THE EVOLUTION AND GENETIC BASIS OF COMPLEX TRAITS IN *DROSOPHILA MELANOGASTER*

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Z. FORREST ELKINS

Dr. Elizabeth G. King, Dissertation Supervisor

DECEMBER 2022

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

THE EVOLUTION AND GENETIC BASIS OF COMPLEX TRAITS IN *DROSOPHILA MELANOGASTER*

presented by Z. Forrest Elkins,
a candidate for the degree of Doctor of Philosophy,
and hereby certify that, in their opinion, it is worthy of acceptance.

_____
Dr. Elizabeth G. King

_____
Dr. Lauren Sullivan

_____
Dr. Rex Cocroft

_____
Dr. Greg Blomquist

For my family, both biological and chosen. Thank you.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Libby King for her patience, guidance, and encouragement. I joined the King lab one year into my time here at Mizzou, and Dr. King could not have been more welcoming and supportive. Whether it was troubleshooting code, executing a large phenotyping experiment, or writing a manuscript, she always knew exactly when to help and when to give me space. Dr. King's knack for mentorship is second to none.

I would not be where I am today without the help of Dr. Troy Zars and Dr. Marco Navarro. Dr. Navarro was the first to reach out to me when I was struggling to adapt to graduate school. He set up a meeting between Dr. Zars and myself that would lead me to Dr. King. And while I only briefly had the privilege to know Dr. Zars, he played no small part in kickstarting my academic career.

I would also like to thank the members of my committee: Dr. Reginald Cocroft, Dr. Lauren Sullivan, and Dr. Gregory Blomquist. Dr. Cocroft has always given me insight into the evolution of animal personality. Dr. Blomquist taught me the math behind population genetics. Dr. Sullivan's enthusiasm for my research, and for dispersal behavior in general, is truly contagious – talking to her about models of dispersal always left me feeling rejuvenated and ready to tackle my research.

I have received a great deal of encouragement and support during my time here. My friends, both local and digital, have always been there for me. Arroyo, Heron, Han, Erin, Em, Pax, Lou – thank you. R Shawn and Christian, thank you for being the best two roommates and friends a person could ask for. Levi and Kevin, Michael, R Shawn – I loved playing DnD with you. I would also like to thank current and former members of my lab: Dr. Enoch Ng'oma, Elizabeth Jones, Ronel Ghidey, Dr. Joe Gunn, Camille Oster, Esdras Tuyishimire, Victoria Hamlin, Dr. Patricka Williams-Simon, Dr. Anna Perinchery-Herman, Paul Petrowski and more.

Finally, I would like to thank my parents for raising me and giving guidance when I need it; my sister Liv for her steadfastness, makeup tips and shopping trips; my brother Sam for our many, *many* overlapping interests; my sister-in-law Haley for her empathy and wit; and Jo Moaton, my best friend. I love you all dearly. My partner, Dr. Arianne Messerman, the love of my life, has been the most encouraging and supportive partner a person could ask for. Spending time with you and our dogs, Ponyo and Raleigh, has soothed my soul more times than I could possibly count.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS……ii

LIST OF TABLES……vi

LIST OF FIGURES……vii

ABSTRACT……ix

CHAPTER 1 • INTRODUCTION

INTRODUCTION……1

LITERATURE CITED……5

CHAPTER 2 • THE GENETIC BASIS OF EXPLORATION BEHAVIOR IN A MULTIPARENT POPULATION OF *D. MELANOGASTER*

ABSTRACT……11

INTRODUCTION……12

METHODS……16

RESULTS……27

DISCUSSION……34

LITERATURE CITED……39

SUPPLEMENTAL MATERIALS……47

CHAPTER 3 • STARVATION RESISTANCE IN AN EXPERIMENTALLY EVOLVED POPULATION OF *D. MELANOGASTER*

ABSTRACT……48

INTRODUCTION……49

METHODS……53

RESULTS.....64

DISCUSSION.....76

LITERATURE CITED.....81

CHAPTER 4 • CODE APPENDIX.....88

VITA.....95

LIST OF TABLES

CHAPTER 2.

Table 1. Significant QTL table.....32

Supplemental table 1. List of recombinant inbred lines.....47

CHAPTER 3.

Table 1. Selection regime schedule.....55

Table 2. Diet recipes for selection experiment and starvation assay.....56

Table 3. Cox proportional hazard models for direct measurement table.....63

Table 4. Cox proportional hazard models for diet treatment table.....64

Table 5. Direct measurement Cox model results table.....68

Table 6. Direct measurement AIC model comparison table.....69

Table 7. Diet treatment Cox model results table.....74

Table 8. Diet treatment AIC model comparison table.....76

LIST OF FIGURES

CHAPTER 2.

Figure 1. Graphical overview of the exploration tendency bulk-segregant experimental design....20

Figure 2. Diagram of MCMC analysis....25

Figure 3. Proportion of exploring flies per recombinant inbred line....28

Figure 4. Linear model visualization of exploration predicted by time of day....29

Figure 5. G' values plotted across the fly genome with two significance thresholds....30

Figure 6. Allele frequency differences between explorers and non-explorers at the 13 significant G' QTL regions....31

Figure 7. Pairwise linkage disequilibrium between the 13 significant G' QTL regions....33

CHAPTER 3.

Figure 1. *Drosophila* Synthetic Population Resource population used in the selection experiment....53

Figure 2. Direct measurement assay design visualization....58

Figure 3. Diet treatment assay design visualization....59

Figure 4. Survival probability predicted by sex in the direct measurement experimental assay.....65

Figure 5. Survival probability predicted by selection regime in the direct measurement experimental assay.....66

Figure 6. Survival probability predicted by sex + selection regime in the direct measurement experimental assay.....67

Figure 7. Survival probability predicted by diet treatment + selection regime in the diet treatment experimental assay.....71

Figure 8. Risk predicted by sex in the diet treatment experimental assay.....72

Figure 9. Cumulative hazard predicted by diet treatment in the diet treatment experimental assay.....73

CHAPTER 4.

Figure 1. Variability in allele frequency estimations due to coverage.....93

Figure 2. Simulated allele frequency differences between exploring and non-exploring flies.....94

THE EVOLUTION AND GENETIC BASIS OF COMPLEX TRAITS IN *DROSOPHILA MELANOGASTER*

ABSTRACT

The current climate crisis, increasing habitat fragmentation, deforestation, the number of extreme weather events, contagious diseases, and more are major environmental stressors that threaten the survival of every species on Earth. It is necessary for organisms to evolve stress resistant traits and phenotypic plasticity to survive these catastrophic effects. To resist stress, organisms will need to invest some amount of energy, leading to trade-offs with other traits. In addition, organisms must develop behavioral adaptations to manage environmental stressors. Behaviors such as dispersal and migration, and stress-resistant traits such as starvation resistance, have evolved as a means of survival.

While researchers have studied behavioral adaptations and stress-resistant traits, the genetic and evolutionary mechanisms and phenotypic plasticity of these traits are not fully known. For my dissertation research, I studied the genetic mechanisms and phenotypic plasticity of two traits: exploration behavior and starvation resistance, respectively, in evolved multiparent populations of *D. melanogaster*.

I used a bulk-segregant analysis (BSA) approach using a multiparent population, the Drosophila Synthetic Population Resource (DSPR), to uncover the genetic basis of exploration tendency in *D. melanogaster*. I defined exploration as the tendency of female fruit flies to move from a starting chamber to a novel fly chamber through a narrow

tube. To identify the source of genetic variability in exploration, I generated 17 pairs of "high exploration" and "low exploration" bulk segregant populations consisting of 40 - 100 female flies and performed whole genome pooled sequencing. I then compared allele frequency differences between these pools to identify regions of the genome implicated in exploration tendency.

In my second chapter I studied starvation resistance in an experimentally evolved population of *D. melanogaster*. Our lab placed twelve replicate populations of *D. melanogaster* on three selection treatments: constant high nutritional diet (CHA), fluctuating nutritional diet (FA), and deteriorating nutritional diet (DA). These three treatments have been ongoing for over 50 generations. For my experiment, a duplicated set of flies from all replicates and treatments were placed on one of three diets for 10 days: high sugar, standard, and low yeast diets. After 22 days post-oviposition (p.o.), flies were placed on nutrition-less agar. Starvation resistance was measured as the time it takes for a fly to die starting the moment it is placed on nutrition-less agar. We link these phenotypic changes to variation in artificial selection pressure and environmental conditions.

In my code appendix, I established a novel method of statistical error estimation due to variation in coverage in pooled sequencing experiments. Coverage is defined as the number of reads at a given location along the genome. Due to the law of large numbers, a higher sequencing coverage value leads to a more accurate allele estimation at that locus.

CHAPTER 1

INTRODUCTION

The current climate crisis is a major environmental stressor that threatens the survival of every species on Earth. Warming climes, rising shores, habitat fragmentation, contagious diseases, wild fires, and an increased prevalence of extreme weather events are a few of the many environmental changes that threaten the survival of species (Kolbert 2014). It is necessary for organisms to evolve resistance and behavioral adaptations to environmental stressors to survive the catastrophic effects of climate change. Resisting stress requires some amount of energy on the organism's part that leads to trade-offs with other energy-consuming traits. Organisms will evolve various resource allocation strategies to compensate. Indeed, resource management is essential for stress resistance, as many stressors involve a lack of resources (i.e., droughts and food scarcity). In addition to resisting stress, organisms develop behavioral adaptations to manage environmental stressors. Behaviors such as dispersal and migration have evolved as a means of survival. While researchers have extensively studied stress resistance and behavioral adaptations, the genetic and evolutionary mechanisms by which organisms resist environmental stressors, as well as the phenotypic plasticity of these stress resistant traits, are not fully

known (Ahmad et al. 2018; Everman et al. 2019; Evans et al. 2021; Mackay 2004).

Phenotypic plasticity occurs when a single genotype gives rise to multiple phenotypes, depending on the environment in which the genotype exists in. Abiotic and biotic environmental factors like temperature, altitude, competition, and access to water and shelter can influence an organism's phenotype (Miner et al. 2005; Whitman and Agrawal 2009). This adds another complex layer to our understanding of the genetics of quantitative traits, as phenotypic plasticity can increase organismal fitness and facilitate evolution (Whitman and Agrawal 2009). In addition, standing genetic variation in a population is important for the evolution of beneficial adaptations such as stress resistance (Orr and Betancourt 2001; Barrett and Schluter 2008).

Drosophila melanogaster is a model organism that has been extensively phenotyped and genotyped for a multitude of stress-resistance traits. These include but are not limited to: desiccation resistance, heat tolerance, resource acquisition, starvation resistance, and physiological changes (McCue 2010; Kennington et al. 2001; Karan and Parkash 1998; Telonis-Scott et al. 2016; Harbison et al. 2005; Goenaga, Fanara, and Hasson 2013; Davidson 1990; King, Roff, and Fairbairn 2011; Long, Macdonald, and King 2014; Williams-Simon et al. 2019; Chippindale et al. 1998; Burke et al. 2010). Their widespread, historic use in genetic studies has made them a model organism for the elucidation of

genetic mechanisms of quantitative traits (Roff and Mousseau 1987; Mackay et al. 2012; Sokolowski 2001; Morgan 1910). In particular, the use of *D. melanogaster* in experimental evolution paradigms has played a major role in the elucidation of the mechanistic underpinning of polygenic, quantitative traits (Garland and Rose 2009; Schlötterer et al. 2015; Kofler and Schlötterer 2014).

Experimental evolution is the process by which traits of interest are artificially selected for in highly controlled laboratory populations (Garland and Rose 2009). A researcher may expose an evolving population to specific conditions in order to artificially guide evolution (Kawecki et al. 2012). As an experimental paradigm, it is highly dynamic – the researcher can select for any trait they desire. Relatively recently, experimental evolution has been done in tandem with a next-generation sequencing technique known as pooled-sequencing (Schlötterer et al. 2015). Pooled-sequencing is a method of DNA sequencing by which large numbers of organisms are pooled and sequenced together (Tilk et al. 2019; Zhu et al. 2012). This method is much more cost-effective than individual sequencing, and with the presence of a reference genome, the pooled DNA sequences can be properly aligned for analysis (Lou et al. 2021).

We utilize artificial selection via experimental evolution to select for various resource allocation patterns and stress-related fitness traits and behaviors. For my dissertation research, I analyzed genotypic and phenotypic data from two stress-related traits: exploration behavior and starvation resistance. First, I

performed a bulk segregant analysis (BSA) of two segregating *D. melanogaster* populations: exploring and non-exploring flies (Pool 2016; Elkins, Storks, and King, n.d.). Second, I performed phenotyping of starvation resistance in an experimentally evolved population of *D. melanogaster*. We selected three populations for three separate resource allocation strategies by placing them on three nutritional regimes: a constant high (CH) nutritional diet, a fluctuating (FA) nutritional diet, and a deteriorating (DA) nutritional diet. We phenotyped for starvation resistance (SR) in two groups: direct measurement and diet treatment. In direct measurement, flies were placed directly onto a nutrition-less agar 'diet,' and time until death was recorded. In diet treatment, we explored phenotypic plasticity by placing flies on one of three diets prior to the SR assay. Once placed onto nutrition-less agar, time until death was recorded for diet treatment flies, as well. Finally, I established a novel method of calculating statistical error in allele frequency estimation due to variability in coverage in pooled sequencing experiments.

LITERATURE CITED

Ahmad, Muhammad, Erin S. Keebaugh, Muhammad Tariq, and William W. Ja. 2018. "Evolutionary Responses of Drosophila Melanogaster under Chronic Malnutrition." *Frontiers in Ecology and Evolution* 6 (April). https://doi.org/10.3389/fevo.2018.00047.

Barrett, Rowan D. H., and Dolph Schluter. 2008. "Adaptation from Standing Genetic Variation." *Trends in Ecology & Evolution* 23 (1): 38–44.

Burke, Molly K., Joseph P. Dunham, Parvin Shahrestani, Kevin R. Thornton, Michael R. Rose, and Anthony D. Long. 2010. "Genome-Wide Analysis of a Long-Term Evolution Experiment with Drosophila." *Nature* 467 (7315): 587–90.

Chippindale, Adam K., Allen G. Gibbs, Mani Sheik, Kandice J. Yee, Minou Djawdan, Timothy J. Bradley, and Michael R. Rose. 1998. "RESOURCE ACQUISITION AND THE EVOLUTION OF STRESS RESISTANCE IN DROSOPHILA MELANOGASTER." *Evolution; International Journal of Organic Evolution* 52 (5): 1342–52.

Davidson, J. K. 1990. "Nonparallel Geographic Patterns for Tolerance to Cold and Desiccation in Drosophila-Melanogaster and Drosophila-Simulans." *Australian Journal of Zoology* 38 (2): 155–61.

Elkins, Z. Forrest, Levi Storks, and Elizabeth G. King. n.d.

Evans, Kathryn S., Marijke H. van Wijk, Patrick T. McGrath, Erik C. Andersen, and Mark G. Sterken. 2021. "From QTL to Gene: C. Elegans Facilitates Discoveries of the Genetic Mechanisms Underlying Natural Variation." *Trends in Genetics: TIG* 37 (10): 933–47.

Everman, Elizabeth R., Casey L. McNeil, Jennifer L. Hackett, Clint L. Bain, and Stuart J. Macdonald. 2019. "Dissection of Complex, Fitness-Related Traits in Multiple Drosophila Mapping Populations Offers Insight into the Genetic Control of Stress Resistance." *Genetics* 211 (4): 1449–67.

Garland, T., and M. R. Rose. 2009. "Experimental Evolution." University of California Press. ucpress.edu. 2009. https://www.ucpress.edu/book/9780520261808/experimental-evolution.

Goenaga, Julieta, Juan José Fanara, and Esteban Hasson. 2013. "Latitudinal Variation in Starvation Resistance Is Explained by Lipid Content in Natural Populations of Drosophila Melanogaster." *Evolutionary Biology* 40 (4): 601–12.

Harbison, Susan T., Sherman Chang, Kim P. Kamdar, and Trudy F. C. Mackay. 2005. "Quantitative Genomics of Starvation Stress Resistance in Drosophila." *Genome Biology* 6 (4): R36.

Karan, Dev, and Ravi Parkash. 1998. "Desiccation Tolerance and Starvation Resistance Exhibit Opposite Latitudinal Clines in Indian Geographical

Populations of Drosophila Kikkawai." *Ecological Entomology* 23 (4): 391–96.

Kawecki, Tadeusz J., Richard E. Lenski, Dieter Ebert, Brian Hollis, Isabelle Olivieri, and Michael C. Whitlock. 2012. "Experimental Evolution." *Trends in Ecology & Evolution* 27 (10): 547–60.

Kennington, W. J., A. S. Gilchrist, D. B. Goldstein, and L. Partridge. 2001. "The Genetic Bases of Divergence in Desiccation and Starvation Resistance among Tropical and Temperate Populations of Drosophila Melanogaster." *Heredity* 87 (Pt 3): 363–72.

King, E. G., D. A. Roff, and D. J. Fairbairn. 2011. "Trade-off Acquisition and Allocation in Gryllus Firmus: A Test of the Y Model." *Journal of Evolutionary Biology* 24 (2): 256–64.

Kofler, Robert, and Christian Schlötterer. 2014. "A Guide for the Design of Evolve and Resequencing Studies." *Molecular Biology and Evolution* 31 (2): 474–83.

Kolbert, Elizabeth. 2014. *The Sixth Extinction: An Unnatural History.* A&C Black.

Long, Anthony D., Stuart J. Macdonald, and Elizabeth G. King. 2014. "Dissecting Complex Traits Using the Drosophila Synthetic Population Resource." *Trends in Genetics: TIG* 30 (11): 488–95.

Lou, Runyang Nicolas, Arne Jacobs, Aryn P. Wilder, and Nina Overgaard Therkildsen. 2021. "A Beginner's Guide to Low-Coverage Whole Genome Sequencing for Population Genomics." *Molecular Ecology* 30 (23): 5966–93.

Mackay, Trudy F. C. 2004. "The Genetic Architecture of Quantitative Traits: Lessons from Drosophila." *Current Opinion in Genetics & Development* 14 (3): 253–57.

Mackay, Trudy F. C., Stephen Richards, Eric A. Stone, Antonio Barbadilla, Julien F. Ayroles, Dianhui Zhu, Sònia Casillas, et al. 2012. "The *Drosophila Melanogaster* Genetic Reference Panel." *Nature* 482 (7384): 173–78.

McCue, Marshall D. 2010. "Starvation Physiology: Reviewing the Different Strategies Animals Use to Survive a Common Challenge." *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* 156 (1): 1–18.

