# DIFFERENTIATION OF GENE EXPRESSION IN BEHAVIORALLY ISOLATED BLACK AND BROWN MORPHS OF *DROSOPHILA ELEGANS*

by

DAIJU HOSHINO

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN BIOLOGY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2007

ACKNOWLEDGEMENTS

This research was supported by a UTA's Research Enhancement Program grant (awarded to Pawel Michalak), and an Animal Behaviour Society grant (awarded to John H. Malone). I thank my primary advisor Pawel Michalak for substantial support for my research and academic career, a PhD candidate student at our laboratory John H. Malone for intensive support for research and discussion, an undergraduate student research assistant Ivan T. Lee for help of fly mating experiments and mRNA extraction, a graduate student Daina Ma for help with flies, two of great committee members Esther Betrán and Jaroslaw Krzywinski for help with master thesis editing, all members from Genome Biology Group for giving me ideas on research, and my family and fiancée for supporting my academic career and growth throughout entire college education.

April 17, 2007

ABSTRACT

DIFFERENTIATION OF GENE EXPRESSION IN BEHAVIORALLY ISOLATED BLACK AND BROWN MORPHS OF *DROSOPHILA ELEGANS*

Publication No. _____

DAIJU HOSHINO, M.S.

The University of Texas at Arlington, 2007

Supervising Professor: Pawel Michalak

Pre-zygotic isolation has been recognized as the primary form of incipient speciation, with behavioral isolation as one of its most significant manifestations. In spite of intensive research on behavioral isolation, its molecular mechanism is still poorly understood. We present microarray data from two morphs, black (OH) and brown (HK), of *Drosophila elegans*. Since only OH females show strong preference to mate with same type of males, these two morphs provide a model system for investigations into the incipient stage of speciation.

We analyzed the expression profiles from heads of mated OH females and non-mated females of both morphs. Subsequent analysis of candidate genes by GoMiner showed that OH unique genes were enriched in the visual system related categories, and

these categories were still enriched with candidate genes whose expression levels were modified after mating. These results suggest that *Drosophila elegans* might use visual cues mainly as mate recognition.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

The new developments in comparative genetics and genomics hold promise for providing answers to many questions about the origin of new species (speciation). Reproductive isolation between species has been recognized as the essential step to speciation (Dobzhansky 1937; Mayr 1963). While post-zygotic isolation has been widely accepted as the main cause of evolutionary divergence of species, pre-zygotic isolation is considered to be involved in incipient speciation (Coyne and Orr 1998). Several theoretical models of pre-mating isolation have been proposed, with assortative mating by direct mate choice playing a central role (Rundle *et al*. 2005). Mate discrimination has been observed in many closely related species, including *Drosophila melanogaster*. Zimbabwe and Cosmopolitan morphs of *D. melanogaster* provide a good model to investigate asymmetrical mating that results in prezygotic isolation. However, in spite of the intensive effort of many researchers, the genetic mechanism of this mating discrimination needs yet to be resolved (Chenoweth and Blows 2006).

Here we use *Drosophila elegans*, another *Drosophila* system that shows partial sexual isolation between two different morphs (Hirai and Kimura, 1997; Ishii et al, 2001; Ishii et al, 2002; Kopp and True 2002). *D. elegans* is an Oriental species from the *D. melanogaster* subgroup and is characterized by distinct morphological differences

between geographically isolated populations - in particular brown and black body color (Bock and Wheeler 1972; Kopp and True 2002). Several numbers of strains have been described in *D. elegans*, and female choice experiments showed the asymmetrical behavioral isolation between the black morph and the brown morph. Females Black morphs of *D. elegans* preferably mate with their own male and discriminate against brown morphs of *D. elegans*, while females brown morphs of *D. elegans* mate randomly (Hirai and Kimura 1997). The correlation between adaptive changes and their traits in *D. elegans* has been investigated, but there was virtually no attempt to determine the molecular basis of morphological and behavioral differences between the black and brown morphs (Kalmus 1941; Jacobs 1968).

In our previous study, we identified a number of candidate genes potentially involved in sexual isolation between Zimbabwe morphs and Cosmopolitan morphs of *D. melanogaster* (Michalak et al. 2007). In this study, we have used similar methodology to show that there also are transcriptional differences between two morphs of *D. elegans* associated with their behavioral isolation. We hypothesized that expression of genes which respond to external stimuli, such as olfactory, auditory, or visual cues, would be different at the transcriptional level, providing good candidate genes for behavioral isolation and incipient speciation.

Here, we present analysis of the differential expression of genes between two strains of *D. elegans,* OH (black morph) and HK (brown morph). Using Affymetrix microarray technology, we analyzed expression profiles from female heads to compare the differential expression of genes between the two morphs. We found that several

genes relating to visual system were upregulated in the OH strain and genes involved in neuronal network were more expressed after mating behavior OH female. Some of these genes were involved in olfactory guidance behavior.

# CHAPTER 2

# MATERIALS AND METHODS

## 2.1 Flies

Fly stocks were reared in uncrowded cultures at 24°C with a 12 hours light-dark cycle on standard sugar-yeast-agar medium. Two strains of *D. elegans*, OH as the black morph and HK as the brown morph, were provided by John True and Masahito Kimura.

## 2.2 Mating Experiment

Virgin females and males were collected and aged for 7 days in dark. Around 9 to 10 am on the day 7, females of one strain were put into a chamber containing both same strain males and different strain males for one hour. Copulating flies were aspired into separate vials, and females were categorized into two groups; one was labeled as mated with OH male and the other was labeled as mated with HK male. Females of the HK strain were treated in the same way.

## 2.3 RNA Collection

Heads of 7-day old mated and non-mated OH females as well as non-mated HK females were used for RNA extraction. We focused on transcriptional changes

associated with mating in the OH strain, for it is the discriminating group of *D. elegans*, which would be more informative than that of the HK strain. For mated females, RNA was extracted within 30 minutes after mating. Twenty female heads were used per sample. At least five micrograms of total RNA were extracted with RNApure and the manufacture's protocol (GenHunter Corp., Nashville, TN), followed by purification with GeneChip Sample Cleanup Modules (QIAGEN).

## 2.4 Affymetrix *Drosophila* GeneChip® Profiles

By using the ENZO BioArray High Yield RNA transcript labeling kit, cRNA was synthesized and biotin-labeled in an in vitro transcription reaction. The target cRNA was hybridized to Affymetrix GeneChip® Drosophila arrays. Although this microarray tool was designed for studying expression of *D. melanogaster* transcripts, we show that the genetic material from *D. elegans* also successfully hybridizes to these arrays, reflecting relatively close taxonomic positions of these two species (Ashburner et al. 1984; Schawaroch 2002; Kopp 2006). Probes for a particular transcript are present in 28 cells on the array chip. Half of those cells possess Perfect Match (PM) sequence and the other half contains Mismatch (MM) sequence. Affymetrix Microarray Suit (MAS) 5.0 software computes the average of the logarithm of the PM:MM ratio for each probes. This software has also the detection algorithm which calculates the Discrimination Score from each probe pair and a Detection P value from the one-sided Wilcoxon Signed Rank text that assigns a P (Present), M (Marginal), or A (Absent) call

in relation to a threshold value Tau = 0.015. Arrays were hybridized and scanned at the Microarray Core Facility at the University of Texas Southwestern Medical Center in Dallas. Overall, we obtained 4 chips of non-mated OH (nMOH), 2 chips of non-mated HK (nMHK), and 2 chips of mated OH (MOH).

## 2.5 Statistical Analysis

For mating experiment, replicated goodness-of-fit test (replicated G-test) was used to test deviations from random mating for both OH and HK females. The expected ratio was calculated by multiplying the ratio of the male by the total number of mated females for each replicate.

For microarray data analysis, in order to minimize the effect of inefficient hybridization due to sequence divergence, only transcripts detected as P call were analyzed. To identify up- and down-regulated genes, the Rank Product (RP) analysis was adopted (Breitling et al. 2004). Two comparisons were analyzed, one between the strains, and the other between before and after mating of the OH strain. Since we had 4 chips for nMOH and 2 chips for nMHK and MOH, 8 sets of fold change were obtained for each comparison. Ranks were given to each list of fold change from the highest to the lowest and the rank product values (RP-values) were calculated by multiplying the ranks for each gene (*l*i) and dividing them by the total number of the genes (n) to the power of the number of sets of fold change (k), which is shown by the following formula;

RP-value (for each gene) = $\frac{\prod_{i=1}^{k} li}{n^k}$

The total number of the genes was 605 and the power of the fold change sets was 8. Significance level was obtained by converting rank product values of each gene to E-value in analogy to the BLAST results (Breitling et al. 2004). In order to set the threshold for E-value, we permutated the ranks of each fold change list, and then got new RP-values. This simulative permutation tests were done 1000 times. These newly generated RP-values were compared to the original RP-values, and the number of times exceeding original RP-values by newly generated ones was counted and divided by the total number of permutation, which gave us E-value for each gene. Genes whose E-values were less than the conventional criteria (0.05) were selected as candidate genes for further analysis.

In order to investigate biological significance of these differentially expressed genes, we analyzed them also in the context of Gene Ontology (GO). GO is organized in a hierarchy of annotated terms, and this facilitates to explore the biological significance of candidate genes (Beissbarth 2006). Recently, several programs to find GO categories enriched with candidate genes have been developed and become available, and GoMiner is one of them (Lomax 2004). GoMiner was developed especially for biological interpretation of microarray data. It adopts the one-sided Fisher's exact test to assess the null hypothesis; there was no difference between the proportion of candidate genes falling into the category and that of candidate genes not falling into the category. Even though p-values were obtained by Fisher's exact test, the

result potentially includes many false positives. Therefore, to obtain lists of the GO categories enriched with up-regulated genes and down-expressed genes individually for each comparison, False Discovery Rate (FDR) with q-values were also estimated. This q-value is obtained by resampling algorithm (Zeeberg et al. 2003; Zeeberg et al. 2005).

# CHAPTER 3

# RESULTS

## 3.1 Mate-By-Female Choice Test

We performed mate-by-female choice test over ten days; five days for OH female and other five days for HK female. Each day, we had different total number of flies. Table 1 summarizes the number of mated males, unmated females, and unmated males (Table 3.1). Although previous studies reported that black morph (OH) females showed very strong preference to homotypic males and brown morph females did not (Hirai and Kimura 1997; Ishii et al. 2001; Ishii et al. 2002; Kopp and True 2002), our results suggested that brown morph females also preferably mates with black morph males ($G=16.250$, $df=5$, $p<0.01$) (table 3.1 and 3.3). Anova-like replicated G-test results are shown on table 3.2 and 3.3.

Table 3.1: Summary of multiple female choice tests

| | OH Female | | | | | HK Female | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Replicates | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| x OH Male | 42 | 53 | 37 | 59 | 48 | 19 | 10 | 12 | 24 | 12 |
| x HK Male | 16 | 5 | 9 | 24 | 11 | 13 | 6 | 14 | 8 | 15 |
| Unmated Female | 17 | 6 | 12 | 10 | 5 | 10 | 8 | 7 | 13 | 4 |
| Unmated OH Male | 36 | 27 | 23 | 16 | 38 | 15 | 4 | 15 | 9 | 12 |
| Unmated HK Male | 64 | 72 | 52 | 36 | 75 | 22 | 10 | 16 | 34 | 24 |

Table 3.2: Total G and all of the each replicate show the significance non-random mating by OH female

OH Female

| Tests | df | G |
|---|---|---|
| Pooled | 1 | 96.191* |
| Heterogeneity | 4 | 11.074* |
| Total | 5 | 107.264* |

| Replicates | df | G |
|---|---|---|
| 1 | 1 | 12.748* |
| 2 | 1 | 44.526* |
| 3 | 1 | 18.758* |
| 4 | 1 | 8.452* |
| 5 | 1 | 25.031* |
| Total | 5 | 109.515* |

(* indicates significance.)

Table 3.3: Total G and only the replicate 4 show the significance non-random mating by HK female

HK Female

| Tests | df | G |
|---|---|---|
| Pooled | 1 | 8.577* |
| Heterogeneity | 4 | 7.673 |
| Total | 5 | 16.250* |

| Replicates | | |
|---|---|---|
| 1 | 1 | 1.312 |
| 2 | 1 | 1.616 |
| 3 | 1 | 0.015 |
| 4 | 1 | 12.695* |
| 5 | 1 | 0.453 |
| Total | 5 | 16.092* |

(* indicates significance.)

## 3.2 Differentially Expressed Gene Between OH Female and HK Female

Candidate genes were categorized into 4 different groups: group I, containing genes overexpressed in nMOH relative to nMHK, group II with genes underexpressed in nMOH relative to nMHK; group III with genes overexpressed in nMOH relative to MOH, and group IV with genes underexpressed in nMOH relative to MOH. As a result of the analysis by MAS 5.0, on average 1581 probes had P-call for each chip. We chose the probe sets that had P-call for all 8 chips and the other probe sets. This is a very conservative way to analyze microarray data, for it was shown that this filtering process increased the ratio of true positive to false positive (McClintick and Edenberg 2006). A total of 605 genes met these criteria. Of these 605 genes, differentially expressed genes

were produced by the RP-analysis, which is very powerful method with a small sample size (Jeffery et al. 2006). As a result, there were 91 genes in the group I and 91 genes in the group II (E-value < 0.05) (Table 3.4, 3.5, and 3.6).

Table 3.4: Summary of the number of genes differentially expressed in each contrast

| Category | Number of Genes |
|---|---|
| I: Up-regulated genes in non-mated OH compared to non-mated HK | 91 |
| II: Down-regulated genes in non-mated OH compared to non-mated HK | 91 |
| III: Up-regulated genes in non-mated OH compared to mated OH | 75 |
| IV: Down-regulated genes in non-mated OH compared to mated OH | 76 |

Table 3.5: Group I: Up-regulated genes in non-mated OH compared to non-mated HK

| Gene Symbol | RP-value | E-value | Function |
|---|---|---|---|
| | | | |
| for | 4.212E-20 | 0 | cyclic-nucleotide dependent protein kinase |
| maps | 1.651E-18 | 0 | |
| Lsp2 | 4.368E-17 | 0 | larval serum protein |
| beta-Spec | 1.950E-14 | 0 | actin cross-linking |
| BG | 5.640E-13 | 0 | |
| CG15381 | 1.383E-12 | 0 | |
| CG6287 | 2.883E-12 | 0 | enzyme |
| emc | 5.354E-12 | 0 | transcription co-repressor |
| Gdi | 5.411E-12 | 0 | GDP-dissociation inhibitor |
| wupA | 7.960E-12 | 0 | cytoskeletal structural protein |
| Act5C | 2.135E-11 | 0 | motor protein |
| CG15105 | 4.315E-11 | 0 | transcription factor |
| Pdk | 5.527E-11 | 0 | pyruvate dehydrogenase (lipoamide) kinase ; EC:2.7.1.99 |
| CG13666 | 8.690E-11 | 0 | |
| CG15893 | 1.451E-10 | 0 | |
| Taf110 | 1.617E-10 | 0 | transcription activating factor |
| Gbeta13F | 2.469E-10 | 0 | heterotrimeric G protein |
| CG7351 | 3.311E-10 | 0.001 | lysophospholipase |
| CG1534 | 3.664E-10 | 0 | |
| CG9318 | 3.763E-10 | 0 | |
| CG6963 | 5.047E-10 | 0 | |
| CG12431 | 5.308E-10 | 0 | |
| CG3534 | 9.407E-10 | 0.001 | xylulokinase |
| Gst2 | 9.515E-10 | 0 | glutathione transferase ; EC:2.5.1.18 |
| Rac1 | 9.638E-10 | 0 | RHO small GTPase |
| CG5827 | 1.148E-09 | 0.004 | structural protein of ribosome |
| Atpalpha | 1.376E-09 | 0 | sodium/potassium-transporting ATPase ; EC:3.6.1.37 |
| CG10078 | 1.394E-09 | 0 | enzyme |
| CG14503 | 1.487E-09 | 0.002 | |

(Table 3.5 - continued)

| | | | |
|---|---|---|---|
| CG12203 | 2.049E-09 | 0 | electron transfer |
| CG14844 | 2.914E-09 | 0.003 | |
| CG8723 | 3.068E-09 | 0.003 | enzyme |
| Arr2 | 3.196E-09 | 0 | |
| CG9057 | 3.523E-09 | 0.001 | |
| Cam | 3.627E-09 | 0 | calcium sensing |
| Argk | 3.685E-09 | 0 | arginine kinase ; EC:2.7.3.3 |
| CG8895 | 3.844E-09 | 0 | |
| SdhB | 4.163E-09 | 0 | succinate dehydrogenase-(ubiquinone) ; EC:1.3.5.1 |
| CG13943 | 4.914E-09 | 0 | |
| CG8818 | 6.379E-09 | 0.001 | |
| fax | 6.935E-09 | 0 | |
| CG9894 | 8.426E-09 | 0 | |
| eIF-4a | 1.096E-08 | 0.001 | ATP dependent RNA helicase |
| CG7510 | 1.192E-08 | 0.002 | |
| CG10640 | 1.250E-08 | 0 | enzyme |
| CG9336 | 1.348E-08 | 0.002 | |
| CG10664 | 1.764E-08 | 0.001 | enzyme |
| CG13315 | 1.905E-08 | 0.004 | |
| Arf79F | 2.080E-08 | 0.028 | GTP binding |
| Sdc | 2.349E-08 | 0.001 | heparin sulfate proteoglycan |
| robl | 2.538E-08 | 0.002 | motor |
| Pfk | 2.769E-08 | 0.003 | 6-phosphofructokinase ; EC:2.7.1.11 |
| CG6056 | 3.206E-08 | 0.01 | transporter |
| da | 3.534E-08 | 0.001 | specific RNA polymerase II transcription factor |
| CG15567 | 4.142E-08 | 0.004 | |
| Pu | 5.200E-08 | 0.003 | GTP cyclohydrolase I ; EC:3.5.4.16 |
| CG7888 | 5.420E-08 | 0.004 | transporter |
| CG16745 | 6.601E-08 | 0.001 | |
| BcDNA | 8.564E-08 | 0.014 | |
| CG10083 | 9.261E-08 | 0.005 | cell cycle regulator |
| CG13233 | 9.691E-08 | 0.007 | |
| igl | 1.249E-07 | 0.035 | calmodulin binding |
| fng | 1.323E-07 | 0.023 | |
| CG18290 | 1.511E-07 | 0.009 | |
| Dp1 | 1.958E-07 | 0.007 | single-stranded DNA binding |
| CG3662 | 2.252E-07 | 0.028 | structural protein |
| Act57B | 2.274E-07 | 0.013 | muscle motor protein |
| CG3874 | 2.645E-07 | 0.022 | transporter |
| BcDNA | 3.520E-07 | 0.013 | |
| Pabp2 | 3.577E-07 | 0.04 | poly(A) binding |
| BcDNA | 3.604E-07 | 0.005 | cytoskeletal structural protein |
| CG12143 | 3.943E-07 | 0.044 | |
| Gdh | 4.554E-07 | 0.02 | glutamate dehydrogenase (NAD(P) +) ; EC:1.4.1.3 |
| CG15081 | 4.608E-07 | 0.031 | |
| Rpn6 | 5.024E-07 | 0.028 | multicatalytic endopeptidase ; EC:3.4.99.46 |
| CG1548 | 5.212E-07 | 0.012 | |
| G-salpha60A | 6.441E-07 | 0.041 | heterotrimeric G protein |
| CG2159 | 7.345E-07 | 0.018 | diacylglycerol kinase |
| Ace | 8.220E-07 | 0.029 | acetylcholinesterase ; EC:3.1.1.7 |
| blw | 8.620E-07 | 0.029 | hydrogen-transporting ATP synthase ; EC:3.6.1.34 |

