METABOLOMIC ANALYSIS OF AGING AND LONGEVITY

by

JESSICA MARIE HOFFMAN

(Under the Direction of Daniel Promislow)

ABSTRACT

Recent advances in the biology of aging have uncovered many genetic and environmental factors that influence aging and longevity. However, even with these discoveries, we can still only explain a small proportion of the variance seen in longevity within and among species. My thesis attempts to close this gap by employing metabolomics, the study of all the small molecules in an organism. I first examine how the metabolome changes with age, sex, and genotype in the fruit fly *Drosophila melanogaster*. I find the metabolome is highly variable (and predictive of) all three factors, and I find dramatic age related changes occurring in metabolic pathways associated with monoamine neurotransmitters. I then take a comparative approach across the *Drosophila* genus to determine how the metabolome is associated with longevity and resistance to oxidative stress in the largest comparative metabolomics study to date. I find that the metabolome varies considerably among species, and that much of the variability is explained by interspecific variation in body weight. Within sexes, a large fraction of the metabolome is associated with species differences in longevity and oxidative stress resistance. Despite these strong correlations, the metabolome appears to contain only a very weak phylogenetic signal. Finally, while *Drosophila* studies have the potential to

give many new insights into the basic biology of aging, results in the species can be limited in their translational impact to humans. In this light, I attempt to understand the metabolome's association with age in a non-human primate, the common marmoset. I use a longitudinal approach and find the metabolome is highly predictive of age, and I can identify metabolites that show consistent age-related trajectories within individuals. Many of these changes are associated with oxidative stress and xenobiotic metabolism, both of which are already known to be associated with age in other model organisms. This suggests that the association between age and specific metabolic pathways might be conserved across millions of years of evolution. Taken together, the results of this thesis demonstrate the powerful potential metabolomics has to discover the underlying biochemical pathways influencing natural variation in aging and longevity.

INDEX WORDS: Drosophila; marmoset; metabolomics; aging; longevity

METABOLOMIC ANALYSIS OF AGING AND LONGEVITY

By

JESSICA MARIE HOFFMAN

B.S., Clemson University, 2008

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY`

ATHENS, GEORGIA

2015

METABOLOMIC ANALYSIS OF AGING AND LONGEVITY

by

JESSICA MARIE HOFFMAN

Major Professor:	Daniel Promislow
Committee:	Kate Creevy
	David Hall
	Paul Schliekleman
	Shaying Zhao

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2015

ACKNOWLEDGEMENTS

Thank you to all family, friends, and colleagues who supported me over the past five years of my graduate career.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

The largest risk factor of mortality of an organism is its age; however, the genetic and biochemical determinants of longevity are largely unknown. Advances in genetic sequencing over the last decade have enabled researchers to use large scale genetic mapping techniques to try to discover genes associated with age. However, these studies are only able to determine genes that explain a small portion of the genetic variance seen in longevity (reviewed in Murabito et al., 2012), and the metabolic consequences of these genes are often unknown.

I am attempting to fill this gap in our knowledge by using high sensitivity metabolomics. Metabolomics is the study of all small molecules in an organism (less than 1000 Daltons), and it is the least studied of the major "omics" in biology. Metabolomics studies have the potential to unlock the molecular mechanisms that govern many different phenotypes of interest, and here I attempt to determine how aging and longevity in multiple organisms are associated with the metabolome. If we can find a link between metabolites and aging/longevity, we have the potential to bridge the gap between genotype and phenotype.

My dissertation work is focused on understanding the link between aging and metabolites in three separate studies. My first study attempts to discover the metabolome's association with age in one species, the fruit fly, *Drosophila melanogaster*. Then, I move

onto a comparative study of eleven different *Drosophila* species to determine how the metabolome changes across species and correlates with longevity. This evolutionary approach allows me to determine if distantly related, long-lived *Drosophila* species have similar metabolomic profiles. While the fruit fly provides an excellent model to understand the basic biology of aging, they have over 800 million years of evolution between them and humans, which can potentially limit their translational abilities. Therefore, my final work focuses on the effects of age on the metabolomic profiles of the common marmoset, *Callithrix jacchus*, a small short-lived primate, which is less than 40 million years diverged from humans.

AGING

Age is the greatest risk factor for morbidity and mortality of an organism; and as such, aging has been extensively studied over the last hundred years. In humans, vast differences in longevity can be seen with about 25% of the variation due to genetic factors (Herskind et al., 1996; Hjelmborg et al., 2006). Different theories of aging are commonly argued in biology, including antagonistic pleiotropy, mutation accumulation, and oxidative stress (reviewed in Hughes and Reynolds, 2005), and many of these theories rely on specific genes affect aging. Yet the discovery of genes that are significantly associated with longevity has been rare.

With the advent of next generation sequencing technologies, researchers believed they would be able to determine the genetic basis of aging and longevity. The most widely used technique in human aging studies has been association mapping using single nucleotide polymorphism (SNPs) to correlate these genetic markers with the a long and

short lived individuals (reviewed in Murabito et al., 2012). However, even with thousands of individuals in a study, SNP mapping often fails to find any gene that can be repeatably found to be associated with longevity over multiple studies. The greatest repeatability found has been with the APOE gene (Deelen et al., 2014; Ewbank, 2004), yet even this gene is only able to explain a percent of the variance in longevity, leaving most of the genetic variation in longevity (~25%) unaccounted.

METABOLOMICS

The failure of GWAS leaves a gap in our knowledge about the genetic variants that underlie aging and longevity, so a new technique is needed to help us bridge the gap between genotype and phenotype. Mapping individual metabolites has the potential to give new insights into genes associated with longevity, and the study of all the metabolites in an organism is referred to as metabolomics.

Metabolomics is among the least studied of all the major "omics", but the metabolome has the potential to be revolutionary in how we think about the genotype to phenotype map (Figure 1.1). The metabolome sits right below the phenotype in the genotype-phenotype pathway, and all lower levels affect it: the genome, transcriptome, and proteome. The metabolome also incorporates the environment more than other levels, as the environment can shift internally derived metabolites as well as many metabolites come from external sources. There is a large component of the environment in aging variation, and by understanding the metabolome we may be more easily able to tease apart the genetic and environmental variation that affect aging and longevity.

Metabolomic studies are primarily carried out by one of two technologies: mass spectroscopy and NMR (nuclear magnetic resonance) spectroscopy. Here, I focus on mass spectroscopy technologies, as those were employed for this dissertation. Mass spectroscopy involves detection of molecules and their abundance based on differences in their chemical properties. Samples are run through a mass spectrometer (mass spec) and are separated based on their mass to charge ratios (reviewed in Mishur and Rea, 2012). To correctly separate the most metabolites, different columns and ionization modes are utilized in the mass spec. First, both positive and negative ionization modes are always used as they can detect two different sets of metabolites (Dettmer et al., 2007), and different columns that samples are run through are also employed for the detection of different metabolites. The two most commonly used columns are a reverse phase C18 column which can detect hydrophobic, often organic, molecules and an anion exchange (AE) column which detects negatively charged metabolites. The use of multiple columns allows for the detection of thousands of metabolites; however, many metabolites will go undetected in the mass spec.

Recently, metabolomic studies have moved from only a hundred or so molecules to the ability to detect thousands (Jones et al., 2012), which has become a problem for metabolite identification. Each metabolite separated in the mass spec has an individual mass to charge ratio (m/z) based on its chemical make-up; however, two metabolites with the same chemical formula will have the same m/z, for example glucose and galactose. Two metabolites with the same m/z can have different retention times in the mass spec to help separate their identities; however, no annotation programs are yet able to incorporate retention time in metabolite identification. The only other way to differentiate two

identical m/z requires expensive mass spectrometry/mass spectrometry (MS/MS) techniques, which break up metabolites into pieces to determine true identity. However, MS/MS is very costly and time consuming, such that it is not feasible on large-scale datasets

Recently, studies have begun to map genes to individual metabolites using SNP mapping. These studies have found that individual SNPs can explain a large proportion of the variance seen in an individual metabolite (Suhre and Gieger, 2012). If that metabolite is correlated with a phenotype of interest, it can point to new candidate genes that may be influencing that phenotype. An individual metabolite is only a small part of the larger phenotype, so by mapping metabolites, researchers can begin to break down the phenotype into smaller parts to find new candidate genes. Using metabolomics to find metabolites associated with aging, and then mapping them back to the genome might be the breakthrough needed to finally begin to understand the genetic basis of aging and longevity.

METABOLOMICS AND AGING

Early metabolomics studies tended to focus on biomarkers of specific diseases (e.g. Qiu et al., 2010; Slupsky et al., 2010), where metabolomic profiles of diseased individuals are compared to healthy counterparts; however, recently more metabolomics studies have begun to look at the metabolic consequences of aging and longevity, predominately in model organisms. Studies have found specific metabolites associated with age in worms (Fuchs et al., 2010), flies (Hoffman et al., 2014), and mice (Wijeyesekera et al., 2012).

In chapter two of this thesis, I focus on understanding the metabolomic correlates of aging in the fruit fly, *Drosophila melanogaster*. In addition to being one of the most studied

model organisms in biology, the fruit fly is an ideal organism to understand aging and longevity. With a relatively short lifespan (~90 days maximum), longevity experiments can be easily completed and repeated, and genetic manipulations of candidate genes are possible. While the fruit fly provides an excellent model of aging, few studies have attempted to determine the metabolomic changes that occur with natural aging in the fly.

While previous metabolomics studies have been instrumental in our understanding of the associations between metabolites and age, most studies only analyze several dozen to a couple hundred metabolites, which represent only a small proportion of the entire metabolome (Jones et al., 2012). The entire metabolome of an organism consists of tens of thousands of metabolites, and my work centers on using high sensitivity metabolomics techniques to get a better understanding of aging on the entire metabolome of an organism.

In chapter two, I determine how the metabolomic profile of the fruit fly, *Drosophila melanogaster*, is changed with age, sex, and genotype and all their two-way interactions. I find that hundreds of metabolites are associated with age, and many of these metabolites are in metabolic pathways involving monoamine neurotransmitters and the carnitine shuttle. Interestingly, I also find large interaction effects between age and both sex and genotype, suggesting there are many metabolites whose age association is sex or genotype specific. These results indicate if we can determine metabolites that vary with age across more genotypes, we might be able to map specific metabolites to the genes that regulate them.

COMPARATIVE STUDIES AND AGING

While studies of individual species are important in our understanding of aging and longevity, they give us little insight into how the aging phenotype may have evolved. In this light comparative studies have long been employed to understand the variation (or lack of) that occurs between species, and how this variation evolved in response to adaptation (Harvey and Pagel, 1991). Comparative studies have played extensive roles in studies of longevity, senescence, and mortality (e.g. Promislow, 1991; Promislow and Harvey, 1990; Promislow and Haselkorn, 2002); however, little is known about how biochemical processes influence these traits across species.

Comparative studies analyzing metabolomic profiles have been extremely limited and often focus on only clustering patterns of metabolites. One study on only a couple dozen metabolites in sponges was able to show that metabolites could separate two distinct groups of sponges studied (Ivanisevic et al., 2011). However, a large metabolomics study of seven vertebrate species was not able to recapitulate their correct species phylogeny (Park et al., 2012), with results suggesting metabolomic profiles of human were more closely related to pigs than other primates. Yet a second study in primates suggests the metabolome of primate brains shows large evolutionary conservation with little changes due to the environment (Bozek et al., 2014). These contradictory results suggest we still have a long way to go to completely understand how the metabolome evolves and affects different phenotypes. This gap in our knowledge can be remedied by studying a group of closely related species that have sequenced genomes (necessary to determine how protein rates correlate with metabolite concentrations). The *Drosophila* species group

provides an ideal model to study how evolution affects the metabolome and its association with longevity.

Ever since the 12 *Drosophila* genomes project was completed (Clark et al., 2007), dozens of *Drosophila* comparative studies have been completed to look at how the genome evolves across the phylogeny (i.e., Greenberg et al., 2008; Larracuente et al., 2008; Stark et al., 2007). These studies have shown that rates of non-synonymous to synonymous protein change rates are smaller for proteins that catalyze enzymatic reactions compared to non-enzymatic proteins (Larracuente et al., 2008), and levels of protein constraint are higher in essential metabolic pathways compared to non-essential ones (Greenberg et al., 2008). While these studies give us insights into how evolution works on proteins involved in metabolic pathways, they do not show how the individual metabolites are affected by these changes in protein sequence. This is important to understand as selection works on the protein sequences, which catalyze metabolic reactions, so metabolites are the end product of selective forces. Yet these effects are unknown. To begin to understand the effects of selection on metabolite, we first need to know how metabolomic profiles vary between species.

One study has attempted to determine how metabolite levels vary in relation to changes in protein sequence. Using 35 genes in human erythrocytes cells, previously defined metabolic flux ratios were discovered to correlate with the ratio of non-synonymous to synonymous changes in gene sequences (Colombo et al., 2014). The researchers found high flux enzymes were more constrained compared to those enzymes with low flux rates. However, this study only focused on previously published flux distributions, and it did not compare these values across different species (though multiple

primates were used to determine rates of protein evolution). There remains a huge gap in our knowledge of how individual metabolite quantities, ratios, and variances across species are influenced by rates of protein evolution, and there is no clear way to correctly analyze these types of data, similar to problems seen with comparative transcriptome analyses (Dunn et al., 2013; Schreiber et al., 2009).

In chapter three, I complete the largest metabolomic comparative study to date. I analyze metabolomic data from eleven *Drosophila* species to begin to understand how metabolite concentrations vary both within and between species. I find little phylogenetic signal in the metabolome, and I discover large effects of sex, species, and body weight on overall metabolomic profiles. I also find many metabolites that vary with longevity and resistance to oxidative stress, but they tend to be associated in a sex specific manor. My results suggest there is potentially great power in using comparative metabolomic studies to understand the evolution of life history traits.

THE COMMON MARMOSET AS A MODEL OF HUMAN AGING

Drosophila provide an excellent group of species to study aging due to the fact that they have a relatively short lifespan while being fairly easy to manipulate. However, fruit flies are not always the best model for human aging for two reasons. First, they are not phylogenetically closely related to humans (800 million years of divergence time), which suggests aging correlates in *Drosophila* might not be found in humans unless they are highly conserved, such as the insulin signaling pathway (Tatar et al., 2003). Second, to completely understand the physiological changes that influence aging, longitudinal studies

must be completed that follow individuals across their entire life. However, longitudinal studies looking at physiological traits in flies are impossible due to destructive sampling, and longitudinal studies in humans can be done but are difficult to accomplish due to long lifespans (Blumenthal, 1985). To remedy these two problems, non-human primates have been studied, as their lifespans are shorter than humans, yet they are more closely related phylogenetically to human than other model organisms. Here, I analyze metabolomic profiles from a non-human primate model of aging, the common marmoset, to understand the biochemical changes that occur with age.

The common marmoset, *Callithrix jacchus*, has recently been proposed as a new model non-human primate for aging research (Fischer and Austad, 2011). The average lifespan of the marmoset is relatively short (6-12 years), even compared to other non-human primates, and they present with many clinical diagnoses similar to humans (Nishijima et al., 2012; Tardif et al., 2011). The genome of the marmoset was recently published (Worley et al., 2014) making it an ideal species to study aging with potential translational properties to humans.

Previous longitudinal studies in primates have focused primarily on the effects of different treatments on aging, especially calorie restriction (e.g., Colman et al., 2014), and previous metabolic longitudinal studies tend to focus on responses to different conditions (Li et al., 2012). While these studies are pivotal in gaining information about different interventions that may delay aging, they do not look at the biochemical changes that occur during the natural aging process.

While the marmoset provides an ideal non-human primate to study aging and understand the metabolomic correlates that occur with age, little is known about these

biochemical changes that occur with age within the species. Previous research has shown that many basic blood chemistry parameters can be correlated with age (Kuehnel et al., 2012), and metabolomic correlates with specific, pre-defined metabolites of interest were found to be associated with age (Roede et al., 2012). However, the effects of age on the entire metabolome of the marmoset are unknown.

While no longitudinal marmoset metabolomics studies have been completed, a couple of studies have attempted to determine longitudinal metabolite changes in humans, yet these studies have often only followed individuals over only a couple of months (e.g. Nicholson et al., 2011). One recent study was able to follow individuals over a seven year time period discovering that metabolomic profiles tend to be conserved over the time period in individuals; however they did not determine if specific metabolites were changing with age across individuals (Yousri et al., 2014). And, even seven years is a relatively short time to study changes for an individual that can easily live beyond 80 years.

In chapter four, I complete one of the largest metabolomic analyses in the marmoset. I take a longitudinal approach following a colony of marmosets over a period of 17 months in which metabolomic blood profiles were analyzed from three different time points. Examining over 2000 metabolites, I was able to discover that the metabolome is highly predictive of age, and as such the metabolome has the potential to become a new biomarker of aging. Using a longitudinal approach allowed me to determine more metabolic pathways associated with age than using individual time points on their own. This is one of the longest longitudinal metabolomics studies completed to date, and it shows the importance of using longitudinal studies when trying to determine metabolic correlates with natural aging.

SUMMARY

Here, I lay the foundation for using metabolomics to understand the biochemical changes that influence aging and longevity. If we can determine specific metabolites that are correlated with longevity, then we have the potential to understand the genetic basis of these traits by discovering the genes influencing the metabolite. The results of this thesis have the ability to inform researchers about new hypotheses that could be integral to understanding the differences in longevity between the sexes and across species. Here I present the most comprehensive analysis of the metabolome on aging and longevity to date.

LITERATURE CITED

Blumenthal, H.T. (1985). Normal Human Aging - the Baltimore Longitudinal-Study of Aging - Shock,Nw, Greulich,Re, Andres,R, Arrenberg,D, Costa,Pt, Lakatta,Eg, Tobin,Jd. J Gerontol *40*, 767-767.

Bozek, K., Wei, Y.N., Yan, Z., Liu, X.L., Xiong, J.Y., Sugimoto, M., Tomita, M., Paabo, S., Pieszek, R., Sherwood, C.C.*, et al.* (2014). Exceptional Evolutionary Divergence of Human Muscle and Brain Metabolomes Parallels Human Cognitive and Physical Uniqueness. Plos Biol *12*.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N.*, et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*, 203-218.

Colman, R.J., Beasley, T.M., Kemnitz, J.W., Johnson, S.C., Weindruch, R., and Anderson, R.M. (2014). Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. Nat Commun *5*.

Colombo, M., Laayouni, H., Invergo, B.M., Bertranpetit, J., and Montanucci, L. (2014). Metabolic Flux Is a Determinant of the Evolutionary Rates of Enzyme-Encoding Genes. Evolution *68*, 605-613.

Deelen, J., Beekman, M., Uh, H.W., Broer, L., Ayers, K.L., Tan, Q., Kamatani, Y., Bennet, A.M., Tamm, R., Trompet, S.*, et al.* (2014). Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Hum Mol Genet *23*, 4420-4432.

Dettmer, K., Aronov, P.A., and Hammock, B.D. (2007). Mass spectrometry-based metabolomics. Mass Spectrom Rev *26*, 51-78.

Dunn, C.W., Luo, X., and Wu, Z.J. (2013). Phylogenetic Analysis of Gene Expression. Integr Comp Biol *53*, 847-856.

Ewbank, D.C. (2004). The APOE gene and differences in life expectancy in Europe. J Gerontol a-Biol *59*, 16-20.

Fischer, K.E., and Austad, S.N. (2011). The Development of Small Primate Models for Aging Research. Ilar J *52*, 78-88.

Fuchs, S., Bundy, J.G., Davies, S.K., Viney, J.M., Swire, J.S., and Leroi, A.M. (2010). A metabolic signature of long life in Caenorhabditis elegans. Bmc Biol *8*.

Greenberg, A.J., Stockwell, S.R., and Clark, A.G. (2008). Evolutionary Constraint and Adaptation in the Metabolic Network of Drosophila. Mol Biol Evol *25*, 2537-2546.

Harvey, P.H., and Pagel, M. (1991). The comparative method in evolutionary biology (Oxford).

Herskind, A.M., McGue, M., Holm, N.V., Sorensen, T.I.A., Harvald, B., and Vaupel, J.W. (1996). The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet *97*, 319-323.

Hjelmborg, J.V., Iachine, I., Skytthe, A., Vaupel, J.W., Mcgue, M., Koskenvuo, M., Kaprio, J., Pedersen, N.L., and Christensen, K. (2006). Genetic influence on human lifespan and longevity. Hum Genet *119*, 312-321.

Hoffman, J.M., Soltow, Q.A., Li, S.Z., Sidik, A., Jones, D.P., and Promislow, D.E.L. (2014). Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, Drosophila melanogaster. Aging Cell *13*, 596-604.

Hughes, K.A., and Reynolds, R.M. (2005). Evolutionary and mechanistic theories of aging. Annu Rev Entomol *50*, 421-445.

Ivanisevic, J., Thomas, O.P., Lejeusne, C., Chevaldonne, P., and Perez, T. (2011). Metabolic fingerprinting as an indicator of biodiversity: towards understanding inter-specific relationships among Homoscleromorpha sponges. Metabolomics *7*, 289-304.

Jones, D.P., Park, Y., and Ziegler, T.R. (2012). Nutritional Metabolomics: Progress in Addressing Complexity in Diet and Health. Annu Rev Nutr *32*, 183-+.

Kuehnel, F., Grohmann, J., Buchwald, U., Koeller, G., Teupser, D., and Einspanier, A. (2012). Parameters of haematology, clinical chemistry and lipid metabolism in the common marmoset and alterations under stress conditions. J Med Primatol *41*, 241-250.

Larracuente, A.M., Sackton, T.B., Greenberg, A.J., Wong, A., Singh, N.D., Sturgill, D., Zhang, Y., Oliver, B., and Clark, A.G. (2008). Evolution of protein-coding genes in Drosophila. Trends Genet *24*, 114-123.

Li, X.L., Chen, Y.N., Liu, J.P., Yang, G., Zhao, J.M., Liao, G.N., Shi, M.M., Yuan, Y.J., He, S.R., Lu, Y.R.*, et al.* (2012). Serum metabolic variables associated with impaired glucose tolerance induced by high-fat-high-cholesterol diet in Macaca mulatta. Exp Biol Med *237*, 1310-1321.

Mishur, R.J., and Rea, S.L. (2012). Applications of mass spectrometry to metabolomics and metabonomics: Detection of biomarkers of aging and of age-related diseases. Mass Spectrom Rev *31*, 70-95.

Murabito, J.M., Yuan, R., and Lunetta, K.L. (2012). The Search for Longevity and Healthy Aging Genes: Insights From Epidemiological Studies and Samples of Long-Lived Individuals. J Gerontol a-Biol *67*, 470-479.

Nicholson, G., Rantalainen, M., Maher, A.D., Li, J.V., Malmodin, D., Ahmadi, K.R., Faber, J.H., Hallgrimsdottir, I.B., Barrett, A., Toft, H.*, et al.* (2011). Human metabolic profiles are stably controlled by genetic and environmental variation. Mol Syst Biol *7*.

Nishijima, K., Saitoh, R., Tanaka, S., Ohsato-Suzuki, M., Ohno, T., and Kitajima, S. (2012). Life span of common marmoset (Callithrix jacchus) at CLEA Japan breeding colony. Biogerontology *13*, 439-443.

Park, Y.H., Lee, K., Soltow, Q.A., Strobel, F.H., Brigham, K.L., Parker, R.E., Wilson, M.E., Sutliff, R.L., Mansfield, K.G., Wachtman, L.M.*, et al.* (2012). High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. Toxicology *295*, 47-55.

Promislow, D.E.L. (1991). Senescence in Natural-Populations of Mammals - a Comparative-Study. Evolution *45*, 1869-1887.

Promislow, D.E.L., and Harvey, P.H. (1990). Living Fast and Dying Young - a Comparative-Analysis of Life-History Variation among Mammals. J Zool *220*, 417-437.

Promislow, D.E.L., and Haselkorn, T.S. (2002). Age-specific metabolic rates and mortality rates in the genus Drosophila. Aging Cell *1*, 66-74.

Qiu, Y.P., Cai, G.X., Su, M.M., Chen, T.L., Liu, Y.M., Xu, Y., Ni, Y., Zhao, A.H., Cai, S.J., Xu, L.X.*, et al.* (2010). Urinary Metabonomic Study on Colorectal Cancer. J Proteome Res *9*, 1627-1634.

Roede, J.R., Uppar, K., Yongliang, L., Promislow, D.E.L., Wachtman, L.M., and Jones, D.P. (2012). Characterization of plasma thiol redox potential in a common marmoset model of aging. Redox Biology *1*, 387-393.

Schreiber, A.W., Sutton, T., Caldo, R.A., Kalashyan, E., Lovell, B., Mayo, G., Muehlbauer, G.J., Druka, A., Waugh, R., Wise, R.P.*, et al.* (2009). Comparative transcriptomics in the Triticeae. Bmc Genomics *10*.

Slupsky, C.M., Steed, H., Wells, T.H., Dabbs, K., Schepansky, A., Capstick, V., Faught, W., and Sawyer, M.B. (2010). Urine Metabolite Analysis Offers Potential Early Diagnosis of Ovarian and Breast Cancers. Clin Cancer Res *16*, 5835-5841.

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N.*, et al.* (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature *450*, 219-232.

Suhre, K., and Gieger, C. (2012). Genetic variation in metabolic phenotypes: study designs and applications. Nat Rev Genet *13*, 759-769.

Tardif, S.D., Mansfield, K.G., Ratnam, R., Ross, C.N., and Ziegler, T.E. (2011). The Marmoset as a Model of Aging and Age-Related Diseases. Ilar J *52*, 54-65.

Tatar, M., Bartke, A., and Antebi, A. (2003). The endocrine regulation of aging by insulin-like signals. Science *299*, 1346-1351.

Wijeyesekera, A., Selman, C., Barton, R.H., Holmes, E., Nicholson, J.K., and Withers, D.J. (2012). Metabotyping of Long-Lived Mice using H-1 NMR Spectroscopy. J Proteome Res *11*, 2224-2235.

Worley, K.C., Warren, W.C., Rogers, J., Locke, D., Muzny, D.M., Mardis, E.R., Weinstock, G.M., Tardif, S.D., Aagaard, K.M., Archidiacono, N.*, et al.* (2014). The common marmoset genome provides insight into primate biology and evolution. Nat Genet *46*, 850-857.

Yousri, N.A., Kastenmuller, G., Gieger, C., Shin, S.Y., Erte, I., Menni, C., Peters, A., Meisinger, C., Mohney, R.P., Illig, T.*, et al.* (2014). Long term conservation of human metabolic phenotypes and link to heritability. Metabolomics *10*, 1005-1017.

FIGURE LEGEND

Figure 1.1. Molecular pathway between genotype and phenotype. Interactions occur within and among the different levels, and the environment can affect all levels.

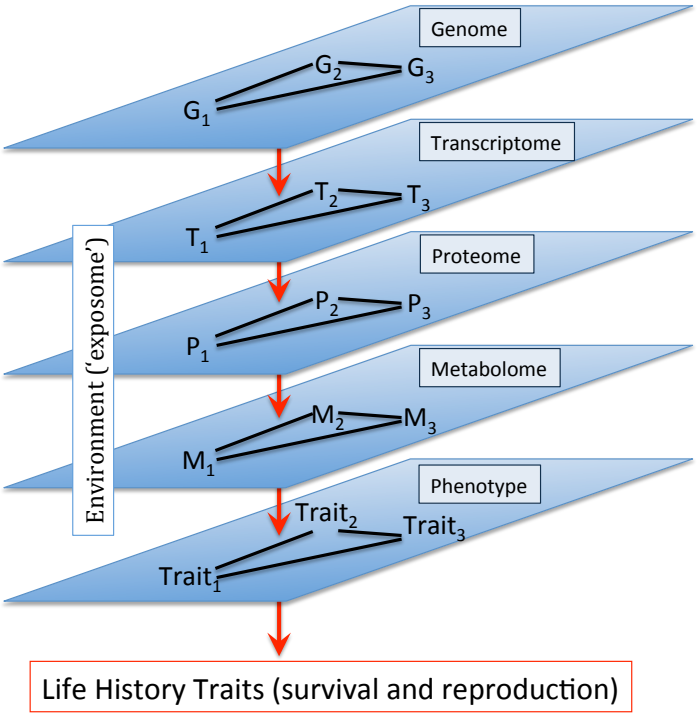
Genome
G1
G2
G3
Transcriptome
T1
T2
T3
Proteome
P1
P2
P3
Metabolome
M1
M2
M3
Phenotype
Trait1
Trait2
Trait3
Environment ('exposome')
Life History Traits (survival and reproduction)

Figure 1.1

CHAPTER 2

EFFECTS OF AGE, SEX AND GENOTYPE ON HIGH-SENSITIVITY METABOLOMIC PROFILES IN THE FRUIT FLY, *DROSOPHILA MELANOGASTER*[1]

[1]Hoffman J.M., Soltow Q.A., Li S., Sidik A., Jones D.P., and Promislow D.E.L. 2014. *Aging Cell*. 13(4): 596-604. Copyright 2014. Reprinted here with permission of publisher.

ABSTRACT

Researchers have used whole genome sequencing and gene expression profiling to identify genes associated with age, in the hope of understanding the underlying mechanisms of senescence. But there is a substantial gap from variation in gene sequences and expression levels to variation in age or life expectancy. In an attempt to bridge this gap, here we describe the effects of age, sex, genotype and their interactions on high-sensitivity metabolomic profiles in the fruit fly, *Drosophila melanogaster*. Among the 6800 features analyzed, we found that over one-quarter of all metabolites were significantly associated with age, sex, genotype or their interactions, and multivariate analysis shows that individual metabolomic profiles are highly predictive of these traits. Using a metabolomic equivalent of gene set enrichment analysis, we identified numerous metabolic pathways that were enriched among metabolites associated with age, sex, and genotype, including pathways involving sugar and glycerophospholipid metabolism, neurotransmitters, amino acids, and the carnitine shuttle. Our results suggest that high-sensitivity metabolomic studies have excellent potential not only to reveal mechanisms that lead to senescence, but also to help us understand differences in patterns of aging among genotypes and between males and females.

INTRODUCTION

Lifespan is a highly heritable trait. Over the past 20 years, researchers working on lab-adapted organisms have been able to identify evolutionarily conserved genetic pathways which, when knocked down or overexpressed, are able to dramatically increase lifespan. These successes underscore two critical questions: first, at the molecular level, what are

the underlying mechanisms by which these genes affect longevity; second, at the population level, do these same genes account for standing variation in longevity in natural populations?

These questions are complicated by the fact that the age at which an individual dies depends not only on its genotype, but also on a lifetime of effects accumulated through environmental exposure, the environment-specific response of genes, and the downstream physiological consequences of these complex factors. Fortunately, whole-genome sequencing and genome-wide association (GWA) studies now make it possible to identify segregating alleles that affect complex phenotypes such as body height, diabetes, schizophrenia, and even longevity (Jeck et al., 2012). But GWA studies suffer from numerous challenges, and these are further compounded in analyses of lifespan. First, alleles identified in GWA studies typically explain just 0.1-1.0% of the variation in complex traits (Park et al., 2010). Second, the genetic basis of lifespan appears, at least in part, to differ between the sexes (Burger and Promislow, 2004). Third, lifespan includes a substantial degree of stochasticity, varying dramatically even among genetically identical individuals raised in a constant and identical environment (Kirkwood et al., 2005). Finally, and perhaps most importantly, lifespan is a highly composite trait potentially influenced by the functional decline of many underlying processes. To fully understand the genetics of lifespan, we need to understand the genetics not simply of age at death, but rather of the underlying causes of death.