Miner, Benjamin G., Sonia E. Sultan, Steven G. Morgan, Dianna K. Padilla, and Rick A. Relyea. 2005. "Ecological Consequences of Phenotypic Plasticity." *Trends in Ecology & Evolution* 20 (12): 685–92.

Morgan, T. H. 1910. "SEX LIMITED INHERITANCE IN DROSOPHILA." *Science* 32 (812): 120–22.

Orr, H. A., and A. J. Betancourt. 2001. "Haldane's Sieve and Adaptation from the Standing Genetic Variation." *Genetics* 157 (2): 875–84.

Pool, John E. 2016. "Genetic Mapping by Bulk Segregant Analysis in Drosophila: Experimental Design and Simulation-Based Inference." *Genetics* 204 (3): 1295–1306.

Roff, D. A., and T. A. Mousseau. 1987. "Quantitative Genetics and Fitness: Lessons from Drosophila." *Heredity* 58 (Pt 1) (February): 103–18.

Schlötterer, C., R. Kofler, E. Versace, R. Tobler, and S. U. Franssen. 2015. "Combining Experimental Evolution with Next-Generation Sequencing: A Powerful Tool to Study Adaptation from Standing Genetic Variation." *Heredity* 114 (5): 431–40.

Sokolowski, M. B. 2001. "Drosophila: Genetics Meets Behaviour." *Nature Reviews. Genetics* 2 (11): 879–90.

Telonis-Scott, Marina, Carla M. Sgrò, Ary A. Hoffmann, and Philippa C. Griffin. 2016. "Cross-Study Comparison Reveals Common Genomic, Network, and Functional Signatures of Desiccation Resistance in Drosophila Melanogaster." *Molecular Biology and Evolution* 33 (4): 1053–67.

Tilk, Susanne, Alan Bergland, Aaron Goodman, Paul Schmidt, Dmitri Petrov, and Sharon Greenblum. 2019. "Accurate Allele Frequencies from Ultra-Low Coverage Pool-Seq Samples in Evolve-and-Resequence Experiments." *G3* 9 (12): 4159–68.

Whitman, Douglas, and Anurag Agrawal. 2009. "What Is Phenotypic Plasticity and Why Is It Important?" In *Phenotypic Plasticity of Insects*. Science Publishers.

Williams-Simon, Patricka A., Christopher Posey, Samuel Mitchell, Enoch Ng'oma, James A. Mrkvicka, Troy Zars, and Elizabeth G. King. 2019. "Multiple Genetic Loci Affect Place Learning and Memory Performance in Drosophila Melanogaster." *Genes, Brain, and Behavior* 18 (7): e12581.

Zhu, Yuan, Alan O. Bergland, Josefa González, and Dmitri A. Petrov. 2012. "Empirical Validation of Pooled Whole Genome Population Re-Sequencing in Drosophila Melanogaster." Edited by Jeong-Sun Seo. *PloS One* 7 (7): e41901.

CHAPTER 2

THE GENETIC BASIS OF EXPLORATION BEHAVIOR IN A MULTIPARENT POPULATION OF *D. MELANOGASTER*

Z. Forrest Elkins, Levi Storks, Elizabeth G. King

ABSTRACT

The ability of animals to move throughout their environment to find food, mates, and suitable habitat is critical to their survival and reproduction. However, this behavior can be energetically expensive and potentially costly. As a result, individuals often vary widely in their overall motility, exploration, and dispersal tendency. We used a bulk-segregant analysis (BSA) approach using a multiparent population, the Drosophila Synthetic Population Resource (DSPR), to uncover the genetic basis of exploration tendency in *Drosophila melanogaster*.

Our measurement of exploration tendency was the tendency of female fruit flies to move from a starting chamber to a novel fly chamber. We first demonstrated our measure of exploration tendency has a genetic basis by assaying 20 recombinant inbred lines (RILs) to estimate the broad-sense heritability of exploration tendency ($H2$=0.4). To identify the source of this genetic variability, we generated 17 pairs of "high exploration" and "low exploration" pools consisting of 40 - 100 female flies and performed whole genome sequencing. We compared allele frequency and haplotype frequency differences

between these pools to identify regions of the genome implicated in exploration tendency.

INTRODUCTION

The rapid rise in global temperatures, increasing habitat fragmentation, deforestation, an increase in the number of annual extreme weather events, contagious diseases, and more have made the Earth more perilous than ever for living organisms (Kolbert 2014). During the current sixth mass extinction event, the organisms and species that survive will be the ones that are able to adapt. Beneficial behavioral adaptations such as predator avoidance, foraging for food, and dispersal can confer higher rates of reproductive success to organisms that exhibit the beneficial trait (Alcock and Rubenstein 2009). In the face of climate change and habitat destruction, organisms must develop behavioral adaptations to survive. One such behavioral adaptation, exploration tendency, is an important trait for an organism to have in their survival toolkit. By exploring novel environments, organisms can find habitable environs to escape to when their previous habitat is rendered inhospitable or destroyed.

Ecologists and behavioral researchers have observed and documented exploration behavior and its variability both within and between populations (Dingemanse *et al.* 2002; Arvidsson *et al.* 2017; Moran *et al.* 2017; Bengston *et al.*

2018; Lemanski *et al.* 2019; Arika *et al.* 2019; Mouchet *et al.* 2021). In particular, the variance in dispersal distance within metapopulations and its contribution to gene flow have been extensively studied in various species, including Glanville fritillaries and the Western European great tit (Fountain *et al.* 2018; Mouchet *et al.* 2021). In addition to metapopulation dynamics, there is a broad body of research on exploration behavior as a personality trait commonly linked with foraging behavior (Verbeek *et al.* 1994; Sokolowski *et al.* 1997; Fitzpatrick *et al.* 2005; Arvidsson *et al.* 2017; Lemanski *et al.* 2019). As a sub-phenotype of dispersal, exploration tendency is commonly characterized as an individual's movement through a novel space (Verbeek *et al.* 1994; Dingemanse *et al.* 2002; Réale *et al.* 2007).

While we know a lot about exploration behavior and its effect on population structure, one key area is the genetic mechanism that controls variation in exploration tendency. There have been extensive studies on genetic variance between populations and its relation to behavioral phenotypes of exploration and dispersal, which indicate that exploration is a heritable trait with genetically-based differences between populations (Rudin *et al.* 2018). Researchers have also identified that exploration behavior leads to upregulation of molecules implicated in neuroplasticity (Ramírez-Amaya *et al.* 2005; Paul *et al.* 2020). However, the exact genetic mechanisms that underlie exploration behavior are still relatively unknown.

One reason for our lack of knowledge of the genetic mechanism is that it is difficult to assess specific allele frequencies in natural populations. With the advent of next-generation sequencing, researchers can more cheaply and effectively peel back the curtain on the relationship between phenotype and genotype, particularly with polygenic traits like exploration tendency. Mass groups of organisms can be assayed and sequenced efficiently enough to provide sufficient power in implicating causative loci. Experimental evolution techniques such as pool-seq and bulk segregant analysis (BSA), as well as various other mapping paradigms have arisen over the past three decades (Baldwin-Brown *et al.* 4/2014; Michelmore *et al.* 1991; Nuzhdin and Turner 2013; Kofler and Schlötterer 2014; Long *et al.* 2015; Stanley *et al.* 2017). Bulk segregant approaches came about as a method to optimize our ability to detect genotypic differences underlying phenotypes while minimize sequencing costs (Michelmore *et al.* 1991; Magwene *et al.* 2011; Pool 2016). We use BSA, in which two samples (or bulks) from a single population have segregated into two opposing phenotypic groups whose genotypes can then be compared, to quickly observe genetic differences between "high" and "low" explorers. One downside of the BSA. Technique is that it is impossible to calculate linkage disequilibrium as the individual organism IDs are removed from the genetic data (Lynch *et al.* 2014). However, we can infer haplotype data from the DSPR's fully-sequenced parent lines to identify linkage disequilibrium in our experimental population (King *et*

al. 2012, 07/2012). In addition, the usage of inferred haplotype data increases the accuracy of our allele frequency differences between the two pooled populations (Tilk *et al.* 2019).

To study the genetic components of exploration behavior, we looked at haplotype structure and single nucleotide polymorphisms (SNPs). We use a bulk-segregant analysis (BSA) approach to identify the underlying genetic components of exploration tendency in a synthetic population of *D. melanogaster*. Bulk-segregant analysis is a quick and affordable method for looking at genetic differentiation between populations that display extreme forms of a given phenotype (Magwene *et al.* 2011). The method works by sorting organisms by their extreme phenotype. In our case, flies are sorted as explorers (E) and non-explorers (NE). Since the phenotype is binary, we can reasonably expect genetic differentiation to be present between the two populations if the phenotype in question is heritable. We use this method to answer the following questions: 1) is exploration tendency a heritable trait, 2) if it is, what are the QTLs (quantitative trait loci) that contribute to variation in exploration tendency, and 3) what is the overall genetic architecture involved in exploration tendency?

METHODS

Experimental Population

The DSPR is a collection of two populations of *D. melanogaster* recombinant inbred lines (RILs) formed by intercrossing two sets of eight inbred founder lines (Elizabeth G. King, Macdonald, and Long 07/2012; E. G. King et al. 2012; Long, Macdonald, and King 2014). Here, we used lines from the pB population. The population was crossed for 50 generations to create offspring whose genetic material is a mosaic of the founders followed by 25 generations of inbreeding to create the RILs. The founders consist of lines from all over the world to gain a wide variety of genetic variation that exists in these populations. Therefore, any given RIL contains a mosaic of the founders' haplotypes. The eight founder haplotypes have been completely sequenced, and a hidden Markov model informs us which of the founder haplotypes exist at any given position along the genome in our experimental populations (King et al. 2012, Long et al. 2014).

Exploration chamber design

The exploration cages consist of two chambers (A and B) connected by a long, narrow tube. The two chambers, made of nylon and measuring 9.0cm x 9.5cm, are connected by a tube that is 14.5cm long with a diameter of 0.5cm. The tube is in the center of the inward sides of the two boxes. No food or smell was placed

inside either chamber. Chambers were placed randomly inside the environmental chamber. The cages were modeled after those used by Simon et al. (2011).

Heritability

Fly rearing

We replicated out twenty recombinant inbred lines (RILs) from our DSPR fly population. After two generations of flies that were fed a standard yeast diet, offspring 14-16 days post-oviposition (po) were used for the heritability assay.

Experimental design

To test for the heritability of exploration tendency, a group of 30 male and 30 female flies aged 14-16 days post-oviposition (po) were isolated from a single RIL, anesthetized with carbon dioxide, and placed on the floor of the A chamber of the apparatus 24 hours after anesthetization. The apparatus was then placed on a rack in the brightly lit environmental chamber at 23°C and 50% humidity. Flies were initially acclimated for 1 hour during which they were confined to the A chamber. After this acclimation period, A and B chambers were connected and the assay proceeded for 6 hours during which the flies were able to explore the B chamber. The assays took place between 08:00 and 19:00. Upon completion of the assay, the number of male and female flies in each chamber was counted and an exploration index was calculated as the percentage of total flies counted in the B

chamber after the assay. We performed the assay for 20 RILs with each replicated 4-5 times. We then calculated the overall proportion of flies found in the A chamber and B chamber per RIL. By graphing the proportion of flies found in the B chamber compared to the A chamber in order of least to greatest, we can visually see that some RILs are more exploratory than others. We also ran a simple linear regression model to test whether or not the time at which the assay began influences exploration tendency (`lm(proportion explorers ~ time)`).

We estimated broad-sense heritability by calculating the genetic and phenotypic variation in exploration in the RILs from a linear mixed model in R (R Core Team 2022). We fit phenotype by genotype using an analysis of variance (`aov`) model in R (`aov(phenotype~genotype)`). We used the estimated genetic variance over the total variance in RIL means to get our estimated broad-sense heritability of RIL means.

Bulk segregant analysis (BSA)

Fly rearing

In our BSA, we used a series of inbred lines (supplemental materials) (King et al. 2012). We then outbred the population by mixing the inbred lines for our BSA experiment (Figure 2). Flies from the outbred population were placed on a

control yeast diet similar to the control diet commonly used in studies (Bass *et al.* 2007; Stanley *et al.* 2017). We used the SAFPro Relax + YF 73050 brand of yeast. It contains 45-60 g of protein and 30-38 g of carbohydrate per 100 g inactivated yeast (Lesaffre Yeast Corp., Milwaukee, USA). The control diet was stored at 4 ^{o}C and used within 2 weeks of the preparation date. When the female offspring were 14-16 days po, we placed them in the exploration assay.

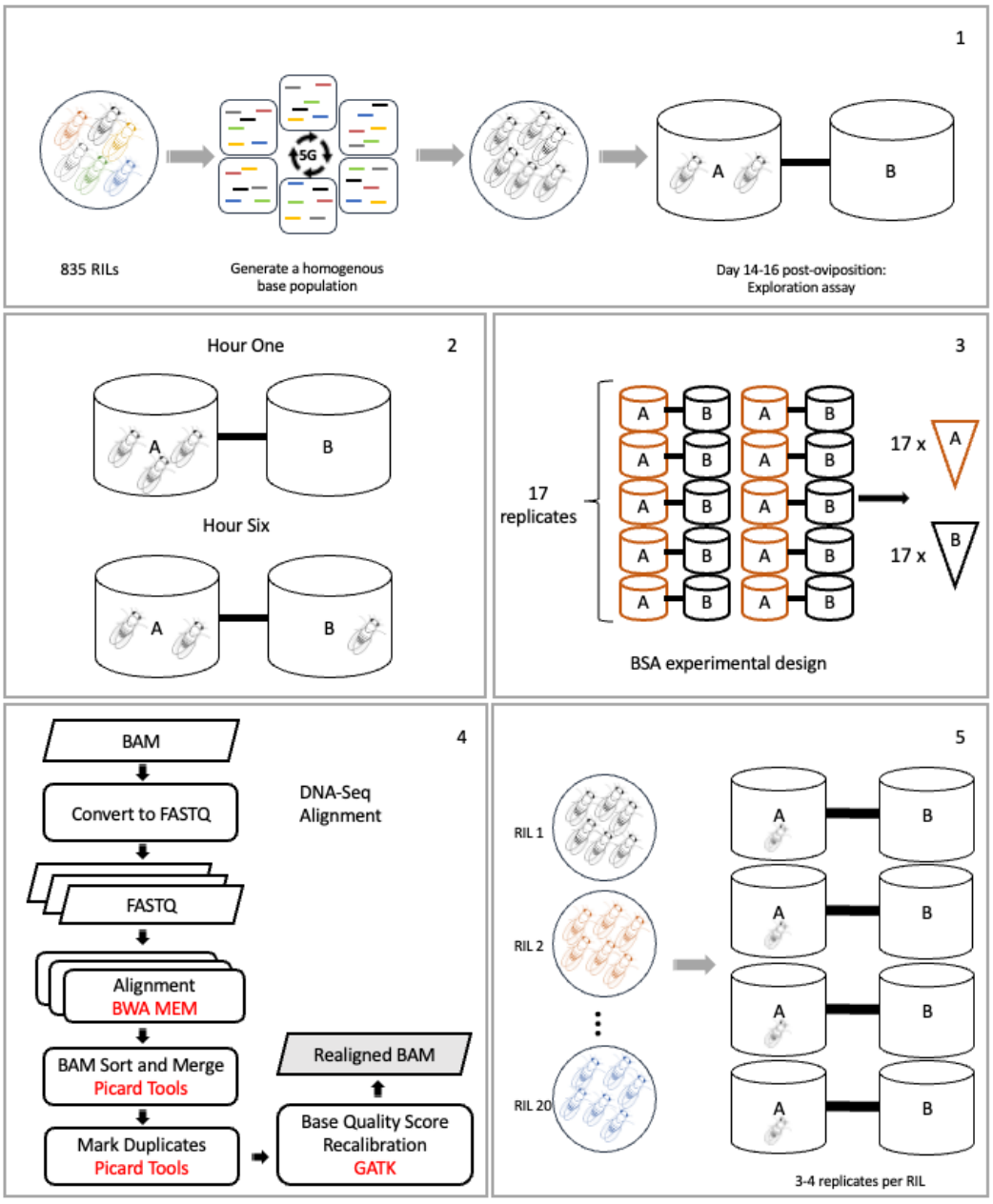


Figure 1. The DSPR population started from eight inbred founders representing global standing variation in Drosophila melanogaster. They were mixed en masse for 50 generations and about 800 RILs from the DSPR were inbred from this population. The RILs were then re-mixed to create a base population for experimental evolution (modified from Ng'oma et al. 2021). 1) Details the creation of our outbred population that populates our exploration chamber assay. 2) Flies were placed into the exploration chamber and left there for a total of six hours. 3) Each

experimental replicate contained ten exploration chambers with flies gathered equally from the A and B chambers -- A denoting non-explorers and B denoting explorers. We ran a total of seventeen replicates. After all seventeen replicates were completed, we had a total of 34 pooled samples, 17 per both sides of the chamber. 4) The Burrows-Wheeler Alignment pipeline we used to obtain our genetic data. 5) To calculate the heritability of our exploration assay, we took 20 RILs and placed 30 male and 30 female flies into the A chamber. At the end of 6 hours, we counted the number of flies that we found in both chambers. We performed four replicate assays per RIL, for a total of 20 RILs x 4 replicates = 80 experimental chambers.

We used samples taken from an outbred population of the DSPR to select for exploration in a single generation. We sampled from an outbred population made up of the B synthetic population of the DSPR. We utilized an outbred population to increase the genetic variation between flies in our experiment, thereby allowing us to observe genetic differences between flies showing extreme phenotypic values in either direction within a single generation.

BSA Assay

Sets of 60 female flies were used to seed each cage. Individuals were placed into chamber A of the exploration cages. The assay lasted for six hours. Flies found within chamber B after the duration of the experiment are labeled as "explorers," while flies collected from chamber A are labeled as "non-explorers" (Figure 3). Both explorers and non-explorers are then pooled within their

respective groups in equal numbers to create matched pools. Because the number of non-explorers always exceeded explorers, we used a random set of the non-explorers to match the number of explorers for each round. There was a total of 10 cages per trial. Seventeen trials were conducted. The result was a total of 34 samples, with two samples per trial (Figure 2).

We conducted bulk-segregant analyses (BSA) by designating the extreme forms of the phenotype – E and NE – as populations. We then used a next-generation pooled sequencing approach to compare any genetic differentiation between explorers and non-explorers. We looked for differentiation at both the single nucleotide polymorphism (SNP) and haplotype level.

For each trial, the explorer flies from each cage were combined and their DNA extracted into a single vial. This process was repeated for the non-explorers. There was a total of 17 trials, resulting in 34 total samples. Each sample contained 30-100 flies. The number of flies in the explorer and non-explorer vials for a single trial are equal. If more flies were collected from chamber A, the number placed into the vial matched the number of flies collected from chamber B.

