ACCELERATED EVOLUTION OF TISSUE-SPECIFIC MODULATION OF GENE EXPRESSION IN SIBLING SPECIES DROSOPHILA SIMULANS AND D. SECHELLIA

ACCELERATED EVOLUTION OF TISSUE-SPECIFIC MODULATION OF GENE EXPRESSION IN SIBLING SPECIES DROSOPHILA SIMULANS AND D. SECHELLIA

By

MADELINE LOOMER, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Masters of Science

McMaster University

MASTERS OF SCIENCE (2011) MCMASTER UNIVERSITY
(Biology) Hamilton, Ontario

TITLE: Accelerated evolution of tissue-specific modulation of gene expression in sibling species *Drosophila simulans* and *D. sechellia*

AUTHOR: Madeline Loomer, B.Sc. (York University)

SUPERVISOR: Professor Rama S. Singh

NUMBER OF PAGES: [xii], 111

ABSTRACT

Comparison of tissue-specific gene expression profiles between sibling species and their F1 interspecific hybrids are expected to reveal important information about the mechanisms involved in speciation and species divergence. In the present study, the expression of 40 candidate genes were analyzed using qRT-PCR in the testis, ovary and head tissues (both male and female) among the hybrids and their parental species. The expression patterns of these genes were profiled via quantification of misexpression (both over- and under-expression) relative to that of their parents as well as across tissues. We set out to answer several questions as well as test the following hypotheses: (1) Do *Drosophila* sibling species differ in tissue-specific distribution of gene expression? (2) Do males and females differ in tissue specific distribution of expression? (3) Do reciprocal crosses differ in patterns of gene misexpression suggesting X-effect? (4) Do sex and non-sex genes differ in extent of gene misexpression? The results of this study gave rise to two important findings. First, it was found that while the majority of genes showed head and testis expression in *Drosophila melanogaster*, more genes showed head and ovary expression in *D. sechellia*. Second, we observed differences in gene misexpression between reciprocal *D. simulans* and *D. sechellia* hybrid females, suggesting the role of maternal effect. Thus, these findings supply a wealth of data regarding tissue-specific expression in both fertile females and sterile male hybrids, the former of which have largely been ignored, as well as advance our understanding of the process of species divergence and speciation. Ultimately, this thesis will provide a contribution to the field of gene regulatory evolution.

ACKNOWLEDGEMENTS

I first would like to thank my supervisor, Dr. Rama Singh, for his direction and support throughout the duration of my Masters. Dr. Singh's guidance during lab meetings and, always appreciated, trips to the Faculty Club served as encouraging environments to discuss ideas and work through the current week's tribulations. I will always be eternally grateful for the opportunity and for one of the most important learning experiences of my educational career.

An equal amount of thanks are owned to Dr. Carlo Artieri and Dr. Wilfried Haerty. It is not possible to truly express how indebted I am to both these researchers. Their steadfast guidance and endless patience has been invaluable and will forever be appreciated. I would also like to thank Dr. Richard Morton for his insightful discussions and challenging questions, making my experience all the more rewarding.

I would like to express thanks to the members of the faculty of Biology who made up my supervisory committee, Drs. Ben Evans, and Bhagwati Gupta, for their insightful critiques and helpful advice.

Thanks must also go to the members of the Singh lab. To Abha Ahuja, thank you for all of the talks, both work-related and personal. Thanks to Abhishek Pai and Scott Devito for the implicit aid and, most importantly, the enjoyable working environment. I will miss all of you very much.

Appreciation must also be expressed to my peers in graduate studies, Andrea Morash, Iqbal Setiadi, Derek Alsop, Adam Bewick, and Emily Smith. Thank you for all of your insight and assistance. On the technical side, I must thank Dr. Jonathon Stone for allowing me access to his lab and his imaging microscope. As well, I must thank Dr. Grant McClelland, and again his student Andrea Morash, for allowing me to use their qPCR equipment, in addition to instructing me in its proper use. I must also thank Emily Smith and Luke Bowley from Joanna Wilson's lab.

Special thanks to my family for their love and support, and to my husband Avi, thank you for the unwavering love and encouragement. You inspire me every day to reach my goals and to be the best version of myself.

I would like to dedicate my M.Sc. thesis to my Zaidys Cyril Loomer and Ben Parker, my Poppa Norman Fox as well as my Bubby Lorraine Loomer. Although they have passed, they continue to inspire me, whether it be in study or the way in which they treated others and lived their lives. I hope I have made you all proud and will continue to strive to do so throughout my life.

TABLE OF CONTENTS

LIST OF FIGURES

Supplementary Figures

LIST OF TABLES

Supplementary Tables

INTRODUCTION

"No one definition (of species) has as yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species. Generally the term includes the unknown element of a distinct act of creation. The term "variety" is almost equally difficult to define; but here community of descent is almost universally implied, though it can rarely be proved."

Charles Darwin (1859), On the Origin of Species

1.1 On the Origin of Species

1.1.1 Diversity of Life and the Development of an Evolutionary Theory

The diversity of life has endured as one of mankind's most intriguing and investigated mysteries. Charles Darwin summarized it best when he referred to the origin of species as "...that mystery of mysteries..." (Darwin 1859). This vast diversity of life had long been recognized amongst plants and animals through a typological concept of species dating back to the ancient Greeks (Mayr 1970); the recognition that organisms are similar within groups and distinctively different from other groups (Singh 2000). The foundation of naturalistic thinking on biology was conceptualized by the Greek philosopher Anaximander (611-547 B.C.), where he introduced the notion that all living things were related and could change over time (Wright 1984; Gould 1989). Another well known Greek philosopher, Aristotle, laid the foundation for a natural classification deemed Scala Naturae, translated to mean Ladder of Life, which was developed to elucidate the visible diversity of life. Aristotle also proposed the theory of spontaneous generation, which suggests that life emerges from non-living matter (Wiener 1973; Kutschera and Niklas 2004). This theory of spontaneous generation remained prominent and there was little advancements made towards the currently accepted ideologies of evolution for what would be close to two millennia (Wright 1984). This was thought to, at least in part, be the result of Biblical literalism, that the Bible was without error and its contents where entirely factual. This, in turn, placed a

great deal of constraints on "scientists" throughout the Middle Ages who countered the more popular natural theology of the time, which supported the concept of species as static and fixed entities. In order to progress to the now widely accepted theory of evolution, the notion of spontaneous generation needed to be disproven (Wright 1984; Levine and Evers 1999).

The first documented assault on spontaneous generation was conducted in 1668 (1680) through an experiment conducted by an Italian physician, Francesco Redi (1626-1697). He intended to prove that the prevailing belief that maggots arose from rotting meat was false, and that they in fact were produced by flies lying eggs on said meat. While this was a good first experiment in the pursuit to disprove spontaneous generation, the theory remained strong (Levine and Evers, 1999). The dispute of this theory continued throughout the eighteenth century, however it was not until the nineteenth century that Louis Pasteur (1822-1895), through his work with bacteria, was the scientific world convinced of the mendaciousness of spontaneous generation (Wright 1984; Levine and Evers, 1999).

At the same time, doubts in the traditional views of the "truth", referring to the inerrancy of the Bible, began to develop by the middle of the eighteenth century. The field of geology introduced the theory of uniformitarianism, a principle that states that the same processes that exist in the present are those which had acted in the past are sufficient to account for (geologic) change (Gould 1987). This theory proved to be widely accepted, and its fundamental idea was adopted by naturalists and applied to the perceptible differences between species. Ultimately, lending to the beginnings of the development of a modern view of evolution (Gould 1987; Levine and Evers, 1999).

Thought to be some of the first to put forth a general theory of evolution, was the French

philosopher, Pierre Louis Maupertuis (1698-1759), with his concept of heredity as a collection of a number of separate particles, and the French naturalist, Georges-Louis Leclerc, Comte de Buffon (1707-1786), who hypothesized the idea that species could change and believed to have spearheaded the modern doctrine of evolution (Mayr 1970; Bowler 1973). Around the same time, Erasmus Darwin (1731-1802), grandfather of Charles Darwin, a British physician, proposed that life had changed overtime on the basis of inheritance through acquired characters, yet his theory was conjectural and did not present a mechanism for which species could change (Wright, 1984).

Near the beginning of the 19th century, thought to be one of the most influential evolutionists at the time, Jean Baptiste de Lamarck (1744-1829), presented the first mechanism for evolutionary change. He conceptualized the theory of inheritance of acquired characteristics, commonly referred to as Lamarckism. This theory suggests that: (1) any organism may pass on acquired characteristics, such as the infamous example of the giraffes neck, to their offspring, and (2) the law of use and disuse, which proposed the notion that an organism may lose a particular characteristic if unnecessary and develop more useful characteristics in its place (Gould 2002). While this was a step towards unveiling a mechanisms for such change in species, it was heavily disputed. One palaeontologist by the name of Georges Cuvier(1769-1832), who was instead an advocate for catastrophism, proposed a theory which suggests that geological events formed as a result of some great catastrophe (Levine and Evers, 1999). After this, there was a number of others who presented their own theories, such as Richard Owen (1804-1892), a comparative anatomist, who called attention to the homology between species. However, despite many attempts to develop a mechanism for the noted change among species, no one person could present a clear and undisputed theory.

1.1.2 Darwin's Revolutionary Theory

"There is a grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one: and that whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved."

Charles Darwin, On the Origin of Species (1859)

The absence of a widely-accepted theory ended with the publication of On the Origin's of Species (1859) put forth by Charles Darwin (1809-1882), where he presented his revolutionizing theory of evolution, and most significantly, his theory of natural selection, a mechanism for change and dynamic force of his evolutionary theory. The process of natural selection describes the means by which certain heritable characteristics, or "traits", lend to the likelihood of an organism's survival and, in turn, promote its ability to successfully reproduce in a population over successive generations (Darwin 1859). Darwin arrived at this theory through his observations made during his almost five year journey aboard the H.M.S. Beagle to South America, were he spent time accumulating data and forming hypotheses (Van Wyhe 2008). However, while this theory indicated the necessity of some form of trait heritability, it was not entirely unique. No one had previously presented the effects of selection acting on small random variations in an organism's reproductive capacity over an extensive period of time as clearly (Wright 1984). As Darwin summarized in On the Origin of Species (1859):

"If during the long course of ages and under varying conditions of life, organic beings vary at all in the several parts of their organisation, and I think this cannot be disputed; if there be, owing to the high geometrical powers of increase of each species, at some age, season, or year, a severe struggle for life, and this certainly cannot be disputed; then, considering the infinite complexity of the relations of all organic beings to each other and to their conditions of existence, causing an infinite diversity in structure, constitution, and habits, to be advantageous to them, I think it would be a most extraordinary fact if no variation ever had occurred useful to each being's own

welfare, in the same way as so many variations have occurred useful to man. But if variations useful to any organic being do occur, assuredly individuals thus characterised will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance they will tend to produce offspring similarly characterised. This principle of preservation, I have called, for the sake of brevity, Natural Selection."

Charles Darwin, On the Origin of Species (1859)

Most unique to Darwin's theory of evolution was his recognition of the existence of a single "superior" trait, and that, through the mechanism of natural selection, such a trait could prevail over the previously dominating traits. Additionally, he showed that this cycle of dominance continued over time and thus lead to advancements (Wright 1984). It can be said that Darwin's theory relied upon logical inferences taken from general observations. He reported that variability could be detected in all species under surveillance and at least some of the observed variability is hereditary (Wright 1984). From this point on, thinking shifted from biological species as inherent properties of nature to products of a major force generating gradual change.

In spit of his revolutionizing theory, Darwin, while presenting a solution to the species problem, did not solve the problem of speciation, contrary to his belief. This largely had to do with his coupling of the problems of change, this being evolution, and that of diversity, noted as speciation (Singh 2000). Therefore, the mechanism producing the visible groupings of organisms remained unsolved, this being the enigma of speciation.

1.1.3 Natural Mechanism of Speciation: The Missing Link

While Darwin's theory received international interest and recognition - the majority of the educated public generally accepted that evolution occurred in some form - there remained some weaknesses and difficulties with his theory. His failure to explain the source of variability with his theory of natural selection, as well as his inability to describe the method in which traits transcend generations, prevented an absolute acceptance of his theory. The response to this quandary came in 1866, when Gregor Mendel (1822-1884) used Pisum sativum, garden peas, to observe recurring patterns of inheritance. Through his studies, he was able to trace characteristics through successive generations (Weiling 1991; Yon Rhee 1999). With his experiments with garden peas and their multiple characteristics, and later his work with honeybees, Mendel became the first person to introduce the concept of heredity.

Unfortunately, during his life, Mendel's work was overlooked, predominately due to most scientists supporting the concept of blending inheritance and perhaps as a result of his work being ahead of its time (Janick 1989). Mendel's theory was given new life, however, thirty-four years after its discovery in the early twentieth century and his journal entitled, Experiments with Plant Hybrids, has endured as one of the most influential publications and laid the ground work for modern genetics (Yon Rhee 1999).

1.1.4 The Marriage of Darwinism and Mendelism

"His words, his calculations were to take a sudden belated flight out of the dark tomblike volumes and be written on hundreds of university blackboards, and go spinning through innumerable heads."

Loren Eiseley, Darwin's Century. Evolution and the Men Who Discovered It (1961)

The words of Loren Eiseley describe the rediscovery and consequential impact of Gregor Mendel's work in the 1900's, after it went unacknowledged for over three decades. Most would attribute the precursor to the rediscovery of Mendel's work to the German evolutionary biologist, August Weismann (1834-1914) (Novák 2008). He proposed the rudimentary idea of the division of cells into germ and soma cells, "germinal selection", and that the germ cells possessed heritable information. The vast amount of his work was dedicated to the research of development, heredity, and variation (Kutschera and Niklas 2004; Novák 2008). Following Weismann's work, was the revival of Mendel's theory which has been credited to three independent researchers — Hugo Marie de Vries (1848-1935), mainly known for his proposed concept of genes, as well as Carl Correns (1864-1933), and Erich von Tschermak (1871-1962), published their own experimental work drawing the same conclusions as Mendel (Janick 1989).

The full significance of Mendel's work was not fully unveiled until the combined efforts of R.A. Fisher (1890-1962), J.B.S. Haldane (1892-1964), and Sewall Wright (1889-1988) along with a number of other researchers, in the 1930s and 1940s, who are credited with the 'marriage' of Darwinism and Mendelism, where natural selection is the mechanism of change and genes are the units on which it functions (Olby 1997) (Figure 1.1). It was at this time that Mendelian

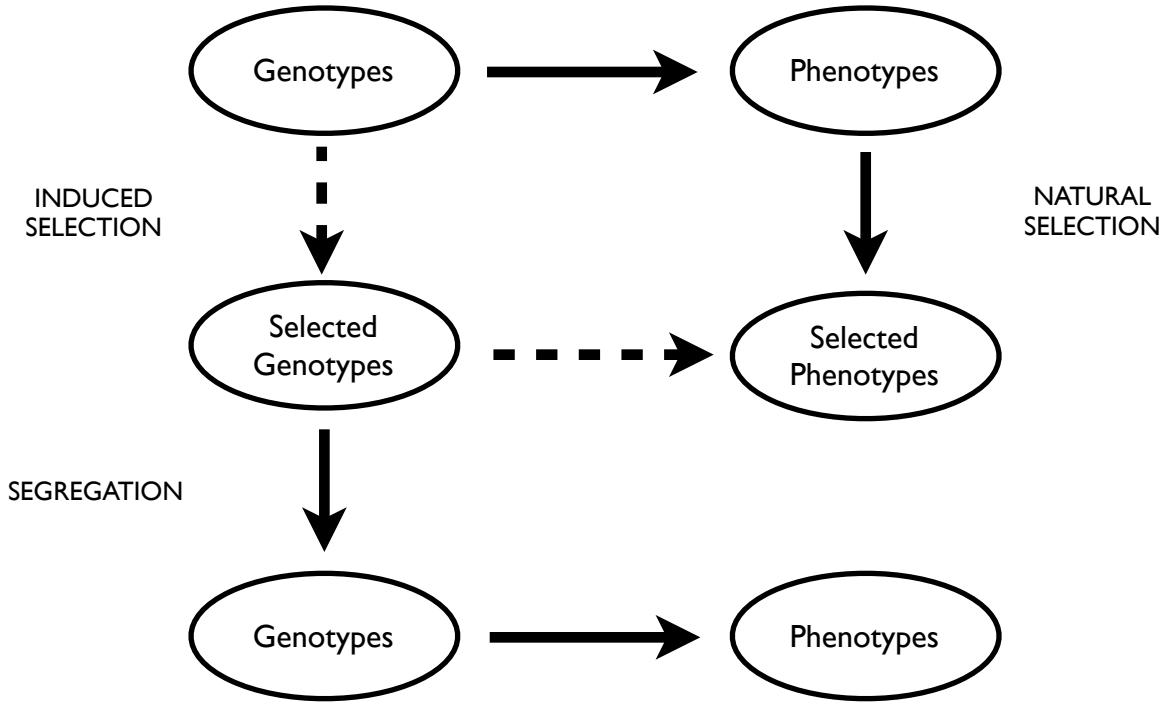


Figure1.1: A simplified scheme of Darwinian evolution by natural selection incorporating Mendelian indirect inheritance. Adapted from Christiansen 1990.

genetics had emerged as the basis of variation observed among organisms required by Darwin's theory of evolution and was subsequently deemed the "modern evolutionary synthesis", which persists to this day (Kutschera and Niklas 2004) (Figure 1.2).

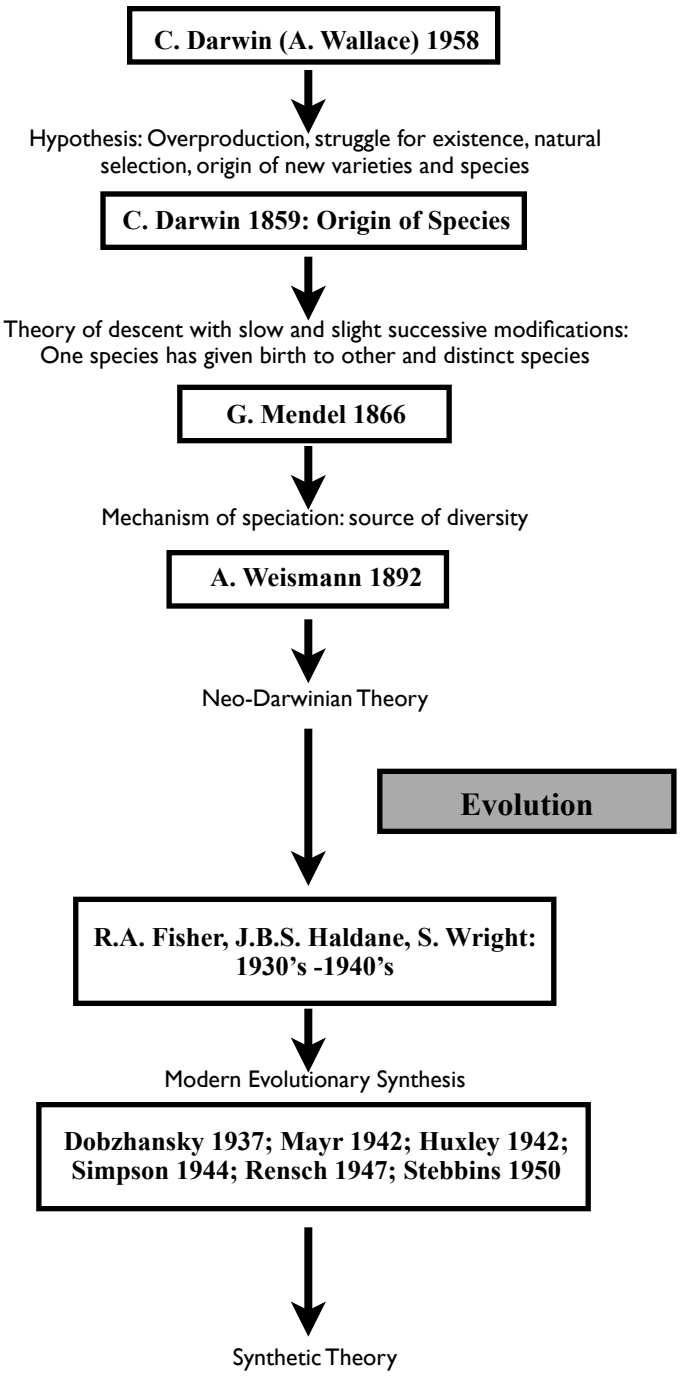


Figure 1.2: Scheme illustrating the historical development of the concept of evolution: from the hypothesis of Darwin, through the Neo-Darwinian theory, to modern evolutionary synthesis, to the synthetic theory. Adapted from Kutschera and Niklas 2004.

1.2 Modern Synthesis and Speciation

"The grossest blunder in sexual preference, which we can conceive of an animal making, would be to mate with a species different from its own..."

Fisher, R.A. The Genetical Theory of Natural Selection. (1958)

1.2.1 Genetics of Speciation

Over one hundred and fifty years ago, through his revolutionizing paper, Charles Darwin (1859) introduced the notion of natural selection, yet the underlying mechanisms which exists in order to allow all species to evolve in isolation remains one of the great conundrums within the field of biology (Coyne 1992; Kulathinal and Singh 2008). The inability to unveil the fundamental mechanisms necessary for the process of speciation was believed to result from the lack of a working definition of a species (Mayr 1942). The development of an operational definition for a species began when E. Mayr and T. Dobzhansky approached the act of defining species and imposed a more 'Modern Synthesis' perspective of a species (Mallet 2010). A group was characterized as a species through reproductive isolation from other such groups (Dobzhansky 1935; 1937; Mayr 1940; 1942; 1963). Even with this now widely accepted definition of a species, many questions remained concerning the specific workings of the theory, such as: How much variation among the genes are necessary to constitute a new species? As well as, the question of whether it is simply standard changes in allele frequencies or is it more complex, requiring novel evolutionary processes (Coyne 1992).

The difficulty in answering such questions is the consequence of speciation being such a slow process, thus making direct observation difficult. However, due to what has been determined about the process of speciation, it is believed that it may be investigated through

patterns it leaves behind (Coyne 1992). Studies surrounding these patterns of speciation have been investigated, dating back to 1936 in a paper by Dobzhansky, yet the importance of the study of speciation in the field of evolutionary biology was not fully realized until the last thirty years, largely due to the advent of molecular technologies (Carson 2000).

1.2.2 Genetics of Reproductive Isolation

Darwin was able to propose a theory of evolution which is still universally accepted in present day, however, his proposed mechanism for speciation was erroneous. Currently, the mechanism of species formation is one of the most prominent, yet most elusive, enigmas in evolutionary biology (Dronamraju 2011; Palumbi 1994). Ernst Mayr (1904-2005) proposed a solution to the 'species problem', referring to the ability of multiple species to evolve from a single common ancestor. In his publication in 1942, entitled Systematics and the Origin of Species, Mayr moved away from the predominate belief of grouping individuals based on morphological similarities and presented a new definition for the notion of species, which he entitled the 'Biological Species Concept' (BSC) (de Queiroz 2005): ''species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups'' (Mayr 1942).

Mayr acknowledged that he was not the first to suggest such a definition for species, citing publications by K. Jordan and E.B. Poulton where they indicated that species were syngamic, suggesting they formed 'reproductive communities'. However, this concept of species became commonly accepted during the advent of the Modern Evolutionary Synthesis, independently introduced by S. Wright, T. Dobzhansky, G. Simpson and, most notably, Ernst

Mayr (Mallet 2004).

In light of this novel definition of a species, it prompted many to inquire after the source of species inability to interbreed. The resolution to this query came with a publication by Theodosius Dobzhansky in 1937 entitled, *Genetics and the Origin of Species*, where he outlined the evolution of 'reproductive isolation' as a key player in the formation of species. The mechanisms of reproductive isolation can be classified in two types: pre-zygotic, referring to mechanisms which take place prior to fertilization — such barriers as temporal isolation (different reproductive periods), behavioural isolation (courtship) and mechanical isolation (prevention of copulation through dissimilar reproductive organs) — and post-zygotic, referring to mechanisms which take place post fertilization — referring to reduced sterility and inviability of species hybrids (Orr and Presgraves 2000). The latter of the two mechanisms will be highlighted and remain the focus of this thesis.

1.2.3 Bateson-Dobzhansky-Muller Model of Speciation

Prior to the 1930s, many biologists accepted evolution as fact, yet many cast doubt over the ability of natural selection of random mutation to justify the observed degrees of adaptation between organisms (Haldane 1932). The works of R.A. Fisher (1930), S. Wright (1931) and J.B.S. Haldane (1932) renewed confidence in Darwin's evolutionary theory through the use of mathematical analysis - which married the newly develop science of genetics with the established theory of evolution (Leigh 1990). These three researchers focused on mathematical analysis of gene frequencies in light of mutation, selection, and genetic drift. Unanimously, they concluded selection to be the vehicle of change, over the previously favoured notion of mutation

(Nei 1987). As a result of this revolution, a number of scientists published works — Dobzhansky (1937), Huxley (1942) and Mayr (1942), to name a few — based on the growing understanding of genetics to illuminate the causes of evolution, thus creating the Modern Evolutionary Synthesis or "neo-Darwinian synthesis".

However, while the idea of post-zygotic isolation as a major player in speciation had been presented, questions remained regarding the causation behind such hybrid sterility (Johnson 2002). The foundation for answering this was independently presented by T. Dobzhansky (1937), H.J. Muller (1940, 1942) and, while generally overlooked, W. Bateson (1861-1924) (Queiroz 2005). It has been said that Bateson first presented a solution to explain the genetic basis of reproductive isolation less than a decade after the rediscovery of Mendel's work, yet Theodosius Dobzhansky is credited with the first proposed evidence of such genetic incompatibilities as the source of hybrid dysfunction. He did so through his empirical work creating backcrosses with hybrids between Drosophila pseudoobscura and D. persimilis (Ehrman 1960; Tamura et al. 2004; Noor and Feder 2006). He concluded that the sterility observed in these backcrossed hybrids resulted as a direct response to epistatic interactions between chromosomes (Dobzhansky 1936; Orr 1997; Bomblies and Weigel 2007).

Around the same time, Hermann J. Muller supplied a more theoretical approach to Dobzhansky's empirical treatment of the theory in an essay published in 1942, which has been noted to have greatly contributed to the conceptualization of speciation genetics (Orr 1997; Johnson 2002). In his essay, Muller argued that incompatibles at interacting loci completed the process of speciation, opposed to the more commonly accepted idea that it began the process (Muller 1942; Carlson 1971). Muller was also credited with the ability to demonstrate that

incompatible interactions among loci require more than two interacting factors, when in actuality could consist of multiple interactions, deemed ‘complex’ (Johnson 2002). Lastly, and one of the most significant contributions made by Muller, was his introduction of the ‘Dominance Theory’, which ultimately lead to the explanation of the theory of Haldane’s rule (see below). Muller proposed the concept for the potential of a unique consequence in the inheritance of incompatible genetic factors which would occur if one or both of these factors were found on the sex chromosome, with particular emphasis on the heterogametic sex (i.e. XY) which only inherits one the these factors (Orr 1997).

The amalgamation of Dobzhansky and Muller’s theories was formalized into a simple model, deemed the Dobzhansky-Muller model, which has come to describe the development of post-zygotic isolation in allopatry as a by-product of natural evolutionary divergence (Orr and Turelli 2001).

“ Two taxa separated by an adaptive valley can evolve even though no genotype ever passed through the valley.”

Orr, A. *Haldane’s Rule.* (1997)

Genetic substitutions, or mutations, accumulate as two lineages diverge from their common ancestor and will not cause significant disruption to a species viability due to natural selection and its rejection of deleterious alleles (Orr 1997; Orr and Turelli 2001). However, these same mutations within a different genetic background, possible when two species hybridize, may not be able to support such negative epistatic interactions, thus leading to sterility or inviability and ultimately giving rise to postzygotic isolation (Orr and Turelli 2001) (Figure 1.3).

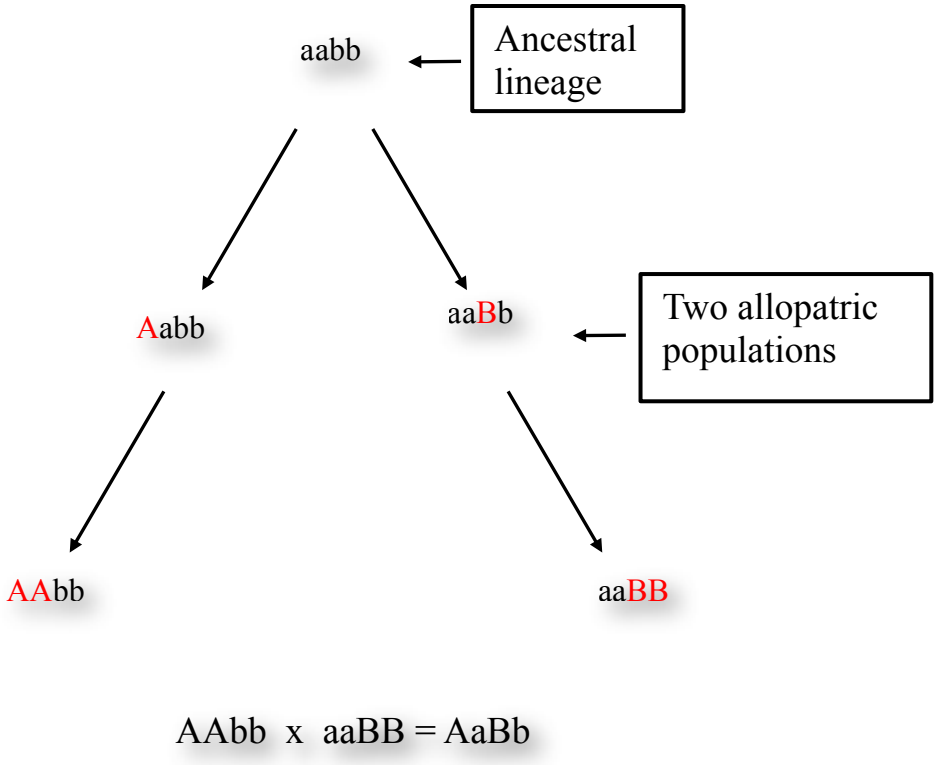


Figure 1.3: Graphical illustration of the 'Bateson-Dobzhansky-Muller' model of the evolution of genetic hybrid incompatibility. At two epistatic loci, a and b, diverging lineages may evolve substitutions, depicted in red, that are harmless in their native context, negatively impact hybrid offspring when brought together. This can occur when these loci are in the heterozygous state and cause the F1 hybrids to have reduced fertility or viability. Adapted from Bomblies and Weigel 2007.

1.2.4 Haldane's Rule

The theory proposed by Dobzhansky and Muller underlying postzygotic isolation has been thought to, at least in part, explain an almost universal generalization formulated by the British evolutionary biologist, J.B.S. Haldane (1892-1964) (Coyne 1985; Ridley 2004). The phenomena of hybrid sterility and inviability had been documented, and its key role in speciation via postzygotic isolation was eventually acknowledged. However, it was not until J.B.S. Haldane proposed a rule, later to be known as Haldane's rule, would indicate this phenomena's role in speciation (Wu et al. 1996). Haldane's rule has been regarded as one of the most uncontested and best-supported empirical generalizations in the field of evolutionary biology regarding patterns in speciation (Johnson 2002; Coyne 1994). Originally stated:

"When in the offspring of two different animal races one sex is absent, rare, or sterile, that sex is the heterozygous [heterogametic] sex."

J. B. S. Haldane. *Sex-ratio and unisexual sterility in hybrid animals*. (1922)

In this case, 'heterozygous' refers to the heterogametic sex, such as the XY in males among *Drosophila* and mammal species or, conversely, ZW in females of *Lepidoptera* and birds (Johnson 2000). Haldane was one of the first to discover hybrid sterility and inviability and it has been deemed as one of the strongest patterns in evolutionary biology, thus prompting the suggestion of its fundamental role in characterizing speciation.