Here we suggest that many of the challenges that we face in our attempts to define the pathways that account for age-related declines in function, and for genetic variation in these declines, can be resolved through the use of high-resolution metabolomics (Mishur

and Rea, 2012). If we can decompose the physiological processes that influence morbidity and mortality to their constituent components (i.e., the metabolome), we will be an important step closer to bridging the gap between genotype and phenotype (Figure 2.1). The metabolome is effectively a functional intermediate between genotype and phenotype.

Previous work illustrates how the metabolome can serve as a strong bridge between genotype and phenotype. While allelic variation typically explains only a small fraction of the variation in complex phenotypes, GWA studies of the metabolome have found genetic variants capable of explaining up to 60% of the variance in the concentration of individual metabolites (Suhre et al., 2011). The metabolome also appears to be a sensitive indicator of age-related physiological changes both in invertebrates and vertebrates (e.g., Sarup et al., 2012; Yu et al., 2012). Moreover, different mutants that extend longevity share common metabolomic signatures (e.g., Caenorhabditis elegans: Fuchs et al., 2010; Mus musculus: Wijeyesekera et al., 2012).

While we have learned much from these initial studies, most have been limited by the use of relatively low-sensitivity metabolomics technology, and by limited genetic information. Studies of age and the metabolome have hitherto been carried out using standard metabolomic methods, which typically measure concentrations of several hundred metabolites, at most. This represents one percent or less of all the circulating metabolites found within an individual animal (Jones et al., 2012). This would be the equivalent of measuring just 300 genes in a 'genome-wide' screen in humans. Moreover, most of these studies have not been able to distinguish metabolomic variation that is due to genetic differences among individuals from variation due to non-genetic physiological differences.

Here we demonstrate the power of new, state-of-the-art high-resolution metabolomics (Jones et al., 2012) to characterize the aging metabolome. Hardware and software advances in high-resolution metabolomics now support extremely sensitive measurement of tens of thousands of metabolites (Jones et al., 2012; Uppal et al., 2013), and these high-resolution studies can be used to determine the effects of age on the entire metabolome (De Guzman et al., 2012). We use high-resolution metabolomics to ask three specific questions. First, is the metabolome a sensitive indicator of physiological state in the aging fly? Second, is the high-sensitivity metabolome a useful biomarker of age, can it predict the age of an individual, and moreover, can it explain sex- and genotype-specific variation in age. And third, can metabolomic studies reveal novel pathways associated with natural variation in aging?

METHODS

Fly stocks and culturing conditions

All analyses were carried out using a set of 15 inbred fruit fly genotypes randomly chosen from the Drosophila Genome Reference Panel (DGRP, Mackay et al., 2012) (Table S2.1). The DGRP consists of 192 fully-inbred and fully-sequenced strains of *Drosophila melanogaster*, and is freely available from the Bloomington *Drosophila* Stock Center. Flies were maintained in incubators at 24°C on a 12/12 light-dark cycle at ~50% humidity. For all procedures, flies were maintained on standard yeast-molasses-agar-cornmeal medium.

Collection of known-age flies

Prior to the onset of the study, fly cultures were expanded to include 4 bottles per genotype, at which point 150 virgin males and females were collected over a 72-hour period under light CO_2 anesthesia. For each sex and each of the 15 genotypes, we placed an average of 27 individuals in each of 5 40-ml glass vials, for a total of 4032 flies distributed among 150 vials. Flies were transferred to new vials very two days without anesthesia, at which time the number of dead flies in each vial was recorded.

At seven time points (days 3, 10, 24, 36, 51, 66 and 81), we collected two samples of three flies from each unique genotype-sex cohort without anesthesia, placed each sample in a 1.5-ml Eppendorf tube, instantly froze the samples in liquid nitrogen, and then placed these samples in a -80°C freezer until the end of the experiment. Not all genotypes survived to age 81 d, and in some cases at later ages only one sample of three flies per genotype and sex was collected.

Metabolomic analysis

Each frozen fly sample was homogenized using a Pellet Pestle® Motor in 150 µl acetonitrile in water (2:1 v/v) containing an isotopic standard mix (Soltow et al., 2011) and refrozen. Immediately before analysis, the samples were thawed, vortexed, and centrifuged at 14,000 rpm for 10 min at 4 ºC. Extracts (100 µl) were randomized, placed in a refrigerated autosampler and 10 µl volumes were analyzed in duplicate with dual chromatography-mass spectrometry (DC-MS) platforms (Soltow et al., 2011), one using an anion exchange (AE) column (Hamilton PRPX-110S, 2.1 mm ×10 cm, Reno, NV) and the other using a C18 column (Higgins Analytical, Targa, 2.1 mm x 10 cm, Mtn View, CA). C18 and anion exchange chromatography separate molecules based upon different chemical

properties. C18 is also termed "reverse phase" chromatography because the 18-carbon units are hydrophobic, retaining and separating chemicals with partial hydrophobic character. Anion exchange, on the other hand, has positive charges on the column, which retain and separate negatively charged molecules. The conditions used result in about 30% overlap between columns in chemicals detected.

Samples were fractionated with a formate or acetonitrile gradient, respectively, ionized with electrospray ionization in the positive mode, and detected with an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) with *m/z* from 85 to 2000 and 30,000 resolution. Data were extracted using apLCMS (Yu et al., 2009) as *m/z* features, where an *m/z* feature is defined by *m/z* (mass/charge), RT (retention time) and ion intensity (integrated ion intensity for the peak). As part of quality control, we also generated a 'fly standard', consisting of a large volume of identical sample taken from 350 flies, which was run alongside samples daily to evaluate reproducibility. While some overlap in metabolites between the two columns is expected, data from both columns cannot efficiently be combined for analysis due to very different column chemistries which can lead to the same metabolite showing different ion intensities and retention times. Therefore, we analyzed data from both columns separately.

Data Analysis

Quality control

Many metabolites show significant stochastic variation even within samples, and thus are likely to be uninformative. To minimize the impact of 'noisy' metabolites, prior to data analysis, we carried out a set of quality control procedures to limit our analysis to the most

informative metabolites. First, we only included metabolites with a signal to noise ratio (SNR_i = mean/sample standard deviation) ≥ 15. Second, we log-transformed the data, which led to a normal distribution of concentrations across all metabolites. Third, we removed any metabolites that were missing from more than 5% of either all male samples or all female samples. Fourth, we used the LSimpute imputation procedure (Bo et al., 2004) to estimate the values of missing samples. Fifth, we limited our analysis to metabolites with a mass-charge ratio of less than 900, as data collection parameters were optimized for m/z ≤ 900. Finally, once all these procedures were complete, we normalized the data such that each sample had a mean value of 0.

Metabolite-specific analysis

All statistical analyses were carried out using the statistics package R.

We used a general linear model to test for the effects of age (A), sex (S) and genotype (G) on metabolite intensity (Y):

$$Y = \mu + A + S + G + A\times S + A\times G + G\times S + \varepsilon \quad (1)$$

treating all predictors as fixed effects, where μ is the grand mean and ε is the residual error. We treated age as an ordered factor (effectively a categorical variable, with the proviso that we know that age 3 < age 10, age 10 < age 24, and so forth). Due to small sample sizes and not all genotypes being present, age 81 flies were removed, leaving 274 samples for metabolite-specific analyses.

By including all parameters in the model, we are asking, for example, whether sex has a significant effect on metabolite intensity after controlling for the effects of age and

genotype. To determine significance for each factor in the model shown in Equation 1, we set the false discovery rate (Benjamini and Hochberg, 1995b) at 1% using all P-values associated with that specific factor. To obtain P-values for the effects of genotype and of its interactions with age or sex, we carried out a likelihood ratio test using the `lrtest` function in the *epicalc* package in R.

We followed Hoffmann and Parsons (1988) to obtain approximate estimates of narrow-sense heritability (h^2) from the intraclass correlation (t) between lines. Here, $t = \sigma_b^2/(\sigma_b^2 + n\sigma_w^2)$, where σ_b^2 and σ_w^2 are the between-line and within-line variances, respectively, and n = 3 is the number of individuals within each sample. Within and between genotype variances were determined using the `lme` function in the package `nlme` in R, treating age and sex as fixed effects and genotype (line) as a random effect. While Hoffmann and Parsons (1988) and subsequent authors suggest various equations to convert t to h^2, here we present only the intraclass correlation.

Metabolome-wide analysis

To determine the degree to which metabolomic profiles could be used to predict sex or age, we used the sparse partial least squares discriminant analysis function (`splsda`) as implemented in the R package *mixOmics* (for sex and genotype), and partial least squares regression as implemented by the `mvr` function in the *pls* package in R (for age). We set the number of components as min (k-1), where k is the number of classes (2 for sex, 7 for age, 14 for genotype). For `splsda`, the number of metabolites included in the model was determined using 10-fold cross validation in *mixOmics*. We chose the number of metabolites to include in each analysis based on that number which minimized the

Classification Error Rate (CER), using 10-fold cross validation. Supervised classification schemes with relatively low numbers of samples and high numbers of variables can lead to overfitting (Westerhuis et al., 2008). Accordingly, we compared observed Classification Error Rates (CER) with mean CER (± 1 s.e.) obtained from ten permutation tests in which the classification group (sex or age) was sampled randomly without replacement. This comparison tells us whether the observed classification is any better than one would expect by chance.

We used partial least squares to predict sample age using training and testing sets. In this case, training sets consisted of 2/3 of all samples, and the model derived from this set was used to predict the classification for the remaining 1/3 of the samples. We obtained R^2 values using ten-fold cross validation and compared these values with R^2 values where the class variable was permuted.

Metabolite annotation and tests of metabolite enrichment

In this study, we adopted a novel approach to perform putative metabolite annotation and enrichment analysis in one step (Li et al., 2013). Theoretically, most metabolites from mass spectrometry can match multiple metabolite compounds. The program that we have used here, *mummichog*, selects the most probable metabolites from these multiple candidates, based on the enrichment of metabolic networks and pathways, since a biological process is expected to favor a more connected network over random distributions. This approach has been validated on multiple experimental data sets (Li et al., 2013). The fly metabolic network model from the MetaCyc database (Caspi et al., 2012) and the reference model in *mummichog* (Li et al., 2013) were used to test the enrichment of

pathways and networks. The null distribution in pathway analysis is based the permutation of metabolite selections, which takes into full consideration of the aforementioned multiple-matching problem.

We first determined subsets of metabolites to look for pathway enrichment. We ran a model for both columns separately of only age, sex, and their interaction, and took the 250 most positively and 250 most negatively associated metabolites based on their *P*-value for each parameter (age, sex, and the interaction between the two). This gave us 12 different subset lists (six from each column) that were then run through the program *mummichog* for pathway enrichment analysis.

For age and genotype interactions, we ran a likelihood ratio test as described above to determine those metabolites with a significant age-genotype interaction. We chose 250 metabolites with the smallest p-value for both the C18 and AE columns for enrichment analysis.

RESULTS

We analyzed metabolomics data from 15 inbred lines from the *Drosophila* Genome Reference Panel (DGRP, Mackay et al., 2012) (Table S2.1). After quality control (see Experimental Procedures), our final dataset consisted of 293 biological samples, each of which was run twice through each of two columns. Each sample provided measures for 3091 features from an anion exchange (AE) column, and 3714 features by reverse phase (C18) liquid chromatography, for a total of 6805 features (note that some overlap occurs between columns) (see Experimental Procedures below, and Soltow *et al*. 2011, for a description of the differences between these two columns). Of the features detected, we

were able to obtain putative molecular matches for 877 metabolites in the AE column and 717 metabolites in the C18 column based on mass-charge (m/z) ratio using the metabolite prediction program *mummichog* (Li et al., 2013). In our individual metabolite analysis, we define statistically significant effects on m/z concentrations using a conservative false discovery rate α = 0.01 (Benjamini and Hochberg, 1995b). Seven of the fifteen lines used here were putatively positive for *Wolbachia* infection (Table S2.1). Including *Wolbachia* status in our statistical models had no appreciable effect on our conclusions, and so is not discussed further.

We divide the presentation of results into three sections. First, we examine the broad effects of age, sex and genotype (including measures of heritability) on individual metabolites. Second, we identify metabolites that show significant interaction effects, looking in particular at age × sex and age × genotype interactions. Third, we describe specific types of metabolites that are enriched among all metabolites significantly associated with these main effects and their interactions.

Main effects

The metabolome appears to be highly sensitive to physiological state, showing dramatic variation in response to sex, age, and genotype (Table 2.1). The proportion of metabolites significantly associated with the traits we measured varied from a low of 1% of C18 metabolites increasing with age to a high of almost 14% of AE metabolites that vary among genotypes. In each case, effects of one variable were measured after controlling statistically for the effects of the other two variables. For example, our test for the effect of genotype held the effects of sex and age constant, effectively testing if there are significant

differences in the intercept of *m/z* intensity versus age among genotypes. Figure 2.2 shows two metabolites with confirmed identities with significant age effects and one with significant sex effects. Note that sex effects are likely underestimated, because we only included metabolites present in at least 95% of male samples and 95% of female samples.

As a separate measure of genetic effects, we used the intraclass correlation (*t*) among the fifteen inbred lines as a measure of heritability (the percentage of total variation due to genetic effects). We identified almost 300 metabolites with $t \geq 5\%$ (134 C18 metabolites, 149 AE metabolites), with maximum value of $t = 28.8\%$ among AE metabolites (*m/z* = 693.2182) and $t = 32.3\%$ among C18 metabolites (*m/z* = 179.0844). The full distribution of heritabilities is shown in Figure S2.1.

Supervised multivariate analysis revealed that the metabolome is strongly predictive of sex and age. Partial least squares-discriminant analysis differentiated almost completely between the metabolome of males and females (Figure 2.3) and among flies of different ages (Figure S2.2). Similarly, using partial least squares regression, we found the metabolome to be an accurate predictor of age. For both males and females, a model based on a random sample of one third of all individuals was able to explain between 78% and 92% of the variance in age of the remaining two-thirds of samples (Figure 2.4).

Interaction effects

In addition to the substantial proportion of metabolites that showed significant age-specific changes in intensity, or that differed significantly between sexes or genotypes, we also found numerous metabolites associated with interaction effects (i.e., metabolites for which the slope of change with age differs significantly between the sexes or among

genotypes (Table 2.1)). One example of a known metabolite with an age × sex interaction is shown in Figure 2.2.

Taken together, we found a total of 995/3091 metabolites (32.2%) in the AE column and 995/3714 (26.8%) in the C18 column whose concentration was effected by age, sex, genotype, or some interaction thereof.

Metabolite Enrichment

Using the program *mummichog* (Li et al., 2013), we asked whether subsets of metabolites associated with a particular trait or trait combination were enriched for specific classes of metabolites. Our analysis revealed a large number of metabolic pathways whose constituent metabolites were overrepresented among all metabolites associated with sex, age, and their interactions. The complete list is shown in Table S2.2.

Among metabolites that change with age, we identified four groups that are notable for being strongly enriched for certain pathways, and being of specific interest from an aging perspective. The first group, illustrated in Figure 2.5, includes metabolites associated with sugar and glycerophospholipid metabolism. These two pathways are connected via interactions with 4'-phosphopanthothenate. Metabolites associated with glycolysis, including G6P, F6P, and 'feeder' molecules F1P, G1P, galactose 1-phosphate, trehalose 6-phosphate, and mannose 6-phosphate, show a clear decline with age. Metabolites associated with glycerophospholipid metabolism show both increases and declines with age, and include sphingosine, phosphoryl ethanolamine, 1-L-myo-inositol-1-phosphate, and phosphoryl choline.

The second group of molecules is the carnitines, illustrated in Figure S2.3. These molecules, which make up the carnitine shuttle, are a key component of fatty acid metabolism, and show an almost uniform, highly significant decline with age.

The third and fourth groups, described in detail in Table 2.2, include amino acids and neurotransmitters, including molecules involved in monoamine metabolism as well as well as the inhibitory neurotransmitter γ-aminobutyric acid (GABA) and its precursor arginine.

Among the 150 most heritable metabolites, we observed enrichment for two pathways related to tryptophan metabolism—tryptophan degradation (AE: n = 4 observed / 8 total, P_{adj} = 0.0034; C18: n = 2 observed / 7 total, P_{adj} = 0.023) and serotonin and melatonin biosynthesis (AE: n = 4/6, P_{adj} = 0.0028; C18: n = 2/7, P_{adj} = 0.023).

DISCUSSION

Beginning with Pletcher's groundbreaking study on the transcriptome of aging fruit flies (Pletcher et al., 2002), systems biological studies of aging in the fruit fly have focused primarily on the transcriptome. Here we turned our attention to the metabolome, in an attempt to determine the degree to which the metabolome a) is associated with physiological state; b) can predict age, sex and genotype-specific variation in aging; and c) might reveal novel metabolic pathways associated with natural genetic variation in aging.

While previous studies have looked at the effects of age, sex, or genotype on metabolomic variation, this study is unique not only in its scope (i.e., the very large number of metabolites analyzed), but also in its design. By including age, sex, and genotype, we were able to measure the independent effects of all three of these fundamental biological parameters simultaneously, as well as of their interactions, and to identify specific

metabolic pathways associated with changes due to age, sex, and genotype. This study also takes advantage in particular of new advances in high-resolution metabolomics (Jones et al., 2012). With the availability of this new system, we can now identify tens of thousands of features from samples as small as a few flies (2-3 mg of tissue, in this case).

We found that the metabolome is, indeed, highly sensitive to physiological state as influenced by sex, by age and by underlying genotype, with as much as one-third of the entire metabolome responding significantly to these factors or their interactions. Moreover, the metabolome proved to be strongly predictive of sex and age. These findings hold out the hope that metabolomic profiles might be a powerful biomarker of age. Further studies are needed to determine whether samples that are younger or older than predicted by their metabolomic profile would be relatively longer- or shorter-lived, respectively.

While this study set out to identify metabolites affected by age, the most dramatic effects were found with respect to sex. In fact, we have likely underestimated the proportion of metabolites that differ between the sexes, as we excluded all metabolites that were absent in more than 5% of samples from either sex. Thus, metabolites that were found consistently in one sex, but less so or not at all in the other sex, would have been excluded from our analysis. Our finding of sex differences in multiple classes of metabolites is consistent with earlier studies on flies showing sex differences in lipid profiles (Parisi et al., 2011; Scheitz et al., 2013).

Given these extensive differences, and the ease with which we are able to carry out genetic manipulations in *Drosophila*, we should be able to use the fly as a model system to identify the genetic basis of sex-specific differences in the metabolome.

We found between 7% (C18) and 13% (AE) of all metabolites differed among genotypes. This finding holds out the promise of addressing two critical issues. First, given the strong genetic signature of variation in metabolites, we should be able to identify the genetic basis of individual variation in metabolites in *Drosophila*, as others have done in humans (e.g., Kettunen et al., 2012; Suhre and Gieger, 2012). Second, and of central interest to aging research, we should be able to determine the extent to which genotype-specific metabolite profiles are correlated with lifespan (and, of course, other traits of interest). These results also fit well within the context of earlier work showing the effects both of life-extending mutants (e.g., Fuchs et al., 2010; Wijeyesekera et al.) and of selection on lifespan (Sarup et al., 2012) on metabolite profiles. Whether these treatments can actually reverse the effects of age on the metabolome requires more detailed analysis.

Perhaps the most unexpected finding was the considerable number of metabolites that showed significant age-×-genotype interactions (Table 2.1). This finding was facilitated by the fact that we were able to measure metabolites at seven different ages for 15 different genotypes. Scaled up to a larger sample of genotypes, we are confident that future studies should enable us to identify single genes associated not only with baseline metabolite levels (e.g., Kettunen et al., 2012; Suhre et al., 2011), but also with the degree to which metabolite concentrations change with age, and the impact of these changes on other age-related traits.

The strong effect of a sex × age interaction on metabolomic profiles also holds out great promise for our ability to better understand why sexes differ in longevity and in patterns of age-associated disease. Previous studies have looked at the degree to which age and sex affect metabolite concentration (Psihogios et al., 2008; Slupsky et al., 2007), though these

studies have not looked specifically at *interactions* between sex and age. Going a step further, we might ask whether we can identify genes that affect sex-differences in age-specific trajectories of specific metabolites. While such a 'third-order' analysis is complex, its feasibility is hinted at by the existence of numerous metabolites with statistically significant three-way interaction between sex, age and genotype (C18: $n = 56$; AE: $n = 34$).

Our findings suggest not only that the metabolome might be a useful biomarker of physiological state, but also that it might reveal novel pathways associated with age. The age-related decline in carnitines was perhaps the most consistent pattern that we observed here. Carnitines are critical for the transfer of fatty acids into the mitochondrion, where they undergo β-oxidation, generating acetyl-CoA, which then enters the TCA cycle. In their study of aging in mice, Houtkooper *et al*. (2011) found a similar age related decline not only in acylcarnitines, but also in genes associated with fatty acid metabolism. Similar declines have also been seen in humans (e.g., Gomez et al., 2012), and numerous studies point to the ameliorative effects of supplemental carnitines on senescence (e.g., Noland et al., 2009). Our findings are consistent with another recent study in *Drosophila* showing age-related changes in fatty acid profiles (Moghadam et al. 2013).

Amino acid (AA) balance is thought to explain the dramatic response of lifespan to dietary restriction (Grandison et al., 2009), and AAs are important activators of the aging-related Target of Rapamycin (TOR) pathway. We found AA enrichment among metabolites that increased with age, that decreased with age, that were higher in males, and that interacted between age and sex or age and genotype. These last two groups suggest that amino acid levels might be useful in predicting differences in patterns of aging between males and females, or among genotypes. In support of this potential, AAs were among the

metabolites with the highest heritability, a finding consistent with previous studies on the heritability of the metabolome (Kettunen et al., 2012; Rhee et al., 2013; Suhre et al., 2011). Interestingly, the AAs that we identified as having the highest heritability were similar to three of four AAs found to be highly heritable in humans, including isoleucine, proline and glutamine (Suhre et al., 2011).

Among metabolites that declined with age, we also observed enrichment of those associated with glycolysis and glycophospholipid metabolism. These two groups were linked through pantothenic acid (Vitamin B), which is required for the conversion of pyruvate (from glycolysis) to acetyl CoA, part of phospholipid biochemistry. Our data suggest that at least in flies, glycolysis declines with age. We saw that overall, membrane phospholipids declined with age, though some metabolites, such as phosphatidylethanolamine and phosphatidylcholine, moved in opposite directions (Figure 5), an observation consistent with previous studies (Kostal et al., 2011). Studies in flies have shown strong effects of thermal stress on membrane phospholipids (e.g., Overgaard et al., 2008). Future studies might benefit from a focus on the degree to which the well-established effect of temperature on lifespan is affected by changes in phospholipid biochemistry.

In this study, biogenic amines were associated with age and sex, and tryptophan metabolism in particular showed high levels of heritability. Previous studies in flies have suggested that both dopamine and serotonin signaling might be important regulators of aging (De Luca et al., 2003; Vermeulen et al., 2006), either directly, or indirectly through their interaction with insulin signaling pathways (Toivonen et al., 2009).

Interestingly, we also saw enrichment for the tryptophan degradation pathway, to kynurenine, among metabolites with the highest heritability. Recent work has implicated this pathway to aging in both worms (Coburn et al., 2013) and flies (Oxenkrug et al., 2011).

Given the enrichment of these pathways among the most heritable metabolites, our results offer hope that these pathways might help to explain variation in important fitness traits, including survival, in natural populations. This might be especially true of sex-specific differences in aging. Biogenic amines were found to be common not only among metabolites that differed between males and females, but also among metabolites whose age-specific trajectories were sex-specific (Table S2.1).

There are, of course, some limitations to this work. First and foremost, we do not yet have a curated *Drosophila* metabolome. While we are able to obtain putative matches for approximately one-quarter of all metabolites, these matches often carry considerable uncertainty. In some cases, a single mass-charge ratio might match 10 or more putative known molecules. Once the fly metabolome is curated, we will be able to more accurately describe the dynamics of fly metabolic pathways. Moreover, in limiting our assay to metabolites with *m/z*-ratios of less than 900, we have eliminated many components of lipid metabolism, which is of interest to aging studies (Barzilai et al., 2001).

Second, our study relied on whole flies. We know that metabolomic profiles differ among tissues in the fly (Chintapalli et al., 2013). Moreover, our whole-body analysis includes the gut, and so includes potentially confounding metabolites from the large bacterial flora found in the fly gut (Corby-Harris et al., 2007).

Third, while this is the largest-scale analysis of age-related change in the metabolome to date, the data represent just one experiment (albeit a large-scale one). Changes that appear

to be related to age might, in fact, be due to secular environmental trends that occurred over the 12-week course of the experiment (an often unstated concern of cohort aging studies).

Our work has established the power and potential of the fly metabolome as a model to both explain and predict variation in aging. In light of our findings, there are several avenues for future research of immediate interest. To begin with, we have established that there is substantial genetic variation in the metabolome, and that a large number of specific metabolites are not only affected by age, but also show age-effects that vary by genotype. Thus, we are confident that we can use metabolomic variation to tie together genotype with phenotype, identifying genes that affect metabolites, and metabolites that affect and/or reflect variation in lifespan. In this way, large-scale metabolomic studies hold out much promise in helping us to complete the genotype-phenotype map for aging. Furthermore, here we have focused on measures of individual metabolites. Recent work suggests that as individuals age, we see changes not only in the levels of specific molecules, but also in the way that these molecules interact with one another within intracellular networks (Soltow et al., 2010). The metabolomic profiles we have described here should allow us to examine age-specific changes in network structure, and in so doing, to generate novel hypotheses regarding pathways that are robust or frail in the face of the aging process. Finally, and as we mentioned earlier, we must now make it an urgent priority to curate the metabolomes of all model organisms that are used in aging research. A recent curation of the yeast metabolome has been made available (Jewison et al., 2012). To this we now need to add the worm and the fly.

ACKNOWLEDGEMENTS

This work was supported by grants NIH T32 GM007103 (JH), NIH R01AG038756 (DJ and DP), NSF 1021720 (DP) and an AFAR/EMF Breakthroughs In Gerontology grant (DP). We are grateful for helpful comments from Yousin Suh and members of the ISEM (Institut des Sciences de l'Evolution—Montpellier).

LITERATURE CITED

Barzilai N, Gabriely I, Gabriely M, Iankowitz N , Sorkin JD (2001). Offspring of centenarians have a favorable lipid profile. *J Am Geriatr Soc*. **49**, 76-79.

Benjamini Y , Hochberg Y (1995). Controlling the false discovery rate - A practical and powerful approach to multiple testing. *Journal Of The Royal Statistical Society Series B-Methodological*. **57**, 289-300.

Bo TH, Dysvik B , Jonassen I (2004). LSimpute: accurate estimation of missing values in microarray data with least squares methods. *Nucleic Acids Res*. **32**, e34.

Burger JM , Promislow DEL (2004). Sex-specific effects of interventions that extend fly life span. *Sci Aging Knowledge Environ*. **2004**, pe30.

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P , Karp PD (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*. **40**, D742-753.

Chintapalli VR, Al Bratty M, Korzekwa D, Watson DG , Dow JA (2013). Mapping an Atlas of Tissue-Specific *Drosophila* melanogaster Metabolomes by High Resolution Mass Spectrometry. *PLoS One*. **8**, e78066.

Coburn C, Allman E, Mahanti P, Benedetto A, Cabreiro F, Pincus Z, Matthijssens F, Araiz C, Mandel A, Vlachos M, Edwards SA, Fischer G, Davidson A, Pryor RE, Stevens A, Slack FJ, Tavernarakis N, Braeckman BP, Schroeder FC, Nehrke K , Gems D (2013). Anthranilate fluorescence marks a calcium-propagated necrotic wave that promotes organismal death in *C. elegans*. *PLoS Biol*. **11**, e1001613.

Corby-Harris V, Pontaroli AC, Shimkets LJ, Bennetzen JL, Habel KE , Promislow DE (2007). Geographical distribution and diversity of bacteria associated with natural populations of *Drosophila* melanogaster. *Appl Environ Microbiol*. **73**, 3470-3479.

De Guzman JM, Ku G, Fahey R, Youm YH, Kass I, Dixit VD , Kheterpal I (2012). Chronic caloric restriction partially protects against age-related alteration in serum metabolome. *Age*. **35**, 1091-1104.

De Luca M, Roshina NV, Geiger-Thornsberry GL, Lyman RF, Pasyukova EG , Mackay TF (2003). Dopa decarboxylase (Ddc) affects variation in *Drosophila* longevity. *Nat Genet*. **34**, 429-433.

Fuchs S, Bundy JG, Davies SK, Viney JM, Swire JS , Leroi AM (2010). A metabolic signature of long life in *Caenorhabditis elegans*. *BMC Biol*. **8**, 14.

Gomez LA, Heath SH , Hagen TM (2012). Acetyl-L-carnitine supplementation reverses the age-related decline in carnitine palmitoyltransferase 1 (CPT1) activity in interfibrillar mitochondria without changing the L-carnitine content in the rat heart. *Mechanisms of ageing and development*. **133**, 99-106.

Grandison RC, Piper MD , Partridge L (2009). Amino-acid imbalance explains extension of lifespan by dietary restriction in *Drosophila*. *Nature*. **462**, 1061-1064.

Hoffmann A , Parsons P (1988). The analysis of quantitative variation in natural populations with isofemale strains. *Genetique, selection, evolution*. **20**, 87-98.

Houtkooper RH, Argmann C, Houten SM, Canto C, Jeninga EH, Andreux PA, Thomas C, Doenlen R, Schoonjans K , Auwerx J (2011). The metabolic footprint of aging in mice. *Scientific reports*. **1**, 134.

Jeck WR, Siebold AP , Sharpless NE (2012). Review: a meta-analysis of GWAS and age-associated diseases. *Aging cell*. **11**, 727-731.

Jewison T, Knox C, Neveu V, Djoumbou Y, Guo AC, Lee J, Liu P, Mandal R, Krishnamurthy R, Sinelnikov I, Wilson M , Wishart DS (2012). YMDB: the Yeast Metabolome Database. *Nucleic Acids Res*. **40**, D815-820.

Jones DP, Park Y , Ziegler TR (2012). Nutritional metabolomics: progress in addressing complexity in diet and health. *Annu Rev Nutr*. **32**, 183-202.

Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikainen LP, Kangas AJ, Soininen P, Wurtz P, Silander K, Dick DM, Rose RJ, Savolainen MJ, Viikari J, Kahonen M, Lehtimaki T, Pietilainen KH, Inouye M, McCarthy MI, Jula A, Eriksson J, Raitakari OT, Salomaa V, Kaprio J, Jarvelin MR, Peltonen L, Perola M, Freimer NB, Ala-Korpela M, Palotie A , Ripatti S (2012). Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet*. **44**, 269-276.

