ABSTRACT

Title of Document: PATTERNS OF SEX-BIASED GENE EXPRESSION AND GENE PATHWAY EVOLUTION IN *DROSOPHILA*

Kawther Abdilleh, Doctor of Philosophy, 2014

Directed By: Dr. Carlos A. Machado, Associate Professor
Department of Biology

Sexual dimorphism, or the phenotypic differences that exist between males and females of the same species, is widespread throughout nature. Sexually dimorphic traits are primarily generated by differences in gene expression between the sexes, commonly known as sex-biased gene expression. In this dissertation, I explore the evolutionary patterns and consequences of sex-biased gene expression across *Drosophila* species.

The most obvious sexually dimorphic characteristics exist in adult stages and consequently patterns of sex-bias in early *Drosophila* development have not been well-studied. In chapter 1, I examine patterns of sex-biased gene expression during ontogeny in two closely related *Drosophila* species belonging to the *D. pseudoobscura* group (*D. pseudoobscura* and *D. persimilis*). This study provides insight into global patterns of sex-bias gene expression throughout development between species.

The visual pathway in *Drosophila* shows abundant evidence for sex-biased and species-specific differential gene expression. In chapter 2, across 12 different *Drosophila* species, I

examine rates of protein sequence evolution of genes in this pathway to determine if observed differences in gene expression correlate with rates of evolutionary change at the level of protein sequence. As a whole the visual pathway in *Drosophila* exhibits strong conservation at the level of protein sequence over 65 million years of evolutionary time suggesting that observed differences in levels of transcription are the result of differences in the underlying regulatory mechanisms.

The comparative molecular evolutionary analysis of the visual pathway revealed a novel isoform-specific lineage-specific duplication event of the key signal transduction activator gene *G-alpha-q*. In *D. melanogaster*, *G-alpha-q* is present as a single-copy and alternatively spliced in a tissue- and isoform-specific manner. The same gene is duplicated in an isoform-specific manner in the species belonging to the subgenus Drosophila such that each duplicate appears to retain the exon complement of only one of the splice-variants. In chapter 3, using experimental and computational approaches, I examine the evolution of the gene structure and expression of these novel isoform-specific duplicates. This analysis revealed a mechanism by which duplicate genes can evolve novel functions and expression patterns (including sex-biased expression patterns) while retaining their ancestral functions.

PATTERNS OF SEX-BIASED GENE EXPRESSION AND GENE PATHWAY EVOLUTION IN *DROSOPHILA*

By

Kawther Abdilleh

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Associate Professor Carlos A. Machado, Chair
Associate Professor Najib M. El-Sayed
Associate Professor Eric S. Haag
Professor Thomas D. Kocher
Professor Leslie Pick, Dean's Representative

Dedication

To my Grandpa Saeed Hassan

Acknowledgements

I feel very fortunate to have worked with many people who have helped me tremendously these past few years.

I'd like to thank my advisor Carlos A. Machado for making my transition into his lab a seamless one. I am sincerely grateful for his patience, guidance and his overall support. I'm thankful he allowed me to complete the work I'd begun in my early graduate studies and gave me the freedom to explore new interests of my own.

I am very thankful for the tremendous help, support and encouragement of Tom Kocher, Karen Carleton and my committee members including Leslie Pick, Eric Haag and Najib El-Sayed

I'd like to also thank members of the Machado lab past and present including Kevin Nyberg, Carlos Flores-Lopez, Laura Rascon, Erik Mohlhenrich, Levi Teitz, Kacie Baker, and Yiressy Izaguirre Baday for valuable discussion and advice about science and life in general. I'd also like to acknowledge Kwame Okrah for help with statistics and R. Thanks to Harley King and Sarah Moore for their support, encouragement and friendship.

Last but certainly not least; I'd like to thank my family including my parents, siblings, aunts and grandparents for all their love, support and undying encouragement. I could not have done any of this without them.

Table of Contents

List of Tables

Chapter 1:

Chapter 2:

Chapter 3:

List of Figures

Chapter 1

Chapter 2

Chapter 3

INTRODUCTION

Understanding the mechanisms driving phenotypic diversity both within and between species has remained an active area of research in evolutionary biology since Darwin. Classically, phenotypic divergence between species and between the sexes across many different taxa has been examined in the pursuit of distinguishable morphological and/or behavioral characteristics and traits between the organisms under investigation (Brush 1970; Cuervo and Moller, 1999; Figuerola and Green, 2000). However, many of these studies rely on measurements of very specific traits and often cannot capture all of the differentiating characteristics between species and between the sexes. Since the majority of morphological, physiological and behavioral traits of an organism are the result of the interactions of many genes that are regulated at the level of transcription, gene expression itself can be thought of as a molecular phenotype (Madan Babu et al. 2004; Davidson and Erwin 2006; Ben-Tabou de-Leon and Davidson 2007). Gene expression fluctuates throughout the course of an organism's life span; these changes can be divergent between species, between the sexes and even between individuals (White et al. 1999; Jin et al. 2001; Arbeitman et al. 2002;Ranz et al. 2003; Schadt et al. 2003; Whitehead and Crawford 2005; Ellegren and Parsch 2007; Jiang and Machado 2009; Pavey et al. 2010; Assis et al. 2012; Romero et al. 2012; Parsch and Ellegren 2013). Thus measuring and comparing gene expression changes is a precise quantitative way to determine the global transcriptional changes that play a role in phenotypic diversity.

Much of the genome is highly conserved at the level of protein coding sequence

even between deeply divergent species (King and Wilson, 1975; Rubin et al. 2000) yet there is great morphological, physiological and behavioral disparity between closely related species and between the sexes. It has been shown that these divergences are due in large part to differential regulation of gene expression during development (reviewed in West-Eberhard 2005; Carroll 2005; Carroll 2008). The contrast between the sexes is particularly striking given the near identical nature of their genomes. This means then that the sexually dimorphic traits observed in nature are also the result of differential regulation of expression patterns unique to one sex over the other during development, or so-called sex-biased gene expression (reviewed in Williams and Carroll 2009).

The most obvious sexually dimorphic features are often exhibited in the adult organism. For this reason, a majority of sex-biased gene expression studies have been limited to comparisons between adult males and females (Ranz et al. 2003; Jiang and Machado 2009; Assis et al. 2012; Naurin et al. 2011; Pointer et al. 2013). However, these are just snapshots in time of a particular gene's status of sex-bias. Very few studies have been conducted to get a full picture of the developmental trajectory of sex-biased genes (Mank et al. 2010; Magnusson et al. 2011; Zhao et al. 2011; Perry et al. 2014). As a result, the dynamics of sex-biased gene expression throughout an organism's life cycle are not well understood. For example, how early in development does sex-biased gene expression manifest itself? Once sex-bias is acquired, does a gene remain sex-biased throughout life? Or do switches in sex-bias occur within an organism's lifespan? Additionally, because no such studies have been conducted between even closely related species, it is difficult to say if observed patterns of sex-biased gene expression throughout development are species-specific or if they are evolutionarily conserved. Thus, an

exploration of global dynamics of sex-biased gene expression throughout development is critical to furthering our understanding of sex-biased gene expression evolution.

Genes involved in generating sexually dimorphic traits and other complex phenotypes do not act alone but rather within pathways that mediate critical developmental, physiological and organismal processes. Changes in the activity of gene pathways are therefore responsible for generating changes to the phenotype during an organism's life cycle and for generating changes that differentiate between the sexes and species. Thus, understanding the general properties of gene pathways is of fundamental importance to biology. However little is known about how even well characterized pathways differ between individuals or between species.

An ideal pathway to study the evolution of such changes would be well characterized both at the biochemical and genetic levels. The phototransduction signal transduction cascade, which is responsible for the visual response in *Drosophila*, is one such pathway. It has been shown that genes within the phototransduction pathway may be modulated at the level of transcription to exploit the different ecological, physiological and behavioral demands of the sexes and different *Drosophila* species (Ranz et al. 2003; Ma et al. 2006; Landry et al. 2007). However, it is not known if those changes at the gene expression level correspond to changes at the level of protein coding sequence. Outstanding and fundamental questions include: How do the collective genes as a pathway evolve over time? What are the selective forces acting on the genes of a pathway and how are they distributed from a small to large evolutionary scale? Is there coordinated evolution among the interacting partners of a gene pathway?

Of course, divergence of the individual components of pathways can occur at levels beyond differences in transcription and protein sequence change. It is expected that as evolutionary distance increases, changes to the topology including the distribution (number and type) of genes within a pathway can occur. Gene loss and gene duplication events occur over time and often in a lineage-specific manner. Duplicate genes are thought to be crucial to the growth and elaboration of gene pathways over evolutionary time and are subject to distinct modes of evolution (Ohno 1970; Force et al. 1999; Lynch and Connery 2000). Moreover, gene duplications can gain novel expression patterns including sex-biased gene expression and are thought to play a role in generating phenotypic diversity (Force et al. 1999). An investigation of duplicate gene evolution in the context of a gene pathway requires a fine-scale analysis to establish the evolutionary forces acting on them.

Broadly, this dissertation is a comparative genomic analysis of patterns of sex-biased gene expression and protein evolution across multiple species of *Drosophila.* In chapter 1, I examine patterns of sex-biased gene expression during ontogeny in two closely related *Drosophila* species belonging to the *D. pseudoobscura* group (*D. pseudoobscura* and *D. persimilis*). Not much is known about sex-biased gene expression evolution throughout development and this study provides insight into global patterns of sex-bias throughout development between species.

The phototransduction visual pathway has shown evidence of evolutionary change at the level of gene expression both between species and between sexes (Ranz et al. 2003; Ma et al. 2006; Landry et al. 2007). In order to determine if observed divergences in gene expression are the result of changes in protein coding sequence, in

chapter 2, I examine the rate of protein evolution for all 35 genes in the pathway across the 12 sequenced species to establish the strength and type of selection (purifying, positive or neutral evolution) on each member gene in order to determine to what degree the proteins have retained their ancestral functions relative to *D. melanogaster*.

In chapter 3, using experimental and computational approaches, I examine the evolution of the gene structure and expression of a novel lineage-specific duplication event that produces descendent duplicate copies of an alternatively spliced ancestral gene. This analysis revealed that the gene duplicates have undergone a complex mode of evolution that would suggest a mechanism by which duplicate genes can evolve novel functions and expression patterns while retaining their ancestral functions.

CHAPTER 1

Sex-biased gene expression evolution throughout development in the *Drosophila pseudoobscura* group

I. INTRODUCTION

Sexual dimorphism, the phenotypic differences between males and females of the same species, is pervasive in nature. While the phenotypic diversity between species is often simply understood in the context of natural selection and drift, disentangling how natural selection, drift, and sexual selection contribute to phenotypic diversity between sexes is usually much more difficult. Since Darwin's time, understanding the mechanisms that underlie sexually dimorphic phenotypes has been a central theme in evolutionary biology.

In most species, with the exception of sex chromosomes, males and females of the same species have almost identical genomes. Yet, in many cases, males and females of the same species differ substantially in morphology, physiology and behavior. In addition, as a result of differences in reproductive-related interests and strategies, males and females typically face distinct evolutionary pressures (Chippindale et al. 2001; Rice 1984). Traits advantageous to one sex may be disadvantageous to the other, leading to genomic conflicts between the sexes known as sexual antagonism (reviewed in Cox and Calsbeek, 2009).

Sexual conflict at loci can be resolved through sex-biased gene expression. Genes are considered sex-biased if they exhibit statistically significant differential expression patterns between the sexes. Sex-biased genes are classified into three categories based on

their direction of sex-bias – genes more highly expressed in males are male-biased, genes more highly expressed in females are female-biased and genes that are equally expressed in both males and females are unbiased. While sex-biased gene expression has been extensively studied and characterized across many phylogenetically distinct taxa, (Ranz et al. 2003; Marinotti et al. 2006; Yang et al. 2006; Mank et al. 2007; Jiang and Machado 2009; Hale et al. 2011; Ayers et al. 2013; Martins et al. 2013; Trabzuni et al. 2013; Vicoso et al. 2013; Gossmann et al. 2014; Smith et al. 2014; Stuglik et al. 2014) much more remains unknown about the evolution and mechanisms of sex-biased gene expression changes throughout an organism's life-cycle.

In many species, sexually dimorphic traits are observed later in development; most especially after sexual maturation. As such, most analyses of sex-biased gene expression evolution have focused on adults (Ranz et al. 2003; Jiang and Machado 2009; Assis et al. 2012; Naurin et al. 2011; Pointer et al. 2013) and consequently not much is known about sex-biased gene expression throughout development. As of now, very few analyses examining sex-biased gene expression dynamics across multiple stages of development have been conducted (Mank et al. 2010; Magnusson et al. 2011; Zhao et al. 2011; Perry et al. 2014). Moreover, the studies that have been conducted thus far have been conducted on a single species and as a result it is not fully understood if observed sex-biased expression patterns during development are species-specific or evolutionarily conserved. Remaining questions include: at what stage in development do we first see evidence of sex-biased gene expression? Do sex-biased genes remain sex-biased throughout development? Are these patterns conserved between closely related species?

Using a comprehensive comparative approach, we aimed to answer these questions with a development time-course transcriptome profiling dataset of males and females of two closely related *Drosophila* species, *D. pseudoobscura* and *D. persimilis*. These species are partially sympatric species in the western part of North America (Dobzhansky and Epling 1944). They diverged from one another ~0.5 MYA and hybridize at low frequencies in the wild (Dobzhansky 1973; Powell 1983; Schaeffer and Miller 1991; Wang et al. 1997; Machado et al. 2002; Machado and Hey 2003). Both species belong to the obscura group in the greater Sophophora subgenus within the *Drosophila* phylogeny and last shared a common ancestor with *D. melanogaster* ~25-40 MYA (Singh et al. 2009). Thus far, sex-biased developmental transcriptome studies in *Drosophila* have solely been conducted on *D. melanogaster*. Our analyses here are unique in that they are the first to examine sex-biased gene expression evolution throughout development in a comparative manner and provide insights into the developmental transcriptome dynamics of two other *Drosophila* species.

II. RESULTS

Custom Agilent oligonucleotide arrays were designed to capture gene expression profiles from four developmental time-points; two larval stages (L1 and L3), a pupal stage and 7 day old adult virgins from males and females of *D. pseudoobscura* and *D. persimilis*. After data filtering, a total of 15,473 predicted protein-coding genes were obtained representing 92% and 93% of the annotated *D. pseudoobscura* (16,787 CDS in release v. 2.29) and *D. persimilis* genomes, respectively (16,878 CDS in release v 1.3).

i. Sex-Biased Gene Expression Increases as a Function of Age in the *D. pseudoobscura* group.

The phenotypic features that differentiate the sexes are most prominent after sexual maturation, most especially during adulthood. Thus, greater numbers of sex-biased genes are expected as the reproductive interests of males and females diverge.

To determine if the number of sex-biased genes increases as development progresses, we performed differential expression analysis for each stage of development between the sexes. Using a sex-bias criterion of fold difference ≥ 2 and $P_{adj} < 0.05$, we observe a steady increase in the number of genes exhibiting sex-biased expression as development progresses with the greatest level of sex-bias occurring in the adult organism (Table 1-1 and Fig 1-1). Interestingly, in both species, we observe more male-biased genes in the L3 and pupal stages of development and more female-biased genes at adulthood ($P = 0.03$; χ^2_1) (Table 1-1 and Fig 1-1). This is consistent with reports that male pre-gonadal tissue in L3 and pre-pupal stages are dominated by primary spermatocytes, structures that synthesize the RNA necessary for spermatid development (Bodenstein 1950). In contrast, the complementary structures in females, the oocytes, are not present yet in pre-gonadal tissue of both L3 and the pre-pupal stages (Bodenstein 1950). Our observation of more female-biased genes in the adult stages of both species is also consistent with previous reports that also observed higher proportions of female-biased genes in comparison to male-biased genes in *D. pseudoobscura* adults (Zhang et al. 2007; Jiang and Machado 2009; Assis et al. 2012).

To determine if patterns of sex-bias are generally conserved between species, we identified genes that were commonly female-biased or male-biased in both *D.*

pseudoobscura and *D. persimilis* throughout each stage of development. Here again, we found that the fraction of genes commonly sex-biased (both male-biased and female-biased) between species increases also as a function of ontogenetic time with the greatest overlap in number of genes in both males and females occurring in the adult stages (Table 1-2). In sum, patterns of sex-bias throughout development appear to be consistent in both species suggesting that an increase in sex-bias throughout development is conserved within the *D. pseudoobscura* group.

Gene ontology (GO) analyses to determine functional overrepresentation of sex-biased genes revealed that female-biased genes for both species were enriched for biological processes including neurogenesis, mitotic spindle organization and elongation as well as female-specific processes including eggshell chorion gene amplification and oogenesis ($P_{adj} < 0.05$). Male-biased *D. pseudoobscura* and *D. persimilis* genes were enriched for those involved in oxidation-reduction process, phototransduction, humoral immune response, sperm motility and spermatogenesis ($P_{adj} < 0.05$) (Table A1, Appendix I).

Table 1-1: Number of Sex-biased Genes Throughout Development

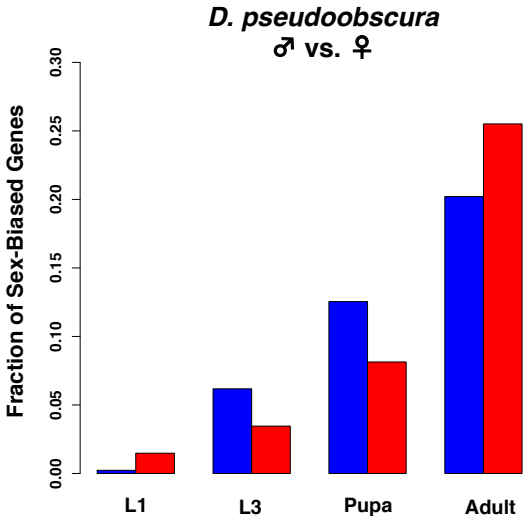
Stage	*D.pseudoobscura*			*D. persimilis*		
	MB	**FB**	**UB**	**MB**	**FB**	**UB**
L1	36 (0.2%)	229 (1.4%)	15208 (98.3%)	63 (0.4%)	29 (0.2%)	15381 (99.4%)
L3	956 (6.2%)	535 (3.5%)	13982 (90.3%)	1016 (6.6%)	588 (3.8%)	13869 (89.6%)
Pupa	1941 (12.5%)	1259 (8.1%)	12273 (82.6%)	1215 (7.9%)	270 (1.7%)	13988 (90.4%)
Adult	3127 (20.2%)	3947 (25.5%)	8399 (54.3%)	3168 (20.5%)	3665 (23.7%)	8640 (55.8%)

Shown are the raw number and percentage of total genes that were differentially expressed between the sexes at each stage using the criterion of fold difference ≥ 2 and $P_{adj} < 0.05$. MB= male-biased, FB= female-biased, UB= un-biased.

Table 1-2: The number of sex-biased genes that were commonly sex-biased in both species at each stage.

Developmental Stage	Male-Biased	Female-Biased
L1	2	7
L3	361	19
Pupa	938	161
Adult	2770	3376

A



B

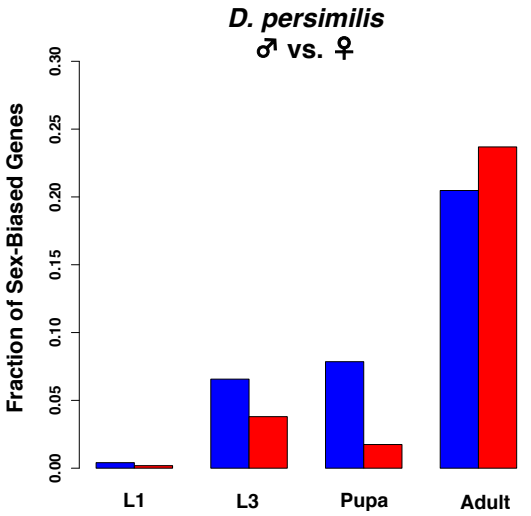


Figure 1-1. Fraction of sex-biased genes at each stage of development in *D. pseudoobscura* and *D. persimilis*
Histogram representation of the distribution of sex-biased genes at each stage of development in *D. pseudoobscura* (A) and *D. persimilis* (B).
Genes were considered differentially expressed between the sexes if they exhibited a fold difference ≥ 2 and $P_{adj} < 0.05$.
Red = female-biased, Blue = male-biased.

ii. The Dynamic Nature of Sex-Biased Gene Expression Throughout Development

Patterns of sex-biased gene expression are dynamic; gains, losses and switches (male-biased to female-biased or vice versa) of sex-bias have been observed in comparisons of adults of closely related *Drosophila* species (Ranz et al. 2003; Zhang et al. 2007; Jiang and Machado 2009; Assis et al. 2012). However it is not fully understood if sex-bias is as dynamic within an organism's life cycle – that is do sex-biased genes remain sex-biased throughout all stages of development?

While we observed an increase in sex-bias (Fig. 1-1) as the organisms develop towards sexual maturation, we did not observe a large set of genes that remain either female-biased or male-biased from L1 to adulthood in the males and females of both species (Table 1-3).

Table 1-3: Genes that maintain sex-bias throughout all the stages of development (L1-Adult)

Species	Gene Name (*D. mel.* orthologs)	Chromosomal/Scaffold Location (in *D. pse* and *D. per*)
D. pse females	*Sxl*	XL
	CG11382	XL
D. pse males	*Ant2,*	XL
	CG7946-PA	XL
	e(r)-PA	4
	can-PA	4
	CG4415	4
	ctp-PB	4
	CG6650	XL
D. per females	*Sxl*	scaffold_25
	Dper\GL20534	scaffold_25
D. per males	*CG4415*	scaffold_8
	Dper\GL26827	scaffold_25

In both *D. pseudoobscura* and *D. persimilis* females, exactly two genes maintained female-bias throughout all stages of development (Table 1-3). One of those genes is the ortholog of the *D. melanogaster* master switch gene for somatic sex - determination *Sex-lethal* (*Sxl*). *Sxl* is actively expressed throughout the life cycle of females, as it is required for proper female sexual development (reviewed in Penalva and Sanchez, 2003). In addition to *Sxl*, in *D. pseudoobscura* females, the ortholog for the *D. melanogaster* gene, *CG11382,* maintains female-bias throughout development. In *D. melanogaster*, *CG11382* is an un-annotated gene that is highly expressed in larval and pupal stages and is putatively involved in imaginal disc-derived wing morphogenesis. Genes involved in imaginal disc-derived wing morphogenesis were significantly enriched ($P_{adj} < 0.05$) in our GO analyses of female-biased genes. As there is no female-specific gene expression data currently available for *D. melanogaster* at the larval stages, it is not possible to say at this time whether this gene is sex-biased or un-biased in *D. melanogaster* at this stage. Interestingly, this same gene is female-biased only in the pupal stage in *D. persimilis* females and remains un-biased in the other stages suggesting a disparity in sex-bias patterns between even these closely related species.

In both *D. persimilis* and *D. pseudoobscura* males, the *D. melanogaster* ortholog, *CG4415*, retains male bias throughout development (Table 1-3). Tissue-specific expression data in *D. melanogaster* generated by the modENCODE project (modENCODE Consortium, 2010) indicate this gene is highly expressed in both L3 wandering larvae and adult testes and moderately expressed in adult ovaries. Since the gene is expressed at moderate to high levels in both adult gonads, it is unlikely that it plays a role in either female or male gonad-specific functions in *D. melanogaster*. This

suggests that this gene may have gained a novel male-biased or male-specific function in the ancestor of *D. pseudoobscura* and *D. persimilis*.

In both *D. persimillis* males and females, one of the two genes that maintained sex-bias throughout development was found only within the *D. pseudoobscura* group. The *Dper\GL20534* gene in *D. persimilis* females and the *Dper\GL26827* gene in *D. persimilis* males have no orthologs in *D. melanogaster*. Also in both cases, the same gene gains sex-bias expression at the L3 stage and remains sex-biased until adulthood in *D. pseudoobscura.* While we do not rule out the possibility of annotation errors, genes that are typically sex-biased are fast evolving at the sequence level in comparison to unbiased genes and their orthologs are often difficult to identify within the genomes of divergent species.

The raw number of genes that remain sex-biased throughout development (from L1- Adult) is highly dependent on the statistical framework used to define sex-bias (in our case fold difference ≥ 2 and $P_{adj} < 0.05$). However, even under a more relaxed measure of sex-bias (no fold difference, just a significant adjusted *P*-value cutoff of $P_{adj} < 0.001$), we find that only ~1.5% of sex-biased genes (for example: 154 out of a total 10,537 female-biased genes in *D. pseudoobscura* and 141 out of a total 9913 female-biased genes in *D. persimilis*) remain sex-biased throughout development.

In addition, the low numbers of genes that retain sex-bias from the L1 to the adult stage is likely the result of the limited number of genes that exhibit sex-biased gene expression during the L1 stage. More genes remain sex-biased from the L3 to adult stage (812 male-biased genes and 11 female-biased genes in *D. pseudoobscura* and 383 male-

biased genes and 17 female-biased genes in *D. persimilis* ($\chi^2_{1=,}$ P= 2.26 × 10^{-16}). The significant difference in the retention of sex-bias from the L3 to adult stages observed between males and females of both species is likely due to the delay in gametogenesis in females (Bodenstein, 1950). Taken together, these results suggest that most sex-biased genes do not maintain sex-bias throughout development; they are subject to gains, losses and possible switches in sex-bias.

Genes that switch in their directionality of sex-bias (i.e. from male to female-bias or vice versa) are of particular interest. In the context of development, these genes suggest shifts in the direction of sexually antagonistic selection during ontogeny. We found significantly fewer gene switches from male to female bias than from female to male bias in both species (48/3553 M-F vs. 240/5560 F-M in *D. pseudoobscura* ,χ^2_1, P=2.73 ×10^{-14} and 29/4060 M-F vs. 113/4257 F-M in *D. persimilis*, $\chi^2_{1=}$, P= 3.38 × 10^{-11} Table A2, Appendix I). GO term analyses indicate the genes that switch from female to male are enriched for processes including the oxidation-reduction process, phototransduction and metabolism (P < 0.003). Genes that switched from male-to-female bias were enriched for humoral immune response, mitotic spindle organization, neurogenesis, and oocyte maturation (P < 0.007) (Table A2, Appendix I). An examination of the overlap of genes that switched their sex-bias direction in both species revealed 39 genes that switched from female to male-biased and only 5 genes that switched from male to female-biased (Table A3, Appendix I). In sum, these results suggest that sex-biased gene expression is labile during development; marked by gains, losses and switches throughout the organism's life cycle.

iii. Patterns of Sex-Biased Gene Expression Throughout Development

Previously conducted sex-biased transcriptome studies in *Drosophila* have suggested that anywhere from ~50-80% of genes show evidence of sex-biased expression (Ranz et al. 2003; Jiang and Machado 2009; Assis et al. 2012). This suggests that a substantial number of genes between the sexes exhibit divergent expression patterns between the sexes. Since the majority of studies have been conducted on adult stages, it is not clear if this is also the case in pre-adult stages. How diverged are the expression profiles of orthologous genes between the sexes throughout development?

One way to measure the similarity in expression profiles between two genes is to compute Pearson's correlation (*r*) (Huminiecki and Wolfe 2004; Bhardwaj and Lu, 2005;Yang et al. 2005). This measure allows for the quantitative determination of conservation between two gene expression profiles. Genes with divergent expression profiles throughout development will have a *r* of -1 (or close to -1), genes with no relationship have a *r* of 0 (or close to 0) and genes that have conserved expression profiles throughout development have a *r* of 1 (or close to 1). Using Pearson's correlation, we compared expression profiles of orthologous genes throughout development between the sexes in both *D. pseudoobscura* and *D. persimilis* to determine the fraction of orthologs between the sexes that have conserved gene expression profiles throughout development. We expected that un-biased genes would have the most positively correlated expression profiles and genes exhibiting sex-bias should demonstrate the most divergent expression profiles (they should be negatively correlated).

We find that more than half of orthologous genes (~59% between *D.*

pseudoobscura sexes and ~65% between *D. persimilis* sexes) have positively correlated expression profiles ($r > 0.5$) throughout development and ~21% (*D. pseudoobscura* sexes) and 18% (*D. persimilis* sexes) have negatively correlated expression profiles ($r < 0$) (Table 1-4 and Figure 1-2). Of the genes with r > 0.5, there were more genes with near identical expression profiles ($r > 0.8$) between *D. persimilis* sexes than *D. pseudoobscura* sexes (8,113/15,473 compared to 6,977/15,473 , χ^2_1, $P < 2.2\times10^{-16}$). These results suggest that the vast of majority of orthologs between the sexes have very similar expression profiles throughout development while ~20% have divergent expression profiles and another ~20% have low to no correlation.

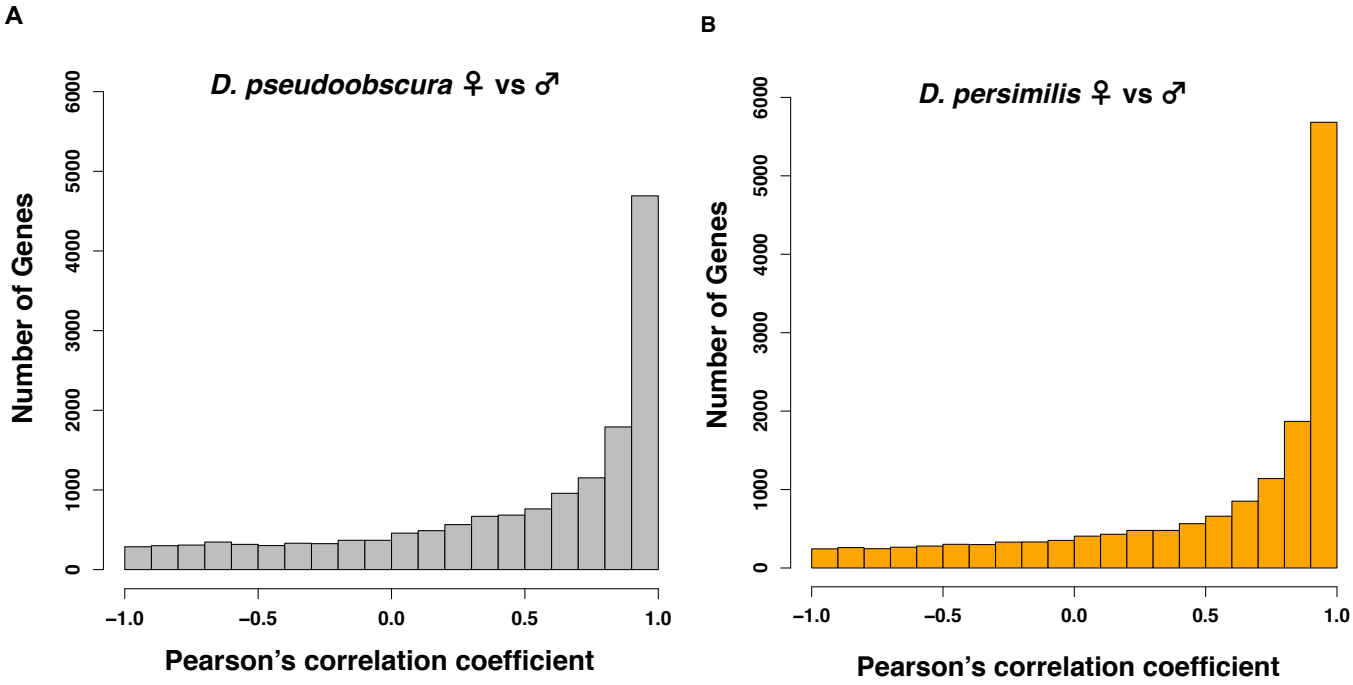


Figure 1-2: **Comparative analysis of gene expression profiles between sexes throughout development**

Gene expression profiles throughout development were compared between the sexes in both *D. pseudoobscura* and *D. persimilis*. Histograms depicting the distribution of Pearson's correlation coefficients shown here. (A) Frequencies of Pearson's correlation coefficients (r) in comparison between *D. pseudoobscura* males and females (B) Frequencies of Pearson's correlation coefficients (r) in comparison between *D. persimilis* males and females.

We expected that the orthologous genes with divergent expression profiles (negatively correlated expression profiles, $r < 0$) between the sexes would be enriched for sex-biased genes and unbiased genes would mostly have low-to-no correlations ($0<r<0.5$) or would be highly correlated ($r<0.5$). Interestingly, we found that genes with low to no correlation in gene expression profiles were almost 50% (1060/2865) male-biased and almost 50% (1283/2865) female-biased and the remainder were unbiased. More than half of genes with positively correlated expression profiles were unbiased (56%, 5222/9355). Indeed, we find that in *D. pseudoobscura*, 87% (2,857/3,253) of orthologous genes that have negatively correlated expression profiles between the sexes exhibit sex-biased gene expression. Of those, more than 60% (1934/2857) are female-biased. Accordingly, genes with negatively correlated expression profiles were enriched for the same functions previously observed for female-biased genes, namely oogenesis and neurogenesis ($P < 0.05$). This was consistent and also observed between *D. persimilis* females and males.

Table 1.4: Number of genes that have Pearson's correlation coefficients within each threshold in comparisons between *D. pseudoobscura* males and females and *D. persimilis* males and females

Comparison	$r<0$	$0<r<0.5$	$r>0.5$
***D. pse* M v F**	3253	2865	9,355
***D. per* M v F**	2,910	3,538	10,204

An arbitrary Pearson's correlation coefficient threshold was defined to divide the genes according to their level of conservation of expression profiles between the sexes. Highly positive correlations ($r > 0.5$), low-to-no correlations ($0<r<0.5$) negative correlations ($r <0$).
M = males, F = females.

While these results suggest that nearly 20% of orthologs between males and females in both species have divergent expression profiles throughout development, these results do not say at what stage in development the greatest divergence occurs or when in development the divergence first becomes apparent. To determine at which stage in development divergence in expression between the sexes is first observed, we examined the correlation of expression of orthologs at each individual stage. We find that orthologous genes have very highly correlated expression patterns at the earliest L1 stage (R^2 = 0.97 for *D. pseudoobscura* and R^2 = 0.99 for *D. persimilis*) (Figure 1-3). However, we note a gradual increase in expression divergence with the greatest divergence occurring between the sexes at the adult stage (R^2 = 0.27 for *D. pseudoobscura* and R^2 = 0.32 for *D. persimilis*), consistent with our expectation based on greater sexual dimorphism later in life.

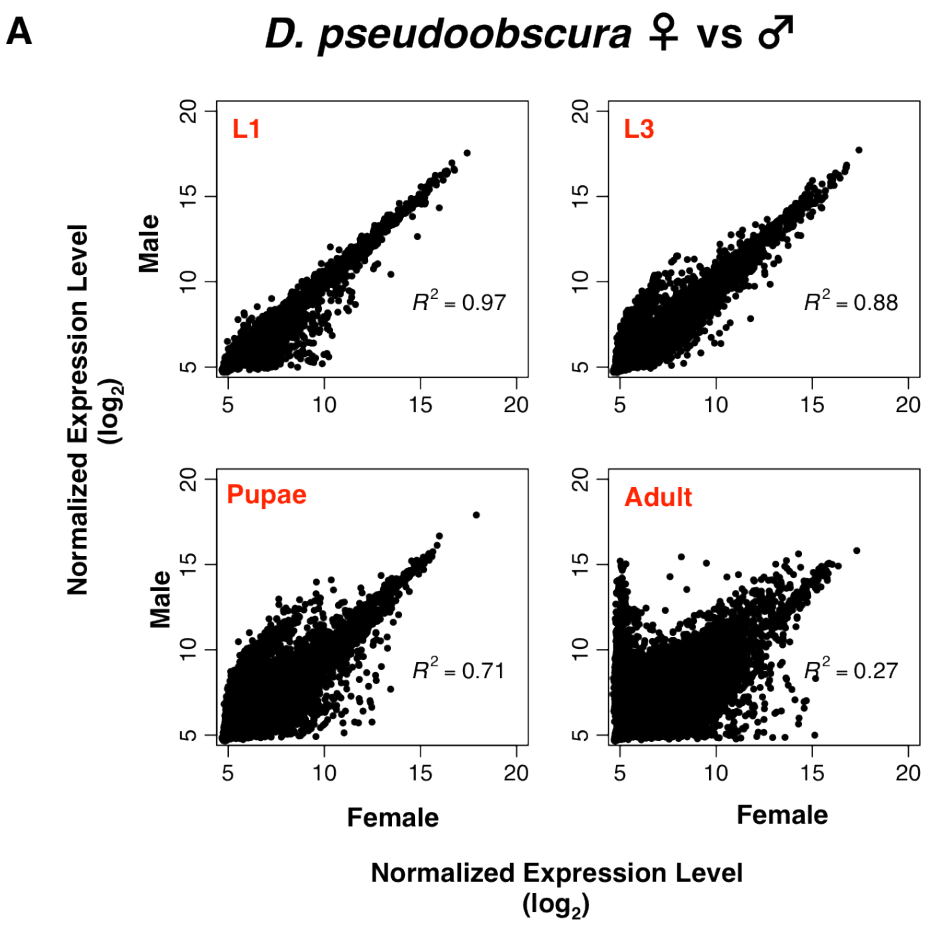


Figure 1-3.
Gene expression correlation at each stage between sexes

Log_2 normalized expression values for orthologs were compared at each stage between males and females

(A) Comparison between *D. pseudoobscura* males and females
(B) – (on next page)
D. persimilis males and females

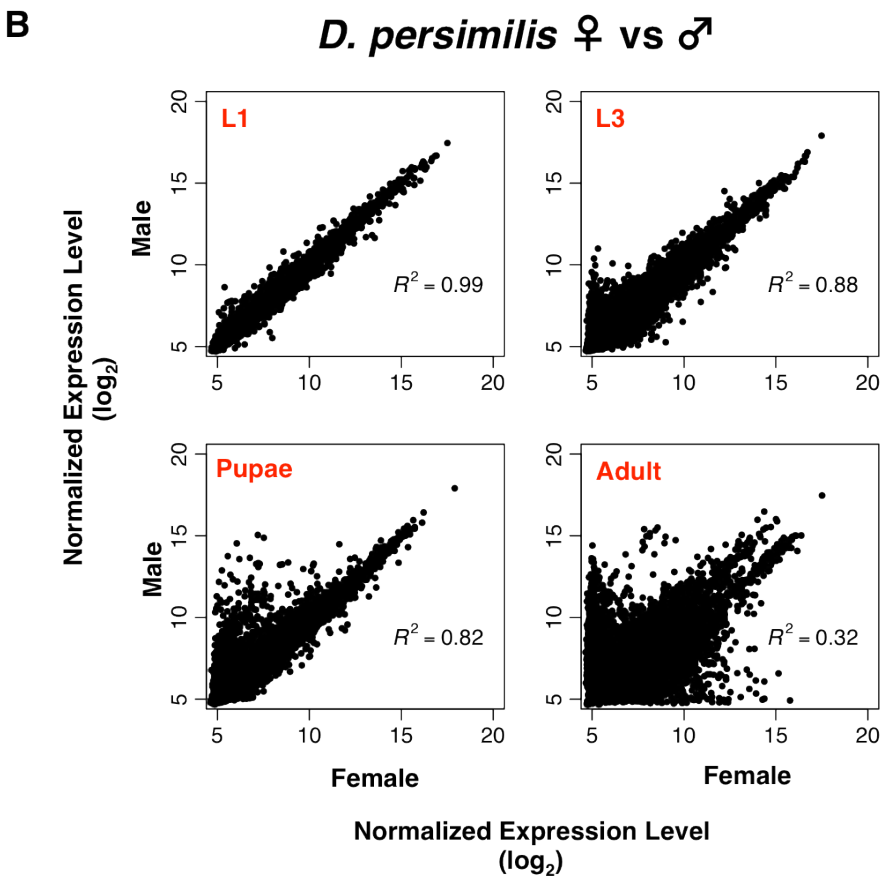


Figure 1-3. (continued) Gene expression correlation at each stage between sexes

Log_2 normalized expression values for orthologs were compared at each stage between males and females

(B) *D. persimilis* males and females

iv. Sex-biased gene expression evolution during development

A hallmark of male-biased genes is that they are typically fast evolving both at the level of gene expression and at the level of coding sequence (Meiklejohn et al. 2003; Zhang et al. 2004). As a result of the great deal of sequence divergence, orthologs of male-biased genes are often difficult to identify within the genomes of distantly related species. To determine if male-biased genes have fewer matches in the *D. melanogaster* genome compared to female-biased and un-biased genes, we performed a reciprocal blast search for male-biased, female-biased and unbiased genes against the *D. melanogaster* genome. We were only able to identify orthologs for ~ 50% of male-biased genes compared to ~79% for female-biased genes and 97% of un-biased genes. This difference

is statistically significant by χ^2 test ($P < 2.2 \times 10^{-16}$). The male-biased genes for which we could identify *D. melanogaster* orthologs were enriched for cellular maintenance, metabolic processes as well as male-specific functions including sperm competition and regulation of post mating female receptivity ($P_{adj} < 0.03$).

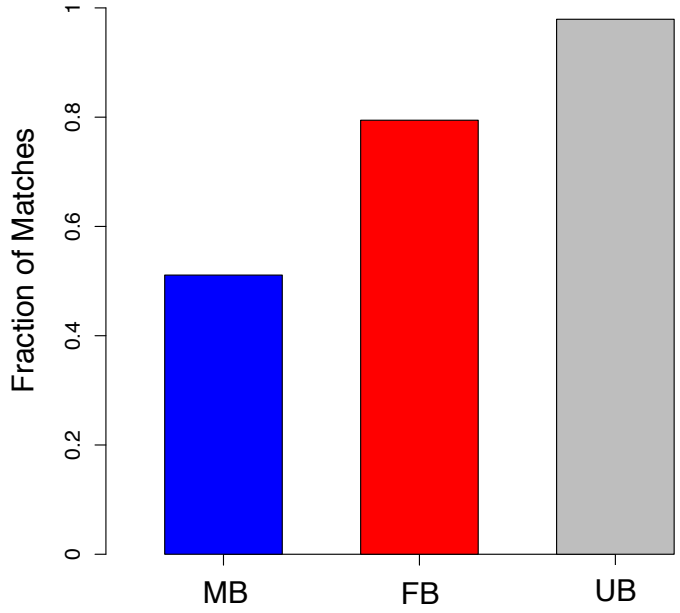


Figure 1-4.
Fraction of sex-biased genes with matches in the *D. melanogaster* genome

tBlastx was performed with an e-value cutoff of 1×10^{-10} to determine genes with matches in the *D. melanogaster* genome

MB = male-biased
FB = female-biased
UB = un-biased

v. De-masculinization of the X-chromosome

It has been shown in a number of *Drosophila* species including *D. pseudoobscura* that sex-biased genes are non-randomly distributed across the genome (Ranz et al. 2003, Parisi et al. 2003, Sturgill et al. 2007, Jiang and Machado, 2009, Meisel et al. 2012, Assis et al. 2012). In particular, a deficit of male-biased genes on the X-chromosome in comparison to female-biased and un-biased genes has been observed. The phenomenon that leads to this deficit of male-biased genes on the X-chromosome has come to be known as the de-masculinization of the X-chromosome.

In *D. pseudoobscura*, an autosome fused to the ancestral X chromosome ~ 8–12

MYA and became the right arm of the X chromosome (XR) or the so-called neo-X chromosome (Steinemann et al. 1984; Richards et al. 2005). Thus in *D. pseudoobscura*, it is expected that the level of de-masculinization is lower in the neo-X compared to the ancestral X (XL) which has had more time to purge male-biased genes compared to the neo-X chromosome. Thus, we would expect to find more male-biased genes on the XR in comparison to the XL.

We compared the proportions of male and female-biased genes for each stage of development on both arms of the X chromosome (XL and XR) of *D. pseudoobscura*. We find more male-biased and female biased genes on autosomes than on either arm of the X-chromosome at each stage of development (χ^2, $P < 2.2\times10^{-16}$) (Figure 1-5). In addition, we observed a statistically significant difference between male and female-biased genes, with significantly fewer male-biased than female-biased genes on both XL and XR compared to female-biased (XL: χ^2, $P = 9.116\times10^{-11}$ and XR: χ^2, $P = 7.093\times10^{-11}$) (Figure 1-6).