(Table 3.5 - continued)

| | | | |
|---|---|---|---|
| Rab1 | 8.720E-07 | 0.044 | RHO small GTPase |
| Vha55 | 9.428E-07 | 0.04 | hydrogen-transporting ATPase ; EC:3.6.1.35 |
| maps | 1.145E-06 | 0.013 | |
| CG18638 | 1.193E-06 | 0.049 | |
| CG18282 | 1.304E-06 | 0.034 | |
| PyK | 1.416E-06 | 0.013 | pyruvate kinase ; EC:2.7.1.40 |
| CG9172 | 1.511E-06 | 0.03 | enzyme |
| BG | 1.652E-06 | 0.048 | |
| CG1200 | 1.733E-06 | 0.028 | |
| Rh6 | 1.892E-06 | 0.042 | light-sensitive visual pigment |
| mts | 2.901E-06 | 0.043 | protein phosphatase type 2A catalyst |

Table 3.6: Group II: Down-regulated genes in non-mated OH compared to non-mated HK

| Gene Symbol | RP-value | E-value | Function |
|---|---|---|---|
| | | | |
| Dbi | 1.835E-16 | 0 | acyl-CoA or acyl binding |
| CG13779 | 1.966E-15 | 0 | |
| CG11765 | 3.804E-15 | 0 | enzyme |
| CG10652 | 5.189E-14 | 0 | |
| CG8857 | 7.741E-14 | 0 | structural protein of ribosome |
| RpL18A | 9.310E-14 | 0 | large-subunit cytosol ribosomal protein |
| BcDNA:GH05536 | 5.770E-12 | 0 | |
| CG9026 | 6.343E-12 | 0 | enzyme |
| CoVa | 9.225E-12 | 0 | cytochrome-c oxidase ; EC:1.9.3.1 |
| RpS20 | 1.415E-11 | 0 | small-subunit cytosol ribosomal protein |
| RpL9 | 7.681E-11 | 0 | large-subunit cytosol ribosomal protein |
| CG7613 | 1.012E-10 | 0 | enzyme |
| BcDNA:GH10614 | 3.861E-10 | 0.001 | enzyme |
| CG14449 | 4.414E-10 | 0 | |
| RpL32 | 5.733E-10 | 0.001 | large-subunit cytosol ribosomal protein |
| Elf | 7.267E-10 | 0.001 | cytosolic translation release factor |
| MtnA | 7.891E-10 | 0.001 | Cu/Cd binding |
| CG5991 | 1.096E-09 | 0.001 | enzyme |
| CG11218 | 1.375E-09 | 0 | ligand binding or carrier |
| CG18598 | 1.454E-09 | 0 | |
| CG13771 | 1.465E-09 | 0 | |
| Ubi-p63E | 1.891E-09 | 0.013 | protein degradation tagging |
| RpL13 | 2.107E-09 | 0.001 | large-subunit cytosol ribosomal protein |
| CG2862 | 3.542E-09 | 0.004 | enzyme inhibitor |
| CG4572 | 4.883E-09 | 0.001 | peptidase |
| CG7430 | 5.539E-09 | 0 | enzyme |
| RpL1 | 7.457E-09 | 0.002 | large-subunit cytosol ribosomal protein |
| CG5841 | 9.180E-09 | 0 | apoptosis inhibitor |
| RpS4 | 9.891E-09 | 0.002 | small-subunit cytosol ribosomal protein |
| CG8057 | 1.191E-08 | 0.005 | protein kinase |
| RpS17 | 1.724E-08 | 0.003 | small-subunit cytosol ribosomal protein |
| CG6631 | 2.487E-08 | 0.006 | motor |
| Gbp | 2.520E-08 | 0.002 | GTP binding |

(Table 3.6 - continued)

| | | | |
|---|---|---|---|
| RpL14 | 2.680E-08 | 0.009 | large-subunit cytosol ribosomal protein |
| oho23B | 2.833E-08 | 0.006 | ribosomal protein |
| Jra | 2.949E-08 | 0.005 | RNA polymerase II transcription factor |
| smp-30 | 3.003E-08 | 0.001 | |
| CG8343 | 3.876E-08 | 0.008 | ligand binding or carrier |
| CG7738 | 4.119E-08 | 0.002 | |
| CG7630 | 4.139E-08 | 0.005 | |
| CG10363 | 4.475E-08 | 0.011 | |
| CG14320 | 5.007E-08 | 0.007 | |
| dock | 5.170E-08 | 0.009 | SH3/SH2 adaptor protein |
| RpS27A | 5.179E-08 | 0.011 | protein degradation tagging |
| Yp1 | 5.370E-08 | 0.001 | yolk protein |
| CG1883 | 5.817E-08 | 0.02 | structural protein of ribosome |
| CG7360 | 6.261E-08 | 0.003 | receptor |
| RpS25 | 6.364E-08 | 0.005 | small-subunit cytosol ribosomal protein |
| CG9274 | 6.439E-08 | 0.006 | |
| RpS12 | 8.914E-08 | 0.025 | small-subunit cytosol ribosomal protein |
| CG15231 | 9.104E-08 | 0.006 | |
| Elongin-C | 9.736E-08 | 0.008 | transcription factor |
| sta | 9.740E-08 | 0.043 | small-subunit cytosol ribosomal protein |
| Rack1 | 9.931E-08 | 0.021 | protein kinase C binding protein |
| CG9836 | 1.005E-07 | 0.006 | |
| RpS18 | 1.216E-07 | 0.006 | small-subunit cytosol ribosomal protein |
| RpS9 | 1.248E-07 | 0.011 | small-subunit cytosol ribosomal protein |
| CG17280 | 1.357E-07 | 0.011 | enzyme |
| HLHmgamma | 1.455E-07 | 0.004 | transcription factor |
| Pgi | 1.740E-07 | 0.024 | phosphogluconate dehydrogenase (decarboxylating) ; EC:1.1.1.44 |
| RpS13 | 1.926E-07 | 0.019 | small-subunit cytosol ribosomal protein |
| CG14214 | 2.140E-07 | 0.015 | transporter |
| CG15222 | 2.533E-07 | 0.009 | |
| ninaA | 2.555E-07 | 0.024 | cyclophilin |
| eas | 2.940E-07 | 0.016 | ethanolamine kinase ; EC:2.7.1.82 |
| Dlc90F | 3.862E-07 | 0.018 | motor |
| Takr99D | 6.123E-07 | 0.018 | tachykinin receptor |
| CG9904 | 6.430E-07 | 0.026 | |
| CG8332 | 7.923E-07 | 0.037 | structural protein of ribosome |
| CG14975 | 8.146E-07 | 0.036 | |
| CG5081 | 9.584E-07 | 0.024 | transporter |
| CG6867 | 9.624E-07 | 0.047 | structural protein |
| Hnf4 | 9.704E-07 | 0.016 | ligand-dependent nuclear receptor |
| CG4561 | 1.114E-06 | 0.048 | tyrosine--tRNA ligase |
| CG9282 | 1.648E-06 | 0.034 | structural protein of ribosome |
| CG9674 | 1.671E-06 | 0.021 | enzyme |
| CG13255 | 1.720E-06 | 0.037 | |
| CG7275 | 1.815E-06 | 0.026 | signal transduction |
| ATPsyn-b | 1.924E-06 | 0.034 | hydrogen-transporting ATP synthase ; EC:3.6.1.34 |
| CG15235 | 1.962E-06 | 0.036 | signal transduction |
| CG16844 | 2.266E-06 | 0.02 | |
| CG10622 | 2.299E-06 | 0.046 | |
| CG3792 | 2.472E-06 | 0.041 | |

(Table 3.6 - continued)

| | | | |
|---|---|---|---|
| CG9354 | 2.597E-06 | 0.044 | structural protein of ribosome |
| BcDNA:GH12504 | 2.758E-06 | 0.029 | transmembrane receptor |
| GlyP | 2.895E-06 | 0.04 | phosphorylase ; EC:2.4.1.1 |
| mbf1 | 3.284E-06 | 0.042 | |
| CG8505 | 3.632E-06 | 0.032 | structural protein |
| CG7528 | 5.653E-06 | 0.046 | |

Further analysis of the group I genes by GoMiner showed that 30 categories of Biological Process, 7 of Cellular Component, and 4 of Molecular Function were significantly enriched (q-value < 0.05) (Table 3.7). Remarkably, enriched Biological Process categories included those related to visual system, such as response to radiation, eye morphogenesis, and eye photoreceptor cell differentiation. Under the category of Cellular Component, rhabdomere was also shown to be significantly enriched (q-value < 0.05). The genes in those categories include *da, mts, fng, rac1, emc, arr2, rh6,* and *cam.* As for Molecular Function, the group I genes were enriched in nucleoside-triphosphatase activity, hydrolase activity, and monovalent ionic cation transporter activity (q-value < 0.05).

Table 3.7: *D. elegans* between species (OH and HK), enriched GO categories of overexpressed genes in nMOH (Group I)

| Description | Over | Total | p-value | q-value |
|---|---|---|---|---|
| **Biological Process** | | | | |
| cytoskeleton organization and biogenesis | 13 | 353 | 0 | 0.003 |
| organelle organization and biogenesis | 15 | 563 | 0 | 0.008 |
| cell organization and biogenesis | 22 | 1092 | 0 | 0.0083 |
| cell differentiation | 12 | 417 | 0.0001 | 0.0332 |
| main pathways of carbohydrate metabolism | 5 | 82 | 0.0006 | 0.0456 |
| morphogenesis | 12 | 524 | 0.0011 | 0.0459 |
| cellular carbohydrate metabolism | 7 | 222 | 0.0024 | 0.0465 |
| muscle fiber development | 3 | 33 | 0.0026 | 0.0466 |
| skeletal muscle fiber development | 3 | 33 | 0.0026 | 0.0466 |
| skeletal muscle development | 3 | 33 | 0.0026 | 0.0466 |
| eye photoreceptor fate commitment (sensu Endopterygota) | 3 | 32 | 0.0023 | 0.047 |

(Table 3.7 - continued)

| | | | | |
|---|---|---|---|---|
| eye photoreceptor cell fate commitment | 3 | 32 | 0.0023 | 0.047 |
| phosphorylation | 10 | 369 | 0.0009 | 0.0472 |
| eye photoreceptor cell differentiation | 4 | 66 | 0.0022 | 0.0475 |
| carbohydrate metabolism | 9 | 373 | 0.0036 | 0.0476 |
| organismal physiological process | 15 | 690 | 0.0004 | 0.0478 |
| eye morphogenesis (sensu Endopterygota) | 5 | 117 | 0.0029 | 0.0479 |
| muscle contraction | 4 | 75 | 0.0035 | 0.0482 |
| development | 19 | 1180 | 0.0025 | 0.0482 |
| compound eye morphogenesis (sensu Endopterygota) | 5 | 110 | 0.0022 | 0.0483 |
| compound eye development (sensu Endopterygota) | 5 | 110 | 0.0022 | 0.0483 |
| eye-antennal disc morphogenesis | 5 | 122 | 0.0034 | 0.0485 |
| eye morphogenesis | 5 | 122 | 0.0034 | 0.0485 |
| spermatid development | 3 | 29 | 0.0018 | 0.0489 |
| spermatid differentiation | 3 | 29 | 0.0018 | 0.0489 |
| response to radiation | 4 | 45 | 0.0005 | 0.0493 |
| generation of precursor metabolites and energy | 10 | 423 | 0.0024 | 0.0497 |
| energy derivation by oxidation of organic compounds | 5 | 94 | 0.0011 | 0.0497 |
| photoreceptor cell differentiation (sensu Endopterygota) | 4 | 74 | 0.0033 | 0.0499 |
| regulation of cell shape | 4 | 74 | 0.0033 | 0.0499 |
| **Cellular Component** | | | | |
| rhabdomere | 4 | 18 | 0 | 0.005 |
| plasma membrane | 11 | 451 | 0.0011 | 0.0477 |
| cellular_component | 44 | 4014 | 0.0024 | 0.0487 |
| membrane part | 18 | 962 | 0.0006 | 0.0492 |
| intracellular | 33 | 2683 | 0.0034 | 0.0493 |
| cell part | 43 | 3735 | 0.0009 | 0.0498 |
| cell | 43 | 3735 | 0.0009 | 0.0498 |
| **Molecular Function** | | | | |
| nucleoside-triphosphatase activity | 11 | 421 | 0.0006 | 0.0435 |
| hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides | 11 | 442 | 0.0009 | 0.0448 |
| hydrolase activity, acting on acid anhydrides | 11 | 442 | 0.0009 | 0.0448 |
| monovalent inorganic cation transporter activity | 4 | 74 | 0.0033 | 0.0499 |

7 categories of Biological Process, 22 of Cellular Component, and 3 of Molecular Function were shown to be significantly enriched by group II genes (q-value < 0.05) (table 3.8). In contrast to the group I genes, the group II genes were mainly enriched in biosynthesis and metabolism. Notably, categories related to ribosome

constituents were significantly enriched with this group of genes, and 27 genes had molecular function of nucleic acid binding (q-value < 0.05).

Table 3.8: *D. elegans* between species (OH and HK), enriched GO categories of underexpressed genes in OH (Group II)

| Description | Under | Total | p-value | q-value |
|---|---|---|---|---|
| **Biological Process** | | | | |
| protein biosynthesis | 23 | 397 | 0 | 0 |
| cellular biosynthesis | 28 | 641 | 0 | 0 |
| macromolecule biosynthesis | 24 | 420 | 0 | 0 |
| biosynthesis | 29 | 692 | 0 | 0 |
| cellular macromolecule metabolism | 31 | 1707 | 0.0004 | 0.0166 |
| cellular protein metabolism | 30 | 1650 | 0.0005 | 0.0191 |
| protein metabolism | 30 | 1733 | 0.0012 | 0.0415 |
| **Cellular Component** | | | | |
| protein complex | 33 | 1302 | 0 | 0 |
| intracellular non-membrane-bound organelle | 23 | 582 | 0 | 0 |
| non-membrane-bound organelle | 23 | 582 | 0 | 0 |
| small ribosomal subunit | 14 | 62 | 0 | 0 |
| eukaryotic 48S initiation complex | 14 | 38 | 0 | 0 |
| eukaryotic 43S preinitiation complex | 14 | 54 | 0 | 0 |
| ribonucleoprotein complex | 22 | 272 | 0 | 0 |
| cytosolic small ribosomal subunit (sensu Eukaryota) | 14 | 38 | 0 | 0 |
| cytosolic large ribosomal subunit (sensu Eukaryota) | 8 | 46 | 0 | 0 |
| ribosome | 22 | 164 | 0 | 0 |
| cytosolic ribosome (sensu Eukaryota) | 22 | 84 | 0 | 0 |
| cytosol | 23 | 156 | 0 | 0 |
| cytoplasm | 36 | 1201 | 0 | 0 |
| intracellular organelle part | 31 | 1172 | 0 | 0 |
| cytosolic part | 22 | 106 | 0 | 0 |
| cytoplasmic part | 30 | 952 | 0 | 0 |
| organelle part | 31 | 1172 | 0 | 0 |
| large ribosomal subunit | 8 | 87 | 0 | 0.0001 |
| intracellular part | 44 | 2595 | 0 | 0.0011 |
| intracellular | 44 | 2683 | 0.0001 | 0.002 |
| intracellular organelle | 37 | 2193 | 0.0003 | 0.0109 |
| organelle | 37 | 2193 | 0.0003 | 0.0109 |
| **Molecular Function** | | | | |
| structural molecule activity | 26 | 623 | 0 | 0 |
| structural constituent of ribosome | 22 | 163 | 0 | 0 |

(Table 3.8 - continued)

| | | | | |
|---|---|---|---|---|
| nucleic acid binding | 27 | 1424 | 0.0006 | 0.0197 |

## 3.3 Differentially Expressed Gene in OH Female Due to Mating Effect

The group III contained 75 genes and the group IV contained 76 genes (E-value < 0.05) (Table 3.4, 3.9, and 3.10). Notably, even though a few GO categories were reported as enriched with the group III genes (4 categories of Biological Process, 8 of Cellular Component, and 1 of Molecular Function), many GO categories were enriched with the group IV genes (table 3.11 and 3.12). 64 categories of Biological Process were significantly enriched (q-value < 0.05), and these categories are related to the visual system, such as eye photoreceptor cell differentiation, eye-antennal disc development, and R8 cell differentiation, and also neural network system, such as regulation of synaptic structure and function, regulation of neurogenesis, regulation of axonogenesis, and glutamate metabolism. Genes from the visual system-related categories included *emc, h, mts, rac1,* and *rho1,* and genes from the other neuronal categories included *rab5, slob, rac1, rho1, gdh,* and *CG9674.* Enriched categories of Molecular Function were mainly related to hydrolase activity, pyrophosphatase activity, and nucleoside-triphosphatase activity.