Kirkwood TB, Feder M, Finch CE, Franceschi C, Globerson A, Klingenberg CP, LaMarco K, Omholt S , Westendorp RG (2005). What accounts for the wide variation in life span

of genetically identical organisms reared in a constant environment? *Mechanisms of ageing and development*. **126**, 439-443.

Kostal V, Korbelova J, Rozsypal J, Zahradnickova H, Cimlova J, Tomcala A , Simek P (2011). Long-term cold acclimation extends survival time at 0 degrees C and modifies the metabolomic profiles of the larvae of the fruit fly *Drosophila melanogaster*. *PLoS One*. **6**, e25025.

Li S, Park Y, Duraisingham S, Strobel FH, Khan N, Soltow QA, Jones DP , Pulendran B (2013). Predicting network activity from high throughput metabolomics. *PLoS computational biology*. **9**, e1003123.

Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF, Anholt RR, Barron M, Bess C, Blankenburg KP, Carbone MA, Castellano D, Chaboub L, Duncan L, Harris Z, Javaid M, Jayaseelan JC, Jhangiani SN, Jordan KW, Lara F, Lawrence F, Lee SL, Librado P, Linheiro RS, Lyman RF, Mackey AJ, Munidasa M, Muzny DM, Nazareth L, Newsham I, Perales L, Pu LL, Qu C, Ramia M, Reid JG, Rollmann SM, Rozas J, Saada N, Turlapati L, Worley KC, Wu YQ, Yamamoto A, Zhu Y, Bergman CM, Thornton KR, Mittelman D , Gibbs RA (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature*. **482**, 173-178.

Mishur RJ , Rea SL (2012). Applications of mass spectrometry to metabolomics and metabonomics: Detection of biomarkers of aging and of age-related diseases. *Mass spectrometry reviews*. **31**, 70-95.

Noland RC, Koves TR, Seiler SE, Lum H, Lust RM, Ilkayeva O, Stevens RD, Hegardt FG , Muoio DM (2009). Carnitine insufficiency caused by aging and overnutrition

compromises mitochondrial performance and metabolic control. *J Biol Chem*. **284**, 22840-22852.

Overgaard J, Tomcala A, Sorensen JG, Holmstrup M, Krogh PH, Simek P , Kostal V (2008). Effects of acclimation temperature on thermal tolerance and membrane phospholipid composition in the fruit fly *Drosophila melanogaster*. *J Insect Physiol*. **54**, 619-629.

Oxenkrug GF, Navrotskaya V, Voroboyva L , Summergrad P (2011). Extension of life span of *Drosophila melanogaster* by the inhibitors of tryptophan-kynurenine metabolism. *Fly*. **5**, 307-309.

Parisi M, Li R , Oliver B (2011). Lipid profiles of female and male *Drosophila*. *BMC Res Notes*. **4**, 198.

Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ , Chatterjee N (2010). Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet*. **42**, 570-575.

Pletcher SD, Macdonald SJ, Marguerie R, Certa U, Stearns SC, Goldstein DB , Partridge L (2002). Genome-wide transcript profiles in aging and calorically restricted *Drosophila melanogaster*. *Curr Biol*. **12**, 712-723.

Psihogios NG, Gazi IF, Elisaf MS, Seferiadis KI , Bairaktari ET (2008). Gender-related and age-related urinalysis of healthy subjects by NMR-based metabonomics. *NMR Biomed*. **21**, 195-207.

Rhee EP, Ho JE, Chen MH, Shen D, Cheng S, Larson MG, Ghorbani A, Shi X, Helenius IT, O'Donnell CJ, Souza AL, Deik A, Pierce KA, Bullock K, Walford GA, Vasan RS, Florez

JC, Clish C, Yeh JR, Wang TJ , Gerszten RE (2013). A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab*. **18**, 130-143.

Sarup P, Pedersen SM, Nielsen NC, Malmendal A , Loeschcke V (2012). The metabolic profile of long-lived *Drosophila melanogaster*. *PLoS One*. **7**, e47461.

Scheitz CJ, Guo Y, Early AM, Harshman LG , Clark AG (2013). Heritability and inter-population differences in lipid profiles of *Drosophila melanogaster*. *PLoS One*. **8**, e72726.

Slupsky CM, Rankin KN, Wagner J, Fu H, Chang D, Weljie AM, Saude EJ, Lix B, Adamko DJ, Shah S, Greiner R, Sykes BD , Marrie TJ (2007). Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. *Anal Chem*. **79**, 6995-7004.

Soltow QA, Jones DP , Promislow DE (2010). A network perspective on metabolism and aging. *Integr Comp Biol*. **50**, 844-854.

Soltow QA, Strobel FH, Mansfield KG, Wachtman L, Park Y , Jones DP (2011). High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. *Metabolomics*. **9**, 132-143.

Suhre K , Gieger C (2012). Genetic variation in metabolic phenotypes: study designs and applications. *Nat Rev Genet*. **13**, 759-769.

Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, Altmaier E, CardioGram, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmuller G, Kottgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes HW, Milburn MV, Prehn C, Raffler J, Ried JS, Romisch-Margl W, Samani NJ, Small KS, Wichmann HE, Zhai G, Illig T, Spector TD, Adamski J, Soranzo N , Gieger C (2011). Human

metabolic individuality in biomedical and pharmaceutical research. *Nature*. **477**, 54-60.

Toivonen JM, Gems D , Partridge L (2009). Longevity of Indy mutant *Drosophila* not attributable to Indy mutation. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, E53; author reply E54.

Uppal K, Soltow QA, Strobel FH, Pittard WS, Gernert KM, Yu T , Jones DP (2013). xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics*. **14**, Article 15.

Vermeulen CJ, Cremers TI, Westerink BH, Van De Zande L , Bijlsma R (2006). Changes in dopamine levels and locomotor activity in response to selection on virgin lifespan in *Drosophila melanogaster*. *Mechanisms of ageing and development*. **127**, 610-617.

Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, van Duijnhoven JPM , van Dorsten FA (2008). Assessment of PLSDA cross validation. *Metabolomics*. **4**, 81-89.

Wijeyesekera A, Selman C, Barton RH, Holmes E, Nicholson JK , Withers DJ (2012). Metabotyping of long-lived mice using 1H NMR spectroscopy. *J Proteome Res*. **11**, 2224-2235.

Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, Romisch-Margl W, Lattka E, Gieger C, Soranzo N, Heinrich J, Standl M, Thiering E, Mittelstrass K, Wichmann HE, Peters A, Suhre K, Li Y, Adamski J, Spector TD, Illig T , Wang-Sattler R (2012). Human serum metabolic profiles are age dependent. *Aging cell*. **11**, 960-967.

FIGURE LEGENDS

Figure 2.1. To complete the map from genotype-phenotype (G-P), we face numerous hurdles. First, the genome consists of complex epistatic networks of genes. Second, each gene can have pleiotropic effects on multiple traits. Third, the causal path between a gene and its downstream phenotype is far from direct, including effects on the transcriptome, proteome and metabolome, and influenced by environmental effects. This figure suggests that metabolome can provide a valuable intermediate stage to complete the G-P map, identifying genetic and environmental factors that influence metabolomic profiles, and then correlation metabolomic profiles with phenotypes of interest.

Figure 2.2. Specific metabolites associated with age (A, C), sex (D) and their interaction (B) ($P < 4 \times 10^{-11}$ in all cases). Blue dots indicate males and red dots indicate females; the value on the age-axis for (B) and (D) is shifted between the sexes for illustrative purposes only. Confirmed metabolite identities include (A) Oleoylcarnitine; (B) 1−Oleoylglycerophosphoethanolamine; (C) Stearoylcarnitine; and (D) Glutamine.

Figure 2.3. First component value from partial least squares discriminant analysis showing strong separation of males and females. (A) AE column (minimum classification error rate (CER) based on 10-fold validation = 0.017, using four metabolites. After permuting the class variable (sex), CER = 0.385 ± 0.008 [mean ± standard error, *n* = 10 permutations]). (B) C18 column (minimum CER = 0.021 based on 25 metabolites. After permutation, CER = 0.371 ± 0.015 [mean ± s.e., *n* = 10 permutations]).

Figure 2.4. Predicted versus observed age, for C18 males (A), AE males (B), C18 females (C) and AE females (D). In each case, two thirds of all samples were chosen at random to construct a model using partial least squares regression. This model was then used to predict values for the remaining one-third of the samples. Predictions were repeated twenty times in each case, and the figures shown here represent cases close to the mean R^2 value. The red line represents a 2^{nd}-order polynomial fit to the data, and the dashed grey line is the isometric line. These results are consistent with R^2 values obtained from 10-fold cross validation scores based on sparse PLS on the full data set (R^2 values: AE males: 0.92; AE females 0.83; C18 males: 0.88; C18 females: 0.78). In all four cases, R^2 values for permuted data sets were less than 0.05.

Figure 2.5. The network shown here represents output from *mummichog* analysis with color hue determined by the sign and size and color intensity determined by the magnitude of the regression coefficient in the age model (blue is positive, red is negative). The metabolites are putatively annotated based on *m/z* ratio. This particular module is enriched for metabolites associated with glycolysis, for metabolites that feed the glycolytic pathway, and for metabolites associated with glycophospholipid metabolism ($P = 0.04$, see text for details).

Figure S2.1. Histogram of intraclass correlations among all metabolites for AE (top) and C18 (bottom) columns.

Figure S2.2. Sparse partial least squares discriminant analysis is able to distinguish samples of different ages, where age is treated as a factor rather than a continuous variable as in Figure 2.4, with relatively high accuracy.

Figure S2.3. Carnitine shuttle pathway from *mummichog* analysis of metabolites, with color and intensity determined by the sign and magnitude of the regression coefficient in the age model (blue is negative, red is positive).

Table 2.1. Number and percentage of metabolites associated with age, sex, genotype and their interactions for AE and C18 columns.

Parameter	# of AE metabolites	% of AE total	# of C18 metabolites	% of C18 total
Increase with Age	76	2.5%	36	1.0%
Decrease with Age	96	3.1%	167	4.5%
Increase in Males	267	8.6%	133	3.6%
Increase in Females	381	12.3%	431	11.6%
Genotype	423	13.7%	292	7.9%
Age×Sex	71	2.3%	167	4.5%
Age×Genotype	137	4.4%	542	14.6%
Sex×Genotype	26	0.8%	15	0.4%

Table 2.2. List of specific amino acids and neurotransmitters associated with different age parameters. Age ↑ and Age ↓ refer to metabolites that either increase or decrease with age, respectively. A × S and A × G refer to metabolites with significant interaction effects between age and sex or age and genotype, respectively. Effects are all significant after correction for multiple comparisons with a False Discover Rate of 0.01 (see text). Note that isoleucine and leucine co-elute under our conditions and have identical mass-charge ratios so cannot be distinguished.

Metabolite	Effect
Amino Acids	
Tryptophan	Age ↑, Age ↓, A × G
Isoleucine/Leucine	Age ↑, Age ↓, A × G
Arginine	Age ↑
Phenylalanine	Age ↑
Methionine	A × S
Proline	A × S, A × G
Threonine	A × S, A × G
Aspartate	A × G
Glutamine	A × G
Neurotransmitters	
N-acetyl-serotonin	Age ↓
4-alpha-Hydroxytetrahydrobiopterin	Age ↑, A × S
5-hydroxy-L-tryptophan	Age ↑, A × S
Tetrahydrobiopterin	A × G
L-Dopa	Age ↓, A × S
Dopamine	Age ↑
GABA	Age ↑, Age ↓

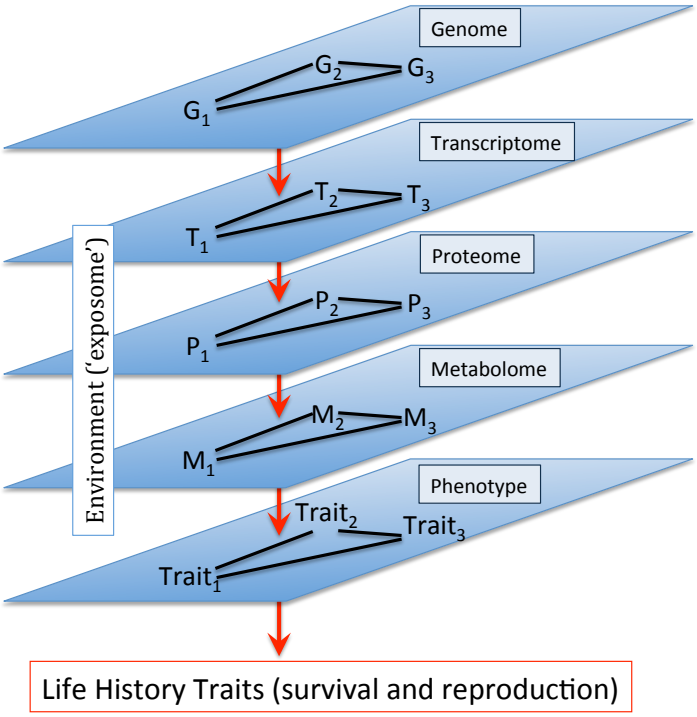
Genome
G1
G2
G3
Transcriptome
T1
T2
T3
Proteome
P1
P2
P3
Metabolome
M1
M2
M3
Phenotype
Trait1
Trait2
Trait3
Environment ('exposome')
Life History Traits (survival and reproduction)

Figure 2.1

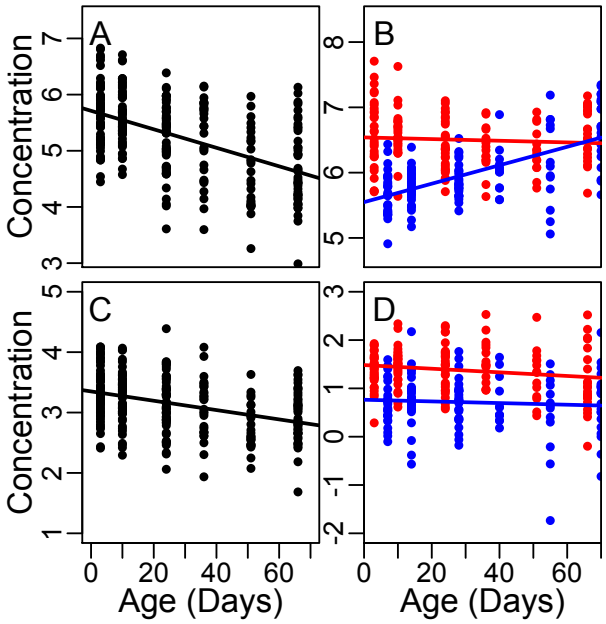
A
B
C
D
Concentration
Concentration
Age (Days)
Age (Days)
0
20
40
60
3
4
5
6
7
8
1
2
-2
-1
0
1
2
3

Figure 2.2

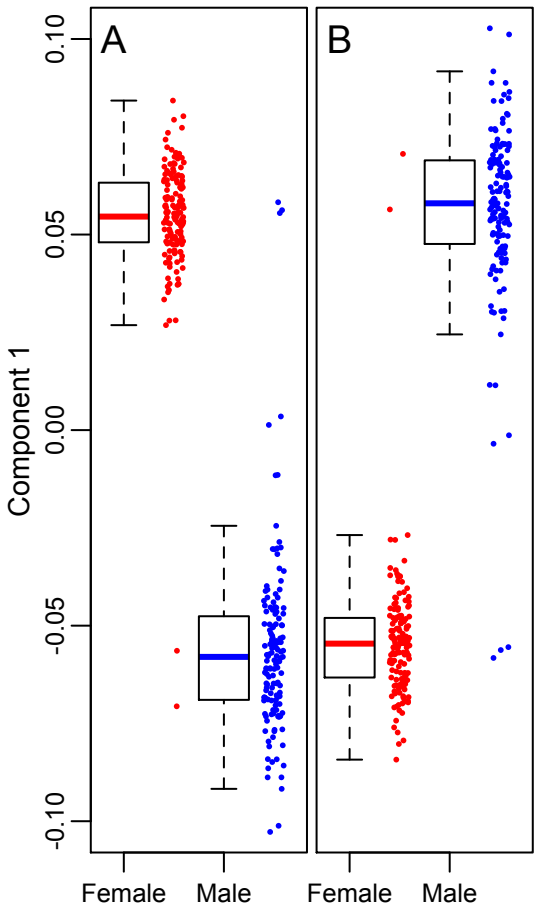

Figure 2.3

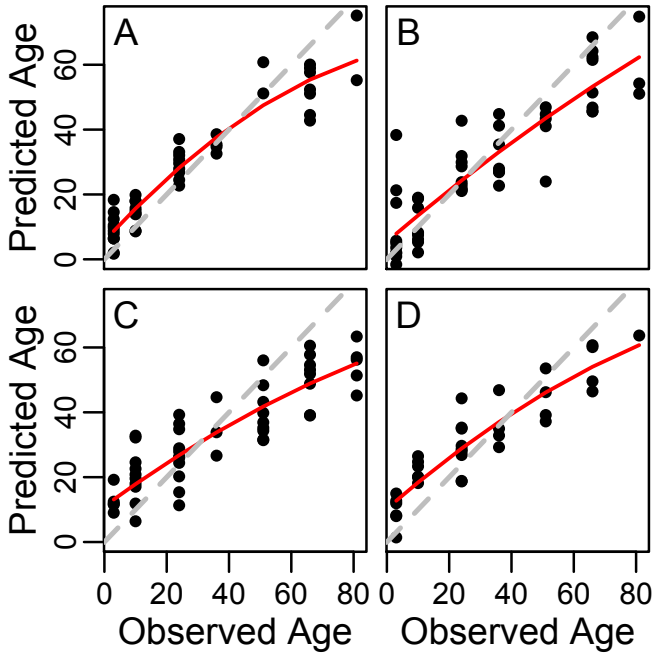
A
B
C
D
Predicted Age
Predicted Age
0
20
40
60
0
20
40
60
80
Observed Age
Observed Age

Figure 2.4

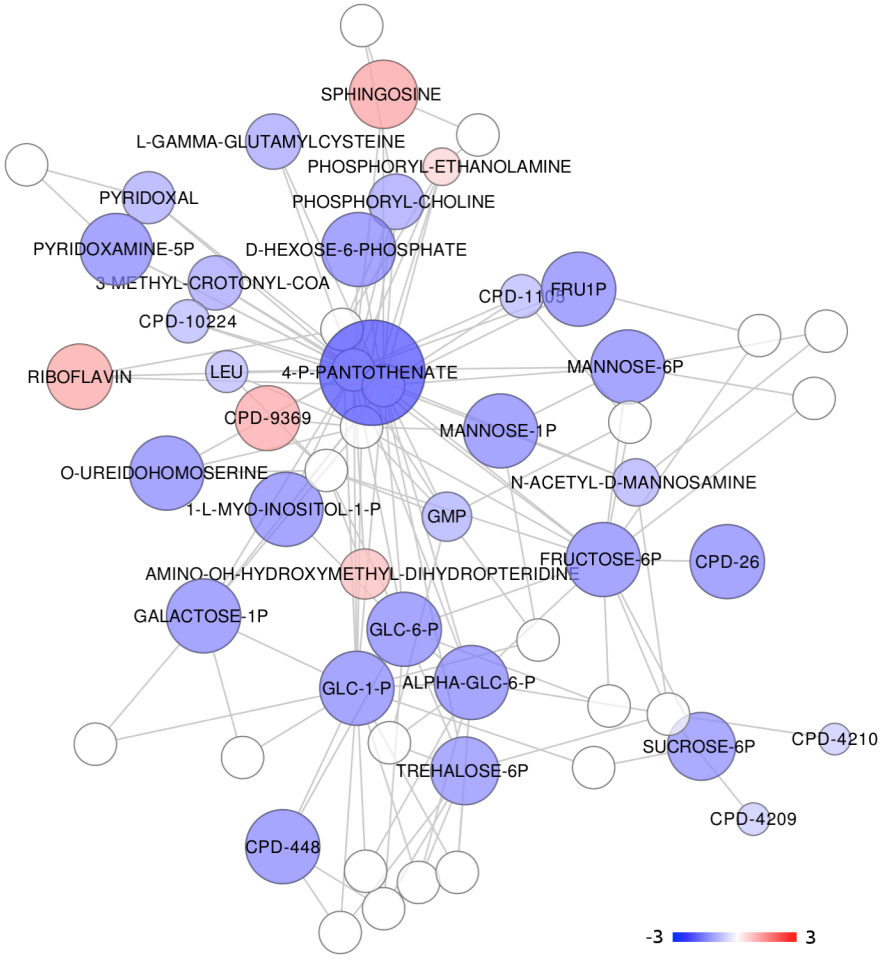
SPHINGOSINE
L-GAMMA-GLUTAMYLCYSTEINE
PHOSPHORYL-ETHANOLAMINE
PYRIDOXAL
PHOSPHORYL-CHOLINE
PYRIDOXAMINE-5P
D-HEXOSE-6-PHOSPHATE
3-METHYL-CROTONYL-COA
CPD-11103
FRU1P
CPD-10224
RIBOFLAVIN
LEU
4-P-PANTOTHENATE
MANNOSE-6P
CPD-9369
MANNOSE-1P
O-UREIDOHOMOSERINE
N-ACETYL-D-MANNOSAMINE
1-L-MYO-INOSITOL-1-P
GMP
FRUCTOSE-6P
CPD-26
AMINO-OH-HYDROXYMETHYL-DIHYDROPTERIDINE
GALACTOSE-1P
GLC-6-P
GLC-1-P
ALPHA-GLC-6-P
SUCROSE-6P
CPD-4210
TREHALOSE-6P
CPD-4209
CPD-448
-3
3

Figure 2.5

SUPPLEMENTARY INFORMATION

Table S2.1. Bloomington Stock Center numbers, DGRP numbers, and wolbachia status for those inbred lines analyzed in this study.

Bloomington	DGRP no.	Core40?	Wolbachia
25174	RAL-208	Core40	No
25181	RAL-315	Core40	No
25185	RAL-358	Core40	No
25186	RAL-360	Core40	Yes
25187	RAL-362	Core40	Yes
25188	RAL-375	Core40	No
25190	RAL-380	Core40	Yes
25192	RAL-399	Core40	No
25197	RAL-517	Core40	No
25198	RAL-555	Core40	Yes
25201	RAL-712	Core40	Yes
25204	RAL-765	Core40	No
25206	RAL-786	Core40	Yes
25210	RAL-859	Core40	Yes
25745	RAL-714	Core40	No

Table S2.2. *Mummichog* enrichment analysis for sets of analytes significantly associated with specific factors. For sake of clarity we only include pathways that are enriched with at least three analytes.

Column	Significant predictor	Pathway	Obs.	Ref.	P-value	Adjusted Pvalue
C18	Age genotype interaction	Zymosterol Biosynthesis	4	12	0.0257	0.0008
C18	Age genotype interaction	Purine deoxyribnucleoside degradation	3	6	0.016	0.0008
AE	Age genotype interaction	arginine biosynthesis IV	4	12	0.1088	0.0054
AE	Age genotype interaction	citrulline-nitric oxide cycle	3	7	0.085	0.0059
AE	Age genotype interaction	isoleucine biosynthesis from threonine	3	8	0.1206	0.0086
AE	Age genotype interaction	tRNA charging pathway	5	20	0.2016	0.0099
AE	Age genotype interaction	uridine-5'-phosphate biosynthesis	3	10	0.2039	0.0175
AE	Down with age	Isoleucine degradation I	4	8	0.0331	0.0009
AE	Down with age	Purine Deoxyribonucleosides degradation	4	8	0.0331	0.0009
AE	Down with age	2-methylbutyrate biosynthesis	3	5	0.0379	0.0014
AE	Down with age	Beta-carboline biosynthesis	3	5	0.0379	0.0014
AE	Down with age	Dopamine Degradation	4	10	0.075	0.0016
AE	Down with age	Serotonin and melatonin biosynthesis	3	6	0.0663	0.0021
AE	Down with age	Coenzyme A biosynthesis	3	6	0.0663	0.0021
AE	Down with age	Leucine degradation I	4	11	0.1026	0.0022
AE	Down with age	L-carnitine biosynthesis	3	7	0.1018	0.0031
AE	Down with age	Glycolysis II	3	10	0.2374	0.0103
AE	Down with age	Glycolysis III (Thermotoga)	3	11	0.2879	0.0147
AE	Down with age	Glycolysis I	3	11	0.2879	0.0147
AE	Down with age	5-aminoimidazole ribonucleotide biosynthesis I	3	11	0.2879	0.0147
AE	Down with age	FormylTHF biosynthesis I	3	12	0.3392	0.0205
AE	Down with age	Gluconeogenesis	3	13	0.3902	0.0278
AE	Down with age	tRNA Charging pathway	4	20	0.457	0.0301
C18	Down with age	Dolichyl-diphosphooligosaccharide biosynthesis	4	4	0.0009	0.0007
C18	Down with age	Zymosterol Biosynthesis	6	12	0.0092	0.0008
C18	Down with age	Purine Deoxyribonucleosides degradation	3	6	0.0685	0.0026
C18	Down with age	Acetylneuraminate biosynthesis I	3	7	0.105	0.0037
C18	Down with age	Dopamine Degradation	3	8	0.1473	0.0053
AE	Increase in Females	Catecholamine biosynthesis	5	9	0.0181	0.0006
AE	Increase in Females	Choline biosynthesis III	3	4	0.026	0.0011
AE	Increase in Females	Betaxanthin biosynthesis (via dopamine)	3	4	0.026	0.0011

AE	Increase in Females	Phosphatidylcholine biosynthesis	3	6	0.0953	0.003
AE	Increase in Females	4-hydroxyproline degradation I	3	7	0.1432	0.0049
AE	Increase in Females	Isoleucine biosynthesis from threonine	3	8	0.1971	0.0077
AE	Increase in Females	Salvage pathways of pyrimidine ribonucleotides	3	8	0.1971	0.0077
C18	Increase in Females	Dolichyl-diphosphooligosaccharide biosynthesis	4	4	0.0003	0.0021
C18	Increase in Females	Zymosterol Biosynthesis	5	12	0.0127	0.0025
C18	Increase in Females	Acetylneuraminate biosynthesis I	3	7	0.0507	0.0046
C18	Increase in Females	Dopamine Degradation	3	8	0.0736	0.0055
AE	Increase in Males	Arginine degradation X (arginine monooxygenase pathway)	3	4	0.0152	0.0003
AE	Increase in Males	Arginine degradation VI (arginase 2 pathway)	3	7	0.091	0.0011
AE	Increase in Males	Catecholamine biosynthesis	3	9	0.1706	0.0024
AE	Increase in Males	tRNA Charging pathway	4	20	0.4196	0.0099
C18	Increase in Males	L-carnitine biosynthesis	3	6	0.032	0.0065
AE	Negative age sex interaction	Betaxanthin biosynthesis (via dopamine)	3	4	0.0492	0.001
AE	Negative age sex interaction	tRNA Charging pathway	7	20	0.2052	0.0021
AE	Negative age sex interaction	Catecholamine biosynthesis	4	9	0.1606	0.0024
AE	Negative age sex interaction	Lipoate biosynthesis and incorporation II	3	6	0.1653	0.0037
AE	Negative age sex interaction	Serotonin and melatonin biosynthesis	3	6	0.1653	0.0037
AE	Negative age sex interaction	5-aminoimidazole ribonucleotide biosynthesis I	4	11	0.2799	0.0061
AE	Negative age sex interaction	Isoleucine biosynthesis from threonine	3	8	0.3157	0.0111
AE	Negative age sex interaction	Purine Deoxyribonucleosides degradation	3	8	0.3157	0.0111
AE	Negative age sex interaction	Folate Transformations	3	10	0.4682	0.027
AE	Positive age sex interaction	NAD biosynthesis from 2-amino-3-carboxymuconate semialdehyde	3	8	0.0458	0.0004
C18	Positive age sex interaction	Dolichyl-diphosphooligosaccharide biosynthesis	4	4	0.0003	0.0003
C18	Positive age sex interaction	Zymosterol Biosynthesis	5	12	0.0127	0.0004
C18	Positive age sex interaction	Acetylneuraminate biosynthesis I	3	7	0.0507	0.0013
C18	Positive age sex interaction	Dopamine Degradation	3	8	0.0736	0.0019
AE	Up with age	Serotonin and melatonin biosynthesis	3	6	0.0901	0.0038
AE	Up with age	tRNA Charging pathway	6	20	0.1719	0.0041
AE	Up with age	Isoleucine biosynthesis from threonine	3	8	0.1876	0.0109
AE	Up with age	Isoleucine degradation I	3	8	0.1876	0.0109
AE	Up with age	uridine-5'-phosphate biosynthesis	3	10	0.3014	0.0274
AE	Up with age	Leucine degradation I	3	11	0.36	0.0408
AE	Up with age	5-aminoimidazole ribonucleotide biosynthesis I	3	11	0.36	0.0408

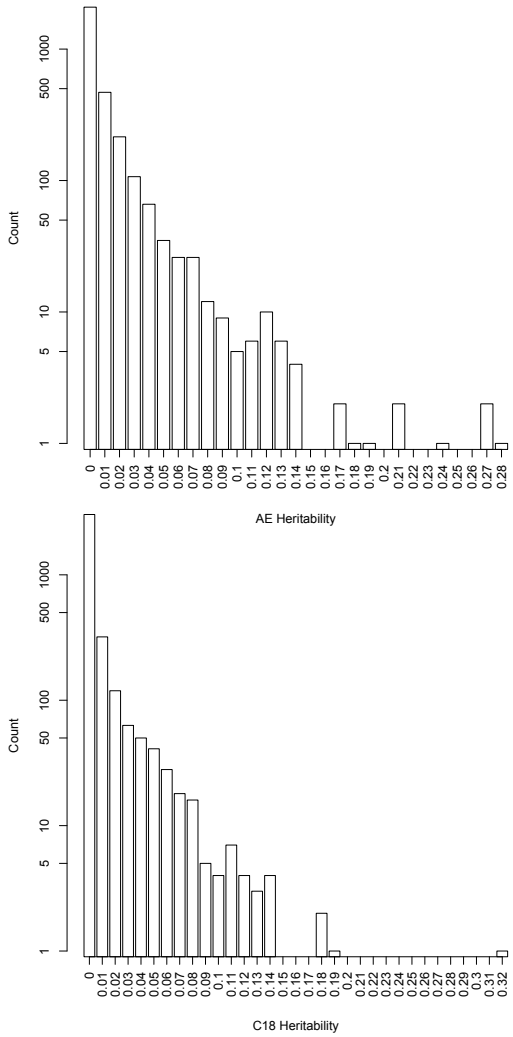
Count
1
5
10
50
100
500
1000
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
0.11
0.12
0.13
0.14
0.15
0.16
0.17
0.18
0.19
0.2
0.21
0.22
0.23
0.24
0.25
0.26
0.27
0.28
AE Heritability
Count
1
5
10
50
100
500
1000
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
0.11
0.12
0.13
0.14
0.15
0.16
0.17
0.18
0.19
0.2
0.21
0.22
0.23
0.24
0.25
0.26
0.27
0.28
0.3
0.31
0.32
C18 Heritability

