The Genetics of Interspecific Variation in Pheromone Profile in *Drosophila simulans* and *D. sechellia*

Elliott Cantor

Supervisor: Dr. A Civetta

A thesis submitted in partial fulfilment of the Honours Thesis Course 05.4111/6
Department of Biology
The University of Winnipeg
2002

Abstract

The cuticle of *Drosophila* consists of long chain, unsaturated hydrocarbons that have pheromonal properties. These pheromones are used for recognition of the opposite sex and are responsible for mating recognition between closely related species. The chromosomal basis for the difference in pheromone profile between two closely related species *D. sechellia* and *D. simulans* has been previously examined, yet the specific genes involved in their production and level of expression remain unknown. Using gas chromatography and mass spectrometry, the pheromone profiles of over 90 lines of *D. simulans* with randomly introgressed *D. sechellia* chromosomal fragments (introgressed lines) were determined for both males and females. Variations in pheromone profile among introgressed lines were quantified with respect to a known quantity of internal standard. The introgressed lines were previously scored for twenty-three molecular markers along the three major chromosomes of *Drosophila*. Variation in the genetic nature of each marker (*D. simulans* vs. *D. sechellia*) and the phenotypic variation in pheromone profile were associated using linear regression analysis and interval mapping. Significant quantitative trait loci (QTLs) that explain variation in male and female pheromone production were located within the X chromosome (chromosome 1). *Pbrb2* and *Or19a* were identified as potential candidate genes for 7-tricosene production.

Acknowledgements

I would like to thank Dr. Civetta for taking me on as his project student this year and being the best supervisor I could ask for. You were always available to answer any questions I had regarding my research. I appreciate all the valuable time you were able to spend with me in order to further my learning in the field of genetics. Thank you to my committee member Dr. Wiegand, and a special thank you to my committee member Dr. Vanderwel who trained me on the GC-MSD and familiarized me with the extraction protocol. Without your guidance, this project would not have been possible. I would also like to thank Dr. Moodie for doing a wonderful job organizing the Honours Thesis Course.

Table of Contents

Abstract **ii**

Acknowledgements **iii**

Table of Contents **iv**

List of Figures **v**

List of Tables **vi**

Introduction **1-8**

Methods **9-15**
- Making the *D. simulans /D. sechellia* IG lines 9
- Rearing conditions 9
- Collection of virgin flies 10
- Hydrocarbon extraction and gas chromatography-mass spectrometry (GC-MS) protocol 10
- Hydrocarbon identification and analysis 12
- QTL analysis 13
- Candidate gene loci 15

Results **16-29**
- Hydrocarbon identification 16
- Phenotypic variation 19
- Statistical tests 21
- Marker-trait associations 21
- Interval mapping 25
- Associations made with chromosome 1 25
- Selected candidate genes 26

Discussion **30-36**
- Chromosome 1: Genetic associations with tricosene levels and potential candidate genes 30
- Chromosome 2 32
- Chromosome 3 33
- Resolving the molecular marker map 33
- Other pheromonal hydrocarbons 34
- Future Directions 35

Conclusions **37**

References **38-39**

List of Figures

Figure 1: Gas chromatograms of the cuticular hydrocarbon profiles of 5-7 day old virgin males (A) and females (B) of IG line Q15. Peaks are numbered according to increasing retention times. See Table 3 for corresponding peak identification. **17**

Figure 2: Number of IG lines showing different relative concentrations of tricosene to hexacosane standard in males (A) and females (B). **20**

Figure 3: Interval map of male chromosome 1 showing an association between tricosene levels and a genetic region between markers 6 and 7 (Approximate genetic position 1-57). **27**

Figure 4: Interval map of female chromosome 1 showing an association between tricosene levels and a genetic region close to marker 5 (Approximate genetic position 1-42.7). **28**

Figure 5: Interval map of female chromosome 2 showing an association between tricosene production and the approximate genetic position 21, which lies between markers 8 and 9. **29**

List of Tables

Table 1: Major pheromonal hydrocarbons produced by four closely related species of *Drosophila* (++ = predominant, + = present, 0 = trace/none) (from Coyne *et al.*, 1994). **3**

Table 2: Recipe for preparation of *Drosophila* media (Prepares 20 bottles or 120 vials). Add cornmeal, yeast and agar to cold water and mix well. Add to boiling water and mix well. Add molasses and mix well. Let stand and cool to 60ºC. Add tegosept and propionic acid. **9**

Table 3: Identification of the cuticular hydrocarbons of 5-7 day old virgin flies seen in Figure 1. **18**

Table 4: Summary of statistical analyses performed on tricosene data in both male and female IG lines. Data demonstrates normal distribution as denoted by the S-value. **22**

Table 5: Linear regression analysis of male tricosene levels with respect to the 23 scored molecular markers. **23**

Table 6: Linear regression analysis of female tricosene levels with respect to the 23 scored molecular markers. **24**

Introduction

In *Drosophila*, as well as all other terrestrial insects, the cuticle is present to secrete substances with a dual purpose: to protect from environmental elements such as wind and water, and to be used for chemical recognition. The insect cuticle is secreted from the epidermis and is composed of a procuticle and epicuticle. Pore canals in the procuticle branch off to form a network of canals in the epicuticle, supplying the surface with its waxy coating (Chapman, 1982).

In the cases of *Drosophila* and other insects, the compounds secreted to the surface include hydrocarbons (Chapman, 1982). The chemical structure of these compounds can vary significantly with respect to carbon chain-length and branching between different groups of insects, often dependent on the climate in which these insects live. Insects such as *Drosophila* that live in hot regions tend to have longer hydrocarbon chains that are more resistant to melting at higher temperatures (Chapman, 1982). This means they are less volatile, giving them the ability to maintain their coating under warmer conditions. Hydrocarbons, which are hydrophobic in nature, are also known for their desiccation-preventing properties (Jallon and David, 1987).

The cuticular hydrocarbons of *Drosophila* species include odd-numbered long chain hydrocarbons that range from 23 to 29 carbons in length. In addition to serving as a protective coating, certain unsaturated hydrocarbons also serve as pheromones. These pheromones are implicated in the mating process as a method of recognizing conspecifics and discriminating against heterospecifics. During the mating ritual, males track females, often touching them with their forelegs. The males vibrate their wings, lick the abdomen of the female and attempt copulation (Savarit *et al.*, 1999). The forelegs pick up chemical

signals by touch, while volatile pheromones released in the process are received by olfactory receptor neurons in the antennae (de Bruyne *et al.*, 1999).

Many experiments have been conducted on the effects of hydrocarbons on mating behaviour in *Drosophila*. Antony and Jallon (1982) compared the effects of alkanes, monoenes and dienes on mating behaviour in *Drosophila melanogaster* in an attempt to determine the chemical basis for sex recognition. They found both the monoenes and the dienes to be stimulatory, while the alkanes were non-stimulatory. The stimulatory compounds are successful in inducing male courtship, with certain unsaturated compounds causing a longer wing vibration response than others.

Little is known regarding the behavioural effects of male-produced pheromones on females. In response to female-produced pheromones, males court females by singing, which involves rapid wing vibration. During this process, volatile hydrocarbons are released from males and received by females; however, their effects on female mating behaviour remain unknown (Antony and Jallon, 1982).

Closely related species of *Drosophila* vary significantly in their hydrocarbon profile, which allows for conspecific mating recognition (Coyne *et al.*, 1994). Table 1 shows the hydrocarbons produced by four species of the *D. melanogaster* complex.

Table 1: Major pheromonal hydrocarbons produced by four closely related species of *Drosophila* (++ = predominant, + = present, 0 = trace/none) (from Coyne *et al.*, 1994).