Table 1.5: Chromosomal distribution sex-biased genes throughout development

	Male-biased			Female-biased		
	XL	**XR**	**Autosome**	**XL**	**XR**	**Autosome**
L1	7	4	25	35	43	151
L3	116	78	762	107	118	310
Pupae	286	256	1399	227	257	775
Adult	486	421	2214	702	766	2479

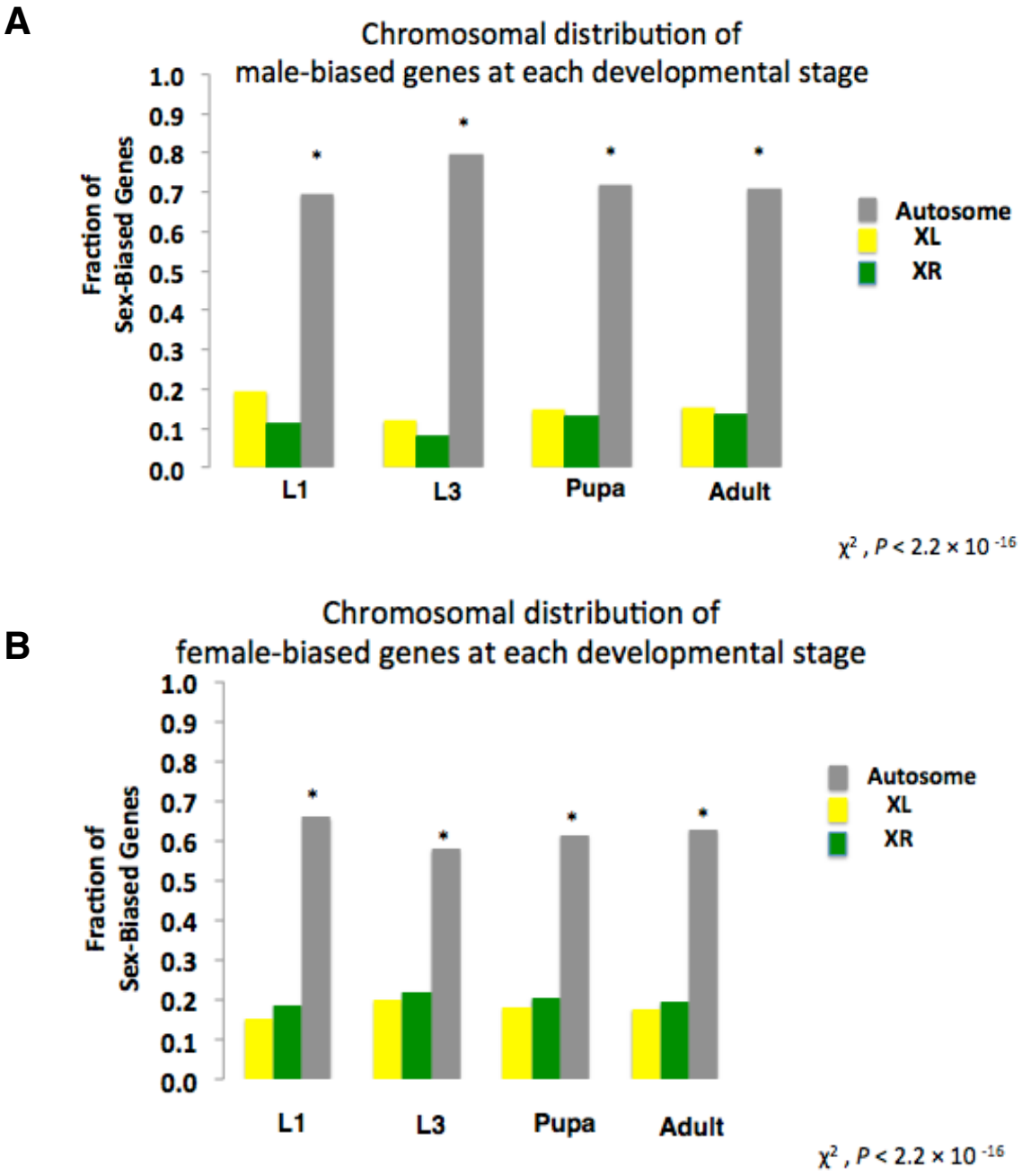


Figure 1.5: Chromosomal distribution of male biased and female-biased genes at each developmental stage.
Distribution of male-biased genes (A) and female-biased genes (B) on autosomes and both arms of the X chromosome at each stage of development.

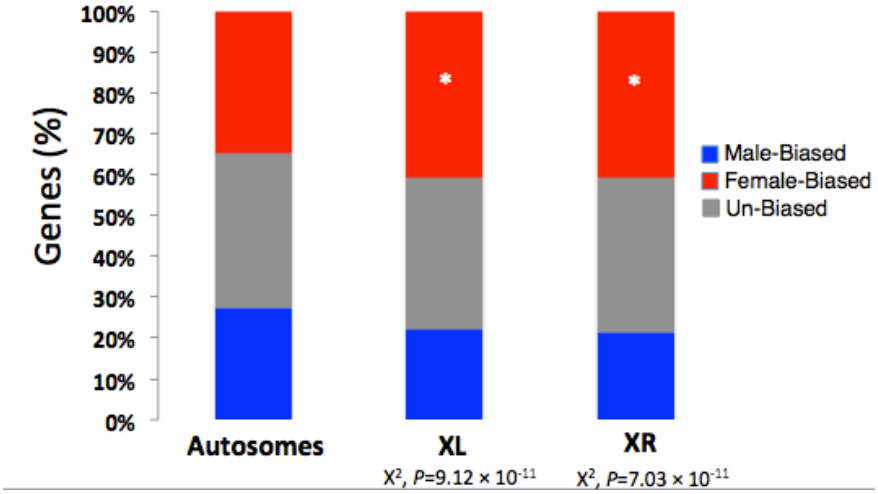


Figure 1-6: Proportion of male-biased and female-biased genes on the arms of the X chromosome
Distribution of male-biased and female-biased genes on both arms of the X chromosome and autosomes controlled for total number of genes on each chromosome.

vi. Developmental Transcriptome Evolution Between Species

Thus far, we have examined gene expression dynamics between the sexes of both species throughout development. With our data, we can also make comparisons between the males of the two species and the females of the two species throughout development to determine if the observed gene expression patterns between the sexes are also exist between the species.

We performed differential expression analyses (fold difference ≥ 2 and $P_{adj} < 0.05$) on the same sex of both species (males of the two species and females of the two species) and observed an increase in the number of differentially expressed genes from the L1 to pupal stage between males of the two species and the females of the two

species (Figure 1.7, Table 1.6,). In both cases, we also observe a decrease in the number of differentially expressed genes in adults compared to those observed in the pupal stage (Figure 1.7, Table 1.6).

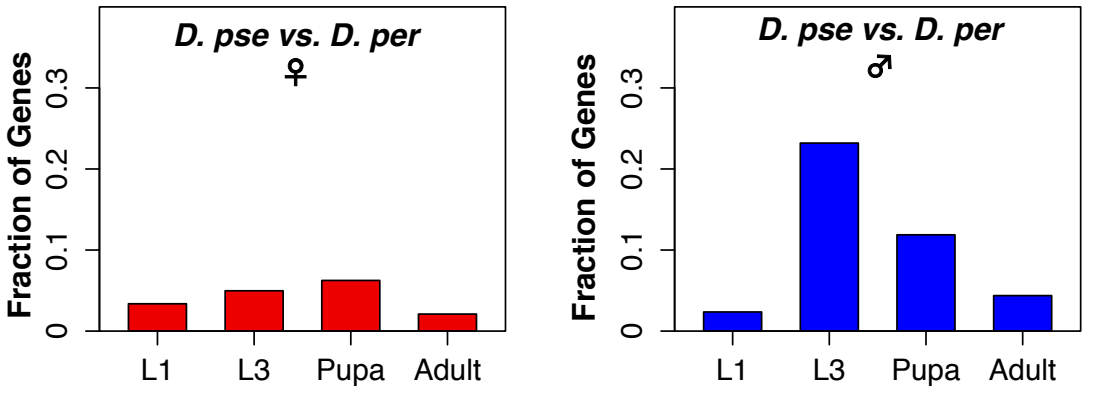


Figure 1-7. Fraction of genes differentially expressed between the males of the two species and the females of the two species.

Genes were considered differentially expressed if they exhibited a fold difference ≥ 2 and $P_{adj} < 0.05$.

Table 1-6.
Number of genes differentially expressed between the males of the two species and between the females of the species at each stage of development

Developmental Stage	*D.pse vs. Dper* Males	*D. pse vs D. per* Females
L1	428	609
L3	4229	896
Pupa	2171	1131
Adult	785	368

Aside from the L3 males, less than 20% of orthologous genes were differentially expressed at each stage of development for both comparisons (Figure 1.7). Most of the differential expression between males of the two species and females of the species occurs in the L3 and pupal stages (Figure 1.7, Table 1.6). This is of particular interest, as the L3 and pupal stages are when gametogenesis takes places in both males and females.

Oogenesis begins in the pupal stage and spermatogenesis begins after the L1. Pre-gonadal tissue of males is dominated by primary spermatocytes by the L3 stage (Bodenstein 1950).

To quantify levels of conservation of the developmental transcriptomes between the two species, we also compared the expression profiles of orthologous genes using Pearson's correlation. We find that expression profiles throughout development are highly conserved between the males of the two species and likewise between the females of the two species. 85% (13,009/15,473) and 87% (13,495/15,473) of orthologous genes between the males of the two species and the females of the two species respectively had positively correlated ($r > 0.5$) transcriptional profiles throughout development (Figure 1.8, Table 1.7). Moreover, in both sexes, of those genes with positively correlated expression profiles, a substantial fraction have near identical expression profiles ($r > 0.8$) throughout development (11,344 out of 13,495 (84%) between females and 10,045 out of 13,009 (77%) between males). These results indicate that a significant proportion of the developmental transcriptomes between the species have conserved transcriptional profiles throughout development suggesting strong conservation of underlying regulatory mechanisms.

Table 1-7. Number of genes in each Pearson's correlation coefficient threshold for comparisons between the males of the two species and the females of the two species.

Comparison	$r < 0$	$0 < r < 0.5$	$r > 0.5$
D.pse vs. *D. per* Female	832	1,146	13,495
D. pse vs. *D. per* Males	870	1,594	13,009

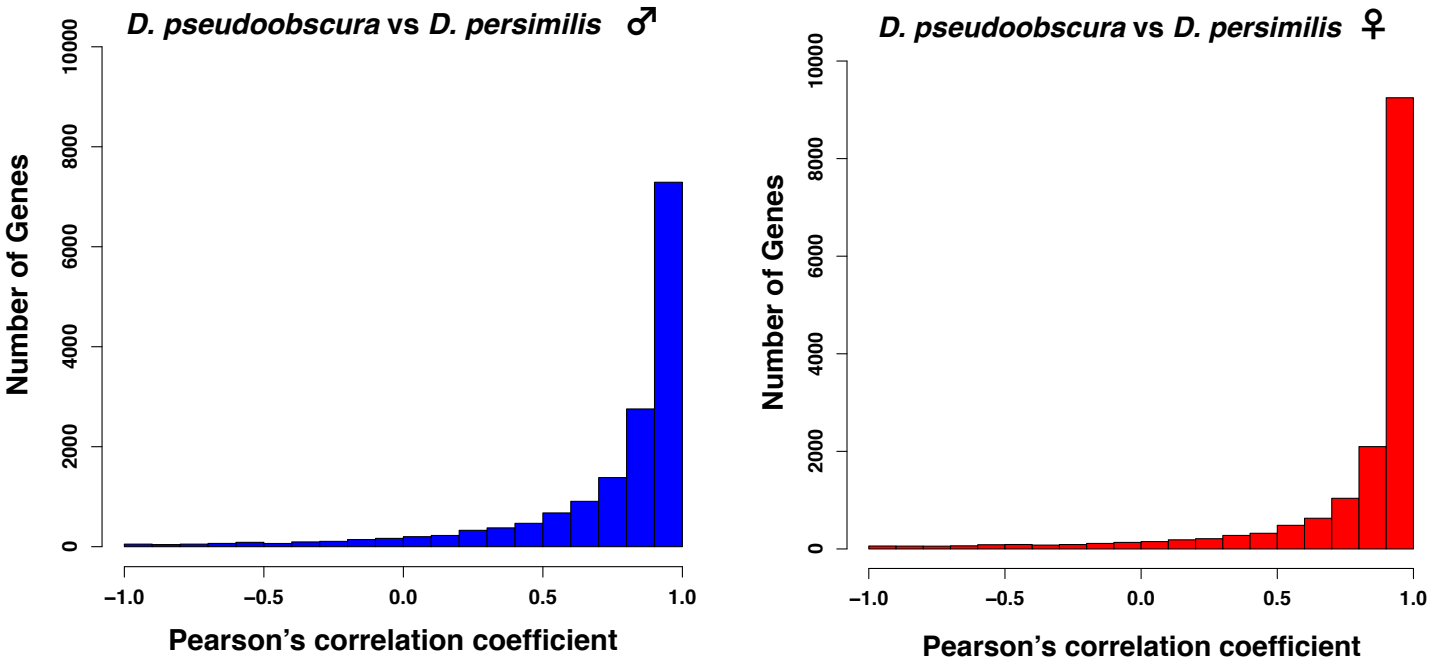


Figure 1-8. Comparative analysis of gene expression profiles between species throughout development.
Gene expression profiles throughout development were compared between the males of *D. pseudoobscura* and *D. persimilis* and females of *D. pseudoobscura* and *D. persimilis*. Histograms depicting the distribution of Pearson's correlation coefficients are shown here. (A) Frequencies of Pearson's correlation coefficients (*r*) in comparison between males of the two species (B) Frequencies of Pearson's correlation coefficients (*r*) in comparison between females of the two species.

To determine when in development we might observe expression divergence between males of the two species and females of the two species, we examined the correlation of expression of orthologs at each individual stage of development. We observed very strong conservation of expression between the females of the two species and the males of the two species at each individual stage of development, all R^2 values were generally greater than 0.8 except in the comparison between males at the L3 stage (R^2 = 0.73) (Figure 1-9). These results are in agreement with our earlier findings of a higher number of differentially expressed genes at the L3 stage between males of the two

species suggesting that there might be species-specific male differences resulting in divergence of gene expression at this stage.

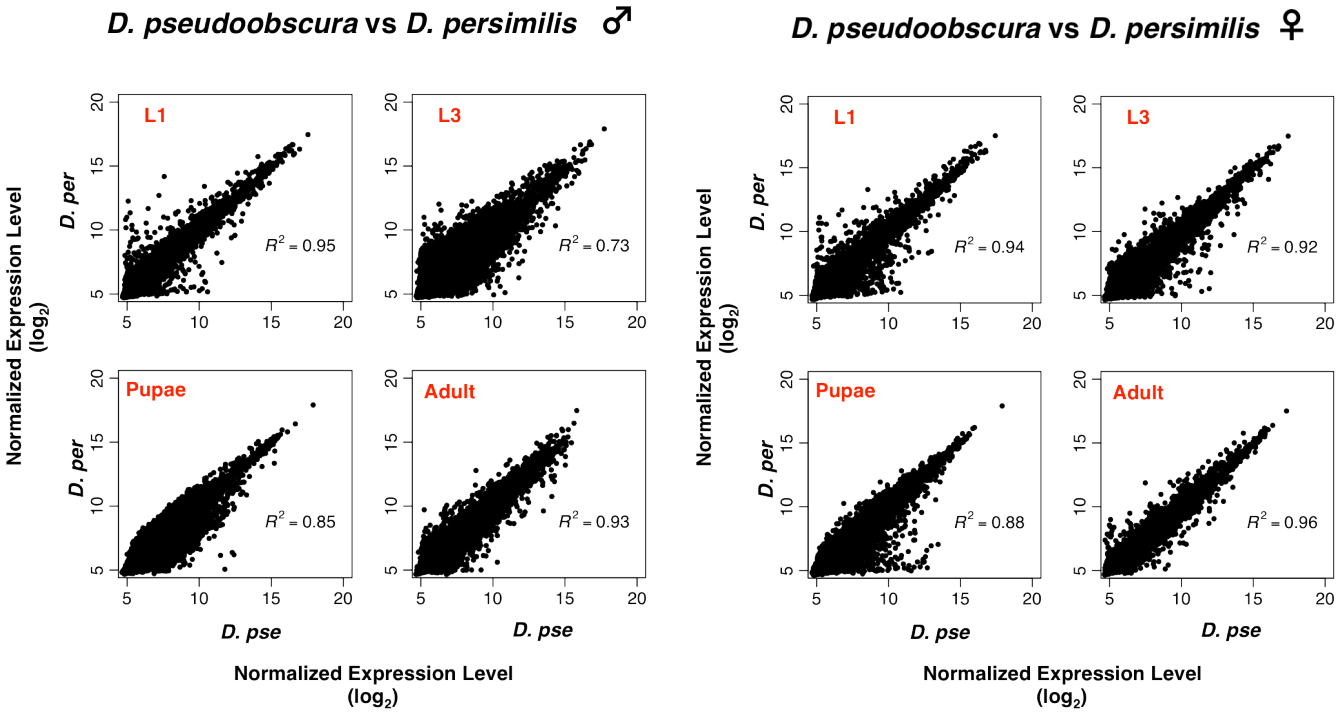


Figure 1-9: Broad conservation of gene expression at each stage between species
Orthologous gene expression was correlated at each stage for males of the two species (A) and females of the two species (B). Gene expression values are $\log_2$ normalized.

vii. The Developmental Transcriptomes of the *D. pseudoobscura* group are more conserved between species than between sexes

In order to get a global picture of developmental transcriptome dynamics, we compared patterns of gene expression between the sexes and the species across all stages of development. We performed hierarchical clustering on a subset of genes (n=6,967) that are commonly expressed at all stages in males and females of both species (Figure 1-10).

First, rather than clustering by species, we observe gene expression profiles

clustering by stage suggesting that developmental stage-specific expression has been conserved between the species. Conservation of developmental stage-specific expression has previously been observed between species belonging to the *D. melanogaster* subgroup (Artieri and Singh, 2010). However, here the inclusion of separate male and female expression profiles for each species allows for a fine-scale determination of sex-by-species gene expression relationships over developmental time.

At our earliest sampled developmental time (L1), expression profiles of males and females of the same species cluster together. This suggests that any sex-biased gene expression at this stage is negligible, as the sexes have very conserved gene expression profiles. Patterns of gene expression between the species are generally very conserved during the L1 stage as well. After L1, beginning at L3 until the adult stage, we observe clustering by sex rather than by species. That is, gene expression profiles of the same sex cluster together from the L3 stage to the adult stage. This suggests that at or by the L3 stage, divergent expression patterns between the sexes become apparent.

Additionally, consistent with our earlier findings, the L3 male gene expression profiles do not cluster together reflecting the earlier observed divergence in expression and comparatively higher fraction of differentially expressed genes between the males of the two species at this stage. Overall, our results indicate that gene expression profiles for the sexes are most conserved at our earliest sampled developmental time-point (L1). Sex-biased gene expression is observed after the L1 stage, from L3 to adult stages in both species. Once sex-biased gene expression becomes apparent, gene expression profiles diverge between the sexes such that there is greater conservation in gene expression profiles between the same sexes of different species than between sexes of the same

species. These results then suggest that sex-biased gene expression plays a tremendous role in the evolution of the developmental transcriptome.

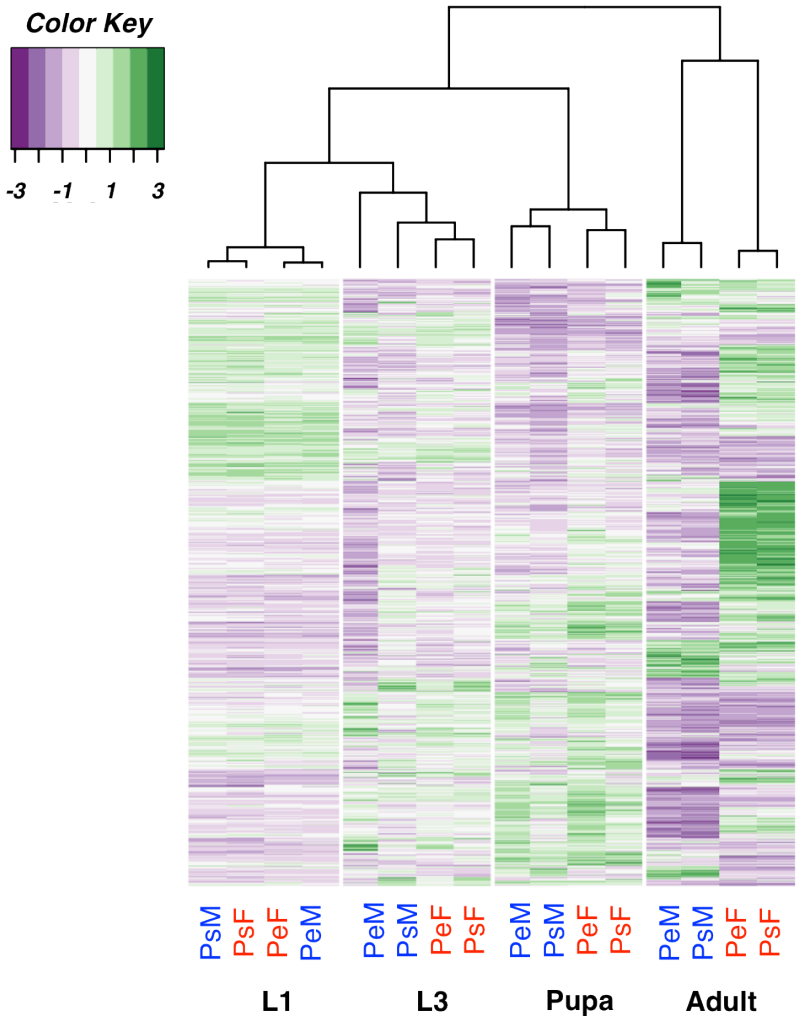


Figure 1-10: Heatmap representation of global gene expression dynamics across development in males and females of *D. pseudoobscura* and *D. persimilis*

Gene expression profiles (n=6967) were hierarchically clustered across samples (sex, species and developmental stage) to determine global developmental transcriptome dynamics. Hierarchical clustering was performed using Spearman's correlation as a distance metric. Clustering relationships were bootstrapped for 2000 replicates and the resulting dendrogram was supported above the 95% confidence interval at each node. Gene expression profiles for males of both species are denoted in blue and and females in red.

III. DISCUSSION

The first developmental transcriptome study of *D. melanogaster* was conducted over a decade ago (Arbeitman et al. 2002). Since then, advances in high-throughput transcriptome profiling technologies have allowed not only for an in-depth analysis of developmental gene expression and regulation but also for the identification of novel

transcripts (both protein-coding and non-coding RNAs) and alternative isoforms from multiple stages of development (Daines et al. 2011; Graveley et al. 2011). While these studies have served to illuminate the complexities of the developmental transcriptome, they have by and large been devoid of sex-specific gene expression data from earlier stages of development. Thus, they have not been able to shed light on the dynamics of sex-bias and sex-specific expression during early *Drosophila* development. While the importance of incorporating sex into *Drosophila* developmental transcriptome analyses is quickly becoming realized (Perry et al. 2014; Perry and Mank, in press), these single-species studies are not comparative in nature and therefore can not inform if observed patterns are evolutionarily conserved.

In this study, using a comparative approach, we examined the developmental transcriptomes of the males and females of *D. pseudoobscura* and *D. persimilis*. This study is to our knowledge the first to explore in a comparative manner how sex-biased gene expression shapes the evolution of the developmental transcriptome in two non-melanogaster *Drosophila* species.

i. Sex-Biased Gene Expression Increases as a Function of Ontogenetic Time

We observed a steady increase in the number of genes exhibiting sex-biased expression as development progresses with sex-biased patterns greatest in the adult stages. In agreement with previously reported estimates for *D. pseudoobscura*, we found more female-biased genes in the adult stage compared to male-biased genes in both species (Figure 1-1) (Zhang et al. 2007; Jiang and Machado 2009; Assis et al. 2012). Additionally, we observed more male-biased genes in the earlier pre-adult L3 and pupal

stages for both species (Figure 1-1). This is also in agreement with a recent study on sex-biased gene expression evolution in pre-gonadal tissue of *D. melanogaster* in the L3 and pre-pupal stages (Perry et al. 2014). Differences in the timing of gametogenesis between males and females (delayed onset of oogenesis in comparison to spermatogenesis which is already active by the L3 stage) are thought to be the cause of the increased frequency of male-biased genes at these earlier stages (Bodenstein 1950). All these patterns are consistent between species suggesting they are evolutionary conserved.

ii. The Transient Nature of Sex-Biased Gene Expression During Development

The trajectory of a gene from early development to adulthood is variable with regards to sex-biased expression. We observed gains, losses and at times switches in sex-bias with very few genes maintaining sex-biased expression patterns throughout all stages of development (Table 1-3). Our results here are in agreement with a study that surveyed sex-biased gene expression over a range of early to late developmental time points in birds (Mank et al. 2010) in which a limited number of genes that maintained sex-bias throughout development was also observed. In addition, consistent with reports in *D. melanogaster* (Perry et al. 2014), we find that more male-biased genes maintained sex-bias throughout development in comparison to female-biased genes (Table 1-3). These results suggest that most sex-biased expression is in fact stage-specific. Moreover, this pattern was consistent between *D. pseudoobscura* and *D. persimilis* suggesting that stage-specific sex-biased gene expression during development is conserved in the *D. pseudoobscura* group.

Genes that switch their status of sex-bias from female to male bias or vice versa during the course of development are of particular interest. Consistent with recent reports in *D. melanogaster* and the silkworm *B. mori* (Zhao et al. 2011; Perry et al. 2014), we observed more female-to male bias switches than vice versa male to female bias switches. Of the genes that switched from male to female-biased, the majority of them switched after the L3 and pupal stages and became female-biased in adults. Our comparative analysis of the overlap of genes that switched from male to female-biased in both *D. pseudoobscura* and *D. persimilis* revealed only 5 genes in common between the two species including the ortholog of *D. melanogaster*, *wispy (*Table A3 Appendix I). In *D. melanogaster*, extensive experimental evidence suggests that *wispy* is required for oogenesis and proper egg development in adult females (Brent et al. 2000; Cui et al. 2008). It is also a maternally deposited gene required for embryogenesis. Consequently, in *D. melanogaster*, this gene is highly expressed in the embryo stage and in the adult ovaries. Since the publically available modEncode data for *D. melanogaster* (modEncode consortium, 2010) does not include gene expression data for males at the L3 and pupal stages, it is not possible to say if this gene exhibits male-biased expression patterns during early development in this species. Examples like the *wispy* gene underscore the importance of incorporating sex-specific and sex-biased gene expression data in developmental transcriptome evolution analyses. Knowledge of a gene's full expression history throughout ontogeny is crucial to furthering our understanding of gene function.

iii. Evolutionary dynamics of male-biased genes

The evolutionary dynamics of male-biased genes have been very well studied (reviewed in Ellegen and Parsch, 2007). Sex-biased genes are fast evolving at the sequence level and it is often difficult to identify orthologous genes between divergent genomes. Consistent with previous studies (Meiklejohn et al. 2003), we show that male-biased genes show faster rates of protein evolution compared to female or un-biased genes (Figure 1-4). More unbiased and female-biased genes have orthologous matches within the *D. melanogaster* genome compared to male-biased genes (Figure 1-4). Furthermore, we observed a deficit of male-biased genes on both arms of the X chromosome (XL and XR) compared to the autosomes (Figure 1-6).

Between the males of the two species, we found the greatest divergence in expression and greatest number of differentially expressed genes occurring at the L3 stage (Figure 1-7). By the L3 stage, primary spermatocytes dominate male pre-gonadal tissue (Bodenstein 1950). It has recently been reported that primary spermatocytes manufacture and store tubulin, a protein necessary for sperm tail formation (Lattao et al. 2012). Differences in sperm tail length have been reported between even closely related *Drosophila* species including between *D. pseudoobscura* and *D. persimilis* (Lattao et al. 2012). While we do not completely rule out technical anomalies relating to developmental timing at this stage, it is possible that the observed differential expression between the males of the two species at the L3 stage might be driven by the development of species-specific sexual traits. Tubulin subunits make up microtubules (reviewed in Mandelkow and Mandelkow, 1995) and a GO analysis of L3 differentially expressed genes between males of the two species shows an enrichment of genes associated with

the microtubule complex (P<0.0004).

This study allowed not only for a comparative examination of sex-biased gene expression dynamics throughout development between *D. pseudoobscura* and *D. persimilis*, but it also allowed for the opportunity to determine if patterns observed between the sexes are also observed between species. As expected we find that gene expression profiles throughout development diverged as the sexes grow towards sexual maturation, with the greatest divergence occurring in the adult stage. This finding was consistent with the steady increase in the number of sex-biased genes throughout development (Figure 1-1). Our results suggest that the greatest proportion of sex-biased gene expression occurs after the L1 stage in *D. pseudoobscura* and *D. persimilis* (Figure 1-10). In contrast, we find that gene expression profiles were significantly more conserved between the species than between the sexes. Given that these species are very closely related, it is perhaps not surprising that expression divergence is very minimal at all stages of development (Figure 1-9). The stages in which we observed slight divergences in expression between males of the two species and females of the two species correspond with times in development where gametogenesis is beginning suggesting that species-specific sexual trait development may be causing the divergence in gene expression at these stages. Taken together, these results show that sex-biased gene expression plays a tremendous role in the evolution of the development.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

Our results here provide further insights into interspecific and sex-biased gene expression evolution during development. Furthermore, they highlight the importance of incorporating sex-bias and sex-specific gene expression analyses into all developmental

transcriptome studies. This is critical as otherwise unknown expression patterns can be determined and the full-scale knowledge of a gene's developmental trajectory gives a better context and further understanding of gene function. Thus far, sex-biased genes have been examined as solitary units, but it is known that many genes working in concert within developmental regulatory networks generate sexually dimorphic traits. An important next step to further our understanding of the genetic basis of sexually dimorphic phenotypes includes determining how sex-biased genes interact within networks during development.

V. MATERIALS & METHODS

i. Drosophila Strains and RNA preparation

At least three isofemale inbred lines were used for both species (*D. pseudoobscura*: MV2-25, Flagstaff 18, Mather 10, Mather 32; *D. persimilis*: MSH 3, Mather G, MSH 42). The MV2- 25 (Baylor) and MSH3 lines are the genome sequence strains of *D. pseudoobscura* and *D. persimilis*, respectively (Clark et al. 2007; Richards et al. 2005). Information about the other lines can be found in Machado et al. (2002). Flies were reared on cornmeal/yeast medium and maintained at 20°C with a 12:12-h light/dark cycle. Flies for each line were collected at four different life stages: 1st instar larvae (24-32 hours after laying eggs), 3rd instar larvae (120-128 hours after laying eggs), yellow pupae (8-16 hours after puparium formation) and adult (7 day old virgin flies). The following protocol was used to genotype the larvae/pupae by sex: samples were directly transferred into a 0.2-mL PCR tube containing 14.5 μL SB buffer (100 mM Tris-Cl, pH 8.2, 1 mM EDTA, and 25 mM NaCl) on ice and homogenized using a pestle. Eleven microliters from the single larvae/pupae extract were stored individually in 96-well plate

wells at -80°C, each containing 40 μL TRIzol reagent (Invitrogen, CA, USA). Proteinase K (Invitrogen, CA) was added to the remaining 3.5μ L extract to a final concentration of 200 μg/mL, and samples were then incubated at 37°C for 30 min, followed by 95°C for 2 min to inactivate proteinase K. After incubation with proteinase K, 1 μL of the extract was used for PCR using the following conditions: 95°C 1 min; 95 °C 15 s, 60 °C 30 s, 65 °C 1 min, 40 cycles; 65 °C 5 min. Primers for 2 Y-linked *D. pseudoobscura* specific genes were used to genotype by sex: (CG12218Y-: forward, GCAGTCGAACCAGTGCAAT; reverse, GTGCGGGCAATGGATAAT) (CG10274Y: forward (F1), CTGTGGCAAGCGGTTCGTG; reverse (R2), CACGTCGCGGATCCTTGGGTA) (Carvalho and Clark, 2005). Ten micro-liters from each PCR were run on a 1.2% agarose gel in TAE buffer [0.5× Tris-acetate-EDTA buffer]. The same sex of each line was pooled for RNA extraction. All RNA was extracted using Trizol. RNA quantity and quality were separately measured on a NanoDrop spectrophotometer (Thermo Scientific) and a Bioanalyzer 2100 (Agilent Technologies). High-quality RNA samples (260/280>1.85; 260/230>1.7) from all lines of each species were pooled in equal amounts and further purified using the RNeasy kit (Qiagen). RNA from multiple lines was pooled to avoid drawing conclusions about interspecific differences based on data from a single isofemale line and to focus on major differences that are likely consistent across multiple lines.

ii. Microarray design and Hybridization

Custom 4-plex 44K oligonucleotide arrays were designed using the Agilent eArray platform and synthesized by Agilent Technologies (Hughes et al. 2001). The array

consists of 45,220 spots including positive and negative controls, with 55-60mer oligonucleotides representing 18,850 unique consensus gene predictions from the *D. pseudoobscura* genome using the GLEAN prediction combiner software (Elsik et al. 2007). The oligos were later re-annotated based on the annotation of the *D. pseudoobscura* genome (version 2.29). Oligos were designed to be identical between the sequenced *D. pseudoobscura* and *D. persimilis* genome lines (MV2-25 (Baylor) and MSH3) to eliminate or reduce possible artifacts in the estimation of gene expression differences between species due to inefficient cross-species hybridizations. Species-specific probes were designed for 1,469 genes for which probes with identical sequences could not be found. An average of 2.17 probes per gene are included in the array. Fluorescent cRNA was synthesized from 200 ng of total RNA using Agilent's low-RNA input fluorescent linear amplification kit following manufacturer's protocols (Agilent Technologies). Labeled RNAs were cleaned with RNeasy columns (Qiagen), and cRNA yields were quantified on a NanoDrop spectrophotometer. Two separate labeling reactions per sample were pooled and hybridized to the arrays following Agilent's protocols. We conducted single color (Cyanine 3-CTP dye) labeling and hybridizations. Three different replicates were hybridized for each species and sex. Adult samples were scanned using the Genepix 4000B scanner, larval and pupal samples were scanned using the Agilent technologies scanner G2505C. Labeling reactions, hybridization and scanning of the data were done at the Genomics Core of the Arizona Cancer Center (The University of Arizona).

iii. Microarray Statistical Analyses

All microarray statistical analyses were conducted using R (Ihaka and Gentleman 1996)

and the Bioconductor package, Limma (Smyth 2004). Raw data from adult and larval/pupal arrays were combined and quantile normalized using the between-array normalization function in Limma. Probes from the combined normalized dataset were filtered to retain those that had significant expression above background in at least 2 of the replicate slides and were 20% brighter than negative controls. Probes for the same gene were averaged using Limma's avereps function. Limma uses a linear model to assess differential expression. Data was fit to the model using a cell means design containing 3 factors (sex, species and developmental stage). Linear model: $Y \sim N(\mu, \sigma^2)$ where Y is expression level for any given gene and μ = stage+ sex+ species. An empirical Bayes analysis was applied to the data to determine species, stage and sex-biased differential expression. Genes were considered differentially expressed between the sexes and between the species if they exhibited a fold difference ≥ 2 and $P_{adj} < 0.05$. Differential *P* values were adjusted for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995).

iv. BLAST analyses

A reciprocal blast search was performed between the *D. melanogaster* genome (v.5.55) and the *D. pseudoobscura* sequences corresponding to the 15,473 genes from our microarray dataset using tBlastx (Altschul et al 1997). Genes were retained if they were the best reciprocal tblastx hits between both *D. melanogaster* and our *D. pseudoobsura* genes exhibiting a *e*-value $< 1 \times 10^{-10}$.

v. Expression Correlation Analyses

All 15,473 orthologous genes were used for all correlation analysis. The expression profiles throughout development of the orthologous genes were correlated between the sexes and between species using Pearson's correlation coefficient. An arbitrary Pearson's correlation coefficient threshold was defined to divide the genes according to their conservation of expression profiles between the sexes and between species. Positive correlations ($r > 0.5$), low-to-no correlations ($0<r<0.5$) negative correlations ($r <0$).

vi. Clustering Analyses

6,967 orthologous genes that are broadly expressed across all stages and between both sexes and species were clustered across developmental stages using Spearman's correlation as a distance metric for hierarchical clustering in R. The resulting clustered dendrogram was bootstrapped using the pvclust package in R with 2,000 bootstrap replicates (Suzuki and Shimodaira 2006). All clustering relationships were significant above the 95% confidence level. A heatmap was generated in R to visualize the clustering relationships among the stages.

vii. Gene Ontology (GO) Analyses

All GO term enrichment analyses were conducted using the web-based program Genecodis3 (http://genecodis.cnb.csic.es) (Tabas-Madrid et al. 2012).

CHAPTER 2

NETWORK-LEVEL MOLECULAR EVOLUTIONARY ANALYSIS OF THE PHOTOTRANSDUCTION PATHWAY ACROSS 12 SPECIES OF *DROSOPHILA*

(This chapter was submitted and provisionally accepted for publication in Genome Biology and Evolution in March 2012)

I. ABSTRACT

Genes do not act in isolation, but rather within networks to coordinate most critical biological functions. Despite their importance, it is not well understood how networks of interacting genes diverge between species. This study examines the evolution of the phototransduction pathway across 12 *Drosophila* species that share a common ancestor 65 million years ago. The genes that comprise this network are highly conserved at the protein sequence level across the species. Moreover, while rates of protein evolution were generally low throughout the network, relatively relaxed rates of non-synonymous substitution were observed on genes involved in chromophore biosynthesis. Using a comparative genomic approach, the lineage-specific duplication event of the key signal transducer molecule, *Galpha49B* was uncovered in the species belonging to the subgenus Drosophila (*D. mojavensis*, *D. virilis* and *D. grimshawi*). These results suggest that the *Drosophila* phototransduction pathway is characterized by evolutionary change at one level of biological organization in the face of strong functional constraint at others.

II. INTRODUCTION

Most critical biological processes within an organism's life span are mediated by the complex interactions of many genes. Interactions in the form of gene networks are also ultimately responsible for generating much of the phenotypic diversity observed in

nature. These networks are often dynamic entities and how they evolve within or between species is not well understood, even for well-characterized gene networks. The *Drosophila* phototransduction pathway is one of the most well-studied and well-characterized gene networks to date. As the first sensory modality to be investigated at the genetic level (Pak et al. 1969) the general topology of the network is well-known, as is the function of almost all its genes at the genetic, biochemical and physiological levels (Zuker 1992; Hardie et al. 2001; Wang and Montell 2007).

Phototransduction is the process of turning light energy hitting the eye into a nerve impulse and it requires approximately 35 genes encoding protein products ranging from structural and transport proteins to G-proteins and ion channels (Figure 2-1, Table 2-1) (Wang and Montell 2007). A photon of light is sufficient to activate the proteins involved in the phototransduction process. The genes that make up this network can be divided into categories based on their contributions to the phenotypic goal of converting light energy into a nerve impulse. A number of genes in the network are involved in the activation and inactivation of rhodopsin molecules (*arr1*, *arr2*, *rdgC, ninaE, Rh2, Rh3, Rh4, Rh5, Rh6, sun, ninaA*) while others are involved in the biosynthesis of the chromophore, the molecule which is bound to rhodopsin and absorbs light (*ninaB, ninaD, pinta, santamaria)*. Other members of the network function in: regeneration of the PIP_2, an important regulator of the ion channels (*norpA, CdsA, Pis, rdgB, laza, Pld, rbo, stops, inaE*), ion-channel specific genes (*trp, trpl, CalX, inaF*), genes coding for the subunits of the heterotrimeric G-protein (*Galpha49B, GBeta76C, Ggamma30A*) and key genes of the *inaD* signalling complex (*ninaC, inaD, inaC*).

Table 2-1: Genes in the *Drosophila* phototransduction pathway

Gene	Function	Chromosomal location in *D. melanogaster*
arr1	Minor arrestin	2L
arr2	Major arrestin	3L
calX	Na^{+}/Ca^{2+} exchanger	3R
cdsA	Phosphatidatecytidylyltransferase activity	3L
Galpha49B	α subunit of G-protein	2R
GBeta76C	β subunit of G-protein	3L
Ggamma30A	γ subunit of G-protein	2L
inaC	Protein kinase C	2R
inaD	PDZ domain containing scaffolding protein	2R
inaE	Lipoprotein lipase activity	X
inaF	Putative regulator of TRP	X
laza	Lipid phosphate phosphohydrolase	3L
ninaA	Cyclophilin	2L
ninaB	Beta-carotene-15,15'-oxygenase	3R
ninaC	Protein kinase/myosin III	2L
ninaD	Class B scavenger receptor	2L
ninaE	Major rhodopsin	3R
norpA	Phospholipase C-β (PLC- β)	X
pinta	Retinoid binding	3R
Pis	Phosphoinositide Synthase	X
Pld	Phospholipase D	2R
Rbo	Putative lipase in PIP_2 pathway	2R
rdgA	Diacylgylcerol kinase	X
rdgB	Phosphatidylinositol transfer protein	X
rdgC	Rhodopsin phosphatase	3L
Rh2	Violet absorbing Rhodopsin	3R
Rh3	UV absorbing Rhodopsin	3R
Rh4	UV l absorbing Rhodopsin	3L
Rh5	Blue absorbing Rhodopsin	2L
Rh6	Green absorbing Rhodopsin	3R
santa maria	Class B scavenger receptor	2L
stops	Deactivation of signaling	3R
sun	Tetraspanin	X
trp	Major light sensitive cation channel	3R
trpl	Light sensitive cation channel	2R

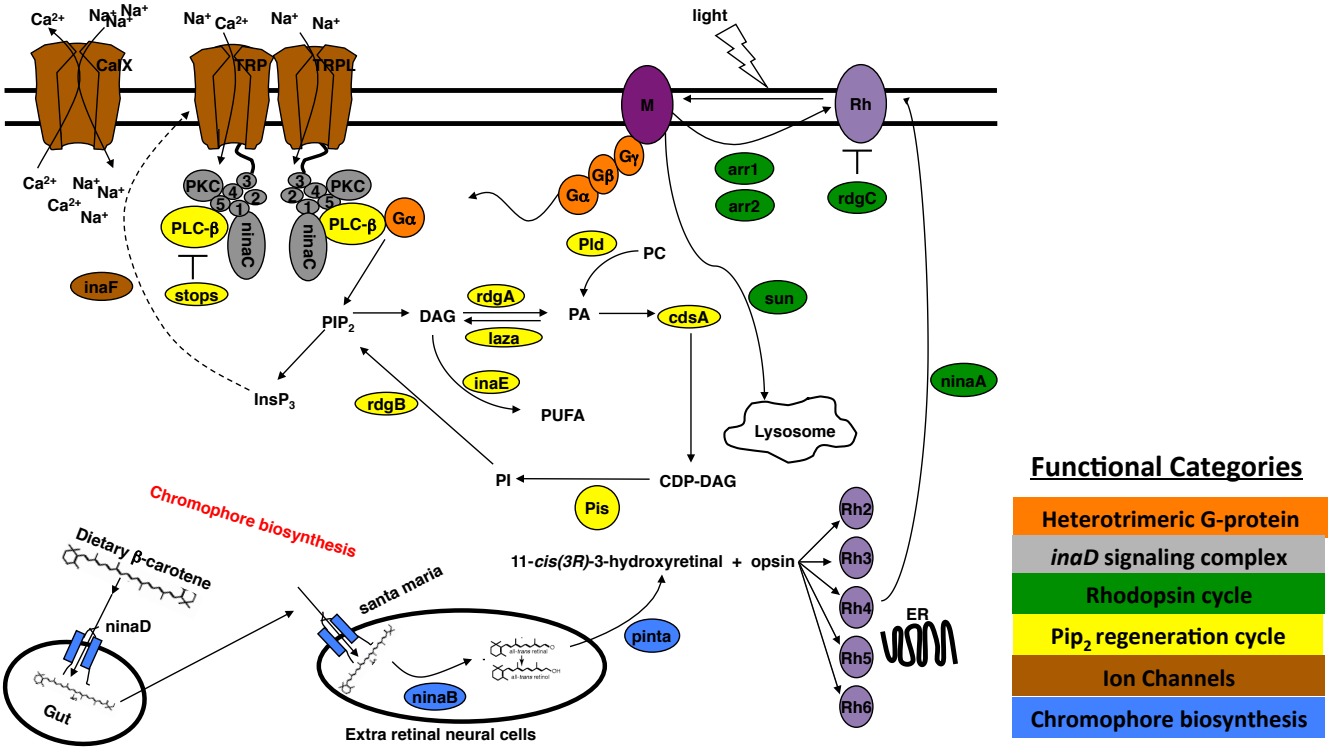


Figure 2-1. The *Drosophila* phototransduction pathway

The *Drosophila* phototransduction pathway is represented here. The genes in this network have been grouped into functional categories and the color scheme depicted represents membership of a gene to a particular functional group. Light energy converts rhodopsin (Rh1-Rh6) to the active state, metarhodopsin. Metarhodopsin activates the heterotrimeric G-protein (G-αβγ), resulting in the subsequent dissociation of the G-α subunit. The G-α subunit, which is encoded by the *Galpha49B* gene, activates an effector enzyme, phosphoinositide-specific phospholipase C or PLC-β. The PLC-β protein (encoded by the *norpA* gene) cleaves PIP_2 (phosphatidylinositol-4, 5-biphosphate) to produce two second messenger molecules, DAG (diacylglycerol) and $InsP_3$ (inositol triphosphate). These second messengers work through an as of yet undetermined mechanism leading to the opening of the Na^+ and Ca^{2+} permeable channels, TRP and TRPL. Another Na^+ and Ca^{2+} permeable channel, *CalX* colocalizes with TRP and is required to counter the influx of Ca^{2+} into the photoreceptor cell. The genes *rdgB*, *Pis* and *CdsA*, which are involved in the regeneration of PIP_2 and the genes *laza,* and *rdgA*, which are involved in the regeneration of DAG, ensure the constant production of DAG and InsP3. A few genes in the network are involved in inactivating the phototransduction process. The arrestin genes *arr1* and *arr2* bind and inactivate rhodopsin. The *rdgC* gene also acts in inactivation by dephosphorylating rhodospin. A number of the genes including *trp*, *ninaC* and *inaC* are assembled onto INAD (encoded by the *inaD* gene), a scaffold protein made up of five PDZ domains. *inaF* serves as a regulator of calcium-channel activity. The genes *ninaA* and sunglasses (*sun*) are both involved in the transport of rhodopsin; *ninaA* from the endoplasmic reticulum to the cell membrane and sun from the cell membrane to the lysosome. The genes *ninaB*, *ninaD*, *pinta* and *santamaria* are involved in the conversion of dietary β-carotene to the chromophore bound to the rhodopsin molecule (11-*cis*-3-hydroxyretinal).