Sequencing

We extracted the DNA using the Qiagen Puregene extraction method. The DNA yield was between 80-120 ng/uL, with 50-85 uL per sample. DNA was

suspended in a DNA hydration solution. The genetic library was prepared by the University of Missouri DNA core. The 34-plexed pool of DNA was sequenced with an Illumina NextSeq 500, using a paired-end read (150 base pairs) for roughly 20x coverage. We aligned the sequenced DNA reads with a Burrows-Wheeler Alignment (BWA) pipeline that utilized the *Drosophila* reference genome Release 6 to call our SNPs.

SNP Calling

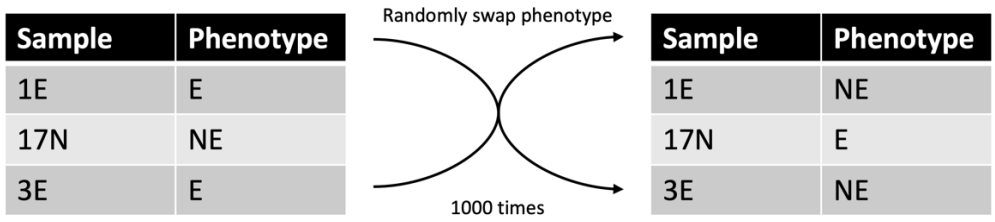
After mapping our sequence reads to the reference genome using BWA 0.7.15, we utilized Picard 2.7.1 tools to assign reads, find any duplicate reads and mark them, and create a sequence dictionary. We then used Samtools 1.3.1 to index our reads, and then used the HaplotypeCaller from GATK 3.8 with a –ploidy tag of 8 and the –variant tag for all 34 .vcf files to place our reads with raw SNP calls into a single output file. We then used the GATK 3.8 VariantFiltration method to filter out SNPs with LowQual values under 30.0, LowVQCBD values under 5.0, and FisherStrand values under 60.0. Finally, we used our list of founder SNPs to filter out any SNPs in the experimental population not found within the founders.

G' analysis

Since we used a BSA-seq approach, our genetic data consisted of varying levels of sequencing coverage. To counteract this, we used a statistic developed by Magwene et. al, 2011. This statistic, the G' statistic, takes coverage into account when calculating allelic differences between bulks. We prepared our genetic data with position along the genome, chromosome, and the counts of both alleles. *G'* values were calculated using the code established by Magwene et. al 2011 (https://bitbucket.org/pmagwene/bsaseq/src/master/).

Significance testing

We calculated significant *G'* values using the Monte Carlo method. To simulate a null distribution of *G'* values, we randomly assigned our observed data to either the E or NE bulks. This process was iterated 1000 times. This gave us simulated, randomly calculated *G'* values across the genome (https://bitbucket.org/pmagwene/bsaseq/src/master/).



a)

```
python3 ../bsacalc.py
-L ../Gtest10D.txt ../Gtest7D.txt ../Gtest3N.txt
-H ../Gtest5N.txt ../Gtest3D.txt ../Gtest1D.txt
-o ../sim_N.out
```

b)

Figure 2. Diagram of our MCMC analysis. A) First, observed data is randomly shuffled into two groups: explorers (E) and non-explorers (NE). This process is repeated 1000 times for our 34 total samples (17 E and 17 NE). B) Then, we use the randomly shuffled data to calculate *G'* values. In the end, we run 1000 commands, one per simulation.

Once we obtained our simulated data, we calculated the false discovery rate (FDR) and family-wise error rate (FWER) for the data. To calculate the FWER, we found the maximum *G'* values from each simulation, for a total of 1000 maximum *G'* values. Then, we used the quantile function in base R (https://www.r-project.org) to find the *G'* value demarcating the top 5% of these values (R Core Team 2022). This number became our FWER threshold.

To calculate FDR, we created several thresholds within the simulated data. These thresholds started at 20 and increased by 10, ending with 50 as our highest threshold. We then counted the number of "peaks," or *G'* values, above each of these thresholds. We then took the number of simulated values above the current chosen threshold and divided it by the number of observed values above that

threshold to get our FDR. Proceeding in this way, we narrowed down our threshold until we arrived at an FDR of 0.05. The *G'* value at which our FDR is 0.05 became the threshold of significance.

Regions of interest

After finding the 13 loci above the G' FDR threshold, we grabbed the position, SNP name, chromosome arm, and *G'* values of those loci. Then, we used the 13 genetic loci to find genes close to those locations on flybase (https://flybase.org). We observed, on average, the genes found within 20 kilo base pairs (kbp) of the implicated loci. For some, we extended our search out to 100 kbp. We also used our linkage disequilibrium analysis to see if any of our implicated loci were in regions of high LD with other locations along the genome. For the sake of clarity, we also plotted the raw mean allele frequencies around the 13 loci.

Linkage disequilibrium

Linkage disequilibrium (LD) was calculated in the parent population using PLINK ver.1.90b. First, we used the sequenced DNA of our founder genomes to create required pedigree (.ped) and mapping (.map) files. The .map file contains the chromosome name, SNP, and its location in both centimorgans (cM) and base pairs. The .ped file contains the pedigree information of the sequenced flies. This

includes individual ID, mother and father IDs (which were set to 0 as parents are unknown), sex of the flies (2 for female), and phenotype (set to 0 for unknown).

We then passed our .map and .ped files through PLINK with the `'--make-bed'` tag to create a binary file. This file makes the calculation of R^2 linkage disequilibrium values quick and easy. We then made pairwise comparisons among the SNPs underlying our 13 *G'* values of interest. We used the thirteen G' SNP IDs to index the parent genome's pedigree and mapping files for our LD calculation. We made a total of 13x13=169 LD comparisons to look for linkage between our significant hits.

RESULTS

Heritability

We calculated a broad-sense heritability of $H^2 = 0.4$, 95% CI [0.29,0.57]. In addition, we can clearly see that some RILs exhibit higher exploration behavior than others (Figure 3). We saw an extreme difference between the lowest and highest explorers, too, with low exploring lines almost never visiting the B chamber while the highest exploring lines would have half the flies explore the B chamber. Our linear regression model, `lm(proportion explorers ~ time)`, ran on our RIL exploration data, indicates that time does not predict

exploration tendency (R^2 = 7.51e-08, F(1,96) = 7.209e-06, p = 0.9979) (Figure 4).

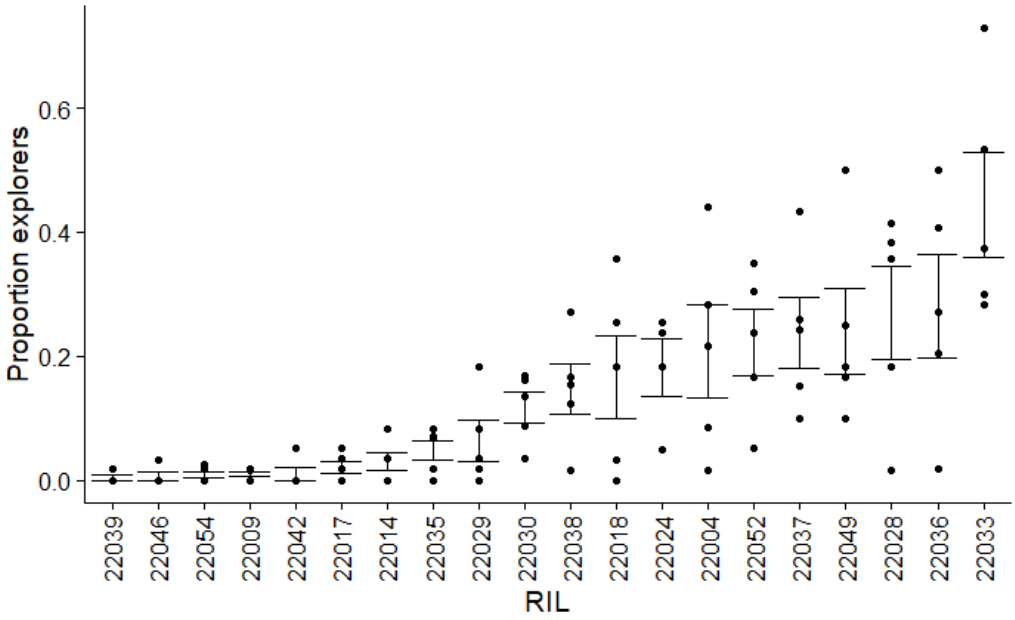


Figure 3. The proportion of flies found in the B chamber of our exploration assay, sorted from lowest to highest. Flies' exploration behavior can be predicted by which RIL the flies come from. This confirms that exploration behavior has a genetic component.

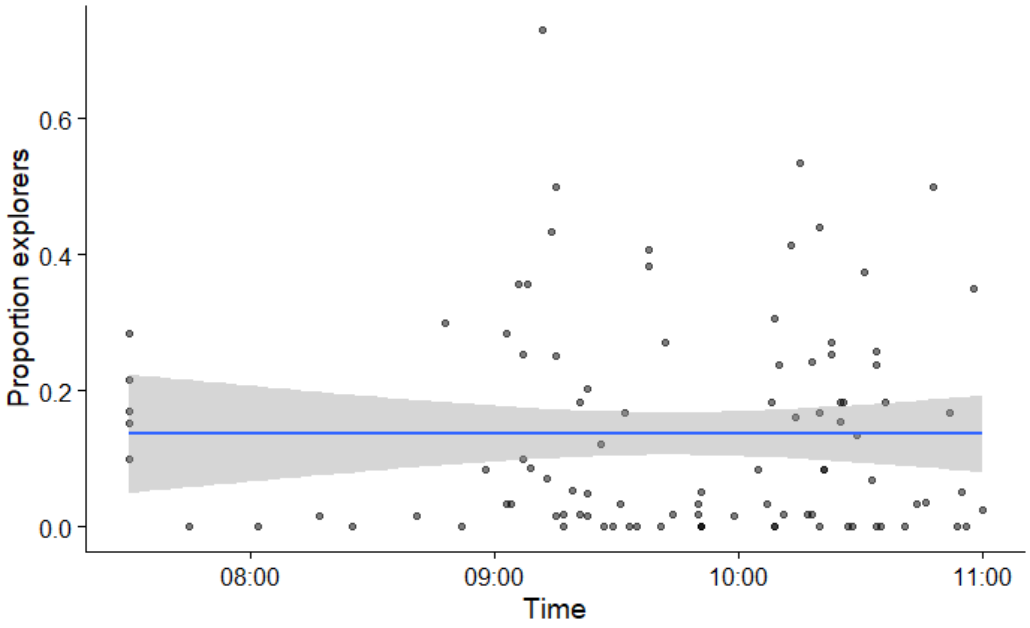


Figure 4. The proportion of flies found in the B chamber of our exploration assay predicted by the time at which the assay was ran. Exploration tendency is not significantly predicted by time (R^2 = 7.51e-08, F(1,96) = 7.209e-06, p = 0.9979).

G' analysis

We found no significant peaks utilizing an FWER of 0.05. However, we did find thirteen significant G' values after calculating the false discovery rate.

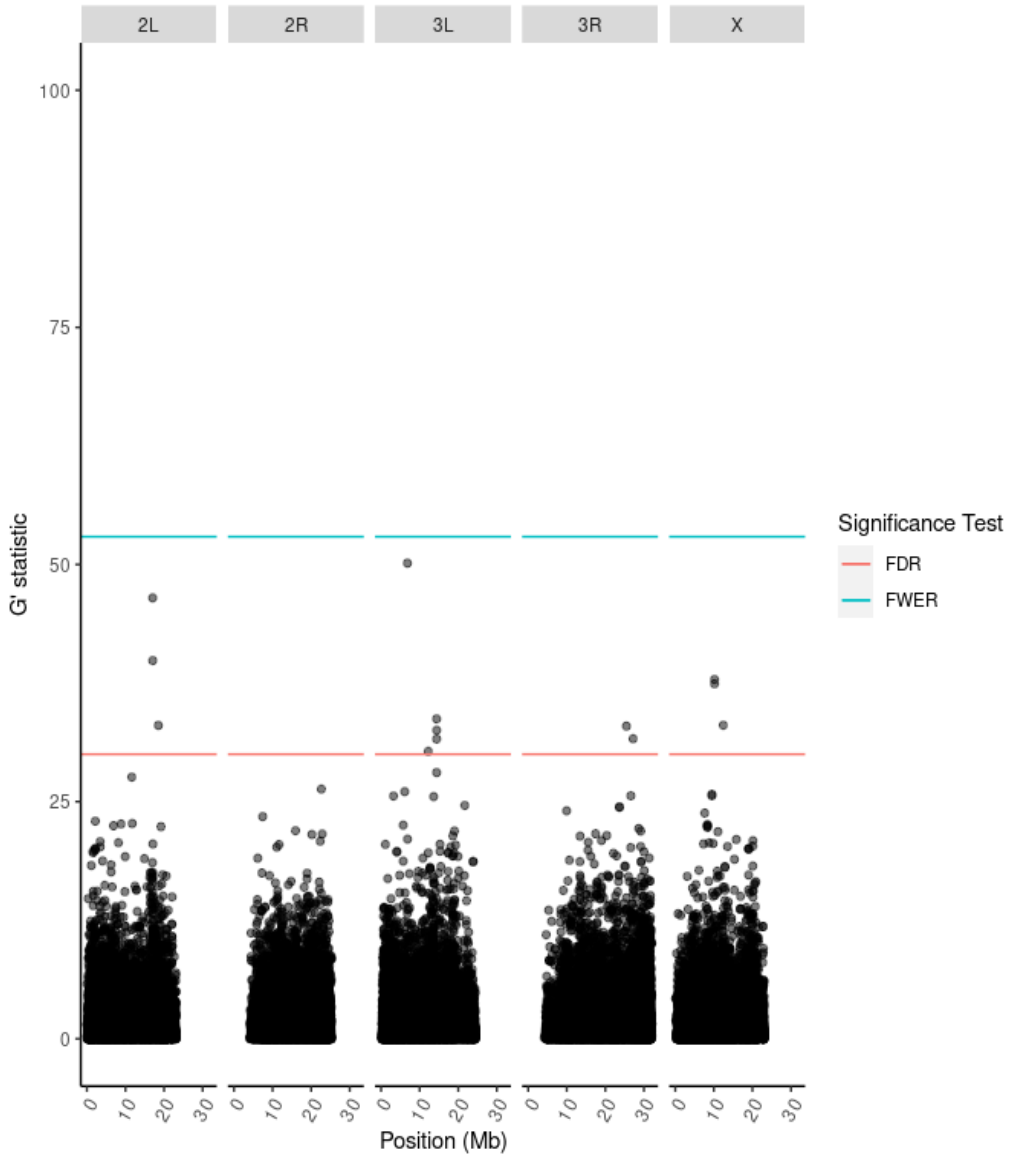


Figure 5. G' values plotted across the entire genome. The blue line denotes family-wise error rate of 0.05 for our data. There were no observed data points above the FWER threshold. The red line denotes a false discovery rate of 0.306. We found 13 loci above our false discovery rate.

Regions of interest

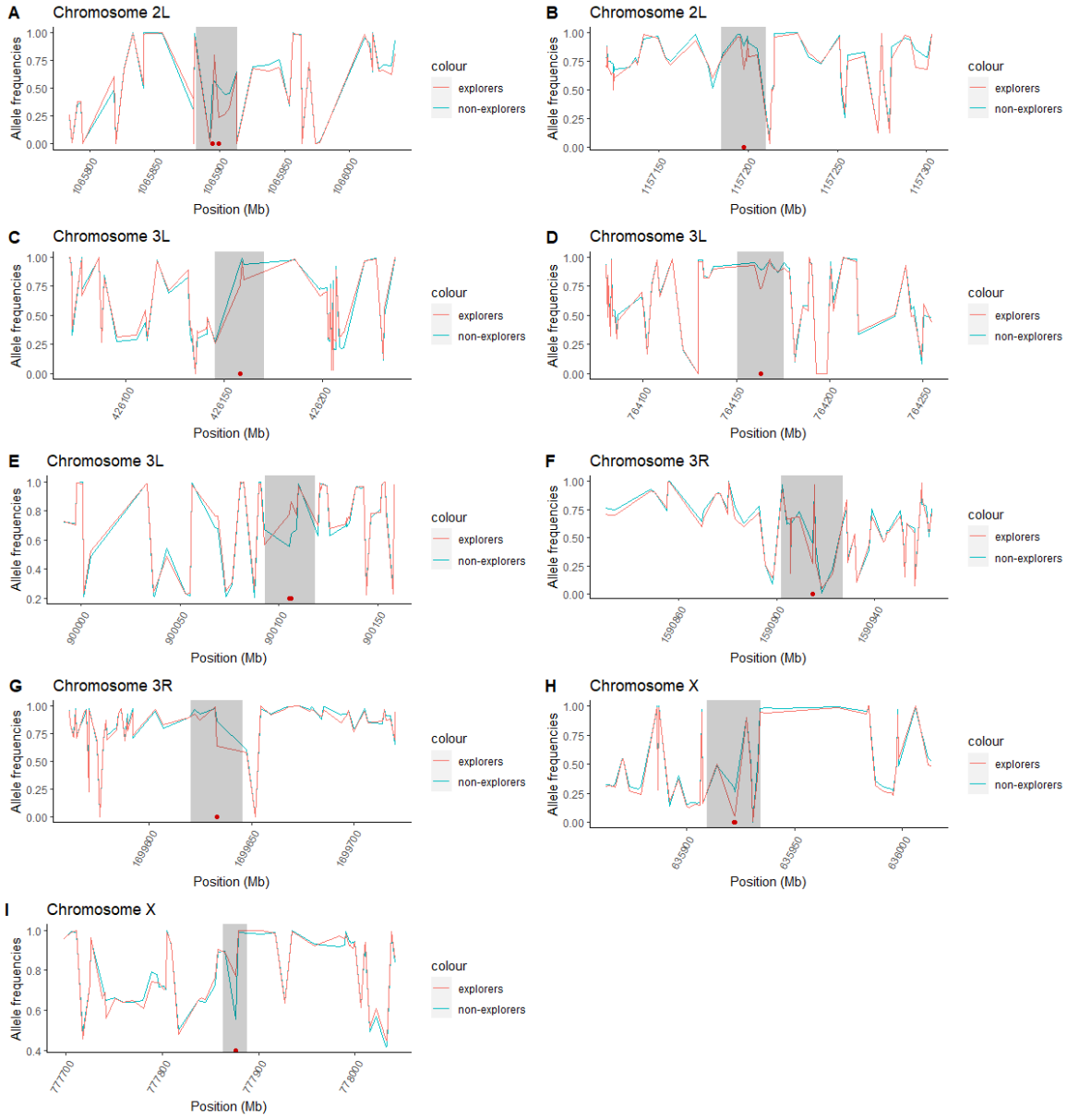


Figure 6. Each plot shows mean allele frequencies of explorers in red and non-explorers in blue across the genome. The red dots centered in the grey rectangles indicate the positions of the 13 significant G' QTLs. Plots A, E and H contain 2, 3, and 2 significant QTLs, respectively.

