THE GENETIC BASIS OF SEXUAL DIMORPHISM IN *DROSOPHILA* AND PRIMATES

A Thesis
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
MASTER OF SCIENCE

---

by
Nichole Howard Rigby
July 2016

Thesis Approvals:

Rob J. Kulathinal, Thesis Advisor, Department of Biology
David Liberles, Department of Biology

ABSTRACT

Sexual dimorphism, i.e., differences in morphology, physiology, and behavior between conspecific males and females, is ubiquitous, extensive, and often species-specific, indicative of its rapidly evolving nature. Ever since Darwin first described a general theory of sexual selection to explain the extraordinary differences between males and females of the same species, biologists have proposed a variety of mechanisms ranging from runaway selection to good genes to sexual conflict. While a popular approach is studying the effects of sexual selection on different components of fitness, the results of these studies are generally difficult to interpret and are typically not generalizable across populations, let alone taxa.

Recent advances in the "omics" field are transforming the way that we study patterns and processes involved in sexual selection. At the molecular level, sexual dimorphism is present in gene expression differences between the sexes, providing a powerful framework to study sexual selection. By studying genes that are sex-biased in expression, we will better understand the underlying genetic basis of traits that are sexually dimorphic. Already, studies of sex-biased genes in model organisms, particularly *Drosophila*, have revealed that male-biased genes are among the most rapidly evolving functional classes of genes. However, while a number of intrinsic factors appear to correlate with evolutionary rate (e.g., gene expression level, codon bias), it is unclear whether any of these factors drive the rapid divergence of male-biased genes. Another important discovery is the prevalence of sex-biased gene expression. However, even with widespread sexual dimorphism at the phenotypic level, it remains unknown the extent to which sex-biased gene expression exists in humans and their primate relatives. In fact,

studies of sexual dimorphism on a molecular level in primates have been very few, even though understanding this phenomenon in humans could further our knowledge of the nature of sex-biased phenotypes and diseases.

In this thesis, I advance our knowledge of the genetic bases and mechanisms that shape sexual dimorphism. First, I review a classic framework that biologists have traditionally applied to define and partition fitness measures between males and females in the model system, *Drosophila*. Second, I apply a molecular framework to compare the relative roles of intrinsic factors on the evolutionary rate of rapidly evolving male-biased genes in *Drosophila*. Third, I review the current state of our knowledge of sexual dimorphism and sex-biased gene expression in humans. Fourth, I present a bioinformatics framework to identify the extent of sex-biased expression in primate tissue and to examine the selective forces involved in their evolution. Overall, I demonstrate the effectiveness of using a functional comparative genomics approach in studying the nature of sexual dimorphism at the molecular level across multiple taxa.

ACKNOWLEDGMENTS

First, I would like to thank my thesis advisor Rob Kulathinal of the Department of Biology at Temple University for guidance and support. I would also like to thank David Liberles of the Department of Biology at Temple University as the second reader of this thesis. I am gradeful for his valuable comments on this work.

Second, I owe many thanks to all of the individuals who assisted me in this work. I thank Coryandar Gilvary and Craig Stanley, Jr. for assistance in data acquisition, mining, and transformation. I also thank Jessica Chin and Nancy Dordal for their dedication to literature searches and data quality assurance. I am very grateful to Fang Zhou and Claudia Weber for help with data analysis in this work. I would also like to thank Christie Rockwell and Nora Engel for their valuable feedback on my work with human and primate data. I thank all of the faculty and staff in the Department of Biology at Temple University, including Ralph Hillman, Richard Waring, Evelyn Vleck, Toni Matthews, Regee Neely, Brenda Mims, Tania Stephens, Derrick Love, and Steve Widing.

I thank my fellow graduate students for their moral support, especially Alanna Durkin, Samuel Georgian, Craig Stanley, Jr., Bobby Daugherty, Carlee Cunningham, and Stephen Lengyel. I am grateful to have had such a wonderful cohort of Carlos Gomez, Joseph Kelich, Steven Schnell, and Krishna Mudumbi. It has been a privledge to know all of you.

I would like to acknowledge my funding sources, which made this work possible. I thank Temple University for support though teaching assistantships and university scholarships. I would also like to thank the National Science Foundation, through which my research assistantship was funded.

Last, but of course not least, my deepest graditude goes to my family. The support, love and guidance of my parents have motivated me to pursue my goals and given me the confidence to be who I am today. My husband, Graham Rigby, has truly been my rock throughout this endeavor. This work would not have been possible without his endless love, patience, and encouragement. Also I must thank my English Bulldog, Beefcake, who layed on the couch with me, and kept me company all throughout the writing process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 The prevalence and causes of sexual dimorphism

Males and females can be strikingly different in appearance and behavior. This phenomenon, referred to as sexual dimorphism (SD), arose over a billion years ago with the rise of anisogamy, and has remained a defining characteristic across much of life. Although it has become common among taxa, SD can also be quite species-specific, reflecting its rapid evolutionary nature. The existence of phenotypic SD has inspired a large body of work aimed at providing explanations for its origin, prevalence, and consequences.

Darwin initially introduced an explanation for the existence and maintence of SD when he attempted to explain the existence of elaborate secondary sexual characteristics of male animals (Darwin, 1871). He proposed that SD is the result of sexual selection, a type of natural selection that affects an organism's ability to acquire mates. Like natural selection, sexual selection can influence the fitness of individuals, which can confound methods for estimating fitness for males and females. Darwin explained fitness as an organism's ability to survive and reproduce (Darwin, 1859). Yet, males and females have differing strategies for pursuing mates and survivorship, making the study of fitness with relation to reproductive success and its associated costs and benefits, highly nuanced.

Several hypotheses can explain the evolution of SD and demonstrate the complexity of individual fitness. Female mate choice, for example, is a mechanism resulting from a female trait that biases mating towards males with specific desired

characters (Maynard-Smith, 1987; Kirkpatrick and Ryan, 1991; Andersson, 1994). As females in a population continue to choose males with these traits, the desired traits become more prevalent and pronounced among males in the population. This also results in differential selection on sex-specific traits, promoting the evolution of SD. Because mate choice may be energetically costly for females (Alatalo et al., 1987; Pomiankowski, 1987; Gibson and Bachman, 1992; Wigby and Chapman, 2005), this will only increase female fitness if costs are outweighted by benefits gained from the mating (Friberg and Arnqvist, 2003). These fitness costs can be remedied by direct benefits gained from mating (e.g., improved resource control or improved paternal care; Price et al., 1993), or indirect benefits (e.g., 'good genes'; Houle and Kondrashov, 2002). Sexual conflict – a conflict between the evolutionary interests of males and females – can also produce differential selection between the sexes (Parker, 1979; Parker, 2006), resulting in complex fitness differences. Analyses that examine changes in fitness in a population must consider appropriate measures or proxies for fitness, given the selective forces acting on males and females in a system.

1.2 Sexual dimorphism at the molecular level

In populations with SD, males and females appear phenotypically different, despite a common genome. Selective pressures acting on the phenotypes of organisms can result in signatures of change at the molecular level. Sexually dimorphic traits are the product of differential expression of genes between the two sexes i.e., sex-biased genes (Ellegren and Parsch, 2007), that can be readily identified by examining differences in expression level between males and females via RNA-seq or microarray technology. This

approach, combined with methods for estimating the extent of selection on genes and their associated proteins (e.g., dN/dS or K_a/K_s) can provide insight into the evolution of SD on a molecular scale.

Such genomic analyses have provided valuable insight to a number of commonly observed trends among sex-biased genes. First, sex-biased gene expression appears to be pervasive, existing in fruit flies (Jin et al., 2001; Parisi et al., 2003; Ranz et al., 2003), nematodes (Reinke et al., 2000; Thoemke et al., 2005), and mice (Yang et al., 2006). Second, sex-biased appears to be highly tissue-specific, with reproductive tissues exhibiting the largest proportion of sex-biased genes (Ellegren and Parsch, 2007). Third, sex-biased genes, especially those that are male-biased, exhibit higher rates of evolution (i.e., dN/dS) than unbiased genes (Ellegren and Parsch, 2007). Yet, the cause(s) of this rapid evolution of male-biased genes and male reproductive-related proteins remain unclear.

To understand why male gene's rapidly evolve, one approach would be to examine the relative contribution of intrinsic factors that constrain and promote protein evolution in this class of genes. Intrinsic factors refer to the selective pressures and/or constraints that are present at the molecular level such as those found in the transcriptional and translational machinery. A "core set" of intrinsic factors have been previously identified as potential correlates of rapid evolution in protein sequences (Larracuente et al., 2008). A major factor is gene expression level, which has been shown to have a strong negative correlation with dN/dS (Drummond et al., 2005; Drummond et al., 2006; Popescu et al., 2006; Lemos et al., 2005). This is likely driven by purifying

selection, assuming that highly expressed genes may experience increased non-specific binding and translational errors (Chi and Liberles, 2016). Another major factor that commonly correlates with evolutionary rate is the codon adaptation index (CAI). CAI consistently appears to be negatively correlated with dN/dS of protein sequences (Pal, 2001; McInerney, 2006; Drummond et al., 2006), potentially due to selection on translational efficiency (McInerney, 2006). By examining the contributions of these intrinsic factors and others (e.g., functional specificity, protein-protein interactions, gene length, chromosome location) to protein evolution, it may be possible to reveal the correlates of rapid evolution in male-biased genes on a molecular level.

While Drosophila and its extensive set of genomic resources is an excellent model to understand these patterns of sex-bias, primates and particularly humans have lacked in both resources and analyses. This knowledge gap in human sex-biased gene expression patterns is concerning given that many diseases and complex phenotypes have significant sex interaction effects.

1.3 Objectives

Understanding how SD evolves at the molecular level can provide valuable insight into the mechanisms and forces that have shaped sexually dimorphic characters in the recent past. I present four chapters highlighting phenotypic and molecular approaches to studying SD in both insects and primates. These chapters will address the relative contributions of factors shaping evolutionary change of reproductive-related genes, methods for experiementally testing these factors, and expanding this work from *Drosophila* to humans and their closest relatives.

First, I review the fitness framework that enables reseachers to experimentally test for the role of adaptation among partitioned phenotypic traits via estimates of fitness componenets. I discuss empirical methods for measuring fitness in *Drosophila,* as well as commonly used proxies used for estimating fitness. I additionally address important caveats of their use, which include complex trade-offs and correlations. These approaches lend their hand towards understanding the relative role of extrinsic forces of selection on various components of fitness between males and females.

Second, I use a molecular approach to compare selective forces that contribute to the rapid evolution of male reproductive genes observed in *Drosophila*. Molecular and cellular processes such as codon bias, pleiotropy, and epistasis that can impose varying evolutionary constraints on protein evolution. The relative roles of these correlates among functional gene classes are not well understood. I collect a comprehensive set of genic parameters including gene length, chromosomal distribution, tissue specificity, protein-protein interactions, sex-bias, codon usage, and total expression to determine their relative contributions in the evolutionary rate of male reproductive-related genes.

Third, in an already published manuscript (Rigby and Kulathinal, 2015) in the Journal of Cellular Physiology, I switch species systems and review sexually dimorphic phenotypes in humans and the current state of our knowledge on sex-biased gene expression in humans. Despite societal anthropocentrism, our understanding of this phenomenon in humans is lacking. Here, I expose the current gaps in the field and the potential implications of filling these knowledge gaps.

Fourth, I present a bioinformatics framework to identify sexually dimorphic genes and genomic elements at the molecular level in humans and primates in order to examine the selective forces involved in their evolution. To demonstrate, I apply independent metrics to identify differentially expressed genes between the sexes on the primate liver, and reveal the difficulty in developing a clear definition of a sex-biased gene. Using genes identified as sex-biased by a combination of metrics, I then examine the evolutionary rates of sex-biased genes to determine whether sexually-biased genes are rapidly evolving in primates.

CHAPTER 2

THE APPORTIONMENT OF FITNESS: MEASUREMENT, TRADEOFFS, AND APPLICATION OF FITNESS COMPONENTS IN *DROSOPHILA*

2.1 Abstract

The concept of fitness is central to our understanding of natural selection and adaptation. As a population-level phenomenon, fitness can be broadly defined as the relative ability of an organism to survive and reproduce within its environment. Partitioning fitness into separate life history components may provide stronger biological insight into the precise targets of natural selection. However, applying this phenotypic approach is not straightforward. First, the methods used to examine how relatively fit an organism is can be laborious and difficult. Second, there are a number of drawbacks that researchers should be cognizant of when designing experiments. Factors such as temperature, nutrition, and age can influence life history traits that are implemented as fitness components. Lastly, trade-offs between fitness components related to survival and reproduction exist as well. In this essay, I will examine the concept and measurement of fitness and its components in the model system, *Drosophila*. I will discuss the major fitness components related to survival and reproduction with respect to *Drosophila,* and how they are measured and applied in empirical investigations. The potential tradeoffs between these components will also be addressed, as these may have implications for future studies. This chapter provides a theoretical and empirical framework to study differences in the strength of selection between both males and females. This phenotypic approach of addressing sex-specific selection is complementary to molecular-based

approaches that partition components of fitness through patterns of gene expression and will be applied in the next chapter.

2.2 Introduction

For natural selection and adaptation to occur, differences in fitness between individuals in a population must exist (Darwin 1859; Orr 2009). These differences in fitness dictate how quickly a population will change in allele frequency and how strongly selection will act upon a given trait. Yet, despite its importance to evolutionary processes, the concept of fitness is neither straightforward nor completely universal among evolutionary biologists.

The idea of fitness was first introduced by Charles Darwin (1859). Darwin's concept of fitness encompassed an organism's ability to survive and reproduce (Darwin 1859). His description, however, was entirely qualitative and lacked the means to quantitate fitness between individuals (Demetrius and Ziehe, 2007). In 1930, Fisher addressed this issue in, *The Genetical Theory of Natural Selection*, a culmination of some of his greatest contributions to evolutionary biology (Plutynski, 2006). This piece vindicated Darwin's ideas and directly tied fitness with genetic variation and the rate of evolutionary change (Fisher, 1930; Plutynski, 2006). In this work, Fisher makes the statement that the "rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time" (Fisher, 1930). Using this paradigm, he developed a quantitative method to address fitness. Fisher was first to apply the Malthusian parameter, also referred to the intrinsic rate of increase, as a measure of fitness (Roff, 2008). Since then, fitness has been examined mathematically, theoretically, and

empirically by a number of studies, across many taxa, including the genetic model of *Drosophila melanogaster*.

Drosophila has been used in the laboratory for more than a century in order to understand a variety of biological phenomena (Jennings, 2011). It is well known that there are a number of advantages to using this system, including its short generation time, the ease and low cost of maintaining laboratory cultures, and the abundance of sophisticated tools with which it can be genetically modified. For these reasons, *Drosophila* has been a popular experimental system to study fitness and its components. In this chapter, I focus on the knowledge and application of fitness measures with respect to Drosophila.

Below, I discuss the concept of fitness and its components, focusing on its application to the *Drosophila* system. I first address the definition of fitness and how fitness is measured. Then, I discuss partitioning of fitness into its fitness components. The role of common fitness components in *Drosophila* is examined, as well as, a review of the current literature pertaining to the knowledge about these components. Tradeoffs between fitness components that may confound the ability of researchers to draw meaningful conclusions from these measurements are also addressed.

2.3 Defining Fitness

The concept of fitness has inspired and mystified evolutionary biologists since Darwin's pivotal work, *On the Origin of Species*. Confusion surrounding this idea partially stems from the difficulty in finding a suitable, universal definition. Many definitions have been proposed by biologists for the concept of fitness (Orr, 2009;

Barker, 2009). These vary in terms of the relationship to individual, population, generation and environment (Barker, 2009). Yet, it is generally agreed that fitness refers to the relative ability of an organism to survive and reproduce in its environment (Orr, 2009; Barker, 2009). Given this definition, a genotype with higher fitness will produce more offspring, and therefore, increase in frequency in subsequent generations. While in principle, this concept and its implications may be rather intuitive; the practice of measuring fitness is not historically simple.

2.4 Measuring fitness

In order to measure fitness accurately, one must monitor aspects of both survival and reproduction across the entire lifespan on an organism. Fitness can be measured at the level of an individual or a population through two ways: absolute fitness or relative fitness. The method for measuring fitness (absolute or relative) and the level at which it is measured (population or individual) depends on the biological question being addressed.

2.4.1 Absolute fitness

Absolute fitness is the total number of gene copies or offspring that an individual produces during its lifetime. It can be measured for a population over a single generation as the ratio between the number of individuals with a given genotype after selection to those before selection. Absolute fitness can also be calculated as the product of fecundity (offspring production) and the probability of survival on an individual or population level.

2.1.2 Relative fitness

Relative fitness is a measure of fitness that has been normalized relative to another fitness value. Essentially, it can be thought of as standardized absolute fitness. At the level of an individual, relative fitness is expressed as the ratio of absolute fitness of that individual to the absolute fitness of another individual. At the population level, relative fitness is the average number of surviving progeny of one genotype compared to the average number of surviving progeny of other genotypes, calculated over a single generation.

2.4.3 Implications of population and individual fitness measures

It is not difficult to imagine that changes in fitness at the individual level can translate to changes at the population level. This can be seen when examining the effects of genetic diversity on fitness. On an individual level, increased homozygosity has been shown to cause decreases in fitness in laboratory and natural settings (Westemeier et al., 1998; Boakes et al., 2007; Fritzshe et al., 2006; Johnson and Dunn, 2006). At a population level, increases in homozygosity and decreases genetic diversity can be associated with a decrease in population fitness (Westemeier et al., 1998). This is because on an individual level, an increase in homozygosity reveals deleterious, recessive mutations. Such a decline in population fitness, due to changes at the individual level can further lead to increased extinction risk through a genetic extinction vortex (Gilpin and Soule, 1986).

2.4.4 Pitfalls in measuring fitness

Estimating fitness in natural populations and environments involves extensive fieldwork over many years (Ellegren and Sheldon, 2008). Due to this difficulty, fitness is most often measured under well-controlled laboratory conditions that, while easy to monitor, do not accurately represent an organism's natural environment. According to Haymer and Hartl (1982), the optimal conditions for obtaining an accurate measure for fitness include: the understanding of a working definition for fitness, inclusion of as much of the organism's life cycle as possible, experimentation under uniform environmental conditions, and manageable protocols. Yet, even under laboratory conditions these requirements can still be difficult to meet. For example, for an organism with a long generation time, even laboratory experiments can take many months or years to complete. Due to the difficulty of measuring fitness, biologists often measure components of fitness and use them as proxies for fitness.

2.5 Fitness components

Partitioning fitness into components can be a useful way to reduce the complexity of fitness and its measurement. Fitness components are traits in which an increase in value correlates with an increase in fitness, given all other traits are constant (Charlesworth, 2000). These are often life history traits – lifespan, viability, fertility, fecundity, etc. – that can be measured and are assumed to be correlated to fitness. As the data for the genetics of fitness are difficult to obtain for diploid organisms (Fowler et al., 1997), most quantitative genetics studies in *Drosophila* use fitness components in

laboratory conditions (Charlesworth, 2015) to make predictions about the fitness of populations, strains, and genotypes.

2.5.1 Drawbacks to studying individual fitness components

Despite the benefits that come with substituting fitness components for the measure of fitness, there are some drawbacks. First, fitness components are different across taxa (Orr, 2009). For example, consider the fitness component of mating success. In a sexual species this measure is applicable, however, in an asexual species mating success cannot be determined (Orr, 2009). Second, researchers often subdivide fitness components arbitrarily and subjectively (Orr, 2009). For example, one researcher may divide the component of survival into egg survival, juvenile survival, and adult survival. Yet, another may decide to divide survival into the number of days the organism is living. Third, predictions made based on fitness components may not mirror those in wild populations (Service and Rose, 1985). Fourth, different components of fitness can be negatively correlated (Partridge and Fowler, 1992, 1993), which can lead researchers to make incorrect assumptions about the overall fitness of the organism. In addition to these issues, there can be other tradeoffs between various fitness components. These will be discussed in more detail later, in the context of the *Drosophila* system.

2.6 Fitness components of *Drosophila*

Various life history traits that are commonly used as fitness components of *Drosophila* will be discussed below. The measurement of these components is commonly carried out in a laboratory setting, remote from the natural population and potentially representing an incomplete portrait of wild conditions (Prout, 1971). Yet, such laboratory

studies are necessary for creating a uniform environment. Because much of the literature on fitness components in *Drosophila* has been in the laboratory, that will be the focus here. Traditional fitness components of Drosophila can be broken into two categories: survival-related fitness components and reproductive-related fitness components (Figure 2.1). While survival fitness components can be measured in the same way for males and females, some reproductive components require differential methods of measurement due to differences in mating strategy between the sexes.

2.6.1 Viability

Viability in *Drosophila* is traditionally determined in terms of egg-to-adult viability. This is reported as a percentage or fraction, and is measured as the proportion of eggs that develop into an adult fly. Viability can easily be determined in the laboratory by counting the number of eggs placed into a container, and after several days, counting the number of viable adults emerged. In order to study differences between the viability of different genotypes, this proportion can be statistically compared between control and experimental study groups.

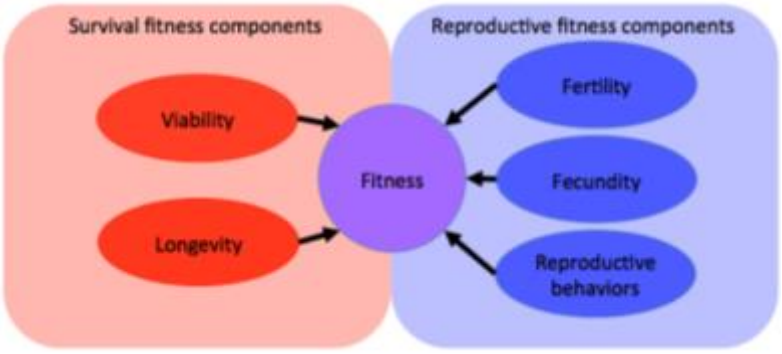


Figure 2.1 *Graphical representation of fitness components*. These components contribute to fitness and are used as proxies for estimating fitness. Survival related fitness components are shown in red, while reproductive related fitness components are shown in blue.