Figure S2.1

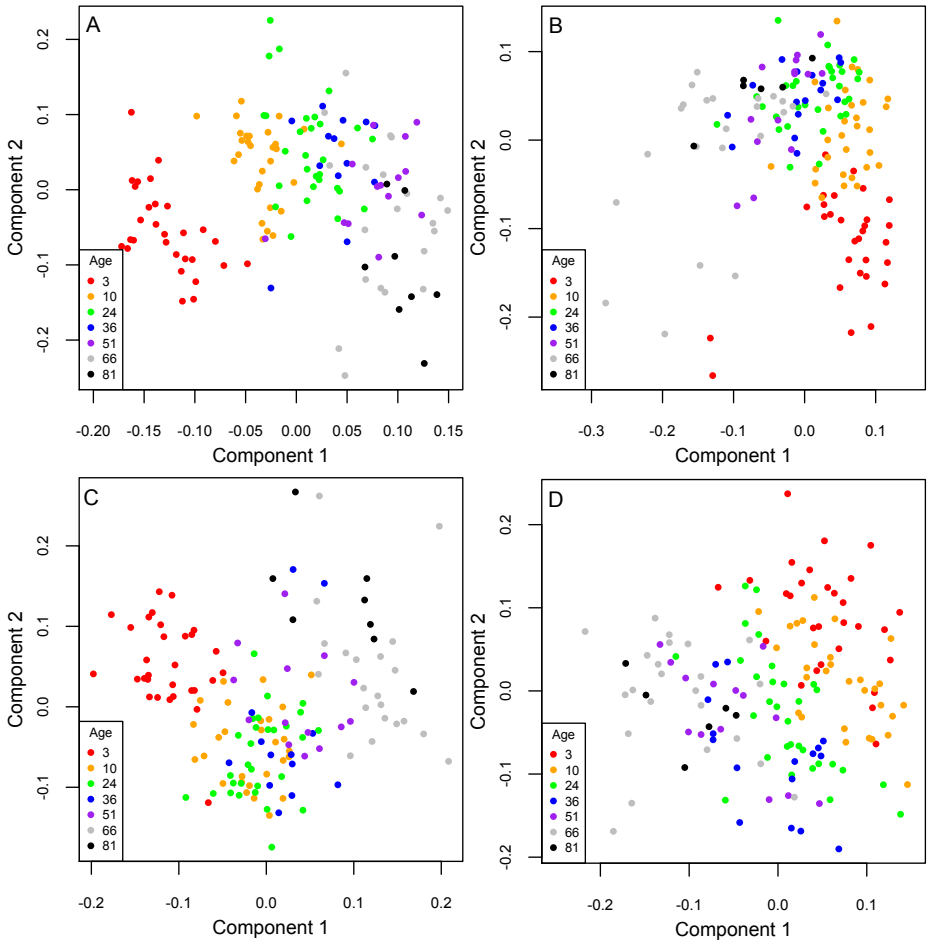

Figure S2.2

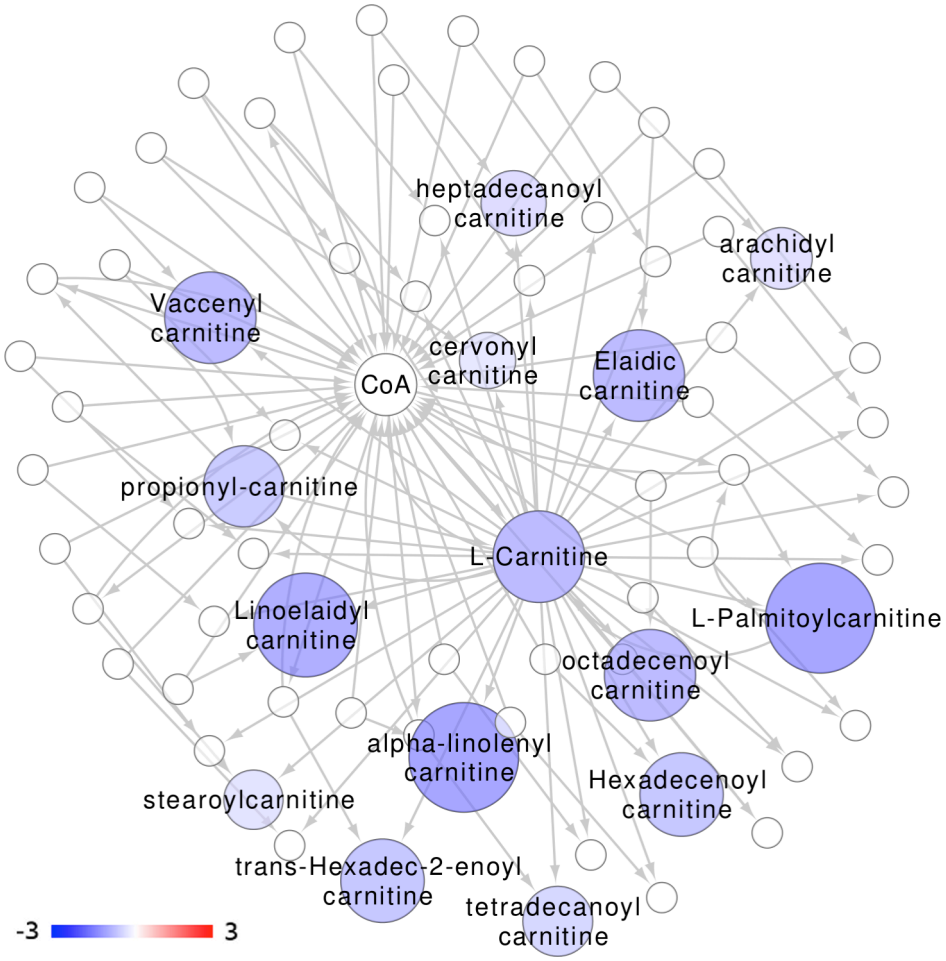
heptadecanoyl carnitine
arachidyl carnitine
Vaccenyl carnitine
cervonyl carnitine
Elaidic carnitine
CoA
propionyl-carnitine
L-Carnitine
Linoelaidyl carnitine
L-Palmitoylcarnitine
octadecenoyl carnitine
alpha-linolenyl carnitine
Hexadecenoyl carnitine
stearoylcarnitine
trans-Hexadec-2-enoyl carnitine
tetradecanoyl carnitine
-3
3

Figure S2.3

CHAPTER 3

A COMPARATIVE ANALYSIS OF THE *DROSOPHILA* METABOLOME[2]

[2]Hoffman J.M., Samuelson A., and Promislow D.E.L. To be submitted to *Evolution.*

ABSTRACT

Comparative studies have been often been utilized to understand how life history traits evolve, and recently, comparative "omics" have come to the forefront as a powerful way to determine the molecular causes and consequences of adaptation and evolution. However, while comparative genomics and transcriptomics have been heavily studied, metabolomics, the study of all the small molecules in an organism, has been largely ignored, and the role of metabolites in shaping life history evolution is unknown. To fill this gap, we studied variation in the metabolome across eleven *Drosophila* species. Further, we determined the associations between the metabolome and both longevity and resistance to oxidative stress. We first discovered that individual metabolites did not contain phylogenetic signal, but the entire metabolome taken together showed some concordance between the known *Drosophila* phylogeny and one estimated from the metabolome alone. We then found that a large proportion of the metabolome is correlated with sex, species, and body weight, and many metabolites are associated with longevity in a sex specific manner. Finally, we found that amino acids and TCA cycle metabolites are enriched among all metabolites associated with longevity. Here, we have shown the power of metabolomics to help us understand the molecular mechanisms that underlie variation across species in important fitness traits, including longevity and oxidative stress resistance.

INTRODUCTION

The comparative method has long been employed to address a wide range of biological questions, helping us to understand species' relatedness, how species adapt to new environments, and how and why phenotypic traits coevolve among groups of species (Harvey and Pagel, 1991). The earliest comparative studies focused on the measurement of individual phenotypic traits (e.g. Damuth, 1981; Harcourt et al., 1981), including many life history traits (e.g. Harvey et al., 1991; Promislow, 1991; Promislow and Harvey, 1990; Promislow and Haselkorn, 2002), and drew important conclusions that were often missed by studying a single species. Recently, "omics" methods have come to the forefront of comparative studies, with the goal of understanding adaptation at the molecular level. Comparative genomics has taught us largely about how genetic sequences evolve (e.g. Clark et al., 2007; Scally et al., 2012), as well as how transcriptomes change across organs and species (e.g. Brawand et al., 2011; Lu et al., 2008) However, little is known about the newer "omics" technologies, including the proteome, the microbiome, and the focus of this study, the metabolome.

Metabolomics, the study of all the small molecules in an organism, integrates both the genome (through proteins regulating metabolite flux) and the environment (both responses to the environment and molecules from external sources). As such it has the potential to inform us about how life history traits evolve at the molecular level by discovering how metabolic correlates associated with life history traits change across species. By identifying metabolic associations with life history traits across species, we can learn about the molecular mechanisms that underlie life history traits. A metabolomic perspective is imperative to completely understand how adaptation has affected species at

the biochemical level; however, the literature on comparative metabolomic studies is scarce.

While there is large potential for understanding the evolution of the metabolome by utilizing comparative studies, few have actually attempted to determine how the metabolome has evolved across taxa. One study in sponges discovered that metabolomic profiles could distinguish two distinct groups of species, already known to be different (Ivanisevic et al., 2011), though this study only looked at a small subset of metabolites. A study across seven mammal species determined the metabolome could not recapitulate the phylogeny of the species (Park et al., 2012), and surprisingly, principal component analysis suggested that humans are more similar to mice than to new world primates. At the same time, other studies have shown that the metabolome reflects genetic divergence across primate species and does not appear to be largely influenced by the environment (Bozek et al., 2014), but there may be signatures of adaptation to different diets within primates (Blekhman et al., 2014). Across the limited comparative metabolomics literature, we see conflicting results that suggest we still do not understand the effects of different evolutionary forces on the metabolome. These studies also did not examine how adaptation via life history trait evolution may be associated with changes in metabolomic profiles. Here, we suggest that studies on closely related *Drosophila* species might allow us to begin to grasp how the selection has shaped the metabolome and its connections to life history traits.

To identify coevolutionary patterns between life history traits and the metabolome, the fruit fly genus *Drosophila* offers a potentially powerful model. Flies are often thought of as a classic molecular genetic model with most studies focusing on one species, *Drosophila*

melanogaster. However, the *Drosophila* genus has also been studied in a comparative light to understand phenotypic adaptation. Early *Drosophila* comparative studies usually focused on very specific phenotypes (e.g. Capy et al., 1993; Coyne and Orr, 1989; Kambysel and Heed, 1971; Partridge et al., 1987; Schnebel and Grossfield, 1983). But recently, with the advent of whole genome sequencing, the *Drosophila* genus became the first to receive large genomic comparative analyses with the sequencing of 12 closely related species (Clark et al., 2007).

Since the sequencing of the 12 *Drosophila* genomes, the *Drosophila* genus has become a powerful model for comparative genomics, and researchers now understand the genetic basis of adaptation and speciation (e.g. Clark et al., 2007; Lu et al., 2008; Stark et al., 2007) to a degree not previously possible. With regards to metabolic evolution in *Drosophila*, researchers have found that rates of sequence evolution differ for enzymatic versus non-enzymatic proteins (Larracuente et al., 2008); as well as differences in evolution rates for essential versus non-essential proteins across the *Drosophila* phylogeny (Greenberg et al., 2008). However, the association of life history traits with metabolic pathways and individual metabolites is unknown.

To begin to fill this gap in our knowledge, we completed the largest comparative metabolomic study to date using eleven different *Drosophila* species. We assess the degree to which metabolomic variation in steady state levels is captured within species versus among species. We then ask if phylogenetic signatures exist within individual metabolites as well as the extent to which metabolomic profiles recapitulate the true phylogeny. Finally, we determine if individual metabolites and metabolic pathways are associated with life history variation among species.

METHODS

Species and lines

Eleven species of *Drosophila* with sequenced genomes were utilized for this project: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. ananassae, D. erecta*, *D. yakuba*, *D. willistoni*, *D. pseudoobscura*, *D. persimilis*, *D. mojavensis*, *and D. virilis*. All species except *D. mojavensis* were obtained from the UCSD Drosophila species stock center (https://stockcenter.ucsd.edu/). *D. mojavensis* lines were obtained from Luciano Matzkin at the University of Alabama at Huntsville. To enable us to distinguish species differences from genotype affects, we obtained three wild type lines from each species with the exception of *D. yakuba* and *D. erecta*, of which we were only able to obtain one wild type strain. Our final longevity analysis consisted of only two lines of *D. mojavensis*, *D. persimilis*, and *D. willistoni* due to insufficient collection of one genotype of each species.

Media

To control for the many effects diet has on the metabolome, all lines were reared on the same diet, consisting of banana, molasses, corn syrup, yeast, agar, methylparaben (to reduce fungal growth), ethanol, and water.

For the oxidative stress assay, we used three different types of media. Control medium consisted of agar, dextrose, propionic acid (to reduce fungal growth), and water. For the paraquat medium, we combined agar, dextrose, propionic acid, and water. Paraquat was then dissolved in water and added in replacement of some water to make a final 1.3 mM paraquat medium. Finally, for the H_2O_2 medium we combined agar, dextrose, propionic

acid, and water. We then used freshly opened hydrogen peroxide in place of some water such that the medium had a concentration of 2% H_2O_2.

Longevity Assay

Flies were placed in bottles on banana medium and allowed to mate and lay eggs for 48 hours. Adults were removed from bottles, and the eggs were allowed to develop at 24°C and approximately 60% humidity. Virgin males and females were collected under CO_2 anesthesia and placed into vials with banana food within 8 hours of eclosion, except *D. virilis*, which was collected within 12 hours of eclosion. Approximately 20 flies were placed in each vial, and collection took place over 48 hours. Total number of vials and flies analyzed per line are given in Table S1. Flies were transferred onto new food three times over a seven day period before the start of the longevity assay. On the seventh day, flies were transferred into vials randomized and scanned into the *Drosophila* longevity program dLife (Linford et al., 2013). dLife allows for the researcher to be blind to the vial identities, such that biases in calling of deaths are minimized. Flies were then transferred onto new food three times a week and dead flies were counted and recorded into dLife at each transfer. Recording in dLife continued until all flies in all vials were deceased.

Body weight analysis

Flies were collected simultaneously during the longevity collection for body weight analysis. When flies were five days old, they were frozen in groups of five from each sex and line with three replicate samples for each line by sex combination. Weights were

averaged over the three replicates and then divided by five to attain an average weight per fly (Table S3.1), giving us only a single weight value for each genotype.

Oxidative Stress Assay

Flies were reared as described for the longevity assay until eclosion, and after the collection of virgins, flies were held on banana food for two days after the last day of collection (48 hours of collection). Flies were placed on one of three treatments as described above: control food, 1.3mM paraquat, or 2% hydrogen peroxide.

Most lines consisted of one vial of control food and three vials each of paraquat and H_2O_2 food with each vial consisting of approximately 20 flies. Total numbers of flies analyzed for oxidative stress resistance are in Table S3.1. Flies were randomized into their respective food vial using dLife as in the longevity assay. After the first 24 hours, flies were then checked every 12 hours; at each checkpoint the number of dead flies were counted and recorded in dLife. After 5-7 days, flies were transferred onto new food to prevent the food from drying out. After 3 weeks, the longevity study was ended with those flies still alive being censored for final oxidative stress analysis.

Metabolomics Assay

Flies were reared and collected as virgins as described in the longevity assay. Flies were held on banana food and transferred every two days until they were five days old. They were then placed in 1.5 mL tubes in groups of 10 flies, and tubes were flash frozen in liquid nitrogen and placed in a -80°C freezer to be extracted.

Fly extractions began with homogenizing tissue samples in 200uL water using a tissuelyser (TissueLyser II, Qiagen) for 6 minutes at 25 Hz. Samples were then incubated for 30 minutes in 800uL methanol on dry ice. The suspension was then tissuelysed again for 10 minutes at 25Hz, and then spun at 13000 rpm for 10 minutes at 4°C. The supernatant was then removed and placed in a new tube for metabolomics analysis. These supernatants were then dried using a speedvac at 30°C. Metabolomics analysis was completed at the Raftery lab at the University of Washington using liquid chromatography coupled to quadrupole time of flight mass spectroscopy in negative ion mode.

Data analysis

All statistical analyses were completed in the program R (R Core Team, 2013) unless otherwise stated.

Lifespan: We first determined average, median, and maximum longevities by sex for each unique genotype. Results of the longevity experiment were compared across lines to determine differences that exist both within and between species. We used a Kaplan-Meier analysis to plot longevity curves for each sex and species and to determine if their survival curves were significantly different.

Oxidative stress analysis: First we determined correlations between resistance to paraquat and body weight, median longevity, and maximum longevity using a Pearson correlation. Then we repeated the analysis for hydrogen peroxide. This analysis allowed us to discover if the ability to resist oxidative stress was the same across different stressors.

Metabolomics analysis: Metabolomics data were first log-transformed to normalize data. We then removed all metabolites that were not present in at least 80% of samples.

We first determined if a phylogenetic signal was present in the metabolome using Pagel's K (Pagel, 1994), an estimate of phylogenetic signal for individual phenotypes. Using the known *Drosophila* species phylogeny, we used the `picante` package (Kembel et al., 2010) in R to discover Pagel's K for each individual metabolite, using species metabolite concetration mean as the trait value. We corrected for multiple comparisons using a false discovery rate (FDR) at $\alpha=0.05$ (Benjamini and Hochberg, 1995). We then determined if the entire metabolome could recapitulate the true species phylogeny. We first took the species' mean of each individual metabolite. Then using this matrix, computed a distance matrix based on Euclidian distance, and ran the resulting distance matrix through a neighbor-joining tree building algorithm (Saitou and Nei, 1987) in the `ape` package (Paradis et al., 2004). We then tested for concordance between the true species phylogeny and our metabolomic phylogeny using a permutation test (999 permutations) of Kendall's W (Kendall and Smith, 1939), which is a non-parametric test of concordance between two distance matrices. The null hypothesis is no concordance exists between the two matrices, and test statistics range from zero (no concordance) to one (completely concordant).

To discover the association between body weight, sex, and species on each individual metabolite, we ran an ANOVA for each metabolite including body weight, sex, species, and genotype to determine which metabolites were associated with sex and body weight. To correct for multiple comparisons, a factor was considered significant if it passed a FDR at $\alpha=0.05$. Samples with missing data were removed from this analysis.

Next, we sought to know how metabolite values were associated with longevity measures. We ran an analysis of variance with median longevity as the dependent variable with body weight, metabolite concentration, and species as fixed predictor effects. The

sexes were run separately with missing data removed from the analysis. Then we ran the same analysis except instead of median longevity we used maximum longevity, paraquat median longevity, or hydrogen peroxide median longevity. For the oxidative stress resistance models, we also included median longevity as a predictor. For all the ANOVAs described above we again used an FDR of 0.05.

For all the individual factors analyzed in the ANOVAs, we then ran metabolic enrichment analysis to determine if metabolites associated with a specific factor were enriched for specific metabolic pathways. This was completed using the enrichment program *mummichog* (Li et al., 2013), in which mass to charge ratios are used for annotation and enrichment. For this analysis, we utilized the *D. melanogaster* reference metabolic pathway. Pathways were considered to be significant if their adjusted p-value was less than 0.05 and if the number of metabolites found in a pathway was greater than two.

We then completed an unsupervised principal component analysis (PCA) to determine if specific components of the entire metabolomic profile were associated with sex, species, or median longevity. We used the `made4` package (Culhane et al., 2005) to get coordinates for the first 6 principal components, and we determined associations with body weight, sex, and species using a linear model. For this analysis, missing data were imputed using the java program LSimpute (Bo et al., 2004).

RESULTS

Longevity results

Final analysis consisted of 26 genotypes across 11 species. Measure of fly longevity for each sex and genotype are shown in Table S3.1. In nine species females had significantly greater longevity than males (Figure S3.1), and in one species, *D. ananassae*, males lived significantly longer. *D. erecta* showed no significant differences between the sexes , and *D. virilis* was the longest-lived species in all natural longevity measures with a maximum longevity of 170 days (Table S3.1, Figure 3.1). Body weight was significantly associated with median and maximum longevity in both males and females (Pearson correlation: $P<0.0001$ for both). We also found that both median and maximum longevity varied among species, and among genotypes within species (ANOVA: $P<0.0001$ for both).

While there were significant differences found in longevity between species and sexes (Figure 3.1), we also observed large differences within species (Table S3.1). We also found that coefficients of variations differed over two fold across species from a low of 24% in *D. simulans* females to a high of 61% in *D. erecta* females (Table 3.1).

Our oxidative stress analysis showed large variation between different stressors. Strong correlations were seen between median longevity and paraquat resistance (Figures 3.2 and 3.3); however, there was no significant correlation between hydrogen peroxide resistance and median longevity. Also, body weight was significantly correlated with paraquat resistance (Pearson correlation: $P=0.0005$) but not hydrogen peroxide resistance (Pearson correlation: $P=0.083$, Figures 3.2 and 3.3). Even though the correlations between natural longevity and paraquat and hydrogen peroxide were different, we found a

correlation between survival on paraquat and hydrogen peroxide in both sexes (Figures 3.2 and 3.3).

Metabolomics results

Our final metabolomic dataset consisted of 253 metabolites across 159 samples. Of these metabolites 89 had a known metabolite identity, 127 had a known chemical formula but no structural identity, and 36 had a mass to charge ratio only. Looking at the variance averages across each individual metabolite, we found the variance within a genotype was low on average (mean=0.042, min=0.0029, max=0.2834). A larger variance was seen within species (mean=0.059, min=0.0033, max=0.3433), and the largest variance was seen across all species (mean=0.08, min=0.0035, max=0.466).

First we looked for a phylogenetic signal in the metabolome. Overall, we find that for individual metabolites, there is little phylogenetic signal, and for no metabolites did Pagel's K reach statistical significance after controlling for multiple comparisons. We also failed to find a significant phylogenetic signal in median as well as maximum longevity (Pagel's K P-value >0.05). The entire metabolome also could not recapitulate the *Drosophila* phylogeny (Figure 3.4). However, it could correctly place together two sets of sister taxa. *D. mojavensis* and *D. virilis* as well as *D. pseudoobscura* and D. *persimilis.* Our statistical test of concordance between the true phylogeny and our metabolomic phylogeny returned a Kendall's W statistic of 0.599 with a p-value of 0.089. A W statistic of 0.599 suggests the metabolomic phylogeny shows some concordance with the true species phylogeny, but the non-significant p-value indicates the test was unable to reject the null hypothesis that there is no congruence between the two phylogenies.

Using analysis of variance, we discovered a large number of individual metabolites associated with sex and body weight (Table 3.2, Figure 3.5). When we analyzed the sexes separately, we found about 30% of the metabolome is associated with median longevity (Table 3.2, Figure 3.6 and 3.7); however there is less than 40% overlap in the metabolites associated with longevity in each sex. We also find large numbers of metabolites associated with maximum longevity and resistance to oxidative stress (Table 3.2).

We carried out enrichment analysis among metabolites that were significantly associated with sex, species, body mass, and longevity (Table 3.3). *Mummichog* was able to putatively identify 205 (81%) of all the metabolites analyzed. We were unable to determine any metabolic pathways that were significantly different among species, but we were able to discover two pathways associated with body weight and eight metabolic pathways enriched for sex differences (adjusted p-value<0.05 for all pathways). Using our sex-specific analyses, we were able to find pathways associated with median longevity (3 in female sand 4 in males), maximum longevity (4 in females), resistance to paraquat (1 in males), and resistance to hydrogen peroxide (3 in males) (Table 3.3). For many of the different longevity factors we find pathways associated with amino acids, arginine degradation, and the TCA cycle.

Our unsupervised clustering analysis (PCA) showed large effects of body weight, sex, and species on metabolomic profiles (Table S3.2). PC1 and PC2 explained 20% and 13%, respectively, of the variation seen in the metabolome, and PC1 and PC2 were both associated with body weight, sex, and species. Figures 3.8 and 3.9 show PCs associated with sex and species, respectively.

DISCUSSION

As was expected, our comparative longevity analysis found large differences between species in median and maximum longevity, and this is consistent with previous *Drosophila* studies (Durbin and Yoon, 1986; Schnebel and Grossfield, 1983). It is worth noting that *D. virilis* did live to a maximum of 170 days, which appears to be longer than any previously published longevity measure in the genus. Interestingly we also discovered that extensive variation exists within species with respect to longevity and response to stressors. Even though all our lines were wild-type derived, extreme differences were often seen within species. Most comparative studies only use one genotype of a species, and here we show this may not be valid for many species because one line may not represent the large amounts of variation seen within a species. This could potentially confound studies as researchers may actually be comparing genetic variation not interspecific variation. The results of this study suggest that all future comparative studies, especially of life history traits, need multiple genotypes per species to be analyzed

We also found the metabolome is highly associated with sex, species, body weight, and longevity. While 70% of the metabolites in the metabolome varied among species, only 21% varied among genotypes. This suggests most of the variation in the metabolome is contained among species more than within species. While this result is not unexpected, this is the first study to demonstrate differences in metabolomic variances within and among species.

Not surprisingly, we discovered sex is highly correlated with the metabolome, both through individual metabolites and our principal components analysis. Many previous studies have found large effects of sex on the metabolome in different species (e.g. Hoffman

et al., 2014; Lawton et al., 2008; Slupsky et al., 2007), and our enrichment analysis suggests sex differences in the metabolome are reflected in large part by variation in amino acids including arginine, proline, serine, and alanine. Proline metabolic pathways made up 25% of the enriched sex pathways in our analysis. Very little research appears to be done on the sex differences in proline, but one previous study found proline increases with age in females but not males (Kouchiwa et al., 2012). This is consistent with our results, where metabolites found in our *mummichog* analysis in the proline pathways had significantly higher concentrations in females compared to males.

We discovered that almost 30% of the metabolome is correlated with longevity in each sex (Table 3.2). Of the metabolites associated with longevity, less than 40% of the metabolites are the same across the two sexes, suggesting the metabolic influences on aging and longevity are potentially different across the two sexes. Similar patterns are seen in our maximum longevity and resistance to oxidative stress analysis with many different metabolites seen associated in the two sexes. This is consistent with previous work in *D. melanogaster* that showed large age-by-sex interactions (Hoffman et al., 2014), and these results suggest metabolomics may have the power to understand why we often seen large differences in aging and longevity between the two sexes.

Our enrichment analyses on longevity found many metabolic pathways associated with longevity and resistance to oxidative stress. We found large effects of biosynthesis and degradation of different amino acids, especially proline, aspartate, and arginine. These results indicate that both the biosynthesis and degradation of different amino acids may be affecting longevity and oxidative stress resistance in *Drosophila.* Future studies of

manipulations of these metabolites are necessary to understand to what extent they are directly influencing longevity.

The TCA-cycle also showed up regularly in our enriched longevity pathways. Previous work has shown the citric acid cycle may play an important role in regulating longevity and well as paraquat resistance (DaCunha and DeOliveira, 1996). Interestingly we see TCA cycle enrichment in female maximum longevity and male paraquat resistance.; however, in neither sex does resistance to hydrogen peroxide show TCA cycle enrichment. Previous studies have shown that paraquat causes reactive oxygen species production by altering electron transport in the mitochondria (Bus and Gibson, 1984), while hydrogen peroxide cause reactive oxygen species production throughout the cell. This might explain, at least in part, why TCA-cycle metabolites are associated with paraquat resistance but not hydrogen peroxide resistance. Interestingly, with the exception of arginine degradation, no metabolic pathways were enriched in both paraquat and hydrogen peroxide resistance groups. The difference in the mechanisms of paraquat and hydrogen peroxide might also explain why paraquat resistance correlates with natural longevity while resistance to hydrogen peroxide does not. This coincides with the fact that most longevity studies with an oxidative stressor tend to utilize paraquat (e.g. Bjedov et al., 2010; Zou et al., 2000).

This is the largest comparative study of the metabolome to date. While we found that individual metabolites generally failed to have any significant phylogenetic signal, we did discover the entire metabolome could recapitulate some parts of the known phylogeny. At the level of individual metabolites, this could be due to a lack of power—we have relatively few species, and individual metabolite levels are highly sensitive to environmental as well as genetic and phylogenetic variation. The metabolome-wide

analysis likely eliminates at least some of this environmental variation, leading to some evidence for a phylogenetic signal. Our phylogenetic test of congruence found evidence for some similarities between the true species phylogeny and the metabolome derive phylogeny, and while the p-value was not statistically significant (P=0.089), it is suggestive of some phylogenetic concordance. The metabolome derived phylogeny places sister species *D. pseudoobscura* and *D. persimilis* together as well as *D. virilis* and *D. mojavensis* (they are not sister species but are most closely related of the species analyzed here). We also discover *D. melanogaster* and *D. simulans* group together, but *D. sechellia* does not cluster with these two species, as would be expected based on evolutionary history. *Drosophila sechellia* has evolved the ability to survive on *Morinda citrifolia* (Jones, 2005), which contains high levels of oxalic acid. This adaptation might have caused a shift in this species' metabolic pathways such that their metabolomic profile is no longer similar to related species. It is also interesting to note that all these species were reared in the same laboratory environment, so this metabolomic shift appears to be an evolved character, not a response to the environment. By studying those individual metabolites that are catalyzed my proteins from recently discovered genes (Huang and Erezyilmaz, 2015) that detoxify oxalic acid, we may be able to determine the metabolic consequences that cause *D. sechellia* to have such a distinct metabolomic profile.

In this study, we only analyzed those metabolites that were present in at least 80% of all samples, such that all metabolites analyzed here were present in at least some samples in all 11 species. However, the original data set included almost 2000 metabolites in one or more samples. Future analyses will be completed on the entire dataset to discover if specific metabolites are found on one branch of the phylogeny but not others. If we can

determine metabolites showing this pattern, they may give us insight into how metabolic pathways evolve or are regulated.

Our phylogenetic reconstruction places *D. virilis*, *D. mojavensis*, *D. pseudoobscura*, and *D. persimilis* as most closely related to each other with respect to their metabolomes; a pattern also observed in our PCA analysis. These four species differ from the other species in our study in that they have markedly slower development time (3 week generation verses 2 weeks for all the other species). This suggests that the metabolome potentially contains signatures of physiological pathways related to development. Previous research has shown that the metabolome is highly variable during development (An et al., 2014), and the adult metabolome is significantly affected by developmental temperature (Hariharan et al., 2014). However, all previous developmental metabolomic studies have only looked at *D. melanogaster*. While not analyzed in this study, future analyses should look at how the metabolome changes in response to different natural developmental conditions and how shifts in developmental environments might shape the metabolome of different species.