Chromosome	Position	Gene	Associated phenotype
2L	17054316, 17054384	*CG15136*	Abnormal flight
2L	18515161	*Ugt201D1*	Enables UDP-glycosyltransferase activity
		CG10211	Involved in response to oxidative stress
3L	6818526	*vvl*	Specification of cell fates, patterning and immune defense
3L	12226612	*app*	Regulation of fat signaling abnormal locomotive behavior
3L	14401683, 14401693, 14401703	*Dscam2*	Abnormal neuroanatomy, size, body color
3R	25454633	*Men*	Abnormal heat stress response, abnormal sleep
3R	27194130	*G14369, CG14370*	Little to no information
X	10174752, 10174756	*CG32767*	Involved in several transcription factor activities, expressed in wing hinge primordium and wing pouch
		CG15465	Little to no information
X	12446014	*Btnd*	Flightless, abnormal heat stress response
		Efr	Manifests in wing vein
		sqh	Involved in cytokinesis and tissue morphogenesis
		dtn	Abnormal heat stress response

Table 1. List of regions of interest by chromosome arm, position (Mb), gene name and associated phenotype. Information was collated from FlyBase (https://flybase.org).

We found several regions of interest that appeared above our false discovery rate significance threshold. On chromosome arm 2L, we found two loci of interest involved in abnormal flight. On chromosome arm 3L, we found one locus of interest suspected to be involved in abnormal locomotive behavior. On the X chromosome arm, we found several loci involved in heat response and variation in wing vein phenotypes. The genes in question on the X chromosome were *Btnd*, *Efr*, *dtn* (detonator) and *sgh* (spaghetti squash).

Linkage disequilibrium

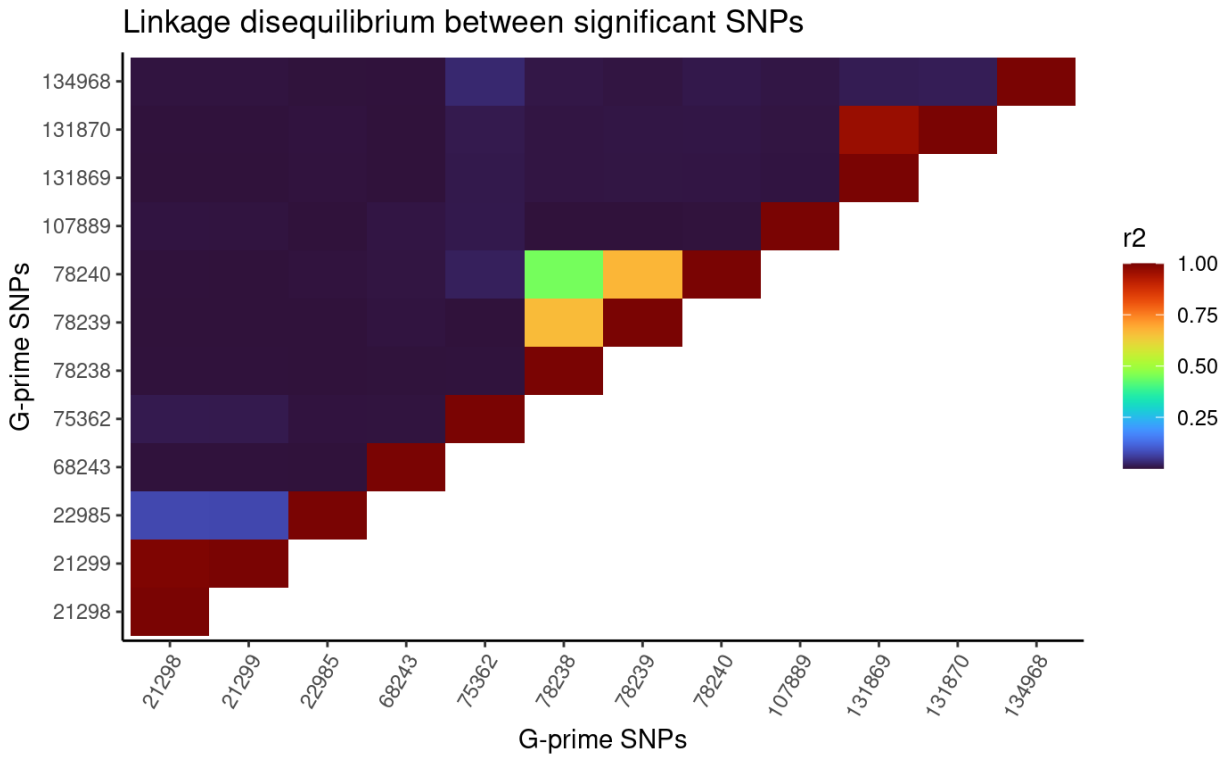


Figure 7. Pairwise linkage disequilibrium calculated between the 13 significant SNPs obtained from the G' analysis. Plotted heatmap values are R^2 values calculated with the parent

population data, where numbers closer to 1 indicate SNPs in high LD and numbers closer to 0 indicate SNPs in low LD.

Our analysis of linkage disequilibrium indicates that significant QTLs further from each other are not in LD. We can see that QTLs close to each other, like SNPs 78240 and 78238, are in LD with each other. This makes sense, as SNPs that are physically close to each other are more likely to be in high LD than SNPs far apart from each other.

DISCUSSION

We found 15 genes of interest underlying our 13 significant G' hits. These genes are located on chromosome arms 2L, 3L, 3R and X. The full list of implicated genes is listed in Table 1. Among the 15 genes, 4 are involved to some extent with flight or locomotive phenotypes. In addition, 5-6 of the genes are involved in the Scer\GAL4 pathway. This pathway is involved in abnormal stress and temperature responses in *D. melanogaster.* There are two additional genes involved in abnormal temperature stress response as well. One implicated QTL encodes for the cell-surface protein Down syndrome cell adhesion molecule 2 (Dscam2). The original Dscam protein was discovered in a region of human chromosome 21 critical for Down's syndrome (Yamakawa *et al.* 1998). Dscam2 is expressed in the *D. melanogaster* nervous system, and has been primarily studied

due to its involvement in the *Drosophila* visual system and the role it plays in the formation of neuronal synapses (Millard *et al.* 2007; Lah *et al.* 2014; Bosch *et al.* 2015; Odierna *et al.* 2020). Perhaps our non-exploring flies were unable to visually identify the passage opening due to changes in their visual system caused by Dscam2.

Of the 13 significant loci implicated by G' in exploration behavior, we found 3 of them to be in high LD. We compared implicated regions along the genome with known LD within those regions. Implicated regions with low LD are more likely to house true QTLs that contribute to exploration. Conversely, implicated regions with high LD may house a QTL, or they might be linked to the true alleles that underlie exploration tendency. The 3 loci in high LD were all physically clustered together. This matches our expectations, as genes that are physically close are physically linked, and any selection on one of the loci will invariably affect the linked loci. Overall, we found that exploration behavior is a heritable, polygenic trait with several loci of interest. The QTLs we identified have been previously implicated in other behavioral traits, such as locomotive behavior and flight.

We used a multiparent population of *D. melanogaster* (the DSPR) that gave us higher power and insight into our experimental results. Since the experimental population is descended from the eight founding lines, our

experimental population exhibits one of eight possible haplotypes at any given position along the genome (Burke *et al.* 2010; Long *et al.* 2014; King *et al.* 07/2012). This increases the power we have in determining QTLs of interest that actually underlie exploration tendency (Kessner and Novembre 2015). In addition, we used linkage disequilibrium across the genome calculated from the parent population in comparison to the significant *G'* analysis results. The DSPR has been used by multiple research groups to identify the genetic basis of traits like longevity, resistance to toxins, learning and place memory, transposable element abundance, relative fitness, nicotine resistance, and fungal resistance (Cridland *et al.* 2013; Marriage *et al.* 2014; Najarro *et al.* 2015; Highfill *et al.* 2016; Everman *et al.* 2019; Williams-Simon *et al.* 2019; Singh *et al.* 2022; Riddle and Farnoudi).

One of the key benefits of using the DSPR in this set of experiments is our ability to utilize linkage disequilibrium (LD) to inform our results. By adding LD to our analysis, we can determine which significant signals are in high LD with other regions of the genome. This tells us that our observed significant signal could either a) underlie our trait of interest, or b) merely be in high LD with the true region underlying our trait. However, this combined analysis does not tell us which signal is the true QTL. It does give us more information than we had previously, adding to the overall power of our experimental design to detect

significant QTLs. The ability to reduce the level of uncertainty in our results due to LD is a unique aspect of this technique.

We calculated LD between the 13 loci of interest that we discovered in our G' analysis. We expected high LD to occur around the loci that are close to each other, as loci that are physically close to each other are physically linked as well. What this indicates is that if one of the loci in high LD is selected for, the loci that it is in LD with will change in frequency as well. We found three of our thirteen loci of interest were in high LD (Figure 7).

We also used a bulk segregant analysis approach, allowing us to view genetic differences within a single generation due to the nature of our binary phenotype. Since our exploration assay result is binary, we can exploit this to rapidly view genetic differences via BSA-seq (Michelmore *et al.* 1991).

We found that exploration tendency is a complex trait with several loci across the genome contributing to it in our synthetic population. This coincides with previous QTL mapping studies conducted with the DSPR, where fungal resistance, sleep, cold tolerance, nicotine resistance, and more are polygenic traits controlled by several alleles of small effect (Crawford 2013; Marriage *et al.* 2014; Smith and Macdonald 2020; Riddle and Farnoudi). While to our knowledge no other QTL mapping studies have been conducted on exploration behavior as defined in this study, related traits have been found to be polygenic

(Saastamoinen *et al.* 2018). It follows that the genetic mechanism controlling exploration tendency would also be polygenic in nature.

LITERATURE CITED

Alcock J., and D. R. Rubenstein, 2009 *Animal Behavior, 11th ed.* Sinauer Associates.

Arika W. M., C. M. Kibiti, J. M. Njagi, and M. P. Ngugi, 2019 Effects of DCM Leaf Extract of Gnidia glauca (Fresen) on Locomotor Activity, Anxiety, and Exploration-Like Behaviors in High-Fat Diet-Induced Obese Rats. Behav. Neurol. 2019: 7359235.

Arvidsson L. K., F. Adriaensen, S. van Dongen, N. De Stobbeleere, and E. Matthysen, 2017 Exploration behaviour in a different light: testing cross-context consistency of a common personality trait. Anim. Behav. 123: 151–158.

Bass T. M., R. C. Grandison, R. Wong, P. Martinez, L. Partridge, *et al.*, 2007 Optimization of dietary restriction protocols in Drosophila. J. Gerontol. A Biol. Sci. Med. Sci. 62: 1071–1081.

Bengston S. E., R. A. Dahan, Z. Donaldson, S. M. Phelps, K. van Oers, *et al.*, 2018 Genomic tools for behavioural ecologists to understand repeatable individual differences in behaviour. Nat Ecol Evol 2: 944–955.

Bosch D. S., B. van Swinderen, and S. S. Millard, 2015 Dscam2 affects visual perception in Drosophila melanogaster. Front. Behav. Neurosci. 9: 149.

Burke M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature 467: 587–590.

Crawford P. J., 2013 The genetic basis of variation in thermal plasticity in Drosophila melanogaster

Cridland J. M., S. J. Macdonald, A. D. Long, and K. R. Thornton, 2013 Abundance and distribution of transposable elements in two Drosophila QTL mapping resources. Mol. Biol. Evol. 30: 2311–2327.

Dingemanse N. J., C. Both, P. J. Drent, K. van Oers, and A. J. van Noordwijk, 2002 Repeatability and heritability of exploratory behaviour in great tits from the wild. Anim. Behav. 64: 929–938.

Everman E. R., C. L. McNeil, J. L. Hackett, C. L. Bain, and S. J. Macdonald, 2019 Dissection of Complex, Fitness-Related Traits in Multiple Drosophila Mapping Populations Offers Insight into the Genetic Control of Stress Resistance. Genetics 211: 1449–1467.

Fitzpatrick M. J., Y. Ben-Shahar, H. M. Smid, L. E. M. Vet, G. E. Robinson, *et al.*, 2005 Candidate genes for behavioural ecology. Trends Ecol. Evol. 20: 96–104.

Fountain T., A. Husby, E. Nonaka, M. F. DiLeo, J. H. Korhonen, *et al.*, 2018 Inferring dispersal across a fragmented landscape using reconstructed families in the Glanville fritillary butterfly. Evol. Appl. 11: 287–297.

Highfill C. A., G. A. Reeves, and S. J. Macdonald, 2016 Genetic analysis of variation in lifespan using a multiparental advanced intercross Drosophila mapping population. BMC Genet. 17: 113.

Kessner D., and J. Novembre, 2015 Power Analysis of Artificial Selection Experiments Using Efficient Whole Genome Simulation of Quantitative Traits. Genetics 199: 991–1005.

King E. G., C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen, *et al.*, 2012 Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res. 22: 1558–1566.

King E. G., S. J. Macdonald, and A. D. Long, 07/2012 Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. Genetics 191: 935–949.

Kolbert E., 2014 *The Sixth Extinction: An Unnatural History.* A&C Black.

Lah G. J.-E., J. S. S. Li, and S. S. Millard, 2014 Cell-specific alternative splicing of Drosophila Dscam2 is crucial for proper neuronal wiring. Neuron 83: 1376–1388.

Lemanski N. J., C. N. Cook, B. H. Smith, and N. Pinter-Wollman, 2019 A Multiscale Review of Behavioral Variation in Collective Foraging Behavior in Honey Bees. Insects 10. https://doi.org/10.3390/insects10110370

Long A. D., S. J. Macdonald, and E. G. King, 2014 Dissecting complex traits using the Drosophila Synthetic Population Resource. Trends Genet. 30: 488–495.

Lynch M., D. Bost, S. Wilson, T. Maruki, and S. Harrison, 2014 Population-Genetic Inference from Pooled-Sequencing Data. Genome Biol. Evol. 6: 1210–1218.

Magwene P. M., J. H. Willis, and J. K. Kelly, 2011 The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing. PLoS Comput. Biol. 7: e1002255.

Marriage T. N., E. G. King, A. D. Long, and S. J. Macdonald, 2014 Fine-mapping nicotine resistance loci in Drosophila using a multiparent advanced generation inter-cross population. Genetics 198: 45–57.

Michelmore R. W., I. Paran, and R. V. Kesseli, 1991 Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences 88: 9828–9832.

Millard S. S., J. J. Flanagan, K. S. Pappu, W. Wu, and S. L. Zipursky, 2007 Dscam2 mediates axonal tiling in the Drosophila visual system. Nature 447: 720–724.

Moran N. P., K. D. Mossop, R. M. Thompson, D. G. Chapple, and B. B. M. Wong, 2017 Rapid divergence of animal personality and syndrome structure across an arid-aquatic habitat matrix. Oecologia 185: 55–67.

Mouchet A., E. F. Cole, E. Matthysen, M. Nicolaus, J. L. Quinn, *et al.*, 2021 Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird. Proc. Natl. Acad. Sci. U. S. A. 118. https://doi.org/10.1073/pnas.2024994118

Najarro M. A., J. L. Hackett, B. R. Smith, C. A. Highfill, E. G. King, *et al.*, 2015 Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. PLoS Genet. 11: e1005663.

Odierna G. L., S. K. Kerwin, L. E. Harris, G. J.-E. Shin, N. A. Lavidis, *et al.*, 2020 Dscam2 suppresses synaptic strength through a PI3K-dependent endosomal pathway. J. Cell Biol. 219. https://doi.org/10.1083/jcb.201909143

Paul M. H., M. Choi, J. Schlaudraff, T. Deller, and D. Del Turco, 2020 Granule Cell Ensembles in Mouse Dentate Gyrus Rapidly Upregulate the Plasticity-Related Protein Synaptopodin after Exploration Behavior. Cereb. Cortex 30: 2185–2198.

Pool J. E., 2016 Genetic Mapping by Bulk Segregant Analysis in Drosophila: Experimental Design and Simulation-Based Inference. Genetics 204: 1295–1306.

R Core Team, 2022 R: A Language and Environment for Statistical Computing

Ramírez-Amaya V., A. Vazdarjanova, D. Mikhael, S. Rosi, P. F. Worley, *et al.*, 2005 Spatial exploration-induced Arc mRNA and protein expression: evidence for selective, network-specific reactivation. J. Neurosci. 25: 1761–1768.

Réale D., S. M. Reader, D. Sol, P. T. McDougall, and N. J. Dingemanse, 2007 Integrating animal temperament within ecology and evolution. Biol. Rev. Camb. Philos. Soc. 82: 291–318.

Riddle M., and N. Farnoudi, Quantitative Trait Locus Mapping Identifies Candidate Fungal Resistance Genes in Drosophila melanogaster. The Journal of Undergraduate Research in Natural Sciences and Mathematics 31.

Rudin F. S., L. W. Simmons, and J. L. Tomkins, 2018 Social cues affect quantitative genetic variation and covariation in animal personality traits. Evolution. https://doi.org/10.1111/evo.13661

Saastamoinen M., G. Bocedi, J. Cote, D. Legrand, F. Guillaume, *et al.*, 2018 Genetics of dispersal. Biol. Rev. Camb. Philos. Soc. 93: 574–599.

Singh A., A. Hasan, and A. F. Agrawal, 2022 An investigation of the sex-specific genetic architecture of fitness in Drosophila melanogaster. bioRxiv 2022.08.04.502812.

Smith B. R., and S. J. Macdonald, 2020 Dissecting the Genetic Basis of Variation in Drosophila Sleep Using a Multiparental QTL Mapping Resource. Genes 11. https://doi.org/10.3390/genes11030294

Sokolowski M. B., H. S. Pereira, and K. Hughes, 1997 Evolution of foraging behavior in Drosophila by density-dependent selection. Proc. Natl. Acad. Sci. U. S. A. 94: 7373–7377.

Stanley P. D., E. Ng'oma, S. O'Day, and E. G. King, 2017 Genetic Dissection of Nutrition-Induced Plasticity in Insulin/Insulin-Like Growth Factor Signaling and Median Life Span in a Drosophila Multiparent Population. Genetics 206: 587–602.