Species	Sex	7-tricosene	7-pentacosene	7,11-heptacosadiene
D. melanogaster	Male	++	+	0
	Female	0	0	++
D. sechellia	Male	++	+	0
	Female	0	0	++
D. simulans	Male	++	+	0
	Female	++	+	0
D. mauritiana	Male	++	+	0
	Female	++	+	0

Coyne *et al.* (1994) conducted tests on *Drosophila simulans* and *D. sechellia* in order to determine whether pheromones play a role in the mating behaviours that exist within and between species. These two species were selected in the study because of a sexual polymorphism that exists between them. *D. simulans* males and females both produce 7-tricosene (7-T) as the predominant hydrocarbon. For this reason, they are said to be monomorphic. *D. sechellia*, conversely, are dimorphic with males producing mainly 7-T and females producing predominantly 7,11-heptacosadiene (7,11-HD)

(Antony *et al.*, 1985; Ferveur and Sureau, 1996; Coyne and Charlesworth, 1997; Ferveur *et al.*, 1997). It has been shown that males from the sexually monomorphic species (*D. simulans*) do not court females from the sexually dimorphic species (*D. sechellia*); however, *D. sechellia* males may court *D. simulans* females due to the familiarity of the compounds that they themselves produce (Coyne *et al.*, 1994). They also report that *D. simulans* males, despite not courting *D. sechellia* females, do tend to court F1 female hybrids to some extent. *D. simulans* males tend to ignore their female counterparts when these females are coated with the *D. sechellia* pheromone 7,11-HD. They also tend to court *D. sechellia* females that have picked up high levels of 7-T due to crowding. This is strong evidence implicating hydrocarbons as a major factor in determining mating preference.

The genetic basis behind differences in pheromone profiles between closely related species of *Drosophila* has been studied. Coyne *et al.* (1994) investigated differences in major hydrocarbons between *D. simulans* and *D. sechellia* using multiple markers. *D. sechellia* males with markers for homozygous recessive eye colour on the X chromosome (*zinfandel*) and the 2nd chromosome (*cinnabar*) were crossed with *D. simulans* females heterozygous for a *Hairless* dominant allele on the 3rd chromosome. Backcrossing *Hairless* F1 female progeny with *D. sechellia* males yielded eight genotypically variable progeny with all the possible phenotypic marker combinations. The 7-T to 7,11-HD ratios were measured, and using analysis of variance (ANOVA) an association was made between the relative quantity of these hydrocarbons and the *Hairless* marker on the 3rd chromosome. Also, high levels of 7-pentacosene (7-P) were

associated with the X chromosome. No significant interactions in analysis of variance suggest the lack of epistatic interactions between markers.

Despite finding only two chromosomes significantly associated with variation in the pheromone profile between *D. simulans* and *D. sechellia*, Coyne (1996) later determined that the X chromosome, 2nd chromosome and the 3rd chromosome were significant in affecting the ratio of 6-tricosene (6-T) to 7-T in *D. simulans*. Genetically marked *D. simulans* females were crossed with *D. sechellia* males and their F1 female progeny was backcrossed to the same marked *D. simulans* males. As a result, 32 genotypically different lines were created, 10 of which were scored. Two markers on chromosome 2 were associated with the ratio of 6-T to 7-T, having a distance between them greater than 100 map units. The same conclusion was reached for the third chromosome, which also associated two distant markers. In all, at least five gene loci were postulated to be responsible for tricosene ratios, including possible epistatic gene interactions among genes within chromosome 3.

Finding gene loci responsible for the pheromone production is a cumbersome task. However, employing mapping techniques that involve a large number of molecular markers randomly spread throughout chromosomes increases the opportunity of narrowing down associations between phenotypes and markers. Ferveur (1991) mapped variations in the ratio of 7-T to 7-P in *D. simulans* using strains of flies with multiple molecular markers. An association was made with 7-T to 7-P ratios and the 2nd chromosome at genetic position 65.3 ±1.0 known as *Ngbo*. It was found that variation between the two known alleles of *Ngbo* caused marked variation in 7-P levels. In the same way, Ferveur and Jallon (1992) identified *kété*, a locus on the X chromosome at the

genetic position 18.5 responsible for 7-T levels in flies of both sexes. Although these gene regions have been identified, it is still unknown as to how much epistasis, pleiotropy and other factors play in the expression of these genes.

Through the use of statistical methods, it is possible to associate markers with phenotypes. The genetic factors responsible for these phenotypes are referred to as quantitative trait loci (QTL). With this tool, a higher-density molecular map is beneficial because it results in better localization of QTLs. MacDonald and Goldstein (1999) carried out experiments on backcrossed lines of *D. simulans* and *D. sechellia* to test for QTLs for multiple phenotypes. Testis length, cyst length, tibia length, sex comb tooth number and posterior lobe of the genital arch area were all measured in 20 lines of flies. Thirty-eight markers spread across the genome at an average distance of 8.4cM were used to make associations with the phenotypes collected. The most notable of their findings was a significant association between a marker and the shape and size of the posterior lobe of the genital arch. This is due to the fact that the size of the genital arch is noticeably variable between these two species. With a QTL detected, they were able to reveal a candidate gene for posterior lobe shape and size using FlyBase.

FlyBase (http://flybase.bio.indiana.edu) is a search database found on the Internet that has readily available information about the molecular and biological role of genes based on their cytological position in *D. melanogaster*. It can be used to search for potential candidate genes at or near the position of experimentally detected QTLs that may be responsible for the quantitative trait. Genes whose identities are described on the database and have related functions to that of the phenotype are potentially good candidate genes.

MacDonald and Goldstein (1999) and Civetta *et al.* (2002) use interval mapping to associate gene regions with phenotypic variations. Using the QTL data, a continuous probability of association, a likelihood ratio (LR), is scored across the entire chromosome using chi-square. This statistical method fills in the gaps between molecular markers where significant regions of the genome would have otherwise gone undetected using only a marker-based analysis. With this information, candidate genes can be drawn from FlyBase and investigated further.

Most of what we know today about the *Drosophila* genome comes mainly from studies of *D. melanogaster*. In this study, however, I examined phenotypic variation in two very closely related species, *D. simulans* and *D. sechellia*. These species separated only about 0.4-0.9 million years ago, having branched off from their common *D. melanogaster* ancestor about 2.5-3.4 million years ago (Lemeunier *et al.*, 1986). Although the genomes of *D. simulans* and *D. sechellia* have not been fully sequenced like their *D. melanogaster* relative, their genetic compositions are homologous as demonstrated by the similar banding patterns of their polytene chromosomes, which are used to obtain their cytological map position (Griffiths *et al.*, 1996). Since *D. simulans* and *D. sechellia* can interbreed to produce viable female offspring, they are effective tools for marker-based genetic analyses that study variations between species.

One contributor to species isolation is changes in mating preference that evolve over time. As demonstrated by the behavioural experiments (Antony and Jallon, 1982; Coyne *et al.,* 1994), a major component of mating preference and recognition involves the pheromone profile. Therefore, targeting the genes responsible for variations in pheromone profiles between these two sister species may give us some insight into the

genes involved in the isolation these species. Potentially, these genes may possess homologues in other related species, and may help uncover the current genetic mystery of speciation.

How do different species of *Drosophila* produce different hydrocarbon profiles? What is the origin of such diversity? The answers to these questions potentially lie in the genes responsible for hydrocarbon production. If it is possible to find the genes that code for hydrocarbon production in *Drosophila*, then it may be possible to predict the origin of such phenotypic variation.

In this study, I used a set of introgressed (IG) lines consisting of mainly *D. simulans* genome with randomly introgressed *D. sechellia* chromosomal fragments. Along the three major chromosomes, twenty-three molecular markers were previously scored (*D. simulans* vs. *D. sechellia*). My goal was to collect the hydrocarbon profile from all the IG lines, male and female, and look at variation that may exists. With this data I identified QTLs that significantly associated with phenotypic variations in pheromone profile. Ultimately, candidate gene loci responsible for variation were identified using FlyBase.

Materials and Methods

Making the *D. simulans /D. sechellia* IG lines

The IG lines used for experimentation were previously constructed by crossing *D. simulans* females with *D. sechellia* males. The fertile F1 female progeny were doubly backcrossed with *D. simulans* males. The progeny were sib-mated in single pairs for 14 generations, losing heterozygosity in the process. As a result of the backcrosses, each IG line has a predicted average of 1/8 of their genome from *D. sechellia* and 7/8 from *D. simulans*. Due to random recombination and crossing-over events, each IG line likely has different chromosomal fragments of *D. sechellia* as part of its genome (Civetta *et al.*, 2002).

Rearing conditions

The flies were reared in plastic bottles containing a corn meal-based medium with the addition of tegosept and propionic acid to prevent bacterial growth (Table 2). Throughout experimentation, the bottles were stored at 24ºC and were subjected to 12-hour cycles of light and darkness.

Table 2: Recipe for preparation of *Drosophila* media (Prepares 20 bottles or 120 vials). Add cornmeal, yeast and agar to cold water and mix well. Add to boiling water and mix well. Add molasses and mix well. Let stand and cool to 60ºC. Add tegosept and propionic acid.

