proper phenotypic expression. The findings of this study implicate *fru* and *dsx* as having a possible influence on both of these sexually dimorphic traits, however the results are not conclusive. The QTL results indicate the possibility that *fru* may influence the naturally occurring variation of mean IPI between *D. simulans* and *D. sechellia*.

The results of the RI line analysis indicate that both *fru* and *dsx* may be influential on mean IPI, however the possible influence of QTL present at other linked loci is not

accounted for. Previous experiments have attempted to assess the importance of single loci on trait variation successfully. The molecular transfer of the *D. simulans period* loci to *D. melanogaster* proved it was responsible for the species-specific difference in song rhythm (WHEELER *et al.* 1991). However this molecular transfer would be problematic, due to the large size of *fru* (140 Kb), and also the fact that it is not being transferred into *D. melanogaster*.

The most recent QTL assessment of the genetic determinants of sex comb tooth number between *D. simulans* and *D. mauritiana* indicated two QTL associated with *dsx* and *Scr*, however the same study failed to find any apparent gene expression differences at these loci between both species (GRAZE *et al.* 2007). *takeout* is a known target gene of both *fru* and *dsx* and is also involved in the expression of male courtship behaviour. *CG2016* is a member of the *takeout* gene family and previous QTL analysis has associated this locus with sex comb tooth variation. Furthermore expression differences were found in *CG2016* between *D. simulans* and *D. mauritiana* within different leg tissues (GRAZE *et al.* 2007) . However it does not appear to be under sex specific regulation since previous analysis has also shown it to be equally expressed between male and female adult fly heads (DAUWALDER *et al.* 2002).

Recently more attention has been focussed on the slight changes within pre-existing regulatory elements, and the impact they may have on evolutionary diversity (SIMPSON 2007). The question still remains as to whether it is the species-specific differences within the gene coding regions or perhaps cis-regulatory differences that are responsible for the variation of trait expression. Phenotypic variation can be achieved with a minimal amount of sequence coding variation through the complex

timing and expression of regulatory genes involved in development pathways (CARROL 2005). It is possible that the sex determination genes may be similar, in that the species-specific changes are due to subtle temporal and spatial interactions affecting regulatory changes occurring throughout *Drosophila* development. Furthermore the co-evolution of both trans and cis-regulatory elements within species may underlie the mis-expression of certain phenotypic traits in hybrids (LANDRY *et al.* 2005). The sex-specific behavioural and morphological differences are achieved through alternative splicing of both *fru* and *dsx*, thus retaining the gentic coding integrity of these important developmental genes. However it is possible that subtle differences in the sequences of these sex determination genes at specific splice sites between different *Drosophila* species may also contribute to alterations of their interspecific expression.

CHAPTER 4

Sequence analysis of *fruitless*

4.1 **Abstrac**t

In this study an analysis was conducted of sequence variation of all *fru* proteins between ten recently sequenced Drosophilid genomes including *Drosophila melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. pseudoobscura*, *D. grimshawi, D. annasae*, *D. virilis* and *D. mojavensis*. These represent different species groups within the genus *Drosophila*, from a wide range of geographical locations and show vast ecological and behavioural diversity. The PAML program was used to detect the possible influence of natural selection on sequence divergence. There was no significant positive selection detected at the BTB functional domain and the sequences encoding for this domain were extremely conserved, indicating that strong purifying selection constraints operate within this domain. Positive selection was found to be acting on the exon encoding for the Zinc-finger C domain. This domain is present in two protein isoforms including the male sex-specific isoform FRU^{MC}, and the common non-sex-specific isoform FRU^{ComC}. Surprisingly positive selection was also found at the Zinc-finger D domain, this is exclusive to just one protein isoform (D) and is thought to be involved in the non-sex-specific vital functions of *fru* (GAILEY *et al.* 2006). It appears that the positive selection detected at the *fru* coding regions of the Zinc-finger DNA binding domains may account for some of the adaptive evolution found within sexually dimorphic traits in a number of *Drosophila* species.

4.2 Introduction

Within *Drosophila* the *fruitless* gene is roughly 140kb in size, and produces numerous proteins relating to the BTB/POZ (Broad-Complex, Tramtrack, Bric-a-brac, Poxvirus and Zinc finger) family of transcription factors (ITO *et al.* 1996; RYNER *et al.* 1996; WEN *et al.* 2000). *fru* specifically encodes for the BTB-Zinc finger domain family of proteins (ITO *et al.* 1996; RYNER *et al.* 1996) and is unique compared to the other BTB-ZF genes for its differential sex expression. The *fru* gene has the potential to produce a possible 15 transcripts through the use of four promoters and alternative splicing at the 5' and 3' ends. The complexity of the exonic splicing that occurs within *fru* allows the differentiation of its sexually dimorphic expression, controlling both sex-specific and non-sex-specific functions (ANAND *et al.* 2001; GAILEY *et al.* 2006; GOODWIN *et al.* 2000; ITO *et al.* 1996; RYNER *et al.* 1996; SONG *et al.* 2002; USUI-AOKI *et al.* 2000).

4.2.1 Sex-specific splicing of *fru*

The three male sex specific isoform transcripts (FRU^{MA}, FRU^{MB} and FRU^{MC}) are produced from the P1 distal promoter and are spliced at the default splice acceptor site near the 5' end, which results in an amino-terminal extension, proceeding the BTB domain. Each male specific transcript has a specific zinc finger domain (Zn-F-A, B or C), as a result of alternative splicing at the 3' end (see Figure 1) (GOODWIN *et al.* 2000; ITO *et al.* 1996; LEE *et al.* 2000; RYNER *et al.* 1996; SONG *et al.* 2002; USUI-AOKI *et al.* 2000). The male specific transcripts are expressed in the CNS, and mutational analysis has found that the expression of male specific *fru* proteins are involved in the formation of the male specific Muscle of Lawrence [MOL] and all the

stages in male courtship behaviour (GAILEY *et al.* 1991; GOODWIN *et al.* 2000; ITO *et al.* 1996; RYNER *et al.* 1996; VILLELLA *et al.* 1997).

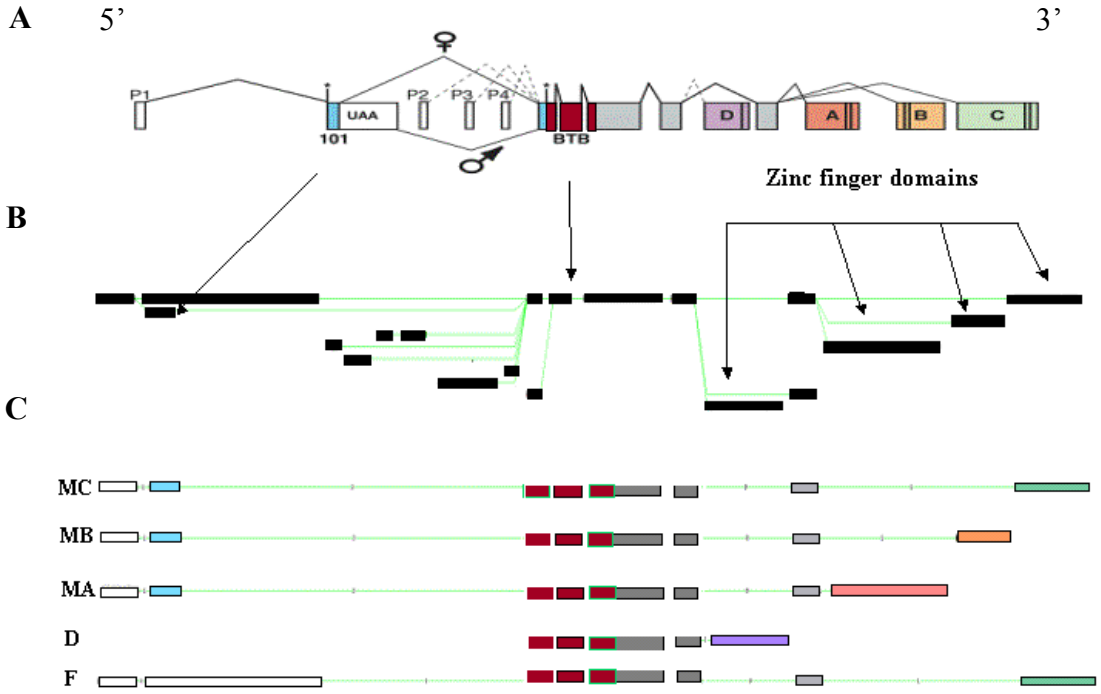


Figure 1. Schematic representation of the structure and splicing of the *fru* gene. (A) Shows the P 1-4 alternative promoters. The broken line shows the splicing from the non-sex-specific promoters (P2-4). The solid black lines define the splicing of the male specific P1 transcripts that contain an additional 101 amino acid domain shown in blue, present in all 3 *fru* male (FRU^{M}) specific transcripts. The BTB domain [red] is present in all 13 transcripts and involved in the non-sex specific vital function of *fru*. The 4 zinc finger domains are situated at the 3' end labelled A, B, C and D. The white boxes represent the un-translated exons. The regions shaded in grey represent other translated non-sex-specific exons (BILLETER *et al.* 2006). (B) The black boxes represent all of the *fru* exons and the green lines show the splicing pattern resulting in the formation of all *fru* transcripts (http://flybase.bio.indiana.edu/). (C) The 3 male transcripts (FRU^{MA}, FRU^{MB} and FRU^{MC}), incorporating the male specific exon (blue) spliced at the P1 promoter, each male transcript contains an alternative zinc finger domain (A, B or C). Below the male specific transcripts are two non-sex-specific transcripts, common to both sexes, F (FRU^{ComC}) and D (FRU^{ComD}).

In females the presence of Tra and Tra-2 proteins direct the splicing to a second downstream acceptor site, this prevents the translation of the male specific transcripts within female flies (HEINRICHS *et al.* 1998; RYNER *et al.* 1996; USUI-AOKI *et al.*

2000). The transcripts spliced at the more proximal end involving the second, third and fourth promoters are present in both male and female pupae and adult flies and are involved in *fru's* more vital (non-sex-specific) functions (GOODWIN *et al.* 2000; LEE *et al.* 2000; RYNER *et al.* 1996). These vital functions control the embryonic neural development, and embryos lacking functional *fru* proteins show abnormal growth of FasII, and BP102- positive axons, which are important for the formation of the normal axonal pathways within the CNS (SONG *et al.* 2002).

4.2.2 Structure and function of the BTB domain

The BTB domain is found at the amino terminal, which is separated by several hundred amino acids from the Zn-F (Zinc finger) DNA binding domain situated at the carboxyl-terminal end (GAILEY *et al.* 2006; PRIVÉ *et al.* 2005). Their general functions include dimerization, transcription repression and formation of high molecular weight DNA protein complexes (HUYNH and BARDWELL 1998; LI *et al.* 1999; MELNICK *et al.* 2000; SIEGMUND and LEHMANN 2002). The basic structure of the dimerization domain has been determined from X-ray chrystallography analysis of the BTB- PLZF (Promyelocytic Leukaemia Zinc Finger) protein domain. These experiments revealed a novel alpha/beta homodimeric fold, which allows for dimeric interactions to occur at two surfaces of the protein subunit (AHMAD *et al.* 1998; LI *et al.* 1999). The structure of the BTB domain also contains several highly conserved features, which include a charged pocket, a hydrophobic monomer core, an exposed hydrophobic surface at the base of the dimer and two negatively charged patches (AHMAD *et al.* 1998; LI *et al.* 1999; MELNICK *et al.* 2000).

The BTB domains interact with other BTB proteins, and it is likely that a network of BTB interactions exist, possibly through the formation of high order multimers (PRIVÉ *et al.* 2005). Analysis of the BTB domain protein has found the integrity of the dimer interface (interactive binding region) to be very sensitive to mutation (LI *et al.* 1999), and a number of point mutations at key residues are known to severely disrupt the proper function of the BTB domain. These mutations can result in the mis-folding and non-functionality of the protein, and one particular amino acid replacement (Mutation R49Q: arginine to glutamine) results in the domain activating rather than repressing transcription (MELNICK *et al.* 2000).

4.2.3 Mutations of *fru* and their affects on courtship behaviour

Various mutations associated with disrupting courtship behaviour (*fru* 2, 3, 4, and fru^{sat}) are produced by P- element insertions, placed between the P1 promoter and the common coding region of *fru*. The P- element insertions contain splice acceptor sites and disrupt the wild type splicing of the *fru* transcripts. The *fru* 1 mutation is caused by an inversion breakpoint at the distal P1 promoter region, and insitu-hybridisation experiments revealed this mutation results in abnormal distributions and levels of these male specific transcripts within the dorsal lateral protocerebrum and abdominal ganglion. This mutation also influences mean interpulse interval (IPI) of courtship song (GOODWIN *et al.* 2000; MELNICK *et al.* 2000; VILLELLA *et al.* 1997; WHEELER *et al.* 1988). These mutational experiments led to the implication of *fruitless* as a candidate gene for courtship behaviour in other organisms.

4.2.4 Conservation of *fru* across several insect species

Orthologues of *fru* are present in many different species across different insect orders, including Odanata, Hymenoptera, and more recently it has been discovered in Orthoptera in the sub family Gomphocerinae. It has been theorized that the *fru* gene is highly conserved due to its regulatory actions on the terminal effector genes, which act further down the sex-determination pathway (WILKINS 1995). A study comparing the functional domains of *fru* across several insect species including *Drosophila melanogaster*, *D. pseudoobscura*, *D. virilis*, *Apis mellifera*, *Tribolium castaneum* and *Anopheles gambiae*, showed the coding regions of the BTB and Zinc finger domains to be highly conserved (GAILEY *et al.* 2006). It is estimated to be around 250 million years since *Anopheles* and *Drosophila* diverged. However the *fru* ortholog present in *Anopheles gambiae* has a structure and conservation with functional domains similar to that of *fru* within *Drosophila*. Experimental evidence has revealed that the male specific mosquito isoform *Ag* Fru$^{\mathbf{MC}}$ also undergoes sex specific splicing, and its ectopic expression causes the development of the male specific MOL in female mosquitoes (GAILEY *et al.* 2006). The MOL is particular to more ancient species that pre-date the radiation of the sub family Drosophilinae and sexual behaviour and formation of MOL within *A. gambiae* appears to be governed by very similar processing to that of *Drosophila*. The sexually dimorphic differences between the sexes are also induced by male specific splicing resulting in three male specific mosquitoe isoforms; *Ag* Fru ^{MA}, *Ag* Fru^{MC}, *Ag* Fru^{MC} (GAILEY *et al.* 2006).

In *Drosophila* the enhancer response elements at which the Tra and Tra-2 proteins bind, contain nearly identical tandem repeat sequences within the alternatively spliced exons of both *fru* and *dsx* (HEINRICHS *et al.* 1998; LAM *et al.* 2003). The alternatively

spliced *fru* exons within *A gambiae* showed significant sequence similarity with these enhancer response elements found in *Drosophila* (GAILEY *et al.* 2006). It is possible that these regions could act as putative binding sites for the Tra and Tra-2 protein complex within *A. gambiae*. The other important sex determination gene *dsx*, which controls the expression of the sexual dimorphic morphology within Drosophilidae is highly conserved within nematodes, rodents, and humans (ZARKOWER 2001). It has been established that the basic structure of *fru* is conserved in several insect species, however the full extent of *fru* homology across more diverse taxa is as yet not known (GAILEY *et al.* 2006).