Through his observations, Haldane formulated this rule which demonstrates, almost universally, that speciation through postyzgotic isolation occurs similarly across very distinct taxa. This finding warranted a great deal of attention due to the significance of an implied

fundamental similarity which existed among the genetic events causing speciation (Coyne 1992; Wu and Davis 1993; Orr 1997; Orr and Presgraves 2000). However, a clear explanation of Haldane's rule eluded the scientific world and thus, a need to better understand the mechanistic basis of this observation was required. Several hypotheses have been proposed as a means to explain this noted sterility and inviability among heterogametic hybrids, yet three of these hypotheses are at the forefront: The dominance theory, "faster-male" theory and the "faster-X" theory (Orr 1997).

As stated above, the dominance theory was first introduced by H. J. Muller as a means to correlate the Dobzhansky-Muller model of speciation with a feasible explanation for Haldane's rule (Muller 1942; Orr 1993; 1995; 1997). The basis of the theory centers around the notion that heterogametic hybrids are affected by both recessive and dominant genes which are linked to the X-chromosome and function in incompatible interactions, while at the same time, homogametic hybrids are only affected by the dominant genes (Coyne 1985). While Muller's theory was not always accepted, it was formalized and further refined towards the end of the 20th century by Allen Orr and Michael Turelli in a number of their papers (see Orr 1993; 1995; Turelli and Orr 1995; 2000; and Orr and Turelli 1996).

The "faster male" hypothesis focuses on the forces employed by sexual selection which cause faster evolution of male genes as well as a larger incidence of hybrid male sterility (Wu and Davis 1993; Presgraves 2002). Studies have produced evidence of a faster evolution among male reproductive characters over those in females due to a greater intensity of sexual selection acting upon males (Orr 1997). Therefore, while the dominance theory constitutes the observance of Haldane's rule among species possessing a degenerate Y chromosome, in instances where a

particular species lacks a degenerate Y chromosome, the faster male theory accounts for Haldane's rule when the dominance theory could fall short (Presgraves 2002). The faster male theory does however have deficiencies which is two-fold: (1) The theory is unable to explain hybrid inviability due to the fact that a male versus a female lethal does not exist. (2) Furthermore, the theory cannot account for sterility among species such as birds and Lepidoptera, where the female is heterogametic (Orr 1997).

Lastly, the "Faster-X" theory was offered as yet another explanation of Haldane's rule. A paper by Charlesworth, Coyne and Barton (1987) suggested as a by-product of the large X-effect — the idea that X-linked genes have a disproportionate effect on hybrid fitness — was that the X-linked genes would also be affected among heterogametic hybrids (Orr 1997). This proposed effect could be possible under the stipulation of rapid evolution of the loci on the X chromosome over those located on autosomes (Coyne 1992). Having said that, the faster-X theory may not be capable on its own to explain Haldane's rule due to the fact that female hybrids possessing double the number of X chromosomes lending to doubling the effect of recessive mutations (Orr 1993; 1997). In combination with the notion of dominance, however, it has been thought to intensify its effects (Orr 1997).

1.2.4.1 The Molecular Era: Renewed Interest in Speciation

By the middle of the twentieth century, soon after an expansion of molecular biology, the discovery of protein polymorphism was a significant step towards uncovering the mysteries of speciation. Following this discovery, emphasis was placed on aspects which may impact protein variation, such as, mutation, drift, and natural selection (Whitehead and Crawford 2006). Until

this point in history, never had there been a means to measure "genetic distance" (King and Wilson 1975). Several methods were developed to measure organismal diversity with the understanding of protein polymorphism, including comparisons of proteins via electrophoresis and sequencing techniques, to name a few (King and Wilson 1975; Coulthart and Singh 1988; Singh 1990).

It was later proposed that it may not necessarily be the proteins themselves producing the observed variation, but rather elements that control gene expression. These elements have been classified as either *trans*-regulatory elements (transcription factors) or *cis*-regulatory elements. The former bind to *cis*-regulatory elements, while the latter are sequences throughout DNA (Whitehead and Crawford 2006; Wittkopp et al. 2008). It was suggested that the observed differences in gene regulation are at the root of variation between closely related species, however this notion was relatively unexplored due to a lacking in the area of molecular technology to support further exploration. This all changed with the emergence of molecular technologies such as, DNA microarrays and real-time polymerase chain reaction (real-time PCR) (White et al. 1999; Orr and Presgraves 2000; Whitehead and Crawford 2006; Moehring et al. 2007; Catron and Noor 2008). However, despite the variations within the regulatory architecture between species there still remained difficulties when attempts were made to obtain a clear picture within the genetic backgrounds of pure species (Ranz et al. 2003; Rifkin et al. 2003; Ranz et al. 2004). A departure was needed from the conventional comparisons of the gene-expression profiles. This came in the form of species interspecific hybrids and the genetics of postzygotic isolation (Turelli and Orr 2000; Ranz et al. 2004; Clark 2006; Ortiz-Barrientos et al. 2007).

1.2.4.2 Role of Interspecific Hybrids

Central to the evolution of hybrid incompatibilities is the understanding of the molecular basis of these newly generated epistatic interactions (Landry et al. 2007; Wray et al. 2003; Gibson and Weir, 2005). However, while natural populations harbor extensive variation in gene expression (Wray et al., 2003; Ranz and Machado, 2006), it appears as cryptic genetic variation (CGV) (Gibson and Dworkin, 2004) allowing for sibling species to exhibit consistent phenotypes yet possessing distinctive regulatory architecture (Britten and Davidson, 1969; Moehring et al., 2006; Landry et al., 2005; Orr and Presgraves, 2000; Artieri et al., 2007). This occurrence is a consequence of stabilizing selection (Haerty and Singh, 2006), a process termed 'developmental systems drift' (DSD) by True and Haag (2001), and thus has given rise to compensatory changes which has extended to widespread conservation of gene expression levels between sibling species (Artieri et al., 2007; Ranz et al., 2004). However, within these genetic backgrounds, interspecific hybrids no longer have the ability to obscure sequence divergences due to the amalgamation of incompatible gene products which have been encoded by different genomes and are now present in the same cells (Ranz et al., 2003; 2004; Rifkin et al., 2003; Landry et al., 2005). As a result, interspecific hybrids have allowed the unique opportunity to uncover the incompatibilities which exist within the regulatory networks between species and ultimately demonstrate what had previously been observed at the morphological level (Eberhard 1985; Joly et al. 1991) now at the molecular level (Clark 2006).

1.3 The Study of Speciation via Gene Expression

1.3.1 Regulatory pathways and epistatic interactions

Ubiquitous to all species are the presence of regulatory pathways which aid in the realization of specific phenotypes (Johnson and Porter 2000). This allows species to remain in distinct groupings or develop into unique species. Interspecific hybrid incompatibilities may develop due to isolated populations having a specific trait — determined via regulatory pathways — which experience directional natural selection causing the trait to change (Johnson 2008). A large number of traits associated with reproductive isolation, specifically postzygotic isolation, pertain to reduced hybrid viability and sterility are associated with sex-related genes and are often involved in the formation of new species (Coyne and Orr 1989; Presgraves 2002; Michalak and Noor 2003). Distinctive to hybrid dysfunction is the requirement of epistasis — non-additive genetic interactions among distinctive loci — which has been found to have gene interactions which can produce complex patterns (Johnson and Porter 2000). Due to the paucity of reproductive isolation, despite its potential to occur with relative ease, the notion of phenotypes thought to be developmentally additive or multiplicative seems to be less feasible (Johnson and Porter 2000; Johnson and Porter 2002; Michalak and Noor 2003). Instead, it has been proposed that divergence in regulatory pathways appears to be the most promising hypothesis (Johnson and Porter 2000). Therefore, if one may identify these “failures” in the regulation of gene expression one may be able to, first, establish if they are associated with hybrid sterility and inviability, and second determine which specific genes or genetic pathways may be responsible (Michalak and Noor 2003).

1.3.2 Genome-Wide Expression Profiling

The notion of transcriptional dynamics acting as a mechanism of hybrid sterility and inviability was first put forth as theoretical models (Johnson and Porter 2000). They described complex interactions between transcription factors (*Trans*-elements) and regulatory elements (*cis*-elements) which would be revealed through disruptions in hybrid genomes. Soon to follow, empirical work was conducted based on the previous theoretical models through genome-wide analyses of comparisons in expression profiles between pure species and their F1 interspecific hybrids (Mavarez et al. 2009). In its infancy, studies (Reiland and Noor 2002) used such techniques as differential display — which compared and identified changes in gene expression at the level of mRNA between a pair of cell samples — however, this technique lacked sensitivity and only a single gene could be identified as having differential expression. With advancements in technology, techniques involving oligonucleotide analysis, cDNA microarrays as well as quantitative real-time polymerase chain reaction (qRT-PCR) were employed in a number of studies (Michalak and Noor 2003; Barbash and Lorigan 2007, Ranz et al. 2004; Haerty and Singh 2006; Moehring et al. 2007 and Michalak and Noor 2004, respectively). These studies confirmed ideas presented in the theoretical models suggesting patterns of gene misexpression in interspecific hybrids (Ortiz-Barrientos et al. 2007). Two main conclusions emerged: (1) a large number of genes display an expression level in their hybrids under that observed in the pure species. This confirms the notion that gene regulatory divergence significantly contributes to Dobzhansky-Muller incompatibilities. The other conclusion (2) describes a greater proportion of the differentially expressed genes to experience a male-bias

(higher or exclusive expression in male somatic or gonadal tissues) (Johnson and Porter 2000, 2002; Landry et al. 2007; Ortiz-Barrientos et al. 2007 and Mavarez et al. 2009).

1.4 Sexual Dimorphism at the Genome Level

Sexual dimorphism is the phenotypic differences between the sexes within a species and has been recognized as a major contributor to observed biological diversity among sexual organisms (Parisi et al. 2004). The morphological variation between the sexes has been well documented — some of the most commonly used examples are the plumage of the male peacock's tail and the birds of paradise — and has raised questions over its incidence due to the maladaptive nature of many of these traits (Singh and Kulathinal 2005). In accordance with Charles Darwin's theory of natural selection, the phenomenon of sexual dimorphism seems paradoxical. To circumvent this paradox, Darwin (1859) proposed the theory of sexual selection as a response to the overwhelming evidence suggesting the greatest degree of variation existing among males (Huxley 1938). Furthermore, with recent advancements in molecular technology, genome-wide expression profiles have uncovered extensive molecular sexual dimorphism, deemed "sex-biased" gene expression (Connallon and Clark 2011). Therefore, what was first observed at the morphological level has been substantiated at the molecular level.

1.4.1 Sexual Selection and its Role in Speciation

Sexual selection can be defined as the occurrence of a more pronounced secondary sexual traits in one sex over the other of the same species (Andersson 1994). More specifically, these secondary sexual traits are predominately found in males of a species as a direct response to

pressures exerted by female choice (Singh and Kulathinal 2005). As stated above, this variance between the sexes is not only at the morphological level but, despite both sexes developing from the same underlying genome, have been found to be present at the genomic level and is a direct response to evolutionary changes among male-specific (male-biased) genes (genes with a higher or exclusive expression within the male sex) (Ellegren and Parsch 2007; Mavarez et al. 2009 and Connallon and Clark 2011). It has been shown that these male-specific genes are undergoing a faster rate of evolution and contribute significantly to speciation (Wu et al. 1996; Singh and Kulathinal 2000; Haerty and Singh 2006 and Mavarez et al. 2009).

1.4.2 Sex-Specific Expression Patterns

The sexes within a species develop from the same genome (with the exclusion of the Y/ W-chromosome) yet experience differential phenotypic and genomic patterns. Thus, there is a necessity to exploit differential divergence of these shared genes (Ellegren and Parsch 2007 and Connallon and Clark 2011). There are two ways in which the sexes can differentially utilize shared genes: (1) each sex can experience alternative splicing of a coding sequence or (2) undergo sex-specific selection giving rise to the evolution of sex-specific (both male- and female-specific) genes (Connallon and Clark 2011). Genome-wide analyses have been used to identify and compare male- and female-specific levels of gene expression and have found that male-biased genes consistently exhibit a greater degree of divergence between species at the level of expression (Parisi et al. 2004 and Ellegren and Parsch 2007). A study by Conallon and Knowles (2005) described male-specific genes displaying increased expression in males when compared to females suggesting a faster rate of evolution driven by positive selection (Gallach et al. 2011).

1.5 Rapidly Evolving Genes (REGs)

A number of recent gene expression studies have support this notion of rapid evolution of sex-specific genes (Ranz et al. 2003, 2004; Meiklejohn et al. 2003; Parisi et al. 2004; and Haerty and Singh 2006). Using the technique of cDNA microarrays, these studies revealed that between sibling species of *Drosophila*, sex-specific expression is rapidly evolving within the entire genome and contributes to ~ 30-50% of genes displaying divergence in expression between the sexes (Ranz et al. 2003; 2004; Parisi et al. 2003; Singh and Kulathinal 2005 and Ortiz-Barrientos et al. 2007).

1.5.1 Tissue-Specific Expression

With the discovery of these rapidly evolving genes it become important to identify which genes were affected and where these genes were localized. A study by Coulthart and Singh (1988) used a method of two-dimentional electrophoresis to investigate the localization of such rapidly evolving genes and discovered a large proportion of these genes were found in the reproductive tissues. Furthermore, another study conducted by Civetta and Singh (1995) found a battery of proteins within both the testis and ovary tissues to exhibit a greater divergence when compared to proteins of somatic tissues. More recently, studies comparing the DNA of varying *Drosophila* species have noted a rapid evolution in sex and reproduction-related genes (SRR-genes) between closely related sibling species (Singh and Kulathinal 2000; Swanson and Vacquier 2002; Haerty and Singh 2006). Later, in a paper by Ranz et al. (2003), the observed rapid evolution in reproductive tissues was also found to be the case at the transcriptional level, displaying patterns of sex-biased expression between sibling species. Interesting, it was found

that many of these rapidly evolving genes are expressed in more than one tissue despite the known differences which exist in the evolutionary rates of tissue-specific and shared genes (Jagadeeshan and Singh 2005; Khaitovich et al. 2005). Consequently, it would be advantageous to investigate the degree to which specificity influences the rates of evolution among these genes when looking at tissue-specific expression.

1.6 Thesis Overview

1.6.1 The Species

Two sibling species of the *simulans* clade were the focus of this study and were retrieved from the *Drosophila* species stock center in Arizona; the stock numbers are given in brackets. *D. simulans* (14021-0248.197) and *D. sechellia* (14021-0251.25) (from Joffreville, Madagascar and Seychelles island, respectively). *Drosophila simulans* is a cosmopolitan species, like the thoroughly investigated *D. melanogaster*, and *D. sechellia* is an island species restricted to the three known Seychelles islands and it has been suggested that *D. sechellia* may have resulted from a colonization by a proto-*simulans* ancestor (Coyne and Kreitman 1986). These sibling species diverged ~ 250, 000 years ago and it has been shown that the *D. simulans* is more closely related to *D. melanogaster* (Figure 1.5) (Caccone et al. 1988; Catron and Noor 2008; McDermott and Kliman 2008). Both species have identical chromosome banding sequencing and are nearly indistinguishable other than the shape of the posterior process of the male genital arch (Lemeunier and Ashburner 1984 and Coyne and Kreitman 1986). Due to *D. sechellia* specialized food preference to a toxic fruit *Morinda citrifolia*, it has been found to be difficult to rear this

species under laboratory conditions and therefore can be a difficult species to maintain (Dworkin and Jones 2009). The interspecific hybrid crosses between these species have varying fitness — The reciprocal crosses produce sterile F1 males and fertile females, however the cross between ♀*D. simulans* x ♂ *D. sechellia* is a 'successful' cross while it has been suggested that the cross between ♀*D. sechellia* x ♂ *D. simulans* is 'difficult'(specifically, only 1 out of 7 crosses are successful) (Coyne and Kreitman 1986 and Barbash 2010).

1.6.2 Objectives

The purpose of this thesis has been to investigate a number of intriguing questions that aim to supply a wealth of data regarding tissue-specific expression in both fertile females and sterile male hybrids, as well as advance our understanding of the process of species divergence and speciation. Cardinal to this study was the comparison of expression profiles of previously studied genes which were categorized according to *D. melanogaster* expressed sequence tags (ESTs), with those determined in this thesis, using *D. simulans* and *D. sechellia* hybrids. This will allow for possible insight into the variation in expression patterns between species. Furthermore, this thesis intends to reveal important information exhibited through unique expression patterns when looking at four different tissue types — male and female heads, ovary and testis — two species and their reciprocal crosses. We can answer several questions as well as test the following hypotheses: (1) Do *Drosophila* sibling species differ in tissue-specific distribution of gene expression? (2) Do males and females differ in tissue specific distribution of expression? (3) Do reciprocal crosses differ in patterns of gene misexpression suggesting X-effect? (4) Do sex and

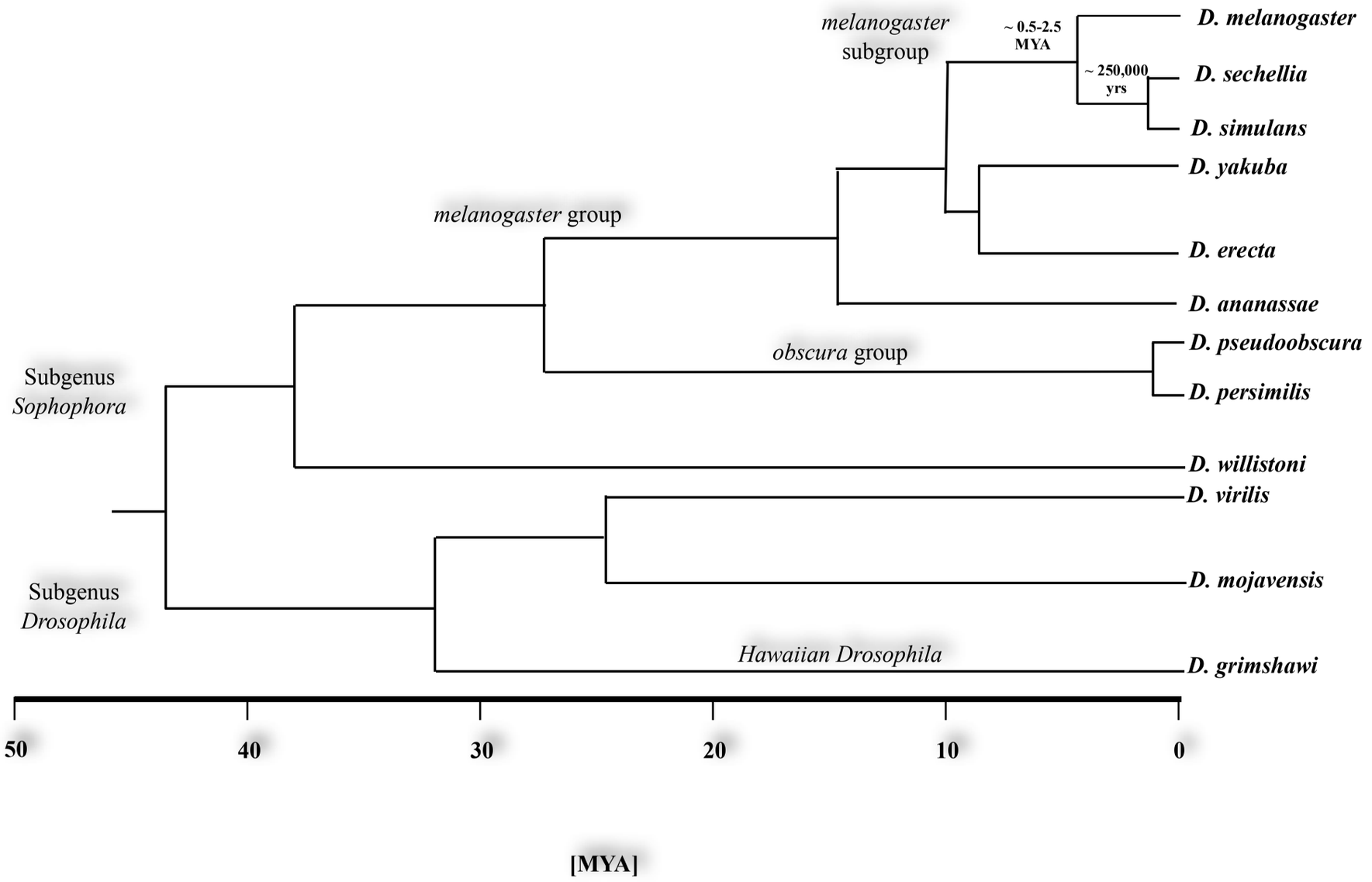


Figure 1.5: The phylogeny of the genus *Drosophila* adapted from Powell 1997. Within the *melanogaster* subgroup, *D. melanogaster*, *simulans*, and *sechellia* are closely related species. The sibling species, *D. simulans* and *D. sechellia* diverged from *D. melanogaster* ~ 0.5 - 2.5 million years ago and diverged from each other ~ 250,000 years ago (Caccone et al. 1988; Catron and Noor 2008; McDermott and Kliman 2008).

non-sex genes differ in extent of gene misexpression? Ultimately, this thesis will provide a notable contribution to the field of gene regulatory evolution and speciation.

MATERIALS AND METHODS

2.1 *Drosophila* Species

Two sibling species of the *simulans* clade were the focus of this study and were retrieved from the *Drosophila* species stock center in Arizona; the stock numbers are given in brackets. *D. simulans* (14021-0248.197) and *D. sechellia* (14021-0251.25) (from Joffreville, Madagascar and Seychelles island, respectively) were received in two vials for both species, one containing 30 adult individuals and the other containing a combination of larvae and pupae (Table 2.1). Flies were reared on cornmeal-molasses-yeast medium at 25ºC. Due to *D. sechellia* specialized food preference to a toxic fruit *Morinda citrifolia*, it has been found to be difficult to rear this species under laboratory conditions. To impede this, *D. sechellia* stocks were maintained on medium which had a yeast mixture, yeast mixed with lukewarm tap water, painted across a small portion of the medium (Table 2.2). As a further measure, greater than one-day-old food supplies were scored along the surface as a means to aid female egg implantation.

Table 2.1: List of *Drosophila* stocks

Species Stock List	Obtained From
Drosophila simulans (14021-0248.197)	UC San Diego Drosophila Stock Center Joffreville, Madagascar
Drosophila sechellia (14021-0251.25)	UC San Diego Drosophila Stock Center Seychelles Island

2.2 Parental Species and Tissue Collections

Both males and females of pure species were collected at the pupa stage, approximately 6-9 days after hatching, and sexed under a dissecting microscope (Vista Vision Stereozoom) via the observed presence or absence of sex-combs along the first tarsal segment of the *Drosophila* forelimb. This method was performed over the customary anaesthetized method of sexing flies as a way to achieve more accurate collections of virgin specimens. The flies were then separated into 35mL glass vials with cotton plugs and approximately one inch of medium with no more than 30 individuals in a single vial as a means to prevent density dependent growth effects. Flies were maintained for four days (96 hours), at which point the ovary, testis, and heads of both males and females were dissected in a 1x PBS (phosphate buffered saline) buffer solution, pooling 20 of each tissue in respective 0.2mL tubes and submerged immediately in 100µl of RNALater (Ambion) and placed at -80ºC for later use (see below) (Figure 2.1) (Table 2.3). Flies that emerged prior to pupae collection, and therefore their virginity in question, were discarded.

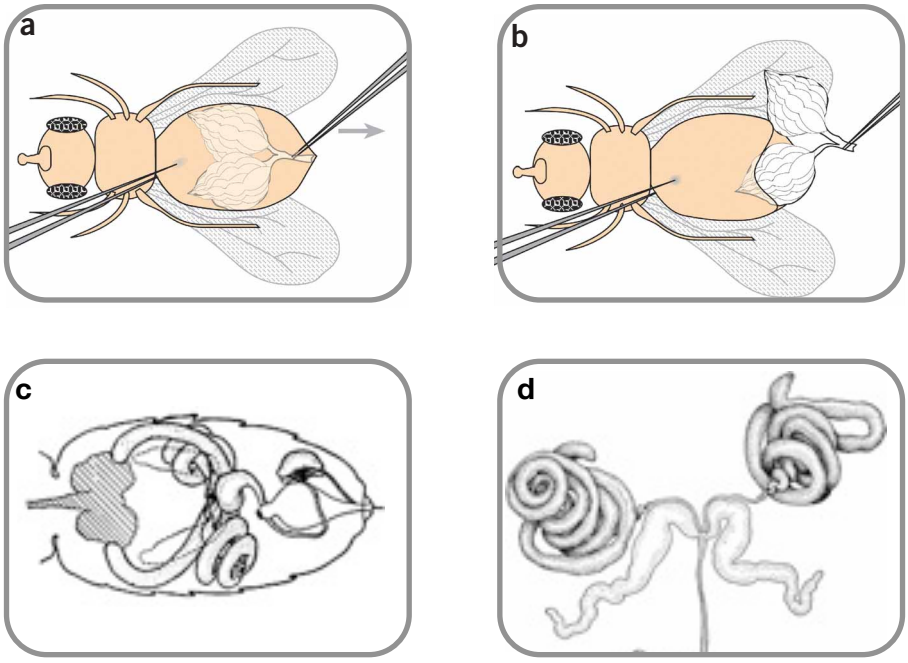


Figure 2.1: Schematic illustration of the testis, ovary and head tissue dissections. (a) The ovaries of anesthetized adult females were dissected using two sets of forceps in 1xPBS solution. (b) the left forceps secure the distal end of the thorax and the right forceps pinch the distal end of the abdomen and pulls away delicately. (c) The testis of anesthetized adult males were dissected using two sets of forceps in 1xPBS solution. (d) the left forceps secure the distal end of the thorax and the right forceps pinch the distal end of the abdomen and pulls away delicately and the gut and accessory glands were removed. The head tissue was removed in a similar manner simply by squeezing the head away from the body using the forceps. The images in (a) and (b) were adapted from Prasad et al. 2007 Figure 1 and the images in (c) and (d) were adapted from www.FlyBase.org under the 'ImageBrowse' tab.

2.3 *Drosophila* Hybridizations and Tissue Collections

Both virgin male and female flies were collected from parental pure species stocks with one day and four day periods prior to mating, respectively. Hybridization crosses were constructed from mating ten virgin four-day-old females of one species with one-day-old virgin males from the other species in a 1:1 sex ratio. The two reciprocal hybridizations took place in 35mL glass vials with cotton plugs and approximately one inch of medium. These crosses were then transferred to a new vial containing the same medium every 3-4 days as a way to attain as many hybrids per cross as possible. The empty vials were kept and observed daily for larvae and pupae formation. Upon discovery, the cross would be discarded along with any progeny that may have ensued. Hybrid pupae were collected from each cross, sexed, and maintained for a four-day period in the same manner described above. After this 96-hour time period, 150 individuals were dissected for each hybrid cross following the same protocol described above.

With respect to the newly emerged F1 hybrids, in contrast to the procedure of pure species, they were never discarded. They were maintained in separate glass vials which possessed the same cornmeal-yeast-molasses medium and retained for approximately 15-20 days. With the knowledge that these F1 hybrid males are sterile (Lachaise et al. 1986) one could deduce whether or not the parental females were virgins through the observation of the materialization of larvae. If any were found at the 20 day mark, all individuals collected from the specific cross were discarded.

2.4 Total RNA Extraction and cDNA Synthesis

To obtain the required quantities of RNA for qRT-PCR assays, total RNA was extracted using Trizol (Invitrogen, Carlsbad, CA) from four-day-old individuals from both pure species and their F1 reciprocal hybrids from a pool of 60 tissues in each category: male head, female head, ovary, and testis (dissected without accessory glands). Some tissues were homogenized using a handheld homogenizing instrument; these included the head tissues due to their tough exoskeleton, as well as the ovaries. The testes were simply vortexed in Trizol solution. Samples of RNA were checked for their quality and concentration and in ng/µl using a NanoDrop ND1000 speterophotometer and the integrity of the RNA was assessed using the Agilent 2100 bioanalyzer.

The RNA was then treated with DNase I (Qiangen), used to cleave unwanted DNA that may have remained in the samples. The samples then were converted into cDNA using qScript™ cDNA Synthesis Kit (Qunata Biosciences), and stored at -20°C until qRT-PCR (see below).

2.5 Gene Selection and Primer Design

2.5.1 Gene Selection

Prior to this study, cDNA microarray hybridization data was generated from studies of gene expression in hybrid male whole body, as well as male whole body and testes, between *D. simulans* and *D. sechellia* (Artieri and Singh, 2009; Haerty and Singh, 2006). Within these studies, the genes under investigation were selected from *D. melanogaster* wildtype expressed sequence tags (ESTs). They were assigned to specific tissues as their site of expression based on an arbitrary cut-off of at least three ESTs per tissue type. Following analysis, these genes were

deemed as either misexpressed or non-misexpressed in hybrids when compared to the expression patterns of both pure species, as per the criteria used in Hearty and Singh (2006). From this data set, a total of 36 genes were selected from the category of genes found to be misexpressed in hybrids. A total of 2 genes were selected from the category of genes found to be non-misexpressed in hybrids. Two additional genes were selected from a gene set described as ovary-biased, and were generated from female hybrids where the ovary samples included the ovary, lateral oviduct, and most of the common oviduct (Parisi, et al. 2004). These genes were found to be misexpressed in the hybrid. A total of 40 genes were selected as candidate genes and were divided into five categories: 2 Non-misexpressed; 2 Testis-specific; 2 Head-specific; 2 Ovary-biased; 32 Mixed-tissue, which were further subdivided into four categories: 14 HOT (head, ovary, and testis), 1 OH (ovary and head), 4 OT (ovary and testis) and 13 TH (testis and head).

A sample set of 40 genes were chosen to explore tissue-specific expression between parental species and their F1 hybrid and among tissue within the hybrid: genes were chosen at random, using a PerlScript software program, from list of previously determined misexpressed genes in hybrids between species of the genus *Drosophila* (Haerty et al. 2006; Artieri et al. 2007; Parisi et al. 2004). This randomized selection was done to eliminate a bias in our results. Furthermore, these randomly chosen genes were subdivided into specific categories: 2 were testis-specific (T), 2 were ovary-biased (O), 2 were head-specific (H), 32 were mixed tissue (OH, OT, TH, HOT) and 2 were non-misexpressed genes within these hybrids.

2.5.2 Primer Design

Using the *Drosophila* genome annotation project (http://rana.lbl.gov/drosophila.supplementary Table 1), we were able to retrieve coding sequences for *D. melanogaster*, *D. sechellia* and *D. simulans* for all 40 genes of interest as well as the internal control gene, RpL32, from the FlyBase web site (www.flybase.org). For each gene, under the drop-down tab orthologs there are two categories: Genome-wide drosophilid orthologs and Curated drosophilid orthologs. The genomes under the curated category take precedence over genome-wide, however both were used. The GenBank was selected to retrieve the genome region, which was then placed in a file written in a FASTA format. Sequences were aligned using ClustalW2 (www.ebi.ac.uk/clustalw/) according to the protein sequence alignments. Primers were designed based on the observed conserved regions, 3' most constituently expressed exon, through the use of the online program Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi), which were then tested against the aligned sequences using BioEdit Sequence Alignment Editor software (see Supplementary Table § 3.1 for a complete list of investigated genes and their forward and reverse primers). All primers were designed to generate primer products between 100-120 base pairs.

2.6 Quantitative Real-Time Polymerase Chain Reaction

2.6.1 Design and Products

Expression data was collected using quantitative RT-PCR using PerfeCTa® SYBR® Green SuperMix, Low ROX™ (Quanta Biosciences catalog No. 95056). As there appears to be no consensus regarding endogenous control selection, RpL32 was chosen, which according to

the qRT-PCR data was expressed at expressed at very similar levels in all samples (species/ tissues). A total of 10ng/sample of RNA was extracted as described above from pools of 20 tissues in each category for both parental species and their F1 reciprocal hybrids. cDNA was produced using the Quanta Bioscience qScript™ cDNA Synthesis Kit (catalog No. 95047) using the manufacturer's protocol. To assay for genomic DNA and reaction contamination, a blank (DEPC-treated water) was used as a negative control. No amplification was observed in any of the negative control runs. The Stratagene MX3000P® RT-PCR system and the instrument default cycling conditions were used. Each RNA sample was assayed twice for every mRNA in separate RT reactions. The threshold cycle (CT) ratios between the target RNA and the endogenous control were calculated.