Ingredient	Amount
Cold water	170mL
Corn meal	65g
Yeast	13g
Agar	6.5g
Hot water	760mL
Molasses	45.5mL
99% Propionic acid	5mL
10% Tegosept (50 g methyl-hydroxybenzoate per 500 mL 95% ethanol	20mL

Collection of virgin flies

Flies were collected first by setting them up in bottles of fresh media. Parent flies were removed when the first signs of larvae were visible. Virgin flies eclosed from their pupae stage on or after day 12, and were collected as needed over a period of three days. Newly emerged flies remain virgins for 8 hours at 24ºC (Greenspan, 1997) therefore the bottles were emptied if left too long between collections. Virgin flies were dumped from the bottles into a vial filled with CO_2 gas. This temporarily knocked out the flies, making them immobile and easy to manipulate without flying away. Collected flies were separated according to sex using a light microscope and placed into plastic vials containing similar media in groups of five. The flies were then stored at 24ºC under the same conditions and allowed to age 5-7 days before the hydrocarbon extraction procedure.

Hydrocarbon extraction and gas chromatography-mass spectrometry (GC-MS) protocol

Before collecting the data on the IG lines, preliminary trial runs of GC-MS were performed on female *Drosophila melanogaster* from a well-established laboratory line (Canton-S) whose hydrocarbon profile had been well documented and identified by Antony and Jallon (1982). Because this data was easily replicable, I predicted that the hydrocarbon profile of the IG lines could be obtained using the same protocol.

In order to measure the concentration of each hydrocarbon extracted, a hexacosane internal standard was prepared at a concentration of 40µg/mL in hexane (Antony and Jallon, 1982).

Each sample was prepared by delivering 400µL of hexane via a glass syringe into a 7mL glass vial. Twenty-five microlitres of 40µg/mL (1µg) hexacosane standard was added to each vial. Five virgin flies aged 5-7 days of the same IG line and sex were anaesthetized with CO_2, placed in the 7mL vial, and vortexed at speed 7 for 1 minute. A small amount of anhydrous granular sodium sulphate drying agent was added to each vial to remove any water from solution. Using a Pasteur pipette, the organic solvent was transferred to a 1.5mL glass GC vial, leaving the drying crystals behind. The crystals were then washed with 50µL hexane and the remaining solvent was added to each GC vial. All GC vials were properly labelled and stored at -20ºC until ready for GC-MS. Before running GC-MS, the samples were evaporated under a slow stream of nitrogen gas and redissolved in 25µL hexane and pipetted into glass inserts that sat inside each GC vial.

GC-MS analyses were performed on a Hewlett Packard 5980 GC and a Hewlett Packard 5970 MSD. Helium gas flowed through the column with a head pressure of 10 psi and a flow rate near 41mL/min. The column was a 30 metre long SPB-5 non-polar phase useful for most types of analysis, including hydrocarbons. The injection port temperature was 250ºC and the oven program started at 120ºC and climbed at a rate of 2º/min until 250ºC at which point the temperature held steady for 15 minutes. The temperature was then programmed to rise sharply to 280ºC for 5 minutes to burn out the column. Each run took approximately 88 minutes to complete. The injection needle was washed once with acetone and twice with hexane before automatically injecting 1µl of sample. During a sequence of runs, the vials lined up for automatic injection were prevented from evaporating by running cold water through the hollow holding station.

The MS covered a mass range of 50-550 atomic mass units (AMU) taking 1.3 scans/sec. The data obtained was saved for further analysis.

Hydrocarbon identification and analysis

The program used for hydrocarbon data collection was MS ChemStation G1034C version C.03.00. The pheromone profile of each IG line was analyzed separately. First, one chromatogram was loaded and programmed to auto-integrate the area under each significant peak recorded. The auto-integrate function automatically calculated this area based on the position of the baseline before and after each peak. The resulting integrations were listed and saved to a notepad (.txt) file. Next, the mass spectrum of each peak was analyzed, noted and the identity of this peak was recorded in the notepad file.

There were two ways to identify each group of peaks and the subset of peaks within each grouping. The data for the IG lines were compared to the data collected on the Canton-S strain. With this method alone, the identity of each peak was evident due to homologies between species. Each peak also had characteristic MS fragmentation patterns, which could be compared with those of the Canton-S strain. To further verify the identity of each peak, chromatographs were subjected to a library search. The MS ChemStation software is equipped with a database library Wiley138 that can identify a GC peak with a certain probability based on the fingerprint of its corresponding mass spectrum.

Once each peak was identified, the file was imported to a Microsoft Excel worksheet. Each set of data for the subsequent IG lines was saved to a different

worksheet in the same file. The final result was two Microsoft Excel files, each containing the entire data set collected for each IG line; one for male profiles and one for female profiles.

QTL analysis

QTL Cartographer version 1.30 (Wang *et al.*, 2001) was used to calculate the mean values of tricosene levels in males and females, including standard deviation. Each data set was also statistically analyzed to see if it fit a pattern of normal distribution (S). S relies on the values of kurtosis and skewness, which are deviations from normality (Table 4). Sokal and Rohlf (1995) estimate skewness by:

$$skw(y) = \frac{n^2}{(n-1)(n-2)}\left(\overline{y^3} - 3\bar{y}^2\bar{y} + 2\bar{y}^3\right)$$

where $\bar{y}$ is the mean value of tricosene and n is sample size. The standard error of skewness is estimated by:

$$SE(skw) = \sqrt{\frac{6}{n}}$$

Skewness, or asymmetry, is the degree to which the values are skewed toward one tail of the normal curve, thus varying the median from the mean. Kurtosis is given by the following equation:

$$kur(y) = \frac{n^2(n+1)}{(n-1)(n-2)(n-3)}\left(\overline{y^4} - 4\overline{y^3}\bar{y} + 6\overline{y^2}\bar{y}^2 - 3\bar{y}^4\right)$$

Standard error estimated for kurtosis is:

$$SE(kur) = \sqrt{\frac{24}{n}}$$

Kurtosis describes a more complicated asymmetry of the normal curve. The value of kurtosis is given as the proportion of values found in the center and tails of the curve compared to the shoulders.

The coefficient of skewness is:

$$k_3 = \frac{skw(y)}{s^3}$$

The coefficient of kurtosis is:

$$k_4 = \frac{kur(y) - 3s^4}{s^4}$$

The test statistic for normality (S) is:

$$S = \left(\frac{nk_3^2}{6}\right) + \left(\frac{nk_4^2}{24}\right)$$

S follows a chi-square distribution. When $S < \chi_2^2;0.95 = 5.99$, the distribution is normal.

Linear regression analysis was performed on all the collected data with respect to each of the twenty-three scored molecular markers. The equation for linear regression is:

$$y = b_0 + b_1x + e$$

where y = phenotypic score, x = marker score, b_0 = intercept and b_1 = slope. The slope of each regression indicates the direction of the allele. A positive slope associates an increase in tricosene concentration with the *D. simulans* marker, while a negative slope associated higher levels of tricosene with the *D. sechellia* marker.

Interval mapping was used to make associations with gene regions between markers by giving a continuous range of likelihood ratios of marker-trait association. Likelihood ratio scores follow a chi-squared distribution and were plotted against the genetic position of the corresponding chromosome, yielding candidate gene loci at points where LR>3.84, which is a 95% confidence interval with 1 degree of freedom (d.f.).

Candidate gene loci

Candidate genes were selected using FlyBase based on their proximity to significant molecular markers and marker intervals associated with the scored phenotype. Likely candidates were selected from genes of known function that are potentially related to pheromone production or expression.