The BTB and Zn-F domain protein coding sequences of *fru* are highly conserved across dipteran insects and the alternative splicing of *fru* produces the numerous transcripts responsible for its functional diversity (ANAND *et al.* 2001; GAILEY *et al.* 2006; GOODWIN *et al.* 2000). In most recorded cases *fru* has been found to exist as a single copy gene, however recent evidence indicates the possibility of several copies of *fru* in the grasshopper genome (USTINOVA and MAYER 2006). Grasshoppers belonging to the genus *Chorthippus* are known to produce mating song, which acts as a premating barrier preventing hybridisation (VON HELVERSEN and VON HELVERSEN 1994). A partial fragment of the BTB domain of *fru* was cloned in *Chorthippus biguttulus*, *C. brunneus* and *C. mollis* (USTINOVA and MAYER 2006). The translation of the cloned region of *fru* revealed interesting similarities as well as differences to the structure of *fru* found in other insect species. The BTB domain and at least one Zinc finger domain appeared conserved, the repetitive linker sequence between both domains was less conserved, more closely resembling *Apis mellifera*. There is evidence that alternative splicing operates in grasshoppers, however only two

transcripts are found compared to several within *Drosophila*. It is not known if they are subject to sex specific splicing and there is no apparent evidence of any transcription factor binding sites or regulatory signals within the non-coding 5' region of *fru* in grasshoppers.

The most interesting finding was that numerous copies of *fru* exist within individuals, indicating that there are several paralogues of *fru* in the grasshopper genome (USTINOVA and MAYER 2006). Gene duplication events are common, however in most cases the duplication of genes is followed by their rapid degradation. However in rare cases its possible that duplicated gene copies can evolve to fulfil different functional roles (LYNCH and CONERY 2000). Within the grasshopper *fru* paralogues may fulfil the genes functional diversity that is otherwise achieved through alternative splicing of the single copy of *fru* in Drosophilid species (USTINOVA and MAYER 2006)

Genes that are differentially expressed between the sexes often show the most divergence among species (PRÖSCHEL *et al.* 2006; ZHANG *et al.* 2004). *fru* is an important gene involved in the early stages of neuronal development and has vital non-sex-specific functions, as well as being involved in controlling the expression of sexually dimorphic traits (SONG *et al.* 2002). It is likely that both purifying and positive selection operate at different rates on different exons depending on the sex and non-sex-specific functions of *fru*. For example exonic regions associated with the sexual dimorphic expression of courtship behaviour may experience more positive selection than the exonic regions associated with *fru* vital function. Deciphering the genetic basis of species-specific differences in sexual dimorphic traits has to start with assessing the apparent variation within coding regions of a gene, here we assess the

coding variation of fru in 10 recently sequenced Drosophilid genomes (CLARK *et al.* 2007) including *Drosophila melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. pseudoobscura*, *D. grimshawi, D. annasae*, *D. virilis* and *D. mojavensis*.

4.3 Materials and methods

4.3.1 Section 1: Assessment of the variation of the BTB domain sequence within and between species of the *melanogaster* subgroup

The initial assessment of the sequence variation was carried out on a region of the BTB domain. Sequence data for 8 of the 9 species within the *D. melanogaster* subgroup were obtained by using a Blast search from GenBank (www.ncbi.nlm.nih.gov). An alignment of the 8 available species was made using ClustalX (version 2.0) (THOMPSON *et al.* 1997) and forward and reverse primers were then designed within the conserved regions using Primer3. These primers were then used to sequence the same region for *D. santomea* and 9 African *D. melanogaster* populations. A standard PCR protocol was used to amplify the sequences (see table A2.3 in the appendix section for the standard PCR reaction and the cycling conditions). The DNA samples were cleaned using either the QuickStep PCR Purification Kit (www.edgebio.com) or the QIAquick Gel Extraction Kit (www.QIAGEN.COM), depending on the purity of the product. The purified products were then sent for sequencing at the Dundee University Sequencing Department (www.dnaseq.co.uk). The forward and reverse sequences for each DNA sample were aligned using ChromasPro program (version 1.22) and the edited sequences were entered into GeneWise (BIRNEY *et al.* 2004). The program was used to predict the gene structure using the *D. melanogaster* translated protein as a template, and indicates the correct splice sites between the introns and exons for each sequence.

This was then confirmed using the BioLign program (version 2.0.7) by manually edititing out the introns and translating the exonic coding regions. The sequence alignments and GeneWise translation reports for the BTB domain can be seen in the supplementary material in the appendix section (Tables A4.1 and A4.2). A molecular phylogenetic tree was also created for the *melanogaster* species subgroup with the inclusion of *D. santomea*, from *fru* exon and intron sequences spanning a section of the genomic region, which encodes for the BTB domain. The sequences were aligned using BioLign and the online program Treecon (VAN DE PEER and DE WACHTER 1994) was used to construct a neighbour-joining phylogenetic tree, using *D. pseudoobscura* as an outgroup.

4.3.2 Section 2: Assessment of selection acting upon *fru* proteins within 10 sequenced Drosophilid genomes

The amino acid sequences of the 13 *fruitless* transcripts of *D. melanogaster* were obtained from FlyBase (http://flybase.bio.indiana.edu/). Each protein was used as the query in a TBLASTN (GERTZ *et al.* 2006) search against the following published Drosophilid genomes; *Drosophila melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. pseudoobscura*, *D. grimshawi, D. annasae*, *D. virilis* and *D. mojavensis* (CLARK *et al.* 2007). The identification and reconstruction of the *fru* gene followed the procedure outlined in Gardiner *et al.* (2008). Ininitially a TBLASTN search was carried out. This uses an algorithm which matches the query amino acid sequence with nucleotide sequence data. The search focuses on windows of scaffolds made up of small sequence units (contigs) constructed from ordered overlapping clones. The E-value associated with each hit indicates the probability of the nucleotide composition matching that of the query protein. The hits with low E values and high

bit scores represented the best blast search results. The co-ordinates of these regions were retained and extracted using a BioPerl script (STAJICH *et al.* 2002). The nucleotide sequences were then entered into GeneWise and the homologous protein of *D. melanogaster* was used as a template for the construction of the probable gene structure of each sequence. The GeneWise algorithm incorporates parameters that account for gene structure and sequencing error (BIRNEY *et al.* 2004). The presence of improbable frame shift mutations, or start and stop codons indicated the possibility of pseudogenes. GeneWise reports that showed truncated translations were manually corrected to include the start and stop codons. The resulting sequences of orthologues were then assessed for selection within the exonic regions of significant amino acid variation.

The computer package PAML (Phylogenetic Analysis by Maximum Likelihood) (YANG 1997) was used to test for selection at different coding regions of the *fru* gene, across the 10 Drosophild genomes. The program uses an algorithm, which searches for the presence of codons with significantly different ratios of non-synonymous (*dn*) to synonymous (*ds*) substitutions. A ratio (ω) greater than 1 indicates positive selection and a ratio less than 1 showed that purifying or stabilizing selection is operating. The likelihood ratio is made by comparing 2 pairs of nested models (ANISIMOVA *et al.* 2001). These models conform to the null (H_0) and alternative (H_1) hypothesis. The null hypotheses represented by Model 7 assumes that no selection is operating on the coding sequence, and is set with a beta distribution from 0-1 and estimates this distribution with 10 classes of sites with differing ω (ratios) values. The alternative hypothesis represented by Model 8 adds a new class of sites and assumes positive selection is present ($\omega > 1$). The likelihood ratio is obtained by the formula

LR= - 2(ln H_0 – ln H_1). The P-value of the log likelihood was obtained assuming the chi squared distribution with 2 degrees of freedom. The Bayes Empirical Method (YAND *et al.* 2005) was applied to find the sites within coding regions experiencing significant selection. Genomic sequence data was incomplete for *D. simulans* and *D. sechellia* in the published genomes (CLARK *et al.* 2007), for the exonic region encoding for the zinc finger C domain. These regions were manually sequenced.

4.4 Results 1: Sequence conservation of the fruitless BTB domain within the *D. melanogaster* species subgroup.

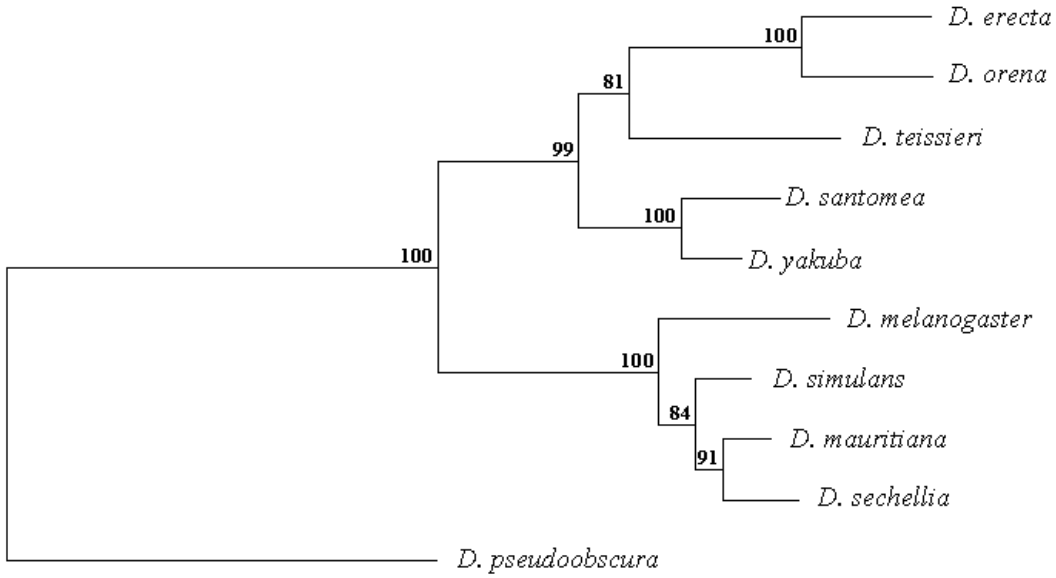


Figure 2. Neighbour-joining tree for the *melanogaster* species subgroup (using *D. pseudoobscura* as the outgroup). Bootstrap values were out of 100, and indicate the streangth of support for each node.

The neighbour-joining tree (Figure 2) confirms the status of *D. santomea* as a close relative of *D. yakuba*. The translation of the exonic coding regions of the BTB domain showed complete conservation of the protein within all of the *D. melanogaster* species subgroup and all *D. melanogaster* African strains. The phylogenetic relationship represented in figure 2 was therefore derived exclusively from silent site changes within exons as well as the variation existing in the intronic regions of *fru*.

4.4.1 Results 2: Analysis of positive selection on all *fru* protein domains, across several sequenced Drosophilid genomes.

Table 1. PAML (YANG 1997) Analysis for all 13 *fru* protein isoforms.

Protein Isoform	**Ln** Model 7	**Ln** Model 8	(2*log) - 2(ln H_0 – ln H_1).	**P-value**	**P1**	**ω (ratio)**
A	-5080.24825	-5080.21429	0.067912	n/s		
B(FRU^{MC})	-8003.4887	-7999.71218	7.553054	0.025	0.01367	1.54695
C	-10109.6746	-10109.6746	6.00E-06	n/s		
D	-7747.0129	-7732.43851	29.14878	0.001	0.04611	1.58629
E(FRU^{MA})	-11250.4308	-11249.4746	1.912246	n/s		
F	-6758.08697	-6754.9998	6.174332	0.05	0.00666	1.9134
G(FRU^{MB})	-7939.60726	-7939.6074	-0.000272	n/s		
H	-6819.03459	-6819.03473	-0.000268	n/s		
I	-9761.52609	-9761.50709	0.038016	n/s		
J	-9403.95478	-9402.0478	3.813958	n/s		
K	-6766.00517	-6766.00526	-0.000196	n/s		
L	-8956.64093	-8956.63726	0.007346	n/s		
M	-9591.52174	-9588.94942	5.144644	0.1	0.00271	6.63541

The results (Table 1) were derived from analysing the exonic nucleotide sequences of all 13 *fru* isoforms from 10 sequenced species genomes. Significant positive selection is evident in 3 *fru* protein isoforms, B, D and F. The Likelihood Ratio (LR) was calculated as follows; LR= - 2(ln H_0 – ln H_1), and the P-value of the log likelihood was obtained assuming the chi squared distribution (with 2 degrees of freedom). The P1 values refer to the proportion of sites undergoing positive selection and this appears to be greatest for protein isoform D. The Bayes Empirical Method (YAND *et al.* 2005) was used to locate the sites at which there were significant amino acid replacements (see the appendix section table A4.3 for details). These results indicated that positive selection was present within two protein coding regions, the Zinc-finger C and D domains.

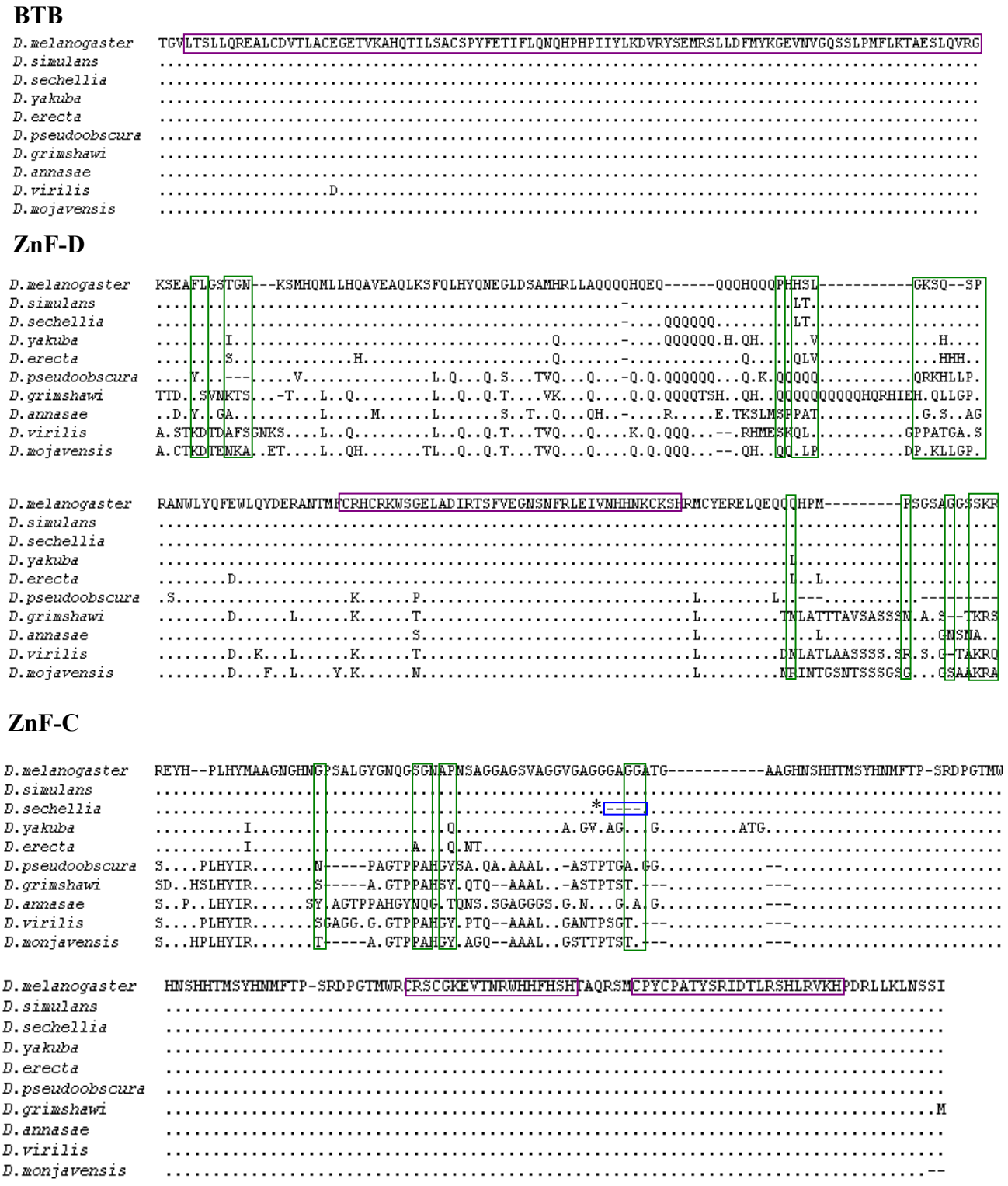


Figure 3. Protein sequence alignment of the *fru* functional BTB domain and the ZnF-D and ZnF-C domains across 10 sequenced Drosophilid genomes. The horizontal rectangular boxes (purple) on the *D. melanogaster* sequence indicate the domain sequences. The vertical boxes (green) indicate the sites at which positive selection was detected, using the Bayes Empirical Method (YANG *et al.* 2005).