2.6.2 Normalization of Expression Data

Several variables need to be controlled for in gene-expression analysis, such as the amount of starting material and differences between tissues in overall transcriptional activity. Various strategies have been applied to normalize these variations. Under controlled conditions of reproducible extraction of good-quality RNA, the gene transcript number is ideally standardized to the number of cells, but accurate enumeration of cells is often precluded, for example when starting with solid tissue. Another frequently applied normalization scalar is the RNA mass quantity, especially in northern blot analysis. The quality of RNA and related efficiency of the enzymatic reactions are not taken into account. Moreover, in some instances it is impossible to quantify this parameter, for example, when only minimal amounts of RNA are available from microdissected tissues. Probably the strongest argument against the use of total

RNA mass for normalization is the fact that it consists predominately of rRNA molecules, and is not always representative of the mRNA fraction. Further drawbacks to the use of 18S or 28S rRNA molecules as standards are their absence in purified target mRNA transcripts. The latter makes it difficult to accurately subtract the baseline value in real-time RT-PCR data analysis.

To date, internal control genes are most frequently used to normalize the mRNA fraction. This internal control - often referred to as a housekeeping gene - should not vary in the tissues or cells under investigation, or in response to experimental treatment. However, many studies make use of these constitutively expressed control genes without proper validation of their presumed stability of expression.

It is generally accepted that gene-expression levels should be normalized by a carefully selected stable internal control gene. However, to validate the presumed stable expression of a given control gene, prior knowledge of a reliable measure to normalize this gene in order to remove any nonspecific variation is required.

The first step of normalization was the quantification of my collected RNA. The collected tissues were treated with a RNA extraction technique using Trizol. The RNA integrity was assayed through a Bioanalyzer (Agilent RNA 6000) using 1 µl aliquot of each tissue sample and the given concentration was recorded in nanograms per microliter (µL). To equalize the relative RNA concentrations the lowest value was divided by all other values (i.e. 60/60 ng/µl and 60/73.3 ng/µl etc).

2.7 Analysis of qRT-PCR Gene Expression Data

To analyze results of quantitative real-time PCR (qRT-PCR) a standard curve was generated for each gene that was under investigation. A standard curve was designed by creating a dilution series: undiluted, 1:10, 1:100, 1:1000. A pool of the tissues were used for each standard curve and two samples of RNA were plotted for each dilution (Figure 2.2). An equation was determined for each gene based on the generated slope (m) and y-intercept data. This was then used to calculate the Log10 Input Amount (Ct value - y-intercept/slope) and the Input Amount (remove the log: POWER(10, [Log10 Input Amount]). This input amount was established for all genes and for each tissue and species type (the same was done for the HKG) for each technical replicate. The final input amount was established by determining the mean of the two input amounts for each technical replicate. As stated above, the gene of interest (GOI) mean input amount was divided by the HKG mean input amount, giving the GOI/HKG ratio, which was used as the relative expression levels.

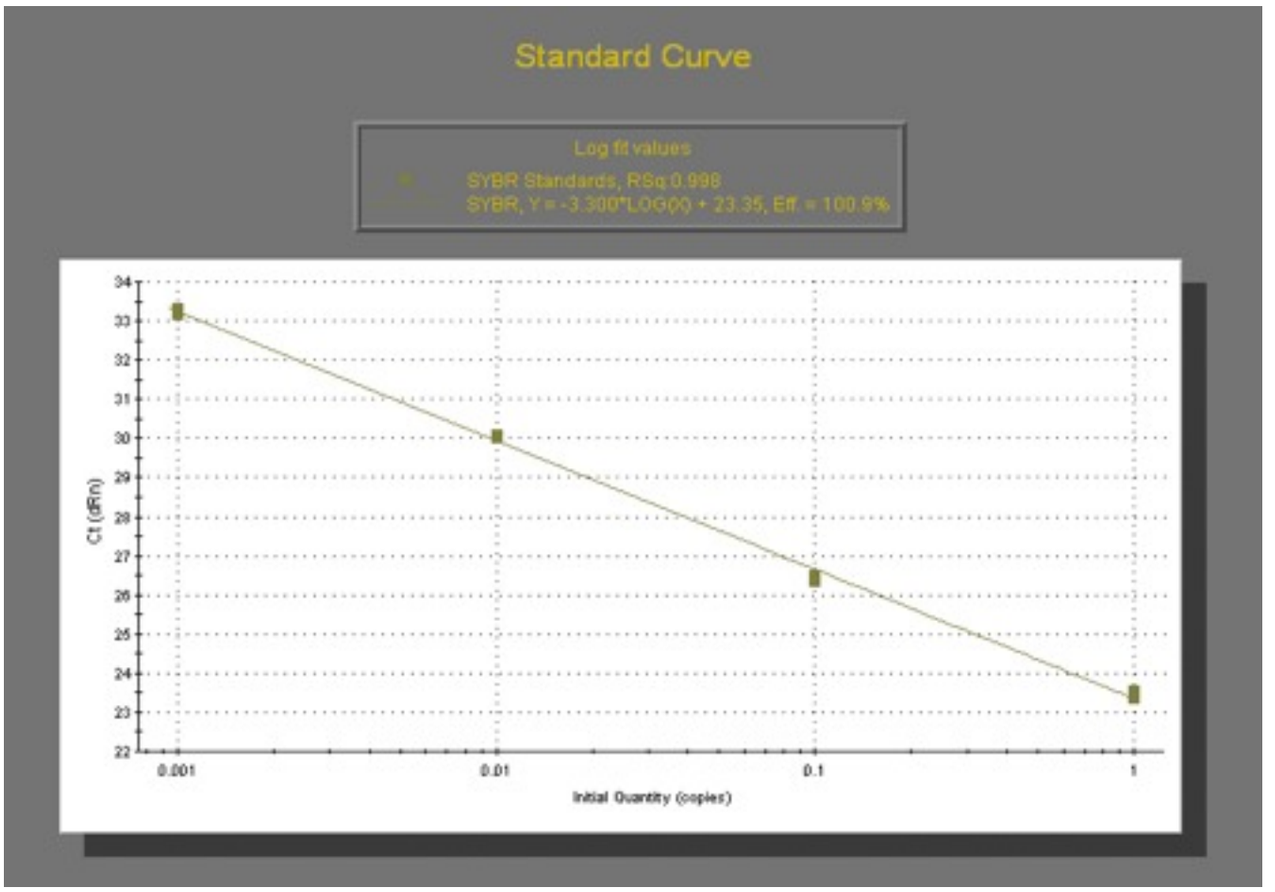


Figure 2.2: Standard Curve — A standard curve is used to quantitate alterations in mRNA levels using qRT-PCR. An ideal curve has both technical replicates fits along a straight line and has a correlation coefficient of 1.0. In the example above "RSq" is 0.998. Once the correlation coefficient has been determined to be sufficient the slope and y-intercept are recorded for each gene (individually) and used to calculate the input amount.

2.8 Biological Replicates

Two biological replicates were generated within the present study for all samples types — the testis, ovary and male and female head tissues within the pure species and reciprocal interspecific crosses for all 40 genes of interest. Upon initial review, the biological replicates were presumed to possess extreme variation in expression data. As a consequence, it was decided that both could not be used to described the observed results, as they were too variant in their respective data sets. The biological replicate that was used in the current study was chosen in response to the following criteria: First, the biological replicate that possessed the least amount of extreme values (outliers) relative to their respective data sets. Second, upon review of the values generated for the internal control gene (RpL32), the biological replicate that demonstrated the least amount of variation between sample types for this gene. As a result, one biological replicate met the above criteria more closely than the other and, therefore, was chosen.

2.9 Tissue Morphology

A digital camera (Nikon SM21000) connected to a dissecting microscope and a computer were used to capture images of the parental and their F1 reciprocal hybrid tissues. EOS™ software package was used for image capturing. Each dissected tissue type — testis, ovaries and male and female head tissues — was captured for both F1 reciprocal hybrids.

RESULTS

3.1 Conservation of Gene Distribution Among the *Drosophila* Triad Species

With advancements in technology the ability to scan and analyze species genomes was revolutionizing to the field of biology. Yet, uncovering the biological relevance and critical information from collected sequence data have proven to be an ongoing process with much to be investigated and explored.

In this study, gene expression profiles for 40 genes were analyzed through quantitative real-time PCR across four tissues types among two parental species, *Drosophila simulans* and *D. sechellia*, and their two reciprocal hybrids. The selected genes were chosen from a collection of genes determined to be misexpressed in the genetic background of *D. melanogaster.* They were annotated according to their presence in various expressed sequence tag (EST) libraries and were assigned specific tissue categories based on *D. melanogaster* ESTs. One of the first undertakings of this study was to compare the *D. melanogaster* EST data with the observed expression profiles determined in this study of the closely related sibling species, as a means to determine if gene expression patterns are conserved between closely related species.

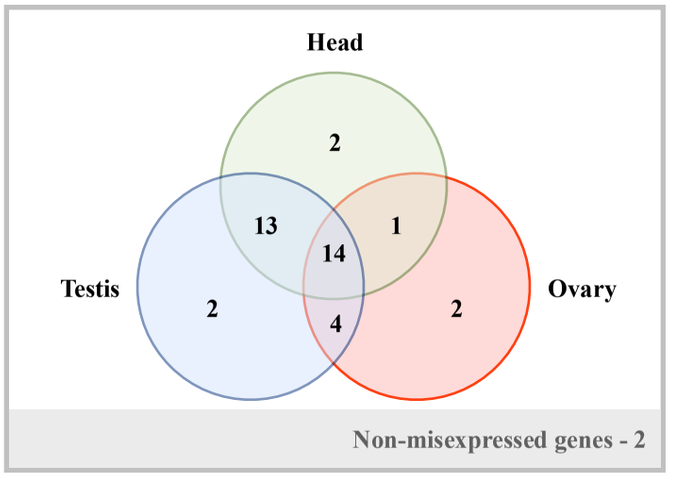
3.1.1 Variation in Tissue-Specific Distribution of Gene Expression

Expressed sequence tags (ESTs) have been created for the *D. melanogaster* species allowing the ability to assign genes of interest to specific tissues. Based on these ESTs, previous research investigating gene expression had employed these sequence tags to classify a collection of genes demonstrating biased expression in a particular tissue or group of tissues by setting an

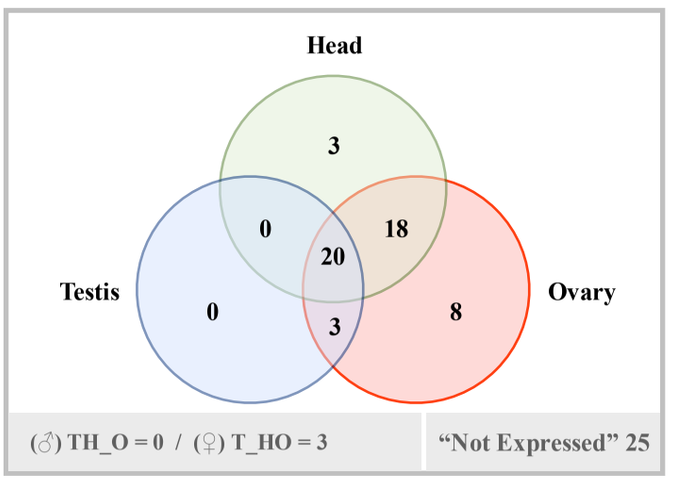
arbitrary cut-off of three ESTs per tissue or percent of the genome (see Haerty et al. 2007 and Parisi et al. 2004, repsectively). For the species of interest within this study, such ESTs do not exist, and as a consequence, genes were selected for this study based on *D. melanogaster* data. Therefore, when comparing gene distribution throughout the investigated tissues, *D. melanogaster* data was determined in accordance with its EST data, while both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing an expression magnitude cut-off above 1.0 of the logarithmic GOI/HKG ratio (see Supplementary Figure §.3.1a).

3.1.1.1 Gene Distribution Among *Drosophila* Species: A 'Global' Analysis

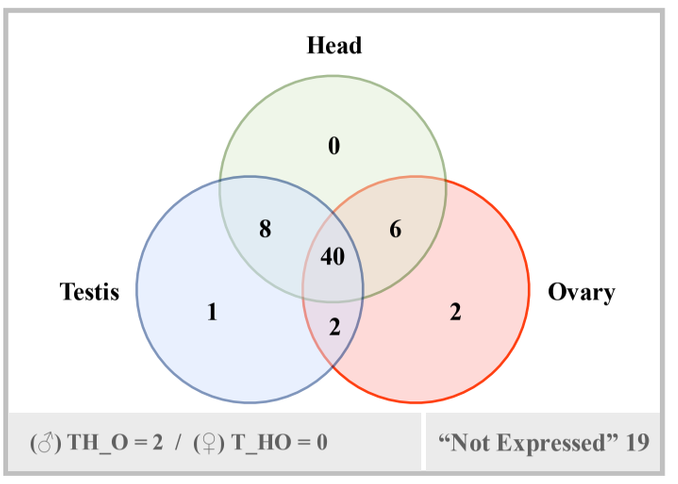
Determined within this study, a 'global' analysis of gene distribution — the overall trend in the number of genes expressed within each tissue or mixed-tissue grouping — was conducted and a number of patterns emerged (Figure 3.1a,b and c). In this study, one of the goals is to observe genes which are expressed in multiple tissue types and determine whether or not these genes are expressed similarly in these tissues. Therefore, out of the 40 genes selected, genes found to be expressed in all three tissue types (head, ovary and testis) designated as (HOT), were the focus having a total of 14 genes in this category. Another significant mixed tissue category, consisting of 13 genes, possessed genes which had ESTs in both the head and testis (TH). This grouping was of particular interest as a means to test the evolution of reproductive (testis) genes relative to 'nonreproductive' (head) genes. When referring to the literature, a number of studies have noted that the testis transcriptome possesses a greater extent of rapidly evolving genes when compared to the head tissue (Coulthart and Singh 1988; Civetta and Singh 1995 and Haerty et al.



[a] Gene assignment based on ***D. melanogaster*** EST data adapted from Artieri and Singh, (2009); Haerty and Singh, (2006) and Parisi et al. (2004).



[b] Gene assignment based on both male and female ***D. sechellia*** expression data.



[c] Gene assignment based on both male and female ***D. simulans*** expression data.

Figure 3.1: Venn diagram representation of gene distribution in [a] *D. melanogaster*, [b] *D. sechellia* and [c] *D. simulans*. Each circle depicts a specific-tissue category: Head (green), Testis (blue), and Ovary (red). The mixed-tissue genes are depicted within their respective overlapping sections of the diagram. Both male and female gene expression was illustrated in one venn diagram for both *D. sechellia* and *D. simulans*, thus having a total of 80 genes represented (the 40 genes of interest for each sex).

2007). Furthermore, the rates of evolution among these 'testis-biased' genes exceed rates found in rapidly evolving 'head-biased' genes (Jagadeeshan and Singh 2005 and Haerty et al. 2007). The remaining genes were selected as follows: 4 ovary and testis (OT) genes, 1 ovary and head (OH) gene, and 2 genes for the head-, ovary- and testis-biased categories, respectively. The last two genes were ***not*** found to be misexpressed in any of the above tissues when looking at the expression of the hybrid cross between *D. melanogaster* and *D. simulans*, and was deemed a 'non-misexpressed' gene (Haerty and Singh 2006) (see Figure 3.1a).

The parental gene expression data among the *Drosophila* triad illustrated greater similarity between *D. melanogaster* and *D. simulans* species by means of tissue expression distribution for the genes of interest (Figure 3.1 a and c). Most interesting, the majority of the observed genes fell within the HOT and TH tissue categories, having 40 (42 including the male specific HOT category) and 8 genes, respectively. Similarly, the gene distribution within *D. melanogaster* had 14 HOT-specific and 13 TH-specific genes. The remaining tissue categories, while not completely equivalent, demonstrated comparable gene numbers — Head: mel (2), sim (0); Testis: mel (2), sim (1); Ovary: mel (2), sim (2); OT mel (4), sim (2); OH mel (1), sim (6).

Conversely, the other sibling species, *D. sechellia*, exhibited a clear departure in tissue expression distribution from the other two species (Figure 3.1b). While there were found to be 20 (23 including the female specific HOT category) genes which occurred in the HOT category, there was a complete lack of genes within the TH category. Interestingly, the OH category possessed 18 genes, a notable increase from *D. melanogaster* having only 1 gene. Moreover, there was a greater number of genes, a total of 8, found in the ovary tissue category. The remaining tissue categories were as follows — Head: mel (2), sec (3); Testis: mel (2), sec (0); OT

mel (4), sec (3).

Between the sibling species, one would expect greater similarities in gene expression than either with *D. melanogaster*, due to their more recent divergence of ~ 250,000 years ago over that of both species from *D. melanogaster* ~ 0.5 - 2.5 MYA (Caccone et al. 1988; Catron and Noor 2008; McDermott and Kliman 2008). This was found in a study conducted by Artieri and Singh (2009) where they investigated an increase in regulatory conservation throughout metamorphosis among the species of the *D. melanogaster* subgroup. During the larval, late pupal and adult stages there were fewer differentially expressed genes and it was only in the early pupal stage that these species appeared to have significantly more genes differentially expressed when compared to either species with *D. melanogaster* (see Artieri and Singh 2009 — Figure 1). Despite this, however, within this study it was found that these sibling species demonstrated a greater extent of differential gene expression than *D. simulans* experienced when compared to *D. melanogaster*. This finding may have resulted due to a variance in methodology. The study conducted by Artieri and Singh 2009 was carried out using microarray technology, while this study employed qRT-PCR, which has been suggested to be a more sensitive methodology (Catron and Noor 2008).

Thus, looking globally at gene distribution, distinct patterns emerge illustrating a more similar gene distribution between *D. melanogaster* and *D. simulans* and a vastly distinct gene distribution demonstrated by *D. sechellia*. This finding is consistent with previous results in the literature (Caccone et al. 1988; Caccone et al. 1996) which support a more recent divergence between *D. melanogaster* and *D. simulans* through the use of DNA-DNA hybridization experiments using single-copy nuclear DNA(scnDNA), than the divergence of the former and *D.*

sechellia. This could be a consequence of effective population size, were the *D. melanogaster* and *D. simulans* species originating as cosmopolitan human commensal and *D. sechellia* as an island species in the Seychelles (Tsacas and Bachli, 1981 and Lachaise et al. 1986; Coyne and Kreitman 1986).

3.1.1.2 Gene Distribution Among *Drosophila* Species: Gene Specific Analysis

At the global level of analysis, there was found to be fewer genes differentially expressed between *D. simulans* and *D. melanogaster* over that observed between the latter and *D. sechellia.* Furthermore, the sibling species did not display similar tissue distributions across genes of interest. Having noted this, however, while the analysis of global tissue distribution reports the number of genes observed within each tissue category, it neglects to identify the specifics of each gene of interest. By constructing a table listing *D. melanogaster*, *simulans*, and *sechellia* and their specific tissue expression patterns for all 40 genes of interest, one can distinguish specific genes sharing the same tissue expression patterns (Table 3.1). It was found that 30% of genes shared the same tissue expression patterns between *D. simulans* and *D. sechellia*, 15% shared the same tissue expression patterns between *D. simulans* and *D. melanogaster*, 10% shared the same tissue expression patterns between *D. sechellia* and *D. melanogaster*, and 7.5% shared the same tissue expression patterns amongst all three species. The remaining 37.5% of genes lacked any similarities in tissue expression. Therefore, it was found that the sibling species had the highest degree of shared tissue expression followed by *D. melanogaster* and *D. simulans*, which is supported by the divergences suggested in a number of established *Drosophila* phylogenies (Powell 1997; Tamura et al. 2004; Stark et al. 2007; Larracuente et al. 2008).

Table 3.1: Specific gene distribution among *D. melanogaster*, *D. sechellia* and *D. simulans*. Genes are classified into tissue categories: (H) head, (O) ovary, (T) testis, (HOT) head, ovary and testis, (TH) testis and head, (OT) ovary and testis, (OH) ovary and head, and (NE) not expressed based on the 1.0 GOI/HKG ratio cut-off (see result section) — The red box indicates genes which show similar tissue distribution between *D. simulans* and *D. sechellia*.; the blue box indicates genes which show similar tissue distribution between *D. simulans* and *D. melanogaster*; the green box indicates genes which show similar tissue distribution between *D. sechellia* and *D. melanogaster*; the purple box indicates genes which show similar tissue distribution between all species of the three species.

Genes	*D. simulans*	*D. sechellia*	*D. melanogaster*
FBgn0030706	MALE: HOT FEMALE: OT	MALE: NE FEMALE: O	H
FBgn0032702	HOT	HOT	H
FBgn0020392		MALE: OT	O
	HOT	FEMALE: HOT	
FBgn0038964	MALE: NE	HOT	
	FEMALE: O		O
FBgn0032821	TH	HOT	T
FBgn0034532	HOT	MALE: NE FEMALE: O	T
FBgn0038607		MALE: OT	TH
	HOT	FEMALE: HOT	
FBgn0028902	HOT	HOT	TH
FBgn0052112	OH	MALE: OT FEMALE: HOT	TH
FBgn0031853	TH	MALE: NE FEMALE: O	TH
FBgn0002036	MALE: HOT FEMALE: OT	MALE: NE FEMALE: OH	TH
FBgn0042710	HOT	MALE: NE FEMALE: O	TH
FBgn0011273	HOT	OH	TH
FBgn0032973	MALE:NE FEMALE: O	NE	TH
FBgn0038266	NE	NE	TH
FBgn0029947	HOT	NE	TH
FBgn0028579	HOT	MALE: NE FEMALE: O	TH
FBgn0036763	HOT	NE	TH
FBgn0038334	HOT	OH	TH
FBgn0039336	NE	NE	OT
FBgn0031227	HOT	MALE: NE FEMALE: O	OT
FBgn0036710	OH	OH	OT
FBgn0015562	HOT	HOT	OT

Genes	*D. simulans*	*D. sechellia*	*D. melanogaster*
FBgn0035878	HOT	HOT	HOT
FBgn0034689	HOT	HOT	HOT
FBgn0022338	TH	HOT	HOT
FBgn0051108	TH	HOT	HOT
FBgn0034176	NE	HOT	HOT
FBgn0032474	HOT	MALE: NE FEMALE: H	HOT
FBgn0011604	NE	H	HOT
FBgn0015799	NE	NE	HOT
FBgn0032248	HOT	OH	HOT
FBgn0032451	OH	OH	HOT
FBgn0034689	NE	MALE: NE FEMALE: O	HOT
FBgn0039360	NE	NE	HOT
FBgn0037369	HOT	OH	HOT
FBgn0050104	HOT	MALE: OH FEMALE: O	HOT
FBgn0030574	HOT	OH	NMG
FBgn0031988	NE	NE	NMG
FBgn0034158	MALE: T FEMALE: NE	OH	OH

3.2 Males vs. Females: Observed Expression Patterns Across Species

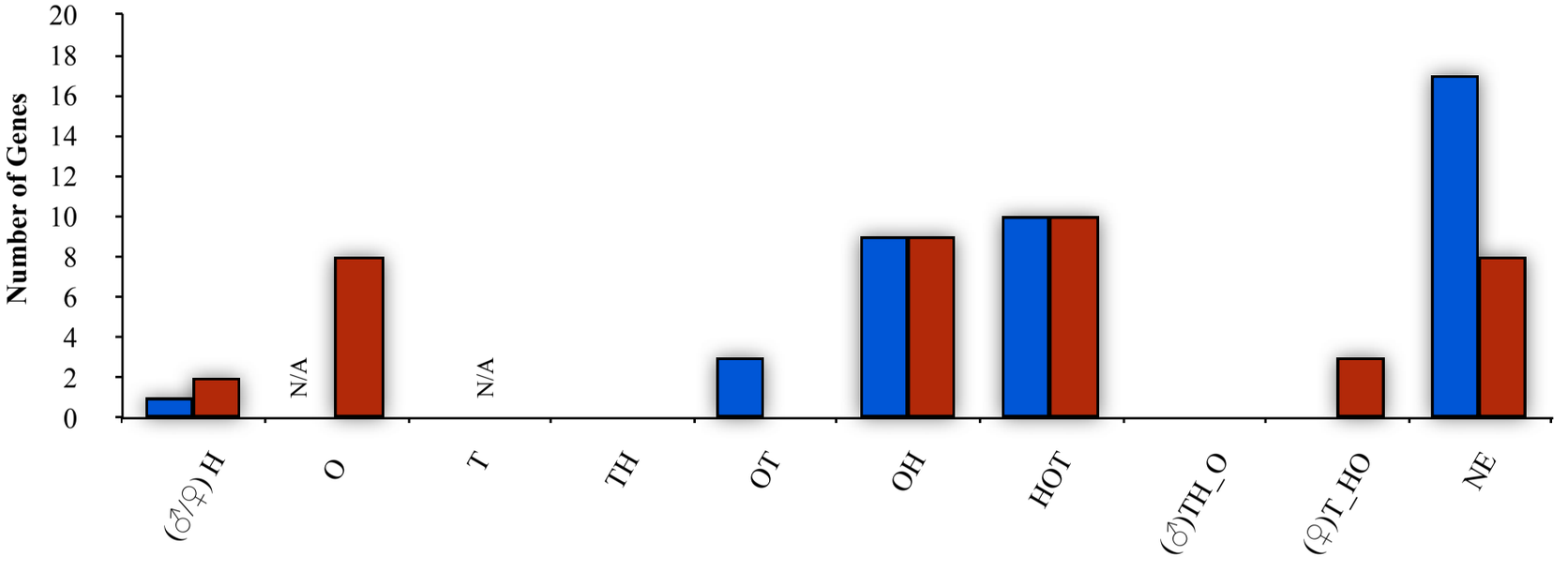
Unique to this study was the focus on both sexes rather than the previously predominate affinity towards the male sex (Andrews et al. 2000; Michalak and Noor 2003; 2004; Haerty and Singh 2006 and Moehring et al. 2007). This has largely been contributed to the discovery of sex-biased genes, with particular emphasis on male-biased genes, demonstrating a faster rate of evolution. This was inferred via higher variation in expression patterns suggesting an experience of a higher rate of positive selection (Coulthart and Singh 1988; Civetta and Singh 1999; Meiklejohn et al. 2003; Michalak and Noor 2003 and Oliver 2003).

However, through the investigation of these rapidly evolving genes (REGs) as well as gene expression variation between sexes and across species, one may be able to infer whether or not one sex may be driving the evolution of the other, via cryptic female choice, male-sex drive or the possibility of a coevolution between the sexes may emerge.

3.2.1 Tissue-Specific Distribution between the Sexes: Parental Species

The parental gene expression data demonstrates a strong correlation in gene distribution between males and females of both parental species, *D. sechellia* and *D. simulans,* among the tissue-specific categories (Figure 3.2). This result, while interesting, does not prove unexpected given the comprehensive analyses revealing that gene expression profiles of parental species appear more similar to one another than either is to that of the hybrid (Ranz et al. 2003; Rifkin et al. 2003 and Ranz et al. 2004). This result is a consequence of compensatory changes which act to ‘eclipse‘ the prevailing interspecific differences between species, however, in the genetic

[a]



[b]

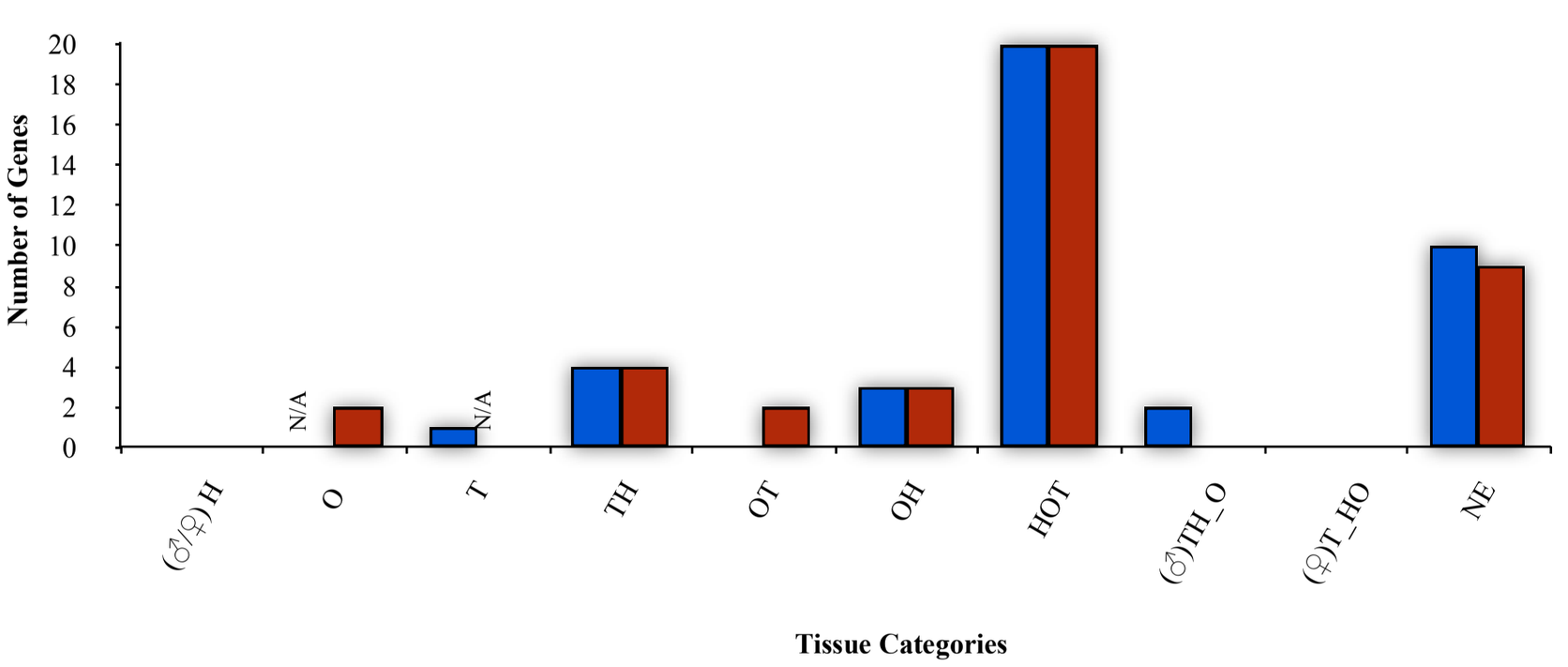


Figure 3.2: Parental species tissue-specific distribution between the sexes of *Drosophila sechellia* **[a]** and *D. simulans* **[b]**. Male and female distribution are depicted in blue and red, respectively, across the tissue categories: (H) male and female head; (O) ovary; (T) testis; (TH) testis and head; (OT) ovary and testis; (OH) ovary and head; (HOT) head, ovary and testis expressed for both sexes; (TH_O) head, ovary and testis for the male sex only; (T_HO) head, ovary and testis for the female sex only; and (NE) not expressed.