Table 3.9: Group III: Up-regulated genes in non-mated OH compared to mated OH

| Gene Symbol | RP-value | E-value | Function |
|---|---|---|---|
| | | | |
| Lsp2 | 8.529E-17 | 0 | larval serum protein |
| CG4347 | 2.226E-16 | 0 | UTP--glucose-1-phosphate uridylyltransferase |
| CG5827 | 2.732E-16 | 0 | structural protein of ribosome |
| CG8505 | 1.098E-15 | 0 | structural protein |
| CG9282 | 2.058E-15 | 0 | structural protein of ribosome |
| BG:DS07851.3 | 9.334E-15 | 0 | |
| CG3534 | 1.681E-12 | 0 | xylulokinase |
| CG2827 | 4.021E-12 | 0 | enzyme |
| CG13779 | 2.037E-11 | 0 | |
| CG3560 | 4.124E-11 | 0 | NADH dehydrogenase |

(Table 3.9 - continued)

| | | | |
|---|---|---|---|
| CG10320 | 2.687E-10 | 0.001 | enzyme |
| CG6770 | 2.872E-10 | 0.001 | |
| Mp20 | 6.022E-10 | 0 | calcium binding |
| CG15201 | 6.903E-10 | 0 | |
| dock | 1.789E-09 | 0 | SH3/SH2 adaptor protein |
| CG9112 | 1.971E-09 | 0.003 | DNA binding |
| CG4019 | 2.044E-09 | 0 | transporter |
| CG8360 | 2.495E-09 | 0 | enzyme |
| CG12030 | 2.860E-09 | 0.001 | enzyme |
| CG13550 | 2.903E-09 | 0.002 | motor |
| CG15199 | 3.076E-09 | 0.002 | |
| CG5642 | 4.013E-09 | 0.003 | |
| CG5995 | 5.500E-09 | 0 | |
| ATPCL | 5.569E-09 | 0.007 | ATP-citrate (pro-S)-lyase ; EC:4.1.3.8 |
| TRAP1 | 6.028E-09 | 0 | chaperone |
| hig | 6.222E-09 | 0.001 | cell adhesion |
| CG5168 | 6.984E-09 | 0.005 | signal transduction |
| CG11901 | 7.223E-09 | 0.003 | |
| CG6188 | 1.013E-08 | 0.001 | |
| CG14214 | 1.062E-08 | 0.001 | transporter |
| CG18606 | 1.695E-08 | 0.008 | |
| Qcr9 | 1.911E-08 | 0.005 | ubiquinol-cytochrome c oxidase ; EC:1.10.2.2 |
| RpL9 | 2.214E-08 | 0.011 | large-subunit cytosol ribosomal protein |
| CG15499 | 2.243E-08 | 0.015 | |
| CG2922 | 2.507E-08 | 0.012 | |
| CG1458 | 3.031E-08 | 0.008 | |
| CG8723 | 3.252E-08 | 0.021 | enzyme |
| CG13328 | 3.431E-08 | 0.006 | |
| CG4962 | 3.467E-08 | 0.002 | |
| RhoBTB | 3.862E-08 | 0.022 | RHO small GTPase |
| Pdh | 5.616E-08 | 0.013 | enzyme |
| eIF1A | 6.612E-08 | 0.004 | |
| CG12079 | 7.913E-08 | 0.009 | enzyme |
| CG12269 | 8.780E-08 | 0.017 | |
| CG5973 | 9.275E-08 | 0.01 | ligand binding or carrier |
| RpL29 | 1.517E-07 | 0.02 | large-subunit cytosol ribosomal protein |
| Ip259 | 1.694E-07 | 0.01 | |
| CG9298 | 1.890E-07 | 0.024 | |
| eas | 1.992E-07 | 0.029 | ethanolamine kinase ; EC:2.7.1.82 |
| CG12131 | 2.424E-07 | 0.04 | translation factor |
| CG9354 | 3.176E-07 | 0.006 | structural protein of ribosome |
| CG9034 | 3.955E-07 | 0.021 | |
| CG5165 | 4.027E-07 | 0.035 | enzyme |
| CG10794 | 4.133E-07 | 0.013 | defense |
| CG8369 | 4.432E-07 | 0.012 | |
| CG7712 | 5.277E-07 | 0.02 | NADH dehydrogenase |
| His4r | 5.499E-07 | 0.034 | DNA binding |
| CG4769 | 5.632E-07 | 0.025 | electron transfer |
| CG12264 | 5.954E-07 | 0.038 | enzyme |
| Mlc1 | 6.243E-07 | 0.018 | muscle motor protein |
| CG15731 | 6.414E-07 | 0.025 | |

(Table 3.9 - continued)

| | | | |
|---|---|---|---|
| CG7834 | 6.509E-07 | 0.013 | electron transfer |
| CG1318 | 6.656E-07 | 0.017 | |
| CG7584 | 9.819E-07 | 0.023 | ligand binding or carrier |
| CG8309 | 1.006E-06 | 0.041 | |
| CG11979 | 1.381E-06 | 0.044 | enzyme |
| Sec61beta | 2.016E-06 | 0.023 | transporter |
| CG13078 | 2.053E-06 | 0.044 | |
| CG1214 | 2.114E-06 | 0.04 | |
| CG17272 | 2.993E-06 | 0.041 | calcium binding |
| CG16994 | 3.115E-06 | 0.037 | |
| Sap-r | 3.206E-06 | 0.031 | |
| CG8707 | 4.238E-06 | 0.045 | signal transduction |

Table 3.10: Group IV: Down-regulated genes in non-mated OH compared to mated OH

| Gene Symbol | RP-value | E-value | Function |
|---|---|---|---|
| | | | |
| Ubi-p63E | 1.114E-22 | 0 | protein degradation tagging |
| fax | 3.429E-15 | 0 | |
| CG12431 | 1.120E-14 | 0 | |
| CG6963 | 5.808E-14 | 0 | |
| Taf110 | 4.069E-13 | 0 | transcription activating factor |
| Cam | 6.539E-13 | 0 | calcium sensing |
| CG9894 | 3.983E-12 | 0 | |
| h | 1.427E-11 | 0 | specific RNA polymerase II transcription factor |
| CG15223 | 2.357E-11 | 0 | |
| Act5C | 2.495E-11 | 0 | motor protein |
| CG8895 | 2.618E-11 | 0 | |
| Sdc | 6.098E-11 | 0 | heparin sulfate proteoglycan |
| CG13943 | 7.807E-11 | 0 | |
| Pu | 2.776E-10 | 0 | GTP cyclohydrolase I ; EC:3.5.4.16 |
| Gdi | 2.889E-10 | 0 | GDP-dissociation inhibitor |
| CG18282 | 5.359E-10 | 0 | |
| Slob | 5.628E-10 | 0.001 | signal transduction |
| CG12004 | 1.016E-09 | 0.003 | |
| Atpalpha | 1.309E-09 | 0 | sodium/potassium-transporting ATPase ; EC:3.6.1.37 |
| CG14844 | 2.741E-09 | 0 | |
| | 3.033E-09 | 0.004 | |
| Pbprp3 | 4.041E-09 | 0.004 | pheromone binding |
| 26/29kD-proteinase | 4.117E-09 | 0 | cysteine-type endopeptidase |
| Ace | 4.483E-09 | 0.002 | acetylcholinesterase ; EC:3.1.1.7 |
| CG9140 | 5.628E-09 | 0.001 | enzyme |
| CG9057 | 5.828E-09 | 0.002 | |
| emc | 5.935E-09 | 0.005 | transcription co-repressor |
| TM4SF | 6.020E-09 | 0.002 | |
| CG14503 | 6.435E-09 | 0 | |
| CG7085 | 6.758E-09 | 0.001 | |
| beta-Spec | 7.422E-09 | 0.005 | actin cross-linking |
| CG18598 | 8.244E-09 | 0.001 | |

(Table 3.10 - continued)

| | | | |
|---|---|---|---|
| CG9318 | 9.806E-09 | 0.005 | |
| CG7181 | 1.221E-08 | 0.007 | |
| CG5889 | 2.413E-08 | 0.003 | |
| CG13666 | 2.421E-08 | 0.003 | |
| CG12324 | 2.542E-08 | 0 | structural protein of ribosome |
| CG15567 | 3.103E-08 | 0.006 | |
| Rac1 | 3.116E-08 | 0.002 | RHO small GTPase |
| BcDNA:GH08860 | 3.502E-08 | 0.012 | enzyme |
| PhKgamma | 3.702E-08 | 0.01 | phosphorylase kinase catalyst |
| sta | 4.867E-08 | 0.025 | small-subunit cytosol ribosomal protein |
| CG14975 | 4.916E-08 | 0.017 | |
| CG13233 | 5.003E-08 | 0.004 | |
| Rab5 | 5.055E-08 | 0.014 | RAB small GTPase |
| CG18502 | 5.453E-08 | 0.003 | |
| Arf79F | 5.506E-08 | 0.002 | GTP binding |
| CG12566 | 6.035E-08 | 0.009 | |
| Elf | 7.756E-08 | 0.006 | cytosolic translation release factor |
| CG15381 | 8.143E-08 | 0.008 | |
| CG15893 | 8.149E-08 | 0.015 | |
| CG9674 | 9.040E-08 | 0.011 | enzyme |
| CG2233 | 9.152E-08 | 0.015 | |
| Pabp2 | 9.305E-08 | 0.004 | poly(A) binding |
| CG13159 | 1.042E-07 | 0.004 | |
| CG7981 | 1.477E-07 | 0.008 | |
| AnnX | 1.541E-07 | 0.012 | calcium-dependent phospholipid binding |
| CG7510 | 1.891E-07 | 0.019 | |
| nAcRalpha-96Ab | 2.073E-07 | 0.027 | nicotinic acetylcholine-gated receptor-channel |
| CG3186 | 2.124E-07 | 0.017 | translation factor |
| mts | 2.289E-07 | 0.012 | protein phosphatase type 2A catalyst |
| Elongin-B | 2.290E-07 | 0.021 | enzyme |
| Gbeta13F | 2.461E-07 | 0.042 | heterotrimeric G protein |
| Nlp | 2.604E-07 | 0.019 | DNA binding |
| CG2768 | 3.956E-07 | 0.018 | transporter |
| | 5.800E-07 | 0.011 | |
| Gdh | 7.052E-07 | 0.038 | glutamate dehydrogenase (NAD(P)+) ; EC:1.4.1.3 |
| Rho1 | 8.498E-07 | 0.025 | RHO small GTPase |
| CG1534 | 8.906E-07 | 0.017 | |
| CG11676 | 1.173E-06 | 0.031 | transcription factor |
| CG8029 | 1.296E-06 | 0.019 | enzyme |
| CG13255 | 1.532E-06 | 0.017 | |
| CG13053 | 2.396E-06 | 0.032 | |
| CG5991 | 2.419E-06 | 0.03 | enzyme |
| Vha55 | 2.645E-06 | 0.032 | hydrogen-transporting ATPase ; EC:3.6.1.35 |
| CG15890 | 4.766E-06 | 0.049 | transporter |

Table 3.11: *D. elegans* mating effect, enriched GO categories of overexpressed genes in non-mated OH (Group III)

| Description | Over | Total | p-value | q-value |
|---|---|---|---|---|
| **Biological Process** | | | | |

(Table 3.11 - continued)

| | | | | |
|---|---|---|---|---|
| ATP synthesis coupled electron transport | 5 | 51 | 0 | 0.003 |
| ATP synthesis coupled electron transport (sensu Eukaryota) | 5 | 50 | 0 | 0.0035 |
| oxidative phosphorylation | 6 | 109 | 0.0001 | 0.0119 |
| SRP-dependent cotranslational protein targeting to membrane, translocation | 2 | 4 | 0.0003 | 0.0389 |
| **Cellular Component** | | | | |
| cytoplasm | 22 | 1201 | 0 | 0.002 |
| cytoplasmic part | 19 | 952 | 0 | 0.0025 |
| mitochondrial electron transport chain | 5 | 59 | 0 | 0.0093 |
| cytosolic large ribosomal subunit (sensu Eukaryota) | 4 | 46 | 0.0003 | 0.0315 |
| intracellular organelle part | 18 | 1172 | 0.0004 | 0.035 |
| organelle part | 18 | 1172 | 0.0004 | 0.035 |
| mitochondrial membrane part | 5 | 93 | 0.0004 | 0.042 |
| translocon complex | 2 | 5 | 0.0005 | 0.0437 |
| **Molecular Function** | | | | |
| electron carrier activity | 5 | 48 | 0 | 0.0047 |

Table 3.12: *D. elegans* mating effect, enriched GO categories of underexpressed genes in non-mated OH (Group IV)

| Description | Under | Total | p-value | q-value |
|---|---|---|---|---|
| **Biological Process** | | | | |
| synaptic transmission | 8 | 196 | 0 | 0.005 |
| regulation of synapse structure and function | 3 | 11 | 0 | 0.0063 |
| cell organization and biogenesis | 19 | 1092 | 0 | 0.007 |
| neurophysiological process | 11 | 470 | 0.0001 | 0.021 |
| transmission of nerve impulse | 9 | 311 | 0.0001 | 0.0237 |
| endocytosis | 5 | 93 | 0.0003 | 0.0275 |
| eye photoreceptor cell differentiation (sensu Endopterygota) | 4 | 64 | 0.0007 | 0.0282 |
| eye photoreceptor cell differentiation | 4 | 66 | 0.0008 | 0.0283 |
| membrane organization and biogenesis | 3 | 27 | 0.0006 | 0.0287 |
| cell differentiation | 9 | 417 | 0.0011 | 0.0289 |
| compound eye morphogenesis (sensu Endopterygota) | 5 | 110 | 0.0006 | 0.0296 |
| compound eye development (sensu Endopterygota) | 5 | 110 | 0.0006 | 0.0296 |
| eye photoreceptor fate commitment (sensu Endopterygota) | 3 | 32 | 0.0011 | 0.0297 |
| eye photoreceptor cell fate commitment | 3 | 32 | 0.0011 | 0.0297 |
| eye-antennal disc morphogenesis | 5 | 122 | 0.001 | 0.0309 |
| eye morphogenesis | 5 | 122 | 0.001 | 0.0309 |
| photoreceptor cell differentiation (sensu Endopterygota) | 4 | 74 | 0.0012 | 0.032 |
| organelle organization and biogenesis | 11 | 563 | 0.0006 | 0.0324 |
| organismal physiological process | 12 | 690 | 0.001 | 0.0324 |

(Table 3.12 - continued)

| | | | | |
|---|---|---|---|---|
| neuromuscular synaptic transmission | 3 | 25 | 0.0005 | 0.0325 |
| cell-cell signaling | 9 | 383 | 0.0006 | 0.0326 |
| eye morphogenesis (sensu Endopterygota) | 5 | 117 | 0.0008 | 0.0332 |
| cellular physiological process | 40 | 4965 | 0.0009 | 0.0334 |
| morphogenesis | 10 | 524 | 0.0014 | 0.0334 |
| restriction of R8 fate | 2 | 6 | 0.0006 | 0.0335 |
| glutamate metabolism | 2 | 6 | 0.0006 | 0.0335 |
| dorsal closure, amnioserosa morphology change | 2 | 6 | 0.0006 | 0.0335 |
| actin cytoskeleton organization and biogenesis | 4 | 82 | 0.0018 | 0.0341 |
| actin filament-based process | 4 | 82 | 0.0018 | 0.0341 |
| photoreceptor cell differentiation | 4 | 81 | 0.0017 | 0.0344 |
| eye-antennal disc development | 5 | 133 | 0.0015 | 0.035 |
| cellular morphogenesis | 7 | 277 | 0.0017 | 0.0351 |
| eye development (sensu Endopterygota) | 5 | 144 | 0.0021 | 0.0352 |
| cytoskeleton organization and biogenesis | 8 | 353 | 0.0016 | 0.0352 |
| photoreceptor fate commitment (sensu Endopterygota) | 3 | 38 | 0.0018 | 0.0354 |
| photoreceptor cell fate commitment | 3 | 38 | 0.0018 | 0.0354 |
| R8 cell fate commitment | 2 | 9 | 0.0014 | 0.0355 |
| eye development | 5 | 149 | 0.0025 | 0.0361 |
| generation of neurons | 5 | 149 | 0.0025 | 0.0361 |
| establishment of wing hair orientation | 2 | 11 | 0.0021 | 0.0369 |
| glial cell migration | 2 | 11 | 0.0021 | 0.0369 |
| cytoplasm organization and biogenesis | 3 | 42 | 0.0024 | 0.037 |
| R8 cell differentiation | 2 | 12 | 0.0026 | 0.038 |
| energy reserve metabolism | 2 | 12 | 0.0026 | 0.038 |
| regulation of axonogenesis | 2 | 12 | 0.0026 | 0.038 |
| system development | 9 | 478 | 0.0028 | 0.039 |
| wing hair organization and biogenesis | 2 | 13 | 0.003 | 0.041 |
| endosome transport | 2 | 13 | 0.003 | 0.041 |
| neurogenesis | 5 | 158 | 0.0032 | 0.0417 |
| axonogenesis | 4 | 101 | 0.0038 | 0.0429 |
| neurite morphogenesis | 4 | 101 | 0.0038 | 0.0429 |
| neuron morphogenesis during differentiation | 4 | 101 | 0.0038 | 0.0429 |
| epidermis morphogenesis | 2 | 14 | 0.0035 | 0.0433 |
| epidermal cell differentiation | 2 | 14 | 0.0035 | 0.0433 |
| trichome organization and biogenesis (sensu Insecta) | 2 | 14 | 0.0035 | 0.0433 |
| hair cell differentiation | 2 | 14 | 0.0035 | 0.0433 |
| vesicle-mediated transport | 6 | 242 | 0.0041 | 0.0457 |
| actin filament organization | 3 | 52 | 0.0044 | 0.0458 |
| regulation of neurogenesis | 2 | 15 | 0.004 | 0.0461 |
| glutamine family amino acid metabolism | 2 | 16 | 0.0046 | 0.0482 |
| development | 15 | 1180 | 0.0048 | 0.0484 |
| oxidative phosphorylation | 4 | 109 | 0.005 | 0.0486 |
| generation of precursor metabolites and energy | 8 | 423 | 0.0048 | 0.0489 |
| secretory pathway | 5 | 179 | 0.0054 | 0.0497 |
| **Cellular Component** | | | | |

(Table 3.12 - continued)

| | | | | |
|---|---|---|---|---|
| plasma membrane | 11 | 451 | 0.0001 | 0.0154 |
| plasma membrane part | 8 | 278 | 0.0003 | 0.0266 |
| cytoplasm | 16 | 1201 | 0.0021 | 0.0352 |
| membrane part | 13 | 962 | 0.0056 | 0.0494 |
| **Molecular Function** | | | | |
| GTPase activity | 6 | 108 | 0.0001 | 0.0165 |
| hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides | 10 | 442 | 0.0004 | 0.0242 |
| hydrolase activity, acting on acid anhydrides | 10 | 442 | 0.0004 | 0.0242 |
| oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor | 2 | 3 | 0.0001 | 0.0275 |
| pyrophosphatase activity | 10 | 427 | 0.0003 | 0.0298 |
| nucleoside-triphosphatase activity | 10 | 421 | 0.0002 | 0.0317 |
| monovalent inorganic cation transporter activity | 4 | 74 | 0.0012 | 0.032 |
| GTP binding | 5 | 137 | 0.0017 | 0.0336 |
| guanyl nucleotide binding | 5 | 138 | 0.0018 | 0.036 |
| oxidoreductase activity, acting on the CH-NH2 group of donors | 2 | 10 | 0.0018 | 0.0367 |
| protein binding | 14 | 984 | 0.0024 | 0.0374 |
| transcription corepressor activity | 2 | 14 | 0.0035 | 0.0433 |
| ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism | 3 | 51 | 0.0041 | 0.0454 |

## 3.4 Overlap of Candidate Gene Among the Group I, II, III, and IV

5, 36, 8, and 8 genes overlapped between the group I and the group III, I and IV, II and III, and II and IV, respectively. To determine the biological significance of these overlapping genes, we checked the shared GO categories. Interestingly, GO categories, which were showed to be enriched by both groups I and IV (table 3.13) still included the biological processes related to the visual system, such as eye photoreceptor cell differentiation and eye-antennal disc morphogenesis. These shared genes were also enriched in Molecular Function categories of nucleoside-triphosphatase activity, hydrolase activity, and monovalent inorganic cation transporter activity. The shared genes between other gene groups were not enriched in any interesting GO categories (table 3.14).