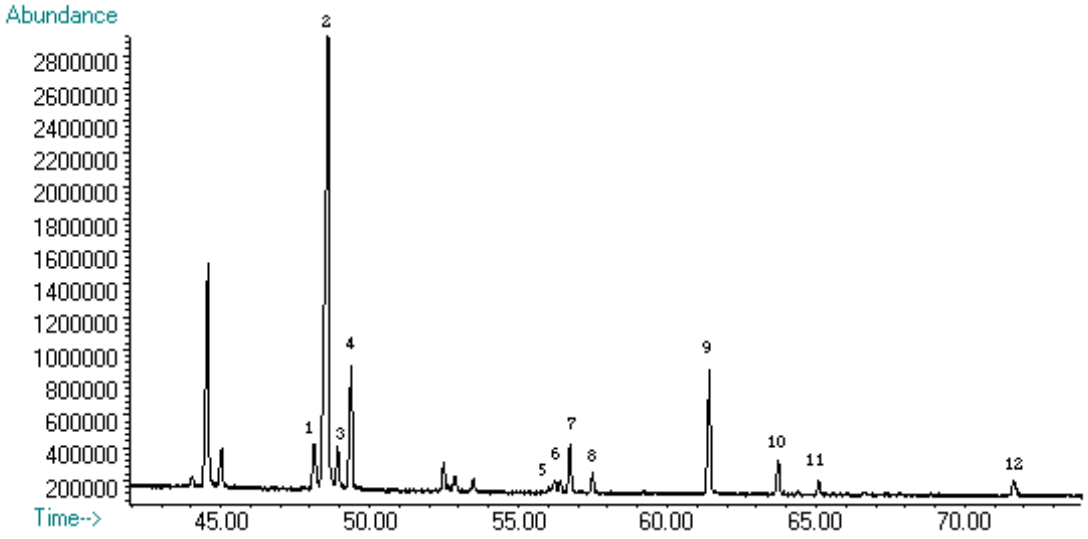
Results

The pheromone profiles of 92 male and 94 female IG lines were obtained, accounting for a total of 930 scored flies. Comparing the male and female phenotypes seen in the gas chromatographs (Figure 1), there were almost no major differences in the types of hydrocarbons expressed. This was especially true in the pheromone regions where the peaks are numbered (Figure 1). Despite this observation, the amount of each hydrocarbon in each sample varied between males and females, calculated as the area under each peak.

Hydrocarbon identification

Peak 2 seen in Figure 1 was consistently found throughout all IG lines, male and female. In all IG lines it was the most abundant of all hydrocarbons. Its identity as tricosene was determined in comparison with the data obtained by Antony and Jallon (1982) in which the pheromone profile of *D. melanogaster* Canton-S strain was identified (Table 3). Peak retention time and shapes corresponded almost identically between strains. As well, the available database library Wiley 138 identified peak 2 as tricosene to a 93% probability. Since 7-tricosene is known to be the most abundant of the tricosenes (Coyne, 1996), we postulate this to be the identity of peak 2 and use its area in the tricosene ratio calculations. Peaks number 1 and 3 are likely tricosene isomers, having similar retention times, but relatively little abundance. Possible identities for these peaks include 6-tricosene, 9-tricosene or methylated carbon-22 compounds. The areas under these peaks were not analyzed. The mass of the compounds for peaks 1-3, as determined by its mass spectrum, was 322 AMU.

A)



B)

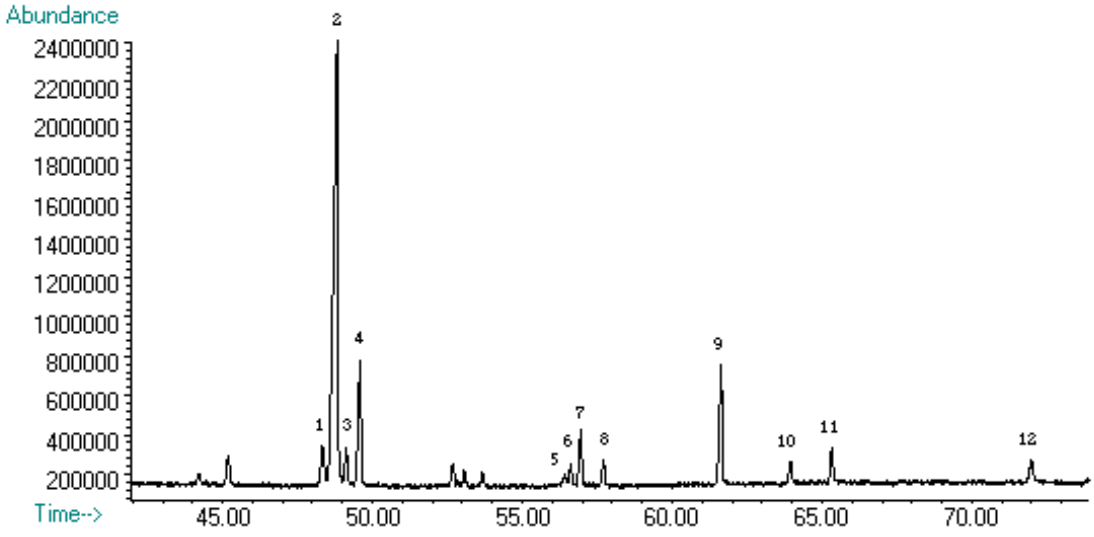


Figure 1: Gas chromatograms of the cuticular hydrocarbon profiles of 5-7 day old virgin males (A) and females (B) of IG line Q15. Peaks are numbered according to increasing retention times. See Table 3 for corresponding peak identification.

Table 3: Identification of the cuticular hydrocarbons of 5-7 day old virgin flies seen in Figure 1.

Hydrocarbon	GC peak #
Tricosene	1, 2, 3
Tricosane	4
Pentacosadiene	5
Pentacosene	6, 7
Pentacosane	8
Hexacosane (standard)	9
Heptacosadiene	-
Heptacosene	-
Heptacosane	-
Nonacosadiene	-
Nonacosene	-
Nonacosane	-

Peaks 5-8 represent the carbon-25 compounds. Pentacosenes were identified in some of the IG line profiles, but it was not considered for analysis due to its low and often indistinguishable abundance relative to the baseline and poor separation of peaks in the carbon-25 region.

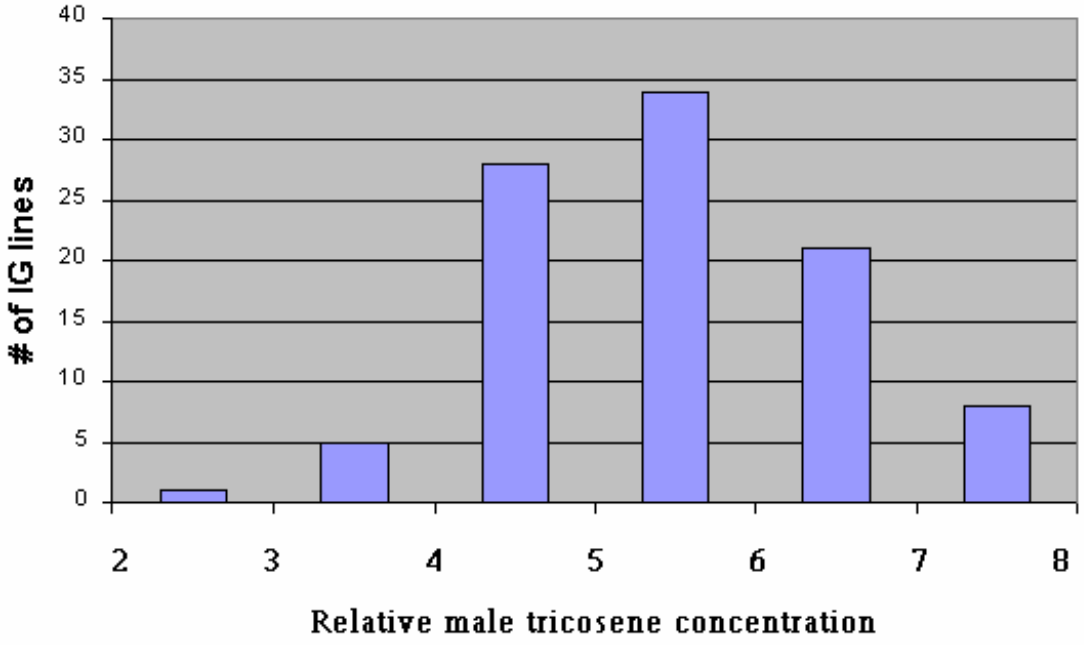
Using this method, heptacosadienes were not positively identified, thus could not be analyzed in this study. It is probable that due to the high concentration of *D. simulans* genes that make up the IG lines, the levels of heptacosadiene (and nonacosadiene) were undetectable.