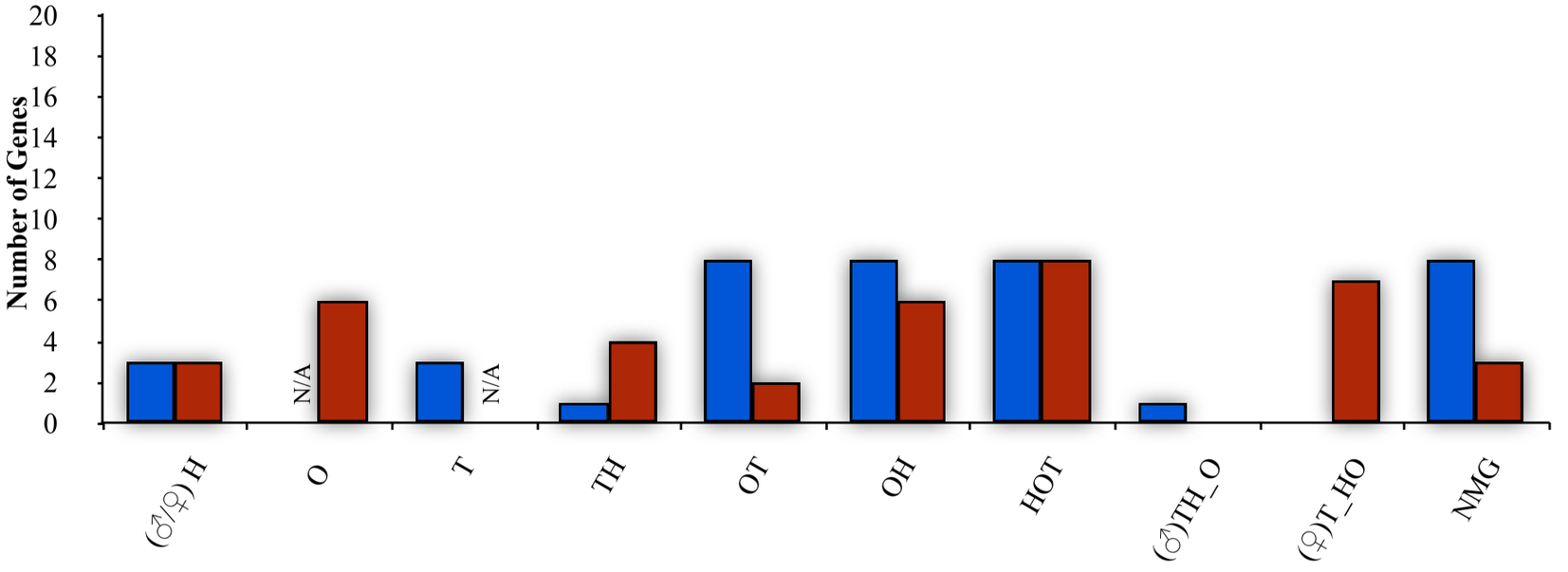
backgrounds of interspecific hybrids, these changes may no longer be concealed as a result of a lack in genetic coadaptation (Dobzhansky 1937; Muller 1942; Orr 1997; Johnson and Porter 2000; Orr and Presgraves 2000; Porter and Johnson 2002).

3.2.2 Tissue-Specific Distribution between the Sexes: Hybrids

A greater opportunity to observe differential gene expression between the sexes resides within the reciprocal hybrids. A large number of studies have suggested ~ 30 - 50% of the *Drosophila* transcriptome are sex-biased — a gene is deemed sex-biased when its level of expression significantly varies between the sexes (Ranz et al. 2003; 2004; Parisi et al. 2003; Ortiz-Barrientos et al. 2007 and Gallach et al. 2011). Furthermore, there has been specific emphasis on male-biased gene expression as a result of: (1) a greater difference in expression than that found in female- and non-sex-biased genes (Haerty et al. 2006) as well as (2) a greater proportion of genes experiencing sex-biased (Ranz et al. 2003; 2004; Parisi et al. 2003; Graveley et al. 2010 and Daines et al. 2011).

Surprisingly, when looking at the male and female tissue-specific gene expression data for both interspecific crosses, unlike previous research, the majority of genes did not show a large extent of variation between the sexes, where the variance was established by a factor of 2 (see Castillo-Davis et al. 2003) (Figure 3.3). For the cross between ♀ *D. simulans* x ♂ *D. sechellia,* genes which experienced differential expression (71/80) displayed variation between the sexes for 33.8% of the genes and a total of 63.4% of genes which displayed parallel expression for both sexes. This is disparate from the study conducted by Castillo-Davis et al. (2003) were they compared the gene expression profiles of adult *D. melanogaster* and *D. simulans* species as a means to reveal the evolution of sex-biased regulation of genes. They found that out of the 2238

[a]



[b]

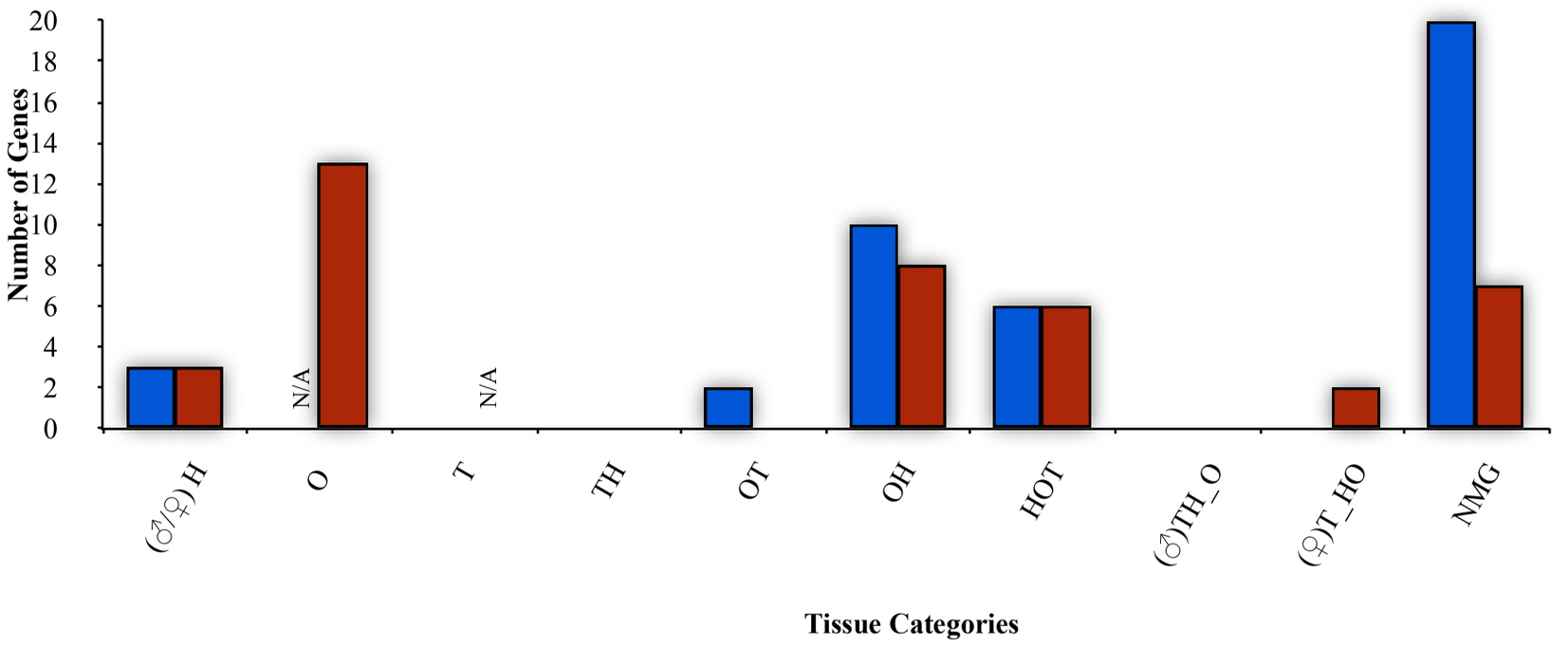


Figure 3.3: Reciprocal hybrid tissue-specific distribution between the sexes of ♀ ***simulans*** **x** ♂ ***sechellia*****[a]** and ♀ ***sechellia*** **x** ♂ ***simulans*** **[b]**. Male and female distribution are depicted in blue and red, respectively, across the tissue categories: (H) male and female head; (O) ovary; (T) testis; (TH) testis and head; (OT) ovary and testis; (OH) ovary and head; (HOT) head, ovary and testis expressed for both sexes; (TH_O) head, ovary and testis for the male sex only; (T_HO) head, ovary and testis for the female sex only; and (NMG) non-misexpressed.

genes surveyed only 380 genes, or 16.6%, of differentially expressed genes displayed parallel differences in expression for both sexes. Moreover, there were 1903 out of the same gene set, or 83.4%, of genes which exhibited sex-specific expression. Similarly, in the reciprocal cross between ♀ *D. sechellia* x ♂ *D. simulans*, it was found for the genes which experienced differential expression (53/80) only 24.5% and 71.7% of genes exhibiting variation between the sexes and parallel differences for both sexes, respectively.

Therefore, contrary to previous studies, the results in here have the majority of differentially expressed genes demonstrating parallel expression for both sexes and fewer genes which have observed variation between the sexes.

3.2.3 Comparison: Male and Female Head Expression

It is known that males and females utilize the same basic set of genes and achieve dimorphism at the morphological and molecular level. Thus, each sex possesses either quantitative difference in the level of gene expression and/or undergo sexual selection lending to the evolution of genes with male- and female-specific functions (Connallon and Clark 2011). A number of studies have been conducted investigating sex-biased regulation among *Drosophila* species (Arbeitman et al. 2002 and Parisi et al. 2003) and have found this regulation to be the result of expression within the germ-line tissues. Specifically, it was suggested that transcription in testis and ovaries are responsible for the majority of sex-biased expression (Meiklejohn et al. 2003). To validate this theory, dissected testes and ovaries, along with gonadectomized males and females of the species *D. melanogaster* and *D. simulans* were analyzed and genes were classified based on their variation in expression between testes, ovaries and both male and

female somatic tissues (Meiklejohn et al. 2003; Parisi et al. 2003; 2004 and Ranz et al. 2003). It was determined that male-biased expression experienced the greatest degree in variation over both female-biased and non-sex-biased genes.

In this study, the 'somatic tissues' — both the male and female heads, were dissected and their expression was compared between the sexes (Figure 3.4). The results followed what was to be anticipated, revealing a close correlation between the head tissues across both pure species and their reciprocal hybrids.

3.2.4 Comparison: Testes and Ovaries Expression

Upon the comparison between the testis and ovary tissues, an unexpected result emerged. For pure species, it was found that the ovary tissue had a higher level of expression over the testis tissue, which had a relatively low level of gene expression (Figure 3.5a and b). For the reciprocal hybrids, they exhibit varied expression between the gonadal tissues (Figure 3.5c and d). The cross between ♀ *D. simulans* x ♂ *D. sechellia* demonstrates a higher level of expression in the testis over ovary tissues, while the reciprocal cross ♀ *D. sechellia* x ♂ *D. simulans,* exhibits the same patterns observed in the two pure species.

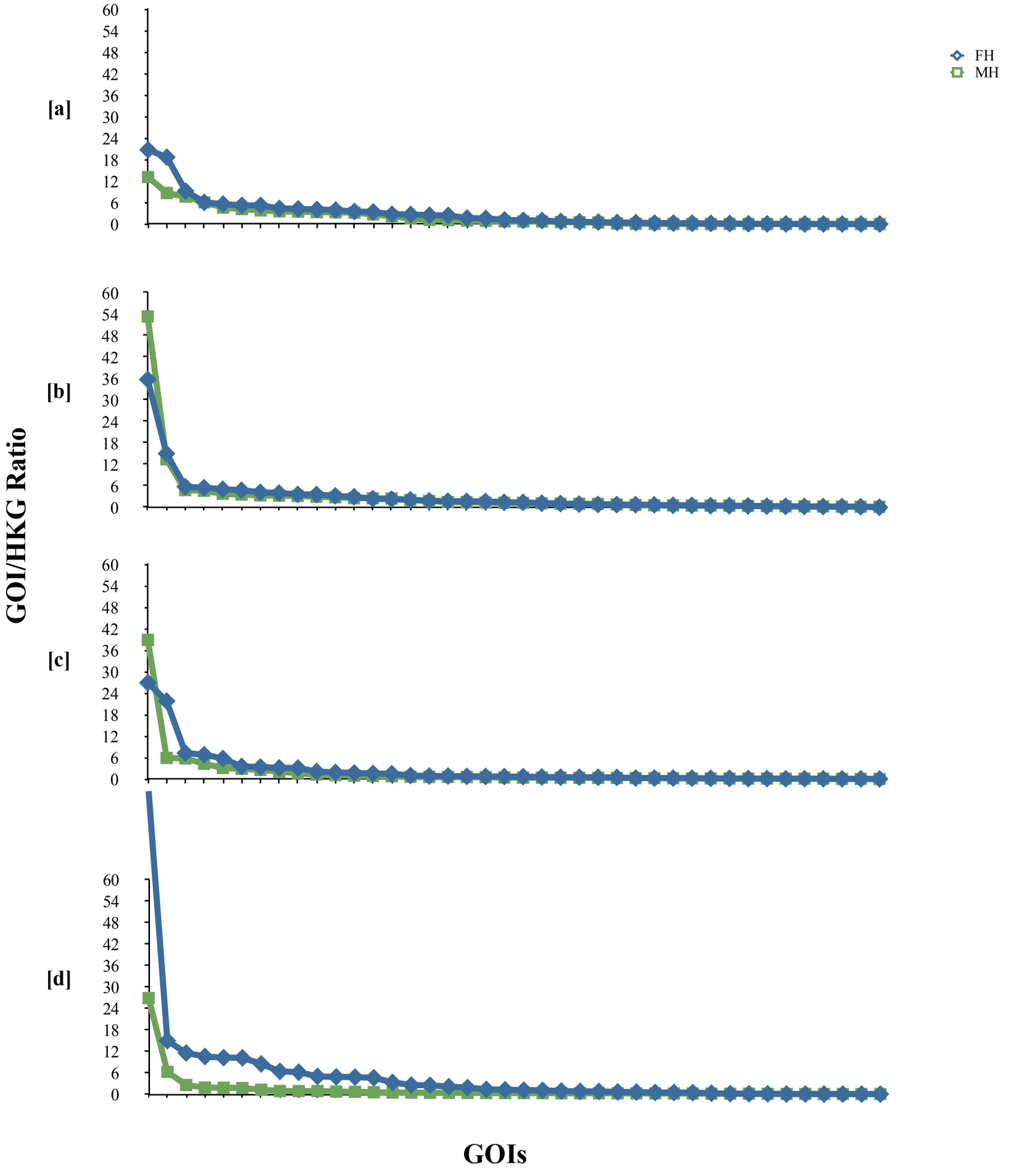


Figure 3.4: Male and female head tissue expression of both parental species: *Drosophila sechellia* [a] and *D. simulans* [b] as well as their reciprocal hybrids*:* ♀ ***simulans*** **x** ♂ ***sechellia*** [c] and ♀ ***sechellia*** **x** ♂ ***simulans*** [d]. Male and female head tissues are depicted in square green and diamond blue data points, respectively. These graphical representations act to compare the relative gene expression of all genes of interests between both the male and female head tissues.

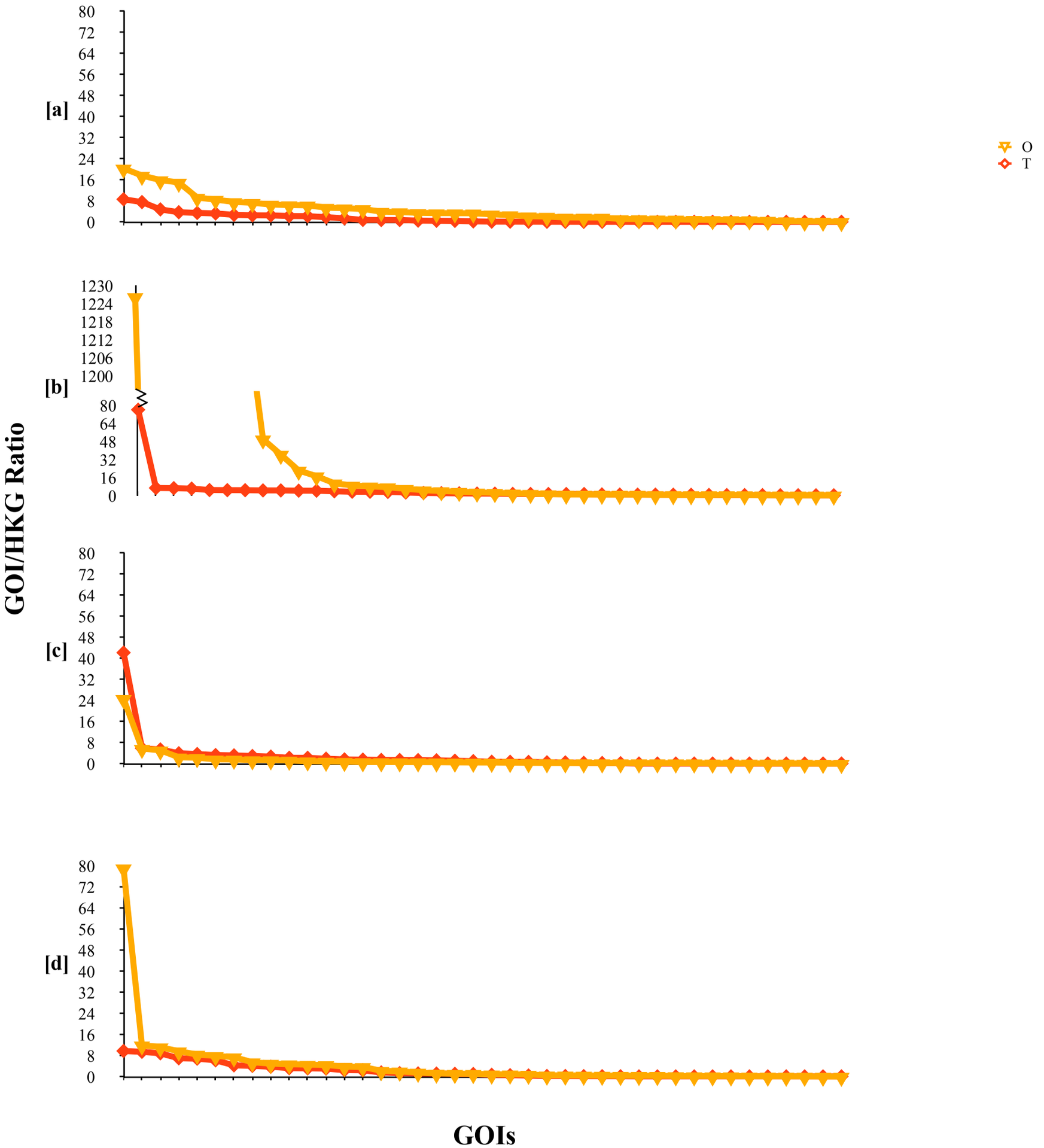


Figure 3.5: Ovary and testis tissue expression of both parental species: *Drosophila sechellia* [a] and *D. simulans* [b] as well as their reciprocal hybrids*:* ♀ ***simulans*** **x** ♂ ***sechellia*** [c] and ♀ ***sechellia*** **x** ♂ ***simulans*** [d]. The ovary and testis tissues are depicted in square yellow and diamond red data points, respectively. These graphical representations act to compare the relative gene expression of all genes of interests between both the male and female gonadal tissue.

3.3 Variation between Reciprocal Hybrids

To date, the predominance in the literature, for the most part, does not focus on the analysis of differential expression between reciprocal hybrids between species. The majority of studies discuss two or more pure species and their (one-way) F1 hybrid (Ranz et al. 2004; Haerty et al. 2007; Cantron and Noor 2008 and Artieri et al. 2009). Yet, a paper by Dobzhansky (1935) indicates that reciprocal crosses can frequently produce diverse results which can lend to insight into such phenomenon as maternal effect, an area of study largely overlooked.

3.3.1 Observed Patterns of Expression Data

As a means to observe the relative gene expression levels across the testes, ovaries and both male and female head tissues for all the genes of interest a graphical representation was plotted from highest to lowest GOI/HKG ratios (Figure 3.6a and b). The graph illustrating relative gene expression for ♀ *sechellia* x ♂ *simulans* demonstrated a correlation between the male-specific (testis and male head) tissues and female-specific (ovary and female head) tissues, respectively (Figure 3.6a). This pattern of relative gene expression was similarly observed in the reciprocal cross, ♀ *simulans* x ♂ *sechellia* (Figure 3.6b). One notable difference relates to the magnitude of relative expression for male and female tissue types between hybrids. The cross between ♀ *sechellia* x ♂ *simulans* demonstrated male-specific tissues to have a higher relative level of expression over that observed in female-specific tissues. The reverse was observed in the reciprocal cross, showing the female-specific tissues having a notably higher relative level of expression compared to the male-specific tissues.

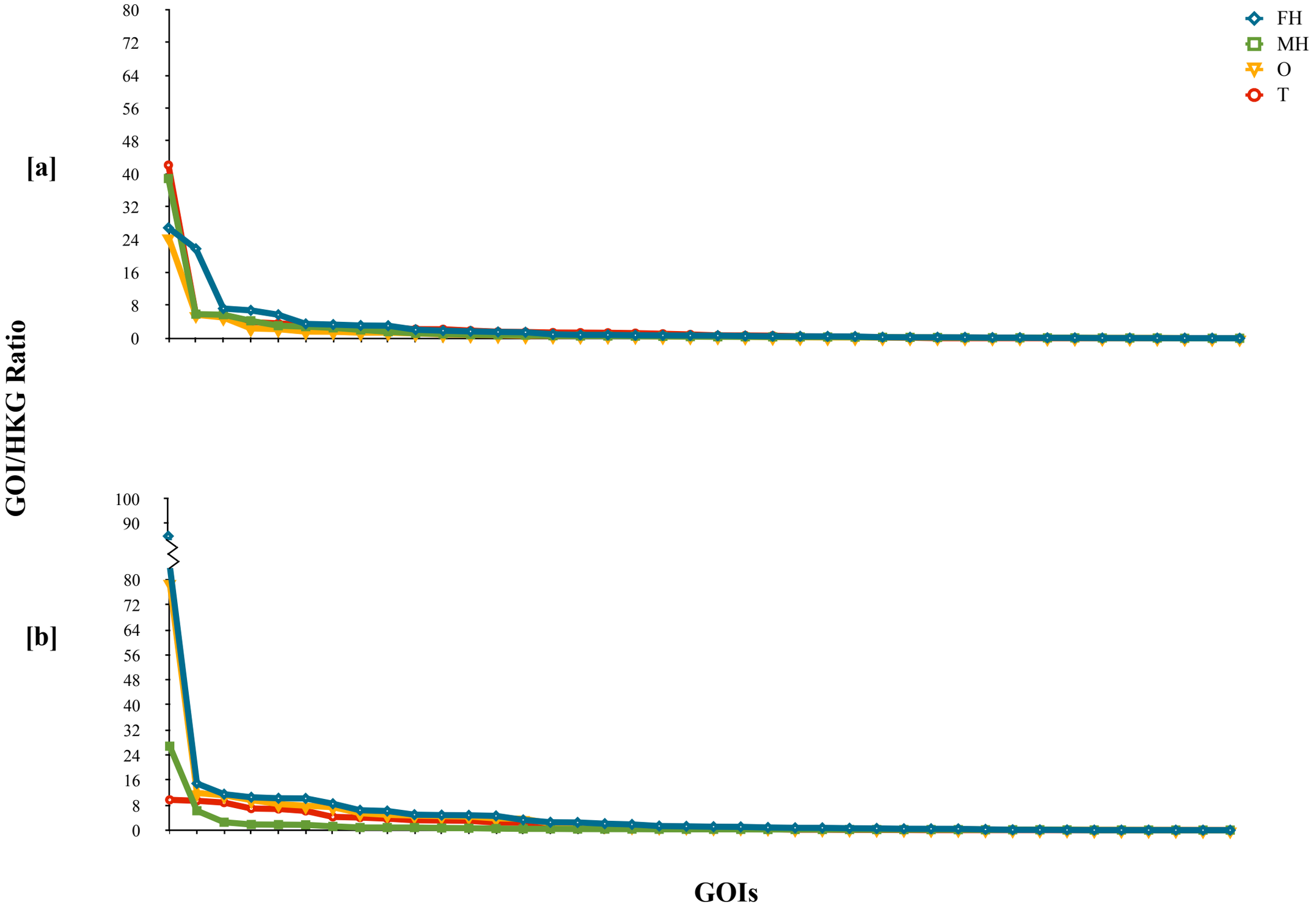
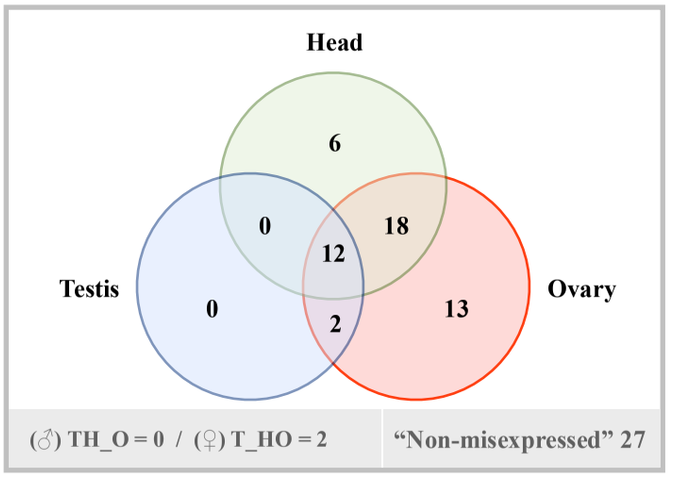


Figure 3.6: Graphical representation of the relative gene expression levels across all four tissue types — male and female head, ovary and testis tissues — for all 40 genes of interest plotted from highest to lowest GOI/HKG ratio data. Graph illustrating gene expression for ♀ ***sechellia*** **x** ♂ ***simulans*** [a] as well as ♀ ***simulans*** **x** ♂ ***sechellia*** [b].

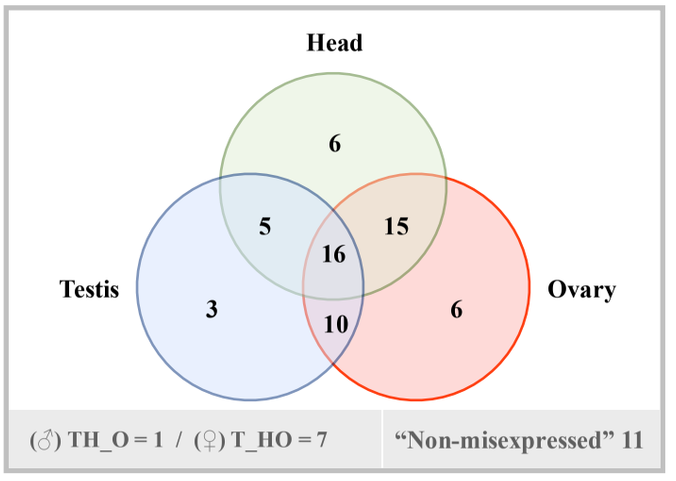
Specifically, looking at the gene distribution across tissue-specific categories, variant results were observed between the reciprocal crosses (Figure 3.7). In the cross between ♀ *sechellia* x ♂ *simulans,* the majority of genes are concentrated amongst tissue categories central to female-specific tissues (those tissue categories involving ovaries), with the exception of the OT category having only 2 genes. The remaining tissue categories possessed few, if any, genes — Head: 6; OT: 2; Testis: 0; TH: 0 (Figure 3.7a). Conversely, while a large number of genes congregate within the female-specific tissue categories, several genes occupy tissue categories associated with male-specificity having the tissue categories as follows — Head: 6; Testis: 3; Ovary: 6; HOT: 16; OH:15; OT: 10 and TH: 5 (Figure 3.7b). These findings could be attributed to a strong pressure exerted by the maternal genome on their respective crosses. This can be further supported by the observation that suggests the reciprocal hybrids having tissue-specific gene distribution with a large degree of similarity to that found in their maternal parental species, respectively ([Figure3.2a — Figure3.3b] and [Figure3.2b — Figure3.3a]).

3.3.2 Observed Maternal Effect

The examination of the phenomenon deemed ‘maternal effect’ dates back to the 1930s, with two original papers by Dobzhansky (1935) and Dobzhansky and Sturtevant (1935) noting its relative evolutionary significance. Despite these attempts to highlight its significance, maternal effect received little attention until the late 1980s (Mousseau et al. 2009). This fact may be accredited to the prevailing argument favouring cytoplasmatic inheritance, which has since been discredited (Dobzhansky 1935). In a paper put forth by Roach and Wulff (1987), gave rise to the “first [empircial] support for the near-ubiquitous role of maternal effects in



[a] Gene assignment based on expression data from the hybrid cross *D. sechellia* x *D. simulans*.



[b] Gene assignment based on expression data from the hybrid cross *D. simulans* x *D. sechellia*.

Figure 3.7: Venn diagram representation of gene distribution in [a] *D. sechellia x D. simulans*, [b] *D. simulans x D. sechellia*. Each circle depicts a specific-tissue category: Head (green), Testis (blue), and Ovary (red). The mixed-tissue genes are depicted within their respective overlapping sections of the diagram. Both male and female gene expression was illustrated in one venn diagram for both hybrids, thus having a total of 80 genes represented (the 40 genes of interest for each sex).

plants...” (Mousseau et al. 2009). This, along with a number of additional studies which followed, outlined the potential for maternal effect to act as an evolutionary response to selection.

To investigate the degree of maternal effect, the relative expression levels for genes which were classified as ‘H’, ‘O’ and ‘OH’ were plotted for both the reciprocal hybrids (Figure 3.8). For the cross between ♀ *sechellia* x ♂ *simulans*, the ‘H’ tissues had the fewest number of genes (4) and the ‘O’ tissues had 13 genes, which suggest a difference between the tissues by a factor of 3. The ‘OH’ tissue category had 16 genes and was not notably different from the ‘O’ tissue category. However, there was an observed difference between the ‘OH’ and ‘H’ tissue categories by a factor of 4. For the cross between ♀ *simulans* x ♂ *sechellia*, the ‘H’ and ‘O’ tissues had the fewest number of genes, both having 7 genes respectively, and the ‘OH’ tissue category had 22 genes. There was an observed difference between the ‘OH’ and the ‘H’ and ‘O’ tissue categories by a factor of 3. Between the reciprocal hybrids, none of the tissue categories had large degrees of variance between them.

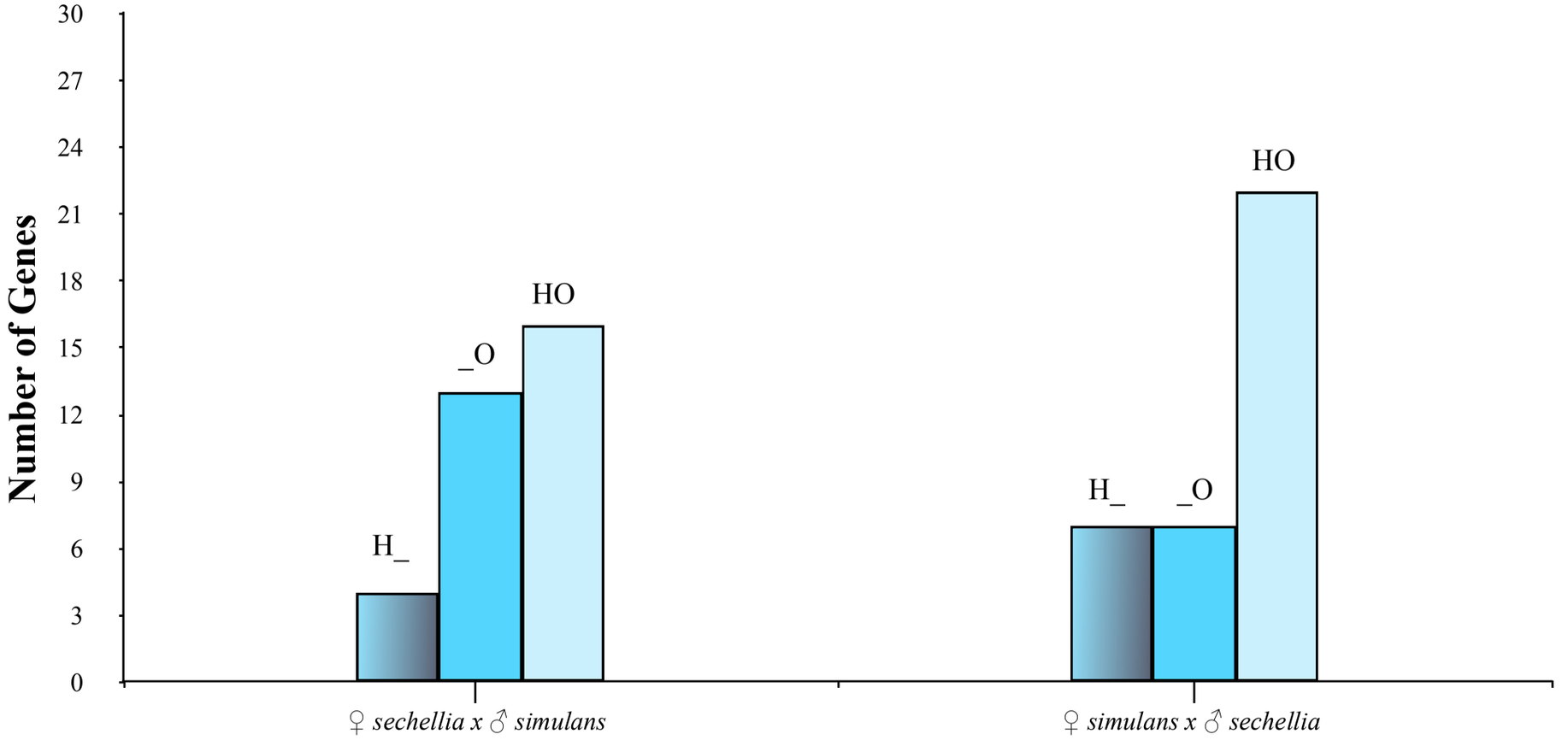


Figure 3.8: Maternal effect illustrated via tissue-specific modulation between the reciprocal crosses, ♀ ***sechellia*** **x** ♂ ***simulans*** and ♀ ***simulans*** **x** ♂ ***sechellia*** for female head and ovary tissues — (H_) only the head tissue is misexpressed; (_O) only the ovary tissue is misexpressed; (HO) head and ovary are both misexpressed.