* The blue rectangular box highlights a deletion found in the sequenced *D. sechellia* ZnF-C domain

Figure 3 shows that the BTB domain is highly conserved with only one amino acid replacement occurring in *D. virilis*. Significant selection was detected in the exonic regions coding for the ZnF-C and D protein isoforms. No significant positive selection was detected at the functional domain sequences highlighted by the purple rectangular boxes. The ZnF-D proteing coding region has the highest incidence of sites at which positive selection was detected. The ZnF-C exonic coding region also has as a high level of positive selection occurring preceeding the functional domain sequence. There is also a deletion of four amino acids within the *D. sechellia* ZnF-C exon, this occurs at a region where positive selection was detected.

4.5 Discussion

The analysis of *fru* protein sequences showed the BTB domain to be strongly conserved in several populations and all the species within the *D. melanogaster* subgroup. The conservation of the *fru* BTB domain across several distantly related species (*melanogaster*, *obscura*, *virilis*, *repleta* and *grimshawi*) (Figure 3) also implies purifying selection operates on this important domain, which is present in all *fru* protein isoforms. This is in agreement with several studies confirming strong functional constraints on the BTB domain in numerous insect species (DAVIS *et al.* 2000; GAILEY *et al.* 2000; USTINOVA and MAYER 2006). The phylogenetic tree (Figure 2) confirms the status of *D. santomea* (recently discovered on Sao Tome island in the Gulf of Guinea) as a sister species of *D. yakuba* (COYNE *et al.* 2002; LACHAISE *et al.* 1988). There was no positive selection detected for the exon encoding the male sex-specific transcripts produced at the P1 promoter.

Significant positive selection was detected within the coding regions of the two separate zinc finger domains ZnF-D and ZnF-C (Figure 3). The detection of positive selection at the ZnF-D domain is an intriguing finding, since it is exclusive to the one transcript involved in non-sex-specific function and little is known about the functional properties of this domain (GAILEY *et al.* 2006).

The ZnF-C domain and it is present in both male sex specific proteins ($Fru^{MC)}$ and non-sex-specific proteins ($Fru^{ComC)}$. Experiments disrupting the functionality of Fru^{MC} show that it is associated with the sexual dimorphic expression of all stages involved in courtship. It is also the only isoform controlling the correct formation of the Muscle of Lawrence (BILLETER *et al.* 2006). The sequenced *D. simulans* and *D. sechellia* ZnF-C coding region revealed an amino acid deletion (GAGG) in *D. sechellia*, spanning one of the regions at which positive selection was detected (highlighted by the blue box in figure 3). However it is impossible to assess the potential influence the coding variation detected has on the inter-specific variation of sexually dimorphic traits amongst the diverse range of species included in this study, without careful manipulative experiments. It is possible that the sequence variation detected may alter the regulatory activity of *fru* without disrupting the properties of the functional DNA binding domain, containing the C2H2 residues (highlighted by the purple rectangular box in figure 3). The results show this functional domain to be highly conserved, however it is thought that the divergence of transcription factor activation sites (whilst conserving the DNA binding domain) may also contribute to species diversity (LEVINE and TJIAN 2003).

There are a number of evolutionary strategies that allow genes to be differentially expressed, whilst conserving their functional coding domains, such as gene duplication events, resulting in the formation of numerous paralogues, which may increase the gene's functional diversity (CARROL 2005). It has been suggested that such duplication events may drive the differential expression of *fru* in grasshoppers, which have been found to contain several *fru* paralogues (USTINOVA and MAYER 2006). In Drosophila *fru* exists as a single copy and its differential expression between the sexes has been achieved by alternative splicing and the use of numerous promoter regions (GOODWIN *et al.* 2000), which produces various transcripts, whilst retaining the integrity of some functional domains of *fru*.

More recently emphasis has been made on the importance of alterations of sequences within cis-regulatory elements (CARROL 2005) and the evolution of these elements is especially important in the alteration of expression of pleiotropic genes in which coding variation would result in the alteration of many interrelated phenotypic traits. These elements have co-evolved within species, and control the gene expression, resulting in the evolution of diverse phenotypes using a minimum amount of sequence variation, within the protein coding regions of genes (WITTKOPP *et al.* 2002). Theoretically the role of regulatory elements in the evolutionary process could be likened to altering the architect plans whilst still conserving the integrity of the building materials (structural coding regions). It would be a logical assumption that the interspecific expression differences of pleiotropic regulatory genes such as *fru* are more likely to be controlled at the cis-regulatory regions.

However positive selection detected at the level of amino acid polymorphisms for many traits may drive many important differences between species and there is a large body of empirical evidence to support coding sequence adaptation in morphological and physiological traits (HOEKSTRA and COYNE 2007). Experimental analysis has found that the rapid evolution of many male biased genes linked to reproductive functions, such as the difference in the accessory gland protein *Acp26Aa* between *D. mauritiana* and *D. simulans*, is driven by positive selection acting on protein coding regions (TSAUR *et al.* 2001). In fact all known "speciation genes" such as *OdsH* (a homebox gene from a family of transcription factor-encoding genes) connected with post-mating reproductive isolation, show a rapid evolution within protein coding regions between different species (ORR *et al.* 2004; WU and TING 2004). The positive selection detected within the coding region of the ZnF-C domain may affect the regulatory activity of *fru* and the presence of alternative zinc finger domains in each male specific *fru* isoform suggests they may each have a different binding specificity interacting with several different target genes (BILLETER *et al.* 2006; GAILEY *et al.* 2006). There is growing evidence that changes in transcription factor proteins are more common and may evolve more rapidly than previously thought. These changes may alter the expression of downstream targets without having detrimental effects on the downstream pathway (BUSTAMANTE *et al.* 2005; HOEKSTRA and COYNE 2007; HSIA and MCGINNIS 2003). It is likely that adaptation and speciation progress through a combination of structural (protein coding) and cis-regulatory mutations, however the emphasis on their importance is a matter of some debate. This study indicates that structural mutations of *fru* may drive phenotypic variation of courtship differences observed between several Drosophilid species. The highest level of conservation was found within the coding sequence of the BTB domain. This level of constraint was not

found within the ZnF-C domain, which is known to be involved in the sexually dimorphic expression of both morphological and behavioural traits. Therefore it appears that the divergence and constraint of the coding regions of *fru* may depend on their function role. The exonic region encoding for the ZnF-C domain, involved in male sex-specific traits, appears to be under less functional constraint, and subject to greater divergence than exonic regions involved with *fru*'*s* more vital function. However it is intriguing as to why such a high level of positive selection was found at the ZnF-D domain, since it is present in only one *fru* protein isoform, and little is known about its functional role.

.

CHAPTER 5

Conclusions and future work

Mutational screening has identified numerous candidate genes associated with traits influencing sexual isolation within *Drosophila*. It is known that the sex determination hierarchy controls the sexual differentiation of *Drosophila* morphology and courtship behaviour. The expression of Tra protein determines the sexual expression of both *fru* and *dsx*, which in turn control the respective critical downstream terminal effector genes involved with sexual morphology and behaviour (BAKER *et al.* 2001; BILLETER *et al.* 2002; BURTIS 1993). The induced expression of the transgene (*UAS-tra*) in males causes the feminisation of their brain and CHC profiles (SAVARIT *et al.* 1999). The results of the QTL analysis on the candidate genes implicated in affecting CHCs, showed that of all the sex determination genes a QTL associated with *dsx* exerts the strongest influence on affecting naturally occurring variation of 7-tricosene. This QTL peak was just below the significance threshold (Chapter 2, Figure 1), however this is still an indication that there may be allelic differences between these two species at the *dsx* locus influencing differences in the levels of 7-tricosene.

The two large affect QTL detected on the third chromosome affecting 7,11-HD (a CHC present in females of dimorphic species *D. melanogaster* and *D. sechellia*) were closely associated with two desaturase candidate genes, *desatF* and *eloF*. Experiments involving the RNAi knockdown of both these genes in *D. melanogaster* show that they play a crucial role in female cuticular hydrocarbon synthesis. The induced expression of the transgene (*UAS-tra*) in *D. melanogaster* males resulted in expression of *desatF* and *eloF* (CHERTEMPS *et al.* 2007; CHERTEMPS *et al.* 2006),

which suggests that both of these terminal effector genes are controlled by the sex determination hierarchy, and the species specific differences in the levels of 7,11-HD may be due to the sequence variation at these terminal effector genes.

More recently, experiments involving the RNAi knockdown of *desatF* in African strains of *D. melanogaster* have shown that it may also be involved in 5, 9 production and suggests that the geometric positioning of a second desaturase (four carbons along from the first desaturase position) for both 5, 9 and 7, 11 diene compounds may be determined by the structural conformation of this desaturase enzyme (LEGENDRE *et al.* 2008). In the same study results from the analysis of sequence variation between *D. simulans*, *D. melanogaster* and *D. sechellia* showed conservation within coding regions compared to the rapid divergence of the promoter regions. In fact the *D. simulans* promoter region was 10-15% longer than that of *D. melanogaster*, which suggest that the extra sequence may contain inhibitory elements, down regulating its transcription (LEGENDRE *et al.* 2008). This indicates that the loss of expression of 7,11-HD in *D. simulans* has possibly evolved through the loss of function of the *desatF* gene, through the corruption of the promoter sequence region.

The candidate gene approach has been very successful in discovering the common functional role (often shared across numerous species and taxa) of several genes, and this approach implied the *desat2* locus has a major affect on the differences in pheromone profiles between *D. melanogaster* populations. The *desat2* allele is functional in African populations but non-functional in Cosmopolitan strains, due to a 16 bp deletion in its promoter region (DALLERAC *et al.* 2000; TAKAHASHI *et al.* 2001) and it is thought that this dysfunctional allele gives the cosmopolitan *D. melanogaster*

populations an adaptive advantage to colder geographic regions (GREENBERG *et al.* 2003). It is thought that ecological adaptation is the main driver of population divergence and speciation (GREENBERG *et al.* 2003; SCHLUTER 2000) and this example of ecological adaptation agrees with Muller's hypothesis that pre-mating reproduction occurs as a by-product of genetic divergence through allopatric speciation (MULLER 1942; SINGH and HOENIGSBERG 2004). However the ecological advantage caused by the loss of function of *desatF* in *D. simulans* is not as apparent. The sequence changes for *desat2* and *desatF* have resulted in their loss of function, and it may also be sequence variation in the promoter region of *desatF*, which causes its loss of function. This would agree with the theory suggesting cis-regulatory changes are more commonly associated with the elimination of traits than the evolution of new ones (HOEKSTRA and COYNE 2007).

The overall genetic architecture of the interspecific differences of pheromone profiles between *D. simulans* and *D. sechellia* appears to be polygenic with numerous QTLs contributing to the genetic variation. However unlike the genetic architecture of post-mating reproductive isolation, there are a few QTLs that exert a relatively large affect. It is also likely that more numerous minor affect QTLs (not detected within this study) contribute to the rest of the total phenotypic variation. The four markers associated with the greatest genetic effect on specific CHCs were *forked*, *desatF*, *Mtn* and *pros* (*eloF* is situated between *Mtn* and *pros*). The QTLs associated with the marker *desatF* were influential on both diene compounds (7,11-HD and 7,11-PD). The strong epistatic interaction detected between the QTLs associated with the markers *desatF* and *Mtn* for 7,11-HD suggests the possible importance of the co-expression of two QTLs for the production of 7,11-HD. *eloF* is a candidate gene associated with the

synthesis of long chain female CHCs and possibly may interact with *desatF*. The QTL associated with the marker *forked* appears to exert a strong influence on the interspecific variation of all monoene compounds.

The recent RNAi experimental findings on the role of *eloF* in the expression of long chain female specific CHCs, indicates that the scoring of *eloF* and its incorporation into this QTL study would provide more conclusive evidence of its affect and interaction. The addition of further markers near *forked* on the X-chromosome would also be necessary to increase the resolution of this region and possibly indicate the QTL or QTLs involved in the interspecific variation of monoene CHCs.

There should certainly be a follow up to the sequencing analysis of *desatF* incorporating all published genomes (CLARK *et al.* 2007), assessing the variation within coding as well the cis-regulatory regions. The inclusion of *D. mauritiana* would be interesting to see the extent of the sequence variation within the promoter regions of both monomorphic species within the *simulans* clade.

In the light of recent findings on the role of *desatF* in the expression of 5,9 dienes (LEGENDRE *et al.* 2008) and *eloF* in the expression of female specific long chain cuticular hydrocarbons (CHERTEMPS *et al.* 2007), perhaps a multiple precise gene targeting experiment (including *desat2, eloF* and *desatF*) from African to Cosmopolitan *D. melanogaster* populations might prove the importance of the interaction of multiple loci for the production of these diene CHCs. This may also explain why the first attempt of the transgenic insertion of *desat2* (GREENBERG *et al.* 2003) did not result in the appropriate assortative mating found between these two

populations and perhaps why the replication of this experiment failed (COYNE and ELWYN 2006).

In chapter three the QTL analysis on mean IPI detected eight significant QTLs, most of which show a relatively small affect from all three chromosomes, whereas the MIM detected three specific QTL of a relatively intermediate size affect (on each chromosome). The MIM implicated *fru* as the most influential marker situated on the third chromosome, associated with interspecific variation of mean IPI. However a *D. sechellia* allele present at the QTL associated with *fru* caused shorter mean IPI's, which is the opposite to the expected direction of affect. However this QTL could be one of a number of other genes situated within the genomic region between the markers *nos* and *fru*. The marker *Dga* (47A) was associated with the most influential QTL on the second chromosome and a *D. sechellia* allele present at this locus resulted in longer IPIs, this is the appropriate direction for the *D. sechellia* phenotype and this marker is most closely association with the candidate gene *croaker* (45E). The *croaker* allele is known to affect the pulse cycles of courtship song and flight (YOKOKURA *et al.* 1995). The epistatic interaction involving the QTL on the second chromosome with the marker *forked* on the X-chromosome suggests that there may be an interaction involving the other candidate gene *maleless* (42A6) and a QTL or QTLs situated on the X-chromosome. This is speculative, however scoring *maleless* may further validate this hypothesis. Intriguingly the significant QTL on the X-chromosome is involved in both epistatic interactions affecting mean IPI (involving *fru* and *Dga*) and this same region (associated with the marker *forked*) is also influential on the interspecific difference of 7-tricosene. The marker *forked* (15F4-9) is not closely associated with any candidate genes involved with affecting mean IPI or

pheromone profiles within *D. melanogaster*, however it is located very close to that of *OdsH* (16D1) a known "speciation" gene (expressed in males) and involved with post-mating reproductive isolation (PEREZ and WU 1995; TING *et al.* 1998; WU and TING 2004). This genomic region on the X-chromosome is involved with both sexually dimorphic behavioural traits (pheromone and song) included in this study. Including markers around this region of the X-chromosome might be insightful.