Tilk S., A. Bergland, A. Goodman, P. Schmidt, D. Petrov, *et al.*, 2019 Accurate Allele Frequencies from Ultra-low Coverage Pool-Seq Samples in Evolve-and-Resequence Experiments. G3 9: 4159–4168.

Verbeek M. E. M., P. J. Drent, and P. R. Wiepkema, 1994 Consistent individual differences in early exploratory behaviour of male great tits. Anim. Behav. 48: 1113–1121.

Williams-Simon P. A., C. Posey, S. Mitchell, E. Ng'oma, J. A. Mrkvicka, *et al.*, 2019 Multiple genetic loci affect place learning and memory performance in Drosophila melanogaster. Genes Brain Behav. 18: e12581.

Yamakawa K., Y. K. Huot, M. A. Haendelt, R. Hubert, X. N. Chen, *et al.*, 1998 DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. Hum. Mol. Genet. 7: 227–237.

SUPPLEMENTAL MATERIALS

Recombinant inbred lines	
22004	22035
22009	22036
22014	22037
22017	22038
22018	22039
22024	22042
22028	22046
22029	22049
22030	22052
22033	22054

Supplemental table 1. List of recombinant inbred lines used for the outbred population.

CHAPTER 3

STARVATION RESISTANCE IN AN EXPERIMENTALLY EVOLVED POPULATION OF *D. MELANOGASTER*

Z. Forrest Elkins, Jordyn Moaton, Elizabeth G. King

ABSTRACT

An organism's access to food fluctuates over the course of its life. In many populations, organisms experience prolonged lengths of time with limited or no access to food. In such cases, an organism's ability to resist starvation is essential to its survival. Starvation resistance (SR) can differ among individuals due to both genetic differences as well as environmental differences. We used an experimental evolution framework to assay for SR in three different populations of *D. melanogaster* selected for three different resource allocation patterns. Our goal was to see how three different diets affected SR in each of the three evolved populations. We placed twelve replicate populations of *D. melanogaster* on three selection treatments: constant high availability (CHA), fluctuating availability (FA) and deteriorating availability (DA). Flies in the CHA treatment were fed a high sugar diet for their entire lifespan. Flies in the FA treatment are in turn given a standard diet followed by a low yeast diet followed by a standard diet. Flies in the DA treatment start with a standard diet and then are put on a low yeast diet until eggs are collected for the next generation. These three treatments

have been ongoing for over 50 generations. For our experiment, flies from all replicates and treatments were placed on one of three diets for 10 days: high sugar, standard and low yeast diets. After 22 days post-oviposition (p.o.), flies were placed on nutrition-less agar. SR was measured as the time it takes for a fly to die starting the moment it is placed on nutrition-less agar. We determined how these different selection treatments led to the evolution of plasticity in SR in response to diet. We link these phenotypic changes to variation in artificial selection pressure and environmental conditions.

INTRODUCTION

Nearly all species experience periods of food scarcity or nutritional imbalance (McCue 2010). An organism's ability to survive these periods of starvation and malnutrition is important for its overall fitness. Organisms without the ability to resist periods of stress exhibit reduced survival (Lee and Jang 2014). Starvation resistance is correlated with various stress response traits, longevity and energy storage (Chippindale *et al.* 1996; Djawdan *et al.* 1998; Harshman *et al.* 1999; Schwasinger-Schmidt *et al.* 2012; Lee and Jang 2014; Everman *et al.* 2019). Since starvation resistance allows organisms to survive periods of little to no food, it comes as no surprise that it would evolve alongside desiccation resistance, reduced fecundity and general stress resistance (Service *et*

al. 1985; Hoffmann and Parsons 1989b; Rose *et al.* 1992; van Herrewege and David 1997; Hoffmann and Harshman 1999).

We suspect that starvation resistance evolves alongside resource allocation strategies. Previous research found that lines of *D. melanogaster* selected for starvation resistance tended to store more lipid, develop more slowly, show lower rates of early-age reproduction, and grow to be larger than control lines (Harshman *et al.* 1999). Flies selected for starvation resistance were also able to resist desiccation, suggesting that the mechanism underlying starvation resistance is linked to other stress-resistant traits (Hoffmann and Parsons 1989b; Harshman *et al.* 1999). For example, artificially selecting for starvation resistance negatively correlates with fecundity, and positively correlates with longevity, consistent with the Y-model of trade-offs (Chippindale *et al.* 1993). Starvation resistance has also been found to negatively correlate with cold resistance another stress-related trait, in female *Drosophila* (Hoffmann *et al.* 2005). There is even evidence to suggest pre-adult temperatures mediate resistance to starvation, with lower ambient temperatures resulting in reduced resistance (Pijpe *et al.* 2007). Starvation resistance is a phenotypically plastic trait, i.e., a single genotype can give rise to multiple phenotypes depending on the environment surrounding it (Pijpe *et al.* 2007). Environmental differences in access to lipids, or different environments that lead to varying lipid levels, can lead to increased resistance to starvation (Djawdan *et al.* 1998). Changes in lipid content directly contribute to

starvation resistance in *D. melanogaster*, as well (Chippindale *et al.* 1996; Djawdan *et al.* 1998; Hoffmann and Harshman 1999). Therefore, it would make sense that various resource allocation strategies, with variable levels of energy and lipid storage, would lead to differences in starvation resistance. While we know a lot about starvation resistance and its relation to other stress-related traits, we do not know how it co-evolves in response to consistent, cross-generational variation in resources. The genetic mechanisms underlying starvation resistance and phenotypic plasticity of starvation resistance are generally unknown, as well.

Experimental evolution is a powerful paradigm by which lab populations are placed in highly controlled environments to select for specific traits (Garland and Rose 2009). The evolved organisms can then be resequenced, and changes in the allele frequency of single nucleotide polymorphisms (SNPs) or genomic haplotypes that occur between the unevolved and evolved populations are attributed to the artificial selection (Baldwin-Brown *et al.* 2014). We can also compare phenotypic trait values between populations that were placed on differing artificial selection regimes. This method allows us to address our unknown questions: 1) we can artificially evolve populations on drastically different resource availabilities and assay for starvation resistance, and 2) we can place evolved organisms on a specific diet prior to the starvation resistance assay to assess phenotypic plasticity.

We utilize the *Drosophila* Synthetic Population Resource (DSPR), a multi-parent intercross mapping panel of *D. melanogaster*, a powerful tool that allows us to dissect phenotypic variation and its underlying genetic basis to address this question (King *et al.* 2012, 07/2012; Long *et al.* 2014). We used a subset of flies from the DSPR to perform an evolve and resequence experiment where three distinct resource allocation strategies were selected for via cross-generational variation in access to nutritional resources. We then assayed for starvation resistance in flies from these three selection lines. This directly addresses the co-evolution of starvation in response to cross-generational variation in resources. Finally, we conducted two different starvation assays: a 'direct measurement' experiment where flies were placed directly onto the starvation resistance assay, and a 'diet treatment' experiment, where flies were placed onto one of three diets for ten days prior to the assay. The diet treatment assay allows us to address our third unknown – the contributions of the organism's nutritional environment on phenotypic plasticity in starvation resistance.

METHODS

Base population

We used flies taken from the *Drosophila* Synthetic Population Resource (DSPR). These flies are an eight-way multi-parent population comprised of >800 recombinant inbred lines (RILs). The development and genetic properties of the DSPR can be found in (Elizabeth G. King, Macdonald, and Long 07/2012; E. G. King et al. 2012). The founder lines were crossed for 50 generations and then inbred for 25 generations, forming the set of recombinant inbred lines. We crossed 835 of these RILs together for 5 generations to form an outbred panel by placing 5 mated females from each line in an acrylic cage. The outbred panel served as our base population for our artificial evolution experiment.

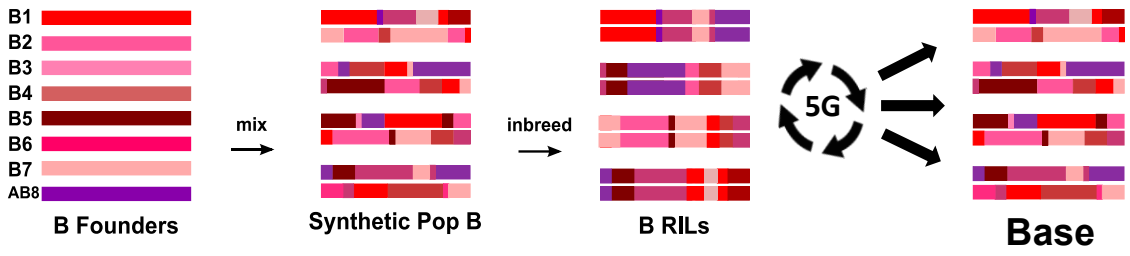


Figure 1. This population started from eight inbred founders representing global standing variation in *Drosophila melanogaster*. They were mixed *en masse* for 50 generations and 835 recombinant inbred lines (RILs) were inbred from this population. The RILs were then re-mixed for five generations to create a base population for experimental evolution.

Experimental evolution

Flies from the outbred population were used in our artificial selection experiment. We selected for three different resource allocation patterns in our outbred populations of flies. The fluctuating availability (FA) selection regime consists of a diet that fluctuates over the course of the fly's lifetime. Prior to eclosure, they are placed on a control (C) diet (Bass *et al.* 2007). After six days on the control diet, they are transferred to a dietary restriction (DR) diet (Bass *et al.* 2007). After four more days, they are placed back onto the control diet. Eggs are collected from the flies after four days on the control diet. The deteriorating availability (DA) selection regime consists of a diet that deteriorates over time. These flies are placed onto a control diet prior to eclosure and stay on that diet for ten days. They are then placed onto the DR diet, and eggs are collected for the next generation after four days. The final selection regime is a high sugar (HS) dietary regime. These flies are placed onto a high sugar diet prior to eclosure and remain on that diet for fourteen days before eggs were collected from their offspring.

	Days post-oviposition													
	8	9	10	11	12	13	14	15	16	17	18	19	20	21
FA	C	C	C	C	C	C	DR	DR	DR	DR	C	C	C	C
DA	C	C	C	C	C	C	C	C	C	C	DR	DR	DR	DR
CH	HS	HS	HS	HS	HS	HS	HS	HS	HS	HS	HS	HS	HS	HS
			Eclosure											Egg collection

Table 1. Selection regime schedule. Flies eclosed from their eggs on day 10 post-oviposition (p.o.). Flies on the fluctuating availability selection regime were placed on a control diet for 6 days, then onto the DR diet for 4 days, and spent the last 4 days on the control diet again, with egg collection occurring on the last day. Flies on the deteriorating availability selection regime were placed on the control diet for 10 days before being moved to the DR diet for the final 4 days, with egg collection occurring on the last day. Flies on the constant high availability selection regime were kept on the HS diet for the entire 14-day period, with egg collection occurring on day 21 p.o.

Starvation resistance assays

We ran two separate SR assays using flies that were duplicated out of our selection population. We then maintained the duplicated flies for two generations, and we used the second generation's offspring in our two experimental assays. The first and second generation of duplicated flies were kept on a standard yeast diet to remove maternal and transgenerational effects of the selection regimes on our experimental flies. In the direct measurement, eggs collected from the second

generation of duplicated flies were placed onto a standard diet until eclosure. Once the flies eclosed, they were placed directly onto the nutrition-less agar. We will refer to these as direct measurement flies (Figure 2).

In the second assay, eggs collected from the third generation of duplicated flies were placed on one of three diets for ten days before they were transferred to the nutrition-less agar. These flies will be referred to as diet treatment flies (Figure 3). The nutrition-less agar consists of 1000 mL of water, 15 g of agar, 12 mL of acid mix (330 mL water/259 mL propionic acid/31 mL phosphoric acid), and 2 g of Tegosept dissolved in 20 mL of 95% ethanol.

	Maintenance	**DR**	**C**	**HS**	**Nutrition-less**
Water (ml)	1066	1000	1000	1000	1000
Agar (g)	6.25	10	10	10	15
Dextrose (g)	86.26	-	-	-	-
Sucrose (g)	-	50	50	342	-
Yeast (g)	21.6	100	200	200	-
Cornmeal (g)	40.8	-	-	-	-
Tegosept (g)	1.8	2.7	2.7	2.7	2
Ethanol (ml)	7.3	11	11	11	20
Acid mix (330 ml water/ 259 ml proprionic acid/ 31 ml phosphoric acid)	-	-	-	-	12

Table 2. The four diets used in the selection experiment, plus the nutrition-less agar diet used in our starvation resistance assay.

Direct measurement

Flies were duplicated from the three experimental selection regimes for two generations before they were placed onto the nutrition-less agar starvation resistance assay. The time until death (counted in 12-hour increments) for these direct measurement flies were recorded. Flies were labeled by their RIL line, replicate number, and sex (ex: 1FA-1-M). Each vial contained 10 flies of a single sex. Flies were sorted and sexed using pre-sterilized paintbrushes with a 10% bleach mix. To sort and sex the flies, we used CO2 to put them to sleep. The CO2 bed was sterilized using a Kim wipe that had been dampened with a 10% bleach mix. Each vial was checked post-sorting to make sure the flies did not die from the CO2 bed.

The experimental vials were placed in a temperature- and light-controlled chamber. After the number of dead flies were counted every 12 hours, the vials' placement in the chamber was rotated to control for environmental effects.

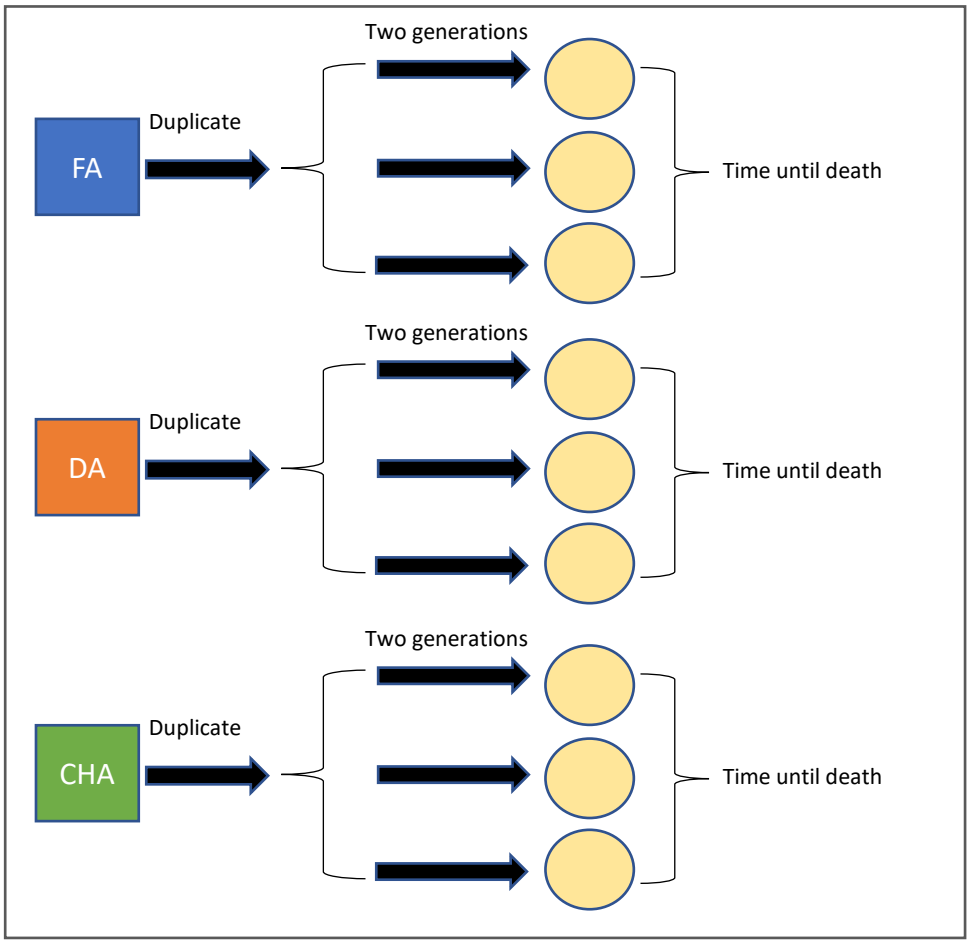


Figure 2. Flies were duplicated from the three experimental selection regimes for two generations before they were placed onto the nutrition-less agar starvation resistance assay. The time until death (counted in 12-hour increments) for these direct measurement flies were recorded.

Diet treatment

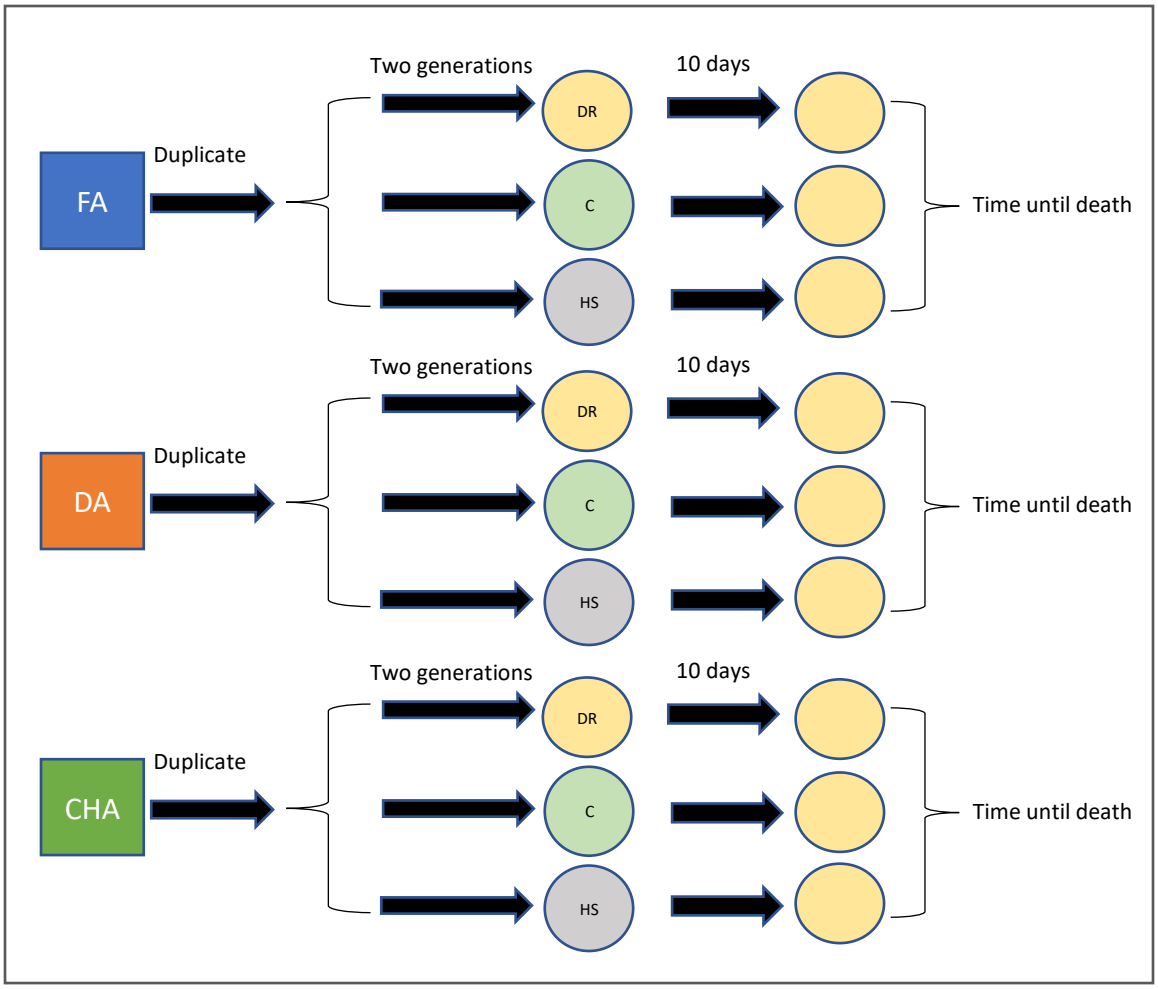


Figure 3. Flies were duplicated from the three experimental selection regimes for two generations before they were placed on one of three diets – dietary restrictive, control, and high sugar diets. After ten days, all flies were placed onto the nutrition-less agar starvation resistance assay and the time until death (counted in 12-hour increments) for these diet treatment flies were recorded.