3.4 Sex vs. Non-Sex Genes: Extent of Gene Misexpression

It has been previously determined by a number of studies (Ranz et al. 2003, 2004) that ~ 30 - 50% of genes in the *Drosophila* transcriptome exhibit sex-biased expression. Many of these genes have expression found in reproductive organs (testis and ovary) (Arbeitman et al. 2002 and Parisi et al. 2003). Therefore, it is important to look at the variation observed in this study between the isolated testis, ovary, and female and male head tissues between both pure species and their reciprocal hybrids.

3.4.1 Tissue Modulation

In a means to identify the varying extent of gene expression between sex (testis and ovary) and non-sex (male and female head) genes, the observed tissue-specific modulation was plotted between the reciprocal hybrids for all tissues, with male head and testis as one comparison and female head and ovary as the other (Figure 3.9). Between the male head and testis gene expression, the cross between ♀ *sechellia* x ♂ *simulans* was found to have 35% of its gene expression displaying tissue-specific modulation, specifically 30% in the male head and 5% in the testis. In the reciprocal cross, ♀ *simulans* x ♂ *sechellia*, 55% of gene expression demonstrated tissue-specific modulation, with an equal distribution of 27.5% for both male head and testis. The unexpected testis expression finding in the former hybrid may be the by-product of the difficult nature of the hybridization between the female *D. sechellia* with the male *D. simulans*. The tissue morphology (see below) between the hybrids reveals a slightly more atrophic appearance of the this cross relative to the testis of the other hybrid and the pure species. In both crosses, the hybrid males exhibited a higher proportion of misexpressed genes in the head

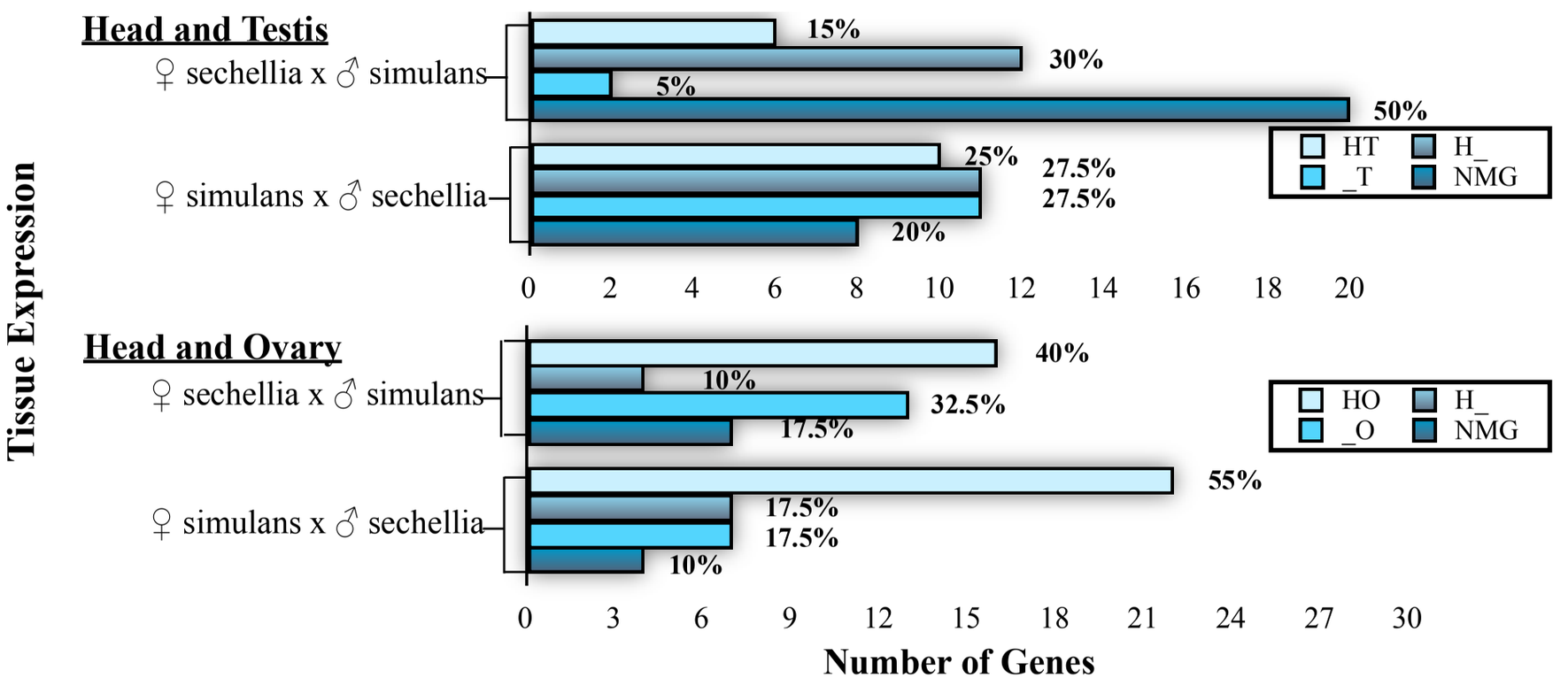


Figure 3.9: Tissue-specific modulation of gene expression between the reciprocal crosses, ♀ ***sechellia*** **x** ♂ ***simulans*** and ♀ ***simulans*** **x** ♂ ***sechellia*** for [a] male head and testis tissues — (HT) head and testis are both misexpressed; (NMG) within a gene both tissues demonstrate non-misexpression; (H_) only the head tissue is misexpressed; (_T) only the testis tissue is misexpressed. The female was also investigated via the [b] female head and ovary tissues — (HO) head and ovary are both misexpressed; (_O) only the ovary tissue is misexpressed.

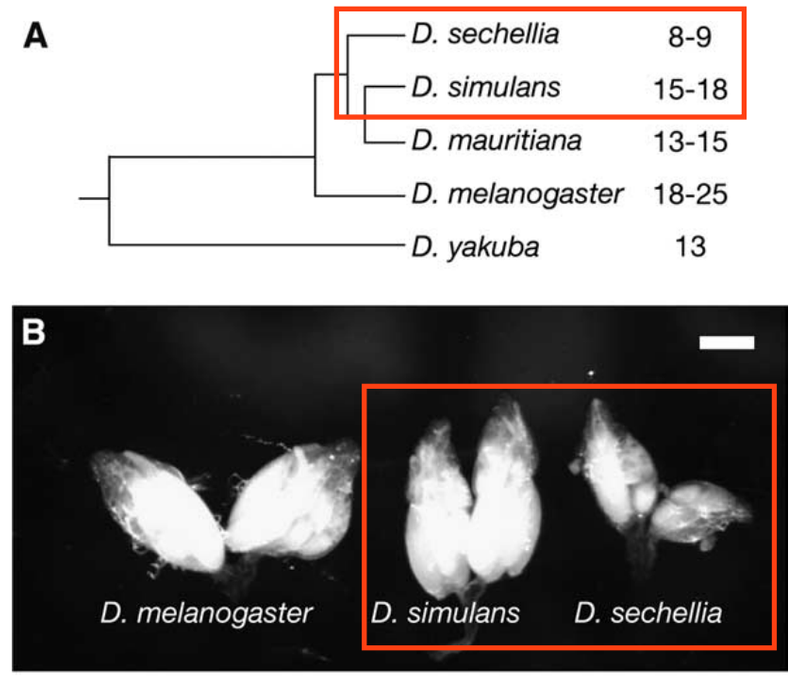
tissue.

For the comparison between the female head and ovary gene expression, the cross between ♀ *sechellia* x ♂ *simulans* out of the 42.5% of genes demonstrating tissue-specific modulation, 10% were found in the female head and 32.5% in the ovary, while in the reciprocal cross, ♀ *simulans* x ♂ *sechellia*, out of the 35% of genes showing tissue-specific modulation, both the head and ovary tissues exhibited 17.5% equal expression. This result was similarly found in a study looking at synonomous (d*S*) and nonsynonomus (d*N*) substitutions of rapidly evolving genes (REGs) between *D. melanogaster* and *D. simulans,* where they found that there was no the d*N* estimates of both ovary and head significant difference between REGs (d*Novary*=0.0147 ± 0.006, d*Nhead*=0.0149 ± 0.007) (Jagadeeshan and Singh 2005). Thus, hybrid females display a higher proportion of misexpressed genes in head and ovary tissues.

3.5 Tissue Morphology

The gonadal tissue morphology between the reciprocal hybrids, ♀ *simulans* x ♂ *sechellia* and ♀ *sechellia* x ♂ *simulans*, gives insight into the workings of the relative difficulty of the crosses. The cross between ♀ *simulans* x ♂ *sechellia* has been found to be a 'successful cross' and the ovary tissue of the females of this cross resemble that of their maternal parent (Figure 3.10a and b). The testis of this cross were found to have both properly developed testis and seminal vesicles (refer to the paper by Zeng and Singh 1993). This was similarly observed in a paper by Orgogozo et al. (2006) were the cross between ♀ *simulans* x ♂ *sechellia* had a closer resemblance in ovariole number to *D. simulans* (Figure 3.10a). Conversely, when examining the morphology of the ovary and testis of the reciprocal cross, ♀ *sechellia* x ♂ *simulans*, the ovary, while not completely atrophied, appear less affluent than those observed in the other cross

[a]



[b]

[c]

Figure 3.10: Gonadal Tissue Morphology — **[a]** Adapted from Orgogozo, Broman and Stern (2006) Figure 1. Variation in ovariole number in *Drosophila simulans* and *D. sechellia*. (A) Phylogeny showing range in the number of ovarioles per ovary for each species (B) Ovarian morphology (C) Mean ovariole number and the standard deviation for the sibling species and their F1 hybrid ♀ *D. simulans* x ♂ *D. sechellia*. The red boxes indicate data which is of interest to this study. **[b]** Gonadal morphology of the ovary (left) and testis (right) of the reciprocal hybrid cross — ♀ *D. simulans* x ♂ *D. sechellia* **[c]** Gonadal morphology of the ovary (left) and testis (right) of the reciprocal hybrid cross — ♀ *D. simulans* x ♂ *D. sechellia.*

(Figure 3.10c). The ovaries from this cross are more comparable to those found in *D. sechellia* (Figure 3.10a). Furthermore, the testis of this cross appears to be atrophic, with smaller testis size and reduced seminal vesicles (Figure 3.10c).

DISCUSSION

4.1 Sibling Species Differ in Tissue-Specific Distribution of Gene Expression

The results obtained for tissue-specific distribution of gene expression between the species under investigation were of interest. When analyzing the genes at a global level — a comprehensive analysis of the number of genes expressed within each tissue or mixed-tissue grouping — gene expression data among the three species, *Drosophila melanogaster*, *D. simulans* and *D. sechellia*, illustrated greater resemblance between the *D. simulans* and *D. melanogaster* species upon the comparison of distribution across tissues of the genes of interest. The other sibling species, *D. sechellia*, displayed a clear departure in tissue-specific distribution from the other two species. The observed result depicting a greater homogeneity between the *melanogaster* and *simulans* species, relative to former and *sechellia*, is consistent with the reported phylogeny of the *D. melanogaster* subgroup (Powell 1997). While this result is expected, the finding that between the sibling species, *D. simulans* and *D. sechellia*, relative to *D. melanogaster*, there was not a high level of homogeneity was unusual. One would expect a greater proportion of similarities to exist between the sibling species as a consequence of their more recent divergence, ~250,000 years ago, relative to either from *D. melanogaster* ~0.5 - 2.5 MYA (Caccone et al. 1988; Catron and Noor 2008; McDermott and Kliman 2008). Possible reasons for such an unusual result may be due to (1) variation in methodology used to analyze relative gene expression among this study versus previous experiments, or (2) perhaps it is due to the *D. sechellia* species having origins of an island species while both *D. melanogaster* and *D. simulans* share a cosmopolitan origin (Tsacas and Bachli, 1981 and Lachaise et al. 1986; Coyne

and Kreitman 1986).

That being said, however, when looking at specific genes and their relative tissue-distributions one can distinguish trends for each individual gene across species (Table 3.1). It was found that for genes which had similarities in tissue expression between at least two of the species, the majority of genes shared the same tissue expression patterns between the sibling species, *D. simulans* and *D. sechellia*. Furthermore, as support for the proposed phylogenies (Powell 1997; Tamura et al. 2004; Stark et al. 2007; Larracuente et al. 2008), *D. simulans* had a greater number of genes sharing tissue expression patterns with *D. melanogaster*, while *D. sechellia* had fewer genes sharing tissue expression.

Therefore, it was found that while the number of genes expressed in specific tissue categories may appear distinct between closely related species, the analysis of specific gene tissue expressions allowed for a clear observation of gene expression patterns between species.

4.2 Males and Females Differ in Tissue-Specific Distribution of Gene Expression

Among the pure species within this study, a strong correlation was found in gene distribution between males and females among tissue-specific categories. This result, demonstrating that pure species generally display similar expression patterns, parallels those found in a number of other studies (Ranz et al. 2003; Rifkin et al. 2003 and Ranz et al. 2004). This result is a consequence of compensatory changes which act to ‘eclipse‘ the prevailing interspecific differences between species (Dobzhansky 1937; Muller 1942; Orr 1997; Johnson and Porter 2000; Orr and Presgraves 2000; Porter and Johnson 2002).

A greater opportunity to observe differential gene expression between the sexes resides within the reciprocal hybrids. Previously determined, approximately half of the *Drosophila* transcriptome is composed of sex-biased genes — a gene is deemed sex-biased when its level of expression significantly varies between the sexes (Ranz et al. 2003; 2004; Parisi et al. 2003; Ortiz-Barrientos et al. 2007 and Gallach et al. 2011). Specifically, male-biased genes have shown to have a greater difference in expression than both female- and non-sex-biased genes as well as possess a greater proportion of genes experiencing sex-biased (Ranz et al. 2003; 2004; Parisi et al. 2003; Haerty and Singh 2006; Graveley et al. 2010 and Daines et al. 2011). With this in mind, surprising results were found when observing the male and female tissue-specific gene expression data for the F1 interspecific crosses. The majority of genes did not display significant variation between the sexes. In the cross between ♀ *D. simulans* x ♂ *D. sechellia*, more than half of the genes analyzed displayed parallel expression between the sexes. Similar results were found in the reciprocal cross, ♀ *D. sechellia* x ♂ *D. simulans*, having the majority of genes exhibiting parallel patterns of expression for both sexes.

To investigate this further, comparisons were made between the expression patterns found in male and female head tissues. As expected, based on previous studies, the 'somatic tissues' (male and female head) revealed a close correlation between the sexes across both pure species and their reciprocal hybrids (Meiklejohn et al. 2003; Parisi et al. 2003; 2004 and Ranz et al. 2003). Conversely, when comparing the testis and ovary tissues, an unexpected result emerged. For pure species, it was found that the ovary tissue had a higher level of expression over the testis tissue, which had a relatively low level of gene expression. For the reciprocal hybrids, they exhibit varied expression between the gonadal tissues. The cross between ♀ *D.*

simulans x ♂ *D. sechellia* demonstrates a higher level of expression in the testis over ovary tissues, while the reciprocal cross ♀ *D. sechellia* x ♂ *D. simulans,* exhibits the same patterns observed in the two pure species. Previous studies had found that testis tissue generally has a much higher degree of expression relative to ovary and head tissues (Parisi et al. 2004 and Ranz et al. 2004 and Haerty and Singh 2006). This result may be the consequence of: (1) unlike previous studies which employed the less sensitive technique of microarrays, this study analyzed expression via qRT-PCR which has been found to have a greater degree of sensitivity, or (2) there may be an underlying problem with the procedure utilized in tissue collection/preservation. It may be possible that during the duration of the dissection of the testis the RNA could have degraded slightly from the time of extraction until it was placed in RNAlater©. Furthermore, the samples were placed on ice instead of being flash frozen.

4.3 Reciprocal Crosses Differ in Patterns of Gene Misexpression

Analysis of reciprocal hybrids is lacking throughout the literature due to the majority of studies focusing on one F1 hybrid when making comparisons between pure species (Ranz et al. 2004; Haerty et al. 2007; Cantron and Noor 2008 and Artieri et al. 2009). Despite this lack of attention in the literature, a paper presented by Dobzhansky (1935) suggests the potential significance regarding the study of reciprocal crosses as they were noted to produce diverse results. Such divergence could be the result of such a phenomenon known as ‘maternal effect’ which, until the late 1980s, had largely been ignored (Mousseau et al. 2009). As expected, looking at the overall gene distribution based on expression levels across all tissues types

between the hybrids, varying results were observed between the hybrids. The results found between the hybrids suggest a strong pressure exerted by the maternal genome on the respective cross. This was further supported by the observation that suggests the reciprocal hybrids possessing tissue-specific gene distribution with a noted degree of similarity to that found in their maternal species.

4.4 Sex and Non-Sex Genes Differ in the Extent of Gene Misexpression

As previously stated, a number of studies have determined that ~ 30 - 50% of genes in the *Drosophila* transcriptome display sex-biased expression and the majority of the genes exhibiting a sex-bias are found to be expressed in reproductive tissues (testis and ovary) (Arbeitman et al. 2002 and Parisi et al. 2003 and Ranz et al. 2003, 2004). Within this study it was found that the hybrid males demonstrate a higher proportion of misexpressed genes in the head tissue, while the hybrid females demonstrate a higher proportion of misexpressed genes in both the head and ovary tissues. These results are unexpected and the reasons are two-fold: (1) In male hybrids it should be expected that there would be a higher proportion of misexpressed genes in the testis, yet there were only 5% of genes displaying tissue-specific modulation. (2) In female hybrids, while it has been previously determined that the expression between ovary- and head-specific genes are not significantly divergent, it has also been observed that divergent expression levels are far lower than those found in testis-specific genes (see Supplementary Figures §4.7-4.10) (Civetta and Singh 1995; Jagadeeshan and Singh 2005 and Haerty and Singh 2006).

A significant difference between this study and those previously conducted has much to

do with the methodology employed. A paper published by Catron and Noor (2008) which discussed the misleading nature and relative sensitivity of former experimental designs. When investigating hybrid misexpression a number of the studies used whole adult bodies (Michalak and Noor 2003 and Moerhring et al. 2007). By analyzing expression for tissue-specific expression patterns using a whole adult body rather than the specific tissue of interest the tissue-specific expression may be overlooked or erroneously inferred as a response to disruption in other tissues. Another study improved on this method by exclusively analyzing the testis for testis-specific expression patterns using microarrays (Haerty and Singh 2006). However, it has been demonstrated that microarrays are less sensitive due to the use of extensive pooling and an RNA amplification step (Gilad et al. 2005 and Moerhring et al. 2007). As a result, Catron and Noor suggested the next step would be to use biologically replicated qRT-PCRs on cDNA pools of individual tissue types as a means to examine the extent of misexpression between species (Catron and Noor 2008).

4.5 Biological Replicates

Two biological replicates were conducted for all 40 genes of interest across the two pure species and their two reciprocal crosses in all four tissue types. However, upon review of these two replicates, the first replicated set, deemed biological replicate 1 (BR1), appeared to contain extreme values that may have resulted from a technical or procedural error. Furthermore, when comparing the values of the housekeeping gene expression between the two replicates, the values found in BR1 were more variant than those found in BR2. As a result, analysis was conducted on

BR2 only (See Supplementary Figures §4.11-4.14 for a plotted comparison between both biological replicates).

4.6 Future Directions

Due to the employment of only one biological replicate it would prove paramount to have a second replicate for this study to increase the significance of the results. It may be interesting to compare the two biological replicates produced in this study to see the amount of difference it would make to the final results. As well, it would be interesting to conduct a functional analysis of all the genes to investigate the possibility of overrepresentation of a particular functional category among the differentially expressed genes. This may give insight into the varying degrees of expression across tissues. Finally, one could use this study as a useful foundation for a better understanding of gene regulation and its role in speciation. The future direction of this study necessitates the targeting of 10 genes out of the 40 studied here and perform the same methodology using three biological replicates. Ultimately, this study will supply a wealth of data regarding tissue-specific expression in both sexes, supply insight into the variation between reciprocal hybrids as well as advance our understanding of the process of species divergence and speciation.

REFERENCES

Andersson, M. (1994). *Sexual selection*. New Jersey: Princeton University Press.

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

Artieri, C., Haerty, W. and Singh, R. S. (2007). Association between levels of coding sequence divergence and gene misregulation in *Drosophila* male hybrids. *J. Mol. Evol.* **65**: 697-704.

———. (2009). Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of *Drosophila*. *BMC Biology* **7**: 42-47.

Barbash, D. (2010). Genetic testing of the hypothesis that hybrid male lethality results from a failure in dosage compensation. *Genetics*. **184**: 313-316.

Barbash, D. and Lorigan, J. (2007). Lethality in *Drosophila melanogaster/Drosophila simulans* species hybrids is not associated with substantial transcriptional misregulation. *J Exp Zoolog B Mol Dev Evol.* **308**: 74–84.

Bomblies, K. and Weigel, D. (2007). Arabidopsis- a model genus for speciation. *Current Opinion in Genetics and Development* **17**: 500-504.

Bowler, Peter. 1973. Bonnet and Buffon: Theories of Generation and the Problem of Species. *Journal of the History of Biology* 6, no. 2: 259-281.

Britten, R. and Davidson, E. (1969). Gene regulation for higher cells: a theory. *Science*. **165**: 349-357.

Caccone, A., Amato, G. and Powell, J. (1988). Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. *Genetics*. **118**: 671-683.

Carlson, P. (1971). A genetic analysis of the rudimentary locus of Drosophila melanogaster. *Genet. Res.* **17**: 53-81.

Carson, H. L. (2000). Sexual selection in populations: the facts require a change in the genetic definition of the species. pp. 495-512. Cambridge University Press, Cambridge.

Catron, D. and Noor, M.A.F. (2008). Gene expression disruptions of organism versus organ in *Drosophila* species hybrids. *PLoS ONE*. **3**: e3009.

Charlesworth, B., Coyne, J. A. and Barton, N. (1987). The relative rates of evolution of sex chromosomes and autosomes. *American Naturalist*. **130**: 113–146.

Civetta, A. and Singh, R. S. (1995). High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. *J. Mol. Evol.* **41**: 1085-1095.

———. (1999). Broad-sense sexual selection, sex gene pool evolution and speciation. *Genome.* **42**: 1033-1042.

Clark, Andrew. (2006). Genomics of the evolutionary process. *Trends Ecol Evol.* **21**: 316–321.

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

Connallon, T. and Clark, A. (2011). Association between sex-biased gene expression and mutations with sex-specific phenotypic consequences in *Drosophila*. *Genome Biol. Evol.* **3**: 151-155.

Coulthart, M. B. and Singh, R. S. (1988). High level of divergence of male reproductive tract proteins between *Drosophila melanogaster* and its sibling species, *Drosophila simulans*. *Molec. Biol. Evol.* **5**: 182-191.

Coyne, Jerry. (1985). The genetic basis of Haldane's rule. *Nature*. **314**: 736–738.

———. (1992). Genetics and speciation. *Nature*. **355**: 511-515.

———. (1994). Ernst Mayr and the origin of species. *Evolution.* **51**: 19-30.

Coyne, J. A. and Kreitman, M. (1986). Evolutionary genetics of two sibling species *Drosophila simulans* and *D. sechellia*. *Evolution*. **40**: 673-691.

Coyne, J. A. and Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution*. **43**: 362-381.

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

Darwin, Charles. (1859). *On the Origin of Species By Means of Natural Selection or The Preservation of Favoured Races in the Struggle for Life (First Edition)*. John Murray, London.

de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proc. Natl. Acad. Sci.* **102**: 6600-6607.

Dobzhansky, Theodosius. (1935). The Y Chromosome of *Drosophila pseudoobscura*. *Genetics*. **20**: 366-376.

———. (1936). Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics*. **21**: 113-135.

———. (1937). *Genetics and the Origin of Species*. Columbia University Press, New York.

Dronamraju, K. (2011). *Haldane, Mayr, and Beanbag Genetics*. Oxford University Press, New York.

Dobzhansky, T. and Sturtevant, A. H. (1935). Further data on maternal effects in Drosophila pseudoobscura hybrids. *Proc. Natl Acad. Sci. USA.* **21:** 566–570.

Dworkin, I. and Jones, C. (2009). Genetic changes accompanying the evolution of host specialization in *Drosophila sechellia*. *Genetics*. **181**: 721-736.

Eberhard, W. G. (1985). *Sexual selection and animal genitalia*. Harvard University Press, Cambridge.

Ellegren, H. and Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet.* **8**: 689–698.

Ehrman, L. (1960). Genetics of hybrid sterility in Drosophila paulistorum. *Evolution*. **14**: 212-223.

Eisley, Loren. (1961). *Darwin's Century. Evolution and the Men Who Discovered it.* New York: Anchor Books.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

———. (1958). *The Genetical theory of natural selection. (ed. 2).* Dover: New York.

Gallach, M. Domingues, S. and Betran, E. (2011). Gene duplications and the genome distribution of sex-biased genes. *International Journal of Evolutionary Biology*. **2011:** Article ID 989438.

Gibson, G. and Dworkin, I. (2004). Uncovering Cryptic Genetic Variation. *Nature Reviews Genetics*. **5**: 681-690.

Gibson, G. and Weir, B. (2005). The quantitative genetics of transcription. *Trends Genet*. **21**: 616–623.

Gilad, Y., Rifkin, S., Bertone, P., Gerstein, M. and White, K. (2005). Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. Genome Res. 15:674–680.

Gould, Stephen J. (1987). *Time's Arrow, Time's Cycle: Myth and metaphor in the discovery of geological time.* Harvard University Press.

———. (1989). *Wonderful life: the Burgess Shale and the nature of history.* W. W. Norton & Company, Inc.

———. (2002). *The Structure of Evolutionary Theory.* Harvard: Belknap Harvard.

Graveley, B., Brooks, A., Carlson, J. et al. (2010). The developmental transcriptome of Drosophila melanogaster. *Nature*. **471**: 473-479.

Haerty, W. and Singh, R. S. (2006). Gene regulation divergence is a major contributor to the evolution of Dobzhansky-Muller incompatibilities between species of *Drosophila*. *Mol. Biol. Evol.* **23**: 1707-1714.

Haerty, W., Jagadeeshan, S., Kulathinal, R.J., Wong, A., Ravi Ram, K., Sirot, L.K., Levesque, L., Artieri , C.G., Wolfner, M.F., Civetta, A. and Singh, R. S. (2007). Evolution in the fast lane: Rapidly evolving sex-related genes in *Drosophila*. *Genetics.* **177**: 1321-1335.

Haldane, J. B. S. (1922). Sex-ratio and unisexual sterility in hybrid animals. *Journal of Genetics*. **12**: 101–109.

———. (1932). *The causes of evolution*. Harper and Brothers, New York.

Huxley, Julian. (1938). Darwin's theory of sexual selection and the data subsumed by it, in the light of recent research. *American Naturalist*. **72**: 416-433.

———. (1942). *Evolution: the modern synthesis*. Harper.

Jagadeeshan, S. and Singh, R.S. (2005). Rapidly Evolving Genes of *Drosophila*: Differing levels of Selective Pressure in Testis, Ovary and Head Tissues Between Sibling Species. *Mol Biol Evol.* **22**: 1793 – 1801.

Janick, Jules. (1989). Gregor Mendel. In: Classic Papers in Horticultural Science. pp. 406-412. Prentice Hall, Englewood Cliffs, NJ.

Joly, D., Bressac, C., Devaux, J. and Lachaise, D. (1991). Sperm length diversity in *Drosophilidae*. *Drosophila Information Service.* **70**: 104-108.

Johnson, Norman A. (2000). *Gene interaction and the origin of species. Epistasis and the Evolutionary Process*. Oxford University Press, New York.

———. (2002). Sixty years after "Isolating mechanisms, evolution, and temperature": Muller's legacy. *Genetics.* **161**: 939-944.

Johnson, N. A. and Porter, A. (2000). Rapid speciation via parallel, directional selection on regulatory genetic pathways. *Journal of Theoretical Biology.* **205**: 527-542.

Khaitovich, P. Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G. Lachmann, M. and Paabo, S. (2005). Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*. **309**: 1850-1854.

King, M. C. and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science*. **188**: 107-116.

Korbert, Gaye. (1998). *What Darwin Didn't Know: Gregor Mendel and the Mechanism of Heredity*. WGBH Educational Foundation.

Kottler, M. (1979). Hugo de Vries and the Rediscovery of Mendel's Laws. *Annals of Science. An International Review of the History of Science and Technology from the Thirteenth Century London.* **36**, (5): 517-538.

Kulathinal, R. J. and Singh, R. S. (2008). The molecular basis of speciation: from patterns to processes, rules to mechanisms. *Journal of Genetics*. **87**: 327-338.

Kutschera, U. and Niklas, K. (2004). The modern theory of biological evolution: an expanded synthesis. *Naturwissenschaften*. **91**: 255-276.

Landry, C., Lemos, B., Rifkin, S. A., Dickinson, W. and Hartl, D. (2007). Genetic properties influencing the evolvability of gene expression. *Science* **317**: 118–121.

Landry, C., Wittkopp, P., Taubes, C., Ranz, J., Clark, A. and Hartl, D. (2005). Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics*. **171**: 1813-1822.

Larracuente, A., Sackton, T., Greenberg, A., et al. (2008). Evolution of protein-coding genes in Drosophila. *Trends in Genetics.* **24**: 114–123.

Leigh, E. (1990). Genes, bees and ecosystems: The evolution of a common interest among Individuals. *Trends in Ecology and Evolution.* **6**: 257-262.

Lemeunier, F. and Ashburner, M. (1984). Relationships within the melanogaster species subgroup of the genus *Drosophila (Sophophora).* IV. The chromosomes of two new species. *Chromosoma.* **89**: 343-351.

Levine, R. and Evers, C. (1999). *The Slow Death of Spontaneous Generation (1668-1859).*Washington, D.C.: National Health Museum.

Mallet, James. (2004). Poulton, Wallace and Jordan: how discoveries in *Papilio* butterflies led to a new species concept 100 years ago. *Systematics and Biodiversity*. **1**: 441-452.

———. (2010). Group selection and the development of the biological species concept. *Philosophical Transactions of the Royal Society.* **365**: 1853-1863.

Mayr, Ernst. (1940). Speciation phenomena in birds. *American Naturalist*. **74**: 249-278.

———. (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

———. (1963). *Animal species and evolution.* Harvard University Press, Cambridge.

———. (1970). *Populations, Species, and Evolution*. Cambridge: Harvard University Press.

———. (1981). *The growth of biological thought*. Cambridge: Harvard University Press.