The recombinant inbred line analysis on mean IPI showed that flies homozygous for *D. sechellia* at the *fru* locus had significantly longer IPIs. This may indicate underdominance of the heterozygote at the *fru* locus, since the QTL analysis found flies heterozygous at the QTL associated with the *fru* locus to have significantly shorter mean IPIs. This implies that there may be allelic variation at the *fru* locus influencing interspecific variation of this trait or the possibility of linked loci. The opposite was true for *dsx* homoygotes at the *D. sechellia* allele, since shorter mean IPIs were recorded. None of the RI lines tested were homozygous at both of these loci and in the light of recent findings, that the expression of both *fru* and *dsx* are necessary for the sexually dimorphic expression of male courtship song, testing their combined effects would be interesting. However this study was limited in the number of RI lines that tested positive for *D. sechellia* homozygotes at the sex determination loci. It would be worthwhile repeating this experiment, increasing the number of RI lines, and the chance of possibly finding a line homozygous for *D. sechellia* at both the *fru* and *dsx* loci. The combined affects of both genes could then be assessed on the interspecific difference in mean IPI.

The results from the RI line analysis on the interspecific difference in sex comb teeth between *D. simulans* and *D. sechellia*, showed all of the sex determination markers as significant, confirming that the region between *tra* (73A10) and *fru* (91A7-91B3) may be influential on this trait and also confirms previous analysis, that it is likely multiple loci are involved (GRAZE *et al.* 2007; MACDONALD and GOLDSTEIN 1999; TATSUTA and TAKANO-SHIMIZU 2006; TRUE *et al.* 1997).

Both QTL and RI line results strongly implicate that *fru* may be involved in the interspecific difference in mean IPI and the final data chapter carried out an analysis for positive selection on all of the *fru* coding regions. The results indicated significant selection at two of the zinc finger DNA binding domains. The positive selection detected at the ZnF-C domain was particularly interesting since this is present in one of the three male specific isoforms, and therefore may contribute to the interspecific differences found in courtship behaviour between the species included in this analysis. The high level of conservation of the BTB and ZnF protein coding domain sequences across many insect species is not as apparent at the predicted sites of the *fru* promoter regions, and more recent studies have shown that there appears to be considerably less sequence conservation within the cis-regulatory regions associated with the *fru* promoter regions P1-P4, suggesting that these cis-regulatory regions are not experiencing the same functional constraints (BILLETER and GOODWIN 2004; GAILEY *et al.* 2006; GOODWIN *et al.* 2000; LEE and HALL 2001). In this study an assessment of the sequence variation of the promoter regions of *fru* has not been made, though it is likely that both structural and regulatory sequence divergence may account for the variation in the expression of sexually dimorphic traits between different Drosophilid species. It is also probable that the target genes regulated by *fru* expressed in the CNS

involved with courtship traits may well have species-specific sequence differences at the coding and regulatory levels affecting their expression. Within *Drosophila melanogaster* the gene *yellow* controls the pattern of pigmentation within a number of body parts (CARROL 2005). The species-specific expression variation is controlled by the evolution of cis-regulatory elements (SIMPSON 2007). The *yellow* locus is also associated with wing extension during courtship and mating success, and studies on *yellow* indicate that the zinc-finger transcription factors encoded by *fru* bind to the cis regulatory regions, controlling its expression in the neuroblasts associated with male specific behaviour. In fact both *dsx* and *fru* are candidate genes for regulating this downstream target gene (DRAPEAU *et al.* 2006; DRAPEAU *et al.* 2003). Another known target gene of both *fru* and *dsx* is *takeout* which produces transcripts from the fat body tissue associated with the adult male brain and is also involved in normal sexual behaviour (DAUWALDER *et al.* 2002).

Future work should include a thorough assessment of the sequence variation in the promoter regions of *fru* as well as sequence analysis of variation with a number of terminal effector genes regulated by both *dsx* and *fru*. Ideally transgenic manipulation experiments are the best way to test the true affects that structural and regulatory mutations may have on phenotypic differences. However such experiments are technically difficult to carry out whilst maintaining the correct genetic background, which is especially important for regulatory genes such as *fru* considering the post-transcriptional interactions that shape the final phenotypic expression. However it is certainly possible for a new more extensive QTL analysis, incorporating markers based on findings from mutational screening, RNAi experiments and sequence analysis. A QTL implicated as having a major affect on interspecific variation of a

trait is rare. More often these large affect QTL are not acting in isolation and the apparent large affect is frequently due to the composite affect of several QTLs interacting, which often depends on a finely tuned genetic background.

References

AHMAD, K. F., C. K. ENGEL and G. G. PRIVE, 1998 Crystal structure of the BTB domain from PLZF. Proc Natl Acad Sci U S A **95:** 12123-12128.

AMREIN, H., and N. THORNE, 2005 Gustatory perception and behavior in Drosophila melanogaster. Curr. Biol **15:** 673–684.

ANAND, A., A. VILLELLA, L. C. RYNER, T. CARLO, S. F. GOODWIN *et al.*, 2001 Molecular Genetic Dissection of the Sex-Specific and Vital Functions of the Drosophila melanogaster Sex Determination Gene fruitless. Genetics **158:** 1569-1595.

ANISIMOVA, M., J. P. BIELAWSKI and Z. YANG, 2001 Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. Mol Biol Evol **18:** 1585-1592.

ARNQVIST, G., M. EDVARDSSON, U. FRIBERG and T. NILSSON, 2000 Sexual conflict promotes speciation in insects. Proc Natl Acad Sci U S A **97:** 10460-10464.

ASHBURNER, M., 1989 Drosophila: A Laboratory Manual. Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY.

BAKER, B., B. TAYLOR and J. HALL, 2001 Are Complex Behaviors Specified by Dedicated Regulatory Genes? Reasoning from Drosophila. Cell **105:** 13-24.

BAKER, B. S., 1989 Sex in flies: the splice of life. Nature **340:** 521-524.

BAKER, B. S., and K. A. RIDGE, 1980 Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster. Genetics **94:** 383-423.

BARTON, N. H., and B. CHARLESWORTH, 1984 Genetic Revolutions, Founder Effects, and Speciation. Annual Review of Ecology and Systematics **15:** 133-164.

BASTEN, C. J., B. S. WEIR and Z.-B. ZENG, 1997 *QTL Cartographer: A reference manual and tutorial for QTL mapping.* North Carolina State University, Raleigh, NC.

BEN-SHAHAR, Y., A. ROBICHON, M. B. SOKOLOWSKI and G. E. ROBINSON, 2002 Influence of gene action across different time scales on behavior. Science **296:** 741-744.

BENNET-CLARK, H. C., and A. W. EWING, 1969 Pulse Interval as a critical parameter in the courtship song of D. melanogaster Animal behaviour **17:** 755-759.

BERNSTEIN, A. S., E. K. NEUMANN and J. C. HALL, 1992 Temporal analysis of tone pulses within the courtship songs of two sibling Drosophila species, their interspecific hybrid, and behavioral mutants of D. melanogaster (Diptera: Drosophilidae). J. Insect Behav **5:** 15-36.

BILLETER, J.-C., S. F. GOODWIN and K. M. C. O'DELL, 2002 Genes mediating sex-specific behaviors in Drosophila. Genetics **47:** 87-115.

BILLETER, J. C., E. RIDEOUT, A. DORNAN and S. GOODWIN, 2007 Control of Male Sexual Behavior in Drosophila by the Sex Determination Pathway. Current Biology **16:** 766-776.

BILLETER, J. C., A. VILLELLA, J. B. ALLENDORFER, A. J. DORNAN, M. RICHARDSON *et al.*, 2006 Isoform-specific control of male neuronal differentiation and behavior in Drosophila by the fruitless gene. Curr Biol **16:** 1063-1076.

BIRNEY, E., M. CLAMP and R. DURBIN, 2004 GeneWise and Genomewise. Genome Res **14:** 988-995.

BLOMQUIST, G. J., J. W. DILLWORTH and T. S. ADAMS, 1987 *Biosynthesis and endochrine regulation of sex pheomone production in Diptera*. Academic Press, London.

BOAKE, C. R., 2005 Sexual selection and speciation in Hawaiian *Drosophila* Behav Genet **35:** 297-303.

BOAKE, C. R. B., S. J. ARNOLD and F. BREDEN, 2002 Genetic tools for studying adaptation and the evolution of behavior. American Naturalist **160:** S143–S159.

BRADSHAW, H. D., J. K. G. OTTO, B. E. FREWEN, J. K. MCKAY and D. W. SCHEMSKE, 1998 Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (Mimulus). Genetics **149:** 367-382.

BURTIS, K. C., 1993 The regulation of sex determination and sexually dimorphic differentiation in Drosophila. Curr Opin Cell Biol **5:** 1006-1014.

BURTIS, K. C., and B. S. BAKER, 1989 Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell **56:** 997-1010.

BUSTAMANTE, C. D., A. FLEDEL-ALON, S. WILLIAMSON, R. NIELSEN, M. T. HUBISZ *et al.*, 2005 Natural selection on protein-coding genes in the human genome. Nature **437:** 1153-1157.

BUTLIN, R. K., and M. G. RITCHIE, 1994 Mating behaviour and speciation, pp. 43–79 in *Behaviour and Evolution,*, edited by P. B. J. SLATER and T. R. HALLIDAY. CAMBRIDGE UNIVERSITY PRESS, UK.

CAMPESAN, S., Y. DUBROVA, J. C. HALL and C. P. KYRIACOU, 2001 The nonA gene in Drosophila conveys species-specific behavioral characteristics. Genetics **158:** 1535-1543.

CARROL, S. B., 2005 Evolution at Two Levels: On Genes and Form. Plos Biology **3:** 1159-1166.

CARSON, H. L., and A. R. TEMPLETON, 1984 Genetic Revolutions in Relation to Speciation Phenomena: The Founding of New Populations. Annual Review of Ecology and Systematics **15:** 97-132.

CHAN YONG, T. P., and J.-M. JALLON, 1986 Synthèse de novo d'hydrocarbures potentiellement aphrodisiaques ches les Drosophile. C. r: hebd. Séanc. Acad. Sci. **303:** 197–202.

CHEN, E. H., and B. S. BAKER, 1997 Compartmental organization of the Drosophila genital imaginal discs. Development **124:** 205-218.

CHERTEMPS, T., L. DUPORTETS, C. LABEUR, R. UEDA, K. TAKAHASHI *et al.*, 2007 A female-biased expressed elongase involved in long-chain hydrocarbon biosynthesis and courtship behavior in Drosophila melanogaster. Proc Natl Acad Sci U S A **104:** 4273-4278.

CHERTEMPS, T., L. DUPORTETS, C. LABEUR, M. UEYAMA and C. WICKER-THOMAS, 2006 A female-specific desaturase gene responsible for diene hydrocarbon biosynthesis and courtship behaviour in Drosophila melanogaster. Insect Molecular Biology **15:** 465-473.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963-971.

CIVETTA, A., and R. S. SINGH, 1998 Sex and speciation: Genetic architecture and evolutionary potential of sexual versus nonsexual traits in the sibling species of the Drosophila melanogaster complex. Evolution **52:** 1080-1092.

CLARK, A. G., M. B. EISEN, D. R. SMITH, C. M. BERGMAN, B. OLIVER *et al.*, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature **450:** 203-218.

CLINE, T. W., and B. J. MEYER, 1996 Vive la difference: males vs females in flies vs worms. Annu Rev Genet **30:** 637-702.

COBB, M., and J.-F. FERVEUR, 1996 Evolution and genetic control of mate recognition and stimulation in Drosophila. Behav. Proc. **35:** 35–54.

COBB, M., and J.-M. JALLON, 1990 Pheromones, mate recognition and courtship stimulation in the *Drosophila melanogaster* species sub-group. Anim. Behav. **39:** 1058-1067.

COLEGRAVE, N., H. HOLLOCHER, K. HINTON and M. G. RITCHIE, 2000 The courtship song of African *Drosophila melanogaster*. Journal of Evolutionary Biology **13:** 143-150.

COOK, R., 1979 The courtship tracking of Drosophila melanogaster. Biol. Cybern **34:** 91–106.

COOK, R. M., 1977 Behavioral role of the sexcombs in Drosophila melanogaster and Drosophila simulans. Behav Genet **7:** 349-357.

COYNE, J. A., 1985 Genetic studies of three sibling species of Drosophila with relationship to theories of speciation. Genet Res **46:** 169-192.

COYNE, J. A., A. P. CRITTENDEN and K. MAH, 1994 Genetics of a pheromonal difference contributing to reproductive isolation in Drosophila. Science **265:** 1461-1464.

COYNE, J. A., and S. ELWYN, 2006 Does the desaturase-2 locus in Drosophila melanogaster cause adaptation and sexual isolation? Evolution **60:** 279-291.

COYNE, J. A., S. Y. KIM, A. S. CHANG, D. LACHAISE and S. ELWYN, 2002 Sexual isolation between two sibling species with overlapping ranges: Drosophila santomea and Drosophila yakuba. Evolution **56:** 2424-2434.

COYNE, J. A., C. WICKER-THOMAS and J. M. JALLON, 1999 A gene responsible for a cuticular hydrocarbon polymorphism in Drosophila melanogaster. Genet Res **73:** 189-203.

DALLERAC, R., C. LABEUR, J. M. JALLON, D. C. KNIPPIE, W. L. ROELOFS *et al.*, 2000 A Delta 9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America **97:** 9449-9454.

DAUWALDER, B., S. TSUJIMOTO, J. MOSS and W. MATTOX, 2002 The Drosophila takeout gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. Genes Dev. **16:** 2879–2892.

DAVIS, A. W., E. G. NOONBURG and C. I. WU, 1994 Evidence for Complex Genic Interactions between Conspecific Chromosomes Underlying Hybrid Female Sterility in the Drosophila Simulans Clade. Genetics **137(1):** 191–199.

DAVIS, T., J. KURIHARA, E. YOSHINO and D. YAMAMOTO, 2000 Genomic organisation of the neural sex determination gene fruitless (fru) in the Hawaiian species Drosophila silvestris and the conservation of the fru BTB protein-protein-binding domain throughout evolution. Hereditas **132:** 67-78.

DEMIR, E., and B. J. DICKSON, 2005 Fruitless splicing specifies male courtship behaviour in Drosophila Cell **121:** 785-794.

DOBZHANSKY, T., 1937 *Genetics and the Origin of Species*. Columbia University Press, New York.

DRAPEAU, M. D., S. A. CYRAN, M. M. VIERING, P. K. GEYER and A. D. LONG, 2006 A cis-regulatory sequence within the yellow locus of Drosophila melanogaster required for normal male mating success. Genetics **172:** 1009-1030.

DRAPEAU, M. D., A. RADOVIC, P. J. WITTKOPP and A. D. LONG, 2003 A gene necessary for normal male courtship, yellow, acts downstream of fruitless in the Drosophila melanogaster larval brain. J Neurobiol **55:** 53-72.

ETGES, W. J., and M. A. AHRENS, 2001 Premating isolation is determined by larval-rearing substrates in cactophilic Drosophila mojavensis. V. Deep geographic variation in epicuticular hydrocarbons among isolated populations. American Naturalist **158:** 585-598.

ETGES, W. J., and L. L. JACKSON, 2001 Premating isolation is determined by larval rearing substrates in cactophilic Drosophila mojavensis part 6 - Epicuticular hydrocarbon variation in Drosophila mojavensis cluster species. Journal of Chemical Ecology **27:** 2125-2149.

FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction To Quantitative Genetics*. Longmans Green, Harlow, Essex, UK.