Studies measuring this fitness component have largely examined the environmental, nutritional, and mutational conditions under which fruit flies can survive, with the aim of understanding how these influences play a role in evolution. For example, Kristensen et al. (2015) analyzed temperature effects on egg-to-adult viability in both laboratory and natural settings for *D. melanogaster*, showing that the evolution of viability is more constrained at higher temperatures than lower temperatures, which has implications for global climate change. In terms of nutrition, *D. melanogaster* flies given medium that was protein rich and carbohydrate poor have been shown to exhibit reduced egg-to-adult viability compared to those on standard medium after several generations (Kristensen et al., 2010), suggesting a role for nutrition in this fitness component as well. In another study, Fry and Nuzhdin (2003) measured egg-to-adult viability to support the prediction that transposable element insertions have a greater dominance (on average) in their viability effects than point mutations.

2.6.2 Longevity

Longevity in Drosophila is traditionally determined at the level of the adult stage of the organism. This is a life-history trait that is also related to survival, but unlike viability, it is not commonly measured in juvenile stages in *Drosophila*. Longevity is the duration of the lifespan of an adult individual. For *Drosophila*, this is typically measured in the number of days from emergence to death, and can be monitored in the laboratory for the lifetime of individual flies. As the maximum lifespan of *D. melanogaster* can exceed 90 days, even lifespan studies in this model have the potential to last several months.

Similar to viability, longevity has also been shown to be affected by nutrition. For example, diet dilution (i.e., dietary restriction) has been shown to increase longevity 50% or more in *D. melanogaster* (Tatar, 2007; Mair et al., 2004; Partridge et al., 2005; Piper and Partridge, 2007). Yet, more recent studies have indicated that the relationship between nutrition and longevity is more complex than previously thought. A number of studies indicate that this increase in longevity may be due to the content of specific macronutrients in the food, rather than actual calorie counts (Mair et al., 2005; Vigne and Frelin, 2007). Furthermore, the effects of nutrition on longevity appear to be age-dependent as Vigne and Frelin (2007) showed that 30-day-old flies to not respond to dietary restriction by increased longevity.

In addition to fitness and its evolutionary applications, measuring longevity in *Drosophila* has other applications to aging and age-related diseases. For example, Jeon *et al* (2015) used *D. melanogaster* as a model to examine muscle aging and indicated that there was a strong correlation between age-dependent muscle damage and lifespan. Furthermore, female longevity is traditionally used as a proxy for the detection of sexual conflict in populations of *Drosophila* (Arbuthnott et al., 2014; Fowler and Partridge, 1989; Rice, 1996). This proxy for sexual conflict is traditionally used because male-induced reductions in female longevity have been shown to correlate with reduced total fitness of females (Edward et al., 2011), and in *D. melanogaster*, males with the largest influence on female longevity tended to have the greatest fitness (Rice, 1996). These relationships indicate that male fitness gains are associated with a reduction in female

fitness (Chapman, 2001), and support the use of longevity as a proxy for fitness in detecting ongoing sexual conflict.

2.6.3 Fertility

Fertility refers to an individual's ability to produce offspring. For *Drosophila* females this can be determined by counting the number of eggs produced over the lifetime of the individual. For males, fertility would refer to a male's ability to produce offspring through the transfer of sperm. Yet, determining the amount of sperm transferred in mating is difficult and can only be determined by dissection of females (Lupold et al., 2010). This process, however, would prove extremely difficult if the female has previously been mated, as the sperm from additional males may still be present in her reproductive tracts (Lupold et al., 2010). Due to these challenges, male fertility is often quantified as the number of sired offspring. This is a reasonable method if the aim is to use fertility as a proxy for fitness, as *Drosophila* male fitness has been demonstrated to largely depend on number of matings and average number of progeny sired (Bateman, 1948).

Fertility and sterility studies are central to our understanding of the genes and factors that influence reproduction and hybridization during the process of speciation. Many genes have been indicated as essential for male fertility, and thus, a male's ability to reproduce. Over 2,000 alleles across the *D. melanogaster* genome appear to be involved in male fertility, 40% of which also lead to changes in female fertility when homozygous in D. melanogaster females (Wakimoto et al., 2004). Additionally in

females, fertility also appears to have a nutritional basis, in which lifetime egg production increases with higher food levels (Chapman and Partridge, 1996).

2.6.4 Fecundity

Like fertility, fecundity is also related to the offspring production of an individual. Fecundity, however, refers to an organism's capacity to produce offspring and cause population growth. For *Drosophila* females, this can be determined by counting the number of viable progeny produced from eggs laid. Fecundity is traditionally not calculated for males, as the number of viable progeny produced is usually used as a fertility measure in males (see above section).

Fecundity is proposed to be a major element in female fitness (Roff, 1992). The fecundity of a female is influenced by her age (Lieps et al., 2005), mates (Markow and Ankey, 1984; Partridge et al., 1986; Pitnick 1991), environment, and genetics (Lieps et al., 2005). For example, females mated to males with larger body sizes appear to have greater fecundity (Pitnick, 1991). Additionally, females taken from natural populations that differ in their environmental conditions show different fecundity responses to temperature changes (Lazzaro, 2008).

2.6.5 Other fitness components related to reproductive behavior

Outside of the fitness components already discussed, there are several behavioral components of fitness that are commonly measured in *Drosophila* with respect to mating. As these life-history traits are related to reproduction, they have been indicated as components of fitness. Furthermore, courtship behavior is important to evolution and speciation for its role in premating isolation. These components include courtship

latency, courtship duration, copulation latency, and copulation duration. Courtship latency is the time from which flies are introduced to one another to when the process of courtship begins, while courtship duration is the time from which courtship begins to the time at which flies begin copulating. Similarly, copulation latency is the time from which flies are introduced to when the copulation begins, and copulation duration is the total time for which copulation takes place. As males are responsible for performing the courtship behavior sequence in *D. melanogaster*, components related to courtship are male-specific. Copulation related components, however, are traditionally determined for a pair of mating flies. One can imagine that if a male dedicates more time to courtship (thereby increasing his courtship duration), he may sire more offspring, thereby increasing his fitness. The connection between courtship and fitness, however, has not been experimentally confirmed in *Drosophila*. These behavioral components can be measured by observing or computationally tracking flies in a behavioral arena with a program such as Ctrax (Branson 2009).

Experimental conditions will greatly affect the measurement of these components. For example, the behavioral arena size and the light conditions will influence courtship latency. A larger chamber and lower light conditions will lead to a longer courtship latency, as it will require males more time to locate females (Ejima and Griffith, 2007). Additionally, courtship latency has been shown to be age-specific, with young males taking significantly longer to initiate courtship than older males (Eastwood and Burnet, 1977). So to ensure sexual naivety, these experiments are traditionally carried out with virgin flies.

Related to courtship duration, is the courtship index. This is measured as the total duration of the male's courtship behavior as a fraction of the experiemental observation period (Ejima and Griffith, 2007). For wild-type D. melanogaster, the courtship index ranges from 0.3 to 0.9, depending on lighting and arena conditions (Ejima and Griffith, 2007). As there are many behaviors involved in the male courtship display (e.g., orientation, wing extension, chasing) the courtship index between males may be the same, but the amount of time spent on any one behavior may vary (Ejima and Girffith, 2007). Increasing the granularity of the courtship index may provide insight into individual-specific patterns of courtship behavior.

2.7 Tradeoffs between fitness components

Tradeoffs represent when an advantageous change in one trait leads to a detrimental change in another. These have been discussed with respect to many biological phenomena (Charnov and Krebs, 1974; Gadgil and Bossert, 1970; Partridge and Farquhar, 1981), and may also be present in fitness components. Biologists sometimes draw conclusions about the fitness of an organism based on a single component, however, if tradeoffs exist this could lead to erroneous conclusions. Researchers should consider potential tradeoffs between fitness components, and between fitness components and other influential abiotic factors, like those discussed above, when experiments are designed. The most discussed tradeoff is with reproduction.

Reproduction is often associated with costs in sexually reproducing species (Williams, 1966), therefore, it is not surprising that there is evidence of tradeoffs between survival and reproductive fitness components (Figure 2.1). Longevity has historically

been expected to have tradeoffs with reproductive effort (Tatar, 2007; Piper and Partridge, 2007; Partirdge et al., 2005; Hunt et al., 2004). This is certainly true for effort associated with the performance of reproductive behaviors. In particular, male flies that more frequently perform reproductive behaviors, specifically courtship, tend to show a decrease in longevity (Cordts and Partridge, 1996; Partridge and Farquhar, 1981). There is also a growing amount of evidence suggesting that sexual reproduction and mating is costly to female insects, including *Drosophila*. Females continuously exposed to males (and therefore courtship and seminal fluid) consistently live longer than those intermittently exposed to males despite equal amounts of egg production (Chapman et al., 1995; Chapman and Partridge, 1996; Partridge et al., 1987), suggesting an negative interaction between reproductive behavior and longevity for females as well. From these types studies, there appears to be no interaction between egg production (fertility) and longevity (Chapman et al., 1995; Chapman and Partridge, 1996; Partridge et al., 1987; Barnes et al., 2008).

Longevity and fecundity have also been suggested to have a tradeoff relationship. Several studies using artificial selection methods show that extended lifespan decreases early fecundity in *D. melanogaster* (Zwaan et al., 1995; Rose, 1984; Partridge et al., 1999). Yet, other studies have shown that this relationship can be uncoupled through genetic processes such as recombination. For example, Khazaeli and Curtsinger (2013) show that the correlation between longevity and fecundity can be broken down through the generation of fruit fly genotypes in highly recombinant lines.

2.8 Conclusions and future directions

Natural selection provides a mechanism enabling species to adapt to their biotic and abiotic environments. An understanding of fitness and its various components is helping us to tease apart the selective forces that influence traits. The genetic model system, *Drosophila*, has been a premiere model for the study of fitness. As with many other taxa, partitioning fitness into components has facilitated the study of fitness. Due to the possibility of tradeoffs between fitness components, however, in the future it will be important to develop complete knowledge about the correlation between these variables.

This classical phenotypic approach in studying the strength of selection on various components has been largely superseded by a genomics approach based on gene ontologies. By surveying their temporal-spatial patterns of expression, we can classify genes according to when (e.g., embryo vs. adult), where (e.g., gonadal vs. brain), and to whom (e.g., male vs. female) they are expressed. Both phenotypic and genomic approaches can complement each other.

CHAPTER 3

INTRINSIC FACTORS AFFECTING THE EVOLUTIONARY RATE OF RAPIDLY EVOLVING MALE GONADAL GENES IN *DROSOPHILA*

3.1 Abstract

A longstanding goal of evolutionary biology is to understand what promotes and constrains evolutionary rate among proteins. A well-known evolutionary pattern ubiquitous across taxa is that protein sequences of reproductive-related genes are generally more rapidly evolving, with male proteins evolving significantly faster than female reproductive genes and non-reproductive genes. While this pattern is usually explained through extrinsic factors such as stronger sexual selection on male genes, intrinsic factors including protein length, codon bias, and expression levels unique to testes-expressed genes may also explain the rapidity of male gene evolution. However, the relative roles of these extrinsic vs. intrinsic evolutionary parameters in promoting rapid evolution remains unknown. Here, we use a multiple linear regression approach to examine the relative roles of correlates in the rate of sequence evolution of reproductive proteins in *Drosophila*. We find that these models best predict variation of dN/dS (ω) in male gonadal genes, with codon adaptation index (CAI), length, and chromosome as the most significant contributors. Sex-bias and CAI are significant predictors of female gonadal genes while for both reproductive and non-reproductive genes, CAI has the greatest relative importance among correlates, indicating its important contribution to evolutionary rate in the form of ω. However, while significantly lower codon bias in male genes reduces evolutionary constraints in both dN and dS, it only partially explains the

rapid evolution of male genes, highlighting the importance of extrinsic factors such as selection on male traits in driving higher rates of protein evolution.

3.2 Introduction

A remarkable amount of variation exists in the evolutionary rates of proteins, with important implications to evolutionary biology. A wealth of functional and comparative genomic data has provided insight into the function of genes and their proteins on both ends of the evolutionary spectrum (Stanley and Kulathinal 2016). A commonly found pattern ubiquitous across taxa is that reproductive-related genes and proteins are regularly identified as among the most rapidly evolving (Wong and Wolfner, 2012), with a higher than average proportion of amino-acid substitutions between species (Swanson and Vacquier, 2002). In ciliates of the genus *Euplotes*, an alignment of amino acids involved in sexual conjugation show less than ten conserved amino acids across different mating types (Luporini et al., 1995). In mammals, sperm-egg interacting proteins are also evolving rapidly (Makalowski and Boguski, 1998; Swanson et al., 2003; Torgerson et al. 2003). In the diatom. *Thalassiosira*, the extracellular matrix protein *Sig1* is upregulated during mating and believed to be involved in the mating process. This gene is divergent between species and distinguishes Atlantic and Pacific Ocean populations (Armburst and Galindo, 2001). Even plants exhibit rapid evolution of reproductive genes. Pollen coat components of *Arabidopsis thaliana* involved in plant mating show high variability (Mayfield et al., 2001).

The rapid evolution of reproductive proteins has been extensively studied in *Drosophila*. In this group, proteins from reproductive tissues are more divergent than

those from non-reproductive tissues (Civetta and Singh, 1995; Singh and Kulathinal, 2000; Vacquier, 1998). Specifically, male reproductive genes are evolving faster than non-reproductive genes (Civetta and Singh, 2005; Swanson et al., 2001). Much work has been focused on male accessory gland proteins, which are transferred from the male to the female reproductive tract via ejaculate (Wolfner, 1997). These accessory gland proteins are demonstrated to be twice as diverse between species as non-reproductive proteins (Civetta and Singh, 1995).

This trend raises the question: why are reproductive genes evolving so rapidly? On one hand, the rapid evolution of male reproductive genes could be attributable to adaptive evolution, with extrinsic processes of sexual selection (e.g., sperm competition) promoting amino acid change. On the other hand, the observed rapid evolution may be due to a lack of functional constraint unique to male genes (Swanson and Vaquier, 2002) with genes rapidly accumulating substitutions due to relaxed purifying selection. It has been highly debated whether elevated rates of substitution, as observed through studies of ω, are indicative of positive selection or relaxed purifying selection (Li and Gojobori, 1983; Zhang et al., 1998; Van de Peer et al., 2001; Zhang, 2003). Therefore, understanding what drives rapid evolution in male-reproductive genes will require a more thorough examination of the factors driving these elevated observations of ω.

Several intrinsic factors have been identified as potential predictors in the evolutionary rate of proteins, and are known to correlate with protein sequence evolution. Intrinsic factors refer to evolutionary constraints acting at the molecular and cellular level and include protein length, chromosomal location, and codon bias among others

(Larracuente et al., 2008). Here, we attempt to untangle the importance of each of these factors in driving the rapid evolution of reproductive proteins. First, we classify genes of *Drosophila melanogaster* based on their presence in reproductive tissues in the sexes. We then use multiple linear regression to disentangle the role of seven evolutionary rate correlates by comparing sequence divergence between *D. melanogaster* and *D. simulans*.

3.3 Methods

3.3.1 Tissue-specific gene expression and functional classification

As a measure of functional pleiotropy, we estimated tissue specificity for each gene using τ (Yannai et al., 2005), which has been shown to have strong positive correlation with dN/dS (Wright et al., 2004; Ingvarsson, 2007; Duret and Mouchiroud, 2000; Liao et al., 2006). Gene expression levels for multiple tissues of *D. melanogaster* were obtained from FlyAtlas (Chintapalli et al., 2007) in December of 2015 from http://flyatlas.org/data .html. Similar to Meisel et al., 2012, we focused our analysis on the following 14 adult tissues: brain, eye, thoracicoabdominal ganglion, salivary gland, crop, midgut, tubule, hindgut, heart, fatbody, ovary, testis, male accessory gland, and virgin spermatheca, eliminating compound tissues such as 'head' and 'carcass' from our analysis. Since the FlyAtlas data is referenced in the original data under Affymetrix probes, in order to determine the tissue expression for each gene the Affymetrix annotation file "Drosophila_2.na23.annot.short.csv" was also downloaded from the above web address, as recommended by FlyAtlas. The expression level for each Affymetrix probe in each of the 14 tissues was set to 0 unless the probe was determined to be present in at least two of the four replicate FlyAtlas arrays. As a single gene may be represented

by multiple probes in the FlyAtlas dataset, we averaged over all probes and arrays for each gene in each tissue. To determine the degree of tissue specificity for each gene we estimated τ (Yannai et al., 2005) following the approach of Meisel (2012) as

$$\tau = \frac{\sum_{i=1}^{N} 1 - \frac{\log S_i}{log S_{max}}}{N-1},$$

where S_i is the signal intensity of tissue i and S_{max} is the maximum signal intensities of all tissues for the gene. N refers to the total number of tissues. For this calculation, taking the log of the expression values reported by FlyAtlas transforms the data to reduce large differences between the maximal tissue expression and tissue expression values within the expression profile of the gene (Weber and Hurst, 2011). Genes with $\tau > 0.9$ were conservatively designated as tissue-specific, and for these we recorded the tissue in which expression was highest. The genes were then classified into 5 non-overlapping categories as follows:

a. male gonadal-specific: genes specific to the male testis or accessory gland;
b. female gonadal-specific: genes specific to the female ovary or spermatheca;
c. other specific: genes with values above the tau specificity cutoffs with highest expression in a tissue other than those listed above;
d. non-specific: genes with values below the tau specificity cutoff of 0.90;
e. all genes: includes all genes in dataset.

3.3.2 Total gene expression levels

Because gene expression levels are found to have a strong negative correlation with dN/dS (Drummond et al., 2005; Drummond et al., 2006; Pal et al., 2001; Popescu et al., 2006; Lemos et al., 2005; Marais et al., 2004), we estimated total expression for each

gene. We initially attempted to estimate total expression by summing the raw expression values obtained from tissue-specific FlyAtlas data. We found, however, that the correlation between this total expression value and dN/dS was not as strong as expected (Spearman's rank, $\rho = -0.244$, $P < 0.001$). We suspect that this is due to greater noise arising from microarray platforms. We instead chose to implement total expression as determined from RNA-seq data. Gelbart and Emmert (2013) calculated expression levels from modENCODE (Celniker et a., 2007) RNA-seq expression data as reads per kilobase of exon model per million mapped reads (RPKM) for over 25 *D. melanogaster* life stages. To obtain expression levels of genes across life stages, the file containing these calculations, gene_rpkm_report_fb_2016_01.tsv.gz, was downloaded from FlyBase (Attrill et al., 2016) in January of 2016. In this dataset, adult male and female flies aged 1 day, 5 days and 30 days post-eclosion are reported separately. To obtain a single value for adult flies at each of these time points, RPKM values for each gene at each time point were averaged between the male and female samples. We then calculated total expression by adding together all RPKM expression levels from all stages for each gene from this dataset. The observed correlation between this total expression value and dN/dS was indeed stronger than what we had observed with the FlyAtlas data (Spearman's rank, $\rho = -0.335$, $P < 0.001$).

3.3.3 Sex-biased gene expression

Sex-biased genes, especially those that are male-biased, show rapid protein evolution (Ellegren and Parsch, 2007). In order to include sex-biased expression as one of our correlates, sex-biased expression estimates for *D. melanogaster* were obtained from

SEBIDA (Gnad and Parsch 2006; SEBIDA Dmel v3.2) in January of 2016. The SEBIDA database provides processed microarray data comparing gene expression levels from male and female flies. We used a meta-analysis reported by the database providing the male-to-female ratio of expression level, P values and false discovery rate (FDR) estimated by Bayesian inference (Townsend and Hartl, 2002). This meta-analysis takes sex-specific measurements from whole flies (Innocenti and Morrow, 2010; Wyman et al., 2010; Ayroles et al., 2009; McIntyre et al., 2006; Gibson et al., 2004; Parisi et al., 2004; Ranz et al., 2004; modENCODE 2011) and gonads (Parisi et al., 2003; Gan et al., 2010) to generate the male-to-female (M:F) estimates of gene expression. An FDR cutoff of 10% was used to classify sex-bias in the SEBIDA dataset. To determine the extent of sex-bias of each gene, regardless of sex, we first set any M:F estimate with an FDR greater than 10% as 1.0, indicating unbiased expression between males and females. Then, we transformed all M:F values by taking the absolute value of the $\log_{10}$ of each, using this as the level of sex-bias for that gene. Thus, the sex-bias value ranges from 0 to infinity, with 0 indicating an unbiased gene and increasing values indicating increasing levels of sex-bias.

3.3.4 Protein length and chromosomal location

Protein length is known to have a weak, but statistically significant, negative correlation with protein evolution (Comeron et al., 1999; Duret and Mouchiroud, 1999). As a proxy for protein length, we included the length of the coding sequence (CDS) of each gene as a predictor in our model. In addition to length, chromosomal locations of genes have also been demonstrated to influence dN/dS of genes, with X-linked genes

having higher evolutionary rates of substitution (Mackay et al., 2012). We obtained gene lengths and locations from FlyBase (Attrill et al., 2016). The file dmel-all-CDS-r6.09.fasta was downloaded from Flybase.org in February of 2016. The chromosomal location and length of each gene transcript was extracted using a custom python script. For genes with multiple transcripts, the transcript used to calculate the rates of evolution in flyDIVaS was chosen in order to maintain consistency.