While the *Drosophila* phototransduction pathway has long represented a model in the study of G-protein mediated signaling pathways (Hardie et al. 2001), there is recently abundant evidence of natural variation in this network at the level of gene expression between the sexes and between species. Microarray analyses of whole transcriptome profiles of *D. melanogaster* and *D. simulans* showed evidence of sex-biased and species-specific differential expression of genes involved in phototransduction (Ranz et al. 2003; Arbeitman et al. 2002; Ma et al. 2006). In addition, a detailed qPCR study that examined the gene expression levels of 22 genes in the phototransduction pathway in *D. melanogaster* and *D. simulans* found a significant difference in the expression of the key ion channel gene, *trpl* in *D. simulans* males when compared to *D. melanogaster* males (Landry et al. 2007). In particular, the *trpl* gene is on average expressed 4.7-fold higher in *D. simulans* males compared to *D. melanogaster* males. This divergence in expression between the species is of particular interest as it is thought to allow flies to see better under dim light conditions (Bahner et al. 2002).

However, very little is known about protein sequence divergence of the genes that make up the phototransduction pathway across the *Drosophila* phylogeny. While a few studies have characterized rates of protein evolution of certain genes involved in the network (Carulli and Hartl 1992; Haerty et al. 2007), none have looked at the network as a whole across multiple *Drosophila* species. This study presents a comprehensive protein evolution analysis of all of the genes in the network across the 12 sequenced *Drosophila* species, spanning 65 million years of evolutionary time. A comparative genomic approach taking advantage of the complete genomes of *D. melanogaster* (Adams et al. 2000) and 11 other *Drosophila* species (Clark et al. 2007) was undertaken to determine

rates of protein evolution for all genes in the network. While these analyses indicated generally strong constraint at the level of protein sequence, we found evidence of evolutionary change in the form of a lineage-specific duplication event of the key signal transducer molecule, *Galpha49B*. The duplication of this key member gene is of particular interest as it possibly suggests lineage-specific network topological growth.

III. RESULTS

i. Strong Purifying Selection on all Network Proteins

Amino acid divergence of proteins in the network relative to *D. melanogaster* was generally very low. Percent amino acid identity ranged from an average (across the network) of 85% in the most distant species, *D. grimshawi*, to 96% in the close relative, *D. simulans*. The vast majority (98%) of proteins (379/385) show >75% amino acid identity relative to *D. melanogaster* in all 11 species indicating preservation at the protein sequence level and suggestive of strong functional conservation (percent identities for each gene in Appendix 2, Table A2-1). The selective forces acting on the network were determined using a codon-based model of sequence evolution (Yang 2007) to estimate rates of protein evolution for all 35 genes in the network across the 12 species. In particular, site-specific models M8 ($0< d_N / d_S <1$ and $d_N / d_S >1$) and M8a ($0< d_N / d_S <1$ and $d_N / d_S =1$) were compared. Rates of protein evolution were estimated across all 12 species as well as 4 phylogenetic groupings including: the subgenus Sophophora ($n = 9$), subgenus Drosophila ($n = 3$), the melanogaster group ($n = 6$) and the melanogaster subgroup ($n = 5$). For all five groupings, d_N/d_S was typically $<<1$ for all genes indicating strong purifying selection. Mean d_N/d_S values for the network ranged from 0.053 in the melanogaster subgroup to 0.144 for the species belonging to the subgenus Drosophila

(Figure 2-2; Appendix II, Table A2-5). Tests for positive selection revealed no statistically significant evidence of positive selection acting on any member proteins ($P > 0.05$) (data not shown).

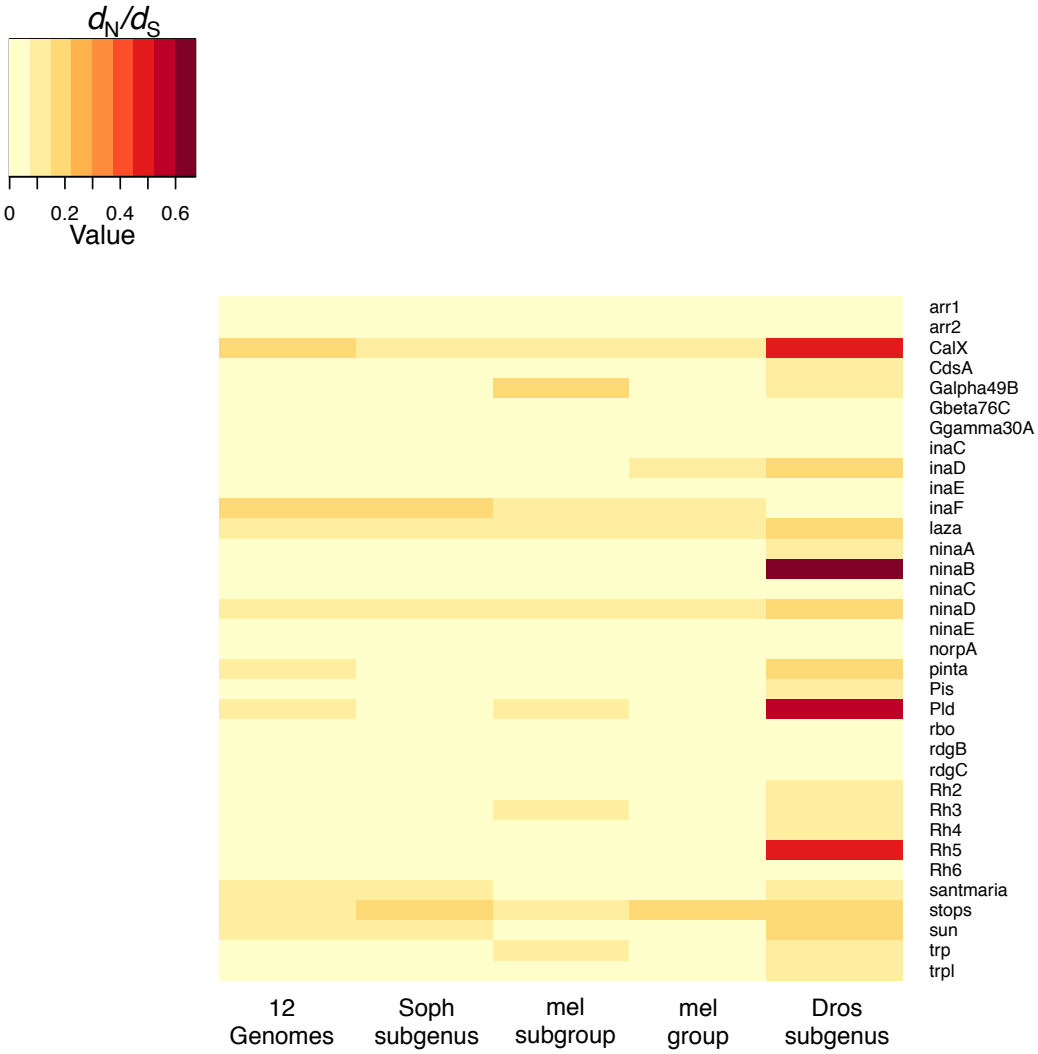


Figure 2-2. Heatmap of d_N/d_S values of phototransduction genes across 12 *Drosophila* species.
Maximum likelihood estimates of the rate of protein evolution (d_N/d_S) for phototransduction members across *Drosophila* species for each phylogenetic grouping (Soph = Sophophora subgenus, mel subgroup= melanogaster subgroup,mel group = melanogaster group, Dros subgenus = Drosophila subgenus). d_N/d_S is typically << 1 indicating strong purifying selection throughout the network and functional conservation of most proteins.

ii. Weaker Purifying Selection on Chromophore Biosynthesis Proteins

To determine if the strength of purifying selection varies among the different parts of the network, the 35 member genes were divided into 6 distinct functional categories (Figure 2-1). Mean rates of non-synonymous substitution (d_N) among the 6 different functional groupings were then compared. Rates of d_N were significantly different among functional categories in the subgenus Sophophora ($\chi^2 = 11.4$, df = 5, $P = 0.04$) and the melanogaster subgroup ($\chi^2 = 11.9$, df = 5, $P = 0.03$) by Kruskal-Wallis one-way analysis of variance. Pairwise Mann-Whitney U-tests revealed a significantly elevated mean rate of non-synonymous substitution in genes involved in chromophore biosynthesis in all 5 phylogenetic groups ($P < 0.05$), however, none were significant after Bonferroni correction ($m_{tests} = 15$, $P > 0.05$). Similar results were found using d_N/d_S.

iii. Duplication of *G-alpha-q* in subgenus Drosophila ~45–60MYA

Computationally based comparative analyses revealed that the annotated duplications of *ninaC* in *D. simulans and D. persimilis* [(*dsim_GLEANR_7199, dsim_GLEANR_7200)*, (*dper_GLEANR_21046*, *dper_GLEANR_21047*)] and *rdgA* in the obscura group [(*dpse_GLEANR_1145, dpse_GLEANR_1148*), (*dper_GLEANR_13146*, *dper_GLEANR_13149*)] were the result of errors in gene prediction and ortholog identification (Table 3-1, Appendix). For both *ninaC* and *rdgA*, two genes were predicted where there is actually one; since both predicted genes exhibited sequence homology with the full length locus in *D. melanogaster*, they were annotated incorrectly as paralogs. The *ninaC* putatively annotated duplicate genes are tandemly arrayed but are separated by only 5 bp; moreover, the number of exons in the two genes sum together to make up exactly the 13 exons of *ninaC* in *D. melanogaster* (Figure 2-3).

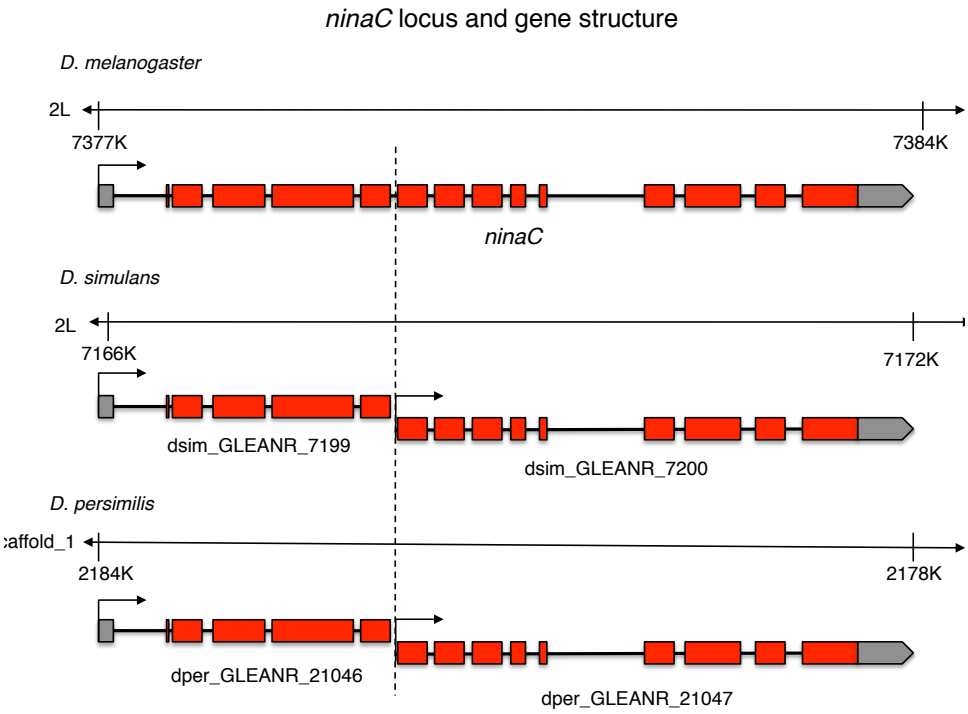


Figure 2-3. *ninaC* locus and gene structure

In *D. simulans* and *D. persimilis*, annotations of the locus *ninaC* report two genes where there is really one: the reported *ninaC* duplicate genes are tandemly arrayed in both *D. simulans* and *D. persimilis* and are separated by only 5 bp. Red (dark) boxes indicate exons and grey boxes depict 5' and 3' UTRs. The number of exons in the two genes sum together to make up exactly the 13 exons of *ninaC* in *D. melanogaster* in both *D. simulans* and *D. persimilis*.

A similar pattern is found at the *rdgA* locus (data not shown). In the case of *rdgA*, the sequence features that caused the prediction algorithm to err and split the gene are shared between related species, resulting in an apparent, but false, pattern of lineage-specific gene duplication.

In contrast, comparative analysis of the lineage-specific duplication of *Galpha49B*, which belongs to the *G-alpha-q* class of G-alpha proteins, in the subgenus Drosophila (Figure 2-4) does not show this pattern. Specifically, duplicate pairs of *Galpha49B* in *D. mojavensis, D. virilis* and *D. grimshawi* are tandemly arrayed, but are separated by >6 kb and span ~11 kb of genomic space in all three species. Both genes have an open reading frame (ORF) of exactly 1062 base pairs that encodes a 353 amino acid chain similar in length to the canonical *G-alpha-q* in *D. melanogaster*, *Galpha49B* (Lee et al. 1990). We designate these genes *G-alpha-q-1* and *G-alpha-q-2,* respectively, in each species. Since the duplication of the *G-alpha-q* gene appears to be unique to the

species belonging to the Drosophila subgenus based on the phylogeny of Singh et al. 2009, this event is estimated to have occurred between ~45-60 MYA (Figure 2-4).

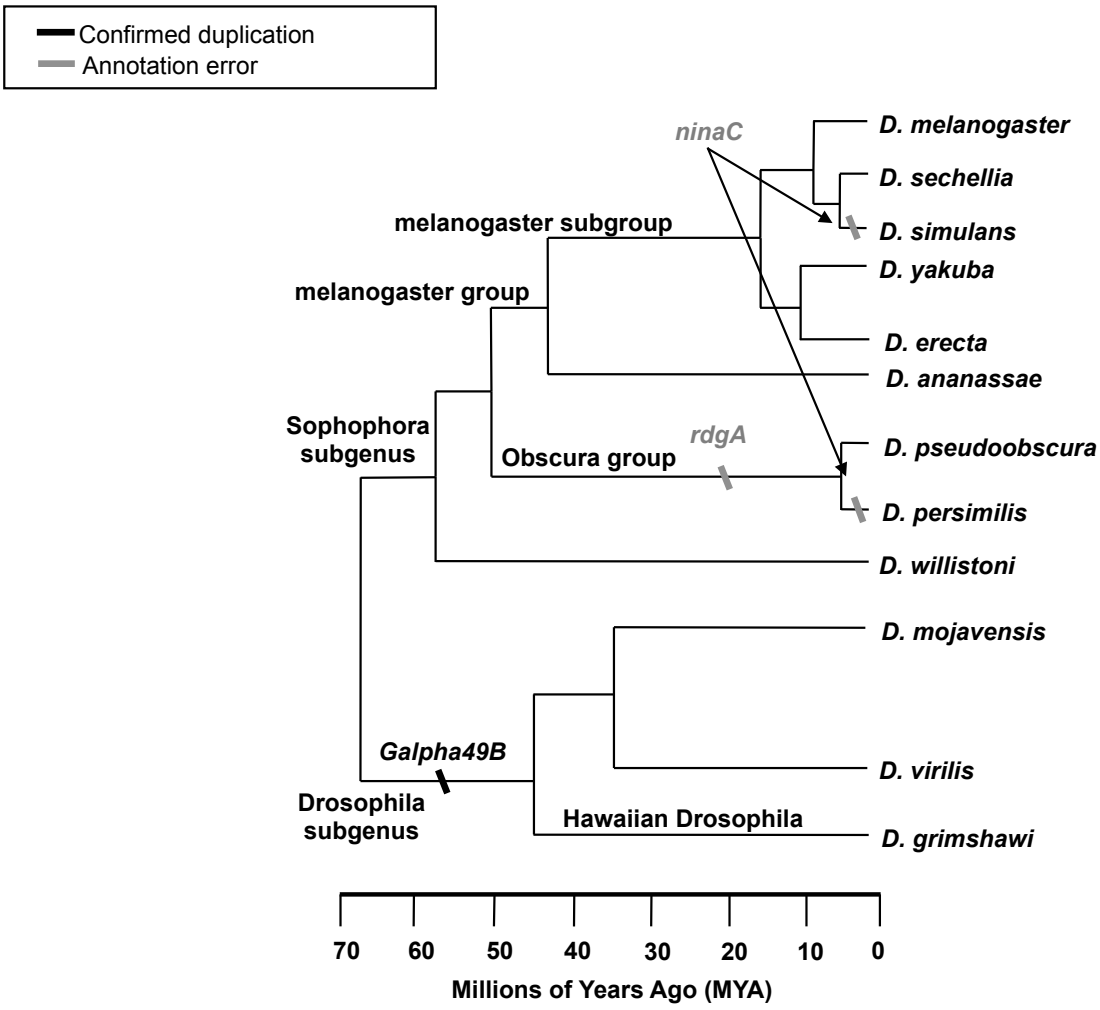


Figure 2-4. Duplications mapped on to phylogeny of genus Drosophila. Phylogeny showing 11 *Drosophila* species that diverged from *D.melanogaster* 3-65 million years ago modified from Singh et al. 2009. The duplication of the *Galpha49B* gene (black bar) occurred in the subgenus Drosophila (*D. mojavensis*, *D. virilis* and *D. grimshawi*,). The genes *rdgA* and *ninac* were incorrectly predicted to be products of lineage specific duplication events (grey bar).

IV. DISCUSSION

Divergence in gene networks between species can occur at all levels of genomic architecture, including at the level of protein coding sequences, level of transcriptional regulation or even at the level of network topology. As a result, the evolutionary forces acting on gene networks may be both diverse and complex in nature. Using a comparative genomic approach, this study examined how natural selection has molded the protein evolution of the phototransduction across 12 distinct *Drosophila* species that diverged from one another 3-65 MYA.

Many gene network evolution studies have focused on comparing rates of protein evolution to network topology in both larger protein-protein interaction networks (Fraser et al. 2002; Lemos et al. 2004; Lemos et al. 2005; Hahn and Kern 2005) and smaller physiological, metabolic and/or biochemical networks in both plants (Rausher et al. 1999; Lu and Rausher 2003; Ramsay et al. 2009; Flowers et al. 2009; Livingstone and Anderson 2009) and animals (Riley et al. 2003; Flowers et al. 2007; Alvarez-Ponce et al. 2009; Montanucci et al. 2011, Yu et al. 2011). In many of these studies, the linear topology of the networks has allowed for a comparison of rates of protein evolution relative to position in the network. Consequently, a number of these studies have reported evidence of a gradient of purifying selection distributed along the network with more relaxed selection on downstream member proteins (Rausher et al. 1999; Riley et al. 2003; Ramsay et al. 2009; Wu et al. 2010; Yu et al. 2011) while others have found contrary results suggesting downstream genes are more constrained (Alvarez-Ponce et al. 2009; Cui et al. 2009; Montanucci et al. 2011; Alvarez-Ponce et al. 2010, Jovelin and Phillips 2011). The overall lack of consensus coupled with the possibility that other factors such

as expression level may more strongly affect rates of network evolution (Jovelin et al. 2009) suggests patterns of purifying selection are likely to be pathway-or-network specific.

Because the phototransduction pathway is comprised of two cycles and several pathway-like structures (Figure. 2-1), it does not possess an obvious linear gradient. Therefore, in this study, the genes that make up the network were organized based on functional categories and groups. Conservative, non-parametric tests revealed weak evidence for relaxed purifying selection in genes involved in the biosynthesis of the chromophore in some lineages. One explanation for more relaxed purifying selection on these genes may be that while variation in proteins responsible for chromophore biosynthesis does affect the visual response (Wang and Montell 2007; Wang et al. 2007) none of these proteins are directly involved in real-time channel activation or inactivation (Wang et al. 2007) and so may be under a different, weaker selective regime. A larger sample size (more species) will help clarify this pattern. These results here suggest that as a whole the phototransduction pathway genes exhibit strong conservation at the level of protein sequence over 65 million years of evolutionary time. However, these findings do not rule out the fact that divergence in these genes between species can be found at other levels of genomic architecture including at the level of the *cis*-regulatory regions. A more detailed analysis of the non-coding regions of these phototransduction genes is necessary to make this determination.

In addition to protein sequence change, changes to the topology of a network, including the distribution (number and type) of genes, can occur. In particular, duplicate genes, well-known to be critical for evolutionary innovation (Ohno 1970; Force et al.

1999; Lynch and Connery 2000), are crucial to the growth of gene networks over evolutionary time (Teichmann and Babu 2004). In the phototransduction pathway, there is potential evidence of topological network growth in the form of a lineage-specific duplication; while 34 of 35 network member genes maintain a 1:1 orthology relationship across the 12 species, one gene appears to be duplicated in a lineage-specific manner, the signal transducer, *Galpha49B*.

Are both *G-alpha-q-1* and *G-alpha-q-2* duplicate genes involved in phototransduction, or has one copy acquired a new function? The former would constitute authentic lineage-specific topological growth of the phototransduction pathway, whereas the latter would constitute evolutionary co-option of a member of the phototransduction pathway for another function, possibly as the activator of another signal transduction network. Unfortunately, comparative sequence data alone is insufficient to differentiate between these two alternatives. Detailed experimental work in *D. mojavensis, D. virilis* and/or *D. grimshawi* will be necessary to establish the precise functional role of *G-alpha-q-1* and *G-alpha-q-2* as it relates to phototransduction and the visual response.

While drastic protein sequence divergence appears to have been selected against in the *Drosophila* phototransduction pathway, hallmarks of adaptive evolution at shorter divergence times (3-6 MYA) have been previously reported in this network at the level of mRNA expression. The ion channel gene *trpl* shows a 4-fold up-regulation in expression in *D. simulans* males when compared to *D. melanogaster* males (Landry et al. 2007). Here, this study reports evidence of potential lineage-specific topological growth of the network in the form of the duplication of *Galpha49B* in the Drosophila subgenus.

Detailed functional analysis on the duplicate genes in this lineage is necessary to determine if this is the case.

Overall, these results suggest that the evolution of the *Drosophila* phototransduction pathway appears to be characterized by evolutionary change at one level of biological organization (e.g. at the level of gene expression and potentially at the level of network topology) despite strong constraint at other levels (e.g. primary protein sequence). Future comparative studies that examine levels of variation and divergence of different biological entities— proteins, small RNAs, regulatory motifs, etc.— in other known gene networks will help establish what types and amounts of genetic variation are possible and which are phenotypically relevant.

V. CONCLUSIONS AND FUTURE DIRECTIONS

Here, we found evidence of strong purifying selection across ~65 MY of evolutionary time suggesting that observed differences at the level of transcription between species do not correlated with changes at the level of protein sequence. An analysis of the *cis*-regulatory regions of these genes across the *Drosophila* phylogeny is necessary to provide insight into the underlying regulatory mechanisms governing the observed differences in gene expression.

VI. MATERIALS &METHODS

i. Comparative Genomic Analysis of Homologous Genes

Protein-coding nucleotide sequences for each species were obtained from the *Drosophila* 12 Genomes Consortium AAAwiki (http://rana.lbl.gov/drosophila/) based on predicted orthologous genes defined by fuzzy reciprocal best hits and synteny criteria (Clark et al. 2007). 366 of 385 expected orthologous genes (35 phototransduction genes

×11 species) were reported for the 11 sequenced non-melanogaster species; 3 genes were predicted to have paralogs in six different species: *ninaC (D. simulans and D. persimilis)*, *rdgA (D. pseudoobscura and D. persimilis)* and *Galpha49B (D. mojavensis, D. virilis, and D. grimshawi)* (Clark et al. 2007). We recovered all 15 missing orthologous genes (12 genes among 6 species) computationally. Four annotated genes and 6 recovered genes exhibited obvious errors in annotation and/or sequence, including missing exons, missing or ambiguous sequences, internal stop codons, and frameshift errors (Table 2-2) Missing orthologous genes were identified using BLAST (Altschul et al. 1997) and BLAT (Kent 2002; http://genome.ucsc.edu). We used an *in silico* comparative genomic approach to identify missing exons, frame-shift mutations, internal stop codons and other DNA sequence and gene structure errors in non-*D. melanogaster* genes. Iterative local and global alignment of whole genes, exons, introns, and other sequence features among species using BLAST, BLAT, Emboss Needle (Rice et al. 2000; http://www.ebi.ac.uk/emboss/align/; Needleman and Wunsch 1970) and custom PERL scripts were used to identify and correct corrupt or missing DNA sequences and gene structures. *D. melanogaster* genome annotations obtained from FlyBase (FB2009_07 Dmel Release 5.20) were used as a reference for establishing gene structure and syntenic relationships. We used a strict parsimony criterion to infer missing sequences, i.e., only nucleotides that were identical among all species examined were inferred. All other missing nucleotide sequences were coded as "N" (ambiguous).

Nine *D. simulans* genes could not be recovered by our comparative method because they contained long (>30% gene length) runs of ambiguous nucleotides. These nine genes (*ninaA, ninaC, norpA, rdgA, rdgB, rdgC, Rh3, trp* and *trpl*) were individually

re-assembled *de novo* using the original DNA sequencing reads (obtained at http://www.dpgp.org/syntenic_assembly/ by BLAST with the corresponding *D. melanogaster* ortholog) followed by comparative genome assembly using *D. melanogaster* as a guide (S. Salzberg, personal communication) allowing for up to 13% polymorphism among reads. Each re-assembled *D. simulans* gene was then subjected to *in silico* comparative genomic reconstruction as above. *rdgA* contained too many sequencing errors and could not be re-assembled. Full sequences of all reconstructed genes are in Appendix II.

Table 2-2: Phototransduction genes with annotation errors

Gene	*D. simulans*	*D. sechellia*	*D. ananassae*	*D. pseudoobscura*	*D. persimilis*	*D. willistoni*	*D. virilis*
GBeta76C		D					
inaC		C					
inaD							C,D
inaF					C,D		
ninaA	A						
ninaC	A,B				B,C		
norpA	A,B				C,D		
pinta					C,D		
Pld		C,D			C,D		
rdgA	A			B	B		
rdgB	A						
rdgC	A						
Rh3	A,B	D					
Rh5		C,D					
santa-maria					D		
stops		C			D		
sun			C			C	C
trp	A						
trpl	A,C						

A - *De novo in silico* reassembly required.
B - Predicted to be duplicated according to AAAwiki (*Drosophila* 12 Genomes Consortium, 2007).
C - No predicted ortholog, manual reconstruction required.
D - Missing exons, frameshifts, or internal stop codons.

iii. Protein Evolution

Protein coding sequences were globally aligned using CLUSTALW v2.02 (Larkin 2007) using the amino acid translation of each sequence followed by back-translation into nucleotides. Maximum likelihood estimates of the rate of nonsynonymous substitution (d_N) and synonymous substitution (d_S) between alignments were calculated using a codon-based model of sequence evolution (Goldman and Yang 1994; Yang 2007). Likelihood ratio tests for positive selection were performed on all genes in the phototransduction pathway by comparing twice the difference in log-likelihood between model M8a and model M8 in PAML v4.2 (Yang 2007; Swanson et al. 2003). Positive selection was inferred if both $2(ln\ L1 - ln\ L2) \geq 6.63$ ($P < 0.01$, $\sim\chi^2_1$) and d_N/d_S was greater than 1 in at least one of the site classes.

iv. Functional Grouping and Statistics

Genes were grouped based on the categories described in Wang and Montell (2007). The functional classes include: biosynthesis of the chromophore (*ninaB, ninaD, pinta, santamaria*), rhodopsin cycle (*arr1*, *arr2*, *rdgC, ninaE, Rh2, Rh3, Rh4, Rh5, Rh6, sun, ninaA*), genes in the PIP_2 regeneration pathway (*norpA, CdsA, Pis, rdgB, laza, Pld, rbo, stops, inaE*), ion-channel specific genes (*trp, trpl, CalX, inaF*), genes coding for the subunits of the heterotrimeric G-protein (*Galpha49B, GBeta76C, Ggamma30A*) and key genes of the *inaD* signalling complex (*ninaC, inaD, inaC*). Mean differences in d_N and d_N/d_S in each group were calculated across all 12 species as well as for four non-mutually exclusive phylogenetic groupings: Sophophora (*D. melanogaster, D. sechellia, D.*

simulans, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. persimilis, D. willistoni), the melanogaster subgroup (*D. melanogaster, D. sechellia, D. simulans, D. yakuba, D. erecta*), the melanogaster group (*D. melanogaster, D. sechellia, D. simulans, D. yakuba, D. erecta, D. ananassae*) and members of the subgenus Drosophila (*D. mojavensis, D. virilis, D. grimshawi*). Significant differences in d_N and d_N/d_S among groups were determined using Kruskal-Wallis one way analysis of variance and Mann-Whitney U-tests on pairwise comparisons of groups followed by Bonferroni correction. All statistical analyses were performed using R (Ihaka and Gentleman 1996, http://www.r-project.org).

CHAPTER 3

ISOFORM-SPECIFIC DUPLICATION OF AN ALTERNATIVELY SPLICED GENE IN *DROSOPHILA*

I. INTRODUCTION

Alternative splicing and gene duplication are both mechanisms that contribute to protein diversity. Alternative splicing is a regulated process that results in the generation of multiple mRNA transcripts from a single locus. Rates of alternative splicing differ between species and are thought to not only play a role in transcriptome and proteome complexity but in phenotypic diversity as well (Brett et al. 2001; Calarco et al. 2007; Barbosa-Marois et al. 2012). For example, a recent study showed evidence that almost every gene in the human genome with more than one exon is alternatively spliced, while in *D. melanogaster* about 60% of genes are alternatively spliced (Wang et al. 2008; Graveley et al. 2011). There are a number of different types of alternative splicing that determine the resulting mRNA transcript structure (Figure 3-1).

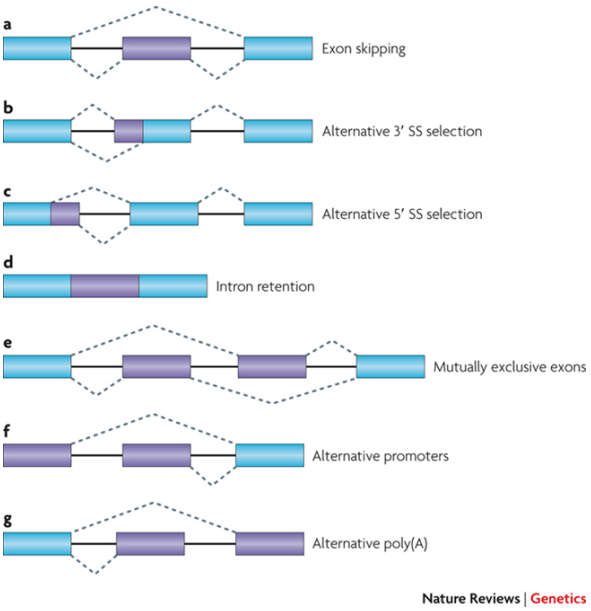


Figure 3-1: Types of Alternative Splicing
(Keren et al. 2010)

Constitutive exons are shown in blue and alternatively spliced regions in purple. Introns are represented by solid lines, and dashed lines indicate splicing options.

Of particular interest is the strategy that leads to a transcript with mutually exclusive exons (Figure 3-1E). In this case, the gene is composed of clusters of internal exons that are spliced in a mutually exclusive manner. Interestingly, it has been shown that these clusters of internal exons are typically of the same length and have high sequence similarity, thus in contrast to some of the other types of alternative splicing, mutually exclusive splicing does not alter the size and length of the resulting ORF (Wang et al. 2008; Pohl et al. 2013; Hatje and Kollmar 2013). It is possible to imagine clusters of internal exons with different tissue or developmental-stage specific expression allowing a single locus to generate transcripts with diverse spatial-temporal expression patterns without altering transcript size or ORF. While alternative splicing via mutually exclusive splicing is much less common than the other types (Pohl et al. 2013), the evolutionary advantages for maintaining this strategy are clear.

In contrast to the process of alternative splicing, gene duplications typically arise either as a result of an error during DNA replication/repair or are generated by retrotransposition events. Nonetheless, they are thought to be crucial to the evolution of novel traits within organisms (Ohno 1970; Walsh 1995; Lynch and Connery, 2000). The widespread existence of gene families across phylogenetically diverse taxa suggests that the retention of gene duplicates is a frequent occurrence (Ohta 2000; Elemento et al. 2002; Cannon et al. 2004; Shoja and Zhang 2006; Hahn et al. 2007; Hanada et al. 2008; Fan et al. 2008; Jacquemin et al. 2013). Once they are fixed, duplicate genes are subject to distinct evolutionary fates. In the first possibility, one of the genes maintains the original function while the other accumulates random, often deleterious mutations rendering the gene nonfunctional. Alternatively, one of the duplicates acquires a new

function as a result of beneficial mutations or neo-functionalization. The third possible fate is that both genes accumulate degenerative mutations resulting in the ancestral gene function being subdivided among the duplicate pairs, a process known as sub-functionalization (Force et al. 1999).

The evolutionary model of duplication-degeneration-complementation (DDC) suggests that many duplicate genes are maintained through sub-functionalization as a result of degenerative mutations that occur specifically within regulatory elements resulting in subdivision of the ancestral gene expression (Force et al. 1999). However, other complex patterns of sub-functionalization involving changes to coding gene structure have been reported. For example, a single ancestral gene that is alternatively spliced can give rise to duplicate descendent copies that contain the exon complement of mutually exclusive splice variants. Aside from one example between plants (Rohmann et al. 2009) and another between yeast species (Marshall et al. 2013), thus far this phenomenon, which I have termed "isoform duplication", has mainly been reported between very distantly related species (for example human-fish which diverged from each other > 400 MYA) (Lister et al. 2001; Altschmied et al. 2002; Yu et al. 2003). Here, I present evidence of an isoform duplication event that occurred ~45-65MYA between *Drosophila* species.

In *D. melanogaster*, sensory signal transduction pathways such as the phototransduction pathway are mediated via the interaction between G-protein coupled receptors (GPCRs) and downstream heterotrimeric G-proteins. These heterotrimeric G-proteins are made up of a GDP-bound alpha subunit and a heterodimer composed of beta and gamma subunits (Gilman 1987; Sprang 1997). Sensory signals are transmitted from

the GPCR resulting in activation of the alpha subunit of the intracellular heterotrimeric G-protein. In the *Drosophila* phototransduction pathway, a photon of light is sufficient to activate the GPCR, rhodopsin, which in turn activates the alpha subunit of the heterotrimeric G protein. Thus, the G-alpha subunit acts a critical intermediary ensuring the signal is transmitted downstream through the pathway (Lee et al. 1994; Scott et al. 1995; Hardie 2001).

There are six annotated and experimentally validated genes that encode G-alpha subunits in the *D. melanogaster* genome. These genes have been categorized into classes depending on their molecular function (Table 3-1). The G-alpha subunit protein that belongs to the class q (G-alpha-q) is the only one that has been shown to function in the phototransduction pathway. As an integral part of the phototransduction pathway, it is among the most well-characterized and well-studied of the G-alpha subunits.

Table 3-1: Annotated and experimentally validated G-alpha subunits in the *D. melanogaster* genome.
Modified from Katanayeva et al. 2010

G-alpha subunits	*D.melanogaster* Gene IDs	Molecular Function/Biological Process
G-alpha-q	Galpha49B	Phototransduction, olfaction
G-alpha-o	Galpha47A	Neuro-synaptic growth, ventral cord development
G-alpha-s	Galpha60A	Wing disc development, neuromuscular development
G-alpha-i	Galpha65A	Establishment of glial blood brain barrier
G-alpha-f	Galpha73B	Regulation of hemocyte proliferation
Concertina	cta	Gastrulation, regulation of embryonic cell shape

In *D. melanogaster*, the *G-alpha-q* gene, annotated as *Galpha49B*, is located on the right arm of the 2nd chromosome. Expression of this gene has been localized to a number of different sensory tissues including but not limited to the brain, photoreceptors

and olfactory and gustatory neurons (Talluri et al. 1995; Alvarez et al. 1996; Kalidas et al. 2002). This gene is alternatively spliced via mutually exclusive exons that are expressed in a tissue-specific manner (Figure 3-2A). It has two main isoforms; one that has been shown to be expressed in a multitude of sensory biology organs and tissues and has been implicated in chemosensory, olfaction and gustatory pathways, while the other has been shown to be exclusively expressed in photorecetors and is involved only in phototransduction (Lee et al. 1990; Lee et al. 1994; Talluri et al. 1995; Alvarez et al. 1996; Ratnaparkhi et al. 2002; Kain et al. 2008).

The single alternatively spliced gene, *Galpha49B* in *D. melanogaster* gave rise to duplicate copies that contain the exon complement of mutually exclusive splice variants in the species belonging to the subgenus Drosophila (*D. mojavensis*, *D. virilis* and *D. grimshawi*). Here I examine the evolution of the gene structure and expression of these novel lineage-specific duplicate genes.

II. RESULTS

1 – *G-alpha-q* gene structure in the Drosophila subgenus and outgroup species

Comparative genomic analysis of the *G-alpha-q* locus across 12 species of *Drosophila* revealed a novel isoform-specific duplication event in the subgenus Drosophila that occurred ~45-65MYA (Figure 3-3). In *D. mojavensis*, *D. virilis* and *D. grimshawi* each duplicate appears to retain the exon complement of only one of the splice variants (Figure 3-2B). More specifically, each duplicate in *D. mojavensis*, *D. virilis* and *D. grimshawi* consists of the five constitutively expressed exons along with its own complement of mutually

exclusive exons (Figure 3-2B).

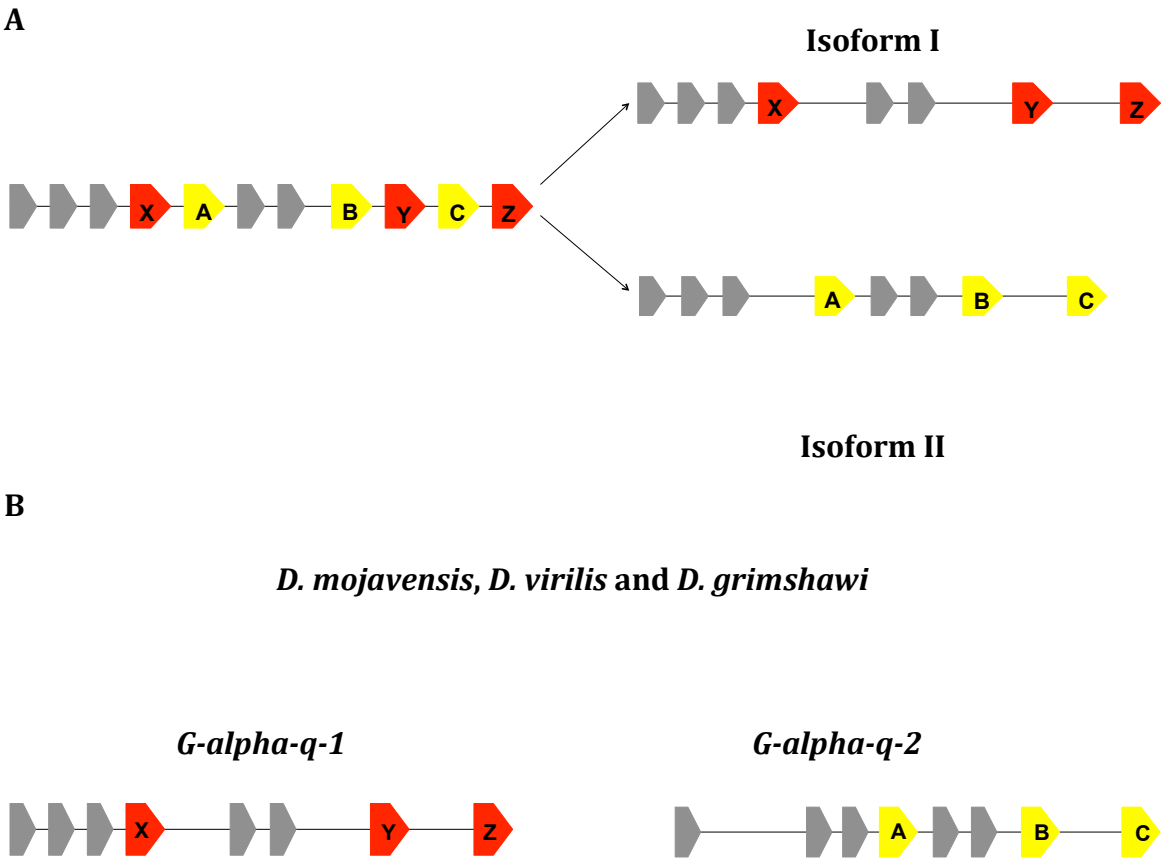


Figure 3-2. Gene exon-intron structure of *G-alpha-q* in *D. melanogaster* and the species of the Drosophila subgenus

G-alpha-q is a single-copy gene that is alternatively spliced with phototransduction and non-phototransduction isoforms in *Drosophila melanogaster.* The gene consists of five constitutively expressed exons (colored in grey) and two sets of mutually exclusive exons (colored in red and yellow). Isoform I (red exons denoted X,Y and Z) has been shown to be involved in a number of different sensory signal transduction pathways including taste and smell. While isoform II (yellow exons denoted A, B and C) is the rhodopsin-binding isoform that is expressed at high levels in the eye and is involved in phototransduction. This same gene is present as two duplicate copies in the subgenus *Drosophila.* Each duplicate copy retains the exon complement of only one of the splice-variants (isoform duplication).