Phenotypic variation

Figure 2A shows the distribution of tricosene levels observed within the male IG lines. Variations ranged from just over 2:1 to almost 8:1 tricosene to standard, proving that some IG male lines possess over 3.8 times more tricosene than others. Crude calculations of tricosene levels showed that levels range from 16.5ng 7-T/fly to 62.8ng 7-T/fly. These values are similar to those detected by Coyne (1996) for 7-T in various strains of *D. sechellia*, and are more than an order of magnitude less than the values for *D. simulans*.

The female IG lines demonstrate a wider range of 7-T levels. Variations range from just over 1:1 to over 10:1 tricosene to standard, some having more than 8 times more tricosene than others. Crude calculations of tricosene levels in the extreme phenotypes show they range from 9.8ng 7-T/fly to 82.2ng 7-T/fly. Like the male IG lines these levels were much lower than those reported in the literature for *D. simulans* flies (Coyne, 1996).

A)



B)

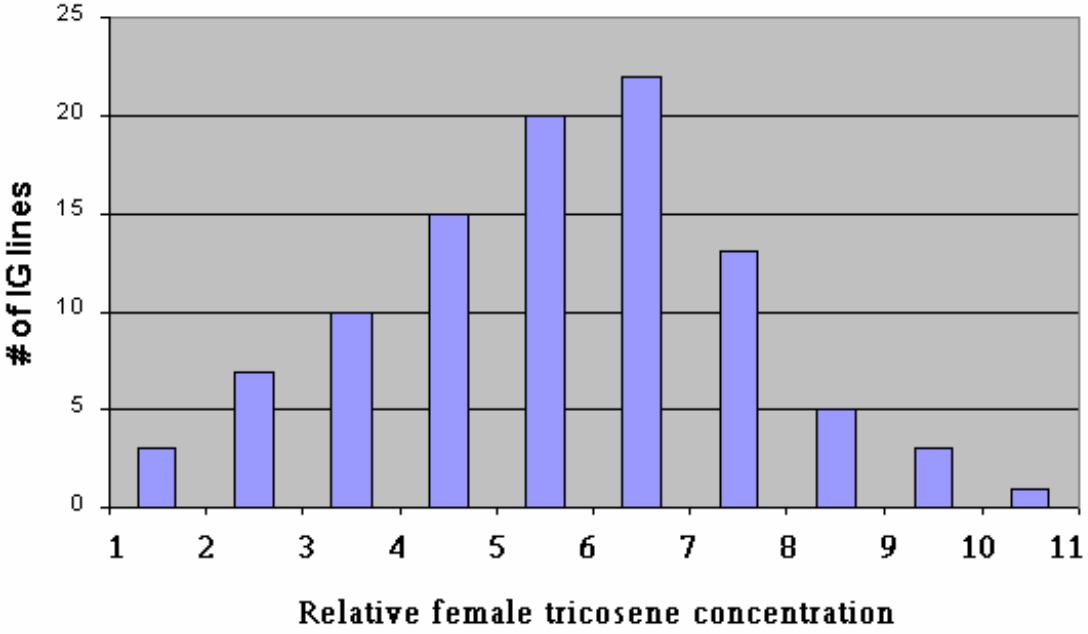


Figure 2: Number of IG lines showing different relative concentrations of tricosene to hexacosane standard in males (A) and females (B).

The mean values for tricosene ratios in males and females were 5.4628 ±1.0623 and 5.6291 ±1.8726, respectively. The similar ranges of these values within error were expected due to the *D. simulans* sexual monomorphism that exists between males and females.

Statistical tests

The distribution of the 92 male and 94 female data sets were tested for normality using QTL cartographer. As seen in Table 4, the S values for male and female tricosene ratios were 1.4177 and 0.1158, respectively, following a chi-square distribution ($S < \chi_2^2;0.95 = 5.99$). The values for skewness were –0.2079 for male and –0.3230 for female lines, while kurtosis values were 4.4560 for males and 35.1544 for females.

Marker-trait associations

Using a linear regression analysis, associations were made between markers and the phenotype. In males, marker 7 on the X chromosome has a significant effect on the phenotype, as seen by the F-statistic of 5.713 (Table 5), which is statistically significant at a 5% level ($p = 0.018$). The positive b_1 value of 0.319 represents the slope of the regression. This positive b_1 value indicates that having a *D. simulans* allele at this marker significantly increases the tricosene concentration. In females, marker 3 on the X chromosome made a significant association with tricosene values, as seen by the F-statistic of 5.348 where $p = 0.022$ (Table 6). The negative b_1 value indicates that having a *D. sechellia* allele at this marker significantly increases the tricosene concentration. Marginal significance ($0.05 < p < 0.10$) was found for markers in both males and females.

Table 4: Summary of statistical analyses performed on tricosene data in both male and female IG lines. Data demonstrates normal distribution as denoted by the S-value.

Statistic	**Male**	**Female**
Sample size	92	94
Mean trait value	5.4628	5.6291
Standard Deviation	1.0623	1.8726
Skw	-0.2079	-0.3230
SE skw	0.2554	0.2526
Kur	4.4560	35.1544
SE kur	0.5108	0.5053
k_3	-0.1734	-0.0492
k_4	0.4996	-0.1411
S (5%: 5.99, 1%: 9.21)	1.4177	0.1158

Table 5: Linear regression analysis of male tricosene levels with respect to the 23 scored molecular markers.

Chrom.	Marker	Cytological position (*mel*)	Genetic position (*sim/sec)*	b0	b1	F(1,n-2)	P
1	1	3C2	1-3.6	5.461	0.019	0.023	0.878
1	2	4F1-2	1-8.9	5.482	-0.049	0.029	0.866
1	3	7C4-5	1-14.3	5.541	-0.214	0.392	0.532
1	4	10A2-4	1-23.7	5.426	0.098	0.137	0.712
1	5	13E3-8	1-42.7	5.486	-0.093	0.464	0.497
1	6	16E4-F1	1-49	5.461	-0.113	1.035	0.31
1	7	19E3	1.57.9	5.385	0.319	5.713	0.018**
2	8	22C3-D1	2-13	5.396	0.17	0.221	0.638
2	9	26A5	2-22	5.234	0.587	3.648	0.057*
2	10	27F5-6	2-28.5	5.397	0.198	1.055	0.305
2	11	31D11-E1	2-47.8	5.228	0.57	0.283	0.595
2	12	35B10-C1	2-64	5.479	-0.07	0.272	0.603
2	13	38E5-6	2-71.9	5.217	0.604	0.63	0.428
2	14	50C23-D3	2-100.5	5.464	-0.008	0.004	0.95
2	15	53F13-54A2	2-115.3	5.364	0.247	0.421	0.517
2	16	58A4-B1	2-146.5	5.49	-0.109	0.605	0.438
3	17	62B4	3-0.2	5.49	-0.088	0.27	0.604
3	18	65C3	3-20.3	5.512	-0.186	1.722	0.191
3	19	67B10	3-30.8	5.457	0.025	0.032	0.857
3	20	75D7-E1	3-71.7	5.528	-0.216	1.767	0.185
3	21	84C1-4	3-134.6	5.453	0.026	0.005	0.942
3	22	90C10-D1	3-97	5.542	-0.23	1.375	0.242
3	23	94E11-F1	3-141.5	5.505	-0.127	0.46	0.499

** $P < 0.05$

* $0.05 < P < 0.1$

Table 6: Linear regression analysis of female tricosene levels with respect to the 23 scored molecular markers.