3.3.5 Codon usage bias

For a given gene, the codon adaptation index (CAI) reflects its synonymous codon bias (Sharp and Li, 1987). Significant negative correlations between evolutionary rate and CAI have been observed previously (Sharp and Li, 1986). CAI for each gene was calculated using codonW program (http://codonw.sourceforge.net/), using *D. melanogaster*-specific reference values as determined by Carbone et al., 2003 (http://www.ihes.fr/~carbone/materials/genomes/Dmelanogaster/ wv.txt).

3.3.6 Protein-protein interactions

While there is some controversy in the role protein-protein interactions play in determining evolutionary rate, we included a metric estimated from the protein interaction network of *D.melanogaster* to represent the role of intramolecular epistasis in our model. Such epistatic effects can be inferred from a protein-protein interaction (ppi) network (Sun and Kardia 2010). The network measure of closeness (defined as the average number of nodes connecting a protein to all others) takes both direct and indirect interactions into account (Hahn and Kern 2005), providing a comprehensive estimate of epistasis within the protein-protein interaction network. We obtained the ppi network for

D. melanogaster from the High-Quality Interactions Database (HINT; Das and Yu, 2012), which includes known ppi fron the literature, filtered for high-quality interactions only. Using the Python module Networkx, we calculated closeness centrality (a measure of closeness, normalized by the total number of proteins in the network) for each gene in our dataset.

3.3.7 Rates of evolution

Rates of evolution (dN, dS, and dN/dS) for each gene were obtained from flyDIVaS at flydivas.info (flyDIVaS_v1.2; Stanley and Kulathinal, 2016). We examined the rates of evolution of orthologous genes between the closely related species, *D. melanogaster* and *D. simulans*, in order to maximize the number of genes with complete data in the dataset.

3.3.8 Statistical analyses

Any gene with incomplete data (tissue specificity, sex-biased expression, dN/dS, gene length, chromosome, CAI) was removed from the final dataset and subsequent analyses. Mann-Whitney-Wilcoxon tests were used to compare differences in determinants between functional groups. In order to test for the effects of seven different factors on protein divergence (length, chromosome, tissue specificity, total expression, CAI, sex-bias) we performed an independent multiple linear regression model for each functional category of genes. For the protein length of each gene, we used the total length of the CDS of the gene, as obtained from FlyBase. Chromosome location, a categorical determinant, was transformed into binary dummy variables (1 or 0; X-linked or autosomal). Tissue-specificity was measured as τ, as calculated from the FlyAtlas data

(described above). Sequence divergence between *D. melanogaster* and *D.simulans*, ω, was considered the response variable in each model, with the seven determinants (no interaction terms included) as main effects. Features were normalized to have a mean of zero and a variance of one. The response variable, ω, was log transformed and extreme small values were removed. We observe pairwise correlations between some of these variables, and the multiple linear regression models intended to disentangle these effects. We additionally used custom R scripts to then determine the relative importance of each of the determinants.

3.4 Results

3.4.1 Functional classification of genes

We analyzed a total of 8,686 genes with complete data in our final data set (Table 3.1), representing over 60% of the genes in Drosophila. We characterized these genes into functional categories based on their tissue specificity, τ, estimated based on standardized microarray data from multiple tissue obtained from FlyAtlas (reference). Of these genes, the largest portion were categorized as non-specific (6,644 genes; 76.5%), while the smallest portion were female-biased (143 genes; 1.6%; Table 3.1). As expected, we observed that the mean dN and ω of male gonadal genes is significantly higher than that of non-specific genes on average (Table 3.1; Figure 3.1). Male gonadal genes also have a significantly higher ω than all genes on average ($P < 0.01$).

Table 3.1

Summary of gene categories and mean values of determinants

Category	Number of genes	Mean dN	Mean dS	Mean ω	Mean length	Mean closeness	Mean τ	X:A ratio	Mean sex-bias	Mean CAI	Mean total expression
Male gonadal-specific	1,263	**0.0439***	**0.1787***	**0.2871***	**1,187***	0.2287	**0.985***	1 : 6.65	**0.670***	**0.306***	**254.72***
Female gonadal-specific	143	**0.0269***	0.1492	**0.1903***	1,772	0.2349	**0.962***	**1 : 2.83***	**0.499***	**0.332***	**963.71***
Other specific	636	**0.0193***	0.1500	**0.13386**	**1,358***	0.2263	**0.965***	1 : 7.42	**0.026***	0.359	**1,140.84***
Non-specific	6,644	0.0153*	0.1343*	0.1149*	1,727*	0.2291	0.382*	1 : 6.00	0.174*	0.362*	1,776.27*
All genes	8,686	0.0200	0.1422	0.1425	1,622	0.2290	0.523	1 : 6.08	0.241	0.353	1,571.04

Note. Bold values in specific categories significantly different from non-specific genes, Mann-Whitney-Wilcoxon $P < 0.001$; * values significantly different from all genes Mann-Whitney-Wilcoxon $P < 0.001$. For X:A ratio, R proportions test $P<0.001$.

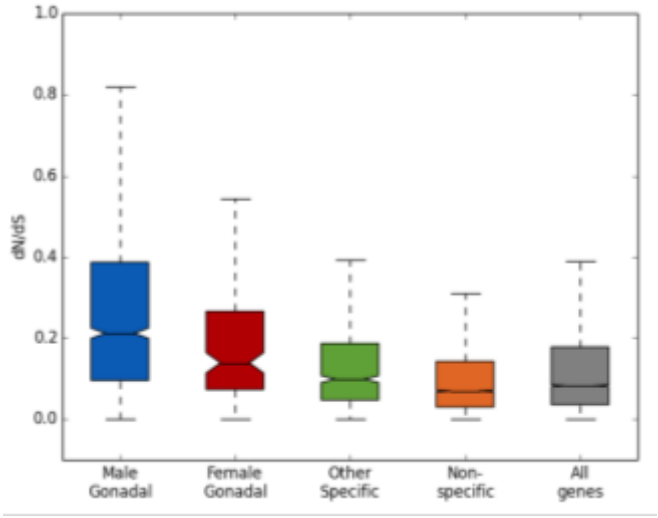


Figure 3.1. *Distribution of dN/dS (ω) for functional gene classes*. Notched boxplot displaying the distribution of dN/dS values between *D. melanogaster* and *D.simulans* for genes in each functional class.

3.4.2 The role of evolutionary determinants

To assess the relative contribution of each evolutionary determinant on dN/dS we performed multiple linear regression for each functional class of genes. Table 3.2 displays the results of these models. The coefficients of determination (multiple R^2) of the models are highly significant (Table 3.2). We find that length, chromosome, and CAI

are significant contributors to dN/dS in male gonadal genes, while sex-bias and CAI are significant contributors for female gonadal genes (Table 3.2; $P < 0.05$). For all genes and non-specific genes, we observed that all variables except closeness had a significant impact (Table 3.2).

Table 3.2

Linear regression model for divergence

	Male gonadal				Female gonadal			
Linear model	Coefficient		P-value		Coefficient		P-value	
Multiple R^2	0.3356		**< 2.2 e-16*** **		0.2815		**0.0006816** **	
Variable	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
Tau	0.2251	0.4885	0.461	0.6451	0.8460	0.9723	0.870	0.3871
Length	-0.3204	0.0741	-4.325	**1.83e-5*** **	-0.07982	0.1078	-0.740	0.4615
Sex-bias	-0.0033	0.0291	-0.113	0.9103	0.1646	0.0795	2.071	**0.0418**
CAI	-0.7172	0.0505	-14.198	**< 2e-16*** **	-0.71016	0.14472	-4.907	**5.35e-6*** **
Closeness	0.0580	0.0376	1.541	0.12397	-0.0572	0.1142	-0.501	0.6182
Total Expression	-0.3156	0.6034	-0.523	0.6012	0.5673	0.3311	1.713	0.0909
Chromosome	0.4644	0.1128	4.116	**4.48e-5*** **	-0.0428	0.2492	-0.172	0.8640

Note. Bold values indicate $P < 0.05$, asterisks indicate level of significance: <<0.001***,<0.001**, <0.01 *.

Table 3.2

Linear regression model for divergence (continued)

	Other specific				Non-specific			
Linear model	Coefficient		P-value		Coefficient		P-value	
Multiple R^2	0.2285		**2.688e-10*** **		0.2084		**< 2.2 e-16*** **	
Variable	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
Tau	0.6964	0.5256	1.325	0.1865	0.1252	0.0205	6.118	**1.04e-9*** **
Length	-0.1062	0.0789	-1.345	0.1798	-0.0893	0.0143	-6.240	**4.86e-10*** **
Sex-bias	0.0268	0.1120	0.240	0.8108	0.0599	0.0203	2.958	**0.00311** **
CAI	-0.44963	0.0625	-7.200	**9.02e-12*** **	-0.5392	0.0182	-29.690	**< 2e-16*** **
Closeness	-0.0956	0.0591	-1.691	0.1068	-0.0115	0.0146	-0.785	0.4323
Total Expression	0.8715	0.1467	5.942	**1.07e-8*** **	0.2198	0.0240	9.153	**< 2e-16*** **
Chromosome	0.3510	0.1881	1.866	0.0633	0.3312	0.0432	7.653	**2.46e-14*** **

Note. Bold values indicate $P < 0.05$, asterisks indicate level of significance: <<0.001***,<0.001**, <0.01 *.

Table 3.2

Linear regression model for divergence (continued)

	All genes			
Linear model	Coefficient		P-value	
Multiple R^2	0.2681		**< 2.2 e-16*** **	
Variable	Estimate	Std. Error	t value	Pr(>|t|)
Tau	0.1754	0.0143	12.266	**< 2.2 e-16*** **
Length	-0.1058	0.0136	-7.774	**9.29e-15*** **
Sex-bias	0.0868	0.0141	6.156	**8.07e-10*** **
CAI	-0.5511	0.0163	6.156	**< 2.2 e-16*** **
Closeness	-0.115	0.0132	-0.870	0.384
Total Expression	0.2418	0.0233	10.364	**< 2.2 e-16*** **
Chromosome	0.3457	0.0391	8.831	**< 2.2 e-16*** **

Note. Bold values indicate $P < 0.05$, asterisks indicate level of significance: <<0.001***,<0.001**, <0.01 *.

The estimation of the relative importance of each correlate on the linear model is revealed in Figure 3.4. CAI appears to have the largest relative importance for all classes of genes. CAI was also a significant contributor for each class in our linear models (Table 3.2). Figure 3.4 illustrates the degree to which our correlates correlate with one another for all genes in the dataset. We found significant, negative correlations between CAI and τ ($P<0.001$), length ($P<0.001$), sex-bias ($P<0.001$), and ω ($P<0.001$). We also noted significant, positive correlations between ω and sex-bias ($P<0.001$) and τ ($P<0.001$).

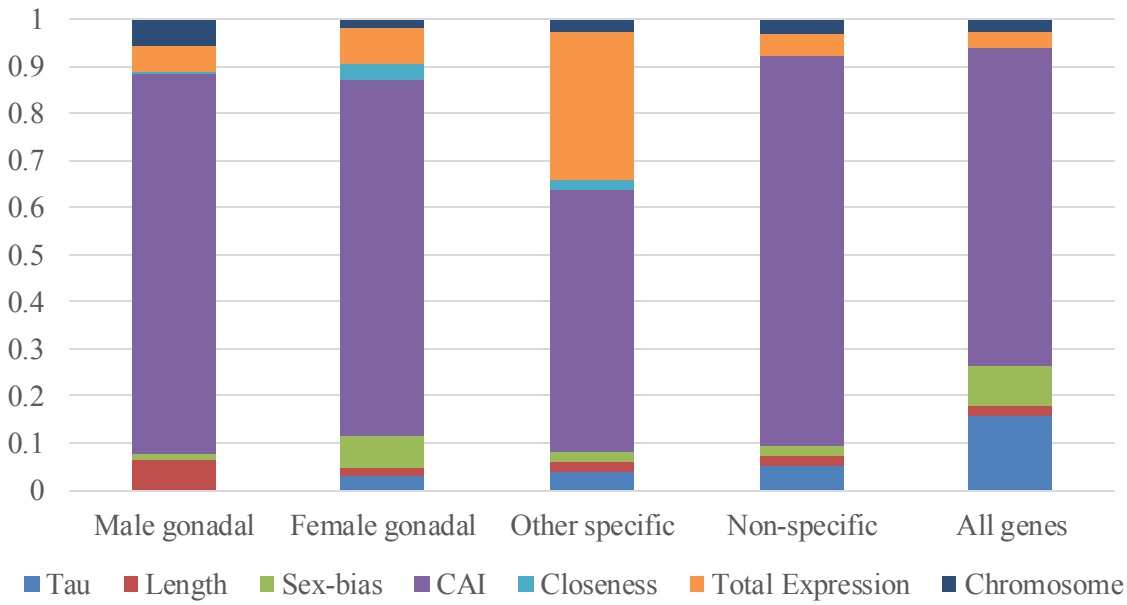


Figure 3.2. *Relative importance metrics for each functional category*. Stacked bar chart showing the relative importance of each evolutionary correlate, as calculated from multiple linear regression models.

In order to visualize the distribution of CAI with relation to dN/dS and to determine if CAI explains higher dN/dS solely in male genes, we divided CAI values for each function class into five groups based on their value (Figure 3.3). We noticed that for

male gonadal genes, as well as, non-specific and all genes, low values of CAI have the highest dN/dS.

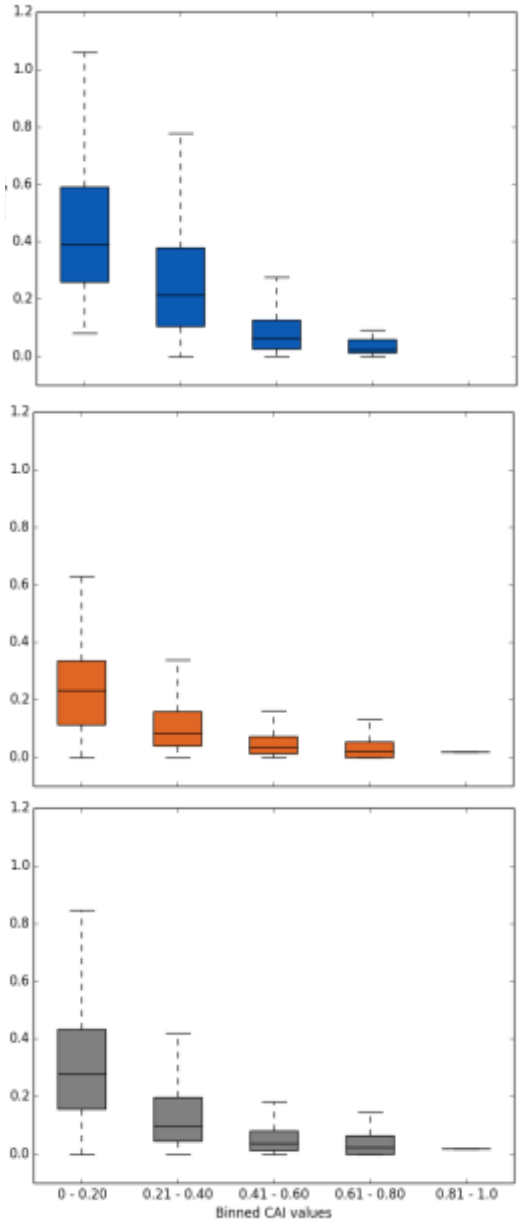


Figure 3.3 *Binned CAI values for functional classes*. Boxplots of CAI values binned into 5 groups for male gonadal genes (top, blue), non-specific genes (middle, orange), and all genes (bottom, gray).

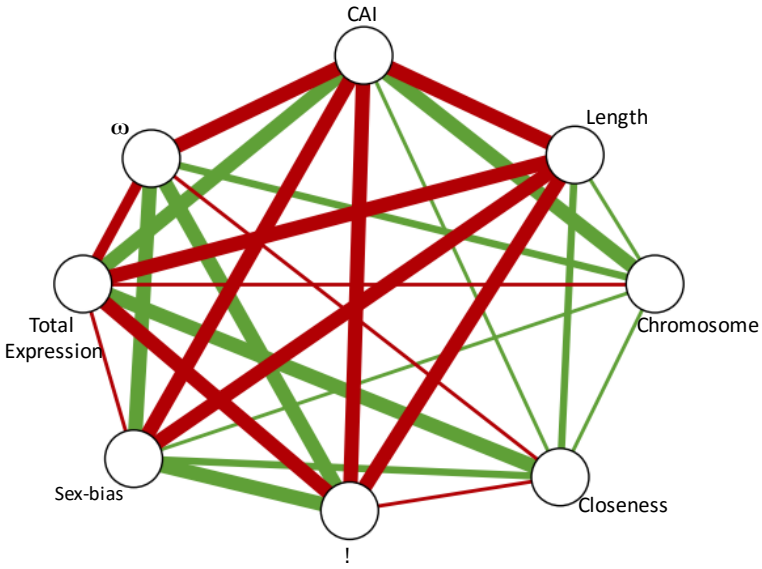


Figure 3.4. *Correlation among evolutionary correlates for all genes*. Network figure depicting correlations between correlates. Green edges (lines between nodes) indicate positive correlations, while red edges indicate negative correlations. Edge width increases with increasing correlation.

3.5 Discussion

Male reproductive genes are among the most rapidly evolving classes of genes (Civetta and Singh 1995; Singh and Kulathinal 2000; Vacquier 1998). Studies in protein evolution have revealed a number of factors that correlate with rates of sequence evolution, but the relative importance of these determinants in the evolutionary rates of reproductive genes has been largely unknown. Here, we have used independent multiple linear regression models to disentangle the effects of these evolutionary correlates on sequence evolution in reproductive and non-reproductive genes.

We first developed a functional classification of genes, grouping male and female gonadal-specific genes based on their tissue specificity. Our classifications proved robust, as we observed expected trends in dN/dS values among these functional classes. Male gonadal genes exhibited the highest rates of sequence divergence, with female gonadal

genes exhibiting the next to highest rates (Figure 3.2; Table 3.1). We observed the lowest rates of divergence among non-specific genes, which could also be expected (Figure 3.2; Table 3.1). These non-specific genes are broadly expressed throughout the organism and likely play important roles related to cellular and molecular function.

Second, we performed five, independent multiple linear regressions, one for each of our functional classes. In these regression models, we used our seven correlates (length, chromosome, closeness, τ, sex-bias, total expression, and CAI) as explanatory variables and dN/dS as the response variable. For each of our linear models, we found that codon bias was a significant contributor for all functional classes, and it provided the largest relative importance for all functional classes (Table 3.2; Figure 3.2). The magnitude of its relative contribution and the negative correlation between codon bias and evolutionary rate that we observe here, has also been documented in the past (Pál et al., 2001; Drummond et al., 2006; Sharp and Li, 1986). While it is unclear whether the correlation indicates causality, one can imagine that CAI could influence evolutionary rate. Slightly selectively beneficial substitutions may generate a synonymous codon that could be translated more quickly (McInerney, 2006). On the other hand, male gonadal genes with significantly lower codon bias may reflect the testis environment and the large amounts of sperm that need to be produced without strong selection for translational accuracy.

In addition to trends observed with relation to CAI, we also observe a relationship between ω and length. Our linear regression model shows that length is a significant contributor to ω in male gonadal genes, though its relative importance is less than 10% in

our model. Interestingly, male gonadal genes on average are significantly smaller than non-specific genes (Table 3.1) as well. This could be due to the fact that testis genes are often novel due to strong selective pressures from male-male competition in *Drosophila,* and, thus, lack network connectivity and functional constraint (Hansen and Kulathinal, 2013).

With this analysis, we confirm that CAI provides a substantial negative correlate to overal dN/dS in *D. melanogaster*. CAI's contribution to evolutionary rate is also significant with male gonadal genes, but certainly not entirely predictive (Figure 3.3). With just greater than 30% of the variation in omega explained by the linear regression model, missing extrinsic factors such as selection on male reproductive genes may provide some of the remaining contribution.

CHAPTER 4

GENETIC ARCHITECTURE OF SEXUAL DIMORPHISM IN HUMANS

4.1 Abstract

Males and females differ across a broad spectrum of morphological, physiological, and behavioral characters. In fact, sexually dimorphic traits typically contribute the largest component of phenotypic variance in most taxa that use sex to reproduce. However, we know very little about the mechanisms that maintain these dimorphic states and how these sexually dimorphic traits evolve. Here, we review our current knowledge of the underlying genetic basis of sexual dimorphism in humans. First, we briefly review the etiology of sex differences starting from sex determination's initial switch early in embryogenesis. We then survey recent sex-biased transcriptomic expression literature in order to provide additional insight into the landscape of sex-biased gene expression in both gonadal and non-gonadal tissues: from overall prevalence to tissue specificity to conservation across species. Finally, we discuss implications of sex-biased genetic architecture to human health and disease in light of the National Institute of Health's recently proposed initiative to promote study samples from both sexes.

4.2 Introduction

In most mammalian species, individuals can be defined as either male or female based on the respective presence or absence of a Y chromosome. This distinction is concealed within each individual cell, however, more conspicuous phenotypic differences between females and males are common and readily observed. Sex differences that are

not directly associated with primary reproductive organs (i.e., distinctive male and female gonads) such as behavior, morphology, and physiology, feature prominently in most species. In humans, for example, adult females generally possess developed mammary glands, broader pelves, and thinner body hair compared to their male counterparts (Plavcan, 2012). On an evolutionary level, these divergent sex-specific traits, often called "secondary reproductive characters", are also thought to play a large role in social interactions, mate choice, and reproductive fitness (e.g., Darwin, 1871).