Table 3.13: Genes overlapping between Group I and Group IV

| Gene ID | Gene Symbol | Function | Map |
|---|---|---|---|
| 151557_i_at | maps | | |
| 143361_at | beta-Spec | actin cross-linking | 16C1-16C1 |
| 151193_f_at | CG15381 | | 22C1-22C1 |
| 153703_at | emc | transcription co-repressor | 61D1-61D2 |
| 141214_at | Gdi | GDP-dissociation inhibitor | 30B8-30B9 |
| 143058_f_at | Act5C | motor protein | 5C5-5C5 |
| 148396_f_at | CG13666 | | 66C8-66C8 |
| 154623_at | CG15893 | | 5E4-5E4 |
| 151747_s_at | Taf110 | transcription activating factor | 72D6-72D6 |
| 154331_at | Gbeta13F | heterotrimeric G protein | 13F5-13F5 |
| 153560_at | CG1534 | | 9E2-9E2 |
| 155059_at | CG9318 | | 38E3-38E4 |
| 154428_at | CG6963 | | 89B17-89B19 |
| 151244_f_at | CG12431 | | 47C7-47C7 |
| 143597_at | Rac1 | RHO small GTPase | 61F5-61F5 |
| 142520_at | Atpalpha | sodium/potassium-transporting ATPase ; EC:3.6.1.37 | 93B1-93B1 |
| 151213_f_at | CG14503 | | 55C5-55C5 |
| 151048_f_at | CG14844 | | 88C3-88C3 |
| 145098_at | CG9057 | | 13A11-13A11 |
| 153465_at | Cam | calcium sensing | 48F5-48F5 |
| 142892_at | CG8895 | | 25B10-25C1 |
| 151224_f_at | CG13943 | | 50F4-50F4 |
| 154421_at | fax | | 72F4-72F4 |
| 145664_at | CG9894 | | 23B1-23B1 |
| 153525_at | CG7510 | | 74E2-74E2 |
| 142723_at | Arf79F | GTP binding | 79F5-79F5 |
| 153619_at | Sdc | heparin sulfate proteoglycan | 57D12-57E5 |
| 150941_f_at | CG15567 | | 100C3-100C3 |
| 154812_at | Pu | GTP cyclohydrolase I ; EC:3.5.4.16 | 57C5-57C6 |
| 151241_f_at | CG13233 | | 47D1-47D1 |
| 143566_i_at | Pabp2 | poly(A) binding | 44A7-44A8 |
| 154218_at | Gdh | glutamate dehydrogenase (NAD(P)+) ; EC:1.4.1.3 \| inferred from direct assay | 95C12-95C13 |
| 152974_at | Ace | acetylcholinesterase ; EC:3.1.1.7 | 87E3-87E3 |
| 153041_at | Vha55 | hydrogen-transporting ATPase ; EC:3.6.1.35 | 87C5-87C5 |
| 144587_f_at | CG18282 | | 5E1-5E1 |
| 154353_at | mts | protein phosphatase type 2A catalyst | 28D2-28D2 |

Table 3.14: GO categories shown to be enriched by both enriched GO categories of Group I and those of Group IV

| Description |
|---|
| **Biological Process** |
| cytoskeleton organization and biogenesis |

(Table 3.14 - continued)

organelle organization and biogenesis
cell organization and biogenesis
cell differentiation
morphogenesis
eye photoreceptor fate commitment (sensu Endopterygota)
eye photoreceptor cell fate commitment
eye photoreceptor cell differentiation
organismal physiological process
eye morphogenesis (sensu Endopterygota)
development
compound eye morphogenesis (sensu Endopterygota)
compound eye development (sensu Endopterygota)
eye-antennal disc morphogenesis
eye morphogenesis
generation of precursor metabolites and energy
photoreceptor cell differentiation (sensu Endopterygota)

**Cellular Component**

plasma membrane
membrane part

**Molecular Function**

nucleoside-triphosphatase activity
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
hydrolase activity, acting on acid anhydrides
monovalent inorganic cation transporter activity

# CHAPTER 4

# DISCUSSION

## 4.1 Does HK Female Also Show Mating Preference?

All of the previous works on *D. elegans* done by other research groups reported that black morph females of *D. elegans* have preference to select their mates and brown morph females of *D. elegans* do not (Hirai and Kimura 1997; Ishii et al. 2001; Ishii et al. 2002; Kopp and True 2002). Our result was consistent with theirs in terms of the black morph female preference, but, on the other hand, it suggested that HK females, one of the brown morphs in *D. elegans*, have slight mating preference to OH males (Table 3.2 and 3.3 and Figure 4.1). However, this result is most likely because of the few sample size of HK group and an outlier of the data. Among 5 replicates in the HK group, only replicate 4 showed significant deviation from random mating ($G=12.695$, $df=1$, $p<0.001$) while others do not ($0.015<G<1.616$, $df=1$, $p>0.05$). This replicate makes total G-value significant. Even though it is still possible that HK females preferably mate with OH males, it is safe to conclude that HK females do not show sexual isolation. As for OH females, there is not the shadow of a doubt that OH females show the strong preference to select their mate ($G=107.264$, $df=5$, $p<0.001$). Figure 4.1 illustrates this mating preference by OH and HK morphs of *D. elegans*.

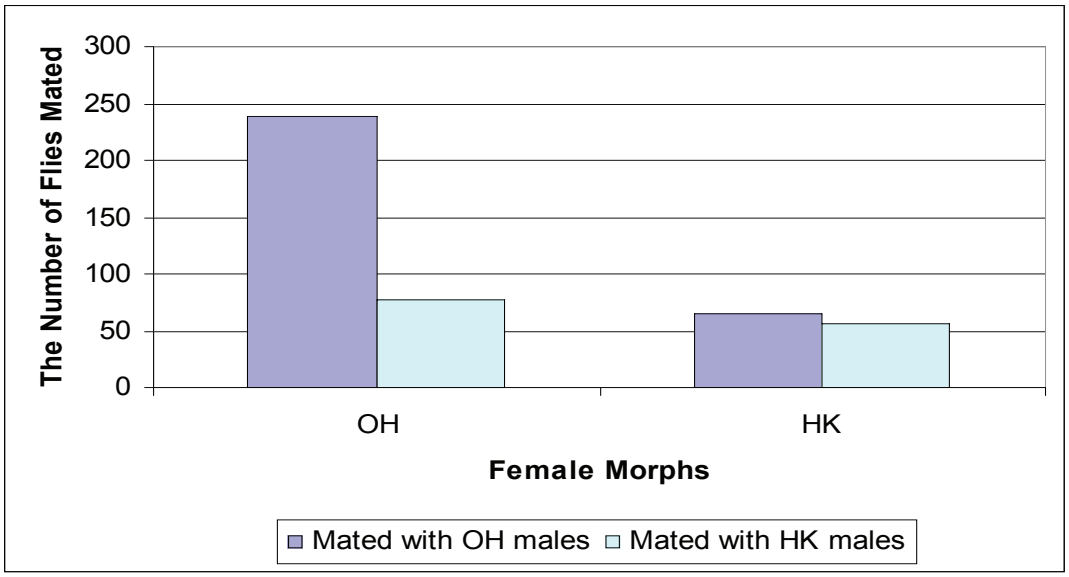



Figure 4.1: Multiple Female Choice Mating Reveals That OH Females Have a Strong Preference to Mate With Own Males

## 4.2 Sequence Divergence Between *D. elegans* and *D. melanogaster*

Since we hybridized RNA from *D. elegans* on the *D. melanogaster* microarray chips, we expected that our analysis contained sequence divergence problems. To overcome this problem, we used a filtering procedure to exclude poorly hybridized probes so that we minimized the effect of sequence divergence. Moreover, both *D. elegans* morphs are expected to have been equally diverged from *D. melanogaster* and this should effectively result in canceling out the sequence divergence effects. In order to estimate when *D. elegans* subgroup was branched out from *D. melanogaster* subgroup, we calculated synonymous substitution rate (Ks) of the coding region of *yellow* gene among 6 different *Drosophila* species, including *D. melanogaster, D. elegans, D. sechellia, D. yakuba, D. erecta,* and *D. ananassae*. We used *yellow* gene since it is the only gene whose mRNA has been completely sequenced in *D. elegans*.

Coding regions of *yellow* gene of these 6 species were obtained from the GenBank on the National Center for Biotechnology Information (NCBI) WebSite. We estimated Ks by Kumar method with MEGA 4 and the result showed that there were about 72% differences between *D. elegans* and *D. melanogaster* (Table 4.1).

Table 4.1: Estimation of Ks by Kumar method. Calculation was done by MEGA 4

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1: D. elegans | | | | | | |
| 2: D. melanogaster | 0.716 | | | | | |
| 3: D. sechellia | 0.708 | 0.053 | | | | |
| 4: D. yakuba | 0.714 | 0.199 | 0.179 | | | |
| 5: D. erecta | 0.705 | 0.182 | 0.168 | 0.172 | | |
| 6: D. ananassae | 1.422 | 1.526 | 1.581 | 1.243 | 1.496 | |

Table 4.1 also suggested that *D. elegans* subgroup was split from *D. melanogaster* subgroup after *D. ananassae*, which is consistent with other researches (Schawaroch 2002; Tamura et al. 2004; Kopp 2006). If we calculate a divergence time at a rate of 1 % synonymous substitution per million year, which is a standard estimated rate, *D. elegans* was branched out from *D. melanogaster* about 36 million years ago (MYA). This estimated divergence time is very close to the divergence distance between *D. melanogaster* subgroup and *D. takahashii* subgroup, which is 35.6 ± 8.7 MYA (Tamura et al. 2004). Since we just used one gene, *yellow,* for sequence divergence time between *D. elegans* and *D. melanogaster*, it is not enough to conclude the divergent time between these two species. However, this estimation provided us an idea how much these two species are diverged, giving us no big surprise by obtaining only 605 probes after filtering process for microarray analysis.

4.3 Some OH Strain-Unique Candidate Gene Are Involved in the Visual System

Since OH females, unlike HK females, preferably mate with homotypic males, if we can identify genes differentially expressed in OH females relative to HK females, those genes could be very good candidates for this phenomenon. Ishii et al. (2001) showed that this behavioral isolation is caused by different kinds of cuticular hydrocarbon concentration, not by their body color (Ishii et al. 2001). However, they did not show which genes are involved in recognizing these sexual signals. Our analysis showed that a total of 182 genes were differentially expressed between OH and HK female heads, of which 91 genes were unique to OH (the group I genes), and the others were unique to HK (the group II genes). The genes unique to OH are most likely involved in sexual discrimination. In order to see the biological significance of these OH unique genes, we used GoMiner to find enriched GO categories. As a result, very interesting GO categories were found to be enriched by OH unique genes. Even though Ishii *et al.* reported that visual cues were not essential for *D. elegans* mate discrimination, our result showed that some of our candidate genes (*da, emc, fng, mts, rac1, arr2, cam, rh6,* and *g-s60αa*) were enriched in the visual system categories, which may have two possible explanations. First, this could be due to our experimental design. Since we took female out from the dark environment right before mating experiment, this could have affected the result. However, if this was true, those visual system relating genes should have expressed to relatively same amount in both strains (OH and HK). Besides, although our previous paper performed same method for mating experiments, none of these genes were shown to be differentially expressed (Michalak

et al. 2007). The other explanation is that these genes are possibly involved in female courtship behavior. It is known that locomotion of female slows down when she is more receptive to male. This transition is presumed as the consequence of detecting some sexual cues, such as male pheromones or courtship songs. After female stops locomotion, male attempts further courtship behavior, such as circulating around female (Greenspan and Ferveur 2000). In *D. elegans*, it's possible that OH females select their mates via visual cues at either of these courtship steps.

## 4.4 Genes Involved in the Visual System Are Still Up-Regulated in Mated OH Female

Even though the significant number of the genes involved in the visual system was more expressed in nMOH compared to nMHK, some of them were even higher expressed in MOH, such as *emc, mts,* and *rac1*. Additionally, *h* and *rho1* appeared to be highly expressed in MOH over nMOH. Even though the other genes relating to the visual system, *da* and *fng*, were not significantly over-expressed in MOH, because of these two new genes, GO analysis showed that those categories are still significantly enriched. Although this phenomenon is still interesting, the visual system relating genes which turned off after mating are more interesting, because those genes could be good candidates for mate discrimination factors in *D. elegans*, for it is known that female decreases their receptivity towards male after mating (Tompkins and Hall 1983; Ejima et al. 2001). If *D. elegans* females use visual recognition as sexual cues a number of the

vision-related genes should be turned off after mating so that they do not consume any energy to select their mate.

## 4.5 OH Unique Genes Are Significantly Enriched in Biochemical Process

Cognition is a very complicated system, so it is unlikely that a single gene explains whole behavior difference. However, interestingly, a group of genes which are involved in biochemical process were shown to be unique to OH female, and these GO categories are hydrolase activity (*arf79F, atpalpha, blw, CG8801, eif-4a, Gβ13f, g-s60αa, rab1, rac1, robl,* and *vha55*), cation transporter activity (*atpalpha, blw, CG10664,* and *vha55*), and phosphorylation (*argk, blw, cam, CG10664, CG12203, CG9172, for, pdk, pyk,* and *vha55*). This result suggests that recognition systems of OH female and HK female are significantly different from each other, and it is very likely that some of these genes are involved in assortative mating. In contrast to OH gene profile, HK genes are more enriched in structural and organelle constituents, such as involved in ribosome synthesis.

## 4.6 Transcriptional Changes Associated With Mating

Significant female behavior changes, such as less receptivity toward males or decrease in female sexual attractiveness after mating have been observed (Fuyama and Ueyama 1997; Tram and Wolfner 1998; Ejima et al. 2001). These female behavior

changes are due to sperm and seminal fluid injection and mating behavior itself, which possibly regulate transcriptional changes. By comparing the transcriptional expression between pre-mating and after-mating female, we can identify genes associated with these behavioral changes. Unfortunately, the genes we obtained from the analysis with *D. elegans* did not overlap with any of the genes in the list from our previous work in which we used *D. melanogaster* Zimbabwe and Cosmopolitan strains (Michalak et al. 2007). McGraw *et al.* and Lawniczak and Begun performed similar whole-genome wide analysis to identify mating associating genes (McGraw et al. 2004; Lawniczak and Begun 2004). Even though our gene list did not match any candidate genes from Lawniczak and Begun, there were some overlapping with McGraw's candidate genes (table 4.2). Interestingly, some of these overlapping genes were regulated in opposite directions, so those genes could be specific to each species.

Table 4.2: Genes overlapping with McGraw's candidate genes

| Gene Symbol | Function | Group |
|---|---|---|
| **Genes Modulated by Sperm** | | |
| CG8505 | structural protein | III |
| CG3560 * | NADH dehydrogenase | III |
| Mp20 | calcium binding | III |
| dock | SH3/SH2 adaptor protein | III |
| CG6188 | | III |
| Qcr9 * | ubiquinol-cytochrome c oxidase ; EC:1.10.2.2 | III |
| CG13328 * | | III |
| CG12079 * | enzyme | III |
| eas | ethanolamine kinase ; EC:2.7.1.82 | III |
| CG5165 * | enzyme | III |
| CG10794 | defense | III |
| CG7712 * | NADH dehydrogenase | III |
| CG4769 * | electron transfer | III |
| Mlc1 * | muscle motor protein | III |
| CG1318 * | | III |
| Sec61beta | transporter | III |
| CG13078 | | III |

(Table 4.2 - continued)

| | | |
|---|---|---|
| Pu ? | GTP cyclohydrolase I ; EC:3.5.4.16 | IV |
| Pbprp3 | pheromone binding | IV |
| CG12566 * | | IV |
| CG9674 | enzyme | IV |
| Pabp2 * | poly(A) binding | IV |
| CG13255 | | IV |
| CG13053 * | | IV |
| **Genes Regulated by Acps** | | |
| Lsp2 * | larval serum protein | III |
| CG3534 * | xylulokinase | III |
| CG2827 * | enzyme | III |
| ATPCL * | ATP-citrate (pro-S)-lyase ; EC:4.1.3.8 | III |
| CG8723 * | enzyme | III |
| **Gene Regulated by Non-Sperm/Non-Acps Components of Mating** | | |
| CG4347 * | UTP--glucose-1-phosphate uridylyltransferase | III |
| CG12030 * | enzyme | III |
| 142042_at | | III |
| CG7584 * | ligand binding or carrier | III |
| h * | specific RNA polymerase II transcription factor | IV |
| beta-Spec | actin cross-linking | IV |
| CG5889 | | IV |
| PhKgamma | phosphorylase kinase catalyst | IV |
| CG7981 * | | IV |

* Same direction of up and down regulation.