Ten flies were placed into a vial containing nutrition-less agar. Flies were labeled by their RIL line, replicate number, sex and diet (ex:1FA-1-M-C). Each vial contained 10 flies of a single sex. Flies were sorted and sexed using pre-

sterilized paintbrushes with a 10% bleach mix. In order to sort and sex the flies, we used CO2 to put them to sleep. The CO2 bed was sterilized using a Kim wipe that had been dampened with a 10% bleach mix. Each vial was checked post-sorting to make sure the flies didn't die from the CO2 bed.

The experimental vials were placed in a temperature- and light-controlled chamber. After the number of dead flies were counted every 12 hours, the vials' placement in the chamber was rotated to control for environmental effects.

Phenotypic analysis

At the end of the experiment, we had the time (in 12-hour increments) at which flies from every vial died (i.e., the time that they resisted starvation). We then utilized R and, more specifically, the `tidyverse` package to clean and prepare our data for analysis and visualization (Wickham *et al.* 2019; R Core Team 2022). From our recorded data, we calculated the total time, per sample, when all flies in a sample had died. We right-censored data from vials in which flies never experienced death due to starvation. This would occur in flies that died from methods other than starvation (i.e., a fly was crushed by the vial's foam plug during the transfer of flies to nutrition-less agar), or other experimental error. Since the flies, prior to censorship, did not experience death due to starvation, we can use that knowledge in our survivorship analysis. These

flies still contribute to our survival analysis as censored data, which is better than dropping those vials from our analysis completely.

To conduct survival analyses, including a Cox Proportional Hazard (CPH) model, we needed to calculate several factors from our gathered data. These factors included the selection treatment, diet, sex, and batch number, as well as our right-censored data and total time until complete death per vial.

We conducted survival analyses on our cleaned and prepped SR phenotype data (Clark et al. 2003b, 2003a; Bradburn et al. 2003a, 2003b; Zabor 2022; Therneau 2022).

We then fitted our survival data using the Kaplan-Meier (KM) method using the `'ggsurvfit::survfit2'` function (Clark et al. 2003a; Kaplan and Meier 1958; Sjoberg and Baillie 2022; R Core Team 2022). This method estimates survival probability nonparametrically. Each vial is observed at regular time intervals, with the assumption that the occurrence of an event (in our case, the death of all ten flies in a given vial) happens independently of each other. Therefore, the probabilities of surviving across time intervals can be multiplied together. This gives us a cumulative survival probability where each vial contributes information to the KM method if no event has occurred. It also allows censored data to contribute to the calculations (Kaplan and Meier 1958).

Next, we fitted our data using the Cox proportional hazard model (Andersen and Gill 1982). The Cox regression model is semi-parametric and can

fit univariable and multivariable regression models that have survival outcomes. The Cox model assumes 1) non-informative censoring and 2) proportional hazards. Cox regression models output what is known as a hazard ratio (HR). The HR represents proportional hazards between two groups at a given time point. For our data, the HR can be interpreted as the proportional, instantaneous rate of occurrence of death due to starvation in flies still at risk of starvation. HR values less than 1 indicate a reduced hazard of death, and an HR greater than 1 indicates an increased hazard of death.

Once we fitted our Cox proportional hazard models, we then used Akaike's Information criterion (AIC) to perform a model comparison (Mazerolle 2020). We used the model comparison to find the Cox model that best fit our data.

Direct measurement

For the direct measurement flies, we fitted the survival data with eight different models using the `'survival::Surv'` and `'survival::coxph'` functions in R (Therneau 2022; R Core Team 2022). For our first fitted model, we wanted to see how the selection ID fit our data. For our second model, we fitted our data by selection ID and sex. Then, to make sure that our two batches didn't influence our survival data, we fitted our data by batch ID. We then used the `'AICcmodavg::aictab'` function in R to perform our model comparison (Mazerolle 2020; R Core Team 2022).

Cox proportional hazard models
`Survival ~ Selection + Sex + Batch`
`Survival ~ Selection`
`Survival ~ Selection + Sex`
`Survival ~ Sex`
`Survival ~ Selection * Sex`
`Survival ~ Batch`

Table 3. Cox proportional hazard models that we ran on our direct measurement data.

Diet treatments

For the diet treatment flies, we fitted the survival data with seven models, then performed model comparison to find the best-fitting model. First, we wanted to see how diet predicted our survival data. For our second fitted model, we wanted to see how diet plus the selection ID fit our data. For our third model, we fitted our data by diet, selection ID, and sex. Finally, to make sure that our two batches did not influence our survival data, we fitted our data by batch ID.

Cox proportional hazard models

`Survival ~ Diet + Selection + Sex + Batch`
`Survival ~ Diet`
`Survival ~ Diet + Selection`
`Survival ~ Diet + Selection + Sex`
`Survival ~ Selection`
`Survival ~ Sex`
`Survival ~ Diet * Selection`

Table 4. The seven Cox proportional hazard models that we ran on our diet treatment data.

RESULTS

Direct measurement

In our direct measurement flies, sex was a major predictor of starvation resistance according to both Kaplan-Meier and Cox proportional hazard models. Males were always more likely to die at earlier time periods than female flies ($\chi_1^2 = 25.08$, p = 5.507e-07) (Figure 6). In fact, male flies were three times more likely to die at any given time point during the starvation resistance assay than female flies (Table 1).

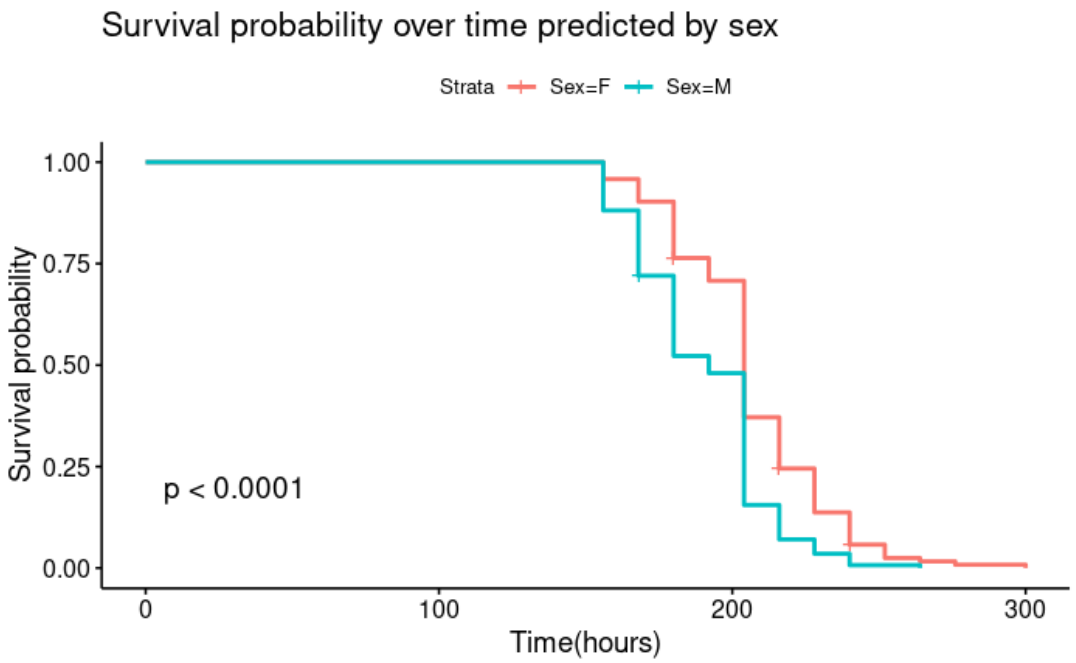


Figure 4. Sex was a significant predictor ($\chi_1^2 = 25.08$, p = 6.732e-09) of survival in the direct measurement (DM) flies.

In our direct measurement flies, the selection regime significantly contributed to differences in survival probability during the starvation resistance assay ($\chi_2^2 = 12.09$, p = 0.002371). Flies from the CH and FA selection regimes were much more likely to survive at any given time point during the assay than flies from the DA selection regime (Figure 7). Unlike our flies from the diet treatment, where the contributions of selection regime appeared to be washed out by diet treatment, we clearly saw differences in survival due to selection regime in our direct measurement flies.

We also saw that patterns of starvation resistance due to selection regime remained the same regardless of sex (Figure 8). When we predict survival

probability by both sex and selection regime, males and females on the CH and FA selection regimes resisted starvation for longer periods of time than flies from both sexes on the DA selection regime. Adding to this, males always died quicker than females, with male flies on the CH selection regime resisting starvation at a similar rate as female flies on the FA selection regime.

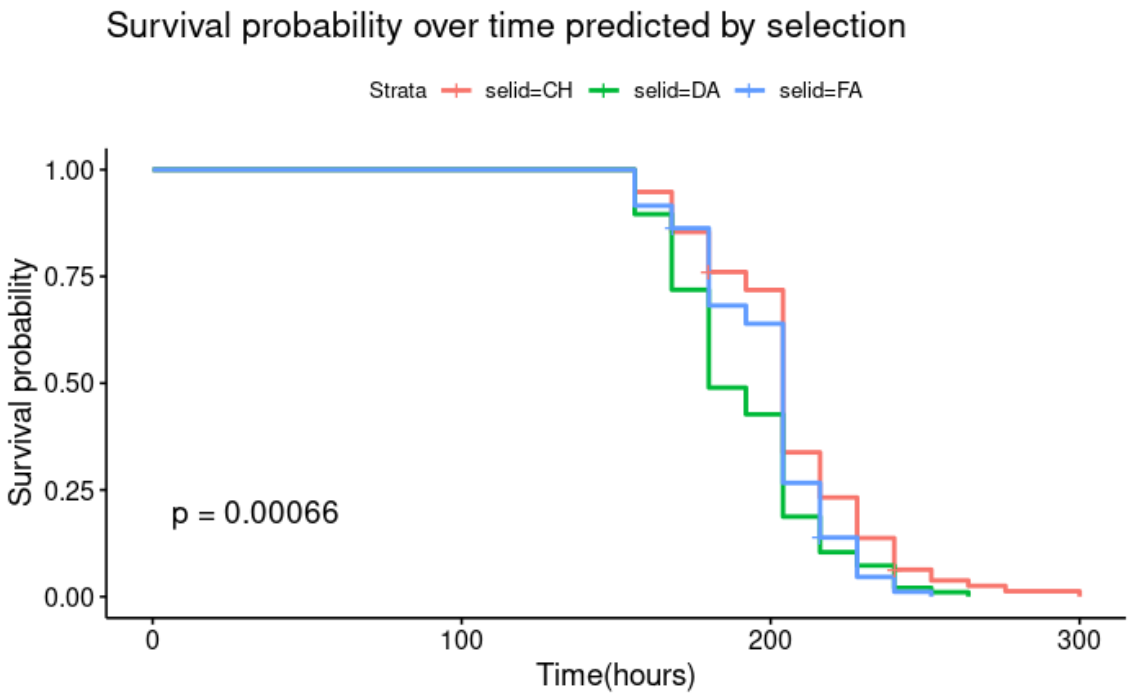


Figure 5. Selection was a significant predictor of survival in the direct measurement (DM) flies. Flies from both the CH and FA selection regimes resisted starvation for longer periods of time than DA flies, with CH flies surviving the longest.

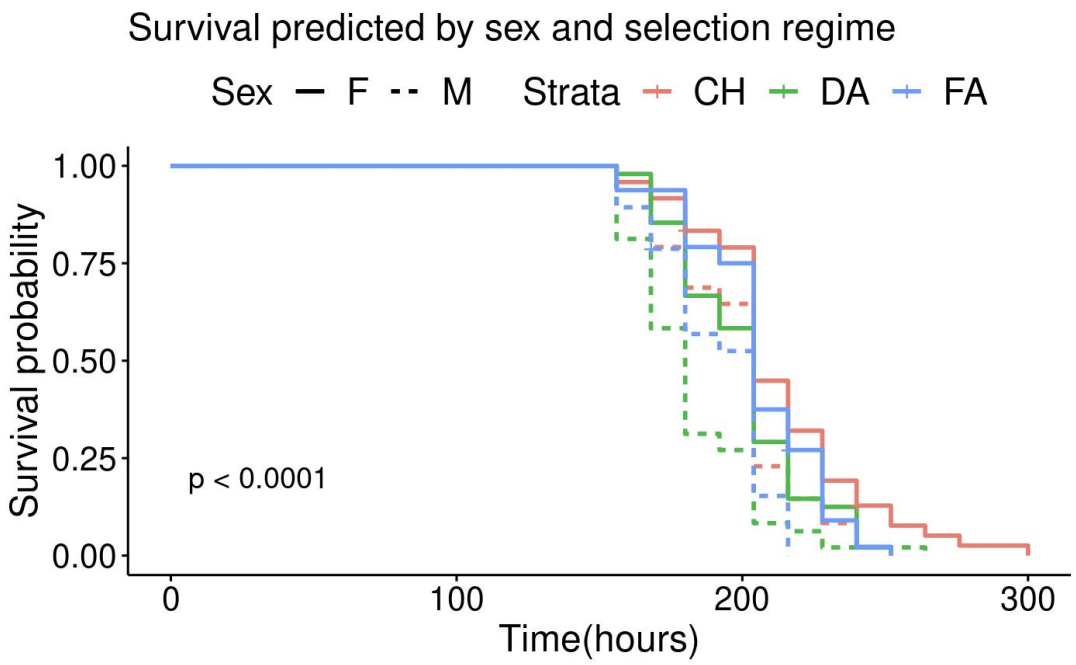


Figure 6. Selection and sex together were a significant predictor ($\chi_3^2 = 39.54$, p = 1.335e-08) of survival probability. Males, regardless of selection regime, died quicker than females. However, our selection regime results are the same regardless of sex. Both males and females on the CH and FA selection regimes resisted starvation longer than males and females on the DA selection regime.

Cox proportional hazard model

We fit Cox proportional hazard models to our survival data to see how selection and sex altered the hazard ratio, which in our case, is the likelihood that a fly will die from starvation at a given time. Hazard ratios (HR) < 1 indicate a reduced hazard of death due to SR. HR > 1 indicates an increased hazard of death due to SR. The Cox proportional hazard model uses one category from each dependent variable as a 'baseline' hazard around which the other categories

are compared to. For selection regime, the Cox proportional hazard model used the CH regime as its baseline. Compared to CH, flies from the DA regime were 1.73 times as likely to die due to starvation at a given point in time. Likewise, FA flies were 1.41 times as likely to die due to starvation than CH flies. For sex, female flies were 'baseline,' and male flies were found to be 1.9 times more likely to die due to starvation than females at a given point in time. We see selection treatment significantly influence starvation resistance ($\chi_2^2 = 12.09$, p = 0.002371), while selection + sex + batch was our top predictor for survival ($\chi_4^2 = 43.9$, p = 6.732e-09).

Characteristic	**HR[1]**	**95% CI[1]**	**p-value**
Selection			
CH	-	-	
DA	1.73	1.30, 2.31	<0.0001
FA	1.41	1.05, 1.88	0.022
Sex			
F	-		
M	1.90	1.50, 2.42	<0.001

[1] HR = Hazard Ratio, CI = Confidence Interval

Table 5. Cox proportional hazard model for direct measurement data. Hazard ratios (HR) < 1 indicate reduced hazard of death. HR > 1 indicate an increased hazard in death. A HR of 1 indicates neither a reduction nor an increase in likelihood of death. In the selection treatments, HR values of 1.73 and 1.41 imply that roughly 1.7 and 1.4 times as many flies in the deteriorating availability (DA) and fluctuating availability (FA) selection regimes, respectively, die in

comparison to the constant high (CH) availability selection regime. We also saw a higher hazard ratio for male flies, with male flies dying at 1.9 times the rate of female fly death.

Model comparison

We used AICc from the AICcmodavg package in R to compare our Cox proportional hazard models (Mazerolle 2020). From this comparison, we found that an additive model of selection + sex + batch best fit our data, and an interaction between selection and sex fit our data third best (Table 3). There is not much evidence to suggest that a more complex interaction occurs between selection and sex.

Model	K	AICc	Delta AICc	AICc Weight	Cum. Wt.	LL
Selection + Sex + Batch	4	2612.94	0.00	0.74	0.74	-1302.40
Selection + Sex	3	2615.25	2.31	0.23	0.97	-1304.58
Selection * Sex	5	2619.27	6.33	0.03	1.00	-1304.53
Sex	1	2625.64	12.69	0.00	1.00	-1311.81
Selection	2	2640.65	27.71	0.00	1.00	-1318.31
Batch	1	2647.53	34.58	0.00	1.00	-1322.76

Table 6. AICc model comparison of seven models that we used to fit our direct measurement data. Selection + Sex + Batch fit our data the best out of all our Cox proportional hazard models, with Selection + Sex and Selection * Sex fitting second and third best, respectively.

Diet treatments

We can see that flies in the diet treatments experiment were heavily influenced by their 10-day pre-SR assay diet treatment (Figure 9). Flies on the DR and HS diets both were more resistant to starvation than the C diet flies, although we did see an outlier fly from a C diet treatment living abnormally long on the SR assay.