Caveats

This work sets the stage for a new and powerful approach to understanding the underlying mechanisms of interspecific variation and adaptation. While the scope of this work is large, there were several limitations. First, flies used for metabolomics sampling were not collected at the same time as the flies in the longevity analysis. Because the two samples came from different experiments, there potentially could be noise caused by the differences in the laboratory environment between the two experiments. However, we still

find many metabolites associated with median and maximum longevity, so it appears the different times of the experiment did not affect our ability to detect associations between metabolites and longevity. Second, because of destructive sampling, we could not measure metabolomic profiles for the exact flies used in our longevity analysis. Therefore, we can only use statistical measures of longevity (median, maximum) to find metabolomic correlations. If we could sample the individual flies and measure their true longevity, we might be better able to find metabolomic correlates with longevity. However, this is not possible in flies; but it could potentially be conducted across different species of captive rodents. Lastly, we sampled all flies for metabolomics at the same age, and then asked what associations with longevity were found. We implicitly assume that the best predictors of life history traits are early-age steady-state levels. However, it might be that age-related trajectories of individual metabolites are more indicative of changes in physiology and life history traits (e.g. Hoffman et al., 2014). We might be able to find longevity correlates if we sample flies at different ages and correlate the age-related changes in metabolite levels with longevity.

Finally, while greater than any previous study, our sample sizes within genotypes, among genotypes, and among species were smaller than we would have liked. We only analyzed three replicate samples per sex-by-genotype combination. While each sample did contain ten flies, this might not be enough to actually represent the environmental variation seen within inbred genotypes. Increased metabolomic sampling would give much more power to this study. Also, we attempted to use three wild type genotypes for each species; however, for two species (*D. erecta* and *D. yakuba*) we only had one genotype. This could have affected our phylogenetic reconstruction because for *D. erecta* and *D. yakuba* we

only ended up with six metabolomic samples per species, which could potentially explain why these two species were relatively isolated in the metabolome-derived phylogeny. Ideally, we need more genotypes for these two species to place the two species into their correct place on the metabolomic phylogenetic reconstruction. The lack of genotypes for these two species also meant we had fewer flies to make overall species longevity measures. Lastly, we had vastly different numbers of flies analyzed per genotype and treatment (Table S3.1). Many species had hundreds of flies per longevity analysis while others only had tens. This could have biased the longevity measures in the metabolomic analysis correlates due to small sample sizes for some of the species and genotypes.

Conclusion

Here we have presented the largest comparative metabolomics study ever completed, and we are the first to quantify the metabolomic variation that exists at the species and genotype level. We show the metabolome cannot recapitulate the entire phylogeny. However, related species often group together, and some potentially large adaptive changes are seen. We also find the metabolome is highly associated with sex, species, and body weight, and we find large effects of longevity and resistance of oxidative stressors, though they are different between the sexes. This study is the first to show the enormous power comparative metabolomic studies have to understand the metabolic consequences of evolution, and to generate new hypotheses about why aging and longevity vary across species.

ACKNOWLEDGEMENTS

This work was partially funded by NSF1402604 awarded to JMH. We want to thank the Matzkin lab at the University of Alabama at Huntsville for *D. mojavensis* lines, the Rabinovitch lab at the University of Washington for use of their equipment, and all members of the Promislow lab who helped collect the thousands of flies analyzed in this experiment.

LITERATURE CITED

An, P.N.T., Yamaguchi, M., Bamba, T., and Fukusaki, E. (2014). Metabolome Analysis of Drosophila melanogaster during Embryogenesis. Plos One *9*.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

Bjedov, I., Toivonen, J.M., Kerr, F., Slack, C., Jacobson, J., Foley, A., and Partridge, L. (2010). Mechanisms of Life Span Extension by Rapamycin in the Fruit Fly Drosophila melanogaster. Cell Metab *11*, 35-46.

Blekhman, R., Perry, G.H., Shahbaz, S., Fiehn, O., Clark, A.G., and Gilad, Y. (2014). Comparative metabolomics in primates reveals the effects of diet and gene regulatory variation on metabolic divergence. Sci Rep-Uk *4*.

Bo, T.H., Dysvik, J., and Jonassen, I. (2004). LSimpute: accurate estimation of missing values in microarray data with least squares methods. Nucleic Acids Res *32*.

Bozek, K., Wei, Y.N., Yan, Z., Liu, X.L., Xiong, J.Y., Sugimoto, M., Tomita, M., Paabo, S., Pieszek, R., Sherwood, C.C.*, et al.* (2014). Exceptional Evolutionary Divergence of Human

Muscle and Brain Metabolomes Parallels Human Cognitive and Physical Uniqueness. Plos Biol *12*.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M.*, et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-+.

Bus, J.S., and Gibson, J.E. (1984). Paraquat - Model for Oxidant-Initiated Toxicity. Environ Health Persp *55*, 37-46.

Capy, P., Pla, E., and David, J.R. (1993). Phenotypic and Genetic-Variability of Morphometrical Traits in Natural-Populations of Drosophila-Melanogaster and Drosophila-Simulans .1. Geographic Variations. Genet Sel Evol *25*, 517-536.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N.*, et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*, 203-218.

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

Culhane, A.C., Thioulouse, J., Perriere, G., and Higgins, D.G. (2005). MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics *21*, 2789-2790.

DaCunha, G.L., and DeOliveira, A.K. (1996). Citric acid cycle: A mainstream metabolic pathway influencing life span in Drosophila melanogaster? Exp Gerontol *31*, 705-715.

Damuth, J. (1981). Population-Density and Body Size in Mammals. Nature *290*, 699-700.

Durbin, E.J., and Yoon, J.S. (1986). Longevity in the Drosophila-Virilis Species Group .1. The Drosophila-Virilis Phylad. Ohio J Sci *86*, 14-17.

Greenberg, A.J., Stockwell, S.R., and Clark, A.G. (2008). Evolutionary Constraint and Adaptation in the Metabolic Network of Drosophila. Mol Biol Evol *25*, 2537-2546.

Harcourt, A.H., Harvey, P.H., Larson, S.G., and Short, R.V. (1981). Testis Weight, Body-Weight and Breeding System in Primates. Nature *293*, 55-57.

Hariharan, R., Hoffman, J.M., Thomas, A.S., Soltow, Q.A., Jones, D.P., and Promislow, D.E.L. (2014). Invariance and plasticity in the Drosophila melanogaster metabolomic network in response to temperature. Bmc Syst Biol *8*.

Harvey, P.H., and Pagel, M. (1991). The comparative method in evolutionary biology (Oxford).

Harvey, P.H., Pagel, M.D., and Rees, J.A. (1991). Mammalian Metabolism and Life Histories. Am Nat *137*, 556-566.

Hoffman, J.M., Soltow, Q.A., Li, S.Z., Sidik, A., Jones, D.P., and Promislow, D.E.L. (2014). Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, Drosophila melanogaster. Aging Cell *13*, 596-604.

Huang, Y., and Erezyilmaz, D. (2015). The genetics of resistance to Morinda fruit toxin during the postembryonic stages in Drosophila sechellia.

Ivanisevic, J., Thomas, O.P., Lejeusne, C., Chevaldonne, P., and Perez, T. (2011). Metabolic fingerprinting as an indicator of biodiversity: towards understanding inter-specific relationships among Homoscleromorpha sponges. Metabolomics *7*, 289-304.

Jones, C.D. (2005). The genetics of adaptation in Drosophila sechellia. Genetica *123*, 137-145.

Kambysel, M.P., and Heed, W.B. (1971). Studies of Oogenesis in Natural Populations of Drosophilidae .1. Relation of Ovarian Development and Ecological Habitats of Hawaiian Species. Am Nat *105*, 31-&.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., and Webb, C.O. (2010). Picante: R tools for integrating phylogenies and ecology. Bioinformatics *26*, 1463-1464.

Kendall, M.G., and Smith, B.B. (1939). The problem of m rankings. Ann Math Stat *10*, 275-287.

Kouchiwa, T., Wada, K., Uchiyama, M., Kasezawa, N., Niisato, M., Murakami, H., Fukuyama, K., and Yokogoshi, H. (2012). Age-related changes in serum amino acids concentrations in healthy individuals. Clin Chem Lab Med *50*, 861-870.

Larracuente, A.M., Sackton, T.B., Greenberg, A.J., Wong, A., Singh, N.D., Sturgill, D., Zhang, Y., Oliver, B., and Clark, A.G. (2008). Evolution of protein-coding genes in Drosophila. Trends Genet *24*, 114-123.

Lawton, K.A., Berger, A., Mitchell, M., Milgram, K.E., Evans, A.M., Guo, L.N., Hanson, R.W., Kalhan, S.C., Ryals, J.A., and Milburn, M.V. (2008). Analysis of the adult human plasma metabolome. Pharmacogenomics *9*, 383-397.

Li, S.Z., Park, Y., Duraisingham, S., Strobel, F.H., Khan, N., Soltow, Q.A., Jones, D.P., and Pulendran, B. (2013). Predicting Network Activity from High Throughput Metabolomics. Plos Comput Biol *9*.

Linford, N.J., Bilgir, C., Ro, J., and Pletcher, S.D. (2013). Measurement of Lifespan in Drosophila melanogaster. Jove-J Vis Exp.

Lu, J., Shen, Y., Wu, Q., Kumar, S., He, B., Shi, S., Carthew, R.W., Wang, S.M., and Wu, C.I. (2008). The birth and death of microRNA genes in Drosophila. Nat Genet *40*, 351-355.

Pagel, M. (1994). Detecting Correlated Evolution on Phylogenies - a General-Method for the Comparative-Analysis of Discrete Characters. P Roy Soc B-Biol Sci *255*, 37-45.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289-290.

Park, Y.H., Lee, K., Soltow, Q.A., Strobel, F.H., Brigham, K.L., Parker, R.E., Wilson, M.E., Sutliff, R.L., Mansfield, K.G., Wachtman, L.M.*, et al.* (2012). High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. Toxicology *295*, 47-55.

Partridge, L., Hoffmann, A., and Jones, J.S. (1987). Male Size and Mating Success in Drosophila-Melanogaster and Drosophila-Pseudoobscura under Field Conditions. Anim Behav *35*, 468-476.

Promislow, D.E.L. (1991). Senescence in Natural-Populations of Mammals - a Comparative-Study. Evolution *45*, 1869-1887.

Promislow, D.E.L., and Harvey, P.H. (1990). Living Fast and Dying Young - a Comparative-Analysis of Life-History Variation among Mammals. J Zool *220*, 417-437.

Promislow, D.E.L., and Haselkorn, T.S. (2002). Age-specific metabolic rates and mortality rates in the genus Drosophila. Aging Cell *1*, 66-74.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing *Vienna, Austria*.

Saitou, N., and Nei, M. (1987). The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. Mol Biol Evol *4*, 406-425.

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T.*, et al.* (2012). Insights into hominid evolution from the gorilla genome sequence. Nature *483*, 169-175.

Schnebel, E.M., and Grossfield, J. (1983). A Comparison of Life-Span Characteristics in Drosophila. Exp Gerontol *18*, 325-337.

Slupsky, C.M., Rankin, K.N., Wagner, J., Fu, H., Chang, D., Weljie, A.M., Saude, E.J., Lix, B., Adamko, D.J., Shah, S.*, et al.* (2007). Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. Anal Chem *79*, 6995-7004.

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N.*, et al.* (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature *450*, 219-232.

Zou, S., Meadows, S., Sharp, L., Jan, L.Y., and Jan, Y.N. (2000). Genome-wide study of aging and oxidative stress response in Drosophila melanogaster. P Natl Acad Sci USA *97*, 13726-13731.

FIGURE LEGENDS

Figure 3.1. Longevity curves for each species and sex analyzed. Each species is represented by a separate color, and females (solid lines) and males (dashed lines) are pictured separately.

Figure 3.2. Correlations body weight, median longevity, paraquat resistance, and hydrogen peroxide resistance for all genotypes of females.

Figure 3.3. Correlations body weight, median longevity, paraquat resistance, and hydrogen peroxide resistance for all genotypes of males.

Figure 3.4. Plot of true species phylogeny (left) and metabolite phylogeny (right). The true species phylogeny was used to test for phylogenetic signal among individual metabolites. The metabolite phylogeny was constructed using a neighbor-joining algorithm. Lines between trees connect the same species. Kendall's W test of congruence p-value 0.08

Figure 3.5. Plot of two metabolites affected by weight (left) and sex (age). $P<0.0001$ for both metabolites. Red lines/dots represent females, blue lines/dots represent males.

Figure 3.6. Plot of four individual metabolites associated with longevity in females. ANOVA $P<0.0001$ for all metabolites after controlling for body weight and species.

Figure 3.7. Plot of four individual metabolites associated with longevity in males. ANOVA $P<0.0001$ for all metabolites after controlling for body weight and species.

Figure 3.8. PCA plot of metabolome colored by sex. Females (blue). Males (red). Axes shown are PC 1 and PC 2 that explain 20.3% and 13.0% of the variance respectively. Axis 2 is associated significantly with sex (Table S2, $P<0.0001$).

Figure 3.9. PCA correlates with species. PC2 and PC4 are shown, These PCs were chosen because of best visualization of the species. Colors represent different species. Both axes are significantly correlated with species (Table S2).

Figure S3.1. Longevity plots for each species (all genotypes combined). Colors denote sexes: females (red) and males (blue). P-values indicate survival differences from Kaplan-Meier analysis of sex.

Table 3.1. Coefficients of variation for longevity for all sexes and species.

Species	Sex	Coefficient of Variation
D. ananassae	male	36
D. ananassae	female	45
D. erecta	male	31
D. erecta	female	61
D. melanogaster	male	29
D. melanogaster	female	42
D. mojavensis	female	41
D. mojavensis	male	51
D. persimilis	female	31
D. persimilis	male	45
D. pseudoobscura	female	38
D. pseudoobscura	male	55
D. sechellia	male	26
D. sechellia	female	40
D. simulans	female	24
D. simulans	male	38
D. virilis	female	35
D. virilis	male	45
D. willistoni	female	34
D. willistoni	male	39
D. yakuba	female	31
D. yakuba	male	49

Table 3.2. All metabolites found to be significant with different factors. For the sex specific factors, we controlled for body weight and species.

Factor	Metabolites Significant	Proportion Metabolites
Sex	104	41.11%
Species	176	69.57%
Genotype	53	20.95%
Body Weight	145	57.31%
Females		
Median Longevity	71	28.06%
Maximum Longevity	104	41.11%
Paraquat Resistance	128	50.59%
H2O2 Resistance	61	24.11%
Males		
Median Longevity	77	30.43%
Maximum Longevity	98	38.74%
Paraquat Resistance	61	24.11%
H2O2 Resistance	76	30.04%

Table 3.3. Pathway enrichment within groups of metabolites associated with different predictors. Only those pathways with adjusted p-values < 0.05 are shown. Letter in factor level indicates which sex was affected: F- females and M- males.

Pathway	Factor	Factor Mets	Total Mets	P-value
Gluconeogenesis	Weight	5	5	0.0152
Salvage pathways of adenine	Weight	4	4	0.0364
FormylTHF biosynthesis I	Sex	3	3	0.0091
Gluconeogensis	Sex	4	5	0.0113
Arginine degradation VI (arginase 2 pathway)	Sex	4	5	0.0113
Amino Acids	Sex	7	11	0.0186
Proline biosynthesis II	Sex	3	4	0.0352
Proline biosynthesis III	Sex	3	4	0.0352
Arginine degradation I (arginase pathway)	Sex	3	4	0.0352
TCA cycle variation III	Sex	3	4	0.0352
Tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	Median Longevity -F	3	3	0.0004
Salvage pathways of adenine	Median Longevity -F	3	4	0.0007
Amino Acids	Median Longevity - F	3	11	0.0444
Amino Acids	Median Longevity- M	7	11	0.0062
FormylTHF biosynthesis I	Median Longevity- M	3	3	0.0092
Aspartate degradation	Median Longevity- M	3	3	0.0092
Glycogen degradation	Median Longevity- M	3	4	0.0279
Tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	Maximum Longevity- F	3	3	0.0073
Arginine degradation X (arginine monooxygenase pathway)	Maximum Longevity- F	3	3	0.0073
Salvage pathways of adenine	Maximum Longevity- F	3	4	0.0239
TCA cycle variation III	Maximum Longevity- F	3	4	0.0239
TCA cycle variation III	Paraquat Resistance-M	3	4	0.0019
Tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	H2O2 Resistance- M	3	3	0.0039
Amino Acids	H2O2 Resistance- M	5	11	0.0207
Arginine degradation VI (arginase 2 pathway)	H2O2 Resistance- M	3	5	0.0211

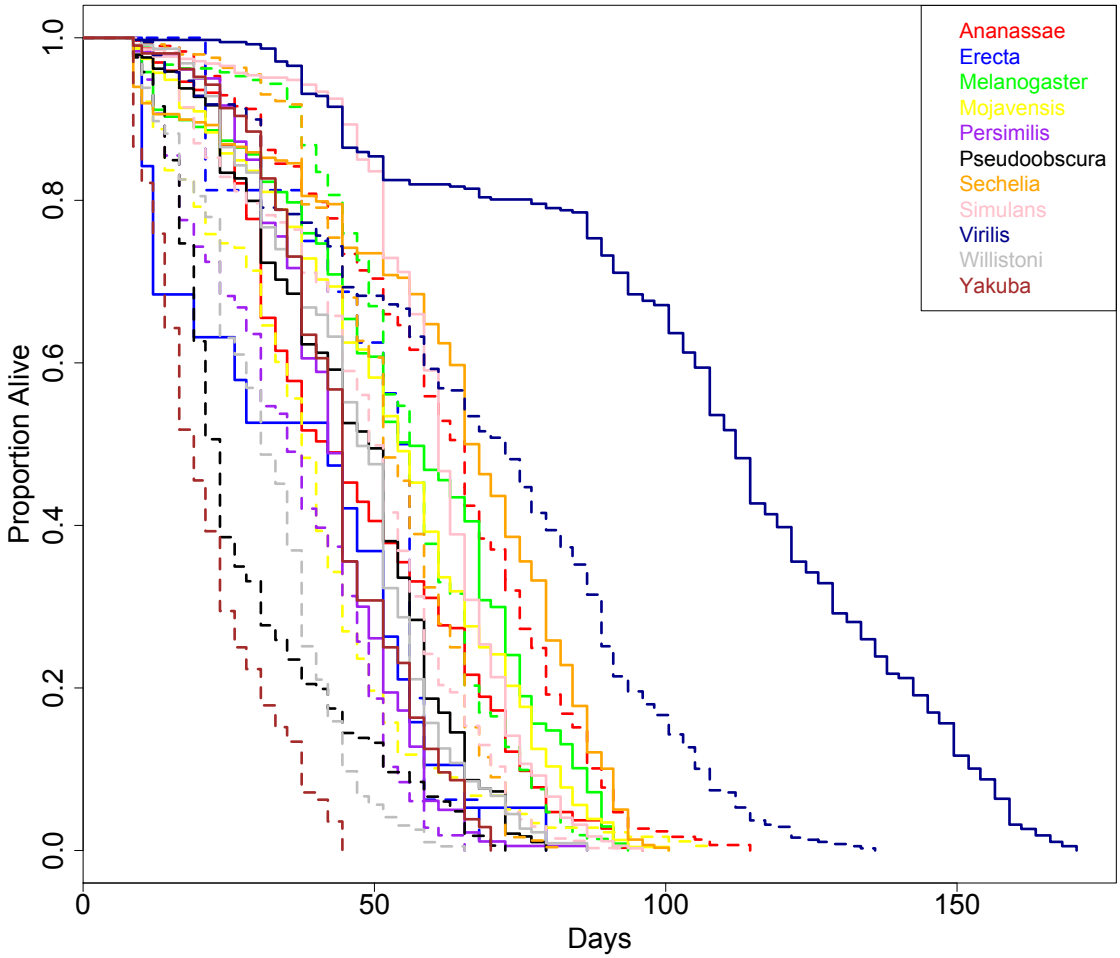
Ananassae
Erecta
Melanogaster
Mojavensis
Persimilis
Pseudoobscura
Sechelia
Simulans
Virilis
Willistoni
Yakuba
Proportion Alive
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
Days

Figure 3.1

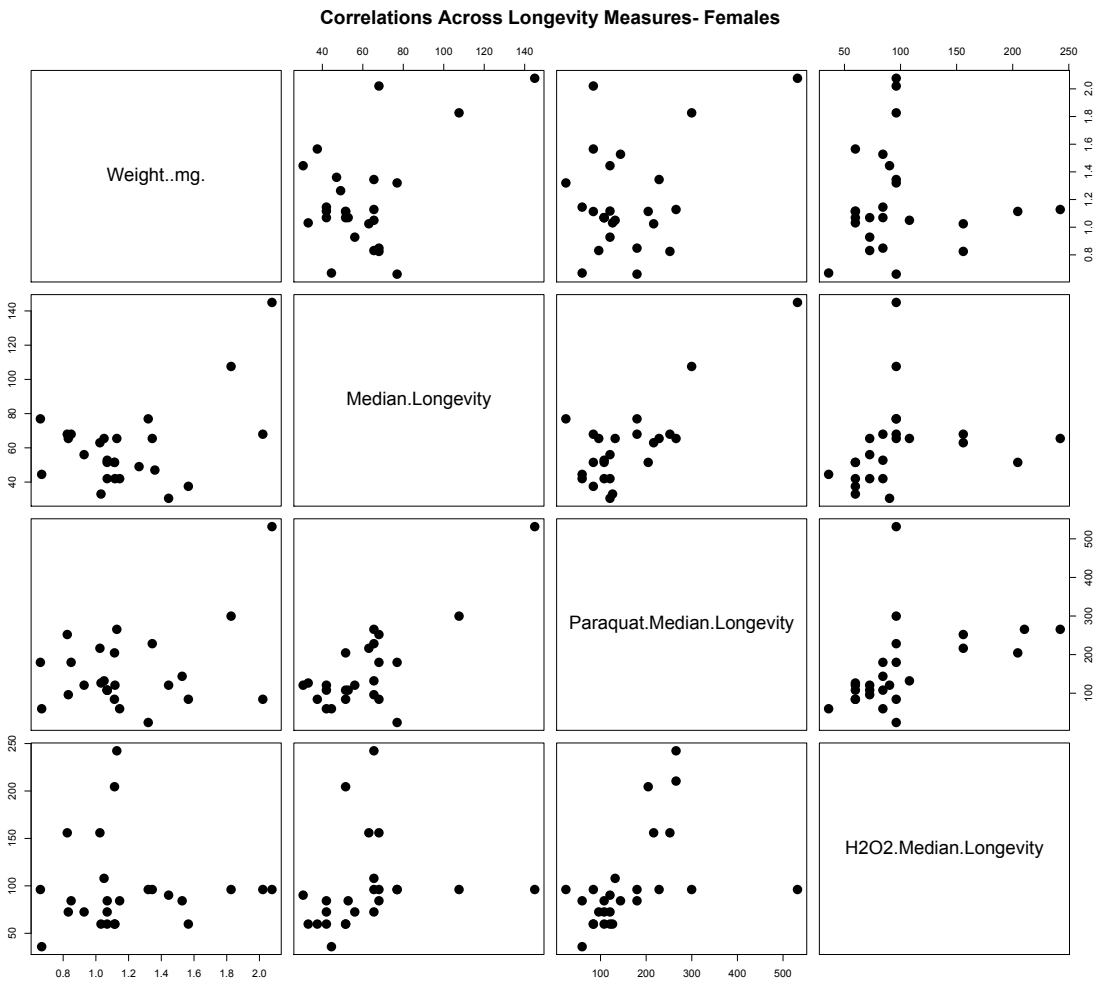
Correlations Across Longevity Measures- Females
Weight..mg.
Median.Longevity
Paraquat.Median.Longevity
H2O2.Median.Longevity

Figure 3.2

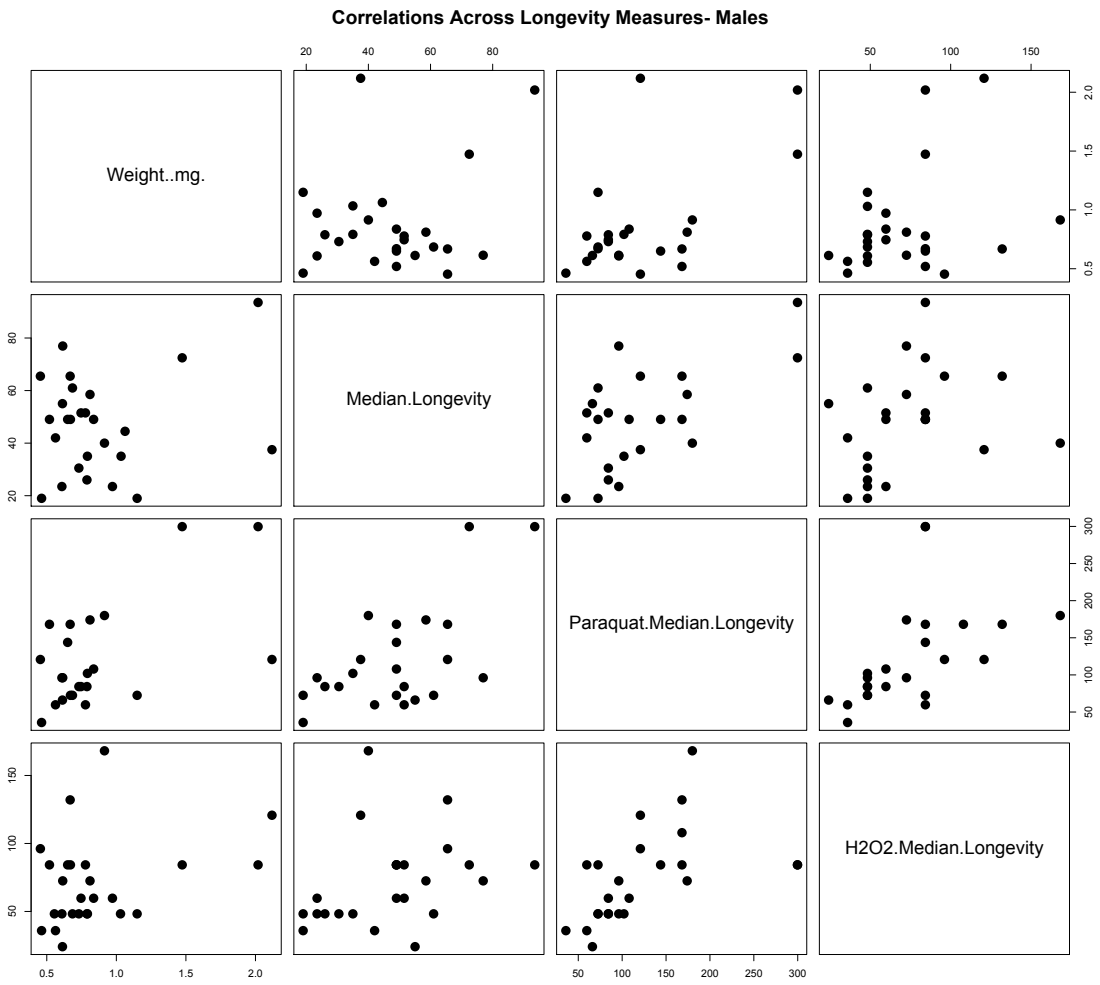
Correlations Across Longevity Measures- Males
Weight..mg.
Median.Longevity
Paraquat.Median.Longevity
H2O2.Median.Longevity

Figure 3.3

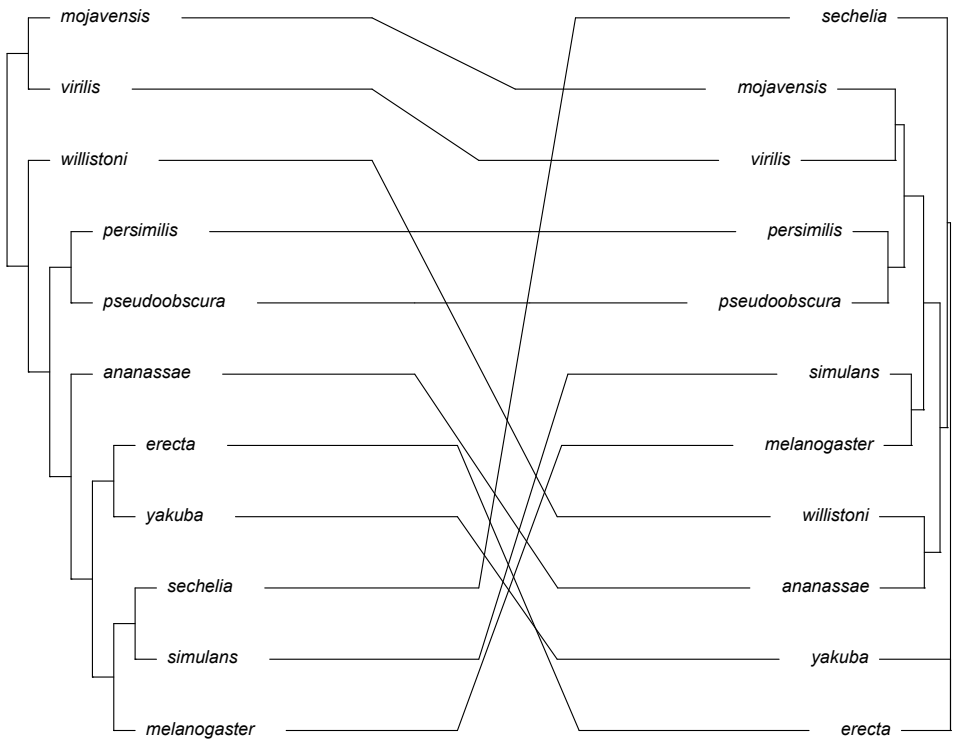
mojavensis
virilis
willistoni
persimilis
pseudoobscura
ananassae
erecta
yakuba
sechelia
simulans
melanogaster
sechelia
mojavensis
virilis
persimilis
pseudoobscura
simulans
melanogaster
willistoni
ananassae
yakuba
erecta