Mavarez, J., Audet, C. and Bernatchez, L. (2009). Major disruption of gene expression in hybrids between young sympatric anadromous and resident populations of brook charr (Salvelinus fontinalis Mitchill). *J. Evol. Biol.* **22**: 1708-1720.

McDermott, S. and Kliman, R. (2008). Estimation of isolation times of the island species in the *Drosophila simulans* complex from multilocus DNA sequence data. *PLoS ONE*. **3**: e2242.

Meiklejohn, C., Parsch, J., Ranz, J. and Hartl, D. (2003). Rapid evolution of male-biased gene expression in *Drosophila*. *PNAS*. **100**: 9894-9899.

Michalak, P. and Noor, M.A.F. (2003). Genome-wide patterns of expression in Drosophila pure-species and hybrid males. *Mol. Biol. Evol.* **20**: 1070-1076.

———. (2004). Association of misexpression with sterility in hybrids of *Drosophila simulans* and *D. mauritiana*. *Journal of Molecular Evolution*. **59**: 277-282.

Moehring, A. Llopart, A. Elwyn, S., Coyne, J. and Mackay, T. (2006). The genetic basis of postzygotic reproductive isolation between *Drosophila santomea* and *D. yakuba* due to hybrid male sterility. *Genetics*. **173**: 225-233.

Moehring, A., Teeter, K. and Noor, M.A.F. (2007). Genome-Wide Patterns of Expression in Drosophila Pure Species and Hybrid Males. II. Examination of Multiple-Species Hybridizations, Platforms, and Life Cycle Stages. *Mol. Biol. Evol.* **24**: 137-145.

Mousseau, T., Uler, T., Wapstra, E. and Badyaev, A. (2009). Evolution of maternal effect: past and present. *Phil. Trans. R. Soc. B*. **364**: 1035-1038.

Muller, Hermann. (1940). Bearing of the Drosophila work on systematics. In *The New Systematics*, J. Huxley, ed. Oxford: Clarendon.

———. (1942). Isolating mechanisms, evolution and temperature. *Biol. Symp.***6**: 71-125.

Nei, Masatoshi. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Noor, M. and Feder, J. (2006). Speciation genetics: Evolving approaches. *Nature Reviews Genetics*. **7**: 851-861.

Novák, Jakub. (2008). Alfred Russel Wallace's and August Weismann's evolution: A story written on butterfly wings. Princeton University. New Jersey.

Olby, Robert. (1997). Mendel, Mendelism and Genetics. *History of Science.* **17**: 55-72.

Orr, H. Allen. (1993). Haldane's rule has multiple genetic causes. *Nature*. **361**: 532-533.

———. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*. **139**: 1805-1813.

———. (1997). Haldane's rule. *Annual Review of Ecology and Systematics*. **28**: 195-218.

Orr, H.A. and Presgraves, D.C. (2000). Speciation by postzygotic isolation: forces, genes, and molecules. *BioEssays.* **22**: 1085-1094.

Orr, H. A. and Turelli, M. (1996). Dominance and Haldane's rule. *Genetics*. **143**: 613-616.

———. (2001). The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution*. **55**:1085-1094.

Ortiz-Barrientos, D., Counterman, B. and Noor, M. (2007). Gene expression divergence and the origin of hybrid dysfunctions. *Genetica*. **129**: 71-81.

Palumbi, S. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annu. Rev. Ecol. Syst.* **25**: 547-572.

Parisi, M., Nuttall, R. Edwards, P. et al. (2004). Survey of ovary-, testis-, and soma-biased gene expression in Drosophila melanogaster adults. *Science*. **299**: 697-700.

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

Philip P. Wiener. (1973). "Spontaneous Generation". *Dictionary of the History of Ideas*. New York: Charles Scribner's Sons.

Porter, A. and Johnson, N. A. (2002). Speciation despite gene flow when developmental pathways evolve. *Evolution*. **56**: 2103-2111.

Powell, J. R. (1997). *Progress and prospects in evolutionary biology: the Drosophila model*. Oxford University Press, Oxford.

Presgraves, Daven. (2002). Patterns of postzygotic isolation in Lepidoptera. *Evolution*. **56**: 1168-1183.

Ranz, J., Castrillo-Davis, C., Meiklejohn, C. and Hartl, D. (2003). Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science*. **300**: 1742-1744.

Ranz, J. M. and Machado, C. A. (2006). Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol.* **21**: 29-37.

Ranz, J., Namgyal, K., Gibson, G. and Hartl, D. (2004). Anomalies in the level of expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. *Genome Research*. **14**: 373-379.

Reiland, J. and Noor, M.A.F. (2002). Lack of qualitative RNA misexpression in sterile male F1 hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *BMC Evolutionary Biology*. **2**: 16.

Ridley, Mark. (2004). *Evolution*. 3rd ed. Blackwell Scientific Publishing, Malden, MA.

Rifkin, S.A., Kim, J., and White, K.P. (2003). Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**: 138–144.

Sapp, Jan. (1990). *Experimental inquiries: historical, philosophical, and social studies of:The Nine Lives of Gregor Mendel.* Netherlands. Kluwer Academic Publishers. p137-166.

Singh, Rama, S. (1990). *Patterns of species divergence and genetic theories of speciation. In: Topics in Population Biology and Evolution*. pp. 231-265. Springer-Verlag, New York.

———. (2000). *Toward a unified theory of speciation. In Evolutionary genetics: from molecules to morphology.* R. S. Singh and C. Krimbas (ed.) pp. 570-604. Cambridge University Press, New York.

Singh, R. S. and Kulathinal, R. (2000). Sex gene pool evolution and speciation – a new paradigm. *Genes and Genetic Systems*. **75**: 119-130.

———. (2005). Male sex-drive and masculinization of the genome. *BioEssays.* **27:** 518-25.

Stark, A., Lin, M.F., Kheradpour, P., et al. (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature.* **450**: 219–232.

Swanson, W. and Vacquier, V. (2002). The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3**: 137–44.

Tamura, K., Nei, M. and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the Neighbor-Joining method. *Proc Natl Acad Sci USA*. **101**:11030–11035.

Tamura, K., Subramanian, S. and Kumar, S. (2004). Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. *Mol Biol Evol*. **21**: 36–44.

Terrall, Mary. (2002). *The Man who flattened the Earth: Maupertuis and the sciences in the Enlightenment*. The University of Chicago Press.

True, J. and Haag, E. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol Dev*. **3**: 109–119.

Turelli, M. and Orr, H. A. (1995). The dominance theory of Haldane's rule. *Genetics*. **140**: 389-402.

———. (2000). Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*. **154**: 1663-1679.

Van Wyhe, John. (2008). *Darwin: The Story of the Man and His Theories of Evolution*. London: Andre Deutsch Ltd.

Weiling, F. (1991). Historical study: Johann Gregor Mendel 1822–1884. *American journal of medical genetics*. **40**: 1–25.

White, O., Eisen, J., Heidelberg, J. et al. (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans*. *Science*. **286**: 1571–1577.

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

Wiener. P. (1973). *Spontaneous generation*. *Dictionary of the History of Ideas*. New York: Charles Scribner's Sons. (ed.)

Wittkopp, P. J., Haerum, B. K. and Clark, A. G. (2008). Independent effects of *cis*- and *trans*-regulatory variation on gene expression in *Drosophila melanogaster*. *Genetics*. **178**: 1831-1835.

Wray, G., Hahn, M., Abouheif, E., Balhoff, J., Pizer, M., Rockman, M. and Romano, L. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**: 1377-1419.

Wright, Sewall. (1931). Evolution in Mendelian populations. *Genetics*. **16**: 97–159.

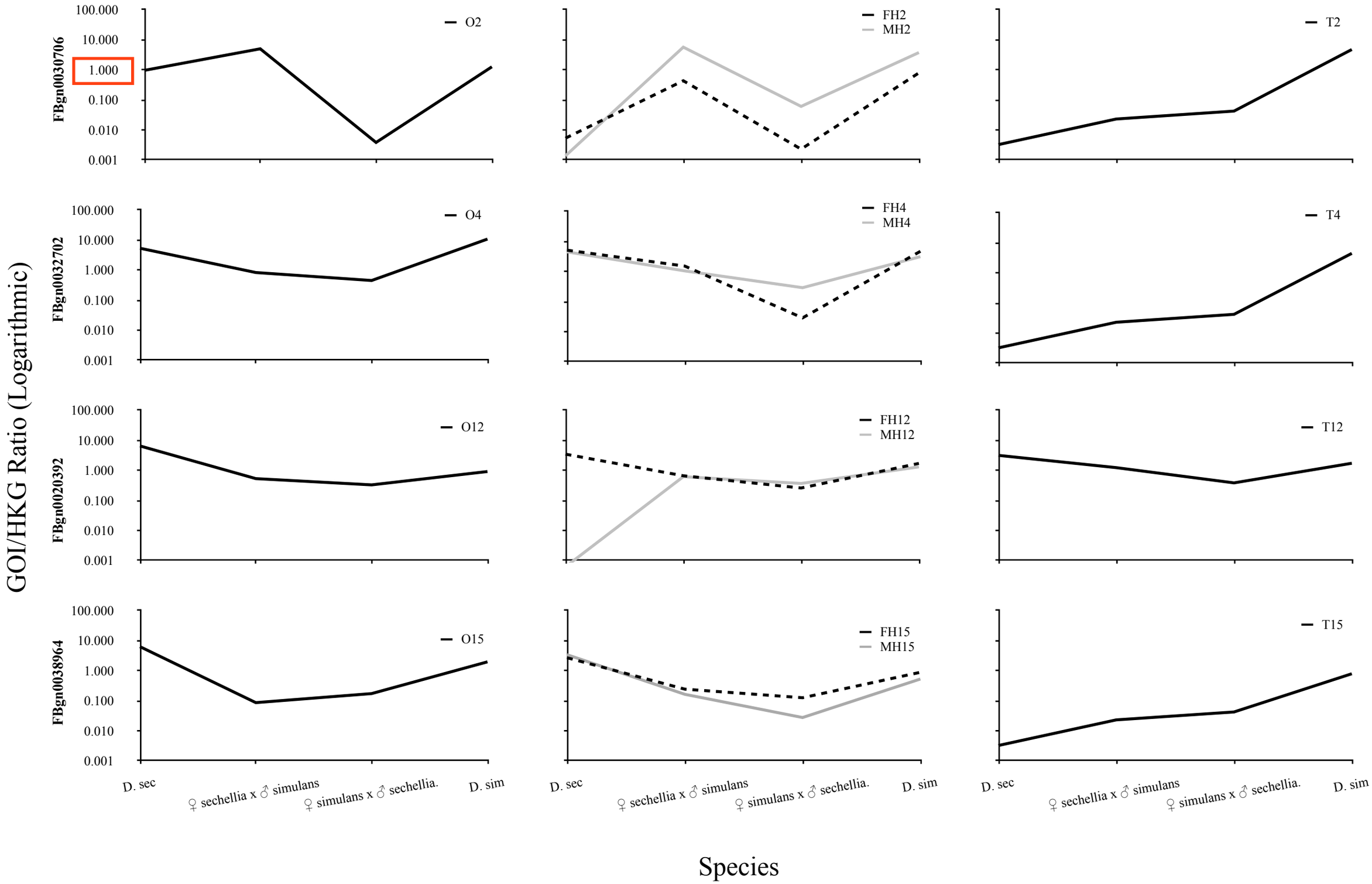
———. (1984). *Evolution and the Genetics of Populations, Volume 1: Genetic and Biometric Foundations*. The University of Chicago Press.

Wu, C. I. and Davis, A. (1993). Evolution of postmating reproductive isolations — The composite nature of Haldane's rule and its genetic bases. *American Naturalist*. **142**: 187-212.

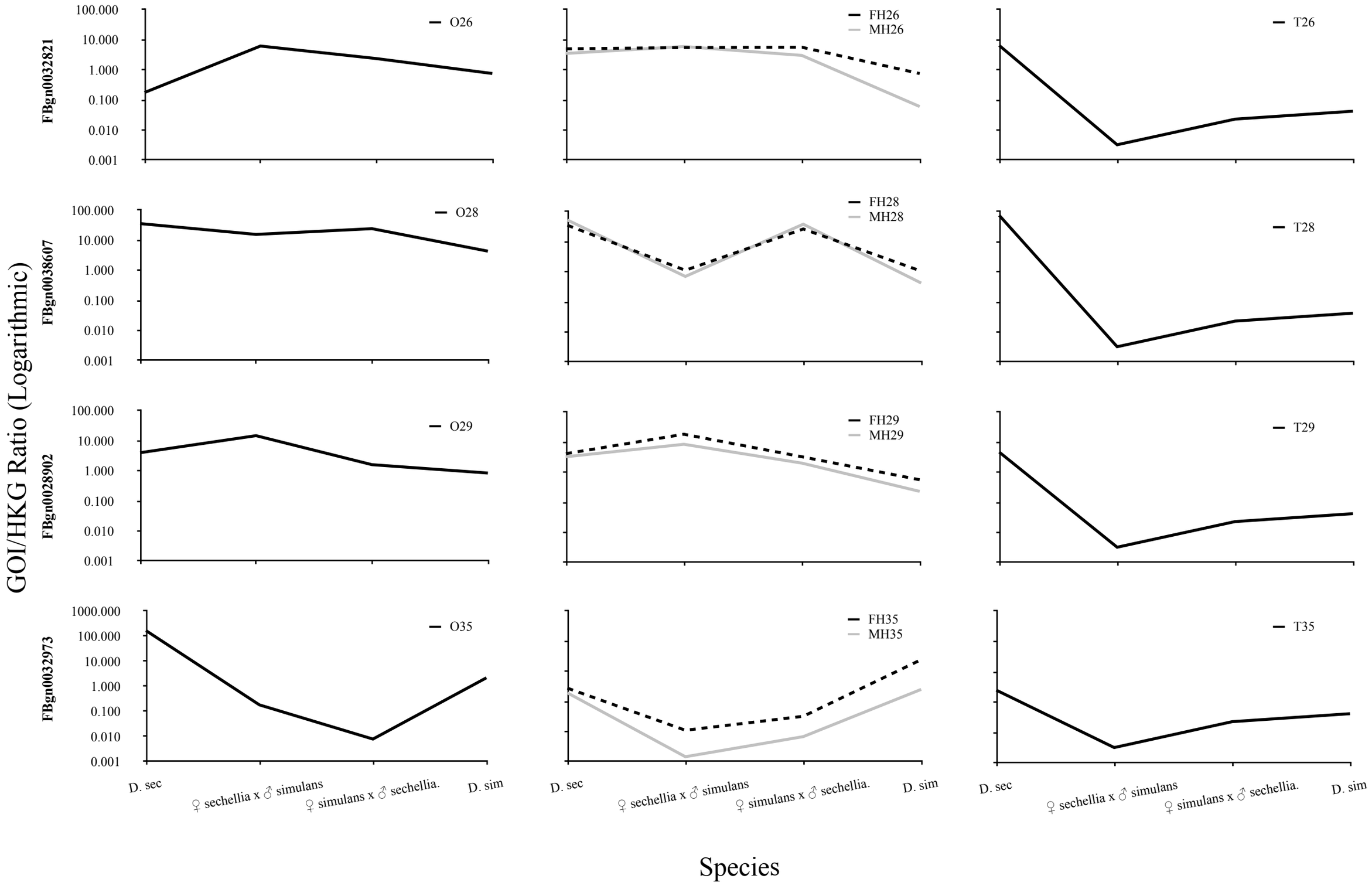
Wu, C. I., Johnson, N. and Palopoli, M. (1996). Haldane's rule and its legacy: Why are there so many sterile males? *Trends in Ecology and Evolution* **11**: 281-284.

Yon Rhee, Seung. (1999). *Gregor Mendel (1822-1884)*.Washington, D.C.: National Health Museum.

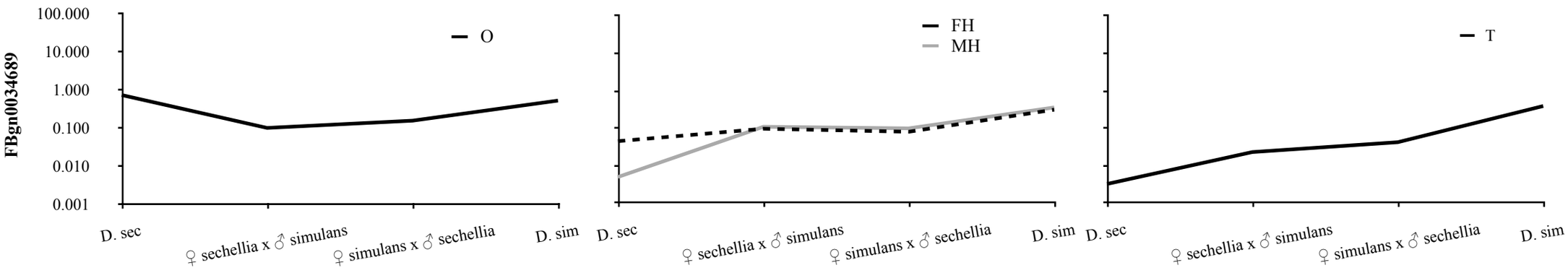
Zeng, L.-W. and Singh, R. S. (1993). A combined classical genetic and high resolution two-dimensional electrophoretic approach to the assessment of the number of genes affecting hybrid male sterility in *Drosophila simulans* and *Drosophila sechellia*. *Genetics.* **135**: 135-147.



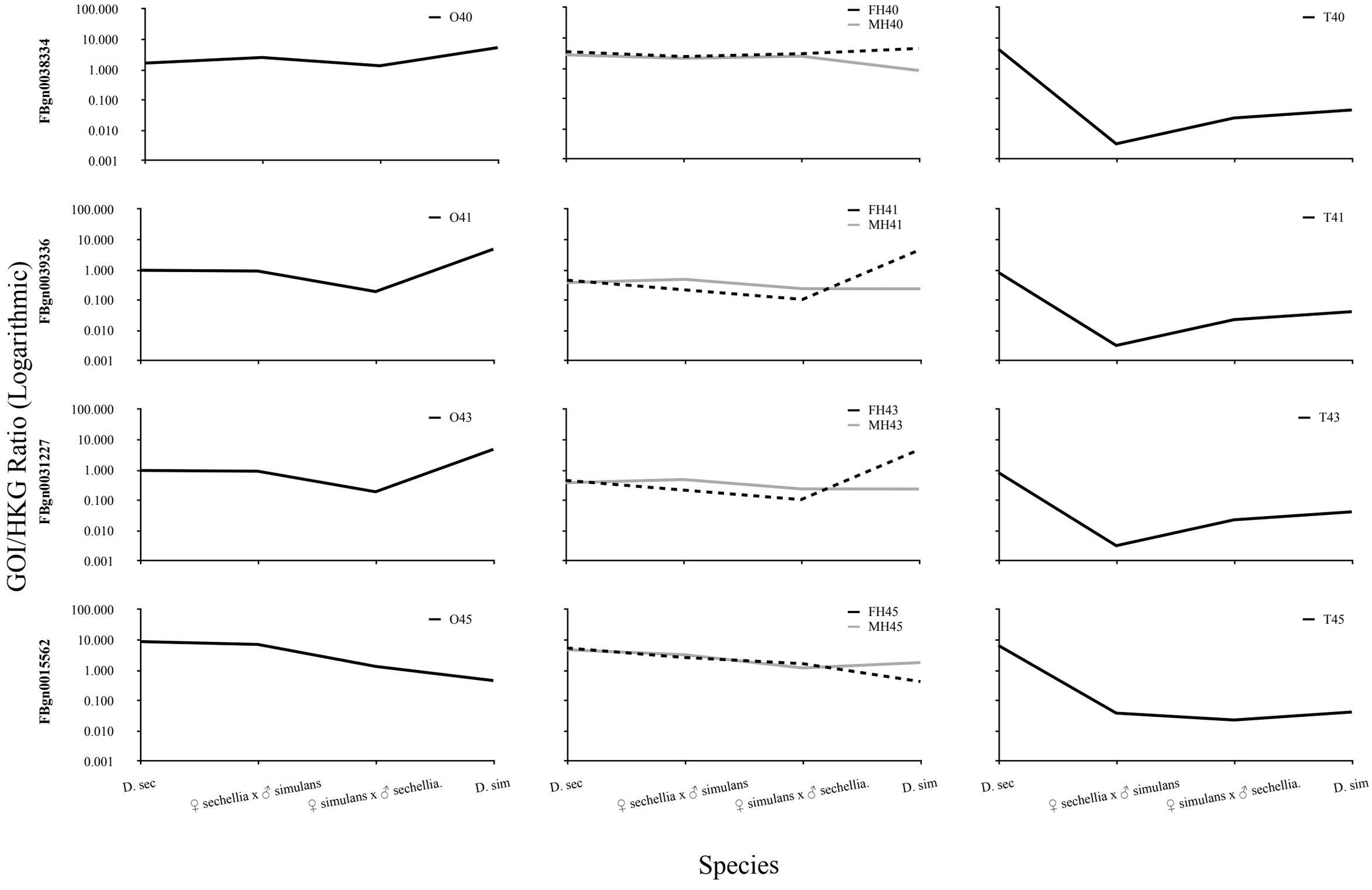
Supplementary Figure §4.1a: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



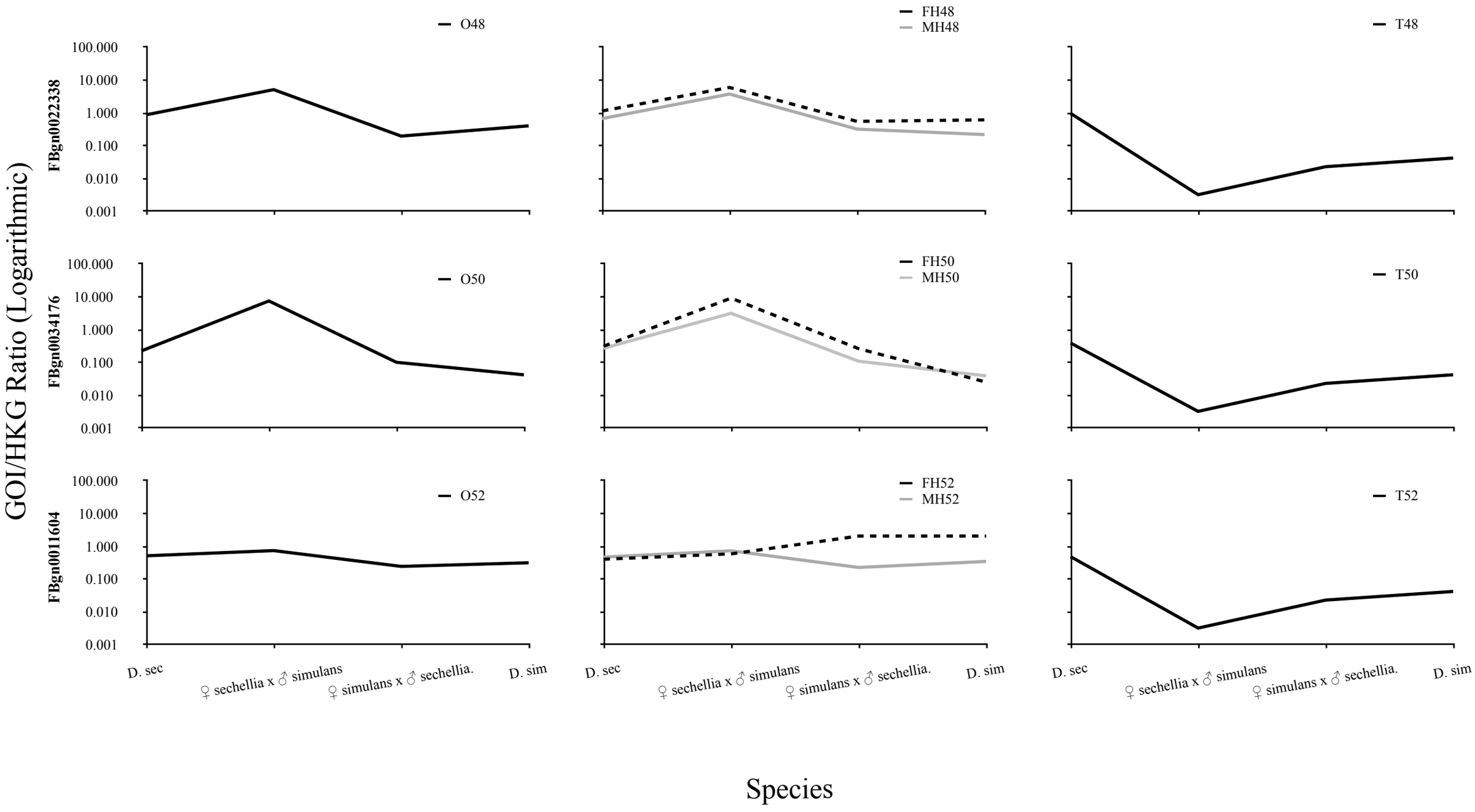
Supplementary Figure §4.1b: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



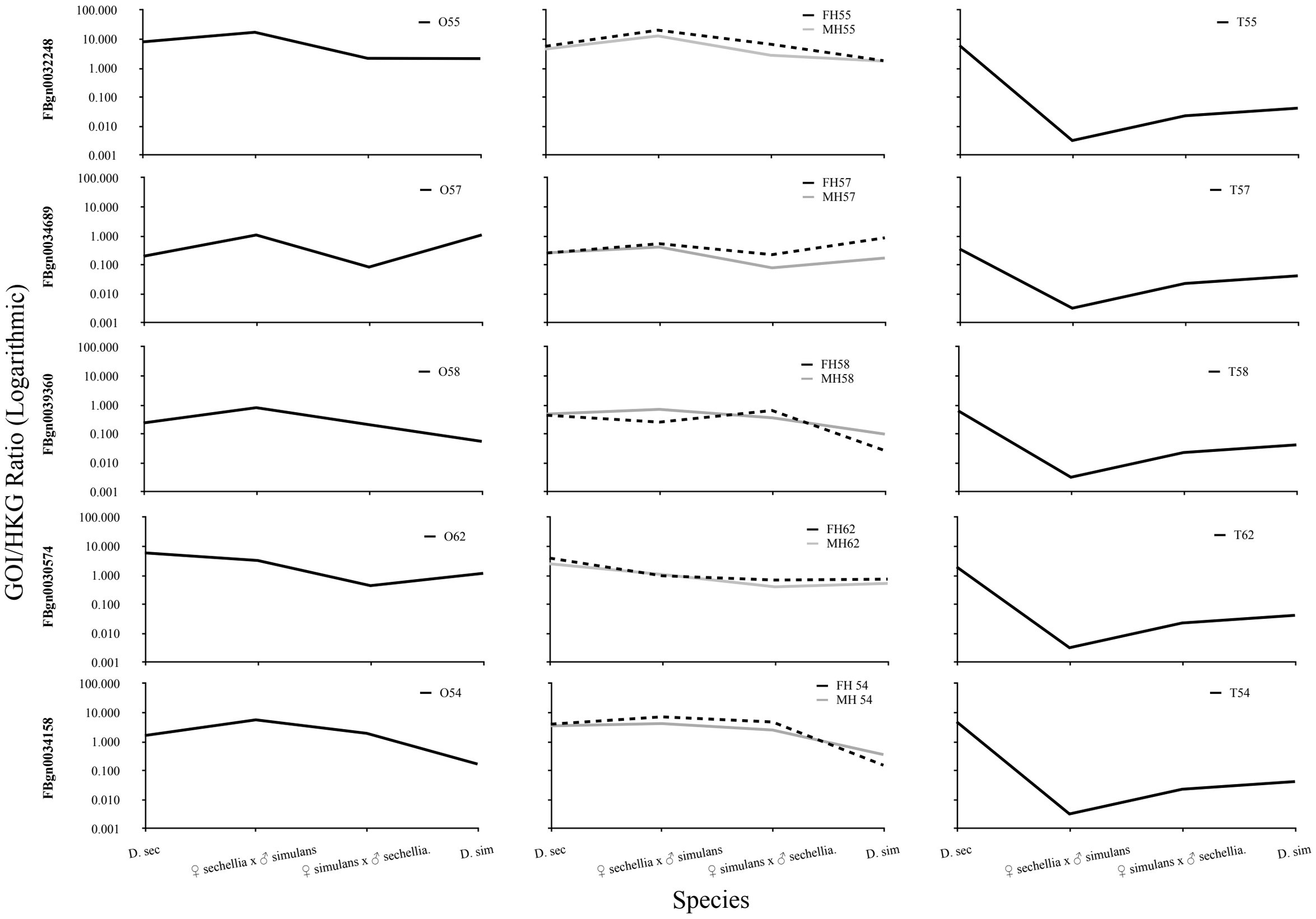
Supplementary Figure §4.1c: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



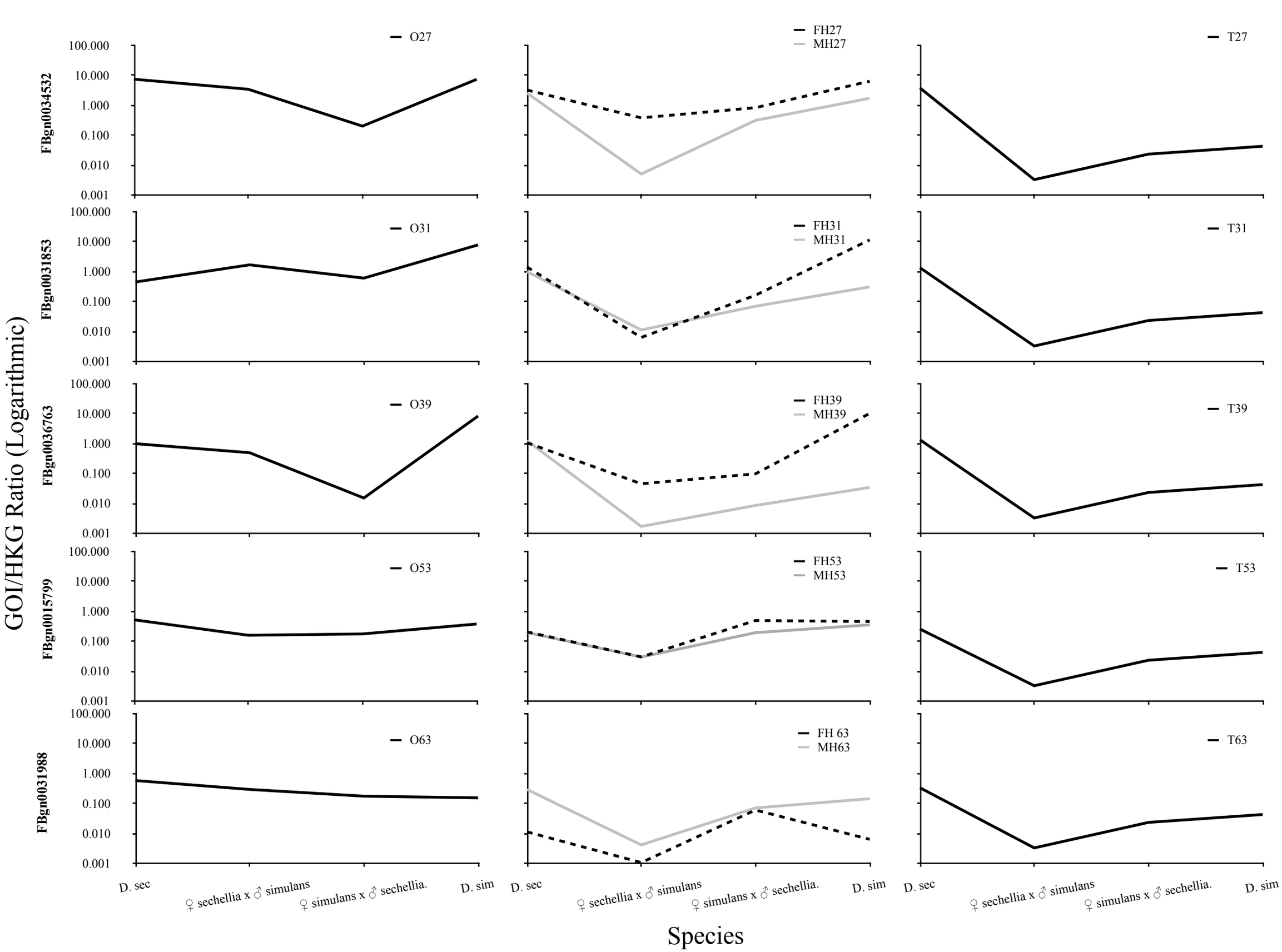
Supplementary Figure §4.2a: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