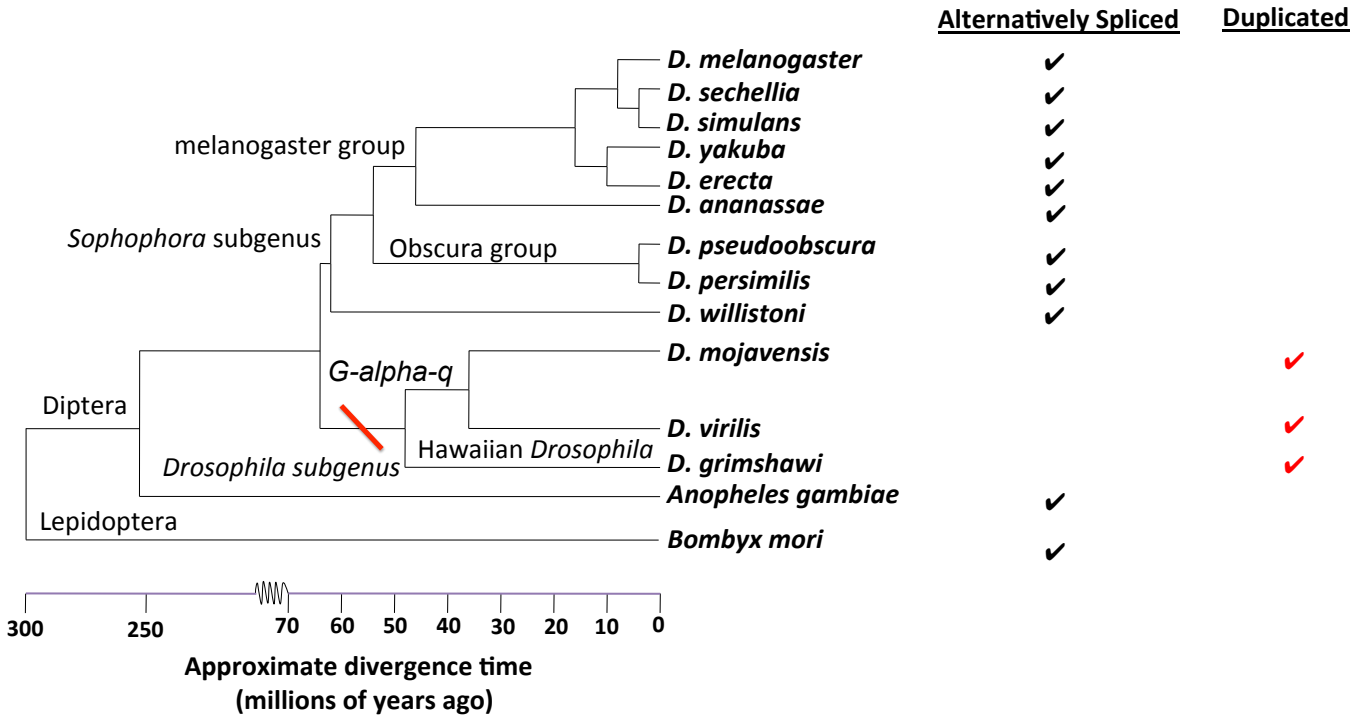


Figure 3-3. Lineage-specific isoform duplication of the *G-alpha-q* gene.
The duplication of the *G-alpha-q* gene (indicated in red) occurs in the Drosophila subgenus (*D. mojavensis*, *D. virilis* and *D. grimshawi*) and is estimated to have occurred ~ 45-65 MYA. The gene is alternatively spliced in all other *Drosophila* lineages as well as in the outgroup species, *Anopheles gambiae* and *Bombyx mori*.

The duplicate pairs are tandemly arrayed but are separated by ~6kb intergenic space and span ~11kb of genomic space in all three species. Both genes have an ORF length of exactly 1062 base pairs, which encodes a 353 amino acid chain similar in length to the *D. melanogaster Galpha49B* gene (Lee et al. 1990). In addition, synteny is conserved at this locus between *D. melanogaster* and these species; the genes are flanked on the 5' end by the *Amph* gene and by the *muskelin* gene on the 3' end. A multiple sequence alignment of the translated sequences of the *D. melanogaster* alternatively spliced isoforms and the duplicate genes from the three species shows generally high sequence similarity suggesting overall conservation of motifs and domains (Figure 3-4).

Structural studies on the *G-alpha-q* protein have shown that the last 3 amino acids are crucial for determining receptor specificity (Conklin et al. 1993). The rhodopsin binding isoform in *D. melanogaster* and its corresponding duplicate gene in *D. mojavensis*, *D. virilis* and *D. grimshawi* have an NLG motif and likewise the non-phototransduction isoform and its duplicate complement have an NLV motif (Figure 3-4). Taken together, these results illustrate that duplicate genes in *D. mojavensis*, *D. virilis* and *D. grimshawi* share structural features with their complementary splice variants in *D. melanogaster*.

To determine if the gene is alternatively spliced or duplicated in other insect species, I examined the annotated structure of the orthologous locus in mosquito, *Anopheles gambiae* and silkworm, *Bombyx mori*. These species were chosen because: 1) functional studies on *G-alpha-q* have been conducted in both and 2) they diverged from Drosophilids ~250 MYA and ~300 MYA respectively allowing for the determination of whether the ancestral gene was alternatively spliced or duplicated ~ 300 MYA. (Sakurai et al. 2004; Miura et al. 2005; Hull et al. 2010; Weinstock et al. 2006). The *G-alpha-q* gene in these species is also alternatively spliced via mutually exclusive splicing suggesting that the isoform duplication of *G-alpha-q* is lineage-specific (Figure 3-3).

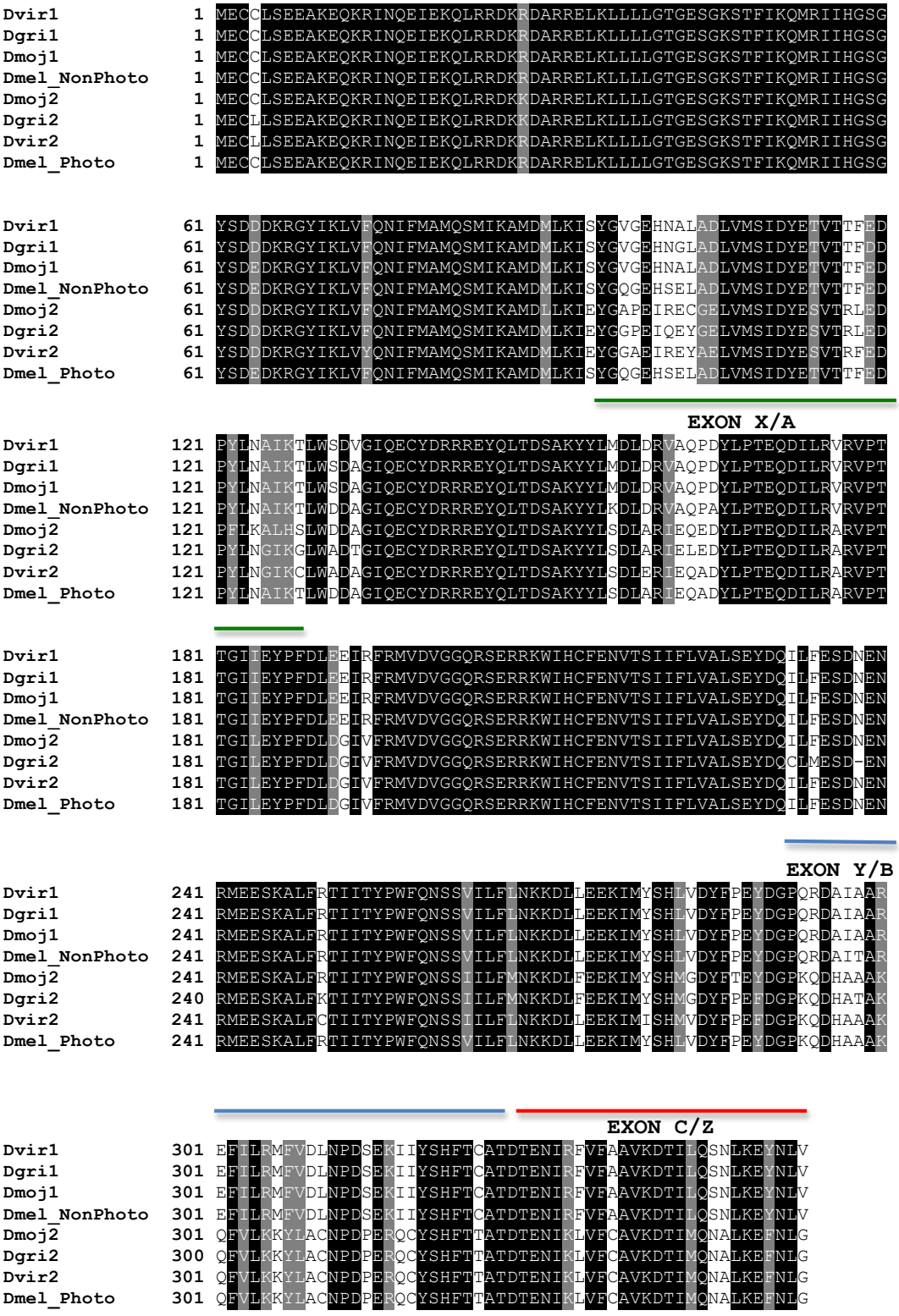


Figure 3-4. Sequence alignment of *G-alpha-q* protein sequence.

Shading in black indicates regions of identical sequence. Shading in grey indicates an amino acid change to one that has similar biochemical properties. Shading in white indicates a change to an amino acid that doesn't share biochemical properties. Regions of the protein that encode the mutually exclusive exons are marked on the alignment.
Very high sequence identity is observed generally throughout the protein. The motif at the C-terminus end of the protein has been shown to determine receptor binding. Each isoform and its complementary duplicate share the same amino acid at this residue.

ii. *G-alpha-q* duplicate genes are transcriptionally active in the Drosophila subgenus

One way to explore the consequences and causes behind the duplication is to examine mRNA transcript abundance of the duplicates in different tissues to determine if their expression patterns correspond with those observed in the alternatively spliced gene. I designed cross-species (*D. mojavensis* and *D. virilis*) duplicate specific primers and splice variant specific primers (*D. melanogaster*) in order to make accurate and quantitative comparisons between the different species and tissues. I quantified absolute mRNA transcript abundance of each duplicate gene and *D. melanogaster* splice variant using qPCR. As *D. grimshawi*, a Hawaiian *Drosophila*, is extremely difficult to rear in the lab, all qPCR analyses of the duplicate genes were conducted on *D. mojavensis* and *D. virilis* males and females. First, to determine if the duplicate genes are both transcriptionally active, I quantified absolute mRNA transcript abundance from heads and bodies of *D. mojavensis* and *D. virilis* males and females. I was able to detect expression of both duplicate genes in the heads and bodies of both species suggesting they are transcriptionally active (Figure 3-5).

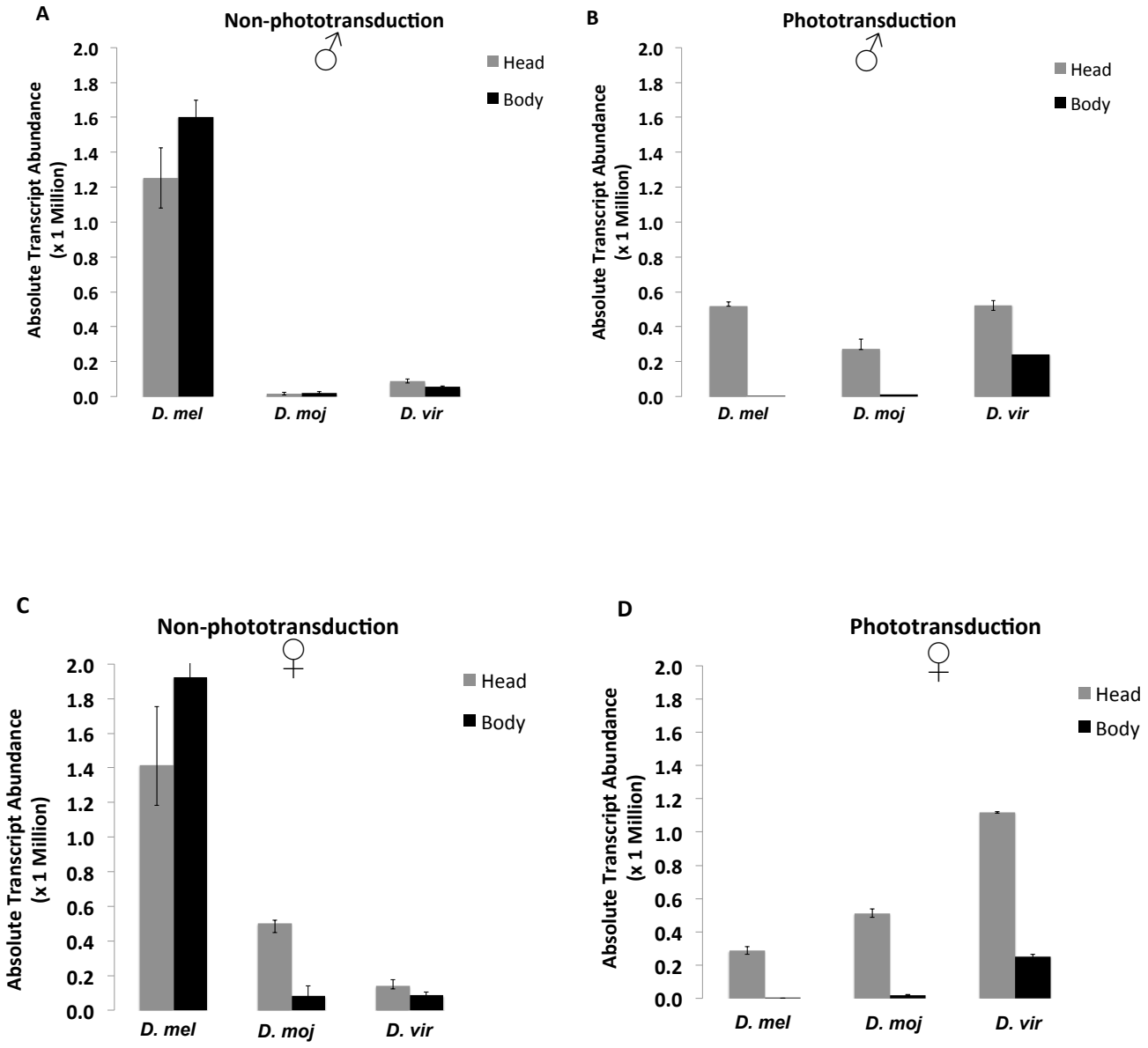


Figure 3-5 Absolute transcript abundance of *G-alpha-q* alternatively spliced variants in *D. melanogaster* and the duplicates in *D. mojavensis* and *D. virilis*

The following bargraphs (A-D) compare the absolute transcript abundances (quantified via qPCR) of the *G-alpha-q* phototransduction specific and non-phototransduction splice variants (*D. melanogaster*) and duplicates (*G-alpha-q-1 and G-alpha-q-2 in D. mojavensis* and *D. virilis*). Absolute transcript abundance was quantified separately from heads and bodies of males and females for all 3 species. Bars represent means calculated for 3+ replicates. Error bars indicate SEM.

To determine if expression patterns of the duplicate genes were consistent with expression patterns of the alternatively spliced variants in *D. melanogaster*, I compared absolute mRNA transcript abundance from heads and bodies of *D. melanogaster*, *D. mojavensis* and *D. virilis*. I found expression of the phototransduction-specific duplicate in *D. mojavensis* and *D. virilis* was largely consistent with the expression pattern of its complementary alternatively spliced variant (Figure 3-5 B,D). Of the two duplicate genes, *G-alpha-q-2* was expressed significantly higher in the heads compared to the bodies of *D. mojavensis* and *D. virilis* males and females ($P = 0.01$; Welch's two-sample t-test) (Figure 3-5 B, D) suggesting that it likely maintains the role of phototransduction pathway activator in these species. As the non-phototransduction variant in *D. melanogaster* has been shown to be involved in a number of sensory biology pathways, I expected to see high expression of this isoform in both heads and bodies. Expression of the non-phototransduction isoform in males and females of *D. melanogaster* was not significantly different between both the head and body ($P = 0.10$; Welch's two-sample t-test) (Figure 3-5 A, C). Compared to the transcript abundance observed in *D. melanogaster* however, there was significantly lower expression of the non-phototransduction duplicate gene in both the heads and bodies of *D. mojavensis* and *D. virilis* males and females ($P = 0.0001$; Welch's two sample t-test) (Figure 3-5 A,C). The significantly reduced expression of this duplicate gene compared to its complementary splice-variant suggests a possible divergence in expression patterns. Since I only assayed expression from heads and bodies of adult males and females, it is possible that this duplicate copy is expressed at earlier stages of development or in a very specific tissue. Fine-scale developmental and tissue-specific transcriptional profiling is necessary to

determine more precise expression profiles for the non-phototransduction duplicate gene in *D. mojavensis* and *D. virilis.*

iii. Sex-biased gene expression of the *G-alpha-q* gene in *Drosophilids*

Transcriptional profiling of the duplicate genes also revealed two interesting patterns of sex-bias. Firstly, females of both *D. mojavensis* and *D. virilis* express the phototransduction-specific duplicate at two-fold higher levels than males whereas the reverse is true between *D. melanogaster* males and females (P = 0.009; Welch's two-sample t-test) (Figure 3-5B,D). Genes exhibiting reversals of sex-bias have been previously reported and their proportion seems to increase as the evolutionary distance between species increases (Metta et al. 2006; Grath et al. 2009). These types of genes are very interesting as they suggest switches in the direction of sexually antagonistic selection during species divergence (Ellegren and Parsch, 2007). Secondly, I observed significantly higher expression of the non-phototransduction duplicate gene in the heads of *D. mojavensis* females than in both *D. mojavensis* males (P =0.008; Welch's two-sample t-test) and *D. virilis* females (P=0.02;Welch's two-sample t-test) suggesting the divergence in expression is likely an example of species-specific female-biased gene expression (Figure 3-5 C). In contrast to *D. virilis*, which is a generalist species, both *D. mojavensis* males and females feed exclusively on rotting cactus in the deserts of North America (Markow and O'Grady, 2007). *D. mojavensis* females also oviposit on this substrate (Newby and Etges, 1998). Although more detailed experimental work is necessary to make this determination, I hypothesize the higher expression of the non-phototransduction duplicate in *D. mojavensis* females may be due to increased expression of this gene in chemosensory (olfactory or gustatory neurons) organs for reasons related

to oviposition. Evidence to support this hypothesis includes a report that *G-alpha-q* may interact with the GPCR taste receptor, *tre* in *D. melanogaster* (Isono et al. 2002, Kain et al. 2008). It has been shown that in *D. melanogaster,* taste perception is strongly sexually dimorphic and plays a role not only in pheromone perception but also in oviposition (Clyne et al. 2000; Bray and Amrein, 2003; Joseph et al. 2009; Swarup et al. 2014). In sum, I show evidence of substantial sex-bias gene expression of both duplicate genes. These results further illustrate the importance of taking sex-biased and sex-specific gene expression into account in any study that examines any sort of species divergence.

III. DISCUSSION

Alternative splicing and gene duplication, as the major sources of protein diversity, both contribute to the complexity of genomes. While it is generally thought that these two processes are independent of one another, here I show evidence of a single alternatively spliced ancestral gene that gave rise to duplicate copies that contain the exon complement of mutually exclusive splice variants. The *G-alpha-q* subunit of the G-protein in *D. melanogaster* is alternatively spliced in a tissue specific manner. One splice variant is expressed in the photoreceptors and another is expressed in a number of chemosensory organs (Lee et al. 1990; Talluri et al. 1995; Alvarez et al. 1996). In striking contrast, in the species belonging to the Drosophila subgenus, the orthologous *G-alpha-q* gene underwent a complex lineage-specific isoform duplication event ~ 50-60 MYA.

According to theory, duplicate genes generated via isoform duplication events are likely maintained as a result of sub-functionalization – where each duplicate accumulated degenerative mutations such that combined they perform the ancestral function. An

extension to this theory has been proposed suggesting that after prolonged sub-functionalization, a gene may change its interacting partners or expression patterns and undergo neo-functionalization – a process called sub-neofunctionalization (He and Zhang 2005). These isoform duplication events afford the new genes the opportunity to free up constraint and disengage from alternative splicing to optimize their function in a new tissue or to alter their expression patterns.

Thus, two possible evolutionary fates exist for duplicate genes that arise from isoform duplication events, sub-functionalization and sub-neofunctionalization. I explored this hypothesis by comparing the expression patterns of the alternatively spliced isoforms in *D. melanogaster* with the isoform duplicates in *D. mojavensis*, In the case of the *G-alpha-q* duplicates, if the locus has undergone subfunctionalization, we expect that the mRNA transcript abundance of the duplicates should conform to the pattern of the alternatively spliced gene.

Here, while I showed evidence that expression of the phototransduction specific duplicate was by and large analogous to its corresponding alternatively spliced variant in *D. melanogaster*, there was significant difference in expression between the non-phototransduction specific duplicates and their corresponding splice variant in *D. mojavensis* and *D. virilis* males and females. As the non-phototransduction splice variant in *D. melanogaster* is expressed in multiple tissues, we might expect that its corresponding duplicate would be subject to sub-neofunctionalization and would exhibit a novel expression pattern. Indeed, I found that the non-phototransduction duplicate is only expressed at moderately high levels in the heads of *D. mojavensis* females and not in *D. mojavensis* males or *D. virilis*. Moreover, the phototransduction specific duplicate in

both *D. mojavensis* and *D. virilis* also revealed a female-biased expression pattern. This pattern of female-biased gene expression, which appears to be a switch from *D. melanogaster's* male-biased expression, is another example of sub-neofunctionalization. Together, these results indicate that gene duplication followed by sub-neofunctionalization can provide an evolutionary substitute to alternative splicing as a way to produce unique tissue-specific protein products, even in closely related species.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

Functional work:

The functional consequences of the duplicate genes in *D. mojavensis* and *D. virilis* remain to be investigated. New genome editing technologies like the CRISPR-Cas9 system (Gratz et al. 2013) might make it possible to determine the functional consequences of the isoform duplication event in these non-model *Drosophila* species. Potential experiments include knocking out the non-phototransduction duplicate in *D. mojavensis* and *D. virilis* to determine phenotypic consequences. In addition, the non-phototransduction duplicate can be inserted into *D. melanogaster* to determine if it can function in the role of phototransduction pathway activator.

Genomics/Bioinformatics:

This isoform duplication event is novel and appears to be rare. Thus far, only a few examples of so-called isoform duplications have been shown in highly diverged species and an analysis examining how widespread this pattern of gene structure evolution among diverse phyla has not been carried out. It might be that as the phylogenetic distance between species increases, the percentage of isoform duplication

events also increases. Comparative genomic analyses on phylogenetically distinct taxa with reliable gene structure information and expression data are needed to determine the breadth and frequency with which these sorts of duplication events occur.

V. MATERIALS & METHODS

i. Fly species and tissue preparation

D. mojavensis (CI 12 IB-4 g8) and *D. virilis* (15010-1051.87) were obtained from University of California San Diego Drosophila Stock Center (San Diego, CA; http://stockcenter.ucsd.edu). All species were reared in wide fly vials (Genesee Scientific, San Diego, CA) at 25°C in a controlled 12h:12h light:dark environment with UVA/UVB bulbs (Zoo Med Laboratories, San Luis Obispo, CA) for several generations until 80–100 flies were available for RNA extraction. *D. virilis* was reared on standard cornmeal media and *D. mojavensis* was reared on banana-based media containing *Opuntia* cactus powder (UC San Diego Drosophila Stock Center). Adult (2-5 day old) male and female flies were used for all experiments. 100 males and 100 females of *D. mojavensis and D. virilis* were transferred to transparent 15ml tubes (BD Biosciences, Carlsbad, CA), 10–15 flies per tube, midway (6h) through the day light cycle. Tubes were returned to the light chamber for 1h for CO_2 anesthesia recovery and then flash-frozen *in situ* by rapid submersion in liquid nitrogen. All fly handling was done using CO_2.

ii. mRNA Expression Quantification

Absolute quantification of mRNA transcript abundance was determined for duplicate genes of *Galpha49B* using real-time quantitative PCR (rt-qPCR) using duplicate-specific primers to *G-alpha-q-1* and *G-alpha-q-2* in two species, dmoj_GLEANR_5329 and

dmoj_GLEANR_5330 in *D. mojavensis* and dvir_GLEANR_7621 and dvir_GLEANR_7622 in *D. virilis*. Primers: *G-alpha-q-1*: Forward: CGTGTGCCGACAACGGGGA Reverse: GCGCAGTATGAATTCACGGG. *G-alpha-q-2:*Forward: CAGGATATTCTGCGTGCTCGTGT Reverse: CGTGGTTCCAAAACTCATCGAT). Primers specific to the *D. melanogaster* isoforms were also designed. Primers: Dmel_NonPhotoTrans: Forward: GCAATAACGGCCCGAGAGTTTA. Reverse: GGTTCGATTGCAGAATTGTGTCC. Primers for Dmel_isoform_PhotoTrans : Forward: CAAAACAGGATCACGCTGCG Reverse: GTATCTTTGACAGCGCAG. Total RNA extractions of both head and body (*n* = 50) were performed using the SV Total RNA Isolation system (Promega, Madison, WI) following the procedure of Landry, Wittkopp *et al.* (2005). Two RNA extractions were performed per sample. 500ng total RNA per extraction from each of 2 reverse transcription cDNA syntheses (Qiagen Quantitect Reverse Transcription Kit, Valencia, CA) using both poly(dT) and random hexamers, for a total of four reverse transcription reactions per experiment. cDNA products from the 4 reactions were pooled and diluted 1:10. 5 µl of the 1:10 dilution of cDNA was used for each rt-qPCR reaction. DNA template controls of known concentration were created in order to quantify absolute transcript abundance as follows: PCR products for each duplicate gene were individually cloned into plasmids and transformed into *E. coli* using the Topo-TA Cloning kit (Invitrogen, Carlsbad, CA). Next, DNA concentration of linearized plasmid in ng/µl was determined in quadruplicate using a low-volume spectrophotometer (NanoDrop, Wilmington, DE). The absolute number of plasmid copies per µl was then determined by calculating the mass of each plasmid + amplicon for each gene, using the formula:

amplicon mass = total number of base pairs × nucleotide mass (660 g/mol of nucleotides). Dilutions (1:10) of linearized amplicon-incorporated plasmids were then used to construct standards for each gene. The mean range of standards across genes was 1000–1 billion molecules per µl. Sample transcript abundances were calculated by interpolation of standard values. Both sample and control cDNA for each gene was analyzed in 96-well plate format on the Roche Lightcycler 480 instrument (Indianapolis, IN). cDNA for all treatments was stored at 80°C.

iii. Bioinformatics Analyses

G-alpha-q CDS in *D. melanogaster* (annotated *Galpha49B*) was obtained from Flybase (www.flybase.org) and used to BLAST against the *D. mojavensis*, *D. virilis*, *D. grimshawi* genomes to determine putative *G-alpha-q* orthologs. Initial BLAST search suggested more than one *G-alpha-q* ortholog in *D. mojavensis*, *D. virilis* and *D. grimshawi*. A reciprocal BLAST search of this gene was conducted for each genome against itself to rule out false positives. CDS for duplicate genes in *D. mojavensis*, *D. virilis* and *D. grimshawi* were translated and aligned using Clustalw (Larkin et al. 2007). G-alpha-q gene structures for *Anopheles gambiae* and *Bombyx mori* were obtained from VectorBase and Silkworm respectively (Megy et al. 2012; Wang et al. 2005).

Cross-species duplicate specific primers were designed for the duplicate genes in *D. mojavensis* and *D. virilis*. The primers were BLAST against both genomes to ensure they would not amplify any other region. Splice-variant specific primers were also designed for *D. melanogaster*. Those primers were also BLAST against the *D. melanogaster* genome to ensure it would not amplify any other region.

APPENDIX I

Table A1 GO term enrichments from male and female-biased genes

Female-biased Genes in *D. pseudoobscura* enriched for the following GO terms

Items	GO: Biological Process	*P*	P_{adj}
GO:0022008	neurogenesis	1.94E-67	3.70E-64
GO:0007052	mitotic spindle organization	8.48E-29	8.09E-26
GO:0000398	nuclear mRNA splicing, via spliceosome	3.32E-23	2.11E-20
GO:0048477	oogenesis	8.18E-20	3.90E-17
GO:0007067	mitosis	1.40E-17	5.35E-15
GO:0006367	transcription initiation from RNA polymerase II promoter	1.71E-13	5.44E-11
GO:0006260	DNA replication	5.76E-13	1.57E-10
GO:0000381	regulation of alternative nuclear mRNA splicing, via spliceosome	1.50E-12	3.58E-10
GO:0007095	mitotic cell cycle G2/M transition DNA damage checkpoint	2.25E-12	4.77E-10
GO:0007049	cell cycle	1.05E-10	2.00E-08
GO:0042254	ribosome biogenesis	1.77E-09	3.08E-07
GO:0006357	regulation of transcription from RNA polymerase II promoter	2.24E-09	3.57E-07
GO:0006281	DNA repair	7.06E-09	1.04E-06
GO:0006325	chromatin organization	1.37E-08	1.87E-06
GO:0006364	rRNA processing	2.20E-08	2.80E-06
GO:0051276	chromosome organization	2.36E-08	2.82E-06
GO:0007306	eggshell chorion assembly	4.42E-08	4.97E-06
GO:0006342	chromatin silencing	3.82E-07	2.92E-05
GO:0008360	regulation of cell shape	8.44E-07	6.19E-05
GO:0007076	mitotic chromosome condensation	9.64E-07	6.82E-05
GO:0006911	phagocytosis, engulfment	1.03E-06	7.02E-05
GO:0007307	eggshell chorion gene amplification	1.07E-06	7.07E-05
GO:0006974	response to DNA damage stimulus	1.18E-06	7.53E-05

Male -Biased in *D. pseudoobcura* enriched for following GO terms:

GO: ID	Biological Process	*P*	P_{adj}
GO:0055114	oxidation-reduction process	6.24E-09	6.87E-06
GO:0018401	peptidyl-proline hydroxylation to 4-hydroxy-L-proline	0.00028055	0.0257405
GO:0006508	proteolysis	2.22E-07	0.000122245
GO:0006629	lipid metabolic process	1.64E-06	0.00060093
GO:0007602	phototransduction	0.000135499	0.021312
GO:0005975	carbohydrate metabolic process	0.000243327	0.0267903
GO:0006626	protein targeting to mitochondrion	0.00015414	0.0212135
GO:0007018	microtubule-based movement	0.000495237	0.0363504
GO:0006096	glycolysis	4.85E-06	0.00133517
GO:0019731	antibacterial humoral response	0.000700368	0.0481941
GO:0006099	tricarboxylic acid cycle	7.72E-05	0.0170017
GO:0006810	transport	0.000259282	0.0259517
GO:0030317	sperm motility	0.000371249	0.0291961
GO:0006959	humoral immune response	0.000296754	0.0251327
GO:0046692	sperm competition	8.86E-05	0.0162569
GO:0006006	glucose metabolic process	0.000172803	0.0211396

Female-Biased in *D. persimilis* enriched for the following GO terms

GO:ID	Biological Process	*P*	P_{adj}
GO:0022008	neurogenesis	9.21E-54	1.91E-50
GO:0007052	mitotic spindle organization	5.00E-26	5.17E-23
GO:0048477	oogenesis	3.12E-16	2.15E-13
GO:0007067	mitosis	8.21E-15	4.24E-12
GO:0000398	nuclear mRNA splicing, via spliceosome	9.93E-14	4.11E-11
GO:0006367	transcription initiation from RNA polymerase II promoter	4.82E-11	1.66E-08
GO:0042254	ribosome biogenesis	6.64E-10	1.96E-07
GO:0000022	mitotic spindle elongation	4.14E-09	1.07E-06
GO:0006260	DNA replication	5.86E-09	1.35E-06
GO:0000381	regulation of alternative nuclear mRNA splicing, via spliceosome	8.76E-08	1.81E-05
GO:0006270	DNA-dependent DNA replication initiation	1.02E-07	1.92E-05
GO:0007095	mitotic cell cycle G2/M transition DNA damage checkpoint	1.59E-07	2.75E-05
GO:0007306	eggshell chorion assembly	1.04E-06	0.000143884
GO:0045793	positive regulation of cell size	1.41E-06	0.00018224
GO:0006325	chromatin organization	1.60E-06	0.000194784
GO:0045214	sarcomere organization	3.37E-06	0.000366623
GO:0007305	vitelline membrane formation involved in	3.35E-06	0.000385322

	chorion-containing eggshell formation		
GO:0006261	DNA-dependent DNA replication	5.72E-06	0.000563966
GO:0008360	regulation of cell shape	5.55E-06	0.000574481
GO:0006366	transcription from RNA polymerase II promoter	6.42E-06	0.000604209
GO:0006396	RNA processing	1.07E-05	0.000923063
GO:0006364	rRNA processing	1.07E-05	0.000958695
GO:0006468	protein phosphorylation	2.09E-05	0.00172814
GO:0007076	mitotic chromosome condensation	2.73E-05	0.00217622
GO:0046331	lateral inhibition	3.37E-05	0.00248829
GO:0006357	regulation of transcription from RNA polymerase II promoter	3.33E-05	0.00255029
GO:0007184	SMAD protein import into nucleus	5.12E-05	0.00365179
GO:0006974	response to DNA damage stimulus	6.70E-05	0.00420266
GO:0007317	regulation of pole plasm oskar mRNA localization	6.10E-05	0.00420386

Male *D. persimilis* enriched for the following GO terms

GO:ID	**Biological Process**	***P***	P_{adj}
GO:0055114	oxidation-reduction process	1.01E-06	0.00110569
GO:0006626	protein targeting to mitochondrion	9.62E-06	0.00150163
GO:0006096	glycolysis	9.62E-06	0.00150163
GO:0005975	carbohydrate metabolic process	4.74E-06	0.00172852
GO:0051603	proteolysis involved in cellular protein catabolic process	9.22E-06	0.00201494
GO:0007018	microtubule-based movement	4.28E-06	0.0023367
GO:0007602	phototransduction	8.94E-06	0.00244286
GO:0007140	male meiosis	4.80E-05	0.00582726
GO:0046692	sperm competition	4.37E-05	0.00596928
GO:0007291	sperm individualization	6.61E-05	0.00722309
GO:0006006	glucose metabolic process	9.96E-05	0.00989561
GO:0018401	peptidyl-proline hydroxylation to 4-hydroxy-L-proline	0.000132328	0.0120529
GO:0030317	sperm motility	0.000201069	0.0169053
GO:0006511	ubiquitin-dependent protein catabolic process	0.000420841	0.0306653
GO:0007283	spermatogenesis	0.000405994	0.0316965
GO:0006959	humoral immune response	0.000679248	0.0436717
GO:0006508	proteolysis	0.000644141	0.0440029

TABLE A2: GO analyses for genes that switched from female-to-male bias

GO:IDs	Biological Process	*P*	P_{adj}
GO:0055114	oxidation-reduction process	0.000200067	0.00648789
GO:0006629	lipid metabolic process	0.00417904	0.0499286
GO:0007602	phototransduction	0.000104103	0.0047263
GO:0016059	deactivation of rhodopsin mediated signaling	1.93E-05	0.00146254
GO:0005975	carbohydrate metabolic process	0.000743923	0.0187634
GO:0045214	sarcomere organization	3.38E-06	0.00038373
GO:0007519	skeletal muscle tissue development	0.000585024	0.0166
GO:0042049	cellular acyl-CoA homeostasis	0.00297283	0.0449888
GO:0070588	calcium ion transmembrane transport	0.00297283	0.0449888
GO:0006006	glucose metabolic process	5.99E-05	0.00339956
GO:0050962	detection of light stimulus involved in sensory perception	1.77E-06	0.000402784
GO:0030239	myofibril assembly	0.00393211	0.0495882
GO:0030845	inhibition of phospholipase C activity involved in G-protein coupled receptor signaling pathway	0.000147373	0.0055756
GO:0016062	adaptation of rhodopsin mediated signaling	0.000870104	0.0179558
GO:0006071	glycerol metabolic process	0.00393211	0.0495882
GO:0006874	cellular calcium ion homeostasis	0.00393211	0.0495882
GO:0060537	muscle tissue development	0.000870104	0.0179558
GO:0006936	muscle contraction	0.00297283	0.0449888
GO:0006497	protein lipidation	0.00143856	0.0272127

TABLE A2: GO analyses for genes that switched from male-to-female bias

GO ID	GO: Biological Process	*P*	P_{adj}
GO:0007052	mitotic spindle organization	7.90E-06	0.000893054
GO:0001556	oocyte maturation	1.90E-05	0.00107626
GO:0006959	humoral immune response	1.98E-05	0.000747362
GO:0007056	spindle assembly involved in female meiosis	3.80E-05	0.00107453
GO:0000212	meiotic spindle organization	0.000132484	0.00299414
GO:0050830	defense response to Gram-positive bacterium	0.000142535	0.00268441
GO:0022008	neurogenesis	0.000452698	0.00730783
GO:0007018	microtubule-based movement	0.00175849	0.0248387
GO:0019731	antibacterial humoral response	0.00198849	0.0249666
GO:0007067	mitosis	0.00203967	0.0230483
GO:0016321	female meiosis chromosome segregation	0.00300559	0.0178754
GO:0007049	cell cycle	0.00400342	0.0215422
GO:0050829	defense response to Gram-negative bacterium	0.00400342	0.0215422
GO:0042742	defense response to bacterium	0.00421905	0.0216706
GO:0007147	female meiosis II	0.00764789	0.0288071
GO:0046594	maintenance of pole plasm mRNA location	0.00764789	0.0288071
GO:0006446	regulation of translational initiation	0.0101845	0.0319681
GO:0007144	female meiosis I	0.0101845	0.0319681
GO:0046012	positive regulation of oskar mRNA translation	0.0101845	0.0319681

Table A3

Genes with *D. melanogaster* orthologs that switch from M-F bias in both *D. pseudoobscura* and *D. persimilis*

***D. pse* gene name**	***D. mel* gene name**	***D.mel* FBGN**
Dpse\GA13566	CG15202	FBgn0030271
Dpse\GA14834	CG18190	FBgn0034403
Dpse\GA20032	RpS5b	FBgn0038277
Dpse\GA26658	wisp	FBgn0260780

Genes with *D. melanogaster* orthologs that switch from F-M bias in both *D. pseudoobscura* and *D. persimilis*

***D. pse* gene name**	***D. mel* gene name**	***D.mel* FBGN**
Dpse\Est-5B	Est-6-PA	FBgn0000592
Dpse\GA27207	AttA-PA	FBgn0012042
Dpse\GA28633	CG14096	FBgn0036871
Dpse\GA23745	Aats-asn-PA	FBgn0086443
Dpse\GA23746	Aats-asn-PA	FBgn0086443
Dpse\GA15129	sls-PX	FBgn0086906
Dpse\GA29314	Rdl	FBgn0004244
Dpse\GA18512	Pdh	FBgn0011693
Dpse\GA10902	TM4SF	FBgn0020372
Dpse\GA19674	fau	FBgn0020439
Dpse\GA25973	CG16704	FBgn0031558
Dpse\GA12705	CG14022	FBgn0031700
Dpse\GA10180	CG10237	FBgn0032783
Dpse\GA21243	nrv3	FBgn0032946
Dpse\GA11590	CG12374	FBgn0033774
Dpse\GA14498	CG17386	FBgn0033936
Dpse\GA10635	CG10911	FBgn0034295
Dpse\GA10968	CG11388	FBgn0034959
Dpse\GA13274	CG14823	FBgn0035734
Dpse\GA23704	CG6910	FBgn0036262
Dpse\GA12764	CG14109	FBgn0036364
Dpse\GA18419	CG4770	FBgn0038751
Dpse\GA24253	CG16926	FBgn0040732
Dpse\GA26199	Pif1B	FBgn0046874
Dpse\GA24366	Unc-89	FBgn0053519
Dpse\GA25953	CG34172	FBgn0085201
Dpse\GA24888	CG34225	FBgn0085254
Dpse\GA27541	CG42508	FBgn0260234
Dpse\GA23436	CG43897	FBgn0264489

APPENDIX II

Reconstructed Phototransduction Sequences

>**Dana_sun**
ATGGAGAAGTCGTTCTCCATAACGCCGTGGAAGTACGGCCTACTAGCCACTTGCATTCTCATTGTGACATG
TAATGTGTTCTTCTTTTCTTGCGGCGTTGTGACCTGGGGATCGGCAGTATCTCGCTCCGGGAGCTATTCGG
CAGCTCTATGTGGAGTGTCTGTCTTTGGAGTGGCCTTTCTCGGAATGTACGTGTCCCTGAAGGAGTCCCAT
AAGTATTCCGTATACTACTTGATATGCGGTGGATTGGTGATATCCGCCTTGGGTGCCTATCTTTTTGTCTT
CACGGCGATGCGGACGGAGCTGATGGGCCGGTTCGAGGAGCGCATGCGTCGCCTGTTTGCGGAGAAGACCC
ACCACGATGACAAGATGCAGCCCGTGCATAGCCTCTTCGGGTGCTGCGGAATGGATGGACCGCAGGACTAT
CTAAGGACGGAGCACGGAGCTCTGCCCAGCAGCTGTTGCTATGCCTTCGACTGCTCTAATCCCAGTCATGT
TTACGAAGAGGGCTGCTCCACCAAAGCAATAACAAACCTTAGGATGCAGGCTGATCTAAACTATTATAGCT
GCCTGGCCATTATTTCCCTCGAGTTCATGGGCTTGTTCACTGCTTACTACATGGGCAAGGCCCGGAAGCAC
GCCAAAACCAAGGTGAAGGACGAGGAGACTCCCATGAACGATTAA