FANG, S., A. TAKAHASHI and C.-I. WU, 2002 A Mutation in the Promoter of Desaturase 2 Is Correlated With Sexual Isolation Between Drosophila Behavioral Races. Genetics **162:** 781-784.

FARRALL, M., 2004 Quantitative genetic variation: a post-modern view. Hum. Mol. Genet. **13:** R1-7.

FERVEUR, J. F., 2005 Cuticular hydrocarbons: Their evolution and roles in Drosophila pheromonal communication. Behavior Genetics **35:** 279-295.

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection.* Oxford,UK:Oxford University Press.

FISHER, R. A., 1931 The evolution of dominance. Biological Reviews **6:** 345–368.

FISHER, R. A., 1958 *The genetical theory of natural selection*. Dover, New York.

FITZPATRICK, M. J., 2004 Pleiotropy and the Genomic Location of Sexually Selected Genes. American Naturalist **163:** 800-808.

FITZPATRICK, M. J., Y. BEN-SHAHAR, H. M. SMID, L. E. VET, G. E. ROBINSON *et al.*, 2005 Candidate genes for behavioural ecology. Trends Ecol Evol **20:** 96-104.

FLYBASE, 2003 The FlyBase database of the Drosophila genome projects and community literature. Nucleic Acids Res **31:** 172-175.

GAGE, M. J., G. A. PARKER, S. NYLIN and C. WIKLUND, 2002 Sexual selection and speciation in mammals, butterflies and spiders. Proc Biol Sci **269:** 2309-2316.

GAILEY, D. A., J. C. BILLETER, J. H. LIU, F. BAUZON, J. B. ALLENDORFER *et al.*, 2006 Functional conservation of the fruitless male sex-determination gene across 250 Myr of insect evolution. Mol Biol Evol **23:** 633-643.

GAILEY, D. A., K. H. SUN, S. OHSHIMA, J. H. LIU, M. EYASSU *et al.*, 2000 A phylogeny of the Drosophilidae using the sex-behaviour gene *fruitless*. Hereditas **133:** 81-33.

GAILEY, D. A., B. J. TAYLOR and J. C. HALL, 1991 Elements of the fruitless locus regulate development of the muscle of Lawrence, a male-specific structure in the abdomen of Drosophila melanogaster adults. Development **113:** 879-890.

GARDINER, A., D. BARKER, R. K. BUTLIN, W. C. JORDAN and M. G. RITCHIE, 2008 Drosophila chemoreceptor gene evolution: selection, specialization and genome size. Molecular Ecology **17:** 1648-1657.

GENTILE, G., A. DELLA TORRE, B. MAEGGA, J. R. POWELL and A. CACCONE, 2002 Genetic differentiation in the African malaria vector, Anopheles gambiae s.s., and the problem of taxonomic status. Genetics **161:** 1561-1578.

GERTZ, E. M., Y. YU, R. AGARWALA, A. A. SCHÄFFER and S. F. ALTSCHUL, 2006 Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol **4:** 41-55.

GLEASON, J. M., 2005 Mutations and natural genetic variation in the courtship song of Drosophila. Behav Genet **35:** 265-277.

GLEASON, J. M., J. M. JALLON, J. D. ROUAULT and M. G. RITCHIE, 2005 Quantitative trait loci for cuticular hydrocarbons associated with sexual isolation between Drosophila simulans and D. sechellia. Genetics **171:** 1789-1798.

GLEASON, J. M., S. V. NUZHDIN and M. G. RITCHIE, 2002 Quantitative trait loci affecting a courtship signal in Drosophila melanogaster. Heredity **89:** 1-6.

GLEASON, J. M., and M. G. RITCHIE, 2004 Do quantitative trait loci (QTL) for a courtship song difference between Drosophila simulans and D. sechellia coincide with candidate genes and intraspecific QTL? Genetics **166:** 1303-1311.

GLOOR, G., and W. ENGELS, 1992 Single-fly DNA preps for PCR. Dros. Inf. Serv **71:** 148-149.

GOODWIN, S. F., B. J. TAYLOR, A. VILLELLA, M. FOSS, L. C. RYNER *et al.*, 2000 Aberrant splicing and altered spatial expression patterns in fruitless mutants of Drosophila melanogaster. Genetics **154:** 725-745.

GORCZYCA, M., and J. C. HALL, 1987 The Insectavox, an integrated device for recording and amplifying courtship songs of Drosophila. Dros. Inf. Serv **66:** 157-160.

GRATTEN, J., D. BERALDI, B. V. LOWDER, A. F. MCRAE, P. M. VISSCHER *et al.*, 2007 Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. Proceedings of the Royal Society B-Biological Sciences **274:** 619-626.

GRAZE, R. M., O. BARMINA, D. TUFTS, E. NADERI, K. L. HARMON *et al.*, 2007 New candidate genes for sex comb divergence between Drosophila mauritiana and Drosophila simulans. Genetics **176**: 2561-2576

GREENBERG, A. J., J. R. MORAN, J. A. COYNE and C. I. WU, 2003 Ecological adaptation during incipient speciation revealed by precise gene replacement. Science **302:** 1754-1757.

GREENBERG, A. J., J. R. MORAN and C. I. WU, 2006 Proper control of genetic background with precise allele substitution: A comment on Coyne and Elwyn. Evolution **60:** 623-625.

GREENSPAN, R. J., and J.-F. FERVEUR, 2000 Courtship in Drosophila. Annu. Rev. Genet. **34** 205–232.

GRIFFITHS, A. J. F., W. M. GELBART, J. H. MILLER and R. C. LEWONTIN, 2000 *Modern Genetic Ananlysis*. W.H. Freeman and company.

GROOT, A. T., J. L. HOROVITZ, J. HAMILTON, R. G. SANTANGELO, C. SCHAL *et al.*, 2006 Experimental evidence for interspecific directional selection on moth pheromone communication. Proceedings of the National Academy of Sciences of the United States of America **103:** 5858-5863.

HAAG, E. S., and J. R. TRUE, 2001 Perspective: From mutants to mechanisms? Assessing the candidate gene paradigm in evolutionary biology. Evolution Int J Org Evolution **55:** 1077-1084.

HALDANE, J. B. S., 1922 Sex ratio and unisexual sterility in hybrid animals. J. Genet **12:** 101-109.

HEINRICHS, V., L. C. RYNER and B. S. BAKER, 1998 Regulation of sex-specific selection of fruitless 5' splice sites by transformer and transformer-2. Mol Cell Biol **18:** 450-458.

HENDERSON, N. R., and D. M. LAMBERT, 1982 No significant deviation from random mating of worldwide populations of Drosophila melanogaster. Nature **300:** 437-440.

HEY, J., and R. M. KLIMAN, 1993 Population genetics and phylogenetics of DNA sequence variation at multiple loci within the Drosophila melanogaster species complex. Mol Biol Evol **10:** 804-822.

HOEKSTRA, H. E., and J. A. COYNE, 2007 The locus of evolution: evo devo and the genetics of adaptation. Evolution Int J Org Evolution **61:** 995-1016.

HOIKKALA , A., S. PAALLYSAHO, J. ASPI and J. LUMME, 2000 Localization of genes affecting species differences in male courtship song between Drosophila virilis and D. littoralis. Genet. Res., Camb. **75:** 37-45.

HOLZER, B., A. JACOT and M. W. G. BRINKHOF, 2003 Condition-dependent signaling affects male sexual attractiveness in field crickets, Gryllus campestris. Behav. Ecol. **14:** 353-359.

HSIA, C. C., and W. MCGINNIS, 2003 Evolution of transcription factor function. Curr Opin Genet Dev **13:** 199-206.

HUTTUNEN, S., J. ASPI, A. HOIKKALA and C. SCHLÖTTERER, 2004 QTL analysis of variation in male courtship song characters in Drosophila virilis. Heredity **92:** 263–269.

HUTTENEN, S., C. VIERA and A. HOIKKALA, 2002 Nucleotide and repeat length variation at the nonA gene of the Drosophila virilis group species and its effects on male courtship song. Genetica **115:** 159-167.

HUYNH, K. D., and V. J. BARDWELL, 1998 The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. Oncogene **17:** 2473-2484.

INGRAM, K. K., P. OEFNER and D. M. GORDON, 2005 Task-specific expression of the foraging gene in harvester ants. Mol Ecol **14:** 813-818.

ITO, H., K. FUJITANI, K. USUI, K. SHIMIZU-NISHIKAWA, S. TANAKA *et al.*, 1996 Sexual orientation in Drosophila is altered by the satori mutation in the sex-determination gene fruitless that encodes a zinc finger protein with a BTB domain. Proceedings of the National Academy of Sciences **93:** 9687-9692.

JALLON, J.-M., and J. R. DAVID, 1987 Variations in cuticular hydrocarbons among the eight species of the Drosophila melanogaster subgroup. Evolution **41:** 294–302.

JALLON, J.-M., and C. WICKER-THOMAS, 2003 Genetic studies on pheromone production in *Drosophila*, pp. 253-281 in *Insect Pheromone Biochemostry and Molecular Biology*, edited by G. J. BLOMQVIST and R. G. VOGT. Elsevier.

JALLON, J. M., G. LAUGE , L. ORSSAUD and C. ANTONY, 1988 Female pheromones in Drosophila melanogaster are controlled by the doublesex locus. Genetical Research **51:** 17-22.

JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136:** 1447-1455.

JOLY, D., BAZIN, C., ZENG, L-W. & SINGH, R. S., 1997 Genetic basis of sperm and testis length differences and epistatic effect on hybrid inviability and sperm motility between *Drosophila simulans* and *D. sechellia.* Heredity **78:** 354-362.

KANESHIRO, K. Y., 1989 *The dynamics of sexual selection and founder effects in species formation*. Oxford University Press, New York.

KANESHIRO, K. Y., and C. R. BOAKE, 1987 Sexual selection and speciation: issues raised by Hawaiian Drosophila. Trends Ecol Evol **2:** 207–212.

KAWANISHI, M., and T. K. WATANABE, 1980 Genetic variations of courtship song of Drosophila melanogaster and D. simulans. Jpn. J. Genet. **55:** 235-240.

KERSEY, M. J., 1988 The principles of QTL analysis (a minimal mathematics approach). Journal of Experimental Botany **49:** 1619-1623.

KIM, S. C., and L. H. RIESEBERG, 1999 Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. Genetics **153:** 965-977.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, , Cambridge.

KOEPFER, H. R., and E. J. FENSTER, 1991 Asymmetrical Mating Patterns between Geographic Strains of Drosophila mercatorum: A Test of the Kaneshiro Hypothesis. Evolution **45:** 455-458.

KONDRASHOV, A. S., and M. SHPAK, 1998 On the origin of species by means of assortative mating. Proc Biol Sci **265:** 2273-2278.

KOPP, A., R. M. GRAZE, S. XU, S. B. CARROLL and S. V. NUZHDIN, 2003 Quantitative trait loci responsible for variation in sexually dimorphic traits in Drosophila melanogaster. Genetics **163:** 771-787.

KOROL, A., E. RASHKOVETSKY, K. ILIADI and E. NEVO, 2006 Drosophila flies in "Evolution Canyon" as a model for incipient sympatric speciation. Proc Natl Acad Sci U S A **103:** 18184-18189.

LABEUR, C., R. DALLERAC and C. WICKER-THOMAS, 2002 Involvement of desat1 gene in the control of Drosophila melanogaster pheromone biosynthesis. Genetica **114:** 269-274.

LACHAISE, D., M.-L. CARIOU, J. R. DAVID, F. LEMEUNIER and L. TSACAS, 1988 Historical biogeography of the Drosophila melanogaster species subgroup. Evol. Biol **22:** 159-225.

LAM, B. J., A. BAKSHI, F. Y. EKINCI, J. WEBB, B. R. GRAVELEY *et al.*, 2003 Enhancer-dependent 5'-Splice Site Control of fruitless Pre-mRNA Splicing. J. Biol. Chem **278:** 22740-22747.

LANDE, R., 1981 Models of Speciation by Sexual Selection on Polygenic Traits. PNAS **78:** 3721-3725.

LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. J. DALY *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174-181.

LANDRY, C. R., P. J. WITTKOPP, C. H. TAUBES, J. M. RANZ, A. G. CLARK *et al.*, 2005 Compensatory cis-trans Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of Drosophila. Genetics **171:** 1813-1822.

LEE, C.-G., K. A. CHANG, M. I. KURODA, and J. HURWITZ, 1997 The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation. EMBO J **16:** 2671-2681.

LEE, G., M. FOSS, S. F. GOODWIN, T. CARLO, B. J. TAYLOR *et al.*, 2000 Spatial, temporal, and sexually dimorphic expression patterns of the fruitless gene in the Drosophila central nervous system. J Neurobiol **43:** 404-426.

LEE, G., and J. C. HALL, 2001 Abnormalities of male-specific FRU protein and serotonin expression in the CNS of fruitless mutants in Drosophila. J Neurosci **21:** 513-526.

LEGENDRE, A., X.-X. MIAO, J.-L. DA LAGE and C. WICKER-THOMAS, 2008 Evolution of a desaturase involved in female pheromonal cuticular hydrocarbon biosynthesis and courtship behavior in Drosophila. Insect Biochemistry and Molecular Biology **38:** 244-255.

LEMEUNIER, F. D. J., TSACAS L, ASHBURNER M (Editor), 1986 *The melanogaster species group. In: Ashburner M, Carson HL, Thompson JN (eds)* Academic Press New York.

LEVINE, M., and R. TJIAN, 2003 Transcription regulation and animal diversity. Nature **424:** 147–151.

LI, X., H. PENG, D. C. SCHULTZ, J. M. LOPEZ-GUISA, F. J. RAUSCHER, 3RD *et al.*, 1999 Structure-function studies of the BTB/POZ transcriptional repression domain from the promyelocytic leukemia zinc finger oncoprotein. Cancer Res **59:** 5275-5282.

LINDHOLM, A., and F. BREDEN, 2002 Sex chromosomes and sexual selection in poecilid fishes. American Naturalist **160:** S214–S224.

LOUIS, J. D., and J. R. David 1986 Ecological specialization in the Drosophila melanogaster species subgroup: A case study of D. sechellia. Acta Oecol. **7:** 215-229.

LYNCH, M., and J. S. CONERY, 2000 The evolutionary fate and consequences of duplicate genes. Science **290:** 1151-1155.

MACDONALD, S. J., and D. B. GOLDSTEIN, 1999 A quantitative genetic analysis of male sexual traits distinguishing the sibling species Drosophila simulans and D. sechellia. Genetics **153:** 1683-1699.

MACHADO, C. A., R. M. KLIMAN, J. A. MARKERT and J. HEY, 2002 Inferring the history of speciation from multilocus DNA sequence data: the case of Drosophila pseudoobscura and close relatives. Mol Biol Evol **19:** 472-488.

MACKAY, T. F., 2001 Quantitative trait loci in Drosophila. Nat Rev Genet **2:** 11-20.

MACKAY, T. F. C., 1995 The genetic-basis of quantitative variation: numbers of sensory bristles of Drosophila melanogaster as a model system. Trends in Genetics **11:** 464-470.

MANNING, A., 1959 The sexual isolation between Drosophila melanogaster and Drosophila simulans. Anim. Behav. **7:** 60–65.

MARCILLAC, F., Y. GROSJEAN and J. F. FERVEUR, 2005 A single mutation alters production and discrimination of Drosophila sex pheromones. Proc Biol Sci **272:** 303-309.

MARKOW, T. A., and P. O'GRADY, 2005 Evolutionary Genetics of Reproductive Behavior in Drosophila: Connecting the Dots. Annu. Rev. Genet. **39:** 263–291.

MAYR, E., 1963 *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.