Figure 3.4

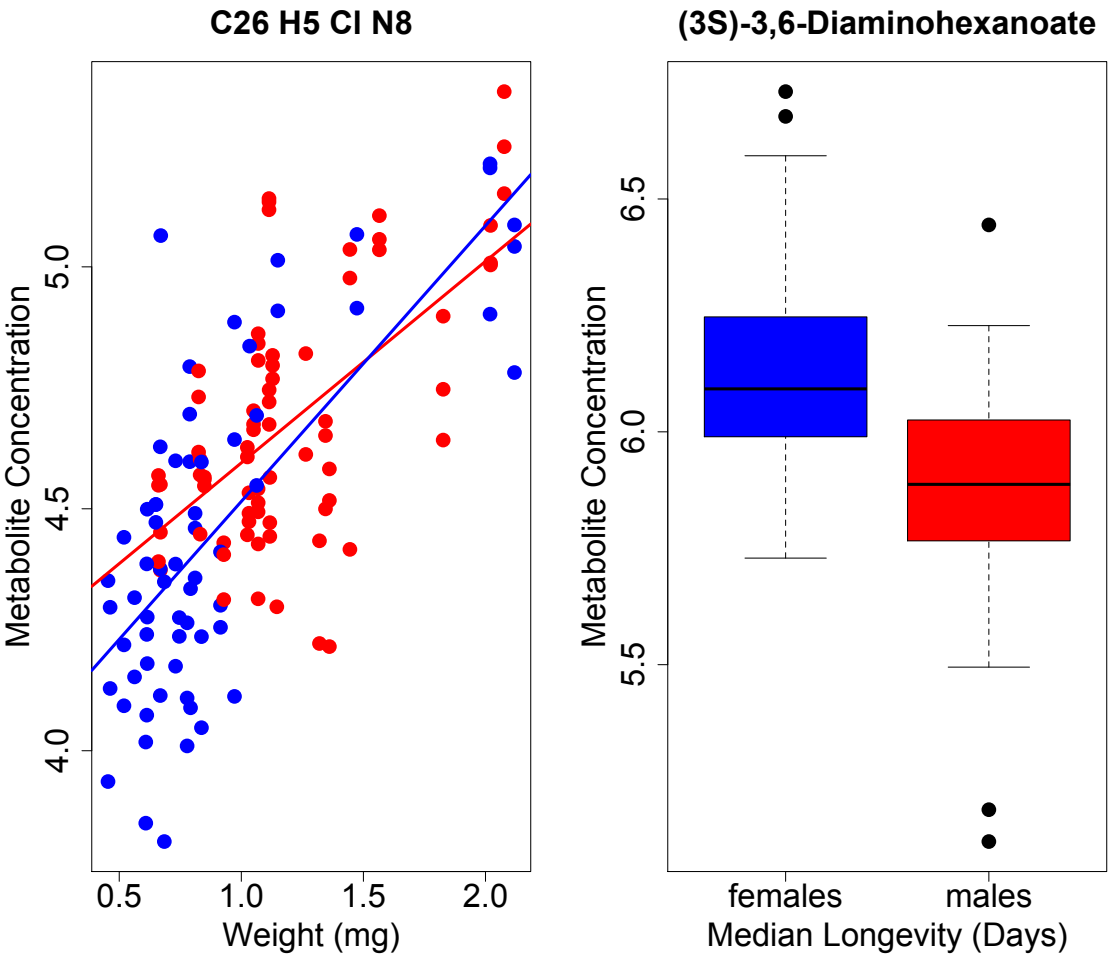
C26 H5 Cl N8
Metabolite Concentration
4.0
4.5
5.0
0.5
1.0
1.5
2.0
Weight (mg)
(3S)-3,6-Diaminohexanoate
Metabolite Concentration
5.5
6.0
6.5
females
males
Median Longevity (Days)

Figure 3.5

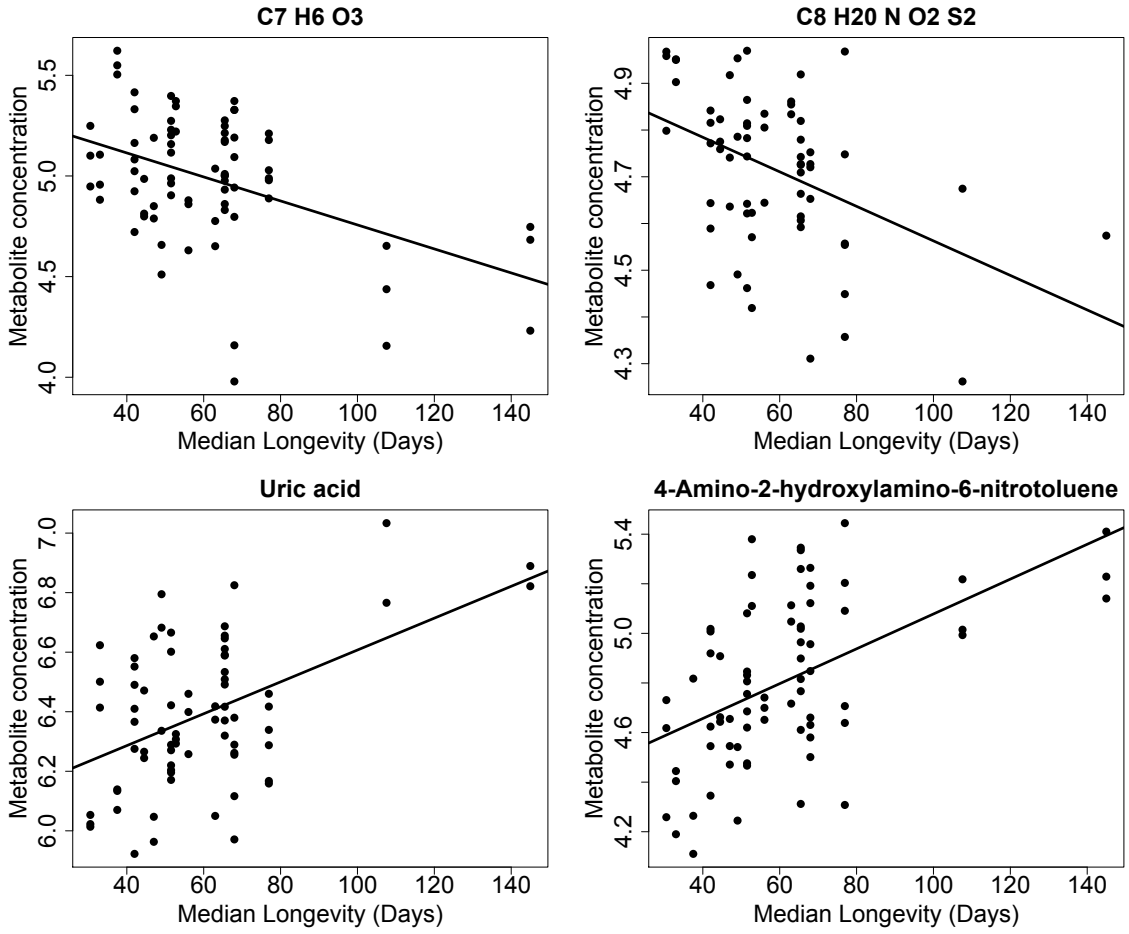
C7 H6 O3
Metabolite concentration
Median Longevity (Days)
C8 H20 N O2 S2
Metabolite concentration
Median Longevity (Days)
Uric acid
Metabolite concentration
Median Longevity (Days)
4-Amino-2-hydroxylamino-6-nitrotoluene
Metabolite concentration
Median Longevity (Days)

Figure 3.6

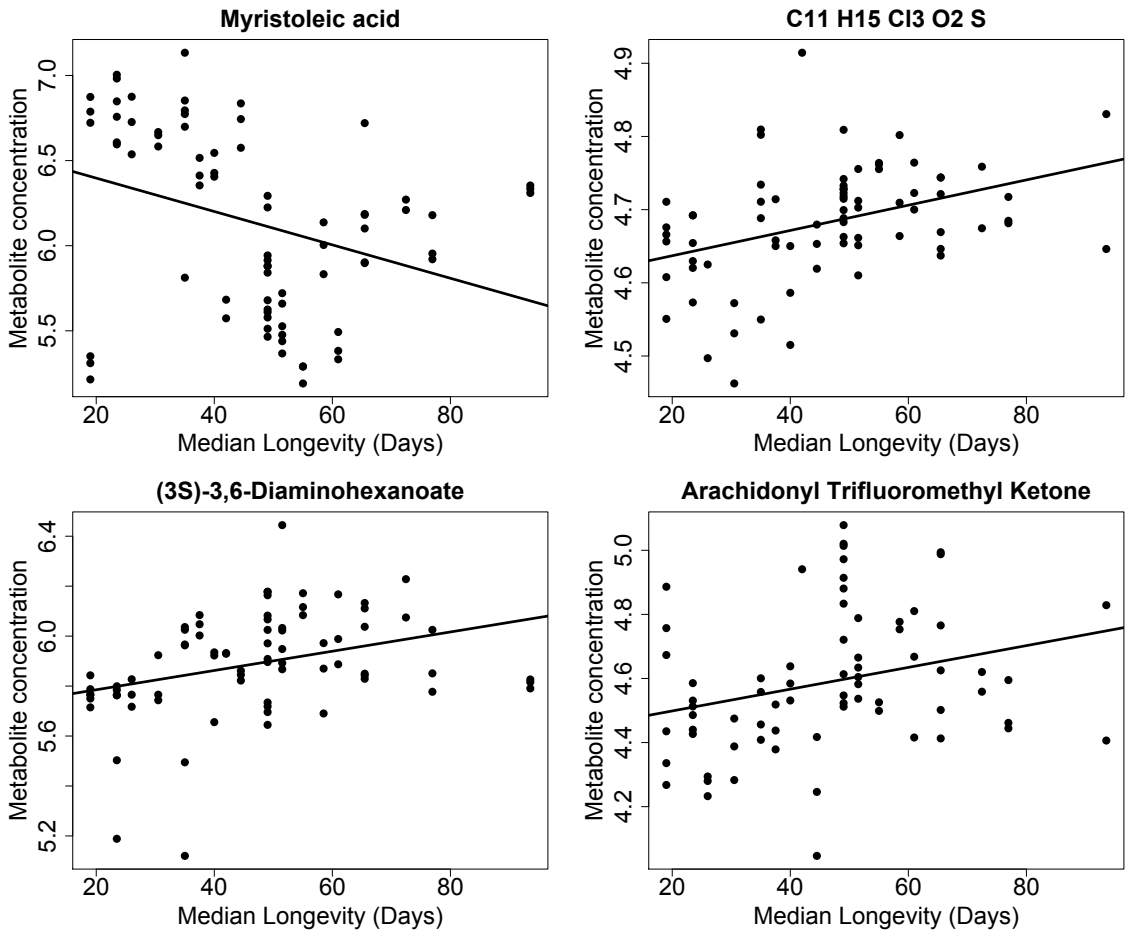
Myristoleic acid
Metabolite concentration
Median Longevity (Days)
C11 H15 Cl3 O2 S
Metabolite concentration
Median Longevity (Days)
(3S)-3,6-Diaminohexanoate
Metabolite concentration
Median Longevity (Days)
Arachidonyl Trifluoromethyl Ketone
Metabolite concentration
Median Longevity (Days)

Figure 3.7

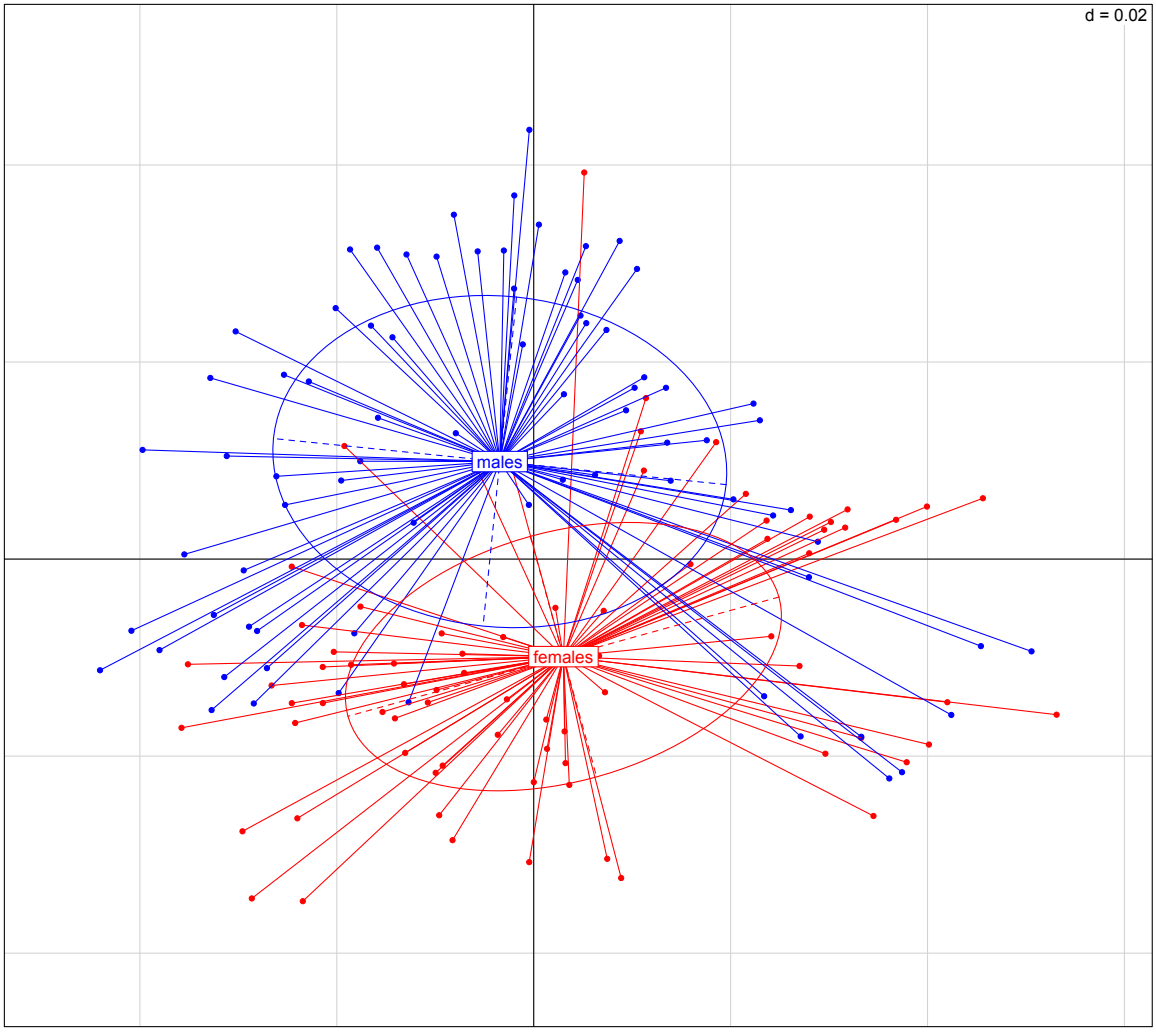

Figure 3.8

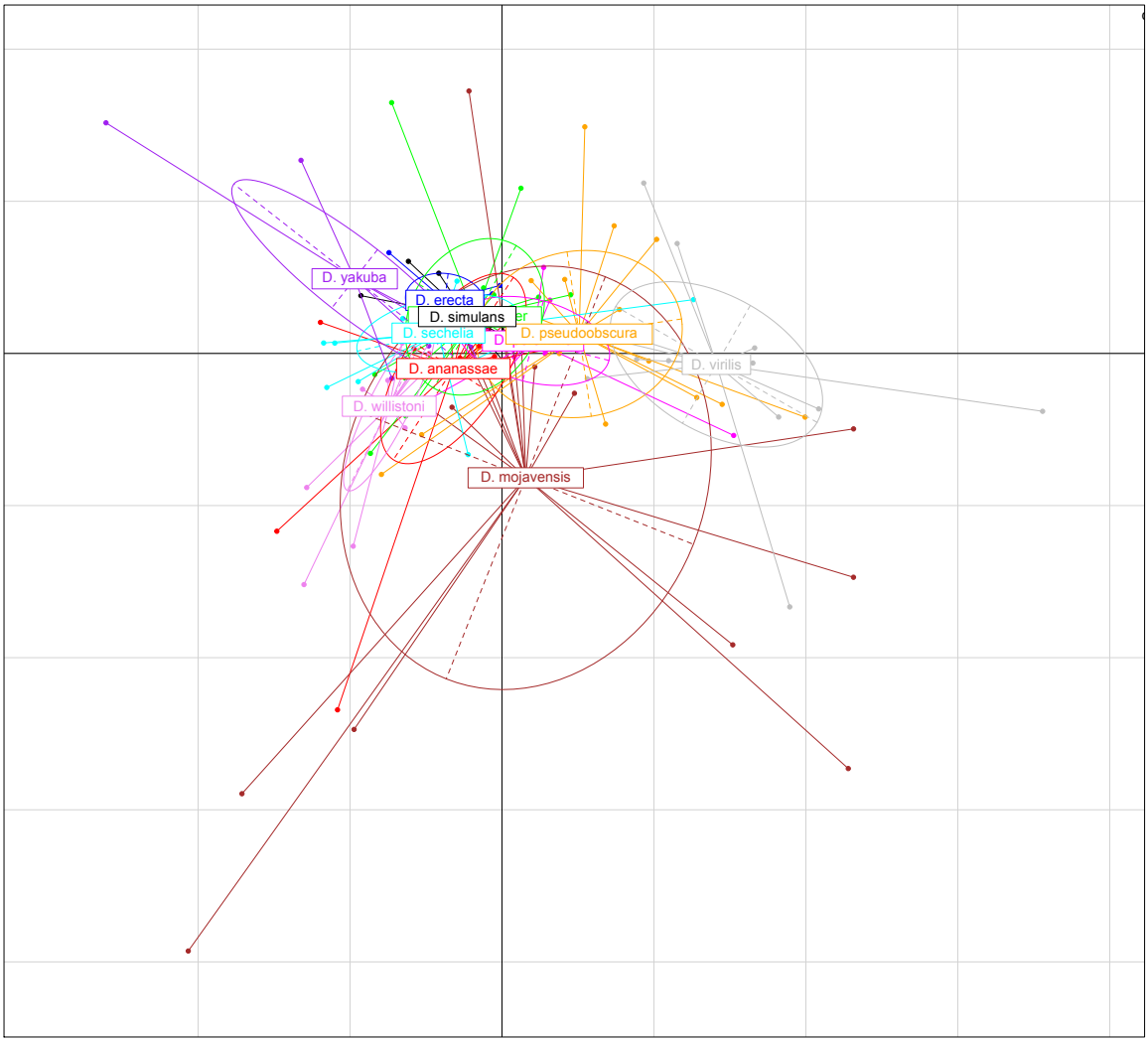
D. yakuba
D. erecta
D. simulans
D. sechelia
D. pseudoobscura
D. virilis
D. ananassae
D. willistoni
D. mojavensis

Figure 3.9

SUPPLEMENTARY INFORMATION

Table S3.1. Genotypes, numbers of flies, and basic longevity measures for all species analyzed. All longevity measures are median unless otherwise stated.

P. Longevity represents paraquat longevity. Natural longevity measures are in days.

Oxidative stress longevities are in hours.

Species	Line	Sex	(mg)	Longevity Flies	H2O2 flies	Paraquat flies	Longevity	Max Longevity	H_2O_2 Longevity	P. Longevity
D. ananassae	Ana-00	F	1.44	118.00	34.00	18.00	30.51	44.50	90.20	120.76
D. ananassae	Ana-00	M	0.84	115.00	43.00	40.00	49.03	100.52	59.72	107.92
D. ananassae	Ana-12	F	1.11	117.00	111.00	94.00	51.51	96.04	204.50	204.50
D. ananassae	Ana-12	M	0.67	86.00	104.00	98.00	65.48	114.49	132.04	168.14
D. ananassae	Ana-15	F	0.85	61.00	95.00	73.00	67.97	90.99	84.26	179.95
D. ananassae	Ana-15	M	0.61	96.00	78.00	76.00	76.95	114.49	72.50	96.14
D. erecta	Ere-00	F	1.12	19.00	32.00	19.00	42.00	79.46	59.72	120.76
D. erecta	Ere-00	M	0.61	16.00	37.00	20.00	55.02	67.97	24.10	66.11
D. melanogaster	Mel-14	F	0.83	35.00	57.00	40.00	67.97	74.95	155.93	252.07
D. melanogaster	Mel-14	M	0.52	34.00	124.00	112.00	49.03	84.00	84.26	168.14
D. melanogaster	Mel-56	F	0.66	88.00	147.00	119.00	76.95	93.53	96.14	179.95
D. melanogaster	Mel-56	M	0.45	102.00	100.00	89.00	65.48	93.53	96.14	120.76
D. melanogaster	Mel-61	F	1.15	114.00	19.00	20.00	42.00	62.99	84.26	59.72
D. melanogaster	Mel-61	M	0.67	76.00	38.00	37.00	49.03	60.99	84.26	72.50
D. mojavensis	Moj-1002	F	1.13	111.00	14.00	19.00	65.48	100.52	242.37	265.51
D. mojavensis	Moj-1002	M	0.91	78.00	19.00	20.00	40.00	60.99	168.14	179.95
D. mojavensis	Moj-2008	F	1.26	121.00	NA	NA	49.03	86.50	NA	NA
D. mojavensis	Moj-2008	M	1.03	100.00	NA	NA	35.01	107.56	NA	NA
D. persimilis	Per-17	F	1.07	91.00	132.00	115.00	42.00	65.48	72.50	107.92
D. persimilis	Per-17	M	0.79	95.00	97.00	80.00	26.01	58.52	48.25	84.26
D. persimilis	Per-23	F	1.36	89.00	NA	NA	47.02	86.50	NA	NA
D. persimilis	Per-23	M	1.06	119.00	NA	NA	44.50	65.48	NA	NA
D. pseudoobscura	Pse-217	F	1.57	132.00	13.00	20.00	37.51	65.48	59.72	84.26
D. pseudoobscura	Pse-217	M	1.15	56.00	19.00	18.00	19.00	42.00	48.25	72.50
D. pseudoobscura	Pse-33	F	1.11	67.00	19.00	7.00	51.51	79.46	59.72	84.26
D. pseudoobscura	Pse-33	M	0.97	33.00	7.00	0.00	23.49	58.52	59.72	NA
D. pseudoobscura	Pse-96	F	1.07	90.00	143.00	132.00	52.76	79.46	84.26	107.92

D. pseudoobscura	Pse-96	M	0.73	77.00	146.00	138.00	30.51	72.46	48.25	84.26
D. sechellia	Sec-07	F	0.83	104.00	79.00	80.00	65.48	93.53	72.50	96.14
D. sechellia	Sec-07	M	0.68	91.00	43.00	36.00	60.99	79.46	48.25	72.50
D. sechellia	Sec-28	F	1.05	120.00	69.00	81.00	65.48	100.52	107.92	132.04
D. sechellia	Sec-28	M	0.75	100.00	81.00	80.00	51.51	81.95	59.72	84.26
D. sechellia	Sec-31	F	1.32	74.00	19.00	19.00	76.95	93.53	96.14	24.10
D. sechellia	Sec-31	M	0.78	53.00	20.00	13.00	51.51	79.46	84.26	59.72
D. simulans	Sim-006	F	0.93	108.00	62.00	59.00	56.03	72.46	72.50	120.76
D. simulans	Sim-006	M	0.56	111.00	54.00	30.00	42.00	67.97	35.89	59.72
D. simulans	Sim-286	F	1.34	118.00	55.00	58.00	65.48	96.04	96.14	228.37
D. simulans	Sim-286	M	0.81	119.00	37.00	38.00	58.52	86.50	72.50	174.05
D. simulans	Sim-287	F	1.02	121.00	52.00	34.00	62.99	86.50	155.93	216.42
D. simulans	Sim-287	M	0.65	109.00	60.00	37.00	49.03	93.53	84.26	143.81
D. virilis	Vir-00	F	2.02	126.00	138.00	107.00	67.97	128.64	96.14	84.26
D. virilis	Vir-00	M	2.12	133.00	158.00	150.00	37.51	93.53	120.76	120.76
D. virilis	Vir-08	F	2.08	120.00	59.00	61.00	144.97	170.52	96.14	531.64
D. virilis	Vir-08	M	2.02	126.00	59.00	59.00	93.53	135.94	84.26	299.72
D. virilis	Vir-48	F	1.83	131.00	18.00	22.00	107.56	166.01	96.14	299.72
D. virilis	Vir-48	M	1.47	119.00	20.00	21.00	72.46	131.08	84.26	299.72
D. willistoni	Wil-02	F	1.03	99.00	27.00	26.00	33.01	86.50	59.72	126.40
D. willistoni	Wil-02	M	0.61	74.00	19.00	16.00	23.49	58.52	48.25	96.14
D. willistoni	Wil-15	F	1.07	124.00	21.00	20.00	51.51	86.50	59.72	107.92
D. willistoni	Wil-15	M	0.79	121.00	23.00	16.00	35.01	65.48	48.25	102.03
D. yakuba	Yak-00	F	0.67	104.00	57.00	53.00	44.50	69.97	35.89	59.72
D. yakuba	Yak-00	M	0.46	112.00	60.00	38.00	19.00	44.50	35.89	35.89

Table S3.2. First six principal components and associations with sex, species and body weight. Numbers under factors represent p-values from linear model.

PC	Body weight	Sex	Species	% Variance Explained
1	4.82E-24	0.041	8.32E-14	20.32%
2	1.74E-13	2.03E-11	3.87E-09	12.95%
3	0.271	0.009	3.19E-24	6.28%
4	0.082	1.12E-07	2.64E-25	5.91%
5	0.010	0.002	0.001	3.95%
6	0.467	5.48E-05	5.17E-12	3.27%

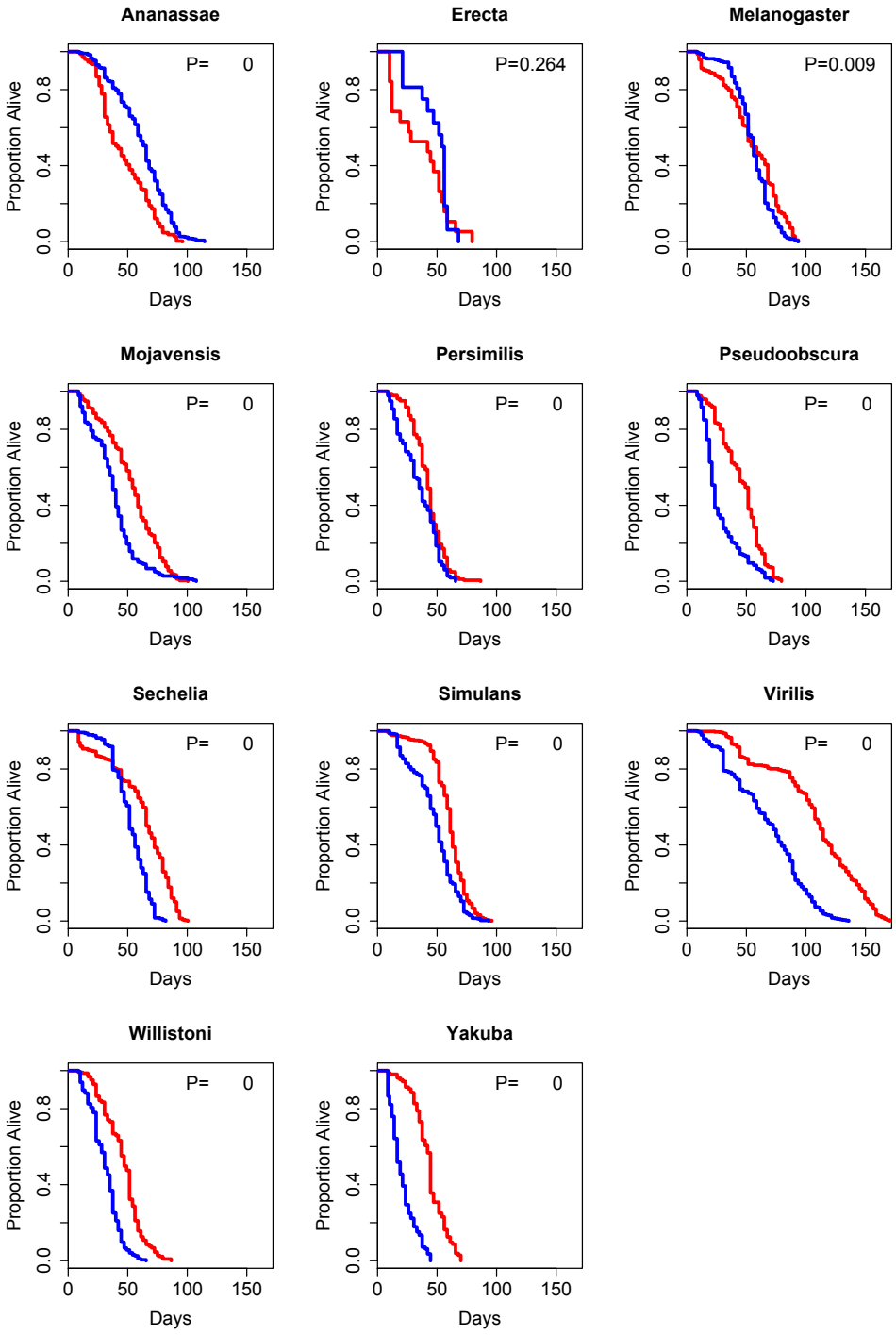
Ananassae
P= 0
Erecta
P=0.264
Melanogaster
P=0.009
Mojavensis
P= 0
Persimilis
P= 0
Pseudoobscura
P= 0
Sechelia
P= 0
Simulans
P= 0
Virilis
P= 0
Willistoni
P= 0
Yakuba
P= 0
Proportion Alive
0.0
0.4
0.8
0
50
100
150
Days

Figure S3.1

CHAPTER 4

A LONGITUDINAL ANALYSIS OF THE EFFECTS OF AGE ON THE METABOLOME IN THE COMMON MARMOSET, *CALLITHRIX JACCHUS*[3]

[3]Hoffman J.M., Tran V., Wachtman L., Jones D.P., and Promislow D.E.L. To be submitted to *Experimental Gerontology*.

ABSTRACT

Aging and longevity are both complex phenotypes that are difficult to study in humans due to their long lifespan, and as such non-human primate models have been extensively employed to understand the factors that influence aging and longevity. While most non-human primate studies have focused on the rhesus macaque, the common marmoset has recently been presented as new model of aging. However, little is known about the metabolic changes that occur throughout the life of the marmoset. Here we utilize high sensitivity metabolomics to understand the biochemical changes that are associated with age in the common marmoset. We tracked 2,104 metabolites over a 17-month period (3 separate time points), and we discover hundreds of metabolites that are associated with age after controlling for sex and genotype. We also discover the entire metabolome is both correlated with and predicative of age, and we suggest the metabolome has the potential to be an excellent biomarker of aging. We find significant enrichment in age-associated metabolites for metabolic pathways involved in oxidative stress and xenobiotic metabolism, suggesting these biochemical pathways might play an important role in the basic mechanisms of aging in primates. We show the power of using a longitudinal approach to understand the biochemical changes that occur with age.

INTRODUCTION

Many studies have attempted to determine biomarkers of age and mortality of individuals; however, most have been unsuccessful with the most accurate biomarkers only explaining a percent or so of the variation seen in aging and longevity. One reason for the

failure of these studies may be that they are usually cross-sectional. There are molecular changes that occur throughout an individual's life that may have profound effects of mortality, yet are not discovered in traditional aging studies. The most meaningful biomarkers might be those whose long-term trajectories, rather than static values at one time point, predict pathology or death. In this light, the gap in our knowledge can be filled at least in part by looking at longitudinal studies across the life of different individuals. Many longitudinal studies in humans have identified genetic and environmental correlates with longevity (e.g. Colditz and Hankinson, 2005; Ferrucci, 2008); however, these studies take many decades to complete. Therefore, we were interested in developing a model system to more easily identify longitudinal changes in systems biology phenomena that could predict healthspan. With this goal in mind, we believe non-human primates could play a pivotal role.

Small, relatively short-lived non-human primates offer a powerful, translational model to understand the causes and correlates of aging. Though most studies have utilized large primates, such as the rhesus macaque, recent studies have pointed to the common marmoset (*Callithrix jacchus*) as an ideal non-human primate model of aging (Fischer and Austad, 2011; Tardif et al., 2011). Marmosets have age-associated pathologies also seen in humans, and they have a relatively short lifespan (average 8-12 years and maximum 16.5-21.5) compared to other well-studied primates (Nishijima et al., 2012; Tardif et al., 2011), and half that of the most studied the rhesus macque.