>Dgri_Galpha_1
ATGGAGTGCTGTTTATCGGAAGAAGCCAAGGAACAAAAGCGGATCAATCAGGAAATCGAGAAACAGCTGCG
CCGCGACAAGCGAGATGCGCGACGTGAGCTCAAGCTCCTCCTACTGGGCACGGGCGAGTCCGGCAAATCAA
CGTTCATCAAACAGATGCGTATTATACACGGCAGCGGGTACTCGGACGATGACAAACGTGGATACATTAAA
CTGGTTTTCCAAAACATATTCATGGCCATGCAGTCCATGATCAAAGCCATGGATATGCTCAAAATATCATA
TGGTGTTGGCGAGCATAACGGACTTGCTGATCTGGTGATGAGCATTGACTATGAAACAGTGACAACATTTG
ATGATCCATATTTGAATGCCATAAAAACGCTTTGGTCCGATGCTGGCATACAGGAATGTTACGATCGCCGA
AGAGAATATCAGTTGACAGATTCCGCTAAATATTATCTGATGGATCTCGATCGTGTGGCTCAACCTGACTA
TTTACCCACTGAGCAAGACATTTTAAGAGTTCGTGTGCCGACAACGGGGATCATTGAGTATCCCTTTGACT
TAGAAGAGATCAGATTTAGAATGGTTGATGTCGGCGGTCAGCGTTCGGAGCGAAGAAAGTGGATTCATTGC
TTTGAGAATGTTACCTCAATAATCTTTTTGGTAGCGCTATCTGAATACGATCAAATTTTGTTTGAATCTGA
TAATGAGAACCGCATGGAGGAATCAAAAGCTTTATTTAGAACTATTATAACATACCCGTGGTTCCAAAATT
CATCAGTTATTCTATTTCTTAACAAGAAAGATTTGTTAGAGGAGAAAATAATGTATTCGCATTTGGTGGAC
TATTTTCCTGAATACGATGGTCCACAGCGAGATGCAATAGCAGCCCGTGAATTCATACTGCGCATGTTTGT
AGACTTAAATCCAGATTCCGAAAAGATTATCTATTCTCATTTCACCTGTGCTACGGATACTGAAAATATAA
GGTTTGTGTTTGCAGCTGTAAAGGACACGATTCTGCAATCCAATCTTAAGGAATACAATTTGGTCTAA
>**Dgri_Galpha_2**
ATGGAGTGCTTATTATCGGAGGAAGCCAAGGAACAAAAGCGAATCAATCAGGAAATCGAGAAACAGCTGCG
TCGTGATAAGAAAGATGCCAGACGTGAACTCAAGCTACTTCTACTAGGCACTGGTGAATCTGGAAAATCGA
CCTTCATTAAACAGATGCGTATTATCCACGGTAGTGGTTATTCAGATGACGACAAGCGTGGTTATATTAAA
CTGGTGTTCCAAAACATATTCATGGCCATGCAGTCCATGATCAAAGCCATGGATATGCTCAAAATAGAATA
TGGTGGTCCAGAGATTCAAGAATACGGTGAACTGGTGATGAGCATTGATTATGAATCGGTTACACGACTTG
AGGATCCCTATTTAAATGGTATCAAAGGCCTCTGGGCCGATACTGGAATACAGGAGTGCTATGATCGCCGA
AGAGAATATCAGTTGACAGATTCAGCTAAATATTATCTGAGCGATTTAGCGCGCATTGAACTGGAAGACTA
TTTGCCCACTGAGCAGGATATTCTGCGTGCTCGTGTTCCCACGACTGGAATACTTGAGTATCCCTTTGATT
TGGACGGCATTGTTTTCAGAATGGTTGACGTCGGAGGTCAGCGTTCAGAGAGAAGAAAGTGGATTCATTGC
TTTGAAAATGTTACTTCTATTATATTTTTGGTGGCGCTATCTGAATACGATCAATGCTTGATGGAATCTGA
TGAGAATCGTATGGAAGAATCAAAGGCTTTATTTAAAACTATCATTACATACCCGTGGTTCCAAAACTCAT
CGATTATTCTCTTTATGAACAAGAAAGATTTGTTTGAAGAGAAAATAATGTATTCGCATATGGGTGACTAT
TTTCCAGAGTTCGATGGTCCAAAACAGGACCATGCCACGGCTAAGCAATTTGTTCTCAAAAAATACTTGGC
ATGCAATCCAGATCCAGAGCGTCAGTGCTATTCGCATTTCACAACAGCCACAGATACCGAAAACATCAAAC
TTGTGTTCTGCGCTGTCAAAGATACAATTATGCAAAATGCACTTAAGGAATTCAATTTGGGTTAA
>Dmoj_Galpha_1
ATGGAGTGCTGTTTATCGGAAGAGGCCAAGGAACAAAAGCGGATCAATCAGGAAATCGAGAAACAGCTGCG
CCGTGACAAGCGAGATGCGCGACGTGAACTTAAACTGCTCCTACTGGGCACGGGCGAGTCCGGCAAATCAA
CGTTCATCAAACAGATGCGTATTATCCATGGCAGCGGGTACTCGGACGAGGACAAACGCGGCTACATTAAG
CTGGTGTTCCAAAATATATTCATGGCCATGCAGTCAATGATCAAAGCCATGGATATGCTCAAAATATCATA
CGGCGTTGGCGAGCATAATGCGCTCGCTGATCTGGTGATGAGCATTGACTATGAAACAGTTACAACATTTG

AGGATCCCTATCTAAATGCCATAAAAACGCTGTGGTCTGATGCTGGCATACAGGAGTGCTACGATCGCCGA
AGAGAATATCAGTTGACCGACTCCGCTAAATATTATCTGATGGATCTCGATCGTGTGGCTCAACCTGATTA
TTTACCCACTGAACAAGACATTTTAAGAGTGCGTGTGCCGACAACGGGGATCATTGAGTATCCCTTTGACT
TAGAAGAGATCAGATTTAGAATGGTGGATGTCGGTGGTCAGCGCTCGGAGCGAAGGAAGTGGATTCATTGC
TTTGAGAATGTTACCTCAATAATATTTTTGGTTGCACTATCCGAATACGATCAAATCTTGTTTGAATCTGA
TAATGAGAACCGCATGGAGGAATCAAAAGCTTTATTTAGAACTATTATTACATACCCGTGGTTCCAAAATT
CATCAGTAATTCTATTCCTTAACAAGAAGGATTTGTTAGAGGAGAAAATTATGTATTCGCATTTGGTGGAC
TATTTTCCTGAATACGATGGTCCACAGCGAGATGCAATCGCAGCCCGTGAATTCATACTGCGCATGTTTGT
AGACTTAAATCCAGATTCCGAAAAGATTATCTATTCTCATTTCACCTGTGCTACAGATACTGAAAATATAA
GGTTTGTGTTTGCAGCTGTAAAGGACACGATTCTGCAATCCAATCTTAAGGAATACAATTTGGTCTAA

>Dmoj_Galpha_2

ATGGAGTGCTGCTTATCGGAAGAGGCCAAAGAACAGAAGCGTATCAATCAGGAAATTGAAAAACAGCTACG
CCGTGATAAGAAAGATGCTCGTCGCGAACTTAAACTGCTTTTACTGGGTACTGGTGAATCTGGGAAATCAA
CTTTTATTAAGCAAATGCGTATTATTCACGGCAGCGGCTACTCGGACGATGACAAGCGTGGCTATATTAAG
CTGGTGTTCCAAAATATATTCATGGCCATGCAGTCAATGATCAAAGCCATGGATCTGCTCAAAATAGAATA
TGGCGCACCAGAGATTCGCGAATGTGGTGAGCTGGTGATGAGCATTGACTACGAATCTGTTACAAGATTGG
AGGATCCATTTTTGAAGGCCCTACATTCTCTTTGGGACGATGCTGGCATACAGGAGTGCTATGATCGCCGA
AGAGAATACCAGTTAACCGATTCTGCTAAATACTATCTGAGCGATTTGGCGCGTATTGAACAGGAAGATTA
TTTGCCCACTGAACAGGATATTCTGCGTGCTCGTGTGCCCACAACTGGAATACTGGAGTATCCATTCGATT
TGGATGGCATCGTCTTCAGAATGGTGGACGTTGGCGGTCAGCGTTCGGAAAGAAGAAAGTGGATTCATTGC
TTTGAGAATGTTACATCGATTATATTCTTGGTGGCGCTCTCTGAATATGATCAGATCTTGTTTGAGTCTGA
CAACGAGAATCGTATGGAGGAATCGAAAGCTTTGTTCAGAACCATTATTACATACCCGTGGTTCCAAAACT
CATCGATCATTCTATTTATGAACAAGAAAGATTTGTTCGAAGAGAAAATTATGTATTCGCATATGGGTGAC
TATTTTACTGAATATGATGGTCCAAAACAAGATCACGCAGCAGCGAAGCAGTTTGTGTTGAAGAAATACCT
GGCCTGCAATCCGGATCCAGAGCGTCAATGTTATTCGCATTTCACCACAGCCACAGATACCGAAAACATCA
AGCTTGTGTTCTGCGCTGTCAAAGATACAATTATGCAAAATGCACTTAAGGAATTCAATTTGGGTTAA

>Dper_ninaC

ATGTATTTACCGTATTCGCAATTGCCGGATCCAACGGATAAGTTCGAGATCTATGAAGAAATAGCCCAGGG
TGTCAATGCCAAAGTATTTCGGGCCAAGGAATTGGAGACCGATCGTCTGGTGGCATTAAAGATACAGCACT
ACGATGAAGAGCATCAGGTGTCCATCGAGGAGGAGTATCGTACTTTGCGAGACTATTGTGCCCACCCCAAT
CTCCCGGAGTTCTATGGCGTGTACAAGCTATCGAAGCCGAATGGTCCTGATGAGATTTGGTTCGTTTTGGA
GTACNNNATANNNTCCTTACAGTACTGTGCTGGAGGAACAGCCGTAGATATGGTTAATAAGCTCCTCAAAC
TGGACCGTAGGATGAGAGAGGAGCACATCGCCTACATCATCCGGGAGACGTGTCGGGCAGCCATCGAGCTG
AATCGCAATCATGTCCTGCATCGGGACATCCGTGGCGACAACATTTTGCTCACGAAAAATGGACGGGTGAA
GCTCTGCGACTTCGGTCTGTCCCGTCAAGTGGATTCGACTTTGGGGAAGCGAGGCACCTGCATTGGTTCGC
CGTGCTGGATGGCCCCCGAAGTGGTGGCTGCCATGGAGTCCCGCGAGCCCGATATCACCGTTCGGGCAGAT
GTCTGGGCCCTGGGTATCACCACCATTGAGTTGGCCGATGGAAAGCCCCCGTTTGCTGACATGCATCCCAC
TCGAGCCATGTTCCAGATCGTGCGGAATCCACCGCCGACACTGATGCGACCCACGAATTGGTCCCAGCAGA
TCAATGACTTTATTGCCGAGTGTCTGGAGAAGAGTGCCGAGAATCGCCCCATGATGGTCGAGATGGTAGAG
CATCCATTCCTCACGGAGCTCATTGAGAACGAGCAGGAGATGCAGTCGGATCTTGCAGAGATGCTGGAGCT
TTCGAGGGATGTGAAGAGCCTGTACAAGGAGCCGGAACTATTTGTGGATCGCGGCTATATCAAGCGCTTCG
ATGGGAAGCCGGAGCGCATGTATCCTGAAGATTTGGCAGCCCTGGAAAACCCCGTGGATGAGAGCATCATT
GATTCGCTGCTAAATCGCATGATGACTGGTGAATCCTATAGTTTTATTGGAGATATCCTGCTGTCCCTGAA
TTCGAATGAGATCCAAAGCGAATTCCCTCTGCAGTTCCATGCCAAATACCGATTCAAGTCTCGCTCGGAGA
ACCAACCCCATATCTTCTCCGTGGCCGATATTGCCTACCAGGATATGCTCCATCATCGCGAGCCCCAGCAC
ATTGTCCTTTCTGGGGAGAGCTACTCGGGCAAGTCGACCAACGCCCGTCTGCTGATCAAGCATCTGTGCTA
TCTTGGGGCTGGCAATCGGGGTGCCACTGGGCGCGTGGAGACCTCCATTAAGGCCATCCTGATGATGGTCA
ATGCCGGCACTCCCGTCAACAATGACTCCACCCGCTGTGTCCTGCAGTATTGTCTGACTTTCGGCAAGACC
GGAAAGATGAGTGGAGCGGTATTCAATATGTATATGCTGGAGAAGCTGCGAGTGGCCACGTGTGACGGCAC
CCAGCACAATTTCCACATTTTCTACTATTTCTATGACTTCATCAACCAGCAGAATCAGCTGAAGGACTACA
ATCTGAAAGCGGACCGCAACTATCGCTATCTGCGCGTGCCACCAGATGCTCCCCCAACCAAGATCAAGTAT
CGTCGCGACGATCCCGAGGGTAATGTGGAGCACTACCGTGAGTTTGAGAACATTCTCAGGGATCTGGACTT
TAACCACAAGCAGCTGGAGACTGTGCGAAAGGTGCTGGCTGCCATCTTAAACATTGGAAACATTCGCTTCC
GACAGTCCGGCAAGTACGCCGAAGTCGAGAACACGGATATAGTATCGCGTATTGCCGAGCTCTTGAGGGTG
GATGAGAAGAAATTCATGTGGTCCTTAACAAACTTCATAATGGTCAAGGGAGGTGTGGCCGAAAGACGGCA
GTACACCGCAGAGGAGGCTCGCGATGCACGCGATGCTGTGGCCAGCACCATCTACTCCCGATTTGTGGACT
TTATTATCAACAGGATTAACATGAACATGTCGTTTCCAAGAGCTGTCNNNGGCGACACAAACGCCATCATT

ATCCATGATATGTTCGGCTTTGAGTGCTTCAATCGGAATGGCATGGAGCAGCTAATGATCAACACTCTGAA
CGAACAGATGCAATATCACTACAACCAACGCATTTTCATTAGCGAAATGCTCGAAATGGAGGCCGAGGATA
TAGACACATTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGATCGCGTTTC
GGAGAAACACAGCCAGTTTGTGAAGAAGCACACTGCCACTGAGATATCCGTTGCTCATTATACGGGCCGCA
TCATCTATGATACTCGCTCCTTCACGGACATCAACCGCGACTTTGTCCCCCCGGAAATGATTGAGACGTTC
CGCTCTTCGCTGGACGAGAGCATAATGCTCATGTTCACCAATCAACTGACGAAAGCTGGCAATCTGACAAT
GCCTTTTGAGGCTGTCCAGCATAAGGATGAAACTGAACGCAGGTCCTACGCCCTCAACACTTTGAGTGCTG
GATGCATCTCCCAGGTGAACAATTTACGTACTCTGTCGGCCAATTTCCGCTTCACCTGCCTGACTCTGCTC
AAAATGCTAAGCCAAAATGCCAGTCTCGGCGTGCACTTTGTCCGTTGCATACGAGCTGATTTGGAGTACAA
GCCGCGTGCGTTCCACTCTGATGTGGTGCAGCAACAGATGAAGGCTCTGGGTGTCCTGGACACGGTCATTG
CCCGGCAAAAGGGCTACAGCTCTCGACTGGGCTTCGATGAGTTCCTCAGACGGTATCAATTCCTAGCCTTT
GACTTTGATGAGCCTGTGGAAATGACAAAAGAGAATTGTCGTCTCTTATTTTTACGTCTCAAGATGGAGGG
TTGGGCCTTGGGCAAGTCCAAAGTCTTTTTGCGCTACTACAACGATGAGTTCTTAGCCAGGAATTACGAGC
TTCAAGTGAAGAAAGTCATCAAAGTGCAGTCCATGATGCGGGCACTTTTAGCCCGAAAACGTGTCAAAGGT
GGCAAGGTCTTTAAAGTGGGCAAAAAGGGTCAGGAGCATTCCGATTTGGCAGCTGCCAAAATACAAAAAGC
GTATCGCGGGTTCCGTGACCGGGTCCGTCTGCCGCCGCTGGTCAATGAGAAGACGGGCCAACGGAATGAGA
ACACCGCCGACTTTATTCGACCCTATGCCAAGAAATGGCGAGAGAAGTCCATCTTTCAGGTCCTGCTGCAC
TACAGAGCGGCCCGCTTCCAGGACTTTGTCAATCTCTCACAACAGGTAGTACATATTTACAACCAACGGAT
GGTCGGTGGGCTGAACAAATGCACCCGGTCGGTGCCCTTCGAGCGCATTAATATGCGGGAGGTAAACTCCT
CGCAGCTGGGACCCCTTCCTGTTCCGATGAAGAAGATGCCATTCCGCCTGGATCAGATACCCTTCTATGAC
ACACAGTACATGGTGGATCCGGCCAATTCTATATCTCGTCAAACGTTTCCCAATCAGTTGCTGGTGCAAAA
TCTGGAAGATGATGAAGCCTGGGATAGTCCTCTACAGCGCAATCCTTCGATGACTTCGTGTGCCATGACCT
ACAATGCTTATAAGAAAGAACAGGCTTGCCAGACCAATTGGGATCGTATGGGCGAAAGCGATAATATATAC
AATCAGGGGTTCTTTAGAGATCCACAGCAGTTAAGAAGGAGAAACCAAGTGCAACAAAACATGAATGCTTA
CAACAATGCCTATAACAACTACAACAGCAGCAACAATCCCAACTGGGGCCGCTCAGGATCCCGTCGTAATT
CCCTCAAGGGCTACGCGGCGCCACCGCCGCCGCCACCGATGCCTTCATCCAACAACTACTATCGAAATAAT
CCCAGTCAACAGCCTCAAAGGAACTATCAGCAACGCTCCTCGTATCCACCTTCAGATCCCATAAGAGAGCT
CCAGAACATGGCACGCAACGACGGAGACAACTCCGAGGATCCTCCATTCAACTTCAAGGCCATGCTGCGCA
AGACAAACTATCCAAGGGGCTCAGAGACCAGCACCTATGACTTTAACAATCGCCGGGGGTCGGACGCCGGG
GATCAGCAGCACACATTCCAGGCACCCAAACTACGCTCCACTGGACGCCGCTTCGAAGATGAGGGGGGCTA
TAAAACCTCGGCGGGCAACTATGGGGCTGCCCGCAACTTTGGCCAGCAGCGGGCTCCTCCTTTGAACCAGA
GCTCCTCCAGTGTGGGTCGCAGCTTTGAGAACAGCAATGCAAGGTCCTTCGAGCAGGCGGGCTCCTATGTT
GAAGAGGAGATTGCACCAGGGGTGACCCTATCCGGATATGCGGTGGACATATAA

>Dper_norpA

ATGACCAAGAAGTACGAGTTCGATTGGATCATCCCGGTGCCGCCGGAATTGACCACTGGCTGTGTCTTCGA
CCGTTGGTTCGAGAACGAGAAGGAGACCAAGGAGAATGACTTTGAGCGAGATGCCCTCTTCAAGGTCGATG
AATATGGGTTCTTTTTGTACTGGAAGAGCGAGGGTCGGGATGGCGATGTCATCGAGCTGTGCCAAGTGAGC
GACATTCGTGCGGGCGGCACTCCCAGGGATCCCAAGATACTGGATAAGGTGACCAAAAAGAATGGAACCAA
CATACCGGAGCTGGACAAGCGATCTCTCACTATTTGCTCGAACACGGATTATATCAATATAACATATCACC
ATGTTATCTGCCCCGATGCGGCAACGGCCAAGAGTTGGCAAAAGAATTTGCGTATCATAACGCATAACAAT
CGCGCCACAAATGTGTGTCCGCGCACCAATCTGATGAAGCAGTGGATGAGACTAAGTTATTGCGTGGAGAA
GAGCGGCAAGATACCGGTTAAGACGCTGGCCAAGACCTTCGCCTCGGGCAAGACCGAGAAGCTGGTCTACA
CGTGCATCAAGGATGCGGGTCTGCCAGACGACAAAAATGCCACCATGACCATGGATCAGTTCACTTTCGAT
AAGTTCTATGCCCTCTACCACAAGGTCTGTCCCCGTAACGACATCGAGGAGCTCTTCACGTCGATGACCAA
GGGTAAACAGGACTTTATCAGTCTCGAGCAGTTTATACAGTTCATGAACGAGAAGCAGAGGGATCCGCGCA
TGAACGAGATCCTCTTTCCGCTCTACGAGGAGAAGCGCTGCACGGAGATCGTAAACGACTATGAGCTGGAC
GAAGAGAAGAAGAAAAGCGTTCAACTATCCCTGGACGGATTCAAGCGCTATCTGATGTCCGATGAGAACGC
ACCCGTGTTTTCTGGATCGCCTGGACTTCTACATGGAGATGGACCAGCCACTGGCCCACTACTACATCAATA
GCTCGCATAACACCTACTTGTCCGGCCGACAGATCGGCGGAAAGAGCTCCGTGGAGATGTACCGCCAGACG
CTGCTTGCCGGCTGTCGGTGTGTGGAGCTCGACTGCTGGAACGGCAAGGGGGAGGACGAGGAACCGATTGT
CACCCACGGTCACGCCTACTGCACGGAAATCCTCTTCAAGGACTGCATCCAGGCCATCGCGGACTGTGCGT
TCGTCTCGTCCGAGTATCCGGTGATCCTGTCCTTCGAGAACCATTGCAACCGGGCCCAGCAATACAAGCTG
GCCAAATACTGTGATGACTTCTTCGGGGAGCTGCTCCTGAAAGAACCGCTCGCAGATCATCCGCTGGATCC
GGGCCTGCCCCTGCCGCCGCCGTGCAAGCTGAAGAAAAAGATACTCATCAAGAACAAGCGCATGAAGCCCG
ACGTGGAGAAGGTTGAGCTGGAGCTCTGGCTGAAGGGCGAGCTGAAGACCGATGACGATCCCGAGGAGGAC
GCCAGTGCCGGGAAGCCGCCGGAGGCTGCGGCCGCTCCCGCACCAGTTCCCGAAGCCGCAGCCGCTGCCGA

CGGTGCTGCCGAAGGAGGTGGCGGTGCCGAGGCGGAGGCAGCGGCGGCCAACTACAGTGGCTCCACCACCA
ATGTACATCCCTGGCTTTCATCGATGGTCAATTACGCGCAGCCCATTAAGTTCCAGGGCTTCGATAAGGCA
ATTGAGAAGAACATTGCCCACAACATGTCCTCGTTTGCGGAATCGGCGGGCATGAACTACCTGAAGCAGAG
CTCCATTGATTTCGTCAATTACAACAAGCGCCAGATGTCGCGCATCTACCCGAAGGGCACCAGGGCGGACT
CCTCGAACTACATGCCGCAGGTGTTCTGGAACGCCGGCTGCCAGATGGTGTCGCTGAATTTCCAGACCTCC
GATCTGCCCATGCAGCTGAACCAGGGCAAGTTCGAGTACAATGGCGGATGCGGATACCTGCTGAAGCCGGA
TTTTATGCGGCGCGCGGACAAGGATTTCGATCCATTCGCCGACGCACCCGTGGACGGTGTGATCGCTGCCA
TGTGCTCGGTGAAGGTAATCGCCGGACAGTTCTTGTCGGACAAGAAGGTGGGCACCTACGTGGAGGTGGAC
ATGTTTGGCCTGCCCTCGGACACCGTGAAGAAGGAGTTCCGCACCCGTCTGGTGCCCAACAACGGATTGAA
TCCGGTCTACAACGAGGACCCGTTCGTGTTCCGCAAGGTGGTCCTGCCGGATCTGGCCGTGCTGCGGTTCG
GCGTCTATGAGGAGAGCGGCAAGATGATCGGTCAGCGAATCCTCCCGCTGGACGGACTCCAGCCGGGATAC
CGCCATGTGTCGCTGCGCACAGAGGCCAACTTCCCGATGTCGCTTCCCATGCTGTTTGTGAACATCGAATT
GAAGATTTACGGGCCTGATGGCTTTGAGGACTTCATGGCCATGCTGTCGGANNNNNNNNNNNNNNNNNNN
NNNNNCAGCAGCAGAACGAGCAGATGAAGNCACTNGGCATCGAGNAGCAGAGCGGCGGNGCCGCCCGNGAT
GCCGGCAAGGCNAAGGAGGAGGAGAAGAAGGAGCCGCGNCTGGTCTTCGAGCCGGTCACGCTGGAGTCGTT
GCGCCAAGAGAAGGGCTTCCAGAAGGTGGGCAAGAAACAGATCAAGGAGCTGGACACGCTGCGCAAGAAGC
ACGCCAAGGAGCGCACCTCCGTCCAGAAGACCCAGAACGCGGCCATCGACAAGCTGATCAAGGGCAAGAGC
AAAGACGACATCCGCAACGATGCAAACATCAAGAACGCCATCAACGACCAGACCAAGCAGTGGACCGACAT
GATCGCCCGCCACCGCAAAGAGGAGTGGGACATGCTGCGCCAGCATGTCCAGGACTCGCAGGATGCCATGA
AAGCGCTCATGCTGACCGTGCAGGCTGCTCAGATCAAACAACTGGAGGATCGTCATGCCGATCTGAAGGAA
CTGAATGCCAAGCAGGCCAAGACCTCAGCGGATACCGCCAAGGAGGTACAGAACGACAAGACTCTCAAGAC
CAAGAACGAGAAGGATCGTCGGCTGCGCGAGAAGCGTCAGAACAATGTCAAGCGATTCATGGAGGAGAAGA
AGCAAATCGGCGTGAAGCAGGGACGCGCCATGGAGAAGTTGAAGCTGGCGCATTCGAAACAGGTCGAGGAA
TTTAGTACCGATGTGCAAAAGCTAATGGATATGTACAAAATCGAGGAGGAGGCCTATAAGACGCAAGGAAA
AACGGAATTTTATGCCTAA

>Dper_pinta

ATGTGGTCGCGCAGCAGCAGTGCAAGGCGGCAAGTCACCGCCATAGACAGTGATCCCGAGAAGGTATTGAG
GCAGGTGCAGGAGCTGAACGACTGGCTAGACGCCAATCCGCAGATCAACTGATGCAAAACTTTCGAGAATT
TGCATATTTTTCTGCGCACCTCCAAATTTGATGTGGAACGCACCAAGAAGAAGCTAAAGACATTCTATCAG
ATGCGTGCTGAGCGAACCGAATGGTTTCGATAATCGCGATCCCAACTTACCCGAAATCCAGGAGCTGCTGG
AACTGGGCGTTTTTCTTCCCGTTGGCCCCGGACTCAGACAATGGTTGTATGTGTTGTGGTCATTCGGACGG
CGGCCCATGATCCGAAGAAACACACACAGAACAACGTCTTCAAAACCAGTAAAATGATATTGGACTTGCTG
TTGAAGCTCGATCCGGAGAGCTGTGCTAAGGGCATGGTTGCCATACTAGATATGCAGGGAGTTCAATTGGG
TCATGCCCTGCAGCTGAATCCAAAGCTCATTAAACGATCGGTGGAGAGCTGGACTGCCTATCCCTGCCAGC
CCAAGCTATTGGAGTTTACTAACGCTCCGCGTCATGTCAACATCTTTCTGAACACTTTCCGGATCTTCATG
ACACCCAAGATACGGTCCCGAGTGATGGTCCGCCGCGAGGGAACCTCTGTGAACAGTGCGCACTTGCCAAA
AGAATTGGGTGGTCAGGGCCCGAGCTACGCCGAACTCGCCATCAAATGGAAACGGTTGCTGGAGGAGAATT
CCTGATTTCTATGTGGAACAAGACAAATATAAAAGCAAGCTCAAGTAA

>Dper_pld

ATGGCCAAGTCCACGGATACCGTCGCCAGCAGCAGCGACCACAACGAGGCGGCAGAGGCGGCGATGGCGGT
GACTGCAACCTTGCAGCATCTGCAGGTGCGACGTCATCGCCGCAGCATTTCACAGCTGATGGCTAGCATGG
CCGAATACCCCGCGGATGATGTGTACCGGCCCACGAACTTCGCGGGCTACGTGAATCGGGGCTACGATGAT
ATAGACAGTTCCTATTACTTTGCCCAGTTCGAGGCCATGACGGAGGAGCGCAATGGAGCGGCCCTGCCGCC
GTACACCAACGATTCGGATGCCTCCGCGGAGGACAACCGCAACGAGGATGAGGACCAGGAGGACGAGGAGC
AGGACCCGGACTGCACGGACGATGGCACTGTGGTGGAGCATCATAACCGATGCCTGCCCGAGTTCCAGTTC
TCGCTGGTCGACTCGGAGTACGATGAGACGCTGGGGTTCCCCGACTCCGTGACCATCCTGTCGAATGTGGG
CGACAAGCCGGTGCTGGTGCAGCGCAAGGAGACCACCGACGACGACGACGAGGAGTACGACGACGATGAGA
ACAACAGCGTGGTGATGCGCCAGGAGATACCCTTCACGAGCATCTACGGGCCCAGCGTGAAGTTCAACTCC
TTTCAGCGGAAGGTCTTCATCCCGGGCCGCGAGATTCACGTGCGGATCGTGGACACCGAGCGGAGTGTGAC
CACGCATCTGCTGAACCCCAATCTATACACCATTGAGCTCACCCACGGACCCTTCAAGTGGACCATCAAGC
GGCGCTACAAGCACTTCAATTCGCTCCACCAGCAACTGAGCTTCTTCCGCACCTCCCTGAACATACCCTTC
CCCAGTCGCAGTCACAAGGAGAAGCGGACCACGCTCAAGGCCACCGCCAGCCAGATGGCCGATGAGTCGAC
GCTCAAGGACCTGCCCGCCCACACCAAGCAGGTGAAGCAGACGAGCACGCCCCAGAACAATGGCAGCAGGC
ACCAGAACGGACCCGGCACTATCGTCAGTCCCAGTCACAACTCGATTCTGTCGGGCCTCAATCCGCGGCGC
ATCCAAAAGAAGCGCAAGAAGAAGAAAAAGAAGAAGCTGCCGCGGTTCCCCAACCGCCCGGAGAGTCTCGT
GACGGTGGAGAATCTGGGCGTGCGGATCAAGCAGCTGGAGGACTATCTGTACAATCTGCTGAACATCAGCT
TGTACCGATCGCACCACGAAACGGTAAACTTTGTGGAGGTCTCCAATGTGTCCTTTGTGCCCGGGATGGGC

CTCAAGGGCANGGAGGGCGTAATACTCAAGCGCACGGGCTCGACGCGACCCGGCCAGGCGGGCTGCAACTT
CTTTGGGTGCTTCCAGAAGAACTGCTGCGTGCGCTGCAACTACTTCTGCTCGGACGTGGTGTGTGGCACGT
GGCGCAATCGCTGGTTCTTCGTGAAGGAGACCTGCTTCGGCTACATCCGACCCACCGACGGCAGCATCCGG
GCCGTGATCCTGTTCGACCAGGGCTTCGATGTGTCGACGGGCATCTACCAGACGGGCATGCGGAAGGGACT
GCAGGTGCTGACCAACAACCGGCACATTGTGCTCAAGTGCTGGACAAAGCGCAAGTGCAAGGAGTGGATGC
AGTACCTGAAGCACACCTCCAACTCGTATGCCCGCGACTTCACGCTGCCCAATCCGCACATGTCCTACGCC
CCCATGCGGGCCAACACGCAGGCCTCCTGGTACGTGGACGGCTCCCAGTACATGTCCGCCGTCGCCGATGG
CCTCGAGGCGGCCACCGAAGAGATCTACATTGCGGACTGGTGGCTCAGCCCGGAGATCTACATGAAGCGGC
CCGCTCTGGATGGCGACTACTGGCGCCTGGACAAGATCCTCCTCCGCAAGGCCGAGCAGGGGGTGCGGGTC
TTCGTGCTGCTCTACAAGGAGGTGGAGATGGCCCTGGGCATCAATAGCTACTACAGCAAGTCCACGCTGGC
CAAGCATGACAACATCAAGGTCATGCGGCATCCGGACCATGCACGCGGTGGCATTCTCCTCTGGGCGCACC
ACGAGAAGATCGTGGTGATCGACCAGACGTACGCCTTCATGGGCGGCATCGATCTGTGCTACGGCCGCTGG
GACGATCATCATCATCGTCTCACAGATCTCGGAAGCATTTCCACGGCCTCCGTTTCGGGCAGCACGCGCCG
CACGCCCAGCGGCTACTTCAACAAGGACGACGTGGACTCGGCCTTTGGCTCGCGCAAGTCCTCGAGGAATG
CCCACCACGAGACCCCGCCCAAGGCTGATAGGGCGGCCTCGCCAGAGAAGCTGGAGGCACCGNNNNNNNNN
NNNNNNNNNCCGCCCCCGCCAGTCAACAGCATGGAGTTGAGGGTCCTCAAACCGGGCGATCGACTGCTGAT
ACCCGCCATGCTGTCTAGTGGCGCTGGCGAGGATCCCGACGAGAACATGGCCCTGGAGGGAATGAAGCTGA
ACACGCCAGAGATGGAGCGGAAGAACATGCTCGATCGGCTGAAGAACAACGCCATGAAGGGGGCTCGCATG
GGCAAGGACTTCATGCACCGCCTGACGGCCAACGAGACGACGTTGGAGGGGCCAGAGGATGGAGGTGTCTT
CAGCATAGACACAGTGGATGCCACCATGGATCCCGACATGACTATGGCCTCCTCCTCTGCTACAGCCACAG
CTAATGCCACCGCCACAGCCACAGCCACCGATGTCGCCACTGCCACAGCTAATGCCACGGCCAACTCCACG
CAATTATTCAGCGAATTCTATGGCCAGGCCAAGTACTGGTTCGGCAAGGACTACTCGAACTTCATACTCAA
GGACTGGATGAACCTCAACTCGCCGTTCGTTGACATTATCGATCGCACGACCACACCGCGCATGCCCTGGC
ACGATGTCGGCCTCTGCGTGGTGGGCGCTTCGGCGCGGGATGTGGCCCGCCACTTTATACAGCGCTGGAAC
GCCATGAAGCTGGAGAAGCTGCGGGACAATACGCGCTTTCCCTACCTCATGCCGAAGAGCTACCACCAGAT
CCGCCTCAACTCGCATCTGCAGCAGCTGCAACAGCGCCGCCAGCAGAGGGTCACGTGCCAGCTGCTGCGCA
GCGTCTCCGCCTGGAGCTGCGGCTTCATCGAGGCGGACCTCGTGGAGCAGAGCATCCACGACGCCTACATC
CCGACGATCACCAAGGCCCAGCACTANGTGTACATTGAGAATCAGTTCTTCATCACAATGCAGCTGGGCAT
GGGCGTGCCGGGGGCCTACAACAACGTGCGGAATCAGATGGGGGAGACACTTTTCAAGCGCATTGTGCGGG
CCCATAAAGAACGGAAACCCTTCCGTGTGTATGTGATAATGCCACTCCTTCCGGGATTCGAGGGCGACGTG
GGCGGCAGCACGGGCATCGCTGTGCGGGCCATCACCCACTGGAACTATGCCTCCATTTCCAGGNGACGCAC
TGCTATACTGACACGTCTCCAGGAGGCCGGCGTCTCGGAGCCGCACAACTACATCTCATTCCACAGTCTAC
GCAATCATTCCTTCCTGAACAGCACGCCCATCACCGAACTGATTTACGTGCATTCCAAGCTGCTCATCGCC
GACGATCGGGTGGTCATCTGCGGCTCTGCGAACATCAACGACCGCTCGATGATTGGGAAGCGCGACTCGGA
GATTGCGGCCATCATCATGGACGAGGAGTTCGAGGATGGCCGCATGAATGGCAAAAAGTATCCGAGCGGCA
TCTTTGCCGGTCGCCTGCGAAAGTATCTCTTCAAAGAGCATTTAGGCCTCCTGGAGGGCGAGGGCTCTGCG
CGCTCTGATCTGGACATCAGCGATCCCGTTTGCGACGCCTTCTGGCACGGCACCTGGCGACAGATCTCCAC
GAGGAACACCGAGATCTACGACGAGGTGTTCAAGTGCATTCCCACGGACTTCGTGAAGACATTCGCCAGTC
TGCGCAAGTACCAGGAGGAGCCGCCGCTCTCCAAGACCGATCCAGAGATGGCCGCCCATCGGGTCACAGAA
ATCCAGGGCTTTCTGGTCAACCTGCCATTGGAGTTCCTCAACAAGGAGGTGCTCACGCCGCCGGGCGCCAG
CAAGGAGGGCGTCCTACCTACCTCTGTTTGGACTTAG

>Dper_stops

ATGTTTTTGCAGTTGCTCGATGCATTTCAGCGGCTGCAAGTCAATAAAACGACAATTTGTGTGGATTCGCT
GCTGGTTGCCATTGCCGCAACGCTTCTACGGCTTCCGGCGGTCATTTCTAGTAGACCCGGCACGGACACGG
GCACGGGAACGGGCCAGCTACCAGTGCAATCCGAGGGCGACGCGTCGCCAACTGGTCATGGACTACATCTT
CGAGTGCTTCTTCGAGGACACATTCGATCAGATCACCCGCAACGGCCTGCAGGATCGACAGTCGCGTCGGG
ACGTGCTCGACCATCTCAGTTCTATAATCAAGGGCTGCAGCGAGGGCCAGAACACGCAGACAGATGAGGTG
GCTGGCATAGCCGTGGTGGCGGCCATGCGATATCACCGACTGGCCAAGGAGGATAATGGCGGAAAGGTATG
CCTCATGGGCAAGTATCACAACATCCTGTACATCGCCCTGCGCACCTGCTGGGACTGGGGTGTGCGCGACT
CCGAGGTGGTCGTCCAACTCTTGGTGGCCATTTATGAGTGCGAGAAGACTTTCGAGCGGATATTCCTGGGC
GCCTTGTTTGGGCCACATGCTCCGCACTTTATCGCCGGTTGGCGCAGCGACTTTCAGGATCAGCATGAAAA
CGTCCGAGCCGTGGTTTACTTTCTGAAGCATGCCACACGCGAGAAGCTCATGTTGCCCGTCTGGATACCGC
GCTTCGAGGAGGAGAGACAGCTTCGATTCATTGATGTGCCCATTGAGTCCTGTGGCAGGTCCTCGCCGCTG
CGCATCGCCCTGCAGGCAAATGCTCCCGAGCTGCTGCTCATTCTCCTGAGATACGGCGCCTCACCGCATCC
CCCGGACGGAGGCACCTCGGTAATCCTGGCACTGCTCGACAAACTAATTGAGAATGGCCGCAATTACAGCT
ACGAACTGGTGCTGTGCCTCAAGATACTGTTGCGAAACGTGGCCATGATTGAGATGCCCTTCAAGCCCCTG
TTGTATGGCTCACGCAGGGAGATGTTCTTCGAGCGCTACGGCAGACTGTTAATCGACAAAATCATTGGCAA

GGAGCAGGTGTATGGTGTGCCCACACTGCGTCACCTATGTCGCTGCTGCATTCGCGATGTCCTCCGGGAGC
AGAATCAACTGCCCAACGGCATCGACACTTTGAGGCTGCCAATGCGACTGCAGCGCTATATTGACCTAACG
GAAGAGCTGGAGGGAGTCGGGGGACAGCCAGAGCCACCAGAGCAGCCAAAGCCCAAGCCGGGCACTCGACT
GGCAAGTGGCATGCAGCGGATGGACACCATCAGCAGCATTACCGAGAGCGAGGACGAGAGCGAGAACCGTG
TCCAGCTGCTGGAGGCAGTCACGGATGCGGAAAAGGCCGCGGCAGCCGCCTTTGTGGCCGCCAGCGATGGG
GCAATCACTGCCGAGGCCAAGCAGGCGGCTGTGGCCGCCTATGTGGCAGCCAGCAGACCCGAAGCCGATCC
CGATTTCCCACCTAAGCCCCAGCAGGACGATGACGACGACGGGGGCCCTGCCCCAGAGCTTAGCTAG

>Dsec_GBeta76C

ATGCCGAAAATTGACCCAGAAACACAAAAACTCTACGACGAAATCAATGGCATGATACAGAAGTTTAAGGA
TGACCAAAAGAGCAAGGCTGATTGCACGTTGGCCGATAAATGCGGCGACATGGGCGATGTGCCGAAGATTC
GATTCTCATCCAAGAAGATTCTCAAGGGCCACATCAACAAGGTGAATTCGGTGCACTTTGCCGGGGACTCG
CGGCACTGCGTCACTGGATCGCTGGACGGCAAGCTGATCATCTGGGACACCTGGACCGCCAACAAGGTGCA
GATCATCCCCTTGAGATCCGCCTGGGTGATGACGGTGGCCTTCTCGCCGTCGGGTAACTTTGTGGCCTGCG
GAGGAATGGACAACCAGTGCACCGTCTACGATGTAAACAATCGTGATGCCTCCGGAGTGGCTAAGATGGTC
AAGGAGCTGATGGGCTACGAGGGATTCCTCAGCTCCTGTCGCTTCCTCGACGACGGCCATCTGATCACCGG
CTCGGGAGATATGAAAATNTGCCATTGGGATCTGGAGAAGGGTGTCAAGACAATGGATTTCAATGGACACG
CTGGTGATTGCCATTGGGATCTTGAGAAGGGCGTCAAGACAATGGATTTCAATGGACACGCTGGTGATATC
GCTGGCTTATCGCTTTCGCCAGATATGAAGACCTATATCACTGGCTCTGTGGACAAAACTGCCAAGTTGTG
GGATGTGCGCGAGGAGGGCCACAAGCAGATGTTCTTCGGTCATGACATGGATGTGTCATCCGTCTGCTACC
ATCCCAGTGGCTTCGGGTTCGCCTCCTGCTCGGAGGACCAGACGGCGCGCATGTACGACCTGCGGGCGGAC
CAGCAAATCGCCCAGTATGAGCCGCCCCAGAAGAACACTGGCTTCACGTCGTGCGCTCTATCGACCAGCGG
GCGCTACCTCATGTGCGGCGGAATTGAGGGCAACGTGCACTCGTGGGACACGATGAAGCAGCGGCACACGG
GCACAGGCACACTGTCTGGTCACGAGAACCGCATCACTTGCATCAGCCTGTGCCCCAATGGCATGTGCCTG
GCGTCCACCAGTTGGGATCAGCAAGTGCGTCTGTGGCTCTAA

>Dsec_inaC

ATGGCTGCGGCTGCTGTTGCCACACCGGGGGCCACTGTGCTCCCTCCCTCGGCGCCCAGTGCCGCTCCAGG
GGCCAAGGCTCCAGCGGCTGGAGCTGGCAAAGGTCCAGGAAATCTACTGGAGATCACTGGCGAAGCAAACA
TCATCAACTATATGAAGAATCGCCTCAGGAAGGGTGCGATGAAGCGAAAGGGTCTCGAAATGGTCAATGGT
CATCGTTTCGGTGTGAGATTCTTTAAGAATCCAACCTACTGTGGTCATTGCAAGGATTTCATCTGGGGCTT
TGGCAAGCAAGGTTTTCAGTGCGAGGAGTGCCGCTTCAACATTCATCAGAAGTGCTGCAAGTTTGTGGTAT
TCAAGTGTCCTGGCAAGGACACGGACTTCGATGCAGACTGCGCCAAGGTGAAGCACGGCTGGATCTCCACC
ACCTACACCACACCGACATTCTGTGACGAGTGCGGACTGCTGCTCCATGGAGTCGCCCATCAGGGCGTCAA
GTGCGAGAATTGCAACCTCAATGTGCATCATGCCTGCCAGGAGACAGTGCCACCGATGTGCGGTGCGGACA
TCAGCGAGGTGAGGGGTAAGCTACTACTCTACGTCGAGCTTAAGGGCAATAACTTGAAAGTTGATATTAAG
GAGGCAGCTAACCTCATTCCCATGGATACCAACGGGTTTAGCGACCCCTATATTGCAGTGCAAATGCATCC
GGACCGTTCGGGCCGCACCAAGAAGAAGACCAAGACCATTCAGAAAAACCTCAATCCAGTGTTCAACGAGA
CCTTTACATTTGATTTGCAGCCGCAGGATCGTGACAAGCGTTTGCTCATCGAGGTCTGGGATTGGGATCGA
ACTTCTCGCAATGACTTCATGGGTTCGTTCTCGTTCAGCCTGGAGGAACTGCAAAAGGAGCCAGTTGATGG
CTGGTACAAGTTCCTGTCGCAAGTGGAGGGTGAACATTATAACATACCCTGTGTGGATGCCTTTAATGATA
TAGCTCGCTTAAGGGATGAAGTTAGGCACGATCGTCGTCCGAATGAAAAGCGTCGCATGGATAACAAGGAT
ATGCCGCATAACATGAGCAAACGCGATATGATTCGAGCTGCTGATTTCAACTTTGTTAAAGTTATTGGCAA
GGGATCCTTTGGCAAAGTCCTGCTAGCCGAGCGGCGTGGCACGGATGAGCTGTATGCCGTGAAGGTCCTGC
GAAAGGATGTCATTATCCAAACGGACGACATGGAGCTGCCCATGAATGAGAAGAGAATTCTGGCTCTATCC
GGTCGTCCTCCTTTCCTCGTCTCCATGCACTCATGCTTCCAGACTATGGACCGATTGTTTTTCGTAATGGA
ATACTGTAAGGGTGGAGATCTTATGTATCATATGCAGCAGTATGGAAGGTTCAAGGAGTCTGTGGCTATTT
TTTACGCTGTTGAAGTGGCTATTGCTCTGTTTTTCCTCCACGAGCGTGATATTATCTATAGAGATCTGAAG
CTGGACAACATTTTACTGGATGGTGAGGGACACGTAAAGCTGGTGGACTTTGGGTTGAGCAAAGAAGGCGT
TACCGAGCGGCAGACCACACGAACTTTTTGTGGTACTCCGAATTATATGGCTCCCGAAATTGTTAGCTACG
ATCCGTACTCCATCGCCGCTGACTGGTGGTCTTTTGGTGTTTTGCTTTTCGAATTCATGGCAGGCCAGGCT
CCATTCGAAGGAGATAATGAAAATACAGTCTTTAGGAACATAAAGGACAAGAAGGCTGTGTTTCCAAAGCA
TTTCAGCGTGGAGGCCATGGATATAATTACGAGCTTCCTGACCAAAAAGCCCAACAATAGATTGGGTGCAG
GGCGCTATGCCAGACAGGAGATTACCACGCATCCTTTCTTCCGCAATGTGGACTGGGATAAGGCAGAGGCA
TGCGAAATGGAGCCACCTATTAAGCCAAAGATTAAACATCGCAAGGATATAAGCAACTTTGATGATGCCTT
TACCAAGGAGAAGACGGACTTAACTCCCACGGATAAGCTTTTCATGATGAATCTGGACCAGAACGATTTTA
TCGGATTCTCTTTCATGAATCCCGAGTTTATAACCATCATTTAA