MCKEOWN, M., 1992 Sex differentiation: The role of alternative splicing Genetics and Development **2:** 299-304.

MEIKLEJOHN, C. D., J. PARSCH, J. M. RANZ and D. L. HARTL, 2003 Rapid evolution of male-biased gene expression in Drosophila. Proc Natl Acad Sci U S A **100:** 9894-9899.

MELNICK, A., K. F. AHMAD, S. ARAI, A. POLINGER, H. BALL *et al.*, 2000 In-depth mutational analysis of the promyelocytic leukemia zinc finger BTB/POZ domain reveals motifs and residues required for biological and transcriptional functions. Mol Cell Biol **20:** 6550-6567.

MILLER, C. W., and A. J. MOORE, 2007 A potential resolution to the lek paradox through indirect genetic effects. Proc Biol Sci **274:** 1279-1286.

MULLER, H. J., 1942 Isolating mechanisms, evolution and temperature. Biol. Symp. **6:** 71–125.

NAGOSHI, R., M. MCKEOWN, K. BURTIS, J. M. BELOTE and B. S. BAKER, 1988 The Control of alternative splicing at the genes regulating sexual differentiation in D. melanogaster. Cell **53:** 229-236.

NAGOSHI, R. N., and B. S. BAKER, 1990 Regulation of sex-specific RNA splicing at the Drosophila doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. Genes Dev. **4:** 89-97.

NAISBIT, R. E., C. D. JIGGINS, M. LINARES, C. SALAZAR and J. MALLET, 2002 Hybrid sterility, Haldane's rule and speciation in Heliconius cydno and H. melpomene. Genetics **161:** 1517-1526.

NG, C.S., and A. KOPP, 2008 Sex Combs are Important for Male Mating Success in Drosophila melanogaster. Behav Genet **38:** 195-201.

NOOR, M. A. F., 2003 Genes to make new species. Nature **423:** 699-700.

NUZHDIN, S. V., and S. G. REIWITCH, 2000 Are the same genes responsible for intra- and interspecific variability for sex comb tooth number in Drosophila? Heredity **84 (Pt 1):** 97-102.

NUZHDIN, S. V., M. L. WAYNE, K. L. HARMON and L. M. MCINTYRE, 2004 Common pattern of evolution of gene expression level and protein sequence in Drosophila. Mol Biol Evol **21:** 1308-1317.

ORR, H. A., 1995 The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics **139:** 1805-1813.

ORR, H. A., 1998 The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution **52:** 935-949.

ORR, H. A., and S. IRVING, 2001 Complex epistasis and the genetic basis of hybrid sterility in the Drosophila pseudoobscura Bogota-USA hybridization. Genetics **158:** 1089-1100.

ORR, H. A., J. P. MASLY and D. C. PRESGRAVES, 2004 Speciation genes. Current Opinion in Genetics & Development **14:** 675-679.

PANHUIS, T. M., R. BUTLIN, M. ZUK and T. TREGENZA, 2001 Sexual selection and speciation. Trends Ecol Evol **16:** 364-371.

PATTATUCCI, A. M., and T. C. KAUFMAN, 1991 The homeotic gene sex combs reduced of Drosophila melanogaster is differentially regulated in the embryonic and imaginal stages of development. Genetics **129:** 443-461.

PEREZ, D. E., and C. I. WU, 1995 Further Characterization of the Odysseus Locus of Hybrid Sterility in Drosophila: One Gene is Not Enough. Genetics **140:** 201-206.

PETRIE, M., and G. ROBERTS, 2007 Sexual selection and the evolution of evolvability. Heredity **98:** 198-205.

POLAK, M., W. T. STARMER and L. L. WOLF, 2004 Sexual selection for size and symmetry in a diversifying secondary sexual character in Drosophila bipectinata Duda (Diptera: Drosophilidae). Evolution **58 (3):** 597-607.

POMIANKOWSKI, A., and A. P. MØLLER, 1995 A Resolution of the Lek Paradox Proceedings: Biological Sciences **260:** 21-29.

PRESGRAVES, D. C., 2003 A fine-scale genetic analysis of hybrid incompatibilities in Drosophila. Genetics **163:** 955-972.

PRICE, C. S., C. H. KIM, J. POSLUSZNY and J. A. COYNE, 2000 Mechanisms of conspecific sperm precedence in Drosophila. Evolution Int J Org Evolution **54:** 2028-2037.

PRIVÉ, G. G., A. MELNICK, K. F. AHMAD and J. D. LICHT, 2005 *Zinc Finger Proteins From Atomic Contact to Cellular Function* Landes Biosciences, Georgetown, Texas.

PRÖSCHEL, M., Z. ZHANG and J. PARSCH, 2006 Widespread Adaptive Evolution of Drosophila Genes With Sex-Biased Expression. Genetics **174:** 893–900.

PURVES, W. K., G. H. ORIANS, C. H. HELLER and D. SADAVA, 2000 *Life the science of biology*. W.H. Freeman and company.

RAYMOND, C. S., C. E. SHAMU, M. M. SHEN, K. J. SEIFERT, B. HIRSCH *et al.*, 1998 Evidence for evolutionary conservation of sex-determining genes. Nature **391:** 691-695.

REEVE, H. K., and D. W. PFENNIG, 2002 Genetic biases for showy males: are some genetic systems especially conducive to sexual selection? . Proceedings of the National Academy of Sciences of the USA **100:** 1089–1094.

RICE, W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution **38:** 735–742.

RIDEOUT, E. J., J. C. BILLETER and S. F. GOODWIN, 2007 The sex-determination genes fruitless and doublesex specify a neural substrate required for courtship song. Curr Biol **17:** 1473-1478.

RITCHIE, M. G., E. J. HALSEY and J. M. GLEASON, 1999 Drosophila song as a species-specific mating signal and the behavioural importance of Kyriacou & Hall cycles in D. melanogaster song. Anim. Behav **58:** 649-657.

RITCHIE, M. G., and C. P. KYRIACOU, 1994 Reproductive isolation and the period gene of Drosophila. Mol Ecol **3:** 595-599.

RITCHIE, M. G., and C. P. KYRIACOU, 1996 Artificial selection for a courtship signal in Drosophila melanogaster. Anim. Behav **52:** 603-611.

ROELOFS, W. L., and A. P. ROONEY, 2003 Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. Proceedings of the National Academy of Sciences of the United States of America **100:** 14599-14599.

ROUAULT, J., P. CAPY and J. M. JALLON, 2000 Variations of male cuticular hydrocarbons with geoclimatic variables: an adaptative mechanism in Drosophila melanogaster ? Genetica **110:** 117-130.

RYAN, M. J., and A. STANLEY, 1993 Species Recognition and Sexual Selection as a Unitary Problem in Animal Communication. Evolution **47:** 647-657.

RYNER, L. C., S. F. GOODWIN, D. H. CASTRILLON, A. ANAND and A. VILLELLA, 1996 Control of male sexual behaviour and sexual orientation in Drosophila by the fruitless gene. Cell **87:** 1079-1089.

SANCHEZ, L., and I. GUERRERO, 2001 The development of the Drosophila genital disc. Bioessays **23:** 698-707.

SAVARIT, F., and J. F. FERVEUR, 2002 Genetic study of the production of sexually dimorphic cuticular hydrocarbons in relation with the sex-determination gene transformer in Drosophila melanogaster. Genet Res **79:** 23-40.

SAVARIT, F., G. SUREAU, M. COBB and J. F. FERVEUR, 1999 Genetic elimination of known pheromones reveals the fundamental chemical bases of mating and isolation in Drosophila. Proc. Natl Acad. Sci. USA **96:** 9015–9020.

SCHILTHUIZEN, M., 2000 Dualism and conflicts in understanding speciation. Bioessays **22:** 1134-1141.

SCHLUTER, D., 2000 Ecological character displacement in adaptive radiation. Am. Nat. **156:** S4-S16.

SEEHAUSEN, O., P. J. MAYHEW and J. J. M. VAN ALPHEN, 1999 Evolution of colour patterns in East African cichlid fish Journal of Evolutionary Biology **12:** 514-534.

SHANKLIN, J., E. WHITTLE and B. G. FOX, 1994 Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. Biochemistry **33:** 12787-12794.

SHOREY, H. H., 1962 Nature of The Sound Produced by Drosophila melanogaster during courtship. Science, NY **137:** 677-678.

SIEGMUND, T., and M. LEHMANN, 2002 The Drosophila Pipsqueak protein defines a new family of helix-turn-helix DNA-binding proteins. Dev Genes Evol **212:** 152-157.

SIMPSON, P., 2007 The stars and stripes of animal bodies: evolution of regulatory elements mediating pigment and bristle patterns in Drosophila. Trends Genet **23:** 350-358.

SINGH, S. R., and H. F. HOENIGSBERG, 2004 Reproductive isolation and modes of speciation in Drosophila. Genet.Mol.Res **3(2):** 213-250.

SONG, H.-J., J.-C. BILLETER, E. REYNAUD, T. CARLO, E. P. SPANA *et al.*, 2002 The fruitless gene is required for the proper formation of axonal tracts in the embryonic central nervous system of Drosophila Genetics **162:** 1703-1724

SPEITH, H. T., 1952 Mating Behaviour within the Genus Drosophila (Diptera). Am. Nat **99:** 395-474.

STAJICH, J. E., D. BLOCK, K. BOULEZ, S. E. BRENNER and S. A. CHERVITZ, 2002 The BioPerl toolkit: Perl modules for the life sciences. Genome Res **12:** 1611-1618.

SUN, S., C. TING and C.-I. WU, 2004 The normal function of a speciation gene, Odysseus, and its hybrid sterility effect. Science **305:** 81-83.

SUREAU, G., and J. F. FERVEUR, 1999 Co-adaptation of pheromone production and behavioural responses in Drosophila melanogaster males. Genet Res **74:** 129-137.

TAKAHASHI, A., S. TSAUR, J. A. COYNE and C.-I. WU, 2001 The nucleotide changes governing cuticular hydrocarbon variation and their evolution in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **98:** 3920-3925.

TAO, Y., and D. L. HARTL, 2003 Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana: III.Heterogeneous accumulation of hybrid incompatibilities, degree of dominance, and implications for haldanes rule. Evolution **57:** 2580-2598.

TATSUTA, H., and T. TAKANO-SHIMIZU, 2006 Genetic architecture of variation in sex-comb tooth number in Drosophila simulans. Genet. Res **87:** 93–107.

TAYLOR, B. J., and J. W. TRUMAN, 1992 Commitment of abdominal neuroblasts in Drosophila to a male or female fate is dependent on genes of the sex-determining hierarchy. Development **114:** 625-642.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS, 1997 The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research **25:** 4876-4882.

TILLMAN, J. A., S. J. SEYBOLD, A. J. RUSSELL and G. J. BLOMQUIST, 1999 Insect pheromones—an overview of biosynthesis and endocrine regulation. Insect Biochemistry and Molecular Biology **29:** 481–514.

TING, C.-T., S.-C. TSAUR, M.-L. WU and C.-I. WU, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282:** 1501-1504.

TOKUNAGA, C., 1962 Cell lineage and differentiation on the male foreleg of Drosophila melanogaster. Dev Biol **4:** 489-516.

TOMARU, M., M. DOI, H. HIGUCHI and Y. OGUMA, 2000 Courtship song recognition in the Drosophila melanogaster complex: hetero-specific songs make females receptive in D. melanogaster, but not in D. sechellia. Evolution **54:** 1286–1294

TOMPKINS, L., R. W. SIEGEL, D. A. GAILEY and J. C. HALL, 1983 Conditioned courtship in Drosophila and its mediation by association of chemical cues. Behav Genet **13:** 565-578.

TRUE, J. R., J. J. LIU, L. F. STAM, Z. B. ZENG and C. C. LAURIE, 1997 Quantitative genetic analysis of divergence in male secondary sexual traits between Drosophila simulans and Drosophila mauritiana. Evolution **51:** 816-832.

TSAUR, S.-C., C.-T. TING and C.-I. WU, 2001 Sex in Drosophila mauritiana: A Very High Level of Amino Acid Polymorphism in a Male Reproductive Protein Gene, Acp26Aa. Mol Biol Evol **18:** 22-26.

TURELLI, M., and H. A. ORR, 1995 The dominance theory of Haldane's rule. Genetics **140:** 389–402.

USTINOVA, J., and F. MAYER, 2006 Alternative Starts of Transcription, Several Paralogues, and Almost-Fixed Interspecific Differences of the Gene fruitless in a Hemimetabolous Insect. J Mol Evol **63:** 788–800.

USUI-AOKI, K., H. ITO, K. UI-TEI, K. TAKAHASHI and T. LUKACSOVICH, 2000 Formation of the male-specific muscle in female Drosophila by ectopic fruitless expression. Nat. Cell Biol **2:** 500-506.

VAN DE PEER, Y., and R. DE WACHTER, 1994 TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Bioinformatics **10:** 569-570.

VENARD, R., C. ANTONY and J.-M. JALLON, 1989 *Drosophila chemoreceptors*. Plenum,, New York.

VILLELLA, A., D. A. GAILEY, B. BERWALD, S. OSHIMA, P. T. BARNES *et al.*, 1997 Extended reproductive roles of the fruitless gene in Drosophila melanogaster revealed by behavioural analysis of new fru mutants. Genetics **147:** 1107–1130.

VILLELLA, A., and J. C. HALL, 1996 Courtship anomalies caused by doublesex mutations in Drosophila melanogaster. Genetics **143:** 331-344.

VON HELVERSEN, O., and D. VON HELVERSEN, 1994 *Forces driving coevolution of song and song recognition in grasshoppers*. In *Neural Basis of Behavioural Adaptations* (ed. K. Schildberger and N. Elsner), pp. 253–284.

VON SCHILCHER, F., 1977 A mutation which changes courtship song in Drosophila melanogaster. Behav. Genet. **7:** 251-259.

VON SCHILCHER, F., 1976 The function of pulse song and sine song in the courtship of Drosophila melanogaster. Anim Behav **24:** 622-625.

WALTHOUR, C. S., and S. W. SCHAEFFER, 1994 Molecular Population Genetics of Sex Determination Genes: The transformer Gene of Drosophila melanogaster. Genetics **136:** 1367-1372.

WATERBURY, J. A., L. L. JACKSON and P. SCHEDL, 1999 Analysis of the Doublesex Female Protein in Drosophila melanogaster: Role in Sexual Differentiation and Behavior and Dependence on Intersex. Genetics **152:** 1653-1667.

WEN, Y., D. NGUYEN, Y. LI and Z. C. LAI, 2000 The N-terminal BTB/POZ domain and C-terminal sequences are essential for Tramtrack69 to specify cell fate in the developing Drosophila eye. Genetics **156:** 195-203.

WHEELER, C. J., W. L. FIELDS and J. C. HALL, 1988 Spectral analysis of Drosophila courtship songs : D. melanogaster, D. simulans, and their interspecific hybrid. Behav. Genet **18:** 675-703.

WHEELER, D. A., C. P. KYRIACOU, M. L. GREENACRE, Q. YU, J. E. RUTILA *et al.*, 1991 Molecular transfer of a species-specific courtship behaviour from Drosophila simulans to Drosophila melanogaster. Science **251:** 1082-1085.

WICKER-THOMAS, C., C. HENRIET and R. DALLERAC, 1997 Partial characterization of a fatty acid desaturase gene in Drosophila melanogaster. Insect Biochem Mol Biol **27:** 963-972.

WICKER-THOMAS, C., and J. M. JALLON, 2000 Control of female pheromones in Drosophila melanogaster by homeotic genes. Genet Res **78:** 235-242.