As we explore the landscape of male and female states, previously unknown differences are being discovered. For example, on a physiological level, men are able to synthesize serotonin, the neurotransmitter commonly associated with pleasant moods, at a greater mean rate that women (Nishizawa et al., 1997). Also, women are able to metabolize certain drugs more efficiently due to greater levels of cytochrome P450 in their livers (Anderson, 2001; Tullis et al., 2003). In addition, women are more likely to develop diseases such as major depression, anxiety, and multiple sclerosis (Ngo et al., 2014), while men are more likely to be diagnosed with coronary artery disease (Kannel and Feinleib, 1972; Gordon et al., 1978; Patrick et al., 1982; Lerner and Kannel, 1986), and attention deficit hyperactive disorder (Arnold,1996; Gaub and Carlson, 1997; Arcia and Conners, 1998). While more research is needed to disentangle genetic and environmental sex-specific effects, it is an intriguing possibility that innate physiological differences between males and females may play a large role in sex differences in disease onset, susceptibility, prevalence, and treatment responses.

Also intriguing is that fact that these sex differences exist despite the fact that males and females share a common genome, with the exception of a handful of genes residing on the male-specific Y chromosome. In most species, once the sex of an individual is established during development, sexually dimorphic characters emerge due to differences in gene expression levels (Rinn and Snyder, 2005; Connallon and Knowles, 2005) and the differential expression of sex-specific alternative transcripts and isoforms (McIntyre et al., 2006; Telonis-Scott et al., 2009). Genes that exhibit differential expression between males and females are referred to as sex-biased. A sex-biased gene may be additionally designated as male- or female-biased if it harbors greater expression in males or females, respectively (Ellegren and Parsch, 2007). Identifying sex-biased gene expression will, ultimately, help uncover the genes responsible for sexually dimorphic phenotypes (Ellegren and Parsch, 2007). Furthermore, understanding the extent of sex-bias across the genome may also provide new insight into sexually dimorphic physiology and susceptibility to disease.

Unfortunately, studies of sexually dimorphic expression in humans, particularly on a genome-wide scale, are still few and far between. Currently, most research on sex-biased gene expression has been carried out using model organisms such as fruit flies (Jin et al., 2001; Parisi et al., 2003; Ranz et al., 2003), nematodes (Reinke et al., 2000; Thoemke et al., 2005), and mice (Yang et al., 2006). While model organisms, especially the mouse, have provided invaluable information for understanding human health and disease, they may not provide the ideal model system to unlock genetic mechanisms involved in sex-biased gene expression. Unlike Drosophila (Zhang et al., 2007), sex-

biased gene expression appears not to be conserved across mammalian species (Si et al., 2009), thus, a thorough knowledge of sex-biased expression in humans, and not just its close evolutionary proxies, is critical to understanding its role in human evolution and disease.

Here, we attempt to motivate new insight into the general genetic architecture of genes involved in global sex-biased transcription in humans. First, we briefly review the developmental mechanisms involved in sex determination that commence during embryogenesis. We then integrate emerging new data of genome-wide sex-biased expression across different tissues including the most sexually dimorphic tissues, the ovaries and testis, as well as non-gonadal tissues. Finally, we examine the implications of sex-biased gene expression on our understanding of sex differences on human health and preview recent steps that the National Institutes of Health (NIH) has proposed to promote research in this relatively uncharted area.

4.3 Determining sex in humans

Sex-biased gene expression first arises during embryogenesis as sexually dimorphic morphology and physiology starts to develop. From fertilization, every individual cell can be characterized as male or female based on the sex chromosomes that are present. It is not until approximately 1.5 months into human embryonic development, however, that sex differences can be detected. These differences begin with an individual's genetic (i.e., chromosomal) endowment that determines the gonadal sex of the organism and the development of reproductive tissues.

Much of what we know about the sex determination pathway in humans derives from mouse models and studies of human disorders. At the early epiblast stage, male and female embryos are indistinguishable morphologically. They contain an identical collection of cells, called the bipotential gonad, capable of becoming either testes or ovaries. The formation of this non-differentiated, bipotential gonad requires the expression of two genes, Wilms' tumor 1 (WT1) and Steroidogenic Factor-1 (SF1) (Eggers and Sinclair, 2012, and references therein). The expression of the sex-determining region Y (SRY) gene, on the Y chromosome, signals the bipotential gonad to develop into testes by instructing the expression of downstream genes necessary for normal testes formation (Figure 4.1) including GATA binding protein 4 (GATA4), SRY-related HMG box 9 (SOX9), Zinc finger protein multitype 2 (FOG2), WT1, SF1, and Anti-Mullerian hormone (AMH) (reviewed in Eggers and Sinclair, 2012). In females, the absence of SRY expression leads to ovary formation via the expression of a different set of genes including Wingless-type MMTV integration site family, member 4 (WNT4), R-spondin 1 (RSPO1), and Forkhead box L2 (FOXL2) (reviewed in Eggers and Sinclair, 2012).

Once formed, ovaries and testes become the primary regulators of mammalian sexual differentiation by secreting sex-specific hormones that regulate downstream developmental processes. Thus, these reproductive tissues impose body-wide and long-lasting phenotypic effects (Figure 4.1). For example, the sex steroids testosterone and estradiol, differ in expression levels between the sexes throughout adult life (Ober et al., 2008) and affect complex biological pathways from metabolism to reproduction. While

some sex difference in morphology and disease may be attributable to the presence of sex hormones (Manwani et al., 2014) or the Y chromosome (Charchar et al., 2002; Charcher et al., 2003; Bellott et al., 2014), transcriptome analyzes hold the promise of comprehensively uncovering the gene expression differences that underlie these adult phenotypic and developmental sex differences in mammals.

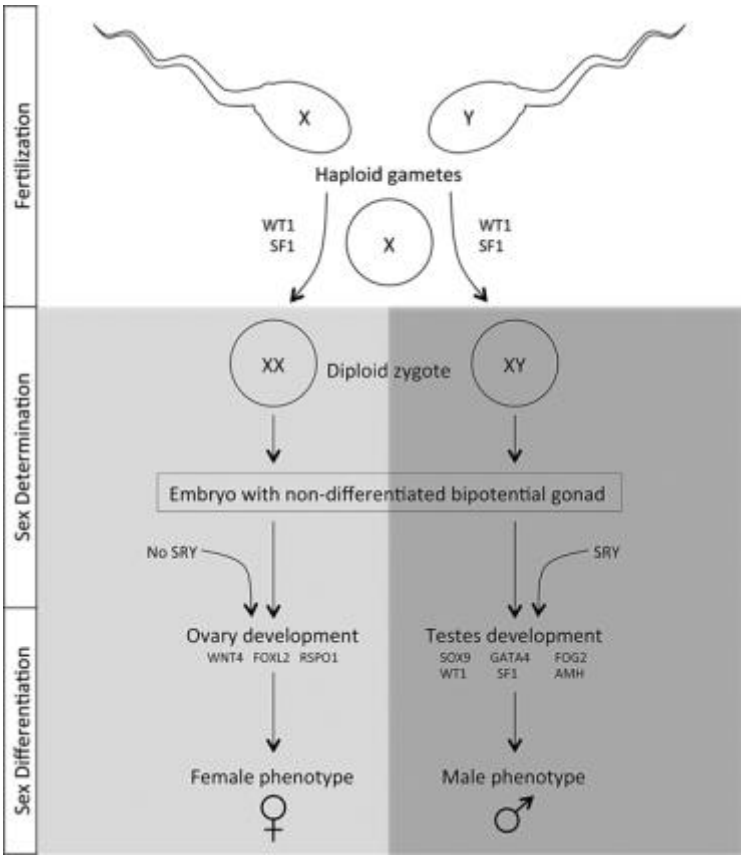


Figure 4.1 *Sex determination and sexual differentiation in humans*. Haploid gametes fuse during fertilization (top) to produce a diploid zygote. Diploid zygotes become an embryo with a non-differentiated, bipotential gonad through the WT1 and SF1. Sex determination in females (pathway highlighted in light grey) is dependent on the absence of a Y chromosome and lack of SRY expression, leading to ovary formation via the expression of WNT4, RSPO1, and FOXL2. Sex determination in males (pathway highlighted in dark grey) is based on expression of the sex-determining region Y (SRY) gene, which signals testes development through expression of GATA4, SOX9, FOG2, WT1, SF1, and AMH. Secretion of sex steroids and other sex-specific hormones by the developing gonads leads to further sex differentiation, ultimately, leading to the maturation of an adult human with a distinctly male or female phenotype.

4.4 Sex-biased gene expression in gonadal tissues

It is clear that divergent gonadal development and sex-specific hormones play a critical role in the early etiology of sex differences in mammals. As the most dimorphic feature in taxa with two distinct sexes, the testes and ovaries are essential in maintaining the dimorphic character state across both development and generations. Since these dimorphic tissues are involved in the ultimate of sex-specific functions—the development of sperm and eggs—one would predict that the greatest amount of sex-biased expression would be found among those gonadal tissues.

In our survey of the recent literature, there currently is no genome-wide analysis of sex-bias across human reproductive tissues. However, using observations in mice transcriptomes as an evolutionary proxy to understand the extent and prevalence of sex-bias in humans, it was found that murine ovaries and testes harbor the largest amount of sex-biased expression (Rinn et al., 2004). In this particular study, testes were functionally enriched for immunosuppression processes while ovarian genes were enriched for drug and steroid metabolism. Yet, rodents may not provide the best basis for comparison, as sex-biased gene expression profiles appear to be largely species-specific (Si et al., 2009). New data are needed to determine sex differences in gene expression in human reproductive tissues, including those tissues not directly associated with the testis and ovary.

4.5 Sex-biased gene expression in non-gonadal tissues

In contrast to sex-specific gonadal tissues, somatic tissues provide functions common to both sexes such as metabolism and physiology and generally lack

conspicuous sexual dimorphism at the morphological level. Although these tissues and functions are shared between both sexes, evidence suggests that sexual dimorphism in gene expression exists (Isensee et al., 2008; Reinius et al., 2008; Si et al., 2009; Zhang et al., 2011; Michael et al., 2011; Jansen et al., 2014; Xu et al., 2014). So, what is the extent of sex-bias in gene expression in human somatic tissues? In particular, how much sex-bias is present in tissues that appear not to bestow a dimorphic phenotype? Furthermore, are there differences in the number of sex-biased genes between tissues? And are sex-biased genes common between tissues or conserved between species? Below, we attempt to answer these questions from recent genome-wide analyzes examining sex-biased gene expression in a handful of major somatic organs (Table 4.1).

Table 4.1

A survey of recent genome-wide analyses examining sex-biased gene expression in somatic organs

Study	Tissue	Male Samples	Female Samples	% of genes with sex-biased expression
Isensee et al. 2008	Heart	< 40y: 4 50-65y: 5	<40y: 3 50-65y: 5	< 40y: 0.70% (93/13,169) 50-65y: 0.95% (125/13,169)
Si et al. 2009	Kidney	9	10	Glomeruli 0.25% (26/10,561) Tubuli 0.48% (50/10,478)
Zhang et al. 2011	Liver	112	112	3.75% (1,249/33,250)
Michael et al. 2011	Minor salivary gland	4	5	1.07% (360/33,717)
Reinius et al. 2008	Occipital cortex	4	4	2.92% (1,349/46,128)
Jansen et al. 2014	Peripheral blood	1,814	3,427	3.15% (582/18,495)
Xu et al. 2014	Prefrontal cortex	32	14	10.02% (1,489/14,851)
Welle et al. 2008	Skeletal Muscle	15	15	1.94% (336/17,282)
Roth et al. 2002	Skeletal Muscle	5	5	20% (210/1,000)

Note. Numbers in the parentheses indicate the total number of genes or transcripts with sex-biased expression over the total number included in the study.

4.5.1 Liver

As one of the most homogeneous organs, the liver is responsible for the filtration of blood, metabolism of drugs, and detoxification of chemicals. Some of the first studies to assay molecular differences between the sexes in mammals compared the livers of male and female rats (Gustafsson et al., 1983; Roy and Chatterjee, 1983) spurring additional research into differences in the expression of genes involved in metabolism in rodent models. While these early studies focused on single genes, recent work has examined sex-biased genes in the liver across the genome. Zhang et al. (2011) identified 1,249 sex-biased genes (from ~20,000 genes in humans) using 224 human non-tumorous liver samples obtained from subjects undergoing liver surgery. Functional gene ontology (GO) clustering of these genes revealed that lipid metabolism was among the top molecular functions represented, affirming physiological differences seen between the sexes.

4.5.2 Heart

As the human transcriptome varies from tissue to tissue, variation is also expected across different developmental stages. Isensee et al. (2008), using microarrays in normal human left ventricular myocardial samples, examined sex-biased gene expression among different age groups by comparing nine individuals greater than 40 years old to fourteen individuals between the ages of 50 and 65 years. These different age groups controlled for pre- and postmenopausal effects. Both the <40 year age group and the 50–65 year old age group exhibited less than one percent of sex-bias genes.

4.5.3 Brain

Differences in the brain are among the most intriguing contrasts between males and females. Physical differences have been noted between the sexes in gray and white matter volume (Gur et al., 1999) and total brain volume (Ruigrok et al., 2014). On a functional level, differences in cognitive function, attention, and memory have been found (Gur et al., 2012). In addition, studies have shown that there are differences in the incidence, severity, and progression of well-known neurodegenerative diseases, such as Alzheimer's disease and dementia (Li and Singh, 2014). Xu et al. (2014) examined sex-biased expression in the prefrontal cortex, an area of the brain responsible for cognition, and found the largest fraction of sex-biased genes in non-gonadal tissues seen in such studies (Table 4.1). While the prefrontal cortex has been shown to have the largest extent of sex-biased expression, the occipital cortex expresses only about 3% of its genes in a sex-specific manner (Table 4.1; Xu et al., 2014; Reinius et al., 2008). This difference between tissue subtypes is also seen in the kidney, where tubuli contain nearly twice the number of sex-biased genes compared to the glomeruli (Table 4.1; Si et al., 2009).

These tissue-specific surveys, while relatively few, reveal modest numbers of non-gonadal sex-biased genes, relative to testes and ovaries, with the prefrontal cortex of the brain harboring the largest amount of sex-biased gene expression of the somatic tissues.

4.6 Non-conserved nature of sex-bias at tissue- and species levels

Whether sex-biased genes are specific to particular tissues or are common across all tissues and developmental stages is another important question. Two microarray

studies have made direct comparisons in mammalian gene expression among multiple tissues, but only in mice. While both studies used different metrics to identify sex-biased expression, there are similarities in their results. In the mouse liver, Rinn et al. (2004) and Yang et al. (2006) observed less than 1% of genes with sexually dimorphic expression at greater than three-fold change in the liver and the brain (specifically the hypothalamus in Rinn et al., 2004). Both studies also observed little overlap of sex-biased genes across tissues indicating tissue specificity. Perhaps, the best systematic study of sex-biased expression across multiple tissues in humans to date is a meta-analysis by Jansen et al. (2014). In this study, the authors compared the sexually dimorphic transcriptome in peripheral blood to that found in skeletal muscle by Welle et al. (2008) and Roth et al. (2002), and in the liver by Zhang et al. (2011). Jansen and colleagues noted little overlap between sex-biased genes of these different tissues. However, such a meta-analyzes may be prone to biases specific in each study (see below) and, ultimately, will underestimate the number of common sex-biased genes.

An evolutionary approach to sexually dimorphic gene expression may also provide insight into the development of sexually dimorphic gene expression and the relative amount of dimorphism to expect. Again, however, only a few such studies exist. Reinius et al. (2008) published a comparative study into sex-biased expression across primates using samples from the occipital cortex of the brain. The occipital cortex is a potential candidate for sex differences on the molecular level as it is responsible for higher behavior functions. In this study, researchers compared expression levels from humans, cynomolgus macaques, and common marmosets. At 2.92%, the human occipital

cortex appears to contain a lesser extent of sexual dimorphism than the prefrontal cortex. Of the 1,349 differentially expressed genes between the sexes in humans, approximately only 6% of these were conserved in their bias between humans and macaques, and only two genes were conserved across all three species. In the kidney, Si et al. (2009) observed nine transcripts that were conserved in their bias between mouse and human kidneys. These genes represented only 13% of all human sex-biased genes and 0.8% of mouse sex-biased genes.

This lack of conservation seems to be consistent with what has also been shown in other tissues as well. However, this low conservation in sex-bias significantly differs from what has been observed in the genus Drosophila, where the majority of genes are observed to maintain their sex-bias across phylogenetically distant species (Zhang et al., 2007). It is still unclear as to why we see decreased conservation of biased genes in mammals, and future comparative studies should be able to shed light on this question. The non-conservative nature of sex-biased expression between species may also allude to population-level differences. In Drosophila, such differences in sex-biased expression profiles have been identified between populations (Huylmans and Parsch, 2014) indicating that sex-bias shifts can occur quite rapidly.

4.7 Challenges to studying sex-biased gene expression in humans

More work is clearly needed to understand the prevalence and conservation of sex-biased genes between tissues and across mammalian species. However, a number of technical challenges must be overcome before we can get a true picture of human sex-biased architecture. First, most studies use different metric and statistical cutoffs for

determining whether a gene is differentially expressed or male- or female-biased. This variance in metrics makes cross-tissue and cross-species comparisons extremely difficult. In Drosophila, it has been demonstrated that different metrics for determining sex-biased genes produce different lists of genes with little overlap (Assis et al., 2012). This seems to be the case for human tissues as well, judging by the differences in amount of sex-biased gene expression found between two skeletal muscle studies (Table 4.1). Future analyzes would benefit from a standard or composite metric for determining differences in sex-biased gene expression.

Other factors such as environmental, dietary, and age differences among the surveyed samples may confound the study of sex-biased expression. The human transcriptome responds to changes in environment and lifestyle by varying gene expression levels (Maretty et al., 2014) yet most studies of sex-related gene expression do not control for these factors. In addition, current studies primarily sample adults since at sexual maturity males and females are the most sexually dimorphic, but studies in Drosophila indicate that juvenile sex-bias occurs as well (Arbeitman et al., 2002; Perry et al., 2014). The changes in sex-biased expression across the development of an organism will provide more insight into the dynamic nature and regulation of these genes. In a similar vein, differences in sampling protocols can profoundly affect the analyzes and results of gene expression studies.

Lastly, all of the aforementioned studies were carried out using microarray technology, which is prone to systematic error such as gene and transcript ascertainment biases and cross-species divergence biases (Marioni et al., 2008). As next-generation

sequencing has dramatically decreased in cost over the last decade, whole transcriptome sequencing has made sequence data more accessible. RNA sequencing (RNA-seq) provides a far more nuanced picture of the sex-biased transcriptome, allowing for the identification of novel transcripts and splice sites. In addition, many researchers look to freely available data from publicly supported databases (e.g., NCBI's SRA database and European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) ArrayExpress) to help leverage their results. In the case of sex-biased expression, the bank of available RNA-seq datasets useful for examining the phenomenon in humans is quickly growing (Figure 4.2, Supplementary Table 1). Unfortunately, the most understudied of sex-biased tissues—testis and ovary—are still lacking in available data.

4.8 Future advances

The need for gender inclusion in research is quickly becoming apparent as we learn about sex-biases in disease prevalence and drug efficacies. Women and men are disproportionally affected by diseases including stroke (Roy-O'Reilly and McCullough, 2014), dementia (Li and Singh, 2014), and coronary artery disease (Gordon et al., 1978; Kannel and Feinleib, 1972; Patrick et al., 1982; Lerner and Kannel,1986). In terms of cancer, men exhibit a higher prevalence of certain types of cancer, even when environmental variables are controlled (Kiyohara and Ohno, 2010, and references therein). A large number of diseases show profound differences in sex-bias for mortality, as seen from reorganized data from the World Health Organization (Figure 4.3). Some of these differences are specific to geography, potentially indicating an environmental component (Figure 4.3). Yet, many of these sex differences co-vary with geography and

ethnicity, which is consistent with the results found in studies attempting to relate ethnicity to gene expression using HapMap data (Tabassum et al., 2013). Future studies of sex-biased gene expression will benefit by sampling multiple populations living under varying environmental conditions and lifestyles.