Among the overlapping genes, *dock* and *pbprp3* are quite interesting. *Dreadlocks (dock)* has been identified as a component of a signaling cascade involved in photoreceptor axon guidance, which composed of three Src homology 3 (SH3) domain and single SH2 domain (Clemens et al.1996; Garrity et al. 1996). Recent works reveal that its adapter protein interacted with Pak, a serine/threonine kinase, to form a signaling pathway that mediates the response of olfactory axon (Ang et al. 2002). Therefore, *dock* is involved in both visual and olfactory system. This gene turns off after mating in *D. elegans* (Group III), suggesting that it possibly connects to sexual isolation behavior. The other gene, *pbprp3* is more fascinating. This gene is named as

pheromone binding protein (PBP)-relating proteins since it is related to moth PBPs and odorant binding proteins (OBPs), and it is expressed in the main olfactory organ of the antennae (Pikielny et al. 1994). *Pbprp3*, also called as *Olfactory Specific-F* (*OS-F*), is highly conserved genes among 4 species of *Drosophila* and have an annotated duplicated gene, *OS-E*, which is very diverged (Sánchez-Gracia et al. 2003). The exact function of OBPs is not well understood yet, but many recent researches have shown that OBPs play an important role in odorant recognition. Besides, OBPs might be involved in inactivating odorants and other volatile molecules (Park et al. 2000; Hekmat-Scafe et al. 2002). Since this gene is more expressed after mating (Group IV), it is possibly used to inhibit an olfactory signal cascade to reduce receptivity of females towards male.

## 4.7 Biochemical Process Profile in Neural Circuit in *D. elegans* Is Modified by Mating

GO analysis of the genes turned on after mating (the group IV genes) revealed that female behavioral changes were associated with neural network changes, for many of the neural network relating genes were more expressed in MOH, such as *ace,arf79f, β-spec, cam, fax, gdi, nacrα-96b, pbprp3, rab5, rac1,* and *slob*. These genes are involved in neurophysiological processes, such as synaptic transmission, transmission of nerve impulse, generation of neurons, and axonogenesis. Especially, *fax* and *cam* are reported to affect olfactory behavior (Anholt et al. 2003), so these genes are good candidates to explain female behavior change after mating. Moreover, these two genes

are more expressed in OH female compared to HK female (Group I), suggesting us that these could related to OH female choosiness in terms of olfactory.

## 4.8 Problems Need to Be Addressed

So far, our analysis showed many interesting results, there are some problems that we need to discuss. First, in GoMiner's result, there are some categories, shown to be enriched by our candidate genes, which seem completely irrelevant such as spermatid differentiation. This is because of categorization of GO. Since most of genes are assigned to some different biological processes, cellular components, and molecular functions based on GO, this GO categorization might have given us noisy data. Second, for our analysis, we just used the P-call probes, excluding all other probes. These P-call probes are representing the genes whose sequences have been highly conserved, probably around 36 million years, between *D. elegans* and *D. melanogaster* in their evolutional history. By filtering out M-call and A-call probes from microarray, we have definitely missed many important genes, which could possibly participate in major role of sexual discrimination in *D. elegans* system. However, it is still quite interesting to see that each morph of *D. elegans* has retained dramatically different portion of genome with *D. melanogaster*.

## 4.9 Conclusion and Future Direction

Our analysis showed that OH strains, black morph, of *D. elegans* might use visual cues as their sexual isolation. In order to confirm this hypothesis, we will perform choice-by-female test under dark conditions. We might be able to perform that experiment under infrared conditions, but since *D. elegans*' habitat is over flowers, they might be influenced by infrared.

Incipient speciation is one of the big mysteries in evolutionary biology. Unveiling the mechanisms of sexual isolation, one of the pre-zygotic isolation, would lead us to answering fundamental questions of speciation. In our study, we showed some candidate genes are possibly involved in sexual discrimination behavior, especially visual and olfactory relating genes. Further characterization of these candidate genes will be necessary to confirm their involvement in sexual isolation.

# APPENDIX A

## A PUBLISHED PREVIOUS RESEARCH

# GENE EXPRESSION POLYMORPHISM IN *DROSOPHILA* POPULATIONS

**Pawel Michalak[1], John H. Malone, Ivan Timothy Lee, Daiju Hoshino & Daina Ma**

Address:

DEPARTMENT OF BIOLOGY
UNIVERSITY OF TEXAS AT ARLINGTON

[1] Corresponding author:
Box 19498
Arlington, TX 76019-0498
PHONE: 817-272-2979
FAX 817-272-2855
EMAIL: michalak@uta.edu

Key words:

Gene expression profiles, mating discrimination, behavioral isolation, *desaturase2*

Running title: GENE EXPRESSION POLYMORPHISM IN *DROSOPHILA*

**ABSTRACT**

Changes in gene expression have long been recognized as critical to evolutionary processes. However, the extent of natural polymorphism in gene expression has yet to be assessed, which currently opens a new area of active research. We present microarray and quantitative real-time PCR data from Cosmopolitan and Zimbabwe morphs of *Drosophila melanogaster*. These morphs provide a useful model for investigations into the incipient stages of speciation because Zimbabwe females tend to preferentially mate with their own males and discriminate against Cosmopolitan males, while Cosmopolitan females mate indiscriminately. We analyzed expression profiles from heads of mated and nonmated females and identified 45 candidate genes whose expression levels were associated with the behavioral morphs and were modified by mating. Genes with altered transcription levels were randomly distributed across the genome and fell into diverse categories of biological activities. Several candidate genes, such as *desaturase2* and *Odorant receptor 63a*, were additionally subjected to quantitative RT-PCR analysis. Notably, *desaturase2*, which has been invoked to play a role in sexual isolation between Cosmopolitan and Zimbabwe and predicted to be translational-inactive in Cosmopolitan due to a major deletion, was found to be upregulated in Zimbabwe and downregulated, but still expressed, in Cosmopolitan.

## INTRODUCTION

Rapid differentiation and speciation are often associated with striking behavioral and morphological changes and at the same time limited mutations in functional genes. To explain this pattern, changes in regulatory regions have been proposed as a major source of biological differences between species, potentially also contributing to speciation (King and Wilson 1975). Identifying changes in gene regulation during the earliest stages of speciation therefore becomes an important direction for understating how new species form.

Much evidence suggests that behavioral isolation is the first step in the formation of new species. First, behavioral isolation is consistently higher than postzygotic isolation among sympatric but not among allopatric species (Coyne and Orr 1989, 1997; Sasa *et al.* 1998). Second, reproductive isolation breaks down in the absence of behavioral isolation (Seehausen *et al*. 1997) and finally, comparative studies in insects and birds imply a positive correlation between indices of sexual selection and species richness (reviewed in Coyne and Orr 2004). Because of the importance of behavioral isolation to the formation of new species, it is then natural to explore variation in gene expression among populations that show signs of behavioral isolation.

In order to achieve this goal, it is important to distinguish expression variation related to genetic variation within a species (i.e. polymorphism), genetic variation between species (i.e. endogenous factors) and variation in expression that may occur due to the complex behaviors that isolate species (exogenous factors like sexual stimuli). A mechanistic link between altered gene expression and behavioral changes has been well established in a wide range of taxa at all three levels (Bucan and Abel 2002; Rankin

2002; Robinson *et al*. 2005). At the polymorphism level, DNA microarray studies have established that natural genetic variation produces significant differences in gene expression within a species (Cowles *et al*. 2002; Oleksiak *et al*. 2002; Rifkin *et al*. 2003; Cheung *et al*. 2003; Nuzhdin *et al*. 2004; Ranz and Machado 2006; Whitehead and Crawford 2006). At the endogenous level, novel social behavior in closely related whiptail lizards is associated with modified gene expression of a steroid hormone receptor in the brain (Woolley *et al*. 2004). Genome-wide expression profiling of relative transcript abundance in the brain indicated differences between domestic dogs and wolves (Saetre *et al*. 2004), as well as between humans and chimpanzees (Enard *et al*. 2002; Cáceres *et al*. 2003; Gu and Gu 2003; Preuss *et al*. 2004; Uddin *et al*. 2004). At the exogenous level, various labor occupations in the honeybee, such as nursing and foraging, are associated with massive changes in brain transcript levels (Whitfield *et al*. 2003a, 2003b; Grozinger *et al*. 2003; Goodman 2004; Cash *et al*. 2005). Transcription activation is crucial in mediating various forms of mate recognition as exemplified by changes in gene expression and neural circuits in the forebrain of zebra finches presented with conspecific and heterospecific birdsongs (Mello *et al*. 1992; Jarvis *et al*. 2002; Clayton 2004). Sexual experience was observed to dramatically affect expression profiles in the brain of female hamsters as well (Bradley *et al*. 2005). A microarray comparison between mated and nonmated females and between courted (exposed to males, but nonmated) females and unexposed females in *Drosophila. melanogaster* detected dozens of differentially expressed genes (Lawniczak and Begun 2004).

Substantial variation in gene expression among natural populations, between biological species, and the clear link between sexual behavior and transcriptional

alterations, allow us to hypothesize that gene expression changes may also underlie behavioral isolation. Two morphs of *Drosophila melanogaster* (Cosmopolitan and Zimbabwe) offer a model system to address variation in gene expression as it relates to behavioral isolation. Behavioral isolation in *Drosophila* is exerted through courtship behavior, consisting of sequential actions that exchange auditory, visual, and chemosensory signals between males and females (Greenspan and Ferveur 2000). *Drosophila melanogaster* Cosmopolitan and Zimbabwe morphs provide an example of incipient speciation (Wu *et al*. 1995; Hollocher *et al*. 1997a,b; Ting *et al*. 2001) because while there is little deviation from random mating between individuals from various Cosmopolitan populations (Henderson and Lambert 1982, but see Capy *et al*. 2000 and Korol *et al*. 2000), females from some Zimbabwean populations (Sengwa Wildlife and Harare) preferentially mate with their own males and discriminate against Cosmopolitan males. Subsequent experiments with Zimbabwean flies found considerable polymorphism in mating behavior, ranging from pure M-type ("Cosmopolitan"-like) to pure Z-type ("Zimbabwe") and mapped the trait to several genomic regions with the largest impact on the third chromosome (Hollocher *et al*. 1997a,b; Ting *et al*. 2001). Therefore, we hypothesized that fixed differences in gene expression as well as population specific patterns of gene expression due to external stimuli (i.e. exposure to mates) cause behavioral differences that are evolutionarily important.

We used DNA microarrays to identify genes differentially expressed between Cosmopolitan and Zimbabwe populations of *Drosophila melanogaster* and simultaneously, differentially expressed due to mating behavior. This transcriptomics-based approach provides a powerful entrée towards discovery and functional

characterization of genes responsible for incipient stages of speciation. Thus far, there have been no such genes identified despite the fact that estimates of their numbers were obtained from genetic mapping and there have been a number of genes identified that are responsible for hybrid defects (reviewed in Orr *et al*. 2004 and Michalak and Noor 2006). Comparing expression profiles from the heads of mated and nonmated females, we identified a number of genes likely associated with mating (and postmating) effects.

## MATERIALS AND METHODS

### *Flies*

Stocks were reared in uncrowded cultures at 24°C with a 12-h light-dark cycle on Carolina 4-24® formula. Three stocks of *Drosophila melanogaster* Cosmopolitan were used as biological replicates: E10019 from Fukui, Japan (Ehime Drosophila Stocks), EC175 from Ecuador (courtesy of Esther Betrán), and 14021-0231.23 from Crete (Tucson Stock Center, TSC). Four stocks of *Drosophila melanogaster* Zimbabwe were used (courtesy of Jerry Coyne and Tami Panhuis): Z29, Z30, Z49, and Z53.

### *Experimental mating*

Although the pattern of asymmetric isolation between M and Z types has been found highly repeatable by independent researchers (e.g., Wu *et al*. 1995; Hollocher *et al*. 1997a-b; Alipaz *et al*. 2001; Ting, Takahashi, and Wu 2001; Fang *et al*. 2003; Panhuis *et al*. 2003; Takahashi and Ting 2004), we conducted 43 female multiple choice mating experiments to confirm that it persists in our laboratory. During each mating experiment, n (>20) 7-day old females from the same stock were given a choice between n conspecific and n heterotypic males. Z and M males were marked by a dot on their scutella with red and blue colored sharpie markers. Colors were alternated between

experiments but they were found to have no effect on mating frequencies. Replicates were aggregated together within each female type and organized into contingency tables.

*RNA collection*

For RNA collection, virgin females were collected, separated, and aged for 7 days. On day 7, half of them were transferred to bottles with heterotypic males for mating and the other half used as nonmated controls. Copulating pairs were aspirated into individual vials and then males were removed after 30 min and simultaneously frozen with females along with corresponding nonmated control females for comparison. This scheme ensured that female age and time of exposure to treatments were identical in both groups. RNA was extracted from heads only, after 20 females per sample were collected. At least five micrograms of total RNA were extracted using RNApure and following the manufacture's protocol (GenHunter Corp., Nashville, TN). After purification with GeneChip Sample Cleanup Modules (QIAGEN), concentrations were measured with the NanoDrop® ND-1000 UV-Vis and checked for integrity via denaturing gel electrophoresis following the protocol of Liang and Pardee (1997).

*Affymetrix* Drosophila *GeneChip® profiles*

A total of 13 Affymetrix Drosophila GeneChip® arrays (9 for Z and 4 for M females) were used. Double-stranded cDNA was synthesized with a T7-(dT)24 primer; cRNA was synthesized and biotin-labeled in an in vitro transcription reaction using the ENZO BioArray HighYield RNA Transcript Labeling Kit. The target cRNA was hybridized to Affymetrix GeneChip® Drosophila Genome Arrays that allow assays of the relative abundance of >13,500 mRNA transcripts. Probes for a particular transcript are present in 28 cells on the array chip. Half of those cells possess Perfect Match (PM)

sequences that are complementary to the reference transcript. The other half possesses Mismatch (MM) sequences that are complementary to the reference transcript except for a homomeric base mismatch at the 13$^{th}$ position. The number of instances in which the PM hybridization signal is larger than the MM signal is computed along with the average of the logarithm of the PM:MM ratio (after background subtraction) for each probe set using Affymetrix Microarray Suite (MAS) 5.0 software. The software Detection algorithm calculates the Discrimination Score from each probe pair (PM vs. MM), assesses probe saturation, calculates a Detection P value from a one-sided Wilcoxon Signed Rank test, and assigns a Present, Marginal, or Absent Call in relation to a threshold value Tau=0.015.

*Gene expression analysis*

Prior to the statistical analysis, MAS 5.0 data were scaled to 10,000 and $\log_2$-transformed. Additionally, Affymetrix CEL files were reanalyzed with DNA-Chip Analyzer (dChip; Li and Wong 2003). dChip is a software package that implements a statistical model-based expression analysis to account for probe variability and enables both low-level and high-level analyses (see Li and Wong 2003 for details). Expression values were generated from PM-MM using Li-Wong transformation and normalization (Li and Wong 2001a, 2001b). dChip normalizes all arrays to a common baseline array having the median overall brightness (Li and Wong 2003). Normalization is based on an invariable set of probes identified through iterative procedures. The CEL images were visually checked for local contamination and data were additionally filtered to include only those 3237 transcripts that had more than 50% present calls among all samples.

ANOVA contrasts implemented in dChip were used to find the intersection of two candidate gene lists, one differentiating all Z and M and the other differentiating mated and nonmated Z females. Candidate transcripts from the intersection represent mating-related genes that significantly differ between the Z and M morphs. They may include genes involved in sexual discrimination but transcripts underlying postmating effects will likely be included as well. We focused on contrasting mated and nonmated Z females, as we reasoned that comparing females from the discriminating group would be more informative than those from the nondiscriminating group (M). Although expression profiles of M females may also be changed due to their mating status, this change will not be associated with mating discrimination against foreign males (in contrast to Z females), and as such will fail to contribute to phenotypic (behavioral) differences between Z and M groups. Therefore, Z females were mostly compared with nonmated M females to assure a cost-effective approach. We did not address the problem of gene expression polymorphism within M or Z groups. Instead, we aimed to detect inter-group differentiation between M and Z. For that reason, we used geographically distant M populations to provide a wide coverage of genetic differentiation and thus assure a statistically conservative approach. The moderated t-statistic based on an empirical Bayes method and implemented in the limma package (Irizarry 2005) was used to estimate the false discovery rate (FDR) and FDR-adjusted P-values. Additionally, the SAM thresholding procedure (Tusher *et al*. 2001) was used to generate the plots for FDR (Fig. 1).

Hierarchical clustering was used to infer patterns of coregulation among the candidate genes. The distance between two genes is defined as 1 - r where r is the

Pearson correlation coefficient between the standardized expression values (mean 0 and standard deviation 1) of the two genes across the samples used (Li and Wong 2003). Standardization and clustering methods followed Golub *et al*. (1999) and Eisen *et al*. (1998). To detect overrepresented functional groups, information from the Gene Ontology (GO) database (Gene-Ontology-Consortium 2001) was integrated with the expression data using the EASE application (Hosack *et al*. 2003), and probability values were assigned to each observed GO representation relative to GO fractions for all Affymetrix probe sets.

*Quantitative RT-PCR*

Based on their statistical significance and potentially relevant functional classification, four candidate genes (Table 2) were selected for quantitative real-time fluorescent polymerase chain reaction (QRT-PCR) analysis to confirm the microarray results with independent RNA sampling and methodology. For QRT-PCR, 100 ng of total RNA prepared as described above were reverse transcribed and amplified in two-step reactions using Promega Access RT-PCR system (cat. # A1250) and Promega protocols. QRT-PCR reactions contained 5 µL AMV/ Tfl5 Reaction Buffer, 1 µL 25mM $MgSO_4$, 1.5 µL 0.0001% SYBR Green, 1 µL 20 µM downstream and upstream primers each, 0.5 µL dNTP mix (10 mM each), 0.5 µL TflDNA Polymerase, 0.5 µL AMV Reverse Transcriptase, and complemented to a total of 25 µL with DEPC-treated water. All PCRs were run on the same BIO-RAD iCycler with MyiQ$^{TM}$ Optical Module, and included an initial 45 min at 48ºC for reverse transcription, 2 min at 95ºC, 35 cycles of 95ºC for 30 s, 55ºC for 1 min and 72ºC for 2 min. CT-values of the target gene were corrected by β-Actin values as a normalizer. To test for contamination with genomic DNA, negative

controls for each sample contained RNA and all other components, excluding the AMV Reverse Transcriptase. Three-way ANOVA of threshold cycle ($C_T$) values was used to analyze differential expression with Group (M and Z), mating status, and PCR reaction as fixed factors.

**RESULTS**

Intergroup mate choice tests confirmed the expected pattern of asymmetric behavioral isolation between Z and M (Fig. 2). All 19 replicates of tests with Z females exposed to both types of males resulted in preference for conspecific males (range 60-100%), compared to 24 replicates of tests with M females out of which 10 resulted in preference for heterotypic males (<84%), 6 for conspecific males (<71%), and 8 no preference (50% of each type). Thus, M females showed a slight preference to mate with Z males but this tendency was not significant (contingency table $\chi^2$= 2.4955, df=1, *P*=0.1142). Conversely, Z females strongly discriminated against heterotypic (M) males ($\chi^2$=19.7195, df=3, *P*=0.0002).