We believe that by studying the biochemical changes that occur throughout the life of the marmoset, we may be able to better detect specific, accurate biomarkers of aging. Metabolomics, the study of all the small molecules in an organism, allows for

understanding of all metabolic changes that occur over time. Previous metabolomics studies in primates have used cross-sectional approaches to identify metabolites correlated with age (Muehlenbein et al., 2003), and longitudinal studies of metabolites in primates have been confined to earlier developmental stages (Higley et al., 1992). Thus, the effects of natural aging global metabolite profiles in primates are unknown.

Few studies have attempted to utilize metabolomics in the common marmoset. Prior studies of age-related biochemical changes in marmosets have focused on specific subsets of blood chemistry parameters (Kuehnel et al., 2012) and metabolites (Roede et al., 2012). However, the effects of age on global metabolome profiles are unknown. The metabolome represents all the small molecular components of metabolic pathways, and previous research in model organisms has shown that the analysis of metabolites has the potential to determine metabolic pathways that might be implicated in aging (Fuchs et al., 2010; Hoffman et al., 2014; Houtkooper et al., 2011). Previous primate metabolomics research has usually focused on biomarkers of specific diseases, not the affects of natural aging (Liu et al., 2013; Patterson et al., 2011), leaving a potentially important gap in our understanding of the metabolic pathways associated with natural aging.

Here, we take advantage of the marmoset's short lifespan to gain insight into the longitudinal changes in the metabolome, and use this system as a potential model for human aging metabolomics. While the common marmoset has been proposed as a new non-human primate model of aging, very little is known about how its biochemical makeup changes with age, other than basic blood chemistry measures (Kuehnel et al., 2012). We present the first longitudinal study of age related changes in the metabolome of a large

colony of marmosets, and we suggest the metabolome has the potential to be an excellent biomarker of aging.

METHODS

Marmosets

Our final dataset consisted of 229 unique marmosets with some metabolomics data of which 84 had data in each of the three time points. The animals ranged in age from 1-17 years with an average age of 6.3 years, and age distributions in each time point are shown in Figure 4.1.

Sample Collection

Marmosets were housed at the New England Primate Research Center and were maintained as described in (Roede et al., 2012). Blood samples were collected at three different time points over a 17-month period (June 2012, October 2012, and November 2013) during routine physical exams of the animals under sedation with 0.2mL of ketamine as described in Roede et al. (2012). Data were collected from well over 100 animals at each time point (156 in June 2012, 175 in October 2012, and 144 in November 2013).

Metabolomic analysis

Metabolites were analyzed by high-resolution mass spectrometry (MS; LTQ-Velos Orbitrap, Thermo Fisher) coupled to liquid chromatography (LC) using a reverse-phase C18 column (Soltow et al., 2013). Briefly, 50 µL of plasma was added to 100 µL of acetonitrile and 2.5 µL containing stable isotope standards. Samples were mixed, incubated at 4^{o}C for 30 min, and centrifuged to remove protein. Supernatants were analyzed in triplicate by LC-MS (Go et al., 2014; Soltow et al., 2013). Data were extracted using `apLCMS` (Yu et al., 2009) with `xMSanalyzer` (Uppal et al., 2013) as *m/z* features, where an *m/z*

feature is defined by *m/z* (mass-to-charge ratio), retention time, and ion intensity (Johnson et al., 2010).

Data analysis

All data analyses were completed using the statistics package R (R Core Team, 2013) unless otherwise stated. For each time point, the data structure originally consisted of over 20,000 metabolites measured for each individual, and each individual sample was run in triplicate technical replicates. We first removed those metabolites with a high within replicate variance as described in (Hoffman et al., 2014), using a signal-to-noise ratio of 15. Then all metabolomics data were log-transformed. These control measures were executed separately for each time point.

Next, to ensure we were analyzing the same metabolites across the three replicate time points, we used the R package `xMSanalyzer` (Uppal et al., 201b), which uses both the mass to charge (*m/z*) ratio and the column retention time to combine the same metabolites from different datasets. Metabolites were considered identical across the time points if they had *m/z* ratios within 10 parts per million of each other and their retention times varied by less than 10 seconds.

To determine the effects of age on metabolite intensity we used a general linear model controlling for the effects of both sex and weight on metabolite intensity. For this analysis, each metabolite within each time point was run individually. Metabolites can show heteroscedasticity due to increasing variances in older animals. To correct for this potential problem, we first ran a Breuch Pagan test on our linear model to test for heteroscedasticity. If none was found, we continued using our linear model, and if heteroscedasticity was discovered, we controlled for it by using a covariance matrix of the

linear model employed in the `lmtest` package (Zeileis and Hothorn, 2002). To correct for multiple comparisons, we used a conservative false discovery rate (FDR) (Benjamini and Hochberg, 1995a) at $\alpha = 0.01$. We then determined the number of sampling time points at which each metabolite appeared to be correlated with age.

Among metabolites associated with age within each time point, we looked for specific metabolic pathways that were overrepresented using the pathway enrichment program *mummichog* (Menni et al., 2013). *Mummichog* annotates *m/z* features and determines specific metabolic pathway enrichment using the MetaCyc database (Caspi et al., 2014). In particular, we set *mummichog* to query the "human" reference metabolic pathway, as it is the mostly closely related organism to the marmoset for which metabolite profiles and pathways have been well defined. P-values are calculated from a permutation test. For this analysis, we considered pathways significantly enriched for a factor if the *mummichog*-adjusted p-value was less than 0.05.

To determine if the changes observed across ages within a time point continued across time within each individual, we sought metabolites with significant longitudinal changes in intensity across all time points. The longitudinal analysis was carried out by implementing a random effects model in the `nlme` package (Pinherio et al., 2012), with age, sex, and weight as fixed effects and individual as a random effect. Thus, we wanted to determine if the changes with age continued across time within each individual. For this analysis, each individual sample was centered to a mean of zero to be able to better compare the same metabolites across different time points. To correct for multiple comparisons we again used an FDR of 0.01. *Mummichog* was then run on the significant

metabolites to determine if the same pathways were enriched for age as when the time points were run individually.

After running these analyses on individual metabolites, we then ran both unsupervised (principal components analysis) and supervised (partial least squares discriminant analysis) clustering analysis on each time point individually. For these analyses only, samples with missing data were converted to zero. First, we ran principle components analysis (PCA) using the `made4` package (Culhane et al., 2005) and determined if the principle components axes were correlated with age. Then, we used partial least squares discriminate analysis (PLSDA) as implemented in the `DiscriMiner` package (Sanchez and Determan, 2103) to discover if metabolomic profiles could predict the response of age in each time point. For this analysis the sexes and time points were run separately. For each group, we used 200 randomly drawn metabolites, and set the PLS component number to 10. For each group, this analysis was repeated ten times. We then repeated this analysis using randomly permuted ages to control for the fact that PLSDA might predict age even for randomly permuted data, given the large number of metabolites.

RESULTS

After quality control and combining of similar metabolites across time points, we were left with 2104 metabolites found in each time point to use for further statistical analyses. Of these metabolites, we were able to putatively annotate only 311 using the program *mummichog*.

We first identified those metabolites that changed with age within each individual time point. Different numbers of metabolites were found to vary with age when controlling

for weight and sex across the three time points (615 metabolites in June 2012, 885 in October 2012, and 171 in November 2013). Two specific metabolites are shown in Figure 4.2. Fifty-six metabolites were significantly affected by age in all three time points. However, they did not necessarily change in the same direction. Eight metabolites were found to always increase with age, while two were found to decrease with age in each time point (Figure 4.3). This left 46 metabolites that while significantly affected by age their slopes were in different directions across the three time points. Interestingly, of these 46 metabolites, over half (27) showed a pattern of increasing with age in the October 2012 and November 2013 time points while decreasing with age in the June 2012 time point.

At the level of individual metabolites, surprisingly few showed consistent age-related correlations across time points. However, at the pathway level, our *mummichog* pathway enrichment did find similar pathways involved in the changes with age. We were able to find a total of 28 metabolic pathways that were significantly enriched for metabolites that changed with age in at least one time point. Among these, six metabolic pathways that were enriched for age related metabolites in each of the June 2012 and October 2012 time points (Table 4.1). Only one metabolic pathway was found to be significantly changing with age in the November 2013 time point.

We then ran a random effects model on all three time points combined; using individual as a random effect, we found 382 metabolites that significantly changed with age (example in Figure 4.4). These metabolites were significantly enriched for 26 different metabolic pathways (Table 4.1). Of these 26 pathways, five were found to also be associated with age in the June 2012 and October 2012 time points individually. One metabolic pathway of the five (arsenate detoxification I) was found to be associated with

age within each of the three time points as well as the longitudinal data (across all three time points). Two of these five pathways are directly involved in oxidative stress (ubiquinol biosynthesis and lipoate biosynthesis). Xenobiotic metabolism had one significant pathway (arsenate detoxification) in June 2012, October 2012, and the longitudinal data, while a second xenobiotic pathway (nicotine degradation) was significant in June 2012 and the longitudinal analysis.

After our analysis of individual metabolites we ran global unsupervised and supervised clustering analyses on each time point individually. In our unsupervised principle components analysis, we found that PC1 in both the June 2012 and October 2012 time points was significantly associated with age (Figure 4.5), and explained very similar proportions of the variance at both time points (13.7% and 13.9% respectively). For the November 2013 time point PC5 was significantly correlated with age (3.4% of the variance explained). Our PLSDA shows that the metabolome can accurately predict the age of an animal. One example, from June 2012 females, is shown in Figure 6. R-squared values for observed versus predicted ages varied from 0.60±0.02 (mean ± 1 s.e., November 2013 males) to 0.82±0.02 (November 2013 females). For permuted age data, r-squared values varied from 0.38±0.05 (June 2012 females) to 0.58±0.05 (Nov 2013 females). In all cases, PLSDA was able to fit age with far greater accuracy for actual data than for randomly permuted age data (Figure S4.1).

DISCUSSION

We have shown that high-sensitivity metabolomic analysis of marmoset blood samples reveals many metabolites significantly associated with age. The metabolome of the

marmoset is highly variable, and finding repeatability across different time points can be difficult. However, by using metabolic pathway enrichment analyses we have been able to find specific metabolic pathways that were associated with age across all time points, and many of these metabolic pathways were also found in our random effects longitudinal model.

Many of the pathways significantly correlated with age have been found to be associated with age in other organisms. Metabolism related to monoamine biosynthesis (i.e. catecholamines and dopamine) has previously been implicated in aging flies (Hoffman et al., 2014), and it has long been known that changes in dopamine and catecholamines change dramatically with age in rhesus macaques (Goldmanrakic and Macbrown, 1981). While the monoamine result was not completely repeatable across the time points, our results combined with previous research suggests there maybe some evolutionary conservation of metabolic pathways that are influenced by aging. Future studies need to test perturbations of these metabolites and metabolic pathways to determine their direct effect on aging.

While the variation across marmosets and time points was often large, we were able to more accurately distinguish metabolites that significantly change with age by combining all the time points in a random effects model. We found five metabolic pathways associated with age across two time points (June 2012 and October 2012) and also in the longitudinal analysis. One of these pathways was also significantly enriched in the November 2013 dataset (arsenate detoxification).

Interestingly, the most repeatable metabolic pathways found to be associated with age in the marmoset are involved in oxidative stress, a major area of aging research. Both

ubiquinol biosynthesis and lipoate biosynthesis are implicated in mitochondrial function with ubiquinol being involved with mitochondrial electron transport while the lipoate metabolic pathway is directly involved in oxidative stress resistance. Previous work in aged rats has shown that supplementation of lipoic acid can reduce oxidant production in the heart to levels seen in young rats (Suh et al., 2001), and ubiquinol ratios have been suggested to be an accurate representation of oxidative stress in humans (Yamashita and Yamamoto, 1997). These results suggest that the changes in the ability to cope with oxidative stress may be major contributing factors of aging in the marmoset.

Xenobiotic metabolism also appears to be a largely associated with the aging metabolome in the marmoset. Both arsenate detoxification and nicotine degradation (which was significant in June 2012 and the longitudinal study) are involved in the xenobiotic metabolic pathway, and previous research in mice has shown that longer-lived strains exhibit increases in expression of genes linked to xenobiotic metabolism (Amador-Noguez et al., 2007; Steinbaugh et al., 2012). Furthermore, there is evidence that the breakdown of xenobiotics can lead to the production of free radicals (Chignell, 1985), which in turn might accelerate aging. Therefore, in the marmosets it appears changes in the ability to counter oxidative stress and break down the oxidative stressors themselves maybe fundamental factors involved in the aging process.

Strikingly, of the 56 metabolites that are significantly affected by age in all three time points, close to half (27) show a pattern of common slopes in October 2012 and November 2013, and opposite slopes in June 2012 (Figure 4.3). This potentially could be caused by a seasonal effect in the animals; however, to test this striking time-by age hypothesis additional time points are required. Surprisingly, even though we find

significant changes between metabolite concentrations in the different time points, the underlying enriched pathways are relatively consistent across time points, suggesting that changes in metabolic pathways may be more important biologically than the individual metabolites.

There was relatively little overlap among time points in the specific metabolites associated with age. However, our unsupervised principal component analysis suggests some repeatability in the importance of age in the metabolome. For both June 2012 and October 2012, the first principal component (PC) was significantly correlated with age. Our PCA analysis for the November 2013 dataset also found the metabolome to be associated with age, but only for the fifth PC, which explained a much smaller portion of the overall variance in the metabolome. Overall, our PCA analysis correlates nicely with our individual metabolite analysis, where more metabolites were found to be significantly associated with age in the June 2012 and October 2012 time points as opposed to the November 2013 sample. Interestingly, the June 2012 dataset shows a distinct pattern between young (<9 years) and old marmosets. This suggests there could be an environmental difference between the young and old animals such that something happened to the old marmosets when they were younger that did not affect the young marmosets. However, this same pattern is not observed in either of the later time points, so this environmental explanation might be incorrect, and the pattern in the June 2012 time point is caused by something unknown to us.

Our PLSDA shows that metabolomic profiles can fairly accurately predict age of an individual marmoset. This suggests the metabolome has the potential to be an ideal biomarker of aging. Previous studies have suggested the epigenome is a powerful

biomarker of aging (Hannum et al., 2013; Horvath, 2013), and our analysis, as well of previous work in flies (Hoffman et al., 2014), present the metabolome as potentially a superior biomarker of aging. Metabolomic profiling is both cheaper and easier than methylome assays, and metabolomic results have the potential to influence specific targeted interventions that cannot be accomplished with the epigenome. To determine the power of the metabolome as a biomarker, it will be interesting to note in the future, if those marmosets whose predicted age is younger than actual age, live longer than the rest, and vice versa for those animals with a predicted age higher than biological age.

While we were able to analyze thousands of metabolites and find hundreds that significantly change with age, there were several limitations to this study. First, metabolomic profiles are known to change on a daily basis and even cyclically within the day (Queiroz, 1974). However, we could not draw marmoset samples on the same day much less the same time of day. These daily changes could explain, in part, the large variances seen in metabolite concentrations even within animals of the same age, which might have limited our ability to detect metabolites that change with age. However, this does suggest that those metabolites we did find to change with age are probably due to actual signal and not background noise, making our findings all the more impressive, given what we assume to be a large number of unmeasured variables affecting metabolite levels in our samples.

Second, blood samples were drawn while the primates were anesthetized with ketamine, which could potentially disrupt normal metabolomic profiles. While the effects of ketamine on blood metabolomics are unknown, previous research in macaques has shown that ketamine does not change blood hormones levels and has less pronounced effects than

other forms of anesthesia (Zaidi et al., 1982). This suggests that while ketamine probably does have some unknown effects on the metabolome, its overall metabolomic effects should be fairly minimal. To determine the actual effects of ketamine, we need to compare populations with blood samples drawn with and without ketamine as an anesthetic.

Next, the parameters we used to combine metabolites across time points were conservative (especially the retention time). Each time point had over 10,000 metabolites left after quality control, but only 2104 could be conservatively said to be the same across all three time points, which suggests that we probably discarded many metabolites that were actually found in all three datasets. If we had utilized more metabolites for the study, we might have been better able to find metabolic pathways that changed with age repeatedly across all time points.

Finally, our study was limited by the lack of metabolite annotation matches available. Very little is known about specific marmoset metabolites, so we used the human metabolome as the reference. However, *mummichog* was only able to annotate a small proportion of the metabolites used in this study (~15%). The lack of annotation of all 2104 metabolites used in the study coupled with the few metabolites found to be associated with age in the November 2013 time point might be the main reason we were unable to determine metabolic pathways associated with age in that specific time point. Our analyses suggest that better annotation of the marmoset metabolome is needed to more accurately determine the effects of natural aging.

While we find large effects of xenobiotic metabolism, we cannot discount the idea that some of the metabolites may have been misannotated. Nicotine degradation and arsenate detoxification are both results of compounds that we would not normally expect

to be present in marmosets. However, we are utilizing high sensitivity metabolomics, which has the capability to detect more rare metabolites than targeted analyses. If any of the marmoset caretakers were smokers, then the marmosets may have indirectly been exposed to some nicotine products, which could be detected in our high sensitivity analysis, and arsenic can be found naturally in some groundwater. Therefore, there is a real possibility that these results are true signals in the marmoset metabolome.

Here we have presented the first large-scale longitudinal metabolomics study in a non-human primate. The metabolome of the marmoset is highly variable and finding repeatability across time points proved difficult, yet we were still able to determine many metabolic pathways that are associated with natural aging. Many of the most repeatable metabolic pathways were involved in oxidative stress resistance in some capacity, suggesting oxidative stress may be a major contribution to marmoset aging. We believe longitudinal studies are underutilized as a method for determining changes in metabolites and metabolic pathways that are involved in the aging process, and future metabolomics studies should try to incorporate multiple measurements of the same individuals. The marmosets in this colony will continue to be followed, and as the animals age and die, we will hopefully be able to determine long-term changes in metabolomic profiles that are predictive of risk of morbidity and mortality.

ACKNOWLEDGEMENTS

This work was funded in part by NIH grant AG038746 to DPJ and DELP. JMH was supported in part by NIH training grant T32 GM007103.

LITERATURE CITED

Amador-Noguez, D., Dean, A., Huang, W.D., Setchell, K., Moore, D., and Darlington, G. (2007). Alterations in xenobiotic metabolism in the long-lived Little mice. Aging Cell *6*, 453-470.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., Holland, T.A., Keseler, I.M., Kothari, A., Kubo, A.*, et al.* (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res *42*, D459-D471.

Chignell, C.F. (1985). Structure Activity Relationships in the Free-Radical Metabolism of Xenobiotics. Environ Health Persp *61*, 133-137.

Colditz, G.A., and Hankinson, S.E. (2005). The Nurses' Health Study: Lifestyle and health among women. Nat Rev Cancer *5*, 388-396.

Culhane, A.C., Thioulouse, J., Perriere, G., and Higgins, D.G. (2005). MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics *21*, 2789-2790.

Ferrucci, L. (2008). The Baltimore Longitudinal Study of Aging (BLSA): A 50-Year-Long Journey and Plans for the Future. J Gerontol a-Biol *63*, 1416-1419.

Fischer, K.E., and Austad, S.N. (2011). The Development of Small Primate Models for Aging Research. Ilar J *52*, 78-88.

Fuchs, S., Bundy, J.G., Davies, S.K., Viney, J.M., Swire, J.S., and Leroi, A.M. (2010). A metabolic signature of long life in Caenorhabditis elegans. Bmc Biol *8*.

Go, Y.-M., Uppal, K., Walker, D.I., Dury, L., Strobel, F.H., Baudichon-Cortay, H., Roede, J.R., and Jones, D.P. (2014). Mitochondrial Metabolomics Using High-Resolution Fourier-Transform Mass Spectrometry, in Mass Spectrometry in Metabolomics Methods and Protocols. In Methods in Molecular Biology, D. Raftery, ed.

Goldmanrakic, P.S., and Macbrown, R. (1981). Regional Changes of Monoamines in Cerebral-Cortex and Sub-Cortical Structures of Aging Rhesus-Monkeys. Neuroscience *6*, 177-187.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y.*, et al.* (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Mol Cell *49*, 359-367.

Higley, J.D., Suomi, S.J., and Linnoila, M. (1992). A Longitudinal Assessment of Csf Monoamine Metabolite and Plasma-Cortisol Concentrations in Young Rhesus-Monkeys. Biol Psychiat *32*, 127-145.

Hoffman, J.M., Soltow, Q.A., Li, S.Z., Sidik, A., Jones, D.P., and Promislow, D.E.L. (2014). Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, Drosophila melanogaster. Aging Cell *13*, 596-604.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol *14*.

Houtkooper, R.H., Argmann, C., Houten, S.M., Canto, C., Jeninga, E.H., Andreux, P.A., Thomas, C., Doenlen, R., Schoonjans, K., and Auwerx, J. (2011). The metabolic footprint of aging in mice. Sci Rep-Uk *1*.

Johnson, J.M., Yu, T.W., Strobel, F.H., and Jones, D.P. (2010). A practical approach to detect unique metabolic patterns for personalized medicine. Analyst *135*, 2864-2870.

Kuehnel, F., Grohmann, J., Buchwald, U., Koeller, G., Teupser, D., and Einspanier, A. (2012). Parameters of haematology, clinical chemistry and lipid metabolism in the common marmoset and alterations under stress conditions. J Med Primatol *41*, 241-250.

Liu, J.P., Wang, D., Chen, Y.N., Sun, H.J., He, S.R., Wang, C.S., Yang, G., Shi, M.M., Zhang, J., Ren, Y.*, et al.* (2013). H-1 NMR-based metabonomic analysis of serum and urine in a nonhuman primate model of diabetic nephropathy. Mol Biosyst *9*, 2645-2652.

Menni, C., Kastenmuller, G., Petersen, A.K., Bell, J.T., Psatha, M., Tsai, P.C., Gieger, C., Schulz, H., Erte, I., John, S.*, et al.* (2013). Metabolomic markers reveal novel pathways of ageing and early development in human populations. Int J Epidemiol *42*, 1111-1119.

Muehlenbein, M.P., Campbell, B.C., Richards, R.J., Svec, F., Phillippi-Falkenstein, K.M., Murchison, M.A., and Myers, L. (2003). Dehydroepiandrosterone-sulfate as a biomarker of senescence in male non-human primates. Exp Gerontol *38*, 1077-1085.

Nishijima, K., Saitoh, R., Tanaka, S., Ohsato-Suzuki, M., Ohno, T., and Kitajima, S. (2012). Life span of common marmoset (Callithrix jacchus) at CLEA Japan breeding colony. Biogerontology *13*, 439-443.

Patterson, A.D., Bonzo, J.A., Li, F., Krausz, K.W., Eichler, G.S., Aslam, S., Tigno, X., Weinstein, J.N., Hansen, B.C., Idle, J.R.*, et al.* (2011). Metabolomics Reveals Attenuation of the SLC6A20 Kidney Transporter in Nonhuman Primate and Mouse Models of Type 2 Diabetes Mellitus. J Biol Chem *286*, 19511-19522.

Pinherio, J., Bates, D., DebRoy, S., Sarkar, D., and Team, R.C. (2012). nlme: Linear and Nonlinear Mixed Effect Models. R package version 31-118.

Queiroz, O. (1974). Circadian-Rhythms and Metabolic Patterns. Annu Rev Plant Phys *25*, 115-134.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing *Vienna, Austria*.

Roede, J.R., Uppar, K., Yongliang, L., Promislow, D.E.L., Wachtman, L.M., and Jones, D.P. (2012). Characterization of plasma thiol redox potential in a common marmoset model of aging. Redox Biology *1*, 387-393.

Sanchez, G., and Determan, C. (2103). DiscriMiner: Tools of the Trade for Discriminant Analysis. R package version 01-29.

Soltow, Q.A., Strobel, F.H., Mansfield, K.G., Wachtman, L., Park, Y., and Jones, D.P. (2013). High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. Metabolomics *9*, S132-S143.

Steinbaugh, M.J., Sun, L.Y., Bartke, A., and Miller, R.A. (2012). Activation of genes involved in xenobiotic metabolism is a shared signature of mouse models with extended lifespan. Am J Physiol-Endoc M *303*, E488-E495.

Suh, J.H., Shigeno, E.T., Morrow, J.D., Cox, B., Rocha, A.E., Frei, B., and Hagen, T.M. (2001). Oxidative stress in the aging rat heart is reversed by dietary supplementation with (R)-alpha-lipoic acid. Faseb J *15*, 700-706.

Tardif, S.D., Mansfield, K.G., Ratnam, R., Ross, C.N., and Ziegler, T.E. (2011). The Marmoset as a Model of Aging and Age-Related Diseases. Ilar J *52*, 54-65.

Uppal, K., Soltow, Q.A., Strobel, F.H., Pittard, W.S., Gernert, K.M., Yu, T.W., and Jones, D.P. (2013). xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. Bmc Bioinformatics *14*.

Yamashita, S., and Yamamoto, Y. (1997). Simultaneous detection of ubiquinol and ubiquinone in human plasma as a marker of oxidative stress. Anal Biochem *250*, 66-73.

Yu, T.W., Park, Y., Johnson, J.M., and Jones, D.P. (2009). apLCMS-adaptive processing of high-resolution LC/MS data. Bioinformatics *25*, 1930-1936.

Zaidi, P., Wickings, E.J., and Nieschlag, E. (1982). The Effects of Ketamine Hcl and Barbiturate Anesthesia on the Metabolic-Clearance and Production-Rates of Testosterone in the Male Rhesus-Monkey, Macaca-Mulatta. J Steroid Biochem *16*, 463-466.

Zeileis, A., and Hothorn, T. (2002). Diagnostic checking in regression relationships. R News *2*, 7-10.

FIGURE LEGENDS

Figure 4.1. Age distributions for each of three time points. Y-axis represents rankings of marmosets from oldest to youngest for females (left) and males (right). Each point on the x-axis represents the time points from which blood serum samples are taken (1-3 time points per animal).

Figure 4.2. Plot of two specific metabolites correlation with age from the June 2012 time point.

Figure 4.3. Plot of three individual metabolites across three time replicates. Rows correspond to specific metabolites whose *m/z* ratio is shown on the y-axis. Columns

correspond to the three time points analyzed. Each plot shows a different time point with a significant change with $P<1x10^{-3}$.

Figure 4.4. Specific metabolite with longitudinal changes with age. Each black line represents the changes in metabolite concentration across an individual marmoset. Red line is the overall linear model of all data.

Figure 4.5. Plot of PCs significantly associated with age. June 2012 and October 2012 are both PC1 while November 2013 is PC5. P-values< $3x10^{-9}$.

Figure 4.6. Sunflower plot of PLSDA observed vs. predicted ages. Points represent individual marmosets from June 2012. Red sunflowers indicate multiple marmosets based on the number of leaves shown.

Figure S4.1. Differences in variances explained in different time points for observed (orange) and randomized (blue) data. Males are represented on the left with females on the right.

Table 4.1. Metabolic pathways enriched for age in each time point individually and all combined for the longitudinal analysis. Pathways shown in bold have significant age effects in June 2012, October 2012, and the longitudinal analysis.