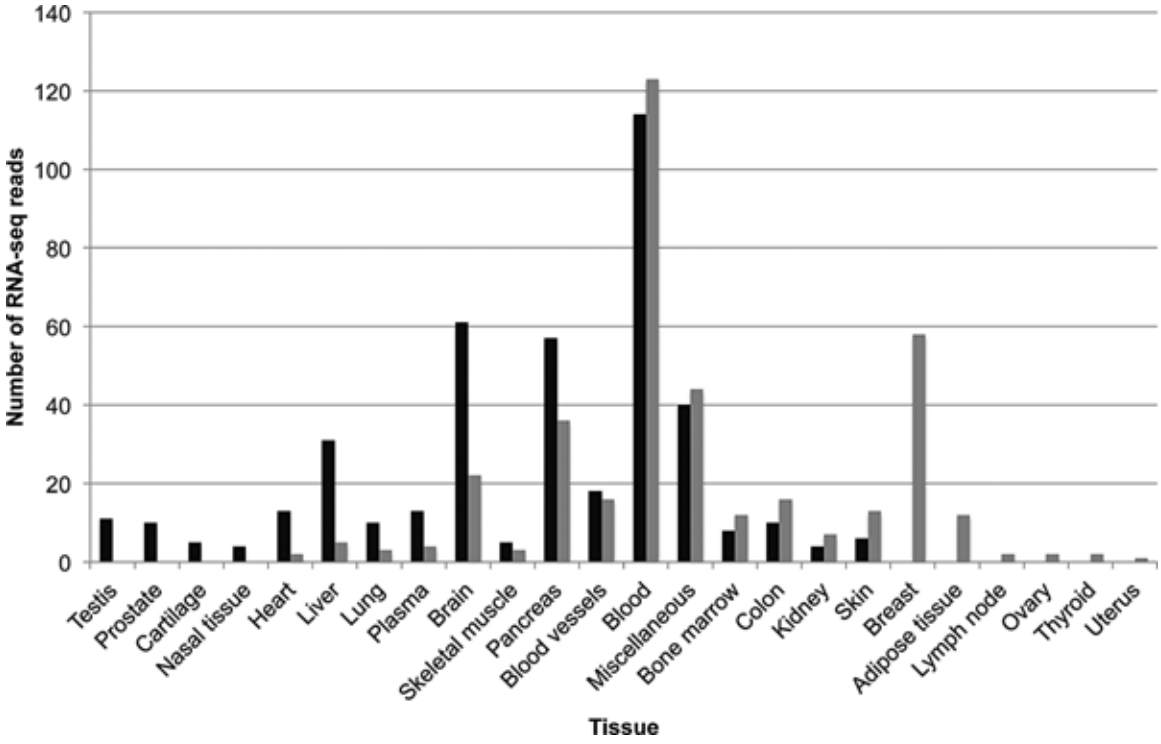


Figure 4.2. *Number of male and female human adult RNA-seq reads currently available in NCBI for various tissues*. Bar graph depicting the number of adult male and female RNA-seq reads publicly available through NCBI's Sequence Read Archive (SRA) database (Leinonen and Sugawara, 2011). Number of reads were determined by searching NCBI's SRA database for the query term, "homo sapiens[orgn: txid9606]", in December of 2014 and filtering for RNA results that were publicly available. Reads were excluded if: neither sex nor tissue was specified, the sample was not from adult tissue, or if the sample tissue was unhealthy (e.g., cancer, tumor, lesion). Black bars represent male samples (420 in total), while grey bars represent female samples (383 in total).

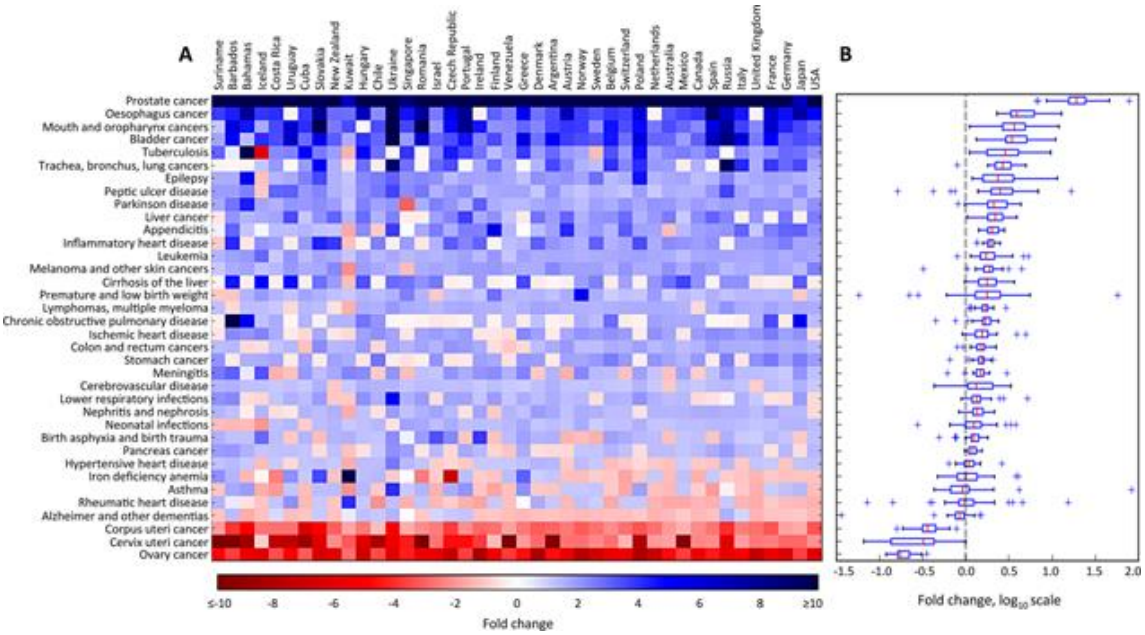


Figure 4.3. *Sex-biased death rate across disease per country*. A: Heatmap showing ranked fold changes (see legend) between males and females in age-standardized death rates for diseases (rows) across countries (columns). Data was obtained from the World Health Organization for the year, 2008. Nations are ranked from left to right by increasing 2008 GDP and diseases ranked by average fold change of death rates across all countries. Negative values (red) indicate a female-biased death rate while positive values (blue) indicate a male-biased death rate. B: Boxplot showing the median and variance in the fold changes on a log scale for each disease in all countries. Each box extends from the lower to upper quartile values of the data, with a line at the median. Data outside of the first and third quartiles indicated by blue crosses. Gray dashed line at zero indicates the fold change at which there is no bias towards either sex. Male-biased values are to the right of the dashed line and female-biased values are to the left.

Recently, the National Institutes of Health announced an initiative to promote gender inclusion in research and funding (Clayton and Collins, 2014), recognizing the dearth of female samples (see Figure 4.2) and lack of sex-biased analyzes in preclinical and basic research (Mogil and Chanda, 2005; Beery and Zucker, 2011). The preference for male study subjects was previously justified due to women having 28-day hormonal cycles that can affect multiple genetic pathways, adding difficult-to-control variables to the study (Wizemann and Pardue, 2001). However, sex-bias differences in disease and

drug response that may potentially have an effect on over half of our species' population serves as a substantiated call to promote gender parity in research sampling and analysis. To encourage this initiative, the NIH plans ask potential grantees to specify their efforts in providing equal attention to male and female cells and tissue samples in research and analyzes, and reviewers will be trained to seek out this information. These new efforts are good news to the field of sex-biased gene differences, as more human-based samples and studies are needed to understand this dynamic phenomenon and its impacts on health and disease.

CHAPTER 5

EVOLUTIONARY DYNAMICS OF SEX-BIASED GENE EXPRESSION IN PRIMATES

5.1 Abstract

Sexually dimorphic traits are extensive across taxa and among the most rapidly evolving characters between species. Primates, including humans, are no exception. Studying the evolution of genes that exhibit sex-biased gene expression may uncover important details about the evolutionary processes that drive and/or maintain sexually dimorphic features. However, little work has been done to examine sex-biased gene expression and its evolution in humans and their close primate relatives. Here, as an initial phase of a larger study to investigate the evolution of sex-biased genes in humans and mammals, we estimate baseline levels of sexually differentiated gene expression in a sexually monomorphic tissue. We apply four independent estimates to identify whether sex-biased proteins expressed in the liver are conserved or rapidly evolving between humans, chimpanzees and macaques. Using evolutionary rate estimates, we test for evidence of rapid evolution of sex-biased proteins, a pattern observed in a variety of other taxa. We find that defining sex-biased expression is difficult, as metrics for identifying differentially expressed genes result in differing sets and numbers of sex-biased genes. However, each of the metrics consistently identify relatively small numbers of sex-biased genes in these species, indicating that the amount sex-biased expression in primates is not as extensive as that observed in other species such as *Drosophila*. Furthermore, We find that sex-biased proteins in humans do not show significantly higher rates of evolution

than unbiased proteins. This, however, could be the result of a number of caveats related to the use of primate data.

5.2 Introduction

Sexual dimorphism exists in a large variety of taxa, including humans and non-human primates. Morphologically, primates are sexually dimorphic in body mass, hair, skin color, and tooth and skeletal shape (Plavcan 2012). In addition to this, differences between males and females exist in behavior and aggression as well (Bernstein 1978). These differences persist although the sexes share a common genome, with the exception of a handful of genes on the Y chromosome. The differences observed between the sexes, therefore, are due to the differential expression of genes between the sexes (Connallon and Knowles, 2005; Rinn and Snyder 2005; Ellegren and Parsch, 2007). Sex-biased genes – those exhibiting differences in gene expression between males and females – provide an underlying genetic basis to sexually dimorphic traits.

Relatively little work has been done to understand sex-biased expression in humans and close primate relatives. Most studies of sexually dimorphic expression have been carried out using model organisms such as fruit flies (Jin et al., 2001; Parisi et al., 2003; Ranz et al., 2003; Singh and Kulathinal 2005), nematodes (Reinke et al., 2000; Thoemke et al., 2005), and mice (Yang et al., 2006). Studies that have examined sex-biased gene expression in humans have mainly used microarray technology, which can introduce systematic error in the form of ascertainment bias (Marioni et al., 2008). These studies (reviewed in Rigby and Kulathinal 2015) report a relatively small occurrence of sex-biased expression in human tissues that ranges from 0-20%.

Understanding sex-biased genes may provide new insight into sexually dimorphic physiology and susceptibility to disease. There is an increasing amount of evidence that the majority of human diseases have sex-biased or sex-specific effects (Ober et al., 2008). These differences may exist in disease prevalence, age of onset, or severity. Well known examples of such diseases include cardiovascular disease (Choi et al., 2007), Alzheimer's dementia (Li and Singh, 2014) and stroke (Roy-O'Reilly and McCullough, 2014).

Comparative studies of sex-biased expression in other organisms, particularly *Drosophila*, have revealed some common trends. One frequent observation is that proteins encoded by sex-biased genes exhibit greater sequence divergence than unbiased genes (Ellegren and Parsch, 2007). Specifically, male-biased genes have higher turnover rates and evolve faster than female-biased and unbiased genes (Meikeljohn et al., 2003; Zhang and Parsch, 2005; Zhang et al., 2007). There is some evidence that this may also be true in some primate tissues as well (Reinius et al., 2008; Wyckoff et al., 2000), but additional analyses are needed to determine whether this is true for multiple tissues.

We used RNA-seq data from published liver samples in three primate species (*Homo sapiens*, *Pan troglodytes* and *Macaca mulatta*) with three goals in mind. First, we aimed to determine the extent of sex-biased gene expression in the somatic tissue of humans and primates. Second, we intended to develop an empirical framework for detecting sex-biased gene expression in humans. Third, we aimed to estimate the evolutionary rate of sex-biased genes in somatic tissues in order to determine if these genes show the common trend of rapid evolution compared to unbiased genes.

Our reasoning for choosing the liver to examine sex-biased expression is four-fold. First, the liver is one of the most homogeneous tissues in terms of cellular composition (Balashova and Abdulkadyrov 1984). This prevents the introduction of bias due to differences in cell content from tissue sampling. Second, sexually dimorphic gene expression in the mouse has been most extensively studied in the liver (Khil et al., 2004; Zhang et al., 2012; Yang et al., 2006; Bur et al., 2009), which facilitates downstream comparisons within mammals. Third, among freely available published RNA-seq samples, those sampled from the liver are the most abundant. Fourth, the liver presents a monomorphic tissue that likely is not under sex-specific selection for fertility, so a baseline level of sex-biased gene expression can be estimated.

Since different metrics for determining sex-biased genes can produce different lists of sex-biased genes (Assis et al., 2012), we applied four widely-used, independent metrics to identify sex-biased genes in each species. With the goal of understanding the evolutionary dynamics of sex-biased gene expression among these species, we compared the levels of sex-bias and amino acid substitutions of one-to-one orthologs between humans and chimpanzees.

5.3 Methods

5.3.1 Gene Expression Data

We analyzed published RNA-seq data sequenced by Illumnia GA IIx instruments from three studies (Brawand et al., 2011; Blekhman et al., 2010; Perry et al., 2012). Single- and paired-end RNA-seq reads for liver samples in three primate species (*H. sapiens*, *P. troglodytes* and *M. mulatta*) were downloaded from NCBI's SRA database

(Table 5.1; a complete list of sample and accessions can be found in Appendix B). Sequences were trimmed for quality and sequencing adapters using Trimmomatic (v0.36; Bolger et al., 2014). Using a 4 base-pair sliding window approach, Trimmomatic was used to trim portions of reads with a phred33 score below an average value of fifteen. For samples with paired-end reads, the 2nd read in each pair was discarded to prevent bias in mapping between samples, effectively making all samples single-end.

Table 5.1

Summary table of species and sample numbers

Tissue	*H. sapiens*		*P. troglodytes*		*M. mulatta*	
	Male	Female	Male	Female	Male	Female
Liver	10	9	9	9	9	9

Human, chimpanzee, and macaque CDS sequences were downloaded from Ensembl (Ensembl release 84 – March 2016; Flicek et al., 2014). For genes with multiple transcripts, the CDS of the longest transcript was used for transcriptome alignment. Trimmed RNA-seq reads were aligned to the respective transcriptomes with Burrows-Wheeler Aligner (BWA version 0.7.5a-r405; Li and Durbin, 2009) using default parameters. Output files from BWA were converted from SAM to BAM format and sorted using SAMtools (version 0.1.19; Li et al., 2009). Read counts and the number of fragments per kilobase per million mapped (FPKM) were determined using eXpress (version 1.5.1; Roberts and Pachter, 2012).

5.3.2 Identification of Sex-Biased Genes

Assis et al., (2012) noted that different metrics for differential expression identify different numbers and sets of sex-biased genes in *Drosophila*. Because there is no consistent metric for identifying and defining a sex-biased gene (Ellegren and Parsch,

2007; Meisel, 2011), we chose to compare the results of several metrics. To identify genes differentially expressed between the sexes in each tissue of each species, we applied four widely used, independent metrics, DESeq2 (Love et al., 2014), CuffDiff (Trapnell et al., 2013), edgeR (Zhou et al., 2014) and a 2-fold change difference in FPKMs between males and females.

5.3.3 Conservation and evolution of Sex-Biased Genes

In order to determine the extent of conservation of sex-biased genes across the three species, we obtained single copy orthologs for human, chimpanzee, and macaque from Ensembl (release 84; Flicek et al., 2014). We then examined whether or not each of these orthologs were labelled as sex-biased in the three species by our differential expression metrics.

To understand divergence and selective pressures acting on primate genes, we first obtained single copy orthologs for human and chimpanzee from Ensembl (release 80; Flicek et al., 2014). Translated nucleotide sequences for each gene were aligned using MUSCLE (version 3.0; Edgar, 2004). To reduce impact of misaligned regions around insertions (Markova-Raina and Petrov, 2011), +/- six nucleotides surrounding insertions

Table 5.2

Number of sex-biased genes identified in each species by various metrics

Metric	*H. sapiens*			*P. troglodytes*			*M. mulatta*		
	MBG	FBG	UBG	MBG	FBG	UBG	MBG	FBG	UBG
edgeR	16	65	22,672	139	353	18,252	61	41	21,802
DESeq2	19	55	15,158	98	203	12,626	93	54	14,962
Cuffdiff	27	59	14,761	174	239	13,980	80	44	16,768
2-fold	416	466	19,413	362	544	17,852	313	241	21,350

Note. MBG: male-biased genes; FBG: female-biased genes; and UBG: unbiased genes.

were masked using 'N' and corresponding codons were ignored in final analyses. K_a/K_s, a commonly used homolog to dN/dS, was calculated using the sequinR (version 3.1.4; Charif and Lobry, 2006).

5.4 Results

5.4.1 Identification of sex-biased genes

Depending on the metric used, we identified 0.4-5.6% of sex-biased genes in humans, 2.3-4.9%of sex-biased genes in chimps, and 0.5-2.5% 0f sex-biased genes in macaques (Table 5.2). In humans and chimpanzees, we consistently see more female-biased genes than male-biased genes identified by the metrics, however, this is not the case for macaques. Our 2-fold approach for identifying differentially expressed genes between the sexes, returned the highest number of sex-biased genes across all species (Table 5.2).

The overlap of genes identified by our four metrics is small, particularly for male-biased genes. For humans, we observed only 8 and 27 genes identified as male- and female-biased in all metrics, respectively (Figure 5.1). In chimpanzee and macaque samples, 70 and 32 genes were labelled as male-biased by all four metrics, respectively. It appears, therefore, that male-biased genes were more robustly identified than female-biased genes.

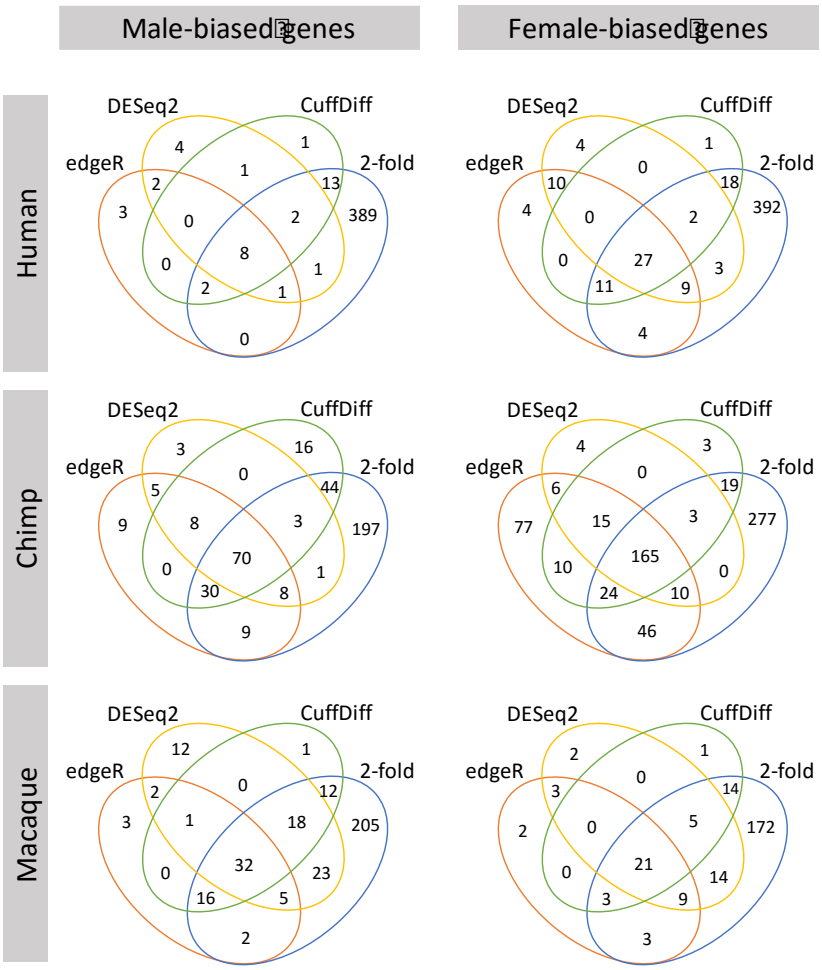


Figure 5.1. *Comparison of four metrics used to classify sex-biased genes*. Venn diagrams showing the overlap of male-biased genes (left) and female-biased genes (right) among the four metrics used.

Due to the lack of overlap among metrics, we see that there are a large number of genes that were identified as sex-biased by at least one metric (Figure 5.2), with fewer genes identified as the number of metrics increases. Rather than choosing one metric for labelling sex-biased genes in our following analyses, we defined sex-biased genes as those identified by any combination of $n=2$ metrics, similar to Assis et al., (2012). Assis et al. (2012), noted that as the variance in sex-biased expression and distance between

male- and female-biased genes increases with *n*. Thus, in choosing too few *n* metrics the number of false positive rate would be too great, yet in including too many metrics, the false negative rate would be large (Assis et al., 2012).

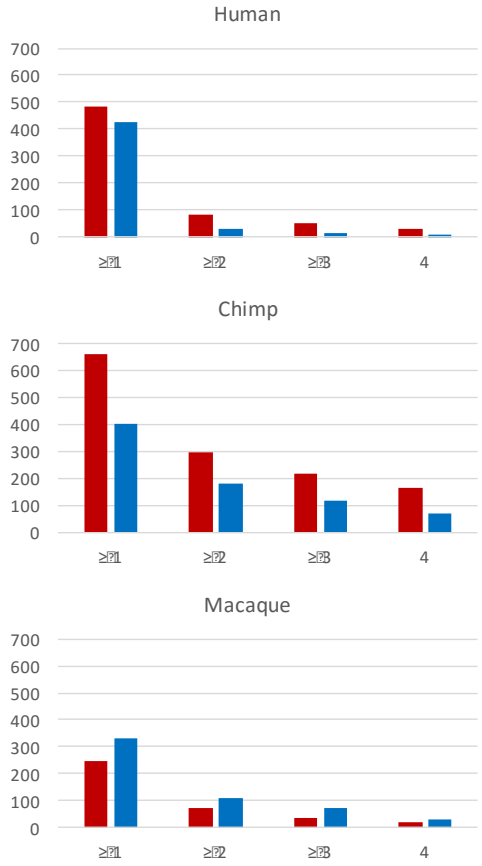


Figure 5.2. *Comparison of sex-biased genes identified by n metrics in each species*. Stacked bar graph showing the number of male- and female-biased gene identified by *n* metrics.

5.4.2 Evolution of sex-biased genes in primates

To study the evolution of sex-biased genes in our three primate species, we obtained single-copy orthologs and the human-chimp K_a/K_s values for each. Among single-copy orthologs, we notice very little overlap among sex-biased genes in these species (Figure 5.3). Only one ortholog is identified as male-biased in all species and

female-biased in all species (Figure 5.3). In males, the ortholog for human phospholisase A2 group IIA (PLA2G2A, ENSG00000188257) was male-biased in all samples. In females, the ortholog for human insulin-like growth factor binding protein 5 (IGFBP5, ENSG00000115461) was female-biased in all samples. To study the sequence divergence, we calculated evolutionary rate as K_a/K_s between humans and chimpanzees for male-, female-, and unbiased genes in humans. We found no significant differences in evolutionary rate between sex-biased and unbiased genes.