We obtained four nonmated and five mated Z transcriptome profiles compared with three nonmated and one mated M profiles. Knudsen (2002) showed that four replicates are sufficient to keep the rate of false positives low. Replicates were biologically diversified to represent variation within group (M stocks originated from different continents). We also reasoned that pooling of mRNA from multiple females (see Materials and Methods) effectively increases the level of biological replication and reduces variability between arrays (see also Kendziorski *et al*. 2005). We found that 1065 genes were differentially expressed between Z and M, 609 genes were differentially expressed between mated and nonmated Z, and only 45 genes were simultaneously

different ($P < 0.05$) in both contrasts (Table1). The intention behind this two-way contrast was to identify Z-specific transcript changes associated with mating effects, thus manifest in mated Z females but not nonmated Z and any M. We reasoned that if expression of a gene responsible for sexual discrimination in Z females is modulated through mating, nonmated Z (rather than mated Z) females will be more M-like. Thus of particular interest are those 17 genes, such as *Odorant receptor 63a*, whose expression difference between mated and nonmated Z was in the same direction, respectively, as between Z and M (Fig. 3).

We assessed whether probe sets with significantly altered transcript abundance due to our experimental conditions were randomly distributed across the five major chromosome arms. We used a chi square goodness-of-fit test to check for significant deviation of observed from expected numbers of affected loci on each chromosome. By conventional criteria ($P < 0.05$), this deviation could not be considered significant ($\chi^2$= 9.387, df = 4, $P$ = 0.0521; after including the fourth chromosome the significance decreases even further, $\chi^2$= 10.181, df = 4, $P$ =0.070). Three biological processes (physiological, cellular, and development), four molecular functions (catalytic activity, binding, transporter activity, and transcription regulator activity), and one cellular component (cell) were present among GO groups represented by the 45 genes, but none of them were statistically overrepresented according to EASE ($P < 0.05$).

Using QRT-PCR, we tested expression changes in four candidate genes: *Odorant Receptor 63a*, *desaturase2*, *CG10120*, *CG1812* (Table 2). We confirmed that *Odorant Receptor 63a* was downregulated in Z females relative to M and at the same time suppressed in mated Z females compared with nonmated Z females. For *desaturase2* and

*CG10120*, we validated the difference between Cosmopolitan and Zimbabwe to be consistent with microarray results (Fig. 4) but there was no significant difference between mated and nonmated Z females. Neither contrast was statistically significant for *CG1812* (Table 2).

**DISCUSSION**

Behavioral isolation in *Drosophila* is exerted through courtship behavior, consisting of sequential actions that exchange auditory, visual, and chemosensory signals between males and females (e.g., Greenspan and Ferveur 2000; Markow and O'Grady 2005). Given the complexity of mating behavior, it should not be surprising that candidate genes represent a variety of biological functions. One of the candidate genes, *Odorant receptor 63a*, plays a critical role in the olfactory system and mate recognition in *Drosophila*. Odorant receptors constitute the molecular basis for the detection of volatile odorous molecules and the perception of smell (Clyne *et al*. 1999; Vosshall *et al.* 1999, Vosshall *et al*. 2000). The members of the odorant receptor gene family (~60) in the *Drosophila* genome are considerably divergent, with an average amino acid identity of ~20%, and each of them encodes a putative seven-transmembrane domain protein of about 380 amino acids (Vosshall *et al*. 2000). Odorant receptors are expressed in small subsets of olfactory receptor neurons in the olfactory sensory organs of adult *Drosophila*, mostly the antenna and the maxillary palp (Clyne *et al*. 1999). Despite the extreme confinement and patchiness of expression, we were able to validate the microarray expression pattern of *Odorant receptor 63a* with quantitative RT-PCR (Table 2). This gene was downregulated in Z females relative to M and at the same time suppressed in mated Z females compared with nonmated Z females (Fig. 3A).

Another receptor, *Nmda1* (N-methyl-D-aspartate receptor-associated protein), represents a class of glutamate receptors that are of central importance in synaptic plasticity. NMDA receptors allow $Ca^{2+}$ influx and are thought to trigger $Ca^{2+}$ dependent postsynaptic processes involved in long term potentiation and depression (Sucher *et al*. 1996). This gene has been reported among transcripts typical for neuroblast lineage development (Brody *et al*. 2002). Receptor activity was also changed at *neurexin* putatively involved in cell adhesion (Arbeitman *et al*. 2004). Two other genes with changes in expression, *elk* and *CG15088*, are involved in potassium transport. In mammals, sexual interactions have been known to cause extensive changes in channels, signal transduction and neurotransmission (Bradley *et al*. 2005).

Notably, we found that transcription of *desaturase2* (*CG5925*), a gene with stearoyl-CoA 9-desaturase activity involved in cuticle hydrocarbon biosynthesis, was also altered in our experimental conditions. Its regulation pattern clustered together with *CG7763* (putative sugar binding activity), *geko* (*CG13695*), *hoe1* (*CG12787*, producing a membrane component with transporter activity), CG6208 (soluble NSF attachment protein activity involved in vesicle-mediated transport), *CH5* (NEDD8 activating enzyme involved in axonogenesis), and *Dsor1* involved in signal transduction (Fig. 3B). This gene has been suggested to be involved in differential adaptation to climate as well as behavioral isolation between Z and M lines (Fang *et al*. 2002; Greenberg *et al*. 2003; Greenberg *et al*. 2006, but see Coyne and Elwyn 2006a,b). Moreover, functional variation at *desaturase2* locus has a distinct geographic pattern: females from African (including Z) and Caribbean populations produce 5,9-heptacosadiene (5,9-HD) as the predominant cuticle hydrocarbon. This product seems to be absent from most

Cosmopolitan females as a result of a 16-bp deletion at the 5' end of the gene, largely contributing to the polymorphism of cuticle hydrocarbons (Dallerac *et al*. 2000; Takahashi *et al*. 2001). The loss-of-function allele is widely distributed over the world and the nucleotide diversity at the locus as well as experimental data suggest that this spread may have been driven by positive selection for increased cold resistance (Takahashi *et al*. 2001; Greenberg *et al*. 2003; Greenberg *et al*. 2006, but see Coyne and Elwyn 2006a,b). Although downregulated relative to Z females, *desaturase2* transcripts were still present in M females in contrast to the report by Dallerac *et al*. (2000). Using RT-PCR, we confirmed a difference between Cosmopolitan and Zimbabwe consistent with the results of microarray analysis (Tables 1 and 2, Fig. 4) but the difference between mated and nonmated females was not significant.

In addition to *desaturase2*, transcription levels of 10 other enzymes were related to divergence between the M and Z morphs, including *CAH1* with carbonic anhydrase activity and *Pi3K68D* with phosphatidylinositol-4-phosphate 3-kinase activity. Seven transcription factors or other DNA binding gene products were altered in the experimental conditions. These include *trithorax* which is required to maintain the proper spatial pattern of expression for multiple homeotic genes of the Bithorax and Antennapedia complexes (Kuzin *et al*. 1994). *Trithorax* has also been reported to interact genetically with *brm*, *ph-p*, *Pc*, *mod(mdg4)*, *Asx* and 32 other genes (FlyBase Report). However, none of these occurred among the candidate genes from Table 1 or at the cluster of coregulated transcripts (not shown). One of the two most closely coregulated genes with *trithorax* was *CG6701* that has two conserved domains related to DNA and RNA helicases and helicase subunits (DNA replication, recombination, and repair;

Celniker *et al*. 2002). The other tightly coregulated gene (*CG3558*) has not been functionally characterized. In adults, *trithorax* is known to affect the development of sex combs, which are male specific chaetae located on the prothoracic tarsal segment of the prothoracic leg, in addition to influencing development of at least eight other tissues.

At least three loci on the third chromosome have been mapped for female discrimination between Zimbabwe and Cosmopolitan populations (Ting *et al*. 2001) and this raises a question about the genomic distribution of expression changes we observed. Sexual conflict provides a plausible mechanism leading to a nonrandom distribution of genes with sex-biased patterns of expression. Specifically, the evolution of sex-biased gene expression may alleviate sexual antagonism which has been shown to be extensive in adult *Drosophila* (Chippindale *et al*. 2001). Transcription level is an excellent measure of sex-biased function and global expression profiles do provide evidence of extensive sex-biased expression in *Drosophila* and *C. elegans* (Andrews *et al*. 2000; Reinke *et al*. 2000; Jiang *et al*. 2001; Jin *et al*. 2001; Swanson *et al*. 2001; Arbeitman *et al*. 2002; Kelly *et al*. 2002; Meiklejohn *et al*. 2003; Parisi *et al*. 2003; Ranz *et al*. 2003, Michalak and Noor 2003). As the X chromosome is depauperate for male-specific genes probably due to demasculinization or feminization effects of selection (Parisi *et al*. 2003; Ranz *et al*. 2003; Oliver and Parisi 2004; Mackay *et al*. 2005), it is thus plausible that X chromosome-linked transcripts have a disproportionate contribution to sexual isolation. However, we observed no chromosomal overrepresentation of analyzed loci across the genome.

Our results also suggest substantial epistasis (many interacting genes involved) and pleiotropy (same genes, such as *trithorax*, affect multiple traits) underlying

differences between populations of *D. melanogaster*. Further studies are evidently required to determine whether any candidate genes reported here are directly responsible for behavioral isolation between Cosmopolitan and Zimbabwe groups. The genetic architecture underlying behavioral isolation may be extraordinarily complex, as exemplified by mapping genes responsible for reinforcement in *Drosophila pseudoobscura* and *D. persimilis*; Ortiz-Barrientos *et al*. (2004) found that the genetic architecture of basal female mating discrimination (between allopatric populations) is different from that of reinforced mating discrimination (between sympatric populations).

In sum, we demonstrated significant differentiation of gene expression in two major morphs of *Drosophila melanogaster*, Cosmopolitan and Zimbabwe. As some of those differences were additionally modified by female mating status, we believe that the list of identified candidate genes may be very useful in further studies of sexual isolation between these two *Drosophila melanogaster* morphs. Prior to our study, only one gene, *desaturase2*, has been invoked to play a role in sexual isolation between Cosmopolitan and Zimbabwe flies. Remarkably, this gene was among 44 other candidate genes detected here with microarrays. However, there are several limitations in the presented analysis, such as relatively high FDR resulting from the moderate sampling and rather low fold changes. Also due to the experimental setup, expression changes in the candidate genes likely represent postmating effects, such as insemination, in addition to mating discrimination.

**ACKNOWLEDGEMENTS**

This research was supported by an UTA's Research Enhancement Program grant (awarded to PM) and an Animal Behaviour Society grant (awarded to JHM). We thank D.L. Hawkins for statistical assistance and three anonymous reviewers for their constructive comments.

**REFERENCES**

Alipaz JA, Wu C-I, Karr TL (2001) Gametic incompatibilities between races of *Drosophila melanogaster*. Proceedings of the Royal Society of London Series B-Biological Sciences, **268**, 789-95.

Andrews J, Bouffard GG, Cheadle C, Lu J, Becker KG, Oliver B (2000) Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. Genome Research, **10**, 2030-2043.

Arbeitman MN, Furlong EE, Imam F *et al*. (2002) Gene expression during the life cycle of *Drosophila melanogaster*. Science, **297**, 2270-2275.

Arbeitman MN, Fleming AA, Siegal ML, Null BH, Baker BS (2004) A genomic analysis of Drosophila somatic sexual differentiation and its regulation. Development, **131**, 2007-2021.

Bradley KC, Boulware MB, Jiang H, Doerge RW, Meisel RL, Mermelstein PG (2005) Changes in gene expression within the nucleus accumbens and striatum following sexual experience. Genes Brain and Behavior, **4**, 31-44.

Brody T, Stivers C, Nagle J, Odenwald W (2002) Identification of novel *Drosophila* neural precursor genes using a differential embryonic head cDNA screen. Mechanisms of Development, **113**, 41-59

Bucan M, Abel T (2002) The mouse: genetics meets behaviour. Nature Review Genetics, **3,** 622-630.

Capy P, Veuille M, Paillette M, Jallon JM, Vouidibio J, David JR (2000) Sexual isolation of genetically differentiated sympatric populations of *Drosophila melanogaster* in Brazzaville, Congo: the first step towards speciation? Heredity, **84**, 468-475.

Cáceres M, Lachuer J, Zapala MA *et al*. (2003) Elevated gene expression levels distinguish human from non-human primate brains. Proceedings of the National Academy of Sciences of the United States of America, **100**, 13030-13035.

Capy P, Veuille M, Paillette M, Jallon JM, Vouidibio J, David JR (2000) Sexual isolation of genetically differentiated sympatric populations of *Drosophila melanogaster* in Brazzaville, Congo: the first step towards speciation? Heredity, 84, 468-475.

Cash AC, Whitfield CW, Ismail N, Robinson GE (2005) Behavior and the limits of genomic plasticity: power and replicability in microarray analysis of honeybee brains. Genes Brain and Behavior, **4**, 267-271.

Celniker SE, Wheeler DA, Kronmiller B *et al*. (2002) Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence. Genome Biology, **3**, 0079.

Chapman T, Liddle LF, Kalb JM, Wolfner MF, Partridge L (1995) Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. Nature, **373**, 241-244.

Chapman T, Bangham J, Vinti G *et al*. (2003) The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. Proceedings of the National Academy of Sciences of the United States of America, **100**, 9923-9928.

Cheung VG, Conlin LK, Weber TM *et al*. (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. Nature Genetics, **33**, 422-425.

Chippindale AK, Gibson JR, Rice WR (2001) Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America, **98**, 1671-1675.

Clark AG, Aguade M, Prout T, Harshman LG, Langley CH (1995) Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. Genetics, **139**, 189-201.

Clayton DF (2004) Songbird genomics: methods, mechanisms, opportunities, and pitfalls. Annals of the New York Academy of Sciences, **1016**, 45-60.

Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR (1999) A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. Neuron, **22**, 203-204.

Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. Nature Genetics, **32**, 432-437.

Coyne JA, Orr HA (1989) Patterns of speciation in *Drosophila*. Evolution, **43**, 362-381.

Coyne JA, Orr HA (1997) "Patterns of speciation in *Drosophila*" revisited. Evolution, **51**, 295-303.

Coyne JA, Orr HA (2004) *Speciation*. Sunderland, MA: Sinauer Associates.

Coyne JA, Elwyn S (2006a) Does the *desaturase-2* locus in *Drosophila melanogaster* cause adaptation and sexual isolation? Evolution, **60**, 279-291.

Coyne JA, Elwyn S (2006b) Desaturase-2, environmental adaptation, and sexual isolation in *Drosophila melanogaster*. Evolution, **60**, 626-627.

Dallerac R, Labeur C, Jallon JM, Knipple DC, Roelofs WL, Wicker-Thomas C (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.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster Analysis and Display of Genome-Wide Expression Patterns. Proceedings of the National Academy of Sciences of the United States of America, **95**, 14863-14868.

Enard W, Khaitovich P, Klose J *et al*. (2002) Intra- and interspecific variation in primate gene expression patterns. Science, **296**, 340-343.

Fang S, Takahashi A, Wu C-I (2003) A mutation in the promoter of desaturase2 is correlated with sexual isolation between *Drosophila* behavioral races. Genetics, **162**, 781-784.

Gene-Ontology-Consortium (2001) Creating the gene ontology resource: design and implementation. Genome Research, 11, 1425-1433.

Golub TR, Slonim DK, Tamayo P *et al*. (1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science*, 286, 531-537.

Greenberg AJ, Moran JR, Coyne JA, Wu C-I (2003) Ecological adaptation during incipient speciation revealed by precise gene replacement. Science, **302**, 1754-1757.

Greenberg AJ, Moran JR, Wu C-I (2006) Proper control of genetic background with precise allele substitution: A comment on Coyne and Elwyn. Evolution, **60**, 623-625.

Greenspan RJ, Ferveur JF (2000) Courtship in *Drosophila*. Annual Review of Genetics, **34**, 205-232.

Grozinger CM, Sharabash NM, Whitfield CW, Robinson GE (2003) Pheromone-mediated gene expression in the honey bee brain. Proceedings of the National Academy of Sciences of the United States of America, **100**, 14519-14525.

Gu J, Gu X (2003) Induced gene expression in human brain after the split from chimpanzee. Trends in Genetics, **19**, 63-65.

Heifetz Y, Lung O, Frongillo EA Jr, Wolfner MF (2000) The *Drosophila* seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. Current Biology, **10**, 99-102.

Henderson NR, Lambert DM (1982) No significant deviation. from random mating of world-wide population of *Drosophila melanogaster*. Nature, **300**, 437–440.

Hollocher H, Ting CT, Pollack F, Wu C-I (1997a) Incipient speciation by sexual isolation in *Drosophila melanogaster*: variation in mating preference and correlation between sexes. Evolution, **51**, 1171-1181.

Hollocher H, Ting CT, Wu ML, Wu C-I (1997b) Incipient speciation by sexual isolation in *Drosophila melanogaster*: extensive genetic divergence without reinforcement. Genetics, **147**, 1191-1201.

Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA (2003) Identifying biological themes within lists of genes with EASE. Genome Biology, 4, P4.

Irizarry RA (2005) From CEL files to annotated lists of interesting genes. Biostatistics, **4**, 249-264.

Jarvis, ED, Smith VA, Wada K *et al*. (2002) A framework for integrating the songbird brain. Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology, **188**, 961-980.

Jiang M, Ryu J, Kiraly M, Duke K, Reinke V, Kim SK (2001) Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. Proceedings of the National Academy of Sciences of the United States of America, **98**, 218-223.

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. Nature Genetics, **29**, 389-395.

Kaplan WD, Tinderholt VE, Gugler DH (1962) The number of sperm present in the reproductive tracts of *Drosophila melanogaster* females. Drosophila Information Service, **36**, 82.

Kelly WG, Schaner CE, Dernburg AF, Lee MH, Kim SK, Villeneuve AM, Reinke V (2002) X-chromosome silencing in the germline of *C. elegans*. Development, **129**, 479-492.

Kendziorski C, Irizarry RA, Chen KS, Haag JD, Gould MN (2005) On the utility of pooling biological samples in microarray experiments. Proceedings of the National Academy of Sciences of the United States of America, **102**, 4252-4257.

King MC, Wilson AC (1975) Evolution at two levels in Humans and Chimpanzees. Science, **188**, 107-116.

Knudsen S (2002) A Biologist's Guide to Analysis of DNA Microarray Data. Wiley-Liss, New York.

Korol A, Rashkovetsky E, Iliadi K, Michalak P, Ronin Y, Nevo E (2000) Nonrandom mating in Drosophila melanogaster laboratory populations derived from closely adjacent ecologically contrasting slopes at "Evolution Canyon". Proceedings of the National Academy of Sciences of the United States of America, **97**, 12637-12642.