As for the interplay between diet treatment and selection regime, we did notice that the two combined additively to significantly contribute to SR survival in both Kaplan-Meier and Cox proportional hazard models. While diet treatments exhibit a much more significant effect on SR survival than diet plus selection regime, and selection regime exhibits a much less significant effect on SR survival than both diet and diet plus selection regime, the additive interaction between the two is significant in our Cox proportional hazard analysis (diet – $\chi_2^2 = 15.4$, p = 5e-04; selection – $\chi_2^2 = 0.01$, p = 1; diet + selection – $\chi_4^2 = 16.76$, p = 0.002). We observe that flies from the CH and FA selection regimes, plus the HS and DR diets, resist starvation for longer periods of time than flies from the DA selection regime and C dietary treatment.

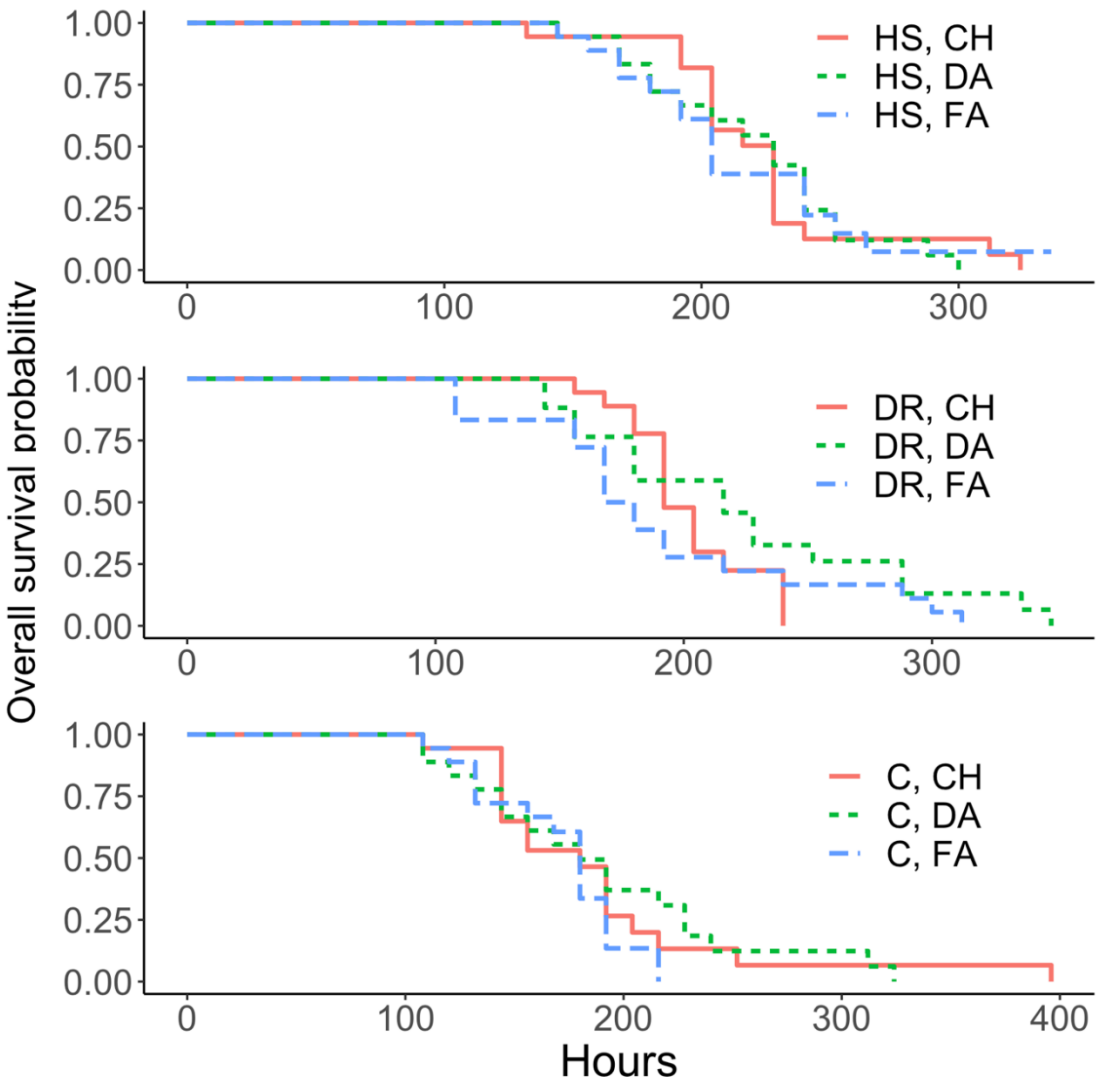


Figure 7. The overall survival probability predicted by an additive model of diet + selection. Each row denotes diet treatment, with the first row containing flies on the HS diet treatment, the second row the DR diet treatment, and the third row the C diet treatment. Each plot shows differences in survival between the three selection regimes, denoted by color and linetype.

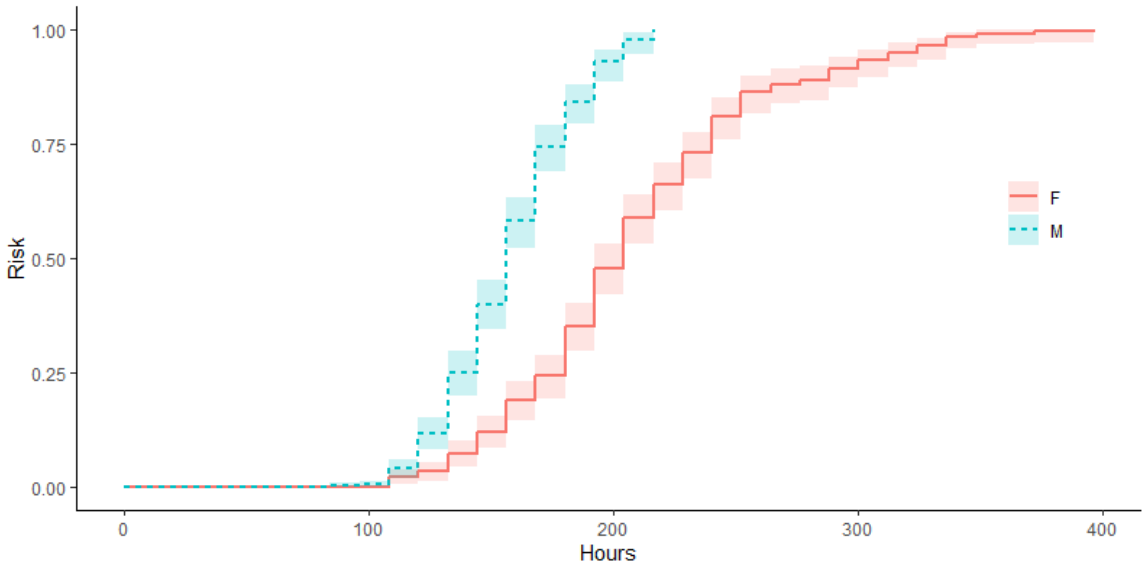


Figure 8. The risk (1 – survival probability) at a given point of time between male and female diet treatment flies. As time went on, the risk of death for males approached 1 prior to the 200-hour mark, while the risk of death for females approached 1 between 300 and 400 hours. As you can see, female flies lived much longer than male flies, regardless of diet or selection regime. The difference in survival probability between males and females is significant in our KM analysis ($p < 0.001$).

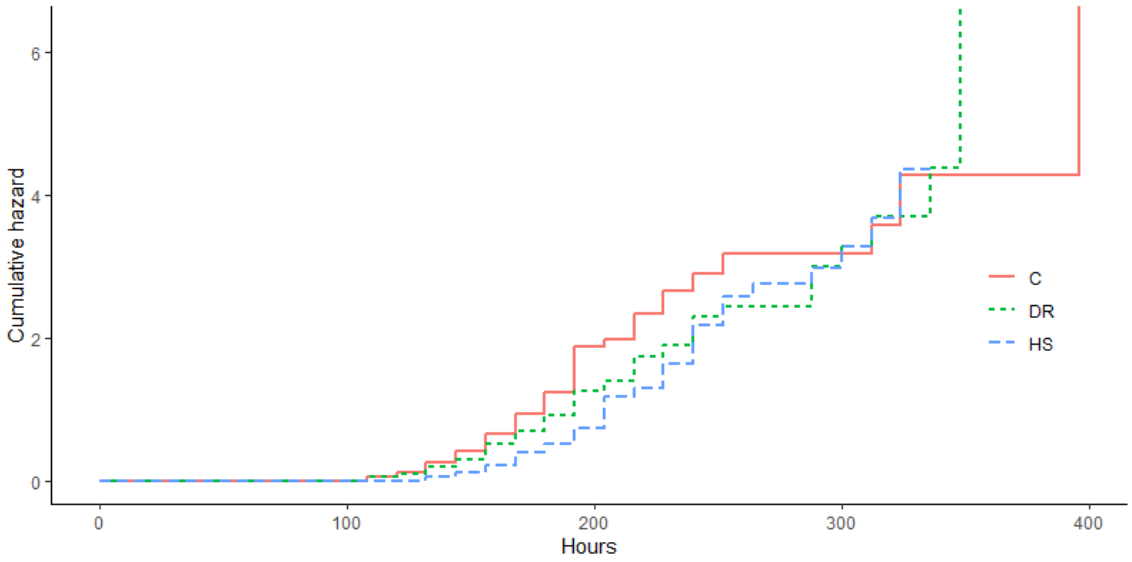


Figure 9. The cumulative hazard (-log (survival probability)) over time for flies on three separate dietary treatments. The cumulative hazard for flies from the control diet accumulated the quickest, with flies from the DR and HS diet treatments accumulating hazard at a slower rate. Survival probability between diets was significant in our KM analysis ($p < 0.001$).

Cox proportional hazard model

Our Cox proportional hazard model gives us more insight into our data (Table 7). In the different diet treatments, HR values of 0.72 and 0.49 imply that roughly 0.7 and 0.5 times as many flies on the dietary restrictive and high sugar diets, respectively, die in comparison to the control diet. This matches what we see in Figure 11, where flies from the C diet accumulate hazard at a faster rate than flies from the other diet treatments.

Characteristic	HR[1]	95% CI[1]	p-value
Diet			
C	-	-	
DR	0.72	0.54, 0.96	0.027
HS	0.49	0.37, 0.67	<0.001
Selection			
CH	-	-	
DA	1.16	0.87, 1.54	0.3
FA	1.27	0.94, 1.70	0.12
Sex			
F	-	-	
M	3.34	2.52, 4.43	<0.001

[1] HR = Hazard Ratio, CI = Confidence Interval

Table 7. Cox proportional hazard model for diet treatment data. Hazard ratios (HR) < 1 indicate reduced hazard of death. HR > 1 indicate an increased hazard in death. An HR of 1 indicates neither a reduction nor an increase in likelihood of death. In the diet treatments, HR values of 0.72 and 0.49 indicate that roughly 0.7 and 0.5 times the amount of flies in the diet restriction (DR) and high sugar (HS) diet treatments, respectively, die in comparison to the control (C) diet treatment. In the selection treatments, HR values of 1.16 and 1.27 imply that roughly 1.7 and 1.4 times as many flies in the deteriorating availability (DA) and fluctuating availability (FA) selection regimes, respectively, die in comparison to the constant high (CH) availability selection regime. However, this result was not statistically significant in our Cox proportional hazard model ($p = 0.3, 0.12$). We also saw a higher hazard ratio for male flies, with male flies dying at 3.34 times the rate of female fly death.

Likewise, in the selection treatment, HR values of 1.16 and 1.27 imply that roughly 1.2 and 1.3 times as many flies in the deteriorating availability (DA) and fluctuating availability (FA) selection regimes, respectively, die in comparison to the constant high (CH) availability selection regime. Finally, roughly 3.3 times as many males die than female flies. In our diet treatment, diet overshadows selection in contributing to starvation resistance (diet – $\chi_2^2 = 15.4$, p = 5e-04; selection – $\chi_2^2 = 0.01$, p = 1; diet + selection – $\chi_4^2 = 16.76$, p = 0.002). Overall Cox proportional hazard model trends indicate that in both direct measurement and diet treatment, batch ID does not significantly predict resistance to starvation ($\chi_1^2 = 0.04$, p = 0.8). Sex was our main predictor for starvation resistance, as females were much more likely to live longer than males (direct measurement – $\chi_1^2 = 25.08$, p = 5.507e-07; diet treatment – $\chi_1^2 = 209.5$, p =< 2e-16).

Model comparison

We used `AICc` from the `AICcmodavg` package in R to compare our Cox proportional hazard models (Mazerolle 2020). From this comparison, we found that an additive model of diet + selection + sex best fit our data, and an interaction between diet and selection with sex as an additive effect fit our data second best (Table 4). This indicates a more complex relationship between diet and selection regime may be occurring.

Model	K	AICc	Delta AICc	AICc Weight	Cum. Wt.	LL
Diet + Selection + Sex	5	2626.07	0.00	0.81	0.81	-1307.94
Diet * Sel. + Sex	9	2629.03	2.96	0.19	1.00	-1305.23
Diet	2	2694.03	67.96	0.00	1.00	-1345.00
Diet + Selection	4	2696.76	70.69	0.00	1.00	-1344.32
Diet * Selection	8	2702.52	76.46	0.00	1.00	-1343.03
Sex	1	6160.77	3534.71	0.00	1.00	-3079.38
Selection	2	6372.28	3746.21	0.00	1.00	-3184.13

Table 8. AICc model comparison of seven models that we used to fit our diet treatment data. The model that best explains SR is the Diet + Selection + Sex Cox proportional hazard model, followed closely by the Diet * Selection + Sex Cox proportional hazard model.

DISCUSSION

We found that the top-performing model for diet treatment was Diet + Selection + Sex (AICc = 2626.07), while the second-best model was Diet * Selection + Sex (AICc = 2629.03). The interaction term between diet and selection indicates that there is a more complex relationship between the two that is not only additive. Put another way: while an additive effect of diet, selection and sex has a significant impact on a fly's hazard of death, diet and selection interact with each other with the additive effect of sex to significantly impact a fly's hazard of death as well, albeit to a slightly lower degree. Our top performing models according to our AIC model comparison for direct measurement flies was

Selection + Sex + Batch (AICc = 2612.94), with Selection + Sex nearly identical to it (AICc = 2615.25). However, according to our KM model, batch was not a significant predictor of survival probability (p = 0.1).

Flies from different selection regimes are not statistically different from each other when it comes to SR in the diet treatment flies. However, we observed a statistically significant interaction between diet and selection. Since we observed significant changes in SR due to selection in our direct measurement flies, it seems likely that diet overshadows selection in our diet treatment flies. This is not what we expected, as we predicted that SR would clearly co-evolve alongside differing resource allocation patterns.

While we do see that slightly less flies die from the CHA selection regime compared to the other two, this difference is not statistically significant in our diet treatment flies. The greater predictors of SR, rather, are sex and diet. Male flies were dying from starvation at over three times the rate of death in females at any given time. Likewise, flies that had access to a high sugar diet and flies that had a dietary restricted diet for the first 10 days post-eclosure were dying less than flies on the control diet. We did note some instances of cannibalism in our experimental population, a phenomenon that has been documented in starvation resistance assays before (Vijendravarma *et al.* 2013; Ahmad *et al.* 2015, 2018). It was rare, but present. In future studies on starvation resistance,

we counteract fly cannibalism by placing the remaining living flies on fresh nutrition-less agar every 24 hours.

In sum, the diet treatment results are interesting because they indicate two different methods of SR – flies on the HS diet might have more fat stores to access during the period of starvation, while flies on the DR diet might have allocated the limited resources, they did have access to towards survival – an excess of wealth compared to a survival tactic. Female flies resisted starvation longer regardless – if a fly was male, they were three times as likely to die at any given moment. In both diet treatment and direct measurement flies, an additive Cox proportional hazard model of three terms fit our data best. This gives support to selection treatment playing a significant role in starvation resistance.

The phenotypic variation we observed, and the variables that our Cox, Kaplan-Meier and AIC models indicated as significantly different in predicting SR, reinforces the conclusions of previous research. Starvation resistance did co-evolve alongside cross-generational variation in resource availability. Flies from certain diets and selection regimes survived for longer periods of time than others. Selection on certain resource allocation strategies confers resistance to starvation, just as it confers longevity and other forms of stress resistance (Harshman *et al.* 1999). In the future, we will address the underlying genetic mechanisms of resistance to starvation in our resequenced fly population.

Previous studies have found that flies selected for starvation resistance exhibit a concurrent increase in body size (Brown *et al.* 2019). Flies from the CH selection regime are much larger than flies from the DA and FA selection regimes. The CH flies resisted starvation for significantly longer periods of time than the other selection regimes in both direct measurement and diet treatment paradigms, as well. Even in the diet treatment flies, those placed on an HS diet prior to the SR assay resisted starvation for significantly longer periods of time than flies placed on the control or DR diets. This makes sense, as increased body size has been shown to be a fitness-related trait conferring stress resistance (Ewing 1961).

Other studies show that animals develop starvation resistance by reducing their energy expenditure – a 'thrifty' resource allocation pattern whereby animals store energy in the form of lipids (Hoffmann and Parsons 1989a; Marron *et al.* 2003; Rion and Kawecki 2007; Aggarwal 2014). This coincides nicely with our observed results, where flies from the FA selection regime resist starvation longer than flies from the DA selection regime. We suspect that these flies have a 'thriftier' resource allocation strategy, where fluctuating availability of resources leads to lower metabolic rates and higher rates of lipid storage when access to resources is more abundant.

However, previous research from our lab suggests that plastic nutrient allocation strategies under different diets in flies do not affect crucial parts of

nutrient signaling pathways like the insulin/TOR signaling pathway (Ng'oma *et al.* 2020). Phenotypic plasticity in resource allocation strategies, and related traits such as starvation resistance, are not solely due to allele frequency differences in metabolism and nutrient-sensing mechanisms. Therefore, the actual mechanism of genetic control underlying resource allocation and starvation resistance are still unknown.

Our next steps are to resequence the evolved population of *D. melanogaster* to find allele and haplotype frequency differences between our unevolved and evolved populations. We will also look for regions of significant differentiation between selection lines and diet treatment groups with high and low resistance to starvation, which will give us insight into possible genes of interest underlying SR. We plan to further inform our observed phenotypic differences with our genotypic analyses. Significant genetic differentiation between phenotypic extremes, for example, can be reasonably attributed to SR. This will further elucidate the genetic mechanism by which SR is controlled.

LITERATURE CITED

Aggarwal D. D., 2014 Physiological basis of starvation resistance in Drosophila leontia: analysis of sexual dimorphism. J. Exp. Biol. 217: 1849–1859.