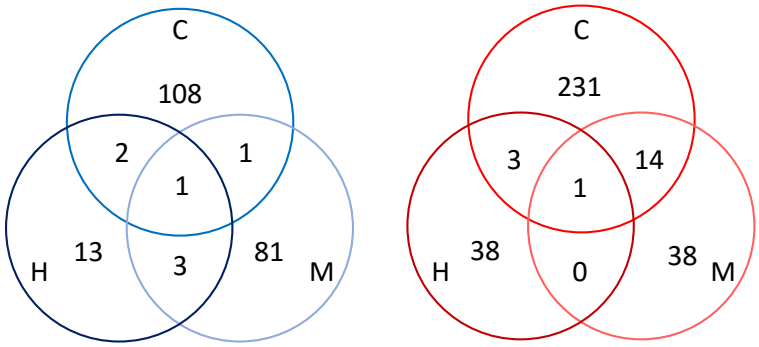


Figure 5.3. *Conservation of sex-biased genes across primates*. Venn diagrams showing the conservation of male- (left, blue) and female-biased (right, red) genes among single copy orthologs of the three species.

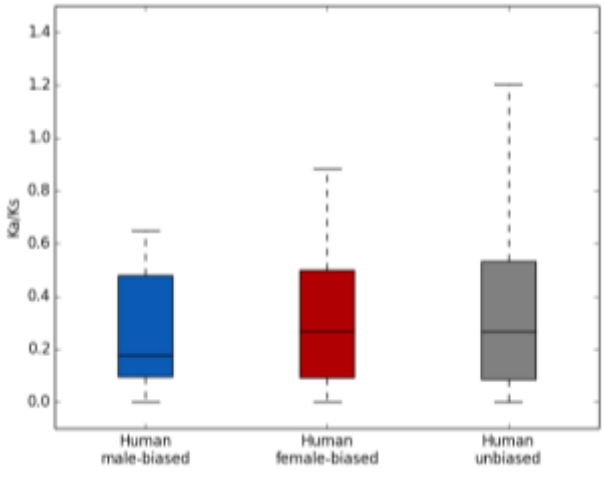


Figure 5.4. *Sequence divergence in sex-biased and unbiased genes*. Boxplot of K_a/K_s values between humans and chimpanzees for male-biased, female-biased and unbiased genes.

5.5 Discussion

Despite its broad implications for our understanding of the evolutionary processes that maintain and promote sex, as well as the origins of sex disparities in disease (Rigby and Kulathinal 2015), little is known about the evolution of sex-biased expression in humans. Here, we have examined sex-biased expression across a single somatic tissue of humans, chimps and macaques in order to explore the extent and evolutionary dynamics of baseline sex-biased expression in primates.

For all species in this analysis, less than 5% of all genes were identified as being sex-biased by any of our 4 independent metrics. This is consistent, for humans at least, with previous microarray studies, which report similarly small proportions of sex-biased gene expression (Rigby and Kulathinal, 2015). In other taxa, much larger proportions of sex-biased expression have been reported. In *Drosophila*, proportions of over 50% have even been reported (Ellegren and Parsch 2007). This disparity could be due to differences in the level of sexual dimorphism between these organisms, however, it could also be a result of a lack of definition for sex-biased gene expression. Currently, there is no clear definition or singular metric for identifying differentially expressed genes between males and females, which makes it impossible to make accurate comparisons between studies. Adopting a single method for identifying sex-biased genes is easier said than done, as there are a plethora of methods for carrying out differential expression analyses.

In order to reduce both false positives and negatives in the identification of sex-biased genes, we chose to follow the protocol of Assis et al. (2012) to define sex-biased genes as those labelled biased by at least two metrics. When we examined one-to-one

orthologs among primate species, we noted very little conservation of sex-bias across organisms. In fact, only one gene was identified as male- or female-biased in all three species. The gene PLA2G2A, which was noted to be male-biased in all three species, is an enzyme that catalyzes the hydrolysis of phosphoglycerides, causing the release of free fatty acids and lysophospholipids. In one study, mutations in PLA2G2A were noted to be associated with changes in levels of an important secretory phospholipase (Shuvalova et al., 2015) associated with an increased risk of coronary heart disease (Boekholdt et al., 2005; Mallat et al., 2007). This is particularly interesting because it is well known that there are sex differences in the risk, onset, and prevalence of heart disease in humans (Klein, 1996). In all three species, the gene IGFBP5 was labelled as female-biased. Interestingly, this gene has been noted to have roles pertaining to breast cancer (Ghoussaini et al., 2014), a commonly female-biased disease. It is unclear from our analysis whether these sex-biased genes represent an ancestral condition, and further research will be required to determine this.

In *Drosophila*, it is well-established that male-biased genes exhibit higher sequence divergence (Ellegren and Parsch, 2007). We examined the evolutionary rate of single copy orthologs in humans and chimpanzees to determine if this trend also existed in humans. Unlike Drosophila, we found no significant difference between the K_a/K_s ratio of male-biased and unbiased genes. Other studies have found a signature of rapid evolution in male-biased genes in primate lineages within the brain (Reinius et al., 2008) and the gonads (Wyckoff et al., 2000). We suspect that we may not see this trend in our

study due to the fact the primate liver may not be under as much sexual selection as those tissues that are involved in reproduction and reproductive behavior.

The absence of this trend may also be due to one or more potential caveats associated with our analysis. For example, the difficulty in obtaining biological samples from primates resulted in the lack of readily available, high-quality primate RNA-seq samples and, as a result, our analysis contains unequal and small sample sizes, particularly for females. This lack of female samples is a direct result of an inequality in the availability of female RNA-seq samples through NCBI (Rigby and Kulathinal 2015), which has been recently brought to light by the NIH (Clayton and Collins, 2014). Furthermore, our data mainly consists of single-end RNA-seq reads, which map poorer than paired-end data. Additionally, the study of gene expression levels can be complicated by environmental, dietary, and age differences amongst samples (Maretty et al., 2014), which we have been unable to control for here.

This study provides an important initial step towards our understanding of sex-biased expression in humans and other primates. The fraction of sex-biased genes in these organisms is much less than previous observations in other taxa, particularly *Drosophila*. A more appropriate comparison is with mouse tissue samples which has more publically available sequences, at present. We plan to perform tissue-by-sex comparisons in a future study when higher quality primate samples are more abundant. Such a comparison would provide greater power to detect evolutionary rate differences among sex-biased genes in mammals.

CHAPTER 6

CONCLUDING REMARKS

Sexual dimporphism is extensive and prevalent across taxa. Understanding this phenomenon on both a phenotypic and molecular level will provide insight into the mechanisms and forces that have shaped sexual dimorphic characters in the recent past. In this thesis, I have advanced the current knowledge about sex-biased gene expression from *Drosophila* to primates.

First, I examined the literature for theoretical and experimental approaches to measuring fitness in *Drosophila*. I discuss how the study of fitness has been facilitated by the partitioning of fitness into its compnents. The measurement of these components can differ for males and females, due to the mating behavior of this system. I point out, however, that due to the possibility of tradeoffs between fitness components, it will be important to develop complete knowledge about the correlation between these variables when performing a study.

Second, I used a linear regression analysis to determine the relative contributions of intrinsic factors to the evolutionary rate of reproductive-related genes. The intrinsic correlates of evolutionary rates tested explain only ~20-30% of variation in dN/dS. This suggests that either extrinsic factors that may influence evolutionary rate (i.e., external adapted forces) have a stronger effect on evolutionary rate or there are other intrinsic factors (i.e., evolutionary constraints at the cellular level) at work. As expected, I observe that male gonadal genes have a higher evolutionary rate than non-gonadal genes. Also, of the intrinsic factors that I tested, CAI has the largest relative contribution to dN/dS in

male gonadal genes. This measure of codon usage bias influences the largest portion of the variation seen in the evolutionary rate of both reproductive-related and non-reproductive genes. It was found that there was significantly less codon bias in male genes, with this lack of evolutionary constraint explaining higher dN/dS.

Third, I reviewed current literature in sex-biased gene expression in humans. While few studies of sex-biased gene expression in humans exist (primarily using microarray technology), those that have examined these genes find low proportions of sex-biased expression across human tissues. Unfortunately, little sex-specific data has been produced in order to advance the study of sex-biased expression in humans. I discussed how the National Institute of Health has recognized this problem and plans to resolve this issue in the future by requiring equal inclusion of male and female samples in future studies.

Finally, to expand the knowledge of sex-biased expression in humans, I identified sex-biased genes on a tissue-specific level in humans and primates using multiple independent metrics. I found littte overlap in the number and sets of sex-bised genes among metrics, but even less overlap across species. Only one gene was male-biased and another female-biased, across humans, chimpanzees and macaques. These genes were associated with commonly-known sex-biased diseases, but further study will be necessary to determine if the sex-biased nature of these genes is involved in the causality of these diseases. Additionally, I examined the evolutionary dymanics of sex-biased genes identified in humans by comparing the K_a/K_s. ratios of unbiased, male- and female-biased genes. No differences were seen between the biased and unbiased classes, which

could be attributable to small sample size or other factors. More investigations of sex-biased gene expression in humans, primates and other mammals are sorely needed to understand the etiology of sexually dimorphic traits and disease.

REFERENCES CITED

Alatalo RV, Carlson A, Lundberg A. The search cost in mate choice of the pied flycatcher. Animal Behaviour. 1988 Feb 1;36(1):289-91.

Anderson GD. Sex differences in drug metabolism: cytochrome P-450 and uridine diphosphate glucuronosyltransferase. The journal of gender-specific medicine: JGSM: the official journal of the Partnership for Women's Health at Columbia. 2001 Dec;5(1):25-33.

Andersson M. Evolution of condition-dependent sex ornaments and mating preferences: sexual selection based on viability differences. Evolution. 1986 Jul 1:804-16.

Armbrust EV, Galindo HM. Rapid evolution of a sexual reproduction gene in centric diatoms of the genus *Thalassiosira*. Applied and environmental microbiology. 2001 Aug 1;67(8):3501-13.

Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP. Gene expression during the life cycle of *Drosophila melanogaster*. Science. 2002 Sep 27;297(5590):2270-5.

Arcia E, CONNERS KC. Gender Differences in ADHD?. Journal of Developmental & Behavioral Pediatrics. 1998 Apr 1;19(2):77-83.

Arbuthnott D, Dutton EM, Agrawal AF, Rundle HD. The ecology of sexual conflict: ecologically dependent parallel evolution of male harm and female resistance in *Drosophila melanogaster*. Ecology letters. 2014 Feb 1;17(2):221-8.

Arnold LE. Sex differences in ADHD: Conference summary. Journal of abnormal child psychology. 1996 Oct 1;24(5):555-69.

Assis R, Zhou Q, Bachtrog D. Sex-biased transcriptome evolution in *Drosophila*. Genome biology and evolution. 2012 Nov 1;4(11):1189-200.

Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ, FlyBase Consortium. FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*. Nucleic acids research. 2015 Oct 13:gkv1046.

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF. Systems genetics of complex traits in *Drosophila melanogaster*. Nature genetics. 2009 Mar 1;41(3):299-307.

Balashova VA, Abdulkadyrov KM. Cellular composition of hemopoietic tissue of the liver and spleen in the human fetus. Arkhiv anatomii, gistologii i embriologii. 1984 Apr;86(4):80-3.

Barker, J. S. F. in Adaptation and Fitness in Animal Populations (eds van der Werf, J., Graser, H.-U., Frankham, R. and Gondro, C.) 3-14 (Springer, Heidelberg, 2009).

Barnes AI, Wigby S, Boone JM, Partridge L, Chapman T. Feeding, fecundity and lifespan in female *Drosophila melanogaster*. Proceedings of the Royal Society of London B: Biological Sciences. 2008 Jul 22;275(1643):1675-83.

Beery AK, Zucker I. Sex bias in neuroscience and biomedical research. Neuroscience & Biobehavioral Reviews. 2011 Jan 31;35(3):565-72.

Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho TJ, Koutseva N, Zaghlul S, Graves T, Rock S, Kremitzki C. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature. 2014 Apr 24;508(7497):494.

Bernstein IS. Sex differences in the behavior of nonhuman primates. Social Science & Medicine. Part B: Medical Anthropology. 1978 Jan 1;12:151-4.

Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. Sex-specific and lineage-specific alternative splicing in primates. Genome research. 2010 Feb 1;20(2):180-9.

Boakes EH, Wang J, Amos W. An investigation of inbreeding depression and purging in captive pedigreed populations. Heredity. 2007 Mar 1;98(3):172-82.

Boekholdt SM, Keller TT, Wareham NJ, Luben R, Bingham SA, Day NE, Sandhu MS, Jukema JW, Kastelein JJ, Hack CE, Khaw KT. Serum levels of type II secretory phospholipase A2 and the risk of future coronary artery disease in apparently healthy men and women The EPIC-Norfolk Prospective Population Study. Arteriosclerosis, thrombosis, and vascular biology. 2005 Apr 1;25(4):839-46.

Branson K, Robie AA, Bender J, Perona P, Dickinson MH. High-throughput ethomics in large groups of *Drosophila*. Nature methods. 2009 Jun 1;6(6):451-7.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW. The evolution of gene expression levels in mammalian organs. Nature. 2011 Oct 20;478(7369):343-8.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Apr 1:btu170.

Bur IM, Cohen-Solal AM, Carmignac D, Abecassis PY, Chauvet N, Martin AO, Van Der Horst GT, Robinson IC, Maurel P, Mollard P, Bonnefont X. The circadian clock components CRY1 and CRY2 are necessary to sustain sex dimorphism in mouse liver metabolism. Journal of Biological Chemistry. 2009 Apr 3;284(14):9066-73.

Carbone A, Zinovyev A, Képès F. Codon adaptation index as a measure of dominating codon bias. Bioinformatics. 2003 Nov 1;19(16):2005-15.

Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM, Micklem G. Unlocking the secrets of the genome. Nature. 2009 Jun 18;459(7249):927-30.

Chintapalli VR, Wang J, Dow JA. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nature genetics. 2007 Jun 1;39(6):715-20.

Chapman T. Seminal fluid-mediated fitness traits in *Drosophila*. Heredity. 2001 Nov 10;87(5):511-21.

Chapman T, Liddle LF, Kalb JM, Wolfner MF, Partridge L. Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. Nature. 1995 Jan 19;373(6511):241-4.

Chapman T, Partridge L. Female fitness in *Drosophila melanogaster*: an interaction between the effect of nutrition and of encounter rate with males. Proceedings of the Royal Society of London B: Biological Sciences. 1996 Jun 22;263(1371):755-9.

Charchar FJ, Tomaszewski M, Strahorn P, Champagne B, Dominiczak AF. Y is there a risk to being male?. Trends in Endocrinology & Metabolism. 2003 Jun 30;14(4):163-8.

Charchar FJ, Tomaszewski M, Padmanabhan S, Lacka B, Upton MN, Inglis GC, Anderson NH, McConnachie A, Zukowska-Szczechowska E, Grzeszczak W, Connell JM. The Y chromosome effect on blood pressure in two European populations. Hypertension. 2002 Feb 1;39(2):353-6.

Charlesworth B. Causes of natural variation in fitness: evidence from studies of *Drosophila* populations. Proceedings of the National Academy of Sciences. 2015 Feb 10;112(6):1662-9.

Charlesworth B, Hughes KA (2000). The maintenance of genetic variation in life-history traits. Evolutionary Genetics from Molecules to Morphology, eds Singh RS, Krimbas CB (Cambridge Univ Press, Cambridge, UK), pp 369–392

Charif D, Lobry JR. SeqinR: a contributed package to the R project for statistical computing. Structural approaches to sequence evolution: molecules, networks and populations. U. Bastolla, M. Porto, HE Roman and M. Vendruscolo.

Charnov EL, Krebs JR. On clutch-size and fitness. Ibis. 1974 Apr 1;116(2):217-9.

Chi PB, Liberles DA. Selection on protein structure, interaction, and sequence. Protein Science. 2016 Feb 1.

Choi BG, McLaughlin MA. Why men's hearts break: cardiovascular effects of sex steroids. Endocrinology and metabolism clinics of North America. 2007 Jun 30;36(2):365-77.

Civetta A, Singh RS. High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and Drosophila virilis group species. Journal of Molecular Evolution. 1995 Dec 1;41(6):1085-95.

Clayton JA, Collins FS. NIH to balance sex in cell and animal studies. Nature. 2014 May 15;509(7500):282-3.

Comeron JM, Kreitman M, Aguadé M. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics. 1999 Jan 1;151(1):239-49.

Connallon T, Knowles LL. Intergenomic conflict revealed by patterns of sex-biased gene expression. Trends in Genetics. 2005 Sep 30;21(9):495-9.

Cordts R, Partridge L. Courtship reduces longevity of male *Drosophila melanogaster*. Animal Behaviour. 1996 Aug 31;52(2):269-78.

Darwin C. 1871. The descent of man, and selection in relation to sex. London: John Murray.

Darwin, C. (1859). On the origin of the species by natural selection.

Das J, Yu H. HINT: High-quality protein interactomes and their applications in understanding human disease. BMC systems biology. 2012 Jul 30;6(1):92.

Demetrius L, Ziehe M. Darwinian fitness. Theoretical population biology. 2007 Nov 30;72(3):323-45.

de Souza N. Model organisms: Mouse models challenged. Nature methods. 2013 Apr 1;10(4):288-.

Drummond DA, Raval A, Wilke CO. A single determinant dominates the rate of yeast protein evolution. Molecular biology and evolution. 2006 Feb 1;23(2):327-37.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. Proceedings of the National Academy of Sciences of the United States of America. 2005 Oct 4;102(40):14338-43.

Duret L, Mouchiroud D. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Molecular biology and evolution. 2000 Jan 1;17(1):68-070.

Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis, Drosophila*, and *Arabidopsis*. Proceedings of the National Academy of Sciences. 1999 Apr 13;96(8):4482-7.

Dye RV, Miller KJ, Singer EJ, Levine AJ. Hormone replacement therapy and risk for neurodegenerative diseases. International journal of Alzheimer's Disease. 2012 Apr 4;2012.

Eastwood L, Burnet B. Courtship latency in male *Drosophila melanogaster*. Behavior genetics. 1977 Sep 1;7(5):359-72.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research. 2004 Mar 1;32(5):1792-7.

Edward DA, Fricke C, Gerrard DT, Chapman T. Quantifying the life-history response to increased male exposure in female *Drosophila melanogaster*. Evolution. 2011 Feb 1;65(2):564-73.

Eggers S, Sinclair A. Mammalian sex determination—insights from humans and mice. Chromosome research. 2012 Jan 1;20(1):215-38.

Ejima A, Griffith LC. Measurement of courtship behavior in *Drosophila melanogaster*. Cold Spring Harbor Protocols. 2007 Oct 1;2007(10):pdb-rot4847.

Ellegren H, Parsch J. The evolution of sex-biased genes and sex-biased gene expression. Nature Reviews Genetics. 2007 Sep 1;8(9):689-98.

Ellegren H, Sheldon BC. Genetic basis of fitness differences in natural populations. Nature. 2008 Mar 13;452(7184):169-75.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L. Ensembl 2014 Nucleic Acids Res. 2014, 42. D749–55

Fowler K, Semple C, Barton NH, Partridge L. Genetic variation for total fitness in *Drosophila melanogaster*. Proceedings of the Royal Society of London B: Biological Sciences. 1997 Feb 22;264(1379):191-9.

Fowler K, Partridge L. A cost of mating in female fruitflies. Nature. 1989 Apr 27;338(6218):760-1.

Fritzsche P, Neumann K, Nasdal K, Gattermann R. Differences in reproductive success between laboratory and wild-derived golden hamsters (Mesocricetus auratus) as a consequence of inbreeding. Behavioral Ecology and Sociobiology. 2006 Jun 1;60(2):220-6.

Gadgil M, Bossert WH. Life historical consequences of natural selection. American Naturalist. 1970 Jan 1:1-24.

Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, Chen X. Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. Cell research. 2010 Jul 1;20(7):763-83.

Gaub M, Carlson CL. Gender differences in ADHD: a meta-analysis and critical review. Journal of the American Academy of Child & Adolescent Psychiatry. 1997 Aug 31;36(8):1036-45.

Gelbart WM, Emmert DB. FlyBase high throughput expression pattern data beta version. A Database of Drosophila Genes &Genomes. 2010.

Ghoussaini M, Edwards SL, Michailidou K, Nord S, Cowper-Sal R, Desai K, Kar S, Hillman KM, Kaufmann S, Glubb DM, Beesley J. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. Nature communications. 2014 Sep 23;4.

Gibson RM, Bachman GC. The costs of female choice in a lekking bird. Behavioral Ecology. 1992 Dec 21;3(4):300-9.

Gibson G, Riley-Berger R, Harshman L, Kopp A, Vacha S, Nuzhdin S, Wayne M. Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. Genetics. 2004 Aug 1;167(4):1791-9.

Gilpin ME. Minimum viable populations: processes of species extinction. Conservation biology: the science of scarcity and diversity. 1986:19-34.

Gnad F, Parsch J. Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression. Bioinformatics. 2006 Oct 15;22(20):2577-9.

Gordon T, Kannel WB, Hjortland MC, McNamara PM. Menopause and coronary heart disease: the Framingham Study. Annals of Internal Medicine. 1978 Aug 1;89(2):157-61.

Gur RC, Turetsky BI, Matsui M, Yan M, Bilker W, Hughett P, Gur RE. Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. The Journal of neuroscience. 1999 May 15;19(10):4065-72.