Chrom.	Marker	Cytological position (*mel*)	Genetic position (*sim/sec)*	b0	b1	F(1,n-2)	P
1	1	3C2	1-3.6	5.624	0.047	0.046	0.83
1	2	4F1-2	1-8.9	5.535	0.242	0.221	0.638
1	3	7C4-5	1-14.3	6.238	-1.565	5.348	0.022**
1	4	10A2-4	1-23.7	5.373	0.682	2.659	0.104
1	5	13E3-8	1-42.7	5.54	0.378	2.701	0.102
1	6	16E4-F1	1-49	5.639	-0.365	3.649	0.057*
1	7	19E3	1.57.9	5.505	0.456	3.387	0.067*
2	8	22C3-D1	2-13	5.594	0.093	0.048	0.827
2	9	26A5	2-22	5.371	0.633	0.893	0.346
2	10	27F5-6	2-28.5	5.54	0.256	0.518	0.472
2	11	31D11-E1	2-47.8	5.58	0.121	0.036	0.849
2	12	35B10-C1	2-64	5.616	0.054	0.051	0.822
2	13	38E5-6	2-71.9	5.513	0.286	0.134	0.714
2	14	50C23-D3	2-100.5	5.62	0.095	0.196	0.658
2	15	53F13-54A2	2-115.3	5.35	0.701	1.639	0.202
2	16	58A4-B1	2-146.5	5.617	0.047	0.038	0.846
3	17	62B4	3-0.2	5.526	0.335	1.44	0.231
3	18	65C3	3-20.3	5.605	0.093	0.141	0.708
3	19	67B10	3-30.8	5.625	0.019	0.007	0.935
3	20	75D7-E1	3-71.7	5.651	-0.074	0.074	0.785
3	21	84C1-4	3-134.6	5.897	-0.657	0.963	0.327
3	22	90C10-D1	3-97	5.754	-0.373	1.427	0.233
3	23	94E11-F1	3-141.5	5.546	0.235	0.457	0.5

** $P < 0.05$

* $0.05 < P < 0.1$

In males, marker 9 on the 2nd chromosome had an F-statistic of 3.648 (Table 5) and a positive b_1, indicating the same directionality as the other male marker. The female data marginally associated with both the 6th and 7th marker on the X chromosome, demonstrating positive and negative b_1 values, respectively.

Interval mapping

Using interval mapping techniques, regions along the genome that have a high gene-trait associations were detected. In Figure 3, 4 and 5 likelihood ratios exceed 3.84 (5% significance at n-1 d.f.) and were denoted with arrows. In Figure 3, tricosene production among male IG lines was found to be significant at genetic position 57 (LR = 4.13), which lies in a region between markers 6 and 7, very close to the 7th marker.

In female IG lines, a significant association was made with genetic position 42.7 along the X chromosome. This region is right at marker 5, which was not detected using linear regression analysis. Another female association was made at genetic position 21 on the second chromosome, which lies between the 8th and 9th marker, just flanking the 9th marker. This gene region was marginally associated with male tricosene levels using linear regression analysis.

Associations made with chromosome 1

A significant finding in this study was the association made between tricosene levels and marker 7 on male chromosome 1. Among the male IG lines, this was the only significant marker association found using linear regression analysis. Interval mapping also associated male tricosene levels with genetic position 1-57, which lies in a region

between markers 6 and 7, flanking marker 7. Markers 6 and 7 on chromosome 1, although demonstrating opposing directionality, made marginal associations among female IG lines.

As determined by the p-values in Table 6, the region between markers 6 and 7 is not the only region where associations can be made within chromosome 1. Marker 3 is strongly associated with tricosene levels in female IG lines, showing the opposite directionality compared to marker 7, the only significant male marker. Marker 5 also makes an association with tricosene levels in the female IG lines. Between markers 3 and 7, all the p-values range from 0.022 to 0.104. Although not statistically significant ($p<0.05$), markers 4, 5, 6 and 7 are all worth investigating further because none of the other p-values on any of the other markers was below 0.202.

Selected candidate genes

FlyBase database searches were conducted for cytological intervals where QTLs explaining variation in pheromone production were detected. The gene *Pheromone-binding protein related protein (Pbprp2)* at cytological position 19D2 was selected as a candidate gene for tricosene production or expression in the IG lines. Its function as a pheromone binding protein has been putatively established, but its function as a pheromone-producing gene has yet to be determined. The gene *Odorant receptor 19a* (*Or19a*) at cytological position 19C1 was also identified as a candidate gene for tricosene production.

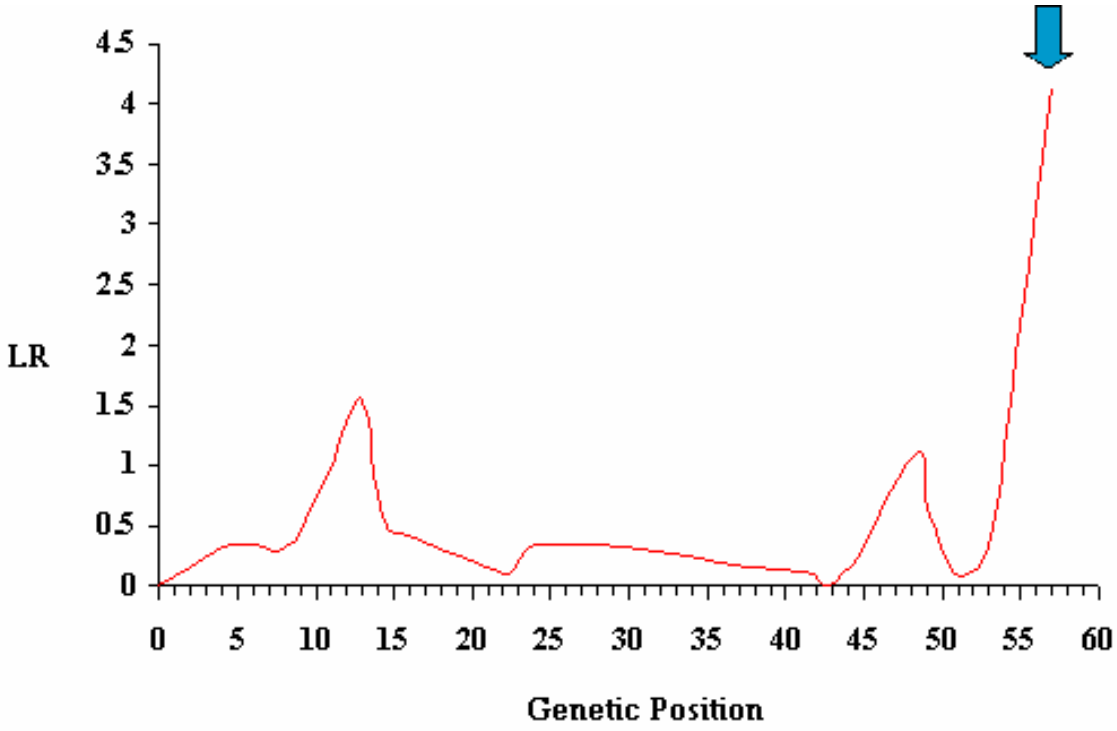


Figure 3: Interval map of male chromosome 1 showing an association between tricosene levels and a genetic region between markers 6 and 7 (Approximate genetic position 1-57).

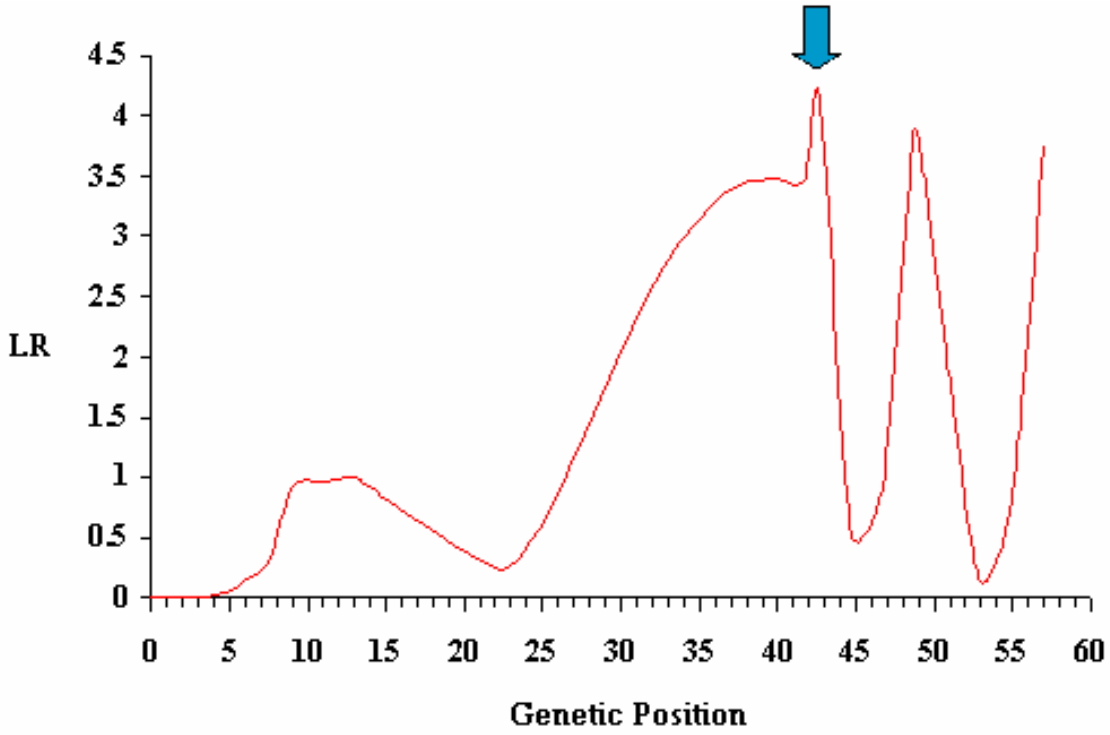


Figure 4: Interval map of female chromosome 1 showing an association between tricosene levels and a genetic region close to marker 5 (Approximate genetic position 1-42.7).