WICKER, C., and J.-M. JALLON, 1995 Influence of ovary and ecdysteroids on pheromone biosynthesis in Drosophila melanogaster (Diptera: Drosphilidae). Europ. J. Ent **92:** 197–202.

WIENER, P., and M. W. FELDMAN, 1993 The effects of the mating system on the evolution of migration in a spatially heterogeneous population. Evol Ecol **7:** 251–269.

WILKINS, A. S., 1995 Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. Bioessays **17:** 71-77.

WILLIAMS, M. A., A. G. BLOUIN and M. A. NOOR, 2001 Courtship songs of Drosophila pseudoobscura and D. persimilis. II. Genetics of species differences. Heredity **86:** 68-77.

WITTKOPP, P. J., J. R. TRUE and S. B. CARROLL, 2002 Reciprocal functions of the Drosophila Yellow and Ebony proteins in the development and evolution of pigment patterns. Development **129:** 1849-1858.

WRIGHT, S., 1932 "The roles of mutation, inbreeding, crossbreeding and selection in evolution". Proceedings of the VI International Congress of Genetrics **1:** 356-366.

WU, C.-I., and C.-T. TING, 2004 Genes and speciation. Nature Review Genetics **5:** 114-122.

WU, C. I., and H. HOLLOCHER, 1998 *Subtle is nature: The genetics of species differentiation and speciation*. Oxford University Press, New York.

YAND, Z., W. S. W. WONG and R. NIELSEN, 2005 Bayes empirical inference of amino acid sites under positive selection. Mol Biol Evol **22:** 1107-1118.

YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences **13:** 555-556.

YAWSON, A. E., D. WEETMAN, M. D. WILSON and M. J. DONNELLY, 2007 Ecological zones rather than molecular forms predict genetic differentiation in the malaria vector Anopheles gambiae s.s. in Ghana. Genetics **175:** 751-761.

YOKOKURA, T., R. UEDA and D. YAMAMOTO, 1995 Phenotypic and molecular characterization of croaker, a new mating behavior mutant of Drosophila melanogaster. Jpn J Genet **70:** 103-117.

ZARKOWER, D., 2001 Establishing sexual dimorphism: conservation amidst diversity? Nat Rev Genet **2:** 175-185.

ZENG, Z.-B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457-1468.

ZENG, Z. B., C. H. KAO and C. J. BASTEN, 1999 Estimating the genetic architecture of quantitative traits. Genet Res **74:** 279-289.

ZHANG, Z., T. M. HAMBUCH and J. PARSCH, 2004 Molecular evolution of sex-biased genes in Drosophila. Mol Biol Evol **21:** 2130-2139.

ZHANG, Z., and J. PARSCH, 2005 Positive Correlation Between Evolutionary Rate and Recombination Rate in Drosophila Genes with Male-Biased Expression. Mol Biol **22:** 1945-1947.

Appendix

Table A2.1. Additional markers used and their cytological locations

Gene	Abbreviation	Location [a]	Size relation [b]
doublesex	*dsx*	84E5-84E6	Fnu4H I cuts sim
fruitless	*fru*	91A7-91B3	sech > sim
desat1	*desat1*	87B10-87B11	sim > sech
desat2	*desat2*	87B10-87B10	Hae II cuts sech
desatF	*desatF*	68A1-68A1	Nde I cuts sech

[a] Cytological locations were obtained fromFlybase(http://flybase.bio.indiana.edu) and are for *D. melanogaster*

[b] Size of PCR products *D. simulans* (sim) relative to *D. sechellia* (sech) or restriction enzymes used to digest PCR products

Table A2.2. Sequences of Primers for Scoring Candidate Genes

dsx	forward	5′-CCAACATTGAAGAAGCTTCC-3′	reverse	5′-GTCCACCCCCGTCATAGATA-3′
desatF	forward	5′-CCACCCAATACCAAGGACAC-3′	reverse	5′-GTGCCAGGACACATTGAGTG-3′
desat1	forward	5′-TTTATCAGAGGCACGCATTG-3′	reverse	5′-CTAAACAAATCGGCCGACAC-3'
desat2	forward	5′-TTTGCCTTCTAATCGGTTCC-3′	reverse	5′-TCCGAGAATTTGTGGTGGAC-3'
fru	forward	5′-TGTGCAAATCAGGGATAC-3′	reverse	5′-GCTCTGGCATAGTTTGTTTCG-3′

Table A2.3 PCR reaction contains the following:

0.3 pmole/ µl each primer (forward and reverse) 1X buffer 0.32 mM dNTP* 1.5 mM Mg Cl2* 5 ng to 1 µg DNA 0.2 U/µl Taq enzyme
Typical cycling conditions: First step denature the DNA 92°C for 2 minutes Then 30 cycles Denaturing temperature 92°C for 10 seconds Annealing temperature 52°C for 15 seconds Extension temperature 72°C for 1 minute

Figure A4.1 CLUSTAL W (1.83) multiple sequence alignment

```
yakuba          CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATCTGACC 60
santomea        CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATCTGACC 60
teissieri       CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATCTGACC 60
simulans        CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATTTGACC 60
sechellia       CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATTTGACC 60
mauritiana      CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATTTGACC 60
melanogaster    CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATTTGACC 60
erecta          CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATCTGACC 60
orena           CAAGGAGCGATGGACCAGCAATTCTGCTTGCGCTGGAACAATCATCCCACAAATCTGACC 60
                ************************************************** *****

yakuba          GGCGTGCTCACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
santomea        GGCGTGCTCACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
teissieri       GGCGTGCTCACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
simulans        GGCGTGCTCACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
sechellia       GGCGTGCTCACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
mauritiana      GGCGTGCTCACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
melanogaster    GGCGTGCTAACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
erecta          GGCGTGCTCACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
orena           GGCGTGCTCACCTCACTGCTGCAGCGGGAGGCGCTATGCGACGTCACGCTCGCCTGCGAG 120
                ******** ************************************************

yakuba          GGCGAAACAGTCAAGGTGCGTCTCTG---------AGATACATTTGGAAGTATATATGTA 171
santomea        GGAGAAACAGTCAAGGTGCGTCTCTG---------AGATACATTTGGAAGTATATATGTA 171
teissieri       GGCGAAACAGTCAAGGTGCGTCTATG---------AGATACTTTTGGAAGTATATA---- 167
simulans        GGCGAAACAGTCAAGGTGCGTCTCTG---------AGATACACTTGCAGATATAGATGTA 171
sechellia       GGCGAAACAGTCAAGGTGCGTCTCTG---------AGATACACTCGAAGATATAGGTGTA 171
mauritiana      GGCGAAACAGTCAAGGTGCGTCTCTG---------AGATACACTTGAAGATATAGATGTA 171
melanogaster    GGCGAAACAGTCAAGGTGCGTCCATG---------AGATACACTTAAAGATATAGATATA 171
erecta          GGCGAAACAGTCAAGGTGCGTCTCCGTCTCCGGATGGACATATTTGAAAACACAT-TGTA 179
orena           GGCGAAACAGTCAAGGTGCGTCCCTG-----GAATAGACATATTTGAAAACATGC-TGTA 174
                ** ******************   *          ** *   *   *   *

yakuba          CATAGCTGACATAATT-CGTATTCTATC-----ACAGGCTCACCAGACCATCCTGTCAGC 225
santomea        CATATCTGACATAATT-CGTATTCTATC-----ACAGGCTCACCAGACCATCCTGTCAGC 225
teissieri       CATAGCTGACATAATT-CATATTCTTCC-----ACAGGCTCACCAGACCATCCTGTCAGC 221
simulans        CATATCTGACATTATTTCGTATTCATTCTAATCGCAGGCTCACCAGACCATCCTGTCAGC 231
sechellia       CATATCTGACATTATTTCATATTCATTCTAATCGCAGGCTCACCAGACCATCCTGTCAGC 231
mauritiana      CATATCTGACATTATTTCATATTCATTCTAATCGCAGGCTCACCAGACCATCCTGTCAGC 231
melanogaster    CATATCTGACATTATTTCATATTCATTCTA-TTGCAGGCTCACCAGACCATCCTGTCAGC 230
erecta          CTTATCCGACATAAT--CACTTTCTGTC-----ACAGGCTCACCAGACCATCCTGTCAGC 232
orena           CATATCTGACATAATG-CACTTTCTGCC-----ACAGGCTCACCAGACCATCCTGTCAGC 228
                * ** * ***** **  *   ***   *      **************************

yakuba          CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 285
santomea        CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 285
teissieri       CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 281
simulans        CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 291
sechellia       CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 291
mauritiana      CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 291
melanogaster    CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 290
erecta          CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 292
orena           CTGCAGTCCGTACTTCGAGACGATTTTCCTACAGAACCAGCATCCACATCCCATCATCTA 288
                ************************************************************

yakuba          CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 345
santomea        CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 345
teissieri       CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 341
simulans        CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 351
sechellia       CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 351
mauritiana      CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 351
melanogaster    CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 350
erecta          CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 352
orena           CTTGAAAGATGTCAGATACTCAGAGATGCGATCTCTGCTCGACTTCATGTACAAGGGCGA 348
                ************************************************************
```

```
yakuba          GGTCAACGTGGGTCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 405
santomea        GGTCAACGTGGGTCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 405
teissieri       GGTCAACGTGGGTCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 401
simulans        GGTCAACGTGGGTCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 411
sechellia       AGTCAACGTGGGTCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 411
mauritiana      GGTCAACGTGGGTCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 411
melanogaster    GGTCAACGTGGGCCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 410
erecta          GGTCAACGTGGGTCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 412
orena           GGTCAACGTGGGTCAGAGTTCGCTGCCCATGTTTCTCAAGACGGCCGAGAGCCTGCAGGT 408
                 *********** ************************************************

yakuba          ATGTGTAAATT-GGGGTTTAC--TTTAAGTAATAAGGGATATTCGGAGGGCATTCGGTAA 462
santomea        ATGTGTAAATT-TGGGTTTAC--TCTAAGTTA---------------------------- 434
teissieri       GTGTGTAAATT-GGGATTTAC--TCTTAATAA---GGGATATCCGGAGTTCATTCGGTGA 455
simulans        ATGTGCAAATCAGGGATACAT--TTTAAGGGAC-----ACAACTTAAGGACATCCTTTTT 464
sechellia       GTGTGCAAATCAGGGATACAT--TTTAAGGGAC-----ACAAATTTAGGATATCCTTTTT 464
mauritiana      GTGT--AAATCAGGGATATATATTTTAAGGGAC-----ACAACTTTAGGACATCCTTTTT 464
melanogaster    GTGTGCAAATCAGGGATCTAT--TTTAAGGGAC-----ACGACTTTAGGACATCCTTTTT 463
erecta          GCGT--AGATC-AGATTTTGC--TCTGAGTAATAGGGGATATCCGGCGGACATTCGGTTG 467
orena           GTGT--AGATC-AGATTTTGC--TCGAAGTAATAAGGGTTATCCGGAGGACATTCGTTTA 463
                  **  * **   *  *       *   *   *

yakuba          AGCGATCGTCTACAG-------------------AACCAGAAACTTGCGATTTATTAGTT 503
santomea        AGCGATCGTCTACAG-------------------AACCAGAAACTTGCGATTTATTAGTT 475
teissieri       ACCAATCGTTTACAG-------------------AACCAGAAACTTGCGATTTATTTGTT 496
simulans        -GATTATACTAAAG-------------------TCATTCAGAGGCTTTAAAGTATGGCGAA 505
sechellia       -GATTGCATTTAAG-------------------T---TTAGAAGCTTTAAAGTATGGCGAA 502
mauritiana      -GATTGTATTAAAG-------------------TCATTCAGAAGCTTTAAAGTATGGCGAA 505
melanogaster    TGCTTGTATTTAAGGCATAAAGAATATCAGAGTCATTCACAAGCTTTAAAGTATGGCGAA 523
erecta          CACAATATTATATATTG---------------TAGAATATTATATTTTTAATTCAAACACT 513
orena           --------TATATACT----------------CGTAATATTATAATTTTAATTGAAACACT 500
                        *                                *   **   * *

yakuba          AAATTT---AATAGAAGAAT---TGTTGTTTTGTCATAGAC------------------- 538
santomea        AAATTT---AATAGAAGAAT---TGTTGTTTTGTCATAGAC------------------- 510
teissieri       AAATTT---AATAGCAGAGTAAATATTTTATTGTTCTTTAT-------------------- 534
simulans        GAAATA---AGTTGTCGGCTTACTTTAAATTTAAA------------------------- 537
sechellia       GAAATA---AGTTTTCGGCTTACTTTAAATTTAAAAAAAACAATTGTATAAACAGAAAAT 559
mauritiana      GAAATA---AGTTGTCGGCTTACTTTAAATTTATAA------------------------ 538
melanogaster    GAAATA---TGTTGTCGGCTTACTTTAAATTAGTTAT----------------------- 557
erecta          CGTTTATAGAACCAGAAACTTGCGGTTTATTTGTTAAAAGTAATAAAAAAG--------- 564
orena           CGCTTATAGCACCAGAAACTTGCAATTGATTTGTTATCTGCAATGAAAAAG--------- 551
                    *             *     *    *

yakuba          ------------GATCTT-AGTATTACTTTCAAGAAA-------GTATGTCAGTTAACAT 578
santomea        ------------GATCTTTAGTATTAATTTCAAGAAA-------GTATGTCAGTTTACAT 551
teissieri       ------------AATTGTAGTTCTTGTATTTGTCATT-------CAGTTTCAAC-AAAGT 574
simulans        ------------TATAATGCGAATCTTACGAAACAAA-------GTATTCCAGAGCACAG 578
sechellia       TATTTTATTATATATTATGCGAATCTTACGAAACAAA-------CTATGCCAGAGCACAG 612
mauritiana      -------TTATGTATAACCAGAAATTTACTTTATTAATTATTATGTATGCCAGAGCACAT 591
melanogaster    ------------TATTAAGCGAGTCTTGCGAAACAAA-------GTGTGCCAGACCACAT 598
erecta          --------TAAATAGTTC-ATTATTTTCTGAATTGAATTGAACTTTTGAACATTGCACGT 615
orena           --------TTAATATTTT-ACTATTTTTTGAATTGAACTGATCATTTGAACATTGAACAT 602
                        *                                          **    *

yakuba          TTAATA----GTCTTGTTTTT-----AAATGGGAGAATGCCCGAATTTTTGTCCGCTTTC 629
santomea        TTAATA----GTCCAGGGGGT-----AAATGGGGGAATGCCCGAATTTGTGTCCCCTCTT 602
teissieri       ATGTCA----GTTCA-CTTTT-----AATAGGGGGAATGCCCGAATTTGCGTCCGCT-TC 623
simulans        TTAATA----GTCTTGCTTAT----CAAACTGGGGAACTCCCGAATTTGTGTCCGCT--A 628
sechellia       TTAATA----GTCTTGCTTATTTATCAAACTGGGGAACTCCCGAATTTGTGTCCGCT--A 666
mauritiana      TTAATA----ATCTTGCTTAT----CAAACTGGGGAACTCCCGAATTTGTGTCCGCT--A 641
melanogaster    GTAATAAATAGTCTTGCTTAT----CAAACTGGGGAATTCCCGAATTTGTGTCCGCT--A 652
erecta          -TAATA----GTCTTGTTTTT----TAAATCGGGGAAAGCCCGAACTTGCGTCTGCT--G 664
orena           -TAAGA----GTCTTGTTATT----TAAATCGGGGAAAGCCCGAATTTGTGCCTGCT--T 651
                 *   *     *       *     **   ** ***  ****** **  * *  **

yakuba          CCCCTCTATCAACAATTCAACATGAACTTATAATAACATGTATATTACTTGGGACTCTTG 689
santomea        AACCGCCATCAACAAT-CAACATGAACTTATAATAACGTGTATATTTAATGGGGA-CTTG 660
teissieri       CCCCTCTATCAACAATTCAACATGAACTTCTAATAACATGTATATTATTTGCGACTCTTG 683
simulans        ACCCACCACCAACAATTCAATATGAACTTATAATAATATGTACGTTATTTGGGATTTTTG 688
sechellia       ACCCACCACCAACAATTCAATATGAACTTATAATAATATGTACGTTATTTGGGATTTTTG 726
mauritiana      ACCCACCACCAACAATTCAATATGAACTTATAATAACATGTACATTTTTTGGGATTTTTG 701
melanogaster    CCCCAACAACAACAATTTAATATGAACTTATAATAATATGTACGTTATTTGGGACTTTTC 712
erecta          CCCCTCCAGCAACAATTTAATATAAAGCTCTAATAAAATGTATATTATTCC--AC-TTTG 721
orena           CCCCTCCACCAACAATTCAATATAGAGCTCTAATAAAATGTATATTATTCCGCAC-TTTG 710
                  **   * *******  ** **  *  * ******  ****  **           **
```

```
yakuba          CAGGTGCGTGGTCTCACAGATAACAACAATCTGAACTACCGCTCCGATTGCGACAAGCTG 749
santomea        ------------------------------------------------------------
teissieri       CAGGTGCGTGGTCTCACAGATAACAACAATCTGAACTACCGCTCCGATTGCGACAAGCTG 743
simulans        CAGGTGCGTGGTCTCACAGATAACAACAATCTGAACTACCGCTCCGACTGCGACAAGCTG 748
sechellia       CAGGTGCGTGGTCTCACAGATAACAACAATCTGAACTACCGCTCCGACTGCGACAAGCTG 786
mauritiana      CAGGTGCGTGGTCTCACAGATAACAACAATCTGAACTACCGCTCCGACTGCGACAAGCTG 761
melanogaster    CAGGTGCGTGGTCTCACAGATAACAACAATCTGAACTACCGCTCCGACTGCGACAAGCTG 772
erecta          CAGGTGCGTGGTCTCACAGATAACAACAATCTGAACTACCGCTCCGATTGCGACAAGCTG 781
orena           CAGGTGCGTGGTCTCACAGATAACAACAATCTGAACTACCGCTCCGATTGCGACAAGCTG 770

yakuba          CGCGACTCGGCGGC 763
santomea        --------------
teissieri       CGCGACTCGGCGGC 757
simulans        CGCGATTCGGCGGC 762
sechellia       CGCGATTCGGCGGC 800
mauritiana      CGCGATTCGGCGGC 775
melanogaster    CGCGATTCGGCGGC 786
erecta          CGCGATTCGGCGGC 795
orena           CGCGATTCGGCGGC 784
```