Kuzin B, Tillib S, Sedkov YA, Mizrokhi LI, Mazo A (1994) The *Drosophila* trithorax gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene fork head. Genes & Development, **8**, 2478-2490.

Liang P, Pardee AB (1997) Differential display: a general protocol. In: Methods in molecular biology: Differential display methods and protocols (ed. Liang P, Pardee AB), pp. 3-12. Humana Press, New Jersey.

Lawniczak MK, Begun DJ (2004) A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. Genome, **47**, 900-910.

Li C, Wong WH (2001a) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proceedings of the National Academy of Sciences of the United States of America, **98**, 31-36.

Li C, Wong WH (2001b) Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. Genome Biology, **2**, 1-11.

Li C, Wong WH (2003) DNA-Chip Analyzer (dChip). In: The Analysis of Gene Expression Data: Methods and Software (ed. Parmigiani G, Garrett ES, Irizarry RA, Zeger SL), pp. 120-141. Springer, New York.

Mackay T F, Heinsohn SL, Lyman RF, Moehring AJ, Morgan TJ, Rollmann SM (2005) Genetics and genomics of *Drosophila* mating behavior. Proceedings of the National Academy of Sciences of the United States of America, **102**, 6622-6629.

Markow TA, O'Grady PM (2005) Evolutionary genetics of reproductive behavior in Drosophila: connecting the dots. Annual Review of Genetics, **39**, 263-291.

Mayr E (1963) Animal Species and Evolution. Belknap Press, Cambridge, MA.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL (2003) Rapid evolution of male-biased gene expression in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America, **100**, 9894-9899.

Mello CV, Vicario DS, Clayton DF (1992) Song presentation induces gene expression in the songbird forebrain. Proceedings of the National Academy of Sciences of the United States of America, **89,** 6818-6822.

Michalak P, Noor MAF (2003) Genome-wide patterns of expression in *Drosophila* pure–species and hybrid males. Molecular Biology and Evolution, **20**, 1070-1076.

Michalak P, Noor MAF (2006) Genetics of reproductive isolation and species differences in model systems. In: Evolutionary Genetics (ed. Fox C, Wolf JB), pp. 387-398. Oxford University Press, USA.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM (2004) Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. Molecular Biology and Evolution, **21**, 1308-1317.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. Nat Genetics, **32**, 261-266.

Oliver B, Parisi M (2004) Battle of the Xs. Bioessays, **26,** 543-548.

Orr HA, Masly JP, Presgraves DC (2004) Speciation genes. Current Opinion in Genetics and Development, **14,** 675-679.

Ortiz-Barrientos D, Counterman BA, Noor MAF (2004) The genetics of speciation by reinforcement. Public Library of Science (PLoS) Biology **2**:e416.

Panhuis TM, Swanson WJ, Nunney L (2003) Population genetics of accessory gland proteins and sexual behavior in Drosophila melanogaster populations from Evolution Canyon. Evolution, **57**, 2785-2791.

Parisi M, Nuttal R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Okiver B (2003) Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. Science, **299**, 697-700.

Preuss TM, Caceres M, Oldham MC, Geschwind DH (2004) Human brain evolution: insights from microarrays. Nature Reviews Genetics, **5**:850-860.

Rankin CH (2002) From gene to identified neuron to behaviour in *Caenorhabditis elegans*. Nature Reviews Genetics, **3**, 622-630.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science, **300**, 1742-1745.

Ranz JM, Machado CA (2006) Uncovering evolutionary patterns of gene expression using microarrays. Trends in Ecology and Evolution, **21**, 29-37.

Reinke V, Smith HE, Nance J *et al*. (2000). A global profile of germline gene expression in *C. elegans*. Molecular Cell, **6**, 605-616.

Rice WR, Linder JE, Friberg U, Lew TA, Morrow EH, Stewart AD (2005) Inter-locus antagonistic coevolution as an engine of speciation: assessment with hemiclonal analysis. Proceedings of the National Academy of Sciences of the United States of America, **102**, 6527-6534.

Rifkin SA, Kim J, White KP (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. Nature Genetics, **33**, 138-144.

Robinson GE, Grozinger CM, Whitfield CW (2005) Sociogenomics: Social life in molecular terms. Nature Reviews Genetics, **6**:257-270.

Saetre P, Lindberg J, Leonard JA *et al*. (2004) From wild wolf to domestic dog: gene expression changes in the brain. Brain Research Molecular Brain Research, **126**, 198-206.

Sasa M, Chippindale PT, Johnson NA. 1998. Patterns of postzygotic reproductive isolation in frogs. Evolution, **52**, 1801-1810.

Seehausen O, van Alphen JJM, Witte F (1997) Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection. Science, **277**, 1808-1811.

Sucher NJ, Awobuluyi M, Choi YB, Lipton SA (1996) NMDA receptors: from genes to channels. Trends in Pharmacological Sciences, **17**, 348-355.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America, **19**, 7375-7379.

Takahashi A, Tsaur SC, Coyne JA, Wu C-I (2001) The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America, **98**, 3920-3925.

Ting CT, Takahashi A, Wu C-I (2001) Incipient speciation by sexual isolation in *Drosophila*: concurrent evolution at multiple loci. Proceedings of the National Academy of Sciences of the United States of America, **98**, 6709-6713.

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America, **98**, 5116-5121.

Uddin M, Wildman DE, Liu G *et al*. (2004) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. Proceedings of the National Academy of Sciences of the United States of America, **101**, 2957-2962.

Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axe R (1999) A spatial map of olfactory receptor expression in the *Drosophila* antenna. Cell, **96**, 725–736.

Vosshall LB, Wong A, Axel R (2000) An Olfactory Sensory Map in the Fly Brain. Cell, 102, 147-159.

Whitehead A, Crawford DL (2006) Variation within and among species in gene expression: raw material for evolution. Molecular Ecology, **15**, 1197-1211.

Whitfield CW, Cziko AM, Robinson GE (2003a) Gene expression profiles in the brain predict behavior in individual honey bees. Science, **302**, 296-299.

Whitfield CW, Brillet C, LeConte Y, Robinson GE (2003b) Behavioral plasticity and gene expression in the brain: microarray analysis of the influences of environment and genotype on behavioral maturation in the honeybee. Society for Neuroscience Abstracts, **627**, 15.

Wolfner MF, Harada HA, Bertram MJ *et al*. (1997) New genes for male accessory gland proteins in *Drosophila melanogaster*. Insect Biochemistry and Molecular Biology, **27**, 825-834.

Woolley SC, Sakata JT, Crews D (2004) Tracing the evolution of brain and behavior using two related species of whiptail lizards: *Cnemidophorus uniparens* and *Cnemidophorus inornatus*. Institute of Laboratory Animal Resources (Ilar) Journal, 45, 46-53.

Wu C-I., Hollocher H, Begun DJ, Aquadro CF, Xu Y, Wu ML (1995) Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. Proceedings of the National Academy of Sciences of the United States of America, **92**, 2519-2523.

**Table 1. Genes differentially expressed between Zimbabwe (Z) and Cosmopolitan (M) and simultaneously affected by the mating experience (v – nonmated Z, nv – mated Z).** † symbols denote genes chosen for the QRT-PCR analysis and asterisks (*) denote P-values <0.05 after adjustment for FDR.

| Locus | Activity description | Z/M fold change | P-value (Z – M contrast) | nv/v fold change | P-value (nv – v contrast) |
|---|---|---|---|---|---|
| *desaturase2*† | fatty acid desaturase | 1.873 Z > M | 0.009* | 1.655 nv < v | 0.006 |
| *CG11357* | enzyme | 1.072 Z > M | 0.004 | 1.053 nv < v | 0.049 |
| *CG12400* | NADH dehydrogenase | 1.052 Z < M | 0.024 | 1.034 nv < v | 0.026 |
| *CG11836* | endopeptidase | 1.645 Z < M | 0.001 | 1.140 nv > v | 0.016 |
| *CG15547* | enzyme | 1.122 Z < M | 0.013* | 1.125 nv > v | 0.004* |
| *CG10120*† | deoxyribonuclease | 1.126 Z < M | 0.003* | 1.179 nv > v | 0.0004 |
| *CG4199* | enzyme | 1.026 Z < M | 0.036* | 1.032 nv > v | 0.0007* |
| *CAH1* | carbonate dehydratase | 1.203 Z > M | 0.034 | 1.219 nv < v | 0.002 |
| *CG3027* | beta-ureidopropionase | 1.026 Z < M | 0.009* | 1.020 nv < v | 0.020 |
| *Aats-gln* | Glutaminyl-tRNA synthetase | 1.271 Z > M | 0.044 | 1.353 nv < v | 0.009 |
| *SamDC* | adenosylmethionine decarboxylase | 2.040 Z < M | 0.001 | 1.465 nv > v | 0.025 |
| *Rpn11* | proteasome regulatory particle | 1.489 Z < M | 0.004 | 1.307 nv < v | 0.017 |
| *Arf72A* | ARF small GTPase | 1.016 Z < M | 0.029 | 1.021 nv > v | 0.023 |
| *Elk* | potassium channel | 11.05 Z > M | 0.0004 | 1.815 nv < v | 0.042 |
| *Pi3K68D* | Phosphotidylinositol 3 kinase | 1.268 Z > M | 0.023 | 1.292 nv > v | 0.035 |
| *CG15088* | neurotransmitter transporter | 1.036 Z < M | 0.006 | 1.036 nv < v | 0.034 |
| *CG7050* | Neurexin, cell adhesion | 1.858 Z < M | 0.004* | 1.662 nv < v | 0.028 |
| *Cyp4p2* | cytochrome P45 | 1.026 Z < M | 0.039* | 1.029 nv > v | 0.044* |
| *CHORD* | Zink ion binding | 1.245 Z < M | 0.023 | 1.392 nv < v | 0.028 |
| *CG1812*† | transcription factor | 1.296 Z < M | 0.000001* | 1.124 nv > v | 0.008* |
| *CG14730* | DNA binding | 2.034 Z < M | 0.019 | 4.675 nv < v | 0.030 |
| *Bowl* | RNA polymerase II transcription factor | 1.077 Z < M | 0.003* | 1.076 nv > v | 0.039 |
| *Trithorax* | DNA binding | 1.089 Z < M | 0.011* | 1.098 nv > v | 0.048* |
| *prospero* | RNA polymerase II transcription factor | 1.042 Z < M | 0.023 | 1.049 nv > v | 0.012* |
| *CG14962* | nucleic acid binding | 1.033 Z < M | 0.010 | 1.036 nv > v | 0.020 |
| *Gcn2* | elongation-factor-2 kinase | 1.391 Z < M | 0.004 | 1.390 nv > v | 0.043 |
| *CG14641* | RNA binding | 1.281 Z < M | 0.008 | 1.231 nv > v | 0.018 |

| | | | | | |
|---|---|---|---|---|---|
| *CG15636* | DNA binding | 1.029 Z > M | 0.010 | 1.036 nv > v | 0.036 |
| *Schnurri* | RNA polymerase II transcription factor | 1.633 Z > M | 0.030 | 1.433 nv > v | 0.023 |
| *Odorant receptor 63a*[†] | olfactory receptor | 1.856 Z < M | 0.009* | 2.420 nv < v | 0.026 |
| *CG3814* | N-methyl-D-aspartate receptor-associated protein | 1.449 Z < M | 0.001 | 1.268 nv > v | 0.046 |
| *CG17988* | --- | 1.036 Z < M | 0.009 | 1.033 nv > v | 0.009 |
| *CG13339* | --- | 1.040 Z > M | 0.039 | 1.049 nv < v | 0.015 |
| *CG10127* | --- | 1.03 Z > M | 0.011 | 1.034 nv < v | 0.023 |
| *CG4398* | --- | 1.198 Z > M | 0.0003 * | 1.056 nv < v | 0.014 |
| *CG13301* | --- | 1.058 Z < M | 0.046 | 1.083 nv > v | 0.022 |
| *CG8844* | --- | 1.027 Z < M | 0.009* | 1.033 nv < v | 0.035 |
| *CG7530* | --- | 1.027 Z < M | 0.019 | 1.028 nv > v | 0.020* |
| *Msta* | --- | 1.350 Z > M | 0.004 | 1.207 nv < v | 0.020 |
| *CG3173* | --- | 1.017 Z > M | 0.002 | 1.014 nv > v | 0.045 |
| *CG14899* | --- | 1.401 Z > M | 0.017 | 1.373 nv > v | 0.008* |
| *CG15929* | --- | 1.098 Z > M | 0.017* | 1.093 nv < v | 0.040 |
| *CG12765* | --- | 1.037 Z < M | 0.039* | 1.024 nv < v | 0.037 |
| *CG14701* | --- | 1.032 Z > M | 0.013 | 1.027 nv > v | 0.041 |
| *CG17496* | --- | 1.043 Z > M | 0.049 | 1.071 nv > v | 0.032 |

**Table 2. Differences in threshold cycle ($C_T$) values from quantitative RT-PCR for four candidate genes.** Lower $C_T$-values represent higher expression levels. Results of 3-way ANOVA are shown, with the PCR reaction effect and interactions (not significant) not shown. The Z-M difference is subtraction of mean $C_T$-value for Cosmopolitan from the mean $C_T$ for Zimbabwe. The nv-v is between mean $C_T$-values for mated (nv) and nonmated (v) Zimbabwe females.

| Gene | df | Z-M difference | F | P-value | nv-v difference | F | P-value |
|---|---|---|---|---|---|---|---|
| *Odorant Receptor 63a* | 1, 17 | 0.135 | 50.33 | 0.0001 | 0.080 | 12.85 | 0.002 |
| *desaturase2* | 1, 15 | 0.100 | 9.64 | 0.007 | -0.002 | 0.00 | P>0.05 |
| *CG10120* | 1,8 | -0.072 | 5.51 | 0.047 | -0.014 | 0.21 | P>0.05 |
| *CG1812* | 1,16 | -0.009 | 0.07 | P>0.05 | 0.048 | 1.81 | P>0.05 |

**FIGURE LEGENDS**

Fig.1. SAM plot of the false discovery rate (FDR) against the fixed rejection region (delta) and the number of significant genes for two contrasts: (A) Zimbabwe versus Cosmopolitan and (B) mated versus nonmated Zimbabwe.

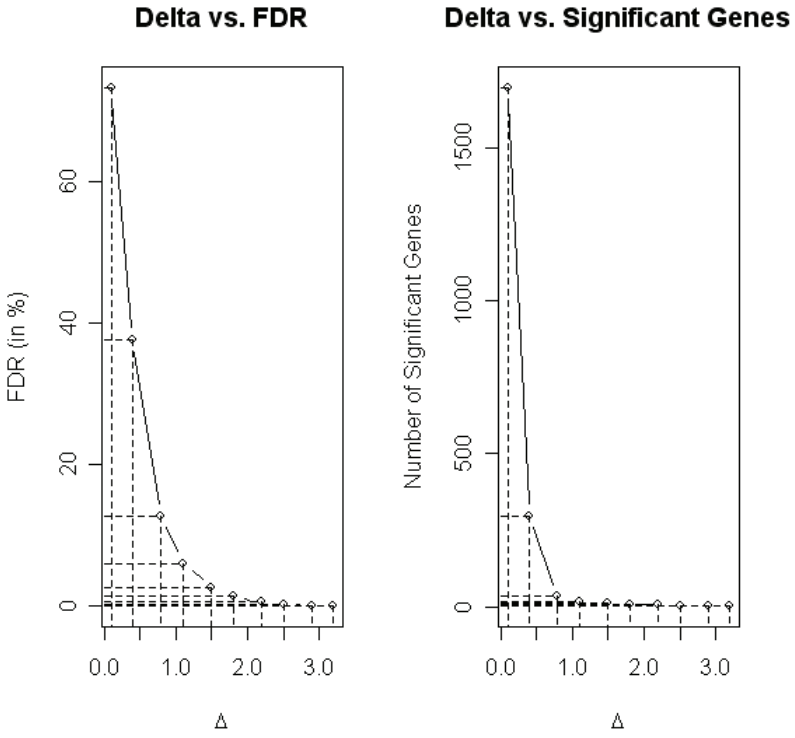
Fig. 2. Average mating frequencies from the multiple choice mating tests. Third bar in each group (Paired) represents percent of females paired.

Fig. 3. Hierarchical clustering of genes that were differentially expressed between Zimbabwe (Z) and Cosmopolitan (M) and simultaneously between mated (nv) and nonmated (v) Zimbabwe females: (A) *Odorant receptor 63a* and (B) *desaturase2* (*CG5925*). Clustering at the top of the figure groups treatments (morphs and mating status), whereas clustering on the right panel of the figure groups genes with similar expression levels, ranging from overexpression (red color) to underexpression (green color).

Fig. 4. An example of quantitative real-time PCR amplification of *desaturase2*. Black curves with circles represent Zimbabwe samples and red lines with squares correspond to Cosmopolitan samples.