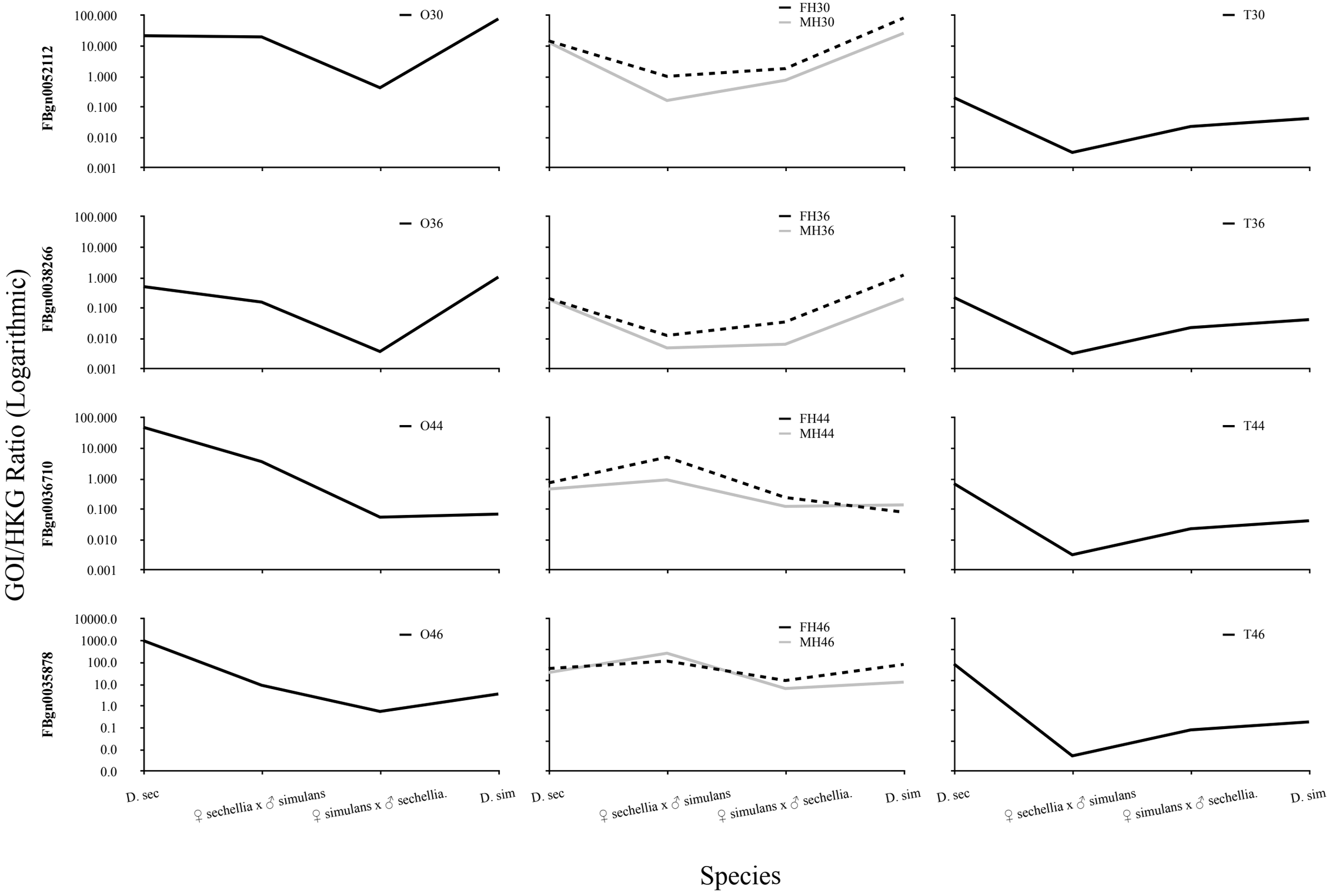
Supplementary Figure §4.2b: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



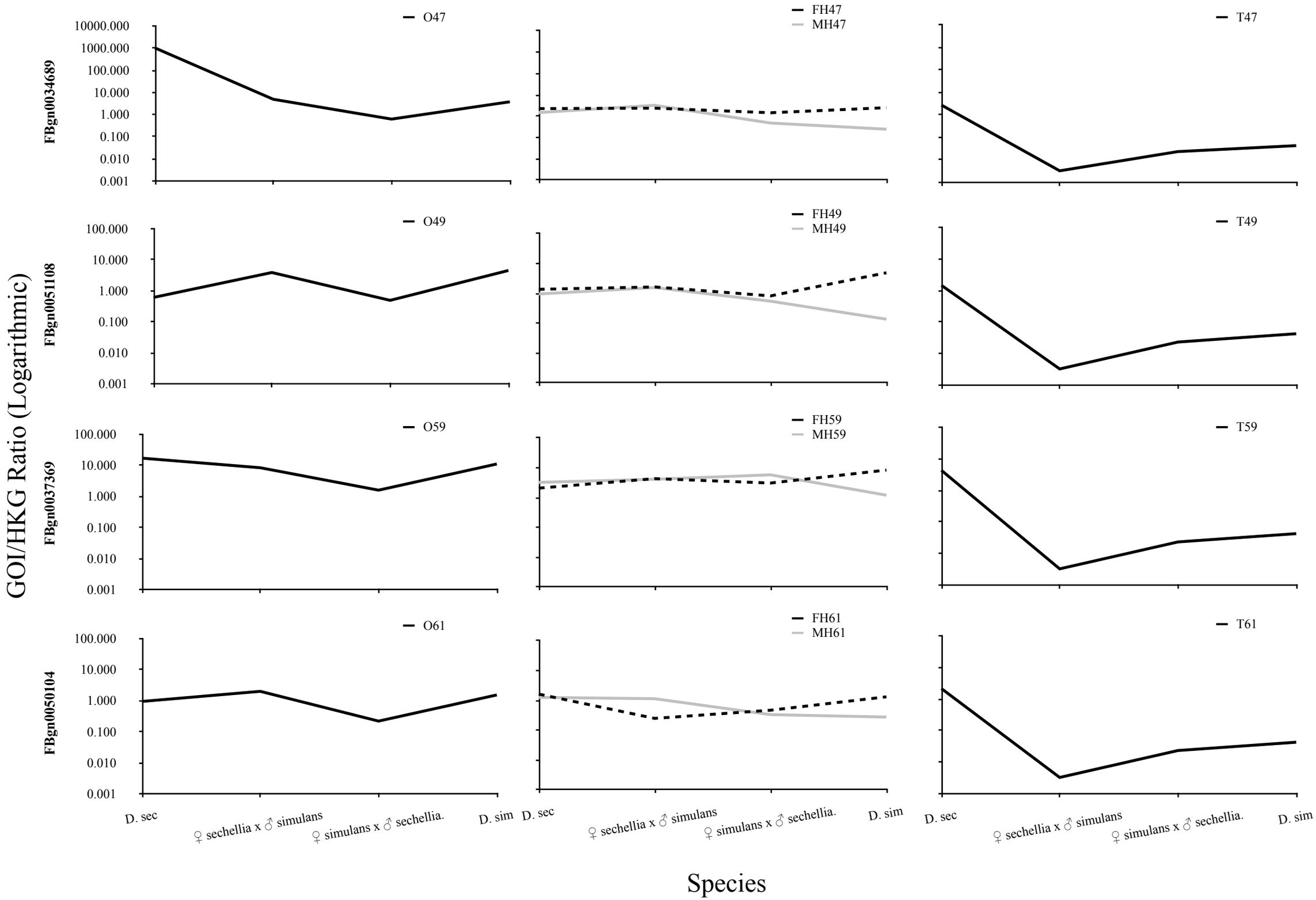
Supplementary Figure §4.3: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



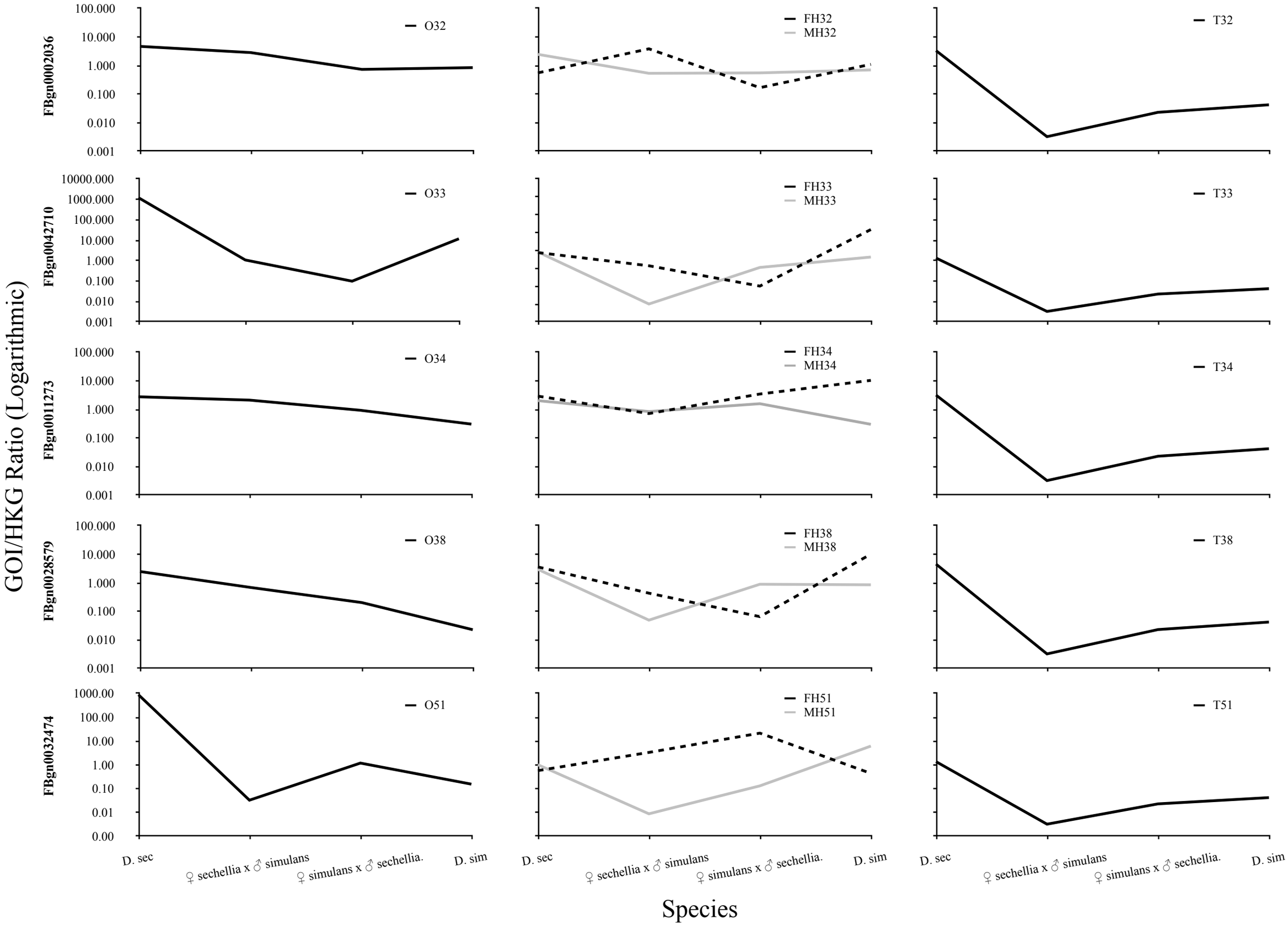
Supplementary Figure §4.4: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



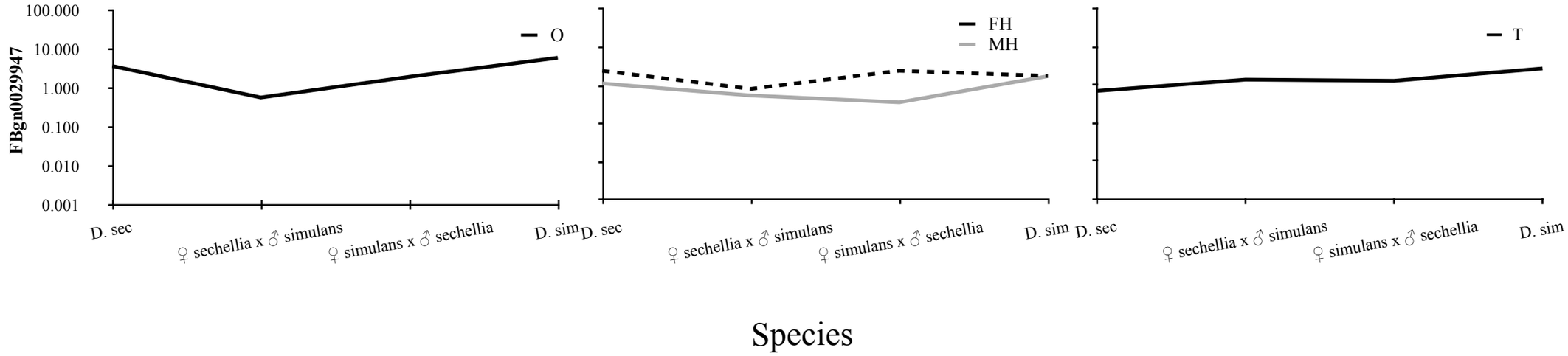
Supplementary Figure §4.5a: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



Supplementary Figure §4.5b: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.

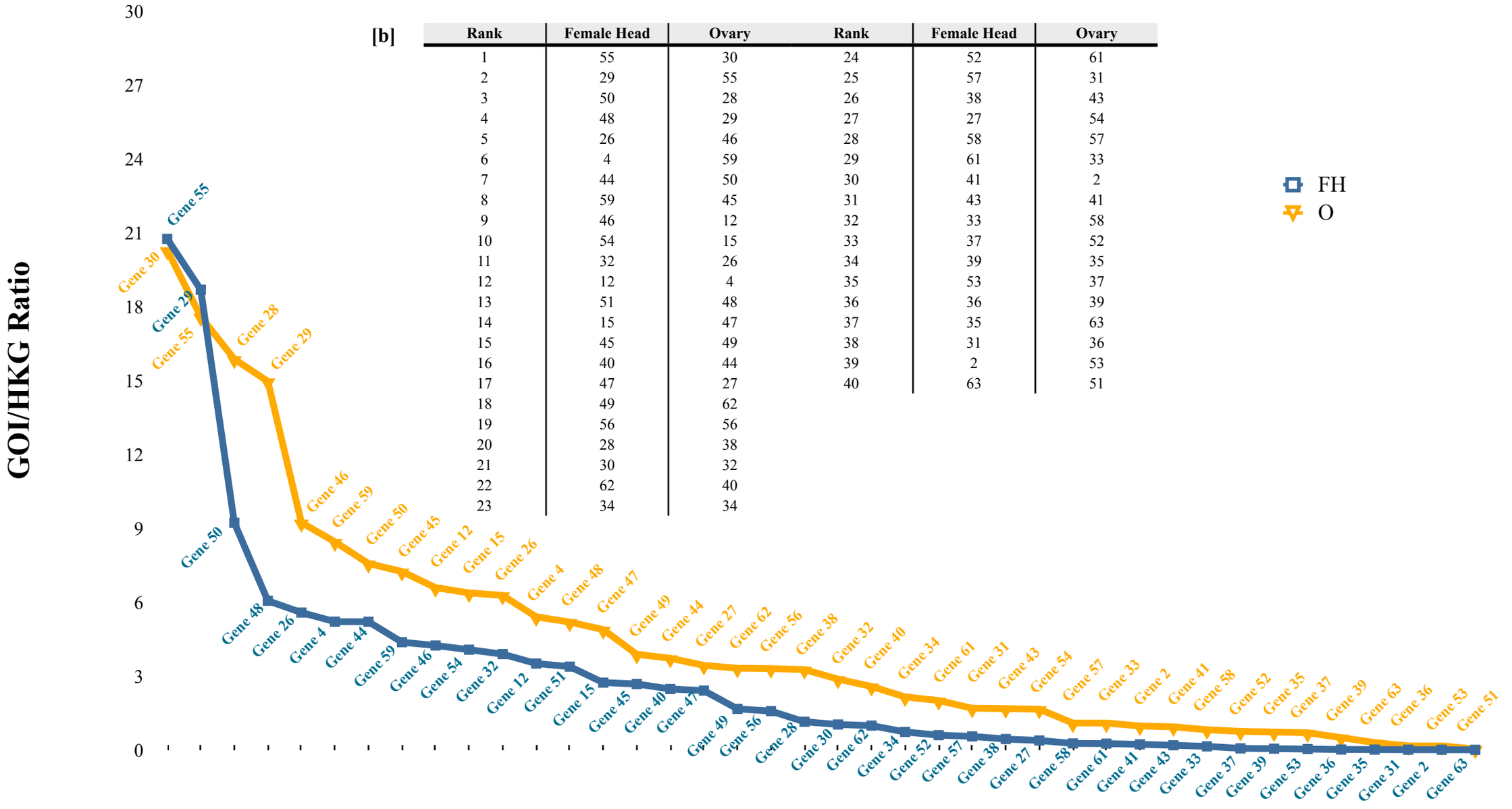


Supplementary Figure §4.6a: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.



Supplementary Figure §4.6b: Expressed via GOI/HKG ratio, testis, male and female head and ovary tissue data for the 40 genes of interest amongst the four species types: parentals, ***D. sechellia*** and ***D. simulans*** (parentals flank the reciprocal hybrids) and the two reciprocal hybrids, ♀ ***sechellia*** **x** ♂ ***simulans*** (hybrid 1) and ♀ ***simulans*** **x** ♂ ***sechellia*** (hybrid 2). Both *D. simulans* and *D. sechellia* data were determined by the observed expression patterns in this study as well as establishing a magnitude cut-off of expression above 1.0 of the logarithmic GOI/HKG ratio illustrated by the red box.

[a]

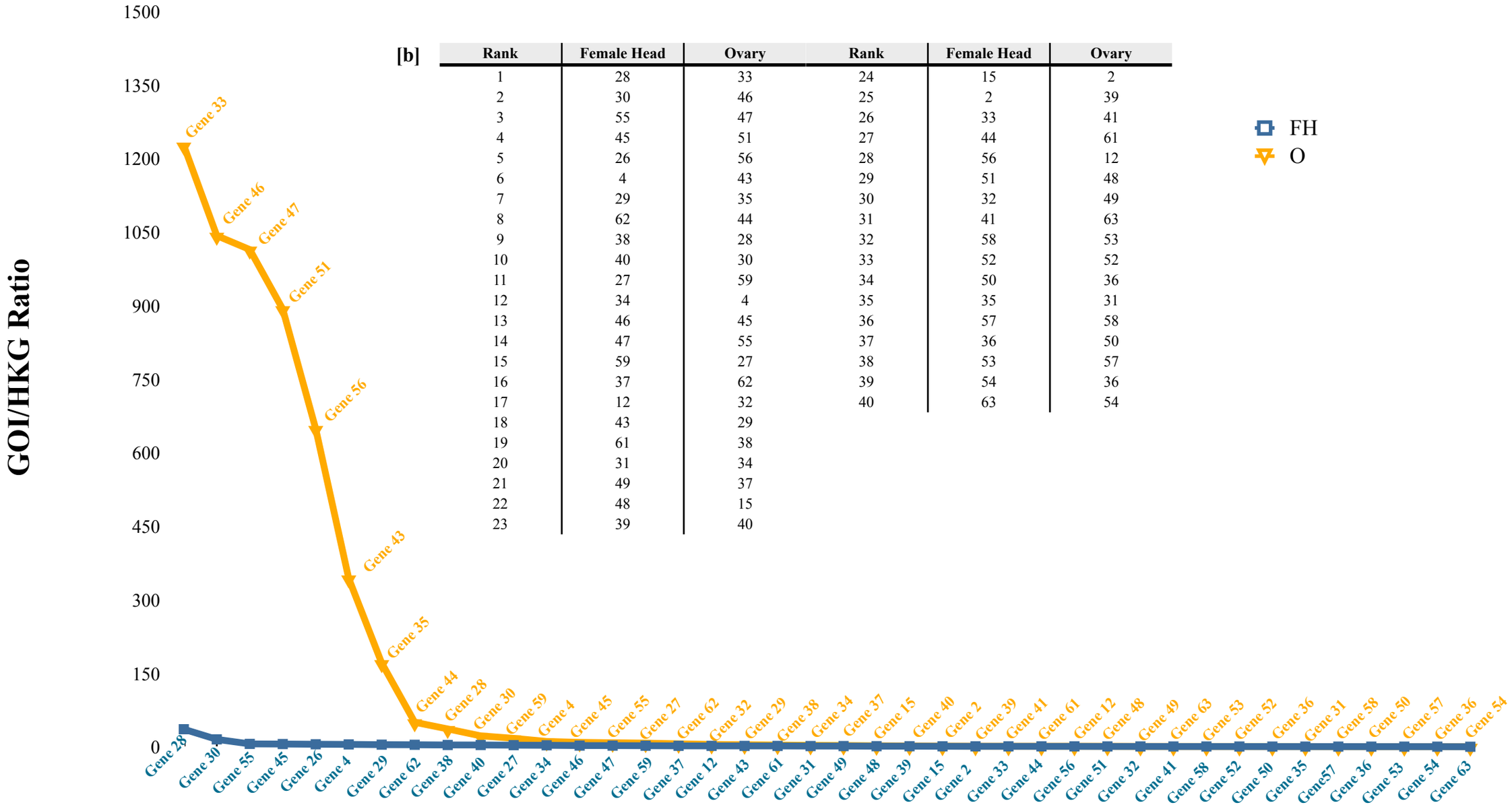


[b]

Rank	Female Head	Ovary	Rank	Female Head	Ovary
1	55	30	24	52	61
2	29	55	25	57	31
3	50	28	26	38	43
4	48	29	27	27	54
5	26	46	28	58	57
6	4	59	29	61	33
7	44	50	30	41	2
8	59	45	31	43	41
9	46	12	32	33	58
10	54	15	33	37	52
11	32	26	34	39	35
12	12	4	35	53	37
13	51	48	36	36	39
14	15	47	37	35	63
15	45	49	38	31	36
16	40	44	39	2	53
17	47	27	40	63	51
18	49	62			
19	56	56			
20	28	38			
21	30	32			
22	62	40			
23	34	34			

Supplementary Figure §4.7: Female head vs. ovary tissue data - ***Drosophila sechellia***. Graphical illustration [a] comparing the relative gene expression of all genes of interest between female sex (O) and non-sex (H) tissues through the use of the GOI/HKG ratio. A table representation [b] aligns the genes of interest — which are arbitrarily assigned a number based on their relative primer generation — between female head and ovary tissues and ranked them from most highly expressed (1) to the lowest expressed gene (40).

[a]

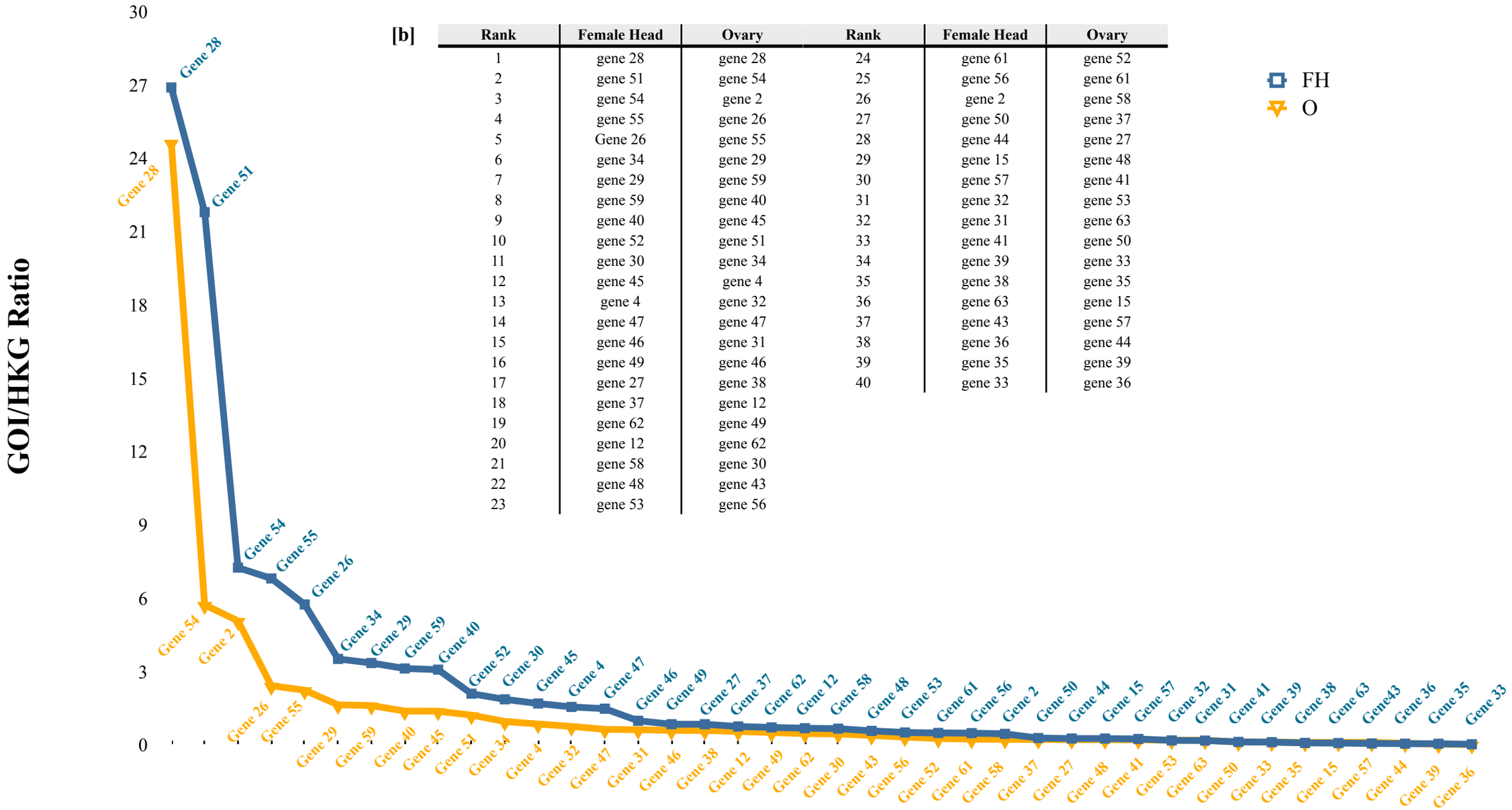


[b]

Rank	Female Head	Ovary	Rank	Female Head	Ovary
1	28	33	24	15	2
2	30	46	25	2	39
3	55	47	26	33	41
4	45	51	27	44	61
5	26	56	28	56	12
6	4	43	29	51	48
7	29	35	30	32	49
8	62	44	31	41	63
9	38	28	32	58	53
10	40	30	33	52	52
11	27	59	34	50	36
12	34	4	35	35	31
13	46	45	36	57	58
14	47	55	37	36	50
15	59	27	38	53	57
16	37	62	39	54	36
17	12	32	40	63	54
18	43	29			
19	61	38			
20	31	34			
21	49	37			
22	48	15			
23	39	40			

Supplementary Figure §4.8: Female head vs. ovary tissue data - ***Drosophila simulans***. Graphical illustration [a] comparing the relative gene expression of all genes of interest between female sex (O) and non-sex (H) tissues through the use of the GOI/HKG ratio. A table representation [b] aligns the genes of interest — which are arbitrarily assigned a number based on their relative primer generation — between female head and ovary tissues and ranked them from most highly expressed (1) to the lowest expressed gene (40).

[a]

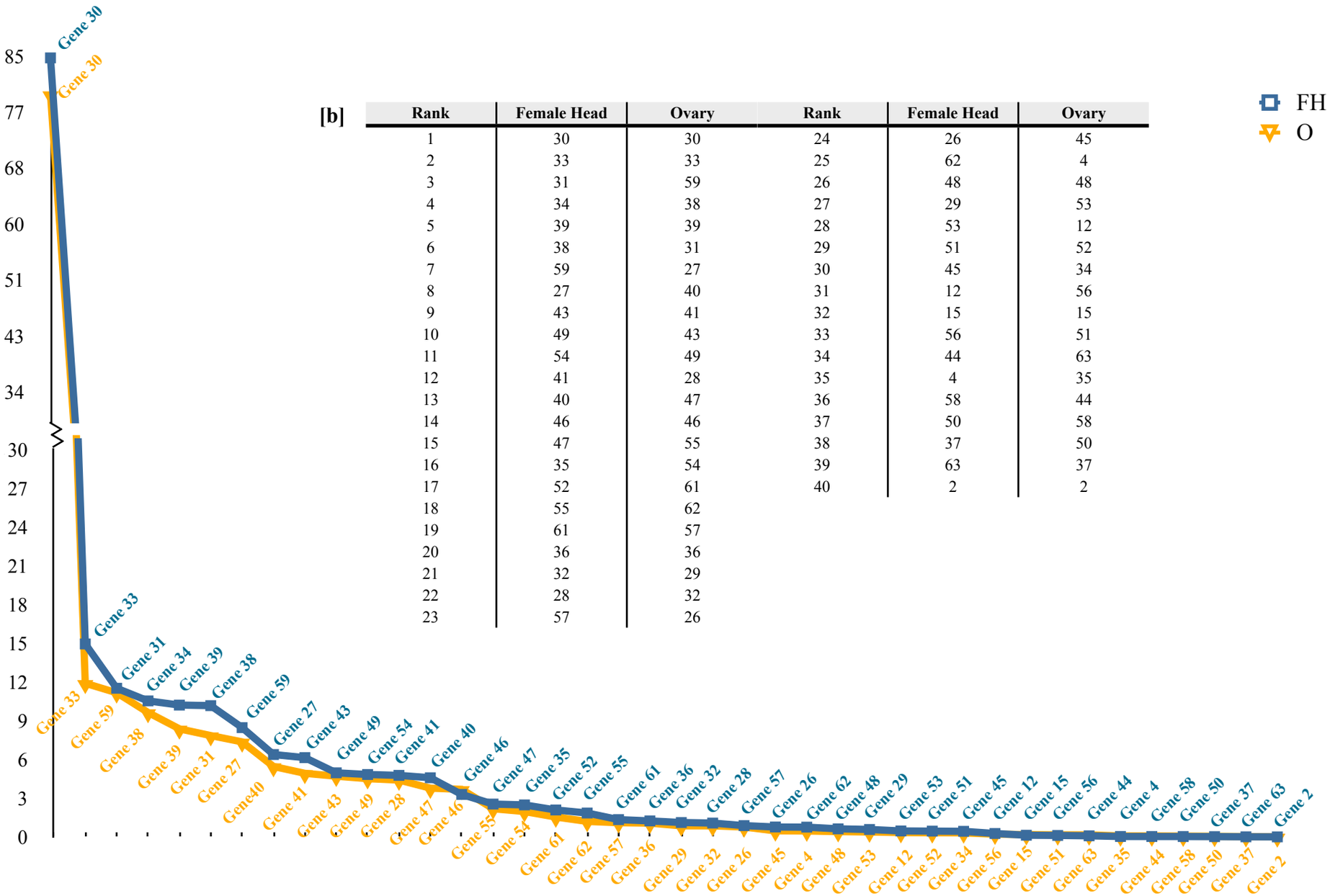


[b]

Rank	Female Head	Ovary	Rank	Female Head	Ovary
1	gene 28	gene 28	24	gene 61	gene 52
2	gene 51	gene 54	25	gene 56	gene 61
3	gene 54	gene 2	26	gene 2	gene 58
4	gene 55	gene 26	27	gene 50	gene 37
5	Gene 26	gene 55	28	gene 44	gene 27
6	gene 34	gene 29	29	gene 15	gene 48
7	gene 29	gene 59	30	gene 57	gene 41
8	gene 59	gene 40	31	gene 32	gene 53
9	gene 40	gene 45	32	gene 31	gene 63
10	gene 52	gene 51	33	gene 41	gene 50
11	gene 30	gene 34	34	gene 39	gene 33
12	gene 45	gene 4	35	gene 38	gene 35
13	gene 4	gene 32	36	gene 63	gene 15
14	gene 47	gene 47	37	gene 43	gene 57
15	gene 46	gene 31	38	gene 36	gene 44
16	gene 49	gene 46	39	gene 35	gene 39
17	gene 27	gene 38	40	gene 33	gene 36
18	gene 37	gene 12			
19	gene 62	gene 49			
20	gene 12	gene 62			
21	gene 58	gene 30			
22	gene 48	gene 43			
23	gene 53	gene 56			

Supplementary Figure §4.9: Female head vs. ovary tissue data - ♀ ***sechellia*** **x** ♂ ***simulans***. Graphical illustration [a] comparing the relative gene expression of all genes of interest between female sex (O) and non-sex (H) tissues through the use of the GOI/HKG ratio. A table representation [b] aligns the genes of interest — which are arbitrarily assigned a number based on their relative primer generation — between female head and ovary tissues and ranked them from most highly expressed (1) to the lowest expressed gene (40).