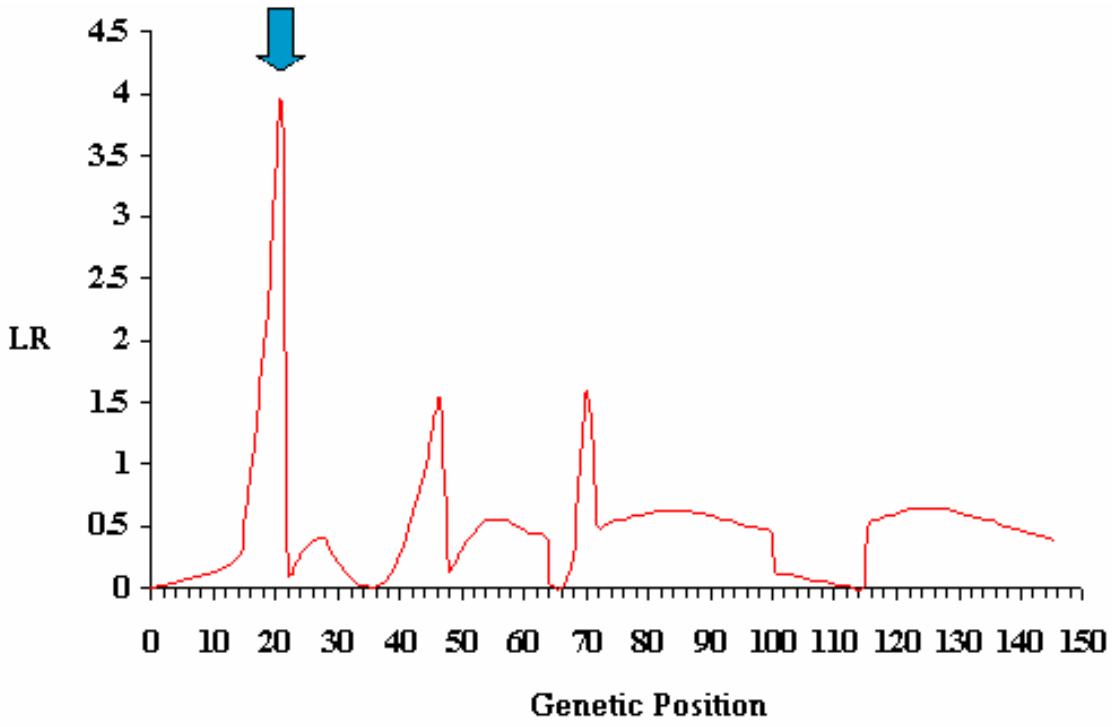


Figure 5: Interval map of female chromosome 2 showing an association between tricosene production and the approximate genetic position 21, which lies between markers 8 and 9.

Discussion

Chromosome 1: Genetic associations with tricosene levels and potential candidate genes

The positions of the significant male and female markers (females 3-7, male 6-7) overlap between markers 6 and 7, where there could potentially be similar genes responsible for tricosene levels in both sexes. The directionality of marker 7 in both sexes indicates that the *D. simulans* allele is responsible for increased tricosene levels. In females, we know that *D. simulans* alleles produce more tricosene (Table 1); therefore, a gene or group of genes that was detected in this region could be responsible for this phenotype.

Looking at the same region, marker 6 demonstrates that the *D. simulans* allele is responsible for lower levels of tricosene among IG females, as demonstrated by the negative b_1 value. Thus, there is likely more than one gene affecting tricosene levels in IG females. In *D. simulans*, as well as *D. sechellia* males, this gene or group of genes could prove to regulate the levels of tricosene produced, while in *D. sechellia* females, the effects of this gene could be inhibited by the presence of other genes, allowing for expression of the normal 7,11-HD phenotype. These other genes could be those that fall in the range of nearby markers (3-7).

The significance of markers 3-7 can be interpreted in different ways. Firstly, these marker-trait associations (other than one near marker 5) were not detected using interval mapping techniques. It is possible that the significance of the linear regression analysis across the range of markers could be due to the influence of neighbouring markers that amplify the probability of association in an attempt to locate major genes. Using interval mapping, this phenomenon may have been alleviated by focusing in on

small map intervals. Conversely, it is possible that due to lack of density of the marker map, many smaller, highly significant regions within these markers were left undetected and only the flanking regions were identified with marginal significance. Regardless, tricosene levels are highly associated with chromosome 1. Candidate genes for tricosene production or expression are likely to be found in this region.

FlyBase was used to probe for candidate genes in this region: two candidates were identified. The gene *Pbprp2* is located on chromosome 1 at the cytological position 19D2, corresponding to a genetic position that flanks marker 7. It is a good candidate for affecting tricosene levels because of its supposed function as a pheromone/odorant-binding protein, which was determined by its structural similarity to other taste receptors. This gene is expressed in the antennal segment 3, maxillary palpus and proboscis, which are regions where pheromonal hydrocarbons are received and interpreted. It may be worthwhile to consider studying this gene in more detail. This could be accomplished by analyzing the phenotypes between lines of flies with variable alleles for this gene. As well, a higher resolution marker map could prove to make more significant marker-trait associations.

The other candidate gene is located at cytological position 19C1, which is located between markers 6 and 7 on chromosome 1. The gene is known as *Or19a* and is characterized as an integral membrane protein that functions as an olfactory receptor. Although involvement in the production of pheromones has not yet been identified, this receptor could prove to initiate a signaling pathway ultimately leading to the production or regulation of tricosene levels. Of course, odors in the form of pheromones are passed from one fly to another and are received by odorant receptors as a part of the mating

recognition process. For this reason, odorant receptors are good candidates for being implicated in pheromone production or expression.

Chromosome 2

Interval mapping detected a region on female chromosome 2 (genetic positions 2-21) that was not detected using linear regression analysis. Strangely, linear regression analysis detected a marginally significant association between tricosene levels and male marker 9 (genetic position 2-22) while interval mapping in this region lacked any points of interest. This is not likely a coincidental set of events; it could be a gene region where tricosene levels may be controlled. At marker 9 in both males and females, the directionality of the marker suggests that the *D. simulans* allele increases tricosene levels. Since genetic positions detected (2-21 and 2-22) are relatively close together, it can be assumed that there is one gene in the region controlling tricosene levels in both males and females. If such is the case, it follows that the *D. simulans* allele would be most suitable for expressing this phenotype because both male and female *D. simulans* predominantly produce tricosene as its major pheromonal hydrocarbon. Chromosome 2 may prove important in finding genetic regions responsible for pheromonal variations between the closely related species *D. simulans* and *D. sechellia*.

The study by Ferveur (1991) associated the genetic position 2-65.3 with variation in the ratio of 7-T to 7-P. In this study, however, no associations near genetic position 2-65.3 were made with respect to tricosene levels alone. Therefore, it is likely that the association detected by Ferveur was due to changes in 7-P levels that altered the ratio of

7-T to 7-P. Further studies dealing with variation in the levels of 7-P should be aware of this gene region as a potential candidate loci for pheromone-producing genes.

Using FlyBase, the genetic region between markers 8 and 9 was probed for candidate genes for tricosene production. No candidates were found, however, this interval within chromosome 2 is a good region to create a higher resolution molecular marker map. This way, new and more powerful associations could potentially be made between markers and pheromones.

Chromosome 3

Chromosome 3 did not show any significant associations with tricosene levels in either the male or female IG lines. This result was surprising considering that Coyne *et al.* (1994) used chromosomal markers to make associations with 7-T to 7,11-HD ratios. The discrepancy between this study and that of Coyne *et al.* (1994) may be attributed to varying levels of 7,11-HD caused by genes in chromosome 3 that were detected by Coyne. If such is the case, an effort should be made to look further into chromosome 3 for potential candidate genes for 7,11-HD production in *D. sechellia*.