Ahmad M., S. U. Chaudhary, A. J. Afzal, and M. Tariq, 2015 Starvation-Induced Dietary Behaviour in Drosophila melanogaster Larvae and Adults. Sci. Rep. 5: 14285.

Ahmad M., E. S. Keebaugh, M. Tariq, and W. W. Ja, 2018 Evolutionary responses of Drosophila melanogaster under chronic malnutrition. Front Ecol Evol 6. https://doi.org/10.3389/fevo.2018.00047

Andersen P. K., and R. D. Gill, 1982 Cox's Regression Model for Counting Processes: A Large Sample Study. Ann. Stat. 10: 1100–1120.

Baldwin-Brown J. G., A. D. Long, and K. R. Thornton, 2014 The Power to Detect Quantitative Trait Loci Using Resequenced, Experimentally Evolved Populations of Diploid, Sexual Organisms. Mol. Biol. Evol. 31: 1040–1055.

Bass T. M., R. C. Grandison, R. Wong, P. Martinez, L. Partridge, *et al.*, 2007 Optimization of dietary restriction protocols in Drosophila. J. Gerontol. A Biol. Sci. Med. Sci. 62: 1071–1081.

Bradburn M. J., T. G. Clark, S. B. Love, and D. G. Altman, 2003a Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. Br. J. Cancer 89: 431–436.

Bradburn M. J., T. G. Clark, S. B. Love, and D. G. Altman, 2003b Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit. Br. J. Cancer 89: 605–611.

Brown E. B., M. E. Slocumb, M. Szuperak, A. Kerbs, A. G. Gibbs, *et al.*, 2019 Starvation resistance is associated with developmentally specified changes in sleep, feeding and metabolic rate. J. Exp. Biol. 222. https://doi.org/10.1242/jeb.191049

Chippindale A. K., A. M. Leroi, S. B. Kim, and M. R. Rose, 1993 Phenotypic plasticity and selection in Drosophila life-history evolution. I. Nutrition and the cost of reproduction. J. Evol. Biol. 6: 171–193.

Chippindale A. K., T. J. F. Chu, and M. R. Rose, 1996 COMPLEX TRADE-OFFS AND THE EVOLUTION OF STARVATION RESISTANCE IN DROSOPHILA MELANOGASTER. Evolution 50: 753–766.

Clark T. G., M. J. Bradburn, S. B. Love, and D. G. Altman, 2003a Survival analysis part I: basic concepts and first analyses. Br. J. Cancer 89: 232–238.

Clark T. G., M. J. Bradburn, S. B. Love, and D. G. Altman, 2003b Survival analysis part IV: further concepts and methods in survival analysis. Br. J. Cancer 89: 781–786.

Djawdan M., A. K. Chippindale, M. R. Rose, and T. J. Bradley, 1998 Metabolic reserves and evolved stress resistance in Drosophila melanogaster. Physiol. Zool. 71: 584–594.

Everman E. R., C. L. McNeil, J. L. Hackett, C. L. Bain, and S. J. Macdonald, 2019 Dissection of Complex, Fitness-Related Traits in Multiple Drosophila Mapping Populations Offers Insight into the Genetic Control of Stress Resistance. Genetics 211: 1449–1467.

Ewing A. W., 1961 Body size and courtship behaviour in Drosophila melanogaster. Anim. Behav. 9: 93–99.

Garland T., and M. R. Rose, 2009 Experimental evolution. University of California Press.

Harshman L. G., A. A. Hoffmann, and A. G. Clark, 1999 Selection for starvation resistance in Drosophila melanogaster: Physiological correlates, enzyme activities and multiple stress responses. J. Evol. Biol. 12: 370–379.

Herrewege J. van, and J. R. David, 1997 Starvation and desiccation tolerances in Drosophila: comparison of species from different climatic origins. Ecoscience 4: 151–157.

Hoffmann, and Parsons, 1989a An integrated approach to environmental stress tolerance and life-history variation: desiccation tolerance in Drosophila. Biol. J. Linn. Soc. Lond.

Hoffmann A. A., and P. A. Parsons, 1989b Selection for increased desiccation resistance in Drosophila melanogaster: additive genetic control and correlated responses for other stresses. Genetics 122: 837–845.

Hoffmann A. A., and L. G. Harshman, 1999 Desiccation and starvation resistance in Drosophila: patterns of variation at the species, population and intrapopulation levels. Heredity 83: 637–643.

Hoffmann A. A., R. Hallas, A. R. Anderson, and M. Telonis-Scott, 2005 Evidence for a robust sex-specific trade-off between cold resistance and starvation resistance in Drosophila melanogaster. J. Evol. Biol. 18: 804–810.

Kaplan E. L., and P. Meier, 1958 Nonparametric Estimation from Incomplete Observations. J. Am. Stat. Assoc. 53: 457–481.

King E. G., C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen, *et al.*, 2012 Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res. 22: 1558–1566.

King E. G., S. J. Macdonald, and A. D. Long, 07/2012 Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. Genetics 191: 935–949.

Lee K. P., and T. Jang, 2014 Exploring the nutritional basis of starvation resistance in Drosophila melanogaster, (G. Davidowitz, Ed.). Funct. Ecol. 28: 1144–1155.

Long A. D., S. J. Macdonald, and E. G. King, 2014 Dissecting complex traits using the Drosophila Synthetic Population Resource. Trends Genet. 30: 488–495.

Marron M. T., T. A. Markow, K. J. Kain, and A. G. Gibbs, 2003 Effects of starvation and desiccation on energy metabolism in desert and mesic Drosophila. J. Insect Physiol. 49: 261–270.

Mazerolle M. J., 2020 AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c)

McCue M. D., 2010 Starvation physiology: reviewing the different strategies animals use to survive a common challenge. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 156: 1–18.

Ng'oma E., P. A. Williams-Simon, A. Rahman, and E. G. King, 2020 Diverse biological processes coordinate the transcriptional response to nutritional changes in a Drosophila melanogaster multiparent population. BMC Genomics 21: 84.

Pijpe J., P. M. Brakefield, and B. J. Zwaan, 2007 Phenotypic plasticity of starvation resistance in the butterfly Bicyclus anynana. Evol. Ecol. 21: 589–600.

R Core Team, 2022 R: A Language and Environment for Statistical Computing

Rion S., and T. J. Kawecki, 2007 Evolutionary biology of starvation resistance: what we have learned from Drosophila. J. Evol. Biol. 20: 1655–1664.

Rose M. R., L. N. Vu, S. U. Park, and J. L. Graves Jr, 1992 Selection on stress resistance increases longevity in Drosophila melanogaster. Exp. Gerontol. 27: 241–250.

Schwasinger-Schmidt T. E., S. D. Kachman, and L. G. Harshman, 2012 Evolution of starvation resistance in Drosophila melanogaster:

measurement of direct and correlated responses to artificial selection. J. Evol. Biol. 25: 378–387.

Service P. M., E. W. Hutchinson, M. D. MacKinley, and M. R. Rose, 1985 Resistance to environmental stress in Drosophila melanogaster selected for postponed senescence. Physiol. Zool. 58: 380–389.

Sjoberg D. D., and M. Baillie, 2022 ggsurvfit: Easy and Flexible Time-to-Event Figures

Therneau T. M., 2022 *A Package for Survival Analysis in R.*

Vijendravarma R. K., S. Narasimha, and T. J. Kawecki, 2013 Predatory cannibalism in Drosophila melanogaster larvae. Nat. Commun. 4: 1789.

Wickham H., M. Averick, J. Bryan, W. Chang, L. McGowan, *et al.*, 2019 Welcome to the tidyverse. J. Open Source Softw. 4: 1686.

Zabor E. C., 2022 Survival Analysis in R

CHAPTER 4
CODE APPENDIX

Z. Forrest Elkins

Background

Pooled sequencing has become a cost-effective and accurate method of DNA sequencing. As a result, we lose individual genetic data identifiers. The sequenced DNA are a representative sample of the pool instead of the individual. While this method reduces sequencing cost without sacrificing accurate allele & haplotype frequency reads, it does introduce extra sources of statistical error. Since the DNA reads are a representative sample, we wind up with varying amounts of sequenced DNA reads along the genome. In addition, the resulting allele & haplotype frequencies are influenced by the formation of the sample.

Here, I establish a novel method of statistical error estimation due to varying amounts of coverage. Coverage is defined as the number of reads at a given location along the genome. Due to the law of large numbers, a higher sequencing coverage value leads to a more accurate allele estimation at that loci.

Simple pairwise comparisons

Simulate a pairwise comparison between alleles at a single location without error. Here, calculate the absolute allele frequency difference between segregated bulks. The allele frequencies for the 'high' and 'low' bulks are assumed to be true – i.e., there is no error in allele frequency estimation.

```
# load dependencies
library(ggplot2)
library(tidyverse)
library(purrr)
library(cowplot)
set.seed(172452)
```

```
# establish pairwise parameters

## allele frequencies for high and low populations
popH <- 0.3
popL <- 0.09
## absolute difference in allele frequencies
diff <- abs(popH - popL)
## scale data logarithmically
logdiff <- -log10(diff)
## print pairwise comparison value -- 'true' value
logdiff
```

```
[1] 0.6777807
```

Estimating 'effective coverage'

A simulated calculation of coverage using the math detailed in Tilk et al. 2019. In other words, my own miniature HAF-pipe functions. Our variable `cvg_e` is the 'theoretical coverage at which binomial sampling of reads would be expected to contain

the observed amount of error from estimated frequencies.' Put another way, the estimated and theoretical root mean squared error (RMSE) rates will equal each other under 'effective coverage.'

This method provides a level of sequencing coverage that researchers should aim for in their genetic data to minimize error in their observed allele frequencies. It does not give an estimation of that error as a function of coverage.

```
# modulate pairwise comparison with coverage as a source of error

# n is the number of sites, which in our case is just one
n <- 1

## estimated and true allele frequencies at one site
AFtrue <- rbinom(100,100,0.38) / 100
AFest <- rbinom(100,100,0.43) / 100

## effective coverage -- Tilk et al. 2019
cvg_e <- sum(AFtrue * (1 - AFtrue)) / sum((AFest - AFtrue) ^ 2)
cvg_e
```

```
[1] 30.42092
```

```
## theoretical root mean squared error
RMSEthe <- sqrt( sum(AFtrue * (1 - AFtrue)) / (cvg_e * n) )
RMSEthe
```

```
[1] 0.8772115
```

```
## estimated RMSE
RMSEest <- sqrt( sum((AFest - AFtrue) ^ 2) / n)
RMSEest
```

```
[1] 0.8772115
```

In this case, with the code above, error introduced by variation in coverage would be evaluated via the magnitude of difference between theoretical and estimated RMSE values.

> Tilk, Susanne, Alan Bergland, Aaron Goodman, Paul Schmidt, Dmitri Petrov, and Sharon Greenblum. 2019. "Accurate Allele Frequencies from Ultra-Low Coverage Pool-Seq Samples in Evolve-and-Resequence Experiments." *G3* 9 (12): 4159–68.

Coverage and pairwise allele frequency comparisons

The following is a novel method of estimating error as a function of coverage. Simulate a pairwise comparison between alleles at a single location with coverage error. To do this, we use a random binomial distribution that takes in coverage and true allele frequency and outputs the estimated allele frequency.

First, we simulate a single allele frequency estimation given coverage and true allele frequency:

```
# how to simulate coverage
## rbinom(1, cvg, AFtrue) / cvg
## output is estimated AF

cvg <- 20
trueAF <- 0.43

estAF <- rbinom(1,cvg,trueAF) / cvg
estAF
```

```
[1] 0.4
```

Next, we establish a function that calculates estimated allele frequency when it is given coverage and true allele frequency:

```r
# function calculating estimated allele frequency
estAF <- function(cvg,trueAF){
  est <- rbinom(1,cvg,trueAF) / cvg
  return(est)
}
```

Finally, we can scale our code and plot the results:

```r
# run this n times, scale up
n <- 1000
tru <- 0.43
cvg <- sample(2:150,n,replace = TRUE)
af <- tibble(
  "cvg" = cvg,
  "trueAF" = rep(tru,times=n)
)

af$estAF <- af %>% pmap(estAF) %>% unlist()

# plot estimated allele frequency by coverage
plt <- af %>% ggplot(aes(x = estAF, y = cvg)) +
  geom_point(alpha = 0.15,colour="red") +
  geom_smooth(method = "gam", colour = "red",se=FALSE) +
  geom_vline(xintercept=tru, size = 0.8,linetype = "dashed") +
  labs(x = "Estimated allele frequency", y = "Coverage") +
  annotate("label",x = tru, y = -50, label = "True allele frequency") +
  theme_cowplot()
plt
```

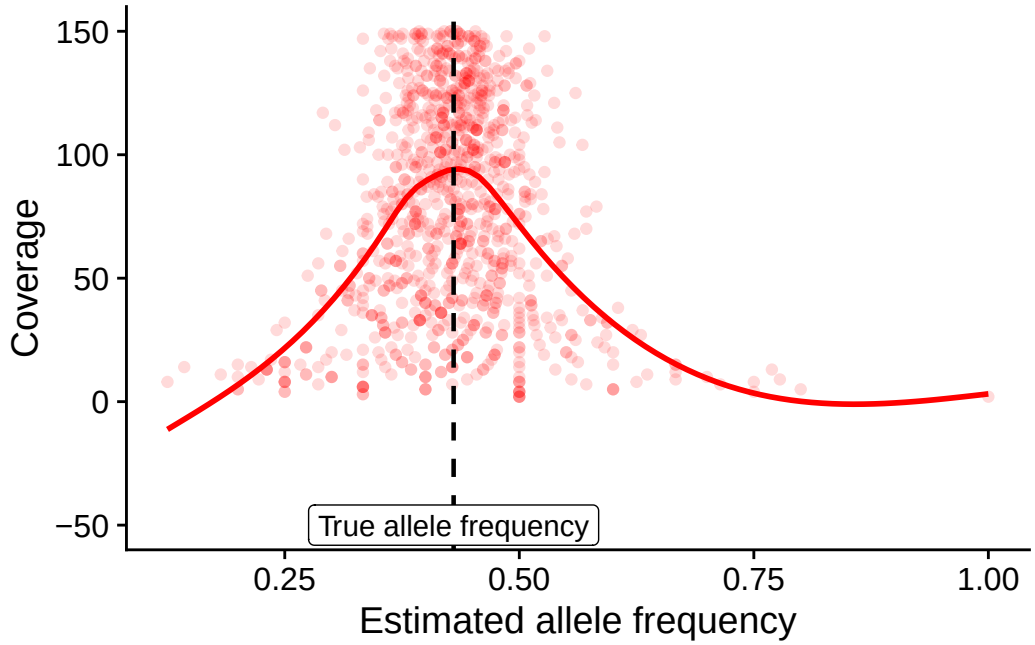


Figure 1: Variability in allele frequency estimations due to coverage.

We can see the level of coverage and its effect on allele frequency estimation in the plot above.

Simple plot of random allele frequencies

Pull allele frequencies from a distribution for 'low' and 'high' poolseq populations and plot them.

```
## the following code is from my professional website
expl_freqs <- readRDS('expl-freqs.Rds')
arm <- expl_freqs %>%
  filter(chrom=='2L') %>%
  slice(1:800)

D <- (sample(400:600,800,replace = TRUE))/1000
N <- (sample(300:600,800, replace = TRUE))/1000

sampledata <- tibble(
  "pos" = arm$pos,
```

```
  "N" = N,
  "D" = D
)

plt <- ggplot(data=sampledata, aes(pos/16)) +
  geom_line(aes(y = N, colour = "non-explorers")) +
  geom_line(aes(y = D, colour = "explorers"))
plt + labs(x = "Position (Mb)")
```

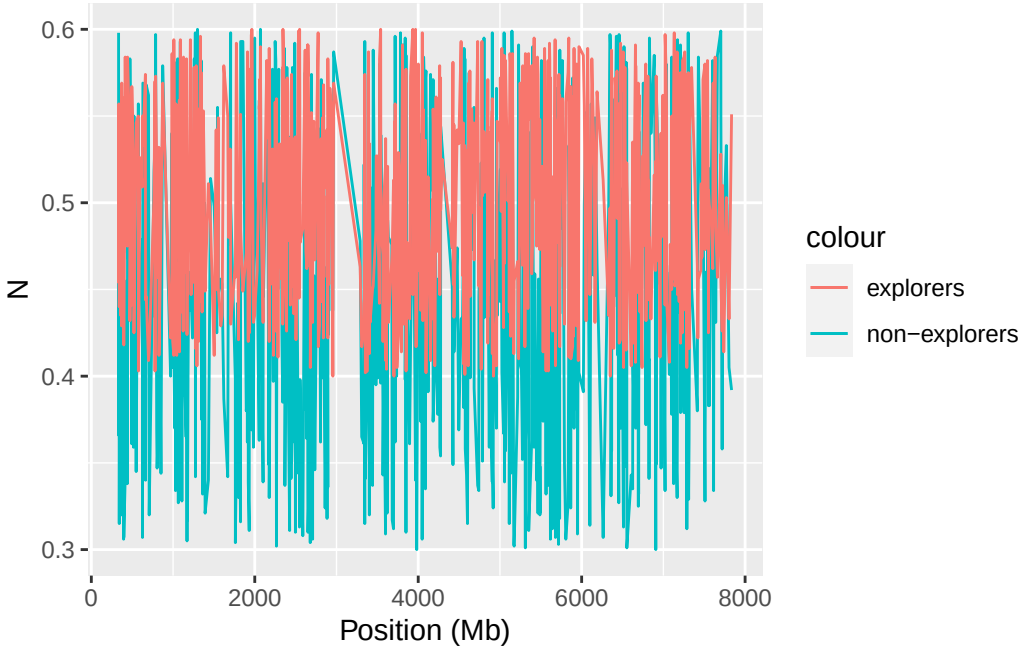


Figure 2: Allele differences between exploring and non-exploring flies.

VITA

Z. Forrest Elkins was born 01 June 1993 in Columbus, Ohio to Jodi and Stephen Elkins. They graduated from St. Joseph-Ogden High School in 2012, and thereafter attended Drake University in Des Moines, Iowa. After earning a B.S. in Computer Science and a B.S. in Psychology in 2016, Zoe was accepted into the Ph.D. program in the Division of Biological Sciences at the University of Missouri-Columbia. They started under the advisement of Dr. Lorin Milescu, and in 2017 became a mentee of Dr. Elizabeth G. King. For their dissertation research, Zoe focused on the genetic basis of exploration and starvation resistance in the common fruit fly, *D. melanogaster*. They developed and utilized open-source statistical pipelines written in R and python to analyze and visualize large datasets of genomic data. While Zoe's next steps are to be determined, they are excited to continue their journey as a lover of all data (dataphile). They look forward to expanding and deepening their statistical toolkit.