Figure A4.2 GeneWise Reports for D.Melanogaster Splicing

D. melanogaster

```
protein             1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                      MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                      MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
melanogaster       10 agcctttctaaccaataggcatccccggctggacgtgggaga
                      taaatgtggaaaccatcgttccttagactgatctcgagacta
                      gcgaccgcgcttcatgccgacaggggggacccgcccgcaacg

protein             43                          AHQTILSACSPYFETIFLQNQHPHPI
                                                AHQTILSACSPYFETIFLQNQHPHPI
                                                AHQTILSACSPYFETIFLQNQHPHPI
melanogaster       136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccacccccca
                      <0-----[136  :  207]-0>caacttccggcatactttaaaacact
                                          tcgccgacctgccggtcagcgtatcc

protein             69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                      IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                      IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
melanogaster      286 attaggattgactccgtatagggaggcatccatcaaggacc
                      tataatgacatgcttattaagatatgagctctttaccagta
                      ccgatcacaggatgcccgcgcgccgcgtggcgtcggcgcgg

protein            110                          VRGLTDNNNLNYRSDCDKLRDSA
                                                VRGLTDNNNLNYRSDCDKLRDSA
                                                VRGLTDNNNLNYRSDCDKLRDSA
melanogaster       409 GTGTGTG  Intron 2   CAGgcgcagaaacatctgtgaccgtg
                      <0-----[409  :  715]-0>tggtcaaaataagcagaatgacc
                                          gttcatcctgccccccggctgg
```

D. simulans

```
protein
                1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                  MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                  MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
simulans         10 agcctttctaaccaataggcatccccggctggacgtgggaga
                    taaatgtggaaaccatcgttccttagactgatctcgagacta
                    gcgaccgcgcttcatgccgccaggggggacccgcccgcaacg

protein          43                          AHQTILSACSPYFETIFLQNQHPHPI
                                             AHQTILSACSPYFETIFLQNQHPHPI
                                             AHQTILSACSPYFETIFLQNQHPHPI
simulans        136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccacccccca
                    <0-----[136  :  208]-0>caacttccggcatactttaaaacact
                                             tcgccgacctgccggtcagcgtatcc


protein          69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                    IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                    IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
simulans        287 attaggattgactccgtatagggaggcatccatcaaggacc
                    tataatgacatgcttattaagatatgagctctttaccagta
                    ccgatcacaggatgcccgcgcgccgtgtggcgtcggcgcgg


protein         110                          VRGLTDNNNLNYRSDCDKLRDSA
                                             VRGLTDNNNLNYRSDCDKLRDSA
                                             VRGLTDNNNLNYRSDCDKLRDSA
simulans        410 GTATGTG  Intron 2   CAGgcgcagaaacatctgtgaccgtg
                    <0-----[410  :  691]-0>tggtcaaaataagcagaatgacc
                                             gttcatcctgccccccggctgg
```

D. sechellia

```
protein           1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                    MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                    MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
Drosophila       10 agcctttctaaccaataggcatccccggctggacgtgggaga
                    taaatgtggaaaccatcgttccttagactgatctcgagacta
                    gcgaccgcgcttcatgccgccaggggggacccgcccgcaacg


protein          43                          AHQTILSACSPYFETIFLQNQHPHPI
                                             AHQTILSACSPYFETIFLQNQHPHPI
                                             AHQTILSACSPYFETIFLQNQHPHPI
Drosophila      136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccacccccca
                    <0-----[136  :  208]-0>caacttccggcatactttaaaacact
                                             tcgccgacctgccggtcagcgtatcc


protein          69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                    IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                    IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
Drosophila      287 attaggattgactccgtatagggaggcatccatcaaggacc
                    tataatgacatgcttattaagatatgagctctttaccagta
                    ccgatcacaggatgcccgcgcaccgtgtggcgtcggcgcgg


protein         110                          VRGLTDNNNLNYRSDCDKLRDSA
                                             VRGLTDNNNLNYRSDCDKLRDSA
                                             VRGLTDNNNLNYRSDCDKLRDSA
Drosophila      410 GTGTGTG  Intron 2   CAGgcgcagaaacatctgtgaccgtg
                    <0-----[410  :  729]-0>tggtcaaaataagcagaatgacc
                                             gttcatcctgccccccggctgg
```

D. mauritiana

```
protein            1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
Mauritiana        10 agcctttctaaccaataggcatccccggctggacgtgggaga
                     taaatgtggaaaccatcgttccttagactgatctcgagacta
                     gcgaccgcgcttcatgccgccaggggggacccgcccgcaacg


protein           43                          AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
Mauritiana       136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccacccccca
                     <0-----[136  :  208]-0>caacttccggcatactttaaaacact
                                              tcgccgacctgccggtcagcgtatcc


protein           69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
Mauritiana       287 attaggattgactccgtatagggaggcatccatcaaggacc
                     tataatgacatgcttattaagatatgagctctttaccagta
                     ccgatcacaggatgcccgcgcgccgtgtggcgtcggcgcgg


protein          110                          VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
Mauritiana       410 GTGTGTA  Intron 2   CAGgcgcagaaacatctgtgaccgtg
                     <0-----[410  :  704]-0>tggtcaaaataagcagaatgacc
                                              gttcatcctgccccccggctgg
```

D. teisserie

```
protein            1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
teissieri         10 agcctttctaaccaacaggcatccccggctggacgtgggaga
                     taaatgtggaaaccatcgttccttagactgatctcgagacta
                     gcgaccgcgcttcatgccgccaggggggacccgcccgcaacg


protein           43                          AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
teissieri        136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccacccccca
                     <0-----[136  :  198]-0>caacttccggcatactttaaaacact
                                              tcgccgacctgccggtcagcgtatcc


protein           69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
teissieri        277 attaggattgactccgtatagggaggcatccatcaaggacc
                     tataatgacatgcttattaagatatgagctctttaccagta
                     ccgatcacaggatgcccgcgcgccgtgtggcgtcggcgcgg


protein          110                          VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
teissieri        400 GTGTGTG  Intron 2   CAGgcgcagaaacatctgtgaccgtg
                     <0-----[400  :  686]-0>tggtcaaaataagcagaatgacc
                                              gttcatcctgcccctccggccgg
```

D. yakuba

```
protein            1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
yakuba            10 agcctttctaaccaacaggcatccccggctggacgtgggaga
                     taaatgtggaaaccatcgttccttagactgatctcgagacta
                     gcgaccgcgcttcatgccgccaggggggacccgcccgcaacg


protein           43                          AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
yakuba           136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccacccccca
                     <0-----[136  :  202]-0>caacttccggcatactttaaaacact
                                              tcgccgacctgccggtcagcgtatcc


protein           69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
yakuba           281 attaggattgactccgtatagggaggcatccatcaaggacc
                     tataatgacatgcttattaagatatgagctctttaccagta
                     ccgatcacaggatgcccgcgcgccgtgtggcgtcggcgcgg


protein          110                          VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
yakuba           404 GTATGTG  Intron 2   CAGgcgcagaaacatctgtgaccgtg
                     <0-----[404  :  692]-0>tggtcaaaataagcagaatgacc
                                              gttcatcctgcccctccggccgg
```

D. erecta

```
protein            1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
erecta            10 agcctttctaaccaacaggcatccccggctggacgtgggaga
                     taaatgtggaaaccatcgttccttagactgatctcgagacta
                     gcgaccgcgcttcatgccgccaggggggacccgcccgcaacg


protein           43                          AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
erecta           136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccacccccca
                     <0-----[136  :  209]-0>caacttccggcatactttaaaacact
                                              tcgccgacctgccggtcagcgtatcc


protein           69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
erecta           288 attaggattgactccgtatagggaggcatccatcaaggacc
                     tataatgacatgcttattaagatatgagctctttaccagta
                     ccgatcacaggatgcccgcgcgccgtgtggcgtcggcgcgg


protein          110                          VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
erecta           411 GTGCGTA  Intron 2   CAGgcgcagaaacatctgtgaccgtg
                     <0-----[411  :  724]-0>tggtcaaaataagcagaatgacc
                                              gttcatcctgcccctccggctgg
```

D. orena

```
protein            1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
orena             10 agcctttctaaccaacaggcatccccggctggacgtgggaga
                     taaatgtggaaaccatcgttccttagactgatctcgagacta
                     gcgaccgcgcttcatgccgccaggggggacccgcccgcaacg


protein           43                          AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
orena            136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccaccccca
                     <0-----[136  :  205]-0>caacttccggcatactttaaaacact
                                              tcgccgacctgccggtcagcgtatcc


protein           69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQ
orena            284 attaggattgactccgtatagggaggcatccatcaaggacc
                     tataatgacatgcttattaagatatgagctctttaccagta
                     ccgatcacaggatgcccgcgcgccgtgtggcgtcggcgcgg


protein          110                          VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
                                              VRGLTDNNNLNYRSDCDKLRDSA
orena            407 GTGTGTA  Intron 2   CAGgcgcagaaacatctgtgaccgtg
                     <0-----[407  :  713]-0>tggtcaaaataagcagaatgacc
                                              gttcatcctgcccctccggctgg
```

D. santomea

```
protein            1 MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
                     MDQQFCLRWNNHPTNLTGVLTSLLQREALCDVTLACEGETVK
santomea          10 agcctttctaaccaacaggcatccccggctggacgtgggaga
                     taaatgtggaaaccatcgttccttagactgatctcgagacta
                     gcgaccgcgcttcatgccgccaggggggacccgcccgaaacg


protein           43                          AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
                                              AHQTILSACSPYFETIFLQNQHPHPI
santomea         136 GTGCGTC  Intron 1   CAGgccaactgtacttgaatccaccccca
                     <0-----[136  :  202]-0>caacttccggcatactttaaaacact
                                              tcgccgacctgccggtcagcgtatcc


protein           69 IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQV
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQV
                     IYLKDVRYSEMRSLLDFMYKGEVNVGQSSLPMFLKTAESLQV
santomea         281 attaggattgactccgtatagggaggcatccatcaaggaccg
                     tataatgacatgcttattaagatatgagctctttaccagtat
                     ccgatcacaggatgcccgcgcgccgtgtggcgtcggcgcgga
```

Figure A4.3 Bayes Empirical Bayes (BEB) analysis (YAND et al. 2005) Positively selected sites (*: P>95%; **: P>99%).

Fru B Isoform

Position	Pr(w>1)	post mean +- SE for w
41 L	0.868	1.383 +- 0.316
91 A	0.816	1.334 +- 0.366
246 T	0.651	1.151 +- 0.497
249 L	0.521	0.956 +- 0.596
252 A	0.811	1.328 +- 0.371
661 G	0.717	1.242 +- 0.427
672 S	0.938	1.451 +- 0.207
673 G	0.567	1.073 +- 0.509
678 I	0.954*	1.465 +- 0.178
680 *	0.787	1.304 +- 0.394
683 *	0.880	1.394 +- 0.302
684 *	0.624	1.123 +- 0.508
685 *	0.857	1.372 +- 0.329
686 *	0.631	1.145 +- 0.484

Fru F Isoform

Position	Pr(w>1)	post mean +- SE for w
319 R	0.611	1.114 +- 0.509
547 G	0.708	1.235 +- 0.435
558 S	0.770	1.294 +- 0.400
561 A	0.902	1.420 +- 0.274
562 P	0.966*	1.478 +- 0.165
572 G	0.908	1.426 +- 0.265
573 G	0.516	0.961 +- 0.589

Fru D Isoform

Position	Pr(w>1)	post mean +- SE for w
145 S	0.548	1.054 +- 0.543
151 L	0.739	1.277 +- 0.445
474 Y	0.784	1.324 +- 0.417
475 L	0.833	1.369 +- 0.400
478 A	0.975*	1.509 +- 0.208
479 G	0.833	1.379 +- 0.372
480 N	0.591	1.113 +- 0.523
532 S	0.851	1.392 +- 0.373
534 P	0.716	1.259 +- 0.448
535 A	0.954*	1.491 +- 0.239
536 T	0.907	1.449 +- 0.299
537 G	0.714	1.251 +- 0.463
538 G	0.825	1.368 +- 0.386
539 S	0.935	1.475 +- 0.266
540 S	0.928	1.468 +- 0.277
541 A	0.781	1.324 +- 0.421
542 G	0.781	1.321 +- 0.423
543 A	0.701	1.246 +- 0.456
544 A	0.946	1.485 +- 0.251
545 L	0.746	1.285 +- 0.453
546 G	0.918	1.458 +- 0.293
548 S	0.820	1.355 +- 0.414
549 S	0.585	1.093 +- 0.544
632 Q	0.630	1.171 +- 0.480
636 P	0.533	1.056 +- 0.520
641 N	0.511	0.958 +- 0.622
644 A	0.608	1.140 +- 0.504
646 R	0.905	1.445 +- 0.313
647 P	0.762	1.301 +- 0.435