Resolving the molecular marker map

Based on the results obtained in this study combined with the results of other studies, there is more reason to continue increasing the density of the molecular marker map. Chromosome 1 has revealed a region of genes between markers 3 and 7 that may have likely candidate loci involved in tricosene expression. Chromosome 2 has potential for identifying candidate genes in the genetic region 2-21, including other genes for 7-P

production while chromosome 3 could potentially house genes involved in 7,11-HD production. In all of these regions, a denser marker map could aid in making more precise associations between markers and phenotypes.

Although only found in males, the Y chromosome is a good place to look for genes involved in pheromone production, considering the large variations that exist between males and females of sexually dimorphic species. Also, chromosome 4 should not be ignored even though it contains few genes. Only one quarter of the genes on this chromosome have been identified so far, making it possible for pheromone producing genes to be present. Recently, researchers at University of Chicago dispelled the classic genetic idea that all species of *Drosophila* carry the same chromosome 4 due to a lack of recombination (Wang *et al.*, 2002). They detected polymorphisms in a 200 kilobase region of the chromosome, suggesting that chromosome 4 has a variable evolutionary history.

Other pheromonal hydrocarbons

The hydrocarbon profiles of the IG lines clearly show that tricosene is the most abundant of all the cuticular hydrocarbons detected by GC-MS. However, this does not rule out the potential effects and influences of other pheromonal hydrocarbons. For example, 7,11-nonacosadiene (7,11-ND) is known as a potent pheromone that alters male mating behaviour in some species of *Drosophila* (Ferveur *et al.*, 1996). The dose of 7,11-ND required to stimulate males is very minute compared to the levels of 7,11-HD, meaning it is likely that although undetected, these hydrocarbons could be present. In order to properly visualize and identify these compounds, extractions need to be carried

out using more flies per extraction, or by using less hexane for dilution. This would result in more distinguishable peaks from the background.

A large portion of the literature pertaining to marker-based studies on pheromones examines the ratios of one hydrocarbon to another and their significance (Antony and Jallon, 1982; Coyne, 1994; Coyne and Charlesworth, 1997; Ferveur 1991). This type of analysis, unfortunately, was not performed in this study due to the inability to positively identify other peaks such as 7,11-HD and 7,11-ND. Although 7-P peaks were visible, they were not consistent throughout the IG lines, and often could not be properly distinguished from the baseline reading. Again, a different extraction method needs to be used to analyze such compounds.

There is, however, an advantage to measuring the quantity of individual hydrocarbons without analyzing ratios. Using an internal standard (hexacosane), this study was able to monitor absolute variations in tricosene levels without having to analyze ratios of one hydrocarbon to another. Associations made in this study between markers and traits were due to tricosene alone, narrowing down the possible function of the detected candidate gene region to the production of one hydrocarbon, not two or more.

Future Directions

In order to provide more solid grounds for the results presented, replicas of the hydrocarbon profiles of all the IG lines must be determined using GC-MS. If this were possible, it would show that the hydrocarbon profiles of each IG line are not chance measurements, but are found by consistent experimental techniques that can be

reproduced with confidence. In the near future, I plan to repeat the extraction protocol used in this study for three more sets of each IG line using the same GC-MS method.

Conclusions

1. Chromosome 1 shows highly significant associations between male tricosene production and a genetic region between markers 6 and 7. Female tricosene levels are marginally significant in this marker region. *Pbprp2* and *Or19a* were selected as potential candidate genes for tricosene production.

2. Associations between tricosene levels and chromosome 2 were detected in both males and females near the genetic region 2-21, which may control tricosene levels in *D. simulans*; however, no candidate genes were identified.

3. Chromosome 3 detected no significant marker-trait associations.

4. Based on the literature and this study, unidentified genes in chromosome 2 and chromosome 3 may be good candidates for 7-P and 7,11-HD production, respectively.

5. Increasing the resolution of the molecular marker map will be critical in narrowing down regions of the genome associated with the production of 7-T and other pheromones.

References

Antony, C., J-M. Jallon. 1982. The Chemical Basis for Sex Recognition in *Drosophila melanogaster*. Journal of Insect Physiology 28 (10): 873-880.

Chapman, R. F. 1982. The Insects: Structure and Function 3rd Edition. Harvard University Press, Cambridge, MA: 502-510.

Civetta, A., H. M. Waldrip-Dail, A. G. Clark. 2002. An introgression approach to mapping differences in mating success and sperm competitive ability in *Drosophila simulans* and *D. sechellia*. Genetical Research, Cambridge 79: 65-74.

Coyne, J. A. 1996. Genetics of a Difference in Male Cuticular Hydrocarbons Between Two Sibling Species, *Drosophila simulans* and *D. sechellia*. Genetics 143: 1689-1698.

Coyne, J. A., A. P. Crittenden, K. Mah. 1994. Genetics of a Pheromonal Difference Contributing to Reproductive Isolation in *Drosophila*. Science 265: 1461-1464.

Coyne, J. A., B. Charlesworth. 1997. Genetics of a Pheromonal Difference Affecting Sexual Isolation Between *Drosophila mauritiana* and *D. sechellia*. Genetics 143:1015-1030.

de Bruyne, M., P. J. Clyne and J. R. Carlson. 1999. Odor coding in a model olfactory organ: the Drosophila maxillary palp. Journal of Neuroscience. 19 (11): 4520-4532.

Ferveur, J-F. 1991. Genetic Control of Pheromones in *Drosophila simulans*. I. *Ngbo* a Locus on the Second Chromosome. Genetics 128: 293-301.

Ferveur, J-F., J-M. Jallon. 1993. Genetic Control of Pheromones in *Drosophila simulans*. II. *kété*, a Locus on the X Chromosome. Genetics 133: 561-567.

Ferveur, J-F., M. Cobb, H. Boukella, J-M. Jallon. 1996. World-wide variation on *Drosophila melanogaster* sex pheromone: behavioural effects, genetic bases and potential evolutionary consequences. Genetica 97: 73-80.

Ferveur, J-F., Savarit, F. 1997. Genetic feminization of pheromones and its behavioral consequences in *Drosophila* males. Sceince 276 (5318): 1555-1558.

Ferveur, J-F., G. Sureau. 1996. Simultaneous influence on male courtship of stimulatory and inhibitory pheromones produced by live sex-mosaic *Drosophila melanogaster*. Proceedings of the Royal Society of London B263: 967-973.

Greenspan, R. J. 1997. Fly Pushing: The Theory and Practice of *Drosophila* Genetics. Cold Springs Harbor Laboroatory Press. Plainview, NY: 19-21.

Griffiths, A. J. F., J. H. Miller, D. T. Suzuki, R. C. Lewontin, W. M. Gelbart. 1996. An Introduction to Genetic Ananlysis. 6th Edition. W.H. Freeman and Company. New York, NY: 214-216.

Jallon, J-M., J. R. David. 1987. Variations in Cuticular Hydrocarbons Among Eight Species of the *Drosophila melanogaster* Subgroup. Evolution 41 (2): 294-302.

Lemeunier, F., J. R. David, L. Tsacas, M. Ashburner. 1986. The *melanogaster* species group. Genetics and Biology of *Drosophila* (3E): 147-256.

MacDonald, S. J., D. B. Goldstein. 1999. A Qunatitative Genetic Analysis of Male Sexual Traits Distinguishing the Sibling Species *Drosophila simulans* and *D. sechellia.* Genetics 153: 1683-1699.

Savarit, F., G. Sureau, M. Cobb, J-F Ferveur. 1999. Genetic elimination of known pheromones reveals the fundamental chemical bases of mating and isolation in *Drosophila*. Proceedings of the National Academy of Science, USA 96: 9015-9020.

Sokal, R. R., F. J. Rohlf. 1995. Biometry 3rd Edition. W. H. Freeman and Company, New York, NY: 98-115

Wang, S., C. J. Basten, Z-B. Zeng. 2001. QTL Cartographer, Version 1.30. Department of Statistics, North Carolina State University, Raleigh, N.C.

Wang, W., K. Thornton, A. Berry, M. Long. 2002. Nucleotide Variation Along the *Drosophila melanogaster* Fourth Chromosome. Science 295 (5552): 134-137.