>Dsec_Pld

ATGATGGCCACGAGCTTGGATACCAAAACCAGCAAGAACCACGAGGCCCAGCAACTGCAGCTGCCCCGGCA
GGCGGCATCGGGCGCCAACTTGCCACATTTGCAGGTCCGACGTCATCGCCGCAGCATCTCACAGCTGATGG
CCAGCATGGTTGAATACCCCACTGACGATGTGTACCGGCCGGCGCACTACGGCGGCTACGTTAACAGAGGC
TACGATGATATAGACAGTTCCTACTATTTTGCCCAGTTCGAGGCGATGGCAGATGCCGGCACCGTTGGAGG
CGCCTTGCCGCCCTACGCCCTTACCAACTCGGACGAAGAACTTGGCAGCGGGGAGGAGGACGCGTCGGAGG
AGAACTCAAACAATGAAGAGGGGGAGGGGGTGTTCCGGGACTGCACAGACGAAGCGGTCGTCGAGCATCAT
AACCGCTGCCTGCCGGAATTTCAGTTCTCCCTGGTCGATTCTGAGTACGATGAGACCCTCGCTTTTCCTGA
TTCTGTGACCATTCTATCCAACGTGGGCGACAAGCCGGTGCTGGTGGAGCGCAAGGAGACGGACGATGATG
AGGAGGAGTTTGACGACGAGGAAAACAACAGTGTAGTCCTGAGACACGAAATACCATTTACTAGCATATAC
GGACCGAGCGTCAAGTTCAACTCGTTCCAGCGCAAGGTCTTCATCCCGGGCCGTGAGATTCATGTTCGGAT
CGTCGATACGGAGCGTAGCGTAACCACACATCTGCTTAACCCCAATCTGTACACAATCGAGCTGACCCACG
GTCCCTTCAAGTGGACGATCAAGAGGCGATACAAGCACTTTAACTCGTTGCACCAGCAGCTCAGCTTTTTC
CGCACCTCGCTCAACATTCCTTTTCCCAGTCGCAGCCACAAGGAGAAGCGTACCACGCTGAAGGCCACAGC
CAGAGAGATGGCTGACGAGTCCACCCTAAAGGACCTTCCTTCCCACACCAAGGTCAAACAAACTAGCACTC
CGCTGAGGCCTGAAGGCAGAAACAGTAAAATCGCGGGCAGTAACGCCGACAATGCCATGGCTATCATCAGT
CCCAATCACAGCTCCATTCTGGCGGGCCTAACACCACGGCGCATTCAAAAGAAGCGCAAAAAAAAGAAGAA
ACGGAAGCTACCGCGATTCCCAAACCGACCGGAGAGTCTGGTCACCGTGGAGAATCTGAGCGTCCGAATAA
AACAGCTGGAGGACTACTTGTACAACCTGCTGAACATCAGCTTGTACCGATCTCACCATGAAACGCTAAAC
TTCGTTGAAGTGTCTAATGTGTCCTTCGTTCCGGGAATGGGAATTAAGGGCAAGGAAGGCGTGATTTTAAA
GCGAACTGGATCGACGAGACCCGGGCAAGCAGGATGCAATTTCTTCGGGTGCTTTCAAAAGAACTGCTGTG
TGCGCTGCAACTACTTTTGCTCCGACGTAGTTTGCGGCACGTGGCGGAACCGATGGTTTTTCGTAAAAGAA
ACCTGCTTCGGCTACATCCGTCCAACAGACGGCAGCATCCGGGCAGTGATCCTCTTTGACCAGGGCTTCGA
CGTCTCCACGGGTATCTATCAGACGGGCATGCGCAAGGGTTTGCAGGTACTGACGAACAACCGTCACATTG
TGCTCAAGTGCTGGACACGTCGTAAGTGTAAGGAGTGGATGCAATACCTCAAGAACACTGCCAACTCGTAT
GCGCGCGACTTCACCCTGCCCAATCCGCACATGTCCTTCGCTCCGATGCGCGCCAACACCCATGCCACGTG
GTATGTAGACGGCGCCCAGTATATGTCAGCCGTGGCCGACGGCTTAGAGGCAGCCCTCGAAGAGATCTATA
TTGCCGACTGGTGGCTCAGTCCCGAGATTTACATGAAGCGACCCGCACTCGACGGAGACTACTGGCGATTG
GACAAGATCCTGTTGCGCAAGGCCGAACAGGGAGTGCGTGTCTTCGTGCTGCTCTACAAGGAGGTTGAAAT
GGCACTTGGCATAAACAGCTACTACAGCAAGTCCACGCTGGCCAAGCACGAAAACATCAAGGTCATGCGTC
ATCCGGACCATGCTAGAGGAGGTATTTTGCTTTGGGCACATCACGAAAAGATGGTCGTAATCGACCAAACC
TATGCGTTCATGGGAGGTATTGATTTGTGCTATGGACGTTGGGATGATCACCTCCATCGGCTAACGGACCT
GGGAAGCATATCTACGGCGTCTTTTTCTGGCAGCACGCGTCGAACGCCAAGCTTGTACTTCACCAAAGACG
ACACGGACTCAGCTTTCGGATCACGTAAGTCCTCGCGAAACGCTCACTACGATACCTCCGCCAAGGAAAGG
CCACCGTCCCCACCCCCGGATGAGCCCAGTACTAGCATAGAGTTGAAAACTCTAAAGCCTGGCGATCGCCT
GCTTATACCGTCTACGCTCGTTTCCAGTCCGGGAGAAACTCCCGTAGAATCGGGAATCGCTTTAGAAGGGA
TGAAACTCAACACCCCTGAAATGGAGCGTAAGAACGTACTCGATCGCCTGAAGAACAACGCGATGAAGGGC
GCCCGTATGGGCAAGGACTTTATGCACCGTCTAACAGCCACTGAGACGGAGGAAAAATCTGCCGAGGTGTA
CACTATCGAGTCCGAGGATGCTACGGACCACGAAGTCAACCTTAACATGGCTTCAGGTGGGCAGGAAGTGG
CGATTACCACTAGCAGTACACAAATACTCAGTGAGTTCTGCGGCCAGGCCAAGTACTGGTTCGGCAAGGAT
TACTCCAACTTTATACTTAAAGACTGGATGAACCTAAACTCGCCGTTCGTGGATATCATAGATCGAACAAC
AACACCGCGGATGCCATGGCACGACGTGGGTCTGTGTGTGGTGGGTACTTCCGCTAGGGATGTGGCACGCC
ACTTCATTCAGCGCTGGAATGCCATGAAGCTGGAGAAACTACGCGATAACACGAGGTTTCCCTATCTGATG
CCGAAAAGCTATCACCAAGTGAGGCTCAATCCTAACCTTCAGCAAAACCGTCAGCAGCGGGTCACGTGCCA
GCTACTTCGGAGCGTCTCCGCCTGGAGCTGCGGCTTTATAGAGGCGGATCTTGTGGAGCAAAGCATCCACG
ATGCCTACATCCAGACGATCACAAAGGCGCAGCACTACGTGTACATCGAAAATCAATTTTTTATCACGATG
CAGTTGGGCATGGGTGTGCCAGGCGCTCATAACAATGTGCGGAATCAAATCGGGGAAACACTTTTTAAACG
TATTGTTCGGGCGCACAAGGAACGGAAGCCTTTCCGAGTTTATGTGATTATGCCGCTTCTACCGGGCTTTG
AGGGCGATGTGGGTGGCAGTACTGGGATAGCAGTCAGAGCCATTACACACTGGAACTATGCATCCATTTCC
AGGGGACGCACAGCAATTTTGAACCGCCTGCAGGAGGCGGGTATTGCCAATCCGCAAAACTATATCTCATT
CCACAGCCTGCGTAACCATTCTTTTTTGAATAACACACCCATAACGGAGTTGATATATGTCCACTCAAAGC
TCTTGATAGCCGACGATCGCGTTGTGATCTGCGGCTCGGCAAACATTAACGATCGGTCTATGATCGGAAAG
CGGGACTCCGAGATAGCGGCTATTCTAATGGACGAGGAGTTTGAGGACGGACGCATGAATGGCAAGAAGTA
TCCGAGCGGAGTGTTTGCCGGACGCCTTCGAAAATACCTTTTTAAAGAACACTTAGGGCTCCTGGAAAGCG
AAGGTTCCAGTCGGTCTGACCTGGACATTAACGATCCTGTTTGCGATAAGTTTTGGCACGGCACATGGCGT
AGGATTTCAATGCAGAACACAGAGATTTACGACGAAGTGTTTAAGTGCATCCCCACTGACTTTGTAAAGAC
CTTTGCCAGCCTTCGCAAATACCAGGAGGAGCCTCCTCTTGCCAAAACCGCCCCTGAGCTAGCTGCCAACA

GAGCCAACGACATTCAGGGTTACTTGGTCGACCTGCCATTGGAATTTCTGAACAAGGAGGTCCTTACGCCG
CCTGGAACTAGTAAGGAGGGCCTAATCCCTACCTCTGTATGGACATAG

>Dsec_rdgA

ATGCAGCAGCAGCAACAACCGTCCATCGACCAGCTGCCCGAGCCCACAGCTTCGACCAGCAATTCGGCCAC
AACTAAACCGACAATTGCAACAGCAACCACATCGACAACAACAACGTCTGGCAACAACTTCCACCAGCAAC
TGCAAGCAACAACGGCAGCCACCATGCAACGCCTGCGAACAACCTTTACGCGATCACGCACTCCAACAGGT
GCCGAAATGAAGATGCAGAATAGCCTGGAGGTGCCCAAGCAGGTGCAGGTGCGCTCCGCCTCTTTCGACGA
GATGCAACTGGAGTCGCAGCGTGCCTCCTCCAGGCAGGTGCGTTCAGCCTCCTTCGACGAGATGCAACTGG
AGTCGCAGCGTGCCTCCTCCAGTCTGCTCAAGCAGCAATCCTCATCATCGGCCTCGGCAGACGAACGCTCC
TCGGAGGCTGGTCTGCTCAAGCAGCAGTCCTCATCATCGGCCTCGGCGGACGAACGCTCCTCAGAGGCTGG
CTTCCTGCAGGTTCCATTGGCGGCCCATCAGCAAAGATCCCATAGCTTCGATTCGGCAACCTTCCTGCAGG
TTCCATTGGCGGCCCATCAGCAAAGATCCCATAGCTTCGACTCGGCAACTGCCTCGGCGGGCAGCGATGAC
TCGGGCACCTTTTTGGAGGTGCCACGACGGCTGAAGGCTGCCTCGGCGGGCAGCGATGACTCGGGCACCTT
TTTGGAGGTGCCCAGACGGCTGAAGGTTCGACGCAGTTCGTCAACCAAGACACCGCCGCCTTGCATCCATT
GTCACTATCTGGAGGATCGACGCAGTTCGTCCACCAAGACACCGCCGCCTTGCATCCATTGTCACTATCTG
GAGGAGTATGAGCGTCGTATGACCGCCGAACAGCGTTACTTCATCGATCATCGCGAGTTAACGGCGTATGA
GCGTCGGATGACCGCCGAGCAGCGTTACTTTATCGATCATCGCGAGCTGACAGCTTTGAGCTACAGCAATA
CCAGCTCGGAGGCGAGTGAGGATGAGGATGAAGTTGAGGGCCATTTGAGCTACAGCAATACCAGCTCGGAG
GCGAGTGAGGATGAGGATGAAGTGGAGGGCCATAACGCGGAGGAAGAGGAGGAGGGCAGTGCGGCCATCGA
GGATGCCGAGGAGGAGACCACTAATGCGGAGGAAGAGGAGGAGGTCTGTGCGGCCATCGAGGATGCCGAGG
AGGAGACAACGGAGGCAGCTACCGAAGAAGCGGATGAAGATCCCCGCACCGAGGTTGAGTCCGAGCATGAT
GAGGCGGCTACCGAGGAGGCGGATGAAGATCCCCGCACCGAGGTTGAGTCCGAGCATGATCACGATCCGGA
TGACGATGCCGCATTGGAGATGGATATTCGCATTGGCAACATGAGCCATCACGATCCGGATGACGATGCCG
CATTAGAGATGGACATCCGCATTGGCAACATGAGCCAAGGCAGCAGCATCGAGGAGTCGCGCGCCCGCCTA
CCGCGACAAATGCGGCGTCACACCATAGGCAGCAGCATCGAGGAGTCGCGCGCCCGCCTACCGCGACAAAT
GCGGCGCCACACCATCGGTAGCAGCTCGGTCACCTCGGCCTCCGAGGATGAGGGCTTGGAGGGCTCGGACA
ATGGCGGAAGCAGCTCGGTCACCTCGGCCTCCGAGGATGAGGGCTTGGAGGGCTCAGACAATGGTTCGCCG
CATTTTGGCAACACCTTGTTGCCACCACAGCCGACAACACCATGTGGCATAACTTCGCCGCATTTCGGCAA
CACCTTGTTGCCACCACAGCCGACAACACCTTGTGGCATAACCTTCACGCTAAGTCCAACCAATGGGGATT
ATCCTTCGCCGCCGCATCTGCCCTTAGATCCCTTCACGCTAAGTCCCACCAATGGGGATTATCCTTCGCCG
CCGCATCTCCCCTTGGATCCCGGCTCGCCGCCCATTTCGCCCTGCTCCTCCAATTCGGGTAGATTGCCAGC
TCTGGCCCCCGGATCGCCGCCCATTTCGCCCTGCTCCTCCAATTCGGGAAGATTGCCAGCTCTGGCACCGA
TCATTTCCACGCCCTGCTCCTCGGCGGATGCCGATGATGCTGGTGCCGCTATGGGTTTGACCATTTCCACG
CCCTGTTCCTCGGCGGATGCCGATGATGCTGGTGCAGCTATGGGCTTGCCAGTCCGAGCAAGACGTCGTTC
GATTTCGCGACAGGAAGCGATATTCGTGGAGCCCACGCCAGTTCGAGCAAGACGTCGCTCGATTTCGCGAC
AGGAGGCGATATTCGTGGAGCCCACTGGTAACTCACTGGAGAACGTTTCGCACGAGGAGGTGGACAACAGC
AATACCAAGTCATCTGGAAACTCACTGGAGAACGTTTCGCACGAGGAGGTGGACCACAGCAACACCAAGTC
ATCGGTGGACACGGCTGATTCACTGGACGAGGCCTCGACCATGGCCACGTGTGGCTCACCCGGGGTGGACA
CTGCTGATTCACTGGACGAGGCCTCGACAATGGCCATGTGTGGCTCACCCGGAGCCGCCGGTGGCAGTGGT
GCCTCCTCCAGTCATCATAACGCCTTTGTTGTGCGAGATATAGCCGCCGGAGGCAGTGGTGCCTCCTCCAG
TCATCATAACGCCTTTGTGGTGCGGGATATTTACCTAATGGTACCGGATTTGAAGCGAGATCGCGCCGCCT
CCGTGGATTCGTGCTTCTCTTACCTAATGGTACCGGATTTGAAGCGAGATCGCGCCGCCTCCGTGGATTCG
TGCTTCTCCAAGCTCTCGTCGAATGCCAAAACGGAGGAGCTGCAACCAAGTGCCGATGGTTGCTTCCTCAA
GCTCTCGTCGAATGCCAAAACGGAGGAGCTGCAACCAAGGGCCGATGGTTGTTTCCTTACCGTACCGAATA
TAAATGCAACTCGATCGCGATCCGTGGACATCGTGCTACCAACTGATACCGTACCGAATATAAATGCAACT
CGATCGCGATCCGTGGACATCGTGCTACCAACTGACGAGCAAGCCAGATATAAGGCTTTATCCATGACCGG
ATCCACTGTCACCTATGCAGACGGCGAGCAAGCCAGATACAAGGCTTTGTCCATGACCGGATCCACTGTCA
CTTACGCAGACGGGCGTACCGCTTCGGCCAGCAACTCCCGTAGACCGATACGGATTGTGCCCGATTGGACG
GAGAATGCCGTGCAGGGCGAGCACTACTGGAAGCCCTCCTCCGCCTCCGGCGATCTCTGCTGCCTGAACGA
GGAGTGCATTAAAAGCGGCCAACGGATGAAGTGCTCCGCCTGCCAACTGGTTGCCCACCACAACTGCATTC
CGTTCGTCAACGAGAAGAGCACCCTGGCCTGCAAGCCCACCTATCGGGATGTGGGCATCCGGCAGTACCGG
GAGCAGACCACAACCCACCACCACTGGGTGCATCGCAAGCTGGAGAAGGGCAAGTGCAAGCAGTGCGGCAA
GTTCTTTCCTATGAAGCAGGCGGTGCAAAGCAAGCTATTTGGGTCGAAGGAGATTGTGGCGTTGGCCTGTG
CGTGGTGCCACGAGATCTACCACAACAAGGAGGCCTGTTTTAACCAGGCCAAAATCGGCGAGGAATGTCGT
CTGGGGCAACTATGCGCCCATCATTGTGCCGCCATCGTGGATTGTGAAGCTGCCAACCAAGGGCAACTTCA
AGTCCTCGATACGCGTGAGCAACAAGAACAATGCCGCATCGGGATCAGGANNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNCAAGAGCAAGAAGCAGACGCAACGGCGGCAGAAGGGCAAGGAGGAGAAGAAGGAG

CCGCGAGCGTTCATTGTGAAGCCCATACCATCGCCGGAGGTAATCCCGGTCATTGTGTTTATTAATCCCAA
ATCGGGCGGCAATCAGGGCCACAAGCTGCTGGGAAAGTTCCAGCATCTGCTCAATCCGCGACAGGTTTTCG
ATCTTACGCAAGGCGGGCCCAAAATGGGCCTGGACATGTTCCGCAAGGCGCCCAATCTGCGGGTGCTGGCT
TGCGGTGGCGACGGCACCGTCGGCTGGGTCCTTTCCGTCCTGGATCAAATCCAACCGCCGCTCCAGCCAGC
TCCGGCGGTCGGTGTCCTGCCGCTGGGCACTGGAAACGATCTCGCTCGCGCTCTCGGATGGGGAGGGGGCT
ACACGGACGAACCGATCGGGAAGATACTGCGCGAGATCGGGATGTCCCAGTGCGTGCTAATGGATCGCTGG
CGGGTGAAGGTCACGCCCAACGATGACGTCACCGATGACCATGTCGATCGCAGCAAGCCCAATGTCCCATT
GAACGTCATCAACAACTATTTCTCCTTCGGCGTGGACGCGCACATCGCCCTGGAATTCCACGAGGCACGTG
AGGCTCATCCGGAGCGCTTCAATTCGCGCCTGCGCAACAAGATGTACTACGGCCAGATGGGCGGCAAGGAT
CTGATCCTGCGCCAGTACCGCAACCTCTCCCAATGGGTGACGCTCGAGTGCGATGGACAGGATTTTACCGG
AAAACTGCGCGATGCCGGCTGCCATGCCGTGCTCTTCCTAAACATTCCCAGCTATGGCGGTGGCACCCATC
CATGGAACGACTCCTTTGGCGCCTCCAAGCCATCGATTGATGACGGACTGATGGAGGTTGTTGGCCTGACC
ACCTACCAGCTGCCCATGCTGCAGGCTGGCATGCATGGCACCTGCATTTGCCAGTGCCGCAAGGCCAGGAT
CATCACCAAGCGCACCATTCCGATGCAGGTGGATGGAGAGGCATGTCGCGTCAAGCCATCGGTGATCGAGA
TCGAGCTGCTCAACAAGGCACTGATGCTGTCCAAAAGGAAGCAAGGACGCGGCGATGTTCAAGTGAATCCC
CTGGAGAAGATGCAGCTGCACATCTTACGCGTGACCATGCAACAGTACGAGCAGTACCACTATGACAAGGA
GATGCTCCGCAAATTGGCCAATAAGCTGGGCCAAATCGAAATCGAATCTCAGTGCGATCTGGAGCATGTGC
GCAACATGCTCAACACCAAGTTCGAGGAATCCATCTCCTATCCAAAGGTCTCACAGGATTGGTGCTTCATC
GACTCCTGCACCGCTGAGCACTACTTCCGGATCGATCGTGCCCAGGAGCACCTGCACTACATCTGCGACAT
TGCCATCGATGAGCTGTACATCCTGGACCACGAGGCGGCCACCATGCCGCAGACTCCGGACCAGGAGCGCT
CCTTTGCAGCCTTCTCGCAGCGGCAGGCACAGAACGAGCGGCGCCAAATGGATCAGGCCCAGGGCCGCGGC
CCGGGCAGCACCGATGAGGATTTGCAAATTGGCTCCAAGCCCATAAAAGTGATGAAGTGGAAGAGCCCTAT
ACTGGAGCAAACTTCCGATGCCATACTACTAGCAGCCCAGAGCGGCGATCTCAATATGTTACGTGCACTAC
ATGAACAAGGATACTCACTACAGTCAGTGAACAAGAACGGACAGACGGCGCTGCACTTCGCCTGCAAATAC
AACCACAGGGACATTGTCAAATATATCATAGCGAGCGCCACACGACGCCTCATCAACATGGCCGACAAGGA
GCTTGGACAGACGGCGCTTCACATTGCCGCCGAGCAGAATCGCCGCGACATCTGCGTGATGCTGGTTGCCG
CCGGCGCCCATTTGGACACTCTTGACTCCGGCGGAAATACGCCCATGATGGTGGCCTTCAACAAGAACGCC
AACGAGATAGCCACCTATTTGGAAAGTAAGCAGGGGACGCAGCCCGTCGACGGCTGGCTCGACGACTAG

>Dsec_Rh3

ATGGAGTCCGCTAACGTGTCGTCGAGTCTATTTGGCAACGTGTCCACCGCGCTGCGGCCGGAGGCGCGGCT
CTCCGCCGAAACGCGTCTGCTGGGCTGGAATGTGCCGCCGGAGGAGCTGCGTCACATTCCCGAGCACTGGC
TAACTTACCCGGAACCGCCCGAATCGATGAACTACCTGCTGGGCACTCTATACATCTTCTTCACCCTGATG
TCGATGCTGGGCAATGGACTGGTGATTTGGGTCTTCTCGGCAGCCAAATCGCTGCGAACTCCTTCCAATAT
ACTGGTCATCAATCTGGCCTTCTGCGACTTCATGATGATGGTCAAGACTCCAATATTCATCTACAATAGTT
TCCACCAGGGATATGCACTGGGTCATCTGGGATGCCAGATCTTTGGAATCATTGGCTCCTATACGGGAATC
GCTGCCGGTGCCACCAATGCGTTTATAGCCTACGATCGATTCAATGTCATCACGCGACCCATGGAGGGCAA
GATGACGCATGGCAAGGCCATTGCCATGATCATATTCATCTACATGTACGCCACTCCATGGGTGGTTGCCT
GCTACACGGAGACTTGGGGCCGTTTTGTGCCGGAGGGATATCTGACGTCCTGCACCTTTGACTATCTCACC
GATAACTTCGATACGCGCCTCTTTGTGGGCTGCATCTTCTTCTTCAGCTTTGTGTGCCCCACCACGATGAT
CACGTATTACTATTCCCAGATTGTGGGCCATGTCTTTAGCCACGAGAAAGCACTGCGGGATCAGGCCAAGA
AGATGAACGTGGAATCGCTGCGCTCGAATGTGGACAAAAGCAAGGAGACGGCGGAAATCCGGATAGCCAAA
GCGGCCATCACCATCTGCTTCCTGTTCTTTTGCTCATGGACGCCGTACGGAGTTATGTCGCTGATTGGCGC
CTTTGGGGATAAGACCCTTTTGACGCCCGGAGCCACAATGATTCCCGCCTGTGCCTGCAAAATGGTGGCCT
GCATCGATCCGTTCGTGTACGCCATAAGTCACCCCAGATACCGCATGGAGCTGCAGAAGCGATGTCCCTGG
CTGGCGCTCAACGAAAAGGCGCCGGAATCGTCGGCTGTCGCCTCCACCAGTACCACCCAGGAACCACAGCA
GACCACCGCCGCCTAA

>Dsec_Rh5

ATGCACCTCAACGGCCCAAGTGGGCCACAGGCCTATGTGAACGATAGCTCCGGAGATGGCAGCATCTTTCC
CATGGGACACGGATATCCGGCGGAGTACCAGCACATGGTGCACGCCCACTGGCGGGGATTCCGAGAGGCCC
CCATCTACTATCACGCCGGTTTCTACATTGCCTATGTCGTCCTCATGCTGTCGAGTATCTTCGGAAATGGA
CTGGTCATCTGGATATTCTCCACCTCCAAATCCCTGCGTACTCCTTCTAACCTATTGATACTCAACCTGGC
CATTTTCGATCTGTTCATGTGCACCAATATGCCGCATTATCTCATCAACGCCACAGTTGGCTATATAGTGG
GCGGAGATCTGGGCTGCGATATATACGCTTTGAACGGAGGCATCTCCGGGATGGGAGCCTCCATAACGAAT
GCCTTTATAGCATTCGATCGGTACAAGACAATATCTAATCCAATAGACGGACGCTTGAGCTACGGACAGAT
TGTACTGCTGATACTCTTCACCTGGTTGTGGGCCACTCCGTTCTCCGTGCTGCCATTGTTCCAGATCTGGG
GACGTTACCAGCCGGAGGGCTTCCTGACCACCTGCTCCTTCGATTACCTGACCAACACGGACGAAAACCGA
CTATTCGTGCGCACGATCTTCGTGTGGTCCTACGTGATACCGATGACCATGATCCTGGTGTCCTACTACAA

GCTCTTCACCCACGTCCGCGTGCACGAGAAGATGCTGGCGGAGCAGGCCAAGAAAATGAACGTAAAGTCAC
TGTCGGCCAATGCCAATGCGGATAACATGAGTGTGGAGCTACGGATCGCCAAGGCAGCCCTCATAATATAC
ATGCTCTTCATCCTGGCATGGACACCCTACTCCGTGGTGGCACTCATTNNNTGCTTCGGGGAGCAGCAGCT
AAGCACGCCCATCGTCTCTATGCTTCCGTGCCTGGCCTGCAA

>Dsec_stops

ATGCTTTCGCAGTTGCTACAGGCTTTCCAGCGGCTGGTAAGTCAATTAAAAAGTCAATTTGTGTGGGTTCG
CTGCTGGTTGCCAATACCACGGCATTTCTACGGCTTCCGCCGGGAAACTGAGTCACGGGCACGGCTGCGTC
TGCGGCTGCGGGAACGACTGCGGCGGAATCCCGCCGAACGGGCCAGTCACCAGCGGAACCGGCGGACCCAG
CGCCGCCATCTGGTCATGGACTACATCTTCGAGTGCTTCTTCGACGACACCTTCGAGCAGATCTCGCGCAA
CGGCCTGCAGGACCGCCAGTCGCGGCGAGATGTCCTCGACCACCTGAGCTCCATCATCAAGGGCTGCAGCG
GCGGCCAGAACTCGCAGACGGACGAGGTGGCCGCCATCGCGGTGACGGCGGCCATGAGGTACCACCGAATG
GCCAAGGAGCAGAACGGCCAGGTCTGCCTCATGGGCAAGTATCACAACATCCTGTACATCGGACTGCGCAC
CTGCTGGGATTGGGGAGTCCGGGACTCGGAGGTGGTCGTCAAGCTGCTGGTGTCCATTTACGAGTGCGAGA
AGACATATGAGCGCATTTTCCTGGGCGCCCTGTTTGGACCACATGCTCCGCACTTCATCGCCGGCTGGCGC
AGTGACTTCCAGGATCAGCACGAGAATGTCCGGGCGATGGTTTACTTTCTGAAGCACGCCACACGCGAGCA
GCTCACATTGCCCGTCTGGATTCCGCGCTTCGAGCAGGAGCGACAGCTCAGATTCATTGATGTCCCCATTG
AGTCCTGCGGCAAGTCCTCGCCCCTTCGCATTGCCCTGCAGGCGAACGCTCCCGAACTGCTGCTCATCCTG
CTGAGATACGGCGCCGCACCGAAGCCGCCGGACGGAGGCGCCTCCGTAATTATAGCGCTGCTCGACAAACT
CATCGAGGATGGCCGCAATTACAGCTTCGAACTGGTCATGTGCCTCAAGATTCTGCTGCGCAATGTGGTGA
TGATTGAGATGCCCTTCAAGCCCCTTTATACGCTGCCCGCAGGGAAATGTTCTTCGATCGCTACGGCCGGC
TTTTGATGGATAAGATTATTGGCAAAGAGCAGGTCTACGGAGTGCCCAGTCTTCGTCATTTATGCCGCTGC
TGTATCCGCGACGTCCTGCGGATGCACAATCAACTGCCCAACGGCATCGACACGTTGAGGCTGCCAAAGCG
CCTGCAACGCTACATCGATCTCACCGAGGAGCTCGAAGGACCTGAGGCACAGGACACGGAGGACAAGGATC
AGCAGAGGGAAAAATCCTCGGCCAAATGCAAGCTGGGTAGCGAGGTGGCCCGGCTGGATACAATCAGCAGT
GCAGGGGAGAGCGAGGACGAGAGCGAGAACCAAGCGCAGCTCCTGCAGGCTGTCACGGATCCGGAAAAGGC
CGCATTGGCGGCTTTCATGGCCACCGGCGATGGAGCGTATACGGAGGAGGCCNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNTGTCCACCCGATTTCGATCCCGATTTTCCGCCAATTTCCCTGCCCGGGGAGCCCAGA
TAG

>Dsim_ninaA

ATGAAGTCATTGCTCAATCGGATAATCCTGTGCAGCGCCTTTCTGGCCGTGGCCAGTGGTCTGAGCTTCAC
GGTGACGTCGCGGATTTACATGGATGTGAAGCACAACAAGAAGCCGGTGGGCAGGATCACGTTTGGGCTCT
TCGGGAAGCTGGCCCCCAAGACGGTGGCCAACTTCCGGCACATTTGCCTGCGCGGCATTAACGGGACCAGC
TACGTGGGCTCGCGATTCCACCGCGTGGTGGACCGCTTCCTCGTCCAAGGCGGCGACATCCTGAACGGCGA
CGGAACTGGCTCCATTAGCATCTATGGAGACTACTTTCCGGACGAGGACAAGGCTCTGGCGGTGGAGCACA
ACAGACCCGGTTACTTGGGCATGGCCAATCGGGGTCCGGACACCAATGGCTGCCAGTTTTATGTGACCACC
GTGGGCGCCAAGTGGCTGGACGGAAAGCACACCGTATTCGGCAAGGTGCTGGAGGGAATGGACACCATCTA
TGCCATTGAGGATGTAAAAACCGATACGGACGACTTCCCTGTGGAACCCGTGGTGATCTCCAACTGCGGCG
AGATCCCCACGGAGCAGTTCGAGTTCTACCCGGACGACTTCAACATACTCGGATGGATAAAGGCGGCTGGT
CTGCCCGTGACCAGCTCCTTCTGCGTTCTGCTCATTTTCCACTACTTCTTCCGCCAGCTCAACATGTACTG
CTGA

>Dsim_ninaC

ATGTATTTACCGTACGCGCAATTGCCGGATCCCACGGATAAATTCGAGATCTACGAGGAGATAGCCCAGGG
CGTCAATGCCAAGGTATTTCGGGCCAAGGAGCTCGACAATGATCGGATCGTGGCCCTAAAGATCCAGCACT
ACGATGAAGAGCACCAGGTTTCCATCGAGGAGGAGTATCGCACTCTGCGGGACTACTGCGACCATCCCAAT
CTCCCCGAGTTCTATGGCGTCTACAAGCTGTCGAAACCAAACGGACCTGATGAGATCTGGTTCGTCATGGA
GTACTGCGCTGGAGGAACAGCAGTGGACATGGTTAATAAGCTTTTGAAACTGGATCGTCGAATGCGAGAGG
AGCACATCGCCTACATCATCCGGGAGACTTGCCGTGCTGCTATCGAGCTGAACCGTAATCATGTCCTGCAT
CGGGACATTCGGGGAGATAACATTTTGTTGACCAAAAATGGACGGGTCAAGCTTTGCGATTTCGGCTTATC
CCGTCAGGTGGATTCCACGCTGGGCAAGAGGGGCACCTGCATTGGATCTCCCTGCTGGATGGCTCCCGAAG
TGGTGTCCGCCATGGAATCACGCGAACCGGACATCACCGTTCGTGCCGATGTCTGGGCCTTGGGAATCACC
ACCATTGAGTTGGCCGATGGAAAGCCACCATTCGCTGACATGCATCCCACTAGGGCCATGTTCCAGATCAT
TCGCAATCCACCGCCGACCTTAATGCGACCCACTAATTGGTCACAGCAGATCAATGATTTTATATCCGAGA
GTCTGGAGAAAAATGCCGAAAATCGTCCAATGATGGTTGAGATGGTGGAGCATCCATTCCTCACGGAGCTC
ATCGAAAACGAGGACGAGATGCGATCGGACATCGCGGAAATGTTGGAACTCTCCCGGGACGTAAAGACTTT
GTACAAGGAGCCGGAATTGTTCGTGGACCGCGGCTATGTGAAGAGGTTTGATGAGAAGCCGGAGAGAATGT
ATCCGGAGGACTTGGCTGCCTTGGAAAATCCCGTGGATGAAAACATAATCGAGTCACTGCGTCATCGCATT
TTGATGGGGGAATCGTACAGTTTCATTGGCGATATCCTGTTATCCTTGAATTCCAATGAAATCAAGCAGGA

ATTTCCGCCGGAGTTCCATGCCAAGTATCGCTTCAAGTCGCGTTCGGAGAACCAGCCGCATATTTTCTCGG
TGGCCGATATCGCCTACCAGGATATGCTGCATCATAAGGAACCGCAACACATCGTGCTCTCCGGAGAGAGT
TACTCGGGCAAGTCCACAAATGCCCGGCTGCTGATCAAGCATCTGTGCTACCTGGGCGATGGAAATCGTGG
GGCCACCGGACGCGTGGAGAGCTCCATCAAGGCCATTCTGATGTTGGTGAATGCAGGAACTCCGGTGAACA
ACGACTCCACCAGATGTGTTCTACAATACTGCTTGACCTTCGGAAAAACCGGCAAGATGAGTGGGGCCGTT
TTCAATATGTATATGCTGGAGAAACTACGAGTCGCCACTACGGATGGAACCCAGCACAACTTCCACATCTT
CTACTACTTCTACGACTTTATCAATCAGCAGAATCAGCTCAAGGAGTACAATCTGAAGGCTGATCGCAATT
ATCGTTATCTGCGAGTTCCGCCAGAGGTTCCGCCATCGAAACTCAAGTATCGTCGTGATGATCCCGAGGGT
AATGTGGAGAGGTACAGGGAGTTCGAGAACATCCTGCGGGACATTGATTTCAATCACAAGCAATTGGAAAC
GGTTCGCAAAGTTCTGGCTGCCATTTTGAATATTGGCAACATCCGTTTCCGGCAGAACGGAAAGTATGCGG
AAGTGGAGAACACCGACATTGTGTCCCGGATCTCCGAGCTCCTCCGCGTGGATGAGAAAAAGTTCATGTGG
TCTCTGACCAACTTCATCATGGTCAAGGGCGGCATCGCCGAGAGGCGACAGTACACCACTGAAGAAGCCAG
GGATGCACGCGATGCGGTGGCCAGCACGCTATACTCTAGATTGGTGGACTTTATAATCAACAGGATCAACA
TGAACATGTCCTTCCCACGGGCTGTGTTCGGCGACACAAATGCCATCATTATCCACGACATGTTTGGCTTT
GAGTGCTTCAATCGAAATGGTCTGGAGCAACTAATGATCAATACTCTAAACGAGCAAATGCAATATCACTA
CAATCAGCGCATTTTTATCAGCGAAATGCTGGAAAAGGAAGCTGAAGACANTGACACCATCAACTTGAATT
TCTATNACAACAAGACGGCCCTGGATAATCTGCTAACCAAACCGGATGGCCTGTTCTACATCATTGACGAT
GCCTCACGCTCCTGTCAAGATCAGGATTTGATTATGGATCGCGTTTCGGAGAAGCACAGCCAGTTTGTGAA
GAAGCACACTGCCACTGAAATATCCGTGGCTCATTACACGGGCCGCATCATCTATGATACGCGCGCCTTCA
CCGACATCAACCGGGACTTTGTGCCGCCGGAAATGATCGAGACCTTCCGCTCCTCCCTGGATGAGAGCATC
ATGCTCATGTTCACCAATCAACTGACCAAGGCTGGCAATCTCACCATGCCATTCGAGGCAGTGCAGCATAA
GGATGAATCTGAACGGAAGTCATATGCCTTGAATACCCTAAGTGCCGGATGCATCTCCCAGGTGAACAATC
TCCGTACCCTGGCGGCCAACTTCCGATTCACCTGCCTGACCCTGCTCAAAATGCTAAGCCAGAATGCAAAT
CTCGGCGTGCACTTCGTCCGCTGCATTCGCGCTGATCTGGAGTACAAGCCCAGATCCTTCCACTCGGATGT
TGTTCAGCAACAGATGAAGGCCTTGGGAGTTCTGGACACTGTTATTGCTCGCCAGAAGGGATTCAGCTCAC
GGCTGCCATTCGAAGAGTTCCTAAGGCGATACCAATTCCTTGCCTTTGACTTTGATGAGCCCGTGGAAATG
ACCAAGGATAACTGTCGTCTATTATTCTTGCGCCTCAAGATGGAAGGCTGGGCCTTGGGCAAGACCAAGGT
GTTCTTGCGCTACTATAACGATGAGTTCTTGGCCAGATTGTATGAGCTGCAGGTGAAGAAGGTCATCAAAG
TGCAGTCCATGATGCGAGCTTTGCTAGCACGAAAACGAGTTAAGGGCGGCAAAGTGTTCAAATTGGGCAAA
AAGGGACCGGAACATCATGATGTGGCCGCGTCCAAAATACAAAAAGCCTTCAGGGGATTCCGCGACCGCGT
CCGCCTCCCTCCGCTAGTCAACGAGAAGTCCGGCCAACTTAACGAGAACACCGTTGACTTTATCCGCCCGT
TTGCCAAGAAATGGCGCGAGAAGTCCATCTTCCAGGTCCTCCTGCACTACCGCGCGGCTCGCTTCCAGGAC
TTTGTCAATCTCTCACAGCAGGTCCACATCTACAACCAACGAATGGTGGCTGGCTTGAATAAATGTACTCG
GGCTGTGCCCTTCGAAAGGATCAACATGCGCGAGGTGAACTCCTCCCAGCTCGGACCACTGCCAGTGCCCA
TAAAGAAGATGCCATTCCGCCTGGATCAGATACCCTTCTACGACACCCAGTATATGGTGGACCCGGCCAAC
TCCATTTCCCGCCAAACCTTCCCCAATCAGCTACTAACGCAGCACATGGAGGACGATGAGCCTTGGGACAG
CCCTCTGCAACGCAATCCCTCGATGACTTCTTGTGCCCTGACCTACAATGCCTATAAGAAGGAGCAGGCTT
GCCAGACCAACTGGGATCGCATGGGCGAGAGCGACAATATCTACAATCAGGGCTACTTCCGAGATCCCCAA
CAGCTGAGGAGAAATCAAATGCAGATGAACATGAATGCCTACAACAATGCCTACAACAGCTATAACAGCAA
CTATAACAACCAGAACTGGGGCGTACATCGATCCGGATCCAGGCGCAACTCCTTAAAGGGCTATGCGGCGC
CTCCACCTCCTCCTCCACCAATGCCCTCATCCAACTATTATCGCAATAATCCGAACCAGCAGCAAAGCAAC
TATCAGCCAAGATCCTCGTATCCACCATCGGATCCAGTTAGAGAACTCCAGAACATGGCTCGCAACGAGGG
TGATAACTCTGAGGATCCACCATTCAACTTTAAGGCCATGCTGCGCAAGACTAACTATCCGAGGGGATCCG
AGACCAACACCTATGACTTTAACAATCGCCGGGGATCGGACAGTGGTGACCAACACACGTTCCAGCCGCCA
AAGCTACGTTCGACAGGTCGCCGGTACCAGGACGATGAGGGCTATAACTCATCATCGGGAAACTATGGTGT
GTCTAGGAACTTTGGCCAGCAGCAGAGGGCTCCGAATTTGAGACAATCTCAGGCCAGCGTGGGTCGCAGCT
TTGAAGATAGCAATGCCAGGTCCTTCGAGGAGGCAGGATCTTATGTCGAGGAGGAGATTGCACCTGGAATC
ACTCTTTCAGGCTATGCCGTCGATATCTAA

>Dsim_norpA

ATGACCAAGAAGTACGAGTTCGATTGGATCATTCCGGTTCCGCCGGAATTGACCACGGGCTGCGTTTTCGA
TCGCTGGTTCGAAAACGAAAAGGAGACAAAGGAGAATGACTTTGAGCGCGATGCCCTCTTCAAGGTCGACG
AATACGGATTTTTTCTGTACTGGAAAAGCGAGGGCCGGGATGGCGATGTCATAGAACTCTGCCAAGTAAGC
GACATTCGTGCGGGCGGAACACGAAAGGATCCGAAGATACTTGATAAGGTGACGAAGAAGAACGGCACCAA
TATACCGGAACTGGATAAGCGATCCCTGACGATCTGTTCGAACACGGACTATATCAATATAACATATCACC
ATGTTATTTGTCCAGATGCGGCAACAGCCAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTGGATGCGATTGAGTTACTGCGTAGAGAA
GAGTGGCAAAATTCCAGTTAAAACGCTGGCCAAAACCTTCGCATCCGGCAAAACGGAGAAATTGGTGTACA