Pathway	June-2012	October-2012	November-2013	Longitudinal
Ubiquinol-10 biosynthesis	**1.94E-04**	**4.82E-03**		**8.42E-05**
Nicotine degradation IV	2.51E-03			8.42E-05
Oxidized GTP and dGTP detoxification	9.08E-03			1.17E-04
GDP-glucose biosynthesis	1.18E-03			2.29E-04
Nicotine degradation III	2.62E-02			5.65E-04
lipoate biosynthesis and incorporation II	**3.40E-03**	**3.14E-02**		**7.25E-04**
GDP-L-fucose biosynthesis II (from L-fucose)	3.40E-03			7.25E-04
Xanthine and xanthosine salvage		3.14E-02		7.25E-04
Biotin-carboxyl carrier protein assembly				7.25E-04
CMP-N-acetylneuraminate biosynthesis I (eukaryotes)		3.14E-02		7.25E-04
Phenylalanine degradation IV (mammalian via side chain)	5.05E-03			8.84E-04
(S)-reticuline biosynthesis II	5.05E-03			8.84E-04
Amino acids	**1.16E-03**	**2.74E-02**		**1.64E-03**
arsenate detoxification I (glutaredoxin)	**9.08E-03**	**5.50E-03**	**5.69E-03**	**1.91E-03**
Methionine degradation I (to homocysteine)	9.08E-03			1.91E-03
Salvage pathways of pyrimidine ribonucleotides	9.08E-03			1.91E-03
tRNA splicing	9.08E-03			1.91E-03
UDP-N-acetyl-D-galactosamine biosynthesis II	9.08E-03			1.91E-03
GDP-mannose biosynthesis	9.08E-03			1.91E-03
Pyrimidine deoxyribonucleotides de novo biosynthesis				1.91E-03
Glycogen biosynthesis II (from UDP-D-Glucose)				1.91E-03
Catecholamine biosynthesis	1.93E-02			4.13E-03
Molybdenum cofactor biosynthesis	1.93E-02			4.13E-03
Zymosterol biosynthesis	**2.62E-02**	**1.99E-02**		**4.75E-03**
Dopamine degradation	3.47E-02			7.69E-03
Bile acid biosynthesis neutral pathway		1.21E-02		1.27E-02

Adenine and adenosine salvage III	3.40E-03	3.14E-02		
Selenocysteine biosynthesis II (archaea and eukaryotes)	3.40E-03			
Spermine and spermidine degradation I	1.93E-02			
NAD biosynthesis from 2-amino-3-carboxymuconate semialdehyde		5.50E-03		
Alpha-tocopherol degradation		3.14E-02		
CDP-diacylglycerol biosynthesis I		3.14E-02		
Trans/trans farnesyl diphosphate biosynthesis		3.14E-02		
Asparagine biosynthesis I		3.14E-02		

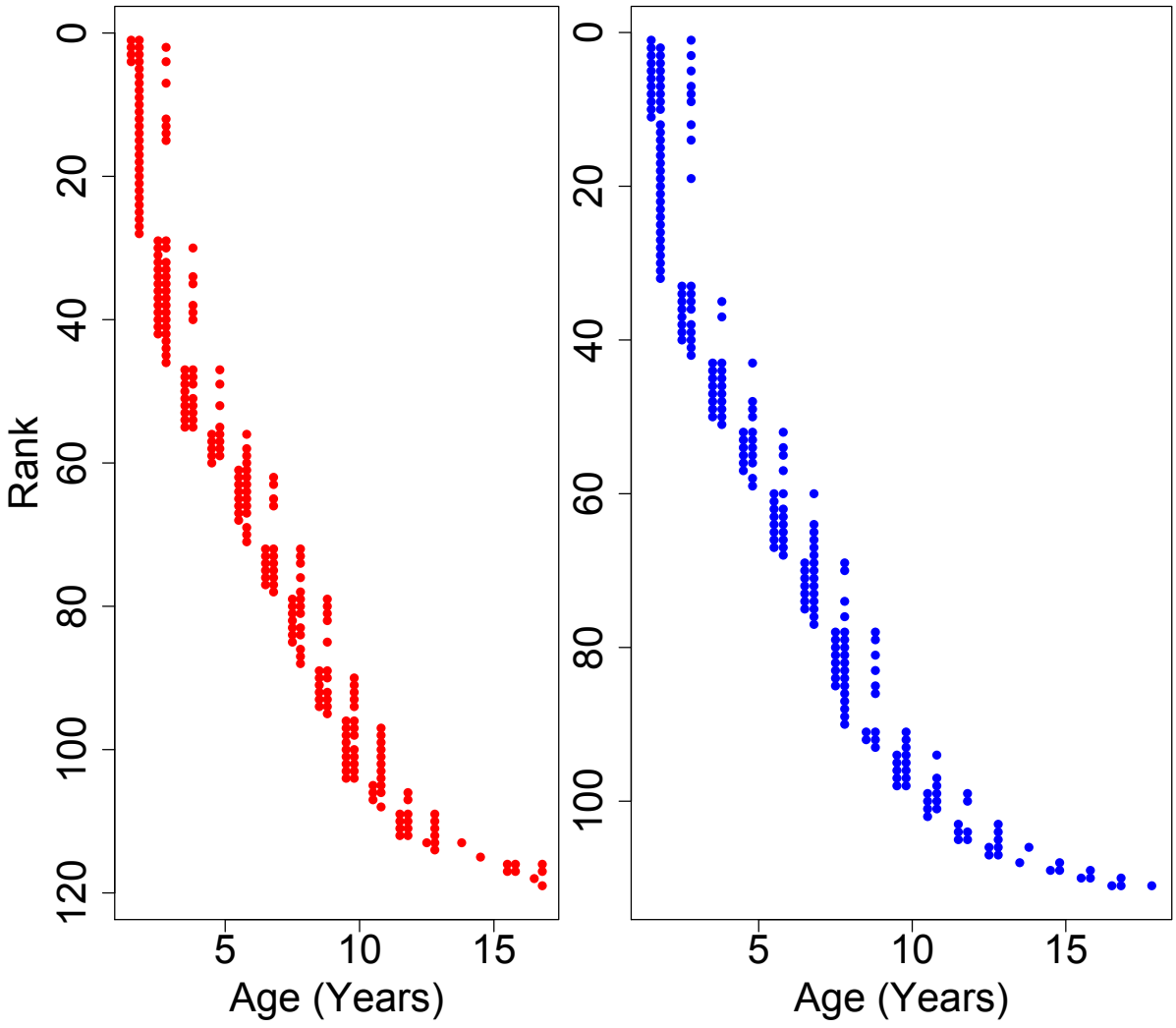
Rank
0
20
40
60
80
100
120
5
10
15
Age (Years)
0
20
40
60
80
100
5
10
15
Age (Years)

Figure 4.1

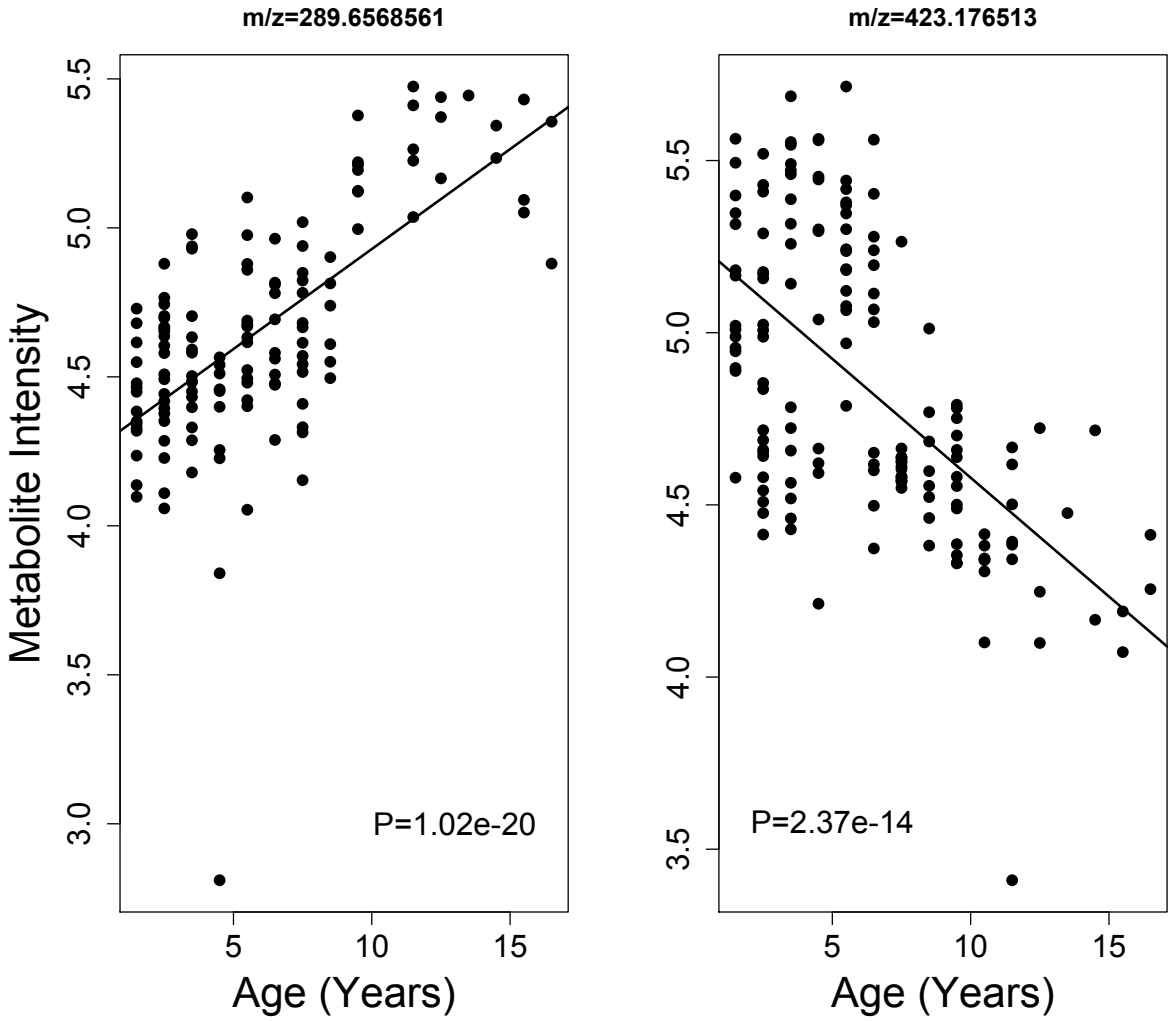
m/z=289.6568561
Metabolite Intensity
3.0
3.5
4.0
4.5
5.0
5.5
5
10
15
Age (Years)
P=1.02e-20
m/z=423.176513
3.5
4.0
4.5
5.0
5.5
5
10
15
Age (Years)
P=2.37e-14

Figure 4.2

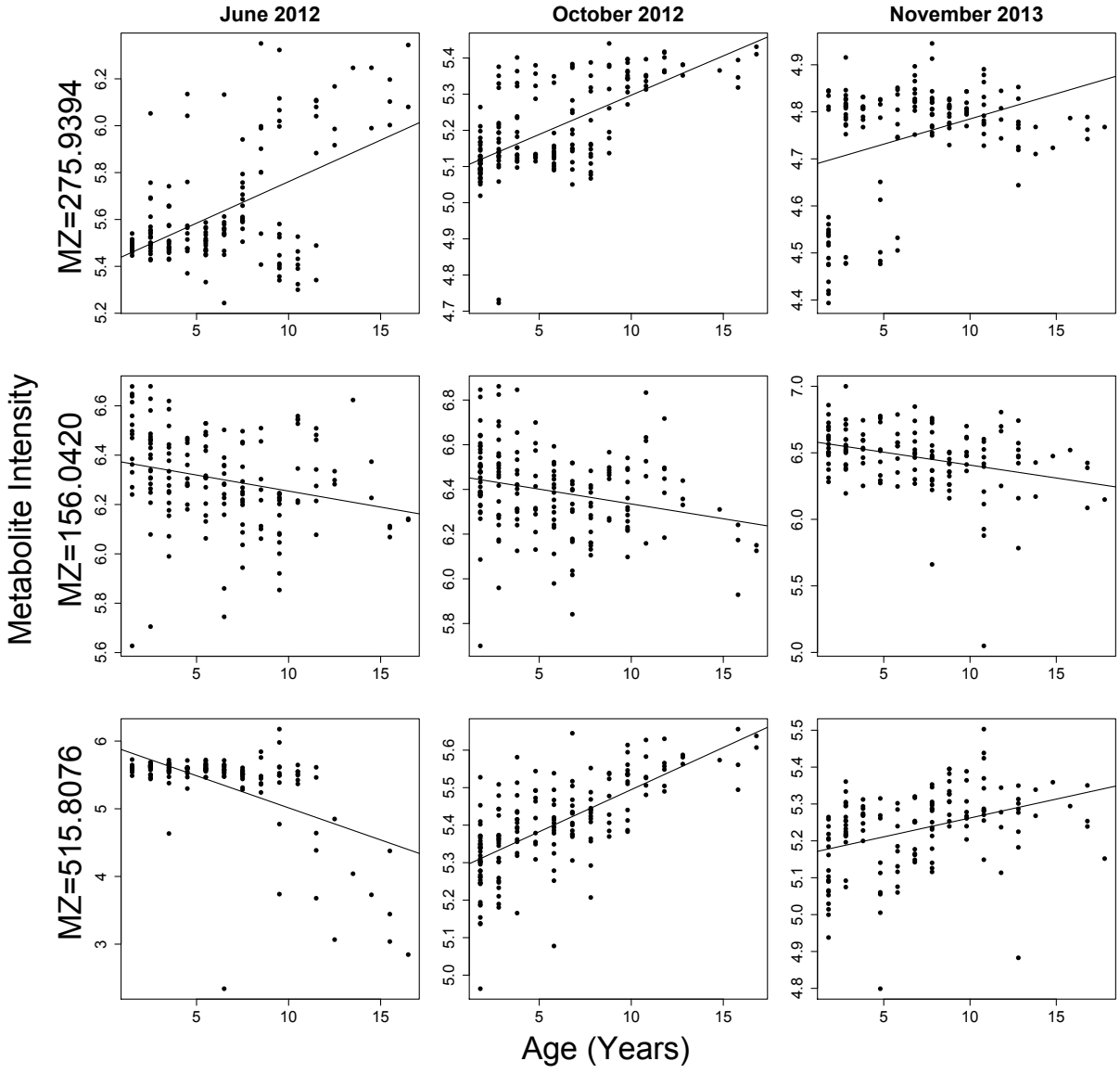
June 2012
October 2012
November 2013
MZ=275.9394
MZ=156.0420
MZ=515.8076
Metabolite Intensity
Age (Years)

Figure 4.3

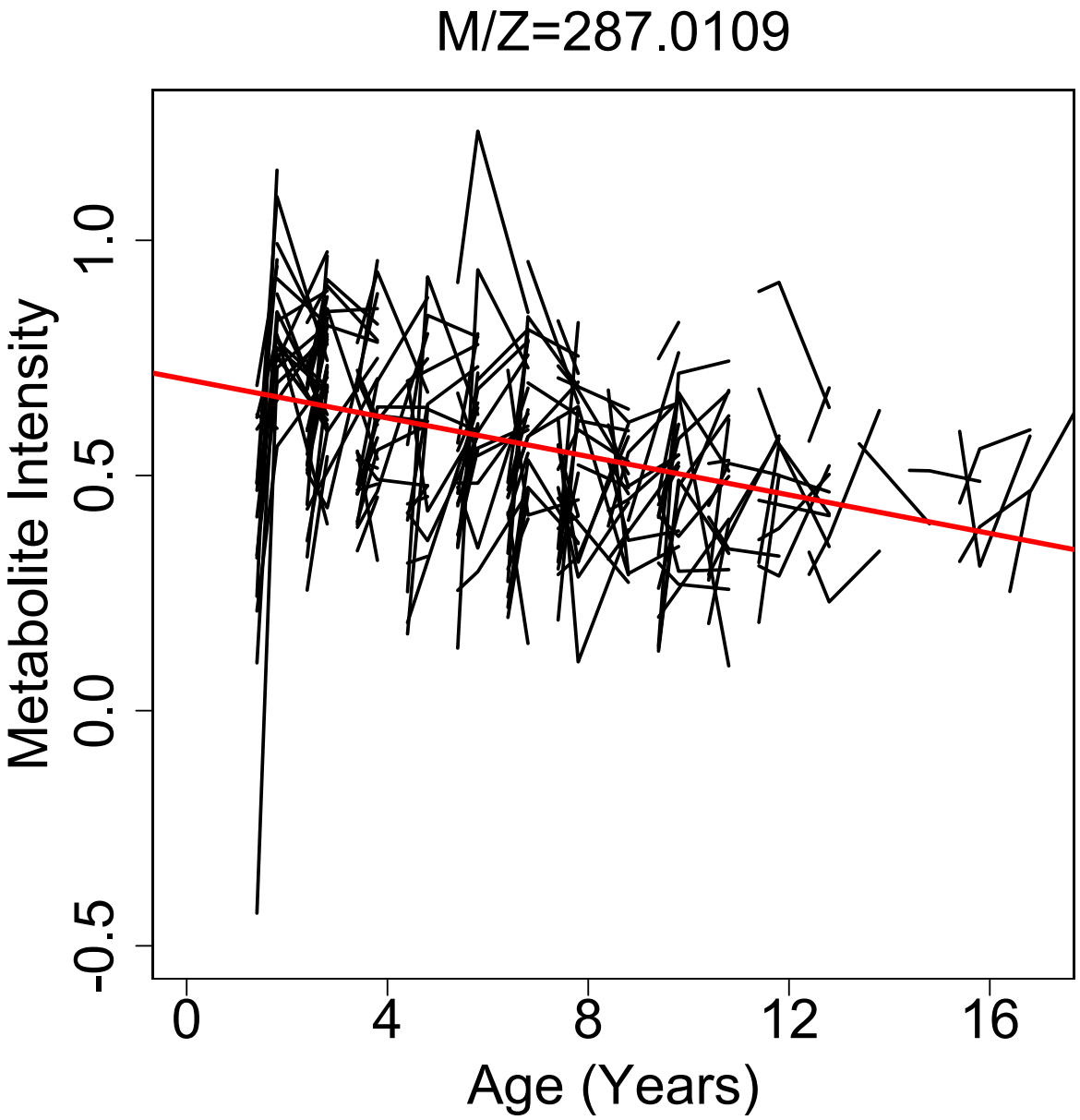
M/Z=287.0109
Metabolite Intensity
-0.5
0.0
0.5
1.0
0
4
8
12
16
Age (Years)

Figure 4.4

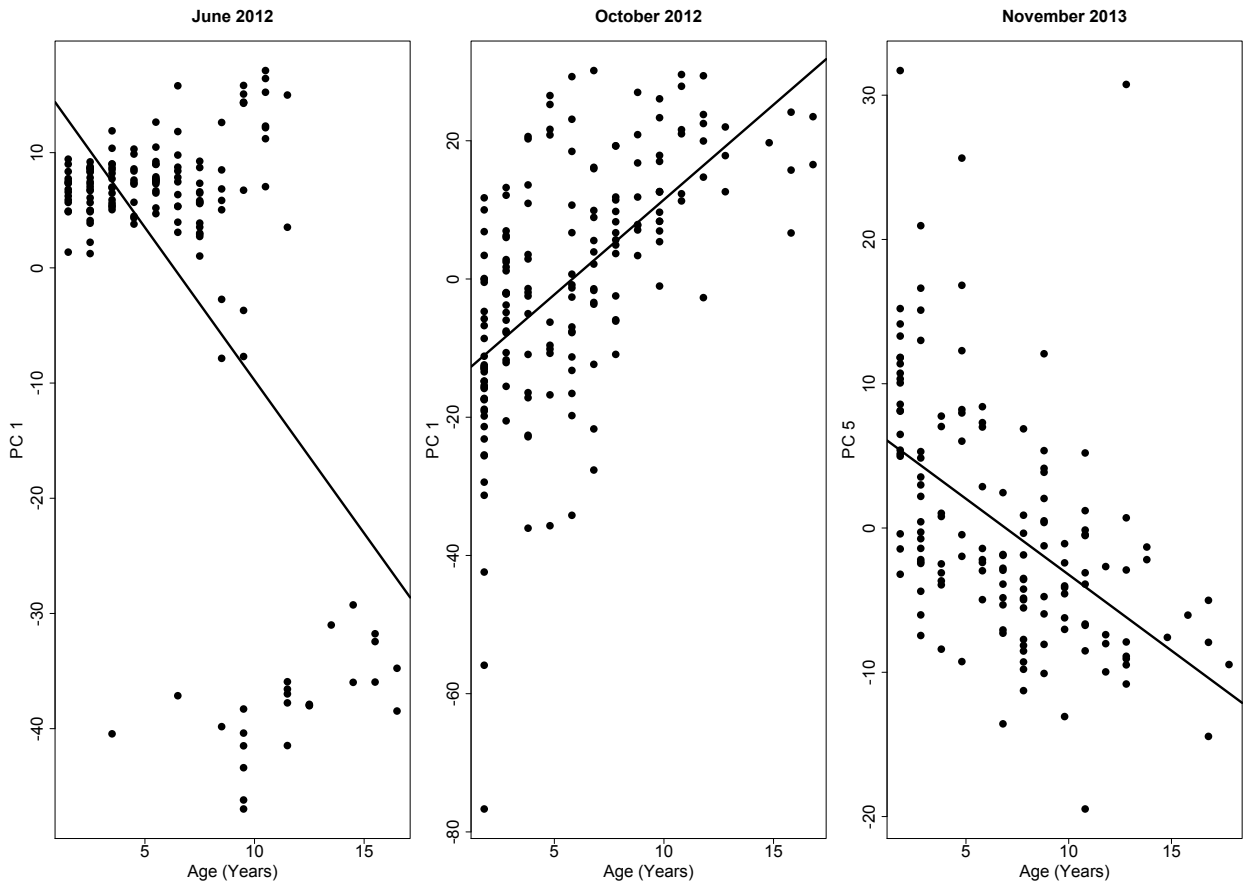
June 2012
October 2012
November 2013
PC 1
PC 1
PC 5
Age (Years)
Age (Years)
Age (Years)

Figure 4.5

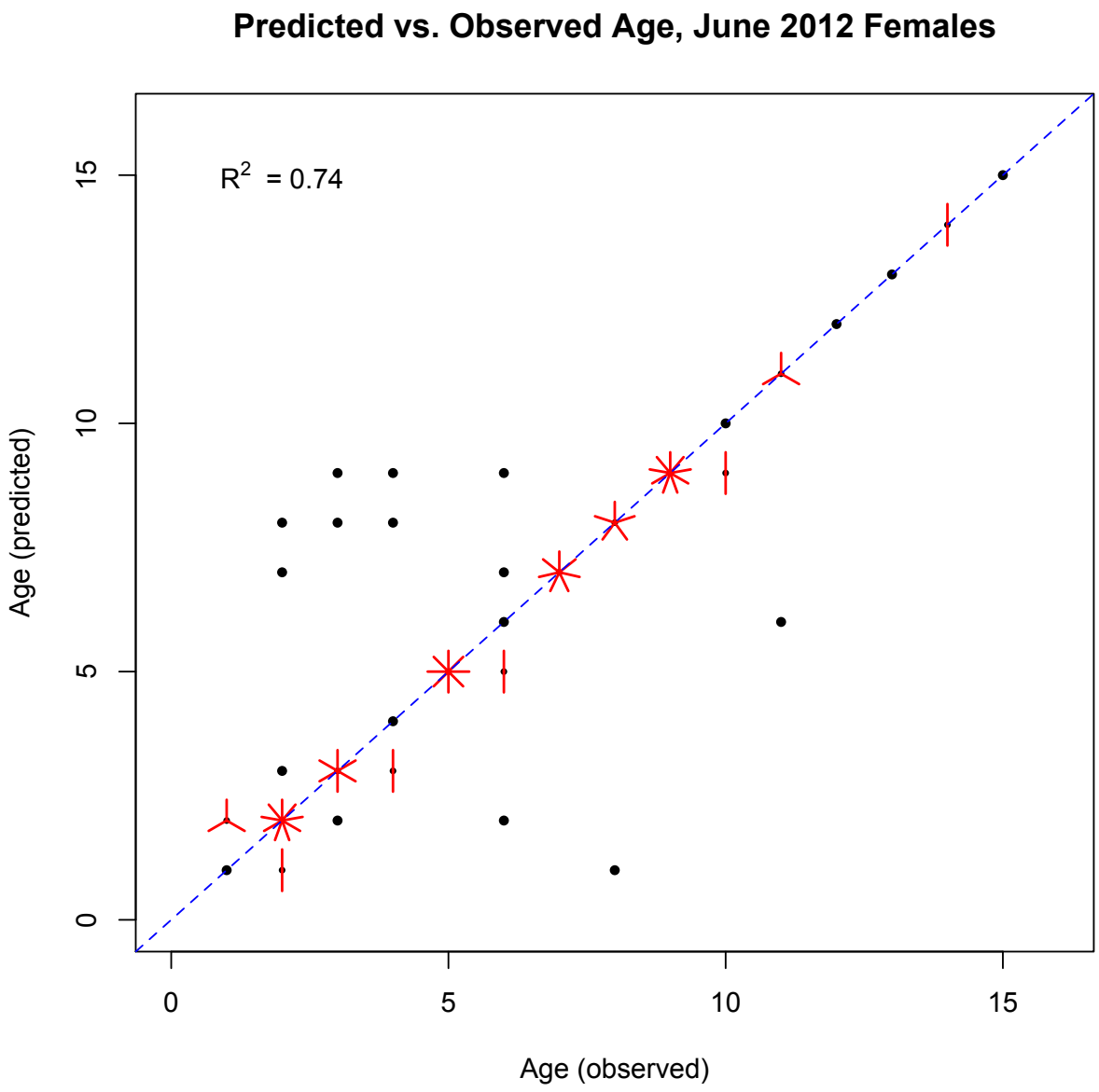
Predicted vs. Observed Age, June 2012 Females
R2 = 0.74
Age (predicted)
Age (observed)
0
5
10
15

Figure 4.6

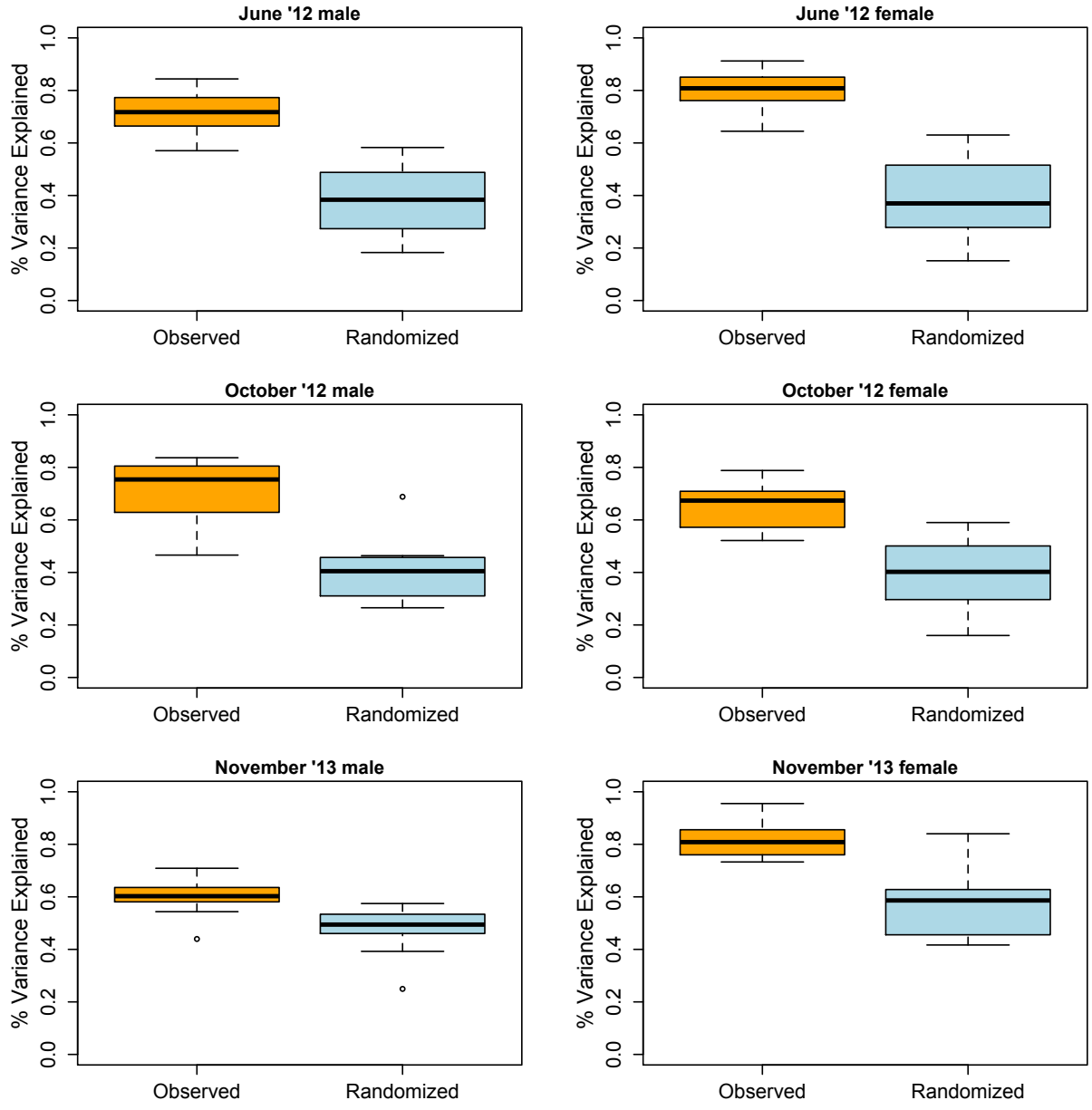
June '12 male
June '12 female
October '12 male
October '12 female
November '13 male
November '13 female
% Variance Explained
0.0
0.2
0.4
0.6
0.8
1.0
Observed
Randomized

Figure S4.1

CHAPTER 5

CONCLUSIONS

Ever since the human genome project was completed 15 years ago, scientists have eagerly jumped into using next generation technologies to determine the effects of the different "omics" on phenotypes of interest, including many studies on aging and longevity. However, while many of these "omics" have been extensively studied, metabolomics (the study of all the small molecules in an organism) has been largely ignored until the past couple of years. In this dissertation, I have attempted to determine the relationship between the metabolome and age and longevity within and across species. I have shown that individual metabolites can be strongly correlated with age, and I speculate that they might potentially influence the longevity of an organism. Morever, the entire metabolome can be a powerful predictor, and potential biomarker, of age.

In my first study, I discovered that the metabolome of *Drosophila melanogaster* is highly associated with age, sex, and genotype. This experiment was one of the first metabolomic analyses to use high sensitivity metabolomics to determine the effects of age on an extremely diverse array of metabolites in an organism. In addition to understanding how individual metabolites vary with age, I found several metabolic pathways associated with age, including some that were already known to be correlated with aging and longevity (monoamine neurotransmitters) and new metabolic pathways that previously not been implicated largely in aging (the carnitine shuttle). One of the most intriguing findings was the different interaction effects with age. First, the effects of age on hundreds of metabolites differed between the sexes.

This suggests that future aging work should look at metabolomic effects of aging on both sexes because they may yield different results, and these results may lead to new insights into why we see differences in longevity between the sexes. Also, we found many age by genotype interactions. Most aging studies focus on only one specific genotype of a species. However, my results indicate genetic variation can greatly affect the metabolome of an organism, suggesting future studies need to incorporate multiple genotypes in aging studies to determine if metabolomic changes with age are relatable to the entire species or are genotype specific. While no actual genetic mapping was done in this study due to insufficient numbers of genotypes, the large numbers of age by genotype interactions suggest that future mapping of metabolites with these interactions might lead to new candidate genes that directly affect aging.

Next, I expanded my earlier work to understand the metabolomic consequences of aging across different species. I have completed one of the first cross species comparative metabolomics studies to focus on the effects of longevity on the metabolome, as well as one of the largest comparative metabolomics studies ever completed. First, it is notable that unlike virtually all other comparative studies, I included multiple genotypes within each species. This enables me to distinguish the potential confounding effects of genetic variation from variation among species. This suggests future comparative studies need to incorporate multiple genotypes of the same species to more accurately determine what is a true species effect instead of a genotype effect. I found individual metabolites do not have a large amount of phylogenetic signal; however, the metabolome combined together is able to recapitulate some parts of the known species phylogeny. Finally, I discovered many metabolites associated with species, sex, and body weight, as well as many metabolites associated with longevity in a sex specific manor. This result suggests, as also seen in my work on *D. melanogaster*, future metabolomic studies

need to incorporate both sexes, as metabolomic profiles appear to be very sex specific, and the biochemical pathways that influence aging and longevity may be different between the sexes. More metabolomic studies on the sexes may also give us new insights into why females are often longer lived than males of the same species. I also discover many metabolic pathways associated with longevity and resistance to oxidative stress, and some of these are the same as in my previous study of only *D. melanogaster* (amino acids). This study only looked at flies from one time point, and some of my previous unpublished work suggests it might be trajectories of metabolites (not static values) that can most influence (or be predictive of) aging. Future comparative metabolomic studies of aging need to incorporate samples from different ages to fully understand the metabolomic consequences of aging.

In my final study, I use a translational model of human aging, the common marmoset, to look longitudinally to discover how the metabolome of an individual changes as it ages naturally. This is the first longitudinal metabolomics study to be completed in the marmoset, as well as the largest longitudinal metabolomics study completed to date. Interestingly, I find substantial evidence for oxidative stress associations with aging in these animals, giving support to the oxidative theory of aging. There also were large numbers of metabolites changing with age associated with xenobiotic metabolism, suggesting the ability of organisms to deal with external xenobiotics may be directly influencing the aging phenotype. I also find the metabolome of the marmoset is highly predictive of age, and as this study is ongoing, it will be interesting to discover if those marmosets whose metabolome predicts a younger age than their biological age live longer than other marmosets. I also hope to be able to predict metabolites associated with morbidity and mortality of these animals. The results of this study suggest to completely

understand the effects of natural aging on an individual, we need to employ longitudinal studies, and the metabolome may be the biomarker of age that has been elusive to researches for years.

Here I have shown the power of the metabolome to determine different metabolic factors that influence aging and longevity, and I have laid the groundwork for future studies to utilize the metabolome as a potential biomarker of aging. Future work needs to expand on this dissertation to begin to fully understand the metabolic causes and consequences of natural aging and longevity. With more metabolomic profiles for more *Drosophila melanogaster* genotypes, we will be able to complete genetic mapping on metabolites associated with age to discover new candidate aging genes that normal SNP mapping studies have not been able to find. Comparative studies have the potential to influences our understanding of the evolution of the metabolome and its association with life history traits. By understanding the biochemical pathways associated with life history evolution, we may be finally able to determine why large variation in longevity exists among related species as well as between the sexes of the same species. Finally, I suggest more longitudinal studies need to be completed to understand the metabolic progression of aging; while I will continue to work on the longitudinal marmosets study, more species need to be included in longitudinal studies to determine if changes seen in the marmosets are actually relatable to other species, both closely and distantly related.