Figure 1

**A.**


Delta vs. FDR
FDR (in %)
0
20
40
60
0.0
1.0
2.0
3.0
Δ
Delta vs. Significant Genes
Number of Significant Genes
0
500
1000
1500
0.0
1.0
2.0
3.0
Δ

**B.**

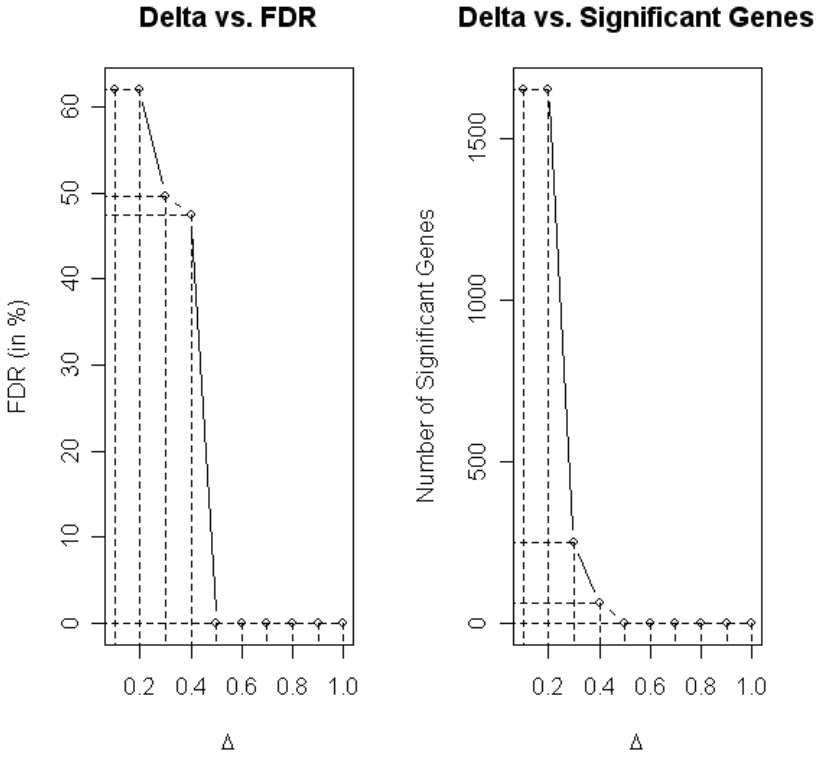

Delta vs. FDR
FDR (in %)
0
10
20
30
40
50
60
0.2
0.4
0.6
0.8
1.0
Δ
Delta vs. Significant Genes
Number of Significant Genes
0
500
1000
1500
0.2
0.4
0.6
0.8
1.0
Δ

Figure 2

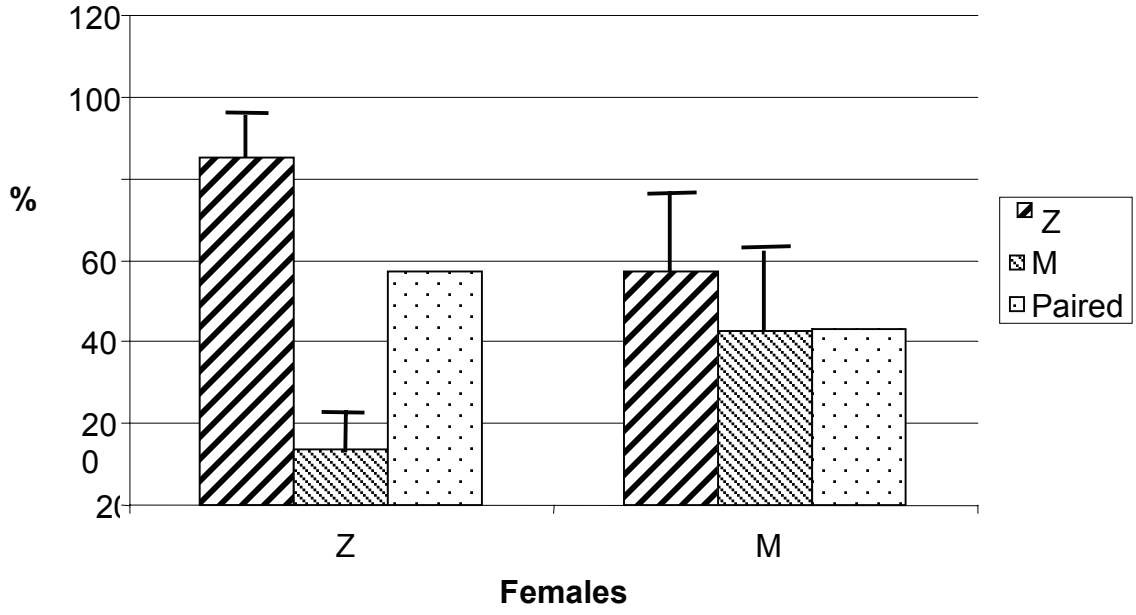
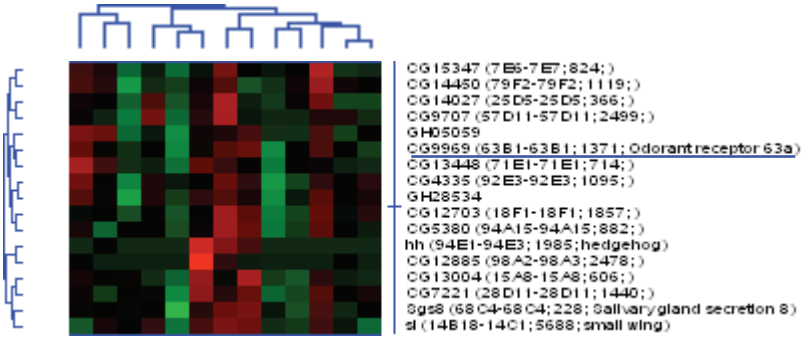


Figure 3


Mv Znv Zv Mv Zv Zv Mnv Mv Znv Zv Znv Znv Znv

**A**.


CG15347 (7E6-7E7;824;)
CG14450 (79F2-79F2;1119;)
CG14027 (25D5-25D5;366;)
CG9707 (57D11-57D11;2499;)
GH05059
CG9969 (63B1-63B1;1371;Odorant receptor 63a)
CG13448 (71E1-71E1;714;)
CG4335 (92E3-92E3;1095;)
GH28534
CG12703 (18F1-18F1;1857;)
CG5380 (94A15-94A15;882;)
hh (94E1-94E3;1985;hedgehog)
CG12885 (98A2-98A3;2478;)
CG13004 (15A8-15A8;606;)
CG7221 (28D11-28D11;1440;)
Sgs8 (68C4-68C4;228;Salivary gland secretion 8)
sl (14B18-14C1;5688;small wing)

**B**.

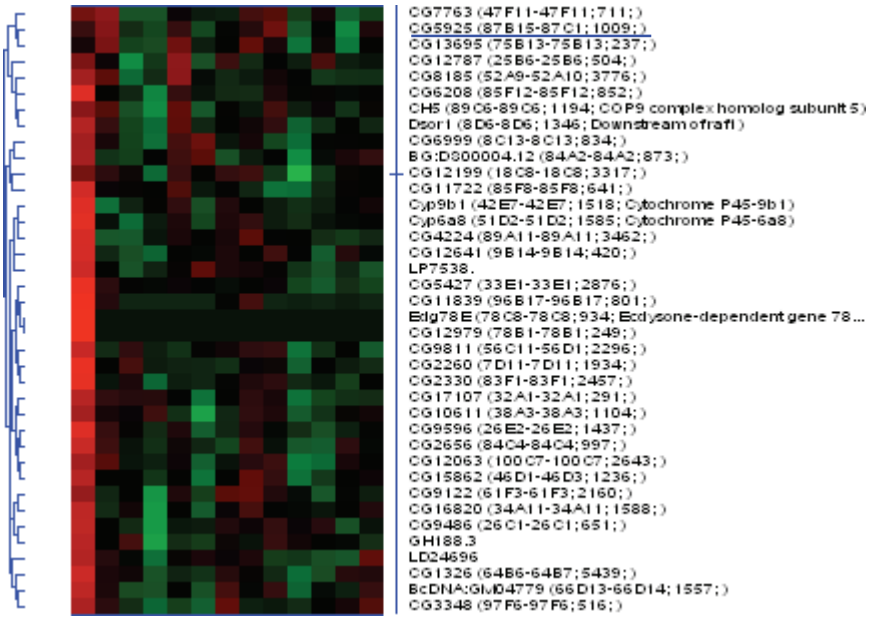
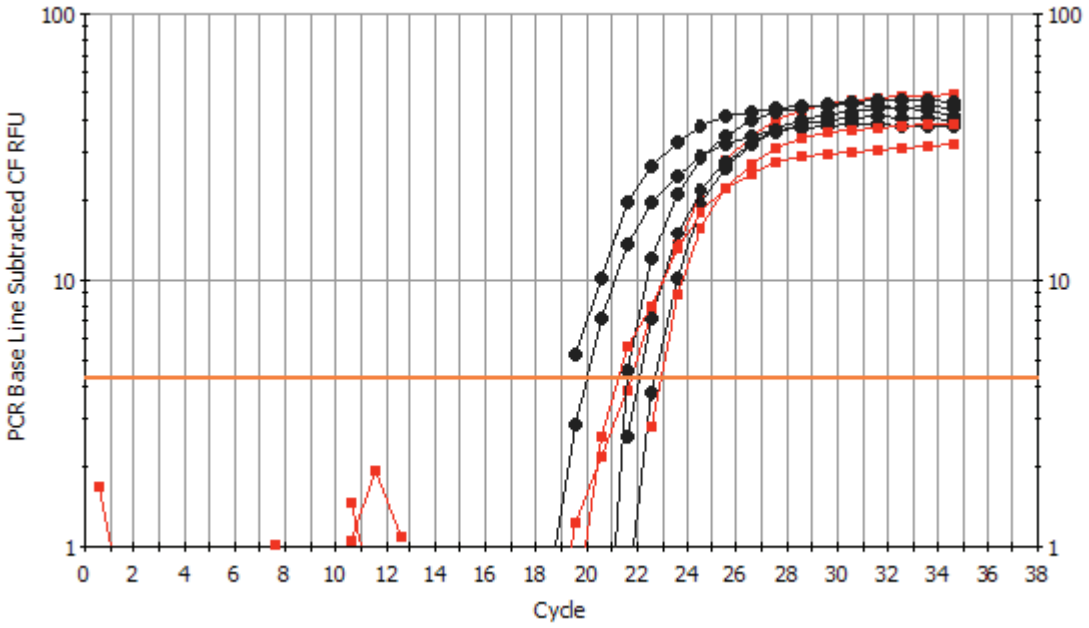

CG7763 (47F11-47F11;711;)
CG5925 (87B15-87C1;1009;)
CG13695 (75B13-75B13;237;)
CG12787 (25B6-25B6;504;)
CG8185 (52A9-52A10;3776;)
CG6208 (85F12-85F12;852;)
CH5 (89C6-89C6;1194;COP9 complex homolog subunit 5)
Dsor1 (8D6-8D6;1346;Downstream of raf1)
CG6999 (8C13-8C13;834;)
BG:DS00004.12 (84A2-84A2;873;)
CG12199 (18C8-18C8;3317;)
CG11722 (85F8-85F8;641;)
Cyp9b1 (42E7-42E7;1518;Cytochrome P45-9b1)
Cyp6a8 (51D2-51D2;1585;Cytochrome P45-6a8)
CG4224 (89A11-89A11;3462;)
CG12641 (9B14-9B14;420;)
LP7538.
CG5427 (33E1-33E1;2876;)
CG11839 (96B17-96B17;801;)
Edg78E (78C8-78C8;934;Ecdysone-dependent gene 78...
CG12979 (78B1-78B1;249;)
CG9811 (56C11-56D1;2296;)
CG2260 (7D11-7D11;1934;)
CG2330 (83F1-83F1;2457;)
CG17107 (32A1-32A1;291;)
CG10611 (38A3-38A3;1104;)
CG9596 (26E2-26E2;1437;)
CG2656 (84C4-84C4;997;)
CG12063 (100C7-100C7;2643;)
CG15862 (46D1-46D3;1236;)
CG9122 (61F3-61F3;2160;)
CG16820 (34A11-34A11;1588;)
CG9486 (26C1-26C1;651;)
GH188.3
LD24696
CG1326 (64B6-64B7;5439;)
BcDNA:GM04779 (66D13-66D14;1557;)
CG3348 (97F6-97F6;516;)

Figure 4


100
10
1
PCR Base Line Subtracted CF RFU
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
Cycle

REFERENCES

Ang LH, Kim J, Stepensky V, Hing H. 2003. Dock and Pak regulate olfactory axon pathfinding in *Drosophila*. Development 130(7):1307-1316.

Anholt RRH, Dilda CL, Chang S, Fanara JJ, Kulkarni NH, Ganguly I, Rollmann SM, Kamdar KPK, and Mackay TFC. 2003. The genetic architecture of odor-guided behavior in *Drosophila*: epistasis and the transcriptome. Nature Genetics 35(2):180-184.

Ashburner M, Bodmer M, Lemeunier, F. 1984. On the evolutionary relationships of *Drosophila melanogaster*. Developmental Genetics 4(4):295-312.

Beissbarth T. 2006. Interpreting experimental results using gene ontologies. Methods in Enzymology 411:340-352.

Breitling R, Armenguad P, Amtmann A, and Herzyk P. 2004. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Letter 573(1-3):83-92.

Chenoweth SF and Blows MW. 2006. Dissecting the complex genetics basis of mate choice. Nature Reviews Genetics 7(9):681-692.

Clemens JC, Ursuliak Z, Clemens KK, Price JV, and Dixon JE. 1996. A *Drosophila* protein-tyrosine phosphatase associates with an adapter protein required for axonal guidance. Journal of Biological Chemistry 271(29):17002-17005.

Coyne JA, Orr HA. 1998. The evolutionary genetics of speciation. Phil. Trans. R. Soc. London B 353(1366):287-305.

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

Ejima A, Nakayamma S, Aigaki T. 2001. Phenotypic association of spontaneous ovulation and sexual receptivity in virgin females of *Drosophila* mutants. Behavior Genetics 31(5):437-444.

Fuyama Y, Ueyama M. 1997. Ovulation and the suppression of mating in *Drosophila melanogaster* female: Behavioral basis. Behavior Genetics 27(5):483-488.

Garrity PA, Rao Y, Salecker I, McGlade J, Pawson T, Zipursky SL. 1996. *Drosophila* photoreceptor axon guidance and targeting requires the dreadlocks SH2/SH3 adapter protein. Cell 85(5):639-650,

Greenspan RJ, Ferveur JF. 2000. Courtship in *Drosophila*. Annual Review of Genetics 34:205-232.

Hekmat-Scafe DS, Scafe CR, McKinney AJ, and Tanouye MA. 2002. Genome-wide analysis of the odorant binding protein gene family in *Drosophila melanogaster*. Genome Research 12(9):1357-1369.

Hirai Y and Kimura MT. 1997. Incipient reproductive isolation between two morphs of *Drosophila elegans* (Diptera: Drosophilidae). Biological Journal of the Linnean Society 61(4):501-513.

Ishii K, Hirai Y, Katagiri C, and Kimura MT. 2001. Sexual isolation and cuticular hydrocarbons in *Drosophila elegans*. Heredity 87(4):392-399.

Ishii K, Hirai Y, Katagiri C, and Kimura MT. 2002. Mate discrimination and cuticular hydrocarbons in *Drosophila elegans* and *D. gunungcola*. Zoological Science 19(10):1191-1196.

Jacob NB. 1968. β-alanine use by ebony and normal *Drosophila melanogaster* with notes on glucose, uracil dopa and dopamine. Biochemical Genetics 1(3):267-275.

Jeffery IB, Higgins DG, and Culhane AC. 2006. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics 7:359.

Kalmus H. 1941. The resistance to desiccation of *Drosophila* mutants affecting body color. Proceedings of the Royal Society of London. 130(859):85-201.

Kopp A. 2006. Basal relationships in the *Drosophila melanogaster* species group. Molecular Phylogenetics and Evolution 39(3):787-798.

Kopp A and True JR. 2002. Evolution of male sexual characters in the Oriental *Drosophila melanogaster s*pecies group. Evolution and Development 4(4):278-291.

Lawniczak MKN, Begun DJ. 2004. A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. Genome 47(5):900-910.

Lomax J. 2005. Get ready to GO! A biologist's guide to the Gene Ontology. Briefings in Bioinformatics 6(3):298-304.

Mayr E. 1963. Animal species and evolution. Cambridge, MA: Harvard University.

McClintick JN, Edenberg HJ. 2006. Effects of filtering by Present call on analysis of microarray experiments. BMC Bioinromatics 7:49.

McGraw LA, Gibson G, Clark AG, Wolfner MF. 2004. Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. Current Biology 14(16):1509-1514.

Michalak P, Malone JH, Lee IT, Hoshino D, Ma D. 2007. Gene expression polymorphism in *Drosophila* populations. Molecular Ecology 16(6): 1179-1189.

Park SK, Shanbhag SR, Wang Q, Hasan G, Steinbrecht RA, and Pikielny CW. 2000. Expression patterns of two putative odorant-binding proteins in the olfactory organs of *Drosophila melanogaster* have different implications for their functions. Cell Tissue Research 300(1):181-192.

Pikielny CW, Hasan G, Rouyer F, and Rosbash M. 1994. Members of a family of *Drosophila* putative odorant-binding proteins are expressed in different subsets of olfactory hairs. Neurons 12(1):39-45.

Rundle HD, Chenoweth SF, Doughty P, and Blows MW. 2005. Divergent selection and the evolution of signal traits and mating preferences. PLoS Biology 3(11):1988-1995.

Sánchez-Gracia A, Aguadé M, and Rozas J. 2003. Patterns of nucleotide polymorphism and divergence in the odorant-binding protein genes *OS-E* and *OS-F*. Analysis in the melanogaster species subgroup of *drosophila*. Genetics 165(3):1279-1288.

Schawaroch V. 2002. Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group (Diptera: Drosophilidae). Biological Journal of the Linnean Society 76(1):21-37.

Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution by mutation clocks. Molecular Biology and Evolution 21(1):36-44.

Tram U, Wolfner MF. 1998. Seminal fluid regulation of female sexual attractiveness in *Drosophila melanogaster*. Proceed National Academy of Science 95(7):4051-4054.

Tompkins L, Hall JC. 1983. Identification of brain site controlling female receptivity in mosaics of *Drosophila melanogaster*. Genetics 103(2):179-195.

Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN. 2003.

GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biology 4(4):R28.

Zeeberg BR, Qin H, Narasimhan S, Sunshine M, Cao H, Kane DW, Reimers M, Stephens RM, Bryant D, Burt SK, Elnekave E, Hari DM, Wynn TA, Cunningham-Rundles C, Stewart DM, Nelson D, Weinstein JN. 2005. High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). BMC Bioinformatics 6:168.

# BIOGRAPHICAL INFORMATION

Daiju Hoshino graduated from Ohta High School (Japan) at March 2001 and entered the University of Texas at Arlington at August 2001 upon completion of the curriculum of the English Learning Institute at UTA. Research interests being speciation genetics, he joined the Michalak Speciation Genetics/Population Genomics Laboratory in the summer of 2003 as an undergraduate research assistant. He received his Bachelor of Science degree in Microbiology from UTA at May 2005 and kept pursuing the Master of Science degree in Biology at the Graduate school of UTA. During working in Michalak Laboratory, he presented a poster at 46$^{th}$ Annual Drosophila Research Conference at Houston at March 2006 and the article on the subject of one of the projects he greatly participated was accepted and published by Molecular Ecology at January 2007. He participated in mainly 4 projects. 1$^{st}$ project was to identify genes involved in behavioral isolation of *Drosophila melanogaster*. Two strains of *Drosophila melanogaster*, Zimbabwe and Cosmopolitan, were used to investigate. 2$^{nd}$ project was to identifying genes involved in hybrid male sterility between *Drosophila simulans* and *Drosophila mauritiana*. 3$^{rd}$ project was investigation of olfactory system of Zimbabwe stain of *Drosophila melanogaster* with antennal removal method. If he received Master of Science degree, he will go back to Japan to teach English and Science to middle school students.