CGTGCATAAAGGATGCCGGTCTGCCCGATGATAAAAACGCAACGATGACCAAGGAGCAGTTCACCTTCGAC
AAGTTCTACGCCTTGTACCACAAGGTGTGTCCCCGAAACGACATTGAGGAGCTCTTCACCTCCATCACCAA
GGGCAAGCAGGACTTCATCAGCTTGGAGCAATTTATCCAGTTCATGAACGACAAACAGCGCGATCCGCGGA
TGAACGAAATCCTGTACCCTCTCTACGAGGAGAAACGCTGCACGGAGATCATCAACGACTACGAGCTAGAT
GAGGAGAAAAAGAAGAACGTTCAAATGTCGCTGGACGGATTTAAGCGCTACCTAATGTCCGACGAAAACGC
ACCCGTATTCCTGGACCGGCTGGATTTCTACATGGAAATGGATCAGCCACTGGCCCATTACTATATCAACA
GCTCGCATAACACCTACCTATCTGGTCGTCAGATCGGCGGCAAAAGTTCCGTGGAAATGTACCGCCAGACA
CTCTTGGCAGGCTGTCGCTGCGTGGAGCTGGATTGCTGGAACGGAAAGGGTGAGGACGAGGAGCCCATCGT
CACCCACGGCCACGCCTACTGCACCGAAATCCTCTTTAAGGACTGCATCCAGGCGATTGCGGATTGCGCCT
TCGTGTCCTCCGAGTATCCGGTAATCCTGTCCTTCGAAAACCACTGCAACCGCGCCCAGCAATACAAGTTG
GCCAAATACTGTGATGACTTCTTTGGCGATCTGCTGCTGAAGGAGCCGCTACCAGATCATCCGCTGGATCC
GGGCCTTCCGTTGCCGCCACCCTGCAAACTGAAGCGTAAGATCCTCATCAAGAACAAGCGAATGAAGCCAG
AAGTGGAAAAGGTCGAGCTGGAGCTCTGGCTGAAGGGCGAACTCAAGACGGATGACGATCCGGAAGAGGAC
GCCAGTGCGGGCAAGCCGCCAGAGGCAGCCGCCGCCCCCGCACCCGCCCCGGAAGCAGCCGCCGCCGCCGA
AGGAGCGGCCGAGGGTGGCGGAGGAGCGGAGGCCGAAGCCGCCGCTGCCAACTACAGCGGCTCCACAACTA
ACGTGCATCCGTGGCTCTCCTCCATGGTCAATTACGCGCAGCCCATCAAGTTCCAGGGCTTCGACAAGGCA
ATCGAAAAGAACATTGCCCACAACATGTCCTCGTTTGCGGAATCGGCGGGCATGAACTACTTGAAGCAGAG
CTCCATCGACTTTGTCAATTACAACAAGCGTCAGATGTCGCGAATTTATCCCAAGGGCACACGAGCGGACT
CCTCAAACTATATGCCGCAGGTGTTCTGGAACGCCGGCTGCCAGATGGTCTCACTCAACTTCCAGAGCTCC
GATTTACCCATGCAACTCAACCAGGGCAAGTTCGAGTATAACGGCGGCTGTGGCTATCTACTGAAACCGGA
TTTCATGCGTCGAGCCGACAAGGATTTTGATCCGTTTGCGGATGCGCCGGTGGACGGCGTGATTGCGGCCC
AGTGTTCCGTGAAGGTGATTGCCGGCCAATTCTTGTCCGACAAGAAAGTGGGCACCTATGTGGAGGTGGAC
ATGTTTGGATTGCCCTCGGACACGGTGAAGAAGGAGTTTCGCACGCGTTTGGTCGCCAATAATGGCCTGAA
TCCAGTTTACAATGAGGATCCCTTTGTGTTCCGCAAAGTGGTCCTTCCGGACTTGGCTGTGCTAAGATTTG
GCGTTTATGAGGAAAGCGGAAAGATTCTGGGTCAACGTATTCTGCCGCTGGACGGTCTACAGGCTGGCTAT
CGGCATGTTTCCCTGCGCACGGAGGCTAACTTCCCCATGTCGTTGCCCATGTTGTTCGTGAATATCGAGCT
AAAGATCTACGTACCTGACGGCTTTGAGGACTTCATGGCCATGTTGTCGGATCCGCGAGGATTCGCCGGTG
CCGCTAAGCAGCAAAACGAACAGATGAAGGCCCTTGGCATAGAAGAGCAGAGCGGCGGTGCCGCCCGAGAT
GCTGGCAAGGCCAAAGAGGAGGAAAAGAAGGAGCCACCACTAGTCTTTGAGCCCGTCACGTTGGAATCTCT
GCGCCAGGAGAAAGGCTTCCAAAAGGTGGGCAAAAAGCAAATCAAGGAGCTCGACACCCTGCGCAAGAAGC
ACGCCAAAGAGCGCACCTCGGTGCAAAAGACCCAGAATGCAGCCATCGACAAGTTGATCAAGGGCAAGAGC
AAAGACGACATTCGCAACGATGCCAACATCAAGAACTCGATCAATGACCAGACCAAACAGTGGACCGACAT
GATCGCCAGGCACCGTAAGGAGGAATGGGACATGCTGCGCCAACATGTCCAGGACTCGCAGGACGCCATGA
AAGCACTGATGCTCACCGTTCAGGCGGCGCAGATCAAGCAGCTGGAGGATCGTCATGCCAGGGACAACGCA
TTCAAGGATCTGAATGCCAAGCAAGCAAAGATGTCGGCGGATACCGCCAAGGAGGTACAGAACGACAAGAC
CTTGAAGACCAAGAACGAGAAGGATCGTCGGCTGCGTGAGAAGCGCCAGAACAATGTGAAGCGCTTCATGG
AGGAAAAGAAGCAAATCGGCGTTAAGCAGGGCCGTGCGATGGAGAAACTAAAGTTGGCGCATTCGAAGCAG
GTCGAGGAATTCAGTACCGACGTGCAAAAGCTTATGGACATGTACAAAATCGAGGAGGAGGCGTATAAGAC
GCAAGGAAAAACGGAATTTTATGCCTAA

>Dsim_rdgA

ATGCAGCAGCAGCAACAACCGTCCATCGACCAGCTGCCCGAGCCCACAGCCTCGACCAGCAATTCGGCAAC
AACTAAGCCGACAACTGCAGCAGCAACCACATCGACAACGACAACGTCTGGCAACAACTTCCACCAGCAAC
TGCAAGCAACAACGGCAGCCACCATGCAACGCCTGCGAACAACCTTCACGCGATCCCGCACTCCCACAGGT
GCGGAAATGAAGATGCAGAATAGTCTGGAGGTGCCCAAGCAGGTGCGCTCAGCCTCCTTCGACGAGATGCA
ACTGGAGTCGCAGCGTGCCTCCTCCAGTCTGCTCAAGCAGCAGTCCTCATCATCGGCCTCGGCGGACGAAC
GCTCCTCGGAGGCTGGCTTCCTGCAGGTTCCATTGGCGGCCCATCAGCAAAGATCCCATAGCTTCGACTCG
GCAACTGCGTCGGCGGGCAGCGATGACTCGGGCACCTTTTTGGAGGTGCCCAGACGTCTGAAGGCTCGGCG
CAGTTCGTCCACCAAGACTCCGCCGCCTTGCATCCATTGTCACTATCTGGAGGAGTATGAGCGTCGTATGA
CCGCCGAGCAGCGTTACTTTATCGATCATCGCGAGCTAACAGCTTTGAGCTACAGCAATACCAGTTCGGAG
GCGAGCGAGGATGAGGATGAAGTGGAGGGCCATAACGCGGAGGAAGAGGAGGAGGGCAGTGCGGCCATCGA
GGATGCTGAGGAGGAGACAACTGAGGCGGCTACCGAGGAGGCGGATGAAGATCCCCGCACCGAGGTTGAGT
CCGAGCACGATCACGATCCGGATGACGATGCCGCATTAGAGATGGACATCCGCATTGGCAACATGAGCCAA
GGCAGCAGCATCGAGGAGTCGCGCGCCCGCCTGCCGCGACAAATGCGGCGCCACACCATCGGTAGCAGCTC
GGTCACCTCGGCCTCCGAGGATGAGGGCTTGGAGGGCTCAGACAATGGTTCGCCGCATTTCGGCAACACCT
TGTTGCCACCACAGCCGACAACACCTTGTGGCATAACCTTCACGCTAAGTCCCACCAATGGGGATTATCCT
TCGCCGCCGCATCTCCCCTTGGATCCCGGCTCGCCGCCCATTTCGCCCTGCTCCTCCAATTCGGGAAGATT
GCCAGCTCTGGCACCGACCATTTCCACGCCCTGTTCCTCGGCGGATGCCGATGATGCTGGTGCAGCTATGG

GCTTGCCAGTCCGAGCAAGACGTCGCTCGATTTCGCGACAGGAGGCGATATTCGTGGAGCCCACTGGAAAC
TCACTGGAGAACGTTTCGCACGAGGAGGTGGACAACAGCAACACCAAGTCATCGGTGGACACGGCTGACTC
ACTGGACGAGGCCTCGACAATGGCCACGTGTGGCTCACCCGGAGCCGCCGGTGGCAGTGGTGCCTCCTCCA
GTCATCATAACGCCTTTGTGGTGCGGGATATTTACCTAATGGTACCGGATTTGAAGCGAGATCGCGCCGCC
TCCGTGGATTCGTGCTTCTCCAAGCTCTCGTCGAATGCCAAAACGGAGGAGCTGCAACCAAGTGCCGATGG
TTGCTTCCTTACCGTACCGAATATAAATGCAACTCGATCGCGATCCGTGGACATCGTGCTGCCAACTGACG
AGCAAGCCAGATACAAGGCTTTGTCCATGACCGGATCCACTGTCACCTACGCAGACGGGCGTACCGCTTCG
GCCAGCAACTCCCGCAGACCGATACGGATTGTGCCCGATTGGACGGAGAATGCCGTGCAGGGCGAGCACTA
CTGGAAGCCCACCTCCGCCTCCGGCGATCTCTGCTGCCTGAATGAGGAGTGCATTAAAAGCGGCCAACGGA
TGAAGTGCTCCGCCTGCCAACTGGTTGCCCACCACAACTGCATTCCGTTCGTCAACGAGAAGAGCACCCTG
GCCTGCAAGCCCACCTATCGGGATGTGGGCATCCGGCAGTACCGGGAGCAGACCACCACCCACCACCACTG
GGTGCATCGCAAGCTGGAGAAGGGCAAGTGCAAGCAGTGCGGCAAGTTCTTTCCTATGAAGCAGGCGGTGC
AAAGCAAGCTATTCGGGTCGAAGGAGATTGTGGCGTTGGCCTGTGCGTGGTGCCACGAGATCTACCACAAC
AAGGAGGCCTGTTTTAACCAGGCCAAAATCGGCGAGGAATGTCGTCTGGGCAACTATGCGCCCATCATTGT
GCCGCCATCGTGGATTGTGAAGCTGCCAACCAAGGGCAACTTCAAGTCCTCGATACGCGTGAGCAACAAGA
ACAATGCCGCATCGGGATCAGGAGGTGGCGGTGCTGGTGGCGGTGCCGGAGGCGGTGGTGGCAAGAGCAAG
AAGCAGACGCAGCGGCGACAGAAGGGCAAGGAGGAGAAGAAGGAGCCGCGAGCGTTCATTGTGAAGCCCAT
TCCATCGCCGGAGGTAATCCCGGTCATTGTGTTTATTAATCCCAAATCGGGCGGCAATCAGGGCCACAAGC
TGCTGGGAAAGTTCCAGCATCTGCTCAATCCGCGCCAGGTTTTCGATCTTACGCAAGGCGGGCCCAAAATG
GGCCTGGACATGTTCCGCAAGGCGCCCAATCTGCGGGTGCTGGCTTGCGGTGGCGACGGCACCGTCGGCTG
GGTCCTTTCCGTCCTGGATCAAATCCAACCGCCGCTCCAGCCAGCTCCGGCGGTCGGTGTCCTGCCCCTGG
GCACTGGAAACGATCTCGCTCGCGCTCTCGGATGGGGAGGGGGCTACACGGACGAACCGATCGGGAAGATA
CTGCGCGAGATCGGGATGTCCCAGTGCGTGCTAATGGATCGCTGGCGGGTGAAGGTCACGCCCAACGATGA
CGTCACCGATGACCATGTCGATCGCAGCAAGCCCAATGTCCCATTGAACGTCATCAACAACTATTTCTCCT
TCGGCGTGGACGCGCACATCGCCCTGGAATTCCACGAGGCACGTGAGGCCCATCCGGAGCGCTTCAATTCG
CGCCTGCGCAACAAGATGTACTACGGCCAGATGGGCGGCAAGGATCTGATCCTGCGCCAGTACCGCAACCT
CTCCCAATGGGTGACGCTCGAGTGCGATGGACAGGATTTTACCGGAAAACTACGCGATGCCGGCTGCCATG
CCGTGCTCTTCCTAAACATTCCCAGCTATGGCGGTGGCACCCATCCATGGAACGACTCCTTTGGCGCCTCC
AAGCCATCGATTGATGACGGACTGATGGAGGTTGTTGGCCTCACCACCTACCAGCTGCCCATGCTGCAGGC
TGGCATGCATGGCACCTGCATTTGCCAGTGCCGCAAGGCCAGGATCATCACCAAGCGCACCATTCCGATGC
AGGTGGATGGAGAGGCATGTCGCGTCAAGCCATCGGTGATCGAGATCGAGCTGCTCAACAAGGCACTGATG
CTCTCCAAAAGGAAGCACGGACGCGGCGATGTTCAAGTGAATCCCCTGGAGAAGATGCAGCTGCACATCTT
GCGCGTGACCATGCAACAGTACGAGCAGTATCACTATGACAAGGAGATGCTCCGCAAATTGGCCAATAAGC
TGGGCCAAATCGAAATCGAATCTCAGTGCGATCTGGAGCATGTGCGCAACATGCTCAACACCAAGTTCGAG
GAATCCATCTCCTATCCAAAGGTCTCACAGGATTGGTGCTTCATCGACTCCTGCACCGCTGAGCACTACTT
CCGGATCGATCGTGCCCAGGAGCACCTGCACTACATCTGCGACATTGCCATCGATGAGCTGTACATCCTGG
ACCACGAGGCGGCCACCATGCCGCAGACTCCGGACCAGGAGCGCTCCTTTGCAGCCTTCTCGCAGCGGCAG
GCGCAGAACGAGCGGCGCCAAATGGATCAGGCCCAGGGCCGCGGCCCGGGCAGCACCGATGAGGATTTGCA
AATTGGCTCCAAGCCCATAAAAGTGATGAAGTGGAAGAGCCCTATACTGGAGCAAACTTCCGATGCCATAC
TACTAGCAGCCCAGAGCGGCGATCTCAATATGTTACGTGCACTACATGAACAAGGATACTCACTACAGTCA
GTGAACAAGAACGGACAGACGGCGCTGCACTTCGCCTGCAAATACAACCACAGGGACATTGTCAAATATAT
CATAGCGAGCGCCACACGACGCCTCATCAACATGGCCGACAAGGAGCTTGGACAGACGGCGCTTCACATTG
CCGCCGAGCAGAATCGCCGCGACATCTGCGTGATGCTGGTTGCCGCCGGCGCCCATTTGGACACCCTTGAC
TCCGGCGGAAATACGCCCATGATGGTGGCCTTCAACAAGAACGCCAACGAGATAGCCACCTATTTGGAAAG
TAAGCAGGGGACGCAGCCCGTCGACGGCTGGCTGGACGACTAG

>Dsim_rdgB

ATGCTGATCAAGGAGTACCGCATTCCGCTGCCCCTCACCGTCGAGGAGTACCGCATCGCCCAGCTCTACAT
GATTGCGAAAAAGAGTCGCGAGGAGAGTCATGGCGAGGGCAGTGGCGTTGAGATAATCATCAATGAGCCGT
ACAAGGATGGACCCGGCGGAAATGGTCAATACACGAAAAAGATCTACCATGTAGGCAATCATCTGCCTGGC
TGGATTAAAAGTGAGTTGAATACCTGTGTTAATTACGTTTTGTATGCTATAATTAATCGTTTGTGCGCAGG
TCTCTTGCCGAAAAGCGCTTTAACCGTTGAGGAGGAGGCATGGAATGCCTATCCGTATACCAGGACTCGCT
ACACCTGTCCGTTTGTGGAGAAGTTCTCGCTGGACATTGAGACATACTACTATCCGGACAATGGCTATCAG
GACAATGTCTTCCAGCTGTCCGGTAGCGATCTGCGTAATCGTATCTGTGTCCTTGCAGACGNNNTTGACAT
TGTCAAGGATCAGCTGTGGGGCGGTGACTATGTGAAGGAGGAGGATCCCAAGCACTTTGTGTCGGACAAGA
CGGGCCGCGGACCCTTGGCCGAGGATTGGCTGGAGGAGTATTGGCGCGAAGTGAAGGGCAAAAAGCAGCCG
ACACCGCGCAACATGTCCCTGATGACCGCCTACAAGATCTGCCGCGTGGAGTTCCGCTACTGGGGCATGCA
GACGAAGCTAGAGAAGTTCATCCATGATGTGGCGCTGCGCAAGATGATGCTGCGTGCCCATCGGCAGGCGA

GCTGGAGCGGCAGACGCAACTGGCCCTGGCCAAGAAGATGGGCGGCGGTGGGCATGGCAGGACGAGTGGTT
CGGCTTGACCATCGAGGATATACGCGCGAGGAGTGCAGCGACGACAGCGTCTCGGAGCCGTATGTCAGCAC
GGCGGCCACCGCCGCCTCCGCAACGGGCAGCGAGCGAAAGAAGTCCGCTCCGGCTGTGCCGCCTATTGTCA
CCCAGCAGCCGCCGAGCGCCGAGGCCAGTTCGGATGAGGAGGGCGAGGAGGAGGNNNATGACGACGAGGAC
GAGAACGATGCCATTGGCACGGGCGTGGATCTGTCTGCCAACCAAGGCGGATCTGCGCAGCGCTCGCGCTC
CCAAAGTATTCAGATGGCCCAGAAGGGAAAGTTCGGGTCAAAGGGCGCCCTGCACTCGCCGGTGGGATCTG
CCCATAGCTTCGATCTCCAGGTGGCTAACTGGCGCATGGAGCGATTGGAAGTGGACTCCAAATCCAATTCG
GATGAGGAATTCTTTGATTGCCTGGACACCAACGAGACGAACTCGCTGGCCAAGTGGAGCTCGCTGGAGCT
GCTGGGCGAGGGCGACGACAGTCCGCCGCCACATGGCGGACCCTNTAGTGCAGCATCGGTGGGTGGGCGTG
GCAACTCGCGGCAAGAGGACAGCATATTCAATCAGGACTTTCTGATGCGCGTAGCCTCGGAGCGCGGCAAC
AAGCGACAGTTACGCTCCTCGGCCAGCGTGGATCGCAGTCACGATTCATCGCCGCCGGGATCACCGAGCAC
ACCGTCGTGTCCCACAACCATTCTCATTCTGGTTGTCCATGCGGGCAGCGTTTTGGATGCGGCCAGCGAGC
TGACCGCCAAGAAATCCGATGTGACCACATTCCGTGGCTCCTTCGAGGCGGTGATGCGACAGCACTATCCC
AGCCTCCTCACCCATGTGACCATCAAGATGGTGCCGTGCCCCTCAATATGCACCGACGCCCTGGGCATTCT
CTCCAGCCTGAGTCCGTACTCCTTTGATGCGTCGCCCTCGGCGGCGGATATACCGAATATAGCCGATGTCC
CCATTGGAGCCATACCACTACTATCCGTGGCATCGCCAGAATTCCACGAGACGGTCAACAAGACGGTTGCC
GCTGCCAATATTGTCTACCATGAGTTTTTGAAATCGGAGGAGGGTCACGGATTCTCCGGCCAGATTGTCAT
GCTGGGCGATTCGATGGGCTCGCTGCTGGCGTACGAGGCCCTCTGCCGATCGAATGGCAGCCAGCCGGGCA
CGGCTTCGGGTGCCTCGAATTCCGGCGGAGATGCGGCCACAAATATAAATACCCACAATCCGTTGAGCGCT
CGTAATTCGCGATTGGACGACGACGAGCGTTTCATCGAAGCCGATCTGGATGCCAAGCGTTTGCTGGTGGC
CCCATCGCCACGTAGACGCCGTTCCAGCTCCTCCAGCGATTCGCGGGCCACCAAATTGGACTTTGAGGTCT
GTGACTTCTTCATGTTCGGATCGCCGCTGTCTGTGGTGCTGGCTGCACGGAAACTTCACGATGCCAAGGCC
GCCCTGCCGCGGCCCAACTGCCACCAGGTCTACAATCTGTTCCATCCAACCGATCCGATCGCCTCGCGCCT
GGAGCCGCTACTGAGCGCCCGGTTTTCTATATTGGCGCCAGTCAATGTCCCACGGTACGCCAAGTATCCGC
TGGGTAATGGCCAGCCATTGCATTTATTGGAGGTCATTCAGTCGCATCCGCAGCACTTTAACGATGGCAAC
AACCTATTGGCTGGTCGCCGTTTGTCGGACGCATCCATGCAGAGCACGATATCGGGTCTGATTGAGAATGT
CTCGCTTAGTACGATCCATGCCCTGCAAAACAAATGGTGGGGCACAAAGCGCTTGGATTTCGCATTATATT
GCCCGGAGGGATTGAGTAACTTCCCTGCTCACGCCTTGCCGCACCTCTTCCATGCCAGCTACTGGGAGAGT
CCGGATGTGATTGCCTTTATTCTACGGCAGATTGGCAAATTCGAGGGCATACCCTTTGTGGGCTCCAACGA
TGACAAGGACAATGCCTCCTTCCATCCCGGACAGCCGAGGGAGAAGTGGATCAAGAAACGGACCTCGGTTA
AGCTGAAAAATGTGGCCGCCAATCATCGGGCCAACGATGTGATCGTGCAGGAGGGCAGGGAGCAGCGCTTG
AATGCGAGATTTATGTACGGACCCCTCGACATGATCACGCTGCACGGTGAAAAGGTGGATGTGCACATTAT
GAAGGATCCGCCGGCGGGCGAGTGGACATTCCTCAGCACCGAGGTGACGGACAAGAACGGTCGCATCTCGT
ACAGCATTCCGGATCAGGTATCCCTTGGCTATGGCATATATCCGGTTAAGATGGTGGTCCGTGGCGATCAC
ACCTCGGTGGATTGCTATATGGCGGTGGTGCCGCCGTTGACCGAATGCGTGGTCTTCAGCATTGATGGCTC
CTTCACCGCTTCGATGTCGGTGACGGGCAGGGATCCCAAGGTGCGTGCCGGAGCTGTCGATGTTTGCCGCC
ACTGGCAGGAGCTGGGCTACCTGCTCATCTACATCACCGGACGACCGGATATGCAGCAGCAACGCGTGGTG
TCCTGGCTGAGCCAGCACAACTTCCCGCACGGCCTGATCTCGTTCGCCGACGGCCTGTCCACCGATCCATT
GGGCCACAAGACGGCCTATCTCAACAATTTGGTTCAGAACCATGGAATCTCAATTACTGCCGCCTACGGCA
GCAGCAAGGACATTAGTGTCTACACGAATGTTGGCATGCGAACCGATCAAATCTTCATCGTGGGCAAGGTT
GGCAAGAAGCTGCAGTCGAATGCCACCGTGCTTAGCGATGGCTATGCCGCCCACTTGGCCGGTTTGCAGGC
TGTGGGTGGTTCGCGTCCGGCGAAGGGCAATGCCCGCATGGTCATTCCCCGCGGATGCTTCAATCTTCCCG
GCCAGACCGCGAATCCGCGGCGCAGAAGGTAAGACGCTGA

>Dsim_rdgC

ATGGATGAGAACGCTATCCGAGCGGCAATCTTCATCCAGAAATGGTATCGCAGGCATCAGGCCCGTCGCGA
AATGCAGAGGCGTTGCAACTGGCAGATCTTCCAGAACCTGGAGTACGCCAGCGAACAGGACCAGGCTGAGC
TGTACAAATTCTTCAATGACCTCATCAAGCACATGCCCCAGGCGGCGGGCAGGAAGAACCAGTACCAAGGA
TCCGCCCATGTTTCCGTTTTGGATGATAAGGACGATCTCGTCGAAGAGTTTGGAGACATTGTGAATGCCAA
AATTGAGTTGCCAATACGAAAAAATCATATCGATCTATTGATTGACGTCTTCCGCAAGAAACGGGGCAACC
GATTGCATCCCAAGTATGTGGCCCTGATTCTACGTGAAGCTGCCAAATCGCTAAAGCAGCTGCCCAATATT
TCTCCGGTGTCCACGGCGGTTTCCCAACAAGTTACGGTTTGCGGGGATCTGCATGGAAAATTGGACGATCT
TCTTGTGGTGCTGCATAAGAACGGTCTTCCCTCATCTTCCAATCCCTACGTGTTCAATGGCGACTTCGTGG
ACAGGGGCAAGCGTGGCTTGGAGGTCCTGCTCCTGCTGCTCTCCCTTTACTTGGCTTTTCCCAACGCAGTT
TTCCTCAACCGCGGGAACCACGAGGACAGTGTCATGAATGCGCGATATGGATTTATCCGGGAGGTGGAAAG
CAAGTATCCCCGGAATCACAAGCGAATCCTGGCCTTCATCGACGAGGTCTACCGATGGCTTCCCCTGGGCT
CCGTACTCAATAGTAGAGTCCTAATCGTGCATGGAGGCTTCTCGGACAGCACCAGCCTGGATCTAATAAAG
AGTATTGACAGGGGAAAGTACGTGTCCATTTTGCGACCGCCACTTACGGATGGCGAACCACTGGACAAAAC

CGAGTGGCAGCAGATCTTCGACATAATGTGGAGTGACCCGCAGGCCACCATGGGATGTGTTCCGAACACTC
TGCGAGGAGCTGGCGTCTGGTTTGGACCGGATGTGACTGACAACTTTCTGCAAAGACATCGACTCAGCTAC
GTCATTCGATCGCACGAGTGCAAGCCCAATGGCCATGAGTTTATGCACGACAACAAGATAATTACTATATT
TTCGGCCTCGAACTACTATGCCATTGGATCCAACAAGGGCGCCTATATCCGGCTGAACAACCAGCTGATGC
CCCATTTCGTTCAGTACATATCGGCGGCATCGCAGACCAAGCGATTATCCTTCAAACAGCGAATGGGCATC
GTGGAGAGTTCGGCACTCAAGGAGTTGGCGGTGCGTATGCGCGATCATCGCGACGAGTTGGAGGACGAGTT
CCGGAAGTATGATCCCAAGGATAGCGGCTATATATCCATATCGCACTGGTGCAAGGTGATGGAGAACGTGA
CCAAGTTGGGGCTGCCCTGGAGACTCCTACGCGATAAGCTGGCACCAGGCACGGATAGTCAAAAGGTTAAC
TATAATAGGACGTTGGATCTGCTGGACACGGATGTTATTGTACTCGAAGCTGAGGCTGATGGCATGTCGGT
AATGGATGCTTTGTATGCCAACAAAGCCAGCCTGGTTGCCATTTTTAATATAATTGATGCCGATAATTCAG
GCGAAATTACCCTAGATGAGTTCGAAACTGCGATTGACCTGCTGGTGGCCCATATGCCAGGTGCCTATTCC
AAGGCGGAAATGCTGGAGAAGTGCAGAATGATGGACCTCAACGGAGATGGCAAGGTGGACTTGAATGAATT
CCTGGAAGCGTTCAGGCTGAGTGACCTGCACAGAAAGGAGCAGCAGGACGAAAACATCAGGAGGCGTTCGA
CTGGCAGACCATCCGTGGCCAAAACCGCCACAGATCCTGTCACTCTCCTGGCCGACAAGATCTCAAAGAAC
ACGCTGGTTGTGGAACACGACATCGATCCCACGGACTGTGAGTCCAAAGTAATCGATCCTAAGAAGTCCTA

>Dsim_Rh3

ATGGAGTCCGCTAACGTGTCGTCGAGTCTATTTGGCAACGTGTCCACCGCGCTGCGGCCGGAGGCGCGGCT
CTCCGCCGAAACGCGTCTGCTGGGCTGGAATGTGCCGCCGGAGGAGCTGCGTCACATTCCCGAGCACTGGT
TAACTTACCCGGAACCGCCCGAATCGATGAACTACCTGCTGGGCACGCTCTACATCTTCTTCACCCTGATG
TCGATGCTGGGCAATGGACTGGTGATTTGGGTTTTCTCGGCAGCCAAATCGCTGCGAACTCCCTCCAATAT
ACTGGTCATCAATCTGGCCTTCTGCGACTTCATGATGATGGTCAAGACTCCGATATTCATCTACAACAGTT
TCCACCAGGGATACGCGCTAGGTCATTTGGGATGCCAGATCTTTGGGATCATTGGCTCCTATACGGGAATC
GCTGCCGGTGCCACCAATGCGTTTATAGCCTACGATCGATTCAATGTGATCACGCGTCCCATGGAGGGCAA
GATGACGCATGGCAAGGCCATTGCCATGATCATATTCATCTACATGTACGCCACTCCATGGGTGGTTGCCT
GCTACACGGAGACTTGGGGACGTTTTGTGCCGGAGGGATATCTGACGTCCTGCACCTTTGACTATCTCACC
GATAACTTCGATACGCGCCTCTTTGTGGGCTGCATCTTCTTTTTAAAAAAGTTGTGTGCCCCACCCCGANN
NTCACGTATTACTATTATCACGATTGTGGCCATGTCTTTAGCCACGAGAAAGCATGGGGATCAGGCCAAGA
AGATGAACGTGGAATCGCTGCGCTCGAATGTGGACAAAAGCAAGGAGACGGCGGAAATCCGGATAGCCAAA
GCGGCTATAACCATCTGCTTCCTGTTCTTTTGCTCGTGGACGCCGTATGGAGTTATGTCGCTGATTGGCGC
CTTTGGGGATAAGACCCTTTTGACGCCCGGAGCCACAATGATTCCCGCCTGTGCCTGCAAAATGGTGGCCT
GCATCGATCCGTTCGTGTACGCCATAAGTCACCCCAGATACCGCATGGAGCTGCAGAAGCGATGTCCCTGG
CTGGCGCTCAACGAAAAGGCGCCGGAATCGTCGGCTGTCGCCTCCACCAGCACCACCCAGGAACCACAGCA
GACCACCGCCGCCTAA

>Dsim_trp

ATGGGCAGCAATACGGAATCCGATGCCGAGAAGGCGTTGGGCTCTCGCCTGGATTACGACCTGATGATGGC
CGAGGAGTACATCCTCAGCGATGTGGAGAAGAATTTCATATTGTCCTGCGAGCGGGGTGACTTGCCAGGTG
TCAAGAAGATCCTCGAGGAGTACCAGGGCACGGACAAGTTCAACATTAACTGCACGGACCCCATGAACCGC
TCCGCCCTCATTTCGGCCATCGAGAACGAGAACTTCGACCTGATGGTGATCCTGCTGGAGCATAACATCGA
GGTGGGCGACGCCCTGCTGCACGCCATCTCGGAAGAGTATGTGGAGGCCGTGGAGGAGCTGCTGCAGTGGG
AGGAGACCAACCACAAGGAGGGCCAGCCATACAGCTGGGAGGCGGTGGACCGCTCCAAGTCCACCTTCACC
GTGGACATCACGCCCCTTATCCTGGCTGCCCACCGCAATAACTACGAGATACTCAAAATCCTCCTGGATCG
CGGGGCCACGCTGCCCATGCCGCACGACGTCAAGTGCGGCTGCGATGAGTGTGTGACCTCCCAGACGACGG
ACTCCCTGCGCCACTCGCAGTCGAGGATCAACGCATACCGCGCCCTGTCCGCCAGCTCGCTGATAGCGCTC
AGCTCCCGGGACCCTGTACTGACCGCCTTCCAGTTGTCCTGGGAACTCAAGCGCCTGCAGGCGATGGAATC
GGAGTTTCGTGCCGAATACACGGAGTGTGCAATGCGTGGCCAGATGGTGCAGGACTTCGTCAGATGGTGCA
GGACTTCGCTCCTGGACCACGCACGCACATCCATGGAGCTCGAGGTGATGCTCAACTTCAACCACGAGCCA
TCCCACGACATCTGGTGCCTTGGCCAGCGGCAAACCCTGGAACGACTGAAGCTGGCCATTCGCTATAAGCA
GAAGACGTTTGTGGCACATCCAAACGTCCAACAATTGTTGGCCGCGATTTGGTACGACGGACTGCCGGGAT
TCCGGCGCAAGCAGGCCTCCCAGCAGCTTATGGACGTCGTGAAGCTGGGATGCAGCTTCCCCATCTACAGC
TTGAAGTACATCCTGGCCCCGGATTCCGAGGGGGCCAAGTTCATGCGCAAGCCCTTTGTCAAGTTCATCAC
GCACTCCTGCTCCTACATGTTCTTCCTGATGCTCCTGGGTGCTGCCTCCCTGAGGGTGGTGCAAATCACCT
TTGAACTCCTTGCCTTTCCCTGGATGCTGACCATGCTGGAGGATTGGCGCAAGCACGAGAGAGGTTCCCTA
CCGGGTCCCATCGAACTGGCAATCATTACCTACATAATGGCTCTGATATTTGAGGAACTGAAATCTTTATA
CTCGGACGGCTTGTTTGAGTACATCATGGATCTTTGGAACATAGTGGACTACATATCGAACATGTTCTATG
TGACGTGGATTCTTTGTAGAGCCACCGCTTGGGTAATCGTCTATATGAATCTAATGAATCGTAATGGATTT
CAGCGCGATCTCTGGTTCCGGGGCATAGATCCTTACTTCCCGAGGGAGCACTGGCATCCGTTCGATCCAAT
GCTTCTATCAGAGGGTGCCTTTGCTGCCGGAATGGTCTTCTCCTATCTGAAGCTCGTCCACATCTTCTCCA

TTAATCCCCACCTGGGACCCTTGCAAGTTTCACTCGGTCGCATGATAATCGACATCATCAAGTTCTTCTTC
ATATACACACTGGTGTTGTTTGCCTTCGGATGTGGTCTCAACCAGTTGCTATGGTACTACGCTGAGCTGGA
GAAGAACAAGTGCTATCACCTGCATCCCGATGTGGCTGACTTTGATGACCAGGAGAAGGCTTGTACCATCT
GGCGAAGATTTTCCAACTTATTCGAAACATCACAATCGCTCTTCTGGGCCTCCTTTGGCCTGGTGGACCTG
GTCTCCTTCGATCTGGCGGGAATCAAGAGCTTCACTCGCTTCTGGGCACTGCTGATGTTCGGCTCCTATTC
GGTTATCAACATCATTGTGCTCCTCAACATGCTGATTGCCATGATGTCCAACTCCTACCAAATCATCTCGG
AGCGAGCCGACACCGAGTGGAAGTTCGCCCGATCCCAGCTGTGGATGAGCTACTTCGAGGATGGCGGCACC
ATTCCACCGCCCTTCAATCTCTGTCCCAACATGAAGATGTTGAGGAAGACGCTGGGGCGAAAGCGACCGTC
ACGAACCAAGAGCTTCATGCGAAAGTCCATGGAACGGGCACAGACGCTGCATGACAAAGTGATGAAGCTGC
TGGTCAGGAGGTACATTACGGCGGAGCAGAGGCGGCGGGACGATTACGGCATTACCGAGGATGATATCATT
GAGGTGCGCCAGGACATCAGCTCCTTGCGGTTCGAGTTGCTGGAGATTTTCACCAACAATAGCTGGGATGT
ACCCGACATTGAGAAGAAGTCGCAGGGAGTTGCTCGAACCACCAAGGGCAAGGTGATGGAACGTCGCATCC
TTAAGGACTTCCAGATTGGCTTCGTCGAGAATCTGAAGCAGGAGATGAGCGAATCTGAGAGCGGACGAGAT
ATATTCTCATCGCTGGCCAAGGTCATCGGCAGAAAGAAGACCCAGAAGGGAGACAAGGATTGGAACGCCAT
TGCGAGGAAGAATACTTTCGCCTCCGATCCCATTGGCTCCAAGCGCTCCTCCATGCAACGTCATAGCCAAC
GAAGCTTGAGGAGGAAGATCATCGAGCAGGCGAATGAGGGTCTCCAGATGAACCAGACCCAGTTGATTGAA
TTCAATCCCAACTTGGGTGATGTGACGCGTGCAACAAGAGTGGCTTATGTCAAGTTCATGCGGAAGAAGAT
GGCTGCCGACGAGGTTTCCTTGGCCGATGATGAGNNNGCTCCAAATGGCGAAGGCGAAAAGAAGCCACTGG
ATGCCTCTGGGTCTCACATGCAGTCGAAAAAGTCCATAACTAGTGGTGGAACTGGAGGAGGGGCTTCTATG
CTGGCTGCAGCTGCTCTAAGAGCATCGGTCAAGAATGTGGATGAAAAATCCGGAGCCAGTGACAAGGATCA
GAAATCCGATGGCAAGCCCGGCACGATGGGCAAGCCAACGGATGACAAGAAGCCAGGTGATGATAAGGATA
AGCAGCAGCCCCCCAAGGACTCCAAGCCGCCAGCAGGTGGTCCCAAGCCCGGGGATCAGAAGCCAACTCCG
GGTGCGGGAGCTCCAAAGCCACAAGCAGCTGGCACTATCAGCAAGCCCGGTGAGCCACAAAAGAAGGACGC
TCCGGCACCACCAACCAAACCTGGAGACACCAAGCCTGCTGCGCCGAAGCCTGGAGAACCCGCCAAGCCCG
AGGCCGCTGCCAAAAAGGAGGAGTCTTCCAAGACCGAAGCTGGCAAGCCGGCAGCCACAAATGGAGCAGCC
AAGAGCGCAGCTCCCTCCGCTCCTTCGGATGCCAAGCCGGACTCCAAACTGAAACCTGGAGCAGCTGGAGC
ACCAGAAGCAACCAAGGCAACCAATGGGGCCTCCAAGCCGGACGACAAGAAGAGCGGTCCGGAGGAGCCCA
AAAAGGCTGCAGGAGACTCCAAGCCAGGAGACGATGCCAAGGACAAGGATAAGAAACCCGGCGACGATAAG
GACAAGAAACCTGGCGACGACAAAGACAAGAAACCTGCCGACAATAATGATAAGAAGCCAGCTGATGACAA
GGACAAGAAGCCGGGTGACGATAAGGACAAGAAGCCGGGTGACGATAAGGACAAGAAGCCGGGTGATGATA
AGGACAAGAAGCCTGCCGATGACAANGACAAGAAGCCAGGAGCAGCTCCTCTGAAGCTGGCGATCAAGGTG
GGTCAGAGCAGTGCCGCAGCTGGTGGAGAACGAGGCAAATCCACGGTCACAGGACGCATGATCTCCGGCTG
GCTCTAA

>Dsim_trpl

ATGGGCCGCAAAAAGAAGCTGCCGACGGGCGTATCGTCTGGCGGCAGCCACGCCTCCTCCGCCCCCAAATC
GGTGGGCGGGTGCTGTGTGCCGCTGGGATTACCACAGCCTCTGCTGCTCGAGGAGAAGAAGTTCCTGCTGG
CTGTCGAGCGCGGCGATATGCCGAATGTGCGCAGGATCCTGCAGAAGGCGCTGCGCCATCAGCACATCAAC
ATCAACTGCATGGACCCCTTGGGTCGTCGGGCCCTCACCCTAGCGATAGACAACGAGAACCTGGAGATGGT
CGAGCTGCTCGTCGTCATGGGGGTGGAGACCAAGGACGCCCTGCTGCACGCCATCAATGCGGAGTTTGTCG
AGGCTGTGGAGCTCCTGCTCGAGCACGAGGAGCTTATCTACAAGGAGGGCGAGCCCTACAGTTGGCAAAAA
GTAGACATAAACACGGCTATGTTTGCGCCGGATATCACGCCTCTGATGCTGGCCGCCCACAAGAACAACTT
TGAGATACTTCGCATTCTGTTGGATCGCGGAGCAGCGGTGCCAGTGCCACATGATATACGTTGCGGTTGTG
AGGAGTGCGTCCGCCTGACCGCGGAGGACTCTCTGAGGCACTCCCTCTCCCGGGTGAACATCTACCGGGCG
CTGTGCAGTCCCTCGCTGATCTGCCTGACCTCCAACGACCCCATCATCACCGCCTTCCAGTTGTCCTGGGA
ACTGAGGAACCTGGCCCTCACCGAGCAGGAGTGCAAGTCAGAGTACATGGACCTGCGAAGGCAGTGCCAGA
AGTTCGCCGTGGATCTGCTGGATCAGACGCGCACCTCCAACGAGCTGGCCATCATCCTGAACTACGATCCG
CAGATGTCGAGCTATGAGCCCGGCGACCGGATGAGCCTGACCCGCCTGGTGCAGGCAATTTCCTACAAGCA
AAAAAAGTTCGTTGCGCACTCGAATATTCAGCAATTACTGTCCTCGATTTGGTATGACGGCCTGCCCGGAT
TCCGGCGCAAGAGCATTGTGGACAAAGTCATCTGCATCGCCCAGGTGGCCGTGCTTTTCCCGCTCTACTGC
CTCATCTACATGTGCGCCCCGAATTGCCGGACGGGCCAGTTGATGCGCAAGCCCTTTATGAAATTCCTGAT
CCATGCCTCCTCCTATTTGTTCTTCCTGTTCATCTTGATTCTGGTATCTCAGCGCGCGGACGATGACTTTG
TACGCATCTTCGGGACGACCAGGATGAAGAAGGAGCTGGCGGAGCAGGAGCTGCGTCAGCGGGGCCAGACA
CCCAGCAAGCTGGAGCTCATCGTGGTGCTATACGTGATTGGGTTCGTCTGGGAGGAGGTGCAGGAGATATT
TGCCGTGGGCATGAAGAGTTATTTGCGCAACATGTGGAACTTCATCGACTTTCTACGCAACAGTCTCTACG
TGAGTGTGATGTGTTTAAGAGCCTTTGCCTACATCCAGCAGGCCACAGAAATCGCAAGGGATCCGCAGATG
GCTTACATACCCCGCGAGAAATGGCACGATTTTGACCCGCAACTAATAGCCGAAGGACTCTTTGCGGCGGC
CAATGTCTTCTCCGCCTTGAAGCTGGTGCACCTGTTCAGCATCAATCCGCACCTGGGTCCGCTGCAGATCT

CCCTGGGTCGCATGGTCATCGACATAGTCAAGTTCTTCTTCATCTACACCCTGGTGCTGTTCGCCTTCGCC
TGTGGACTGAACCAACTGCTCTGGTACTTTGCCGCCCTGGAGAAGAGCAAGTGCTATGTCCTGCCGGGTGG
CGAGGCGGATTGGGGCTCGCACGGCGACTCCTGCATGAAATGGCGCCGGTTTGGCAACCTTTTCGAGTCCT
CGCAGTCGCTCTTTTGGGCCAGCTTTGGGATGGTCGGACTGGACGACTTCGAGCTCAGTGGCATCAAGTCG
TATACGCGATTCTGGGGACTGCTCATGTTCGGCTCGTACAGCGTCATTAACGTGATTGTGCTGCTGAATCT
GCTCATCGCCATGATGTCCAATTCGTATGCCATGATTGATGAGCACTCGGACACGGAGTGGAAGTTTGCCC
GAACCAAGTTGTGGATGAGCTATTTCGAGGACAGTGCCACCCTGCCGCCACCCTTCAATGTCCTGCCCTCC
GTGAAGTGGGTCATCCGGATTTTCCGCAAGTCTAGCAAGACCATCGATCGTCAACGCTCCAAGAAGCGAAA
GGAACAGGAGCAGTTTAGCGAGTATGACAATATAATGCGCTCCCTGGTTTGGCGATATGTGGCCGCCATGC
ATCGCAAGTTCGAGGATAATCCCGTATCGGAGGACGATATCAATGAGGTCAAGAGTGAGATCAATACGATG
CGTTACGAGATGCTCGAGATATTCGAGAACAGCGGAATGGATGTTTCCTCTGCCAACAAAAAGGAGAGGCA
ACCACGACCTCGGCGTATCAAAGTTTGGGAACGGCGGTTAATGAAGGGCTTCCAGGTGGCGCCAGTGCAAA
ATGGCTGTGAGCCCGATGCCTTTGGGAATGTGAACGGGCAGGGAGAAATGCGGGAGATCAAGGTGGAGTCC
ATACCCTCGAAGCCAGCCAAGGAAACCGCCAAGGAGCGGTTTCAACGAGTGGCCCGCACCGTACTCCTTCA
GTCCACTGCCCACAAGTGGAACGTGGTGCTCCGTGCTGCCAAGGATTCTCAAATAGGCCGCTGCACCAAAA
ACGAGCGGAAGAGCCTCCAGAATCTGGGCAGAGCCATCGAAGAGGCCAAGAGGCTTATAATGCTCAATCCA
GGTTGTCCCTCCGGTCGGGAGTCACCGATTCGCATCGAGTTCGAGGATGAGAAGACAAGCACTTTGCTGGA
GCTGCTCAACCAAATCAGTGCTGAGATTAGCGATAGTGAAAAACCTAAAATTAGGCCCACCTGGCGGCCAC
CACTCAAGACAGTGCCTGCCCGGGCAATGGCCGCCAATAACACCAGGTCTCTCACAGCTCCCGAGCTGAAG
ATCAGCAGGAAATCCTCACCAGCTCCCACTCCGACTCCAACTCCGGGAGTTTCCCACACTGCTCTTTCCCA
ATTCCGAAACCGAGAGCTTCCATTGTGTCCCAGCAAGTTAATAGCGAACTCGGCGCCTGCTGCTCCCACTG
CTCCTCCTAAGAAATCAGCACCAACTGCTGCTACACCCACCTATAAGCCCACCACCCACGCCCCCTTCTCC
GTGGAGGGCGGCAACCGGGGAAACACTCGTACCTCAGATGGTGTGCGCTCCGATAATTCGAACTTCGACAT
CCACGTGGTCGATCTGGACGAGAAGGGCGGGCACTTGGGCAGGGATAATGTTTCCGACATTAGCTCCATTG
CCAGCACGAGTCCCCAGCGGCCAAAGCATCGAAACTAA