Gur RC, Richard J, Calkins ME, Chiavacci R, Hansen JA, Bilker WB, Loughead J, Connolly JJ, Qiu H, Mentch FD, Abou-Sleiman PM. Age group and sex differences in performance on a computerized neurocognitive battery in children age 8− 21. Neuropsychology. 2012 Mar;26(2):251.

Gustafsson JA, Mode A, Norstedt G, Skett P. Sex steroid induced changes in hepatic enzymes. Annual Review of Physiology. 1983 Mar;45(1):51-60.

Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Molecular biology and evolution. 2005 Apr 1;22(4):803-6.

Hansen ME, Kulathinal RJ. Sex-biased networks and nodes of sexually antagonistic conflict in *Drosophila*. International journal of evolutionary biology. 2013 Jan 22;2013.

Haymer DS, Hartl DL. The experimental assessment of fitness in *Drosophila*. I. Comparative measures of competitive reproductive success. Genetics. 1982 Nov 1;102(3):455-66.

Houle D, Kondrashov AS. Coevolution of costly mate choice and condition-dependent display of good genes. Proceedings of the Royal Society of London B: Biological Sciences. 2002 Jan 7;269(1486):97-104.

Hunt J, Brooks R, Jennions MD, Smith MJ, Bentsen CL, Bussiere LF. High-quality male field crickets invest heavily in sexual display but die young. Nature. 2004 Dec 23;432(7020):1024-7.

Huylmans AK, Parsch J. Population-and sex-biased gene expression in the excretion organs of *Drosophila melanogaster*. G3: Genes| Genomes| Genetics. 2014 Dec 1;4(12):2307-15.

Ingvarsson PK. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. Molecular biology and evolution. 2007 Mar 1;24(3):836-44.

Innocenti P, Morrow EH. The sexually antagonistic genes of *Drosophila melanogaster*. PLoS Biol. 2010 Mar 16;8(3):e1000335.

Isensee J, Witt H, Pregla R, Hetzer R, Regitz-Zagrosek V, Noppinger PR. Sexually dimorphic gene expression in the heart of mice and men. Journal of molecular medicine. 2008 Jan 1;86(1):61-74.

Jansen R, Batista S, Brooks AI, Tischfield JA, Willemsen G, Van Grootheest G, Hottenga JJ, Milaneschi Y, Mbarek H, Madar V, Peyrot W. Sex differences in the human peripheral blood transcriptome. BMC genomics. 2014 Jan 17;15(1):1.

Jennings BH. Drosophila–a versatile model in biology & medicine. Materials Today. 2011 May 31;14(5):190-5.

Jeon HJ, Kim YS, Park JS, Pyo JH, Na HJ, Kim IJ, Kim CM, Chung HY, Kim ND, Arking R, Yoo MA. Age-related change in γH2AX of *Drosophila* muscle: its significance as a marker for muscle damage and longevity. Biogerontology. 2015 Aug 1;16(4):503-16.

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. Nature genetics. 2001 Dec 1;29(4):389-95.

Johnson JA, Dunn PO. Low genetic variation in the heath hen prior to extinction and implications for the conservation of prairie-chicken populations. Conservation Genetics. 2006 Feb 1;7(1):37-48.

Kannel WB, Feinleib M. Natural history of angina pectoris in the Framingham study: prognosis and survival. The American journal of cardiology. 1972 Feb 29;29(2):154-63.

Khazaeli AA, Curtsinger JW. Pleiotropy and life history evolution in Drosophila melanogaster: uncoupling life span and early fecundity. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2013 May 1;68(5):546-53.

Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. Nature genetics. 2004 Jun 1;36(6):642-6.

Kirkpatrick M, Ryan MJ. The evolution of mating preferences and the paradox of the lek. Nature. 1991 Mar 7;350(6313):33-8.

Kiyohara C, Ohno Y. Sex differences in lung cancer susceptibility: a review. Gender medicine. 2010 Oct 31;7(5):381-401.

Klein W. Gender differences in clinical trials in coronary heart disease: response to drug therapy. European heart journal. 1996 Dec 1;17(12):1786-90.

Kristensen TN, Overgaard J, Loeschcke V, Mayntz D. Dietary protein content affects evolution for body size, body fat and viability in *Drosophila melanogaster*. Biology letters. 2010 Oct 27:rsbl20100872.

Kristensen TN, Overgaard J, Lassen J, Hoffmann AA, Sgrò C. Low evolutionary potential for egg-to-adult viability in *Drosophila melanogaster* at high temperatures. Evolution. 2015 Mar 1;69(3):803-14.

Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. Evolution of protein-coding genes in *Drosophila*. Trends in Genetics. 2008 Mar 31;24(3):114-23.

Lazzaro BP, Flores HA, Lorigan JG, Yourth CP. Genotype-by-environment interactions and adaptation to local temperature affect immunity and fecundity in *Drosophila melanogaster*. PLoS Pathog. 2008 Mar 14;4(3):e1000025.

Leinonen R, Sugawara H, Shumway M. The sequence read archive. Nucleic acids research. 2010 Nov 9:gkq1019.

Leips J, Gilligan P, Mackay TF. Quantitative trait loci with age-specific effects on fecundity in *Drosophila melanogaster*. Genetics. 2006 Mar 1;172(3):1595-605.

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Molecular biology and evolution. 2005 May 1;22(5):1345-54.

Lerner DJ, Kannel WB. Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. American heart journal. 1986 Feb 28;111(2):383-90.

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754-60.

Li WH, Gojobori T. Rapid evolution of goat and sheep globin genes following gene duplication. Molecular biology and evolution. 1983 Dec 1;1(1):94-108.

Li R, Singh M. Sex differences in cognitive impairment and Alzheimer's disease. Frontiers in neuroendocrinology. 2014 Aug 31;35(3):385-403.

Liao BY, Scott NM, Zhang J. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. Molecular Biology and Evolution. 2006 Nov 1;23(11):2072-80.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014 Dec 5;15(12):1.

Lüpold S, Manier MK, Ala-Honkola O, Belote JM, Pitnick S. Male *Drosophila* melanogaster adjust ejaculate size based on female mating status, fecundity, and age. Behavioral Ecology. 2010 Dec 6:arq193.

Luporini P, Vallesi A, Miceli C, Bradshaw RA. Chemical signaling in ciliates. Journal of Eukaryotic Microbiology. 1995 May 1;42(3):208-12.

Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF. The *Drosophila melanogaster* genetic reference panel. Nature. 2012 Feb 9;482(7384):173-8.

Mair W, Sgrò CM, Johnson AP, Chapman T, Partridge L. Lifespan extension by dietary restriction in female Drosophila melanogaster is not caused by a reduction in vitellogenesis or ovarian activity. Experimental gerontology. 2004 Jul 31;39(7):1011-9.

Mair W, Piper MD, Partridge L. Calories do not explain extension of life span by dietary restriction in *Drosophila*. PLoS Biol. 2005 May 31;3(7):e223.

Makałowski W, Boguski MS. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. Proceedings of the National Academy of Sciences. 1998 Aug 4;95(16):9407-12.

Mallat Z, Benessiano J, Simon T, Ederhy S, Sebella-Arguelles C, Cohen A, Huart V, Wareham NJ, Luben R, Khaw KT, Tedgui A. Circulating Secretory Phospholipase A2 Activity and Risk of Incident Coronary Events in Healthy Men and Women The EPIC-NORFOLK Study. Arteriosclerosis, thrombosis, and vascular biology. 2007 May 1;27(5):1177-83.

Manwani B, Bentivegna K, Benashski SE, Venna VR, Xu Y, Arnold AP, McCullough LD. Sex differences in ischemic stroke sensitivity are influenced by gonadal hormones, not by sex chromosome complement. Journal of Cerebral Blood Flow & Metabolism. 2015 Feb 1;35(2):221-9.

Marais G, Domazet-Lošo T, Tautz D, Charlesworth B. Correlated evolution of synonymous and nonsynonymous sites in Drosophila. Journal of molecular evolution. 2004 Dec 1;59(6):771-9.

Maretty L, Sibbesen JA, Krogh A. Bayesian transcriptome assembly. Genome biology. 2014 Oct 31;15(10):1.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome research. 2008 Sep 1;18(9):1509-17.

Markova-Raina P, Petrov D. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. Genome research. 2011 Jun 1;21(6):863-74.

Markow TA, Ankney PF. *Drosophila* males contribute to oogenesis in a multiple mating species. Science. 1984 Apr 20;224(4646):302-3.

Mayfield JA, Fiebig A, Johnstone SE, Preuss D. Gene families from the Arabidopsis thaliana pollen coat proteome. Science. 2001 Jun 29;292(5526):2482-5.

Maynard Smith J. Sexual selection: a classification of models. Sexual selection: Testing the alternatives. 1987:9-20.

McInerney JO. The causes of protein evolutionary rate variation. Trends in ecology & evolution. 2006 May 31;21(5):230-2.

McIntyre LM, Bono LM, Genissel A, Westerman R, Junk D, Telonis-Scott M, Harshman L, Wayne ML, Kopp A, Nuzhdin SV. Sex-specific expression of alternative transcripts in *Drosophila*. Genome biology. 2006 Aug 25;7(8):1.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. Rapid evolution of male-biased gene expression in *Drosophila*. Proceedings of the National Academy of Sciences. 2003 Aug 19;100(17):9894-9.

Meisel RP, Malone JH, Clark AG. Faster-X evolution of gene expression in *Drosophila*. PLoS Genet. 2012 Oct 11;8(10):e1003013.

Meisel RP. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. Molecular Biology and Evolution. 2011 Jun 1;28(6):1893-900.

Michael D, Soi S, Cabera-Perez J, Weller M, Alexander S, Alevizos I, Illei GG, Chiorini JA. Microarray analysis of sexually dimorphic gene expression in human minor salivary glands. Oral diseases. 2011 Oct 1;17(7):653-61.

Mogil JS, Chanda ML. The case for the inclusion of female subjects in basic science studies of pain. Pain. 2005 Sep 1;117(1-2):1-5.

Ngo ST, Steyn FJ, McCombe PA. Gender differences in autoimmune disease. Frontiers in neuroendocrinology. 2014 Aug 31;35(3):347-69.

Nishizawa S, Benkelfat C, Young SN, Leyton M, Mzengeza SD, De Montigny C, Blier P, Diksic M. Differences between males and females in rates of serotonin synthesis in human brain. Proceedings of the National Academy of Sciences. 1997 May 13;94(10):5308-13.

Ober C, Loisel DA, Gilad Y. Sex-specific genetic architecture of human disease. Nature Reviews Genetics. 2008 Dec 1;9(12):911-22.

Orr HA. Fitness and its role in evolutionary genetics. Nature Reviews Genetics. 2009 Aug 1;10(8):531-9.

Pál C, Papp B, Hurst LD. Highly expressed genes in yeast evolve slowly. Genetics. 2001 Jun 1;158(2):927-31.

Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. Science. 2003 Jan 31;299(5607):697-700.

Parker GA. Sexual selection and sexual conflict. Sexual selection and reproductive competition in insects. 1979:123-66.

Parker GA. Sexual conflict over mating and fertilization: an overview. Philosophical Transactions of the Royal Society B: Biological Sciences. 2006 Feb 28;361(1466):235-59.

Partridge L, Prowse N, Pignatelli P. Another set of responses and correlated responses to selection on age at reproduction in *Drosophila melanogaster*. Proceedings of the Royal Society of London B: Biological Sciences. 1999 Feb 7;266(1416):255-61.

Partridge L, Prowse N, Pignatelli P. Another set of responses and correlated responses to selection on age at reproduction in *Drosophila melanogaster*. Proceedings of the Royal Society of London B: Biological Sciences. 1999 Feb 7;266(1416):255-61.

Partridge L, Green A, Fowler K. Effects of egg-production and of exposure to males on female survival in *Drosophila melanogaster*. Journal of Insect Physiology. 1987 Jan 1;33(10):745-9.

Partridge L, Farquhar M. Sexual activity reduces lifespan of male fruitflies. Nature. 1981 Dec 10;294(58):F582.

Partridge L, Fowler K. Direct and correlated responses to selection on age at reproduction in *Drosophila melanogaster*. Evolution. 1992 Feb 1:76-91.

Partridge L, Fowler K. Responses and correlated responses to artificial selection on thorax length in *Drosophila melanogaster*. Evolution. 1993 Feb 1:213-26.

Partridge L, Piper MD, Mair W. Dietary restriction in *Drosophila*. Mechanisms of ageing and development. 2005 Sep 30;126(9):938-50.

Partridge L, Gems D, Withers DJ. Sex and death: what is the connection?. Cell. 2005 Feb 25;120(4):461-72.

Patrick CH, Palesch YY, Feinleib M, Brody JA. Sex differences in declining cohort death rates from heart disease. American Journal of Public Health. 1982 Feb;72(2):161-6.

Perry JC, Harrison PW, Mank JE. The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. Molecular biology and evolution. 2014 May 1;31(5):1206-19.

Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, Pritchard JK. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. Genome research. 2012 Apr 1;22(4):602-10.

Piper MD, Partridge L. Dietary restriction in Drosophila: delayed aging or experimental artefact?. PLoS Genet. 2007 Apr 27;3(4):e57.

Pitnick S. Male size influences mate fecundity and remating interval in *Drosophila melanogaster*. Animal Behaviour. 1991 May 31;41(5):735-45.

Plavcan JM. Sexual size dimorphism, canine dimorphism, and male-male competition in primates. Human Nature. 2012 Mar 1;23(1):45-67.

Plutynski A. What was Fisher's fundamental theorem of natural selection and what was it for?. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences. 2006 Mar 31;37(1):59-82.

Pomiankowski A. The costs of choice in sexual selection. Journal of theoretical Biology. 1987 Sep 21;128(2):195-218.

Popescu CE, Borza T, Bielawski JP, Lee RW. Evolutionary rates and expression level in Chlamydomonas. Genetics. 2006 Mar 1;172(3):1567-76.

Price T, Schluter D, Heckman NE. Sexual selection when the female directly benefits. Biological Journal of the Linnean Society. 1993 Mar 1;48(3):187-211.

Prout T. The relation between fitness components and population prediction in *Drosophila*. I: the estimation of fitness components. Genetics. 1971 May;68(1):127.

Roy-O'Reilly M, McCullough LD. Sex differences in stroke: the contribution of coagulation. Experimental neurology. 2014 Sep 30;259:16-27.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science. 2003 Jun 13;300(5626):1742-5.

Reinius B, Saetre P, Leonard JA, Blekhman R, Merino-Martinez R, Gilad Y, Jazin E. An evolutionarily conserved sexual signature in the primate brain. PLoS Genet. 2008 Jun 20;4(6):e1000100.

Reinke V, Smith HE, Nance J, Wang J, Van Doren C, Begley R, Jones SJ, Davis EB, Scherer S, Ward S, Kim SK. A global profile of germline gene expression in C. elegans. Molecular cell. 2000 Sep 30;6(3):605-16.

Rice WR. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature. 1996 May 16;381(6579):232-4.

Rigby N, Kulathinal RJ. Genetic Architecture of Sexual Dimorphism in Humans. Journal of cellular physiology. 2015 Oct 1;230(10):2304-10.

Rinn JL, Rozowsky JS, Laurenzi IJ, Petersen PH, Zou K, Zhong W, Gerstein M, Snyder M. Major molecular differences between mammalian sexes are involved in drug metabolism and renal function. Developmental cell. 2004 Jun 30;6(6):791-800.

Rinn JL, Snyder M. Sexual dimorphism in mammalian gene expression. Trends in Genetics. 2005 May 31;21(5):298-305.

Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. Nature methods. 2013 Jan 1;10(1):71-3.

Roff DA. Defining fitness in evolutionary models. Journal of genetics. 2008 Dec 1;87(4):339-48.

Roff DA. The evolution of life histories: data and analysis. London: Chapman &. 1992.

Rose MR. Laboratory evolution of postponed senescence in *Drosophila melanogaster*. Evolution. 1984 Sep 1:1004-10.

Roth SM, Ferrell RE, Peters DG, Metter EJ, Hurley BF, Rogers MA. Influence of age, sex, and strength training on human muscle gene expression determined by microarray. Physiological genomics. 2002 Sep 3;10(3):181-90.

Roy AK, Chatterjee B. Sexual dimorphism in the liver. Annual review of physiology. 1983 Mar;45(1):37-50.

Roy-O'Reilly M, McCullough LD. Sex differences in stroke: the contribution of coagulation. Experimental neurology. 2014 Sep 30;259:16-27.

Ruigrok AN, Salimi-Khorshidi G, Lai MC, Baron-Cohen S, Lombardo MV, Tait RJ, Suckling J. A meta-analysis of sex differences in human brain structure. Neuroscience & Biobehavioral Reviews. 2014 Feb 28;39:34-50.

Service PM, Rose MR. Genetic covariation among life-history components: the effect of novel environments. Evolution. 1985 Jul 1;39(4):943-5.

Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic acids research. 1987 Feb 11;15(3):1281-95.

Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. Journal of molecular evolution. 1986 Dec 1;24(1-2):28-38.

Shuvalova YA, Khasanova ZB, Kaminnaya VI, Samoilova EV, Korotaeva AA, Rubanovich AV, Kaminnyi AI. The association of PLA2G2A single nucleotide polymorphisms with type IIa secretory phospholipase A2 level but not its activity in patients with stable coronary heart disease. Gene. 2015 Jun 10;564(1):29-34.

Si H, Banga RS, Kapitsinou P, Ramaiah M, Lawrence J, Kambhampati G, Gruenwald A, Bottinger E, Glicklich D, Tellis V, Greenstein S. Human and murine kidneys show gender-and species-specific gene expression differences in response to injury. PLoS One. 2009 Mar 11;4(3):e4802.

Singh RS, Kulathinal RJ. Sex gene pool evolution and speciation. A new paradigm. Genes & genetic systems. 2000;75(3):119-30.

Stanley CE, Kulathinal RJ. flyDIVaS: A Comparative Genomics Resource for *Drosophila* Divergence and Selection. G3: Genes| Genomes| Genetics. 2016 May 25:g3-116.

Sun YV, Kardia SL. Identification of epistatic effects using a protein–protein interaction database. Human molecular genetics. 2010 Nov 15;19(22):4345-52.

Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. Nature Reviews Genetics. 2002 Feb 1;3(2):137-44.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proceedings of the National Academy of Sciences. 2001 Jun 19;98(13):7375-9.

Tabassum R, Nath A, Preininger M, Gibson G. Geographical, environmental and pathophysiological influences on the human blood transcriptome. Current genetic medicine reports. 2013 Dec 1;1(4):203-11.

Tatar M. Diet restriction in *Drosophila melanogaster*. InMechanisms of dietary restriction in aging and disease 2006 Oct 20 (Vol. 35, pp. 115-136). Karger Publishers.

Telonis-Scott M, Kopp A, Wayne ML, Nuzhdin SV, McIntyre LM. Sex-specific splicing in *Drosophila*: widespread occurrence, tissue specificity and evolutionary conservation. Genetics. 2009 Feb 1;181(2):421-34.

Thoemke K, Yi W, Ross JM, Kim S, Reinke V, Zarkower D. Genome-wide analysis of sex-enriched gene expression during C. elegans larval development. Developmental biology. 2005 Aug 15;284(2):500-8.

Townsend JP, Hartl DL. Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. Genome biology. 2002 Nov 20;3(12):1.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature biotechnology. 2013 Jan 1;31(1):46-53.

Tullis KM, Krebs CJ, Leung JY, Robins DM. The regulator of sex-limitation gene, rsl, enforces male-specific liver gene expression by negative regulation. Endocrinology. 2003 May 1;144(5):1854-60.

Vacquier VD. Evolution of gamete recognition proteins. Science. 1998 Sep 25;281(5385):1995-8.

Van de Peer Y, Taylor JS, Braasch I, Meyer A. The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. Journal of molecular evolution. 2001 Oct 1;53(4-5):436-46.

Vigne P, Frelin C. Diet dependent longevity and hypoxic tolerance of adult *Drosophila* melanogaster. Mechanisms of ageing and development. 2007 Jun 30;128(5):401-6.

Vigne P, Frelin C. Plasticity of the responses to chronic hypoxia and dietary restriction in an aged organism: evidence from the *Drosophila* model. Experimental gerontology. 2007 Dec 31;42(12):1162-6.

Wakimoto BT, Lindsley DL, Herrera C. Toward a comprehensive genetic analysis of male fertility in *Drosophila melanogaster*. Genetics. 2004 May 1;167(1):207-16.

Weber CC, Hurst LD. Support for multiple classes of local expression clusters in *Drosophila melanogaster*, but no evidence for gene order conservation. Genome biology. 2011 Mar 17;12(3):1.

Welle S, Tawil R, Thornton CA. Sex-related differences in gene expression in human skeletal muscle. PloS one. 2008 Jan 2;3(1):e1385.

Westemeier RL, Brawn JD, Simpson SA, Esker TL, Jansen RW, Walk JW, Kershner EL, Bouzat JL, Paige KN. Tracking the long-term decline and recovery of an isolated population. Science. 1998 Nov 27;282(5394):1695-8.

Wigby S, Chapman T. Sex peptide causes mating costs in female *Drosophila* melanogaster. Current Biology. 2005 Feb 22;15(4):316-21.

Williams GC. Natural selection, the costs of reproduction, and a refinement of Lack's principle. The American Naturalist. 1966 Nov 1;100(916):687-90.

Wizemann TM, Pardue ML, editors. Exploring the biological contributions to human health: does sex matter?. National Academies Press; 2001 Aug 2.

Wolfner MF. Tokens of love: functions and regulation of *Drosophila* male accessory gland products. Insect biochemistry and molecular biology. 1997 Mar 31;27(3):179-92.