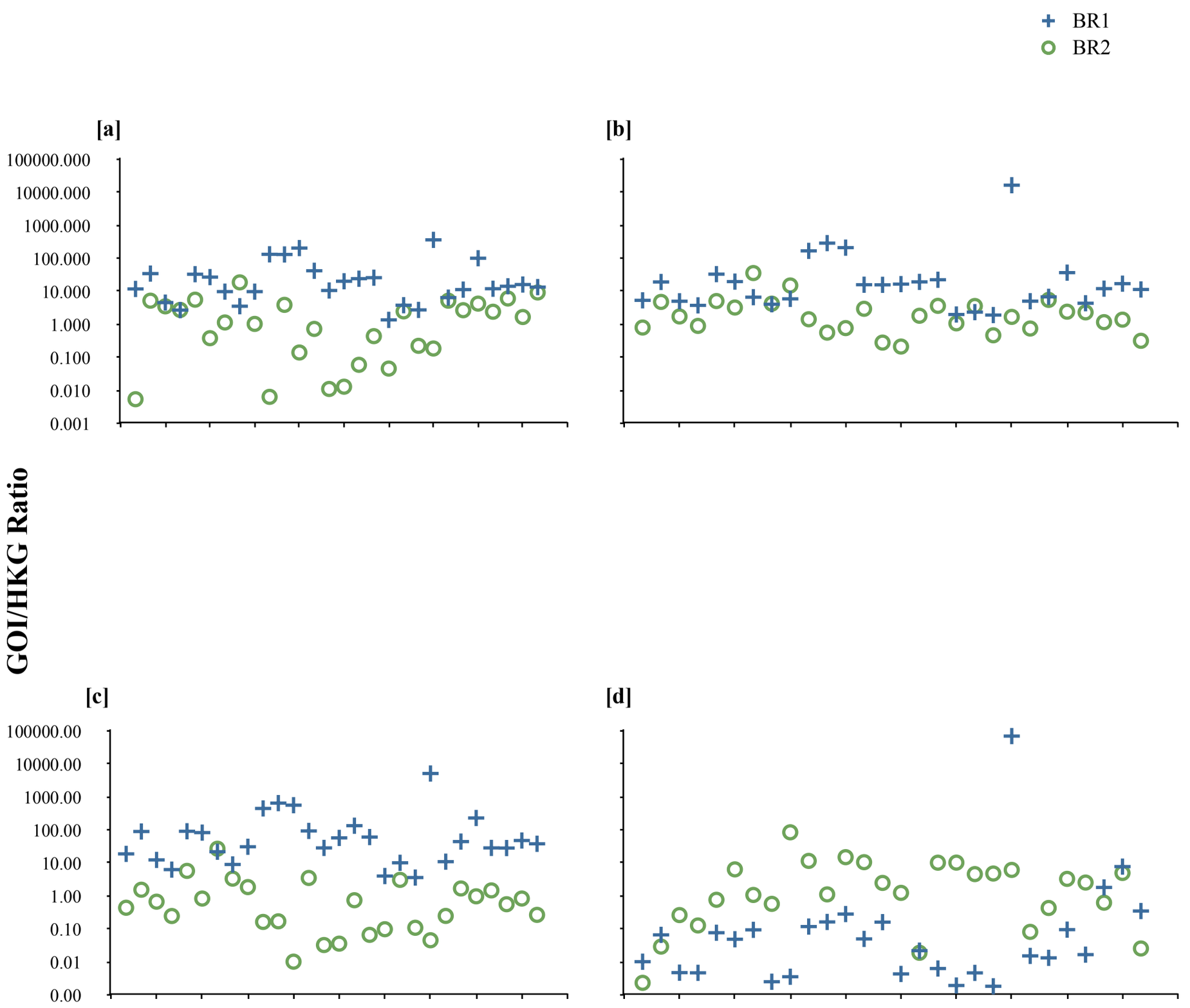
[a]



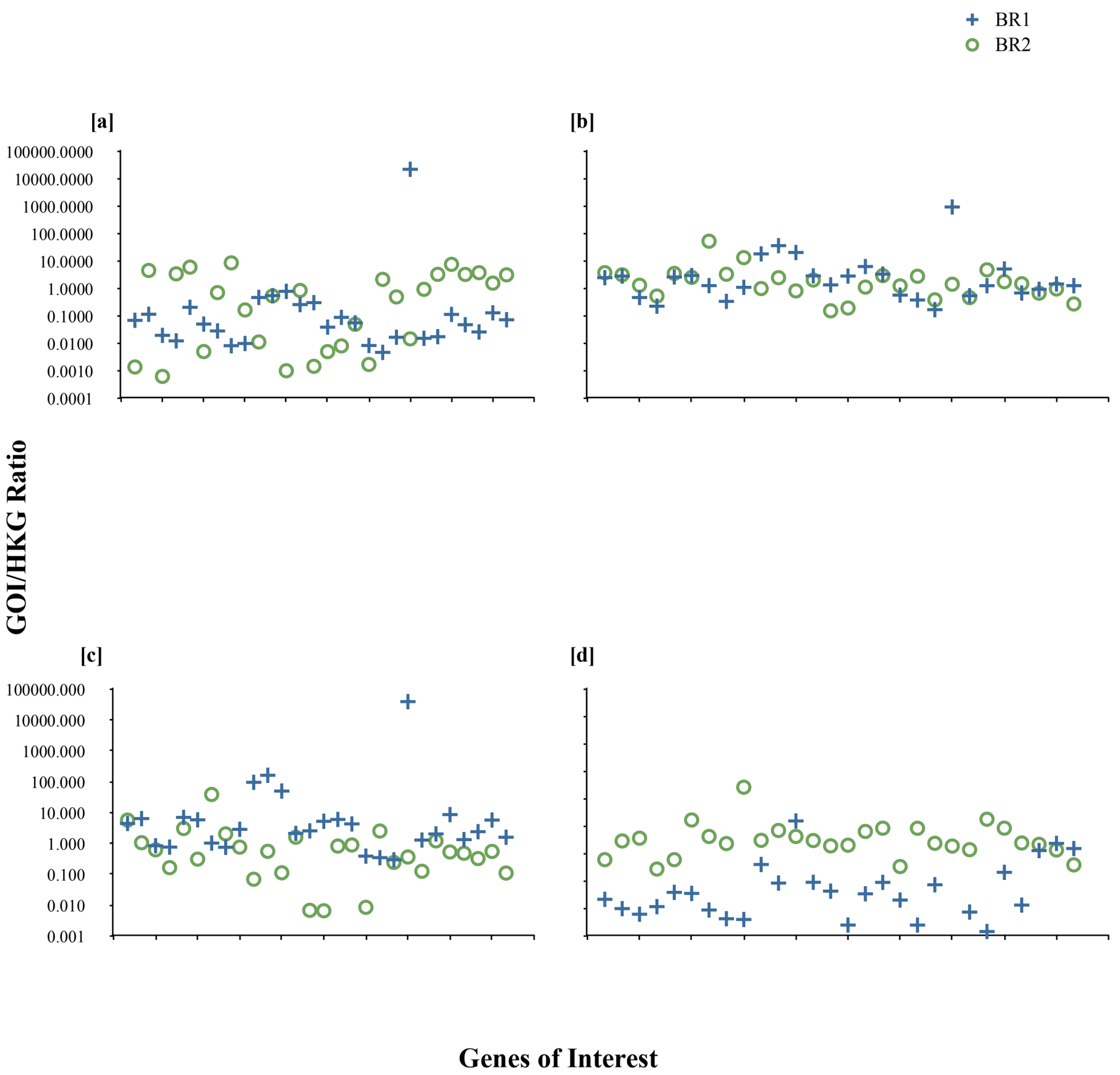
[b]

Rank	Female Head	Ovary	Rank	Female Head	Ovary
1	30	30	24	26	45
2	33	33	25	62	4
3	31	59	26	48	48
4	34	38	27	29	53
5	39	39	28	53	12
6	38	31	29	51	52
7	59	27	30	45	34
8	27	40	31	12	56
9	43	41	32	15	15
10	49	43	33	56	51
11	54	49	34	44	63
12	41	28	35	4	35
13	40	47	36	58	44
14	46	46	37	50	58
15	47	55	38	37	50
16	35	54	39	63	37
17	52	61	40	2	2
18	55	62			
19	61	57			
20	36	36			
21	32	29			
22	28	32			
23	57	26			

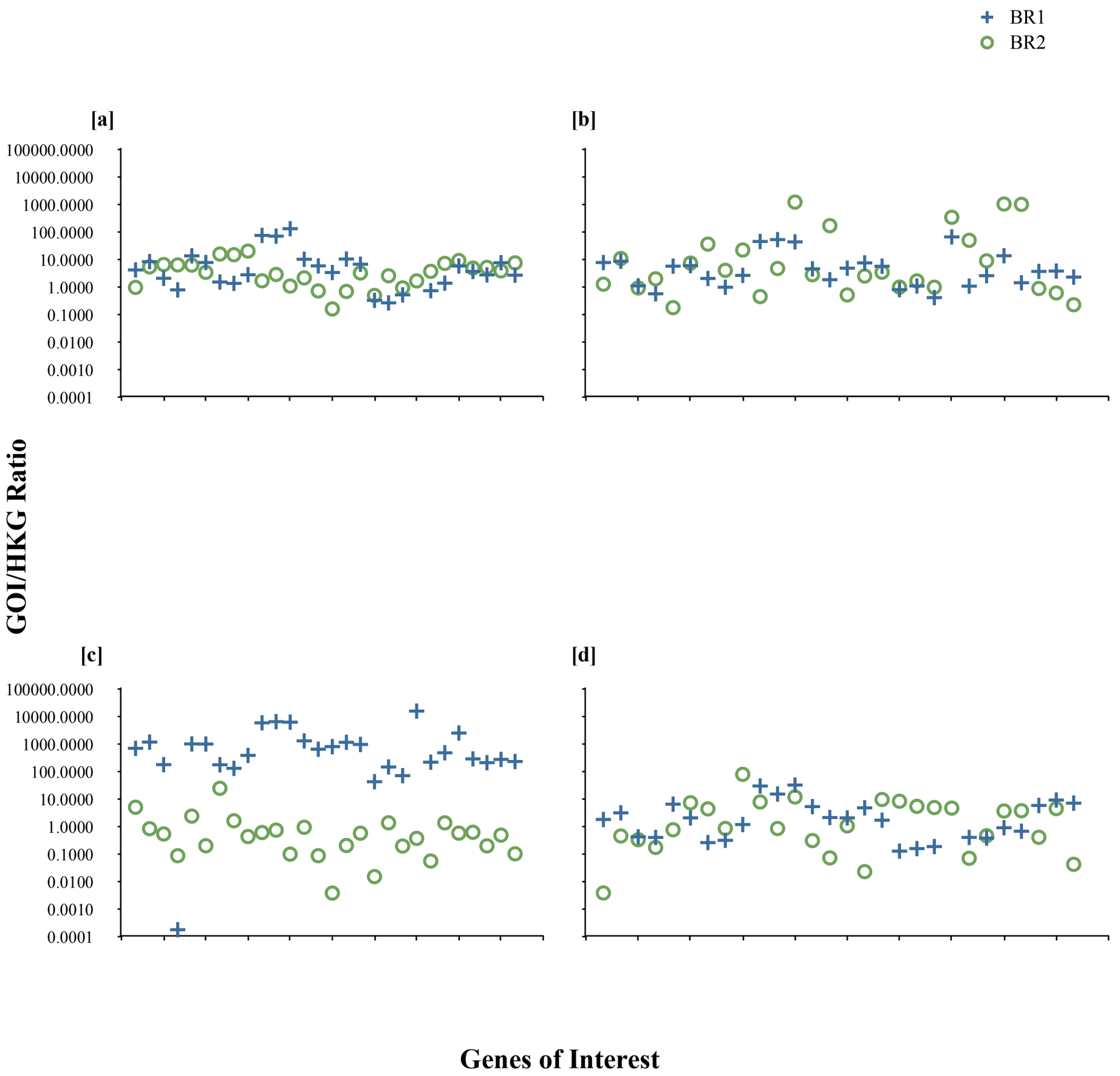
Supplementary Figure §4.10: Female head vs. ovary tissue data - ♀ ***simulans*** **x** ♂ ***sechellia***. Graphical illustration [a] comparing the relative gene expression of all genes of interest between female sex (O) and non-sex (H) tissues through the use of the GOI/HKG ratio. A table representation [b] aligns the genes of interest — which are arbitrarily assigned a number based on their relative primer generation — between female head and ovary tissues and ranked them from most highly expressed (1) to the lowest expressed gene (40).



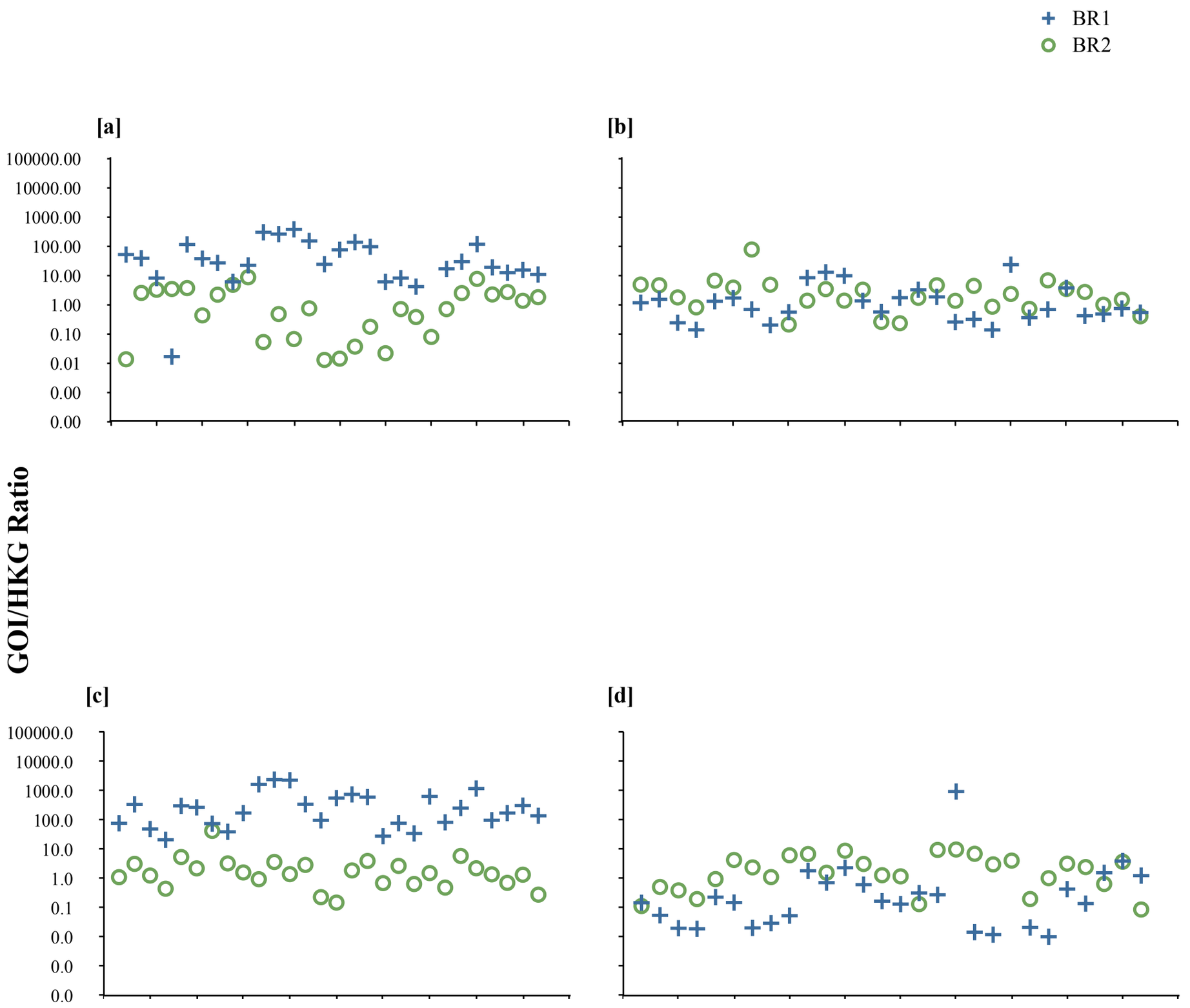
Supplementary Figure §4.11: Plotted comparisons of the biological replicates — Above is the relative gene expression data within the female head tissue for all 40 genes of interest throughout all four species: **[a]** *Drosophila sechellia* **[b]** *D. simulans* **[c]** ♀*simulans* x ♂ *sechellia* and **[d]** ♀ *sechellia* x ♂ *simulans*. The data was represented on a logarithmic scale. (For a list of the genes of interest in the order displayed above refer to **Supplementary Table §.3.3**).



Supplementary Figure §4.12: Plotted comparisons of the biological replicates — Above is the relative gene expression data within the male head tissue for all 40 genes of interest throughout all four species: **[a]** *Drosophila sechellia* **[b]** *D. simulans* **[c]** ♀*simulans* x ♂ *sechellia* and **[d]** ♀ *sechellia* x ♂ *simulans*. The data was represented on a logarithmic scale. (For a list of the genes of interest in the order displayed above refer to **Supplementary Table §.3.3**).



Supplementary Figure §4.13: Plotted comparisons of the biological replicates — Above is the relative gene expression data within the ovary tissue for all 40 genes of interest throughout all four species: **[a]** *Drosophila sechellia* **[b]** *D. simulans* **[c]** ♀*simulans* x ♂ *sechellia* and **[d]** ♀ *sechellia* x ♂ *simulans*. The data was represented on a logarithmic scale. (For a list of the genes of interest in the order displayed above refer to **Supplementary Table §.3.3**).



Supplementary Figure §4.14: Plotted comparisons of the biological replicates — Above is the relative gene expression data within the testis tissue for all 40 genes of interest throughout all four species: **[a]** *Drosophila sechellia* **[b]** *D. simulans* **[c]** ♀*simulans* x ♂ *sechellia* and **[d]** ♀ *sechellia* x ♂ *simulans*. The data was represented on a logarithmic scale. (For a list of the genes of interest in the order displayed above refer to **Supplementary Table §.3.3**).

Supplementary Table § 3.1a: Comparison of misexpression in male and female head tissues among genes of interest from *D. melanogaster* microarray data and those gene expression patterns found through qRT-PCR using the reciprocal F1 cross between *D. simulans* and *D. sechellia*

FlyBase ID	Chromosome allocation	Tissue Type (Based on *D. mel.* ESTs)	Misregulation +/- [Microarray Data]	Sex	Misregulation +/- [qRT-PCR]	
FBgn0030706	**X**	Head	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: +**
FBgn0032702	**2L**	Head	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: +**	**Hy2: +**
FBgn0020392	**3L**	Ovary	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: -**	**Hy2: -**
FBgn0038964	**3R**	Ovary	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: -**	**Hy2: -**
FBgn0032821	**2L**	Testis	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0034532	**2R**	Testis	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: +**	**Hy2: +**
FBgn0038607	**3R**	TH	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: +**	**Hy2: +**
FBgn0028902	**2L**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0052112	**3L**	TH	+	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0031853	**2L**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0002036	**2L**	TH	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: -**	**Hy2: -**
FBgn0042710	**3R**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0011273	**3R**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0032973	**2L**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0038266	**3R**	TH	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: +**	**Hy2: -**
FBgn0029947	**X**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0029947	**X**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0028579	**2R**	TH	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: -**	**Hy2: -**
FBgn0036763	**3L**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**

Supplementary Table § 3.1b: Comparison of misexpression in male and female head tissues among genes of interest from *D. melanogaster* microarray data and those gene expression patterns found through qRT-PCR using the reciprocal F1 cross between *D. simulans* and *D. sechellia*

FlyBase ID	Chromosome allocation	Tissue Type (Based on *D. mel.* ESTs)	Misregulation +/- [Microarray Data]	Sex	Misregulation +/- [qRT-PCR]	
FBgn0038334	**3R**	TH	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0039336	**3R**	OT	+	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0031227	**2L**	OT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0036710	**3L**	OT	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: +**	**Hy2: +**
FBgn0015562	**2R**	OT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0035878	**3L**	HOT	+	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0034689	**2R**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0022338	**3R**	HOT	+	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0051108	**3R**	HOT	+	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: +**	**Hy2: +**
FBgn0034176	**2R**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0032474	**2L**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0011604	**2R**	HOT	+	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0015799	**X**	HOT	+	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0032248	**2L**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0032451	**2L**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0034689	**2R**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: +**	**Hy2: +**
FBgn0034689	**2R**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: +**	**Hy2: +**
FBgn0039360	**3R**	HOT	+	F	**Hy1: +**	**Hy2: +**
				M	**Hy1: -**	**Hy2: -**

Supplementary Table § 3.1c: Comparison of misexpression in male and female head tissues among genes of interest from *D. melanogaster* microarray data and those gene expression patterns found through qRT-PCR using the reciprocal F1 cross between *D. simulans* and *D. sechellia*

FlyBase ID	Chromosome allocation	Tissue Type (Based on *D. mel.* ESTs)	Misregulation +/- [Microarray Data]	Sex	Misregulation +/- [qRT-PCR]	
FBgn0037369	**3R**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0050104	**2R**	HOT	+	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0034158	**2R**	OH	+	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**
FBgn0030574	**X**	NMG	**-**	F	**Hy1: -**	**Hy2: -**
				M	**Hy1: -**	**Hy2: -**
FBgn0031988	**2L**	NMG	**-**	F	**Hy1: +**	**Hy2: -**
				M	**Hy1: +**	**Hy2: -**

Supplementary Table § 3.2: Comparison of misexpression within the male and female gonadal tissue among genes of interest from *D. melanogaster* microarray data and those gene expression patterns found through qRT-PCR using the reciprocal F1 cross between *D. simulans* and *D. sechellia*

FlyBase ID	Chromosome allocation	Tissue Type (Based on *D. mel.* ESTs)	Misregulation +/- [Microarray Data]	Sex	Misregulation +/- [qRT-PCR]	
FBgn0030706	**X**	Head	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** +
FBgn0032702	**2L**	Head	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** +	**Hy2:** +
FBgn0020392	**3L**	Ovary	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** -	**Hy2:** -
FBgn0038964	**3R**	Ovary	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** -	**Hy2:** -
FBgn0032821	**2L**	Testis	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0034532	**2R**	Testis	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** +	**Hy2:** +
FBgn0038607	**3R**	TH	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** +	**Hy2:** +
FBgn0028902	**2L**	TH	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0052112	**3L**	TH	+	O	**Hy1:** +	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0031853	**2L**	TH	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0002036	**2L**	TH	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** -	**Hy2:** -
FBgn0042710	**3R**	TH	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0011273	**3R**	TH	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0032973	**2L**	TH	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0038266	**3R**	TH	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** +	**Hy2:** -
FBgn0029947	**X**	TH	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0028579	**2R**	TH	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** -	**Hy2:** -

Supplementary Table § 3.2b: Comparison of misexpression within the male and female gonadal tissue among genes of interest from *D. melanogaster* microarray data and those gene expression patterns found through qRT-PCR using the reciprocal F1 cross between *D. simulans* and *D. sechellia*

FlyBase ID	Chromosome allocation	Tissue Type (Based on *D. mel.* ESTs)	Misregulation +/- [Microarray Data]	Sex	Misregulation +/- [qRT-PCR]	
FBgn0028579	**2R**	TH	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** -	**Hy2:** -
FBgn0036763	**3L**	TH	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0038334	**3R**	TH	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0039336	**3R**	OT	+	O	**Hy1:** +	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0031227	**2L**	OT	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0036710	**3L**	OT	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** +	**Hy2:** +
FBgn0015562	**2R**	OT	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0035878	**3L**	HOT	+	O	**Hy1:** +	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0034689	**2R**	HOT	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0022338	**3R**	HOT	+	O	**Hy1:** +	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0051108	**3R**	HOT	+	O	**Hy1:** +	**Hy2:** -
				T	**Hy1:** +	**Hy2:** +
FBgn0034176	**2R**	HOT	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0032474	**2L**	HOT	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0011604	**2R**	HOT	+	O	**Hy1:** +	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0015799	**X**	HOT	+	O	**Hy1:** +	**Hy2:** -
				T	**Hy1:** +	**Hy2:** -
FBgn0032248	**2L**	HOT	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0032451	**2L**	HOT	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** -	**Hy2:** -
FBgn0034689	**2R**	HOT	+	O	**Hy1:** -	**Hy2:** -
				T	**Hy1:** +	**Hy2:** +
FBgn0039360	**3R**	HOT	+	O	**Hy1:** +	**Hy2:** +
				T	**Hy1:** -	**Hy2:** -

Supplementary Table § 3.2c: Comparison of misexpression within the male and female gonadal tissue among genes of interest from *D. melanogaster* microarray data and those gene expression patterns found through qRT-PCR using the reciprocal F1 cross between *D. simulans* and *D. sechellia*

FlyBase ID	Chromosome allocation	Tissue Type (Based on *D. mel.* ESTs)	Misregulation +/- [Microarray Data]	Sex	Misregulation +/- [qRT-PCR]	
FBgn0039360	**3R**	HOT	+	O	**Hy1: +**	**Hy2: +**
				T	**Hy1: -**	**Hy2: -**
FBgn0037369	**3R**	HOT	+	O	**Hy1: -**	**Hy2: -**
				T	**Hy1: -**	**Hy2: -**
FBgn0050104	**2R**	HOT	+	O	**Hy1: -**	**Hy2: -**
				T	**Hy1: -**	**Hy2: -**
FBgn0034158	**2R**	OH	+	O	**Hy1: +**	**Hy2: -**
				T	**Hy1: +**	**Hy2: -**
FBgn0030574	**X**	NMG	-	O	**Hy1: -**	**Hy2: -**
				T	**Hy1: -**	**Hy2: -**
FBgn0031988	**2L**	NMG	-	O	**Hy1: +**	**Hy2: -**
				T	**Hy1: +**	**Hy2: -**

Supplementary Table § 3.3: Primers — Complete list of investigated genes and their forward and reverse primers with each genes FlyBase ID and primer CODE (the number assigned to each primer set in this study).

FlyBase ID	Forward Primer	Reverse Primer	Primer CODE
FBgn0030706	GATTTCATGGACGTCGTTTTG	CCAAATTAAACATCGGCAAAGA	2
FBgn0032702	ACCTATGCAACCAGCCAACT	CGAAGGCATTGCGATAAAAG	4
FBgn0020392	GTGTACAATGCCCTCGACCT	CAGCGCCAATTGTAGAGGTA	12
FBgn0038964	GCGCTCTTCCGTGCTCTAAA	ATGCGACCCTTGTTCTTCAG	15
FBgn0032821	GCTGAGCAGCGTGATATGAA	GGTAAAGTGGCAGCTCGATT	26
FBgn0034532	GTGGGGCAATCTGAACAACT	TTGTCAGGGTCATCAAGTCG	27
FBgn0038607	GGAAACGACTATGACCAGTGC	AGATCGCGTCGTGAGTAGTG	30
FBgn0028902	AGCGTATTAAGCTGCGGAAA	CCAACTCGGTATCTGGTCGT	31
FBgn0052112	CTGCATTATTGAGGCTTTCG	TCTCTTTATCAACTGCTGCTCATAGGTT	32
FBgn0031853	ACGGATTCAAGCAGAACCTG	AGCATGAAGTCGCAACCGTA	33
FBgn0002036	CGAGATCTTCCAATCGCTGT	TCCTGCTGCACTGGACACTT	34
FBgn0042710	AACGATACGGTGGGCACTCT	TTCTCCACATAGCACGCATT	35
FBgn0011273	CAACAACAATGGCCAACTGA	AAAATCTTGAATGCCTCACG	36
FBgn0032973	ATGGAGGCATTGTGGCTTTA	GCATTCTTCACATCCCGATT	37
FBgn0038266	AATCGCCAGGATGAGCTTGT	CTCCTCTTCGGTGAGAATGG	38
FBgn0029947	CGAGCGTTCGTCTTATCCTC	CGAATCTGAGGCCATAGTCC	39
FBgn0028579	ACCTTACCTTGGAGGCCCTA	CACACTGCGGTACTTGTGCT	40
FBgn0036763	AGCAATCTGGTCAACATCCA	CGCCACACGGTCCTTATACT	41
FBgn0038334	GCTGATCATGGAGAAGTACCG	TTCGACCACCTTCGACAAAC	43
FBgn0039336	TTGTTATTCCCCGGAAACTG	ATTCCAGGCACTCGGTAGAG	44
FBgn0031227	GGCGATTTGTCTACCGATGT	TGGGGGATTGTTTTCGTCTA	45
FBgn0036710	CTCGGCTGTTTTGCGTAATC	GGTCAAAGAGATCAATTATGTG	46
FBgn0015562	CATTGTGGACTACGCAGTGG	TGCACGTAAGTGTAGGGGAAC	47
FBgn0035878	GACAAGGAGGAGCTGGTGTT	GTCACCTGATAGGCGCTGAG	48
FBgn0034689	GAGTGAGCATGCGGACTACA	TGTCCAGAAATGCGAGATTG	49
FBgn0022338	CTAGTCCTCGATGCCGATCT	CGAGGGCTGTTGGTTACTTG	50
FBgn0051108	CGTTGTCCGCTCTTTTAAGG	AGCACAGGCTAGCACAGCTT	51
FBgn0034176	CGTCCAGGTGGTGGATAAAC	GCAGTCCCTTGAAAGTGCTC	52
FBgn0032474	AGCACTTGGATGGACGAAAC	AAAAACAGGCATACCACTGC	53
FBgn0011604	CAATTCCGCTTTGACTGGTT	TCCAGCTCAATGTTTTCACG	55
FBgn0015799	GCACCTCTGGGACACCTCTA	GATGATCAACGCCGAGTACA	56
FBgn0032248	TGACATTGGTGCCGACTATT	CACTGCAGCTGTGGGATAAG	57
FBgn0032451	CCTCCAGATTCCTCAACGAG	TTCTCTTTGGGCGAGTGAAT	58
FBgn0034689	GTAATGGCCCGACATAAAGG	TCGGAACCCAGCAAAATATC	59
FBgn0039360	GAACGCTAGAAGGCTCATGC	CATACCCAACATGCTGACCA	61
FBgn0037369	GGAATGCCCATCAGAATCTC	CATATTTACCCCCGAGCTGA	62
FBgn0050104	CAATGTCGTCCTGGGAGAGT	CGCCTCGAGATCGTTATTTT	63
FBgn0030574	CACCGCCTGGATATTCTACG	AATGATGATCGGCACCTTGT	28
FBgn0031988	GGACCCGCCTACGTACTCAC	TACTATGCCCGTGGTGAACA	29
FBgn0034158	GCCAACTCGCTGGTGTATTT	TACCATCCATCCAATGGACA	54