>Dvir_Galpha_1
ATGGAGTGCTGTTTATCGGAAGAGGCCAAGGAACAAAAGCGGATCAATCAGGAAATCGAGAAACAGCTGCG
GCGCGACAAGCGAGATGCACGCCGTGAACTTAAACTGCTCCTGCTGGGCACGGGCGAGTCCGGCAAATCAA
CGTTCATCAAACAGATGCGTATTATCCATGGCAGCGGGTACTCGGACGATGACAAACGCGGCTACATTAAG
CTGGTGTTCCAAAATATATTCATGGCCATGCAGTCCATGATCAAAGCCATGGATATGCTCAAAATATCATA
TGGTGTTGGCGAGCATAATGCACTCGCCGATCTGGTGATGAGCATTGACTATGAAACAGTCACAACATTTG
AGGATCCATATTTGAATGCCATAAAAACGCTGTGGTCCGATGTTGGCATACAGGAATGCTACGATCGTCGA
AGAGAATATCAGTTGACAGATTCCGCTAAATATTATCTGATGGATCTCGATCGTGTGGCTCAACCTGATTA
TTTACCCACTGAGCAAGACATTTTAAGAGTTCGTGTGCCGACAACGGGGATCATTGAGTATCCCTTTGACT
TAGAAGAGATCAGATTTAGAATGGTGGATGTTGGCGGTCAGCGTTCGGAGCGAAGAAAGTGGATTCATTGC
TTTGAGAATGTTACCTCAATAATATTTTTGGTAGCGCTATCTGAATACGATCAGATCTTGTTTGAATCTGA
TAATGAGAATCGCATGGAGGAATCAAAAGCTTTATTTAGAACTATTATTACATACCCGTGGTTCCAAAATT
CATCAGTTATTCTATTCCTTAACAAGAAAGATTTGTTAGAGGAGAAAATCATGTATTCGCATTTGGTGGAC
TATTTTCCTGAATACGATGGTCCACAGCGAGATGCAATAGCAGCCCGTGAATTCATACTGCGCATGTTTGT
AGACTTAAATCCAGATTCCGAAAAGATTATCTATTCTCATTTCACCTGTGCTACAGATACTGAAAATATAA
GGTTTGTGTTTGCAGCTGTAAAGGATACAATTCTGCAATCCAATCTTAAGGAATACAATTTGGTCTAA
>Dvir_Galpha_2
ATGGAGTGCTTGTTATCGGAAGAGGCCAAGGAGCAAAAGCGAATCAATCAGGAAATCGAGAAACAGCTGCG
GCGCGACAAGCGAGATGCACGCCGTGAACTTAAGCTGCTCCTACTGGGTACTGGTGAATCGGGCAAATCAA
CATTCATCAAACAGATGCGTATTATCCATGGCAGCGGCTACTCGGACGATGACAAACGCGGCTACATCAAA
CTGGTCTACCAAAACATATTCATGGCCATGCAGTCCATGATCAAGGCCATGGACATGCTCAAAATAGAATA
TGGTGGTGCAGAGATTCGTGAATATGCTGAACTGGTGATGAGCATTGACTATGAATCGGTAACAAGGTTCG
AGGATCCATATTTGAATGGCATCAAATGTCTGTGGGCCGATGCTGGCATCCAGGAGTGCTACGATCGCCGA
AGAGAATATCAGTTGACGGATTCAGCCAAATACTATCTGAGCGATTTGGAGCGCATTGAACAAGCTGATTA
TTTGCCCACGGAGCAGGATATTCTGCGTGCTCGTGTGCCCACAACTGGAATACTGGAGTATCCATTCGATC
TGGATGGCATTGTATTCAGAATGGTGGACGTTGGCGGTCAGCGTTCGGAGAGAAGAAAGTGGATTCATTGC
TTTGAGAATGTGACCTCGATTATATTTTTGGTGGCGCTATCTGAATACGATCAGATCTTGTTTGAATCTGA
TAATGAGAATCGCATGGAGGAATCAAAAGCTTTATTTTGCACTATCATTACATATCCGTGGTTCCAAAACT
CATCGATTATTCTCTTTCTAAACAAGAAGGATTTGTTAGAGGAGAAAATCATGATTTCGCATATGGTCGAC
TATTTTCCAGAATTTGATGGTCCAAAACAAGACCATGCAGCGGCAAAACAGTTCGTGTTGAAAAAATACTT

GGCCTGCAATCCCGATCCAGAGCGTCAATGTTATTCGCATTTCACAACAGCCACAGATACCGAAAACATCA
AACTTGTGTTCTGCGCTGTCAAAGATACAATTATGCAAAATGCACTTAAGGAATTCAATTTGGGTTAA

>Dvir_sun

ATGGAGAAGTCGTTTTCGATAACGCCCTGGAAATATGCGCTGTTCACAACATGCATTCTAATTGCGGGCTT
TAATATACTCTTCTTCTCGTGCGGTGTTGTGACCTGGGGCTCTTCGGCTTCGGTCTACGGTGGCTATTCGT
CTGCCCTCTGCGGCGCTCTTATTTTTGCTGTCACTTTCCTGGGACTTTATGTCGCCCTCAAGGAGTCCTAT
AAGTTCTCCAATTATTATATTATTTGCAGCATTTTGGCTAGCATTGTGCTGGTCATTTACTTTATCATCTT
CTCAAGCATGAAAAGCCAGCTAATGTGGCGGTTCGAGGAGCGCATTAAGCGTTTGTTTGCGGAAAAAACAA
ATCAGAATGACACCATGCAGCCCATACACAGCCTGTTCCGCTGCTGTGGCCTGGAGGGGCCGCAGGACTAT
TTGGCCAAGGAGCATGGCGCCTTGCCGGCCAGCTGCTGTTATGCCTTCGATTGCACTGTGTCCAGGCATAT
CTATCAGGAGGGCTGCGGCACAAAGGCCATTCAGGGCTTGCAATTTCAGGCTGATGTGAATTATTACGCAG
CCATTGCCATATTTGTGCTGGAGCTTTTAAGCATGCTCTTCGCATTTCTCATGGGTAAGGCGCGCAAGCTC
TTGAAGATCAAGGACGACGAGACACCCATTAACGAGGATTAG

>Dwil_sun

ATGGAGAAATCATTCTCAATATCTCCCTGGAAATATGCTCTATTTGCCACATGCATTTTGATTGTGGGCTG
TAACATTTTCTTCTTTTCCTGTGGAGTGGTCACCTGGGGGTCAGCATCGTCGGTCTATGGCAGCTATGCAG
TTGCTTTATGCGGTGTCTCTGTATTTGCGGCTGCTTTTCTAGGTGTCTATGTGGCTCTTAAAGAGTCCTAT
AAATACTCAATTTATTATTTACTCACTACCATATTAGTTATGGGATTGCTGACCACATATTTCTTAACATT
CTCGTCGACCAAAGATCAACTGATGACACAATTCGAGGAACGGATTCAAAGTTTGTTCGATGCAAAAACTC
ATCATTTGGATACCATGCAACCCATACACAGTCTATTTCGTTGCTGTGGCCTACAGGGGCCAAAGGATTAC
TTGTCCGAGCAGAATGAGGGGCAGGGAGCATTACCAGCCAGTTGTTGTTATAGTTTTGATTGTACCAATCC
TTCGCATGTATTTGAAGAGGGTTGCTCGACAAAGGCGACAGGAAATTTGAAATTACAAGCGGAAATTAATT
ATTACGCTTGTGTGGCCATTTTTTGCCTCGAGGTCTTGGGTCTACTGACTGCCTTCTTTATGGGCAAGGCC
CGCAAACATGCCAAAACTAAAATCAAGGACGAAGAATCTCCCATCAATAATGATGATTAA

Table A2-1: Percent Amino Acid Sequence Identity
Numbers reflect percent sequence identity at the amino acid level relative to *D. melanogaster*.

	Dsim	Dsec	Dyak	Dere	Dana	Dpse	Dper	Dwil	Dmoj	Dvir	Dgri
arr1	96	96	96	96	95	94	94	94	94	93	93
arr2	94	93	94	96	92	93	93	92	93	91	92
calX	96	96	94	94	85	77	75	79	68	70	72
CdsA	95	95	95	94	89	90	90	88	89	89	87
Galpha49B	80	80	80	80	80	79	79	79	82	84	82
GBeta76C	100	100	100	100	98	94	94	95	97	96	96
Ggamma30A	99	100	100	100	99	99	100	97	100	97	99
inaC	99	99	99	99	94	92	93	93	93	91	92
inaD	87	88	87	87	80	79	75	76	75	76	75
inaE	97	98	98	98	95	95	92	92	92	93	91
inaF	98	98	98	98	87	86	86	85	84	87	89
ninaA	99	99	98	98	91	93	91	91	84	84	84
ninaB	97	97	96	96	90	84	84	83	82	80	82
ninaC	100	94	93	93	90	88	94	89	88	87	87
ninaD	98	98	93	94	86	79	79	79	75	76	76
ninaE	100	100	100	100	99	99	98	98	99	98	98
norpA	95	88	88	93	88	87	86	88	88	88	87
pinta	99	98	99	97	90	85	87	82	80	80	78
Pis	99	99	98	98	96	94	94	89	88	91	92
Pld	92	96	87	90	83	82	74	80	77	77	75
rbo	97	97	97	96	94	91	91	90	90	90	90
rdgB	87	85	89	89	87	85	84	81	85	85	84
rdgC	100	98	97	97	94	86	87	82	80	80	81
Rh2	99	100	98	98	94	92	92	93	91	92	88
Rh3	96	96	96	94	94	90	90	89	87	88	86
Rh4	97	97	96	96	94	94	94	93	90	91	90
Rh5	99	97	98	96	94	91	91	88	86	89	82
Rh6	97	97	97	97	94	93	93	92	88	88	88
santamaria	91	91	92	89	85	77	75	75	74	75	76
stops	90	85	87	89	81	82	84	82	74	76	73
sun	100	100	100	100	85	85	85	83	74	85	81
trp	85	86	85	86	91	80	80	91	92	91	90
trpl	97	89	87	87	82	79	80	76	75	76	73

Tables A2-2
dN, values for all phototransduction genes across the different lineages

Gene	12 Genomes	Sophophora Subgenus	melanogaster subgroup	melanogaster group	Drosophila Subgenus
			dN		
arr1	0.056	0.020	0.002	0.007	0.030
arr2	0.067	0.027	0.002	0.012	0.031
calX	0.904	0.429	0.038	0.196	0.817
CdsA	0.144	0.119	0.005	0.017	0.057
Galpha49B	0.145	0.024	0.026	0.034	0.196
GBeta76C	0.116	0.062	0.004	0.024	0.036
Ggamma30A	0.104	0.023	0.000	0.000	0.000
inaC	0.213	0.126	0.014	0.045	0.070
inaD	0.327	0.219	0.024	0.124	0.187
inaE	0.138	0.097	0.005	0.225	0.049
inaF	1.194	0.315	0.031	0.032	0.061
laza	0.897	0.700	0.063	0.243	0.270
ninaA	0.384	0.177	0.017	0.091	0.123
ninaB	0.410	0.289	0.038	0.101	1.043
ninaC	0.167	0.097	0.011	0.038	0.047
ninaD	0.702	0.702	0.065	0.173	0.216
ninaE	0.070	0.051	0.004	0.011	0.007
norpA	0.051	0.031	0.023	0.027	0.017
pinta	0.500	0.500	0.025	0.095	0.216
pis	0.249	0.139	0.011	0.032	0.105
Pld	0.510	0.330	0.049	0.028	0.788
rbo	0.120	0.080	0.008	0.147	0.036
rdgB	0.191	0.125	0.029	0.045	0.049
rdgC	0.333	0.182	0.007	0.026	0.099
Rh2	0.181	0.096	0.019	0.056	0.117
Rh3	0.245	0.154	0.072	0.084	0.189
Rh4	0.151	0.060	0.005	0.029	0.129
Rh5	0.337	0.193	0.032	0.058	0.882
Rh6	0.141	0.069	0.004	0.026	0.060
santamaria	0.471	0.471	0.051	0.126	0.181
stops	0.547	0.415	0.042	0.192	2.652
sun	0.824	0.499	0.014	0.137	0.235
trp	0.349	0.238	0.054	0.089	0.100
trpl	0.337	0.270	0.028	0.095	0.115

Tables A2-3
dS values for all phototransduction genes across the different lineages

Gene	12 Genomes	Sophophora Subgenus	melanogaster subgroup	melanogaster group	Drosophila Subgenus
			dS		
arr1	4.299	3.689	0.556	1.511	0.844
arr2	4.147	3.367	0.556	1.364	0.945
calX	6.029	3.753	0.390	1.687	1.585
CdsA	4.084	3.590	0.407	1.529	0.715
Galpha49B	2.927	1.485	0.148	0.460	2.200
GBeta76C	3.433	2.370	0.350	1.089	0.873
Ggamma30A	1.547	1.047	0.074	0.225	0.367
inaC	5.653	4.090	0.538	1.755	1.655
inaD	5.135	4.239	0.438	1.577	1.210
inaE	3.939	3.020	0.401	1.245	0.662
inaF	6.102	6.255	0.355	1.653	1.118
laza	7.374	5.644	0.803	2.977	1.256
ninaA	7.275	6.759	0.706	2.575	1.456
ninaB	5.832	4.279	0.546	1.853	1.552
ninaC	5.346	3.733	0.591	1.920	1.187
ninaD	6.693	4.975	0.475	1.785	1.449
ninaE	3.009	2.243	0.324	0.767	0.549
norpA	4.403	3.134	0.437	1.513	0.998
pinta	6.288	4.465	0.479	2.154	1.118
pis	3.892	3.120	0.347	0.919	1.077
Pld	6.320	4.826	0.447	1.984	1.499
rbo	5.684	4.687	0.447	2.247	1.527
rdgB	4.331	3.147	0.416	1.408	1.058
rdgC	6.373	4.873	0.469	1.488	1.440
Rh2	5.241	3.536	0.518	1.798	1.202
Rh3	5.121	3.157	0.504	1.363	1.788
Rh4	5.383	3.776	0.568	1.736	1.297
Rh5	6.303	4.404	0.542	1.713	1.696
Rh6	4.740	3.294	0.429	1.307	1.043
santamaria	5.774	3.804	0.715	1.795	1.348
stops	3.958	2.720	0.371	1.173	0.986
sun	7.193	6.236	0.695	2.198	1.543
trp	4.860	3.496	0.371	1.518	1.201
trpl	4.891	3.706	0.450	1.451	1.117

Tables A2-4: dN/dS values for all phototransduction genes across the different lineages

Gene	12 Genomes	Sophophora Subgenus	melanogaster subgroup	melanogaster group	Drosophila subgenus
			dN/dS		
arr1	0.013	0.006	0.005	0.004	0.036
arr2	0.016	0.008	0.009	0.004	0.030
calX	0.149	0.111	0.086	0.073	0.212
CdsA	0.035	0.031	0.022	0.012	0.041
Galpha49B	0.049	0.016	0.1749	0.066	0.089
GBeta76C	0.034	0.026	0.011	0.012	0.039
Ggamma30A	0.067	0.060	0.000	0.000	0.015
inaC	0.038	0.031	0.026	0.025	0.039
inaD	0.064	0.052	0.063	0.037	0.140
inaE	0.035	0.030	0.026	0.016	0.064
inaF	0.196	0.127	0.136	0.086	0.243
laza	0.122	0.124	0.082	0.079	0.215
ninaA	0.052	0.025	0.033	0.024	0.085
ninaB	0.069	0.068	0.020	0.069	0.081
ninaC	0.032	0.026	0.103	0.019	0.039
ninaD	0.106	0.100	0.136	0.136	0.184
ninaE	0.023	0.023	0.018	0.010	0.013
norpA	0.012	0.010	0.046	0.044	0.017
pinta	0.080	0.060	0.046	0.052	0.181
pis	0.063	0.045	0.044	0.033	0.098
Pld	0.078	0.066	0.073	0.089	0.127
rbo	0.021	0.017	0.018	0.016	0.023
rdgB	0.044	0.039	0.032	0.069	0.046
rdgC	0.052	0.037	0.018	0.016	0.069
Rh2	0.034	0.027	0.030	0.035	0.073
Rh3	0.047	0.048	0.061	0.143	0.045
Rh4	0.027	0.015	0.015	0.008	0.068
Rh5	0.053	0.043	0.034	0.059	0.065
Rh6	0.030	0.021	0.020	0.008	0.057
santamaria	0.081	0.075	0.070	0.071	0.121
stops	0.138	0.139	0.156	0.112	0.201
sun	0.116	0.081	0.063	0.021	0.153
trp	0.072	0.068	0.057	0.051	0.086
trpl	0.069	0.070	0.063	0.061	0.103

Table A2-5: List of genes in each functional cateogory of the phototransduction pathway and their dN values for each lineage

Functional Categories	12 Genomes	Sophophora subgenus	melanogaster subgroup	melanogaster group	Drosophila subgenus
Biosynthesis of Chromophore			**dN**		
ninaB	0.410	0.289	0.038	0.101	1.043
ninaD	0.702	0.493	0.065	0.173	0.303
pinta	0.500	0.266	0.025	0.095	0.216
santamaria	0.471	0.285	0.051	0.126	0.181
Rhodospin cycle					
arr1	0.056	0.020	0.002	0.007	0.030
arr2	0.067	0.027	0.002	0.012	0.031
rdgC	0.333	0.182	0.007	0.026	0.099
ninaE	0.070	0.051	0.004	0.011	0.007
Rh2	0.181	0.096	0.019	0.056	0.117
Rh3	0.245	0.154	0.072	0.084	0.189
Rh4	0.151	0.060	0.005	0.029	0.129
Rh5	0.337	0.193	0.032	0.058	0.882
Rh6	0.141	0.069	0.004	0.026	0.060
sun	0.824	0.499	0.014	0.137	0.235
ninaA	0.384	0.177	0.017	0.091	0.123
Heterotrimeric G-protein					
Galpha	0.145	0.024	0.026	0.034	0.196
GBeta	0.116	0.062	0.004	0.024	0.036
Ggamma	0.027	0.023	0.000	0.000	0.000
PIP_2 regeneration					
norpA	0.051	0.031	0.023	0.027	0.017
CdsA	0.144	0.119	0.005	0.017	0.057
PIS	0.249	0.139	0.011	0.032	0.105
rdgB	0.191	0.125	0.029	0.045	0.049
laza	0.897	0.700	0.063	0.243	0.270
PLD	0.510	0.330	0.049	0.028	0.788
rbo	0.120	0.080	0.008	0.147	0.036
stops	0.547	0.415	0.042	0.192	2.652
inaE	0.138	0.097	0.005	0.225	0.049
Ion channel associated					
trp	0.349	0.238	0.054	0.089	0.100
trpl	0.337	0.270	0.028	0.095	0.115
calx	0.904	0.429	0.038	0.196	0.817
inaF	0.025	0.315	0.031	0.032	0.061
inaD signalling complex					
ninaC	0.167	0.097	0.011	0.038	0.047
inaD	0.327	0.219	0.024	0.124	0.187
inaC	0.213	0.126	0.014	0.045	0.070

BIBLIOGRAPHY

Adams MD, Celniker SE, Holt RA, Evans CA, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 5461:2185-95.

Altschmied J, Delfgaauw J, Wilde B, Duschl J, Bouneau L, et al. 2002. Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. *Genetics* 161: 259–267.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.

Alvarez, CE, Robison K, and Gilbert W. 1996. Novel Gq alpha isoform is a candidate transducer of rhodopsin signaling in a Drosophila testes-autonomous pacemaker. *Proc. Natl. Acad. Sci. USA* 93:12278–12282

Alvarez-Ponce D, Aguadé M, Rozas J. 2009. Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 *Drosophila* genomes. *Genome Res* 19:234-42.

Alvarez-Ponce D, Aguadé M, Rozas J. 2010. Comparative genomics of the vertebrate insulin/TOR signal transduction pathway: a network-level analysis of selective pressures. *Genome Biol Evol* 3:87-101.

Arbeitman M, Furlong E, Imam F, Johnson E, Null B, Baker B, Krasnow M, Scott M, Davis R., White K. 2002. Gene expression during the life cycle of *Drosophila* melanogaster. Science 298: 2270–2275.

Artieri, CG, Singh RS. 2010. Molecular evidence for increased regulatory conservation during metamorphosis, and against deleterious cascading effects of hybrid breakdown in *Drosophila*. BMC Biology. 8:26.

Assis R, Zhou Q, Bachtrog D.2012. Sex-biased transcriptome evolution in *Drosophila*. *Gen Biol Evol* 4: 1189–1200.

Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH, Oshlack A, Smith CA. 2013. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. Genome Biol. 14:R26.

Bahner M, Frechter S, Da Silva N, Minke B, Paulsen R, Huber A. 2002. Light-regulated subcellular translocation of Drosophila TRPL channels induces long-term adaptation and modifies the light-induced current. *Neuron* 34: 83-93.

Barbosa-Morais, NL et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593.

Ben-Tabou de-Leon S, Davidson EH. 2007. Gene regulation: gene control network in development. Annu Rev Biophys Biomol Struct. 36: 191–212.

Bhardwaj N, Lu H. 2005. Correlation between gene expression profiles and protein-protein interactions within and across genomes. *Bioinformatics* 21:2730-2738

Benjamini Y, Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57, 289–300.

Bodenstein D. 1950. *The postembryonic development of Drosophila.* In: Demerec M, editor. Biology of Drosophila. New York: John Wiley and Sons. p. 275-367

Bray, S. and Amrein, H. 2003. A putative Drosophila pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* 39:1019-1029.

Brett D, Pospisil H, Valcarel J, Reich J, Bork P. 2002. Alternative splicing and genome complexity. *Nat Genet* 30: 29–30

Brush AH. 1970. Pigments in hybrid, variant and melanistic tanagers (birds). *Comp. Biochem. Physiol.* 36:785–93

Calarco JA, Xing Y, Caceres M, Calarco JP, Xiao X, Pan Q, Lee C, Preuss TM, Blencowe BJ. 2007. Global analysis of alternative splicing differences between humans and chimpanzees. *Genes & Dev* 21:2963–2975

Cannon S, Mitra A, Baumgarten A, Young N, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biol*. 4:10.

Carvalho AB, Clark AG. 2005. Y chromosome of *D. pseudoobscura* is not homologous to the ancestral *Drosophila* Y. Science;307:108-110

Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A: GENECODIS: A web-based tool for finding significant concurrent annotations in gene lists. *Genome Biology 2007 81:R3*

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol* 3:e245

Carroll SB. 2008. Evo-Devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* 134:25–36.

Carulli JP, Hartl DL. 1992. Variable rates of evolution among *Drosophila* opsin genes. *Genetics* 132:193-204.

Chippindale AK, Gibson JR, Rice WR. 2001. Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in Drosophila. *Proc Natl Acad Sci USA* 98:1671-1675.

Clark AG, Eisen, MB, Smith, DR, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-128.

Clyne, PJ., Warr, CG., Carlson, JR. 2000. Candidate taste receptors in *Drosophila*. *Science* 287:1830–1834

Conklin, BR., & Bourne, HR. 1993. Structural elements of Gα subunits that interact with Gβγ, receptors, and effectors. *Cell*. *73*4, 631-641.

Connallon T, Knowles LL. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends in Genetics* 21:495-499.

Cork JM, Purugganan MD. 2004. The evolution of molecular genetic pathways and networks. *BioEssays* 26:479-484.

Cox RM, Calsbeek R. 2009 Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *Am Nat* 173:176–187

Cui Q, Purisima EO, Wang E. 2009. Protein evolution on a human signaling network. *BMC Syst Biol* 3:21.

Cuervo JJ, Møller AP. 1999. Ecology and evolution of extravagant feather ornaments. *J. Evol. Biol.* 12:986–98

Daines B, Wang H, Wang L, et al. 2011. The *Drosophila melanogaster* transcriptome by paired-end RNA sequencing. *Genome Res*. 21:315–324.

Davidson, EH, Erwin, DH. 2006. Gene regulatory networks and the evolution of animal body plans. *Science* 311, 796–800

Dobzhansky, T, and Epling T. 1944. *Contributions to the Genetics, Taxonomy, and Ecology of Drosophila pseudoobscura and Its Relatives.*Carnegie Institute of Washington, Washington, DC.

Dobzhansky, T. 1973. Is there gene exchange between *Drosophila pseudoobscura* and *Drosophila persimilis* in their natural habitats? *Am. Nat*. 107 312–314.

Elemento, O, Gascuel, O, Lefranc, MP. 2002. Reconstructing the duplication history of tandemly repeated genes. *Mol Biol Evol*. 19:278-288.

Ellegren, H, Parsch, J.2007. The evolution of sex-biased genes and sex-biased gene expression. Nature Reviews Genetics 8: 689-698.

Elsik CG; Mackey AJ; Reese JT; Milshina N; Roos DS. 2007. Creating a honey bee consensus gene set. *Genome Biol.* 8: R13.

Fan C, Chen Y and Long M 2008. Recurrent tandem gene duplication gave rise to functionally divergent genes in *Drosophila*. *Molecular Biology and Evolution* 25:1451-1458.

Figuerola J, AJ. 2000. The evolution of sexual dimorphism in relation to mating patterns, cavity nesting, insularity and sympatry in the Anseriformes. *Funct. Ecol.* 14:701–10

Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, Eanes WF. 2007. Adaptive evolution of metabolic pathways in *Drosophila*. *Mol. Biol. Evol*. 24:1347–1354.

Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD. 2009. Population genomics of the Arabidopsis thaliana flowering time gene network. *Mol. Biol. Evol*. 26:2475-2486.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531-45.

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. *Science* 296:750-752.

Grath S, Baines JF, Parsch J. 2009. Molecular evolution of sex-biased genes in the Drosophila ananassae subgroup. *BMC Evol Biol*.;9:219.

Gilman, AG. 1987. G Proteins: transducers of receptor generated signals. *Annu. Bev. Biochem* 56:615-649.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.*11:725-36.

Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2014. Selection-Driven Evolution of Sex-Biased Genes Is Consistent with Sexual Selection in Arabidopsis thaliana. Mol. Biol. Evol. 31:574–583.

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM,

Wildonger J, O'Connor-Giles KM. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194:1029–1035.

Graveley BR, Brooks AN, Carlson JW, et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471:473-479.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, Singh RS. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177:1321-1335.

Hahn MW, Kern AD. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol*. 22:803-806.

Hahn, MW, Han, M.V., Han, SG 2007 Gene family evolution across 12 Drosophila genomes. *PLoS Genet*. 3:e197.

Hale MC, Xu P, Scardina J, Wheeler PA, Thorgaard GH, Nichols KM. 2011. Differential gene expression in male and female rainbow trout embryos prior to the onset of gross morphological differentiation of the gonads. BMC Genomics 12:404.

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH. 2008. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol*. 148:993-1003

Hardie RC, Raghu P. 2001. Visual transduction in *Drosophila*. *Nature*. 413:186-93.

Hatje K, Kollmar M 2013 Expansion of the mutually exclusive spliced exome in Drosophila. *Nat Commun* 4: 2460.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169:1157-1164

Hughes, TR, Mao, M,Jones, AR. Burchard, J, Marton, MJ, *et al.* 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol*. 19 342–347.

Hull JJ, Lee JM, Matsumoto S. 2010. Gqalpha-linked phospholipase Cbeta1 and phospholipase Cgamma are essential components of the pheromone biosynthesis activating neuropeptide PBAN signal transduction cascade. *Insect Molecular Biology* 19:553–566.

Huminiecki, L, and Wolfe, KH 2004. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. *Genome Res*. 14:1870–1879.

Ihaka R, Gentleman R. 1996. R: A Language for Data Analysis and Graphics. *J Comput Graph Stat*. 5:299.

Isono K, Ueno, K, Ohta, M, Morita, H. 2002. *Pure Appl. Chem*. Vol. 74, 7:1159–1165

Jacquemin J, Ammiraju JSS, Haberer G, et al. 2013. Fifteen Million Years of Evolution in the Oryza Genus Shows Extensive Gene Family Expansion. Mol. Plant.. doi: 10.1093/mp/sst149

Jiang Z-F, Machado CA. 2009. Evolution of sex-dependent gene expression in three recently diverged species of Drosophila. *Genetics* 183:1175–1185.

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. *Nat. Genet*. 29:389–395.

Joseph RM, Devineni AV, King IF., Heberlein U. 2009. Oviposition preference for and positional avoidance of acetic acid provide a model for competing behavioral drives in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 10627: 11352–11357.

Jovelin R, Phillips PC. 2009. Evolutionary rates and centrality in the yeast gene regulatory network. *Genome Biol* 10:R35.

Jovelin R, Phillips, PC. 2011. Expression level drives the pattern of selective constraints along the Insulin/Tor signal transduction pathway in *Caenorhabditis*. *Genome Biol. Evol*. 3:715–722.

Kain P, Chakraborty TS, Sundaram S, Siddiqi O, Rodrigues V, et al. 2008. Reduced odor responses from antennal neurons of Gqalpha, phospholipase Cbeta, and *rdgA* mutants in *Drosophila* support a role for a phospholipid intermediate in insect olfactory transduction. *J Neurosci* 28: 4745–4755.

Kalidas S, Smith DP. 2002. Novel genomic cDNA hybrids produce effective RNA interference in adult *Drosophila*. *Neuron* 33: 177–184.

Katanayeva, N, Kopein, D, Portmann, R, Hess D. and Katanaev VL. 2010. Competing activities of heterotrimeric G proteins in Drosophila wing maturation. *PLoS ONE* 5, e12331.

Kent WJ. 2002. BLAT— the BLAST-like alignment tool. *Genome Res* 12:656-664.

Keren H, Lev-Maor G, Ast G 2010. Alternative splicing and evolution: Diversification, exon definition and function. *Nat Rev Genet* 11: 345–355

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188: 107–116

Kim PM, Korbel JO, Gerstein MB. 2007. Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. *Proc Natl Acad Sci USA*. 104:20274-20279.

Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. 2005. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. Genetics 171:1813–22.

Landry CR, Castillo-Davis CI, Ogura A, Liu JS, Hartl Daniel L. 2007. Systems-level analysis and evolution of the phototransduction pathway in *Drosophila*. *Proc Natl Acad Sci USA*. 104:3283-32

Larkin MA, Blackshields G, Brown NP, et al.; 13 co-authors. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.

Lattao R, Bonaccorsi S, Gatti M. 2012. Giant meiotic spindles in males from *Drosophila* species with giant sperm tails. *J Cell Sci* 125: 584–588.

Lee YJ, Dobbs MB, Verardi ML, Hyde DR. 1990. dgq: a *Drosophila* gene encoding a visual system-specific G alpha molecule. *Neuron* 5:889-898.

Lee YJ, Shah S, Suzuki E, Zars T, O'Day PM, Hyde DR. 1994. The *Drosophila* dgq gene encodes a Gα protein that mediates phototransduction. *Neuron* 13:1143–1157.

Lemos B, Meiklejohn CD, Hartl DL. 2004. Regulatory evolution across the protein interaction network. *Nat Genet*. 3610: 1059-1060.

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl, DL. 2005. Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Mol. Biol. Evol*. 22:1345–1354.

Lister, J. A., J. Close, and D. W. Raible. 2001. Duplicate mitf genes in zebrafish: complementary expression and conservation of melanogenic potential. *Dev. Biol*. 237:333–344.

Livingstone K, Anderson S. 2009. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. *J. Hered* 100:754-61.

Lu Y, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Mol Biol Evol*. 20:1844-1853.

Lynch M, Connery JS 2000 The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.

Ma P, Castillo-Davis CI, Zhong W, Liu J. 2006. A data-driven clustering method for time course gene expression data. *Nucleic Acids Res*. 34: 1261-9.

Machado, CA, Kliman, RM, Markert, JA, Hey, J. 2002. Inferring the history of speciation using multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* 19: 472–488.

Machado, CA and Hey, J. 2003. The causes of phylogenetic conflict in a classic Drosophila species group. *Proc. R. Soc. Lond. B* 270: 1193–1202.

Madan Babu M, Luscombe NM, Aravind L, Gerstein M ,Teichmann SA 2004. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* 14: 283–29

Magnusson K, Mendes AM, Windbichler N, Papathanos P-A, Nolan T, Dottorini T, Rizzi E, Christophides GK, Crisanti A. 2011. Transcription regulation of sex-biased genes during ontogeny in the malaria vector Anopheles gambiae. *PLoS One*, 6:e21572.

Mank JE, Nam K, Brunström B, Ellegren H. 2010. Ontogenetic complexity of sexual dimorphism and sex-specific selection. *Mol. Biol. Evol*. 27:1570–1578.

Mank JE, Hultin-Rosenberg L, Axelsson E, Ellegren H. 2007. Rapid evolution of female-biased, but not male-biased, genes expressed in the avian brain. *Mol Biol Evol.* 24:2698–2706.

Mandelkow, E and Mandelkow, EM. 1995. Microtubules and microtubule-associated proteins. *Curr. Opin. Cell Biol*. 7, 72-81.

Marinotti O, Calvo E, Nguyen QK, Dissanayake S, Ribeiro JM, et al. 2006 Genome-wide analysis of gene expression in adult Anopheles gambiae. Insect Mol Biol 15: 1–12.

Markow TA, O'Grady PM. 2007.*Drosophila* biology in the genomic age. *Genetics* 177: 1269–1276.

Marshall AN, Montealegre MC, Jiménez-López C, Lorenz MC, van Hoof A. 2013. Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes. *PLoS Genet* 9(3): e1003376.

Martins MJF, Mota CF, Pearson GA. 2013. Sex-biased gene expression in the brown alga Fucus vesiculosus. BMC Genomics 14:294.

Megy, K. et al. 2012. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Research* 40: D729-734.

Meisel RP, Malone JH, Clark AG. 2012.Faster-X Evolution of Gene Expression in *Drosophila*. *PLoS Genet* 810: e1003013

Metta M, Gudavalli R, Gibert JM, Schlötterer C. 2006. No accelerated rate of protein evolution in male-biased Drosophila pseudoobscura genes. Genetics.174:411-420.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci USA* 100:9894-9899

Miura N, Atsumi S, Tabunoki H, Sato R. 2005. Expression and localization of three G protein α subunits, Go, Gq, and Gs, in adult antennae of the silkmoth Bombyx mori. *J Comp Neurol* 485:143-152.

modENCODE Consortium—Roy S, Ernst J, Kharchenko PV, Kheradpour P, Nègre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, et al. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. Science 2010;330:1787-1797.

Montanucci L, Laayouni H, Dall'Olio GM, Bertranpetit J. 2011. Molecular evolution and network-level analysis of the N-glycosylation metabolic pathway across primates. *Mol Biol Evol*. 28:813-23.

Montell C. 1999. Visual transduction in *Drosophila*. *Annu Rev Cell Dev Bio.*15:231 68.

Naurin S, Hansson B, Hasselquist D, Kim YH, Bensch S 2011 The sex-biased brain: sexual dimorphism in gene expression in two species of songbirds. BMC Genomics 12: 37

Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol*. 48:443-53.

Newby BN. and Etges WJ. 1998. Host preference among populations of *Drosophila mojavensis* that use different host cacti. *Journal of Insect Behavior* 11:691-712.

Nogales-Cadenas R, Carmona-Saez P, Vazquez M, Vicente C, Yang X, Tirado F, Carazo JM, Pascual-Montano A. *2009.* GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Research; doi: 10.1093/nar/gkp416*

Ohno, S.. Evolution by gene duplication. 1970. Springer-Verlag.

Ohta T.. Evolution of gene families. 2000Gene 259:45–52.

Pak WL, Grossfield J, White NV. 1969. Nonphototactic mutants in a study of vision of *Drosophila*. *Nature* 222:351-4.

Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, et al. 2003. Paucity of genes on the Drosophila X chromosome showing male-biased expression. *Science* 299: 697–700.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat. Rev. Genet.* 14:83–87.

Pavey S, Collin H, Nosil P, Rogers SM. 2010. The role of gene expression in ecological speciation. Ann. N. Y. Acad. Sci. 1206:110–129.

Penalva L, Sanchez L. 2003. RNA binding protein sex-lethal Sxl and control of Drosophila sex determination and dosage compensation. *Microbiol. Mol. Biol. Rev.* 67: 343–59

Perry JC, Harrison PW, Mank JE. 2014. The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. *Mol. Biol. Evol.*

Pohl M, Bortfeldt RH, Grützmann K, Schuster S 2013 Alternative splicing of mutually exclusive exons–a review. *BioSystems* 114: 31–38

Pointer MA, Harrison PW, Wright AE, Mank JE 2013 Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. *PLoS Genet* 98: e1003697

Powell, JR 1983 Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from Drosophila. *Proc. Natl. Acad. Sci. USA* 80: 492–495

Ramsay H, Rieseberg LH, Ritland K. 2009. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. *Mol Biol Evol*. 26:1045-53.

Ranganathan R, Harris WA, Zuker CS. 1991. The molecular genetics of invertebrate phototransduction. *Trends Neuro* 14:486-93.

Ranz JM, Castillo-Davis CI, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300:1742-5.

Ratnaparkhi A, Banerjee S, Hasan G. 2002. Altered levels of Gq activity modulate axonal pathfinding in *Drosophila*. *J Neurosci* 22:4499-4508.

Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol*. 16:266-74.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 16:276-7.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:735-742.

Riley RM, Jin W, Gibson G. 2003. Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in *Drosophila*. *Mol. Ecol*. 12:1315-23.

Richards S., Liu Y., Bettencourt B.R., Hradecky P., Letovsky S., Nielsen R., Thornton K., Hubisz M.J., Chen R., Meisel R.P., et al. 2005. Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. Genome Res.15:1-18

Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* 13:505–516.

Rohmann KN, Deitcher DL, Bass AH. 2009. Calcium-activated potassium (BK) channels are encoded by duplicate slo1 genes in teleost fishes. *Mol Biol Evol* 26:1509–1521.

Rubin GM. 2000. Comparative Genomics of the Eukaryotes. *Science* 287:2204–2215.

Sakurai T, Nakagawa T, Mitsuno H, Mori H, Endo Y, et al. 2004 Identification and functional characterization of a sex pheromone receptor in the silkmoth *Bombyx mori*. *Proc Natl Acad Sci U S A* 101: 16653–16658.

Schadt EE, Monks SA, Drake TA, et al. 2003. Genetics of gene expression surveyed in maize , mouse and man. 205:1–6.

Schaeffer, SW., Miller, EL. 1991. Nucleotide sequence analysis of Adh genes estimates the time of geographic isolation of the Bogota population of Drosophila pseudoobscura. Proc. Natl. Acad. Sci. U.S.A. 88(14): 6097-6101.

Scott K, Becker A, Sun Y, Hardy R, Zuker C. 1995. *Gqα* protein function in vivo: genetic dissection of its role in photoreceptor cell physiology. *Neuron* 15:919–927.

Shoja V, Zhang L. 2006. A roadmap of tandemly arrayed genes in the genomes of human, mouse, and rat. *Mol Biol Evol*. 23:2134–2141.

Singh, ND, Larracuente, AM, Sackton, TB, Clark, AG. 2009. Comparative genomics on the Drosophila phylogenetic tree. *Annu. Rev. Ecol. Evol. Syst*. 4010:459-480.

Smith G, Chen Y-R, Blissard GW, Briscoe AD. 2014. Complete Dosage Compensation and Sex-Biased Gene Expression in the Moth Manduca sexta. Genome Biol. Evol. 6:526–537.

Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat. Appl. Genet. Mol. Biol. 3.

Sprang SR. 1997. G protein mechanisms: insights from structural analysis. *Annu Rev Biochem* 66:639–678

Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol*. 201:18-20.

Swarup S, Morozova T V, Sridhar S, Nokes M, Anholt RRH. 2014. Modulation of feeding behavior by odorant-binding proteins in *Drosophila melanogaster*. Chem. Senses [Internet] 39:125–132.

Steinemann M, Pinsker W, Sperlich D. Chromosome homologies within the *Drosophila* obscura group probed by in situ hybridization. Chromosoma 1984;91:46-53.

Stuglik MT, Babik W, Prokop Z, Radwan J. 2014. Alternative reproductive tactics and sex-biased gene expression: the study of the bulb mite transcriptome. Ecol. Evol. 4:623–632.

Sturgill D, Zhang Y, Parisi M, Oliver B. Demasculinization of X chromosomes in the *Drosophila* genus. Nature. 2007;450:238–241.

Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540–1542.

Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A. 2012. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Research.* 40:W478-W483

Talluri S, Bhatt A, Smith DP 1995 Identification of a *Drosophila* G protein alpha subunit dGq alpha-3 expressed in chemosensory cells and central neurons. *Proc Natl Acad Sci USA* 92: 11475–11479.

Teichmann SA, Babu MM. 2004. Gene regulatory network growth by duplication. *Nature Gene* 36:492-6.

Trabzuni D, Ramasamy A, Imran S, Walker R, Smith C, Weale ME, Hardy J, Ryten M. 2013. Widespread sex differences in gene expression and splicing in the adult human brain. Nat. Commun. 4:2771.

Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc. Natl. Acad. Sci. U. S. A. 110:6453–6458.

Walsh JB. 1995. How often do duplicated genes evolve new functions? *Genetics* 139:421–428.

Wang, RL, Wakeley, J, Hey, J. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics*, 147: 1091–1106.

Wang J, Xia Q, He X, Dai M, Ruan J, Chen J, *et al.* 2005. SilkDB: a knowledgebase for silkworm biology and genomics. *Nucleic Acids Res* 33:D399-D402.

Wang T, Montell, C. 2007. Phototransduction and retinal degeneration in *Drosophila*. *Pflugers Arch* 454:821-47.

Wang, T, Jiao, Y, and Montell, C. 2007. Dissection of the pathway required for generation of vitamin A and for *Drosophila* phototransduction. *J. Cell Biol*. 177:305–316.

Wang, ET, Sandberg, R, Luo, S, Khrebtukova, I, Zhang, L, Mayr, C, Kingsmore, SF, Schroth, GP, and Burge, CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456, 7221, 470–6.

Weinstock GM, Robinson GE, Gibbs RA, et al. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443: 931–949.

West-Eberhart MJ 2005. Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences, USA Supplement*, 1 102, 6543–6549

White, K.P. 2001. Functional genomics and the study of development, variation and evolution. *Nat. Rev. Genet.* 2: 528–537.

Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: raw material for evolution. *Mol. Ecol.* 15:1197–1211.

Williams TM, Carroll SB 2009. Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nat Rev Genet* 10: 797–804.

Wu X, Chi X, Wang P, Zheng D, Ding R, Li Y. 2010. The evolutionary rate variation among genes of HOG-signaling pathway in yeast genomes. *Biol Direct* 5:46.

Yamada T, Bork P. 2009. Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nature reviews. Mol. Cell Biol*. 10:791-803.

Yang, J, Su, A. I., and. Li, W.-H. 2005. Gene expression evolves faster in narrowly than in broadly expressed mammalian genes. *Mol. Biol. Evol*. 22:2113–2118

Yang X, Schadt EE, Wang S, Wang H, Arnold AP, et al. 2006 Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res 16(8): 995–1004.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586-91.

Yu HS, Shen YH, Yuan GX, Hu YG, Xu HE, Xiang ZH, Zhang Z. 2011. Evidence of selection at melanin synthesis pathway loci during silkworm domestication. *Mol. Biol. Evol*. 286:1785–1799.

Yu, W. P., S. Brenner, and B. Venkatesh. 2003. Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in Fugu. *Trends Genet*. 19:180-183.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus Drosophila. *Nature* 450:233-238

Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in Drosophila. *Mol Biol Evol.* 21:2130–2139.

Zhao M, Zha X-F, Liu J, Zhang W-J, He N-J, Cheng D-J, Dai Y, Xiang Z-H, Xia Q-Y. 2011. Global expression profile of silkworm genes from larval to pupal stages: toward a comprehensive understanding of sexual differences. *Insect Sci.*;18:607-618.

Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, Li X, Ding Y, Yang S, Wang W. 2008. On the origin of new genes in Drosophila. *Genome Res* 18:1446–1455.

Zuker CS. 1992. Phototransduction in *Drosophila*: a paradigm for the genetic dissection of sensory transduction cascades. *Curr Opin Neurobiol*. 2:622-7.

Zuker CS. 1996. The biology of vision of *Drosophila*. *Proc Natl Acad Sci USA*. 93:571-6.