Wong A, Wolfner MF. Evolution of *Drosophila* seminal proteins and their networks. Rapidly evolving genes and genetics systems. Oxford University Press, Oxford. 2012 Jun 28:144-52.

Wright SI, Yau CK, Looseley M, Meyers BC. Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. Molecular biology and evolution. 2004 Sep 1;21(9):1719-26.

Wyman MJ, Agrawal AF, Rowe L. Condition-dependence of the sexually dimorphic transcriptome in *Drosophila melanogaster*. Evolution. 2010 Jun 1;64(6):1836-48.

Xu H, Wang F, Liu Y, Yu Y, Gelernter J, Zhang H. Sex-biased methylome and transcriptome in human prefrontal cortex. Human molecular genetics. 2014 Mar 1;23(5):1260-70.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, Lancet D. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005 Mar 1;21(5):650-9.

Yang X, Schadt EE, Wang S, Wang H, Arnold AP, Ingram-Drake L, Drake TA, Lusis AJ. Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome research. 2006 Aug 1;16(8):995-1004.

Zhang Y, Laz EV, Waxman DJ. Dynamic, sex-differential STAT5 and BCL6 binding to sex-biased, growth hormone-regulated genes in adult mouse liver. Molecular and cellular biology. 2012 Feb 15;32(4):880-96.

Zhang Y, Klein K, Sugathan A, Nassery N, Dombkowski A, Zanger UM, Waxman DJ. Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease. PLoS One. 2011 Aug 12;6(8):e23506.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. Constraint and turnover in sex-biased gene expression in the genus Drosophila. Nature. 2007 Nov 8;450(7167):233-7.

Zhang Z, Parsch J. Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. Molecular Biology and Evolution. 2005 Oct 1;22(10):1945-7.

Zhang J. Evolution by gene duplication: an update. Trends in ecology & evolution. 2003 Jun 30;18(6):292-8.

Zhang J, Rosenberg HF, Nei M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proceedings of the National Academy of Sciences. 1998 Mar 31;95(7):3708-13.

Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic acids research. 2014 Jun 17;42(11):e91-.

Zwaan B, Bijlsma R, Hoekstra RF. Direct selection on life span in *Drosophila* melanogaster. Evolution. 1995 Aug 1:649-59.

APPENDIX A

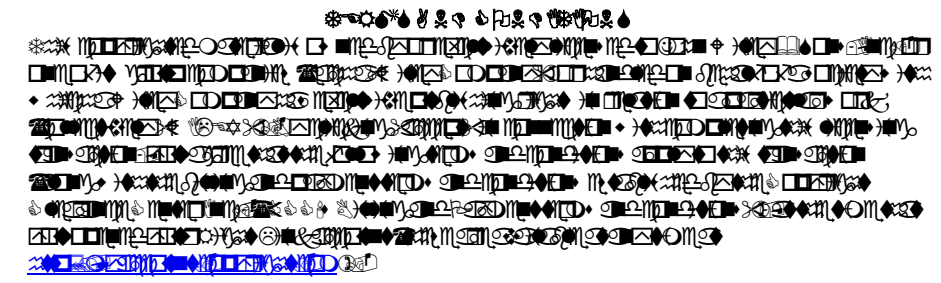
PERMISSION FOR THE INCLUSION OF COPYRIGHTED MATERIAL

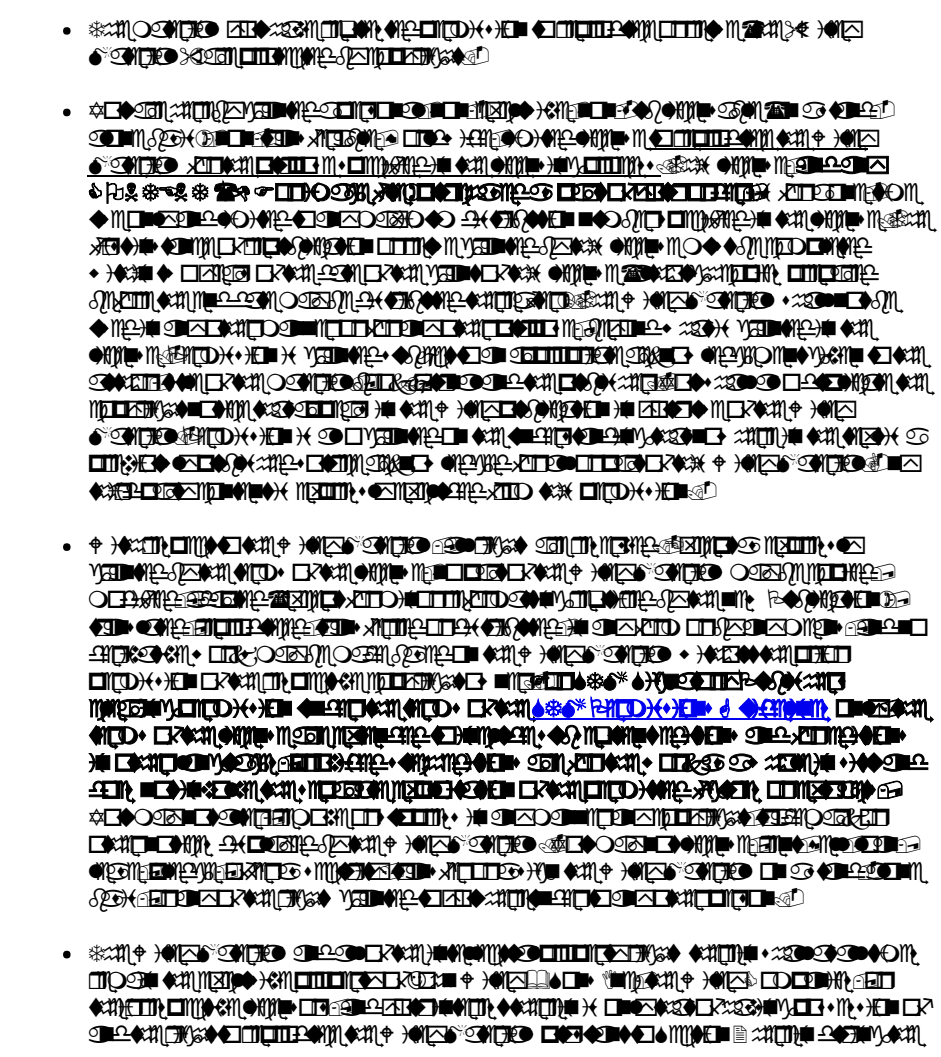
JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

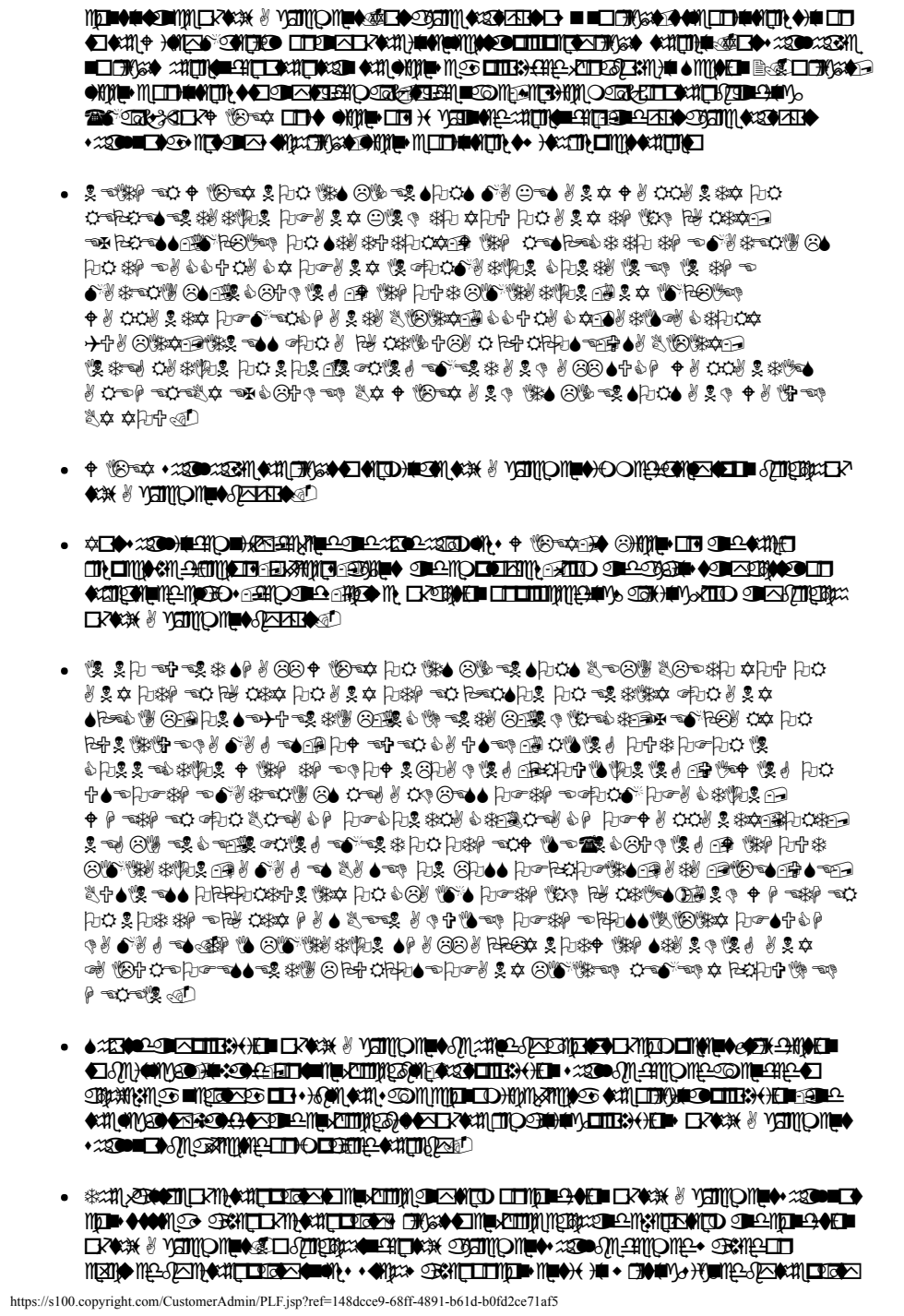
Jun 13, 2016

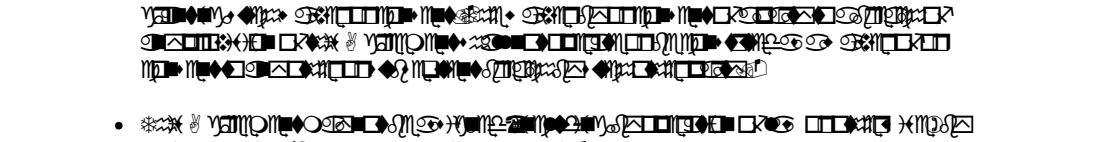
This Agreement between Nichole Rigby ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

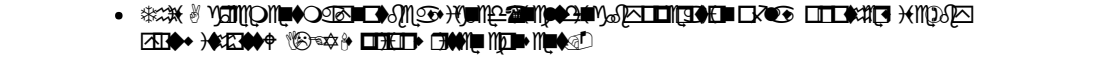
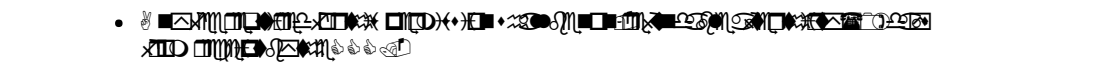
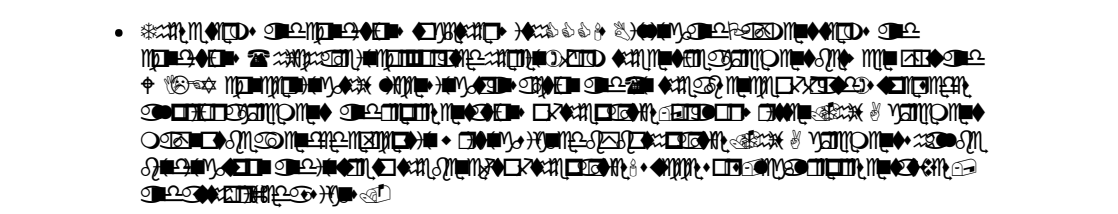
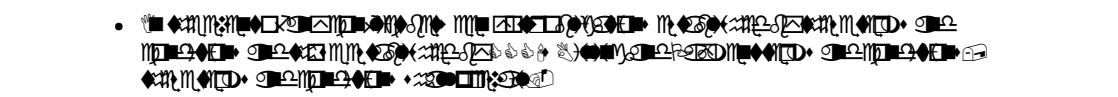
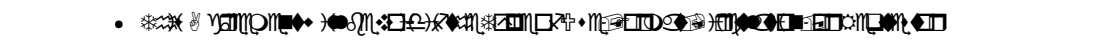
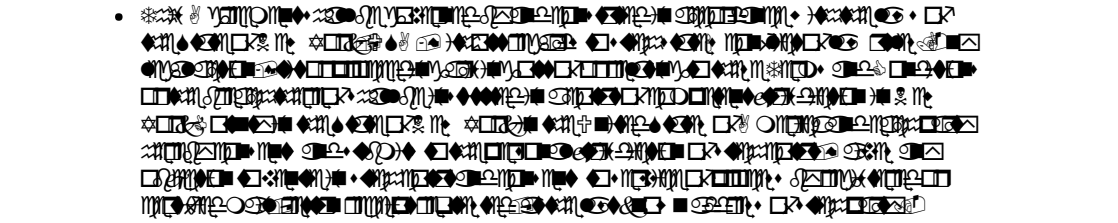
License Number	3886621482325
License date	Jun 12, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Journal of Cellular Physiology
Licensed Content Title	Genetic Architecture of Sexual Dimorphism in Humans
Licensed Content Author	Nichole Rigby,Rob J. Kulathinal
Licensed Content Date	Jun 23, 2015
Licensed Content Pages	7
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Genetic Basis of Sexual Dimorphism
Expected completion date	Jul 2016
Expected size (number of pages)	100
Requestor Location	Nichole Rigby [redacted] PHILADELPHIA, PA 19128 United States Attn: Nichole Rigby
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Nichole Rigby [redacted] PHILADELPHIA, PA 19128 United States Attn: Nichole Rigby
Total	0.00 USD
Terms and Conditions	







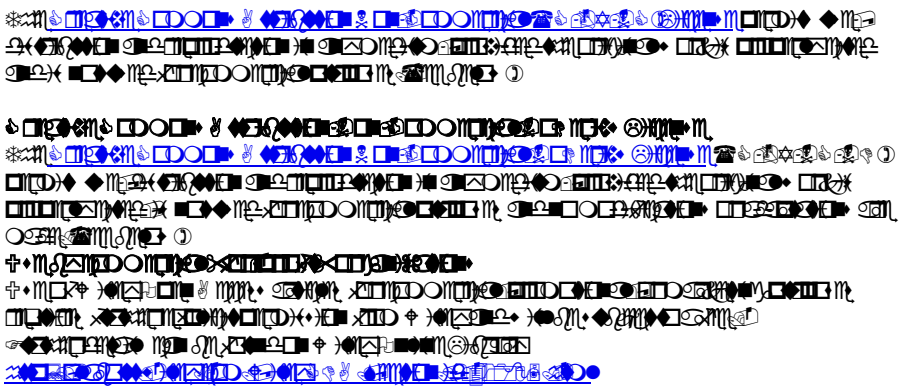


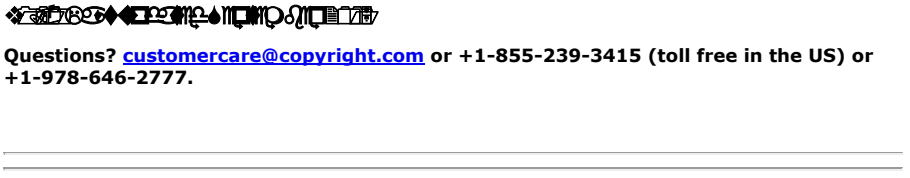
- 
- 
- 
- 
- 
- 
- 

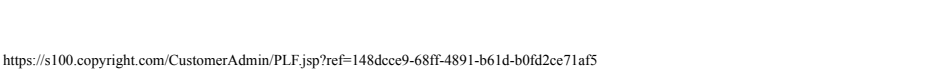
WILEY OPEN ACCESS TERMS AND CONDITIONS

The Creative Commons Attribution License

Creative Commons Attribution Non-Commercial License







APPENDIX B

COMPLETE TABLE OF ALL SAMPLES USED IN PRIMATE STUDY

Appendix B

Complete table of all samples used in primate study

SRR ID	Study ID	Biosample ID	SRS ID	Sex	Species
SRR032116	SRP001558	SAMN00007004	SRS009313	F	*H. sapiens*
SRR032117	SRP001558	SAMN00007005	SRS009314	F	*H. sapiens*
SRR032118	SRP001558	SAMN00007006	SRS009315	F	*H. sapiens*
SRR032119	SRP001558	SAMN00007007	SRS009316	F	*H. sapiens*
SRR357437	SRP008743	SAMN00744084	SRS268623	F	*H. sapiens*
SRR032120	SRP001558	SAMN00007008	SRS009317	F	*H. sapiens*
SRR032121	SRP001558	SAMN00007009	SRS009318	F	*H. sapiens*
SRR361334	SRP008743	SAMN00744094	SRS268633	F	*H. sapiens*
SRR350973	SRP008743	SAMN00722964	SRS265059	F	*H. sapiens*
SRR357412	SRP008743	SAMN00722969	SRS265064	M	*H. sapiens*
SRR306854	SRP007412	SAMN00632246	SRS214088	M	*H. sapiens*
SRR032122	SRP001558	SAMN00007010	SRS009319	M	*H. sapiens*
SRR306855	SRP007412	SAMN00632246	SRS214088	M	*H. sapiens*
SRR032123	SRP001558	SAMN00007011	SRS009320	M	*H. sapiens*
SRR306856	SRP007412	SAMN00632247	SRS214089	M	*H. sapiens*
SRR032124	SRP001558	SAMN00007012	SRS009321	M	*H. sapiens*
SRR032125	SRP001558	SAMN00007013	SRS009322	M	*H. sapiens*
SRR032126	SRP001558	SAMN00007014	SRS009323	M	*H. sapiens*
SRR032127	SRP001558	SAMN00007015	SRS009324	M	*H. sapiens*
SRR032140	SRP001558	SAMN00007028	SRS009337	F	*M. mulatta*
SRR306786	SRP007412	SAMN00632181	SRS214023	F	*M. mulatta*
SRR032141	SRP001558	SAMN00007029	SRS009338	F	*M. mulatta*
SRR032142	SRP001558	SAMN00007030	SRS009339	F	*M. mulatta*
SRR032143	SRP001558	SAMN00007031	SRS009340	F	*M. mulatta*
SRR032144	SRP001558	SAMN00007032	SRS009341	F	*M. mulatta*
SRR357438	SRP008743	SAMN00744085	SRS268624	F	*M. mulatta*
SRR032145	SRP001558	SAMN00007033	SRS009342	F	*M. mulatta*
SRR357411	SRP008743	SAMN00722968	SRS265063	F	*M. mulatta*

Appendix B

Complete table of all samples used in primate study (continued)

SRR ID	Study ID	Biosample ID	SRS ID	Sex	Species
SRR306787	SRP007412	SAMN00632182	SRS214024	M	*M. mulatta*
SRR306788	SRP007412	SAMN00632182	SRS214024	M	*M. mulatta*
SRR357439	SRP008743	SAMN00744086	SRS268625	M	*M. mulatta*
SRR032146	SRP001558	SAMN00007034	SRS009343	M	*M. mulatta*
SRR032147	SRP001558	SAMN00007035	SRS009344	M	*M. mulatta*
SRR032148	SRP001558	SAMN00007036	SRS009345	M	*M. mulatta*
SRR032149	SRP001558	SAMN00007037	SRS009346	M	*M. mulatta*
SRR032150	SRP001558	SAMN00007038	SRS009347	M	*M. mulatta*
SRR032151	SRP001558	SAMN00007039	SRS009348	M	*M. mulatta*
SRR032128	SRP001558	SAMN00007016	SRS009325	F	*P. troglodytes*
SRR032129	SRP001558	SAMN00007017	SRS009326	F	*P. troglodytes*
SRR357413	SRP008743	SAMN00722970	SRS265065	F	*P. troglodytes*
SRR032130	SRP001558	SAMN00007018	SRS009327	F	*P. troglodytes*
SRR032131	SRP001558	SAMN00007019	SRS009328	F	*P. troglodytes*
SRR306823	SRP007412	SAMN00632217	SRS214059	F	*P. troglodytes*
SRR032132	SRP001558	SAMN00007020	SRS009329	F	*P. troglodytes*
SRR032133	SRP001558	SAMN00007021	SRS009330	F	*P. troglodytes*
SRR357406	SRP008743	SAMN00722965	SRS265060	F	*P. troglodytes*
SRR357432	SRP008743	SAMN00744079	SRS268618	M	*P. troglodytes*
SRR306824	SRP007412	SAMN00632218	SRS214060	M	*P. troglodytes*
SRR032134	SRP001558	SAMN00007022	SRS009331	M	*P. troglodytes*
SRR357440	SRP008743	SAMN00744087	SRS268626	M	*P. troglodytes*
SRR032135	SRP001558	SAMN00007023	SRS009332	M	*P. troglodytes*
SRR032136	SRP001558	SAMN00007024	SRS009333	M	*P. troglodytes*
SRR032137	SRP001558	SAMN00007025	SRS009334	M	*P. troglodytes*
SRR032138	SRP001558	SAMN00007026	SRS009335	M	*P. troglodytes*
SRR032139	SRP001558	SAMN00007027	SRS009336	M	*P. troglodytes*